



EDITORIALE

Caro lettore,

*tra poco **Mondo Digitale** compie dieci anni di vita, durante i quali riteniamo abbia svolto un significativo ruolo culturale nel nostro Paese. Alla rivista viene riconosciuta una precisa fisionomia nel panorama editoriale, come veicolo di divulgazione non banale dell'informatica e delle sue applicazioni. In effetti, ogni articolo vuole essere una breve, ma compiuta monografia del tema trattato, con una esposizione rigorosa, ma accessibile anche ai non specialisti.*

Questa formula potrebbe rimanere inalterata anche per i prossimi anni, ma riteniamo sia opportuno cogliere l'occasione del decennale per un ripensamento.

Il motivo lo offre la stessa tecnologia di cui si occupa la rivista, con la straordinaria e tumultuosa evoluzione cui stiamo assistendo.

La profetizzata "fine della carta" è evento ancora irrealistico, ma certamente è in atto una rivoluzione di cui siamo tutti partecipi e che ha molteplici volti, dal web ai tablet agli e-book.

Proprio tenendo conto di ciò, abbiamo pensato di reimpostare la rivista passando a un supporto digitale, valorizzando le opportunità che tale transizione ci consente. In altre parole, l'obiettivo è di mantenere tutti i contenuti che caratterizzano l'attuale rivista, arricchendoli però con quanto il supporto cartaceo non è in grado di offrire. Solo a titolo di esempio, con questa soluzione si potranno mettere a disposizione del lettore tutte le relazioni, i video e le interviste dei più importanti convegni organizzati periodicamente dalla Associazione.

L'evoluzione prospettata non cambia ovviamente l'impostazione generale della rivista, basata su articoli di rassegna di elevato livello, nonché la sua cadenza trimestrale. Ogni nuova uscita sarà segnalata al lettore, oltre che per mail, anche con un pieghevole illustrativo del numero.

*La nuova serie di **Mondo Digitale** sarà disponibile sul sito: <http://mondodigitale.aicanet.net>. Saranno automaticamente **abilitati all'accesso tutti coloro che ricevono l'attuale rivista**.*

*"Panta rei, tutto scorre" – diceva Eraclito – "Non ti puoi bagnare due volte nella stessa acqua del fiume". Anche senza scomodare l'antico filosofo greco, possiamo tutti concordare sul fatto che il cambiamento è connaturato con l'esistenza. A questo principio non si sottrae **Mondo Digitale**, senza però perdere la connotazione culturale che l'ha sinora contraddistinto.*

Franco Filippazzi
Direttore di Mondo Digitale

Giulio Occhini
Direttore di AICA



LA NUOVA BANDA LARGA VIAGGIA SUL FOTONE

Giancarlo Prati

La fotonica è una delle cinque tecnologie abilitanti che l'Europa ha identificato per cambiare il quotidiano dei suoi cittadini e sviluppare la propria competitività a livello globale. Nell'ICT il ruolo della fotonica, già molto significativo, è destinato a diventare pervasivo all'aumentare delle esigenze di banda nelle reti di comunicazione. In questo articolo si passano in rassegna i punti chiave del presente e del futuro inerenti a questa tecnologia per le telecomunicazioni e al ruolo che essa gioca e giocherà nella rete a banda larga di oggi e di domani.

1. Introduzione

Gli sviluppi della fotonica nei decenni futuri porteranno cambiamenti radicali nel mondo della comunicazione e dell'ICT in generale. Essa contribuirà a realizzare sistemi di comunicazione e un'infrastruttura di rete anche nell'area di utente a minore consumo energetico ("*green communication*") e con capacità in grado di fronteggiare qualunque richiesta di traffico oggi prevedibile nelle reti dorsali di trasporto ("*core*" o "*backbone*" o WAN – *Wide Area Network*) con decine di Terabit al secondo, nelle reti metropolitane ("*metro*" o MAN – *Metropolitan Area Network*) con vari Terabit al secondo, nelle reti di accesso con molti Gigabit al secondo, consentendo servizi di alta sofisticazione e qualità in tutte le aree immaginabili di attività. E' nella larghissima banda sia del mezzo trasmissivo, la fibra ottica, che del crescente numero e tipo di componenti che usano fotoni a frequenza infrarossa per l'elaborazione direttamente nel dominio ottico, il punto dove il collo di bottiglia della banda limitata dell'elettronica viene superato. Rispetto a questa, poi, risulta ridotto il consumo energetico per i meccanismi stessi di gestione e interazione dei



flussi di luce infrarossa e la riduzione delle operazioni di conversione da ottico a elettronico e viceversa (conversioni O-E-O), già di per sé costose come investimento in apparati.

La fotonica ha in realtà radici lontane, prende le mosse dal laser a rubino di Thomas Maiman del 1960, dal primo studio sulla fibra ottica di Charles Kao nel 1966, e dagli amplificatori ottici in fibra drogata degli anni '80, in cui la Pirelli con l'Università di Southampton ha giocato un ruolo decisivo. Senza contare che i fotorivelatori, componenti essenziali dei moderni sistemi di comunicazione ottica, hanno nel tubo foto-moltiplicatore degli inizi del secolo scorso il loro antenato. Una sorgente, un canale trasmissivo, un amplificatore, un fotorivelatore: a partire da questi quattro elementi fondamentali per la comunicazione punto-punto ci allarghiamo qui in un *excursus* su dove l'evoluzione della tecnologia fotonica ci sta portando nelle reti di comunicazione, sia nelle dorsali di trasmissione di dati a lunga distanza che nelle reti metropolitane, elementi fondamentali dell'Internet di oggi e di quella del futuro.

Un minimo cenno va al fatto che l'Internet del futuro sarà anche l'Internet delle Cose e che la fibra ottica, oltre che da mezzo di comunicazione, potrà fungere da sensore distribuito, efficace mezzo di monitoraggio della temperatura o dello *stress* e deformazioni di strutture, con rilevanti applicazioni nel campo della sicurezza, dell'energia e dei trasporti. Un altro aspetto dove la fotonica giocherà un ruolo rilevante, ma che di nuovo qui solo accenniamo, è nelle interconnessioni degli apparati, delle schede e dei *chip* che li compongono, ancora una volta per risparmiare energia, diminuire l'ingombro e aumentare la capacità dei collegamenti.

Infine, se la fotonica è avviata nella sua evoluzione sulle orme dell'elettronica, l'integrazione tecnologica dei suoi componenti fondamentali in circuiti fotonici integrati è certamente un passo ormai prossimo. Anzi, la prospettiva più attraente e promettente è proprio l'integrazione completa di circuiti fotonici ed elettronici su un medesimo *chip*. La prospettiva di una fotonica integrata compatibile con le tecnologie attualmente utilizzate su larga scala per l'elettronica e basate su silicio (la cosiddetta "*silicon photonics*") appare oggi concreta e di prossima implementazione, con enormi potenzialità e impatto sul mercato non solo dell'ICT ma anche nel campo dell'energia, dei trasporti, della sicurezza, della biofotonica e della sensoristica in generale.

In questo articolo vengono presentati alcuni aspetti, sia tecnici che tecnico-economici e di visione, del ruolo della fotonica nelle comunicazioni a banda larga del futuro. Dopo un breve richiamo di come la fotonica partecipi alla configurazione dei vari segmenti di rete, si toccherà l'aspetto dell'altissima capacità nel segmento di trasporto, con alcuni argomenti di attualità e qualche considerazione tecnica sulla comunicazione e commutazione fotoniche. Un paragrafo è dedicato alla rete di accesso in fibra ottica con particolare riferimento alla tecnica emergente delle reti ottiche passive a



divisione di lunghezza d'onda. Infine alcuni cenni sono dedicati all'aspetto dei circuiti fotonici integrati, percorso avviato ovunque a livello mondiale e che promette di costituire il futuro vero della fotonica e della rete.

2. La fotonica nella rete

Una tipica rete è costituita da diversi segmenti nei quali la fotonica è presente con soluzioni diverse: il segmento di accesso di utente è collegato ad una o più reti metropolitane (*metro*, *MAN*), per le quali spesso si distingue una rete metro di prossimità (*metro-edge*), a sua volta connessa ad una rete metro di raccolta (*metro-core*), che confluiscono nella rete dorsale (*core* o *backbone*) di trasporto a livello nazionale o globale. Le funzioni di ciascun segmento, e di conseguenza i relativi fondamentali di progetto, sono ciò che determina il ruolo della fotonica con le sue differenti soluzioni nei diversi segmenti.

Il segmento del *backbone* (*WAN*), tipicamente con collegamenti punto-punto tra nodi di una rete magliata (*mesh*), opera il trasferimento del traffico su fibra ottica con alta gerarchia tra regioni, paesi e continenti. Per fronteggiare il continuo aumento delle esigenze di capacità trasmissiva, la fotonica ha messo in campo da tempo la tecnica WDM (*Wavelength Division Multiplexing*) per la trasmissione su fibra ottica a propagazione monomodale. Ogni segnale elettrico, eventualmente aggregato con altri, viene "assegnato" a modulare una portante ottica ad infrarosso di una specifica lunghezza d'onda (intorno ai 1550 nm, la cosiddetta 3^a finestra a bassa attenuazione della fibra) prodotta da uno di *N* laser. Le *N* portanti a diverse lunghezze d'onda così modulate sono poi inviate insieme sulla fibra in una griglia frequenziale a spaziatura costante e standardizzata (50, 100, 200 GHz), mentre in ricezione sono separate tramite filtraggio ottico prima delle rispettive fotorivelazioni necessarie a recuperare i segnali elettrici originari. Gli apparati oggi installati presentano *N* portanti ottiche (denominate in gergo *colori*) che rappresentano a tutti gli effetti *N* canali di comunicazione multiplati in lunghezza d'onda (WDM), oggi prevalentemente del tipo *N*×10Gb/s con progressiva penetrazione del *N*×40Gb/s e con la prospettiva a medio periodo dei 100 Gb/s sulla singola lunghezza d'onda. Per la rete di *backbone*, i *driver* di progetto delle soluzioni fotoniche sono la diminuzione del costo per bit trasmesso e commutato, la lunghezza di tratta (inframmezzata da alcuni amplificatori ottici in fibra drogata all'Erbio) senza rigenerazione elettronica (*reach*), e la possibilità di scalabilità progressiva nel tempo per contenere gli alti costi di investimento tipici di queste soluzioni.

Il segmento *metro-core*, tipicamente configurato con fibre ottiche ad anello (*ring*) per l'interconnessione tra centrali, è tecnologicamente simile al segmento di *backbone*, supporta il WDM, prevalentemente nella versione *N*×2.5Gb/s e *N*×10Gb/s, e gestisce interi flussi di dati da un *host* trasmittente ad uno ricevente (*connection-oriented*). I *driver* di progetto sono la capacità, l'efficienza nell'aggregazione del traffico, l'affidabilità, la flessibilità, la riconfigurabilità e il basso consumo energetico. Anche qui un

tema di interesse per il prossimo futuro è il WDM con capacità di 100 Gb/s su singola lunghezza d'onda, punto di contatto questo con le prospettive del segmento di *backbone*, anche attraverso la soluzione 100Gb Ethernet, di cui si dirà nel seguito.

Il segmento della rete metro in prossimità all'accesso (*metro-edge*) è una rete in fibra che supporta l'interconnessione di diversi PoP (*Point of Presence*) dei vari *Service Providers*, e che può essere di topologie varie (*ring, bus, mesh*), con capacità aggregata che varia a seconda delle esigenze dal WDM alla singola portante, in questo caso prevalentemente di 1 Gb/s piuttosto che 10 Gb/s. Qui il progetto, e quindi la fotonica del sistema di comunicazione, ha come obiettivo il basso costo e il basso consumo energetico. Un altro aspetto desiderabile è la coesistenza della fotonica e dell'elettronica a diversi livelli per rispondere ad una grande varietà di esigenze.

Il segmento di accesso, infine, laddove lo si pensi in fibra ottica per raggiungere l'abitazione dell'utente (FTTH – *Fiber-To-The-Home*), o per raggiungerne la prossimità (FTTB – *Building*, FTTC – *Curb*), si avvia ad essere ampiamente basato sulla tecnologia PON (*Passive Optical Network*), con le soluzioni architetture di oggi quali P2P (fibra ottica per ogni utente fino alla centrale in sostituzione del doppino), EPON (Ethernet PON) e GPON (Gigabit PON) e di domani WDM-PON. Di tali soluzioni e dei rispettivi vantaggi e svantaggi si parlerà ampiamente nel seguito.

Nella pratica odierna le reti PON servono decine di destinazioni entro alcuni chilometri, e fino a 10-20 km nei casi più avanzati, con tecnica TDM (*Time Division Multiplexing*), a valle di uno *splitter* passivo in potenza del segnale sulla fibra primaria, che viene ramificata in venti-trenta fibre. Le tipiche velocità del GPON sono di 2.5 Gb/s complessivi verso l'utenza (*downstream, downlink*) e 1.25 Gb/s complessivi nella direzione opposta (*upstream, uplink*), per l'EPON di 1.25 Gb/s simmetrici nelle due direzioni. Per il futuro prossimo, le WDM-PON, su cui ci soffermeremo, offrono maggiore capacità e maggiore flessibilità di utilizzo. Anche in questo caso, basso costo, ridotto consumo e alta scalabilità sono i fondamentali del progetto anche della componente fotonica, mentre importante si presenta la loro capacità di supportare il crescente accesso disaggregato di diversi operatori al supporto trasmissivo (*unbundling*).

3. La comunicazione ad altissima capacità

A livello mondiale, l'aumento di capacità trasmissiva in generale, e sulla rete *core* in particolare, è determinato sostanzialmente dall'aumento delle esigenze di comunicazione tra IX (*Internet Exchanges*) e ISP (*Internet Service Providers*). A questi, a loro volta, affluiscono flussi crescenti derivanti dal diffondersi di FfTx, e determinati da applicazioni varie quali l'uso di Internet, il commercio elettronico, l'IPTV, il VoIP, la parte cablata delle comunicazioni *wireless*, il *Video-on-Demand*, la collaborazione *online*, l'immagazzinamento di dati in rete, ecc. L'Internet Exchange più trafficato del mondo, quello di Amsterdam (AMS-IX), ha raggiunto nel marzo 2011 un picco di traffico di 1.3 Tb/s (Figura 1), laddove il picco era di poco superiore

a 600 Gb/s nel gennaio del 2009, più che raddoppiandosi quindi in circa due anni.

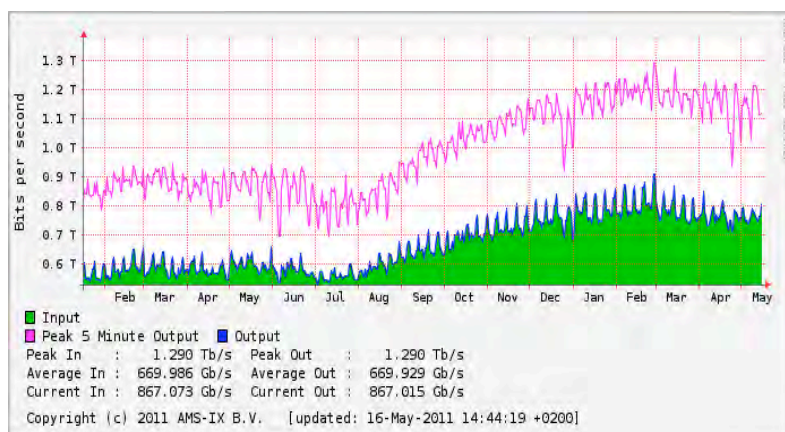


Figura 1

Andamento del traffico nel 2010-2011 presso l'Internet Exchange di Amsterdam (per gentile concessione di The Amsterdam Internet Exchange - AMS-IX)

Una crescita delle esigenze di capacità del 30%-50% annuo (al 30% si raddoppia in circa 2 anni e mezzo), come si può ipotizzare, porterà in breve alla necessità di collegamenti da 5-10 Tb/s, come mostrato in Figura 2, dove si individua anche il punto di saturazione dei sistemi WDM disponibili oggi ($N \times 10\text{Gb/s}$ e $N \times 40\text{Gb/s}$) e domani ($N \times 100\text{Gb/s}$). Vi sono anche proiezioni che indicano in 20 Tb/s la capacità per fibra necessaria nel 2016, presupponendo di fatto un incremento del 75% all'anno (raddoppio ogni 15 mesi).

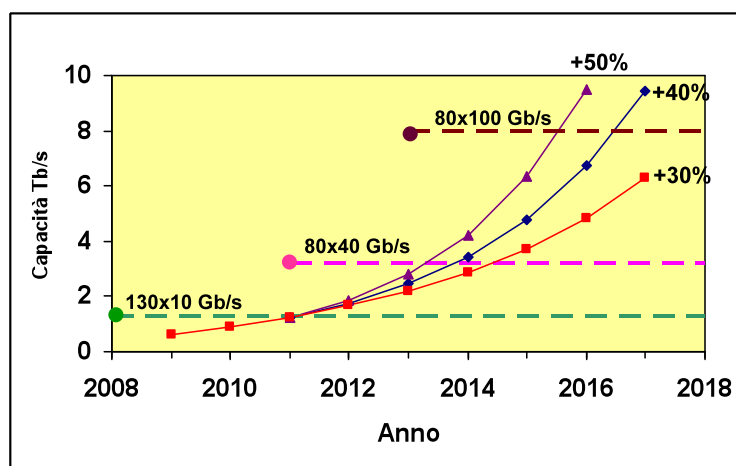


Figura 2

Proiezione delle esigenze di capacità nei collegamenti al variare del tasso annuo di crescita ed effettiva capacità dei sistemi WDM (elaborazione su dati web)

L'evoluzione della tecnica WDM per la lunga distanza negli anni ha seguito due direzioni al fine di accrescere la capacità complessiva: l'una, consistente nel moltiplicare un sempre maggior numero N di canali, spaziando al minimo le rispettive portanti ottiche (*Dense WDM*) per sfruttare al massimo la banda degli amplificatori ottici in fibra drogata¹, l'altra, nell'aumentare la velocità di trasmissione sulla singola portante. I sistemi WDM dell'attuale generazione commerciale per il trasporto offrono, come si è detto, 10 Gb/s e 40 Gb/s per canale, con N che può raggiungere il valore di 130 e di 80 rispettivamente. L'approccio frequente nella pratica è di partire da un numero limitato di schede installate per poi procedere con l'espansione man mano che le esigenze si accrescono.

Cominciano a presentarsi commercialmente sistemi WDM con capacità di 100 Gb/s per ogni lunghezza d'onda, finalizzati anche al supporto della 100Gb Ethernet, che si prevede si affermino rapidamente a partire dal 2013, nel quadro di una generale ripresa del mercato, già manifesta nei dati del 2010 con un aumento inaspettato del 160% del mercato del 10 Gb/s (Figura 3). Infatti, considerando un prevedibile ritmo di crescita delle esigenze del 30% annuo, un sistema WDM da 80×40Gb/s potenziali, che usi cioè al presente un minor numero di canali attivi, nel 2° giro di due-tre anni sarà già saturo (Figura 2). Questo lasso di tempo potrà più che raddoppiarsi operando un *upgrading* a 100 Gb/s per lunghezza d'onda, aumento che fortunatamente resta compatibile con le bande degli amplificatori ottici disposti lungo la fibra.

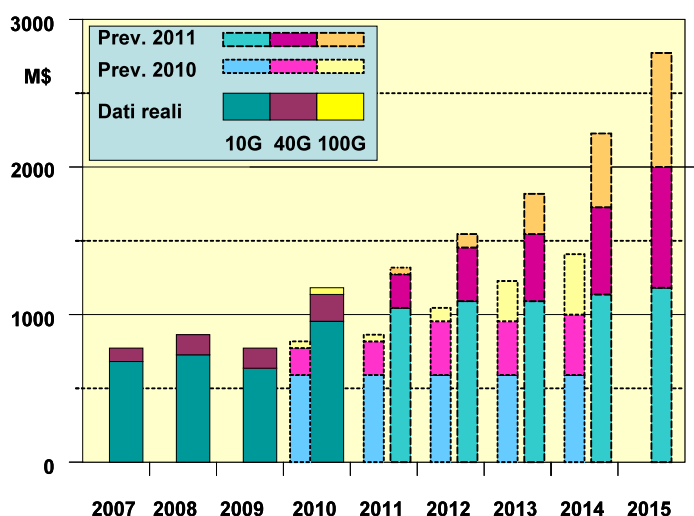


Figura 3
Mercato mondiale dei transceiver a 10, 40, 100 Gb/s (dati reali 2007-2010 e previsioni per il quinquennio successivo elaborate nel 2010 e nel 2011)
(elaborazione su dati web)

¹ Gli amplificatori ottici in fibra drogata con Erblio (EDFA) hanno una banda approssimativamente di 5 THz tra 1525 nm e 1565 nm circa (C-band) in corrispondenza della 3a finestra di bassa attenuazione (<0.2 dB/km) della fibra ottica.

A livello di sperimentazione, NEC (*Nippon Electric Company*) ha presentato nel 2010 un sistema a 640 canali da 107 Gb/s ciascuno, maggiore capacità complessiva per canali a 100 Gb/s. L'operatore ex-monopolista giapponese NTT, a sua volta, ha sperimentato su 240 km un sistema WDM con capacità da 69.1 Tb/s (massima capacità in assoluto, di poco superiore al precedente sistema NEC) realizzato in questo caso con 432 portanti modulate a 160 Gb/s ciascuna. In Figura 4 sono mostrate queste e altre prestazioni significative in funzione di capacità totale (Tb/s) e distanza (migliaia di km), riportandole sulle curve con prodotto capacità×distanza costante. Il valore più elevato di tale prodotto è stato raggiunto da Alcatel-Lucent (2010) con 115 Pb/s.km.

La tecnica di moltiplicazione in lunghezza d'onda per la generazione dello strato ottico della rete ha un punto di forza nel fatto che le interfacce WDM sono indipendenti dalla velocità e dalla modulazione, accettando ogni combinazione di velocità trasmissiva e di tipi di formato. A questo riguardo va sottolineato che il formato di modulazione delle portanti ottiche storicamente usato è stato quello di intensità, cioè una modulazione *on-off* dei laser, che consente una demodulazione diretta mediante un semplice fotodiodo, mentre negli ultimi due-tre anni, con l'aumentare delle esigenze nelle velocità trasmissive, si vanno ripresentando e affermando, dopo un primo interesse negli anni '90 per tutt'altro motivo², formati di modulazione a più alta efficienza spettrale che allocano cioè più di un bit al secondo per ogni Hertz di banda ma che richiedono tecniche di modulazione e demodulazione più sofisticate, come si vedrà nel seguito.

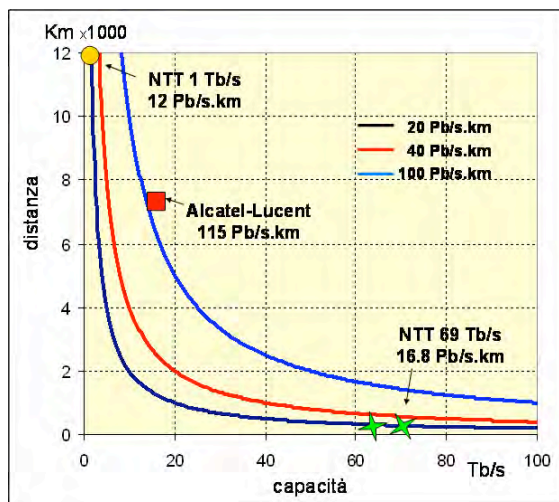


Figura 4

Prestazioni capacità x distanza dei migliori dimostratori (elaborazione da dati web)

² L'interesse per i formati a demodulazione coerente negli anni '90, quando l'amplificazione ottica era ancora agli albori, proveniva dall'allungamento della tratta in fibra per effetto del guadagno di sensibilità al ricevitore dovuto alla potenza immessa dal laser locale.

3.1 WDM e 100Gb Ethernet

La trasmissione sulla singola lunghezza d'onda su fibra ottica nelle reti di trasporto si è venuta storicamente sviluppando come una soluzione in grado di offrire maggiore velocità per merito della enorme banda del portante fisico a beneficio delle reti tradizionali basate sullo standard SDH/SONET (Europa: *Synchronous Digital Hierarchy*, USA: *Synchronous Optical Network*). Lo standard SDH/SONET che, come è noto, prevede di aggregare traffico a differenti velocità con tecnica TDM in un unico flusso sulla portante ottica, è stato concepito a suo tempo per una comunicazione a commutazione di circuito ed è in seguito evoluto a trasportare celle ATM, pacchetti IP e *frames* di Ethernet.

Ad oggi la domanda dominante di traffico è fra *router* IP e può eccedere i 100 Gb/s fra *router* adiacenti, esigenza per la quale la comunicazione su una singola lunghezza d'onda può non bastare, dato che la maggior parte dei collegamenti WDM installati ha capacità sulla singola lunghezza d'onda di 10 Gb/s e 40 Gb/s, come si è detto. Per ottenere un singolo collegamento IP a velocità di 100 Gb/s ("100Gb Ethernet") si opera un'aggregazione di diverse lunghezze d'onda denominata LAG (*Lambda Aggregation*) della griglia WDM. Questa situazione - la combinazione di capacità elevate e tipo dominante di traffico - determina esigenze che l'architettura a standard SDH non fronteggia efficacemente.

L'architettura della rete fotonica del futuro combina due aspetti: la diffusione del 100Gb Ethernet (100GbE) a standard IEEE 802.3 anche nella rete *backbone* per corrispondere al tipo dominante di traffico, e lo standard ITU-T OTN (*Optical Transport Network*) per la gestione efficiente dell'intera rete WDM con un alto livello di trasparenza per i *payload*. Il futuro della fotonica deve quindi assicurare la soluzione "100GbE over OTN". Lo standard OTN prevede pertanto la trasmissione di 100 Gb/s al fine di supportare la 100GbE, anche se la modalità tecnica nello strato fisico con cui questo avviene (WDM-LAG ovvero OFDM-*Orthogonal Frequency Division Multiplexing* ovvero SC-*Single Carrier*), non è precisata ed è lasciata, come di consueto, al costruttore.

La soluzione WDM-LAG per disporre di un canale a 100Gb/s, aggregando preferibilmente 10 flussi WDM paralleli da 10Gb/s, rientra nelle potenzialità tecniche odierne, ed è quanto effettivamente praticato sull'infrastruttura installata. Il mantenimento o il superamento di questo approccio nel futuro da parte di un operatore dipende da vari fattori: l'*hardware* disponibile, la distanza raggiungibile, gli aspetti di gestione delle varie lunghezze d'onda come unico flusso, l'accettabilità del costo per bit trasmesso e commutato che ne risulta. Storicamente la trasmissione seriale anziché parallela a velocità superiori, ha determinato una significativa diminuzione dei costi del bit trasmesso e commutato: nell'SDH l'aumento di un fattore 4 ogni volta, sino al raggiungimento dei 10 Gb/s, ha prodotto una diminuzione del costo del bit trasmesso e commutato di un fattore 2.2 circa, a parità di maturazione delle soluzioni di mercato. E' ancora da stabilire completamente l'effetto del passaggio ai 40 Gb/s, che ha avviato una



significativa diffusione nelle installazioni solo dal 2009, e che forse costituirà un gradino non altrettanto frequentato se l'ambiente Ethernet porterà l'affermazione del 100GbE in tempi brevi, come si prevede. In ogni caso sono a favore della singola lunghezza d'onda a maggiore capacità, rispetto alla soluzione LAG, la diminuzione del tempo di latenza e la gestibilità dei flussi a livello di ROADM (*Reconfigurable Optical Add-Drop Multiplexer*), che è l'apparato fotonico, di cui si dirà in seguito, preposto all'inserzione e all'estrazione di canali dalla griglia WDM in corrispondenza dei nodi della rete.

L'aumento della velocità di trasmissione seriale a 100 Gb/s sulla singola lunghezza d'onda rappresenta per la tecnica fotonica un campo in pieno sviluppo. L'accrescimento puro della velocità di trasmissione mantenendo la modulazione tradizionale di intensità *on-off* del laser determina problemi rilevanti sia nell'effetto degli *impairment* lineari (in particolare la dispersione cromatica) e non lineari (si deve aumentare la potenza di lancio per raggiungere accettabili distanze) che si producono nella fibra a tale velocità, sia nell'elettronica che si richiede nel sistema. Inoltre, al crescere in modo rilevante delle larghezze di banda dei canali WDM da immettere o estrarre aumentano le distorsioni introdotte dai ROADM e quindi diminuisce la distanza operativa.

Per queste ragioni si vanno affermando tecniche di modulazione e di demodulazione diverse da quelle tradizionali dei sistemi in fibra ottica e per formati di modulazione multilivello a più alta efficienza spettrale, che allocano più bit/s per ogni Hertz di banda, in particolare per la modulazione QPSK (*Quadrature Phase Shift Keying*). Il sistema di comunicazione usa fisicamente una sola lunghezza d'onda portante (SC), ma può essere interpretato come facente uso di due portanti in fase e quadratura (*IQ – In-Phase, Quadrature*) alla stessa lunghezza d'onda. Lo schema di principio di un ricevitore coerente per comunicazione ottica nel caso QPSK è mostrato in Figura 5, in cui sono evidenziate la parte "fotonica" e quella "elettronica". Nella pratica, sempre al fine di aumentare l'efficienza spettrale, si usano due coppie di portanti IQ con polarizzazioni tra loro in quadratura (*Dual Pol*), che raddoppiano la capacità del sistema sulla stessa lunghezza d'onda. In questo modo, 100 Gb/s si conseguono con velocità di simbolo di 25 GBaud su ogni polarizzazione, trasmettendo 2 bit/s per simbolo tramite il formato QPSK, spesso nella sua versione differenziale DQPSK.

Senza soffermarsi sulla sorgente laser, sull'amplificatore ottico e sul fotorivelatore, in quanto componenti canonici della comunicazione ottica *on-off*, altri componenti fotonici tipici che si usano per la comunicazione sulla singola lunghezza d'onda con formati multilivello e rivelazione coerente sono il modulatore *Mach Zehnder* (MZ) e il circuito ottico ibrido a 90°. Una sintetica descrizione del principio di funzionamento di questi due dispositivi è riportata nel riquadro 1, al fine di dare un'idea di cosa significhi usare fotoni anziché elettroni per l'elaborazione dei segnali.

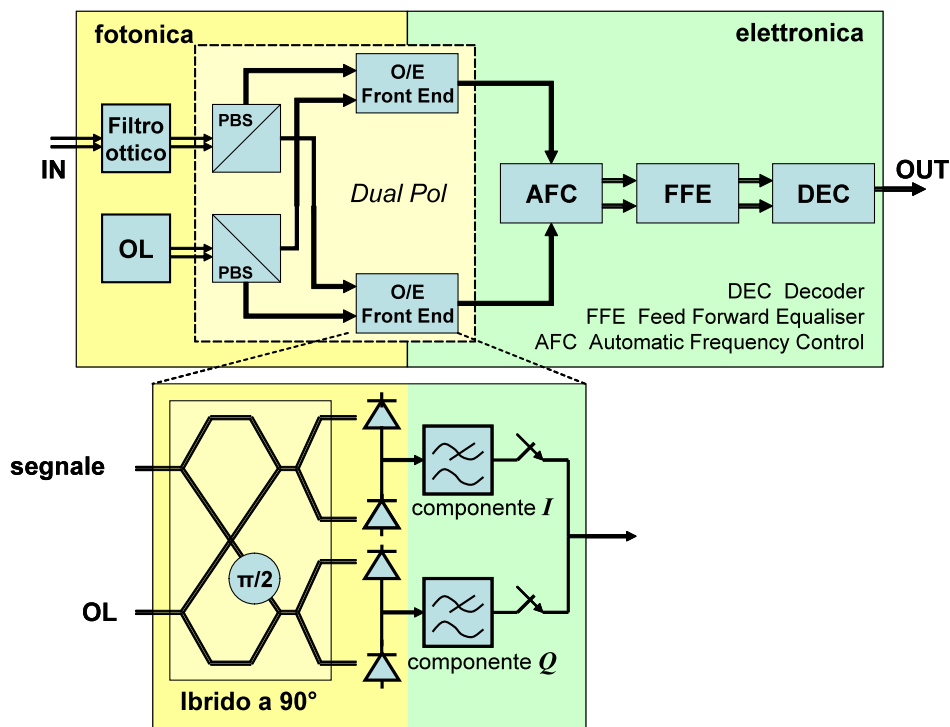


Figura 5

Schema di ricevitore coerente a diversità di polarizzazione

Dal punto di vista tecnico, la differenza basilare tra i sistemi a tradizionale modulazione di intensità e quelli con portanti IQ risiede nel fatto che per i primi il fotorivelatore al ricevitore, generando per sua natura una corrente proporzionale all'intensità del segnale che lo illumina, costituisce una non-linearità rispetto alle alterazioni di ampiezza che la fibra introduce sul segnale modulato trasmesso (l'intensità è il quadrato dell'ampiezza). Nei secondi invece, dove si richiede una rivelazione di tipo coerente che fa uso di un laser al ricevitore come oscillatore locale sommato al segnale della fibra, il fotorivelatore produce una corrente che risulta *linearmente* relazionata, in ampiezza e fase, al segnale ottico ricevuto. Questo permette di affidare all'elettronica post-rivelazione l'elaborazione del segnale necessaria a compensare gli effetti di distorsione della fibra ottica. E' questo un aspetto fondamentale per il conseguimento delle prestazioni volute.

Lo stesso risultato potrebbe ottenersi con i sistemi a modulazione *on-off* e rivelazione di intensità, ma solo con un *processing* nel dominio ottico, cioè *prima* del fotorivelatore, dove la complessità dell'elaborazione fotonica richiesta diventa tecnicamente proibitiva allo stato attuale. Gli altri aspetti positivi dei sistemi coerenti sono l'abbassamento della velocità dei simboli trasmessi (*Baud rate*) con conseguente minore banda richiesta dal canale e minori effetti della dispersione cromatica. Inoltre, l'elaborazione

3.2 Il supercanale al Terabit

L'ascesa verso il Terabit al secondo sulla singola lunghezza d'onda parte dal presupposto che si confermino gli aspetti positivi già citati rispetto alla soluzione LAG: il decremento del costo per bit trasmesso/commutato al crescere della velocità di trasmissione, la diminuzione del tempo di latenza, la gestibilità dei flussi a livello di ROADM. La realizzazione pratica si scontra intanto con i limiti della velocità dell'elettronica quando si pensi di aumentare direttamente la velocità di cifra. Allo stato attuale il gradino intermedio è costituito da 400 Gb/s, valore che di fatto recupera il fattore 4 di accrescimento dello standard SDH, sia pure applicandolo alla base di partenza dei 100 Gb/s necessari allo standard 100G Ethernet.

La comunicazione a 400 Gb/s rappresenta un'evoluzione tecnica delle soluzioni adottate per i 100 Gb/s: uso di formati multilivello (in particolare 16QAM – *Quadrature Amplitude Modulation*) con demodulazione coerente, moltiplicazione *Dual Pol*, efficace *post-processing* elettronico al ricevitore con qualche livello di parallelismo per agevolare l'elettronica, codifica a correzione di errore sul canale ad alto guadagno e bassa complessità (codici LDPC – *Low Density Parity Check*). La corrispondente velocità standardizzata da ITU-T è di 449.219 Gb/s, prevedibilmente sulla griglia standard con spaziatura a 100 GHz. Questi sistemi sono in fase di ricerca e sviluppo nei principali laboratori mondiali: si cita qui come esempio del contributo italiano che l'Istituto di Tecnologie della Comunicazione, dell'Informazione e della Percezione (TeCIP) della Scuola Superiore Sant'Anna di Pisa e il Laboratorio Nazionale di Reti Fotoniche del CNIT (Consorzio Nazionale Interuniversitario per le Telecomunicazioni) hanno sviluppato con Ericsson un dimostratore per un sistema a 448 Gb/s di frontiera che ha superato con successo *field trials* prolungati (Figura 6).



Figura 6
Dimostratore di sistema coerente a 448 Gb/s

Il supercanale a 1 Tb/s richiederà probabilmente soluzioni innovative sul piano sistemistico, diverse da un semplice *scaling* dei fattori qualificanti del 400 Gb/s, al fine di essere tecnicamente vincente per le prestazioni che se ne ricavano. In particolare, la distanza coperta efficacemente a questa velocità (*reach*) con le consuete potenze di lancio in fibra prima che sia necessaria una rigenerazione elettronica si riduce molto, il che porterebbe ad aumentare la potenza, ciò che determina l'insorgere di fenomeni non lineari con rapido degrado delle prestazioni. Il problema è aperto e oggetto di studio a livello mondiale, orientato generalmente ad un *packing* denso di sottocanali a più bassa velocità.

3.3 La commutazione fotonica: ROADM e OXC

Fra le operazioni sui flussi ottici necessarie per la gestione e riconfigurazione della rete da remoto, quella di inserire (*add*), estrarre (*drop*), commutare (*cross-connect*) canali che stanno transitando sulla fibra riveste carattere fondamentale. Il termine "commutazione fotonica" (*photonic switching*) individua nella pratica una capacità di operazioni al riguardo, che si viene via via evolvendo da versioni elementari con poca tecnologia realmente fotonica e gestione elettronica per l'*add* e *drop* dei canali nella griglia WDM verso sistemi in cui esisterà una matrice di commutazione fotonica integrata, in grado di effettuare la commutazione di flussi e, in futuro, di pacchetti tra le diverse portanti WDM.

Il ROADM (*Reconfigurable Optical Add-Drop Multiplexer*) costituisce il commutatore basilare che consente di intervenire in maniera dinamica nelle reti di *backbone*, e ora anche sempre più nelle reti *metro*, per l'inserimento/estrazione della completa informazione trasportata da una qualsivoglia portante WDM mantenendosi nel dominio ottico (Figura 7). L'elaborazione che avviene in un ROADM è un buon esempio dell'attuale livello di elaborazione fotonica dei segnali. In un tipico ROADM il segnale di ingresso WDM è dapprima suddiviso in diversi segnali alle diverse lunghezze d'onda da un AWG (*Arrayed Wavelength Grating*) router, in grado di instradare ogni lunghezza d'onda su una corrispondente fibra in uscita, e, nella direzione opposta, di assemblare le varie lunghezze d'onda in ingresso nell'unico segnale WDM in uscita (si veda il Riquadro 2). Dopo l'AWG uno *switch* fotonico per ogni lunghezza d'onda è in grado di provocare l'estrazione di quel segnale dal ROADM (*drop*), ovvero di farlo proseguire inalterato. Un secondo *switch* può inserire un nuovo flusso (*add*) a quella lunghezza d'onda, se libera, e un secondo AWG ripristina la trama WDM, il tutto senza nessuna conversione O-E-O da ottico in elettronico e viceversa (Figura 8).

I ROADM sono elementi importanti per reti con topologia a bus o ad anello, dove il problema è effettuare *add-drop* di un numero limitato di lunghezze d'onda. Se la rete è più complessa, come ad esempio a topologia magliata e con un elevato numero di nodi e di lunghezze d'onda, o deve gestire un elevato volume di traffico, l'elemento primario di rete diviene l'*Optical Cross Connect* (OXC). Questo ha la funzione, in un nodo in cui vi sia la terminazione di molte fibre, di aggiungere alla funzione di *add-drop* di canali la capacità di instradare flussi che non terminano in quel nodo su

altre fibre, ed eventualmente instradare flussi da lunghezze d'onda in arrivo ad altre in partenza, commutandoli da una qualunque delle prime (anche 80) a una di un certo numero di possibili seconde (solitamente una decina) su qualunque porta in uscita. Va notato che il termine “*optical*” per l'OXC si riferisce storicamente all'oggetto della propria azione, cioè alla sua funzione di commutazione di flussi ottici, e non alla tecnologia che effettivamente usa, ma il futuro della tecnologia dell'OXC è anch'esso sempre più “fotonico”, puntando a mantenere tutte le operazioni nel dominio ottico senza conversione O-E-O, operazione questa molto costosa che richiede la demodulazione e rimodulazione dei flussi.

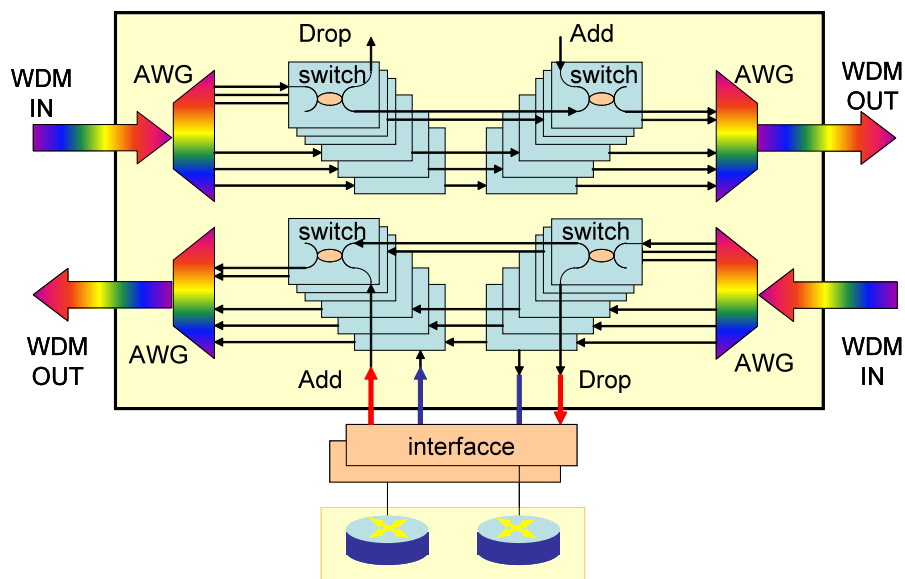


Figura 7

I ROADM nella rete per il controllo remoto di add&drop

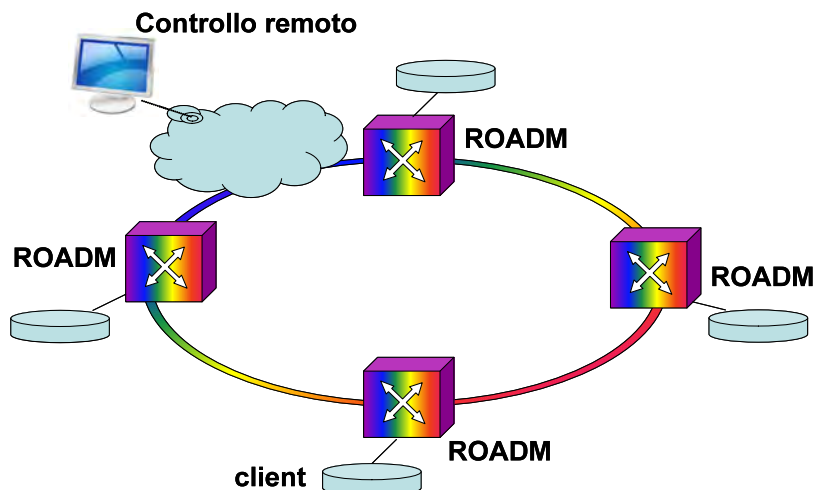


Figura 8

Configurazione tipica di modulo per ROADM

4. L'accesso in fibra ottica

Le soluzioni per l'accesso in fibra ottica, come rappresentato in figura 9, possono classificarsi in tre categorie: l'accesso punto-punto (P2P – *Point-To-Point*), la rete ottica passiva con accesso multiplo a divisione di tempo (TDM-PON, *Time Division Multiplexing - Passive Optical Network*) e la rete ottica passiva a divisione di lunghezza d'onda (WDM-PON). In ciascuna di queste è previsto per ogni utente un canale di andata dalla centrale all'utente (*downlink*) e uno di ritorno dall'utente alla centrale (*uplink*).

Dal punto di vista tecnico, nel caso P2P ogni utenza viene fisicamente collegata alla centrale con una propria fibra, che porta all'utente tutta l'informazione generata dal *transceiver* in centrale con basse perdite in potenza ottica (Figura 9a). Questa soluzione è costosa da vari punti di vista, in quanto richiede molta fibra, l'attestazione di N fibre in centrale con ripartitori ottici di grandi dimensioni per l'elevato numero di collegamenti e N *transceiver* in centrale, uno per ogni utente. Inoltre, in caso di interruzione del cavo ottico per lavori stradali, la problematica di dover operare un giunto di centinaia o migliaia di fibre non è certo un problema di poco conto.

Tuttavia, in regioni dove l'utenza non è molto dispersa, il maggior costo sulla fibra può essere contenuto e la soluzione a diversità spaziale favorisce, per i maggiori gradi di libertà, la coesistenza di diversi operatori sulla rete. Dal punto di vista logico questo approccio è in sostanza e con maggiore flessibilità ripreso e realizzato con la tecnica WDM-PON, che sfrutta appieno le potenzialità della fibra mantenendo notevoli gradi di libertà ed eliminando i difetti citati del P2P.

Nella soluzione TDM-PON, l'accesso avviene tramite uno *splitter* ottico passivo in potenza, che consente la ramificazione di una sola fibra primaria dalla centrale a partire da una certa posizione sul territorio, in diverse fibre dirette agli utenti (solitamente con topologia ad albero - ma sono possibile anche altre topologie), ciascuna delle quali porta in *downlink* la stessa informazione della fibra primaria (*broadcasting*). L'organizzazione dei dati è fatta con moltiplicazione a divisione di tempo (TDM) per servire individualmente ciascuno degli N utenti. Con questa soluzione vi è una sola fibra attestata in centrale e un solo *transceiver*, ma lo *splitter* passivo introduce un frazionamento della potenza e quindi una diminuzione della distanza di copertura tanto maggiore quanto più alto è il numero di utenti serviti. L'unico *transceiver* in centrale deve assicurare una velocità N volte quella che si destina a ciascun utente, il quale accede durante il proprio *slot* temporale al flusso complessivo aggregato che gli perviene. In *uplink* gli *slot* temporali assegnati ai vari utenti ricostruiscono una trama TDMA (*Time Division Multiple Access*) dopo il *combiner*, costituito dallo *splitter* operante in direzione opposta (Figura 9b). Con questa tecnica di accesso i gradi di libertà dei diversi operatori vengono notevolmente ridotti anche rispetto alla situazione attuale poiché deve esistere un operatore di riferimento per assegnare la trama TDM impiegata anche dai concorrenti. In termini regolamentari attuali non sarebbe più possibile il servizio cosiddetto di

unbundling che mantiene ai singoli operatori tutti i gradi di indipendenza e libertà, rimanendosi confinati al solo servizio oggi denominato di *wholesale* nel rapporto con l'operatore dominante. Per le due direzioni di trasmissione di *downlink* e di *uplink* si adottano due portanti ottiche a multiplazione di lunghezza d'onda ad ampia spaziatura, in cui il *downlink* usa una portante in 3^a finestra nell'intorno dei 1550 nm e l'*uplink* una in 2^a finestra intorno a 1300 nm. Vi sono due linee principali di realizzazione ad oggi di questa tecnica, la EPON (Ethernet PON) e la GPON (Gigabit PON) (si veda il Riquadro 3).

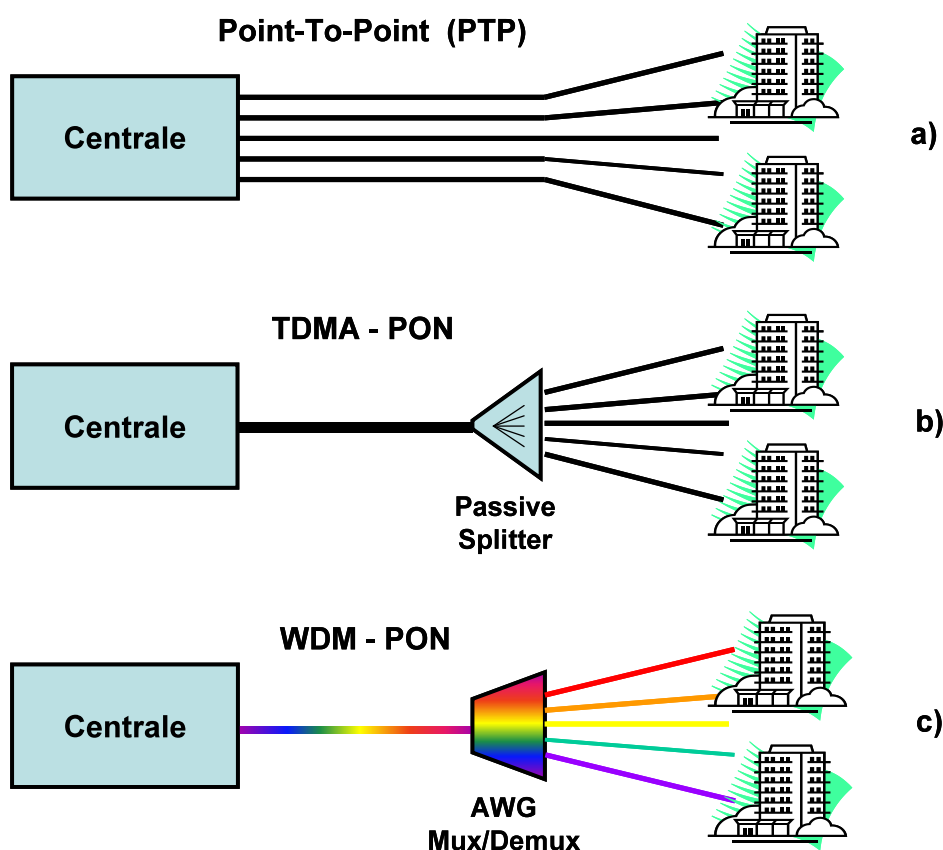


Figura 9
Topologie di reti di accesso

4.1 Le reti ottiche passive WDM-PON

La tecnica WDM-PON rappresenta il futuro delle reti ottiche per l'accesso ed è di fatto una soluzione con topologia ad albero come la TDM-PON, in cui però i segnali lanciati agli utenti dalla centrale sono multiplati in lunghezza d'onda anziché nel tempo. Al posto dello *splitter* ottico di potenza viene inserito ad un punto geograficamente conveniente un demultiplicatore (*demultiplexer*) AWG che instrada sulle fibre secondarie la sola lunghezza d'onda destinata a quell'utenza. Sull'*uplink* l'utenza

trasmette su una lunghezza d'onda assegnata e l'AWG, agendo ora inversamente come *multiplexer*, assembla le varie lunghezze d'onda nella trama WDM *upstream* per l'inoltro alla centrale (Figura 9c). La logica è simile al P2P, dove la moltiplicazione spaziale su più fibre è sostituita dalla moltiplicazione in lunghezza d'onda sul tratto della fibra primaria. Questa soluzione è attualmente in corso di sviluppo a livello mondiale e non è quindi ancora standardizzata, ma una volta fatto questo passo, sarà con ogni probabilità destinata ad una fetta crescente del mercato dell'accesso in fibra a partire dai prossimi quattro-cinque anni, legata al generale e costante espandersi della fibra stessa verso l'utente finale.

La struttura WDM-PON si basa su tre sottosistemi: una OLT (*Optical Line Terminal*) inserita in un PoP (*Point of Presence*) nella rete *metro* a monte, un AWG *demultiplexer* ottico passivo, una ONT (*Optical Network Termination*) presso ciascun utente a cui arriva una fibra, ovvero una ONU (*Optical Network Unit*) se l'unità terminale della fibra è collocata in un armadio nelle vicinanze dell'utente, che viene poi raggiunto con altra tecnica, ad esempio *wireless* (Figura 10).

La principale ragione dell'attenzione verso le WDM-PON risiede nel superamento dei limiti di banda che le TDM-PON presentano, in particolare sull'*uplink*, dove i vari utenti hanno a disposizione il solo proprio *slot* temporale dettato dalla OLT in centrale. Con la WDM-PON ad ogni utente viene assegnata una coppia di lunghezze d'onda, una per il *downstream* e una per l'*upstream*, che in certe soluzioni possono anche coincidere. La moltiplicazione di tutte le lunghezze d'onda del *downlink* è densa (*Dense WDM* - spaziatura di 50 GHz), così come quella delle lunghezze d'onda dell'*uplink*, mentre i due gruppi di portanti o sono moltiplicati tra loro in frequenza in finestre con separazione ampia o sono collocati nella stessa finestra con le stesse lunghezze d'onda, ma propagantesi in verso opposto l'uno rispetto all'altro lungo la fibra (contropropaganti). Tutte le lunghezze d'onda transitano sulla stessa fibra primaria nei due versi fino a che l'AWG *demultiplexer* (in un verso, e *multiplexer* nell'altro) non provvede a separare/riassemblare i segnali di ogni gruppo.

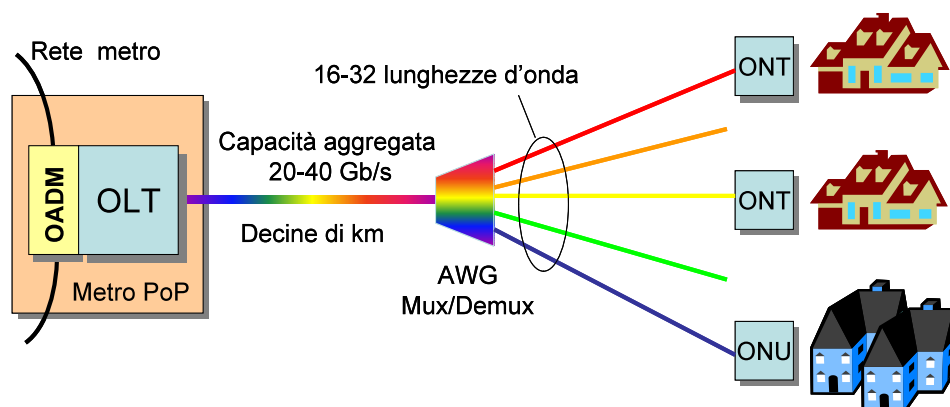


Figura 10
La rete ottica passiva WDM-PON

La flessibilità offerta da una soluzione WDM-PON è evidente: ogni utente può avere capacità diverse e sfruttarle in pieno, il numero di utenti può essere facilmente – entro certi limiti - aumentato, così come le distanze raggiungibili. La differenza sostanziale rispetto alle applicazioni WDM nate per le reti di trasporto risiede nel requisito di estrema economicità: il “modem ottico” della ONU dovrà avere elevate prestazioni, ma ad un costo comparabile con gli apparati ADSL.

Un aspetto rilevante per le WDM-PON, legato alla riduzione dei costi, è quello di individuare soluzioni che consentano di avere delle ONU tutte uguali (*colorless*), e non legate alla lunghezza d'onda specifica dell'utente, cioè “colorate”. Specificità tecnica sostanziale della ONU è che deve trasmettere sulla sua propria lunghezza d'onda, quindi dovrebbe aversi un laser con diversa emissione per ogni ONU, o magari accordabile, ma costoso. Una soluzione per alleviare il costo della ONU, rendendola *colorless*, è quello di far fornire direttamente dalla OLT le portanti ottiche alle ONU, soltanto da modularsi da parte di queste e da rinviare sull'*uplink*, evitando del tutto l'uso di un laser locale. A tale scopo viene usato alla ONU un componente fotonico, detto RSOA (*Reflective Semiconductor Optical Amplifier*), che tipicamente è un amplificatore ottico monolitico in fosforo di Indio (InP) in cui una faccia è ad alta riflettività e che presenta internamente una guida d'onda sagomata, mentre l'altra faccia è a bassa riflettività per consentire il transito in uscita del fascio ottico ricevuto e amplificato (Figura 11). Questo dispositivo consente di realizzare ONU *colorless* a basso costo e con significative potenze d'uscita, dato che l'amplificazione supera i 20 dB.

Per semplificare ulteriormente l'architettura, il segnale di *upstream* di ogni utente può venire generato alla stessa lunghezza d'onda del segnale di *downstream* di quell'utente, sfruttando proprio il segnale modulato in arrivo. Per fare questo, nell'RSOA viene iniettata una frazione del segnale *downstream*, in modo da cancellarne la modulazione di intensità *on-off* sfruttando la saturazione del guadagno del RSOA e consentirne il ri-uso come portante per la trasmissione *upstream*. Per il corretto funzionamento, si richiede che il segnale arrivi all'ONU con un significativo livello di potenza ottica in modo da provocare la saturazione. In tal modo la distanza tra centrale e utente deve essere ridotta (<5 km), portando in frequenti casi a servire solo pochi utenti.

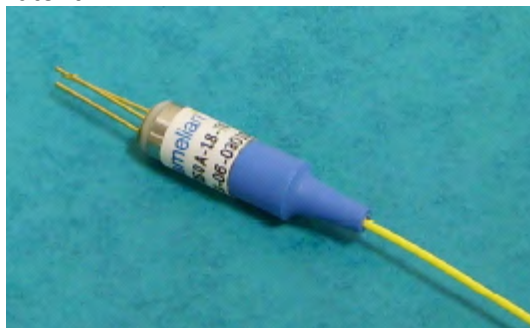


Figura 11
Esemplare di RSOA (per gentile concessione di Amphotonix Ltd)

Una soluzione innovativa per cancellare il segnale modulante in arrivo è stata realizzata e brevettata dall'Istituto TeCIP di Pisa in collaborazione con Ericsson (Figura 12), che ne ha proposto la standardizzazione. Questa soluzione sfrutta una coppia di codifiche di linea che si elidono: il *downstream* è codificato IRZ (*Inverse Return to Zero*) mentre l'*upstream* viene preliminarmente modulato RZ con lo stesso segnale ricevuto. Nella pratica, il formato RZ consente la cancellazione dell'informazione del segnale IRZ mediante una semplice sincronizzazione elettronica al lato utente. Questo porta un aumento della possibile distanza tra centrale e utente con conseguenti economie di scala e vantaggi nell'installazione: nel laboratorio del TeCIP si sono raggiunti 80 km con capacità totale in ciascun verso di 1.25 Gb/s.



Figura 12

Prototipo di WDM-PON ONT con riuso della lunghezza d'onda realizzato a Pisa (per gentile concessione di Ericsson Telecomunicazioni)

5. I circuiti fotonici integrati

I circuiti fotonici integrati (PIC – *Photonic Integrated Circuits*) vengono utilizzati per realizzare in dimensioni estremamente ridotte sia apparati e sistemi già esistenti e che impiegano diversi singoli dispositivi fotonici collegati fra loro (ad esempio: laser, fotorivelatori, amplificatori ottici, AWG, modulatori MZ, filtri, ecc.), sia sistemi di nuova concezione, ottenendo in maniera simile ai circuiti elettronici integrati notevoli risparmi in termini di consumo energetico e ingombro. Molto spesso inoltre utilizzando sistemi ottici integrati è possibile ottenere prestazioni di sistema che non sarebbero raggiungibili con le tecnologie elettroniche. I circuiti fotonici integrati possono essere utilizzati anche per l'integrazione di sensori biologici (*Lab on Chip*) e per la realizzazione di interrogatori di sensori in fibra ottica.



La ricerca e lo sviluppo sperimentale nel campo della fotonica integrata sono considerati strategici a livello globale sia per le potenzialità di innovazione che per le importanti ricadute industriali, e non solo nell'ICT. Michael Lebbby, presidente dell'OIDA (*Optoelectronics Industry Development Association*), ha dichiarato, durante un'audizione nel 2009 al Congresso USA sulla politica industriale della Cina e sul suo impatto sulle industrie e sull'economia statunitensi, che una delle aree su cui egli suggeriva al governo di investire era la tecnologia PIC. Lebbby ha affermato testualmente: *"Invest in photonic integrated circuit (PIC) technology based on both silicon and indium phosphide. PIC devices will transform optoelectronics just as the integrated circuit (IC) transformed semiconductor technology 50 years ago. For PICs, Moore's law, made famous by the semiconductor industry, is just beginning."*

L'uso di dispositivi e infrastrutture di rete basate su fotonica integrata è una delle chiavi di volta per aumentare l'efficienza energetica dei sistemi riducendone allo stesso tempo le dimensioni (*green communications*). Questo importante risultato può essere ottenuto sia implementando per mezzo di circuiti fotonici integrati e miniaturizzati gli attuali sistemi che sono invece realizzati per mezzo di elementi separati e discreti montati su molte schede di controllo individuali, sia realizzando nuovi elementi fotonici multifunzionali. Questi ultimi potranno permettere l'elaborazione ottica dei segnali, sia analogica che digitale (*digital photonics*), in grado di sostituire per molte applicazioni la più dispendiosa, dal punto di vista energetico, elaborazione elettronica con velocità e capacità di banda a questa precluse.

Dal punto di vista tecnologico, l'integrazione può essere monolitica, quando vari componenti funzionali vengono integrati utilizzando un unico materiale. Altrimenti ibrida quando si realizza un *chip* utilizzando materiali diversi, per realizzare funzionalità diverse, che sono poi assemblati mediante un sistema di *mother-daughter board* oppure *bonding*. Tipicamente dispositivi attivi (quali laser e amplificatori) sono assemblati su circuiti ottici passivi, venendo così a rappresentare una sorta di gradino intermedio tra le soluzioni di puro assemblaggio di componenti discreti e quella di totale integrazione monolitica. Un esempio di integrazione ibrida vede tipicamente la tecnologia "*silicon photonics*" utilizzata per la parte passiva con l'applicazione tramite *bonding* di elementi attivi in fosforo di Indio (InP), un semiconduttore composito binario del gruppo III-V.

A quanto è dato vedere al momento, la scelta di tecnologia monolitica o ibrida per le future realizzazioni dipende da moltissimi fattori tecnici, economici e dal contesto applicativo. Se da un lato vi può essere la tendenza ad aggregare monoliticamente per semplicità ciò che può essere fatto (ad esempio un AWG è realizzabile sia in InP che su silicio), dall'altro le prestazioni che si profilano dipendono dalla tecnologia usata: un modulatore per DQPSK realizzato in InP non è competitivo sulla trasmissione a lunga distanza (*long-haul*) con uno realizzato in niobato di

litio LiNbO_3 , mentre può essere conveniente integrarlo con il laser per applicazioni *metro*. E' del tutto probabile che per un periodo non breve l'integrazione monolitica e quella ibrida coesisteranno al meglio di ciascuna. L'utilizzo di circuiti in "*silicon photonics*" per la parte passiva è comunque fortemente favorito dalla possibilità di integrazione con i circuiti elettronici con cui condivide i materiali e i processi tecnologici.

E' comunque certo che le nuove soluzioni di sistemistica interagiranno sostanzialmente con le soluzioni tecnologicamente perseguibili in forma integrata, monolitica o ibrida, influenzandosi a vicenda, e che l'attività futura di R&S per la larga banda dovrà rivolgersi all'integrabilità delle stesse. A riprova della necessità diffusa di affrontare la ricerca in linea con l'evoluzione in corso verso la fotonica integrata, in figura 13 è mostrata una foto al microscopio di un *chip* realizzato per integrazione monolitica contenente una serie di circuiti (un rigeneratore multilunghezza d'onda, un rigeneratore per segnali DPSK e un circuito che realizza una "*1-bit memory*") progettati dall'Istituto TeCIP della Scuola Superiore Sant'Anna di Pisa, *chip* realizzato in InP con la tecnologia della piattaforma europea JePPIX (*Joint European Platform for InP-based Photonic Integrated Components and Circuits*). La figura 14 mostra il banco del TeCIP per la caratterizzazione della componentistica integrata prototipale prima del *packaging*.



Figura 13

Immagine al microscopio di un circuito fotonico integrato multifunzione in fosforo di Indio (dimensioni 3x3 mm)

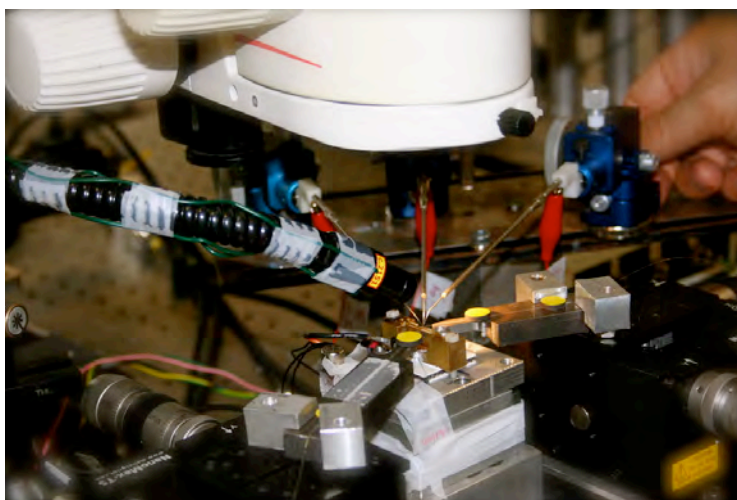


Figura 14

Banco per la caratterizzazione dei circuiti fotonici integrati prima del packaging

Tutte le grandi aziende puntano a sviluppare circuiti fotonici integrati sempre più complessi per un sempre maggior numero di funzioni: Infinera, in particolare, ha fatto dell'integrazione monolitica il proprio cavallo di battaglia, principalmente puntando sulla multiplazione WDM integrata di vari canali per passare da 100 Gb/s totali ($10 \times 10 \text{ Gb/s}$) a 400 Gb/s e oltre. Il numero di componenti per *chip* tende ormai a raggiungere il migliaio nelle soluzioni prototipali più avanzate, come si vede dall'evoluzione negli anni riportata in figura 15. Non vi è dubbio che questa sarà la strada che la fotonica percorrerà nei prossimi decenni.

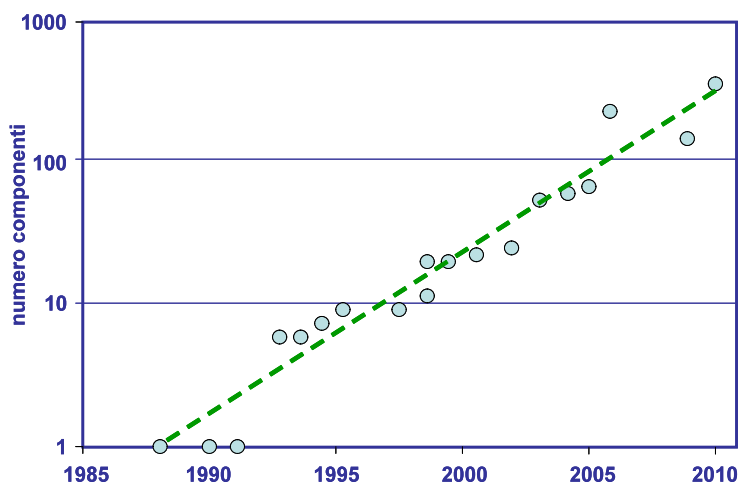


Figura 15

Evoluzione del numero di componenti integrati negli anni

E' interessante evidenziare che oltre alle applicazioni ai grandi apparati di telecomunicazione o elaborazione dati, l'integrazione fotonica troverà

applicazioni anche nel mercato di massa o *consumer*. Si prevede infatti che dispositivi realizzati mediante integrazione fotonica verranno introdotti direttamente nel mercato *consumer*, a dimostrare la capacità di penetrazione di questa tecnologia nei settori economici che richiedono grandi numeri e alte rese. Le periferiche dei *personal computer* potranno essere connesse per mezzo di un cavo ottico ad alta capacità, quale quello proposto da Intel (*Light Peak*), che usa interfacce realizzate con la fotonica integrata. L'aspetto di questo dispositivo, di dimensioni confrontabili con quelle di una comune porta USB attualmente in uso, sarà come rappresentato in figura 16, le capacità di trasmissione saranno invece incomparabili raggiungendo i 10 Gb/s per la prima generazione di dispositivi e i 100 Gb/s per quelle successive.

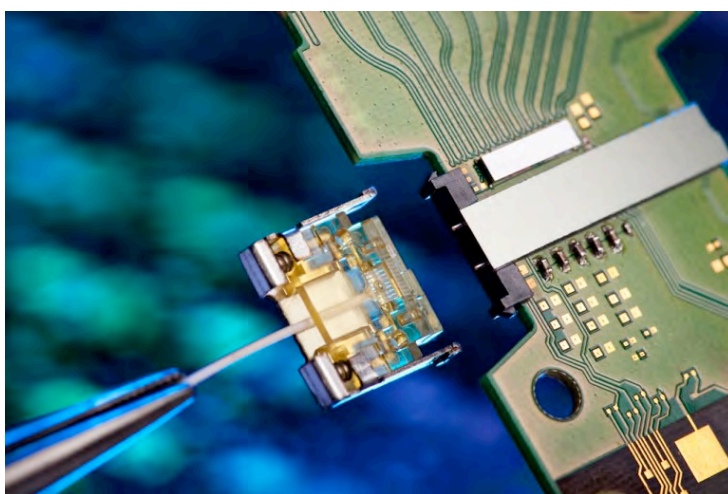


Figura 16
Cavo ottico di interconnessione INTEL ad alta capacità
(per gentile concessione di Intel Corporation)

6. Conclusioni

Già oggi, senza il contributo della fotonica, l'Internet e le comunicazioni in generale non sarebbero possibili con le prestazioni che conosciamo. Le esigenze del futuro prossimo e lontano potranno essere soddisfatte al meglio, sia come prestazioni e servizi all'utente, sia come consumi energetici e impatto ambientale, solo con una rapida evoluzione della tecnologia fotonica. Per effetto di questa evoluzione, le capacità dei collegamenti nelle dorsali stanno rapidamente dirigendosi verso le decine di Terabit al secondo sulla singola fibra e verso il Terabit al secondo sulla singola lunghezza d'onda. Per la rete di accesso, la fotonica sta sviluppando soluzioni a basso costo all'utente per portare molti Gigabit al secondo nei due versi sulla FTTx.

Nello strato fisico, le funzioni svolte nel dominio ottico si arricchiranno rispetto alle attuali, dato che la tecnologia fotonica consentirà di effettuare direttamente molte delle funzioni che sono svolte oggi con un passaggio



O-E-O, cioè dall'ottica all'elettronica e di nuovo all'ottica, riducendo costi e consumo energetico senza il collo di bottiglia in banda dell'elettronica. I circuiti fotonici integrati svolgeranno un ruolo fondamentale in questa evoluzione, coesistendo al meglio con quelli elettronici, favoriti in ciò dalla tecnologia *silicon photonics* che sta sviluppandosi in modo rapido. L'integrazione fotonica, nelle comunicazioni ma anche in molte altre aree, dalla sensoristica alla sicurezza, dalla biofotonica all'illuminazione, rappresenta non solo una delle cinque tecnologie abilitanti individuate dalla Commissione Europea, ma il futuro necessario per l'innovazione strategica e il miglioramento della competitività dell'Europa in tutta l'ICT e in molte altre aree.

Ringraziamenti

L'Autore ringrazia Guido Vannucchi per la preziosa e stimolante revisione e i colleghi Antonella Bogoni, Piero Castoldi, Ernesto Ciaramella, Giampiero Contestabile, Fabrizio Di Pasquale, Enrico Forestieri, Luca Potì per il materiale fornito, gli utili suggerimenti e l'attenta rilettura del manoscritto.

Biografia

Giancarlo Prati è Professore Ordinario di Telecomunicazioni e Direttore dell'Istituto di Tecnologie della Comunicazione, dell'Informazione e della Percezione (TeCIP) della Scuola Superiore Sant'Anna di Pisa. E' stato Direttore e poi Presidente del CNIT (Consorzio Nazionale Interuniversitario per le Telecomunicazioni) dal 1995 al 2010. E' IEEE Fellow e membro del Panel "Telecommunications" dello European Research Council (ERC). Ha condotto ricerche nel settore delle comunicazioni ottiche dal 1978 presso la University of Southern California, il CNR di Pisa e l'Università di Parma. E' autore di oltre 100 pubblicazioni e numerosi brevetti nel settore. E-mail: giancarlo.prati@cnit.it

GLOSSARIO

ADSL	Asynchronous Digital Subscriber Loop
ATM	Asynchronous Transfer Mode
AWG	Arrayed Wavelength Grating
DQPSK	Differential Quadrature Phase Shift Keying
EDFA	Erbium Doped Fiber Amplifier
EPON	Ethernet Passive Optical Network
FFT	Fast Fourier Transform
FTTH	Fiber To The Home
FTTB	Fiber To The Building
FTTC	Fiber To The Curb
FTTx	Fiber To The x
GPON	Gigabit Passive Optical Network
IEEE	Institute of Electrical and Electronic Engineers
IPTV	Internet Protocol TeleVision
IQ	In-phase, Quadrature
IRZ	Inverse Return to Zero
ISP	Internet Service Provider
ITU	International Telecommunication Union
IX	Internet Exchange
LAG	Lambda AGgregation
LDPC	Low Density Parity Check
MAN	Metropolitan Area Network
MZ	Mach-Zehnder
OADM	Optical Add Drop Multiplexer
OFDM	Orthogonal Frequency Division Multiplexing
OIDA	Optoelectronics Industry Development Association
OLT	Optical Line Terminal
ONT	Optical Network Termination
ONU	Optical Network Unit
OTN	Optical Transport Network
OXC	Optical Cross Connect
P2P	Point To Point
PIC	Photonic Integrated Circuit
PLC	Planar Lightwave Circuit
PON	Passive Optical Network
PoP	Point of Presence
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
ROADM	Reconfigurable Optical Add Drop Multiplexer
RSOA	Reflective Semiconductor Optical Amplifier
RZ	Return to Zero
SOA	Semiconductor Optical Amplifier
SC	Single Carrier
SDH	Synchronous Digital Hierarchy
SONET	Synchronous Optical NETwork
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
VoIP	Voice over Internet Protocol
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing

RIQUADRO 1 - Modulatore Mach-Zehnder e ibrido ottico a 90°

Il modulatore MZ è il dispositivo base per la modulazione IQ del fascio ottico e può essere realizzato con tecnologie diverse: niobato di litio LiNbO_3 , silicio su isolante, polimeri organici, e altro. Si tratta di un dispositivo elettro-ottico con due bracci, in cui è suddiviso il fascio entrante e a ciascuno dei quali è applicato un campo elettrico. L'indice di rifrazione in ciascun braccio, e quindi la velocità della luce in esso e il tempo di transito del segnale, dipendono dal campo applicato. Poiché la fase del segnale che lascia il braccio dipende appunto dal tempo di transito, agendo sul campo elettrico dei due bracci e ricombinando i rispettivi fasori all'uscita si ottiene la voluta modulazione del segnale (Figura A.1).

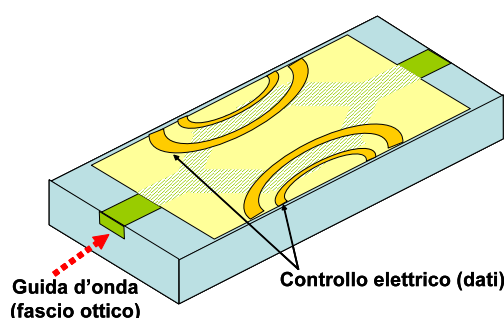


Figura A.1
Modulatore Mach Zehnder

Al ricevitore, l'ibrido ottico a 90° è un dispositivo fotonico a 6 porte consistente in 2 *splitter* lineari, 2 accoppiatori a X e uno *shifter* di fase, interconnessi in modo da ottenere quattro differenti somme fasoriali di due fasci ottici: quello del laser che funge da oscillatore locale e quello del segnale in arrivo dalla fibra ottica (Figura A.2), il tutto ripetuto per ognuna delle due polarizzazioni usate per la trasmissione *Dual Pol.* I quattro segnali sono inviati a una coppia di fotorivelatori bilanciati, configurati cioè in modo da cancellare il proprio rumore intrinseco, che rivelano le correnti delle componenti in fase e in quadratura di quella polarizzazione.

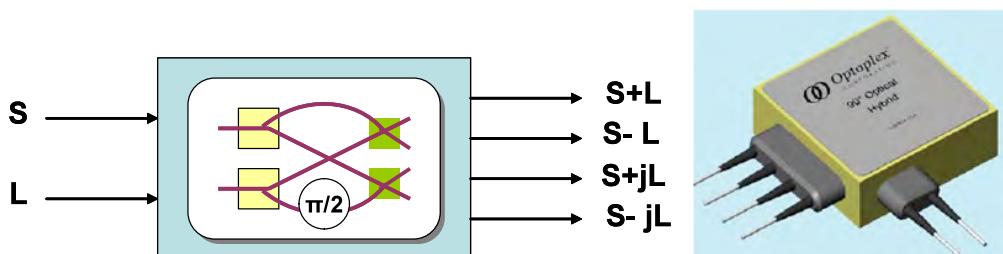


Figura A.2
Circuito ibrido ottico a 90° (per gentile concessione di Optoplex Corporation)

RIQUADRO 2 - Arrayed Waveguide Grating (AWG) e switch fotonico

L'operazione di *demultiplexing* nell'AWG avviene iniettando il fascio ottico WDM proveniente dalla fibra di rete in uno spazio ove si diffonde ad illuminare un certo numero di fibre raggruppate o di guide d'onda ricavate canalizzando un opportuno materiale. Le fibre (o le guide d'onda canalizzate) hanno lunghezze differenti, determinando un'opportuna diversa differenza di fase all'uscita. I segnali ottici così prodotti quindi attraversano un altro spazio interferendo in modo tale che ciò che risulta in uscita su una specifica fibra ottica sia soltanto uno solo dei canali con una certa lunghezza d'onda. Usato nel verso opposto, cioè applicando diverse lunghezze d'onda all'uscita, l'AWG ne effettua la moltiplicazione WDM, che risulta disponibile sull'unica fibra all'ingresso (Figura B.1). I principali materiali per la realizzazione di dispositivi AWG comprendono la silice, il fosforo di Indio, i polimeri, il cosiddetto SOI (*Silicon-On-Insulator*). Quest'ultimo ha il vantaggio di poter venire integrato durante la fabbricazione in un unico chip con altri dispositivi elettronici e di essere quindi poco costoso.

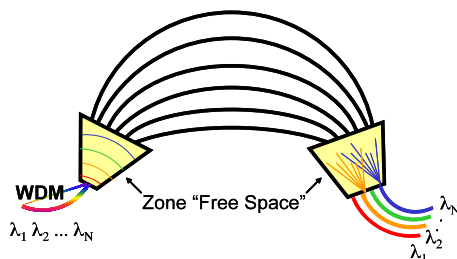


Figura B.1

Schema di principio di un AWG

Per l'operazione di commutazione (*switching*) fotonica vi sono numerose tecniche che sfruttano effetti fisici vari, quali l'effetto magneto-ottico, elettro-ottico, acusto-ottico, termo-ottico, non linearità nel mezzo, cristalli liquidi, e altri ancora. In figura B.2 è mostrato lo schema di un recente switch fotonico basato su effetto termo-ottico, realizzato da NTT Photonics Laboratories³ con tecnologia dei circuiti ottici planari (PLC) in silice su substrato di silicio. La struttura ibrida silice-silicio realizza un interferometro di Mach Zehnder (MZ), in cui uno dei bracci è esposto ad un riscaldatore a film sottile. Quando il riscaldatore è attivato l'interferenza dei fasci dai due bracci in fase di ricombinazione provoca l'uscita sulla porta 2 anziché sulla porta 1.

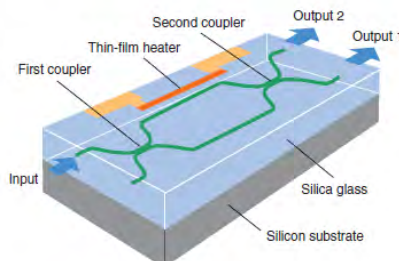


Figura B.2

Struttura di switch fotonico termo-ottico ibrido su silice-silicio (per gentile concessione di NTT Technical Review)

RIQUADRO 3 - GPON ed EPON

Le due varianti di TDM-PON per l'accesso, le GPON e le EPON, la prima standardizzata da ITU-T, la seconda da IEEE, differiscono sensibilmente per caratteristiche operative, e quindi della componentistica e dei sottosistemi fotonici. La rete GPON è basata sulla definizione completa di un nuovo protocollo che accetti una molteplicità di servizi nei loro diversi formati *nativi*, costituendo quindi una soluzione di trasporto multi-protocollare con una propria struttura TDM di *framing* standardizzata, mentre l'EPON ha la sua caratteristica principale nel preservare il modello architetturale di Ethernet, per una connettività *seamless* di qualsiasi tipo di comunicazione basata sull'*Internet Protocol (IP)*, senza introdurre una propria ulteriore struttura di *framing*.

Per effetto di questa diversa impostazione, le due soluzioni, pur simili per architettura, differiscono per tecnologia, contesto di impiego, efficienza, costo e gestione di rete. Le distanze operative sono per entrambe di qualche decina di km: lo standard GPON prevede 60 km per un massimo di 128 utenze, che diventano nella per utenza). Per EPON le capacità sono simmetriche su *downlink* e *uplink*, prevalentemente 1.25 Gb/s, il rapporto di *splitting* potrebbe essere molto superiore in accordo allo standard, ma in pratica è anch'esso di 1:32 con analoghi valori di distanza. Per il prossimo futuro si punta alla capacità pratica 20 km per 32 utenze, con capacità tipiche aggregate di 2.5 Gb/s sul *downlink* e 1.25 Gb/s sull'*uplink* (e quindi 1250/N Mb/s *downlink* di 10 Gb/s, con la 10G-EPON in fase di sviluppo per la produzione, e la 10GPON di ITU-T che ha raggiunto la standardizzazione.

L'elemento tecnologico a comune fra le due tecniche GPON ed EPON è lo *splitter* ottico di potenza 1:N, che presenta un ingresso in fibra su un lato ed N uscite in fibra sull'altro, ciascuna delle quali viene alimentata dallo stesso segnale con 1/N della potenza in ingresso. Lo *splitter* è un componente bidirezionale in quanto agisce da accoppiatore (*coupler*, *combiner*) nella direzione opposta. Una tecnica di realizzazione è basata sulla fusione ad alta temperatura delle fibre ottiche in modo da preservare la monomodalità di propagazione, ottenere l'uniformità del rapporto di *splitting*, e minime perdite dovute alla polarizzazione. Con questa tecnica il rapporto primario di *splitting* ha al più N=4, e per raggiungere valori più alti si mettono diversi *splitter* in cascata. Un'altra possibilità tecnologica, che offre migliori prestazioni, è la realizzazione tramite circuiti ottici planari (*PLC - Planar Lightwave Circuits*) ottenuti depositando opportuni strati di silice SiO₂ su un substrato di silicio e ricavando in essi guide d'onda, con le fibre semplicemente accoppiate in ingresso e in uscita.

Oggi sia per OLT che per ONU dei due tipi di PON sono disponibili sorgenti laser a semiconduttore, compatte e a costo relativamente basso. Per le OLT si adottano laser DFB (*Distributed Feedback Laser*), disponibili sia in 3^a che in 2^a finestra. Si tratta di laser non particolarmente economici, con larghezze di riga stretta e potenza tipica di 1 dBm. Per le ONU sono disponibili laser Fabry-Perot, più economici ma più rumorosi (larghezza di riga maggiore) con potenza tipica di 0 dBm e adatti alla 2^a finestra. I ricevitori possono usare fotodiodi PIN oppure, per distanze maggiori, APD (*Avalanche PhotoDetector*), che hanno più alta sensibilità (8-10 dB in più dei circa 20 dBm dei PIN), ma anche maggior costo.

Nelle PON possono coesistere distanze sensibilmente diverse delle varie ONU dalla OLT: il ricevitore della OLT nelle TDM-PON deve essere in grado di ricevere i *burst* delle singole ONU a potenze anche molto diverse (*near-far effect*), configurandosi automaticamente di volta in volta. In alternativa, si può anche intervenire sulle potenze delle ONU rinunciando alla loro uniformità, con maggiori costi e complessità generale.



SOCIAL WEB: PROSPETTIVE E CONSIDERAZIONI

Marco Gaudina, Mauro Migliardi

L'avvento del Web 2.0 ha reso fruibile ai più uno strumento di comunicazione che, in precedenza, richiedeva un significativo background tecnologico. Tuttavia, questa evoluzione ha causato anche fenomeni come la disgregazione cognitiva, lo svuotamento del concetto di "trust" e l'aumento della frammentarietà dell'informazione stessa. In questo articolo presenteremo l'evoluzione storica del social web e gli aspetti tecnologici che l'hanno veicolata. Affronteremo poi gli aspetti oscuri sopra citati e, parallelamente, i recenti sviluppi della sicurezza informatica nei sistemi di social networking.

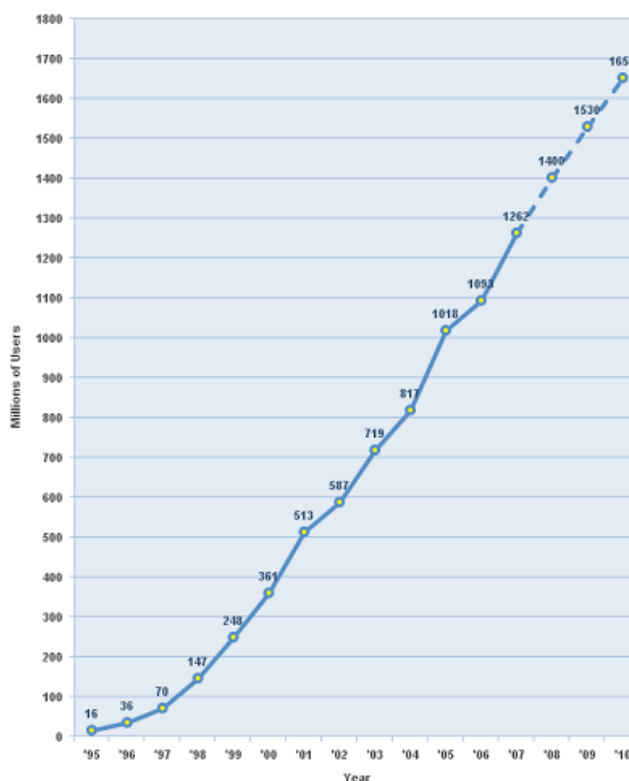
1. Fenomeno Social

L'avvento delle tecniche cosiddette Web 2.0 ha reso completamente fruibile alle masse uno strumento di comunicazione che, in precedenza, era maggiormente utilizzato dagli addetti ai lavori in quanto richiedeva un significativo *background* tecnologico.

In particolare, le potenzialità e i servizi offerti da uno strumento come il World Wide Web sono state negli ultimi anni espansi a tal punto da transire dallo stato di strumento di lavoro riservato alle *elite* tecnologiche a quello di strumento di uso quotidiano per il lavoro e il tempo libero della maggioranza della popolazione. Il numero degli utenti di Internet è cresciuto notevolmente negli ultimi 15 anni (Figura 1): da 16 milioni di utenti (0.4% della popolazione mondiale) nel dicembre del 1995, a 1971 milioni (28.8% della popolazione mondiale) di utenti nel settembre del 2010 [1]. Mentre è relativamente semplice stimare il numero di utenti, risulta estremamente complicato stimare la quantità di informazione nella rete, in primo luogo per la difficoltà di eseguire una somma senza duplicazioni del

traffico in transito nei diversi nodi, in secondo luogo in quanto, in effetti, il traffico da solo non fornisce un'adeguata misura della dimensione del fenomeno Internet e, in effetti, non esiste una metrica adeguata che copra tutti i possibili aspetti di tale fenomeno. Se anche si riuscisse, quindi, a stimare il numero di petabyte effettivamente scambiati, non si riuscirebbe a catturare un effettivo significato di questi scambi.

Un stima qualitativa dell'informazione disponibile può essere formulata valutando il numero di punti di accesso alle informazioni: nel 2008, gli ingegneri di Google hanno annunciato che il motore di ricerca aveva visto fino a quel momento un trilione di URL differenti. Tuttavia, questo numero probabilmente contiene molti collegamenti auto generati che ne falsano quindi l'effettivo valore e, altri studi dello stesso anno, hanno stimato la presenza di almeno cinquanta miliardi di pagine web [2].



Source: www.internetworldstats.com - January, 2008
Copyright © 2008, Miniwatts Marketing Group

Figura 1
Numero di utenti internet negli anni dal 1995 al 2010

L'incredibile crescita dimensionale, però, non può essere giustificata se non si considera parallelamente anche l'evoluzione che ha avuto la modalità di produzione dei contenuti presenti sul Web stesso. Infatti, una delle fonti di

dati per il nuovo Web in crescita più esplosiva è rappresentata dagli utenti stessi che desiderano condividere pezzi della loro esistenza.

Questa evoluzione ha stabilmente e decisamente modificato la posizione di confronto tra utente e risorsa tecnologica ponendo sempre più l'accento sulla partecipazione e dando spazio all'ormai dilagante fenomeno dei Social Networks.

È proprio con l'introduzione del concetto di social network che Internet ha visto e sta vivendo la sua più grande evoluzione dall'introduzione del primo motore di ricerca (Figura 2). Concetti come condivisione, collaborazione e distribuzione dell'informazione, elementi base alla nascita delle reti di calcolatori, sono diventati oggi un elemento diffuso e ben radicato nella coscienza di una sempre più larga base della popolazione. Il fenomeno "social" è in continua espansione e con frequenza quasi giornaliera vengono alla luce nuovi servizi per la condivisione di contenuti, qualsiasi tipo essi siano. Le statistiche di Facebook (dati aggiornati al 31/10/2011) [3], attualmente il social network più importante e diffuso, sono eloquenti:

- più di 800 milioni di utenti attivi;
- 50% di questi si collegano ogni giorno;
- l'utente medio ha 130 amici;
- l'utente medio è connesso a 80 pagine, gruppi ed eventi;
- l'utente medio crea 90 risorse ogni mese;
- più di 30 miliardi di risorse (link a pagine web, notizie, articoli di blog, note, album di foto, etc.) condivise ogni mese.

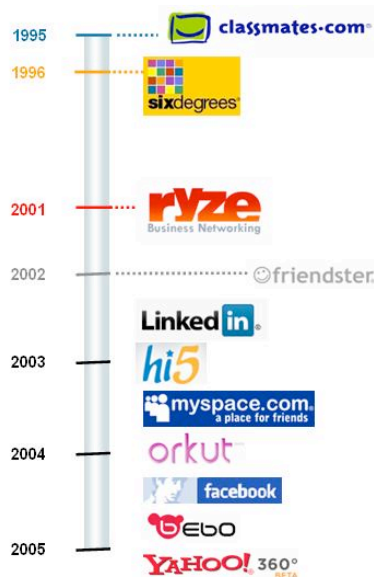


Figura 2

Timeline della nascita di alcuni tra i più noti social network



Boyd ed Ellison [9] hanno definito i siti di social network (SNS) come servizi basati sul web che consentono agli utenti di:

- costruire un profilo pubblico o semi pubblico all'interno di un sistema vincolato;
- articolare una lista di altri utenti con cui condividono una connessione;
- vedere e navigare la loro lista di connessioni e quelle create dagli altri utenti all'interno del sistema.

Gli utenti vogliono condividere informazioni personali con altri utenti e le piattaforme di social networking soddisfano questo bisogno. Questo è uno dei motivi principali del successo dei social network.

Se da una parte la sopra descritta evoluzione, ha contribuito a creare nuovi fenomeni sociali in una ampio spettro di campi di interesse (si pensi al lato ludico, a quello commerciale, a quello medico, a quello sportivo, etc.) dall'altro lato questo stesso fenomeno ha alcuni aspetti oscuri.

Il fenomeno "social" è in continua espansione e con frequenza quasi giornaliera vengono alla luce nuovi servizi per la condivisione di contenuti, qualsiasi tipo essi siano. Se da una parte la sopra descritta evoluzione ha contribuito a creare nuovi fenomeni sociali in una ampio spettro di campi di interesse (si pensi al lato ludico, a quello commerciale, a quello medico, a quello sportivo, etc.) dall'altro lato questo stesso fenomeno ha alcuni aspetti oscuri.

In primo luogo, la semplicità con cui è possibile entrare in relazione con altri utenti tende ad inflazionare il concetto fondamentale di fiducia. "Trust", sulla rete massificata, diventa troppo spesso termine virtuale svuotato di qualunque significato e rischia di inficiare sia gli aspetti tecnici della sicurezza delle infrastrutture informatiche che l'essenza stessa del concetto di condivisione (Riquadro 1).

In secondo luogo, la proliferazione delle esistenze virtuali tende a disgregare l'unità cognitiva dell'utente facendo disperdere la sua capacità di attenzione su miriadi di rivoli di dubbia importanza.

Infine, la proliferazione degli interessi fittizi comporta un aumento dell'errore nella ricerca delle informazioni. Infatti, l'aumento delle possibilità di ricerca sulle sopra citate miriadi di rivoli informativi comporta in modo quasi inevitabile una metodologia di ricerca che si basa su associazioni a brevissimo termine di idee che distraggono troppo spesso dal primario obiettivo.

Nel prosieguo di questo articolo cercheremo di inquadrare l'evoluzione storica del web nella sua dimensione "sociale" mettendo in luce gli aspetti tecnologici che l'hanno veicolata. Affronteremo poi gli aspetti oscuri sopra citati e descriveremo come le linee di ricerca più avanzate in questo campo stanno affrontando dette problematiche. Parallelamente a queste due



tracce principali, analizzeremo il problema della sicurezza informatica nei sistemi di social networking e descriveremo i più recenti sviluppi in questo campo.

2. Il lato oscuro

La semplicità con cui è possibile entrare in relazione con altri utenti, tende a inflazionare il concetto fondamentale di fiducia. "Trust", come precedentemente detto, tende a svuotarsi di quello che è il suo significato reale e rimane solo un ostacolo all'accettare di condividere contenuti con la persona che condivide i propri interessi. Questo, oltre ad avere degli aspetti deleteri dal lato della sicurezza dei dati e dell'identità, porta anche a generare storture nell'essenza stessa del concetto di condivisione. In generale, però, non è possibile appiattire l'utenza su di un solo *modus operandi*, al contrario, il comportamento degli utenti da questo punto di vista risulta variegato. Proviamo a riferirci alla condivisione di notizie e al concetto di "trust" nei confronti delle informazioni fornite. I concetti precedentemente presentati di collaborazione e condivisione hanno portato alla nascita di realtà come Digg, dove l'utente stesso diventa editore della notizia fornendo links a contenuti esistenti, creando addirittura fenomeni di competizione come analizzato in [10]; è doveroso inoltre affiancare un altro protagonista dello scenario informativo attuale ovvero Twitter. In questo caso l'utente è esso stesso la fonte della notizia. Nasce quindi spontanea una domanda, come può l'utente fidarsi dei contenuti presentati da terzi? L'utente medio concentra la sua attenzione sul contenuto della notizia o su colui che la presenta, tralasciando in molti casi la fonte da cui l'informazione stessa viene presentata. Ecco quindi che, a meno di contenuti provenienti da personalità o fonti autorevoli, l'utente si ritrova in una condizione di indecisione sulla veridicità dell'informazione stessa. Sono assai frequenti i casi in cui vengono diffuse informazioni false o poco attendibili. Questo preoccupante fenomeno nasce dal bisogno di protagonismo e affermazione generato negli utenti dall'avvento della globalizzazione informatica. Se in alcuni casi, il pericolo può essere considerato marginale, in molti altri campi dell'informazione sta creando situazioni preoccupanti. Pensiamo solamente al fatto che il cybernauta medio al giorno d'oggi, per avere un parere medico consulta prima il web e, solo in un secondo tempo, il proprio medico. Se da un lato questo ha stimolato all'ingresso in rete di svariati esperti del settore, ha dall'altra parte generato una sorta di superficialità informativa che tende a confondere la quantità con la qualità dell'informazione (Riquadro 2).

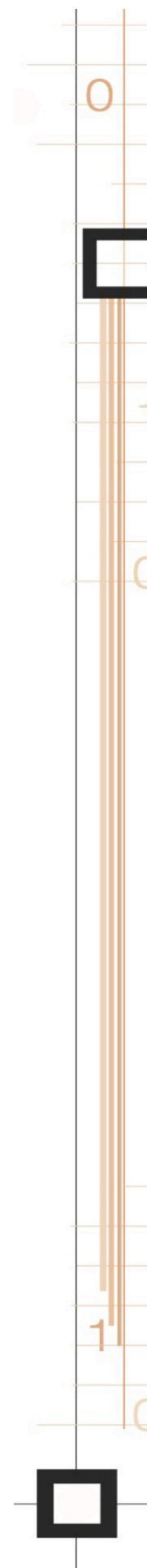
Il fenomeno dell'*overload* informativo, d'altro canto, è spesso postulato essere alla base della generale riduzione dell'*"attention span"* [11]. Una studiosa di psicologia dello sviluppo sostiene che non è solo quello che leggiamo che forma i nostri processi mentali, ma anche il modo in cui leggiamo [12] e, d'altro canto, è ovvio che l'immensa ricchezza di stimoli e informazioni che ci sono resi disponibili in frazioni di secondo dalla rete non può non rappresentare una tentazione fortissima a distogliere la nostra attenzione dal pensiero corrente per inseguire quello successivo.



In modo simile, la proliferazione delle esistenze virtuali (uno dei primi progetti di metapresenza si chiamava efficacemente "*second life*" [13][14], la cui analisi dal punto di vista di interazione sociale può essere trovata in [15]) tende a disgregare l'unità cognitiva dell'utente facendo disperdere la sua capacità di attenzione su miriadi di rivoli di dubbia importanza (Riquadro 3). La maggior parte degli utenti si ritrova ad oggi ad avere più di un *account* (che noi intendiamo come identità) di social networks; la stessa persona fisica si ritrova di conseguenza ad "esistere" su Facebook, su Twitter, piuttosto che su Flickr. Ad ogni istante può "sembrare" una persona diversa a seconda del servizio utilizzato. Questo comporta in molti casi anche la falsificazione della persona stessa poiché, è bene ricordarlo, la partecipazione innesca generalmente un meccanismo di attesa di giudizio dalla comunità che porta alla necessità di presentarsi non tanto al meglio quanto nel modo che ci si aspetta più gradito dalla comunità stessa. I fattori psicologici derivanti, non trattati in questa sede sia per mancanza di spazio che di approfondite competenze, possono creare delle vere e proprie patologie rasentando la dipendenza o l'estraneazione dalla realtà [16].

Da un punto di vista meramente tecnologico, un argomento delicato come quello della sicurezza informatica e della privacy su Internet hanno, ormai da anni, un'ampia e approfondita trattazione [17]. Anche se il *cracking* (ovvero l'utilizzo di tecniche informatiche per eludere blocchi imposti al fine di trarne guadagno) è in continua evoluzione, ogni anno le aziende del settore ICT, dedicano sempre maggiori capitali al concetto di sicurezza. Nei social network le due caratteristiche fondamentali di *Safety* (protezione dai danni) e *Reliability* (capacità di fornire il servizio anche in condizioni di stress), sono fondamentali e ricevono sempre grande attenzione. Inoltre, i modelli di identificazione e accesso sono rigidi e cercano di limitare al massimo l'intrusione di non autorizzati, utilizzando sistemi di conferma come e-mail, o l'inserimento di caratteri non distinguibili da sistemi automatici. Infine, l'utilizzo della crittografia asimmetrica permette di garantire confidenzialità e integrità delle informazioni anche a fronte di tentativi di intrusione. Tuttavia, è ben noto che una catena di sicurezza è forte come il più debole dei suoi anelli e, anche in presenza di un sistema tecnologico perfetto, l'anello debole della catena risiede nel fattore umano [18]. Lo svuotarsi del concetto di "*trust*" precedentemente descritto (i ricercatori stessi riaffermano l'importanza della conoscenza e della fiducia *offline*, nel mondo reale [8]) insieme ad un certo rifiuto della crescente complessità dei meccanismi richiesti dalla sicurezza, porta l'utenza a comportamenti che vanificano qualsiasi tecnologia. Si pensi ai comportamenti stigmatizzati dai *comics*, online e non, come quello di utilizzare una singola password per tutti i sistemi (dal distributore di bevande all'attivazione del contrattacco nucleare) o quello di scrivere le password sul dispositivo relativo.

Analizzare un social network dal punto di vista della privacy [19] è attualmente un argomento molto diffuso e al centro di numerosi dibattiti che





si concentrano, principalmente, sull'uso che può essere fatto dell'associazione dei dati personali con le abitudini on line dell'utente stesso. È infatti ormai consuetudine che i contenuti e le informazioni relative ad un utente vengano fornite a terzi allo scopo di costruire campagne pubblicitarie mirate sui gusti e gli interessi del singolo utente mentre la possibilità di limitare questo uso molto spesso non è facilmente individuabile o addirittura del tutto assente. In Italia, fin dalla legge **547/1993**, sono state introdotte norme a tutela di un corretto utilizzo di un dispositivo informatico denunciando comportamenti di illegittimità mirati all'utilizzo improprio di dati e informazioni. Tuttavia, quello che risulta ignoto ai più è il fatto che, con una semplice ricerca "nome e cognome" su di un motore di ricerca, si ottengono le informazioni base di una persona (età, indirizzo, numero di telefono) poiché esistono diversi siti dedicati ad aggregare ciò che trapela da un social network piuttosto che da un altro. È pur vero che gli stessi social network cercano di impedire questo tipo di rivelazioni e l'indicizzazione dei profili, ma spesso questa schermatura non è efficace. Il concetto di privacy nell'era di Internet risulta quindi essere assai più volatile di quanto il cittadino medio riesca ad immaginarsi in quanto qualunque dato abbia un singolo istante di visibilità su Internet ha un'elevata probabilità di diventare pubblico per sempre. Un altro aspetto difficilmente presente nella coscienza comune è l'intrinseca extra-territorialità della rete.

Si pensi, ad esempio, alla recente campagna di opinione relativa alla chiusura di siti internet che rendevano possibile l'accesso a contenuti coperti da *copyright*. La legge Italiana non può ovviamente avere effetto al di fuori del territorio della Repubblica, quindi qualsivoglia ingiunzione di chiusura rivolta ad un sito estero non avrebbe alcun valore. Inoltre, rendere davvero inaccessibili dal territorio Italiano tali siti richiederebbe l'analisi dei contenuti di tutti i flussi di dati di tutti gli utenti, cosa che, oltre ad essere computazionalmente incredibilmente onerosa, sarebbe essa stessa illegale in quanto rappresenterebbe una plateale violazione della privacy configurandosi alla stessa stregua di un'intercettazione preventiva delle telefonate di tutti i cittadini.

3. Aspetti Tecnologici

I social network più diffusi, ma in generale tutti i moderni servizi web 2.0, presentano similitudini dal punto di vista dell'architettura software. Esistono molti *web framework* che facilitano l'integrazione della componente responsabile della persistenza dei dati con quella responsabile della presentazione della vista agli utenti.

Tra questi possiamo citare Django, Ruby on Rails, Grails. Questi *web framework* sono detti *developer centric*, ovvero puntano la loro attenzione sulla semplicità di sviluppo del codice anche, a volte, a scapito della sua scalabilità sia in termini di crescita della complessità del sistema, sia in termini di crescita delle performance richieste. Infatti, la rapidità di generazione di un sistema funzionante garantita da questi *framework*, deve essere abbandonata nel momento in cui il team di sviluppo cresce, la



complessità del codice cresce, la richiesta di servizio cresce. Ne è un esempio paradigmatico il recente caso di Twitter che ha abbandonato Ruby on Rails a favore di un framework basato su Java. Uno dei punti di forza di *Ruby on Rails*, ad esempio, è la sua capacità di non richiedere esplicitamente una configurazione e di affidarsi alle scelte di *default*.

Se questo è un fattore di elevata semplificazione all'interno di un team di sviluppo "*one man gang*" o poco più, al crescere del numero di persone coinvolte è invece un serio problema, infatti:

1. è estremamente difficile distinguere ciò che si è volutamente tralasciato con ciò che si è distrattamente tralasciato;
2. è estremamente difficile comprendere tutti gli effetti collaterali di una scelta progettuale nel momento in cui essa è espressa semplicemente come *default* senza alcuna giustificazione.

Una ulteriore peculiarità risiede nell'utilizzo di nuove tecnologie per quanto riguarda il lato database. Pur essendo un periodo di transizione, dopo la dimostrazione della congettura di Brewer, cioè l'impossibilità di avere un sistema distribuito che sia contemporaneamente *Consistent*, *Available* e *Partition tolerant* (si parla infatti di CAP theorem) [20], vi è un crescente interesse per i database NoSql. È il caso, ad esempio, di Cassandra [21] nato nei laboratori di Facebook e in seguito diventato di dominio pubblico passando sotto lo sviluppo di *Apache Software Foundation*. Cassandra è un database distribuito, *fault-tolerant*, altamente scalabile e orientato alla gestione di immense quantità di dati. Il prezzo da pagare per queste caratteristiche è la cosiddetta "*eventual consistency*", ovvero l'incapacità di garantire l'atomicità delle modifiche sui dati e la necessità di trovare una soluzione a posteriori nel caso in cui vengano rivelate modificazioni concorrenti.

Arrivando infine al livello di presentazione, uno dei punti tecnologici cardine è lo spostamento di parte della computazione sul lato *client* e il trasporto delle modifiche sul deposito persistente dei dati in modo asincrono. Questo permette di dare una maggior immediatezza e migliorare la reattività delle applicazioni web, ma richiede ovviamente un'elevata affidabilità della connessione tra *client* e *server*, pena la perdita di consistenza tra i dati e la vista dell'utente. Questo tipo di comportamento, introdotto con il *framework* AJAX [22] da Google, è oggi ritrovabile in moltissimi *framework* di presentazione, ad esempio jQuery o Dojo, che stanno prendendo il sopravvento anche su Flash per quanto riguarda le animazioni. Assai spesso, questi *framework*, come AJAX, si basano su javascript. Linguaggio di *scripting* orientato agli oggetti, Javascript, di recente, sta diventando uno standard nella costruzione di siti web dinamici e ricchi di elementi grafici. Infine, anche il cuore stesso del web, il linguaggio HTML, è in evoluzione. Se fino a ieri era pratica comune l'installazione di vari *plugins* per i comuni *browser* per la gestione di elementi multimediali, la recente nascita dello



standard HTML5, sta muovendo il mercato verso una nuova direzione. Con l'introduzione di *tags* (elemento di una pagina web) come `<video>` o `<audio>` si è voluto ampliare notevolmente le capacità multimediali di un contenuto web, andando a limitare la necessità di *plugin* esterni. Le novità introdotte da HTML5 non si limitano ai soli *tags* multimediali ma vanno ad ampliare e migliorare notevolmente le possibilità offerte da una pagina web, fornendo strumenti per semplificare operazioni che precedentemente risultavano complicate come la formattazione del testo.

4. Conclusioni

Uno studio di Forrester, nel 2009, distingueva cinque ere nell'evoluzione del fenomeno social. Nella prima era (inizio 1995, maturità circa 2007) lo strumento si focalizzava sulla condivisione di relazioni. Gli esempi tipici sono tutti gli strumenti nati per riallacciare contatti con persone di cui si è persa traccia o per inserirsi in nuove reti sociali; è importante notare come, all'epoca in cui lo studio fu condotto, questa era l'unica "era" i cui risultati fossero giunti a maturità e ogni altra "era" si riferisse esclusivamente alle proiezioni effettuate dagli autori e in buona parte troppo ottimistiche. La seconda era (inizio 2007, maturità presunta nel 2012) è quella delle "funzionalità sociali", dell'uscita delle funzionalità tipiche del social network dal singolo sito dedicato per diventare il filo conduttore con cui controllare l'utilizzo di tutti i siti visitati dall'utente. Uno dei primi esempi di questa tipologia di evoluzione è il *Facebook Connect* [23][24]. Lanciato nel Maggio del 2007, permette a siti web terzi di integrare l'esperienza di navigazione degli utenti all'interno della rete sociale che loro hanno su Facebook.

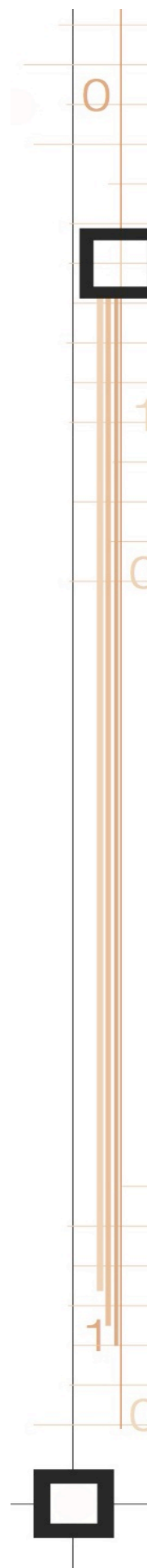
La terza era (inizio 2009, maturità presunta nel 2011) è quella in cui ogni attività sul web diventa "sociale"¹. Seppure siamo ancora lontani da questa visione, il concetto alla base di questo fenomeno è la semplice accettazione del fatto che il modello di moderazione che stava alla base, ad esempio dei *newsgroup* della prima *usenet* o che è utilizzabile in un singolo sito scarsamente popolato, non è in grado di scalare per affrontare lo tsunami di contenuti riversati globalmente nella rete dagli utenti stessi. La soluzione proposta è quella, di fatto, di permettere agli utenti stessi di essere una sorta di moderatore aggregato. Il concetto non è nuovo ed è stato introdotto, seppure in campo diverso, dal progetto Freenet [25] con l'adozione di meccanismi di replicazione e obsolescenza dell'informazione guidati dalle dichiarazioni di interesse degli utenti stessi. La quarta era (inizio presunto 2010, maturità presunta 2012) è quella del "Contesto Sociale", quella, cioè, in cui è possibile contestualizzare completamente i contenuti del Web per renderli accuratamente personalizzati sul singolo utente. Il primo esempio di questo tipo di personalizzazione è stato dato da Google sia tramite la costruzione del profilo degli utenti al fine di migliorare i risultati delle loro ricerche, sia per presentare pubblicità mirata ai loro interessi. Tuttavia, entrambi questi esempi si riferiscono ad un uso dall'alto,

¹ L'accezione qui è sempre quella di sociale come in social network, non sociale nel senso di pubblica utilità.



fatto da parte dei fornitori di servizi e non basato sull'interazione e la partecipazione diretta degli utenti stessi. Il passo successivo, con il coinvolgimento diretto degli utenti nella costruzione dei dati da utilizzare, è ancora ad uno stadio larvale (e.g. il “like” di Facebook o la condivisione di link) o appannaggio di pochi progetti di ricerca ben lontani dalla maturità. La quinta era (inizio presunto 2011, maturità presunta 2013) è quella del “Commercio Sociale”, cioè un'era in cui gli attori sul mercato sono gli utilizzatori stessi e sono loro, tramite i punti di aggregazione del consenso rappresentati dai nuovi media, a definire prodotti e servizi di interesse nell'immediato futuro.

Pur in una visione ottimistica dell'evoluzione temporale, i segnali che provengono dal Web portano a pensare che le direzioni identificate dallo studio siano effettivamente quelle o, almeno, alcune di quelle su cui si sta muovendo il Web stesso. Pertanto, se anche è ancora difficile riconoscere qualcosa di più dei primi larvati accenni all'era del “Contesto Sociale” è evidente che, sia dal punto di vista tecnologico che da quello culturale, i lavori sono in corso. Concludendo, è necessario rammentare che gli aspetti più scintillanti del Social Web portano spesso a dimenticare i lati oscuri e a sottovalutare il fatto che, anche questi, siano di natura sia sociale (e.g. la disgregazione cognitiva e lo svuotamento del concetto di fiducia nelle relazioni), sia di natura culturale (e.g. l'eccessiva fiducia nelle capacità della rete di rispondere in modo tempestivo ed efficace a tutte le nostre richieste) e che comunque sottendono problemi tecnologici (*security* e *reliability* del *cloud*) che non hanno ad oggi sempre risposte soddisfacenti.



RIQUADRO 1 - Trust

In termini informatici, il concetto di trust è alla base di ogni sistema di sicurezza quale, ad esempio, il controllo degli accessi di un sistema informativo [4]. In un sistema informatico, la relazione di *trust* può essere mappata graficamente e si è usualmente espressa in termini di credenziali (*credentials*) e di entità che certificano queste credenziali (*principals*). Uno sviluppo più recente, però, tende ad eliminare l'accentramento richiesto dei *principals* e a fornire invece una struttura completamente reticolare che si incarna nei protocolli definiti dal progetto *Friend of a friend* (FOAF). Il progetto FOAF dichiara di voler costruire una rete di pagine interpretabili da sistemi automatici e contenenti la descrizione delle persone, dei contenuti che creano e delle relazioni tra loro [5][6]. Il progetto FOAF è una componente significativa del Semantic Web in quanto definisce descrizioni interpretabili da sistemi automatici. I protocolli di FOAF si appoggiano sugli standard *W3C Resource Description Framework* (RDF) e *Web Ontology Language* (OWL) per formulare i documenti scambiati dalle entità che interagiscono. Per assicurare la riservatezza delle comunicazioni, FOAF si appoggia allo standard SSL ma non richiede la presenza di certificati garantiti da entità centralizzate (*Certification Authorities*).

Se da un punto di vista meramente tecnico FOAF rappresenta un importante passo in direzione di una rete automaticamente interpretabile (in un saggio del 2007 [7] Tim Berners-Lee ridefinisce il Semantic Web ponendo le relazioni in posizione assolutamente preminente rispetto ai documenti), permane l'aspetto negativo tipico di tutte le informatizzazioni del concetto di *trust*: la dissociazione dall'effettiva relazione interpersonale. In particolare, la semplicità con cui i protocolli di FOAF permettono la costruzione di reti di pseudo-relazioni interpersonali, tende a far perdere completamente quello che è il nesso centrale di un'effettiva relazione di fiducia tra umani: il legame tra la fiducia e il vissuto condiviso. La "leggerezza" dei costrutti sintattici di FOAF si traduce sistematicamente in una "leggerezza" nell'attribuzione dell'attributo di amico (*friend*) a persone di cui, nella realtà, non conosciamo nulla di più di quello che loro ci hanno voluto raccontare di esse stesse. E' esemplare quindi il tentativo di ripristinare il valore oggettivo (in contrasto con quello mediato dagli automatismi) del concetto di *trust* da parte di progetti come Safebook [8] che, al di là di ogni meccanicismo, ribadiscono il fatto che la vera sicurezza proviene dalla garanzia che i nostri *trustee* siano persone di cui siamo realmente e fisicamente a conoscenza, e non semplici entità della rete.

RIQUADRO 2 - Condivisione, *overload* informativo e competizione

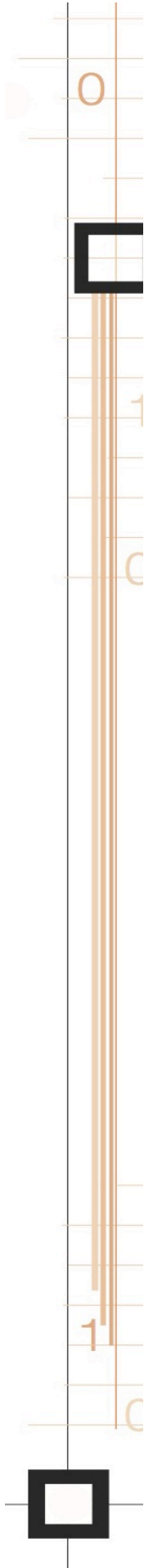
Un recente studio dello University College London ha identificato un cambiamento nelle abitudini di lettura di coloro che utilizzano le risorse on line per le ricerche. Anche in questo caso, il numero di fonti è cresciuto fortemente a scapito dell'approfondimento dedicato al singolo articolo. Se la mancanza di approfondimento può essere mitigata nel caso in cui tutte le sorgenti utilizzate sono, per loro propria natura, affidabili e analiticamente approfondite, la generazione di contenuti sotto forma di semplice condivisione e gara nel formulare prima la notizia, porta inevitabilmente ad un indebolimento del rapporto segnale/rumore nelle nostre fonti. Un indebolimento cui il nostro cervello sembra non essere ben preparato a reagire.

RIQUADRO 3 - Esistenze Virtuali

Si parla di esistenze virtuali per incapsulare in un singolo termine tutti quei fenomeni di costruzione di un *alter ego* sulla rete. Ve ne possono essere di molto semplici, come quella che si costruisce scrivendo in un blog e dipingendosi giorno dopo giorno in modo più o meno rispondente alla realtà fisica, a quelle graficamente sofisticate fornite dai siti in cui i nostri *avatar* possono spaziare liberamente in ambienti tridimensionali. Tutte, però, sono accomunate dal fatto di essere luoghi di aggregazione in cui è possibile a) cercare l'incontro con altri con cui condividere uno specifico interesse e b) presentare una versione di noi stessi ritagliata su quello stesso interesse in modo più o meno sincero. Si può notare come, anche in questa situazione, il concetto di "*trust*" viene a doversi mappare su categorie assai diverse da quelle quotidiane in quanto limitato a priori nei dati su cui basarsi.

Bibliografia

- [1] <http://www.internetworldstats.com/emarketing.html>
- [2] <http://hubpages.com/hub/How-many-webpages-do-you-think-actually-exist-on-the-Internet>
- [3] <http://www.facebook.com/press/info.php?statistics>
- [4] Aa. Vv., Computer Security, http://en.wikipedia.org/wiki/Computer_security
- [5] Aa.Vv., The Friend Of A Friend (FOAF) Project, <http://www.foaf-project.org/>
- [6] Aa. Vv., FOAF (Software), Wikipedia entry, http://en.wikipedia.org/wiki/FOAF_%28software%29
- [7] Berners-Lee T., Giant Global Graph, <http://dig.csail.mit.edu/breadcrumbs/node/215>

- 
- 
- [8] L. A. Cutillo, R. Molva, T. Strufe, *Safebook: A privacy-preserving online social network leveraging on real-life trust*, IEEE Communications Magazine, V. 47 , N. 12, 2009.
- [9] Danah M. Boyd and Nicole B. Ellison: *Social network sites: Definition, history, and scholarship*, Journal of Computer-Mediated Communication, 2007.
- [10] Lerman K.: *User Participation in Social Media: Digg Study, Web Intelligence and Intelligent Agent Technology Workshops*, 2007 IEEE/WIC/ACM International Conferences on.
- [11] N. Carr: *Is Google Making Us Stupid?* Available online at: <http://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/6868/>
- [12] M. Wolf: *Proust and the Squid: The Story and Science of the Reading Brain*, Harper, 1 edition, September 4, 2007.
- [13] S. Kumar, J. Chhugani, Changkyu Kim, Daehyun Kim, A. Nguyen, P. Dubey, C. Bienia, Youngmin Kim: *Second Life and the New Generation of Virtual Worlds*, IEEE Computer, V. 41, N. 9, 2008.
- [14] A.a.V.v. Second Life website: <http://secondlife.com/>
- [15] Qiping Zhang Marksbury, N. Heim, S.: *A Case Study of Communication and Social Interactions in Learning in Second Life*, System Sciences (HICSS), 2010 43rd Hawaii International Conference on.
- [16] K. S. Young: *Internet Addiction: The Emergence of a New Clinical Disorder*, CyberPsychology & Behavior, FALL 1998.
- [17] A. D. Rubin, D.E. Geer Jr.: *A survey of Web security*, IEEE Computer, V. 31, N. 9, 1998.
- [18] F.A. Aloul: *Information security awareness in UAE: A survey paper*, Proc. Of the 5th International Conference for Internet Technology and Secured Transactions, London (UK), Novembre 2010.
- [19] Chi Zhang Jinyuan Sun Xiaoyan Zhu Yuguang Fang: *Privacy and security for online social networks: challenges and opportunities*, Network, IEEE, 24, Volume 24, Pagina 13.
- [20] N. Lynch, S. Gilbert: *Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services*, ACM SIGACT News, Volume 33 Issue 2 (2002), pg. 51-59.
- [21] A.a.V.v., Cassandra web site: <http://cassandra.apache.org/>

[22] L.D. Paulson: *Building rich web applications with Ajax*, IEEE Computer, V. 38, N. 10, 2005.

[23] Aa.Vv., Facebook for Websites:
<http://developers.facebook.com/docs/guides/web/>

[24] Aa.Vv., The Facebook Platform:
http://en.wikipedia.org/wiki/Facebook_Platform

[25] I. Clarke, S.G. Miller, T.W. Hong, O. Sandberg, B. Wiley: *Protecting free expression online with Freenet*, IEEE Internet Computing, V. 6, N. 1, 2002.

Biografie

Marco Gaudina è nato a Genova nel 1981. Dopo aver conseguito la laurea in Ingegneria Informatica con specialistica in Robotica ed Automazione presso l'università di Genova ha realizzato numerosi progetti di carattere internazionale nel settore Industrial Automation. È attualmente PhD student presso l'Istituto Italiano di Tecnologia. I suoi interessi principali riguardano il settore della Human Computer Interaction, specializzandosi in nuove tecnologie volte al miglioramento dell'esperienza utente. I principali fattori del suo tema di ricerca riguardano ma non sono limitati a esperienza sensoriale ed emotiva, feedback sensoriali, graphical user interfaces e interfacce multimodali.

E-mail: marco.gaudina@gmail.com

Mauro Migliardi è nato a Genova nel 1966. Dopo esser stato uno dei principali ricercatori del progetto HARNESS per il meta e Grid computing presso la Emory University di Atlanta, è stato ricercatore universitario presso l'Università di Genova ed è ora Professore associato presso l'Università di Padova. Mauro Migliardi ha pubblicato oltre novanta articoli scientifici soggetti a peer-review ed ha tra i suoi principali interessi di ricerca le tecnologie e le metodologie per la progettazione e lo sviluppo di sistemi software distribuiti complessi.

E-mail: mauro.migliardi@unipd.it



OSSERVAZIONE DELLA TERRA CON SATELLITI RADAR

Enrico Saggese

Il settore delle Osservazioni della Terra via satellite è un campo ormai ben consolidato e nello scenario globale di missioni spaziali il nostro Paese è all'avanguardia per esperienza, competenze e impegno economico. La costellazione satellitare tutta italiana Cosmo-Skymed è dal novembre 2010 operativa in orbita con quattro satelliti tecnologicamente innovativi e avanzatissimi. Il loro utilizzo permette oggi all'Italia di disporre di uno strumento di monitoraggio globale, senza limitazioni geografiche ed in ogni condizione climatica, diurna e notturna. Un grande impegno del nostro Paese frutto di Ricerca & Sviluppo davvero mirata all'eccellenza.

1. Introduzione

Il settore delle Osservazioni della Terra, un campo vastissimo e ben consolidato, è quello in cui il telerilevamento copre una vastissima gamma di applicazioni, dallo studio geofisico dei terreni e dei movimenti tettonici all'analisi meteorologica e atmosferica, dal monitoraggio dei sistemi antropici a quello delle grandi masse oceaniche, dall'indagine sulle risorse agricole e forestali all'utilizzo cartografico e militare. Negli ultimi anni questa innata multidisciplinarietà, aiutata da un comparto industriale in costante sviluppo e dalla crescita di sensibilità di cittadini e governi verso le tematiche della difesa ambientale, ha cominciato ad attrarre sempre più ingenti finanziamenti, divenendo rapidamente un settore strategico e irrinunciabile della scienza orbitale. Tanto che nel 2003 in Europa le Osservazioni della Terra risultavano al primo posto per quantità di investimenti (con il 17,2 % dell'intera impresa spaziale) e negli Stati Uniti al quarto (con l'8,2%), dopo le attività dedicate al Volo Umano, alla Ricerca



quarto (con l'8,2%), dopo le attività dedicate al Volo Umano, alla Ricerca Pura e alle missioni finalizzate allo sviluppo di nuove tecnologie.

Nello scenario globale fiorente di programmi e missioni più o meno compartecipate, il nostro continente è dunque all'avanguardia per esperienza, competenze e impegno economico, e l'Italia anche presa singolarmente non fa eccezione. Non è dunque un caso se il nostro Paese nel 2007 è salito sulla ribalta del settore con un programma ambizioso, innovativo e duraturo, con il lancio del primo satellite della costellazione COSMO-SkyMed. La costellazione che è stata completata nel novembre del 2010 con il lancio del quarto satellite e proprio nei primi mesi del 2011, dopo i normali test di verifica del corretto funzionamento, risulta, pertanto, completamente dispiegata ed operativa. La costellazione Cosmo-SkyMed ben rappresenta la nuova stagione del telerilevamento, sia sul versante tecnologico "*state-of-the-art*", sia sul piano di un importante mutamento nelle politiche spaziali del Vecchio Continente sempre più portata all'utilizzo delle risorse spaziali per arricchire la conoscenza del nostro Pianeta al fine di migliorare la qualità della vita e la sicurezza dei suoi cittadini.

2. Il Programma COSMO - SkyMed

COSMO-SkyMed (*Constellation of small Satellites for Mediterranean basin Observation*) è il più grande investimento italiano nel campo dei sistemi spaziali per l'Osservazione della Terra, ideato e finanziato dall'agenzia Spaziale Italiana (ASI) e dal Ministero della Difesa (MD). Si tratta di un sistema completo il cui scopo è di fornire dati, prodotti di osservazione della terra (immagini) e servizi per un vasto numero di applicazioni sia per scopi di protezione civile che di difesa del territorio.

Infatti, una delle caratteristiche più rilevanti della missione è quella di introdurre nello scenario europeo una nuova dottrina spaziale, emersa negli ultimi anni con sempre maggiore nettezza: il cosiddetto "*Dual Use*". L'utilizzo duale di un satellite o di una costellazione, civile e militare, pur non essendo del tutto inedito in campo mondiale (ad esempio è la strada seguita dagli Stati Uniti per il loro *Global Positioning System*), tende dal punto di vista politico a farne in questo caso un modello di sviluppo incrementandone le applicazioni a diverse tipologie di missione. Per tradizione distante dalle politiche americane di armamento spaziale anche attivo, l'Europa introduce dunque un sistema misto e per così dire "temperato", in cui tuttavia, sotto la pressione internazionale della situazione post-11 settembre, il concetto di *safety* ambientale si compenetra sempre più in quello di *security*. Sfruttando le stesse risorse tecnologiche e ripartendo la spesa, l'Agenzia Spaziale Italiana condivide così un programma estremamente ambizioso con il Ministero della Difesa, realizzando il primo sistema spaziale a vocazione duale, garantendo un complesso equilibrio d'uso e ottenendo per l'Italia un ruolo decisivo nello sviluppo industriale e politico-militare dello Spazio europeo.



Ma i vantaggi del programma non si esauriscono unicamente nel prestigio dell'introduzione pratica del *Dual Use*. Il valore strategico dell'operazione sta nel nuovo posizionamento dell'Italia nello scenario dell'utilizzo dello Spazio europeo, almeno per quanto riguarda il settore delle Osservazioni della Terra, testimoniato anche dalla cooperazione con la Francia nell'ambito dell'accordo intergovernativo ORFEO e dall'accordo con l'Argentina nell'ambito del programma Siasge (Sistema Italiano-Argentino Satellitare per la Gestione delle Emergenze) costituito dalla integrazione operativa della costellazione COSMO-SkyMed e di quella argentina SAOCOM in banda L.

Infine, non va dimenticato l'impulso che con tale sistema si vuole dare all'industria italiana operante nel campo dei servizi a valore aggiunto nell'ambito dell'osservazione della terra. Le industrie nazionali potranno contare su un "asset" nazionale per sviluppare il "know-how" necessario per competere a livello internazionale in un mercato ancora di nicchia che promette di crescere in maniera vertiginosa nei prossimi anni.

Sulla base degli ottimi risultati raggiunti e sul conseguente successo riscontrato a livello internazionale dalla missione, ASI ed MD hanno deciso di dare inizio alle attività di preparazione per la progettazione e realizzazione della seconda generazione del sistema, chiamato COSMO Seconda Generazione (CSG). CSG sarà la missione che fornirà continuità operativa alle applicazioni esistenti sviluppate per COSMO-SkyMed prima generazione (CSK) e, traendo beneficio dall'esperienza maturata in CSK, permetterà di raggiungere prestazioni ancora superiori al fine di mantenere la leadership conquistata nel settore anche negli anni a venire. In tal modo l'Italia con COSMO-SkyMed è uno dei pochi paesi che consente alla propria comunità utenti, fra cui le industrie operanti nel settore, di programmare investimenti nello sfruttamento dei prodotti di EO con un orizzonte di oltre 10 anni (fino al 2023).

2.1. La missione

A partire dallo scenario sopra descritto è evidente che la missione COSMO-SkyMed è stata attentamente progettata per poter servire le esigenze di una vasta varietà di utenti analizzandone le necessità, omogeneizzando le esigenze comuni e fornendo i dati e i servizi richiesti rispettando, dove necessario, il contesto operativo in cui tali dati vengono integrati e interpretati.

Alla luce di tale analisi, l'obiettivo primario perseguito da ASI ed MD nella realizzazione del progetto è stato quello di progettare, sviluppare e operare un sistema di osservazione della Terra capace di consentire:

- la gestione dell'ambiente e del territorio, sia per applicazioni civili che di difesa, attraverso un sistema di monitoraggio e sorveglianza capace di rilevare cambiamenti endogeni ed esogeni del territorio;

- lo sviluppo di prodotti e servizi (per esempio nel campo dell'agricoltura e nella gestione del territorio) che permetta alle aziende italiane di sviluppare il “*know-how*” necessario per confrontarsi a livello mondiale nella fornitura di servizi commerciali a valore aggiunto basati sullo sfruttamento di immagini e dati satellitari.

Il tutto rispettando i necessari vincoli di tempi e costi.

2.2. Il sistema

Il sistema COSMO-SkyMed consiste di una costellazione di 4 satelliti equipaggiati con un sensore RADAR che opera in banda X e di un segmento di terra necessario a controllare il corretto funzionamento dei satelliti e a elaborare e distribuire i dati ricevuti dal sensore agli utenti finali (Figura 1).

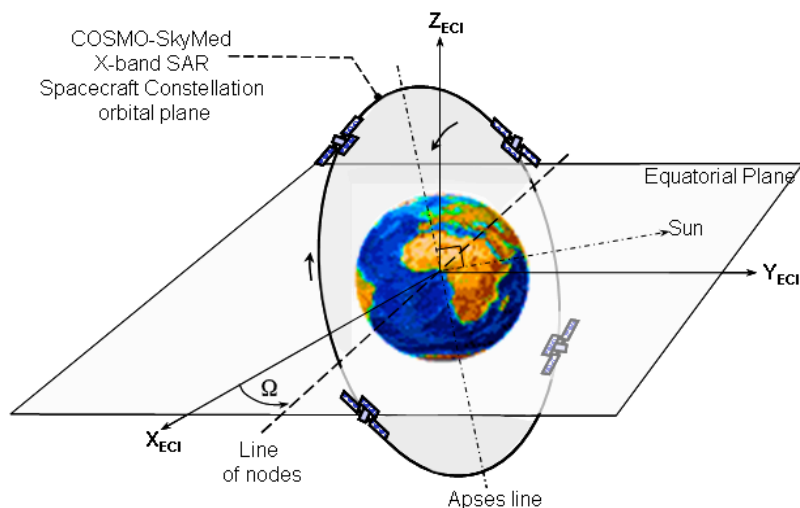


Figura 1

Configurazione orbitale nominale della costellazione COSMO-SkyMed

Le principali caratteristiche del sistema sono di seguito riassunte:

- copertura globale del Globo terrestre con la possibilità di monitorare un qualunque luogo sulla superficie della terra e con una periodicità di poche ore grazie all'opportuna scelta del piano orbitale ed al campo di vista (area di accesso) del sensore SAR;
- tempi di risposta (tempo che intercorre dalla richiesta dell'utente alla fornitura del prodotto) dell'ordine di poche decine di ore nei casi di routine e di solo qualche ora nelle situazioni di emergenza;
- capacità di monitorare un sito in qualunque condizione meteorologica (sole/nubi) e di illuminazione (giorno/notte);

- caratteristiche geometriche (risoluzione, precisione nella geolocalizzazione, ampiezza del campo di vista del sensore) e radiometriche (nitidezza, contrasto) dell'immagine all'avanguardia nel settore (particolarmente indicate per applicazioni interferometriche) in modo da facilitare l'estrazione dell'informazione dai dati e l'interpretazione degli stessi;
- elevate capacità di bordo che consentono a ciascun satellite di riprendere fino a 450 immagini al giorno;
- “Dual Use” del quale si è già parlato, che consente di fornire dati e informazioni sia per applicazioni civili che per quelle di difesa con i relativi requisiti di integrità e confidenzialità.

2.3. Capacità del sistema

Grazie alla versatilità dello strumento SAR ciascun satellite della costellazione COSMO-SkyMed può riprendere scene con diverse caratteristiche ampiezza (*swath*) / risoluzione. Le capacità di ripresa, vanno, infatti, da immagini con piccolo *swath* ad alta risoluzione a quelle con ampio *swath* a risoluzione minore a seconda delle modalità di acquisizione programmate (Figura 2). Le caratteristiche delle varie modalità di ripresa disponibili sono riassunte in tabella 1.

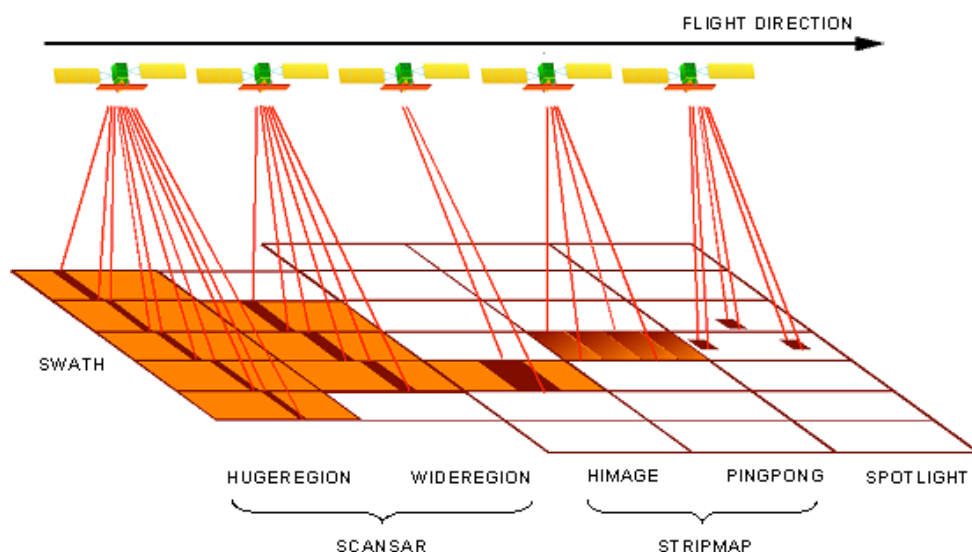


Figura 2
I modi operativi del SAR di COSMO-SkyMed

	Risol. [m]	Swath [km]	Possibile Polarizzazione. [T/R]		Angoli di Incidenza che delimitano l'area di accesso[°]
SPOTLIGHT	1 x 1	10 x 10	Singola.	selezionabile fra HH o VV	25-50 (nominale) - 20-59.5 (esteso)
STRIPMAP HIMAGE	3 x 3 5 x 5	40 x 40	Singola.	HH o HV o VH o VV	
PING-PONG	15 x 15	30 x 30	Alternata	2 pol. selezionabile fra HH, VV, VH e HV	
SCANSAR WIDE REGION	30 x 30	100 x 100	Singola.	HH or HV o VH o VV	
SCANSAR HUGE REGION	100 x 100	200 x 200	Singola.	HH o HV o VH o VV	

Tabella 1
Caratteristiche dei modi operative del SAR di COSMO-SkyMed

Da non dimenticare sono poi le capacità del *ground segment* di COSMO-SkyMed conosciute come IEM (*Interoperability, Expandability, and Multi-Mission*). In sintesi tali capacità consentono al sistema di “interfacciarsi” con altri sistemi di osservazione della terra, di espandersi consentendo ad altri partner di accedere alle risorse messe a disposizione dal segmento di terra e di “interagire” a livello di missione con altri sistemi di osservazione della terra al fine raggiungere una più completa integrazione per eseguire, per esempio, riprese combinate della stessa scena con satelliti appartenenti ad altre missioni e una elaborazione congiunta e centralizzata dei dati ricevuti. Tali capacità sono proprio quelle sfruttate nell'implementazione di accordi istituzionali per lo sfruttamento di immagini satellitari con altre nazioni.

3. Esempi applicativi

Le caratteristiche di COSMO-SkyMed e in particolare la capacità di ripresa in ogni condizione di tempo e illuminazione, l'alta risoluzione spaziale e radiometrica delle immagini SAR, la possibilità di sviluppare applicazioni interferometriche, l'alta frequenza di riveduta di una determinata regione, rendono il sistema particolarmente idoneo alle applicazioni nel capo della



gestione del rischio e delle emergenze. Infatti, in occasione di disastri, i soccorritori (governi, agenzie predisposte, organizzazioni umanitarie, etc.) hanno l'immediata esigenza di avere un quadro generale della situazione attraverso mappe tematiche ricavate prima e dopo l'evento per poter organizzare ed indirizzare in modo efficiente il proprio intervento. In questi ultimi anni sono state tante le occasioni in cui il sistema COSMO-SkyMed è stato chiamato a fornire il proprio contributo nel caso di "allarmi" ambientali e ogni volta il suo contributo è stato ampiamente riconosciuto a livello internazionale.

Sin da quando è stato lanciato il primo satellite, COSMO-SkyMed ha fornito un fondamentale, spesso unico, contributo sia nella gestione di crisi dovute a disastri naturali, sia nello studio dei mutamenti del territorio a seguito di tali eventi e nella gestione della fase susseguente alla crisi stessa. In particolare le sue caratteristiche hanno consentito di eseguire una mappa completa delle aree di crisi in poco tempo al fine di valutare quasi in tempo reale e su una vasta porzione di territorio quali fossero le aree maggiormente pericolose e danneggiate a causa di terremoti, inondazioni, frane, smottamenti, versamenti di petrolio in mare ed altri fenomeni legati a cause naturali o all'intervento dell'uomo.

Nei paragrafi successivi il contributo fornito da COSMO-SkyMed nei vari campi applicativi è descritto attraverso esempi tratti dalla vita reale di questi ultimi anni.

3.1 Terremoti: China-Sichuan (Maggio 2008), L'Aquila-Italy (Aprile 2009)

China-Sichuan Earthquake (Maggio 2008)

Il terremoto che ha investito la regione cinese di Sichuan il 12 Maggio 2008 è stato uno dei più catastrofici degli ultimi anni e ha lasciato alle sue spalle circa 68000 morti e circa 4.8 milioni di cinesi senza tetto. L'epicentro del terremoto era situato a 80 km a Nord-Est della città di Chengdu, la capitale della regione. Gli effetti sarebbero potuti essere ancora più devastati visto che nella regione vivono circa 15 milioni di cinesi.

A causa delle concomitanti condizioni climatiche con forti temporali sulla zona, è stato impossibile effettuare voli aerei sulla zona per verificare lo stato delle infrastrutture e individuare le zone più colpite dal sisma. Le autorità cinesi hanno richiesto il soccorso e la cooperazione delle autorità internazionali e l'Italia, grazie a COSMO-SkyMed ha potuto fornire dopo meno di 24 ore le prime immagini della città di Guan-Xian che contava 60.000 abitanti ed era risultata fino ad allora inaccessibile.

Tali immagini erano completate dalle prime immagini della diga sul fiume Min che sovrasta la città e che aveva causato parecchie preoccupazioni ai soccorritori che non avevano nessuna informazione riguardo la sua integrità. Nei giorni successivi sono state acquisite ed elaborate dal sistema decine di immagini sulla regione, in particolare sulle zone montuose,

impervie e inaccessibili attraverso i mezzi convenzionali. Grazie a questo enorme contributo fornito dalle immagini satellitari di COSMO-SkyMed, le autorità cinesi hanno richiesto una mappatura completa della regione che è stata eseguita in meno di 12 giorni, a riprova delle reali utilità delle performance offerte dal sistema di cui si è parlato nei paragrafi precedenti.

La figura 3 mostra a sinistra la prima immagine di Guan-Xian con la sua diga in evidenza (in rosso) e un'immagine multi temporale di Chengdu a destra. Questo tipo di immagine permette agli esperti di individuare attraverso opportuni codici di colore quali cambiamenti sono intervenuti nella scena fra un'immagine ripresa prima dell'evento e una ripresa dopo l'evento consentendo di scoprire le zone che hanno subito maggiori danni o più in generale quelle che necessitano di immediati accertamenti per scoprire l'entità e la pericolosità di tali modifiche.

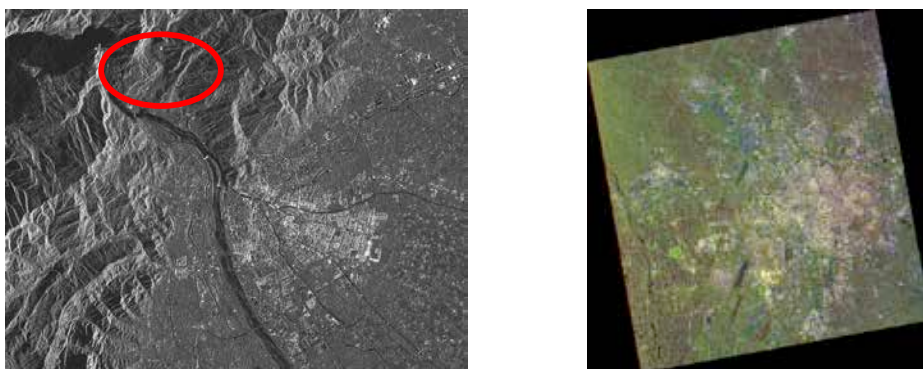


Figura 3

Immagine di Guan-Xian con la diga evidenziata (sinistra) e immagine multi temporale di Chengdu (destra)

L'Aquila - Abruzzo Earthquake (Aprile 2009)

Il 6 Aprile 2008 un terremoto di 5.8 gradi di magnitudo ha colpito la regione Abruzzo nel centro Italia. La scossa principale si è verificata alle 3:32 ora locale e il suo epicentro è stato localizzato vicino L'Aquila, il capoluogo della regione, la cui provincia ha sofferto dei maggiori danni.

Grazie all'elevata frequenza di rivisita della costellazione (a suo tempo composta da soli 3 satelliti), COSMO-SkyMed è stato il primo sistema di osservazione della terra a riprendere la zona dopo il sisma fornendo le prime immagini a poche ore dall'evento.

Sin da subito i dati forniti dal sistema sono stati utilizzati per monitorare gli effetti del terremoto e sono state utilizzate per calcolare una mappa di deformazione del terreno dovuta al terremoto. Nel corso del primo mese, dopo la scossa del 6 aprile, grazie alla estrema flessibilità del sistema e alle sue capacità di rivisita, sono state acquisite più di 100 immagini della zona colpita (3 al giorno in media) da diverse orbite e con diversi angoli di vista. Inoltre la disponibilità di immagini di archivio ha reso possibile applicare sofisticate tecniche di elaborazione per estrarre un modello digitale della faglia tettonica (già nota ai geologi) e della sua dislocazione (Figura 5). Il

piano di faglia con un'inclinazione di circa 50° passa sotto la città de L'Aquila. Durante il terremoto il piano di faglia è scivolato in basso per un massimo di 90 cm a 4 km di profondità sotto la città de L'Aquila producendo uno smottamento superficiale visibile nell' interferogramma mostrato in Figura 4.

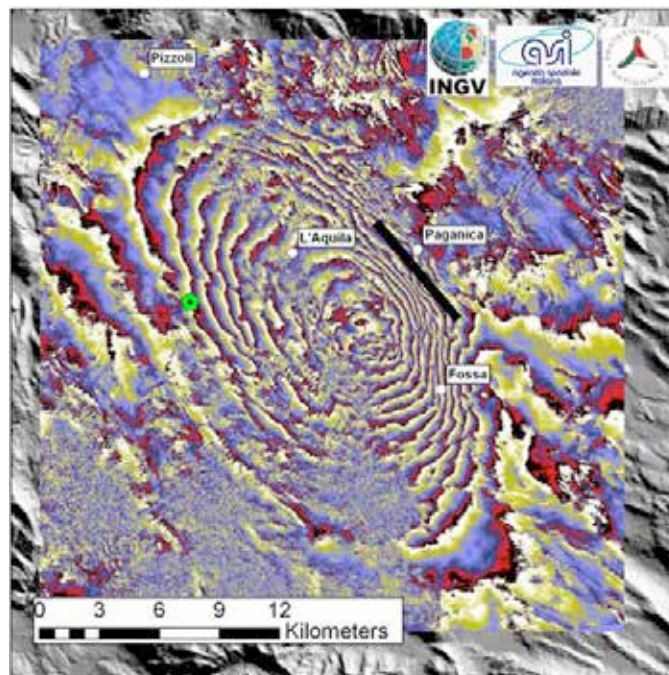


Figura 4

Interferogramma differenziale, il punto verde indica il luogo con maggiore intensità mentre la linea nera indica la faglia paganica

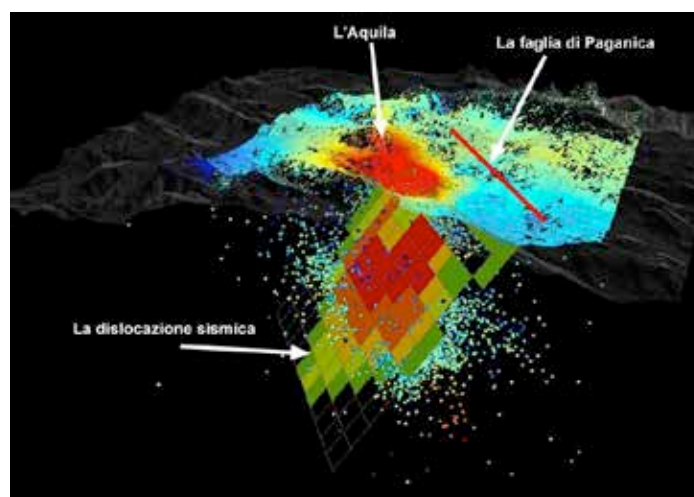


Figura 4

Ricostruzione digitale 3D del piano di faglia estratto utilizzando dati COSMO-SkyMed

3.2 Inondazioni: Haiti (2008), Bangladesh (2009), Australia (2011)

Uragani Hanna e Ike - Haiti (Settembre 2008)

Nel 2008 l'isola di Haiti è stata colpita da quattro uragani che hanno completamente devastato il territorio e le colture, causando circa 800 morti. Le immagini di CoSMO-SkyMed hanno testimoniato le devastazioni provocate dagli uragani nominati Hanna e Ike proprio all'inizio del 2008, fornendo, nuovamente, un importante supporto per la stima dei danni subiti dalle zone inondate all'organizzazione ONU World Food Programme (Figura 6). Le mappe delle aree inondate, uno strumento indispensabile per gli attori coinvolti in simili disastri, sono state rese disponibili poche ore dopo la loro acquisizione e sono state intensivamente utilizzate per determinare l'estensione dell'inondazione, la sua evoluzione e l'individuazione di vie di accesso sicure per i servizi di soccorso.

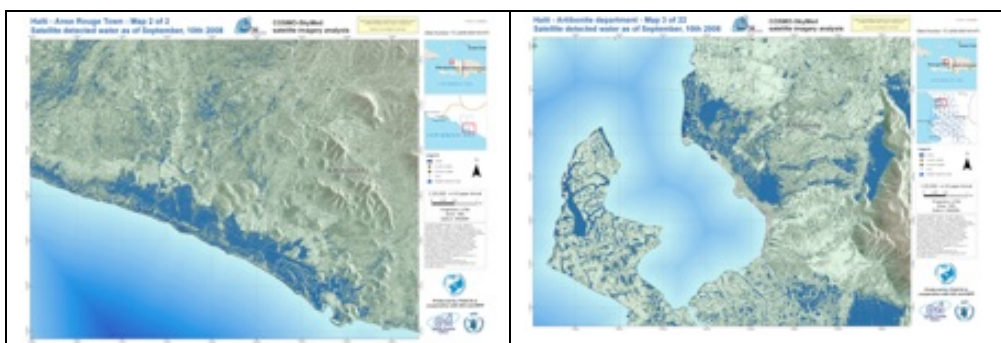


Figura 6

Mappe delle aree inondate di Haiti basate su dati COSMO-SkyMed

Bangladesh (Maggio 2009)

Il ciclone Aila ha colpito le coste occidentali del Bengala il 25 Maggio 2009, lasciando una scia di distruzione alle su spalle, causando la morte di 190 persone e ferendone migliaia. Una potente tempesta con venti che soffiavano a 110 km/hr accompagnata da forti precipitazioni ha preceduto il ciclone. Circa un milione di persone sono rimaste senza tetto e circa 4.8 milioni di persone sono state in qualche modo direttamente interessate dai danni causati da questo evento straordinario con circa 3 milioni di acri di coltivazioni che sono state distrutte.

La costellazione COSMO-SkyMed ha acquisito un gran numero di immagini per lo più ScanSAR ad ampio *swath* e bassa risoluzione. Anche in questo caso, sebbene il sistema fosse ancora incompleto (3 satelliti operativi) è stato evidente il vantaggio di avere una costellazione di satelliti che lavora in sinergia, infatti in meno di 72 ore sono state acquisite 8 immagini delle zone interessate dalle inondazioni causate dal ciclone. L'analisi delle immagini ha consentito di individuare le regioni ancora interessate da inondazioni 5 giorni dopo l'evento, evidenziando le aree dove l'inondazione stava regredendo (poligoni arancioni in Figura 7) e permettendo di estrarre

informazioni molto dettagliate per mappe tematiche da fornire all'organizzazione internazionale dell'ONU World Food Programme che coordinava le operazioni di soccorso.

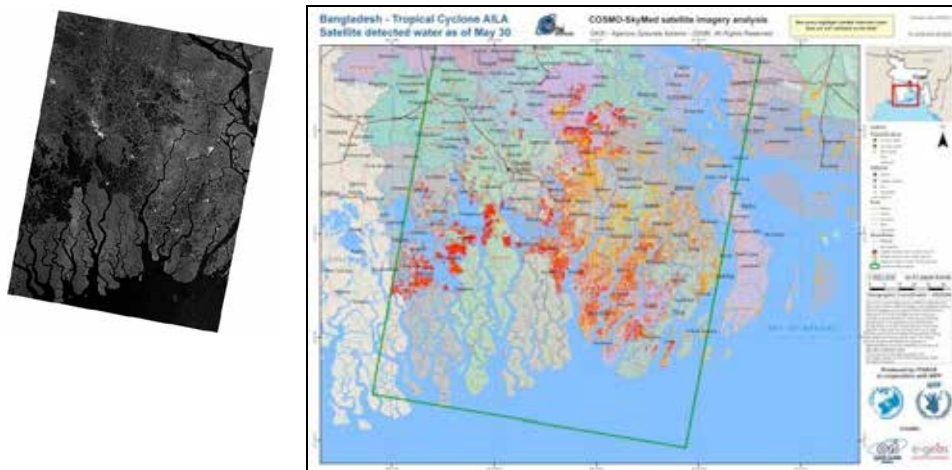


Figura 7

(Sinistra) Prima immagine di cosmo SkyMed del Bangladesh (le zone più scure sono quelle interessate da inondazioni) - (Destra) Mappa tematica delle aree inondate creata con dati di COSMO-SkyMed

Australia (Dicembre 2010 – Gennaio 2011)

Una serie di inondazioni ha colpito lo stato del Queensland in Australia a dicembre 2010 e nei primi mesi del 2011. L'inondazione ha causato l'evacuazione forzata di migliaia di persone da città e paesi della regione. Almeno 70 paesi e più di 200.000 persone sono state colpite, in circa tre quarti della nazione è stato dichiarato lo stato di calamità naturale e circa 50 persone sono rimaste uccise.

Dalle immagini COSMO-SkyMed attraverso consolidate procedure sono state estratte le mappe delle regioni alluvionate (Figura 8 e Figura 9). Tali mappe sono state fornite alle autorità competenti meno di sei ore dopo la loro acquisizione in modo da supportare le operazioni di evacuazione e l'organizzazione degli aiuti.

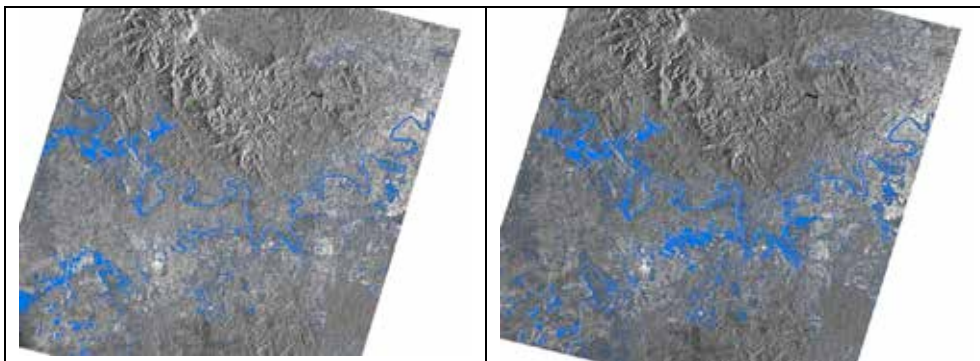


Figura 8

Evoluzione dell'inondazione visibile da due immagini della stessa area riprese a distanza di un giorno (11 e 12 gennaio 2011)

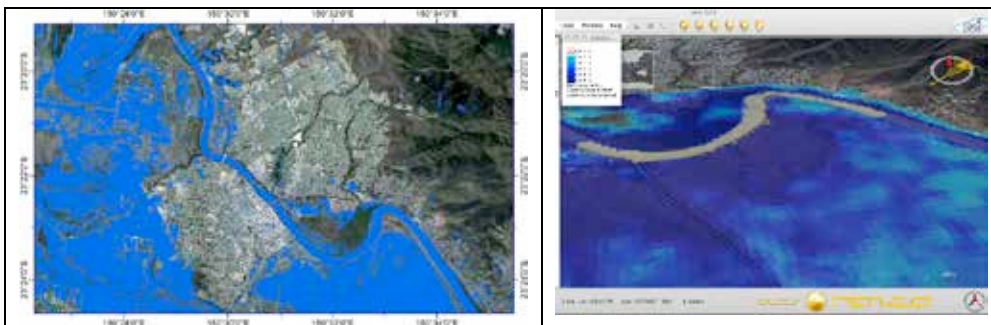


Figura 9

Mappe (2D e 3D) della regione inondata basate su immagini RADAR di CSOMO-SkyMed

3.3 Monitoraggio delle coste e inquinamento marino: Golfo del Messico – Luoisiana (Aprile – Giugno 2010)

Nel mondo moderno basato sul petrolio, la maggior parte del trasporto di idrocarburi avviene via mare e gli incidenti sono inevitabili. Ogni anno vengono disperse in mare circa 4.5 tonnellate di petrolio con un notevole impatto sull'eco-sistema mondiale. È, però, interessante notare come della quota totale di versamenti di petrolio in mare solo il 7% è dovuto al naufragio di petroliere o a incidenti su piattaforme petrolifere, mentre la causa con maggior impatto è l'intenzionale scaricamento di petrolio in mare durante le operazioni di pulizia delle navi cisterna seguita dal versamento proveniente dalle raffinerie di terra.

Il sottile strato di petrolio che galleggia sulle acque attenua fortemente l'increspatura del mare dovute ai venti superficiali e tale fenomeno è facilmente visibile attraverso le immagini RADAR dove la chiazza di petrolio viene vista proprio come una macchia scura sulla superficie grigia del mare.



Il sistema COSMO-SkyMed implementa la modalità di acquisizione ScanSAR (ad ampio *swath* e bassa risoluzione) che è l'ideale per il monitoraggio dell'inquinamento marino causato da idrocarburi, permettendo di collegare (attraverso l'elaborazione dell'immagine satellitare) una chiazza di petrolio all'imbarcazione che lo ha generato e consentendo in tal modo di perseguire e punire queste operazioni fraudolente.

Tra i vari esempi di monitoraggio dell'inquinamento da idrocarburi effettuato da COSMO-SkyMed a partire dal lancio del primo satellite, l'evento più significativo è certamente quello relativo al più grande disastro di questo tipo nella storia dell'umanità, vale a dire l'esplosione della piattaforma Deepwater Horizon avvenuta il 20 Aprile 2010 a largo delle coste della Louisiana nel Golfo del Messico. L'esplosione della piattaforma della British Petroleum ha creato una falla da cui è fuoriuscito greggio alla velocità di 200.000 galloni al giorno per la durata complessiva di circa 3 mesi ininterrottamente. La perdita è stata, alla fine quasi del tutto fermata il 15 Giugno dopo che circa 780.000 m³ di greggio erano stati riversati in mare.

Durante l'evoluzione degli eventi COSMO-SkyMed ha fornito preziosi dati al *Center for Southeastern Tropical Advanced Remote Sensing* (CSTARS) dell'università di Miami consentendo di registrare l'evoluzione della enorme chiazza di greggio e acquisendo diverse immagini al giorno dell'area colpita. È da rilevare che la CSTARS ha usato un numero di immagini di COSMO-SkyMed tre volte maggiore di quello di un qualunque altro sistema satellitare, fornendo le mappe ricavate dall'elaborazione di queste immagini alle autorità competenti del Governo degli Stati Uniti e del Messico. La figura 10 mostra alcune immagini rilevanti di tale evento e la sua evoluzione nel corso del primo mese.

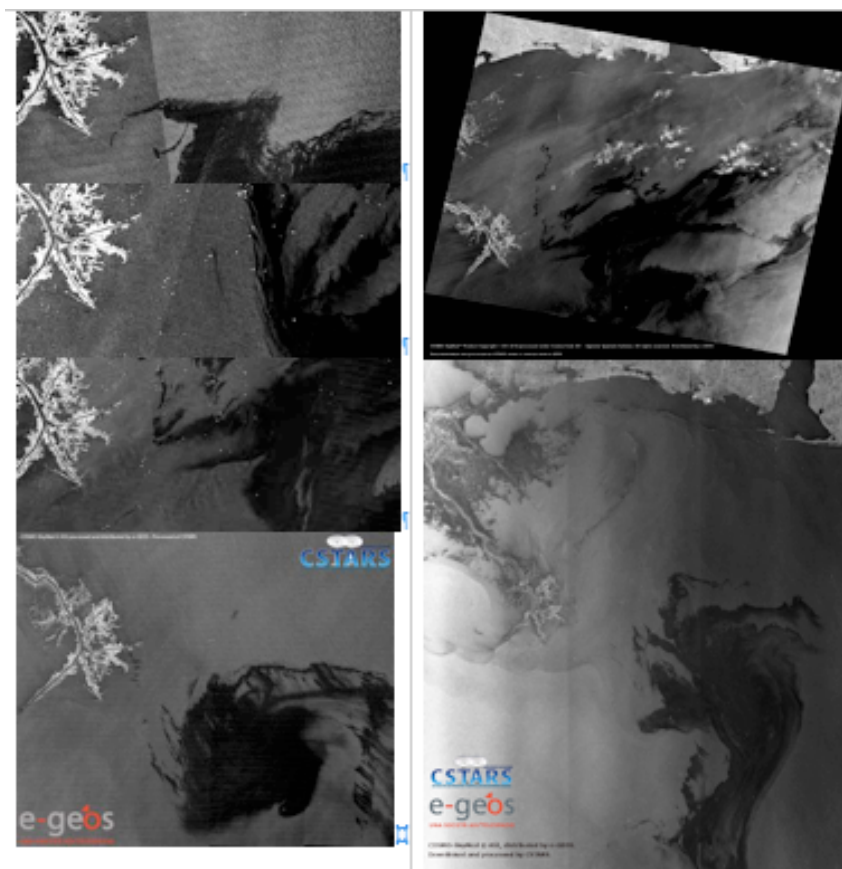


Figura 10

Immagini dell'evoluzione della chiazza di petrolio risultante dall'esplosione della Deepwater horizon nel Golfo del Messico (Aprile - Maggio 2010)

3.4 Monitoraggio di colture e delle foreste

COSMO-SkyMed costituisce una importante risorsa per applicazioni di monitoraggio del territorio agricolo e delle foreste, in quanto fornisce mappe delle colture con informazioni sul ciclo di crescita e della qualità di prodotti utili per definire i migliori periodi di irrigazione o fertilizzazione dei campi. In aggiunta a ciò, la costellazione di satelliti contribuisce a monitorare costantemente lo stato delle foreste che costituiscono i polmoni del nostro pianeta e a misurarne la velocità di deforestazione.

La figura 11 mostra alcuni esempi di immagini multi temporali che evidenziano il ciclo di crescita delle colture agricole, mentre la figura 12 evidenzia alcuni esempi di deforestazione in Sudamerica e in Europa.

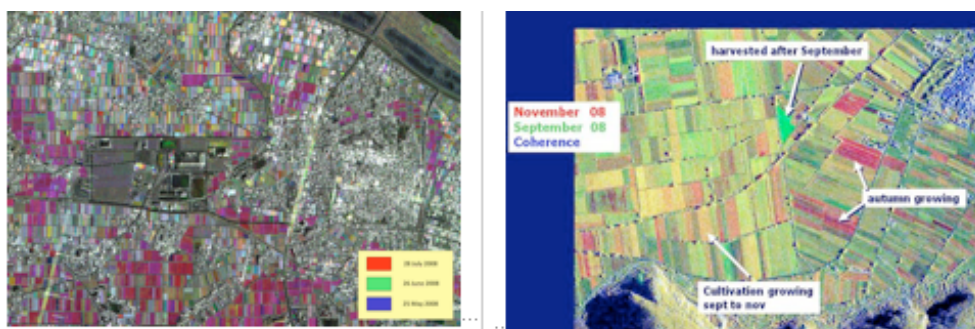


Figura 11

(Sinistra) Immagine multi temporale che mostra il ciclo di crescita delle piantagioni di riso nella regione di Kumagaya (Japan) – (Destra) Immagine multi temporale di un territorio agricolo in Abruzzo (Italia)

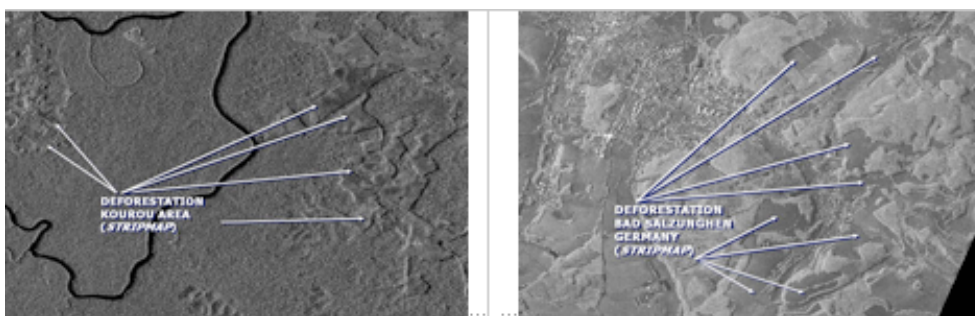


Figura 12

(Sinistra) Immagine con segni evidenti di deforestazione nell'area di Kourou (Guyana Francese) - (Destra) Immagine di deforestazione in Europa (Bad Salzungen – Germania)

3.5 Monitoraggio delle aree urbane

Grazie alle caratteristiche del sensore SAR e alla capacità di riprendere immagini ad alta risoluzione, con COSMO-SkyMed è possibile eseguire riprese di edifici, città, strutture ed infrastrutture come strade e ponti. In generale è pertanto possibile effettuare il monitoraggio di aree urbane per studiarne i cambiamenti sia indotti dall'attività umana che da cause naturali. In particolare la tecnica dell'interferometria differenziale (DInSAR) consente di misurare anche millimetrici movimenti del terreno. Inoltre la tecnica dei *permanent scatterers* (PS) permette la realizzazione di mappe di deformazione su vaste aree delle città consentendo per esempio di misurare gli effetti dello sbancamento effettuato per costruire una linea metropolitana e di prevenire crolli di edifici.

La Figura 13 mostra la mappa di subsidenza della città di Roma dove è possibile apprezzare le zone della città che sono più stabili e quelle invece maggiormente soggette a movimenti. Ovviamente, come tutte le misure geodetiche, tali mappe devono essere interpretate da geologi, geofisici o

ingegneri in modo da estrarre il fattore di rischio relativo all'area sotto osservazione: una zona rossa sulla figura non fornisce da sola sufficienti ragioni di preoccupazione così come una zona verde non è garanzia di sicurezza.

Nel caso di COSMO-SkyMed, l'alta risoluzione delle immagini consente di identificare addirittura singoli edifici e di estrarre da una serie di immagini, riprese per esempio nel corso di un anno, l'evoluzione degli spostamenti subiti dalla struttura in Figura 14.

Usando la stessa metodologia è possibile effettuare la stima dell'altezza degli edifici, rendendo pertanto questa tecnica particolarmente idonea per rilevare cambiamenti nella configurazione degli edifici a fini catastali. La Figura 15 mostra un esempio di tale applicazione con il relativo errore nella misura realizzato attraverso dati acquisiti dalla costellazione COSMO-SkyMed.

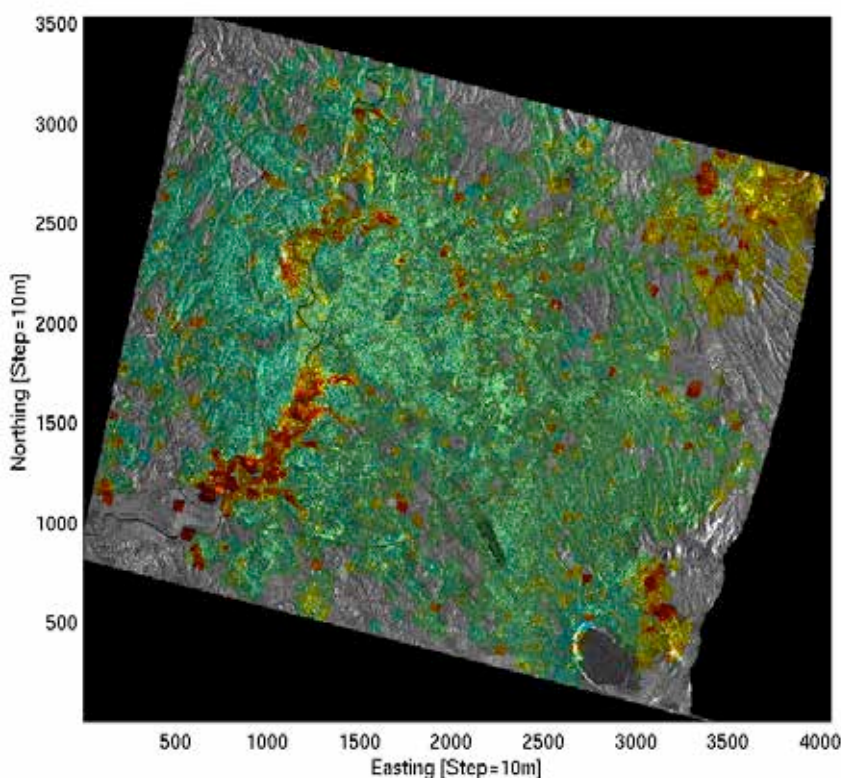


Figura 5

Dati SAR: mappa di subsidenza della città di Roma (verde: zone stabili; rosso: zone maggiormente soggette a movimenti) (elaborazione T.R.E.)

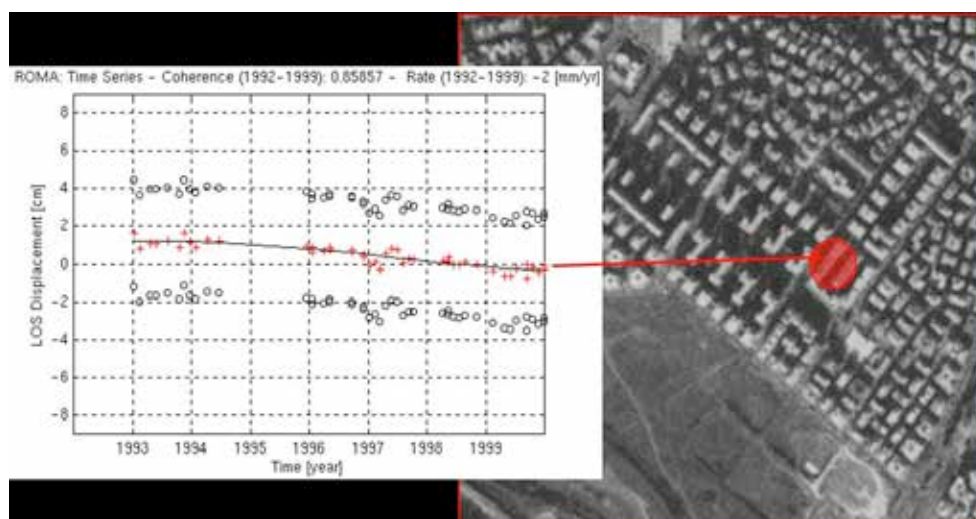


Figura 6
Dati SAR: linea di evoluzione degli spostamenti di un particolare edificio nel corso degli anni (elaborazione T.R.E.)



Figura 7
Dati COSMO-SkyMed: stima dell'altezza dell'edificio (Elaborazione e-Geos)

4. Cooperazioni internazionali

Sin dalla sua fase di progetto e sviluppo la missione COSMO-SkyMed è stata concepita per supportare al meglio le presenti e le future collaborazioni con sistemi di monitoraggio satellitare di altre nazioni. Esempi delle attuali cooperazioni internazionali in corso sono:

- SIASGE: il sistema italo-argentino per la gestione delle emergenze che mira ad integrare il sistema COSMO-SkyMed che fornisce immagini e dati acquisiti dal RADAR in banda X con il sistema argentino SAOCOM che fornisce immagini e dati acquisiti mediante un RADAR in Banda L.
- ORFEO: il sistema italo-francese che mira allo sfruttamento congiunto delle immagini RADAR in banda X fornite da COSMO-SkyMed con le immagini ottiche fornite dai satelliti francesi Pleiades.

L'esperienza nell'implementare simili cooperazioni è unica e universalmente riconosciuta come dimostrano i riconoscimenti ricevuti da varie organizzazioni internazionali nell'ambito dei progetti GMES, Galileo e MUSIS.

Inoltre nel settembre 2009 l'Agenzia Spaziale Italiana e l'Agenzia Spaziale Giapponese (JAXA) hanno firmato un *Memorandum of Understanding* (MoU) per lo scambio di dati fra COSMO-SkyMed e ALOS per il monitoraggio ambientale e per la gestione di eventi disastrosi. Durante i primi mesi di cooperazione sono state eseguite due esercitazioni per il coordinamento operativo dei due sistemi al fine di acquisire immagini su una stessa area e processare i dati risultanti al fine di integrare le informazioni estraibili dalle immagini provenienti dai due sistemi.

Particolarmente significativo è stato il contributo dato da COSMO-SkyMed nell'emergenza causata dal terremoto di circa magnitudo 9.0 (scala Richter) verificatosi in Giappone lo scorso venerdì 11 Marzo 2011 alle ore 14:46 locali (06:46 ora italiana), e dal conseguente tsunami che ha investito alle ore 15:57 locali la costa nord-orientale del Giappone. Nell'ambito della cooperazione con JAXA ASI, a seguito dell'alert inoltrato da JAXA stessa, ha immediatamente intrapreso una intensa attività di coordinamento e ottimizzazione delle richieste di acquisizione di immagini COSMO-SkyMed sulle aree più colpite. Le prime immagini acquisite sulle zone dell'evento in Giappone sono subito state rese disponibili alla JAXA e va sottolineato come l'Agenzia Spaziale Italiana, allo scopo di offrire il massimo supporto alla nazione colpita dal terribile evento sismico, abbia fornito un numero di immagini molto superiore al quantitativo direttamente richiesto dall'Agenzia Spaziale Giapponese. Per la tempestività, la quantità e la qualità delle informazioni ottenute, JAXA e il Governo del Giappone hanno ripetutamente ringraziato l'ASI per il fondamentale aiuto fornito in occasione di un evento così catastrofico per il Giappone.

5. Conclusioni

COSMO-SkyMed è uno dei principali esempi di sistemi di osservazione della terra ad uso duale, cooperante e interoperante con altri sistemi omologhi di altre nazioni con lo scopo di fornire servizi integrati multi missione alla comunità di utenti più ampia possibile.

Queste caratteristiche progettuali sono fortemente radicate nella estrema versatilità degli elementi cardine del sistema: la costellazione di 4 satelliti, il *payload* RADAR ad alta risoluzione e il segmento di terra.

Queste caratteristiche fanno di COSMO-SkyMed un sistema capace di fornire una “conoscenza istituzionale” in modo da formulare le corrette decisioni, rispondere e gestire in maniera appropriata le situazioni di crisi che si dovessero prospettare in qualunque area del mondo permettendo di garantire la sicurezza e la salute alla popolazione.

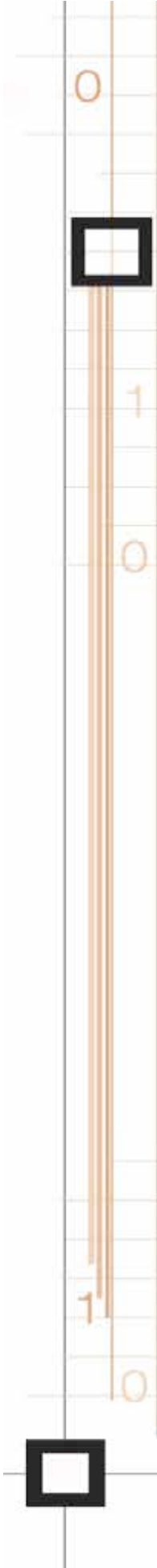
COSMO-SkyMed rappresenta uno dei maggiori contributi italiani a tali esigenze, un sistema all'avanguardia che rappresenta una pietra miliare nella realizzazione di sistemi di questo tipo.

Inoltre, il programma fornisce una reale opportunità per la ricerca sia nel campo delle tecniche di elaborazione dati per l'estrazione di informazioni dalle immagini satellitari sia per quanto riguarda la tecnologia per la realizzazione di sistemi satellitari in accordo con la strategia nazionale per lo sviluppo di la realizzazione di sistemi duali. Tutti gli investimenti effettuati hanno infatti restituito sufficienti ritorni che hanno permesso all'Italia e alla sua industria di assumere una posizione di rilievo a livello internazionale

Infine con COSMO-SkyMed ASI e il Ministero della Difesa stanno agendo come dei veri e propri *spin-off* nei confronti di industrie che forniscono servizi a valore aggiunto nel campo dello sfruttamento dei dati provenienti da sistemi satellitari di osservazione della Terra con l'auspicio di fornire impulso ad un mercato potenzialmente florido e ai conseguenti benefici sociali che deriverebbero dal suo sviluppo.

Bibliografia

- [1] F. Covello, F. Battazza, A. Coletta, E. Lopinto, L. Pietranera, G. Valentini, S. Zoffoli: *COSMO-SkyMed an existing opportunity for observing the Earth*, Journal of Geodynamics 49 (2010) 171–180.
- [2] A. Coletta, G. Valentini G., A. Capuzi, F. Caltagirone, M. De Carlo, G. De Luca, F. Battazza, F. Covello: *Il programma COSMO-SkyMed: Descrizione della missione e del sistema e primi risultati*, Rivista Italiana di Telerilevamento, Vol. 40, no.2, pp5-13, 2009.
- [3] F. Battazza, A. Ciappa, A. Coletta, F. Covello, G. Manoni, L. Pietranera, G. Valentini: *COSMO-SkyMed Mission: a set of X-band SAR Applications conducted during 2008*, International Journal of Remote Sensing, 2009.
- [4] A. Coletta, G. Angino, F. Battazza, F. Caltagirone, F. Impagnatiello, G. Valentini, A. Capuzi, S. Fagioli, R. Leonardi: *COSMO-SkyMed Program: Utilization and Description of an Advanced Space EO Dual-Use Asset*, Proceedings of ENVISAT Symposium 23-27 April 2007, Montreux (Switzerland).

- 
- 
- [5] A. Ciappa, L. Pietranera, A. Coletta, Xingwei Jiang: *Surface transport detected by pairs of COSMO-SkyMed ScanSAR images in the Qingdao region (Yellow Sea) during a macro-algal bloom in July 2008*, Journal of Marine Systems, 2009.
- [6] S. Salvi, S. Vignoli, M. Serra, V. Bosi, and the SIGRIS Team: *Use of COSMO-SKYMED data for seismic risk management in the framework of the ASI-SIGRIS project*, Proceedings of IGARSS 2009.
- [7] F. Caltagirone, A. Capuzi, R. Leonardi, S. Fagioli, G. Angino, F. Impagnatiello: *COSMO - SkyMed: An Advanced Dual Use System for Earth Observation*, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium - July 23-27, 2007, Barcelona (Spain).
- [8] A. Coletta, G. Angino, F. Battazza, F. Caltagirone, F. Impagnatiello, G. Valentini, A. Capuzi, S. Fagioli, R. Leonardi: *COSMO-SkyMed Program: Utilization and Description of an advanced Space EO Dual-Asset*, Proceedings of 58th International Astronautical Congress 2007, Hyderabad, India, 24-29 September 2007.

Biografia

L'Ing. Enrico Saggese è dal 3 luglio 2009 Presidente dell'Agenzia Spaziale Italiana (ASI), che in precedenza dal luglio 2008 aveva diretto in qualità di Commissario Straordinario. Le sue esperienze professionali lo vedono da oltre trent'anni impegnato nel settore aerospaziale con importanti incarichi di vertice in aziende legate sia alla progettazione e realizzazione di sistemi spaziali, che di vendita di servizi ed applicazioni da essi derivati.



II CASO WIKILEAKS

Esempio di libertà e democrazia nel cyberspazio o strumento di condizionamento psicologico?

Antonio Teti

Non vi sono più dubbi sulle straordinarie capacità della rete quale strumento di informazione - disinformazione globale, ma la maggiore peculiarità risiede nella straordinaria azione di condizionamento e persuasione che può essere esercitata nel cyberspazio. Uno degli esempi in tal senso, è rappresentato da Wikileaks, il più discusso, osannato e criticato portale informativo che negli ultimi anni ha creato non poco scompiglio a governi, personaggi noti, organizzazioni diverse e strutture istituzionali di molteplici nazioni. Julian Assange ne è il dominatore assoluto e ha dimostrato come il web possa trasformarsi in un'arma mediatica di inaudita potenza.

1. Il fenomeno Wikileaks

Sono in molti a sostenere che Internet, soprattutto grazie ai *social network*, stia soppiantando i media classici (radio, tv e giornali) nell'azione di comunicazione e condizionamento delle masse a livello mondiale. Per qualsiasi individuo o organizzazione, il web rappresenta la piattaforma di comunicazione ideale su cui costruire il proprio consenso o per conseguire il proprio scopo, obiettivi che hanno lo stesso comune denominatore: la società del *cyberspazio*. Se poi la comunicazione di massa online può anche essere arricchita da sensazionalismi o annunci spettacolari, il livello di notorietà (e quindi anche di credibilità prodotta) può raggiungere livelli inaspettati. Al momento, uno degli esempi più interessanti, è senz'altro



costituito dal più noto e discusso portale web dedicato proprio alla diffusione di documenti più o meno riservati: il portale di Wikileaks.

Personalmente ritengo che il successo mondiale, ottenuto soprattutto negli ultimi mesi da questo discusso portale, rappresenti uno dei più straordinari esempi di condizionamento mediatico in rete che siano mai stati concepiti finora. Dietro il più famoso portale di “informazione libera” come viene concepito dal mondo della rete, si cela un’organizzazione *no-profit* (così viene definita sul sito www.wikileaks.org) che ha deciso di fare dell’informazione vera e della verità oltre ogni limite, la sua personalissima bandiera.

Da ciò ne deriva che il naturale palcoscenico di Wikileaks sia rappresentato dal *cyberspazio* e dalla sua popolazione virtuale. La *mission* di Wikileaks è chiaramente indicata nella citazione del Time Magazine, che appare nella pagina *Introduction: “Poter diventare un importante strumento giornalistico come il Freedom Information Act¹”*.

Nata da un’idea di un personaggio particolarmente complesso, l’australiano trentanovenne Julian Assange, l’organizzazione riesce a finanziare tutte le sue attività (che sono prevalentemente quelle di pubblicare documenti scottanti e riservati), grazie ad un complesso e nebuloso sistema di contribuzioni anonime e volontarie. Un ulteriore alone di mistero è costituito non solo dalla nebbiosità che circonda l’intera infrastruttura organizzativa di Wikileaks e il suo fondatore, ma anche dalla complessa e capillare infrastruttura dei sistemi informatici su cui vengono raccolti e custoditi i documenti reperiti dal portale. Naturalmente, l’aspetto più inquietante è senz’altro riconducibile alla facilità con cui Julian Assange riesce ad avere documenti apparentemente segreti o scottanti.

Sembra che questi documenti vengano acquisiti da Assange senza particolari difficoltà per essere riversati, successivamente, in rete con l’unico intento di *“tutelare e garantire la verità assoluta e la democrazia della collettività mondiale”*. Anche l’enigmatico fondatore conduce una vita di assoluta riservatezza, in apparenza lontano da telecamere e riflettori, ma sempre pronto a conquistare l’attenzione dei media quando si tratta di fare dichiarazioni eclatanti o lanciare mirate accuse su ipotetici complotti rivolti contro la sua persona. A pochissime individui è consentito avere contatti con lui e tutte le sue comunicazioni si svolgono attraverso algoritmi di cifratura dei messaggi.

¹ Il Freedom of Information Act (FOIA) è una legge emanata negli stati Uniti il 4 luglio 1966 dal presidente Lyndon B. Johnson che impone la libertà di informazione. Essa obbliga le amministrazioni pubbliche a rispettare una serie di norme per consentire a chiunque di accedere ad informazioni sull’operato del [Governo, ivi compreso](#) l’accesso totale o parziale a documenti classificati.

Wikileaks, pur essendo attivo da anni, solo nel 2010 conquista improvvisamente un ruolo primario su giornali e televisioni, grazie alla pubblicazione di un video che lo fa balzare alla ribalta delle cronache: la ripresa video di un attacco aereo in Iraq.

La data è del 5 aprile 2010. In quel giorno nel sito viene presentato un video registrato da un elicottero militare da combattimento statunitense durante un'azione condotta a Baghdad nel 2007. Nel video viene riportata una sequenza in cui si vede l'elicottero aprire il fuoco su un gruppo di presunti terroristi visibilmente armati. Nel gruppo, in realtà, ci sono anche alcuni giornalisti, erroneamente scambiati per terroristi ostili per via delle macchinette fotografiche portate a tracolla e munite di teleobiettivi facilmente scambiabili per armi da fuoco. Nella brevissima azione militare perdono la vita 11 persone tra cui un fotografo e un autista della Reuters. E' il classico triste caso di "friend fire"², ma il video viene presentato in modo da esaltare una apparente superficialità da parte delle forze americane nella conduzione dell'attacco, cosa che suscita un sentimento di orrore e di scalpore a livello mondiale. Le reazioni di protesta e di sdegno giungono da ogni angolo del mondo, soprattutto perché nel video si vede anche il ferimento di due bambini. Nel corso dei mesi successivi, gli scoop di Wikileaks si susseguono. Sul portale di Assange vengono pubblicati ulteriori documenti particolarmente imbarazzanti, soprattutto per il governo USA, sulla guerra in Iraq ma anche relativi al conflitto in Afghanistan. Nelle migliaia di documenti vi sono anche informazioni che coinvolgono altri paesi occidentali e lo sgomento maggiore è provocato dal fatto che la quasi totalità dei documenti pubblicati viene ritenuta segreta o *classificata* cioè munita di un livello di riservatezza che ne dovrebbe garantire la non diffusione al pubblico. Lo stupore e l'imbarazzo generale per le rivelazioni fatte da Wikileaks sono incontenibili e la figura del suo fondatore comincia ad assumere un ruolo di particolare rilievo e importanza nei sistemi di comunicazione internazionale.

2. L'anima di Wikileaks: Julian Assange

Senza alcun dubbio, i punti di forza di questa organizzazione dedita alla diffusione di documenti riservati riconducibili ad avvenimenti che si verificano nel mondo intero, sono riferibili a due elementi caratteristici: il primo è rappresentato dalla capacità di entrare in possesso e di rendere disponibili in Internet, documenti in grado di mettere in forte imbarazzo uomini, governi e organizzazioni mondiali; il secondo è la figura del suo fondatore, un personaggio misterioso e alquanto discutibile, considerato da alcuni come un difensore della libertà di stampa e di informazione e da altri un autentico pericolo pubblico.

Si tratta di Julian Assange, che nasce nel 1971 a Queensland, in Australia da una poliedrica coppia di artisti dedita della gestione di una compagnia teatrale. Dopo un'infanzia non proprio tranquilla (la madre si risposa nel

² *Friend fire* (fuoco amico). In guerra e in tutti i conflitti a fuoco si verifica spesso, ed è riconducibile a tutti quegli eventi in cui le forze amiche si confondono con quelle nemiche.



1979 con un musicista appartenente ad una setta dal nome Santiniketan Park Association, anche conosciuta come The Family e The Great White Brotherhood e diretta da Anne Hamilton-Byrne), il giovane Julian nel 1987 abbandona la famiglia (la madre nel frattempo ha un altro figlio) e inizia un lungo e solitario peregrinare, trasferendosi da una città all'altra e frequentando più università in tempi diversi. Autodidatta nella sua formazione e nei suoi interessi, Assange, concentra i suoi studi su diverse discipline scientifiche come la matematica e le neuroscienze, aggiungendo anche studi in filosofia. Dal 2003 al 2006 frequenta l'Università di Melbourne, ma non riesce a conseguire una laurea.

Appassionato di informatica, da giovanissimo diventa un *hacker* esperto, e nel 1980 fonda un gruppo chiamato "*International Subversives*" con il quale, sembra riuscire a penetrare il sistema informatico dello United States Department of Defense.

Nel 1991 subisce la prima perquisizione, con relativo arresto, da parte della Polizia Federale Australiana. L'ipotesi accusatoria si basa su di una segnalazione che lo identifica come il responsabile di numerosi accessi non consentiti a computer appartenenti ad una università australiana e ad una azienda di telecomunicazioni canadese (Nortel). Viene arrestato e nel 1992 si dichiara colpevole di ben 24 accuse di pirateria informatica, ma viene rilasciato su cauzione (2.100 dollari australiani) per aver dimostrato disponibilità a collaborare con le forze dell'ordine.

In linea con la sua precocità, nel 1989, Assange si sposa con la sua convivente a soli diciotto anni e diventa padre, salvo poi separarsi dalla moglie nel 1991 in seguito all'arresto, seguito da una lunga battaglia legale.

Nel 1995 Assange programma *Strobe*, software *open-source* dedicato al *port scanning*. Nel 1997 collabora alla stesura del libro "*Underground: Tales of Hacking, Madness and Obsession on the Electronic Frontier*". All'anagrafe è un cittadino australiano, ma si sposta continuamente.

Nel 2006 fonda Wikileaks e ne diventa il padre-padrone affermando "*Io sono il cuore e l'anima di questa organizzazione, il suo fondatore, filosofo, portavoce, sviluppatore, organizzatore, finanziatore, e tutto il resto*". Il 29enne australiano, forte di una infrastruttura informatica di tutto rispetto, ben presto comincia a manifestare le sue intenzioni: trasformare il suo portale in un'organizzazione online, finanziata da contributi anonimi e volontari, in grado di fornire informazioni riservate o segrete con l'obiettivo "*...di portare importanti novità e informazioni al pubblico. Una delle nostre attività più importanti è quella di pubblicare notizie e materiale originale che si affianchi alle nostre storie in modo che i lettori e gli storici che ci visitano possano verificare le prove della verità*" [2].

La vita di Assange subisce una trasformazione sostanziale: si sposta continuamente come un fuggiasco da un aeroporto all'altro, pochissime persone hanno un contatto diretto con lui, si registra sotto falso nome, evita di utilizzare carte di credito e *dulcis in fundo* inizia a tingersi i capelli.



Trascorre lunghi periodi in Australia, Kenya, Tanzania, Islanda (luogo in cui acquista una casa). Si esibisce in convegni, seminari e teleconferenze come al New Media Days 2009 a Copenaghen e nel 2010 al Simposio Logan Investigative Reporting presso la UC Berkeley Graduate School of Journalism. Nel 2010, suscita un particolare stupore la sua presenza sull'emittente Al Jazeera, in cui rilascia una dichiarazione clamorosa: annuncia la pubblicazione di una mole impressionante di documenti riconducibili alla guerra in Iraq e in Afghanistan, alcuni dei quali classificati dai servizi di sicurezza di alcuni paesi occidentali. Il 21 giugno 2010 il poliedrico australiano partecipa ad una audizione a Bruxelles, apparendo in pubblico per la prima volta dopo quasi un mese. Viene peraltro inserito, come componente, all'interno di un comitato che affronta il problema della censura in rete. In alcune dichiarazioni rilasciate al quotidiano britannico The Guardian, riferisce di aver subito censure e minacce e che, pur non temendo per la sua incolumità, non ha intenzione di recarsi negli Stati Uniti nemmeno in futuro, su consiglio dei propri legali.

Viene definito in svariati modi da strutture e organizzazioni che lo adorano e lo denigrano, ad esempio *"hacker infame"*, *"personaggio intrigante"* fino a *"combattente per la libertà di Internet"*. A maggio del 2010, i suoi guai giudiziari hanno una nuova ripresa: sbarca in Australia e il suo passaporto gli viene ritirato. Il 20 agosto dello stesso anno, in Svezia un pubblico ministero emette un mandato di cattura nei suoi confronti per sospetta violenza sessuale esercitata su due donne svedesi. Il giorno seguente l'accusa di stupro viene ritirata insieme con il mandato d'arresto. In base ad alcune fonti di informazioni, molto vicine alle presunte vittime coinvolte, sembra che almeno inizialmente, nessuna delle due donne fosse intenzionata a denunciare Assange. Sembra che una delle due fosse intenzionata solo a dichiarare il presunto stupro alla polizia, senza però che ciò determinasse l'attivazione di un'azione giudiziaria contro il patron di Wikileaks. Inoltre la seconda donna sembra abbia avuto solo il compito di accompagnare la sua amica alla polizia, ma poi sarebbe rimasta anch'essa coinvolta nell'episodio, in funzione delle pressioni di alcuni agenti che avrebbero insistito nell'inserimento della sua presenza nel fatto criminoso.

Un aspetto singolare è che nessun colloquio si sia svolto tra la polizia e il pubblico ministero con lo stesso Assange, per verificare la sua versione dei fatti. Nonostante la vicenda sia stata pubblicata da un quotidiano svedese, sono rimasti oscuri i motivi che hanno condotto il pubblico ministero ad emettere il mandato di cattura per poi ritirarlo frettolosamente.

Il pubblico ministero si è limitato a dichiarare al quotidiano Guardian, che il mandato non è stato un errore, ma nuove informazioni ricevute hanno determinato il suo ritiro. Assange, che ha negato categoricamente ogni addebito circa le accuse a lui attribuite, sostiene che in molti lo avevano avvertito che *«qualcuno avrebbe usato sporchi trucchi per delegittimarlo»*. Non sono molti i paesi disposti ad ospitare il fondatore di Wikileaks. Nella sua nazione di origine non sentono la sua mancanza e nel Regno Unito,



dove attualmente risiede, la legge gli consente, grazie alla sua cittadinanza australiana, di rimanere per altri mesi. In Islanda la situazione non è per nulla diversa, anche perché, secondo il parere di Assange, il paese nordico sarebbe troppo “vicino” agli Stati Uniti, che a suo dire lo vorrebbero morto.

La sua madre patria ha apertamente preso le distanze da lui, lasciando anche intendere di essere disponibile a contrastare le azioni di Wikileaks, soprattutto dopo la pubblicazione dei documenti inerenti il conflitto in corso in Afghanistan. Negli USA si starebbe valutando, in funzione della norma sullo spionaggio risalente al 1917, di mettere sotto accusa Assange.

Sembra, peraltro, che in seguito alla diffusione di alcuni documenti, in cui sarebbero presenti molti nomi di collaborazionisti del governo afgano, un esponente dei talebani (identificato con lo pseudonimo di Zabiullah Mujahid) avrebbe dichiarato di aver attivato una sorta di commissione di indagine per confrontare questi nominativi con quelli inseriti in una loro lista di ricercati. Il portavoce talebano avrebbe affermato che, in funzione delle risultanze della commissione, saranno presi *provvedimenti* sulle persone presenti in entrambe le liste.

La scelta di pubblicare questi documenti senza eliminare i nomi delle fonti afgane è pesantemente condannata non solo dal governo statunitense, ma anche da quelle strutture che storicamente hanno sempre sostenuto e difeso il lavoro svolto da Wikileaks. Prime tra tutte, Amnesty International e Report Senza Frontiere, oltre a numerosi esponenti dei governi di diversi paesi. Questa presa di distanza da strutture un tempo “amiche” di Julian Assange, potrebbe riconfigurare uno scenario del tutto nuovo: il passaggio di Wikileaks da icona della democrazia e della libertà di informazione a mero strumento di condizionamento e perseguimento di finalità perseguite da oscuri personaggi e strutture occulte.

Dalle ultime notizie di fine 2011, si apprende che a febbraio prossimo i giudici della Corte Suprema britannica esamineranno la richiesta di estradizione della Svezia, paese in cui permane l'accusa di molestie e violenza sessuale. Nel contempo, il militare americano Bradley Manning, il soldato statunitense accusato di aver passato documenti segreti a WikiLeaks, che sarà giudicato da un tribunale statunitense, rischia 50 anni di carcere per trafugamento di documenti riservati appartenenti al Dipartimento di Stato.

3. Strutture costose, finanziamenti misteriosi

Attualmente la struttura operativa di Wikileaks, può contare su un organico di cinque impiegati e un numero imprecisato di collaboratori (sarebbero circa quaranta), più alcune centinaia di *attivisti* (sostenitori disposti a collaborare con la struttura per la ricerca e fornitura di informazioni).

Per quanto concerne l'aspetto finanziario, la formula dell'autofinanziamento, sposata nella fase iniziale della creazione del sito, produce risultati così



deludenti da non consentire neanche la copertura delle spese annuali di circa 200.000 dollari. Formula che determina, nel giro di poco tempo, una situazione di debito così catastrofica da ridurre Wikileaks quasi sull'orlo del fallimento. Nel giro di qualche mese, però, la situazione economica subisce un'inversione di rotta grazie ad una massiccia campagna di richiesta di finanziamenti, che evidenzia la maggiore delle potenzialità del sito: la possibilità di poter creare situazioni di imbarazzo e difficoltà a governi e paesi di tutto il mondo, grazie alla diffusione di documenti e informazioni riservate e scottanti.

Quindi, inizialmente i contributi vengono assicurati per una metà da piccole donazioni online e per l'altra metà da finanziatori occulti disposti ad elargire somme che raggiungono livelli che vanno ben oltre le piccole somme raccolte.

Nel 2009 Wikileaks sospende momentaneamente le attività lasciando online solo un appello per riscuotere aiuti attraverso la Wau Holland³ (una fondazione tedesca che si occupa interamente della gestione delle donazioni). La Wau Holland attiva un sistema di finanziamento del tutto particolare. I fondi questa volta sono raccolti dalla fondazione tedesca (che conserva della vecchia gestione l'alone di assoluta riservatezza) e Wikileaks, per ottenere il denaro, deve limitarsi a presentare delle ricevute di spese alla Wau Holland. Dato che la legge tedesca impone alle fondazioni la massima trasparenza sulle spese sostenute, Wikileaks si affida ad altre organizzazioni per produrre una serie di variegate ricevute che vengono tutte presentate all'incasso alla Wau Holland. Sembra che fino ad ora, la Wau Holland abbia distribuito 50.000 euro a un conto di Wikileaks in Germania, soltanto sulla base della presentazione di ricevute poco convincenti. Questo sistema di finanziamento viene descritto da Daniel Schmitt, portavoce di Wikileaks, e Hendrik Fulda, consigliere della fondazione. Generalmente l'importo delle donazioni si aggira sui 20 euro, ma la donazione più corposa sembra sia quella di 10.000 euro - come asserisce lo stesso Fulda - elargita da un finanziatore tedesco dopo la pubblicazione dei documenti di guerra sull'Afghanistan. Nella complessa e inestricabile maglia dei rapporti internazionali di Assange, non compare solo la Germania. In un'intervista rilasciata dallo stesso fondatore di Wikileaks al Wall Street Journal, egli asserisce *"Siamo registrati come una biblioteca in Australia, come una fondazione in Francia, come un*

³ Wau Holland Foundation è una fondazione creata in memoria del filosofo e visionario Wau Holland che ha sviluppato raffinati concetti come l'etica degli hacker e la passione per l'utilizzo delle tecnologie informatiche. La fondazione (come è anche specificato sul sito www.wauland.de), è collegata al Chaos Computer Club (CCC) che è la più grande e influente organizzazione di hacker a livello mondiale. Ha sede in Germania e sembra che i suoi iscritti ammontino a circa 4.000. Nel 1989, il CCC è stato indirettamente coinvolto nel primo caso di cyberspionaggio che avrebbe fatto parlare i giornali di tutto il mondo. Un gruppo di hacker tedeschi, guidati da [Karl Koch](#) (che era stato affiliato al CCC), è stato arrestato a seguito di intrusioni nelle reti del Governo Americano e di molte società di computer e per la vendita del codice sorgente di un sistema operativo al sovietico KGB (fonte Wikipedia).



quotidiano in Svezia”, affermando inoltre che avrebbe anche il controllo di due organizzazioni *no profit* presenti negli Stati Uniti, le quali lavorerebbero come “testa di ponte” per il suo portale web (organizzazioni delle quali però non viene rivelata alcuna informazione).

Sembra che nell’aprile del 2010, in seguito alla pubblicazione del video della strage dei civili in Iraq avvenuta nel 2007, le elargizioni al portale di Assange siano aumentate vertiginosamente. Ad ottobre 2010, Wikileaks pubblica qualcosa come 400.000 documenti ritenuti segreti dalla Casa Bianca e dal Pentagono che riguardano stragi di innocenti e presunte verità nascoste sulla guerra in Iraq. Le donazioni decollano al punto tale da portare lo stesso Assange a dichiarare che dall’inizio del 2010, l’importo delle donazioni ammonterebbero a circa un milione di dollari. Tuttavia i nomi e le strutture che finanziano il più noto e discusso sito web, rimangono rigorosamente avvolti nel mistero, scatenando numerose critiche anche da parte degli storici sostenitori di Wikileaks.

Lo stridente contrasto tra il sostegno della verità assoluta e la trasparenza di informazione sventolata dal sito, e la coltre di nebbia che avvolge i nomi dei finanziatori, diventa così imbarazzante da portare i vertici di Wikileaks a sostenere che la diffusione di queste informazioni getterebbe in una situazione di grande pericolo tutti coloro che sostengono economicamente l’associazione.

Senza alcun dubbio, le pubblicazioni dei video e dei documenti imbarazzanti per le strutture governative statunitensi, basterebbero a garantire l’immediata chiusura del sito e, quanto al fondatore di Wikileaks, una probabile lunga permanenza nelle patrie galere degli Stati Uniti. Tuttavia esiste un problema sostanziale: il portale web di Wikileaks non risiede sul territorio USA, quindi non rientra nella giurisdizione americana. E come se non bastasse l’organizzazione *Pirat Partiet* (una sorta di partito, nato in Svezia che mira a raggruppare tutti i pirati informatici del mondo) ha offerto al fondatore di Wikileaks, la disponibilità di tutti i servizi informatici di cui dispone. Sembra tuttavia che Wikileaks abbia declinato l’offerta, grazie al fatto che la struttura operativa può contare su una infrastruttura informatica di tutto rispetto.

Nonostante la coltre di segretezza che avvolge l’organizzazione di Assange, è noto che i numerosi server (computer connessi in rete) sui quali risiedono gli enormi database informativi, sono ubicati presso un ex bunker atomico, situato proprio in Svezia e risalente al periodo della Guerra Fredda.

I server sono ospitati a Stoccolma presso la Bahnhof, uno dei più grandi provider di servizi Internet fondata nel 1994 da Oscar Swart⁴ a Uppsala, nonché il primo Internet Service Provider indipendente della Svezia. La Bahnhof ha sedi a Stoccolma, Göteborg, Uppsala, Borlänge e Lund. A circa 98 metri di profondità, in una collina rocciosa al centro di Stoccolma, ha sede il Pionen White Mountains, un ex bunker nucleare, straordinario cimelio della Guerra Fredda, sono attualmente custoditi i sistemi di Wikileaks. Gli enormi ambienti climatizzati in cui sono alloggiati i sistemi, sono protetti in strutture ad accesso controllato e sono muniti di porte di accesso spesse 40 cm in grado di resistere all'esplosione di un ordigno all'idrogeno (Figure 1 e 2).



Figura 1
Il bunker di Pionen White Mountains

⁴ Oscar Swart è anch'egli un personaggio interessante. Imprenditore, scrittore e blogger nasce a Solna (Stoccolma) nel 1959 ed è il pronipote dell'ex primo ministro svedese Carl Swartz, nonché cugino di secondo grado di Eva Swartz, amministratore delegato della rivista *Natura e Cultura* e del giornalista Richard Swartz. Nel 1994 fonda Bahnhof, ma nel 2004 lascia l'azienda per dirigere un blog (*Texplorer*) in cui tratta argomenti di Information Technology e libertà di stampa in rete. Anch'egli è membro del Pirate Partiet. Come Assange, anche Swartz ha problemi con la giustizia. A marzo del 2005 la polizia svedese sequestra quattro server ubicati presso i locali della Bahnhof nel tentativo di rintracciare materiale protetto da copyright. La società dichiara che i sistemi che detengono questo materiale sono di proprietà di alcuni componenti del personale interno e che il materiale in questione è memorizzato nei sistemi da qualcuno che collabora con Antipiratbyrån, un'organizzazione svedese che lotta contro la violazione del copyright.



Figura 2
Ingresso del bunker di Pionen White Mountains

Per garantire l'energia elettrica 24 ore su 24 vengono impiegati due motori di un ex sommergibile tedesco, capaci di erogare una potenza di 1,5 megawatt. Inoltre, per garantire una temperatura costante negli ambienti dove sono collocati centinaia di *server* che costituiscono l'anima del Pionen, viene anche riutilizzato il sistema di raffreddamento originale di un altro modello di sottomarino. Per quanto concerne la connettività ad Internet, il centro può contare su di una triplice connessione a strutture ridondanti, fisicamente separate l'una dall'altra, peculiarità che consente al Pionen di essere considerato come uno dei centri più protetti e collegati in rete a livello europeo. Un'altra delle peculiarità del sito svedese, risiede nel livello di benessere ambientale che si vive all'interno del centro. Al di là dell'altissimo livello di tecnologie implementate, una particolare attenzione è riservata all'aspetto umano: all'interno di tutti i locali, sono presenti delle serre che conferiscono l'atmosfera di un ambiente all'aperto; il sistema di climatizzazione garantisce una temperatura gradevole e l'impianto di illuminazione è studiato per simulare livelli di luminosità diversi in funzione delle diverse ore del giorno (Figura 3). Dulcis in fundo, viene riservata anche una particolare attenzione all'aspetto scenico della struttura, attraverso alcune piccole cascate d'acqua e un acquario con acqua salata in grado di contenere fino a 2600 litri.

Anche l'aspetto energetico viene tenuto in debita considerazione: per raffreddare i sistemi informatici viene utilizzata, soprattutto durante l'inverno, l'aria fredda proveniente dall'esterno.

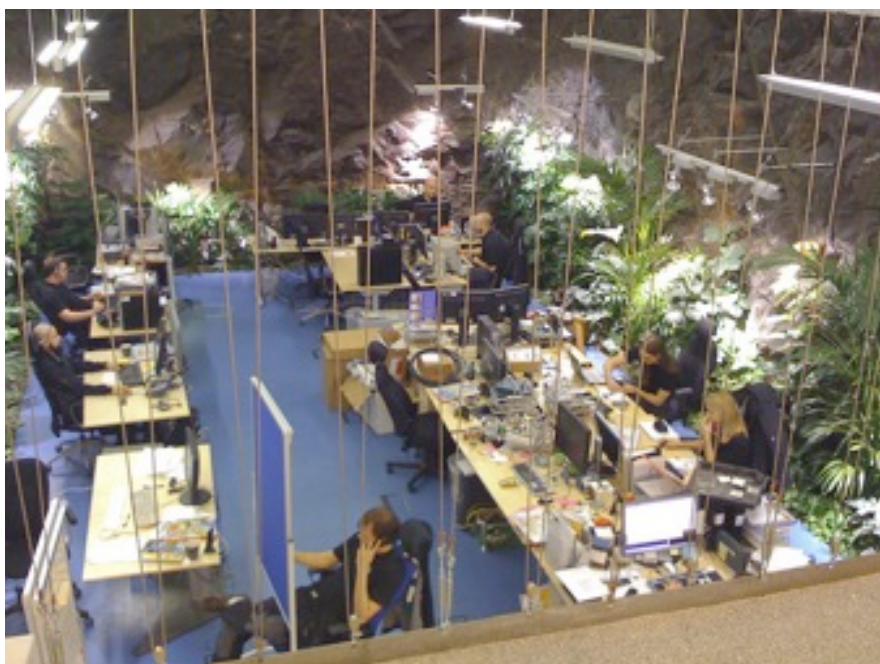


Figura 3
Bunker di Pionen White Mountains

In funzione di quanto descritto, risulta evidente che l'intera infrastruttura di Wikileaks, per il suo funzionamento, fagociti enormi quantità di denaro, sia per quanto concerne i costi di gestione ordinaria, sia per quel che riguarda i costi riconducibili alla ricerca delle informazioni a livello mondiale. Quest'ultimo, è un costo che non va certo sottovalutato. E' impensabile che, per assicurarsi una mole di documenti così interessante e imponenti, Assange non sia costretto a sborsare fiumi di denaro. Se solo considerassimo gli importi riconducibili alla copertura delle spese sostenute dai collaboratori sparsi per tutto il globo terrestre, saremmo già in grado di poter effettuare una superficiale stima del livello dei costi sostenuti.

E' altresì inconcepibile che l'intera organizzazione si possa reggere sui finanziamenti minimali anche di migliaia di simpatizzanti. Pertanto non vi sono dubbi sulla presenza di potenti finanziatori in grado di erogare ingenti somme di denaro per finanziare questo straordinario strumento di informazione/disinformazione mondiale. Rimane pertanto da capire quali siano le reali motivazioni che si celano dietro le rivelazioni del portale.

Ad ulteriore sostegno della tesi che intorno all'organizzazione di Assange giri tanto denaro, c'è la notizia in cui si afferma che Wikileaks necessiti di circa 200.000 dollari annui, per le spese di affitto delle sedi, per la gestione e manutenzione dei server e per i costi legati ai viaggi e al personale.

Pare inoltre che tutti i fondi siano custoditi presso un conto corrente bancario gestito dalla finanziaria Moneybookers, gestita a sua volta dal gruppo finanziario Investcorp, ubicato nel Barhein, ma registrato in Gran Bretagna.



Proprio su questo punto il governo statunitense non ci vede chiaro e, a luglio 2010, decide di inserire il conto corrente del sito nella *watchlist* dei depositi sospettati di finanziare il terrorismo internazionale. Ad agosto, Assange riceve una comunicazione dalla Moneybookers in cui apprende che la finanziaria ha deciso di sospendere il suo conto per compiere una serie di accertamenti. Nella comunicazione si legge che sono stati registrati alcuni accessi da parte di indirizzi IP inseriti in due diverse liste nere, statunitensi e australiane.

Assange non si scompone più di tanto e per contro, il 16 ottobre annuncia che Wikileaks sta per pubblicare qualcosa come 400.000 documenti militari riservati sulla guerra in Iraq ottenuti direttamente da un database collocato in Iraq.

Dopo la precedente pubblicazione di luglio di 77.000 *war logs* (diari di guerra) sul conflitto in Afghanistan, il 22 ottobre Assange fa esplodere l'ennesima bomba mediatica: pubblica i documenti in cui vengono riportati i dati sulle vittime del conflitto iracheno. Le cifre sono dettagliate: 109.032 mila uccisi, con 15 mila casi non riportati (66.081 civili, 23.984 ribelli, 15.196 soldati iracheni, 3.771 militari alleati). Ma le informazioni non si limitano all'esposizione dei dati da necrologio, vengono pubblicate altre informazioni, ma stavolta colpiscono il governo britannico. E' sempre il Guardian a fornirle: sembra che i soldati di Sua Maestà conducessero metodi di interrogatorio che avrebbero violato la convenzione di Ginevra, e tutto ciò sarebbe documentato dalla presenza di manuali segreti in cui si spiega come gestire gli interrogatori ai prigionieri.

A questo punto l'ONU (Organizzazione delle Nazioni Unite) decide di assumere una posizione sulla questione e chiede che vengano effettuate delle indagini sulle attività militari condotte in Iraq.

Alcuni giorni dopo è la Russia ad intervenire. Nel sito *Lifenews.ru*, alcuni esperti di sicurezza russi asseriscono che un team di esperti informatici potrebbe intervenire per bloccare definitivamente le attività di Wikileaks. Questa decisione potrebbe essere motivata dal rischio che il sito possa divulgare documenti riservati che possano costituire una minaccia per lo stato russo. Sembra peraltro che questa dichiarazione sia la diretta conseguenza dell'annuncio del portale di Assange di pubblicare materiale riservato riguardante la Russia e la Cina. In tal senso, sembra che la strategia perseguita da Assange sia assimilabile alla famosa *legge del taglione* in uso nelle popolazioni antiche, che consiste nella possibilità riconosciuta a una persona che abbia ricevuto un'offesa, di infliggere all'offensore una pena analoga proporzionata all'offesa ricevuta. In altri termini occhio per occhio, dente per dente!

4. Chi si cela dietro Wikileaks?

A questo punto sarebbe logico porsi alcune domande: perché un uomo come Julian Assange decide di fare tutto ciò esponendo la sua stessa vita a rischi e pericoli di ogni genere? Come mai il sito Wikileaks che ha



provocato danni incalcolabili a governi e strutture militari non è stato ancora messo “off-line”? E infine, chi finanzia l'imponente infrastruttura IT di Wikileaks? A chi giova e chi danneggia l'azione condotta dal portale?

Dare una risposta a tutte queste domande significherebbe conoscere tutti i segreti che si celano dietro l'organizzazione e il suo fondatore. Tuttavia è possibile fare alcune considerazioni attraverso l'analisi di una serie di informazioni.

Come sappiamo, l'obiettivo apparentemente perseguito dal portale del giovane australiano, è sostanzialmente quello di difendere la verità assoluta mediante la diffusione di informazioni provenienti da tutti coloro che sono intenzionati al conseguimento di questa nobile missione. La formula del *giornalismo partecipativo* non è nuova, ma l'elemento convincente utilizzato da Assange è quello di garantire la riservatezza dei collaboratori in tutte le fasi di trasmissione dei documenti che vengono successivamente pubblicati. Questa certezza crea le basi per lo sviluppo di uno strano *rapporto fiduciario* (azione degna di una vera e propria *psyops* o *operazione psicologica*), che persuade una parte del pubblico a collaborare con Wikileaks, e un'altra a credere ciecamente alle finalità benefiche del progetto. Va inoltre sottolineato che questa garanzia di anonimato assicurata ai collaboratori, lascerebbe intendere che in tutte le fasi di trasmissione dati che avvengono in rete, Wikileaks non effettui alcun controllo sull'indirizzo di rete (IP number) del mittente. Ipotesi poco credibile, soprattutto per una struttura che si sente particolarmente minacciata a livello mondiale.

Bisogna anche aggiungere che non tutti i collaboratori di Wikileaks sono disposti a collaborare gratuitamente, soprattutto quando parliamo di informazioni che possono sconvolgere i rapporti politici, istituzionali, economici e militari di diversi Paesi. Quindi è facile dedurre che tra i simpatizzanti del portale web non manchino di certo coloro che possono intravedere facili e corposi guadagni dalla vendita di documenti che possono sembrare interessanti per il perseguimento degli obiettivi di Assange.

E ormai chiaro che per il sostentamento delle sue attività, Wikileaks necessita di consistenti quantità di denaro, e dato che non è noto da chi e a quanto ammontino le contribuzioni annuali che confluiscono sui suoi conti correnti, appare quantomeno contraddittoria la sua immagine di simbolo della trasparenza assoluta.

Altro aspetto contraddittorio è dato dall'eccessiva risonanza mediatica che viene attribuita ogni volta alle rivelazioni di Wikileaks. Innanzitutto, a ben guardare le informazioni pubblicate non sono poi così eclatanti come le si vorrebbero far apparire: sono molti gli esperti informatici che sostengono che circa il 40% dei documenti presenti su Wikileaks è facilmente rintracciabile con i motori di ricerca, mediante apposite analisi strutturate.



Un altro esempio è dato proprio da una delle ultime pubblicazioni che ha destato particolare scalpore su tutti i media: i dati resi pubblici sul sito riconducibili al numero delle vittime del conflitto iracheno. I dati pubblicati sul portale di Assange riferiscono che il numero delle vittime ammonta a 109.000, ma non è poi così distante da quello fornito, mesi prima, dal portavoce del Ministero della Sanità iracheno, che è pari a 104.000.

Un'altra notizia pubblicizzata di recente si riferisce all'utilizzo, in Afghanistan, di missili portatili a ricerca di calore utilizzati contro gli aerei della coalizione. Anche in questo caso, nulla di nuovo sotto il sole. I missili sembrerebbero essere assimilabili agli Stinger che gli americani hanno fornito ai mjaheddin di Osama Bin Laden per combattere contro i sovietici negli anni Ottanta. E' ben noto a tutto il pubblico mondiale che l'Afghanistan sia la nazione in cui si concentra il più grande traffico di armi a livello internazionale. Alquanto strano appare invece il fatto che ben poca attenzione venga riservata alla questione del traffico di droga. Che l'Afghanistan sia il paese dell'Asia centrale in cui ha sede il centro nevralgico del commercio internazionale di stupefacenti, è noto a tutti, ma Wikileaks non ha mai fatto grandi annunci o pubblicato documenti interessanti su queste attività criminose. Perché?

Anche l'Italia è stata oggetto di attenzione da parte del portale di Assange. Tra i documenti pubblicati negli ultimi mesi, c'è anche un rapporto che riguarda la morte del funzionario del SISMI, Nicola Calipari, avvenuta il 4 marzo 2005, per salvare la vita alla giornalista Giuliana Sgrena. Nelle pagine del documento viene riportata la testimonianza di un certo Sheikh Husayn (che risulta essere l'ex capo di una cellula terroristica di Al Qaeda a Bagdad, successivamente arrestato dai servizi segreti giordani), in cui asserisce di aver egli stesso consegnato la giornalista Giuliana Sgrena in cambio di un riscatto di 500.000 dollari.

Oltre all'annuncio della dichiarazione, ampiamente divulgato soprattutto sui media nazionali, non vi è alcuna prova o indizio certo che possa supportare la testimonianza di Husayn. Non viene indicato il nome di chi avrebbe attivato le trattative e nemmeno per conto di chi; non viene precisato chi sarebbe stato incaricato della consegna del denaro e in che modo si sarebbe svolta l'operazione di consegna; infine non viene detto dove sarebbe avvenuto lo scambio.

Inoltre, dopo la consegna, il terrorista asserisce di aver telefonato al ministro degli interni iracheno, sostenendo che l'auto che trasportava gli italiani era una Toyota Corolla blu che era stata imbottita di esplosivo per compiere un determinato attentato. L'inganno avrebbe funzionato al punto tale che la polizia irachena avrebbe comunicato la questione ai militari americani, provocando la ben nota reazione di questi ultimi che avrebbero aperto il fuoco, uccidendo il funzionario italiano, proprio per evitare una strage.



In tal senso, mi sembra doveroso fare alcune considerazioni. Innanzitutto la Sgrena ha sempre sostenuto che i suoi rapitori non appartenevano ad Al Qaeda ma erano uomini della “Resistenza”. In effetti se i rapitori fossero stati membri del gruppo terroristico, ad una mera operazione di scambio di denaro, avrebbero certamente preferito un’azione di massima visibilità mediatica ottenibile mediante l’eliminazione della giornalista (tragico epilogo di sequestri che hanno coinvolto diversi giornalisti stranieri). Seconda riflessione: perché mai questo terrorista Husayn avrebbe comunicato agli americani (tramite il ministro iracheno) la presenza di un’auto bomba diretta verso un posto di blocco? Cosa ci avrebbe guadagnato? Anzi, un gesto del genere avrebbe sicuramente compromesso la sua stessa sicurezza in qualità di leader di una cellula terroristica. E poi, è veramente plausibile sostenere che Calipari, il quale non era di certo uno sprovveduto, abbia interagito per una trattativa così delicata, con un personaggio così bizzarro? Dulcis in fundo, come è possibile pensare che il capo di una cellula terroristica comunichi con la polizia irachena? Altro elemento contraddittorio è rappresentato dal fatto che Wikileaks asserisce che l’auto in cui viaggiavano Calipari, la Sgrena e l’autista, era una Corolla blu mentre in realtà si trattava di una Corolla bianco-grigia.

Insomma le imprecisioni e le incongruenze abbondano e aprono ampi margini di dubbio e di perplessità sull’attendibilità delle rivelazioni effettuate. In questo scenario, dai tratti oscuri e controversi, si delinea ancora una volta l’unica verità assoluta: sulla rete si combatte una vera e propria guerra delle informazioni, o per meglio dire, la guerra della disinformazione. I giornali e le televisioni sono solo un appendice di quel complesso e sconfinato mondo del cyberspazio, che non può essere controllato e censurato in alcun modo. I prossimi conflitti informativi, saranno combattuti sempre di più nel cyberspazio, fino a quando, tra dieci o venti anni, i tradizionali media non saranno completamente surclassati.

Robert Fisk, famoso giornalista del quotidiano *The Independent*, sostiene che il giornalismo investigativo di tipo *old style*, che si basa sul meticoloso e lungo lavoro dei reporter che operano sul campo, sia ormai messo completamente in discussione dai mezzi di comunicazione fruibili in rete.

Seguendo questa ipotesi di scenario, il quesito che bisogna porsi è il seguente: è possibile sostenere che le rivelazioni che si basano su documenti incompleti e imprecisi e di dubbia provenienza, ma opportunamente plasmati ed abilmente trasmessi alle masse, possano costituire l’ossatura base su cui costruire la verità assoluta? Se ciò fosse vero (ed io sono convinto che lo sia) saremmo di fronte ad una situazione di alta pericolosità.

Stiamo parlando dell’utilizzo del cyberspazio come strumento ideale per condurre delle *psychological operations*, che consentirebbero, ad un manipolo di potenti eminenze grigie, di conseguire i loro specifici obiettivi, manipolando le menti e quindi i destini dell’uomo e delle stesse nazioni.

Non a caso, il pubblico non è portato a riflettere sul fatto che molti dei documenti pubblicati da Wikileaks possono essere considerati utili da un punto di vista informativo, ma solo se opportunamente analizzati e riscontrati in funzione di dati certi e sostanziali. Quindi i dati “grezzi” non devono essere considerati come informazioni vere e reali, al contrario possono servire solo a mistificare la realtà dei fatti. Se poi aggiungiamo che non è dato sapere quali siano le fonti dalle quali provengono tali documenti, è facile che le stesse siano soggette a distorsioni, omissioni o particolari autocensure.

Alcuni mesi fa, Julian Assange in un'intervista rilasciata al quotidiano El Pais, si definisce come l'incubo di governi, multinazionali e grandi banche, e persino della Chiesa di Scientology. Interrogato sulla rivoluzione digitale e sul giornalismo indipendente, ha asserito *“Forse arriveremo a un sistema con un maggior controllo giudiziario e accordi internazionali per reprimere la libertà di stampa, o forse andremo verso un nuovo standard in cui la gente si aspetta e pretende un'informazione più aggressiva rispetto al potere; e un contesto commerciale che renda redditizie inchieste di questo genere; e un contesto legale che le protegga”*. E poi ha aggiunto *“Con le nostre azioni di oggi stiamo determinando il destino del contesto mediatico internazionale dei prossimi anni”*. Sembra quindi che dietro il liberalismo informativo e all'esaltazione ossessiva della verità assoluta dell'inventore di Wikileaks, si celino ben altri obiettivi...

Bibliografia

- [1] <http://www.wikileaks.org>
- [2] <http://www.nytimes.com/2010/10/24/world/24assange.html>
- [3] <http://www.guardian.co.uk/media/2010/oct/26/wikileaks-fox-iraq-war-logs>
- [4] <http://www.digitaljournal.com/article/296979>
- [5] <http://gizmodo.com/5626381/this-is-the-nuclear-bunker-where-wikileaks-will-be-located>
- [6] http://www.huffingtonpost.com/christopher-alessi/wikileaks-founder-lashes_b_774413.html
- [7] http://www.washingtonpost.com/wp-dyn/content/article/2010/10/23/AR2010102301316_2.html
- [8] <http://www.veteranstoday.com/2010/10/23/gordon-duff-ho-hum-more-wikileaks-chickenfeed/>
- [9] <http://www.caffenews.it/?p=11929>
- [10] http://www.corriere.it/esteri/10_ottobre_23/iraq-wikileaks-olimpio_89682e28-de70-11df-99d6-00144f02aabc.shtml
- [11] http://www.corriere.it/esteri/10_luglio_26/afghanistan-documenti-segreti_081f12fe-9876-11df-a51e-00144f02aabe.shtml
- [12] <http://www.ilpost.it/2010/08/23/come-si-mantiene-wikileaks/>
- [13] <http://semidiceviprima.com/?p=9123>



Biografia

Antonio Teti è responsabile del supporto tecnico informatico della Direzione Generale dell'Università "G. d'Annunzio" di Chieti-Pescara. Esperto di ICT Governance e ICT Security, ha accumulato una esperienza ventiquennale in studi, ricerche e consulenze in aziende pubbliche, private e istituzioni governative. È stato docente di Informatica presso diverse università italiane tra le quali l'Università Cattolica del Sacro Cuore di Roma, l'Università di Teramo e l'Università della Calabria. E' Presidente Onorario della Società Italiana delle Scienze Informatiche e Tecnologiche (SISIT). E' membro della European Academy of Sciences and Arts, professional member dell'Association for Computing Machinery (ACM) e della New York Academy of Sciences (NYAS). Tra i suoi libri più noti si citano: EUCIP – Il manuale per l'informatico professionista (U. Hoepli); Business and Information System Analyst – Il Manuale per il Manager IT (U. Hoepli); Network Manager – Il manuale per l'Amministratore di Reti e Sistemi (U. Hoepli); Management dei servizi IT: dal modello ITIL all'ISO/IEC 20000 (Il Sole 24 Ore); Sistemi informativi per la Sanità – ECDL Health" (APOGEO); Il futuro dell'Information & Communication Technology: tecnologie timori e scenari della global network revolution (Springer); PsychoTech il punto di non ritorno – La tecnologia che controlla la mente - Springer .

E-mail: teti@unich.it



FORMAZIONE INFORMATICA E POSSIBILI RISPARMI NELLA PUBBLICA AMMINISTRAZIONE LOCALE

Antonio Piva, Michele Baggi

L'articolo presenta la ricerca sperimentale, svolta nel nord est, sui possibili risparmi per la Pubblica Amministrazione locale. Sono stati prima verificati i costi per l'assistenza tecnica informatica negli Enti locali coinvolti nella sperimentazione e quindi calcolati i risparmi dopo aver formato, in base alla certificazione ECDL, il personale. Oltre al costo dovuto alla scarsa conoscenza degli strumenti informatici si è considerato anche il potenziale miglioramento della produttività dei dipendenti nella Pubblica Amministrazione locale e nel Sistema Sanitario, per arrivare ad un risparmio annuale di circa 87,00 Euro per cittadino.

1. Introduzione

Per stimare i possibili risparmi nella Pubblica Amministrazione Locale dovuti dal buon uso dell'informatica da parte degli utenti, si è svolta una ricerca che ha verificato i risparmi derivanti dalla formazione e certificazione ECDL del personale.

Innanzitutto, nell'anno 2010 è stata effettuata una ricerca sperimentale sui costi dell'assistenza tecnica informatica negli Enti Pubblici Locali del Nord Est. E' stata quindi calcolata sperimentalmente l'incidenza della formazione e certificazione ECDL del personale su tali costi, ottenendo, di conseguenza, i possibili risparmi.

La prima parte della ricerca ha verificato la quantità di interventi di assistenza tecnica richiesti, dovuta alla scarsa conoscenza degli strumenti informatici da parte dei dipendenti degli Enti Pubblici Locali, e quindi risparmiabile attraverso l'aumento delle competenze informatiche degli



stessi. Tale ricerca sperimentale è stata condotta in alcuni Comuni del Nord Est, quali il Comune di Udine, il Comune di Caorle e una rete di undici Comuni del Medio Friuli aventi come capofila il Comune di Codroipo. Un gruppo di dipendenti, per ciascuno di questi Comuni, ha poi partecipato a un periodo di formazione sui contenuti della certificazione ECDL core, per poi ottenere la certificazione. Con questa ricerca sperimentale si è anche voluto verificare la validità della formazione e certificazione ricevuta dai dipendenti, misurando la diminuzione del numero di interventi di assistenza richiesto da parte degli stessi. Si è riscontrato che a fronte di un aumento della conoscenza e delle capacità informatiche dei dipendenti è corrisposto un minor numero di richieste di assistenza. Dai dati ottenuti dalla ricerca sperimentale sul numero di interventi di assistenza tecnica negli Enti Pubblici Locali, è stato possibile calcolare il costo annuale per l'assistenza tecnica dovuta alla scarsa conoscenza degli strumenti informatici da parte degli utenti, sostenuto dagli Enti Pubblici Locali in Friuli Venezia Giulia.

Infine, attraverso i dati riscontrati dalla SDA Bocconi, riguardanti l'incremento della produttività dei dipendenti della Pubblica Amministrazione locale e Sanità, è stato possibile calcolare il costo dovuto alla scarsa conoscenza degli strumenti informatici e quindi i possibili risparmi con la formazione e certificazione ECDL e il potenziale miglioramento della produttività dei dipendenti nella Pubblica Amministrazione locale e nel Sistema Sanitario. Questi risparmi sono stati stimati a livello regionale globale, per dipendente e anche calcolati come risparmio/beneficio per singolo cittadino.

2. Gli interventi di assistenza tecnica derivanti dalla scarsa capacità degli utenti nell'uso dell'informatica

Al fine di verificare quanti siano gli interventi tecnici negli Enti Pubblici Locali del Nord Est derivanti dalla scarsa capacità degli utenti nell'uso degli strumenti informatici d'ufficio, e pertanto risparmiabili attraverso l'aumento delle competenze informatiche degli stessi, si è condotta una ricerca sperimentale coinvolgendo dodici comuni del Friuli Venezia Giulia e un comune della provincia di Venezia. Fra i comuni del Friuli Venezia Giulia vi sono il comune di Udine, capoluogo di provincia, che conta quasi 100.000 abitanti, e la rete di undici comuni del Medio Friuli facenti capo al comune di Codroipo. I comuni della rete sono Codroipo, Basiliano, Lestizza, Mortegliano, Castions Di Strada, Talmassons, Bertolo, Varmo, Camino Al Tagliamento, Sedegliano, Mereto Di Tomba, e insieme contano complessivamente circa 52.000 abitanti, dei quali, solo il comune di Codroipo ne conta 15.704. Gli altri comuni del gruppo hanno un numero di abitanti che varia dai 5.472 di Basiliano ai 1.676 di Camino al Tagliamento e quindi risultano di dimensioni piccole o molto piccole.

Da un punto di vista strutturale il Friuli Venezia Giulia si compone di 219 comuni, dei quali il 42,4% di dimensioni molto piccole (meno di 2.000 abitanti), il 47% di piccole dimensioni (da 2.000 a 10.000 abitanti), l'8,6% di



medie dimensioni (da 10.000 a 30.000 abitanti) e solo l'1,8%, corrispondente ai quattro capoluoghi di provincia, contano più di 30.000 abitanti. La regione Friuli Venezia Giulia conta complessivamente 1.231.257 abitanti, di cui i comuni coinvolti nella ricerca ne coprono 163.558, corrispondente al 13,3%.

Il tredicesimo comune coinvolto nella ricerca è il comune di Caorle che fa parte di quell'area geografica posta a nord-est della provincia di Venezia, conosciuta come Veneto Orientale. Il comune conta 12.016 abitanti, corrispondente a circa l'1,8% del totale provinciale; infatti, la provincia di Venezia, facente parte della regione Veneto, conta 44 comuni, prevalentemente di piccole e medie dimensioni, per un totale di 858.915 abitanti¹.

Complessivamente, il campione di comuni considerato nella ricerca è quindi variegato e rappresentativo della situazione nel Nord Est; infatti, esso consiste di un comune capoluogo, due comuni di medie dimensioni e dieci comuni di piccole dimensioni.

Durante un periodo temporale di dodici mesi, da Aprile 2009 a Marzo 2010, sono stati raccolti i dati riguardanti gli interventi di assistenza tecnica informatica nei comuni campione. Assieme ai referenti dei sistemi informativi degli Enti Pubblici Locali e i tecnici operativi che hanno prestato il servizio di assistenza, tali dati sono stati catalogati distinguendo da un lato gli interventi di assistenza dovuti alla poca dimestichezza informatica degli utenti e, quindi, considerati interventi evitabili e dall'altro gli interventi veramente necessari². Durante il processo di catalogazione si è reso evidente che molti interventi non erano inseribili in una sola delle suddette categorie, infatti, è frequente che un intervento comprenda diverse operazioni, alcune squisitamente tecniche e altre di assistenza di base imputabili alle scarse conoscenze informatiche degli utenti. In questi casi l'intervento è stato suddiviso in due parti conformemente al tipo di operazioni richieste (ad esempio, un intervento di sette ore per "operazioni manutenzione server e assistenza generica" è stato suddiviso in un intervento di sei ore riguardante la manutenzione del server, catalogato come necessario e un intervento di un'ora per l'assistenza all'utente, catalogato come evitabile), salvaguardando così la consistenza e l'attendibilità del dato registrato.

Si riportano, a titolo esemplificativo, alcune fattispecie di richieste di assistenza da parte degli utenti, imputabili alle scarse conoscenze informatiche degli stessi: difficoltà nell'impostare la stampa fronte/retro, reindirizzamento di una e-mail, problemi con gli allegati delle e-mail,

¹ La fonte del numero di abitanti dei comuni del Friuli Venezia Giulia e della provincia di Venezia è il rapporto ISTAT 2009.

² Da questa ricerca sono esclusi tutti gli interventi tecnici riguardanti operazioni di manutenzione su infrastrutture informatiche centralizzate quali dispositivi di rete o server, mentre l'attenzione è stata focalizzata sull'assistenza tecnica verso gli utenti finali.

rimozione destinatario da una e-mail, problemi con la formattazione del testo in Word, disabilitazione del controllo di ortografia in Word, difficoltà nella gestione delle tabelle di Excel³. Interventi dovuti a rotture dell'hardware o riguardanti la configurazione della rete, delle stampanti o l'installazione di nuovo hardware o software, sono stati considerati interventi necessari.

I risultati dell'analisi del numero totale di interventi di assistenza tecnica informatica presso i comuni campione, rivela che nel Comune di Udine circa il 38% degli interventi è classificato come evitabile. Questo valore aumenta a circa l'82% nei Comuni del Medio Friuli e circa all'83% nel Comune di Caorle. In tabella 1 vengono riportati i valori numerici e le percentuali delle diverse categorie di interventi tecnici.

	INTERVENTI EVITABILI		INTERVENTI NECESSARI		INTERVENTI TOTALI
UDINE	1800	38,3%	2904	61,7%	4704
MEDIO FRIULI	1670	81,6%	377	18,4%	2047
CAORLE	491	82,9%	101	17,1%	592
TOTALE	3961	54%	3382	46%	7343

Tabella 1

Numero di interventi tecnici eseguiti presso i comuni campione nel periodo Aprile 2009 - Marzo 2010, suddivisi per categoria e totali

Dall'analisi dei dati emersi si riscontra che nel Comune capoluogo di Udine, con circa 1.000 dipendenti, e pertanto fortemente strutturato, il numero di interventi evitabili è di 1800 su un totale di 4704 interventi, che corrisponde a circa il 38% del totale. Questa percentuale risulta essere molto inferiore rispetto ai comuni di media piccola dimensione. Inoltre, in relazione alla dimensione dell'Ente, il Comune di Udine dispone di alcuni dipendenti di riferimento, che fungono da *focal point* nelle diverse aree dell'ente, in quanto possiedono una buona conoscenza e dimestichezza con gli applicativi informatici d'ufficio e pertanto possono filtrare, almeno in parte, le richieste di assistenza informatica da parte del personale dell'area di competenza fornendo il primo supporto ai colleghi meno preparati nell'uso dei sistemi informatici d'ufficio.⁴

³ Questi argomenti sono previsti all'interno del programma (syllabus) della certificazione ECDL.

⁴ Si vedano anche le considerazioni nel paragrafo successivo.

Per quanto riguarda, invece, i comuni di media e piccola dimensione, la percentuale del numero di interventi di assistenza tecnica evitabile è molto superiore. Infatti, nei dodici comuni appartenenti alla rete degli Enti Pubblici Locali del Medio Friuli e Caorle la media è dell'82,2%.

