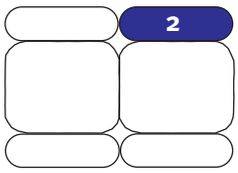
IL COMPUTER TRA COMPLESSITÀ E NARRAZIONE

Giuseppe O. Longo

La tesi esposta in questo articolo è che gli uomini sono creature della narrazione: ciascuno di noi racconta e si fa raccontare un seguito di storie nel tentativo di dare un senso al mondo e alla sua presenza nel mondo. A ben guardare, l'arte, il mito, la filosofia e la stessa scienza sono tutte forme di narrazione. Anche il computer, ultimo arrivato sulla scena, si inquadra in questo contesto: quali storie ci racconta questa macchina straordinaria? [1]



1. INTRODUZIONE

I nostri sensi sono fatti per tener fuori il mondo

Gregory Bateson

L'uomo, come specie e come individuo, si trova proiettato in un mondo vasto e rimbombante, dal quale rischia di essere sopraffatto per la ricchezza eccessiva degli stimoli e delle sollecitazioni. Per sopravvivere ha adottato una strategia che ha segnato tutto lo sviluppo della civiltà: ha costruito modelli semplificati del mondo. Ricorrendo a forme diverse di narrazione - il mito, la poesia, il racconto, la filosofia, la scienza - l'uomo ha costruito mondi (o meglio modelli del mondo) semplificati, dove abitare in modo più confortevole e meno precario. Spesso accade che questi modelli siano presi per il mondo vero: tale è la forza di suggestione della parola, del pensiero, della logica, della formalizzazione matematica, della riflessione filosofica, insomma della narrazione, che queste realtà secondarie, fatte di descrizione e interpretazione, si sono interposte come un velo tra noi e il mondo, occultandolo e sostituendolo. Ma, prima ancora, è il nostro apparato sensoriale che filtra la realtà e ce ne restituisce un'immagine distorta e impoverita, ricostruita. Su questa immagine, poi, noi lavoriamo, estrapolando e interpolando, per tentare di conoscere il mondo e ciò che esso contiene.

Le narrazioni non ci forniscono soltanto modelli significativi della realtà. Con esse infatti ci costruiamo anche un'immagine del nostro sé: impresa interminabile, perché il sé è mutevole e ambiguo, molteplice e sfuggente. Interminabile e tuttavia necessaria, l'impresa, perché ciascuno di noi deve offrire a sé e agli altri un'immagine unitaria, quell'immagine che si riassume nel pronome personale "io", protagonista dei nostri ricordi e dei nostri progetti. Naturalmente l'immagine del sé e l'immagine del mondo sono strettamente intrecciate e sono correlate agli scopi che perseguiamo.

Tra le due polarità, il mondo e l'io, che pure

[1] La tesi presentata nell'articolo è più ampiamente argomentata nel libro dell'Autore: *Il senso e la narrazione*, ed. Springer Italia, 2008.

sono indissolubili, si è via via creata una contrapposizione e l'attenzione dei poeti e dei filosofi si è concentrata ora sul vasto universo ora sull'interiorità dell'essere umano. La prima forma di attenzione ha dato luogo alla scienza e la seconda alla letteratura. Entrambe conservano i segni della loro comune origine narrativa: anche la scienza è fatta di racconti, benché si sia creata un linguaggio suo proprio. Scienza e narrazione forniscono entrambe, con stili e intenti diversi, immagini e interpretazioni della realtà. In questo senso sono complementari e nessuna delle due può rivendicare il primato sull'altra.
L'assiduo impegno che ci dovrebbe avvicinare alla conoscenza è sostenuto dall'uso di strumenti sempre più raffinati: epistemologia e tecnologia si legano in modo inestricabile. Questo intreccio è diventato ancora più stretto da quando è stata inventata quella straordinaria macchina della mente che è il *calcolatore digitale*.

2. CONOSCENZA LETTERARIA E CONOSCENZA SCIENTIFICA: L'IO E IL MONDO

Tra noi e il mondo c'è sempre un filtro creativo
Gregory Bateson

La letteratura e la scienza tentano entrambe di comprimere la lussureggiante complicatezza del mondo percepito, anche se questa riduzione è più evidente nella scienza (anzi, *nelle scienze*) e soprattutto nella fisica. Ma le origini e gli intenti comuni sono stati dimenticati e tra questi due strumenti di descrizione, interpretazione e conoscenza si è creato un conflitto, che si riassume nel titolo di un fortunato libro di Charles P. Snow, "*Le due culture*", ripubblicato di recente.
Non c'è dubbio che, dopo secoli di sudditanza, la conoscenza scientifica abbia oggi preso il sopravvento rispetto a quella umanistica. Animati da un senso di rivalsa nei confronti di letterati, artisti e filosofi, alcuni scienziati giungono perfino a ritenere che la scienza sia l'unica matrice di conoscenza e additano nel suo metodo la sola via per comprendere il mondo, fino a sconfinare, talvolta, in un'assurda e pericolosa caricatura della scienza, lo scientismo, che vorrebbe applicare il paradigma *scientifico* a tutta la realtà, compreso l'io. Ma sull'io la scienza ha poco da dirci, e qui entra in gioco il valore conoscitivo della letteratura, che è stato sostenuto, tra gli altri, proprio da un fisico teorico argentino di origine italiana, Ernesto Sabato, il quale nel 1945, a 34 anni, abbandonò una brillante carriera di ricercatore per dedicarsi al romanzo. Sabato argomenta con eloquenza che la letteratura non è né un passatempo né un'evasione, ma il modo forse più complesso e profondo di indagare la condizione umana. Trattandosi di un fisico, è interessante riportare una sua osservazione sul linguaggio della logica:

... un linguaggio che agli uomini concreti non serve. Primo perché l'esistenza non è logica e non può servirsi di simboli inequivocabili, creati per rispondere ai principi di identità e contraddizione; e poi perché l'uomo concreto non solo non si propone di comunicare verità astratte, ma ha bisogno di esprimere sentimenti ed emozioni, cercando di agire sull'anima degli altri, incitandoli alla simpatia o all'odio, all'azione o alla contemplazione.

Il linguaggio dell'uomo concreto, che è anche il linguaggio del romanzo, è assurdo e contraddittorio, ma anche efficace e potente. È un linguaggio che dice e non dice, che mostra e nasconde, mentre il linguaggio della scienza si sforza di tutto dire e di tutto illuminare, creando un paesaggio che vorrebbe essere senza ombre e seguendo una marcia orientata dall'ignoranza alla conoscenza che rifiuta a priori qualunque *ignorabimus*. Al contrario, la conoscenza fornita dal romanzo e in genere dalla letteratura non è progressiva: pur essendo gli autori inseriti in una dimensione storica e pur essendo ciascuno figlio del proprio tempo, non si può certo dire che Leopardi abbia superato Dante o che Kafka abbia superato Flaubert.
Insomma, scienza e romanzo offrono conoscenze diverse: dissimili come intenzioni e come esiti, ma *entrambe intersoggettive* e non una soggettiva, il romanzo, e l'altra oggettiva, la scienza. Una letteratura puramente soggettiva non potrebbe esistere, o meglio sarebbe incomunicabile. Allo stesso modo una scienza puramente oggettiva non potrebbe esistere perché ogni conoscenza po-

stula un soggetto che di quella conoscenza sia titolare e protagonista attivo, filtrante e costruttivo. Inoltre anche la scienza va *comunicata* ai singoli, adattandosi alle loro capacità e limitazioni.
Nel libro "*Il senso e la narrazione*" sostengo che anche la scienza è una forma di narrazione, benché col tempo abbia assunto forme specifiche. È una narrazione il cui esito conoscitivo è profondo e rigoroso, ma affilato e circoscritto. D'altra parte la letteratura fornisce sì conoscenze, ma non tanto sistematiche e formali da costituire la base per previsioni certe del comportamento umano. Entrambe le impostazioni sono legittime, nessuna esaurisce la complessità dell'io nel mondo, ma ciascuna fornisce qualche verità su di esso. Ed è forse l'insufficienza di entrambe queste narrazioni che ci fa sperare in una lunga vita sia della scienza sia del romanzo.

3. UN COLPO DI MANOVELLA

Il genio paga sempre per le sue doti
Henry James

La spinta alla conoscenza del mondo ha utilizzato strumenti capaci di farci risparmiare tempo e fatica: strumenti "economici" basati sull'apprendimento di tecniche particolari, sulla ripetizione di procedimenti, sulla meccanizzazione di certe operazioni. Così, da un lato, si è costituito un patrimonio di metodi conoscitivi che può essere trasmesso da una generazione all'altra, estendendo a tutti la possibilità di fornire contributi e, dall'altro, si è tolto il monopolio della scoperta alle persone dotate di capacità eccezionali, magari trasferendolo alle macchine: esorcizzando il genio e meccanizzando il pensiero, si è potuto accelerare il tentativo di sistemazione razionalcomputante del mondo.
Questa tendenza, che si manifesta pienamente nell'età moderna e caratterizza soprattutto la contemporaneità, è sfociata nell'impresa *dell'intelligenza artificiale forte*, la quale vorrebbe estendere il metodo scientifico, basato sul riduzionismo, fino a comprendere il pensiero dell'uomo: suggestionati dalla potenza (letterale e metaforica) del computer, alcuni hanno ipotizzato che il pensiero fosse riconducibile all'esecuzione di algoritmi. Trasferendo poi questi algoritmi su un supporto diverso dal cervello, il calcolatore digitale, si sarebbe attuato il sogno della macchina pensante. Ma su questa strada si sono incontrati ostacoli insormontabili, che derivano soprattutto dalla pretesa dell'intelligenza artificiale di ignorare la dimensione corporea dell'intelligenza umana e quindi la sua interazione dinamica con il mondo.
L'intelligenza artificiale funzionalistica si basa su un riduzionismo mentalista secondo il quale ciò che conta è lo scheletro logico, non la carne del supporto o i panni dei contenuti. Le cose in realtà non stanno proprio così: la struttura logica non è tutto, e il supporto materiale ha un'importanza straordinaria, perché la sua struttura fisica interagisce in maniera inestricabile con la funzione e la modifica (per esempio, come accade nel computer, introducendo ritardi temporali e trasformando i rapporti logici in rapporti di causa-effetto). Inoltre, per quanto riguarda gli esseri umani, i contenuti influiscono in modo determinante sul modo di ragionare e sull'efficacia e rapidità del ragionamento, e i contenuti hanno a che fare con la struttura, il corpo, l'ambiente e la comunicazione. La stessa struttura logica può far da supporto a contenuti diversissimi, ma se il contenuto è fami-