A livello globale, la percentuale di interventi di assistenza evitabili sul totale delle richieste è del 54%, il che significa che più della metà degli interventi richiesti dai dipendenti è dovuta alla poca dimestichezza degli stessi con gli strumenti informatici.

I grafici presenti nelle figure 1 e 2 illustrano la suddivisione percentuale del totale di interventi tecnici richiesti fra evitabili e necessari, rispettivamente nei comuni di grandi dimensioni e nei comuni di medie e piccole dimensioni, e il totale generale.

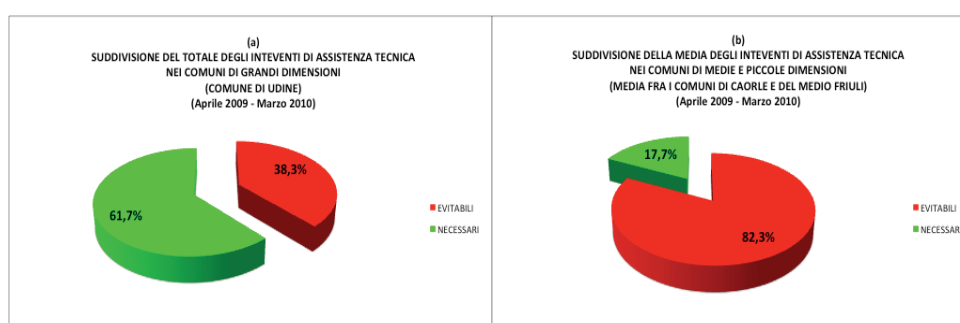


Figura 1

Suddivisione percentuale del totale degli interventi tecnici (a) nei comuni di grandi dimensioni, come nel caso del Comune di Udine, e (b) nei comuni di medie dimensioni, ottenuto dalla media dei valori dei Comuni di Caorle e del Medio Friuli

Considerando il numero di dipendenti di ciascun ente è possibile analizzare il numero medio di interventi, totali ed evitabili, richiesti annualmente per dipendente. Dalla ricerca si evince che in media vengono richiesti circa 5 interventi all'anno, di cui 3 evitabili. Nei comuni del Medio Friuli la media annua di interventi totali richiesti dai dipendenti è la più alta, pari a 8,12 interventi, mentre per i Comuni di Udine e Caorle tale valore scende attorno a circa 5 interventi annui. Considerando solo gli interventi evitabili, la tendenza è confermata; infatti, i comuni del Medio Friuli riportano la media più alta pari a 6,63 interventi evitabili per dipendente, seguito dal Comune di Caorle con 4,09 interventi e infine il Comune di Udine con un valore di molto inferiore, pari a 1,84 interventi evitabili annui per dipendente. I dati sono riportati in tabella 2.

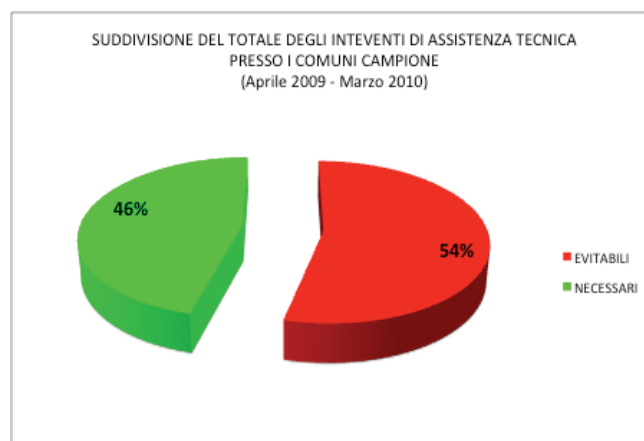


Figura 2

Suddivisione percentuale del totale degli interventi tecnici, eseguiti presso i comuni campione, nelle due categorie “evitabili” e “necessari”

	NUMERO DIPENDENTI ⁵	NUMERO TOTALE INTERVENTI	INTERVENTI ANNUI PER DIPENDENTE	NUMERO TOTALE INTERVENTI EVITABILI	INTERVENTI ANNUI EVITABILI PER DIPENDENTE
UDINE	976	4704	4,82	1800	1,84
MEDIO FRIULI	252	2047	8,12	1670	6,63
CAORLE	120	592	4,93	491	4,09
TOTALE	1348	6803	5,45	3961	2,94

Tabella 2

Calcolo del numero medio annuo di interventi tecnici, totali ed evitabili, richiesti per ciascun dipendente presso i comuni campione

3. Incidenza della formazione ECDL sul risparmio degli interventi di assistenza tecnica

Per valutare l'incidenza della formazione sull'uso degli strumenti d'informatica d'ufficio da parte degli utenti e quindi sul numero di interventi di assistenza tecnica informatica richiesti dai dipendenti degli Enti Pubblici Locali, sono stati organizzati dei corsi di formazione per i dipendenti dei comuni campione sui contenuti previsti dalla certificazione ECDL. Dopo il periodo di formazione, sono stati misurati gli interventi tecnici richiesti dai dipendenti formati per valutare sperimentalmente i benefici della

⁵ Fonte: uffici personale e sistemi informativi dei Comuni di Caorle, Udine, e dei comuni considerati del Medio Friuli.



formazione sul numero di tali interventi dovuti alla scarsa conoscenza degli strumenti informatici da parte dei dipendenti stessi.

Nello specifico, sono stati organizzati tre corsi, della durata di 40 ore ciascuno, i quali prevedevano la formazione sui seguenti moduli della certificazione ECDL Core: l'uso del computer e gestione dei file (modulo 2), elaborazione di testi (modulo 3), foglio elettronico (modulo 4), database (modulo 5), strumenti di presentazione (modulo 6) e reti informatiche (modulo 7).

Per il Comune di Udine il corso è stato organizzato dall'Istituto Tecnico Industriale "A. Malignani" e ha avuto 17 corsisti fra i dipendenti comunali. Per la rete di comuni del Medio Friuli il corso è stato organizzato dall'Istituto scolastico "Linussio/Ceconi" in collaborazione con Qnet s.r.l. e ha avuto 17 corsisti scelti in modo da rappresentare tutti gli undici comuni della rete del Medio Friuli. Per il Comune di Caorle il corso è stato organizzato da Formazione Unindustria Treviso ed ha avuto 14 corsisti.

Al fine di comprendere la percezione dei corsisti rispetto alle proprie conoscenze sui contenuti dei moduli della certificazione ECDL, sono stati utilizzati dei questionari di valutazione. Ai dipendenti è stato chiesto di autovalutare la propria conoscenza degli applicativi per l'elaborazione testi, il foglio elettronico, strumenti di presentazione, database, posta elettronica e internet browser. La scala di giudizi proposti nei questionari prevedeva cinque livelli di conoscenza: nulla, scarsa, appena sufficiente, buona e ottima.

Dai risultati dei questionari di valutazione è emerso che, prima di iniziare il corso, i dipendenti degli Enti Pubblici Locali considerati valutava di possedere una conoscenza appena sufficiente degli applicativi per l'elaborazione dei testi, dell'uso della posta elettronica e degli internet browser. La conoscenza del foglio elettronico e degli strumenti di presentazione risultava scarsa mentre quella dei database era quasi nulla. I risultati globali dei questionari di autovalutazione sottoposti ai dipendenti prima di iniziare il corso di formazione, sono illustrati nel grafico della figura 3.

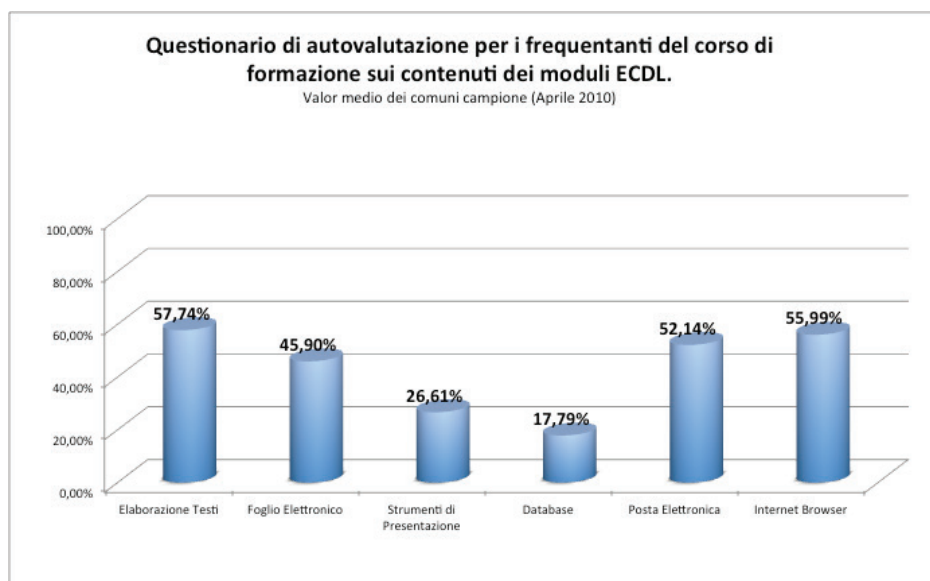


Figura 3

Valori medi dei risultati dell'autovalutazione rappresentanti la conoscenza media dei partecipanti al corso sui contenuti dei moduli previsti dalla certificazione ECDL, prima del corso di formazione. I giudizi dei corsisti sono stati normalizzati su scala percentuale

Calcolando la media globale dei giudizi risultanti dai questionari di autovalutazione, si ottiene che la conoscenza media dei contenuti dei moduli previsti dalla certificazione ECDL, da parte dei corsisti prima del periodo di formazione, era piuttosto scarsa. Il valor medio dei giudizi è pari a 42,69%.

Considerando i valori medi per singolo comune, si ottiene che i dipendenti del comune di Udine hanno valutato di possedere una conoscenza media dei contenuti della certificazione ECDL appena sufficiente, pari a 50,98%. Per i comuni del Medio Friuli e di Caorle i valori si abbassano rispettivamente a 44,36% e 32,74%, corrispondenti a una conoscenza scarsa. I risultati dei questionari di autovalutazione sottoposti ai dipendenti dei comuni considerati sono riportati in tabella 3.

	APPLICATIVI INFORMATICI PER L'UFFICIO						Media
	Elab. Testi	Foglio Elettronico	Strumenti di Presentazione	Database	Posta Elettronica	Internet Browser	
UDINE	63,24%	55,88%	39,71%	30,88%	57,35%	58,82%	50,98%
MEDIO FRIULI	61,76%	51,47%	29,41%	11,76%	54,41%	57,35%	44,36%
CAORLE	48,21%	30,36%	10,71%	10,71%	44,64%	51,79%	32,74%
MEDIA	57,74%	45,90%	26,61%	17,79%	52,14%	55,99%	42,69%

Tabella 3

Risultati percentuali dell'autovalutazione dei partecipanti al corso di formazione rispetto alla conoscenza degli argomenti dei moduli della certificazione ECDL

La differenza fra l'autovalutazione dei dipendenti del Comune di Udine e quella degli altri comuni è dovuta al fatto che, come accennato nel precedente paragrafo, il Comune di Udine dispone di alcuni dipendenti maggiormente competenti nell'uso degli strumenti informatici, i quali fungono da *focal point* per i colleghi meno esperti. Tali referenti sono proprio i dipendenti che hanno partecipato al corso di formazione e che sono stati scelti appositamente dall'ufficio Personale e Sistemi Informativi del Comune con l'obiettivo di accrescere ulteriormente le loro conoscenze e capacità, migliorando così la citata strategia *focal point*. Nei comuni del Medio Friuli c'è stata solo una parziale selezione del personale, inserendo alcune persone che avevano già una buona conoscenza degli strumenti informatici, mentre gli altri avevano un livello di conoscenza più basso. Nel Comune di Caorle hanno partecipato al corso le persone con non molta dimestichezza nell'uso del computer e degli applicativi informatici d'ufficio.

Al termine del corso di formazione sostanzialmente tutti i partecipanti hanno iniziato a sostenere gli esami dei diversi moduli per ottenere la certificazione ECDL e alcuni si sono certificati poco dopo.

Per valutare l'incidenza della formazione sul numero di interventi di assistenza tecnica di natura informatica, sono stati considerati e analizzati gli interventi di assistenza richiesti, limitatamente dai dipendenti che hanno ricevuto la formazione, prima e dopo il periodo del corso. Il periodo precedente alla formazione fa riferimento a un intervallo di tempo della durata di un anno, da Aprile 2009 a Marzo 2010, mentre il periodo successivo corrisponde a un intervallo di otto mesi, da Maggio 2010 a Dicembre 2010. Dal confronto dei dati ottenuti dall'analisi realizzata prima e dopo la formazione, è stato possibile riscontrare la validità della certificazione ECDL quale metodo di formazione di conoscenze e competenze adeguate alle necessità operative dei dipendenti comunali e quindi a una loro maggiore produttività e un conseguente risparmio per gli Enti Pubblici Locali.

Limitatamente ai dipendenti che hanno ricevuto la formazione, il numero totale di interventi di assistenza richiesti, nei comuni campione, prima del periodo di formazione, sono così suddivisi: 71 evitabili e 121 necessari, per un totale di 192 richieste di assistenza. Gli interventi evitabili risultavano quindi il 37% del totale, che è inferiore al dato globale riportato in tabella 1, il quale indicava che nei comuni campione ben il 54% degli interventi era dovuto alla scarsa conoscenza degli strumenti informatici da parte dei dipendenti⁶.

I dati che si riferiscono agli interventi registrati prima della formazione dei dipendenti sono riportati in tabella 4.

⁶ Il 54% indicato in tabella 1 si riferisce a tutti i dipendenti e non solo a quelli selezionati per la formazione ECDL.

NUMERO INTERVENTI (APRILE 2009 -MARZO 2010)		
EVITABILI	NECESSARI	TOTALI
71	121	192
37%	63%	

Tabella 4

Suddivisione del totale di interventi tecnici richiesti dai dipendenti che hanno ricevuto la formazione ECDL nei comuni campione, prima della formazione

Nei sei mesi successivi alla formazione, da Maggio 2010 a Ottobre 2010, sono stati monitorati gli interventi richiesti dallo stesso gruppo di dipendenti nei comuni campione. Ne risulta che sono stati richiesti globalmente 9 interventi dovuti alla poca dimestichezza con gli strumenti informatici e 60 necessari, per un totale di 69 richieste di assistenza.

Gli interventi evitabili risultano quindi solo il 13% sul totale. I dati riguardanti gli interventi registrati dopo la formazione sono riportati in tabella 5.

NUMERO INTERVENTI (MAGGIO 2010 - OTTOBRE 2010)		
EVITABILI	NECESSARI	TOTALI
9	60	69
13%	87%	

Tabella 5

Suddivisione del totale di interventi tecnici richiesti dai dipendenti che hanno ricevuto la formazione ECDL nei comuni campione, nei sei mesi successivi alla formazione

Sebbene la percentuale globale degli interventi evitabili sul totale degli interventi tecnici, e quindi dei dipendenti, fosse stata riscontrata nel 54% dei casi, la percentuale degli interventi tecnici evitabili relativi ai dipendenti oggetto della formazione risultava essere del 37% prima del corso. Dopo la formazione si è passati dal 37% al 13%.

Il grafico presente nella figura 4 illustra il cambiamento del rapporto percentuale degli interventi fra evitabili e necessari, prima e dopo la formazione dei dipendenti.

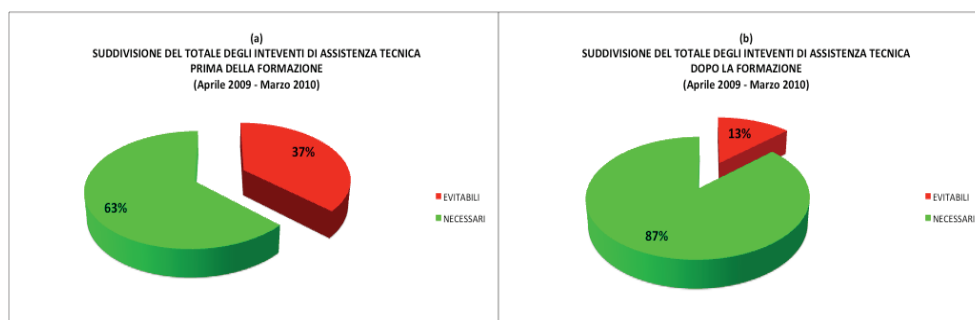


Figura 4

Suddivisione percentuale del totale degli interventi tecnici, richiesti dai dipendenti che hanno ricevuto la formazione ECDL, (a) prima della formazione e (b) dopo la formazione

Come detto in precedenza, i dipendenti del Comune di Udine avevano già in partenza una buona formazione informatica, di conseguenza i miglioramenti maggiori si sono apprezzati laddove il livello di conoscenza iniziale era più basso. I dati emersi indicano che nel comune capoluogo, si passa da un 35% di interventi evitabili a un 15%. Nei comuni di medie e piccole dimensioni si osserva che la percentuale globale degli interventi evitabili sul totale degli interventi tecnici, e quindi dei dipendenti, era stata riscontrata, prima della formazione, nel 82,3%, mentre la percentuale degli interventi tecnici evitabili relativi ai dipendenti oggetto della formazione risultava essere del 46,6%. Per questi ultimi, dopo la formazione, si è passati da un circa 47% di interventi evitabili a un 6,3% (si veda la Figura 5).

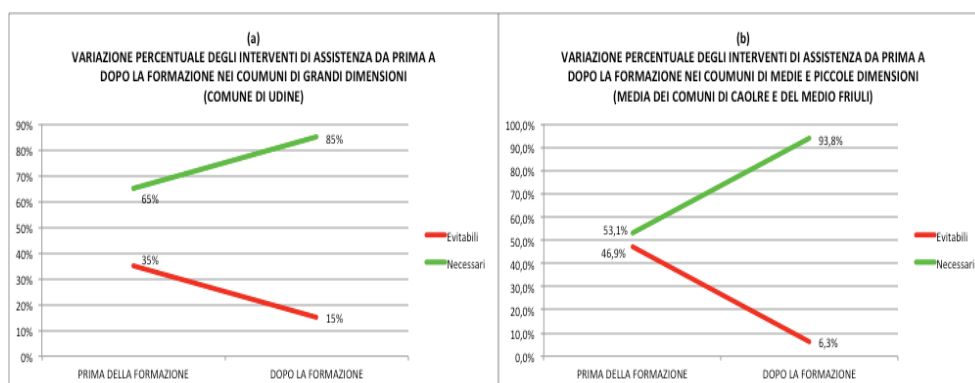


Figura 5

Riduzione percentuale, per categoria, degli interventi tecnici (a) nei comuni di grandi dimensioni, come nel caso del comune di Udine, e (b) nei comuni di medie dimensioni, ottenuto dalla media dei valori dei comuni di Caorle e del Medio Friuli

Considerano solamente il numero di interventi tecnici evitabili, l'incidenza della formazione ECDL appare interessante. La riduzione del numero degli interventi evitabili nei singoli comuni è variegata, si passa da una riduzione del 57% degli interventi evitabili sui totali a Udine, fino all'87% nei comuni del Medio Friuli.

Ritornando ai valori globali degli interventi tecnici eseguiti presso i comuni campione, si ha che il 37% degli interventi prima del periodo di formazione era evitabile, mentre successivamente alla formazione, questo valore si è ridotto al 13%. Si riscontra quindi che, grazie alla formazione ECDL dei dipendenti, il numero di richieste di assistenza dovute alla scarsa conoscenza degli strumenti informatici si è ridotto globalmente del 64,7%. Il grafico della figura 6 mostra la percentuale di interventi evitati e residui sugli interventi dovuti alla scarsa conoscenza degli strumenti informatici da parte dei dipendenti, conseguente alla formazione ECDL.

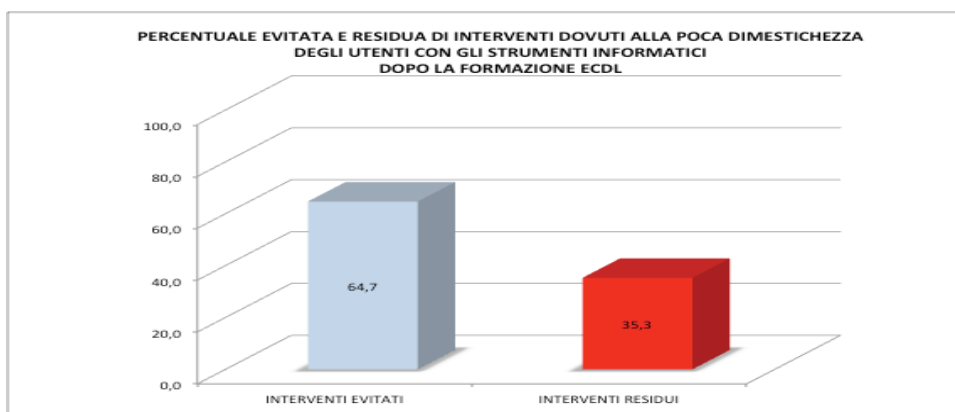


Figura 6

Percentuale degli interventi evitati e residui sugli interventi dovuti alla poca dimestichezza degli utenti con gli strumenti informatici, dopo la formazione ECDL del personale dipendente

Oltre alla diminuzione degli interventi evitabili sui totali, si osserva anche una riduzione del numero globale di richieste di assistenza tecnica dovuto al fatto che le persone formate hanno chiamato di meno l'assistenza perché sono riuscite con le proprie conoscenze a risolvere i problemi che, prima della formazione, non erano capaci di risolvere. Infatti, il numero totale di richieste di assistenza prima della formazione ammontava a 192, mentre nel periodo successivo alla formazione il numero totale di richieste di assistenza è sceso a 69.

Il periodo monitorato dopo la formazione ha una durata di sei mesi, quindi per fare un confronto corretto con i dati relativi al periodo precedente la formazione, che abbraccia un intero anno, abbiamo moltiplicato per due i

valori registrati dopo la formazione, ottenendo così un totale di interventi pari a 138⁷.

Il risultato ottenuto dimostra, quale ulteriore risparmio, che il numero globale di richieste annuali di assistenza tecnica è diminuito di circa il 28%.

Il grafico della figura 7 mostra la riduzione annuale del numero globale di richieste di assistenza nei comuni campione, dovuta alla maggior conoscenza degli strumenti informatici da parte dei dipendenti dopo la formazione.

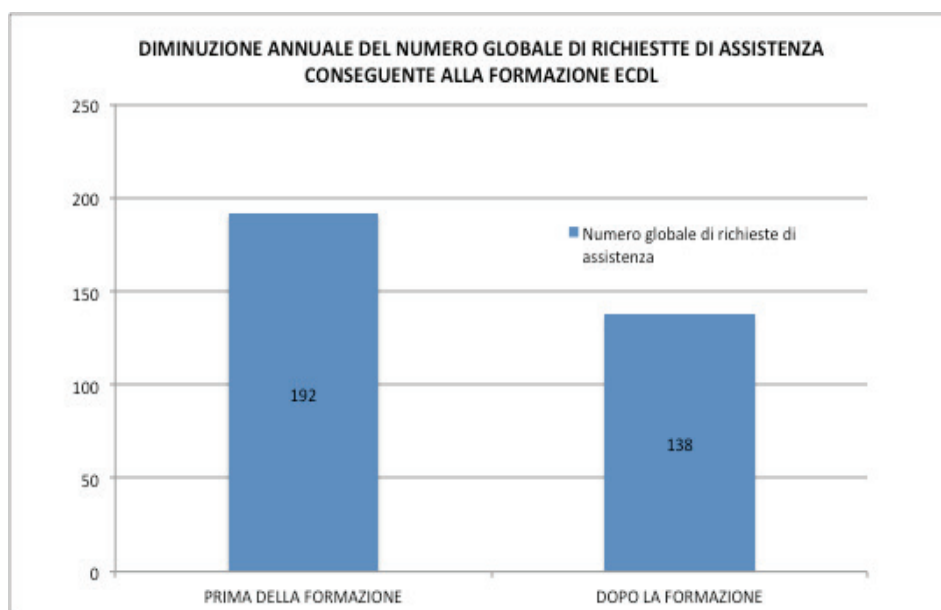


Figura 7

Riduzione annuale del numero globale di richieste di assistenza come conseguenza alla formazione ECDL

4. I costi per l'assistenza tecnica negli Enti Pubblici Locali in Friuli Venezia Giulia

In questa sezione si vuole calcolare il costo dell'assistenza tecnica informatica sostenuto da Comuni, Province e Regione nel Friuli Venezia Giulia, dovuto alla scarsa conoscenza degli strumenti informatici da parte dei dipendenti di questi Enti Pubblici Locali.

Nel paragrafo 2 è stato calcolato il numero medio annuo di interventi di assistenza tecnica per dipendente nei Comuni coinvolti nella ricerca e il

⁷ Il periodo successivo alla formazione comprende i sei mesi, da Maggio 2010 ad Ottobre 2010, includendo anche il mese di Agosto in cui le normali attività sono a regime ridotto.

numero medio annuo di tali interventi dovuti alla poca dimestichezza dei dipendenti con gli strumenti informatici. La ricerca mostra che tali valori sono: 4,82 interventi annui per dipendente nel Comune di Udine, mentre gli interventi evitabili sono 1,84 per dipendente. Nei Comuni del Medio Friuli sono stati calcolati mediamente 8,12 interventi annui per dipendente, e 6,63 quelli evitabili. I valori emersi sono già stati riportati in precedenza nella tabella 2.

Dalla media dei due valori risulta che annualmente ogni dipendente richiede in media 6,47 interventi di assistenza tecnica, mentre quelli dovuti alla poca dimestichezza con gli strumenti informatici sono annualmente in media pari a 4,23 a dipendente. Questi risultati sono riportati nella tabella 6.

MEDIA INTERVENTI ANNUI PER DIPENDENTE IN FRIULI VENEZIA GIULIA	MEDIA INTERVENTI ANNUI PER DIPENDENTE
INTERVENTI EVITABILI	4,23
INTERVENTI TOTALI	6,47

Tabella 6

Il valore medio degli interventi annui per dipendente, evitabili e totali, calcolato sui valori dei Comuni di Udine e del Medio Friuli

Avendo a disposizione questi dati e moltiplicandoli per il numero di dipendenti negli Enti Pubblici Locali del Friuli Venezia Giulia, è possibile calcolare il numero medio annuo di interventi di assistenza richiesti in tutti gli Enti Pubblici Locali della Regione.

Il numero totale di dipendenti negli Enti Pubblici Locali che comprendono i Comuni, le Comunità Montane, le Province e la Regione, è 15.644⁸. Da una recente ricerca condotta da AICA e SDA Bocconi sui costi dell'ignoranza informatica negli Enti Pubblici Locali in Italia, risulta che mediamente il 95% [1] degli impiegati negli Enti Pubblici utilizza normalmente il computer nello svolgimento delle proprie mansioni, quindi nel nostro caso consideriamo un numero di dipendenti pari a 14.862. Si ottiene così che complessivamente vengono richiesti dai dipendenti 96.157 interventi di assistenza tecnica all'anno, di cui 62.866 dovuti alla scarsa conoscenza degli strumenti informatici. La tabella 7 riassume schematicamente il calcolo di tali valori.

⁸ Valore fornito dalla Regione Friuli Venezia Giulia relativo all'anno 2009.

NUMERO DI INTERVENTI ANNUI IN F.V.G.	NUMERO INTERVENTI ANNUI
INTERVENTI EVITABILI	62.866
INTERVENTI TOTALI	96.157

Tabella 7

Il numero totale di interventi negli Enti Pubblici Locali in Friuli Venezia Giulia

Per il calcolo dei costi dell'assistenza tecnica si è considerato il numero di ore necessarie per portare a termine gli interventi registrati. Nel Comune di Udine, sono state necessarie 1824 ore e 20 minuti per eseguire i 4704 interventi di assistenza richiesti nel periodo da Aprile 2009 a Marzo 2010. Nei Comuni del Medio Friuli le ore necessarie per eseguire i 2047 interventi richiesti nello stesso periodo sono state pari a 1431 e 30 minuti. I dati che si riferiscono alle ore di assistenza fornite ai Comuni di Udine e del Medio Friuli sono riportati in tabella 8.

	ORE DI ASSISTENZA
UDINE	1824h, 20m
MEDIO FRIULI	1431h 30m

Tabella 8

Numero di ore di assistenza tecnica corrispondenti agli interventi eseguiti nei Comuni di Udine e del Medio Friuli, nel periodo da Aprile 2009 a Marzo 2010

Per ottenere la media di ore necessarie per eseguire un intervento di assistenza, si è diviso il totale delle ore impiegate in assistenza per il numero di interventi eseguiti. Si è così riscontrato che in media sono necessari circa 23 minuti per eseguire un intervento di assistenza tecnica nel Comune di Udine, mentre vengono impiegati circa 42 minuti per eseguire un intervento di assistenza nei Comuni del Medio Friuli. Dai due valori si evince una media di 32,61 minuti per intervento. Il calcolo del numero medio di minuti richiesti per gestire gli interventi di assistenza nei Comuni considerati è schematizzato in tabella 9.

TEMPO NECESSARIO PER INTERVENTO	ORE DI ASSISTENZA	NUMERO INTERVENTI	MEDIA MINUTI PER INTERVENTO	MEDIA GLOBALE MINUTI PER INTERVENTO
UDINE	1824h 20m	4704	23,27	32,61
MEDIO FRIULI	1431h 30m	2047	41,96	

Tabella 9

Calcolo del numero medio di minuti necessari per gestire un intervento nei Comuni di Udine e del Medio Friuli

Per trovare il costo economico medio di un intervento di assistenza tecnica si può moltiplicare il tempo medio richiesto per gestire un intervento per il costo orario del personale interno o esterno all'ente che ha prestato l'assistenza tecnica. Il costo del personale dedicato all'assistenza tecnica informatica è stato stimato in 24,25 euro all'ora⁹. Si è così ottenuto che un intervento di assistenza costa mediamente 13,2 euro¹⁰. I dati riguardanti il costo medio di un intervento sono riportati in tabella 10.

COSTO MEDIO PER INTERVENTO	MEDIA MINUTI PER INTERVENTO	COSTO MEDIO ORARIO DEL PERSONALE CHE HA PRESTATO ASSISTENZA TECNICA	COSTO MEDIO PER INTERVENTO
	32,61	24,25 euro	13,2 euro

Tabella 10

Il costo medio per intervento è ottenuto moltiplicando il tempo medio necessario per intervento, per il costo orario del personale dedicato all'assistenza tecnica

Sapendo che il costo medio che viene sostenuto dagli Enti Pubblici Locali per la gestione di un singolo intervento di assistenza tecnica è pari a 13,2 Euro, è stato possibile ottenere il costo medio annuo per l'assistenza

⁹ L'azienda, che fornisce assistenza tecnica ai comuni del Medio Friuli presi in esame, ha un compenso di 25,00 Euro all'ora. Il comune di Caorle ha fornito lo stesso dato per il costo orario dei suoi dipendenti. Per il costo medio orario dei dipendenti dei comuni del Friuli Venezia Giulia, il dato è stato elaborato nel seguente modo: il numero di dipendenti dei Comuni, Province e Regione del Friuli-Venezia Giulia è 15.078 (ANCI FVG 2009 e Direzione Centrale Pianificazione Territoriale, Autonomie Locali e Sicurezza FVG 2008); il costo annuo del personale dipendente di Comuni, Province e Regione è 602.159.002,21 € (ANCI FVG 2007). Ricavando così il costo annuo per dipendente e dividendolo per le 1.699,2 ore lavorative annuali (dato reperito dalle informazioni del Comune di Udine riguardanti il contratto comparto unico dei dipendenti degli Enti Locali della regione FVG), risulta un costo medio orario di 23,502 € per dipendente.

¹⁰ Non sono stati considerati i tempi e i costi viaggio.



tecnica dovuta alla scarsa conoscenza degli strumenti informatici da parte dei dipendenti degli Enti Pubblici Locali del Friuli Venezia Giulia, moltiplicando il costo del singolo intervento per il numero medio annuo di interventi evitabili, stimati in 62.866, come già riportato in tabella 7. Il costo medio annuo per l'assistenza evitabile in Friuli Venezia Giulia risulta quindi pari a 829.831 Euro.

Il costo medio calcolato per singolo intervento, riportato in tabella 10, corrisponde in realtà solo al costo sostenuto dall'Ente Locale per la soluzione, da parte dei tecnici informatici referenti dell'assistenza, del problema sollevato dal dipendente che ha richiesto l'intervento.

Dalla ricerca sperimentale è emerso che questo costo è spesso un costo esterno che l'Ente Locale deve sostenere per i servizi del professionista informatico o dell'azienda esterna che fornisce l'assistenza tecnica. Molto spesso, infatti, nei comuni piccoli e medi l'assistenza tecnica è fornita da un ente esterno, come ad esempio succede nei Comuni considerati nel Medio Friuli, dove l'assistenza tecnica informatica è affidata all'azienda Qnet s.r.l. di Codroipo. Anche nei comuni più grandi e strutturati del Friuli Venezia Giulia, molto spesso il servizio di assistenza viene dato in *outsourcing*. Nel caso del Comune di Udine, per esempio, il servizio di gestione Help-Desk è sostanzialmente in *outsourcing*.

Il costo per una richiesta di assistenza tecnica non si riduce però al solo costo del professionista informatico che interviene per risolvere il problema, ma comprende anche il costo interno dato dall'improduttività del dipendente che attende che l'intervento richiesto venga portato a termine. Di conseguenza il costo per la gestione di un intervento di assistenza è almeno il doppio, dovuto al tempo di attività della persona chiamata a risolvere il problema, e il tempo di inattività del dipendente che ha sollecitato l'intervento. In realtà, il tempo di inattività del dipendente che ha richiesto l'assistenza è ancor più alto se si considera il tempo per l'attesa dell'arrivo del referente informatico per l'assistenza tecnica che può essere a volte maggiore del tempo necessario per l'esecuzione dell'intervento stesso.

Considerando quindi un costo almeno doppio per intervento, si ottiene un valore globale pari a 26,4 Euro. Moltiplicando il costo medio globale di un singolo intervento per il numero medio annuo di interventi evitabili si ottiene un costo globale annuo per l'assistenza evitabile in Friuli Venezia Giulia di almeno 1.659.662 Euro. I dati riguardanti il costo annuo per l'assistenza evitabile in regione sono riportati in tabella 11.

COSTO MEDIO GLOBALE ANNUO PER L'ASSISTENZA TECNICA EVITABILE IN FRIULI VENEZIA GIULIA	COSTO MEDIO GLOBALE PER INTERVENTO	NUMERO MEDIO ANNUO DI INTERVENTI EVITABILI	COSTO GLOBALE ANNUO
	26,4 euro	62.866	1.659.662 euro

Tabella 11

Il costo totale per l'assistenza tecnica evitabile in Friuli Venezia Giulia è dato dal prodotto fra il costo medio per intervento e il numero medio annuo di interventi evitabili in regione

È interessante a questo punto comprendere quale sia il costo annuo potenzialmente risparmiabile per ogni dipendente degli Enti Pubblici Locali. Il valore, ottenuto dividendo il costo annuo riportato in tabella 11 per il numero totale di dipendenti negli Enti Pubblici Locali del Friuli Venezia Giulia, è risultato pari a 111,5 Euro per dipendente.

5. I possibili risparmi nella Pubblica Amministrazione Locale attraverso la formazione e certificazione ECDL – il caso Friuli Venezia Giulia

Nella recente ricerca condotta da AICA in collaborazione con SDA Bocconi, pubblicata nel volume “L’ignoranza informatica – il costo nella pubblica amministrazione locale [1]”, è stato effettuato uno studio a livello nazionale sui costi dell’ignoranza informatica negli Enti Pubblici Locali.

Da questa ricerca è emerso che il ritorno economico della formazione informatica dei dipendenti degli Enti Pubblici Locali, oltre alla riduzione del tempo perso per la poca dimestichezza con gli strumenti informatici, comprende anche il miglioramento della produttività individuale nel lavoro al computer, calcolato in 3.828 Euro annui per dipendente. Tale dato non ci sorprende, infatti, dalle ricerche precedentemente svolte da AICA e SDA Bocconi risulta che il miglioramento della produttività, conseguente alla formazione e certificazione ECDL del personale, è di circa 7.100 euro nel settore bancario, circa 3.450 euro nel settore sanitario nazionale, circa 2.160 euro nel settore aziendale italiano [2] e 1520 Euro nella pubblica amministrazione centrale [3] che, globalmente potrebbero far risparmiare al sistema economico italiano fino a 18 miliardi di Euro all’anno.

Se al miglioramento della produttività nella Pubblica Amministrazione locale (3.828 Euro annui) sommiamo il costo annuo risparmiabile per dipendente, dovuto all’assistenza tecnica e all’improduttività, calcolato in 111,5 Euro, si ottiene un risparmio annuo di quasi quattromila Euro per dipendente, come evidenziato in tabella 12.

RISPARMIO POTENZIALE ANNUO PER DIPENDENTE NEGLI ENTI LOCALI IN FRIULI VENEZIA GIULIA	COSTO ASSISTENZA TECNICA RISPARMIABILE PER DIPENDENTE	MIGLIORAMENTO POTENZIALE DELLA PRODUTTIVITÀ PER DIPENDENTE	RISPARMIO COMPLESSIVO PER DIPENDENTE
	111,5 euro	3.828 euro	3939,5 euro

Tabella 12

Risparmio potenziale annuo per dipendente dato dalla somma del costo evitabile per assistenza tecnica e improduttività e dal miglioramento potenziale della produttività

Moltiplicando il potenziale risparmio annuo per dipendente per il numero di dipendenti degli Enti Pubblici Locali in Friuli Venezia Giulia che utilizzano normalmente il computer nel loro lavoro, otteniamo un beneficio potenziale annuo pari a 58.548.849 Euro.

Dividendo il valore del beneficio complessivo per il numero di abitanti in Regione, pari a 1.231.257¹¹, si ottiene un potenziale beneficio annuo per cittadino del valore di circa 47,5 Euro. In tabella 13 sono riportati il beneficio complessivo annuo per la Pubblica Amministrazione locale e il beneficio annuo per cittadino in Friuli Venezia Giulia.

VALORI ANNUI DEI BENEFICI DELLA FORMAZIONE NELLA P.A. LOCALE IN FVG	BENEFICIO COMPLESSIVO PER LA P.A. LOCALE	BENEFICIO PER CITTADINO
	58.548.849 euro	47,5 euro

Tabella 13

Quadro riassuntivo dei benefici annui della formazione e certificazione ECDL del personale nella Pubblica Amministrazione locale in Friuli Venezia Giulia

Nel quadro dei benefici derivanti dalla formazione non c'è solo quello economico ma ve n'è un altro apprezzabile sotto il profilo motivazionale dei dipendenti. Infatti, la formazione contribuisce ad aumentare l'efficacia del lavoratore nello svolgimento dei propri compiti e crea le condizioni per una maggiore soddisfazione sul lavoro, ad esempio, diminuisce il senso di frustrazione che deriva dall'incapacità di usare appropriatamente gli strumenti informatici e dalla sensazione di "perdere" tempo con il computer. A fronte di un costo sostenuto una tantum per un corso e per la certificazione ECDL, che può ammontare a circa 300-400 Euro per ogni partecipante, si ottiene un ritorno nell'ordine di 4.000 Euro all'anno per ogni addetto "formato e certificato".

¹¹ Fonte: rapporto ISTAT 2009.

Si sa che formare non significa fornire semplicemente una serie di nozioni, ma creare un progetto a lungo termine che rinforzi le competenze del soggetto. Considerata la velocità di evoluzione delle tecnologie e degli applicativi di produttività individuale, è necessario programmare, successivamente alla prima formazione, dei corsi di aggiornamento (anche di breve durata) per il personale, ogni 3 anni circa, seguendo il paradigma della formazione continua.

Considerando, quindi, che l'investimento sulla formazione nell'uso dell'informatica garantisce un beneficio apprezzabile almeno nel triennio successivo, si ottiene che a fronte di un costo sostenuto di circa 300-400 Euro per la formazione e certificazione ECDL di un dipendente, si ha un ritorno di circa 11.800 Euro in tre anni per dipendente. Il beneficio netto è pertanto di circa 11.400 Euro per dipendente.

Risulta evidente che la formazione ha un ritorno d'investimento elevatissimo¹². Inoltre, si può notare come già nel corso del primo anno rientri completamente il costo sostenuto per la formazione. In tabella 14 sono riportati il costo sostenuto per il corso e la certificazione ECDL, il ritorno conseguente nel triennio successivo, e il beneficio netto, per dipendente.

COSTO DELLA FORMAZIONE ECDL E BENEFICIO CONSEGUENTE PER DIPENDENTE	COSTO PER FORMAZIONE E CERTIFICAZIONE ECDL PER DIPENDENTE	RITORNO PER DIPENDENTE (calcolato in 3 anni)	BENEFICIO NETTO PER DIPENDENTE (calcolato in 3 anni)
	300-400 euro	11.800 euro	11.400 euro

Tabella 14
Ritorno dell'investimento sulla formazione e certificazione ECDL

6. I possibili risparmi nel Sistema Sanitario della Regione Friuli Venezia Giulia

Il Servizio Sanitario Regionale conta 20.190¹³ dipendenti suddivisi fra il personale sanitario, professionale, tecnico e amministrativo. L'utilizzo dell'informatica fra il personale nei suddetti ruoli è chiaramente diverso, l'utenza informatica risulta costituita in gran parte dal personale impiegatizio che usa questa tecnologia per finalità gestionali e contabili. Da uno studio condotto da AICA e SDA Bocconi sui costi dell'ignoranza informatica nel Sistema Sanitario Nazionale risulta che, per ciascun ruolo professionale, la percentuale di personale che utilizza l'informatica in maniera rilevante nel proprio lavoro va dal 90% del personale impiegatizio

¹² $ROI = \text{Valore del beneficio} / \text{Costo sostenuto} = 3.939,5 \times 3 / 400 = 29,55 = 2.955\%$

¹³ Valore fornito dalla Regione Friuli Venezia Giulia relativo all'anno 2010.

al 20% del personale tecnico e infermieristico. I valori percentuali per ciascun ruolo sono esposti in tabella 15 [4].

CATEGORIE DI ADDETTI	UTILIZZATORI DI INFORMATICA
Personale impiegatizio	90%
Personale tecnico e infermieristico	20%
Personale medico	80%
Medici di base	76%
Personale delle farmacie	100%

Tabella 15

Percentuale del personale che utilizza l'informatica per ciascun ruolo professionale

La suddivisione del personale dipendente del Sistema Sanitario Regionale in Friuli Venezia Giulia nei diversi ruoli professionali, è riportato in tabella 16.

CATEGORIE DI ADDETTI	PERSONALE
Personale amministrativo	1.937
Personale tecnico e infermieristico	4.400
Personale medico	13.811
Altre figure professionali	42
Totale	20.190

Tabella 16

Suddivisione del personale del Sistema Sanitario Regionale in Friuli Venezia Giulia nei diversi ruoli professionali

Da questi dati si evince che sul totale dei dipendenti del Sistema Sanitario Regionale in Friuli Venezia Giulia, 13.672 fanno un uso considerevole dell'informatica nel loro lavoro (senza contare che il personale che non figura tra questi, utilizza comunque gli strumenti informatici, ma in maniera non prevalente rispetto alle attività previste dal loro ruolo).

Da una rielaborazione dei dati della già citata ricerca sul costo dell'ignoranza informatica nella sanità italiana, risulta che il valore annuale dell'aumento di produttività è di circa 3.450 Euro per dipendente. Estendendo anche al personale sanitario il costo annuo risparmiabile calcolato per dipendente negli Enti Pubblici Locali, dovuto all'assistenza tecnica e all'improduttività, pari a 111,5 Euro, si ottiene un risparmio annuo complessivo di circa 3.560 Euro a dipendente. Considerando i dipendenti che fanno uso dell'informatica, si ottiene un potenziale beneficio annuo di circa 48.672.320 Euro per l'intero Sistema Sanitario Regionale.

Dividendo il valore del beneficio complessivo per il numero di abitanti in Regione, si ottiene un risparmio potenziale annuo per cittadino del valore di circa 39,5 Euro. I benefici annui potenziali per il Sistema Sanitario Regionale e per cittadino sono evidenziati in tabella 17.

VALORI ANNUI DEI BENEFICI DELLA FORMAZIONE NELLA SANITÀ IN FRIULI VENEZIA GIULIA	BENEFICIO COMPLESSIVO PER IL SISTEMA SANITARIO REGIONALE	BENEFICIO PER DIPENDENTE	BENEFICIO PER CITTADINO
	48.672.320 euro	3.560 euro	39,5 euro

Tabella 17

Quadro riassuntivo dei benefici annui della formazione e certificazione ECDL del personale nel Sistema Sanitario Regionale in Friuli Venezia Giulia

7. Conclusioni

Considerando complessivamente i costi dovuti alla scarsa conoscenza degli strumenti informatici e quindi risparmiabili con la formazione e certificazione ECDL e il potenziale miglioramento della produttività dei dipendenti nella Pubblica Amministrazione locale e nel Sistema Sanitario Regionale in Friuli Venezia Giulia, si ottiene un potenziale risparmio annuo di circa 107.220.000 Euro. La tabella 18 riassume i potenziali risparmi annui in Friuli Venezia Giulia.

RISPARMI POTENZIALI ANNUI IN FRIULI VENEZIA GIULIA	
Pubblica Amministrazione locale	58.548.849 euro
Sistema Sanitario Regionale	48.672.320 euro
Totale	107.221.169 euro

Tabella 18

Quadro riassuntivo dei risparmi potenziali in Friuli Venezia Giulia attraverso la formazione e certificazione ECDL del personale

Come illustrato in tabella 13, il risparmio potenziale nella Pubblica Amministrazione locale si traduce in un beneficio per cittadino di circa 50 Euro all'anno che, se sommato al beneficio annuo per cittadino dovuto al risparmio potenziale nella Sanità Regionale, di circa 40 euro (si veda la tabella 17), permette di raggiungere un valore di quasi 90 Euro all'anno. Il calcolo del risparmio complessivo per cittadino è illustrato in tabella 19.

RISPARMO POTENZIALE ANNUO PER CITTADINO IN FRIULI VENEZIA GIULIA	
Derivante dai risparmi nella P.A. locale	47,5 euro
Derivante dai risparmi nel S.S.R.	39,5 euro
Totale	87 euro

Tabella 19

Risparmio potenziale annuo per cittadino considerando i risparmi possibili nella Pubblica Amministrazione locale e nel Sistema Sanitario Regionale

Questo valore può essere facilmente esteso ad altre realtà provinciali o regionali sul territorio nazionale per ottenere una stima dei benefici che si possono ottenere grazie alla formazione e certificazione ECDL del personale.

Bibliografia

[1] P. Bielli, P.F. Camussone, F. Sala: *L'ignoranza informatica – il costo nella pubblica amministrazione locale*, Egea 2011.

[2] P.F. Camussone, G. Occhini, F. Sala: *Le conoscenze informatiche in Italia*, Egea, 2006.

[3] P.F. Camussone, F. Sala, P. Bielli: *L'ignoranza informatica. Il costo nella pubblica amministrazione centrale*, Egea, 2008.

[4] E. Borgonovi, P.F. Camussone, G. Occhini: *L'ignoranza informatica: il costo nella sanità*, pag. 145, McGraw Hill, 2004.

Biografie

Antonio Piva laureato in Scienze dell'Informazione, Vice Presidente dell'ALSI (Associazione Nazionale Laureati in Scienze dell'Informazione e Informatica) e Presidente della commissione di informatica giuridica. Docente a contratto di diritto dell'ICT e qualità all'Università di Udine. Consulente sistemi informatici e Governo Elettronico nella PA locale, valutatore di sistemi di qualità ISO9001, Ispettore e Coordinatore AICA Nord Est.
E-mail: antonio@piva.mobii.it

Michele Baggi, laureato con lode in Informatica presso l'Università di Udine, Ph.D. con lode in Informatica presso l'Università di Siena e l'Università Politecnica di Valencia con marchio di garanzia europea "Doctor Europæus". Certificato EUCIP Core, autore di molte pubblicazioni, ha maturato diverse esperienze internazionali e presso gli Enti Locali.
E-mail: friulanbags@gmail.com

A Smart Cities Vision Of The Future: An Evaluation Approach

Patrizia Lombardi¹, Silvia Giordano²

¹Politecnico di Torino, School of Architecture and Planning
Viale Mattioli 39, 10125 Torino (TO)
patrizia.lombardi@polito.it

²SITI Innovation Research Centre, Politecnico di Torino
Corso Castelfidardo 30/A, 10138 Torino (TO)
silvia.giordano@polito.it

Although there is no agreement on the exact definition of “Smart city”, a number of main dimensions of a smart city has been identified through a literature review exercise (Giffinger et al., 2007; Van Soom, 2009; Fusco Girard et al., 2009). These include: smart governance; smart economy; smart people; smart living; smart environment.

In order to explore the concept of smart city, a revised triple helix model has been recently proposed which focuses on the production of knowledge by Universities and Government and the production of innovation which are patented by Industry and University as an index of intellectual capital (Etzkowitz, 2008; Caragliu et al, 2009; Deakin, 2010; Lombardi et al., 2011). This model presupposes that the three helices operate in a complex urban environment, where market demand, governance, civic involvement and citizens’ characteristics, along with cultural and social capital endowments, shape the relationships between the traditional helices of university, industry and government.

Based on the components above and on an adjusted version of the triple helix approach which includes civil society as major actor (Etzkowitz and Zhou, 2006), this paper suggests a novel framework for classifying smart city performance indicators. This framework includes more than 60 indicators classified in the five aforesaid clusters: smart governance; smart economy; smart people; smart living; smart environment. The source of data are both literature review including EU projects’ reports and Urban Audit dataset and indicators selected from statistics of European commission, European green city index, TIS-SUE, Trends and Indicators for Monitoring the EU Thematic Strategy on Sustainable Development of Urban Environment and smart cities ranking of European medium-sized cities.

This framework will be augmented using the Analytic Network Process to model cluster and begin measuring the performance of smart cities. This model allows interactions and feedbacks within and between clusters, providing a process to derive ratio scales priorities from elements (Saaty, 2005). This interconnected model will be used for investigating the relations between smart cities components, actors and strategies to which the smart cities are moving to.

Synthetically, the methodology involves the following main steps (Saaty, 2001; 2005).

- I. Structuring the decision-making model. This activity involves an identification of both of the elements constituting the decision problem and their relationships. The network model is constituted by various clusters of elements, and alternatives or options from which to chose.*
- II. Developing pairwise comparison of both elements and clusters to establish relations within the structure..*
- III. Achievement of the final priorities.*

The final model provides a comprehensive decision making approach which is able to evaluate Smart cities’ performances in order to enhance their strategy. In this paper, four policy visions of a European Smart city are suggested, as reported in the “Urban Europe” Joint Programme Initiatives (see report by: P. Nijkamp, K. Kourtik, 2011). The four visions are:

- Connected City (smart logistic & sustainable mobility),*
- Entrepreneurial City (economic vitality),*

- *Liveable City (ecological sustainability) and*
- *Pioneer City (social participation & social capital).*

A pioneer application of the model is illustrated. This exercise has been conducted within a focus group, involving a number of experts in different disciplines. The results show that the Entrepreneurial City is the policy vision with higher priorities in all the sectors considered in the model, i.e. Universities, Government, Industry and Civil Society.

This vision assumes that in the current and future global and local competition, Europe can only survive, if it is able to maximize its innovative and creative potential in order to gain access to emerging markets outside Europe. The policy implication of such a result are finally discussed, involving all the followings:

- *A high degree of entrepreneurial activities and a constant flow of new firm creation is a prerequisite for finding a new role within the new global economic landscape. Innovation and creativeness are thus the necessary ingredients for entrepreneurial cities in Europe.*
- *Special emphasis has to be given to new architectures, building technologies, intra-urban mobility solutions, public space management, e.g. for lighting or citizen information management, integrated urban energy planning and management and ICT-based solutions that offer various opportunities for new urban design and management.*
- *New requirements for efficient, effective and reliable infrastructures (such as energy, ICT, water, waste treatment and management etc) may occur. Since an appropriate infrastructure is essential for cities' attractiveness for companies and people alike and therefore their economic development, emphasis has to be given to the determination of these requirements within the scope of cities as complex systems.*

The next authors' task will be to develop a rating system based on the ANP model in order to verify whether: the cities are "smart"; if not, whether they are moving in the right direction.

Keywords: *Analytic Network Process, smart cities, triple helix approach, sustainable community development.*

1. Introduction

Traditional neoclassical theories of urban and regional growth are often inadequate in analyzing the urban transformations of the 21st century. Theories of regional competitiveness try to fill this gap by focusing not only on the city's endowment of hard infrastructures (physical capital), but also, and increasingly, on the availability and quality of knowledge communication, social and environmental infrastructures (intellectual, social and environmental capital).

The application of information technology (ICT) in the context of future cities is often indicated by the notion of Smart City. This concept has been quite fashionable in the policy arena in the recent years. Compared with the concept of digital city or intelligent city, the main focus is not limited to the role of ICT infrastructure, but mainly on the role of human capital/education, social and relational capital and environmental issues. These are considered important drivers of urban growth.

In order to explore the concept of smart city, a revised triple helix model has been recently proposed which focuses on the production of knowledge by Universities and Government and the production of innovation which are patented by industry and university as an index of intellectual capital (Etzkowitz, 2008; Caragliu et al, 2009; Deakin, 2010; Lombardi et al., 2011). This model presupposes that the three helices operate in a complex urban environment, where market demand, governance, civic involvement and citizens' characteristics, along with cultural and social capital endowments, shape the relationships between the traditional helices of university, industry and government. The results of the above study has shown the analysis to baseline the development of smart cities in terms of their dual roles as generators of intellectual capital, creators of wealth and regulator of standards (University, Industry, and Government), as well as supporting the social learning and knowledge-transfer abilities which are needed to meet the requirements of their regional innovation systems.

Although this analysis has been a useful start for understanding the main “governance” component of smart cities, it has a number of limitations as follows:

- it does not consider other fundamental aspects of a smart cities, related to ecological sustainability;
- it does not recognize the numerous relationships and feed backs between categories which are dependant upon the interconnected and systemic nature of the aspects involved;
- it does not take into account an essential actor of a smart city, especially in the field of sustainable development policies, planning and management, such as the Civil Society.

The paper proposes a different analytical model which involves the civil society as one of the main key actor, alongside the university, the industry and the government (Etzkowitz and Zhou, 2006). This new triple helix framework is used as a “control hierarchy” in a Analytical Network Process (ANP) application with the aim to identify and evaluate the performances of a smart city. The next two sections illustrate both the adopted model and the performance indicators used for this exercise, alongside the main results obtained from this pilot application.

2. Modelling the Smartcity's performance

Although there is no agreement on the exact definition of “Smart city”, a number of main dimensions of a smart city has been identified through a literature review exercise (Giffinger et al., 2007; Van Soom, 2009; Fusco Girard et al., 2009). These include: smart governance; smart economy; smart people; smart living; smart environment.

Based on the components above and on an adjusted version of the triple helix approach which includes civil society as major actor (Etzkowitz and Zhou, 2006), a novel framework for classifying smart city performance indicators has been developed, as shown in Table (1). As anyone can notice, both the main components/activities and the main actors/helices of a Smart city are represented.

Clusters Revised Triple Helix	SMART Governance	SMART Economy	SMART Human Capital	SMART Living	SMART Environment
	INDICATORS				
UNIVERSITY	No of universities, research centres in the city	Public expenditure on R&D - % of GDP per head of city	% of population aged 15-64 with secondary level education living in Urban Audit	% of professors & researchers involved in international projects and exchange	An assessment of the ambitiousness of CO2 emissions reduction strategy
	No. courses entirely downloadable from the internet / Total No. Courses	Public expenditure on education - % of GDP per head of city	% of population aged 15-64 with high education living in Urban Audit	Number of grants for international mobility per year	An assessment the extensiveness of city energy efficiency standards for buildings
		Number of research grants funded by international projects	% of inhabitants working in education and in research & development sector	% of accessible courses for PWD	
GOVERNEMENT	E-government on-line availability (% of the 20 basic services which are fully available online)	Gross Domestic Product per head of city	Voter turnout in national and EU parliamentary elections	Proportion of the area in recreational sports and leisure use	Total annual energy consumption, in gigajoules per head
		Debt of municipal authority per resident	Share of female city representatives	Green space (m2) to which the public has access, per capita	Efficient use of electricity (use per GDP)
	Percentage households with computers	Median or average disposable annual household income	City representatives per resident	Number of public libraries	Total annual water consumption, in cubic metres per head
		Unemployment rate		Number of theaters & cinemas	Efficient use of water (use per GDP)
	Percentage households with Internet access at home	Energy intensity of the economy - Gross inland consumption of energy divided by GDP		Health care expenditure - % of GDP per head of city	Area in green space (m2)
					Geenhouse gas emissions intensity of energy consumption
				Tourist overnight stays in registered accommodation in per year per resident	An assessment of the comprehensiveness of policies to contain the urban sprawl and to improve and monitor environmental performance
				Urban population exposure to air pollution by particulate matter - micrograms per cubic metre	

CIVIL SOCIETY	E-government usage by individuals aged 16 to 74 who have used the Internet, in the last 3 months, for interaction with public authorities)	% of projects funded by civil society	Foreign language skills	Total book loans and other media per resident	The total percentage of the working population travelling to work on public transport,
			Participation in Life-long learning (%)	Museums visits per inhabitant	An assessment of the extent to which citizens may participate in environmental decision-making
			Individuals' level of computer skills	Theatre & cinema attendance per inhabitant	An assessment of the extensiveness of efforts to increase the use of cleaner transport
			Individuals' level of internet skills		% of citizens engaged in environmental and sustainability oriented activity
INDUSTRY	Number of research grants funded by companies, foundations, institutes / No annual scholarships	Employment rate in: - High Tech & creative industries - Renewable energy & energy efficiency systems - Financial intermediation and business activities - culture & entertainment industry - commercial services - transport and communication - hotels and restaurants	Patent applications per inhabitant	Number of enterprises adopting ISO 14000 standards	The percentage of total energy derived from renewable sources, as a share of the city's total energy consumption, in terajoules
					Combined heat and power generation - % of gross electricity generation
					Rate of rycycled waste per total kg of waste produced
					Total CO2 emissions, in tonnes per head (2)
					% of new buildings and reneovation which were
		All companies (total number)	Employment rate in knowledge-intensive sectors	Rate of people undertaking industry based training	
		Number of local units manufacturing High Tech & ICT products			
		Companies with HQ in the city quoted on national stock market			
		Components of domestic material consumption			

Table 1 Smart Cities components and indicators

The source of data for this table are both literature review including EU projects' reports and Urban Audit dataset and indicators selected from statistics of European commission, European green city index, TISSUE, Trends and Indicators for Monitoring the EU Thematic Strategy on Sustainable Development of Urban Environment and smart cities ranking of European medium-sized cities. This includes more than 60 indicators classified in the five aforesaid clusters.

Furthermore, the relations between the Smart Cities components have been identified by using a focus group where the participants were selected on the basis of their specific competences in the various fields involved by a smart city (components highlighted in Table 1 by columns).

This modeling exercise has been supported by an Analytic Network Process (ANP). This consists of clusters (*i.e.* groups of homogeneous elements of a decision problem), elements (*i.e.* nodes of the network), interrelationship between clusters, and interrelationship between elements. It allows interactions and feedback within and between clusters and provides a process to derive ratio scales priorities from the elements (Saaty, 2005).

Synthetically, the ANP methodology involves the following main steps (Saaty, 2001; 2005).

- I. Structuring the decision-making model. This activity involves an identification of both of the elements constituting the decision problem and their relationships. The network model is constituted by various clusters of elements, and alternatives or options from which to chose. Each element can have influence and inter-dependence relations: this can be a "source", that is an origin of a path of influence, or a "sink", that is a destination of paths of influence. There are two kinds of interdependences: between elements related to different clusters ("outer" or external connections) and within the same cluster ("inner" or internal relations). The latter relation is identified as a "loop".
- II. Developing pairwise comparison of both elements and clusters to establish relations within the structure. In this step, a series of pairwise comparisons are made by participants to the decision making process (usually experts, managers and citizens representatives) to establish the relative importance of decision elements with respect to each component of the network. In pairwise comparisons, a ratio scale of 1-9 number is used (named, fundamental scale or Saaty' scale). The numerical judgments established at each level of the network form pair matrixes which are used to derive weighted priority vectors of elements (Saaty, 2001).

- III. Achievement of the final priorities. In order to obtain the global priority vector of the elements, including the alternatives, the mathematical approach encompass the use of “supermatrices” (portioned matrices composed by sub-matrices consisting of priority weight vectors of the elements which have been evaluated). A final supermatrix is obtained at the end of the process, containing the global priority vector of the elements.

As required by step I, a complex model was developed which involves all the mentioned clusters of a smart city, i.e. *Smart Governance* (related to participation); *Smart Human Capital* (related to people); *Smart Environment* (related to natural resources); *Smart Living* (related to the quality of life) and *Smart Economy* (related to competitiveness).

The relationships between indicators (and clusters) has been identified by using a “control hierarchy” (Saaty, 2001) composed by the four axes of the adopted Triple Helix, i.e. university, industry, government and civil society. Figure 1 shows this control hierarchy.

Each axe is organised by a sub-network consisting of:

- the five clusters representing the above mentioned smart cities component/activities including the relative selected indicators;
- a cluster of alternatives. These are identified in the four policy visions (or prototypes) of smart cities in Europe, as derived from the “Urban Europe” Joint Programme Initiatives (see report by: P. Nijkamp, K. Kourtik, 2011): *Connected City* (smart logistic & sustainable mobility), *Entrepreneurial City* (economic vitality), *Liveable City* (ecological sustainability) and *Pioneer City* (social participation & social capital).

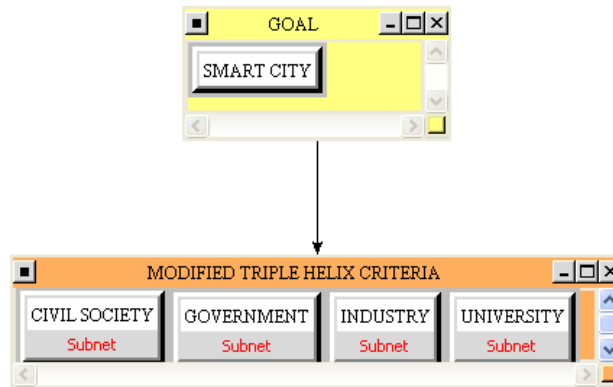


Fig. 1 The main network

As an example, Figure 2 shows the Civil Society sub-network.

It can be noted that both Smart Governance and the Smart Economy clusters include only one element, respectively: “*E-gov usage by individuals*” and “*Percentage of projects funded by civil society*”. In the Smart Environment and in the Smart Human Capital there are inner connections (as shown by a loop in the Figure 2 below), because the indicator named “*Foreign language skills*” influences other nodes included in the same cluster, such as “*Individuals level of computer skills*” and “*Individuals level of internet skills*”. In addition, bi-directional relationships have been identified between Smart Economy and Smart Environment (e.g. “*Percentage of projects funded by civil society*” is the cause of the “*Percentage of citizens engaged in environmental and sustainability oriented activities*”) and between Smart Human Capital and Smart Living (the indicator named “*Participation in life-long learning*” is connected by all smart living nodes, including “*Museums visit per inhabitant*”, “*Theatre & cinema attendance per inhabitant*” and “*Total book loans and other media per resident*”).

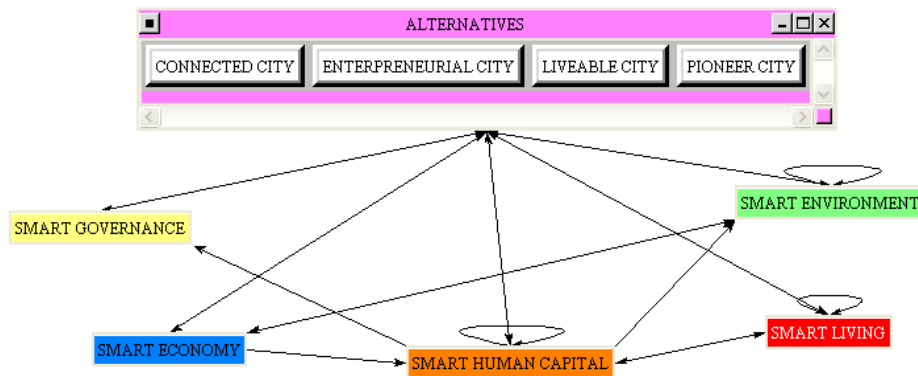


Fig. 2 The Civil Society sub-network

3. Assessing the Smartcity's performance

As required in Step II of the ANP methodology, an assessment exercise was conducted with the aim to develop pair wise comparisons, of both elements (or nodes) and clusters;

Figure 3 below shows one of the several pair matrixes which are composed and used to derive weighted priority vectors of elements (Saaty, 2001). In each pairwise comparisons matrix, a ratio scale of 1-9 number is used. In particular, Figure 3 shows the cluster comparison matrix for the alternatives.

Cluster comparisons for "ALTERNATIVES"

File Computations Misc Help

Graphic Verbal Matrix Questionnaire

SMART ENVIRONMENT is moderately more important than SMART ECONOMY

1. SMART ECONOMY	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART ENVIRONMENT
2. SMART ECONOMY	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART GOVERNANCE
3. SMART ECONOMY	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART HUMAN CAPITAL
4. SMART ECONOMY	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART LIVING
5. SMART ENVIRONMENT	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART GOVERNANCE
6. SMART ENVIRONMENT	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART HUMAN CAPITAL
7. SMART ENVIRONMENT	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART LIVING
8. SMART GOVERNANCE	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART HUMAN CAPITAL
9. SMART GOVERNANCE	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART LIVING
10. SMART HUMAN CAPITAL	>=9.5	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	>=9.5	No comp.	SMART LIVING

Fig. 3 Pair wise cluster comparison using the Saaty's scale

The achievement of the final priorities of all the elements included in the model is obtained in step III. This includes the overall priorities of the alternatives obtained by synthesizing the priorities of the alternatives from all the sub networks. The final results are:

1. Entrepreneurial City (48%)
2. Pioneer City (20%)
3. Liveable City (17%)
4. Connected City (13%)

The Entrepreneurial City. This image assumes that in the current and future global and local competition, Europe can only survive, if it is able to maximize its innovative and creative potential in order to gain access to emerging markets outside Europe; cities are then spearheads of Europe's globalization policy.

The Pioneer City. This image refers to the innovative ‘melting pot’ character of urban areas in the future, which will show an unprecedented cultural diversity and fragmentation of lifestyles in European cities; this will prompt not only big challenges, but also great opportunities for smart and creative initiatives in future cities, through which Europe can become a global pioneer.

The Liveable City. The final image addresses the view that cities are not only energy consumers (and hence environmental polluters), but may – through smart environmental and energy initiatives (e.g., recycling, waste recuperation) – act as engines for ecologically-benign strategies, so that cities may become climate-neutral agents in a future space-economy; cities in Europe are then attractive places to live and work.

The Connected City. The image of a connected city refers to the fact that in an interlinked (from local to global) world, cities can no longer be economic islands in themselves (‘no fortresses’), but have to seek their development opportunities in the development of advanced transportation infrastructures, smart logistic systems and accessible communication systems through which cities become nodes or hubs in polycentric networks (including knowledge and innovation networks).

4. Conclusions and further steps

This paper has illustrated an on-going study in the field of Smart cities’ evaluation. The analysis started from a revised notion of triple helix considering that civil society usually plays a prominent role toward the realization of sustainable development in cities (Etzkowitz and Zhou, 2006).

In order to assess the connections between Smart city development and this institutionalization of the Triple Helix, an ANP model has been developed. This interrelated model is used for investigating the relations between smart cities components, actors and visions, or strategies to which the smart cities are moving to.

The results show that *the Entrepreneurial City* is the policy vision with higher priorities in all the sectors considered in the model, i.e. Universities, Government, Industry and Civil Society. Some relevant urban planning and policy implications of this vision are:

- A high degree of entrepreneurial activities and a constant flow of new firm creation is a prerequisite for finding a new role within the new global economic landscape. Innovation and creativeness are thus the necessary ingredients for entrepreneurial cities in Europe.
- Special emphasis has to be given to new architectures, building technologies, intra-urban mobility solutions, public space management, e.g. for lighting or citizen information management, integrated urban energy planning and management and ICT-based solutions that offer various opportunities for new urban design and management.
- New requirements for efficient, effective and reliable infrastructures (such as energy, ICT, water, waste treatment and management etc) may occur. Since an appropriate infrastructure is essential for cities’ attractiveness for companies and people alike and therefore their economic development, emphasis has to be given to the determination of these requirements within the scope of cities as complex systems.

The next authors’ task will be to develop a rating system based on the ANP model in order to verify whether: the cities are “smart”; if not, whether they are moving in the right direction.

5. References

- Caragliu, A., Del Bo, C. and Nijkamp, P. (2009). Smart Cities in Europe, paper presented to the *Creating Smarter Cities Conference*, Edinburgh Napier University, March 2009.
- Deakin, M. (2010) SCRAN: the Smart Cities (inter) Regional Academic Network Supporting the Development of a Trans-National Comparator for the Standardisation of eGovernment Services, in Reddick, C. ed. *Comparative E-government: An Examination of E-Government Across Countries*, Springer Press, Berlin.
- Etzkowitz H. and Zhou C. (2006). Triple Helix twins: innovation and sustainability, *Science and Public Policy*, vol. 33, n°. 1, pp. 77–83
- Ezkowitz, H. (2008). *The Triple Helix: University, industry and Government*, Routledge, London.
- Fusco Girard, L., Lombardi, P., Nijkamp, P. (2009). Creative Urban Design and Development (special issue), *International Journal of Services Technology and Management: Vol. 13, No. 2/3/3*, pp.111-115.
- Gabe, T.M. (2006). Growth of Creative Occupations in U.S. Metropolitan Areas, *Growth and Change*, vol. 37, no. 3, pp. 396-415.

- Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanovic, N., Meijers, E. (2007). Smart cities – Ranking of European medium-sized cities, Centre of Regional Science, Vienna UT, October 2007. Retrieved November 25, 2010, from <http://www.smart-cities.eu>
- Hollands, R. (2008). "Will the real smart city please stand up?", *City*, 12 (3), 303-320.
- Nijkamp, P., Kourtik, K. (2011). Joint Programming Initiative (JPI) on Urban Europe. Global challenges and local responses in the urban century, *a scoping document*, 27, VU University Amsterdam, 2011.
- Lombardi P., Giordano S., Caragliu A., Del Bo C., Deakin M., Nijkamp P., Kourtit K. and Farouh H. (2011). An Advanced Triple-Helix Network Model for Smart Cities Performance, In Y. Ozge (Ed.), *Green and Ecological Technologies for Urban Planning: Creating Smart Cities*, IGI Global (in press).
- Lombardi P., Cooper I, Paskaleva K & Deakin M. (2009). The Challenge of Designing User-Centric e-Services: European Dimensions. In C.Reddick (Ed), *Strategies for Local E-Government Adoption and Implementation: Comparative Studies*. Idea Group Publishing, Hershey.
- Nijkamp, P. and Kourtik, K. (2011). Joint Programming Initiative (JPI) on Urban Europe. Global challenges and local responses in the urban century. *A scoping document*, 27, VU University Amsterdam, 2011.
- Quality of Life in Twelve of New Zealand's Cities, Report 2007. Retrieved November 25, 2010, from <http://www.qualityoflifeproject.govt.nz>
- Saaty, T., (2001), *The Analytic Network Process*, RWS Publications, 4922 Pittsburgh, PA 15213.
- Saaty, T.L. (2005). *Theory and Applications of the Analytic Network Process*, RWS Publications, Pittsburgh.
- Van Soom, E. (2009). Measuring levels of supply and demand for e-services and e-government: a toolkit for cities. *Smart Cities Research Brief*, N. 3, Retrieved February 25, 2009, from <http://www.smartcities.info/research-briefs>.

A Service Oriented Solution for Interoperable Networks in Smart Public Spaces¹

Andrea Acquaviva, Chiara Aghemo, Laura Blaso, Daniele Dalmasso,
Enrico Macii, Giovanni Fracastoro, Anna Osello,
Edoardo Patti, Anna Pellegrino, Paolo Piumatti

Politecnico di Torino

Corso Duca degli Abruzzi 24, 10129 Torino

andrea.acquaviva@polito.it, laura.blaso@polito.it, daniele.dalmasso@polito.it,
enrico.macii@polito.it, giovanni.fracastoro@polito.it, anna.osello@polito.it,
edodardo.patti@polito.it, anna.pellegrino@polito.it, paolo.piumatti@polito.it

In this work we present the design of a service-oriented software infrastructure for monitoring and managing energy consumption in public buildings and spaces. We focus on the development of a software layer enabling the interoperability between heterogeneous nodes aimed monitoring energy and environmental parameters to ensure the best possible comfort conditions with the most efficient use of energy. In this paper we present a case study where the proposed infrastructure has been deployed, in three different environment: i) A new building; ii) An existing modern building; iii) A historical building. We describe their different requirements and we discuss how the wireless sensor network monitoring infrastructure can be exploited to develop energy control policies. Finally we present some preliminary monitoring results.

1.Introduction

One of the major challenges in today's economy concerns the reduction in energy usage and CO2 footprint in existing public buildings and spaces without significant construction works, by an intelligent ICT-based service monitoring and managing the energy consumption. Special attention is paid to historical buildings, which are typically less energy efficient and impose tight deployment constraints to avoid damage by extensive retrofitting.

In this context, key challenges regards the design and development of a monitoring and control infrastructure to provide control of appliances to effortlessly optimize energy efficiency usage without compromising comfort or convenience and offering decision makers strategies and tools needed to plan energy saving measures.

¹ This paper has been produced as part of the SEEMPubS project, which is a Seventh Framework Programme (FP7)

In existing modern buildings, a Building Management System (BMS) is often present to control Heating, Ventilation and Air Conditioning (HVAC) and lighting. Typically this guarantees a coarse grain control. However, the advance of sensing technologies enables a finer grain, more pervasive monitoring of energy consumption and environmental parameters such as temperature, relative humidity, luminance, to take advantage of natural sources (i.e. daylight and solar energy) to ensure the best possible comfort conditions with the most efficient use of energy.

Wireless Sensor Networks (WSNs) represent a key solution in this field. However, to enable a widespread usage of this technology and to make possible a complete and effective control in this scenario, it is mandatory to enable the interoperability between heterogeneous networks and devices, characterized by various types of hardware components and interconnection technologies. On the user-side, in order to ease the development of energy control policies, a smart, flexible and easy to extend interface to operators must be designed. Furthermore, such an interface can be exploited to expose the energy utilization information collected by monitoring devices to final users in order to promote green-behaviours.

To achieve interoperability, middleware solutions have been recently proposed [Fummi et al., 2010], [Bonino and Corno, 2008a], [Jahn et al., 2009b], enabling a uniform and hardware-abstracted view of the different components and networks, providing to the developers a set of APIs or methods to be used for control and configuration of the underlying networks.

In this work we present the design and implementation of a service-oriented infrastructure for public space monitoring and its possible exploitation to create services and applications across heterogeneous devices to develop an energy-aware platform. In its core, the software infrastructure is composed of a middleware, a database and a network interface layer. It is designed to enable interoperability between heterogeneous sensor and actuator networks. Moreover, being web service-based, it implements a hardware independent and flexible user interface. We discuss the specific characteristics and requirements of the case study of public spaces we explored, consisting of both modern and historical buildings, and we outline how the proposed infrastructure can help implementing energy control and awareness policies. Finally we show preliminary monitoring results provided by some of the deployed sensors.

The rest of the paper is organized as follows. In Section 2 we review some background literature. In Section 3 we present the proposed software infrastructure. The considered case study is described in Section 4. Preliminary monitoring results and considerations are reported in Section 5.

2. Background

Solutions for smart building management and context-awareness are increasingly widespread. These projects mainly focus on providing energy solutions addressing specific issues in a vertical way.

The AIM [Capone et al., 2009] project aims at the development of technologies for the monitoring and management of energy consumption in

households in real time. Buildings habitants should be able to manage their home networks. For this purpose, selected system functionalities are exposed by means of services to external networks via a gateway. Policy management, device detection and configuration of the system are examples of such functionalities.

The EnergyPulse [Jahn et al, 2009a] project was developed on the basis of LinkSmart (earlier called Hydra) middleware [Jahn et al, 2009b], [Lardies et al, 2009]. LinkSmart is designed to support the development of networks of embedded systems, simplifying the interconnection of heterogeneous devices that communicate using different protocols. In the EnergyPulse project, through the LinkSmart middleware is possible to build an intelligent network that seamlessly integrates heterogeneous embedded devices in the same network.

Concerning alternative solutions for interoperability, in the domain of Home Automation, Corno et al. have developed a Dog home gateway [Bonino and Corno, 2008a], for the interoperation between various domotic devices. The ontology is designed to explicit representation of commands, accepted by domotic devices, and notifications, generated by suitable control devices. The process of abstraction is done by the definition of associations between controller and controlled devices. This information is the basis for designing device interoperation solutions.

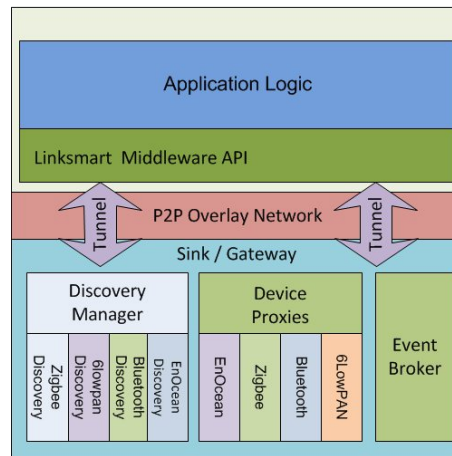


Fig. 1 - LinkSmart Architecture.

In addition to research projects, the industry provides a broad range of solutions for energy monitoring, and power measuring, such as Google [Powermeter], Greenbox [Greenbox] or The PowerTab [PowerTab]. Typically, industrial solutions are proprietary and depend on the peculiar smart meters and sensors used. Some of these solutions are even able to monitor energy consumption at a single device level.

All these smart applications are able to manage only the aspect of energy consumption, missing a wider vision about other parameters that can affect the efficiency of the building, i.e. air quality, temperature, etc.

3. Software Infrastructure

In this section we describe the web service-oriented software infrastructure designed to manage heterogeneous and commercial wireless sensor nodes belonging to different WSNs. We adopted the LinkSmart middleware to enable hardware independent access to the nodes. Hence, the software should automatically associate virtual end-nodes to real physical devices making them available to higher layers.

The Figure 2 shows the scheme of the proposed infrastructure. The heterogeneous WSNs communicate directly with the software running on a PC-Gateway (GW). All the network information is independent from the hardware, network topology and communication protocols and it is directly accessed through a dedicated interface. This interface is the same for every WSN. On the gateway, data are parsed and stored in an integrated database (DB), which is the second layer of the infrastructure.

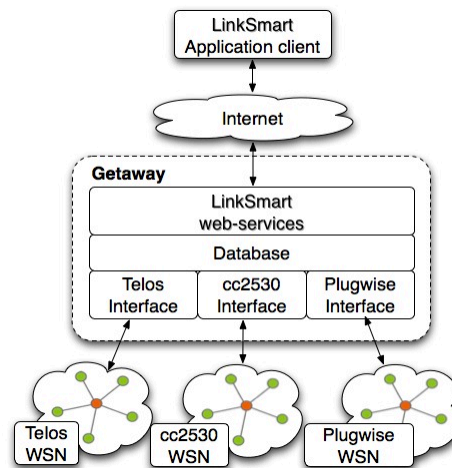


Fig. 2 - Software infrastructure scheme to handle heterogeneous wireless sensor networks.

The local database provides flexibility and reliability to the whole infrastructure with respect to possible backbone network congestions or failures, since data are locally stored. As Database Management System (DBMS) we adopted SQLite [SQLite]. Moreover, the stored data can be easily exported to others DBMS or formats.

The web-services layer has been implemented using the LinkSmart middleware. It interfaces the WSNs to the web, facilitating remote management and control. Furthermore, it exports to the application-client layer all the environmental data stored in the database, previously collected by the wireless sensor nodes. Thus information is available at application-client layer to be post-processed.

Particular emphasis was given to the possibility to reconfigure each node, changing, for instance, some parameters about power management. Thus, the

end-user sends the new configuration via web-services to the GW and stores it in the DB. Then, the new settings will be automatically sent to the receiver mote, when it will wake up from the sleeping period, through the specific WSN software Interface. The configurable parameters change depending on the hardware and the Operating System running on the end-node. However, it is worth noting that the software infrastructure allows the user to choose the appropriate setting without knowing the real physical hardware related to the virtual device.

In a nutshell, the proposed web-based infrastructure is a software that makes transparent to the end-user the underlying WSNs, abstracting all the information about the hardware, the protocol stack and embedded operating system. Furthermore, by the use of web-services, it makes the interoperability with third-party software easy.

4. Case Study

4.1. Devices and Network Topology

In this work we considered three different networks. Two of them are based on commercial wireless sensor devices and one on custom motes. To enable power efficiency, in all cases the firmware configures the motes to sleep most of the time, quickly wake-up upon an event, perform the required data processing and return to sleep after data transmission.

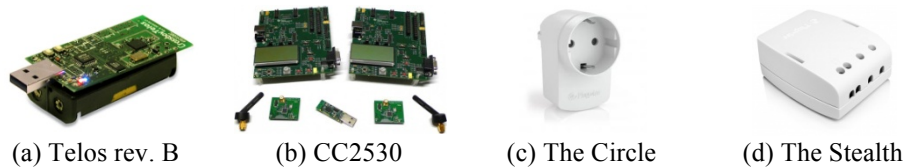


Fig. 3 – The Wireless Sensor Nodes.

In this work, three different networks made of three commercial wireless sensor devices have been considered as case study, namely: i) Crossbow Telos rev. B mote [TelosB] (Figure 3a), ii) customized mote built on Texas Instruments' CC2530 system on chip [CC2530] (shown in Figure 3b) with two additional sensors: a Texas Instruments TMP125 temperature sensor and an Avago ADPS-9300 photodiode sensor and iii) Smart-Plug provided by Plugwise Company [Plugwise] that provides details of energy consumption per appliance connected to the mains and remotely switch connected devices on or off via software interface. Specifically the adopted devices are the Plugwise Circle (Figure 3c) and Plugwise Stealth (Figure 3d).

The TelosB runs TinyOS [TinyOS], which implements a 802.15.4 protocol stack. On the other side, CC2530 and Plugwise motes run a ZigBee 2007 protocol stack.

In this case study, it has been implemented a star network for each commercial wireless sensor node, consisting of a Coordinator (CO), one for every device network, always connected to a PC-Getaway (GW) via USB, and

several End-Nodes (EN). The information flow is always from the EN to the CO and vice versa. Hence the ENs monitor the environment at regular time intervals, then they send the collected information to the Coordinator. The CO forwards all the data coming from the WSN to the PC-Getaway. Finally, the GW stores the information collected in a customized database and interfaces all the WSNs to the web through LinkSmart middleware web-services.

Considering the opportunity to configure remotely, these parameters are an example of the type of settings that can be made:

- *sampling period*: the time interval between two consecutive sampling and transmission of environmental information;
- *radio wake up interval*: the time interval between two consecutive wakes of the radio module to check for possible presence of packets to be received;
- *active sensors*: those sensors that must be activated depending on the environmental information to be collected. Concerning CC2530, the user can choose between Temperature and Light; Humidity and Battery level are added for TelosB motes.

On CC2530, sampling period and radio wake up interval are coincident. These setting will be exported as a virtual End-Node form the GW to the upper layers transparently trough web-services. Therefore the user can choose the right settings without knowing the real physical hardware related to the virtual device.

4.2. Environment and Building Description

The Politecnico di Torino campus buildings, which were selected as case study in this research, are located in three different sites in the city. The short description below reveals the main features of each building and the reason of their representativeness as examples of a wide range of similar existing buildings in Europe.

- The historical campus building (The Valentino Castle) dates back to the beginning of the 16th century; two grand staircases lead to the first floor, where thirteen rooms, with their rich stuccoes and commemorative allegorical fresco paintings, are the evidence of the ancient splendour of the 17th century.
- The old campus site is the main campus for the Engineering Faculties. It was inaugurated in November 1958 and its aesthetics is based on the rigor of its lines. All the buildings on this site were built in relation with their use: the heavy laboratories, with their factory-like function, have metal or concrete structures, while the other ones (classrooms and offices) are concrete only.
- The modern campus site comes from a complex refurbishment of a former industrial area. It includes the area's most interesting industrial buildings that were regained and new buildings that were built demolishing other less appealing structures.

Obviously, each building requires a specific solution for the installation of new sensors and the control of existing HVAC and lighting systems. The

modern campus is equipped with a basic building management system (Desigo by Siemens) and the energy saving can be improved by simply optimizing the present system regulation. In the main campus it is feasible to install new sensors in a wired system. In the historical building, the historical value of paintings and stuccoes does not allow an easy installation of sensors and each room requires a specific solution and only wireless sensors can be considered.

The diagram in Figure 4 shows the connection among building construction period, existing technologies, sensors installation cost and sensors installation difficulty. An historical building normally is characterized by little existing technologies, and high construction work costs closely related to the difficulty of new technologies installation in order to preserve painting, stuccos and wood/marble floors. Instead, a new building normally is originally characterized by new technologies and it required only integrations in an easy way using false ceilings and floating floors. In the middle there are existing buildings where new technologies have to be significantly improved, and cost and difficulty are closely related to civil work required.

Within each campus buildings, some representative rooms were selected to implement the BMS with the new sensors network infrastructure. The rooms were chosen as significant case studies on the basis of the following criteria: representativeness with respect to the Campus and other Public buildings; energy saving potential, supposed according to their architectural, services and occupancy characteristics. Both private and public spaces were selected, such as classroom, student offices, single private offices and open plan offices.

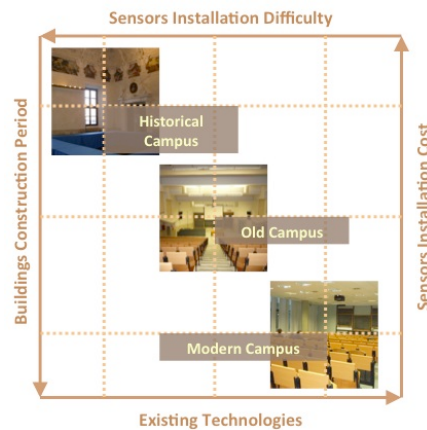


Fig. 4 –Diagram of all elements considered as case study.

4.3. Control and monitoring solutions for energy saving

To achieve the goal of raising the buildings energy performance, new solutions for controlling lighting plants, electrical appliances and air conditioning services and for monitoring environmental conditions and energy consumptions have been defined.

As an example, for a small office with low absence probability and high

daylight availability, the proposed lighting control strategy includes both daylight harvesting and occupancy control. The lights will be switched on and off when user presence or absence is detected by an indoor occupancy sensor and will be dimmed up or down to integrate daylighting in order to achieve the required lighting conditions (defined according to lighting standards or users preferences). A photosensor that measure the indoor light quantity, such as TelosB or similar, will be used to implement the described control. For heating, the principle is to switch off the fancoil's fan, hence reducing thermal energy consumption, when, during working hour, the user absence is detected by the occupancy sensor or when the electric (+ lighting) power, measured by the energy monitoring devices, is sufficient to cover the heating need. Furthermore the measurement of the internal air temperature, for instance by means of TelosB nodes, will allow activating or deactivating the heating, by comparing this temperature to the set point value previously defined to respond to comfort requirements.

In the following diagrams an example of control logics for heating and lighting are presented.

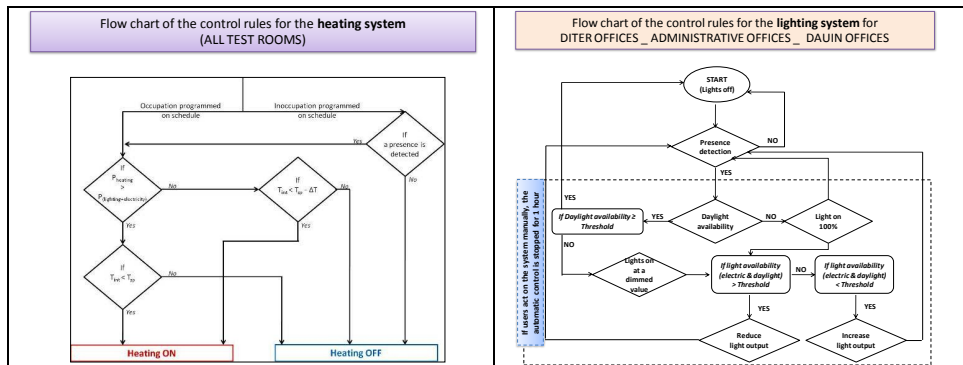


Fig. 5 - Diagrams representing an example of the control logics proposed to manage the heating and lighting plants.

To implement the described control rules in the different buildings of the Politecnico, several sensors, actuators and controllers are needed and the designed architecture has to take into account the constraints and limits that arise from the characteristics of the buildings. The new system could include both wired and wireless sensors and actuators as described in the following schemes.

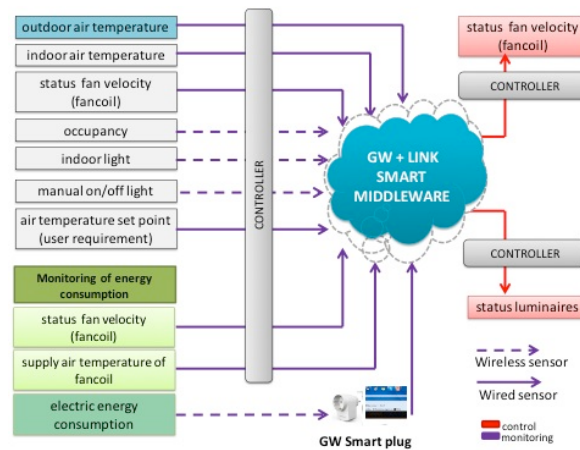


Fig. 6 - Wired and wireless sensors and actuator scheme designed to control lighting and heating in a room of the old campus site

5. Preliminary monitoring results

The overall control and monitoring system is currently under deployment in the Politecnico campus. Currently, a small network of TelosB motes have been installed in some of the selected rooms, to test their reliability in monitoring the microclimatic conditions. The End-Nodes were placed in the amphitheatre Classroom 3 of Politecnico di Torino at a height of 4 meters from the seats plane. They have been programmed to send a data every 15 minutes and to enter into low power mode during periods of inactivity.

In the following graph (Figure 7) the indoor air temperature and the relative humidity measured in a large classroom during the period from the 14th of September to the 7th of October are shown. In this period the HVAC service was deactivated, as the cooling season was already ended and the heating season is going to start the 15th of October. Even so the data demonstrate the effectiveness of the wireless TelosB sensors and network in recording the microclimatic conditions as can be seen from the comparison to the data measured outside during the same period (figure 7). It should be also noted that the classroom was not used until the 3rd of October. With the beginning of lessons the presence of people contributed in rising the indoor temperature and relative humidity with respect to the trend shown by outdoor air temperature data.

6. Conclusions

In this paper we discussed the main challenges and requirements concerning the deployment of sensing and monitoring infrastructure inside public buildings.

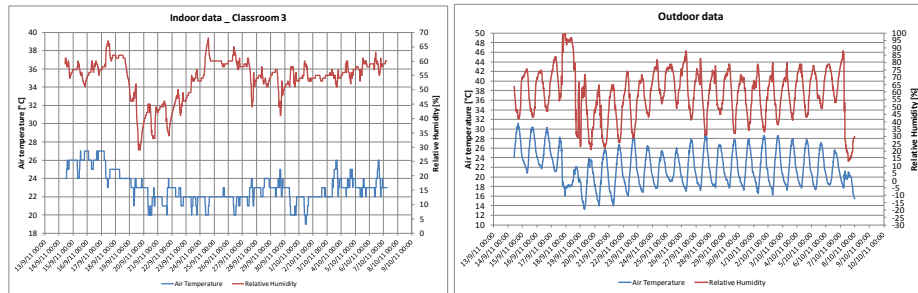


Fig. 7 - Microclimatic conditions measured inside Classroom 3 and outside the building.

We presented a software infrastructure aimed at ensuring the interoperability between heterogeneous networks. These networks have been deployed on various types of buildings, with the final purpose of enabling the development of energy management solutions. This study made evident that a multidisciplinary approach is needed to efficiently exploit hardware and software technologies to create new smart building and to bring intelligence to historical ones.

Bibliografia

[Bonino and Corno, 2008a] Bonino, D., Corno, F., The DOG Gateway: Enabling Ontology-based Intelligent Domotic Environments. Proc of IEEE Transactions on Consumer Electronics 54(4), 2008.

[Capone et al, 2009] A. Capone, M. Barros, H. Hrasnica, and S. Tompro, A NewArchitecture for Reduction of Energy Consumption of Home Appliances. Proc. of TOWARDS ENVIRONMENT, European conference of the Czech Presidency of the Council of the EU, 2009.

[Fummi et al, 2010] Fummi F., Perbellini G., Quaglia D.: Trenti R., Exploration of network alternatives for middleware-centric embedded system design. Proc. of 13th Euromicro Conference on Digital System Design , Lille, France 2010, 2010.

[Greenbox] Greenbox project web page, <http://getgreenbox.com/>. Accessed on September 22th, 2011.

[Jahn et al, 2009a] Jahn, M., Jentsch, M., Prause, C.R., Pramudianto, F., Al-Akkad, A., Reiners, R., The Energy Aware Smart Home. Proc of 5th International Conference on Future Information Technology (FutureTech), 2010.

[Jahn et al, 2009b] M. Jahn, F. Pramudianto, and A. Al-Akkad, Hydra Middleware for Developing Per- vasive Systems: a Case Study in the eHealth Domain. Proc. of 1st International Work- shop on Distributed Computing in Ambient Environments (DiComAe 2009). 2009: Paderborn, Germany.

[Lardies et al, 2009] Lardies, F.M., et al. Deploying Pervasive Web Services over a P2P Overlay. Proc of 18th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises. 2009, IEEE Computer Society.

[Plugwise] Plugwise web page <http://www.plugwise.com>. Accessed on September 22th, 2011.

[Powermeter] Google Powermeter web page, <http://www.google.org/powermeter/>. Accessed on September 22th, 2011.

[PowerTab] The PowerTab project web page, <http://www.energy-aware.com/our-products/ ihd/>. Accessed on September 22th, 2011.

[SQLite] SQLite web page <http://www.sqlite.org/>. Accessed on September 22th, 2011.

[TelosB] Crossbow Telos rev B datasheet http://bullseye.xbow.com:81/Products/Product_pdf_files/Wireless_pdf/TelosB_Datasheet.pdf. Accessed on September 22th, 2011.

[TinyOS] TinyOS web page <http://www.tinyos.net/>. Accessed on September 22th, 2011.

A Terascale, Cost-Effective Open Solution for Academic Computing: Early Experience of the DAUIN HPC Initiative

Federico Della Croce, Elio Piccolo, Nicolò Nepote
Dipartimento di Automatica e Informatica – Politecnico di Torino
C.so Duca Degli Abruzzi 24, 10129 Torino (TO)
[federico.dellacroce,elio.piccolo,nicolo.nepote]@polito.it

Supercomputing plays an increasingly important role in academic research, due to its many interdisciplinary applications. This creates a growing demand for computational resources that the current national and European HPC infrastructure cannot always satisfy. Access to those systems is in fact controlled by calls that require months of planning. However, the creation of smaller HPC facilities inside universities or even single departments, presents problems such as difficulty to find system administrators with appropriate technical skills and high costs of HPC systems. This paper describes the experience of the DAUIN HPC Initiative at the Politecnico di Torino, which included the creation of a cluster reaching 1.0 TFLOPS with a very low budget and using mainly open-source software. The ease of access to computing resources, shared among the research groups of the department, made it possible to achieve encouraging results in less than six months.

1. Introduction

1.1 Context and Challenges

Even at a time of economic stagnation and global financial crisis, investments in supercomputers [Kindratenko and Trancoso, 2011] do not decrease, as evidenced by the evolution of the *Top500* list [www.top500.org]. The European Union is moving in the same direction with the *PRACE project* [www.prace-project.eu], which effectively unifies the network of existing High Performance Computing infrastructures in member countries. Two petascale supercomputers are already in use in France and Germany.

Today HPC [Dongarra, 2006] is seen as an essential supporting service to any research project involving the solution of complex problems for which sequential algorithms are not appropriate. Non-parallel algorithms are unable to fully exploit the multi-core architecture of modern CPUs, now scaling towards 16 cores per chip. Over the past three decades, software parallel transformation has always been a major field of interest [Dongarra, 2011].

Large computing centers, however, usually attract large research projects. These projects typically need grants for more than 10^6 core-hours. More often the funding of such research projects is subject to the grant from a computing center. Access to these centers is regulated by calls that are held every 2 or 3 months, requiring careful planning for submission of requests and a lot of documentation. The search for more agile and efficient systems for granting and accounting core-hours is still open [Luque et al, 2011].

This article focuses on the needs of another type of research project requiring less core-hours (between 10^3 and 10^4) to be granted within a very short time. These projects are unlikely to be served by large HPC networks like PRACE or similar. From now on we will refer to these projects as "small projects", where "small" is meant in terms of lower computational requirements.

The solution often used by the management of small projects includes the purchase of a preconfigured entry-level HPC system (this definition is border-line because they often don't scale up to the TFLOPS) sized to meet the computational needs of the project, with some drawbacks:

- the system is non-shared: it is used exclusively for the project for which it has been purchased (or projects of the same research group) and then left unused for an undefined period of time,
- the system can be fully exploited only by investing time in studying the management software or by requesting assistance from the manufacturer,
- those systems typically have a proprietary Management/OS layer.

1.2 Our Proposal

The solution we present in this paper includes the sharing of computing resources at departmental (or campus) level, by means of the creation of a computing center financed by the department itself and by the research groups. This kind of approach encounters some open issues, such as high costs for the hardware, difficulties in finding the right staff to manage the computer center, a general distrust of shared resources and the need for qualified engineering support to research projects that our solution addresses, as depicted in the following sections.

The paper is organized as follows: Section 2 exposes a case study whose subject is the HPC initiative of the Department of Control and Computer Engineering at the Politecnico di Torino: the focus is on the technical side, describing all the software and hardware being used, analyzing the cost/benefit ratio in terms of money spent and energy consumption. Finally, Section 3 draws our conclusions, discussing a possible generalization of this case study.

2 Case Study: The DAUIN HPC Initiative

The *DAUIN HPC Initiative* started in 2008 with the first experiments and feasibility studies using cheap hardware and free or open source software. The result was a specification document in which a roadmap was established to set up a Tier-II cluster (a low level computing center facility, following the European PRACE classification) called CASPER-3 (*Cluster Appliance for Scientific Parallel Execution and Rendering*). This cluster should be able to reach 1.0 TFLOPS of computational power, with a very reduced budget. The initiative aims to provide computational resources and technical support to both academic research and student labs. The main focus is on research groups belonging to the Department of Control and Computer Engineering, but the initiative is open for collaborations with other departments and research centers, including institutions outside the Politecnico di Torino. Since 2010 the Labinf [www.labinf.polito.it] staff has been in charge of this project, and the operational phase began.

2.1 Requirements, Funding and Objectives

The research groups were directly involved in requirements definition through a series of individual meetings in which they also suggested a financial contribution. The resulting list of general requirements follows:

1. single core *clock frequency* ≥ 3.0 GHz,
2. for each computational node of the cluster: $4 \leq N_{CORES} \leq 8$, $RAM/N_{CORES} \geq 2GB$,
3. memory with low latency and high clock,
4. *InfiniBand*® interconnection,
5. maximization of the computational power (FLOPS) within the budget constraint,
6. granular management of execution queues (with different priorities depending on projects and/or users),
7. use of open source software for the Management/OS layer and possibly for the application layer too,
8. creation of a fast ($\leq 48h$) request/grant mechanism,
9. availability of an engineering service for any hosted project.

For all the previously listed requirements there was a budget of € 13,500 to be funded by both the DAUIN and the research groups. In addition to meeting the needs of academic computing, as described before, the DAUIN HPC Initiative addresses some additional targets, including:

1. reaching the terascale region with CASPER-3 within 2011,
2. making CASPER-3 to be multipurpose, covering as many HPC applications as possible,
3. using open source software,
4. making the DAUIN emerge as a source of HPC-related knowledge within the Politecnico di Torino,
5. showing the research group managers the benefits of computing resources sharing,
6. encouraging researchers to work in teams with the computing center staff,
7. enhancing the technical skills of the department staff.

2.2 Hardware Implementation

Item	Description
CPU Type	Intel Nehalem (core i7)
CPU Specs	3.46 GHz with 8 MB cache
interconnection	InfiniBand DDR 20 Gb/s
network	Gigabit Ethernet 1 Gb/s
N_{CORES} per node	8 (4 physical plus HT)
total N_{CORES}	40 (20 physical plus HT)
nodes	5 + frontend
memory	120 GB (5x24) DDR3 1066
storage	8 TB RAID 5

Tab.1 – CASPER-3 Original hardware configuration.

CASPER-3 is a MIMD distributed shared memory heterogeneous Linux cluster with InfiniBand DDR interconnection, consisting of "bare metal" (not branded) computational nodes whose configuration has been designed by the staff of the initiative. This allowed to maximize the number of nodes in accordance with the budget constraint, to have sufficient control over the components choice in order to meet the requirements, to ensure full hardware compatibility with the designed OS (see section 2.3) and to get

everything directly from common hardware retailers, without intermediaries. In its original configuration, the cluster was composed as shown in Tab.1.

The respect of the core frequency requirement has led to the temporary exclusion of server-oriented CPUs, because of their high price-per-GHz. The Intel *Hyper Threading*® technology (referred as “HT” in Table 1) is to be seen as an added value: it can be useful for certain types of applications, such as rendering, but not in terms of CPU-intensive activities, like the Linpack benchmark [www.netlib.org].

The ratio RAM/N_{CORES} is equal to 3GB, slightly above the minimum requirement. The cost for the interconnection layer has a strong impact on the final cost, but InfiniBand has a great impact on supercomputer performance and scalability over MPI [Shipman et al, 2006]. The choice fell on *Mellanox*, the market leader for InfiniBand, that provides an updated version of the OFED drivers that is fully compatible with Linux.

A master node (usually referred as the “frontend”) with reduced performance, was set up separately for the sole purpose to host the scheduler, to act as NAS, as a network server and as an interface to the users.

The six computational nodes composing CASPER 3, the network devices, and everything needed for wiring are housed in a single rack hosted inside a preexisting DAUIN server room, occupying a total of 16U (see Fig.1).

2.3 Management and Application Layers

The choice for the OS fell on the well known ROCKS Clusters Linux distribution [www.rockcluster.org], formerly known as NPACI-ROCKS, that equips some of the *Top500* supercomputers. ROCKS has registered excellent performances on non-server-oriented CPUs [Sacerdoti et al, 2004] and integrates very well with InfiniBand [Sajjipanon and Ngamsuriyaroj, 2008]. Among ROCKS features we include:

- it is well maintained and based on the enterprise class CentOS distribution,
- it integrates both OS and Management layers,
- the Management layer is strictly dependent from the OS version,
- it is very easy to install and setup,
- it has an automatic node discovery utility,
- it comes with a lot of on board preconfigured software, including complete parallel environments, schedulers and monitoring systems,
- customization can be done only by adding compatible RPM packages or by manually building the software,
- it has lots of add-ons available from many computing centers around the world,
- there is no way to manually upgrade the distribution (until a new version is released).

The scheduling functionality is provided by the Oracle Grid Engine (formerly known and referred here as SGE) of which a fully GPL-compatible version is also available. SGE is almost a standard among the *Top500* systems and represents the state of the art in supercomputers job scheduling [Yanyong et al, 2003].

CASPER 3 has three default queues available under SGE:

1. **all.q** max 40 CPU (8 x 5 nodes) - hyper threading ON - max job duration = 96h - low priority - mutual exclusion with noht.q - subordinated to urgent.q: it is a long-run default queue for any job.
2. **noht.q** max 20 CPU (4 x 5 nodes) - hyper threading OFF - max job duration = 96h - low priority - mutual exclusion with all.q - subordinated to urgent.q: used in case HT may degrade the performance.

3. **urgent.q** max 6 CPU (2 x 3 nodes) - hyper threading OFF -max job duration = 1h - high priority - preemptive towards both all.q and noht.q: used to test feasibility, performance, interconnection, MPI operation, etc. before switching to the long-run queues.

All the sub-queues corresponding to the individual nodes are enabled with the *exclusive* function that allows users to increase the RAM/N_{CORES} ratio when sending a job. This behavior leaves some unused cores within the node. Other queues can be created in the future, with special permissions based on research groups or even individual users. By using ROCKS for the OS/Management layer, the initial set of libraries and scientific HPC-oriented software installed on CASPER-3 is very extended. Other software are commonly added when requested by research groups. The licenses of those additional software may not meet our will to use mostly open-source programs.

2.4 Engineering Support, Grant Mechanism and Cluster Monitoring

Technical support is provided by the DAUIN staff, during the normal office hours. Beside phone calls and emails, users can contact the supporting team through remote chats and remote calls. Engineering services includes:

- supporting the users during the start-up phase of the project,
- understanding the purpose of the research project,
- maximizing the benefits of HPC on the project,
- explaining the usage of the scheduler queues to the users,
- installing, integrating, configuring and testing new software,
- supporting the users for any cluster related issues.

The grant mechanism for core-hours is very simple:

1. The applicant verifies his requisites on the DAUIN HPC Initiative website [dauin-hpc.polito.it] and than submits a usage request through a pre-shaped email,
2. the usage request is reviewed by the staff for a feasibility approval,
3. the usage request is reviewed by the HPC initiative's overall supervisor for final approval,
4. the granting or denying of system usage is confirmed via email within 24 hours.

Finally, the users can monitor the evolution of their submitted jobs by two means:

- through the SGE command line on the frontend node of CASPER 3,
- through the Ganglia web interface on the DAUIN HPC Initiative website. Ganglia is a widely used, open-source scalable distributed monitoring system for high-performance computing systems, in bundle with the ROCKS cluster distribution.

2.5 Benchmarks

Linpack [www.netlib.org] is a software library developed to perform linear algebra operations. It was written in Fortran by Jack Dongarra, Jim Bunch, Cleve Moler, Stewart and Gilbert, and was developed specifically for the supercomputers of the '70s and early '80s. Linpack uses the BLAS2 library to perform vector and matrix operations.

The Linpack benchmarks are used to measure computer performance in floating point operations. They measure how quickly is solved a dense system of linear equations $Ax=b$ of $N \times N$ size. The solution requires about $2/3 \cdot N^3 + 2 \cdot N^2$ floating point operations. The test result is directly given in MFLOPS.

To analyze massively parallel supercomputers, High Performance Linpack, a portable version of the Linpack benchmark, is typically used. HPL is the test used to generate the *Top500* list [www.top500.org]. The test is carried out for different values of N in order to achieve the highest possible R_{max} . R_{max} is a measure of the highest speed of the system while running HPL, while R_{peak} represents the theoretical peak performance.

Both figures are given in MFLOPS. Linpack has been criticized [Anderson, 2011] for not sufficiently stressing the interconnection layer, but actually Linpack instead focuses on the processing unit and memory cache. The network is highly used only when dramatically increasing the system memory, because the maximum size of the problem also increases.

However, the memory usually available in modern supercomputers is around only 1.0 GB per core only. In the case of CASPER-3, having a memory of 3 GB per core, the same test repeated with and without InfiniBand enabled gave slightly different results (20% better with InfiniBand enabled).

HPL was compiled and optimized for the CASPER-3 Intel Nehalem architecture using different BLAS implementations. GotoBLAS gave the best experimental results.

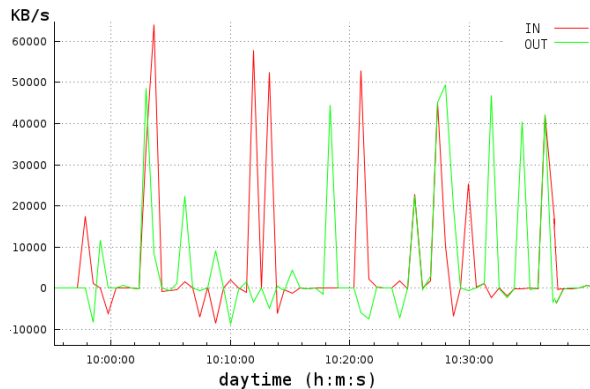


Fig.1 – A typical CASPER-3 node bandwidth usage in KB/s during a complete HPL execution over InfiniBand.

In Fig.1 the typical network usage of a single node during the test execution is visible. The peak is near 65 MB/s, that represents only 2.6% of the total available bandwidth. The total aggregate bandwidth usage is around 250 MB/s, which corresponds to 10% of the capability. The best Linpack values achieved with such HPL configuration were, in GFLOPS: $R_{peak} = 615$ and $R_{max} = 580$.

The calculation of the interesting *Green500* index [Feng and Cameron, 2007] is not feasible in our environment because we cannot modify the temperature and humidity. However we can calculate a similar index with sufficient approximation. The *Green500* index is the value of the ratio R_{max}/R_{max_Power} expressed in MFLOPS/W. We can compute the value of R_{max_Power} as the total theoretical power consumption of CASPER-3 assuming all the components to be at their maximum power consumption (according to each item's datasheet). The resulting *Green500* index expressed in MFLOPS/W is: $R_{max} / R_{max_Power} = 196.61$.

Such ratio is pretty good because it corresponds to the 167th position on June 2011 *Green500* list and we can compute the cost per GFLOPS in €: $budget / R_{max} = 23, 27$.

From these results we see that a doubling of CASPER-3 computational nodes within the end of 2011 will suffice in order to safely reach the TFLOPS region, respecting our time constraints.

2.6 Experience with Hosted Research Projects

A variety of research activity is supported by DAUIN HPC Initiative. This ranges from fluid mechanics to computer vision. During the past five months CASPER-3 was used to solve problems related to eleven research projects, belonging to the fields in Tab.2.

Field	Core-hrs	Core-secs	%
Bioinformatics	385	1386162	0.3
Optimization	25870	93131071	18.0
Networking	14182	30051424	9.8
Fluid Dynamics	5761	20738328	4.0
Computer Vision	11612	41803956	8.1
Virtual Heritage	4824	17366184	3.3
Teaching	8443	30394800	5.9
Unallocated	72923	262524300	50.6
Total	144000	518400000	100.0

Tab.2 – Breakdown of core-hours on CASPER-3 depending on the scientific field, until Sept. 6, 2011.

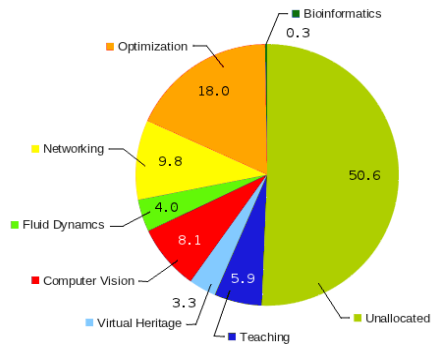


Fig.2 – Percentage breakdown of core hours on CASPER-3 depending on the field, until Sept. 6, 2011.

The total core hours breakdown on CASPER-3 is also plotted in Fig.2, where we observe a quite heterogeneous distribution. We notice also that roughly half the core hours are tagged as “unallocated”. We can partially explain this figure by experimentally observing that in the networking, computer vision and virtual heritage fields, the SGE’s exclusive function has been widely used, thus filling-up the memory while leaving 50% of the cores in idle state. Among all the hosted research projects, we can list some relevant use cases:

- A parallel implementation of the heuristic algorithm proposed in [Della Croce and Grosso, 2012] for the 0/1 multidimensional knapsack problem has been tested using the CPLEX optimizer for different values of “set parallel” and “set threads” with the same “timelimit” constraint, with interesting experimental results.
- The first application of Matlab under CASPER 3 was done in the field of pattern recognition [Bottino and Laurentini, 2011] with a study on the intrinsic dimensionality of attractive face profiles.
- For the “Re-living Las Vegas” project [Antognozzi et al, 2009], a multi-user mixed-reality edutainment environment in the virtual heritage field, CASPER 3 has been used for rendering 360° high-resolution static images. In this case a new dispatcher for distributed rendering has been developed internally by the staff to integrate Blender with Yafaray seamlessly.
- Researchers from the Department of Aerospace Engineering simulated energy spectrum power-law decay of linearized perturbed shear flows using Matlab on CASPER 3. A paper on that work is currently in review state. This collaboration with the DIASP was the first, but there are others ongoing with researchers from other departments.
- The Ph.D. course of “Parallel and Distributed Computing” at Politecnico di Torino has used the CASPER 3 facilities for the laboratory part. The main laboratory topic concerned the use of OpenMPI over InfiniBand. This is a typical example of the kind of teaching activity that is supposed to be hosted by the DAUIN HPC Initiative.

After only five months of activity, the DAUIN HPC Initiative helped in the completion of one paper that has already been published, and other papers are in review or in preparation. This is an encouraging result if we think about the actual performance of CASPER-3 and the departmental size of the computing center.

2.7 Usage Statistics

The Ganglia monitoring system can effectively collect a lot of data during all the lifetime of a cluster. We present here an interesting subset of CASPER-3 related data, in order to better explain in which way the cluster has been used during the first five months of activity. In Fig.3 the cluster workload is compared to the total available number of cores. This is directly linked to the core-hours distribution chart in Fig.2.

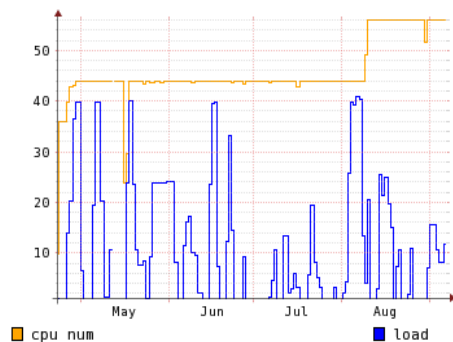


Fig.3 – CASPER-3 workload compared to available number of CPUs, until Sept. 6, 2011.

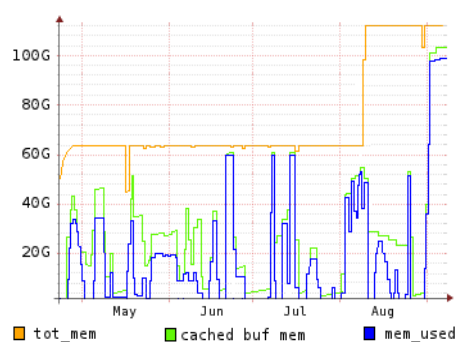


Fig.4 – CASPER-3 memory usage in GB compared to total, cached and buffered memory, until Sept. 6, 2011.

In Fig.4 the cluster memory (RAM) usage is presented. From this graph, compared with the one in Fig.3, we understand that the total memory size often represents a bottleneck. Future additions to CASPER-3 will provide more memory-per-node to address this issue. The NAS disk size on the frontend node is more than enough for the needs of the research projects hosted so far, as shown in Fig.5. Finally, in Fig.6 and Fig.7 the InfiniBand interconnect usage is shown. The five-months chart in Fig. 6 shows a peak around the first decade of June, due to several executions of the Linpack benchmark. To appreciate a more typical data collection of the bandwidth usage over a shorter period, we can look at Fig. 7.

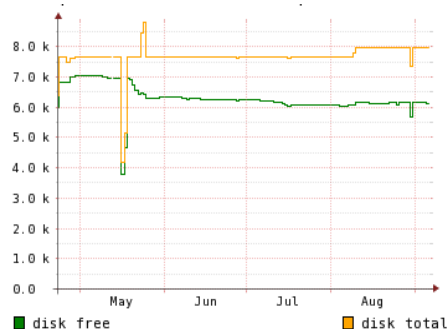


Fig.5 – CASPER-3 free NAS space in TB compared to the total NAS capacity, until Sept. 6, 2011

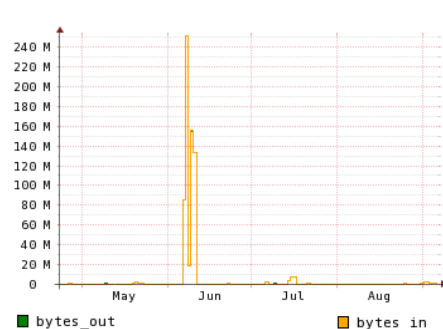


Fig.6 – CASPER-3 network throughput in MB/s (inbound and outbound), until Sept. 6, 2011.

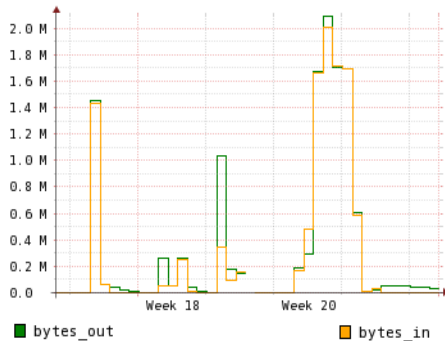


Fig.7 – A typical CASPER-3 memory usage in MB/s over a short time window (since Aug. 6, 2011 until Sept. 6, 2011)

We notice, as expected, that the interconnect layer is not an issue for our implementation, being used for less than 10% of its capability during intensive computations. The InfiniBand technology will easily support further evolution of CASPER-3 both in terms of increased number of nodes and in terms of increased size of the total memory.

2.8 Further Developments

After the early experience described above, the DAUIN HPC Initiative, in order to expand and improve CASPER-3's capabilities to better fit the needs of the current and upcoming users of the system, found a new way of self-funding based on a mechanism called "fellowship". According to this, fellow research groups volunteer the buying of new computational nodes for their temporary exclusive use, and then donate them once their projects are concluded. Following this schema, two important agreements were concluded in September 2011, that will add four new computational nodes to CASPER-3 within the end of 2011, reaching a total of 76 cores and 432 GB memory. The new nodes are designed using different requirements, following the suggestions from current and past users of the system, namely for each computational node of the cluster: $12 \leq N_{\text{CORES}} \leq 32$ and $4\text{GB} \leq \text{RAM} / N_{\text{CORES}} \leq 8\text{GB}$ and for each core of the node: $2.0\text{GHz} \leq \text{clockspeed} \leq 3.0\text{GHz}$. The first AMD based dodeca-core, 2.6 GHz, 48 GB, node has been already put in place since the first decade of August, 2011. In fact, in Fig.3 and Fig.4 a big increment of "n_cpu" and "tot_mem" can be seen since August 2011.

Furthermore, additional funds are expected from institutional funders.

3 Conclusions

This paper presents the design and implementation of CASPER-3, a cost-effective open solution for academic high performance computing, designed and adopted by the DAUIN HPC Initiative at Politecnico di Torino to address the needs of small research projects requiring computational power between 10^3 and 10^4 core-hours, in a very short time. These projects are unlikely to be served by larger computing centers. Our analysis of the early experience of this initiative shows that all the objectives were reached while guaranteeing the time and budget constraints, achieving encouraging results in terms of productivity and appreciation by the research groups, also outside the department. New collaborations are now starting, and an expansion of CASPER-3 is expected in the coming months, making it able to approach the TFLOPS region. While this paper

concentrates on a specific well defined HPC initiative, we believe that the proposed solution may apply to other similar environments within the national academic landscape. Furthermore, we are interested in other effective ways of applying parallel computation to help improve the research activity of the whole campus.

References

- [Anderson, 2011] M. Anderson, "Better Bechmarking for Supercomputers", IEEE Spectrum, vol. 48, no. 1, pp. 12-14, Jan. 2011.
- [Antognozzi et al, 2009] M. Antognozzi, A. Bottino, F. De Santi, V. Lera, M. Locatelli, D. Cook, "RE-LIVING LAS VEGAS: A multi-user, mixed-reality edutainment environment based on the enhancement of original archival materials", Proc. 2009 IEEE Int. Conf. on Virtual Environments, Human-Computer Interfaces and Measurements Systems, pp. 292-297, 2009.
- [Bottino and Laurentini, 2011] A. Bottino, A. Laurentini, "The Intrinsic Dimensionality of Attractiveness: a Study in Face Profiles", Submitted to Pattern Recognition Letters, in review, 2011.
- [Della Croce and Grosso, 2012] F. Della Croce, A. Grosso, "Improved core problem based heuristics for the 0/1 multi-dimensional knapsack problem", Comput. & Op. Res, 39, pp. 27-31, 2012.
- [Dongarra, 2006] J. Dongarra, "Trends in high performance computing: a historical overview and examination of future developments", IEEE Circuits and Devices Magaz. vol. 22, , pp. 22-27, 2006.
- [Dongarra, 2011] J. Dongarra, "Architecture-Aware Algorithms and Software for Peta and Exascale Computing", Proc. 2011 IEEE Int. Symp. on Parallel and Distrib. Processing, pp. 507, May. 2011.
- [Feng and Cameron, 2007] W. Feng, K.W. Cameron, "The Green500 List: Encouraging Sustainable Supercomputing", IEEE Computer, vol. 40, no. 12, pp. 50-55, Dec. 2007.
- [Kindratenko and Trancoso, 2011] V. Kindratenko, P. Trancoso, "Trends in High-Performance Computing", IEEE Computing in Science & Engineering, vol. 13, no. 3, pp. 92-95, Sept. 2011.
- [Luque et al, 2011] C. Luque, M. Moreto, F. Cazorla, R. Gioiosa, A. Buyurktosunoglu, M. Valero, "CPU Accounting for Multicore Processors", IEEE Trans. on Comput., forthcoming.
- [Sacerdoti et al, 2004] F.D. Sacerdoti, S. Chandra, K. Bathia "Grid Systems Deployment and Management Using Rocks", Proc. 2004 IEEE Int. Conf. on Cluster Comput, pp 337-345, Sept. 2004.
- [Sajjipanon and Ngamsuriyaroj, 2008] P. Sajjipanon, S. Ngamsuriyaroj, "Web Services for MPI-Based Parallel Applications on a Rocks Cluster", Proc. 2008 Asia-Pacific Services Computing Conf., pp. 265-270, Dec. 2008.
- [Shipman et al, 2006] G.M. Shipman, T.S. Woodall, R.L. Graham, A.B. Maccabe, P.G. Bridges, "Infiniband Scalability in Open MPI", Proc. 2006 Int. Symp. on Parallel and Distributed Processing, pp. 10, Apr. 2006.
- [Yanyong et al, 2003] Z. Yanyong, H. Franke, J. Moreira, A. Sivasubramaniam, "An Integrated Approach to Parallel Scheduling Using Gang-Scheduling, Backfilling, and Migration", IEEE Transactions on Parallel and Distributed Systems, vol. 14, no. 3, pp. 236-247, Mar. 2003.

Web References

- [dauin-hpc.polito.it] "The DAUIN HPC Initiative", <http://dauin-hpc.polito.it>.
- [www.labinf.polito.it] "LABINF - Laboratorio di Informatica Avanzata", <http://www.labinf.polito.it>.
- [www.netlib.org] "LINPACK libraries and benchmarks on NetLib", <http://www.netlib.org/linpack/>.
- [www.prace-project.eu] "PRACE: Partnership for Advanced Computing in Europe", <http://www.prace-project.eu>.
- [www.rocksclusters.org] "The ROCKS Cluster Distribution", <http://www.rocksclusters.org>.
- [www.top500.org] "The Top500 Supercomputer List", <http://www.top500.org>.

Acquisizione del traffico IP da una rete ad elevato throughput utilizzando un sistema Open Source: tecnologie e problematiche

Marco Mezzalama, Gianluca Oglietti¹

Politecnico di Torino – DAUIN

Corso Duca degli Abruzzi, 24 - 10129 Torino (TO)

marco.mezzalama@polito.it

¹Politecnico di Torino – AreaIT

Corso Duca degli Abruzzi, 24 - 10129 Torino (TO)

gianluca.oglietti@polito.it

L'attività di acquisizione del traffico IP di una rete di calcolatori, attività di fondamentale importanza durante le normali operazioni di monitoring, di troubleshooting e nell'ambito della Sicurezza Informatica, è divenuta negli ultimi anni sempre più complessa a causa non solo del continuo e costante incremento della quantità dei dati scambiati ma anche e soprattutto a causa dell'elevato throughput raggiunto (all'interno di alcuni data center aziendali infatti il throughput è aumentato addirittura di 100 volte negli ultimi 4 anni passando ad esempio dai 100 Mbit/s del 2007 ai 10 Gbit/s del 2011). Questo costante e continuo aumento della velocità del flusso di dati non è però coinciso con un equivalente aumento del clock delle singole CPU che, al contrario, si sono evolute verso architetture di tipo multi-core. Una tale diversità evolutiva ha fin da subito evidenziato una serie di problemi legati soprattutto all'elevatissimo numero di operazioni di I/O che i sistemi avrebbero dovuto gestire a questi livelli di throughput.

1. Introduzione

Le interfacce di rete (Network Interface Controllers, NIC), nonostante il continuo e costante incremento delle velocità delle connessioni dati, sono state per molto tempo equiparate a delle normali periferiche a cui assegnare compiti relativamente semplici come ad esempio l'acquisizione dei singoli pacchetti IP dal livello fisico (modello ISO/OSI) e la copia del loro contenuto nella memoria centrale. Tutto il carico computazione relativo all'elaborazione dei singoli pacchetti veniva quindi demandato alla sola CPU di sistema. Questo approccio è rimasto praticamente immutato per circa 2 decenni e cioè fino a quando il considerevole aumento del throughput sulle reti informatiche ha iniziato a farne emergere una serie di gravi limitazioni strutturali sia in presenza di sistemi con

architetture mono-core che utilizzando architetture multi-core.

In un primo tempo, quando cioè la quasi totalità dei sistemi di acquisizione utilizzati erano realizzati con architetture mono-core ed il throughput di rete non era superiore ai 100Mbit/s, le migliorie che sono state via via apportate ai NIC avevano come unico obiettivo quello di alleggerire il carico computazionale della CPU di sistema. La principale modifica in questa direzione è stata infatti l'introduzione, direttamente all'interno delle schede di rete, di una piccola unità di calcolo specializzata nell'elaborazione dei pacchetti. Parallelamente sono stati poi modificati alcuni aspetti dello stack TCP/IP per diminuire il numero delle richieste di interrupt inviate dalla scheda di rete alla CPU (Interrupt Coalescing) o per ottimizzare alcuni aspetti della trasmissione, o della ricezione, di particolari famiglie di pacchetti IP come ad esempio quelli aventi una dimensione molto superiore al MTU (Maximum Transmission Unit) imposto dalla rete utilizzata (Large Segment Offload).

Con l'aumentare del throughput e con il superamento della soglia del Gbit/s le ottimizzazioni appena descritte si sono rilevate però insufficienti per consentire ai sistemi di acquisizione di gestire una così elevata mole di dati anche in presenza di architetture computazionalmente molto performanti come i sistemi multi-core di tipo NUMA (Non Uniform Memory Access). Per questa ragione sono state apportate una serie di modifiche che hanno interessato profondamente tutta la catena di acquisizione dei pacchetti: le schede di rete (con l'introduzione dei sistemi multi-coda), i driver dei NIC (con lo sviluppo dei driver TNAPI), l'architettura delle schede madri (con l'implementazione di nuove tecnologie come il DCA) e lo stack di rete a livello del kernel del sistema operativo (con lo sviluppo del socket PF_RING).

2. Background

Nei paragrafi seguenti verranno descritte brevemente le principali migliorie apportate negli ultimi anni a tutta la catena di acquisizione dei pacchetti IP.

2.1 Evoluzione hardware in sistemi mono-core

Le interazioni fra la CPU, la memoria di sistema e la scheda di rete, fin dall'avvento dei primi NIC con throughput massimo di 10Mbit/s, sono sempre state relativamente complesse e hanno richiesto lo studio e l'implementazione di numerosi meccanismi per risolvere sistematicamente tutti i problemi via via riscontrati. Il principale di questi meccanismi è la tecnica del Direct Memory Access (DMA) [1] che permette di escludere l'intervento del processore durante le operazioni di lettura o scrittura effettuate dall'interfaccia di rete sulla memoria di sistema. Le interazioni tipiche presenti in un sistema che utilizza il DMA sono mostrate in Figura 1.

I problemi legati a questo genere di architettura sono dovuti al fatto che l'unico processore di sistema, oltre a doversi occupare dell'esecuzione del codice dei programmi a livello utente e a livello kernel, deve anche eseguire buona parte del codice presente nello stack TCP/IP. Durante l'esecuzione del codice poi, in presenza di un elevato throughput, il processore viene

Acquisizione del traffico IP da una rete ad elevato throughput utilizzando un sistema Open Source continuamente interrotto dalle richieste di interrupt della scheda di rete: queste interruzioni non fanno altro che generare dei “context switch” della CPU che comportano una repentina diminuzione delle prestazioni dell'intero sistema (il processore trascorre più tempo a gestire gli interrupt della scheda che ad eseguire codice utente).

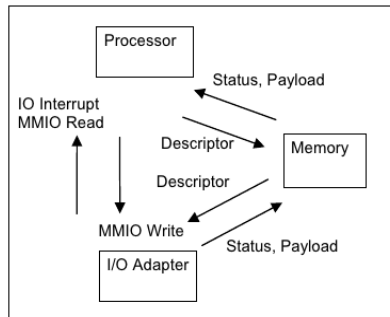


Figura 1 - Iterazioni fra CPU, memoria di sistema e scheda di rete in una architettura mono-core

Per diminuire il carico sul processore di sistema è stata introdotta su tutte le schede di rete una piccola unità di calcolo specializzata sulla quale è stata via via spostata l'esecuzione di buona parte del codice relativo allo stack TCP/IP. Il calcolo del checksum di ogni pacchetto trasmesso e/o ricevuto (Checksum Offload) e la segmentazione dei pacchetti (Large Segment Offload), operazione necessaria per inviare i dati secondo il MTU (Maximum Transmission Unit) della rete, sono due esempi relativi al codice migrato.

Per diminuire il numero di interrupt generati dalla scheda di rete è stato poi introdotto il meccanismo dell'Interrupt Coalescing (o Interrupt Moderation) che permette alla scheda di non contattare il processore ad ogni pacchetto TCP/IP ricevuto ma di contattarlo invece solo quando ne sono stati ricevuti un certo numero (o quando è stato raggiunto un determinato time-out).

Il Checksum Offload, il Large Segment Offload e l'Interrupt Coalescing vengono comunemente raggruppati ed indicati come meccanismi di Traffic Onload Engines (TOE) [2].

Tutti questi meccanismi, assieme all'incremento delle prestazioni delle CPU e delle memorie di sistema, hanno permesso ai normali server di gestire con una discreta efficienza un traffico di rete avente un throughput massimo che, nell'arco dell'ultimo decennio, è passato da 10 Mbit/s a 1 Gbit/s. Questi meccanismi però, anche se utilizzati con le ultimissime CPU messe in commercio, non risultano essere più sufficienti in presenza di throughput ancora più elevati.

2.2 Evoluzione hardware in sistemi multi-core

La necessità di dover incrementare il throughput delle reti informatiche, a causa principalmente dell'avvento di quello che viene definito “Web 2.0” (maggiori contenuti multimediali disponibili in Internet, pagine web più

complesse, VoIP, servizi in real time, P2P, ecc), ha messo in luce tutti quelli che sono i limiti imposti dai meccanismi di tipo TOE. Con l'aumentare del throughput infatti questa soluzione, oltre ad essersi rivelata poco scalabile, poco flessibile e addirittura costosa, si è rivelata essere anche strutturalmente inadatta a sfruttare al meglio le caratteristiche e le potenzialità messe a disposizione dai sistemi multi-core di ultima generazione come ad esempio l'architettura di tipo Non Uniform Memory Access (NUMA) di Intel mostrata in Figura 2.

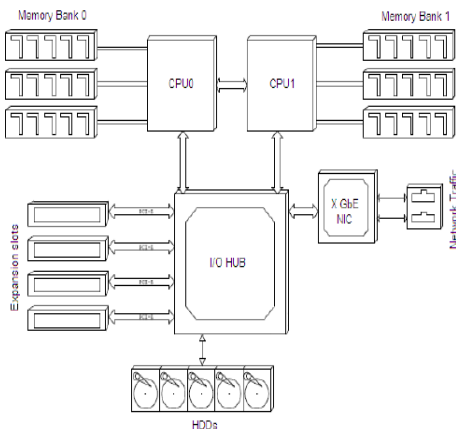


Figura 2 - Architettura di un sistema NUMA di Intel

Le architetture NUMA della famiglia Intel (in questo articolo verranno trattate solo architetture di questo tipo a causa della loro ampia diffusione sul mercato e per alcuni interessanti meccanismi in esse implementati) sono caratterizzate dal fatto che ognuna delle CPU presenti nel sistema contiene al suo interno un controller dedicato per indirizzare una parte della memoria di sistema e dal fatto che per comunicare fra di loro e con il chipset viene utilizzato un bus ad elevate prestazioni chiamato QuickPath Interconnect (QPI).

Se da un lato queste caratteristiche aumentano notevolmente le prestazioni del sistema (diminuendone la latenza) dall'altro la presenza contemporanea di due o più controller per la gestione della memoria di sistema, ne richiede un attento utilizzo. Usando in modo scorretto un'architettura di tipo NUMA si può infatti incorrere ad esempio alla saturazione del bus QPI, a problemi di coerenza della memoria cache dei singoli core e all'incremento sostanziale della latenza durante le operazioni di lettura (o scrittura) in memoria. I nuovi miglioramenti introdotti recentemente ai vari componenti della catena di acquisizione dei pacchetti IP cercano quindi di sfruttare al meglio le caratteristiche messe a disposizione dalle architetture NUMA (e più in generale da tutte le architetture multi-core) per ottenere una maggiore scalabilità e flessibilità rispetto ai meccanismi TOE precedentemente utilizzati.

Per poter sfruttare le caratteristiche messe a disposizione dalle architetture di tipo NUMA sulle nuove schede di rete ad alte prestazioni (con throughput di 1 Gbit/s e 10 Gbit/s) sono state introdotte alcune tecnologie il cui scopo principale

Acquisizione del traffico IP da una rete ad elevato throughput utilizzando un sistema Open Source è quello di distribuire in modo ottimizzato il flusso dei dati acquisito su tutti i core presenti sul sistema.

Le principali tecnologie supportate dalle schede di rete di ultima generazione, quindi, sono [3, 4, 5, 6, 7]: Multiple descriptor queues (per gestire più code di pacchetti sulla stessa scheda), Receive Side Scaling o RSS (per suddividere i pacchetti fra le varie code), Virtual Machine Device Queue o VMDq (per assegnare alla medesima coda tutti i flussi di dati diretti verso una stessa macchina virtuale), Extended Message-Signaled Interrupt o MSI-X (per assegnare gli interrupt generati da una coda ad un particolare core), Receive Side Coalescing o RSC (per aggregare più pacchetti provenienti dallo stesso flusso in un pacchetto più grande), Low Latency Interrupts o LLI (per assegnare ad un particolare tipo di traffico IP una maggiore priorità) e Header Splitting and Replication (per ottimizzare la gestione dei pacchetti in memoria).

Le idee che stanno alla base di queste tecnologie sono quelle di:

- creare all'interno della scheda di rete alcune code indipendenti di pacchetti;
- assegnare in modo univoco ogni pacchetto ricevuto alla coda corretta;
- assegnare ad ogni core di sistema (o macchina virtuale) la coda di pacchetti corretta;
- facilitare l'interazione fra il sistema, le code e i vari core.

Queste tecnologie non sono però ancora sufficienti per poter utilizzare al meglio le potenzialità messe a disposizione dai sistemi di calcolo con architettura NUMA. L'elevato numero di interrupt generato dalle code di pacchetti riducono infatti notevolmente le performance del sistema in quanto i vari core devono occupare molti dei loro cicli di clock per effettuare le operazioni di I/O necessarie a spostare i pacchetti dalle code alla memoria di sistema.

Per risolvere questo problema Intel è andata a modificare alcuni componenti dell'architettura NUMA introducendo una tecnologia, o meglio, un'insieme di tecnologie chiamate Input/Output Acceleration Technology (I/OAT) [8, 9, 10]. Queste nuove tecnologie contribuiscono ad aumentare le prestazioni di un sistema di acquisizione di pacchetti IP andando ad ottimizzare l'utilizzo della memoria centrale e della memoria cache di ogni processore. Le tecnologie che sono alla base dell'I/OAT sono principalmente due: Intel QuickData Technology e Direct Cache Access (DCA).

L'Intel QuickData Technology è stata introdotta in quanto buona parte del tempo necessario ad elaborare un pacchetto appena ricevuto veniva impiegato (inutilmente) dalla CPU per spostare dati all'interno della memoria di sistema. In una situazione normale infatti, quando il processore esegue la copia o lo spostamento dei dati contenuti nella memoria di sistema, deve rimanere di fatto bloccato fino al termine dell'operazione. Usando questa tecnologia invece il processore non deve attendere il completamento dell'operazione di copia in quanto essa viene demandata al motore DMA inserito direttamente nel controller. Come è possibile osservare in Figura 3 la CPU, una volta istruito il controller (AMC) sull'operazione da eseguire, non rimane in stallo ad attendere che il dato venga copiato ma può iniziare ad elaborare l'istruzione successiva

(ad esempio può iniziare ad elaborare il prossimo pacchetto). Sarà poi lo stesso controller ad avvisare il processore quando la copia richiesta in precedenza è stata ultimata.

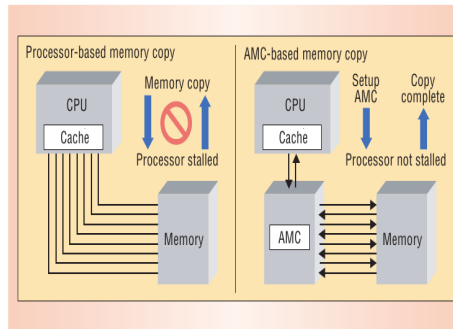


Figura 3 - Intel QuickData Tecnology

Il Direct Cache Access è stato introdotto per migliorare le prestazioni del sistema durante il trasferimento dei dati appena acquisiti dalla scheda di rete verso la memoria cache del processore designato per la loro elaborazione.

In un normale sistema di acquisizione infatti, come è possibile osservare dalla Figura 4, il dato acquisito dalla scheda di rete deve transitare per la memoria di sistema (DMA) prima di poter raggiungere la memoria cache del processore e aggiungendo quindi all'operazione una notevole latenza. Oltre ai problemi di latenza bisogna poi tenere in considerazione un ulteriore overhead dovuto ai meccanismi che mantengono la coerenza dei dati contenuti nella cache. Il DCA risolve questi problemi permettendo alla scheda di rete di scrivere il dato appena acquisito direttamente all'interno della cache del processore corretto senza dover passare dalla memoria di sistema. I principali benefici del DCA sono due: la considerevole riduzione del tempo necessario al dato appena acquisito per raggiungere il cuore dell'unità di calcolo e la riduzione della banda utilizzata dalla memoria di sistema.

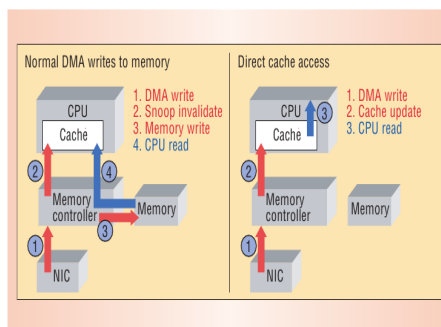


Figura 4 - Direct Cache Access

2.3 Evoluzione Software – il socket PF_RING

Uno dei primi problemi che hanno interessato il sistema operativo Linux in

Acquisizione del traffico IP da una rete ad elevato throughput utilizzando un sistema Open Source ambito networking è stato riscontrato nel ramo di sviluppo 2.4 del kernel (in modo particolare nelle versioni precedenti alla 2.4.20) ed era caratterizzato da un'evidente degradazione delle prestazioni del sistema in presenza di un elevato traffico di rete (dell'ordine dei 100 Mbit/s). La causa di questo problema è stata attribuita al fatto che, in queste condizioni, il sistema operativo era impegnato per la quasi totalità del tempo a gestire gli interrupt generati dal NIC all'arrivo dei pacchetti a scapito dei processi in esecuzione a cui era invece riservato il poco tempo restante.

Per risolvere questo problema e per migliorare le prestazioni del sistema, a partire dal ramo di sviluppo 2.5 di Linux, è stata introdotta una nuova interfaccia di comunicazione fra il kernel e i driver dei dispositivi di rete chiamata NAPI (New API) che, fra le varie caratteristiche, introduceva l'utilizzo di una tecnica di mitigazione degli interrupt basata sul metodo del Polling. La caratteristica principale di questa nuova tecnica, infatti, è che i driver del dispositivo non possono inviare un interrupt al sistema in un qualsiasi istante di tempo ma solo dopo essere stati interrogati, ad intervalli più o meno regolari, dal kernel stesso (Interrupt Mitigation).

Alcuni studi sull'acquisizione passiva dei pacchetti IP su reti ad elevato throughput [11] effettuati negli ultimi anni da Luca Deri hanno però evidenziato che molti pacchetti non vengono acquisiti dal sistema in quanto il kernel Linux con interfaccia NAPI impiega un tempo non trascurabile per spostare fisicamente un pacchetto dal device al software in userspace. Per risolvere questo problema Deri ha quindi introdotto un nuovo socket chiamato PF_RING [13] che permette di diminuire il tempo impiegato dal pacchetto ad attraversare le code e le strutture dati presenti all'interno del kernel Linux. Questo socket, come è possibile osservare in Figura 5, è basato su di un buffer circolare che viene allocato in memoria durante l'inizializzazione del socket stesso e viene deallocato solamente alla sua chiusura eliminando quindi la necessità di effettuare operazioni di allocamento della memoria ogni volta che vengono ricevuti dei pacchetti dalla scheda di rete.

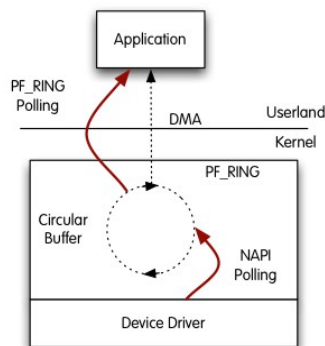


Figura 5 - Il socket PF_RING

Quando un pacchetto viene ricevuto dall'interfaccia di rete (tramite il NAPI Polling) esso, invece di essere gestito dagli strati superiori del kernel, viene

inserito direttamente all'interno del buffer circolare che viene poi esportato a livello utente in modo tale che le applicazioni possano accedervi direttamente per prelevare i pacchetti. Più applicazioni possono accedere allo stesso socket: in questo caso il PF_RING fornirà un buffer circolare ad ogni applicazione evitando quindi l'insorgere di eventuali interferenze fra di esse (la presenza di applicazioni intrinsecamente lente non andranno a rallentare applicazioni più veloci durante la fase di lettura dei pacchetti presenti nei buffer circolari).

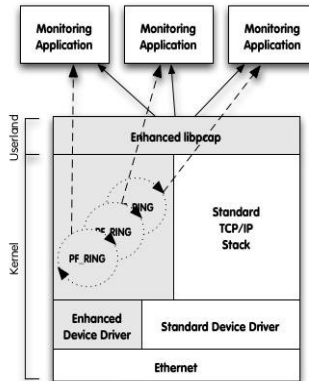


Figura 6 - PF_RING e Clustering

Il codice alla base del socket PF_RING offre poi una serie di funzionalità molto utili durante l'acquisizione passiva del traffico IP di una rete informatica. In particolare infatti il socket permette di filtrare i pacchetti inviati dal driver della scheda prima che vengano inseriti all'interno del buffer circolare, e di incrementare le performance dei software di acquisizione ed elaborazione dei pacchetti grazie ai meccanismi di bilanciamento e di clustering. Come è possibile osservare in Figura 6, questi meccanismi permettono di inviare ad una applicazione solo una porzione dell'intero flusso di pacchetti e di instradare il flusso rimanente alle altre applicazioni che fanno parte del cluster. Un cluster di applicazioni può essere configurato attraverso delle semplici regole in base al flusso dei dati o attraverso la tecnica del round-robin.

Il socket PF_RING può essere utilizzato da un qualsiasi device driver ma per alcune periferiche sono stati sviluppati alcuni driver ad hoc (nel seguito indicati come PF_RING-aware) che permettono di copiare i pacchetti appena acquisiti direttamente all'interno del socket stesso senza utilizzare altri meccanismi intermedi messi a disposizione dal sistema operativo. Esistono infine alcuni driver, sempre sviluppati da Deri, che permettono di mappare la memoria della scheda di rete a livello utente (driver PF_RING DNA, Direct NIC Access): in questo modo, come è possibile osservare in Figura 7, i pacchetti vengono copiati all'interno del buffer circolare direttamente dal processore presente sulla scheda di rete senza passare dall'interfaccia NAPI del sistema operativo. Questa ulteriore ottimizzazione però può essere utilizzata solamente da un'applicazione alla volta.

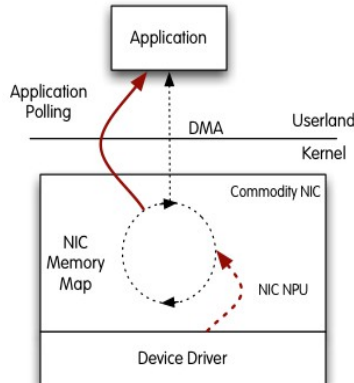


Figura 7 - PF_RING e driver PF_RING DNA

Le applicazioni che già si interfacciano alle schede di rete utilizzando le librerie libpcap, per poter utilizzare il socket PF_RING, non devono essere riscritte o modificate ma devono essere semplicemente ricompilate utilizzando delle particolari librerie chiamate libpcap-PF_RING che permettono di accedere in modo trasparente al nuovo socket PF_RING.

Il socket PF_RING, per garantire la compatibilità con tutti i driver sviluppati in precedenza, può essere avviato in tre diverse modalità di funzionamento configurabili tramite il parametro "transparent_mode":

- modalità 0: in questa modalità i pacchetti sono inviati al buffer circolare utilizzando i meccanismi standard del kernel. Questa modalità è quella di default ed è utilizzabile da tutti i driver delle schede di rete.
- modalità 1: in questa modalità i pacchetti sono inviati al buffer circolare direttamente dal driver PF_RING-aware della scheda di rete, una copia del pacchetto viene comunque ancora passata al kernel utilizzando i meccanismi standard.
- modalità 2: in questa modalità i pacchetti sono inviati al buffer circolare direttamente dal driver PF_RING-aware della scheda di rete.

La modalità 2 risulta essere la più performante ma è necessario sottolineare che essa comporta l'assenza di connettività per le schede di rete che la utilizzano in quanto i pacchetti ricevuti non vengono inviati al kernel tramite i meccanismi standard.

2.4 Evoluzione Software – i driver TNAPI

Come si è visto nei paragrafi precedenti, le moderne schede di rete ad elevato throughput hanno una struttura interna che permette di suddividere il flusso dei pacchetti acquisiti su più code utilizzando ad esempio alcune funzionalità hardware come il Receive Side Scaling (RSS) di Intel. Queste tecniche però, essendo nate soprattutto per essere utilizzate in ambienti virtualizzati, non sono comunemente sfruttate per il monitoraggio del traffico di

rete anche perché i driver a disposizione non permettono ai programmi a livello utente di sfruttare appieno questo tipo di parallelizzazione. Quello che solitamente accade infatti, come si può vedere in Figura 8, è che i normali driver disponibili per queste periferiche fondono i vari flussi di dati provenienti dalle diverse code della scheda in un unico flusso offrendo ai software (anche se multi-thread) un solo punto di accesso ai dati acquisiti (ad esempio la classica interfaccia eth1). Questo è sicuramente un collo di bottiglia di notevole entità.

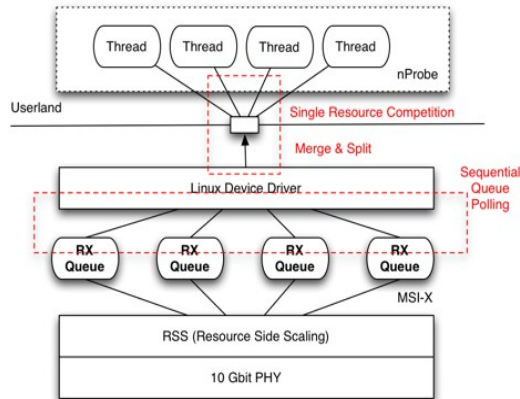


Figura 8 - Driver standard con schede multi-coda

Si pensi ad esempio ad un software multi-thread (ottimizzato cioè per essere eseguito su architetture multi-core) dove i vari thread devono sincronizzarsi fra di loro (di solito utilizzando i semafori) per poter accedere all'unica sorgente di pacchetti disponibile: all'aumentare del numero di thread, contrariamente alle attese, è possibile osservare che le prestazioni del software diminuiscono considerevolmente. Questo accade essenzialmente per due motivi:

- i vari thread trascorrono la maggior parte del tempo a scambiarsi fra loro i segnali di sincronizzazione. Solo un thread alla volta infatti può trovarsi nello stato di “ready” (può cioè accedere all'unica fonte di pacchetti disponibile) mentre tutti gli altri devono obbligatoriamente essere in uno stato di “busy” (sono cioè in uno stato di attesa dove non eseguono alcuna operazione)
- la modalità di accesso alla fonte dei dati, essendo puramente casuale,

non permette di assegnare ad uno stesso core sia l'interrupt generato dalla coda della scheda di rete in cui transita un determinato pacchetto che il thread dell'applicazione che lo elaborerà. Questa casualità strutturale comporta quindi la mancata ottimizzazione della cache dei vari core rendendo necessario un uso intensivo della memoria di sistema (a causa dei numerosi context switch) con una conseguente degradazione delle prestazioni dell'applicazione.

Per risolvere questi problemi Deri ha sviluppato un nuovo driver multi-thread (disponibile solo per alcune periferiche di ultima generazione) chiamato TNAPI [12, 14] (Thread-NAPI) che, come è possibile osservare in Figura 9, associa un

Acquisizione del traffico IP da una rete ad elevato throughput utilizzando un sistema Open Source thread di accesso ai dati ad ognuna delle code disponibili sull'interfaccia di rete.

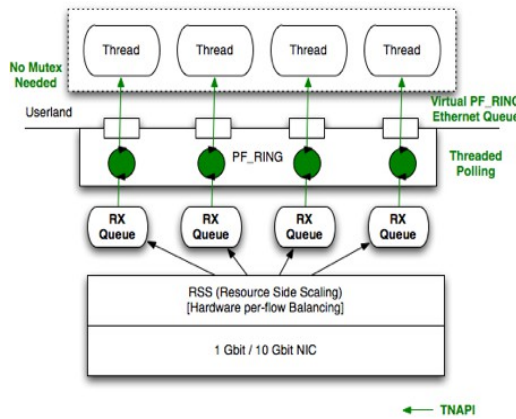


Figura 9 - Driver TNAPI con schede multi-coda

In questo modo è possibile introdurre il concetto di adattatore di rete virtuale (indicato ad esempio con il device linux `eth2@4` che sta ad indicare la coda numero 4 dell'interfaccia di rete `eth2`) fornendo quindi l'accesso alle singole code a livello utente. Le varie applicazioni, in presenza dei driver TNAPI, potranno quindi accedere ai pacchetti acquisiti dal NIC in due modi: utilizzando il device fisico (es. `eth1`) per accedere ai pacchetti acquisiti da tutte le code o utilizzando i device virtuali (es. `eth1@3`) per accedere alle singole coda senza la necessità di sincronizzare i vari thread con i semafori.

2.5 SMP IRQ Affinity

Un'importante fattore di cui bisogna tenere conto quando si vuole acquisire del traffico di rete ad elevato throughput è l'associazione fra gli interrupt generati dai NIC (o dalle code presenti nelle schede) all'arrivo dei pacchetti IP e il core che dovrà elaborarli. Tale associazione prende il nome di IRQ affinity o SMP affinity [15, 16] e può essere effettuata in diversi modi in base a quale dei seguenti obiettivi si intende raggiungere:

- massimizzare il livello di coerenza della cache di ogni core
- migliorare la distribuzione degli interrupt fra i vari core

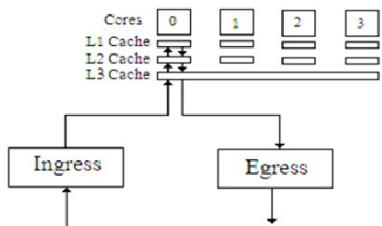


Figura 10 - SMP Affinity per l'ottimizzazione della cache

In Figura 10 è possibile osservare la configurazione di IRQ Affinity da

adottare per massimizzare la coerenza della cache mentre in Figura 11 è rappresentata la configurazione di IRQ Affinity per migliorare la distribuzione degli interrupt fra i vari core.

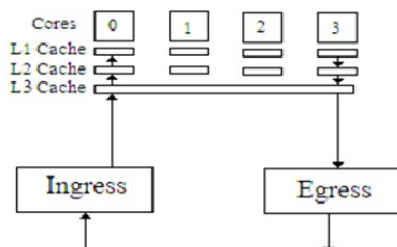


Figura 11 - SMP Affinity per migliorare la distribuzione degli interrupt

Le due configurazioni sono sostanzialmente differenti in quanto la prima tende ad assegnare gli interrupt ad uno stesso core (migliorando le prestazioni di un sistema NUMA ma peggiorando le prestazioni in caso di elevato throughput) mentre la seconda a distribuirli fra tutti i core disponibili (migliorando le prestazioni in caso di elevato throughput ma peggiorandole in presenza di un sistema NUMA). La configurazione che deve essere adottata nella pratica dipende da molti fattori come ad esempio il tipo di applicazione in esecuzione sui core, il tipo di traffico da acquisire ed il suo throughput: in alcuni casi le prestazioni migliori si ottengono massimizzando il livello di coerenza delle cache mentre altre volte migliorano distribuendo gli interrupt su più core, altre volte ancora utilizzando una configurazione ibrida.

In Figura 12 è possibile infine osservare la configurazione consigliata che dovrebbe essere adottata in presenza di un sistema Simultaneous Multi Threading (SMT) di Intel dove ogni core fisico, utilizzando la tecnologia HT (Hyper Threading) è affiancato da un core logico (il numero dei core utilizzabili è doppio rispetto al numero di core fisici). In questo caso è consigliabile, quando possibile, associare tutti gli interrupt diretti al core fisico e al relativo core logico in quanto condividono la stessa memoria cache.

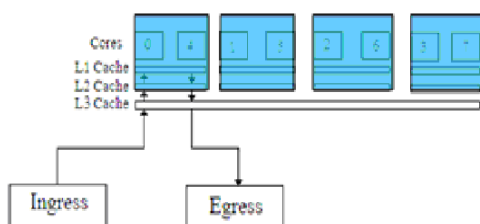


Figura 12 - SMQ Affinity consigliata con sistema SMT

3. Realizzazione di un sistema di acquisizione

Per poter effettuare alcuni test sulle tecnologie analizzate nei paragrafi precedenti si è realizzato un sistema di acquisizione per reti ad elevato throughput (fino a 10Gbit/s).

Acquisizione del traffico IP da una rete ad elevato throughput utilizzando un sistema Open Source

Nei paragrafi seguenti quindi sono state descritte le caratteristiche principali dei componenti (scheda di rete, server, sistema operativo e driver) che sono stati utilizzati per la realizzazione del sistema di acquisizione del traffico IP.

3.1 Scheda di rete

La scheda di rete utilizzata è il modello “X520-SR2” [17] di Intel.

Questo NIC è composto da due interfacce in fibra da 10 Gbit/s (tramite adattatori GBIC) ed è basato sul chip 82599 che risulta essere uno degli ultimi chip messi in commercio da Intel. Le caratteristiche principali di questa scheda sono:

- Multiple queues: 128 code di pacchetti in ricezione e 128 code in trasmissione (vengono attivate solo una quantità di code pari al numero di core del sistema)
- Receive (and Transmit) Side Scaling per i protocolli IP (v4 e v6), TCP/UDP
- Checksum Offload per i protocolli IP (v4 e v6), SCTP, TCP e UDP
- Header Splits and Replication per il traffico ricevuto
- Advanced Packet Filtering (filtraggio dei pacchetti in hardware)
- Low Latency Interrupts
- Intel Direct Cache Access (DCA)
- Supporto agli interrupt MSI-X
- Interfaccia PCI-Express 2.0 (5.0 GT/s)
- Processore dedicato: Intel 82599

3.2 Server di acquisizione

Il server utilizzato è il modello “SYS-6026T-6RFT+” della casa produttrice SuperMicro.

In un primo tempo si era pensato di utilizzare un server della casa produttrice HP appartenente alla famiglia dei DL-380 ma, purtroppo, analizzando il BIOS del sistema ci si è subito accorti che non supportava alcune tecnologie fondamentali come l'I/OAT di Intel. Per questo motivo è stato scelto il server SuperMicro che, per contro, permette una configurazione molto più puntuale (attraverso il BIOS) di molti parametri dell'hardware del sistema. Le principali caratteristiche di questo server sono:

- CPU: 2 x Intel Xeon X5560 @ 2.80GHz (16 core)
- Architettura NUMA con bus QPI (6.4 GT/s)
- RAM: 12 GB DDR3 @ 1333 MHz ECC
- Chipset: Intel 5520 e ICH10R
- Supporto completo alla tecnologia I/OAT 3 e VMDq

3.3 Sistema operativo, driver e software

Per garantire la massima versatilità e sicurezza possibile, sul server SuperMicro “SYS-6026T-6RFT+”, è stato installato il sistema operativo Open

Source GNU/Linux Debian 6.0.2 (Squeeze, stable) per architetture a 64bit con kernel Linux 2.6.38.8 compilato manualmente per abilitare tutti i moduli necessari al corretto funzionamento del sistema (supporto ad architetture NUMA, modulo ioatdma per la gestione della tecnologia QuickData di Intel, modulo dca per il supporto alla tecnologia Direct Cache Access, ...).

Per poter utilizzare tutte le funzionalità messe a disposizione dalla scheda di rete scelta sono stati utilizzati i driver TNAPI (versione 3.3.8) e il socket PF_RING (versione 4.7.0 – revisione 4701) [18].

Per le varie prove infine sono stati utilizzati applicativi messi a disposizione dalla distribuzione Debian e applicativi presenti all'interno del pacchetto del socket PF_RING distribuito da Deri.

Va evidenziato quindi il fatto che il sistema di acquisizione è stato realizzato interamente utilizzando solo software Open Source.

4 Problematiche riscontrate

La scelta di utilizzare solamente codice Open Source all'interno di un progetto di questo tipo (utilizzo di nuove tecnologie ad elevate prestazioni) è dettata dal fatto che questa tipologia di software porta con sé notevoli benefici. Si pensi, ad esempio, all'opportunità di poter accedere al codice sorgente del software per poterlo migliorare correggendo, praticamente in “real time”, gli eventuali errori di programmazione individuati durante le varie fasi di test o aggiungendo via via ad esso tutte le funzionalità di cui si ha bisogno: operazioni che, nella stragrande maggioranza dei casi, risultano essere assolutamente impossibili da realizzare utilizzando software a sorgenti chiusi.

Durante i test effettuati sul sistema di acquisizione appena descritto sono stati però evidenziati alcuni problemi in qualche modo legati alla tipologia di software utilizzato. Tali problemi verranno descritti nei seguenti paragrafi.

4.1 Evoluzione del software Open Source

Una delle caratteristiche peculiari di un software Open Source è la sua modalità di evoluzione. Nella maggior parte dei casi infatti un software di questo tipo è caratterizzato da un'elevatissima velocità di evoluzione soprattutto durante le prime fasi del suo sviluppo. Una così estrema rapidità evolutiva risulta essere però un problema considerevole soprattutto quando il software Open Source viene impiegato in ambienti in cui è necessario utilizzare nuove tecnologie (si pensi ad esempio all'I/OAT di Intel o al socket PF_RING) o dispositivi hardware relativamente recenti (si pensi alle schede di rete 10Gbit/s che implementano la tecnologia multi-coda).

Uno dei problemi riscontrati con maggior frequenza, durante i numerosi test effettuati sul sistema di acquisizione di pacchetti IP implementato, è stato quello di dover continuamente aggiornare una buona parte del software utilizzato a causa delle numerose e frequenti modifiche apportate dagli sviluppatori al codice del socket PF_RING o al codice del driver TNAPI. Il continuo aggiornamento di questi software si è rivelato assolutamente necessario in quanto le modifiche apportate erano focalizzate principalmente a correggere

Salvamente, a causa di un semplice aggiornamento del codice del socket PF_RING, era possibile ad esempio riscontrare problemi di compatibilità con i driver TNAPI. Aggiornando quindi i driver TNAPI, per allinearli alla versione del socket PF_RING installata, a volte venivano però riscontrati problemi di compatibilità con il kernel in uso. Aggiornando infine il kernel del sistema operativo era ancora possibile riscontrare alcuni problemi con altre parti del sistema che risultavano essere compatibili, ad esempio, solo con una versione precedente del kernel Linux.

4.2 Kernel Panic

```
[76051.431130] kernel BUG at net/pv4/inetpeer.c:381
[76051.488428] invalid opcode: 0000 [#1] SMP
[76051.537630] last sysfs file: /sys/bus/pci/drivers/ixgbe/uevent
[76051.607385] CPU 2
[76051.629288] Modules linked in: pf_ring ixgbe microcode loop uhci_hcd ioh401m igb ehci_hcd usbcore sg psmouse rtc_cmos rtc_core sr_mod
i2c_i801 i2c_core rtc_lib dca evdev mdio serio_raw [last unloaded: pf_ring]
[76051.874470]
[76051.882326] Pid: 0, comm: kworker/0:1 Tainted: G W 2.6.38.8-superechelon #1 Supermicro X8DTU-6+/X8DTU-6+
[76052.015641] RIP: 0010:[<ffffffff12791f8>] [<ffffffff12791f8>] cleanup_once+0x13f/0x210
[76052.113527] RSP: 0018:ffff800bf4839f0 EFLAGS: 00010287
[76052.177055] RAX: ffff801b4c4b000 RBX: ffffffff13e74e0 RCX: 0000000000000001
[76052.262381] RDY: ffff800bf483ad0 RSI: 0000000000000000 RDI: 0000000046d553d5
[76052.347707] RBP: ffffffff13e74e8 R08: 0000000046d74848 R09: 00000000000000012
[76052.433032] R10: ffffffff13e74e0 R11: 0000000000000013 R12: ffff800bf483ad0
[76052.518358] R13: ffffffff130b080 R14: ffff80171c3a8a8 R15: ffff80171c3a880
[76052.603683] FS: 0000000000000000(0000) GS: ffff800bf480000(0000) knlGS:0000000000000000
[76052.700427] CS: 0010 DS: 0000 ES: 0000 CR0: 000000008005003b
[76052.769144] CR2: 000077fd7826ad00 CR3: 00000000013b9000 CR4: 0000000000000060
[76052.854470] DR0: 0000000000000000 DR1: 0000000000000000 DR2: 0000000000000000
[76052.939797] DR3: 0000000000000000 DR6: 00000000ffff0000 DR7: 0000000000000400
[76053.025122] Process kworker/0:1 (pid: 0, threadinfo ffff801b90a8000, task ffff801b907d780)
[76053.126018] Stack:
[76053.150101] ffff800bf483a40 ffff8017d7a9d00 ffffffff13e74e0 ffffffff13e74e8
[76053.239058] ffff800bf483a40 ffff800bf483bc0 ffff800bf483ad8 ffff801b1279404
[76053.328016] 0000000000000000 000000001383aa170 ffffffff13e74e0 ffff80033818c400
[76053.416975] Call Trace:
[76053.446250] <IRQ>
[76053.471473] [<fffff8b1279404>] ? inet_getpeer+0x1a4/0x260
[76053.539153] [<fffff8b1277d37>] ? ip_route_input_slow+0x307/0x790
[76053.614099] [<fffff8b127a1b9>] ? ip4_frag_init+0x89/0xa0
[76053.680742] [<fffff8b12add87>] ? inet_frag_find+0x97/0x230
[76053.749456] [<fffff8b127a316>] ? ip_defrag+0xd6/0xb80
[76053.818294] [<fffff8b1279d28>] ? ip_local_deliver+0x28/0x1a0
[76053.883779] [<fffff8b1257639>] ? netif_receive_skb+0x2f9/0x3d0
[76053.9587823] [<fffff8b125bfa8>] ? netif_receive_skb+0x78/0xb0
[76054.029517] [<fffff8b125c52b>] ? napi_gro_receive+0xbb/0xd0
[76054.099272] [<fffff8b125c0d8>] ? napi_skb_finish+0x38/0x50
[76054.167993] [<fffff8a0155a6c>] ? ixgbe_receive_skb+0x1cc/0x1e0 [ixgbe]
[76054.249164] [<fffff8b1184641>] ? unmap_single+0x31/0x90
[76054.314769] [<fffff8a0159526>] ? ixgbe_poll+0x636/0x14a0 [ixgbe]
[76054.389712] [<fffff8a0159e21>] ? ixgbe_poll+0xf3f/0x14a0 [ixgbe]
[76054.464657] [<fffff8b125c676>] ? net_rx_action+0x86/0x170
[76054.532363] [<fffff8b12407dc>] ? dma_issue_pending_all+0x7c/0xb0
[76054.607282] [<fffff8b103e6f1>] ? __do_softirq+0x91/0x140
[76054.673924] [<fffff8b100334c>] ? call_softirq+0x1c/0x30
[76054.739527] [<fffff8b100520d>] ? do_softirq+0x4d/0x80
[76054.803053] [<fffff8b10047ac>] ? do_IRQ+0x5c/0xd0
```

Congresso Nazionale AICA 2011

```
[76054.862429] [<fffff812d10d3>] ? ret_from_intr+0x0/0xe
[76054.926993] <EOI>
[76054.952219] [<fffff8119a524>] ? intel_idle+0xb4/0x110
[76055.016783] [<fffff8119a504>] ? intel_idle+0x94/0x110
[76055.081348] [<fffff8123f101>] ? cpuidle_idle_call+0x81/0xf0
[76055.152140] [<fffff81001640>] ? cpu_idle+0x50/0xa0
[76055.213590] Code: 20 02 0f 95 c1 31 c0 8d 4c 49 01 31 f6 41 8b 7c 07 10 45 8b 44 05 10 44 39 c7 75 c0 ff c6 48 83 c0 04 39 ce 7c e7 4d 39 ef 74 18 <0f>
0b eb fe 48 c7 c7 d0 74 3e 81 e8 b1 7b 05 00 83 c8 ff e9 6e
[76055.446005] RIP [<fffff8127918f>] cleanup_once+0x13f/0x210
[76055.515865] RSP <ffff800bf4839f0>
[76055.557910] ---[ end trace 06a70dacbf2d08ad ]---
[76055.613128] Kernel panic - not syncing: Fatal exception in interrupt
[76055.689113] Pid: 0, comm: kworker/0:1 Tainted: G   DW  2.6.38.8-superechelon #1
[76055.781808] Call Trace:
[76055.811081] <IRQ> [<fffff812ce4d4>] ? panic+0x92/0xa18a
[76055.877930] [<fffff810393c1>] ? kmsg_dump+0x41/0x100
[76055.941458] [<fffff8100661b>] ? oops_end+0x9b/0xa0
[76056.002959] [<fffff810037f4>] ? do_invalid_op+0x84/0xa0
[76056.069549] [<fffff8127918f>] ? cleanup_once+0x13f/0x210
[76056.137230] [<fffff810030d5>] ? invalid_op+0x15/0x20
[76056.200757] [<fffff8127918f>] ? cleanup_once+0x13f/0x210
[76056.268436] [<fffff81279404>] ? inet_getpeer+0x14a/0x260
[76056.336115] [<fffff81277d37>] ? ip_route_input_slow+0x307/0x790
[76056.411059] [<fffff8127a1b9>] ? ip4_frag_init+0x89/0xa0
[76056.477702] [<fffff812add87>] ? inet_frag_find+0x97/0x230
[76056.546420] [<fffff8127a316>] ? ip_defrag+0xd6/0xb80
[76056.609945] [<fffff81279d28>] ? ip_local_deliver+0x28/0x1a0
[76056.680739] [<fffff81257639>] ? __netif_receive_skb+0x2f9/0x3d0
[76056.755686] [<fffff8125bfa8>] ? netif_receive_skb+0x78/0x80
[76056.826478] [<fffff8125c52b>] ? napi_gro_receive+0xbb/0xd0
[76056.896232] [<fffff8125c0d8>] ? napi_skb_finish+0x38/0x50
[76056.964952] [<fffffa0155a6c>] ? ixgbe_receive_skb+0x1cc/0x1e0 [ixgbe]
[76057.046124] [<fffff8118461a>] ? unmap_single+0x31/0x90
[76057.111730] [<fffffa0159526>] ? ixgbe_poll+0x636/0x14a0 [ixgbe]
[76057.186674] [<fffffa0159e21>] ? ixgbe_poll+0xf31/0x14a0 [ixgbe]
[76057.261619] [<fffff8125c676>] ? net_rx_action+0x86/0x170
[76057.329298] [<fffff812407dc>] ? dma_issue_pending_all+0x7c/0xb0
[76057.404243] [<fffff8103e6f1>] ? __do_softirq+0x91/0x140
[76057.470884] [<fffff8100334c>] ? call_softirq+0x1c/0x30
[76057.536488] [<fffff8100520d>] ? do_softirq+0x4d/0x80
[76057.600015] [<fffff810047ac>] ? do_IRQ+0x5c/0xd0
[76057.659391] [<fffff812d10d3>] ? ret_from_intr+0x0/0xe
[76057.723954] <EOI> [<fffff8119a524>] ? intel_idle+0xb4/0x110
[76057.795995] [<fffff8119a504>] ? intel_idle+0x94/0x110
[76057.860561] [<fffff8123f101>] ? cpuidle_idle_call+0x81/0xf0
[76057.931352] [<fffff81001640>] ? cpu_idle+0x50/0xa0
```

Sono state effettuate numerose prove per identificare la causa di questo errore di sistema. In particolare per prima cosa si è deciso di verificare se la causa del problema fosse il codice sviluppato da Deri. Per questa ragione è stato subito disinstallato il socket PF_RING e, di conseguenza, il driver TNAPI è stato sostituito dall'ultima versione del driver standard, per la scheda di rete Intel X520-SR2, disponibile sul sito del produttore. Anche in questo caso, inviando una grande quantità di pacchetti frammentati al server di acquisizione è stato possibile riscontrare l'errore di kernel panic riportato in precedenza con il conseguente blocco del sistema.

Una volta esclusa l'ipotesi che la causa del problema fosse da attribuire al nuovo codice del socket PF_RING o ai driver TNAPI sono state quindi effettuate delle ulteriori prove andando a modificare la IRQ Affinity fra le code della scheda di rete e i core disponibili sui due processori installati sul server. Da questi ultimi test si è ottenuto un risultato abbastanza interessante. Assegnando infatti ad un unico core tutti gli interrupt generati dalle diverse code di acquisizione dei pacchetti non è mai stato riscontrato alcun blocco del sistema. Assegnando invece l'interrupt generato da ogni coda di acquisizione ad un diverso core si è osservato che, sistematicamente dopo alcuni secondi di utilizzo, il server si bloccava con il relativo errore di kernel panic.

Analizzando il trace generato in concomitanza al kernel panic è possibile effettuare alcune ipotesi sulle probabili cause del problema. Una delle più attendibili associa l'errore ad una non corretta gestione dei pacchetti frammentati da parte del codice del kernel di Linux che, in ambienti multi-core,

Acquisizione del traffico IP da una rete ad elevato throughput utilizzando un sistema Open Source

si occupa di ricostruire il pacchetto originario partendo dai frammenti acquisiti dalla scheda di rete. Probabilmente infatti il codice del deframmentatore dei pacchetti IP implementato all'interno del kernel è stato realizzato quando non era pensabile di poter utilizzare schede di rete multi-coda per acquisire pacchetti IP da una rete ad elevato throughput su sistemi multi-core. Andrebbe quindi rivista questa parte di codice del kernel soffermandosi soprattutto sulle tecniche utilizzate per sincronizzare le operazioni di deframmentazione dei pacchetti fra i diversi core in ambienti in cui vengono utilizzate tecnologie particolari (come l'I/OAT di Intel) sviluppate per ottimizzare il funzionamento di dispositivi avanzati come le schede di rete multi-coda introdotte solo di recente.

5 Conclusioni

L'utilizzo di codice Open Source in sistemi tecnologicamente avanzati, come può essere un sistema di acquisizione dei pacchetti IP in reti ad elevato throughput, porta sicuramente a notevoli benefici dovuti principalmente alla filosofia della “condivisione” che sta alla base di questo tipo di software. Permettendo a più persone di poter accedere ai sorgenti del software è infatti possibile ottenere un incremento considerevole della velocità con cui il software stesso si evolve verso una condizione di stabilità, di sicurezza e di completezza. Il più delle volte però, per raggiungere questa condizione, è necessario attraversare fasi intermedie in cui utilizzare software Open Source può risultare quantomeno problematico. Si pensi ad esempio all'incompatibilità che può nascere fra alcuni programmi appena aggiornati e altri software presenti nel sistema. Oppure si pensi al continuo stravolgimento che può subire un software durante la sua primissima fase evolutiva: la nuova versione può stravolgere completamente o in parte la versione utilizzata fino a quel momento. In particolare si sta osservando un comportamento analogo a quello descritto in questo ultimo esempio per il socket PF_RING e i driver TNAPI. Deri, infatti, negli ultimi mesi ha introdotto una nuova versione del socket chiamata PF_RING_DNA e una nuova versione dei driver TNAPI chiamata TNAPIv2; versioni che differiscono in modo considerevole da quelle utilizzate per realizzare il sistema di acquisizione descritto e che, per questo motivo, non sono state trattate in questo articolo.

L'individuazione di un problema a livello del kernel Linux, in grado di bloccare completamente un sistema, è sicuramente una scoperta molto importante ma, per valutarne la gravità, deve essere effettuato uno studio per determinare le cause del problema e le condizioni che si devono verificare per generare il kernel panic. Per quanto riguarda le cause del problema, come si è visto in precedenza, è ragionevole pensare che si possano attribuire quasi certamente al codice del deframmentatore di pacchetti IP implementato all'interno del kernel Linux. Le condizioni che generano il kernel panic invece sono assolutamente certe infatti:

- è necessario aver installato nel sistema una scheda di rete multi-coda di ultima generazione
- è necessario che la IRQ Affinity sia configurata in modo tale da

assegnare un diverso core ad ognuno degli interrupt generati dalle diverse code di acquisizione della scheda di rete

A questo punto per generare il kernel panic è sufficiente inviare un flood di pacchetti frammentati al server di acquisizione utilizzando ad esempio il software hping3.

Da quest'ultima analisi risulta essere chiaro che la probabilità che l'errore individuato a livello del kernel di Linux possa causare un kernel panic su un sistema di uso comune è considerevolmente bassa. Ciò non toglie però che il deframmentatore presente all'interno del kernel di Linux andrebbe controllato o riscritto in modo da poter essere utilizzato correttamente anche in sistemi di acquisizione ad elevato throughput impiegati in ambienti reali dove, peraltro, il verificarsi di un flood di pacchetti IP frammentati analogo a quello generato durante i test è quantomeno improbabile.

Bibliografia

- [1] "Direct cache access for high bandwidth network I/O"; Huggahalli R., Iyer R., Tetrick S.; Computer Architecture; 2005.
- [2] "TCP onloading for data center servers"; Regnier G., Makineni S., Illikkal I., Iyer R., Minturn D., Huggahalli R., Newell D., Cline L., Foong A.; Computer; November 2004.
- [3] "Design Considerations for Efficient Network Applications With Intel® Multi-core Processor-based Systems on Linux"; Gasparakis J., Waskiewicz P.; Intel White Paper; July 2010.
- [4] "Receive Side Coalescing for Accelerating TCP/IP Processing" Srihari Makineni, Ravi Iyer, Partha Sarangam, Donald Newell, Li Zhao, Ramesh Illikkal, Jaideep Moses; Lecture Notes in Computer Science; 2006.
- [5] "Improving Network Performance in Multi-Core Systems"; Intel White Paper; 2007.
- [6] "Intel 82576 Gigabit Ethernet Controller Datasheet (rev 2.61)"; Intel; 2010.
- [7] "MSI-X – the right way to spread interrupt load"; A. Sandler; <http://www.alexonlinux.com/msi-x-the-right-way-to-spread-interrupt-load>; 2009.
- [8] "Benefits of I/O Acceleration Technology (I/OAT) in Clusters"; Vaidyanathan K., Panda D.K.; Performance Analysis of Systems & Software; 2007.
- [9] "Accelerating High-Speed Networkin with Intel I/O Acceleration Technology"; Intel White Paper; 2006
- [10] "Intel I/O Acceleration Technology"; <http://www.intel.com/go/ioat/>; Intel.
- [11] "Improving Passive Packet Capture: Beyond Device Polling"; Luca Deri; 2004.
- [12] "Exploiting Commodity Multi-core Systems for Network Traffic Analysis"; Luca Deri, Francesco Fusco; 2010.
- [13] "PF_RING"; Luca Deri; http://www.ntop.org/PF_RING.html; 2011.
- [14] "Why TNAPI (Threaded NAPI)?"; <http://www.ntop.org/TNAPI.html>; Luca Deri; 2011.
- [15] "SMP affinity and proper interrupt handling in Linux"; A. Sandler; <http://www.alexonlinux.com/smp-affinity-and-proper-interrupt-handling-in-linux>; 2008.
- [16] "Why interrupt affinity with multiple cores is not such a good thing"; A. Sandler; <http://www.alexonlinux.com/why-interrupt-affinity-with-multiple-cores-is-not-such-a-good-thing>; 2008.
- [17] "Intel Ethernet X520 Server Adapters"; Intel; <http://www.intel.com/Assets/PDF/prodbrief/322217.pdf>; 2008.
- [18] "PF_RING User Guide – Linux High Speed Packet Capture"; Luca Deri; https://svn.ntop.org/svn/ntop/trunk/PF_RING/doc/UsersGuide.pdf; 2011

Ambienti Intelligenti a supporto della Sicurezza Personale

Francesco De Angelis, Roberto Gagliardi, Fausto Marcantoni, Alberto Polzonetti,

Scuola di Scienze e Tecnologie – Università di Camerino
Via Madonna delle Carceri, 9 – 62032 Camerino (MC)
(nome.cognome@unicam.it)

Questo lavoro presenta due realizzazioni prototipali adatte per valutare la rilevanza degli ambienti intelligenti ed il relativo impatto economico ed organizzativo in settori determinanti come la natura e la sicurezza nei luoghi di lavoro.

1. Introduzione

Nell'importante articolo "Il computer per il 21° secolo" [Weiser,91], Mark Weiser ha creato una visione di computer onnipresenti, che sarebbe servita alle persone nella loro vita di tutti i giorni a casa e lavoro, operando invisibilmente in background e liberandoli dalla attività di routine. Mentre i principi fondamentali esposti da Weiser sono stati oggetto di molte ricerche negli ultimi anni [ISHUC,99], le implicazioni relative alla diffusione di tale tecnologia nella società, sono molto meno approfondite. Mai lo sviluppo della tecnologia dell'informazione ha avuto l'esplicito obiettivo di cambiare la società, mentre la "vision" associata ad "ubiquitous computing" propone espressamente di trasformare il mondo e la nostra società.

Sempre di più i progetti di ricerca trovano sviluppo nel mondo reale (sia in prototipo pubblico e nel mercato consumer, come i sistemi di localizzazione o dispositivi elettronici portatili), e il campo di "ubiquitous computing" ha, nel corso degli ultimi anni, iniziato ad entrare nella coscienza di gran parte del pubblico. Recenti discussioni sulla privacy dei consumatori con riguardo all'uso commerciale di tag RFID hanno dimostrato che il pubblico si rende conto a poco a poco che questo continuo sviluppo può avere un impatto a lungo termine sulla vita quotidiana, con conseguenze di vasta portata per i valori etici della società.

Questo lavoro riporta alcuni dei risultati di un progetto di ricerca interdisciplinare denominato "Vivere in un ambiente intelligente", che mirava a immaginare le conseguenze sociali, organizzative ed economiche di "ubiquitous computing", attraverso lo sviluppo (e di esecuzione parziale) di un insieme di scenari prototipo. Questi scenari sono stati scelti nel contesto dell'ambiente e del lavoro come paradigmi in grado di analizzare e valutare le implicazioni tecnologiche di una "u-società." Mettendo insieme le loro diverse prospettive, il progetto volto a fornire alcune delle risposte attese dalla comunità scientifica e

società in generale: quali sono le aree che possono essere influenzati dalla distribuzione su larga scala di tecnologie di ubiquitous computing - in modo positivo e negativo? E quale potrebbe essere la conseguenze a lungo termine?

Per mostrare come applicazioni di vasta portata potrebbero essere, presentiamo due implicazioni prototipali a nostro avviso rilevanti.

Nel quadro di un approccio territoriale integrato che richiede il coinvolgimento del partenariato istituzionale e socio-economico e la promozione del rafforzamento della "governance" delle politiche di sviluppo, è stato individuato il Parco come una delle dimensioni di sviluppo da valorizzare e attore protagonista di sviluppo integrato tra l'ambiente, il turismo, l'agricoltura e la cultura, con la finalità di dare rilevanza al ruolo dei piccoli comuni nel contesto delle realtà e delle economie rurali. I cambiamenti profondi del paradigma economico che possono derivare da un grande spiegamento dello sviluppo di tecnologie informatiche più comuni, i problemi di privacy potenziali associati con esso, e in generale la socio-compatibilità e l'affidabilità delle future applicazioni basate su nuove tecnologie potranno essere oggetto di studio futuro. La sicurezza sul lavoro rappresenta un importante tema nella società odierna e delinea uno dei principali indicatori di sostenibilità dello sviluppo del territorio. La prevenzione e protezione nell'ambito lavorativo costituiscono obiettivi sociali dai molteplici vantaggi a breve e lungo termine, non solo per il dipendente, ma anche per le aziende.

2. Gli ambienti oggetto di indagine

Di seguito sono indicati tutti gli stili da adottare per i paragrafi di primo, secondo e terzo livello. Si sconsiglia di utilizzare ulteriori livelli.

2.1 Monitoraggio Boschivo

Il problema degli incendi boschivi non è ancora risolto e si tratta di un problema di impegno civile, di conoscenza e soprattutto di una nuova cultura che ponga la salvaguardia della natura al primo posto. Come dato rilevante al fine di dimensionare il fenomeno occorre ricordare che, secondo il Corpo Forestale dello Stato [CONECOFOR], nel solo 2009 si sono verificati complessivamente 5.422 incendi boschivi che hanno percorso 73.355 ettari. Per il danno causato dalla diminuzione della produzione di prodotti del sottobosco, si giunge a valutare un costo complessivo annuo di oltre 500 milioni di euro. E' quindi chiaro come il problema debba essere affrontato per consentire una riduzione della spesa e un miglior rapporto con la natura da parte dei cittadini in quanto il fenomeno degli incendi può essere prevenuto con le attuali tecnologie sensoristiche che possono consentire una individuazione immediata dei principi d'incendio.

L'obiettivo del progetto di monitoraggio ambientale AIB (Antincendio Boschivo) che gli autori stanno portando avanti mira alla prevenzione degli incendi con l'obiettivo di salvaguardare l'ambiente e ridurre i costi connessi alla sua gestione in parchi, comunità montane ed enti locali. Da un punto di vista tecnologico, il rischio di incendio boschivo può, infatti, essere determinato tramite l'analisi di precursori di evento noti in letteratura [Di;Joo 2007] come la

temperatura rilevata, l'umidità relativa, la velocità del vento, le precipitazioni previste/avvenute, la natura del terreno e la tipologia di vegetazione coinvolta. L'applicazione di modelli statistici per la previsione di incendi non riescono tuttavia a monitorare in maniera continuativa e “sul campo” l'insorgere di precursori che scatenano l'incendio. Inoltre, una consistente porzione di incendi boschivi è di origine dolosa quindi imprevedibile a priori.

Un'analisi delle dinamiche di sviluppo e propagazione di un incendio fa sì che si possano distinguere in esso diversi elementi necessari e diverse fasi che si susseguono in ordine temporale. Innanzitutto, affinché un incendio di qualsiasi natura possa svilupparsi, sono necessarie tre componenti: (i) combustibile, (ii) comburente (ossigeno), e (iii) calore (temperatura). I tre elementi devono essere contemporaneamente presenti. In assenza anche di un solo elemento, l'incendio non può svilupparsi. Nei casi più comuni di incendio boschivo si conosce come lo sviluppo di un incendio ha luogo e come si possa distinguere tale evento monitorando le grandezze fisiche dei tre componenti necessari per l'insorgere di incendi. Questo monitoraggio è efficace nel rilevare i principi di incendio ma occorre che il rilevamento venga fatto prima che l'incendio inizi la sua vasta propagazione. Infatti, in ogni incendio si distinguono le seguenti fasi (Figura 1):

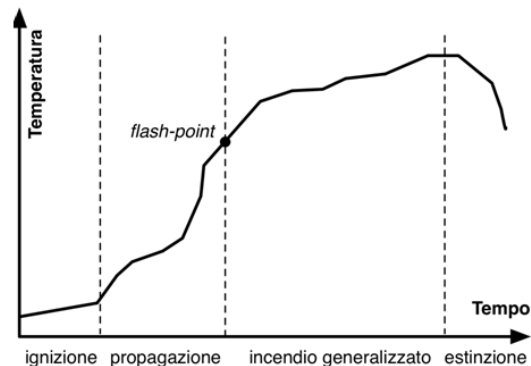


Figura 1 Fasi di propagazione di un incendio

1. Fase di ignizione: fase principale dell'incendio, le sostanze combustibili, iniziano il processo di combustione;
2. Fase di propagazione: caratterizzato da bassa temperatura e scarsa quantità di combustibile coinvolta;
3. Flash-point: da origine al flash-over, brusco innalzamento della temperatura ed aumento massiccio della quantità di materiale che partecipa alla combustione. E' considerato il punto di “non ritorno” nello scoppio di incendi;
4. Incendio generalizzato: gran parte del materiale presente partecipa alla combustione;
5. Estinzione e/o raffreddamento: conclusione della combustione. Il raffreddamento è la fase post-conclusiva dell'incendio e che comporta il raffreddamento della zona interessata.

Nella dinamica di un incendio si suppone di poter intervenire con misure attive atte alla prevenzione prima della fase di flashover (o entro pochi minuti da esso), superata la quale occorre applicare approcci strutturali massicci per lo spegnimento di vaste aree.

Allo scopo di monitorare i parametri ambientali tipici della fase di propagazione si propone l'utilizzo di una Wireless Sensor Network (WSN) in grado di tracciare le variazioni di temperatura, umidità e anidride carbonica presenti negli spazi boschivi unita ad un software di controllo in grado di rilevare potenziali pericoli di incendio e segnalarli alle autorità competenti. La WSN non sarà intrusiva essendo basata su dispositivi mimetizzati per avere un basso impatto sull'ambiente monitorato e consentirà precisi interventi in caso di insorgere di incendi. Per una effettiva applicazione della rete si necessita di: (i) integrare sul nodo wireless la sensoristica preposta alla rilevazione di temperatura, umidità, e anidride carbonica; (ii) sviluppare un nodo a basso consumo alimentato con batterie di lunga durata (almeno due anni); (iii) sia efficiente dal punto di vista del funzionamento (riducendo al minimo i falsi allarmi); (iv) costituire una soluzione a basso costo adottata dagli enti (e certificata dagli organismi preposti).

Il prototipo realizzato include il software necessario e consente il monitoraggio della WSN dispiegata sul territorio consentendo il recupero dei dati e l'invio di comandi ai nodi sensore. I dati rilevati sono poi oggetto di analisi real-time per la generazione di allarmi agli operatori responsabili in grado poi di allertare le autorità competenti con tempi di risposta inferiori ai normali avvistamenti o segnalazioni.

2.2 Sicurezza su i luoghi di lavoro

Il fenomeno degli incidenti sul lavoro ha causato nel 2008, a fronte di oltre 23 milioni di lavoratori, più di 870.000 infortuni di cui 1.120 mortali. I dati relativi al primo trimestre del 2009 riportano un miglioramento rispetto all'anno precedente, ma nonostante questo, ci si trova di fronte ad una situazione sempre preoccupante. Il costo sociale ed economico subito dalle aziende è ingente e destinato a crescere se si analizzano le previsioni per il 2012 (stimato a 51,9 mld. di euro, pari a circa il 3% del PIL italiano). Anche se molto è stato fatto e si sta facendo in termini di formazione, normativa, e adozione di ulteriori certificazioni su base volontaria come OHSAS, occorre ancora investire nel settore e individuare soluzioni adeguate per garantire la sicurezza delle persone coinvolte nelle attività aziendali e ridurre notevolmente i costi associati all'infortunistica.

L'introduzione di ambienti intelligenti nelle aree lavorative, che abbiamo denominato smart workplaces, ha l'obiettivo di realizzare ambienti di lavoro sicuri poiché proattivi rispetto al verificarsi di situazioni di pericolo. Diversamente dalle attuali soluzioni, che nella maggior parte dei casi sono in grado di segnalare semplicemente il verificarsi di una situazione di pericolo dopo che questa si è verificata, il progetto vuole prevenire tali situazioni predicendo e gestendo potenziali rischi, soprattutto nei casi in cui questi sono dovuti al verificarsi in modo concomitante di situazioni diverse, che singolarmente non sarebbero pericolose.

Il concetto di sicurezza viene elevato verso la proattività in quanto il sistema che si intende realizzare ha la finalità di rendere noti eventuali rischi prima che possibili incidenti si verifichino. L'idea alla base del progetto è lo sviluppo di un framework metodologico e tecnologico (i) per stabilire le strategie di progettazione di smart environment aziendali, (ii) per sviluppare tecniche di predizione in grado di individuare pattern ambientali e comportamentali potenzialmente rischiosi, segnalandoli e avviando un piano di risposta adeguato, e (iii) per apprendere nuovi pattern implementando una sorta di "sistema immunitario" aziendale in grado di rilevare anticipatamente fattori di rischio e segnalarli per supportare la sicurezza degli individui.

Il progetto adotta il paradigma degli smart environment [Poslad S,2009], ossia luoghi fisici dotati di sensoristica integrata nell'ambiente e interconnessa in rete, per compiere un continuo monitoraggio dell'ambiente lavorativo e una costante prevenzione degli incidenti, finalizzato all'individuazione dei possibili fattori di rischio. Nella visione tipica degli smart environment i dispositivi fisici lavorano a supporto degli individui nelle loro attività quotidiane nella maniera il più naturale possibile. Gli oggetti di uso comune sono dotati di sensori e connessioni in modo da poter dialogare tra loro e con l'ambiente circostante. Tale ambiente si definisce come: "a physical world that is richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives, and connected through a continuous network" [Cook,Sajal,2005]. Generalmente uno smart environment (i) include dispositivi integrati nell'ambiente, (ii) fornisce informazioni context-aware riconoscendo le persone ed il contesto in cui si trovano, (iii) può essere personalizzato in base alle necessità delle persone e dell'ambiente stesso, (iv) è adattivo in quanto cambia in base alle situazioni, ed (v) è in grado di anticipare i bisogni delle persone.

3.Caratteristiche tecniche

3.1 Specifiche per reti di sensori

La WSN prevede la realizzazione di una rete decentralizzata tra nodi sensore, che sia economica, adattabile e resistente, poiché ogni nodo deve solamente trasmettere un segnale al massimo fino al nodo successivo. I nodi fungono da ripetitori per trasmettere il segnale dai nodi più vicini agli altri fino a raggiungere speciali nodi di raccordo con una rete wireless in cui si troverà il server dotato di applicativo di gestione. I nodi sensore sono stati implementati utilizzando le soluzioni tecnologiche su protocollo IEEE 802.15.4/ZigBee associate a sensori in grado di rilevare le grandezze fisiche di temperatura, umidità, e anidride carbonica. Parte della realizzazione del prototipo è stata focalizzata sul contenimento dei consumi attestando la potenza richiesta dall'apparato sui 90 uA che garantiscono una durata del nodo sensore di circa due anni con normali batterie AA. All'interno della rete, particolari nodi fungono da router nei confronti dei nodi sensore. Questi ultimi sono stati implementati con la stessa tecnologia ma necessitano di alimentazione costante che proviene da un pannello solare appositamente predisposto. Infine, tra i nodi

router uno o più nodi fungono da raccordo verso una rete wireless che connette la WSN alla rete Internet. La connessione su dorsali esistenti distribuite nel territorio (come quelle fornite dalla rete GARR, dalle comunità montane, etc.) rappresenta in questo ambito una componente necessaria al corretto ed efficace funzionamento del sistema proposto che necessita di essere integrato in importanti reti territoriali.

Il software applicativo di supporto alla rete prevede la realizzazione del sistema di gestione dei singoli nodi con la possibilità sia di recuperare i dati da essi, sia di modificare parametri di funzionamento (es: intervallo di recupero dati). L'innovatività della soluzione proposta a livello software risiede nell'introduzione di algoritmi di apprendimento che possano "capire" le caratteristiche dell'insorgere dell'incendio sia tramite apposite prove in ambiente sicuro e controllato, sia tramite l'analisi di episodi che posso avvenire durante il normale funzionamento della rete. [Di;Joo 2007]]

Ogni sensore sarà georeferenziato all'atto dell'istallazione che avverrà all'interno del bosco direttamente su alberi ad una altezza da terra di circa 4 metri. Particolare attenzione andrà posta, per il dispiegamento della WSN, sul corretto posizionamento geografico dei sensori. Questo può essere previsto tramite software cartografico identificando in linea di massima i punti necessari alla sorveglianza in relazione alla conformazione del terreno e alla sua morfologia.

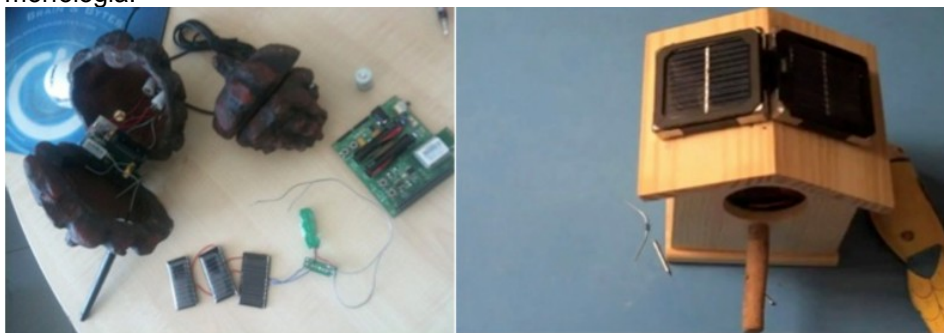


Figura 2 - Immagine della “pigna” nodo sensore e della casetta per router e accentratore

3.2 Il primo prototipo: lo Smart Helmet

Al fine di raggiungere l'obiettivo prefissato si evidenzia l'importanza di monitorare il comportamento del singolo individuo in relazione all'ambiente che lo circonda e ai dispositivi di protezione individuale (DPI) necessari alla sicurezza personale nei vari ambienti in cui un individuo può trovarsi. Il primo prototipo, sviluppato nel contesto degli smart workplaces e presentato in questo paper, riguarda un sistema di monitoraggio delle aree “a rischio” di una impresa con annessa gestione dei DPI dei singoli individui e monitoraggio real-time delle policy di sicurezza stabilite.

Il prototipo, denominato Smart Helmet, ha lo scopo di rilevare e monitorare la presenza e gli spostamenti del personale tra le differenti aree aziendali. Il sistema permette di definire aree in base alle necessità

Ambienti Intelligenti a supporto della Sicurezza Personale dell'azienda, ai differenti livelli di rischio che possono presentare, alla tipologia del lavoro in esse svolto, e ai DPI richiesti. Il software permette una definizione delle aree di lavoro presenti nell'azienda con policy di sicurezza ad esse associate.

Nel caso in esame, la gestione dei DPI prevede sia la presenza di uno o più lettori RFID connessi in rete e posti in corrispondenza degli ingressi alle aree sensibili, sia la disponibilità di tag RFID ad identificare i singoli dispositivi e le persone che li indossano. In questo caso, si considerano luoghi di lavoro separati e singolarmente monitorati da tali apparecchiature in cui ad ogni ingresso o uscita dalle aree protette i lettori sono in grado di rilevare i movimenti degli individui e dei DPI ad essi associati. In questo modo è possibile avere costantemente una chiara visione di chi si trova in ogni area e verificare la presenza di eventuali punti deboli nella sicurezza delle persone in quanto vengono evidenziati gli individui non regolarmente protetti dai richiesti e necessari DPI. Il sistema consente un controllo real-time di ogni evento di ingresso/uscita ed è in grado di riconoscere la violazione delle policy di sicurezza aziendale scelte permettendo ai responsabili della sicurezza di intervenire in tempo in caso di situazioni a rischio.

Il sistema è quindi in grado di evitare che banali dimenticanze o incuranze si tramutino in situazioni di effettivo pericolo per la persona. In ogni caso il sistema tiene traccia di ogni accesso/uscita riguardanti ogni singola area e li considera come eventi avvenuti e di cui tener traccia. Per ogni evento vengono riportate informazioni dettagliate circa la data e l'ora in cui si sono verificati, la persona che li ha generati, il tempo di permanenza all'interno di un'area, i Dispositivi di Protezione Individuale rilevati e, ovviamente, il risultato del controllo dei permessi e dei requisiti sopra citati. Si tratta di un controllo accessi avanzato che non si limita alla verifica del permesso di entrata per una persona in una determinata area, ma si preoccupa di accertare che la persona abbia con sé i giusti dispositivi di sicurezza e che questi gli appartengano.

Smart Helmet è composto da diverse componenti (Figura 1). Una prima componente è rappresentata dall'hardware che include le apparecchiature utilizzare per il monitoraggio degli ingressi, lettori RFID e tag; mentre il componente software principale si occupa della gestione dei dati ed è strettamente collegato allo strumento di Complex Event Processing [Luckham D 2002] per il controllo real-time degli eventi.

Il software è dotato di una complessa interfaccia di gestione che fornisce funzionalità quali: l'amministrazione delle tre entità chiave: personale, DPI e aree; l'associazione tra personale e DPI per stabilire le relazioni di proprietà; la definizione dei requisiti per le singole zone, ovvero quali e quanti sono i Dispositivi di Protezione Individuale richiesti per poter effettuare l'accesso; la definizione delle policy di sicurezza per stabilire chi può accedere alle varie aree; la disposizione coerente delle differenti zone all'interno dello spazio aziendale (ad esempio adiacenze e inclusioni); l'organizzazione del personale in gruppi, ai quali possono essere associati dei permessi ereditati automaticamente dai membri; ed infine uno strumento di reportistica per l'esportazione dei dati.

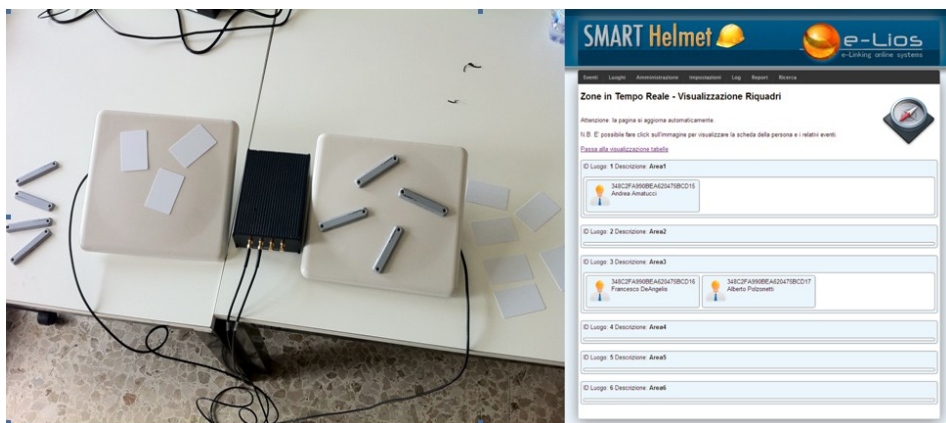


Figura 3: Vista dell'hardware utilizzato e screenshot di parte del software implementato

4. Conclusioni

Il sistema di monitoraggio proposto si caratterizza per un basso impatto sull'ambiente con un elevato ritorno in termini di accuratezza delle rilevazioni effettuate utili non solo per la prevenzione antincendi ma anche per statistiche di più ampio respiro sulla flora dei parchi italiani. Per una effettiva validazione della tecnologia proposta e del sistema software, un prototipo del sistema è in corso di deploy in alcune aree boschive della Comunità Montana dell'Esino-Frasassi (sita a Fabriano (AN)). In questo caso la connessione verso la rete Internet è garantita dalla rete territoriale "Marche Way". Il bosco scelto per il deploy è a maggioranza di conifere e a tal proposito la sperimentazione prevede l'utilizzo di contenitori mimetici a forma di pigna per un impatto visivo e ambientale minimo. I router, che necessitano di pannello solare, sono stati realizzati come cassette per gli uccelli di piccole dimensioni come evidenziato nella Figura 2. Gli autori vogliono estendere la rilevazione ambientale implementando il riconoscimento e l'apprendimento di "pattern ambientali" in maniera tale da adattare il sistema ad un sempre maggior numero di casi e facilitarne l'adozione ad ampio spettro negli enti locali.

L'idea di un ambiente di lavoro intelligente orientato alla sicurezza dei lavoratori è senz'altro ambizioso e articolato. Il caso presentato con il prototipo Smart Helmet è soltanto uno dei differenti moduli che compongono l'idea di Smart Workplaces proposta. Si parla di identificazione e controllo dei requisiti di accesso in arre protette per quanto riguarda i Dispositivi di Sicurezza Individuale (DPI). E', tuttavia, importante tenere presenti altri aspetti della gestione della sicurezza del personale in un ambiente di lavoro. Ne è un esempio l'integrazione di sensori per la rilevazione di temperatura, pressione, elettricità, sostanze chimiche, gas, eccetera. Sono molti, infatti, i fattori che si dovranno tenere in considerazione per ottenere un quadro completo di prevenzione dei rischi che possono compromettere la sicurezza delle persone.

Lo sforzo nella direzione della sicurezza va verso ambienti che presentano strumenti, materiali e macchinari che necessitano di particolare attenzione. Si aprono, così, numerosi scenari diversi e possibili applicazioni in differenti ambienti lavorativi da affrontare con la tecnologia a radiofrequenza e sensori appropriata alla circostanza. Gli autori vogliono estendere il prototipo realizzato ed utilizzare il trend di crescita tecnologia per introdurre innovazioni con il giusto livello di intelligenza e proattività negli ambienti di lavoro della vita quotidiana volte a salvaguardare la vita, la salute e il benessere degli individui.

5. Bibliografia

CONECOFOR (CONTrolli ECOsistemi FORestali), Corpo Forestale dello Stato,
<http://www3.corpoforestale.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/94>

Cook, Diane; Das, Sajal, "Smart Environments: Technology, Protocols and Applications" Wiley-Interscience, 2005, ISBN 0-471-54448-5

Di Ma; Joo Er Meng; , "A survey of machine learning in Wireless Sensor networks From networking and application perspectives," Information, Communications & Signal Processing, 2007 6th International Conference on , vol., no., pp.1-5, 10-13 Dec. 2007

International Symposium on Handheld and Ubiquitous Computing (HUC 99), www.teco.edu/huc/.

Luckham D., "The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems", Addison-Wesley, 2002, ISBN: 0-201-72789-7

Poslad S., "Ubiquitous Computing: Smart Devices, Environments and Interactions", Wiley, 2009, ISBN: 0-470-03560-9

Weiser M., "The Computer for the 21st Century", *Scientific American*, 265(3):94–104, September 1991.

An Albanian e-Health infrastructure

Gianluca Zangara ¹, Angelo Marcheggiani ², Francesca Valentino ¹, Francesco Di Blasi ¹

¹ Consiglio Nazionale delle Ricerche – Istituto di Biologia ed Immunologia Molecolare
“Alberto Monro”

Via Ugo La Malfa, 153 – 90146 - Palermo
zangara@ibim.cnr.it, valentino@ibim.cnr.it, diblasi@ibim.cnr.it

² Azienda Sanitaria Regione Molise
Via Ugo Petrella, 1 – 86100 - Campobasso
angelomarcheggiani@libero.it

The health situation in Western Balkan countries, especially Albania, is poor but improving, as the rest of the Balkan area. The ‘Welfare and Health Cooperation in the Balkans’ project is focused on the health sector in order to optimize resources and increase the quality of Public Health Service. The system we are developing consists of a communication environment based on state-of-the-art communication technologies for shared management of knowledge. As a result, a network will be created to exchange best practices, promote innovative learning and have a second opinion from expert staff through a telemedicine system.

1. Introduction

In the last few years, Albania has undergone a profound transformation, characterized by economic growth and the typical epidemiological changes that accompany any transition phase such as the increasing of pathologic diseases and the decreasing of infection diseases [Institute of Statistics and Institute of Public Health, 2010].

Such economic progress resulted in an inadequate health care system, as the latter also had to adjust to many changes, namely the increasing age of population as well as changes in their lifestyle and, consequently, an increase in chronic diseases and a decrease in infectious diseases.

Through the Lisbon Strategy, the European Union and its member states planned several actions to deal with low productivity and poor social conditions in both European and neighbouring territories. Some actions focus on economic development, whilst some other on welfare and health care.

The ‘Welfare and Health Cooperation in the Balkans’ project, a part of a European strategic intervention in Western Balkans, is aimed at providing institutional support and technical assistance to improve Albanian Health Services. The project is divided into multiple sub-technical units defined as work-packages.

Public and private Italian partners, such as regional governments and non-governmental organisations, participate in specific work-packages of the project.

The availability of cheap computer equipment and modern telecommunication technologies enabled us to implement large-scale telemedicine systems in this area where a new health care approach is developing.

Actually we have deployed the network infrastructure and set-up serious software modules that Albanian and Italian partners are using. In this phase we are collecting usage data to enhance software modules and add other appliances to the infrastructure and also we are planning to add other network-enabled medical devices.

To identify the health care needs of population, the Catholic University of Campobasso carried out an epidemiological analysis choosing the population of Shkodër. During the preliminary phase of the project, questionnaires [Gianfagna et al., 2010] were administered to health care personnel. More specifically, the questionnaires were administered to a group of 30 general practitioners and nurses in the age bracket 27-65, geographically distributed between urban and rural areas. Results of questionnaires showed that both groups of health care professionals had specific learning needs in several medical subjects. Moreover, they generally require to refer to a specialist consultation if they have to examine patients having chronic cardiovascular diseases. According to the results obtained from the questionnaires, the system developed was specifically focused on topics such as cardiology, urology and diabetes.

The geographical area in which the project was developed has been chosen in the city of Shkodër. Subsequently, in order to design a solution that could be enlarged to a metropolitan area or a region, have been chosen two nearby small towns – about 15km away from downtown. Currently, the country's communication infrastructure is underdeveloped and not up-to-date. Hence, the choice to include two rural towns was made to facilitate communication between these places and the city, which is still considered as the reference point in health care when it comes to severe conditions.

In this area have been identified: the Regional Hospital of Shkodër, four family planning clinic, a Mental Health Institute and two health centers - one for each rural municipality.

2. Network infrastructure

To put in communication health centers an Intranet was implemented. Each health center involved in project has been equipped with a broadband Internet access and a router to allow access to the internal network.

Internet access was delivered in three different ways: in Shkodër Regional Hospital via a optic fiber broadband access, in four family planning clinic and Mental Health Institute (located within Shkoder city) via ADSL technology, while in rural countries that are located in a digital divide area we provided internet access via Wi-Fi connection.

Virtual Private Network technology was used to make the Intranet, using IP Security (IPSec) standard to protect communication flows between two hosts or routers. IPSec is the most widely used standard to deploy VPNs and it is supported from a large number of devices and software clients, so the architecture created is easily scalable.

Each health care center was also equipped with a multimedia workstation with a multifunctional printer and a Voice Over IP Phone. VOIP technology advantages of free telephone call among the same network and low telephone call costs versus national, international and mobile phones by using Internet connection. Having all flat rate broadband Internet access, health care structures can reduce phone call costs. We also experienced that a hardphone is more friendly than a softphone (such as Skype or some other software) for a user with poor I.T. expertise.

3. The software

One of the objectives of the work-package is to create an environment for communication, by means of state-of-the-art communication and networking technologies for shared management of knowledge [Di Blasi et al., 2010]. As a result, a network to exchange knowledge and best practices will be set up.

The platform enables effective cooperation between distant project participants, besides laying the foundation to establish a specialized Knowledge base on which training courses and/or continuing medical education can also be provided.

Key features of the system are:

- Easy data entry by operators;
- Categorization of data entered by tagging;
- Consistency of search results with respect to the parameters;
- Establishment of a groupware between users;
- Creation of a public web site where to publish information of public interest.

Knowledge base is one of the system components that allow users to add useful content for staff training. Moreover, this section contains web pages that patients may browse to find information to treat diseases and guidelines about healthy lifestyles.

Wiki is another section where operators can keep an archive of best practices on various medical subjects.

A system of resource management allows operators from peripheral health centres to reserve resources, i.e. diagnostic and therapeutic equipment, which are usually available only in hospitals located in urban areas.

4. Telemedicine application

In one of the health center of Shkodër was installed an ultrasound scanner for gynecological applications.

This model of ultrasound scanner is equipped with Ethernet connection for interfacing with other storage devices based on DICOM (Digital Imaging and Communication in Medicine) [National Electrical Manufacturers Association, 2009] standard. On a workstation placed in the office we have installed a DICOM server, i.e. a software able to receive images from devices that support this technology. This software has a database that catalogues all images for a patient. This information is entered by healthcare staff on ultrasound scanner before the visit.

Accessing the workstation over the Internet, using a remote control software, you can allow a specialized physician situated many kilometers away, to see the pictures and give an opinion on possible diseases detected.

This method of intervention for telemedicine is called second opinion, and allows you to have an expert opinion from a distance using digital communication infrastructures.

5. Used technologies

To create the VPN, Cisco ASA devices where used. To join Italian partners, software client IPSec enabled were setup.

The communication environment that uses Web 2.0 technologies to access the database is also available for access by 'mobile' devices, such as PDAs and smartphones.

Health care facilities in rural municipalities were provided with multimedia workstations equipped with a printer so as to give each health care centre the opportunity to access the software. Facilities in Shkodër had already been equipped with workstations to access the Intranet. Furthermore, to ensure mobile access, we provided two HTC 7 Mozart smartphones to two GP of rural country. Smartphones are equipped with Windows Mobile 7 operating system. These devices provide access to the Intranet via Wi-Fi connection, available in all health care facilities, and available in almost all the surrounding areas through UMTS / GPRS.

Moreover, the availability of broadband access, such as Wi-Fi, UMTS or 3GPP, made it possible to have a faster connection to the server.

The system was created using only open source software. The server is equipped with the Linux operating system Debian 5.0, Apache 2 web server with support for PHP5, and the MySQL 5.0 database server.

'KnowledgeTree - Community version 3.7' was used as a document repository. Although software producer say that this version is not designed for production use, we strongly tested and setup the system to be reliable.

The latter allows users to store documents by making them immutable as required by law. It is also useful for producing documents thanks to a workflow system that helps engage multiple users in document editing.

The collaboration system between health care members is based on the eGroupWare 1.6 software, which is available in a GPL version too. It was customized in order to make useful content for the staff immediately accessible. It provides a shared address book, bookmarks that record a lot of institutional sites, and a resource management system to reserve and coordinate the use of resources such as medical equipment. Moreover, a website accessible to all Internet users was also set up, which contains information about treatment and diagnosis of diabetes.

Ultrasound scanner installed in health center of Shkodër is a General Electric Logiq C5 equipped with convex, linear and phased array sector probes.

To acquire and store DICOM images from Logiq C5, Conquest DICOM Server was installed. It was available with freeware license, both for Windows and Linux platform. To access Windows workstation and consult DICOM images stores for patient diagnosis, Ultra VNC with encryption plugin was installed to enforce the weak security of the VNC protocol. The system is all behind firewall to ensure strong protection of sensible data.

6. Conclusion

After a six month trial period, the system's impact of use will be evaluated through the administration of questionnaires. Results will be compared with the previous ones obtained.

During the design phase of the project, it was envisaged to implement the same or similar systems in other countries. Hence, many planning choices were taken with the aim of making the system scalable both in terms of users and geographic distribution.

References

Zangara G., Valentino F., Piceno A., Di Blasi F., A telemedicine environment for Balkans. IADIS International Conference eHealth 2011. Rome, Italy, 2011.

Institute of Statistics, Institute of Public Health, Albania Demographic and Health Survey (ADHS) 2008-09, Tirana, Albania, 2010.

National Electrical Manufacturers Association. Digital Imaging and Communications in Medicine (DICOM). Virginia, USA, 2009.

Di Blasi F., Valentino F., Zangara G., Le nuove frontiere dell'e-health: la situazione attuale e le prospettive future, Il progetto integrato Welfare and Health Cooperation in Balkans a supporto delle politiche sanitarie e di welfare nell'area adriatica, Campobasso, Italy, 2010.

Gianfagna F., Zhabaku S., De Gaetano G., Report Analisi Epidemiologica, Campobasso, Italy, 2010.

An Ontology-based Approach to Open Innovation

Claudia Diamantini, Michele Missikoff, Domenico Potena
Dipartimento di Ingegneria dell'Informazione
Università Politecnica delle Marche - via Brecce Bianche, 60131 Ancona, Italy
{diamantini, missikoff, potena}@diiga.univpm.it

An ontology is conceived to capture and manage the conceptual knowledge of a given domain in a structured and rigorous fashion. In its essence, innovation is commonly perceived as a process, i.e., a set of activities that are able to intervene and change for better one or more business elements (e.g., products or business processes); therefore the knowledge concerning innovation is not easily represented in a structured way, it is better represented as a workflow. However, an innovation process is different from a usual business process. Innovation belongs to the category of creative endeavours that exhibit loosely structured processes, difficult to be modelled and managed with the 'usual' methods. For these reasons, the present paper proposes a process reification method and develops the concept of innovation as an evolving entity. Such an entity, referred to as Innogotchi, is structured and managed according to an ontology representation method, achieved with the Topic Maps standard that exhibit a number of convenient features for our purpose.

1. Introduction

Time passes but the attention to innovation is not showing any sign of decline, probably the opposite is true. But innovation is an elusive term that is often used in an imprecise way, or seizing only part of its substance. In fact, if we consider largely accepted definitions, it is easy to see that they are able to capture only part of the essence of the term innovation. For instance, consider the initial definition given by Wikipedia that reports: "innovation generally refers to the creation of better or more effective products, processes, technologies, or ideas that are accepted by markets, governments, and society." It is very schematic. Reading further, the text offers other perspectives. For instance: "innovation is the catalyst to growth", is the definition from an economic point of view; then, from an organizational point of view: "innovation may be linked to positive changes in efficiency, productivity, quality, competitiveness, market share, and others."

Another important aspect, in seizing the essence of innovation, is represent-

Congresso Nazionale AICA 2011

ed by the 'source of innovation', i.e., where the innovation originates. According to the literature, if we focus on technology-based business innovation, there are typically three modes to push forward innovation: (i) market pull, when it originates to respond to a specific market need; (ii) technology push, when new technological solutions offer the opportunity to improve a given business; (iii) co-creation, when innovation stems from a combined effort of the two above. Another dimension is represented by the scope and impact of the innovation. Where its influence is it impacting? Among the most important target we have: production processes and means, products (goods, services, or both), markets and marketing strategies, organizations and business models, to name a few.

In its essence, innovation is commonly perceived as a process, i.e., a set of (partially ordered) activities that are able to intervene and change (for better) one or more business elements. But an innovation process is different from a usual business process we find in an enterprise. The latter is (supposedly) well defined and specifies to a good level of detail the activities, their sequencing, who is expected to perform them, what are the required and committed resources, etc. Innovation belongs to the category of creative endeavours that exhibit partially structured processes (if we can still call it a process), difficult to be modelled and managed with the 'usual' methods and tools (i.e., Business Process representation methods, such as BPMN, EPC, and related tools.) This is because innovation belongs to the categories of 'wicked problems' [Buchanan, 1992].

For the above reasons, in proposing a method aimed at supporting business innovation, we intend to abandon the traditional process-oriented approach to adopt an entity-centered perspective [Van der Aalst, 1999]. In our approach, innovation is seen as a complex structure that progressively evolves, growing from an initial seed, e.g., a creative intuition, through successive stages, such as proof of concept, prototype, until it reaches the final stage of an industrial-strength product (or process, or marketing strategy, etc.), where each stage can be more or less defined. At each stage, the innovation process requires the acquisition of new knowledge in order to proceed forward and reach a new stage. But, given a stage, what are the successive stages is not deterministically specified. Therefore, we will have a fuzzy membership function to determine what is the stage that an innovation structure has reached and, when moving ahead, the next stage is non-deterministically reached.

2. A structural approach to innovation modelling

As anticipated, innovation concerns the entire process of moving new and valuable ideas into the enterprise, having an impact (direct, e.g., on the commercialised products, or indirect, e.g., on the production processes) on the marketplace [Fitzgerald et al, 2010]. Here we introduce a method that adopts an entity-oriented approach able to specify the structural nature of an innovation and, starting from the latter, its operational nature according to an underlying computational model. To represent the ‘innovation entity’ we adopt a Topic Maps modelling method.

2.1 Reification of an Innovation Process

A business innovation process consists in a (non deterministic) sequence of activities aimed at gathering and organizing a rich knowledge structure. In our proposal we will concentrate on the knowledge structure that is progressively built rather than in the process that is necessary to activate in order to achieve it.

The knowledge about a new business innovation case typically starts to be created moving from a creative idea, i.e., a preliminary innovation embryo. Then it is necessary to activate further investigations to understand the feasibility, the cost and benefits, if the market (and which market) is ready for absorbing the innovative product, etc. There are several studies in the literature that propose different variants of business innovation processes. For instance, [Fitzgerald et al, 2010] proposes the following steps: (1) discovery, (2) invention (3) development, (4) product, (5) market, and (6) profit. Another proposal, due to SAP, includes the following steps: (1) invent, (2) define, (3) develop, (4) deploy, (5) optimise. Both proposals are very sketchy specifications of a business innovation process; getting more specific, we discover that an innovation process is much more intricate, far from exhibiting the linearity reported above. However, despite their differences, they all exhibit a common trait: all the reported activities are tightly connected to the acquisition, understanding, organization, integration, enrichment, distribution, validation of a large amount of knowledge, typically organised in a predefined set of documents.

Our objective is to make such knowledge structures explicit. In our perspective, innovation is seen as a complex object that starts rather minimal (e.g., representing a creative idea or an intuition, a business opportunity or need) and then needs to progressively grow acquiring additional elements by means of specific studies, data acquisition, engineering investigation, etc. In essence, the ‘innovation entity’ is like a sort of artificial creature that once conceived needs to

Congresso Nazionale AICA 2011

be fed with knowledge to progressively grow until it reaches the maturity, i.e., an industrial strength. To better fix the ideas, we wish to use the metaphor of *Tamagotchi*, that is a Japanese term referring to a digital creature and an electronic game that requires one to take care of the creature, feeding it to make it grow, evolving to more mature forms. Hence, we adopt the name *Innogotchi* for our case that concerns business innovation.

We will consider the *Innogotchi* as a knowledge entity that, in order to grow and successfully develop in all its parts, needs to be fed, and the food is essentially knowledge. As an example, assume we are considering a fragment of a product innovation process, where the knowledge to be 'fed' to our *Innogotchi* is represented by the following list:

- **Innovative idea**, that includes an idea description.
- **Feasibility study**, providing a preliminary account of objectives, technical and financial viability, IPR, etc.
- **Business model**, defining the economic sustainability and the manner by which the novel product will be delivered to customers.
- **Technical specification**, structural layout and components (Bill of Materials), materials, building methods, standards, etc.
- **Business plan**, including a market analysis, cost-benefit analysis, financial viability (including break-even analysis), critical risk factors (including SWOT and PEST analysis), an exit strategy.
- **Marketing and sales strategy**, including selling modes, pricing, warranties, after-sale services.
- **Production planning**, including manufacturing, testing and QA methods, sourcing, logistics and delivery strategies.

As an example of non-determinism in the innovation process, a business model is typically defined during business plan definition, but often it is taken into consideration at earlier phases, for instance when a feasibility study is performed. In turn, a (preliminary) feasibility study can be started as soon as the innovative idea is sketched, or it can be delayed until the technical specification is given.

The above points can be structured according to the knowledge they need to gather and, in addition, the specific methods and procedures that need to be adopted to collect it; all this is reported in the *Innogotchi* structure. As anticipated, the above knowledge cannot be acquired in any possible order since there will be dependencies and priorities: also those represent knowledge that must be fed to the *Innogotchi*.

2.2 Innovation as a Topic Map

Our objective is to capture, model, and organise the knowledge necessary to develop an innovative idea until it reaches the maturity and completeness of an industrial strength stage (be it a product or a process.) There is a great variety of knowledge modelling methods, graphical, such as E-R or UML, rather than in linear form, such as Prolog, RDF, or OWL, that may be adopted. But here we need to create a rich knowledge repository that includes the structural representation but also constraints, and procedures to acquire the needed knowledge, with also the possibility to easily connect instances and information resources, like documents, URLs, images, and so forth. For our purpose, a suitable approach is represented by the Topic Maps [Maicher e Park, 2005] that, in addition to a powerful modelling paradigm, offers a good choice of management tools.

Topic Maps is a standard for the representation and interchange of knowledge. It is based on a sound mathematical basis (Common Logic) and has been defined as an ISO standard (ISO/IEC 13250:2003).

Basic elements of a Topic Map are:

- **Topics**, representing any subject of discourse, from people, countries, and organizations to software modules, individual files, and events. (E.g., The graduate student Aldo.);
- **Occurrences**, representing information resources relevant to a particular topic. (E.g., Aldo's Facebook page.)
- **Associations**, representation of a relationship between one or more topics, (E.g. Aldo is enrolled at Stanford University).

Topics, occurrences and associations can be categorized according to their kind. (E.g. the topic "Aldo" belongs to the category "graduate student"). Categories are named types in the Topic Maps notation. In practice, the set of topic, occurrence and association types form the conceptual or ontological level of the topic map, typically organised according to a super-type / sub-type relationship. Then, an instance-class relationship is established between a topic and its type. The capability of a Topic Map to simultaneously represent concepts, instances as well as information resources related to instances, within a unified frame, is one of the most powerful features of its knowledge representation paradigm.

Topic Maps can be associated to graphical representations where topics and information resources are represented as nodes while occurrences and associations are represented as arcs (see Fig.1 for a simple example).

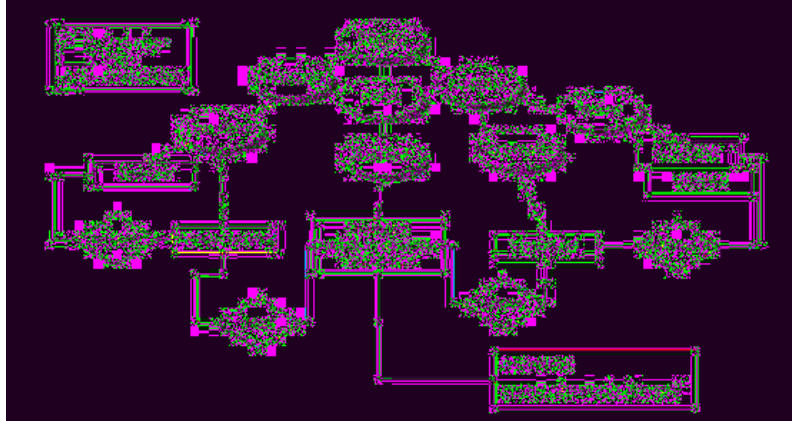


Fig.1 - An example of Topic Map for the Chernobyl Disaster [Sowa, 2006].

3. Innogotchi: an evolving knowledge structure for innovation

Here we elaborate the knowledge structure, sketchy reported in Section 2.1 in the form of a bullet list, as a fragment of a Topic Map. It is conceived as a core structure that progressively evolves, capturing the knowledge gathered in the different stages of development of the initial idea. The proposed method consists in the *Innogotchi* knowledge structure that needs to be progressively fulfilled, by providing the required knowledge. Each component, besides the structure (e.g., the information that needs to be provided) is also associated to one or more procedures (i.e., subprocesses) that can be activated in order to collect and/or produce the required knowledge. Such procedures are very diverse, depending on the kind of knowledge, and may require skilled people (e.g., market analysts), specific tools (e.g., PEST analysis tool), and/or the access to dedicated information resources (e.g., web sites with figures on the market of interest). Furthermore, dependencies among *Innogotchi* sections are also reported. This structure is represented in Fig.2, where the reported map has been drawn by using *Ontopia* (<http://www.ontopia.net>), one of the most widespread Topic Maps management tools.

An Ontology-based Approach to Open Innovation

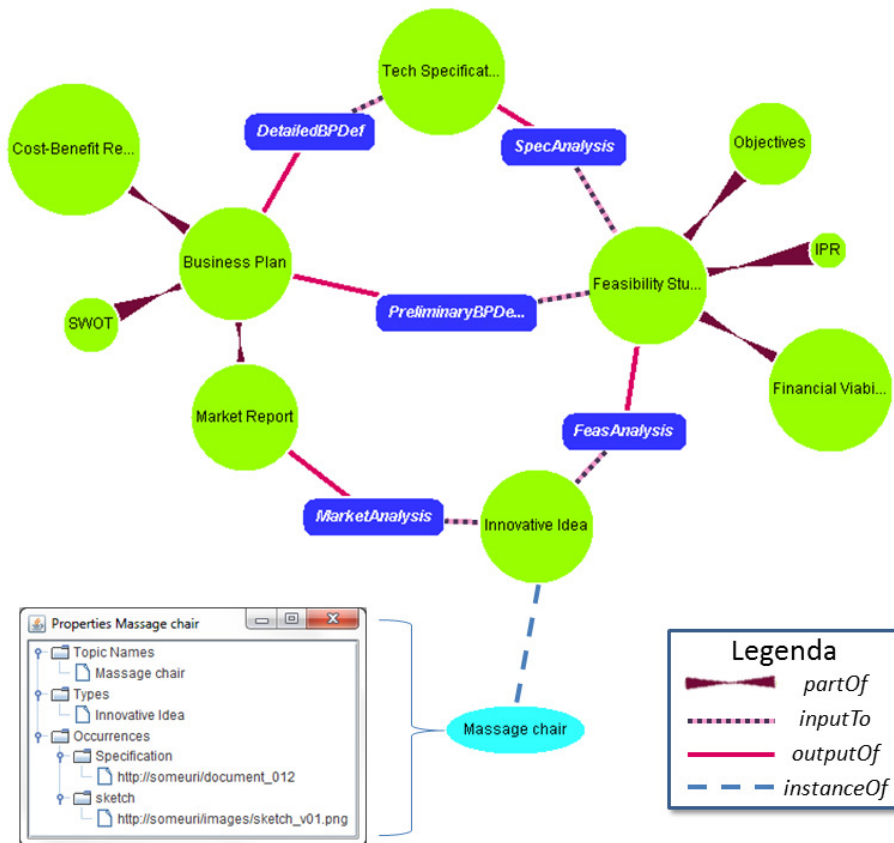


Fig.2 - A fragment of the Topic Map for Innovation Knowledge

In order to give an example, here only the ontological level is rendered, and a single instance (the innovative idea "Message chair") and its properties are reported as an example. In particular, the broken line represents an *instanceOf* relationship between a topic and its topic type, hence de facto this kind of lines divides the conceptual level of the Topic Map, from the level of instances. Also note the panel, which describes the properties of the instance, and in particular a pair of occurrences: the URI of the document describing the idea specification and the URI of the sketch of the new chair.

Congresso Nazionale AICA 2011

At the ontological level, green circles represent knowledge types, whereas blue rounded boxes represent activity types. Arc shapes represent different association types. Bowtie arcs represent *partOf* relationships among knowledge types, while straight and dotted linear arcs respectively represent the *output_of* and *input_to* associations between a knowledge type and an activity. The definition of these associations is a crucial design choice since it allows to define constraints in the growth of the *Innogotchi*, showing what knowledge is needed to produce another. Although a similar dependency constraint could be represented by directly linking two different knowledge types, the explicit use of activities serves two goals: (i) it allows a richer description of activity features, for instance the fact that two or more knowledge types are needed to perform the activity and (ii) it allows to recognize and represent the existence of different instances of same activity type, that is of different modes to move to the next stage, with different efficiency and effectiveness levels. This represents also the non-determinism inherent in the innovation process and, besides being a support for the incremental evolution of a specific *Innogotchi*, it also provides different “ways to innovation”.

4. Conclusion

In this paper we presented a few preliminary ideas aimed at addressing the problem of business innovation from a declarative knowledge representation perspective, instead of the more traditional Business Process approach. Innovation is widely recognised as a complex loosely defined process and it shares the difficulties typical of the wicked problems: when you start you are unable to foresee how the process will evolve in its later stages and if eventually it will successfully end. Furthermore, each innovation endeavour is different from its predecessors, although it is possible to identify a number of ‘knowledge invariants’, i.e., elements that need to be acquired in order for the focused innovation to progress and mature towards an industrial and/or commercial result. In this paper we proposed to see an innovation as an entity, referred to as *Innogotchi*, which is fed with the knowledge necessary to evolve along a successful innovation path. So our specification is concentrated on the structural aspects of the *Innogotchi* and its components, represented according to the Topic Maps standard, rather than the process necessary to achieve it. The growth and achievement of an *Innogotchi* can be performed according to a computational model driven by its structure with the objective of *Innogotchi* completion.

The future of this work will concentrate on a better definition of an example of *Innogotchi* and the specification of the computational model.

An Ontology-based Approach to Open Innovation

References

[Buchanan, 1992] Buchanan R., Wicked problems in design thinking. Design issues, 8, 2, 1992, 5-21.

[Fitzgerald et al, 2010] Fitzgerald E., Wankerl A., Schramm C., Inside Real Innovation: How the Right Approach Can Move Ideas from R&D to Market - And Get the Economy Moving, World Scientific Publishing Company, 2010.

[Maicher e Park, 2005] Maicher L., Park J., Charting the topic maps research and applications landscape, in First International Workshop on Topic Maps Research and Applications, Leipzig, Germany, October 6-7, 2005.

[Sowa, 2006] Sowa J., Concept Mapping, AERA Conference, San Francisco, 10 April 2006, <http://www.jfsowa.com/talks/cmapping.pdf>.

[Van der Aalst,1999] Van der Aalst W., On the automatic generation of workflow processes based on product structures, Computers in Industry, 39, 1999, 97–111.

Analisi di una pila protocollare open source per sistemi in tempo reale a basso costo

Ivan Cibrario Bertolotti, Tingting Hu
CNR – Consiglio Nazionale delle Ricerche, IEIIT
C.so Duca degli Abruzzi, 24 – 10129 Torino (Italia)
ivan.cibrario@ieiit.cnr.it, tingting.hu@ieiit.cnr.it

La maggior parte dei dispositivi embedded dev'essere oggi in grado di offrire funzioni quali, ad esempio, configurazione, diagnosi ed aggiornamento remoto dei programmi applicativi mediante una rete di comunicazione. L'avere a disposizione software open source per le pile protocollari TCP/IP e UDP/IP ha reso questo compito relativamente agevole, anche per applicazioni a basso costo. Le caratteristiche di queste pile protocollari per quanto concerne l'esecuzione in tempo reale sono invece ancora poco note, anche perché le applicazioni già citate non necessitano di queste informazioni per funzionare. Questo contributo analizza le prestazioni in tempo reale di una pila protocollare open source e mostra che esse sono accettabili per molte applicazioni industriali, a patto di realizzare un corretto adattamento verso il sistema operativo e l'interfaccia di rete.

1. Introduzione

Molti sistemi embedded odierni sono connessi ad una rete di comunicazione di qualche tipo, allo scopo di fornire funzioni utili quali, ad esempio, configurazione, diagnosi ed aggiornamento remoto dei programmi applicativi. Dal punto di vista software, questa operazione è resa agevole ed economica dalla disponibilità di pile protocollari open source per TCP/IP e UDP/IP che hanno ormai raggiunto un livello di maturità sufficiente per essere adottate in applicazioni reali. Fra di esse ricordiamo, ad esempio, uIP and lwIP [Dunkels, 2003].

Alcuni sistemi operativi in tempo reale per applicazioni embedded, ad esempio eCos e RTEMS, integrano a loro volta una pila protocollare, di norma derivata dalla ben nota pila open source Berkeley BSD [McKusick et al, 1996]. Al tempo stesso, la maggior parte dei microcontrollori in commercio dispone di un'interfaccia Ethernet in grado di funzionare con l'aggiunta di pochi componenti esterni e quindi con costi molto bassi.