liare o gradevole il soggetto ragiona in modo rapido e felice, se il contenuto è ostico o troppo astratto, il soggetto s'impunta e si avvilisce. Ciò indica che la rappresentazione (narrazione) che del mondo fornisce il computer non è adeguata alle esigenze degli umani.
Il sogno di meccanizzare il pensiero e, soprattutto, di esorcizzare le misteriose e inquietanti capacità del genio ha accompagnato tutto lo sviluppo dell'età moderna: si tratta di smontare i riposti ingranaggi della mente superiore per consentirne la riproduzione artificiale illimitata (del resto questo smascheramento, questo smontaggio, questa scomposizione riduzionistica è ciò che si propone la scienza in tutti i settori della ricerca). Così, sventrato il giocattolo e rivelato il meccanismo che fa muovere il genio, tutti potranno imitarlo.
Per rendersi conto della pervasività di questo sogno, basta pensare ai tentativi di Raimondo Lullo, Giulio Camillo Delminio, Leibniz e tanti altri di costruire dispositivi, mentali o materiali, capaci di affiancarsi alla memoria individuale e di agevolare i ragionamenti, di manifestare insomma capacità cognitive. Anche la storia dei dispositivi da calcolo si può interpretare in questa prospettiva: che cosa sono in fondo le macchine di Pascal, di Leibniz e, per menzionare le più ambiziose, la *Macchina Analitica* di Charles Babbage e la *Macchina di Turing*, se non estroflessioni cognitive, più o meno raffinate ma sempre di natura automatica, tendenzialmente capaci di fornirci con un colpo di manovella tutte le proposizioni "vere", tutti i risultati "esatti", tutti i teoremi "dimostrabili"? La stessa *geometria analitica* di Cartesio è una protesi mentale che, grazie a ricette meccaniche, consente anche ai meno dotati di dimostrare i teoremi più ardui della geometria, che altrimenti richiederebbero immaginazione, intuito e talento.
La macchina mentale che, con un colpo di manovella, dovrebbe generare tutte le proposizioni vere eredita il carattere sintattico della logica, cioè prescinde dal significato delle proposizioni, dal loro legame con il mondo, dal loro *senso* per gli umani. Così, abbandonata alla sintassi, essa provoca un'alluvione dalla quale ci si può salvare solo scartando le proposizioni insignificanti e banali e conservando quelle interessanti e significative. La discriminazione, tuttavia, può essere compiuta solo in base a quei criteri che si erano voluti evitare ricorrendo alla macchina mentale: criteri soggettivi, perché "insignificante" e "significativo" sono sempre relativi a un individuo particolare, alle sue caratteristiche, alla sua storia, alla sua intuizione e capacità, ai suoi interessi. Il cieco automatismo della macchina non consente di adeguare le proposizioni generate ai *contenuti* della vita e dell'esperienza, quindi la complessità della persona e lo spettro del genio, cacciati dalla porta, rientrano dalla finestra.
È proprio qui che, a prescindere dai teoremi di limitazione alla Gödel, la narrazione del mondo offerta dal computer deve fare i conti con le narrazioni degli uomini. Le proposizioni vere, infatti, non sono sempre *significative*. E ciò che è significativo per Tizio può non esserlo per Caio. Insomma, il criterio che guida gli uomini nella vasta e intricata foresta delle proposizioni (o dei fatti descritti dalle proposizioni) è quello del *senso*, e la sua amministrazione non può essere affidata a una macchina (e neppure a un altro essere umano): deve essere assunto in prima persona, da cui l'importanza dell'"io". Quindi non è solo il genio, ma è ciascun individuo che rientra dalla finestra per riprendere il suo posto centrale di deputato alla scelta e allo smistamento delle proposizioni.