In ultima analisi, la connessione in rete di un sistema embedded, anche molto economico, è oggi relativamente semplice. L'unica eccezione di rilievo è rappresentata dallo sviluppo del modulo di controllo del dispositivo Ethernet,

noto anche come device driver, attività che richiede una conoscenza dettagliata del funzionamento del dispositivo stesso.

Il compito è reso più agevole dagli esempi di codice normalmente forniti insieme alla pila protocollare. Tali esempi sono costruiti ad arte per essere strutturalmente il più semplici possibili, di immediata comprensione, nonché facilmente adattabili ad ogni tipo di interfaccia Ethernet, qualunque siano le sue caratteristiche.

Questo modo di procedere è senza dubbio vantaggioso per quanto riguarda i tempi e la difficoltà di sviluppo del codice, ma ha spesso l'effetto collaterale di produrre codice inefficiente. La conseguenza ultima è che, almeno per applicazioni relativamente poco sofisticate quali, ad esempio, un semplice web server per la configurazione e gestione remota di un dispositivo embedded, la maggior parte dell'attività di sviluppo di una pila protocollare si focalizza sul "far funzionare" il codice, anche a costo di penalizzarne notevolmente le prestazioni, soprattutto per quanto riguarda l'esecuzione in tempo reale.

Lo scopo di questo contributo è mostrare invece come la pila protocollare lwIP [Dunkels, 2003] versione 1.3, purché adattata in modo opportuno ad un sistema operativo in tempo reale e ad una interfaccia Ethernet tipica di un microcontrollore odierno, può raggiungere un livello di prestazioni e determinismo sufficiente a costruire su di essa una rete di campo basata su un'interfaccia Ethernet standard sul lato dispositivo, come ad esempio Modbus/TCP [Modbus-IDA, 2006a, Modbus-IDA, 2006b] e Ethernet/IP [IEC, 2010].

La scelta di lwIP è motivata dal fatto che, come indicato in [Dunkels, 2003], essa rappresenta un buon punto di compromesso fra ingombro (in termini di memoria e tempo di esecuzione) e ricchezza delle funzioni offerte. Inoltre, questa pila protocollare è già stata usata con successo in passato per sistemi embedded distribuiti [Pérez et al, 2004], anche se non sono disponibili dati sperimentali sulle prestazioni ottenute in tal caso.

Il sistema operativo in tempo reale adottato per l'analisi è **FreeRTOS**, versione 6.1.0. Si tratta di un sistema operativo ben noto fra gli sviluppatori e già documentato in passato come adatto all'uso in piccoli sistemi embedded, ad esempio in [Cena et al, 2010]. Per quanto riguarda la piattaforma hardware, gli esperimenti descritti in questo contributo sono stati condotti su un microcontrollore NXP LPC1768 [NXP, 2010] a 99MHz, dotato di una CPU ARM Cortex-M3 [ARM, 2010]. La CPU è in grado di eseguire la maggior parte delle istruzioni in un ciclo di clock, ossia circa 10ns.

Il corpo del contributo è organizzato in questo modo: il paragrafo 2 descrive in che modo la pila protocollare lwIP è stata adattata al sistema operativo FreeRTOS, mentre il paragrafo 3 fornisce ulteriori dettagli sul modulo di controllo del dispositivo Ethernet realizzato per lwIP; il paragrafo 4 presenta i risultati sperimentali ed il paragrafo 5 contiene alcune note conclusive.

2. Interfaccia lwIP – sistema operativo

La Fig. 1 illustra la struttura generale del sistema di misura; le parti evidenziate in grigio sono quelle sviluppate ex novo ed oggetto di discussione in

questo contributo. In particolare, l'**interfaccia verso il sistema operativo** dev'essere realizzata caso per caso ed ha il compito di fornire a lwIP un'interfaccia uniforme verso alcuni servizi di comunicazione e sincronizzazione fra thread resi disponibili dal sistema operativo in uso quali, ad esempio, code di messaggi e semafori binari. Essa fornisce inoltre un insieme uniforme di funzioni per creare nuovi thread e delimitare **regioni critiche** cosiddette "veloci", caratterizzate in altre parole da un tempo di esecuzione massimo dell'ordine del centinaio di cicli di clock.

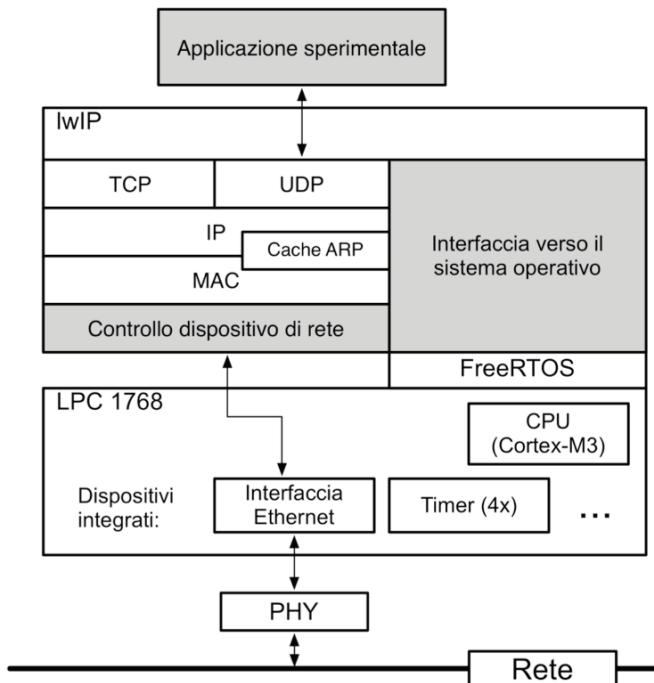


Fig. 1 – Struttura del sistema di misura

L'approccio usato per realizzare questa interfaccia non differisce sostanzialmente da altri già descritti in letteratura, in quanto la maggior parte dei servizi di sistema richiesti da lwIP ha un'immediata corrispondenza con quanto FreeRTOS offre.

Un'importante eccezione è tuttavia costituita dal modo in cui le regioni critiche anzidette sono state realizzate, in quanto la scelta intuitivamente più immediata, ricorrere ad un semaforo di mutua esclusione, si è rivelata inadeguata in questo caso. Le misure hanno evidenziato, infatti, che sulla macchina in uso l'esecuzione di una primitiva semaforica di FreeRTOS richiede circa 300 cicli di clock. Di conseguenza, l'uso di tali primitive per proteggere regioni critiche di durata massima pari a circa 100 cicli di clock avrebbe comportato un overhead inaccettabile.

Per questa ragione, le regioni critiche di lwIP sono state realizzate mediante disabilitazione selettiva delle interruzioni di priorità minore rispetto a quelle dell'interfaccia Ethernet. Questo meccanismo, sebbene insufficiente a risolvere il problema di mutua esclusione nella loro forma più generale, si è dimostrato in questo caso estremamente **efficiente**, tale da poter essere realizzato in pochi cicli di clock, nonché **corretto**, in quanto il campo di applicazione di FreeRTOS è limitato a sistemi monoprocesso. Infine, l'impatto negativo sulla ordinaria **schedulazione** dei thread è estremamente limitato, dato che la massima estensione di tali regioni critiche corrisponde ad un tempo di esecuzione dell'ordine di 1 μ s.

3. Controllo del dispositivo di rete in lwIP

Come illustrato in Fig. 1, il modulo di controllo del dispositivo di rete consente alla pila protocollare lwIP di dialogare con l'interfaccia Ethernet sottostante. In questo caso particolare, esso è costituito da quattro parti principali, evidenziate in grigio in Fig. 2.

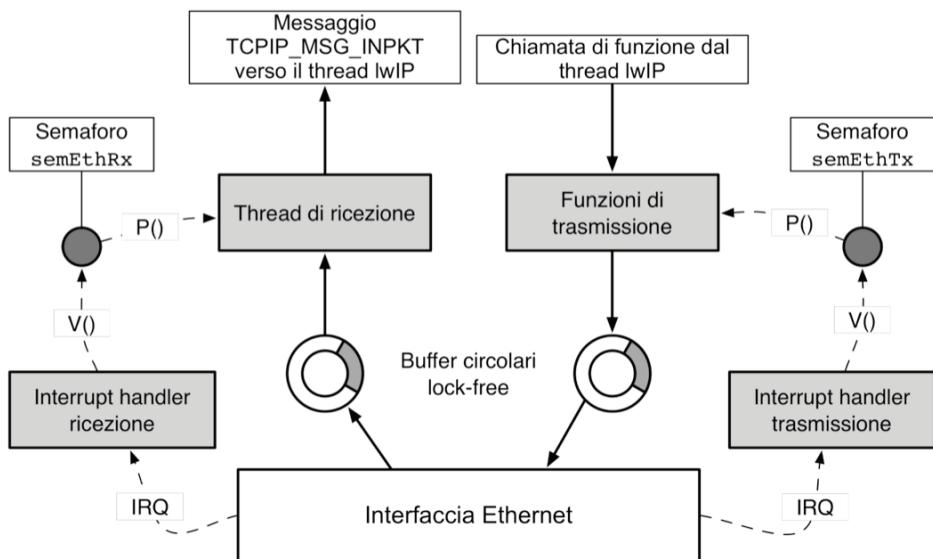


Fig. 2 – Componenti principali del modulo di controllo dispositivo di rete

In particolare considereremo:

- Due routine di gestione delle interruzioni, o **interrupt handler**, eseguite su richiesta del dispositivo per indicare l'avvenuta ricezione e l'avvenuta trasmissione di una trama Ethernet, rispettivamente.
- Un insieme di **funzioni di trasmissione**. Tali funzioni sono chiamate dal thread principale della pila protocollare quando occorre trasmettere una trama Ethernet. Esse hanno il compito di prelevare

la trama da trasmettere dal buffer di lwIP indicato come argomento, passarla all'interfaccia e dare inizio alla trasmissione.

- Un **thread di ricezione** autonomo rispetto alla pila protocollare. Questo thread ha il compito di prelevare le trame Ethernet ricevute dall'interfaccia e passarle al thread principale della pila protocollare mediante una coda di messaggi (per la sincronizzazione) ed un buffer di lwIP (per il trasferimento dei dati).

La **sincronizzazione** fra l'interfaccia Ethernet ed il modulo di controllo del dispositivo avviene mediante due semafori contatori, uno per il flusso di ricezione (`semEthRx`) e l'altro per il flusso di trasmissione (`semEthTx`). Il **trasferimento dati** vero e proprio si basa invece su una coppia di buffer circolari wait-free [Herlihy, 1991] residenti in un'area di memoria dedicata e condivisi fra l'interfaccia Ethernet ed il modulo di controllo.

La principale differenza di progetto fra questo modulo di controllo ed altri moduli simili, disponibili per sistemi operativi di uso generale, consiste nell'aver ridotto la dimensione delle routine di gestione delle interruzioni al minimo indispensabile, in modo che l'unica operazione di rilievo per quanto riguarda i tempi di esecuzione fosse una sola primitiva di sincronizzazione semaforica V().

Ciò è di fondamentale importanza quando si opera con sistemi operativi monolitici, quali ad esempio FreeRTOS, in quanto essi implicitamente assegnano alle routine di gestione delle interruzioni una priorità maggiore di qualunque thread presente nel sistema e non è possibile modificare in alcun modo questa assegnazione. In un tale contesto, le routine di gestione delle interruzioni rappresentano perciò una inevitabile eccezione alle regole di assegnazione delle priorità imposte dall'algoritmo di schedulazione in tempo reale. Di conseguenza, renderle per quanto possibile snelle ne minimizza l'impatto sul sistema.

Gli effetti di questo approccio sono particolarmente evidenti all'interno del flusso di ricezione delle trame Ethernet, dove la maggior parte dell'elaborazione non avviene all'interno della routine di gestione delle interruzioni, bensì in un thread di ricezione indipendente, la cui priorità può essere impostata a piacere. Lungo il flusso di trasmissione, non è invece stato necessario isolare le funzioni di trasmissione in un thread ulteriore, in quanto esse sono già chiamate nel contesto di un thread, ossia il thread principale di lwIP.

4. Analisi dei risultati sperimentali

Lo scopo delle misure consiste nel determinare il contributo dei vari componenti di lwIP illustrati in Fig. 1 al tempo totale di attraversamento della pila protocollare in entrambe le direzioni (ricezione e trasmissione) da parte di trame Ethernet contenenti un datagram UDP. Questa particolare configurazione è stata scelta in quanto essa costituisce la modalità tipica di scambio di dati di processo, ad esempio, in Ethernet/IP [IEC, 2010].

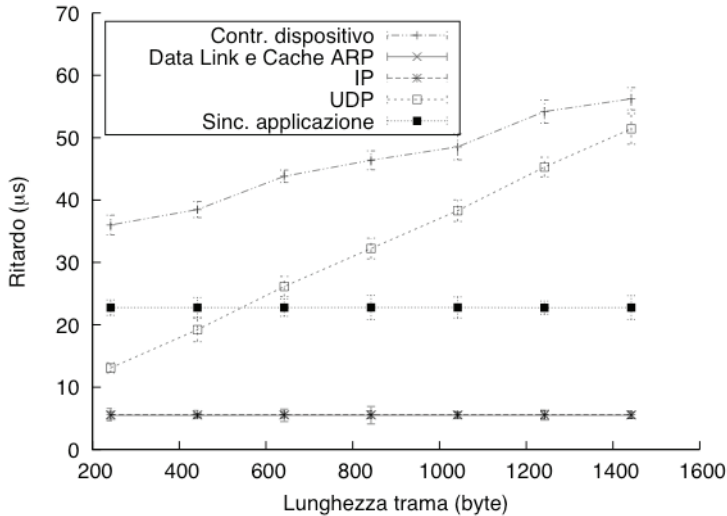


Fig. 3 – Ritardi lungo il flusso di ricezione delle trame

A tal fine, il codice sorgente dei moduli di lwIP nonché del modulo di controllo dispositivo descritto al paragrafo 3 sono stati modificati in modo da memorizzare i tempi di transito delle trame Ethernet in un'apposita struttura dati. La base dei tempi è costituita da uno dei timer a 32 bit interni al microcontrollore, con frequenza di aggiornamento pari a 11MHz, corrispondente ad una risoluzione di circa 90ns.

Le modifiche al codice non ne alterano in modo significativo le temporizzazioni, in quanto la memorizzazione dei tempi di transito è effettuata in circa 10 istruzioni macchina, quindi con un tempo complessivo di esecuzione dello stesso ordine di grandezza della risoluzione del timer.

Durante le misure, è stato usato un generatore di traffico esterno per iniettare trame Ethernet nel flusso di ricezione della pila protocollare. L'applicazione di prova, raffigurata nella parte superiore di Fig. 1, provvede a ritrasmettere le trame ricevute al mittente, fornendo quindi l'opportunità di effettuare misure anche lungo il flusso di trasmissione.

Il numero di campioni raccolti durante le misure è limitato dalla scarsa quantità di memoria RAM disponibile nel sistema. Tenendo conto della memoria necessaria al funzionamento di FreeRTOS, lwIP e dell'applicazione di prova, è stato possibile riservare 16KByte di RAM ai campioni, sufficienti a tracciare circa 400 trame. Per ridurre ulteriormente la quantità di memoria necessaria sono state effettuate due misure distinte per il flusso di ricezione e di trasmissione, rispettivamente.

La Fig. 3 illustra il contributo dei principali componenti della pila protocollare al ritardo delle trame lungo il flusso di ricezione, in funzione della dimensione delle trame stesse. Sono stati considerati, in particolare:

- Il modulo di controllo del dispositivo descritto al paragrafo 3.

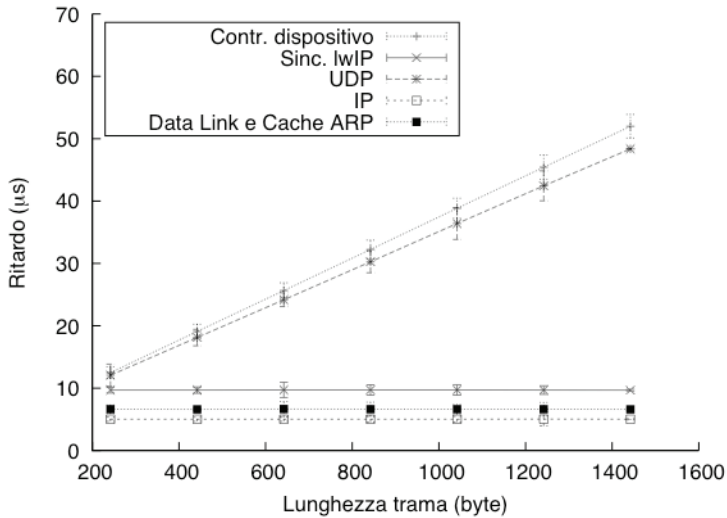


Fig. 4 – Ritardi lungo il flusso di trasmissione delle trame

- Le restanti operazioni effettuate da lwIP a livello data link, ma indipendenti dal dispositivo come, ad esempio, la verifica dell'header Ethernet e la gestione della cache ARP [Plummer, 1982].
- L'elaborazione a livello IP e UDP.
- La sincronizzazione con l'applicazione di prova, per informarla della disponibilità di nuovi dati.

In Fig. 4 è rappresentato invece l'andamento dei ritardi lungo il flusso di trasmissione, anche questa volta in funzione della lunghezza della trama. I componenti considerati sono gli stessi già descritti ma, in luogo della sincronizzazione della pila protocollare con l'applicazione, è stata considerata la sincronizzazione duale, vale a dire la richiesta di trasmissione di un datagram effettuata dall'applicazione verso la pila protocollare.

Il tempo di accesso alla **cache ARP** è di particolare importanza per il flusso di trasmissione, in quanto essa contiene la corrispondenza fra l'indirizzo IP di destinazione ed il relativo indirizzo MAC, senza cui la trasmissione è impossibile. Per contro, è possibile disattivare l'aggiornamento automatico della cache ARP in base alle trame ricevute e, di conseguenza, ridurre i ritardi lungo il flusso di ricezione, impostando la variabile di configurazione `ETHARP_TRUST_IP_MAC` di lwIP a `OFF`.

La prima e, probabilmente, più importante considerazione che è possibile formulare a partire dai dati sperimentali è il notevole impatto delle operazioni di **sincronizzazione** fra il thread principale della pila protocollare ed i thread applicativi sul ritardo complessivo. Tale ritardo è pari a circa 23μs per la ricezione e circa 10μs per la trasmissione, indipendentemente dalla lunghezza della trama. Si tratta quindi di un'importante indicazione, di cui è opportuno tener conto durante l'ottimizzazione del sistema.

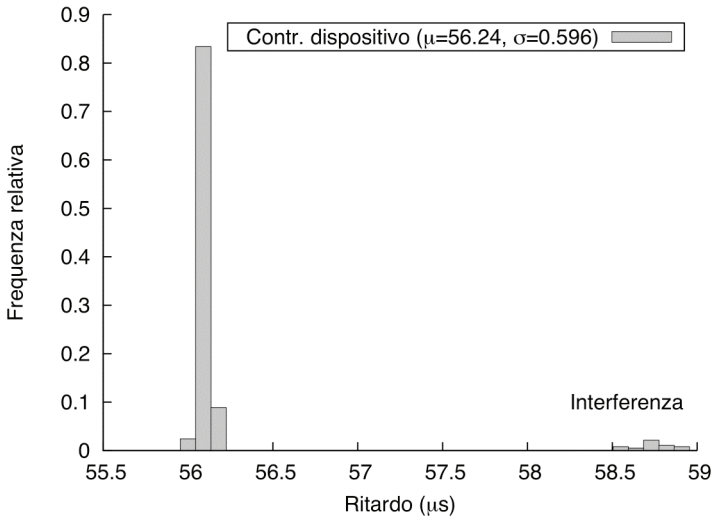


Fig. 5 – Istogramma dei ritardi dovuti al controllo dispositivo in ricezione

Un ulteriore aspetto d'interesse è rappresentato dal fatto che il ritardo dovuto all'elaborazione a livello UDP è proporzionale alla lunghezza della trama, in entrambe le direzioni. Questa peculiarità dipende dalla configurazione standard della pila protocollare lwIP, che prevede la generazione e la verifica incondizionate del **checksum UDP**.

A differenza del checksum a livello IP, che comprende solo l'intestazione [Postel, 1981], il checksum UDP comprende sia l'intestazione, sia il payload completo [Postel, 1980]. Sia la generazione, sia la verifica del checksum UDP comportano pertanto la scansione completa del payload, un'operazione onerosa per cui il sistema in esame non fornisce alcuna assistenza in hardware.

Qualora la comunicazione UDP avvenga mediante un singolo segmento Ethernet, può essere accettabile prendere in considerazione solo gli errori di trasmissione introdotti dai livelli fisico e data link, per cui il Frame Check Sequence (FCS) di Ethernet, generato e verificato in hardware, già fornisce un livello di protezione sufficiente.

In questo caso, senza comunque violare le specifiche di protocollo dato che la gestione del checksum UDP [Postel, 1980] è opzionale [Braden, 1989], l'overhead legato all'elaborazione UDP può essere ridotto considerevolmente impostando le variabili di configurazione di lwIP `CHECKSUM_CHECK_UDP` e `CHECKSUM_GEN_UDP` (per il flusso di ricezione e di trasmissione, rispettivamente) a `OFF`.

La dipendenza del ritardo introdotto dal modulo di controllo del dispositivo rispetto alla lunghezza della trama non è invece sorprendente. Tale modulo esegue, infatti, una copia memoria-memoria delle trame che lo attraversano, durante il loro trasferimento da/verso i buffer circolari condivisi col dispositivo.

Le barre di errore verticali visibili in Fig. 3 e Fig. 4 corrispondono ad una differenza di $\pm 3\sigma$ rispetto al valor medio delle misure, dove σ è la deviazione standard dei dati sperimentali. La principale causa di variabilità nei ritardi nel sistema esaminato è l'interferenza causata dal cosiddetto **tick** di FreeRTOS, un'interruzione periodica con frequenza di 1KHz usata dal sistema operativo per mantenere traccia del tempo trascorso e portare a termine ogni operazione legata ad esso.

Tale interferenza può essere meglio caratterizzata mediante l'istogramma di Fig. 5. Esso rappresenta la frequenza relativa dei ritardi introdotti dal modulo di controllo del dispositivo nel flusso di ricezione, ma tutti gli altri moduli esibiscono un comportamento analogo. Come si può osservare nella parte destra del grafico, l'entità dell'interferenza è alquanto ridotta ed è compresa fra 2.5 e 3 μ s.

5. Conclusioni

In questo contributo, si è mostrato in che modo è possibile adattare la pila protocollare lwIP al sistema operativo FreeRTOS e come dev'essere progettato e realizzato il modulo di controllo del dispositivo Ethernet per ottenere risultati soddisfacenti dal punto di vista dell'esecuzione in tempo reale. Il microcontrollore usato per gli esperimenti è molto simile ad altri, diffusamente usati per sistemi in tempo reale a basso costo; i risultati ottenuti sono quindi applicabili ad una vasta gamma di dispositivi e applicazioni.

I risultati sperimentali discussi nel paragrafo 4 mostrano che la pila protocollare è caratterizzata da un elevato determinismo a tutti i livelli e che l'unica sorgente rilevante di interferenza è dell'ordine dei 3 μ s. Per trame di lunghezza limitata (dell'ordine di 200 byte), spesso tipiche della comunicazione in tempo reale, il ritardo medio totale è di circa 83 μ s per il flusso di ricezione e 46 μ s per quello di trasmissione.

Ciò è particolarmente importante perché mostra che la pila protocollare è adatta non solo per applicazioni con requisiti di esecuzione in tempo reale limitati, come ad esempio un server web o altri tipi di interfaccia utente remota, già realizzati in passato, ma anche per fornire supporto al lato dispositivo di un protocollo per reti industriali basato su Ethernet.

Al tempo stesso, l'analisi ha evidenziato altri aspetti interessanti del comportamento della pila protocollare, utili al fine di migliorarne ulteriormente le prestazioni.

Gli sviluppi futuri dell'attività descritta in questo contributo riguarderanno in primo luogo l'estensione del sistema di misura in modo da poter raccogliere un numero molto più elevato di campioni, con l'ausilio di una memoria di massa, in modo da poter identificare con sicurezza anche anomalie di temporizzazione a probabilità molto bassa.

Successivamente, si prevede inoltre di valutare l'uso della pila protocollare lwIP per le funzioni di comunicazione di un'applicazione in tempo reale tipica dell'ambiente industriale, come ad esempio il master EtherCAT [IEC, 2007] open source IGH EtherLab [IGH, 2010].

Bibliografia

[ARM, 2010] ARM Ltd., CortexTM-M3 Technical Reference Manual DDI 0337H, rev. r2p0, Cambridge, United Kingdom, 2010.

[Braden, 1989] Braden R. (ed), Requirements for Internet Hosts – Communication Layers, RFC 1122, Marina del Rey, CA, 1989.

[Cena et al, 2010] Cena G., Cesarato R., Cibrario Bertolotti I., An RTOS-based Design for Inexpensive Distributed Embedded System, in Proc. of the 19th IEEE International Symposium on Industrial Electronics, Bari, Italy, 2010, 1716-1721.

[Dunkels, 2003] Dunkels A., Full TCP/IP for 8-bit architectures, in Proc. of the 1st International Conference on Mobile Applications, Systems and Services, San Francisco, CA, 2003.

[Herlihy, 1991] Herlihy M., Wait-free synchronization, ACM Transactions on Programming Languages and Systems, 11, 1, 124-149, 1991.

[IEC, 2007] IEC 61158-3/4-121.0, Industrial Communication Networks—Fieldbus Specifications—Part 3–12: Data-Link Layer Service Definition—Part 4–12: Data-Link Layer Protocol Specification—Type 12 Elements, Geneva, Switzerland, 2007.

[IEC, 2010] IEC 61784-1, ed. 3.0, Industrial Communication Networks – Profiles – Part 1: Fieldbus Profiles, Geneva, Switzerland, 2010.

[IGH, 2010] IgH, EtherLab—EtherCAT master, Available online, at: <http://www.etherlab.org/>.

[McKusick et al, 1996] McKusick M. K., Bostic K., Karels M. J., Quarterman J. S., The Design and Implementation of the 4.4BSD Operating System. Addison-Wesley, Reading, MA, 1996.

[Modbus-IDA, 2006a] Modbus-IDA, MODBUS Application Protocol Specification V1.1b. Modbus Organization Inc., Hopkinton, MA, 2006.

[Modbus-IDA, 2006b] Modbus-IDA, MODBUS Messaging on TCP/IP Implementation Guide V1.0b. Modbus Organization Inc., Hopkinton, MA, 2006.

[NXP 2010] NXP B.V., UM10360 – LPC17XX User manual, rev. 2, Eindhoven, The Netherlands, 2010.

[Pérez et al, 2004] Pérez S., Vila J., Alegre J. A., Sala J. V., A CORBA Based Architecture for Distributed Embedded Systems Using the RTLinux-GPL Platform, in Proc. of the 7th IEEE International Symposium on Object Oriented Real-Time Distributed Computing, Vienna, Austria, 2004, 285-288.

[Plummer, 1982] Plummer D.C., An Ethernet Address Resolution Protocol – or – Converting Network Protocol Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware, RFC 826, Cambridge, MA, 1982.

[Postel, 1980] Postel J., User Datagram Protocol, RFC 768, Marina del Rey, CA, 1980.

[Postel, 1981] Postel J. (ed), Internet Protocol—DARPA Internet Program Protocol Specification, RFC 791, Marina del Rey, CA, 1981.

“Analisi tecnico-economica di un modello Smart Lighting per l’illuminazione pubblica”

Bucci¹ F., Annunziato² M., Iacobone¹ F.A., Meloni² C.

¹Università degli Studi Roma Tre

Via della Vasca Navale n. 79 - 00146 Roma

bucci.fabrizio@gmail.com, iacobone@uniroma3.it

²ENEA

C.R. Casaccia, Via Anguillarese n. 301, 00123 Roma

mauro.annunziato@enea.it, claudia.meloni@enea.it

In questo articolo proponiamo un modello di illuminazione pubblica innovativa, basato su tecnologie di Smart Lighting e lo abbiamo comparato con gli approcci comunemente usati nel settore. L'analisi è stata svolta qualificando su un caso reale (Castelnuovo Magra - SP) le varie strategie di intervento. Successivamente è stata effettuata un'estrapolazione dello studio su scale territoriali più estese. I risultati dell'analisi comparativa dimostrano la validità e competitività della soluzione proposta (Smart Lighting).

1. Introduzione

Nell’ottica della riduzione dei consumi dell’energia elettrica, l’illuminazione pubblica (di seguito I.P.) riveste un’attenzione crescente sia per l’elevato volume energetico complessivo (con circa 6,1 TWh l’illuminazione pubblica rappresenta più del 22% dell’intero fabbisogno energetico nazionale per l’illuminazione [TERNA, 2007]), sia perché rappresenta la terza voce di bilancio per i Comuni italiani e sia per i consistenti margini di risparmio energetico potenziale (che oscillano intorno al 30-40%). Nuove tecnologie sono emerse recentemente (come i LED e i sistemi di controllo avanzato) che promettono elevate efficienze energetiche ma la loro introduzione va supportata attraverso approfondite analisi tecnico-economiche per rendere credibile la promessa di prestazioni superiori.

L’obiettivo di questo lavoro è dedicato in particolare alla valutazione della efficacia e della competitività di soluzioni tecnologicamente innovative che si collocano nei contesti più ampi di “Smart City” [Annunziato, 2011] in quanto offrono la possibilità di trasformare la rete di illuminazione pubblica in una infrastruttura elettrica e digitale con diffusione capillare nella città e di integrare nella rete I.P. servizi cosiddetti “Smart” [Art Valley Association, 2010].

Lo studio si è basato su dati raccolti nel contesto del progetto Lumière dell’ENEA (www.progettolumiere.enea.it) che mira alla diffusione sul territorio di strumenti che facilitino i Comuni nel realizzare interventi di efficientamento degli impianti di I.P. [Gozo et al, 2011] attraverso la creazione di strumenti e metodologie di valutazione di tipo tecnico-economico degli impianti I.P. di piccoli e medi Comuni.

L'analisi delle tecnologie esistenti [Bucci, 2011] ha consentito di individuare quattro strategie di intervento che sono state qualificate su un caso reale; queste vanno dalla semplice sostituzione delle lampade poco efficienti, alla regolazione di linea, al telecontrollo punto-punto [Cecchini, 2006], all'illuminazione adattiva.

Dal confronto è emerso che la strategia più interessante, sia dal punto di vista del risparmio energetico che dal punto di vista della competitività è quella del controllo adattivo della illuminazione, ovvero la fornitura di illuminazione in modo proporzionale alla domanda (flussi veicolari e flussi di persone). Tale approccio, ancora in fase di sperimentazione pilota, può produrre mediamente un risparmio energetico di oltre il 40% rispetto al 15% [Annunziato et al, 2011] di risparmio che si ottiene con un controllo del flusso luminoso tradizionale, ottenendo in questo modo la soluzione ottimale che coniuga il risparmio energetico con il tempo di ritorno dell'investimento.

2. Lo stato dell'arte

Al fine di ottenere il massimo risparmio energetico conseguibile e di scegliere le varie soluzioni da confrontare sono state analizzate le tecnologie esistenti e quelle innovative, sia dal punto di vista delle sorgenti luminose sia da quello della gestione del flusso luminoso [Bucci, 2011]. Lo studio si è concentrato in particolare sugli aspetti di gestione e controllo del flusso luminoso e quindi sull'energia consumata dall'intero sistema di illuminazione pubblica. Sono stati analizzati e comparati diversi approcci commerciali alla regolazione del flusso (reattori ferromagnetici biregime, regolatori di flusso centralizzati e alimentatori elettronici dimmerabili), ai sistemi di telecontrollo e telegestione e ad un approccio innovativo basato sui sistemi adattivi.

Per avere un'idea della situazione di partenza del sistema italiano, sono stati analizzati i dati raccolti in ambito del progetto Lumière (Network di 400 Comuni di cui circa 200 hanno fornito informazioni sul loro impianto di illuminazione) al fine di sviluppare metodologie di valutazione ed indicatori di efficienza per l'illuminazione pubblica per piccoli e medi Comuni.

Da quest'analisi è emerso che, per quanto riguarda la potenza installata, ben l'80% non risulta controllata da alcun sistema di gestione del flusso luminoso (stabilizzatori e regolatori) e che più del 40% delle sorgenti luminose è rappresentato dalle inefficienti [Cammarata, 2003] lampade ai vapori di mercurio (di seguito HG).

In questo quadro le attuali strategie di intervento si articolano sulle seguenti scelte.

1. Sostituzione HG-SAP (sodio alta pressione, di seguito SAP): prevede la semplice sostituzione delle lampade al mercurio con le SAP (servizio luce 1 di Consip). È stata la scelta fino ad oggi maggiormente realizzata.
2. Controllo di linea: oltre alla sostituzione delle lampade, prevede l'installazione di orologi astronomici e di regolatori di linea del flusso luminoso su quei quadri elettrici che gestiscono una potenza sufficiente da giustificare il costo di investimento. La regolazione di linea permette di operare in modo uniforme su tutte le lampade alimentate dal quadro che

viene regolato. È la soluzione che oggi viene maggiormente proposta ai Comuni da parte di chi effettua opere di efficientamento di impianti di illuminazione in quanto la più competitiva dal punto di vista economico.

3. Telegestione punto-punto: oltre alla sostituzione delle lampade, prevede la installazione di una rete digitale (tipicamente PLC – Power Line Communication, di seguito PLC) per effettuare il telecontrollo del flusso luminoso separatamente per ogni lampada (un regolatore per ogni punto luce e uno per ogni quadro). Questa soluzione è abilitante agli "Smart Service" in quanto permette di utilizzare il lampione per installare altri sensori e sistemi e di acquisire i dati tramite linea PLC.

3. L'approccio Smart Lighting

L'approccio "Smart Lighting" è derivato dalla telegestione punto-punto ma introduce uno degli elementi tipici trasversali di tutte le applicazioni cosiddette di "Smart City", cioè il principio di adattività e di "Energy on demand".

Tale approccio si basa sulla presenza nella rete IP di lampioni "intelligenti e multifunzionali", ossia equipaggiati con sensori di varia natura (tipicamente telecamere) che attraverso sistemi di comunicazione avanzati (power line ad alta velocità o reti zigbee), interagiscono con un sistema intelligente in grado di ritracciare continuamente il profilo di attività (persone, veicoli, emissioni) della strada in base al quale attivare una regolazione adattiva ed automatica per l'intero anno, del flusso luminoso punto-punto ed altre funzionalità.

Questo approccio, oltre a potenziare notevolmente le prestazioni energetiche dei sistemi punto-punto, apre notevoli prospettive di evoluzione della rete IP in quanto permette l'aggregazione sulla linea della illuminazione di molti "Smart Services" con la conseguenza di abbattere i costi e moltiplicare le prestazioni. Infatti, la gestione autonoma ed adattiva dell'illuminazione è il primo passo verso uno sfruttamento a 360 gradi delle potenzialità che tale apparecchiatura può offrire. Tra gli "Smart Services" più interessanti abbiamo:

- monitoraggio del traffico ed infomobilità;
- supporto alla gestione di flotte di mezzi pubblici e di parcheggi;
- videosorveglianza, sicurezza e gestione delle emergenze;
- rilevamento di inquinamento acustico e ambientale;
- servizi di connessione wireless e servizi informativi;
- gestione, monitoraggio e diagnostica di edifici pubblici;
- caricabatterie per veicoli elettrici.

Ai fini del confronto con le altre soluzioni per la illuminazione pubblica, nel seguito indicheremo con la soluzione Smart Lighting una soluzione simile alla telegestione punto-punto ma che introduce in ogni strada una telecamera (una ogni 500-1000 m) per il rilievo del flusso veicolare e pedonale ed un sistema di controllo che regola il flusso luminoso, secondo normativa [Annunziato et al, 2011], con regolazione del flusso correlato al flusso di veicoli.

Nel rapporto tecnico "Sviluppo sistemi intelligenti per la gestione della "Smart Street" [Annunziato et al, 2011] è illustrata tutta la metodologia e la stima dei risparmi energetici connessi all'uso di questa tecnologia. Da questo rapporto si è visto che è possibile effettuare le seguenti ipotesi: il controllo di

linea consente di ridurre del 20% il flusso luminoso per il 45% delle ore di accensione, la telegestione del 33% per il 45% delle ore e l'illuminazione adattiva del 45% per il 93% delle ore. Inoltre la stabilizzazione della tensione consente di ridurre del 10% la potenza istantanea e gli orologi astronomici fanno passare il numero di ore annue di funzionamento da 4198 a 4000.

4. Analisi tecnico-economica su un caso reale delle strategie di intervento

Per confrontare le quattro soluzioni citate è stato selezionato il Comune di Castelnuovo Magra (SP) oggetto di un approfondito audit energetico del sistema illuminotecnico che garantisce la base di dati necessaria per sviluppare un confronto tecnico-economico. Il Comune presenta queste caratteristiche:

- Abitanti: 8251. Estensione: 15 km².
- Punti luce: 868, di cui: lampade HG: 615 (71%), lampade SAP: 141 (16%)
- Potenza totale: 134 kW. Energia assorbita: 562 MWh/y
- Spesa: potenza impegnata: 4,8 k€/y, componente energia: 78,6 k€/y, manutenzione: 25,2 k€/y, spesa totale: 108,7 k€/y.

La tabella 1 riporta nel dettaglio come variano, rispetto allo stato attuale, i parametri tecnici ed economici per le quattro soluzioni presentate.

	Situazione attuale	Sostituzione HG-SAP	Controllo di linea	Telegestione pto-ptto	Smart Lighting	
Potenza totale	134,6	119,7	119,7	119,7	119,7	kW
Energia annua	562	499	408	386	288	MWh/anno
Consumo equivalente	105	93	76	72	54	TEP/anno
Risparmio energia	-	63	155	176	274	MWh/anno
	-	11	27	31	49	%
Δ emissioni CO₂	-	37	91	104	162	tCO ₂ /anno
Spesa totale energia	83.490	74.198	61.333	58.355	44.608	€/anno
Spesa manutenzione	25.225	21.148	16.147	14.137	12.450	€/anno
Spesa annua totale	108.715	95.346	77.480	72.492	57.058	€/anno
Risparmio spesa totale	-	13.369	31.235	36.223	51.657	€/anno
	-	11	29	33	48	%
Costo di investimento	-	121.192	148.365	230.942	307.742	€
Flussi di cassa	-	13.369	31.235	36.223	51.657	€/anno
VAN	-	27.449	198.916	171.807	266.608	€
TRA	-	11,48	5,37	7,51	6,94	anni
Livello di comfort	3	3	1	2	3	

Tabella 1 - Comparazione tecnico-economica tra le varie soluzioni

La riduzione dei consumi energetici va dall'11% (sostituzione lampade HG) al 49% (illuminazione adattiva) (da 63 a 274 MWh/anno) e quella delle emissioni da 37 a 162 tonnellate di diossido di carbonio l'anno.

La diminuzione dei costi fissi è garantita dalla minor potenza installata grazie all'utilizzo di lampade SAP; quella dei costi variabili dalla riduzione del numero di ore annue di accensione dell'impianto e, quando previsto, dagli effetti positivi della stabilizzazione e regolazione del flusso luminoso. La diminuzione dei costi per la manutenzione deriva dalla maggior durata di vita delle lampade SAP, che riduce il numero di sostituzioni annue, dalla riduzione delle ore di vita equivalenti, per le lampade regolate, e dalla riduzione dei costi di ricerca dei guasti delle lampade telegestite. In più la riduzione nel numero annuo di sostituzioni delle lampade porta come beneficio quello di dover smaltire meno materiale e soprattutto di non dover smaltire lampade contenenti mercurio, altamente pericoloso.

Grazie alle azioni di efficientamento energetico la spesa complessiva del Comune passa dagli attuali 108.715 euro a 95.346 euro mediante la semplice sostituzione delle lampade obsolete e a 77.480 euro se si effettua anche la regolazione di linea del flusso luminoso. Le altre due soluzioni consentono risparmi ancora più marcati e riducono la spesa annua rispettivamente a 72.492 euro (telegestione punto-punto) e a 57.058 euro (Smart Lighting).

5. Analisi comparativa

Il grafico di Fig. 1 riporta l'andamento del valore attuale netto (di seguito VAN) in funzione del tempo; qui è possibile leggere, oltre ovviamente al VAN al termine della durata dell'iniziativa economica, il tempo di ritorno attualizzato (di seguito TRA) e il costo di investimento.

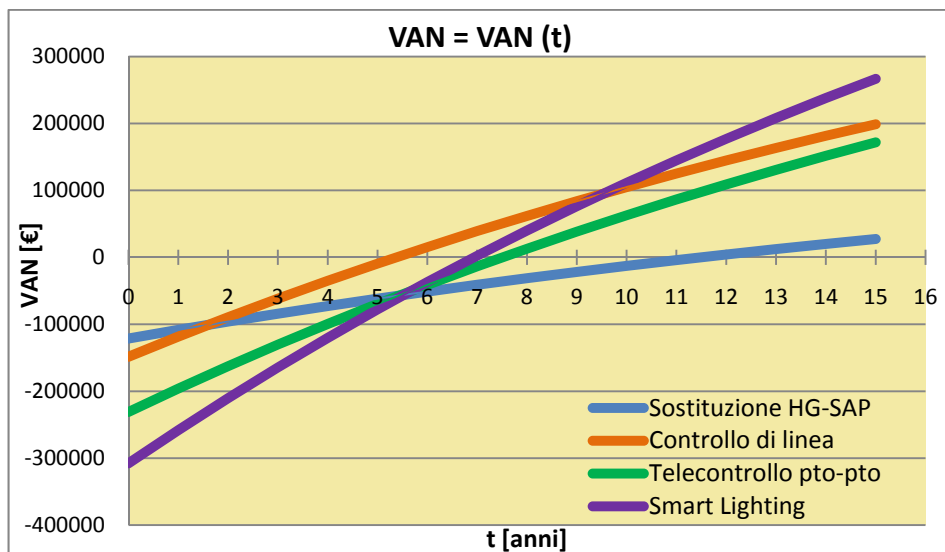


Fig. 1 - VAN nei quattro casi considerati

Come si può notare dalla linea blu la soluzione che fino ad oggi è stata maggiormente praticata è in realtà quella meno conveniente, in virtù del basso VAN e dell'alto TRA.

Il controllo di linea presenta un VAN di circa 200.000 euro e un TRA di quasi 5 anni e mezzo, mentre il telecontrollo punto-punto risulta meno efficace dal punto di vista economico, con un VAN di 170.000 euro e un TRA di quasi 7 anni e mezzo. L'illuminazione adattiva (Smart Lighting) è invece molto competitiva, con un VAN e un TRA rispettivamente di circa 270.000 euro e di poco meno di 7 anni. I costi di investimento vanno dai 120.000 euro per la sostituzione delle lampade, ai 150.000 euro per il controllo di linea, ai 230.000 euro per il telecontrollo punto-punto, ai 300.000 euro per il controllo.

In Fig. 2 sono rappresentati il risparmio energetico percentuale, il TRA e il livello di comfort (rappresentato dalla dimensione delle bolle) che le varie soluzioni possono garantire. Il controllo di linea, che dal punto di vista economico sembrerebbe migliore del telecontrollo, in realtà ha un risparmio energetico e un livello di comfort più basso rispetto alle altre due soluzioni.

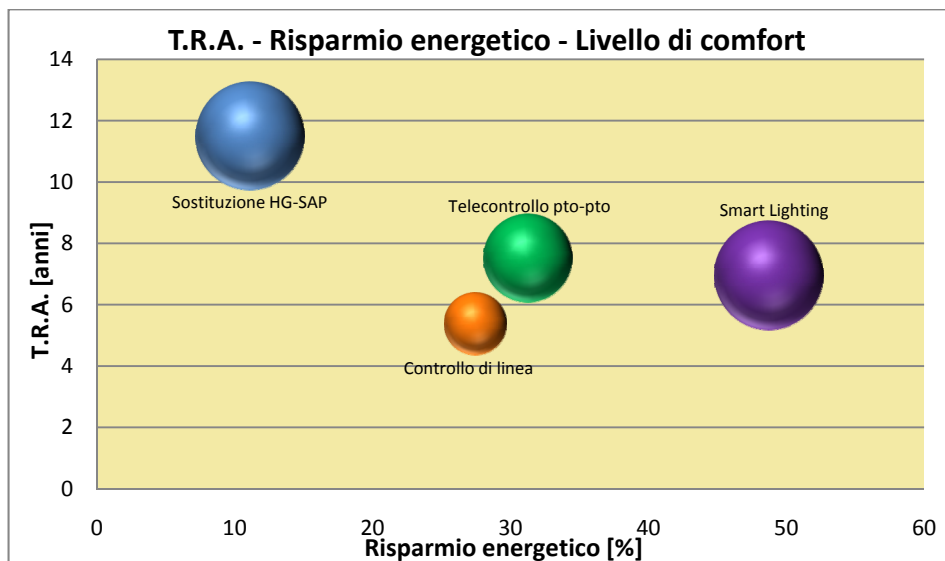


Fig. 2 - Risparmio energetico e TRA dell'investimento nei quattro casi considerati

Da questa analisi possiamo ritenere che l'ultima soluzione, quella "Smart", si possa considerare come una logica e naturale evoluzione competitiva della telegestione punto-punto piuttosto che come una sua alternativa. L'illuminazione adattiva non solo è soluzione complessivamente migliore dal punto di vista economico, ma lo è anche dal punto di vista del risparmio energetico e della qualità del servizio offerto.

Tale soluzione porta risparmi tanto maggiori quanto più il picco di traffico nelle ore di regolazione col controllo di linea o col telecontrollo si discosta da tutti gli altri valori di flusso in tali ore, in quanto è proprio su tale picco che le strategie tradizionali impostano il loro valore di riduzione del flusso luminoso.

Inoltre, oltre al controllo del flusso luminoso, la Smart Lighting può agevolmente integrare altri due servizi smart come il monitoraggio del traffico e la videosorveglianza nei punti di rilievo delle telecamere e potrà integrarne altri di grande interesse economico, sociale, e ambientale con investimenti decisamente sostenibili.

6. Estrapolazione dello studio su scale territoriali più estese

Come ultima riflessione è possibile pensare a quali effetti avrebbe sul sistema Paese una riqualificazione dell'illuminazione pubblica di tutti i Comuni che hanno aderito al Network Lumière.

A tale scopo è stato sviluppato un modello di calcolo per simulare l'estrapolazione degli effetti della applicazione di tecnologie di illuminazione adattiva basato sullo studio fatto su Castelnuovo (vedi tabella 2). Per i Comuni di cui si dispongono i dati, la potenza totale, comprensiva delle perdite, è di 14,3 MW e il 40,8% dei punti luce è dotato di lampade ai vapori di mercurio; inoltre mediamente gli impianti rimangono accesi per 4.200 ore l'anno. Il modello di calcolo considera che mediamente le lampade SAP con cui si sostituiscono quelle HG hanno una potenza nominale del 25% in meno a parità di flusso luminoso emesso e che è possibile controllare all'incirca il 75% della potenza totale.

	Castelnuovo	Paesi aderenti al Network Lumière	Italia	
Situazione iniziale				
Potenza installata iniziale	134	14.339	1.452.381	kW
Spesa potenza imp. iniziale	4.826	516.189	52.285.714	€/anno
Energia annua iniziale	561.886	60.222.055	6.100.000.000	kWh/anno
Spesa energia iniziale	78.664	8.431.088	854.000.000	€/anno
Situazione finale				
Potenza installata finale	119	12.877	1.304.336	kW
Spesa pot. impegnata finale	4.289	463.573	46.956.100	€/anno
Energia annua finale	287.990	31.126.881	3.152.897.610	kWh/anno
Spesa energia finale	40.319	4.357.763	441.405.665	€/anno
Risparmi				
Δ potenza installata	-15	-1.462	-148.045	kW
Δ spesa per pot. installata	-537	-52.616	-5.329.614	€/anno
Risparmio energetico	-274	-29.095	-2.947.102	MWh/anno
Risparmio energetico %	-49	-48	-48	%
Risparmio spesa energia	-38.345	-4.073.324	-412.594.335	€/anno

Tabella 2 - Estrapolazione dello studio a scale territoriali più estese

Benché le ipotesi iniziali siano abbastanza semplificative, i risultati ottenuti sono molto espliciti (vedi tabella 2): con l'illuminazione di tipo adattivo si potrebbe avere un risparmio energetico del 48%, con una riduzione di 29 GWh l'anno di energia elettrica assorbita dalla rete. Tale valore rappresenta quasi lo 0,1 ‰ dell'intero fabbisogno nazionale, pari a circa 310 TWh l'anno [TERNA, 2007]. Questo comporterebbe una riduzione nelle emissioni annue di anidride carbonica di 17.193 tonnellate con conseguente beneficio per tutta la collettività.

In termini economici i risparmi annui sono tutt'altro che trascurabili; la riqualificazione comporterebbe per i Comuni una riduzione annuale di 4.000.000 euro solo per quanto riguarda la componente energia della bolletta.

Con questo modello di calcolo sono stati calcolati, anche se come indicazione puramente qualitativa, gli effetti di una riqualificazione di tutti gli impianti di illuminazione pubblica in Italia. Questa proiezione si basa su alcuni dati e su alcune ipotesi che sono conseguenza dei dati raccolti nel progetto Lumière.

In Italia i consumi di energia elettrica per l'illuminazione pubblica sono di circa 6,1 TWh l'anno cui corrisponde una spesa di circa 854 milioni di euro; supponendo lo stesso numero di ore/anno di accensione dei Comuni del Network (4.200) ne deriva una potenza installata di 1.452 MW.

Se si ipotizza che la percentuale di lampade al mercurio sia del 41% e che quelle al sodio abbiano una potenza inferiore del 25%, la potenza installata, grazie a tale sostituzione, diminuisce di 148 MW (10%).

Telegestendo il 75% della potenza installata con un sistema di tipo adattivo i risparmi energetici ed economici risulterebbero notevoli; l'energia assorbita diminuirebbe di 2,9 TWh l'anno (48%), le emissioni di anidride carbonica di 1.741.000 tonnellate l'anno e la relativa spesa di quasi 413 milioni di euro. Per capire l'entità del risparmio energetico basta confrontarlo con il fabbisogno nazionale (310 TWh/anno) e notare che rappresenta quasi l'1% di tale valore.

Per cercare conferma dei dati a livello nazionale, questi sono stati ricalcolati in altri due modi; più precisamente considerando i risultati, decisamente attendibili, ottenuti per i Comuni del Network ed estendendoli all'intera Nazione mediante il rapporto tra gli abitanti e le superfici. Mediando i risultati si è visto che non si discostano troppo da quelli sopra riportati e riassunti in tabella 2.

7. Conclusioni

In conclusione è possibile affermare che un'illuminazione di tipo adattivo, che sfrutta le tecnologie ICT e che è basata sulla trasmissione di dati digitali, avrebbe ripercussioni decisamente positive su tutto il sistema Paese e per tutte le Amministrazioni Comunali, dal punto di vista energetico, ambientale ed economico (vedi tabella 3).

Comuni del Network Lumière	Proiezione Italia
•Potenza installata: 14 MW •ΔPotenza installata: - 1,5 MW •Energia attuale: 60 GWh/anno •ΔEnergia: - 48% (-29 GWh/anno) •ΔEmissioni: -17.193 tCO₂/anno •ΔSpesa energetica: -4,0 MLN €	•Potenza installata: 1.419 MW •ΔPotenza installata: - 148 MW •Energia attuale: 6100 GWh/anno •ΔEnergia: - 48% (-2947 GWh/anno) •ΔEmissioni: -1.741.503 tCO₂/anno •ΔSpesa energetica: -413 MLN € •ΔEnergia sui consumi nazionali: - 1%

Tabella 3 - Quadro riassuntivo dell'estrapolazione dello studio

Nell'ottica della riduzione dei consumi dell'energia elettrica è stata posta l'attenzione sull'illuminazione pubblica. Dal quadro sullo stato attuale in termini di tecnologie e strumentazione è emersa la necessità di sviluppare una metodologia che promuova l'innovazione tecnologica e di conseguenza l'efficientamento energetico. Il tempo di ritorno economico è infatti inversamente proporzionale all'implementazione massiccia di strumentazione sugli apparati di illuminazione pubblica. Inoltre lo sfruttamento delle onde convogliate consente una gestione centralizzata delle informazioni e dell'elaborazione dati senza nessun costo aggiuntivo in termini di cablaggio e una possibilità computazionale molto elevata, difficilmente raggiungibile elaborando i dati in loco.

L'obiettivo di questo lavoro è stato quello di effettuare una valutazione sulla efficacia e competitività di soluzioni tecnologiche innovative per l'illuminazione pubblica che permettano di integrare nella rete servizi smart, creando strumenti e metodologie di valutazione di tipo tecnico-economico e qualificando le strategie di intervento su un caso reale.

Sono state presentate le tecnologie di gestione del flusso luminoso, dal punto di vista dei sistemi di accensione/spegnimento, della stabilizzazione e regolazione della tensione di alimentazione e del telecontrollo del flusso.

In prospettiva tali tecnologie potranno essere integrate con altri sensori, come ad esempio quello per il monitoraggio della qualità dell'aria, e potrà essere sviluppato un sistema che controlli una sottorete (Smart Street o Smart Square) e successivamente un insieme di sottoreti (rete urbana).

Il sistema di controllo adattivo supera il concetto della semplice regolazione oraria di flusso adattando l'energia erogata all'effettiva domanda del momento in modo tale da consumare energia soltanto nella misura e nel momento in cui l'energia è effettivamente richiesta (Energy on Demand). È stato dimostrato su un caso reale che la strategia di controllo adattiva può produrre mediamente un risparmio di oltre il 40% rispetto al 15% di risparmio che si ottiene con un controllo del flusso luminoso tradizionale non adattivo (sia con il controllo di

linea che con il controllo punto-punto). Il controllo avviene attraverso la regolazione oraria sulla base di un modello predittivo dei flussi veicolari.

L'illuminazione di tipo adattivo, che sfrutta le tecnologie ICT e che è basata sulla trasmissione di dati digitali, avrebbe ripercussioni decisamente positive su tutto il sistema Paese e per tutte le Amministrazioni Comunali, dal punto di vista energetico, ambientale ed economico.

L'applicazione di tecnologie di "Smart Lighting" aumenterebbe considerevolmente l'efficienza energetica del nostro Paese e garantirebbe notevoli risparmi economici; inoltre, come benefico effetto collaterale, potrebbe dare un contributo significativo per il raggiungimento degli obiettivi delle direttive europee e mondiali, come quelle imposte dal protocollo di Kyoto, che prevedono la drastica riduzione delle emissioni di gas serra. Consentirebbe più elevati risparmi in termini monetari non solo a livello nazionale ma anche alle Amministrazioni Comunali e creerebbe un indotto di tecnologie accompagnato da un notevole sviluppo non solo nel settore dell'illuminazione, ma anche in quello dei trasporti, nella gestione degli edifici, nella sicurezza e in generale nella qualità dei servizi offerti ai cittadini e turisti.

In conclusione l'introduzione di un sistema di controllo autonomo ed adattivo, a fronte di costi iniziali non elevati, può portare ad un incremento significativo del risparmio energetico sulla scala dell'intero distretto urbano.

Bibliografia

[Annunziato, 2011] Annunziato M., Le Smart Cities e le sinergie europee del consorzio European Energy Research Alliance, QualEnergia, giugno-luglio 2011, anno IX, n.3.

[Annunziato M. et al, 2011] Annunziato M. et al, Sviluppo sistemi intelligenti per la gestione della "Smart Street", rapporto tecnico ENEA-MSE, Ricerca di Sistema Elettrico PAR 2008-2009, 2011.

[Art Valley Association, 2010] Art Valley Association, Le città intelligenti. Scrivendo un manuale di citymatica per amministratori lungimiranti e cittadini esigenti, Art Valley Association 2010.

[Bucci, 2011] Bucci F., Analisi tecnico-economica di un modello di illuminazione pubblica innovativa basato su tecnologie innovative abilitanti alla SMART CITY, Tesi di laurea in Ingegneria Meccanica, Università degli Studi di Roma Tre, maggio 2011.

[Cammarata, 2003] Cammarata G., Illuminotecnica, Dipartimento di Ingegneria Industriale e Meccanica, Università di Catania, 2003.

[Cecchini, 2006] Cecchini L., Il telecontrollo nell'illuminazione stradale, in "Luce & Design", Maggio 2006.

[Gozo N. et al, 2011] Gozo N. et al, Percorso e soluzioni economico-finanziarie per la realizzazione di interventi di riqualificazione dell'impianto di illuminazione pubblica, Linee Guida, Accordo di Programma ENEA-MSE, 2011.

[Terna, 2007] Terna, "Previsioni della domanda elettrica in Italia e del fabbisogno di potenza necessario, anni 2007 – 2017", Documento Tecnico, 30 Settembre 2007.

Approccio Semantico per un Trattamento Documentale Massivo in Domini Specialistici

Flora Amato, Valentina Casola, Antonino Mazzeo, Nicola Mazzocca, Sara Romano

Università degli studi di Napoli Federico II

Dipartimento di Informatica e Sistemistica

Via Claudio 21, 80127 Napoli

{flora.amato, casolav, mazzeo, nicola.mazzocca, sara.romano}@unina.it

Negli ultimi anni si è assistito ad un crescente interesse per le tematiche di gestione documentale, sostenuto anche da decreti legge ispirati alla reingegnerizzazione dei flussi documentali. In questo lavoro si propone una metodologia di trattamento semantico del flusso documentale, finalizzata all'individuazione della struttura dei documenti, per poter realizzare meccanismi di accesso alle informazioni efficaci ed efficienti e identificare le sezioni sensibili dei documenti, al fine di implementare politiche di controllo accessi a grana-fine.

1. Introduzione

In numerosi contesti applicativi, come quello giuridico, medico ed amministrativo, la gestione documentale occupa notevoli risorse sia umane che tecniche. Numerose applicazioni stanno nascendo negli ultimi anni per supportare gli operatori in tutto il ciclo di vita di un documento, dalla sua ricezione alla lavorazione, fino alla sua chiusura.

La caratteristica principale di tali strumenti, è che essi sono nati per supportare il processo tradizionale, manuale, basato prevalentemente sul forte utilizzo della carta ma, dati i numerosi vincoli normativi e tecnologici, non riescono completamente a sostituire il processo tradizionale cartaceo.

Sebbene esistano numerosi strumenti software che sono in grado di gestire completamente il flusso documentale, essi non sono adottati in toto, il perché è abbastanza intuitivo se si pensa ai vincoli normativi che impongono la conservazione a lungo termine di atti cartacei, se si pensa alla necessità di avere documenti firmati, anche da più soggetti, se si pensa alla difficoltà di alcuni operatori nell'utilizzo delle tecnologie informatiche e, non da ultimo, se si pensa all'esperienza comune di sfiducia nelle caratteristiche di sicurezza ed affidabilità sia delle informazioni digitalizzate che dei loro supporti (cd masterizzati e non più leggibili). Di fatti, dal punto di vista dell'operatore medio, conservare e proteggere la carta è "più sicuro" che proteggere dati conservati su un server non direttamente accessibile (leggasi vicino alla scrivania).

Lo stato attuale vede una forte convivenza di informazioni digitali e informazioni cartacee che rendono non solo costoso e poco gestibile l'intero processo, ma lo rendono in parte anche fallace. Il motivo della non efficienza è

legato al fatto che numerosi documenti, sebbene digitali, non sono strutturati affatto; gli operatori, infatti, non sono interessati a strutturare i loro documenti.

Gli operatori dei domini applicativi menzionati, sono abituati a pensare ai loro documenti come a degli oggetti monolitici la cui protezione è legata alla protezione del supporto stesso, così come avviene per i documenti cartacei. Essi non sono dunque abituati a strutturare i loro documenti perché non comprendono che grazie alle tecnologie oggi disponibili, strutturare le informazioni significa anche garantire una loro migliore gestione e una più adeguata protezione delle stesse.

Un esempio significativo è dato dalle cartelle cliniche: i dottori e i loro assistenti in molti contesti scrivono ancora su fogli bianchi contenuti misti quali i dati anagrafici del paziente, le sue precedenti malattie, la diagnosi attuale e la terapia da seguire. Se il dottore ha utilizzato un sistema di editing informatico, dovrà proteggere l'intero file e, per rispetto della privacy del paziente, nessun altro operatore sanitario dovrebbe poter leggere quel file. Gli attuali meccanismi di protezione delle informazioni permettono di implementare politiche di controllo accessi a grana-fine, ovvero le risorse da proteggere non sono necessariamente l'intero file ma possono essere parti dello stesso. Questo implica che anche i documenti possono essere protetti nelle loro singole parti nel modo più appropriato, a condizione che sia resa esplicita la loro struttura, ovvero sia possibile identificare le singole parti che lo compongono.

Dunque, non solo la gestione ma anche la sicurezza delle informazioni è fortemente legata a come tali informazioni sono state strutturate.

A tale scopo, da alcuni anni stiamo investigando l'adozione di tecniche di analisi semantica per analizzare testi ed estrarre automaticamente informazioni rilevanti anche da documenti destrutturati; l'idea è quella di estrarre informazioni, concetti e relazioni complesse da documenti destrutturati al fine di individuare specifiche sezioni "critiche" e, in questo modo, strutturare i dati in esso contenuti.

In questo lavoro, si vuole presentare un framework riconfigurabile in grado di elaborare testi destrutturati di varia tipologia (testuali, multimediali,...) purché appartengano ad un dominio specifico, che sia in grado di strutturare in maniera semiautomatica le informazioni di interesse. Saranno definite delle regole di trasformazione che permetteranno di caratterizzare le varie parti di un documento. A tale scopo useremo una metodologia di classificazione delle risorse in grado di localizzare le parti critiche di un documento che devono essere protette dall'accesso non autorizzato da parte di diversi soggetti. Di seguito, presenteremo non solo l'architettura del framework ma anche un caso studio nel dominio della telemedicina per mostrare come parti riservate di una cartella clinica possano essere protette con politiche di sicurezza a grana fine anche se le informazioni della cartella sono destrutturate.

2. Il processo di trattamento semantico delle informazioni su domini specialistici

Come detto precedentemente, una delle maggiori limitazioni nell'adozione di strumenti informatici a supporto della gestione documentale, è causata dalla non strutturazione dei documenti prodotti da molteplici operatori. In questo paragrafo, vogliamo illustrare un modello generale di framework per l'analisi dei documenti in grado di estrarre in maniera semi-automatica informazioni rilevanti, concetti e relazioni complesse che permettano di strutturare documenti destrutturati.

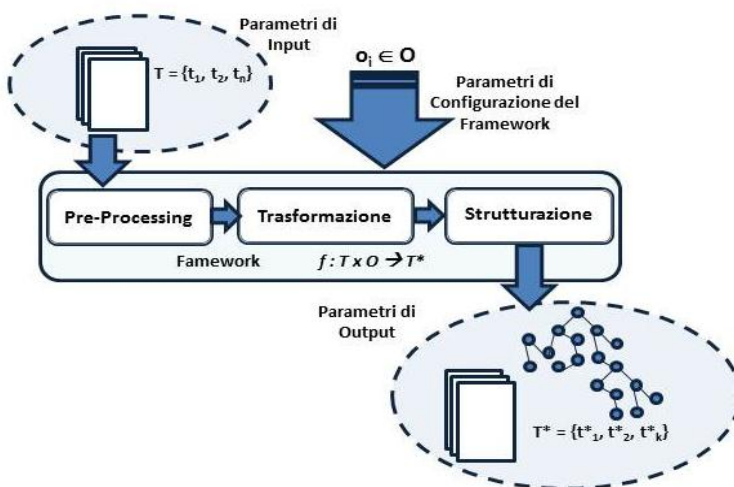


Figura 1 – Framework per l'elaborazione dei testi

In Figura 1 è riportato lo schema del framework proposto. Come si vede, esso è formato da tre blocchi principali: (i) il modulo di Pre-processing per estrarre gli elementi testuali dai file in ingresso; (ii) il modulo di Trasformazione in grado di trasformare gli elementi testuali in base ad un insieme di regole di trasformazione che dipendono dal dominio applicativo e che sono selezionate nell'ambito dei possibili parametri di configurazione; (iii) il modulo di Strutturazione che elabora ed eventualmente formatta i dati strutturati in base al loro utilizzo finale. Il dominio applicativo determina la tipologia di ingressi consentiti e l'insieme di tecniche ed algoritmi che meglio si adattano allo specifico dominio.

Il framework può dunque essere caratterizzato da una funzione di trasformazione $f: T \times O \rightarrow T^*$ e dall'insieme dei documenti in ingresso $T = \{t_1, t_2, \dots, t_n\}$, l'insieme delle possibili strutture dati $T^* = \{t^*_1, t^*_2, \dots, t^*_k\}$ in cui i contenuti documentali possono essere strutturati e l'insieme $O = \{o_1, o_2, \dots, o_s\}$ dei parametri di configurazione che permettono di selezionare le più appropriate tecniche di trasformazione dei documenti, come sarà illustrato

nel prossimo paragrafo in cui analizzeremo un'istanza architetturale di tale framework.

2.1 Architettura del Sistema

I documenti appartenenti ai domini specialistici sono costituiti da sezioni, ognuna concernente un aspetto: è possibile di fatto individuare la struttura implicita dei documenti, identificando appunto le sezioni che li costituiscono. Al fine di applicare meccanismi di protezione a grana fine sui documenti, è necessario rilevare quali sono le sezioni rilevanti da proteggere.

Nel presente lavoro proponiamo l'architettura di un sistema di gestione documentale, basato su metodologie di trattamento semantico (Amato et al.,2010). Il sistema proposto, illustrato in fig. 2, è di fatto, un'istanza del framework presentato nel paragrafo precedente.

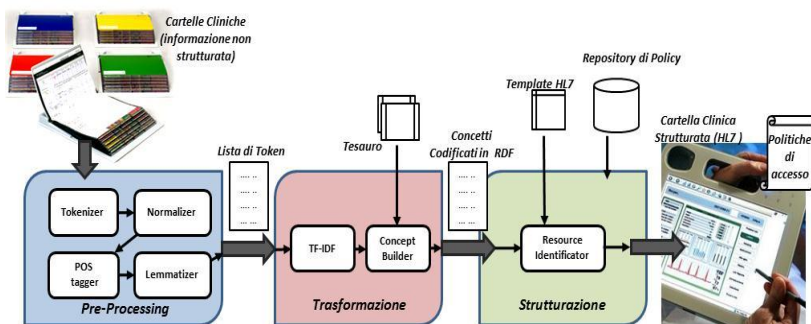


Figura 2– Architettura del sistema di trattamento semantico delle informazioni applicato al dominio medico

Per illustrare le caratteristiche del sistema ne considereremo l'applicazione al dominio medico delle cartelle cliniche: il sistema è in grado di processare cartelle cliniche e produrre in output le sezioni di cui sono composte, associare poi il livello di confidenzialità con cui ogni singola sezione del documento deve essere trattata.

In questo modo è possibile applicare meccanismi di protezione a grana fine sulle cartelle cliniche, proteggendo solo le parti sensibili dei documenti, oppure garantendo l'accesso a sezioni critiche al solo personale autorizzato, permettendo al contempo l'accesso ad altre sezioni del documento agli altri utenti.

Per il dominio specialistico in esame è possibile individuare (in accordo con gli standard internazionali di rappresentazione delle informazioni mediche HL7) differenti tipologie di sezioni, come le sezioni inerenti le Informazioni sull'anagrafica, le informazioni prelevate al momento dell'ammissione del paziente nella struttura medica e le informazioni diagnostiche e cure prescritte.

L'architettura proposta è composta da quattro componenti principali: Pre-Processing, Identificazione dei Concetti, Strutturazione e Controllo degli Accessi; tali componenti sono descritti di seguito.

Pre-Processing: questo componente è responsabile dell'estrazione del testo dai file, e della conseguente individuazione delle unità lessicali che saranno processate dagli altri componenti. Il Componente di Pre-Processing implementa le procedure di *Tokenizzazione* e *Normalizzazione*. Le procedure di *Tokenizzazione* sono finalizzate alla segmentazione delle frasi in token, unità atomiche di analisi, costituite da termini semplici oppure abbreviazioni, acronimi ed espressioni alfanumeriche. La *tokenizzazione* prevede attività di analisi dei segni grafici, per individuare l'insieme dei simboli di alfabeto, distinguendoli dai simboli separatori, esso effettua la disambiguazione dei segni di punteggiatura per separare stringhe contigue (ad es la stringa c'era, deve essere analizzato come giustapposizione di due termini separati da apice). Per gli scopi di tokenizzazione si utilizzano sia glossari contenenti espressioni idiomatiche del dominio specialistico trattato (nel nostro esempio, medico) che grammatiche che implementano regole euristiche per la combinazione dei token. La combinazione di glossari e grammatiche ha mostrato, nei risultati sperimentali, un alto livello di accuratezza, anche in presenza di testo ricco di abbreviazioni ed acronimi quale quello medico. Le procedure di *Normalizzazione* sono finalizzate alla riduzione di tutte le variazioni ortografiche di un dato termine, esse sono responsabili di

- ridurre tutte le parole in lettera minuscola, gestendo in maniera opportuna le parole che, scritte in minuscolo o maiuscolo assumono significati differenti, (ad esempio un nome comune riportato in maiuscolo in quanto posto ad inizio frase può essere convertito in minuscolo, ma il nome di persona "Rosa" se riportato erroneamente in minuscolo potrà essere confuso con il nome del fiore);
- separare le parole composte o prefissate, che possono contenere segni non alfabetici come il "-" (ad esempio è possibile trovare il termine Data-base in luogo di database);
- uniformare le date che possono comparire in differenti formati ("1 Gennaio 1948" o "01/01/48");
- uniformare gli acronimi e le abbreviazioni ("USA" in luogo di "U.S.A.", "pag" per "pg").

Identificazione dei Concetti: questo modulo è finalizzato all'identificazione dei termini rilevanti, utilizzando metodi statistici e lessicali. I termini estratti saranno poi associati ai relativi concetti. Queste informazioni saranno formalizzate in opportune ontologie di dominio. I termini estratti saranno utilizzati per la caratterizzazione delle sezioni di interesse (Amato et al.,2009): l'identificazione delle sezioni avviene, infatti, proprio in base alla presenza/assenza di particolari concetti presenti in essa.

I termini sono estratti in base alla loro rilevanza semantica, valutata in base alla categoria sintattica di appartenenza e al valore dell'indice TF-IDF (Term Frequency - Inverse Document Frequency) (Salton et al.,1988) ad essi associato sulla base del numero di occorrenze del termine nel documento e della sua distribuzione su tutto il corpus. Una volta estratti i termini rilevanti, si

eseguono operazioni di clustering su essi, volte a raggrupparli in “synset” (insiemi di termini considerati semanticamente equivalenti, nel senso che godono della proprietà di poter essere scambiati in una frase senza che questa perda di significato). Ogni synset identifica un concetto. In questo modo è possibile designare, e in generale, ricercare concetti nei documenti, indipendentemente dal particolare termine utilizzato nel documento per riferirsi ad esso.

In lavori precedenti l'uso dei concetti finalizzato all'individuazione delle sezioni è già stato sperimentato, sia nel dominio medico (Amato et al.,2010b) che in quello giuridico (Amato et al.,2008). Per la definizione dei synset facciamo ricorso a risorse esterne, specialistiche del dominio medico, in particolare utilizziamo l'ontologia del dominio medico “Unified Medical Language System” (UMLS) (Indberg et al.,1993) e il tesaurus di termini medici “Mesh” (Autori vari, 2010). Il ricorso a risorse esterne permette di riferirsi ai concetti utilizzando anche termini differenti da quelli impiegati nella base documentale. Infatti sarà possibile per un utente, che interroga il sistema con una query in linguaggio naturale, recuperare i documenti di interesse, anche se essi contengono concetti espressi con termini diversi da quelli utilizzati nella query, in quanto le risorse esterne riportano, per un dato concetto, tutti i termini utilizzati per riferirsi ad esso, indipendentemente dal fatto che il termine sia comparso o meno nella base documentale.

Strutturazione: questo modulo è responsabile dell'identificazione della struttura dei documenti, riconoscendo le sezioni che lo costituiscono. Al fine di riconoscere le sezioni si esegue un processo di classificazione, utilizzando, come features, la presenza o l'assenza dei concetti identificati nelle fasi precedenti. A tal fine è stato implementato un sistema di voting che utilizza la combinazione di tre differenti tipologie di classificatori: Nave Bayes (Aas et al.,1999), Decision Tree (Ridgeway et al.,1998) e K-Nearest Neighbor (Yuan et al.,2008). I risultati dei singoli classificatori sono combinati e in caso di disaccordo, alla sezione in esame sarà assegnata la tipologia indicata dalla maggioranza dei classificatori. Questo modulo è in grado di individuare la tipologia delle sezioni, da cui dipende il livello di protezione con cui essa deve essere gestita.

La tipologia di una sezione è determinata dall'insieme di concetti presenti (o assenti) in essa.

Ogni sezione costituisce quindi una risorsa, da gestire con livelli di protezione dipendenti dalla sua tipologia (ad esempio, se una sezione è identificata come diagnosi sarà accessibile al personale medico, ma non a quello amministrativo, che invece potrà accedere alle sezioni identificate come anagrafica).

Questo processo automatico di identificazione delle risorse necessita della supervisione di un esperto di dominio nella fase di start-up: infatti l'esperto deve stabilire quali sono i concetti che caratterizzano le sezioni (ad esempio i concetti “paziente” e “sintomo” sono associabili alla sezione “diagnosi”). Tale associazione è propedeutica al processo di classificazione automatica delle risorse. In figura 3 è illustrato l'output del sistema, che è in grado di riconoscere la tipologia della sezione in analisi come “Diagnosi” grazie alla presenza in essa

Processi di Trattamento Semantico delle Informazioni per domini specialistici dei concetti rilevanti “paziente”, “diagnosi” e “terapia”, come sarà mostrato nella discussione del caso di studio.

Controllo degli Accessi: Questo modulo è responsabile di regolare l'accesso alle informazioni, mediante l'assegnazione, ad ogni risorsa, della policy di accesso appropriata. Infatti, quando la tipologia di una risorsa è identificata, è possibile applicare ad essa la relativa politica di accesso, in tal modo si realizza un sistema di protezione su tutta la base documentale, e, implementando una politica di accesso a grana fine, permettere agli utenti di accedere solo alle parti di documento a cui sono abilitati.

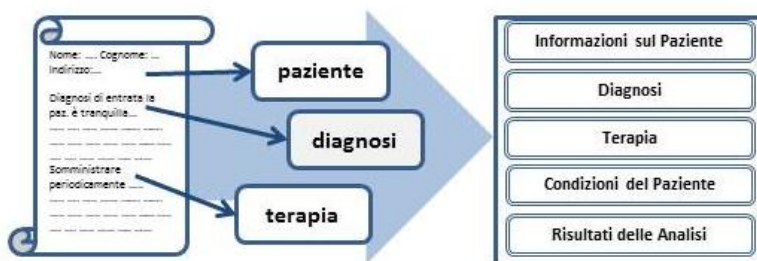


Figura 3 – Esempio di identificazione di risorse appartenenti ad una cartella clinica

Al fine di regolare gli accessi, il sistema prevede tre tipologie di utente: (i) Amministratore, che può gestire l'accesso al sistema di nuovi utenti e inserire nuovi documenti; (ii) Esperto di dominio che può controllare e raffinare gli output del sistema, al fine di effettuare operazioni di validazione sia sulla lista di concetti impostati come rilevanti dal sistema che sulle risorse individuate e classificate; (iii) Utenti standard, a cui possono essere associati differenti profili, autorizzati ad accedere a particolari risorse in base al loro ruolo nella struttura medica in cui il sistema è stato adottato.

3. Caso di studio: strutturazione di documenti in ambito sanitario

Per il trattamento documentale in ambito sanitario, il framework descritto nel paragrafo 2 è stato istanziato selezionando opportunamente i parametri di selezione σ_i . In questo modo vengono selezionati gli opportuni algoritmi che implementano ciascun modulo del sistema. L'input fornito al sistema è costituito dal corpus (formato da cartelle cliniche destrutturate) che verrà poi opportunamente strutturato mediante i moduli che implementano il sistema. La strutturazione delle cartelle cliniche rappresenta un passo fondamentale per l'individuazione delle sezioni che le compongono. Tali sezioni rappresentano, nel nostro lavoro, le risorse da proteggere e la loro individuazione permette di applicare ad essi le opportune regole di sicurezza. La metodologia semantica descritta nel paragrafo 2 permette di individuare la presenza/assenza di concetti rilevanti che identificano le risorse (sezioni della cartella clinica) da proteggere (Amato et al.,2010b). Come mostrato in figura 2, il sistema è composto da diversi moduli, ognuno dei quali implementa gli

algoritmi necessari allo scopo di strutturare documenti sanitari destrutturati. Gli algoritmi del modulo di Pre-processing sono realizzati ricorrendo a tool per l'analisi del testo in lingua italiana TaLTaC2 (Bolasco et al.,2010). Anche il modulo di Trasformazione utilizza le procedure di TaLTaC2 per il calcolo dell'indice TF-IDF mentre il componente Concepts Builder, responsabile dell'individuazione dei concetti, è stato implementato dal nostro gruppo di ricerca. Tale componente raggruppa i termini rilevanti (quelli aventi valori di TF-IDF più alti) in concetti sfruttando le informazioni sulle relazioni semantiche tra termini provenienti da un tesaurus di dominio medico (Autori vari, 2010).

Infine, il componente "Resource Identification" del modulo di Strutturazione, a partire dai concetti individuati dal modulo di Trasformazione, identifica le sezioni critiche di ciascun documento utilizzando un algoritmo di classificazione offerto dal tool KNIME (Berthold et al.,2006).

Per poter descrivere meglio le fasi della metodologia semantica implementata, si consideri il seguente frammento di una cartella clinica:

"La Signora si presenta con un anamnesi di precedenti ricoveri presso differenti reparti di questo ospedale. Inquietata ed a tratti aggressiva, manifesta un forte stato d'ansia e dolori allo stomaco. Vista la storia clinica di patologie ansiogene del paziente, le sono stati somministrati 10mg di Maalox."

La prima fase della metodologia consiste nell'estrazione dei termini che avviene grazie al modulo Pre-processing (si rimanda la lettura per i dettagli di tale fase a (Amato et al.,2010)). A questo punto il modulo di Trasformazione individua i termini rilevanti mediante il calcolo del valore TF-IDF associato a ciascun termine, che è stato estratto nella fase precedente. In questa fase vengono selezionati tutti quei termini che hanno il valore di TF-IDF oltre una certa soglia (che nel nostro caso è 3.0): "paziente"(4.1), ansia"(4.2), "dolori allo stomaco"(3.8), "aggressiva" (3.1), "storia clinica"(4.8), and Maalox" (4.7). Questi termini vengono poi raggruppati in synset, ossia in gruppi di termini che sono sinonimi tra loro, sfruttando le informazioni provenienti dal tesaurus. Ogni synset rappresenta un concetto ed ogni concetto viene associato a una sezione del documento così come mostrato in Tabella 1.

Termini estratti	Synset	Concetti	Sezione associata
Paziente	Ammalato, degente, malato, paziente	Paziente	Diagnosi
Ansia	Ansioso, ansiogeno, ansia	Ansia	Diagnosi
dolori allo stomaco	Dolori_allo_stomaco, mal_di_stomaco	Dolori allo stomaco	Diagnosi
Aggressiva	Aggressivo, aggressiva	Aggressivo	Stato paziente
Storia clinica	Storia clinica, cartella clinica	Storia clinica	Stato paziente
Maalox	Maalox	Maalox	Terapia

Tabella 1 – Associazione tra i termini estratti e i concetti

Nell'esempio sono stati individuati i concetti dei termini: "paziente", ansia", "dolori allo stomaco", "aggressiva", "storia clinica", e "Maalox". Tali concetti sono stati codificati in formato RDF e l'insieme delle risorse individuate (sezioni delle cartelle cliniche) è stato codificato, mediante il modulo di Strutturazione, nel formato standard per la codifica delle cartelle cliniche, HL7.

All'interno dell'esempio di cartella clinica descritto sopra, la presenza dei concetti "Paziente", "Ansia" e "Dolori allo stomaco" (che si trovano nell'ultima parte dell'esempio) e l'assenza di concetti che tipicamente appartengono ad altre sezioni, classifica il frammento di cartella clinica analizzato come sezione "Diagnosi". Questa risorsa potrà quindi essere accessibile solo da chi avrà i permessi. In un meccanismo di controllo degli accessi basato su regole, tali permessi sono associati ad una regola e vengono assegnati da un amministratore di sistema in accordo con le regolamentazioni legali e aziendali (ad esempio legge sulla privacy). In precedenti lavori abbiamo definito delle policy di sicurezza per tale dominio che vedevano coinvolti i seguenti attori: Dottore, Personale amministrativo, Infermiere e Paziente. A titolo esemplificativo possibili policy basate sul ruolo degli attori, per le sezioni "Diagnosi" possono essere: "lettura", "scrittura". Nell'esempio preso in considerazione in questo lavoro, per la sezione "Diagnosi" individuate, le policy di sicurezza sono:

R1: {Dottore, Diagnosi, (Lettura, Scrittura)}

R2: {Infermiere, Diagnosi, (Lettura, ¬Scrittura)}

R3: {Personale amministrativo, Diagnosi, (¬Lettura, ¬Scrittura)}

R1 indica che un medico può effettuare operazioni di Lettura e Scrittura sulla risorsa "Diagnosi", mentre un Infermiere può solo effettuare operazioni di lettura (R2) e il Personale amministrativo non può effettuare alcuna operazione su tale risorsa (R3).

4. Conclusioni

Al giorno d'oggi, molti sistemi si basano sul trattamento documentale senza beneficiare delle nuove tecniche per la strutturazione dei dati; questo impone diversi limiti tra cui, per esempio, quello di proteggere i dati in modo corretto applicando politiche di controllo degli accessi a grana fine. In effetti i sistemi di trattamento documentale che sono attualmente a disposizione si limitano a digitalizzare i documenti cartacei rendendoli disponibili agli utenti. Ma in questo modo, essendo il documento trattato come un blocco monolitico, risulta difficile applicare le adeguate policy di sicurezza che garantiscano la privacy e la riservatezza dei dati. Per questo motivo, in questo lavoro abbiamo proposto la formalizzazione di un framework che sia in grado di analizzare e trasformare dati testuali ed abbiamo inoltre proposto l'implementazione di un'architettura il cui funzionamento è basato sull'adozione di una metodologia semantica. La metodologia semantica sviluppata è in grado di recuperare informazioni da specifiche parti del documento che possono essere utili per la classificazione, la sicurezza, ecc. Attraverso l'architettura proposta, abbiamo illustrato l'applicabilità del framework su un caso di studio in modo da mettere in evidenza le potenzialità dell'approccio proposto da punto di vista della sicurezza e abbiamo mostrato come è possibile definire policy di sicurezza a grana che regolino l'accesso ai dati, in origine destrutturati, che sono stati strutturati sfruttando la metodologia semantica.

Noi riteniamo che la formalizzazione del framework proposta in questo lavoro sia molto promettente e, per tale motivo, gli sviluppi futuri che intendiamo intraprendere mirano all'investigazione sull'uso di altre metodologie volte ad

integrare diversi approcci di analisi, che potrebbero essere il punto di partenza per lo sviluppo di nuove applicazioni.

Bibliografia

K. Aas and L. Eikvil, Text Categorisation: A Survey, Report No 941, Norwegian Computing Center (1999), ISBN 82-539-0425-8. 1999

F. Amato, A. Mazzeo, V. Moscato, A. Picariello, Information Extraction from Multimedia Documents for e-Government Application- s Book Chapter of Information Systems. pp. 101 108, Springer, 2010.

F. Amato, A. Mazzeo, V. Moscato, A. Picariello. "A System for Semantic Retrieval and Long Term Preservation of Multimedia Documents in E-Government Domain". To Appear in International Journal of Web and Grid Services, Vol. 5, No. 4, Inderscience Publishers, pp. 323.338(16), 2009.

Amato F., Mazzeo A. , Penta A. , Picariello A. , Knowledge Representation and Management for E-Government Documents, Book Chapter of E-Government Ict Professionalism and Competences Service Science, pp.31 40, Springer 2008.

Amato F, Casola V, Mazzeo A, Romano S. A semantic based methodology to classify and protect sensitive data in medical records. In: Information Assurance and Security (IAS), 2010 Sixth International Conference on. Atlanta, GA, p. 240-246, ISBN: 978-1-4244-7407-3.

Berthold MR, Cebon N, Dill F, Di Fatta G, Gabriel TR, Georg F, Meinl T, Ohl P (2006) KNIME: The Konstanz Information Miner, In: Proceedings of the 4th annual industrial simulation conference, Workshop on multi-agent systems and simulations, Palermo.

Bolasco S. (2005). Statistica testuale e text mining: alcuni paradigmi applicativi, Quaderni di Statistica, Liguori Ed., 7, p. 17-53.

HL7 CDA Release 2.0 2005. The HL7 Version 3 Standard: Clinical Data Architecture, Release 2.0, ANSI Standard.

L. Indberg , D . A . B . , H Umphreys , B . L . M C C Ray , A . T . (1993) . The unified medical language system . Methods of Information in Medicine , vol. 32 , pp. 281 291 .

Autori vari, Medical Subject Headings of National Library of Medicine. <http://www.nlm.nih.gov/mesh/> .

G. Ridgeway, D. Madigan, T. Richardson. Interpretable boosted naive bayes classification. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. 1998.

Salton, Gerard and Buckley, C. (1988). "Term-weighting approaches in automatic text retrieval". Information Processing & Management 24 (5): 513523.

The Medical Subject Headings comprise the National Library of Medicine's www.nlm.nih.gov/mesh/

P. Yuan, Y.Chen, H. Jin, L. Huang, MSVM-kNN: Combining SVM and k-NN for Multi-class Text Classification. IEEE International Workshop on Semantic Computing and Systems, 2008. WSCS '08. Pp. 133-140. ISBN: 978-0-7695-3316-2. 2008.

Architettura e sostenibilità nel Mediterraneo

G. Piero Cossu, Daniela Ludoni¹, Gianraffaele Loddo², Marco Pittaluga³

*Dipartimento di Architettura, Università degli Studi di Cagliari
Piazza D'Armi 19, 09123 Cagliari (CA)*

picossu@tiscali.it

¹*Dipartimento di Architettura, Università degli Studi di Cagliari*

dlludoni@alice.it

²*Dipartimento di Architettura, Università degli Studi di Cagliari*

grloddo@unica.it

³*Dipartimento di Architettura, Università degli Studi di Cagliari*

mpittalu@unica.it

Progetto, costruzione e gestione dell'edificio, vitali alla architettura altamente sostenibile, si codificano da decenni anche con riguardo agli effetti sull'economia edilizia, avendo contribuito il settore ad inquinamento ed esaurimento di risorse non rinnovabili. Le ulteriori relazioni sperimentali tra materia e contorno vagliano su qualità ed innovazione prestazionale di materiali e costruito.

Fra molteplici materiali performanti, il versatile vetro circoscrive, in particolare, l'indagine del Dipartimento di Architettura di Cagliari. Per le prestazioni fornite ai fini energetici di risparmio ed efficienza, i ricercatori proseguono l'indagine in ambito climatico locale, con l'edificazione di test room, a seguito dei promettenti esiti delle simulazioni sui vetri elettrocromici. Di rilevante applicazione per effetti energetici e di comfort, gli elettrocromici consentono di ridurre l'apporto necessario alla climatizzazione estiva nelle zone calde del Mediterraneo e, per traslucenza variabile, forniscono codici lessicali di un'architettura finita e coerente.

1. Introduzione

Fra i problemi ambientali, il cambiamento climatico sollecita l'umanità per gli esiti causati. Al contempo, l'instabilità economica genera effetti in ambiti vasti e coinvolge l'occupazione e gli equilibri sociali. Spicca la dipendenza tra ereditati meccanismi produttivi e consumismo, inidonea per processi irreversibili e residua fruibilità di risorse. In ambito edilizio, la disarticolazione si palesa nello specifico delle certificazioni energetiche di costruito e di suoi componenti industriali. Per causa di incoerenti relazioni fra procedure ed operatori di deboli impostazioni di metodo correnti, essa compromette il fine di più alti gradi d'efficienza. L'innovazione tecnologica agevola l'ottimizzazione dei componenti, pur se talvolta gli operatori non riconoscono le articolazioni da adattare alle variabili esterne, garantendo flessibilità e armonia all'organismo. Per identificare la sequenza di criteri operativi, efficaci all'articolazione d'insieme, l'analisi rivela

un metodo che simula gli effetti su ogni elemento, verificandoli in opera. In analogia ai criteri sperimentali, si perfeziona l'apporto di ogni componente e delle fasi del processo. Tale più congrua impostazione di metodo programma l'efficienza energetica del processo, nelle prassi esecutiva e gestionale. Per validare il processo, il metodo sperimentale abbozza modelli articolati, potenziati con tecnologie avanzate, testando ogni stadio di coerenza. La ricerca d'avanguardia si indirizza su più efficaci processi produttivi, congrui a disponibili energie rinnovabili, compresi risparmio ed efficienza energetica. Il metodo introduce verso coeve sperimentazioni, caratterizzate da ibridazione di tecniche (biomimicking) e avanzamento tecnologico. Finora l'avanguardia ha orientato il progetto dei componenti sulle esigenze d'organicità; oggi deve strumentalizzarsi alla sostenibilità globale, inclusi costruzione ed esercizio. Qui rilevano tecniche edili e forme d'architettura, impostate su gradi di reversibilità ed efficienza di variabili configurazioni dello spazio naturale ed edificato. Tenuità e leggerezza qualificano l'organismo, in coerenza a tecnica ed ambiente. Sulle logiche sperimentali dei codici contemporanei dell'architettura, il ricorso al rendimento tecnologicamente avanzato di materiali innovativi consente il dissolversi della materia apparente, con traguardi di traslucenza variabile di involucro e struttura. Come la trasparenza è materia e forma di progetto e costruito, anche l'efficienza energetica ne è materia e forma, percepibile fino al compimento del ciclo di vita utile dell'organismo edilizio. La sperimentazione inquadra l'involucro edilizio, interfaccia dinamica, responsabile del comfort giacché reattivo a micro e macroclima. Comparate performanti pelli artificiali ed avanzate tecnologie, l'analisi conduce fino ai vetri elettrocromici. Il progresso dei componenti edilizi identifica per essi affinamenti prestazionali conseguibili su coerenti fasi costruttive e gestionali.

2. Materiali e vetri elettrocromici

I materiali elettrocromici (EC) o *elettrodi ad avvicinamento di ioni* sono dei cromogenici: materiali intelligenti adattabili all'ambiente per l'attivarsi di reazioni chimico-fisiche. Generata una differenza di potenziale elettrico (*d.d.p.*), cristalli liquidi (LCs), Suspended Particle Device (SPD) ed EC persistono in una reazione reversibile delle performance di trasmissione spettrale. Nelle tipiche applicazioni edili, i materiali cromogenici ricorrono a sottilissimi film, sovrapposti agli strati, anche intervallati da interstizi, di vetri o sintetici trasparenti. Il controllo della frazione di calore e luce in ingresso ottimizza l'efficienza ed il comfort, riducendo i consumi per climatizzare ed illuminare. Tuttavia il ricorso ai LCs è arduo per inafferrabili livelli intermedi di traslucenza e per i consumi energetici di trasparenza. Questo aspetto ostacola anche l'impiego di vetri SPD, pur con permanenti gradi di traslucenza intermedia. Le prestazioni dei vetri EC o *smart windows* stimolano l'avanzamento tecnologico dell'industria. Le tecnologie a film sottile (*sputtering*), sol-gel o a sedimentazione di vapori chimici producono vari tipi di sistemi EC: sottili pellicole solide di EC con elettroliti, film solidi EC con gel elettrolitici, soluzioni di EC in elettroliti in stato gel o liquido. Per tutti i requisiti del sistema cambiano con l'attivarsi di un campo elettrico. La conversione per assorbimento dallo stato iniziale al finale (*switching*) si verifica

applicando una tensione elettrica. La conseguente d.d.p. ($1\div 5V$) tra EC e elettrolita mobilita gli ioni, addizionandoli o sottraendoli. L'esodo influenza la trasmissione nella gamma delle radiazioni visibili e invisibili e altera lo spettro cromatico del sistema. I livelli restano in carica più o meno stabilmente fino a memorie a lungo termine ($12\div 48h$). Il diverso colore persiste con la tensione garante dello scambio ionico. Per un tempo utile, ciò scherma la radiazione solare nel campo visibile ($380\div 780\text{ nm}$) e dell'infrarosso ($780\div 2500\text{ nm}$). Il sistema EC si combina su tre o cinque strati diversi con spessori fino a 10^{-3} mm (vedi Fig.1). Nella fattispecie, l'elettrolita costituisce lo strato centrale, perdendo ioni per d.d.p. Sulle sue facce esterne si stratificano l'elettrodo EC attivo e il negativo (ECE) accumulando gli elettroni. Gli strati esterni del conduttore trasparente fanno presa sul supporto vitreo o sintetico (PC, PMMA, PET), a protezione dal degrado causato da agenti atmosferici e inquinanti. Configurazioni ed effetti mutano ricorrendo ad elettroliti ad accumulo di ioni, o combinando EC di colori diversi in stato ridotto/ossidato. Il comportamento EC si manifesta in materiali inorganici, quali ossidi di metalli di transizione, e in polimeri organici, di differenti contrasto EC, efficienza di colorazione e velocità di switching. Alla riduzione del WO_3 si apprezzano valori notevoli di questi parametri, il che ne motiva il maggior impiego, insieme alle variazioni d'intensità del visibile. Fra gli altri EC: MoO_3 , TiO_2 , Nb_2O_5 ; viologeni e polimeri conduttori: PANI, PPy, PT, ecc. [Loddo, 2010a]

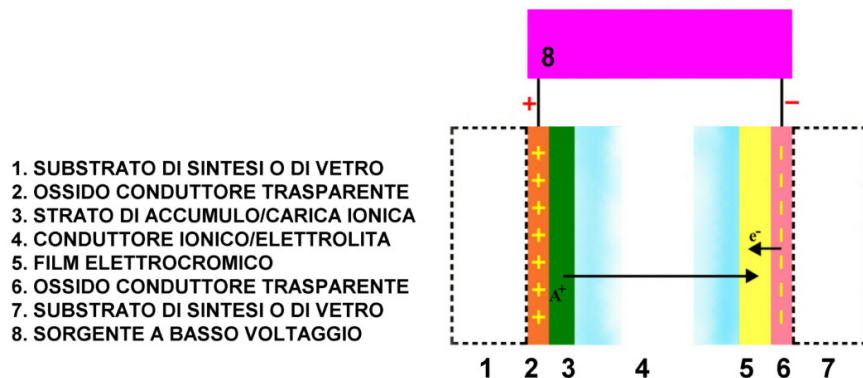


Fig. 1. Il dispositivo EC

Progredite nanotecnologie hanno ammesso sviluppi a fini militari e aerospaziali, con applicazioni per vetture ed edilizia. Nel 2000, la NASA (con Jet Propulsion Laboratory-California Institute of Technology e Ashwin-Ushas Corporation Inc., NJ), indagava su materiali a cambiamento di fase, leggeri e coibenti per sonde a controllo termico flessibile destinate a diversi ambienti. Per la navicella, il J.P.L. ideò un dispositivo EC a IR con un polimero conduttore, garantendo il controllo termico delle bocchette di ventilazione. La ridotta tensione ($1,5V\text{ CC}$) avrebbe alterato la capacità di riflessione, secondo il range spaziale simulato. Per fini militari poi ci si sarebbe avvalsi di EC ai IR di polimeri mimetizzabili. Il mercato automobilistico rilevò sul ciclo di vita utile dell'EC e sulle reazioni alle alte temperature, trascurando le dimensioni ridotte dei vetri EC. Infatti, la tecnologia consente da subito il controllo attivo di specchi di EC organici e prototipi per tettucci. Anche in edilizia l'avanzamento sperimentale e i nuovi modelli conseguirono all'implementazione tecnologica. Allestiti i vetri EC per interagire all'ambiente, essi avrebbero contribuito al maggior comfort

termico e visivo interno, avversando interferenza luminosa e surriscaldamento e riducendo i consumi. Asahi Glass® (J) testò prototipi 10x10 cm e commercializzò lastre 30x30cm, assemblando al Seto Bridge Museum (Kojima, Okayama-Pref., J, 1988) 196 vetri EC (40x40cm), garanti della vista sottomarina, passando dal blu al trasparente a tempo di musica. Si ricorse ad un EC di WO_3 amorfo e colorato con ioni Li^+ e ad un elettrolita polimerico. Lastre poco più ampie (45x40cm) s'installarono nella Daiwa House (Mita-City, Hyogo-Pref., J, 1988).

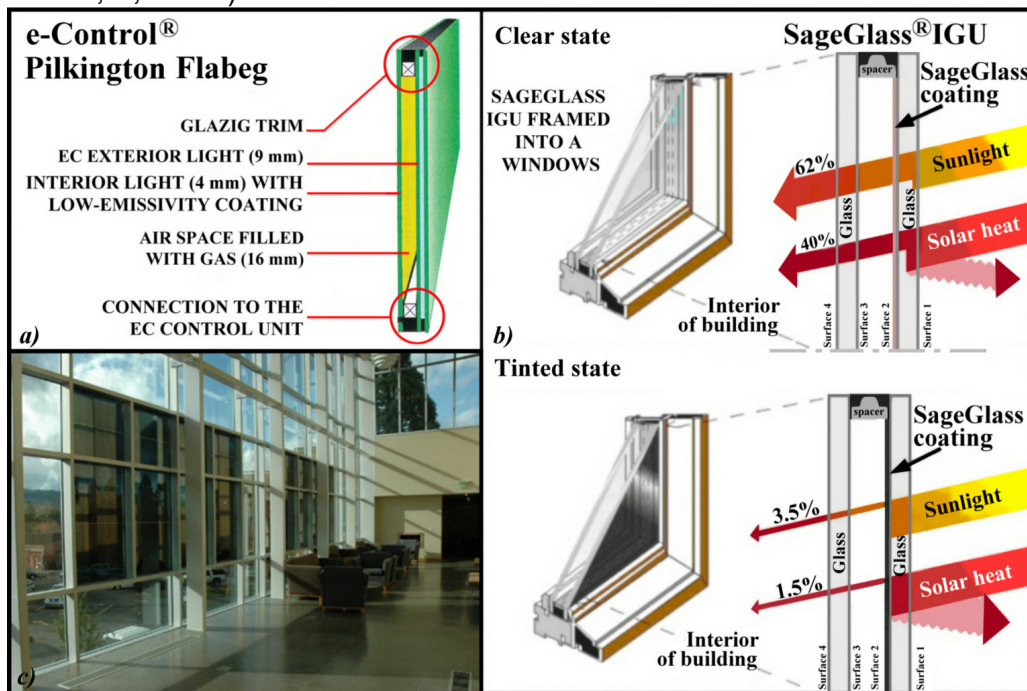


Fig. 2. a) E-control® Pilkington Flabeg. Fonte: [Dorado, 2001]; b) IGU SageGlass®. c) Shiley Hall, University of Portland (OR). Vetri EC attivi controllare abbagliamento e guadagno termico solare. Fonte b), c): Electrochromics Inc.

A Düsseldorf, il Glasstech '98 fu l'occasione dell'E-control ®-Pilkington Flabeg, usato nella Stadtparkasse di Dresda. Si trattava di un vetrocamera con film EC che garantiva anche da rottura. Aderente all'interno dell'E-control® (sp. 9mm), il film si separava dai 4mm del vetro basso emissivo (*low-E*) con una camera d'aria (16mm) colma d'Argon (vedi Fig.2a). Con un controllo esterno silenzioso e a basso consumo (2W/mq), il fattore di TL variava da 50 (trasparente) a 15% (scuro) e il fattore solare g da 40 a 15%. Il progresso dei vetri EC per l'edilizia riguarda performance e resistenza, pur senza conseguire lastre estese; l'offerta commerciale varia sugli alterni successi d'impresa. Fra le aziende all'avanguardia emergeva la SAGE Electrochromics (Faribault, MN) per dimensioni, tecnologie e qualità di prodotto. Il SageGlass® Classic™ Heat Treated Sealed Insulated Glass Units (IGUs) (vedi Fig.2b) ha valori estivi e invernali di trasmittanza termica (U) di 0,28 e 0.29W/m²K; l'inversione tra stadio attivo e inattivo porta: TL da 3,5 a 62%; g da 0,09 a 0,48; la riflettanza interna ed esterna da 10 a 15% e da 6 a 21%. Oltre allo Standard Double-Pane®, SAGE commercializza anche un High-R Triple-Pane®, con accresciuta resistenza termica (7,35m²K/W), limitando al 50% l'energia altrimenti necessaria in stato chiaro. Circa la U del SageGlass® rileva che negli U.S.A. i metodi di calcolo sono diversi da quelli europei, secondo riferimenti su condizioni di clima

standardizzati dal National Fenestration Rating Council e dall'European standard European Committee for Standardization. Diversi involucri di edifici destinati al terziario ricorrono a vetri EC (vedi Fig.2c), con sempre più interesse del settore residenziale, per incrementi di produzione e di mercato e ridotti costi.

3. Performance dei vetri EC in clima mediterraneo

In ambito mondiale la ricerca integra i vetri EC all'involucro edilizio, giacché sistema flessibile ai bisogni locali di operatori ed utenti. Algoritmi di controllo e strumenti di simulazione, diagnosi e gestione integrano le fasi di processo, imponendo maggiore sensibilità verso gli operatori, per afferrare i conseguibili benefici. Per una tecnologia coerente alle prestazioni pretese, la ricerca s'indirizza su tecnologie produttive di vetri EC di maggiori dimensioni, integrati con regolazioni automatizzate, anche di domotica avanzata [Loddo, 2010b]. Il successo impone test su vetri EC in vari climi, più freddi di quello californiano dove all'inizio hanno mostrato efficacia, e in quelli temperati caldo-subtropicali. Progresso, cooperazioni, produzioni e nuovi mercati motivano investimenti in rapida espansione in Europa. Conseguiti livelli intermedi tra stadi trasparente e colorato del SageGlass® e lastre più ampie ($1,5 \times 3,5 \text{ m}^2$) si rapporta l'avanzamento tecnologico su accresciuti livelli di comfort e di produzione. Nel 2010, SAGE Electrochromics Inc. destina \$100 milioni di credito d'imposta alla produzione di tecnologie per l'Europa, investendo la Saint Gobain su un nuovo impianto di produzione con capacità iniziale di 4 milioni di metri quadri annui. Alle potenziate performance del prodotto si associa la possibilità di acquisire crediti conformi agli standard L.E.E.D. Tutto ciò motiva l'interesse del Dipartimento di Architettura (DiArch) di Cagliari per le SageGlass®, testando l'efficacia dei vetri EC all'efficienza energetica dell'involucro in clima mediterraneo, dove è intenso l'apporto solare. In Sardegna il clima è temperato caldo. Sulla costa l'influsso marino genera inverni miti e estati prolungate. Il caldo marca il Sud dell'Isola, con maggiore umidità relativa attenuata solo dall'azione dei venti da W-NW. In estate s'accresce il disagio per saturazione dell'umidità relativa e vapori sospinti da S-SE sul litorale densamente abitato. La tradizione costruttiva locale ha privilegiato ingenti masse murarie d'inerzia termica, garanti del raffrescamento estivo per l'inclinazione zenitale dei raggi solari. Tuttavia, odierne architetture fanno presa sulle coste mediterranee. Spesso si tratta di involucri che con vetrate comunque estese ed orientate, esaltano il rapporto interno/esterno, pur se a qualunque costo di gestione. Solo tipologie costruttive autoctone, valorizzate di recente, e il ricorso a più coerenti principi bioclimatici accordano più ammissibili pratiche costruttive e gestionali. La riflessione sull'uso di ampi vetri, al di là del simbolismo architettonico, rileva sui benefici invernali e sul surriscaldamento estivo, cosicché opportunamente selezionato il tipo e l'orientazione, i vetri devono garantire, in sinergia con l'involucro, gradi d'efficienza congrui alle qualità locali, sfruttando l'innovazione dell'EC. Per tutto ciò, il Sud dell'Isola assurge a riferimento dell'intero ambito climatico Mediterraneo: la ricerca si ubica nell'area baricentrica di Cagliari, relazionando poi i parametri ad ambiti attigui con adattamenti correlati. La fase preliminare ha simulato le performance di vetri EC con codici di calcolo in

regime dinamico, risultando stimolante l'entità del conseguibile risparmio energetico teorico. La simulazione del 2008 ha riguardato un retrofit energetico su una torre per uffici in Cagliari. Lo studio, concentrato sui suoi livelli più alti a curtain-wall, ha concepito, in particolare, la sostituzione dei vetri con quelli EC (vedi Fig.3), elaborando un più adeguato sistema di facciata, secondo gradi d'efficienza e prestazione, pesati sulla destinazione d'uso.

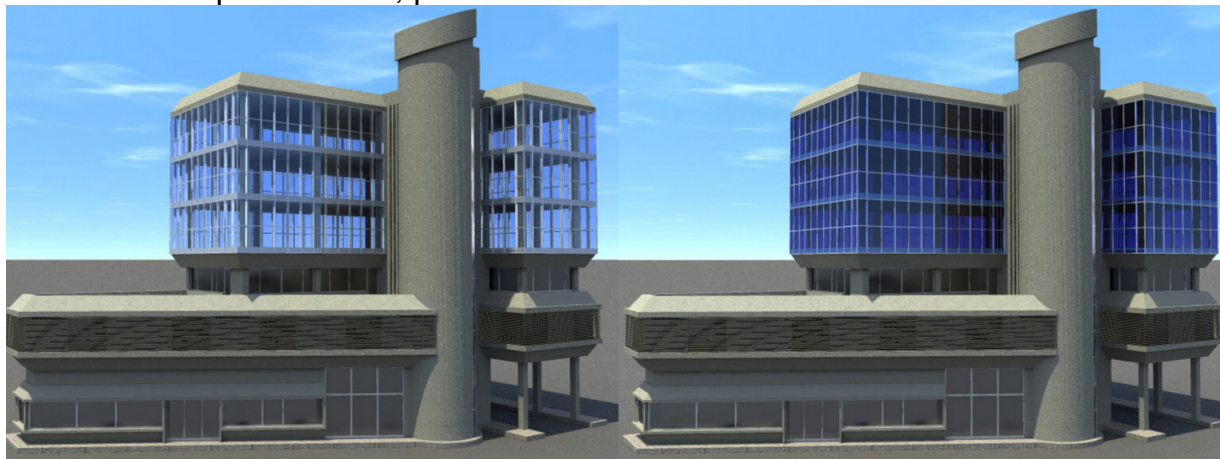


Fig. 3. Rendering dell'involucro della torre con vetri EC in stati Off ed On.

La simulazione [Loddo, 2010a] ha valutato i consumi per climatizzare la torre, elaborando il confronto di modelli diversi (stratigrafia, entità di superficie vetrata, ricorso a vetrate commerciali low-E ed a vetri SageGlass®). Impostati gli input sui modelli, noti radiazione in ingresso e consumi di climatizzazione, il calcolo ha privilegiato l'uso di vetri EC in estate. Sui modelli con superfici di vetro incrementate o meno, la simulazione estiva sui vetri EC ha rilevato costi energetici ridotti del 76 e 57% rispetto ai low-E. Per accertare se l'uso dei vetri EC e il risparmio sulla climatizzazione ammortizzassero, in tempi ragionevoli, il maggior costo dei pannelli EC rispetto ai tradizionali, la preliminare analisi costi-benefici ha stimato solo i benefici direttamente monetizzabili, rivelando che per il costo dei vetri EC, l'intervento di retrofit non sarebbe stato ammortizzabile in 15÷20 anni con il solo risparmio energetico. La ricerca sarebbe tuttavia proseguita, prevedendo un calo dei costi dei vetri EC.

Con l'iniziale contributo della Fondazione Banco di Sardegna, il DiArch avvia la sperimentazione, cooperando con la SAGE Electrochromics che fornisce know-how, tecnologia e il SageGlass® e concorre al sistema di controllo ed al trasporto. Gli studiosi coinvolgono il Dipartimento di Ingegneria Elettrica ed Elettronica (D.I.E.E.) di Cagliari ed alcune imprese locali. L'apporto sinergico tra ricerca universitaria e impresa promuove sviluppo e trasferimento di tecnologie dai laboratori all'imprenditoria, secondo competenze essenziali a fondare la sperimentazione. Il D.I.E.E. fornisce strumenti di rilievo e software per programmare i dati raccolti, nonché la pratica del protocollo sperimentale, indagando sui vetri EC (efficienza, alimentazione, influenza di climatizzazione artificiale su microclima e interferenze del macroclima). L'apporto d'impresa riguarda Laborvetro s.r.l. e Ziro Immobiliare s.r.l., entrambe con sedi a Cagliari, ma operative oltre i confini regionali. Le prestazioni dei vetri EC suscitano interesse e attese che impongono alla ricerca tempi medio-lunghi, rendendo indisponibile la torre. L'esigenza di test room con identici requisiti geometrici, d'uso ed orientamento, sempre disponibili, impongono d'impostare un modello

geometrico e costruttivo, utilizzabile a lungo termine, svincolato dal contesto. Ciò reclama l'ulteriore analisi energetica da impostare sul modello, propedeutica alla fase sperimentale [Loddo, 2011]. L'analisi rende disponibili dati che, utili all'edificazione delle camere, nonché comparabili e verificabili durante i test, riguardano le caratteristiche dei componenti edilizi EC, accertandone l'attendibilità in relazione alle prestazioni locali. Per l'immediata disponibilità del sito, le test room si localizzano sulla copertura di una sede del DiArch (vedi Fig.4).



Fig. 4. Cagliari, sede del DiArch la cui copertura accoglie le camere di prova

L'analisi energetica condiziona la sperimentazione avviata allestendo a luglio 2011 le due test room, una con vetri EC e l'altra con convenzionali low-E. I dati periodicamente rilevati si elaboreranno per confrontare risultati e finalità. Oltre agli esiti della simulazione, l'obiettivo sperimentale inquadra anche un più attendibile contronto di costi e benefici, conseguente all'uso di vetri EC e in attinenza al valore medio-basso del rapporto tra superfici finestrate e opache, e all'ambito climatico, procedendo ad indentificare un range ove sia ottimale la convenienza d'uso dei vetri EC. Le camere A e B si configurano geometricamente su due cubi identici di 4,00m di lato, altezza utile di 2,70m e copertura piana appena inclinata. Per dimensioni modelli e aperture vetrate si conformano agli standards edilizi comunali. Il contesto non condiziona l'ubicazione sul lastrico solare, orientando le finestre sugli assi di soleggiamento massimo (S-E, S-O) e distinguendo fra portafinestra di m 0,80x2,10 (h) e finestra di m 1,50x1,00 (h), con quoziente di 0,05 tra vetri e superfici opache. I modelli s'impostano su due sistemi costruttivi. In uno le chiusure verticali sono di blocchi di laterizio alveolare a fori verticali (vedi Fig.5a), di alto potere isolante termo-acustico; un intonaco termico ne rifinisce le due facce. Le chiusure orizzontali sono in latero-cemento (sp. 24cm), stratificate (sp. 30cm) per l'isolamento termo-acustico, oltre che per l'impermeabilizzazione superiore. Il balloon frame caratterizza l'altro tipo costruttivo (vedi Fig.5b): le chiusure verticali sono su telai a montanti di legno lamellare incollato (sez. 16x6cm), con una pelle stratificata spessa 40cm (vedi Fig.5c) Uguali strati si usano nelle chiusure orizzontali inferiore e superiore, oltre il manto impermeabile nell'ultima e nella prima il pavimento sopraelevato di 15cm dalla quota esterna. I due modelli, ciascuno simulante i due sistemi, sono uguali per dimensioni e tipo

costruttivo; mutano per tipi di vetri, pur con uguali telai in alluminio a TT (vedi Fig.6, punto 8).

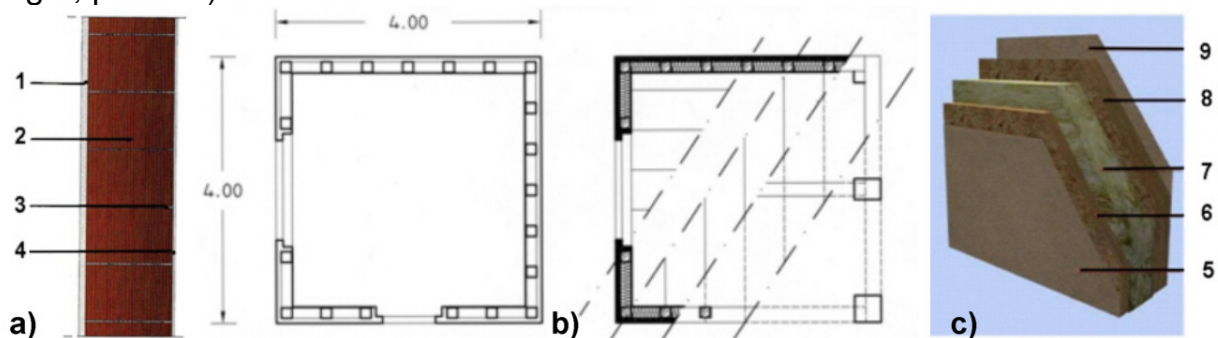


Fig. 5. a) Sezione di parete in laterizio alveolare. Esterno: 1. intonaco termico (sp. 3cm, densità ρ 720Kg/mc) e pittura, 2. blocchi di laterizio alveolare (sp. 30cm, ρ 700Kg/mc), 3. malta cementizia, 4. intonaco termico e pittura. b) Pianta della test room, sistema a balloon frame: sinistra, schema di telaio; destra, sequenza costruttiva dalla base alle pareti. c) Sezione di parete in pannelli di fibra di legno. Esterno: (5) intonaco termico (sp. 3cm); (6) pannello in fibre orientate di legno di pino (sp. 6cm); (7) lana di roccia basaltica (sp. 16cm, ρ 40Kg/mc); (8) pannelli in fibre orientate di legno di pino (sp. 6+6cm); (9) intonaco termico (sp. 3cm).

Il SageGlass® s'impone su (vedi Fig.6a): (1) lastra di vetro temprato trasparente (sp. 6mm); (7) film multilayer sottile di ossidi metallici; (2) intercapedine satura al 90% di Argon (sp. 12,7mm); (6) distanziatore in acciaio inox, a sezione cava con sali disidratanti; (4) sigillante di prima barriera; (5) isolamento del bordo inferiore del distanziatore con sigillante siliconico e PIB; (3) lastra di vetro chiaro temprato (sp. 6mm). I fattori luminosi ed energetici sono quelli forniti dalla Electrochromics Inc. per l'IGU.

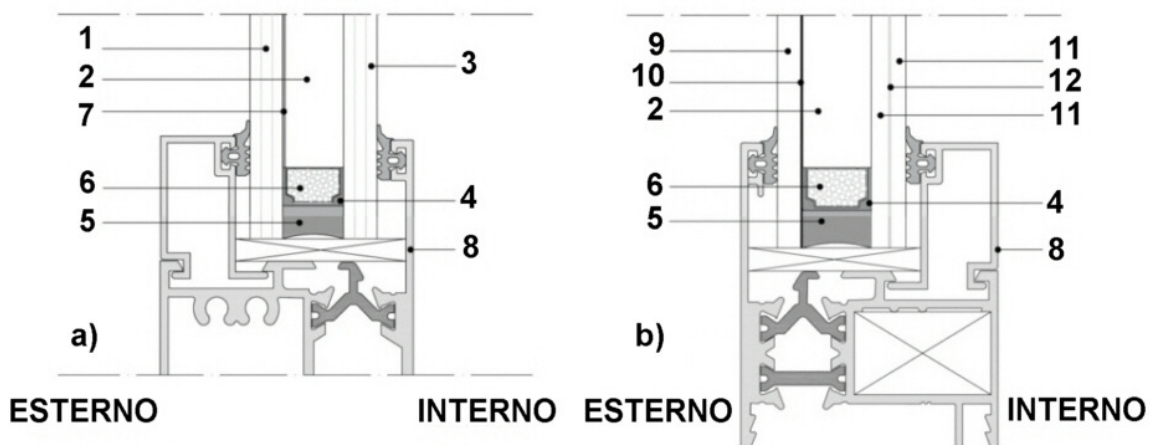


Figure 6. Telai con: a) pannello SageGlass®; b) pannello P-Planitherm®

Per il low-E (vedi Fig.6b) s'impone un P-Planitherm di vetrocamera isolante (sp. nom. 22,4mm), con (9) lastra Planitherm® Saint Gobain (sp. 4,0mm), basso-emissivo per il Planitherm Ultra N® (10); (2) intercapedine satura di Argon al 90% (sp. 12,0mm); (6) distanziatore in acciaio inox a sezione cava e sali disidratanti; (4) sigillante di prima barriera (sp. $\geq 0,5$ mm); (5) intercalare a 3mm almeno dal bordo inferiore del distanziatore, protetto da sigillante butilico di seconda barriera. La superficie interna è un vetro stratificato di (11) due lastre Planilux® Saint Gobain (sp. 3+3mm) con interposto (12) un film di PVB (sp. 0,38mm).



Figure 7. a) Rendering del modello; b) edificazione in corso delle test room (2011)

L'analisi sui due sistemi di ogni modello si avvale dell'Ecotect Analysis-Autodesk®-2010, procedendo dal modello tridimensionale delle camere (vedi Fig.7a). Nella simulazione, oltre ai dati orari e climatici locali al 2010, s'impostano gli input di superfici trasparenti ed opache per i tipi costruttivi. Per unicità delle test room (dimensioni limitate, assenza di utenti) si esclude la climatizzazione forzata, valutando le ore di comfort coerente ai soli apporti d'involucro. Le specifiche energetiche di vetri EC e low-E rilevano, per l'apertura a S-E, sui dati dell'involucro a muratura portante, secondo il trend della radiazione solare che, nelle sue componenti, è espresso nel corso della giornata più calda dell'anno (15 luglio). L'analisi di dettaglio dei parametri secondo i termini percentuali del valore di picco alle 11,00 a.m., rivela una frazione trasmessa d'energia radiante ridotta del 99% circa dai vetri low-E ai vetri EC completamente colorati. La riduzione si attesta al 35% dai low-E agli EC in stato chiaro, crescendo al 98,9% da vetri EC in stato chiaro a quelli in stato colorato. Analogamente rileva la riduzione della quota d'energia radiante assorbita: 46% da low-E a vetri EC in stato colorato; 58% per quelli EC da stato colorato a chiaro. Per la frazione incidente d'energia radiante, l'analisi simulata contestualmente anche per tutto l'anno, indica il massimo in agosto (48.764 Wh/m^2), costante nei tre casi. Per la quota assorbita d'energia radiante, al picco d'agosto si registrano: 5.415 Wh/m^2 per i low-E; 7.332 Wh/m^2 per i vetri EC in stato chiaro; 3.011 Wh/m^2 per gli EC colorati. Significativa è la frazione trasmessa: 22.061 Wh/m^2 per i low-E; 14.620 Wh/m^2 per i vetri EC in stato chiaro; 133 Wh/m^2 per gli EC colorati. Si attestano i valori precedenti (15 luglio). Ripetuta l'analisi per le A e B per il balloon frame, i risultati sono ancora confrontabili con i precedenti, rilevando una notevole riduzione della frazione d'energia trasmessa per i vetri EC allo stato colorato, dell'ordine del 99% rispetto ai low-E.

In sintesi (vedi Tabella 1) le percentuali dei guadagni che, in un anno solare, si devono ai fattori lì designati, rivelano su valori incidenti di apporti interni (corpi illuminanti, dispositivi, strumentazioni, utenza), avulsi da involucro e suoi componenti. Le prestazioni energetiche globali dell'edificio e quelle particolari dei suoi elementi influenzano la radiazione solare diretta, riducendosi il valore del 99% circa tra i low-E e i vetri EC allo stato colorato. Ulteriore interesse rileva l'analisi sui differenti tipi d'involucro: l'uso e l'efficacia dei vetri EC pare indipendente dal sistema costruttivo, il che ne accresce la versatilità, consentendo di estenderne le applicazioni a vaste tipologie edilizie, allo smorzarsi del rapporto costi-benefici a favore degli ultimi.

	GUADAGNI (%)					
	CONDUZIONE	SOLARE INDIRETTA	SOLARE DIRETTA	CONVEZIONE	APP. INTERNI	FLUSSI ATT.
VETRI LOW-E	0,7	6,5	18,4	2,9	70,7	0,7
VETRI EC (stato chiaro)	0,9	7,0	12,7	3,1	75,6	0,8
VETRI EC (stato compl. colorato)	1,06	8,0	0,1	3,6	86,4	0,9

Tabella 1. Guadagni per conduzione dall'involucro, radiazione solare indiretta e diretta, per convezione, per apporti interni e flussi tra attigue zone termiche.

4. Conclusioni

L'analisi preliminare segnala la performance energetica dei vetri EC, per la capacità dinamica e reversibile di adattare le caratteristiche ottiche e termiche del componente edilizio, secondo il contorno. Essi concorrono, in sinergia a tutti i componenti, ad un involucro di ottimale risparmio energetico e benessere ambientale. Ciò sfruttando gli apporti della radiazione solare, secondo destinazione d'uso ed esigenze dell'utenza, con incisiva riduzione dei consumi. Poiché nel clima mediterraneo le calde estati condizionano i risparmi energetici, per l'esigenza di raffrescare ambienti di lavoro e residenziali, l'indagine sui vetri EC rivela conseguibili condizioni di efficienza energetica, superiori alle tipiche soluzioni convenzionali, motivando l'approfondimento. Per tale persuasione, la ricerca prosegue con l'edificazione delle test room, iniziata questa estate 2011, e avviando la sperimentazione sui modelli costruiti per verificare l'attendibilità dei dati d'analisi. L'approfondimento ulteriore riguarda sia la tipologia dello involucro, secondo il rapporto tra superficie vetrata e opaca, definendone un range ottimale all'uso dei vetri EC, e testando la versatilità d'uso dei vetri EC nel recupero e nel retrofit energetico del patrimonio edilizio del Moderno.

5. Riferimenti bibliografici

Dorado C., Dossier de productos, in Tectonica, Vidrio (I), Vol. 10, ATC Ediciones, España, Madrid, 2001, 68-106.

Loddo G., Ludoni D., Cossu G.P., Pittaluga M., Contemporary sustainable architecture: knowledge, prospects and testing of EC windows, in Proc. of Engineered transparency, International Conference at Glasstec, Germany, Düsseldorf, 2010a, 391-402.

Loddo G., Ludoni D., Pittaluga M., Cossu G.P., EC glazing in contemporary architecture: examples of use and opportunities for the future, in Proc. of 37th IAHS World Congress on Housing, University of Cantabria, Spain, Santander, 2010b.

Loddo G., Ludoni D., Cossu G.P., Pittaluga M., Pinna M.L., Experimenting and testing the use of EC windows in the Mediterranean climate, in Conference Proc. of the G.P.D., Finland, Tampere, 2011, 415-420.

Arginare la Disa(gia)bilità: proposte tecnologiche innovative per favorire, attivare e riabilitare la comunicazione delle persone disabili

Raffaella Conversano, Gaetano Manzulli¹, Maurizio Binacchi²
Media Educator – Docente Specializzata I grado
I.C. "A. Moro", via Alfieri 12 - 74021 Carosino (TA)
raffaellaconversano@in.it

¹Direttore Dipartimento di Informatica
Docente di Sistemi di Trasmissione ed Elaborazione dati
ITI "Pacinotti" – via L. Trasimeno s/n – 74100 Taranto
gaetano.manzulli@pacinottitaranto.it

²Docente a contratto di "Tecnologie audiovisive applicate all'E-Learning"
Facoltà di Scienze della Comunicazione,
Università degli Studi di Roma "La Sapienza"
ingbinacchi@libero.it

La società contemporanea è caratterizzata da molteplici linee di frattura che segnano distanze e differenze tra gruppi sociali; le condizioni di salute e le menomazioni nelle funzioni e strutture del corpo costituiscono uno dei motivi principali di disuguaglianza e discriminazione cui si associano spesso forme di "stigmatizzazioni" gravemente lesive della stessa dignità della persona. Se partecipare con successo ai processi di apprendimento ed entrare efficacemente in relazione con i coetanei è oggi condizione assolutamente indispensabile per non essere discriminati, di contro si assiste ad una continua identificazione della persona e delle sue potenzialità con la sua diagnosi che, spostando di fatto l'attenzione sul danno ad essa associato, porta a concepire in modo errato la condizione di "disabilità" come una conseguenza della malattia o disturbo della persona. E' per arginare tali errati comportamenti di disagio approccio, che stiamo operando ad un intervento progettuale specifico che non operi per compensare ciò che con la disabilità manca nell'individuo, ma per potenziarne ciò che la stessa ha risparmiato integrando, all'interno di un percorso software/hardware, un sistema in grado di far interagire la persona con l'ambiente informatico virtuale per prevenire/compensare/alleviare o eliminare la menomazione e lo svantaggio che da esso ne deriva. Nato come ricerca complessa di soluzioni con l'apporto integrato di varie competenze soprattutto tecnologiche, esso mira alla soluzione ottimale di "universalità" degli ausili collegato alle

condizioni di vita del disabile, in un processo di sviluppo delle sue abilità e competenze senza introdurre altri elementi di recisione e/o di contrapposizione fondendo, in un connubio operativo che ne argini le difficoltà di approccio, il sistema cognitivista con quello costruttivista, ridefinendo la visione di empirismo teorico formulata da de Kerckhove come psicotecnologia in pragmatismo tecnologico del principio di "zona di sviluppo prossimale" di Vygotskij.

1. Introduzione

L'uso esplicitamente teorico e contenutistico delle tecnologie dell'istruzione in ambito scolastico, ancora oggi, elude la reale finalità della loro presenza cioè, far cogliere l'oggettività della mediazione applicativa eolutiva di cui questi sono portatori nelle dinamiche quotidiane di comunicazione, modificando a priori il concetto trasmissivo della scuola; difatti, chiunque può apprendere se messo nella condizione di poter comunicare e decodificare in modo personale i propri codici di interpretazione dei messaggi.

L'importanza di far comprendere appieno ciò che questi strumenti permettono, soprattutto se applicati al campoolutivo dei bisogni di autonomia delle persone con difficoltà, è stata l'ispirazione operativa del nostro lavoro attraverso la rilettura efficace di una teoria che, per molti versi, è ancora "nuova", di provenienza straordinaria e che potrebbe suscitare risvolti ancora imprevedibili nella sua applicazione.

Affermare che la sua applicabilità e la sua rilettura, dal nostro punto di vista possa mettere in crisi tutto ciò che era stato teorizzato e realizzato sino ad ora, per nome e per conto dell'uso delle nuove tecnologie, è a dir poco riduttivo, scuotendo dalle fondamenta quanto sino ad oggi affermato da teorici ed empiristi.

Capire come dentro ciascun individuo etichettato "disabile" dai "normali", indipendentemente dalla gravità delle patologie che li affliggono, siano celate grandi potenzialità di idee e soluzioni per la vita sociale di tutti, è stato il valore preminente che abbiamo scelto di approfondire e trasmettere.

2. Le nuove frontiere

La disabilità non è una scelta: è un trauma che nessuno vorrebbe mai vivere, la cui comparsa sconvolge l'individuo nel profondo della sua soggettività trasformando le relazioni con gli altri.

Difatti, il deficit impone in modo predominante ed immediato la ridefinizione del paradigma relazionale e comunicativo con gli altri dove, non si tratta di negare la menomazione in quanto tale ed il suo oggettivo vincolo che di fatto limita alcune funzioni ed azioni, ma di superare, quando non eliminare del tutto, l'esito di pregiudizi, culture e relazioni sociali incapaci di vedere oltre la menomazione, individui nel senso pieno.

Il contributo che le nuove tecnologie hanno posto in essere in questo campo, ha permesso di superare molti limiti favorendo l'espressione del "sé" in molte dimensioni dove, l'infinita panoplia che queste offrono, soprattutto in campo mediale, comincia lentamente a rivolgere lo sguardo verso il settore della interrelazione comunicativa funzionale all'apprendimento.

Il mondo dei media è cambiato, così come è cambiato il modo in cui ci si avvale della comunicazione per confrontarsi con la realtà dove, modificando la cornice socio-culturale si è trasformato, di seguito, il profilo, lo stile di vita, le abitudini, i modi di essere, di pensare e di comportarsi nel quotidiano degli individui. In questo senso di cambiamento l'innovazione e la "corsa" tecnologica, se da un lato hanno consentito una democratizzazione culturale, per l'aumento delle opportunità di accesso alla "competenza" comunicativa, dall'altro stanno segnando inevitabilmente nuovi divari e gap socioculturali.

Capire l'intensità, la radicalità del senso di mutamento che stiamo attraversando significa esplorare i territori di interazione fra processi formativi e percorsi di comunicazione, soprattutto riguardo all'universo giovanile, in una società che anticipa di per sé l'idea stessa di futuro con i suoi sviluppi tecnico-scientifici ed economici, illuminanti per definire la nuova dimensione dell'uomo contemporaneo che, tuffato in un caos di virtualizzazione e dematerializzazione dell'esperienza, vive una sorta di nomadismo nuovo in cui gli individui, pur stando fermi, vedono cambiare tutto intorno a loro.

Disorientamento ed incertezza rendono necessario così riflettere sul mondo attuale della conoscenza e della divulgazione del sapere che, in passato, era prevalentemente rivolto ad assicurare la continuità dell'adattamento alla vita attraverso il passaggio per generazioni del patrimonio di conoscenze acquisite nel tempo, interrogandosi sul ruolo che essa ha assunto o "dovrebbe" assumere in una società in continuo movimento.

Si rende necessario riformare il modo di trasmissione, veicolazione e condivisione dei saperi attraverso una analisi applicativa nuova del nostro modo di conoscere, pensare ma soprattutto insegnare aiutando questa nuova epoca, ancora in gestazione, a vivere nelle nostre menti dando vita a una forma di cultura che si presenti come identità fluida, area di confine in cui emergono differenze e scambi preziosi.

3. La Dis(agi)abilità

Nelle quotidiane realtà scolastiche emergono contemporaneamente problemi legati alle difficoltà di apprendimento, di inclusione ed integrazione mentre di contro, si annaspa nel cercare la soluzione in percorsi di mediazione tra saperi dei docenti e le competenze apportate dagli esperti del settore che operano sul territorio, visti quasi sempre come uniche vere risorse.

Difatti, convinti assertori che "si educa attraverso ciò che si dice, di più attraverso ciò che si fa, ancor più attraverso ciò che si è", la disabilità in sé pone in essere la necessità di elaborare una riflessione condivisa sulle problematiche e la consapevolezza di essa secondo una rilettura in chiave olistica: puntare ai diversi modi e stili di apprendimento come risorsa poiché, la patologia clinica che determina in qualche modo una ulteriore sfumatura di diversità, deve diventare la linea guida da osservare e tenere come punto di riferimento per

calibrare e adeguare il metodo e la didattica, volti a favorire il processo di apprendimento e di insegnamento.

Il non tener conto di ciò o il non rispetto di questa regola basilare, porta a confondere la incapacità di approccio con tutti nella prassi quotidiana di insegnamento, banalizzando la diagnosi clinica come una barriera dell'individuo verso il sociale e chiudendolo nello stereotipo comune di disabilità che noi abbiamo inteso come "dis(agi)abilità"

Infatti, con il termine coniato di "disagiabilità", si ridefiniscono quelli che sono i veri ostacoli che i bambini, gli alunni e le persone disabili in genere vivono e/o incontrano: non tanto il diverso approccio abile con la realtà circostante, cioè "usare" il quotidiano in modo diverso, quanto le ansie, le angosce e le incapacità di approccio che costoro scatenano in chi li circonda, che evidenziano da una parte le difficoltà legate al deficit e, dall'altra gli atteggiamenti di paura e l'ambivalenza dell'ambiente che interagiscono con il grado potenziale di autonomia raggiungibile, pur nella condizione di svantaggio.

Con questo termine si focalizza con maggiore chiarezza l'attenzione sulla osservazione della situazione di difficoltà del soggetto con peculiarità abilitative diverse, per "vedere" il problema e le possibili ulteriori soluzioni di autonomia nella loro globalità, allo scopo di affrontarli al meglio creando uno spazio, anche virtuale, in cui esperienze e professionalità possano trovare la giusta espressione nel rispetto di ruoli e competenze.

Questi concetti sono stati ribaditi nell'intervista radiofonica che la Dr.ssa Conversano ha rilasciato all'interno della trasmissione "Diversi da Chi?" andata in onda su Rai Radio1 il 20/11/2010, il cui podcast può essere riascoltato al link pubblicato nella sitografia.

Il forte punto di riferimento nella revisione di questo nuovo termine indicatore è stato quello di intervenire sulle false interpretazioni comportamentali di chi, suo malgrado, si avvicina al rapporto con realtà umane dalle condizioni abilitative diverse; infatti, in un soggetto disabile, la gravità della sua diagnosi deficitaria, con ritardo più o meno grave, non deve impedire il rispetto e lo sviluppo ampiamente potenziale della sua residua parte funzionale per una corretta vita normale: in altre parole, a causa del suo deficit psichico o fisico che sia, non deve essere individuato "come colui che non è in grado di capire o di vivere" secondo i comuni canoni di normalità.

In campo scolastico il sistema dell'istruzione è investita da una nuova missione: dare vita a nuove forme di cittadinanza dai confini sfumati, mettendo in crisi gli originari obiettivi tesi solo a creare linguaggi omogenei; la scuola, oggi più che mai, ha bisogno di riflettere su ciò che funziona, per individuare quali siano le condizioni, le idee ed i percorsi ottimali da attuare per far incontrare le esigenze di apprendimento degli alunni con la sensibilità e la capacità di insegnamento dei docenti interrogandosi sul ruolo che dovrebbe assumere in una società in continuo movimento.

L'integrazione scolastica delle diverse abilità diventa, quindi, il patrimonio di buone prassi da rivestire con l'abito della sperimentazione - arginare la dis(agi)abilità appunto - rinnovando la spinta creativa tecnologica come propensione metodologica a formulare ipotesi spendibili sul campo, da

sottoporre al vaglio della prova, per rilevarne la validità e/o eventuali misure alternative.

4. Il Progetto

Questo progetto nasce come naturale evoluzione del progetto “Il labirinto incantato” presentato nella edizione Didamatica2010, e del progetto “La Mediateca Virtuale come strumento per lo sviluppo e l'integrazione degli alunni affetti da disabilità”, realizzato con una rete di istituti scolastici di Taranto, e rivolto a circa 50 alunni con disabilità intellettive, ma anche ad alunni con disturbi specifici e con difficoltà di apprendimento (DSA, dislessia, disgrafia, discalculia), dovute spesso ad un ritardo maturazionale, a scarso bagaglio esperienziale o basso investimento motivazionale.

Si tratta, in sostanza, di una sperimentazione di un progetto didattico orientato fortemente all'adozione di specifiche tecnologie informatiche per la didattica e all'uso di piattaforme software tecnologicamente avanzate per favorire l'integrazione di alunni con diverse disabilità intellettive.

La consapevolezza che le modalità diverse ed innovative nell'attività di sostegno possano agevolare la formazione e la didattica consentendo possibilità di recupero mirate e accessibili, ha portato, i docenti autori della iniziativa, a progettare un sistema didattico e tecnologico destinato a perseguire obiettivi specifici qui di seguito indicati:

- Ricerca di soluzioni e sviluppo innovativo di nuove tecnologie della Informazione e Comunicazione applicate alla didattica ;
- Sostegno agli alunni con invalidità permanente o temporanea rendendo l'accessibilità e la fruibilità delle informazioni conformi agli standard W3C;
- Sviluppo dell'autonomia nell'apprendere e coinvolgimento esperienziale e motivazionale della persona disabile.

La logica perseguita nell'ambito progettuale è stata quella di realizzare uno strumento per favorire lo sviluppo dell'autonomia negli apprendimenti di allievi con disabilità che, oltre alle problematiche di apprendimento, presentano difficoltà nelle applicazioni delle conoscenze e nello svolgimento di compiti o di uso di strumenti che non sono in grado di attuare autonomamente.

Il progetto si prefigge lo scopo di potenziare le capacità logiche e di astrazione, consolidando le abilità di base e le capacità operative reali dei soggetti destinatari; attraverso l'utilizzo di una tecnologia innovativa da un lato, ma estremamente familiare alle generazioni attuali dall'altro, permette di diminuire le difficoltà cui l'alunno disabile va incontro, riducendone il divario con il gruppo classe.

In tali allievi, infatti, è spesso presente un basso livello di autostima ed immaturità psico-affettiva e comportamentale per cui, nel realizzare qualunque attività didattica, essi sono di solito frenati dal timore di sbagliare e/o di non riuscirci ricercando, di conseguenza, costantemente la guida dell'insegnante distaccandosi dalla realtà del gruppo classe.

Come sostiene l'Ing. Manzulli nella sua intervista radiofonica rilasciata all'interno della trasmissione "Area di Servizio" del 2 Ottobre 2011 in onda su Rai Radio1, il cui podcast è riascoltabile al link indicato nella sitografia, ogni disabile è "straniero" nel mondo intero in quanto spesso vive il suo isolamento anche tra i propri connazionali. Pertanto, le problematiche affrontate in questo articolo si rivolgono sia ai disabili italiani che stranieri, accomunati in questo intervento didattico.

L'idea progettuale, quindi, si fonda sul presupposto di rendere i contenuti, spesso difficilmente comprensibili o esperibili quando si utilizzano gli approcci didattici tradizionali, più facilmente assimilabili dalla popolazione disabile, privilegiando modalità e situazioni coinvolgenti e usuali fornite dall'uso di tecnologie ormai diventate consuete seppur innovative nella popolazione studentesca; del resto, la comprensione del messaggio didattico, quando non riferito al bagaglio esperenziale in tutte le patologie disabilitanti, è fortemente inficiata dalla labilità attentiva, per cui l'alunno ha bisogno di continui richiami per focalizzare l'attenzione: l'uso invece di una tecnologia "familiare" e comune ai propri coetanei, riduce le distanze tra il disabile ed il mondo che lo circonda, annullando l'isolamento e favorendo la sua integrazione.

Si rileva, inoltre, in tali alunni una bassa prestazione in compiti di memoria di lavoro, pertanto è importante fornire uno strumento che dia l'opportunità di autocontrollo dei risultati e quindi la possibilità di correzione dell'errore in tempo reale, inducendoli a riflettere in merito, in conseguenza dell'immediatezza con cui viene presentato il feed back, positivo o negativo.

In tal modo, quindi, riuscendo autonomamente a manipolare ed a trasformare le informazioni, questo strumento aiuterebbe il fruitore ad agire in modo più autonomo, attivo e costruttivo.

Dal punto di vista della innovazione tecnologica, l'obiettivo è quello di favorire l'apprendimento ricorrendo all'uso delle modalità di comunicazione adottate dai modelli più recenti di telefonia mobile quali gli smartphone poiché, questo strumento, se impiegato in attività di comunicazione e di applicazioni didattiche, consente la realizzazione di un ambiente di apprendimento le cui caratteristiche sono l'immediatezza di comprensione, la facilità di impiego e l'adozione di meccanismi coinvolgenti dal punto di vista motivazionale. Infatti, questa generazione di allievi già usa internet e le nuove frontiere della comunicazione, come naturale spazio di apprendimento e di gioco, tanto che sono stati definiti "digital natives" [Marc Prensky, Digital Game-Based Learning, 2001] in quanto cresciuti all'interno di un mondo già digitale.

Il coinvolgimento dell'allievo disabile nell'utilizzo didattico di tali strumenti consentirebbe, da un lato, la possibilità di abbattere le "barriere digitali" spesso dovute alla non sempre pratica fruizione della tecnologia, in modo da sconfiggere il "digital divide" (isolamento tecnologico) di cui egli è spesso vittima e, dall'altro, di superare l'isolamento personale, adottando un linguaggio e strumenti attuali vicini al modo di comunicare nel quotidiano.

L'uso ormai imprescindibile delle nuove tecnologie della comunicazione consente, quindi, ambienti di apprendimento davvero innovativi riformulando i paradigmi che sono alla base dei processi educativi: se il tradizionale paradigma si basava sul trasferimento della conoscenza dal docente al

discente, nella nuova lettura esso si appoggia sui meccanismi costruttivisti della conoscenza.

Difatti, gli ambienti di apprendimento "virtuali" favoriscono nuovi sistemi di apprendimento basati sui modelli relativistici della conoscenza dove, grazie alla cooperazione ed alla comunicazione basata sull'uso di immagini, suoni e messaggeria elettronica gli studenti scoprono modi creativi per migliorare le proprie conoscenze, utilizzando la comunità intesa come luogo di incontro per la produzione, la distribuzione e la gestione del sapere ma anche come spazio per la socializzazione (social learning) e dimensione virtuale di confronto e dialogo.

La prima esperienza realizzata con l'introduzione dell'uso del cellulare a scuola fu condotta in una classe multietnica variamente composta che vedeva la presenza di allievi con disabilità varie.

Dopo aver condotto una indagine tecnica sui tipi di cellulare che loro avevano e su come funzionava la loro funzione agenda fu creato uno "slang" comune che andasse bene per tutti. In modo del tutto semplice e naturale ci rendemmo conto che gli allievi erano diventati più autonomi (digitare è più facile che scrivere) e tutti non perdevano occasione di impegnarsi in questa nuova metodologia didattica pur di usare il proprio telefono. Inoltre in una successiva esperienza teatrale a scuola, gli allievi messaggiavano le varie correzioni alle battute del copione, procuravano col bluetooth i vari file sonori in mp3, realizzavano le foto del backstage partecipando attivamente alla creazione dello spettacolo.

Tradotto in termini pedagogici, tale metodologia didattica prevede, oltre ad un'attenzione particolare all'aspetto relazionale ed affettivo (con l'attivazione di meccanismi in grado di favorire dinamiche di gruppo), anche l'esposizione dei vari argomenti e percorsi programmati in forma semplificata per facilitare sia l'apprendimento che la conquista dell'autonomia.

Gli insegnanti possono così scegliere i contenuti e fissare gli obiettivi minimi, con interventi specifici coordinati dall'insegnante di sostegno non solo, ma ricorrendo tra le varie metodologie da utilizzare soprattutto al tutoring, alla didattica multimediale, orientativa ed interattiva dal carattere multidisciplinare, possono in particolare svolgere attività di orientamento nel percorso scolastico dei ragazzi diversamente abili.

E' in questa ottica che si sostanzia la nuova missione della scuola: orientare la sua spinta pedagogica come educatività di azione.

In riferimento alla attuale ricerca progettuale, dopo il successo dell'impiego delle tecniche sopracitate, la nostra attenzione è stata posta su alcuni limiti che l'attuale tecnologia smartphone possiede quando gli utenti sono disabili.

Per esempio, per gli ipovedenti la ricchezza grafica di uno smartphone può essere un esercizio scoraggiante in quanto il display risulta essere di dimensioni ridotte rispetto a quello di un computer; analogamente chi è affetto da disabilità motoria degli arti non può usare le mani correttamente, sfruttando quindi la proprietà capacitiva di un display di smartphone (sensibilità alla piccola corrente dei polpastrelli delle dita).

Davanti a queste difficoltà palesi, l'idea progettuale ha visto la realizzazione di una piattaforma software che, attraverso un sistema operativo dedicato,

ricostruisce su un computer o su un tablet PC l'esatta riproduzione di uno smartphone e delle sue peculiari caratteristiche operative. In tal modo, con gli strumenti già disponibili come accessori del PC per disabili (per ogni tipologia di disabilità esiste un ausilio specifico, uno strumento capace per agevolare i movimenti o migliorare la visione dello schermo), tutti potranno utilizzare funzioni come navigare con il touch del mouse mentre a video tutto sarà esattamente come nello schermo del telefono: ogni funzione, compreso quella telefonica potrà essere attivata dal PC, grazie ad un opportuno interfacciamento con schede SIM.

L'interfacciarsi con il computer è una cosa per niente stupefacente. Sono tanti i programmi che consentono già oggi di pilotare i più comuni cellulari con il computer, addirittura Mac OS X consente di scrivere direttamente da tastiera SMS con tantissimi telefoni Bluetooth ma, per un disabile, la semplicità e la flessibilità dell'interfaccia di uno smartphone direttamente sul computer significherebbe poter utilizzare il cellulare grazie ai sistemi di controllo o visualizzazione che già usa abitualmente, senza subire le limitazioni delle soluzioni oggi disponibili.

5. Conclusioni

Studiare ed indagare come le tecnologie possano aiutare e supportare il mondo della scuola e del Welfare in generale, predisponendo al meglio i percorsi di inclusione e di integrazione dei soggetti con difficoltà abilitative e di apprendimento non può essere, comunque, limitato ed affrontato solo ed esclusivamente da una prospettiva tecnologica; è necessario valutare la storia delle persone, il superamento delle varie barriere, quella delle relazioni e della comunicazione in primo luogo, le lontananze culturali, gli approcci alla disabilità che ne modifica le aspettative ed il comportamento dei familiari, valutare le diagnosi fatte e gli altri numerosi elementi che necessitano di approfondimento e organizzazione per predisporre percorsi inclusivi.

La scoperta attraverso l'esperienza è un cammino complesso e difficile, pieno di ostacoli, impedimenti e rallentamenti, non esiste alcuna garanzia di successo né guide capaci di suggerire percorsi da seguire, la pratica e l'azione sono le uniche chiavi accessibili per procedere oltre il limite, contribuendo a consolidare e a definire le attività sperimentali già intraprese.

Gli unici prerequisiti per continuare questo iter sono l'impegno, la costanza, la volontà, la disponibilità al rischio e a mettersi continuamente in discussione, sapendo da dove si parte ma ignorando dove si potrà arrivare, dove l'esplorazione è certamente la prima fase dell'apprendimento e dell'esperienza, ma ad essa tuttavia, dovrà far seguito l'impegno e la continuità che si pongono alla base del sapere.

"C'è un solo bene...la conoscenza, ed un solo male...l'ignoranza" affermò Socrate, del resto, se l'homo sapiens sopravviverà al XXI secolo, ciò dipenderà molto dal modo con cui avrà imparato ad imparare, interrogandosi sul ruolo che deve assumere in una società in continuo movimento.

E sembra oggi quasi naturale concludere con una frase di Steve Jobs tratta dal suo indimenticato discorso ai laureandi della Stanford University nel 2005: “Se non avete ancora trovato ciò che fa per voi, continuate a cercare, non fermatevi!”. Un suggerimento regalatoci dall'uomo che ha inventato il futuro rendendo più facile la nostra vita e che oggi sentiamo di fare nostro per dare migliore qualità alla vita dei nostri amici meno fortunati.

Bibliografia

- [1] ALBANESE O., FERRARI E., FIORILLI C., GARBO R., SALA R., (a cura di) Ricerche per fare. Percorsi di mediazione e di inclusione. Edizioni Junior, luglio 2009
- [2] CONVERSANO R., Interagire per crescere – Interazione tra Media e Formazione. EDIZIONI PUGLIESI, 2005
- [3] CONVERSANO R., Progetto di sperimentazione in M.E. – La Media Education nella scuola elementare. EDIZIONI PUGLIESI, 2005
- [4] CONVERSANO R., Efficacia dell'uso delle Tecnologie nel processo di insegnamento-apprendimento. Rimini 11/13 novembre 2005 Atti Convegno Internazionale “La qualità dell'integrazione scolastica”
- [5] de KERCKOVE D., La coscienza planetaria, in «Mass Media. Rivista bimestrale di comunicazione», VI, n. 1, 1987.
- [6] de KERCKOVE D., L'estetica dei media e la sensibilità spaziale, in «Mass Media. Rivista bimestrale di comunicazione», IX, n. 4, 1990.
- [7] de KERCKOVE D., Skin of culture, trad. it di M. Carbone, La pelle della cultura: un'indagine sulla nuova realtà elettronica, Genova, Costa & Nolan, 1996.
- [8] VYGOTSKIJ L., Pensiero e linguaggio. Ricerche psicologiche, a cura di L. Mecacci, 10^a ed., Roma-Bari, Laterza [1990], 2008.
- [9] VYGOTSKIJ L., Il processo cognitivo - Raccolta di scritti a cura di Michael Cole, Sylvia Scribner, Vera John-Steiner, Ellen Souberman, Ed. Bollati Boringhieri, 1987-2002
- [10] VYGOTSKIJ L., Immaginazione e creatività nell'età infantile, Editori Riuniti, Roma 1972
- [11] SCHIANCHI M., La terza Nazione del Mondo. I disabili tra pregiudizio e realtà, Editore Feltrinelli, Serie Bianca Feltrinelli Milano 2009
- [12] A cura di DE POLO G., PRADAL M., BORTOLOT S., ICF-CY Servizi per la Disabilità, FrancoAngeli Milano 2011
- [13] Marc Prensky, Digital Game-Based Learning, 2001

Link utili:

RAI RADIO 1 : “Area di servizio” - Intervista a Gaetano Manzulli

<http://www.rai.it/dl/radio1/2010/popup.html?t=Area%20di%20Servizio%20del%20%20ottobre%202011&p=Area%20di%20Servizio%20del%20%20ottobre%202011&d=&u=http%3A%2F%2Fwww.radio.rai.it%2Fpodcast%2FA41316018.mp3>

RAI RADIO 1 : “Diversi da chi?” – Intervista a Raffaella Conversano

www.radio.rai.it/radio1/diversidachi/view.cfm?Q_EV_ID=321903

HANDIMATICA 2010 – Seminario Migranti e disabili: tecnologia delle mediazioni e dei mediatori

<http://www.youtube.com/watch?v=DCU0A74cwfo>

Aspetti di Sicurezza nelle Reti per la Safety

Marco Cereia, Stefano Scanzio

CNR-IEIIT

C.so Duca degli Abruzzi, 24; I-10129 Torino, Italia

{marco.cereia, stefano.scanzio}@ieiit.cnr.it

L'importanza dei sistemi di safety negli ambienti industriali è cresciuta costantemente negli ultimi anni. Un aspetto particolarmente interessante di questi sistemi riguarda il modo in cui i relativi dispositivi possono essere connessi in maniera semplice, economica ed efficace negli impianti. La tendenza è adottare tecniche di comunicazione digitali convenzionali, in grado di trasmettere sullo stesso cavo sia i dati safe sia quelli standard, esponendo però i sistemi di safety allo stesso tipo di rischi di sicurezza informatica di cui soffrono le comuni reti d'ufficio. Questo contributo affronta alcuni problemi di sicurezza che possono affliggere le reti di comunicazione per la safety e, in particolare, il protocollo SafetyNET p, mostrando come i protocolli di safety non solo siano esposti alle stesse minacce di sicurezza delle reti convenzionali, ma possano addirittura mostrarsi più sensibili di queste in situazioni specifiche, come ad esempio nel caso di attacchi di tipo denial of service (DoS).

1. Introduzione

La sicurezza funzionale, anche conosciuta col termine “safety”, è oggi un aspetto sempre più importante negli ambienti di produzione industriali. Di conseguenza, è necessario poter disporre di contromisure opportune per evitare che guasti inevitabili, anche ad un singolo componente dell'impianto, possano arrecare danni alle attrezzature, indurre rischi reali per l'incolumità delle persone o provocare disastri ambientali.

È bene rimarcare fin da subito che il termine italiano “sicurezza” può portare ad ambiguità di interpretazione quando utilizzato per i moderni impianti industriali. Difatti, in tale contesto sono rilevanti sia la sicurezza funzionale (safety), ovvero la capacità del sistema di non arrecare danni a persone o cose, sia la sicurezza informatica (security), ovvero la capacità del sistema di evitare attacchi di tipo intenzionale. Nel primo caso possiamo affermare, in modo forse semplicistico ma efficace, che la safety protegge l'uomo dalle macchine (in caso di guasti alle stesse), mentre nel secondo la security protegge le macchine dall'uomo (in caso di comportamenti malevoli). Nel seguito di questo contributo, per indicare la sicurezza funzionale sarà utilizzato il termine safety.

Ci si può chiedere come mai solo oggi si è giunti a considerare in modo congiunto le problematiche di safety e di security negli ambienti di produzione automatizzati. La risposta è che, sempre più i relativi apparati di controllo si

basano su soluzioni (piattaforme di calcolo, sistemi di comunicazione, formato dei dati, metodologie di sviluppo) che derivano, più o meno direttamente, dal mondo ICT, e ne ereditano tanto i vantaggi (riduzione dei costi, maggiore compatibilità, un livello di integrazione impensabile fino ad alcuni anni or sono) quanto i difetti (nel caso considerato, una potenziale vulnerabilità verso gli attacchi informatici).

Nel caso di sistemi che si estendono su aree di grandi dimensioni, è necessario prevedere un apposito sistema di collegamenti in grado di supportare la trasmissione dei segnali inerenti alla *safety* dai sensori a uno o più controller e, da questi, agli attuatori. Recentemente, l'avvento delle reti *real-time Ethernet* (RTE) ha favorito l'introduzione sul mercato di diversi prodotti per la *safety* (e delle relative tecnologie) che sfruttano Ethernet per la comunicazione. La maggior parte di queste soluzioni può essere vista come l'estensione di specifiche soluzioni industrial Ethernet, volta a supportare caratteristiche legate alla *safety*.

Combinare comunicazioni *safe* (cioè, legate alle funzionalità di *safety*) e *unsafe* (tutte gli altri tipi di comunicazioni, inclusi i dati di processo in tempo reale e quelli di parametrizzazione) ha diversi vantaggi, il principale dei quali è la possibilità di semplificare enormemente i cablaggi riducendo di conseguenza i costi di installazione. Sfortunatamente, dal momento che le caratteristiche delle comunicazioni *safe* e *unsafe* sono molto diverse fra loro, questo approccio introduce una serie di problemi che devono essere efficacemente risolti in vista di un impiego su vasta scala di questo tipo di soluzioni nell'industria. Infatti, nel primo caso il requisito fondamentale è quello di portare il sistema in uno stato *safe* entro un tempo massimo prestabilito e con una probabilità elevatissima (in teoria, con la completa certezza) non appena un malfunzionamento viene rilevato in una parte dell'impianto. Nel secondo caso, invece, bisogna soddisfare i vincoli prestazionali tipici delle reti di controllo in tempo reale, che includono ad esempio tempo di ciclo, *jitter*, capacità trasmissiva (dati di processo scambiati per unità di tempo), nonché, flessibilità, configurabilità, *availability* e così via. In particolare, l'*availability* rappresenta la frazione del tempo totale per la quale il sistema opera in modo corretto (cioè, svolge in modo adeguato l'attività produttiva).

Far sì che questi due tipi di informazioni (*safe* e *unsafe*) condividano lo stesso supporto trasmissivo, espone i sistemi di *safety* a possibili minacce di sicurezza, cosa impossibile quando le reti di *safety* erano fisicamente separate dalla rete di fabbrica o quando i sistemi di automazione erano concepiti come sistemi autonomi e isolati. L'obiettivo principale di questo contributo è analizzare come un comportamento malizioso (ovvero, un attacco informatico alla sicurezza) possa influenzare una soluzione reale di *safety*, ovvero il protocollo SafetyNET p [Safety Network, 2009]. Questo argomento è stato recentemente trattato in [Akerberg e Bjorkman, 2009] e [Hansen, 2009].

Il contributo è così strutturato: la Sezione 2 introduce il concetto di *black channel* e descrive sinteticamente il protocollo SafetyNET p. La Sezione 3 illustra la configurazione sperimentale utilizzata per analizzare il comportamento del protocollo, mentre la Sezione 4 fornisce alcuni dettagli sulla sequenza di

messaggi scambiati nella rete. In fine, la Sezione 5 tratta alcuni problemi di sicurezza riscontrati sul protocollo e la Sezione 6 riepiloga i risultati ottenuti.

2. Comunicazioni safe

La capacità di garantire un certo livello di *safety* in un sistema distribuito di automazione, indicato nella specifica IEC 6158-1 [IEC, 2010b] come *safety integrity level* (SIL), richiede che anche i guasti inerenti al sottosistema di comunicazione vengano prontamente ed efficacemente rilevati. Difatti, il livello di *safety* complessivo è determinato fondamentalmente dall'elemento meno affidabile della catena di controllo. L'incapacità di scambiare dati correttamente in una rete reale può essere legato tanto a cause temporanee (ad esempio, un'interferenza elettromagnetica di entità tale da causare errori di ricezione) quanto a cause permanenti (ad esempio, un cavo interrotto o un connettore rotto). È bene notare come, nonostante il tasso di errore tipico nelle reti cablate sia in genere piuttosto basso, gli errori di trasmissione sono eventi inevitabili nel mondo reale e, a maggior ragione, negli ambienti industriali caratterizzati dalla presenza di notevole rumore elettromagnetico.

Nel caso di dispositivi preposti all'elaborazione dell'informazioni di *safety* (ad esempio, un *safe PLC*), si adottano normalmente delle architetture interne ridondate (cioè, in cui esistono più istanze replicate delle funzionalità h/w) al fine di soddisfare i requisiti di *safety* complessivi dell'applicazione. Invece, sebbene questo possa sembrare poco intuitivo, la ridondanza dei supporti trasmissivi non è strettamente necessaria al fine di realizzare una rete di comunicazione *safe*.

L'approccio normalmente utilizzato per implementare comunicazioni *safe* su canali intrinsecamente *unsafe* (di fatto, le reti industriali convenzionali) è conosciuto come "*black channel*". Da un punto di vista pratico, uno strato di *safety* viene aggiunto tra il sistema di comunicazione sottostante e i programmi applicativi. Tale strato ha il compito di implementare le funzioni di trasmissione legate alla *safety* e di monitorare l'integrità del canale di comunicazione. Così facendo, è possibile soddisfare anche i requisiti di *safety* più stringenti (in ambito industriale), come quelli previsti dal livello SIL 3.

Tale idea è stata inizialmente introdotta nell'ambito dei meccanismi di segnalazione ferroviaria ed è formalizzata nella specifica IEC 62280-1 [IEC, 2002]. Questa tecnica offre due vantaggi: primo, è possibile sfruttare le tecnologie di comunicazione esistenti (molto spesso già presenti negli ambienti di fabbrica); secondo, dispositivi *safe* e *unsafe* possono essere fisicamente connessi allo stesso supporto trasmissivo. Tutto questo porta a vantaggi notevoli in termini di riduzione dei costi. In aggiunta, tutta una serie di funzioni non legate alla *safety* possono essere rese disponibili anche ai dispositivi *safe*. Per esempio, poter disporre di funzioni avanzate di diagnostica (che, a rigore, non sono legate direttamente alla *safety*) può migliorare da un lato l'*availability* dell'impianto e, dall'altro, rendere più efficaci le contromisure per la *safety* riducendo la probabilità di guasto dei singoli componenti (manutenzione preventiva e reattiva).

L'approccio del *black channel* si basa su un certo numero di meccanismi, che possono essere usati per rilevare diversi tipi di errori di comunicazione. Nel

Errori di comunicazione	Misure di safety				
	Numeri di sequenza	Time-out	Identificatore di connessione	Integrità dei dati	Integrità dei dati alternativa
Corruzione	-	-	-	X	-
Ripetizione accidentale	X	-	-	-	-
Sequenza incorretta	X	-	-	-	-
Perdita	X	X	-	-	-
Ritardo inaccettabile	-	X	-	-	-
Inserimento	X	-	X	-	-
Camuffamento	X	-	X	-	X
Indirizzamento	X	-	X	-	-
Errori nella memoria degli switch	X	-	X	X	-

Tab.1 – Misure per il rilevamento degli errori di comunicazione

caso di SafetyNET p, gli errori che possono essere rilevati dalle misure di safety messe in atto dal protocollo sono specificati in Tab.1.

2.1 Il protocollo SafetyNET p

SafetyNET p è un protocollo di comunicazione aperto introdotto da Pilz GmbH & Co. KG, [IEC, 2011] che si basa sulla tecnologia di trasmissione Ethernet ed è adatto sia per lo scambio di dati di controllo in tempo reale sia per le applicazioni orientate alla *safety*. Le applicazioni di *safety* trasferiscono i propri dati (*safe*) su un canale (*safe channel*) certificato secondo la specifica SIL 3 dello standard IEC 61508-1 [IEC, 2010b], reso disponibile dal protocollo sullo stesso cavo utilizzato per le comunicazioni *unsafe*.

SafetyNET p distingue fra due distinti modelli di comunicazione: *Real Time Frame Line* (RTFL), utilizzato per scambio di dati ad alta velocità all'interno di una stessa cella di lavorazione e *Real Time Frame Network* (RTFN), utilizzato per connettere il sistema di automazione alla rete di ufficio o per consentire la comunicazione fra celle di automazione distinte, allorché i dati da scambiare abbiano requisiti laschi di tempo reale (oltre 1 ms). Il livello di trasporto utilizzato da RTFN è UDP/IP, e i dispositivi in rete sono identificati per mezzo del loro indirizzo IP. Poiché le differenze fra RTFN e RTFL si concentrano principalmente a livello trasporto, in questo contributo verrà considerata solo il modello di comunicazione RTFN. SafetyNET p implementa due canali distinti per lo scambio dei dati: il canale ciclico, o *Cyclic Data Channel* (CDC) è usato per trasferire dati di processo con cadenza ciclica. Il canale dei messaggi, o *Message Channel* (MSC) è usato per le comunicazioni acicliche, dove le tempistiche di scambio non sono critiche, come ad esempio nel caso di informazioni concernenti la diagnostica o la parametrizzazione dei dispositivi.

Per lo scambio di dati ciclici, RTFN adotta un modello *publisher/subscriber*, dove ogni singolo *publisher* produce dati per uno o più consumatori. Ogni dato

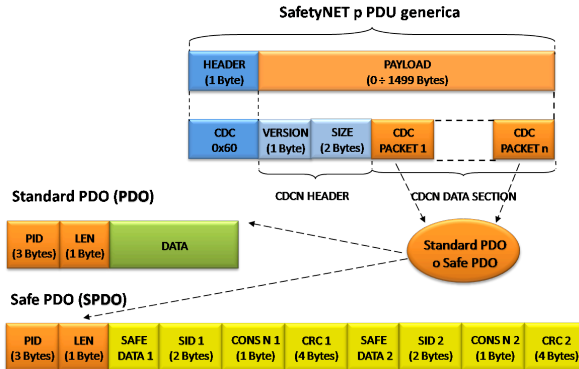


Fig.1 – Struttura generale di una PDU SafetyNET p e di un messaggio CDC

è trasmesso mediante connessioni punto-punto distinte per ogni *subscriber*. In RTFN, ogni *Protocol Data Unit* (PDU) di SafetyNET p è incapsulata in un *datagram* UDP con porta di destinazione 0x9C40.

La struttura generale delle PDU SafetyNET p [IEC, 2010a] consiste di una intestazione (*PDU header*) seguita da un campo dati (*PDU payload*) come mostrato in Fig.1. L'intestazione contiene un codice per identificare il tipo di canale usato per la comunicazione o, in alternativa, il messaggio di gestione della comunicazione contenuto nella PDU. Il contenuto del campo dati dipende direttamente dal codice specificato nell'intestazione.

I messaggi di gestione della comunicazione sono usati per effettuare l'inizializzazione della rete e per stabilire le connessioni punto-punto fra i dispositivi. Tra i messaggi definiti in questa classe, particolarmente importante è la coppia "*subscribe request/acknowledge*", utilizzata dai dispositivi per effettuare la sottoscrizione ad uno specifico dato pubblicato da un nodo remoto. Il *payload* nella richiesta di sottoscrizione contiene la lista dei PID per i quali il dispositivo chiede la sottoscrizione, mentre il *payload* nel messaggio di conferma contiene la lista dei PID per i quali il *publisher* ha accettato la richiesta di sottoscrizione.

Il *payload* dei messaggi ciclici CDC è formato da una sezione dati che contiene la sequenza di pacchetti che devono essere trasmessi sul canale ciclico, preceduta da un'intestazione (*CDC header*), che contiene informazioni relative alla versione del protocollo e alla dimensione della sezione dati, come mostrato in Fig.1. Il canale *safe* per scambi di dati ciclici è implementato a questo livello: in base al tipo di dati da trasferire, ovvero *safe* o *standard (unsafe)*, la struttura dei pacchetti contenuti nella sezione dati varia, come mostrato in Fig. 1.

Nel caso di comunicazioni standard, i dati di processo sono trasmessi per mezzo di *Process Data Object* (PDO), la cui struttura è particolarmente semplice: un campo su 3 byte che codifica il PID associato al dato di processo da trasmettere, un campo su 1 byte che contiene la lunghezza (LEN, in byte)

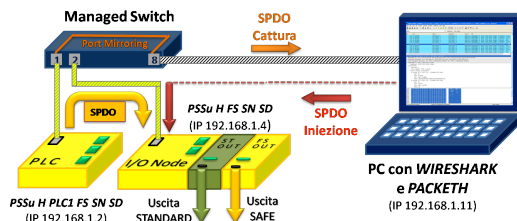


Fig.2 – Configurazione sperimentale

del singolo PDO, inclusiva anche dei campi PID e LEN ed un campo di dimensione variabile, che contiene il dato di processo vero e proprio (*data*).

Nel caso delle comunicazioni *safe*, i dati di processo sono trasmessi per mezzo di *Safe Process Data Object* (SPDO), con una struttura un po' più complessa: in aggiunta ai campi di un PDO, sono presenti un campo su 2 byte che codifica il *Safe Identifier* (SID) del trasmettitore, un campo su 1 byte che codifica un numero progressivo, o *Consecutive Number* (CN), ed un campo su 4 byte che contiene un codice di controllo su 32-bit (CRC), che copre i campi PID, LEN, *safe data*, SID e CN. Infine, i campi *safe data*, SID, *consecutive number* e CRC sono replicati, per rendere il messaggio ridondante.

I campi che differenziano un SPDO da un PDO sono stati introdotti al fine di implementare alcune delle misure di *safety*, presentate in Tab.1, volte a rilevare gli errori di trasmissione. L'identificatore di connessione è implementato per mezzo della coppia (PID, SID), mentre il numero di sequenza corrisponde al campo *consecutive number*. Il campo CRC implementa il meccanismo base per assicurare l'integrità dei dati, mentre il meccanismo di integrità dei dati alternativa è implementato per mezzo dei campi ridondati (*safe data*, SID, *consecutive number* e CRC).

3. Configurazione sperimentale

Il sistema sperimentale utilizzato per analizzare il comportamento del protocollo, rappresentato in Fig.2, è costituito da due nodi SafetyNet p, formati rispettivamente da un PLC (PSSu H PLC1 FS SN SD) e da un dispositivo di I/O (PSSu H FS SN SD). Entrambi i dispositivi, prodotti da Pilz GmbH & Co. KG, reagiscono al rilevamento di un errore sul canale *safe* portando a zero le uscite *safe*. Al dispositivo di I/O sono stati collegati due moduli con uscite digitali, uno per segnali standard e l'altro per segnali *safe*. I due nodi sono stati connessi per mezzo di uno switch di rete con supporto al *port mirroring*. Lo switch è stato configurato per replicare il traffico che attraversa la porta 1, alla quale è collegato il PLC, sulla porta 8, alla quale è stato collegato un Personal Computer (PC). Sul PC sono stati installati l'analizzatore di rete Wireshark (a cui è stato aggiunto il plug-in sviluppato da Pilz GmbH & Co. KG per la dissezione del protocollo SafetyNET p) ed il generatore di pacchetti ethernet packETH. Con questa configurazione è possibile sia analizzare il traffico di rete

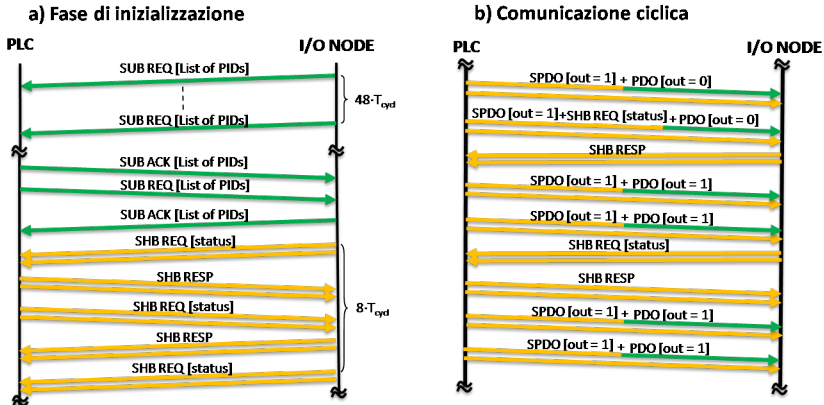


Fig.3 – Flusso dei messaggi scambiati tra il PLC ed il nodo di I/O

durante la comunicazione fra i due nodi, sia iniettare nella rete pacchetti mirati, per studiare gli effetti provocati sull'*availability* del sistema dalla presenza di traffico anomalo nella rete.

Il PLC esegue un'applicazione ciclica, che genera un'onda quadra su una delle uscite digitali standard del nodo di I/O e, allo stesso tempo, mantiene a livello logico alto una delle uscite digitali safe dello stesso nodo. Per l'applicazione è stato scelto un tempo ciclo T_{cycl} di 10 ms e l'onda quadra è stata generata con periodo pari a $12 \cdot T_{cycl}$ e duty cycle del 50% (il valore dell'uscita viene invertito ogni 6 cicli).

4. Analisi della comunicazione

Per comprendere meglio i meccanismi di una comunicazione basata sul protocollo SafetyNET p è stato analizzato il flusso dei messaggi tra il PLC ed il nodo di I/O utilizzando la configurazione sperimentale descritta nella Sezione 3.

Appena alimentati, i dispositivi entrano in una fase di inizializzazione, rappresentata in Fig.3-a, durante la quale il nodo di I/O invia messaggi *Subscribe Request* (SUB REQ) al PLC per richiedere la sottoscrizione ai PDO ed ai SPDO di cui necessita. I messaggi SUB REQ contengono la lista dei PID che identificano univocamente i PDO ed i SPDO richiesti. Nella comunicazione analizzata, il dispositivo di I/O sottoscrive quattro tipi di messaggio pubblicati dal PLC: i messaggi che contengono lo stato logico a cui portare l'uscita digitale *safe*, i messaggi che contengono lo stato logico a cui portare l'uscita digitale standard e i messaggi *Safe Heartbeat Request* (SHB REQ) e *Safe Heartbeat Response* (SHB RESP). L'ultima coppia di messaggi è utilizzata durante una comunicazione *safe* tra due nodi per monitorare lo stato della loro connessione: il messaggio SHB REQ è inviato da un nodo per comunicare il suo stato corrente (*bootup*, *pre-operational*, *operational*) al destinatario e l'avvenuta ricezione viene confermata mediante l'invio di un messaggio SHB RESP. Quando il PLC entra nello stato *pre-operational*, conferma le richieste di

sottoscrizione inviate dal nodo di I/O trasmettendogli un messaggio SUB ACK che contiene la lista di PID richiesti. Allo stesso tempo, il PLC inizia ad inviare messaggi SUB REQ al dispositivo di I/O per sottoscrivere la ricezione dei suoi messaggi SHB REQ e SHB RESP. Non appena il nodo di I/O trasmette al PLC una messaggio di conferma con la lista dei PID richiesti, i due dispositivi iniziano a scambiarsi messaggi di tipo SHB, utilizzando i corrispondenti PID che si sono comunicati durante la fase di sottoscrizione. I messaggi SHB hanno la struttura di un SPDO (rappresentata in Fig.2) e, come si può notare in Fig.3-a, vengono inviati in duplice copia.

Quando il PLC entra nello stato *operational*, ha inizio la comunicazione ciclica, rappresentata in Fig.3-b. Ad ogni ciclo il PLC invia un messaggio che contiene sia il SPDO che il PDO per pilotare rispettivamente l'uscita digitale *safe* e quella standard. Come per i messaggi SHB, ogni SPDO è inviato due volte. Quando gli istanti di invio di un messaggio SHB REQ e di un messaggio che contiene un SPDO coincidono, entrambi vengono trasmessi all'interno dello stesso frame UDP.

5. Problemi di sicurezza

Durante una comunicazione sul canale *safe* il sistema deve essere in grado di rilevare errori e reagire portandosi in uno stato *safe* entro un dato tempo massimo. Ogni volta che il sistema si porta nello stato *safe*, il processo di produzione viene fermato. Questo comportamento è accettabile se l'evento che ha causato la reazione è un errore reale, ma se tutti i componenti del sistema stanno funzionando correttamente e la causa della transizione verso lo stato *safe* è esterna (ad esempio un attacco alla rete) la riduzione di *availability* in un impianto costituisce un ingente danno economico ingiustificato.

Di seguito viene presentato un insieme di esperimenti mirato a valutare se e come il canale di comunicazione *safe* può essere sfruttato da un attaccante per compromettere l'*availability* del sistema, iniettando nella rete messaggi considerati validi (ben formati) dai nodi, ma la cui parte *safe* è corrotta (fa scattare una delle misure di *safety*). La configurazione sperimentale utilizzata per le prove è quella descritta nella Sezione 3 e il software packetETH è stato utilizzato per iniettare messaggi sul canale *safe*.

Nel primo esperimento si è tentato di portare a termine il più semplice tipo di attacco, ovvero la simulazione di una ripetizione di messaggio. Per mezzo di Wireshark è stato catturato un messaggio ciclico trasmesso dal PLC al nodo di I/O, contenente un SPDO valido. Successivamente, il messaggio è stato inviato nuovamente al nodo di I/O. L'errore sul canale *safe* è stato rilevato dal sistema e l'uscita digitale *safe* è stata portata a zero in accordo con la politica *fail-safe*, mentre il canale di comunicazione standard ha continuato a funzionare correttamente, senza alcuna alterazione del segnale sulla sua uscita digitale.

Nel secondo esperimento si è tentato di provocare un errore di integrità dei dati. Anche in questo caso, è stato utilizzato un messaggio catturato durante la comunicazione tra il PLC ed il nodo I/O. Il messaggio è stato corrotto modificando il valore contenuto nel campo SDATA del SPDO originale, senza però ricalcolare il valore del CRC. Inoltre, considerato che ogni SPDO contiene due copie identiche del campo SDATA, l'esperimento è stato ripetuto in due

diverse condizioni: prima sono state modificate entrambe le copie del campo in maniera identica e successivamente ne è stata modificata solamente una copia. Nel primo caso, il messaggio iniettato è stato ignorato dal nodo di I/O. Visto che entrambe le copie del campo SDATA sono state corrotte allo stesso modo, il nodo di I/O ha potuto determinare senza dubbio che il pacchetto ricevuto è stato alterato, quindi ha scartato il pacchetto corrotto ed ha atteso la ricezione del prossimo pacchetto prima di decidere se portare o meno il sistema nello stato *safe*. Il messaggio seguente è stato ricevuto correttamente e nella giusta sequenza, quindi la comunicazione è proseguita. Nel secondo caso, il nodo di I/O ha ricevuto un messaggio con due copie diverse del campo SDATA. Non potendo determinare in maniera univoca la causa della corruzione, il sistema è stato portato nello stato *safe*, segnalando un problema grave: la comunicazione è stata interrotta e l'uscita digitale *safe* è stata portata a livello logico basso. Inoltre, anche la comunicazione standard si è interrotta.

I risultati delle prove hanno dimostrato che i meccanismi di *safety* del protocollo possono essere sfruttati da un attaccante per compromettere l'*availability* del sistema in maniera subdola, nascondendo l'attacco in maniera da renderne difficile il rilevamento da parte dei meccanismi comunemente impiegati per rafforzare la sicurezza delle reti [Cheminod et al, 2011] [Treytl et al, 2004]. I messaggi inviati per portare a termine un attacco e le loro conseguenze sono gli stessi che potrebbero essere generati in seguito al guasto di un componente del sistema. L'insieme minimo di informazioni che un attaccante deve conoscere comprende gli indirizzi dei dispositivi nella rete ed i PID utilizzati durante la comunicazione. Gli indirizzi possono facilmente essere ottenuti inviando in *multicast* un messaggio di servizio "*scan network read request*", previsto dal protocollo, al quale ogni dispositivo risponde indicando il proprio indirizzo IP. Per quanto riguarda i PID invece, non esiste un messaggio di servizio per richiedere la loro trasmissione. L'unico modo per ottenerli, da noi conosciuto al momento, è intercettare la comunicazione per mezzo di un attacco "*man in the middle*", molto difficile da attuare se la rete è ben protetta.

6. Conclusioni

Questo contributo analizza una specifica implementazione del concetto di *black channel* in una rete industriale basata su tecnologia di trasmissione Ethernet, cioè SafetyNET p. Questo tipo di soluzioni permette di realizzare gli scambi di dati inerenti alla *safety* e al controllo convenzionale dell'impianto sullo stesso supporto di comunicazione, con una riduzione tangibile dei costi di installazione e manutenzione.

Si è investigato se (e come) i meccanismi di *safety* possono essere sfruttati per effettuare attacchi informatici ad un impianto di produzione. In particolare, l'attenzione si è focalizzata su un attacco DoS particolarmente subdolo, che può essere difficile da rilevare da parte dei moderni *intrusion detection system* (IDS) e che può portare a perdite economiche non trascurabili prima di venire individuato. Dal punto di vista della *safety*, il protocollo SafetyNET p si è comportato come previsto. Occorre tuttavia rimarcare che la convinzione (purtroppo non rara) che i protocolli di *safety* possano tollerare in qualche modo anche gli attacchi informatici è totalmente infondata.

È bene notare come il comportamento descritto non sia da ascrivere alla specifica soluzione analizzata (la quale, per la verità, ha sempre dimostrato di riuscire a gestire correttamente le funzionalità di *safety* in condizioni canoniche, ovvero in assenza di attacchi informatici). Piuttosto, le limitazioni evidenziate sono intrinseche del concetto di *black channel*. Questo implica che, sfortunatamente, esse non possono essere risolte in modo semplice, ma richiedono l'inclusione di contromisure specifiche per la sicurezza direttamente a livello del protocollo.

I risultati da noi ottenuti, per quanto potenzialmente interessanti, sono da considerarsi preliminari. Le attività in corso di esecuzione prevedono la realizzazione di un generatore programmabile in grado di iniettare in rete specifiche sequenze di pacchetti (ben formati e corrotti), in modo da analizzare il comportamento del sistema di *safety* in un insieme di scenari e condizioni sufficientemente ampio e significativo.

Bibliografia

[Akerberg e Bjorkman, 2009] Akerberg J., Bjorkman M., Exploring network security in profisafe, Proc the 28th International Conference on Computer Safety, Reliability, and Security, 2009, 67–80.

[Cheminod et al, 2011] Cheminod M., Pironti A., Sisto R., Formal vulnerability analysis of a security system for remote fieldbus access. IEEE Transactions on Industrial Informatics, 7, 1, 2011, 30-40.

[Hansen, 2009] Hansen K., Security attack analysis of safety systems, Proc of the 14th IEEE Conference on Emerging Technologies Factory Automation, 2009, 1-4.

[IEC, 2002] IEC 62280-1, Railways applications - Communication, signaling and processing systems - Part 1/2 – Safety related communication in closed/open transmission systems, 2002.

[IEC, 2010a] IEC 61158-4. Industrial Communication Networks - Fieldbus specifications - Part 4-18: Data-link layer protocol specification - Type 18 elements, 2010.

[IEC, 2010b] IEC 61508-1. Functional safety of electrical/electronic/programmable electronic safety-related systems - Part 1 - General requirements, 2010.

[IEC, 2011] IEC 61784-3-18. Industrial Communication Networks - Profiles - Part 3 - 18: Functional safety fieldbuses - Additional specifications for CPF 18 - SafetyNET p (FSCP 18/1), 2011.

[Safety Network, 2009] Safety Network International e.V. SafetyNET p Real-time Ethernet, 2009. [online]. <http://www.safety-network.de/en/technology/safetynet.html>.

[Treytl et al, 2004] Treytl A., Sauter T., Schwaiger C., Security measures for industrial fieldbus systems - state of the art and solutions for ip-based approaches, in Proc. of the 5th IEEE International Workshop on Factory Communication Systems, 2004, 201 – 209.

Automatic model-driven construction of communication-based software applications

Antonio Natali

Alma Mater, University of Bologna
via Venezia 52, 47023 Cesena, Italy
antonio.natali@unibo.it

Computer-based communications are enabled by several supports, but are still not reflected into proper language abstractions. The consequent abstraction gap can be today automatically overcome by model-driven software development techniques based on the definitions of proper meta-models and related software factories. This work presents a factory of this kind, by showing how it can promote the production of custom IDEs able to transmit the knowledge and the know-how of a company and to improve both the production process and the quality of the products.

1. Introduction

Computer-based communications are today possible not only by human beings but also by things at anytime and from anywhere. Thus, software applications become increasingly distributed and communication takes an increasingly central role. The problem is that current programming languages do not provide the new ways of thinking required by communications-based applications. The lack of a reference language with a clear semantics related to communication is particularly harmful during the phases of analysis and design, since sentences like "a sends a message to b" are not adequate to understand what the communicating partners have to do or what they expect from each other; it is then often difficult to evaluate the impact of a communication patterns over the business logic.

Communications-based applications are an example of a recurrent problem in software development: the need to design and build software that is conceptually related to an infrastructure that has to overcome some "abstraction gap" rather than to express the business logic. To avoid the risk of facing the same problems several times, each time in a different way, the current software technology allows a company to define with a minimal effort a proper, application-related, conceptual space by introducing domain specific languages (DSL) inherently associated with the application needs and with the concrete organization of the final system. Since a DSL can be associated with an explicit meta-model, the model-driven software development (MDS) approach [Stahl and Voelter 2006] can be used to create a custom software factory that

automatically builds an application-specific software infrastructure between the application level and some concrete support for communications.

The aim of this work is to present (section2) a software factory of this kind, (called Contact)based on the stack of meta-models described in [Natali and Molesini 2010a]. Section 3 introduces a simple case-study, to give an example of the impact of Contact on the production process and on the structure (and quality) of the final product. In this section, the product is a prototype (in which messages are exchanged via shared memory only) built in a very short time in order to achieve as soon as possible customer's feedbacks. Section 4 shows how this prototype can be transformed into a distributed system based on connection-based protocols without changing the application code and how non-trivial application problems can be managed by a proper model-driven cooperative work between *application designers* (experts in the application domain) and system designers (experts in the technology). Finally, section 5 is devoted to the conclusions.

2. Communication platforms: the Contact approach

Today there are many supports for communications: starting from libraries that enable the usage of protocols (TCP, UDP, HTTP, etc), one can use general purpose communication platforms such as *Java Message Service* [JMS 2011] *Eclipse Communication Framework* [ECF 2011], etc. and supports related to specific technologies such as *Web Services* (e.g. Axis [Axis 2011]), FIPA [IEEE FIPA 2009] (e.g. Jade [Bellifemine et al 2005]) and so on. Each communication support provides an application programming interface (API) that defines de-facto a conceptual space, but without providing any explicit meta-model. Usually this space includes concepts like *connection*, *publish-subscribe*, *(a)synchronous calls*, *time (un)coupling* and so on. The problem is that these concepts reflect low-level mechanisms and they are often not adequate to express the application logic at a right level of abstraction. Thus, another set of concepts is often required, leading to the need of some other software layer, to keep the application code independent of specific communication API. In the Contact approach this layer is represented by a meta-model called *corecl* that captures the basic idea that, in order to communicate, two or more entities must share a common, abstract medium. In the Contact terminology, the common medium is named *sharedSpace*; it provides basic operations to *write* (out), *read* (rd, rdw, selectOne) or *consume* (in, inMany) information, as defined by an interface named *ILindaLike* :

```
public interface ILindaLike {
    public IMessage in(String query) throws Exception;
    public IMessage rd(int LastMsgNum, String queryMsg) throws Exception;
    public IMessage rdw(int LastMsgNum, String queryMsg) throws Exception;
    public void out(String msg) throws Exception;
    public Vector<IMessage> inMany(Vector<String> tokens) throws Exception;
    public IMessage selectOne(int msgNum, Vector<IMessage> queries)
}.
```

The idea is that each communication pattern required at application level can be expressed as a sequence of *ILindaLike* operations (technically that are a Java representation of an instance of the *corecl* meta-model) and that this sequence defines the operational semantics of this pattern. The high level communication pattern required by the application logic are instead expressed by a proper symbolic specification language (the *Contact* language). Another intermediate meta-model, called *medcl*, is introduced to capture (in an operational perspective) the conceptual space of *Contact*. Thus, the formal meaning of each *Contact* expression is defined by two model-to-model (M2M) transformations, as shown in Figure 1: the first translates each *Contact* operation type into an instance of some *medcl* operation; the second translates a *medcl* operation type into a sequence of *ILindaLike* moves. Each *ILindaLike* operation sequence can be in its turn automatically mapped onto some specific communication API by introducing a proper model-to-code (M2C) transformer, that works as the back-end of a compiler with reference to some specific machine.

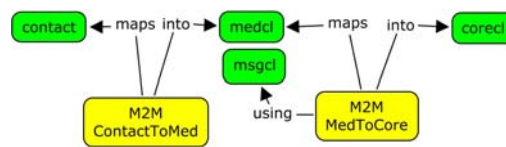


Figure 1 – The contact M2M transformation chain

This (meta)model-based transformation chain is the core of the *Contact* software factory and of the related development environment (called *Contact-ide*). The advantage of this approach is that the part that defines the semantics of the application language remains unchanged, while the porting of the application on a different communication support can be done by writing another M2C transformer. Moreover, the code related to the business logic does not change, since the intermediate layer required to adapt the application code to some concrete platform is automatically generated by the *Contact-ide*.

3. Impact of the *Contact* factory

Once properly defined, a custom language becomes a tool able to promote reasoning and cooperation, while its related factory (the *Contact-ide* in our case) provides an environment able to play an active role in driving cognitive processes in all the phases of a software production. To give an example, we make reference here to a simple case-study: the definition of a ping-pong system in which two players interact by sending each other a message. The game is observed by an arbitrator which can interrupt at any time the action of the players.

Since this specification is rather unclear, a careful analysis phase is required to understand the nature of the communication patterns. The *Contact* specification language provides a vocabulary that allows an analyst to express

different kinds of messages, each associated with a different set of obligations / expectations by the partners (called *Subjects*). A *Contact::Message* type is defined as the specialization of some element of a message ontology that defines two main message categories: *OutOnlyMessage* and *OutInMessage*; moreover, each message type is associated with a pair of high level operations (one to send information and one to receive information) like those written in bold in Table 1.

MsgType	Meaning
Signal	a (<i>OutOnlyMessage</i>) message that a subject can emit on the sharedSpace; a signal can be sensed by $N \geq 0$ time-uncoupled subjects
Dispatch	a (<i>OutOnlyMessage</i>) message that a subject can forward to one specific receiver subject and with the expectation that it will serve it.
Invitation	a (<i>OutInMessage</i>) message that a subject can ask to $ND \geq 1$ receivers, with the expectation to acquire $NR \geq 0$ acknowledgments (ack); a receiver will accept an invitation and will reply with an ack to the sender.
Request	a (<i>OutInMessage</i>) message that a subject can demand to $ND \geq 1$ receivers, with the expectation to acquire $NR \geq 0$ responses; a receiver will grant a request and will answer to the sender with a response.

Table 1 – Informal meaning of Contact message types

The result of the requirement analysis (a cooperative work with the customer) can be expressed by the following Contact specification:

ContactSystem PingPongSystem; //Definition of the entities Subject ping ; Subject pong ; Subject arbitrator ; //Definition of the message types Dispatch ball ; Invitation stop ; //High-level communication patters ping forward ball to pong ; ping serve ball ; ping accept stop ; pong forward ball to ping ; pong serve ball ; pong accept stop ; arbitrator ask stop to ping, pong ;	This Contact specification describes a system composed of three entities qualified as <i>Subjects</i> . The subject named ping forward (sends) a <i>dispatch</i> named ball to the entity pong, that declares to serve (receive) it. The subject named pong works in the same way. Both ping and pong declare to accept (receive) the <i>invitation</i> named stop that the subject named arbitrator ask (send) to them. From the technical point of view, this specification describes a model (instance of the Contact meta-model) that must be written according to the concrete syntax of the Contact language. A sentence like pong accept ball is syntactically not correct, since only an invitation can follow the “accept” keyword.
---	---

A Contact specification (like any model, instance of a explicitly defined meta-model) can be validated by a proper set of custom semantic rules. For example, a model-checking rule could check that for each message sent in the system there is at least one receiver.

The syntax of Contact is defined by using the Xtex2-0 framework [Xtext 2011], that automatically creates a parser, a syntax-driven editor and a Java

representation of the Contact meta-model (source in figure 1) by exploiting the *Eclipse Modeling Framework* (EMF) [Steinberg et al 2009] .

Since the Contact-ide (deployed as a set of Eclipse plugins) automatically executes the transformation chain, the specification of the PingPongSystem is translated into a set of code modules (Java abstract classes) that provide for each subject a set of ad-hoc methods for the communication it requires. The application designer can exploit the generated supports as an extension of the implementation language and build in a very short time a working prototype that exploits the common memory as communication support .

3.1 A first single-machine prototype

Each Contact subject can be designed and implemented as a finite state machine (FSM) whose transitions are associated to generated, application-specific communication methods. The application designer can exploit the generated support of each subject to write clear, structured and technology-independent code. For example (the arbitrator is not considered here):

<pre> public class Ping extends PingSupport{ public void doJob() throws Exception { while(true){ send(i); receive(i); } } protected void send(int i) throws Exception{ String msg = getName()+"_hit_"+i; hl_ping_forward_ball_pong(msg); } protected void receive(int i) throws Exception{ IMessage m = hl_ping_serve_ball(); showMsg(m); } } </pre>	The method named hl_ping_forward_ball_pong is inherited by the PingSupport and provides the operation required to send the dispatch to pong. The method named hl_ping_serve_ball is also inherited by PingSupport and provides the operation required to receive the dispatch .
--	---

The communication methods generated into the support class of each subject reflect the stack of Contact meta-models and the consequent decomposition of the application-support infrastructure into a set of layers, each related to a proper meta-model and built by a proper software factory. The general structure of the application-specific communication methods can be sketched as follows:

Output operation
<pre> <AnswerType> hl_<SenderName>_<OpName>_<messageld>_<ReceiverName>(String M){ ILindaLike support = PlatformExpert.findOutSupport(<ReceiverName>,<messageld>,<SenderName>); <AnswerType> answer= <SenderName>_<OperationName>_<messageld>_<ReceiverName>(M,support); return answer; } </pre>
Input operation
<pre> <ReturnType> hl_<ReceiverName>_<OperationName>_<messageld>(){ ILindaLike support = PlatformExpert.findInSupport(<ReceiverName>,<messageld>,<WithAnswer>); <ReturnType> answer=<ReceiverName>_<OperationName>_<messageld>(support); return answer;} </pre>

Table 1 – Communication operation scheme

Each **hl** xxx method calls a corresponding low-level operation (that defines the operational semantics in terms of ILindaLike operations) by giving to it the support to perform the required communication actions (see section 4). The types of the object returned by a *Contact* operation are:

AnswerType	ReturnType
void : when no reply is expected by the caller. IAcquireOneReply : when the caller expects to receive back just one reply. IAcquireManyReply : when the caller expects to receive one or more reply.	IMessage : when the receiver does not have to send any reply to the caller. IMessageAndContext : when the receiver must create and send a reply to the caller.

The objects returned by a *Contact* operation allow an application designer to check for the availability and/or to consume a message or a reply when it is required by the business logic.

public interface IMessageAndContext { public void replyToCaller (String msg); public IMessage getReceivedMessage (); }	<i>getReceivedMessage</i> returns the received message, while <i>replyToCaller</i> sends the reply (ack, answer etc.) to the Subject that has sent that message.
---	--

public interface IAcquireOneReply{ public boolean replyAvailable (); public IMessage acquireReply () ; } }	<i>acquireReply</i> is a blocking operation that returns a reply message if it is available (i.e when <i>replyAvailable</i> .returns true)
--	--

public interface IAcquireManyReply { public int numOfReplyExpected (); public int numOfReplyReceived (); public IMessage acquireReply (int n);}	<i>numOfReplyReceived</i> return the number of reply received, while <i>numOfReplyExpected</i> returns the number of reply that could be received.
---	--

Low-level concepts like (a)synchronous call, connection, time (un)coupling, etc. are now embedded into the semantics of the message abstractions, like the *if* and *goto* statements are embedded into iteration constructs.

4. From the prototype to a distributed system

To obtain a version of the system able to run on several computational nodes, the *Contact* specification language is extended, so allow a receiver subject to express information about some concrete communication technology.

```
ContactSystem PingPongSystem;
..
ping forward ball to pong ;
ping serve ball support=TCP [ host= "localhost" port= 8050] ;
ping accept stop support=TCP [ host= "localhost" port= 8050];
pong forward ball to ping ;
pong serve ball support=TCP [ host= "localhost" port= 8060] ;
pong accept stop support=TCP [ host= "localhost" port= 8060];
arbitrator ask stop to ping, pong ;
```

The task of the PlatformExpert called within a **hl_** communication support method (see table 1) is to dynamically find (or build) at run-time the right communication support. The fact of using some specific support (the TCP protocol in this case) does not affect the business logic and should have also minimal impacts on the application code. In this case, to transform the first prototype in a distributed version no change at all is required in the application code: all the work is done by the factory.

In general, the transition from a system working into a single machine to a distributed system cannot be completely transparent to the application logic. The strategy adopted in Contact is that the content of a message handled at application level can include also information set by the underlying platform, e.g. an *exception-message*. However, since distributed systems can pose several problems, the point here is not how they are solved by Contact, but that the system designer and the application designer must cooperate to overcome the abstraction gap by finding the best way to make these problems detectable and manageable at application level. The result of this cooperation can be the extension of some meta-model or of some part in the code-generation, according to (evolving) application needs. In our case-study, for example, each player should manage several kinds of messages, by reacting as soon as possible to a stop message sent by the arbitrator, that is supposed to take its decisions by observing the messages exchanged by the players.

4.1 Message observers

Differently from what happens in many connection-based supports, all the messages exchanged within a Contact system are stored in a local knowledge base implemented as a logic theory where each message is represented as a Prolog term `mmm/2` that takes the following form:

`mmm(channelId(subjectId,msgId,content,msgNum),seqNum)`

The system-level message components are defined by the `msgcl` meta-model (see figure 1) and made available at application level by the operations of the `IMessage` interface:

<pre>public interface IMessage extends IBasicMessage{ public String msgEmitter(); public String msgId(); public String msgContent(); public String msgNum(); }</pre>	<pre>public interface IBasicMessage { public String getMsgName(); public String getMsgContent(); public int getMsgSeqNum(); }</pre>
--	---

The message interfaces reflect the two layers composing the Contact runtime system and most of the cooperative work between application and system designers, as summarized in the box hereunder.

The message set as a logic theory: The message representation has been defined by the system designer to support the different Contact communications forms, by allowing also reasoning and inspection on messages. The structure of the functor *channelId* reveals how the *sharedSpace* (an abstraction that inherently provides many-to-many communications among time-uncoupled subjects) can implement also P2P messages (like dispatches, invitations, requests). If two subjects named *s1* and *s2* declare to send a message with *msgId*=*m* to a same subject *r*, conceptually they make reference to a same logical channel with *channelId*=*r_m*. The *channelid* *s1_r_m* and

s2_r_m are used to "carry" an answer/reply. A (TCP) network connection is viewed as an alternative communication medium. Since each message describes its own logical connection, a single connection can be used to support all the P2P communications among the subjects running on the two connected nodes.

The local message theory is defined as an "observable" object according to the *Observer* design pattern [Gamma et al. 1995]. Thus, the arbitrator subject can be implemented as an observer of the *sharedSpace* that must decide how and when send the stop *invitation* (a oneToMany message) to all the players. For example:

```
public synchronized void update(Observable source, Object obj) {
    IBasicMessage mb = new BasicMessage(msg);
    IMessage m = new Message(mb.getMsgContent());
    if(m.msgEmitter().equals( getName())) return; //drops its own messages
    if( localDesction() ) IAcquireManyReply mr = hl_arbitrator_askM_stop( "reason" );
}
```

4.2 Multi-input

While playing, both the players should also handle the message stop sent by the arbitrator. This problem could be solved at application level, by splitting each player into several parallel parts, each waiting for a specific message type. But this extra-complexity can be removed from the application code, by a further step in the cooperative work between application designers and system designers. In Contact this approach has lead to the introduction of two new operations, that are automatically injected by the Contact-ide into the support class of each subject that can receive many input messages:

```
Vector<IMessage> prepareInput() ;
IMessage selectOneInput(int LastMsgNum,Vector<IMessage> queries) ;
```

The method *prepareInput* builds a list of queries that must be given as argument to the *selectOneInput* method that returns (in non-deterministic way) a message that can match (unify) with one of the queries. Thanks to these operations (whose source code is made available to the application designer by the Contact-ide), the application code can be changed by giving to each subject behavior a message-driven (and no more control-driven) state transition structure.

<pre>public class Ping extends PingSupport{ boolean goon = true; public void doJob() throws Exception { while(true){ if(!goon) break; interact(i); } }} </pre>	<pre>protected void interact(int i) throws Exception{ send(i); IMessage m = selectOneInput(prepareInput()); elab(m,i); } protected void elab(IMessage m, int i) throws Exception{ if(m.msgId().equals("stop")) goon = false; else if(m.msgId().equals("ball")) ... }} </pre>
--	---

In order to react as soon as possible to the `stop` message, a player can recur to inspection of the message space by using the low-level `ILindaLike.rd` primitive that reads a messages without removing it from the *sharedSpace*.

4.3 Smart devices

In the era of the “internet of things”, it is quite common that a distributed software application should run on an heterogeneous set of computational nodes, including some smart device. Let us suppose here that one of the two our players must be installed on a Android [Android 2011] phone.

Android provides its own conceptual space for the communication (the *intent* concept) between its own basic components (*activities* and *services*) but must recur to “commodity objects” (sockets) to express the communications with the outside world. But, since the Android virtual machine is compatible with Java, a Contact player can be ported “as it is” on Android: the only thing to do is to allow this player to interact with some specific gui-compliant Android component (an *activity* in particular) to perform its output. This goal can be achieved either introducing a new gui-compliant subject in the Contact specification or by extending the software factory by generating a new version of the *showMsg* operation, based on message-passing. In the second case, the message type that best fit with the Android conceptual space is the *dispatch*, since any Android activity is associated to a message-queue that must be used by other threads to send information to that activity. Another possibility is to obtain the result via a M2M mapping of the Contact conceptual space into a proper representation of the Android conceptual space, like that reported in [Natali and Molesini 2010].

5. Conclusions

Modern communication platforms provide powerful mechanism to *enable* communication but in actual software applications often arises the need of abstractions allowing to *reason* about communications both at development time and at run time. The “abstraction gap” is then not limited to the need of introducing high-level concepts to capture the logic of communications, but it also requires more smart infrastructures, able, for example, to provide a global view of what happens within a system.

In this context, a communication based software application should be viewed as the result of the cooperative work of application designers and system designers. The main vision behind the work presented in this paper, is that the relationships among these kinds of co-workers should be mainly derived from the artifacts that comprise a model-driven project. In fact, the conceptual space related to the meta-models and the related (abstract) language semantics become a crucial aspect not only for the common understanding required in every cooperation, but also to achieve an appropriate expressive power during software production.

From the technical point of view, the idea of model-driven cooperation around custom languages finds an excellent support in production ecosystems like

Eclipse and XText, that have been used for the development of Contact, the system presented in this paper. The goal of model-driven software co-construction is then within the reach of many companies that can build custom artifacts (e.g. IDEs) able to explicitly represent and transmit the knowledge and the know-how of the company both in some specific application domain and in the (relationship with the) supporting technologies.

The Contact experience shows that a modern company can not only implement advanced forms of communications like those proposed by FIPA (without the need of adopting an agent-based paradigm) but it allows also to build in automatic way new “advanced” infrastructures, able to enable run-time reasoning on communications and to use message spaces as a run-time knowledge base to achieve more ambitious application goals.

6. References

- [Android 2011] Google: Android site. <http://developer.android.com/index.html>
- [Axis 2011] Axis home page. <http://ws.apache.org/axis/>.
- [Bellifemine et al 2005] F. Bellifemine, F. Bergenti, G. Caire, and A. Poggi. JADE - A Java Agent Development Framework. Multiagent Systems, Artificial Societies, and Simulated Organizations, 15(2):125{147, 2005.
- [ECF 2011] Eclipse Communication Framework Project Home, <http://www.eclipse.org/ecf/>
- [Gamma et al. 1995] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, Design Patterns: Elements of Reusable Object-Oriented Software. Addison Wesley, 1995.
- [IEEE FIPA 2009] IEEE FIPA Design Process Documentation and Fragmentation Homepage. http://www.pa.icar.cnr.it/cossentino/_pa-dpdf-wg/ (2009)
- [JMS 2011] JMS home page. <http://java.sun.com/products/jms/>
- [Natali and Molesini 2010a] Natali, A., Molesini, A.: Towards model-driven communications. In Ardil, C., ed.: World Academy of Science, Engineering and Technology. Volume 64., Rome, Italy, Academic Science Research (2010) 69{80 International Conference on Software Engineering and Technology (ICSET 2010).
- [Natali and Molesini 2010b] Natali, A., Molesini, A.: Using Domain Specific Languages to describe the architecture of software systems: the case of Android applications In Coccoli, M., ed.: Eclipse-IT 2010. Volume 90-103., Savona, Italy, Eclipse Italian Community (2010) 29-40 The Fifth Workshop of the Italian Eclipse Community.
- [Stahl and Voelter 2006] Stahl, T., Voelter, M.: Model-Driven Software Development. Wiley (2006)
- [Steinberg et al 2009] Steinberg, D., Budinsky, F., Paternostro, M., Merks, E.: EMF: Eclipse Modeling Framework, 2nd Edition. Addison-Wesley Professional (2009)
- [Xtext 2011] The Eclipse Foundation: XText site: Home page. <http://www.eclipse.org/Xtext/> (2011)

Building Smart City Views through Web Information Clouding^{*}

Stefano Montanelli¹, Silvana Castano¹, and Lorenzo Genta¹

Università degli Studi di Milano
DICO - Via Comelico, 39 - 20135 Milano
{stefano.montanelli,silvana.castano,lorenzo.genta}@unimi.it

Abstract. In this paper, we present the **MI-Search** project for the construction of *smart city views* based on web information clouding. By smart city view, we mean an user-centered way of organizing and delivering information about the events of a given city. The **MI-Search** architecture and related clouding techniques are discussed in the paper as well as real examples concerned with the city of Milan.

1 Introduction

Techniques for search and exploration of information about city events and points of interest have become very popular thanks to the success of recent Web 2.0 and Semantic Web technologies. Tools and mobile apps are now being appearing to satisfy the growing need of dynamic users to retrieve targeted tourist information by only relying on device geo-coordinates and integrated maps [9, 8]. Many other useful information are also available to the users for exploration on separate delivery channels such as traditional web sites, social networks, and RSS feeds [6]. In the current scenario, what is really missing is a single comprehensive platform where all such kinds of information are appropriately acquired with the goal of delivering them in a smart, intuitive picture for the final user fruition.

To this end, in this paper, we present the **MI-Search** project co-funded by Regione Lombardia and Fastweb S.p.A. to enforce the construction of *smart city views* based on web information clouding. By smart city view, we mean an user-centered way of organizing and delivering information about the events of a given city. In **MI-Search**, a smart city view is built by relying on *clouding techniques* developed to enable similarity-based classification of all the web data about a target event extracted from different sources and systems. The **MI-Search** architecture and related clouding techniques are discussed in the paper as well as real examples for the city of Milan, that we selected for pilot experimentation in **MI-Search**.

2 The MI-Search architecture

The **MI-Search** architecture is shown in Figure 1. The core component of the

^{*} This work is funded by Regione Lombardia and Fastweb S.p.A. in the framework of the *Dote in Ricerca* project.

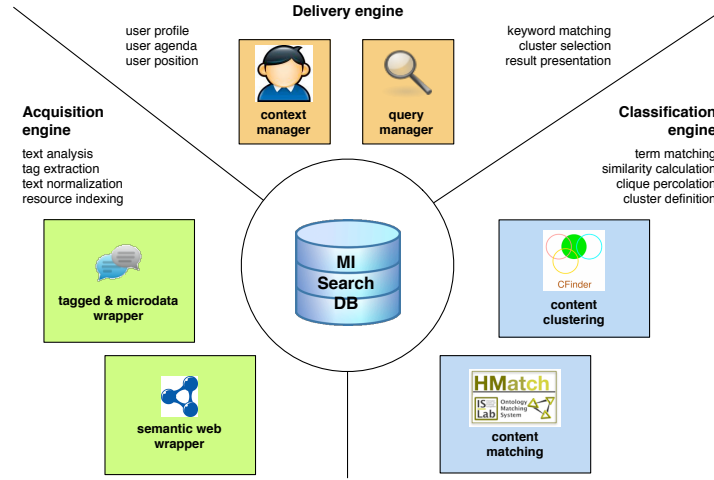


Fig. 1. The MI-Search architecture

architecture is the MI-Search-DB repository storing all the different kinds of web data/contents considered in the project according to a uniform representation. In particular, MI-Search-DB is implemented as a PostgreSQL relational database, and it allows the representation of both events and information resources acquired from external web datasources, such as websites, social networks, and RSS feeds (Section 3). On top of MI-Search-DB, we have the following engines:

- *Acquisition engine*. It has the role of populating the MI-Search-DB with the web data/contents for the construction of smart city views. This engine relies on appropriate wrappers, namely the *tagged & microdata wrapper* and the *semantic web wrapper*, to convert the data items extracted from external web datasources into the reference data model of the MI-Search-DB repository (Section 4).
- *Classification engine*. It has the role of building *similarity clusters* by aggregating similar web data items stored in MI-Search-DB. This engine is based on i) a *content matching* module to detect similar items in the repository that refer to the same topic, and on ii) a *content clustering* module to aggregate in a single cluster those data items that can have a different nature, but are similar in content (Section 5).
- *Delivery engine*. It has the role of handling the user requests/queries. This engine is based on i) a *context manager* to derive information about the user from her/his preferences, agenda, and geo-position, and ii) a *query manager* to select the information in the city view to deliver to the user by exploiting the similarity clusters and the user context (Section 6).

In particular, a **smart city view** is defined as a collection of similarity clusters about the events of a selected target city. A smart city view is built on top of a set of acquired web data/contents for effective delivery to the interested users in any possible situation. We call **web information clouding** the process of building a smart city view by relying on acquisition and classification engines.

Technical details about MI-Search-DB and MI-Search engines are discussed in the following sections.

3 The MI-Search-DB repository

The conceptual ER schema of MI-Search-DB is shown in Figure 2. In the schema,

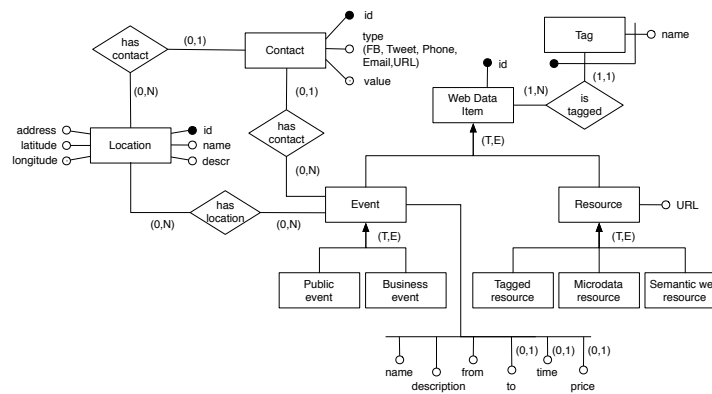


Fig. 2. The schema of the MI-Search-DB repository

we note that any kind of acquired web information is represented as a *web data item* (entity *Web Data Item*). Web data items are classified in events (entity *Event*) and resources (entity *Resource*).

Event. Events are classified in public events (entity *Public Event*), that represent official initiatives like art exhibitions or concerts, and business events (entity *Business Event*) that represent commercial initiatives directly inserted in the MI-Search system by business users. An event is characterized by attributes that describe its temporal frame (i.e., *from-date*, *to-date*, *time*, and *frequency*) and other features, like *description* and *price* (where needed). The entity *Event* is associated with the entities *Contact* and *Location* to represent the different contact-points for the event (e.g., Phone, Facebook page, Twitter channel) and the geo-coordinates where the event takes place, respectively.

Resource. Resources are web data items acquired from outside the system and they are distinguished in **Tagged Resource**, **Microdata Resource**, and **Semantic Web Resource**. Tagged resources are traditional web resources (i.e., web pages) and they are characterized by a raw structure with few metadata. Microdata resources are posts/comments coming from news feeds and microblogging systems (e.g., Facebook¹, Twitter² posts). A microdata resource is characterized by a short textual content and a set of metadata/properties, like title, author, and creation date, that are commonly employed to describe publishing items. Semantic web resources are instances/individuals coming from RDF(S) knowledge repositories and OWL ontologies (e.g., Freebase³, DBpedia⁴). This kind of resources are characterized by a structured description composed of a set of assertions denoting its specification in the web document of origin.

Tag. Each web data item, either event or resource, is associated with a set of tags (entity **Tag**) denoting the keywords that most appropriately characterize/represent the event/resource. In most cases, tags are already specified in the external datasources. In addition, tags can be automatically identified and inserted in MI-Search-DB through an extraction/normalization process (see Section 4).

4 The acquisition engine

The goal of the acquisition engine is to populate the MI-Search-DB repository with web data items, describing either events or resources, extracted from external web datasources. In particular, the acquisition engine has the role of associating the acquired web data items with a set of featuring tags to be subsequently used for classification. We note that, before insertion in the entity **Tag** of MI-Search-DB, a tag is submitted to a normalization procedure for word-lemma extraction and for compound-term tokenization [5, 10]. The acquisition engine is based on the *tagged & microdata wrapper* and on the *semantic web wrapper*.

Tagged & microdata wrapper. This wrapper is used to insert tagged and microdata resources in the **Tagged Resource** entity and it is associated with a set of tags extracted from bookmarking and social annotation systems (e.g., Delicious⁵, Flickr⁶). A microdata resource is inserted in the **Microdata Resource** entity and it is associated with a set of tags extracted from the textual content of the resource and from other available metadata/properties, like the title. The tagged

¹ <http://www.facebook.com/>.

² <http://www.twitter.com/>.

³ <http://www.freebase.com/>.

⁴ <http://www.dbpedia.org/>.

⁵ <http://www.delicious.com/>.

⁶ <http://www.flickr.com/>.

& microdata wrapper is also employed for event insertion since public events are usually published on electronic walls and distributed as RSS feeds⁷.

Semantic web wrapper. This wrapper is used to insert semantic web resources in the **Semantic Web Resource** entity and it is associated with a set of tags extracted from literals, property names, and property values extracted from the RDF/OWL assertions of the resource specification (see [4] for more technical details).

Example. In Figure 3, we consider two acquisition examples. Figure 3(a) shows



(a)

Description: Ricostruire la genesi...

Price: 9 Euro biglietto intero

From: 11/03/2011

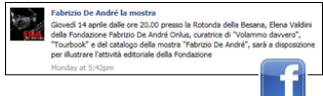
To: 15/05/2011

Date/time: Martedì-domenica 9.30-19.30

Tags: allestimento, andrè, besana, de, fabrizio, mostra, rotonda...

Contact: [http://blog.milano-italia.it/...](http://blog.milano-italia.it/)

Location: Rotonda della Besana, Milano, Italy



(b)

URL: <http://www.facebook.com/...>

Tags: andrè, besana, de, fabrizio, fondazione, mostra, rotonda...

Fig. 3. Examples of web resource acquisition

a RSS post published on a well-known electronic wall about events planned in the city of Milan (<http://blog.milano-italia.it/>). This is an example of public event related to the art-exhibition about the singer Fabrizio de Andrè located at Rotonda della Besana in Milan. Besides the featuring attributes expected in MI-Search-DB for a public event, contact and location information are also provided. Figure 3(b) shows a comment posted on the Facebook social network published by a user that visited the art-exhibition about Fabrizio de Andrè. This is an example of microdata resource featured by its URL on the web as expected in MI-Search-DB. Moreover, either the public event and the microdata resource, are associated with a set of tags extracted by the tagged & microdata wrapper from the two posts as a sort of synthetic characterization of each web data item.

⁷ In some situations, an event can be inserted in MI-Search-DB by an user through a manual association of a set of featuring tags to the event. This usually happens with business events.

5 The classification engine

The goal of the classification engine is to detect and build the similarity clusters to use for content delivery to the final users. The classification engine is based on the *content matching* module and on the *content clustering* module.

Content matching module. This module has the goal to evaluate the degree of similarity between pairs of web data items stored in the MI-Search-DB repository. Given two web data items wdi_i and wdi_j , the *similarity coefficient* $\sigma(wdi_i, wdi_j) \in [0, 1]$ denotes the level of similarity of wdi_i and wdi_j based on their matching tags. We define $\text{Tag}_{wdi} = \{tag_1, \dots, tag_m\}$ as the set of tags associated with the web data item wdi in MI-Search-DB. The similarity coefficient $\sigma(wdi_i, wdi_j)$ is calculated as follows:

$$\sigma(wdi_i, wdi_j) = \frac{2 \cdot |tag_x \sim tag_y|}{|\text{Tag}_{wdi_i}| + |\text{Tag}_{wdi_j}|}$$

where $tag_x \sim tag_y$ denotes that $tag_x \in \text{Tag}_{wdi_i}$ and $tag_y \in \text{Tag}_{wdi_j}$ are matching tags according to a string matching metric that considers the structure of tag_x and tag_y . For σ calculation, we employ our matching system HMatch 2.0, where state-of-the-art metrics for string matching (e.g., l-Sub, Q-Gram, Edit-Distance, and Jaro-Winkler) are implemented [3]. Two web data items wdi_i and wdi_j are similar ($wdi_i \approx wdi_j$) if $\sigma(wdi_i, wdi_j) \geq th$ where $th \in (0, 1]$ is a matching threshold denoting the minimum level of similarity required to consider two web data items as matching items. Based on the experimental results obtained with HMatch 2.0 on real datasets, a default value $th = 0.3$ is recommended when string matching techniques are employed for similarity evaluation of web data items.

Content clustering module. This module has the goal to determine the similarity clusters that constitute a smart city view. Similarity clusters are built by relying on a clique percolation method (CPM) [7]. This method receives in input a graph $\mathcal{G}$ where nodes are the web data items stored in the MI-Search-DB repository and edges are defined between any pair of similar nodes $wdi_i \approx wdi_j$. An edge (wdi_i, wdi_j) in $\mathcal{G}$ is labeled with the value of $\sigma(wdi_i, wdi_j)$. The CPM returns a set of similarity clusters where each cluster collects a region of nodes in $\mathcal{G}$ that are more densely connected to each other than to the nodes outside the region. The CPM is based on the notion of *k-clique* which corresponds to a complete (fully-connected) sub-graph of k nodes within the graph $\mathcal{G}_s^+$. Two k -cliques are defined as *adjacent k-cliques* if they share $k - 1$ nodes. The CPM determines clusters from k -cliques. In particular, a cluster, or more precisely, a k -clique-cluster, is defined as the union of all k -cliques that can be reached from each other through a series of adjacent k -cliques. More technical details about the CPM and the construction of similarity clusters can be found in [4].

Example. We consider the example shown in Figure 3. We call wdi_1 the public event of Figure 3(a) and wi_2 the microdata resource of Figure 3(b). The similarity

coefficient of wdi_1 and wdi_2 is $\sigma(wdi_1, wdi_2) = 0.35$ due to the matching tags in Tag_{wdi_1} and Tag_{wdi_2} . With a matching threshold $th = 0.3$, the web data items wdi_1 and wdi_2 are considered as matching items and an edge (wdi_1, wdi_2) is set in the graph $\mathcal{G}$ that is passed to the CPM method for calculation of the similarity clusters. An example of similarity cluster is shown in Figure 4. Besides the web data items wdi_1 and wdi_2 , the cluster of Figure 4 contains

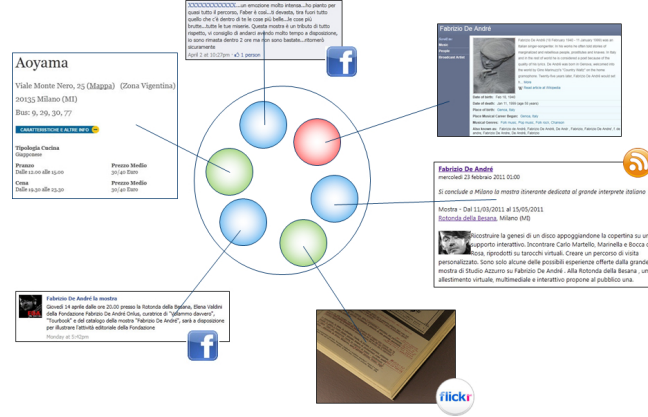


Fig. 4. An example of similarity cluster

a Flickr image taken from the exhibition (tagged resource), another Facebook user comment (microdata resource), the Freebase page about Fabrizio de Andrè (semantic web resource), and the contact information of the Aoyama restaurant, a sushi restaurant that published in MI-Search a discounted dinner offer associated with the art-exhibition at Rotonda della Besana (i.e., business event).

6 The delivery engine

The delivery engine supports users during their surfing activities of the smart city views and related similarity clusters. The delivery engine is based on the *context manager* and on the *query manager*.

Context manager. This manager has the role of handling the user-related information like the user profile, the user agenda, and the user position. For user context representation, many different models and solutions are available in the literature according to the level of detail that needs to be supported [2, 1]. In general, the context manager allows the user to insert and update her/his interests. Moreover, extracting tools are defined to enable the context manager to exploit the user meetings planned in the agenda and the user geo-coordinates sensed by a GPS device. In MI-Search, the user context is also in charge of extracting featuring keywords from the user context that describe the user interests.

Query manager. This manager has the role of handling the user requests. In some cases, queries can be manually specified by the user interested in a certain argument. More frequently, queries are automatically generated by the manager to extract interests, planned meetings, and geo-coordinates about the user from her/his personal context. In both cases, a query consists in a list of keywords describing the argument/topic of interest for which information needs to be retrieved. The query manager adopts keywords to select the similarity clusters that are more interesting for a certain argument. Given a query, the associated keywords are used to perform a string matching operation against the tags featuring the resources in the similarity clusters. The clusters that have sufficient matching resources (according to a fixed threshold) are selected as pertinent for the query and they are delivered to the user. The choice of using string matching for selecting the pertinent clusters is motivated by the need of reactivity to guarantee in content delivery for satisfying the user expectations.

Example. We consider an user that plan to visit the art-exhibition about Fabrizio de Andrè at Rotonda della Besana. The following note is inserted in the agenda of her/his mobile device⁸:

What : mostra su Fabrizio de Andrè
Where : Rotonda della Besana, Milano
When : 30/04/2011 - 6PM

The following keywords are extracted from the note by the query manager: *andrè, besana, fabrizio, milano, mostra, rotonda*. These keywords are matched against the tags of the web data items in the MI-Search-DB repository. The items shown in Figure 3 are found to match the keywords in the query, thus the similarity cluster of Figure 4 is delivered to the user as a result. The overall content of this cluster is provided to the user for browsing and exploration of the available information about the art-exhibition.

7 Use case scenarios

In MI-Search, we distinguish two different kinds of user categories: the *citizen* and the *business users*. Citizens are users interested in receiving information and they exploit the MI-Search technology for obtaining contents and suggestions about their personal interests and events in the agenda. Business users are users interested in public events and other possible situations that are suitable for promoting their business activities. In this respect, the following three main scenarios will be supported in MI-Search:

Search-2-me scenario. This is the typical scenario of citizens and it is triggered when new personal events are planned by the citizen in the agenda. By exploiting the user agenda, the MI-Search technology discovers the user interests and it can

⁸ The structure of the note can vary from one mobile agenda to another. Here, we consider an abstract format where only classical information are represented.

provide a complete set of information about a planned event. In particular, MI-Search retrieves spontaneous information and user-generated contents related to the considered event, like comments from other similar citizens and special user offers joint with the participation to the event. As an example, we consider the example of the art-exhibition event about **Fabrizio de Andrè** located in Milan at **Rotonda della Besana**. A citizen can plan to visit this exhibition and thus a personal event is inserted in the agenda for a certain date. Through specialized websites (e.g., <http://www.fabriziodeandrelamostra.com/>), MI-Search automatically provides to the citizen all the available information about the exhibition and about the singer. Moreover, other information are extracted by the MI-Search technology from social networks (e.g., Facebook, Twitter) to provide comments of other users that previously visited the exhibition.

Me-2-search scenario. This is the typical scenario of business users and it is triggered by the user when she/he start browsing the available suggestions that the system provides as potentially interesting opportunities for promoting the user business. The user can browse a suggestion list of public events that can be interesting from the business point of view and she/he can decide to insert in the system a new business offer joint with a suggested event. For example, we consider a business user that has a sushi restaurant located in Milan, Viale Montenero (near to **Rotonda della Besana**). When the user starts browsing the possible suggestions, the art exhibition about **Fabrizio de Andrè** at **Rotonda della Besana** is retrieved (due to a geo-locality proximity). The business user can decide to insert in the system a special menu price for the exhibition visitors. Such an offer will be linked to the art exhibition event and it will be visualized by citizens that plan to visit the exhibition.

Recommend-2-me scenario. This is a basic scenario of citizens and it is permanently active without requiring any triggering event. The recommend-2-me scenario is based on the user profile to suggest events and/or (promotional) initiatives that can be potentially interesting. In this scenario, the citizen periodically receives a report with a list of upcoming events, either public and business events, that match her/his preferences for possible selection (and subsequent insertion in the personal agenda). As an example, we consider a citizen who specified an interest for sushi restaurants in her/his profile. Receiving the periodic report of interesting upcoming events, the citizen becomes aware of the special menu price of the sushi restaurant in Viale Montenero joint with the art exhibition about **Fabrizio de Andrè**. The citizen can decide to visit the exhibition with the goal to subsequently take advantage of the special sushi offer. A personal event is inserted in the user agenda to plan the visit and to receive further information about the event (see the search-2-me scenario).

8 Concluding remarks

In this paper, we presented the main features of the MI-Search project for the construction of smart city views through web information clouding. The MI-

Search architecture and related core-component engines are presented in the paper as well as real examples applied to the city of Milan. Ongoing research work is devoted to study the problem of periodically refreshing the contents of the MI-Search-DB repository and to complete the acquisition of a dataset about the city of Milan to be employed for experimentation. Moreover, matching techniques combining both string-based techniques and position-based techniques are currently under development as well as techniques for content delivery based on similarity cluster exploitation. Finally, next-future activities will be focused on the development of a mobile prototype based on the presented ideas.

References

1. Bolchini, C., Curino, C., Quintarelli, E., Schreiber, F.A., Tanca, L.: Context Information for Knowledge Reshaping. *International Journal on Web Engineering and Technology* 5(1), 88–103 (2009)
2. Bolchini, C., Curino, C.A., Quintarelli, E., Schreiber, F.A., Tanca, L.: A Data-Oriented Survey of Context Models. *SIGMOD Record* 36(4), 19–26 (2007)
3. Castano, S., Ferrara, A., Montanelli, S.: Matching Ontologies in Open Networked Systems: Techniques and Applications. *Journal on Data Semantics V* (2006)
4. Castano, S., Ferrara, A., Montanelli, S.: Thematic Exploration of Linked Data. In: *Proc. of the 1st VLDB Int. Workshop on Searching and Integrating New Web Data Sources (VLDS 2011)*. Seattle, USA (2011)
5. Castano, S., Varese, G.: Next Generation Data Technologies for Collective Computational Intelligence, chap. Building Collective Intelligence through Folksonomy Coordination. Springer (2011)
6. Easley, D., Kleinberg, J.: *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press (2010)
7. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. *Nature* 435 (2005)
8. Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., Vakali, A.: Cluster-based Landmark and Event Detection on Tagged Photo Collections. *IEEE Multimedia Magazine* 18(1) (2011)
9. Schmeiß, D., Scherp, A., Staab, S.: Integrated Mobile Visualization and Interaction of Events and POIs. In: *Proc. of the 18th Int. ACM Conference on Multimedia*. Firenze, Italy (2010)
10. Sorrentino, S., et al.: Schema Normalization for Improving Schema Matching. In: *Proc. of the 28th Int. ER Conference*. Gramado, Brazil (2009)

Case Study on Educational Systems Evolution through a Smart Solution

Nicola Convertini¹, Giovanni Dimauro¹, Lorena Pocatilu²

¹Dipartimento di Informatica – Università di Bari, Italy

dimauro@di.uniba.it, nicolaconvertini@di.uniba.it

²Informatics in Economy Department – Academy of Economic Studies,
Bucharest, Romania

lpocatilu@yahoo.com

Knowledge society emphasizes the opportunity to put into practice the elements studied. All these can be done by involving all stakeholders in the educational preparation of future generations. To achieve an effective correlation between all parts of the education system it is necessary a smart education system. To highlight the need of smart system we start with the study of implications of modern technology in educational activities. To analyze the impact of online communication through modern technology in higher education of economics and business research we use the support of the students how have achieved a third-year, from the Faculty of Economics Cybernetics, Statistics and Informatics, Faculty of Commerce and the Faculty of Marketing from Academy of Economy Studies from Bucharest. Our goal was to stress the need to implement a smart education system to improve performance in higher education.

1.Introduction

This paper aims to analyze the concrete support of smart solutions for communication, use and storage of data to improve educational activity in higher education.

The first part of the paper introduces the elements of smart planet putting emphasis on the educational system. This part also highlights the advantages and the significant changes that smart systems can make possible in our daily life and work.

The second part characterizes a smart educational system and identifies the three elements that are the base of it interconnection (a resource sharing technology education), instrumentation (accumulation of necessary data) and intelligence (making decisions that enhance the learning process).

The third part presents a case study about the use of existing computer in the educational process. On the other hand, in this part we identify some points where the infrastructure can be upgraded to improve performance in higher

education economics and business. The case study is based on 100 responses from students in the third year of the Faculty of Economics Cybernetics, Statistics and Informatics, Faculty of Commerce and the Department of Marketing from the Academy of Economic Studies, Bucharest. Students had completed a questionnaire containing 10 questions on the use of online communications and the Internet in the educational process.

The fourth part presents the conclusions and underlines the importance of the use of cloud computing in higher educational system.

2. Smart Planet

A smart planet means to use new technologies advances to improve all aspects of our life or to make world to work better. We can define the concept of smart planet like new solutions for a better manage of infrastructure and resources to cater to the existing and modernity needs of the peoples [Naphade, 2011].

Technologies have become more widespread, at all levels of our society. The study of smart planet concept is significant because it's a quickly growing trend and a lot of researchers believe that it could introduce new important changes to our society. Only few years ago, smart planet was an unknown concept now it spreads like wildfire in every city and organization.

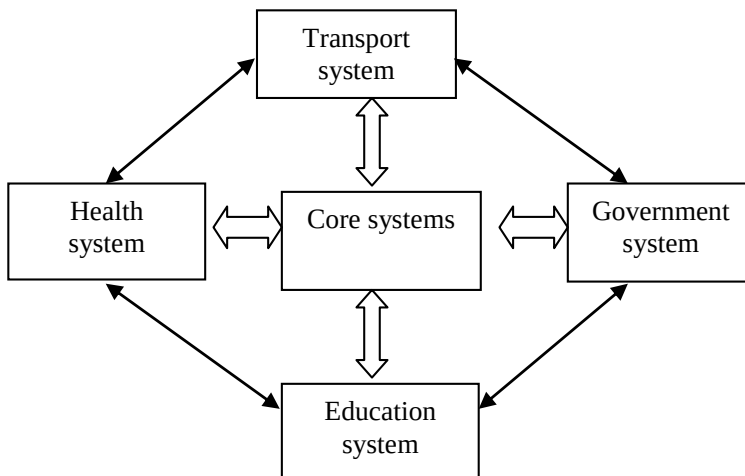


Fig. 1 – The core systems of the smart city and of the smart planet

To realize a smart planet we need smart cities. Smart city can integrates and optimizes different systems in order to achieve a new level of productivity, quality and efficiency. In our vision a smart city is based on use of the new technologies in all activities to improve the quality of work and life, to reduce cost and to increase general efficiency. All of these activities are profoundly

Case Study on Educational Systems Evolution through a Smart Solution influenced by the core/the base systems (fig. 1) of a city [Dirks et al., 2010]: education, health, transport, and government services.

This complex systems are systems of systems defined as “a complex structure in which the overall mission is performed through reliance on behaviors of component systems, all or partially external to the mission system m” [Parks, Jung and Ramotowski, 2004]. Today evolution is based on the production of knowledge and then it is more important to use efficiently smart solutions.

The transformation of city in smart city has been recognized of researchers [Dirks et al., 2010], [Yi-Chen, 2005], [Naphade, 2011] as a new stage of urban development. In our day to construct a smart city has become an international trend. Several ongoing projects around the world illustrate the opportunities and challenges of this transformation.

Many cities from Europe are implementing smart cities by starting with a transformation of vertical. So, they start with one area from the core system and are focus on: efficiency, performance and quality of life.

On the other hand is China, which building smart cities, are in view the all the core systems of a city. They start to create a strategic development plan for a smart city through surveying, analyzing, summarizing and other methods. In China, many cities have committed themselves to the construction and development of smart city. To build a smart city, they invest in a big number of projects in all categories and they have in view to make new cities. In China smart cities are seen as the solution to their urban growth woes. They start to use the smart solutions to build greener and more efficiency cities.

Education is one of the fundamental components that generate solutions to economy problems. Well educated and skilled people are the key elements for creating, sharing, disseminating and using knowledge effectively. A good economy requires a smart education system which is flexible and promotes creative, critical thinking, innovation.

3. Smart educational systems

Changes that occur each time and complexity of our society, which uses every new type of electronic communication solutions, have resulted in continued growth in the volume and diversity of service carried out in any field [Yi-Chen, 2005].

The existences of a smart educational system - focus on the efficient use of existing infrastructure and modernisation of it where is necessary – is considered crucial during an economic crisis when funds are limited. But most importantly, an education system should focus learning on the two key components: student and teacher.

Through an smart educational system, according to the latest data published by IBM, [IBM, 2010], teachers can analyze student data electronically - from academic results, to information on mobility and presence. This information will help teachers identify individual student needs and to individualize instruction as to improve the process of their results, while retaining confidentiality. Smart systems are also ensuring that universities do not bear alone the burden of education but also to other interested parties (as for example parents, banking,

government, public administration, firms and companies). They allow the inclusion of educational organizations and companies interested in each area through access to existing data - from academic results, state of completion of individual or group projects, until the stage of practical information and skills specific to each student - transforming universities into an ecosystem focused on student learning.

A smart educational system [Dirks et al., 2010] is based on three elements interconnection (a resource sharing technology education), instrumentation (accumulation of necessary data) and intelligence (making decisions that enhance the learning process).

In our society efficient interconnection can be achieved through a smart solution like cloud computing system.

Smart solutions for education have in view to improve the quality, performance and efficiency of learning. At the moment and for the near future, a smart solution for educational system is based on cloud computing. There are seven main reasons that could be invoked in support of investment in intelligent educational system, [BECTA, 2008]: flexibility, availability, accessibility, low cost, low time, strong connection and minimal hardware requirements.

We can describe the impact of the cloud computing on educational systems for each essential characteristic [NTSC, 2011].

Table 1 Benefits of application of cloud computing in education

On-demand self-service A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.	Institutions may decide to increase or decrease the allocated resources for training depending on the number of students and services they want to deliver
Broad network access Capabilities are available over the network and accessed through standard	Institutions can provide services ubiquitous (ubiquitous learning)
Resource pooling The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to	The Dynamic allocation of resources and the reduced waste results considerable savings in costs

consumer demand., and virtual machines.	
Rapid elasticity Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.	Educational institutions can deliver services without time delay.
Measured Service Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts.	Educational institutions can monitor the level and quality of each service offered monitor user activity on cloud infrastructure deciding which and how many services offer in the budget.

In this vision, the support of flexibility is an important issue. The hybrid cloud model (fig.2) is a good choice because it allows you to display only selected data and applications without sacrificing, for authorized subjects the features described in table 1. The cloud infrastructure is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

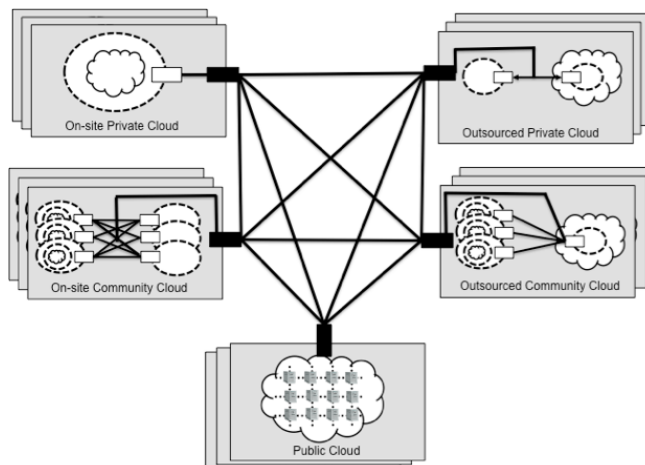


Fig. 2 – Hybrid cloud model

While the student in past, [Nicolescu, 2010], [OECD, 2008], learned how to manipulate objects and memorized actions, the student in the modern society learns how to think, learn, apply information to a task and produce knowledge.

For a successful implementation of smart educational systems in universities, teachers and technology developers should consider the following significant challenges: mobility, information and accessibility.

Smart educational system profoundly changed how research and creative activity are carried out through distributed access to research networks, grids and cloud computing, simulation, or virtual world research can be performed in real time. He also changed research, learning and innovation, by connecting systems so as to record, analyze and integrate data and respond intelligently to their specific needs. In short, we can revitalize the educational system so that they can become smarter and more efficiently. In this process, universities can increase and sustain the quality of the results of their students.

4. Case Study – methodology

To analyze the impact of online communication through modern technology in higher education of economics and business research we are using the support of the students how have achieved a third-year, from the Faculty of Economics Cybernetics, Statistics and Informatics, and the Faculty of Marketing from Academy of Economy Studies from Bucharest. In this case study we have tried to record the opinions of the implications of modern technology in educational activities, be they good or bad (well, normally there are all the answers). Our goal was to render the most interesting ideas that came from the results of that survey.

We have set up a 10 questions form for this analysis. The first eight questions were addressing the subject's opinion on implications of modern technology in educational activities. The questions dealt with both quantitative

Case Study on Educational Systems Evolution through a Smart Solution and qualitative issues. The last two questions were added in order to define the profile of the subject, and two open questions were appended at the end.

The subjects of the survey were selected from third year students at colleges said that they used as support for their educational website <http://acs.ase.ro/> (about 900 students). Given the uneven distribution of the visitors during one day, the observation was made during a week in order to record all the aspects related to sample of visitors. An algorithm was created to help our attempt to get the real structure of the group of visitors. Thus, each 40th visitor was asked to reply to the survey. Number 40 was chosen to be able a sample of 100 responders.

The answers were collected online, through a server side application and they were stored in a database. Following the collection, the data was processed through an application server side which returned statistical reports.

The questions have in view: whether online communication between teacher-student is essential in today's society, where online communication change student-teacher relationship, if students are interested in on-line documentation sources, which are the major advantages of using online resources, which are disadvantages use online resources, frequency and educational site uses elements from the site showing the most interest, if students are interested in working on real examples, if students are interested in online teaching projects line, where students consider beneficial relationship through the site by specialized companies for practical training.

It is essential for local development to have a smart strategy for a city. This strategy will help determine where and when to invest, will highlight the key steps and return on investment and can help define a timetable for integration / implementation of all systems, intelligent solution. This strategy should begin with analyzing the existing infrastructure.

5. Case Study – analyses

According to the replies, are a very large numbers of students who believe that in education is essential the online communication with the teacher 98.99%. We believe that this attitude represents the openness to new technologies.

The large number of those who believes that the online communication between teacher and student relationship is improved (table 2) helped us to believe that intelligent educational systems (based on the interconnection between educational partners) will contribute to proper conduct of educational activities.

Table 2 On-line communication and teacher-student relationship

User Rating	Percentage
improved	91,46%
unchanged	8,42%
worse	0,02%

It notes the large number of those who considered beneficial on-line access to resources necessary for their training.

It is generally noted a positive trend at the conduct of activities on-line documentation, so that the vast majority consider it useful activity 87.34% overall. Those who have acknowledged the usefulness of the online documentation option selected argued, in response to the open question related to the major advantages of using online resources. Responses show that most students prefer to use online resources to go smoother than the library. This case can be explained by:

- library books do not provide current information;
- libraries program that does not always correspond with the program students;
- during the research a book is much larger than the on-line documentation;
- can search a book is much smaller than the search in an electronic document;
- the possibility of accessing resources from different locations, without being obliged to be in a certain place.

Users also highlight the possibility of obtaining additional information on a particular topic through online communication with the teacher helping them to improve their knowledge.

In the open question related to the disadvantages of using online resources responses could be grouped in the following categories:

- Accessibility - some students have highlighted the difficulty of connecting to the Internet in some areas;
- Encouraging absenteeism - another problem identified by students was related to the hours they are considering participating in them are sufficient resources online.

Table 3 Frequency of use and percentage of users

Frequency of use	percentage of users
Several times a day always	3,81%
Several times a day during exams	27,83%
Several times a day during the project	68,36%

A special importance should be given the frequency with which the student returns to the site for educational training, documentation and information. There is a more pronounced recovery of students several times a day (table 2) during this project confirmed the need for correlation with the economic reality of educational activity. We must to highlighted the large number of students assigned to the site for documentation and information 78.34%, for training (13.16%), for links to other sites (9%) (Fig. 3).

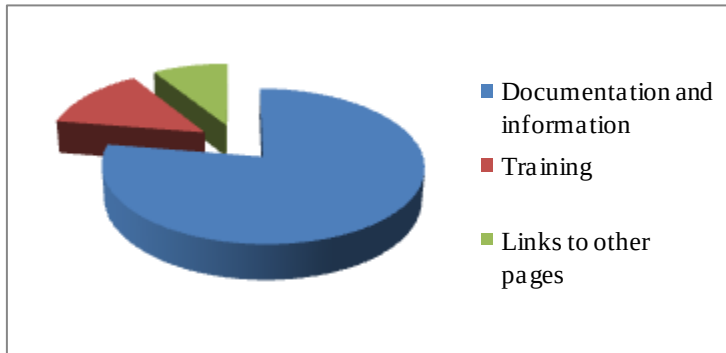


Fig. 3 – Information search site

The eighth question is one that emphasizes the importance for higher education to improve education in economics and business in conjunction with the economic reality of educational activity. Most students (99.18%) have stressed the interest in real examples to verify the theoretical learning.

When asked if they prefer to deliver projects on-line majority declares that it is very useful on-line feedback received from the teacher (97.32%). The small number (2.68%) of those who believes that are favoured by the classical teaching is represented by those who didn't use frequently online communication.

The last question is one that stresses the need for a cluster to enable the completion of educational partnerships between universities and businesses in industry, agriculture, construction, transport, communications, and computer science. Through this educational cluster may be correlated with the practice and theoretical work can be resolved the question of the difficulty of carrying out internships. Students found that it is entirely beneficial relationship through the site by specialist firms of practical training.

6.Conclusions

Based on the presented case study, online communication with the teacher is considered essential by the majority of students in an educational process. Also, the study stressed out the importance to improve economics and business oriented higher education by linking it with the economic reality of the business using smart systems. Students also considered as beneficial the existence of partnerships between universities and businesses, industry, agriculture, construction, transport, communications, IT that allows them to perform internships and much more.

Analysis of responses confirmed expectations about online communication with the teacher majority considered essential in the educational process has also highlighted the importance of higher education, improve education in economics and business in conjunction with the economic reality of educational activity, but the important conclusion is the very high percentage of those who consider the existence of beneficial partnerships between universities and

businesses in industry, agriculture, construction, transport, communications, computer to perform internships.

The development of smart educational systems depends on the evolution and quality of the global telecommunications network and equipment that connects to it. In many countries the telecommunications systems infrastructure that currently exists, allows the development of smart educational systems. It is recommended to involve all interested parties in improving the quality and performance of higher education, to support or to implement intelligent educational systems.

7.Acknowledgment

This work was supported from the European Social Fund through Sectoral Operational Programmer Human Resources Development 2007- 2013, project number POSDRU/89/1.5/S/59184“Performance and excellence in postdoctoral research in Romanian economics science domain”.

8.References

[Bathelt, 2007] Bathelt, H., Malmberg, A., Maskell, P., “Clusters and Knowledge Local Buzz, Global Pipelines and the Process of Knowledge Creation”, DRUID, Copenhagen Business School, ISBN 87-7873-128-3, 2007

[Batagan, 2009] Batagan, L., Mobile Devices and e-Learning, Economy Informatics, pp. 48-54, vol. 9, no. 1/2009, ISSN 1453-1305.

[BECTA, 2008] BECTA, Next Generation Learning, Becta, 2008, [Online], Available: publications.becta.org.uk/download.cfm?resID=37348

[Etzkowitz, 2002] Etzkowitz, H., The Triple Helix of University-Industry-Government: Implications for Policy and Evaluation, Working Paper 2002: 11, Science Policy Institute, [Online], Available: http://www.sister.nu/pdf/wp_11.pdf

[IBM, 2010] IBM, Education for a Smarter Planet - Cloud computing, virtualization and student data analytics can make our systems smarter, [Online], Available: http://www.ibm.com/smarterplanet/us/en/education_technology/nextsteps/index.html

[Naphade, 2011] Naphade, M., Banavar, G., Harrison, C., Paraszczak, J., Morris, R. Smarter Cities and their innovation challenges, Innovation Technology for Computer Professionals, Computer, June 2011, Volume 44, Number 6, pp 32-39

[Nicolescu, 2010] Nicolescu, L., Dima, A. M., „The quality of educational services – institutional Case Study from the Romanian Higher Education”, Transylvanian Review of Administrative Sciences, Nr.29, pp. 100-108, 2010

[NTSC, 2011] NTSC, National Institute of Science and Technology, ” DRAFT Cloud Computing Synopsis and Recommendations” Retrieved 29 September 2011. <http://csrc.nist.gov/publications/drafts/800-146/Draft-NIST-SP800-146.pdf>

[OECD, 2008] OECD, The Future of the Internet Economy , OECD Ministerial Meeting of the future of the Internet Economy, Seoul, Korea, 17-18 June 2008, [Online], Available: <http://www.oecd.org/dataoecd/20/41/40789235.pdf>

[Parks, et al., 2004] Parks, R. C., Jung, R. A. and Ramotowski, K. O., Attacking agent-based systems, 2004 IEEE 1st Symposium on Multi-Agent Security and Survivability , 2004, pp. 31.

Cloud Computing: a survey

Nicola Sfondrini¹, Gianmario Motta²,
University of Pavia

Department of Computer Engineering and System Science

Via Ferrata 1, 27100 Pavia (PV) – Italy

¹nicola.sfondrini01@ateneopv.it, ²motta05@unipv.it

We present a survey on the adoption of Cloud Computing. The purpose of our survey is an overview of benefits and issues. We found that approaches to Cloud Computing reflects the size of the company and not surprisingly cloud is managed by consolidated ITIL and COBIT standards while it would require a specific service level management.

1. Introduction

Our analysis is based on data collected through the portal questionpro.com (<http://cloudcomputing-unipv.questionpro.com>) in the period 07/11-09/11 and during these two months 35 companies responded to the survey. Differently from other surveys carried out previously, we have imposed two important limitations to the interviewed companies:

- use at least cloud systems under evaluation or test;
- no cloud vendors were accepted to avoid their conditioned responses.

2. Survey findings

2.1 Positioning

To properly contextualize our analysis we provide below a brief introductory analysis to correctly position our sample. Our sample consists of companies from 17 different countries with an obvious majority of Italian companies (40%) due to frequent relationships of our laboratory with these companies.

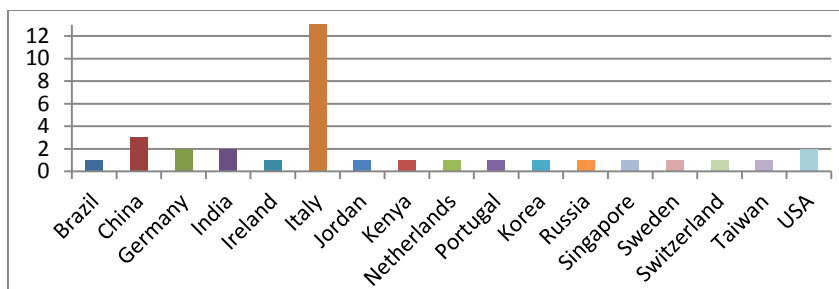


Figure 1 - Geographical Distribution

The sample was then classified according to different business features.

Figure 2 shows the classification of the companies surveyed on the basis of industry sector showing a quite homogeneous distribution. The slight majority of the category "Educational Services" ($\approx 14\%$) can be explained by our close relationship with foreign universities while the width of the sectors "Telecommunications" ($\approx 11\%$), "Media" ($\approx 9\%$) and "Banks" ($\approx 9\%$) reflects, in our opinion, the areas where Cloud Computing has a high strategic importance. In fact, the high frequency resources utilization, both by internal personnel and by customers, mainly localized in specific time slots, can direct these categories of companies to adopt cloud technologies. The voice "Others" includes two companies of Access Service Provider and Physical Security.

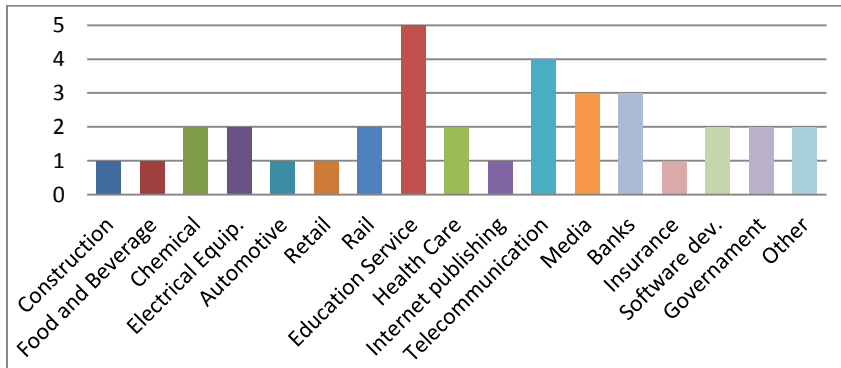


Figure 2 - Business Areas Distribution

We have also analyzed the sample by number of employees. Figure 3 shows a significant distribution towards the two extremes of the categories that we have considered. In fact, the fields "50 or less" ($\approx 23\%$), "1.001-10.000" ($\approx 37\%$) cover about 50% of our sample.

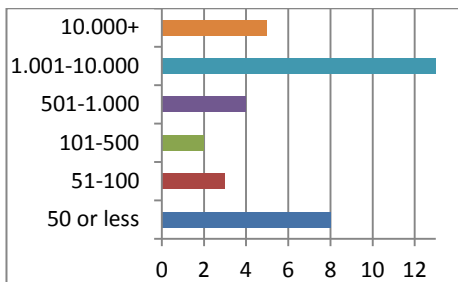


Figure 3 - Organization's employees

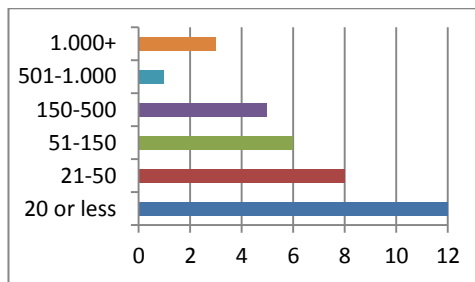


Figure 4 - IT department's headcount

This unexpected factor could lead to important considerations about the types of companies that are more directed to the adoption of cloud technologies. The limited robustness of our survey and the reduced sample does not allow us to express specific considerations, but opens new and important scenarios for future research. The sample was also classified according to the number of employees of the IT department. This sub-division has allowed us to have an important snapshot of the complexity of the IT

infrastructure. Only a limited number of the surveyed companies has a large IT department and most IT departments have less than 50 FTE (56%). Figure 4 shows that only few IT departments are of really large size. IT departments with over 1000 FTES are in companies in which IT has a strategic role (i.e. "Telecommunications" and "Media"). Finally the sample was analyzed by company's revenues. Even in this case, we have a strong concentration in two extremes with a clear majority of companies with an income of over \$ 200 M (or M€) as shown in Figure 5.

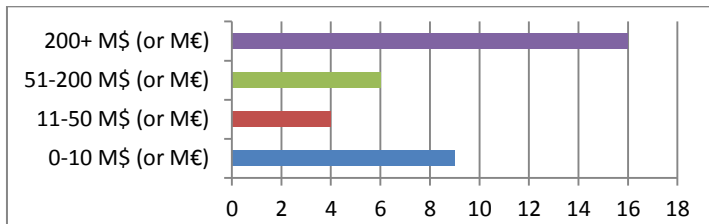


Figure 5 - Sales turnover or equivalent

2.2 Benefits and Issues provided by Cloud Computing

Our survey has considered the different viewpoints and benefits and issues that may result from the adoption of Cloud Computing. The uncertainty emerging by the diverse definitions on Cloud Computing has led us to ask companies to choose the definition that best describes cloud computing. The proposed definitions are:

1. A new buzzword
2. The repackaging of outsourcing and grid computing
3. A new technological approach for the business evolution
4. A revolutionary approach to IT

The difference between the visions proposed by the first two statements, which have a negative connotation, and the other two, which have a strong positive acceptance, it is immediately evident. In addition, the results are particularly of interest because, despite having investigated companies that use cloud systems, we have obtained a huge majority in the choice between two conflicting answers.

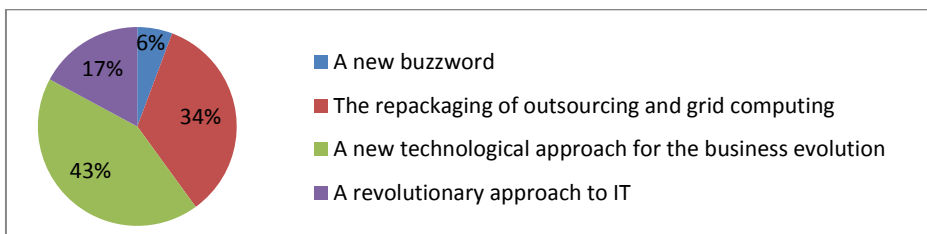


Figure 6 - Opinions on Cloud Computing

Figure 6 shows that 34% of companies consider cloud computing as the

renaming of existing technologies while 43% consider this technology a revolutionary approach. This result indicates again a high degree of uncertainty and confusion that characterizes this technology across enterprises. The adoption of a new technology always involves a number of advantages and issues that, on one side, are tangible and easily measurable by specific variables, and on the other side, are intangible and related to the sensations of users. Based on the characteristics of cloud systems emerged from our SLR, we have provided a list of advantages and problems to which responders had to assign a mark from 1 to 5. The benefits to be evaluated were:

- **Lower Energy Consumption (Greener IT):** turning off unused resources avoids the presence of idle servers allowing a greener IT;
- **Cost transparency (pay-as-you-go):** the user knows at all times its actual cost which is calculated on the real use of resources:
- **Easy to use:** the web interfaces facilitate the usability of the applications;
- **Shortest deploy time:** time to configure goes down with clouds;
- **Faster response to business request:** the dynamic resources allocation allows to be more reactive to any changes in business requirements;
- **Easy upgrades/updates:** any updates performed on cloud systems is immediately available to all its users;
- **Cost reduction:** the pay-as-you-go payment system allows you to use, and therefore pay, only the resources actually used;
- **Lower initial investment:** the use of a public cloud eliminates all of the cost of purchase and operation of enterprise servers
- **Flexibility of use:** cloud systems are strongly oriented to providing services in this way, their remodeling and the resulting change of use;
- **Scalability of use:** resources are dynamically allocated based on the level of requests received by the system allowing to manage any workload peaks.

Figure 7 shows the opinions expressed by the companies showing a strong positive advice on the proposed statements.

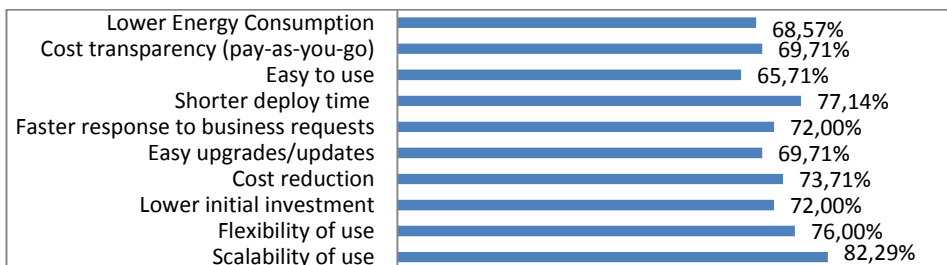


Figure 7 - Cloud Computing Benefits

In fact, all the voices have a degree of consensus over 65%. Considering the entries with the highest rating, over 75%, we notice the presence of two items that are strongly related to the main features of a physical cloud infrastructure: "Flexibility" (76%) and "Scalability" (≈82%). Another voice above this threshold evaluation is to reduce deployment time (≈77%) and is strongly correlated with the two previous entries. By lowering the threshold to 70% we can note that the

focus shifts from technical to strictly economic factors: “cost reduction” ($\approx 74\%$), “lower initial investment” (72%) and “fast response to business requirements” ($\approx 72\%$). This order of consent may be explained, in our opinion, as a logical sequence of events. In fact, the technical advantages are immediately perceived in the initial stages of implementation and use, while the effective economic benefits should be considered only in the medium or long term. Instead, the issues to be evaluated were:

- **Data security:** the access to data through Internet and the lack of exact knowledge of the location of data storage involves several security issues;
- **Loss of control:** the external management of infrastructure and its dynamics strongly decreases the ability to control it;
- **Hard integration with existing applications:** integration with existing applications, non-migratable directly on cloud systems can be complex;
- **Privacy:** the management of sensitive data by a provider increases concerns on their security;
- **Cumbersome SLA definition:** the complexity of cloud infrastructure complicates the definition and the respect of the SLA's parameters;
- **Reliability of cloud providers:** entrust complete control of enterprise systems to an outside company requires a good level of communication;
- **Performance measurement:** the measurements of the real system's performance are complex due to the low degree of control available on it;
- **Overlap of multiple legal regulation:** the use of a foreign provider and the location of infrastructure in a foreign country complicates the precise definition of the laws applicable in the event of violation of contract;
- **Difficult to compose different services:** the composition of services offered by different providers is still difficult due to lack of specific standards.

As in the case of the benefits provided by cloud systems, although with a lower degree, companies are shown in agreement with the problems we listed. The lower average percentage shows how this technology involves more benefits than problems and is therefore convenient for their business. In fact, Figure 8 shows that there is no percentage above 75% and shifts the lower limit from 65% , presented in the assessment of benefits, at 55% .



Figure 8 - Cloud Computing Issues

The values beyond 70% show two problems that are related to service level management, "Performance measurement" ($\approx 73\%$) and "Loss of Control" ($\approx 71\%$). These factors appear to be the downside of cloud computing. In fact, the

possibility to entrust the whole systems management to an external provider and the highly dynamic infrastructure greatly complicates the process of management and control over it. Around 70% we find instead three issues: "Privacy" ($\approx 70\%$), Data Security ($\approx 70\%$) and "Cumbersome in the SLA definition." This fact confirms the findings of our SLR highlighting challenges such as performance in the first place, and those relating to security are critical issues in the adoption of cloud systems. These four factors, combined together, lead the complexities in the definition of robust and accurate SLA that, in fact, achieves a rating of approximately ($\approx 70\%$).

2.3 Delivery models

We have also assessed various technical approaches with which companies are facing this technology. The different delivery models available allow companies to design systems that reflect their needs and respond quickly to business changes. These different models reflect different degrees of abstraction and a diverse level of user control that increases gradually down the three layers. Figure 9 shows the percentage of use of the diverse delivery models in the surveyed enterprise revealing a clear majority in favor of SaaS (63%). In addition, SaaS has only 9% of absence indicating the large propensity of companies to its use. This phenomenon can be explained by the wide range of public cloud SaaS services that have even anticipated their appearance compared to the same cloud.

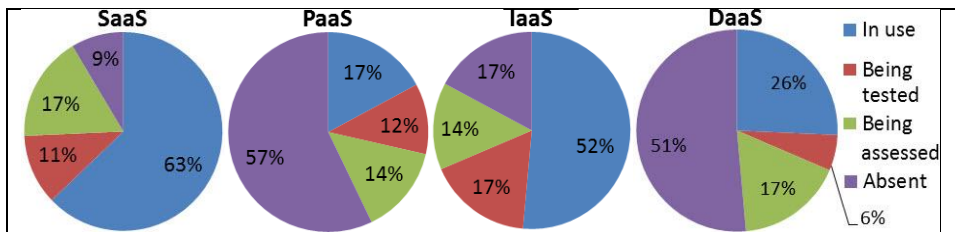


Figure 9 - Delivery Models Diffusion

A good level of use is also obtained by IaaS (52%) that enables companies to have better control over the underlying infrastructure and, even in this case, the percentage of absence is quite low. The use of PaaS is instead strongly reduced 17%. The low prevalence of this delivery layer, highlights as a lack of development standards, and the consequent forced bond to the used development platform, can lead to only use the two extreme levels of the pyramid. This factor can open new and important scenarios in which developers of PaaS systems must agree on the creation of standards and the possibility of direct interaction of different systems if they want to take space in this emerging market. Finally, the low level of use of the DaaS model (26%) may be explained by the presence of security and privacy problems that this layer presents hindering its mass adoption by companies. Next the analysis has increased its precision and depth by analyzing the trends of adoption of cloud systems from 2007 to present as shown in Figure 10.

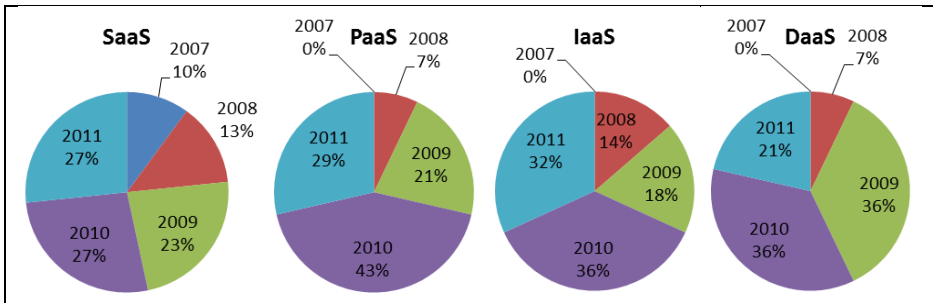


Figure 10 - Year of Adoption

All delivery models have an increasing trend from 2007 to 2010, which is the year with the highest percentage of cloud systems adoption, and only in 2011 there is a slight decrease in this trend. These results may be due to their consolidation in the market and the increasing levels of maturity.

2.4 Deployment models

The delivery models can be deployed in four different ways that present diverse strengths and weaknesses. For example, Private cloud allows a greater degree of control but implies an initial expense for the infrastructure construction; conversely, the public cloud allows a significant reduction in costs and time, but it entrusts the entire management to an external provider. As shown in Figure 11 we can easily note that the private cloud is the most common deployment model (50%). This result can be explained through two different perspectives:

- companies with an existing internal IT infrastructure has changed their systems in "cloud logic" to exploit its benefits;
- companies that can afford to build a solid IT infrastructure may be more attracted to this deployment model to maintain direct control over their systems.

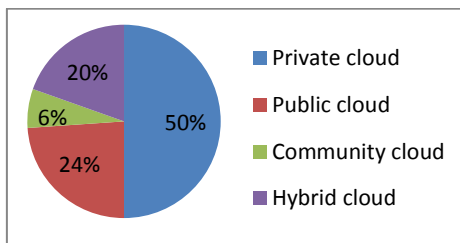


Figure 11 - Deployment Models

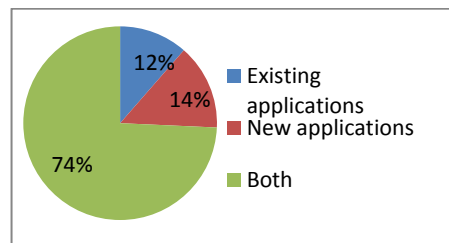


Figure 12 - Deployed applications

Public Cloud (24%) and Hybrid Cloud (20%) show an almost even balance. This statement, however, has an important implication because hybrid cloud need an internal infrastructure that can be expanded to satisfy any peak of demands, on public cloud resources. Another kind of hybrid cloud utilization is based on the diversification of services offered into the two different platforms: a company can maintain inside its internal infrastructure the most critical services

and sensitive data, providing the remaining services through a public cloud. The deployment model with the lowest adoption rate is the Community Cloud (6%) that involves the use of a shared cloud platform among companies that have similar needs. Even in this case the low percentage may find a possible interpretation of the attention that companies demonstrate to the problems of QoS control and security management. It's so clear that the issues, even with a rating lower than the benefits, as shown in the above paragraph, strongly affect the choice of the deployment model. The survey suggested two simple but useful questions to measure the impact that cloud computing involves application portfolio of the companies that have adopted it. Figure 12 shows how cloud computing has a very strong impact on the application portfolio by encouraging the enterprise to the use of new applications and, at the same time, convincing them to migrate existing ones on the new technology platform. Only 12% of companies use cloud computing to migrate their applications, taking advantage of its benefits, and 14% apply this technology by extending their portfolio with new applications, increasing the capabilities provided to both employees and end users. The second question was designed to investigate the type of applications that are more used leveraging the cloud paradigm. The decision to deploy these applications on cloud platforms are closely tied to specific benefits and problems of this new technology. Figure 13 shows a clear majority of companies that use CRM systems on cloud platforms ($\approx 22\%$). This phenomenon can be explained by the use of the front-end of this type of application by end users, focused in specific times of day. The use of cloud systems allows to cope with peak of demands, located at this particular time slot, allocating additional resources and remove them at the end of it.

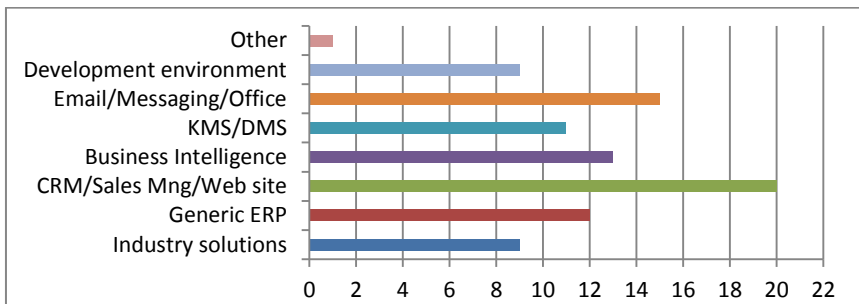


Figure 13 - Deployed applications

Other types of widely used applications on cloud platforms are the "office suites" ($\approx 17\%$) and business intelligence systems ($\approx 15\%$). The explanation for this result is quite simple and obvious:

Office suites are currently the most popular SaaS applications and allow to share and editing documents in real time by multiple users;

Instead, the business intelligence systems need large amounts of resources to perform complex statistical and mathematical calculations. The frequency of these calculations is not, however, very high and the use of allocated resources only as they are needed enables a significant cost savings.

Finally, under the voice "Other", a university that has participated in the survey, indicated the use of Cloud Computing to create an infrastructure of high performance computing used for scientific computing.

2.5 Service level management

The last part of our survey analyzes the various approaches to cloud systems management. First, we investigated what kind of disaster recovery systems are adopted by the surveyed companies. Figure 14 shows a clear majority of direct management by the company (57%) followed by a mixed approach (31%), with both internal management and external (supplied by a provider). This result is directly related to the widespread use of hybrid and private cloud (70% if considered together) in our sample and shows that cloud computing still requires consolidated systems for managing critical situations. In fact, just 9% entrusts the management of these situations to an external supplier by shifting, in this way, the responsibility for monitoring the quality of service to a third party while only 3% does not include the application of such systems. A strategic point for service level management is the adoption of specific frameworks to guide the company both in building infrastructure to provide its business services and in the management of their performance levels. Figure 15 shows the percentage of adoption of the framework by the surveyed companies. A predictable result highlights the great majority of use of ITIL (47%) followed by Cobit (17%). In fact, these two frameworks are currently the most widely used also for the management of traditional IT infrastructure. The use of specific frameworks for web services, WS-Agreement and WSLA, is reduced only to 6% showing how companies are still strongly tied to traditional approaches and are more inclined to manage the IT infrastructure rather than directly control the service levels provided to the end user.

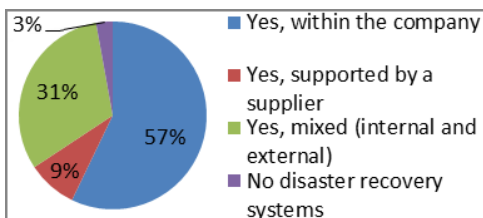


Figure 14 - Disaster Recovery

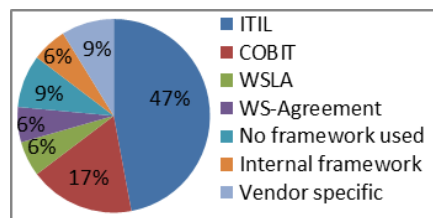


Figure 15 - Reference Framework

An approach modeled on the new technology is indicated by 9% of use of vendor-specific frameworks and 6% of use of internal framework. Instead, a completely unexpected result is the lack of framework for the management of IT systems into 9% of companies surveyed.

3. Conclusion

Our survey starts by the results obtained in previous studies and has the main purpose of mitigating some of their possible limitations and to provide an updated overview of the cloud computing in a company's vision (Bitcurrent 2011, Feuerlicht, Burkon and Sebesta 2011, Motamedian 2011). The survey

reveals an extremely important fact: after an initial distrust against this technological paradigm, today there is a strong tendency to consider the cloud systems as a key evolution step of enterprise computing. Despite this growing interest in this new technology on the market there are a wide range of heterogeneous products supported by a lack of an official definition. Some academic studies have preferred to describe the cloud with a specific set of characteristics rather than a simple definition but creating, at the same time, a broad vision in which any IT service, usable through the Internet and a based on the pay-as-you-go payment method, can be classified as a cloud computing service (Mell and Grance 2009). The paradigm of cloud, if properly and consciously applied, can be an important opportunity for the creation of an IT system able to respond quickly to changing business requirements. However, there are still some important issues that are highlighted in all the different clusters in our sample, thus slowing its global spread. This result proves that cloud computing, even if it provides a lot of benefits, it is not the "Eldorado" of enterprise systems but need a robust controls to ensure the quality of service. For this reason it is necessary that companies have a service-oriented architecture (SOA) in order to mitigate the risks of adoption of cloud computing, and to fully exploit the benefits carried by its adoption (Linthicum 2009). The results about the adoption of different levels of delivery and deployment still show the need for direct management of infrastructure by the companies, through the adoption of private clouds and hybrids, while the low adoption of PaaS reflects the lack of standardization which can lead to the risk of being bound with the vendor's proprietary technology. For this reason, the decisions about cloud computing adoption require an accurate cost-benefit analysis and considerations about the capabilities of the available infrastructure and the flexibility of the services (Vouk 2008, Kim et al. 2009).

Bibliography

- Bitcurrent. 2011. Bitcurrent Cloud Computing Survey 2011: Cloud Adoption, concerns, and motivations.
- Feuerlicht, G., L. Burkon & M. Sebesta (2011) Cloud Computing Adoption: What are the Issues?
- Gartner. 2011. Gartner Says Worldwide Software as a Service Revenue Is Forecast to Grow 21 Percent in 2011.
- Kim, W., S. D. Kim, E. Lee & S. Lee. 2009. Adoption issues for cloud computing. 3-6. ACM.
- Lachal. 2011. Trends to Watch: Cloud Computing Technology, in Ovum Trends Brief.
- Linthicum, D. S. 2009. Cloud computing and SOA convergence in your enterprise: a step-by-step guide. Addison-Wesley Professional.
- Mell, P. & T. Grance (2009) The NIST definition of cloud computing. National Institute of Standards and Technology, 53.
- Motamedian, B. (2011) Users' Perception Towards "Cloud Computing".
- Vouk, M. A. (2008) Cloud computing—Issues, research and implementations. Journal of Computing and Information Technology, 16, 235-246.

Cloud Computing Literature Review

Gianmario Motta¹, Nicola Sfondrini²

University of Pavia

Department of Computer Engineering and System Science

Via Ferrata 1, 27100 Pavia (PV) – Italy

¹motta05@unipv.it, ²nicola.sfondrini01@ateneopv.it

We present a review on Cloud Computing through a systematic analysis of 79 articles, published in first rank journals and conferences in 2008-2011. We identify topics, issues, research strategies in the literature. The paper is structured in two parts: the first one describes what Systematic Literature Review is, and the second one illustrates the step to evaluate and select research work.

1. Introduction

SLR can be defined as “explicitly formulated, reproducible and up-to-date summary” (Egger, Smith and O’Rourke 2001) that includes and extends the statistical results of a meta-analysis methodology. As opposed to narrative reviews, it is based on a structured method that is explicitly specified.

2. Systematic Literature Review (SLR): model & method

Our SLR includes:

1. Question formulation
2. Source selection
3. Quality assessment & data extraction
4. Synthesis and summary
5. Results interpretation

2.1 Question Formulation

Our objective is to identify initiatives, experiences and viewpoints on Cloud Computing. Our key question *What cloud computing really is and what level of maturity has reached* is refined into Research Questions (RQs):

- **RQ1:** What aspects are common in definitions of Cloud Computing?
- **RQ2:** What are the most discussed topics on this technology?

From SLR we expect:

- Overview on the "state of the art" of Cloud Computing
- The identification of gaps in current research, solutions, trends and future research and suggestions to the community of researchers.

Papers are classified into classes shown in Table 1 (Yang and Tate 2009) (Sriram and Khajeh-Hosseini 2010). Eventually, 79 articles were classified and

analyzed by year of publication, geographical area, research method and publication.

Primary Domains	Sub-domains	Contents
General Overview	Introduction, Definition	Guidelines, definitions, introductions and general aspects.
Technological Issues	Network, Storage	Technology and components, mechanisms and architectures
Service Level Management	Performance, Security	Performance studies to refine workflow setting, load balancing and security.
Business Issues	Cost, Legal Issues	The economic value for providers and users and legal issues
Case studies	/	Real world applications and scenarios
Other		Papers that cover multiple domains

Table 1 – Cloud issues classification framework

2.2 Source Selection

The SLR has been performed through English-based web search engines:

- IEEE Computer Society
- ISI Web of Knowledge
- ACM Digital Library
- Google Scholar
- Science direct
- The SCOPUS Database

The choice of using only sources in English reflects the small number of relevant documents in other languages. Those sources are appropriate since they contain the work published by journals, conferences and workshops of high impact research community. We have tried several search strings. In particular, have been used the keyword “Cloud Computing”.

2.3 Studies Selection

This selection is performed by inclusion and exclusion criteria arising on the research question. In a first step selection criteria were interpreted liberally and exclusions were only made because of the title, abstract and introduction. We will consider only the following “study types”:

- **Case Study:** an exhaustive investigation on actual practices;
- **Theoretical:** a study that includes introductions to a particular subject
- **Surveys:** a study with a statistical treatment of collected data;
- **Simulation:** a study that reports simulation methods and related results;
- **Position Paper:** a study that presents opinions about a specific issue;
- **Literature Review:** a study that collects information on a particular.

Our methodology includes three selection's stages:

- In a **first stage**, the search string runs on selected sources. An initial set of studies is obtained by reading of the title, abstract and introduction;
- In a **second stage**, exclusion criteria reflect this rules: short papers, non-

English papers, non-International Conference or Workshop papers;

- In a **third stage**, selection is based on compliance of the contents to research questions.

3. Results: a statistical analysis

3.1 Time distribution

There are no significant articles before 2008, because the term “Cloud Computing” was coined in 2006, and only at the end of 2007 researchers started to consider this topic. The articles considered in our research range of time from 2008 to first quarter 2011. Our selection criteria considers the number of citations received and the impact factor. The reduced amount of time between the publication of considered articles and our analysis, however, raises an issue on the real value of citations. To normalize results, we weight citations by dividing the number of publications of each year for the number of years available to be cited as shown in Table 2. The number of articles published in 2011 has not been considered because too recent.

Year	# of Publication	%	Weighted # of Publication	%
2008	18	23%	6	12%
2009	38	48%	19	40%
2010	21	27%	21	44%
2011	2	2%	2	4%
Total	79	100%		100%

Table 2 - Annual counts of research publications

Figure 1 shows the distribution of articles by comparing the real and weighted values. The steady growth of publications is apparent.

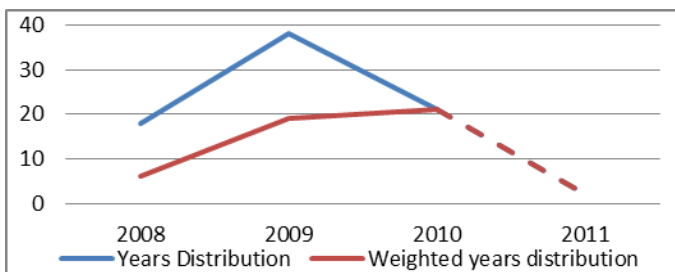


Figure 1 - Annual counts of research publications

3.2 Geographic distribution

The geographical classification highlights a predictable overwhelming America (Table 3). Table 4 shows the dominance of U.S. in research with 94% of the articles and a marginal role of Canada.

Continents	Publications	%
Europe	19	24%
America	48	61%
Africa	1	1%
Asia	6	8%
Oceania	5	6%
Total	79	100%

Table 3 - Counts of publications organized by continents

America	Publications	%
USA	45	94%
Canada	3	6%
Total	48	100%

Table 4 - Counts of publications in America

Europe (Table 5) has a quarter of the selected articles while the ROW accounts 15%. UK (37%) and Germany (32%) provide the main contribution, followed by a surprising Netherlands with over 20%. Table 6 shows Asian countries in which China scientific community has the major contribute (total 3).

UE Country	Publications	%
UK	7	37%
Germany	6	32%
Italy	1	5%
Holland	4	21%
Greece	1	5%
Total	19	100%

Table 5 - Counts of publications in UE

Asia	Publications	%
China	1	17≈%
Honk Kong	1	17≈%
Japan	2	33≈%
Iran	1	17≈%
Taiwan	1	17≈%
Total	6	100%

Table 6 - Counts of publications in Asia

3.3 Topic-Oriented Classification

Table 7 shows that “General Issues” (28%) and “Service Level Management” (26%) are the most frequent topics while only the 8% of articles propose cloud technologies for new applications. On Table 8, we observe the shift of research focus from introductory, distributed in the first two years, to the performance and quality of service, that become a key research in 2010.

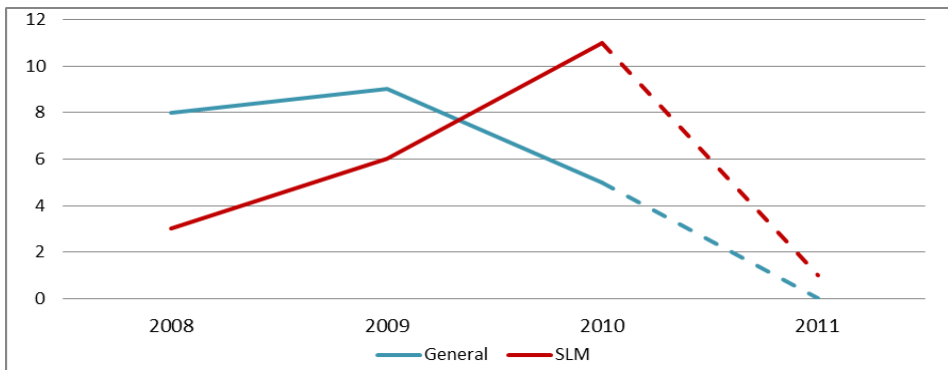
Primary Domains	# of Publications	%
General Issues	22	28%
Technological Issues	11	14%
Service Level Management	21	26%
Business Issues	16	20%
Uses Cases	6	8%
Other (SLR)	3	4%
Total	79	100%

Table 7 – Primary domains distribution

Years	General Issues	Service Level Management
2008	8	3
2009	9	6
2010	5	11
2011	0	1
Total	22	21

Table 8 - Topics distribution comparison by years

In Figure 2 we observe a growth of articles on SLM. In 2010, introductory articles drop about 54% while articles on security and performance increase of approximately 55%.

**Figure 2 - Topics distribution comparison by years**

The articles have been further divided into two sub-domains for a deeper analysis as shown in Table 9.

General Issues	Quantity
Introduction	15
Definition	7
Total	22

Technological Issues	Quantity
Network	6
Storage	5
Total	11

SLM	Quantity
Performance	12
Security	9
Total	21

Business Issues	Quantity
Cost	10
Legal issues	6
Total	16

Table 9 – Sub-domains articles distribution

"General Issues" mainly contain introductory articles (68%), while a smaller part (32%) focus on definition. Despite their lower proportion, definition articles reflect that cloud computing is still an undefined and somewhat vague concept.

The domain "Technological Issues" shows a solid balance of the two sub-domains "Network" (55%) and "Storage" (45%). The domain "SLM" shows a slight majority of articles on performance, that is of great interest to researchers; for, cloud performance is critical to ensure high levels of service and to exploit cloud potential. A substantial part are articles on security (43%) showing a series of new adoption problems. Finally, the domain "Business Issues", with the 62% of articles about cost and economics of cloud systems, confirms a great interest on cost saving. An important part are legal issues (38%) raised by the peculiar architecture of Cloud Computing that, alike security issues, slows down mass adoption. Figure 3 is a graphical representation to show "on the fly" the distribution of articles on the topics considered.

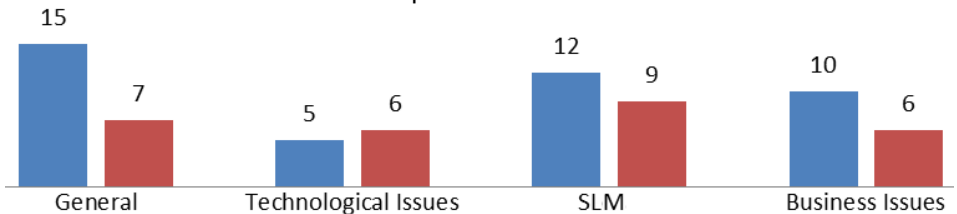


Figure 3 – Sub-domains distribution

4. Conclusions

We have illustrated the current academic research landscape on cloud computing, and classified the most discussed topics and has highlighted relevant trends. According to our statistical analysis, the number of researches shall increase exponentially thanks also to the growing economic interest on this technology. In the analysis, we have found different definitions that still show conceptual uncertainty. On the other side, National Institute of Standards and Technology describes cloud computing by a set of specific features and delivery/deployment models. In the analysis we have found that almost all articles, even in different domains, have at least addressed technological issues related to the adoption of cloud systems. These issues are analyzed, in most cases, in the provider's perspective, while the quality of the service from the end-user's viewpoint is almost neglected. Relevant work on the effective use of PaaS, that might be a strategic element in the future competition of vendors, is still scarce. Also standards are lacking, such as unified APIs, and interoperability of diverse systems so far is not granted. The analysis of business impact looks in an early stage. For, the assessment of benefits is not yet mature. Also, critical ethical and legal issues are still to be solved and, therefore, may delay or prevent a mass adoption. In short, this paper has highlighted the directions of academic research. However, vendors tend to tailor cloud characteristics on their offerings (Pastaki Rad et al. 2009). The analysis of this very different point of view is a target of a further research.

Bibliography

Armbrust, M., A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin & I. Stoica (2010) A view of cloud computing. Communications of the ACM, 53, 50-58.

Armbrust, M., A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin & I. Stoica (2009) Above the clouds: A Berkeley view of cloud computing. EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28.

Aymerich, F. M., G. Fenu & S. Surcis. 2008. An approach to a Cloud Computing network. 113-118. IEEE.

Buyya, R., C. S. Yeo, S. Venugopal, J. Broberg & I. Brandic (2009) Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems, 25, 599-616.

Calheiros, R. N., R. Ranjan, A. Beloglazov, C. A. F. De Rose & R. Buyya (2011) CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software: Practice and Experience, 41, 23-50.

Carolan, J., S. Gaede, J. Baty, G. Brunette, A. Licht, J. Remmell, L. Tucker & J. Weise (2009) Introduction to Cloud Computing architecture-White Paper.

Chappell, D. (2009) Introducing the windows azure platform.

Chun, B. G. & P. Maniatis. 2009. Augmented smartphone applications through clone cloud execution. 8-8. USENIX Association.

Dash, D., V. Kantere & A. Ailamaki. 2009. An economic model for self-tuned cloud caching. 1687-1693. IEEE.

De Assuncao, M. D., A. Di Costanzo & R. Buyya. 2009. Evaluating the cost-benefit of using cloud computing to extend the capacity of clusters. 141-150. ACM.

Delic, K. A. & J. A. Riley. 2009. Enterprise Knowledge Clouds: Next Generation KM Systems? , 49-53. IEEE.

Dornemann, T., E. Juhnke & B. Freisleben. 2009. On-demand resource provisioning for BPEL workflows using Amazon's elastic compute cloud. 140-147. IEEE Computer Society.

Egger, M., G. D. Smith & K. O'Rourke (2001) Rationale, potentials, and promise of systematic reviews. Systematic reviews in health care: Meta-analysis in context, 3-19.

Ekanayake, J. & G. Fox (2010) High performance parallel computing with clouds and cloud technologies. Cloud Computing, 20-38.

Erdogmus, H. (2009) Cloud computing: does nirvana hide behind the nebula? IEEE software, 4-6.

Foster, I., Y. Zhao, I. Raicu & S. Lu. 2008. Cloud computing and grid computing 360-degree compared. 1-10. IEEE.

Ganon, Z. & I. Zilbershtein. 2009. Cloud-based Performance Testing of Network Management Systems. 1-6. IEEE.

Greenberg, A., J. Hamilton, D. A. Maltz & P. Patel (2008) The cost of a cloud: research problems in data center networks. ACM SIGCOMM Computer Communication Review, 39, 68-73.

Grossman, R. & Y. Gu. 2008. Data mining using high performance data clouds: experimental studies using sector and sphere. 920-927. ACM.

Grossman, R. L. (2009) The case for cloud computing. IT professional, 11, 23-27.

Grossman, R. L., Y. Gu, M. Sabala & W. Zhang (2009) Compute and storage clouds using wide area high performance networks. *Future Generation Computer Systems*, 25, 179-183.

Gu, Y. & R. Grossman. 2008. Exploring data parallelism and locality in wide area networks. 1-10. *IEEE*.

Hazelhurst, S. 2008. Scientific computing using virtual high-performance computing: a case study using the Amazon elastic computing cloud. 94-103. *ACM*.

Hoefer, C. & G. Karagiannis (2010) Taxonomy of cloud computing services.

Hoffa, C., G. Mehta, T. Freeman, E. Deelman, K. Keahey, B. Berriman & J. Good. 2008. On the use of cloud computing for scientific workflows. 640-645. *IEEE*.

Hou, Y., M. Zafer, K. Lee, D. Verma & K. K. Leung. 2009. On the mapping between logical and physical topologies. 1-10. *IEEE*.

Jaeger, P. T., J. Lin, J. M. Grimes & S. N. Simmons (2009) Where is the cloud? Geography, economics, environment, and jurisdiction in cloud computing. *First Monday*, 14.

Johnson, J. L. (2009) SQL in the Clouds. *Computing in Science & Engineering*, 12-28.

Joint, A., E. Baker & E. Eccles (2009) Hey, you, get off of that cloud? *Computer Law & Security Review*, 25, 270-274.

Khajeh-Hosseini, A., I. Sommerville, J. Bogaerts & P. Teregowda (2011) Decision Support Tools for Cloud Migration in the Enterprise. *Arxiv preprint arXiv:1105.0149*.

Khajeh-Hosseini, A., I. Sommerville & I. Sriram (2010) Research challenges for enterprise cloud computing. *Arxiv preprint arXiv:1001.3257*.

Klems, M., J. Nimis & S. Tai (2009) Do clouds compute? a framework for estimating the value of cloud computing. *Designing E-Business Systems. Markets, Services, and Networks*, 110-123.

Lagar-Cavilla, H. A., J. A. Whitney, A. M. Scannell, P. Patchin, S. M. Rumble, E. De Lara, M. Brudno & M. Satyanarayanan. 2009. SnowFlock: rapid virtual machine cloning for cloud computing. 1-12. *ACM*.

Lenk, A., M. Klems, J. Nimis, S. Tai & T. Sandholm. 2009. What's inside the Cloud? An architectural map of the Cloud landscape. 23-31. *IEEE Computer Society*.

Li, X., Y. Li, T. Liu, J. Qiu & F. Wang. 2009. The method and tool of cost analysis for cloud computing. 93-100. *IEEE*.

Llorà, X., B. Ács, L. S. Auvil, B. Capitanu, M. E. Welge & D. E. Goldberg. 2008. Meandre: Semantic-driven data-intensive flows in the clouds. 238-245. *IEEE*.

Mathew, R. & R. Spratz. 2009. Test Automation on a SaaS Platform. 317-325. *IEEE*.

Matos, M., A. Sousa, J. Pereira, R. Oliveira, E. Deliot & P. Murray (2009) CLON: Overlay networks and gossip protocols for cloud environments. *On the Move to Meaningful Internet Systems: OTM 2009*, 549-566.

Mei, L., W. Chan & T. Tse. 2008. A tale of clouds: paradigm comparisons and some thoughts on research issues. 464-469. *Ieee*.

Mell, P. & T. Grance (2009) The NIST definition of cloud computing. National Institute of Standards and Technology, 53.

Mika, P. & G. Tummarello (2008) Web semantics in the clouds. IEEE Intelligent Systems, 82-87.

Napper, J. & P. Bientinesi. 2009. Can cloud computing reach the top500? , 17-20. ACM.

Nelson, M. R. (2009) The cloud, the crowd, and public policy. Issues in Science and Technology, 25, 71-76.

Okuhara, M., T. Shiozaki & T. Suzuki (2010) Security Architecture for Cloud Computing. FUJITSU Sci. Tech. J, 46, 397-402.

Ostermann, S., A. Iosup, N. Yigitbasi, R. Prodan, T. Fahringer & D. Epema (2010) A performance analysis of EC2 cloud computing services for scientific computing. Cloud Computing, 115-131.

Pallickara, S., J. Ekanayake & G. Fox. 2008. An Overview of the Granules Runtime for Cloud Computing. 412-413. IEEE.

Paquette, S., P. T. Jaeger & S. C. Wilson (2010) Identifying the security risks associated with governmental use of cloud computing. Government Information Quarterly, 27, 245-253.

Pastaki Rad, M., A. Sajedi Badashian, G. Meydanipour, M. Ashurzad Delcheg, M. Alipour & H. Afzali (2009) A Survey of Cloud Platforms and Their Future. Computational Science and Its Applications–ICCSA 2009, 788-796.

Patel, P., A. Ranabahu & A. Sheth. 2009. Service Level Agreement in Cloud Computing.

Pearson, S. 2009. Taking account of privacy when designing cloud computing services. 44-52. IEEE Computer Society.

Robison, W. J. (2010) Free at What Cost?: Cloud Computing Privacy Under the Stored Communications Act. Georgetown Law Journal, 98.

Singh, A., M. Korupolu & D. Mohapatra. 2008. Server-storage virtualization: integration and load balancing in data centers. 53. IEEE Press.

Sotto, L. J., B. C. Treacy & M. L. McLellan (2010) Privacy and data security risks in cloud computing. Electronic Commerce & Law Report, 15, 186.

Srinivasamurthy, S. & D. Liu (2010) Survey on Cloud Computing Security.

Sriram, I. & A. Khajeh-Hosseini (2010) Research agenda in cloud technologies. Arxiv preprint arXiv:1001.3259.

Stein, L. D. (2010) The case for cloud computing in genome informatics. Genome Biol, 11, 207.

Streitberger, W. & T. Eymann (2009) A simulation of an economic, self-organising resource allocation approach for application layer networks. Computer Networks, 53, 1760-1770.

Takabi, H., J. B. D. Joshi & G. Ahn (2010) Security and Privacy Challenges in Cloud Computing Environments. Security & Privacy, IEEE, 8, 24-31.

Timmermans, J., V. Ikonen, B. C. Stahl & E. Bozdog. 2010. The Ethics of Cloud Computing: A Conceptual Review. 614-620. IEEE.

Tsai, C. L., U. C. Lin, A. Chang & C. J. Chen. 2010. Information security issue of enterprises adopting the application of cloud computing. 645-649. IEEE.

Vaquero, L. M., L. Rodero-Merino, J. Caceres & M. Lindner (2008) A break in the clouds: towards a cloud definition. ACM SIGCOMM Computer Communication Review, 39, 50-55.

Varia, J. (2008) Cloud architectures. White Paper of Amazon, jineshvaria. s3.amazonaws.com/public/cloudarchitectures-varia.pdf.

Voas, J. & J. Zhang (2009) Cloud Computing: New Wine or Just a New Bottle? IT professional, 11, 15-17.

Vouk, M. A. (2008) Cloud computing—Issues, research and implementations. Journal of Computing and Information Technology, 16, 235-246.

Walker, E. (2009) The real cost of a CPU hour. Computer, 42, 35-41.

Wang, C., Q. Wang, K. Ren & W. Lou. 2010. Privacy-preserving public auditing for data storage security in cloud computing. 1-9. IEEE.

Wang, L., J. Tao, M. Kunze, A. C. Castellanos, D. Kramer & W. Karl. 2008. Scientific cloud computing: Early definition and experience. 825-830. IEEE.

Wang, Q., C. Wang, J. Li, K. Ren & W. Lou (2009) Enabling public verifiability and data dynamics for storage security in cloud computing. Computer Security—ESORICS 2009, 355-370.

Weinhardt, C., A. Anandasivam, B. Blau & J. Stöber (2009) Business models in the service world. IT professional, 11, 28–33.

Wilson, M. S. 2009. Constructing and managing appliances for cloud deployments from repositories of reusable components. 16-16. USENIX Association.

Woo, D. H. & H. H. S. Lee (2009) PROPHET: goal-oriented provisioning for highly tunable multicore processors in cloud computing. ACM SIGOPS Operating Systems Review, 43, 102-103.

Yang, H. & M. Tate (2009) Where are we at with cloud computing?: a descriptive literature review.

Yeo, C. S., S. Venugopal, X. Chu & R. Buyya (2010) Autonomic metered pricing for a utility computing service. Future Generation Computer Systems, 26, 1368-1380.

Yigitbasi, N., A. Iosup, D. Epema & S. Ostermann. 2009. C-meter: A framework for performance analysis of computing clouds. 472-477. IEEE Computer Society.

Youseff, L., M. Butrico & D. Da Silva. 2008. Toward a unified ontology of cloud computing. 1-10. IEEE.

Yu, S., C. Wang, K. Ren & W. Lou. 2010. Achieving secure, scalable, and fine-grained data access control in cloud computing. 1-9. IEEE.

Zhang, Q., L. Cheng & R. Boutaba (2010) Cloud computing: state-of-the-art and research challenges. Journal of Internet Services and Applications, 1, 7-18.

Zissis, D. & D. Lekkas (2010) Addressing cloud computing security issues. Future Generation Computer Systems.

Cloud Computing Reference Architecture

Raffaele Stifani
IBM Italia S.p.A.
Cirocnvallazione Idroscalo, 20090 Segrate (Milano)

Today Chief Information Officers and business leaders from industry and governments around the globe are challenged to stay competitive and meet business objectives. Aligning the enterprise IT model with the business, controlling costs and keeping pace with the rapid rate of innovation are critical. As a consequence, interest in cloud computing is very strong. Early adopters of cloud computing have already realized significant benefits: saving in IT labour costs, reducing server and application provisioning cycle times, improving quality, lowering end-user IT support costs. IBM has recently submitted a subset of the IBM Cloud Computing Reference Architecture to the Cloud Architecture Project of The Open Group. The IBM Cloud Computing Reference Architecture is based on real world input from many cloud implementations and provides guidelines for creating a cloud environment.

1. Introduction

Today's CIOs and business leaders from industry and governments around the globe are challenged to stay competitive and meet business objectives. Aligning the enterprise IT model with the business, controlling costs and keeping pace with the rapid rate of innovation are critical.

As a consequence, interest in cloud computing is very strong. In one IBM study, between 45 and 55 percent of IT leaders surveyed reported that they are implementing or planning to implement a private cloud project (see reference 1) and 91 percent of the 2,000 IT professionals responding to the most recent IBM Tech Trends Survey anticipate that in five years cloud computing will overtake traditional on-premises computing as the primary way organizations acquire IT (see reference 2).

Early adopters of cloud computing, including IBM's own IT organization and IBM customers, have already realized significant benefits including: reducing IT labour cost (in configuration, operations, management and monitoring); reducing server and application provisioning cycle times from weeks to minutes; improving quality (eliminating software defects); lowering end-user IT support costs.

With the growing adoption of cloud computing there has been a concurrent growth in the number and variety of vendors in the marketplace with cloud offerings. As this marketplace becomes more crowded, it is more challenging for CIOs to select the right cloud vendor to meet the organization's needs. And as the applications deployed on private or public clouds move up the value chain into the enterprise computing realm, the selection criteria are shifting from an emphasis on price to considerations such as security, reliability, scalability, control and tooling, and a trusted vendor relationship.

This paper describes the proven cloud computing reference architecture that IBM employs in building private clouds for clients, private clouds that house IBM internal applications and the IBM Cloud that supports our public cloud service offerings. The same reference architecture is reflected in the design of IBM cloud appliances (bundled hardware and software for cloud implementation) and IBM cloud service management software products.

2. Best practices and continuous improvement

Developed by the IBM cloud computing architecture board (comprising technology leaders from IBM Research and IBM's software, systems and services organizations), the Cloud Computing Reference Architecture is derived from an extensive client interaction combined with IBM's extensive capabilities and experience in building industrial strength IT systems and SOA solutions.

Moreover, a process of continuous improvement helps to ensure that both real world experiences and technology advancements from IBM Research are integrated into IBM's cloud reference architecture and management components.

This reference architecture provides specifications not only for the physical components of a cloud implementation (network, compute, storage, virtualization), but as importantly for the software components required to run operational and business management processes. It also defines governance policies tailored for the environment or organization. IBM's Cloud Computing Reference Architecture provides detailed documentation of all components, including how to realize each component for cloud-scale efficiencies.

It is a comprehensive, tightly integrated compilation of documents that describe in detail how to build a management platform for delivering any type of service and includes use cases, non-functional requirements, components, operations, security, performance and scalability, resiliency, consumability considerations, cloud service creation guidance and much more.

3. Principles

An architectural principle is an overarching guideline or paradigm driving decisions across the entire architecture development process. IBM established three principles that guide IBM cloud architects in defining the detailed components of each module:

- o Efficiency principle - Design for cloud-scale efficiencies and time-to-deliver/time-to-change metrics when realizing cloud characteristics such as elasticity, self-service access and flexible sourcing. Objective is to drive down costs per service instance hour and time to response by orders of magnitude.
- o Lightweight principle - Support lean and lightweight service management policies, processes and technologies with an eliminate-standardize-optimize evolutionary approach. Objective is a radical exploitation of standardization in cloud environments to reduce management costs.
- o Economies-of-scale principle - Identify and leverage commonality in cloud service design. Objective is to optimize sharing of management components and infrastructure across cloud services to reduce capital expense, operating expense and time to market.

4. IBM Cloud Computing Reference Architecture

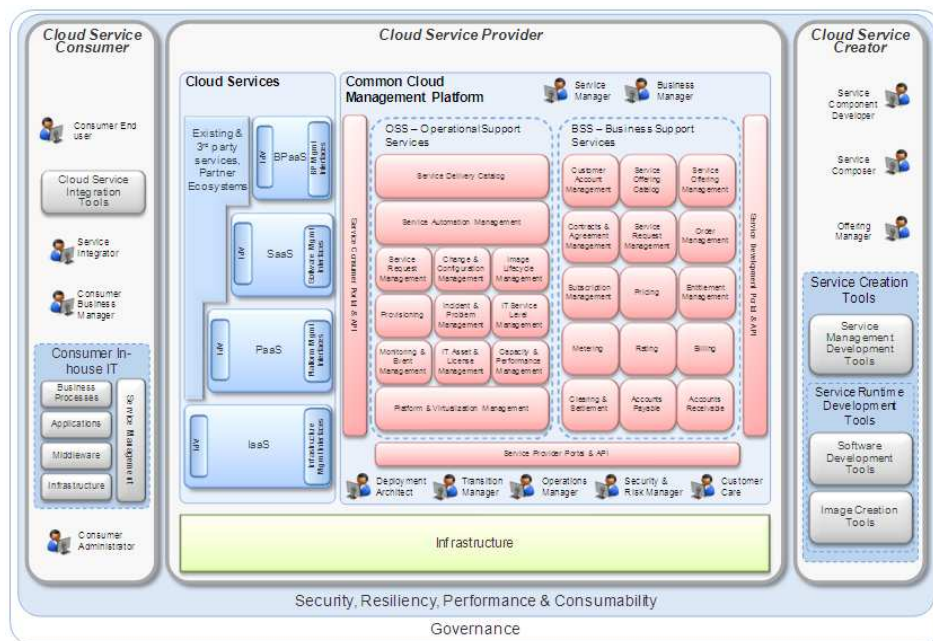


Figure 1 - IBM Cloud Computing Reference Architecture Overview

The essential characteristics for Cloud Computing are on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured Service.

The IBM Cloud Computing Reference Architecture (CCRA) defines the fundamental architectural elements constituting a cloud computing environment. The Cloud Computing Reference Architecture is structured in a modular fashion in a sense that on its highest level of abstraction, the main roles and the corresponding architectural elements are defined allowing to drill down for each of these elements as needed.

5. Main roles

The IBM Cloud Computing Reference Architecture defines three main roles: Cloud Service Consumer, Cloud Service Provider and Cloud Service Creator. Each role can be fulfilled by a single person or by a group of people or by an organization. These roles intend to capture the common set of roles typically encountered in any cloud computing environment. Therefore it is important to note that depending on a particular Cloud Computing scenario or specific cloud implementation, there may be project-specific sub-roles defined.

A **Cloud Service Consumer** is an organization, a human being or an IT system that consumes service instances delivered by a particular cloud service. The service consumer may be billed for all (or a subset of) its interactions with cloud services and the provisioned service instances. A service consumer can also be viewed as a kind of super-role representing the party consuming services. The cloud service consumer browses the service offering catalog and triggers service instantiation and management from there.

The **Cloud Service Provider** has the responsibility of providing cloud services to Cloud Service Consumers. A cloud service provider is defined by the ownership of a Common Cloud Management Platform (CCMP). This ownership can either be realized by truly running a CCMP by himself or consuming one as a service. People acting in the role of a Cloud Service Provider and a Cloud Service Consumer at the same time would be a partner of another cloud service provider reselling cloud services or consuming cloud services and adding value add functionality on top, which would in turn be provided as a cloud service.

The **Cloud Service Creator** is responsible for creating a cloud service, which can be run by a Cloud Service Provider and by that exposed to Cloud Service Consumers. Typically, Cloud Service Creators build their cloud services by leveraging functionality which is exposed by a Cloud Service Provider. Management functionality which is commonly needed by Cloud Service Creators is defined by the Common Cloud Management Platform architecture. A Cloud Service Creator designs, implements and maintains runtime and management artifacts specific to a cloud service. Just like the Cloud Service Consumer and the Cloud Service Provider, the Cloud Service Creator can be an organization or a human being. It is also typical that operations staff responsible for operating a cloud service is closely integrated with the development organization developing the service.

6. Cloud Service Consumer

From the perspective of a Cloud Service Consumer, it is important to be able to integrate cloud services with their on-premise IT. The functionalities of the **Cloud Service Integration Tools** are specifically relevant in the context of hybrid clouds, where seamless integrated management, usage and interoperability of cloud services in integration with on-premise IT is critical.

Besides IT capabilities consumed as cloud services, consumers of such IT may continue to have **in-house IT**, which can be managed in a traditional non-cloud fashion. In case functionalities of the existing in-house IT should be integrated with cloud services consumed from a cloud service provider, the aforementioned cloud service integration tools are required. Consumer in-house IT exists across all layers of the technology stack (infrastructure, middleware, applications, business processes, service management): therefore integration with cloud services can take place on each of these layers.

7. Cloud Service Provider

In the Cloud Service Provider the most important elements are: Cloud Services, Common Cloud Management Platform and the Infrastructure.

3.3 Cloud Service

Cloud Services represent any type of (IT) capability which is provided by the Cloud Service Provider to Cloud Service Consumers, implementing all cloud characteristics (on demand self-service, broad network access, resource pooling, rapid elasticity, measured services). There are four categories of Cloud Services: Infrastructure, Platform, Software or Business Process Services. In contrast to traditional (IT) services, cloud services have attributes associated with cloud computing, such as a pay-per-use model, self-service usage, flexible scaling & shared of the underlying IT resources.

The management functions defined as part of the Common Cloud Management Platform architecture are responsible for delivering instances of Cloud Services of any category to Cloud Service Consumers, the ongoing management of all Cloud service instances from a provider perspective and allowing Cloud Service Consumers to manage their Cloud Service instances in a self-service fashion.

For all cloud services there is software required implementing cloud service specifics: for IaaS there are typically hypervisors installed on the managed hardware infrastructure, for PaaS there is a multi-tenancy enabled middleware platform, for SaaS a multi-tenancy end-user application and for BPaaS a multi-tenancy enabled business process engine.

Depending on the nature of the respective cloud service, there are cloud service specific management interfaces exposed, which are typically used by the management services as defined in the Common Cloud Management Platform architecture. Each cloud service also typically exposes APIs towards the cloud service consumer for programmatic use.

3.3 Cloud Service Models

There are four **Cloud Service Models** within the context of the IBM Cloud Computing Reference Architecture: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS) and Business-Process-as-a-Service (BPaaS). IaaS, PaaS and SaaS are defined according to the National Institute of Standards and Technology (NIST - U.S. Department of Commerce); BPaaS is an IBM-specific definition.

Infrastructure as a Service (IaaS) - The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).

Platform as a Service (PaaS) - The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

Software as a Service (SaaS) - The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

Business Process as a Service" (BPaaS) - Business process services are any business process (horizontal or vertical) delivered through the cloud service model (multitenant, self-service provisioning, elastic scaling and usage metering or pricing) via the Internet with access via web-centric interfaces and exploiting web-oriented cloud architecture. The BPaaS provider is responsible for the related business function(s).

3.3 Infrastructure

Infrastructure represents all infrastructure elements needed on the cloud service provider side, which are needed to provide cloud services. This includes facilities, server, storage, and network resources, how these resources are wired up, placed within a data center, etc. The infrastructure element is purely scoped to the hardware infrastructure, therefore it does not include software running on top of it such as hypervisors (hypervisors are generally specific to an IaaS offering and therefore belong to that particular cloud service). Consequently it also does not include any virtualization management software.

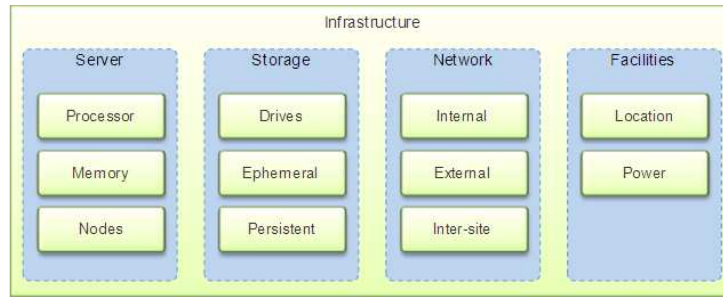


Figure 2 - Infrastructure of Cloud Computing Reference Architecture

3.3 Common Cloud Management Platform

The **Common Cloud Management Platform (CCMP)** exposes a set of Operational Support Services (OSS) and Business Support Services (BSS) focused on the management issues. These OSS and BSS functions must be exploited by Cloud Services to run within the context of the respective cloud service provider (and the corresponding CCMP installation).

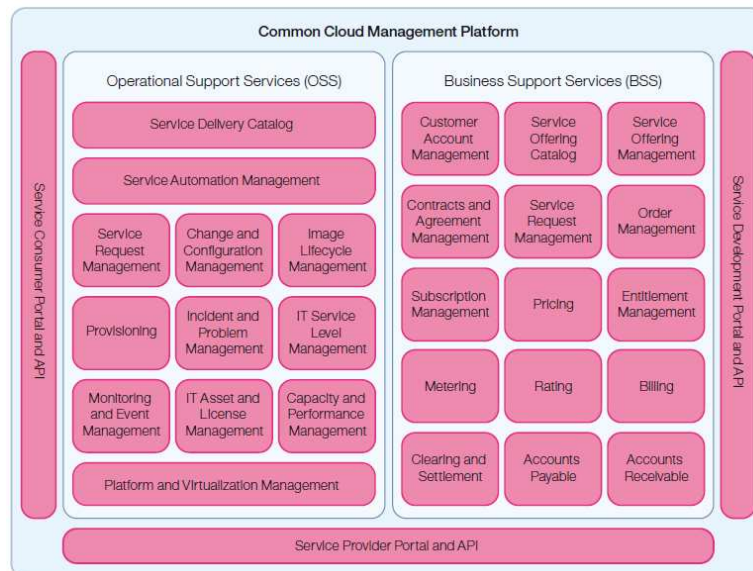


Figure 3 - Common Cloud Management Platform

Besides OSS and BSS, the CCMP also includes User Interfaces serving the three main roles defined in the CC RA:

1. a Service Consumer Portal to be used by Cloud Service Consumers for self-service delivery and management (the actual cloud service instances are used via a cloud service specific User Interface),
2. a Service Provider Portal serving Cloud Service Provider internal users and administrators for daily operations
3. a Service Development Portal used by Cloud Service Creators.

The CCMP functionalities are accessible via APIs exposed by the CCMP internal components.

The CCMP exposes a set of services which can be used within the context of a specific cloud service. The management services exposed by the CCMP to cloud service creators are not to be confused with the cloud services developed by cloud service creators. Cloud Service Creators are strongly encouraged to use the management services provided by the CCMP in order to enable the economies-of-scale required for achieving the extremely high degrees of efficiency associated with any cloud computing environment.

The CCMP is defined as a general purpose cloud management platform to support the management of any category of cloud service across IaaS, PaaS, SaaS and BPaaS. The CCMP is split into two main elements: the Operational Support Services (OSS) and Business Support Services (BSS).

3.4.1 Business Support Services

Business Support Services (BSS) represents the set of business-related services, exposed by the Common Cloud Management Platform, which are needed by Cloud Service Creators to implement a cloud service.

The motivations for using these commonly defined Business Support Services functions are the same as the ones for reusing all other services defined in the context of Common Cloud Management Platform:

- using standard component services will reduce the unique development required, reducing development time and costs
- using standard components will result in a cloud service more likely to be compatible with other cloud services, providing the client with even more economies of scale and efficiencies
- standard components and user interfaces will reduce the complexity, simplify operations and make the cloud solutions easier to use and more consumable.

3.4.1 Operational Support Services

Operational Support Services (OSS) represents the set of operational management and technical-related services exposed by the Common Cloud Management Platform (CCMP), which are needed by Cloud Service Creators to implement a cloud service.

Many management domains shown in the Operational Support Services (OSS) can also be encountered in traditionally managed data centers

(monitoring and event management, provisioning, incident and problem management, etc.) while other components are new and pretty specific to the degrees of automation and efficiency associated with clouds (service automation, image lifecycle management, etc.). Particularly for the traditional management domains it is important to note that conceptually they are the same in the cloud world and in the traditional world, whereas in a cloud world these domains are generally implemented in radically different ways taking advantage of the high degrees of homogeneity in a cloud.

8. Cloud Service Creator

Service Development Tools are used by the Cloud Service Creator to develop new cloud services. This includes the development of runtime artifacts and management-related aspects (monitoring, metering, provisioning, etc.). Runtime artifacts refer to any capability needed for running what is delivered as-a-service by a cloud deployment (J2EE enterprise applications, database schemas, analytics, golden master virtual machine images, etc.).

In the context of a particular infrastructure-as-a-service or platform-as-a-service offering, there may also be tooling to develop artifacts which are specific to the particular cloud service.

9. Security, resilience, performance and consumability

Security, Resiliency, Performance and Consumability are cross-cutting aspects spanning the Common Cloud Management Platform (CCMP), the hardware infrastructure and Cloud Services. These non-functional aspects must be viewed from an end-to-end perspective including the structure of Common Cloud Management Platform (CCMP) by itself, the way the hardware infrastructure is set up (in terms of isolation, network zoning setup, data center setup for disaster recovery, etc.) and how the cloud services are implemented.



Figure 4 – Security, Resiliency, Performance & Consumability

10. Conclusion

Today's CIOs and business leaders are challenged to stay competitive and meet business objectives: aligning the enterprise IT model with the business, controlling costs and keeping pace with the rapid rate of innovation are critical.

For CIOs it is challenging to select the right cloud computing reference architecture to meet the organization's needs. Examining the cloud solution provider's reference architecture should be a standard element of a CIO's cloud vendor evaluation strategy.

Cloud computing must be enabled with effective security, resiliency, service management, governance, business planning and lifecycle management. These are the components of an effective and comprehensive cloud architecture that will enable the enterprise to control the environment more effectively, optimize productivity, reduce associated labour costs and ensure a resilient, safe environment for business users. IBM Cloud Computing Reference Architecture:

- o is based on open standards
- o delivers robust security, governance, compliance and privacy capabilities
- o combines powerful automation and management services with rich business management functions for fully integrated management of cloud infrastructure and cloud services
- o supports the full spectrum of cloud service models, including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) and Business Process as a Service (BPaaS)
- o enables the flexible scaling and resiliency required for successful cloud economics and return on investment
- o facilitates seamless integration into existing customers' environments
- o is based on industry-leading expertise for building services and service oriented architectures.

Moreover, a process of continuous improvement helps to ensure that both real-world experiences and technology advancements from IBM Research are integrated into IBM Cloud Computing Reference Architecture.

References

- [1] IBM Market Insights. Cloud Computing Research, July 2009.
- [2] 2010 IBM Tech Trends Survey. October, 2010.
www.ibm.com/press/us/en/pressrelease/32674.wss
- [3] The Open Group SOA Reference Architecture.
http://www.opengroup.org/soa/drafts/refarch.htm#_SOA_Reference_Architecture
- [4] Peter Mell and Tim Grance, The NIST Definition of Cloud Computing, Version 15, October 7, 2009. <http://www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf>
- [5] Raffaele Stifani, Il Cloud Computing: sorge una nuova era, Atti del Congresso Nazionale AICA 2009, AICA
- [6] Raffaele Stifani, 2010, Le potenzialità del Cloud Computing, Atti del Congresso Nazionale AICA 2010, AICA

Conformità di politiche di sicurezza in ambienti industriali

Manuel Cheminod, Marco Cereia
IEIT-CNR, C.so Duca degli Abruzzi 24, I-10129 Torino, Italia
{manuel.cheminod,marco.cereia}@ieiit.cnr.it

Il miglioramento tecnologico, il continuo aumento dei livelli di interconnessione interna ed esterna nei sistemi informatici offrono enormi potenzialità per quello che riguarda flessibilità e facilità di gestione delle reti. Tuttavia, gli stessi avanzamenti tecnologici rendono i sistemi ICT sempre più esposti ai rischi legati agli attacchi informatici. Uno degli strumenti utili per gestione di tali rischi si trova nella definizione di politiche di sicurezza. La progettazione e successiva implementazione di tali politiche presentano però serie difficoltà, soprattutto in ambiti peculiari come quelli industriali. In questo contributo si dimostra come politiche di sicurezza in ambienti industriali possano essere definite ed analizzate tramite uno strumento di analisi automatica al fine di verificare l'effettiva coerenza tra i requisiti di sicurezza del sistema e l'implementazione delle politiche definite. Tale approccio viene esemplificato tramite l'ausilio di un caso di studio.

1. Introduzione

La protezione di sistemi ICT è un argomento la cui importanza è in costante crescita e nel quale vengono continuamente investite ingenti risorse umane e finanziarie. L'incremento del grado di interconnessione all'interno di questi sistemi e tra sistemi diversi (favorito anche dallo sviluppo tecnologico di Internet) ha portato a indubbi benefici in termini di gestione e di possibilità di servizi offerti incrementando, nello stesso tempo, l'esposizione a rischi legati ad attacchi informatici. La protezione dei componenti di tali sistemi assume, quindi, una rilevanza strategica al fine di evitare perdite in termini monetari ma anche in termini umani e ambientali.

La complessità dei sistemi ICT e dei suoi componenti richiede un approccio alla sicurezza su più livelli in grado di coinvolgere tutte le fasi di sviluppo di un progetto: dalla pianificazione/progettazione iniziale, all'implementazione, al processo di continua gestione e mantenimento [Avizienis et al, 2004].

Nei sistemi industriali la complessità aumenta sotto certi aspetti a causa dell'integrazione sempre più spinta tra sistemi di natura diversa (gestione, produzione e altri) che migliora l'efficienza della gestione ma implica requisiti di sicurezza più stringenti. Architetture classiche [McLean et al, 1983] devono venire rafforzate o sostituite per affrontare i nuovi problemi di queste infrastrutture [Bessani et al, 2008][Miller, 2005].

Da un punto di vista tecnico, meccanismi come quelli legati a firewall, liste di controllo di accesso (ACL) e gestione dei privilegi, sono abbastanza flessibili e potenti per essere usati nella protezione di elementi hardware e software. Il numero di elementi coinvolti, tuttavia, rende la progettazione e implementazione coerente ed efficace di questi meccanismi un problema complesso. Inoltre le contromisure classiche usate in sistemi ICT non sono direttamente applicabili in tali ambienti in quanto esistono sia affinità tra i due tipi di sistemi, sia differenze considerevoli. L'obiettivo principale in un sistema ICT, per esempio, è generalmente legato alla segretezza delle informazioni maneggiate mentre in sistemi industriali è la disponibilità delle risorse ad essere la priorità.

Nel progetto di sicurezza di un sistema, il primo passo richiede la definizione di determinati requisiti di sicurezza a seguito della quale è possibile proseguire nella progettazione e gestione del sistema. Per affrontare la complessità di quest'ultimo si può ricorrere ad un approccio diviso su due livelli: alto e basso. Ad alto livello sono definite delle **politiche** in maniera indipendente dall'effettiva implementazione tecnica delle stesse. A basso livello, invece, sono presi in considerazione i dettagli del sistema. L'approccio su due livelli presenta una criticità insita nella verifica di consistenza tra i due livelli divisa in:

- **controllo di conformità delle politiche:** si verifica la coerenza interna nelle regole di alto livello e l'effettiva aderenza delle politiche rispetto ai requisiti di sicurezza definiti;
- **controllo di implementazione delle politiche:** si verifica la coerenza tra le politiche di alto livello e la loro implementazione nel sistema.

Esistono numerosi lavori di ricerca che affrontano il problema della definizione di politiche di sicurezza [Sandhu et al, 1996][Damianou et al, 2001] e anche il controllo di conformità delle politiche [Hu et al, 2007][Hughes e Bultan, 2008]. D'altra parte, la ricchezza di scelte tecniche possibili nella progettazione di un sistema [Samarati e De Capitani di Vimercati, 2001] rende difficilmente automatizzabile la verifica della traduzione dal livello alto al livello basso. In questo contributo viene affrontato il problema della definizione e verifica di politiche di sicurezza in un ambiente industriale tramite l'utilizzo di uno strumento di analisi automatica. I due aspetti, controllo di conformità e di implementazione, vengono affrontati con lo stesso strumento su uno stesso caso di studio.

2.Politiche di sicurezza in ambienti industriali.

La definizione di regolamentazioni e politiche che controllano le interazioni tra i diversi elementi in un sistema è un'importante strumento per la gestione della sicurezza informatica. Uno dei primi problemi da affrontare nella definizione di politiche risiede nella necessità di usare un linguaggio formale e sufficientemente generale. Sono disponibili diversi modelli utili a questo proposito, per esempio esistono diversi approcci per la gestione delle politiche di accesso [Samarati e De Capitani di Vimercati, 2001]. L'approccio Role Based Access Control (RBAC) è il più comunemente usato [Ferraiolo et al, 2001] e la definizione formale è supportata da alcune proposte come il linguaggio Ponder e il linguaggio XACML.

Ruolo	Operazione	Risorsa
Amministratore	<i>qualsiasi</i>	<i>qualsiasi</i>
Impiegato	login, logout	autenticazione
Consulente	consulta	historian
Dirigente	<i>qualsiasi</i>	autenticazione
Operaio	stop	PLC
Tecnico	diagnostica	PLC
Supervisore	<i>qualsiasi</i>	PLC
Supervisore	<i>qualsiasi</i>	server di controllo
Macchina	<i>qualsiasi</i>	historian

Tabella 1 Privilegi dei ruoli

2.1 Descrizione del modello

Il caso di studio usato in questo contributo viene descritto seguendo un approccio top-down ovvero da un alto livello di astrazione fino ai dettagli implementativi. Questo riflette in parte il progetto di un sistema ICT.

Il più alto livello di astrazione permette di individuare le principali classi di agenti e risorse nel sistema. In una generica rete industriale possiamo identificare tre “zone”: la zona esterna, la zona aziendale, la zona di campo. La zona esterna include tutto ciò che non è controllabile dal sistema ed in generale si identifica questa zona con Internet. La zona aziendale include le postazioni di lavoro di impiegati e dirigenti ed include anche i servizi informatici legati alla gestione aziendale ed i servizi comuni (come il servizio di posta elettronica). È sempre più comune trovare in questa zona anche servizi “pubblici” come server web. La zona di campo, invece, include i controllori logici programmabili (PLC), i dispositivi di campo, e così via. Seguendo alcune *best practices* è già possibile individuare una prima politica di livello generale:

Politica 1: deve esistere una netta distinzione tra i ruoli operanti nelle diverse zone. Non deve essere possibile, per esempio, che un agente operi con lo stesso ruolo in una zona e in un'altra.

I prossimi elementi introdotti sono le **risorse** e le relative **operazioni** che possono essere eseguite. In questo contributo viene considerato uno scenario tipico basato sulle raccomandazioni presenti in [Stouffer et al, 2008] usando, per la definizione delle politiche, la terminologia introdotta in [Sandhu et al, 1996]. In Tab.1 vengono riportate le risorse e relative operazioni per il caso di studio. La risorsa “autenticazione”, per esempio, rappresenta il sotto-sistema di autenticazione dell'azienda, a prescindere dall'implementazione tecnica effettiva. Le operazioni relative sono “login” e “logout” (che permettono ad un utente di ottenere e rilasciare una particolare identità), e “assegna” e “revoca” (operazioni per la gestione dinamica dei ruoli nel sistema). La risorsa “historian”, invece, rappresenta il server dello storico (*historian*) che generalmente tiene traccia delle attività della rete di campo presentandole poi alla parte gestionale per le decisioni strategiche. Le operazioni in questo caso

sono tipicamente di consultazione e scrittura. Infine, la risorsa "PLC" rappresenta i vari PLC presenti nella rete e le operazioni relative sono quelle comuni di gestione e mantenimento.

In questo contributo, l'approccio RBAC classico (in cui gli utenti sono correlati a ruoli e ruoli sono correlati a privilegi) è arricchito dall'attributo **livello di fiducia** per ogni utente definendo tre valori possibili: alto, medio, basso. Questa caratteristica viene usata nella prossima politica:

Politica 2: gli utenti che accedono alle risorse "historian" e "server di controllo" nella zona di campo devono avere un livello di fiducia "alto".

A questo livello di dettaglio appartiene anche la definizione dei **ruoli** nel sistema. Questi definiscono le classi di utenti e sono strettamente collegati alla zona in cui andranno ad operare. Mantenendo una certa generalità è possibile elencare una serie di ruoli principali: "ospite", "impiegato", "consulente", "dirigente", "operaio", "tecnico", "supervisore". Vale la pena specificare che "ospite" è un qualsiasi agente non autenticato cui si permette di accedere al sistema per una qualche ragione. Nel modello vengono aggiunti due ruoli peculiari: "remoto" e "macchina". Il ruolo "remoto" rappresenta un agente che non si trova fisicamente all'interno del sistema e può rappresentare, per esempio, un tecnico che fornisce assistenza remota oppure un impiegato che lavora da una sede esterna dell'azienda. Il ruolo "macchina" viene invece incluso per rappresentare le interazioni dei vari PLC all'interno della rete. Da un certo punto di vista, infatti, è possibile modellizzare un PLC come un agente che agisce nel sistema. In Tab.2 sono mostrati i privilegi assegnati ai vari ruoli: sono elencate le operazioni che i ruoli possono eseguire sulle varie risorse.

A questo punto, gli elementi introdotti permettono di definire alcune **proprietà** che si desidera siano soddisfatte nel sistema.

- **Pr1:** nessun "ospite" può interagire con i PLC della rete di campo;
- **Pr2:** nessuno può eseguire l'operazione "diagnostica" su un PLC se questo è operativo ("online");
- **Pr3:** nessun impiegato può accedere al server di controllo.

Queste proprietà definiscono cosa non deve mai accadere e quindi devono essere verificate in ogni stato possibile in cui può trovarsi il sistema.

La proprietà Pr2 è importante poiché programmi di diagnostica possono causare danni se eseguiti mentre l'apparecchio è in linea. Solitamente questo tipo di operazioni vengono svolte ponendo gli apparecchi in stato "offline".

La politica 2 permette di definire due ulteriori proprietà desiderate:

- **Pr4:** utenti con livello di fiducia inferiore a "alto" non possono accedere alla risorsa "historian";
- **Pr5:** utenti con livello di fiducia inferiore a "alto" non possono accedere alla risorsa "server di controllo".

3.Uno strumento automatico di analisi

L'analisi proposta in questo contributo è un'analisi esaustiva del modello al fine di verificare un determinato set di proprietà. L'analisi viene supportata da uno strumento automatico sviluppato seguendo un approccio di programmazione logica ed implementato nel linguaggio Prolog. L'architettura dello strumento è schematicamente mostrata in Fig.1. Dall'alto verso il basso

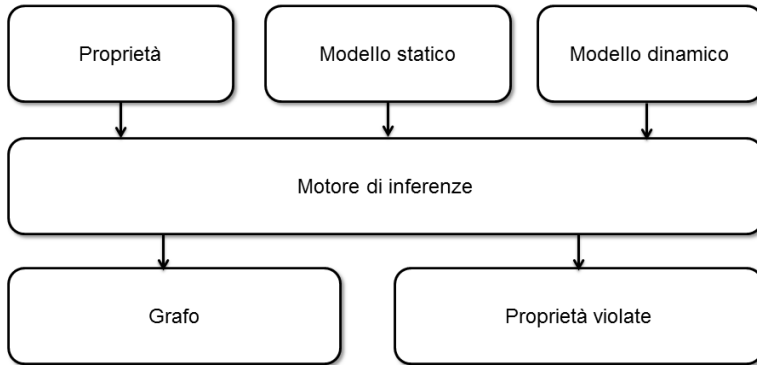


Figura 1 Architettura

sono distinguibili le tre fonti di input (**modello statico**, **modello dinamico** e **proprietà**), il blocco principale e i due blocchi dei risultati (**grafo** e set di proprietà **violate**). Il modello statico rappresenta lo stato iniziale del sistema e include la definizione delle variabili in gioco, i loro valori iniziali e i vincoli cui sono soggette (vincoli sul dominio delle variabili). Il modello statico, per esempio, è composto dall'insieme dei ruoli individuati, dalle risorse (con le loro caratteristiche) e dai vari utenti e privilegi relativi. Il modello dinamico, invece, descrive le possibili evoluzioni (transizioni) del sistema in termini di cambiamento di stato. Questo include cambiamenti nei valori delle variabili ma anche cambiamenti nelle relazioni tra i vari elementi come, per esempio, l'attribuzione dinamica di ruoli agli utenti.

Le proprietà vengono formalizzate tramite regole in Prolog, ovvero:

$L_0 :- L_1, \dots, L_n$

dove $L_0, L_1, \dots, L_n$ sono predicati, L_0 è la testa della regola ed il resto rappresenta il corpo della regola. La regola è valida (quindi la proprietà è soddisfatta) se sono veri tutti i $L_1, \dots, L_n$ predicati.

I vari input sono quindi combinati e analizzati dal motore di inferenze al fine di verificare le proprietà nello stato iniziale e in tutti quegli stati in cui può trovarsi il sistema applicando il modello dinamico. L'evoluzione degli stati viene registrata e organizzata in un digrafo i cui nodi rappresentano uno stato del sistema e i cui archi rappresentano una particolare transizione. Verificando le proprietà per ogni nodo è possibile individuare l'insieme di proprietà violate ed analizzando il grafo è possibile risalire alle cause di una particolare violazione. Seguendo le transizioni all'indietro fino allo stato iniziale è infatti possibile ottenere tutte le tracce (successione di stati e transizioni) di una violazione.

3.1 Aspetti tecnici

La complessità dell'analisi dipende direttamente dalla grandezza del modello e dai suoi elementi. In particolare la parte dinamica influenza il numero di stati raggiungibili e quindi la cardinalità del grafo generato. Per garantire la finitezza dell'analisi in termini di numero di stati viene imposto un vincolo sui valori delle variabili: i valori devono appartenere ad un insieme finito di valori, ovvero il dominio delle variabili è discreto e finito. La finitezza dell'analisi in termini

temporali, invece, è garantita attraverso l'introduzione di un controllo di duplicazione degli stati: ogni nuovo stato generato nell'analisi viene confrontato con tutti gli stati precedenti, in caso di due stati equivalenti questi vengono collassati in un unico nodo nel grafo. In questa maniera si evitano generazioni infinite degli stessi stati senza perdere la completezza nelle tracce generate.

4. Esempio di controllo di conformità

Viene ora presentato un esempio di analisi ad alto livello, ancora indipendente dall'effettiva implementazione delle politiche di sicurezza. Si aggiunge un ulteriore ruolo: "sviluppatore". Si suppone che, in passato, uno sviluppatore esterno abbia seguito lo sviluppo di una macchina (PLC) nella rete industriale e che tale attività sia finita. Durante il periodo in questione, lo sviluppatore ha avuto accesso alle operazioni di diagnostica locale sui PLC. Questo privilegio viene modellato andando ad arricchire il contenuto di Tab.2. Il dominio della variabile "ruolo" viene quindi arricchito dal valore "sviluppatore".

A questo punto viene dichiarata l'esistenza di un utente di nome bob asserendo (in Prolog) il fatto: *user(bob)*.

L'analisi del sistema con la configurazione definita fino a questo punto mostra una violazione della proprietà Pr1. La traccia della violazione permette di distinguere i seguenti passi (semplificati per chiarezza):

- *ruoli_utente(bob,[])*.
- *ruoli_utente(bob,[ospite, sviluppatore])*.
- violazione della proprietà **Pr1**.

Il primo passo mostra che nello stato iniziale l'utente bob non appartiene a nessun ruolo. Nel secondo passo all'utente bob risultano associati due ruoli ("ospite", "sviluppatore"). Questo porta evidentemente alla violazione della proprietà Pr1 in quanto un utente "ospite" risulta in grado di accedere alla risorsa PLC. È importante a questo punto chiarire il significato del predicato "ruoli_utente". La forma corretta è: *ruoli_utente(Utente, Lista)*. Questo predicato asserisce il fatto che l'utente "Utente" (in Prolog la lettera maiuscola identifica una variabile) sia associato ai ruoli presenti nella lista "Lista". L'associazione tra un utente ed uno o più ruoli è mutevole nel tempo e viene quindi descritta nel modello dinamico. Non sono state specificati vincoli per quello che riguarda i ruoli dell'utente bob e, per questo motivo, lo strumento di analisi ha effettivamente provato ogni combinazione utente/ruoli possibile evidenziandone una che ha portato ad una violazione. Questa situazione può effettivamente verificarsi, si pensi ad esempio ad un ex-sviluppatore che viene invitato (come "ospite") nell'azienda ottenendo accesso fisico ad una vecchia macchina su cui aveva lavorato. Se non viene rimossa esplicitamente l'associazione del ruolo "sviluppatore" con l'utente "bob", risulta teoricamente possibile per lui agire impropriamente sulla risorsa PLC. Questo tipo di risultato evidenzia la necessità di vincolare l'assegnazione del ruolo "sviluppatore" andando quindi a descrivere un vincolo nel modello dinamico.

È possibile evidenziare un ulteriore problema nella definizione delle politiche. Nel caso in cui un PLC incorra in un problema grave, la prima azione da intraprendere è una fermata di emergenza seguita dall'intervento di personale adeguato ("tecnico" che ha il permesso di eseguire "diagnostica" sul

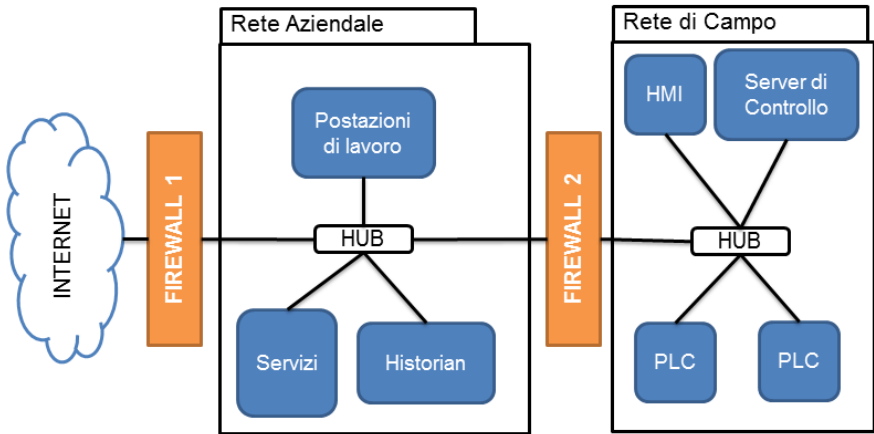


Figura 2 Topologia di rete

PLC). Non sono inusuali situazioni in cui si ricorre ad assistenza tecnica remota e questo può essere rappresentato dalla politica seguente:

Politica 3: in caso di errore grave in un PLC, è possibile assegnare il ruolo "tecnico" ad un utente con ruolo "remoto".

In questo scenario avviene un'assegnazione dinamica di un ruolo eseguita attraverso delega ed in particolare solo chi detiene il ruolo di "amministratore" è abilitato a tale operazione e solo nel caso il cui lo "stato" del PLC sia "errore". Questa regola è formalizzata in Prolog nella seguente:

*delega_ruolo(amministratore, Utente, remoto, tecnico) :-
valore_attributo_risorsa(plc, stato, errore).*

La proprietà di alto livello Pr2 è invece formalizzata come segue:

*proprietà2 :-
valore_attributo_risorsa(plc, stato, online) →
non(utente_può(User, diagnostica, plc)).*

Quindi: il predicato proprietà2 è vero (valido/soddisfatto) se il fatto che il valore dell'attributo "stato" per il "plc" è "online" **implica** che **nessun** utente possa accedere alla diagnostica del "plc". Il dominio della variabile "stato" del "plc" è ristretto ai valori: normale,online,offline,errore. Il valore del predicato "utente_può" dipende dal ruolo che possiede l'utente. Essendo l'associazione dei ruoli una caratteristica dinamica, ne consegue che anche "utente_può" dipende dallo stato corrente del sistema e quindi il suo valore (per uno stesso utente) può variare nel tempo.

L'analisi del modello definito fino a questo punto evidenzia la violazione della proprietà **Pr2** riportando la seguente traccia:

- ruoli_utente(bob,[remoto]).
- cambiamento_valore_variabile(plc,stato,errore).
- cambiamento_valore_variabile(plc,stato,offline).
- ruoli_utente(bob,[remoto,tecnico]).
- cambiamento_valore_variabile(plc,stato,normale).
- violazione della proprietà **Pr2**.

Firewall	Origine	Protocollo	Destinazione
1	qualsiasi	http	servizi
	qualsiasi	ssh	qualsiasi
2	historian	sql	server di controllo
	historian	http	server di controllo
	server di controllo	sql	historian
	server di controllo	http	historian
	qualsiasi	ssh	server di controllo

Tabella 3 Regole dei firewall

Questa traccia rappresenta lo scenario in cui: bob inizialmente ha il ruolo “remoto”; si verifica un “errore” nel “plc”; il plc cambia stato e diventa “offline”; viene assegnato dinamicamente il ruolo “tecnico” a bob (già con ruolo “remoto”); il problema nel plc viene risolto e il plc diventa “normale”; bob mantiene il ruolo “tecnico” e può quindi accedere alla diagnostica del plc violando la proprietà Pr2. Questa violazione evidenzia un problema nel modello dinamico del sistema. Manca una regola/politica per cui il ruolo “tecnico” sia effettivamente temporaneo e venga rimosso una volta risolto un problema su un plc. Questa politica è descritta come:

Politica 3’: quando lo stato del plc ritorna normale, il ruolo temporaneo di “tecnico” deve essere rimosso.

Questa politica viene formalizzata andando a specificare restrizioni nel modello dinamico del predicato “ruoli_utente”.

5. Esempio di controllo di implementazione

L'esempio di analisi mostrato finora evidenzia l'utilità dell'analisi automatica effettuata su politiche di alto livello, senza includere i dettagli tecnici di una eventuale implementazione. Tuttavia questo tipo di analisi non basta nei sistemi moderni. La complessità e dimensione delle reti moderne impone un controllo di coerenza tra le politiche di alto livello e l'effettiva implementazione. Inoltre è d'obbligo controllare il mantenimento di questa coerenza a ogni minima modifica della configurazione della rete (per esempio anche solo una regola di un firewall che viene cambiata).

In questa sezione viene quindi mostrata una possibile implementazione delle politiche di alto livello e ne viene verificata la coerenza. Si suppone di implementare le politiche in un sistema già esistente in cui si vogliano limitare al minimo le modifiche e le aggiunte di componenti. La rete considerata è mostrata in Fig.2 e la sua struttura è stata presa da [Stouffer et al, 2008]. La rete è suddivisa in tre sezioni: Internet, rete Aziendale e rete di Controllo. Il flusso di dati tra le sezioni è regolato da due firewall. Le regole definite per questi due elementi sono mostrate in Tab.3. Vengono elencate solo le comunicazioni permesse, tutto il resto si suppone bloccato. Vale la pena notare come l'unico flusso di dati tra la zona Aziendale e la zona di Controllo avviene tra il server Historian e il server di Controllo.

La proprietà **Pr3** è formalizzata come la seguente regola:

proprietà3 :-

ruolo_utente(Utente,impiegato) →

non(utente_accede(Utente,server_di_controllo)).

Il predicato “utente_accede” tiene conto della topologia fisica e logica (configurazione dei firewall). È possibile introdurre la topologia nel modello dinamico descrivendo eventuali algoritmi di routing.

L'analisi del sistema così definito evidenzia violazioni per le proprietà **Pr3**, **Pr4** e **Pr5**. Una delle tracce riporta i seguenti passi:

- bob ha il ruolo di “impiegato”
- bob manda pacchetti al “server di controllo” marcando “historian” come mittente.
- violazione della proprietà **Pr3**.

La forzatura dell'indirizzo mittente (attacco “ip spoofing”) è possibile perché non c'è alcun elemento nella zona Aziendale in grado di impedirlo. Bob, di fatto, impersona la risorsa “historian” e dialoga con il server di controllo nella zona di Controllo. Una differente traccia evidenzia un problema diverso:

- bob ha il ruolo di “impiegato”,
- l'amministratore assegna a bob il ruolo di “consulente”
- violazione della proprietà **Pr4**,
- bob sfrutta un problema nel software (vulnerabilità) della risorsa “historian” compromettendone l'integrità (tramite un'attacco “sql injection”),
- violazione delle proprietà **Pr3** e **Pr5**.

La Tab.1 mostra che ad un “consulente” è permesso di accedere alla risorsa “historian”, tuttavia avviene una violazione in quanto l'utente bob non possiede un livello di fiducia sufficiente. La compromissione dell'integrità della risorsa “historian” porta ad una situazione più complessa. Si suppone infatti che il “server di controllo” acceda ai dati dell’“historian” ma essendo questi dati compromessi dall'utente bob, allora è possibile asserire, con un certo grado di approssimazione, che bob ha acceduto al server di controllo violando la proprietà **Pr5** e **Pr3** (in quanto il ruolo di bob è “impiegato”). Gli attacchi “ip spoofing” e “sql injection” sono problemi reali ([Stouffer et al, 20008]) e sono inclusi nell'analisi come rischi, non come eventi certi.

Per risolvere i problemi evidenziati vi sono due strade percorribili: modificare il modello dinamico e la topologia della rete. Il modello dinamico, in effetti, può essere arricchito includendo restrizioni sull'assegnazione dinamica dei ruoli. Per esempio può essere richiesto che un utente candidato per l'assegnazione del ruolo “consulente” dimostri di avere un livello di fiducia “alto”. L'attacco “ip spoofing” può essere evitato, invece, seguendo alcune *best practices* inserendo una quarta zona tra quella Aziendale e di Controllo. Questa soluzione implica l'aggiunta di un ulteriore firewall e può essere dispendiosa per l'azienda. In effetti, l'analisi automatica può essere intesa come analisi di rischio evidenziando le problematiche e dando la possibilità di verificare l'efficacia di eventuali cambiamenti prima ancora di applicarli nella pratica.

6.Conclusioni

In questo contributo è stato presentato un approccio all'analisi di politiche di sicurezza in ambienti industriali tale da tenere in considerazione sia la parte di progettazione che la parte d'implementazione. È stato descritto un esempio

partendo dalle politiche di alto livello fino alla loro implementazione nelle regole dei firewall presenti. È stato mostrato come lo strumento automatico possa includere la gestione di “delega” di ruoli e il concetto di accessibilità nella rete (a livello fisico e logico). È stato mostrato, infine, come l'evoluzione delle relazioni tra i ruoli, gli utenti e i privilegi relativi possa essere inclusa in un modello dinamico.

Bibliografia

[Avizienis et al, 2004] A. Avizienis, J. C. Laprie, B. Randell, C. Landwehr, “Basic Concepts and Taxonomy of Dependable and Secure Computing”, IEEE Trans. Dependable and Secure Computing, 1, 1, 2004, 11-33.

[Bessani et al, 2008] A. N. Bessani, P. Sousa, M. Correia, N. F. Neves, P. Verissimo, “The Crucial Way of Critical Infrastructure Protection”, IEEE Security / Privacy, 6, 6, 44-51, 2008.

[Chakraborty and Ray, 2006] S. Chakraborty, I. Ray, “TrustBAC: integrating trust relationships into the RBAC model for access control in open systems”, in 11th ACM Symp. on Access Control Models and Technologies (SACMAT 2006), 2006, 49-58.

[Damianou et al, 2001] N. Damianou, N. Dulay, E. C. Lupu, M. Sloman, “The Ponder Policy Specification Language”, in Int. Workshop on Policies for Distributed Systems and Networks (POLICY 2001), 1995, 2001, 18-38.

[Ferraiolo et al, 2001] D. F. Ferraiolo, R. S. Sandhu, S. Gravila, D. R. Kuhn, R. Chandramouli, “Proposed NIST standard for role-based access control”, ACM Trans. Inf. Syst. Secur., 4, 3, 2001, 224-274.

[Hu et al, 2001] V. C. Hu, E. Martin, J. Hwang, and T. Xie, “Conformance Checking of Access Control Policies Specified in XACML”, in 31th IEEE Int. Conf. on Computer Software and Applications (COMPSAC 2007), 2, 2007, 275-280.

[Hughes e Bultan, 2008] G. Hughes, T. Bultan, “Automated verification of access control policies using a SAT solver”, Int. Journal on Software Tools for Technology Transfer (STTT), 10, 6, 2008, 503-520.

[McLean et al, 1983] C. McLean, M. Mitchell, E. Barkmeyer, “A Computer Architecture for Small-Batch Manufacturing”, IEEE Spectrum, 20, 5, 1983, 59-64.

[Miller, 2005] A. Miller, “Trends in Process Control Systems Security”, IEEE Security / Privacy, 3, 5, 2005, 57-60.

[Samarati e De Capitani di Vimercati, 2001] P. Samarati, S. De Capitani di Vimercati, “Access Control: Policies, Models, and Mechanisms”, in Tutorial Lectures at Int. School on Foundations of Security Analysis and Design (FOSAD 2000), 2171, 2001, 137-196.

[Sandhu et al, 1996] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, C. E. Youman, “Role-based access control models”, IEEE Computer, 29, 2, 1996, 39-47.

[Stouffer et al, 2008] K. Stouffer, J. Falco, K. Scarfone, Guide to Industrial Control Systems (ICS) Security, NIST – National Institute of Standard and Technology, 2008.

Data quality e Open data

Domenico Natale
Commission UNINFO JTC1/SC7 "Software Engineering"
c/o Corso Trento 13, Torino – Italy
nataledo@tiscali.it

Scopo del presente lavoro è quello di descrivere il concetto di qualità dei dati, introducendo il recente standard ISO/IEC 25012, e di sottolineare l'importanza di disporre nei vari contesti della massima qualità dei dati, anche al fine di supportare l'Open data. Si considera come l'Open data stesso potrà favorire a sua volta il raggiungimento della qualità dei dati con il contributo dell'utente finale, spesso unico detentore della conoscenza del dato reale nelle sue varie evidenze e co-presenze in diversi sistemi informatici, ancora non tutti integrati tra loro.

1. Introduzione

Le nostre vite stanno aumentando in complessità e questo è spesso causato dalla diminuzione della qualità dell'informazione o dal fatto che essa è difficile da controllare. La qualità dei dati è una componente principale della qualità dell'informazione e parecchi processi della vita sociale dipendono dalla qualità dei dati. L'esistenza di una bassa qualità contribuisce infatti a una informazione insoddisfacente, se non contraddittoria, a risultati non utilizzabili e a rendere gli utenti insoddisfatti.

Ma come si definisce la qualità dei dati? Il modello di qualità dei dati, come definito nello standard ISO/IEC 25012 "Data quality model" [ISO, 2008], è una componente chiave per comprendere il concetto di qualità dei dati, spesso utilizzato in modo vago. Questo modello può contribuire a garantire la semplicità dei contenuti, nelle interfacce utente, nei processi di business, ecc., particolarmente quando è necessario collegare diversi database per scopi di interoperabilità. Tale standard internazionale, pubblicato dall'ISO nel dicembre 2008, e ancora non molto conosciuto, riguarda la qualità del prodotto e non la qualità di un processo. Definisce un modello di qualità per i dati memorizzati in un sistema informatico e categorizza gli attributi di qualità in quindici caratteristiche. Lo standard fornisce per ogni caratteristica una definizione completa e un esempio di misura.

La quantità dei dati gestiti dai sistemi informativi sta aumentando in tutto il mondo. Particolarmente sul web, dove i dati cambiano di continuo, è necessario prestare la massima attenzione agli standard che possono aiutare a gestire tali complessità [Natale, 2011].

In passato la complessità è stata esaminata [Giuliani et al, 1989], [Natale, 1991], [Natale, 1994], [Fasolino et al, 1999] nel campo dell'ingegneria del

software con particolare riferimento al codice (numero delle LOC, commenti, istruzioni decisionali, strutture “one in/out”, modularità). Fin dal 1990 si è data attenzione alla complessità delle interfacce grafiche e alla navigazione uomo-macchina HCI [Mantovani, 1995], [Nielsen, 2000], [Natale e Stortone, 1998]. Più recenti andamenti tendono a completare l'analisi della complessità considerando anche il contenuto dell'informazione rappresentata nelle applicazioni software, il valore e significato del dato stesso. La complessità nel campo HCI non è solo dovuta alle difficoltà incontrate durante la navigazione e l'esecuzione di compiti, ma anche al fatto di incontrare errori nei dati resi. Si sottolinea questo avvertimento perché dati e metadati non corretti spesso appaiono tuttora nelle applicazioni software, interfacce e servizi IT.

E' quindi utile nei test includere la verifica di qualità dei dati, in particolare relativamente alla accuratezza, completezza, coerenza, accessibilità e comprensibilità, come di seguito definite. Questi sono aspetti del contenuto/valore dell'informazione e nulla hanno a che fare con gli artefatti software sviluppati per visualizzare l'informazione.

Come afferma Norman [Norman, 1988] la conoscenza è nella mente umana, mentre l'informazione è fuori di essa: nei testi scritti e elementi culturali. Per questa ragione è importante verificare se l'informazione è conforme il più possibile a caratteristiche attese di qualità, consentendo di evitare o ridurre confusioni, stress o problemi nella mente umana e nei suoi schemi mentali/logici. Questo significa, come rappresentato di seguito (Fig. 1), che un'adeguata coerenza dovrebbe essere presente tra l'informazione percepita dall'utente come attinta dall'ambiente reale, la qualità della sua conoscenza con i processi cognitivi, il software e le interfacce che a volte attingono da fonti diverse [Natale, 2011].

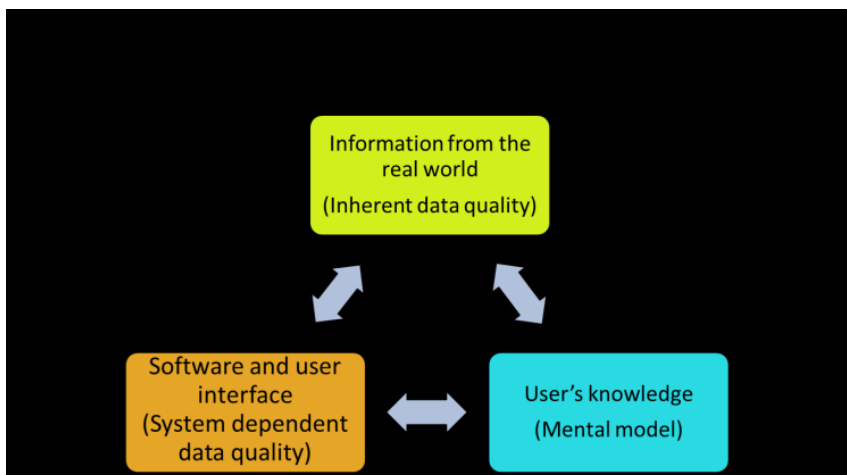


Fig. 1 Modello contestuale: informazioni, software, utenti

Dati e informazioni errati e approssimati sono anche presenti in vari settori della gestione delle informazioni: nei segnali stradali o nei segnali disposti in aree pubbliche come stazioni, aeroporti, ospedali, ecc., sia gestiti manualmente che da supporti elettronici. La quantità in continua crescita di tali dati, digitali e non, richiede di essere gestita usando modelli standardizzati, anche in questi settori che vanno dalle indicazioni stradali, agli orari dei mezzi di trasporto, al grado di inquinamento delle città, al traffico.

E' necessario e utile in questa prospettiva avere un modello in mente, con le caratteristiche da considerare, per migliorare i processi che sottendono la qualità dei dati strutturati e delle informazioni destrutturate. Il modello dovrebbe essere tenuto in mente per evitare gli errori nei dati e migliorare i processi che gestiscono tali dati. Investire in un modello di qualità può aiutare a capire meglio possibili errori nei valori dell'informazione ed evidenziare le possibili cause della bassa qualità, misurandone i costi.

Tener presente un modello di qualità dei dati è una parte essenziale di questa analisi [Natale, 2009]. Il modello è il risultato di una iniziativa italiana di ricerca e una "best practice" in aziende italiane [Sogei, 2010] e nel mondo [Natale et al, 2001], [Scannapieco e Catarci, 2002], sviluppata in ambito internazionale con il lavoro condotto nell'ISO-SC7 da parte del gruppo di lavoro WG6, dal 2004 al 2008.

2. Il modello di qualità dei dati

L'ISO/IEC 25012 categorizza gli attributi di qualità in quindici caratteristiche considerate da due differenti punti di vista: inerente i dati e dipendente dal sistema. Alcune caratteristiche possono essere interessate da entrambi gli aspetti.

Ogni caratteristica può essere valutata nel suo specifico contest d'uso. Il dato è definito come "rappresentazione reinterpretabile di informazione in un modo formalizzato utilizzabile per comunicazione, interpretazione, processo". Ogni caratteristica è di uguale importanza. Il dato può essere considerato la base dell'informazione e della conoscenza digitale; lo standard tiene conto di tutti i tipi di dati, come i testi, numeri, immagini, suoni.

I due punti di vista che raggruppano le quindici caratteristiche di qualità dei dati sono:

1. **Inerente**: quando la qualità del dato si riferisce al valore del dato stesso, a prescindere dal supporto di rappresentazione; generalmente è inerente il dominio gestito da esperti;

2. **Dipendente dal sistema**: quando la qualità del dato è ricercata e preservata all'interno di un sistema informatico; la qualità dei dati dipende in tal caso dal dominio tecnologico che automatizza i dati (hardware, dispositivi, software, ecc.); generalmente la responsabilità di tale punto di vista è dei tecnici del sistema.

Di seguito si riportano le definizioni delle quindici caratteristiche, necessariamente abbreviate rispetto a quelle originali:

Inerente

Accuratezza: il dato rappresenta il valore reale
 Attualità: il dato è del giusto tempo (è aggiornato)
 Coerenza: il dato non presenta contraddittorietà rispetto ad altri dati
 Completezza: il dato ha valori per tutti gli attributi attesi
 Credibilità: il dato ha fonte autentica

Inerente e dipendente dal sistema

Accessibilità: il dato è accessibile, particolarmente dai disabili
 Comprensibilità: il dato è leggibile e interpretabile per l'utente
 Conformità: il dato rispetta convenzioni e regolamenti locali
 Efficienza: il dato consente le prestazioni attese
 Precisione: i dati sono esatti ed è discriminante
 Confidentiality: data are accessible and interpretable by authorized users
 Riservatezza: il dato è accessibile da utenti autorizzati
 Tracciabilità: il dato è soggetto a cronologie di accesso e modifica

Dipendente dal sistema

Disponibilità: il dato ha attributi che consentono l'accesso autorizzato
 Portabilità: il dato è installabile, sostituibile o migrabile tra sistemi
 Ripristinabilità: il dato è recuperabile da copia, anche per system failure

La seguente visualizzazione (Fig. 2) mostra un'interpretazione del modello ISO [Natale, 2010] (la versione originale è una semplice lista di caratteristiche).

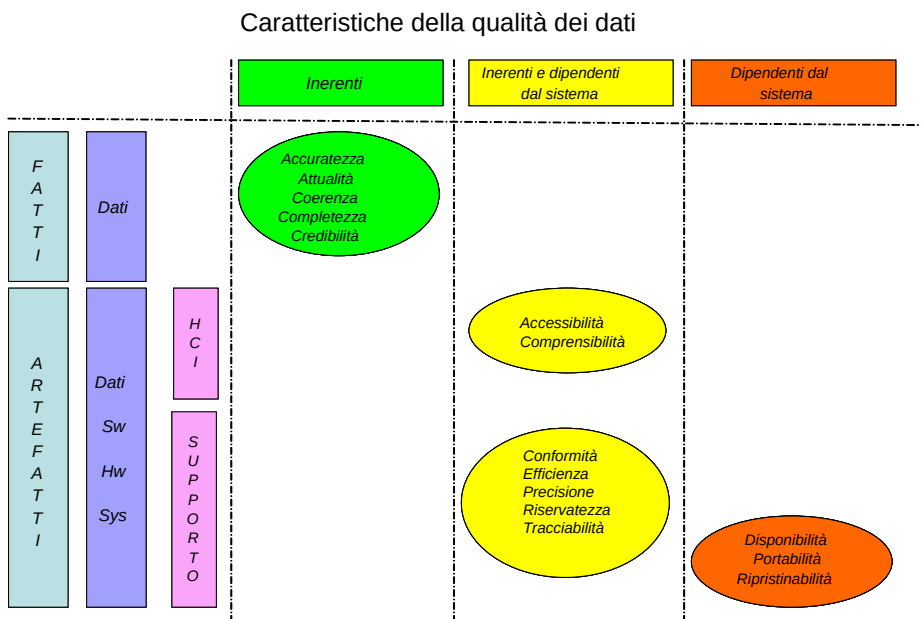


Fig. 2 Modello concettuale di Data quality

Le caratteristiche del punto di vista inerente sono connesse con i fatti della realtà, quelle dipendenti dal sistema sono connesse ad artefatti (supporti, dispositivi, ambienti dove i dati sono ospitati), le altre sono corrispondenti ad ambedue.

In particolare si intravedono varie relazioni con temi attuali. L'accessibilità e l'usabilità sono connesse alla disciplina della HCI. La riservatezza e tracciabilità al Codice della Privacy. La disponibilità, portabilità e ripristinabilità al tema del Disaster recovery/Business continuity e del Cloud computing.

La visualizzazione di queste caratteristiche mostra che alcune sono connesse al dominio, altre alle tecnologie utilizzate o a entrambi i casi. In altre parole ciò significa che per mantenere alti livelli di qualità è necessario controllare sia i significati e i contenuti dei dati stessi, che disporre di un'adeguata infrastruttura fisica e tecnologica.

Ogni caratteristica può essere misurata definendo un algoritmo, un metodo specifico di misura, un livello di accettabilità, in base al contest d'uso. Un grafico Radar può essere utile per gestire le misure e stabilire la distanza tra il valore conseguito e quello atteso [Natale, 2010].

Nel seguente grafico (Fig. 3), concernente un caso esemplificativo, si riportano i valori misurati di correttezza, completezza, credibilità, riservatezza. Le altre caratteristiche non appaiono ancora misurate.

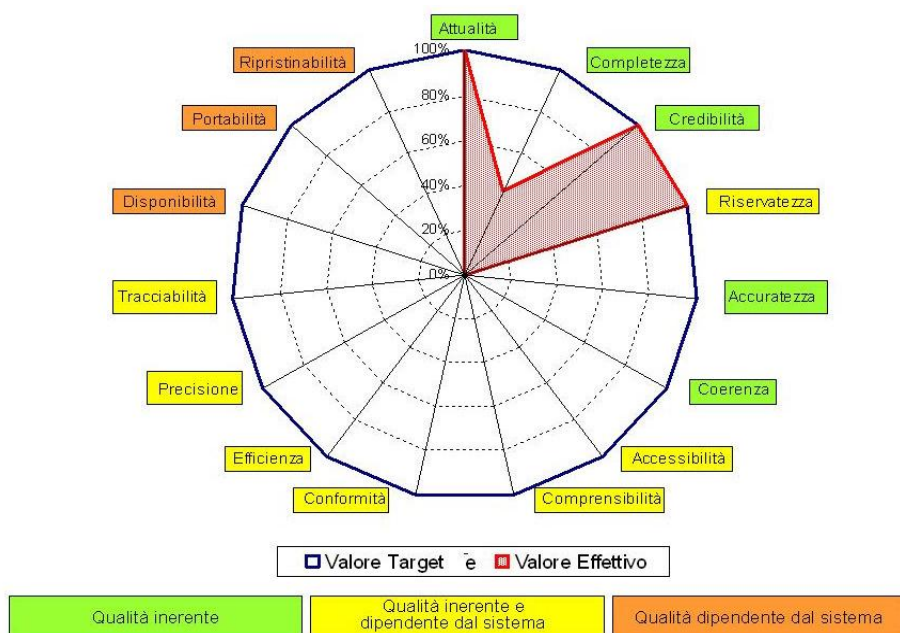


Fig. 3 Data chart

L'importanza di controllare la qualità dei dati è cruciale per migliorare i processi nei casi in cui causano scarsa qualità. I migliori risultati si ottengono con misurazioni aggiornate automaticamente e tenute sotto controllo in data base e grafici Radar disponibile online in rete. Seguendo un concetto di "Total quality" chi apre un processo dovrebbe essere in grado di chiuderlo. L'utente dovrebbe poter "creare" e seguire il proprio dato, sia a livello singolo, se riservato, sia a livello statistico. In questa prospettiva appare utile la pratica dell'Open data.

3. Open data

Si sta affermando sempre più l'approccio, specialmente in ambito pubblico, di rendere disponibili i dati on-line, possibilmente in formato aperto, e si stanno susseguendo seminari e convegni sull'argomento [Tiscornia, 2011] in rapida evoluzione. La diffusione di internet rappresenta una grande facilitazione rispetto ai tempi passati e gli utilizzatori, utenti finali, stanno aumentando ogni giorno. Nascono così nuovi servizi per organizzazioni e imprese e i dati pubblicati e catalogati offrono a molti spunti creativi di analisi, studi e proposte.

Possono verificarsi diversi livelli di "apertura" dei dati: dal puro PDF all'uso di EXCEL, o meglio XML e RDF [Truscello, 2011]. Secondo Tim Berners-Lee possono verificarsi cinque livelli evolutivi tendenti a facilitare la trasmissione e il riuso dei dati [Trentino, 2011]:

- dati pubblicati in qualsiasi formato;
- dati in formato strutturato proprietario (es. Excel);
- dati in formato strutturato non proprietario (es. CSV);
- dati in formato W3C aperto (es. RDF);
- contestualizzazione dei dati che consentono il link verso altri dati in rete (linked open data) [W3C].

Dati interni divengono così esterni e dati che in passato rimanevano inutilizzati possono offrire a studiosi e semplici cittadini nuove opportunità di conoscenza di fenomeni di interesse pubblico: ambiente, salute, mobilità, inquinamento, turismo, vita cittadina, ecc. [Frezza, 2011a]. Esperienze di Open data si stanno avviando in tutto il mondo:

- San Francisco Data-SF;
- New York Data Sets;
- Data Vancouver;
- Berlino APP4;
- Barcellona;
- Marsiglia;
- UK;
- Helsinki;
- Australia;
- Regione Piemonte;
- Regione Sardegna.

Secondo recenti rapporti possono citarsi esperienze in varie altre regioni e province: Emilia Romagna, Liguria, Lombardia, Sicilia, Toscana, Veneto,

Provincia di Trento. Così come stanno moltiplicandosi associazioni e siti internet orientati verso le pratiche Open data [Morando, 2011].

I benefici della diffusione dei dati riguardano cittadini, Pubbliche amministrazioni e imprese, facendo aumentare un senso di trasparenza e partecipazione, credibilità e vicinanza degli organi istituzionali alle persone [Frezza, 2011b]. Le informazioni si diffondono con maggiore tempestività e dovrebbero rispettare le caratteristiche di qualità dell'ISO 25012 come in particolare: accessibilità, comprensibilità, efficienza, coerenza tra differenti fonti, tenendo inoltre conto la conformità alle regole della riservatezza e Privacy. Una soluzione alla inopportunità di diffusione di dati singoli è quella della produzione di dati statistici che offrono i livelli di quantificazione di fenomeni, utili a vari fini, senza entrare nel merito di casi individuali.

I costi di diffusione dei dati aperti dovrebbero essere altamente compensati dalle risposte esplicite o implicite dei comportamenti umani, più consapevoli e meno casuali. Potrebbe così arricchirsi la quantità di informazioni che si muove dal mondo reale agli utenti attraverso il software con informazioni di individui o gruppi [Fig. 4].

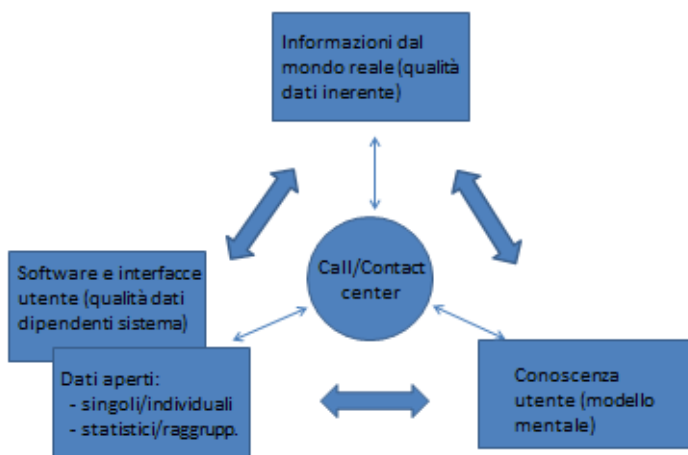


Fig. 4 Modello contestuale aperto: informazioni, software, utenti

Appare evidente che l'Open data può implicare diversi aspetti:

- per i dati singoli/individuali, la diffusione sarebbe limitata ai diretti interessati che avrebbero la possibilità di seguire meglio via web l'andamento dei propri dati, evitando di recarsi agli sportelli;

- per i dati statistici/raggruppati, la diffusione potrebbe essere pubblica ed offrire panoramiche di comportamento con vantaggi di orientamento e prese d'atto di conformità ai fenomeni medi.

Il fenomeno andrebbe monitorato caso per caso con avvii gradualisti e partendo da casi semplici [Morando 2011]: i dati e le informazioni dovrebbero essere particolarmente accessibili e comprensibili per gli utenti al fine di evitare richieste di chiarimento o di comunicazioni correttive che porterebbe ad un incremento eccessivo e aggravio degli impegni di personale amministrativo.

Una visione positiva della pratica dell'Open data suggerisce di poter patrimonializzare i feedback degli utenti, attraverso la possibilità di ricezione delle correzioni di dati: una sorta di creazione di Call/Contact center in grado non solo di offrire informazioni e chiarimenti, ma anche acquisire variazioni e precisazioni. Oppure una serie di funzionalità via internet per consentire un dialogo semplice, facilitato ed efficiente tra cittadino e amministrazione o impresa, come già avviato da alcune Regioni e Comuni [Wikipedia, 2011], [Google, 2011].

Come dire che l'Open data può essere a due vie considerando la possibilità di uscita e di entrata di informazioni. In una tale prospettiva tutto ciò che riguarda dichiarazioni periodiche di dati, verrebbe ad essere sostituita da un flusso continuo di informazioni accessibili e controllabili dagli interessati. I censimenti potrebbero almeno in parte in futuro essere dedotti dai sistemi informatici, come pure le dichiarazioni a vari fini socio-economici-fiscali.

Nell'epoca moderna sembra quasi che gran parte dei problemi finanziari, bancari, economici, inflazionistici, ecc. siano anche dovuti a ritardi e incompletezze delle informazioni conosciute: tardive, tracciate con difficoltà e poco sintetizzate con indicatori sintetici di avvertimento che dovrebbero funzionare in tempo reale. Appare a volte che i vertici decisionali prendano atto a posteriori di fenomeni quantitativi il cui dinamismo spesso non è stato possibile monitorare nel giusto tempo o in modo utile. Un esame dei dati principali che occorrerebbe conoscere, a vari livelli, con maggiore tempestività potrebbe inoltre portare ad una riforma consapevole e condivisa dei processi che li governano.

4. Conclusioni

L'Open data presuppone la disponibilità di dati di alta qualità ed il raggiungimento della qualità dei dati si può avvalere delle pratiche di Open data. In tale prospettiva la società si muove dalla mentalità del controllo a quello della partecipazione, di conoscenze specifiche e ulteriori capacità elaborative, e verso un'accelerazione dello sviluppo governato da dati certi e tempestivi. Nel complesso i risultati attesi da questi indirizzi possono essere così riassunti, senza alcuna pretesa di citare un elenco esaustivo:

- interfacce web altamente usabili che presentino i dati in modo comprensibile e aperto, riducendo errori di acquisizione dati;
- riduzione del ricorso di moduli cartacei;

- workflow amministrativi per la trasparenza e tracciabilità delle pratiche burocratiche;
- elevata comunicazione inter-amministrativa tra i gestori dei dati;
- semplificazione dei processi burocratici;
- miglioramento della qualità in uso per l'utente finale.

L'attuale andamento va quindi dalla forma ai contenuti, dal riuso di software al riuso di dati, dalla facilità della navigazione sul web alla disponibilità di conoscenza per avviare un interesse comune di qualità delle informazioni digitali. Si sta incrementando il circolo virtuoso che va dalla fotografia dei dati reali, alla sua rappresentazione integrata nei sistemi informatici [Batini e Scannapieco, 2006], ognuno con la propria competenza e responsabilità, alla visualizzazione presso l'utente finale che interpreta, eventualmente corregge, ed utilizza per i vari scopi istituzionalmente previsti.

Bibliografia

[Batini e Scannapieco, 2006] C. Batini e M. Scannapieco, Data Quality, Springer, 2006

[Fasolino et al, 1999] A.R. Fasolino, D. Natale, A. Poli, A. Alberigi Quaranta, Metrics in the development and maintenance of software, Journal of Software Maintenance, 1999

[Frezza, 2011a] E. Frezza (a cura di), Incontri ICT 2011, Convegno Finaki, Taormina, Giugno 2011

[Frezza, 2011b] E. Frezza, Dall'Open Government all'Open Data, Convegno Roma Capitale, Roma, Luglio 2011

[Giuliani et al, 1989] P. Giuliani, D. Natale, Metodologia della complessità, Il Pensiero Informatico, Periodico del Gruppo Finsiel, n. 2, 1989

[Google, 2011] Google: Open data Comuni d'Italia, 2011

[ISO, 2008] ISO/IEC 25012, Software Engineering – Software Product Quality Requirements and Evaluation (SQuaRE) – Data Quality Model, 2008

[Mantovani, 1995] G. Mantovani, L'interazione Uomo-Computer, Il Mulino, 1995

[Natale, 1991] D. Natale, On the impact of metrics application in a large scale software maintenance environment, IEEE Computer Society, Proceedings of the Conference on Software Maintenance, Sorrento, 1991

[Natale, 1994], D. Natale, Misure, metriche e indicatori, Congresso Annuale AICA, Palermo, 1994

[Natale e Stortone, 1998] D. Natale , C. Stortone External metrics for quality in use and usability, Workshop, University of Maryland, College Park, 1998

[Natale et al, 2001] D. Natale, M. Scannapieco, P. Angeletti, T. Catarci, G. Raiss, Qualità dei dati e standard ISO/IEC 9126: Analisi critica ed esperienze nella Pubblica Amministrazione Italiana, Workshop AIPA, Sistemi in rete nella PA, Sogei e Università di Roma La Sapienza, in conjunction with the VLDB Very large data base Conference, Roma, 2001

[Natale, 2009] D. Natale, La qualità dei dati e l'ISO/IEC 25012, Magazine "Unificazione&Certificazione", UNI, Milano, 2009

[Natale, 2010] D. Natale ISO/IEC Modelo de Calidad de datos y Data Governance, Capitulo 3, in C. Calero, Ma. A. Moraga, M. G. Patini, Calidad del producto y proceso software, Ra-Ma, Madrid, 2010

[Natale, 2011] D. Natale, Complexity and data quality, Poster e Atti Conferenza 13-16 settembre CHItaly 2011, Alghero, 2011

[Nielsen, 2000] J. Nielsen, Designing web usability: the practice of simplicity, New Riders Publishing, Usa, 2000

[Norman, 1988] Donald A. Norman, The psychology of everyday things, 1988

[Scannapieco e Catarci, 2002] M. Scannapieco, T. Catarci, Data quality under the Computer Science perspective, Università di Roma La Sapienza, IASI-CNR, Journal of ACM, 2002

[Sogei, 2010] Ufficio stampa Sogei, Data quality: Sogei una best practice italiana e internazionale, Semestre Europeo, pp. 72-73, Gangemi Editore, n. 1, 2010

[Tiscornia, 2011] D. Tiscornia (a cura di), Seminario, Gli open data nel contesto italiano, <http://www.cnr.it/cnr/events/CnrEventi?IDn=2527>, Firenze, 2011

[Truscello, 2011] Dal web 2.0 al web 3.0: aspetti tecnici e osservazioni, Vega, Periodico di cultura, Didattica e Formazione Universitaria, www.vegajournal.org

[Trentino, 2011] www.trentinoopendata.eu

[Morando, 2011] Opendatamanual.org, www.datagov.it, spaghettiopendata.org, it.ckan.net, linkedopendata.it, op. cit. pag.73, Federico Morando, Nexa, Public Sector Information/Open data opportunità e sfide, I Venerdì dell'AGCOM, 23 settembre 2011

[Wikipedia, 2011] Wikipedia, http://it.wikipedia.org/wiki/Dati_aperti, 2011

[W3C] <http://www.w3.org/2011/gld/charter.html> Government Linked Data Working Group Charter

e-health in Italy: current state and future prospects of statistical information

Claudio Di Carlo, Elisabetta Santarelli
Ministry of Economic Development, Department of Communications
Viale America 201, Rome - ITALY
claudio.dicarlo@sviluppoeconomico.gov.it
elisabetta.santarelli@sviluppoeconomico.gov.it

ICT is increasingly spreading through the health system in Italy, with the consequences of improving quality and quantity of supplied services. This paper presents a review of the main technologies and data sources currently available in Italy on e-health. The original contribution of the work consists in the proposal of a set of standard statistical indicators to evaluate the diffusion of e-health in the Italian health system.

1. Introduction and aim of the work

The application of ICT in healthcare is known by the term **e-health**, which combines information technology, telecommunications and medicine. The term was created with the aim of applying the principles, potentialities and developments of electronic commerce to the sphere of health and health systems [Eysenbach, 2001]. A more formal definition of e-health is given by Eng [2001]: e-health is "the use of emerging interactive information and communication technology, especially the Internet, to improve or enable health and healthcare". According to the definition of the European Commission [2004: 4], e-health "includes the application of information and communication technologies across the whole range of functions that affect the healthcare sector", which includes doctors, hospital managers, nurses, data management specialists, social security administrators and, of course, the patients.

Therefore, the notion of e-health is very broad, ranging from IT and new technologies to health and healthcare. It is a multidimensional concept. The first two dimensions concern the use and supply of common e-health services: appointments, check-ups, bookings, emergencies, ambulance services, etc. The third dimension concerns IT infrastructure and equipment for the provision of these services. The last aspect is cultural: the use and provision of e-health can only take place within an adequate technological skills among health personnel.

e-health tools and solutions involve products, systems and services that go beyond the most popular Internet-based applications. These tools have positive impacts on the lives of citizens, the working conditions of health personnel and the activity of health offices and national governments. If combined with appropriate organizational changes and the acquisition of new skills, e-health

can lead to more effective prevention and treatments, lower-costs and more accessible and rapid services for the patients (for example, reducing waiting list for medical examinations and visits). The electronic applications can support health personnel in making diagnoses, collecting non-invasive images, preparing for surgery. Savings and increased productivity in health systems are due to the reduction of medical errors, the attenuation or elimination of unnecessary treatments, decrease and/or elimination of paperwork.

e-health can also greatly benefit the economy and productivity of a country by creating jobs and producing incentives for employment [Stroetmann et al, 2006]. e-health is emerging, in fact, as a new industry alongside pharmaceuticals and medical equipment, in which, among other things, ICT has played, and continues to play, a key role. e-health can drive innovation in a variety of disciplines: computer science, medicine, business economics and statistics. The business-economic approach to e-health underlines the need to integrate new technologies, the clinical and administrative processes of healthcare organizations, the skills and culture of health professionals and citizens [Buccoliero, 2010]. e-health is therefore one of the channels through which ICT stimulates economic growth [Di Carlo and Santarelli, 2011].

e-health development cannot be planned without adequate knowledge of the characteristics of the phenomenon itself, which has to be analyzed qualitatively and quantitatively. To this end, relevant and reliable data are needed. The aim of this paper is to describe the main statistical sources on e-health currently available in Italy, analyze the critical issues and review their potential for providing information. A set of statistical indicators will then be put forward to measure the extent of e-health in Italy, which may be used for the quantitative analysis and exploration of the phenomenon.

The paper is organized as follows. Section 2 looks at the diffusion and characteristics of e-health in Italy. Section 3 focuses on issues relating to data interoperability and standardization. Section 4 gives an overview of data sources on e-health currently available in Italy. Section 5 proposes a set of indicators to set up an e-health statistical information system. Section 6 concludes the paper.

2. e-health implementation in Italy

e-health strategies in Italy are in line with European guidelines, in which ICT is a tool not only for better diagnoses and treatments, but also for simplified and easy access to universal services.

In Italy, healthcare is the responsibility of the regions. In 2010 there were 146 Aziende Sanitarie Locali (ASL) (local health centers) and 178 Aziende Ospedaliere (AO) (hospitals) [Ministry of Health, 2011]. About 650,000 medical professionals and nurses were employed in these public structures, including 57,000 general practitioners (GPs) and paediatricians (Pediatri di Libera Scelta PLS). In all, there were about 18,000 public and private pharmacies [Between, 2010].

The Ministry of Health sets national directives and guidelines on e-health for ASLs to implement new ways of organizing and providing services, rationalize

investment and create synergy within a single institutional-strategic e-health framework [Ugenti et al, 2011].

In our country public health expenditure in 2008 amounted to 106.65 billion euros, or 6.8% of GDP [Ministry of Health, 2011]. On the other hand, an increasingly aging population is creating and will continue to create a growing demand for services both quantitatively and qualitatively. Moreover, in terms of social equity our welfare system will have to try to facilitate access to services and treatment, take into account the increasing mobility of both patients and professionals and reduce the disease burden, i.e. the impact of an illness in terms of mortality, disability and economic costs. The fundamental aim is to provide citizens with a better quality of life by reducing the cost of services, through prevention, de-hospitalization and optimization of the hospital network and more effective controls on health spending [Ministry of Public Administration and Innovation - MPAl - and Confindustria, 2009].

Italy took an important step in this direction with the 2012 e-Government Plan presented in 2008, with the aim of simplifying and digitizing primary health services by 2012 (digital prescriptions and medical certificates, online booking systems), creating the necessary infrastructure for the provision of services closer to the needs of citizens, improving service cost-quality and eliminating waste and inefficiency. The government plan includes substantial investments, in particular to create a GPs' network, digitized prescriptions, dematerialization, electronic health records (EHR) and online booking systems [Ugenti et al, 2011]. At the local level GPs form the backbone of the national health service; therefore, a network that includes GPs and the various health facilities (local health centers, hospitals) is essential. The family doctor will play an increasingly active and integrated role in the cycle of patient treatment through the digitization of medical certificates and prescriptions and online booking of appointments.

A survey of the European Commission [2008a] shows that Italy is in line with European countries in the use of computers by GPs, but it is a long way behind in terms of online health applications (only 3% of GPs exchange administrative data with others, compared with an average of 10% for EU 27). This is due to a lack of networking between GPs and other health system actors.

Digital medical certificates, namely the electronic transmission of certificates, instead of paper certificates, by GPs to the INPS (Italian National Institute for Welfare System) and employers, would reduce costs significantly. According to recent MPAl estimates in the early months of implementation (April 2011), the "Digital Certificates Operation" was almost universally adopted by GPs, producing estimated savings of 500 million Euros over the year [Donzelli, 2011].

Digital prescriptions also simplify things for citizens and provide a real-time control of spending: the doctor fills the prescription on the computer and sends it electronically to the central system, giving the patient the prescription number (a paper copy may be requested); the patient receives the prescription at the pharmacy by presenting a health electronic card and the prescription number; the pharmacy checks the prescription online, consigns the medicine and communicates this to the central system.

The **EHR** is a collection of individual data derived from medical records in electronic format from different health centers, hospitals, GPs and PLS. The EHR contains information from medical records about the clinical state of the patient: diagnosis, admission and discharge, specialist check-ups, results of examinations and radiographic images. They are put online for authorized operators and the citizen concerned. The EHR is set up by the regions and autonomous provinces, with the consent of the patient, to support and optimize prevention, treatment, rehabilitation, emergencies and administration [Ministry of Health, 2010]. Lombardy, Tuscany, Emilia Romagna, Friuli and Sardinia have been the first regions to start planning activities to implement EHR.

Currently, there are different regional **phone booking systems** (CUP that stands for Single Booking Centre) for a small number of health facilities in regional areas. The government plans to set up an articulated network of CUPs to allow citizens to book healthcare services throughout the country as part of a larger health reform project. In essence, the plan is to set up a supra-regional telephone and online booking system. In this context, to facilitate access to the CUP it is necessary to increase the number of alternative booking channels: internet portals, mobile phones with internet connections, online pharmacies (at present examinations can be booked via the web only in some health centers and for just a few types of services).

The development of these technological innovations in the national health system is still very patchy due largely to a lack of coordination. According to a 2010 survey conducted by Netics, Italian regions have different degrees of "digital capacity" (the "e-readiness" indicator contains 11 specific indices) ranging from 0.28 in Calabria to 0.82 in Emilia Romagna [Colli Franzone, 2011].

Another important area of e-health applications is **telemedicine**, namely distance ICT assistance. It primarily concerns a series of "classic" electronic services such as teleconsultation, telemonitoring, electronic second opinion, etc., but in its broadest sense it also involves creating links between centers with different specializations, dialogue with equipment in the patient's home and assistance in remote or isolated areas. Telemedicine is currently a limited branch of e-health but it is an extremely important field of application, because it can provide a new type of service, high quality and affordable or at reduced costs, introducing new organizational models such as the "home hospital" and the "virtual ward".

3. Interoperability and data standards in healthcare

To analyze the characteristics of a phenomenon and to understand its effect on society, the first indispensable step is to find out as much as possible about its features, the ultimate aim being the planning of effective intervention policies. In the case of e-health, the data necessary for the study of the phenomenon are stored on computer systems operated by individual health centers. System architecture and structure are particularly important to guarantee high-quality data and, then, statistical analyses.

ICT system **interoperability**, namely the capacity of two or more systems to exchange data, is essential for efficient business processes and this is

especially true in healthcare, which involves the work of government offices, institutions, companies, enterprises, and medical and other professionals. Interoperable ICT systems allow data of different formats to be exchanged: accounting data can be transmitted electronically, health offices or research centers can have access to data on the incidence or prevalence of a disease during a specified period of time, just to name a few examples.

A **standard** is a technical specification that is adopted by different actors in a certain sector of the market through a joint agreement. Standards may be official if made compulsory by law or voluntary if they are the product of a voluntary agreement between parties. Furthermore, business standards are defined by individual companies or groups of companies; standards can also be open to the initiatives of institutions and individuals.

The lack of interoperability and commonly accepted standards compromises service quality and increases healthcare system costs. This also leads to uncompetitive national ICT industries when compared to the rest of the world. Both in Europe and Italy the lack of standards undoubtedly affects the availability and quality of data, the possibility of integrating and comparing them across time and/or space, which represents a barrier to market development and e-health policies. There are different kinds of barriers to the spread of standards. Political barriers: national and local health systems often have different standards and, in general, health policies offer few incentives to encourage homogenization. Definitional barriers: health and medicine are complex areas, whose many dimensions are difficult to define and classify because of cultural, social and contextual differences (medical knowledge changes and evolves rapidly, e-health standards must adapt to them and this requires continuous and costly reviews). User barriers: healthcare professionals in hospitals and health centers focus mostly on process efficiency rather than the spread of electronic systems, of which they usually have little knowledge [European Commission, 2008b].

According to the results of a 2007 European Commission survey on e-health experts (health ministries, competent national authorities, hospitals, universities and research institutes, ICT companies, health professional associations in the IT field), there is a broad consensus about investing more on the spread of uniform standards and, in general, on e-health interventions [European Commission, 2008b]. The lack of uniform standards hinders the distribution of homogeneous data in different areas, which makes it difficult to conduct a comparative analysis of medical, health, and epidemiological data. In contrast, the use of common standards and information systems can have undoubted positive impacts on economic growth and development, and the global competitiveness of ICT companies that provide health services. A US study conducted in 2007 estimated that the implementation of full system interoperability would save 5% of the country's annual health expenditure [European Commission, 2008b]. Though percentages may be different, this is true everywhere, due to the benefits of IT-health standardization: consider, for example, the savings made by avoiding treatments that do not improve health or are not appropriate to patients' condition.

In general, common standards would allow e-health data to have all the characteristics of 'quality' statistical data: relevance, accuracy, clarity, consistency, completeness, timeliness and comparability.

4. Current state of e-health statistics in Italy

To promote effective e-health policies it is necessary to measure the level of its access and dissemination in terms of both supply and demand. Considering that Italy is at an early stage of implementation and use of e-health, data and indicators on the use and supply of these services are still scarce and, where available, have problems of accuracy, comparability and timeliness. There is not enough available data to provide a comprehensive description of the spread of ICT in healthcare. This is due mainly to two reasons: firstly, the fact that health services are decentralized, making it difficult for national surveys to effectively perceive the different regional systems; secondly, the complex mechanisms of data collection, which are slow to adapt to the heterogeneous and changeable situations of healthcare.

However, available data and statistical indicators on the extent and speed of adoption of IT tools are of primary interest in a context where the use of ICT in healthcare is being developed.

A correct and reliable measurement of the spread of e-health in advanced societies is an important tool to guide the choices and decisions of policy makers. Current statistical information needs in the e-health field are represented in the circle in Figure 1 [Ronchi, 2010] on the spread of ICT innovations over time, the corresponding level of ICT activity and the possible future evolution of these information needs.

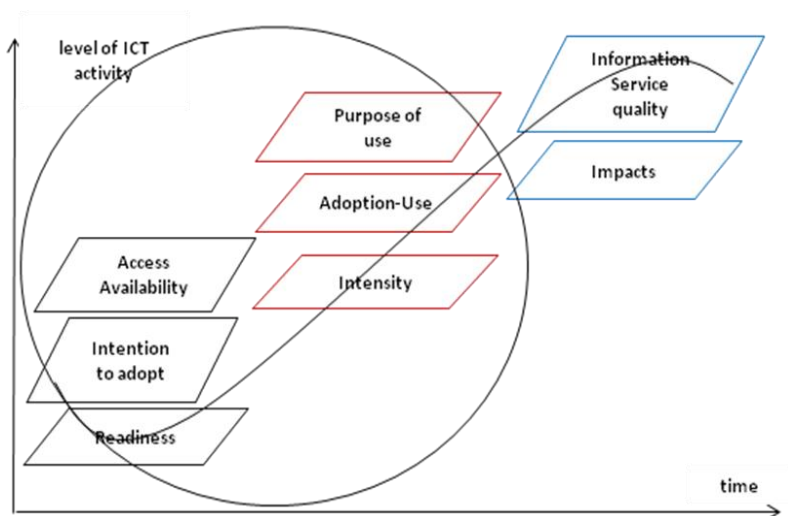


Fig.1 - Main information needs

To date, the most common e-health indicators in major OECD countries concern the use of ICT tools for health services and for administrative and bureaucratic practices. The information provided by e-health indicators is essential to implement health policies that benefit society as a whole. For example, indicators on EHR adoption measure the level of healthcare quality and quantity, so that adherence to clinical guidelines and quality criteria can be monitored, results of the performance of systems measured and the spread of diseases checked. Indicators on the level of user satisfaction can influence policies on the need for financial incentives to promote the spread of healthcare digitization [Ronchi, 2010]. The availability of comparable data and indicators allows comparisons between the policies of different countries, so as to evaluate their effectiveness, the links between incentives and results, the analysis of cost reduction and efficiency in terms of time and space.

To this end the OECD has been offering member countries guidance on implementing a "model survey" to assess the spread and impact of e-health on the whole cultural and socio-economic system. The question of data comparability and statistics should be guaranteed by standard modules that are comparable across time and space [Ronchi and Spiezia, 2011].

The situation of e-health statistics in Italy is similar to that of most European countries: e-health data are currently produced and managed by administrative, management and clinical information subsystems, which are the responsibility of a variety of local centers. Since initiatives are mostly entrusted to regional projects, the aims of the surveys are often distinct and data is collected according to different criteria (sampling design, type of interview, definition and classification of variables, processing of missing data).

The central administrative source for the collection and management of e-health data is the Ministry of Health, through the New Health Information System (NSIS). In addition, individual health system bodies (regions, ASL, AO) have their own information systems, as also ISTAT (Italian National Statistical Institute). Apart from administrative data, e-health indicators are calculated on the basis of specific surveys. Here are some of the main Italian surveys which, to our knowledge, were carried out up to 2010 on various aspects of e-health.

In 2002 Ce.Ri.S.Ma.S. (that stands for Centre for Study and Research on Healthcare Management) conducted a study entitled "Health.Net" to analyze the presence of health institutions (public and private) on the Internet. The results showed that in 2002 less than one centre in two was on the Internet (47%) and that their presence on the web often lacked important structural features (site map, internal search engine, foreign language version) [Baraldi and Memmola, 2003].

Again in 2002, a survey by Confservizi showed that almost all health centers spent less than 1% on ICT: in particular, 93% of centers had their own website, 40% had a call center, 24 % had web contacts [Rossi Mori, 2002].

In 2009, a survey was carried out on healthcare centers with the aim of collecting data and developing a methodology for the evaluation of technological innovation as part of the LITIS project (Level of Technological Innovation in Healthcare). The results show that services, such as online payment of fees or bookings, were still not widespread in Italy (respectively 7%

and 22% of health centers in the survey offered these services) [Rossi Mori and Tamburis, 2010].

In 2010, a survey was carried out by the Ministry of Health and "La Sapienza" University of Rome [2010] with the aim of identifying the needs of citizens in terms of online information on healthcare. The results show that GPs were still the primary source of information for health problems, followed by the Internet. Respondents showed a widespread interest in the online publication of health promotion campaigns (e.g. blood and/or organ donations, occupational safety, responsible use of drugs), healthy lifestyles and health system benefits.

In general, the results of these surveys, conducted independently and for different needs, show that innovation in the health sector in Italy is still fragmentary. Therefore, it is fundamental to implement actions for the harmonization and integration of local subsystems, defining standard variables and common classifications so as to guarantee prompt, reliable and comparable data. This would allow data to be integrated, linked and compared. Technical documentation on data producing process, time references, classification, information type, results, conceptual models and datasets should be updated, harmonized, generalized and easily accessible. Data anonymity would guarantee privacy for citizens, private institutions and national health centers. Among the most important tools for data harmonization is the **metadata register**, i.e. a collection of systematic descriptions of microdata, datasets and indicators. The metadata register includes clinical and administrative variables, the nomenclature used by health personnel, the classification used for institutional information flows, data processing type (e.g. linkage of different archives) [Rossi Mori and Consorti, 2002]. In general, statistical analysis can be oriented towards planning, management, control, epidemiological surveillance, effectiveness of treatments and evaluation of health policies.

The EHR is certainly one of the e-health tools with the highest information potential. For doctors, epidemiologists and scientists the EHR is a mine of information used to study people's health status over time, risk factors, effectiveness of therapies and outcomes of treatments. The EHR is an ideal tool for studying the health of a population, life styles, and the use of health services, because it follows changes over time in health status, lifestyles, medication and exposure to risk factors.

From the point of view of demand and use of e-health, appropriate questions could be included in surveys traditionally carried out by ISTAT on households. In particular, in 2005 the survey "Aspects of Daily Life" included an Eurostat questionnaire on the use of ICT. Respondents were asked whether they used the Internet to seek health information, interact with government offices and buy medicine online. Some more specific questions could be included in this survey on the use of e-health: how often the Internet is used for health information, booking of appointments or making payments, receiving online medical reports, and carrying out medical checks with the aid of IT tools. This would produce savings in both time and money, since it involves only adding some more questions to a survey already under way.

5. A proposal for e-health indicators in Italy

The aim of this section is to propose a set of indicators to measure the level of ICT implementation in health infrastructure and to evaluate electronic services for citizens and the extent of utilization. Detailed e-health knowledge can be used to set up national policies to guide and harmonize the policies of the various local health centers, which are often fragmented. In addition, choices can be assessed, the quality of the new e-health service can be checked, and the use of the technological services by citizens can be analyzed.

The studies carried out so far seem to indicate that overall the phenomenon of technological innovation in individual healthcare centers is evolving much faster than central and regional directives and harmonization: in the medium and long term there is the risk of an even greater fragmentation of electronic service quality and typology. Consequently, it is at present difficult to measure the different e-health services offered by individual health facilities used by citizens. As described in the introduction, technological development in health is moving in several directions representing the dimensions of a broader concept of e-health, which by definition is multidimensional. Each dimension expresses different aspects of technological progress and consists of several sub-dimensions that in turn include single aspects of innovation.

An analysis of all the areas of ICT development in healthcare led to the definition of four basic dimensions: the supply of e-health services, the use of e-health services, the e-health network and the technological know-how of health personnel. Figure 2 shows the relationships between the different dimensions: the network of health centers and the expertise of health personnel are the basis on which e-health services may be implemented. e-health services are provided by health organizations and used by citizens, public and private companies.

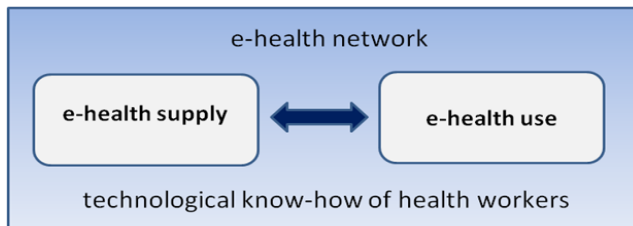


Fig.2 - The dimensions of e-health

The **e-health supply** dimension is part of the institutional task of the national health system to deliver services to citizens, which, with the adoption of e-health, would be more effective and efficient. To measure this dimension we suggest the **eSupply indicator**, which measures the level of technological services offered to citizens in terms of the following sub-indicators: online bookings, medical reports, payment of fees, EHR, communications and telemedicine services (see Tab. 1). The indicator assesses both ICT service typology (bookings, EHR) and the technological level of implemented services (CUP bookings, web, etc.).

Bookings
Medical reports
Payment of fees
Electronic health records
Online communication
Telemedicine services

Tab. 1 - Set of eSupply indicators

The dimension related to **e-health use** measures the use of technological services by citizens. The more these services are publicized, simple to use, cheap, and tailored to citizens' needs, the more they will be used. For this dimension, we propose the **eUse indicator**, which assesses the level of use of technological services by citizens and consists of the same sub-indexes as *eSupply* from the perspective of the citizen (see Tab. 2). This index indicates the use by citizens of the main e-health services, such as medical reports of online analyses, information on diagnoses on the institutional website, illnesses.

Bookings
Medical reports
Payment of fees
Electronic health records
Online communication
Telemedicine services

Tab. 2 - Set of eUse indicators

The **e-health network** dimension concerns aspects of basic infrastructure for the provision of digital health services and is formed by regional networks that link the various local healthcare centers, GPs and PLS. The network is crucial for the provision of quality digital services accessible to all citizens, such as a single booking system that operates throughout the country and applies to all health centers. To quantify this dimension we suggest using the **eNet indicator**, which assesses the extent to which health facilities are connected to the network, i.e. sharing a single standard language, and includes the following sub-indicators: networked local health centers, pharmacies, GPs and PLS, existence of a regional health registry (see Tab. 3).

Local health centers network
Pharmacies network
GPs and PLS network
Regional health register

Tab. 3 - Set of eNet indicators

The eNet indicator expresses the "network effect", that is, the capacity of the health system to offer standard integrated services for all citizens. This dimension includes the regional health register, i.e. essential information that is online for all residents in the region, which represents another element in the network.

The last aspect to be considered in the management of e-health services is the **technological know-how** of health workers, which measures the degree of technological expertise and knowledge among medical staff. The more expert the personnel, the more technological services can be exploited to their full potential. To measure this dimension we propose the **eKnow-how indicator** which takes into account the following aspects: use of digital signatures, ability to compile of digital certificates and prescriptions, electronic medical reports, distance learning and whether the staff has the ECDL Health (see Tab. 4).

Digital signature
Digital certificates
Digital prescriptions
Electronic medical reports
Distance learning
ECDL health

Tab. 4 - Set of eKnow-how indicators

The proposed indicators may be used as tools to improve governance and support the choices and decisions of policy makers in the field of e-health; they provide valuable information to understand the technological phenomenon and, in particular, to assess the validity of ICT services activated by national health system facilities.

The indicators can be calculated using a technique in which for every macro indicator (eSupply, eUse, eNet, eKnow-how) a multiplicity of sub-indicators are transformed into a single numerical value, which can be used for comparisons over time and in different geographical areas. These standard indicators are applicable at different territorial levels and can be used for international comparisons. Moreover, through them we can keep track of the evolution of ICT technologies and assess the level of e-health reached in terms of service typology and quality.

In a follow-up to this paper, we shall test the above methodology on selected cases of best practices to apply the technique used to create them.

6. Conclusions

The time is ripe to extend the use of ICT to both healthcare facilities and the life of Italian citizens. The full implementation of e-health is a challenge for Italy and, more generally, for Europe as a whole: what is needed is commitment from central and local health authorities, a common national strategy supported by appropriate legislation and, generally, a new cultural approach to innovation and technology.

The benefits of e-health for our citizens and country, in terms of improving quality, appropriateness of treatments, governance and reduction of costs cannot be produced without the integration of clinical, economic and organizational data. To this end, there is a clear need for accurate, timely, clear and comparable data to carry out cost-benefit analyses, policy evaluation, and national and international comparisons. These analyses can help identify best practices, assess quantitatively and qualitatively the benefits of e-health, calculate budget savings, disseminate infrastructure and "successful" institutions, and improve people's health. In a citizen oriented system, existing data must be available to both patients and health facilities, each for their own needs and requirements. Therefore, the data must be open and transparent especially in this field. Citizens must be able to access their personal data, for example if they wish to get a second medical advice about an illness. The community (i.e. institutions, citizens, and the business world) must be able to find out information on people's health, government policies for the promotion and protection of health, governance and health management.

The biggest challenge for e-health is to support the everyday management and integration of clinical, organizational, and economic data, both for clinical governance and to support administrative decisions. Local and national authorities can encourage and promote common frameworks through the harmonization and integration of information systems, through shared definitions, classifications and standards and IT infrastructure. In this way, the indicators proposed in this paper may be constructed reliably, promptly and regularly. These indicators may be used to conduct thorough evaluations of the effectiveness of health systems, analyze savings and the benefits of e-health on the country's economic and social system. The adoption of the proposed indicators can help assess the monetary value of the intangible assets provided by e-health which do not have a market value: savings in time, improved safety and appropriateness of treatments, greater attention from healthcare staff, things which are usually measured through proxies. A widespread knowledge of these indicators (hopefully on an ongoing basis and at regular intervals) can provide a better understanding of the benefits of an online health system and, thus, the spread of a culture that is more oriented to the use of new technologies by both citizens and health personnel. This paper, therefore, represents an important contribution to e-health literature for the proposal of described indicators.

In the future it is hoped that the current fragmentary information on e-health in Italy and in most European countries can be improved by implementing common IT infrastructure. As proposed by Rossi Mori and Consorti [2002], the information should be channeled into a national "e-health observatory" linked to a European observatory. To support this initiative, common policies and guidelines are needed, as well as the involvement of all the institutions concerned (governments, hospitals, private companies, citizens and professional associations): in short, a widespread innovation-oriented cultural approach.

References

- Baraldi S., Memmola M., Health.net. Sanità e Internet. Ce.Ri.S.Ma.S. Syntesis report, 2003.
- Between, Sanità elettronica. Rapporto 2010.
- Buccoliero L., E-health 2.0. Tecnologie per il patient empowerment. Mondo Digitale, 36, 2010, 3-17.
http://www.mondodigitale.net/Rivista/10_numero_4/Buccoliero_p_3_17.pdf
- Colli Franzone P., Piano e-Gov 2012. Le sfide per la sanità. Paper presented at the e-health Conference, Rome, april 14, 2011.
- Di Carlo C., Santarelli E., Il ruolo dell' ICT nella crescita economica in Italia, Mondo Digitale, 37, 2010, 3-8.
www.mondodigitale.net/Rivista/11_numero_1/Di_Carlo_p_3_8.pdf
- Donzelli P., Il piano eGov 2012. Obiettivo Salute. Paper presented at the e-health Conference, Rome, april 14, 2011.
- Eng T.R., The e-health landscape: a terrain map of emerging Information and Communication Technologies in health and healthcare, The Robert Wood Johnson Foundation: Princeton, NJ, 2001.
- European Commission, e-health. Making healthcare better for European citizens: an action plan for a European e-health area, 2004.
- European Commission, Benchmarking ICT use among General Practitioners in Europe, 2008a.
- European Commission, ICT standards in the health sector: current situation and prospect, Special Study n. 1, 2008b.
- Eysenbach G., What is e-health?. Journal of Medical Internet Research, 3, 2, 2001.
- Ministry for Public Administration and Innovation, Confindustria, Osservatorio Italia Digitale 2.0, 2009.
- Ministry of Health, Il Fascicolo Sanitario Elettronico - Linee guida nazionali, 2010.
- Ministry of Health, Sapienza University of Rome, Linee guida per la comunicazione online in tema di tutela e promozione della salute, 2010.
- Ministry of Health, <http://www.salute.gov.it>, 2011.
- Ronchi E., ICTs in the health sector: towards and OECD model survey. Working Paper DSTI/ICCP/IIS(2010)6, 2010.
- Ronchi E., Spiezia V., Project on benchmarking adoption and use of ICT in the health sector: towards an OECD model survey. An update. Working Paper DSTI/ICCP/IIS(2011)7, 2011.
- Rossi Mori A. L'ICT nelle aziende sanitarie analisi dei dati sulle aziende sanitarie aderenti alla FIASO nella rilevazione effettuata dalla Confservizi nell'estate 2002.
- Rossi Mori A., Consorti F., A reference framework for the development of e-health, <http://www.e-osiris.it/e-library/databaseOnStrategies.htm>, 2002.
- Rossi Mori A., Tamburis O., LITIS: uno strumento di valutazione e programmazione. eHealthcare, 2, 2010, 8-19.

Congresso Nazionale AICA 2011

Stroetmann K. A., Jones T., Dobrev A., Stroetmann V. N., eHealth is worth it. The economic benefits of implemented eHealth solutions at ten European sites, 2006.

Ugenti R., Di Minco L., Gilardi M.C., Le strategie nazionali in materia di sanità elettronica. e-healthcare, 3, 10, 2011, 14-21.

E-Maintenance of Carpigiani Ice Cream Making Machines

Roberto Lazzarini¹, Cesare Stefanelli², Mauro Tortonesi², Giovanni Virgilli¹

¹R&D Department, Carpigiani Group

Via Emilia 45, 40011 Anzola dell'Emilia (BO)

{robertolazzarini,giovannivirgilli}@carpigiani.it

²Dipartimento di Ingegneria, Università degli Studi di Ferrara

Via Saragat 1, 44122 Ferrara (FE)

{cesare.stefanelli,mauro.tortonesi}@unife.it

The integration of information and communication technologies in the machines maintenance process has the goal of automating management operations, like diagnostics and prognostics, by exploiting the remote monitoring and control of manufacturing devices. E-maintenance has initially interested large plant machinery, but recent ICT advances are making it affordable also for mass-produced equipment such as household and similar appliances. This paper presents the Carpigiani e-maintenance solution for ice cream machines, that provides remote monitoring of machines, automatic notification of malfunctions, diagnostics and prognostics functions, remote assistance interventions, and automated reporting of production data. The Carpigiani e-maintenance solution controls more than 2000 machines and it is significantly improving the after-sale service to Carpigiani customers.

1. Introduction

E-maintenance broadly refers to the application of ICT to the manufacturing process, with the main goal of automating management operations, like diagnostics and prognostics, by exploiting the remote monitoring and control of manufacturing devices [Chrysosolouris et al, 2009] [Muller et al, 2008] [Campos, 2009]. The adoption of proactive e-maintenance solutions could reduce failures and unscheduled maintenance shutdowns.

Most e-maintenance research has so far focused on condition-based monitoring of large plant machinery, in particular in the energy industry and in the heavy one [Kunze, 2003] [Garcia et al, 2006]. The size and cost of those machines justify and encourage the development of sophisticated e-maintenance solutions to improve efficiency. However, the costs to implement machine-installed monitoring functions are significantly decreasing, thanks to recent advances in computation, communication, and software technologies. Relatively high performance COTS microprocessors are becoming affordable

for large scale adoption, thus increasing the computational capabilities available for machine-installed monitoring equipment, while keeping the hardware costs reasonably low with regard to appliance production costs. In addition, the improvements in communication technology now provide several solutions, e.g., Internet and mobile connectivity, that can significantly facilitate the remote access to machines. Finally, modern software development technologies for distributed and embedded applications simplify the realization of complex management functions such as diagnostics and prognostics.

These advances suggest the adoption of e-maintenance solutions also for lower cost machines, such as the household and similar appliances, at least for higher-end appliances such as food processing machines and industrial cleaning ones. Household and similar appliances are mass produced equipments that can be deployed in extremely heterogeneous environments, such as homes, shops, and offices where they operate unattended with no on-site technical support personnel available. It is therefore of utmost importance to detect failures automatically, stop the machine to prevent potential damage, and call immediately for technical support. Prompt failure detection and resolution is indispensable to limit downtimes of appliances, and therefore to preserve revenues.

This paper focuses on the management of a specific class of household and similar appliances, the ice cream machines, and specifically the ones produced by Carpigiani. Ice cream machines are complex and high-duty systems with a long (usually more than 10 years) life cycle. During their lifetime ice cream machines require many maintenance interventions, which are performed on-site and scheduled according to a combination of run-to-failure and periodic maintenance policies. Considered that ice cream machines are often installed in remote areas, on-site interventions for maintenance operations is usually very expensive, for both technical personnel costs and (possibly long) machine downtimes with related revenue loss.

The Carpigiani e-maintenance solution, called Teorema, allows to reduce maintenance costs and to offer an innovative after-sales service to customers. Teorema permits the remote monitoring of each machine operational state (number of ice creams served, engine hours of work, temperatures, etc.) and processes the collected data in order to provide several services, such as the automatic notification of malfunctions, interactive diagnostics analysis, some prognostics functions, etc. In addition, Teorema provides Carpigiani customers with a reporting service for the real time access to the production data of their machines (number of served ice creams, ingredient exhaustion, etc.).

The Teorema system is a client/server Web based application built on a telemetry platform composed of a machine-installed remote control hardware and a centralized monitoring station. The remote control hardware is packaged in a kit that comes pre-installed in new ice cream machine models and can be installed in older models as well.

Project Teorema is already in production, installed in more than 2.000 machines in 25 countries across the entire world. The experience demonstrates that Teorema enables the adoption of an innovative after sales process with

significant benefits for both Carpigiani and its customers. Teorema already represents a strategic asset for Carpigiani to win over competition in an increasingly difficult global market.

2.Carpigiani Machines: the need for E-Maintenance

Carpigiani is the world leader of the ice cream machines market, with more than 50% market share for artisanal ice cream making machines. Carpigiani's ice cream machines are complex and high-duty systems which operate unattended, perform critical food processing operations and have a long (usually more than 10 years) life cycle. As a result, notwithstanding their high-quality construction, during their lifetime ice cream machines require numerous maintenance interventions, which are performed on-site and are scheduled according to a combination of run-to-failure and periodic maintenance policies. Because of the high revenue of ice cream making machines, it is of utmost importance to realize timely detection of failures and deliver prompt technical support interventions.

However, Carpigiani's ice cream machines are installed all over the world, very often in remote areas, and on-site maintenance interventions can be very expensive, in terms of both machine downtimes (and consequently loss of revenues) and significant travel expenses for the technical support. One of Carpigiani's most important customer is in Norway, and his machines are typically installed in remote locations, sometimes more than 500 Km away from the technical service company. Teorema has been successfully deployed on all of the of Norwegian customer's Carpigiani machines, with significant benefits due to its proactive effectiveness.

In addition to maintenance interventions, Teorema has another important function: machines need a constant supply of ice-cream base mix, in order to ensure continuous operation. As a result, customers need to place a special attention in supplying their ice cream machines before they are empty. In this case, Teorema can send a specific warning to the customer, advising him of the need for adding ingredients.

The advantages that the adoption of an e-maintenance platform could bring in the management of ice cream machines are evident. E-maintenance solutions have an enormous, disruptive potential that could completely change the after sales process in the ice cream machines market.

From a technical point of view, the instruments provided by e-maintenance platforms would enable the restructuring of such maintenance processes as monitoring, diagnosis, and prognosis. In fact, they would foster a paradigm shift from a "fail and fix" maintenance strategy to a more cost-effective "predict and prevent" one [Campos, 2009] [Lee et al, 2006]. A significant share of maintenance interventions could be performed remotely, and on-site interventions could be optimized.

From an organizational point of view, the real-time provisioning of both maintenance as well as production data to customers represents a terrific

opportunity to foster the adoption of more effective sales processes and to support more informed business strategy decisions.

3.The Teorema Project

The development of the Teorema solution for the e-maintenance of Carpigiani ice cream machines has required non trivial efforts. Ice cream machines are small manufacturing systems deployed on customers' premises, with unique characteristics with regard to monitoring and communications, such as limited access to machine status data and low cost constraints, that significantly differentiate them from energy industry and heavy industry applications. This effectively prevents the adoption of traditional Supervisory Control And Data Acquisition (SCADA) [Fan et al, 2005] platforms and calls for the development of new e-maintenance solutions based on Internet technologies.

The Teorema e-maintenance solution is realized as a client/server Web application built on top of a telemetry platform. Teorema enables the remote monitoring and control of Carpigiani ice cream machines and offers several e-maintenance services that have changed the after sales process of the ice cream machines. Figure 1 shows the architecture of the Teorema platform, identifying the placement of e-maintenance features between a Central Monitoring Station and machine-installed remote control kits.

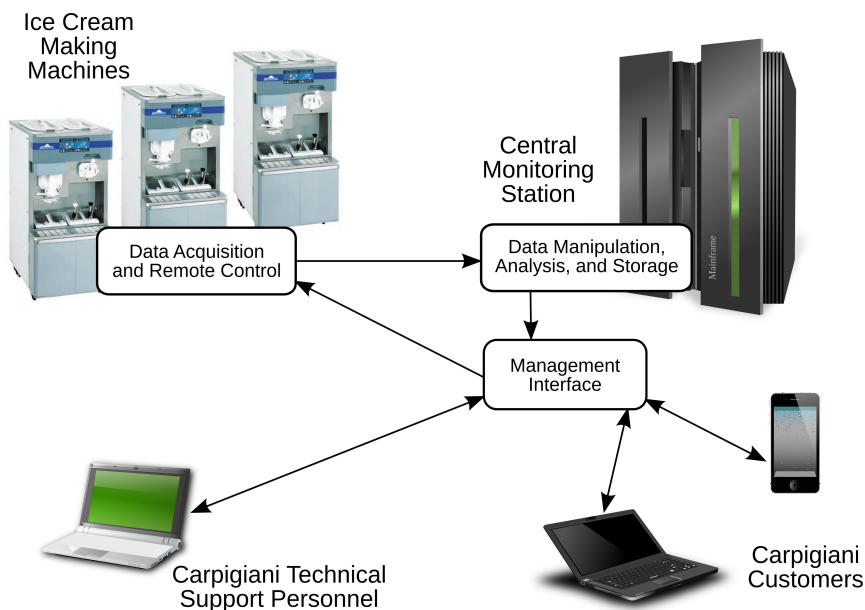


Fig. 1 - The architecture of the Teorema e-maintenance platform.

3.1. Teorema Operations

Teorema provides four main functions: automated monitoring, self-diagnostics and self-prognostics, remote management, and customer data access.

The automated monitoring function permits periodical and/or on-demand reporting of operational state (including failure notifications), from the ice cream machines to the centralized management application. Teorema supports the synoptic visualization of the machine state (programming table, firmware version, analysis of alarms, diagnostics of failures, events, machine access log), updating values in real-time. Time-varying metrics, such as hours of work and temperatures, can be graphically represented. In addition, the system provides an advanced reporting mechanism, that sends to each registered technician a daily report on the state of the machines under his responsibility.

The self-diagnostics and self-prognostics function realizes both the automated diagnosis at the machine side and the centralized just-in-time analysis of data collected from the machines to detect failures, anomalous operations, or machines and components that are likely to fail. In case of unexpected behaviour in a machine, Teorema immediately notifies the technician who is in charge of its management.

The remote management function allows to establish on-demand real-time remote monitoring sessions to help technicians in quickly identifying the cause of faults and in deciding how to proceed for the resolution. In addition, Teorema provides a remote reconfiguration function which enables the automation of firmware and configuration upgrades for ice cream machines.

Finally, Teorema provides Carpigiani customers with a direct access to production data, such as machines productivity (number of served ice cream cones, ingredient exhaustion, etc.), reliability, and geographic location.

Providing production data to customers represents a significant innovation in the ice cream machines market, as it opens up the way to new business models. Customers do not simply buy a manufacturing product, but subscribe to a comprehensive service that monitors machines for both management and production. The servitisation of business models is a very effective way to align a company's offer to its customers' needs [Oliva and Kallenberg, 2003]. In this context, Teorema is a major step for Carpigiani, traditionally a product-based company, towards a more service-oriented offer.

Customers can access Teorema information and services anywhere, anytime, and from a wide range of devices with different screen sizes, CPU capabilities, and input interfaces, e.g., laptops, netbooks, PDAs, and smartphones. In fact, Teorema provides a Web-based, multi-modal interface that tailors data visualisation according to the characteristics of the user terminal device.

3.2. The Teorema Architecture

Teorema adopts a client/server architecture, with machine-installed kits that integrate with the ice cream machine hardware and act as clients sending

maintenance data to the Teorema server, a centralized data monitoring and control station (currently located in the Carpigiani headquarters).

Different Teorema operations are executed at either the client or the server side, with a carefully designed placement that is decided considering the requirements in terms of computational capability and communication costs. For instance, the diagnostics operation is distributed across both client and server components, while the monitoring follows a centralised approach.

Diagnostic functions are split between the ice cream machines and the central monitoring station, because the limited CPU power available on the machines requires to run locally only simple self-diagnostic functions (based on either predefined threshold values for monitored variables or alarm triggered by machine-installed sensors) and to transmit maintenance data to the central monitoring station for aggregation and further processing. On its turn, the Teorema central station performs just-in-time analysis of the received data via computationally expensive diagnostics and prognostics algorithms, further increasing the capabilities to timely detect failures and unexpected behaviours.

The monitoring of machines is, instead, centralized. In fact, only the central management station has the storage capacity to record all the management data and the computational power to process them to provide aggregation and visualization functions. The transmission of maintenance data from the ice cream machines to the central station occurs according to several conditions. The machine-installed diagnostics functions automatically report to the central station in case of malfunctions. In addition, periodical transmissions of maintenance data are also scheduled. Finally, the machines also provide a control interface for remote technical support interventions, enabling on-demand, real-time transmission of monitoring data.

Most of the data received by the central station consist of discrete data sets, also known as “value type” monitoring data in maintenance literature [Jardine et al, 2006], which are more difficult to handle than continuous time series data and present some limits to the applicability of automated anomaly detection mechanisms for real-time monitoring systems [Chandola et al, 2009]. In addition, ice cream machines also record a list of events, including events related to the normal operating conditions of the machines such as commands issued by the machine operators, completion of processes such as pasteurization (in compliance with the food safety preventive approach HACCP), etc., and transfer them to the central management station. This information, also known as “event data” in maintenance literature, is used to realize higher-order detection of malfunctions or incorrect machine usage through complex event correlation techniques [Martin-Flatin et al, 2007].

4. Teorema in Action

Project Teorema is already in production with more than 1 Million of connections and 1.6 Millions of data received from machines. The functions provided by Teorema, as well as the technologies adopted for their implementation, are proving robust and effective. Most importantly, Teorema is

proving capable of reengineering the after-sales process, to provide an innovative service to Carpigiani customers. On the one hand, Teorema improves the maintenance process and significantly reduces after-sales assistance costs, by exploiting remote monitoring and control and the diagnostic/prognostic functions. On the other hand, Teorema allows Carpigiani to provide its customers with direct and real time access to production data (thus opening up to an entirely new business model of after sale service).

4.1. Improvements to the Maintenance Process

The deployment of Teorema on a growing number of machines allowed the optimization of assistance interventions, that now are significantly better planned and scheduled. In fact, the information provided by the Teorema diagnostics and prognostics functions enable to identify which components are likely to break in underperforming machines and which type of assistance intervention is required for failed ones. Therefore, technicians now perform proactive assistance interventions and always carry with them the needed spare parts, thus reducing the need for further on-site operations.

In addition, Teorema has enabled a more effective use of work time from expert technicians, who now can be much more productive as they do not have to spend most of their time travelling to gain on-site access to the ice cream machines. The ubiquitous Web-based access to all maintenance data also facilitates the cooperation of technical support personnel operating on-field with the experts of the Carpigiani management division working at the Carpigiani HQ.

Since the deployment of the first remotely-controlled machines, Teorema has proved to be an extremely effective a diagnostic tool, promptly identifying faults and, in some cases, even identifying worn down components and machines that are likely to break in the near future. For instance, in one machine Teorema detected the upcoming failure of the beater blades component by identifying a negative trend in the ratio between the number of served ice cream cones and the number of working hours for the refrigeration engine.

Teorema has also allowed to perform several remote maintenance operations. For instance, recently a UK-based customer requested the customization of 90 already installed Carpigiani ice cream machines. Using Teorema, Carpigiani technical personnel managed to accomplish this task by simply upgrading the machine firmware online. The Teorema remote configuration functions have also significantly facilitated the machine installation and configuration procedures. In fact, machines now do not need to be pre-configured in the factory or on-site, but are simply installed and configured online. In addition, in the last few months we have performed 10 remote reconfigurations of Teorema-controlled machines to tune their working parameters, e.g., scheduling of the pasteurization process, according to their deployment location.

Overall, the adoption of Teorema has lead to very significant savings. In the last month, around 20% of the assistance interventions for Teorema-controlled

machines have been performed remotely. This has brought to a reduction in maintenance costs of almost 25%, and we estimate that in the long run reductions could stabilize around 40%.

Finally, from the Carpigiani's customers perspective, the promptness improvement in maintenance intervention enabled by Teorema translated into much smaller machine downtimes, and therefore significantly decreased loss revenues due to faults.

4.2. Customer Access to Production Data

The first customers that have adopted Teorema are heavily using the new customer oriented service for production data analysis. By accessing real-time data on the production, customers can have a timely and significantly higher resolution view of the ice cream sales processes. Carpigiani customers are using Teorema to verify the productivity of their machines, in order to assess which are the most productive ones, to taking actions for improving the productivity of shops with lower sales figures and, in some cases, they are considering whether to install more machines at shops with high sales figures.

Carpigiani customers also use Teorema to make sure that their machines do not experience costly downtimes due to improper or untimely management by operating personnel, e.g., unjustified delays in the load of ingredients or in the cleaning of the machine. For instance, in a recent maintenance intervention a Carpigiani technician received an alarm from a Teorema-enabled machine and managed to trace the source of a problem to an obstruction in the ventilation grid – a misuse that can lead to serious damage to the machine. The technician immediately contacted the customer, who confirmed that there was a folded cardboard box leaning on the back of the machine and removed it, thereby fixing the problem.

Finally, customers take advantage of Teorema to monitor the state of their ice cream machines and immediately access information about the machines' assistance history, including up-to-date reports on the status of ongoing maintenance interventions. This feature grants customers a transparent view of the Carpigiani after-sales service operations, allowing to evaluate its efficiency and its conformance to the subscribed service level agreements.

The features provided by Teorema are particularly appealing for franchise chain owners, that can also reap additional benefits from the integration of production data in their enterprise information system. For instance, this could significantly facilitate the management of supplies for ice cream machines.

Customers access to Teorema mostly via laptops, but the first smartphone access have already demonstrated the correct design of the interface that can adapt to a wide range of devices.

5.Implementation

The development of Teorema is exclusively based on Open Source technologies, as they allow Carpigiani to protect its investments by not restricting the software platform to a single proprietary vendor. In particular, the

technologies used in Teorema are highly innovative, albeit already mature for the industrial use and very well supported.

To remotely control ice cream machines, Teorema leverages on a machine-installed kit that is realized as a custom board powered by ARM-based Ethernet platform. The Ethernet board runs Nut/OS, an embedded operating system allowing a single application and providing some level of concurrency based on the cooperative multithreading model.

The Nut/OS programming environment provides an almost complete compatibility with ANSI C, supporting most functions of the standard C library and thereby easing the porting of programs to/from other platforms. Nut/OS also provides the Nut/Net TCP/IP API for communications, which is similar to the traditional BSD Sockets API.

As a result, while the capabilities of the Ethernet platform are limited, the adoption of an ANSI C-like programming environment and of the ARM hardware platform allows Carpigiani to later extend the CWebGate kit to a full fledged Linux-based system while preserving important parts of the current software and hardware design.

The centralized monitoring and control station makes extensive use of high performance and high scalability Open Source solutions, running on top of Debian Linux.

All the functions provided by Teorema functions are available from a Web-based interface, realized as a Ruby on Rails application. Ruby on Rails is a framework for the development of Web 2.0 applications based on the Ruby programming language and designed to enable agile programming practices. The Ruby on Rails characteristics allow the rapid development of new features, minimizing their time-to-deployment. This makes Ruby on Rails a perfectly suited framework for a dynamic and rapidly-evolving platform such as Teorema [Vooslo and Kourie, 2008] [Backle and Kirchberg, 2007].

6. Conclusions

The availability of mature and cost effective information and communication technologies are making possible the adoption of e-maintenance solutions for household and similar appliances. This paper has presented Teorema, an innovative e-maintenance platform for ice cream machines, that has allowed Carpigiani to achieve considerable savings through a reduction of on-site maintenance interventions as well as the restructuring of the maintenance process. In addition, Teorema has significantly improved the quality of after-sales assistance service provided to customers. We have received a very positive feedback from Carpigiani's customers with regards to the capability to access their after-sales maintenance data in real time. We have been reported that in some cases Teorema has been a key factor in the customer decision to buy Carpigiani's machines.

Future Teorema developments will also consider the automated construction of knowledge from customer reports and telemetry data, by exploiting data mining techniques. The evolution of on-board processors will also allow to

implement more sophisticated diagnostics functions directly on the ice cream machines. We expect that in the next future e-maintenance solutions like Teorema will be investigated and put in place for other household and similar appliances markets.

Bibliografia

[Chryssolouris et al, 2009] Chryssolouris G., Mavrikios D., Papakostas N., Mourtzis D., Michalos G., Georgoulas K., Digital manufacturing: history, perspectives, and outlook. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 223, 5, 2009, 451-462.

[Muller et al, 2008] Muller A., Crespo Marquez A., Iung B., On the concept of e-maintenance: Review and current research. Reliability Engineering and System Safety, 93, 8, 2008, 1165-1187.

[Campos, 2009] Campos J., Development in the application of ICT in condition monitoring and maintenance. Computers in Industry, 60, 1, 2009, 1-20.

[Lee et al, 2006] Lee J., Ni J., Djurdjanovic D., Qiu H., Liao H., Intelligent prognostics tools and e-maintenance. Computers in Industry, 57, 6, 2006, 476-489.

[Kunze, 2003] Kunze U., Condition telemonitoring and diagnosis of power plants using web technology. Progress in Nuclear Energy, 43, 1-4, 2003, 129-136.

[Garcia et al, 2006] Garcia M. C., Sanz-Bobi M. A., del Pico J., SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a windturbine gearbox. Computers in Industry, 57, 6, 2006, 552-568.

[Oliva and Kallenberg, 2003] Oliva R., Kallenberg R., Managing the transition from products to services. International Journal of Service Industry Management, 14, 2, 2003, 160-172.

[Jardine et al, 2006] Jardine A., Lin D., Banjevic D., A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20, 7, 2006, 1483-1510.

[Chandola et al, 2009] Chandola V., Banerjee A., Kumar V., Anomaly detection: A survey. ACM Computing Surveys, 41, 3, 2009.

[Martin-Flatin et al, 2007] Martin-Flatin J. P., Jakobson G., Lewis L., Event Correlation in Integrated Management: Lessons Learned and Outlook. Journal of Network and Systems Management, 15, 4, 2007, 481-502.

[Vooslo and Kourie, 2008] Vosloo I., Kourie D., Server-Centric Web Frameworks: An Overview. ACM Computing Surveys, 40, 2, 2008.

[Backle and Kirchberg, 2007] Bachle M., Kirchberg P., Ruby on Rails. IEEE Software, 24, 6, 2007, 105-108.

[Fan et al, 2005] Fan R., Cheded L., Toker O., Internet-based SCADA: a new approach using Java and XML. Computing & Control Engineering Journal, 16, 5, 2005, 22-26.

Energy production from multiple renewal sources in small towns: a question of effective communication

Andrea F. Villa¹, Agostino Villa²
¹*Link Campus University*

Via Nomentana, 35 – 00162 Roma
E-mail: villandrea@libero.it

²*Politecnico di Torino – DIGEP*
C.so Duca degli Abruzzi, 24, 10129 Torino
E-mail: agostino.villa@polito.it

A wide and effective utilization of renewable energy sources is conditioned on understanding their usefulness and convenience by people living in any community. This means two needs: on one side, for the population potentially interested to implement an energy production system in their small town or village, to understand the usefulness of integrating different renewable energy sources (e.g., solar, wind, biomasses); on the other, for the designers of a new co-generation installation, to understand the crucial role of an effective communication of the plant utilization, such to make the population actively involved in control of the plant installation and the operations management. This contribution will approach both aspects. First, a simple integration scheme of multiple energy sources utilization will be outlined, together with the main features of this integrated system. Then, a survey of different communication initiatives in some European countries, where specific projects have been started to make clear how to install, utilize, monitor and control the most diffused systems for renewable energy production, will be presented.

1.Introduction

The installation of an energy production plant using renewable sources is hardly conditioned on a good acceptance of the population living in the area. On the other side, a common utilization of renewable energy, mainly in small towns and/or villages, is becoming a necessity for saving costs and keeping as clean as possible the ambient.

The distrust of population against any new installation of energy production is generated by two main causes: on one hand, they are unable to understand how renewable energy sources can be used to generate energy any time, through an integrated utilization of more sources together; on the other, they do not receive an effective communication of the use and management of an

energy production plant based on renewable sources, of its usefulness and convenience.

This contribution will approach these two aspects. First, a simple integration scheme of multiple energy sources utilization will be outlined, together with the main features of this integrated system. Then, a survey of different communication initiatives in some European countries, where specific projects have been started to make clear how to install, utilize, monitor and control the most diffused systems for renewable energy production, will be presented. Just as an example, the Institute of Energy and Sustainable Development (IESD) of the De Montfort University in Leichesther is promoting “inspiration for action” events for local communities, to learn about the steps they could take to start local green groups and heard about practical actions these groups could carry out, such as saving energy in community buildings or organizing energy saving events (see at: <http://www.iesd.dmu.ac.uk/news/index2010.htm>). The Communities Cutting Carbon project is a two-year Knowledge Transfer Partnership (KTP) project, involving the IESD and the Rural Community Council (RCC) for Leicestershire and Rutland. This clearly shows that the dissemination of knowledge about renewable energy and energy saving requires a deep, clear and continuous communication program, by using language to be immediately understood by a variety of people and appealing communication systems.

2. Renewable-Energy Multi-Source Multi-Producer (REMSMP) System

“ Usually, when urban planners start the design of a new settlement, they look for pre-existing landmarks, such as roads, railways, rivers, etc. and align the new buildings and streets accordingly. Very rarely do they look for the most ancient pre-existing landmarks: the path of the sun and prevailing winds” [Butera, 2008].

As society becomes more aware of the need to protect and preserve our planet, many people are looking at the benefits of renewable energy resources. These energy sources are the key to a cleaner and greener future.

Traditional energy sources, like oil, cause damage to our environment through air pollution and waste, which is one of the main reasons why scientists and researchers are looking at renewable energy sources. Indeed, gas emissions come primarily from the combustion of fossil fuels in energy. The world carbon dioxide emissions are expected to increase, following the present trend of energy use, by 1.8 percent annually between today and 2030. More than 50% of world population today lives in a urban systems and cities contribute to 67% of world's primary energy demand and emissions. Given the projected rate of urbanization, cities are expected to increase this share to 73% in 2050. Buildings are one of the largest end of energy; in OECD countries, the building sector accounts for 25-40% of the final energy demand, 33% in commercial buildings and 67% in residential. Worldwide energy consumption for buildings is expected to grow 45% from today to 2025 [Vettorato & Zambelli, 2009].

Renewable energy refers to a source of power harvested from a sustainable and long-lasting sources, here listed in the following. Solar power refers to energy harvested from the sun. Photovoltaic cells draw in heat from the sun, as a form of kinetic energy. This energy is then transferred to a generator, which converts it into electricity. Solar water heating (SWH) systems comprise several innovations and many mature renewable energy (or SHW Solar Hot Water) technologies. Wind-mills capture the power of the wind and send the energy to a turbine that converts it into electricity. Geothermal energy uses the heat from the ground. The ground collects a lot of heat during the day, which creates steam. That steam can be captured by a generator and used to produce power. Bioenergy, or biomass, occurs when natural products such as wood is burned. The burning lets off heat. Ethanol fuel is also a form of bioenergy created by breaking down plants into starches and sugars, fermenting and creating a wood alcohol. Currently, hydropower is the most widely used source of alternative energy, produced from the flow of water. By creating a dam, water can be directed to flow through turbines, which use the force of the water to create electricity.

Renewable energy production may contribute to a local economy in three different ways: (i) by creating jobs; (ii) by creating local tax-income from sold-off energy; (iii) by lowering energy costs.

A region, village, or city is called "energy autarkic" if it covers the main part of its demand for heat and electricity from locally produced renewable energy (wind, sun, water, biomass/gas). Furthermore this production is "sustainable" in the sense of the our common future [Brundtland Report by the United Nations, 1987] only if energy production does not result in any significant reduction of food production, by creating a concurrence for agricultural space.

In this light, [Greenbang, 2009] asserts that small-town renewable energy has big potential: small-scale renewable energy schemes in small, rural communities could play a big role in helping to reduce greenhouse gas emissions and meet national climate-change targets.

The technology of an integrated renewable energy system is a local grid, which combines several renewable local producers (wind-turbines, solar panels, waterturbines, biomass, and biogas) and bypasses intermediaries like the big grid monopolies by linking production directly to consumption.

However, in a renewable energy multi-source system, some technical and managerial problems are still open because of the necessity of integrating several energy sources.

Then a model of a **Renewable-Energy Multi-Source Multi-Producer (REMSMP) System** must take account of: (a) the source variation in time; (b) the necessity of interconnecting several producers; (c) to be able to satisfy the required energy demand by the local social-economic ambient (typically, a small town of about 1500-3000 citizens).

In addition, the REMSMP System Model must be designed in order to : 1. optimize the integration of the different Energy Producers, to maximize the efficiency of the whole local grid; 2. design the best possible control strategy of the Energy Producers operations, to minimize any waste and pollution.

A simplified scheme of a REMSMP system is illustrated in the following Figure 1, where three main phases are integrated into a multi-sources energy production, namely: (a) selection of the energy source more usable at the current time; (b) set of steps ("production chain") to produce energy from the renewable source in input; (c) control of both the source utilization and the energy production.

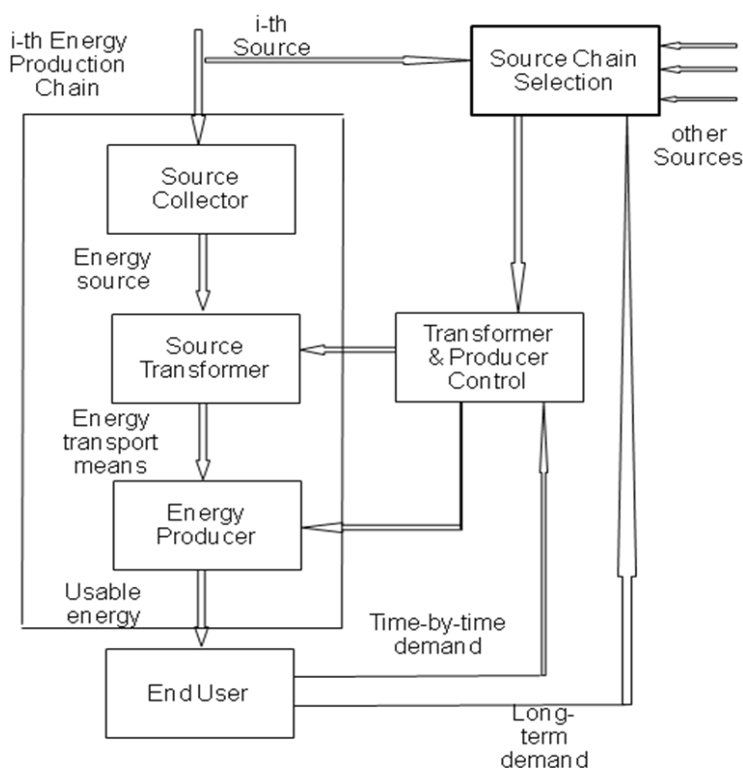


Fig. 1. Scheme of the Renewable-Energy Multi-Source Multi-Producers system

The REMSMP System Model here proposed includes two different loops (and related management levels):

1. external loop, to optimally select the energy sources to be contemporarily used in order to maximize the efficiency of the whole local grid. This loop is designed by solving a Constrained 0-1 Linear Programming (C-LP) Problem, [Gill et al., 1981] where the cost to be maximized is the global efficiency of the set of selected Energy Production Chains with the constraint to satisfy the average mid-term demand.

2. internal loop, to optimize the control strategy of each Energy Production Chain, in order to minimize any waste and pollution. This loop can be designed

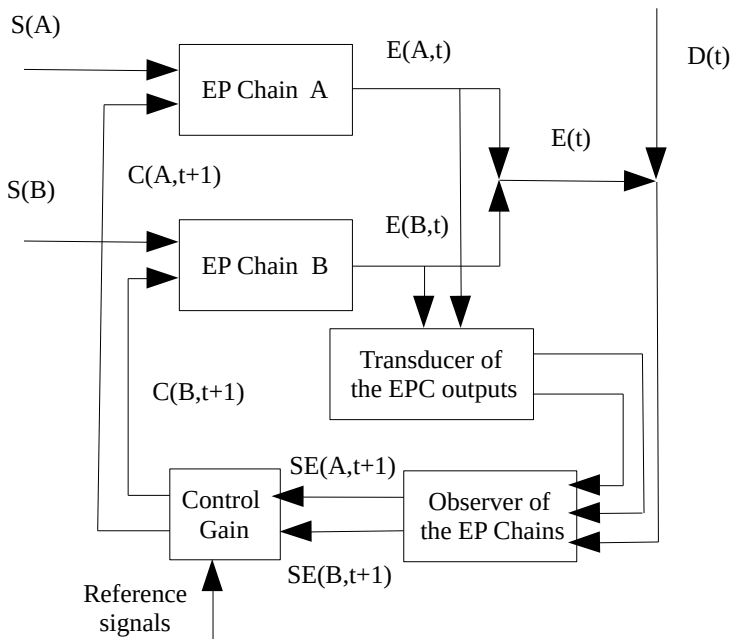
by solving an Optimal Control Problem [Brandimarte & Villa, 1995] for each energy production system over a short-term horizon.

The main aspects of the two-level management/control strategy are the following:

(A) The most complex problem is the mid-term selection of parallel Energy Production (EP) Chains, because: (a) the most profitable EP Chains must be selected over a period in which the respective sources are high and not significantly variable; (b) the Source Transformers in all selected EP Chains must be „compatible“, i.e. the energy produced by each one must be „summed“ to the others.

(B) The time by time control of the selected set of Energy Production (EP) Chains, can be approached according to the following control diagram (see Figure 2).

Fig. 2. Real time control scheme of the Energy Production Chain (EPC)



In summary, Figure 2 makes evidence of the operations of the REMSMP system when two different sources could be managed, namely source A, denoted by $S(A)$, and source B – $S(B)$. From each one, the system could generated energy at time t , i.e. $E(A,t)$ and $E(B,t)$ respectively, whose integration will give rise to the energy usable for satisfy the demand $D(t)$. An “observer” [see Luemberger, 1969] will forecast the energy generation in the near future, such that the control gain could regulate each chain accordingly.

The application of the above sketched feedback control to a bench of different renewable energy sources will allow to track each source evolution in time with a very low involvement of personnel. However, some crucial conditions have to be taken into account, because the installation and management of a system using multiple energy sources is revealing to be convenient for a population of end-users living at least in a small town or a village. Then, the form of organizing the necessary technical infrastructure must be “participative” to the extent that local consumers take a share in property relations of the production and distribution facilities and engage in relevant decisions.

From the above main issue, besides the technological analysis of integrated renewal energy systems (first part of this paper), two further R&D lines have to be developed:

- evaluate by which communication approaches it could be possible to make understandable the convenience of renewable energies to citizens → this is discussed in the second part of this contribution (see also [Villa & Andretta, 2011];
- evaluate which cooperation management approaches could be applied to design and implement renewable energy systems in living ambient: this is outlined in the paper by [Bellomo & Villa, 2011].

3. Effective communication to counteract nimby.

No longer a mere suggestion of what might be, renewable energy (RE) today is hitting a tipping point, with far-reaching implications. In fact, today for the first time, understanding the scale and patterns of RE development has become essential to any full analysis of trends that will shape the global economy and the health of the planet. Thanks to the policies of several governments; the acceleration of private investments and the advances of technology, RE is breaking into the mainstream of the economic markets.

But any development in the installation of new energy source production systems is today obstructed by the people mistrust: the recent disasters in nuclear plants have greatly increased the so called “nimby” opinion. Nimby is an acronym for the phrase “*not in my back yard*”. The term (or the derivative Nimbyism) is used pejoratively to describe opposition by residents to a proposal for a new development close to them. The term was coined in 1980 by Emilie Travel Livezey, and was popularized by British politician Nicholas Ridley, who was Conservative Secretary of State for the Environment. Projects likely to be opposed include (but are not limited) to tall buildings, chemical plants, industrial parks, wind turbines, incinerators, mobile phone networks masts, nuclear waste dumps, new highways and railways. Two weeks ago a report written by Bankitalia stated that the Nimby “syndrome” is growing up in Italy. It’s a sign of the lack of confidence of people in the politic and economic choices of the Government; at the same time is a signal of the difficulties of the Government in communicate his decisions about RE to the citizens. Last week about 5 to 10.000 people participated to protest near Turin against the high-speed railway line between Piedmont region (north-west of Italy) and France.

Then, an effective communication is crucial in order to explain to the people the change and the benefits of RE [OECD, 2009]. An example of the communication problems with local population. Often the placement of the wind turbines at the centre of the landscape makes hard contentious issues between government & people. The role of the media often exacerbates the situation, repeating misstatements put forward by W-P's opponents (noise; landscape beauty; pollution created by the transport system; etc.). On the other side, the proponents of W-P are too quick to dismiss any problems raised, levelling the charge of NIMBY. To empower communities to make such decisions demands a more mature and responsible approach from the media, the wind industry and pressure groups on both sides of the debate. The noise of the rotating blades of the turbines can be a real problem: UK, USA & Germany's governments created specific institution (called "Noise Working Group") in order to analyze & check the situation, according the guidelines of the World Health Organization. At the same time wind industry began to refine the blade design, with the goal to reduce the noise. The Italian-Israeli architect David Fisher in 2008 designed a new kind of blades: in late 2009 this solution were employed to build the new Dubai's private hotel skyscraper.

In 2009 the Genoa administration planned to create a "wind farm" in the Apennine's mountains (between Genoa & Turin). In this area there are few villages: local people (organized by the "green party") began to protest against the project. Why? Because the lorries picking up the turbines will create pollution along the road. Since 2009 until now the project is blocked. Probably young generations will be more "open-minded" & able to solve the problems.... (see the following Figure 3).



Fig. 3. A toy box to build a windmill

Helping people change the way they think and act on energy must be a goal for the UE governments since some time [European Commission, 2011]. In fact, appropriate energy advice is one of the keys to a sustainable future: it's

essential that people can understand what they can do to improve the energy efficiency of their homes, offices and public buildings. Many governments (above all the British) in the last few years created specific agencies able to talk with people, local authorities and private enterprises, helping them to understand and engage with the important issues linked with the R-E. It's crucial to understand that only with the support, cooperation and active participation of communities we will see real changes in the way we use and source energy. Individuals take their cues from the actions of others around them and, through organised collective action, people can achieve far more than they can acting in isolation.

In summary the goal of the R-E communication is to 'change the way people think and act on energy'. Main actions can be by: giving energy advice, encouraging dialogue between wind developers and communities, spreading the word about energy efficiency, exploring new ways of communicating the issues of climate change, running workshops and training sessions, and mainly, by empowering young people with knowledge about climate change, energy use and renewable energy, such to empower their potential as energy activists in their own homes and schools. The goal is to enable young people to think about sustainable energy and what they can do, both in their future lives and right now as influential members of their home and school communities. The Ministry of Education in UK between 2000 and 2004 delivered a range of innovative programmes used in more than 1.000 schools around the Country, reaching some 18,000 children of the primary schools.

The illustration of some experiences can show how an effective communication is necessary, because implementation of renewable energy systems imply people sensitiveness and cooperation. Darmstadt University of Technology in Germany won the 2007 Solar Decathlon in Washington, D.C. with this passive house designed specifically for the humid and hot subtropical climate. But day-lighting features such as the oculus at the top of the Pantheon, in Rome, Italy, have been in use since antiquity (see the following Figure 4-a and 4-b).



Fig. 4. (a) Passive house – Darmstadt; (b) Pantheon oculus.

Some other examples: from the Solar House #1 of Massachusetts Institute of Technology in the United States, built in 1939, using seasonal thermal storage for year-round heating, to the solar water disinfection in Indonesia; from the Solar Bowl in Auroville, India, which concentrates sunlight on a movable receiver to produce steam for cooking, to the cell park which concentrates sunlight from a field of heliostats on a central tower. All these are renewable energy applications which can be disseminated everywhere to make people known about the convenience as well as their role of warranters of a sustainable development.

4. Some concluding remarks

Perception holds great importance in energy policy decision-making.

It is the essence of democratic society that elected officials should act upon the views of the public. Thus where public and consumer reaction to a certain type of energy is negative, politicians may in many cases have less inclination to pursue that energy type as they would if public perception were favourable. This can be a serious problem if public perception is not informed of key facts and long-term implications of policy decisions. This lack of awareness can lead to distorted perceptions that can in turn impact on the course of policy-making.

If effective communication of sound facts is not in place, debates can easily become misinformed and can lead to policy decisions that may jeopardise the energy security of a country or region.

To have a clear information of the opportunity, the convenience as well as of the potential problems of a new installation for producing energy from renewal sources is of crucial essential importance today, when the plant for energy co-generation from multiple sources are becoming a real opportunity for small towns and villages [Kunze & Busch, 2011]. The integration scheme illustrated in the Section 2 of the paper indeed is a basic structure for multiple-sources plants which could be built in the very large number of Italian sites: the reader must note that, only in the Piedmont region, more than one thousand of the 1206 municipalities are small or very small towns, and all could be equipped with many-sources plants, e.g. with biomass, and/or solar, and/or wind, and/or geothermic.

In any site the installation of a common co-generation plant should be viewed by population and politicians as the real way to save both energy and the ambient. Then, the cooperative development of the right plant, and its management, must become the main criteria for renewable energy source utilization.

References

Bellomo D., Villa A., *From the new opportunities offered by renewable energy sources towards new integrated management of territory*. Proc. Int. Conference "Renewable energy sources and sustainable development", Paneuropean University Apeiron, Banja Luka, BiH, June 2011

Brandimarte P., Villa A., Advanced models for manufacturing systems management, CRC Press, Boca Raton, FL, 1995.

Brundtland Report by the United Nations, 1987

Butera F. in: Urban Energy Transition: From Fossil Fuels to Renewable Power, Droege P., ed., Elsevier Science; 1 edition, Oxford, 2008.

European Commission, FP7 Cooperation Work Programme: ENERGY.2012.7.1-1: Integration of variable distributed resources in distribution networks, July 2011.

Gill P.E, Murray W., Wright M.H., Practical optimization, Academic Press, London, 1981

Greenbang (2009): Small-town renewable energy has big potential: http://www.greenbang.com/small-town-renewable-energy-has-big-potential_11766.html

Kunze C., Busch H., The Social Complexity of Renewable Energy Production in the Countryside, Electronic Green Journal, 1(31), 2011.

Luenberger D., Optimization by vector space methods, Wiley, New York, 1969

OECD, Addressing Perception and Communication in the Energy Sector, Business and Industry Advisory Committee to the OECD, France, www.biac.org, 2009.

Vettorato & Zambelli, Estimation of energy sustainability at local scale, 45th ISOCARP Congress 2009

Villa A.F., Andretta D. Renewable energy dissemination in communities: a question of effective communication, Proc. Int. Conference "Renewable energy sources and sustainable development", Paneuropean University Apeiron, Banja Luka, BiH, June 2011

Federating e-identities across Europe, or how to build cross-border e-services

Diana Berbecaru¹, Antonio Lioy¹, Marco Mezzalama¹,
Giorgio Santiano², Enrico Venuto², Marco Oreglia²

¹Dip. di Automatica e Informatica

²Integrazione Processi e Sistemi Informativi

Politecnico di Torino,

Corso Duca degli Abruzzi 24, 10129, Torino, Italy

first.last @ polito.it

This work discusses the main challenges and requirements of technical and legal authentication interoperability in e-services involving organizations from different countries. We present the Stork cross-border authentication framework that satisfies these requirements by establishing interoperability between existing European national eID infrastructures in a federated manner. As a sample application, we present the pilot for Student Mobility that has been developed to evaluate and demonstrate the functionality of the Stork authentication framework.

1. Introduction

During the last decade a lot of effort has been spent to permit on-line execution of several public administration tasks, to reduce costs and save time. Examples are obtaining marital status, birth certificates, or residence attestations or submitting income tax declarations or paying bills online.

Typically, to access an on-line public or private service (e.g. offered by universities, public offices or private companies), a citizen needs first to register directly with the entity providing the service, named the *Service Provider (SP)*. During the registration phase, the SP requires and stores several attributes associated with the citizen (e.g. name, surname, address, age) as well as his unique national identifier, like fiscal code or social security number. The SP must then use additional mechanisms to check that the asserted attributes correspond to the identity claimed by the user, for example by sending via surface mail part of an access password to his home address. Also at registration time, the SP creates and associates the citizen to his credentials (e.g. username and password) that can later be used for authentication purposes.

Credentials come in various forms; apart from the trivial reusable passwords, higher security levels may be obtained by using one-time passwords and cryptographic tokens (like smart-cards). Tokens often contain identification data (e.g., photo or biometric data) and user profile data (e.g., the user ID inside an

organization) in addition to the base cryptographic material (e.g., a digital certificate and the corresponding private key). These tokens are used for authentication and/or digital signing purposes, and are considered more secure than authentication based on username and passwords.

Note that in this simple model, the “producer” of the user profile is also its “consumer”, as the SP itself implements the services requiring the authentication credentials created upon user registration. As a consequence, very often the credential issued by one provider is not automatically recognized by other SPs. Thus nowadays users end up with several different credentials to access various e-services: this creates confusion, opens the way to various attacks, and makes difficult the control of personal data.

Since the above model is costly and inefficient both for users and SPs, the *federated identity management (FIM)* model has been proposed in recent years, allowing users to register and authenticate once and then access various resources across several different domains. In this model, the user registers his identity or profile with one organization (called the *Identity Provider, IDP*), yet manages to access the services offered by other SPs without any further registration. User profiles maintained by one organization can be trusted by another organization, provided that the two organizations established a relationship, called also a “Circle of Trust”. Several different FIM solutions have emerged over the last couple of years such as Microsoft’s Passport [Mic], Liberty Alliance [Liberty], Security Assertion Markup Language (SAML) [OASIS], Cardspace [Infocard] and OpenID [OpenID].

Using FIM in practice, especially in cross-border scenarios, is not only a matter of technology. As the number of IDP and SP increases, the number of trust relationships to be established and maintained becomes quickly unmanageable. Moreover, as explained in [ENISA], when the IDP and the SP belong to different countries (as in a cross-border authentication scenario, where the SP is in country B whereas the IDP is in country A), the SP is actually governed by different laws and business rules than the ones holding for the IDP. Moreover, the IDP and the SP might even use different technologies which may be incompatible [Arora].

In a converging European society, missing interoperability between national *electronic identity (eID)* infrastructures threatens to compromise the success of electronic cross-border services. The European Commission has therefore launched several large scale pilots to address the key issues of technical and legal interoperability of electronic identities to facilitate citizens’ access to e-services across the EU. These initiatives are explicitly mentioned in the e-Government action plan [ECeGov] as enablers of a single European digital market, which is one of the main goals of the European Digital Agenda [ECDigitalAgenda].

To overcome the eID interoperability issues and facilitate secure cross-border authentication, the European project Stork [Stork] implemented a pan-European infrastructure which allows cross-border recognition of national eID solutions. It is important to note that Stork is not introducing any new type of authentication token or registration procedure, rather it uses what is already established in each European country as a legally valid eID. These various eIDs

are mapped to one of four assurance levels defined in Stork to express the reliability of the authentication credentials and registration process.

To validate this approach, Stork is running various cross-border pilots. We'll describe here the Student Mobility application that allows citizens from several EU countries to use their own national eID for on-line application at various universities, including the Politecnico di Torino in Italy. For Italian citizens, this permits the use of various eID, including the CNS (National e-Services Card), CRS (Regional e-Services Card), and CIE (Electronic Identity Card).

The remainder of this paper is organized as follows: Section 2 gives an overview of the basic FIM concepts, Section 3 presents the Stork framework and its operation for a generic cross-border e-service, Section 4 describes the Stork Student Mobility pilot, and Section 5 draws the conclusions.

2. Basic FIM Concepts

Whenever a human is involved in identity interactions, the federated model involves four logical components [Maler]: the user, the user agent (UA, such as a browser or other software application running on a PC, a mobile phone or any other electronic device), the SP, and the IDP (typically a web site that authenticates the users by using credentials such as usernames and passwords or digital certificates).

The main protocols proposed so far for FIM implementation are: OpenID, SAML, which ties into the Liberty Alliance Identity Web Service Framework standard, and the Identity Selector Interoperability Profile specification (often referred as InfoCard) underlying Windows Cardspace. The salient feature of most of these protocols is that they only require a standard web browser as UA while the providers may use several federation protocols, depending on the relationships between partners. Since SAML is a very extensible open standard, it is often used as a basis for several architectures with specific extensions. For example, various protocols of the Liberty Alliance project [Hodges] as well as the Shibboleth project [Shibboleth] build on SAML. The SAML protocol functionality is simple: when the user accesses via the user agent a service provided by an SP, the SP creates an authentication request in SAML format and redirects the user to the IDP to authenticate the user.

Once the user has been authenticated (e.g., by means of username/password or digital certificates), the IDP acting as a SAML Authority creates a SAML assertion in XML format, which is valid for a specific time interval and contains a declaration about the date and method of authentication. In addition, the assertion can include other attribute statements besides the authentication information. Each attribute statement can contain one or more attributes, each of which has associated a name in a specific format. The IDP typically signs the assertion to guarantee its authenticity and integrity. The protocol allows also encrypting the assertion to guarantee its confidentiality. The authentication response containing the issued SAML assertion is sent to the SP using a SAML binding. For example, in case of the web browser Single Sign On (SSO) profile [Cantor] the HTTP POST binding is commonly used, which means the SAML assertion is sent through the user's browser with the HTTP POST

method. Finally, the SP verifies the SAML assertion, including the digital signature applied on the assertion and its temporal validity. If the verification succeeds, the SP provides the requested service to the user.

As part of a federation configuration, the SP must establish a trust relationship with its partner IDP, typically via manual out-of-band exchange of certificates used to protect the SAML messages. Thus, each SP will end up with a list of partner IDP certificates. The SP accepts SAML protocol messages from a partner IDP if the certificate used to sign the SAML assertion validates the signature against its trusted list.

Additionally, the SP might require that the user is authenticated using strong authentication methods (such as a smart-card) in order to grant him access to particular services. Consequently, the SP should individually agree with each IDP about the meaning and the format of the authentication level required by the SP in certain service contexts.

3. The Stork Interoperability Framework

The Stork framework performs a number of basic functions, such as the retrieval of the person's eID and related attributes and the transport of these attributes to a trusted SP.

These services could be either offered by a national proxy server, named PEPS (Pan-European Proxy Services), or they could be provided directly by a dedicated software, which is named Identity Middleware. Thus, the Stork architecture supports two approaches – the proxy (or PEPS-based) approach and the middleware approach – and each European Member State (MS) country can freely decide which one to support. The middleware approach is specifically suitable for those cases in which the smart-card is used for user authentication, as detailed in [Tauber]. Both the PEPS-based and the middleware approach use the SAML 2.0 standard as foundation for authentication and user attribute transfer. However, the SAML format has been extended to support user-related attributes significant for access control services.

A European MS adopting the proxy approach will run a PEPS server, which is part of a centralized European infrastructure used to connect the national identity infrastructures. Each PEPS has two interfaces: the interface used for communicating with the SPs and the IDPs in his own country is MS-specific (thus its implementation is left at MS' choice), whereas the interface used for communicating with the other PEPS servers, based on SAML 2.0, was developed in the project itself. In the Stork framework, the SP contacts its own national PEPS to ask for user authentication and (on the other hand) the IDP will respond to its national PEPS about user authentication requests originating from that PEPS. In addition, in Stork, the SAML authentication request contains an indication of proper authentication assurance level required to get access to an SP service. Currently, four authentication assurance levels are defined, which are related to the assurance required for the user's identity: level 1 (providing minimal assurance) indicates that the user can be authenticated via reusable passwords, level 2 (providing low assurance) demands use of identity

Tokens (software or hardware) and One Time Passwords (OTPs), in a level 3 authentication (providing a substantial assurance) the OTPs are still tolerated, but identity tokens are strongly recommended, whereas in a level 4 authentication only strong crypto tokens (like smart cards with qualified certificates) are accepted.

Let's consider the case – shown in Fig. 1 – of a person trying to access a remote service (provided by a foreign SP) when both countries adopted the proxy approach. The PEPS in the SP country is named also S-PEPS, whereas the PEPS in the citizen's country is named C-PEPS.

To access a web-based service, the user is first asked to authenticate (step 1). As the SP is connected to Stork, it creates a SAML authentication request containing the attributes required for the service and sends it to the S-PEPS through the UA (step 2). In addition, the SP or the S-PEPS provides the citizen with a web page to select the citizen's originating country. Based on this information the citizen is redirected to her national C-PEPS. Upon country selection, the S-PEPS constructs a signed SAML request containing the required attributes mapped to Stork format, which is sent (through the user's browser) to the C-PEPS (step 3).

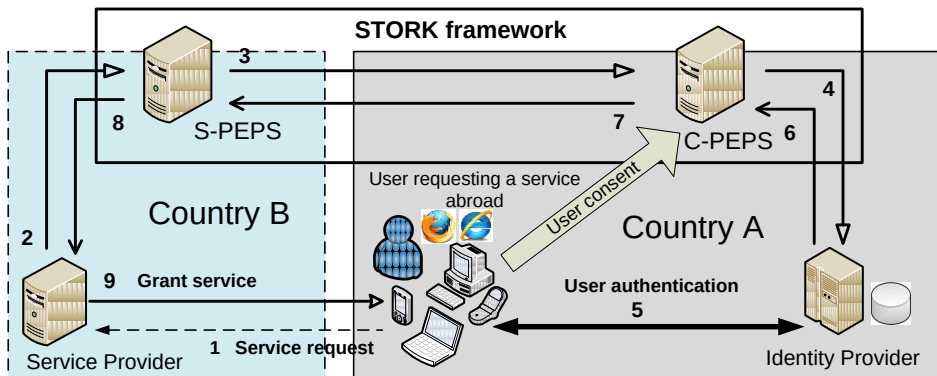


Fig.1 – Using the Stork framework (PEPS-based) in a cross-border e-service.

The C-PEPS selects the appropriate national IDP for user authentication and creates an authentication request where the user attributes required are mapped from the Stork format to the format recognized by the IDP, and sends it to the IDP (step 4). In step 5, the IDP performs the authentication exchange with the user and returns the appropriate response to the C-PEPS (step 6). After validating the response, the C-PEPS maps the received attributes to the Stork format, and (if necessary) derives additional attributes (e.g. "ageOver" derived from birthDate, or an pan-European eID uniquely identifying the person abroad derived from the national electronic identifier) for privacy protection purposes. In addition, the C-PEPS requires also the user's consent to forward his attributes to the S-PEPS. Finally, the C-PEPS creates a signed SAML response (containing the user attributes values) and sends the response to the

S-PEPS through the UA with the HTTP POST method (step 7). The S-PEPS performs similar operations as the C-PEPS, and sends subsequently the newly created SAML response to the SP (step 8), where the certified attributes are extracted and verified to eventually grant access to the requested service (step 9).

The main innovation features of Stork can be summarized as follows:

- the possibility for an European citizen to use her own national e-identity (in whatever form it is, e.g. one-time password, code provided via a mobile phone, smart-card) when accessing a foreign web-based service provided by a service provider connected to Stork;
- the possibility (for a SP) to automatically identify a user from a foreign country and have this identification be performed with a specified level of assurance.
- the possibility to automatically retrieve specific attributes about the user (like name or age) so that she does not need to enter them. Since these attributes are certified by the national eID infrastructures of the European countries participating in the project, the user cannot lie (and this fact is very important for the provider offering the service and requesting identification of the user).
- the high scalability of the Stork infrastructure allowing a SP and an IDP in an European country to easily interconnect with the other providers in the other Member States through the Stork infrastructure.

In addition, a prototype service has been also implemented allowing validation of digital certificates through Stork. This service is of crucial importance for applications or services exploiting digital certificates, like the creation and verification of electronic signatures.

4. Building the Student Mobility e-service with Stork

Based on Stork, we designed and implemented a flexible and scalable student mobility service, allowing both foreign students to apply on-line at Politecnico di Torino using their own national credentials, as well as Italian citizen to apply online at foreign universities integrated in the Stork framework.

Our contribution is twofold. First we integrated in Stork our service for foreign student application at the Politecnico di Torino, and the architecture of this service is illustrated in Fig. 2. On the other hand, we also integrated a real IDP with Stork (as shown in Fig. 3), allowing citizens that have previously registered at Politecnico di Torino's IDP or that have a smart card issued by our organization [PiemonteCard] to access other cross-border on-line services at foreign universities connected to Stork. Examples of such services are access to eLearning platform for students and teachers or applying online to Erasmus programs. As explained in Section 3, the Service Provider portal contains the protected on-line application that can be accessed via common web browsers and requires user authentication.

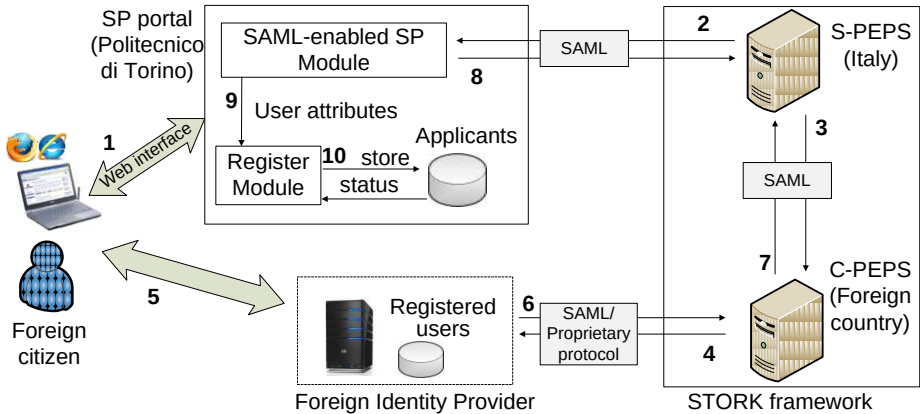


Fig.2 – Integrating the Stork framework into Apply@polito web portal.

Integrating the Stork framework into Politecnico di Torino's Apply@polito web portal. To allow foreign students to apply online at our university, we developed a web-based application running at the university's portal (as shown in Fig. 2), which has been derived from the Demo SP provided in Stork. This application is actually a web application running on an application server such as Apache Tomcat [Tomcat], and it contains several basic components: a web interface initially accessed by the users to apply on-line at our university indicating that the procedure can be fulfilled using their national credentials; a SAML engine (based on OpenSAML [OpenSAML]) allowing the SP to construct SAML authentication requests and to extract user attributes from the SAML authentication responses received from the national PEPS; configuration files containing data required to digitally sign the SAML messages, to indicate the location of the S-PEPS, and the list of user attributes that can be required through Stork. When the user accesses the portal, he will be presented with a web page (step 1) in which he can choose to perform his Application@polito with an European eID. Next, he is presented a page indicating the attributes required to complete the application, some of these attributes are mandatory (i.e. the Given Name, the Surname, the identifier and the date of birth), whereas other are optional (i.e. the residence address, the country of birth, the nationality, the e-mail address and the gender). At the next step, the user is asked to select his country of origin. On selecting the country, the SAML authentication request is sent first to the S-PEPS (step 2) and subsequently to the C-PEPS running in the foreign country (step 3). The format of the messages exchanged in these steps is specified in [Alcalde-Moraño]. At the C-PEPS, the user is required to either authenticate directly (since some PEPS servers embed this functionality in the national proxy) or he is redirected to a specific national IDP (step 4) to complete authentication (step 5). The communication protocol used between IDP and C-PEPS is often based on SAML but it can be even a proprietary protocol (as described in Section 3). The authentication response is sent back through the Stork framework to the SP, where the

dedicated web application extracts the required attributes and pass them to a Registration module (step 9) in charge with storing the data into a relational database, such as an Oracle database (step 10).

The experimental setup put in place for this service consists of a SP machine running the web application derived from the Stork Demo SP application, and the PEPS machine running the PEPS package developed in the project, which contains also a SAML engine based and a set of Java servlets handling the SAML requests and responses arriving from the SP and from the other PEPS respectively.

Providing identity and attribute transfer service through Stork. For this service, we set up a SAML-enabled IDP (as shown in Fig. 3) in charge with authenticating users using any of the three types of authentication methods that have been mapped to the Stork levels of assurance: the username/password (corresponding to level 2), the digital certificates saved in files, named also “soft” certificates (corresponding to level 3) and the digital certificates saved on smart cards (corresponding to level 4). In addition, the IDP is also in charge with providing user attributes, by extracting them from the dedicated database containing the students and teachers registered at the Politecnico di Torino. The IDP is connected to the national PEPS, so that assertions about authentication and user attributes are transferred to the foreign SPs through the Stork framework. In our experimental setup, the IDP machine runs the Shibboleth software (v2.0). We chose Shibboleth because it incorporates a SAML 2.0 engine and it permits easy management of the IDP functionality through dedicated configuration files. We extended the IDP functionality by integrating authentication with digital certificates (either stored in files or on smart-cards) besides the username/password authentication supported by default in the Shibboleth package.

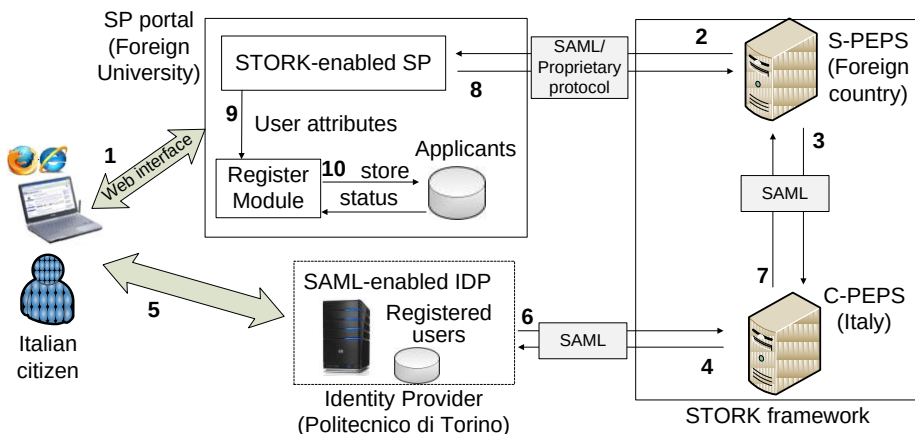


Fig.3 – Providing identity and attribute transfer services through Stork.

5. Conclusions

In this paper, we presented the European e-ID interoperability framework Stork, which offers scalable, secure and privacy-aware cross-border authentication for public and private e-services. Besides the technical aspects, it is important to note that Stork uses legally valid national electronic identifiers and puts the user in control of his personal data.

We have also reported about practical applicability of Stork, focussing on the pilot application for Student Mobility. It demonstrates the main features of Stork, such as integration with the national IDPs and SPs, and the technologies developed to simplify the implementation of a cross-border e-service, such as the one allowing people to apply on-line at European universities by using their own national credentials. The Student Mobility application is currently being piloted in universities from Italy, Austria, Spain, Portugal and Estonia. Our experience, gained in more than nine months of use in a production environment, indicates that Stork is a viable solution to implement cross-border e-services for European citizens.

Acknowledgements. This paper describes work that was developed in the framework of the EU co-funded project Stork (INFSO-ICT-PSP-224993, www.eid-stork.eu).

References

- [Alcalde-Moraño] Alcalde-Moraño J., Hernández-Ardieta J. L., Johnston A., Martinez D., Zwattendorfer B., Stern M., Hepe J., Stork deliverable D5.8.2b Interface Specification, October 2010.
- [Arora] Arora S., National e-ID card schemes: A European overview, Information Security Technical Report, Vol. 13, Issue 2, 2008, 46–53.
- [Shibboleth] Cantor S. (editor). Shibboleth architecture - Protocols and Profiles, <http://shibboleth.internet2.edu/docs/internet2-maceshibboleth-arch-protocols-latest.pdf>, Sept. 2005.
- [Cantor] Cantor S., Kemp J., Philpott R., Maler E., Profiles for the OASIS Security Assertion Markup Language (SAML) v2.0, in OASIS Standard. <http://docs.oasis-open.org/security/saml/v2.0/samlprofiles-2.0-os.pdf>, 15 March 2005.
- [DigitalAgendaEurope] Digital Agenda for Europe, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Brussels, August 2010, http://ec.europa.eu/information_society/digital-agenda/index_en.htm
- [ECeGov] European Commission: The European eGovernment Action Plan 2011-2015, COM(2010) 743, Brussels, 2010.
- [ECDigitalAgenda] European Commission: A Digital Agenda for Europe, COM(2010) 215 final/2, Brussels (2010).
- [ENISA] ENISA Risk Assessment Report: Security Issues in Cross-border Electronic Authentication, February 2010.

- [Hodges] Hodges J, Wason T, Liberty architecture overview, 2003.
- [InfoCard] Microsoft. Identity Selector Interoperability Profile V1.0, <http://download.microsoft.com/download/1/1/a/11ac6505-e4c0-4e05-987c-6f1d31855cd2/Identity-Selector-Interop-Profile-v1.pdf>, September 2007.
- [ISA] Interoperability solutions for European public administrations (ISA) (OJ L 260, 3.10.2009, p. 20). ISA replaces the IDABC programme (Interoperable delivery of pan-European eGovernment services to public administrations, businesses and citizens (OJ L 181, 18.5.2004, p. 25).
- [Liberty] Liberty Alliance. Identity web services framework 2.0, <http://www.projectliberty.org>, 2006.
- [Maler] Maler E., Reed D., The venn of identity: Options and issues in federated identity management. IEEE Security & Privacy, pages 16–23, March/April 2008.
- [Mic] Microsoft. Windows Live ID/Passport Network, <https://accountservices.passport.net/ppnetworkhome.srf?vv=700&lc=1031>.
- [OASIS] OASIS. OpenID Authentication 2.0 - Final. Assertions and Protocols for the OASIS Security Markup Language (SAML) v2.0, OASIS Standard, <http://www.oasis-open.org/apps/org/workgroup/security/>, March 2005.
- [OpenID] OpenID Foundation. OpenID Authentication 2.0. Final, December 5, 2007. http://openid.net/specs/openid-authentication-2_0.html.
- [OpenSAML] OpenSAML libraries, <https://spaces.internet2.edu/display/OpenSAML/Home>.
- [PiemonteCard] Piemonte University System Smart Card. <https://didattica.polito.it/segreteria/sportello/en/Smartcard.html>, 2011.
- [Stork] Secure Identity Across Borders Linked (Stork) project - Towards pan-European recognition of electronic IDs (eIDs) (2008-2011), <http://www.eid-stork.eu>
- [Tauber] Tauber A., Zwattendorfer B., Zefferer T., Mazhari Y., Chamakiotis E., Towards Interoperability: An Architecture for Pan-European eID-Based Authentication Services, in Proc. of EGOVIS 2010, Bilbao, Spain, LNCS Vol. 6267/2010, 120-133.

Il recupero energie in un'azienda manifatturiera

Luisa Spairani
Net Surfing S.r.l.
C.so Vercelli 332P 10015 Ivrea (TO)
spairani@netsurf.it
Davide Gagnor
Tekfor Italy S.p.a
Via Martin Luther King ,6 10051 - Avigliana (TO)
Davide.Gagnor@neumayer-tekfor.com

L'azienda metalmeccanica Tekfor, dopo aver razionalizzato la gestione della manutenzione impianti adottando metodologie e strumenti informatici adeguati, sta ora avviando una nuova fase che aggiorna metodologie e l'IT per il monitoraggio dei consumi elettrici e dell'acqua al fine di individuare politiche convenienti per la riduzione dei consumi stessi.

1. Introduzione

Le aziende che si trovano a dover affrontare una concorrenza sempre più forte e notevoli contrazioni dei margini cercano ogni soluzione possibile per migliorare le proprie performance e aumentare gli utili. Una delle strade spesso inesplorate è la manutenzione degli asset aziendali.