4. LA RICOSTRUZIONE RAZIONALE DEL MONDO

Se sapessi dirlo non avrei bisogno di danzarlo
Isadora Duncan

La tendenza a vedere nel funzionamento del pensiero l'azione di meccanismi precisi, di cui si può costruire una replica esterna, materiata in un dispositivo macchinico (un computer) da cui sporge la famosa manovella, trova un riscontro altrettanto significativo nella convinzione che il mondo tutto sia una macchina che funzioni in base a leggi rigorose, invarianti e universali esprimibili in forma matematica. La capacità (che qualcuno ha definito "irragionevole") della nostra matematica di descrivere la realtà fisica ha dato a lungo l'impressione che il formalismo avesse un valore *ontologico*: le equazioni non sarebbero solo un mezzo

stenografico e allusivo, sarebbero proprio le regole cui i fenomeni *in sé* obbediscono. Di recente questa pretesa metafisica forte si è attenuata, e la matematica è stata ricondotta a funzioni meno colorite di ontologia e più pratiche: con essa costruiamo modelli utili e adeguati ai nostri scopi, ma la "realtà" rimane in una zona irraggiungibile.
Inoltre si ha la sensazione che la nostra matematica sia un prodotto storico e contingente, che affonda le sue radici in un processo evolutivo fisico, biologico e cognitivo che avrebbe potuto dare anche esiti diversi. Sono concepibili, anche se forse non ancora immaginabili, *altre* matematiche e certo quella che possediamo si dimostra singolarmente inadatta ad affrontare problemi che escano dall'ambito fisico nel quale si è raffinata.
Non è fuori luogo aprire qui una breve parentesi sul concetto di *alterità*. Il confronto con l'altro (un'altra lingua, un'altra intelligenza, un altro codice di comportamento...) non solo relativizza la nostra naturale supponenza egocentrica e ci spinge a un salutare confronto, ma - e in primo luogo - ci fa riflettere sulle *nostre* caratteristiche. È grazie al confronto con le altre lingue che possiamo riconoscere le strutture della nostra, è grazie alla costruzione di programmi che giocano a scacchi che possiamo capire meglio come giocano a scacchi gli umani, e in generale è grazie alle ricerche di intelligenza artificiale che possiamo penetrare le caratteristiche della nostra intelligenza; da ultimo, è grazie ai robot che possiamo apprezzare meglio il legame tra il corpo e le attività cognitive.
Chiusa la parentesi, chiediamoci che esito possa avere il tentativo di descrivere il mondo per via razionale, cioè con gli strumenti del formalismo fisico-matematico: tentativo condotto prima a mano e poi con strumenti sempre più raffinati, fino al calcolatore elettronico. Nonostante le speranze di molti, il tentativo presenta limiti precisi, che si possono così riassumere: la razionalità computante non riesce a dar conto del corpo, dell'inconscio, delle emozioni e quindi restituisce un'immagine del mondo gravemente monca; mantenendo il soggetto fuori del quadro, cioè separandolo dall'oggetto, il razionalismo non riesce a descrivere e interpretare ciò che sentiamo di più intimo e importante: il piacere e l'angoscia, la felicità e l'infelicità, il coinvolgimento affettivo e sentimentale, gli aspetti etici ed estetici dell'esistenza e dell'esperienza, la pertinenza di questi aspetti alla persona singola nella sua irripetibile individualità esistenziale; si perde poi la storia, l'evoluzione, la florida esuberanza del vivente, la nascita dell'informazione e del nuovo, tutto il mondo della comunicazione interpersonale, si perde il fondamentale risvolto soggettivo dell'esperienza rappresentato dalla coscienza.
Del resto anche il matematico ricorre alla logica solo nella fase di sistemazione, mentre nella fase creativa, quando insegue con ansia e passione il risultato che pare sfuggirgli, si vale di una panoplia di strumenti tutt'altro che rigorosi e consequenziali, ma carichi di semantica, di balzi intuitivi e di riferimenti metamatematici. Si presenta qui un esempio lampante di coevoluzione culturale retta da una retroazione positiva: la costruzione di un mondo artificiale logico-computante favorisce le caratteristiche logico-computanti degli umani, i quali a loro volta sono spinti ad accentuare queste caratteristiche nella costruzione dell'artificiale, e così di seguito.
La nostra razionalità ha un ambito di applicazione limitato e nella pratica attiva e conoscitiva l'uomo ricorre a robusti strumenti collaterali basati sull'interazione dialogica con gli altri. Intanto siamo animati da una forte volontà di cooperazione comunicativa, poi ci immedesimiamo nei nostri interlocutori, costruendo continue ipotesi su ciò che essi pensano: quindi la nostra mente opera in un *circolo ricorsivo* in cui è impossibile separare le immagini sue proprie da quelle

che essa attribuisce agli interlocutori. Di solito quindi pensiamo, e agiamo, non sulla base di una realtà esterna ma in base alle nostre immagini mentali di quella realtà: la realtà è sempre percepita e ri-costruita e questa ricostruzione riguarda anche ciò che gli altri pensano (o sembrano pensare), fanno o stanno per fare o hanno intenzione di fare. Una conferma fisiologica di questo carattere sociale e condiviso della nostra attività pratica, motoria, cognitiva ed emotiva è stata fornita dalla scoperta, avvenuta negli anni ottanta e novanta, dei *neuroni specchio* (Rizzolatti e Sinigaglia).
L'intelligenza umana s'immedesima nelle menti altrui cercando di rispecchiarle, pesca nel mondo e nell'evoluzione bio-fisico-culturale ed è a spettro larghissimo; all'opposto le macchine hanno prestazioni ristrettissime e sono immerse in un esile universo binario costruito *ad hoc* dal programmatore e non sanno (ammesso che qui il verbo "sapere" abbia senso) che cosa ci possa essere là fuori nel vasto mondo (e non sanno neppure di non saperlo).

5. LE SORPRESE DEL CALCOLATORE

L'uomo è il mezzo di cui un computer si serve per fare un altro computer

Detto moderno

Come ho già accennato, il pensiero umano e la comunicazione umana non si basano solo sulla struttura logica, ma anche sul contesto e sui contenuti: un problema formulato in termini che facciano riferimento alla quotidianità viene risolto in genere con molta più disinvoltura dello stesso problema formulato in termini astratti, e ciò rafforza l'impressione che gli esseri umani non siano molto portati per l'artificialità del formalismo. I linguaggi e gli strumenti astratti e acontestuali sono fragili, anche se raffinati, quelli umani sono rozzi e robusti. Il calcolatore, e anche le "macchine mentali" come la matematica e la logica, sono protesi che suppliscono alla debolezza umana in un settore nel quale ce la caviamo piuttosto male anche perché, in fondo, finora non ha avuto un grande valore di sopravvivenza: comincia ad averlo ora, nel mondo fortemente artificiale che stiamo costruendo. Per converso, queste macchine sono lontane dalla complessa e robusta attività della nostra mente e del mondo che ci circonda e che ci sta a cuore. Se il calcolatore ha un'intelligenza, essa è davvero *artificiale*. Per apprezzare il divario tra la mente naturale e la sua ricostruzione logico-simbolica, basta pensare alle difficoltà che suscita la nozione di "macchina intelligente".
Sul piano pratico, vorrei sottolineare una conseguenza forse inaspettata, ma importantissima, dell'impiego del calcolatore. Costruita per razionalizzare, ordinare e conferire precisione a un mondo confuso e approssimativo, questa macchina ha avuto (anche) un effetto in certo qual modo contrario: consentendo di andare a fondo nei particolari esecutivi delle procedure e di incrementare l'accuratezza dei calcoli, ha messo in luce quelle che sono le inevitabili imprecisioni e vaghezze della realtà conoscibile. Il calcolatore rende complessa - anziché semplificare - la descrizione della realtà e questo suo effetto inatteso lo rende molto diverso dalla macchina matematica tradizionale, la cui stessa fondazione, cioè la costruzione dei numeri, si compie mediante un'*astrazione semplificativa*. In altre parole, il passaggio dagli eventi e dagli oggetti al numero avviene grazie a un sacrificio: alcune differenze (cioè informazioni), anzi quasi tutte le differenze, vengono soppresse e solo alcune vengono lasciate sopravvivere. La regola così cara agli insegnanti per cui si possono sommare solo elementi *omogenei* (anzi la nozione stessa di omogeneità) deriva dall'introduzione di una *relazione di equivalenza* basata sull'eliminazione irreversibile di certe differenze. Alla radice della matematica vi è dunque un degrado irrecuperabile di informazioni o di qualità: *il numero è l'antitesi dell'informazione*.
Del resto anche la fisica galileana si basa sulla soppressione delle cosiddette qualità secondarie, cioè su una forte riduzione dell'informazione. La riduzione non è arbitraria (come non lo è in matematica), anzi mira a ottenere una formulazione rigorosa di una sorta di proto-teoria in parte prefigurata e intuita, anche se in forma vaga. Osserviamo esplicitamente che i due procedimenti di riduzio-

ne e semplificazione in matematica e in fisica, ben armonizzati tra loro, sono alla base del loro accordo, mirabile anche se parziale, che tanto stupisce gli studiosi. In altre parole, la descrizione fisica dei fenomeni viene semplificata quel tanto che consente alla matematica, anch'essa semplificata al punto giusto (di fatto, spesso linearizzata), di darne una descrizione plausibile. Invece il calcolatore ci consente di guardare più a fondo, di tener conto di alcuni termini non lineari, e così il quadro diventa più complesso: vediamo letteralmente un altro mondo. È straordinario che un dispositivo tecnico abbia conseguenze epistemologiche tali da cambiare la nostra visione della realtà! In questa prospettiva, il computer è un *rivelatore e amplificatore di complessità*.
Questa proprietà del computer apre una prospettiva inattesa: la progressiva complessificazione della narrazione offerta da questa macchina potrebbe un giorno, forse, aprirsi al senso e incontrarsi con le nostre narrazioni.