Spesso però non basta avere una gestione computerizzata della manutenzione (CMMS): infatti alle aziende serve aumentare la visibilità delle proprie performance analizzando i dati relativi alle tendenze e alle anomalie principali, eseguendo previsioni sulle problematiche che potrebbero incidere sull'affidabilità e prendendo decisioni sulla base delle prospettive future. Alcune delle funzioni principali necessarie:

- **Ordini di lavoro (ODL)** - pianificazione delle attività, assegnazione del personale, accantonamento dei materiali e registrazione dei costi.
- **Manutenzione preventiva** - monitoraggio delle ispezioni e dei lavori di manutenzione preventiva, anche con istruzioni passo passo o liste di controllo, distinte dei materiali necessari e altre informazioni ad esse collegate.
- **Gestione delle risorse** - registrazione dei dati relativi alle attrezzature e alle proprietà, con indicazione delle specifiche, delle informazioni sulle garanzie, dei contratti di assistenza tecnica, dei pezzi di ricambio, della data di acquisto e della vita utile prevista.
- **Controllo del magazzino** - gestione dei pezzi di ricambio, degli attrezzi e degli altri materiali, compreso l'accantonamento dei materiali stessi.

- **Funzioni avanzate di reporting e analisi** - creazione di report e analisi personalizzate utilizzabili per la previsione delle problematiche che potrebbero verificarsi, in modo da poter intervenire tempestivamente.

Per un'impresa meccanica è altrettanto importante tenere sotto controllo i consumi energetici e possibilmente ridurli. Risorse fondamentali sono luce ed acqua. Per la realizzazione di tutto ciò serve la definizione di una programmazione pluriennale.

E' ciò che ha fatto Tekfor realizzando in pochi anni una rivoluzione interna. La società aveva nome negli anni Ottanta Teksid ed era la Divisione Acciai di Fiat. Poi fu ceduta nel 1999 alla tedesca Neumayer e infine a un gruppo di finanziarie che ha accresciuto il numero di siti produttivi nel mondo.

La sede torinese è specializzata nelle produzioni a caldo, semi caldo e a freddo per la produzione di componentistica in acciaio per motori di automobili come alberi a gomito, pistoni e altre tipologie di parti.

L'esigenza di ottimizzare la produzione ha impattato in buona parte sul reparto che si occupa della manutenzione degli impianti e che occupa decine di persone specialistiche, oltre che a coordinare l'intervento di specialisti dei fornitori delle macchine per la produzione, ad esempio presse, forni e macchine a controllo numerico. Ma la soluzione è stata trovata dal reparto IT che ha realizzato un processo di gestione delle risorse mediante strumenti informatici che hanno permesso di pianificare gli interventi di manutenzione preventiva, coordinare gli interventi, far corrispondere le attività da compiere per la manutenzione agli obiettivi della produzione e, non ultimo, porre a disposizione strumenti che possono anche essere usati dai clienti per verificare la corrispondenza dell'avanzamento della produzione con quanto previsto dagli accordi commerciali siglati con Tekfor.

2. Le gestione della manutenzione degli impianti in Tekfor

Nel 2009, viene avviata l'implementazione della soluzione EAM Enterprise Edition, di Infor, partendo dall'area aziendale delle presse "a caldo", poi si passa alle lavorazioni a freddo e semicaldo.

La parametrizzazione e la creazione dei diversi flussi per il controllo della manutenzione preventiva e la produzione di report con tutte le informazioni dettagliate per il corretto monitoraggio delle diverse fasi del processo di manutenzione sono state le attività più impegnative.

Con l'entrata in funzione di Infor EAM Enterprise Edition, sono stati caricati i dati riguardanti gli impianti con le ore lavorate su ciascun impianto, i guasti e monitoraggi, i particolari utilizzati; oltre alla storia in dettaglio di tutti gli interventi e quindi la possibilità di tracciatura di tutti i componenti dei macchinari. Grazie a tutto ciò si sono potuti stabilire anche i tempi necessari per le sostituzioni dei componenti, le modalità d'intervento per ciascun tipo di macchinario così che ogni differenza da tali parametri possa essere evidenziato, illustrato, motivato e documentato così da arricchire il database di conoscenze di quel macchinario.

Il Recupero Energie in un'azienda manifatturiera

L'adozione di EAM Enterprise Edition si è poi estesa man mano all'intera azienda. I responsabili della produzione oggi immettono nel sistema la richiesta di lavoro per gli interventi manutentivi quando si nota un'anomalia o si riscontra un difetto nella produzione e si è costretti a fermare la macchina. Il responsabile della manutenzione organizza e passa la commessa alla esecuzione. Coloro che intervengono, al termine del lavoro sull'impianto, compilano la maschera con i dettagli: tempo di esecuzione, parti sostituite, commenti ulteriori se per esempio manca un ricambio. Tutte informazioni che arricchiscono la base dati di EAM Enterprise Edition per ciascun macchinario della fabbrica che, inoltre, si sta ampliando con le informazioni sulla documentazione tecnica, gli schemi elettrici e quant'altro nel massimo dettaglio. In seguito è stata avviata l'attività di parametrizzazione e di creazione dei diversi flussi per il controllo della manutenzione preventiva e la produzione di report, con tutte le informazioni dettagliate per il corretto monitoraggio delle diverse fasi del processo di manutenzione.

Dopo la prima fase di caricamento manuale di tutti i dati in Infor EAM Enterprise Edition Tekfor è poi passata all'operatività completa del sistema per vedere le risposte che dava e effettuare il tuning finale con le esigenze aziendali. Il processo e la correlazione con gli altri processi informatici è stato facilitato dal fatto che l'architettura della soluzione Infor EAM è di tipo "event driven" e cioè guidata da eventi, cosa che ne semplifica l'utilizzo in contesti aziendali diversi senza che si debba di volta in volta fare interventi pesanti di riscrittura delle procedure.

La società Net Surfing, partner di Infor, ha sviluppato anche un grande pannello elettronico di sintesi che mostra ai responsabili di manutenzione e produzione l'elenco delle attività da svolgere nella giornata o perché nate durante la notte. Possono così pianificare i lavori e man mano il pannello si colora con lo stato d'avanzamento fino alla fine dell'esecuzione.

Tekfor e Net Surfing assieme stanno ora affrontando una nuova tematica inerente il risparmio energetico. Se i consumi di acqua e energia elettrica non sono monitorati a fine anno ci possono essere sgradevoli sorprese.

Ecco quindi l'inizio di una nuova fase richiesta da Davide Gagnor Project Manager – Ufficio Ingegneria di Manutenzione che coordina il progetto e amministra il sistema Infor EAM per il gruppo Tekfor spa.

3. Una nuova fase progettuale: l'ottimizzazione delle risorse energetiche

Tekfor vuole avere l'opportunità di modificare il proprio comportamento in relazione ai consumi energetici nei propri impianti.

La sfida più difficile riguarda una gestione delle operazioni produttive che contemperino le varie esigenze di fornitura, prezzi, efficienza operativa, immagine aziendale, conformità alle normative e altro ancora.

Una parte dell'energia è impiegata per il funzionamento dell'impianto, ad esempio il riscaldamento, il raffreddamento e l'illuminazione, e non riguarda un particolare processo o operazione. Tuttavia, la maggior parte dell'energia erogata all'impianto è generalmente usata per alimentare macchine, convertire materiali grezzi in prodotti intermedi, generare vapore o facilitare la produzione.

Si può utilizzare meno energia, ad esempio attraverso una schedulazione intelligente della produzione, l'utilizzo di apparecchiature più efficienti o miglioramenti della progettazione. Si può anche utilizzare energia meno costosa, gestendo i punti, le modalità e i tempi di erogazione in modo da concentrarne il consumo quando i costi sono inferiori, come negli orari di minore richiesta.

Tuttavia, l'aspetto più importante, consiste nell'ottimizzare l'utilizzo dell'energia così da conseguire gli obiettivi produttivi nel modo meno costoso e più redditizio, bilanciando allo stesso tempo le molte variabili coinvolte nei processi produttivi. e. Una visione sofisticata di questo genere è impossibile se si considera l'energia semplicemente come costo generale dell'impianto.

Ecco allora il piano in più fasi per l'ottimizzazione energetica.

- **Monitoraggio dell'impianto** – Capire i consumi energetici a livello di impianto per prendere decisioni migliori sui tempi di funzionamento delle apparecchiature
- **Monitoraggio della produzione** – Capire i consumi energetici di ogni macchina nell'area di produzione in tempo reale
- **Controllo** – Implementare l'ottimizzazione automatizzata configurabile della produzione con l'energia come variabile
- **Reazione** – Consentire la reazione ai fattori di mercato esterni per attuare un'ottimizzazione basata sulla fornitura in tempo reale

Con questa metodologia, Tekfor avrà un migliore controllo dei propri impieghi di energia, contribuendo a ridurre l'impatto dei costi energetici sulla redditività e a limitare i rischi associati a fattori esterni che influenzano prezzi e forniture.

Le valutazioni e le verifiche della gestione energetica possono aiutare le aziende a identificare un'ampia gamma di cambiamenti che possono essere introdotti per contribuire alla riduzione dei consumi. Può trattarsi di iniziative semplici, come l'ispezione di una struttura o di un impianto per identificare opportunità rapidamente sfruttabili, oppure di interventi molto più dettagliati, condotti insieme agli altri elementi di questa architettura.

Una volta in possesso di questo quadro generale dell'utilizzo energetico dell'impianto, il personale può identificare e apportare cambiamenti operativi per ridurre i consumi e i costi energetici, ad esempio la riduzione del carico o dell'energia per qualche minuto quando l'impianto si approssima al picco di utilizzo.

Le informazioni raccolte a livello di monitoraggio di impianto aiutano inoltre i produttori a capire e gestire la qualità dell'alimentazione. Attraverso un registro di dati storici, il personale che gestisce l'impianto può individuare i problemi di qualità dell'alimentazione, quali cadute o sbalzi di tensione che possono causare danni alle apparecchiature nell'impianto e problemi di

Il Recupero Energie in un'azienda manifatturiera

rifasamento nella rete elettrica. Conoscendo questi rischi, TEKFOR può proteggere meglio le proprie apparecchiature.

Nell'area di intervento del monitoraggio della produzione, i responsabili dell'impianto possono ricavare informazioni sui consumi. Si sono individuati così i dispositivi da utilizzare e i punti utili di acquisizione dati nelle macchine e nelle linee e la programmazione del sistema informativo per la memorizzazione e l'analisi di tali dati è stata avviata.

4. La soluzione in corso di sviluppo

Quando sarà installato il sistema che estrae informazioni sull'energia dall'area di produzione, Tekfor sarà in grado di separare i dati dei consumi della produzione dai dati dei consumi dell'impianto. In altre parole, sia il management aziendale che quello dell'impianto avranno una visione più chiara della quota esatta dei consumi energetici complessivi dell'azienda assorbita dai processi produttivi rispetto alla quota assorbita dalle funzioni operative quali i data center. Mentre le imprese assegnano tradizionalmente i costi energetici in un impianto in modo uniforme, indipendentemente dalle variazioni nel consumo effettivo, il micromonitoraggio consente di tenere traccia delle spese energetiche e di elaborare proiezioni in base all'uso effettivo in tutto l'impianto.

Ma il fatto più importante è che Tekfor può visualizzare queste informazioni in un dashboard di reportistica, dove i responsabili dell'impianto possono individuare i costi energetici variabili nell'area di produzione e cominciare a considerare i modi per migliorare la redditività.

Con i sistemi di visibilità dei picchi di domanda, un manager può analogamente impedire a un operatore di accendere una macchina a elevato consumo energetico quando l'impianto è prossimo al raggiungimento di un picco di domanda.

Queste conoscenze possono anche aggiungere una nuova dimensione alle equazioni OEE (Overall Equipment Effectiveness, efficienza globale delle apparecchiature) di uso comune, che attualmente tengono conto unicamente della qualità dei prodotti, dei tempi di disponibilità delle apparecchiature e dei tassi di produzione. Attraverso una migliore comprensione dei consumi energetici a livello di impianto, Tekfor potrà modificare i calcoli OEE per includere l'efficienza energetica.

Dopo che i dati dei consumi energetici per la produzione sono memorizzati e analizzati nel sistema informativo, i responsabili dell'impianto possono cominciare a individuare chiaramente gli andamenti dell'utilizzo dell'energia tra vari eventi storici, ad esempio per uno specifico ciclo o lotto, fino a raggiungere una misura affidabile dei consumi di energia a livello unitario di prodotto.

Questa conoscenza consente anche di prendere decisioni aziendali complessive migliori vuole avere l'opportunità di modificare il proprio comportamento in relazione ai consumi energetici nei propri impianti.

5. L'architettura della soluzione

La soluzione in corso di realizzazione si basa sull'utilizzo di contatori sugli impianti per registrare i consumi integrati con la piattaforma di gestione della manutenzione Infor EAM. Su tale piattaforma sono in corso di sviluppo plug-in specifici, vengono configurate ispezioni sulla base dei valori registrati dai contatori e sono sviluppati report specifici

Sugli impianti sono in corso di installazione dei contatori per conteggiare il:

- Consumo elettrico
- Consumo di acqua
- Temperatura

I contatori utilizzati provengono da fornitori diversi: l' Advantech Adam è un apparato configurabile per segnali di ingresso e uscita collegabile a contatori di diverso tipo. Un altro apparato utilizzato è il Ducati Energia.



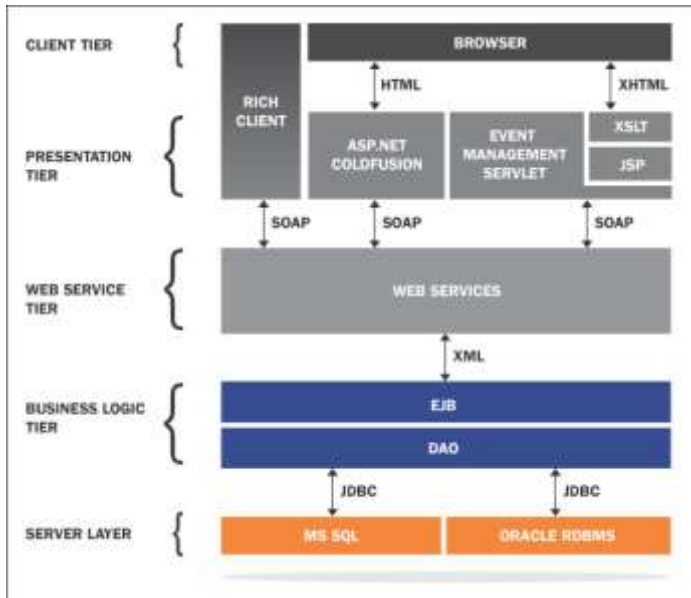
La piattaforma già adottata per la gestione della manutenzione è Infor EAM. Dal punto di vista dell'IT Infor EAM è un'applicazione Internet-based con architettura Web. Infor EAM è accessibile unicamente tramite il browser Web Internet Explorer (versione 6.0 o successive).

L'architettura Service Oriented Architecture (SOA) sta cambiando fundamentalmente il modo in cui un'azienda utilizza l'information technology (IT) per supportare le operazioni di business.

Infor EAM è appunto progettato per indirizzare le esigenze scalabili delle aziende. E' basato sull' industry standard J2EE. La sua architettura service oriented consente ai clienti di costruire una user interface personalizzata con pochi sforzi di codifica e per soluzioni complesse è in grado di operare parallelizzando i servizi.

I componenti architetturali di Infor EAM sono:

Il Recupero Energie in un'azienda manifatturiera



- Infor EAM Client: Computer o Mobile. Un browser cioè una zero-footprint User Interface.

- HTTP Server: riceve una richiesta tramite l' Hypertext Transfer Protocol (HTTP) dal browser e risponde inviando il contenuto

- JSP/Servlet Engine:

Questo componente si occupa di mettere insieme in modo dinamico un documento HTML, che sarà inviato al browser. Tutti gli attributi di presentazione sono applicati su questo livello (es: dimensione del font, immagini, colori, layout dello schermo, ecc.) Non vengono applicate le regole di business qui!

L'engine è costituito da i seguenti blocchi:

- OC4J: questo blocco rappresenta i componenti pdel Servlet Java, eche dirige le richieste di interfaccia utente per i processi di Forms e Report.

- Event Manager: questo blocco rappresenta il cuore del modello Model-View-Controller, gestendo i dati e le richieste di sicurezza del sistema verso Business Manager. Questo componente risolve anche l'azione da intraprendere a seguito di evento scatenato dal browser (es.: scheda fuori di un campo, fare clic su una voce di menu, ecc.) La chiamata distribuita EJB viene eseguita utilizzando Remote Method Invocation (RMI) utilizzando Internet Inter-ORB Protocol (IIOP).

- Tag JSP: questo blocco rappresenta un JSP che crea un frammento XML e lo invia al trasformatore XSLT.

- XSLT trasformatore: questo componente rappresenta un insieme di fogli di stile, che sono user-dipendenti e viene dinamicamente invocato per convertire un risultato XML in un documento HTML.

- EJB Container: Questo componente ha la responsabilità di forzare le regole del business e il meccanismo di aggiungere/modificare/cancellare i dati. Non fa chiamate dirette al database ma interagisce con classi astratte DAO che gestiscono la comunicazione con il DB.

Entity beans: Questo blocco rappresenta gli oggetti istanziati ad un dato momento.

Sessions beans: Questo blocco rappresenta tutte le informazioni di stato delle session che sono necessarie per passare da una chiamata client ad un'altra.

- Java Beans: Questo componente rappresenta un set delle classi Java Bean

DAO: The Data Access Object è un set di classiche collegano l'Oggetto al Relational mapping, e che consente all'EJBs di utilizzare diversi fornitori di database (Oracle o Microsoft). La connessione al DB delle classi DAO avviene tramite i driver Java Database Connectivity (JDBC).

- Cognos: Questo componente rappresenta un set dei Cognos ReportNet Servers

 - Gateway Server
 - Reports Server
 - Content Manager

- Database Server: Questo componente rappresenta il data storage layer

L'architettura rende anche disponibile un Data Access Layer che è responsabile della comunicazione con il DB per consentire lo sviluppo di componenti terze parti. Nasconde agli sviluppatori, la complessità dell'ambiente di accesso ai database. Si occupa anche di presentare un modello uniforme del database astratto, che nasconde accuratamente differenze nel comportamento di Oracle (varie versioni), Microsoft SQL e database DB2.

Uno dei principali vantaggi nell'utilizzare DAL è la capacità di realizzare plug-in di Infor EAM si basa su un set di stored procedure e trigger. DAL permette di accedere a questa funzionalità in modo uniforme, trattando le stored procedure come un oggetto della stessa natura, come tabelle e viste.

La natura DAL light-weight evita un sovraccarico di prestazioni che sono paragonabili a quelle del JDBC standard.

Il framework DAL è utilizzato in combinazione con gli standard entity bean. Con gli Entity bean si possono sviluppare plugin che possono raggiungere alti livelli di scalabilità, utilizzando il meccanismo di caching dei dati dove e quando sia possibile utilizzando le funzionalità di caching avanzate disponibili in tutti gli application server di applicazioni supportate.

Sulla base dell'architettura disponibile Net Surfing sta lavorando in quattro aree specifiche per il monitoraggio dei consumi eccessivi e per la prevenzione guasti dovuti a sovraccarichi di energia.

Il Recupero Energie in un'azienda manifatturiera

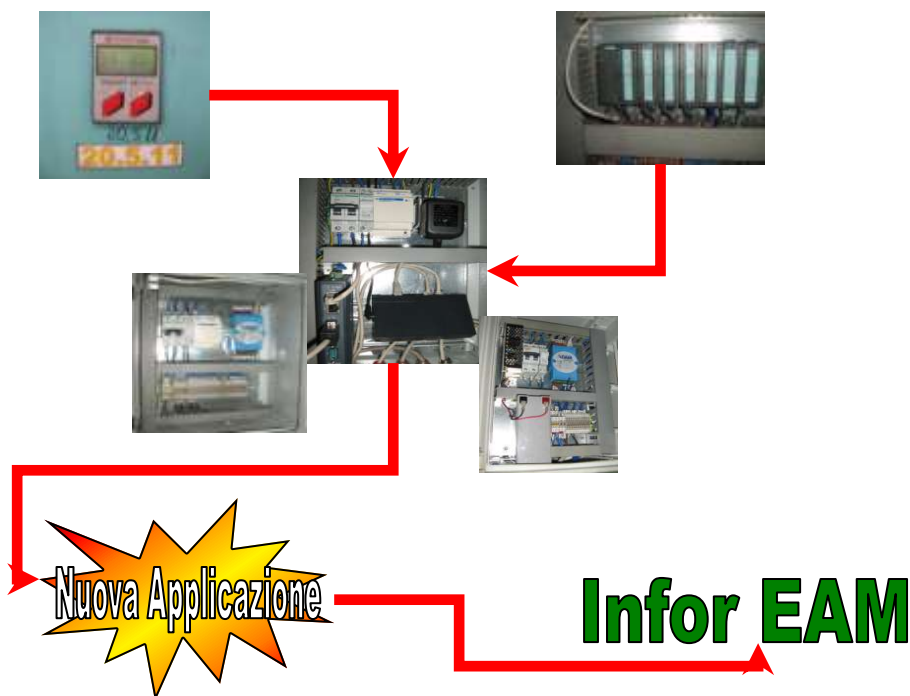
- **Configurazione dei contatori**

E' prevista un'attività di Configurazione dei contatori per gestire gli algoritmi legati alle anomalie, con la possibilità di aggiungere nuovi contatori sugli Asset e di creare delle regole per generare alert .

- **Realizzazione di un plugin per EAM**

Il plugin, realizzato all'interno dell'architettura di INFOR EAM ,acquisisce i dati dei contatori, effettua delle verifiche e se i dati sono anomali, inserisce direttamente un ordine di lavoro (ODL) di manutenzione ispettiva.

La funzionalità, qui semplicemente descritta ,è il cuore dell'attività; si basa sul layer DAL e costituisce la parte concettualmente innovativa del progetto.



- **Ispezioni asset**

La gestione delle ispezioni interagisce con il modulo per la gestione degli ordini di lavoro per offrire maggior controllo sulle funzioni di manutenzioni preventive.

Vengono prima fornite le informazioni sull'ispezioni e quindi le tolleranze alte e basse per le letture dell'ispezione.

Se il risultato di un'ispezione supera la tolleranza alta e bassa, il sistema genera automaticamente un ordine di lavoro correttivo.

Per il monitoraggio dei consumi si configureranno delle ispezioni alle macchine, individuando le formule corrette per la definizione delle tolleranze sia per consumi energetici che di acqua.

Le tolleranze introdotte inizialmente saranno via via affinate con l'esperienza acquisita.



- **Realizzazione di nuovi report per analizzare i dati monitorati**

Saranno realizzati report che evidenziano gli impianti che hanno picchi di consumo, monitorizzandoli nel tempo, grazie alla disponibilità di dati storici.

Sono poi in corso di definizione report di dettaglio che evidenziano i tipi di intervento effettuati, quanti interventi correttivi si sono resi necessari organizzata per data/ora e tipo di impianto.

6. Riferimenti bibliografici

Cominoli F.M., La manutenzione si può anche fare, Pitagora Editrice, Bologna, 2006.

7. Conclusioni

Il progetto dovrebbe concludersi entro fine 2011. La sensibilità di Tekfor al risparmio energetico, combinata ad una gestione accorta dell'IT, consentirà di recuperare energie utilizzando selezionati strumenti e con un investimento relativamente modesto. Una cultura del risparmio diventa così pervasiva senza introdurre stravolgimenti organizzativi e di processo.

Improving business productivity of Knowledge Worker through Social Network Analysis

L. Fortunato, C. Bisconti, A. Corallo

Centro Cultura Innovativa d'Impresa (CCII), University of Salento, Lecce, Italy

Abstract

The role of the knowledge worker has become a key factor to the success of an organization. Nevertheless, knowledge workers perform the business tasks using the tools of knowledge sharing in a de-structured way that do not allow to monitor their productivity. Web 2.0 technologies can have great improvements in the working way of these professional figures, supporting them in the recruitment of the information and in the collaboration with colleagues, avoiding the dispersion of knowledge.

Based on an action research developed in the KPeople project, the paper outlines the use of Social Network Analysis (SNA) to bring out the knowledge of knowledge workers and the flow of information exchanged between them, highlighting the bottlenecks that slow down business processes. This analysis lead to the reorganization and improvement of business processes that could be more easily analyzed and managed, reducing management costs and inefficiencies of the companies.

Keywords

Knowledge worker, Business process management, Knowledge management, Social network analysis

Introduction

Today organizations have to react to many continuous changes of the business world due to economic, technological or consumer-driven requirements.

To compete in this global economy, the companies place more and more emphasis on developing and retaining of their employees and of their human capital.

The corporations are recognizing the importance of their human resources as the principal holders of knowledge.

The human capital is defined as “the recognition that people in organizations and businesses is an important and essential asset who contributes to development and growth, in a similar way to physical assets such as machines and money. The collective attitudes, skills and abilities of people contribute to organizational performance and productivity. Any expenditure in training, development, health and support is an investment, not just an expense” [1].

In an enterprise environment, the “knowledge workers” use their knowledge to carry out the business process with best work performance. The “knowledge worker” is a professional figure that works using his knowledge, information and skills [2]; he makes analysis, defines problems (problem-setting) and identifies solutions (problem solving).

A remarkable difficulty of enterprises is to provide to knowledge workers the tools of knowledge sharing that allow great improvements in their working way and to monitor their productivity.

The intra-organizational knowledge management facilitates the transfer of expertise and the knowledge sharing between knowledge workers in order to make business activities more efficient.

The comprehensive elicitation of knowledge relevant to work processes is a primary objective of Business Process Management (BPM) [3].

Business processes management has emerged as one of the major systematic approaches within organizations to support and optimize their activities and processes, employing methods, policies, metrics, and software tools.

Process management enables organizations to view the progress of the activities in terms of the results, the activities that have yet to be done and the delay respect to the process workflow.

Today, many enterprise have well defined their business processes but hardly these correspond to those that occur in reality. In addition, the knowledge workers often are at the center of numerous processes that have a high degree of complexity and difficulty of management and it is very hard to capture and organize information associate to business processes.

The research project KPeople, carried out by a partnership among the University of Salento (Lecce, Italy), the University of Bari and an Italian computer firm, WebScience, aims to the realization of a collaboration platform supporting knowledge workers in their business processes, facilitating the recruitment of the information and the collaboration with colleagues.

Often the information is retrieved by knowledge worker through informal communication exchange with other knowledge worker which not belong to the traditional organizational hierarchy [4].

“A significant component of a person’s information environment consists of the relationships he or she can tap for various informational needs”... “who you know has a significant impact on what you come to know, as relationships are critical for obtaining information, solving problems and learning how to do your work” [5].

For this reason most organizations recognize the importance of Social Network Analysis (SNA), a powerful diagnostic method used to analyze the nature and pattern of relationships among members of a particular domain. One means of understanding knowledge flows or bottlenecks that slow down business processes, is ‘mapping’ the relationships between employees, with whom they communicate and how often [6].

The Social Network Dashboard (SND), a SNA application developed by Centro Cultura Innovativa d’Impresa (CCII) of University of Salento (Italy), reveals characteristics, composition and structure of existing networks of people, in order to understand how the people work together to respond to business requirements.

This paper aims to obtain a juncture between the management of the business processes, the capture of knowledge flow and the analysis of human relationship, analyzing the integration between Kpeople platform and Social Network Dashboard, in order to obtain a detailed comprehension of business processes through the analysis of human processes.

The paper is structured as follows. In the next section some theoretical definitions of BPM and SNA concept are briefly reviewed in order to outline the starting background of the research activities. Subsequently, the description of the Kpeople research project and the Social Network Dashboard (SND) are undertaken. Finally, some discussions and implications will be presented to understand how SNA application can support the business process and knowledge flow management. The section "Conclusion and future developments" closes the paper.

The role of SNA within BPM

According to Papazoglou [7], a business process is "a set of logically related tasks performed to achieve a well defined business outcome". BPM has presented many definitions over time, which identified the need to enhance processes and allow organizations to operate more efficiently.

Elzinga and Horak [8] state that BPM consists of, "...systematic, structured approaches to analyze, improve, control, and manage processes with the aim of improving the quality of products and services." Zairi [9] describes BPM as, "...a structured approach to analyze and continually improve fundamental activities such as manufacturing, marketing, communications and other major elements of a company's operations".

To capture and model business activities, enterprises use Business Process Modelling tools. The most common notations available in these tool are the Business Process Modelling Notation (BPMN) and the Event-driven Process Chains (EPCs) [10]. With these notation the business activities are inserted within the process workflow and roles and human resources are allocated to the activities. The workflow modeling permits to model employee work processes, attempting to accurately interpret how employees conduct their day to day activities and then find ways to improve them, thus gaining organizational time and consequently cost efficiencies.

But the workflow representation do not actually represents a completely valid association between work task and knowledge workers [11], because often the knowledge workers take care of many other activities that are not required by their process workflow and their relations and interactions often are not related to the tasks that they must perform. There exist differences between workflow prescribed by process and workflows conduct in practice by employees.

With the use of SNA techniques, it is possible to examine the closeness of fit between what managers (or employees) may think employees are doing, and what employees are actually doing.

According to Burt [12], social network is a group of collaborating entities (i.e., actors) that are related to one another. Mathematically, the result of the SNA is a graph wherein each participant in the network is called an actor and depicted as a node in the network. Actors can be persons, organizations, or groups, or any other set of related entities.

Relations between actors are depicted as links between the corresponding nodes. Many social network relations are due to joint participation of actors in business or social activities or membership in collectivities. The common activities create a network of ties among actors. There has been an increased interest in this methodology to analyze the nature and role of informal relationships among individual members in formal business organizations [13, 14].

The increasing adoption of high-speed connections, mobile communication technologies, and real-time collaboration practices has the potential to make knowledge worker communities more effective. With the emergence of the Internet, some of the Web 2.0 applications like wikis and blogs have experienced a rapid growth in recent years. These tools are considered to be highly beneficial applications for supporting collaboration between the knowledge workers because increase and accelerate the knowledge sharing and the ability to work in group.

The most common research methods to measure the human interaction exploit the Web 2.0 technologies by accessing e-mail traffic, phone archives, web site links, online forums, wikis, blogs in addition to a traditional widely used technique of conducting questionnaires with employees or observing them undertaking their tasks in the workplace.

Dynamic social network analysis tools take as input these data and employ text mining, auto-categorization and social network mapping technologies in a unique visual way to discover hidden relationships by mining unstructured data of social networks. Softwares as Ucinet, Jung, Pajek, Condor and Krackplot provide a graphic picture of the relationships of people, teams, and organizations. Moreover they allow the user to create visual maps, movies and adjacency matrices and permit to calculate indicators of collaboration between actors or groups within a communication network [15].

Of late, SNA provides a number of measures, approaches or techniques to use also in business process for evaluating alternative process designs. The workflow mining is an important technique to determine how staff actually undertake their tasks and with whom and to measure workflow communication amongst staff.

The goal of process mining is to extract information about processes from transaction logs. In modern workflow systems, activities are marked with a beginning and end time point; an event log thus retains information on the length of tasks, the time they started and consequently finished and usually some form of identification relating to the personnel who undertook them [16].

The use of SNA in relation to process mining may build a real organizational framework regarding how activities are actually carried out as opposed to how the BPM system represents these [17, 18]. With these techniques, SNA resolutely helps in identifying employees as well as work groups who play central roles in the organization; at the same time SNA is quite capable of discerning holes or bottlenecks in a communication network [19]. Through better understanding of the informal networks existing in the organization, it's possible to know which tasks are effectively undertaken by employees and consequently to better map the business processes.

Hardly there is a one to one mapping between the sociograms and the activities roles defined in EPC or BPMN diagrams. By comparing the results obtained by these two techniques it's possible establish the 'closeness of fit' of knowledge worker to actual versus supposed work tasks [20].

The understanding of knowledge flows in companies means that management is in a better position to determine which knowledge workers are more effective and potentially prevent their loss.

When senior executives acknowledge the networks' importance, they use this dependable tool to be guided in nurturing this human and social capital.

Kpeople project

Kpeople is an industrial research project aimed to support the activities of knowledge workers, increasing their productivity and the ability to create and find the information they need and enabling collaboration with colleagues. To experiment and evaluate new business practices in organizing the tasks assigned to knowledge workers, Kpeople has been deployed, exploiting collaborative Web 2.0 tools, dynamic process composition methods, and semantic engines.

Kpeople system enables workers to improve productivity without changing their habits, while giving management the means to plan and monitor activities through a simple, intuitive web-based application.

The Kpeople system focuses on unstructured and complex processes within an enterprise environment (such as decisional, collaborative, and creative contexts) to improve the management of information and communication and to optimize workspace, recovering the time spent in low-value activities, in particular to find relevant information to execute knowledge tasks, and integrating collaborative workspaces to individual productivity tools (office automation, email, etc.).

These unstructured, or semi-structured processes are a combination of complex activities, based on unspecified procedures, generally performed by individuals or groups of individuals, whose behavior is normally based on the entities that develop the process and on the information exchanged. Unstructured processes are primarily related to the capacity of individual participants, who collaborate actively in the execution of the process, to take decisions in time for the implementation of the process, and for these reasons are hardly predictable.

According to the goals of the project and considering the large number of unstructured processes identifiable within any enterprise, the Kpeople system enables organizations configuring a set of business patterns and supporting the automatic enactment of their workflows.

The pattern is a representative model of a solution to a recurring problem. In the context of business processes, a pattern models a typical and recurrent behaviors and situations, such as delegation, reminder, request for contributions or collaborative decision; the methodology based on the patterns can provide a concrete support to the actors involved in the business processes.

Kpeople introduces this methodology based on patterns in response to two fundamental demands: on the one hand, managing collaboration between various actors who, without a predefined and structured sequence, are expected to work towards completing a particular task; on the other hand, to make the best use of typical Web 2.0 tools within the scope and under the control of the enterprise.

Kpeople uses a flexible, easily configurable, modular design and it is integrated through customized plug-in to the productivity tools usually used by knowledge workers such as communication tools (Mail Clients), legacy Enterprise Systems (ERP, CRM, Business Intelligence Systems), UCC – Unified Communications & Collaboration Systems (VoIP, instant messaging), Document Management System and Web 2.0 facilities.

The Kpeople system was built upon an event-driven architecture which, thanks to custom adapters, is able to trace and store events generated by all enterprise information systems.

Figure 1 shows the Kpeople functional architecture.

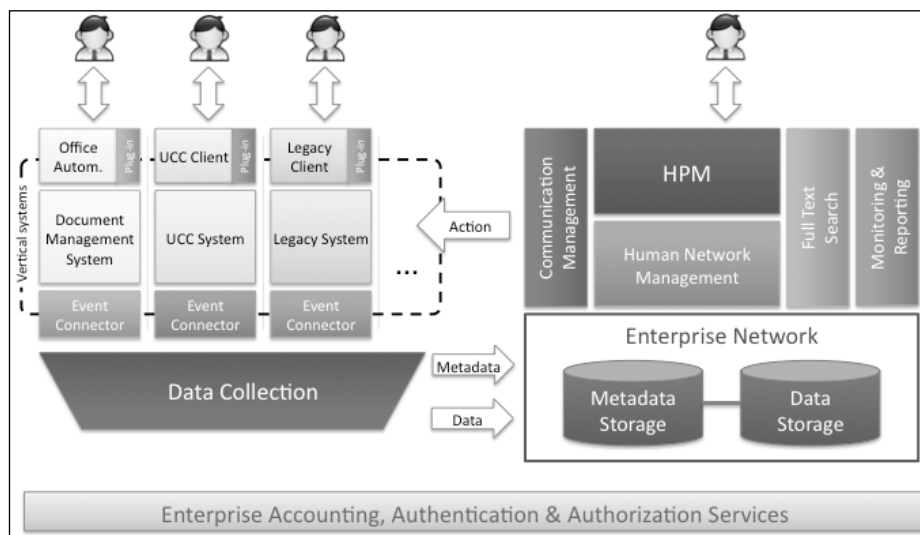


Figure 1: Kpeople System Architecture

Knowledge workers can collaborate with colleagues by exchanging information, files and tasks through the HPM (Human Process Management) tool that allows users to apply patterns and consult the progress of the processes, the activities to be completed, the flow of communication, documents and emails exchanged, and a set of indicators useful to evaluate performances and to identify bottlenecks. All data, information and documents are collected in a common database (Data Storage) enabling easy data retrieval (through Metadata) for knowledge workers and improving their efficiency. Events are connected and clustered using domain ontologies. Event streams may be analyzed by social networks analysis tools.

Figure 2 shows a screenshot of the application of Retrieve Contribution pattern during the configuration phase of Kpeople. Currently the system is being tested by the Italian Association for Computing (AICA) to trace informal activities for new product budgeting, by an Italian large-scale public hospital to manage the deployment lifecycle of internal IT products, by a Brazilian IT company (Elogroup) specialized in BPM tools, by a Hungarian company (John Von Neumann Computer Society) that works in the field of computer driving licensing, and by a Korean company (KPC) working on computer driving licensing too. All users have stated that Kpeople tool is effectively a useful tool for collaborative processes management and information retrieval.

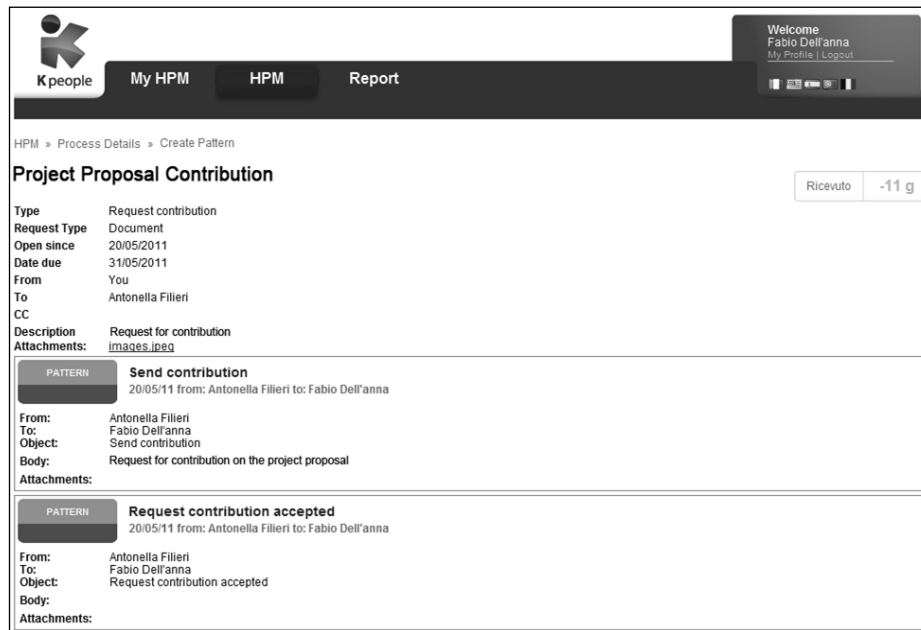


Figure 2: A screenshot of the Kpeople system. The showed interface enables system managers selecting available patterns, and configuring them.

By adding social network analysis functionality to Kpeople tool, the organizations can monitor communications flow between knowledge worker in order to understand the impacts of relationships on their business activities. By identifying and maintaining productive relationships, the organizations could improve the overall performance of activities execution.

The Social Network Dashboard

The Social Network Dashboard is an application designed to aggregate the data existing in a firm, using the techniques of Social Network Analysis. This application allows to understand the relationships established inside a network of companies, entities or others people and performs statistical analysis on the basis of decision-making considerations.

This technology is developed in an environment in which people may easily get in contact with others, allows a quickly analysis of the relationships within the network and at the same time to find useful information for the knowledge worker activities. The application allows users to aggregate all reports and information between people that have already recorded in other applications such as:

- Human Resources Management;
- Management of collaborative relationships within specific projects;
- Management of business skills and opportunities of scouting within company and outside;
- Business process and workflow management;
- Product data management.

The figure 3 shows the architectural model at the base of the Social Network Dashboard.

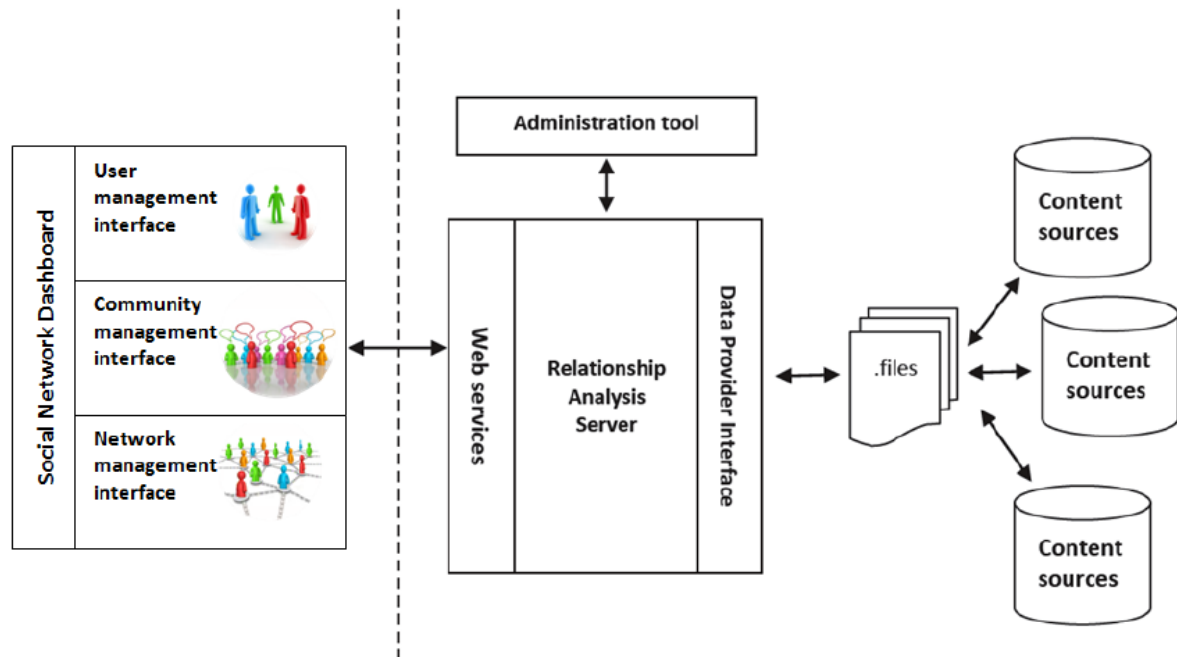


Figure 3: Social Network Dashboard architectural model

The relational analysis server loads the information from various data sources. The administrator uses its own tools to configure the metadata, manage the process of loading, define the level of information security and establish the roles of users for access to data. The analysis server accepts information from any type of systems and it records them in a database-like structure. Then it picks up the information from the database to perform the analysis.

The application consists essentially of the following pages:

- 1) a login page;
- 2) a page for the monitoring of all users including statistics and research tools based on filters;
- 3) a page for the monitoring of all groups and communities;
- 4) a page dedicate to network analysis in which it is possible to visualize the relationships between the actors of a group and define the structural characteristics of a network with the ability to record detailed reports.

A suitable filter system gives valuable information about the dynamics of relationships and "roles" covered by the employees.

SND and Kpeople integration

SND and Kpeople, for what has been said above, are two complementary solutions that together provide all the necessary tools for a detailed analysis of business processes related to human processes.

Kpeople is a BPM solution that captures and reorganizes the information exchanged by knowledge workers and associates these to business processes.

Through the enterprise information tools provided to employees and the tracking of the flow of information, Kpeople is capable of providing all necessary information in order to have a truly representation of relations between employees involved in a workflow process.

SNA Dashboard takes all this information in order to visualize and analyze, using SNA indicators and appropriate mathematical models, the "formal" social network resulting from the business communication.

In addition, SNA Dashboard can analyze all communication channels used by knowledge worker in order to obtain the "informal" network concerning the relations of friendship between the knowledge worker. The resulting graph is the overlapping of the formal enterprise relations related

to the process tasks and the informal relations, so obtaining a truly network at the basis of knowledge worker tasks.

The synergy between the two applications contributes to increase the performance of the organization because it provides the need information for a reverse engineering of business processes and for a more effective management of the knowledge.

SNA Dashboard is designed to be either a module in its own right or a plug-in fully integrated into KPeople application.

Conclusion and future developments

The introduction within the enterprise of technologies for the capture and the sharing of knowledge exchanged by knowledge worker and application tools for the analysis of human communications, can improve the management of the business processes and consequently the overall performance of the organizations. In this paper, the description of integration between KPeople tool and SND application is reported as an example of these useful technologies.

The main goal of this integration is to support the activities of knowledge workers and to manage at best business processes through the monitoring and analysis of social network obtained from the capture of communication exchanges.

The communities and networks are identified as effective environments for the sharing of knowledge and social network analysis provides a valid method to follow this knowledge transfer and to map the real business process workflow.

As future research, the social network data could be modified in order to obtain the structural predictions of future business process changes, useful for encouraging the activities of decision making to improve the performance of knowledge workers activities.

Acknowledgments

The authors wish to thank Webscience s.r.l. (www.webscience.it), Apulia Region and European Community that funded the research as part of the Kpeople project (<http://kpeople.webscience.it>) under the F.E.S.R., P.O. Regione Puglia 2007-2013, Asse I-Linea1.1-Azione 1.1.2. "Aiuti agli Investimenti in Ricerca per le PMI" program.

References

- [1] D. Stockley, 2009, *Human Capital Concept – Definition and Explanation*, Google Website on the Human Capital Management Definition.
- [2] P. F. Drucker, 1973, *Management: Tasks, Responsibilities, Practices*, Harper & Row.
- [3] P. Fettke, 2009, *How Conceptual Modeling is Used*, Communications of the Association for Information Systems.
- [4] K. Chan, J. Liebowitz, 2006, *The synergy of social network analysis and knowledge mapping: a case study*, Int. J. Management and Decision Making.
- [5] R. Cross, A. Parker, 2001, *Knowing What We Know: Supporting Knowledge Creation and Sharing in Social Networks*, Organizational Dynamics.
- [6] P. Busch, 2008, *Tacit Knowledge in Organizational Learning*, IGI-Global Hershey Pennsylvania.
- [7] M. P. Papazoglou, 2003, *Web Services and Business Transactions. World Wide Web: Internet and Web Information Systems*, Kluwer Academic Publishers.
- [8] D. J. Elzinga, T. Horak, 1995, *Business Process Management: Survey and Methodology*, IEEE Transactions on Engineering Management.
- [9] M. Zairi, 1997, *Business process management: a boundary less approach to modern competitiveness*, Business Process Management.
- [10] A.-W. Scheer, 2001, *ARIS – Modellierungsmethoden, Metamodelle, Anwendungen*, Springer-Verlag.
- [11] W. van der Aalst, A. Kumar, 2001, *A Reference Model for Team enabled Workflow Management Systems*, Data and Knowledge Engineering.
- [12] R.S. Burt, 1992, *Structural Holes*, Harvard University Press.
- [13] R. Cross, L. Prusak, 2002, *The people who make organizations go or stop*, Harvard Business Review.
- [14] R. Cross, A. Parker, 2002, *Making invisible work visible: using social network analysis to support strategic collaboration*, California Management Review.
- [15] Hanneman, R., 2002, *Introduction to social network methods*.
- [16] W. van der Aalst, H. Reijers, 2007, *Business process mining: An industrial application*, Information Systems.
- [17] A. Koschmider, M. Song, 2009, *Social Software for Modeling Business Processes*, BPM 2008 Workshops.
- [18] W. van der Aalst, H. Reijers, 2005, *Discovering Social Networks from Event Logs*, Computer Supported Cooperative Work.
- [19] O. Serrat, 2009, *Social Network Analysis*, Knowledge Solutions Asian Development Bank.
- [20] P. Busch, 2010, *Business Process Management, Social Network Analysis and Knowledge Management: A Triangulation of Sorts?*, 21st Australasian Conference on Information Systems.

Integrated mobility of a mobile user

Gianmario Motta¹, Yuwei Yin², Thiago Barroero³, Giovanni Miceli⁴

¹Dipartimento Informatica & Sistemistica, Università di Pavia, Via Ferrata 1, 27100 Pavia (PV), Italia, motta05@unipv.it

²School of Software Engineering, Tongji University, Cao An Gong Lu, Jia Ding Qu, 201808 Shanghai, China, yyw930@gmail.com

³Dipartimento Informatica & Sistemistica, Università di Pavia, Via Ferrata 1, 27100 Pavia (PV), Italia, thiagobarroero@gmail.com

⁴Dipartimento Informatica & Sistemistica, Università di Pavia, Via Ferrata 1, 27100 Pavia (PV), Italia, micelinanni@gmail.com

We present the design of a system that assists consumers in planning, reviewing, monitoring and rescheduling their mobility. Specifically, we address long distance mobility of a frequent traveller. The main novelty is that the system contains an engine that reschedules the consumer's mobility against unanticipated changes. Implementation is on an Android front-end that interacts with transaction processing systems via web services. Potential benefits of this approach look relevant for the customers and for the transportation companies as well. Our system concept is a complementary approach to traffic management systems that support authorities and transportation providers but not the consumer.

1. Introduction: Promise Aware Systems

Promise is a key of services to consumers. When you buy a book on Amazon, you expect that the book shall be delivered within a predefined time, and this expectation is an explicit or implicit element of the contract. Also, when you apply for a government permit, you expect it shall be approved or refused within a given time. We call “promise” such expectation.

In order to fulfill a promise a system should be proactive, i.e. it should be able to manage exceptions and to reschedule resources to overcome the events that undermine the promise. We call “Promise Aware Service System” (PASS) a system that is aware of the promise and are proactive

However, a PASS implies two issues. The first one is to serve an user transaction that involves multiple systems, with web services bridging information islands (e.g. a reservation of journey that involves car renting systems, hotel reservation systems, airline reservation systems). A second issue is to integrate a process, that includes multiple phases performed by different systems of different organizations (e.g. delivering a mortgage loan). Both issues are frequent in e-government, health care, education and alike domains; also they are targeted by European research programs on Internet of

Services [Bitsaki, 2008] [European Commission, 2010]; finally these are a key point in e-commerce transactions as in Amazon (www.amazon.com).

Our paper intends to contribute to define an architecture for PASS by a preliminary research study, that identifies functional requirements and implementation architecture, and shall eventually produce a demo system. The study considers the simple case of a traveler. This approach allows to draw some bottom up assumptions on a possible general PASS paradigm.

2. The traveler case study

Let us consider a simple case. Mr. Rossi lives in Pavia, a small city in Northern Italy, and wants to go to Florence, where his relatives live in a small village not served by public transportation. Now, there are no direct connections from Pavia to Florence. Thus, Rossi shall take the bus to the train station, then a local train to Milan, and, finally, a fast train from Milan to Florence. Rossi books the fast train and alerts relatives. However, things do not happen as planned. Because of traffic jam, the bus is late, but the local train to Milan is late too. Rossi is happy, but when the local train arrives to Milan the fast train has already left. Rossi has to go to the ticket counter and change his ticket. Relatives in Florence pick up Rossi one hour late and have to pay additional parking.

In this case, we can reasonably assume that the promise (P) can be expressed by the duration of the journey. We can also assume that the ability to control exceptions (C) increases the value (V) of to the promise (P): $V = A / P + C$. More precisely, exception control implies three elements: to gather information on actual performance, to analyze deviance and, finally, to undertake appropriate corrections. We call these corrections “compensation”.

In transportation, actual performance is usually gathered by many public systems. In Italy, they include station display panels, station loudspeaker announcements, automatic on board announcing systems, an internet system to display delays (www.viaggiatreneo.it), the general portal www.trenitalia.it where the traveler can book and buy tickets, the call center system, and, finally, the aboard railways staff. However, this information is not structured in one database to be conveyed to the customer nor to customer-linked stakeholders (e.g. relatives) nor to linked actors, (e.g. the renting organization which is waiting for the traveler or the manager the traveler shall meet). Also, the system does not support related compensation actions.

In summary, an ideal PASS for a traveler should:

- Gather real time information from multiple sources
- Merge and analyze information, and display it to the traveler
- Identify appropriate compensation actions

Here below we summarize the functional requirements and the execution architecture of the traveler system.

2.1 Functional Requirements

The journey is conceived as a multi connection path, with a ticket that can be common to multiple travelers. To define the future scenario, we modeled

the flow [Eriksson, 1998] and identified use cases that cover the whole life cycle of a journey:

1. Register
2. Login
3. New journey (= define a new journey)
4. Review plans (= modify journey)
5. Cancel (= delete a stop or the whole journey)
6. Buy (= buy a ticket)
7. Services (= identify and buy side travel services)
8. Status (= visualize the status of the journey)
9. Choose an alternative (the customer chooses among

alternatives proposed by the system)

The key use case is “Choose an alternative” since it implies a compensation as depicted in Figure 1 and Table 1.

Actors	*External system + customer
Description	The customer changes connections in front of external events.
Precondition	The customer has activated monitoring service and an event is happened.
Main Flow	1. The system receives an alert and conveys the information to the customer. 2. The customer goes to Alerts menu in My Trips. 3. The customer clicks on the connection with alerts. 4. The system displays “where, what, time, why” about alert. 5. The customer clicks on the “Other Solutions”. 6. The system displays a list of solutions. 7. The customer select one solution e choose preferences. 8. The system displays the summary. 9. The customer click on “Done” and confirm the change.
Postcondition	The customer can continue his journey continuing to receive information.
Alternative Flow	The customer does not want to change his Journey profile. The customer clicks on “Confirm”. The customer wants another solution. The customer clicks on “Others”.
Exceptions	\
Frequency	Medium
Critical Points	High

Table 1. Choose Alternative Proposal



Figure 1. Choose Alternative Path Mock Up

2.2 Execution Architecture

This system, since it assists almost in real time the traveler, is called Integrated Real Time Mobile Assistant (IRMA). IRMA architecture reflects a web service approach. Thus it includes three components:

1. The IRMA system component, that is responsible for compensation rules and provides a persistence layer and business entities that reflect the domain model.
2. The Android component, that processes information and store local information in a local DB (SQLite)
3. Other service components, that provides information, transaction and miscellaneous services through a service layer

Execution architecture is highly modular, since it uses services and information already existing in the net. Its specific design is the database that structures traveler's needs and, of course, compensation processing. Figure 2 shows the overall architecture of the system.

2.3 Benefits

Benefits of such a system, that increases the value of the service, would be obviously high, from both social and enterprise view points.

The traveler does not spend time to fix issues arising from exceptions (e.g. buy a new ticket, alert relatives) and, moreover, also plans mobility. On the other hand, the system gives a competitive advantage to the enterprise that offers services to the traveler, since it gives much more credibility to the promise and, finally, enables the enterprise to implement a selective policy in compensation actions (e.g. free allowances to gold customers).

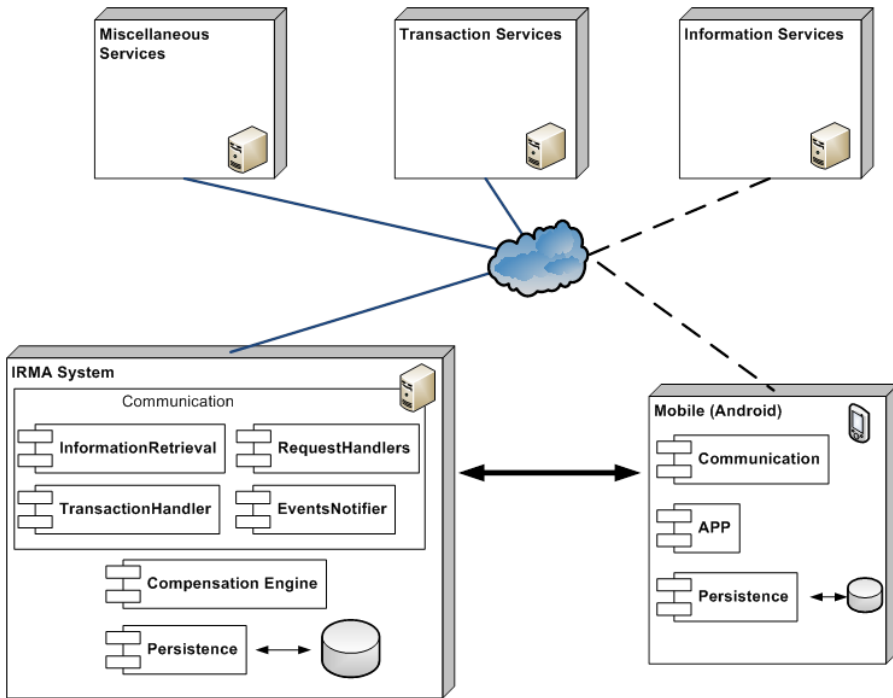


Figure 2. IRMA system overall architecture

2.3 Extensibility

We think that the architecture of IRMA could be used in several cases, namely every time an active user should perform a sequence of activities - more or less as it happens with a finite states machine - and a promise should be maintained. In the table here below we list some potential extension.

Domain	Mobile User	Promise	Source Information Systems
Urban Mobility	Citizen	End-to-end time & cost	Traffic & transportation monitoring
Healthcare	Patient	End-to-end time & cost	Lab & MD agenda
Government	Citizen	End-to-end time & cost	e-Government systems
University agenda	Student	End-to-end time & effort	University curriculum

Table 2. Domain Extensions

3. Discussion

Our case study intends to be a demo a wider concept, that we call Promise Aware Service Systems (PASS). We here discuss some bottom up assumptions on a possible general PASS paradigm, namely (a) promise driven execution (b) performance oriented design (c) execution architecture.

3.1 Promise Driven Execution

Promise driven execution is not a new concept. It is a well known practice of the special domain of Service Level Management systems, where software monitors real time performances of IT “technical services”, e.g. the network [Case, 2007] and IT “business services”, e.g. a seat reservation system [Motta, 2010]. SLM software not only reports execution performances but also fires actions in front of unexpected events, such as incident management.

In short, promise driven execution implies, in front of events or deviances, that the system shall perform appropriate compensation actions. Compensation can be epitomized by the statement:

IF [performance] THEN [action]

The compensation action could be fired either automatically (closed loop) or by the user (open loop). In this latter case, appropriate actions are presented to users, who decide among alternatives. In our specific case study, the compensation is implemented with the following logic. On server side an information retrieval is monitoring events. As soon as an event is detected, the Compensation Engine evaluates the impact on the promise. If the promise is affected, the compensation engine searches alternatives connections and computes the promise of each solution. At this point, the system pushes the solutions through the Events Notifier module on the client and asks to select among alternatives. Based on the selection, the system notifies stakeholders for the new arrival time, and in case of a new connection, buys the tickets for the new connections.

Our solution is, of course, a preliminary ad hoc solution, since it should be designed for each case. Actually, current design approaches do not include promise. To incorporate it, a performance aware language is necessary.

3.2 Performance Oriented Design

Promise, as far as we consider it here, is a business concept that is related to long Running Processes and to complex needs of human users. To incorporate promise, design should start from a clear understating of business context. This already happens with Enterprise Architecture methods. The Open Group Architecture Framework (TOGAF) is a well known framework [The Open Group, 2009], that links Business Architecture, Data Architecture, Application Architecture, Technology [Ackermann, 2003]. A key concept is Business Capabilities, that are delivered by Business Component (BC) [Ernest, 2007]. A BC is a business unit that encompasses a coherent set of activities, supported by assets including people, processes and technology [McDavid, 1999].

A recent research proposes a guide to identify the impact of a new business service on Business Components [Barroero, 2010]. The promise concept can

be considered an operational evolution of the business and/or strategic alignment of IT, and issue that has been popular in the 90's and 00's as shown by Venkatraman's framework [Henderson, 1993] and other recent papers [Reich, 2000].

4. Conclusions & on-going research

By analyzing the implications of a purpose defined case study, we have sketched a design paradigm for systems serving complex user needs that differ from classic approaches in various aspects:

- A traditional Software Engineering approach starts from functional information requirements. Our approach, oriented to business services, starts from business outcome.
- Traditional Software Engineering is conceived to model systems that process short business transactions. Our approach targets long running processes, with Business Process Modeling Notation [Business Process Management Initiative, 2002 & 2004] [OMG, 2004] [Workflow Management Coalition, 2002]
- Classic approaches neglect performance requirements, while our approach is performance aware in architecture, and, eventually, in coding, where we are planning to develop a performance aware extension of a Business Process Execution Language (BPEL) language.

Finally, the case study shows that, in principle, the PASS paradigm can be conceived on the top of existing systems. Existing systems provide services and PASS assist the user in getting the promise.

Of course, the case study is a first step in of our research [Motta 2010]. As a second step of our research program, we plan to implement a demo and to extend our research on performance aware software. Currently we are developing a demo on Android platform, focused on railways services by the next Autumn. The first release will integrate urban mobility. This extension could support visitors and citizens from their arrival by air and train to their life in Milan and vicinity.

References

- Ackermann J., Turowski K., Specification of Customizable Business Components. Euromicro, 29th Euromicro Conference, 2003, 391.
- Barroero T., Motta G., Pignatelli G., Business Capabilities Centric Enterprise Architecture, IFIP Advances in Information and Communication Technology, 326, 2010, 32-43
- Bitsaki, M., Danylevych, O., Van den Heuvel, W.J., Koutras, G., Leymann, F., Mancioppi, M., Nikolaou, C., Papazoglou, M., Model Transformations to Leverage Service Networks. Proceedings of the 4th International Workshop on Engineering Service-Oriented Applications Springer-Verlag, 2008
- Case G., ITIL v3. Continual Service Improvement- Office of Government Commerce, 2007

Eriksson H.E. and Penker M., Business Modeling With UML: Business Patterns at Work, John Wiley & Sons, Inc. New York, NY, USA, 1998

Ernest M., Nisavic J. M., Adding Value to the IT organization with the Component Business Model, IBM System Journal, 46, 3, 2007

European Commission, http://cordis.europa.eu/fp7/ict/ssai/home_en.html, 2010

Henderson, J. C. and N. Venkatraman N., Strategic alignment: leveraging information technology for transforming organizations, IBM Systems Journal, 32, 1, 1993.

McDavid D. W., A Standard for Business Architecture Description, IBM Systems Journal, 38, 1, 1999

Longo A., Motta G., Design processes for sustainable performances: a model and a method, Business Process Management Workshops, Springer, 2006, 399-407

Motta G. , BarroeroT., Telese I., Design of performance aware service systems, The International Joint Conference on Service Sciences (IJCSS), May 25-27, Taipei, Taiwan, 2011

Motta G., Pignatelli G., Longo A., Research issues in designing Services for Quality, Proceedings of the Fourth International Conference on Research Challenges in Information Science (RCIS), 2010, 391-398.

Reich, Blaize.. "Factors That Influence the Social Dimensions of Alignment Between Business and Information Technology Objectives," MIS Quarterly, 24, 1, 2000

The Open Group, TOGAF Version 9. The Open Group Architecture Framework, ISBN 978-90-8753-230-7. Document Number G091. 2009

L'e-Health in Piemonte

Pasquale Marasco, Massimo Berta¹

Fondazione Torino Wireless
Via Francesco Morosini, 19
10128 Torino TO
pasquale.marasco@torinowireless.it

¹ Consulente Fondazione Torino Wireless
Via Biffignandi 37
27029 Vigevano (PV)
info@massimoberta.it

Il presente studio ha lo scopo di realizzare un'analisi del settore della Sanità Elettronica nella Regione Piemonte, prendendo in esame i progetti realizzati sul territorio europeo, quelli concretizzati da altre regioni italiane e le attuali potenzialità del mercato privato in tale settore con riferimento specifico al mondo delle imprese. Lo studio mette in luce le potenzialità che i sistemi di gestione del paziente in modalità deospedalizzata o per monitoraggio a domicilio (telemonitoraggio e teleassistenza), per mezzo di tecnologie di ultima generazione, hanno sui bilanci della Sanità Regionale e sul rapporto, per singole strutture ospedaliere, tra numero di abitanti e posti letto. Risulta fondamentale, per il corretto funzionamento del sistema, il coinvolgimento di tutti gli attori, siano essi enti pubblici, società private, medici o utente/cittadino. In particolare occorre puntare l'attenzione sull'intervento dei medici di medicina generale nei processi di analisi dei risultati.

Riferimenti bibliografici:

1 <http://esante.gouv.fr/>

2 <http://www.sanita.puglia.it/>

3 <http://www.regione.veneto.it/Servizi+alla+Persona/Sanita/>

4 www.regione.lombardia.it

5 <http://www.regione.piemonte.it/sanita/cms/>

6 <http://www.onecare.cup2000.it/>

7 The Criteria Committee of the New York Heart Association. Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels. 9th ed. Boston, Mass: Little, Brown & Co; 1994.

1.Premessa

Il Piemonte rappresenta la terza regione italiana in termini di investimenti pro-capite di tecnologia per la salute. Le linee di intervento di tali investimenti sono da ricercarsi nei benefici che l'ICT può favorire per la cura di pazienti a distanza, come per la gestione ed erogazione di servizi basilari (prenotare una visita o un esame) ma anche nella gestione della spesa in ottica di razionalizzazione dei costi.

Nel contesto europeo, l'e-Health è uno dei sei mercati/aree della "CE Lead Market Initiative" (LMI) ed è considerato un mercato favorevole all'innovazione. L'Europa ha infatti sia un considerevole numero di grandi imprese che di PMI specializzate in soluzioni di e-Health, in particolare sulla telemedicina/assistenza domiciliare e sui sistemi di informazione clinica nel settore dell'assistenza sanitaria primaria.

Nel contesto italiano primeggiano alcune Regioni che attualmente stanno fortemente investendo in progetti di e-Health con un duplice scopo: la riduzione dei costi di gestione del paziente e la riduzione dei posti letto (traguardo 6/4 posti letto per 1.000 abitanti) nel rispetto delle normative nazionali. Regioni come Puglia, Veneto e Lombardia rappresentano casi di eccellenza nell'applicazione di soluzioni di sanità elettronica.

L'e-Health in Piemonte ha visto la nascita di progetti sia in ambito sanitario che sociale. In ambito sociale hanno preso piede sia tipici progetti di telesoccorso sia progetti più evoluti, sviluppati attraverso il supporto della Fondazione Torino Wireless, che comportano, oltre all'invio di un mezzo di soccorso, il monitoraggio ed il controllo di parametri fisici e ambientali legati alla vita del paziente (es. progetti basati sull'utilizzo di strumenti che, al verificarsi di determinate situazioni o al mutare di parametri predefiniti, inviano un allarme automatico alla centrale operativa).

In telemedicina sono stati sviluppati progetti atti a rilevare e trasmettere parametri sanitari clinici per lo più di pazienti affetti da patologie croniche (Servizio di Telemedicina denominato Verbanio-Cusio-Ossola VCO). In un altro caso l'Azienda Ospedaliera Universitaria San Battista di Torino, ha dotato i pazienti del proprio reparto domiciliare di strumenti che consentono il monitoraggio di parametri biometrici mediante l'uso di appositi device cellulari utilizzati come gateway per la trasmissione dei dati.

Partendo da quanto appena descritto, abbiamo verificato la penetrazione di tali servizi in Piemonte ed individuato l'insieme delle principali aziende che propongono sul mercato soluzioni e-Health (hardware, software e servizi); si è voluto quindi formare un cluster di aziende in grado

di offrire un panorama del mercato italiano sia dal punto di vista tecnologico che dei progetti/dimostratori attivi sul territorio nazionale.

Il risultato di tale ricerca è stato quello di fornire indicazioni alla Regione Piemonte su come implementare servizi di e-Health in linea con gli standard del mercato e a prezzi competitivi.

2. Il contesto extraregionale: l'Europa e le altre Regioni italiane

Il recente studio dell'UE sul tema "e-Health strategie nazionali in materia di sorveglianza", conferma che l'Europa è tra i leader mondiali del settore, nonostante la complessità ed il tempo richiesto per la loro diffusione. La complessità dell'e-Health, è stata ampiamente sottovalutata, soprattutto se si considerano gli impatti sulla redditività e sugli equilibri che vengono modificati successivamente alla sua introduzione.

Si possono identificare 4 categorie primarie in cui l'ICT e la Sanità trovano campi di applicazione che permettono di gestire al meglio tutte le cure e i servizi richiesti dal cittadino. In particolare:

1. Servizi di indirizzamento per la gestione della malattia;
2. Attività di supporto alla gestione, amministrazione, logistica e fornitura di beni relativi alla salute e servizi;
3. fornitura di infrastrutture sanitarie della conoscenza, dell'educazione medica e piattaforme di collaborazione;
4. infrastruttura IT per la distribuzione di e-Health (reti di connettività che consentono, ai dispositivi di interagire con i sistemi e le tecnologie), in linea con i provvedimenti e quadri giuridici e regolamentari relativi a questioni di sicurezza e di riservatezza).

Un esempio di implementazione di Telemedicina a livello regionale nel contesto europeo lo si trova in Francia (Bassa Normandia), dove è stata istituita una task-force denominata GCS (Groupement de Coopération Sanitaire¹), che ha l'obiettivo di strutturare, a livello informatico, servizi di Telemedicina.

Il contesto italiano si presenta molto variegato e disomogeneo

La **Regione Puglia**² è da anni fortemente impegnata nel realizzare numerose azioni e progetti di sanità elettronica pienamente in linea con le politiche, le strategie ed i programmi nazionali e comunitari. Citiamo solo alcuni dei progetti attivi: il Portale Regionale della Salute, l'eCUP Puglia, il Sistema Informativo Sanitario Territoriale (SIST), SESIT Puglia, Edotto il Portale Sanitario Elettronico. Un fiore all'occhiello del sistema pugliese è la il servizio di telecardiologia a bordo delle ambulanze, che ha prodotto, nell'arco di 6 anni (2004-2010), risultati notevoli (riduzione della mortalità coronarica del 50% e maggiore appropriatezza delle prestazioni). Tale

servizio ha dato origine alla centrale servizi con la più alta redditività sul territorio nazionale. Un ruolo fondamentale è stato svolto dall'industria farmaceutica che ha finanziato il 100% dello start up del progetto.

La **Regione Veneto**³ ha da tempo messo in atto numerosi progetti ed attività di Sanità Elettronica, anche in ambito europeo, che vedono la Sanità Pubblica evolversi attraverso l'impiego delle tecnologie, come evidenziato dall'Osservatorio Arsenal.IT in uno studio pubblicato nel 2010, che ha evidenziato un forte incremento dei progetti sanitari incentrati sulle componenti tecnologiche. Progetti che sono legati all'ambito Telemedicina.

La **Regione Lombardia**⁴ rivolge da tempo un'attenzione particolare alla telemedicina. L'attuale Piano Socio Sanitario 2010-2014, ritiene terminata con successo la fase di sperimentazione locale dei servizi di telemedicina e si pone come obiettivo di renderli operativi in modo stabile su scala provinciale o regionale. Il piano prevede, per il tramite di servizi di telemedicina operanti su larga scala, "... la necessaria stabilizzazione possibile dei pazienti in stato di cronicità", al fine di prevenire episodi acuti, abbinando ai conseguenti benefici per il cittadino, risparmi per la Regione. Il progetto più importante è il CReG che prevede un modello di cura per i malati cronici per consentirne, anche grazie ai servizi di telemedicina, una gestione domiciliare, riducendo i ricoveri. Con questo progetto la telemedicina consente di passare ad un'assistenza più domiciliare e sul territorio, meno costosa, più efficace e meno disagiata sia per il paziente che per le persone che lo seguono.

3.1 progetti della Regione Piemonte⁵

Il Piemonte rappresenta la terza regione italiana in termini di investimenti pro-capite di tecnologia per la salute; ciò è dovuto proprio ai benefici che l'ICT può favorire nella cura di pazienti a distanza, nella gestione ed erogazione di servizi basilari (prenotare una visita o un esame) ma anche nella gestione della spesa in ottica di razionalizzazione dei costi.

Di seguito, alcuni progetti che seguono le linee di intervento sopra citate e che hanno registrato importanti risultati in termini di soddisfazione dell'utente cittadino⁶.

Il Progetto "**Servizio Monitoraggio Telematico Soggetti Fragili**", ente referente e finanziatore Regione Piemonte (budget di progetto 2.250.000€), ha sviluppato due tipologie di servizi:

- Teleassistenza, servizi audio/video di consulting/compagnia
- Telemonitoraggio patologie croniche, in particolare le cadute e lo stato di immobilità del paziente per periodi prolungati

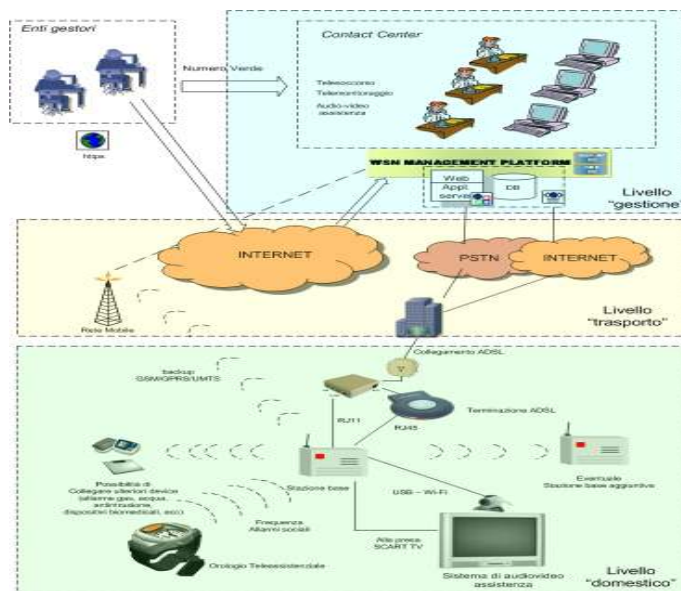


Fig. 1 – Schema Funzionale del dimostratore

Il progetto **“Servizio Telemedicina nel Verbano-Cusio-Ossola”** realizzato dalla Regione Piemonte - Direzione Innovazione, Ricerca ed Università, in collaborazione con la Provincia del Verbano-Cusio-Ossola, l'ASL VCO e il CSI - Piemonte - Direzione Salute, ha sviluppato servizi nell'ambito della teleassistenza e del telemonitoraggio di patologie croniche. Avviato nel gennaio 2009 si concluderà nel dicembre 2011, conta al momento 220 utenti ma punta a servirne circa 500. Costo totale del progetto 2ML€ con un costo medio a carico dell'utente di 220€/mese. Il progetto è rivolto a pazienti cronici che necessitano di assistenza continua, a soggetti in condizione di fragilità, in carico agli Enti gestori dei servizi socio-assistenziali del territorio Piemontese. Il modello di business si basa sull'assunzione che l'assistenza domiciliare dei pazienti cronici comporti sia una riduzione dei costi grazie alla riduzione delle prestazioni improprie (ricoveri, passaggi di pronto, passaggi ambulatoriali), che un miglioramento delle condizioni cliniche del paziente e della sua soddisfazione generale.

Il progetto **“MyDoctor@Home”** ha come referente l'Azienda Ospedaliero - universitaria San Giovanni Battista di Torino e consente il monitoraggio a distanza dei parametri biometrici, mediante l'uso del cellulare come gateway per la raccolta delle misure attraverso Bluetooth ed il successivo invio, mediante rete dati mobile, verso la piattaforma di Telecom Italia. Le funzionalità del progetto possono essere così riassunte:

1. Il servizio permette il monitoraggio a distanza dei parametri biometrici di pazienti con patologie croniche (diabetici, cardiopatici, bronchitici).

2. Il paziente può ricevere, in base ai piani stabiliti dai medici di riferimento, messaggi di promemoria sulle terapie da seguire e sulle misure da effettuare e, utilizzando il cellulare, comunica in modo automatico ad una piattaforma remota il valore dei parametri vitali (ad es. glicemia, pressione del sangue, peso, pressione arteriosa ...).

Il medico curante esegue la refertazione medica delle misure ricevute, supervisiona e controlla l'andamento della terapia ed attiva eventuali azioni correttive (es. variazione terapia, ecc). Elemento cardine per duplicare tale progetto in altri contesti territoriali è il forte commitment dell'Azienda Ospedaliera, in particolare del reparto "a domicilio".

4. La metodologia di analisi

La fase di selezione delle aziende è stata realizzata a seguito di uno screening delle principali realtà italiane e multinazionali e ha visto il coinvolgimento di 22 società che operano su tutto il territorio nazionale. Di seguito la lista delle aziende contattate con l'invio di una "Call for Paper" per partecipare allo studio del settore e-Health per la Regione Piemonte.

1	AEROTEL	12	MEDTRONIC
2	ADITECH	13	MORTARA RANGONI
3	AMD TELEMEDICINE	14	OSPEDALE MONZINO
4	CARD GUARD	15	PHILIPS
5	CARDIO ON LINE EUROPE	16	SIEMENS
6	ELVIA	17	TELBIOS
7	EUROPASSISTANCE	18	INRCA ANCONA
8	HTN	19	TELEMEDICINA RIZZOLI
9	MEDIC4ALL	20	TELEMEDICO
10	MEDICASA	21	TESAN e TBS GROUP
11	MEDITEL	22	FUTURIAINFORMATICA

Fig. 2 - Aziende selezionate per l'Analisi di Benchmarking

Il campione selezionato raccoglie, al suo interno, le principali aree entro le quali si stanno indirizzando gli investimenti sia pubblici che privati.

Nel dettaglio il cluster raggruppa aziende che operano come fornitori di soluzioni hardware, di soluzioni software per la gestione dei processi di raccolta e analisi dei dati in situazioni di paziente remoto o deospedalizzato ed aziende che forniscono servizi sfruttando sia le tecnologie hardware che software.

Gli strumenti utilizzati per l'analisi sono stati principalmente questionari ed interviste dirette con le aziende che, a nostro giudizio, rappresentavano l'eccellenza per alcune aree dell'e-Health nazionale. Come primo passo si è lanciata una "Call for Paper" (avente come punti

centrali la definizione di un panorama tecnologico dell'e-Health e la determinazione di un modello di business sostenibile) alle aziende selezionate e successivamente si è proceduto nella costruzione di 3 tipologie di questionari in base al prodotto/servizio offerto dall'azienda.

Successivamente sono stati inviati, alle aziende che hanno risposto positivamente alla "*Call for Paper*", dei form di compilazione che, oltre a domande di carattere generale, puntavano l'attenzione sulle caratteristiche dell'offerta delle imprese e sui costi e modelli di business applicati. Quest'ultima deriva proprio da una richiesta della Regione Piemonte di individuare le soglie che consentissero, dal punto di vista economico, l'erogazione di servizi che fossero in linea con gli standard qualitativi di mercato ma che avessero il giusto valore economico.

Infine, l'utilizzo delle interviste dirette, con un gruppo ristretto delle aziende selezionate, ha permesso di raccogliere spunti utili da consegnare alla Regione Piemonte come promemoria nella costruzione dei bandi per la sanità elettronica.

4.1 I risultati

I questionari, inviati a 17 aziende, hanno visto una percentuale di risposta di oltre il 70%. Le risposte hanno rivelato che i sistemi sul territorio sono pronti per un'integrazione iniziale della piattaforma e i software attualmente in uso consentono tale integrazione in quanto gestiscono sulla medesima anagrafica eventi di telesoccorso, teleassistenza e telemedicina;

Per quanto riguarda gli aspetti di erogazione del servizio:

- a) *Telesoccorso*: costo medio al giorno di € 0,48/utente;
- b) *Teleassistenza*: costo medio al giorno di 2,00€/ giorno;
- c) *Tele monitoraggio*: costo medio al giorno tra € 4,00/utente ed € 10,00 (dipende dalla gamma di servizi offerti e di parametri rilevati);
- d) *Teleconsulto*: la vastità dei servizi potenzialmente offribili in questo ambito, rende difficoltoso l'individuazione di uno standard;
- e) *Telediagnosi*: servizi on-line offerti al costo di circa € 20 a refertazione immediata. Risposte omogenee.

Per quanto riguarda le soluzioni hardware individuate troviamo:

- soluzioni nel ambito della tele cardiologia e concentratori in grado di fungere da hub per differenti devices di monitoraggio;
- soluzioni rivolte al monitoraggio del benessere fisico;
- soluzioni per device impiantabili nell'ambito cardiologico (defibrillatori, pacemaker e loop) e di controllo del diabete.

5. Conclusioni

La presente analisi nasce dall'esigenza, comune a tutti gli Enti Pubblici, di coniugare la riduzione del costo del comparto sanitario con la necessità di mantenere il livello delle prestazioni erogate ai cittadini.

La strada proposta prevede l'utilizzo della tecnologia per erogare servizi alle persone presso la loro abitazione, facendo muovere i dati ed i tracciati anziché i pazienti, con l'obiettivo di:

- ridurre i giorni di ricovero = riduzione dei costi per la Regione
- ridurre i trasferimenti = riduzione sia dei costi per la Regione che dei disagi per i pazienti
- consentire ai pazienti una maggior permanenza presso la loro abitazione limitando i disagi tipici di ogni ricovero ospedaliero

I diversi pareri raccolti durante le interviste, per i servizi che afferiscono all'area della Telemedicina, Teleassistenza e Teleconsulto, suggeriscono la strutturazione di un Centro Servizi al fine di svincolare le ASL e le Aziende Ospedaliere da carichi di lavoro supplementari; una terza parte in grado di coordinare un'attività di gestione del paziente a domicilio rappresenta un elemento da prendere in considerazione anche in prospettiva della riorganizzazione del sistema sanitario regionale. Accanto a questo modello però, bisogna strutturare un sistema dove la tecnologia sia accentrata ma le competenze/consulti degli specialisti possano provenire da fonti decentrate. Si segnalano due vincoli:

- La dotazione tecnologica di monitoraggio (strumenti per la rilevazione dei parametri, software per l'analisi del dato, cruscotti di lettura) sia presso la centrale che presso il paziente, con l'utilizzo di strumentazioni controllabili a distanza e capaci di interagire con la quasi totalità dei dispositivi hardware presenti sul mercato.
- La formazione del personale, sia esso dislocato presso la centrale sia a diretto contatto con il malato deospedalizzato. La presenza di specialisti (cardiologi, radiologi, medici di medicina generale) consente un pronto intervento ed una tempestiva individuazione di quelle patologie che possono essere causa di ricoveri in strutture.

La struttura ospedaliera può diventare quindi anche centro servizi e fornire un valido contributo medico scientifico in ambito diagnostico.

Un valore aggiunto in tale contesto è rappresentato dalle farmacie che, con la loro capillarità sul territorio, sono in grado di garantire assistenza anche avanzata ai pazienti deospedalizzati ed ai cronici. Alcune delle aziende intervistate vedono le farmacie come un canale di ingresso per la vendita delle loro soluzioni in ottica "win-win" sia per il farmacista, che potrebbe aumentare il suo giro di affari, sia per la sanità pubblica che riduce i ricoveri. La farmacia diventa un "poliambulatorio di monitoraggio virtuale" che raccoglie le informazioni sul paziente per la costruzione di una cartella clinica fruibile da più operatori come medici di base, specialisti, ospedale da cui era stato dimesso. La Telemedicina deve essere vista come una modalità organizzativa che facilita il lavoro dei medici curanti nell'ottica di veicolare alcuni servizi tramite le farmacie, anche in luoghi in cui il servizio pubblico non è presente.

È fondamentale, per il corretto funzionamento del sistema, il coinvolgimento di tutti gli attori del sistema stesso, siano essi enti pubblici, società private, medici o utente/cittadino. In particolare occorre puntare l'attenzione sulla gestione dei medici di medicina generale che mal sopportano un loro coinvolgimento nei processi di analisi dei risultati, anche per una mancata definizione di un compenso extra per il loro consulto.

Accanto ai sistemi di telemedicina, ai fini di un abbattimento delle liste di attesa, nel contesto pugliese si sono rivelati importanti i poliambulatori. Tale compito può esser assolto anche dalle farmacie, che garantiscono la coesistenza di un numero maggiore di punti di contatto tra i cittadini e i medici grazie alla condivisione delle informazioni sul paziente.

Non ritenendo possibile estendere l'erogazione di servizi di telemedicina a tutti i pazienti, si rende necessario individuare un criterio per selezionare a quali cittadini iniziare ad erogare i servizi. Sulla base delle esperienze che abbiamo analizzato proponiamo di coinvolgere in una fase iniziale i pazienti "scompensati". Questa scelta è stata anche determinata dalla disponibilità, per tali patologie, di tecnologie "mature" che abbinano costi accessibili ad un elevato grado di affidabilità. La maggior parte delle organizzazioni intervistate ha suggerito la costituzione di un centro di gestione e refertazione indipendente, soluzione che garantisce equità nei confronti dei pazienti, indipendentemente dalla struttura ospedaliera dove sono stati ricoverati. Tale modello è stato implementato anche dalla Regione Puglia nell'ambito della gestione del servizio 118 regionale, con ottimi risultati, refertando circa 10.000 ECG al mese provenienti dai differenti mezzi di soccorso sanitario, a fronte di una maggior efficienza nei ricoveri.

Per ottimizzare l'erogazione dei servizi di telemedicina domiciliari, i pazienti scompensati possono essere suddivisi in categorie di appartenenza seguendo la classificazione internazionale del New York Heart Association⁷ (NYHA), che prevede 4 classi in funzione dello stato della patologia). Per coloro i quali appartengono alla quarta classe (i casi più gravi), gli esami sono svolti da personale qualificato che si reca al domicilio dotato delle apparecchiature necessarie all'acquisizione dei dati e dei tracciati da inoltrare al centro di refertazione e consulto; per lo svolgimento di queste prestazioni possono essere coinvolte le organizzazioni che già erogano sul territorio servizi di assistenza domiciliare sociale e/o integrata (A.D.I.).

Per gli appartenenti alla prima classe non è previsto un protocollo di intervento specifico. Coloro i quali invece appartengono alle classi seconda e terza, in funzione delle indicazioni mediche vincolanti, si recano a svolgere gli esami previsti presso dei punti di erogazione sul territorio (a titolo di esempio le farmacie").

Il modello di servizi proposto prevede anche la costituzione, all'interno di ospedali specializzati nella cura di patologie cardiovascolari, di Centrali

Operative di Telecardiologia, a supporto della realizzazione di progetti di riabilitazione domiciliare post-cardiochirurgica. In questo caso attraverso valutazioni specialistiche condotte tramite contatti ed esami telefonici, le condizioni cliniche dei pazienti vengono monitorate al fine di cogliere precocemente le alterazioni dello stato di salute, intervenendo tempestivamente con le modifiche terapeutiche del caso. Tale modello è adottato da oltre 10 anni dal Centro Cardiologico Monzino di Milano con ottimi risultati sia sotto l'aspetto sanitario che dal punto di vista dell'ottimizzazione dei giorni di degenza.

Ad eccezione di quest'ultimo caso, in cui è l'ospedale che fornisce al domicilio in comodato d'uso ai propri pazienti la tecnologia necessaria per effettuare i controlli durante il periodo di riabilitazione, nella maggior parte dei casi la tecnologia non rimane nella disponibilità del paziente al suo domicilio. Caso particolare è il progetto MyDoctor@Home dove tutti i macchinari e la strumentazione, di proprietà dell'ospedale, vengono lasciati a casa del paziente. La scelta di non lasciare la tecnologia nella disponibilità del paziente è determinata dai seguenti fattori:

- ottimizzazione del numero di apparati, onde consentirne l'utilizzo a più pazienti
- riduzione dei possibili danni in quanto gli apparati vengono usati da personale professionale
- riduzione della naturale diffidenza da parte dei pazienti verso le nuove tecnologie, poiché l'utilizzo delle stesse è sempre "assistito"
- non è necessario avere le apparecchiature a disposizione 24 ore su 24; in caso di situazioni di emergenza sanitaria, la procedura prevede la chiamata più tempestiva possibile al 118, senza azione alcuna da parte del paziente, infatti le apparecchiature di telemedicina sono omologate come "apparati medici per uso personale a scopo diagnostico".

Un altro aspetto importante per ottimizzare i servizi sul territorio è la costituzione di un'agenzia regionale che abbia i seguenti scopi:

- coordinamento delle iniziative di telemedicina
- erogazione a tutte le strutture territoriali che ne facciano richiesta di dati ed esperienze nell'ambito della telemedicina
- supporto per la stesura di bandi per servizi di telemedicina
- impedire duplicazioni di servizi e/o sperimentazioni sul territorio regionale con conseguente aggravio di costi

Infine nel corso dei colloqui è anche emersa la possibilità di condurre delle sperimentazioni non onerose per l'ente pubblico (a totale carico delle società produttrici della tecnologia), finalizzate a valutare l'impatto economico per la sanità pubblica dell'utilizzo di apparecchiature innovative sia sotto il profilo funzionale che di utilizzo da parte del paziente.

La predizione di flussi di traffico veicolare urbano tramite ensembling di reti neurali

Annunziato¹ M. , Bucci² F. , Meloni¹ C. , Pizzuti¹ S.

¹ENEA

C.R. Casaccia, Via Anguillarese 301, 00123 Roma
{mauro.annunziato, claudia.meloni, stefano.pizzuti}@enea.it

²Università di Roma "La Sapienza"
Piazzale Aldo Moro 5, 00185 Roma
bucci.fabrizio@gmail.com

In questo articolo proponiamo un approccio innovativo per la predizione a breve termine dei flussi di traffico veicolare urbano, basato sul concetto di 'ensembling' di reti neurali e lo abbiamo comparato con gli approcci comunemente usati nel settore. La sperimentazione ha riguardato la predizione ad 1 ora dei flussi reali di tre diverse tipologie di strade urbane ed i risultati dimostrano la validità della metodologia applicata.

1. Introduzione

Oggi giorno ci si sta muovendo sempre di più verso il concetto di 'città intelligenti' (smart cities) [Annunziato, 2011] in cui l'interazione tra le varie componenti della città viene messa a sistema per ottenere ambienti urbani più vivibili ed efficienti. Una delle componenti fondamentali di una smart city è la "Smart Street". Questo è un sistema che poggia sulla rete della illuminazione pubblica di una strada in cui i lampioni sono "intelligenti e multifunzionali", ossia equipaggiati con sensori di varia natura ed attraverso sistemi di comunicazione avanzati (power line ad alta velocità ed interazione web) interagiscono con un sistema intelligente in grado di ritracciare continuamente il profilo di attività (persone, veicoli, emissioni) della strada in base al quale attivare una regolazione adattiva ed automatica per l'intero anno, del flusso luminoso punto-punto ed altre funzionalità. Il sistema si presta particolarmente per importanti arterie veicolari o pedonali cittadine (es: raccordo anulare, strade/piazze principali del centro storico) dove il consumo elettrico per l'illuminazione è particolarmente significativo e la regolazione è critica e permette di salvare grandi quantità di energia. In particolare, i consumi energetici nel settore dell'illuminazione, sia quelli relativi alla pubblica illuminazione sia quelli di competenza delle utenze private, costituiscono una percentuale considerevole dei consumi energetici globali del nostro Paese. Il risparmio energetico e l'inquinamento luminoso sono due argomenti di fortissima attualità, che vedono un tale coinvolgimento degli operatori del campo dell'illuminazione da diventare oggetto di raccomandazioni e normative nazionali ed internazionali. Inoltre il governo italiano ha intrapreso iniziative mirate a promuovere sorgenti luminose

ad alta efficienza e l'impiego di apparecchiature atte a contenere i consumi energetici. Le vigenti disposizioni di legge hanno pertanto stimolato la nascita di nuove tecnologie che permettono di raggiungere le suddette finalità operando sulla accensione-spegnimento, stabilizzazione e regolazione del flusso luminoso e della tensione di alimentazione delle lampade. La maggior parte delle tecnologie citate si configurano come sistemi rigidi e non adattabili tranne, in parte, il caso della telegestione di tipo punto-punto i quali rappresentano il primo passo per la realizzazione di un sistema integrato.

Attualmente si stanno sviluppando delle tecnologie innovative che mirano a superare i limiti mostrati da quelle attuali, che porteranno ad un'illuminazione di tipo adattivo che vede nel "palo intelligente", che potrà integrare servizi smart nella rete di illuminazione pubblica, il suo elemento chiave. L'illuminazione adattiva, essendo basata sul concetto di "energy on demand", necessita di un sistema di controllo e monitoraggio del traffico che sia in grado di estrapolare un indice del flusso veicolare e pedonale. Il sistema di controllo può essere attuato attraverso l'uso di sistemi di misura in tempo reale della domanda ed attraverso l'utilizzo di un sistema di ottimizzazione. Potendo dotarsi di una capacità di modellazione e predizione della evoluzione a breve della domanda stessa sarebbe oltremodo possibile sviluppare un sistema adattivo che evolve in contemporanea alla domanda e generando informazioni sulla domanda stessa di energia. Un'illuminazione di tipo adattivo non solo porterebbe con sé tutti i vantaggi visti in precedenza della telegestione sul singolo punto luce, ma permetterebbe la gestione personalizzata del servizio di pubblica illuminazione consentendo notevoli vantaggi sia sul fronte del risparmio energetico che sui costi di manutenzione, grazie all'eliminazione degli sprechi nelle ore centrali della notte e alla riduzione di interventi "in field" di personale tecnico. Contemporaneamente fornirebbe informazioni utili per il controllo di gestione, misurando l'effettiva resa dei componenti installati e monitorando l'effettuazione dell'intervento e l'efficacia del manutentore. Garantirebbe inoltre una maggiore sicurezza, prevenendo le condizioni di pericolo, un servizio con un elevato livello di qualità, riducendo i reclami ed i tempi di intervento, ed una maggiore tutela dell'ambiente, diminuendo le scorie delle lampade guaste, l'inquinamento luminoso e l'immissione in atmosfera dei gas serra.

Per arrivare a ciò il punto nevralgico è, come già accennato, la capacità di predizione a breve termine di parametri strategici come ad esempio il flusso veicolare con gradi di accuratezza molto elevati. In questo contesto le reti neurali artificiali (RNA) sono una metodologia di modellazione di fenomeni complessi oramai matura che è stata applicata in diversi settori. In particolare, nel settore applicativo del traffico veicolare le RNA sono state utilizzate per la predizione di parametri di traffico [Dougherty e Cobbett,1994] [Taylor e Meldrum, 1995] [Lyons e altri, 1996] [van Lint e altri, 2003] [Annunziato e altri, 2003] [Zheng e altri, 2006] [Çetiner e altri, 2010], nella rilevazione di incidenti [Ritchie e Cheu, 1993] [Cheu e Ritchie,1994] [Ritchie e Cheu, 1995], nel controllo dei comportamenti [Hunt e Lyons, 1994] [Pant e Balakrishnan,1994] [Lyons, 1995], nel controllo del traffico [Nakatsuji, 1994] [Wei e Wang, 2001] e nel monitoraggio [Wan e Dickinson,1992]. In tutte queste applicazioni è stata

sempre applicata la strategia di usare un unico modello globale per descrivere il fenomeno con risultati, in termini di accuratezza, non sempre soddisfacenti.

Per superare gli attuali limiti della strategia con RNA fino ad ora utilizzata in questo settore applicativo, in questo lavoro viene applicata per la prima volta in la strategia dell'ensembling che, invece di usare un unico modello globale, usa e compone tra loro un insieme di modelli.

2. La previsione a breve termine dei flussi di traffico veicolare

Generalmente si parla di previsione a breve termine quando l'orizzonte temporale varia dall'ordine delle ore a quello di una settimana. I dati di flussi di traffico normalmente mostrano una periodicità, spesso più di una, il flusso ad un preciso istante è generalmente dipendente dal valore del flusso nell'ora precedente ma anche nello stesso orario della settimana precedente e così via. Una buona previsione deve essere accurata e, assai importante, presentare un errore massimo il più piccolo possibile. Infatti l'efficacia della gestione dell'illuminazione di una strada, o di una rete stradale, è fortemente influenzata dai picchi di errore ed un predittore con una bassa varianza può essere preferito ad uno che presenta un errore medio inferiore ma picchi di errore più accentuati. Sottostimare la richiesta può avere un impatto negativo sul Demand Response e può rendere il controllo più difficile. Al contrario una sovrastima può creare un surplus inatteso di produzione. In ambedue i casi è chiaro che più alto è l'errore e maggiori sono i costi di gestione, arrivando a situazioni in cui si è costretti a comprare elettricità dalla rete a costi più alti di quanto sarebbero stati grazie ad un'accurata predizione.

In questo contesto, in questa sezione descriviamo le tecniche che sono state utilizzate e comparate per la predizione dei flussi di traffico.

2.1 Metodo 'naive'

Per fornire un confronto significativo tra i vari metodi un modello 'naive' va introdotto per quantificare i miglioramenti introdotti da metodi sicuramente più sofisticati ed "intelligenti". Per dati periodici un modello naive può essere il seguente:

$$X_t = X_{t-S}$$

dove S è il periodo osservabile. Questo modello non fa altro che predire il segnale al tempo t fornendo lo stesso valore osservato esattamente S istanti passati. Nel nostro caso il valore di S è pari a 1, ovvero una ora essendo i dati campionati con frequenza oraria.

2.2 Modellazione statistica

In questo settore applicativo uno dei modelli più comunemente utilizzati è quello del profilo orario medio settimanale. Ovvero, dai dati a disposizione si calcola per i diversi giorni della settimana il valor medio del flusso di traffico ora per ora, ottenendo così un profilo medio di $24 \times 7 = 168$ punti.

2.3 Reti neurali artificiali (RNA) ed ensembling

Le reti neurali artificiali (RNA) sono modelli di calcolo che si ispirano al funzionamento del cervello per risolvere problemi di classificazione, modellazione e predizione di elevata complessità, ovvero dipendenti da un elevato numero di variabili ed altamente non lineari, e pertanto non risolvibili attraverso metodi algoritmici. I primi studi di tali modelli risalgono agli anni '40 con [McCulloch e Pitts, 1943], successivamente negli anni '60 con [Rosenblatt, 1962] e [Minsky e Papert, 1969] e negli anni '80 con [Rumelhart e McClelland, 1986]. Da allora questi modelli sono stati applicati in settori che vanno dall'ingegneria alla sociologia, dalla medicina alla finanza, etc. etc.

L'unità di calcolo elementare di una RNA è il neurone (o nodo), in esso si possono identificare tre componenti principali:

- Un insieme di connessioni in ingresso, ciascuna delle quali viene caratterizzata tramite un peso reale che ne definisce efficacia e tipologia.
- Un sommatore, che calcola la somma pesata dei segnali d'ingresso.
- Una funzione d'attivazione, tipicamente non lineare, che ricevuto come ingresso il segnale calcolato dal sommatore, determina lo stato di uscita del neurone.

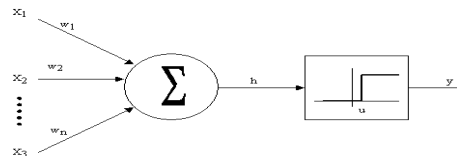


Fig. 1 - Modello di neurone artificiale

In termini matematici, indicando con x_i le componenti del vettore di ingresso, w_i i pesi sulle connessioni entranti del k-esimo neurone, y_k l'uscita del neurone k-esimo, possiamo descrivere il comportamento di un neurone attraverso le seguenti equazioni:

$$y_k = \varphi(\text{net}_k - \theta_k) \quad (1)$$

dove net_k è la combinazione lineare degli ingressi ed è chiamata input netto:

$$\text{net}_k = w^T x = \sum_{j=1}^n w_{jk} x_j \quad (2)$$

Nella (1) abbiamo indicato con θ_k il valore di soglia del k-esimo neurone, detto anche BIAS (Basic Input Attivation System), cioè il grado di "sensibilità" con cui il neurone risponde alle perturbazione dell'input netto e con $\varphi(\cdot)$ la funzione di attivazione. Ogni funzione monotona crescente e continuamente

differenziabile può rientrare tra queste, quella più comunemente utilizzata è il sigmoide:

$$\varphi(x) = \frac{1}{1 + \exp(-x)} \quad (3)$$

L'elemento che caratterizza i diversi tipi di RNA è la topologia delle connessioni dei neuroni. L'architettura più comunemente utilizzata è quella del tipo 'feed-forward'. In tale modello i neuroni sono organizzati a livelli: il primo è detto input layer perché riceve i vettori d'ingresso; poi vi sono uno o più strati nascosti (hidden layer) e uno di uscita (output layer). In tale tipo di rete i neuroni prendono quindi il loro input solo dallo strato precedente ed inviano il loro output solo allo strato seguente. Inoltre neuroni dello stesso strato non sono connessi tra loro. Il meccanismo di determinazione dei pesi delle diverse connessioni viene comunemente indicato addestramento. Tale procedura è a tutti gli effetti un problema di ottimizzazione in cui si richiede di minimizzare l'errore tra il dato reale e quello stimato, ovvero si adotta un meccanismo di 'apprendimento tramite esempi', ovvero un set di dati del fenomeno che si vuole modellare.

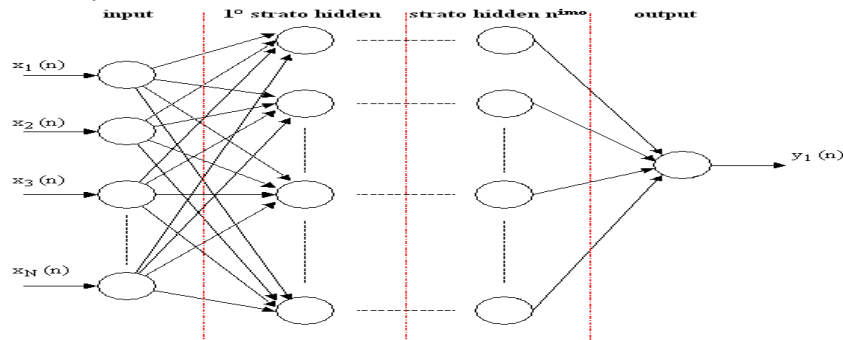


Fig. 2 - Modello di RNA feed-forward

Il concetto generale di 'ensembling' invece è quello di combinare tra loro n modelli in modo tale da ottenere un meta-modello caratterizzato da un'accuratezza superiore rispetto ai singoli modelli:

$Y = f(y_1, \dots, y_n)$, ove y_i è l'uscita dell'i-esimo modello (neurale).

Esistono diversi metodi per far questo [Krogh e Vedelsby, 1995], [Liu e Yao, 1999], [Breiman, 1999], tra questi il più semplice è quello di ottenere il meta-modello come combinazione lineare dei singoli modelli (noto anche come Generalised Ensemble Method - GEM) e come caso particolare fra questi il più semplice (applicato in questo studio) è quello che (noto anche come Basic Ensemble Method - BEM) semplicemente fa la media delle uscite:

$$Y = \frac{1}{n} \sum_{i=1}^n y_i$$

3. Risultati sperimentali

La tipologia di rete neurale utilizzata è feed-forward ad un livello nascosto, con un numero di input pari al numero delle variabili considerate di volta in volta in ingresso, un unico output che rappresenta la stima di consumo energetico di illuminazione e 10 neuroni interni. Le funzioni di attivazione per il livello nascosto e per il livello di uscita sono il sigmoide classico (eq.3), l'algoritmo di training è il BackPropagation con learning rate e momentum adattivi, il criterio di stop è stato fissato al raggiungimento di 100000000 iterazioni.

Il data set si riferisce a dati reali (medie orarie di numero di veicoli) di tre mesi provenienti dalla città di Terni provenienti da 3 diverse strade le cui caratteristiche (secondo il Codice della Strada e il DM 6792 del 05/11/2001) sono riportate nella seguente tabella.

	Descrizione	Portata di servizio per corsia
Strada 1	Strada extraurbana secondaria	600
Strada 2	Strada urbana di scorrimento	950
Strada 3	Strada urbana interquartiere	800

Tab. 1 – Caratteristiche strade

Per sviluppare un modello predittivo neurale il caso più comune da considerare è quello che prende in ingresso i valori passati dell'uscita:

$$x_t = f(x_{t-1}, \dots, x_{t-N})$$

Il numero di nodi di ingresso e quindi il numero di intervalli da considerare come ingresso è una variabile critica per le applicazioni di previsione, poiché un numero troppo basso potrebbe limitare le informazioni disponibili per una previsione accurata mentre un numero troppo alto potrebbe rendere l'addestramento della rete troppo difficile per via della dimensione dello spazio delle soluzioni. E' stata quindi condotta un'analisi preliminare su una strada per determinare il numero ottimale di neuroni di input, ovvero l'ampiezza della storia passata (x_{t-N}). Per questa e tutte le successive elaborazioni il dataset, corrispondente a 3 mesi di misurazioni di valori medi orari di flussi veicolari, è stato partizionato in due sottoinsiemi, uno di training, l'altro di testing. Quest'ultimo è stato composto da 3 settimane, una per ogni mese, e serve a testare le capacità di generalizzazione della rete neurale. Nella tabella seguente vengono riportati i dati sull'accuratezza nella predizione ad un'ora a valle dell'ensembling. L'applicazione di tale metodologia su questa e tutte le successive analisi è stata compiuta su un set di 10 reti neurali.

N (ore)	Errore training	Errore testing
3	4.64%	5.72%
5	3.81%	4.69%
8	3.12%	4.45%

Tabella 2 : analisi finestra temporale input

Da tale analisi si è visto che il modello migliore è quindi quello che ha come input tutti i valori delle precedenti 8 ore e sulla base di questo risultato sono stati condotti tutti gli esperimenti successivi. Nella tabella seguente vengono riportati gli errori medi nominali (ovvero il valore assoluto medio diviso il max-min del range) dei vari metodi messi a confronto. Per le reti neurali accanto all'errore medio (su 10 prove) viene riportata anche la deviazione standard. Inoltre per le reti neurali e per l'ensembling viene riportato tra parentesi anche la differenziazione tra training e testing.

	Naive	Statistico	Reti neurali	Ensembling
Strada 1	8.92%	5.90%	3.74% $\pm$ 0.10% (tr=3.54% $\pm$ 0.12% ts=5.12% $\pm$ 0.26%)	3.29% (tr=3.12%, ts=4.45%)
Strada 2	9.99%	7.14%	4.00% $\pm$ 0.10% (tr=3.81% $\pm$ 0.10% ts=5.22% $\pm$ 0.31%)	3.43% (tr=3.28%, ts=4.37%)
Strada 3	7.66%	5.56%	3.48% $\pm$ 0.09% (tr=3.40% $\pm$ 0.08% ts=3.70% $\pm$ 0.28%)	3.02% (tr=2.98%, ts=3.07%)
Media	8.86%	6.20%	3.74% $\pm$ 0.10%	3.25%

Tabella 3 : comparazione modelli predittivi

Da tali numeri si evince come i metodi neurali siano sensibilmente più precisi dei metodi tradizionali (in particolare statistici). Inoltre, è da notare anche l'apprezzabile miglioramento che si ottiene passando dalle reti neurali al loro ensembling.

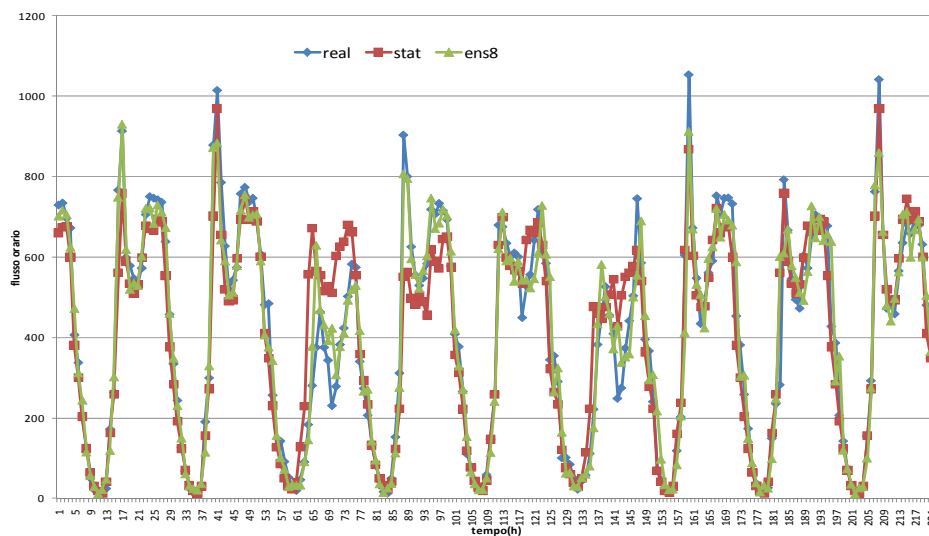


Figura 3 : confronto modelli di predizione strada1

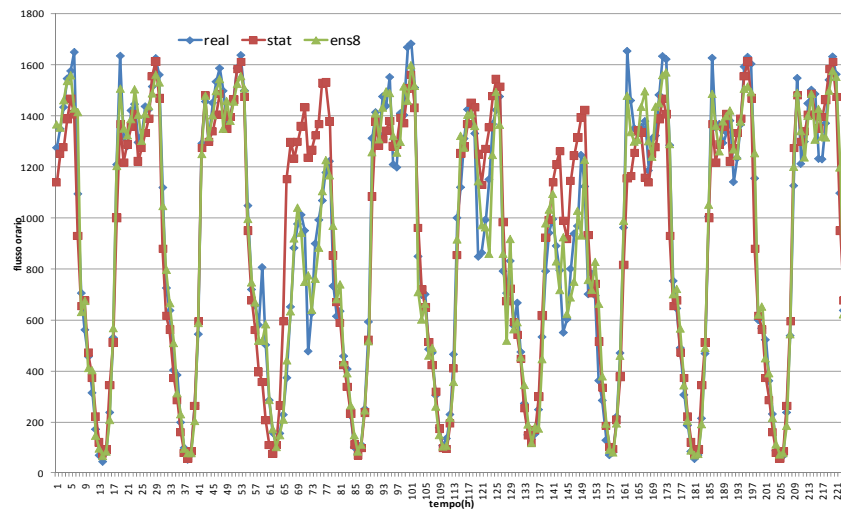


Figura 4 : confronto modelli di predizione strada2

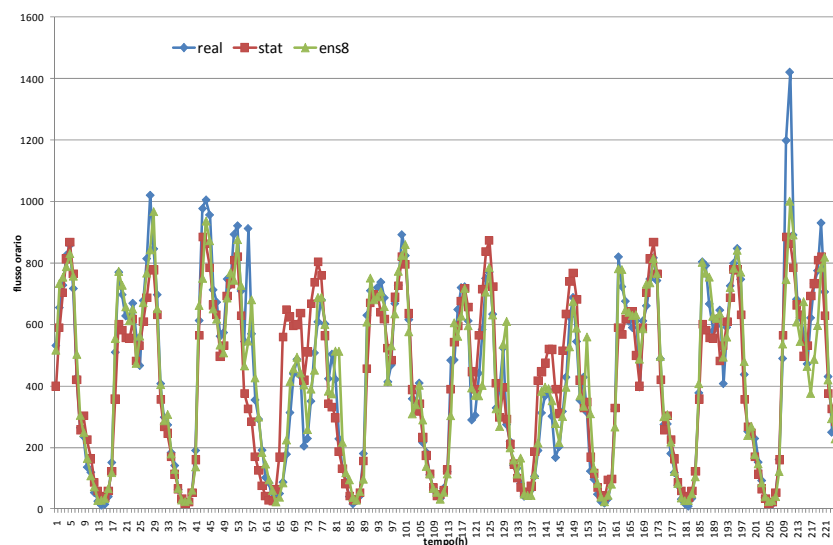


Figura 5 : confronto modelli di predizione strada3

Da tali figure si nota facilmente come la metodologia proposta sia in grado di seguire situazioni anomale (fuori dal caso medio) molto meglio rispetto al metodo statistico. Questo risultato è dovuto al fatto che il modello neurale riesce a catturare la dinamica dei flussi di traffico e quindi è in grado di fornire una predizione sulla base del reale stato corrente di traffico. Questo fatto lo rende quindi molto sensibile anche alle anomalie.

4. Conclusioni

Il lavoro ha affrontato il problema della modellazione di sistemi predittivi a breve termine (1 ora) dei flussi di traffico veicolare. Sono state descritte le metodologie per la modellazione (naive, statistica, reti neurali), quindi sono stati proposti i risultati sulla predizione a 1 ora di flussi di traffico veicolare. La sperimentazione ha prodotto sinteticamente i seguenti risultati : il metodo naive commette un errore del 8.9%, la modellazione statistica il 6.2%, quella dell'ensembling neurale il 3.2%. Tali risultati dimostrano la bontà dell'approccio innovativo sviluppato per la predizione a 1 ora dei flussi di traffico. E' da notare che tale strumento ha un valore scientifico a se stante che lo rende utilizzabile in varie problematiche riguardanti la 'smart street' come ad esempio il controllo adattivo dei semafori e dell'illuminazione.

Bibliografia

Annunziato M. , Le Smart Cities e le sinergie europee del consorzio European Energy Research Alliance, QualEnergia, giugno-luglio 2011, anno IX, n.3, pp.23-26

Annunziato M., Bertini I., Pannicelli A., Pizzuti S. "Evolutionary Feed-Forward Neural Networks for Traffic Prediction", EUROGEN2003, Barcelona, Spagna, Settembre 2003

Arbib M.A., The Handbook of Brain Theory and Neural Networks, The MIT Press, Cambridge (MA),1995.

Çetiner B.G., Sari M. and Borat O. , A NEURAL NETWORK BASED TRAFFIC-FLOW PREDICTION MODEL, Mathematical and Computational Applications, Vol. 15, No. 2, pp. 269-278, 2010

Cheu R.L., and Ritchie S.G., Neural network model for automated detection of lane-blocking freeway incidents. Proc. Int. Conference on Advanced Technologies in Transportation and Traffic Management, Singapore, 245-252 (1994).

Dougherty M.S. and Cobbett M., Short term inter-urban traffic forecasts using neural networks. Proc. Second DRIVE-II Workshop on Short-Term Traffic Forecasting, Delft, The Netherlands, 65-79 (1994).

Haykin S., Neural Networks, a comprehensive foundation (2nd edition), Prentice Hall, New Jersey, 1999.

Hunt J.G. and Lyons G.D., Modelling of dual carriageway lane changing using neural networks, Transportation Research C, 2(4), 231-245 (1994).

Krogh A., Vedelsby J., Neural network ensembles, cross validation and active learning, in Tesauro G., Touretzky D. S., Leen T. K., editors, Advances in Neural Information Processing Systems, 7, 231-238, MIT Press, 1995.

Liu Y., Yao X., Ensemble learning via negative correlation. Neural Networks, 12(10), 1999, 1399-1404.

Lyons G.D., Calibration and validation of a neural network driver decision model, *Traffic Engineering + Control*, 36(1), 10-15 (1995).

Lyons G.D., McDonald M., Hounsell N.B., Williams B., Cheese J., and Radia B., Urban Traffic Management; The Viability of Short Term Congestion Forecasting Using Artificial Neural Networks. *Proc. 24th European Transport Forum, Seminar H(P407)* (1996)

McCulloch W. S., Pitts W., A Logical Calculus of Ideas Immanent in Nervous Activity. *Bull. Mathematical Biophysics*, 5, 1943, 115-133.

Minsky M., Papert S., *An Introduction to Computational Geometry*, MIT Press, Cambridge, Mass, 1969.

Nakatsuji T., et al , Artificial intelligence approach for optimising traffic signal timings on urban road network. *Proc. 5th VNIS Conference*, Yokohama, Japan, 199-202 (1994).

Pant P.D., and Balakrishnan P., Neural Network For Gap Acceptance at Stop-Controlled Intersections, *ASCE Journal of Transportation Engineering*, 120 (3), 432-446 (1994).

Ritchie S.G. and Cheu R.L., Simulation of freeway incident detection using artificial neural networks. *Transportation Research C*, 1(3) 203-217 (1993).

Ritchie S.G., Cheu R.L. and Khan S.I., A comprehensive system for incident detection on freeways and arterials, *Proc. Annual Meeting of ITS America - Intelligent transportation: serving the user through deployment*, Washington D.C., March, 1, 617-622 (1995).

Rosenblatt R., *Principles of Neurodynamics*, Spartan Books, New York, 1962.

Rumelhart D.E., McClelland J. L., *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, MIT Press, Cambridge, Mass, 1986.

Taylor C. and Meldrum D., Freeway traffic data prediction using neural networks. *Proc. 6th VNIS Conference*, Seattle, WA, 225-230 (1995).

Van Lint J.W.C., Hoogendoorn S.P., van Zuylen H.J., Freeway Travel time Prediction with State-Space Neural Networks, *Transportation Research Record* 1811, pp30-39, National Academies Press, Washington D.C. (2003)

Wan C.L., and Dickinson K.W., Computer vision and neural networks for traffic monitoring, *Proc. 6th International Conference on Road Traffic Monitoring and Control*, IEE, London, 143-147 (1992).

Wei W. and Wang M., Traffic Signal Control using Fuzzy and Neural Network, *8th International Conference on Neural Information Processing ICONIP*, Shanghai, China (2001)

Zheng W., Der-Horng L., and Shi Q. , Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach, *JOURNAL OF TRANSPORTATION ENGINEERING*, 114-121 (2006)

La ridondanza come supporto tecnologico al business delle imprese

Carlo Muzzi

AICA - Associazione Italiana per l'Informatica ed il Calcolo Automatico
Piazzale Rodolfo Morandi 2, 20121 Milano
muzzi@acm.org

L'implementazione della ridondanza nei sistemi che automatizzano i molteplici aspetti delle attività umane costituisce oramai esigenza primaria anche nelle infrastrutture a supporto del business delle imprese; in tali contesti è utile disporre di strumenti grafici che assistono le aziende nel comprendere i vantaggi effettivamente ottenibili e valutare consapevolmente gli investimenti da realizzare.

1. Introduzione

Abbiamo già esposto in un precedente Congresso Nazionale AICA [Muzzi, 2009], un formalismo e una metodologia atta a gestire la **ridondanza dei sistemi nel mondo reale**. Questo articolo, che costituisce la naturale evoluzione del precedente, descriverà una metodologia che assiste nell'utilizzo rigoroso del modello formale basato sulla **matrice di ridondanza** utilizzando una semplice rappresentazione grafica – imperniata su quanto definiremo **quadrante di ridondanza** – che permetterà di manipolare visivamente la ridondanza stessa.

In particolare tratteremo di come tale rappresentazione grafica, pur preservando un approccio formalmente rigoroso, faciliterà il lavoro dei progettisti dei sistemi ridondati e permetterà ai committenti (o finanziatori) degli stessi di maggiormente “comprendere” come i conseguenti investimenti supporteranno il business delle proprie imprese.

2. Cenni sulla matrice di ridondanza

La **matrice di ridondanza** [Muzzi, 2009] di un sistema composto da n nodi che interagiscono mutualmente, rappresenta la capacità (o l'incapacità) di un singolo nodo del sistema di sopperire all'eventuale malfunzionamento degli altri nodi.

Per favorire la lettura di quanto tratteremo nell'articolo premettiamo una sintesi degli elementi salienti che costituiscono il modello formale a base della matrice di ridondanza (eventuali dettagli e dimostrazioni potranno approfondirsi nell'articolo già richiamato).

DEFINIZIONE 1 - Per matrice di ridondanza **M** si intende una matrice quadrata di n righe e n colonne rappresentabile dalle notazioni seguenti:

$$\mathbf{M} = \mathbf{M}_{n \times n} = [m_{ij}] = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix} \quad \text{con } n > 1$$

Dove i singoli elementi m_{ij} della matrice rappresentano proprio i livelli di ridondanza reciproca esistenti tra gli n nodi interagenti che costituiscono l'intero sistema; in particolare ogni generico elemento potrà assumere i seguenti valori:

$m_{ij} = 1$ se $i \neq j$ e le funzionalità del nodo i sono replicate dal nodo j

$m_{ij} = 0$ in tutti gli altri casi

DEFINIZIONE 2 - La sequenza degli n elementi costituenti una generica riga i -esima della matrice di ridondanza $\mathbf{M}$ può indicarsi con $r^i = (m_{i1}, m_{i2}, \dots, m_{in})$.

DEFINIZIONE 3 - Con S^i introduciamo una funzione che esprime la somma di tutti gli elementi della riga r^i della matrice $\mathbf{M}$, ossia:

$$S^i = S(r^i) = \sum_{j=1}^n m_{ij} = m_{i1} + m_{i2} + \dots + m_{in}$$

DEFINIZIONE 4 - Con $\{S^i\}$ indichiamo un insieme di n elementi ognuno corrispondente alla funzione S^i di ogni singola riga (o colonna) della matrice $\mathbf{M}$.

PROPOSIZIONE 1 - La sommatoria di tutti i valori $\{S^i\}$ di una matrice di ridondanza $\mathbf{M}$ è un numero intero compreso tra 0 e $n(n-1)$:

$$0 \leq \sum_{i,j=1}^n m_{ij} \leq n(n-1)$$

PROPOSIZIONE 2 - In una matrice di ridondanza $\mathbf{M}$ con tutti gli n nodi **completamente ridondati**, la sommatoria di tutti i valori $\{S^i\}$ è un numero intero pari a $n(n-1)$:

$$\sum_{i,j=1}^n m_{ij} = n(n-1)$$

3. Un modello per rappresentare la ridondanza nei sistemi business-oriented

Abbiamo già richiamato le possibilità offerte dalla matrice di ridondanza come strumento formale per "misurare" l'effettivo grado di ridondanza di un sistema realmente utilizzato nel mondo reale; in particolare dalla PROPOSIZIONE 1 abbiamo visto che in un sistema composto da n nodi il calcolo discendente

La ridondanza come supporto tecnologico al business delle imprese dalla sommatoria degli elementi di $\{S^i\}$ restituisce un numero naturale g compreso tra 0 e $n(n-1)$. Siccome tale valore sintetizza in un solo numero il grado di ridondanza di cui gode un sistema è utile introdurre quanto segue:

DEFINIZIONE 5 - Il *grado di ridondanza* di un sistema di n nodi è il numero $g \in \mathbf{N}$ così definito:

$$g = \sum_{i,j=1}^n m_{ij} \quad \text{dove } 0 \leq g \leq n(n-1)$$

Dalla PROPOSIZIONE 2 possiamo affermare che in un sistema di n nodi tra essi tutti ridondati il grado di ridondanza è massimo ed assume il limite superiore della disequazione $0 \leq g \leq n(n-1)$; denotiamo tale valore come segue: $g^{max} = n(n-1)$.

Vedremo ora come il grado di ridondanza g costituirà l'elemento cardine intorno al quale si svilupperà il modello proposto per rappresentare la ridondanza ed in particolare esamineremo l'importanza assunta dal valore limite g^{max} nella realizzazione di sistemi composti da più nodi tutti ridondati.

3.1 Un caso di riferimento

Per meglio comprendere la trattazione introduciamo un esempio che svolga da caso di riferimento; nello specifico, consideriamo il caso di un'azienda di produzione industriale che assembla una serie cospicua di semilavorati e materie prime per ottenere dei prodotti finiti personalizzati in base alle esigenze puntuali di ogni singolo cliente.

Ai fini della nostra analisi non è necessario precisare la tipologia dei prodotti realizzati dall'azienda; è però importante considerare che il prezzo medio di ogni prodotto è relativamente basso e direttamente rivolto al mercato dell'home consumer. Si è dunque nelle circostanze di dover trattare in modo puntuale una quantità considerevole di singoli ordinativi tutti diversi nelle varie fasi di produzione e distribuzione.

Per una tale azienda lo strumento della vendita on-line è essenziale e strategico perché è il solo che permette di raggiungere un bacino di clientela internazionale sufficiente a garantire la continuità del business; operare in un mercato h24 ha però richiesto una revisione del sistema informativo in modo da ridonarlo adeguatamente.

In sintesi, il team incaricato della revisione del sistema, ha applicato i principi metodologici esposti in [Muzzi, 2009] primariamente identificando i contesti aziendali significativi sui quali agire, specificatamente individuando due aree di potenziale intervento: l'area del sistema gestione degli ordini (essenziale per operare nel mercato di riferimento) e quella del sistema di produzione (essenziale per produrre il bene richiesto da ogni singolo cliente).

Come indicato dalla metodologia gli approfondimenti sulle due aree hanno effettivamente condotto alla prevista **proliferazione degli elementi ridondabili**; è quindi emerso chiaramente che ridondare tutti gli elementi significativi del sistema avrebbe richiesto cospicui investimenti. che avrebbero oltrepassato il budget del committente.

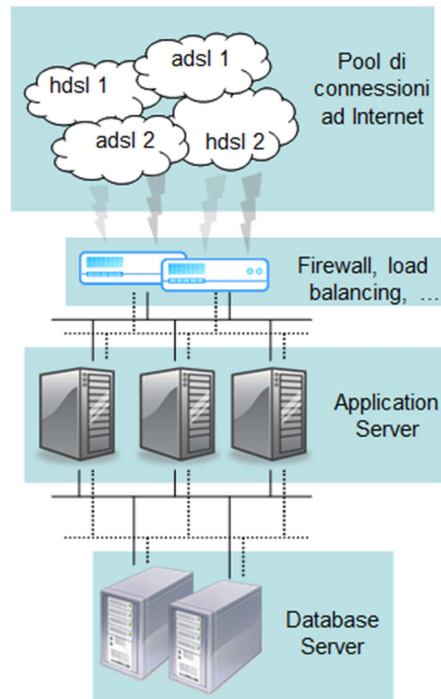


Fig. 1 - La gestione ordini ridondata in un'azienda di produzione

Applicando ancora i principi del modello il team tecnico ha potuto condividere col committente la decisione di **selezionare solo alcuni elementi da effettivamente ridondare**; in particolare è stato deciso di non agire sull'area di produzione (perché avrebbe significato duplicare l'intero stabilimento industriale) ma di limitarsi all'area della gestione ordini: intervento meno oneroso ma in grado di permettere l'effettuazione di azioni commerciali e di marketing su clienti coinvolti in crash di produzione.

La Fig.1 sintetizza la soluzione finale approntata, realizzando un sistema ridondato basato su quattro componenti (partizioni) significative:

- più **application server** per garantire l'operatività in real-time e h24 di una quantità cospicua di clienti sparsi per il mondo per ognuno dei quali occorre una significativa capacità computazionale (la natura del business del committente impone, per la composizione e la trattazione di ogni singolo ordine, anche l'elaborazione di contenuti grafici in tre dimensioni)
- un **pool di connessioni ad Internet** che aggrega accessi di tipo eterogeneo forniti da provider differenti per offrire una connettività sufficientemente alta per soddisfare le esigenze di interazione con la clientela e per fornire quell'accesso alla rete oramai parte essenziale della vita di ogni organizzazione
- **Firewall, load balancing, ...**: necessari per l'aggregazione della banda Internet, la distribuzione del carico sui diversi application server, gli aspetti di

La ridondanza come supporto tecnologico al business delle imprese
sicurezza informatica e la ridondanza della rete interna al sistema degli
ordini (espressa in figura con cammini di rete duplicati)
– l'esigenza assoluta di garantire la gestione di tutti gli ordine dei clienti
richiede un cluster di **Database server**

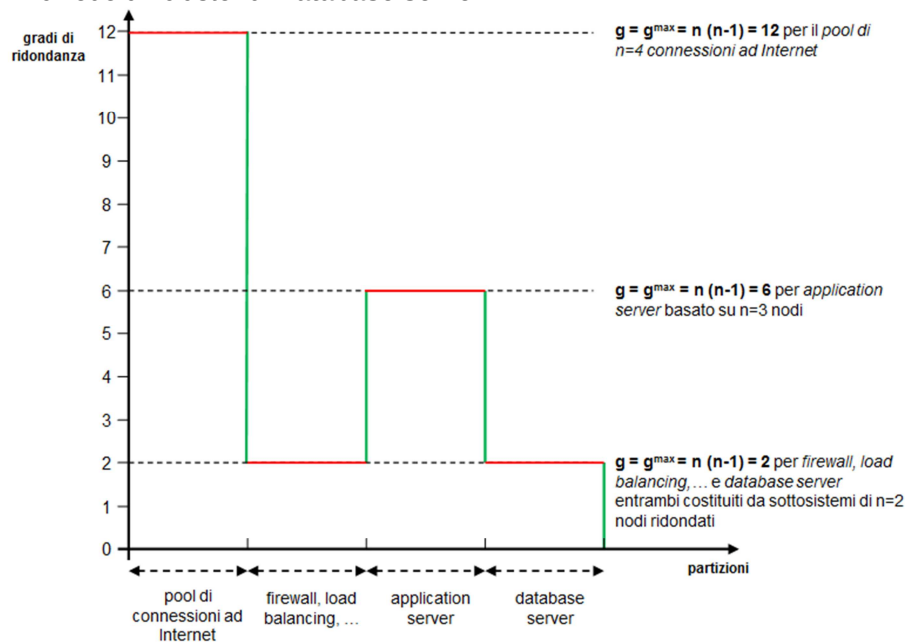


Fig. 2 – Le basi della rappresentazione proposta

Vediamo ora come ognuna delle quattro componenti costituenti tale sistema partizionato possano essere graficamente rappresentate all'interno del sistema cartesiano tracciato in Fig. 2. Ignoriamo, per ora, come suddividere l'asse delle ascisse e limitiamoci semplicemente a suddividerla in 4 intervalli di equa dimensione, ognuno rappresentante una delle 4 partizioni del nostro sistema. Approfondiamo invece la scelta attuata per l'asse delle ordinate; su quest'asse rappresenteremo i diversi gradi di ridondanza (g) di ognuno dei sottosistemi costituenti il nostro sistema partizionato (tutti i di sottosistemi sono completamente ridondati quindi $g = g^{max}$). Rappresenteremo ognuno di questi sottosistemi come una proiezione verso l'asse delle ascisse dell'intervallo rappresentato sull'asse orizzontale fino ad un'altezza che corrisponde al grado di ridondanza dello stesso sottosistema (vedi i segmenti in rosso). Sul grafico così ottenuto procediamo quindi a connettere tutti gli intervalli tracciando degli ulteriori segmenti (in verde nel grafico), tutti paralleli all'asse verticale, che interconnettono il punto finale di un intervallo con il punto di inizio dell'intervallo successivo: il risultato ottenuto è una nuova funzione continua che circoscrive uno spazio piano costituente un sottoinsieme del I quadrante di un sistema di riferimento cartesiano.

Notiamo che tale funzione continua deriva anche dall'aver tracciato le linee verticali di interconnessione; questa aggiunta, che potrebbe anche ritenersi

superflua rispetto alla rappresentazione formale del grado di ridondanza, permette invece di maggiormente apprezzare in termini grafici lo spazio occupato nel quadrante da ogni singolo sottosistema. Le possibilità offerte da quest'arricchimento grafico sono immediate in termini di facilità di analisi soprattutto a supporto di lettori non esperti (come i committenti) in quanto si ottiene una sorta di **grafico ad istogrammi** che fornisce molte utilità aggiuntive: maggiore comprensione dei rapporti esistenti tra i diversi sottosistemi pur enfatizzandosi la specificità di ognuno, rapida implementazione di grafici attraverso le funzioni di supporto dei fogli elettronici, ecc. Naturalmente, non essendovi equivalenza tra le basi di ogni rettangolo, queste utilità non sono quelle tipicamente correlate alle tradizionali analisi sulla distribuzione della frequenza [Tague, 2004] tipiche degli istogrammi classici, ma sono invece proprio quelle insite nel trattamento di rettangoli di larghezze differenti.

3.2 Indipendenza della dimensione dell'asse x rispetto alla y

Se esaminiamo la Fig.2 notiamo che tutti i segmenti in rosso, pur esprimendo il grado di ridondanza dei sottosistema corrispondenti, sono tutti della stessa ampiezza. In effetti, la funzione espressa dal grafico è stata tracciata senza avere la necessità di definire in modo rigoroso quale deve essere la dimensione che l'asse delle ascisse deve rappresentare: pertanto qualunque sia l'ampiezza di ogni singolo intervallo orizzontale il valore **g** associato ad ogni sottosistema rimarrà costante.

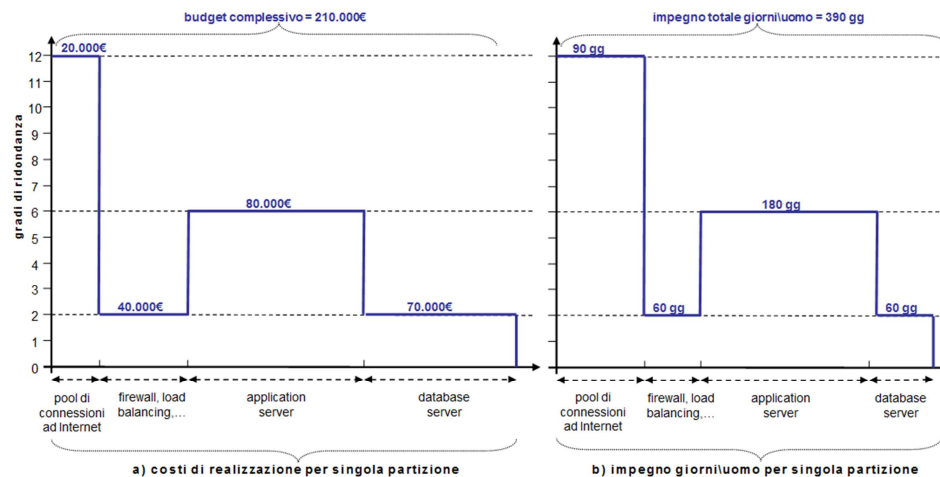


Fig. 3 L'asse orizzontale esprime due possibili dimensioni per lo stesso progetto di ridondanza: a) costi realizzazione b) impegno giornali/uomo per ogni partizione

In altri termini, il grafico proposto continuerà ad esprimere la ridondanza dei sottosistemi **indipendentemente dalla dimensione** scelta per costruire l'ampiezza degli intervalli rappresentati sull'asse orizzontale. È quindi corretto affermare che l'asse delle ascisse potrà rappresentare una qualunque dimensione ritenuta di interesse per il progetto di un sistema ridondato senza

La ridondanza come supporto tecnologico al business delle imprese che ciò infici il significato espresso dall'asse verticale e quindi senza pregiudicare la misura del grado di ridondanza espresso.

Per comprendere appieno la versatilità che discende da una tale indipendenza di dimensione, ipotizziamo che il nostro team di lavoro voglia rendere evidente al committente l'impatto dei costi di realizzazione di ogni singolo sottosistema rispetto al budget complessivo del progetto; in questo caso sarà possibile utilizzare l'asse orizzontale proprio per esprimere una dimensione economica (Fig. 3, grafico a sinistra). Allo stesso modo (Fig. 3, grafico a destra) sarà anche possibile rappresentare l'impegno in giornate uomo per realizzare ogni singola parte del progetto. Questi esempi evidenziano come ogni intervallo non assuma più un'ampiezza equivalente a quella degli altri, bensì una misura proporzionale in funzione del peso che l'intervallo stesso assume rispetto alla variabile scelta per dimensionare l'asse X del grafico.

3.3 Quadrante di ridondanza

Introduciamo ora una definizione formale di quanto esposto con esempi.

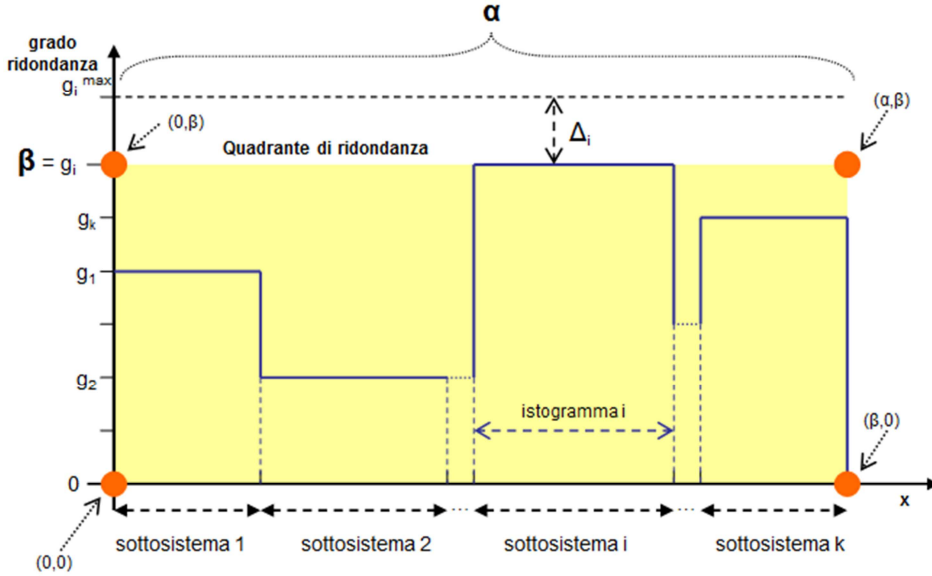


Fig. 5 Un generico quadrante di ridondanza

DEFINIZIONE 6 - Dato un sistema S definito e implementato in modo ridondato attraverso il suo partizionamento in numero finito k di sottosistemi ($1 \leq k, k \in \mathbf{N}$), indicati con g_i il grado di ridondanza dell' i -simo sottosistema e con $w_i(x) \in \mathbf{R}$ il valore che rappresenta il peso assunto sull'asse x da ognuno dei sottosistemi rispetto ad un generico indice di riferimento appositamente scelto per comparare il grado di ridondanza di un sistema, denotiamo con:

$$\alpha = \sum_{i=1}^n w_i(x) \quad \beta = \text{Max} \{g_1, g_2, \dots, g_k\}$$

Definiamo *quadrante di ridondanza* la sezione del I quadrante di un sistema di riferimento cartesiano delimitata dai punti $(0,0)$, $(\alpha,0)$, (α,β) , $(0,\beta)$ dove: i) il segmento dell'asse delle ascisse delimitato dai punti $(0,0)$, $(\alpha,0)$ è suddiviso in k intervalli contigui e disgiunti dove l' i -simo segmento ha ampiezza pari al valore $w_i(x)$; ii) per ogni intervallo i -simo esistente sull'asse orizzontale e delimitato dai punti $(a,0)$, $(b,0)$ si traccia un intervallo parallelo delimitato dai punti (a, g_i) , (b, g_i) ; iii) si traccia un segmento parallelo all'asse verticale che connette il punto finale dell'intervallo i -simo con quello iniziale dell'intervallo $i+1$ -simo

La Fig. 5 sintetizza la rappresentazione grafica di un generico quadrante di ridondanza che è imperniato sulle quattro coppie di punti $(0,0)$, $(\alpha,0)$, (α,β) , $(0,\beta)$ che derivano dal calcolo dei valori α e β . Da notare che nella figura sono presenti degli intervalli che non sono continui (punteggiati) perché rappresentanti i $k-4$ intervalli non indicati nella figura per i vincoli tipici delle rappresentazioni generaliste (nei casi reali i sottosistemi saranno sempre un numero finito e quindi troveranno piena rappresentazione nella figura).

3.4 Altezza limite di un quadrante di ridondanza

Esiste un limite all'altezza massima che ogni intervallo può assumere in un generico quadrante di ridondanza:

PROPOSIZIONE 3 - Per un qualunque intervallo i presente in un generico quadrante di ridondanza, il limite superiore che tale intervallo può assumere rispetto all'asse verticale si ottiene per $g_i = g_i^{\max}$.

DIMOSTRAZIONE. Per il punto ii) della definizione del quadrante di ridondanza, abbiamo che tutti i punti che formano il segmento costituente l'intervallo i -simo (corrispondente al sottosistema i) si trovano alla medesima altezza g_i . Richiamando quanto già discusso in precedenza (DEFINIZIONE 5), sappiamo che $0 \leq g_i \leq n(n-1) = g_i^{\max}$, ne consegue che il limite superiore per g_i si ottiene proprio quando coincide con $g_i^{\max}$. ■

Per comprendere il significato grafico della PROPOSIZIONE 3 riferiamoci nuovamente alla Fig. 5: l'obiettivo di massima ridondanza raggiungibile nel progetto di un sistema ridondato si ottiene quando tutti gli intervalli Δ_i tendono a zero; in termini più formali: $g_i^{\max} = \lim_{\Delta_i \rightarrow 0} g_i$.

Questo limite è particolarmente rilevante nella realizzazione di sistemi ridondati dedicati al core-business delle imprese; in questi contesti, nei quali il raffronto costi e benefici è molto significativo, esso esprime il limite oltre il quale ulteriori investimenti sulla ridondanza **rischiamo di essere sovradimensionati**. Ad esempio, la Fig. 2 evidenzia come già con 2 nodi in cluster il database raggiunga la ridondanza; aumentando a 3 i nodi aumenterebbe anche il livello di resilienza complessiva del sistema che reggerebbe anche al rischio del crash di più nodi; considerando però che il rischio di doppio crash è basso il maggiore investimento potrebbe non sfruttarsi mai. Progettisti e committenti potrebbero dunque scegliere di fermare la ridondanza a 2 nodi o accettare ulteriori

La ridondanza come supporto tecnologico al business delle imprese investimenti a fronte di altri tangibili vantaggi che vanno oltre la ridondanza: ad esempio spostandosi da un cluster di tipo attivo-passivo (in cui solo un nodo lavora) verso uno di tipo attivo-attivo (in cui tutti i nodi funzionanti operano sempre in parallelo) [Feuerlicht e Ates, 2005] che aumentano la capacità di calcolo e la produttività potenziale dell'impresa.

3.5 Rappresentazioni canoniche

Le dimensioni che possono essere rappresentate sull'asse orizzontale di ogni quadrante di ridondanza è potenzialmente elevato; analisi euristiche hanno però evidenziato che vi sono alcuni quadranti specifici particolarmente utili a progettisti e committenti, sono quelle che definiamo come **rappresentazioni canoniche** tra cui evidenziamo: la **banale** utilizzata per elencare i sottosistemi di un sistema ridondato e quindi con intervallo di equa dimensione (Fig. 3); i **costi di messa realizzazione** (Fig. 3a) o i **costi di manutenzione** (es. i tipici costi di maintenance dei sistemi informatici [Swanson, 1976]); l'**impegno delle risorse umane** (Fig. 3b) dedicate ai singoli sottosistemi che, insieme ai **tempi ai approvvigionamento delle risorse hardware e software**, permette di ottenere i **tempi di attivazione** dell'intero sistema (in cui si tiene conto anche della possibilità di svolgere in parallelo alcune attività); la metrica **green** utilizzata per misurare l'impatto che ogni sottosistema ha sull'ambiente durante il proprio ciclo di vita, ad esempio per misurare l'impatto delle misure di riduzione delle emissioni di CO2 su specifici sistemi effettuate per far fronte alle politiche di riduzione dell'effetto serra [Auriemma e D'Ercole, 2007]; infine quella **tipica di un processo** che rappresenta i processi di business dedicati che caratterizzano ogni specifica organizzazione.

4. Conclusioni

Abbiamo discusso di come il quadrante di ridondanza costituisca un espressivo mezzo grafico utilizzabile per facilitare la comprensione e la trattazione delle tematiche che la matrice di ridondanza esprime in termini più formali e così permettere ai soggetti (team specialistici e committenti) coinvolti nei progetti di ridondanza di assumere con più cognizione le conseguenti decisioni.

Riteniamo che tale strumento sia particolarmente utile in un contesto di impresa sempre più impegnata in una competizione internazionale che richiede di affrontare con consapevolezza i rischi cui si è soggetti nella propria attività [Bertozzi et al, 2010], particolarmente tenendo conto che l'adozione dell'approccio ridondato come strumento per la mitigazione o la rimozione dei rischi stessi costituisce una leva richiesta da diverse normative e regolamentazioni nazionali ed internazionali.

In effetti l'opportunità di applicare la ridondanza emerge nello standard [ISO/IEC 27002:2005, 2007] quando viene indicato di considerare l'integrità, la riservatezza e la disponibilità come caratteristiche tipiche della sicurezza delle informazioni; così come la si ritrova anche nella corretta prassi per la gestione dei rischi bancari [Comitato di Basilea per la vigilanza bancaria, 2004] laddove si rappresenta l'esigenza di minimizzare le perdite predisponendo piani di

continuità operativa che garantiscano la prosecuzione delle attività in caso di gravi disfunzioni operative.

Ancora, l'adozione del quadrante di ridondanza è da ritenersi utile perché la rappresentazione grafica, assistendo le organizzazioni nell'assumere con maggior cognizione il grado di ridondanza da implementare, aiuta a meglio comprendere il grado di rischio accettato nelle proprie attività; pertanto, soprattutto nei progetti di sistemi complessi, il modello grafico esposto favorisce la comprensione e quindi la cooperazione tra i vari attori coinvolti riducendo conseguentemente quei cosiddetti rischi di comunicazione contro i quali sono state anche proposte [Covello e Allen, 1998] specifiche regole metodologiche e comportamentali.

Bibliografia

Auriemma M, D'Ercole G., L'impatto delle misure di riduzione delle emissioni di CO₂ sull'attività economica e il lavoro in Europa, Rapporto nazionale finale – Italia, Direzione generale Ambiente della Commissione europea, Bruxelles 2007.

Bertozzi F., Cassia L., De Massis A., La gestione dei rischi operativi nell'impresa - Implicazioni manageriali e impatti organizzativi dall'esperienza progettuale in alcuni casi reali. *Sistemi & Impresa*, 55, 2, 2010, 26-37.

Comitato di Basilea per la vigilanza bancaria, Prassi corrette per la gestione e il controllo del rischio operativo – Febbraio 2003, Banca dei Regolamenti Internazionali, Basilea, 2004.

Covello V.T., Allen H.F., Seven Cardinal Rules of Risk Communication, U.S. Environmental Protection Agency, Washington DC, 1988.

Feuerlicht G., Ates E., Virtualization of Enterprise Data and Services, Proceedings of the 13th International Conference on Systems Integration 2005, Praga, 2005, 305-312.

ISO/IEC 27002:2005, Information technology - Security techniques - Code of practice for information security management, International Organization for Standardization & International Electrotechnical Commission, Svizzera, 2007.

Muzzì C., Ridondanza dei sistemi nel mondo reale, Atti del 47° Congresso Nazionale AICA, Roma, 2009.

Swanson E.B., The Dimension of Maintenance, Proceedings of the 2° IEEE International Conference on Software Engineering, San Francisco, 1976, 492-497.

Tague N.R., The Quality Toolbox - second edition, ASQ Quality Press, USA, 2004.

La scuola italiana segue l'Europa
Eucip Core nei curricula dell'Istituto Tecnico
"Amministrazione Finanza e Marketing,
articolazione Sistemi Informativi Aziendali"

Adriana Fasulo docente di informatica
e i docenti dei Consigli di Classe delle
Classi III e IV dell'Indirizzo Mercurio
dell'Istituto Tecnico Commerciale Statale e per Geometri
'E. Fermi' Pontedera (Pi)
Via Firenze, 51
e-mail: adriana.fasulo@libero.it
e-mail: mail@itcgfermi.it

L'intera istituzione scolastica e in particolare gli istituti tecnici si trovano in questo momento nella necessità di ridefinire e declinare i propri obiettivi, i programmi e le metodologie di insegnamento a partire dai cambiamenti previsti dalla riforma degli istituti tecnici .

Agli istituti tecnici è affidato il compito di far acquisire agli studenti non solo le competenze necessarie al mondo del lavoro e delle professioni, ma anche le capacità di comprensione e applicazione delle innovazioni che lo sviluppo della scienza e della tecnica continuamente produce. In questo contesto è richiesto un aggiornamento permanente dei contenuti e delle metodologie di lavoro evidenziando l'esigenza di attivare una didattica e una valutazione basate sulle competenze, e la formazione di professionisti ICT con competenze certificabili nel mercato del lavoro europeo.

*L'Istituto **ITCG "E.Fermi"** ha voluto precorrere i tempi della riforma introducendo tra i contenuti didattici del triennio Mercurio (dal prossimo a.s. articolazione Sistemi Informativi Aziendali) **l'Eucip Core**, realizzando una sperimentazione con il supporto di **AICA e MIUR** per concorrere alla formazione di tali professionisti.*

1. Importanza delle Certificazioni europee

EUCIP (European Certification of Informatics Professionals) è il sistema europeo di riferimento per le competenze e i profili professionali informatici.

EUCIP è stato sviluppato, con il contributo della Comunità Europea, dalle associazioni professionali informatiche europee raccolte nel CEPIS (Council of European Professional Informatics Societies), tra cui AICA per l'Italia.

È un sistema indipendente dai fornitori, che grazie anche alla disponibilità di un insieme completo di certificazioni delle competenze richieste per ciascun mestiere dell'ICT, rappresenta il riferimento nel mondo delle professioni informatiche, dell'impresa e della formazione.

In questo contesto, EUCIP, con la sua articolazione in ventuno figure professionali, è una certificazione, riconosciuta:

1. a livello internazionale, attestante che chi lo possiede ha le abilità necessarie a gestire un sistema informativo in ambiti organizzativi complessi.
2. dal sistema delle Università italiane (CRUI e CINI), che ha siglato con AICA accordi specifici alla diffusione di EUCIP
3. dal sistema delle imprese (Confindustria) e da grandi organizzazioni pubbliche;
4. è integrata nelle "Linee Guida" di Digit@PA per i contratti ICT della Pubblica Amministrazione
5. è oggetto di collaborazione con i principali fornitori ICT per il raccordo con i rispettivi sistemi di certificazione

La base delle certificazioni del framework EUCIP è **EUCIP CORE**. La certificazione EUCIP Core verifica il possesso di un ampio spettro di conoscenze e abilità basilari che dovrebbero essere comuni a tutti i professionisti informatici, qualunque siano le specializzazioni e le attività svolte.

La certificazione EUCIP Core è un pre-requisito per il conseguimento della certificazione Eucip Livello Professionale.

L'Eucip Core si articola nel seguente modo:

Il syllabus è suddiviso in tre aree di conoscenza o moduli	Pianificazione (plan) Realizzazione (build) Esercizio (operate) Ogni modulo si può definire come un'unità di conoscenza coerente
Nell'ambito di ciascun modulo ci sono varie categorie	Una categoria concettualmente fa riferimento a competenze omogenee
All'interno di ogni categoria vari argomenti (topics)	Gli argomenti comprendono nel complesso tutti gli aspetti principali della categoria

Plan knowledge area

Espone i temi che stanno al confine tra le competenze informatiche in senso lato e altre discipline utili al professionista ICT per muoversi in ambiti lavorativi e aziendali con un buon grado di confidenza su questioni economiche, giuridiche e relazioni interpersonali.

Build knowledge area

Affronta tutti i principali aspetti relativi alla realizzazione e al miglioramento dei sistemi informativi, dal processo di sviluppo di sistemi informatici alla cura degli aspetti grafici e multimediali, dalla progettazione e realizzazione software alla progettazione e gestione di basi di dati.

Operate knowledge area

Si occupa della gestione operativa e del supporto all'esercizio dei sistemi informativi comprendendone gli aspetti tecnologici, le infrastrutture e le architetture.

2. Contesto di riferimento ed esigenza dell'intervento formativo

L'ITCG "E. Fermi" di Pontedera è una realtà storica nel territorio. La scuola opera da molti anni nell'ambito della diffusione delle competenze digitali. Nell'ottica dell'innovazione, negli anni 90, è stato attivato l'indirizzo "Mercurio" che ha formato negli anni generazioni di "ragionieri programmatori". L'Istituto è accreditato come Test Center presso AICA, e offre la possibilità di sostenere gli esami della Patente Europea del computer (ECDL core).

Il continuo sviluppo delle tecnologie dell'informazione e della comunicazione, la nuova riforma degli istituti tecnici, ha posto l'esigenza di un aggiornamento

permanente dei contenuti e delle metodologie di lavoro evidenziando la necessità di attivare una didattica e una valutazione basate sulle competenze.

Le attuali tendenze del mercato del lavoro richiedono dei professionisti con competenze **certificabili e certificate**. In particolare sono richieste figure professionali che riescano a coniugare competenze sui processi aziendali e sui sistemi informativi con competenze ampie e tecnologicamente avanzate nel campo dell'ICT.

Questa esigenza di precorrere i tempi della riforma del secondo ciclo d'istruzione secondaria di secondo grado già in attuazione nelle classi prime dall'Anno Scolastico 2010/2011 è ancora più forte nell'Istituto Tecnico Commerciale con Indirizzo Mercurio (che con la nuova Riforma Scolastica prenderà il nome di Sistemi Informativi Aziendali). I contenuti tecnologici, organizzativi ed economici risentono di una forte esigenza di rispecchiare le richieste del mercato del lavoro. In particolare di definire un profilo professionale in uscita dall'ambiente scolastico che non fissi l'attenzione su singoli prodotti commerciali o su specifici modelli hardware che spesso diventano obsoleti in tempi brevi, per cogliere invece, al loro interno, gli aspetti che permangono, le relazioni con il sistema informativo e i processi aziendali in continua evoluzione. Tali caratteristiche costituiscono la base per la costruzione di un profilo professionale **moderno ed europeo**.

3. Introduzione della sperimentazione

Partendo da questa esigenza di rinnovamento e riorganizzazione, l'Istituto ITCG "E. Fermi" ha raccolto il suggerimento di **AICA e MIUR** di introdurre nei curricula del Corso Mercurio i contenuti proposti dalla certificazione EUCIP Core e di effettuare insieme con altre due Istituti italiani (I.T.C. Romanazzi di Bari e I.T.C. Baffi di Fiumicino) una sperimentazione che si articola nel corso del triennio Mercurio.

Al termine del percorso scolastico, gli studenti che hanno sostenuto e superato gli esami previsti potranno conseguire la certificazione EUCIP Core.

3.1 Destinatari della sperimentazione

Sono destinatari del presente progetto gli studenti dell'ITCG "E. Fermi" frequentanti il triennio informatico.

La sperimentazione è stata attivata a partire dalle classi terze dell'indirizzo Mercurio nell'anno scolastico 2010/2011 e prosegue anche con le classi terze del corrente anno scolastico.

Durante il percorso del triennio, agli studenti sono proposti dei contenuti che fanno parte già dell'attuale percorso ministeriale, anche se sono rivisti dal punto di vista metodologico e riproposti mediante una stretta **interdisciplinarietà** tra alcune

discipline (in particolare Informatica, Economia Aziendale, Lingue, Diritto e Italiano).

Partendo dai contenuti proposti dal Syllabus Eucip Core, sono inoltre introdotti contenuti innovativi dal punto di vista tecnologico, operativo, metodologico, proposti con un linguaggio tecnico evoluto e moderno che avvicinano sempre di più lo studente alla realtà professionale.

L'apprendimento si avvale principalmente dell'orario curriculare diurno. Interventi pomeridiani sono previsti per approfondire alcuni aspetti e per preparare gli studenti allo svolgimento degli esami.

Durante il percorso triennale, gli studenti coinvolti nella sperimentazione devono superare **tre esami in lingua inglese** relativi a 3 moduli di Eucip Core (Plan, Build e Operate). Il superamento di questi esami permette agli studenti di conseguire la certificazione EUCIP Core che come è stato detto prima rappresenta un traguardo importante nel campo delle competenze riconosciute e certificate dell'ICT a livello europeo.

3.2 Obiettivi e organizzazione dell'attività

I contenuti previsti dai syllabus EUCIP-Core in parte riprendono i contenuti dei programmi Ministeriali del triennio programmatori, in parte li completano e li aggiornano introducendo competenze richieste dal mercato del lavoro e non previste dai programmi scolastici attuali.

Il riferimento a uno standard può far sì che le conoscenze acquisite dagli studenti non siano auto-referenziali, ma forniscano competenze ben definite, verificabili e certificabili.

Riassumendo gli obiettivi principali della sperimentazione si possono così sintetizzare:

1. Adeguare i curricoli scolastici partendo dalle nuove esigenze di mercato e dalla necessità di avere competenze standard e riconosciute a livello internazionale
2. Ampliare l'offerta formativa dell'Istituto
3. Fornire agli studenti delle certificazioni spendibili nel mercato del lavoro dopo il diploma come crediti in campo universitario, o come base per proseguire nelle altre certificazioni professionali EUCIP
4. Creare figure professionali specializzate nell'ambito dei sistemi informativi aziendali, con competenze specifiche nell'ambito delle problematiche informatiche

La scansione temporale degli esami prevista nell'arco del triennio è la seguente:

Modulo	Periodo
Modulo A	Primi mesi della classe V
Modulo B	Marzo-Aprile Classe IV
Modulo C	Marzo-Aprile Classe V

Le ore pomeridiane previste nel progetto da effettuare con le classi (con gli studenti che decidono di acquisire la certificazione) avranno l'obiettivo di approfondire alcuni contenuti nell'ottica del superamento dell'esame e di prepararli nella comprensione dei test relativamente ai contenuti e alla lingua.

Requisito importante per l'ottimale integrazione dei contenuti della certificazione all'interno dei curricoli è quello della formazione dei docenti coinvolti nella sperimentazione. Le discipline coinvolte sono informatica, economia aziendale, discipline giuridiche, lingua inglese e italiano. Un determinante supporto alla formazione è fornito da **AICA**.

Per effettuare gli opportuni approfondimenti, ogni docente deve anche operare in autoformazione utilizzando come punto di partenza i libri di testo disponibili sul mercato (vedi bibliografia).

Alcuni di questi docenti devono, inoltre, conseguire la certificazione nei tre moduli previsti dall'EUCIP CORE.

Con il conseguimento della certificazione da parte dei docenti, l'ITCG E. Fermi diventerà centro di competenza accreditato EUCIP Core per gli studenti della scuola e per rispondere alle esigenze di competenze certificabili provenienti dal territorio.

In seguito all'accreditamento, l'ITCG Fermi organizzerà corsi di formazione per coloro che operano nell'ICT oppure sono intenzionati a farlo e vogliono ottenere la certificazione EUCIP Core.

4. Nuova didattica e metodologia: esempi di sinergia cooperativa

L'Eucip Core contiene competenze multidisciplinari per l'applicazione delle tecnologie dell'informazione e della comunicazione nel mondo del lavoro.

Le discipline coinvolte non sono solo quelle legate all'informatica, ma coinvolgono vari aspetti nel campo economico, giuridico e sicuramente non meno importante è l'aspetto linguistico, dal momento che i test sono effettuati in lingua inglese.

All'inizio dell'attività non era chiaro, ai docenti coinvolti, quanto fosse importante l'introduzione di una nuova metodologia e didattica, dopo un anno di sperimentazione qualunque dubbio è stato sciolto ed è stato fondamentale per una buona riuscita dell'attività introdurre una cooperazione stretta tra le discipline sopra elencate.

I docenti d'indirizzo sono stati coinvolti in un'attività di ridefinizione e articolazione dei contenuti da proporre agli studenti nel corso del triennio. Tale fase non è ancora terminata in quanto è soggetta a rivisitazione e cambiamenti legati alla risposta degli studenti. Un aspetto cruciale è stato anche l'identificazione dei contenuti da proporre sia in modo **interdisciplinare**, sia mediante attività **progettuali** sia **laboratoriali**.

Il punto di partenza è quello di coinvolgere in prima persona lo studente ponendolo al centro delle attività, utilizzando metodologie "*student center*". Alcune lezioni si svolgono con il supporto del **CLIL**, esaltando un approccio in cui i contenuti siano pensati e studiati in lingua inglese (ricordo che gli studenti devono affrontare gli esami EUCIP Core con dei test in lingua inglese).

Alcuni docenti d'indirizzo hanno iniziato dei percorsi all'estero per acquisire specifiche competenze nel campo CLIL. Alcune lezioni si svolgono con la presenza contemporanea dell'insegnante della disciplina e con l'insegnante di lingua inglese.

La centralità dello studente è stata esaltata dall'uso quotidiano di lezioni interattive con le **LIM** (ogni classe coinvolta ha una LIM) e con l'uso di piattaforme di e-Learning che supportano lo studente nella fruizione di contenuti digitali.

Uno dei punti di forza della certificazione è che richiede una visione integrata di problematiche aziendali e IT soprattutto nel modulo "**Plan**".

Di seguito un classico esempio: in un'azienda le basi dati tradizionali si configurano come alcune delle sorgenti informative di un *data warehouse*. Questo è alimentato da fonti d'informazioni eterogenee tra loro. Il data warehousing si occupa di gestire la raccolta delle informazioni delle basi dati informative. L'eterogeneità è risolta dal processo *ETL*. Tecniche di analisi chiamate *data mining* permettono di individuare relazioni nascoste fra i dati. Ad esempio scoprire che un certo prodotto A è venduto maggiormente in una determinata zona geografica da un consumatore di mezza età. Le tecniche di data mining individuano associazioni tra prodotti diversi e possono scoprire che il prodotto A è acquistato in un certo periodo dell'anno con un prodotto B.

E' facile intuire come i due mondi si integrino perfettamente:

1. i supporti informatici per memorizzare i dati nel data warehouse e l'uso di tecniche di data mining fornite da strumenti applicativi
2. il mondo del marketing che può utilizzare le informazioni sia a livello direzionale sia strategico per campagne pubblicitarie o addirittura per posizionarsi nel mercato di riferimento.

Lo studente non studia separatamente il data warehouse in informatica e le campagne di marketing nelle discipline economiche. Avrà una visione globale della problematica e del tipo di supporto che la tecnologia può fornire alle strategie aziendali.

Un altro esempio d'integrazione: per soddisfare gli obiettivi strategici di un'azienda, il sistema informativo aziendale è formato da più sottosistemi preposti alla gestione dei processi aziendali (es. Pianificazione personale, Controllo qualità del prodotto, Gestione anagrafica dei dipendenti, Gestione anagrafica dei clienti e fornitori,). Dal punto di vista informatico, l'orientamento generale punta all'adozione di sistemi ERP (Enterprise Resource Planning), per fornire all'azienda un insieme di applicazioni integrate in grado di supportare i diversi processi aziendali.

Lo studente analizza tali processi e il loro inserimento in un contesto più applicativo-tecnologico mediante l'analisi di sistemi ERP.

Quelli forniti sono solo due esempi, ma il mondo informatico e aziendale presentano una stretta integrazione che non può essere certamente esaurita con poche righe. Rappresenta, infatti, uno degli aspetti più critici della sperimentazione.

Come affrontare dal punto di vista didattico metodologico tali integrazioni?

- ✓ predisponendo dei supporti digitali in modo integrato
- ✓ con lezioni da parte del docente della disciplina e di lezioni in copresenza tra docenti di discipline diverse anche di lingua inglese
- ✓ mediante visite aziendali in cui gli studenti possono sperimentare in sul campo l'integrazione di realtà tradizionalmente informatiche e aziendali

Quelli che sono stati elencati sono solo alcuni approcci, sicuramente se ne possono utilizzare altri più efficaci che la sperimentazione potrà evidenziare nel corso degli anni successivi.

I libri attualmente pubblicati sull'argomento, sono molto tecnici, sintetici per alcuni aspetti e non hanno un approccio adatto a uno studente delle scuole superiori. Per fornire agli studenti dei riferimenti completi e di supporto alla loro preparazione che integrino le competenze multidisciplinari, sono in corso di definizione dei materiali

didattici autoprodotti (**Learning Object**) per dare una curvatura vicina a quanto previsto da Eucip Core.

Sempre nell'ottica di un approccio che vede lo studente protagonista primario dell'apprendimento, sono in fase di realizzazione da parte degli stessi dei glossari bilingue fruibili tramite e-books. Gli studenti si occupano anche dell'auto definizione di test in lingua inglese.

La sperimentazione finora ha coinvolto tre istituti. Un aspetto fondamentale è stato la creazione di una rete che ha permesso la condivisione d'idee, metodologie, materiali didattici con gli altri Istituti coinvolti nella sperimentazione. Si spera che nel futuro, altri Istituti possano aderire alla sperimentazione in modo da allargare la condivisione delle esperienze.

Di seguito una tabellina riepilogativa dei punti di forza e degli aspetti critici finora riscontrati in fase di sperimentazione

Punti di forza	Punti critici
Introduzione del CLIL	Iniziale rallentamento dei tempi di apprendimento generati dall'introduzione del CLIL
Uso quotidiano di lezioni interattive con le LIM	Formazione iniziale dei docenti per l'uso di LIM e piattaforme di e-learning (non per tutti è stato un approccio semplice)
Definizione di materiali didattici autoprodotti (Learning Object)	I materiali autoprodotti non sempre sono stabili in quanto soggetti a rivisitazioni e miglioramenti in base ai riscontri avuti con la classe
Coinvolgimento diretto degli studenti nella preparazione di glossari bilingue fruibili tramite e-books e nell'auto definizione di test	I piani di lavoro sono "work in progress" in quanto sono rivisitati periodicamente per adattarsi alle nuove idee e alle nuove proposte del Consiglio di Classe

5 Strumenti di valutazione

La fase di valutazione ha lo scopo di verificare la coerenza e l'adeguatezza dell'intervento formativo.

Gli strumenti utilizzati saranno:

- verifica delle competenze acquisite nei singoli moduli
- valutazione dei miglioramenti nella lingua inglese e della capacità di espressione in inglese di contenuti tecnologici ed economici
- l'interesse degli studenti sulle attività proposte mediante questionari (a metà dell'anno scolastico e alla fine)
- numero di certificazioni acquisite dagli studenti (misurate dopo ogni sessione di esame)
- percentuale di frequenza mantenuta durante gli incontri pomeridiani

Il lavoro è appena cominciato.....

6 Riferimenti bibliografici e sitografici

<http://www.aicanet.it/>

<http://www.cepis.eu/>

"Professione Informatica – Volume unico (nuova edizione)" - Schgor, Franza ed altri- edito da Franco Angeli, 2007

e-book The all-round it professional. Part A. Plan knowledge Area: Use and Management of information systems - Autori e curatori: Paolo Schgor, Raffaele Brambilla, Fabrizio Amarilli editore Franco Angeli

e-book The all-round it professional. Part B. Build knowledge Area: Acquisition, development and implementation of information Autori e curatori: Paolo Schgor , Marina Cabrini , Giovanni Franza , Eugenio Schininà , Giuseppe Zerbi editore Franco Angeli

e-book The all-round it professional. Part C. Operate knowledge Area: Operations and support of information systems Autori e curatori: Paolo Schgor , Raffaele Brambrilla, Marina Cabrini , Giovanni Franza , Eugenio Schininà editore Franco Angeli

“EUCIP - Core Level. Guida alla certificazione per il professionista IT” – Bigini, Brambilla, Cappiello, Plebani, Rizzo – edito da Tecniche Nuove

"EUCIP Il Manuale per l'informatico professionista" - Teti, Cipriano - edito da Hoepli

Uncovering CLIL Mehisto, Marsh Frigols MacMillan Books

Laboratory Assistant Suite platform for biomedical data management and integration

Elena Baralis¹, Andrea Bertotti^{2,5}, Federico Bussolino^{3,5}, Alessandro Fiori¹,
Alberto Grand¹, Enzo Medico^{4,5}, Tommaso Renzulli⁴

¹Dipartimento di Automatica e Informatica, Politecnico di Torino

Corso Duca degli Abruzzi 24, 10129, Torino, Italy

{elena.baralis, alessandro.fiori, alberto.grand}@polito.it

²Laboratory of Molecular Pharmacology, ³Laboratory of Vascular Biology,

⁴Laboratory of Oncogenomics, Institute for Cancer Research and Treatment (IRCC)

Str. Prov. 142 Km. 3.95, 10060, Candiolo, Italy

⁵Department of Oncological Sciences, University of Torino Medical School

Str. Prov. 142 Km. 3.95, 10060, Candiolo, Italy

{andrea.bertotti, federico.bussolino, enzo.medico, tommaso.renzulli}@ircc.it

With the advent of new technologies in the biomedical and molecular oncology fields, research laboratories need systems that manage and analyze the huge amount of heterogeneous and complex data. The analysis of integrated data also permits the discovery of new knowledge related to tumors and the improvement of medical treatments. This paper presents the Laboratory Assistant Suite (LAS), a platform designed to assist the researchers in all laboratory activities, including the integration of heterogeneous data. The modular architecture of the platform allows properly managing each type of data (e.g., molecular, biological) and building complex analyses on integrated data. Finally, the LAS interfaces are designed to ease data management in hostile environments (e.g., in the surgery).

1. Introduction

Nowadays in the post-genomic era, with the advent of new technologies (e.g., microarrays), a huge amount of complex molecular data are generated with high throughput. The management of clinical, molecular and biological data is definitely a challenging task, due to the complexity and heterogeneity of data. Moreover, the analysis of heterogeneous molecular data, integrated with clinical and biological information, is a major prerequisite to discover the main causes of tumors. Indeed, genomic data encode many biological insights whose investigation can be the basis for dramatic advances in the field of molecular oncology.

Many efforts have been devoted to building management systems for this kind of data. One of the main ICT technologies employed in research laboratories is the Laboratory Information Management Systems (LIMS). Many commercial LIMSS are available, but they are usually very expensive and

require a considerable amount of human and economic resources to adapt to specific laboratory requirements [Haquin, 2008]. A number of LIMSs have also been proposed by the researcher community and released as free and open-source projects. For instance, TreeSNPs [Clément, 2010] manages data generated through single nucleotide polymorphism (SNP) identification. Differently, Screensaver [Tolopko et al., 2010] supports the storage and the comparison of small molecule and RNA screening data. Other proposed LIMSs focus on managing data associated with mouse colonies. For instance, LAMA [Milisavljevic et al, 2010] has been designed to support the tracking of different animal colonies. Instead, MausDB [Maier et al., 2008] integrates standard mouse colony management, phenotyping workflow scheduling and the management of mouse phenotyping results. However, all these systems are mainly focused on managing few kinds of highly specific data related to a particular set of experiments.

In 2004, the National Cancer Institute (NCI) started the Cancer Biomedical Informatics Grid (caBIG®¹) project [Kuhn et al., 2007], to develop an information infrastructure that enables the interdisciplinary collaboration for cancer research. Many tools have been developed under this project addressing both the management and the analysis of biomedical data. For instance, caTissue provides biospecimen inventory, tracking, and basic annotation capabilities for biospecimen resource facilities. Differently, VISDA [Wang et al., 2007] is an analytical tool focused on clustering approaches applied to high dimensional and complex biomedical datasets. However, the employment of caBIG® tools in research labs presents many critical issues. As pointed out in [London & Chatterjee, 2010], the end user data accessibility may be compromised by object model complexity and overly generic terminologies.

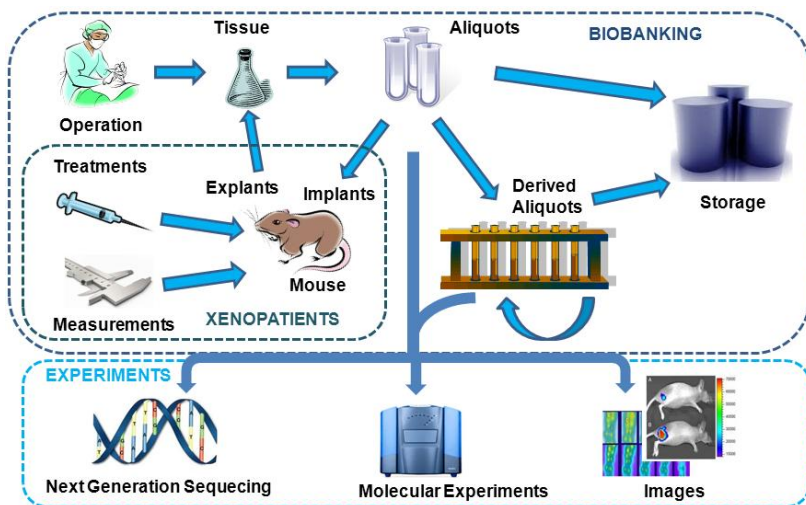


Fig. 1 – Dataflow including the xenopatient approach

¹ <https://cabig.nci.nih.gov/>

To overcome the main issues of all aforementioned approaches and address functionality requirements of research laboratories, we propose the Laboratory Assistant Suite (LAS) platform that assists the researchers in different laboratory activities. Its modular architecture allows managing different kinds of raw data (e.g., biological, molecular) and tracking experimental data. Furthermore, the platform supports the integration of different resources and aids in performing a variety of analyses in order to extract knowledge related to tumors. User interfaces have been designed to enable managing data in hostile environments where researchers need to minimize their interactions with the system (e.g., in sterile conditions). The paper is organized as follows. Section 2 describes the context in which the LAS has been studied and developed. Section 3 discusses the main features of the system architecture, while Section 4 presents the main characteristics of three modules and demonstrates the effectiveness of the proposed system in real use-cases. Finally, Section 5 draws conclusions and presents future developments.

2. Context

Several procedures are performed every day in research laboratories to analyze different biological and medical aspects of tumors, with the aim of discovering new knowledge and improving the therapies. To model our environment we analyzed a subset of procedures developed and adopted in the research laboratories at the Institute for Cancer Research and Treatment (IRCC). In particular, we focused our attention on the procedures that are involved in the xenopatient experimental pipeline [Bertotti et al., 2011].

This approach is based on the serial transplantation of human tumor specimens in immunocompromised animals. The aim is to help in translating the correlative information emerging from data integration into clinically relevant and functionally validated biomarkers. In the following we describe the main activities of a research study including the xenopatients approach (see Fig. 1).

Tumour specimens are initially collected from surgical interventions. From each individual patient-derived material a set of aliquots is generated (i.e., vital, RNA Later and snap-frozen). According to the characteristics of the aliquots and the purpose of the research study in which they are collected, different operations can be performed: (i) storage in a dedicated container (e.g., freezers), (ii) extraction of derived aliquots such as DNA and RNA, and (iii) implantation in immunocompromised animals (i.e., xenopatients). Different procedures can be applied on the implanted animals. For instance, researchers can define treatment protocols to monitor the evolutions of the tumor mass in the animal by means of measurements and evaluate the response to drugs. Moreover, they can generate new tissue samples from these animals for further analyses. Indeed, all the (derived) aliquots can be exploited for experimental analyses with different technologies. For instance, the expression values of thousands of genes can be analyzed by means of the microarray technology. Differently, with next generation sequencing techniques, the complete DNA sequence of a sample can be extracted and analyzed.

To the aim of managing and integrating all this information, a robust but flexible data management platform is needed. In particular, different types of information (e.g., biological data, molecular data, procedure tracking data, sample tracking data), some of which can be highly complex, should be independently managed by the platform but, at the same time, hardwired and interconnected, in order to permit integrated analyses. Moreover, user interfaces should be very friendly on one side, and perfectly fit the actual procedures on the other, in order to avoid hindering the experimental pipeline. In addition, working with biological samples implies that many data should be entered by the user in a hostile environment (e.g., working with gloves in sterile conditions with potentially infectious samples).

3. Architecture

Since the laboratory-related procedures can be categorized into different levels according to data complexity and activity purpose, the LAS architecture has been extensively based on a three-tier design pattern [Eckerson, 1995], both at the system-wide and the software module levels. This is a well-established architectural paradigm in software engineering, which targets flexibility and reusability by breaking up an application into tiers. Each tier addresses a specific issue (e.g., presentation, business logic, persistent data management) and interacts with the other tiers by means of well-defined interfaces. We modeled the platform architecture following this paradigm in order to reflect the main activities performed in the research laboratories which are described in Section 2.

In the general architecture of LAS, each tier includes a set of fully-fledged applications, or *modules*. While the lower tier is mainly concerned with the collection of experimental data, the modules and data managed by the upper tiers are characterized by an increasing level of abstraction. Lower tiers can service requests generated by the upper tiers and provide the data needed to carry out complex tasks (e.g., data integration and/or analysis).

The **operative tier** is responsible for collecting, storing, and tracking raw experimental data. These include data from several sources, such as xenopatient management, microarray experiments, imaging, and tissue collection, each handled by a specific software module. Modules belonging to this tier are meant to work in close interaction with the researchers in a laboratory environment. Graphical user interfaces (GUIs) are thus explicitly tailored to ease data entry operations and assist the researchers throughout their experiments. The interaction is designed to be especially lean with the aid of special input devices, such as touch-screen notepads and barcode readers, available in the IRCC research institute.

The **integration tier** is aimed at integrating different types of raw experimental data by means of complex queries. The integrated data can be browsed or visualized as graphs (e.g., genealogy trees) or diagrams (e.g., Venn diagrams). In addition, they can be fed to the analysis tier and enriched by means of annotations. For instance, a population of xenopatients can be annotated as responsive to a given drug according to statistical analyses.

Similarly, publicly available collections of experimental data can be imported into the system and analyzed to produce annotations. To ensure a high degree of flexibility, annotations need not follow a predefined scheme (e.g., a relational database table) but can include any kind of information the user deems relevant. To this aim, an ad-hoc, non-relational database is exploited to store annotations. Moreover, virtual experiments on molecular data can be defined by complex queries and submitted to related operative modules to be managed.

The **analysis tier** allows the definition of a workflow for the analysis of integrated data. Like with popular data mining tools (e.g., Rapid Miner [Mierswa et al., 2006]), users can design their analysis session by selecting and cascading one or more analysis building blocks. Blocks map to diverse analysis tools, including commercial off-the-shelf bioinformatics software (e.g., GenePattern [Reich et al., 2006]) as well as data mining algorithms developed within the research institute. The analysis process ultimately generates annotations, which are sent to the integration tier and stored in the database, and it can optionally export data for visualization with external tools.

The browser-based interface of each LAS module is implemented in Python and exploits Django², an open-source web application framework which follows the Model-View-Controller [Gamma et al., 2002] architecture. According to the data structure and type, a relational DBMS (MySQL³) and/or a non-relational database engine (MongoDB⁴) are exploited to provide persistent storage for the experimental data and the annotations. The database also stores administrative data (e.g., user privileges), which can be easily managed through an ad-hoc interface. Different user privileges can be set in order to grant users different levels of access to the application functionalities. Furthermore, each module tracks all user activity by means of audit trail procedures.

A set of APIs (Application Programming Interface) is provided to expose some of the functionalities offered by each application and permit data exchange among different applications. All the APIs have been implemented with the Piston⁵ Django library.

To provide better flexibility, all interface elements that prompt the user to choose from a set of predefined options (e.g., drop-down lists, radio buttons, check-boxes) are populated with values drawn from the database. This way, the interface can be easily adjusted, as new options become available in the experimental environment or old ones are discontinued, by simply updating the corresponding table through the administrative interface. Moreover, the system makes wide use of Ajax and JavaScript to enhance interface responsiveness and dynamically update pages in response to user interaction.

² <https://www.djangoproject.com/>

³ <http://www.mysql.com/>

⁴ <http://www.mongodb.org/>

⁵ <https://bitbucket.org/jespern/django-piston/>

4. Modules and use cases

The LAS architecture is the result of a joint effort by both IT and biomedical researchers, in an attempt to design a system apt to effectively manage the entire experimental and analysis process in the long run. Since the LAS platform should also start to provide support for the most crucial activities as promptly as possible, some modules, particularly those associated with the operative tier, have been implemented first, while others are still under development. To date, implementation of the BioBanking Management Module and the Living Specimens Management Module is complete, while it is still in progress for the Multi-dimensional Annotation Manager. The main functionalities of each module and some real use-cases are described in the following sections.

4.1 BioBanking Management Module

With the rapid advances in biomedical and genetic technologies, collections of human biological materials have attracted increasing attention from the researcher community. Such collections, named *biobanks*, represent a fundamental resource for the research and the diagnosis of different pathologies, and the study of possible therapeutic applications. Based on the types of biological materials they store, biobanks are commonly divided into *tissue* and *genetic* biobanks. The BioBanking Management Module (BBMM) addresses both issues.

The figure displays three overlapping screenshots of the BBMM sample collection interfaces. The top window, titled 'Insert new sample collection', shows a form for entering sample details: Tumor type (Lung), Hospital (Candilio), Date (07-10-2011), and Operation code (12345). It also includes a 'Tissue collected' section with various checkboxes. The middle window, titled 'You are collecting Peritoneal Met', shows a 'GenealogyID' (LNG0003PM) and a 'Tissue collected' section with checkboxes. The bottom window shows two large grids for sample collection: 'RNA LATER' and 'SNAP FROZEN'. Each grid has columns numbered 1-12 and rows lettered A-H. The 'RNA LATER' grid has a 'LoadPlate' button, and the 'SNAP FROZEN' grid has a 'LoadPlate' button. A 'Barcode' field is also visible on the right side of the bottom window.

Fig. 2 – The BBMM sample collection interfaces

The scope of the BBMM spans a wide range of activities, including management of biological samples and laboratory instrument data, as well as

support to a number of laboratory-related procedures. Aliquots collected from surgical intervention and derived aliquots are managed separately, since the related information and procedures are different. For instance, the module assists the researcher in all derivation steps. The composition of each derived aliquot is suggested according to derivation type (e.g., DNA, RNA) and laboratory rules regulating the volume and concentration of the obtained mixture. Moreover, all kits employed during derivation procedures are tracked in order to manage the warehouse and monitor quality indexes associated to kits and derived aliquots.

To enable the employment of barcode readers and thus streamline data entry procedures, most pieces of experimental equipment (i.e., clinical folders, plates, tubes) are bar-coded by the IRCC prior to their use (see Section 3). Fig. 2 depicts the interfaces for the collection of biological samples. Since this is one of the core functionalities of the BBMM, it is discussed as a representative example. Sample collection consists of two phases. In the first phase (window A) the researcher inputs preliminary information about the collection event and selects one or more sample types. In the next phase (window B), the researcher can add one or more samples for each of the sample types chosen. Particular care has been devoted to the design of this window, so that the graphical interface closely matches the physical working environment of the researcher. Each of the four panels (vital, ffpe, RNA later, snap-frozen) represents a physical working plate in which sample aliquots are stored. Plates can be loaded into the application by means of their barcode. Inserting an aliquot in a given plate position has been designed to be as simple as clicking on the corresponding cell, which can be comfortably done on a touch-screen device without resorting to a mouse. This results in a reduced number of lean interactions with the system, allowing the researcher to focus on the experimental task at hand, rather than on data entry operations.

4.2 Living Specimens Management Module

One of the crucial activities within the IRCC, management of the immunocompromised animals (i.e., xenopatients) is supported by the Living Specimens Management Module (LSMM). In particular, this module enables tracking and monitoring of the xenopatient life cycle, from their acquisition by the research institute to their death.

Mice employed by the IRCC are equipped with RFid tags, allowing quick and easy identification of each animal. The need to load into the system an entire batch of mice at a time benefits from this facility. After mice loading, tumor tissue fragments are implanted into the available xenopatients. The LSMM effectively manages this operation by accessing the BioBank database (through the appropriate APIs) and retrieving the tumor aliquots currently stored in the source plate. Tumor growth within the xenopatient, monitored by either qualitative or quantitative measurement of the tumor mass, can be tracked by means of ad-hoc functionalities. Treatments, aimed at assessing the effectiveness of given drugs, can be assigned to measured xenopatients. Each treatment is composed of several phases, associated with different drugs. Fig. 3 illustrates the interface for defining a new treatment, displayed on a tablet

computing device. For each phase, the drug, the administration mode, the dose, and the administration frequency can be defined. In addition, a Gantt chart allows specifying a time schedule for each phase.

The xenopatient life cycle is ended by an explant, which generates a number of tumor tissue aliquots, made available for future implants. As with implants, explant operations need to cooperate with the BioBanking Management Module, in order to display destination plates and transmit data about generated aliquots.

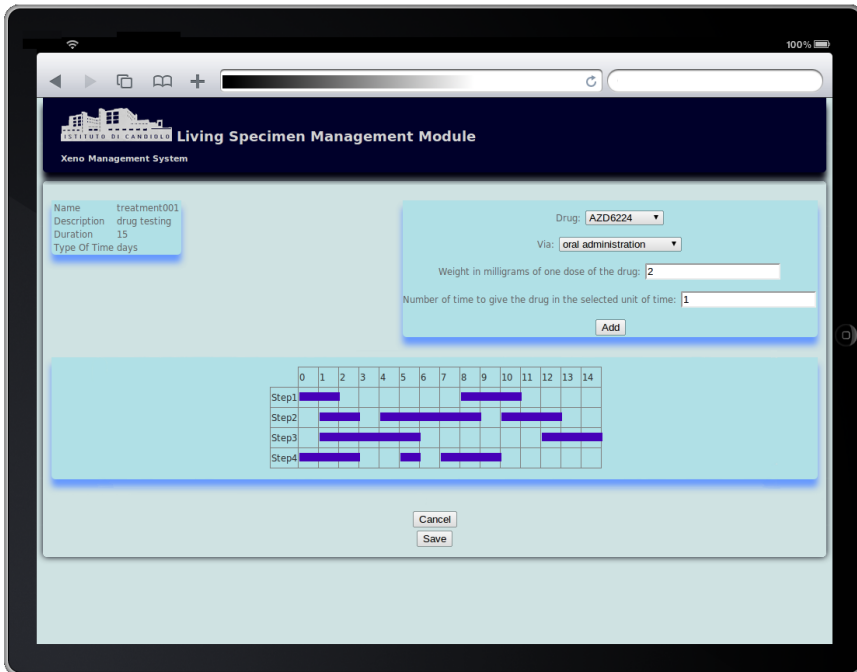


Fig 3. – A tablet PC displaying the treatment definition interface

4.3 Multi-dimensional Annotation Manager

The integration of different biological and molecular data is a fundamental task to discover new knowledge related to tumors. The Multi-Dimensional Annotation Manager (MDAM) aims at extracting from the operative tier all the information of interest for the researcher and/or a study. Since the heterogeneity of data and the different purposes of research studies introduce a high level of complexity, this module is still under development. However, in the following we present the main ideas on which its functionalities will be designed.

The MDAM allows performing heterogeneous and complex queries on all the modules belonging to the operative tier. The query set is defined as a workflow composed of blocks, named *query blocks*. Each query block defines the object that will be retrieved (e.g., aliquot, xenopatient), its related information of interest and the filtering conditions. An internal module of the MDAM evaluates

the workflow in order to detect not properly defined operations (e.g., intersections among disjoint sets of objects) and define an optimal execution plan on the distributed modules. To ease system use, a set of wizards based on predefined workflows is also provided. Moreover, all data requests built by the system to accomplish the task defined by the workflow are hidden to the final user.

Each object returned by a (predefined) workflow is stored, if it is not present, in a non-relational database. This technology allows storing heterogeneous objects without defining a relational and fixed structure for the stored information. Thus, MDAM can easily store different biological and molecular objects and attach other information (i.e., annotations) collected from further analyses. For instance, researchers can assign the response to a drug treatment to a set of objects according to statistical analyses performed on corresponding molecular and biological data. Furthermore, integrated data can be browsed or visualized by means of graph (e.g., genealogy trees) or diagram (e.g., Venn diagrams) representations.

Finally, communication among the modules is performed by means of the APIs defined in each module. Each API of interest for the integration task and its related parameters are registered into the MDAM and a corresponding query block is created.

5. Conclusions and future work

In this paper we present the Laboratory Assistant Suite (LAS) platform designed to assist researchers of biological and biomedical laboratories in all their activities. The modular architecture allows managing heterogeneous and complex data and performing different analyses. Moreover, the graphical interfaces and the web-based architecture enable the researchers to use the platform in hostile environments (e.g., working in sterile conditions). The real use-cases of employment of two operative modules show the effectiveness of the proposed approach in modeling the research environments and managing laboratory-related procedures.

We are currently completing the main functionalities of the MDAM and performing a wide set of experiments to validate the effectiveness of the integration approach. In parallel, we are developing other operative modules of the LAS platform. In particular, we are designing a microarray management module to enable researchers to manage microarray experiments (e.g., hybridization methods), track the material (e.g., chips), define quality controls, and build virtual experiments.

6. Acknowledges

The authors would like to thank Piero Alberto and Emanuele Geda for implementing parts of the LAS platform.

References

[Bertotti et al., 2011] Bertotti A., Migliardi G., Galimi F., Sassi F., Torti D., Isella C., Corà D., Di Nicolantonio F., Buscarino M., Petti C. and others, A molecularly annotated platform of patient-derived xenografts ('xenopatients') identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer, *Cancer Discovery*, 2011.

[Clément et al., 2010] Clément S., Fillon J., Bousquet J., Beaulieu, J., TreeSNPs: a laboratory information management system (LIMS) dedicated to SNP discovery in trees, *Tree Genetics & Genomes*, 6, 3, 2010, 435-438.

[Eckerson, 1995] Eckerson W. W., Three Tier Client/Server Architecture: Achieving Scalability, Performance, and Efficiency in Client Server Applications. *Open Information Systems*, 10, 1995.

[Gamma et al., 2002] Gamma E., Helm R., Johnson R., Vlissides J. and others, Design patterns, 1, 2002, Addison-Wesley Reading, MA.

[Haquin et al., 2008] Haquin S., Oeuillet E., Pajon A., Harris M., Jones A.T., van Tilbeurgh H., Markley J.L., Zolnai Z., Poupon A., Data management in structural genomics: an overview, *Methods in molecular biology*, 426, 2008, 49.

[Kuhn et al., 2007] Kuhn, K. and others, The Cancer Biomedical Informatics Grid (caBIG™): Infrastructure and Applications for a Worldwide Research Community, *Medinfo 2007*, 1, 2007, 330.

[London and Chatterjee, 2010] London J.W. and Chatterjee D., Using the Semantically Interoperable Biospecimen Repository Application, caTissue: End User Deployment Lessons Learned, *IEEE International Conference on BioInformatics and BioEngineering*, 2010, 316-317.

[Maier et al., 2008] Maier H., Lengger C., Simic B., Fuchs H., Gailus-Durner V., de Angelis M.H., MausDB: An open source application for phenotype data and mouse colony management in large-scale mouse phenotyping projects, *BMC bioinformatics*, 9, 1, 2008, 169.

[Mierswa et al., 2006] Mierswa I., Wurst M., Klinkenberg R., Scholz M., Euler T., YALE: Rapid Prototyping for Complex Data Mining Tasks, in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006.

[Milisavljevic et al, 2010] Milisavljevic M., Hearty T., Wong T.Y.T., Portales-Casamar E., Simpson E.M., Wasserman W.W., Laboratory Animal Management Assistant (LAMA): a LIMS for active research colonies, *Mammalian Genome*, 21, 5, 2010, 224-230.

[Reich et al., 2006] Reich M., Liefeld T., Gould J., Lerner J., Tamayo P., Mesirov J.P., GenePattern 2.0, *Nature genetics*, 38, 5, 2006, 500-501.

[Tolopko et al., 2010] Tolopko A., Sullivan J., Erickson S., Wrobel D., Chiang S., Rudnicki K., Rudnicki S., Nale J., Selfors L., Greenhouse D. and others, Screensaver: an open source lab information management system (LIMS) for high throughput screening facilities, *BMC bioinformatics*, 11, 1, 2010, 260.

[Wang et al., 2007] Wang J., Li H., Zhu Y., Yousef M., Nebozhyn M., Showe M., Showe L., Xuan J., Clarke R., Wang Y, VISDA: an open-source caBIG™ analytical tool for data clustering and beyond, *Bioinformatics*, 23, 15, 2007, 2024.

LITES: Illuminazione stradale intelligente a LED per il risparmio energetico

Fabrizio Lamberti, Andrea Sanna,
Eduardo Andres Henao Ramirez, Claudio Demartini
Politecnico di Torino, Dipartimento di Automatica e Informatica
Corso Duca degli Abruzzi, 24 10129 Torino
fabrizio.lamberti@polito.it, andrea.sanna@polito.it,
eduardo.henaoramirez@polito.it, claudio.demartini@polito.it

La realizzazione di reti per illuminazione stradale basate su tecnologie innovative e su soluzioni di monitoraggio e controllo intelligenti può indubbiamente contribuire ad una gestione più razionale ed efficiente delle risorse energetiche. In questo lavoro viene presentata una piattaforma software basata su soluzioni web e su tecniche di computer grafica e realtà virtuale progettata per la simulazione e la supervisione remota di infrastrutture pubbliche di illuminazione basate su punti luce a LED intelligenti in uno scenario di città digitale.

1. Introduzione

Il concetto di città digitale è oggi inteso come un meccanismo per mettere a disposizione dei diversi attori coinvolti nella gestione e nello sviluppo di scenari urbani (ingegneri, architetti, gestori dei servizi, ecc.) sistemi informativi in grado di supportarli in maniera efficace e sostenibile nelle varie fasi di pianificazione, progettazione, costruzione, ecc. Le città digitali, una combinazione di sistemi informativi territoriali, reti di sensori per la raccolta distribuita di informazioni, rappresentazioni tridimensionali ed altro possono infatti costituire un prezioso supporto nell'esecuzione di molteplici attività di analisi e supervisione basate su meccanismi di monitoraggio, visualizzazione, simulazione e controllo. Nel corso degli ultimi anni sono stati sviluppati diversi prototipi di città digitali utilizzati in scenari eterogenei: controllo del traffico, progettazione di edifici, prevenzione di disastri naturali, ecc. Un contesto applicativo che potrebbe beneficiare in modo significativo dell'integrazione in uno scenario di città digitale è quello dell'illuminazione pubblica stradale. L'illuminazione stradale deve essere in grado di offrire agli utenti molto più che un mero senso di comfort. Occorre infatti garantire il rispetto di specifici criteri fotometrici al fine di permettere ad utenti diversi di svolgere molteplici attività visuali di natura differente. Inoltre, dal momento che gli aspetti legati al monitoraggio ed al contenimento dei consumi assumono un ruolo sempre più centrale nella gestione del suddetto contesto, l'individuazione di sinergie tra l'obiettivo di fornire l'illuminazione giusta al momento giusto ed approcci in grado di garantire chiari risparmi energetici senza mettere in pericolo la sicurezza pubblica o lo svolgimento delle normali attività quotidiane sta ricevendo sempre maggior attenzione. Quindi, da un lato,

molte attività di ricerca si sono concentrate sull'analisi delle effettive possibilità di impiego di soluzioni per l'illuminazione efficiente, basate ad esempio sulla tecnologia LED. Nell'illuminazione stradale sono in genere utilizzate luci fluorescenti, a scarica, ecc. in quanto generalmente più potenti di quelle a LED. Tuttavia, oltre ad essere caratterizzate da una maggiore qualità di illuminazione e da una più elevata durata e robustezza, le luci a LED presentano un vantaggio fondamentale rispetto alle tecnologie tradizionali: sono totalmente regolabili in termini di potenza (e quindi di luce prodotta), con un consumo lineare ed una risposta istantanea. In particolare, la potenza può essere ridotta del 70% e più e riportata a quella originale non appena necessario, ad esempio in risposta ad un'improvvisa richiesta di illuminazione. Dall'altro lato, nel tempo sono state analizzate diverse strategie mirate ad ottimizzare tutti i fattori di efficienza legati alle diverse tecniche di illuminazione, al fine di individuare strategie efficaci per fornire la giusta quantità di luce in base ad una serie di parametri monitorati e legati, ad esempio, alle specifiche condizioni ambientali. Tuttavia, al fine di beneficiare effettivamente dei vantaggi offerti dalla tecnologia LED e di soluzioni per il monitoraggio e l'attuazione remota dimostrando in maniera concreta la loro efficacia nel produrre un significativo risparmio di energia elettrica nel contesto di soluzioni per l'illuminazione stradale, è ora necessario che vengano sviluppate architetture specifiche in grado di sfruttare la possibilità di regolare la potenza in tempo reale ed in maniera dinamica. Per questo motivo, negli ultimi anni sono stati finanziati diversi progetti aventi l'obiettivo di analizzare i suddetti aspetti e testare soluzioni concrete in condizioni di funzionamento reale.

A questo proposito, nel presente lavoro sono presentate in particolare le attività svolte dal Politecnico di Torino nell'ambito del progetto Europeo denominato LITES (www.lites-project.eu). Nell'ambito di tale progetto, l'obiettivo è quello di sviluppare una tecnologia per la costruzione di reti per l'illuminazione pubblica stradale basate su punti luce intelligenti. I punti luce sono dotati di sensori in grado di monitorare diversi parametri di contesto oltre che di capacità di elaborazione integrata per implementare regole di illuminazione locali. I dati raccolti dai sensori sono poi trasmessi ad un centro di gestione remoto che permette di controllare e supervisionare il funzionamento della rete attraverso Internet. Il centro di gestione utilizza moderne tecniche di computer grafica per permettere il monitoraggio in tempo reale della configurazione di illuminazione, l'impostazione dei parametri dei diversi punti luce ed il controllo delle regole di illuminazione locale. Mediante l'analisi dei dati di monitoraggio raccolti dai sensori (con diverse configurazioni di illuminazione ed in molteplici situazioni ambientali, correlate ad esempio allo stato del traffico, alle condizioni meteo, ecc.) su di un periodo di tempo significativo, il centro di gestione può consentire inoltre la definizione di regole di illuminazione globali in grado di adattarsi a condizioni di funzionamento eterogenee. Infine, il centro di gestione può essere utilizzato come strumento di simulazione per pianificare una certa configurazione di illuminazione per un determinato sito, per studiare il comportamento della rete in risposta a particolari eventi individuati dai sensori (pedone o autoveicolo in movimento, variazione nelle condizioni di luce, ecc.), per stimare il risparmio di energia elettrica nelle condizioni di funzionamento previste e, infine, per impostare i diversi parametri di funzionamento in maniera tale da consentire alla

rete di fornire la miglior illuminazione possibile ed offrire agli utenti il maggior confort possibile. Il lavoro è così articolato: nel paragrafo 2 viene riportata una panoramica delle soluzioni esistenti per la realizzazione e la gestione di scenari virtuali, con particolare riferimento a quelle soluzioni basate nello specifico su tecniche di visualizzazione digitali. Nel paragrafo 3 viene discussa l'architettura generale del sistema sviluppato dal partenariato del progetto LITES e viene illustrato in particolare il ruolo del centro di gestione remoto. Nel paragrafo 4 vengono descritti gli aspetti implementativi del componente di visualizzazione realizzato per gestire la simulazione, il monitoraggio ed il controllo delle reti di illuminazione pubblica stradale in una serie di siti pilota.

2. Stato dell'arte

Recentemente, diversi pacchetti software e numerose tecnologie sono state sviluppate con l'obiettivo specifico di consentire la creazione di città digitali. Da un lato, le attività di ricerca in questo ambito si sono concentrate sulla identificazione delle tecnologie più idonee per ricreare in maniera realistica le caratteristiche specifiche del territorio, della vegetazione, degli edifici, ecc. al fine di garantire una ricostruzione fedele del mondo reale garantendo al tempo stesso prestazioni adeguate per il particolare scenario applicativo [Poullis e Suya, 2009]. Dall'altro lato, diversi lavori in letteratura si sono concentrati sulla sperimentazione di soluzioni di visualizzazione adeguate da adottare per lo specifico contesto di riferimento. In particolare, è possibile individuare alcuni casi d'uso ricorrenti [Batty et al, 2001] che spaziano dalla gestione delle situazioni di emergenza, alla pianificazione urbana e territoriale, dalla amministrazione dei servizi alla gestione del turismo, ecc. Ad esempio, in [Hai et al, 2008], le tecnologie della realtà virtuale sono state utilizzate per creare un ambiente di visualizzazione tridimensionale in grado di garantire all'utente un elevato grado di immersività nel coordinare le attività di un gruppo di soccorso nella gestione delle emergenze. In [Jianchun et al, 2010], diverse tecniche di computer grafica sono utilizzate per sviluppare un sistema grafico da utilizzarsi per l'ottimizzazione e la simulazione del layout di una rete di servizi sotterranei. In [Pack, 2010] sono invece mostrati alcuni esempi di come le moderne tecniche di visione artificiale possano essere utilizzate per affrontare le problematiche connesse alla rappresentazione dei dati e la gestione di sistemi di trasporto pubblico e di viabilità in genere; nello specifico, vengono sintetizzati i risultati di alcuni progetti di ricerca che si sono focalizzati, tra le altre cose, sulla progettazione virtuale e la costruzione di infrastrutture per i trasporti, la visualizzazione in tempo reale dei dati connessi ad incidenti e la visualizzazione su larga scala delle condizioni del traffico. Infine, in letteratura è possibile individuare alcuni interessanti applicazioni delle suddette tecnologie nel settore del turismo, focalizzate sia sull'analisi e la visualizzazione di dati connessi a situazioni contemporanee che a scenari di rilevanza storica. Ad esempio, in [Calori et al, 2009] vengono riportati i risultati del progetto Roma Virtuale, nel quale mediante rappresentazioni tridimensionali è possibile passare da una visualizzazione della struttura urbana attuale a quella relativa a precedenti periodi storici.

Sebbene alcuni dei lavori in letteratura utilizzino sistemi per la visualizzazione basati su motori grafici pubblicamente disponibili, la maggior parte ha fatto uso di ambienti grafici sviluppati appositamente per la particolare applicazione, al fine di permettere agli sviluppatori di utilizzare strumenti e sorgenti di dati standard (e già disponibili) e di concentrarsi quindi sugli aspetti essenziali e le funzionalità ritenute utili per il particolare servizio da offrire all'utente finale. Ad esempio, in [Chevrier et al, 1999] viene affrontato specificatamente il problema della simulazione luminosa, e viene presentato un ambiente software per gestire illuminazione naturale ed artificiale nella progettazione e costruzione di edifici. Per la realizzazione del suddetto ambiente sono state sviluppate soluzioni specifiche in grado di superare le limitazioni dei software esistenti, non in grado, ad esempio, di gestire una navigazione tridimensionale di scene complesse illuminate contemporaneamente da centinaia di sorgenti di luce. Le limitazioni dei software esistenti appaiono ancora più costrittive nel momento in cui si considerano scenari caratterizzati da vincoli ancora più stringenti, legati ad esempio alla necessità di controllare in tempo reale i parametri di illuminazione attraverso sistemi di monitoraggio e di attuazione remoti e di includere nella visualizzazione parametri ambientali liberi di variare in modo dinamico (come richiesto nello specifico dai requisiti del progetto LITES).

3. La soluzione LITES

La rete LITES si basa su una serie di punti luce LED intelligenti. Ogni punto luce è composto da una matrice di LED, un sistema di alimentazione, un insieme di sensori ed un'unità di controllo (e comunicazione). I dati rilevati dai sensori e relativi all'intensità di luce, alla presenza di moto, alla temperatura ed ai consumi di corrente sono processati dall'unità di elaborazione di bordo, la quale è responsabile della generazione degli adeguati segnali di controllo indirizzati al modulo di alimentazione per la regolazione dinamica della potenza erogata. In questo modo l'intensità della luce emessa dal punto luce può essere variata in tempo reale (Fig.1).

I dati di monitoraggio e le informazioni relative allo stato dei diversi punti luce sono inviati ad un armadio elettrico tramite comunicazione su rete elettrica (PLC). L'armadio, da un lato, funge da concentratore. Dall'altro, si comporta da gateway per la trasmissione delle informazioni raccolte al centro di gestione remoto attraverso la rete Internet. Il centro di gestione si occupa della ricezione e della memorizzazione delle informazioni in arrivo in una base di dati centralizzata. Presso il centro di gestione, un sistema intelligente elabora i dati raccolti al fine di generare statistiche sul comportamento della rete e identificare strategie di illuminazione ottimizzate. L'interfaccia del centro di gestione remoto consente inoltre agli operatori (responsabili del servizio di illuminazione, squadre di manutenzione, ecc.) di supervisionare le funzionalità del sistema, permettendo agli stessi di ricevere notifiche e allarmi, di accedere ai dati sui consumi, di controllare la stabilità e l'efficacia delle regole di illuminazione, di identificare eventuali malfunzionamenti oltre che di simulare il comportamento del sistema in condizioni controllate (mediante un componente in grado di emulare il comportamento di un concentratore reale).

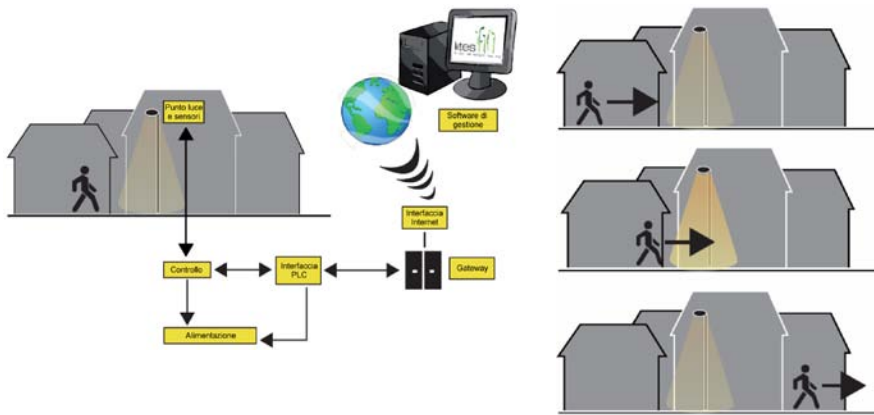


Fig.1 - Architettura generale della soluzione LITES

Sebbene la soluzione LITES sia composta da diversi moduli comunicanti tra loro, in questo lavoro l'attenzione è focalizzata specificatamente sul centro di gestione remoto e, nello specifico sul lato utente dello stesso (sono quindi tralasciati gli aspetti legati all'implementazione della base di dati, alla trasmissione delle informazioni ed all'elaborazione automatica delle stesse), realizzato integrando varie soluzioni web e tecniche di computer grafica.

In particolare, data l'eterogeneità delle informazioni da visualizzare (regole per il controllo dei parametri di illuminazione, dati provenienti dai sensori, allarmi, ecc.) e tenuta in considerazione la necessità di selezionare una strategia di presentazione di tali informazioni adatta a personale specializzato e non, in grado di mescolare simulazione su larga scala in tempo reale e generazione di statistiche, per l'implementazione del centro di gestione è stata selezionata la tecnologia tridimensionale. Sulla base di precedenti esperienze nel dominio considerato (ad esempio [Pack 2010]), la tecnologia tridimensionale è stata considerata, come confermato anche dai risultati preliminari nell'ambito del presente progetto, in grado di consentire lo sviluppo di un'interfaccia utente in grado di permettere agli operatori di avere a prima vista informazioni efficaci sullo stato generale della rete. Facendo ricorso alla suddetta tecnologia, la topologia di un particolare sito nel quale è stata installata la tecnologia LITES può essere ricostruita utilizzando tecniche tipiche della grafica al calcolatore, e l'effetto di illuminazione ricreato in un contesto virtuale. In questo modo, diventa possibile implementare funzionalità di monitoraggio efficaci arricchite da meccanismi di controllo potenti e facili da utilizzare. Ad esempio, un operatore può visualizzare l'impatto di un (eventuale) guasto nel sistema di illuminazione, supervisionare il comportamento dei sensori e studiare l'effetto di possibili aggiustamenti alle regole di controllo della potenza in termini di consumo di energia elettrica ma anche di confort luminoso.

4. Il centro di gestione remoto

Il centro di gestione remoto è stato progettato per la supervisione di quattro siti pilota coinvolti nel progetto, rappresentati dai campus dell'Università Tecnica di Riga (Lettonia) e dell'Università di Aveiro (Portogallo) e da due quartieri residenziali rispettivamente nelle città di Bordeaux (Francia) e Piaseczno (Polonia). Ciascun sito sarà dotato di 50 punti luce intelligenti.

4.1 Costruzione dei siti pilota digitali

Allo stato attuale, il centro di gestione permette di lavorare sulle versioni virtuali dei primi tre siti pilota. Per la ricostruzione tridimensionale sono stati utilizzate informazioni cartografiche, disegni tecnici e fotografie rese disponibili dai partner di progetto o reperiti mediante siti pubblici (Fig.2). In Fig.3 e Fig.4 è possibile valutare i risultati delle fasi di ricostruzione confrontando gli stessi scenari rispettivamente nella loro rappresentazione reale e virtuale. In Fig.5 sono infine riportati alcuni rendering delle geometrie realizzate. Le informazioni relative ai modelli virtuali (inclusi quindi il terreno, gli edifici, le strade, i marciapiedi, ecc.) sono memorizzate nella base di dati assieme ai dati di monitoraggio e di simulazione.



Fig.2 - Fasi della ricostruzione virtuale del sito pilota di Bordeaux (Francia)

4.2 Implementazione software

L'approccio perseguito nello sviluppo del centro di gestione remoto è finalizzato al soddisfacimento degli specifici requisiti utente e dei relativi vincoli di progetto. In particolare, il centro di gestione è stato progettato mediante tecnologie web, sfruttando unicamente componenti open source, avendo in mente un utilizzo multi-modale su diversi tipi di dispositivi (inclusi quelli mobili) e con diverse modalità di interazione, inclusi comandi vocali e gesti. Il software è quindi rilasciato come applicazione Java eseguibile dal web.

L'interfaccia utente consiste di una finestra di visualizzazione tridimensionale e di alcuni pannelli di controllo (Fig.6). La finestra di visualizzazione consente di accedere in modo intuitivo a tutte le funzionalità di monitoraggio in tempo reale, mentre i pannelli di controllo sono utilizzati per la gestione dei parametri di simulazione e la visualizzazione delle statistiche di funzionamento.

La finestra di visualizzazione è stata sviluppata utilizzando Multi-Touch for Java (MT4j), una libreria flessibile per la creazione di applicazioni dotate di interfacce grafiche bidimensionali e tridimensionali controllabili mediante gesti e basate su Java OpenGL.

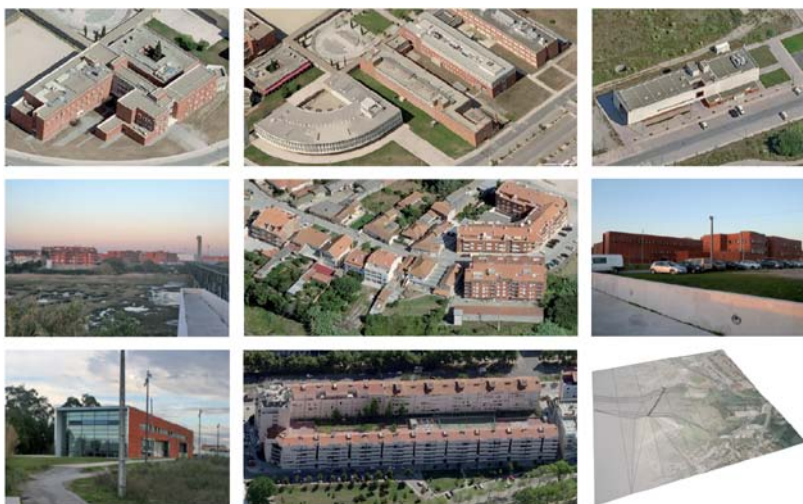


Fig.3 - Fotografie relative al sito pilota di Aveiro (Portogallo)



Fig.4 - Ricostruzione virtuale del sito pilota di Aveiro (Portogallo)



Fig.5 - Ricostruzione virtuale del sito pilota di Riga (Lettonia)

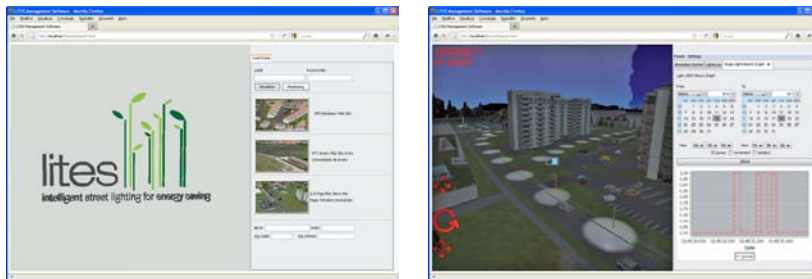


Fig.6 - Interfaccia grafica del centro di gestione remoto

In particolare, MT4j include il supporto per manipolare componenti geometrici primitivi, per elaborare formati standard, per gestire componenti grafici di base/personalizzati e per incorporare elementi complessi. MT4j è in grado di sfruttare l'eventuale acceleratore grafico disponibile sulla piattaforma, permettendo di raggiungere prestazioni significative garantendo una elevata portabilità del codice. Il rendering della scena è gestito da uno shader sviluppato ad hoc in grado di gestire le 50 sorgenti luminose (corrispondenti ai punti luce) previste nei diversi siti pilota. Lo shader permette di configurare la singola sorgente luminosa in termini di colore, posizione, orientamento ed apertura del cono interno/esterno. In base ai test sperimentali effettuati con una scheda grafica consumer di ultima generazione, lo shader consente di visualizzare scene composte da centinaia di migliaia di poligoni e centinaia di punti luce a più di trenta fotogrammi al secondo all'interno di un browser web.

Un modulo specifico realizzato utilizzando le API Java Speech permette l'interazione con il centro di gestione mediante comandi vocali. Per raggiungere i suddetti obiettivi di multi-modalità e consentire di operare sul sistema anche attraverso dispositivi dotati di limitate risorse di elaborazione, il centro di gestione è stato dotato di un modulo di visualizzazione remota. Il sistema può quindi essere configurato per distribuire a dispositivi client remoti il solo risultato del processo di rendering tridimensionale sotto forma di un flusso video/di immagini. Infine, sono state sperimentate tecniche alternative di interazione basate su sistemi per il riconoscimento dei gesti (ad esempio mediante sensori ottici di profondità), sul tracciamento nello spazio di dispositivi personali (ad esempio mediante la telecamera dei moderni telefoni cellulari) e su ambienti interattivi immersivi basati su stereoscopia attiva (Fig.7).

4.3 Monitoraggio e controllo

Durante il monitoraggio, le condizioni di luce dell'ambiente virtuale possono essere configurate per riprodurre quelle del sito pilota. I dati raccolti dai sensori sono infatti mappati sui parametri di illuminazione Java OpenGL, creando una rappresentazione piuttosto realistica dell'ambiente controllato. Il sistema di gestione è in grado di aggiungere notifiche visive direttamente all'interno della scena tridimensionale per mostrare all'operatore i consumi istantanei del singolo punto luce, gli eventi rilevati dai sensori, eventuali situazioni di allarme, ecc. Attraverso una telecamera virtuale, l'operatore può navigare la rete ed

avere una visione chiara dello stato dell'intera infrastruttura. Mediante l'interfaccia fornita, l'operatore può selezionare un singolo punto luce (o un gruppo di punti luce) ed impostare le regole di illuminazione locale definendo gli orari di accensione e spegnimento, la potenza minima, la potenza da fornire in risposta ad un particolare evento, ecc. Le regole per il controllo attivo e dinamico della potenza possono essere variate con estrema facilità, al fine, ad esempio, di ottimizzare i consumi o di implementare particolari politiche di illuminazione dettate dagli standard locali di riferimento. Mediante l'interfaccia, l'operatore può infine variare istantaneamente l'intensità di luce, la temperatura di colore ed altri settaggi del punto luce, oppure intervenire per gestire particolari situazioni degne di attenzione (ad esempio legate al verificarsi di situazioni di emergenza, a particolari condizioni meteo, ecc.).

4.4 Simulazione

Mediante i pannelli di controllo, l'operatore può configurare scenari di illuminazione virtuali variando i diversi parametri legati al funzionamento dei punti luce, incluse la posizione, l'altezza, l'orientamento, la temperatura di colore, ecc. Come nella modalità di monitoraggio e controllo, l'operatore può configurare le regole di illuminazione da utilizzare nella simulazione. Inoltre, al fine di valutare l'impatto della particolare configurazione in termini di comfort, sicurezza, risparmio energetico, ecc. in condizioni realistiche, l'operatore può configurare il comportamento di pedoni e autoveicoli sintetici i quali possono essere movimentati su una rete di traiettorie invisibili. Il comportamento di un singolo pedone o di un gruppo di pedoni può essere configurato in termini di velocità, propensione ad aggregarsi e fermarsi, attraversare la strada, ecc. Infine, è possibile considerare nella simulazione condizioni ambientali variabili, configurando l'intensità di luce nelle diverse ore del giorno, ecc.



Fig.7 - Accesso in mobilità e mediante ambienti virtuali immersivi

4.5 Analisi e statistiche

I pannelli di controllo permettono all'operatore di accedere ai dati di monitoraggio e/o di simulazione. Le informazioni generate nelle suddette fasi sono archiviate nella base di dati. Mediante opportuni filtri, l'operatore può accedere ai dati in tempo reale oppure alle relative informazioni statistiche. In questo modo l'operatore può, ad esempio, verificare l'impatto economico di una specifica configurazione adottata per il particolare sito e individuare pattern di

illuminazione ricorrenti (legati, ad esempio, a determinate condizioni di traffico in una particolare ora del giorno) da utilizzarsi per ottimizzare le regole di illuminazione locale impiegate per il controllo dei punti luce nella rete analizzata. Nella visione del progetto, i dati di simulazione e monitoraggio dovranno essere utilizzati in contemporanea per verificare/dimostrare l'efficacia della soluzione proposta confrontando in maniera numerica siti virtuali e siti reali basati su soluzioni a LED intelligenti o su sistemi/tecnologie di illuminazione tradizionali.

5. Conclusioni

In questo lavoro è stata presentata una piattaforma open source basata su tecnologie web e grafica tridimensionale pensata per supportare la migrazione dei tradizionali sistemi di illuminazione pubblica stradale nella prospettiva delle città digitali. La piattaforma è progettata per consentire il controllo remoto di quattro siti pilota nei quali, nei prossimi mesi dell'anno, saranno installate reti di punti luce intelligenti basate sulla tecnologia sviluppata nell'ambito del progetto LITES. La sperimentazione sul campo permetterà di verificare la sostenibilità della soluzione proposta, verificando le opportunità di impiego della tecnologia LED in ambito urbano e di metodi per la regolazione attiva della potenza.

Ringraziamenti

Il presente lavoro è stato realizzato nel contesto del progetto LITES "Led-based intelligent street lighting for energy saving" (G.A. no. 238916), finanziato dalla Commissione Europea nell'ambito del programma CIP - ICT - PSP.

Bibliografia

Batty, M., Chapman, D., Evans, S., Haklay, M., Kueppers, S., Shiode, N., Hudson-Smith, A., Torrens, P., Visualizing the City: Communicating Urban Design to Planners and Decision-Makers. ESRI Press and Center Urban Policy Research, Rutgers University, Redland, 2001.

Calori, L., Camporesi, C., Pesacrin, S., Virtual Rome: A FOSS approach to WEB3D, in Proc. 14th International Conference on 3D Web Technology, 2010, 177-180.

Chevrier, C., Bur, D., Perrin, J., A simulation system for architectural projects, in Proc. 6th International IBPSA Conference on Building Simulation, 1999, 1085-1092.

Hai, W., Zabuwalla, S., Zeferjahn, N. H., Yadegar, J., Towards a real-time 3D situational awareness visualization for emergency response in urban environment, in Proc. 10th IEEE International Symposium on Multimedia, 2008, 1-8.

Jianchun, H., Jinxing, H., Qingyuan, T., Shanshan, G., Layout optimization of urban underground pipeline based on 3D digital city, in Proc. Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science, 2010, 279-283.

Pack, M., Visualization in transportation: Challenges and opportunities for everyone. IEEE Computer Graphics and Application, 30, 4, 2010, 90-96.

Poullis, C., Suyu, Y., Photorealistic large-scale urban city model reconstruction. IEEE Transactions on Visualization and Computer Graphics, 15, 4, 2009, 654-669.

Location-aware Overlay Scheme for Vehicular Networks

Marco Picone, Michele Amoretti, Francesco Zanichelli, Gianni Conte

Distributed Systems Group – Università degli Studi di Parma

Via Usberti 181a, 43124 Parma

picone.m@gmail.com, {michele.amoretti, francesco.zanichelli, gianni.conte}@unipr.it

In this paper we illustrate a location-aware overlay network called Distributed Geographic Table (DGT) and we show how a vehicular network may take advantage of it. Moreover, we propose a possible implementation based on a novel middleware, namely Sip2Peer.

1. Introduction

An open traffic information platform that collects, processes, analyzes, displays and stores real-time traffic status information from a variety of independent data sources has not been implemented and deployed, yet. To manage a very high number of users would require abundant resources to provide an adequate QoS. If the whole information is managed by centralized servers, system scalability can be hardly guaranteed, unless practically unlimited computational and network resources can be deployed (as feasible for Google, TomTom and few other global service providers).

Granted that users have to agree on allowing their localization and tracking, we foresee location-based services (LBS) supported by an infrastructure implemented as a network of distributed software entities, with flat or hierarchical organization, but always without relying on a central entity. Any centralized approach would hardly scale for such highly variable contexts, with mobile nodes changing their location very quickly, in a wide and highly populated area. Indeed, search results would be incomplete or outdated with high probability.

Like other authors (see the Related Work section) we believe that a partially or fully decentralized approach can increase the accuracy of information and the rate at which they are retrieved by users. Moreover, it may allow to update and publish information directly, with low cost and high scalability. Last but not least, it may simplify the process of joining the virtual community and publish new services.

In this paper we describe a general framework called Distributed Geographic Table (DGT), that defines a peer-to-peer strategy for mobile node localization, and a particular instance that supports applications in which every node requires to be constantly updated about its neighbors. Compared to a centralized approach, DGT is more scalable, since its performance (in terms of

responsiveness, completeness and robustness) remains valuable also for a large number of nodes and when the nodes' dynamics are very high.

The rest of the paper is organized as follows. In section 2 we discuss related work. In section 3 we illustrate the DGT framework. In section 4 we describe a possible implementation strategy, based on a novel middleware. Finally, in section 5 we conclude the paper with proposals for future work.

2. Related Work

An almost complete overview of existing and emerging technologies and solutions for distributing and aggregating sensor data in vehicular networks has been proposed by Uichin and Gerla [Lee2010]. An example is the CarTel mobile sensor computing system [Hull2006], that allows to collect, process, deliver and visualize data retrieved from sensors placed on mobile units such as vehicles. MobEyes [Lee2006] proposes a strategy for harvesting, aggregating and distributing sensed data, by means of proactive urban monitoring services provided by vehicles that continuously discover, maintain and process information about events of the urban scenario. Messages and summaries are routed to vehicles in the vicinity, to achieve common goals, such as providing police cars with the trajectories of specific "target" cars.

A fundamental issue for VSNs is connectivity: different wireless access and communication methods have been evaluated, including Dedicated Short-Range Communication (DSRC), WiMax/802.16e, WLAN, as well as cellular systems. The usage of cellular communication network reduces the problem of implementing a working traffic information system (TIS), but introduces, on the other side, the issue of collecting data and distribute them to interested users.

A common approach is based on the client/server paradigm, where all data generated by vehicles are stored in a central server or a server farm on the Internet. Hull et al. in [Hull2006] pointed out the major technical challenges of this solution, that are mostly related with the huge amount of simultaneous updates and queries generated by movements and requests of users (each car is a source of queries and sends its own measurements regularly).

For these reasons, in recent years researchers started investigating architectures based on the peer-to-peer (P2P) paradigm, to build a distributed TIS where cars are not only consumers but also producers of information. Rybicki et al., with PeerTIS [Jedrzej2009], have shown P2P architectures where participating cars are peers organized in a distributed hash table (DHT). The main idea is that each node is responsible for a certain part of the ID space, and consequently for a certain number of road segments. Up to now one of the unsolved issue is that obtaining full information about planned and alternative routes is expensive in terms of bandwidth consumption. Our approach can cope with this problem.

3. Distributed Geographic Table

A decentralized P2P overlay is structured (i.e. based on the Decentralized Structured Model DSM) if its topology is controlled and shaped in a way that resources (or resource advertisements) are placed at appropriate locations [Amoretti, 2009].

The Distributed Geographic Table (DGT) is a structured overlay scheme where each participant can efficiently retrieve node or resource information (data or services) located near any chosen geographic position. In such a system the responsibility for maintaining information about the position of active peers is distributed among nodes, for which a change in the set of participants causes a minimal amount of disruption.

In a DGT, the *distance* between two nodes is defined as the actual geographic distance between their locations in the world. The *neighborhood* of a geographic location is the group of nodes located inside a given region surrounding that location.

The main service provided by the DGT overlay is to route requests to find available peers in a specific area, i.e. to determine the neighborhood of a generic global position. The routing process is based on the evaluation of the region of interest centered in the target position. The idea is that each peer involved in the routing process selects, among its known neighbors, those that presumably know a large number of peers located inside or close to the chosen area centered in the target point. If a contacted node cannot find a match for the request, it does return a list of closest nodes, taken from its routing table. This procedure can be used both to maintain the peer's local neighborhood and to find available nodes close to a generic target.

Regarding the local neighborhood, the general aim of the approach is to have high knowledge of nodes that are close to the peer and a gradually reduced number of known nodes that will be used to forward long range geographic queries.

The DGT addresses mobile peers that frequently/continuously change their geographic position, and may also join and leave the network at any time. As a consequence, the neighborhood of a peer is usually characterized by high dynamism. To preserve the consistency of the DGT, each peer needs to periodically schedule a maintenance procedure of its local neighborhood that compensates the topological changes of the network. The practical usability of the DGT critically depends on the messaging and computational overhead introduced by such a maintenance procedure, whose features and frequency of execution are application-dependent.

A very important issue in real-time localization concerns security and privacy. In DGT, the only data that are shared among peers are their unique identifiers, their IP addresses and ports, as well as their GPS coordinates. At the DGT level no reference exists to sensitive data (e.g. MAC address) that may allow to identify a node's owner. The capability of finding peers that are active and close to a specific geographic position is obtained without sharing any personal data. Sensitive or potentially dangerous information may be added by applications built over the DGT, that may store such data in their data structures, but this

problem potentially affects any P2P scheme and has been addressed by several research works [Balfe2005], [Aringhieri2006], [Kesar,2003], [Kher2005].

3.1 Peer Neighborhood Construction

In this section we illustrate the protocol that builds and maintains the overlay, where each node knows all the active nodes that are available around its GP, in order to provide and realize specific applications and services. An example of application based on such a protocol may be a city monitoring system that uses smartphones of users that move using their vehicles, to monitor the traffic status of the city. By using this system there is no need to deploy powerful servers: mobile peers may cover the whole city area (Fig. 1). Each of them can analyze its region of interest, monitor traffic conditions in real-time, and evaluate the position of peers in order to inform them about accidents and traffic jams, suggesting alternative paths.



Fig.1 – Neighborhood of a vehicle (green marker). Blue markers indicate actually existing nodes and red markers indicate discovered nodes.

Every peer maintains a set of *geo-buckets* (GB), each one being a (regularly updated) list of known peers sorted by their distance from the GP of the peer itself. GBs can be represented as concentric circles, each one having a different (application-specific) radius and thickness. Another list contains the nodes outside the circular area covered by the GBs.

Each peer in the GB set is characterized by

- *Unique ID* - univocally identifies the peer within the DGT;
- *Global Position GP* - latitude and longitude retrieved with a GPS system or with other solutions (e.g. GSM cell-based localization);
- *IP Address* - allowing to identify the node in Internet; if the peer is behind NAT, the IP address may be that of a relay;

- *UDP Port* - on which the peer listens, waiting for connection attempts;
- *Number of known nodes* - used to compare two nodes at the same distance.

3.1.1 Network Join

When a new peer wants to enter the network, it sends a join request, together with its GP, to a bootstrap node, that returns a limited list of references to peers that are geographically close to the joining one. It is important to emphasize that this information is not updated: referenced peers may have moved away from their initial location. It is up to the joining peer to check for the availability of listed peers. This operation is performed not only during the first join of the peer, but also when the peer finds itself to be almost or completely isolated. In these situations (that typically arise when peers enter low density areas), the node may send a new join request to the bootstrap node, in order to obtain a list of recently connected peers that may become neighbors.

3.1.2 Peer Lookup

The main procedure used during peer discovery is `findNodes(GP)`, that returns the β peers that are nearest to the specified global position (GP). By periodically applying `findNodes()` to GP_n , peer n keeps up to date its neighborhood awareness. Such a procedure (with any target GP) may also be executed upon request from another peer.

Node n searches in the geo-bucket associated to the requested GP. The final objective of the lookup is to find the $\alpha < K$ peers that are nearest to the selected GP, including newly connected nodes, as well as mobile peers that have entered the visibility zone. The lookup initiator starts by picking α nodes from its closest non-empty geo-bucket or, if that bucket has less than α entries, it just takes the α closest nodes, by extending the lookup to all its geo-buckets). Such a peer set is denoted as $C_i = \{n1_i, \dots, n\alpha_i\}$ where i is an integer index. The initiator sends parallel `findNodes` requests, using its GP as target, to the α peers in C_i . Each questioned peer responds with β references. The initiator sorts the result list according to the distance from the target position, then picks up α peers that it has not yet queried and re-sends the `findNodes` request (with the same target) to them. If a round of `findNodes` fails to return a peer closer than the closest already known, the initiator re-sends the `findNodes` to K closest nodes it has not already queried. The lookup terminates when the initiator has obtained responses from the K closest nodes, or after f cycles, each cycle resulting with an updated set of nearest neighbors. Thus, the number of find messages is $f\alpha + K$, that depends on the density of peers in the area of interest. A peer is allowed to run a new lookup procedure only if the previous one is completed, in order to reduce the number of exchanged messages and avoid the overlapping of same type of operation.

3.1.3 Position update

Any peer (let us call it A) active in the network can change its geographic position, for many reasons (the user may be walking, driving, etc.). In order to maintain a good level of knowledge accuracy within the network, each node notifies its GP updates to neighbors. In order to reduce the computational and bandwidth impact of this operation, before sending a position update, peer A performs the following two operations, involving every node (called B, for simplicity) in its GBs:

- Peer A checks for the distance between itself and peer B. If such a distance is larger than the radius of the circular area covered by the GBs, it means that peer B is out of the visibility area of A, and for this reason it removes B from its GB and sends to it a RemoveMessage, in order to notify the removing operation. This action is very important because, if peer B does not receive this specific message, it keeps the reference of A in its GBs, but it does not receive new updates because A removed B from its GBs (Fig.2).
- Peer A checks for $d(A) > \epsilon$ (i.e. its displacement is larger than ϵ). If the condition is true, A sends its position update to its neighbors. Parameter ϵ allows to define the accuracy of update messages and can be configured according to application requirements. A small value of ϵ causes a high rate of exchanged messages, but a large value reduces the accuracy of the peer's knowledge and damages the global performance of the protocol.

There is another important aspect related to position updating. In order to improve the performance during the join procedure, each peer sends update message to the bootstrap node if the distance between the peer's actual position and the one that it had when it entered the network has become larger than λ . Such an update is performed only if the peer is moving far from its original area, and helps the bootstrap node to provide more precise information to newcomers (Fig.3).

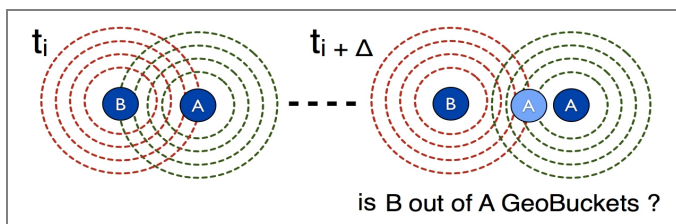


Fig.2 - While moving, peer A finds peer B to be too much far, thus removes it from its GBs, and notifies the operation B

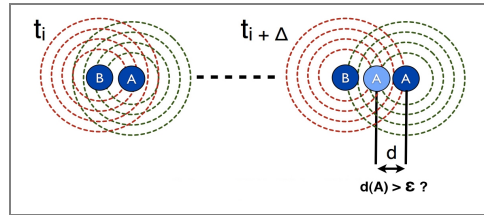


Fig.3 - After a movement peer A checks if it is necessary to notify the new position to its neighbors

3.1.4 Gossiping

To improve the results and the performance of our protocol we added gossip information inside exchanged messages. This approach on one side increases the size of sent packets, but at the same time adds significant knowledge that helps peers to be aware of available nodes in their area. Each peer maintains the references to nodes that have been discovered during the time between two different lookup procedures.

4. Sip2Peer-based implementation

The simulative analysis of sample scenarios based on experimental measurements of coverage and connection throughput, carried out across/around Parma urban area, has given us valuable insights for defining an integrated model that will combine the DGT, user/vehicle mobility and connectivity/coverage types [Picone2010].

Currently we are working on a rich, standard-based software implementation of the DGT. To this purpose, we are using our novel peer-to-peer middleware called Sip2Peer, that is an open-source SIP-based middleware for the implementation of any peer-to-peer application or overlay without constraints on peer nature (traditional PC or mobile nodes) and specific architecture. At this moment Sip2Peer is available for Java SE and Android platforms, but we are working on the iOS release.

The Java and the Android implementation of the library are based on the java SIP stack called MjSip (<http://www.mjsip.org>) that allows to manage the exchange of SIP messages or multimedia streams.

Sip2Peer supports two message formats. It is possible to manage simple text message containing any kind of information like raw data or XML and it is also possible to natively use JSON format. Following this scalable approach, the main class of the Sip2Peer API, i.e. Peer, provides all necessary methods for sending and receiving messages, allowing the developer to select the best solution according to his/her protocol and overlay.

We have designed several modules whose integration defines the behavior of a DGT Peer. Such modules are presented in Fig.4 and described in the following.

- **Sip2Peer Layer (SL):** Represents the communication module providing listeners and methods to receive and send messages from and to other active peers in the network. This layer interacts with the DMH (illustrated below) to route and forward outgoing and incoming messages, providing proper notifications when a packet has been correctly delivered or not.
- **DGT Message Handler/Dispatcher (DMH):** Conveys and manages all DGT messages. It does notify neighbors position updates and redirect subscriptions and event messages from and to the Subscription Manager (detailed below).
- **Geo-Bucket Manager (GM):** Manages the data structures of the peer according to its geographic location. This module interacts with the Location Manager to be notified for a position update, and thanks to DMH notifications it is able to add new discovered peers or remove nodes out of the region of interest.
- **Location Manager (LM):** Subscribes to information and updates about device location through GPS, WiFi or cellular network, trying to minimize the energy consumption according to the context application characteristics.
- **DGT Kernel (DK):** Is the core of a DGT node, implementing the routing strategy and the discovery procedure for neighborhood maintenance, as well as short/long range queries triggered by the user (through the User Interface) or periodically efficiently scheduled by the DK thread according to the application purpose and users settings. It also allows the interaction with DMH and GM to properly disseminate messages and alerts coming from UI or other external inputs.
- **Subscription Manager (SM):** Manages the subscription system of the node, allowing to add or remove subscriptions and handling an filtering incoming events or user queries. It does interact with the UI to notify relevant incoming alerts or messages, and sets preferences about the subscriptions.
- **User Interface (UI):** Presents to the user all required information and interface elements to interact and control DGT functionalities, like dissemination an alert messages about traffic jams or to schedule a query concerning a region of interest. It allows to visualize on a map (or in a dedicated list view) peer/vehicle and neighbor locations, as well as scheduled query results.