6. COMPUTER E NARRAZIONE

Il faticoso ed esaltante cammino della scienza occidentale può essere interpretato come un lungo tentativo di tradurre in conoscenza alta, razionale ed esplicita la massa delle conoscenze implicite (nel mondo e in particolare nel corpo). Ma questo tentativo incappa nell'ostacolo tipico di ogni processo di traduzione, cioè l'*incompletezza*. Rimane pur sempre un residuo intraducibile, una cicatrice insanabile, una macula ostinata che ci ricorda come la traduzione sia un'impresa impossibile, perché vorrebbe o dovrebbe essere un'applicazione totale del mondo su sé stesso. Ma nonostante questo limite intrinseco non possiamo fare a meno di praticare la scienza.
Tentiamo di tradurre nella nostra lingua la lingua del mondo, che però non conosciamo: cominciamo a dubitare che la natura sia, come sosteneva Galileo, un libro "scritto" in termini comprensibili e decodificabili dalla nostra scienza, cioè in caratteri matematici (e poi, quali caratteri: i triangoli o i frattali?). La scienza è solo una nostra interpretazione. La traduzione letteraria, che è certamente più facile perché vuole trasporre un testo da una lingua naturale a un'altra, rende manifesto che la fedeltà è impossibile. Ogni traduzione alla fin fine si rivela un'interpretazione, con tutte le limitazioni intrinseche dell'interpretazione, prima fra tutte quella di non essere mai "vera", unica e definitiva. L'interpretazione è sempre rivedibile, perfettibile, modificabile, storica. Questa consapevolezza ha con-

tribuito, tra l'altro, a un indebolimento sostanziale del concetto di *verità*. In questa prospettiva, il computer può aiutarci a raffinare sempre più le nostre interpretazioni, fornendoci storie sempre più significative.

Bibliografia

[1] Bateson Gregory: *Verso un'ecologia della mente*. Adelphi, Milano, 1976, 2.a ed. 2000.

[2] Gàbici Franco: *Gadda. Il dolore della cognizione*. Simonelli Editore, Milano, 2002.

[3] Kundera Milan: *L'arte del romanzo*. Adelphi, Milano, 1988.

[4] Licata Ignazio: *La logica aperta della mente*. Codice Edizioni, Torino, 2008.

[5] Longo Giuseppe O.: *Homo technologicus*. Meltemi, Roma, 2001, 2.a ed. 2005.

[6] Longo Giuseppe O.: *Il senso e la narrazione*. Springer Italia, Milano, 2008.

[7] Longo Giuseppe O.: Il test di Turing. *Mondo Digitale*, di prossima pubblicazione.

[8] Rizzolatti Giacomo, Sinigaglia Corrado: *So quel che fai: il cervello che agisce e i neuroni specchio*. Raffaello Cortina Editore, Milano, 2006.

[9] Sabato E.: *Lo scrittore e i suoi fantasmi*. Meltemi, Roma, 2000.

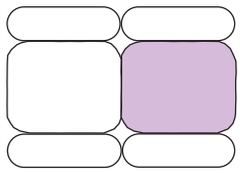
[10] Snow C.P.: *Le due culture*. A cura di A. Lanni, Marsilio, Venezia, 2005.

Giuseppe O. Longo è ordinario di Teoria dell'informazione nella Facoltà d'Ingegneria dell'Università di Trieste. Si occupa di codifica di sorgente e di codici algebrici. Ha diretto il settore "Linguaggi" del Laboratorio della "International School for Advanced Studies" (Sissa) di Trieste e il Dipartimento di Informazione del "Centre Internationale des Sciences Mécaniques" (Cism) di Udine. Socio di vari Istituti e Accademie, si interessa di epistemologia, di intelligenza artificiale e del rapporto uomo-tecnologia. È traduttore, collabora con il Corriere della Sera, con Avvenire e con numerose riviste. È autore di romanzi, racconti e opere teatrali tradotti in molte lingue.
E-mail: longo@univ.trieste.it

KNOWLEDGE MANAGEMENT
L'EREDITÀ DEI SISTEMI ESPERTI

Stefania Bandini

L'esperienza dei Sistemi Esperti, che ha caratterizzato negli anni '80 un approccio pragmatico alla rappresentazione e al trattamento della conoscenza, trova oggi un fecondo spazio di recupero, offrendo alla prospettiva computazionale del Knowledge Management un insieme di metodologie, strumenti e soluzioni atte a incorporare le competenze che supportano l'innovazione. Il lavoro presenta alcune riflessioni sulle nuove prospettive dei Sistemi Esperti e sulla loro evoluzione in una realtà organizzativa e distribuita.



1. INTRODUZIONE

Tra le parole-chiave che caratterizzano le attuali tendenze dello sviluppo di molte applicazioni dell'informatica, spicca il termine *conoscenza* (*knowledge*): basi di dati opportunamente organizzate spesso sono definite, nella letteratura, come *basi di conoscenza*; sistemi di archiviazione e reperimento di documenti, dati e informazioni inerenti alle organizzazioni sempre più spesso vengono riferiti alle attività di *gestione della conoscenza* (*knowledge management*); i sistemi di supporto alla formazione e all'apprendimento, distribuiti, focalizzati all'aggregazione di competenze specifiche e rivolti a *comunità* operative e in continua evoluzione fanno riferimento alla crescita del *patrimonio conoscitivo* condiviso e fruibile grazie alle tecnologie informatiche e telematiche. Questo è solo uno scorcio del panorama dell'automazione che, in modo sempre più pervasivo, traccia il futuro dell'evoluzione del modo in cui la produzione e la fruizione della conoscenza caratterizzerà le nuove generazioni.