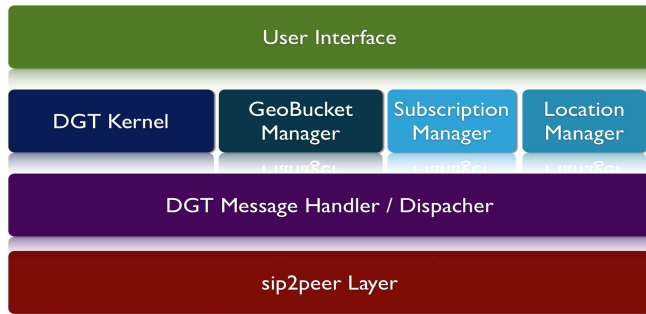


Fig.4 – Sip2Peer-based layered implementation of the DGT.

Fig. 5 illustrates the procedure for joining the network. An Android DGT-Peer sends a request to the Bootstrap peer, that replies with a list of recently connected nodes. The DGT-Peer uses such list to build its initial neighborhood.

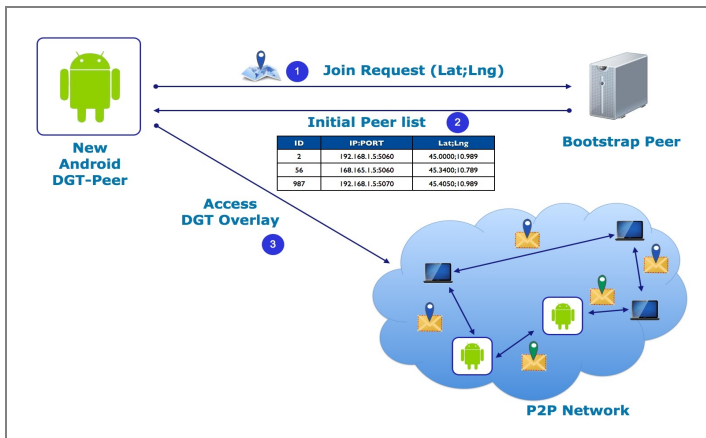


Fig.5 – Joining the DGT network.

5. Conclusions

In this paper we have illustrated the concept of Distributed Geographic Table (DGT), which refers to a decentralized system that allow to retrieve nodes or services located near any chosen geographic position using a distributed responsibility for maintaining information about the positions of active users. We have also presented the particular instance of DGT, based on the peer-to-peer paradigm, that allow peers to localize all available nodes near to their geographic position enabling distributed and low cost application like city and

traffic monitoring. In the last part of the paper we have illustrated the design of the Sip2Peer-based implementation of the DGT architecture.

As future work, we plan to improve DGT protocols, by taking into account the estimation of peer trajectory (e.g. vehicles traveling along highways) in order to reduce the number of exchanged messages. Moreover, we are going to complete the implementation and start the test phase in different scenarios, both urban and extra-urban.

Bibliografia

[Amoretti2009] M. Amoretti, A survey of peer-to-peer overlay schemes: effectiveness, efficiency and security, in Bussum (Ed.), Recent Patents on Computer Science, Vol. 2 No. 3, pp. 195-213, Ed. Bentham Science.

[Amoretti2009b] M. Amoretti, M. Agosti, F. Zanichelli, DEUS: a Discrete Event Universal Simulator, 2nd ICST/ACM International Conference on Simulation Tools and Techniques (SIMUTools 2009), ISBN 978-963-9799-45-5, Roma, Italy, March 2009.

[Aringhieri2006] Aringhieri, R., Damiani, E., de Capitani di Vimercati, S., Paraboschi, S. and Samarati, P. (2006), Fuzzy techniques for trust and reputation management in anonymous peer-to-peer systems, Journal of the American Society for Information Science and Technology, Vol. 57 No. 4, pp. 528-37.

[Balfe2005] Balfe, S., Lakhani, A.D. and Paterson, K.G. (2005), Trusted computing: providing security for peer-to-peer networks, Proceedings of the 5th IEEE International Conference on Peer-to-peer Computing (P2P2005), Konstanz, Germany, August 31-September 2.

[Hull2006] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, S. Madden, Cartel: a distributed mobile sensor computing system, 4th ACM Conference on Embedded Networked Sensor Systems (SenSys), Boulder, Colorado, USA, November 2006.

[Jedrzej2009] R. Jedrzej, B. Scheuermann, M. Koegel, M. Mauve, PeerTIS: a peer-to-peer traffic information system. International Conference on Mobile Computing and Networking (MobiCom), Beijing, China, 2009.

[Kesar2003] R. Kesar, Experimental games for the design of reputation management system, IBM Systems Journal, Vol. 42 No. 3, pp. 498-506.

[Kher2005] S. Kher, A.K. Somani, R. Gupta, Network selection using fuzzy logic, Proceedings of the IEEE Second International Conference on Broadband Networks (BROADNETS 2005), Boston, MA, October 3-7.

[Lee2006] U. Lee, E. Magistretti, B. Zhou, M. Gerla, P. Bellavista, A. Corradi, MobEyes: smart mobs for urban monitoring with vehicular sensor networks, IEEE Wireless Communications ,13 (5), pp. 51-57, 2006.

[Lee2010] U. Lee, M. Gerla, A survey of urban vehicular sensing platforms, Computer Networks, Vol. 54, no. 4: 527-544, March 2010.

[Picone2010] M. Picone, M. Amoretti, F. Zanichelli, Proactive Neighbor Localization Based on Distributed Geographic Table, Proc. of the 8th ACM-SIGMM International Conference on Advances in Mobile Computing & Multimedia (MoMM 2010), Paris, France, November 2010.

Misurare Processi di Business

Roberto Bruni¹, Andrea Corradini¹, Gianluigi Ferrari¹, Tito Flagella²,
Roberto Guanciale¹, Giorgio Oronzo Spagnolo¹

¹ Dipartimento di Informatica, Università di Pisa
Largo B. Pontecorvo 3, 56127 Pisa
[bruni, andrea, giangi, guancio, spagnolo]@di.unipi.it

² Link.it
Via San Martino 51, 56125 Pisa
tito@link.it

La notazione BPMN si sta affermando come standard per la descrizione di processi di business ed è supportata da molti strumenti di sviluppo. La distanza tra la modellazione astratta dei processi in BPMN e la loro realizzazione concreta necessita di strumenti per l'analisi di conformità (tra implementazione e modello) e delle prestazioni (per evidenziare le criticità e guidare la reingegnerizzazione dei processi). In questo articolo descriviamo alcune tecniche per il monitoraggio e la valutazione di processi di business dopo il loro deployment. L'analisi viene effettuata sui log di esecuzione e si basa sul modello formale delle reti di Petri. L'uso delle reti di Petri come modello intermedio tra log e processo BPMN permette di proiettare i risultati sul modello astratto, rendendoli di immediata comprensione per gli analisti. In particolare, presentiamo delle estensioni originali di algoritmi noti che offrono una descrizione più fedele dei fenomeni osservati durante l'esecuzione. Gli algoritmi presentati sono disponibili come plug-in di ProM.

1. Introduzione

Il Business Process Management (BPM) è una disciplina di recente affermazione che stabilisce un cambio di paradigma, da una gestione centrata sui dati a una centrata sui processi, ovvero da “cosa” a “come”. Infatti, BPM affronta la modellazione, l'organizzazione, l'applicazione e l'ottimizzazione delle attività necessarie per raggiungere un determinato obiettivo (es. offrire un determinato servizio, oppure produrre un certo manufatto). In BPM, i processi vengono rappresentati attraverso formalismi grafici, permettendo di comunicare in modo non ambiguo le regole di business, e quindi discuterle o modificarle, tra gli svariati ruoli coinvolti che vanno dagli esperti del dominio di business o del settore, agli architetti software e sviluppatori.

Negli ultimi anni sono state proposte diverse notazioni grafiche, spesso sponsorizzate da grandi consorzi industriali, supportate da varie piattaforme e integrate nei principali ambienti di sviluppo. Tra le notazioni più diffuse citiamo

Event-driven Process Chains (EPC) [Scheer, 1993], Business Process Execution Language (BPEL) [OASIS, 2007] e Business Process Modeling Notation (BPMN) [OMG, 2011], che è diventato recentemente lo standard più diffuso. Il successo di queste notazioni è spesso legato alla loro facilità di utilizzo, alla comprensione intuitiva delle forme grafiche offerte, all'assenza di vincoli strutturali e alla maggiore flessibilità. Come controparte, per favorire questi aspetti si tollera la possibilità di tracciare diagrammi inconsistenti o con ambiguità semantiche.

Dato che i modelli di processi vengono poi tradotti in applicazioni software, la distanza tra il modello astratto e la sua realizzazione è tale da richiedere strumenti di monitoraggio e analisi per monitorare la conformità dell'implementazione al modello concettuale e le prestazioni raggiunte. A questo scopo, diversi modelli formali sono stati sviluppati ed adottati per fornire una semantica rigorosa agli standard industriali (come il π -calculus [Milner et al, 1992], ASM [Börger e Thalheim, 2008], le reti di Petri [Petri, 1962] e in particolare le Workflow Net [van der Aalst, 1998]). Infatti, la descrizione non ambigua del comportamento di un processo è un prerequisito fondamentale per ogni algoritmo o tecnica di analisi.

In questo contributo ci focalizziamo su una specifica fase del BPM, che comprende il *monitoraggio* e la *valutazione*. L'obiettivo di questa fase consiste nel verificare la corretta esecuzione dei processi e misurarne le prestazioni dopo il loro deployment. I risultati di questa fase sono prerequisiti di eventuali ottimizzazioni o re-ingegnerizzazioni dei processi, permettendo di individuare discrepanze tra il processo pianificato e la sua realizzazione e di calcolare parametri (come la latenza e i tempi di sincronizzazione) necessari per dimensionare le risorse di business (per esempio i gruppi di lavoro allocati per le attività). I dati necessari a queste analisi sono di solito forniti dall'infrastruttura di supporto attraverso i log di eventi generati a tempo di esecuzione. Dal punto di vista metodologico, queste analisi richiedono che 1) sia disponibile un modello formale del processo; 2) tutte le attività (di business) importanti siano registrate nei log; 3) gli eventi dei log siano correlati se riguardano la stessa istanza di processo o tenuti separati se riguardano istanze differenti; 4) gli eventi dei log siano ordinati temporalmente.

In questo contributo adottiamo BPMN come linguaggio per descrivere i processi e le reti Petri come modello formale. L'uso delle reti di Petri è giustificato dalla ampia disponibilità di strumenti e tecniche di analisi esistenti. In particolare abbiamo adottato ed esteso la tecnica chiamata *log replay* e implementato le nostre tecniche come plug-in di ProM [Eindhoven Univ. of Technology, 2011]. Nella prossima sezione riassumiamo i principali costrutti di BPMN attraverso l'introduzione di un esempio; inoltre, presentiamo una trasformazione di modelli BPMN in reti di Petri, necessaria ad abilitare le analisi formali. Gli algoritmi di analisi per reti di Petri e la proiezione dei risultati sui modelli BPMN di partenza sono presentati nella sezione 3. Il contributo innovativo è riassunto nelle conclusioni.

Le attività di ricerca qui descritte sono state realizzate nell'ambito del progetto RUPOS (*Ricerca sull'Usabilità delle Piattaforme Orientate ai Servizi*, [Link.it, 2011]) con il parziale supporto della Regione Toscana. La rilevanza di queste tematiche è testimoniata dalla recente istituzione di una Task Force dell'IEEE dedicata a process analysis e process mining [IEEE, 2011].

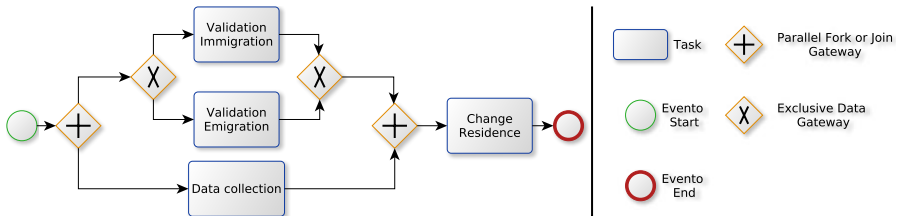


Fig. 1 - modello BPMN per il cambio di residenza

2. Supporto alla modellazione tramite BPMN

BPMN è un linguaggio per modellare processi di business attraverso un'astrazione grafica. In questo articolo non introduciamo formalmente tutti i costrutti di BPMN, ma ne presentiamo i principali attraverso un esempio. Il processo BPMN in Fig. 1 sintetizza le regole necessarie per ottenere un cambio di residenza. Il processo coinvolge quattro distinte **attività** (o **task**): *Validation: Immigration*, *Validation: Emigration*, *Data collection* e *Change Residence*. I cerchi con bordo sottile rappresentano l'inizio del processo, quelli con bordo di spessore maggiore la sua terminazione. I rombi etichettati con **+** sono chiamati **fork/join gateway**. Nell'esempio, il primo rombo descrive un fork dell'esecuzione; *Data collection* è eseguito concorrentemente alle attività di validazione. Il secondo rombo **+** rappresenta invece un join: le attività concorrenti devono essere completate prima di attivare *Change Residence*. I rombi etichettati con **X** sono chiamati **split/merge gateway** e rappresentano una scelta tra due esecuzioni alternative: la procedura di validazione è differente se il cittadino richiede di immigrare o di emigrare. Nel seguito useremo questo esempio per presentare le tecniche sviluppate.

Molti algoritmi di analisi (come quelli discussi nelle prossime sezioni) non sono applicabili direttamente a modelli BPMN, ma sono definiti su reti di Petri. Pertanto definiamo una trasformazione da modelli BPMN a reti di Petri, e successivamente affrontiamo il problema di riportare i risultati di queste analisi sul modello BPMN di partenza. Per trasformare (un sottoinsieme di) modelli BPMN in reti di Petri adottiamo la metodologia presentata in [Dijkman et al, 2008]: le regole di trasformazione corrispondenti sono presentate in Fig. 2. Abbiamo esteso la trasformazione per trattare esplicitamente l'attivazione e la terminazione di attività BPMN. La regola di trasformazione introdotta rappresenta questi eventi attraverso due distinte transizioni. La rete di Petri in Fig. 4 (vedere anche Fig. 7) è ottenuta applicando queste trasformazioni al diagramma BPMN di Fig. 1.

Nel rappresentare la trasformazione abbiamo usato due simboli differenti per due classi di transizioni: (i) le transizioni relative ad attività BPMN sono rappresentate da quadrati non riempiti e sono chiamate **transizioni visibili**; (ii) le transizioni generate per modellare gli altri elementi di BPMN sono rappresentate da rettangoli riempiti e sono chiamate **transizioni invisibili**. Nel proseguo dell'articolo assumiamo che i log dei sistemi contengano solamente gli eventi correlati all'attivazione ed al completamento delle attività.

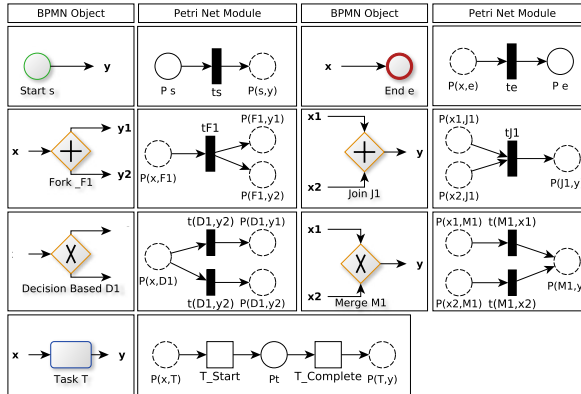


Fig. 2 - Mapping da BPMN a reti di Petri

3. Analisi basata su reti di Petri e log di esecuzioni

Le tecniche presentate in questa sezione permettono di analizzare proprietà di un processo (espresso come rete di Petri) attraverso l'elaborazione di un log di esecuzione. Un log è costituito da una sequenza di eventi, cioè di coppie del tipo $e = (a, t)$ dove a è un'azione registrata dal sistema e t è il suo timestamp (che indicheremo a volte con $\theta(e)$). Gli eventi che appartengono alla stessa istanza di processo sono raggruppati in tracce. Formalmente, una traccia T è una sequenza di eventi $T[1], \dots, T[n]$ tale che $\theta(T[i]) \leq \theta(T[i + 1])$ per tutti gli $i \in [1, n]$. Un log L è un insieme di tracce, cioè la registrazione delle attività svolte da un sistema durante un numero finito di esecuzioni di un processo. Assumiamo che tutte le tracce siano istanze dello stesso processo e che per ogni azione a esista una transizione visibile con lo stesso nome nella rete.

L'algoritmo chiave della nostra metodologia si chiama **log-replay** [Rozinat e van der Aalst, 2008]. Per ogni traccia T , l'algoritmo parte con un token nella piazza iniziale delle rete e per ogni evento (a, t) in T esegue la corrispondente transizione a *in modo non bloccante*, aggiornando il marking della rete di conseguenza. Il termine “*non bloccante*” significa che se la transizione a non è abilitata, i token mancanti vengono creati artificialmente e chiamati **missing token**.

Come descritto nella sezione 2, assumiamo che i log contengano solamente gli eventi correlati all'inizio e al completamento delle attività del processo BPMN. Questo comporta che le transizioni invisibili non siano associate a nessun evento registrato nel log. Di solito, le transizioni invisibili sono considerate *lazy* [van der Aalst et al, 2005], cioè la loro esecuzione è ritardata finché non diventa necessaria all'abilitazione di una transizione visibile corrispondente ad un evento della traccia. Noi introduciamo invece una gestione *eager* delle transizioni invisibili.

Il risultato del log-replay di una traccia T può essere rappresentato da una lista R di coppie (tr, i) , ognuna delle quali rappresenta l'esecuzione della transizione tr per mimare l'evento $T[i]$. La presenza di transizioni invisibili può risultare in più transizioni eseguite per replicare un singolo evento, mentre la presenza di ricorsione può risultare in più occorrenze della stessa transizione.

3.1 Analisi di performance

Sfruttando i timestamp associati agli eventi, la tecnica di log-replay può essere usata per misurare le prestazioni del sistema [van der Aalst e van Dongen, 2002]. L'idea chiave è quella di calcolare l'intervallo di tempo tra produzione e consumo di token in ogni piazza della rete. Questa tecnica può essere applicata solo per le tracce che non richiedono la produzione di missing token, perché tali token non possono avere informazioni temporali. Durante il log-replay possono essere calcolate, per ogni traccia, le seguenti metriche riferite a ciascuna piazza: il tempo di soggiorno ($\bowtie$), cioè l'intervallo di tempo tra l'arrivo e la partenza di un token; il tempo di sincronizzazione ($\times$), cioè l'intervallo di tempo tra l'arrivo di un token in una piazza e l'abilitazione di una transizione nel suo post-set; e il tempo di attesa ($\bowtie$), cioè l'intervallo di tempo tra l'attivazione e l'esecuzione di una transizione nel post-set della piazza (in generale vale $\times + \times = \bowtie$).

3.1.1 Log-replay con transizioni invisibili “eager”

Individuate queste misure da analizzare, e motivati dalla necessità di proiettare tali misure sul modello BPMN di partenza, abbiamo sviluppato un nuovo algoritmo per valutare le performance. Intuitivamente, prima di associare ad ogni piazza le misure, effettuiamo una rielaborazione della sequenza di transizioni ottenuta dal log-replay, in modo da eseguire le transizioni invisibili *il prima possibile*. Di seguito chiamiamo questo tipo di sequenza **eager**. In generale una sequenza è eager se per ogni transizione invisibile $t1$, l'ultima transizione visibile $t2$ che precede $t1$ è necessaria per abilitare $t1$. Il seguente algoritmo trasforma una generica sequenza R di coppie (tr, i) ottenuta dal log-replay in una corrispondente sequenza eager. Con la notazione $R \downarrow k$ indichiamo l'indice (tra 1 e $k - 1$) dell'ultima transizione visibile che precede la posizione k , oppure 0 se tale transizione non esiste.

```

for  $j$  from 2 to size( $R$ ) do
   $tr, i \leftarrow R[j]$ 
  if  $tr$  is invisible then
     $k, done \leftarrow j - 1, false$ 
    while  $k > 0 \wedge \neg done$  do
       $tr_{prev}, i_{prev} \leftarrow R[k]$ 
      if  $tr \cap tr_{prev} \neq \emptyset$  then
         $done \leftarrow true$ 
      else
         $R[k + 1] \leftarrow (tr_{prev}, i_{prev})$ 
         $R[k] \leftarrow (tr, max(R \downarrow k, 1))$ 
         $k \leftarrow k - 1$ 
      end if
    end while
  end if
end for

```

Per esemplificare le metriche calcolate da questa tecnica usiamo la rete di Petri in Fig. 4 e la traccia T in Fig. 3.

Gestendo le transizioni invisibili come *lazy*, il log-replay produce la sequenza $R1 = [(t1, 1), (t2, 1), (DCS, 1), (t3, 1), (IS, 2), (IC, 3), (DCC, 4), (t5, 4), (t7, 4), (CRS, 5), (CRC, 6)]$.

Benché la transizione $t3$ sia abilitata dopo l'evento *DataCollection*, la sua esecuzione è ritardata finché non diventa necessaria per l'attivazione della transizio-

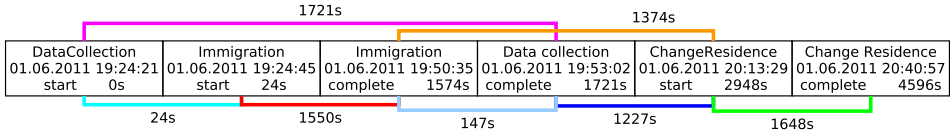


Fig. 3 - Esempio di Log per l'analisi di performance

ne visibile *Immigration*, quindi il token resta nella piazza $p1$ per $24s$ prima che la transizione $t3$ sia eseguita. Analogamente, nonostante la transizione $t5$ sia abilitata dopo l'evento *Immigration*, la sua esecuzione è ritardata finché non diventa necessaria per l'attivazione della transizione visibile *ChangeResidence*, quindi il token resta nella piazza $p4$ per $1374s$ prima che la transizione $t5$ sia eseguita. L'unico tempo di sincronizzazione non nullo è quello in $p6$ ($1227s$) e misura l'intervallo tra l'esecuzione delle transizioni *DCC* e *CRS*, che non sono le due attività concorrenti. È nostra opinione che queste misure non descrivano l'evoluzione del processo in modo preciso.

La sequenza $R1$ viene trasformata dal nostro algoritmo nella sequenza eager $R2 = [(t1, 1), (t2, 1), (t3, 1), (DCS, 1), (IS, 2), (IC, 3), (t5, 3), (DCC, 4), (t7, 4), (CRS, 5), (CRC, 6)]$. La valutazione della performance viene eseguita inserendo un token nella piazza iniziale al tempo 0.

- $t1$ e $t2$ sono le prime due transizioni eseguite e sono entrambe associate all'evento $T[1]$. Quindi, al tempo $\theta(T[1]) = 0$ il token nella piazza iniziale viene consumato e due token prodotti (uno in $p1$ e uno in $p2$) per attivare due esecuzioni concorrenti.
- La transizione $t3$ consuma il token in $p1$ e ne produce uno in $p3$ al tempo $\theta(T[1]) = 0$, attivando una delle due possibili validazioni. Quindi $\times(p1) = 0$;
- La transizione *DCS*, anch'essa eseguita al tempo $\theta(T[1]) = 0$, consuma il token in $p2$ e produce un token in pdc .
- Al tempo $\theta(T[2]) = 24s$ viene eseguita la transizione *IS*, consumando il token in $p3$ e producendone uno in pi . Quindi, $\times(p3) = 24s - 0s = 24s$
- Al tempo $\theta(T[3]) = 1574s$ viene eseguita la transizione *IC*, consumando il token in pi e producendone uno in $p4$. Quindi, $\times(pi) = 1574s - 24s = 1550s$;
- Sempre al tempo $\theta(T[3]) = 1574s$ viene eseguita la transizione invisibile $t5$, consumando il token in $p4$ e producendone uno in $p5$. In questo momento $p5$ non è preset di nessuna transizione attiva ($t7$ dipende dalla piazza vuota $p6$) quindi il suo tempo di sincronizzazione sarà maggiore di zero.
- Al tempo $\theta(T[4]) = 1721s$ viene eseguita la transizione *DCC*, consumando il token in pdc e producendone uno in $p6$. Quindi, $\times(pdc) = 1721s - 0s = 1721s$. In questo momento la transizione $t7$ viene abilitata ed è possibile calcolare i tempi di sincronizzazione dei preset: $\times(p5) = 1721s - 1574s = 147s$ e $\times(p6) = 1721s - 1721s = 0s$
- L'analisi delle ultime tre transizioni permette di calcolare le misure di performance per le rimanenti piazze.

In Fig. 4 è possibile vedere le misure di performance calcolate. Tutte le piazze in bianco hanno tempo di sincronizzazione e di attesa uguale a zero. Il tempo di

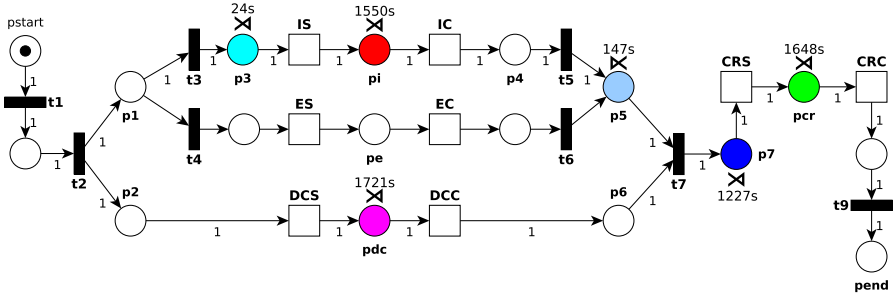


Fig. 4 - Rete di Petri del cambio di residenza con annotata l'analisi di performance

attesa delle piazze pi , pdc , pcr , rappresentano il tempo di esecuzione delle corrispondenti attività BPMN. La sola piazza con tempo di sincronizzazione maggiore di zero è il $p5$ (147s). Questo tempo misura l'intervallo tra l'esecuzione delle transizioni IC e DCC . È importante notare che il tempo di sincronizzazione nella piazza $p5$ è ottenuto grazie alla trasformazione della sequenza $R1$ nella sequenza eager $R2$ (che sposta la transizione $t5$ immediatamente dopo IC).

3.1.2 Proiezione dei risultati di analisi sul modello BPMN

La trasformazione delle sequenze del log-replay in sequenze eager semplifica la proiezione sul modello BPMN dei risultati calcolati. Se una transizione invisibile è abilitata questa viene immediatamente eseguita. Per questo motivo, i tempi di attesa delle piazze che hanno solo transizioni invisibili nei post-set sono sempre nulli. Facendo riferimento alle regole della Fig. 2, le piazze che possono avere tempi di attesa diversi dallo zero sono $P(x, T)$ e P_t , cioè le piazze usate per tradurre le attività BPMN. Allo stesso modo, il tempo di sincronizzazione di una piazza può essere maggiore di zero solo se almeno una transizione nel suo post-set dipende da un'altra piazza. Ne segue che le piazze in Fig. 2 che possono avere tempi di sincronizzazione diversi da zero sono solo $P(x1, J1)$ e $P(x2, J1)$, cioè quelle coinvolte nella modellazione del join gateway.

Sulla base di queste considerazioni si possono riportare le misure di performance della rete di Petri sul modello originale BPMN come segue:

- per ogni task T il tempo di esecuzione è $\times(P_t)$ e il tempo di attivazione è $\times(P(x, T))$
- per ogni ramo concorrente ($i \in [1, 2]$) racchiuso da un fork gateway ($F1$) e un join gateway ($J1$): (i) il tempo di sincronizzazione è $\times(P(x_i, J1))$, ovvero il tempo impiegato per attendere il completamento delle attività concorrenti (ii) i tempi di esecuzione (di seguito $\times(F1i)$) sono la somma di $\times$ di tutte le piazze raggiungibili attraversando il grafico a partire da $P(Fi, yi)$ senza aver visitato $P(xi, Ji)$. Si noti che $\times(P(x_i, J1)) + \times(F1i)$ è costante per ogni ramo di un fork/join. In seguito chiamiamo questa misura come il “tempo di esecuzione” del fork, e lo indicheremo come $\bowtie(F1, J1)$.

La Fig. 5 mostra la proiezione delle analisi di performance sul modello BPMN. Sottolineiamo che le considerazioni appena fatte non valgono per le implemen-

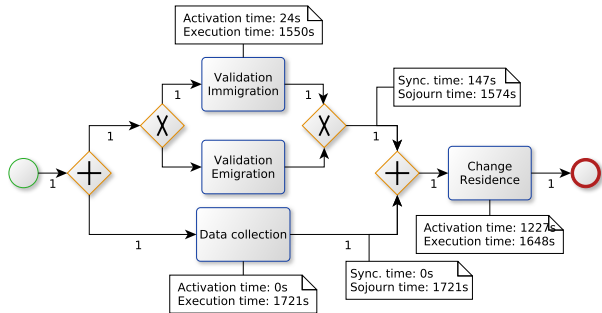


Fig. 5 - BPMN con annotata l'analisi di performance

tazioni preesistenti dell'analisi di performance (come quella di ProM 5.2) che non trasformano le sequenze del log-replay in sequenze eager. In particolare, tali implementazioni permettono di valutare i tempi di sincronizzazione delle piazze coinvolte dai gateway join ($(P(x1, J1)$ e $P(x2, J1))$) solo se sono immediatamente precedute da una transizione visibile.

3.2 Analisi di conformance

Immigration #1 start	Immigration #2 complete	Emigration #3 start	Emigration #4 complete	Data collection #5 start	Change Residence #6 start	Change Residence #7 complete
01.06.2011 17:14:56	01.06.2011 17:31:19	01.06.2011 18:24:33	01.06.2011 18:37:10	01.06.2011 18:25:27	01.06.2011 18:35:21	01.06.2011 18:42:41

Fig. 6 - Esempio di Log

I problemi di conformità possono essere diagnosticati analizzando i token creati artificialmente dal log-replay (i *missing token*) ed i token che non sono stati consumati (i *remaining token*). Il log rappresentato in Fig. 6 fornisce un esempio di esecuzione errata rispetto al processo di Fig. 1. La traccia contiene l'esecuzione di attività mutualmente esclusive (*Immigration* ed *Emigration*) e non contiene gli eventi relativi al completamento dell'attività *DataCollection*. Le misure di conformità prodotte dal log-replay per questa traccia sono mostrate nella Fig. 7. Nella piazza *p3* non è presente alcun token, tuttavia almeno un token è necessario per il replay della transizione *ES* che non è abilitata perché il ramo alternativo *Immigration* è già stato eseguito ed ha consumato il token. Nella piazza *pdc* rimane un token: questa situazione avviene perché nel log manca la terminazione dell'attività *DataCollection*. Dunque la piazza *p7* non riceve alcun token durante il log-replay, la sincronizzazione *t7* fallisce non essendo mai abilitata e quindi resta un token nelle piazze *p4* e *p5*. Il token mancante nella piazza *p8* è necessario per il replay della transizione *CRS*.

Descriviamo ora come riportare i dati di conformance sul modello originale. Il log-replay produce artificialmente i token mancanti solo quando sono necessari per eseguire le transizioni visibili. Pertanto i token mancanti possono essere generati solamente in quelle piazze che hanno nel post-set almeno una transizione visibile. Nella Fig. 2 queste piazze sono $P(x, T)$ e P_t . Inoltre il log-replay esegue le transizioni invisibili solo se la loro esecuzione è necessaria per attivare una

transizione visibile. Per esempio, per ogni esecuzione della transizione iniziale ts il log-replay esegue necessariamente una transizione visibile che consuma il token nella piazza $P(s, y)$. La stessa considerazione si applica a tutte le transizioni invisibili che producono un solo token. Quindi, le sole piazze in cui possono rimanere token sono: (i) quelle nel post-set di una transizione visibile; (ii) quelle nel post-set di una transizione invisibile che produce più di un token. Nella Fig. 2 queste piazze sono P_t , $P(T, y)$, $P(F1, y1)$ e $P(F1, y2)$, cioè quelle coinvolte nella modellazione di una singola attività BPMN e del fork gateway.

In base a queste considerazioni, le metriche di conformità della rete di Petri sono riportate sul diagramma BPMN come segue:

- Per ogni task T un eventuale token mancante in $P(x, T)$ viene definito come “un’esecuzione errata”; la mancanza di un token in Pt viene definita come “un fallimento interno”; la presenza di un token rimanente in Pt viene definita come “un completamento mancante”; e un token rimanente in $P(T, y)$ viene definito come “un’esecuzione interrotta”.
- Per ogni ramo ($i \in 1, 2$) di un fork $F1$, i token rimanenti in $P(F1, yi)$ sono detti “rami interrotti”. Si noti che per ogni esecuzione di $tF1$ un token può rimanere in $P(F1, y1)$ oppure in $P(F1, y2)$, ma non in entrambe le piazze.

Nella Fig. 8 i dati di conformance sono proiettati sul modello BPMN. La figura evidenzia: (i) il mancato completamento di *DataCollection* (ii) l'esecuzione sbagliata di *Emigration*, causata dall'attivazione dell'altro ramo del gateway esclusivo (iii) le esecuzioni interrotte di *Emigration* e *Immigration*, causate dalla mancata sincronizzazione del gateway join (iv) l'esecuzione errata di *ChangeResidence*.

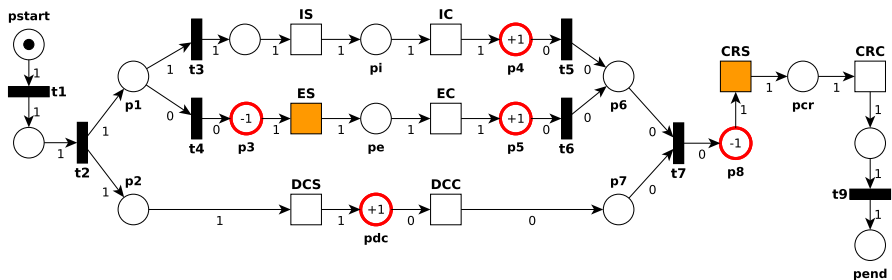


Fig. 7 - Rete di Petri con annotata l'analisi di conformance

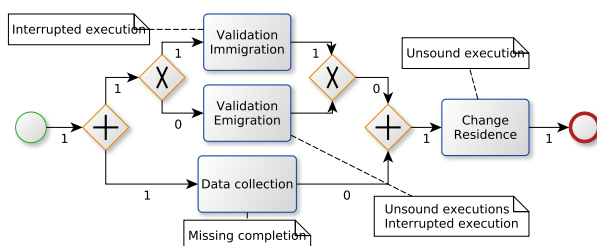


Fig. 8 - BPMN con annotata l'analisi di conformance

4. Conclusioni

Abbiamo presentato l'estensione dell'algoritmo esistente di valutazione di performance in modo da gestire le transizioni invisibili come "eager". La nostra estensione permette la valutazione dei tempi di sincronizzazione anche per le reti di Petri complesse ottenute dai modelli BPMN e semplifica la proiezione delle analisi sul modello di partenza. La possibilità di analizzare modelli BPMN e la presentazione dei risultati come annotazione dei modelli stessi permettono, ai vari attori coinvolti nella gestione dei processi, di trarre beneficio dalle diverse tecniche formali in modo trasparente.

Le tecniche presentate sono state implementate come quattro nuovi plug-in di ProM 6: (i) PerformanceAnalysis utilizza il log-replay esistente, ne trasforma gli output in sequenze eager e valuta i tempi di attesa e sincronizzazione; (ii) BpmnToPetri trasforma modelli BPMN in reti di Petri; (iii) ConformanceToBPMN e (iv) PerformanceToBPMN annotano i modelli BPMN con gli artifact necessari per rappresentare le misure di performance e conformance.

Stiamo attualmente estendendo il framework per integrare tecniche di data mining. In particolare ci stiamo concentrando sulla possibilità di fornire strumenti per individuare le classi di processo che causano colli di bottiglia. Il nostro obiettivo è quello di integrare i linguaggi esistenti per la definizione di Service Level Agreement e verificarne la conformità dei processi.

Riferimenti bibliografici

Börger E., Thalheim B. Modeling workflows, interaction patterns, web services and business processes: The ASM-based approach. In *ABZ*. LNCS, 5238, 2008, 24–38.

Dijkman R. M., Dumas M., Ouyang C. Semantics and analysis of business process models in BPMN. *Information & Software Technology*, 50-12, 2008, 1281–1294.

Eindhoven Univ. of Technology, 2011. *ProM 6.1*. <http://www.processmining.org/>.

IEEE, 2011. *Task Force on Process Mining*. <http://www.win.tue.nl/ieeetfpm/doku.php>.

Link.it, 2011. *Progetto RUPOS*. <http://www.link.it/projects/rupos/>.

Milner R., Parrow J., Walker D. A calculus of mobile processes, I. *Inf. Comput.*, 100-1, 1992, 1–40.

OASIS, 2007. *WSBPel 2.0*. <http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.html>.

OMG, 2011. *BPMN 2.0*. <http://www.omg.org/spec/BPMN/>.

Petri C. A. Fundamentals of a theory of asynchronous information flow. In *IFIP Congress*, 1962, 386–390.

Rozinat A., van der Aalst W. M. P. Conformance checking of processes based on monitoring real behavior. *Inf. Syst.*, 33-1, 2008, 64–95.

Scheer A.-W. Architecture of integrated information systems (ARIS). In *DIISM*. IFIP Transactions, B-14, 1993, 85–99.

van der Aalst W. M. P. The application of Petri nets to workflow management. *Journal of Circuits, Systems, and Computers*, 8-1, 1998, 21–66.

van der Aalst W. M. P., van Dongen B. F. Discovering workflow performance models from timed logs. In *EDCIS*. LNCS, 2480, 2002, 45–63.

van der Aalst W. M. P., de Medeiros A. K. A., Weijters A. J. M. M. Genetic process mining. In *ICATPN*. LNCS, 3536, 2005, 48–69.

Modellazione di consumi elettrici attraverso Ensembling di reti neurali

Annunziato M.¹, Lauro F.², Moretti F.³, Pizzuti S.¹
¹ENEA

C.R. Casaccia, Via Anguillarese 301, 00123 Roma
{mauro.annunziato, stefano.pizzuti}@enea.it

²Università della Calabria,
Dipartimento di Elettronica, Informatica e Sistemistica, Via P. Bucci cubo 42C, 87036
Arcavacata di Rende (CS)
fiorellalauro@gmail.com

³Università degli studi Roma Tre,
Dipartimento di Informatica ed Automazione, Via della Vasca Navale 79, 00146 Roma
fabiomor@gmail.com

In questo articolo proponiamo un approccio innovativo per la modellazione dei consumi elettrici di un edificio, basato sul concetto di 'ensembling' di reti neurali e lo abbiamo comparato con gli approcci comunemente usati nel settore. La sperimentazione ha riguardato la modellazione dei consumi dovuti all'illuminazione di un edificio complesso del settore terziario (situato all'interno del centro ricerche ENEA 'Casaccia') ed i risultati dimostrano la validità della metodologia applicata.

1. Introduzione

Il consumo energetico degli edifici è stimato essere pari a circa il 30–40% dell'energia globale [UNEP-SBC, 2007] e tale settore è responsabile del 40% delle emissioni mondiali di diossido di carbonio (CO₂) [Yudelson, 2010]. In Europa il 54.7% di energia elettrica è impiegata tra il settore residenziale (28.5%) e il settore commerciale e di pubblico servizio (26.2%). In particolare, in Italia il residenziale ha una quota del 22%, mentre il commerciale del 27%, per un totale di 49% di energia consumata (fonte 2008 IEA International Energy Agency). Lo studio della domanda energetica degli edifici è diventato un tema di grande importanza, non soltanto tra i tecnici di settore, grazie alla significativa crescita di interesse verso la sostenibilità energetica in seguito all'emanazione dell'EBP [European Directive, 2003]. L'analisi dei dati di consumo energetico è uno strumento indispensabile per il miglioramento della gestione degli edifici: essa infatti può aiutare nella rilevazione di anomalie riconoscendo episodi in cui si verifica un consumo troppo elevato (monitoraggio) e risulta essenziale negli audit energetici per la stima dei risparmi previsti e per la verifica dei risparmi ottenuti con operazioni di retrofit. Fondamentale è quindi disporre di modelli che permettano di calcolare il consumo per qualsiasi condizione operativa di interesse sia nella fase di progettazione/riqualificazione che nello sviluppo di sistemi di monitoraggio e controllo intelligenti con elevato grado di automazione.

Ottenere delle stime energetiche accurate è la chiave per l'efficienza energetica, con notevoli vantaggi sia economici che ambientali.

Esistono oggi tre approcci principali di sviluppo [Pedersen, 2007] a cui i metodi per la stima del consumo energetico degli edifici fanno riferimento: Approcci Statistici, Programmi di Simulazione Energetica ed Intelligent Computer System. L'utilizzo delle prime due tecniche è consolidato da circa un ventennio mentre lo sviluppo e la verifica dell'ultimo approccio è più recente [Aydinalp et al, 2003]. Ogni metodologia ha diversi ambiti applicativi, così come diversi requisiti in termini di dati in input.

L'Approccio Statistico, noto anche come Modellazione Inversa o Identificativa [Rabl e Rialhe, 1992], si basa sul comportamento dell'edificio. Vengono effettuate delle ipotesi sulla struttura del modello e si tenta di identificarne i parametri attraverso un'analisi statistica. La tecnica di modellazione maggiormente utilizzata è l'Analisi di Regressione Lineare o Multivariata [Pedersen, 2007]. Tale tecnica necessita di molte misure di consumo energetico, ma in compenso di poche informazioni relative ai fattori climatici e al background dell'edificio (sistemi e apparecchiature al suo interno etc.). Questo approccio, a differenza di quello a Simulazione, si presta ad essere facilmente automatizzato. La struttura che sta alla base della metodologia è relativamente semplice e può essere riassunta in poche equazioni; è tuttavia richiesta una procedura statistica sofisticata per far sì che i risultati siano i più affidabili possibili.

Il secondo approccio, anche noto come Modellazione in Avanti [Rabl e Rialhe, 1992], parte dalla descrizione fisica degli edifici (geometria e composizione del tetto, pareti, finestre, massa interna, sistema HVAC, disposizione degli occupanti, etc.) da dare in input a un programma di simulazione. Il modello considerato deve inoltre saper distinguere tra i diversi utilizzi finali di consumo energetico (illuminazione, apparecchiature elettriche, acqua calda domestica, riscaldamento, condizionamento). I Programmi di Simulazione, non appoggiandosi su dati energetici misurati, richiedono in input non solo una grande mole di dati relativi al clima e alla descrizione degli edifici ma anche un elevato grado di dettaglio. Importanti sono anche le informazioni sul comportamento degli occupanti (parametri sociologici). Tutto ciò rende questo approccio piuttosto oneroso da un punto di vista computazionale. Inoltre ci si imbatte spesso nella difficoltà pratica di dover adattare i dati allo specifico formato richiesto dal programma.

L'ultima metodologia, anche nota come approccio ad Intelligenza Artificiale (IA) [Pedersen, 2007], è così chiamata per il tipo di sistemi utilizzati che vanno al di là della semplice programmazione: Expert System e Reti Neurali Artificiali (RNA). Gli Expert System (o Sistemi Esperti) appartengono alla categoria dei Decision Support System: essi "prendono decisioni" interpretando dei dati e scegliendo tra delle alternative, sulla base di una conoscenza rappresentativa di una competenza umana. Le RNA [Arbib, 1995] [Haykin, 1999] sono modelli computazionali stocastici di tipo neuronale (ispirate cioè, come struttura parallela e densamente interconnessa, al cervello umano) in grado di "apprendere". Gli Intelligent Computer System elaborano dati di consumo,

caratteristiche climatiche, parametri sociologici e informazioni sul background degli edifici. Maggiore è l'accuratezza nei dati forniti, migliori sono i risultati che l'algoritmo di risoluzione restituisce, principio ovviamente valido in linea generale anche per gli Approcci Statistici e per i Programmi di Simulazione.

Il campo di applicazione delle metodologie illustrate dipende dalle loro stesse caratteristiche [Pedersen, 2007]. Si ricorre agli Approcci Statistici principalmente nella modellazione energetica che coinvolge più edifici, ad esempio [Caldera et al, 2008] [Rauhala, 1990/91] nella progettazione dello sviluppo di un'area specifica dove sono presenti più utilizzatori finali. La natura dettagliata tipica dei Programmi di Simulazione porta ad applicare questi strumenti di stima energetica ad un singolo edificio (o comunque a pochissimi edifici) [Farahbakhsh et al, 1998] di cui si hanno a disposizione informazioni minuziose: non è raro utilizzare Programmi di Simulazione per la progettazione di nuovi edifici o per la progettazione di interventi di retrofit di edifici già esistenti. Infine, per quanto riguarda l'ultima metodologia, si può dire che la sua applicazione si colloca in qualche modo tra quelle delle due metodologie precedenti: questo approccio risulta infatti adatto sia per uno sia per più edifici [Mihalakakou et al, 2002] [Tso e Yau, 2007] ed è particolarmente interessante per tutte le applicazioni di diagnostica, automazione e controllo ed ottimizzazione della gestione di edifici esistenti.

Nel presente lavoro le modellazioni energetiche realizzate fanno riferimento al consumo elettrico relativo all'illuminazione, così come in [Jadraque Gago et al, 2011], in cui si descrive lo sviluppo e l'applicazione di un modello per il calcolo del consumo di elettricità per illuminazione nel settore residenziale andaluso come prodotto tra la potenza totale nominale utilizzata nelle abitazioni e le ore di consumo, e in [Wong et al, 2010], in cui una delle 4 unità di output del modello neurale implementato è relativa proprio al calcolo della stima giornaliera del consumo per illuminazione per edifici adibiti ad uso ufficio in regioni caratterizzate da clima subtropicale.

2. Tecniche di modellazione

L'obiettivo finale da raggiungere è lo sviluppo di un 'Intelligent Computer System' finalizzato alla diagnostica e controllo applicato ad un singolo edificio (ma estendibile in prospettiva ad una rete) e quindi, come precedentemente descritto, ben si prestano gli approcci IA.

In questa sezione descriviamo quindi le tecniche (IA e non) che sono state utilizzate e comparate.

2.1 Metodo 'naive'

Per fornire un confronto significativo tra i vari metodi un modello 'naive' va introdotto per quantificare i miglioramenti introdotti da metodi sicuramente più sofisticati ed "intelligenti". Per dati periodici un modello naive può essere il seguente:

$$X_t = X_{t-s}$$

dove S è il periodo osservabile. Questo modello non fa altro che predire il segnale al tempo t fornendo lo stesso valore osservato esattamente S istanti passati. Nel nostro caso il valore di S è pari a 168, ovvero una settimana essendo i dati campionati con frequenza oraria. Quindi, ad esempio per stimare il consumo del giorno lunedì 14 alla ora x verrà usato il valore misurato il giorno lunedì 7 alla stessa ora x .

2.2 Regressione lineare semplice

La regressione formalizza e risolve il problema di una relazione funzionale tra variabili misurate sulla base di dati campionari estratti da un'ipotetica popolazione infinita. La formulazione generale è la seguente

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$$

In cui Y_i è la variabile dipendente (da modellare)

$X_{1i}, X_{2i} + \dots + X_{ki}$ sono le i -esime osservazioni di ciascuno dei k regressori;

$\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \dots + \beta_k x_{ki}$ è la retta di regressione;

β_0 è il valore atteso di Y quando tutte le X sono pari a zero;

β_1 è il coefficiente angolare di X_1 , β_2 è il coefficiente angolare di X_2 , (tenendo costanti gli X_k non presi in considerazione), etc.;

u_i è l'errore statistico.

Da questa formulazione generale discende il caso semplificato che abbiamo applicato nel presente lavoro $y_i = a + bx_i + \varepsilon_i$

Quindi la variabile dipendente y viene "spiegata" attraverso una relazione lineare della variabile indipendente x (cioè: $a + bx$) e da una quantità casuale.

Il problema della regressione si traduce nella determinazione di a e b in modo da esprimere al 'meglio' la relazione funzionale tra y e x .

2.3 Reti neurali artificiali (RNA)

Le reti neurali artificiali (RNA) sono modelli di calcolo che si ispirano al funzionamento del cervello per risolvere problemi di classificazione, modellazione e predizione di elevata complessità, ovvero dipendenti da un elevato numero di variabili ed altamente non lineari, e pertanto non risolvibili attraverso metodi algoritmici. I primi studi di tali modelli risalgono agli anni '40 con [McCulloch e Pitts, 1943], successivamente negli anni '60 con [Rosenblatt, 1962] e [Minsky e Papert, 1969] e negli anni '80 con [Rumelhart e McClelland, 1986]. Da allora questi modelli sono stati applicati in settori che vanno dall'ingegneria alla sociologia, dalla medicina alla finanza, etc. etc.

L'unità di calcolo elementare di una RNA è il neurone (o nodo), in esso si possono identificare tre componenti principali:

- Un insieme di connessioni in ingresso, ciascuna delle quali viene caratterizzata tramite un peso reale che ne definisce efficacia e tipologia.
- Un sommatore, che calcola la somma pesata dei segnali d'ingresso.

- Una funzione d'attivazione, tipicamente non lineare, che ricevuto come ingresso il segnale calcolato dal sommatore, determina lo stato di uscita del neurone.

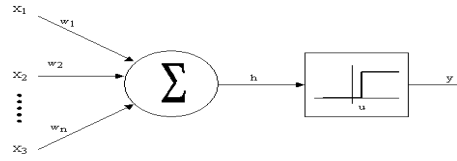


Fig. 1 - Modello di neurone artificiale

In termini matematici, indicando con x_j le componenti del vettore di ingresso, w_i i pesi sulle connessioni entranti del k-esimo neurone, y_k l'uscita del neurone k-esimo, possiamo descrivere il comportamento di un neurone attraverso le seguenti equazioni:

$$y_k = \varphi(\text{net}_k - \theta_k) \quad (1)$$

dove net_k è la combinazione lineare degli ingressi ed è chiamata input netto:

$$\text{net}_k = \mathbf{w}^T \mathbf{x} = \sum_{j=1}^n w_{jk} x_j \quad (2)$$

Nella (1) abbiamo indicato con θ_k il valore di soglia del k-esimo neurone, detto anche BIAS (Basic Input Activation System), cioè il grado di "sensibilità" con cui il neurone risponde alle perturbazione dell'input netto e con $\varphi(\cdot)$ la funzione di attivazione. Ogni funzione monotona crescente e continuamente differenziabile può rientrare tra queste, quella più comunemente utilizzata è il sigmoide:

$$\varphi(x) = \frac{1}{1 + \exp(-x)} \quad (3)$$

L'elemento che caratterizza i diversi tipi di RNA è la topologia delle connessioni dei neuroni. L'architettura più comunemente utilizzata è quella del tipo 'feed-forward'. In tale modello i neuroni sono organizzati a livelli: il primo è detto input layer perché riceve i vettori d'ingresso; poi vi sono uno o più strati nascosti (hidden layer) e uno di uscita (output layer). In tale tipo di rete i neuroni prendono quindi il loro input solo dallo strato precedente ed inviano il loro output solo allo strato seguente. Inoltre neuroni dello stesso strato non sono connessi tra loro. Il meccanismo di determinazione dei pesi delle diverse connessioni viene comunemente indicato addestramento. Tale procedura è a tutti gli effetti un problema di ottimizzazione in cui si richiede di minimizzare l'errore tra il dato reale e quello stimato, ovvero si adotta un meccanismo di 'apprendimento tramite esempi', ovvero un set di dati del fenomeno che si vuole modellare.

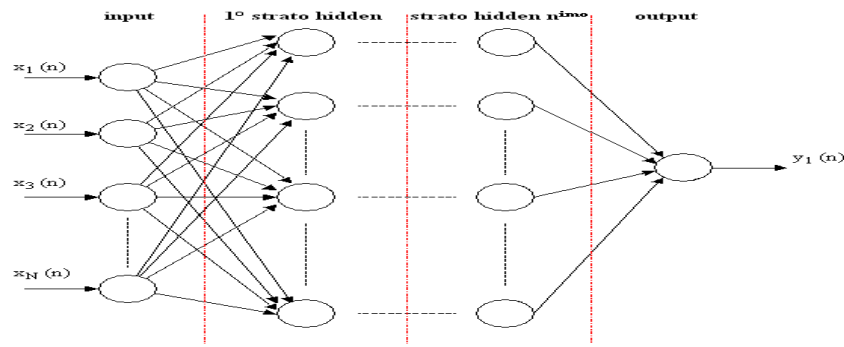


Fig. 2 - Modello di RNA feed-forward

2.4 Ensembling

Il concetto generale di 'ensembling' è quello di combinare tra loro n modelli in modo tale da ottenere un meta-modello caratterizzato da un'accuratezza superiore rispetto ai singoli modelli:

$Y = f(y_1, \dots, y_n)$, ove y_i è l'uscita dell' i -esimo modello (neurale).

Esistono diversi metodi per far questo [Krogh e Vedelsby, 1995], [Liu e Yao, 1999], [Breiman, 1999], tra questi il più semplice è quello di ottenere il meta-modello come combinazione lineare dei singoli modelli (noto anche come Generalised Ensemble Method - GEM) e come caso particolare fra questi il più semplice (applicato in questo studio) è quello che (noto anche come Basic Ensemble Method - BEM) semplicemente fa la media delle uscite:

$$Y = \frac{1}{n} \sum_{i=1}^n y_i$$

3. Risultati sperimentali

In questo paragrafo i metodi appena descritti vengono applicati alla modellazione energetica dell'edificio "C59" situato nel Centro Ricerche ENEA Casaccia di Roma. Nei test sono stati utilizzati i dati seguenti, raccolti all'interno del Centro: dati orari di consumo elettrico relativi all'illuminazione dell'edificio in esame; dati orari climatici, in particolare temperatura ambientale esterna e radiazione solare globale misurate tramite sensore; dati di affluenza, ossia le stime orarie del numero di persone presenti all'interno dell'edificio effettuate utilizzando le letture dei badge all'ingresso principale del Centro. Tutti i dati sono stati raccolti per un intero anno ma nel presente lavoro l'attenzione viene concentrata su un periodo di 3 mesi che va da martedì 1 Settembre 2009 a lunedì 30 Novembre 2009, per un totale di 13 settimane, corrispondenti a 91 giorni e quindi a 2184 campioni orari. I dati inizialmente a disposizione sul consumo illuminazione dell'edificio C59 consistevano in delle rilevazioni

effettuate all'incirca ogni 5 minuti. Al fine di ottenere dei valori medi orari di consumo, questi dati sono stati processati da una routine Matlab che per ogni ora di ogni giorno ha calcolato la media delle rilevazioni effettuate in quell'ora. Le simulazioni sono state eseguite in ambiente MATLAB (R2010a) sfruttando le function messe a disposizione dal toolbox nnet (versione 6.0.4). La tipologia di rete neurale utilizzata è feed-forward ad un livello nascosto, con un numero di input pari al numero delle variabili considerate di volta in volta in ingresso, un unico output che rappresenta la stima di consumo energetico di illuminazione e 10 neuroni interni. Le funzioni di attivazione per il livello nascosto e per il livello di uscita sono quelle di default, tangente iperbolica (tansig) e lineare (purelin) rispettivamente. Anche l'algoritmo di training è quello di default (Levenberg-Marquardt) e così pure i parametri che lo caratterizzano (μ iniziale 0.001, μ massimo 10^{10} , criterio di stop al raggiungimento di 1000 epoche), ad eccezione della suddivisione dei dati di input e di output che è stata così fissata: 90% per il training, 10% per la validazione. Il dataset trimestrale è stato diviso in due parti: la prima consiste in 10 settimane ed è stata utilizzata per la calibrazione (training) dei modelli, la seconda, formata dalle restanti 3 settimane (prese una per ogni mese), è la parte di testing utilizzata per valutare le performance dei modelli. In Tabella 1 sono elencati i diversi tipi di test effettuati, ognuno dei quali è caratterizzato da un diverso set di input. In particolare, con il termine "id giorno" si indica un numero (da 1 a 7) che identifica il giorno della settimana, mentre per "flag" si intende un numero (1 o 0) che distingue i giorni feriali dai festivi. Metà degli esperimenti è stata condotta considerando tutti i giorni nel trimestre di riferimento, l'altra metà invece tenendo conto dei dati relativi ai soli giorni lavorativi.

Test	Numero di input	Variabili in input
Test A	8	mese, giorno, ora, id giorno, flag, affluenza, radiazione globale, temperatura esterna
Test B	7	mese, giorno, ora, id giorno, flag, radiazione globale, temperatura esterna
Test C	4	flag, affluenza, radiazione globale, temperatura esterna
Test D	9	mese, giorno, ora, id giorno, flag, affluenza, radiazione globale, temperatura esterna, ora tramonto
Test E	4	mese, giorno, ora, id giorno (solo lavorativi)
Test F	5	mese, giorno, ora, id giorno, affluenza (solo lavorativi)
Test G	8	mese, giorno, ora, id giorno, affluenza, radiazione globale, temperatura esterna, ora tramonto (solo lavorativi)
Test H	7	mese, giorno, ora, id giorno, affluenza, radiazione globale, temperatura esterna (solo lavorativi)

Tab. 1 – Test effettuati

Per ogni test sono state realizzate 10 prove con le RNA e i calcoli relativi all'ensemble sono stati svolti a partire dalla combinazione degli output di queste

10 RNA. Il criterio utilizzato per confrontare le prestazioni dei modelli considerati è il mean absolute error (MAE). Le tabelle che seguono (Tab. 2 e Tab. 3) illustrano i risultati ottenuti nel training e nel testing: in particolare, per quanto riguarda le RNA, vengono riportati valore medio e deviazione standard (tra parentesi) dei MAE relativi alle 10 prove di ogni test effettuato.

Test	NAIVE	REGR LIN	RNA	BEM
Test A	2.60	1.45	1.02 (± 0.13)	0.88
Test B	2.60	1.45	1.06 (± 0.14)	0.93
Test C	2.60	1.45	1.35 (± 0.02)	1.34
Test D	2.60	1.45	0.98 (± 0.10)	0.85
Test E	2.87	1.54	1.10 (± 0.05)	1.01
Test F	2.87	1.54	1.11 (± 0.06)	1.00
Test G	2.87	1.54	1.02 (± 0.08)	0.89
Test H	2.87	1.54	1.05 (± 0.11)	0.94

Tab. 2 – Risultati training : MAE (kW)

Test	NAIVE	REGR LIN	RNA	BEM
Test A	2.32	1.28	1.30 (± 0.13)	1.04
Test B	2.32	1.28	1.45 (± 0.23)	1.17
Test C	2.32	1.28	1.28 (± 0.04)	1.26
Test D	2.32	1.28	1.54 (± 0.37)	1.13
Test E	2.95	1.37	1.37 (± 0.18)	1.14
Test F	2.95	1.37	1.30 (± 0.13)	1.10
Test G	2.95	1.37	1.32 (± 0.20)	0.98
Test H	2.95	1.37	1.22 (± 0.10)	0.95

Tab. 3 – Risultati testing : MAE (kW)

Come previsto, tutti i modelli proposti presentano risultati migliori rispetto al modello naive. Il modello regressivo lineare, essendo fortemente influenzato dall'occupazione dell'edificio, presenta prestazioni migliori, sia nel training che nel testing, quando si fa riferimento a tutti i giorni della settimana. Si nota come nel testing per i modelli RNA e BEM si ottengano risultati migliori considerando i soli giorni lavorativi. E' interessante notare come il vero valore aggiunto delle reti neurali si abbia quando le si combinano in ensembling. Il risultato globalmente migliore lo si ottiene infatti proprio attraverso questa metodologia con la configurazione H (mae = 0.95). Nella figura successiva viene riportato un confronto tra il segnale reale ed i modelli regressione lineare (regr.lin) e BEM. Anche da tale figura si può osservare come l'ensembling neurale riesca a seguire i picchi molto meglio.

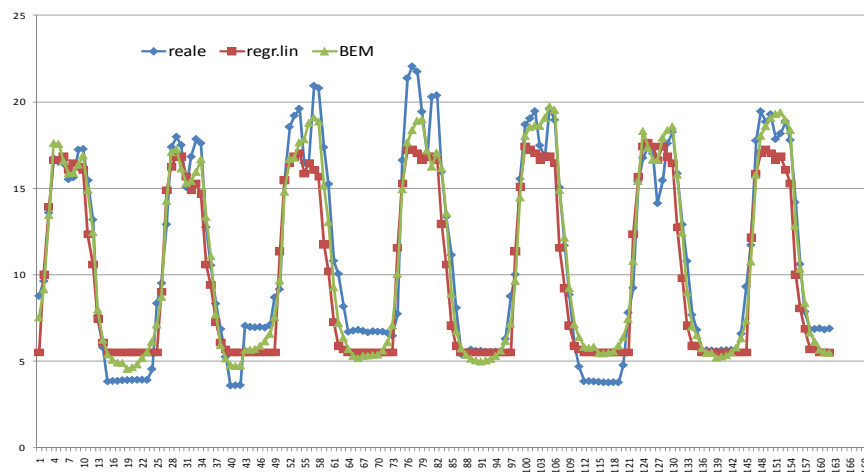


Fig. 3 – Comparazione risultati

4. Conclusioni

In questo articolo abbiamo presentato un approccio innovativo per la modellazione dei consumi elettrici. In particolare l'approccio che abbiamo proposto si basa sull'ensembling di reti neurali e lo abbiamo comparato con gli approcci comunemente usati nel settore. La sperimentazione ha riguardato la modellazione dei consumi dovuti all'illuminazione di un edificio complesso del settore terziario (l'edificio C59 del centro ricerche ENEA 'Casaccia') ed i risultati hanno mostrato come l'approccio proposto sia più preciso rispetto a metodi tradizionalmente utilizzati rendendo ipotizzabile come lavoro futuro una sua applicazione alla modellazione anche di altri parametri di consumo energetico come ad esempio la climatizzazione e la forza elettromotrice.

Bibliografia

- Arbib M.A., The Handbook of Brain Theory and Neural Networks, The MIT Press, Cambridge (MA), 1995.
- Aydinalp M., Ugursal V. I., Fung A. S., Modelling of residential energy consumption at the national level. International Journal of Energy Research, 27, 2003, 441-453.
- Breiman L., Combining Predictors, in Sharkey A.J.C., (ed.): Combining Artificial Neural Nets – Ensemble and Modular Multi-net Systems, Springer, Berlin, 1999, 31-50.
- Caldera M., Corgnati S. P., Filippi M., Energy demand for space heating through a statistical approach: application to residential buildings. Energy and Buildings, Elsevier, 40, 2008, 1972–1983.
- European Union, Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings. Official Journal of the European Communities, January 4, 2003.

Farahbakhsh H., Ugursal V. I., Fung A. S., A residential end-use energy consumption model for Canada. *International Journal of Energy Research*, 22, 1998, 1133-1143.

Haykin S., *Neural Networks, a comprehensive foundation* (2nd edition), Prentice Hall, New Jersey, 1999.

Jadraque Gago E., Ordonez Garcia J., Espin Estrella A., Development of an energy model for the residential sector: electricity consumption in Andalusia, Spain. *Energy and Buildings*, Elsevier, 43, 2011, 1315-1321.

Krogh A., Vedelsby J., Neural network ensembles, cross validation and active learning, in Tesauro G., Touretzky D. S., Leen T. K., editors, *Advances in Neural Information Processing Systems*, 7, 231-238, MIT Press, 1995.

Liu Y., Yao X., Ensemble learning via negative correlation. *Neural Networks*, 12(10), 1999, 1399-1404.

McCulloch W. S., Pitts W., A Logical Calculus of Ideas Immanent in Nervous Activity. *Bull. Mathematical Biophysics*, 5, 1943, 115-133.

Mihalakakou G., Santamouris M., Tsangrassoulis A., On the energy consumption in residential buildings. *Energy and Buildings*, Elsevier, 34, 2002, 727-736.

Minsky M., Papert S., *An Introduction to Computational Geometry*, MIT Press, Cambridge, Mass, 1969.

Pedersen L., Use of different methodologies for thermal load and energy estimations in buildings including meteorological and sociological input parameters. *Renewable and Sustainable Energy Reviews*, Elsevier, 11, 2007, 998-1007.

Rabl A., Rialhe A., Energy signature models for commercial buildings: test with measured data and interpretation. *Energy and Buildings*, Elsevier, 19, 1992, 143-154.

Rauhala K., A simple computer model for estimating the energy consumption of residential buildings in different microclimatic conditions in cold regions. *Energy and Buildings*, Elsevier, 15-16, 1990/91, 561-569.

Rosenblatt R., *Principles of Neurodynamics*, Spartan Books, New York, 1962.

Rumelhart D.E., McClelland J. L., *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, MIT Press, Cambridge, Mass, 1986.

Tso G. K. F., Yau K. K. W., Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. *Energy*, Elsevier, 32, 2007, 1761-1768.

United Nations Environment Programme, *Buildings Can Play Key Role In Combating Climate Change*, 2007.

<http://www.unep.org/Documents.Multilingual/Default.asp?DocumentID=502&ArticleID=5545&l=en>.

Wong S. L., Wan K. K. W., Lam T. N. T., Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*, Elsevier, 87, 2010, 551-557.

Yudelson J., *Greening Existing Buildings*, Green Source/McGraw-Hill, New York, 2010.

Monitoraggio energetico: service robots come alleati?

Roberto Antonini, Gian Piero Fici¹, Marco Gaspardone²

Telecom Italia S.p.a.

Via Reiss Romoli 274, 10148 Torino

roberto1.antonini@telecomitaly.it

gianpiero.fici@telecomitalia.it

marco.gaspardone@telecomitalia.it

Nel presente documento viene presentato il percorso che può essere intrapreso da sistemi di monitoraggio energetico per passare dal semplice utilizzo di sensori fissi all'utilizzo combinato di sensori fissi e sensori montati su robot (sensori mobili). Dal sistema kaleidos, sperimentato e messo in campo da Telecom Italia per l'efficientamento energetico delle centrali telefoniche, ad un sistema che mette insieme sensori fissi e sensori mobili per coprire uno spazio più ampio di requisiti utili al monitoraggio energetico. Il cambio di paradigma "da sensori fissi a sensori mobili", sfrutta la robotica di servizio ed amplia il concetto dell'Internet of Things introducendo fra le "cose" anche i robot.

1. Introduzione

Telecom Italia è uno dei maggiori consumatori di energia elettrica nel paese (circa 2.44 TWh nel 2010), con poco meno dell'1% dell'intera produzione italiana. L'aumento dell'efficienza energetica viene anche incontro a requisiti imposti dall'EU in termini di emissioni e risparmio energetico. E' stato stimato che circa l'80% dell'energia è usata per le telecomunicazioni (Centrali telefoniche, CED, SRB) e metà di questa per il condizionamento.

Ci sono due modi per ridurre il consumo di energia:

- L'uso di apparati di TLC progettati tenendo conto di tutta una serie di parametri ambientali
- Ottimizzare il controllo delle temperature all'interno delle Centrali, CED e SRB

Per migliorare ulteriormente il controllo delle temperature ed aumentare l'efficienza energetica è utile un sistema di monitoraggio real-time e granulare per tenere sotto controllo i consumi di energia, in termini di potenza elettrica assorbita dai vari apparati e le condizioni ambientali dei siti in termini di temperatura ed umidità.

Il dipartimento R&T di Telecom Italia ha costruito un sistema di monitoraggio energetico real-time, basato su Wireless Sensor Network (WSN) ZigBee, per tenere sotto controllo i consumi di energia, le temperature e l'umidità di tutte le

sale di una centrale telefonica.

Dal 2007 il sistema raccoglie dati da circa 190 centrali in tutta Italia, i dati riguardano essenzialmente i consumi elettrici e le temperature. Nel prossimo capitolo (2) approfondiremo il sistema di monitoraggio energetico, che utilizza un certo numero di sensori fissi per rilevare le misure (circa 7000 nodi sensore per tutte le centrali telefoniche), verranno inoltre evidenziate alcune criticità che si presentano nel momento in cui si vuole eseguire un monitoraggio di temperatura più granulare nello spazio. Nel terzo capitolo descriveremo l'approccio alla robotica per giustificare un cambio di paradigma: dal sensore fisso al sensore mobile, montato su un robot che lo muove nell'ambiente; verranno inoltre descritte le possibili interazioni tra sensori fissi e sensori a bordo dei robot e la possibile architettura di una piattaforma di gestione di Telecom Italia che si pone come intermediario (livello di astrazione) tra il "mondo" dei robot e le applicazioni.

2. Monitoraggio energetico (Kaleidos)

Kaleidos è una rete di monitoraggio, a basso costo e ridotta invasività, dei parametri ambientali delle singole sale e dei principali flussi di energia elettrica di una centrale telefonica. Serve a progettare ed ottimizzare il risparmio di energia elettrica attraverso un controllo capillare dei consumi delle singole utenze e delle condizioni ambientali delle sale [Antonini et al, 2010].

2.1 Monitoraggio con sensori fissi.

L'architettura di riferimento di kaleidos è descritta in figura 1:

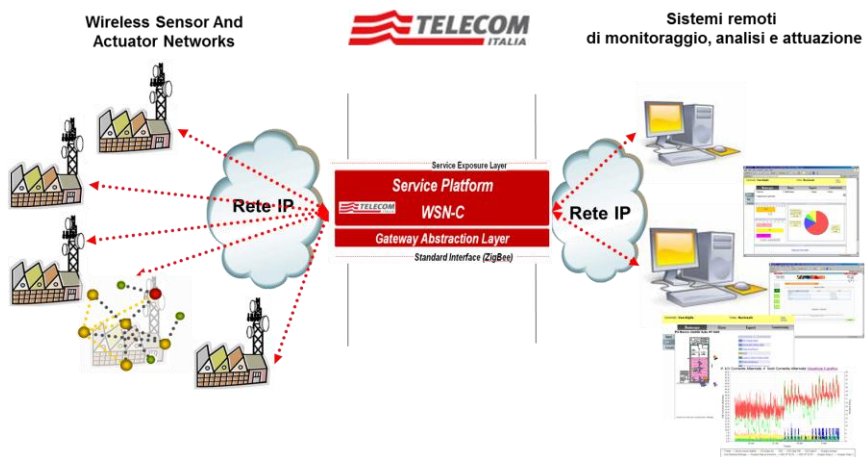


Figura 1: Architettura Kaleidos

Nella parte più a sinistra viene rappresentato il “mondo delle WSN”, nel caso delle centrali telefoniche i nodi sensori, per formare la WSN che copre l'intera centrale, sono rappresentati in figura 2, dove troviamo sensori per il rilevamento delle misure elettriche (corrente alternata CA e continua CC) e sensori per il rilevamento delle temperature (THL).

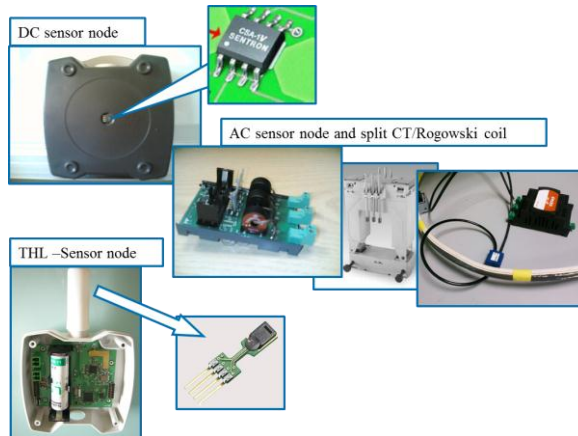


Figura 2: nodi sensori di kaleidos

Nella parte centrale della architettura in figura 1 è presente la parte di “Service Platform”, il livello di astrazione tra le WSN e le applicazioni che utilizzano i dati provenienti dai sensori.

Kaleidos è anche il nome dell'applicazione utilizzata per il monitoring, il provisioning e la diagnostica delle centrali telefoniche monitorate dai sensori in Figura 2, posizionati in diversi punti della centrale. L'applicativo è stato progettato con lo scopo di supportare gli energy manager in azioni di efficientamento energetico.

2.2 Criticità e problematiche

Una normale evoluzione del sistema kaleidos sarebbe stata quella di rilevare i consumi elettrici anche nei CED, in questi ambienti esiste però già un sistema chiamato Building Management System (BMS) per il rilevamento delle potenze assorbite dalle sale server. Un requisito forte nell'ambiente dei CED è quello di rilevare le sacche di calore o hotspot, che si verificano quando non viene disperso o eliminato correttamente il calore emesso dagli apparati rischiando così di causare dei guasti che si possono tradurre in disservizio e costi. Per determinare correttamente un hotspot servirebbe una simulazione della fluidodinamica dell'ambiente della sala server che tenga conto di tutta una serie di parametri (CFD), sia a scopo progettuale che di riqualificazione della sala stessa. Nella figura 4 vengono illustrati gli hotspot e la loro simulazione grazie alla fluidodinamica dell'ambiente che ne permette la prevenzione grazie ad una progettazione adeguata del condizionamento.

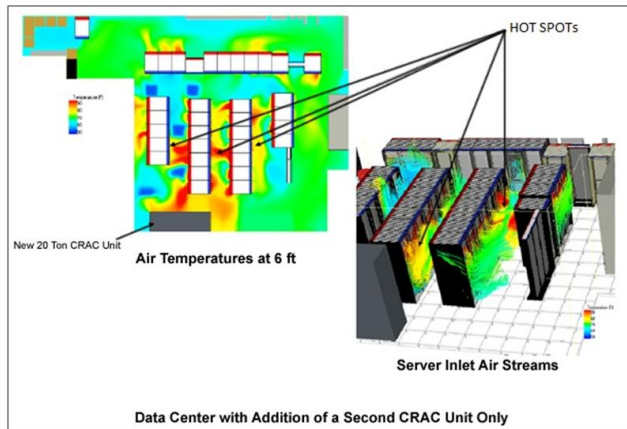


Figura 3: hotspot in un CED

In un contesto come quello descritto nella figura 4 emergono i limiti di un monitoraggio con sensori fissi: per rilevare un hotspot sarebbe necessario “tappezzare” una sala server di sensori di temperatura, perché non è noto a priori dove si possa formare la sacca di calore. Nel prossimo capitolo descriveremo l’approccio alla robotica per giustificare un cambio di paradigma: dal sensore fisso, che come abbiamo visto presenta dei limiti nel soddisfare alcuni requisiti, al sensore mobile, montato su un robot che lo muove nell’ambiente; inoltre verrà descritto un sistema che si avvale dell’interazione tra sensori fissi e sensori mobili.

3. Utilizzo della robotica

La robotica è una scienza che ha raggiunto dei risultati tali che al giorno d’oggi, oltre che essere un campo adatto a studi fantasiosi e affascinanti, può fornire soluzioni concrete per problematiche reali, come quelle descritte in precedenza in questo documento.

Partendo dal 1961, anno in cui è stato installato il primo robot Unimate in una linea di assemblaggio della General Motors, la robotica ha subito una notevole evoluzione ed è passata dalle prime applicazioni, confinate nel solo settore della Robotica Industriale, ad affacciarsi in ambiti estremamente diversificati, quali:

- Robotica Militare: sistemi tattici basati sia su velivoli aerei (UAV – Unmanned Aerial Vehicle), sia su velivoli terrestri, utilizzati per sostituire l’impiego di soldati in situazioni pericolose, oppure sistemi di trasporto utilizzati per supportare operazioni militari (esoscheletri, Legged ATV – All Terrain Vehicle, ecc.)
- Robotica Spaziale: sistemi semiautonomi per attività nello spazio (EVA – Extra Vehicular Activity), oppure sistemi di esplorazione extra-terrestre per la ricognizione di ambienti estremi (pianeti, asteroidi, ecc.)

- Robotica Commerciale: sistemi automatizzati per la gestione di aspetti aziendali e commerciali, come ad esempio: la gestione della logistica dei magazzini, la sorveglianza di aree aziendali, la pulizia di zone industriali, ecc.
- Robotica Medica: sistemi di manipolazione assistita per l'esecuzione di operazioni da remoto (ad esempio il da Vinci Surgical System), oppure sistemi di supporto medico di tipo infermieristico all'interno degli ospedali (distribuzione di farmaci, pasti, risposta alle chiamate e valutazione della gravità, ecc.)
- Robotica Personale: sistemi che permettono di eseguire lavori, di sostituire la presenza umana o di aumentare la tranquillità personale; in questa categoria ricadono applicazioni disparate che vanno dalla pulizia della casa, alla telepresenza, alla teleassistenza per anziani o bambini, fino all'intrattenimento e al gioco.

In particolare da molte fonti la situazione attuale della Robotica Personale viene paragonata a quella dei Personal Computer dei primi anni '80, cioè un campo non ancora completamente maturo, ma che già mostra le potenzialità *game-changing* che il suo impatto può apportare alla vita quotidiana. Ed è in particolare in questo settore della robotica, che permette la creazione di robot autonomi per la navigazione in ambienti semi-strutturati, che possono essere ricercate le soluzioni per il trattamento del problema in oggetto.

3.1 Sensori fissi e sensori “mobili”

Di seguito viene descritto il funzionamento di un sistema composto da:

- sensori di temperatura fissi con connessione wireless,
- robot autonomo dotato di termocamera con connessione *wireless* (sensore “mobile”),
- gateway per la connessione in rete,
- *middleware* di controllo

3.1.1 Fase 1: normali operazioni

Durante le operazioni normali il sistema procede al rilevamento delle temperature ambientali in due modi:

- A) i sensori fissi rilevano la temperatura ad intervalli regolari,
- B) il robot dotato di termocamera si muove nell'ambiente con un percorso semi-casuale prendendo fotografie tramite la termocamera.

Sia i sensori che il robot inviano i propri dati (rilevazioni di temperatura e fotografie termografiche) al middleware di controllo tramite connessioni radio con un gateway locale che provvede all'interlavoro tra le rete wireless e la rete IP.

In figura 5 sono rappresentati i componenti del sistema disposti in un ambiente che rappresenta una sala di un data center vista in pianta.

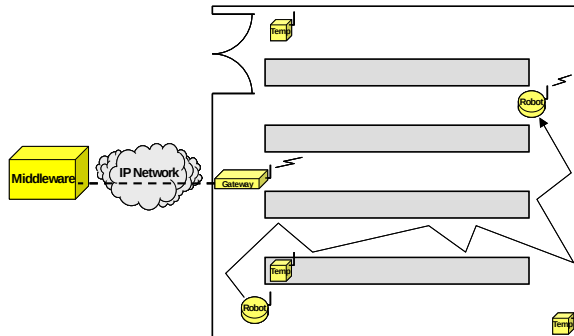


Figura 4: componenti del sistema impegnate nelle operazioni normali

3.1.2 Fase 2: rilevazione temperatura anomala da parte del robot.

Nel caso una fotografia termografica presa dalla termocamera del robot evidenzi una temperatura anomala, il middleware di controllo inizia a monitorare il percorso del robot e lo istruisce in modo che nel suo percorso semi-casuale non si allontani da un'area centrata sulla posizione in cui è stata scattata la fotografia termografica. Allo stesso tempo il middleware di controllo istruisce anche i sensori fissi ad aumentare la frequenza di campionamento delle misure di temperatura. Tutti i dati sono sempre comunicati al middleware centralizzato.

Questa fase di pre-allarme ha una durata configurabile, dopo di che, se non viene rilevato un allarme (vedi fase successiva), il sistema ritorna nella fase di normali operazioni: il confinamento del robot viene rimosso ed il tempo di campionamento dei sensori fissi viene riportato alla normalità. Quando la fase di pre-allarme cessa senza che si passi ad una fase di allarme, per un certo periodo vengono ignorate successive situazioni di pre-allarme per evitare di cadere in una situazione di stallo in cui il robot continua a rilevare la stessa temperatura anomala senza riuscire ad allontanarsi.

In figura 6 viene rappresentato lo stato di pre-allarme del sistema con il confinamento del robot in un'area centrata sul punto in cui è stata rilevata la temperatura anomala.

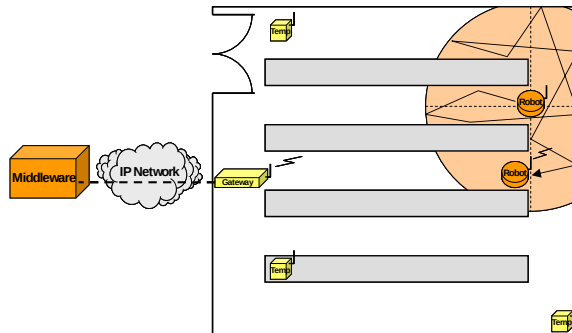


Figura 5: stato di pre-allarme del sistema

3.1.3 Fase 3: rilevazione di un allarme da parte del robot:

Nel caso una fotografia termografica presa dalla termocamera del robot evidenzi una situazione di allarme, il middleware di controllo cancella un'eventuale fase di pre-allarme precedente e inizia a controllare il percorso del robot in modo che esamini approfonditamente un'area ristretta centrata sulla posizione in cui è stato rilevato l'allarme.

Durante questa fase sia i sensori fissi sia il robot continuano a rilevare i dati e a comunicarli al middleware centralizzato.

Anche questa fase di allarme ha una durata configurabile, definita in modo da permettere un'esauriente analisi dei dintorni della posizione in cui è stata rilevata la situazione di allarme, dopo di che il sistema ritorna nella fase di normali operazioni: il confinamento del robot viene rimosso ed il tempo di campionamento dei sensori fissi viene riportato alla normalità.

Quando la fase di allarme cessa, per un certo periodo vengono ignorate successive situazioni di allarme e di pre-allarme per evitare di cadere in una situazione di stallo in cui il robot continua a rilevare lo stesso allarme senza riuscire ad allontanarsi.

In figura 7 viene rappresentato lo stato di allarme del sistema con la cancellazione di una precedente fase di pre-allarme ed il confinamento del robot in un'area ristretta.

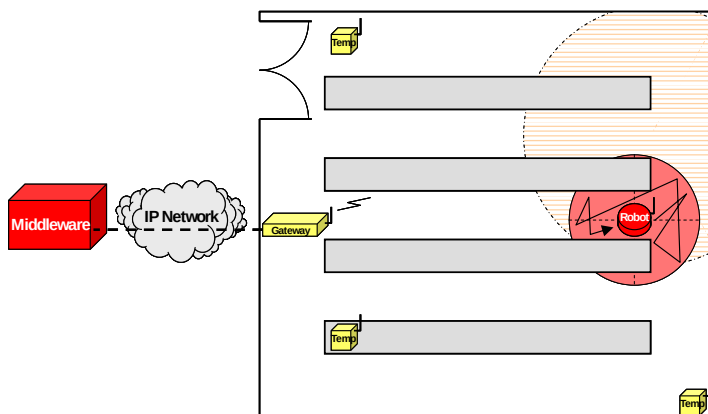


Figura 6: stato di allarme del sistema

3.1.4: Fase 4: rilevazione temperatura anomala da parte di un sensore fisso.

Nel caso un sensore fisso di temperatura evidenzi una temperatura anomala, il middleware di controllo gestisce direttamente il movimento del robot portandolo vicino alla posizione del sensore fisso, dopodiché il sistema entra nella fase di pre-allarme o allarme descritte precedentemente (vedi fase 2 e fase 3) con le stesse modalità tranne che l'evento scatenante è rilevato da un sensore di temperatura fisso invece che da una fotografia termografica.

In figura 8 viene rappresentato lo stato di pre-allarme del sistema rilevato da un sensore di temperatura fisso con il richiamo del robot nella posizione del sensore fisso ed il suo confinamento in un'area centrata su tale posizione.

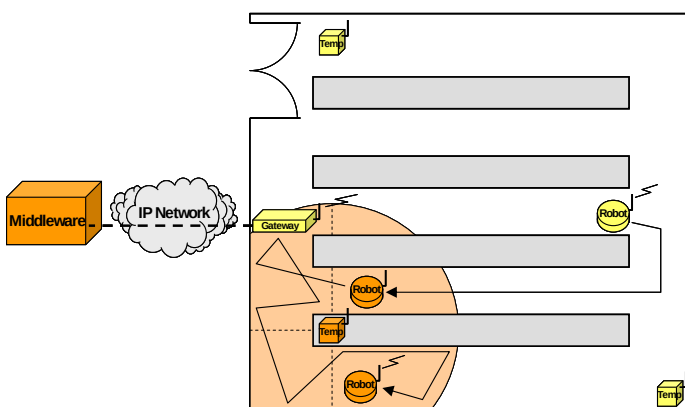


Figura 7 stato di pre-allarme del sistema rilevato da un sensore di temperatura fisso

4 Conclusioni

L'impegno e la ricerca di Telecom Italia negli ultimi anni hanno portato, come visto, al dispiegamento di soluzioni di monitoraggio real time basati su reti di sensori e piattaforme centralizzate di analisi dati.

Ma per poter estendere questi sistemi non solo con sensoristica più evoluta e con dispiegamenti più massivi di reti di monitoraggio e attuazione, TI sta cominciando ad esplorare anche il settore della robotica di servizio che sta evolvendo con soluzioni sempre più evolute e a costi più contenuti. I *service robots* difatti hanno caratteristiche che li rendono idonei alle attività di monitoraggio e attuazione: sono apparati che interagiscono col mondo fisico (si muove, apre, prende, porta, osserva), dotati di mobilità, che eseguono il proprio compito con un certo grado di autonomia.

In questo contesto l'utilizzo congiunto di reti di sensori e di robot di servizio con a bordo ulteriore sensoristica permette, grazie al coordinamento di un middleware centralizzato, di definire scenari di servizio più ampi come quello illustrato nel capitolo precedente.

Con questi approcci la robotica sta diventando un elemento importante nell'IoT (o "Internet of Robots"), punto intelligente, sempre connesso alla rete Internet e alle reti di "oggetti" che lo circondano in modo da interagire con entrambe le reti per portare a termine i suoi compiti al meglio possibile.

Bibliografia

Antonini R., Gaspardone M., Fici G. P., Energy Management of Telecommunication Plants Using Wireless Sensor Network, Proc. of the 14th International Conference on Intelligence in Next Generation Networks (ICIN) - Weaving Applications into the Network Fabric, Berlin, Germany, 2010

Operatori efficienti e robusti a supporto di applicazioni su dati spazio-temporali

Matteo Orsini, Filippo Mortari e Paolino Di Felice

Dipartimento di Ingegneria Elettrica e dell'Informazione
Università degli Studi dell'Aquila – L'Aquila

(matteorsini@yahoo.it, mortari@gmail.com, paolino.difelice@univaq.it)

Da diversi anni si registra una crescente richiesta da parte delle aziende di applicazioni software su dati spazio-tempo dipendenti. Ciò che ancora oggi costituisce un parziale freno al soddisfacimento di tale esigenza è la carenza, al livello del software di gestione dei dati permanenti, di operatori adeguati. In questo articolo si delinea una strada da percorrere per colmare tale lacuna. In concreto si propongono due operatori per oggetti in movimento. La loro prerogativa è calcolare la soluzione con tempo lineare nella dimensione dell'input (efficienza) facendo attenzione che il risultato prodotto assorba al suo interno le molteplici cause di incertezza che complicano la soluzione del problema (robustezza).

1 Introduzione

La diffusione delle tecnologie di posizionamento basate sul GPS ha accresciuto l'interesse nello sviluppo di applicazioni mobili che hanno a che fare con oggetti spaziali la cui posizione cambia nel tempo con continuità. Essi possono essere descritti in termini di punti materiali e la loro modellazione attraverso soluzioni software rappresenta un'esigenza concreta per tutte quelle organizzazioni che vogliono condurre delle analisi sul patrimonio di dati spazio-temporali in loro possesso. In [Di Felice et al, 2010] è stata presentata un'architettura software idonea allo sviluppo di questo tipo di applicazioni, ma per accrescerne l'efficacia è necessario arricchirla con un adeguato pacchetto di operatori spazio-temporali. Questo articolo costituisce una testimonianza concreta di come muoversi in tale direzione. In esso verrà presentato un nuovo operatore, chiamato `time_meet()`, in grado di calcolare le aree in cui le traiettorie definite da due m-points possono essersi incontrate e i periodi in cui tali "rendez vous" possono essersi verificati. Si tratta di un operatore *efficiente* in quanto la soluzione viene calcolata con costo $O(n+m)$, essendo n ed m il numero dei segmenti lineari di cui le traiettorie si compongono. Inoltre, il fatto che le risposte elaborate risultano affette da un certo grado di indeterminazione, lo rende *robusto* e in grado di assorbire le molteplici fonti di incertezza insite nella modellazione dei m-points. `time_meet()` può essere implementato

come User Defined Function (UDF) e come tale aggiunto alla collezione di operatori built-in dei DBMS dotati di estensione spaziale in commercio. Quest'ultima caratteristica ne rende immediato l'aggancio in applicazioni verso basi di dati spazio-temporali.

L'organizzazione dell'articolo è la seguente. Nella Sez. 2 si riparte dall'architettura presentata in [Di Felice et al, 2010] con l'intento di chiarire la collocazione del nuovo operatore. Una panoramica sulle incertezze connesse alla modellazione degli oggetti in movimento viene fornita nella Sez. 3, mentre nella quarta e quinta sezione vengono presentati due possibili algoritmi per `time_meet()`, la valutazione della loro complessità computazionale e i risultati degli esperimenti condotti per misurarne i tempi di esecuzione. Brevi conclusioni sono raccolte nella Sez. 6.

2 Un'architettura software per applicazioni su dati spazio-tempo dipendenti

Quando si vogliono arricchire le applicazioni gestionali *classiche* (in grado, cioè, di automatizzare i processi di business di un'azienda) con estensioni capaci di gestire la “*localizzazione*” di oggetti di interesse, è necessario pensare a soluzioni software in cui i moduli esistenti si integrino con quelli addetti al trattamento dei dati geo-referenziati. A tal proposito, in [Di Felice et al, 2010] è stato presentato un framework, correntemente utilizzato a supporto di questo tipo di problematiche, il quale impone alle applicazioni costruite su di esso l'architettura di Fig.1.

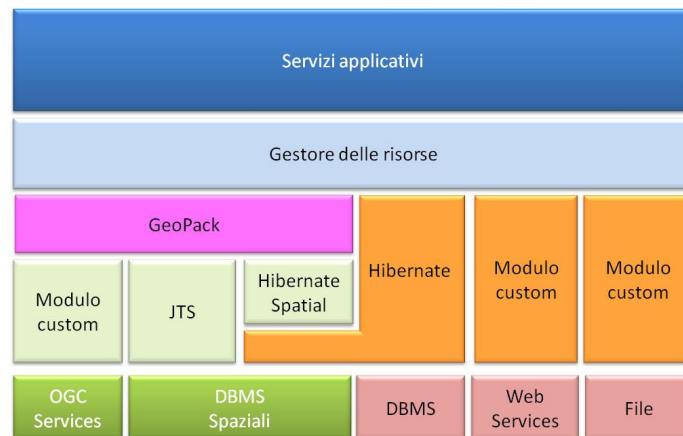


Fig.1 - Architettura a supporto di applicazioni per gestionali su dati spazio-tempo dipendenti

Il passaggio focale di questa soluzione sta nell'offrire una medesima interfaccia ed un unico linguaggio per l'interrogazione di risorse eterogenee, rendendo trasparente all'utilizzatore se l'interazione richiesta coinvolge una base dati spaziale piuttosto che una classica base dati *gestionale*. Dal punto di vista dell'utilizzatore del framework non è, quindi, necessario

comprenderne le dinamiche interne ma è sufficiente la conoscenza delle API e del Gestore delle risorse (Fig.1).

Il contributo offerto da questo articolo, si colloca all'interno della componente "DBMS Spaziali" dell'architettura di Fig. 1 e l'operatore per il calcolo delle intersezioni spazio-temporali fra coppie di traiettorie, rappresenta solo un primo passo verso la definizione di una vasta libreria di funzioni in grado di supportare la modellazione ed interrogazione di m-points e di garantire l'integrazione fra informazione spaziale e temporale permettendo di avere a che fare con *geometrie che si modificano nel tempo continuamente*.

L'esigenza di dotare i DBMS attualmente in circolazione di operatori solidi e performanti per m-points è testimoniata dall'intensa ricerca dell'ultimo decennio. Particolare rilevanza rivestono i risultati raggiunti dal Prof. Güting e dai suoi colleghi (si veda, ad esempio, [Forlizzi et al, 2000] e [Cotelo Lema et al, 2001]), i quali hanno svolto una pionieristica opera di classificazione, progettazione e formalizzazione di ogni tipo di dato ed operatori di rilievo nella prospettiva spazio-temporale, nella quale trovano collocazione la descrizione dei m-points e dell'operatore `intersection()` che ne calcola l'intersezione. Il loro lavoro, di interesse sia dal punto di vista teorico che metodologico, sfortunatamente non aiuta a far fronte alle esigenze quotidiane della maggior parte delle imprese che vogliono sviluppare applicazioni che coinvolgano m-points. Infatti, i loro sforzi di ricerca si concentrano nello sviluppo "da zero" del sistema SECONDO [Güting et al, 2010], la cui progettazione ignora le estensioni spaziali esistenti nei DBMS in commercio, non segue alcuno dei modelli di dati in essi disponibili e non è conforme agli standard OGC, risultando di conseguenza non integrabile in nessuna delle infrastrutture informatiche attualmente utilizzate dalle aziende, depositarie di ingenti quantità di dati spaziali.

Al contrario, il progetto in corso HERMES [Pelekis et al, 2010] mira ad estendere le funzionalità dei DBMS esistenti al fine di aiutare i progettisti e gli sviluppatori di database spazio-temporali nella gestione delle applicazioni sugli oggetti che cambiano posizione, forma e dimensioni nel tempo con continuità. In particolare, le funzionalità spazio-temporali vengono fornite come *data cartridge* su Oracle.

Il nostro lavoro ha le stesse motivazioni alla base del progetto HERMES, ma rispetto ad esso introduce due varianti di rilievo. La prima concerne la scelta di un DBMS open source in luogo di uno proprietario giustificata dal trend che si sta diffondendo negli ultimi anni, prevalentemente nell'ambito della pubblica amministrazione, di spostare l'esercizio di applicazioni su dette piattaforme. La seconda variante concerne l'introduzione di un maggior grado di flessibilità nel processamento delle traiettorie. In HERMES, infatti, in continuità con la ricerca del gruppo del Prof. Güting richiamata sopra, viene adottato un modello (dettagliato in [Frentzos, 2008]) in cui si assume di essere in grado di risalire, per mezzo di una funzione semplice, all'esatta posizione assunta dal m-point in ogni istante della sua esistenza. Si tratta di un approccio che ha un senso in un "mondo perfetto". Sfortunatamente la vita reale è molto più complessa e permeata dall'*incertezza*, la cui rilevanza

nel campo dei m-points è provata da diversi articoli e viene brevemente richiamata nella prossima sezione.

3 Fonti di incertezza nella modellazione dei m-points

Un aspetto cruciale nella modellazione dei m-points concerne l'acquisizione dei dati. Per registrare fedelmente il movimento di un punto dovremmo conoscerne la posizione in ogni istante (su base continua), ma ciò è impossibile poiché anche le più moderne tecniche diffuse nell'ambito delle telecomunicazioni permettono di campionare solo ad istanti discreti di tempo (ad esempio ogni pochi secondi). D'altra parte, la discretizzazione del movimento di un oggetto, necessaria per l'esigenza di rappresentazione finita imposta dal computer, comporta un'approssimazione della traiettoria che introduce un'incertezza nella sua esatta conoscenza.

L'approccio più diffuso in letteratura alla modellazione dei m-points all'interno di una base di dati consiste nel memorizzarne la posizione ad istanti di tempo regolari ed assumere che la direzione del movimento sia costante tra punti contigui, così come la sua velocità. Ne consegue l'impossibilità di conoscere la posizione esatta dell'oggetto in istanti di tempo differenti da quelli di campionamento. Errore che si amplifica quanto più ampio è l'intervallo di campionamento. Inoltre, assumere che tra due punti di campionamento consecutivi il m-point si sia mosso a velocità e direzione costanti appare un'ipotesi troppo vincolante sulla sua *legge del moto*.

In ultimo non bisogna dimenticare che la stessa acquisizione dei dati tramite GPS è influenzata dalle condizioni atmosferiche avverse o dalla presenza di edifici, rilievi montuosi o gole che producono un rimbalzo del segnale che generano una minore accuratezza nell'informazione ricevuta.

A causa delle molteplici fonti di incertezze appena richiamate, è spesso insoddisfacente assumere la correttezza dei valori circa il "*dove*" ed il "*quando*" restituiti dal calcolo dell'intersezione spazio-temporale tra traiettorie. E' questa la ragione per la quale in questo articolo ci si orienta verso una risposta più "flessibile". La flessibilità è ottenuta rimpiazzando le traiettorie per così dire "esatte" (o meglio presunte tali) con traiettorie "incerte" (nozione descritta in [Pfoer e Jensen, 1999] – si veda la Fig.2) e, allo stesso tempo, rinunciando alla conoscenza della legge del moto difficilmente prevedibile nella realtà.

Più in dettaglio, nell'approccio presentato in questo articolo, l'incertezza viene contemplata nel calcolo di una risposta affetta da un certo grado di indeterminazione adottando un metodo di calcolo che restituisca piuttosto che un istante preciso di tempo e uno specifico punto di incontro, un intervallo temporale ($[t_{ini}, t_{end}]$) e un'area geografica in cui *potrebbe* essere avvenuta l'intersezione tra i due m-points. In sintesi, dato che la trattazione dei m-points porta con sé molteplici fonti di incertezza che comunque non è possibile eliminare, pensiamo sia necessario saper convivere con esse.

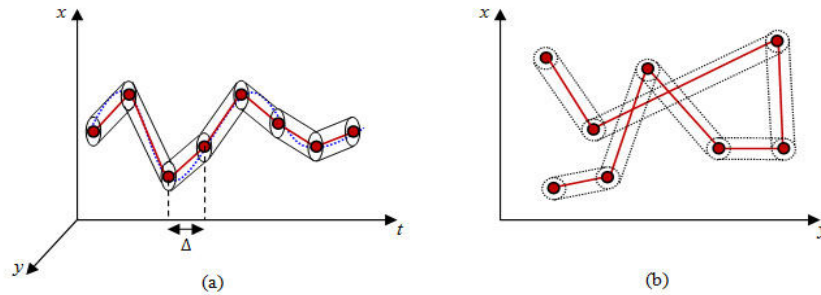


Fig.2 - Una traiettoria *incerta* in 2D+t (a) e la sua proiezione nel piano Cartesiano (b)

4 L'algoritmo *time_meet*

Come anticipato nelle sezioni precedenti, la ricerca riassunta in questo articolo affronta un problema base nel contesto dei m-points: il calcolo delle intersezioni fra coppie di traiettorie in presenza di incertezze. In particolare, vengono discussi due possibili strade per la definizione dell'algoritmo *time_meet*: la prima parte dall'analisi delle geometrie condivise (approccio "spazio-tempo"), mentre la seconda affronta preliminarmente lo studio dello sviluppo temporale delle traiettorie (approccio "tempo-spazio").

In entrambi i casi la traiettoria viene vista come una sequenza di punti ordinati rispetto all'istante in cui essi sono stati acquisiti ($\{ \langle P_1, t_1 \rangle, \langle P_2, t_2 \rangle, \dots, \langle P_{n+1}, t_{n+1} \rangle \}$, con $t_1 < t_2 < \dots < t_{n+1}$). Poste le traiettorie oggetto dell'elaborazione costituite da n ed m segmenti di linea rispettivamente, l'algoritmo utilizzato nell'approccio "spazio-tempo" può essere riassunto come segue:

- **STEP 1:** calcolo dei punti " W_k " di intersezione spaziale fra i m-points sfruttando l'operatore OGC *st_intersection()* [OGC, 2007]. Il costo per il completamento di questo step è determinato dall'algoritmo *Plane Sweep* [13] in esso implementato, ovvero $O((n+m) \log(n+m))$ dove l rappresenta il numero delle intersezioni geometriche.
- **STEP 2:** per ogni punto di intersezione individuato allo STEP 1, si ricerca in quale intervallo di tempo $[t', t'')$ esso sia stato attraversato, essendo t' e t'' i timestamps di campionamento degli estremi del segmento di linea in cui W_k è contenuto. Per il completamento di questa operazione è necessario verificare l'appartenenza di ciascun W_k ad ogni segmento di linea di ciascuna traiettoria, ottenendo una complessità computazionale totale pari a $O(l \cdot (n+m))$.
- **STEP 3:** per ogni punto W_k , si confrontano gli intervalli temporali individuati sulle due traiettorie per determinare, in tempo costante, se esse hanno un periodo $[t_{ini}, t_{fin})$ in comune. In caso affermativo risulta individuata un'intersezione spazio-temporale e, in accordo con la volontà di modellare le fonti di incertezza insite nell'acquisizione dei

punti di campionamento, vengono aggiunti al risultato l'intervallo $[t_{ini}, t_{fin})$ e la regione poligonale ottenuta bufferizzando il punto W_k con un cerchio di diametro proporzionale alla velocità massima raggiunta dai m-points durante l'osservazione.

La $O(g(n))$ per l'algoritmo “spazio-tempo” può essere complessivamente stimata in $O(l^*(n+m))$ ma ai fini della valutazione delle prestazioni di questo approccio, non si deve trascurare l'incidenza negativa del fattore $O((n+m+l) \log(n+m))$, necessario al completamento dello STEP 1.

Passiamo, ora, a presentare l'approccio “tempo-spazio” al calcolo delle intersezioni fra coppie di traiettorie. Come testimoniato anche in [Di Felice e Orsini, 2011], in cui viene discussa questa soluzione progettuale, l'analisi della scena di interesse parte dallo studio dell'evoluzione temporale delle traiettorie coinvolte. L'idea base consiste nell'individuare tutte le coppie di intervalli temporali che si sovrappongono (almeno parzialmente) sui due m-points e accertare, per ciascuna coppia, se i corrispondenti segmenti di linea si intersecano o meno geometricamente. Questa modalità di affrontare la problematica, complementare alla precedente, sfrutta l'ordinamento delle traiettorie rispetto al tempo per completare l'intera elaborazione mediante una loro unica scansione. Il test di intersezione geometrica sui segmenti di linea viene effettuato a seguito di una bufferizzazione degli stessi, al fine di tener conto delle molteplici fonti di incertezza di cui si è già riferito. Segue l'algoritmo che dettaglia le operazioni dell'approccio “tempo-spazio”.

Algoritmo time_meet (integer IdTraj1, integer IdTraj2, float r)

Input: gli interi $IdTraj1$ e $IdTraj2$ che identificano la coppia di traiettorie coinvolta nel test, il valore r che esprime il raggio del buffer costruito intorno ai segmenti di linea proiettati nel piano Cartesiano.

Output: la tabella $result(IdTraj1, IdTraj2, Intersection_geometry, Initial_time, Final_time)$

Metodo:

1. Siano $trjA = \{ \langle A_1, t_1 \rangle, \dots, \langle A_{n+1}, t_{n+1} \rangle \}$ e $trjB = \{ \langle B_1, s_1 \rangle, \dots, \langle B_{m+1}, s_{m+1} \rangle \}$ le traiettorie identificate da $IdTraj1$ e $IdTraj2$ nella base di dati
 2. **FOR EACH** coppia $\{ [t_i, t_{i+1}), [s_j, s_{j+1}) \}$ di intervalli temporali sovrapposti rilevati attraverso la scansione sincronizzata di $trjA$ e $trjB$ **DO**
 3. **IF** ($st_intersects(st_buffer(\overline{A_i A_{i+1}}, r), st_buffer(\overline{B_j B_{j+1}}, r))$) **THEN**
 4. $geom = st_intersection(st_buffer(\overline{A_i A_{i+1}}, r), st_buffer(\overline{B_j B_{j+1}}, r))$
 5. $t_{ini} = \max \{ t_i, s_j \}$
 6. $t_{fin} = \min \{ t_{i+1}, s_{j+1} \}$
 7. **INSERT INTO** $result$ **VALUES** ($IdTraj1, IdTraj2, geom, t_{ini}, t_{fin}$)
 8. **END IF;**
 9. **END FOR**
 10. **RETURN** $result$
-

La scansione sincronizzata (riga 2) consiste nell'avanzamento di segmento in segmento sulle due traiettorie coinvolte nell'elaborazione e in una progressiva verifica della sovrapposizione degli intervalli temporali individuati (con riferimento alla Fig.3, alla prima iterazione $[t_1, t_2]$ e $[s_1, s_2]$ non sono sovrapposti, mentre alla successiva $[t_2, t_3]$ e $[s_1, s_2]$ hanno in comune il periodo $[s_1, t_3]$). Solo quando vi è sovrapposizione temporale fra gli intervalli si procede alla verifica dell'intersezione geometrica fra i segmenti bufferizzati ad essi associati (in Fig.3 non c'è intersezione fra $\overline{A_2A_3}$ e $\overline{B_1B_2}$ e lo stesso vale per $\overline{A_3A_4}$ e $\overline{B_2B_3}$, mentre $\overline{A_4A_5}$ e $\overline{B_3B_4}$ si incontrano geometricamente nella regione R_1). In caso affermativo si aggiungono al risultato l'area individuata con l'operatore `st_intersection()` (riga 4) e il periodo di sovrapposizione (in Fig.3, $[s_3, t_5]$).

Sapendo che l'operatore `st_intersection()` applicato su segmenti di linea unitari risponde in tempo costante ($O(1)$) e che la "camminata" sincronizzata richiede un tempo lineare nel numero dei segmenti di linea delle due traiettorie, segue che nell'approccio "tempo-spazio" la complessità computazionale dell'algoritmo *time_meet* è pari a $O(n+m)$.

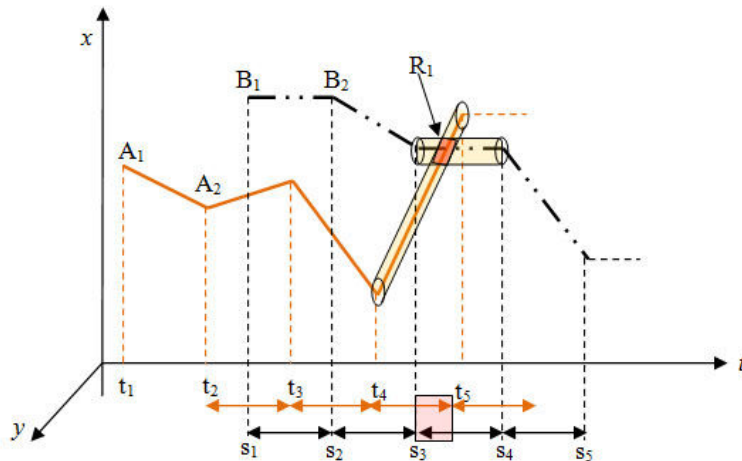


Fig. 3 – Due traiettorie coinvolte nel calcolo dell'intersezione nell'approccio tempo-spazio

5 Realizzazione e misurazione delle prestazioni

Per misurare le prestazioni dell'algoritmo *time_meet* e validarne il funzionamento si è deciso di implementarlo come User Defined Function nel linguaggio procedurale PL/pgSQL. Quindi, lo si è aggiunto all'insieme delle funzioni *built-in* messe a disposizione dal DBMS open-source PostgreSQL (vers. 8.4.3) abilitato con l'estensione spaziale PostGIS (vers. 1.4.2).

La componente geometrica e quella temporale delle traiettorie sono state modellate tramite due strutture separate, rispettivamente una `linestring` (definita in [OGC, 2007] come una curva con interpolazione lineare tra i

punti) e un array di `timestamps with time zone`, con il significato che l'*i*-esimo punto che compone la `linestring` è stato campionato nell'istante specificato all'*i*-esima posizione dell'array. Nel dettaglio, per la registrazione delle traiettorie viene, quindi, usata la seguente tabella:

```
trajectory(
  Pkey:      integer,
  Shape:     geometry,
  Timevalues: timestamp with time zone ARRAY)
```

Le complessità computazionali stimate nella sezione precedente non lasciano dubbi sul fatto che l'approccio da preferire per l'implementazione dell'operatore `time_meet()` sia quello che segue la strada che parte dall'analisi dell'evoluzione temporale dei m-points. Le traiettorie sintetiche utilizzate negli esperimenti seguono due andamenti complementari che possiamo succintamente riassumere nel seguente modo (per approfondimenti si rimanda a [Orsini e Di Felice, 2011]):

- *traiettorie random*: le coordinate dei punti costitutivi sono generate in maniera casuale e gli istanti di tempo associati vengono incrementati progressivamente di una quantità costante predefinita (Δ). Le traiettorie seguono andamenti nervosi e imprevedibili e, all'aumentare della cardinalità dei punti di campionamento, tendono a riempire completamente lo spazio di osservazione (come mostrato in Fig.4a). L'analisi della realtà consente di affermare che le *traiettorie random* sono idonee a simulare gli spostamenti in spazi delimitati (quali riserve per bestiame oppure aree indoor sorvegliate quali, ad esempio, aeroporti per le persone e, infine, aree urbane nel caso si automobili). È corretto aggiungere che le traiettorie random computazionalmente costituiscono una situazione critica. La criticità è causata dal fatto che, all'aumentare del numero di punti, e di conseguenza del periodo di osservazione dei m-points, tali traiettorie riempiono rapidamente l'intero spazio ove essi si muovono;
- *traiettorie semi-random*: l'intervallo di campionamento Δ viene determinato in base a considerazioni sulla velocità massima assunta dai m-points e queste due informazioni vengono congiuntamente utilizzate per stabilire la distanza che può essere coperta dall'oggetto nel periodo Δ . Le traiettorie seguono andamenti più lineari rispetto alle precedenti e sono assimilabili agli spostamenti dei veicoli che si muovono all'interno della rete stradale extraurbana più o meno estesa (come mostrato in Fig.4b).

I tempi rilevati dalle misurazioni di `time_meet()` su traiettorie random sono notevolmente peggiori rispetto a quelli ottenuti su traiettorie semi-random (si passa dall'ordine dei giorni a quello delle ore nel caso *spazio-tempo* e viene dimezzata la quantità di secondi necessaria nel caso *tempo-spazio*). I valori proposti in Tab.1 testimoniano come al crescere della cardinalità (*n* ed *m*) delle traiettorie coinvolte nell'elaborazione, le prestazioni dell'algoritmo *tempo-spazio* siano assolutamente soddisfacenti e,

coerentemente con la valutazione della complessità computazionale, decisamente migliori rispetto a quelle dell'approccio *spazio-tempo*.

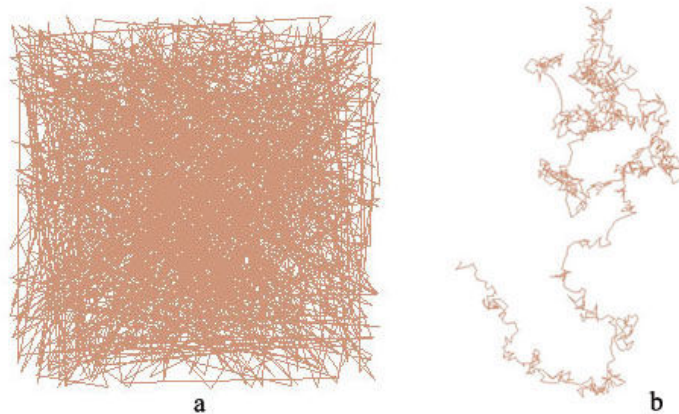


Fig. 4 – Un esempio di traiettoria random (a) e di traiettoria semi-random (b) di cardinalità $n=1000$

n	m	<i>spazio-tempo</i>	<i>tempo-spazio</i>
50	50	$\cong 0,8$ sec	$\cong 0,2$ sec
500	500	$\cong 32$ sec	$\cong 1$ sec
1000	1000	$\cong 8$ min	$\cong 2,5$ sec
3000	3000	$\cong 2$ h	$\cong 10$ sec
5000	5000	$\cong 5$ h 30 min	$\cong 25$ sec

Tab.1 – Tempi di esecuzione di `time_meet()` su traiettorie semi-random e buffer di 25 m.

6 Conclusioni

Consistentemente con l'architettura presentata in [Di Felice et al, 2010], l'operatore `time_meet()`, discusso in questo articolo, costituisce una testimonianza tangibile di come intervenendo al livello del *Gestore delle risorse* si possano creare le premesse per agevolare la successiva implementazione di applicazioni che coinvolgano dei m-points. Aggiungendo `time_meet()` alla collezione di funzioni spaziali messe a disposizione dai principali DBMS commerciali, si rende l'operatore immediatamente fruibile dalle aziende che volessero effettuare investigazioni spazio-temporali sul patrimonio informativo a loro disposizione, senza gravarsi di ulteriori investimenti nel software di supporto.

Si segnala, infine, che in [Di Felice e Orsini, 2011] si fornisce una testimonianza esplicita circa le potenzialità offerte dall'operatore `time_meet()` per condurre indagini che coinvolgano i m-points, che vanno dalla ricerca di quali oggetti si sono incontrati in specifici aree d'interesse in

un predefinito lasso di tempo, alla individuazione delle aree urbane più trafficate in talune ore della giornata. In definitiva, la possibilità di definire nuovi operatori spazio-temporali utilizzando l'approccio proposto in questo articolo rappresenta una prospettiva più che incoraggiante ai fini dello sviluppo di soluzioni informatiche robuste ed efficienti che coinvolgano oggetti in movimento.

Bibliografia

Cotelo Lema, J.A., Forlizzi, L., Güting, R. H., Nardelli, E., Schneider, M., Algorithms for Moving Object Databases. The Computer Journal, 46, 6, 2003, 680-712.

Di Felice, P., Liguori, G., Cestra, G., Un'architettura per lo sviluppo di applicazioni riguardanti dati spazio-tempo dipendenti, in AICA, 2010.

Di Felice, P., Orsini, M., Spatio-temporal intersection of trajectories under uncertainties, in IXX Simposio sui Sistemi Evoluti per Basi di Dati Symposium, Maratea, 2011.

Forlizzi, L., Güting, R.H., Nardelli, E., Schneider M., A Data Model and Data Structures for Moving Objects Databases, in Proc. of the Inter. Conf. on Management of Data, pp. 319–330, 2000.

Frentzos, E. K.: Trajectory Data Management in Moving Object Databases. PhD Thesis, 2008.

Giannotti, F., Pedreschi, D., Mobility, Data Mining and Privacy. Springer, 2008.

Güting, R.H., Behr, T., Duntgen, C., SECONDO: A Platform for Moving Objects Database Research and for Publishing and Integrating Research Implementations. INFORMATIK BERICHT, 356 – 04/2010, University of Hagen, Germany, 2010.

Open Geospatial Consortium Inc.: OpenGIS Implementation Specification for Geographic information - Simple feature access (2005) – Part 2: SQL option (ref. number: OGC 06-104r4), 2007.

Orsini, M., Di Felice, P., Generatori software di traiettorie di oggetti in movimento per valutare le prestazioni di operatori spazio-temporali. ASITA, 2011.

Pelekis, N., Frentzos, E., Giatrakos, N., Theodoridis, Y., Supporting Movement in ORDBMS – the 'HERMES' MOD Engine. Technical Report Series, Departement of Informatics, University of Piraeus, UNIPi-INFOLAB-TR-2010-01, 2010.

Pfoser, D., Jensen, C., Capturing the uncertainty of moving object representation, in SSD, 1999.

Photovoltaic System for Energy Efficiency in Urban Mobility

Monica Donno¹, Aleck Ferrari², Annalisa Scarpelli³,
Alberto Bocca⁴, and Alberto Macii⁴ *

¹ MECT s.r.l., Italy

² INTEGRA renewable energies s.r.l., Italy

³ ACTA s.r.l., Italy

⁴ Politecnico di Torino, Italy

Abstract. Research of real and immediate solutions for pollution reduction requires great efforts for achieving a more sustainable life.

This paper presents a retrofit system for conventional diesel-motor buses, by applying a mechatronic system that includes photovoltaic (PV) modules upon the vehicle. This solution takes place after considering that technology transfer for large bus fleets, from traditional to green transport, requires great investments and, therefore, it is always considered a long term plan.

Experimental results report that the proposed system of 1.76 kWp improves the bus energy efficiency and, thus, leading to greenhouse gas emission mitigation. In fact, it could allow to save several hundred liters of fuel per year, for each bus, after considering annual total solar energy production of 1.4MWh in typical weather conditions in Turin, Italy.

1 Introduction

From the environmental point of view, transformations to accomplishing the goal of a better life quality, especially in urban areas, are usually really slow if considering the so fast modern technology progress. Actions for reducing air pollution are sometime even insufficient with respect to the true requirements established in international agreements on climate change and the following decisions like no.406/2009/EC of the European parliament and of the Council of the European Union [1] on the maximum allowed greenhouse gas emissions for the period from 2013 to 2020.

In Europe, the urban traffic is responsible for 40% of CO₂ emission [2]. Innovation in public transportation is therefore one of the most area of interest for application of renewable energy sources. Thanks to new emerging solutions for electricity generation, a real change is taking place, though very gradually.

In general, towns with large fleets of diesel-motor buses have already begun a program of replacing conventional buses with new generation vehicles (e.g., hybrids, fully electric and eco-fuel buses). However, full transition requires years of investments and

* This paper has been produced as part of the SIMEBUS project, which is co-funded by Regione Piemonte (POR-FESR 2007-2013), "Polo regionale d'innovazione della Meccatronica e dei Sistemi Avanzati di Produzione" (MESAP), Italy.

therefore other solutions of rapid implementation are also needed as near-term applications.

An industrial project, named SIMEBUS (i.e., "Sistemi MEccatronici per l'efficienza energetica nei BUS"), is developing and validating mechatronic systems for retrofitting traditional middle-age buses. The main goal is the installation and testing of mechatronic modules on two bus prototypes in order to improve the energy efficiency of the vehicles and, accordingly, to reduce fuel consumption.

This document describes the solution for the first prototype equipped with Photo-Voltaic (PV) modules. The electrical energy generated can keep the on-board chemical batteries charged or recharge them whenever necessary.

The paper is organized as follows: Section 2 reports a brief background and an overview on previous works about electric vehicles, whereas Section 3 describes the proposed system. Finally, Section 4 reports experimental results and Section 5 draws some conclusions.

2 Background and Previous Work

The most common and traditional electric bus used in the European public transport is the trolley bus. Nevertheless, a new generation of autonomous electric vehicles (i.e., independent from an external power supply) is available on the market. In fact, both fully electric and hybrid-electric buses are running in many towns. However, their presence in modern bus fleets still increases, but not as quickly as needed. Indeed, in England, a funding program (i.e., Green Bus Fund) [3] has been introduced by the government for supporting bus companies and local authorities in order to have new environmentally-friendly vehicles in public transport. For instance, more than hundred diesel-electric hybrid buses are already running in London and this bus fleet is expected to be of three hundred vehicles in 2012.

In Europe, many towns have started to purchase fully electric small-medium size buses, mostly for urban mobility in town centers (e.g., Torino, Roma, Madrid, and Bordeaux). Further, electric buses with supercapacitors (i.e., ultracapacitors or Electric Double Layer Capacitor) [4], have been already developed (e.g., [5]). For example, since 2008 in Shanghai (China), some vehicles powered by both capacitors and lithium batteries are even in service.

About application of photovoltaic modules on buses, proposals in literature are mostly prototypes. Garner [6] proposed the application of solar panels upon the car roof just twenty years ago, in order to generate also energy for internal ventilation. Today, this idea is carried out by some car makers, even in mass production (e.g., Toyota Prius).

In recent years, University of Salerno, Italy, has developed effective design methods and control systems for solar hybrid vehicles [7][8] in part also thanks to a national two-years research program (i.e., PRIN 2006).

However, there are currently quite a lot of proposals for transforming conventional vehicles in hybrids, but mostly for car industry. In fact, photovoltaic installations for urban transport is mostly increasing on static applications like battery charging stations for electric buses, solar-powered shelters lighting, and traffic signs.

In any case, thanks to the new annual European Electric Vehicle Congress (first event this fall in Brussels, Belgium) and, starting from 2012, the IEEE International Electric Vehicle Conference, a more rapid development of these technologies is then expected in the future.

3 The SIMEBUS system

In conventional buses, one or more alternators generate the electrical power. They usually take about $8 \div 10$ kW of mechanical power, which corresponds to $8 \div 12\%$ of the fuel engine power, in order to transform it into electricity with an average conversion efficiency of $30 \div 40\%$. The energy stored in chemical batteries is then returned to continuous loads (e.g., headlights) with efficiency of about $70 \div 80\%$ and to impulsive loads (e.g., door openings) of 60% . But, usually, city driving conditions requires frequent bus stops so that the alternators efficiency decreases and, as a result, the batteries energy is not always sufficient to supplying all the loads.

The mechatronic system, here presented, includes solar panels on the bus roof for generating electrical energy in order to recharge the original lead-acid batteries. In this way, the alternators have a reduced activity and a fuel economy is then achieved. Figure 1 reports a diagram of the system architecture.

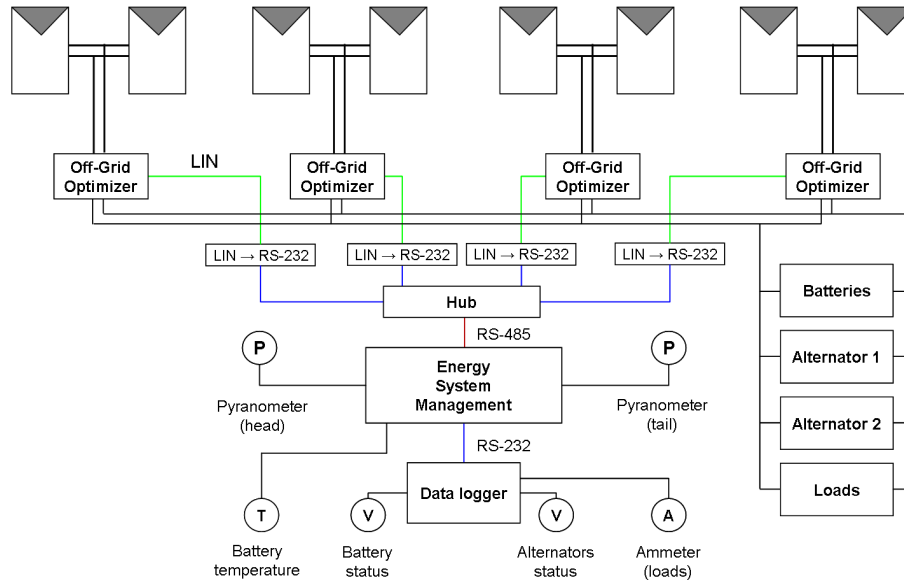


Fig. 1. The Proposed Mechatronic System.

Eight SunPower 220Wp Black PV modules, whose efficiency (η) is 17.7% (their monocrystalline silicon solar cell efficiency is about 23%), have been installed on the first prototype. Total power of the PV system is 1.76 kWp.

Even though total net surface of solar cells is about 9 m^2 , this solution has been adopted but having in mind any possibility of reducing the area in case of over-sizing, after analyzing experimental results.

Typical average solar production on horizontal surface, after considering a PV system at latitude $45^\circ 3'$ longitude North $7^\circ 41'$ East and at 237 m ASL, is about 0.92 and 4.62 kWh per day (i.e., sun hours for each installed kWp), worst and best case respectively (i.e., December and July). These values correspond to 1.62 and 8.13 kWh of produced energy in our case. But, absence of inverters for CC/CA conversion and the adopted technology increase the overall efficiency so that the average total energy improves of about 30%. Therefore, it is estimated to be 2.14 and 10.8 kWh.

Indeed, optimal working point of a PV panel depends on both solar irradiance and temperature. Of course, in urban mobility the solar cells of the system are subject to sudden and various shadowing conditions. So, either partial or full shadowing of a PV panel changes its Maximum Power Point (MPP) or, in other terms, its optimal working point, thus leading to power loss. The proposed system adopts the SolarMagic™ technology by National Semiconductor [9], whose MPP tracking (MPPT) controller has a fast response time for rapid changes of the MPP, in order to recover up to 71% of the lost output power.

This technology is included in an Off-Grid Optimizer (by Bitron), a device that also contains the following items:

- Overttemperature control circuit.
- Surge protection and filters.
- Microprocessor for controlling and programming the battery charging profile.
- Memory for data collection.
- Data communication module for Local Interconnect Network (LIN) bus.

This device can work up to 600Wp@24V or 300Wp@12V and meets the ISO 16750, 2004/104/EC and 2004/108/EC standards.

In a distributed architecture, the system provides four Off-Grid Optimizers (i.e., one device for every two panels that directly supply it). Each one provides information on various parameters through a Local Interconnect Network (LIN) bus.

An energy system management, named IKNOgrid, receives information on the PV modules status through a 4-wire RS-485 line, and also allows remote monitoring of some environmental conditions (i.e., solar irradiance). In fact, two pyranometers are also fixed upon the bus roof, one at the head and the other at the tail of the prototype. Collected data are then sent, through a RS-232 line, to an on-board data logger for analysis and back-up, which also reads load currents, and the batteries and alternators voltage. The whole system allows solar energy production in order to obtaining reduction of the fuel consumption, even the possibility for the bus to keep the engine off during long bus stops (e.g., terminal), but having its auxiliary services on (e.g., ventilation, lights).

4 Experimental results

An Iveco CityClass-Mod.491.12.22 bus (220 HP), with two lead-acid batteries (12V 220Ah each) in serial configuration and two alternators (120A and 55A respectively),

has been equipped with the proposed system. All data reported hereafter are referred to one-hour test.

Figure 2 shows the solar irradiance measured by the two pyranometers.

As you can see from the picture, at the beginning of the test the two curves, corresponding to the solar irradiance of the head and of the tail of the bus, do not have the same behaviour. This is because the bus, before the test, was parked with the head under the sun light and with the tail in a shadowed zone. Then, after this “warm-up” phase, the tail and the head of the bus show the same radiation behaviour.

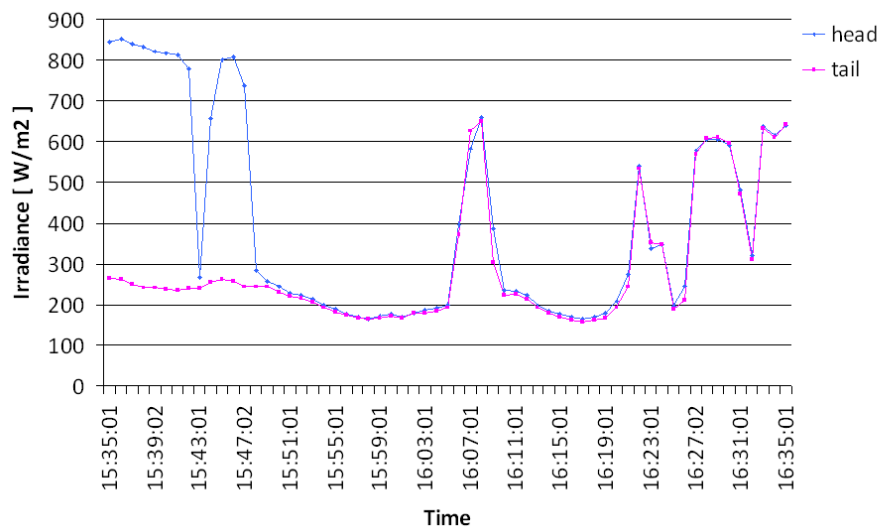


Fig. 2. Solar Irradiance.

Figure 3 shows the absolute temperatures of the Off-Grids. Values should be analyzed after considering that this test was performed in a sunny day at the end of July 2011. Each Off-Grid has a temperature threshold for safety reason, so that any device is turned off when 70°C are reached.

Figure 4 shows the input and output voltage for the MPPT1 optimizer. The output voltage is almost the same for all the devices, therefore only one chart is here reported.

Figure 5 reports the output current for each Off-Grid optimizer. Even though the output voltage is the same for all the Off-Grids, their output currents can be different one from another, as a consequence of the algorithm that manages the batteries charging profile. These values depend on both the maximum available input power (i.e., generated by the PV modules) and the battery voltage status.

Because of the very similar power reports of each optimizer, only one of the charts is here reported in Figure 6, in which it is evident that when the absolute values are reduced, with respect to the maximum admissible, the power efficiency decreases as well.

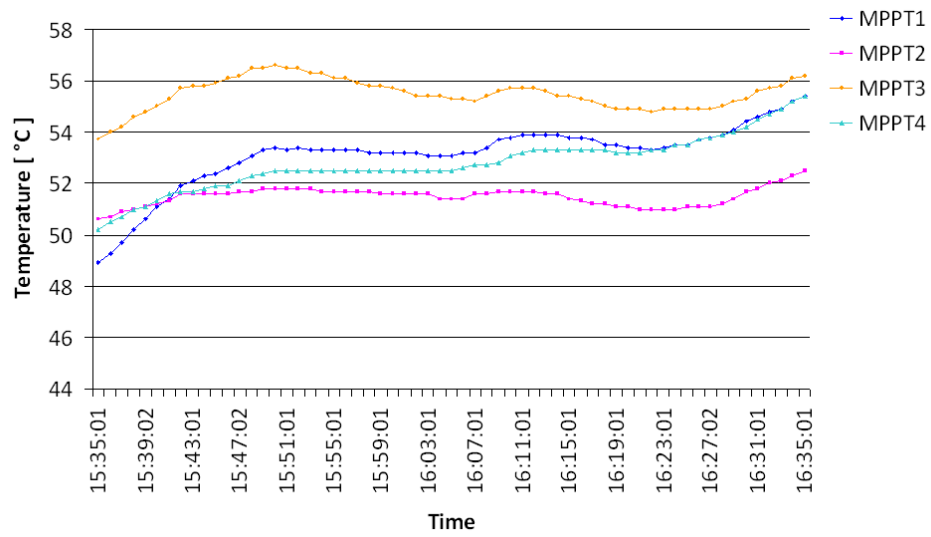


Fig. 3. Off-Grids Temperature.

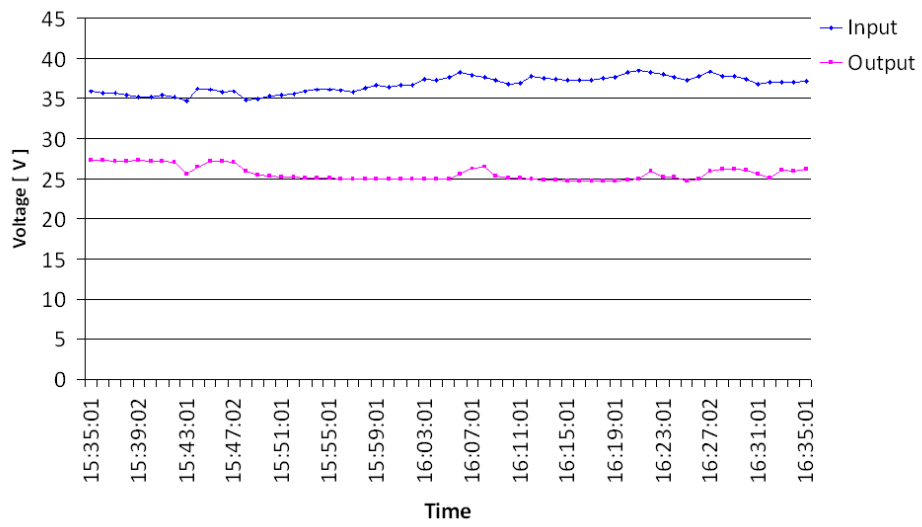


Fig. 4. MPPT1: input/output voltage

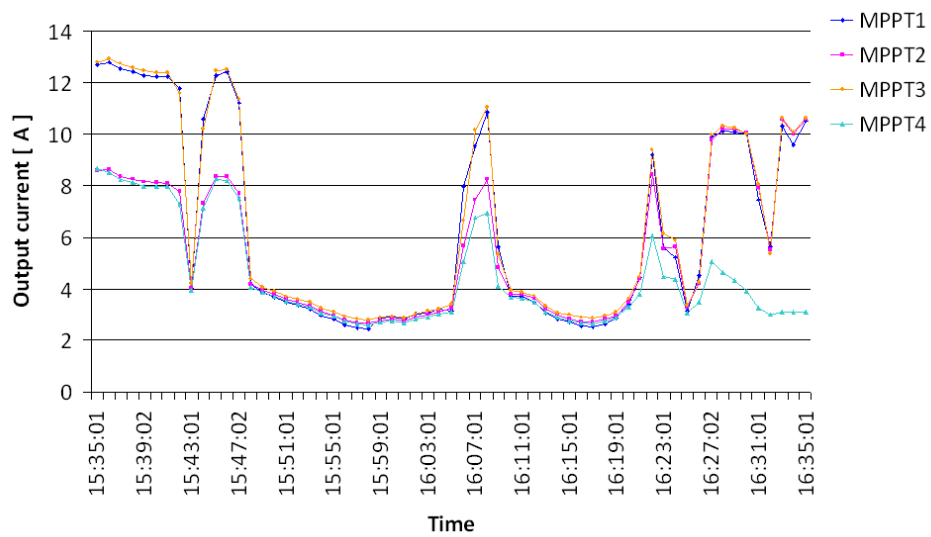


Fig. 5. Output current

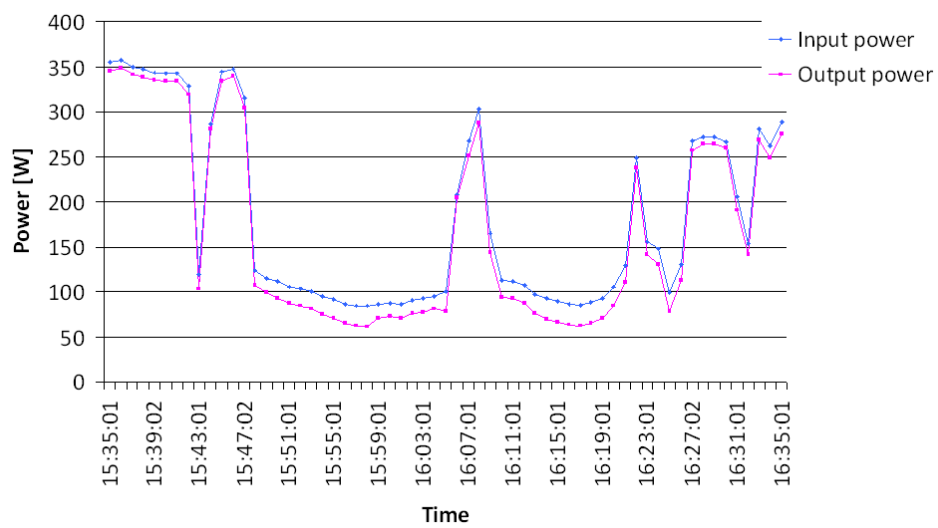


Fig. 6. MPPT1 power report

However, for a better comprehension of the results, Figure 7 reports, in a normalized form, the power balance for the whole system. At the beginning, alternators regularly work until the engine is turned-off. But, at this time, only the PV system can supply the batteries, and therefore the remaining active loads are still powered but without the condition of continuous battery discharge, as a consequence of the available solar energy from the PV modules.

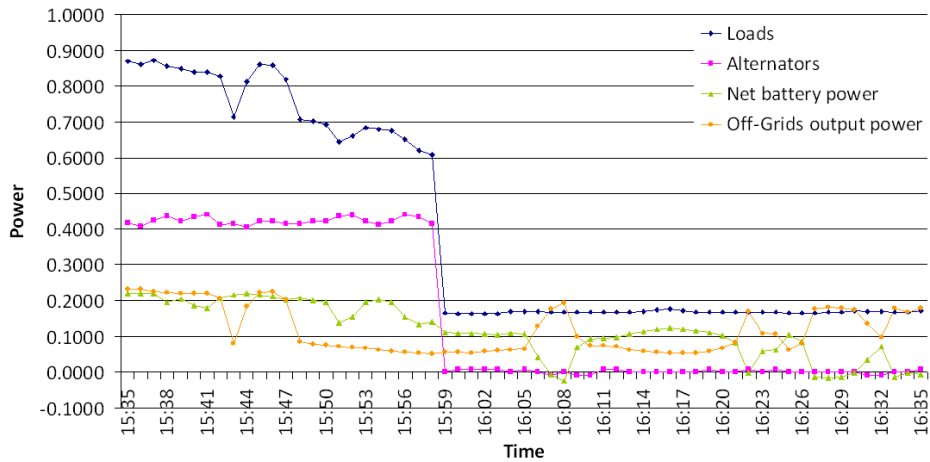


Fig. 7. SIMEBUS power balance

The experimental results confirm the effectiveness of the proposed solution that could lead to save 750 liters of diesel fuel per year, for each bus, after considering a total annual solar energy production of 1.4MWh. If applied to a bus fleet of 500 vehicles, this system could reduce greenhouse gas emissions, of about 650 tons of carbon dioxide (CO_2), 200 kgs of particulate matter (PM), and 3.5 tons of nitrogen oxides (NO_x).

5 Conclusions

In this paper we have presented a mechatronic system as retrofit for traditional diesel-motor buses in public transportation. Such a system includes PV modules for electrical energy generation, in order to recharge the on-board original chemical batteries and, among others, it includes also a management system for remote control.

Experimental results confirm the effectiveness of the proposed solution, which can lead to save several hundred liters of fuel per year, for each bus, and, therefore, to a remarkable greenhouse gas emission reduction.

Next step is to test a second prototype with a different mechatronic system solution, to further improve the total efficiency and, possibly, to disable one of the bus alternators.

References

1. *Official Journal of the European Union*, L 140, pag. 136, 5.6.2009
2. *Official Journal of the European Union*, C 21, pag. 58, 21.1.2011
3. <http://www2.dft.gov.uk/pgr/regional/buses/greenbusfund/>
4. A. Burke, "Ultracapacitor technologies and application in hybrid and electric vehicles", *International Journal of Energy Research*, John Wiley & Sons Ltd., Vol. 34, Issue 2, pp. 133-151, February 2010
5. R. Lu, C. Zhu, L. Tian, and Q. Wang, "Super-Capacitor Stacks Management System With Dynamic Equalization Techniques", *IEEE Transactions On Magnetics*, Vol. 43, No. 1, pp. 254-258, January 2007
6. I.F. Garner, "Vehicle auxiliary power applications for solar cells", *Proceedings of the Eighth International Conference on Automotive Electronics*, pp. 187-191, London, England, 28-31 October 1991
7. I. Arsie, G. Rizzo, M. Sorrentino, "Optimal Design of a Hybrid Solar Vehicle", *8th International Symposium on Advanced Vehicle Control*, Taipei, Taiwan, 20-24 August 2006
8. I. Arsie, R. Di Martino, G. Rizzo, M. Sorrentino, "Energy Management for a Hybrid Solar Vehicle with Series Structure", *17th World Congress of International Federation of Automatic Control*, Seoul, South Korea, 6-11 July 2008
9. P. Tsao, S. Sarhan, I. Jorio, "Distributed max power point tracking for photovoltaic arrays", *Proceedings of the 34th IEEE Photovoltaic Specialists Conference*, pp. 2293-2298, Philadelphia, PA, USA, 7-12 June 2009

Pro Camper Van Attrezzato per Organizzare Co-Pianificazione Nonstop e Rifondazione del Nuovo “Centro Storico” de L’Aquila

Pier Giacinto GALLI
Docta Interamna Universitas
Strada di Piedimonte 32, I 05100 Terni TR
piergiacintogalli@tin.it

Continuo a razionalizzare il discorso sul futuro del Centro Storico della Città de L’Aquila riallacciando tre circostanze: la prima, in occasione del Congresso Nazionale AICA 2010, è stata l’esibizione, nell’area espositiva dell’Aula Magna dell’Università dell’Aquila, di una selezione di lucidi, rilievi di fronti stradali e studi orografici per l’intelligenza delle genesi della forma urbana de L’Aquila, risalenti alla fine degli anni sessanta quand’ero studente presso la Facoltà di Ingegneria Civile Edile dell’omonima Università; la seconda, relativa al “CONTRIBUTO PERVENUTO” il 22 novembre 2010 al FORUM PA tenutosi presso il Campus G. Reiss Romoli dell’Università degli Studi dell’Aquila; la terza, in occasione della manifestazione, tenutasi a L’Aquila il 2 aprile 2011, alla quale ero stato invitato dall’organizzatore Comitato Diritti E Verità: nell’intervista <http://www.laquila99.tv/2011/04/02/diritti-e-verita-piazza-duomo-2-4-2011/> mi si chiedeva un estemporaneo parere sul restauro urbano che contestualizzavo, in 9’ e 44”, dal palco allestito in Piazza Duomo.

1. Introduzione

Il quesito postomi sul “restauro del quarto centro storico italiano” mi dava l’opportunità di ripetere il modo grafo-tecnico e compositivo che occorrono per coordinare le numerose variabili concomitanti alla complessità pianificatoria della ricostruzione. Condizione necessaria e non-sufficiente per rendere concreta la ricostruzione è la progettazione dello smontaggio di caseggiati non ristrutturabili in conformità a analisi Benefici / Costi-della-sicurezza. Smontaggio, al fine di riutilizzarne i materiali di risulta, opportunamente selezionati e preventivamente catalogati, come memorie identitarie che restituiscano a nuova vita insieme di elementi architettonici. Come? Ri-murando i materiali da costruzione e inserendo detti elementi architettonici come decorazioni per evitare la dilapidazione degli stilemi: questa fisionomia edilizia del futuro nuovo centro storico, risponde non solo al criterio di trasmettere il passato al futuro ma anche risponde a esigenze di sostenibilità economica e

risparmio di energia potenziale. Questo modo di ricordare, percorsi e superfici [RIF 3] tra volumi sostanzialmente irrecuperabili, tutela valori associati al concetto di restauro in evoluzione: rimodellare materiali demoliti o smontati per ricomporli in nuove forme edilizie in complementazione con materiali attuali, significa sostanziare nelle più progredite tecnologie edilizie i resti emblematici di habitat medioevale, rinascimentale e barocco, tre concezioni oikologiche non-predisposte a ricevere, nel secolo scorso, l'arrivo in casa dell'acqua, gas, elettricità, automobili, cavi e cavedi, fognature, impianti di depurazione e tecnologici di captazione. Valorizzare i residuati come simboli per ornare ed integrare i nuovi materiali a disposizione della più avanzata edilizia, far così rinascere il *Downtown* dalle sue macerie come la Fenice, ... "l'uccello sacro favoloso dallo splendido piumaggio che aveva l'aspetto di aquila reale", sono cose da artisti raffinati, da tecnologi esperti, da urbanista creativo disimpegnato dalle incombenze civiche che oberano i servizi amministrativi comunali.

2. Visita urbanistica

La visita urbanistica attua la prima fase della procedura logico-pratica che porta alla estensione della diagnosi architettonico-ingegneristica: la registrazione delle informazioni cliniche. Il quadro patologico è in via di definizione perché, sebbene **l'anamnesi clinica** della città possa ritenersi condensata nella mia Tavola di Sintesi dell'anno accademico 1968/69



Fig.1 - Scansione 'A4' presa a caso sulla Tavola (exETA 89x60; 1:2000) di Sintesi (89x70cm) "Semiotica ed etiologia per una tavola morfologica della Città de L'Aquila", A.A. 1968/69. Emergenze di San Pietro, San Silvestro, Santa Maria Paganica

("Semiotica ed etiologia per una tavola morfologica della Città de L'Aquila") tuttavia, **l'esame obiettivo** è in corso di definizione con la raccolta dei segni clinici della città aggiornati al 09-09-2011 (http://www.riplanificazione.altervista.org/files/ambitoagggregati_09_09_2011.pdf) e della "zona rossa" aggiornati allo 11-08-2011 (http://www.riplanificazione.altervista.org/files/zona_rossa_11_ago_2011.pdf) nonché nelle altre cartelle cliniche su - per esempio - crolli, demolizioni parziali

eseguite e da eseguire, o, sui vincoli ... La provvisorietà di siffatti rilevamenti si evince dall'incoerenza delle legende e delle simbologie, dalla incompleta rappresentazione delle reti tecnologiche, dalla quasi inesistenza dell'impianto fognario. L'anamnesi è cittadina, fisiologica, sulle patologie remote e prossime nelle sue svariate sintomatologie. Questo esame fisico o semiologico svolto dall'amministrazione comunale sulla città traumatizzata è perciò ancora incompleto, sia nella sua parte **generale** ovvero morfologica, sia in quella **speciale** cioè rivolta alla coerente registrazione, di tutti i segni e legende, sulle tavole dei vari distretti: la costituzione architettonica, la funzionalità del sensorio e i numerosi segni indiretti sono percepibili dall'occhio clinico dell'urbanista che valuta come un architetto, misura come un ingegnere e prepara le operazioni come una *equipe* chirurgica.

2.1 Diagnosi (discernimento) e terapia urbanistica

Il terremoto è un'emergenza complessa che richiede, tra l'altro, interventi immediati. La diagnosi urbanistica si sviluppa in modo da classificare i vari livelli di ammaloramento dell'edilizia scossa; le terapie d'intervento sono varie [RIF 1]: le terapie ricostruttive si applicano soltanto nello stadio successivo a quelli dello smontaggio dei ruderi, selezione delle macerie, accantonamento dei materiali-di-risultato compresi i detriti triturbili additivabili e calcinabili *in situ*.

Un trattamento terapeutico preparatorio a interventi di chirurgia urbanistica, pena la morte della città in un 'coma farmacologico' che non può essere disturbato ... A prestabilite operazioni di 'raccolta differenziata' e conservativa per il riciclaggio in loco delle macerie non sembrerebbe aver dato il contributo appropriato quell'esasperato 'parentado' - il "Popolo Delle Carriole" - mobilitatosi, in una fuga accelerata verso un futuro, senza discernimento, che sembrerebbe collocarsi al di fuori di un dibattito finalizzato a un coordinamento progettuale, ma che viceversa sembra rispecchiare soltanto la richiesta di una terapia del dolore invocata da gente spaesata e scollata dalle istituzioni locali attive sul fronte delle azioni "a breve" e degli interventi d'edilizia "sostitutiva" senza però avere alcuna chiara idea sul plastico del futuro centro storico da ripensare contestualmente al nuovo impianto fognario funzionale a un diverso sistema di superfici pubbliche e di volumetrie private, prospettabili ai condomini del "centro storico" e alla cittadinanza che entrambi dovrebbero pensare in termini di adeguamento alle nuove realtà dell'auto elettrica e delle tecnologie antisismiche. Questo richiede l'attitudine dell'immaginazione per consentire una programmazione da non confondere con la cosiddetta "riplanificazione" che, in realtà, è puro monitoraggio delle informazioni che, complete e attendibili, consentiranno di sciogliere il nodo gordiano con decisioni ramificate in accordo all'evoluzione dell'algoritmo *Program Evaluation and Review Technique*. La prospettata riutilizzazione di materiali produce nuove forme architettoniche dislocabili in ambiti urbani adeguati al terzo millennio, forme e ambiti sono due idee creative entrambi volte a dominare il futuro con progetti ulteriori alle offerte di Google [RIF 5] di riprodurre una memoria virtuale a tre dimensioni che però è solo uno strumento di lavoro e di transito virtuale verso la realtà materiale.

Una cosa è l'anestesia nel corso di risolutorie operazioni di chirurgia urbanistica, altra cosa è la palliativa terapia del dolore psicologico dovuto alla non accettazione della rimozione di una realtà che sarà inesorabilmente diversa dal ricordo che coincide con la retrospezione sul passato, retrospezione implicita nel Protocollo Di Intesa del 13 ottobre 2010 tra Comune de L'Aquila e MBCA per l'istituzione della commissione per la verifica di ammissibilità delle proposte d'intervento nel centro storico di L'Aquila e frazioni (art. 6 c. 3 DCD 3/2010) con Allegato di Prescrizioni per gli interventi in centro storico di L'Aquila e frazioni. Il Consiglio comunale straordinario, dell'11 luglio 2011, sulla ricostruzione, sembra non abbia risolto l'*impasse* creato dalle Prescrizioni adatte a conformistici esercizi per manutenzioni ordinarie di una città che non abbia mai subito il *vulnus* sismico, per l'appunto, in quell'edilizia di cui si considerava la bonifica già alla fine degli anni '60 con proiezione al 1990.

2.2 Terapie funzionali alle operazioni di chirurgia urbanistica

Le terapie per le "zone rosse" (http://www.riplanificazione.altervista.org/files/zona_rossa_11_ago_2011.pdf) non sono interventi di manutenzione ordinaria e straordinaria o di restauro e risanamento conservativo ma sono piuttosto le ristrutturazioni edilizie: nuove costruzioni che comportano ristrutturazioni urbanistiche [RIF 1] intorno alle emergenze socio-architettoniche, inalienabili cardini per programmi di Ricerca Operativa che scongiurino la 'pompeizzazione' del "Centro Storico" che va visto come un tutto olistico con le "aree di frontiera" onde evitare una 'agrigentizzazione' del *Downtown*-L'Aquila. Questa impietosa esemplificazione occorre nonostante il Servizio comunale apposito alla "riplanificazione" abbia fatto mettere in rete una notevole mole cartografica e documentale aggiornata allo 09.09.2011; tali *file* rappresentano il monitoraggio propedeutico agli interventi a breve. La relativa registrazione di informazioni e dati è in alcuni casi incerta perché a corto di legende, indispensabili per chi volesse abbozzare una qualche pianificazione d'insieme entro cui calare moduli progettuali coerenti a una completa *time table* per altro congelata di fatto dal Protocollo D'Intesa Comune-MBAC e dalle Prescrizioni per gli interventi nei Centri Storici che per l'appunto sembrano conformistici esercizi per manutenzione ordinaria di una città che non abbia mai subito un *vulnus* sismico proprio in quell'edilizia di cui si considerava la bonifica già alla fine degli anni '60 con proiezione al 1990!

3. Il Camper Van nella "zona rossa"

È pensato come centro neuronale di sintesi, una ghiandola pineale, per la stimolazione e controllo progettuale della co-pianificazione *nonstop* fino alla verifica di congruità tra tabelle di marcia e logiche decisionali volte alla produzione dei *kit* grafici eseguibili secondo le tempistiche di cantieri modulari e sequenziali. Cantieri non tanto adibiti a interventi di manutenzione ordinaria e straordinaria o di restauro e risanamento conservativo quanto piuttosto adibiti alla ristrutturazione edilizia ovvero alle nuove costruzioni che comportano ristrutturazioni urbanistiche [RIF 1] intorno alle emergenze socio-architettoniche conosciute come insostituibili perni per programmi d'interventi da studiare per

integrare la “zona rossa” e i suoi interstizi di connessione con le “aree di frontiera”, tutti volumi - aperti e chiusi - da riesaminare, riprogettare, ripianificare ed infine ricostruire. Tutte operazioni che devono essere scaglionate nel tempo, valutate e dirette con algoritmi più evoluti di PERT (*Program Evaluation and Review Technique*) comparabili con le offerte sbalorditive che IBM pubblicizza nei suoi *White Paper* su *Smarter Cities Software Solutions* prospettate per domini cittadini di Gestione Acque, Sicurezza Pubblica, Traffico, Edifici, Energia.

È la distaccata idea di captare *sul campo* tutto il buon senso critico necessario per aiutare il raggiungimento di una specifica più appropriata definizione di “sostenibilità”, concetto in continua evoluzione dal 1968 (Club di Roma) al 2010 (ISO 26000), anno in cui il concetto verte a una concreta gestione della qualità (ISO 9004).

Senza un compiuto piano olistico, che spazi dagli ‘edifici rossi’ del centro storico alle cosiddette “aree di frontiera”, sarà difficile attirare i benefici fiscali previsti per Zone Franche Urbane (ZFU) e i finanziamenti *ad hoc* per questa speciale porzione di territorio comunale; diventerebbe pieno di erbacce il futuro del cosiddetto “Centro Storico” se buoni concetti non siano immessi nella realtà lavorativa.

4. Sviluppi del capitale intellettuale versato nel 1968

Un difficile calcolo, d’interessi, maturati nell’intervallo compreso tra l’originaria scrittura della mia funzione plurivariabile Uomo Urbanista e

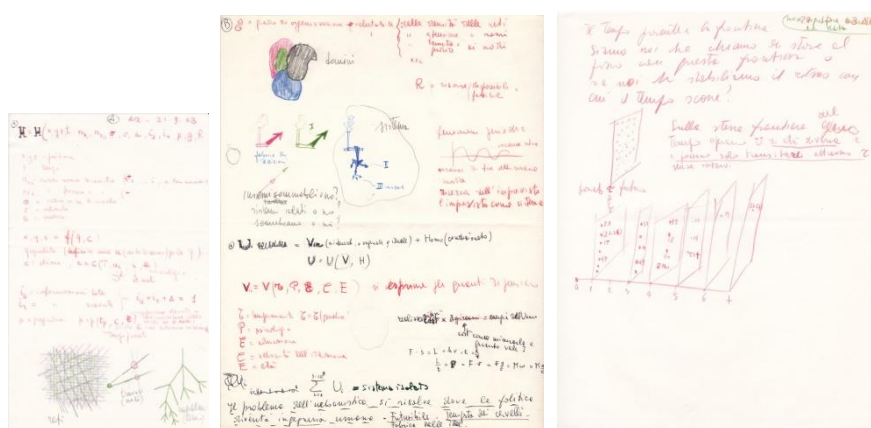


Fig.2 – Riduzione di tre foglietti olografi: formati 14x22cm e 22x28cm nei quali introducevo la funzione $U = U(V, H)$ [Urbanista = U (Vir, Homo)] in data “AQ, 21-9-68”, più foglietto 22x28cm in cui “la data 27.10.68 AQ non significa niente”

l’attuale implementazione di essa come ‘grandezza della fisica’: spiccioli di pensieri affermati nel 1968 e consolidati nella Tabella Logofisica (scienza che studia la natura e il comportamento del pensiero); il reso conto della creazione

di un valore immateriale è contenuto nella terza pagina della relazione [Sulla Natura Di “Eoni Matematici”: Neologismo Scaturito Dai Topoi Impliciti Nel «Paper \[1\]»](#) (DIDAMATICA 2011); nella fattispecie, la Tabella riguarda le qualità umane composite di tutte le combinazioni dimensionali delle grandezze fisiche manifeste e non manifeste come i - per ora - non meglio identificati Eoni Matematici; l'interezza di tutte le grandezza della fisica e di quelle allo stato astratto, di mera combinatorialità dimensionale rintracciabili in una sorta di 'pleroma' matematico, sono le qualità umane richieste ad un Urbanista, qualità che si identificano nella sua capacità di giudizio oltre che nella sua demiurgica funzione metrico - progettuale. L'Urbanista contiene l'interezza o quasi delle grandezze fisiche arricchite di un ethos proba-virale infusogli dalla presenza dello “eone-sophia” [RIF 4]; l'infusione delle dimensionalità del 'pneuma stoico' fa la differenza tra Demiurgo e Urbanista esperto; quindi l'Urbanista si distingue dal “demiurgo” il quale modella tutte le combinazioni dimensionali eccetto quelle in più che consentono al libero Urbanista di appartenere al, di vedere il pleroma gnostico delle combinazioni dimensionali puramente matematiche. La più recente “frontiera ontologica” risiede nel REF 3.

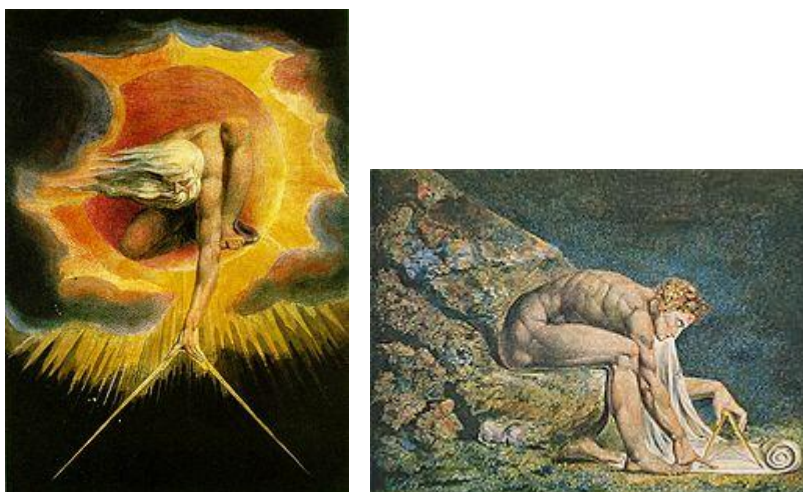


Fig.3 – Di William Blake: The Ancient of Days (1794) e “Newton” (1795)

5.Conclusioni

I servizi statistici e demografici del Comune insieme con altre funzioni istituzionali attinenti alla gestione dello stress da terremoto sono tutte incombenze di *routine* attribuite alla pubblica amministrazione perché diventino fondamenti statuali indispensabili e propedeutici: la qualità delle informazioni e il loro grado di probazione cartografica condizionano lo *start up* pianificatorio; l'urbanista “trae gli auspicj” guardando con gli occhi della mente oltre i ratei statistici pubblicati dal preposto ‘demiurgo’. Infatti, l'Urbanista ordinatore della co-pianificazione implementa i dati disponibili con sopralluoghi *de visu* affinché,

i disegni da elaborare a varie scale siano fra loro compatibili e congruenti, i co-progetti esecutivi siano adeguati alle attese. È evidente che, oltre alle capacità per assolvere i compiti routinari di servizio occorrono successive capacità per far decollare dalla cartografia più rispondente alla realtà, la vera e propria pianificazione; questa distinta attività creativa, idonea per le fasi pianificatorio-progettuali successive agli aggiornamenti della cartografia amministrativa, non può essere riduttivamente burocratica perché il pensiero-creativo-libero-professionale è un'altra cosa. Tra la burocrazia e la creatività si può instaurare la seguente equazione di proporzionalità: l'esecutività della funzione pubblica sta al visibile come le capacità immaginative - inventive dell'Uomo-Urbanista stanno all'invisibile; infatti, il demiurgo pubblica le misure del mondo fenomenico che registra con grandezze fisiche *standard*, diversamente l'urbanista libera la propria creatività perché disegna quelle dimensionalità in più tipiche del 'pleroma', dall'interpretazione del qual 'pleroma' i disegni prendono forma.

L'Urbanista è quella funzione introdotta in Fig.2 (che riporta alcuni di quei fogli, risalenti alla fine degli anni '60, la cui raccolta mi piace chiamare *Codice Aquilano*) ovvero è quella grandezza fisica arbitraria [Sulla Natura Di "Eoni Matematici": Neologismo Scaturito Dai Topoi Impliciti Nel «Paper \[1\]»](#) dove c'è il pensiero di combinazioni dimensionali extra interpretabili come entità astratte che stanno oltre la frontiera ontologica *standard*, in disposizioni topologiche dove, sebbene creativo, raggiungo il mio livello di incompetenza e di incapacità trasduttiva [RIF 3].

Bibliografia

[RIF 1] Antonucci D., Manuale di diritto urbanistico, SE/E35, 2004.

[RIF 2] Galli P. G., "studioso del Centro storico della Città dell'Aquila"
[CONTRIBUTO](#) [PERVENUTO](#) <http://saperi.forumpa.it/story/50975/forum-dellinnovazione-centro>
http://saperi.forumpa.it/sites/all/files/documents/file/FI_centro/WS3_PDF/Pier_Giacinto_GALLI.pdf, 2010.

[RIF 3] Galli. P.G., Some philosophical notes on the mathematical nature of eons and / or on the nature of mathematical eons, © D.I.U., Terni, Italy 18.08.2011.

[RIF 4] Goti F., "Studioso impegnato nella ricerca dello gnosticismo", Eone Gnostico
<http://www.paxpleroma.it/saggi%20di%20filippo%20goti/eonegnostico.html>.

[RIF 5] <http://www.downloadblog.it/post/13641/comefacciamo-google-sketchup-per-ricostruire-laquila> <http://blog.theblogtv.it/2011/06/17/noi-laquila-google-e-il-crowdsourcing-fanno-rivivere-laquila/> <http://www.tomshw.it/cont/news/l-aquila-rivive-in-3d-grazie-a-google-sketchup/28998/1.html>.

Quality of Service in Business Process

Marta Frascaroli, Matteo Lupi, Daniele Sacco
Università di Pavia – Dipartimento di Ingegneria Informatica
Via Ferrata 1, 27100 Pavia

marta.frascaroli01@unipv.it; matteo.lupi01@unipv.it; daniele.sacco01@unipv.it

Evolution of tertiary and quaternary sectors of economy led, in the last years, to a growing service market and induced a new definition of performing service. A competitive service is a profitable service, delivering a better quality than its competitors, thus service competitiveness is based on level of delivered performances. The main objective of this paper becomes to introduce Quality of Service (QoS) awareness into design of a business process. Early stages of BPM programs consider QoS as low priority, taking them into consideration only in the final stages. But performance issues lead to poor reception of the solutions, making them non usable to the users. However, there is lack of an overall framework to support definition of performance requirements within each business process. Our work will develop into definition of a modeling framework to design business processes for quality.

1. Introduction

1.1 Quality of Service

In the context of information technology, QoS is defined as “the set of qualitative and quantitative attributes of a distributed multimedia system that is necessary to achieve the required functionality of an application” [Vogel et al., 1995]. More in general, in computer and network engineering, “QoS is the ability to guarantee a certain level of performance to a data flow” [Evans and Filsfils, 2007]. This definition can be adapted to business process design as follows: “In the Business Process Domain, QoS could be defined as the ability to guarantee a certain level of performance of a service” [Pignatelli, 2010]. Our assumption is that QoS requirements should be described from the business level (what the customer perceives) down to implementation level (how the service is produced). QoS requirements are identified by three key elements. The first element is the Stakeholder. Actually, QoS depends on the stakeholders’ points of view: managers (look at the total cost of the service), end customers (look at the End-to-End QoS and at the effort they incur to get the service) and workers (look at the cost of the process portion they are involved in). The second element is Business Metrics, which include both quality oriented KPIs and SLAs (e.g. time, content and alike variables) and resource oriented KPIs (e.g. cost, productivity and alike). Finally the third element is the Business Process/Service to which the performance is associated. The issue is relevant and uncovered

since most research neglect the stakeholder dimension or do not integrate KPIs.

Designing business processes involves different levels of abstraction, each targeting a diverse scope and a diverse set of tools, models and competences. These three levels make the business process project life cycle: Business level

- Logical level - Execution level

1.2 Strategic literature review

The aim of this section is to produce a State of Art of the research in “Design of Services/Business Process for Performances” applying a systematic analysis, called Systematic Literature Review (SLR), to a selection of mainstream articles in 2003/2010 period. The results of this review will be used to outline the framework for designing business processes for QoS.

The SLR method originated from Biolchini et al. work [Biolchini et al, 2005], which applied a defined protocol template and a set of guidelines to a Literature Review in medical research. Afterwards, Kitchenham and Charters [Kitchenham and Charters, 2007] and Kitchenham et al. [Kitchenham et al, 2009] works, proved that this set of guidelines is suitable also for SLR in Software engineering. Our literature review was performed through the following activities:

1. **Question formulation**
2. **Source selection**
3. **Studies selection**
4. **Information extraction**
5. **Results Summarization**

2. Classification framework

Introducing QoS into business processes design is a complex task which should develop from the conceptual design to the implementation. Thus, in order to incorporate business needs, the framework includes requirements from the conceptual level down to implementation level. This model, in particular, is defined by three abstraction levels: **Conceptual - Logical - Execution**. Moreover, the representation of QoS requirements involves several domains: **Performances - Business Process - Integration between Performances and Business Process - Model Driven Approach** to model the elements of each level and to support the mapping among levels (cross-level domain).

Therefore, the classification model is obtained by crossing domains and levels of abstraction and it is represented by a matrix where each cell represents a particular design segment, as depicted in figure 1.

The purpose of the **Conceptual level** is to collect QoS requirements that reflect the organizational framework and the strategic objectives of a company for a particular given service. On the business process modeling side, company processes are described as black boxes while, on the performance side, conceptual level consists of frameworks which can help defining business performances. QoS requirements, at a conceptual level, are estimated on three aspects:

1. Stakeholders: The stakeholder perspective is important since services are increasingly delivered by complex supply chains and service networks, in which diverse actors cooperate;
2. Business Metrics, which include:
 - a. Quality oriented KPIs and SLA;
 - b. Resource oriented KPIs.
3. Business Process Elements associated to the performance.

KPIs properties can be modeled, at this level, using Dimensional Fact Model (DFM) and other Data Warehouse techniques.

The **Logical level** shows the sequence of activities which can be performed within a process. Activities flow can be described using Workflow Modeling Language such as UML Activity Diagrams, Business Process Modeling Notation or Event Driven Process Chain. In the Logical Level it is also appropriate to describe KPIs and to link the KPI schema to the workflow schema and to the logical rules, in order to have performance oriented activities or activities which trigger compensation actions in order to accomplish expected performances.

The **Execution Level** is the last level of abstraction and it includes execution languages, formats and execution engines. Many execution languages and engines such as Business Process Execution Language (BPEL) have been defined to run the processes and to implement the business rules. Within QoS requirements we introduced at the first level of abstraction, the purpose of this level is to get information on the run-time performances which can be used for several activities like business management control or real-time monitoring.

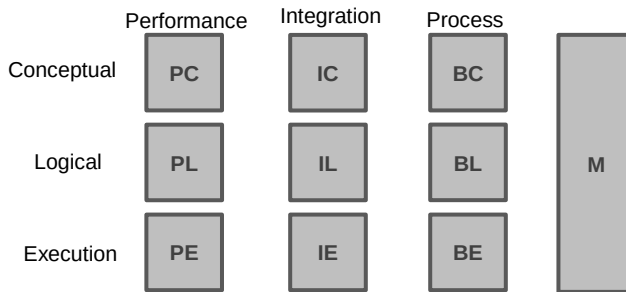


Fig. 1 – Classification model

2.1 Execution of the SLR

Question Formulation: the main purpose of our work is to identify all initiatives and experiences in Service Engineering related to the integration of real-time performance measurement to guide the workflow execution. We defined this Research Question (RQ): *“What is the current level of integration between the current approaches of definition of KPIs and the modeling techniques of business processes?”*

Source Selection: in this section we identify the most relevant types of sources to execute our Literature Review. We considered Journals and Conferences.

Studies Selection: first of all we defined a basic inclusion and exclusion criteria based on the RQ: initially the selection criteria were interpreted liberally and clear exclusions were only made with regard to title, abstract and introduction, moreover we included only studies which publication year is between 2003 to 2010.

Furthermore, in our SLR we considered the following study types:

- Case Study
- Theoretical
- Instrument Development
- Literature Review
- Multiple

Information Extraction: the purpose of this section is to extract from studies all kind of initiative (techniques, methods, models and strategies) useful to evaluate KPI-Business Process Integration. The rating was extracted following two ways of classification:

- General, that provides an overview about the trends of the research field analyzing several dimensions:
 - Year: Ranges from 2003 to 2010
 - Publication type, and research method / study type
 - Geographical area
- Model-Oriented. The starting point is the classification framework we discussed previously, which aims to provide the researchers with a classification and positioning framework.

Result Summarization:

Year	# Publications	Type	# Publications	Geographical Area	# Publications
2003	8	Case Study	19	Europe	89
2004	15	Theoretical	85	America	42
2005	27	Instrum. Development	16	Asia	16
2006	45	Literature Review	24	Oceania	11
2007	36	Multiple	22	Africa	6
2008	18			Others	2
2009	11				
2010	5				

Tab. 1 – Number of publications

As shown in Table 1, we can notice a large increase in the number of publications from 2003 to 2006. This fact demonstrates the raising interest in the topic. Although we observed a variety of research methods, the theoretical was the most common. From the Geographical point of view, we can notice that the contribution of American researchers is at least double than European ones.

2.2 Discussion

2.2.1 Conceptual Level

Within the conceptual level that refers to business processes domain, the most discussed topic is Business Processes Reference Models (BPRMs). In general, a Reference Model (RM) represents a general solution to a particular class of problem in a specific domain. The application of a RM is motivated by the “Design by Reuse” paradigm, well known in software engineering. Fettke et al. [Fettke et al, 2006] performed a very extensive survey on RMs which studied thirty BPRMs. Fettke’s analysis highlights many issues and problems about this topic showing how there is a substantial lack of implementation due to the fact that many BPRMs are not supported by tools and are partly not publicly accessible.

Furthermore only few authors explicit the procedure of their model construction and evaluate languages (such as EPC and UML) before designing a reference model. As a result of this research we present a short list of the most popular RMs:

- IT Infrastructure Library (ITIL) v3 (2007)
- enhanced Telecom Operation Model (eTOM)
- MIT Process Handbook (MIT-PH)

Most of the previous works are based on other frameworks developed before 2003 (our time-limit), like:

- Architecture of Integrated Information Systems (ARIS) (1990)
- Supply-Chain Operations Reference Model (SCOR) (1997)

BPRMs address the need to help the analyst in designing the Business Processes Architecture using normative models. As stated by Becker et al. [Becker et al, 2003], a reference model can be used as a blueprint for information system construction. In our point of view an ideal BPRM should contain: **Business Process - Implementation methodology - KPIs - Best practices – Tools**. Once business processes have been identified within the model, they could be:

1. Implemented and described in order to gain a competitive advantage;
2. Specialized in order to achieve a specific business objective;
3. Measured, managed and controlled.

On the side of the performances domain the conceptual level is the one with the higher number of studies. A rich management literature offers several Business Performances Measurement and Management Frameworks, so, as a result of this research, we present a list of the most popular Framework for Business Performance.

- Six Sigma (Adams et al., 2003)
- HIGO (Longo and Motta, 2006; Motta and Pignatelli, 2008)
- IPM (Xindi and Qingjie, 2006)

Also for the Business Performance Framework, the previous works are based on other frameworks developed before 2003, like:

- Activity-Based Costing (Turney, 1990)
- Balanced Scorecard (Kaplan and David, 1992)

- Total Quality Management (Powell, 1995)
- Economic Value Added (Rogerson, 1997)

2.2.2 Logical Level

At this level of abstraction, most of literature focuses on business process domain while only few papers focus on performances. This aspect of the research could be explained by the fact that once KPIs have been designed at conceptual level their “logical design” could be viewed as a graphical representation of the KPI itself. At this level, the focus is in particular on Business Process Modeling Languages (BPMLs). Literature offers many BPMLs which aim is to describe a business process. Different BPMLs describe different aspects of processes and address different areas of interest.

Among BPMLs the most discussed are:

- Business Process Modeling Notation (BPMN)
- Event-driven Process Chain (EPC)
- UML Activity Diagram
- Integration Definition (IDEF)
- Petri Net
- Yet Another Workflow Language (YAWL)

2.2.3 Execution Level

Most of literature at the execution level of abstraction focuses on business process domain while only few papers focus on performances. Literature within the execution level focus mainly on the construct of workflow and Workflow Management Systems (WfMSs). Workflow is a concept strictly related to business processes and in particular it can be considered as an abstraction of a business process that is made by a number of logical steps, (known as tasks) which may require human involvement (e.g. an approval or the development of custom text) or may be executed automatically.

WfMSs are key tools for organizations motivated to improve competitive advantage, customer service, productivity, and conformity with standards.

Typically, a WfMS is composed by:

- Workflow Language
- Workflow Engine or Workflow Orchestration Engines
- Workflow Modeling Tool

A workflow application uses one or more workflow languages. There are many studies in literature discussing workflow languages and we list the most important:

- XML Process Definition Language (XPDL)
- BPEL
- JPDL

2.2.4 Model driven approach

The most discussed subjects found within this topic of research are articles focusing on the following approaches:

- **Model Driven Engineering (MDE):** Software development methodology which aims to raise the level of abstraction in software specifications and to increase automation in software development.
- **Model Driven Development (MDD):** Software engineering approach consisting of the application of models and model technologies to raise the level of abstraction at which developers create and evolve software, with the goal of both simplifying and standardizing the various activities and tasks which comprise the software life cycle.
- **Model Driven Architecture (MDA):** Software design approach proposed by Object Management Group (OMG) in 2001, that uses models in software development. The main idea is to define system functionalities using a Platform-Independent Model (PIM) with an appropriate Domain-Specific Language (DSL) based on a specific meta-model.
- **Model to Model (M2M) transformations:** Transformations which play a key role as they represent “the process of converting one model to another model of the same system” Miller et al. [Miller et al, 2003] and enable, as stated by Kleppe et al. [Kleppe et al, 2003] “the automatic generation of a target model from a source model, according to a transformation definition”.

Using a model driven approach will fasten integration and transformation between layers. Literature presents some interesting approaches in defining integrated meta-models of business processes and performances and in defining transformation between abstraction levels.

3. Conclusions

3.1 Conceptual Level

In literature there are only few papers that address the integration business processes and business performances at conceptual level. An interesting study is proposed by Buckl et al. [Buckl et al, 2009] which defines a framework for linking the Enterprise Architecture (EA) with its performances, however the study is limited to the availability of IT systems.

An important stream of research is given by the Goal-Process integration. Soffer and Wand [Soffer and Wand, 2005] proposes a formal approach, even if limited to the SCM industry, which applies the concept of soft-goal to processes but its approach lacks in formalize the measurement and neglects the stakeholder dimension. In general, in the field of RMS, there are several gaps that must be filled to gain a really universal framework for designing performing processes, for example:

- enhanced Telecom Operation Model (eTOM) defines a taxonomy for the Telecommunication industry and neglects the dimension of the performances.
- MIT-PH Process Handbook, defines an universal taxonomy but neglects the dimension of the performances.
- SCOR integrates specific KPIs but it neglects the dimension of the stakeholders.

- Information Technology Infrastructure Library (ITIL) integrates specific KPIs but it considers only the customer perspective.

We can summarize considerations on RMs in the following table.

Framework	Financial Perform.	Quality Perform.	Service Perform.	Manager View	Customer View	Worker View	Business Process
HIGO	X	X	X	X	X	X	
BSC	X	X	X	X			
Six Sigma		X		X			
SCOR	X	X					X
MIT-PH							X
eTOM							X
ITIL	X	X	X		X		X

Tab. 2 – Reference models mapping

3.2 Logical level

With regard to integration between performances and workflow we found very few studies. An interesting work is the UML Profile for Schedulability, Performance and Time (SPT) which defines a set of constructs (such as scheduling of tasks and triggers) that introduce the time dimension and the time-awareness into UML models.

Another interesting work is List and Korherr's [List and Korherr, 2006] one, which defines a meta-model that captures a wide range of process elements to evaluate BPMLs, and has been inspired by business process theory, workflow patterns and Workflow Management Coalition (WfMC). The meta-model is categorized according to a framework that includes four perspectives:

- Organizational perspective, that represents where and by whom the process elements are performed. According to WfMC those elements are organizational units, roles and (humans and automatic) resources;
- Functional perspective, that represents the process elements being performed, namely the Activities. They can be either Atomic Activities or Sub-Processes, that are recursively refined by activities;
- Behavioral perspective, that represents the flow of the activities; it describes when and how process elements perform (e.g., sequencing) through feedback loops, iteration, complex decision-making conditions, entry and exit criteria, and so on. The Data Flow connects atomic activities with information resources;
- Informational perspective, that represents the informational entities produced or manipulated by a process, that include data, artifacts, products and objects. The basic elements are resources and events. An Event may start an activity. A Resource is an entity to be produced or consumed by an atomic activity.

Korherr and List [List and Korherr, 2006] extend the framework with the Business Process Context Perspective which provides an overview perspective of the process and describes major business process characteristics:

- Process Type, that can be Core Process, a Support Process or a Management Process;
- Stakeholders, that introduce the Process Customers and Process Owner elements;
- Output or deliverables, that introduce the concept of Deliverables generated by a business process, which are either Services or Products.

In the following table (Table 3) we describe a comparison between BPMLs over description and analysis perspectives.

Language	Organizational Perspective	Functional Perspective	Behavioral Perspective	Informational Perspective	Context Perspective
BPMN	+	+	+	+/-	-/+
EPC	+/-	+/-	+	-/+	-/+
UML AD	+	-/+	+	-/+	-/+
IDEF 3	-	+/-	+	-	-/+
Petri Net	-	-/+	+	-	-/+

Legend: + Notation available; - Notation not available; +/- Several constructs are present; -/+ Few constructs are present.

Tab. 3: Comparison between BPMLs and different perspectives

As shown in the previous table, the functional and the behavioral perspectives are represented in all BPMLs, while organizational and informational perspectives are only partly supported. In particular the informational perspective is better developed in more recent BPMLs like Activity Diagrams, BPMN and EPC along with their support to execution languages.

A definition of business process [Ould and Ould, 1995] describes its purpose as threefold: describe, analyze and enact. Thus a BPML should:

- Describe a process in order to communicate it to other people, to define it or to share it across a group of people.
- Analyze a process for the change (e.g. the ordering of activities, responsibilities) and for the process improvement.
- Enact a process means to translate a schema into executable code.

The big advantage of that generic meta-model is that it can be integrated in every BPML where it is needed; however their work is limited to the graphical definition, as they do not specify how to enact those mechanisms for QoS.

3.3 Execution Level

The evaluation of performances during the execution of a workflow has been treated by few authors in literature mostly in reference to SOA and web services. An interesting approach is proposed by Abate et al. [Abate et al, 2002] that present a language that allows evaluating performances of workflows automated by a WfMS: Workflow Performances Query Language (WPQL). WPQL enables to define measures and to query the workflow engine log files generated by the engine during the system operation in order to compute the measures needed about given workflow entities. Abate's approach presents

several problems: it has not been further developed and the last version of WPQL is more a proof of concept than a real tool.

4. References

- A. Vogel, B. Kerhervé, G. Bochmann, J. Gecsei. Quality of Service Management: a survey. *IEEE Journal of Multimedia Systems*, vol 2 no 2, 1995.
- J. Evans and C. Filsfils. Deploying IP and MPLS QoS for multiservice networks: theory and practice. Morgan Kaufmann Pub, 2007.
- G. Pignatelli. Performance Aware Business Process Design. An integrated approach. University of Pavia, graduation thesis. 2010.
- J. Biolchini, P.G. Mian, A.C.C. Natali, and G.H. Travassos. Systematic review in software engineering. System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES, 679(05), 2005.
- B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering. School of Computer Science and Mathematics, Keele University, and Department of Computer Science, University of Durham, 2007.
- B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman. Systematic literature reviews in software engineering-A systematic literature review. *Information and Software Technology*, 51(1):7–15, 2009.
- P. Fettke, P. Loos, and J. Zwicker. Business process reference models: Survey and classification. In *Business Process Management Workshops*. Springer, 2006.
- J. Becker, M. Kugeler, and M. Rosemann. Process management: a guide for the design of business processes. Springer Verlag, 2003.
- S. Buckl, U. Franke, O. Holschke, F. Matthes, C.M. Schweda, T. Sommestad, and J. Ullberg. A pattern-based approach to quantitative enterprise architecture analysis. In *15th Americas Conference on Information Systems (AMCIS)*, 2009.
- P. Soffer and Y. Wand. On the notion of soft-goals in business process modeling. *Business Process Management Journal*, 11(6):663–679, 2005.
- B. List and B. Korherr. An evaluation of conceptual business process modelling languages. In *Proceedings of the 2006 ACM symposium on Applied computing*, page 1539. ACM, 2006.
- M.A. Ould and MA Ould. Business Processes: Modelling and analysis for re-engineering and improvement. Wiley Chichester, 1995.
- A.F. Abate, A. Esposito, N. Grieco, and G. Nota. Workflow performance evaluation through wpql. In *Proceedings of the 14th international conference on Software engineering and knowledge engineering*, pages 489–495. ACM, 2002.
- J. Miller, J. Mukerji, et al. MDA Guide Version 1.0. 1. Object Management Group, 234, 2003.
- A.G. Kleppe, J. Warmer, and W. Bast. MDA explained: the model driven architecture: practice and promise. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 2003.

Real-Time and Energy Issues in Mobile Health Monitoring Systems

Giorgio Buttazzo, Mauro Marinoni, Gianluca Franchino
Scuola Superiore Sant'Anna
Via Moruzzi 1, 56124 Pisa
{g.buttazzo, m.marinoni, g.franchino}@sssup.it

Health care is one of the areas in our society in which embedded computing systems can have a tremendous impact in reducing costs, improve efficiency, speed up the diagnosis of diseases, and detect critical health conditions. This paper presents an advanced mobile monitoring systems for patients affected by heart failure, focusing on the software methodologies adopted for handling real-time and energy constraints.

1 Introduction

Mobile health monitoring combines medical devices for health care with mobile computing technologies for improving the communication among patients, physicians, and other health care centers.

Recent technological advances in sensors, integrated circuits, and wireless communications have enabled the development of low-cost, miniature bio-medical devices that can be integrated into wireless body area networks (WBANs) for mobile health monitoring. Such networks are capable of sensing, processing, and communicating a set of vital signals and physiological data directly to a medical service center, where they can be stored into a personal file, which can be remotely accessed by the doctor whenever needed.

The advantages of such systems are significant, both from the patient and the society. From the patient point of view, this technology enables the delivery of accurate medical information from anywhere at any time, providing a snapshot of the personal health status to the doctor, without requiring hospitalization. Hence, the patient can continue to carry out his activities in his home, saving a lot of time and also gaining a psychological benefits. Moreover, the acquired data are stored in a personal file and collected to build up a complete and detailed history of the patient, useful for anamnesis, diagnosis, and prognosis. Finally, this technology helps people participating more closely in their own health care, especially for those that are reluctant to go to the hospital or to their doctor for routine medical checkups.

From a social perspective, mobile health care allows saving an enormous amount of money, by avoiding the costs of pre-hospitalization of those patients that need to be monitored but are not critical. Avoiding the pre-hospitalization of

non critical patients has also the benefit of a more efficient usage of the resources, avoiding crowding and long queues in health centers and hospital departments.

A portable health monitoring systems must satisfy several requirements in order to be really useful and accepted by the patient. First of all, it must be small, lightweight, and it must not consume much energy, having an autonomy of at least 24 hours. Such features impose tight constraints on the system architecture, especially on the processing unit, the memory, and the wireless communication device. Given the strong limitations on the processing resources available onboard, the operating system must adopt suitable algorithms for efficient as well as predictable resource management.

Moreover, the system has to carry out several concurrent activities in real-time, most of which require to be periodically executed (e.g., sensory acquisition, filtering, digital signal processing, data integration, storage, and communication). As a consequence, the kernel must handle periodic tasks with timing constraints and must guarantee bounded response times for all time critical processes.

Finally, these systems are typically required to monitor the patient continuously from one up to 3 days, hence battery lifetime is a crucial constraint to be guaranteed. Unfortunately, real-time and energy constraints are contrasting objectives in terms of resources allocation. In fact, to achieve short response times the system should operate at high speeds and have all devices active to reduce access delays, but this generates high power consumption; whereas, in order to last long, the system should to operate at low speed and turn off devices whenever not used, but this generates long latencies. To reduce power consumption while guaranteeing real-time constraints, special energy-aware policies have been used for tasks and message scheduling.

In this paper, we present an advanced mobile monitoring systems for patients affected by heart failure, focusing on the software methodologies adopted for handling real-time and energy constraints.

1.1 Related work

In the last ten years, health care triggered several research projects related to mobile monitoring systems.

At Harvard University, researchers have developed a wireless monitoring systems, called Code-Blue [Mal04], for medical applications, including pre-hospital and in-hospital emergency care. The sensors include portable 2-lead ECG, pulse oximeter, wearable Pluto mote with built-in accelerometer, and a module with accelerometer, gyroscope, and electromyogram sensor for stroke patient monitoring.

At the University of Aarhus, a research team developed and deployed a monitoring infrastructure for hospitals, including a location tracking system and context-aware applications running on interactive displays and mobile phones [Bar06].

MobiHealth [Mobi] aimed at the development of a mobile monitoring system for healthcare, integrating sensors in a body area network (BAN) customized to

the individual patient needs. Physical measurements, such as blood pressure or ECG are measured and transmitted wirelessly from the BAN to the hospital or health call center.

HEARTFAID [Heart] focused on defining an efficient health care delivery organization and management models for improving the processes of diagnosis, prognosis and therapy provision in the heart failure domain. The core of the platform is a clinical decision support system, designed by integrating knowledge representation techniques, hybrid reasoning methods, and advanced tools for the analysis of diagnostic data. In this context, a research team of the project [Col07, Col08, Chi10] investigated the use of semantic web technologies for implementing a real clinical scenario, covering the clinical course of a heart failure patient.

The ASCOLTA project [Asco] is focused on the development of an advanced monitoring systems for heart failure, integrating several physiological parameters like ECG, breath signal, oxygen saturation, blood pressure, inertial data, and patient weight, for early detection of critical conditions.

Stankovic et al. [Sta05] discuss the benefits and the crucial requirements for a pervasive monitoring infrastructure dedicated to assisted living, identifying the main research issues in such a new application domain.

Virone et al. [Vir06, Vir07] proposed an information system for assisted living based on a residential wireless sensor network that detects the presence of patients in specific locations and estimates a set of events of interest.

Wu et al. [Wu07] present the design and implementation of a wireless mobile monitoring system based on Bluetooth technology. Multiple physiological parameters can be monitored in real-time and the users can connect to the medical service center through digital family equipments, such as digital televisions and cell phones.

Wood et al. [Woo08] present a context-aware wireless sensor networks for assisted-living and residential monitoring, integrating environmental, physiological, and activity sensors in a scalable heterogeneous architecture.

Trappey et al. [Tra09] developed a mobile intelligent medical system to support mobile nursing applications and clinical decision support. The system includes an RFID-based module for patient data collection and a Java-based expert system for issuing warnings and diagnostic messages.

Ren et al. [Ren10] present several techniques that can be used to monitor patients effectively and enhance the functionality of telemedicine systems, and discuss how secure strategies can be used to improve the security of mobile healthcare.

Santra [San10] discussed secure strategies that can impede the attacks faced by wireless communications in mobile healthcare systems.

Shahriyar et al. [Sha09] present a mobile health monitoring system that uses a wearable wireless body area network for collecting data from patients and predicts patient's health status. The patients can access their health information by their mobile devices.

2 System architecture

The system considered in this paper is shown in Figure 1 and consists of an embedded platform worn by the patient to monitor his physiological data through a set of sensors. In its full configuration, the platform can include the following components:

- a microcontroller unit (MCU) in charge of acquiring, processing, communicating, and storing data from a set of biomedical sensors;
- a radio transceiver used to connect the system with a remote server or other devices, such as PCs or portable devices. Different wireless standard protocols are supported, as ZigBee, Bluetooth, and WiFi;
- one or more storage components, such as flash memories, SD cards, or USB storage devices;
- a set of biomedical sensors for monitoring the electrocardiographic signal (ECG), oxygen saturation (SPO₂), heart and breath rate;
- a 3-axis accelerometer to detect possible dangerous situations (e.g., falls) and reconstruct the patient activities for a correlation with the ECG trace.
- A user interface used to signal the system status and/or to control the device.

The system has been designed to be modular, in the sense that, depending on the specific application, one or more components can be excluded from the platform to minimize the weight and the energy consumption.. For instance, when the system is used to monitor people affected by heart failure, the platform includes ECG, SPO₂, and breath rate sensors, a SD card to store the acquired data, and a WiFi connection to send the patient status to a remote health-care center. The user interface, kept as simple as possible, includes a switch to turn the device on and off, and a multicolor led and a buzzer to verify the proper operational modes. The system is powered by a rechargeable lithium polymer battery.

3 Real-Time issues

The software of a complex embedded device, such as that described in this work, must be able to handle several concurrent activities with different execution rates. For instance, the ECG signal is typically acquired with a sampling period of 2-4 milliseconds, while the breath rate sensor can be sampled with a period of 50 milliseconds. Other activities with different execution times and activation rates are related to data communication and data storage.

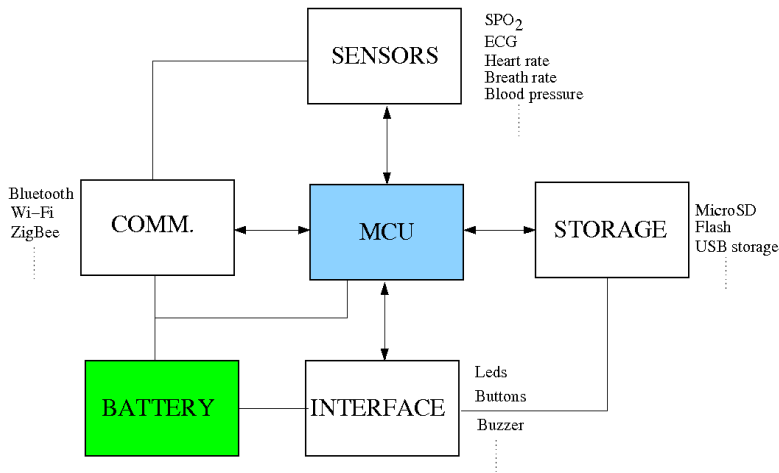


Fig.1 – System architecture

The data storage function is necessary to preserve patient data when the access point is out of reach(e.g. when the patient is not at home, or the communication is temporarily unavailable). The communication between the device and the remote health-care center is handled by a TCP/IP connection. To ensure a timely execution of such different activities and perform an off-line guarantee of the timing properties of the application, the software has to be supported by a real-time operating system. The software developed for the platform presented in this paper has been developed under the ERIKA Enterprise real-time kernel [Erika], which allows achieving high predictable timing behavior with a very small runtime overhead and memory footprint. The kernel is briefly described in the following section.

3.1.1 ERIKA Enterprise

ERIKA Enterprise is an open-source (GPL2 with linking exception) real-time operating system, implementing an Application Programming Interface compatible with the OSEK/VDX standard for automotive embedded controllers. ERIKA includes highly predictable real-time kernel mechanisms and uses innovative programming features to support time sensitive applications on a wide range of microcontrollers and multi-core platforms. In particular, ERIKA supports periodic and aperiodic task scheduling according to fixed and dynamic priorities; interrupt handling for urgent peripherals operation (interrupts always preempt task execution); and time bounded resource sharing through the Immediate Priority Ceiling protocol [Sha90] [But11].

4 Energy management

4.1 General issues

Power management with energy efficiency considerations is not only useful for mobile devices for prolonging their battery duration, but it is also helpful for server systems for the reduction of power bills. Dynamic power due to switching activities and static power due to the leakage current are two major sources of energy consumption in CMOS circuits.

For micrometer-scale semiconductor technology, the dynamic power dominates the power consumption of a processor. However, for technology in the deep sub-micron (DSM) domain, the leakage power consumption is comparable to or even more than the dynamic power dissipation. Reducing the voltage decreases the dynamic power consumption, however it also reduces the maximum operating frequency.

The overall energy consumption of a computing system also depends on other components. Martin et al. [Mar01] derived the following model to describe the power consumption as a function of the processor speed, defined as a normalized frequency ($s = f/f_{max}$):

$$P(s) = K_3 s^3 + K_2 s^2 + K_1 s + K_0$$

When the dynamic power consumption dominates the other components (due to switching and leakage), the K_3 coefficient is greater than the others and, power consumption can be approximated as $P(s) = K s^3$.

Two major classes of power-aware algorithms can be distinguished based on the kind of power they try to reduce. Dynamic Voltage Scaling (DVS) techniques reduce dynamic power by decreasing the supply voltage (and consequently the clock frequency) of the system, whereas Dynamic Peripheral Management (DPM) approaches achieve energy saving by exploiting operational states with reduced energy consumption (e.g., sleep or idle) whenever possible.

To reduce the runtime overhead due to energy management, some algorithms are applied off-line to compute the best energy setup as a function of the task set parameters (e.g., worst-case execution times and periods). The disadvantage of off-line approaches, however, is that they are based on worst-case assumptions and cannot take advantage of situations occurring at run time (such as early completions). For this reason, other methods are applied online to re-compute the parameters after each job termination.

When using DVS techniques in the presence of resource constraints (e.g., under mutually exclusive resources or non preemptive regions), the system can experience scheduling anomalies in which a task could even increase its response time when executed at a higher speed [But06]. Such problems prevent managing the performance of a real-time application as a function of the processor speed.

4.2 Proposed approach

In embedded systems characterized by real-time and energy constraints, selecting the most appropriate energy management policy is not easy, because the result heavily depends on the platform characteristics (e.g., energy modes and profiles of the devices, frequency range and power states of the CPU) and the application constraints (e.g., task deadlines, sensors acquisition delays, communication bandwidth, etc).

A preliminary study carried out on the presented platform showed that to achieve the best results, both DVS and DPM techniques have to be integrated to fit the requirements with the minimum energy cost. The combination of DVFS and DPM is done by forcing a CPU sleep interval followed by an active interval executed at a fixed frequency, selected to minimize the energy (per unit of computation) between the current and the next invocation of the analysis [Mar11]. The approach has been implemented in the Erika as a policy in the Power Manager of the kernel, which is described in the next section.

5 Kernel support

The energy saving module proposed in this section is part of the kernel and interacts with the scheduler, the hardware devices, and the application. While the scheduler selects the next task to execute, the Power Manager chooses an appropriate running configuration (i.e., speed and voltage). It uses a modular approach that allows the user to select a policy customized for a specific device providing a uniform interface.

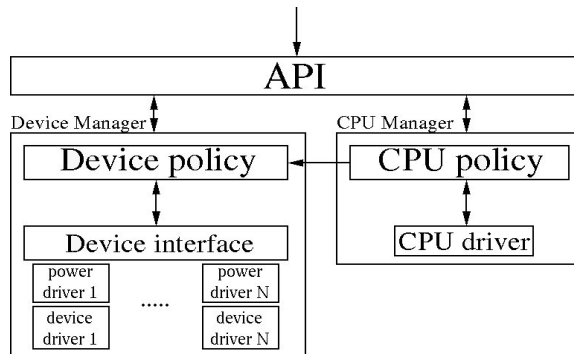


Fig.2 – Architecture of the power management module

The Power Manager consists of three hierarchically organized modules: the Application Programming Interface (API), the CPU Manager and the Devices Manager. Such a modular implementation allows the programmer to easily remove sub-components when not needed by the application, so helping to reduce the footprint.

The API module implements the interface defined for the interaction with the kernel and the applications. The CPU Manager is responsible for the power management of the CPU. Using a set of special callback functions the kernel informs the module about scheduling events.

The CPU policy sub-module implements the energy saving policies, which typically select the best speed to meet the applications constraints, while satisfying a given set of performance requirements. The CPU driver is in charge of setting the CPU parameters.

The Devices Manager handles internal and external peripherals. Inside it, the Device policy sub-module contains all the device policies. For each of them, two stacked components, Power driver and Device driver, abstract the device behavior using a discrete set of states.

6 Performance results

This section presents some experimental results that show that integrating DVFS and DPM approaches leads to lower power consumption in different types of platforms. The experiment compares a pure DPM approach, a pure DVFS approach, and the mixed one proposed by Marinoni et al. [Mar11] on a system that has to remain active during communication slots.

The comparison is done on two different types of architectures: in the first one (denoted as Fully-DPM model) the leakage power consumption is dominant; whereas in the second one (denoted as Fully-DVFS model) the main source of energy consumption is due to dynamic power.

In the experiment, the task set utilization is varied between 0.3 and 0.9 to evaluate the effect of the workload on different scenarios. The results are reported in Figure 3 and show that the mixed approach always produces the best performance in terms of energy reduction and, in the worst case, behaves as good as the other approaches.

Notice that, even in the most favorable condition (left graph), the fully-DPM approach presents poor performance working with energy constraints that prevent from switching the system off in some particular intervals.

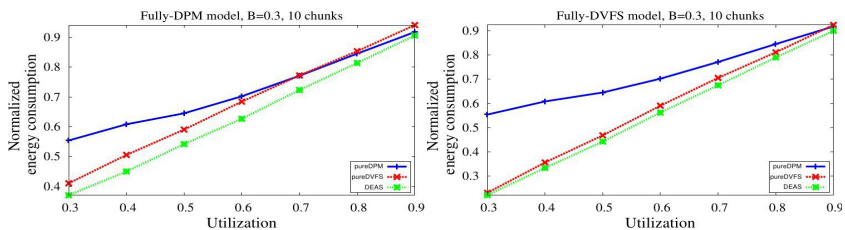


Fig.3 – Experimental results on power management techniques

7 Conclusions

This paper presented an integrated approach for addressing energy consumption and real-time constraints in a small embedded monitoring system for medical applications. Experiments results demonstrate the effectiveness of the method for different types of architectures.

8 Acknowledgments

This work has been carried out within the ASCOLTA project, funded by Tuscany Region under the grant 2009DUA-864500467-1.

9 References

- [Asco] "ASCOLTA: Assistenza domiciliare allo SCOMpenso cardiaco attraverso Le Tecniche Avanzate di comunicazione digitale", Research Project by Tuscany Region, URL: <http://www.ascolta.org/>.
- [Bar06] J. E. Bardram et al., "Experiences from Real-World Deployment of Context-Aware Technologies in A Hospital Environment," Proceedings of Ubicomp 2006, LNCS 4206, pp. 369–386, 2006.
- [But06] G. Buttazzo, "Achieving Scalability in Real-Time Systems", IEEE Computer, 39(5):54-59, May 2006.
- [But11] G. Buttazzo, "HARD REAL-TIME COMPUTING SYSTEMS: Predictable Scheduling Algorithms and Applications", Third Edition, Springer, 2011.
- [Chi10] F. Chiarugi, S. Colantonio, D. Emmanoulidou, M. Martinelli, D. Moroni, O. Salvetti, "Decision Support in Heart Failure Through Processing of Electro and Echocardiograms", Artificial Intelligence in Medicine, Elsevier, Vol. 50, No 2, October 2010.
- [Col07] S. Colantonio, M. Martinelli, D. Moroni, O. Salvetti, F. Perticone, A. Sciacqua, D. Conforti, A. Gualtieri, "An Approach to Decision Support in Heart Failure", 4th Italian Workshop on Semantic Web Applications and Perspectives (SWAP 2007), Bari, Italy, 18-20 December 2007.
- [Col08] S. Colantonio, M. Martinelli, D. Moroni, O. Salvetti, "Decision Support and Image & Signal Analysis in Heart Failure - A Comprehensive Use Case", International Conference on Health Informatics (HEALTHINF 2008), Funchal, Madeira, Portugal, 28-31 January 2008.
- [Erika] The Erika Enterprise real-time kernel: <http://erika.tuxfamily.org>
- [Heart] "HEARTFAID: A knowledge based platform of services for supporting medical-clinical management of heart failure within elderly population", European project IST-2005-027107, URL: <http://lis.irb.hr/heartfaid/>.
- [Mal04] D. Malan et al., "Codeblue: An Ad Hoc Sensor Network Infrastructure for Emergency Medical Care", Proc. of the International Workshop on Wearable and Implantable Body Sensor Networks, 2004.
- [Mar01] T. Martin and D. Siewiorek. "Non-ideal battery and main memory effects on CPU speed-setting for low power". IEEE Transactions on VLSI Systems, 9(1):29-34, 2001.

- [Mar11] M. Marinoni, M. Bambagini, F. Prosperi, F. Esposito, G. Franchino, L. Santinelli and G. Buttazzo, "Platform-aware Bandwidth-oriented Energy Management Algorithm for Real-Time Embedded Systems", Proc. of the 16th IEEE on Emerging Technologies and Factory Automation (ETFA 2011), Toulouse, France, September 2011..
- [Mobi] "MOBIHEALTH: Mobile Health Care", European project, IST-2001-36006, URL: <http://www.mobihealth.org/>.
- [Ren10] Y. Ren, R. W. N. Pazzi, A. Boukerche, "Monitoring patients via a secure and mobile healthcare system", IEEE Wireless Communications, Vol. 17, No. 1, February 2010.
- [San10] T. Santra, "Mobile Health Care System for Patient Monitoring", in Communications in Computer and Information Science, Springer, Vol. 101, pp. 695-700, 2010
- [Sha90] L. Sha, R. Rajkumar, and J. P. Lehoczky, "Priority Inheritance Protocols: An Approach to Real-Time Synchronization", IEEE Transactions on Computers, Vol. 39, No. 9, September 1990.
- [Sha09] R. Shahriyar, M. F. Bari, G. Kundu³, S. I. Ahamed, and M. M. Akbar, "Intelligent Mobile Health Monitoring System (IMHMS)", International Journal of Control and Automation, Vol.2, No.3, September 2009.
- [Sta05] J. A. Stankovic, I. Lee, A. Mok, R. Rajkumar, "Opportunities and Obligations for Physical Computing Systems," IEEE Computer, 38(11), November 2005.
- [Tra09] C. V. Trappey, A. J. C. Trappey, C. S. Liu, "Develop Patient Monitoring and Support System Using Mobile Communication and Intelligent Reasoning", Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, USA - October 2009.
- [Vir02] G. Virone, N. Noury and J. Demongeot, "A system for automatic measurement of circadian activity deviation in telemedicine," IEEE Transactions on Biomedical Engineering, vol. 49, no. 12, pp. 1463-1469, Dec. 2002.
- [Vir06] G. Virone, A. Wood, L. Selavo, Q. Cao, L. Fang, T. Doan, Z. He, and J.A. Stankovic, "An Assisted Living Oriented Information System Based on a Residential Wireless Sensor Network," Proc. of the Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare (D2H2), Arlington, VA, April 2-4, 2006.
- [Vir07] G. Virone, T. Doan, A. Wood, and J. A. Stankovic, "Dynamic Privacy in Assisted Living and Home Health Care," in *Joint Workshop On High Confidence Medical Devices, Software, and Systems (HCMDSS) and Medical Device Plug-and-Play (MD PnP) Interoperability*, 2007.
- [Woo08] A. Wood, J. A. Stankovic, G. Virone, L. Selavo, Z. He, Q. Cao, T. Doan, Y. Wu, L. Fang, R. Stoleru, "Context-Aware Wireless Sensor Networks for Assisted-Living and Residential Monitoring," IEEE Network, 22(4):26–33, July-August 2008.
- [Wu07] K. Wu, X. Wu, "A Wireless Mobile Monitoring System for Home Healthcare and Community Medical Services", The 1st International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2007), Wuhan, China, 6-8 July 2007.

Rischi e opportunità di natura giuridica connessi alla creazione di una piattaforma di condivisione dati per i knowledge workers

Autori: Giovanni Pellerino, professore aggregato di Informatica giuridica presso la Facoltà di Giurisprudenza dell'Università del Salento

Carlo Bucciero, borsista di ricerca presso il Dipartimento di Ingegneria dell'Innovazione dell'Università del Salento

Abstract

L'idea software "K – people" si fonda sulla creazione di una piattaforma di condivisione di dati utilizzabile da parte di tutti i knowledge workers, i quali, in tal modo, diventano i veri protagonisti dei processi aziendali.

Al fine di prevedere e gestire al meglio le implicazioni giuridiche connesse a tale radicale cambiamento metodologico, si rivela fondamentale implementare preventivamente una mappatura dei potenziali rischi giuridici, valutando, altresì, l'esistenza di eventuali opportunità ad essi connesse.

Tale attività di legal risk management consente all'azienda di coniugare innovazione tecnologica web 2.0 e sicurezza in materia legale, tutelando il vantaggio competitivo acquisito.

Sommario

1. Introduzione. – 2. Destinatari. – 3. La Business Intelligence e i rischi di natura giuridica nella conservazione delle informazioni e nella gestione dei flussi informativi – 4. Il Legal Risk Management (LRM): un nuovo approccio metodologico al concetto di rischio giuridico. – 5. Il modello organizzativo ex D. Lgs. 231/2001: uno strumento a tutela dell'azienda. – 6. Legal Risk Management e ICT: gestire i rischi giuridici e cogliere le opportunità dell'innovazione tecnologica. – 7. Il Legal Risk Management come fattore critico di successo: il caso K-people.

1. Introduzione

Scopo del presente lavoro è quello di riflettere sui rischi di natura giuridica e sulle opportunità che l'introduzione di un sistema software di Knowledge management può comportare all'interno di un'azienda¹. Quando parliamo di rischi di natura giuridica, facciamo riferimento ad una serie di conseguenze negative per l'impresa che derivano dal mancato rispetto, spesso inconsapevole, di norme di legge, di direttive, di regolamenti. Tali conseguenze negative sono, sempre e comunque, frutto di decisioni e potrebbero essere evitate se l'azienda fosse messa nelle condizioni di assumere la decisione corretta. In ambito giuridico, assumere la decisione corretta significa effettuare una scelta conforme al diritto. Ma per effettuare una scelta, qualsiasi scelta, occorre avere di fronte delle alternative. Il problema che si pone, dunque, è quello di porre il decisore, all'interno dell'azienda, nelle condizioni di poter scegliere, di essere in grado di osservare le alternative, prima di assumere una decisione rischiosa, potenzialmente contraria al diritto. Il decisore, pertanto, nel caso di studio, deve poter conoscere quali problematiche di natura giuridica possono scaturire dall'utilizzo del sistema di Knowledge management; individuati gli ambiti di criticità, deve essere informato in ordine alla normativa di riferimento; deve essere in grado di evitare di determinare scelte contrarie al diritto che comporterebbero responsabilità personali ed aziendali. E' ovvio

¹ L'indagine ha preso le mosse dall'analisi empirica di un sistema di Knowledge management denominato "K-people".

che i rischi giuridici, come qualsiasi altra categoria di rischio, non possono mai essere del tutto eliminati, per cui ci si trova sempre di fronte ad un rischio residuo che va gestito attraverso delle opportune contromisure.

2. Destinatari

Questo contributo di ricerca pone all'attenzione della comunità scientifica i benefici connessi ad un approccio orientato al legal risk management. Al contempo, la "trasversalità" del metodo, suscettibile di applicazione pratica ad ampio spettro, rende i risultati del lavoro concretamente apprezzabili in campo industriale.

3. La Business Intelligence e i rischi di natura giuridica nella conservazione delle informazioni e nella gestione dei flussi informativi

Come è noto, quando si parla di Business Intelligence ci si riferisce ad un ampio ventaglio di tecnologie, piattaforme software, applicazioni specifiche e processi determinati. Ciò che caratterizza, in linea generale, queste tecnologie è l'obiettivo di assumere più velocemente decisioni migliori, attraverso la conversione di dati in informazioni, utilizzando un approccio razionale per la gestione. Per assumere decisioni migliori più velocemente i *decision makers* hanno necessità di disporre di una enorme mole di informazioni che risultino utili ai propri scopi e, soprattutto, facilmente accessibili. Ma c'è spesso un grosso gap tra le informazioni che i *decision makers* richiedono e le montagne di dati che le aziende raccolgono ogni giorno. Tale fenomeno si definisce *gap di analisi*.

Per colmare questo *gap di analisi*, le organizzazioni realizzano importanti investimenti nello sviluppo dei sistemi BI per convertire *dati grezzi* in *informazioni utili*. I più efficaci sistemi BI accedono ad enormi volumi di dati e forniscono istantaneamente i sottoinsiemi pertinenti ai *decision makers* in una forma in cui questi soggetti posano facilmente metterli in relazione. Indipendentemente dal modello di processo che le piattaforme di BI impiegano per catturare i dati grezzi e convertirli in informazioni utili, non possono fare a meno di utilizzare database che, generalmente, prendono il nome di *data warehouse* (magazzino dati). Tali database, solitamente, raccolgono i dati che provengono dai diversi sistemi gestionali che supportano le attività dell'azienda. Una volta caricati i dati nel *data warehouse*, i processi di produzione di report e analisi business sono responsabili della loro estrazione e dell'assemblaggio in un formato agevole per le applicazioni business (per esempio fogli di calcolo, strumenti di diagrammazione e di grafica business) e di fornire queste informazioni agli utenti. Per implementare questo processo, vi è una categoria di software, chiamata in modo generico *strumenti front-end*, che raccoglie le informazioni dal *data warehouse* e le presenta agli utenti sottoforma di report e di viste interattive. Inoltre, i dati possono essere esaminati utilizzando una tecnologia di analisi più avanzata come il *data mining* per individuare modelli nei dati che spieghino il comportamento o che scoprano le tendenze difficili da vedere ad occhio nudo. Mentre le specificità di questi processi dipenderà dalle necessità degli utenti, il loro obiettivo principale è mettere dati utili nelle mani dei *decision maker* per aiutarli a prendere più velocemente decisioni migliori. I sistemi di generazione più recente si concentrano sull'ottimizzazione dei flussi comunicativi e sull'efficienza nel recupero delle informazioni con conseguente incremento nell'efficacia delle attività svolte.

In ogni caso, a prescindere dalla struttura del sistema, gli strumenti offerti dalla BI e le possibilità di analisi che sono ad essa associate favoriscono senz'altro un incremento di capacità di assumere decisioni corrette all'interno delle organizzazioni. Tuttavia, per l'impresa, assumere la decisione apparentemente più corretta sotto il profilo economico e finanziario, attraverso l'utilizzo di sistemi di BI, potrebbe significare incorrere in rischi di altra natura quali, ad esempio, quelli giuridici, se non opportunamente studiati e valutati. Tali rischi, non facilmente eliminabili, in quanto connaturati alla struttura stessa del sistema, possono essere gestiti adeguatamente adottando idonee strategie di contenimento e specifiche politiche di sicurezza (*Security policy*).

4. Il Legal Risk Management (LRM): un nuovo approccio metodologico al concetto di rischio giuridico

L'approccio metodologico alle problematiche di natura giuridica connesse a qualsiasi processo o prodotto è spesso caratterizzato dall'assenza di un vero e proprio studio analitico preventivo e prognostico, consistendo sostanzialmente in una attività, successiva al verificarsi dell'evento negativo, finalizzata unicamente ad evitare il peggioramento delle conseguenze pregiudizievoli ad esso connesse. La decennale esperienza in tema di responsabilità amministrativa degli enti, ex D. Lgs 231/2001, è esemplare in tal senso. L'adeguamento a questa disciplina è stato, infatti, assolutamente insufficiente da parte di numerosissime aziende, determinandone la responsabilità per una vasta serie di reati commessi, o tentati, a loro vantaggio da soggetti apicali ovvero da dipendenti. Ciò dimostra che le problematiche di natura giuridica non possono essere affrontate unicamente in una fase successiva al loro accadimento, dovendo al contrario trovare una adeguata forma di gestione quando le medesime sono ancora sotto forma di rischi potenziali.

Il *legal risk management*² nasce proprio da questa diffusa necessità, costituendo un approccio metodologico innovativo alla gestione delle criticità di natura giuridica basato su una serie di attività tese ad evitare l'accadimento dei rischi ad esse connessi o, quantomeno, a minimizzarne gli impatti negativi. Gli elementi di innovazione³ vanno tipicamente ricercati nella capacità di gestire preventivamente ed in maniera integrata tali problematiche; in tal modo, è possibile avere una visione globale delle situazioni di rischio di natura giuridica connesse all'attività da svolgersi ed approntare un piano di gestione unitario ed integrato, con un importante risparmio in termini di tempo, di costo e di organizzazione. Questo approccio metodologico produce, infatti, molteplici ed evidenti benefici per l'azienda, creando nuove opportunità: agire preventivamente al fine di evitare il sorgere del contenzioso, o quantomeno al fine di ridurne al minimo le possibilità di accadimento, producendo un evidentissimo risparmio di

² Il Legal risk management è una disciplina quasi del tutto inesplorata nel panorama scientifico del diritto italiano. Ad oggi non esistono specifiche pubblicazioni italiane sull'argomento. A livello comunitario si segnalano articoli e pubblicazioni scientifiche a firma di J. Trzaskowski.

³ Cfr. J. Trzaskowski, 2005, *Legal Risk Management - Some Reflections*, Article, DJØF Publishing.

risorse, non solo economiche; efficientare i processi, rendendoli più sicuri, riducendone i tempi e scegliendo le risorse umane più idonee agli scopi da perseguire; migliorare la qualità dei prodotti, rendendoli più efficaci sul mercato. In sintesi, essere in grado di guardare al concetto di rischio come opportunità, ossia approcciarsi ad esso come evento che, attraverso una adeguata gestione, può essere potenzialmente produttivo di una situazione di vantaggio.

Ad una gestione integrata dei rischi giuridici, consegue la naturale evoluzione del ruolo tradizionale del giurista che, in tali condizioni, non può più essere considerato semplicemente il protagonista di una eventuale attività giudiziale o di consulenza potendo, e dovendo, fornire un contributo fattivo strategico determinante in ambito di risk management. Sicché, deve essere in grado di coniugare agli aspetti tecnici della professione legale adeguate competenze manageriali e strategiche: analizzare ed interpretare i dati giuridici, adattarli al contesto di riferimento, identificare i rischi giuridici e, conseguentemente, implementare un'attività di analisi e di gestione dei medesimi, cogliendo, al contempo, eventuali opportunità connesse. Il giurista assume, così, le vesti del legal risk manager.

In sintesi, un approccio orientato al legal risk management si rivela premiante per l'azienda⁴, garantendole benefici economici concreti ed aumentandone la stabilità, soprattutto in un periodo di tempo medio - lungo. L'attività svolta in relazione all'introduzione del sistema K-people⁵ è la dimostrazione pratica di tale assunto nel settore ICT: un approccio integrato alla gestione dei rischi giuridici ha consentito all'azienda di innovare in sicurezza, di tutelare il vantaggio competitivo acquisito, di accrescere il proprio know how, nonché di cogliere e tutelare i vantaggi connessi alle nuove opportunità di business⁶.

5. Il modello di organizzazione, gestione e controllo ex D. Lgs. 231/2001: uno strumento a tutela dell'azienda

Un'attività orientata ai principi del legal risk management non può prescindere, nel nostro Paese, dallo studio e dall'analisi del D. Lgs. 231/2001, normativa che ha introdotto nel diritto italiano la responsabilità amministrativa a carico di società, enti e associazioni, anche prive di responsabilità giuridica, per i reati commessi da parte dei soggetti in posizione apicale, sottoposti, nonché nelle ipotesi in cui l'autore del reato non sia stato identificato o non sia imputabile. L'accertamento di tale responsabilità avviene tramite gli strumenti propri del processo penale, con sanzioni a carico dell'azienda che possono avere natura pecuniaria, interdittiva (anche in via cautelare), nonché comportare la confisca del

⁴ Cfr. J. Trzaskowski, 2005, *Legal Risk Management in Electronic Commerce - Managing the risk of cross-border law enforcement Book* (PhD thesis), Ex Tuto Publishing.

⁵ Il software K – people è una innovativa piattaforma di condivisione dati per i knowledge workers introdotta da Webscience, società operante da decenni nel settore ICT. Questo software ha costituito oggetto di studio per il progetto pilota dell'attività di ricerca in tema di legal risk management condotta dal dr. Carlo Bucciero, con il supporto scientifico del Centro Interdipartimentale di Studi sul Rischio e del Dipartimento di Ingegneria dell'Innovazione dell'Università del Salento, con i tutor accademici Prof. Giovanni Pellerino, docente di informatica giuridica, e Prof. Roberto Paiano, docente di sistemi informativi.

⁶ Cfr. J. Trzaskowski, 2006, *Legal Risk Management in a Global, Electronic Marketplace*, Article, *Scandinavian Studies in Law*, Volume 49 (A Proactive Approach).

profitto del reato e la pubblicazione della sentenza di condanna⁷. A fronte di sanzioni così significative sulla vita dell'impresa, la medesima normativa indica espressamente gli strumenti attraverso i quali è possibile contenere i rischi giuridici potenzialmente connessi alla sua entrata in vigore, sicché l'imprenditore è chiamato a formalizzare un sistema di regole e controlli sulle attività concretamente svolte e sui modelli di organizzazione esistenti al fine di prevenire i rischi connessi alla propria attività. Soprattutto alla luce della giurisprudenza più recente, sempre più orientata a punire con severità quegli enti che non si sono ancora uniformati a tale disciplina, l'adeguamento agli standard minimi di sicurezza previsti dalla normativa è diventata una condizione assolutamente necessaria ed imprescindibile per poter svolgere, in sicurezza, un'attività imprenditoriale.

Ebbene, la costruzione dei processi aziendali in modo conforme a quanto previsto dal D. Lgs. 231/2001 è una parte fondamentale ed imprescindibile di quell'approccio metodologico innovativo al concetto di rischio giuridico che prende il nome di legal risk management e si caratterizza per la sua applicabilità trasversale⁸.

Il settore dell'ICT, più di ogni altro caratterizzato da una fortissima e costante tendenza al cambiamento e alla sperimentazione, proprio in ragione di questa caratteristica peculiare, è particolarmente esposto a tale tipologia di rischio. A dimostrazione di tale assunto, un approccio orientato al legal risk management, condotto con riferimento al software K-people, ha evidenziato l'esistenza di numerose importanti criticità in materia di protezione dei dati personali, di tutela del diritto d'autore ed, al tempo stesso, ha messo in rilievo la necessità di approntare quegli standard minimi di sicurezza previsti e disciplinati dal D. Lgs. 231/2001. Detta legge prevede, in particolare, l'adozione del modello di organizzazione, gestione e controllo, sebbene in termini di facoltatività e non di obbligatorietà. Tuttavia, la mancata adozione di tale modello espone l'ente alla responsabilità per gli illeciti realizzati da amministratori e dipendenti, pertanto la sua creazione, di fatto, diviene necessaria per poter beneficiare dell'esimente. Ad ogni modo, prescindendo dall'aspetto strettamente giuridico - sanzionatorio, la formalizzazione di un sistema di regole volte a ribadire la *compliance* dell'impresa non solo rispetto a norme giuridiche, ma anche a valori di tipo etico, rappresenta per l'azienda una vera e propria opportunità: un modello organizzativo che rende trasparenti le procedure interne, oltre a garantire l'esenzione dalla responsabilità amministrativa, infatti, costituisce un'occasione di sviluppo per le aziende, migliorandone l'immagine pubblica e riducendo i costi di transazione derivanti da eventuali azioni legali. Anche nella consapevolezza di tali vantaggi, sempre più soggetti giuridici si sono dotati di codice etico e modello organizzativo.

⁷ Il D. Lgs. 231/2001 è una normativa di portata epocale per la vita giuridica delle aziende: introduce nell'ordinamento giuridico italiano la cultura dei controlli interni nelle aziende come strumento di prevenzione di una vasta serie di reati. La norma prevede sanzioni in capo alle aziende, come persona giuridica, responsabile per non avere impedito ai propri dipendenti di commettere reati nell'interesse della società.

⁸ Cfr. R. Nielsen, S. Sandfeld Jakobsenog, J. Trzaskowski, 2004, *Legal Risk Management in Cross-Border Electronic Commerce*, Article, EU Electronic Commerce Law, Djøf Publishing.

Tutte le considerazioni che precedono inducono a ritenere il modello di organizzazione gestione e controllo, ex D. Lgs. 231/2001, uno strumento concreto a tutela dell'azienda che integra perfettamente quell'approccio metodologico innovativo al rischio giuridico, denominato legal risk management⁹.

6. Legal Risk Management e ICT: gestire i rischi giuridici e cogliere le opportunità dell'innovazione tecnologica

La disciplina del legal risk management rappresenta un approccio metodologico innovativo alla gestione dei rischi giuridici d'impresa che, potenzialmente, è applicabile a qualunque settore di attività. Questa metodologia di analisi e di gestione delle criticità di natura giuridica connesse ad un business, ad un processo o ad un prodotto è, infatti, un'attività, di per se stessa, neutrale che ben si adatta a qualsiasi contesto operativo. Il progetto pilota di tale attività di management ha trovato applicazione pratica nella realtà mutevole e variegata dell'ICT, settore, per definizione, in costante divenire, che si è prestato ottimamente alle aspettative di ricerca¹⁰. Detta scelta progettuale non è stata casuale, essendo quello dell'Information and Communication Technology, il settore idoneo, più di ogni altro, ad essere destinatario di un'attività di gestione dei rischi giuridici: l'innovazione tecnologica, infatti, impatta fatalmente con il contesto giuridico di riferimento che, tra l'altro, spesso non muove parallelamente all'innovazione medesima. In tal modo, l'azienda che fa dell'innovazione tecnologica il suo *core business* si trova, a dover essere in grado di gestire adeguatamente, al contempo, rischi ed opportunità. In relazione ai primi, un approccio di LRM ne consente una eliminazione o, quantomeno, un controllo del loro accadimento, minimizzandone significativamente gli impatti negativi; con riferimento alle opportunità eventualmente connesse a processi o a prodotti innovativi, al contrario, una corretta ed adeguata gestione ne massimizza il profitto. Per questi motivi, il legal risk management appare uno strumento particolarmente adeguato al soddisfacimento delle esigenze delle aziende operanti nel mercato dell'information and communication technology. A conferma di tale assunto, si osservino i risultati estremamente positivi che il connubio tra diritto e innovazione tecnologica ha prodotto nel caso del progetto K-people, consentendo all'azienda committente di produrre innovazione in piena sicurezza e mantenendo inalterato il vantaggio competitivo¹¹ connesso

⁹ Confindustria ha istituito l'Osservatorio di giurisprudenza in materia di responsabilità amministrativa da reato degli enti, al fine di "mettere a disposizione delle imprese una banca dati delle principali pronunce massimate in materia di decreto legislativo n. 231/2001, organizzate per tematiche rilevanti". (www.confindustria.it).

¹⁰ Cfr. J. Trzaskowski, 2003, *Cross-Border Law Enforcement in the Information Society* (v. 0.82) Article, DJØF Publishing.

¹¹ Vantaggio competitivo è un termine utilizzato nel marketing e divenuto centrale nella gestione strategica (o *management strategico*), anche a seguito dei contributi di Michael Porter, accademico ed economista statunitense. Per vantaggio competitivo si intende ciò che costituisce la base delle performance superiori registrate dall'impresa, solitamente in termini di profittabilità, rispetto alla media dei suoi concorrenti diretti nel settore di riferimento, in un arco temporale di medio - lungo termine. Per Enrico Valdani, è "la capacità distintiva" (o competenza distintiva) "di un'impresa di presidiare, sviluppare e difendere nel tempo, con maggiore intensità dei rivali, una capacità *market driving* o una risorsa critica che possono divenire fattori critici di successo".

e conseguente all'innovazione medesima. Le specifiche di progetto relative alla piattaforma software K-people sono state implementate sulla scorta delle criticità identificate nel corso della preventiva attività di legal risk management; in tal modo, gli impatti negativi conseguenti all'eventuale accadimento dei rischi giuridici connessi al prodotto software medesimo sono stati evitati o, quantomeno, fortemente ridimensionati. Sono state realizzate soluzioni adeguate a prevenire il rischio in materia di privacy, intellectual property e responsabilità amministrativa degli enti ex D. Lgs. 231/2001. Ciò ha reso il prodotto più sicuro ed, al contempo, ha incrementato il know how dell'azienda, attribuendole un valore aggiunto. Pertanto, l'esperienza K-people, concretatasi in una gestione dei rischi giuridici a monte della realizzazione pratica dell'idea innovativa, rappresenta un esempio positivo per tutte le realtà operanti nel settore ICT: innovare in sicurezza, investendo preventivamente in un'attività necessaria e proficua per l'azienda. Il raggiungimento di tutti gli obiettivi prefissati nel caso K-people dimostra l'efficacia dell'approccio metodologico che, ragionevolmente, può essere ritenuto utilizzabile, con i medesimi risultati, per prodotti o processi analoghi nel settore ICT e non solo. Questa esperienza dimostra, altresì, concretamente che i fattori di rischio non sono, di per se stessi, negativi, potendo essere, al contempo, portatori di opportunità¹² per l'azienda. A titolo di esempio, in tema di responsabilità amministrativa degli enti ex D. Lgs. 231/2001, un approccio di legal risk management impone all'azienda l'onere di munirsi di un modello organizzativo, di un codice etico, nonché di un organismo di vigilanza; al contempo, però, tali adempimenti le consentono di sgravarsi da qualsivoglia forma di responsabilità per una vasta serie di reati commessi, o tentati, a suo vantaggio da soggetti apicali ovvero da dipendenti, tutelandone la sicurezza giuridica e migliorandone l'immagine pubblica.

La necessità di adeguarsi alle norme di diritto, ovviamente, è una circostanza particolarmente sentita da quelle aziende che fanno dell'innovazione tecnologica il proprio core business, dovendosi confrontare con una realtà costantemente mutevole che le costringe fisiologicamente ad adeguate "protezioni" giuridiche¹³. L'innovazione tecnologica, infatti, è portatrice di rischi ed opportunità: "mission" del legal risk management è l'implementazione di un'attività integrata volta ad eliminare i primi e a sfruttare le seconde. La creatività tipica dei prodotti ICT necessita in modo particolare di una attività manageriale orientata, tramite il controllo dei rischi giuridici d'impresa, alla tutela del vantaggio competitivo intrinsecamente connesso all'introduzione di qualsiasi prodotto innovativo rispetto allo stato dell'arte. Tale approccio metodologico permette di coniugare i benefici di una gestione preventiva dei rischi giuridici con le

¹² Cfr. M. Giorgino, 2008, *Il risk management nelle imprese italiane. Come progettare e costruire sistemi e soluzioni per la gestione dei rischi d'impresa*, Il Sole 24 Ore.

¹³ Cfr. F. Bravo, 2010, *Responsabilità delle società da reato ex d.lgs. 231/2001. Il controllo sociale della criminalità economica e i reati informatici*, Clueb.

¹⁴ Fattore critico di successo è un termine che indica un elemento necessario ad un'organizzazione o ad un progetto per realizzare la sua missione. Secondo Jack F. Rockart il fattore critico di successo è costituito da quel numero limitato di aree i cui risultati, ove siano soddisfacenti, assicurano all'organizzazione un rendimento competitivo positivo.

nuove opportunità di business, costituendo, in tal modo, un autonomo fattore critico di successo¹⁴ per l'azienda.

7. Il Legal Risk Management come fattore critico di successo: il caso K-people

L'attività di legal risk management, attraverso l'implementazione di un'attività integrata idonea a gestire i rischi giuridici e a sfruttare le relative opportunità di business, costituisce un vero e proprio fattore critico di successo per l'azienda. Come è ben noto, in tema di management, con fattore critico di successo si intende quell'elemento indispensabile affinché un'azienda o un progetto possa realizzare la propria "mission": tutte le aziende modernamente strutturate, infatti, implementano una pianificazione strategica finalizzata al raggiungimento di obiettivi prefissati, facendo leva proprio su quei fattori critici di successo che possono determinare per l'azienda stessa un vantaggio competitivo. Tradizionalmente, sono considerati fattori critici di successo, ad esempio, la crescita dei margini di profitto, la creazione di prodotti o servizi di qualità, la soddisfazione delle esigenze dei clienti, nonché la capacità di acquisirne di nuovi, la creazione di nuove relazioni strategiche con connesse opportunità di business; ad essi, poi, va ad aggiungersi la sostenibilità da parte dell'azienda, ossia la capacità di mantenere indefinitamente tali fattori critici di successo.

Il legal risk management, attraverso le sue tipicità, garantisce la sostenibilità dei fattori critici di successo dell'azienda sotto il profilo giuridico, diventando, di per se stesso, un autonomo fattore critico di successo.

Anche in ambito legal, il punto di partenza per lo svolgimento di tale attività è costituito dall'analisi swot¹⁵, notissimo strumento di pianificazione strategica, che consente di dar vita a strategie creative a favore dell'azienda, tramite l'individuazione di quei fattori interni ed esterni strumentali al raggiungimento dell'obiettivo: fattori interni, ossia, punti di forza e di debolezza interni all'organizzazione e fattori esterni, cioè, opportunità e minacce presenti all'esterno della medesima. In tal modo, diventa possibile prevedere e gestire razionalmente non solo i rischi giuridici, ma anche le future opportunità di business, tutelando i fattori critici di successo dell'azienda medesima; il rischio, infatti, nella sua accezione più moderna ed innovativa, non viene valutato da un punto di vista meramente negativo¹⁶, ben potendo essere portatore di concrete opportunità per l'azienda. Prevedere e gestire un rischio giuridico, infatti, contribuisce ad incrementare il patrimonio aziendale in termini di know how: in assoluta convergenza con le esigenze della nostra moderna knowledge society¹⁷, l'attività di legal risk management, andando ad integrare il patrimonio di conoscenza dell'azienda, concorre al suo miglioramento continuo e si pone in linea con le esigenze di un mercato sempre più competitivo ed orientato

¹⁵ La *swot analysis* è una tecnica attribuita ad Albert Humphrey, che ha guidato un progetto di ricerca alla Università di Stanford fra gli anni Sessanta e Settanta. Conosciuta anche come Matrice TOWS, è uno strumento di pianificazione strategica usato per valutare i punti di *forza* (Strengths), *debolezza* (Weaknesses), le opportunità (Opportunities) e le minacce (Threats) di un progetto o in un'impresa o in ogni altra situazione in cui un'organizzazione o un individuo deve prendere una decisione per raggiungere un obiettivo.

¹⁶ Cfr. G. Pellerino, 2007, *Le origini dell'idea del rischio*, Pensa Multimedia.

¹⁷ Cfr. V. D'Agnese, 2003, *La costruzione e la gestione dei saperi nella knowledge society*, Pensa Multimedia.

all'acquisizione di informazioni strategiche¹⁸. Il settore ICT, più di ogni altro, fonda il proprio business sulla capacità di fruire di quel patrimonio multidisciplinare di conoscenze che sono sottese all'innovazione tecnologica e che vanno al di là delle mere competenze tecniche, attraversando settori diversi ma connessi tra loro; possedere una base di conoscenza superiore a quella dei competitors consente di innovare prima di loro, ricavando un vantaggio competitivo: il legal risk management diventa, così, patrimonio dell'azienda, costituendo un autonomo fattore critico di successo. Nel caso K-people, l'approccio orientato al legal risk management nell'implementazione di questa piattaforma software di condivisione dati ha consentito a Webscience di poter gestire oculatamente e mantenere inalterato il vantaggio competitivo connesso all'introduzione sul mercato ICT di tale prodotto innovativo. Infatti, la previsione dei rischi giuridici connaturati a questo prodotto software ha messo in condizione Webscience di poter adottare preventivamente tutti gli opportuni accorgimenti finalizzati alla tutela del medesimo prodotto, dei processi aziendali connessi ed, ovviamente, dell'azienda. Non solo. L'identificazione dei suddetti rischi ha evidenziato quelle opportunità di business che spesso si accompagnano a situazioni di rischio. In quest'ottica, Webscience ha implementato fattivamente un approccio innovativo al rischio, come evento portatore di conseguenze non necessariamente negative. Sicché, all'esito di un'attività preventiva di analisi dei rischi giuridici connessi al software K-people, sono state individuate delle aree critiche. Successivamente, studiando nel dettaglio le specifiche di progetto, sono state focalizzate le specifiche criticità che, andando ad impattare con le disposizioni di legge vigenti nelle su dette materie, necessitavano di un adeguamento concreto. In tal modo è stato possibile intervenire, in tempo reale, sul prodotto software e sui relativi processi aziendali, accrescendone la sicurezza e contribuendo alla stabilità dell'azienda. Tutte queste attività hanno creato nuove ed importanti opportunità: Webscience è stata avviata alla cultura della gestione integrata dei rischi giuridici; ha accresciuto il proprio know-how, trovandosi, in tal modo, nella condizione di saper gestire situazioni di rischio analoghe con un netto risparmio in termini di tempo, energie e risorse; il prodotto software K-people ha trovato una adeguata tutela giuridica, sicché l'azienda ha potuto beneficiare della propria innovativa idea di business, conservando inalterato il vantaggio competitivo ad essa connesso; le problematiche in termini di privacy¹⁹ relative all'utilizzo del software sono state gestite razionalmente attraverso una suddivisione dei ruoli e delle responsabilità tra i soggetti apicali all'interno dell'azienda, la predisposizione di adeguate contromisure tecnico-giuridiche ed il contestuale adattamento delle specifiche del software alle disposizioni di legge; l'azienda è stata sensibilizzata a conformarsi ai dettami di cui al D. Lgs. 231/2001 ed orientata alla predisposizione di quegli strumenti tecnico giuridici necessari alla prevenzione della responsabilità amministrativa da reato: modello di organizzazione, gestione e controllo, codice etico e

¹⁸ Cfr. R.M. Grant, 1994, *L'analisi strategica nella gestione aziendale*, Il Mulino.

¹⁹ Cfr. L. Bolognini, D. Fulco, P. Paganini, 2010, *Next privacy. Il futuro dei nostri dati nell'era digitale*, Etas.

²⁰ Cfr. C. Sunna, 2009, *Problemi e soluzioni per la tutela della privacy nella costruzione di banche dati integrate*, Cacucci.

organismo di vigilanza. In sintesi, questo approccio al concetto di rischio giuridico ha cambiato radicalmente la strategia dell'azienda, abituandola a valutare gli impatti giuridici connessi ad ogni singola attività svolta e a contemperarli con i benefici prodotti; ha diminuito sensibilmente il rischio di contenzioso; ha aumentato la capacità di transigere le liti insorgende, limitandone gli effetti negativi; ha migliorato, conseguentemente, il lato pubblico dell'azienda, costruendole addosso un'immagine virtuosa, e ne ha reso espliciti i valori.

L'esperienza K-people dimostra, a livello pratico, l'efficacia nel settore ICT²⁰ di un'attività di gestione dei rischi giuridici, nonché la possibilità concreta, per l'azienda, di cogliere le nuove opportunità di business ad essi collegate; a livello teorico, dimostra l'adattabilità di un approccio orientato al legal risk management a contesti, processi e prodotti diversi, nonché la sua capacità di costituire, di per se stesso, un autonomo fattore critico di successo.

Acknowledgements

Unione Europea e Regione Puglia

Questo lavoro è stato realizzato nell'ambito del progetto di ricerca K-People, intervento cofinanziato dall'U.E. – F.E.S.R. sul P.O. Regione Puglia 2007-2013, Asse I-Linea 1.1. - Azione 1.1.2. Aiuti agli Investimenti in Ricerca per le P.M.I.

Bibliografia

- Bolognini L., Fulco D., Paganini P., 2010, *Next privacy. Il futuro dei nostri dati nell'era digitale*, Etas
- Bravo F., 2010, *Responsabilità delle società da reato ex d.lgs. 231/2001. Il controllo sociale della criminalità economica e i reati informatici*, Clueb
- D'Agnese V., 2003, *La costruzione e la gestione dei saperi nella knowledge society*, Pensa Multimedia
- De Santis V.M., 2010, *Manuale del nuovo diritto d'autore*, Editoriale scientifica
- Giorgino M., 2008, *Il risk management nelle imprese italiane. Come progettare e costruire sistemi e soluzioni per la gestione dei rischi d'impresa*, Il Sole 24 Ore
- Grant R. M., 1994, *L'analisi strategica nella gestione aziendale*, Il Mulino
- Nielsen R., Sandfeld Jakobsen S., Trzaskowski J., 2004, *Legal Risk Management in Cross-Border Electronic Commerce*, Article, *EU Electronic Commerce Law*, Djøf Publishing
- Pellerino G., 2007, *Le origini dell'idea del rischio*, Pensa Multimedia
- Sunna C., 2009, *Problemi e soluzioni per la tutela della privacy nella costruzione di banche dati integrate*, Cacucci
- Trzaskowski J., 2003, *Cross-Border Law Enforcement in the Information Society* (v. 0.82) Article, DJØF Publishing
- Trzaskowski J., 2005, *Legal Risk Management - Some Reflections*, Article, DJØFPublishing
- Trzaskowski J., 2005, *Legal Risk Management in Electronic Commerce - Managing the risk of cross-border law enforcement Book* (PhD thesis), Ex Tuto Publishing
- Trzaskowski J., 2006, *Legal Risk Management in a Global, Electronic Marketplace* Article, *Scandinavian Studies in Law*, Volume 49 (A Proactive Approach)

SEMANTIC-PA: Una Piattaforma per la Gestione Semantica dei Contenuti per la Pubblica Amministrazione

Claudia d'Amato, Nicola Fanizzi, Floriana Esposito
Università degli Studi di Bari "Aldo Moro"
Via Orabona, 4 – 70125 Bari

E-mail: [claudia.damato|fanizzi|esposito}@di.uniba.it](mailto:{claudia.damato|fanizzi|esposito}@di.uniba.it)

Michele Carulli, Alessandro Assab, Giuseppe Bux
I&T Servizi S.r.l.

C.so A. De Gasperi, 320 – 70124 Bari

E-mail: [michele.carulli|alessandro.assab|giuseppe.bux}@ietservizi.it](mailto:{michele.carulli|alessandro.assab|giuseppe.bux}@ietservizi.it)

Il panorama attuale dell'eGovernment è caratterizzato da portali che mettono a disposizione dei cittadini servizi e documenti digitalizzati catalogati rispetto a categorie predefinite e rigide. Tali portali hanno uno scarso livello di integrazione tra di essi. In questo articolo si presenta SEMANTIC-PA, un portale caratterizzato dalla immediata fruizione dei contenuti da parte degli utenti con possibilità di integrazione semantica con le altre realtà di interesse. Questo obiettivo è raggiunto mediante: 1) la realizzazione di un modulo in grado di categorizzare automaticamente i contenuti del portale sulla base della loro semantica piuttosto che rispetto a categorie predefinite. 2) La realizzazione di un motore di ricerca semantico che funziona coerentemente anche se risorse annotate semanticamente e/o ontologie di dominio non sono disponibili.

1. Introduzione

Obiettivo dell'eGovernment è rendere snella e veloce l'interazione del cittadino con la Pubblica Amministrazione (PA). Quest'obiettivo è perseguito rendendo disponibili informazioni e servizi 24 ore su 24. Partito come una collezione di siti Internet contenenti per lo più copie digitalizzate di documenti stampati, l'eGovernment si è evoluto rendendo disponibili online, mediante portali, un insieme d'informazioni e servizi che siano direttamente fruibili ai cittadini. Tuttavia, l'integrazione con le molteplici realtà di una regione e/o con altre realtà regionali e nazionali è attualmente pressoché inesistente.

Una costante dei portali di eGovernment è la presenza di un motore di ricerca per il ritrovamento delle informazioni. La ricerca ed il ritrovamento di informazioni è generalmente eseguita mediante matching di parole chiave

specificate dagli utenti rispetto ai contenuti dei documenti e/o delle pagine Web. Questi metodi, alla base anche del funzionamento dei più importanti motori di ricerca quali yahoo e google, sono caratterizzati da due problemi fondamentali: a) l'information overload, dovuto a casi di polisemia; b) il mancato ritrovamento di risorse nel caso di non concordanza di singolari/plurali o di uso di sinonimi. Per ovviare a questi problemi, nuovi metodi possono essere posti in essere muovendo dalla formulazione Web Semantico (SW) [T. Berners-Lee et al, 2001]. Obiettivo del SW è rendere fruibile l'informazione su web, oltre che all'uomo, a macchine, agenti intelligenti, siti web personalizzati e motori di ricerca potenziati semanticamente, in modo da automatizzare, integrare e riusare la conoscenza e le informazioni disponibili. Quest'obiettivo è realizzato arricchendo le risorse¹ con metadati ovvero annotazioni semantiche che fanno riferimento ad *ontologie*² condivise e che hanno la finalità di definire in maniera esplicita il significato delle risorse mediante l'uso di descrizioni formali automaticamente processabili dalle macchine. Elementi fondamentali di una ontologia sono: concetti, relazioni (tra concetti) ed individui (istanze di concetti e/o relazioni). Il linguaggio impostosi come standard de facto nel SW è OWL [OWL] che trova il suo fondamento teorico nelle logiche descrittive (DLs) caratterizzate da una semantica descrittiva e da diversi servizi di ragionamento. Svariati motori di ricerca semantici (ovvero motori in grado di ritrovare informazioni semanticamente rilevanti rispetto alla query specificata) sono stati sviluppati quali: SWSE [SWSE], HAKAI [HAKAI], SHOE [SHOE], SWOOGLE [SWOOGLE]. La loro maggior limitazione è data dall'assunzione di disporre di ontologie di dominio e risorse sul Web annotate semanticamente.

In questo articolo si presentano i principali risultati ottenuti dal progetto POR "Semantic-PA", finanziato dalla Regione Puglia, avente come obiettivo la realizzazione prototipale di una piattaforma software, SEMANTIC-PA, per la PA basata sull'uso di tecnologie del SW. Obiettivo del progetto è integrare informazioni e servizi della PA in un portale, con particolare attenzione alla immediata fruizione dei contenuti da parte degli utenti. Da un lato si categorizzano le informazioni ed i documenti in maniera automatica, su base semantica piuttosto che rispetto a categorie predefinite, non sempre adeguate a qualsivoglia realtà amministrativa. Le categorie apprese sono poi usate per consentire la navigazione a faccette dei contenuti del portale. D'altro canto si mette a disposizione un motore di ricerca semantico che sia in grado di funzionare coerentemente anche nel caso in cui le risorse non siano annotate semanticamente e/o non siano disponibili ontologie dominio. Questi obiettivi individuano le due macro-componenti su cui si articola la piattaforma SEMANTIC-PA: 1) Una **Componente responsabile di:** Aggregazione semantica e Multi-classificazione Automatica dei contenuti del portale rispetto a molteplici punti di vista; 2) Una componente **Motore di Ricerca Semantico.**

¹ Una risorsa è qualunque entità, astratta o concreta (es. Pagina web o servizio web quale ad esempio una prenotazione aerea), identificabile da una URI (Uniform Resource Identifier)

² Una ontologia è "una specificazione formale di una concettualizzazione condivisa" [T. Gruber, 1993]. Essa gioca il ruolo di vocabolario dei meta-dati, dal quale gli annotatori di risorse possono prendere e riusare terminologie e dove le applicazioni che consumano meta-dati possono trovare un riferimento del loro significato.

Il raggiungimento di questi obiettivi costituisce un importante risultato in quanto: 1) rende possibile l'interrogazione semantica del così detto Web 2.0 [T. O'Reilly 2005]; 2) semplifica il problema dell'annotazione delle risorse e della costruzione di ontologie che attualmente rappresentano un collo di bottiglia nel processo di acquisizione della conoscenza funzionale alla realizzazione del SW.

Le due macro-componenti ed i metodi implementati sono di seguito descritti insieme alla discussione di alcuni lavori correlati.

2. Aggregazione semantica e Multi-classificazione Automatica dei contenuti del portale

La componente di Aggregazione Semantica e Multi-classificazione automatica dei contenuti del portale ha come finalità produrre una categorizzazione automatica, su base semantica, dei contenuti informativi del portale piuttosto che sulla base di categorie predefinite. Inoltre, poiché le risorse possono essere rilevanti sotto diversi punti di vista, la categorizzazione è utilizzata per la generazione di percorsi di navigazione a faccette utili anche per l'esplorazione dei risultati di una query. Un sistema di classificazione a faccette, consente la determinazione di classificazioni multiple per uno stesso oggetto, abilitando la categorizzazione dello stesso sotto diversi punti di vista. Inoltre l'uso di faccette apprese automaticamente a partire dalle risorse disponibili consente di mostrare esclusivamente categorie popolate, evitando la visualizzazione di categorie prive di risultati. Questo approccio contrasta con le tassonomie tradizionali, in cui le gerarchie di categorie sono fissate ed immutabili nel tempo.

L'elemento di novità che si introduce in questa componente ed in questo progetto è la categorizzazione dinamica ed automatica delle informazioni, in grado di cambiare al variare (nel tempo) dei contenuti del portale. Questo è realizzato mediante l'applicazione di metodi di clustering (concettuale) gerarchici [A. K. Jain et al., 1999] la cui finalità è la scoperta di gruppi omogenei (cluster) di risorse organizzati gerarchicamente. Le risorse possono appartenere a cluster differenti, a livelli diversi o allo stesso livello della gerarchia, rappresentando i molteplici punti di vista mediante i quali esse possono essere ritrovate e/o esplorate. I cluster sono etichettati con nomi rappresentativi al fine di rendere esplicativo il loro contenuto.

I metodi di clustering sono applicati a risorse annotate semanticamente. Tuttavia, come discusso nella sez. 1, la piattaforma SEMANTIC-PA ha come obiettivo processare, in maniera coerente, anche collezioni di risorse che non siano semanticamente annotate, considerando la potenziale assenza di ontologie di dominio. Per colmare questo gap, la macro-componente di Aggregazione Semantica e Multi-classificazione automatica dei contenuti del portale dispone di una componente che si fa carico di eseguire, in maniera automatica, l'annotazione semantica delle risorse disponibili.

Di seguito si presentano i dettagli del processo adottato, a partire dall'annotazione delle risorse e la successiva aggregazione.

2.1 L'annotazione Automatica delle Risorse

La macro-componente di Aggregazione Semantica e Multi-classificazione automatica dei contenuti del portale sopperisce alla mancanza di risorse annotate semanticamente generando uno o più tag semanticamente attinente ad ogni risorsa. A tale scopo, metodi per la generazione di modelli probabilistici sono stati adottati. Le tecniche statistiche di indicizzazione automatica di documenti basate sul loro contenuto tematico che sono state messe a punto sono diverse, come ad esempio le tecniche di *clustering* [Nigam et al., 2000], *conceptual indexing* (CI) e *latent semantic indexing* (LSI) [Landauer & Dumais, 1997] quest'ultima basata su *latent semantic analysis* (LSA). Più recentemente è stato sviluppato un framework generale basato su *latent Dirichlet analysis* [Blei et al., 2003], detta anche modellazione statistica delle tematiche (*topic*) [Griffiths & Steyvers, 2004]. Tale framework è stato adottato in questo progetto. L'idea di base è che ogni documento di un corpus documentale sia associato ad una distribuzione di probabilità sui topic. Ogni topic è rappresentato da una distribuzione di probabilità multinominale sull'insieme delle parole di un vocabolario (eventualmente estratto dal corpus documentale). Le distribuzioni congiunte documenti-topic e topic-parole sono apprese in maniera automatica tramite tecniche non-supervisionate (ossia prive di un intervento umano di etichettatura e senza presupporre alcuna conoscenza preventiva). Il topic modelling è quindi totalmente data-driven. L'organizzazione statistica sottesa dei topic prevede probabilità non nulle per tutte le parole del vocabolario, ma con una tendenza non uniforme, ovvero con gran parte della massa di probabilità concentrata su un insieme ristretto di parole. Così i topic hanno il vantaggio di essere allineati ai temi dello specifico corpus sul quale viene addestrato il sistema. Inoltre, il modello probabilistico che sovrintende all'organizzazione dei topic consente l'annotazione automatica di ogni parola di un documento col topic che più verosimilmente l'ha generato.

L'Output del processo di annotazione (o tagging) è costituito dai *tag* associati ai documenti del corpus. Un *tag* rappresenta l'etichetta attribuita ad un *topic* (ad esempio la parola nel topic a più alto valore di probabilità). Così un tag può essere associato ad un documento (o parte di esso) anche se il termine che lo rappresenta non compare affatto all'interno del documento.

Di seguito si illustra in dettaglio del processo di tagging a partire dalla generazione del modello probabilistico.

2.1.1 Generazione del modello probabilistico

Il processo di generazione del modello probabilistico si svolge in tre passi: 1) Generazione del vocabolario delle Parole; 2) Generazione dei topic; 3) Campionamento dei Topic (si veda anche Fig. 1).

Creazione Modello

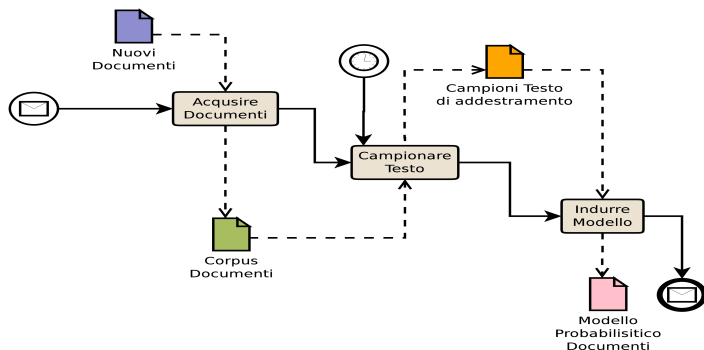


Fig.

1:

Schema di Processo di generazione del modello Probabilistico

Il vocabolario è costituito da un indice delle *parole* presenti nei documenti del corpus documentale. Le *parole* sono unità di base del vocabolario indicizzato, ad es. $\{1, \dots, V\}$ (possono essere rappresentate come vettori con un solo componente pari ad uno e tutti gli altri a zero). Un *documento* è una sequenza di N parole (rappresentato da una bag-of-words (now), ovvero un vettore di occorrenze delle parole). Un *corpus* è un insieme di M documenti. La generazione dei topic del corpus documentale segue un modello generativo probabilistico noto come *Latent Dirichlet Allocation* (LDA) [Blei et al., 2003]. In particolare, la LDA segue il seguente processo generativo per ogni documento del corpus. Ogni *topic* z_i è rappresentato da una distribuzione di probabilità (multinomiale) sulle V parole: $p(w_i | z_j)$, indicato per convenienza con la/le parola/e più probabile/i. In Tabella sono riportati due esempi di *tag/topic* e relative distribuzioni multinomiali (per ogni topic è indicata la distribuzione (parziale) per le parole più probabili mentre i tag sono indicati come intestazioni delle colonne). Si procede quindi col determinare (*inferenza probabilistica*) la distribuzione delle parole dato un documento, in base alle distribuzioni parole-topic e topic-documenti ossia $p(w_i | d) = \sum_{j=1..T} p(w_i | z_j) p(z_j | d)$

HEALTH CARE		FARMING	
health	(0.064)	farm	(0.081)
care	(0.058)	crop	(0.027)
plan	(0.047)	cow	(0.018)
cost	(0.043)	field	(0.015)
insurance	(0.042)	corn	(0.015)
benefit	(0.032)	food	(0.012)
converage	(0.023)	bean	(0.010)
pay	(0.020)	cattle	(0.010)
program	(0.013)	market	(0.010)

Essendo la determinazione esatta di tali distribuzioni generalmente intrattabile, si ricorre ad approssimazioni mediante tecniche di campionamento (metodi *Montecarlo*, *Gibbs sampling*, ecc.). Le associazioni documenti-topic sono determinate a partire dalle distribuzioni topic-documenti risultanti dall'attività di campionamento dei topic a chiusura del processo di generazione del modello probabilistico. Nello stesso modello può essere considerata una gerarchia di concetti preesistente mappabile sulle parole del vocabolario [C. Chemudugunta et al., 2008].

2.1.2 Generazione Tag

Nel corso di questa fase vengono assegnate le etichette (i tag) a ciascun documento del Corpus Documentale. L'assegnazione avviene in maniera automatica a fronte della generazione del modello probabilistico. Viene comunque consentito all'utente di assegnare propri tag ai documenti di interesse. In particolare, l'etichettatura avviene assegnando la/le parola/e più probabile/i nella distribuzione del topic.

2.2 Aggregazione di Tag e risorse

La fase di Aggregazione dei tag è finalizzata alla generazione di una struttura gerarchica di raggruppamenti di tag da utilizzare per la visualizzazione a faccette dei contenuti del portale. L'output di questa fase è una gerarchia in cui il nodo radice contiene l'intero insieme dei tag, mentre i nodi intermedi ed i nodi foglia sono sottoinsiemi di tag, semanticamente correlati ed omogenei, aventi una relazione di generalizzazione/specializzazione con il nodo padre.

Al fine di determinare raggruppamenti omogenei è necessaria una misura per determinare il livello di similarità tra due tag (o gruppi di tag). Si tratta di una funzione matematica che, dati due tag, restituisce un valore generalmente in $[0,1]$ dove 0 indica massima dissimilarità ed 1 massima similarità. La similarità viene stimata sulla base di un insieme di caratteristiche. Intuitivamente, due tag sono tanto più simili quanto maggiore è il numero di caratteristiche comuni; sono tanto più dissimili quanto maggiore è il numero delle caratteristiche non condivise. E' dunque fondamentale determinare le caratteristiche in base a cui confrontare i tag. Nello caso in oggetto, i tag possono essere confrontati sulla base delle loro distribuzioni di probabilità, rendendo possibile l'applicazione di misure di distanza tra vettori (es. Distanza euclidea, cosine distance...) o misure come la *divergenza di Kullback-Leibler* (o divergenza KL) [A. Banerjee et al., 2005]. Una volta definita la misura, un processo di clustering [A. K. Jain et al., 1999] applicato alla collezione di tag è posto in essere al fine di ottenere un insieme di raggruppamenti (cluster) di tag organizzati gerarchicamente. L'algoritmo implementato nella componente in esame è l'algoritmo di clustering gerarchico bottom-up noto con il nome di *average linkage* [A. K. Jain et al., 1999] e di seguito sintetizzato.

1. Sia $C = (t_1, t_2, \dots, t_n)$ l'insieme dei tag da clusterizzare
2. Si calcoli la dissimilarità tutte le coppie di elementi contenuti in C (al primo passo gli elementi sono cluster costituiti da un unico elemento)
3. Si determini la coppia (t_k, t_m) di dissimilarità minima dove la dissimilarità tra cluster è calcolata come la media delle misure di distanza tra tutte le coppie di elementi in un cluster:
$$D(C^1, C^2) = \text{avg}\{D(t^1, t^2) \mid t^1 \in C^1, t^2 \in C^2\} = 1/(|C^1|*|C^2|) \sum \sum D(t^1, t^2)$$
dove la distanza D tra gli elementi t è fissata secondo quanto sopra.
4. Si elimini t_k, t_m da C e si aggiunga l'elemento $C=(t_k, t_m)$ in C
5. Si ripeta i passi da (2) finchè C è costituito da un unico elemento

Questo algoritmo ha il vantaggio di creare raggruppamento con alto livello di granularità (anche perchè ad ogni passo solo due elementi sono raggruppati). Tuttavia questo diventa uno svantaggio quando gli elementi da raggruppare sono in numero elevato. Per ovviare a questo problema un processo di clustering gerarchico partitivo top-down come il K-Means [A. K. Jain et al., 1999] può essere applicato.

Ciascun raggruppamento omogeneo di tag costituisce una faccetta di classificazione e ricerca dei documenti associati ai tag (in quanto si dispone delle corrispondenze tag-documenti). Al fine di rendere la gerarchia più facilmente esplorabile, i cluster possono essere etichettati con dei nomi esplicativi per ognuno di essi. Un possibile modo è utilizzare il centroide di un cluster, ovvero il tag all'interno del cluster avente distanza minima rispetto a tutti gli altri tag contenuti nello stesso cluster. Una esemplificazione del risultato di un processo di clustering è riportato di seguito (dove gli elementi t1,t'2,... rappresentano i tag mentre le etichette che li precedono rappresentano i nomi assegnati ai cluster). Il processo di clustering può essere ripetuto ogni qualvolta

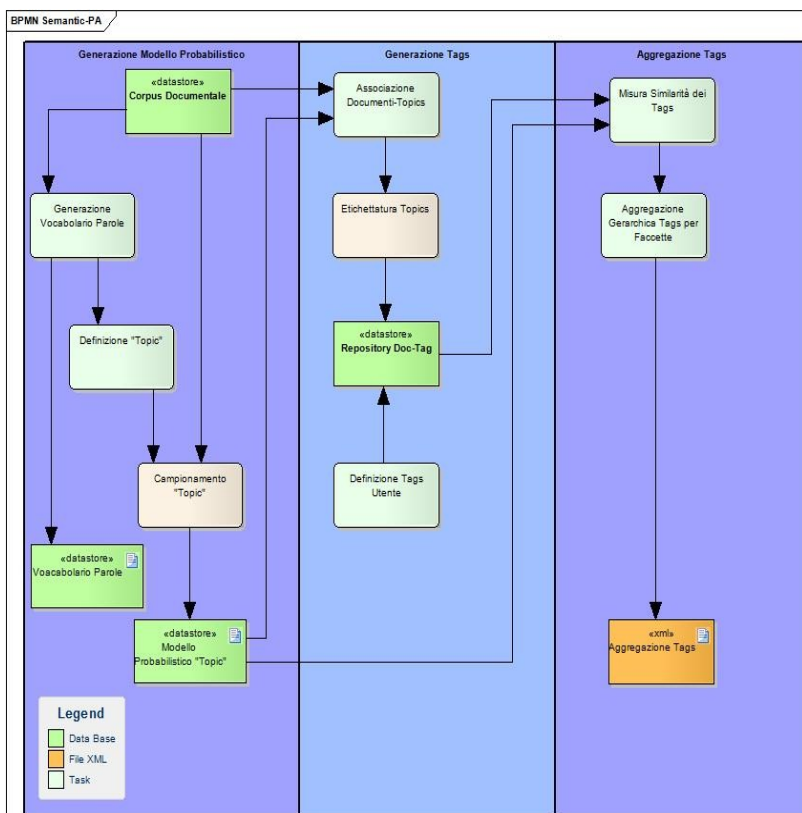


Fig. 2: Processo Semantic-PA di Categorizzazione e Classificazione Semantica di risorse informative

si intenda aggiornare la struttura gerarchica in seguito all'introduzione di un numero sufficiente di nuove risorse tra i contenuti del portale.

In Fig. 2 è sintetizzato il processo di Categorizzazione e Classificazione Semantica di risorse illustrato in questa sezione.

Inquinamento	Salute
Ambientale (t1, t2, ..., tn)	Servizi cittadini(r1,..., rf)
Acqua (t1,t2)	Cliniche (r'1,...,r'd)
Aria (t3,...,tn)	
Sonoro (t'1, t'2, ..., t'k)	
Luminoso (t''1, ..., t''g)	

3. Il Motore di Ricerca Semantico

La macro-componente che realizza il motore di ricerca semantico ha come obiettivo il ritrovamento su base semantica delle informazioni relative a query degli utenti. Diversamente dalle classiche tecniche basate sull'uso di parole chiave, il ritrovamento di informazioni sfruttando la semantica delle richieste e delle risorse disponibili consente di ridurre il numero di informazioni non rilevanti restituite ed il numero di risorse rilevanti ma non ritrovate perché non contenuti i medesimi termini delle query. Nella composizione della query l'utente è supportato da una interfaccia contenente i concetti ed i ruoli della ontologia di riferimento. L'utente sceglie tra di essi per comporre la query (and ed or possono essere usati), quindi il motore di ricerca, restituisce le risorse istanza dei concetti e dei ruoli specificati nella query, mediante l'invocazione di un apposito reasoner (Pellet [PELLET] è stato usato allo scopo).

In questo processo, si suppone di disporre di una ontologia di riferimento, fondamentale non solo per l'interrogazione e la ricerca semantica ma anche per consentire l'integrazione semantica del portale con altre sorgenti esterne. A tale scopo uno SPARQL endpoint può essere messo a disposizione.

Tuttavia, come argomentato nella Sez. 1, uno degli obiettivi fondamentali della piattaforma SEMANTIC-PA è consentire la ricerca semantica anche in assenza di una ontologia di dominio di riferimento. Ciò è reso possibile in quanto la macro-componente che realizza il motore di ricerca semantico si fa carico di creare una bozza di ontologia nel caso essa non sia disponibile. Tale ontologia viene creata a partire dalla gerarchia di tag/categorie (tassonomia) appresa dalla macro-componente di Aggregazione Semantica e Multi-classificazione automatica dei contenuti del portale. Specificatamente la creazione dell'ontologia avviene traducendo la gerarchia appresa nel formato RDF-S/OWL seguendo i seguenti passi. 1) L'etichetta di ogni nodo cluster è definita come il nome di un concetto atomico in RDF-S/OWL; 2) Le URI dei documenti e delle risorse annotati con i tag appartenenti al cluster sono salvate come individui istanza del concetto rappresentato dal cluster (si noti che un documento/risorsa può avere più tag contenuti in cluster differenti per cui le URI delle risorse possono essere istanza di più concetti all'interno della ontologia che si sta creando); 3) Le relazioni padre/figlio tra cluster sono tradotte come

relazioni is-a ovvero relazioni di sussunzione tra i concetti RDF-S/OWL che descrivono i cluster. Il processo di apprendimento e creazione di bozze di ontologie è sintetizzato in Fig. 3. Si noti che questo processo apre le porte a nuove linee di ricerca quali ad esempio l'apprendimento di (bozze) di ontologie da Folksonomie (elementi caratterizzanti del Web 2.0) in quanto i tag prodotti nel processo di annotazione sono del tutto analoghi alle annotazioni prodotte dal cosiddetto *social tagging* [T. O'Reilly 2005].

La bozza di ontologia appresa può essere di interesse anche nel caso in cui una ontologia di dominio sia disponibile. In particolare la bozza di ontologia può essere utilizzata per arricchire/raffinare l'ontologia di dominio con la nuova conoscenza appresa a partire dalla conoscenza disponibile all'interno del portale. Ciò avviene immergendo la bozza di ontologia nella ontologia di dominio. Nuove relazioni di sussunzione tra i concetti della ontologia di dominio ed i concetti della bozza di ontologia possono essere introdotte avvalendosi dell'invocazione di un apposito reasoner (es. pellet, fact, racer).

Si noti che, a causa dell'esiguo numero di ontologie attualmente esistenti, SEMANTIC-PA si muove sull'assunzione che nessuna ontologia di dominio sia disponibile. Nel caso in cui una ontologia di dominio sia disponibile ma non si siano risorse annotate semanticamente, si annotare le risorse con metodi alternativi a quello presentato. Questo aspetto sarà oggetto di investigazioni e sviluppi futuri.

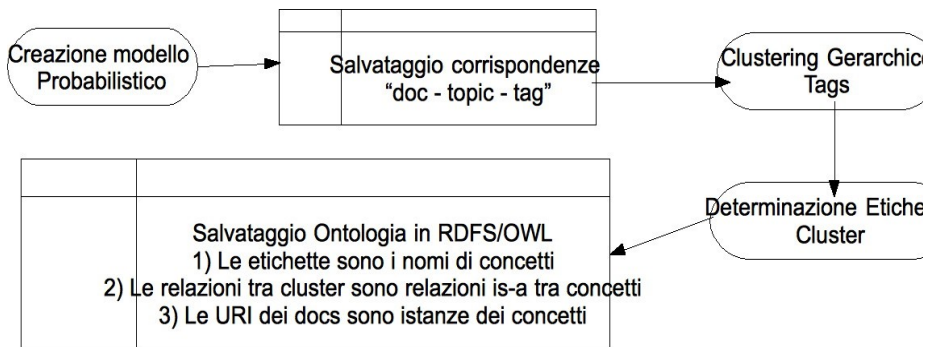


Fig. 2: Schema di Processo per la generazione di bozze di ontologie

4. Conclusioni

In questo articolo è stata presentata la piattaforma SEMANTIC-PA, avente come obiettivo l'integrazione di informazioni e servizi della Pubblica Amministrazione in un portale caratterizzato dalla immediata fruizione dei contenuti da parte degli utenti. Questo obiettivo è stato raggiunto mediante: 1) la realizzazione di un modulo in grado di categorizzare automaticamente i contenuti del portale sulla base della loro semantica piuttosto che rispetto a categorie predefinite. 2) La realizzazione di un motore di ricerca semantico in grado di funzionare in maniera efficace anche nel caso in cui risorse annotate semanticamente e/o ontologie di dominio non siano disponibili.

Diverse sono le migliorie che si possono apportare. A livello di interfaccia, si può prevedere la possibilità di inserimento di query in linguaggio naturale che vengano poi processate al fine di estrapolarne la semantica e mappate rispetto ad una ontologia di dominio o ad una bozza di ontologia appresa. Rispetto all'apprendimento di una bozza di ontologia, un ulteriore livello di granularità può essere ottenuto apprendendo relazioni semantiche non gerarchiche (es. derivate dalle co-occorrenze di tag nelle distribuzioni Tag/Documenti del modello probabilistico). In alternativa, risorse esterne quali ad esempio DBPedia e Wordnet potrebbero essere usate per arricchire la bozza di ontologia appresa.

Bibliografia

[T. Berners-Lee et al, 2001] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. Scientific American, 2001

[A. Banerjee et al., 2005] A. Banerjee, S. Merugu; I.S. Dhillon, J. Ghosh: *Clustering with Bregman divergences*. J. of Machine Learning Research, vol.6 pp.1705-1749. 2005.

[Blei et al., 2003] D. M. Blei, A. Y. Ng., M. I. Jordan. *Latent Dirichlet Allocation*. Journal of Machine Learning Research., 3: 993–1022, 2003

[G. Cesari, 2006] G. Cesari [Portale Multicanale delle PA del territorio della regione Emilia-Romagna: un inizio di rete semantica di informazioni](#), 2006.

[C. Chemudugunta et al., 2008] C. Chemudugunta, A. Holloway, P. Smyth, M. Steyvers: Modeling Documents by Combining Semantic Concepts with Unsupervised Statistical Learning. ISWC 2008, LNCS 5318: pp. 229-244. Springer. 2008.

[HAKIA] HAKIA <http://hakia.com/>

[Griffiths & Steyvers, 2004] T. L. Griffiths M. Steyvers. *Finding scientific topics*. In Proceedings of the National Academy of Science, volume 101, pp. 5228–5235, 2004

[T. Gruber, 1993] T.R. Gruber. A translation approach to portable ontology specifications. Knowledge Acquisition journal, 1993.

[A. K. Jain et al., 1999] A. K. Jain, M. N. Murty, P. J. Flynn: *Data Clustering: A Review*. ACM Computing Surveys, vol. 31, n. 3, ACM, pp. 264-323. 1999.

[Landauer & Dumais, 1997] T. K. Landauer, S. T. Dumais. *A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction and representation of knowledge*. Psychological Review, 104:211-240, 1997

[M.E. Lai, 2005] M.E. Lai. [Classificare per i cittadini. Ovvero, le faccette al servizio dell'usabilità](#). 2005

[Nigam et al., 2000] K. Nigam, A.K. McCallum, S. Thrun, T.M. Mitchell. *Text classification from labeled and unlabeled documents using EM*. Machine Learning, 39(2/3):103–134, 2000

[T. O'Reilly 2005] T. O'Reilly. ["What Is Web 2.0"](#). O'Reilly Network. 2005.

[PELLET] <http://clarkparsia.com/pellet/>

[SHOE] Semantic Search - The SHOE Search Engine
<http://www.cs.umd.edu/projects/plus/SHOE/search/>

[SWOOGLE] SWOOGLE <http://swoogle.umbc.edu/>

[SWSE] Semantic Web Search Engine <http://swse.org/>

SerendipiTAG - We help you to find your path

Lara Marcellin, Alessandra Petromilli¹, Roberto Politi²

CSP- Innovazione nelle ICT
Via Livorno 60, 10144 Torino, Italy
marcellin@csp.it

¹ CSP- Innovazione nelle ICT
Via Livorno 60, 10144 Torino, Italy
petromilli@csp.it

² CSP- Innovazione nelle ICT
Via Livorno 60, 10144 Torino, Italy
politi@csp.it

User involvement in the process of content generation - typical of web 2.0 paradigm – contributes increasing the amount of data available on the net, making information retrieval more difficult. Fortunately this involvement also provides a new kind of tool which can be managed directly by users for content classification: tagging. The contribution of this paper is focused on how the process of information retrieval can be improved through the use of a bottom up approach based on folksonomies. As a result of our study we present our prototype: SerendipiTAG, an information retrieval model which helps users to find contents of interest more easily and through the serendipity process even to find something unexpected.

1. Introduction

Collaborative Tagging, which is the basis of folksonomy systems describes the process by which many users annotate resources with keywords or specific terms in order to provide better organization, retrieval and sharing of the resource. Recently, collaborative tagging has grown in popularity on the web, on sites that allow users to tag bookmarks, photographs and other kinds of content. Considering this wide adoption of tagging systems, and their ability to support different domains and resource types we have focused our research on a clear demonstration of how it is possible to link web content in meaningful ways. While collaborative tagging provides an excellent basis for users to organize and search content, folksonomies are often under threat of invasion by noise. As a matter of fact folksonomies suffer from a crucial problem: the lack of standardisation in the tagging process.

A key implication of their "anything goes" approach is the potential for highly idiosyncratic classifications. The growth of folksonomies has generated a great deal of discussion regarding their potential to interfere with "official" taxonomies

and thus to generate "search noise". However the governance of this problem shouldn't be overestimated. Some studies by Golder and Huberman [Golder and Huberman, 2006] have shown how quickly a tagging system can stabilize itself: with the frequency of appearance of a tag increasing when a content is frequently referred to a fixed value. This is a very important phenomenon: tag convergence takes place when a reference becomes increasingly popular.

In fact, even if a tag is not associated with any kind of structure, it can be aggregated and compared with other tags coming from other users and/or systems. The results can improve the knowledge of the collective and collaborative vision of the web, and be used to investigate new ways to retrieve information.

Another virtue of folksonomy is "serendipity", or the act of unexpectedly encountering something fortunate.

Dr. Merton described it as "the discovery, by chance or sagacity, of valid results which were not sought for" [Merton and Barber, 2004]. When a serendipity process occurs unexpected connections or enlightened discoveries may be made.

Folksonomies can be successfully used to encourage browsing as opposed to a random find, via a collection of interlinked tags which constitute a fantastic source for identifying unexpected finds, which would never have been revealed without it. As a matter of fact folksonomy sites allow a more open exploration of content than the use of a standard search engine.

Spink et al. [Spink et al, 1998] found that partially relevant search results, identified as "containing multiple concepts, [or] on target but too narrow", play a central role during a user's information searching task and problem definition. This theory suggests that there is value in partially relevant results. In education the "zone-of-learnability" refers to texts that provide optimal learning because they are related to, but just distant enough, from what the student already knows. Since search results are inherently related to the query in some way, there may be a similar zone for serendipity.

Furthermore Herlocker [Herlocker et al, 2004] suggests that measures like novelty and serendipity should be used to evaluate the quality of recommendations. Collaborative filtering systems have promoted novelty and serendipity by helping users uncover less popular [Sarwar et al, 2001] and more diverse items [Van Andel, 1994].

2. SerendipiTAG: a retrieval model applied to an enterprise collaborative environment

SerendipiTAG is a search engine based on tags entered by users which suggests alternative research paths in order to provide better results.

During the search task the use of folksonomies guarantees context independency and as a consequence automatic adaptation to different information providers.

Furthermore the fact that tags are entered only by the users of a certain social network allows SerendipiTAG to correctly represent the socio cultural background of this specific community.

With the aim to test this model we decided to use eCollab (<http://ecollab2-demo.csp.it>), a collaborative environment developed by our company. The tags are entered by the users of our collaborative environment, that is based on three open source platforms: Elgg (BLOG), Mindtouch Dekiwiki (WIKI) and Pligg (LINK). This environment integrates three different social tools: a blog, a wiki and a social bookmarking system.

This collaborative environment aims at facilitating the sharing of expertise, ideas and multimedia contents such as documents, videos and images inside a company.



Fig. 1 - shows SerendipiTAG Interface

SerendipiTAG is currently providing an anonymous access to the retrieval model. (see figure 1).

The query results are shown as a list where the items are ordered by date. In order to help the user to understand from which tools (Blog, Wiki or Link) the results are extracted, they are highlighted in three different colours, which refer to one of the three previously mentioned platforms.

The system also analyzes the search results in order to find related tags (tags which coexist in the same search results) which are suggested at the top of the results page in two clouds: the first one shows all the tags related to the query, while the second suggests the name of the experts on that particular topic. Tags may serve as an important liason in establishing community of practice, allowing users with similar interests to identify each other.

The recommendation algorithm analyzes all the contents published in the targeted social tools in order to extract correlations between tags which have been used to describe contents and between users. The correlations are calculated by taking into consideration different parameters such as the co-presence of tags on the same page, the content post date, the number of users that have used a particular tag, the number of social tools in which that particular tag is used in and so on.

With the goal of evaluating satisfaction and user experience during the use of SerendipiTAG, a within-subjects experiment was performed. People subjected to the experiment were users who are familiar with eCollab. They were 10 females and 10 males between the ages of 22 and 40. Subjects were

invited to use SerendipiTAG for a week during their usual search tasks. The questionnaire was based on the System Usability Scale [Brooke, 1996], a simple ten-item scale giving a global view of subjective assessment of usability. In addition to these questions, we asked users to give a description of what they perceived during navigation.

3. Conclusion

As far as the "System usability scale" (SUS scores have a range of 0 to 100) is concerned, the values have a range of 74 to 100. The calculated mean is 83.55 (distribution of values is significant $\sigma = 9.66$). Furthermore the 90% of users declared that they consider SerendipiTAG useful to perform queries that cover a broad topic or that seek a single page either.

As regards the qualitative evaluation, 76% of the users correctly perceived the usefulness of the two tagclouds.

Although we are encouraged by the findings, this research is best viewed as exploratory. Future research would need to include a larger sample and they will be performed after the integration of SerendipiTAG with eCollab in order to verify if the first experiment's results are confirmed after an heavy and daily use of the system.

Bibliografia

[RIF 1] Brooke, J., "SUS: a "quick and dirty" usability scale". In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland. Usability Evaluation in Industry. London: Taylor and Francis, 1996.

[RIF 2] Golder S., Huberman B., Usage Patterns of Collaborative Tagging Systems. In Journal of Information Science, 2006, pp. 198-208.

[RIF 3] Herlocker, J. L., Konstan, J. A., Terveen, L. G. & Riedl, J. T. Evaluating collaborative filtering recommender systems. In Proceedings ACM TOIS 22, 1 January 2004.

[RIF 4] Merton, RK, Barber, EG., The Travels and Adventures of Serendipity: A Study in Historical Semantics and the Sociology of Science. Merton RK. Published by Princeton University Press American Journal of Sociology, 2004.

[RIF 5] Sarwar, B., Karypis, G., Konstan, J. & Reidl, J. Item-based collaborative filtering recommendation algorithms, In Proceedings of the ACM Conference WWW'01, Aarhus, Denmark August 14-18, 2001, 285--295

[RIF 6] Spink, A., Greisdorf, H. & Bateman, J. 1998. From highly relevant to not relevant: Examining different regions of relevance. In IP&M 34, 5, 1998, 599-621.

[RIF 7] Van Andel, P., Anatomy of the Unsought Finding. Serendipity: Origin, history, domains, traditions, appearances, patterns and programmability. The British Journal for the Philosophy of Science, 1994, Volume 45 Issue 2

Simulazione di Esami Radiologici mediante Applicazione della Tecnologia CUDA

N. Balossino, D. Cavagnino, M. Lucenteforte, S. Rabellino, S. Vaglianti
Dipartimento di Informatica
Università degli Studi di Torino
nello.balossino@di.unito.it

Un sistema di addestramento come quello per gli operatori sanitari radiologici acquisisce maggior valenza quando vengano impiegate, a diversi livelli di sofisticazione, le tecnologie della programmazione parallela. Scopo del lavoro svolto è stato la trasformazione del sistema ViRiS (Virtual Radiologic Image Simulator è un simulatore di un'apparecchiatura radiologica [1]), nato per la simulazione di un atto radiologico, in un ambiente interattivo in grado di offrire all'operatore la sensazione di agire in tempo reale in una sala radiologica, riducendo significativamente i tempi di esecuzione. L'aumento delle prestazioni è stato ottenuto sfruttando i nuovi hardware multicore graphics processing units (GPU) e l'ambiente di programmazione Compute Unified Device Architecture (CUDA [2]). La validità del sistema consiste nella possibilità per gli operatori di perfezionare la tecnica radiologica, evitando di esporre il paziente a dosi di radiazioni inutili ai fini dell'esame diagnostico.

1.Introduzione

Sarà illustrata l'architettura di funzionamento di un'applicazione fortemente parallelizzata utilizzando l'ambiente CUDA ed indirizzata alla simulazione interattiva 3D di una unità operativa di radiologia virtuale. Lo scopo è di fornire agli operatori sanitari che lavorano in tale ambito un raffinato strumento di addestramento. Il documento radiografico virtuale è ottenuto simulando l'evento reale in cui il paziente è rappresentato da un modello tridimensionale densitometrico, realizzato mediante l'acquisizione di un insieme di tomografie computerizzate (TC) di soggetti tipo, a cui viene affiancato un modello del sistema sorgente-sensore radiografico. In una precedente fase di ricerca, alcuni degli autori hanno collaborato alla realizzazione di un simulatore di apparecchiatura radiologica denominata ViRiS, illustrato in convegni e riportato su riviste del settore. In entrambe le occasioni sono state messe in luce le limitazioni delle prestazioni e in particolare non solo il limitato livello di interattività ma principalmente i lunghi tempi di creazione dell'immagine radiologica. Le negatività suddette rendono il sistema ViRiS inutilizzabile in tempo reale come richiesto da un'applicazione che si ponga l'obiettivo di essere

uno strumento didattico. Fermo restando il contenuto prettamente medico-radiologico che non è di competenza degli autori, il sistema ViRiS è stato reingegnerizzato al fine di sfruttare la possibilità di parallelizzazione delle operazioni di rendering, vale a dire della ricostruzione e visualizzazione del risultato di un atto radiologico, in modo da fornire un ambiente che presenti le caratteristiche del tempo reale per quanto riguarda le elaborazioni, di un discreto senso di immersione e di interattività, così come avverrebbe trovandosi di fronte a una apparecchiatura radiologica reale.

L'ambiente CUDA permette di sfruttare il parallelismo insito in schede grafiche di ultima generazione (GPU), la cui struttura si basa sulla presenza di un elevato numero di unità di elaborazione (core) e di una adeguata quantità di memoria ad alta velocità (RAM): questa caratteristica permette di ridurre i tempi di rendering tridimensionale dello spazio di interesse, cioè della sala radiologica e del posizionamento del blocco sorgente-rilevatore, coerentemente al tipo di acquisizione che si voglia realizzare.

2.L'architettura del sistema

L'architettura del sistema dal punto di vista del funzionamento è riportato nel Data Flow Diagram di figura 1. Nel diagramma è evidenziata la separazione tra la parte di esecuzione seriale all'interno della CPU del calcolatore, rispetto all'esecuzione parallela sulle GPU presenti nel sistema e che consentono lo sfruttamento della parallelizzazione dell'algoritmo di rendering della radiografia.

Analizziamo ora le unità di elaborazione:

- l'Engine 3D ha lo scopo di rappresentare la scena 3D; questa consiste nel posizionamento del paziente di fronte all'apparecchiatura radiogena in una postura che deve ricalcare quella definita dagli esperti del dominio come opportuna per il segmento anatomico che si vuole sottoporre ad indagine. L'unità deve disporre in input dei parametri che caratterizzano la scena 3D (fantoccio antropomorfo) nonché del modello dell'esame radiologico (tessuti interni del fantoccio);
- l'Engine 2D svolge il compito di visualizzazione dell'istogramma dei livelli di luminanza, della LUT di modifica della distribuzione, dell'anteprima dell'immagine in bassa risoluzione, che anticipa il risultato radiologico, nonché della stessa quando sia stata sottoposta all'applicazione di una LUT;
- il Controllo della Simulazione svolge il compito di acquisire ed interpretare i dati di input immessi dall'utente e pilotare di conseguenza la simulazione, aggiornando le visualizzazioni 2D e 3D; queste informazioni vengono inserite mediante l'utilizzo di una tastiera e di un joypad.

Simulazione di Esami Radiologici mediante Applicazione della Tecnologia CUDA

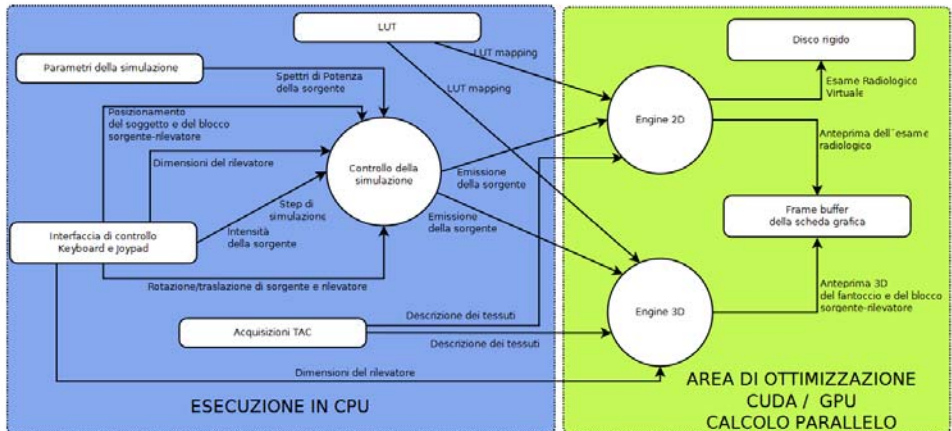


Figura 1 - Data Flow Diagram del sistema

L'aspetto dell'immagine radiografica usata per scopi diagnostici dipenderà sia dalle caratteristiche del paziente, cioè dalle iterazioni tra i raggi X e i diversi tessuti del paziente, sia dalle caratteristiche dei sistemi di rilevazione dei raggi X utilizzati.

Occorre ricordare alcune regole pratiche nell'acquisizione di un'immagine radiografica:

- per ottenere la minima dose al paziente e una buona qualità dell'immagine occorre usare il più piccolo valore di mA (diminuisce così la dose al paziente) e il più alto valore possibile keV (influisce sulla produzione di immagini con maggior dettaglio);
- si dovrebbe utilizzare la più stretta collimazione possibile, al fine di ridurre la radiazione diffusa ed evidenziare soltanto le parti rilevanti del paziente esposte ai raggi.

Da quanto detto è evidente che l'interazione all'interno del sistema virtuale che simuli un'apparecchiatura radiologica deve tener conto delle indicazioni di cui sopra e disporre di un fantoccio virtuale che fornisca i dati relativi ai tessuti oggetto della analisi; essi possono provenire da fonti quali immagini TAC di pazienti reali, da immagini di un fantoccio antropomorfo oppure da immagini che rappresentino un oggetto fisico le cui caratteristiche siano utili ai fini diagnostici.

L'utente del sistema di simulazione può variare a piacimento la posizione del blocco sorgente-rilevatore rispetto al paziente secondo i tre assi di rotazione, la distanza tra emettitore e lastra virtuale ed infine la distanza tra emettitore e paziente.

Anche i parametri di emissione della sorgente radiogena, quali lo spettro di radiazione ed i valori di controllo dell'esposizione, sono direttamente sotto il controllo dell'utente che può sperimentare diverse configurazioni dell'esame radiologico e verificarne l'efficacia in pochi istanti. Proprio questa immediatezza di risultati è l'aspetto più importante nell'esperienza proposta all'operatore di radiologia, che può estendere le sue capacità di analisi delle situazioni

diagnostiche senza coinvolgere soggetti reali. L'estrema velocità nella produzione dei risultati simulati consente di effettuare un numero di prove elevato, facilitando il convergere del percorso di affinamento alla ricerca dei parametri ottimali di esposizione del paziente.

3.La tecnologia CUDA

CUDA è un ambiente di sviluppo e programmazione che consente la scrittura e l'esecuzione di algoritmi che presentano un elevato livello di parallelismo. CUDA è stato sviluppato da NVIDIA [3] per essere eseguito sulle GPU (Graphics Processing Unit) delle sue schede video. Supporta alcuni dei principali linguaggi di programmazione come ad esempio C e MATLAB.

Una GPU è costituita da un elevato numero di core (unità di elaborazione in grado di effettuare operazioni semplici) con associata una memoria RAM ad alte prestazioni. Nel caso di problemi aventi una struttura Single Instruction Multiple Data (SIMD), in cui molti dati devono essere sottoposti alla stessa elaborazione, questi possono essere risolti utilizzando in parallelo le risorse di calcolo delle GPU al fine di ridurre drasticamente i tempi di calcolo.

4.L'applicazione

Seguendo il paradigma di calcolo parallelo suggerito da CUDA, l'applicazione inizialmente provvede a leggere il set di immagini provenienti dalla TAC, in cui ogni pixel fornisce l'indicazione della densità del voxel e quindi del tessuto rappresentato, e provvede a caricare la matrice tridimensionale del fantoccio all'interno della memoria della scheda video.

Il processo parte da una configurazione iniziale standard di posizionamento del blocco sorgente-rivelatore ed inizia a fornire all'operatore una anteprima della radiografia, utilizzando un numero ridotto di passi di scansione, al fine di ottenere una estrema velocità di visualizzazione (maggiore di 25 fotogrammi al secondo) per consentire una reale interattività con il sistema di posizionamento dello strumento radiografico. Per questa anteprima ci si avvale del calcolo attraverso la GPU, in cui ogni singolo pixel dell'anteprima viene calcolato autonomamente secondo lo schema di calcolo, detto kernel, inserito nella GPU. Il kernel sarà eseguito per ogni pixel e, sulla base del numero di core disponibili, ogni pixel sarà calcolato in autonomia da un core.

L'algoritmo all'interno del kernel consente di valutare la luminosità risultante di ogni pixel calcolata sulla retta che lo congiunge con la sorgente radiogena (ipotizzata puntiforme per semplicità). La retta viene suddivisa in K segmenti (step) e per ciascun segmento viene valutata l'intersezione con i voxel del modello tridimensionale, ciascuno dei quali determinerà una attenuazione del segnale in base alla densità; questa modalità di calcolo dell'immagine risultante richiama il metodo di ray-casting noto dalla computer graphics.

Una volta determinati dall'operatore tutti i parametri dell'esame, è possibile ottenere la radiografia simulata, alla risoluzione scelta che, data l'elevata risoluzione, sarà salvata all'interno di un file su disco per una successiva analisi con strumenti di lettura idonei (monitor ad alta risoluzione e dinamica estesa). Il

Simulazione di Esami Radiologici mediante Applicazione della Tecnologia CUDA
procedimento di calcolo della radiografia avverrà in modo simile a quanto descritto per l'anteprima, ma il numero di pixel sarà molto più elevato.

A seguire, a titolo di esempio, due screenshot dell'applicazione:

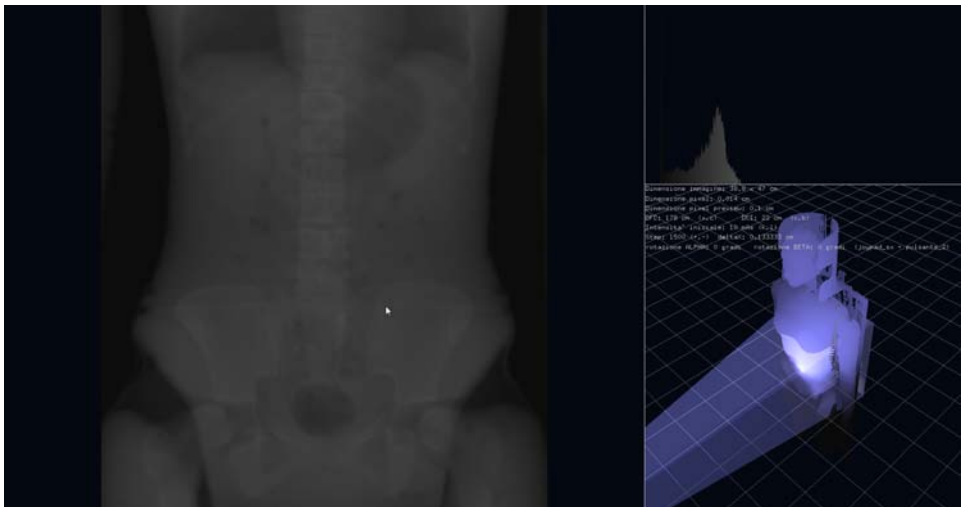


Figura 2 - Immagine frontale

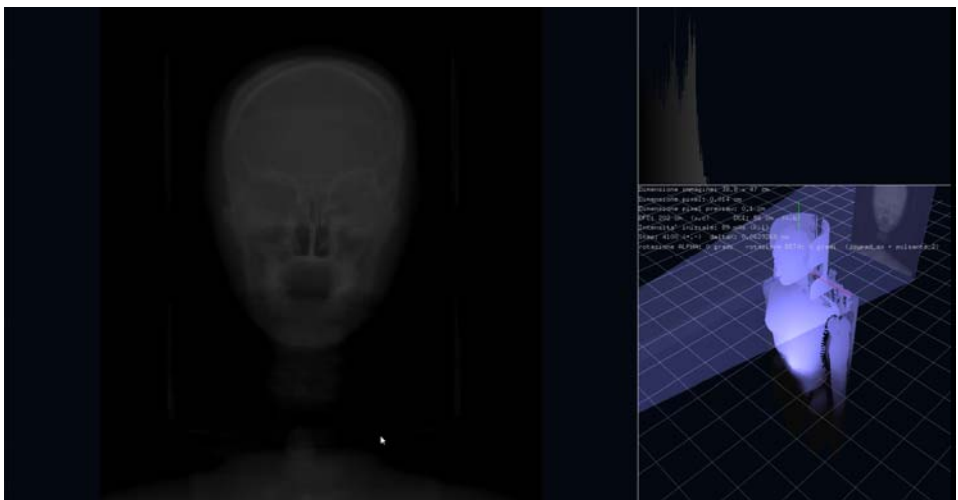


Figura 3 - Spostamento del blocco sorgente-rilevatore

L'applicazione è stata utilizzata su un calcolatore avente la seguente configurazione:

- NVIDIA GTX 460 (architettura Fermi 2.1) con 336 core CUDA
- Clock del bus della scheda grafica: 800 MHz
- Clock processori grafici 1600 MHz
- Memoria RAM scheda grafica 1024 Mbyte DDR5 con bus a 256 bit

- Processore di sistema: CPU Intel CoreDUO a 3600 MHz
- Memoria RAM sistema DDR2 con bus a 400 MHz

Le prestazioni sono state misurate caricando un fantoccio costituito da 400 immagini TC ad alta risoluzione (512×512 con dimensione del voxel di 0.001 cm^3) ed impostando la simulazione con i seguenti parametri:

- dimensione pixel anteprima = 0.1 cm
- dimensione pixel radiografia = 0.014 cm
- dimensione rivelatore $38.8 \times 47 \text{ cm}$
- numero step anteprima = 1500 passi
- numero step radiografia = 5000 passi

si ottengono i seguenti risultati:

- tempo elaborazione anteprima: 0.058 secondi
- tempo di elaborazione radiografia GPU: 7.5 secondi
- tempo elaborazione CPU: 2110 secondi (circa 35 minuti)

5. Conclusioni

Come i risultati analitici evidenziano, l'utilizzo delle architetture parallele basate su GPU consente di ottenere dei reali benefici nei tempi di esecuzione. Il modello fisico adottato è semplificato rispetto al caso reale al fine di rendere l'algoritmo di generazione dell'immagine radiografica facilmente parallelizzabile. L'evoluzione del modello dovrà considerare che nella realtà l'interazione dei raggi X con la materia è più complessa e passare dall'attuale paradigma deterministico ad una visione integrata con paradigmi di tipo probabilistico [4].

6. Ringraziamenti

Gli autori desiderano ringraziare il dott. Osvaldo Rampado per la collaborazione prestata durante lo sviluppo dell'applicazione.

7. Bibliografia e Sitografia

[1] Balossino N., Bianchi S. D., Lucenteforte M., Rabellino S., Rampado O., Ropolo R., Scapola G., "A software tool for the simulation of a digital X-ray imaging system", INTERNATIONAL JOURNAL OF COMPUTER ASSISTED RADIOLOGY AND SURGERY, Volume 3, Supplement 1, 2008, pages 351-352.

[2] <http://www.nvidia.com/cuda>, 2011.

[3] <http://www.nvidia.com>, 2011.

[4] Badal A., Badano A., "Fast Simulation of Radiographic Images Using a Monte Carlo X-Ray Transport Algorithm Implemented in CUDA", GPU Computing Gems Emerald Edition By Wen-mei W. Hwu, Elsevier 2011, pages 813-829.

SMART design: trasposizione efficace ed emulazione efficiente della didattica tradizionale

Luigi Scazzola, Amanda Reggiori¹, Marco Padula²

S.I.TE. - Supporto E-learning, Università Cattolica del Sacro Cuore - Sede di Brescia

Via Trieste 17, 25121 Brescia

luigi.scazzola@unicatt.it

¹FSLLS-Università Cattolica del Sacro Cuore – Sede di Brescia

Via Trieste 17, 25121 Brescia

amanda.reggiori@unicatt.it

²CNR - Istituto per le Tecnologie della Costruzione, US di Milano

Via Bassini 15, 20133 Milano

padula@itc.cnr.it

L'esperienza di e-learning design riferita ha come obiettivo la trasposizione di un corso universitario di poesia di lingua inglese e la realizzazione di un prototipo funzionante. Il corso segue una metodologia consolidata e, mediante il prototipo, è reso in versione da erogare con le tecnologie informatiche. Sono condizioni preliminari la fedeltà ai testi originali, il metodo didattico e di apprendimento, lo stile personale del docente e le convenzioni didattiche adottate.

1. Introduzione

L'applicazione delle tecnologie didattiche alle materie umanistiche è un ambito vasto, complesso e problematico. Normalmente un corso a distanza è progettato ed erogato basandosi sulle caratteristiche dello strumento tecnologico adottato, e su queste è modellata e adattata la didattica. Si ritiene che un approccio SMART in questo ambito significhi sostenere i docenti e i discenti nel trasporre il loro metodo didattico e di apprendimento, lo stile, il loro interrelarsi e interagire, l'uso e il confezionamento dei materiali, preservando, se desiderato, le proprie convenzioni, abitudini, caratteristiche personali, offrendo maggiore semplicità, efficacia e suggestività.

Le peculiarità degli utenti del Web 2.0, a volte viste come un disturbo alla didattica, possono rivelarsi interessanti risorse a cui il docente può fare riferimento, tenendo come obiettivo non l'applicazione del metodo ma il raggiungimento della costruzione di conoscenza. Si tratta di operare una trasposizione dall'insieme dei segni dell'esperienza didattica tradizionale a quelli delle tecnologie didattiche [Giacomantonio 2007], basato sull'organizzazione dei contenuti in **Learning Object** (LO), attraverso un

Learning Management System (LMS), archiviabili e riutilizzabili attraverso un **Learning Object Repository (LOR)**.

La scelta di uno standard per la costruzione e la descrizione dei moduli didattici influenza la scelta della piattaforma didattica e del sistema di archiviazione, e a sua volta il modulo didattico determina come e quali funzioni emulare con gli strumenti tecnologici, in questo caso la piattaforma MOODLE ver.1.7.7, il software di authoring ExeLearning 1.04, l'archivio digitale DOOR 1.8.0, adottando per i **LO** lo standard IMS.CP 1.1.3.

2. Il procedimento di trasposizione

E' necessario analizzare, formalizzare e schematizzare come sono utilizzati ed erogati i contenuti in una lezione tradizionale per progettare una modalità tecnologica, realizzando un'emulazione efficace. Figura 1 schematizza il metodo di lavoro proposto: realizzazione di un **modello ingenuo** della

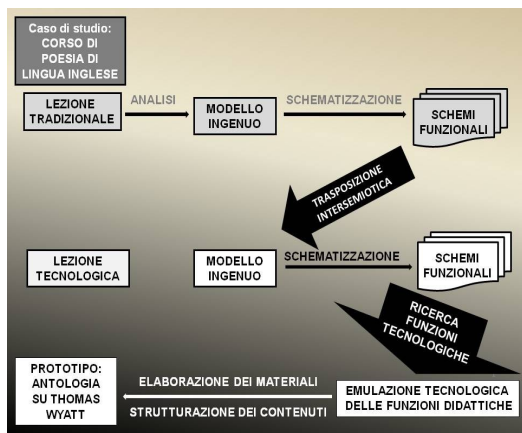


Figura 1: Schema di processo

lezione tradizionale e degli **schemi funzionali** relativi, **trasposizione intersemiotica**, realizzazione di un **modello ingenuo** della **modalità tecnologica** e infine di un **prototipo funzionante**. Modello ingenuo è una rappresentazione schematica in cui per raffigurare i processi sono utilizzate convenzioni grafiche intuitive. Gli attori e i dati sono indicati con icone, le azioni raffigurate in riquadri di tonalità diverse, contenenti gli eventuali sotto processi, uniti da frecce il cui verso definisce la direzione del

flusso dei dati stessi. Negli schemi funzionali il dettaglio descrittivo è aumentato fino a definire i sotto-processi, il flusso dei dati e l'interazione tra gli attori.

3. Trasposizione intersemiotica

Trasposizione intersemiotica in questo ambito è il processo di traduzione di una lezione dall'universo dei segni della didattica tradizionale all'universo digitale; esempi noti sono la traduzione dal libro al cinema, dalla pittura alla musica e, in genere, tutte le attività audiovisive. In [Jakobson 1959] sono distinti tre tipi di traduzione: **infra-linguistica** (all'interno della stessa lingua), **interlinguistica** (da una lingua naturale all'altra) e **intersemiotica** (fra sistemi semiotici diversi). Se si afferma che ogni traduzione è una trasformazione di contenuti e non di sistemi, è opportuno indicare come gli studi

in materia di interpretazione [Neergard 1995, Petrilli 2000, Eco 2000] sottolineino che la traduzione non è la costruzione di equivalenza semantica tra testi spostati da un sistema semiotico all'altro, ma un'interpretazione secondo i più puri dettami pierciani [Calabrese 2000] in cui un'espressione viene sostituita con un'altra che ne è l'interpretante. Di conseguenza si può definire la semiotica come disciplina che si occupa della comprensione dei testi e dello scambio cognitivo fra interpreti dei testi [Calabrese 2000], collegandosi in modo pragmatico alla formulazione di J. Malone [Malone 1988] che definisce nove strategie applicabili alla traduzione: **equation** (equivalenza), **substitution** (sostituzione), **amplification-reduction** (aumento-riduzione), **diffusion-condensation** (ampliamento-restrizione), **divergence** (divergenza), **convergence** (convergenza) e **reordering** (riorganizzazione).

Si riconosce a questa definizione una valenza interdisciplinare utile per descrivere le scelte operate di e-learning design, chiarendo quanto prodotto con gli schemi funzionali, e sottolineando i cambiamenti tra le due modalità con un confronto diretto. L'operazione semiotica è qui applicata ai metodi utilizzati per insegnare (o apprendere) contenuti di poesia in una lingua diversa dall'italiano, operando rispetto alla trasposizione anche una meta-traduzione. Il tratto linguistico e letterario dell'operazione resta di competenza del docente, compito specifico del tecnologo è analizzarne la modalità didattica cercando di descriverla e schematizzarla perché, attraverso un'emulazione efficace, ne vengano mantenuti gli obiettivi didattici.

Parte importante della costruzione del senso è affidata all'interazione degli studenti con i contenuti, secondo un approccio costruttivista [Calvani e Rotta 1999] che, utilizzando servizi forniti dal LMS e dal Web 2.0, si muove verso l'*heutagogy* [Hase 2000], concetto coniato da Stewart Hase, a indicare studio e apprendimento auto-determinato e riprende e sviluppa il concetto di *andragogy* (andragogia) teorizzato da Malcom Knowles [Knowles 1984]. L'andragogia, teoria unitaria dell'apprendimento ed educazione degli adulti volto a comprenderne diversità dei bisogni e interessi rispetto ai bambini, dirige l'attenzione verso le capacità esistenti del discente e le sue motivazioni a migliorare la propria condizione, da sfruttare e potenziare nel contempo dell'apprendimento del contenuto didattico. Nel modello di *heutagogy* si considerano in più i contesti di apprendimento formali e informali, accademici, ludici o non convenzionali, e quindi anche quelli di *Computer-Mediated Communication* (CMC). Confrontiamo le due metodologie per evidenziare cosa e come cambia nella didattica attraverso la trasposizione intersemiotica; il discorso sarà schematico ponendo in maiuscolo le azioni e in grassetto i componenti del modello ingenuo e degli schemi funzionali.

4. Lezione tradizionale nel processo di trasposizione intersemiotica

Il **modello ingenuo** di una lezione tradizionale è schematizzato nel seguito. Nella PREPARAZIONE DEL CORSO il docente definisce gli argomenti

e la **bibliografia** sulla base del programma che vuole svolgere. I dati iniziali possono essere il materiale delle **lezioni degli anni precedenti**, il proprio **archivio personale** o spunti presi da **pubblicazioni** proprie o altrui. La **bibliografia** viene integrata da testi difficilmente reperibili, traduzioni in italiano o quanto possa risultare utile per l'approfondimento per la PREPARAZIONE DEL MATERIALE SUSSIDIARIO. Il corso viene erogato attraverso una serie di **lezioni frontali**, a cui lo studente partecipa con l'ANNOTAZIONE APPUNTI; gli viene fornito **materiale sussidiario** in formato elettronico, scaricabile da un **archivio digitale**, oppure cartaceo, in forma di **dispensa** gestita dalla copisteria. Nella **lezione frontale** il docente presenta il **testo da esporre**, lo legge e analizza il contenuto spiegandone il significato. Lo studente appunta le spiegazioni e le risposte alle domande rivolte al docente e con queste azioni costituisce il materiale per lo STUDIO AUTONOMO con cui si prepara alla VALUTAZIONE DELLA PREPARAZIONE e della comprensione dell'argomento, spesso preceduto da una verifica scritta facoltativa. La comunicazione tra docente e studenti è assicurata dalla presenza alle lezioni oppure dal servizio e-mail. Il materiale cartaceo raccolto nella dispensa può anche venire rivisto alla fine del corso per eventuali pubblicazioni o altri corsi.

Lo **schema funzionale** relativo all'attività di PREPARAZIONE DEL CORSO e del MATERIALE SUSSIDIARIO raggruppa le seguenti attività: nella prima il docente può utilizzare i materiali definiti nello schema ingenuo, per la **definizione degli argomenti** del corso e la **scelta della bibliografia**; il materiale digitale delle **lezioni degli anni precedenti**, archiviate sul proprio PC o su memorie esterne di massa, può essere facilmente adattato alle esigenze contingenti mentre il **materiale cartaceo** di **archivio** o di recente valutazione, necessario per la seconda attività è più difficile da adattare. Può essere utilizzato sia per l'**assemblaggio della dispensa**, sia per la **preparazione delle slide** e opportunamente arricchito da **link** e **contenuti multimediali**; concorrerà all'**assemblaggio della lezione**. Alla fine del processo i dati che il docente utilizzerà sono di tipo eterogeneo; oltre ai **testi bibliografici** per la lezione in presenza c'è il **materiale digitale** da scaricare e quello cartaceo da fotocopiare. Una criticità di questo processo è la fissità del formato cartaceo che richiede supplementi o una completa riedizione della dispensa in caso si rendano necessarie aggiunte o modifiche. Anche gli spunti di approfondimento che il docente indica agli studenti durante la spiegazione perché vengano sviluppati in maniera personale possono diventare materia di formazione per tutti gli studenti solo se li rende disponibili on-line.

Uno schema funzionale della LEZIONE FRONTALE, APPUNTI e STUDIO AUTONOMO evidenzia che nell'erogazione della **lezione frontale** il docente presenta il **testo da esporre**, ne dà **lettura**; facendo l'**analisi del testo**, la **spiegazione dei vocaboli** e la **presentazione del contesto** fornendo le note interpretative necessarie alla comprensione del **commento**. Durante la spiegazione gli studenti possono porre al docente **domande**, a cui risponde direttamente oppure in un momento dedicato. Lo studente prende **appunti** su quanto viene indicato durante la lezione, annotando il **testo originale**

annotando i **riferimenti esplicativi** alle informazioni para-testuali e contestuali che utilizzerà in seguito. Nello **studio autonomo** queste indicazioni, oltre al **download di materiale digitale** e alla **dispensa** gli permettono la costituzione della **fonte di studio**. Il risultato viene infine valutato dal docente attraverso l'**esame**. L'eterogeneità del materiale non strutturato non facilita percorsi personalizzati né il sorgere di **ontologie**, che invece sono connaturate con una sua fruizione interattiva e personalizzata e sono, unitamente alla comprensione del testo, l'obiettivo didattico del corso.

5. Dal metodo al modulo didattico tecnologico

Le attività principali viste finora sono **Preparazione del corso**, **Preparazione del materiale sussidiario**, **Lezione frontale**, **Annotazione appunti** e **Studio autonomo**.

Si pongono alcune questioni: eterogeneità dei materiali presentati, fissità della forma cartacea e rigidità del contesto didattico. Un'emulazione efficace si pone l'obiettivo di produrre materiale standard, facilmente modificabile, ma completo di quanto è necessario per l'azione didattica. I problemi legati alla rigida sincronicità della lezione frontale come unica fonte di conoscenza (assenze degli studenti, problemi di attenzione, classi numerose, etc.) suggeriscono di andare verso un **apprendimento policronico** [Arina, 2008], situazione in cui, pur mantenendo la sua peculiarità didattica, la lezione frontale non è l'unico momento formativo, ma l'apprendimento cresce con lo svilupparsi di connessioni sociali tra gli studenti; apprendimento per **serendipity**, non orientato al raggiungimento di un determinato risultato, ma equamente sul valore della partecipazione a un processo.

I nativi digitali implicano l'adeguamento della didattica e degli strumenti di apprendimento, una sfida che l'emulazione efficace può raccogliere trovando nel concetto di heutagogy [Hase 2000] un punto di partenza importante e nell'approccio policronico [Arina, 2008] il terreno ideale per una declinazione SMART di una metodologia pragmatica.

Intercettare capacità pre-esistenti e sfruttare competenze tecnologiche extra-scolastiche è una necessità di tale approccio. Le possibilità offerte dalla tecnologia inducono a optare per una struttura del modulo didattico misto (**blended-learning**) che non trascuri le competenze e la professionalità del docente e assicuri grande flessibilità del percorso di apprendimento individuale, inserendolo in una forma di apprendimento collettivo (**cooperative learning**).

La **lezione tecnologica** applica i principi appena illustrati attraverso un modulo che utilizza un **LO**, e per le attività di approfondimento alcune applicazioni della piattaforma **LMS**, operando una scelta netta sulla granularità [Wiley et al, 2000 e Wiley et al, 2009]. Questa semplificazione non incide sulla flessibilità e sul carattere multimediale, in quanto in entrambi i casi sono perfettamente supportati ed erogati tutti i formati più utilizzati.

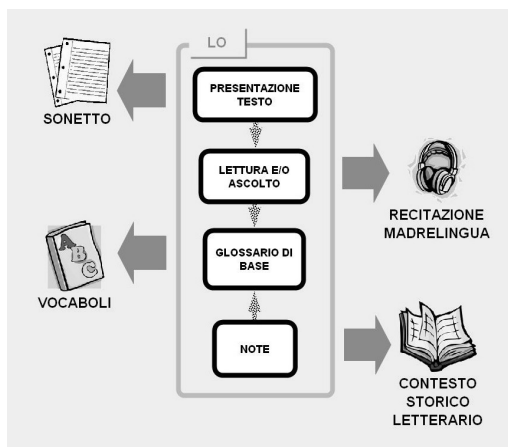


Figura 2: Struttura del LO

Ulteriore caratteristica è fornire oltre al **LO** di base anche traduzioni, versioni in italiano del testo, con caratteristiche diverse (storica autoriale o standard), ma sempre nella stessa forma; cioè singolo testo corredato da un glossario di base e note di contesto, e un file audio con la sua recitazione. La struttura di figura 2 è utilizzata sia dal docente nella lezione in presenza sia dallo studente nello studio autonomo, come attività propedeutica agli approfondimenti.

Un corso antologico prevederà quindi una serie di **LO**, ognuno relativo a un singolo testo poetico, modificabile aggiungendo testi di riferimento, anche in lingue diverse, sempre sotto forma di **LO**. Questo modulo didattico è concepito in modo da essere completato con l'interazione da parte degli studenti, sotto forma di partecipazione ad alcuni servizi della piattaforma, erogati nell'ottica del Social Network del Web 2.0, che verranno indicati e spiegati successivamente.

6. La lezione tecnologica dalla trasposizione intersemiotica

Confrontando il **modello ingenuo** della lezione tecnologica con quello della lezione tradizionale emergono già alcune differenze che si faranno più evidenti del dettaglio degli schemi funzionali successivi. Il formato del materiale è orientato verso il digitale e la **PREPARAZIONE DEL CORSO** si appoggia ad un **archivio digitale dei corsi** integrato il più possibile con la piattaforma LMS, che permetterà poi l'erogazione della **LEZIONE IN PRESENZA**, l'interazione degli studenti con i contenuti nello **STUDIO AUTONOMO** e alla fine del corso l'**elaborazione e l'archiviazione del materiale prodotto**. Nella **lezione in presenza** il docente accede al materiale per le lezioni e lo espone allo studente, il quale oltre a prendere appunti può anche accedere in modo asincrono sia a questo materiale che a quello **sussidiario** per lo **STUDIO AUTONOMO**. La **VALUTAZIONE** della **PREPARAZIONE** offre al docente, oltre alle forme tradizionali (**ESAME ORALE**), strumenti di analisi dell'**interazione didattica** degli studenti con i contenuti, e di strutturazione e **ELABORAZIONE** del **MATERIALE PRODOTTO** durante il corso. Si vuole sfruttare al massimo le caratteristiche di **riusabilità**, **modularità** e **compatibilità** con sistemi operativi e piattaforme diverse dei **LO**, per l'archiviazione sicura e condivisa dei **materiali didattici**, e di successive forme di **PUBBLICAZIONE**, sia digitale che cartacea.

Scendendo nel dettaglio della PREPARAZIONE DEL CORSO, nella DEFINIZIONE degli ARGOMENTI il docente può prendere spunti da fonti cartacee (**pubblicazioni o archivio personale**) ma ha a disposizione un vero archivio di materiale digitalizzato (**archivio LO e archivio corsi**) per la preparazione del corso. Con la **piattaforma LMS** li raccoglie nell'ambiente di produzione in cui crea lezioni e sezioni ad hoc, che andranno poi riempite di contenuti tramite la ricerca dei **LO** necessari, prima nell'archivio locale e successivamente con la RICERCA dei REPOSITORY ONLINE, importando direttamente il **LO**, editabile e completabile per adeguarlo alle esigenze del corso. Al bisogno il docente può generarlo con il software di authoring **ExeLearning** scelto per la sua semplicità d'uso e flessibilità, ma soprattutto considerando la struttura del **LO** progettato per questa emulazione (figura 2). Una volta completato e archiviato è possibile utilizzarlo per la costruzione del corso e delle singole lezioni. Il **materiale sussidiario** della lezione tradizionale è qui completamente strutturato, archiviato e gestito in forma digitale, attraverso un **wiki**, creato dal docente e modificabile in tempo reale a seconda delle esigenze puntuali del corso, a cui lo studente accede in modo autonomo. Segue la progettazione dell'**interazione degli studenti** con i contenuti realizzata attraverso altri servizi della piattaforma, (**forum e glossario condiviso**), attraverso i quali il docente indirizza e sollecita la partecipazione degli studenti alla costituzione dei contenuti del corso.

Nella LEZIONE IN PRESENZA della modalità tecnologica, il docente eroga attraverso la piattaforma la lezione presentando da un lato il **LO** relativo al testo con tutte le sue componenti multimediali (vedi figura 2), e dall'altro l'**interazione con i contenuti** richiesta allo studente nella fase di **studio autonomo**. Durante la **lezione in presenza** fa quindi riferimento ai contenuti del **LO** per la spiegazione di vocaboli o di note che lo studente chiede di approfondire e presenta il wiki **Fonti e riferimenti**, il **Glossario Condiviso** e il **Forum di Approfondimento**. Il dettaglio di questi strumenti è sviluppato nelle attività di STUDIO AUTONOMO e VALUTAZIONE PREPARAZIONE. Diversamente dalla **lezione tradizionale**, lo studente accede non solo a tutto il materiale che il docente progetta ed utilizza per la lezione, ma anche ai **contenuti** creati dall'**interazione degli studenti**, in maniera diversa dalla sincronicità della lezione in presenza. Questa modalità di assemblaggio e fruizione di materiali in modi e in tempi diversi richiede una particolare didattica indicata come **policronica**. Rispetto alla modalità tradizionale si tratta di trasposizione per **amplificazione** (amplification): il **testo** viene trasposto in forma multimediale (il **LO**) mentre il materiale sussidiario, con carattere **contestuale e co-testuale**, viene integrato in tempo reale nel corso secondo le esigenze del docente, la preparazione degli studenti e la loro partecipazione, rendendo possibili anche percorsi differenziati.

Durante lo STUDIO AUTONOMO lo studente rivede gli appunti attraverso l'accesso alla piattaforma e **interagisce con i contenuti**. Eventuali dubbi sul significato di un termine, sorti durante la rilettura del testo, rimandano al **Glossario interattivo condiviso**; lo studente può semplicemente leggere la

traduzione proposta e se insoddisfatto può commentarla o inserirne una alternativa, e per una ricerca più approfondita sul significato del vocabolo, il docente inserisce anche una sezione di link a **dizionari online**. Ogni inserimento personale viene registrato ed è immediatamente disponibile e quindi sottoposto al peer-review degli altri studenti. L'accesso al wiki **Fonti e riferimenti** è propedeutico all'interazione richiesta dal **Forum di approfondimento** in cui lo studente è stimolato a rielaborare e mettere in pratica quello che ha appreso in forma attiva. Il wiki contiene il materiale per la lezione in presenza e la definizione del contesto storico, letterario e culturale, inserito dal docente, modificabile integrabile e sviluppabile in tempo reale e in forma immediata, semplice e multimediale, adattabile non solo allo svolgersi del corso ma anche agli interessi degli studenti che possono usare il wiki per seguire un percorso di approfondimento personalizzato. Nel **Forum** lo studente dà il proprio contributo alle discussioni impostate dal docente, rielaborando non solo la propria conoscenza del testo ma anche quella del contesto, proponendo eventualmente una personale interpretazione dell'opera. Non esiste un equivalente attività così articolata nella **lezione tradizionale**; la **trasposizione intersemiotica** assume in questo caso un duplice carattere: sia di **reordering** (riordino) sia di **diffusion** (ampliamento).

La VALUTAZIONE della PREPARAZIONE dello studente prende in considerazione anche l'interazione con i contenuti del corso e i materiali prodotti in sinergia con gli altri studenti. Il docente può analizzare sia gli apporti personali al **Glossario condiviso**, sia entrare nel merito degli interventi, correggendo eventuali errori di traduzione. La lettura dei post degli studenti nelle discussioni del **Forum di approfondimento**, avendo carattere di risposta aperta, gli permette di valutarne l'espressione e l'elaborazione delle conoscenze acquisite. Può anche assegnare, in caso di scarsa partecipazione, l'elaborazione di ulteriore materiale in forma digitale, caricato in spazi dedicati, che valuterà direttamente, attraverso un esame orale; l'interazione con i contenuti offre quindi una ulteriore opzione di valutazione per il docente, che può sempre utilizzare modalità tradizionali, ma soprattutto vuole intercettare la produzione di conoscenza che lo studente elabora nell'ambito del corso.

L'ELABORAZIONE DEL MATERIALE PRODOTTO durante il corso è una delle attività peculiari della modalità tecnologica. Quanto inserito nel **Glossario condiviso** e nel **Forum di approfondimento** è già in formato digitale, rendendone l'analisi e l'archiviazione più veloce ed efficace. Nel primo caso la correzione finale del docente avviene dopo la **peer-review** degli altri studenti; è possibile manipolare il materiale, editando o cancellando le definizioni inserite, produrre glossari personalizzati, raccogliere tutti i termini che appartengono a un linguaggio di specialità ovvero a un singolo specifico tema, linguistico o letterario. Nel **Forum** le discussioni e i post degli studenti sono divisi per argomento e salvati ed utilizzati come spunti o documentazione per i corsi successivi, valutando i temi scelti e la capacità degli studenti di approfondire e produrre contenuti di conoscenza.

7. Conclusioni

Si evidenziano subito su due punti: il riuso dei LO come prassi consolidata e la strutturazione della conoscenza in chiave modulare, cooperativa e interattiva, cercando di intercettare la naturale inclinazione ad attività di **prosumption** [Toffler 1981] [Ritger e Jurgenson 2006] dei “nativi digitali” anche in ambito educativo. I materiali didattici per la preparazione del corso fanno ora parte di un circuito virtuoso di riuso, allargando la possibilità di articolare la proposta didattica e di approfondire gli argomenti correlati; l'attività attinge all'esperienza didattica condivisa dei corsi precedenti, a un archivio digitale di materiale già strutturato, modulare, espandibile e adattabile alle esigenze correnti. Lo studio autonomo si presenta come un'attività collaborativa che responsabilizza e chiama il discente a farsi componente attivo dell'azione didattica, cercando però di avvicinarsi alle sue modalità espressive e di socializzazione. L'elaborazione del materiale prodotto diventa quindi una fase estremamente importante del ciclo didattico in quanto è possibile non solo preparare in maniera adeguata per l'archiviazione e il conseguente riutilizzo il materiale predisposto per il corso dal docente, ma soprattutto elaborare il risultato dell'interazione degli studenti con i contenuti. Il livello quantitativo e qualitativo di questo materiale è inversamente proporzionale al numero degli studenti, ma basandosi su dinamiche di **social-networking** e il carattere di **prosumer** (anche inconsapevoli) è plausibile contare di raggiungere una massa critica di utenti/studenti che partecipano al **cooperative-learning**; si sottolinea comunque che l'unica competenza in grado di suscitare interesse per questa forma alternativa di didattica è quella del docente. Il prototipo, non ancora utilizzato in un corso accademico regolare, è stato testato con una classe di studenti di linguistica del 1° anno di Laurea Specialistica presso la Facoltà di Scienze Linguistiche e Letterature Straniere dell'Università Cattolica del Sacro Cuore di Brescia, con l'obiettivo di raccoglierne le impressioni di fronte alla simulazione del corso tradizionale di Letteratura Inglese, e allo svolgersi delle attività del corso erogato in modalità tecnologica. Gli studenti hanno mostrato interesse verso strumenti tecnologici che permettessero loro di organizzare e ottimizzare il materiale di studio, valutando positivamente la possibilità di essere valutati anche sulla capacità di elaborare i contenuti proposti nel corso producendo materiale didattico originale, ed efficace l'abitudine ad esprimersi in forme diverse dall'interrogazione orale. E' possibile, perlomeno, valutarne già le potenzialità in termini di articolazione del modulo didattico e in coerenza con i principi enunciati inizialmente: flessibilità verso situazioni didattiche in continuo cambiamento e setting para-didattici, obiettivi articolati e articolabili di costituzione del senso e produzione e fruizione della conoscenza attraverso la riusabilità di materiale didattico adattabile, modulare, strutturato e condiviso. Si ritiene, quindi, che una collaborazione sinergica tra docente e tecnologo permetterebbe l'individuazione di tutti gli strumenti disponibili per l'attuazione nell'immediato di e-learning realizzabile come trasposizione efficiente, efficace e innovativa della didattica tradizionale. E-learning, così concepito, per costituire una valida alternativa alla didattica universitaria indirizzata alla

generazione dei “nativi digitali” che prevede metodo didattico arricchito dalle tecnologie informatiche e multimediali, permanenza dello stile personale del docente, miscellanea di convenzioni didattiche tradizionali e nuove, costituzione di archivi privati e collettivi, condivisione interattiva di contributi, conoscenze e competenze sia top-down sia bottom-up.

Bibliografia

- [1] Arina T., Learning Zeitgeist: The Future of Education is Just-in-Time, Multidisciplinary, Experimental, Emergent, Master New Media, 13/2/ 2008.
- [2] Calabrese O., Lo strano caso dell'equivalenza imperfetta, in Dusi e Nergaard (2000).
- [3] Calvani A., Rotta M., Comunicazione e apprendimento in Internet. Didattica costruttivistica in rete, Edizioni Centro Studi Erickson, Trento, 1999.
- [4] Dusi, N., Nergaard, S., 2000, (a cura di), Sulla traduzione intersemiotica, Versus, nn. 85/86/87, 2000.
- [5] Eco, U., Traduzione e interpretazione, in Dusi e Nergaard (2000).
- [6] Giacomantonio M., Learning Object. Progettazione dei contenuti didattici per l'e-learning, Carrocci, 2007.
- [7] Hase, S., Kenyon C., From Andragogy to Heutagogy, Southern Cross University, 2000 <<http://ultibase.rmit.edu.au/Articles/dec00/hase2.htm>>
- [8] Jakobson, R., Saggi di linguistica generale, Milano, Feltrinelli, 1966.
- [9] Malone, J., The Science of Linguistics in the Art of Translation: Some Tools from Linguistics for the Analysis and Practice of Translation, SUNY Press, Albany, 1988.
- [10] Neergard, S. (a cura di), Teorie contemporanee della traduzione, Milano, Bompiani, 1995.
- [11] Petrilli S., «Traduzione e Traducibilità» in Gensini, S. (a cura di), Manuale della comunicazione, Carrocci, Roma, 2000.
- [12] Ritzer, G., Jurgenson N., Production, Consumption, Prosumption: The Nature of Capitalism in the Age of the Digital “Prosumer”, *Journal of Consumer Culture March 2010 10:13-36*, <<http://www.georgeritzer.com%2Fdocs%2FProduction%2520Consumption%2520Prosumption.pdf>>
- [13] Toffler, A., The Third Wave, Pan Books Ltd, London, 1981.
- [14] Wiley, D. A., Gibbons, A., Recker, M., A Reformulation of the Issue of Learning Object Granularity and its Implications for the Design of Learning Objects, 2000, <<http://www.reusability.org/granularity.pdf>>
- [15] Wiley, D. A., Connecting Learning Objects to Instructional Design Theory: A Definition, a Metaphor, and a Taxonomy, in Wiley, D. A. Wiley, The Instructional Use of Learning Objects: Online Version. December 2009, <<http://reusability.org/read/chapters/wiley.doc>>

Smart technologies per la gestione ambientale e paesaggistica: lo Scatol8® della sostenibilità

Riccardo Beltramo, Sergio Margarita¹

Dipartimento di Scienze Merceologiche, Università di Torino

Corso Unione Sovietica 218 bis, 10134 Torino TO

beltramo@econ.unito.it

¹*Dipartimento di Statistica e Matematica Applicata, Università di Torino*

Corso Unione Sovietica 218 bis, 10134 Torino TO

margarita@econ.unito.it

Il controllo degli aspetti ambientali significativi é un requisito per l'applicazione efficiente ed efficace dei sistemi di gestione ambientale, ma presenta aspetti critici che il sistema di telerilevamento SCATOL8® descritto in questo lavoro si propone di risolvere.

1. Introduzione

L'introduzione degli strumenti di gestione ambientale (SGA) ha portato a verificare le condizioni di applicabilità in realtà organizzative diversificate, mettendo in evidenza criticità nella rilevazione e nel controllo delle variabili ambientali significative. La rilevazione dei dati può interferire con l'operatività quotidiana, specialmente nelle piccole e medie imprese ed avvenire in modo irregolare, determinando così un quadro frammentato del profilo ambientale, al punto da mettere in discussione la possibilità di gestire la variabile ambientale su basi oggettive e di valutare l'efficienza e l'efficacia del SGA. Per contrastare tali difficoltà è stato messo a punto un sistema di telerilevamento di variabili ambientali e paesaggistiche, denominato Scatol8® che, basandosi su hardware e software Open Source, si propone come interfaccia tra l'organizzazione ed il sistema di gestione, automatizzando la raccolta e l'elaborazione dei dati.

2. La gestione ambientale e paesaggistica

La gestione ambientale è stata introdotta a livello internazionale attraverso il regolamento EMAS e le norme ISO14001 e riguarda tutte le organizzazioni economiche. Trattandosi di strumenti volti al miglioramento dell'efficienza e dell'efficacia nell'uso delle risorse ambientali ed economiche delle organizzazioni, sono state avviate attività sperimentali per verificarne l'applicabilità in vari contesti. Le attività realizzate all'interno del Dipartimento di Scienze merceologiche hanno coinvolto organizzazioni economiche di diverso genere ed hanno avuto l'obiettivo di verificare le condizioni di applicabilità dei SGA in realtà organizzative e ter-

ritoriali varie, dalle Piccole e Medie Imprese manifatturiere alle realtà turistico-ricettive e di verificare la possibilità di controllare altre variabili oltre a quelle strettamente ambientali.

Nel 2005 è stato formato un team multidisciplinare per affrontare l'estensione degli aspetti ambientali a quelli paesaggistici e per mettere a punto Linee-guida per la gestione ambiental-paesaggistica, utili per la progettazione e l'implementazione da parte di enti territoriali di un Sistema di Gestione Ambientale-Paesaggistico - SGAP [Beltramo et al, 2011]. La proposta metodologica di integrazione del Sistema di Gestione Ambientale ed Audit - EMAS con la Convenzione Europea del Paesaggio - CEP, è maturata all'interno del "Progetto per la realizzazione e la registrazione di un sistema di gestione ambiental-paesaggistica per l'Unione di Comuni Colline di Langa e del Barolo", finanziato nel periodo 2005-2009 dalla Regione Piemonte, Assessorato Ambiente.

Lo strumento EMAS prevede una struttura formalizzata per l'individuazione e la gestione degli aspetti ambientali significativi legati alle attività di un'organizzazione. Tale struttura, che si è dimostrata valida in termini di qualità ambientale, è stata utilizzata, con le necessarie variazioni, per la gestione degli aspetti paesaggistici significativi, anch'essi legati alle attività svolte dall'ente pubblico.

La possibilità di misurare la qualità ambientale e paesaggistica, e di conseguenza di controllarne l'evoluzione nel tempo, è condizionata dalla scelta di appropriati indicatori. Mentre la qualità ambientale è espressa da parametri chimici, fisici e biologici consolidati, quella paesaggistica risente della soggettività dell'osservatore. Nell'impostazione della ricerca si è operato per individuare indicatori oggettivi o a bassa soggettività, riferiti a variabili legate all'attività della Pubblica amministrazione e per integrarli nel modello gestionale previsto dal Regolamento EMAS. Questa attività è stata propedeutica al primo importante momento di integrazione di EMAS e CEP in SGAP ovvero l'Analisi Ambientale-Paesaggistica. Si tratta di un'indagine ampia e dettagliata sul contesto e sulle attività dell'ente in grado di influire su variabili ambientali e paesaggistiche, alla luce delle norme nazionali, regionali e comunali pertinenti.

In questa fase iniziale sono emerse notevoli carenze conoscitive dovute alla mancanza di sistematiche rilevazioni dei dati ed alla distribuzione dei dati tra soggetti diversi. Per gestire al meglio gli aspetti legati alla sostenibilità ambientale e paesaggistica, tuttavia, non è possibile prescindere dalla conoscenza diretta, precisa ed aggiornata degli aspetti ambiental-paesaggistici significativi della propria organizzazione, ma i metodi di rilevazione classici possono comportare un aggravio in termini di lavoro all'organizzazione, già occupata a trattare i molteplici aspetti dell'attività quotidiana. Simili considerazioni sono valide anche per le Piccole e Medie Imprese, infatti le sperimentazioni hanno messo in evidenza l'importanza del controllo cioè della misurazione delle variabili nel tempo, individuando allo stesso tempo in questa fase una criticità, in particolare nelle piccole realtà economiche. Da questa constatazione è scaturita l'idea di realizzare un dispositivo per il monitoraggio di variabili ambientali e paesaggistiche, funzionale allo svolgimento delle seguenti attività:

- reperimento dati;

- rappresentazione dei profili ambiental-paesaggistici;
- simulazione di scenari alternativi;
- condivisione della conoscenza;
- implementazione del sistema.

Il progetto che ne é derivato, denominato “Scatol8®”, si propone di realizzare un sistema di telerilevamento di variabili ambientali, paesaggistiche e gestionali ed ha avuto un impulso grazie al Progetto Interreg (2007-2013) “V.E.T.T.A., Valorizzazione delle Esperienze e dei prodotti Turistici Transfrontalieri delle medie e Alte quote”, nel quale é stato perfezionato e sperimentato in collaborazione con rifugi alpini. Queste strutture ricettive presentano aspetti ambientali riferibili alle più diffuse organizzazioni, ivi comprese le realtà domestiche, ed aspetti paesaggistici di eccellenza, che contribuiscono a renderli attrattori di un numero crescente di turisti.

La scelta del nome Scatol8® trasmette l'idea di un congegno amichevole, artigianale, facile da comprendere e da impiegare, dunque accessibile. Lo Scatol8® é alla base di un percorso (vedi Fig. 1), che si sviluppa con dispositivi hardware e software Open Source fortemente integrati con il Web.

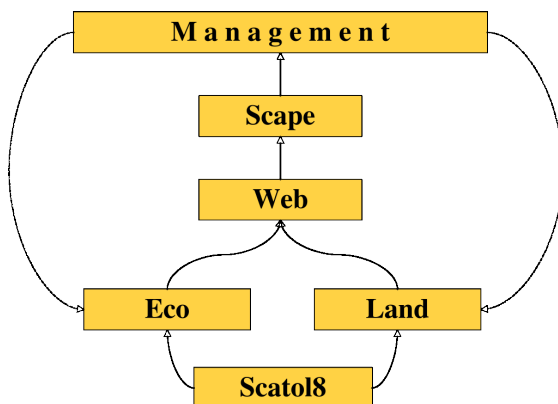


Fig. 1 - Lo Scatol8® e la gestione ambiental-paesaggistica

Il Web é la piattaforma che permette di visualizzare mappe di qualità ambiental-paesaggistica, di archiviare i dati permettendo lo studio delle dinamiche complesse, di condividere i risultati, di sensibilizzare, di formare e di gestire le trasformazioni, attraverso un'attività di comunicazione organizzata. Il paesaggio reale, le trasformazioni territoriali ed ambientali, danno luogo ad un paesaggio nell'internet, dove aspetti quantitativi, percettivi e gestionali vengono assemblati e fatti dialogare, favorendo un percorso condiviso per un miglioramento continuo, grazie all'impiego di modelli dinamici i quali, partendo dal reale portano allo sviluppo di scenari futuribili. Un corredo completo, fatto di strumenti di rilevazione, interfacciati con web attraverso software gestionali, e strumenti di modellizzazione dinamica, dà consistenza all'idea. Al termine del progetto si prevede

l'integrazione tra le variabili rilevate e gli strumenti di gestione volontari più diffusi, fino alla formalizzazione di un Rapporto di Sostenibilità (vedi Fig. 2).

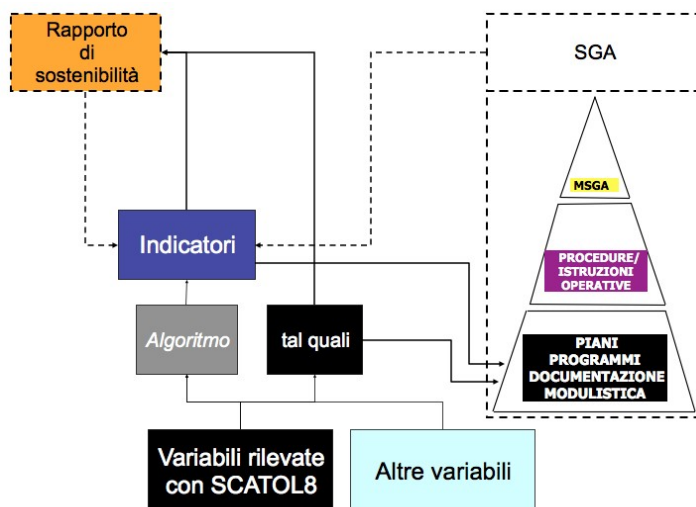


Fig. 2 - Scatol8® e Rapporto di sostenibilità

Lo Scatol8® consiste in una rete (*wired* o *wireless*) formata da microcontrollori e da una varietà di sensori remoti. Il sistema è modulare, può essere ulteriormente sviluppato e potenziato (aumentando il numero degli indicatori o la precisione di alcune misure), con un'incidenza di costo minimo. I dati raccolti sono memorizzati in un database residente su PC e/o essere inviati via Internet, al fine di consentire allo stesso tempo delle misurazioni *in situ* e da postazione remota, in considerazione dell'importanza di entrambi i tipi di osservazioni.

La decisione di orientarsi verso hardware e software Open Source è legata al requisito della accessibilità del sistema Scatol8® che si è voluto comunicare anche attraverso il reimpiego di imballaggi e di componenti di recupero.

Il sistema Scatol8® e l'adozione di codici Open Source sono legati anche ad altre due linee di attività sviluppate dal Dipartimento di Scienze Merceologiche dell'Università di Torino nell'ambito del Progetto V.E.T.T.A., sinergiche fra di loro: l'analisi dell'offerta turistica alle medie ed alte quote del territorio del Verbano Cusio Ossola (VCO), con una particolare attenzione ai rifugi alpini e la definizione di una Carta di Qualità per i rifugi alpini ed escursionistici.

2.1 Analisi dell'offerta

Si è operato secondo modalità distinte, in relazione ai turisti ed alle tipologie di strutture ricettive, per comprendere la percezione dell'offerta turistica da parte dell'ospite nel "territorio pilota", le vallate del VCO, attraverso la somministrazione di un questionario che ha interessato circa 400 intervistati nella stagione

estiva del 2010 ed altrettanti nel 2011, con domande riguardanti sia la struttura ospitante, sia la valle in cui la struttura stessa è localizzata. Inoltre, si è giunti alla valutazione della ricettività turistica del territorio e della sua importanza sul totale delle attività economiche, con approfondimenti sui rifugi alpini.

Nella stagione 2010 sono stati visitati 26 rifugi ed intervistati altrettanti Gestori attraverso una check-list tecnologica e gestionale contenente circa 150 domande, inserite in 5 differenti sezioni, riguardanti aspetti strutturali dell'involucro edilizio e della sua collocazione nell'ambiente, aspetti impiantistici (captazione e trattamento acqua, produzione di energia elettrica e termica, gestione dei rifiuti, ecc...) ed aspetti caratterizzanti le modalità di erogazione dell'offerta turistica (ad esempio l'utilizzo di prodotti alimentari tradizionali del territorio, l'offerta di attività a supporto della ricettività, ecc.). Le risposte e le considerazioni dei gestori sono state inserite in un database realizzato con software Open Source. Nell'estate 2011 l'attività è stata replicata su altre 26 strutture. Oggi la metodologia ideata viene proposta dalla Regione Piemonte agli altri partner, attraverso il progetto V.E.T.T.A., come standard per rilevare le dotazioni impiantistiche e le modalità gestionali dei Rifugi lungo l'arco alpino. Il medesimo database contiene anche le interviste svolte agli ospiti, associate alla struttura nella quale sono state rilevate, rendendo così possibili correlazioni tra efficacia degli impianti e percezione degli ospiti.

Le informazioni contenute nel database rivestono anche la funzione di indirizzare le politiche regionali verso il turismo sostenibile, suggerendo al legislatore interventi migliorativi della compatibilità ambientale e paesaggistica delle strutture. I dati rilevati permettono di formulare valutazioni costi e benefici, economiche ed ambientali, volte ad orientare politiche che, incidendo sull'efficienza delle strutture e degli impianti, elevino la qualità dell'offerta turistica.

2.2 Carta di qualità

Si è operato per definire uno strumento, la Carta di Qualità, in grado di qualificare le strutture sotto il profilo ambientale, per valorizzare il territorio e la cultura, e riavvicinare il turisti alla montagna. Le aree tematiche concernono aspetti legati all'accoglienza, all'impiego di prodotti locali, alle caratteristiche strutturali, di sicurezza ed igiene, di accessibilità e fruibilità e di sostenibilità ambientale. Il tema della Carta di Qualità, quindi, è legato all'analisi dell'offerta dei rifugi espressa nel punto precedente e si basa, in parte, su variabili ivi considerate e in parte a quelle rilevate con lo Scatol8®.

Nell'estate 2011, i criteri individuati per la Carta di Qualità sono stati adottati dall'ISNART per la formulazione del marchio "Ospitalità Italiana per rifugi escursionisti ed alpini" e sono state effettuate le prime visite per il rilascio del marchio a Rifugi del VCO, del Biellese e del Cuneese.

3. Il telerilevamento delle variabili ambientali

Le variabili ambientali da rilevare sono state individuate durante un'analisi ambientale sulla base delle varie relazioni tra l'organizzazione e l'ambiente:

- l'ambiente esterno, inteso come fornitore di materie prime, caratterizzato da parametri in grado di esprimerne la qualità (AE)
- il rapporto tra ambiente esterno ed interno, ovvero il prelievo delle risorse che alimentano l'organizzazione, divise per categorie e misurate in unità fisiche (AEI)
- l'ambiente interno ovvero le trasformazioni delle risorse in semilavorati e prodotti finiti da cui derivano rifiuti, scarichi liquidi, emissioni gassose, dispersioni energetiche da trattare prima dell'immissione nell'ambiente; (AI)
- il rapporto tra ambiente interno ed esterno ovvero l'immissione di inquinanti nell'ambiente (AIE).

Le variabili considerate dallo SCATOL8® sono elencate in Tabella 1:

AE	AEI	AI	AIE
Temperatura	Prodotti	Presenza fumo	Qualità fumi
Innevamento	Livello acqua	Qualità aria	Livello fossa
Precipitazioni	Livello gasolio	Consumo acqua	Qualità scarichi
Velocità del vento	Gas	Consumo gasolio	Quantità rifiuti
Direzione del vento	Energia elettrica	Temperatura	
Irraggiamento		Umidità	
Umidità			
Webcam			

Tab. 1 - Le variabili gestite dallo SCATOL8®

Lo sviluppo dei progetti è avvenuto prima in laboratorio e poi *in situ*. Per la rilevazione delle variabili, sono stati selezionati i sensori sulla base di una valutazione comparativa delle prestazioni. Successivamente, le attività hanno riguardato la realizzazione della rete nelle configurazioni *wired* e *wireless*, la connessione con il PC e la trasmissione dei dati via Internet al server centrale.

4. Hardware e software: la scelta Open Source

Poiché il sistema che si intende sviluppare presenta significative analogie con la domotica o home-automation, l'attività è iniziata con un'analisi dei principali sistemi disponibili sul mercato, per comprenderne l'adeguatezza in relazione alle esigenze progettuali. Questa indagine comparativa ha evidenziato che i sistemi disponibili soddisfano in modo parziale i requisiti in quanto:

- nella maggior parte dei casi, trascurano i parametri rilevanti per la gestione ambientale delle organizzazioni;

- le applicazioni di tipo domestico non prevedono controlli sugli approvvigionamenti e sulla gestione del magazzino, né sugli accessi, aspetti invece rilevanti per un'attività ricettiva;
- prevedono l'allacciamento permanente alla rete elettrica, sono basati su una tecnologia proprietaria che prevede nella quasi totalità dei casi l'interfacciamento con periferiche dello stesso produttore
- sono generalmente costosi.

Per ovviare il più possibile ai limiti emersi, l'interesse si è orientato verso sistemi Open Source per il *physical computing*: Parallax Basic Stamp, Netmedia's BX-24, Phidgets, MIT's Handyboard, Arduino ed altri che offrono funzionalità simili. Tutti sono accomunati dal fatto di racchiudere in un pacchetto *user friendly* gli aspetti tecnici complessi che riguardano la programmazione. La scelta si è indirizzata in un primo tempo verso Arduino per i noti vantaggi in termini di economicità, possibilità di funzionamento in ambiente Windows, Mac OS X e Linux, disponibilità di un ambiente di sviluppo semplice, estendibilità attraverso librerie software Open Source, hardware Open Source e Made in Italy.

5. Aspetti tecnologici

Il passaggio dalla fase prototipale a quella di test sul campo ha comportato problematiche difficilmente affrontabili con Arduino nella sua configurazione base. Fra i limiti incontrati citiamo, ad esempio, il numero di porte, la velocità di comunicazione tra microcontrollore e computer, la limitata capacità di memorizzazione locale. Tra le evoluzioni di Arduino ci siamo pertanto orientati verso Seeduino che aggiunge ad Arduino alcune funzionalità importanti, tra le quali la possibilità di immagazzinare i dati attraverso Flash memory presente sulla scheda del microprocessore. L'architettura del sistema è rappresentata nella Figura 3 che illustra l'articolazione del sistema di rilevamento.

6. L'azione sul campo

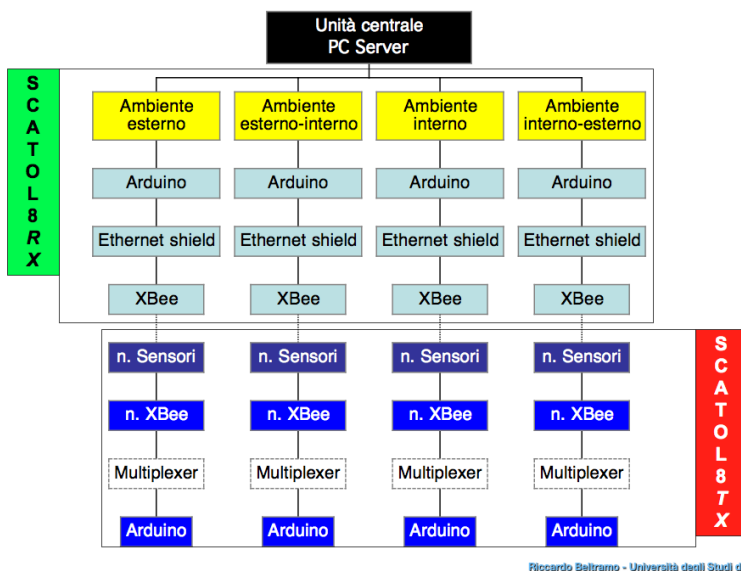
L'attività svolta presso il Dipartimento è stata orientata alla realizzazione di una rete wireless composta da diversi nodi periferici costituiti da microcontrollori, sensori, schede per la trasmissione dati e da un nodo centrale a cui vengono inviati i valori per la loro archiviazione. Scopo principale della fase di sperimentazione è stato di far emergere i problemi hardware e software in vista dell'installazione *in situ*.

Dal 1° Giugno al 31 Agosto 2011 le attività sul sistema di telerilevamento delle variabili ambientali sono entrate nel vivo, con l'installazione e la sperimentazione presso i rifugi dei nodi periferici, il loro collegamento al nodo centrale, la trasmissione dei dati da questo ad un PC ed infine la trasmissione via Internet.

Le problematiche riscontrate sul campo che non erano emerse in laboratorio sono relative ai problemi logistici di installazione, alla propagazione nel caso di comunicazione wireless (nonché la posa dei cavi di collegamento nelle installa-

zioni wired), alla stabilità e all'operatività dei sensori in caso di condizioni operative estreme.

Architettura del sistema - Wireless



Piccardo Beltramo - Università degli Studi di Torino

Fig. 3 - L'architettura del sistema di rilevamento

La scelta dei rifugi presso i quali svolgere gli esperimenti si è basata su quattro criteri:

- la connettività ad Internet;
- l'accessibilità;
- la rappresentatività in termini di variabili ambientali e categoria (escursionistico/alpinistico)
- le difficoltà oggettive legate ai materiali costituenti la struttura e alla distribuzione dei locali.

7. Criticità riscontrate e risultati ottenuti

Nel primo anno di attività del progetto, sulla base di una *check-list* tecnologico-gestionale è stato fatto un censimento di diversi rifugi che ha portato alla costruzione di un database. Sulla base di tali informazioni, sono state scelte tre strutture: il Rifugio Città di Novara, il Rifugio Andolla ed il Rifugio Castiglioni nei quali sono state condotte le esperienze descritte in questo lavoro. Si riportano (vedi Tab. 2) i sensori testati nei tre rifugi, indicando il nodo periferico, lo strumento o il parametro rilevato e la collocazione all'esterno (E) o all'interno (I) del rifugio.

Rifugio Città di Novara			Rifugio Andolla			Rifugio Castiglioni		
Nodo	Sensore	E/I	Nodo	Sensore	E/I	Nodo	Sensore	E/I
1	Livello liquidi o neve	E	1	Produzione solare	E	1	Intrusione	I
2	Anemometro	E	1	Luminosità	E	2	Consumo elettrico	I
2	Direzione vento	E	1	Temperatura	E	2	Gas mq2	I
2	Pluviometro	E	1	Umidità	E	3	Temperatura	I
3	Temperatura	I	2	Consumo elettrico	I	3	Temperatura	I
3	Umidità	I	2	Temperatura	I	3	Temperatura	I
3	Gas mq2	I	2	Umidità	I	4	Luminosità	I
4	Flusso liquidi	I	3	Temperatura	I	4	Temperatura	I
			3	Temperatura	I	4	Umidità	I
			3	Temperatura	I	5	Flusso liquidi	E
			4	Produzione rifiuti	I	5	Temperatura	E
			4	Gas mq2	I	5	Umidità	E
			4	Temperatura	I			
			4	Umidità	I			
			5	Flusso liquidi	I			

Tab. 2 - I sensori testati presso i rifugi

Gli aspetti che hanno presentato le maggiori criticità sono quelli legati alla trasmissione *wireless*, all'affidabilità delle comunicazioni fra i nodi nonché la necessità di ridondanza e di funzioni di sicurezza per gestire i possibili malfunzionamenti dei componenti. Per ridurre i problemi connessi a tali criticità, si è scelto una forte modularità del sistema, in particolare per la trasmissione dei dati, distribuendo le funzionalità di sicurezza e di ridondanza, aumentando così l'affidabilità del sistema nel suo complesso. In caso di non disponibilità del nodo centrale, i nodi periferici immagazzinano localmente le informazioni rilevate, in caso di non disponibilità del PC il nodo centrale opera nello stesso modo. A sua volta il PC invia le informazioni al server centrale in modo asincrono, a seconda della disponibilità del collegamento stesso a Internet. Sul server centrale è disponibile un cruscotto che, attingendo al database delle variabili rilevate, provvede alla loro rappresentazione grafica e alla segnalazione di eventuali anomalie o criticità.

8. Conclusione e futuri sviluppi

Lo SCATOL8® è oggi applicato per alimentare un archivio di registrazioni utili alla gestione ambientale- paesaggistica fornendo gli elementi utili per svolgere un'attività di sensibilizzazione alla sostenibilità. Le linee-guida seguite nello sviluppo del progetto - economicità, riuso di componenti di recupero, packaging di recupero - consentono di mettere l'accento sulla sostenibilità e formulare un'offerta di formazione e consulenza improntate alla sostenibilità stessa: sperimentazione, didattica e gestione come momenti di integrazione di professionalità diverse, orientate a divulgare e promuovere il concetto di sostenibilità.

Quando l'ambiente subisce trasformazioni in seguito all'intervento umano diventa territorio; quando il territorio viene osservato e filtrato dall'esperienza individuale, diventa paesaggio. Gli sviluppi futuri del progetto prevedono l'impiego dello SCATOL8® per monitorare e, nel limite del possibile, controllare le trasformazioni del paesaggio. Può quindi porsi come osservatore in grado di situarsi in punti differenti per abbracciare uno spazio ampio, ed operare così un monitoraggio a scala vasta, oppure uno spazio circoscritto, operando un i monitoraggio micro.

Più specificatamente per gli aspetti paesaggistici, gestendo altri dispositivi quali macchina fotografica e webcam ed integrandosi con sistemi di elaborazione, lo SCATOL8®:

- interviene nella fase di riprese fotografiche, fornendo le coordinate geografiche, l'altitudine nonché controllando l'orientamento della macchina fotografica, per collocare le riprese su mappe;
- interviene nella fase di monitoraggio del territorio, prelevando periodicamente da Google Earth mappe del territorio di interesse o consentendo una navigazione facilitata da interfacce, ad esempio joystick, per vedute paesaggistiche, monitoraggio di eventi, supervisione;
- presidia lo scambio di file tra applicazioni, interfacciandosi con sistemi Open GIS
- gioca un ruolo nella sorveglianza delle situazioni di emergenza reale o potenziale (terreni soggetti a frane, ecc...), associando posizione, ora di rilevamento, immagine e dati rilevati da sensori.li.

9. Riferimenti bibliografici

[Belramo, 2010] The SCATOL8™: an innovation for shifting from Environmental and Landscape Management System (ELMS) to the Eco-Land-Web-Scape Management System (ELWSMS), Romanian Distribution Committee Magazine, 2, 2, 2010, 16-23, <http://www.distribution-magazine.ro/magazine2/>

[Belramo et al, 2011] Belramo R., Duglio S., Quarta M., "Sistema di Gestione Ambientale-Paesaggistico - SGAP: Una metodologia per la gestione integrata dell'Ambiente e del Paesaggio", Aracne Editrice, Roma, 2011

Stack Open-Source & Real-Time

Stefano Scanzio, Manuel Cheminod, Claudio Zunino

CNR-IEIIT

C.so Duca degli Abruzzi, 24; I-10129 Torino (Italy)

E-mail: {stefano.scanzio, manuel.cheminod, claudio.zunino}@ieiit.cnr.it

Personal computer, dotati di sistemi operativi real-time ed open-source, stanno progressivamente sostituendo i sistemi special-purpose nelle applicazioni di controllo in cui il requisito principale è un elevato determinismo. Dall'altro lato, protocolli di rete real-time basati su Ethernet, hanno praticamente rimpiazzato le vecchie tecnologie proprietarie, rendendo le soluzioni open-source basate sull'utilizzo di comuni personal computer convenienti da un punto di vista economico per la realizzazione di applicazioni di controllo distribuito. Due importanti stack protocollari, quello convenzionale di Linux e RTnet, sono stati confrontati dal punto di vista della latenza di trasmissione. In particolare, una maggior rilevanza è stata data all'analisi del ritardo introdotto dai singoli nodi, stack protocollare e scheduler, piuttosto che alla latenza introdotta dalla rete. I risultati hanno confermato come RTnet sia in grado di soddisfare i requisiti temporali della maggior parte delle applicazioni di controllo industriale.

1. Introduzione

L'utilizzo di oggetti prodotti in serie è sempre più comune nelle applicazioni di controllo distribuito. In particolare, dispositivi di controllo basati su piattaforme standard, come ad esempio i personal computer (PC) tradizionali e scheduler real-time come RTAI [Mantegazza et al, 2000], Xenomai [Xenomai, 2011] o RT Patch [Rostedt e Hart, 2007], possono essere adottati come soluzioni poco costose caratterizzate da un ottimo determinismo. In letteratura sono presenti diversi articoli che parlano di PC convenzionali utilizzati proficuamente per applicazioni di controllo real-time in ambito industriale. Ad esempio in [Chiandone et al, 2008] è descritto un sistema di controllo *hard real-time* utilizzato per il *motion-control*. Similmente, in [Ferrarini et al, 2003], un PC convenzionale è stato utilizzato per controllare un'applicazione industriale, sostituendo soluzioni hardware e software proprietarie. In [Cereia et al, 2011], invece, RTAI e RT Patch sono stati confrontati da un punto di vista prestazionale.

Questa tendenza è stata resa possibile dalla presenza di protocolli industriali ad elevato determinismo basati su Ethernet, comunemente conosciuti come *real-time Ethernet* [Decotignie, 2009]. Essi hanno permesso di utilizzare le

normali schede di rete Ethernet per il controllo industriale, sostituendo quasi completamente le soluzioni basate su hardware proprietario.

Il modo più frequente per far comunicare delle applicazioni in sistemi operativi di tipo Linux è attraverso l'interfaccia socket. Nel recente passato diverse pile protocollari (stack) open-source sono state implementate al fine di fornire un supporto alle comunicazioni real-time fornendo a livello applicativo un'interfaccia di tipo socket.

La capacità di assicurare un comportamento deterministico nella comunicazione tra PC interconnessi tramite interfacce Ethernet è stato un argomento molto importante ed investigato negli ultimi anni.

In [Balaji et al, 2005] alcune tecnologie basate sullo stack *Linux* sono state valutate, mentre in [Hou et al, 2005] è stata descritta un'implementazione dei socket eseguita a livello kernel. Lo stack protocollare *RTnet*, invece, è stato valutato in [Zhou et al, 2007] e in [Hanssen et al, 2005]. In questi articoli maggior enfasi è stata data alla comunicazione di rete ed in particolare è stata analizzata la tecnica TDMA usata da *RTnet*. Gli errori di quantizzazione dovuti a TDMA hanno reso impossibile, in questi lavori, la valutazione della latenza interna ai nodi di trasmissione e ricezione.

Questo contributo, a differenza degli articoli reperibili in letteratura, è focalizzato sulla latenza e sull'indeterminismo introdotto dal software e hardware locale ai nodi e non sulla connessione tra i nodi. Per questo motivo i nodi sono stati connessi nel modo più semplice possibile, un cavo di rete punto a punto, al fine ridurre al massimo l'indeterminismo introdotto dal mezzo di comunicazione. In particolare il contributo si focalizza sullo studio delle latenze e dell'indeterminismo introdotto dal software in una piattaforma *real-time* basata su RTAI. Due stack protocollari sono stati confrontati, quello standard di Linux (*Linux*) e *RTnet* [Kiszka e Wagner, 2005] (*RTnet*).

RTAI e *RTnet* saranno descritti nella prossima Sezione del contributo. In Sezione 3 sarà illustrata la configurazione sperimentale utilizzata nell'analisi. Infine, in Sezione 4, saranno riportati i risultati sperimentali ottenuti in differenti condizioni operative.

2. RTAI e RTnet

RTAI è un progetto realizzato dal Politecnico di Milano nel 1996 con lo scopo di realizzare un'estensione *hard real-time* del kernel di Linux. Esso permette di realizzare applicazioni *hard real-time* su computer standard, programmate a livello utente, fornendo un'interfaccia di alto livello.

RTAI è composto da una serie di moduli sviluppati per abilitare le funzionalità real-time. Esso, attraverso un livello di astrazione (RTHAL/ADEOS), intercetta gli interrupt e li smista al kernel Linux o allo scheduler RTAI, dando agli interrupt gestiti da quest'ultimo la priorità massima. RTAI agisce come uno scheduler di processi *hard real-time* o del kernel di Linux stesso. Ai processi *hard real-time*, come nel caso degli interrupt, viene data maggiore priorità. Al fine di installare RTAI, il kernel di Linux deve essere modificato applicando opportune patch fornite dagli sviluppatori.

RTnet è una stack protocollare *hard real-time*. Esso usa RTAI o Xenomai come scheduler. I driver delle schede di rete sono stati riscritti dagli sviluppatori utilizzando un'interfaccia, Real-Time Driver Model [Kiszka, 2005] (RTDM), fornita appositamente da RTAI e Xenomai per lo sviluppo di driver *hard real-time*. RTnet supporta una dozzina di tipi di schede di rete ed implementa i più popolari protocolli di comunicazione (UDP/IP, TCP/IP, ICMP, ecc.) ma in modo deterministico, cioè con una latenza nello spedire o ricevere i pacchetti limitata superiormente. Esso fornisce agli sviluppatori un'interfaccia aderente allo standard POSIX.

RTnet permette, attraverso un livello denominato RTmac, di realizzare un partizionamento temporale del mezzo trasmissivo per evitare indeterminismi dovuti agli accodamenti sugli switch. Tale funzionalità, nota in letteratura con il nome Time Division Multiple Access (TDMA), è stata disattivata in quanto la configurazione sperimentale utilizzata in questo contributo prevede un collegamento punto a punto.

3. Configurazione sperimentale

L'esperimento è composto da due PC identici connessi tramite un cavo di rete Ethernet di tipo *crossover*. Un processo, eseguito sul nodo *produttore* (P), produce ciclicamente, con periodo 500 ms, dei pacchetti di tipo UDP destinati ad un nodo *consumatore* (C). I pacchetti UDP hanno un payload di dimensione 230 ottetti.

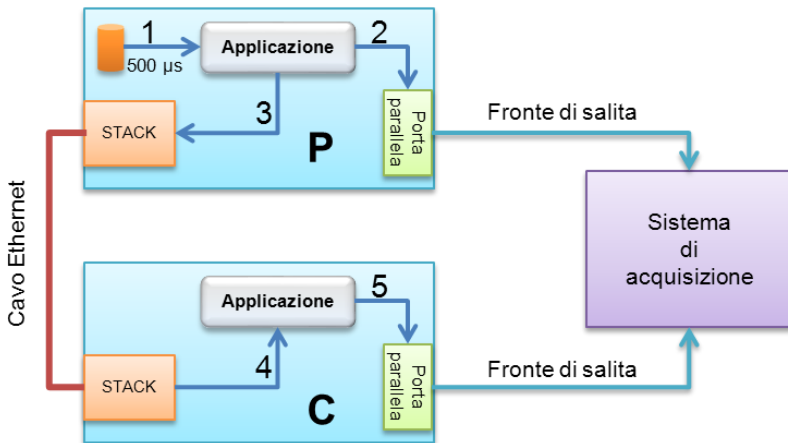


Fig. 1 – Configurazione sperimentale utilizzata

Il programma realizzato per il nodo P, appena rilasciato dallo scheduler RTAI, alza un segnale sulla porta parallela del PC e spedisce il pacchetto UDP usando, in dipendenza dal tipo di esperimento, gli stack protocollari *Linux* o *RTnet*. Il programma di test eseguito sul nodo C, passivamente bloccato sulla funzione di ricezione *recvfrom()*, viene sbloccato dall'arrivo del pacchetto UDP ad esso destinato. Appena sbloccato, viene generato un fronte di salita sulla

porta parallela del nodo C. Entrambi i processi, eseguiti sui nodi P e C, sono realizzati in un contesto *hard real-time* usando le specifiche interfacce di programmazione fornite dallo scheduler RTAI. Opportune funzionalità di RTAI sono state utilizzate per la temporizzazione della spedizione ciclica dei pacchetti eseguita dal nodo P. Nel caso di utilizzo dello stack *RTnet*, tutto il processo di spedizione/ricezione del pacchetto è eseguito in contesto *hard real-time*. Usando lo stack *Linux*, invece, vengono eseguite delle *system call* le quali permettono ad altri task di interferire con l'applicazione dei test. In tal caso il processo non è perciò più caratterizzabile come *hard real-time*. Un sistema di acquisizione esegue quindi i *timestamp* dei fronti di salita dei due segnali generati sulle porte parallele. Al fine di ridurre il più possibile gli errori casuali e di ottenere risultati statisticamente significativi, ogni misura è stata ripetuta 15000 volte.

Ogni campione, ottenuto sperimentalmente, è stato depurato dalla latenza di rete, cioè dal tempo impiegato dalla rete stessa per trasmettere un pacchetto dal nodo P al nodo S. Il tempo di rete T_R è la somma di tre contributi:

$$T_R = T_A + T_{PA} + T_{PR}$$

dove T_A è il tempo impiegato dal meccanismo MAC per ottenere l'accesso al mezzo trasmissivo, T_{PA} è il tempo impiegato dalla rete per trasmettere la sequenza di bit che compongono il pacchetto e T_{PR} è il tempo di propagazione del segnale nel mezzo trasmissivo. Per i collegamenti di tipo *full-duplex* $T_A=0$. Un pacchetto UDP con payload pari a 230 ottetti, tenendo conto degli header protocollari, sul cavo provoca la trasmissione di 284 ottetti, cioè 2271 bit. Se la velocità della rete è pari a 100 Mb/s, il pacchetto sarà trasmesso in 22.72 μ s. Dunque $T_{PA}=22.72 \mu$ s. La velocità di propagazione di un segnale elettrico in un cavo è ~ 5 ns/m. Nei nostri esperimenti tale quantità è trascurabile in quanto è pari a ~ 10 ns. In definitiva il ritardo T_R introdotto dalla rete, è pari a $\sim 22.72 \mu$ s. Siccome in questo contributo l'attenzione è focalizzata sul ritardo introdotto internamente ai nodi di trasmissione e di ricezione, tutti i risultati non comprenderanno il tempo di rete T_R , il quale sarà sottratto da tutte le misurazioni ottenute sperimentalmente.

Sull'insieme di campioni ottenuti, da cui è stato sottratto il ritardo di rete, sono stati calcolati i più importanti indici statistici che sono il valor medio μ (miglior stima del dato reale), la deviazione standard σ (indice della dispersione dei campioni rispetto alla media, caratterizza il determinismo dello stack protocollare) e i valori minimo Min e massimo Max (utili per analizzare il comportamento *real-time* dello stack).

I due computer su cui sono stati eseguiti gli esperimenti sono equipaggiati da un processore AMD Athlon™64 3500+ con frequenza di funzionamento 2.2 GHz. La CPU ha due cache a 4 vie di livello 1 di dimensione 64 KB, una per le istruzioni e l'altra per i dati, e una cache a 4 vie di livello 2 di dimensione 512 KB che viene utilizzata sia per le istruzioni che per i dati. Il sistema è equipaggiato con due banchi di memoria da 512 MB di tipo DDR funzionanti a frequenza 333 MHz. La periferica di comunicazione è una scheda di rete della

Realtek Semiconductors, con chipset RTL8139D. La versione di RTAI è 3.7.1 e il kernel su cui è stato installato ha versione 2.6.24.7. La versione dello stack protocollare *RTnet* è 0.9.12. Esso è stato calibrato con gli opportuni strumenti di calibrazione forniti nel pacchetto d'installazione.

Il *timestamp* sul fronte di salita è stato effettuato con il terminale EL1252 Beckhoff digital input [Beckhoff, 2011]. Per eseguire il *timestamp* tale dispositivo usa una sorgente interna dei tempi di tipo *free running*, la cui risoluzione è 1 ns e la precisione è 10 ns. La differenza tra le latenze d'ingresso delle due porte del dispositivo EL1252 sono state analizzate sperimentalmente ottenendo una discrepanza compresa nell'intervallo [-10,+10] ns. Tale errore è trascurabile in quanto è diversi ordini di grandezza inferiore all'indeterminismo atteso. Il terminale EL1252, così come il protocollo *real-time Ethernet* utilizzato dal dispositivo e il sistema di sincronizzazione sono descritti in [Cena et al, 2010]. Siccome la soglia di 1 logico per il terminale EL1252 è +11 V (e il suo intervallo di funzionamento è 0÷24 V), il segnale è stato amplificato mediante un semplice circuito composto da un transistor (NPN C547C) e da due resistenze tarate per convertire il segnale da 0÷5 V (intervallo tipico di funzionamento della porta parallela) a 0÷24 V. Un'applicazione è stata realizzata per acquisire i campioni ottenuti dal terminale EL1252.

4. Risultati sperimentali

Tre tipologie di esperimenti sono stati effettuati al fine di caratterizzare e confrontare gli stack protocollari *Linux* e *RTnet*, in differenti condizioni di utilizzo. In sottosezione 4.1, il determinismo dei due stack protocollari è stato valutato con tre differenti tipi di carichi interferenti, in sottosezione 4.2 è stato analizzato l'impatto del periodo di generazione dei pacchetti sulla latenza media, infine, in sottosezione 4.3, è stato valutato l'indeterminismo causato dal carico di rete.

4.1 Carichi interferenti

In questo primo set di esperimenti le prestazioni dei due stack protocollari sono state valutate a sistema scarico e con tre tipi di carichi interferenti.

I tre tipi di carichi interferenti sono stati scelti per avere una copertura il più possibile completa e verosimile dei casi reali. Essi sono:

- CPU: un ciclo infinito (*while(1);*), utile per analizzare il tempo di *cambio contesto* dello scheduler
- MEM: un programma che esegue calcoli su grosse moli di dati. Tale carico è importante perché coinvolge il bus di sistema, in cui tipicamente la contesa per l'accesso è arbitrata in modo hardware
- I/O: trasferimento di una grossa quantità di dati verso l'harddisk. Questo carico, oltre ad utilizzare il bus di sistema, genera una grossa quantità di *interrupt*. Gli *interrupt handler*, essendo eseguiti in contesto *kernel*, hanno priorità maggiore rispetto ad ogni processo eseguito in contesto *utente*

Stack	Linux				RTnet			
Tipo di carico	No	CPU	MEM	I/O	No	CPU	MEM	I/O
Media (μ s)	65.8	65.9	124.3	118.1	70.0	68.5	94.5	91.1
Std (μ s)	11.2	28.1	39.0	35.2	6.0	5.2	4.2	4.0
Minimo (μ s)	48.5	52.1	90.1	73.7	59.9	59.7	80.9	79.2
Massimo (μ s)	138.1	1194.8	2673.7	2028.7	108.0	110.5	133.8	124.7

Tabella 1 - Comparazione tra lo stack protocollare di Linux e RTnet in differenti condizioni di carico

Dai risultati riportati in Tabella 1 si può notare come le prestazioni dei due stack, nel caso di sistema scarico, siano paragonabili. *RTnet* è leggermente più deterministico, come testimoniato da una minor deviazione standard sia con sistema scarico (6.0 contro 11.2 μ s), ma soprattutto con sistema carico (4.2 μ s contro 39.0 μ s con carico MEM). La *latenza media*, impiegata da *RTnet* per spedire un pacchetto tra *P* e *C*, è leggermente maggiore rispetto a *Linux*. Infatti, in *RTnet* molte strutture di dati ed algoritmi sono stati ottimizzati per garantire determinismo, chiaramente a discapito delle prestazioni. Già con il solo carico di CPU, quello potenzialmente meno devastante dal punto di vista del *real-time*, appare evidente come lo stack di *Linux* non garantisca prestazioni *hard real-time* in quanto il tempo di esecuzione non è superiormente limitato (massimo di circa 1.2 ms per il carico CPU e di circa 2.7 ms per il carico MEM). Come previsto, lo stack protocollare di *Linux* non permette prestazioni adeguate per il *real-time*, sia da un punto di vista *hard* (la latenza di trasmissione non è limitata superiormente), sia da un punto di vista *soft* (la deviazione standard è troppo elevata per molte applicazioni reali). Per i carichi di tipo MEM e I/O si può notare un incremento della *latenza media* di circa 60 μ s per *Linux* e 25 μ s per *RTnet* rispetto ai casi scarico e CPU. Tale fenomeno è dovuto alla cache della CPU, la quale viene svuotata dai carichi di tipo MEM e I/O. Quando il processo di spedizione pacchetto viene rilasciato nel nodo *P* da RTAI, o un pacchetto è ricevuto dal nodo *C*, il processo di test dovrà essere rimesso in cache prima di essere eseguito, con una non trascurabile perdita di tempo. Tale fenomeno sarà analizzato in dettaglio nella seguente sottosezione.

4.2 Effetti del periodo di generazione

In questa sottosezione verranno valutati gli effetti del periodo di generazione dei pacchetti sulla *latenza media*, facendo variare il periodo di generazione da 5 ms a 8 s. Come si può osservare dai risultati riportati in Fig. 2 la *latenza media* è proporzionale al periodo di generazione per entrambi gli stack protocollari.

Questo comportamento, inatteso, non può essere dovuto a funzionalità attive di *risparmio energetico* per due principali motivi:

- tutte le funzionalità di *risparmio energetico* sono state disattivate sia dal bios dei PC, sia in fase di compilazione del kernel
- solitamente le funzionalità di *risparmio energetico* sono attivate da delle soglie temporali, cioè dopo un determinato periodo di inattività un componente entra in uno stato di risparmio energetico. Questa spiegazione non è coerente con i dati sperimentali riportati nel grafico di

Fig. 2 in cui chiaramente si vede una proporzionalità tra il periodo di generazione dei pacchetti e la *latenza media*

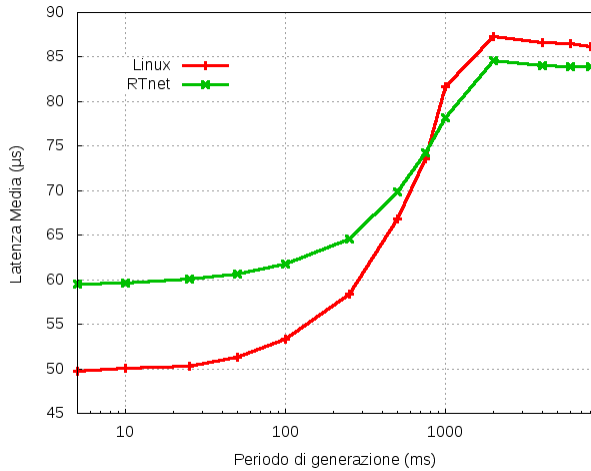


Fig. 2 - Comparazione tra lo stack protocollare di Linux e RTnet con differenti periodi di generazione dei pacchetti

La causa di questo comportamento è la cache della CPU. Inaspettatamente, l'indeterminismo causato dalla cache della CPU sulla *latenza media* è molto grosso e pari a $\sim 35 \mu\text{s}$ per lo stack *Linux* e $\sim 25 \mu\text{s}$ per *RTnet*. Analizzando il grafico di Fig. 2 si osserva che, per periodi di generazione inferiori a 50 ms, il programma di test, e i relativi thread del kernel usati dello stack protocollare ed adibiti al controllo della scheda di rete, risiedono quasi completamente in cache. Siccome il tempo che intercorre tra l'invio di due pacchetti successivi è ridotto, statisticamente pochi processi interferenti potranno essere eseguiti in quel lasso di tempo e perciò, solo una minima parte del programma di test, sarà eliminata dalla cache. Tra 50 ms e 1 s si ha una situazione intermedia in cui un non trascurabile numero di processi interferenti rimpiazza *entry* della cache precedentemente utilizzate dal programma di test. In questa situazione, quando il programma viene rieseguito dalla CPU, il numero di *cache miss*, non trascurabile, sarà proporzionale al periodo di generazione dei pacchetti. Infine, con periodi maggiori di 1 s, praticamente tutto il programma di test e i relativi thread del kernel sono fuori dalla cache. In tal caso il numero di *cache miss* è costante e la *latenza media* è massima e pari a $\sim 86 \mu\text{s}$ per *Linux* e $\sim 84 \mu\text{s}$ per *RTnet*.

4.3 Carico di rete

In questo set di esperimenti si intende valutare l'interferenza del traffico di rete sul determinismo di spedizione e ricezione dei pacchetti da parte del nodo stesso. Sperimentalmente, come nel caso di sottosezione 4.1, un nodo *P* spedisce ciclicamente dei pacchetti UDP, indirizzati al nodo *C*, ogni 500 ms.

Contemporaneamente, un processo con minore priorità di quello di test spedisce, attraverso le stesse interfacce di rete, ogni 1 ms, dei pacchetti UDP. Sperimentalmente il traffico interferente è stato spedito sia nella stessa direzione del traffico analizzato (SD) che nella direzione opposta (OD).

Tipo di stack (tipo di carico di rete)	Media	Std	Minimo (μ s)	Massimo
Linux (assente)	65.8	11.2	48.5	138.1
Linux (1ms DO)	65.8	12.1	43.8	162.5
Linux (1ms SD)	53.9	9.6	47.5	140.9
RTnet (assente)	70.0	6.0	59.9	108.0
RTnet (1ms DO)	70.5	8.6	50.9	148.6
RTnet (1ms SD)	62.4	4.3	44.4	115.5

Tabella 2 - Comparazione tra lo stack protocollare di Linux e RTnet con carichi di rete interferenti con periodo 1ms (DO: direzione opposta, SD: stessa direzione)

Dai risultati sperimentali riportati in Tabella 2 si può notare come, l'esistenza di un carico di rete interferente, agente in direzione opposta rispetto al traffico di test, non porti ad evidenti differenze rispetto all'assenza di carico di rete.

La trasmissione di tipo full-duplex di Ethernet permette di trasmettere, a livello fisico, i pacchetti in direzione $P \rightarrow C$ su cavi diversi rispetto a quelli in direzione $C \rightarrow P$. Così facendo, quando i pacchetti interferenti sono trasmessi in direzione opposta rispetto a quelli di test, non si può avere nessuna interferenza di canale. Dal punto di vista dello stack protocollare, per ogni nodo, parti diverse di software vengono sollecitate dal traffico di test e dal traffico interferente. Il nodo P , ad esempio, usa la parte di spedizione dello stack per quanto riguarda l'applicazione di test, mentre usa la parte di ricezione per quanto riguarda il traffico interferente. Questo permette, nel caso in cui entrambe le parti dello stack protocollare risiedano in *entry* separate della cache, di non avere deterioramenti prestazionali.

Un risultato sorprendente è invece quello riportato per il traffico interferente generato nella stessa direzione di quello dell'applicazione di test. In tal caso la presenza di traffico interferente fa migliorare le prestazioni dell'applicazione di test sia dal punto di vista della *latenza media* (53.9 μ s contro 65.8 μ s per *Linux* e 62.4 μ s contro 70.0 μ s per *RTnet*) sia dal punto di vista del determinismo (una deviazione standard di 9.6 μ s contro 11.2 μ s per *Linux* e di 4.3 μ s contro 6.0 μ s per *RTnet*). Tale comportamento è nuovamente ascrivibile a fenomeni di cache. I pacchetti inviati ad elevata frequenza (1 KHz) nella stessa direzione dei pacchetti inviati dall'applicazione di test vanno infatti ad interessare le stesse parti di codice dello stack protocollare usate dall'applicazione di test. In questo modo l'intero stack protocollare, di entrambi i nodi P e C , rimane completamente in cache perché utilizzato con elevata frequenza. Si noti come i dati ottenuti siano consistenti con i minimi riportati nel grafico di Fig. 2. L'interferenza causata da fenomeni di accodamento nel *ring buffer* di trasmissione della scheda di rete è trascurabile, in quanto mascherata dall'indeterminismo causato dalla cache.

5. Conclusioni

In questo contributo è stata caratterizzata la *latenza* introdotta a livello applicativo da due importati stack protocollari che sono lo stack convenzionale di *Linux* e lo stack *hard real-time RTnet*. In questa esposizione l'attenzione è stata focalizzata maggiormente sullo stack protocollare, piuttosto che sulla rete di comunicazione.

Mentre il comportamento dei due stack protocollari è simile nel caso di assenza di carico interferente, *RTnet* è risultato essere notevolmente più deterministico nel caso della presenza di carichi. *RTnet* ha, infatti, confermato il suo comportamento *hard real-time*, con delle *latenze massime* superiormente limitate. Gli esperimenti con lo stack *Linux*, invece, non hanno evidenziato nessun limite superiore, quindi tale stack non è utilizzabile per applicazioni *hard real-time*. Sfortunatamente la *latenza* di trasmissione dipende dall'hardware e più precisamente dalla cache della CPU. Essa può causare un indeterminismo di $\sim 35 \mu\text{s}$ per lo stack *Linux* e di $\sim 25 \mu\text{s}$ per lo stack *RTnet*. Si è misurato sperimentalmente che il tempo impiegato per spedire un pacchetto UDP, con *payload* pari a 230 ottetti, utilizzando lo stack protocollare *RTnet*, varia da circa 80 a 160 μs . Eliminando il tempo di rete, pari a 22.72 μs , si vede che il tempo speso internamente ai nodi varia da 57 a 137 μs . Tale tempo, e indeterminismo, non può essere trascurato nelle applicazioni reali e spesso è predominante rispetto al tempo di rete.

In definitiva si può concludere che la combinazione scheduler *RTAI* e stack protocollare *RTnet* permette di soddisfare i requisiti richiesti dalla quasi totalità delle applicazioni di controllo industriali.

Bibliografia

[Balaji et al, 2005] Balaji P., Feng W., Gao Q., Noronha R., Yu W., Panda D., Head-to-TOE Evaluation of High-Performance Sockets over Protocol Offload Engines, in IEEE Intl. Cluster Computing, 2005, 1–10.

[Beckhoff, 2011] Beckhoff Automation GmbH, EL1252 2-channel digital input terminal with time stamp, 2011, [online] Reperibile: <http://www.beckhoff.com/>.

[Chiandone et al, 2008] Chiandone M., Cleva S., Menis R., Sulligoi G., Industrial motion control applications using Linux RTAI, in International Symposium on Power Electronics, Electrical Drives, Automation and Motion. 2008, 528–533.

[Cena et al, 2010] Cena G., Cibrario Bertolotti I., Scanzio S., Valenzano A., Zunino C., On the Accuracy of the Distributed Clock Mechanism in EtherCAT, in IEEE International Workshop on Factory Communication Systems, 2010, 43–52.

[Cereia et al, 2011] Cereia M., Cibrario Bertolotti I., Scanzio, S., Performance of a Real-Time EtherCAT Master Under Linux. IEEE Transactions on Industrial Informatics, 2011.

[Decotignie, 2009] Decotignie J.-D., The Many Faces of Industrial Ethernet Past and Present. IEEE Industrial Electronics Magazine, vol. 3, no. 1, 2009, 8–19.

[Hanssen et al, 2005] Hanssen F., Jansen P., Scholten H., Mullender S., RTnet: a distributed real-time protocol for broadcast-capable networks, in Joint Intl. Conference on

Autonomic and Autonomous Systems and Intl. Conference on Networking and Services, 2005.

[Kiszka e Wagner, 2005] Kiszka J., Wagner B., RTnet - a flexible hard real-time networking framework, in 10th IEEE Intl. Conference on Emerging Technologies and Factory Automation. vol. 1, 2005, 448–456.

[Kiszka, 2005] J. Kiszka, "The Real-Time Driver Model and First Applications", in 7th Real-Time Linux Workshop, 2005.

[Ferrarini et al, 2003] Ferrarini L., Veber C., Fogliazza G., Modelling, design and implementation of machining centers control functions with object-oriented techniques, in IEEE/ASME International Conference on Advanced Intelligent Mechatronics, vol. 2, 2003, 1037–1042.

[Huo et al, 2005] Huo Z., Yu Y., Sun N., High performance Sockets over kernel level virtual interface architecture", in 8th Intl. Conference on High-Performance Computing in Asia-Pacific Region, 2005.

[Mantegazza et al, 2000] Mantegazza P., Dozio E. L., Papacharalambous S., RTAI: Real Time Application Interface. Linux Journal, 2000.

[Rostedt e Hart, 2007] Rostedt S., Hart D. V., Internals of the RT Patch, in Proc. of the Ottawa Linux Symposium, vol. 2, 2007, 161–172.

[Xenomai, 2011] Xenomai development team, Xenomai API Documentation 2.5.6, 2011, [online] Reperibile: <http://www.xenomai.org>.

[Zhou et al, 2007] Zhou K., Breinbauer B., Rausch T., Violations of Real Time Communication Constraints caused by Memory Transfers exceeding CPU Cache Limits in RTAI and RTnet, in 5th IEEE Intl. Conference on Industrial Informatics, vol. 1, 2007, 267–272.

Studio empirico sulla relazione tra influence e sentiment in Twitter in un contesto turistico

Donato Barbagallo, Leonardo Bruni, Chiara Francalanci, Paolo Giacomazzi

Dipartimento di Elettronica e Informazione

Politecnico di Milano

Via Ponzio 34/5, 20133 Milano

{barbagallo, bruni, francala, giacomaz}@elet.polimi.it

Data la crescente importanza assunta dai social network, vari studi si sono focalizzati sullo studio degli influencer. Tuttavia, anche se i messaggi pubblicati da questi utenti ottengono ampia diffusione, anche il contenuto stesso del messaggio può essere determinante ai fini della diffusione indipendentemente dalla centralità dell'autore. In questo articolo, questa tesi è supportata tramite la discussione di quattro ipotesi focalizzate sul contenuto del messaggio e sulla sua polarità. Le ipotesi sono validate su un dataset di 500,000 messaggi di Twitter in un contesto turistico.

1. Introduzione

I social media hanno un forte impatto sul modo in cui gli utenti interagiscono online e condividono informazioni. Per gestire questa enorme quantità di informazione si rende necessario lo sviluppo e l'impiego di tecnologie semantiche a supporto ad attività di social media intelligence. Le aziende inoltre sono sempre più consapevoli dei benefici derivanti dall'integrazione di queste attività con i loro processi di business e i relativi strumenti. Questa integrazione rappresenta il primo innovativo passo nell'ambito di ricerche di marketing, brand reputation management e customer care.

In questo contesto, il processo tramite il quale gli utenti creano e condividono opinioni sui brand, prodotti e servizi, ovvero il cosiddetto passaparola (*eWOM*, *electronic word-of-mouth*), sta ricevendo sempre maggiori attenzioni. In ambito web, l'eWOM è stato trasformato da un semplice atto privato di comunicazione di tipo *one-to-one* ad una complessa forma di interazione assolutamente di tipo *one-to-many*. Ed è proprio questo l'aspetto più importante del passaparola online: la diffusione di informazione condivisa attraverso la rete può essere ottenuta in maniera rapida e capillare. Tuttavia le aziende sono consapevoli che il controllo delle dinamiche di condivisione delle informazioni presenta serie difficoltà. La necessità di migliorare questo tipo di controllo è una delle ragioni per il quale c'è un crescente interesse nel comprendere come la struttura di una rete sociale può influenzare le dinamiche di interazione tra utenti e di condivisione delle informazioni.

Alcuni studi precedenti si sono focalizzati sul ruolo degli *influencer*, ovvero sul ruolo di nodi con una posizione di centralità nella rete. In particolare, le

piattaforme di microblogging quali Twitter sono al centro di numerosi studi che mirano a comprendere, da un lato le dinamiche di diffusione dei messaggi all'interno della rete, dall'altro il ruolo dell'autore nella portata di diffusione del messaggio stesso. Un paradigma emergente per lo studio delle reti sociali quale mezzo di comunicazione è quello di *attention economy* [Beck e Davenport, 2001]. Questo paradigma parte appunto dall'osservare come i brand siano sempre più coinvolti in una competizione per guadagnare l'attenzione di possibili clienti. Mentre sui media tradizionali l'attenzione è concentrata non solo sul contenuto ma anche sul modo in cui il messaggio è veicolato, sui social media è il contenuto ad assumere il ruolo principe.

E' nostra opinione che, sebbene le informazioni condivise dagli influencer abbiano indubbiamente una forte diffusione, il contenuto del messaggio assume comunque un ruolo critico e può essere determinante sulla influence del messaggio indipendentemente dalla centralità del suo autore. Occorre fare distinzione tra influencer e *influence*. Mentre i nodi sono influencer dipendentemente dalla loro centralità nella rete sociale, la influence rappresenta l'impatto dei messaggi, la quale dipende non solo dalla struttura della rete ma anche dalla capacità del contenuto dei messaggi di generare interesse. Lo studio di come i contenuti si diffondono attraverso la rete può essere utile per spiegare perché alcuni trend sono adottati più velocemente e con successo rispetto ad altri, fornendo quindi un impagabile valore aggiunto ad attività di business intelligence. Il concetto di influence può fornire accurati suggerimenti su come le aziende possono sfruttare i social media per rafforzare la reputazione del proprio brand.

In questo articolo, vengono definite quattro ipotesi a supporto del fatto che il contenuto dei messaggi assume un ruolo critico e determinante rispetto alla social influence indipendentemente dalla centralità degli autori. Viene inoltre identificato il sentiment dei messaggi come importante aspetto caratterizzante del loro contenuto. In generale, le nostre ipotesi affermano come i messaggi negativi siano maggiormente influenti di quelli positivi. La validazione è stata effettuata su un campione di 500,000 tweet in un contesto turistico.

Il resto dell'articolo è strutturato come segue. La sezione 2 introduce le basi della teoria dei grafi che sono necessarie per caratterizzare influence e influencer. La sezione 3 presenta lo stato dell'arte ed evidenzia gli aspetti innovativi di questo lavoro. La sezione 4 illustra le ipotesi di ricerca, mentre la sezione 5 discute i risultati dei test di validazione. Infine, nella sezione 6 sono discusse alcune conclusioni e presentati possibili sviluppi futuri.

2. Accenni di teoria dei grafi

L'idea alla base di una rete sociale è veramente semplice. Una rete è un insieme di attori (anche detti nodi o elementi) che sono relazionati tra loro in una qualche maniera. Dato un insieme finito di elementi U

$$U = \{X_1, X_2, \dots, X_n\}$$

e un numero finito di relazioni R

$$R_t \subseteq U \times U \text{ dove } t = 1, 2, \dots, r,$$

una rete sociale N è definita dalle tuple composte dall'insieme finito di elementi U e dalle r relazioni tra di loro:

$$N = (U, R_1, R_2, \dots, R_r).$$

Le relazioni possono rappresentare qualsiasi tipo di legame, come ad esempio la relazione padre-figlio, un legame di amicizia o una cooperazione tra impiegati di uno stesso progetto.

Sono state proposte differenti metriche per descrivere la posizione di un nodo all'interno di una rete [Brodka et al, 2009]. La posizione di un nodo è fondamentale per comprenderne il ruolo all'interno della rete. In particolare, le metriche di posizione di *prestigio* e di *centralità* misurano l'importanza del nodo.

Per misurare la posizione di prestigio di un nodo sono tipicamente utilizzate le misure di *indegree centrality* e *proximity prestige*. La prima è una metrica di tipo "locale", in quanto considera solo i nodi adiacenti e tiene in considerazione solo il numero di connessioni in ingresso. La seconda è, al contrario, una metrica di tipo "globale" che definisce la distanza tra un nodo e tutti gli altri nodi della rete.

Relativamente alle metriche di centralità sono state definite, in letteratura, differenti tipi di misure. Tra le più comunemente utilizzate vi sono: *outdegree centrality* (la quale misura la portata di diffusione di un nodo quando vengono fatti transitare messaggi attraverso i suoi archi in uscita), la *closeness centrality* (la quale misura la prossimità di un nodo agli altri nodi della rete) e la *betweenness centrality* (la quale misura il numero di volte che un nodo giace sul percorso a cammino minimo tra altri due nodi della rete).

3. Stato dell'arte

Lo studio delle reti sociali ha avuto inizio negli anni trenta con Simmel, il quale ha costruito la prima teoria capace di spiegare il fenomeno sociale. Nel 1934, Moreno fu il primo a proporre una rappresentazione formale delle reti sociali come combinazione di nodi e archi. In seguito, Harary e Cartwright [Cartwright, 1959] applicarono i concetti della teoria dei grafi alle reti sociali, le quali furono così definite *sociogrammi*. Con l'introduzione degli archi direzionati tra i nodi furono capaci di spiegare pattern sociali anche complessi.

Alla fine degli anni trenta, emersero due differenti scuole di pensiero. L'approccio sociocentrico [Chung et al, 2005] era focalizzato sull'identificazione di sottogruppi di persone all'interno della stessa rete per comprenderne le relazioni. L'approccio egocentrico era focalizzato sullo studio delle community nella loro interezza. Quest'ultimo approccio [Milgram, 1967] enfatizzava l'importanza delle reti sociali come mezzo di condivisione di conoscenza e informazione. In particolare, Milgram introdusse il concetto di "six degrees of separation" [Milgram, 1967], di particolare interesse per comprendere il potere dell'eWOM.

Freeman concentrò l'attenzione sull'identificazione di nodi rilevanti in una rete e sulle relative metriche [Freeman, 1979]. In quest'ottica, il microblogging ha creato nuove opportunità. Uno degli studi più interessanti sul microblogging è "The Million Follower Fallacy" [Benevenuto et al, 2010]. Utilizzando un ampio

dataset di tweet prodotto da circa sei milioni di utenti e considerando il numero di follower come metrica per misurare l'importanza di un utente, questo lavoro analizza la correlazione tra numero di follower, mention e retweet. La conclusione tratta è che la popolarità di un utente influenza debolmente l'attenzione ottenuta da altri utenti misurata come numero di mention e retweet.

Un recente studio nell'ambito dell'Ecology Web Project [Leavitt, 2009] si è focalizzato sulla influence di un insieme composto da 12 utenti molto popolari divisi in tre gruppi: celebrità, news e social media analyst. Gli autori dello studio hanno evidenziato che le celebrità hanno il più alto numero di follower e sono in grado di produrre volumi significativi di risposte con il minimo sforzo (ovvero, di attività). I social media analyst ottengono i più alti valori di influence se si pesano le loro risposte per il numero di follower, tuttavia, questi valori sono raggiunti solo con uno sforzo notevole. Infine, i canali di news hanno la capacità di vedere, in maniera semplice, i loro contenuti inoltrati dagli altri utenti.

Altre ricerche [Kwak et al, 2010ab] [Aus et al., 2010] hanno rilevato che la propagazione di un messaggio in Twitter è maggiore se un *twitterer* è influente, misurando la influence con algoritmi simili al PageRank [Brin e Page, 1998]. Tuttavia, questi approcci non sono stati ulteriormente verificati e non possono essere considerati generali.

4. Ipotesi

Noi sosteniamo che nei social media, incluso Twitter, il contenuto gioca un ruolo fondamentale nel determinare il livello di influenza delle informazioni. In questa sezione, vengono presentate quattro ipotesi che supportano questa tesi. Nelle nostre ipotesi ci focalizziamo sul sentiment dei messaggi, considerandolo come una caratteristica importante del contenuto dei tweet. La letteratura dà prova che gli utenti dei social media tendono ad autopromuoversi generando un numero maggiore di messaggi con sentiment positivo rispetto a quelli con sentiment negativo [Sobel et al, 2009; Berger e Milkman 2010]. Dall'altro lato, i media tradizionali solitamente enfatizzano le notizie negative [Galtung e Ruge 1965]. La nostra prima ipotesi punta a verificare se il sentiment dei messaggi abbia lo stesso impatto in termini di influenza così come nei media tradizionali:

“H1: In Twitter, i post negativi sono più retweettati rispetto ai post positivi. “

La letteratura ha studiato il ruolo dei social media, specialmente Twitter, come una sorgente di notizie [Kwak et al, 2010]. In particolare, la letteratura ha discusso la capacità delle rete sociali di diffondere le informazioni e la relativa volatilità delle informazioni create e “consumate” dagli utenti. [Kwak et al, 2010] hanno mostrato che i topic di tendenza hanno un periodo di attività di una settimana, mentre la metà dei retweet di un certo tweet avviene entro un'ora e il 75% entro un giorno. Nel caso delle emergenze, Twitter è stato spesso visto come un mezzo di informazione affidabile [Mendoza et al, 2010], mostrando picchi di conversazione eccezionalmente alti riguardo argomenti ad alta volatilità, tipici delle emergenze. In queste occasioni, il 33% dei tweet contiene espressioni di opinioni ed emozioni [Qu et al, 2011], che è una percentuale significativamente alta rispetto all'usuale 5% proprio di Twitter [Barbagallo et al, 2011]. Basandoci sui risultati precedenti, la nostra seconda ipotesi stabilisce

che la volatilità è una caratteristica generale dei tweet, comunque, i tweet che contengono opinioni positive sono più volatili rispetto ai tweet che contengono opinioni negative:

“H2: In Twitter, i tweet negativi sono retweettati in media più velocemente dei tweet positivi. “

Diversamente da Facebook, il meccanismo di amicizia proprio di Twitter è diretto. Gli utenti selezionano gli altri utenti da seguire e da cui ricevere delle notifiche in modalità push sulle loro homepage, dette timeline. E' stato dimostrato come questo meccanismo favorisca gli utenti più connessi [Zhou et al, 2010], poiché i messaggi da utenti che non vengono seguiti devono essere ricercati manualmente, ovvero con un meccanismo pull. Nella nostra terza ipotesi, postuliamo che il ruolo dei follower è forte, poiché è noto dalla letteratura del marketing che il passaggio dalla modalità push a quella pull è un meccanismo dal forte impatto. Ciò conduce alla terza ipotesi:

“H3: In Twitter, i messaggi negativi sono più retweettati dai follower rispetto ai non-follower. “

Le analisi condotte da [Bakshy et al, 2011; Benevenuto et al, 2010] hanno mostrato la sussistenza di relazioni tra il numero di follower e la propagazione dei messaggi. Da questi studi risulta chiaro che è importante avere un insieme (anche piccolo) di follower focalizzati su certi argomenti di discussione che possano essere considerati influencer piuttosto che molti utenti “normali”. Secondo la letteratura, gli influencer possono agire come degli hub e generare spontaneamente effetti di propagazione a palla di neve relativamente a notizie, creazione e accelerazione dell'innovazione [Katz e Lazarsfeld, 1955]. La letteratura non ha studiato la relazione tra il tempo di propagazione di un messaggio e la dimensione della rete dell'autore. Comunque, Twitter ha dimostrato di avere caratteristiche differenti nella dinamica dell'influenza così come mostrato da [Benevenuto et al, 2010; Bakshy et al, 2011] e anche utenti con reti non ampie sono capaci di sfruttare la loro influenza efficacemente. Questo ci porta a definire la prossima ipotesi:

“H4: In Twitter, il tempo (minimo, massimo o medio) necessario per un tweet negativo per essere retweettato la prima volta non è correlato al numero di follower dell'autore. “

5. Validazione empirica

Abbiamo raccolto dati relativi ai tweet e ai retweet usando le API di Twitter. Ci siamo focalizzati sul dominio del turismo, raccogliendo tutti i tweet contenenti la parola-chiave “milan” e abbiamo usato uno strumento di analisi semantica per selezionare solo i tweet relativi alla città di Milano [Barbagallo et al, 2011] e ne abbiamo valutato il sentiment. Più specificamente, siamo stati in grado di classificare i tweet secondo varie categorie di interesse per il dominio turismo in accordo al modello di Anholt [Anholt, 2009]. Abbiamo quindi creato due diversi insiemi di dati:

- l'Insieme A, usato per testare H1, contiene tutti i tweet riguardo la città di Milano, scritti tra Gennaio 2011 e Aprile 2011, contiene 676000 tweet (di cui 182000 retweet);

- l'Insieme B, usato per H2, H3 e H4, contiene tutti i tweet sulla città di Milano scritti tra l'1 Febbraio 2011 e il 28 Febbraio 2011. Questo insieme contiene 158000 tweet (di cui 38000 retweet).

Per testare l'ipotesi H1, abbiamo usato l'Insieme A e lo abbiamo analizzato usando il nostro strumento di analisi del sentiment [Barbagallo et al, 2011]. Lo strumento ha trovato 2355 tweet con sentiment, di cui 387 sono stati classificati come negativi e 1968 come positivi. Abbiamo poi considerato il numero di retweet che ciascuno di questi tweet ha ricevuto. Abbiamo quindi testato la normalità della distribuzione dei retweet tramite un test di Kolmogorov-Smirnov (sig. < 0.001). La Figura 1 conferma i risultati del test che mostra la funzione di distribuzione (a) e la densità (b) del numero di retweet di tweet con sentiment. Poiché il campione non è distribuito normalmente, abbiamo testato la differenza delle medie usando un test di Mann-Whitney ($Z = -1.817$, sig. = 0.069) mostrando un debole supporto per l'ipotesi H1.

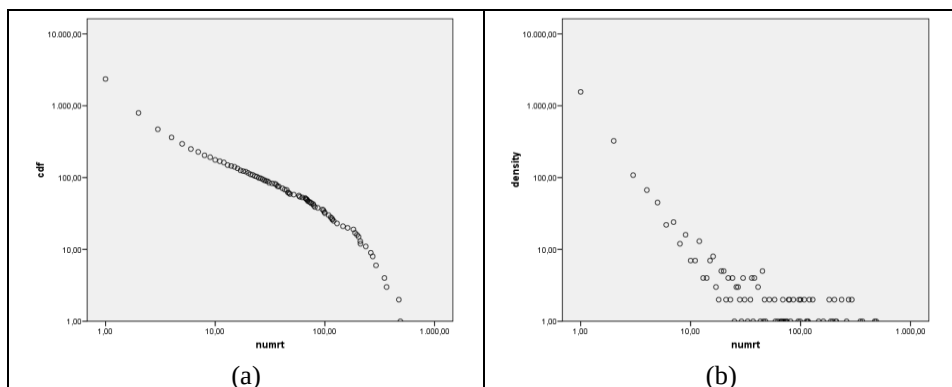


Figura 1. Funzione di distribuzione (a) e densità (b) del numero di retweet per ogni tweet con sentiment

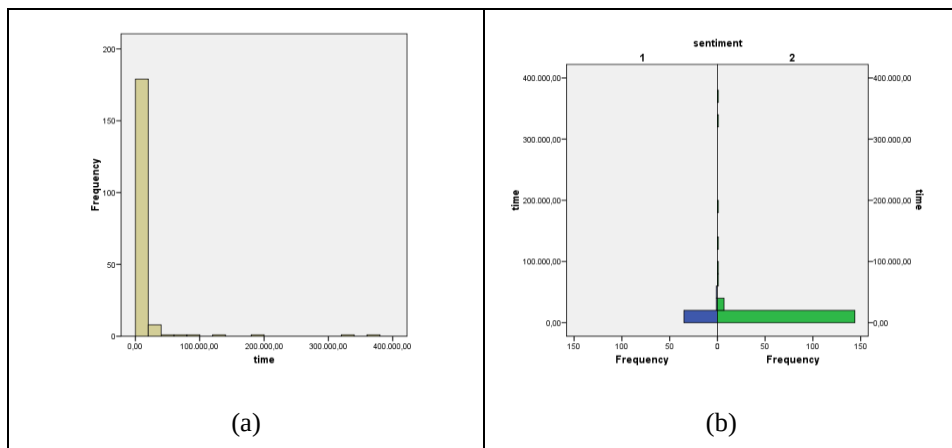


Figura 2. Distribuzione temporale (in secondi) del numero di retweet con sentiment (a) e diversi tempi di retweet rispetto al sentiment (b)

Per testare l'ipotesi H2, abbiamo analizzato le dinamiche di retweet. In particolare, abbiamo considerato le distribuzioni temporali e dei picchi. Più specificamente, la Figura 2a mostra che il 92% dei tweet che contengono il sentiment sono retweettati entro 5 ore, mentre la Figura 2b mostra la differenza dei tempi di retweet tra i tweet con sentiment positivi e negativi. E' possibile osservare che i tweet con sentiment negativo (colonna 1) sono più concentrati nei primi minuti, mentre i tweet positivi (colonna 2) hanno volumi maggiori ma sono retweettati più tardi. Definiamo il tempo t di un picco di retweet come:

$$t = \arg\max (NRT(t)) \quad 0 \leq t \leq T,$$

dove $NRT(t)$ è il numero di retweet di un dato tweet al tempo t e T è il tempo dell'ultimo retweet. Il tempo t è relativo all'istante temporale in cui il tweet è scritto. La Figura 3a riporta la distribuzione dei picchi lungo le prime 4 ore a partire dall'istante temporale in cui il tweet viene creato e mostra il numero cumulato di tweet che hanno il picco dei loro retweet in ciascun istante. E' chiaro che la grande maggioranza dei picchi avviene nel primo minuto. Nel nostro campione, 1168 tweet (il 17%, come si può vedere dalla Tabella 1) hanno il picco entro il primo minuto. La Figura 3b riporta gli stessi dati in una scala log-log. Il grafico che si ottiene è molto simile ad una distribuzione power-law, poiché la decrescita del numero di picchi nel tempo è a coda lunga. Le Figure 2 e 3, quindi, sono utili per avere una idea qualitativa delle dinamiche di retweet. In ogni caso un test di Mann-Whitney non supporta l'ipotesi H2 ($Z = -0.11$, sig. = 0.912).

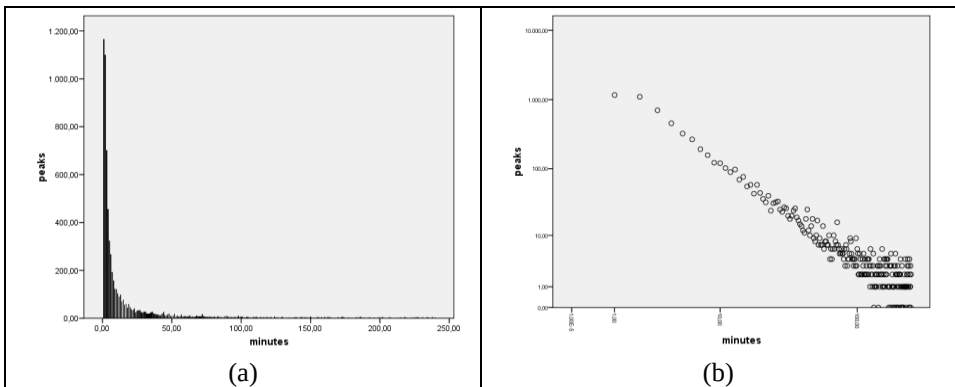


Figura 3. Densità dei picchi di retweet nelle prime 4 ore su scala lineare (a) e log-log (b)

Un'analisi più approfondita può essere condotta osservando i dati riportati in Tabella 1: l'80% dei picchi avviene entro 31 minuti e quasi un terzo avviene dopo due minuti.

Tabella 1. Percentuale cumulata del numero dei picchi di retweet nel tempo

	percentuale cumulata
percentuale di retweet entro 1 minuto	17%
percentuale di retweet entro 2 minuti	32%
percentuale di retweet entro 5 minuti	53%
percentuale di retweet entro 31 minuti	80%
percentuale di retweet prima della media	90%

Per testare l'ipotesi H3, abbiamo considerato l'Insieme B che contiene 18206 retweet che sono stati postati da follower dell'autore del relativo tweet e 13061 che sono stati postati da utenti che non seguivano l'autore nel momento in cui questo ha scritto il messaggio. Un test di Kolmogorov-Smirnov (sig. < 0.001) è stato eseguito per verificare la normalità della distribuzione dei retweet nell'insieme B. Poiché i dati non sono distribuiti normalmente, abbiamo eseguito un test di Mann-Whitney per la differenza delle medie tra i due gruppi, il cui risultato supporta l'ipotesi H3 ($Z = -2.79$, sig. = 0.005).

L'ipotesi H4 è stata testata usando le correlazioni. Poiché i dati non sono distribuiti secondo una normale, abbiamo usato i coefficienti di correlazione di Spearman. Tutte le correlazioni non forniscono supporto per l'ipotesi H4 né per il tempo minimo ($\rho = -0.203$, sig. < 0.001) né per il tempo medio ($\rho = 0.101$, sig. < 0.001) né per il tempo massimo ($\rho = 0.046$, sig. < 0.001).

6. Conclusioni e sviluppi futuri

Mentre vari studi hanno sottolineato le differenze tra i social media e i media tradizionali, i nostri risultati evidenziano alcune interessanti caratteristiche comuni. Anche se il numero assoluto di tweet contenenti sentiment positivo è maggiore del numero di quelli contenenti sentiment negativo, in media i tweet negativi sembrano essere maggiormente retweetati ed attirare maggiore attenzione. Le dinamiche di retweet, comunque, sembrano essere indipendenti rispetto al sentiment contenuto (H2). Le analisi qualitative mostrano come i tweet siano volatili, visto che l'80% dei retweet avvengono entro un'ora da quando sono scritti. Questo aspetto è particolarmente critico per le aziende, perché mostra che il monitoraggio dei social media deve essere un processo che avviene quasi in tempo reale. I nostri risultati sono basati su un campione di tweet di dominio turistico relativi alla città di Milano. E' interessante notare come [Chen et al, 2010] abbiano trovato che il 4.5% dei tweet discutono direttamente di turismo e consigli di viaggio, senza considerare quegli argomenti che possono riguardare indirettamente il turismo, come arte e cultura (31,8%),

notizie locali (54,5%) e sport (18,2%). I risultati suggeriscono che l'influenza di una notizia in uno qualsiasi di questi domini potrebbe avere un impatto indiretto sul turismo, significativamente maggiore se i tweet sono negativi.

Il ruolo del numero di follower è ancora poco chiaro. Studi precedenti indicano che la relazione tra la centralità di un utente e la capacità di stimolare le discussioni è limitata. Questo articolo mostra che la centralità non ha alcuna relazione diretta con il tempo di retweet (H4). D'altro canto è chiaro che avere un numero maggiore di follower aumenta la probabilità di essere retweettati, specialmente per le opinioni negative, come annunciato nell'ipotesi H3.

Negli ultimi anni le città stanno provando a migliorarsi utilizzando la tecnologia per migliorare la qualità della vita dei loro cittadini e per attrarre turisti pubblicizzando i loro servizi. Twitter sembra essere uno strumento molto potente per costruire la reputazione del brand di una città. Scharl [Scharl et al, 2008] ha mostrato che la copertura dei social media e le opinioni influenzano l'immagine di una destinazione turistica. I nostri risultati confermano queste idee e mostrano che mentre il contenuto di un messaggio in termini del sentiment ha un impatto sulla sua diffusione, le dinamiche temporali delle interazioni fra utenti sembrano essere delle caratteristiche proprie dello specifico social medium e indipendenti dal contenuto.

Attualmente stiamo lavorando sulla rilevazione degli argomenti di tendenza al fine di fare delle misure più precise sulla diffusione delle notizie relative agli eventi. Infatti, una delle limitazioni di questo lavoro è quella di considerare solo i retweet che contengono la parola chiave "milan" (nell'accezione di città), mentre lo studio della diffusione degli eventi basato sul meccanismo degli hashtag di Twitter aumenterebbe la generalizzabilità dei nostri risultati.

Bibliografia

Anholt S., *Competitive Identity: The new brand management for nations, cities and regions*. Palgrave Macmillian (eds), 2009.

Aus S., Galuba W., Huberman B.A., Romero D.M., *Influence and passivity in social media*. ACM, 2010.

Bakshy E., Hofman J.M., Mason W.A., Watts D.J., *Everyone's an influencer: quantifying influence on twitter*. Proc. Of the 4th ACM International conference on Web search and data mining (WSDM'11), ACM, New York, 2011, 65-74.

Barbagallo D., Cappiello C., Francalanci C., Matera M., *Semantic sentiment analyses based on the reputation of Web information sources*. Applied Semantic Web Technologies, Sugumaran, V. and Gulla J. A. (eds), Taylor & Francis, 2011.

Beck T., Davenport J., *The attention economy: Understanding the new currency of business*. Harvard Business School press, Cambridge, 2001.

Benevenuto F., Cha M., Gummadi K.P., Haddadi H., *Measuring user influence in Twitter: The million follower fallacy*. Proc. Of the 4th International AAAI Conference on Weblogs and Social Media, 2010, 10-17.

Berger J., Milkman K., *Social transmission, emotion, and the virality of online content*. Wharton Research Paper, 2010.

Brin S., Page L., The anatomy of a large-scale hypertextual Web search engine. Proc. of the 7th International World Wide Web Conference, 1998, 107-117.

Brodka P., Kazienko P., Musial K., User position measure in social networks. Proc. of the 3rd Workshop on Social Network Mining and Analysis, Paris, France, 2009, 1-9.

Cartwright D., Studies in social power, Vol. 6, University of Michigan, 1959.

Chen J., Nairn R., Nelson L., Bernstein M., Chi E., Short and tweet: experiments on recommending content from information streams. Proc. of the 28th International conference on Human factors in computing systems, ACM, New York, 2010, 1185-1194.

Chung K.K., Davies J., Hossain L., Exploring sociocentric and egocentric approaches for social network analysis. Proc. of the International conference on Knowledge Management, Wellington, New Zeland, 2005, 1-8.

Freeman L.C., Centrality in social networks: Conceptual clarification. Social Networks, 1979, 215-239.

Galtung J., Ruge M., The structure of foreign news: The presentation of the congo, cuba and Cyprus crises in fourc Norwegian newspapers. Journal of Peace Research, 1965, 64-90.

Katz E., Lazarsfeld P.F., Personal influence; the part played by people in the flow of mass communications. Free Press, Glencoe, 1955.

Kwak H., Lee C., Park H., Moon S., Finding influential based on the temporal order of information adoption in Twitter. Proc. of the 19th International World Wide Web conference, ACM, Raleigh, North Carolina, 2010a, 1137-1138.

Kwak H., Lee C., Park H., Moon S., What is Twitter, a social network or a news media?. Proc. of the 19th International conference on World Wide Web, ACM, New York, 2010b, 591-600.

Leavitt A., The influential: New approaches for analyzing influence on Twitter, Boston, USA, 2009.

Mendoza M., Poblete B., Castillo C., Twitter under crisis: can we trust what we RT.?. Proc. of the 1st Workshop on Social Media Analytics, ACM, New York, 2010, 71-79.

Milgram S., The small world problem. Psychology Today, 1967, 60-67.

Qu Y., Huang C., Zhang P., Zhang J., Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. Proc. of the ACM 2011 conference on computer supported cooperative work, New York, USA, 2011, 25-34.

Scharl A., Dickinger A., Weichselbraun A., Analyzing News Media Coverage to acquire and structure tourism knowledge. Information technology & tourism, 2008, 3-17.

Sobel K., Jansen B.J., Zhang M., Chowdury A., Twitter power: tweets as electronic word of mouth. Journal of the American Society for information science and technology, 60, 2009, 2169-2188.

Zhou Z., Bandari R., Kong J., Qian H., Roychowdhury V., Information resonance on Twitter: watching Iran. Proc. of the 1st workshop on Social Media Analytics, ACM, New York, 2010, 121-131.

Supporto informatico per indagini preventive di potenziale diffusione di contagio tra bovini

Luca Maselli, Alfonso Scrivo e Paolino Di Felice

Dipartimento di Ingegneria Elettrica e dell'Informazione

Università degli Studi dell'Aquila – L'Aquila

(lucabolo.mas@hotmail.it, giuseppe_scrivo@hotmail.com, paolino.difelice@univaq.it)

Esiste crescente consapevolezza che per tutelare la salute dell'uomo dal consumo di carne infetta occorre monitorare attentamente la salute del bestiame destinato alla macellazione. Questo articolo fornisce un contributo metodologico e tecnologico che rende possibile, avvalendosi di strumenti informatici, diagnosticare con tempestività l'insorgenza di malattie a carattere diffusivo e contagioso, premessa necessaria all'avvio di una campagna di visite veterinarie mirate.

1. Motivazioni

Il rischio di macellare ed immettere sul mercato carne infetta è strettamente connesso al crescente aumento di allevamenti di bestiame intensivi e al notevole incremento degli spostamenti di questi ultimi. Esistono due ragioni prevalenti che inducono a spostare il bestiame: la necessità di trasferirli da un allevatore ad un altro (anche di Paesi differenti) oppure esporli a delle fiere. In entrambi i casi gli spostamenti sono in relazione ad *attività economiche* connesse ad operazioni di compra-vendita. Circa l'entità degli spostamenti essi possono avvenire tanto sul territorio nazionale che nell'ambito dei Paesi della UE. Tipicamente i viaggi del primo tipo sono di breve durata (qualche ora), mentre i secondi richiedono qualche giorno.

La direttiva 91/628/CEE nel 1991 stabilisce che, nel corso del trasporto, gli animali devono essere scaricati, fatti riposare, nutriti e abbeverati a intervalli regolari. Questa direttiva è stata rafforzata da due successivi regolamenti del CE, ovvero:

- il N.1255/97, riguardante i criteri comunitari circa i luoghi destinati ad ospitare temporaneamente il bestiame, che stabilisce le misure sanitarie destinate ad evitare la trasmissione di malattie, nonché l'obbligo di registrare i movimenti degli animali e
- il N.1/2005 *sulla protezione degli animali* coinvolti in lunghi viaggi attraverso i Paesi della UE. Per gli scopi di questo articolo di tale Regolamento è sufficiente richiamare che esso:

- stabilisce *come* debba essere equipaggiato il mezzo di trasporto onde garantire benessere agli animali a bordo, unitamente alla possibilità di nutrirli adeguatamente,
- esplicita, in relazione alla specie, ogni quante ore è necessario far scendere gli animali dal mezzo per concedere loro una pausa (che comunque dev'essere di *almeno* 12 ore - Art.36, comma 2) in luoghi denominati *posti di controllo*.

Sfortunatamente, al di là dei molteplici sforzi di regolamentazione che si è sostenuti da un ventennio in ambito comunitario, la nota dolente denunciata da vari organismi di controllo della salute degli animali e, di riflesso, di quella dell'uomo è che le norme cautelative contenute nei sopra citati regolamenti sono, ad oggi, diffusamente disattese (una testimonianza è reperibile ai punti “(3)” e “(5)” del Reg. CE N.1040/2003).

In attesa che ai regolamenti esistenti, e ad altri che verosimilmente seguiranno, siano affiancati dei controlli sistematici in ambito UE per accertarne l'attuazione piena, resi più efficaci da sanzioni dissuasive per chi disattende le direttive, gli autori ritengono che sia essenziale dotarsi di strumenti informatici cui demandare il compito di condurre analisi preventive di possibili contagi contratti dal bestiame che sia venuto in contatto con capi risultati infetti a seguito di controlli veterinari.

Nel presente articolo si propone una soluzione metodologica e tecnologica in tal senso. Pregio della stessa, oltre al costo bassissimo e ai tempi di risposta immediati, è restringere il numero dei capi sui quali avviare la campagna di visite che, diversamente, andrebbe estesa a tutto il bestiame che ha subito degli spostamenti nel lasso di tempo che è intercorso tra la visita dell'esemplare risultato malato e quella subito prima, nella quale lo stesso era ancora sano. Evidentemente, questa seconda strada è molto dispendiosa in termini di tempo e, in caso di episodi di contagio grave, essa potrebbe rivelarsi poco tempestiva prima di procedere all'abbattimento dei capi infetti, che nel frattempo potrebbero essere stati macellati ed immessi sulle catene della grande distribuzione.

Questo articolo è organizzato come segue. La Sez.2 introduce la terminologia essenziale alla comprensione dello scenario applicativo. In particolare, essa descrive l'organizzazione dei luoghi destinati ad accogliere il bestiame durante le soste a terra e, successivamente, le forme di contagio possibili. La Sez.3 discute modalità differenti attraverso le quali è possibile condurre l'indagine preventiva circa i capi di bestiame potenzialmente contagiati. La Sez.4 riassume la realizzazione tecnologica di quanto proposto nella Sez.3. Nello specifico, la realizzazione consta di una base di dati relazionale corredata da un pacchetto di queries ad hoc che implementano la ricerca del bestiame potenzialmente contagiato. La base di dati raccoglie i dati riguardanti gli animali, i controlli effettuati su di essi, i viaggi e i punti di stazionamento. Si noti che i dati circa le visite veterinarie, gli spostamenti e le soste non sono in tempo reale bensì storicizzati con una qualche periodicità. Aggiornamenti giornalieri oppure settimanali sono esempi di valori plausibili per tale periodicità. Valori più alti sono da sconsigliare perché nei casi di malattie

Supporto informatico per indagini preventive di potenziale diffusione di contagio tra bovini fortemente contagiose potrebbero vanificare la finalità preventiva dell'indagine. La Sez.5 riassume quanto emerso dai primi test delle strategie d'indagine introdotte nella Sez.3 effettuati eseguendo le queries su di un insieme di dati "leggero" ma oculatamente allestito, detto *dataset minimale*. I risultati così ottenuti consentono, altresì, di approntare dei calcoli utili a dare sostegno all'affermazione che la scelta del metodo d'indagine (tra quelli proposti) debba essere guidata dal grado di contagiosità della malattia e dalla sua pericolosità per la salute umana. Brevi conclusioni sono raccolte nella Sez.6.

Si desidera, infine, rimarcare che i destinatari del contributo metodologico e tecnologico contenuto in questo articolo potrebbero essere i Centri di Riferenza Nazionale, localizzati presso gli Istituti Zooprofilattici Sperimentali (IZS) che, come si legge sul sito del Ministero della Salute [3] "rappresentano uno strumento operativo di elevata e provata competenza, nei settori della sanità animale, dell'igiene degli alimenti e dell'igiene zootecnica".

2. Lo scenario applicativo

Il Reg. N.1/2005 riguarda il trasporto di animali vertebrati vivi, ossia: mammiferi, uccelli, rettili, anfibi e pesci. Nel seguito ci si riferirà in senso stretto ai bovini per il largo consumo della carne di tale specie da parte dell'uomo.

Preliminarmente, si procede ad esplicitare la terminologia che verrà adottata, in prevalenza ispirata ai regolamenti richiamati nella Sez.1:

- chiameremo *posto di controllo* un luogo destinato al nutrimento ed al riposo degli animali durante le soste all'interno dei lunghi viaggi. Più "lotti" di bestiame possono essere presenti contemporaneamente in tali luoghi a patto di avere la "medesima qualifica sanitaria certificata" di accompagnamento. Il Reg. 1/2005 (Art. 36, prf. 8, lettera b) stabilisce che i posti di controllo devono essere progettati, costruiti e gestiti in modo da garantire:
 - a) una sufficiente biosicurezza che eviti la diffusione di malattie infettive tra gli animali che in essi transitano;
 - b) che possano essere eseguiti i lavori di pulizia e disinfezione.Nel seguito supporremo i posti di controllo costituiti da un insieme di *recinti* numerati ordinatamente "ragionevolmente" isolati tra di loro, ciascuno dei quali può ospitare un gruppo omogeneo di animali durante la sosta (Fig.1a);
- nel caso di animali in movimento sul territorio nazionale, stante la breve durata dei viaggi (meno di 8 ore), non sono previste pause prima della destinazione. Il Reg. N.1/2005 (Art.2), chiama *centri di raccolta* i luoghi di assembramento del bestiame durante le fiere. Generalmente tali luoghi non soddisfano i requisiti di igiene e "ragionevole" isolamento da contagio dei posti di controllo, pertanto la loro definizione la si deriva da quella di quest'ultimi rimuovendo l'ipotesi che i recinti siano isolati tra di loro (Fig.1b). Si segnala che, ove siano giudicati idonei dai veterinari responsabili, i centri di raccolta possono fungere anche da posti di controllo (Regolamento N.1255/97, Art.4, comma 2).

Nel seguito dell'articolo si chiamerà *punto di stazionamento* tanto un posto di controllo che un centro di raccolta, limitando, per non appesantire la lettura, l'uso della denominazione esplicita ai casi strettamente necessari.



Figura 1 Un punto di controllo (a) ed un centro di raccolta (b) costituiti da 5 recinti

- Chiameremo *lotto* un gruppo di animali di una medesima specie (nello specifico *bovini*) che si sposta insieme e che, quando accade, è raccolto in un medesimo punto di stazionamento. Se, a seguito di un controllo veterinario, accade che un capo di bestiame di un lotto risulta malato e, quindi, infetto, l'intero *lotto* viene detto *malato*. Si parla, invece, di *lotto potenzialmente contagiato* quando si sospetta che i capi al suo interno potrebbero aver subito la trasmissione di microrganismi patogeni da parte di un lotto malato.

E' proprio nei punti di stazionamento che avviene il maggior numero di contagi. Con riferimento ai due tipi di punti di stazionamento, vale a dire: posti di controllo e centri di raccolta, possono manifestarsi episodi di contagio per meccanismi differenti e ciò in conseguenza della ipotesi di ragionevole isolamento tra i recinti che ne scandisce le differenze dal punto di vista del nostro studio.

Forme di contagio possibili nei centri di raccolta

Due recinti si dicono *contigui direttamente* se sono separati da uno steccato o da un muro che però non impedisce il contagio tra gli animali in essi contenuti. Ad esempio, le coppie di recinti (Fig.1b) $\langle i, i+1 \rangle$ con $i=1,2,3,4$, sono contigui direttamente. Due recinti si dicono *contigui indirettamente* se tra di essi c'è almeno un altro recinto contiguo direttamente ad uno o entrambi. Ad esempio, le coppie di recinti $\langle i, i+2 \rangle$, con $i=1,2,3$, (Fig.1b) sono contigui indirettamente.

Da quanto appena descritto discende che sono possibili tre forme di contagio: diretto, indiretto e per contiguità temporale. Il *contagio diretto* si verifica quando due lotti, di cui uno risultato successivamente infetto, sostano, in periodi che siano almeno in parte sovrapposti, in recinti contigui direttamente di un medesimo centro di raccolta.

Il *contagio per contiguità temporale* si verifica quando un lotto sano va a occupare un recinto che in precedenza aveva ospitato un lotto malato e dove, pertanto, potrebbero essere rimasti degli agenti patogeni nell'ambiente. Evidentemente il pericolo reale che il lotto sano venga contagiato dipende dalla contagiosità della malattia del lotto malato, pericolo che comunque decresce con la "distanza temporale" che intercorre tra quando il lotto malato abbandona il recinto e quando quello sano vi entra. La soluzione proposta in questo articolo

Supporto informatico per indagini preventive di potenziale diffusione di contagio tra bovini
 è parametrica, nel senso che essa consente di adeguare il valore di tale intervallo alla contagiosità della malattia.

Il *contagio indiretto*, fra un lotto malato ed uno sano, avviene quando i due lotti hanno occupato recinti contigui indirettamente, ma i recinti che si frappongono fra essi non erano vuoti e, pertanto, i bovini in essi raccolti potrebbero aver fatto da tramite nella trasmissione degli agenti patogeni dal lotto malato al lotto sano.

Forme di contagio possibili nei posti di controllo

Sono due: *contagio per co-presenza* e *contagio per contiguità temporale*. Il primo materializza il pericolo potenziale di contagio cui è sottoposto il bestiame per il semplice fatto di essere in co-presenza con capi malati all'interno di un medesimo posto di controllo (Sez.2). Del secondo si è già detto sopra. Si assume, infine, improbabile lo sviluppo di contagio diretto e/o indiretto in virtù del ragionevole grado di isolamento che esiste tra i recinti dei posti di controllo (Sez.2).

3. Strategie d'indagine

Di seguito vengono discusse tre differenti strategie d'indagine dei lotti potenzialmente contagiati a partire da differenti livelli di dettaglio adottati nell'analisi condotta sui dati a disposizione. La prima delle tre è applicabile a entrambi i tipi di punto di stazionamento della Sez.2, mentre le altre due si contestualizzano ai soli centri di raccolta. In questo secondo scenario va detto che ognuna delle tre strategie ha un suo spazio di usabilità, determinato dalla pericolosità della malattia del lotto e dal personale veterinario impiegabile per condurre la campagna di visite sui lotti potenzialmente infettati in tempi accettabili e comunque tali da poter bloccare, se necessario, l'immissione sul mercato della carne.

Livello di pericolosità	Strategia 1	Strategia 2	Strategia 3
Basso		✓	✓
Medio		✓	✓
Alto	✓		

Tabella 1 Strategie d'indagine e loro utilizzo ottimale

Se si denota con $\{Q_i\}$ (con $i=1,2,3$) l'insieme dei risultati generati dal processamento delle tre strategie d'indagine, vale quanto segue:

$$\{Q_1\} \supseteq \{Q_2\} = \{Q_3\} \quad (1)$$

3.1 Strategia d'indagine 1

Essa considera potenzialmente contagiato un lotto se solo esso ha frequentato lo stesso punto di stazionamento di uno malato. Questa strategia è la sola applicabile delle tre nel caso di stazionamento del bestiame in un posto di controllo mentre, con riferimento ai centri di raccolta, essa è la più cautelativa

delle tre poiché opera al livello del punto di stazionamento disinteressandosi di quali recinti siano stati occupati dai singoli lotti. Ne consegue che il numero di lotti che risulteranno potenzialmente contagiati, sovente, è maggiore di quello che si avrebbe considerando le forme di contagio definite nella Sez.2, la cui definizione fa esplicito riferimento alla collocazione dei lotti nei recinti del centro di raccolta. Poiché i controlli da fare sui lotti potenzialmente contagiati comportano un notevole dispendio di ore uomo, si intuisce perché tale metodo di analisi sia appropriato nei casi in cui la malattia del lotto malato è altamente pericolosa per la salute dell'uomo ed essa ha una elevata contagiosità (Tab.1).

3.2 Strategia d'indagine 2

In essa la ricerca dei lotti potenzialmente contagiati prende in considerazione i recinti ove essi hanno alloggiato. Questa strategia non include, pertanto, tra i lotti potenzialmente contagiati quelli che, pur avendo condiviso lo stesso centro di raccolta con un lotto malato, non hanno subito alcuna forma di contagio in accordo alle definizioni della Sez.2. Questa circostanza può verificarsi, ad esempio, nel caso in cui un lotto sano e uno malato occupano recinti contigui indirettamente ed almeno uno dei recinti che li separa è vuoto e quindi impedisce la trasmissione degli agenti patogeni.

Questo metodo d'indagine è indicato nei casi in cui il lotto malato è affetto da una malattia di media o bassa contagiosità e, pertanto, non particolarmente pericolosa per la salute dell'uomo (Tab.1).

3.3 Strategia d'indagine 3

Questo metodo d'indagine costituisce un raffinamento del precedente ottenuto associando ad ogni lotto che ha sostato in qualche centro di raccolta un valore di probabilità che esso abbia contratto contagio. Questo elemento di novità è importante poiché arricchisce il risultato dell'analisi abbinando ad ognuno dei lotti nell'insieme $\{Q_3\}$ un valore di probabilità di contagio. Il risvolto pratico di questo arricchimento del risultato dell'analisi è che esso suggerisce un ordine da seguire nel condurre la campagna di visite veterinarie e, nei casi in cui questo fosse reputato lecito, fissare una "soglia" di taglio sotto la quale non spingersi. Di seguito si esplicita come calcolare la probabilità di contagio diretto, indiretto e per contiguità temporale (Sez.2).

Probabilità di contagio indiretto (P_{CI})

$$P_{CI} = \left(\frac{0.9}{|i-j|} \right) \times \left(\frac{\cap h}{8} \right), \quad \cap h \leq 8 \quad (2)$$

La (2) evidenzia che il calcolo della probabilità di contagio indiretto si ottiene dal contributo di due fattori. Nel primo, $|i-j| > 1$ esprime la distanza tra i recinti che ospitano, nell'ordine, il lotto sano e quello malato. Nel secondo, $\cap h$ denota l'intervallo di tempo nel quale un lotto sano ed uno malato hanno coabitato in un medesimo centro di raccolta.

Ad esempio, con riferimento al centro di raccolta di Fig.1b ($|i-j| \leq 4$) i valori del primo fattore della (2) sono: 0.45, 0.30, 0.225. A sua volta, il fattore $(\cap h/8)$ aumenta linearmente con $\cap h$ (Fig.2), assumendolo al più valore pari a 1 (si è ipotizzato, infatti, che le ore di permanenza del bestiame in un centro di raccolta non superi le 8 ore). Si noti che $(\cap h/8)$ è una funzione continua definita per qualsiasi $0 < \cap h \leq 8$.

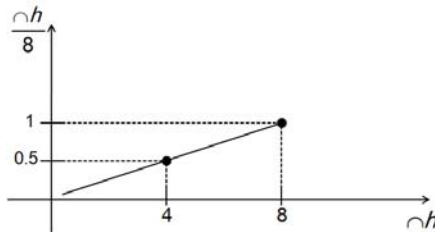


Figura 2 L'andamento del fattore $(\cap h/8)$

Probabilità di contagio indiretto (P_{CD})

La formula per calcolarla discende banalmente dalla (2) ove si ponga $|i-j|=1$.

Probabilità di contagio per contiguità temporale (P_{CCT})

E' definita dalla (3).

$$P_{CCT} = \begin{cases} = 0.75 & \text{se } \Delta h \leq 8 \\ = 0.5 & \text{se } 8 < \Delta h \leq 24 \\ = 0.25 & \text{se } 24 < \Delta h \leq 48 \\ = 0.1 & \text{se } \Delta h > 48 \end{cases} \quad (3)$$

Ovvero ad essa si attribuiscono valori decrescenti, a partire da 0.75, in base alle ore intercorse prima che un lotto sano entri in un recinto che aveva ospitato uno malato.

4. Realizzazione tecnologica

La realizzazione di quanto discusso sinora consta di una base di dati relazionale corredata da un pacchetto di queries ad hoc che implementano la ricerca dei lotti potenzialmente contagiati. La base di dati contiene al suo interno i dati riguardanti gli animali, i controlli effettuati su di essi, i viaggi e i punti di stazionamento. La Fig.3 ne mostra lo schema concettuale.

Lo schema prevede 6 entità e 5 associazioni, di cui una ternaria; le proprietà delle suddette entità sono desumibili dall'esame della Fig.3. La traduzione dello schema concettuale produce la base di dati costituita dalle seguenti 7 tabelle:

LOTTO_ANIMALI	(<u>id_lotto</u> , numero_capi, specie)
VIAGGIO	(<u>id_lotto</u> , <u>partenza</u> , luogo_partenza, luogo_destinazione, arrivo)
RECINTO	(<u>id_ps</u> , numero)
SOSTA	(<u>id_lotto</u> , <u>id_ps</u> , recinto, <u>da</u> , a)
ATTIGUO	(<u>id_ps_1</u> , <u>recinto_1</u> , <u>id_ps_2</u> , <u>recinto_2</u>)

PUNTO_SOSTA (id_ps, nome, comune)
 VISITA (data_visita, id_lotto, esito, diagnosi)

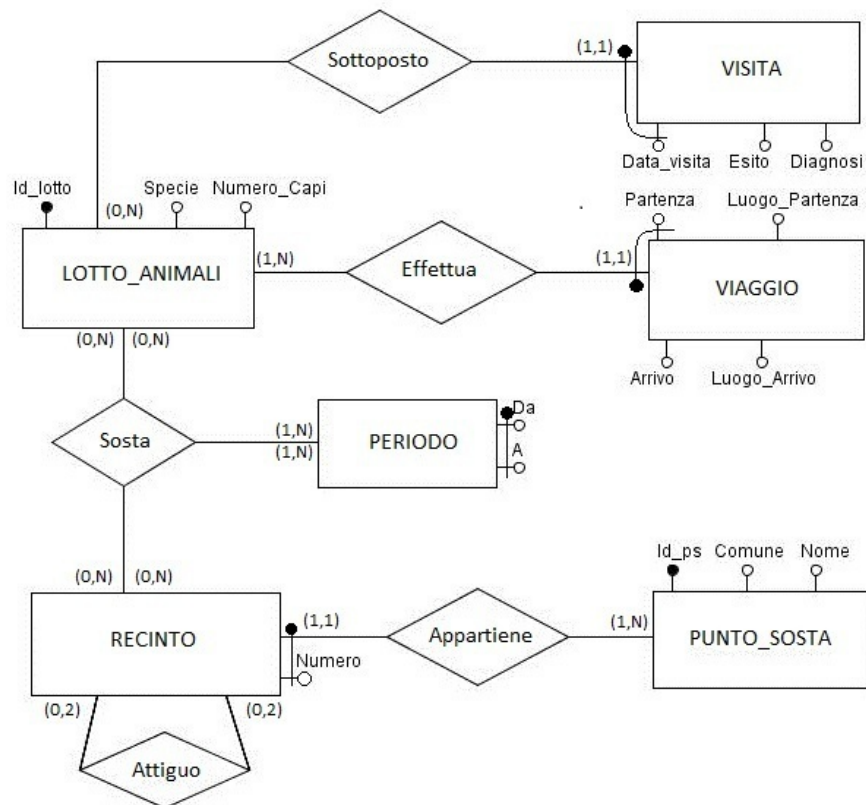


Figura 3 Schema concettuale della base di dati

Quale SGBDR è stato scelto PostgreSQL [4] che, come noto è open source. La consistenza dei dati in ingresso alla BD è garantita da 9 trigger. Mentre la realizzazione delle tre strategie d'indagine ha richiesto 12 viste SQL e 6 User Defined Functions (Tab.2), queste ultime si sono rese necessarie per calcolare la probabilità di contagio dei lotti.

	Num. viste	Tabelle coinvolte	Linee di codice SQL	UDF		Ricorsione
				Num.	Linee codice	
Strategia 1	2	VISITA, SOSTA, VIAGGIO	20			NO
Strategia 2	5	VISITA, SOSTA, VIAGGIO, ATTIGUO	53			SI
Strategia 3	5	VISITA, SOSTA, VIAGGIO, ATTIGUO	49	3	195	SI

Tabella 2 Dati circa la realizzazione delle tre strategie d'indagine

Supporto informatico per indagini preventive di potenziale diffusione di contagio tra bovini
Si segnala, infine, che l'uso della ricorsione ha agevolato indagare il contagio indiretto.

5. Dataset minimale e primi test sulle strategie d'indagine

Per poter validare le strategie d'indagine proposte si è fatto ricorso ad un dataset minimale "sufficientemente esaustivo" da poter diagnosticare *tutti e soli* i casi di contagio potenziale. Il dataset adottato è costituito da 10 lotti che effettuano 14 soste in 3 punti di stazionamento, dei quali il più grande è costituito da 7 recinti. La Tab.3 riporta i lotti potenzialmente contagiati, in accordo alle definizioni della Sez.2, ipotizzando malato il lotto che effettua il maggior numero di soste (codice L02).

Lotto malato	Contagio		
	Diretto	Indiretto	per Contiguità temporale
L02	L04, L05, L06	L03, L07, L09	L01, L08
SUBTOTALI	3	3	2
TOTALE	8		

Tabella 3 I contagi potenziali avvenuti tra i lotti del dataset minimale

La Tab. 4 riunisce i risultati delle tre strategie d'indagine al variare del lotto malato fra i 10 lotti del dataset minimale.

Lotto malato	Numero lotti potenzialmente contagiati secondo la	
	Strategia 1	Strategia 2, Strategia 3
L01	0	0
L02	9	8
L03	4	1
L04	7	4
L05	1	1
L06	1	1
L07	7	4
L08	2	1
L09	7	0
L10	7	4

Tabella 4 Risultati restituiti dalle tre strategie al variare del lotto malato

Essa, oltre a confermare quanto affermato dalla (1), è interessante perché mostra che anche quando sono in gioco pochi lotti vi possono essere scostamenti significativi sul numero dei lotti potenzialmente contagiati, restituiti dalla prima strategia e dalle altre due. Ciò si ripercuote direttamente sulla campagna di visite da espletare la cui durata è data dalla (4):

$$Ore_Veterinario = Numero_lotti \times Numero_Capi \times T_{VISITA} \quad (4)$$

ove Numero_lotti denota quelli potenzialmente contagiati, ciascuno costituito da Numero_Capi, mentre T_{VISITA} denota la durata media necessaria per visitare

un capo di bestiame. Le ore uomo sono un parametro molto importante perché si ripercuotono sulla tempestività con la quale si può confermare l'esistenza di un contagio in atto e quantificarne l'estensione (numero di capi effettivamente malati). Questi ultimi andranno curati o, se necessario, abbattuti, ma comunque non macellati.

Dovrebbe essere chiaro a questo punto il ruolo centrale di coloro cui spetta la responsabilità di scegliere la strategia d'indagine informatica più idonea alla criticità della malattia del lotto. Infatti, tale scelta impatta sul numero di lotti da dover visitare e quindi sui processi decisionali ed operativi che sarà necessario prendere. In questo senso si desidera richiamare il valore aggiunto della strategia 3 rispetto alla 2, la prima delle due, infatti, restituendo oltre ai lotti che potrebbero essere malati anche la probabilità che tale evento si manifesti, consente in casi critici, di limitare le visite a quei lotti per i quali tale valore di probabilità sia al di sopra di una certa soglia. Ad esempio, con riferimento alla Tab.4 nel caso il lotto malato sia L04 i risultati restituiti interrogando la base di dati sono i seguenti: L02 (0.9), L07 (0.9), L10 (0.75) ed L03 (0.5).

6. Conclusioni

Si è presentata una soluzione metodologica e tecnologica idonea a consentire, in seguito all'esito negativo di un controllo su di un lotto di bestiame, di avviare tempestivamente un'indagine con strumenti informatici circa i lotti che sono entrati in contatto con esso e che, quindi, potrebbero essere stati contagiati e pertanto dovranno essere visitati prima possibile da un veterinario. Anche se la soluzione informatica ha meriti oggettivi, tra i quali il suo bassissimo costo, non sfugge che il verdetto definitivo e non opinabile circa lo stato di salute del bestiame lo si avrà solo al termine della campagna di visite sui lotti potenzialmente contagiati.

Come sempre accade quando si propone una metodologia, la sua accettazione va comunque subordinata ad una fase di attenta e prolungata sperimentazione su dati reali onde quantificare gli scostamenti medi tra il numero di capi dichiarato potenzialmente contagiato e quanto riscontrato sul campo dai veterinari.

Bibliografia

[1] Regolamento N. 1/2005 del Consiglio dell'UE del 22.12.2004, "Sulla protezione degli animali durante il trasporto e le operazioni correlate *che modifica le Direttive 64/432/CEE e 93/119/CE ed il Regolamento (CE) n. 1255/97*". Pubblicato sulla Gazzetta ufficiale dell'UE, serie L, n.3, 1.12.2005.

[2] Regolamento N. 1255/97 del Consiglio dell'UE del 25.6.1997 riguardante "i criteri comunitari per i posti di controllo". Pubblicato sulla Gazzetta ufficiale dell'UE, serie L, n.174, del luglio 1997.

[3] www.salute.gov.it

[4] www.postgresql.org

Sviluppo di un servizio per lo Smart & Green building management a partire dall'esperienza del Politecnico di Torino

Piero BOZZA
Politecnico di Torino
C.so Duca degli Abruzzi, 24 10129 TORINO
piero.bozza@polito.it
Sante Lorenzo Carbone
CSI-Piemonte
Corso Tazzoli 215/12b, 10134 Torino
Santelorenzo.carbone@csi.it

Le tecnologie e le soluzioni oggi disponibili per la riduzione dei consumi energetici negli edifici di medie e grandi dimensioni sono numerose e consolidate. Spaziano dagli interventi di riqualificazione dell'involucro alla revisione parziale o totale degli impianti. Si tratta di interventi invasivi che permettono significativi risparmi a fronte di investimenti elevati con tempi di ritorno spesso di parecchi anni. Un approccio "soft" e complementare che può dare benefici significativi con costi di investimento e tempi di ritorno ridotti, consiste nell'adozione di tecnologie ICT non invasive in grado di supportare il monitoraggio costante dei consumi e rendere disponibili in modo tempestivo informazioni utili per ottimizzare la gestione degli impianti o per responsabilizzare e sensibilizzare le persone che operano e lavorano nei locali e spazi dell'organizzazione. L'applicazione estensiva dell'approccio soft al monitoraggio degli edifici e dell'uso sostenibile delle relative risorse fisiche può costituire la base per lo sviluppo di servizi di supervisione e gestione intelligente di patrimoni immobiliari quali università, enti e realtà del territorio.

1 Introduzione

Il Politecnico di Torino è una realtà universitaria pubblica coinvolta a livello nazionale ed internazionale in numerosi progetti di ricerca nell'ambito del risparmio energetico, della eco-compatibilità delle soluzioni tecnologiche, della sostenibilità dei processi di progettazione e realizzazione nei vari campi dell'ingegneria e dell'architettura.

Nell'ambito specifico del risparmio energetico applicato ai sistemi edificio-impianto l'Ateneo torinese si presenta come un'organizzazione complessa costituita da edifici realizzati in periodi molto diversi che vanno dal Castello del

Valentino di epoca seicentesca, già residenza dei reali di casa Savoia, alla recentissima sede di Mirafiori (in classe A+) dotata delle più moderne soluzioni applicate all'involucro, agli impianti ed alla supervisione intelligente.

Si tratta di edifici con problematiche e situazioni impiantistiche spesso eterogenee accomunati dalla presenza di un numero significativo di "occupanti" che vanno dagli studenti ai ricercatori, al personale distribuiti in unità organizzative e centri di responsabilità ben definiti. Ulteriore elemento caratteristico degli edifici dell'Ateneo è costituito dalla presenza capillare della rete dati wireless a disposizione degli studenti e del personale.

Caratteristiche molto simili a quelle del Politecnico possono essere ritrovate in molte delle realtà pubbliche presenti sul territorio (scuole, università, enti, ospedali).

Un problema comune a queste realtà è la consapevolezza della necessità di dover intervenire sui propri edifici ed impianti per razionalizzarne la gestione e per aumentarne l'efficienza energetica. Altrettanto comune è la consapevolezza di dover operare con risorse per gli investimenti limitate che non permette di attivare azioni concrete di riqualificazione significative.

In queste situazioni le tecnologie ICT, applicate al monitoraggio integrato degli impianti e degli edifici, possono diventare una valida ed economica risorsa a livello diagnostico per individuare le criticità, supportare la selezione degli interventi più redditizi e a minor costo di investimento. Si parla infatti di ICT per l'efficienza energetica – ICT4EE.

L'integrazione dei sistemi di monitoraggio con strumenti di comunicazione efficaci quali portali web, messaggistica SMS/mail indirizzati agli occupanti, permette di portare all'utente finale, così come ai responsabili di unità organizzative o all'Energy Manager le informazioni rilevanti utili per indurre comportamenti in grado di ridurre e razionalizzare i consumi energetici (user awareness).

A differenza delle soluzioni tecnologiche di controllo invasive che richiedono la modifica degli impianti, l'approccio basato sulla fornitura di informazioni agli utenti in grado di coinvolgerli ed "attivarli" verso azioni di risparmio energetico, considera la persona come l'attuatore del controllo: si tratta di fornirgli le giuste informazioni in modo che agisca nel modo corretto e di incentivare il suo comportamento nella giusta direzione. In queste situazioni si può parlare a tutti gli effetti di sistema edificio-impianto-utente.

Questi sono i principali presupposti sulla base dei quali il Politecnico ha realizzato negli ultimi anni specifici progetti finalizzati a promuovere lo sviluppo di servizi informativi per il monitoraggio energetico e per la diffusione di informazioni e strumenti utili alla sensibilizzazione dei propri utenti verso comportamenti sostenibili e maggiormente efficienti nell'utilizzo delle risorse dell'Ateneo.

In particolare il progetto Wifi4Energy, che verrà descritto di seguito, ha permesso di realizzare e sperimentare un vero e proprio prototipo di "sistema informativo utente" utilizzabile ed accessibile sia attraverso le tecnologie web che tramite un laboratorio dimostratore (Living LAB) che opera come centrale unitaria di monitoraggio delle energie.

Sulla base dell'esperienza acquisita dal Politecnico nel corso degli ultimi progetti, sarà prossimamente possibile estendere l'utilizzo del Living LAB e della centrale unitaria di monitoraggio per la supervisione intelligente degli edifici e delle risorse scarse in essi contenute (es. le aule), realizzando un sistema/servizio integrato per la gestione sostenibile ed intelligente degli immobili (Smart & Green Building management services).

In questa direzione il Politecnico sta collaborando da luglio 2011 con il CSI Piemonte, attraverso l'accordo di collaborazione "Smart Building e Energy Management", per lo sviluppo sinergico di sistemi e servizi di gestione Smart & Green degli edifici che possano diventare un valido supporto agli enti sul territorio, in coerenza con la vocazione sempre più marcata della nostra Regione di investire in tecnologie che favoriscano la sostenibilità, l'efficienza energetica, la gestione intelligente degli immobili, fino alla gestione smart & green di intere aree urbane (smart cities).

2 Il progetto WiFi4Energy

Tra le varie iniziative realizzate negli ultimi anni dal Politecnico è utile approfondire quella del progetto WiFi4Energy i cui obiettivi, riportati di seguito, sono stati fortemente incentrati sul rafforzamento di strumenti e sistemi per la gestione sostenibile delle risorse energetiche:

- sperimentazione dell'utilizzo delle reti Wireless Sensor Network (WSN) finalizzate al monitoraggio dei consumi energetici [Fracastoro, 2010] integrandole con la rete WiFi presente nel Politecnico;
- realizzazione di una infrastruttura integrata di monitoraggio e di supporto al controllo ambientale finalizzata al risparmio energetico negli edifici [Bozza et al, 2010];
- sviluppo di strumenti per la sensibilizzazione e responsabilizzazione del personale / utenza degli edifici;
- realizzazione del Living LAB, un laboratorio dimostrativo dei benefici del monitoraggio di ambienti, sistemi ed impianti a supporto dell'uso sostenibile delle risorse energetiche;
- realizzazione del portale WEB per il monitoraggio e per la diffusione delle informazioni alle diverse tipologie di utenze;
- introduzione di strategie, logiche e strumenti di contabilizzazione analitica dei consumi per centri di costo ("bolletta virtuale").

Tra le specifiche di progetto va evidenziata la volontà di sviluppare soluzioni in logica non proprietaria ed i cui risultati potessero essere "riusati" da altri enti e soggetti pubblici sul territorio.

Il progetto ha coinvolto ricercatori dei Dipartimenti di Energetica, Ingegneria Elettronica, Ingegneria Elettrica, Ingegneria Automatica e Informatica integrati con specialisti nella gestione degli impianti del Servizio Edilizia e Logistica e nella gestione dei Sistemi Informativi del Centro di Servizi Informatici e Telematici.

Il progetto è stato realizzato con il cofinanziamento della Regione Piemonte.

2.1 La sperimentazione delle reti WSN

Uno degli obiettivi centrali del progetto è consistito nella sperimentazione di sensori, installabili in modo non invasivo, basati su protocolli di comunicazione non proprietari. I sensori dovevano essere idonei ad acquisire i dati provenienti dagli impianti e ambienti e renderli disponibili, attraverso la rete dati wireless e LAN del Politecnico, presso una control room in grado di centralizzare la raccolta e la storicizzazione di tutte le misurazioni effettuate sul campo.

La sperimentazione è stata finalizzata al monitoraggio ambientale e del confort (temperatura operativa interna [T], umidità relativa [RH] e illuminamento [LUX]) di locali in edifici di medio-grandi dimensioni. Si è pertanto focalizzata sull'utilizzo di sensori e reti di sensori WSN (Wireless Sensor Network) in grado di essere configurati secondo una logica gerarchica multi-tree [Gnawali et al, 2009] nella quale sono presenti molti nodi multi-sensore connessi ad elementi di acquisizione dati intermedi costituiti da unità intelligenti dette base station [Apiletti et al, 2010a]. L'adozione di sensori wireless introduce una semplificazione drastica in termini di allestimenti impiantistici ma presenta problematiche non marginali di affidabilità, stabilità della misura, continuità di funzionamento e performance [Sanchez et al, 2009]. In particolare la sperimentazione con nodi innovativi della famiglia Telos (TelosB Mote) sviluppati dall'Università di Berkeley in collaborazione con Intel Research e funzionanti su protocollo IEEE 802.15.4 ha evidenziato seri limiti di autonomia [Polastre et al, 2005] operativa con tempi di vita delle batterie variabili da 5 a 200 giorni). In particolare la sperimentazione ha messo in luce la difficoltà di garantire un'autonomia adeguata della rete in condizioni di comunicazione near real time dei dati dal sensore alla base station [Polastre et al, 2004]. Problematiche simili sono emerse anche con l'utilizzo di sensori custom.

La sperimentazione è stata condotta su diverse aree dell'ateneo presso la sede centrale di C.so Duca degli Abruzzi di Torino, caratterizzate da una specifica destinazione d'uso (aule, uffici, laboratori, locali comuni) e da diverse esposizioni (nord, sud, facciate vetrate o opache, ecc.).

E' stata condotta una specifica azione di monitoraggio tramite rete WSN sull'unità di trattamento aria evoluta (UTA) di una grande aula (Aula 1) al fine di ottimizzarne i consumi e di individuare un modello operativo in grado di rilevare con buona approssimazione il livello di occupazione dell'aula.

Si riporta di seguito (vedi Fig.1) lo schema dell'architettura realizzata per l'integrazione delle reti di sensori WSN con la rete dati LAN Polito.

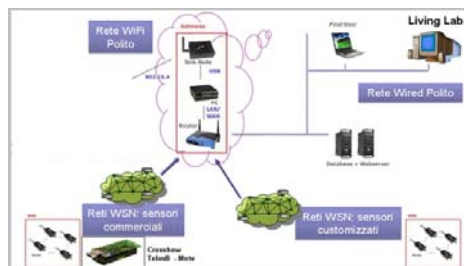


Fig.1: architettura di integrazione delle reti WSN con la rete dati LAN Polito

2.2 Il sistema integrato di monitoraggio (SIM)

Uno degli aspetti più rilevanti che ostacola la possibilità di mettere a disposizione degli utenti che lavorano e operano negli edifici le informazioni salienti sui consumi energetici, è dovuto alla difficoltà di interfacciare i sistemi di supervisione proprietari connessi agli impianti con applicazioni ad alto livello che vengono rese disponibili agli utenti finali (es. un portale WEB). Le organizzazioni reali inoltre, molto spesso dispongono di più impianti eterogenei, sviluppati in tempi diversi e con tecnologie diverse, da fornitori diversi, che rendono ulteriormente complessa e onerosa un'azione di integrazione.

Il progetto ha pertanto affrontato questa complessa problematica sviluppando un sistema integrato di monitoraggio basato su un'architettura modulare multilivello nella quale i singoli componenti possono integrarsi con realtà impiantistiche eterogenee o con diverso livello di maturazione. Si riporta di seguito lo schema logico del SIM (vedi Fig.2):

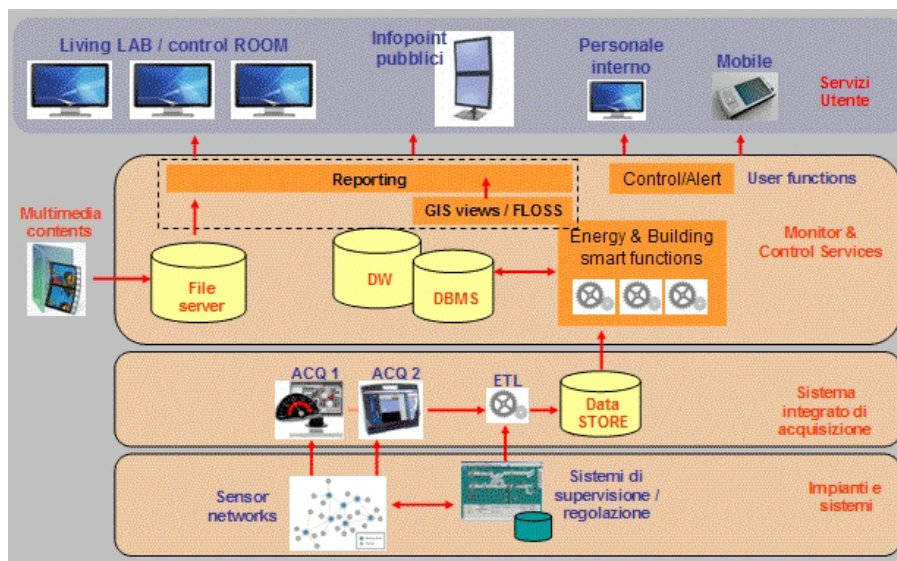


Fig.2: Sistema Integrato di Monitoraggio (SIM): schema logico e componenti

Secondo l'architettura definita i principali componenti del sistema SIM sono:

- **l'impianto tecnologico esistente** (es. elettrico, illuminazione) spesso già dotato di contatori o di sensori per l'acquisizione dei consumi o di altri dati ambientali in specifiche zone/aree/fabbricati;
- **il sistema di supervisione** che effettua il monitoraggio e la regolazione sull'impianto che costituisce l'intelligenza "locale" del sistema e permette la gestione, supervisione e ottimizzazione dei consumi specifici;
- **il sistema integrato di acquisizione** che ha il compito di interfacciare i diversi impianti con il database DBMS centrale attraverso "connettori". In

funzione delle caratteristiche del singolo sistema di supervisione il connettore permette di definire quali dati acquisire, le modalità di acquisizione, la frequenza di acquisizione, i range di validità delle misure e altri parametri specifici del flusso di comunicazione con l'impianto;

- **Il sistema ETL** (extract transform and load) ha il compito di acquisire il dato grezzo proveniente dall'impianto e di attuare una prima strategia di consolidamento del dato nel datastore;
- Il componente **Energy & building smart functions** costituisce il motore logico di più alto livello nel quale vengono effettuate le operazioni di analisi dei dati, di correlazione, di confronto tra sistemi e impianti, di elaborazione per la successiva presentazione di informazioni aggregate per gli utenti finali;
- Il motore può essere corredato di un **modulo aggiuntivo Control/Alert** che ha il compito di generare eventi di notifica, allarme o controllo sulla base di condizioni e regole predefinite;
- Il motore deve inoltre poter **integrare contenuti informativi multimediali** di tipo eterogeneo (es. filmati, presentazioni) per cui è stato previsto un modulo in grado di acquisire questi dati per la successiva presentazione e comunicazione agli utenti attraverso i canali ritenuti più opportuni (web, smartphone, tablet, SMS, mail).

L'insieme dei componenti software descritti, ed in particolare quelli appartenenti ai tre livelli "sistema integrato di acquisizione", "Monitor & control services" e "servizi utente" costituisce un'applicazione complessa ma modulare che rappresenta il Sistema Informativo Utente (SIU).

2.3 Il modello logico dei dati

Il database DBMS costituisce il cuore del sistema informativo di monitoraggio e controllo. E' stato progettato sulla base di un modello logico di dati in grado di rappresentare realtà dell'edificio/impianto molto eterogenee e di diversa dimensione. E' pertanto in grado di acquisire dati provenienti da più siti/sedi/edifici, da più reti di sensori, da più zone. Può gestire qualsiasi tipo di sensore e qualsiasi grandezza da monitorare (temperature, umidità relativa, illuminamento, potenza termica, frigorifera ed elettrica, portata, pressione, stato di presenza/assenza, ecc.). La localizzazione dei sensori può essere integrata con sistemi di facility management o di georeferenziazione.

E' stato inoltre implementato uno schema logico dei dati in grado di rappresentare i sensori come nodi di reti correlabili nello spazio e nel tempo [Baralis et al, 2007] adottando il framework SeReNe [Apiletti et al, 2010] (Selective Representative in Sensor Networks). SeReNe ha lo scopo di ridurre i consumi dei nodi presenti in reti di sensori attraverso l'ottimizzazione delle comunicazioni in fase di trasmissione e ricezione dei dati.

2.4 Il Living LAB

Il processo che porta alla sensibilizzazione ed alla responsabilizzazione degli utenti di una organizzazione verso comportamenti energeticamente virtuosi ed in grado di contenere i consumi, passa attraverso la comunicazione mirata e la diffusione delle informazioni in modo da far crescere progressivamente la cultura della sostenibilità nelle persone.

Tra gli strumenti ritenuti importanti per raggiungere questi scopi di coinvolgimento ed educazione alla sostenibilità energetica vi è quindi il laboratorio dimostratore che il progetto WiFi4Energy ha realizzato e denominato Living LAB.

Si tratta di uno spazio di laboratorio dedicato a "dimostrare" a gruppi di utenti interni ma anche a soggetti esterni pubblici e privati, le potenzialità degli strumenti di monitoraggio integrato e di controllo ed i benefici realizzati dal Politecnico. Presso il Living LAB vengono anche centralizzate tutte le funzioni di supervisione degli impianti e sistemi, accessibili attraverso un portale WEB, così come la strumentazione per l'analisi energetica e per la produzione di reporting, cruscotti ed ogni altro supporto utile per perseguire le azioni finalizzate al risparmio e all'ottimizzazione delle risorse.

Il Living LAB è operativo da fine 2010 presso la sede di C.so Duca degli Abruzzi.

2.5 Il portale WEB Living LAB

Il portale WEB realizzato è stato denominato Portale Living LAB (pLL) in quanto rappresenta anche lo strumento principale per il monitoraggio e la supervisione che vengono gestiti attraverso il laboratorio dimostratore e la control room.

Il portale presenta alcune sezioni informative generali di interesse pubblico relative agli edifici ed in impianti presenti nel Politecnico rendendo disponibili linee guida di carattere generale e guide operative per specifici ambiti di applicazione del contenimento dei consumi. Veicola tutte le funzioni informative, grafici e report di andamento e confronto dei consumi energetici o qualsiasi altro contenuto multimediale attraverso un normale browser (http://www.swas.polito.it/services/w4e_living_lab/).

2.6 Ambiti di monitoraggio e casi pratici

Il portale è suddiviso in quattro sezioni principali che fanno riferimento ai quattro grandi ambiti di monitoraggio: energia, acqua, ambienti interni, ambiente esterno. All'interno di ogni ambito vengono presentati i cruscotti di monitoraggio disponibili e utili all'utente.

Si riportano di seguito alcuni casi significativi di monitoraggio che vengono effettuati attualmente dal sistema SIM-SIU e resi disponibili sul pLL.

- **Monitoraggio dei consumi elettrici** per fasce F1-F2-F3: il sistema monitora e presenta i consumi (kWh) che avvengono in fascia diurna e notturna permettendo di sensibilizzare l'utenza affinché si attivi per "spostarli", quando possibile, in fascia notturna a minor costo. Il monitoraggio viene aggiornato con frequenza di 15';

- **Monitoraggio dei consumi idrici:** la supervisione e presentazione opera in modo analogo a quanto avviene per i consumi elettrici;
- **Controllo dell'accensione e spegnimento dei PC** da parte dell'utente: nell'ambito del progetto WiFi4Energy è stato messo a punto un sistema di monitoraggio e controllo dell'accensione/spegnimento dei PC dell'utente. Il sistema è stato progettato basandosi sui risultati del progetto di ricerca PoliSAVE (<http://polisave.polito.it/>) realizzato dal Dipartimento di Elettronica;
- **Monitoraggio della CO2 per la rilevazione dell'utilizzo delle aule:** il monitoraggio fa parte dell'ambito di supervisione degli ambienti interni e rappresenta una sperimentazione innovativa, basata su modelli teorici e sperimentali consolidati [Jae-Weon Jeong, 2010] di un sistema a basso costo per la rilevazione dello stato di utilizzo ed occupazione di grandi locali (es. le aule). Le aule sono infatti spazi "pregiati" in una università, il cui utilizzo ottimizzato può contribuire ad abbattere i costi di gestione (riscaldamento, raffrescamento, pulizie, ecc.).

La capacità del SIM-SIU è scalabile ed estendibile su nuove reti di sensori o nuovi impianti che vengano introdotti in Ateneo. L'interfacciamento tramite nuovi connettori costituisce l'elemento di integrazione e va progettato ad hoc per nuovi impianti.

3 Conclusioni e sviluppi futuri

Il progetto WiFi4Energy ha permesso di effettuare una ampia sperimentazione sul campo di sensori, reti di sensori, tecnologie innovative per il monitoraggio energetico ed ambientale, mettendo in evidenza i limiti ancora da superare per l'applicazione di reti di sensori wireless basati su protocolli aperti. Ha permesso al contempo di realizzare un prototipo funzionante dell'infrastruttura di servizi e strumenti per la supervisione integrata di edifici, ambienti e impianti utile al monitoraggio e controllo dei consumi energetici e dell'utilizzo di risorse scarse dell'Ateneo.

La nuova sede della Cittadella del Design e della mobilità (Campus di Mirafiori) recentemente inaugurata in ottobre 2011, costituisce un primo significativo ambito di estensione e potenziamento del prototipo realizzato. Grazie alla recente collaborazione avviata con il CSI Piemonte la nuova sede verrà utilizzata per sviluppare e sperimentare una piattaforma integrata di gestione Smart & Green sulla base dell'esperienza fatta con il prototipo dal Politecnico. La piattaforma verrà progressivamente applicata alle altre sedi dell'Ateneo e integrata con il Living LAB costituendo un sistema/servizio applicabile anche in altri contesti e organizzazioni sul territorio.

L'accordo di collaborazione Politecnico-CSI Piemonte permetterà inoltre di sviluppare servizi basati sulla piattaforma tecnologica e favorire nelle PA l'introduzione di sistemi di governance attraverso un utilizzo intelligente delle informazioni disponibili sui propri edifici.

Il tema della governance di sistemi complessi si sta infatti dimostrando sempre più centrale per un utilizzo efficace e sinergico delle risorse sul

territorio. Tra queste rientrano certamente gli edifici e la necessità stringente di un loro utilizzo più efficiente e sostenibile attraverso il costante monitoraggio e controllo dei consumi energetici, del corretto e pieno utilizzo secondo le reali necessità, dei sistemi di sicurezza e delle reti tecnologiche.

Si sintetizzano di seguito i principali obiettivi dell'accordo di collaborazione Politecnico-CSI Piemonte:

- realizzazione di un sistema modulare in grado di integrarsi con i sistemi tecnologici presenti negli enti e nelle PA locali per garantire funzionalità ad alto livello di monitoraggio e gestione dei consumi energetici o di altre risorse ritenute significative;
- implementazione di un servizio di supporto da rendere disponibile agli enti ed alle PA basato sul sistema integrato, erogabile in modo semplice e diretto anche presso utenze che non dispongano di competenze specifiche nell'ambito dell'efficientamento energetico o del facility management;
- favorire la crescita di un centro di competenze in grado di indirizzare soluzioni e azioni di efficientamento e razionalizzazione su scala multi-ente e territoriale;
- creare attraverso un primo sistema realmente utilizzabile in ambito energetico e sugli edifici, le condizioni per lo sviluppo di ulteriori moduli di monitoraggio e supervisione di altre reti e infrastrutture presenti presso gli enti e sul territorio;
- favorire l'interoperabilità tra enti e territorio a partire da informazioni condivise consolidate nei sistemi di monitoraggio e supervisione;
- creare un'infrastruttura in grado di integrare in modo semplice nuove reti e tipologie di oggetti monitorabili (internet of things) di rilevanza per il territorio;
- favorire la crescita della cultura della sostenibilità da parte degli operatori e utenti degli enti e soggetti del territorio attraverso l'utilizzo di strumenti e buone pratiche nella vita quotidiana.

4 Bibliografia

[Apiletti et al, 2010] Apiletti D., Baralis E., Cerquitelli T., Energy saving models for wireless sensor networks, in Knowledge and Information Systems, KAIS international journal, Springer, London, 2010.

[Apiletti et al, 2010a] Apiletti D., Baralis E., Cerquitelli T., Chiusano S., Montrucchio B., Murillo L. M., Rebaudengo M., Sanchez E. R., Tonelli D., Exploiting Wireless Sensor Networks for Monitoring Building Performance, in Proc. of 48th Congresso Annuale Associazione Italiana per il Calcolo Automatico (AICA), L'Aquila, Italy, 29-30 September 1 October 2010, 2010.

[Baralis et al, 2007] Baralis, E., Cerquitelli, T., D'Elia, V.: Modeling a sensor network by means of clustering, in DEXA '07 proceedings of the 18th International Conference on Database and Expert Systems Applications, IEEE Computer Society, Washington, DC, USA, 2007, 177–181.

[Bozza et al, 2010] Bozza P., Blaso L., Fracastoro G.V., L'ICT ed i consumi energetici: il Progetto WiFi4Energy, Torino Software and Systems Meeting - TOSM 2010, Seminario: "Efficienza energetica e Green ICT", Torino, Italy, 26 Maggio, 2010.

[Fracastoro, 2010] Fracastoro G.V., WiFi4Energy - Wireless Sensors Network for Energy Management in Educational Buildings, Convegno AEIT "Building Management and Domotics - Realtà e prospettive", Torino, Italy, 31 Marzo, 2010.

[Gnawali et al, 2009] Gnawali O., Fonseca R., Jamieson K., Moss D., Levis P., Collection tree protocol, In Proc. of the 7th ACM Conference on Embedded Networked Sensor Systems, Berkeley, California, November 04 - 06, 2009), SenSys '09. ACM, New York, NY, 1-14.

[Jae-Weon Jeong, 2010] Jae-Weon Jeong, Anseop Choi, Sang-Tae, Improvement in demand-controlled ventilation simulation on multi-purposed facilities under an occupant based ventilation standard simulation Modelling Practice and Theory, 18, 2010, 51–62.

[Polastre et al, 2004] Polastre J., Hill J., Culler D., Versatile low power media access for wireless sensor networks, in Proceedings of the 2nd international Conference on Embedded Networked Sensor Systems, Baltimore, MD, USA, November 03-05, 2004, SenSys '04. ACM, New York, NY, 95-107.

[Polastre et al, 2005] Polastre J., Szewczyk R., Culler D., Telos: Enabling Ultra-Low Power Wireless Research, in IPSN '05 Proceedings of the 4th International Symposium on Information Processing in Sensor Networks, Piscataway, NJ, USA: IEEE Press, 2005, p.48.

[Sanchez et al, 2009] Sanchez E. R., Chaudet C., Rebaudengo M., Improving preamble sampling performance in wireless sensor networks with state information, in Proceedings of the Sixth international Conference on Wireless on-Demand Network Systems and Services, Snowbird, Utah, USA, February 02-04, 2009. IEEE Press, Piscataway, NJ, 93-100.

Tecniche Model Driven nell'Industria Italiana: Risultati di un'indagine e condivisione di esperienze

Marco Torchiano¹, Federico Tomassetti¹, Filippo Ricca², Alessandro Tiso²,
Gianna Reggio²

¹Politecnico di Torino
C.so Duca degli Abruzzi 24, 10129 Torino
nome.cognome@polito.it

²DISI, Università di Genova
Via Dodecaneso 35, 16146 Genova
nome.cognome@disi.unige.it

Le tecniche di sviluppo del software Model-Driven sono note da alcuni anni e una recente indagine ha dimostrato una loro ampia diffusione nell'industria italiana. La proposta di workshop illustrata in questo documento vuole mettere in comunicazione le aziende interessate in tecniche MD e le realtà accademiche che conducono attività di ricerca e di trasferimento tecnologico su tali tematiche.

1. Introduzione

Il model-driven engineering (MDE) è una tecnica di sviluppo del software fondata sul concetto di modello e definita ormai da diversi anni. Il suo scopo è innalzare il livello di astrazione dal codice ai modelli: in pratica il codice è generato dai modelli tramite trasformazioni automatiche, oppure il modello viene eseguito direttamente.

I principali vantaggi tipicamente ascritti all'approccio MDE sono incrementi in termini di produttività, portabilità, manutenibilità ed interoperabilità [Kleppe et al. 2003]. Tali promesse erano tali circa 10 anni fa quando l'approccio è stato concepito. Ora è importante verificare se tali promesse siano state mantenute e recepite, e se sì in che misura, nell'adozione di tecniche MD nel contesto dell'industria italiana.

Con tale obiettivo gli autori di questa proposta di workshop hanno progettato e condotto un'indagine di larga scala raccogliendo 155 questionari, compilati da professionisti dell'IT, sull'adozione di tecniche MD nelle aziende italiane [Torchiano et al. 2011]. In base a questa rilevazione, il 68% delle aziende utilizza diverse forme di modellazione e tra loro un largo numero usa tecniche MD (48%). Tali proporzioni sono sicuramente degne di attenzione. Altri risultati interessanti riguardanti strumenti, processi utilizzati e maturità nell'adottare MD sono emersi dall'analisi dei questionari raccolti.

2. Obiettivi

Il workshop si pone diversi obiettivi:

- presentazione dei risultati dell'indagine su tecniche MD nell'industria italiana,
- condivisione di esperienze di utilizzo di MD da parte delle aziende,
- identificazione dei fattori di successo/insuccesso dei progetti MD,
- identificazione delle problematiche nell'adozione.