Nella storia dell'informatica e delle sue molteplici sfaccettature (teoriche e pratiche) gli anni '70 e '80 hanno visto configurarsi e crescere attitudini di progettazione di sistemi che, precursori di queste nuove imprescindibili tendenze, hanno iniziato a prendere in considerazione la possibilità di rappresentare e trattare computazionalmente la conoscenza. Ci riferiamo a un periodo storico, inaugurato negli Stati Uniti ma immediatamente recepito sul fronte internazionale, che ha visto la nascita e la crescita dei *Sistemi Esperti*, sistemi software in grado di contenere specifiche conoscenze relative a un dominio conoscitivo circoscritto e di formulare soluzioni a problemi inerenti quel dominio. Questa classe di sistemi è stata sviluppata in seno alla parte applicativa dell'Intelligenza Artificiale: più precisamente, hanno rappresentato una sorta di "rilancio" di questa disciplina dopo il vanificarsi di ambiziose aspettative dichiarate durante la sua fondazione (per esempio, la messa a punto di sistemi inferenziali generici in grado di competere con le prestazioni umane).

I Sistemi Esperti, rivolgendosi a domini circo-

scritti del sapere pratico (diagnostica medica, configurazione di calcolatori, prospezione mineraria, solo per citare i più noti e documentati esempi) hanno rivolto la loro attenzione soprattutto al tema della *rappresentazione della conoscenza* (disciplina con rigorosi fondamenti formali che ancora oggi è uno dei cardini dell'Intelligenza Artificiale), ponendosi domande e fornendo prototipi di soluzioni pratiche di cui ancora oggi possiamo fruire e a cui possiamo attingere per sviluppare le nuove tecnologie di supporto alla società della conoscenza.
Scopo di questo lavoro è di tracciare alcune prospettive in cui l'esperienza accumulata nell'ambito dei Sistemi Esperti può influenzare lo sviluppo delle soluzioni computazionali del *Knowledge Management*, che rappresenta l'aspetto più peculiare e pratico con cui oggi le tecnologie informatiche e telematiche supportano la tesaurizzazione condivisa e la fruizione distribuita per la creazione e la crescita della conoscenza.

1.1. Una base teorica: Knowledge Economy

In una sintesi ben articolata uscita nella raccolta di saggi "*Strategic Learning in Knowledge Economy: Individual, Collective and Organizational Learning Process*" [18], C.K. Prahalad e G. Hamel hanno fornito una chiave interpretativa molto attuale per alcune delle molteplici facce con cui il tema della gestione della conoscenza (KM - *Knowledge Management*) si articola e si sviluppa a fronte dei cambiamenti indotti dall'evoluzione tecnologica e produttiva.
Il cuore del lavoro "*The Core Competence of the Corporation*" di Prahalad e Hamel si innesta su un modello di crescita delle potenzialità di un'impresa basato sull'investimento nella valorizzazione e patrimonializzazione della conoscenza, al fine di rendere dinamico e interattivo il processo d'induzione di bisogni per creare nuove partizioni complesse di mercati, mettendo in risalto la natura dinamica del *Core Knowledge*.
La chiave di volta di questo processo sta nelle capacità che una moderna impresa possiede di proporre prodotti sempre nuovi, in modo da indurre nuovi bisogni e, conseguentemente, creare dinamicamente nuove aree di mercato: nel saggio l'attenzione è posta sul ruolo delle conoscenze specifiche che supportano il core business e permettono l'innesco dei processi d'*innovazione* di prodotto.
In quest'ottica, la conoscenza gioca un ruolo decisivo nel disegno delle strategie di incremento del proprio mercato: da una strategia di conquista di territori, alla creazione vera e propria di nuovi territori, con la proposta di continue innovazioni a supporto delle proprie caratteristiche di competizione.
Per quanto riguarda il primo approccio strategico, gli strumenti classici di tipo statistico-previsionale (con il loro relativo apparato computazionale) offrono modelli di tipo quantitativo molto potenti. Per il secondo, la sfida è quella di doversi creare (in corsa) nuovi strumenti e modelli di tipo qualitativo, quantitativo, concettuale, metodologico e computazionale in grado di rappresentare e dominare una realtà complessa e dinamica dovuta alle nuove condizioni in cui il mercato globale evolve, e in cui la conoscenza, la sua rappresentazione, la sua gestione e la sua trasmissione giocano il ruolo di protagoniste. Il focus sulla conoscenza come valore intangibile da patrimonializzare caratterizza il passaggio dalle tendenze nelle scienze dell'organizzazione mirate all'efficienza (tipiche degli anni '80) a quelle mirate all'efficacia (tipiche degli anni '90).
Dal punto di vista del supporto da parte delle moderne tecnologie informatiche a queste due tendenze, lo sviluppo di nuovi paradigmi computazionali e la diffusione delle reti telematiche hanno permesso di definire dei punti di svolta rivoluzionari, assumendo addirittura un ruolo contemporaneo di traino e di pressione nel disegno e nell'implementazione di nuovi scenari organizzativi per le aziende proiettate nel nuovo millennio.
In un modello organizzativo indirizzato a rendere sempre più competitive le aziende in un momento storico caratterizzato dalle turbolenze della globalizzazione e dalla sua realtà resa possibile proprio grazie allo sviluppo delle tecnologie informatiche, il ruolo di sistemi in grado di rappresentare e gestire la conoscenza (e a garantire in termini di crescente qualità la sua messa in azione per la creazione della spinta innovativa) diventa di primo piano.
Se questo può valere per tutti gli strumenti che il Knowledge Management ha reso disponibili (sistemi orientati al lavoro collabo-