3. Organizzazione/Formato

Il workshop sarà strutturato in tre parti:

1. Presentazione dei risultati del survey da parte dei proponenti del workshop (unità operative di Torino e Genova).
2. Presentazione delle esperienze industriali da parte dei partecipanti industriali che intendono condividere la propria esperienza nell'utilizzo di tecniche MD.
3. Discussione finale.

Diverse aziende del settore saranno direttamente invitate a presentare le proprie esperienze nell'utilizzo di tecniche MD. Gli interessati industriali che intendono condividere con i partecipanti la propria esperienza dovranno presentare agli organizzatori del workshop un riassunto dell'intervento che intendono fare in un position paper di una pagina al massimo. Solo le proposte più interessanti e in linea con gli scopi del workshop saranno accettate.

Il workshop verrà pubblicizzato: (1) a tutti i partecipanti del indagine industriale condotta e riassunta in [Torchiano et al. 2011] mandando via mail inviti mirati e (2) su diverse mailing list riguardanti il software development e il software engineering.

4. Risultati attesi

I risultati che gli organizzatori del workshop si pongono sono:

- creazione di un'embrionale rete di conoscenza focalizzata sulle tecniche MD che possa essere sfruttata in futuro per attività di trasferimento tecnologico, scambio di conoscenza, e ricerca sul campo,
- definizione di un quadro più preciso delle opportunità e dei rischi presentati dall'adozione di un approccio MD sia da un punto di visto organizzativo/gestionale che da un punto di vista tecnologico.

5. CV dei proponenti

- Marco Torchiano è ricercatore presso il Politecnico di Torino, è stato post-doc presso la Norwegian University of Science and Technology (NTNU), Norvegia. Ha conseguito la laurea magistrale e il dottorato

in Ingegneria Informatica presso il Politecnico di Torino. È autore di oltre 80 articoli di ricerca pubblicati in riviste e conferenze internazionali. È coautore del libro *Software Development—Case studies in Java* edito da Addison-Wesley, e curatore del libro *Developing Services for the Wireless Internet* edito da Springer. I suoi principali interessi di ricerca attuali riguardano le notazioni di progetto, le metodologie di test, lo sviluppo basato su componenti Off-The-Shelf, e l'ingegneria del software per dispositivi mobili. Si accosta a tali temi con l'approccio dell'ingegneria del software empirica.

- Federico Tomassetti è studente di dottorato in ingegneria informatica e dei sistemi, ha esperienza come sviluppatore e autore di articoli tecnici.
- Filippo Ricca è ricercatore presso il Dipartimento di Informatica e Scienze dell'Informazione (DISI) dell'Università di Genova, dove ha conseguito la laurea nel 1994 e il dottorato di ricerca nel 2003 e dove è attualmente docente di Ingegneria del Software. La sua attività di ricerca verte principalmente su ingegneria del software empirica, metodi di sviluppo, software testing e SOA. È autore di più di settanta articoli su riviste e conferenze internazionali.
- Alessandro Tiso è studente di dottorato presso il Dipartimento di Informatica e Scienze dell'Informazione (DISI) dell'Università di Genova. Ha conseguito la Laurea in Ingegneria Elettronica presso l'università di Genova nel 1994 dove ha superato l'esame di stato per la professione di Ingegnere. Dal 1989 è insegnante di ruolo di Laboratorio di Informatica Industriale. Ha partecipato a diversi progetti industriali in qualità di consulente.
- Gianna Reggio è Professore associato presso l'Università di Genova dal 1992. Precedentemente è stata Ricercatore presso la stessa università ed ha conseguito il Dottorato di Ricerca in Informatica nel 1988. Ha preso parte a vari progetti di ricerca, tra i quali SAHARA e SALADIN (PRIN) ed è stata co-organizzatrice di molte conferenze e workshop internazionali, tra cui ETAPS 2001 e MODELS 2006. È stata membro del comitato di programma di varie conferenze internazionali, tra cui MODELS (edizioni 2009, 2008, 2007). L'attività di ricerca si è concentrata principalmente sui seguenti temi: Metodi di sviluppo del software, Modellazione e specifica formale di sistemi concorrenti e Semantica dei linguaggi di programmazione. Recentemente ha iniziato ad occuparsi di metodi di sviluppo per applicazioni basate su SOA. È autore di numerose pubblicazioni su riviste scientifiche, conferenze nazionali ed internazionali e monografie.

Bibliografia

[Kleppe et al. 2003] A. G. Kleppe, J. Warmer, and W. Bast. MDA Explained: The Model Driven Architecture: Practice and Promise. Addison- Wesley Longman Publishing Co., Inc, 2003.

[Torchiano et al. 2011] M. Torchiano, F. Tomassetti, F. Ricca, A. Tiso, G. Reggio, Preliminary findings from a State of the Practice Survey in MD*, Proc. International Conference of Empirical Software Engineering and Measurement (ESEM), Banff, Canada, 2011.

Tecnologie semantiche per l'analisi della reputazione online delle destinazioni turistiche

Luisa Mich, Mariangela Franch, Nicola Bortoletto¹, Pamela Negrosanti¹

Dipartimento di Informatica e Studi Aziendali, Università di Trento

Via Inama 5, 38122 Trento

{luisa.mich, mariangela.franch}@unitn.it

¹*Cogito srl*

Via Zeni 8, 38068 Rovereto (Tn)

{nbortoletto, pnegrosanti}@expertsystem.it

L'analisi della reputazione online per quanti operano nel settore del turismo sta diventando un'attività critica per il successo delle strategie di promozione e commercializzazione. La varietà dei siti web su cui gli utenti possono condividere esperienze ed opinioni e il peso che gli User Generated Content (UGC) pubblicati hanno nelle decisioni dei turisti richiedono strumenti in grado di individuare ed analizzare efficacemente grandi moli di fonti e contenuti. Focalizzando sugli UGC di tipo testuale, tecnologie linguistiche semantiche possono essere applicate per supportare un'analisi integrata della reputazione di una destinazione turistica. Per illustrare l'approccio proposto si fa riferimento a uno studio esplorativo realizzato per il Trentino, una destinazione con un'offerta turistica matura e diversificata. Oltre ad ottenere indicazioni importanti sulla reputazione online, evidenziando un 'sentiment' dei turisti complessivamente positivo, l'approccio semantico ha permesso di segnalare aspetti specifici su cui intervenire per migliorarlo, un input utile per i sistemi di Customer Relationship Management delle destinazioni.

1. Introduzione

Gli sviluppi recenti del Web, indicati con la denominazione Web 2.0 [O'Reilly, 2005], si caratterizzano per un ruolo predominante degli utenti, che contribuiscono in maniera rilevante alla creazione di nuovi contenuti, o UGC (*User Generated Content*). Tali contenuti vengono pubblicati su blog o wiki, condivisi su piattaforme per il *social networking*, sono frequentemente accessibili anche da dispositivi mobili e si diffondono con velocità nella *websfera*. In alcuni casi si tratta di recensioni o valutazioni relative a servizi specifici, come avviene ad esempio sui siti delle OTA (*Online Travel Agency*), per gli alberghi o i ristoranti; in altri casi sono commenti pubblicati nei blog di

un'azienda o in quello personale; micro-messaggi, spesso usati per segnalare eventi o informazioni utili, propagando rapidamente le notizie; preferenze e gusti riguardo a specifici prodotti, sono frequenti anche nei siti 'generalisti' (ad esempio, in FaceBook). In questo contesto, da un lato si registrano percentuali crescenti di iscritti ai siti di *social networking* più famosi (si vedano, ad esempio, le statistiche di Alexa, www.alexacom.com e Social Networking Watch, www.socialnetworkingwatch.com) e dall'altro le aziende stanno acquistando maggior consapevolezza dell'impatto della loro reputazione online. Quest'ultimo aspetto trova riscontro nella pubblicazione di articoli sul tema sui quotidiani più autorevoli (a titolo di esempio si possono citare [Pattison, 2009], [Gibson, 2010]), ma anche nell'ampia offerta di servizi di on-line (o web) *reputation analysis* e *monitoring*.

Il giudizio complessivo che emerge online per un prodotto o un servizio viene indicato anche come '*sentiment*' e la sua analisi permette all'azienda di ricavare informazioni utili in termini di *customer relationship management*, sia per comunicare in modo efficace con i clienti (gestione del *complaint* digitale e promozione da parte dei clienti soddisfatti), che per intervenire su prodotti o servizi in modo mirato, e in ultima istanza per le strategie di qualità [Juran e Gryna, 1980], [Juran, 1995]. Per una destinazione turistica, un'analisi costante e mirata dei contenuti pubblicati sul Web che la riguardano può rappresentare un importante vantaggio competitivo, oltre che aumentare la qualità delle relazioni e della comunicazione tra tutti i soggetti coinvolti.

Esistono numerosi strumenti che possono essere usati per l'analisi della reputazione online. Una prima classificazione distingue la ORM (*Online Reputation Management*) rispetto alle citazioni - o menzioni - nei risultati prodotti dai motori di ricerca da quelle pubblicate sugli spazi Web 2.0. Per le prime si possono applicare strumenti di *alert*, quali ad esempio Google Alert o Yahoo Alerts. Sono strumenti che consentono di ricevere una notifica - ad esempio via email con una frequenza definita dall'utente - della pubblicazione sul Web di contenuti relativi alla propria azienda od organizzazione. Si basano su interrogazioni automatiche legate ad un insieme predefinito (e personalizzato) di parole chiave. Il vantaggio di questo tipo di strumenti è che nella versione più semplice sono gratuiti, ma le loro prestazioni sono limitate (con valori bassi sia per la precisione che per il *recall*). Sistemi più sofisticati sono in grado di monitorare le piattaforme di social networking più diffuse, contenuti pubblicati in più lingue, applicando tecnologie di *Information Retrieval* e di analisi linguistica di diversa complessità. Tuttavia, la maggior parte di questi sistemi presenta limiti dovuti prevalentemente ad un approccio sintattico o solo limitatamente semantico alla individuazione delle menzioni e all'analisi del *sentiment*.

In questo lavoro si propone l'applicazione di tecnologie semantiche per realizzare un'analisi degli UGC di tipo testuale che permetta di comprenderne il significato e valutare il giudizio espresso in relazione ad una organizzazione complessa. Per illustrare l'approccio si fa riferimento ad uno studio realizzato nell'ambito del turismo, un settore articolato, non riconducibile ad un singolo dominio [Franch, 2010], in cui l'offerta si caratterizza per il contributo di

numerosi operatori e la domanda, in continua evoluzione, è costituita da segmenti di mercato con esigenze ed aspettative diverse.

Il primo passo prevede la definizione di un modello (*topic model*) che permette di individuare i concetti sui quali si basa la reputazione di una specifica destinazione in base alle strategie promozionali della stessa. A ciascun concetto (*driver*) corrisponde un insieme di indicatori, da ricercare negli UGC rilevanti. L'analisi semantica permette di analizzare più accuratamente anche il sentiment espresso nelle menzioni, evitando i fraintendimenti dovuti all'uso di termini predefiniti in una determinata frase, per cui ad esempio, 'buono' corrisponde a un giudizio positivo, anche in una frase come 'buon quarto d'ora'.

Il resto del lavoro è strutturato come segue. La sezione 2 descrive i sistemi per l'analisi della reputazione online. Nella sezione 3 si illustra l'approccio proposto, evidenziando il ruolo delle tecnologie semantiche. La sezione 4 presenta il caso e infine, la sezione 5 è dedicata alle note conclusive e ai problemi aperti.

2. Sistemi per l'analisi della reputazione online

2.1 Il contesto e le sfide

Per le aziende, la rilevazione di opinioni, giudizi, impressioni, critiche espresse nel Web, può affiancare l'analisi dei dati acquisiti attraverso metodi di ricerca tradizionali per lo studio del grado di soddisfazione del cliente alla base delle attività di *customer care* e in generale dei processi decisionali e strategici. Rispetto agli strumenti tradizionali, che comporendono indagini di mercato, interviste telefoniche, questionari allegati a prodotti e servizi, gli UGB offrono la possibilità di conoscere tempestivamente il giudizio incondizionato di un largo pubblico. Tuttavia, spesso il sentiment espresso su fonti aperte risulta difficilmente reperibile o utilizzabile per le aziende: da una parte per la numerosità dei contenuti che richiede un tempo eccessivo per ottenere una visione d'insieme, dall'altra per la difficoltà di identificare e filtrare i contenuti rilevanti per un'effettiva *marketing intelligence*. Da una conoscenza tempestiva e completa del sentiment dei propri clienti scaturisce la capacità di prevedere nuovi bisogni e tendenze e al contempo di apprendere eventuali segnalazioni o problemi che rispettando il normale iter richiederebbero tempistiche molto più lunghe prima di poter essere presi in considerazione. Tutto ciò è tanto più vero quando ci si focalizza sul mondo dei servizi: è qui che le esigenze del cliente diventano prioritarie e le modalità con cui vengono espresse sono assai diverse e personalizzate. Una mera raccolta di istanze, di occorrenze degli UGC (ottenuta ad esempio utilizzando uno strumento di Alert) offre un punto di vista esclusivamente quantitativo che non delinea la reazione del pubblico nei confronti dell'offerta. Ciò che risulta prioritario, non solo in termini di *time effort*, ma soprattutto in termini di efficacia, è un'analisi qualitativa che offra una visione di insieme sul sentiment relativamente all'offerta e al tempo stesso la possibilità di dettagliare tale sentiment, così da poter avere sotto controllo eventuali situazioni critiche.

L'ambito del marketing territoriale e turistico risulta essere uno dei settori più sensibili per sua natura alle esigenze dei clienti. Tuttavia, non è affatto semplice ricevere un feedback strutturato o strutturabile da parte dei turisti, anche perché ogni giudizio per definizione esprime una visione od emozione soggettiva e in quanto tale esprimibile in diversi modi linguistici. Per quanto riguarda il linguaggio, le problematiche per l'analisi automatica degli UGC possono essere classificate in primo luogo rispetto ad ortografia, sintassi e semantica:

- ortografia: esistono svariate forme 'colorite' e regionali; abbreviazioni, che contraddistinguono il linguaggio degli SMS così come quello di forum e dei blog;
- sintassi: un giudizio può essere espresso in una semplice frase, ad es., "mi piace", oppure può essere il risultato di un ragionamento complesso articolato su più frasi;
- semantica: lo studio del significato delle parole mostra la complessità del linguaggio, continua fonte di potenziale ambiguità; due parole scritte nello stesso modo possono riferirsi a due concetti diversi (come calcio inteso come sport o come elemento chimico); due parole scritte in modo diverso possono fare riferimento allo stesso oggetto (come ad esempio, casa e abitazione).

Un sistema semantico, in grado di comprendere il significato dei testi, permette di:

- focalizzarsi sui contenuti dei testi, e quindi su argomenti, concetti, entità salienti, indipendentemente dal modo in cui sono stati espressi e a prescindere dal formato, archivi, ecc.
- garantire un livello più alto di precisione nell'interpretazione del significato delle parole: le tecnologie più comuni per l'elaborazione di un testo scritto nel linguaggio comune raggiungono un grado di precisione che si aggira fra il 45% e il 55%. Alcuni sistemi per applicazioni specifiche o su domini ristretti- arrivano a percentuali più alte, con punte intorno al 70%. Con l'analisi semantica, la precisione può superare il 90% indipendentemente dal contesto e dalla quantità di documenti.

2.2 Strumenti

L'esigenza di sviluppare sistemi per supportare l'analisi della reputazione online è stata sottolineata da molti autori. Una descrizione dettagliata ed approfondita delle tecnologie applicabili e degli approcci esistenti si trova in [Pang e Lee, 2008]. Un lavoro antecedente che ha posto le basi per affrontare il tema della reputazione online è [Resnick e al., 2000], in cui gli autori fanno riferimento ai sistemi applicati da alcuni siti di successo. I lavori pubblicati più recentemente affrontano i vari aspetti dello sviluppo e realizzazione di un sistema di *online reputation monitoring* e propongono approcci sempre più raffinati sia per l'individuazione delle menzioni rilevanti che per l'analisi del sentiment. Un quadro dello stato dell'arte si può rilevare dalle pubblicazioni di due recenti workshop, dedicati alle tecnologie linguistiche per ORM e all'analisi del sentiment [Antin e al., 2011], [Balahur e al., 2011]. Tuttavia, nessuno dei sistemi proposti adotta un sistema di *natural language processing* semantico

basato su una base di conoscenza sufficientemente allargata da poter essere applicato a domini diversi, sia pur con adattamenti, per tener conto del linguaggio e delle conoscenze specifiche. Molti si limitano al riconoscimento delle cosiddette *named entity* (nomi geografici, indirizzi, ecc.), o recuperano la semantica dei termini nei commenti collegandoli ad un certo numero di concetti; complessivamente si tratta di approcci che potremo definire *semantically-flavoured*. Fra i sistemi commerciali, un'analisi semantica è effettuata da BlogMeter (<http://www.blogmeter.it>); da citare inoltre il progetto realizzato per il Comune di Milano [Barbesino e al., 2010] che si basa su una piattaforma integrata per l'analisi semantica di più fonti. Per quanto riguarda l'analisi della reputazione delle destinazioni turistiche, uno studio recente ha analizzato l'impatto degli UGC restituiti dai motori di ricerca [Inversini e Cantoni, 2011], evidenziando la varietà delle fonti rilevanti per una destinazione.

3. Il nostro approccio

3.1 Il processo di analisi della reputazione online

Il processo di analisi della reputazione online di una destinazione turistica prevede tre fasi principali:

- **individuazione delle fonti web** sulle quali individuare gli UGC collegati alla destinazione tenendo conto dei profili del target dell'offerta turistica e di altri criteri volti ad ottimizzare l'efficacia delle strategie di comunicazione e promozione;
- **definizione e validazione dei concetti** (*driver*) da localizzare all'interno delle opinioni degli utenti Web raccolte;
- **analisi linguistica delle opinioni** per riconoscere e classificare i contenuti rilevanti rispetto agli elementi caratterizzanti l'offerta turistica culturale e valutazione del sentiment.

Per le destinazioni turistiche, i siti di riferimento appartengono a due gruppi (linee guida per l'identificazione delle fonti da monitorare per un hotel sono contenute in [Tourism BC, 2011]):

- social network e altri siti 'generalisti' più usati nelle aree geografiche da cui provengono i turisti (flussi *incoming*), ad esempio Facebook e Twitter perché hanno un largo pubblico (per alcuni mercati sono diversi; ad esempio, per la Cina, alcuni siti sono proibiti ed esistono altri social network più diffusi di quelli citati);
- siti collegati al settore del turismo (denominati anche Travel 2.0), perché frequentati dai turisti per programmare, realizzare e condividere le esperienze di vacanza. Fra questi, Tripadvisor è uno dei più noti, ma ne esistono moltissimi, fra cui si possono citare anche Zoover, Wayn, Triplt.

Il **secondo passo** è il più critico, perché si tratta di focalizzare l'attenzione sulle strategie promozionali della destinazione per stabilire quali sono gli UGC di interesse. Se, ad esempio, si volesse verificare la reputazione online della destinazione rispetto alle attività sportive praticabili per promuovere un'offerta turistica caratterizzata da una vacanza attiva, tipicamente per attirare un segmento di turisti più giovani (un problema tipico delle destinazioni alpine per

le vacanze estive), occorrerà identificare contenuti collegati a tale tema. Si tratta in sostanza di definire un 'topic model'.

A titolo di esempio, la tabella 1 illustra uno dei 'topic' individuati per il modello sviluppato per il caso relativo al Trentino come destinazione turistica culturale. Dalla definizione dei driver di riferimento dipendono le analisi che possono essere effettuate sui dati.

Tabella 1 - Musei ed istituzioni culturali

Brand, Dominio	Line: Musei e Istituzioni	Product, Event	Feature part: Microeventi	Qualità
Cultura-turismo in Trentino	MART (Competitor: Museo Madre Napoli, GAM Torino, Museo di Arte Contemporanea Basilea, Museion Bolzano, Guggenheim Bilbao, MOMA NY)	Le varie mostre	Micro eventi: conferenze, Testimonial, Vernissage	Costi Accessibilità: Parcheggio, navette, collegamenti vari, diversamente abili, location, edificio, estetica, allestimento, contenuti, flessibilità, orari, servizi accessori: book, cataloghi, ristorante, varietà dell'offerta: selezione, divertimento, intrattenimento
	Castello Buonconsiglio (Eventuali Competitor)			
	Elenco principali musei trentini con eventuali commentatori: Museo guerra Rovereto, Casa d'Arte futurista, ecc.			

Come si può vedere dalle tabella, a ciascuno dei driver individuati per descrivere l'offerta culturale del Trentino è stata associata una tassonomia che permette di rappresentare in modo calibrato gli elementi di interesse: sufficientemente dettagliati, ma non troppo, per evitare, rispettivamente classificazioni generiche o forzate o di frammentare l'analisi a cause di un'alta 'sparsità' dei dati.

Il **terzo passo**, analisi e classificazione degli UGC rispetto a contenuto e sentiment, si basa sull'applicazione di tecnologie semantiche in grado di comprendere quali menzioni sono rilevanti per un determinato driver e quale, se presente, è il giudizio dell'autore rispetto ad esso. Le tabelle (tassonomie) definite per una determinata destinazione e strategia della stessa costituiscono la base per la classificazione delle menzioni, per evidenziare, ad esempio, che uno post su un blog riguarda una specifica esposizione (*Product, Event*) di uno dei musei (*Line: Musei e Istituzioni*) nella destinazione, l'acquisto del catalogo (*Qualità*) in occasione del vernissage (*Feature part: Microeventi*). In generale, un driver, è descritto da una quintupla ordinata costituita da:

<Brand/dominio, Line, Product/Service, Feature/Part, Quality>

Ad ognuna delle informazioni riconducibili a un ramo della tassonomia, viene assegnato un giudizio in senso assoluto normalizzandolo secondo uno schema predefinito. Nel caso descritto nella sezione successiva si è applicato lo schema seguente: 1 = very poor; 2 = poor; 3 = fair; 4 = good; 5 = excellent.

3.2 Il sistema per l'analisi semantica dei testi

Rispetto al processo di analisi della reputazione online introdotto nella sezione 3.1, la piattaforma adottata per il progetto, Cogito® (<http://www.expertsystem.net/>) supporta le seguenti attività:

- **reperimento delle informazioni:** agenti software di acquisizione dati, consentono di prelevare e raccogliere tutti i contenuti da analizzare, indipendentemente dalla provenienza e dal formato (*crawling*); in particolare, tali agenti producono post in formato normalizzato aventi la seguente struttura:
 - nome della fonte; localizzatore universale della fonte (URL o equivalente);
 - titolo della discussione (thread); localizzatore universale della discussione (URL o equivalente);
 - localizzazione contestuale della discussione (rispetto alla fonte);
 - titolo del post; localizzazione universale del post (URL o equivalente);
 - localizzazione contestuale del post (rispetto alla discussione);
 - data e ora di pubblicazione; autore; testo; firma dell'autore.
- **individuazione dei driver,** dove per driver si intende una categoria o un insieme di categorie a cui può essere associato un giudizio;
- **analisi linguistica:** sui dati prelevati si svolge l'analisi linguistica completa - morfosintattica, grammaticale, logica, semantica - per estrarre le informazioni necessarie per le successive analisi (un esempio è riportato in Fig.1).
- **interrogazione e reporting:** l'utente finale può visualizzare ed analizzare tutte le informazioni raccolte, attraverso interfacce di interrogazione e monitoraggio. La possibilità di strutturare i dati e renderli facilmente navigabili, permette analisi di vari tipi, sia da parte del top management, che dell'operatore di sportello, senza richiedere competenze tecniche o applicative specifiche.

Per quanto riguarda il sentiment, per interpretare le informazioni individuate e scaricate automaticamente, Cogito applica delle regole che consentono l'identificazione dei giudizi sui driver individuati con la tassonomia. Un esempio è riporta in Fig.2.

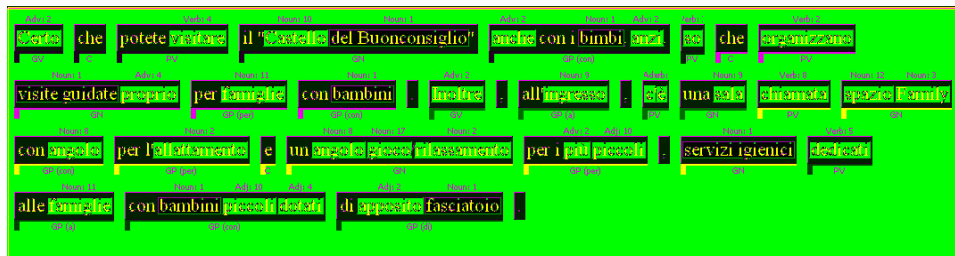


Fig.1 – Esempio analisi linguistica di un post

Per un singolo post, viene effettuata l'**estrazione** e la **normalizzazione** dei dati elementari utili, cioè di tutte le espressioni che possono indicare driver e delle espressioni che possono indicare rating; successivamente, viene

effettuata la **composizione** dei giudizi, ossia l'**attribuzione** dei rating ai driver a partire dai dati elementari estratti nella fase precedente.

```
SCOPE SENTENCE IN SEGMENT(AUTHORIAL_TEXT, Comparison)
{
  // non male, non malaccio
  IDENTIFY(FAIR_ADJ)
  {
    @FAIR_ADJ[KEYWORD[("non male", "non malaccio", "non è male", "non è malaccio", "non
    è male", "non è malaccio", "nella media", "nella norma", "niente di speciale", "niente di che",
    "senza infamia e senza lode", "bellino", "bellina")]]
  }
}
```

Fig.2 – Esempio di regola per la valutazione del sentiment dei post

La Fig.3 riporta l'analisi complessiva per il seguente post:

"Il MART è uno dei pochi edifici di Botta che mi piaccia 😊"

```
BRAND -DOMINIO = MART;
LINE = MUSEI ED ISTITUZIONI;
PRODUCT - EVENTI = (GENERAL);
feature/part - MICROEVENTI = (GENERAL);
quality = ESTETICA EDIFICIO
rating: 4
```

Fig.3 – Analisi dei contenuti e del sentiment di un post

4.11 caso di studio

L'approccio proposto nella sezione precedente è stato applicato in un progetto per il Trentino. l'obiettivo del progetto è l'analisi in tempo reale e in modo continuo del sentiment che i clienti esprimono liberamente nel web a proposito della qualità dell'offerta turistica del Trentino. Le assunzioni principali hanno riguardato due aspetti fondamentali:

- la scelta di ambiti applicativi significativi sul piano strategico e focalizzati su domini trattabili dal punto di vista computazionale: cultura e sport;
- l'adozione di un sistema di analisi del linguaggio che rappresenta lo stato dell'arte per l'area di *Natural Language Processing*, la piattaforma Cogito.

Nella prima fase del progetto si è focalizzata l'attenzione sulle due aree individuate, approfondendo l'analisi per cinque macro-categorie per ognuna di esse: 5 discipline sportive e 5 eventi, creando una sesta categoria sia per le altre discipline sportive che per le attrazioni culturali non incluse nelle voci esplicitate. I post relativi alle prime 5 categorie sono stati analizzati in dettaglio secondo i topic model sviluppati per ciascuno di essi. A tal fine si è personalizzata la rete semantica (SENSIGRAFO®) al fine di adeguarla al linguaggio da interpretare. Un'analisi preliminare ha evidenziato la presenza di

un numero di post relativi alle aree individuate per l'offerta turistica trentina e di alcuni competitor nell'ordine di qualche decina di migliaia di post, indice della fattibilità ed utilità di un sistema automatico di analisi della reputazione online.

Si sono identificate 5 fonti, fra cui Googlegroups e Googleblogs, che da sole sono risultate contenere il 50% dei post sull'offerta culturale del Trentino (sul totale dei giudizi espressi nel mondo web, estrapolati tramite aggregatori che identificano solo i post che esprimono sentiment sul dominio di riferimento).

Lo studio ha permesso di misurare e valutare il positivo ed utile apporto dell'analisi dei cosiddetti UCG per il turismo trentino. Sulla base delle fonti analizzate, sono emerse per la destinazione alcune indicazioni e tendenze precise. Fra queste, una forte domanda di 'natura', 'tranquillità' e 'tipicità', che talvolta stridono con l'organizzazione di grandi eventi. L'analisi dei post individuati potrebbe contribuire a chiarire come l'offerta turistica trentina possa far convivere queste due caratteristiche della domanda turistica. Altre informazioni sono utili sono state ottenute riguardo alla percezione dell'offerta turistica rispetto a quella dei concorrenti.

5. Conclusioni

Punti di forza dell'approccio proposto sono la completa trasparenza del funzionamento del sistema di analisi della reputazione online nei confronti dell'utente finale e l'analisi semantica di testi. Quest'ultima consente di ricavare dagli UGC informazioni utili: per comprendere le esigenze dei turisti, come sono percepiti i *competitors* e verificare se coincidono con quelli attesi; individuare gli aspetti critici su cui intervenire e quelli positivi da valorizzare; identificare 'opinion leader' da contattare e coinvolgere; raccogliere feedback su specifici pacchetti o prodotti dell'offerta turistica. L'individuazione delle *named entity*, e quindi di località, eventi, hotel, ecc. consente di rispondere a domande su quali siano le località o gli eventi citati più frequentemente e se sono quelli attesi; quali siano i siti più suggeriti; se vengono date indicazioni su operatori turistici, ecc.

Sul piano metodologico, queste informazioni confermano l'utilità da parte delle *Destination Management Organization* (DMO) di adottare un efficace sistema di analisi della reputazione online, stabilendo procedure e strumenti per agire a due livelli: operativo, per attività di CRM, rispondendo adeguatamente ai singoli post; strategico, per la pianificazione della promozione e dello sviluppo della destinazione, identificando eventuali trend o minacce per il successo della stessa. L'applicazione di tecnologie semantiche permette a una DMO di svolgere anche un ruolo di coordinamento, importante per il web marketing, laddove venisse utilizzata per segnalare agli operatori sul territorio le opinioni e i commenti che li riguardano (ad esempio, ai singoli alberghi, o alle aziende di promozione turistica d'ambito).

Sviluppi futuri riguardano una valutazione complessiva dei risultati in termini di recall e precision sui tre livelli di applicazione delle tecnologie semantiche: identificazione dei post pertinenti, identificazione dei contenuti nei post e attribuzione del sentiment, laddove i risultati dello studio condotto, benché

preliminari confermano le prestazioni che la tecnologia semantica adottata ha ottenuto negli ambiti applicativi finora affrontati.

Ringraziamenti: si ringraziano Nicola Zeni per la collaborazione alla definizione dei topic model e la Trentino Spa.

Bibliografia

Antin J., Churchill E.F., Chen Bee-Chung, Workshop on online reputation: context, privacy, and reputation management. In Proc. 20th Int. Conf. companion on World wide web - WWW'11, ACM, New York, 2011, 321-322.

Balahur A., Boldrini E., Andres Montoyo A., Patricio Martinez-Barco P. (eds), Proc. 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis - WASSA2011, Association for Computational Linguistics, Portland, Oregon, 2011

Barbesino P., Francalanci C., Petrovich F., Social Media Intelligence Comprendere il "Polso" di un Territorio, Mondo Digitale, AICA, Milano, 45, 2010, 52-63.

Franch M. (ed), Marketing delle destinazioni turistiche. Metodi, approcci e strumenti, McGraw-Hill, Milano, 2010, 277-324.

Gibson M., Repairing Your Damaged Online Reputation: When Is It Time to Call the Experts?, Time Business, April 19, 2011, www.time.com/time/business/article/0,8599,2065597,00.html#ixzz1a7V5Zlbc

Inversini A., Cantoni L., Towards online content classification in understanding tourism destinations' information competition and reputation, Int. J. Internet Marketing and Advertising, 6(3): 282-299, 2011.

Juran J., Gryna Jr F.M., Quality planning and analysis, McGraw-Hill, 1980.

Juran J.M., A History of Managing for quality, ASQC Quality Press, Milwaukee, WI, 1995.

Mich L., La web presence delle destinazioni turistiche, in Garibaldi R., Peretta R. (eds), Facebook in Tourism, eBook, FrancoAngeli, Milano, 2011.

O'Reilly, What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software, 2005, <http://oreilly.com/web2/archive/what-is-web-20.html>

Pang B., Lee L., Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1-135, 2008.

Pattison K., Managing an Online Reputation, The New York Times, July 29, 2009, <http://www.nytimes.com/2009/07/30/business/smallbusiness/30reputation.html?adxnnl=1&pagewanted=all&adxnnlx=1318010523-IMI2YTLQVnuQFQPY4Z3BvA>

Resnick P., Zeckhauser R., Friedman E., Kuwabara K., Reputation Systems, Communications of the ACM, 43(12): 45-48 - 2000.

Tourism British Columbia, Online reputation management, 2011, www.jti.gov.bc.ca/industryprograms/pdfs/OnlineReputationManagementTBEGuide2011_Jun15.pdf

The hitchhiker’s guide to the Network Neutrality Bot test methodology

Simone Basso*+ Antonio Servetti+ Juan Carlos De Martin*+

`simone.basso@polito.it, antonio.servetti@polito.it, demartin@polito.it`

* NEXA Center for Internet & Society – DAUIN – Politecnico di Torino, Italy

+ Internet Media Group – DAUIN – Politecnico di Torino, Italy

Abstract

The Neubot project is based on an open-source computer program, the Neubot, that, downloaded and installed by Internet users, performs quality of service measurements and collects data at a central server. The raw results are published on the web under the terms and conditions of the Creative Commons Zero license. This paper is the guide for researchers and individuals that aims to study, build on and analyze Neubot methodology and results. We provide an exhaustive documentation of Neubot’s HTTP test behavior, along with a discussion of the methodology. Besides that, the article shows an analysis of the Turin-area results (in the May-September time interval) and explains the rationale behind the privacy policy, which allows us to publish results as raw data.

1 Introduction

The debate on “network neutrality” is becoming a more and more relevant topic in economic, technical and political environments [1] [2] [3]. The basic question is whether network operators should be allowed to differentiate the Internet traffic that goes through their infrastructure or whether network neutrality should be explicitly safeguarded by the law, thereby enshrining what has been a characteristic of the Internet since its birth [4].

The ability to block or slow down the traffic, which is used to prevent the spreading of spam, viruses, botnets and other malwares, may also be employed by Internet Service Providers to implement other traffic policies that are very questionable [5]. Apart from the so-called “Great Firewall of China” [6] [7] and other striking blocking efforts [8] (which are beyond the scope of this paper), differentiating technologies are also employed to throttle the traffic of peer-to-peer applications and Content Providers.

The traffic of peer-to-peer applications (e.g. BitTorrent and Emule) may be slowed down to protect underprovisioned shared access networks from congestion, during peak hours or whenever the network load rises above certain threshold. More selectively, differentiation may be employed to throttle “seeding-only” peer-to-peer traffic, especially when the ISP has a transit agreement with the upstream provider and it is not willing to pay money for traffic that does not directly benefit the user. Finally, the differentiation of peer-to-peer applications (e.g. Skype) and Content Providers’ traffic (e.g. YouTube and Netflix) allows to allocate and guarantee bandwidth to the so-called “Managed or Specialized Services”, i.e. services like voice over IP and video on demand, bundled with the Internet-access offer and available to the ISP subscribers only [9] [10].

Technically, network differentiation is a two-step process. At the edge of the network, there is the classification function, which assigns each packet to a traffic class, using different techniques [11] and policies, possibly also accounting for the user’s past behavior. Then, inside the network, packets receive the service level associated with their class. In particular, packets that belong to low priority classes may be diverted on slower and/or more congested links, resulting on a worst

average quality of service. Alternatively, they may be scheduled for forwarding after higher priority packets [12] [13] and dropped with higher probability from router queues in case of congestion [14].

As a consequence, the topic of network neutrality is tightly related to the one of network quality measurement. A nonneutral network path is one where certain traffic classes experience, on the average, more packet losses and/or delays than others. For this reason most of the network neutrality tools available in literature perform network quality measurement, via active probing and/or capturing and analyzing the user's traffic. Then, they compare the measurement results for different protocols and/or services, seeking for significant quality differences [15] [16] [17] [18].

Despite the potential impact of traffic differentiation and the fact that there are some available tools for network quality measurement, most users, developers and network administrators are still not provided with enough information. At the same time, ISPs and Content Providers have access to a wealth of data, respectively because they convey users' traffic and they serve a large number of eyeballs [19]. To start closing this information gap and empower the users, our proposal is the "Network Neutrality Bot" (Neubot) project.

The Neubot project is based on an open-source computer program, the Neubot, that, downloaded and installed by Internet users, performs quality of service measurements and collects data at a *central server* [20]. Neubot runs in the background on the user computer and periodically contacts the *central server*, which redirects it to the closest *test server*. In turn, Neubot connects to the *test server*, performs a *transmission test*, saves the results locally and uploads them to the test server as well [21]. Neubot is designed as a software architecture for network measurements and currently hosts two *transmission tests*. The *speedtest* test uses the HTTP protocol and measures round-trip time, download and upload goodput [22]. It is inspired to Speedtest.net [23] test, hence the name. The *bittorrent* test emulates the BitTorrent peer-wire protocol and measures round-trip time, download and upload goodput. The raw results are published at <http://www.neubot.org/data> under the terms and conditions of the Creative Commons Zero license.

This paper is the guide for researchers and individuals that aims to study, build on and analyze Neubot methodology and results. The major contribution of this work is the exhaustive documentation of the *speedtest* test behavior, along with a discussion of the methodology. Besides that, the article shows a qualitative analysis of the Turin-area results, in the May-September time interval. And explains the rationale behind the privacy policy, which allows to publish results as raw data.

The rest of this paper is organized as follows. In section 2 we describe related research efforts. Section 3 documents and discusses in detail the *speedtest* transmission test. Section 4 is dedicated to the presentation of the measurements in the Turin area. In section 5 we explain the rationale behind the privacy policy. Finally, in section 6 we draw the conclusions.

2 Related work

Early network-neutrality literature and tools focus on detecting port blocking and RST injection. The seminal work of Beverly, Bauer and Berger aims to build a map of port blocking policies, with emphasis on VPN, email and file sharing ports [24]. Electronic Frontier Foundation's Switzerland is a semi peer-to-peer prototypal application designed to detect the modification or injection of packets with the help of a central server [25]. The article of Dischinger et al. describes BTTest, an easy-to-use tool to detect BitTorrent RST injection practices, which has been used to characterize the RST policy of many providers [26]. NNSquad Network Measurement Agent (NNMA) is a prototypal piece of software that detects certain spoofed RST segments [27]. Weaver, Sommer and Paxson's paper provides a rich set of heuristics to detect RST injections and identify the generating device with a certain degree of confidence [28].

More recent works measure various facets of quality and compare those measurements to assess the degree of neutrality. NANO (Network Access Neutrality Observatory) is an agent that passively collects performance data from users' computers and uploads them to a central server. In turn, the server relies on stratification to group clients in clusters where the performance difference

depends on the ISP policy and not on confounding factors [29] [18]. NetPolice (formerly NVLens) is a protocol-aware traceroute-like tool to (i) probe backbone ISP paths with traffic that emulates different applications protocols and (ii) compare the loss rate [30] [17]. DiffProbe is an active probing method to detect whether an ISP is performing delay or loss discrimination comparing the delays and losses experienced by two flows [16]. Glasnost – the evolution of BTTest – compares the throughput of two traffic flows, identical in all aspects other than their packet payloads. It runs the two flows multiple times back-to-back, filters the results to eliminate noise and reports (non)discrimination [15]. WindRider is a mobile application that performs active goodput tests using different ports. It also performs passive tests such as measuring the delay to load web pages. In both cases the results are uploaded to central servers [31].

Other research and nonresearch efforts just focus on quality. BISMark is a OpenWRT-based home router that performs periodic active measurement of the access link performances [32]. Grenouille periodically performs ping, download and upload tests with nearby servers and collects results at a central server [33]. Pathload2 is a tool that measures the available bandwidth of the user’s broadband connection [34].

Completely different from the above projects and tools is Measurement-Lab (M-Lab). Which is a distributed server platform – founded by the New America Foundation’s Open Technology Institute, the PlanetLab Consortium, Google Inc. and academic researchers – where researchers can deploy the server of their active transmission test tools [35].

3 The *speedtest* transmission test

The *speedtest* test takes place between the Neubot and the test server indicated by the master server. Before the actual test, there is the negotiation of the test parameters and the test itself may be delayed to control the load. Then there is the actual test, discussed below. After that, the Neubot sends the test server its results and vice versa, so there is complete information. For a more detailed explanation of the test phases we refer the reader to our previous work [21].

The actual *speedtest* test uses the HTTP protocol and measures round-trip time, download and upload goodput. The round-trip time is estimated using two distinct techniques: the measurement of the time required to connect and the measurement of the request-response latency. The goodput is the application-level throughput [36], calculated dividing the amount of sent (or received) bytes over the elapsed time.

In the following subsections we provide more details regarding the measurements. Then we discuss the limits of our methodology.

3.1 Time required to connect

The time required to connect is the time it takes for the `connect(2)` system call to complete. This approximates the round-trip time because this system call immediately sends a `SYN` segment and returns when it receives the `SYN|ACK` segment.

Neubot uses nonblocking I/O and `select(2)` to dispatch I/O readiness events. Therefore the time required to connect is actually computed performing the difference between the time when the socket becomes writable and the time when `connect(2)` reports that the connection attempt is either complete or `EINPROGRESS`.

Since the test uses two (or four) connections, the result is the average. The current implementation (0.4.1) serializes multiple concurrent connection attempts, so each `connect()` is measured independently of the others. This was not the case before Neubot 0.3.7: in old versions multiple attempts were overlapped. So the average is more noisy in old Neubot versions.

3.2 Request-response latency

The request-response latency is computed performing the difference between the time when the response is received and the time when the request is sent. This is a reasonable overestimation of

the round-trip time because Neubot sends small `HEAD` requests and the server responds with very small bodyless responses. Indeed, per RFC 2616 the response to an `HEAD` request must not contain the body and must consist of the headers only [37].

While the test uses two (or four) connections, only one connection is employed to perform this measurement. The request-response latency is measured back-to-back for ten times. And the result is the average.

3.3 Goodput

The goodput is calculated dividing the amount of bytes sent (or received) over the elapsed time. The download estimation is performed at the receiver, while the upload estimation is performed at the sender. The amount of bytes sent (or received) is initially small and may be adapted after the test, if needed. The elapsed time is the difference between the time when the last byte of the response is received and the time when the first byte of the request is sent. The test is repeated if the elapsed time is less than the target test duration. Before repeating the test, the amount of bytes to send (or receive) is adapted, so that subsequent tests would run for more than the target test duration (under current conditions).

Since the first public release (Neubot 0.3.0) the test has employed two concurrent connections. This was motivated by the fact that we started with just one server, located in Turin, and we wanted to double the range of bandwidth-delay products we could serve. There was also a period (Neubot 0.3.5 – 0.3.6) when Neubot employed four connections, but we decided to step back because the goal was to measure the quality not to maximize the goodput (more on this later).

The target test duration changed significantly since the first public release. One key point is to perform *quick* measurements, because Neubot is a background tool and should not consume too much network resources. For this reason, version 0.3.0 estimated the goodput out of one-second transfers. Unfortunately that was not enough and caused too much variation, even if the mean was consistent with other independent measurements [22]. To counter this problem, Neubot 0.3.5 and Neubot 0.3.6 raised the number of concurrent connections, from two to four. But the variation was still high, so Neubot 0.3.7 returned to the two-connections model but increased the target to five seconds.

The algorithm to adapt the number of bytes to send (or receive) after a test changed since the first public release too. At the beginning the strategy was to double the amount of bytes when the duration was less than the target. That was acceptable for short transfers but is too aggressive for five-seconds transfers. So, starting from Neubot 0.3.7 the amount of bytes is still doubled, but when the duration is less than one second only. Otherwise, it is scaled so that the next test would run for at least five seconds (under current conditions).

Starting from Neubot 0.3.7, the receive buffer is set to the fixed value of 256 KiB, because we have observed that different operating systems scale differently the receive buffer during slow start and we wanted to eliminate a possible confounding factor. We have chosen 256 KiB because it is the maximum-possible socket buffer value on BSD (to increase it more one needs root privileges to raise the limit via `sysctl`).

3.4 Discussion

The “time required to connect” approximation of the RTT is acceptable if neither the `SYN` nor the `SYN|ACK` segments are lost. And if the two operating systems are not heavily loaded, so that the network delay is at least one order of magnitude more than the segment processing delay. The “request-response latency” is less precise because there is always some application delay involved and because there is application overhead caused by HTTP headers. But it allows to collect more than one sample per connection. Of course, both estimates are performed at the beginning of the connection so they may not reflect the actual RTT during the data transfer.

The upload goodput estimate is measured at the sender. This is done for implementation simplicity. But it would be more robust to measure it at the receiver. Because measuring at the

receiver smooths possible peaks, such as the ones caused by drop tail in the home router. And because, more generally, the measurement at the receiver conveys more information.

The methodology always yields an underestimate of the actual goodput. Because the measurement starts when the requester sends the request. Therefore the elapsed time includes at least one idle round-trip time: the one when the connection is waiting for the response to arrive. The impact of that was higher with one-second measurements than it is with five second. Assuming a round-trip of 100 ms the idle time was 10% (or more), while now it is 2% (or more).

The methodology allows to compare connections with similar round-trip time ranges. But the comparison is more problematic when the round-trip times significantly differ. The problem is that the target duration is expressed in seconds, while TCP behavior is typically modeled in terms of “rounds”, where the round duration is one round-trip time [38]. So, connections with very different round-trips run the test for the same number of seconds but certainly not for the same number of rounds.

One-second tests experienced more variation because the percentage of rounds spent in the slow start was higher. In this state, fast recovery is fragile and TCP performances are more loss-sensitive. So, while the goodput measured by a short test may not be a good throughput approximation, the variation is still interesting, because it conveys information on the average loss ratio. In turn, this information can be used to compare the quality of different protocols or different access networks.

The concurrent-connections model has the advantage that the measured goodput is closer to the throughput. The available bandwidth is used more efficiently, because each connection has a smaller window, so it takes less round-trip times to recover after a loss. And because a connection may grow and take more bandwidth when another experiences a loss. As a consequence, measuring a congested network with more concurrent connections yields a better goodput. Which is not desirable, because worst access links should be penalized more, not less.

The advantage of fixed-size receive buffers is that all operating systems should have the same behavior. Moreover, knowing in advance the buffer size simplifies results analysis, because it is easier to classify buffer-limited TCP connections. On the other hand, the problem with this strategy is that most high-speed high-delay connections are likely buffer limited. Especially in countries like Korea and Japan where it is common to have 100 Mbit/s access links. This is not desirable because better access links should be premiated, not penalized.

Finally, the methodology does not account for many confounding factors that may reduce the goodput. It assumes that the user is not heavily using her computer and/or her access link. That no other users are heavily using/sharing the same access link. That the local wireless connection (if any) does not experience many losses. That there is no congestion in the backbone. That the test server is not overloaded. And possibly there are other hidden assumptions and unaccounted confounding factors.

4 Measurements in the Turin area

This section presents an analysis of a subset of the data collected in the period from 30rd May to 13 September 2011.

Table 1 shows the number of tests, the number of Neubots and the date of the first and last test for the whole dataset and for the subset of Neubots geolocated in the Turin area. We use MaxMind <http://www.maxmind.com/> free tools and databases for geolocation and for the identification of the Provider. We focus on the Turin area because our test server is located in the Torino-Piemonte Internet eXchange (TopIX), so we show measurements performed by “nearby” (in terms of RTT) Neubots. Of course, the analysis performed on the Turin area can be repeated for other regions and countries to build a per-geographic-location per-provider results map.

Table 2 shows the number of measurements and the number of Neubots geolocated in the Turin area for each Autonomous System (which is a Provider’s network). In this paper we show the results of the four commercial providers with more tests and more users in the Turin area: Vodafone, Infostrada, Fastweb and Telecom Italia. GARR, the Italian universities network, is

	Whole dataset	Turin area
Number of tests	947729	80196
Number of Neubots	1054	166
First test	30-05-2011	30-05-2011
Last test	13-09-2011	13-09-2011

Table 1: This table shows the number of tests, the number of Neubots and the date of the first and last test for the whole dataset and for the subset of Neubots geolocated in the Turin area.

Autonomous System	Measurements	Neubots
AS2594 CSI Piemonte	77	1
AS44957 OPITEL AS number	45	2
AS8968 BT Italia S.p.A.	1234	2
AS35612 NGI Spa	11	3
AS35719 TEX97 S.p.a	573	3
AS24608 H3G S.p.A.	18	4
AS16232 TIM (Telecom Italia Mobile) Autonomous System	90	4
AS8612 Tiscali Italia S.P.A.	555	7
AS137 GARR Italian academic and research network	26200	25
AS30722 Vodafone N.V.	1162	8
AS1267 Infostrada S.p.A.	794	24
AS12874 Fastweb SpA	24735	44
AS3269 Telecom Italia S.p.a.	24702	65

Table 2: This table shows the number of measurements and the number of Neubots geolocated in the Turin area for each Autonomous System (which is a Provider’s network). The four bottom Autonomous Systems are the ones plotted in Figure 1.

not included, even if it has many tests and users, because its performances are not comparable with the ones of commercial providers, having much more access link capacity. Of course, with the public availability of the results, one can build its own qualitative analysis with more, less or different providers.

Results do not represent users’ broadband connection speed: instead, they represent the average quality experienced by users that perform the same test with different Internet Service Providers and different broadband connections. They are more valuable as a tool for differential analysis than as a tool for measuring the speed magnitude.

The test methodology depends on the version of Neubot and is explained in section 3.3. In addition, the methodology does not account for confounding factors such as poor wireless connectivity and the activity of other users that share the same home network, as discussed in Section 3.4.

Figure 1a shows the cumulative distribution of the request-response latency for the tests performed in the Turin area. The distribution is plotted in the range 0-400 ms and greater values are assigned to the last bin. We use request-response latency because time required to connect is buggy in old versions of Neubot, as explained in Section 3.1. About 20% of Fastweb tests have latency higher than 400 ms: this is caused by 20 Neubots that consistently report very high latencies¹.

Figure 1b shows the cumulative distribution of the download goodput for the tests performed in the Turin area. The distribution is plotted in the range 0-20 Mbit/s and greater values are assigned to the last bin. The goodput is calculated at the receiver, dividing the number of bytes transmitted over the time elapsed from sending the request until receiving the whole response. If we except Vodafone, which has a single 7-Mbit/s-download-512-Kbit/s-upload commercial offer, results are collected by users with different broadband speeds and contracts. It would be interesting to classify the results depending on the highest measured speed and plot the contribution of each class separately.

¹Incidentally, this seems not to be caused by the fiber vs. ADSL split of Fastweb customers. There are many Neubots with less than 1 Mbit/s upload both in the more-than-400-ms camp and in the other camp.

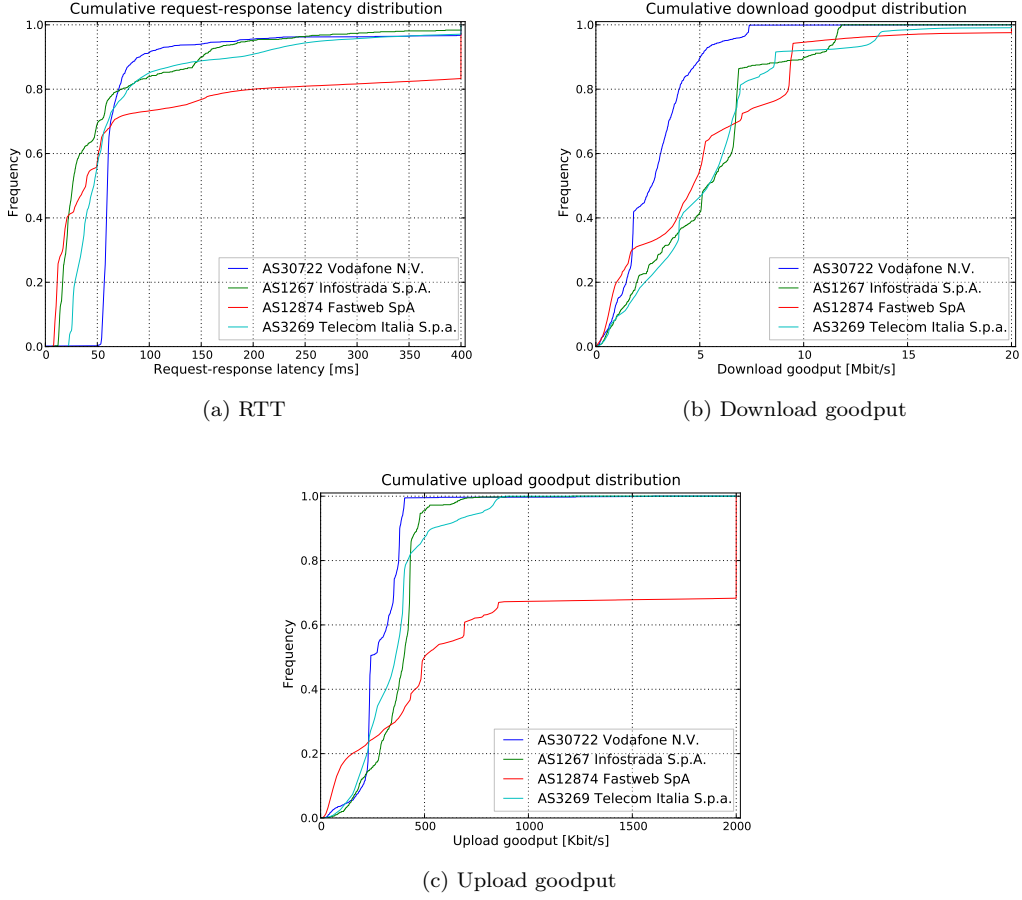


Figure 1: This figure shows the cumulative distribution of request-response latency, download and upload goodput for the tests performed in the Turin area. The distribution is plotted in a given range and greater values are assigned to the last bin. The test methodology changes slightly depending on the version of Neubot and does not account for confounding factors such as poor wireless connectivity and the activity of other users that share the same home network (see Section 3). In addition, one must consider that the maximum goodput that a user can reach depends, of course, on the maximum speed of the access link, as indicated in the contract.

Figure 1c shows the cumulative distribution of the upload goodput for the tests performed in the Turin area. The distribution is plotted in the range 0-2000 Kbit/s and greater values are assigned to the last bin. The goodput is calculated at the sender, dividing the number of bytes transmitted over the time elapsed from sending the request until receiving the whole response. Vodafone apart, results are collected by users with different broadband speeds. For example, 35% of Fastweb tests are above 2000 Kbit/s, clearly due to Fiber-To-The-Premises connections.

Scripts and aggregate data used to produce plots in Figure 1 are available on the Neubot website for other researchers and institution to carry out independent analysis. The whole dataset of raw Neubot results is available online as well. Whenever the user provided the permission to publish (see Section 5), the result retains the original Internet address. Otherwise, the Autonomous System name and estimated geographic location are provided.

5 Privacy policy

The result of a transmission test is a tuple that contains performance metrics, such as goodput and latency, information on the user's computer load, and the user's Internet address. Neubot needs to

collect test results on its servers to be able to study them and publish aggregate data. In addition, the project would like to publish the raw results, to allow other individuals and institutions to reuse them for research purposes.

Both the collection and the publication of the results, and, in particular, of Internet addresses, are instrumental to Neubot goals. The project needs to collect data to accomplish its goal of studying the quality and neutrality of users' Internet connections. In particular, without the Internet address it is not possible to associate the results to an Internet Service Provider and a geographic location. The publication of raw results is very relevant as well, albeit it is not so critical as the collection. The availability of public data allows independent researchers and institutions to study, question and analyze the project methodology and research results. This will benefit Neubot developers and users, as well as the Internet community at large – eventually leading to a better shared understanding of the 'Net. And, again, no serious study of the data can be carried out without knowing the original Internet address.

However, Internet addresses cannot be collected and published as easily as goodput and latency measurements. Indeed, according to many opinions, including the ones of Ohm [39] and of the Article 29 Working Party² [40] [41], Internet addresses are personal data and they must be treated under the provisions of Italian and European Privacy Laws (Decree 196/03: “Codice in materia di protezione dei dati personali”) [42]. This means that, in general, the user's *informed consent* is required for each purpose – collection and publication in this case – of the personal data treatment. There are exceptions to this regime, which allow public institutions, such as the Polytechnic University of Turin, to collect personal data without asking for the permission to do so [43]. However, Neubot does not take advantage of this possibility for two reasons. First, data collected in this way cannot be published freely on the web. Second, providing users a privacy policy and asking for their explicit permission is more transparent and fair.

According to the Law, Neubot installer shows users the privacy policy, informs on personal data treatment procedures and explains users how to exercise their rights. Then, asks users to: (i) assert that they have read the policy; (ii) provide the permission to collect their Internet address for research purposes; (iii) provide the permission to publish their Internet address for reusing them for research purposes. The permission to collect is mandatory, as explained above: if the user does not provide it, the installation aborts. The permission to publish is optional. If it is not provided, the user Internet address will not be published. And the project will publish instead the Internet Service Provider name and the (estimated) geographic location³

6 Conclusion and future work

In this work we present an exhaustive documentation of Neubot's *speedtest* test behavior. The test estimates round-trip time, download and upload goodput. Round-trip time estimation is performed (i) measuring the time it takes for the `connect(2)` system call to complete and (ii) measuring the request-response application-level delay for small requests and responses. The goodput is calculated dividing the amount of bytes sent (or received) over the elapsed time. The test uses two concurrent HTTP connections, but used four connections in the 0.3.5-0.3.6 version range. The download estimation is performed at the receiver, the upload estimation at the sender. Starting from version 0.3.7 the receive socket buffer is fixed at 256 KiB.

The major contribution of this paper is the discussion of the test's methodology. Which does not account for confounding factors such as poor wireless connectivity and the activity of other users that share the same home network. Considering all the caveats, results are still a valuable tool for differential analysis of providers' performances. But they are not an effective measure of the real broadband connection speed.

Future versions of Neubot should dedicate one test to estimate the quality of the network, using

²A body composed of the data protection authority of each EU Member State, the European Data Protection Supervisor and the European Commission.

³We use MaxMind <<http://www.maxmind.com/>> free services to map the Internet address to the Internet Service Provider name and to the estimated geographic location.

just one connection, and another test to approximate the real broadband connection speed, using as many connections as needed. Moreover, given that Neubot has been accepted by Measurement Lab distributed server platform, the fixed buffer limitation should be relaxed (possibly adding another confounding factor), to be able to test high-speed high-latency networks. Finally, the test should be modified to run for a target number of RTTs, not seconds, to allow for comparison of connections with very different RTTs.

The article presents the Neubot measurements in the Turin-area, in the May-September time interval. Measurements have been divided per provider and the four commercial providers with more tests and Neubots have been analyzed. The analysis consists of the cumulative distribution of request-response latency, download and upload goodput. In the latency distribution, it would be interesting to investigate why a significant fraction of Fastweb users systematically experience very high latencies. Vodafone apart, the download and upload goodput distributions convey contributions from users with different broadband connections and contracts. It would be interesting to classify the users depending on their broadband connection and then plot the performances of each class separately, possibly asking users to indicate their broadband connection type.

Of course, the public availability of the raw data – made possible by the project privacy policy – allows independent researchers and institutions to review Neubot methodology and perform alternative analysis.

Acknowledgment

We would like to thank Att. Monica Alessia Senor, Dr. Eleonora Bassi and Dr. Federico Morando for their support and for making suggestions to improve the quality and readability of this paper.

We would also like to thank Mr. Matt Mathis, who prodded us to relax the fixed-buffer limitation, making the case of high-bandwidth high-delay networks in Korea and Japan.

References

- [1] B. Van Schewick, “Towards an economic framework for network neutrality regulation,” *J. on Telecomm. & High Tech. L.*, vol. 5, p. 329, 2006.
- [2] J. Crowcroft, “Net neutrality: the technical side of the debate: a white paper,” *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 1, pp. 49–56, 2007.
- [3] J. Palfrey. A Citizens’ Choice Framework for Net Neutrality. [Online]. Available: <http://blogs.law.harvard.edu/palfrey/2010/11/03/a-citizens-choice-framework-for-net-neutrality/>
- [4] M. Lemley and L. Lessig, “End of End-to-End: Preserving the Architecture of the Internet in the Broadband Era,” *Ucla L. Rev.*, vol. 48, p. 925, 2000.
- [5] S. Jordan, “Some traffic management practices are unreasonable,” in *Computer Communications and Networks, 2009. ICCCN 2009. Proceedings of 18th International Conference on*. IEEE, 2009, pp. 1–6.
- [6] R. Clayton, S. Murdoch, and R. Watson, “Ignoring the great firewall of china,” in *Privacy Enhancing Technologies*. Springer, 2006, pp. 20–35.
- [7] J. Zittrain and B. Edelman, “Internet filtering in China,” *Internet Computing, IEEE*, vol. 7, no. 2, pp. 70–77, 2003.
- [8] Internet blocking booklet. [Online]. Available: <http://www.edri.org/internet-blocking-brochure>
- [9] Netflix to FCC: scary loophole in net neutrality rules. [Online]. Available: <http://arstechnica.com/tech-policy/news/2010/01/netflix-to-fcc-dont-make-managed-services-a-net-neutrality-loophole.ars>
- [10] How will we know when the Internet is dead? [Online]. Available: <http://arstechnica.com/tech-policy/news/2010/11/are-you-on-the-internet-or-something-else.ars>
- [11] H. Kim, K. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee, “Internet traffic classification demystified: myths, caveats, and the best practices,” in *Proceedings of the 2008 ACM CoNEXT conference*. ACM, 2008, p. 11.
- [12] S. Floyd and V. Jacobson, “Link-sharing and resource management models for packet networks,” *IEEE/ACM Transactions on Networking (TON)*, vol. 3, no. 4, pp. 365–386, 1995.
- [13] OpenBSD. Configuring Queueing. [Online]. Available: <http://www.openbsd.org/faq/pf/queueing.html#altq>
- [14] C. Systems. Weighted Random Early Detection (WRED). [Online]. Available: http://www.cisco.com/en/US/docs/ios/11.2/feature/guide/wred_gs.html

- [15] M. Dischinger, M. Marcon, S. Guha, K. Gummadi, R. Mahajan, and S. Saroiu, “Glasnost: Enabling end users to detect traffic differentiation,” in *Proceedings of the 7th USENIX conference on Networked systems design and implementation*. USENIX Association, 2010, pp. 27–27.
- [16] P. Kanuparth and C. Dovrolis, “Diffprobe: detecting ISP service discrimination,” in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–9.
- [17] Y. Zhang, Z. Mao, and M. Zhang, “Detecting traffic differentiation in backbone ISPs with NetPolice,” in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*. ACM, 2009, pp. 103–115.
- [18] M. Tariq, M. Motiwala, N. Feamster, and M. Ammar, “Detecting network neutrality violations with causal inference,” in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*. ACM, 2009, pp. 289–300.
- [19] Youtube Video Speed History. [Online]. Available: http://www.youtube.com/my_speed
- [20] J. De Martin and A. Glorioso, “The Neubot project: A collaborative approach to measuring internet neutrality,” in *Technology and Society, 2008. ISTAS 2008. IEEE International Symposium on*. IEEE, 2008, pp. 1–4.
- [21] S. Basso, A. Servetti, and J. De Martin, “Rationale, Design, and Implementation of the Network Neutrality Bot.” [Online]. Available: <http://www.neubot.org/neubotfiles/aica2010-neubot-paper.pdf>
- [22] —, “The network neutrality bot architecture: a preliminary approach for self-monitoring of Internet access QoS.”
- [23] Speedtest.net - The Global Broadband Speed Test. [Online]. Available: <http://speedtest.net/>
- [24] R. Beverly, S. Bauer, and A. Berger, “The internet is not a big truck: toward quantifying network neutrality,” *Passive and Active Network Measurement*, pp. 135–144, 2007.
- [25] Switzerland Network Testing Tool — Electronic Frontier Foundation. [Online]. Available: <https://www.eff.org/pages/switzerland-network-testing-tool>
- [26] M. Dischinger, A. Mislove, A. Haeberlen, and K. Gummadi, “Detecting bittorrent blocking,” in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*. ACM, 2008, pp. 3–8.
- [27] NNMA – NNSquad Network Measurement Agent. [Online]. Available: <http://www.nnsquad.org/agent>
- [28] N. Weaver, R. Sommer, and V. Paxson, “Detecting forged TCP reset packets,” in *In Proc. of NDSS*. Citeseer, 2009.
- [29] M. Tariq, M. Motiwala, and N. Feamster, “NANO: Network Access Neutrality Observatory,” in *Proceedings of ACM HotNets*. Citeseer, 2008.
- [30] Y. Zhang, Z. Mao, and M. Zhang, “Ascertaining the Reality of Network Neutrality Violation in Backbone ISPs,” in *Proc. of ACM HotNets-VII Workshop*, 2008.
- [31] WindRider – A Mobile Network Neutrality Monitoring System. [Online]. Available: <http://www.cs.northwestern.edu/ict992/mobile.htm>
- [32] BISMark – Monitor and Manage Your Home Network.
- [33] Grenouille.com - la météo du net depuis 2000. [Online]. Available: <http://www.grenouille.com/>
- [34] R. Prasad, C. Dovrolis, M. Murray, and K. Claffy, “Bandwidth estimation: metrics, measurement techniques, and tools,” *Network, IEEE*, vol. 17, no. 6, pp. 27–35, 2003.
- [35] M-Lab — Welcome to Measurement Lab. [Online]. Available: <http://measurementlab.net/>
- [36] Goodput – Wikipedia, the free encyclopedia. [Online]. Available: <http://en.wikipedia.org/wiki/Goodput>
- [37] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, “RFC2616: Hypertext Transfer Protocol–HTTP/1.1,” *RFC Editor United States*, 1999.
- [38] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, “Modeling TCP throughput: A simple model and its empirical validation,” in *ACM SIGCOMM Computer Communication Review*, vol. 28, no. 4. ACM, 1998, pp. 303–314.
- [39] P. Ohm, “Broken promises of privacy: Responding to the surprising failure of anonymization.”
- [40] IP Addresses Are Personal Data, E.U. Regulator Says. [Online]. Available: <http://www.washingtonpost.com/wpdyn/content/article/2008/01/21/AR2008012101340.html>
- [41] Opinion 1/2008 on data protection issues related to search engines. [Online]. Available: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2008/wp148_en.pdf
- [42] Codice in materia di protezione dei dati personali. [Online]. Available: <http://www.garanteprivacy.it/garante/doc.jsp?ID=1311248>
- [43] Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e scientifici. [Online]. Available: <http://www.garanteprivacy.it/garante/doc.jsp?ID=1556635>

Torino Smart City on Line

Sante Lorenzo Carbone
CSI-Piemonte
Corso Tazzoli 215/12b, 10134 Torino
Santelorenzo.carbone@csi.it

Filippo Ricca
CSI-Piemonte
Corso Unione Sovietica 216, 10134 Torino
filippo.ricca@csi.it

Una città si può definire "smart" quando è vista come un sistema composto da sottosistemi interagenti fra loro. I sottoinsiemi sono rappresentati dalle persone, dal mondo economico, dalle istituzioni, dall'ambiente, dall'energia, ecc. Una città "smart" è realizzata dalla capacità tecnologica di dotare ogni "oggetto" fisico o virtuale (edificio, telefonino, frigorifero, metro, semafori, luci,) di intelligenza attraverso strumenti (sensori, nanotecnologia, attuatori, ecc.) interconnessi tra di loro e "governati". La complessità di un sistema integrato di sottosistemi che interagiscono tra loro (sicurezza, acqua, trasporti, salute, economia, infrastrutture, ecc.) necessita di una governance unitaria permanente e dinamica, necessaria a poter condividere le informazioni necessarie. Il documento illustra la piattaforma "Torino Smart City on line" inquadrandola nel contesto dell'iniziativa europea Smart Cities ed esponendone le sue principali funzionalità e la sua mission di ambiente on line condiviso e partecipato per la realizzazione di una città intelligente.

1 Contesto

Con l'iniziativa Smart Cities, la Commissione Europea sosterrà le città che si impegnano a incrementare l'efficienza energetica dei propri edifici, delle reti energetiche e dei sistemi di trasporto in modo tale da ridurre, entro il 2020, del 40% le proprie emissioni di gas serra.

La Città di Torino sta preparando la sua candidatura per diventare una Smart City con l'apporto del Politecnico, dell'Università di Torino e del CSI-Piemonte che intende offrire a tale candidatura il proprio supporto per quanto riguarda le componenti ICT a sostegno delle strategie della Città.

Per predisporre tale candidatura la Città di Torino dovrà pertanto elaborare, lavorando in stretta sinergia con i soggetti pubblici e privati che operano sul

territorio, una piattaforma progettuale, un insieme di azioni che mirino a rendere Torino “città intelligente” in grado di produrre alta tecnologia, ridurre i consumi energetici degli edifici, promuovere trasporti puliti e migliorare in generale la qualità della vita dei suoi abitanti all'insegna delle basse emissioni di anidride carbonica.

Il progetto “*Torino Smart City on line*” coglie la sfida per proporre un sistema di monitoraggio, governance e comunicazione dell'iniziativa che stabilisca una relazione partecipativa on line tra gli enti, i cittadini, le imprese nella coprogettazione di Torino Smart City e dove la tecnologia diviene strumento per il miglioramento della vita nella città.

2 Torino Smart City on line

Con l'iniziativa Smart City Torino, nei prossimi anni, subirà grandi trasformazioni strutturali che la renderanno una “città intelligente” in grado di produrre alta tecnologia, ridurre i consumi energetici degli edifici, promuovere trasporti puliti, assicurare maggior sicurezza ai cittadini, migliorare in generale la qualità della vita dei suoi abitanti.

Una città è “smart” quando consente uno sviluppo economico sostenibile che garantisce nel contempo un alto livello della qualità della vita, un minor spreco di risorse, l'inclusione (partecipazione/collaborazione) di tutti i cittadini.

Una smart city si realizza su sei assi principali : smart economy; smart mobility; smart environment; smart people; smart living; smart governance.

In questo contesto si collocano trasversalmente tutte le tematiche che possono trovare applicazione nella vita quotidiana delle città e delle persone, come l'energia, la sicurezza, lo smart building, la cultura, ecc .

In questa città, i sistemi tecnologici sul territorio sono integrati e collegati, consentendo il monitoraggio e la raccolta di dati provenienti dagli “oggetti” connessi (internet delle cose) rendendoli disponibili alla politica per la governance ed ai cittadini sotto forma di servizi.

La comunicazione rappresenterà un punto cruciale e fondamentale per il successo del progetto Torino Smart City, sia in termini di ritorno di immagine per la città, sia in relazione al consenso e alla partecipazione di cittadini e stakeholders coinvolti.

Sarà necessario far comprendere loro le trasformazioni a cui la città sarà sottoposta nei prossimi anni, accompagnandoli nei processi di cambiamento che una città smart comporta e creando la consapevolezza sul fatto che saremo di fronte a mutamenti di natura culturale che coinvolgeranno gli amministratori, le aziende e i cittadini: vivere in una città intelligente significherà vivere all'interno di una rete di relazioni umane che avranno la capacità di generare saperi e idee attraverso il confronto e aprire le porte a nuove relazioni e a contaminazioni con esperienze di altri progetti, con altre città, con centri di ricerca e istituzioni.

La piattaforma “*Torino Smart City on line*” si pone tre macro obiettivi basati su tre parole chiave:

- **governare** per attuare una piena governance del territorio e di tutte le fasi del progetto Torino Smart City;
- **conoscere** per far comprendere cosa significa vivere in una città smart;
- **partecipare** per creare un ambiente di partecipazione per la progettazione e la realizzazione di Torino smart.

3 Governare

La realizzazione di una città smart è un progetto ambizioso e che coinvolge ogni aspetto ed interesse della città e passa attraverso la definizione di un nuovo modello di sviluppo e trasformazione delle città.

Torino ha identificato una ambito sul quale verrà avviato tale percorso di realizzazione a 360° della visione di Città Intelligente, la cosiddetta “Variante 200” che si colloca a cavallo delle Circoscrizioni 5,6 e 7.



Fig. 1 – La “variante 200” (perimetro in giallo)

In questa zona verrà avviato un progetto dimostrativo pilota nell’ambito di una strategia di sviluppo integrato della Città, che attraverso una serie di interventi chiave prioritari, avvierà per aggregazioni successive la crescita di Torino Smart City.

Gli obiettivi generali sono quelli di realizzare e/o integrare alcuni servizi smart e offrire servizi innovativi nell’ambito della sostenibilità ambientale, dell’efficienza energetica, della governance degli edifici, della condivisione dell’infrastruttura digitale sia come mezzo per la raccolta e diffusione delle informazioni, sia come mezzo per l’erogazione dei servizi innovativi.

Il progetto “*Torino Smart City on line*” sarà il veicolo per introdurre nuove tecnologie abilitanti una miglior sostenibilità ambientale e per attrarre consenso sociale.

Il primo obiettivo del progetto si poggia su quattro driver principali: le reti; lo smart building; il territorio e l’ambiente; la piattaforma.

3.1 Le reti

La città è un insieme di soggetti e oggetti che devono essere inteconnessi tra di loro per poter divenire sistema intelligente.

A partire dalle infrastrutture ICT disponibili, integrate con la capacità trasmissiva delle reti di illuminazione pubblica (power line communication), verrà realizzata una “sensors network”, che consentirà non solo di gestire processi di efficientamento e risparmio energetico, ma abiliterà la capacità di erogare “smart services” al cittadino (es. estensione delle videosorveglianze in aree critiche, disponibilità di informazioni sull’ambiente, informazioni stradali, wifi, diffuso, ecc).

3.2 Smart Building

Lo Smart Building è fondamentale nella costruzione di una Smart City (attraverso l’aggregazione in Smart District) per rendere attuabile una strategia integrata finalizzata ad una maggior efficienza energetica e sostenibilità ambientale.

Sul territorio definito saranno selezionati alcuni edifici, i cui sensori già presenti o che saranno implementati, saranno interfacciati in un sistema di Smart Building & Energy Management, per la raccolta dei dati in “real time”.

Gli edifici disporranno quindi di un sistema di Energy Management che coniuga l’insieme delle azioni tecnico - gestionali finalizzate al monitoraggio e controllo dei sistemi energetici, con l’obiettivo di razionalizzare i consumi di energia (elettricità, gas, ecc,...), assicurando nel contempo la riduzione dell’impatto ambientale (minori emissioni inquinanti, Co2, ecc.), basandosi su un modello di:

- pianificazione: definizione degli obiettivi in modo sistemico
- esecuzione: individuazione degli strumenti necessari al loro raggiungimento;
- controllo: identificazione delle opportunità di miglioramento;
- azione/correzione: riduzione dei costi legati ai consumi ed alle emissioni.

Sarà parzialmente avviato anche l’interfacciamento diretto a POD (point of delivery) su alcuni edifici e l’inserimento di queste informazioni direttamente a sistema.

L’approccio richiede di disporre di informazioni sull’edificio, sia storiche che attuali. Nell’ambito della “Variante 200” verranno raccolti e messi a sistema i dati catastali, i dati geografici, i rilievi (es. da remote sensing), i dati sulla certificazione energetica (SICEE).

3.3 Il territorio e l’ambiente

Un corretto approccio alla sostenibilità ambientale sul tema dell’efficienza energetica non può prescindere da una conoscenza adeguata e diffusa delle condizioni energetiche complessive di un territorio.

Quindi parallelamente ai modelli di dettaglio delle prestazioni degli edifici individuati si avvierà la realizzazione di un modello di valutazioni a larga scala, relative allo stato di fatto dell’insieme di edifici che ricadono nel medesimo territorio definito dalla Variante 200. Coniugato con i dati catastali, con i dati geografici, con i rilievi (remote sensing), con i dati sulla certificazione energetica

(SICEE), i dati relativi alle variabili ambientali (PM10, Co2, ecc), il patrimonio informativo complessivo degli edifici analizzati consentirà una vista “real time” del territorio che permetterà di realizzare e raffinare un modello di lettura, interpretazione e geo-referenziazione dei dati di consumo energetico e di impatto ambientale.

Sarà così possibile ottenere una serie di informazioni sulla situazione attuale dell'edificato che consentirà di osservare la città sotto diverse “viste”.



Fig. 2 –Edifici associati alla classe energetica, fonte SICEE (gli edifici in grigio non ancora sono associati) – dati esempio su Torino periodo maggio 2011.

3.4 La piattaforma

Contrariamente al significato intrinseco di “smart”, il progetto Torino smart City non potrà che essere complesso e oneroso, vista l’elevata numerosità degli attori coinvolti, dei progetti sottesi e degli indicatori quantitativi e qualitativi da monitorare.

La piattaforma “*Torino Smart City on line*” si propone di dare un modello di monitoraggio e di gestionale dinamico, aperto e collaborativo per la governance complessiva del progetto.

Dal punto di vista gestionale la piattaforma integrerà competenze correlate:

- **Gestione della rendicontazione** aspetti amministrativi e di legittimità correlati agli oneri di rendicontazione vs l’EU
- **Valutazione degli impatti e dell’efficacia dell’azione pubblica** (Indicatori di efficacia – KPI) aspetti gestionali e prestazionali di merito correlati a valutare impatti ed efficacia dell’azione pubblica e delle politiche pubbliche
- **Gestione dell’andamento del programma nel suo complesso** (Project Management) aspetti operativi di gestione delle singole linee progettuali, correlate all’andamento del programma nel suo complesso.

A tal fine la piattaforma sarà dotata di

- sistemi in grado di gestire gli iter di finanziamento delle iniziative per favorire la rendicontazione amministrativa in ottemperanza agli obblighi dettati dal Programma Europeo Smart City (Gestione della rendicontazione);
- strumenti in grado di misurare le performance di una politica o di un programma e per assumere decisioni strategiche, operative ed

organizzative, con il coinvolgimento attivo dei portatori di interesse (Valutazione degli impatti e dell'efficacia dell'azione pubblica)

- soluzioni in grado di gestire valutare lo stato avanzamento dei singoli progetti, delle linee e del programma Smart City in ottica di dialogo e coordinamento tra partner e Città di Torino (Gestione dell'andamento del programma nel suo complesso).

Il modello di governance adottato si propone di realizzare un nuovo modello di azione pubblica, che esprima l'esigenza di allargare il consenso sociale e produttivo attraverso l'introduzione di meccanismi e prassi che rendano effettiva la partecipazione dei cittadini all'azione pubblica, in un contesto di trasparenza amministrativa, e che attribuisca chiaramente la responsabilità delle decisioni e delle azioni ai vari attori coinvolti, in modo da migliorarne l'efficacia e l'efficienza, nonché la coerenza

Relativamente al monitoraggio e governo del territorio, la piattaforma consentirà la rilevazione permanente, dinamica e geo-referenziata (Gis 3d) dei fenomeni e di tutte le variabili derivate dai sistemi informativi integrati, dai sensori interfacciati (o dai sottosistemi come ad esempio smart building esistenti), dai modelli di analisi dei dati stessi.

A supporto di tale rappresentazione, la Città di Torino sfrutterà il modello che descrive quasi tutte le dimensioni della politica pubblica cittadina alla base del Cruscotto sulla Sicurezza Urbana che è un sistema permanente di rilevazione dinamica e geo-referenziata dei fenomeni e di tutte le variabili che concorrono a determinare le condizioni di sicurezza dei territori e la percezione d'insicurezza delle persone.

Il modello e la filosofia da cui nasce sono coerenti al modello di "smart city" descritto e caratterizzato da un sistema integrato e gestito unitariamente, di sottosistemi che interagiscono tra loro (sicurezza, acqua, trasporti, salute, economia, infrastrutture, ecc.), allo scopo di consentire uno sviluppo economico sostenibile e garantire un alto livello della qualità della vita, un minor spreco di risorse e una maggior inclusione di tutti i cittadini.

Saranno così integrate tutte le informazioni disponibili o derivabili dal territorio, un insieme di strumenti e metodi (viste, report, elaborazioni 3d) che consentiranno il controllo conoscitivo del territorio, attraverso un'attività di supporto metodologico alla realizzazione di modelli interpretativi dei fenomeni, e l'elaborazione di indicatori di sintesi che saranno georeferenziati sul territorio a diversi livelli: la città, le circoscrizioni, i quartieri, le zone statistiche.

Ciascun indicatore è visualizzato attraverso mappe, tabelle e grafici.

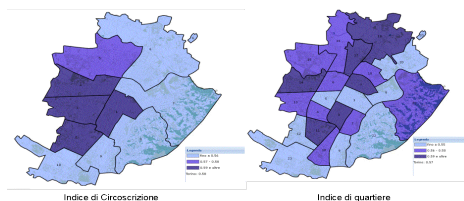


Fig. 3 – un esempio di Cruscotto

4 Conoscere

Oltre alla comunicazione tradizionale, il fulcro più attivo e dinamico di questo flusso di informazioni sarà rappresentato da un ambiente on line, reso disponibile sulla piattaforma di *“Torino Smart City on line”*, che sarà in grado di gestire la comunicazione e la relazione con la città, ma anche di governare tutte le fasi progettuali del progetto “Smart City”.

Per conoscere e vivere all'interno di una città smart i cittadini avranno la possibilità di consultare un ambiente on line di supporto al cambiamento che metterà a disposizione:

- notizie riguardanti lo stato dei progetti e la roadmap per arrivare a Torino Smart City;
- forme di comunicazione visuale (video/animazioni) di breve durata che spieghino con efficacia cos'è una città smart;
- dati rappresentati in forma visuale, attraverso infografiche, che aiutino alla comprensione dei temi legati a “Torino smart”;
- modalità di comunicazione interattiva (serious game / digital storytelling) che assicurino il massimo coinvolgimento degli utenti;
- modalità automatiche che consentano la condivisione dei contenuti sui principali social network con canali dedicati (facebook / twitter / youtube / google+...).

5 Partecipare

Attraverso un ambiente di partecipazione on line sarà possibile attuare forme di collaborazione dal basso per un'efficace formazione del consenso e per trarre benefici dall'intelligenza collettiva.

I modelli di partecipazione individuati si basano su quattro approcci in relazione al tipo di coinvolgimento richiesto:

- coinvolgere per formare consapevolezza sui temi di smart city e sul cambiamento degli stili di vita individuali e collettivi;
- coinvolgere per orientare le politiche pubbliche;
- coinvolgere per raccogliere idee progettuali che attraverso un iter condiviso possano trasformarsi in progetti concreti per la città smart;
- coinvolgere per creare conoscenza e dialogo attivo sui progetti smart già avviati.

Ogni modello individua un processo e/o degli strumenti per la sua realizzazione.

5.1 Formazione della consapevolezza

L'obiettivo della formazione della consapevolezza oltre che essere già insito nell'ambiente legato alla conoscenza e in forme più tradizionali di comunicazione, può essere raggiunto anche attraverso strumenti di edutainment quali i Serious Game per la generalità dei cittadini e il Digital Storytelling per le scuole.

I Serious Game contribuiscono a raccogliere il sensing di un'elevata fetta di popolazione e saranno utilizzati come strumento di conoscenza, promozione e diffusione dei concetti e dei paradigmi dell'essere e diventare "smart", dando la possibilità di "misurare" il proprio livello di consapevolezza dell'essere smart.

Attraverso il serious game, con un approccio ludico, si darà la possibilità di sperimentare progetti, azioni e attività per costruire la nuova città o il nuovo quartiere del futuro (la Torino del 2050) in modo "utopico o distopico", rappresenteranno l'ambiente reale dove sperimentare e attuare comportamenti individuali e collettivi orientati ad un approccio smart anche attraverso una logica di tipo partecipativo. Le modalità dei Serious Game potranno essere molteplici: multiplayer, multidevices, multi-accessibili tramite una logica openid, in ogni caso i risultati, grazie alla realizzazione tramite il web, potranno essere tracciati, analizzati e valorizzati per apportare maggiore conoscenza sul tema trattato. Attraverso il Digital Storytelling si potranno sperimentare nelle scuole nuove dinamiche partecipative e di mash-up culturale, veicolate dal web 2.0, all'interno di contesti didattici formali. "Creare" e "Raccontare" storie in rete sulla città o il quartiere "desiderato" in un ottica smart può diventare così una occasione di apprendimento partecipato che coinvolge docenti, studenti e famiglie. La storia migliore della scuola potrà diventare poi la "storia" migliore del quartiere attraverso strumenti di peer-evaluation con un approccio da "concorso" delle idee il cui risultato finale può essere un vero e proprio video dimostrativo.

5.2 Orientamento delle politiche pubbliche

L'approccio smart passerà anche attraverso un cambiamento di approccio da parte di chi amministra la cosa pubblica. Tale cambiamento comporterà l'adozione del modello "share --> discuss --> vote" finalizzato all'orientamento delle politiche pubbliche.

La Città utilizzerà all'interno della piattaforma *"Torino Smart City on line"* strumenti di consultazione del sentire collettivo secondo il seguente modello:

- condivisione di temi sui quali la Città vuole sentire l'opinione della gente e degli stakeholders;
- discussione sui temi individuati;
- votazione finalizzata all'orientamento della decisione sulla base dei reali bisogni della collettività.

5.3 Raccolta e selezione di idee progettuali

Il modello individuato utilizza strumenti di partecipazione mista, on line e off line, consentendo un coinvolgimento attivo e partecipato per la raccolta di idee progettuali da trasformare in progetti concreti per la città. Sono stati individuati due processi:

- dai Barcamp alla realizzazione di un progetto;
- dal Contest alla realizzazione di un progetto.

Il primo processo prende il via con una serie di Barcamp tematici aperti a tutti i cittadini che hanno interesse a partecipare alla discussione su Torino Smart City.

Il Barcamp viene definito una *non-conferenza* “, nasce dal desiderio delle persone di condividere e apprendere in un ambiente aperto e libero. Il BarCamp è una non-conferenza collaborativa, dove chiunque può “salire in cattedra”, proporre un argomento e parlarne agli altri, con lo scopo di favorire il libero pensiero, la curiosità, la divulgazione e la diffusione dei temi legati al Web. Una non conferenza (unconference) quindi una riunione il cui tema di discussione è deciso dai partecipanti piuttosto che prestabilito in anticipo dagli organizzatori, una riunione aperta i cui contenuti vengono proposti dai partecipanti stessi”. Il lancio dei Barcamp, per la raccolta delle adesioni e l'organizzazione puntuale dell'evento, viene effettuato nell'ambiente on line e tramite i social network. Il Barcamp sarà il luogo dove ognuno potrà presentare idee per Torino smart, le idee più votate al Barcamp verranno successivamente inserite nell'ambiente on line per essere discusse e per partecipare ad un ulteriore processo di selezione. Le migliori idee (le più votate nell'ambiente on line) diventeranno progetti per Torino Smart. Tutte le fasi successive del progetto verranno comunicate e sottoposte a successiva discussione, raccolta di feedback ecc. all'interno dell'ambiente on line attraverso questionari on line, sondaggi, forum tematici.

Il secondo processo è interamente on line e prende il via con l'organizzazione di Contest on line tematici, per raccogliere idee progettuali, aperti a tutti i cittadini che hanno interesse su Torino Smart City.

I Contest sono strumenti che si rivolgono all'intelligenza collettiva per coinvolgere il pubblico nella definizione di idee, soluzioni e proposte attraverso la partecipazione ad una competizione che culmina nell'assegnazione di un premio o un incentivo. Presuppongono un coinvolgimento estensivo del pubblico (raccolta contributi+valutazioni+rating) e hanno anche una funzione promozionale, ad esempio attraverso il reindirizzamento sui social media .

Le idee progettuali raccolte attraverso il Contest vengono valutate e selezionate da una giuria tecnica. I progetti selezionanti ricevono un premio e diventeranno progetti per Torino Smart City.

Anche in questo caso tutte le fasi successive del progetto verranno comunicate e sottoposte a successiva discussione, raccolta di feedback ecc. all'interno dell'ambiente on line attraverso questionari on line, sondaggi, forum tematici.

5.4 Creazione di conoscenza e dialogo attivo

Il coinvolgimento di cittadinanza e stakeholder per creare conoscenza e dialogo attivo verrà realizzato nell'ambiente on line attraverso due processi:

- dal Workshop al Forum di discussione;
- dagli Action Planning al Forum di discussione

La Città e/o gli Stakeholder di progetto organizzano dei Workshop e/o degli Action Planning relativi ad un singolo progetto con l'obiettivo di aprire un tema di discussione on line su specifici aspetti del progetto. I temi individuati vengono proposti all'interno di Forum di discussione on line moderati e aperti a tutti. I

suggerimenti, i feedback e le eventuali criticità che emergono dalle discussioni all'interno dei forum possono servire a diffondere conoscenza e a migliorare in progress l'andamento del progetto

6 Conclusioni

La piattaforma "*Torino Smart City on line*" si propone di realizzare un approccio all'amministrazione della cosa pubblica condiviso e partecipato, allineato alla logica di una città *smart*, dove l'intelligenza collettiva partecipa attivamente ai processi decisionali e contribuisce alla realizzazione concreta di una città intelligente che si rende fruibile attraverso una serie di servizi al cittadino in modo integrato e sui vari devices disponibili: totem, tablet, smartphone: servizi di "realtà aumentata" ; servizi di informativi e di aggiornamento dei servizi pubblici; servizi di "social networking di strada"; servizi legati all'ambiente .

7 Bibliografia

[RIF] Wikipedia - <http://it.wikipedia.org/wiki/BarCamp>

[Saperi PA] – Gianni Dominici, I contest nella PA, scorciatoia o nuovo strumento di partecipazione ?, Editoriale FORUM PA SAPERI PA, 19/07/2011 - <http://saperi.forumpa.it/story/63494/i-contest-nella-pa-scorciatoia-o-nuovo-strumentodi-partecipazione>

Bruner J. (1991), La costruzione narrativa della realtà, in M. Ammaniti, D.n. Stern (a cura di), Rappresentazioni e narrazioni, Laterza, Roma Bari, pp.17-42.

Damasio A. (1999) The Feeling of What Happens: Body and Emotion in the Making of Consciousness,

Harcourt. Taddeo G. (2009), Experiential and narrative elements in the learning processes. Digital storytelling as a cultural re-elaboration practice, REM. Research on Education and Media, v.1, (2), Erickson, Trento.

Gee, J. P. (2008). Learning and games. In Katie Salen (Ed.) The ecology of games: Connecting youth, games, and learning (John D. and Catherine T. MacArthur Foundation series on digital media and learning). Cambridge, MA: The MIT Press.

Klopfer, E., Osterweil, S. & Salen, K. (2009) Moving learning games forward. Retrieved from: http://education.mit.edu/papers/MovingLearningGamesForward_EdArcade.pdf

Koster (2004). A theory of fun. Paraglyph Press. New York, NY.

Lazzaro, N. (2004). Why we play games: Four keys to more emotion without story. Retrieved from: http://www.xeodesign.com/xeodesign_whyweplaygames.pdf

Un'ontologia formale per la creazione automatica di proposte formative: il progetto E-VALUE

Luciano Serafini¹, Volha Bryl¹, Sofia Cramerotti², Chiara Ravagni²,

¹Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, TN
serafini@fbk.eu , bryl@fbk.eu

²Centro Studi Erickson, Via del Pioppeto 24, 38121 Gardolo, Trento, TN
safia.cramerotti@erickson.it , chiara.ravagni@gmail.com

In questo articolo viene illustrato lo schema di rappresentazione e classificazione dei learning object utilizzato all'interno del progetto E-VALUE. Viene inoltre descritto come questo è stato codificato in un'ontologia in linguaggio OWL e vengono mostrati alcuni esempi di query.

1. Introduzione: il progetto E-VALUE

L'individuazione precoce dei fattori di rischio e la successiva messa in atto di percorsi di intervento riferiti alle principali **difficoltà di apprendimento** è un processo che va strutturato seguendo rigorose procedure e parametri di riferimento, basati sulle più accreditate classificazioni e Linee guida condivise dalla comunità scientifica. In particolare la **Consensus Conference sui DSA** (2009) riporta degli indicatori di rischio che hanno mostrato un buon valore predittivo rispetto alle difficoltà di apprendimento; la loro individuazione dovrebbe permettere successivamente di attivare percorsi adeguati di recupero/potenziamento.

Partendo da queste premesse, il **progetto E-VALUE** intende offrire una gamma di **strumenti di valutazione e di materiali per l'intervento educativo-didattico** nei principali ambiti di apprendimento (in particolare lettura, scrittura e calcolo). La finalità è quella di strutturare un sistema integrato per la valutazione di tali abilità e la gestione del processo di intervento didattico.

1.1 Obiettivi e finalità del progetto

L'obiettivo principale del progetto E-VALUE è progettare, sviluppare e sperimentare un sistema informatico integrato – piattaforma multimediale (costituito da un insieme di applicativi web), che consenta la **valutazione delle abilità di apprendimento e la gestione del processo di intervento didattico** all'interno di diversi contesti educativi. Tale sistema è costruito secondo un modello che integra processi differenti, interconnessi sotto il profilo funzionale: procedure di valutazione standard, valutazione dinamica, raccolta di dati normativi e continua "alimentazione" (aggiornamento della banca dati e

confronto con dati normativi attualmente a disposizione), lettura qualitativa oltre che quantitativa dei dati raccolti, impostazione dei percorsi di intervento, metodologie/materiali didattici direttamente spendibili nella progettazione educativo-didattica quotidiana. Tutti questi processi/attività fanno parte di modelli di valutazione/intervento attuati da anni “direttamente sul campo” in scuole di ogni ordine e grado, e che fanno riferimento alle più accreditate modalità di sperimentazione educativo-didattica: percorsi di ricerca-azione, utilizzo di user-group, procedure di test-retest, uso/confronto tra gruppi sperimentali e gruppi di controllo.

In quest'ottica il **sistema implementato all'interno del Progetto E-VALUE** permette quindi:

- la gestione e l'erogazione di servizi, strumenti di rilevazione e contenuti didattici finalizzati alla valutazione psicoeducativa degli alunni in contesti scolastici;
- la conservazione dei contenuti (materiali per la valutazione, materiali per il recupero e il potenziamento), in formato digitale all'interno di appositi archivi al fine di facilitare la ricerca e l'utilizzo delle risorse;
- la gestione delle fasi (workflow) del processo di valutazione attraverso un sistema (workflow management system) in grado di gestire in modo automatico il passaggio dei dati tra le varie fasi diminuendo i tempi di risposta e aumentando l'efficienza del servizio;
- la standardizzazione dei metodi e degli strumenti utilizzati nelle fasi di valutazione /scoring;
- la rappresentatività dei dati rilevati grazie al progressivo ampliamento della banca dati normativa di riferimento.

1.2 Articolazione del Modello E-VALUE

Il modello E-VALUE si articola in:

Valutazione: scelta dello strumento e modalità di somministrazione; somministrazione delle prove secondo la modalità scelta (individuale/collettiva); raccolta delle risposte e invio dei risultati.

Scoring/Restituzione: elaborazione dati quanti-qualitativa (comunità di esperti); restituzione dei risultati in presenza e a distanza; intervento didattico, e strutturazione del percorso di recupero/potenziamento delle abilità/competenze del singolo o del gruppo classe (esperti, referente scolastico).

Intervento: progettazione del percorso di intervento (metodologie-contenuti) da parte del referente scolastico in collaborazione con l'esperto; intervento in laboratori didattici in cui vengono coinvolti tutti gli alunni della classe; tutoraggio e monitoraggio in itinere delle attività di recupero/potenziamento.

Tutto il percorso è preceduto e accompagnato da una formazione continua: iniziale, in itinere (su esito dei risultati e progettazione delle attività) e da momenti di consulenza/tutoraggio, nonché da un'attenta verifica dell'efficacia del progetto educativo-didattico che si sta attuando.

1.3 Analisi dei materiali

Lo scopo dell'analisi dei materiali di intervento didattico è quello di selezionare le parti, da libri o da CD, che intervengono su un determinato processo cognitivo. Per farlo è stato necessario:

- identificare i processi cognitivi
- analizzare il materiale pubblicato dalle Edizioni Centro Studi Erickson e scomporlo in parti autonome
- trasformare il materiale in modo da trasferirlo dal supporto ottico o cartaceo a una struttura in grado di ricomporlo secondo necessità, online e offline.

Nella struttura tradizionale del libro o del CD, infatti, i materiali sono ordinati secondo un percorso didattico progressivo che coinvolge svariati processi cognitivi. Nella struttura del progetto E-VALUE invece i materiali vengono ricombinati in modo nuovo, divisi in parti con contenuto e processo cognitivo uniforme e utilizzabili in determinate condizioni (ad esempio una fascia d'età, oppure singoli alunni piuttosto che l'intera classe).

2. La catalogazione

Si poneva quindi la necessità di predisporre una forma di catalogazione adatta a descrivere il materiale da ogni punto di vista utile al futuro sistema, per la stesura della quale si è tenuto conto dei seguenti aspetti:

- la necessità di descrivere il materiale analizzato dagli esperti in modo completo, dettagliato e bibliograficamente uniforme;
- la collocazione in una struttura di dati da cui fosse sempre deducibile il materiale d'origine e i materiali che da qualunque punto di vista (editoriale, contenutistico, informatico) erano ad esso correlati;
- il bisogno di organizzare la descrizione in una lista di "tag" tali da essere compresi e interpretati automaticamente da un sistema di decisione automatica o da un motore di ricerca per il reperimento di materiale da parte di formatori e collaboratori.

2.1. Descrizione degli oggetti

Si è scelto quindi di produrre una struttura di catalogazione dettagliata che combinasse i campi catalografici del Dublin Core con campi appositamente creati per rispettare le necessità del gruppo di lavoro e del progetto.

I 15 campi del Dublin Core si sono rivelati sufficientemente generici da poter appoggiare ad essi un ampio numero di qualificatori per un maggiore dettaglio nella descrizione. Inoltre sono stati aggiunte alcune categorie di tag più

propriamente amministrative/gestionali, per poter trattare il materiale anche dal punto di vista commerciale e dei contratti con gli autori.

I tag si possono quindi raggruppare in:

- estratti dal Dublin Core
- descrittivi
- gestionali.

La descrizione rispecchia una struttura a faccette spuria, considerando che i campi di tipo contenutistico (quelli che riguardano i processi cognitivi, le funzioni cognitive associate e il livello di competenza) permettono l'accesso ai materiali tramite *query* diverse.

2.2 I vocabolari controllati

Viste le difficoltà nella gestione delle forme di catalogazione precedenti si è deciso di corredare la tabella con una serie di dizionari controllati. Ad ogni campo (tranne i campi autore/titolo e alcuni campi testuali liberi) è stata affiancata una lista di parole che compongono una scelta obbligata per il catalogatore. Questo è necessario per evitare la scelta di termini simili o dello stesso concetto ma in un'altra forma grammaticale (singolare o plurale, aggettivo o sostantivo), fatto che renderebbe molto difficoltoso il *mapping* dei materiali per il DSS. Questi dizionari sono stati concordati durante le riunioni del gruppo di lavoro.

Nel caso dei campi riguardanti il contenuto (area tematica, abilità generale, processo, sottoprocesso) per poter descrivere in modo dettagliato il materiale è stato compilato un amplissimo dizionario controllato con una struttura ad albero su quattro livelli che copre gli ambiti di: matematica, letto-scrittura, metafonologia e abilità linguistiche. Per la stesura del dizionario sono state necessarie periodiche riunioni con esperti delle varie tematiche.

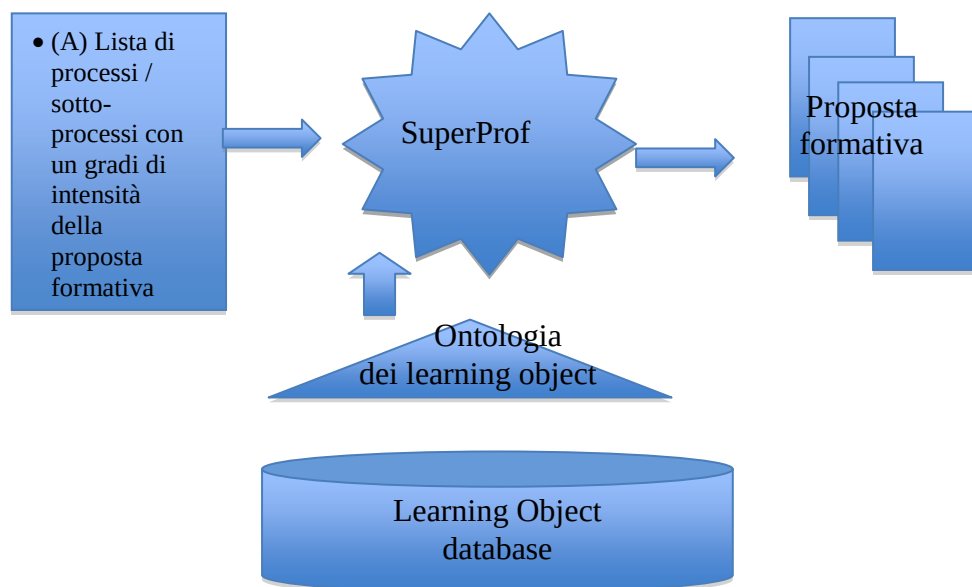
Per riuscire a descrivere l'ambito in modo adeguato si è utilizzato il metodo del brainstorming con esperti, seguito da fasi di test in cui le varie combinazioni di metadati sono state provate nella loro efficacia e modificate quando si rivelavano poco funzionali o incomplete.

3. L'ontologia

Prima di introdurre specificatamente l'ontologia associata al progetto E-VALUE, è importante capire quale sia l'utilizzo di questa ontologia all'interno del progetto. Per questo, nella prima parte di questo paragrafo descriveremo a un livello molto generale, il sistema di supporto alle decisioni per la generazione di proposte formative che verrà realizzato all'interno del progetto.

3.1 Il sistema di supporto alle decisioni

Il sistema di supporto alle decisioni che verrà realizzato in E-VALUE, denominato SuperProf (**S**istema di **S**upporto alle decisioni **PER** la creazione di **PRO**poste **FOR**mativa) ha la funzionalità principale di generare, a partire dalla specifica di una necessità formativa, data in termini di processi/sottoprocessi formativi, una proposta formativa, composta da una insieme di *learning object*. Una visione ad alto livello del sistema di supporto alle decisioni SuperProf è schematizzata nella seguente figura



Il sistema di supporto alle decisioni, SuperProf prende in input una lista di processi/sotto-processi, ognuno dei quali è associato a una serie di parametri che indicano l'intensità dell'intervento formativo necessario in quel processo. In aggiunta a queste informazioni, al sistema vengono forniti una serie di parametri, quali ad esempio: numero totale di *learning object* di cui è composta la proposta formativa, percentuali di mix tra diverse tipologie di *learning object*, che verranno utilizzate da SuperProf per modulare la proposta formativa.

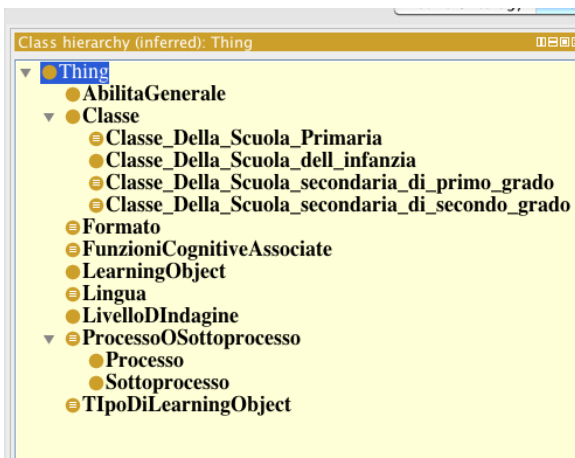
A fronte di questo tipo di input il DSS SuperProf, in una prima fase, seleziona tra tutti i *learning object* disponibili nel catalogo Erickson quelli rilevanti per i processi/sottoprocessi specificati nell'input (A), e successivamente sceglie tra quelli estratti dal database un sottoinsieme che soddisfi nel miglior modo possibile i criteri aggiuntivi (B) specificati nell'input. La classificazione dei *learning object* ha un ruolo fondamentale soprattutto nella prima fase di selezione. Infatti in questa fase, a fronte di una richiesta di attività formativa su un insieme di processi/sottoprocessi, verranno selezionati i *learning object* che sono stati classificati con i relativi tags, o con tags equivalenti. È in questa fase che la codifica della classificazione dei LO in un'ontologia diventa

fondamentale, in quanto permette di realizzare la selezione/estrazione degli LO rilevanti in modo semplice, flessibile ed efficiente.

3.2 L'ontologia di E-VALUE

Il termine Ontologia può avere molti significati ma nell'ambito dei sistemi per la gestione della conoscenza questo termine viene ormai costantemente associato alla definizione data da Gruber (1993) secondo la quale un' ontologia è la caratterizzazione di una concettualizzazione

condivisa. Il linguaggio standard per la caratterizzazione delle ontologie che si è affermato negli ultimi anni è OWL (Ontology Web Language). Questo linguaggio, la cui semantica formale è basata sulla logica descrittiva, permette di descrivere un' ontologia tramite due componenti principali, la T-box (*terminological box*) che contiene la gerarchia delle



classi-sottoclassi dell'ontologia e le relazioni che possono esistere tra le varie classi, e la A-Box (*assertional box*) che contiene gli oggetti dell'ontologia, la loro classificazione (cioè l'appartenenza di un oggetto ad una classe/sotto-classe) e le relazioni tra gli oggetti. L'ontologia di E-VALUE può essere visualizzata utilizzando l'editor di Ontologie *Protegé* disponibile al sito <http://protege.stanford.edu/>

La gerarchia delle classi dell'ontologia di E-VALUE, mostrata nella seguente figura, è composta dalle seguenti classi principali

AbilitaGenerale è la classe che contiene tutte le abilità generali, associate ai vari livelli di indagine, individuate nel modello di classificazione di E-VALUE. Esempi di elementi di questa classe sono: *“Letto-scrittura, decodifica”*, *“Letto-scrittura, comprensione”* e *“Letto-scrittura, compilazione”*.

Classe questa classe contiene tutte le classi dalla scuola dell'infanzia alla scuola secondaria di secondo grado, organizzate in sottoclassi, per ogni periodo scolastico omogeneo. Esempi di istanze di questa classe sono: *1P*, *2P* ... *5P* (le cinque classi della scuola primaria) *1S1*, *1S2*, *1S3* (le tre classi della scuola secondaria di primo grado).

Formato è la classe che contiene tutti i possibili tipi di formato dei *learning object*. Gli elementi appartenenti a questa classe sono: *cartaceo*, *ebook*, *materiale*.

FunzioniCognitiveAssociate è la classe che contiene tutte le funzioni cognitive secondo la classificazione introdotta nella sezione precedente. Alcuni

esempi di elementi di questa classe sono: *FCA_Attenzione_verbale*, *FCA_Abilita_visuo_spaziali*, *FCA_Attenzione_uditiva*, *FCA_Attenzione_visiva*.

LearningObject è la classe che contiene tutti i *learning object* presenti nel catalogo. Ogni *learning object* ha un identificativo unico.

Lingua è la classe che contiene l'insieme delle lingue nelle quali sono scritti i *learning object*.

LivelloDindagine è la classe che contiene i diversi livelli di indagine individuati nello schema di E-VALUE. Alcuni esempi di livelli di indagine sono: *Metafonologia*, *Letto-scrittura*, *Abilità linguistiche*, *Matematica-Geometria*,

ProcessoOSottoprocesso è la classe che contiene tutti i processi raggruppati nei due livelli, processi e sottoprocessi. Esempi di processi sono: *Metafonologia-globale-rime* e *Metafonologia-globale-sillabe*; i relativi sottoprocessi sono *riconoscimento* e *produzione* (sottoprocessi del primo) e *fusione orale* e *fusione orale con supporto visivo* (sottoprocessi del secondo)

TipoDiLearningObject è la classe che contiene tutte le tipologie di *learning object* quali ad esempio: *Libro*, *TLearningObject*, *ILearningObject*, *Articolo_di_rivista*, *Materiale_stand-alone* e *Flash*

Il secondo gruppo di elementi della T-box dell'ontologia di E-VALUE sono le relazioni. In generale le relazioni di un'ontologia sono relazioni binarie che collegano gli oggetti tra loro. Una vista generale delle relazioni (chiamate anche *ObjectProperties*) presenti nell'ontologia di E-VALUE, e fornita nella figura a fianco. Nel seguito ne diamo una descrizione più dettagliata, raggruppandole in due categorie. Nella prima troviamo le relazioni che servono a caratterizzare ciascun *learning object*. Sono relazioni che associano ad ognuno una serie di attributi che lo caratterizzano, come ad esempio: il tipo, il range di classi a cui è indirizzato, la lingua, il formato o il tipo.

haClasseDa e **haClasseA** sono due relazioni che associano ad ogni Learning Object il range di classi a cui esso è indirizzato. Se per esempio un *learning object* LO è indicato per la classe 1°, 2° e 3° elementare, la relazione *haClasseDa* associerà 1P (1° elementare) al LO, e la relazione *haClasseA* associerà 3P a LO.

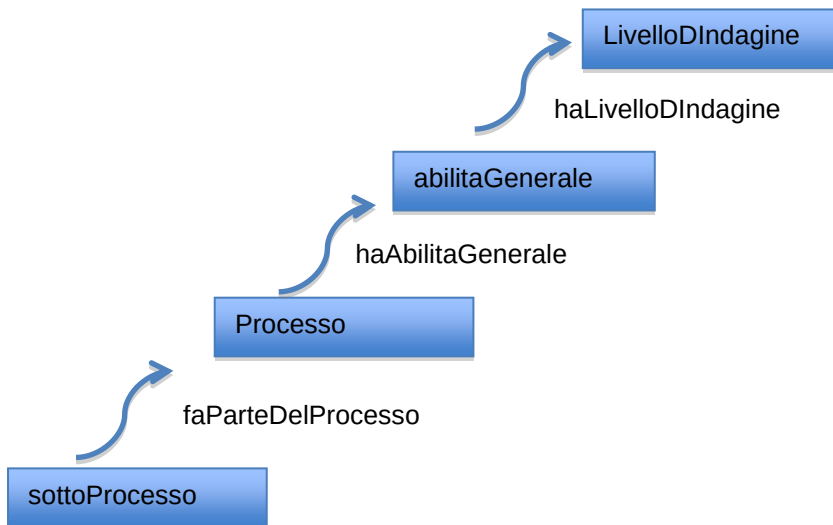
haFormato, **haLingua** e **haTipo** sono tre relazioni che associano ad ogni learning object il suo formato, la lingua in cui è scritto ed il suo tipo.

haProcessoOSottoprocesso è la relazione che associa ad ogni learning object i processi e/o i sotto-processi formativi sui quali si focalizza tale learning object. È da notare che un learning object potrebbe essere associato sia con uno o più processi che con uno o più sottoprocessi.

Il secondo gruppo di relazioni viene utilizzato per rappresentare la struttura gerarchica dei livelli di indagine. Ogni elemento più basso della gerarchia è



collegato al proprio padre con una relazione specifica secondo lo schema seguente



faParteDelProcesso è la relazione che associa ad ogni sottoprocesso il processo a cui appartiene. Ad esempio il sottoprocesso *Letto-scrittura-Decodifica-Lettere-Riconoscimento*, è associato al processo *Letto-scrittura-Decodifica-lettere*.

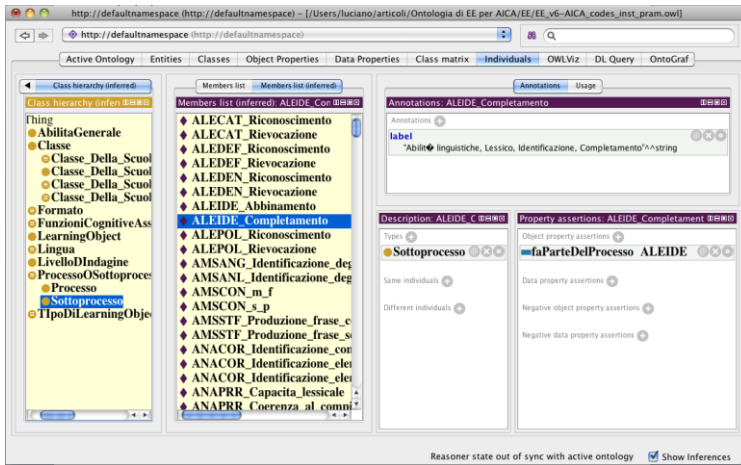
haAbilitaGenerale è la relazione che collega un processo all'abilità generale a cui si riferisce. Per esempio il processo *metafonologia-globale-rime*, è collegato tramite la relazione *haAbilitaGenerale* all'abilità generale *metafonologia-generale*; e così sarà anche per il processo *Metafonologia-globale-sillabe*; mentre il processo *metafonologia-profonda-fonema* sarà collegato da questa relazione all'abilità generale *metafonologia profonda*.

haLivelloDIndagine è la relazione che associa ad ogni abilità generale il livello di indagine a cui fa riferimento. Per esempio l'abilità generale *Letto-scrittura-codifica*, è associata tramite la relazione *haLivelloDIndagine* al livello di indagine *Letto-scrittura*.

Un'ultima relazione che non è riconducibile a nessuna dei due gruppi precedenti è la relazione **haClassePrecedente** che rappresenta l'ordine temporale fra le diverse classi. In particolare questa relazione collega ogni classe, ad esclusione della prima classe della scuola dell'infanzia, con la classe precedente.

La A-box (*assertional Box*) dell'ontologia contiene tutti gli oggetti con la loro caratterizzazione di tipo (cioè la loro appartenenza a una o più classi) e le specifiche relazioni tra gli oggetti. Nella figura seguente viene mostrato uno screenshot di *Protegé* che mostra nel riquadro centrale tutti gli oggetti della classe *sottoprocesso* (evidenziato in blu nella gerarchia delle classi, prima finestra a sinistra) e si focalizza sul processo *Abilità linguistiche, Lessico, Identificazione, Completamento* (evidenziato in blu nella seconda finestra),

mostrando come questo sotto processo sia collegato con il processo relazione *faParteDelProcesso* ALEIDE (*Abilità linguistiche, Lessico, Identificazione*) tramite la relazione *faParteDelProcesso* (ultima finestra a destra)



3.3 Esempi di *query* all'ontologia di E-VALUE

Seguono due esempi di *query* in linguaggio SPARQL (<http://www.w3.org/TR/rdf-sparql-query/>) utilizzato nella prima fase della progettazione del DSS (descritto nella sezione 3.2 come “selezionare tra tutti i *learning object* disponibili nel catalogo Erickson quelli rilevanti per i processi/sottoprocessi specificati nell'input”).

Esempio 1. Seleziona tutte le coppie di *processi amici* in cui il primo processo sia classificato come *Matematica-Geometria*

```
SELECT ?sp ?spa
WHERE {
  ?sp :haProcessoAmico ?spa.
  ?sp :faParteDelProcesso ?p.
  ?p :haAbilitaGenerale ?ag.
  ?ag :haLivelloDIndagine :Matematica-Geometria
}
```

Esempio 2. Seleziona tutti i *learning object* del tipo *Libro*, scritti in italiano e classificati MMASNL (*Metafonologia, Manipolazione di suoni non linguistici*)

```
SELECT ?lo
WHERE {
  ?lo :haProcessoOSottoprocesso ?sp.
  ?sp :faParteDelProcesso ?p.
  ?p :haAbilitaGenerale :MMASNL.
  ?lo :haLingua :ITA.
}
```

```
    ?lo :haTipo :Libro  
}
```

5. Studi correlati

È stata svolta un'analisi sulle forme di catalogazione di materiale digitale utilizzate con maggiore frequenza (ONIX, Dublin Core, gli standard catalografici di SCORM-LOM) dalla quale è emersa la necessità di non limitarsi a scegliere uno di questi sistemi ma di crearne uno che fosse sufficientemente elastico da consentire una catalogazione dettagliata secondo i punti di maggiore interesse per il progetto E-VALUE e con una struttura parzialmente a faccette.

6. Conclusioni

L'ontologia qui presentata intende descrivere e catalogare i materiali che le Edizioni Centro Studi Erickson utilizzano per il progetto E-VALUE, e può essere considerata come un punto di partenza per la creazione di un DSS a scopo formativo.

Un'osservazione importante è che il processo di creazione dell'ontologia non si limita alla traduzione automatica o semiautomatica della classificazione in linguaggio OWL. Si è verificato invece un complesso processo interattivo in cui il lavoro di classificazione è stato verificato e parzialmente rivisto per permettere la rappresentazione dello schema catalografico in linguaggio logico/formale.

Inoltre la fase di creazione dell'ontologia per il progetto E-VALUE si è rivelata anche una "fase di studio" che ha permesso alla parte tecnica del progetto (Fondazione Bruno Kessler) di approfondire la conoscenza in questo settore, e agli esperti del settore (Edizioni Centro Studi Erickson) di avvicinarsi alle logiche della modellazione formale.

7. Bibliografia

AID – Associazione Italiana Dislessia, Comitato Promotore Consensus Conference (a cura di) Disturbi Evolutivi Specifici di Apprendimento. Raccomandazioni per la pratica clinica di dislessia, disortografia, disgrafia e discalculia, Trento, Erickson, 2009.

Baader F. [et al.] (eds.), The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, Cambridge MT, 2003

Gnoli, C, Classificazione a faccette, AIB, Roma, 2004

Gnoli C., Marino V., Rosati L., Organizzare la conoscenza. Dalle biblioteche all'architettura dell'informazione per il Web, Tecniche Nuove, Milano, 2006

Gruber, T. R., A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220, 1993

"OWL 2 Web Ontology Language Document Overview". W3C. 2009-10-27, 2009.

Una metodologia per lo studio di fenomeni intangibili all'interno delle organizzazioni

Raffaele Dell'Aversana¹, Nicola Di Marcoberardino²

¹Banca d'Italia

raffaele.dellaversana@bancaditalia.it

²Bizbrain Group

nicola.dimarcoberardino@bizbrain.it

In questo articolo viene presentata una metodologia per la misurazione e rilevazione dei fenomeni intangibili all'interno delle organizzazioni, la cui conoscenza è indispensabile per coloro che devono prendere decisioni di governance in quanto elementi fondamentali nella costruzione del vantaggio competitivo. Tale metodologia è stata implementata nell'ambito di un progetto europeo ed è stata ampliata successivamente per supportare una grande varietà di casi ed esigenze.

1. Introduzione

La misurazione e rilevazione di quei valori che le moderne teorie economiche e sociali chiamano intangible assets (capitali intangibili) sono fondamentali nella costruzione del vantaggio competitivo ed assumono un significato ancora più pregnante nell'economia della conoscenza, nella quale la disponibilità delle informazioni e del sapere specialistico diventano leva di creazione del valore. Un quadro chiaro di questi valori e dei fenomeni che stanno alla loro base è un'esigenza imprescindibile per prendere decisioni di governance strategica basate su dati di fatto e non su sensazioni. Se consideriamo una realtà concreta come un ospedale, il valore è dato solo per una piccola parte dagli elementi tangibili, come immobili ed attrezzature, mentre la maggior parte è intangibile ed è dato da elementi come risorse umane, spirito di servizio, competenze, il saper fare, le procedure, la reputazione, la cortesia, la fiducia e immagine esterna. Analogamente, anche altre realtà come banche, associazioni, enti locali, camere di commercio e aziende hanno un valore che è costituito, per la maggior parte, da elementi intangibili. Si rende pertanto necessaria una metodologia adeguata per la misurazione, valutazione e gestione di quei fenomeni che stanno alla base degli assets intangibili.

In quest'ottica abbiamo creato un sistema integrato di sperimentazione e controllo, che consente di individuare, studiare e quindi governare **i fenomeni** che si manifestano all'interno delle organizzazioni, e in grado di supportare gli specialisti e gli sperimentatori in tutte le fasi, dalla costruzione del modello, alla rilevazione dei parametri, fino alla successiva elaborazione e seguente presentazione.

La nostra metodologia consente di individuare e quantificare i parametri fondamentali che governano il fenomeno, comporli in un modello riproducibile, controllabile e validato su base statistica, e studiarne le correlazioni di base per sperimentare ed investigare le implicazioni causa-effetto alla base dei cambiamenti che avvengono all'evolversi del fenomeno stesso.

La tecnologia utilizzata, basata su piattaforma web nativa, permette di mantenere in un'unica suite integrata tutte le funzioni/fasi fondamentali alla misurazione: costruzione del framework teorico, raccolta dei dati di input (sondaggi web, focus group, interviste telefoniche o in presenza, data mining su altri sistemi informativi, crawling web, ecc.), elaborazione delle informazioni secondo il framework teorico scelto, pubblicazione dei risultati e messa a disposizione on-demand. Tale approccio permette di abbattere notevolmente i costi di misurazioni di questo tipo, fungendo allo stesso tempo da repository del know-how scientifico e degli strumenti pratici di gestione del fenomeno in esame. Il sistema permette quindi, in modo semplice, la creazione di **osservatori permanenti** pilotati da **specialisti** o **sperimentatori**, per investigare i più svariati ambiti dei fenomeni economici e sociali con forte componente intangibile dove, cioè, sono significative variabili quali: la soddisfazione, le relazioni, l'immagine, le competenze, la capacità di innovazione, la reputazione, il fattore umano, le conoscenze.

2. Architettura

L'idea alla base è semplice:

- Individuare i fenomeni che hanno un impatto strategico sull'organizzazione o che definiscano un ambito di studio
- impostare un sistema di variabili che descriva questi fenomeni in accordo con un framework teorico
- quantificare queste variabili nel tempo attraverso strumenti di rilevazione molteplici, sistematici e strutturati portando alla luce i punti critici e le aree di intervento prioritario
- definire azioni di miglioramento (interventi) in grado di influenzare i fenomeni in esame
- valutare i risultati per investigare se le azioni di miglioramento introdotte nel sistema in studio abbiano avuto l'effetto sperato
- rilevare reiteratamente e procedere al miglioramento sistematico

L'approccio è necessariamente dinamico: per creare "fotografie" della realtà del fenomeno in studio occorre creare un modello che possa essere interrogato in momenti differenti, ottenendo le "istantanee" di cui si ha bisogno e componendo poi queste istantanee in sequenza per cogliere gli andamenti dinamici delle variabili e i trend.

2.1. Il sistema di valutazione

Individuati i fenomeni intangibili strategici per un'organizzazione (ottenibili da un'analisi strategica, ad esempio utilizzando il processo delle Mappe strategiche [Kaplan e Norton, 2005]) si passa all'identificazione delle **variabili** significative del singolo fenomeno che, a loro volta composte e correlate a sistema, creino un modello sufficientemente attendibile del fenomeno in esame. A tale modello daremo il nome di **sistema di valutazione** (vedi Fig. 1), vero cuore pulsante dell'approccio. È nel sistema di valutazione infatti che gli **esperti, in accordo con framework teorici consolidati o sperimentali**, possono strutturare un modello coerente della realtà del fenomeno, andando a definire tre livelli di variabili concatenate: **gli stati** (variabili macroscopiche del fenomeno), **gli indici** (variabili intermedie che specializzano gli stati) e **gli indicatori** (variabili puntuali, che specializzano gli indici e che devono poter essere misurati con strumenti di rilevazione).

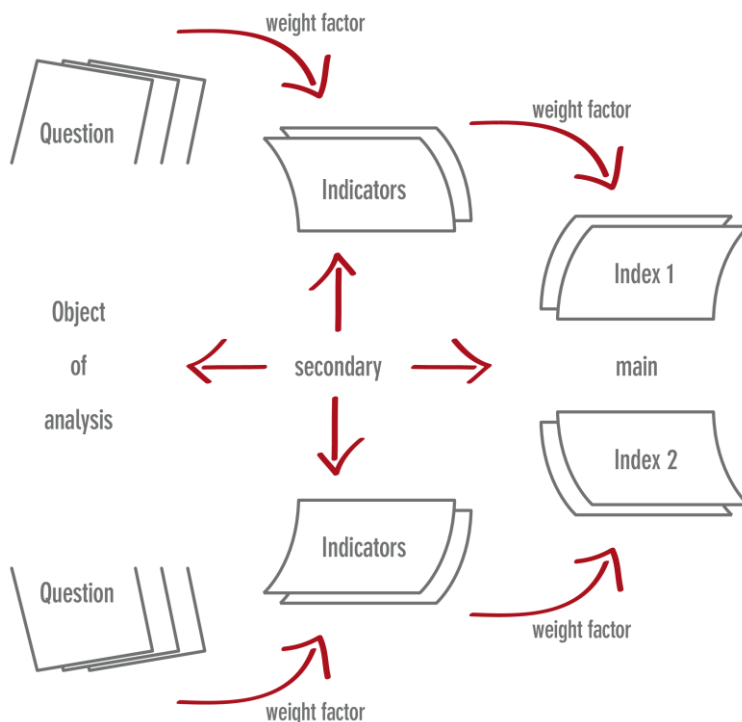


Fig. 1: schema del sistema di valutazione

I valori che **gli indicatori assumono nel tempo** vengono alimentati attraverso l'uso di molteplici strumenti di rilevazione (questionari cartacei,

questionari web, interviste, focus group, data mining su altri sistemi, ecc.), gestiti automaticamente e che permettono allo specialista di scegliere, nell'ambito del sistema di valutazione, il o i metodi che egli ritiene più adatti per il calcolo delle variabili del sistema. I livelli più alti delle variabili (gli indici e gli stati) vengono invece ottenuti per riaggregazione "pesata" dei livelli inferiori utilizzando apposite logiche implementate a seconda del fenomeno in esame.

Parallelamente si procede alla specifica dell'**universo di studio** popolando le cosiddette **anagrafiche oggetto**. Le anagrafiche oggetto rappresentano gli "oggetti" che influenzano e che sono costitutivi dei sistemi alla base dei fenomeni in studio. Se stiamo operando un'analisi di clima organizzativo, ad esempio saranno i dipendenti dell'organizzazione, così come se stiamo invece analizzando i fenomeni alla base di una rete vendita, saranno i clienti, i venditori e la direzione commerciale. Allo stesso modo, nel caso di rilevazioni di sostenibilità, dovranno invece essere presi in considerazione tutti gli stakeholders aziendali (clienti, dipendenti, associazioni di categoria, ONG, ecc), operando un vero e proprio enterprise feedback a 360°. Tali anagrafiche rappresentano allo stesso tempo "coloro che popolano e danno vita al fenomeno" e "coloro che hanno da darci le informazioni necessarie alla rilevazione dei valori delle variabili ad un dato istante". Si pensi all'esempio dell'analisi di clima aziendale, dove i dipendenti di un'organizzazione sono allo stesso tempo i soggetti del fenomeno e coloro che devono essere intervistati per rilevarlo.

Ad ognuna delle possibili tipologie di anagrafiche oggetto (anche radicalmente diverse tra loro) vengono associati due livelli di informazione: un primo set che ne definisce il nome, la tipologia ed eventuali informazioni di contatto (si pensi ad esempio all'anagrafica dei dipendenti di un'organizzazione cui andranno poi somministrati i sondaggi), ed un secondo set che individua le **variabili di stratificazione**, quelle variabili cioè che servono a definirne le particolari specifiche (ad esempio per una persona età, sesso, posizione lavorativa, istruzione, ecc.).

Un modulo di estrazione dei dati permette infine di selezionare un **campione statisticamente significativo** dall'universo individuato. L'uso di un campionatore statistico ci sembra fondamentale in una situazione di regime in cui un alto numero di feedback deve essere acquisito per tenere continuamente aggiornata la situazione delle variabili. Grazie al campionatore è possibile interessare di volta in volta un piccolo numero di "anagrafiche oggetto", evitando di importunarle con contatti troppo frequenti.

Uno schema del sistema di valutazione è mostrato in Fig. 2.

2.2. L'azione di miglioramento e la misurazione del grado di efficacia.

Impostato il modello che permette di seguire dinamicamente l'evolversi del fenomeno, si passa ad introdurre le azioni di miglioramento: azioni che vengono messe in campo per far evolvere il sistema secondo una direzione voluta.

In generale, nelle scienze economiche e sociali, un fenomeno è rappresentato da un sistema aperto, tale per cui esso non è isolabile in laboratorio e si trova esposto all'effetto di tutta una serie di "perturbazioni" provenienti dall'ambiente che lo circonda (politica, moda, modi di pensare, teorie economiche, aspettative sociali, ecc.) e che ne determinano o ne influenzano radicalmente l'evolversi. Se l'obiettivo è quindi quello di studiare che effetto hanno le azioni messe in atto volutamente sul fenomeno (le azioni di miglioramento, appunto) da parte degli attori che cercano di governarlo (datori di lavoro, fornitori, enti locali, istituzioni pubbliche), dobbiamo, applicando una metodologia nota nelle scienze mediche, necessariamente estrarre due campioni statisticamente rilevanti dall'universo in studio: sul primo, che chiameremo gruppo di sperimentazione, agiranno sia le "perturbazioni dell'ambiente esterno" che l'azione di miglioramento prescelta, sul secondo, che chiameremo gruppo di controllo, agiranno invece le sole perturbazioni dell'ambiente esterno.

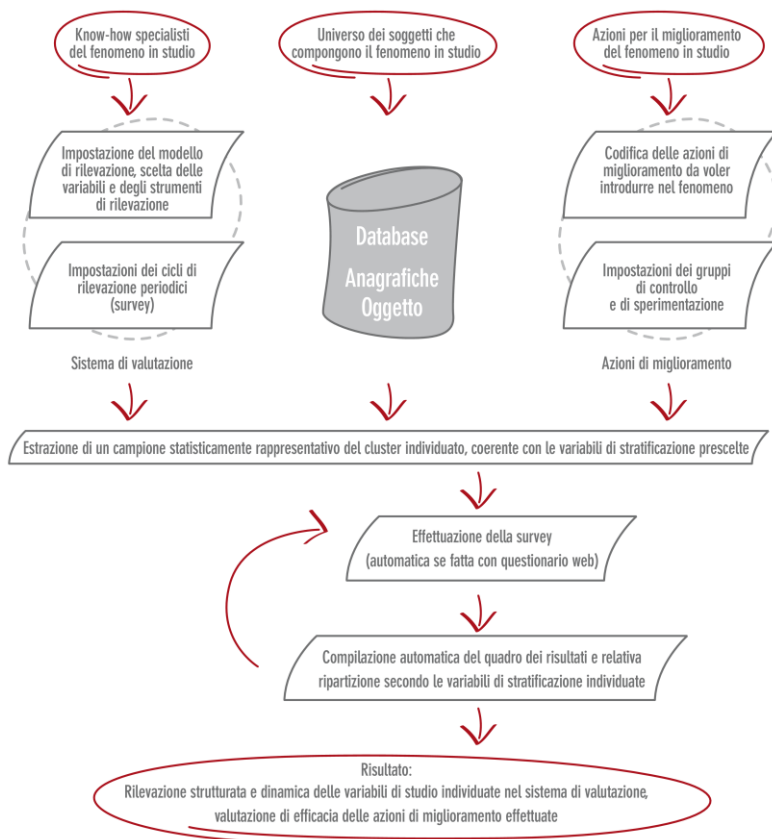


Fig. 2: schema dell'architettura

Effettuando, al termine del periodo di azione, un ciclo di valutazione sui due gruppi, saremo in grado di apprezzare come il sistema è evoluto nel suo complesso e quali siano le differenze prodotte dalla nostra azione di miglioramento [Motterlini e Guala, 2011].

Le indicazioni di nessi di causalità tra “variabili del sistema” ed azioni di miglioramento messe in atto nel ciclo di sperimentazione forniranno una misura del “livello di bontà” di una determinata azione, permettendo all’organizzazione di scartare azioni “poco efficaci” per concentrarsi su quelle che invece ottengono i migliori risultati.

2.3. Analisi multi-universo e multi-sistema di valutazione.

Per poter supportare direttamente le analisi multi-universo e multi-sistema di valutazione, sono stati creati due distinti **livelli di funzionamento**:

- **livello di rilevazione strategica**: un opportuno “centro d’intelligenza”, che chiameremo **osservatorio**, è in grado di seguire i risultati della valutazione a livello corporate centralizzato, aggiornare i pesi e i fattori di correlazione tra i vari parametri, modificare i sistemi di valutazione fungendo da “osservatorio di rilevazione permanente”
- **livello di rilevazione operativa**: ogni unità operativa ha a disposizione una sezione relativa al proprio universo in grado di monitorare i fattori critici nel tempo e rilevare tutti i dati necessari all’elaborazione dei risultati

Combinando l’uso dei due livelli differenti, è possibile portare avanti contemporaneamente rilevazioni su più universi o analisi sullo stesso universo, ma utilizzando più sistemi di valutazione. Questo permette di capire quale sistema di valutazione produce un modello più attendibile e dunque più adatto alla descrizione del fenomeno, permettendo, in sostanza, al sistema di evolversi ed auto apprendere.

2.4. Ambiti di studio

Sono state condotte analisi su una vasta gamma di fenomeni, utilizzando l’approccio proposto. Per fare qualche esempio possiamo ricordare:

- **vendite**: il focus principale è puntato sulla rilevazione e quantificazione delle relazioni intangibili che esistono tra azienda, rete vendita e cliente. [Barnes et al, 2009] Attraverso l’identificazione e misurazione dei parametri che strutturano le relazioni di questo triangolo è possibile tenere sotto controllo l’efficacia della propria azione commerciale, perché, a volte, anche un aumento di fatturato potrebbe essere realizzato a scapito della relazione col cliente o col venditore compromettendo quindi la redditività futura.

- **lavoro:** in quest'ambito i soggetti dell'universo di studio sono i dipendenti e i lavoratori delle organizzazioni, i cosiddetti "clienti interni". Il successo dell'azione di ogni organizzazione dipende oggi più che mai dal contributo delle persone che vi lavorano. Rilevare i parametri che ne regolano soddisfazione, motivazione e potenziale, andando ad investigare le dynamic capabilities [Zollo, 2002] è compito irrinunciabile di ogni ufficio del personale che voglia stare al passo coi tempi e creare vantaggio competitivo per la propria organizzazione.
- **Soddisfazione del cliente:** l'obiettivo principale è investigare cosa il consumatore finale pensa del nostro prodotto/servizio e della nostra organizzazione. Come ci percepisce? Quali sono le qualità della nostra azione che predilige? Quali i nostri difetti? Come possiamo instaurare con lui una relazione fidelizzata?
- **Stakeholders satisfaction:** lo scopo è quello di permettere alle organizzazioni di rilevare la soddisfazione e le necessità dei propri stakeholders, per aumentare gli standard di servizio allo scopo di costruire modalità di fruizione che siano effettivamente efficaci ed efficienti, poter meglio guidare la propria azione strategica e poter quindi operare in maniera concreta ed efficace sulla business health. [Keller e Price, 2010]
- **Sostenibilità:** analisi miranti alla valutazione delle politiche di sostenibilità intraprese dalle organizzazioni, in ambito economico, sociale ed ambientale.
- **Accettazione sociale:** studi miranti alla valutazione dell'accettazione sociale di investimenti pubblici o privati a forte impatto sociale o ambientale.

3. Implementazione

Queste idee sono state implementate in un prodotto commerciale, denominato Bizbrain Life, basato su tecnologia Java Enterprise Edition. Il modello dei dati è di tipo relazionale e l'intera applicazione usa tecnologie web, in modo da poter essere utilizzata tramite i comuni browser. Dal lato server è stato fatto largo uso di tecnologie opensource e platform-independent.

L'applicazione inoltre è stata progettata per essere scalabile, in modo da poter essere distribuita su un insieme di server e soddisfare un gran numero d'accessi contemporanei, garantendo la sicurezza dei dati.

Un appropriato sistema di autorizzazioni consente di stabilire i permessi di ciascun utilizzatore, in modo che i dati, i risultati, gli strumenti di progettazione e i report siano visibili solo agli utenti autorizzati.

Il sistema include uno strumento di reportistica completo, utile per realizzare sia report standard che i cosiddetti “cruscotti direzionali”, in grado di mostrare i risultati raggiunti in modo sintetico e chiaro.

4. Un caso di studio: il progetto OWL

OWL (Improvement of personal life quality and improvement of working life quality of older workers, <http://www.owlproject.net>) è un Progetto Europeo transnazionale, finanziato dal Fondo Sociale Europeo, ESF Article 6 “Innovative approaches to the management of change”, realizzato in collaborazione con partner pubblici e privati internazionali; è stato il primo banco di prova operativo della metodologia descritta in questo articolo, implementata in un software denominato OWLmS e ampliata successivamente in Bizbrain Life. Il progetto OWL, che ha come scopo il miglioramento della qualità della vita dei lavoratori over cinquantacinque, ha permesso di confermare la validità della metodologia in un contesto di grandi aziende ed enti internazionali.

OWLmS ha consentito di gestire le attività del progetto e di fornire uno strumento di utilità per tutte quelle organizzazioni che hanno partecipato alla sperimentazione, in base al principio per cui il miglioramento della vita personale e lavorativa dei lavoratori, oltre che rispondere a un'esigenza di ordine morale, passando attraverso una sensibilizzazione del management sulle problematiche principali, sia in grado di creare un circolo virtuoso: sto meglio, lavoro meglio, produco meglio e di più. Per raggiungere questo scopo, OWLmS, basandosi sulla metodologia esposta, è stato realizzato in modo da essere effettivamente fruibile e utile al management aziendale, permettendo di studiare le correlazioni di base per sperimentare e investigare le implicazioni causa-effetto alla base dei cambiamenti aziendali e della qualità di vita e di lavoro dei lavoratori senior.

Per realizzare le “fotografie” del sistema azienda/lavoratore abbiamo dovuto creare un modello (descritto in questo articolo) che potesse essere interrogato in momenti differenti, per ottenere le “istantanee della realtà” di cui avevamo bisogno. Siamo dunque partiti col definire quali sono i due soggetti principali che compongono il nostro sistema di studio: l'azienda da un lato, con la propria complessità organizzativa, economica, i propri target ed obiettivi, e il lavoratore dall'altro, con la propria personalità, il proprio potenziale, la propria vita.

Un gruppo di esperti, (dall'economia aziendale alla psicologia del lavoro, dalla medicina alla statistica) si è dunque cimentato durante il periodo del progetto, nella creazione di modelli quantificabili (variabili) che fossero in grado di descrivere il fenomeno in studio.

La situazione del lavoratore è stata descritta da una serie di stati che ne definiscono la condizione di vita e di lavoro (condizione fisica e fisiologica, condizione psicologica, performance potenziale ed effettiva, ecc.).

Anche l'azienda a sua volta è stata modellata sulla base di stati specifici (economico/reddituale, articolazione organizzativa, performance, ecc.).

Dopo aver definito tali stati, li abbiamo suddivisi: ad ogni stato principale abbiamo associato un indice, variabile quantitativa definita su un intervallo

percentuale (0%- valore minimo – 100% - valore massimo). Ogni indice è stato a sua volta definito da parametri secondari, espressi quantitativamente dagli indicatori (0%- valore minimo – 100% - valore massimo). Gli indicatori sono investigati per mezzo di domande a risposta chiusa, raccolte in appositi questionari e somministrati ad intervalli definiti ai vari attori del sistema (management, line management, general worker, old worker), tramite focus group e attraverso una serie di dati quantitativi acquisiti direttamente dal sistema.

La riunione di queste **variabili** in un modello coerente, unitario e quantificabile ha dato vita al **sistema di valutazione**.

Tramite l'interpretazione dei punteggi che le variabili assumono ad ogni interrogazione (ciclo) è stato possibile investigare le aree critiche di intervento e le aree di eccellenza, costruire azioni di miglioramento adeguate, pianificare obiettivi.

Ad ogni ciclo, un grafico a radar per ogni stato e indice ha permesso di avere una **lettura visiva** della situazione, tanto aziendale che del lavoratore; grafici ad istogramma ci hanno invece permesso una lettura dei risultati per variabili di stratificazione, ad esempio per classi d'età, per vedere come gli anni d'anzianità influenzano un determinato parametro.

Informazioni statistiche sui dati a livello di indicatori (media, varianza, correlazione tra i dati, ecc.) hanno permesso di condurre analisi statistiche specifiche da parte degli specialisti aziendali e del progetto.

Indicazione di nessi di causalità tra “variabili del sistema” e azioni di miglioramento messe in atto nel periodo di riferimento (ciclo di sperimentazione) basate su test statistici, ci hanno dato una misura del “livello di bontà” di una determinata azione, permettendo al management di scartare azioni poco efficaci per concentrarsi su quelle che ottengono migliori risultati.

Tali variabili sono state alimentate attraverso l'uso degli strumenti di rilevazione citati in precedenza, gestiti automaticamente da OWLmS, che hanno permesso ai vari specialisti di scegliere nel sistema di valutazione il metodo che hanno ritenuto più adatto per il calcolo delle variabili del sistema.

L'uso di tecnologia web nella realizzazione del OWLmS ha permesso, inoltre, di mantenere in un'unica suite integrata anche le funzioni di somministrazione dei questionari su internet e la manutenzione degli altri strumenti di rilevazione, minimizzando i costi nella rilevazione dei dati.

Una logica di gestione a cicli ha permesso di ripetere le rilevazioni in maniera autonoma ad intervalli dipendenti dalla problematica di studio in esame, rendendo il sistema dinamico nel tempo.

Il progetto è arrivato a due importanti risultati: l'individuazione congiunta di metodologie, modelli e processi per risolvere problemi complessi, e la sperimentazione congiunta di metodi alternativi per gestire “l'età lavorativa” attraverso metodi alternativi di organizzare il lavoro ed evidenziare i punti di forza dei lavoratori più anziani.

Nella prima sperimentazione abbiamo testato l'approccio su dieci organizzazioni tra pubbliche e private, sottoponendo ad analisi 811 lavoratori, campione rappresentativo di oltre 8000.

5. Conclusioni e sviluppi futuri

Come messo in evidenza dagli esempi che precedono, l'approccio è in grado di focalizzarsi sulle problematiche aziendali ed organizzative più disparate. Un apposito sistema di valutazione permette infatti in maniera semplice ed immediata di tenere sotto controllo fenomeni anche molto differenti, permettendo allo stesso tempo di considerarli e studiarli in un quadro unitario.

In effetti è proprio lo studio combinato di fenomeni differenti che agiscono sulla stessa organizzazione, tenuti sotto controllo per tempi sufficientemente lunghi da studiarne le dinamiche, che permetterà, in prospettiva, di fare maggior luce su quei fenomeni intangibili, che oggi rappresentano il vero vantaggio competitivo e di distinzione delle aziende.

Bibliografia

[Barnes et al, 2009] Cindy Barnes, Helen Blake and David Pinder, Creating and delivering your value proposition : managing customer experience for profit, Great Britain and the United States in 2009 by Kogan Page Limited, 2009

[Keller e Price, 2010] Scott Keller and Colin Price, Beyond Performance: How Great Organizations Build Ultimate Competitive Advantage, John Wiley & Sons, 2010

[Kaplan e Norton, 2005] Kaplan Robert S.; Norton David P., Mappe strategiche. Come convertire i beni immateriali in risultati tangibili, ISEDI, 2005

[Motterlini e Guala, 2011] Motterlini Matteo; Guala Francesco, Mente mercati decisioni. Introduzione all'economia cognitiva e sperimentale, Università Bocconi Editore, 2011

[Zollo, 2002] M. Zollo (with S. Winter), Deliberate learning and the evolution of dynamic capabilities. Organization Science, 13 (3), 2002

Note

Il contenuto di questo articolo riflette le idee degli autori e le opinioni espresse non impegnano in alcun modo la Banca d'Italia o Bizbrain Group.

Using Geographic Information Systems to Increment the Knowledge of Cultural Landscapes

Amelia Carolina Sparavigna, Roberto Marazzato¹

Dipartimento di Fisica, Politecnico di Torino

Corso Duca degli Abruzzi 24, 10129, Torino, Italy

amelia.sparavigna@polito.it

¹*Dipartimento di Automatica e Informatica, Politecnico di Torino*

Corso Duca degli Abruzzi 24, 10129, Torino, Italy

roberto.marazzato@polito.it

To increment the knowledge of a cultural landscape, the use of Information and Communication Technologies is becoming fundamental. By means of these technologies we can improve the study of areas having ancient or modern cultural evidences, by superimposing data from some territorial information systems on satellite surveys. In particular, we will analyse Geographic Information Systems (GIS), freely available on the World Wide Web. A well-known example is Google Earth, which is offering the possibility of virtual tours to the users of the Web. Moreover, besides being of support for archaeological and other research studies, GIS can be helpful in creating some systems for the management of archaeological and cultural patrimony and for incrementing the revenues from tourism.

1. Introduction

The use of the Information and Communication Technologies is crucial in incrementing our knowledge of cultural landscapes. According to the World Heritage Committee, UNESCO, a cultural landscape is a geographical area that is uniquely representing the combined work of nature and of man [UNESCO, 2005]. In fact, the World Heritage Committee is using specific categories to design and describe how a landscape can be modified or intentionally created by man. Let us consider in this paper only the general point of view, where the cultural landscape is a geographical area combined with archaeological, historical or other social evidences.

The use of new technologies allows capillary recognitions of all the features of a certain territory. For instance, satellite maps provide high-resolution multispectral images, often associated with the land surveys of topographic maps. Superimposed to these images, we can have road maps and the corresponding labels, such as in the Google Maps. Mainly based on user contribution, pictures and other information can be added.

Territorial informative systems, which have a registering system of all local evidences based on the geographical coordinates, can become a good opportunity: it can be oriented to the increase of knowledge, management survey and potential development of our patrimony. In this paper, we will propose in particular the use of Geographic Information Systems (GIS), which are freely available to the users of the World Wide Web. Our discussion on GIS will be oriented to the knowledge of archaeological sites and landscapes.

It is commonly assumed that the study of sites of our past is the task of archaeology. This is partially true, because this study requires an increasingly multidisciplinary approach. For instance, physics and chemistry are necessary for dating, analysis and for the development of restoration techniques. Moreover, geology, photogrammetry, remote sensing and information technology are necessary for the correct identification and for the subsequent management of sites. All of them help in creating the knowledge of cultural landscapes. In fact, logistics can be considered as a good partner too, in the case that we want the optimization of sites preservation and some profits from tourism. Preservation of ancient remains is an important and challenging issue for our society, because several archaeological sites need protection from urbanization or agricultural and mining exploitations. A certain profit for local population could be helpful in the management - often the rescue - of archaeological remains.

2. Remote sensing augmented by GIS

The use of remote sensing in archaeology is based on images obtained by means of aerial or satellite surveys. Depending on the sensors which constitute the survey equipment, we can obtain different information on the surface and on the uppermost layers of the ground. A new opportunity is given by the freely available satellite maps imageries that we find on the World Wide Web. For several areas, we have a high enough resolution: this fact makes the maps suitable for archaeological investigation. Of course, satellite images can be further enhanced and analysed by means of image processing methods. As discussed in [Lasaponara e Masini, 2008], it is the new developments of remote sensing techniques and facilities of the Information and Communication Technologies (ICTs) that produced their strong application in cultural studies. The reference lists the possible remote sensing techniques as: aerial multispectral and hyper-spectral imagery, active airborne data of the Light Detection and Ranging (LIDAR) and Synthetic Aperture Radar (SAR), the detection and spatial reconstruction of sub-surface remains by using Ground-Penetrating Radar (GPR), and magnetic and electrical tomography. Nowadays, the remote sensing is often integrated by means of ICTs with image processing, 3D visualization and virtual reconstruction of landscape and sites in the Geographic Information Systems (GIS). GIS is a system developed to collect and manipulate all types of geographically referenced data [Bolstad, 2005, Chang, 2006]. Besides being a system that stores and displays geographic information, a GIS application allows users to create interactive queries and edit data [Chang, 2006].

For what concerns the aerial photos, their use for archaeological survey is well known [Georgoula et al., 2004]. A large quantity of photos is generally merged to obtain, after being processed with photogrammetric methods, the desired outputs in 2D or 3D models.

Let us note that one of the most important applications of the aerial survey is the detection of buried archaeological remains. Under proper conditions, these remains are best identified when viewed from above. In Ref.4 for instance, the researchers evaluated the possibilities that the high resolution satellite imagery offers in the identification of buried remains. The researchers used the satellite images of Quickbird 2, with a spatial resolution of 0.70 m, covering an extended area of the site of Philippoi in Greece. Applying a statistical evaluation on the fusion of panchromatic and multispectral images, the authors found that the spatial ground resolution of satellite images allows the identification of a large range of archaeological features. Quickbird 2 is the USAF designation of QuickBird, a high-resolution commercial constellation of satellites, owned by DigitalGlobe. The satellites collect panchromatic imagery at a resolution of 0.7 m and multispectral imagery at a lower resolution, approximately 2.5 m [WWW1]. The imagery can be used in GIS packages for analysis or for mapping applications, such as Google Earth.

Another commercial satellite imagery company is GeoEye, based in Virginia [WWW2]. It was the company which launched the first sub-meter commercial IKONOS satellite. It is an Earth observation satellite, offering multispectral and panchromatic imagery. As soon as the IKONOS imagery became available, it was integrated in GIS applications, such as in an archaeological survey of Tuscany [Campana, 2002]. Garrison et al., 2008, performed a detailed evaluation of the use of IKONOS imagery. In their paper, the authors presented an extensive study on the use of IKONOS in detecting some sub-canopy Maya settlements.

Images from DigitalGlobe, GeoEye and others constitute the imagery of Google Earth. This Web service is an example of a popular GIS. Before a detailed discussion on GIS and Google Earth applications, let us consider an example, taken from the site of Google Maps, on how information based on geographic coordinates can appear. Choosing with the Google search engine the archaeological site of Stonehenge for instance, the service provides the satellite image labels and icons of Wikipedia (Fig.1, upper part). To the geographical coordinates we have then associated information. Moreover, dragging the icon of Street View on the image, several dots and lines appear in blue on it (Fig.1, lower part); each dot corresponds to a picture uploaded by the users and the line to a panoramic view. We can check by means of pictures and movies how the landscape is seen from the ground and compare it with the aerial view. It is then quite clear that a GIS service is fundamental to have a precise location of the features of a cultural landscape.

3. Google Earth features

Google Earth [WWW3] implements a 3D graphical browser interface to a huge database of images of the earth's surface, obtained from different imagery

sources, such as DigitalGlobe, EarthSat, GeoEye, IKONOS, and aerial surveys. The spatial resolution of most images is 15m/px, with better defined (2.5m) spots, mostly from CNES and SPOT Image sources. CNES is the Centre National d'Études Spatiales, which is the French government space agency.

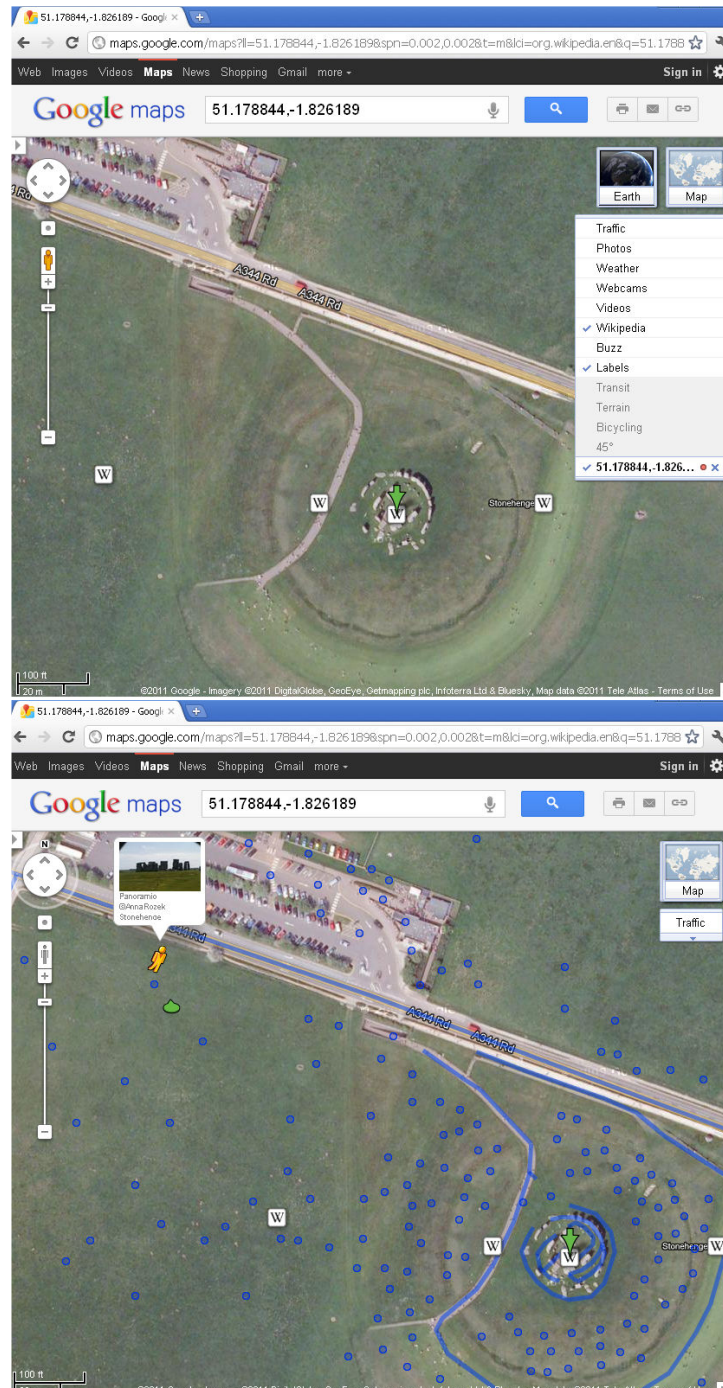


Fig.1 - Google Earth is an example of a popular GIS. For instance, let us choose with the Google search engine the archaeological site of Stonehenge. The service provides the satellite image labels and icons of Wikipedia. Moreover, dragging the icon of Street View on the image, several dots and lines appear in blue on it, each dot corresponds to a picture uploaded by the users and the line to a panoramic view.

Spot Image is a public limited company created in 1982 by CNES and other societies. The company is the commercial operator for the SPOT Earth observation satellites, providing images with different resolutions. Spot Image also distributes from other optical satellites, in particular from Formosat-2 (Taiwan) and Kompsat-2 (South Korea) and from radar satellites (TerraSar-X, ERS, Envisat, Radarsat).

The browser interface of Google Earth allows users to navigate the whole Earth's surface through a 3D pointing system, involving the rendering of the viewer's height (see for instance Fig.2). All visited places can be recorded and then shared by saving them to "placemarks" files. A wide range of superimposed layers can be switched on and off, and a photogrammetry tool is also provided.

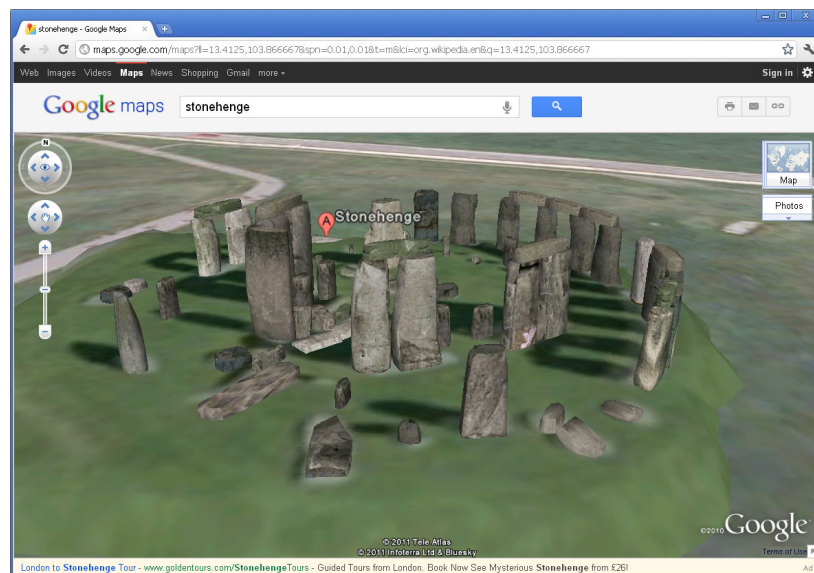


Fig.2 – Stonehenge in 3D as seen by Google Earth.

In the case that the user has GIS data, it is possible to import it into Google Earth Pro, the paid version of Google Earth, to be used as part of the corresponding Google Earth visualizations. In the case that the user has only the free version of Google Earth, there are some tools for converting GIS data to KML. We have for instance, a GIS software, like ESRI ArcGIS and MapInfo, possessing tools to export GIS data into KML format for use in Google Earth. Keyhole Markup Language (KML) is the notation of a mark-up language (XML) for expressing geographic annotation and visualization within Internet-based, two-dimensional maps and three-dimensional Earth browsers. KML was developed for use with Google Earth, which was originally named Keyhole Earth Viewer. It was created by Keyhole, Inc, acquired by Google in 2004. In fact, KML is an international standard of the Open Geospatial Consortium, an international voluntary consensus standards organization, originated in 1994. In it, more than 400 commercial, governmental, non-profit and research organizations worldwide collaborate to encourage the development and implementation of open standards for geospatial content and data sharing.

The fact that a GIS format can be easily converted to KML, through a variety of applications and plugins, widely spread the use of Google Earth among researchers in order to describe and place different sites of archaeological interest [Bousman, 2006, Ur, 2006, WWW4, Conroy et al., 2008, Thomas et al. 2008, WWW5, WWW6, WWW7].

Kris Hirst provides a good guide on Google Earth. As reported in the paper [Kris Hirst, 2011], Google Earth and Google Maps are used in the “searching for crop marks on aerial photos”, which is “a time-tested way to identify possible archaeological sites”. Crop marks are areas where sub-surface archaeological, natural and recent features may be visible from the air. Crop marks arise because vegetation has a different growth. A buried stonewall for instance is negatively affecting the growth above it, whereas, a buried ditch, usually containing more organic matter, provides more nourishment to the plants. Along with soil marks, they can reveal buried archaeological sites not visible from the ground. Soil marks are differences in soil colour that can be the result of archaeological remains. They can be easily seen when fields are ploughed (see Fig.3).

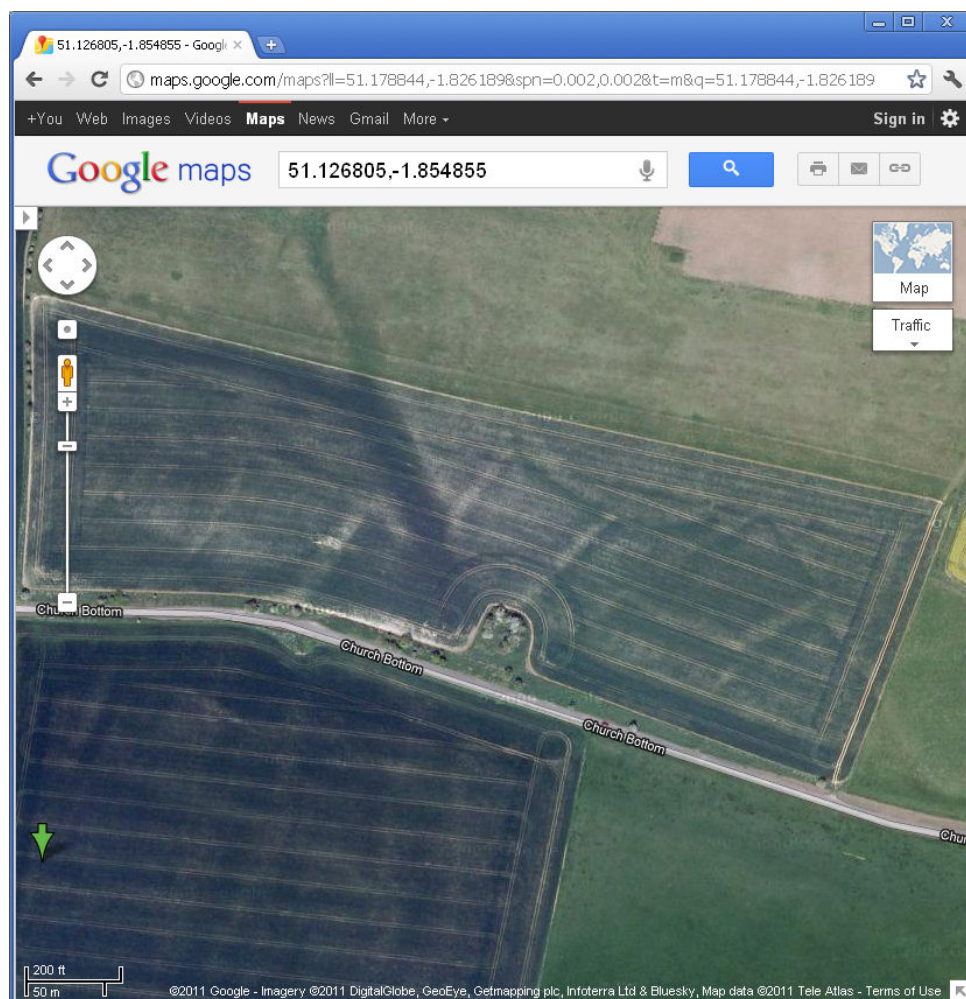


Fig.3 – An example of soil marks near Stonehenge.

Higher resolution satellite imagery is then fruitful for identification of archaeological remains. There are several users posting on the Web their results. Let us just remember two of them, directly observed by one of the authors: the geoglyphs of Titicaca [Sparavigna, 2010], which are earthworks near the Titicaca Lake, Peru, and some stone structures, desert kites and stone circles, in the Syrian Desert [Sparavigna, 2011]. Examples of the archaeological structures in Peru and in the Syrian Desert are provided by Fig.4 and Fig.5, respectively.

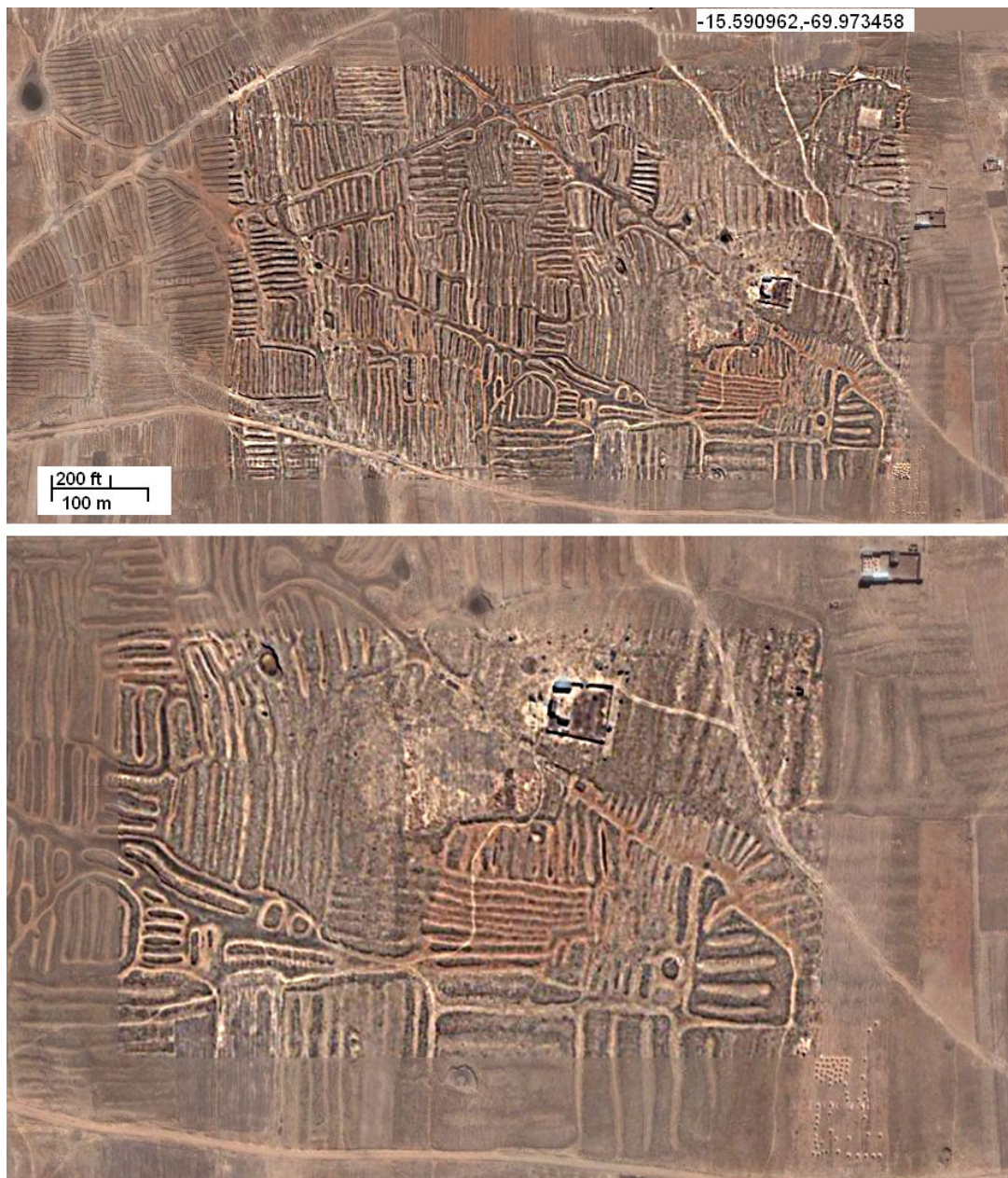


Fig.4 – A geoglyphs near the Titicaca Lake, created by the earthworks of an ancient agricultural system, that of the “raised fields” [Sparavigna, 2010].

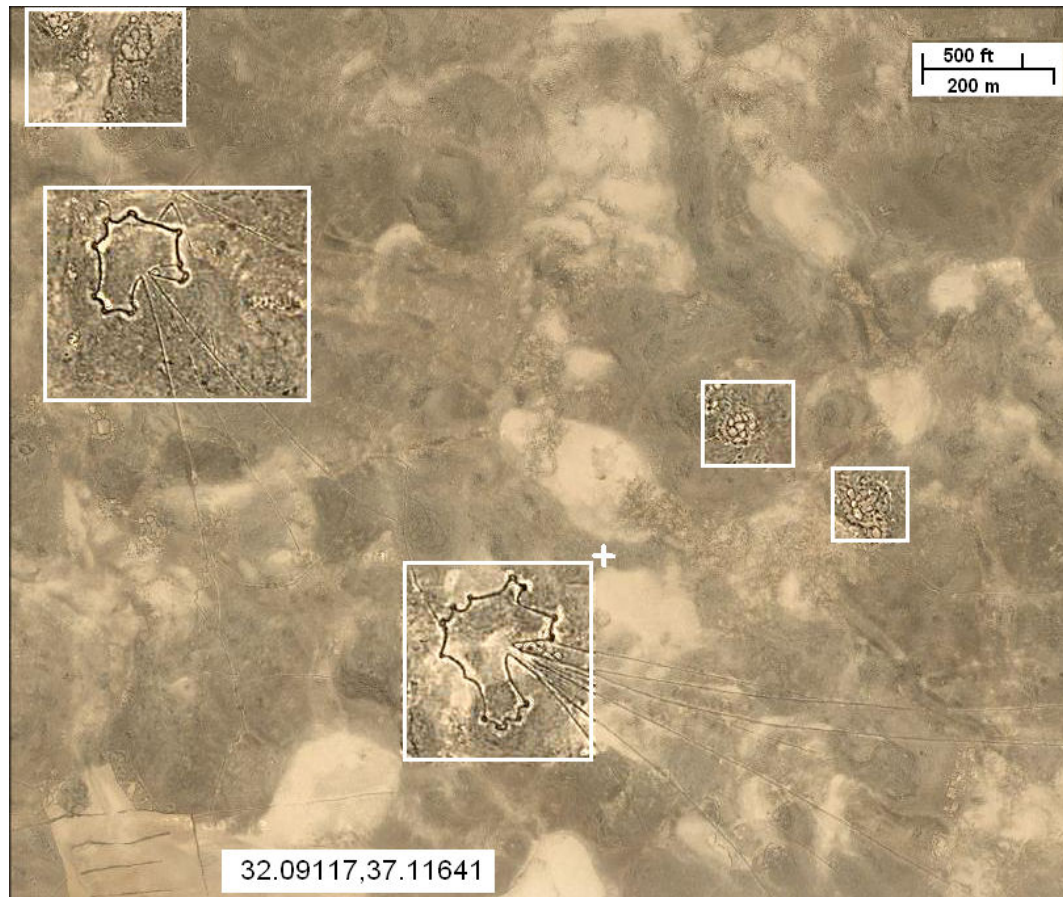


Fig.5 – Desert “kites” and stone circles in the Syrian Desert [Sparavigna, 2011].

4. Conclusions

The high accuracy in placement of the marks and the worldwide dissemination of such data can lead to a possible threat for archaeological heritage. Some concerns had been aroused that pothunters, tomb raiders and looters can use such data to spoil unprotected or unexcavated sites [Conroy et al., 2008]. A good practice in this case is to freely and accurately describe and reference the well known sites with educational and security facilities, leaving new and unexcavated sites with more unspecific geographic data, or provide data with a larger uncertainty. But, let us note that satellites and knowledge are not the principal source for threatening the relicts of our past. Tombaroli know very well the ancient burial sites, sometimes better than archaeologists do, as shown by the case of the Caligula statue [Telegraph, 2011].

On the other side, Google Earth has been shown [Contreras and Brodie, 2010] to be a source for highlighting cases of looting and to document them. Therefore, we can conclude that the benefits to the knowledge of sites, to their studies and preservations are more than the concerns. Further applications of GIS are under developments for geophysics researches.

References

[Bolstad, 2005] Bolstad P., GIS fundamentals: a first text on Geographic Information Systems, White Bear Lake, Eider Press, 2005.

[Bousman, 2006] Bousman B.C., Satellite archaeology for everyone. The SAA Archaeological Record, 6, May 2006, 32-34.

[Campana, 2002] Campana S., Ikonos-2 multispectral satellite imagery to the study of archaeological landscapes: an integrated multi-sensor approach in combination with "traditional" methods, in The Digital Heritage of Archaeology, CAA 2002, Computer Applications and Quantitative Methods in Archaeology, Proceedings of the XXXth Conference, Heraklion, Crete, 2002, 219-225.

[Chang, 2006] Chang K.-T., Introduction to Geographic Information Systems, Boston, McGraw Hill, 2006.

[Conroy et al., 2008] Conroy G.C., Anemone R.L., Van Regenmorter J., Addison A., Google Earth, GIS, and the Great Divide: A new and simple method for sharing paleontological data, Journal of Human Evolution, 55, October 2008, 751-755.

[Contreras and Brodie, 2010] Contreras D.A., Brodie N., Shining light on looting: using Google Earth to quantify damage and raise public awareness, The SAA Archaeological Record, 10, 2010, 30-33.

[Garrison et al., 2008] Garrison T.G., Houston S.D., Golden C., Inomata T., Nelson Z., Munson J., Evaluating the use of IKONOS satellite imagery in lowland Maya settlement archaeology, Journal of Archaeological Science, 35, 2008, 2770–2777.

[Georgoula et al., 2004] Georgoula O., Kaimaris D., Tsakiri M., Patias P., From the aerial photo to high-resolution satellite image: tools for the archaeological research, Proceedings of the XXth ISPRS Congress, Istanbul, Turkey, 2004.

[Kris Hirst, 2011] Kris Hirst K., Google Earth and Archaeology, Serious Science and Serious Fun with GIS, K. Kris Hirst, About.com Guide, <http://archaeology.about.com/>

[Lasaponara and Masini, 2008] Lasaponara R., Masini N., Advances on remote sensing for archaeology and cultural heritage management, in Lasaponara R. and Masini N. (eds) Proceedings of the 1st International EARSeL Workshop CNR, Rome. 2008.

[Sparavigna, 2010] Sparavigna A.C., The geoglyphs of Titicaca, Archaeogate, October 13, 2010, <http://www.archaeogate.org/classica/article/1305/1/>,

[Sparavigna, 2011] Sparavigna A.C., Stone structures in the Syrian Desert, Archaeogate, July 14, 2011, http://www.archaeogate.org/vicino_oriente/article/1445/1/

[Telegraph, 2011] Caligula statue discovered by 'tomb raiders' unveiled in Rome, The Telegraph, July 12, 2011

[Thomas et al. 2008] Thomas D.C., Kidd F.J., Nikolovski S., Zipfel C., The Archaeological Sites of Afghanistan in Google Earth, AARGnews, 37, 2008, 22-30.

[UNESCO, 2005] UNESCO, Operational guidelines for the implementation of the World Heritage Convention, UNESCO World Heritage Centre, Paris. 2005, p.83.

[Ur, 2006] Ur J., Google Earth and archaeology. The SAA Archaeological Record, 6, May 2006, 35-38.

[WWW1] <http://www.digitalglobe.com/index.php/85/QuickBird>

[WWW2] <http://www.geoeye.com/CorpSite/>

[WWW3] <http://earth.google.com>

[WWW4] Ancient Monuments Placemarks, <http://www.jqjacobs.net/archaeo/sites>

[WWW5] <http://www.digitaldigging.co.uk>

[WWW6] <http://www.online-archaeology.co.uk/archaeology-maps-and-gis.aspx>

[WWW7] <http://www.megalithic.co.uk/article.php?sid=2146412594>

50 e non sentirli! Elastic EUCIP

Maurizio Cassi
c.so L.A. Martinetti 42/28, 16149 Genova GE
maurizio.cassi@gmail.com

1. Elastic EUCIP, rivisitazione online e social del modello di erogazione del servizio di qualificazione delle competenze e delle professionalità ICT.

La celebrata ricorrenza dei 50 anni rappresenta l'imperdibile occasione per AICA per avviare a tutti gli effetti l'erogazione di servizi digitali online di fascia alta in modo da rappresentare un esempio rilevante per la comunità professionale e indicare emergenti linee di indirizzo per tutti gli operatori dell'ICT nazionale.

In particolare la rivisitazione di **EUCIP online** come un servizio completamente digitale a più ampio spettro nella logica dei *social network* e del web semantico, renderebbe possibile, partendo dai consolidati ed efficaci servizi di *assessment* già oggi esistenti ed operanti, di accelerarne il processo di adozione da parte degli utenti, facilitando il *self-empowerment* degli utenti stessi che si avvicinano ai servizi di *assessment* e alle certificazioni, contribuendo all'aggiornamento incrementale dei *syllabi*, introducendo **nuovi modelli di business** online.

Utilizzando le tecnologie del web semantico e sociale, valorizzando processi e *user experience*, **Elastic EUCIP** introdurrà quindi importanti vantaggi **a)** per l'istituzione AICA perché contribuisce a disseminare in maniera pro-attiva l'informazione sul servizio di qualificazione professionale delle competenze e delle professionalità estendendo il bacino di utenza, **b)** per gli utenti perché favorisce le auspiccate dinamiche di self-empowerment, **c)** per i Centri di Competenza Business che oggi offrono tradizionali ma poco accattivanti servizi di formazione in aula e di certificazione EUCIP, **d)** per i CIO e per i Direttori del Personale delle aziende potenzialmente interessate ad EUCIP, **e)** al corpo dei *syllabi* esistenti perché ne faciliterebbe l'aggiornamento.

Il servizio potrebbe essere erogato direttamente agli utenti nel *cloud computing* secondo classici modelli di business online e in modalità *Enterprise* quando rivolto alle aziende più conservatrici del panorama ICT italiano.

La gestione del progetto può essere affidata ad una nuova azienda snella che opererebbe riconosciuta e sostenuta sia da AICA ("per conto di") sia dai Centri di Competenza Business già impegnati a proporre EUCIP al mercato: ogni attore dell'attuale ecosistema potrà essere parte del progetto **Elastic EUCIP** anzi ne sarà ulteriormente qualificato!

I servizi digitali sono scelta obbligata per operare secondo gli emergenti scenari dell'ICT: **Elastic EUCIP** è un servizio digitale che in una fase più avanzata, grazie ad opportune evoluzioni, può aspirare al prestigioso palcoscenico globale.

50 e non sentirli! Elastic EUCIP

Algoritmo di supporto decisionale fuzzy logic per l'attivazione di un Audit di Clinical Risk Management

M. Musolino, T. Mariani, G. Olzai, Q. Placevoli, E. Sesti

PRESENTAZIONE: L'audit di risk management si propone di identificare le criticità organizzative in una struttura sanitaria al fine unico di formulare ipotesi di miglioramento della qualità assistenziale e relative proposte operative in seguito ad un incidente clinico. Mira all'identificazione delle cause organizzative, non delle responsabilità individuali, che hanno favorito l'evento avverso a carico del paziente. L'audit di clinical risk management è quindi uno strumento diagnostico di non conformità strutturali/organizzative. Il contenuto innovativo del presente lavoro consiste nell'applicazione della logica fuzzy per la classificazione dei casi di evento avverso o near miss in quanto permette una migliore rappresentazione dell'incertezza intrinseca ai processi decisionali in ambito sanitario rispetto ad un approccio dicotomico di tipo classico.

METODO: Solitamente la significatività di un evento avverso è valutata sulla base delle conseguenze cliniche di un errore umano/guasto tecnico. Il gruppo di studio ha scelto di discriminare la significatività dei casi in base a due variabili: (A) la gravità di natura clinica e (B) la rilevanza in termini di impatto sistemico sull'organizzazione.



maurizio.musolino@aslromab.it
tiziara.mariani14@aslromab.it
olzai.olulii@gmail.com
q.placevoli@sanfilippo.net.roma1
edilio.sesti@aslromab.it

Punteggi per l'attribuzione della Gravità clinica		
Gravità del danno	Descrizione	Score
Altissima	Danni gravissimi o decesso	10
	Danni o disabilità permanenti	9
Alta	Danno con un prolungamento della degenza con esiti alla dimissione	8
	Danno con un prolungamento della degenza senza esiti alla dimissione	7
	Danno che necessita di un trattamento con farmaci importanti	6
Bassa	Danno che necessita di cure semplici e trattamenti minori	5
	Danno che necessita di osservazioni e procedure diagnostiche	4
	Danno lieve che non richiede alcun trattamento	3
	Danno assolutamente trascurabile	2
Nulla	Nessun danno	1

Punteggi per l'attribuzione della Rilevanza organizzativa		
Rilevanza sistemica	Descrizione	Score
Alta	Evento che esprime potenziali evoluzioni catastrofiche dell'organizzazione	4
	Evento che manifesta criticità estese di sistema, criticità profonde che potrebbero far prefigurare il coinvolgimento di più livelli dell'organizzazione, anche lontani dall'area di accadimento dell'evento avverso	3
	Evento che manifesta criticità circoscritte del sistema, che esclude un possibile coinvolgimento di altri livelli e/o aree organizzative	2
Bassa	Evento che si manifesta in relazione all'esecuzione di un'attività semplice o articolata in più azioni, che è privo di implicazioni organizzative di sistema	1

La classe di significatività è rappresentata dalla inferenza fuzzy logic delle due variabili gravità e rilevanza organizzativa e si stratifica in: **Bassa significatività:** l'evento ha bassa gravità e bassa rilevanza; non si procede all'attivazione dell'audit.

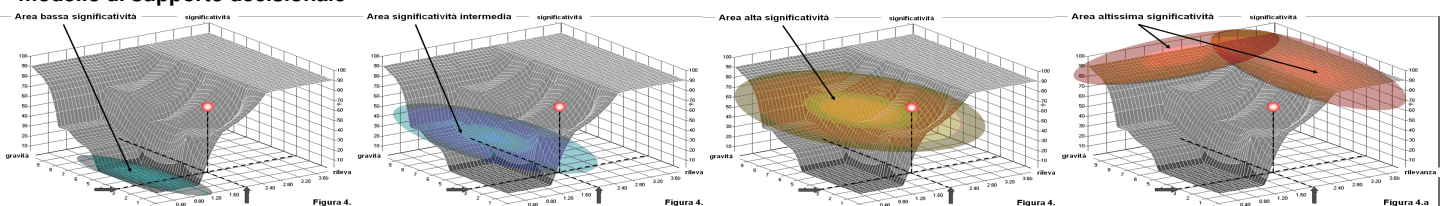
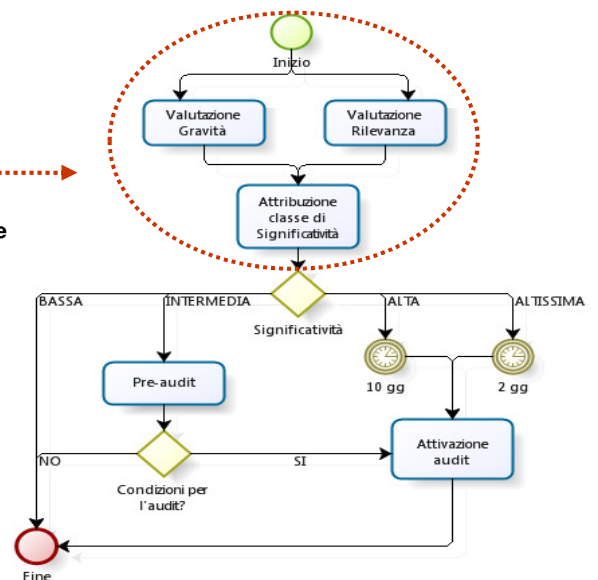
Significatività intermedia: alta gravità e bassa rilevanza; l'evento, seppur grave, ha un debole coinvolgimento con fattori di natura organizzativa e strutturale; si procede ad un pre-audit ricognitivo per valutare la necessità di avviare l'audit.

Significatività alta: bassa gravità e alta rilevanza; l'evento rappresenta un danno lieve o un near miss, ma evidenzia una criticità strutturale; si procede all'attivazione dell'audit.

Significatività altissima (evento sentinella): gravità e rilevanza alte, l'evento comporta dei danni ingenti e coinvolge ad ampio raggio la struttura; l'audit è attivato con la massima urgenza (2 giorni lavorativi).

RISULTATI: Gli autori hanno sottoposto a prova il sistema immettendo in esso i dati relativi agli eventi avversi precedentemente analizzati.

I risultati prodotti dall'utilizzo di tale sistema di supporto decisionale, confrontati con le decisioni precedentemente assunte, hanno riprodotto, nella quasi totalità degli eventi avversi sottoposti a test di confronto, le scelte simili a quelle operate dal risk management aziendale tranne che in un paio di casi analizzati in cui il team decise di non avviare l'audit per effettiva sottostima della significatività. Il sistema è quindi in grado di simulare la modalità esperta per l'attivazione di un audit e rappresentare quindi la base per lo sviluppo di un valido modello di supporto decisionale



Annoluce - Project PRADIGEO

CONGRESSO NAZIONALE AICA 2011

Poster

<http://www.aicanet.it/eventicontestuali/allegati-eventi-2011/CallForPaper2.pdf>

BACKGROUND:

In 2009 **ANNOLUCE** (www.annoluce.net) participated to the Piedmont ICT Regional Call for Bids (POR FESR 2007-2013). Supported by the Department of Literary and Philological Sciences of the University of Turin and the Center of Excellence Geonetlab for the Research in Tele Geomatics and Spatial Information of the University of Trieste, our company got the 5th place in the Innovation area over a total number of 240 participants and it was first in its intervention field (Green web and eco-sustainable practices) after proposing the Project called **PRADIGEO**.

Project **PRADIGEO** provides design and implementation of a computerized platform that acquires, measures, collects, manages, analyzes and publishes data and information, using innovative techniques for environmental and cultural monitoring, tracking and archiving based on a geographical data displacement. Completed development stage (december 2011), it will be defined a list of pilot projects using that system.

FEATURES:

Collect information and data on an environmental and/or cultural context, logical storage to allow the creation of a multimedia consultation system, based on a low cost project consisting of:

- Environment photographically reconstructed with an advanced Virtual Tour System
- 3D objects Capturing and their 3D representation within the context, through photo scanning and texturing and Gigapixel photo
- Geographically accessible web archive
- Data Supply Module for Open Data and sensors feedback analysis

TOPICS:

- Advanced multimedia platforms, ICT tools and cultural content, innovative models of management, 3D digitizing and archiving for the use and promotion of cultural heritage, with particular reference to multimedia technologies and their platforms for the standardization and construction of next-generation digital scenarios that represent cultural heritage (accessible and inaccessible).

Avatar virtuale e e²Commerce

Andrea Carignano, Matteo Gualano¹

Seac02 S.r.l.

Via Avogadro 4, 10121 Torino TO (Cap Città Provincia)

andrea.carignano@seac02.it

¹ Seac02 S.r.l.

Via Avogadro 4, 10121 Torino TO (Cap Città Provincia)

matteo.gualano@seac02.it

1. Un avatar digitale per il *virtual tryon* e l'esperienza di prodotto nello shopping online

Il mercato della vendita di abbigliamento online è stato di circa 42 miliardi di dollari nel 2009 e raggiungerà i 78 miliardi entro il 2014. Tra le inefficienze del business, tuttavia, emerge l'elevato tasso di resi, mediamente del 25%, che implicano una perdita globale considerevole del mercato. Il presente progetto presenta una proposta di soluzione che passa attraverso il concetto di avatar digitale e di una intera piattaforma client server per il supporto ad un servizio di offerta di prova virtuale di abbigliamento.

La piattaforma è basata sul concetto di avatar digitale, un alter ego virtuale depositario delle misure, taglie e forme degli acquirenti. L'avatar è parametrico e personalizzabile dal generico utente che può plasmarlo digitalmente con le proprie misure specifiche secondo un processo definito. Il processo del servizio è quindi completato con un sistema che consente, dato un CAD 2D di un capo, di assemblarne un modello tridimensionale e usarlo per vestire un set di modelli di avatar precalcolati che approssimo in modo esaustivo l'intera popolazione di possibili avatar, intesa come insieme delle possibili combinazioni di misure controllabili. A tal fine sono utilizzati algoritmi di garment fitting e di draping virtuale. Una volta generati i modelli 3D "vestiti", vengono texturizzati per la generazione contenuti destinati al sito dell'online vendor. A seconda del servizio sottoscritto da quest'ultimo saranno generate e accessibili in cloud contenuti 2D in prospettive date, 360° navigabili, 3D, realtà aumentata. Grazie alle sue consolidate esperienze nell'ambito di sviluppo software di realtà virtuale ed aumentata, Seac02 usa tali tecnologie come leva per offrire applicazioni SaaS per l'experience-eCommerce (e²-Commerce) invece di servizi di contenuti per cui sono invece diffuse oggi.

La trasversalità tecnologica è un punto focale del progetto per consentire di proporre ai vari clienti di adottare il sistema dell'avatar sul proprio sito internet (che sarà segnalato come avatar-compatibile), eventualmente a livelli di complessità differente, per offrire il tryon virtuale dei propri capi di abbigliamento o accessori.

Confronto tra la metodologia top-down e bottom-up per la valutazione del contributo del trasporto stradale all'inquinamento atmosferico.

L. Pallavidino, R. Prandi, M. P. Costa¹, A. Nanni¹, A. Bertello², E. Bracco², F. Pavone²

Simularia Srl
Via Principe Tommaso 39, Torino (TO)
info@simularia.it

¹Arianet Srl
Via Gilino 9, Milano (MI)
info@aria-net.it

²Provincia di Torino, Servizio Qualità dell'Aria e Risorse Energetiche
C.so Inghilterra 7, Torino (TO)
aria@provincia.torino.it

Questo studio è finalizzato a descrivere dettagliatamente il contributo del trasporto stradale all'inquinamento atmosferico nell'area metropolitana torinese a partire da dati di traffico e di infomobilità secondo un'approccio bottom-up.

Per la stima delle emissioni, riportata negli inventari regionali, è solitamente preferita una metodologia top-down, che utilizza variabili disponibili a grande scala per la stima delle emissioni a livello regionale e le ripartisce a livello comunale attraverso l'uso di variabili proxy reperibili per questa dimensione amministrativa. Per la stima delle emissioni da trasporto stradale secondo la metodologia top-down, si utilizzano tipicamente il consumo di combustibile e il parco veicolare regionale come variabile a grande scala e il parco veicolare immatricolato in ciascun comune per la ripartizione delle emissioni complessive.

Le emissioni del trasporto stradale dell'area metropolitana torinese sono state ricostruite con una metodologia bottom-up a partire dai dati dei flussi veicolari che la società di gestione del traffico "5T" ha stimato per il 2008 per ciascuna ora dell'anno su una rete di 5125 archi stradali, grazie ai rilievi di un migliaio di sensori e a un modello di assegnazione che opera in tempo reale.

I flussi veicolari sono stati ripartiti tra le diverse categorie veicolari sulla base dei rilievi di monitoraggio del traffico effettuati alle barriere di pedaggio autostradale, su alcuni archi urbani, nonché sulla base delle indagini di mobilità.

I fattori emissivi utilizzati per il calcolo delle emissioni sono quelli elaborati dalla EEA (European Environmental Agency) e dallo IIASA (International Institute for Applied Systems Analysis). Le emissioni ottenute secondo le due metodologie sono state raffrontate e sono state evidenziate nel risultato finale sia le differenze nella quantità stimata che le differenze nella ripartizione tra le diverse categorie veicolari. Le emissioni stimate con l'approccio bottom-up sono state usate come input di un modello fotochimico euleriano per la simulazione della dispersione degli inquinanti in atmosfera attraverso il quale sono state calcolate le concentrazioni degli inquinanti in atmosfera per un raffronto con i dati rilevati dalle centraline di monitoraggio.

DAL MONUMENTO ALLA PIETRA LECCESE IN TEMPO REALE: E' POSSIBILE E SENZA ERRORI

Ing. Ivano De Luca¹, Avv. Giorgia Marrocco², dott. Ugo Pietropaoli³, dott. PierPaolo Abet⁴

¹PI.MAR. – PIETRA MARROCCO – s.r.l. –, S.S. LECCE MAGLIE uscita MELPIGNANO Z.I., 73020 MELPIGNANO (LE), deluca.ivano@inwind.it

²PI.MAR. – PIETRA MARROCCO – s.r.l. –, S.S. LECCE MAGLIE uscita MELPIGNANO Z.I., 73020 MELPIGNANO (LE), direzione@pietraleccese.com

³VALUEUP s.r.l., LABORATORIO ALTAMENTE QUALIFICATO MIUR, via cesco baseggio 122, 00139 ROMA, u.pietropaoli@valueup.it

⁴VALUEUP s.r.l., LABORATORIO ALTAMENTE QUALIFICATO MIUR, via cesco baseggio 122, 00139 ROMA, abetp@hotmail.it

Rilevare ciò che ci circonda è sempre stato un obbligo istintivo da parte dell'uomo, che fotografa e ricorda per avere davanti oggetti, monumenti o scenari particolari e affascinanti. Il mondo è pieno di opere d'arte da fotografare, ma oggi è possibile anche riprodurle con una pietra particolare, la pietra leccese, attraverso il risultato della ricerca effettuata dalla ditta PI.MAR. – Pietra Marrocco – srl, con sede in Melpignano, a Lecce, cofinanziata dalla Regione Puglia attraverso il Programma Operativo 2007-2013, Asse I, Linea 1.1 – Aiuti agli investimenti in Ricerca per le PMI.

Tale progetto di ricerca ha previsto l'integrazione di un sistema di scansione ottica tridimensionale con un dimostratore software in grado di analizzare la rilevazione dal file STL generato dalla rilevazione 3D, comprimerlo per ottimizzarne le dimensioni, selezionare automaticamente il canale trasmissivo più idoneo attraverso un sistema di intelligenza artificiale, e trasmetterlo direttamente nel data center di PI.MAR. srl affinché possa essere rivisto dal personale tecnico specializzato, eventualmente ritoccato, e successivamente ricostruito mediante robot che lavora la pietra leccese.

Tale dimostratore non solo permette di evitare, all'azienda, costosi incarichi ad aziende terze per inserire la codifica digitale degli oggetti rilevati all'interno del robot di produzione per la realizzazione in pietra leccese, ma ha **come importante e consequenziale risolto quello di poter generare una banca dati in cui memorizzare tutte le opere architettoniche e monumentali a livello internazionale (si pensi alla possibilità di ricostruire opere danneggiate da calamità naturali o da attacchi di natura dolosa!)**.

La ricerca ha compreso l'integrazione di questo nuovo processo altamente all'avanguardia affinché possano esserne valutati gli impatti con lo stato attuale di PI.MAR. srl, in termini sia economici che prestazionali, al fine di poterlo validare mediante valutazione basata su indicatori relativi ai singoli processi aziendali. La forza del progetto consiste nell'elemento intelligente che è in grado di analizzare, mediante banda di upload e download, qual è il canale comunicativo al momento più idoneo alla trasmissione (Ethernet, WiFi, GRPS, 3G, ecc) per far in modo che la consegna avvenga nel minor tempo possibile e senza errori di trasmissione, utilizzando codici di correzione automatica o sistemi di alerting.

Il progetto nasce dalla problematica di dover utilizzare aziende esterne di rilevazione, analisi e trasmissione, con conseguente dilatazione dei tempi, aumento dei costi incidenti il lavorato, e percentuale di errori e correzioni. Ora, proseguendo con lo sviluppo e la sperimentazione dell'attività di ricerca realizzata in collaborazione con il laboratorio riconosciuto dal MIUR come altamente qualificato, ValueUp srl di Roma, sarà possibile realizzare un pacchetto software installabile ed utilizzabile su molteplici piattaforme ed integrabile con software di terze parti di scansione per ottimizzare risorse umane e processi aziendali.

La ricerca ha avuto una durata di circa 2 anni ed ha condotto a notevoli risultati che hanno unito il mondo dell'edilizia e del lapideo con la frontiera della ICT.

BIBLIOGRAFIA

- Stefano Margiotta, Sergio Negri, Alla ricerca dell'acqua perduta. Nuove conoscenze del sottosuolo nel Salento leccese, Congedo, 2004.
- Chryssolouris, G., Sheng, P., and Choi, W., Three-dimensional laser machining of composite materials, ASME Journal of Engineering materials and thechnology, vol. 112, pp. 387-392, 1990
- Bertazioli Onelio, Favalli Lorenzo, GSM-GPRS - seconda edizione, Hoepli, Giugno 2002
- La Pietra nell'Arredo Urbano Autori e Curatori: Giorgio Blanco Edizione: 2001

Data quality e Open data

Domenico Natale
Commission UNINFO JTC1/SC7 "Software Engineering"
c/o Corso Trento 13, Torino – Italy
nataledo@tiscali.it

1. Presentazione del poster

Il concetto di **qualità dei dati** è definito dal recente standard ISO/IEC 25012. Con il presente lavoro si sottolinea l'importanza di disporre nei vari contesti della massima qualità dei dati anche al fine di supportare l'**Open data**. Si considera come l'Open data stesso potrà favorire a sua volta il raggiungimento della qualità dei dati. La quantità dei dati contenuta nei sistemi informativi è in continua crescita e sarà meglio governata mantenendo elevati livelli di qualità con il contributo dell'utente finale, spesso unico detentore della conoscenza del dato.

Come afferma Norman la conoscenza è nella mente umana, mentre l'informazione è fuori di essa: nei testi scritti e elementi culturali. Per questa ragione è importante verificare se l'informazione è conforme il più possibile a **caratteristiche attese di qualità**, consentendo di evitare o ridurre confusioni, stress o problemi nella mente umana e nei suoi schemi mentali/logici. Questo significa che un'adeguata coerenza dovrebbe essere presente tra l'informazione percepita dall'utente come attinta dall'ambiente reale, la qualità della sua conoscenza formatasi con i processi cognitivi, il software e le interfacce che a volte attingono da fonti diverse.

L'importanza di controllare la qualità dei dati è cruciale per migliorare i processi che causano scarsa qualità. I migliori risultati si ottengono con misurazioni aggiornate automaticamente e tenute sotto controllo in data base e grafici Radar disponibile online in rete. Seguendo un concetto di "Total quality" chi apre un processo dovrebbe essere in grado di chiuderlo. L'utente dovrebbe poter "crear"e e seguire il proprio dato, sia a livello singolo, se riservato, sia a livello statistico. In questa prospettiva appare utile la pratica dell'Open data.

Si sta affermando sempre più l'approccio, specialmente in ambito pubblico, di **rendere disponibili i dati on-line**. Dati interni divengono così esterni e dati che in passato rimanevano inutilizzati possono offrire a studiosi e semplici cittadini nuove opportunità di conoscenza di fenomeni di interesse pubblico: ambiente, salute, mobilità, inquinamento, turismo, vita cittadina, ecc.. I benefici della diffusione dei dati riguardano cittadini, Pubbliche amministrazioni e imprese, facendo aumentare un senso di trasparenza e partecipazione, credibilità e vicinanza degli organi istituzionali alle persone. L'Open data presuppone la disponibilità di dati di alta qualità ed il raggiungimento della qualità dei dati si può avvalere delle pratiche di Open data. In tale prospettiva la società si muove dalla mentalità del controllo a quello della partecipazione e verso un'accelerazione dello sviluppo governato da dati certi e tempestivi.

Easy Booking PARKING WEB APPS

Giuseppe PITTARI, Pasquale BARREGA

ISE Engineering Lab

Via Umberto I n.37 98063 Gioiosa Marea (ME)

Ise@com2000.it

Il parcheggio della nostra automobile è una necessità quotidiana, che può diventare un problema senza soluzione se la destinazione è un luogo ad elevata frequentazione. L'incertezza di trovare un "posto libero", in prossimità della meta desiderata, determina ritardi inutili negli orari di arrivo, ulteriore traffico locale, parcheggi selvaggi, etc. che ognuno di noi avrà sperimentato sicuramente. ISE, azienda specializzata in Hi-Tec wireless (GPRS / UMTS), ha sviluppato un sistema denominato "**WEB PARKING**" in grado di incontrare tale necessità ed offrire una soluzione. Dettagli pubblicabili sono disponibili all'indirizzo <http://www.com2000.it>

Il sistema si basa su un'architettura **client-server**. Il client è un sistema di trasmissione dati in modalità "trasparent", impiegante un dispositivo della famiglia **ISE M2M power IDE**, che trasmette la disponibilità, ordinata e numerata di ciascun parcheggio, in tempo reale al server e riceve da esso le eventuali opzioni/ricieste degli utenti. Il server ospita il data base generale contenente le info di tutte le strutture di parcheggio rientranti in una determinata area geografica e li mette a disposizione di una WEB App per la gestione. Applicazione che risulta disponibile per vari Browser, anche per palmari e telefonini 3G ormai di uso comune. L'utente ha così a disposizione una "mappatura" del potenziale di parcheggio dell'area geografica complessiva di interesse e può "prenotare" prima dell'arrivo il proprio "posto" in modo da avere certezza e disponibilità immediata. Tale operazione può essere compiuta anche durante la marcia di avvicinamento con l'utilizzo di device mobile GPRS/UMTS da impiegare con le dovute sicurezze richieste dal codice stradale, oppure, in alternativa, compiuta da rete internet "fissa".

Restano salvi i vincoli di convenzione che ciascun ente gestore vuole riservare alla propria clientela per garantire l'impegno anticipato di ciascun posto auto e della relativa tariffazione.

Il gestore : dispone di uno strumento moderno per il servizio che gli garantisce la possibilità di "venderlo" prima del suo utilizzo con ottimizzazioni, ad oggi inimmaginabili, in termini di ricavi e di costi di gestione.

L'utente : ha praticamente dato certezza alla reale disponibilità di servizio e quindi "paga" per qualcosa che avrà a sua disposizione all'arrivo e potrà gestire tempi e costi in maniera flessibile e programmata senza inutili "giri a vuoto" per la città alla ricerca di qualcosa di introvabile. Non stiamo ad elencare vantaggi secondari che pur si potrebbero conteggiare come contributo ecologico alla riduzione del traffico "ridondante", inutilità della segnaletica mobile che fornisce la disponibilità in loco a volte non reale, che eccita la nostra fantasia nella speranza di trovare posto durante l'avvicinamento.



Eucip Core nei curricoli dell'Istituto Tecnico "Amministrazione Finanza e Marketing, articolazione Sistemi Informativi Aziendali"

Adriana Fasulo

Docente di Informatica ITCG E. Fermi Pontedera (PI)

e-mail: adriana.fasulo@libero.it

L'importanza delle certificazioni europee:

EUCIP (European Certification of Informatics Professionals) è il sistema europeo di competenze e profili professionali informatici, sviluppato, con il contributo della Comunità europea, dalle associazioni europee del settore raccolte in un organismo internazionale (CEPIS Council of European Professional Informatics Societies), tra cui AICA per l'Italia.

EUCIP definisce gli standard delle competenze per ciascun profilo raccogliendole in un Syllabus, vero e proprio punto di riferimento nel mondo delle professioni informatiche.

E' riconosciuto sia in ambito universitario (CRUI e CINI) sia in ambito imprenditoriale (Confindustria) che da grandi organizzazioni pubbliche.

EUCIP Core rappresenta la base delle certificazioni del framework EUCIP: esso verifica il possesso di un ampio spettro di conoscenze ed abilità basilari, che dovrebbero essere comuni a tutti i professionisti informatici.

Esigenza dell'intervento formativo:

Il continuo sviluppo delle tecnologie dell'informazione e della comunicazione, la nuova **riforma degli istituti tecnici**, ha posto l'esigenza di un aggiornamento permanente dei contenuti e delle metodologie di lavoro evidenziando la necessità di attivare una didattica ed una valutazione basate sulle competenze.

Le attuali tendenze del mercato del lavoro richiedono dei professionisti con competenze **certificabili e certificate**. In particolare sono richieste figure professionali che riescano a coniugare competenze sui processi aziendali e sui sistemi informativi con competenze ampie e tecnologicamente avanzate nel campo dell'ICT.

I contenuti tecnologici, organizzativi ed economici risentono di una forte esigenza di rispecchiare le richieste del mercato del lavoro, di definire un **profilo professionale in uscita dall'ambiente scolastico** che sappia cogliere le relazioni con il sistema informativo e i processi aziendali in continua evoluzione.

Introduzione della sperimentazione:

Partendo da questa esigenza di rinnovamento e riorganizzazione, l'Istituto ITCG "E. Fermi" ha raccolto il suggerimento di **AICA e MIUR** di introdurre nei curricoli del Corso Mercurio i contenuti proposti dalla certificazione EUCIP Core e di effettuare insieme ad altre due Istituti italiani (I.T.C. Romanazzi di Bari e I.T.C. Baffi di Fiumicino) una sperimentazione che si articola nel corso del triennio Mercurio. Al termine del percorso scolastico, gli studenti che hanno sostenuto e superato gli esami previsti potranno conseguire la certificazione EUCIP Core.

Sono destinatari della presente sperimentazione gli studenti dell'ITCG "E. Fermi" frequentanti il triennio informatico. E' stato attivato a partire dalle classi terze dell'indirizzo Mercurio nell'anno scolastico 2010/2011. Durante il percorso, agli studenti sono proposti sia contenuti che fanno parte già dell'attuale percorso ministeriale, sia nuovi contenuti che sono alla base di competenze professionali. Essi sono rivisti dal punto di vista metodologico e caratterizzati da una stretta **interdisciplinarietà** (Informatica, Economia Aziendale, Lingue, Diritto e Italiano).

Nuova didattica e metodologia:

Gli studenti coinvolti nella sperimentazione devono superare **3 esami in lingua inglese** relativi a 3 moduli di Eucip Core (**Plan, Build e Operate**). Il superamento di questi esami permette agli studenti di conseguire la certificazione EUCIP Core.

L'Eucip Core contiene competenze multidisciplinari per l'applicazione delle tecnologie dell'informazione e della comunicazione nel mondo del lavoro. I contenuti sono proposti sia in modo interdisciplinare, sia mediante attività progettuali che laboratoriali. Alcune lezioni si svolgono con il supporto del **CLIL**, esaltando un approccio in cui i contenuti sono pensati e studiati in lingua inglese.

La centralità dello studente è esaltata dall'uso quotidiano di lezioni interattive con le **LIM** (ogni classe coinvolta ha una LIM) e con l'uso di piattaforme di **e-Learning** che supportano lo studente nella fruizione di contenuti digitali.

Per fornire agli studenti dei riferimenti completi e di supporto alla loro preparazione che integrino le competenze multidisciplinari, sono in corso di definizione dei materiali didattici autoprodotti dai docenti (**Learning Object**).

Lilith, a simple way to analyze your data

Roberto Marchisio (President and CTO)

Hicare Research S.r.l.

Via Livorno, 60 - 10144 Torino (TO)

Tel: +39 011.225.85.51

roberto.marchisio@hicare.com

Lilith is a Business Intelligence tool by Hicare Research developed in 1996, having the capability to analyze data of any type (numeric, but also text and link to multimedia files).

Data can be navigated according to all dimensions (months, years, clients, producers, assets, banks...) and along the hierarchical structure.

The front-end has impressive capabilities of analysis and representation, with more than 20 graph types.

In order to allow the multidimensional analysis with virtually no waiting time also with 500 million records and more (in the range of milliseconds, independent of the size of the data repository), an own internal engine was created, totally integrated in the software.

All data are therefore saved in an optimized way, in order to be ready for exploration.

A brand new set of data may be derived easily at any time, based on actual data.

Any further hierarchical level may be added at any time.

This is very useful in comparisons among actual, budgeted, revised and what-if simulation data.

All analysis may be performed on Lilith Client or on Lilith Webserver, in both cases with write-back capabilities.

In a single suite, sound and independent of external libraries, created in C++ language and therefore multi-platform, Lilith can simplify the use, interpretation and share of data throughout a company or organization.

This makes the product suitable for business control and for marketing simulation over future activities and top-down or bottom-up budget negotiations.

It is suitable besides for bank and insurance performance and trend analysis, and for the storage and analysis of multimedia and physics data.



Monitoraggio energetico: service robots come alleati?

Roberto Antonini, Gian Piero Fici¹, Marco Gaspardone²

Telecom Italia S.p.a.

Via Reiss Romoli 274, 10148 Torino

roberto1.antonini@telecomitaly.it

gianpiero.fici@telecomitalia.it

marco.gaspardone@telecomitalia.it

1. Il monitoraggio energetico e la “Service Robotics” ai tempi dell’Internet of Things

Telecom Italia SpA (TI) è uno tra i primi consumatori Italiani di elettricità tanto da impiegare l’1% dell’intera produzione italiana. Il miglioramento dell’efficienza energetica nelle applicazioni industriali e domestiche, rappresenta un modo per soddisfare i requisiti dell’Unione Europea in termini di emissioni e di risparmio energetico ed è stato incluso tra le principali azioni da intraprendere nel quadro del cambiamento climatico “20-20-20”.

Analizzando i consumi di TI circa l’80% dell’energia è utilizzata per far funzionare la rete di Telecomunicazione (centrali telefoniche, data center, stazioni radio-base...) e circa la metà di questa energia viene spesa nel condizionamento dell’aria. Negli ultimi anni TI ha introdotto diversi sistemi di monitoraggio e controllo in tempo reale dei suoi consumi energetici utilizzando massivamente le **reti di sensori** e implementando al suo interno una piattaforma di aggregazione, analisi e fruizione di tutti i dati collezionati dai vari punti misura. Un dispiegamento dell’Internet of things (IoT) che ha dato sicuramente il suo risultato in termini di saving energetico, previsione dei consumi futuri, miglioramento nella progettazione dei nuovi impianti.

Oggi la ricerca di TI ha cominciato a guardare oltre l’utilizzo della sola sensoristica e di sistemi di attuazione installati in punti prestabiliti in quanto, in alcuni contesti, gli spazi o gli apparati da monitorare sarebbero tali da non permetterne una copertura capillare. Ecco così farsi strada in questo settore la **robotica di servizio**. Un *service robot* difatti ha caratteristiche che lo rendono idoneo alle attività di monitoraggio e attuazione: è un apparato che interagisce col mondo fisico (si muove, apre, prende, porta, osserva), è un apparato dotato di mobilità, è un apparato che esegue il proprio compito con un certo grado di autonomia.

Grazie alla disponibilità di soluzioni sempre più evolute e a costi sempre più contenuti la robotica sta così diventando un elemento importante nell’IoT, punto intelligente, sempre connesso alla rete Internet e alle reti di “oggetti” che lo circondano in modo da interagire con entrambe le reti per portare a termine i suoi compiti al meglio possibile. Se la sfida sarà raggiunta, il monitoraggio energetico avrà acquisito un nuovo alleato nella sfida del saving.

Musei virtuali, immagini reali

Roberto Marazzato, Amelia Carolina Sparavigna¹
Dipartimento di Automatica e Informatica, Politecnico di Torino
Corso Duca degli Abruzzi 24, Torino, Italy
roberto.marazzato@polito.it

¹*Dipartimento di Fisica, Politecnico di Torino*
Corso Duca degli Abruzzi 24, Torino, Italy
amelia.sparavigna@polito.it

1. Introduzione

Al concetto di museo virtuale si tende ad abbinare una collezione di oggetti digitali di tipo artistico o culturale, accessibile per via telematica. Un museo virtuale può quindi essere una collezione d'immagini di quadri o disegni, se siamo nell'ambito artistico figurativo, oppure una collezione di fotografie, video o ricostruzioni 3D di statue, ambienti architettonici o siti archeologici. Ovviamente, nella definizione di museo virtuale troviamo sia la visualizzazione di quanto si trova nelle sale di un museo tradizionale, sia la collezione di risorse espressamente creata per la rete Internet, detta "museo on-line".

Sempre più numerosi sono i musei reali che si sono dotati di spazi sul World Wide Web. Un interessante esempio è il MAO di Torino [1]. Un esempio invece di museo virtuale on-line è dato dal sito [2], che presenta la documentazione fotografica di diversi musei che possiedono collezioni di antichi reperti egizi e sudanesi.

2. L'implementazione

Dal punto di vista della realizzazione del museo, essa è collegata direttamente all'interfaccia di navigazione scelta. Spesso il sito Web si deve adattare alla catalogazione delle collezioni seguendo l'esempio dei database museali. Le tecnologie usate sono in genere Web standard, con immagini in formato. La collezione delle immagini è tra gli aspetti più onerosi della realizzazione, in particolare l'acquisizione delle immagini reali. Anche il passare queste immagini in un formato unico adatto al Web, con Photoshop o GIMP, agendo su ogni singola immagine, può diventare altrettanto oneroso.

In effetti, basta identificare le operazioni base che si fanno di solito sulle immagini e creare un software che le parametrizzi e automatizzi efficacemente, con la possibilità di un'eventuale ambientazione dei reperti. Il lavoro di ricerca propone due soluzioni: una è quella di automatizzare completamente il processo, l'altra è quella di ridurre al minimo e dotare di strumenti specifici la parte da fare necessariamente a mano.

3. Riferimenti

1. Museo d'Arte Orientale di Torino, <http://www.maotorino.it/>
2. Virtual Egypt, <http://www.accessingvirtualegypt.ucl.ac.uk/index.php>

PAZIENTI ONLINE: IL CASO WWW.CIDP.IT

Massimo Marra, Raffaella Longo
Associazione Italiana Neuropatie Periferiche ONLUS
Corso del Popolo 1, 35131 Padova (Pd)
info@cidp.it

Le neuropatie disimmuni sono un gruppo di malattie del sistema nervoso periferico la cui causa si ritiene essere legata ad una aggressione del sistema immunitario contro antigeni del nervo periferico. Tale gruppo di malattie si è molto ampliato negli ultimi anni e la CIDP (polineuropatia cronica infiammatoria demielinizzante) è tra queste. La CIDP esordisce talvolta in modo acuto ma più frequentemente presenta dall'inizio un andamento cronico. In circa la metà dei casi (soprattutto nei giovani) la malattia ha un andamento cronico recidivante, mentre negli altri pazienti ha un andamento cronico progressivo sin dall'inizio. La prognosi a lungo termine è buona in circa la metà dei pazienti.

Il sito www.cidp.it nasce dopo l'esperienza familiare degli autori alla fine del 2009 con l'obiettivo di informare su una malattia sconosciuta ai più e ampiamente sotto-diagnosticata.

Attualmente il sito conta circa 400 visite al giorno ed è diventato il punto di riferimento per pazienti, specialisti e case farmaceutiche. Sul sito vi è una periodica pubblicazione di abstract tratti da ncbi spesso accompagnate da interviste ai maggiori specialisti italiani e internazionali.

Oltre all'informazione sulla malattia, il sito è un importante strumento di comunicazione per rivendicare il diritto alla cura di immunoglobuline; cosa invece che è spesso negata per mere questione di budget sanitario.

Nel periodo gennaio 2010 – luglio 2011 le pagine visitate del sito sono state 72629 per 20503 visitatori differenti con una media di 3,54 pagine/visita; il bounce rate del sito è di 56,70% con un tempo medio di permanenza di 2,50 minuti e la percentuale di nuove visite pari a 81,74%. Quasi il 5% dei visitatori proviene da paesi esteri e tra questi si segnalano Svizzera, Stati Uniti, Gran Bretagna, Germania, Francia, Belgio, Austria e Spagna. Per le città italiane, le visite più frequenti provengono da Roma, Milano, Torino, Napoli, Firenze, Padova, Bologna, Catania, Palermo e Cagliari.

Il traffico diretto è pari al 7,54%, quello derivante da siti referenti è del 3,70% mentre l'88,76% proviene da motori di ricerca. Google in particolare genera da solo l'83,60% del traffico del sito e le parole più ricercate sono nell'ordine: cidp, neuropatia periferica, polineuropatia demielinizzante, cidp polineuropatia, neuropatia idiopatica, poliradicolonevrite infiammatoria demielinizzante cronica, poliradicolonevrite, cosa sono le immunoglobuline, cidp sintomi, neuropatia cura, cidp terapia, rf0180, cortisone.

Le pagine più visitate sono nell'ordine: home page, sintomi, cos'è una neuropatia periferica, cura della neuropatia, centri di riferimento regionali CIDP, patogenesi CIDP, informazioni utili per pazienti e famiglie, come avere l'esenzione.

L'obiettivo per il prossimo anno è quello di potenziare il posizionamento sui motori di ricerca del sito per le keyword relative ai sintomi.

Phase Changes Materials and Controls in Thermodynamic Models

Amelia Carolina Sparavigna
Dipartimento di Fisica, Politecnico di Torino
Corso Duca degli Abruzzi 24, Torino, Italy
amelia.sparavigna@polito.it

1. Introduction

The study of the thermal behaviour of macroscopic systems is quite important because of its usefulness in simulating the temperature behaviour and heat exchanges of local environments. Due to the current condition of an increasing average temperature coming from the global warming, these simulations could help in offering new solutions to reduce the energy consumption and prevent side effects. Here I am proposing and discussing the thermal behaviour of models, which simulate macroscopic structures having some parts consisting of Phase Changes Materials (PCMs). These are materials able to store the thermal energy. Among the various methods for energy storage, those based on the latent thermal energy of PCMs are widely considered as able to provide highly effective systems [1].

2. Models and simulations

As a possible approach to simulate the behaviour of macroscopic volumes, which include some energy storage systems with PCMs, I am proposing the use of models composed of several parts, each obeying the laws of thermodynamics. These parts are interacting with heat exchanges. Some parts are in connection with the external environment. The thermal behaviour of the models is obtained by means of a simulation based on lumped elements, where the description of spatially distributed physical systems is realized through a topology consisting of discrete entities. Under certain assumptions, the simulation with lumped elements, originally developed for electrical systems, is suitable to solve and determine the behaviour of a distributed system [2]. This approach has been already proposed for systems under period conditions [3]. Here, besides PCMs, passive and adaptive controls are included in the models to study the temperature optimization inside specific environments.

3. References

1. B. Zalba, J.M. Marin, L.F. Cabeza and H. Mehling, *Applied Thermal Engineering*, Vol. 23, No. 3, 2003, pp. 251-283.
2. M.G. Davies, *Building Heat Transfer*, John Wiley & Sons, Inc., New York, 2004.
3. A.C. Sparavigna, S. Giurdanella, M. Patrucco, *Energy and Power Engineering*, Vol.3, 2011, pp.150-157.

Reliable On-Chip Power Distribution Using Graphene Nanoribbons

Sandeep Miryala, Andrea Calimera, Enrico Macii, Massimo Poncino
Politecnico di Torino

C.so Duca degli Abruzzi, 24, 10129 Torino

{sandeep.miryala; andrea.calimera; enrico.macii; massimo.poncino}@polito.it

In state-of-the-art Integrated Circuit (IC) technologies, power supply is distributed across the chip through copper metal wires organized as multilayer grids. Although effective in delivering stable V_{dd} and **ground** levels, copper-based structures suffer of many wear-out mechanisms which limit their reliability during the lifetime of the IC. Among such mechanisms, **Electromigration** (EM) is certainly the most critical.

EM refers to the gradual displacement of the metal atoms of a conductor from one point to another due to the current carried by the conductor itself. This mechanism leads to the formation of **voids** at some points in the metal line and **hillocks** in other points, which, in turn, result in the creation of **open-circuit** of individual wires or **short-circuits** among adjacent wires. The maximum current density (J_{EM}) that can safely flow in a wire strictly depends on the material ($J_{EM}=10^7 A/cm^2$ for copper). A common design practice is to size the cross-section of the metal wires such that $J_{MAX} < J_{EM}$, where J_{MAX} is the maximum equivalent dc current expected to be drained by the circuit divided by the cross-sectional area of a metal wire itself. However, as well stated in the International Technology Roadmap for Semiconductors (ITRS), due to aggressive geometry scaling and higher integration densities of future nodes (i.e., 22nm and beyond), the on-chip power demand will exceed the physical limit copper metal wires can tolerate. Needless to say, this poses a severe limitation to the design of the power distribution network of the next generation of ICs.

Emerging technologies based on new one-dimensional carbon nanomaterials, like **Graphene Nanoribbons** (GNRs), offer potential solutions to the reliability problem. Besides showing longer mean-free path, which traduces in higher carrier mobility, the ballistic transport that governs current flows makes such materials resilient to EM. Recent results have shown a maximum current density that is one order of magnitude larger than that of copper (i.e., $10^8 A/cm^2$). Enabling the graphene technology in the VLSI domain could have a true impact on the semiconductor industry, and may open new research scenarios.

In this work, we investigate on the effective capability of GNRs to replace copper in realizing on-chip power distribution networks. More specifically, we prove that, if properly sized, power-grids based on multi-layer GNRs can show smaller equivalent resistance than copper, thus allowing smaller voltage drops, higher speed, lower power consumptions, and, most important, immunity to EM. The results obtained through our simulation framework, which allows to emulate real current work-load through grids of GNRs, confirm this claim by showing a IR-drop that is 30% smaller than in the case of copper.

SMART design: trasposizione efficace ed emulazione efficiente per la didattica tradizionale

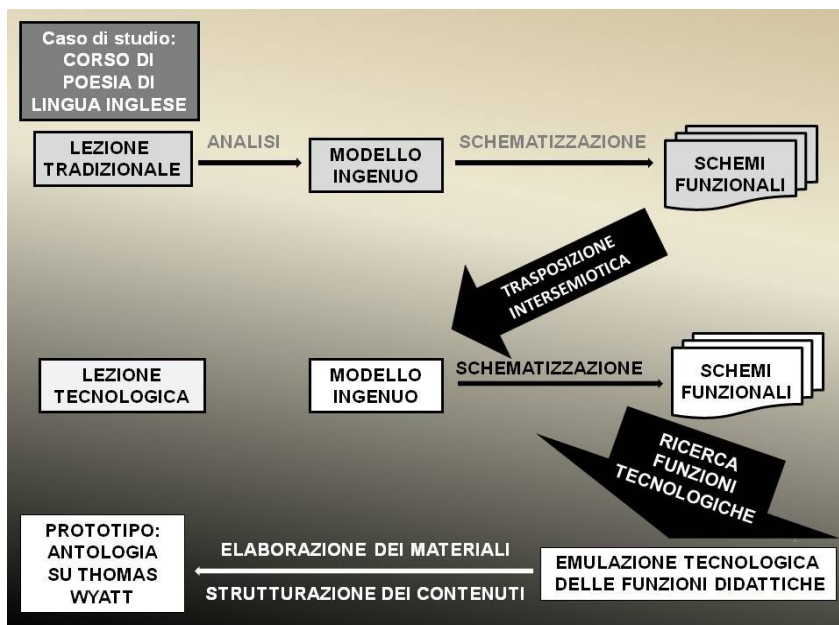
Luigi Scazzola, Amanda Reggiori¹, Marco Padula²

S.I.TE - Supporto E-learning, Università Cattolica del Sacro Cuore - Sede di Brescia
Via Trieste 17, 25121 Brescia
luigi.scazzola@unicatt.it

¹FSLLS-Università Cattolica del Sacro Cuore – Sede di Brescia
Via Trieste 17, 25121 Brescia
amanda.reggiori@unicatt.it

²CNR - Istituto per le Tecnologie della Costruzione, US Milano
Via Bassini 15, 20133 Milano
padula@itc.cnr.it

L'esperienza di e-learning design riferita ha come obiettivo la trasposizione di un corso universitario di poesia di lingua inglese e la realizzazione di un prototipo funzionante. Il corso segue una metodologia consolidata e attraverso il prototipo è reso in versione da erogare con le tecnologie informatiche. Sono condizioni preliminari la fedeltà ai testi originali, il metodo didattico e di apprendimento, lo stile personale del docente e le convenzioni didattiche adottate.



Schema del processo di trasposizione

SMARTPHONE-CENTRIC REMOTE MONITORING ARCHITECTURE FOR PATIENTS SUFFERING FROM HEART FAILURE

Igor Bisio, Fabio Lavagetto, Mario Marchese, Andrea Sciarrone

Department of Communications, Computer and System Science – University of Genoa, Italy

e-mails: { igor.bisio, fabio.lavagetto, mario.marchese, andrea.sciarrone }@unige.it

THE CONSIDERED PATHOLOGY

The Heart Failure (HF) is a chronic disease that alternates intense and weak phases and requires repeated and frequent hospital treatments. The use of automatic instruments for a remote and ubiquitous monitoring of biological parameters, relevant with respect to the HF pathophysiology, offers new perspectives to improve the patients' life quality and the efficacy of the applied clinical treatment.

THE REMOTE MONITORING ARCHITECTURE

The architecture described in this poster represents a practical realization of the mentioned automatic remote monitoring instruments where the smartphones play a crucial role and it is not employed just to communicate with traditional client-server interactions. It is the sink of information sent by several sensors such as chest strips, weight steps and pulse oximeters (traditional approach) and it is also an intelligent and autonomous sensor of the patients' motion based on an **activity recognition** algorithm with high level of accuracy (innovative approach).

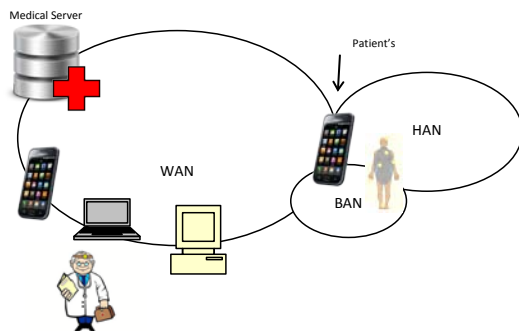


Figure 1. Remote Monitoring General Architecture

The architecture is divided into three networks: the Body Area Network (BAN), the Home Area Network (HAN) and a Wide Area Network (WAN). The relationship among the mentioned networks is reported in Fig. 1. The BAN is composed by the wearable sensors aimed at monitoring specific parameters. A HAN (simply called home network) is a common residential Local Area Network (LAN), which is usually used for communications between digital devices typically deployed at home. Here, the HAN is composed by (not wearable) sensor nodes aimed at monitoring specific clinical parameters such as weight and oxygen saturation of patients. Finally, the WAN is the network employed to transfer information from the BAN and the HAN to a server, called Medical Server (MS), where the parameters of all the monitored patients are stored and made available to the medical staff.

THE PATIENTS' SMARTPHONE ROLE

It plays the most important role in the proposed architecture since it provides several functions useful for the monitoring aim of

the overall architecture. From the technical viewpoint, all functions are grouped into a unique software application. In Fig. 2 a scheme of the smartphone functions is proposed. Smartphone-centric architecture means that several operations, usually performed by other (dedicated) devices, are concentrated in a unique system that is the smartphone. In more technical detail, the smartphone has four macroscopic functions: *i) Interacting*, *ii) Sensing*, *iii) Processing* and *iv) Communicating*. Each of them can be divided in several microscopic functions with more specific tasks.

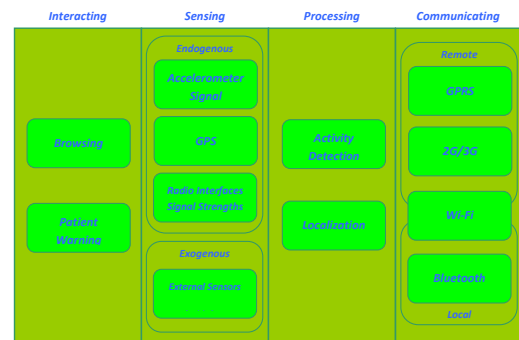


Figure 2. The Smartphone Functions

THE ACTIVITY RECOGNITION TASK

It is based on the sensing, processing and classification of data provided by the smartphone-embedded accelerometer and is designed to recognize four different classes of physical activities of HF suffering patients. The evaluated activities are: 1) **Idle**, 2) **Still**, 3) **Walking** and 4) **Running**. The first class is useful to recognize if the patient has abandoned the smartphone. The Still class represents the case in which the patient has the smartphone and he/she is spending his/her time sedentarily (i.e., he/she is sitting or standing). In case of the last two classes the patient is walking and running, respectively.

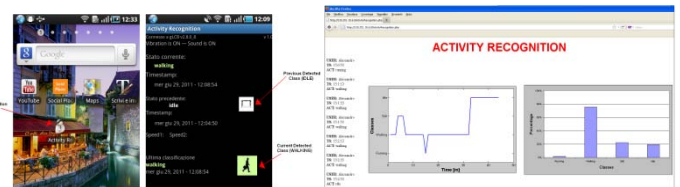


Figure 3. Detected activity view and MS web interface

ACKNOWLEDGEMENT

The authors wish to deeply thank Dr. Guido Gigli, Head of the Cardiology Department of the Hospital of Rapallo (Genoa), for his precious support and suggestions about the clinical issues considered in this poster and Dr. Pietro Jacassi of CAP Research S.p.A. for the valuable cooperation in the framework of this activity.



Synthesis and Optimization of Neuromorphic Hardware

Andrea Calimera, Alberto Macii, Enrico Macii, Massimo Poncino
Politecnico di Torino

C.so Duca degli Abruzzi, 24, 10129 Torino
{andrea.calimera; alberto.macii; enrico.macii; massimo.poncino}@polito.it

A true understating of the human brain could open the doors to a new era of discoveries and innovations. The knowledge of the dynamics which govern neural information processing may in fact help to treat many surgical diseases of the nervous system; moreover, it would laid the basis for a new generation of bio-inspired ICT systems that exhibit what we use to call intelligence.

Although we are still quite far from this visionary goal, in the last years a new branch of neuroscience, i.e., the **Neuromorphic Engineering**, is substantially helping biomedical scientists in this complicated task. Neuromorphic engineering is a concept developed in the late 1980's, describing the use of very-large-scale integration (VLSI) systems to mimic neuro-biological architectures. The activities which fall under neuromorphic engineering can be divided into main classes: **neuromorphic modeling**, and **neuromorphic computation**. Basically, the former provides the knowledge of the biological algorithms, while the latter translates the algorithms into electrical circuits. This is an iterative process, where, as more knowledge evolves yielding improved algorithms, the electrical circuits are revised and improved accordingly, thus providing new valuable feedbacks for improving the accuracy of algorithms and models. A key aspect of this closed-loop flow is the availability of a **smart** computing hardware, where the word **smart** encompasses a plurality of characteristics and special properties, such as **high efficacy of computing**, **low power consumption**, and **easy implementation flow**.

Modern **Application Specific Integrated Circuits** (ASICs) represent potential candidates for fitting neuromorphic applications into silicon. It well known that ASICs can accommodate very complex logic functions in a small silicon area, alwo providing high computation speed and large storing capacity in the same die. Moreover, the use of Electronic Design Automation (EDA) tools, which have been object of extensive research in the last decades, allows to aggressively optimize the power consumptions while reducing the Prototype Turnaround Time (PTT). Nevertheless, implementing efficient neuromorphic HW requires additional features that standard implementation flows do not support. Our long-term research goal is to adapt, and enhance when necessary, standard EDA tools to neuromorphic computing. More specifically we are focusing on the logic synthesis problem and, in particular, on the development of dedicated neuromorphic primitives that make the **Technology Mapping** algorithms more efficient in terms of power, area, and performance optimization. Preliminary results have shown that using dedicated neuro-libraries instead of classical Boolean ones, one can implement the same neuro-model with an architecture that consumes 64% less power at no cost of speed penalties.

Tecnologie per il risparmio energetico: il ruolo delle ICT

Carolina Lucchesini, Stefano Quaranta
Smart-e srl
Via Italia 54, 10093 Collegno (TO)
info@smart-e.it

La crescente domanda di energia e il costante richiamo alla riduzione dei consumi da parte delle istituzioni europee ha determinato un'attenzione sempre più rilevante sul tema dell'efficienza energetica e della riduzione dei consumi.

Per questa ragione, è decisiva la sinergia tra il mondo dell'impresa e i centri di ricerca per implementare quelle tecnologie ICT che possano determinare una trasformazione strutturale nell'uso dell'energia da parte degli utilizzatori finali (che siano singoli cittadini, industrie, terzo settore) senza pregiudicare stile di vita o esigenze produttive.

L'impiego delle ICT nel settore del risparmio energetico consente un approccio sistemico al tema, permettendo agli utenti di entrare in **connessione diretta con la totalità dei propri consumi e provvedere a una riduzione degli stessi**, senza interventi di grossa portata o modifiche strutturali nei processi produttivi o negli edifici. L'ICT Energy Management costituisce, infatti, **il trait d'union tra utilities e utenti finali**, permettendo a questi ultimi di conoscere in maniera dettagliata l'andamento dei consumi della propria azienda, del proprio ufficio, del proprio macchinario secondo differenti variabili (tempo, edifici, sistemi, dispositivi, attacco/attacco carichi) e di intervenire adeguatamente.

Smart-e, società che offre soluzioni per il risparmio energetico, ha elaborato una **piattaforma di Energy Management (Smart-e Energy Toolbox)** in grado di mettere in pratica il concetto di **Internet of Energy**: Smart-e Energy Toolbox è un sistema che connette ogni dispositivo di misurazione in campo con un **software web based** in grado di ordinare i dati a seconda della variabile di analisi desiderata e generare report informativi su base temporale, a disposizione dell'utente in qualunque momento. Un sistema di monitoraggio dei consumi che ha come obiettivo primario quello di dare vita a una **rete energetica integrata** in cui sensori, dispositivi (per linee elettriche, gas, acqua, vapore, aria compressa) e composizioni di sistemi aggregati comunicano con l'utente attraverso la rete internet. Tale tecnologia ICT permette un'analisi dettagliata delle modalità di utilizzo dell'energia, per **fotografare lo stato dei consumi** di processi produttivi/edifici/aziende, individuare le principali attività energivore e stabilire gli interventi idonei per ottimizzare i consumi e **razionalizzare l'uso dell'energia**.

Considerando che la sola introduzione di energia rinnovabile non sarà in grado di soddisfare il fabbisogno energetico mondiale e la crescita dei consumi nei prossimi decenni, la cosiddetta **White Energy** costituisce l'unica soluzione in grado di contenere la crescente richiesta energetica e limitarne l'impatto a livello ambientale. In tal senso, l'innovazione nel campo delle ICT rappresenta il primo passo verso la creazione di una nuova generazione di sistemi energetici e nuove pratiche per il consumo dell'energia.

Uptu

Alvise Rossano,
At Work S.r.l.
Via Borgone 48 Bis/B, 10139 Torino TO
a.rossano@atwork.it

Social network for security and civic engagement

In everyday life, people face dangerous situations: emergencies, injuries, natural hazards, roads in poor condition and so on; while P.A. and other related entities find it difficult to provide a quick and adequate response and to communicate directly with citizens when these issues arise.

Uptu is a Social Network that allows to report and receive real time multimedia geo-located information related to safety, people, environment, animals and properties, sharing it with the community that can take concrete actions to solve them. Uptu's web and mobile (iOS and Android Apps for the 1st release) components will be launched worldwide in 6 languages and it will incorporate others as it expands geographically.

Social network offers a specific collaborative environment thought to involve P.A. and other related entities, local media, associations and no-profit organizations to develop a community working on social proposals and issues with an effective approach. Project keywords are related to "strengthen local community approach", citizen's active participation, e-democracy active citizenship but beyond e-government.

The first release is expected on Nov. 2011 after 10 months of start-up activities. From the beginning an innovative rewarding and visibility strategy will be executed to boost users' attraction and incite active participation. After reaching the expected critical mass, revenue models will be incorporated: advertising, additional value-added services for end-users and 3rd parties, serious games and services developed by key partners.

Currently the competitive environment is composed by worldwide several initiatives, most of them deriving from Mysociety project dated 2003. Uptu differentiates by offering an international collaborative environment where everyone is encouraged to be an active user: creating and making part of groups, sharing thoughts, solving issues, gaining a responsible status, etc. The real time factor and integration that Uptu will offer to partners are relevant features that allow the community to create a quicker response to social problems.

Uptu project aims to create an integrated model for sharing information, services and participation and a place to enable social innovation.