rativo, sistemi dedicati alla rappresentazione e strutturazione della conoscenza, sistemi eterogenei integrati di supporto alla decisione [7]), il punto centrale del modello proposto dai due prestigiosi autori (*Core Knowledge*) rievoca molti degli sforzi, di ricerca e applicativi, che negli anni '80 hanno caratterizzato quella classe di sistemi conosciuti come *Sistemi Esperti*, nella variegata compagine delle proposte e delle esperienze che hanno fatto da corollario alla loro evoluzione.
In modo semplicistico, possiamo annoverare questi sistemi come quelli dedicati alla rappresentazione e strutturazione della conoscenza, ma cerchiamo di approfondire meglio in che modo i Sistemi Esperti possono essere intesi in una prospettiva di Knowledge Management.
Torniamo a Prahalad e Hamel: nell'ottica di supportare la gestione del *Core Knowledge* (quella conoscenza che caratterizza l'oggetto del core business di un'azienda – il suo prodotto - e che, attraverso i suoi possessori integrati in una realtà organizzativa che ne riconosce e supporta il ruolo, permette la creazione delle innovazioni necessarie per rimanere sul mercato o dare impulso alla nascita di nuovi mercati), vanno tenuti in conto alcuni fondamentali fattori. Si tratta di avere a disposizione strumenti concettuali, metodologici e computazionali in grado di:
- ❑ permettere una fase di analisi mirata, metodologicamente strutturata e fortemente adattabile al contesto conoscitivo e sociale;
- ❑ mettere a disposizione metodi e tecniche per l'individuazione, il trasferimento e l'acquisizione delle conoscenze interne di valore per il core business (*Core Competence*);
- ❑ permettere di catturare aspetti non strutturati, qualitativi e ad alta combinatorietà delle conoscenze coinvolte nella fase creativa di risoluzione di problemi;
- ❑ fornire strumenti metodologici, concettuali e formali per la rappresentazione di un insieme coordinato e dinamico di conoscenze focalizzate alla soluzione di problemi complessi;
- ❑ fornire una gamma di strumenti computazionali creati per il trattamento della conoscenza e facilmente integrabili con tutte le componenti di un sistema di Knowledge Management.

Molti (e altri) di questi strumenti trovano risonanza nel "parco attrezzi" che l'area dei Sistemi Basati sulla Conoscenza (a cui i Sistemi Esperti appartengono [13]) ha sviluppato e affinato dalla loro nascita (inizi degli anni '80) fino ai giorni nostri. Molti di questi strumenti, più in generale, sono naturalmente "percolati" in ambiti più tradizionali dell'informatica, così come è stato per molti strumenti nati nell'ambito dell'Intelligenza Artificiale, e hanno influenzato il consolidamento di molti approcci computazionali al trattamento della conoscenza di nuova generazione. Questi fattori sono spesso, però, adombrati dalle critiche che, a partire dagli anni '90, sono state rivolte a questa classe di sistemi, che possiamo qui grossolanamente sintetizzare in questi punti:
- ❑ mancanza di un approccio sistematico e metodologicamente formalizzato dell'Ingegneria della Conoscenza;
- ❑ onerosità della fase di acquisizione delle conoscenze esperte;
- ❑ "faziosità" dell'approccio verso strumenti computazionali tipici dell'Intelligenza Artificiale;
- ❑ manutenzione e aggiornamento delle basi di conoscenza (per una trattazione più specifica di questo tema rimandiamo alla nota [3]);
- ❑ isolamento: le conoscenze contenute in un Sistema Esperto avevano il compito di supportare o sostituire "campioni" di conoscenze esperte, senza tenere conto del contesto sociale in cui la conoscenza si sviluppa, evolve e viene fruita.

2. CONOSCENZA IMPLICITA ED ESPLICITA

È nota l'enfasi che nei Sistemi Esperti (e in molti approcci del Knowledge Management [17]) viene data ai problemi di acquisizione, rappresentazione e trattamento delle conosciute tacite ed esplicite. Sintetizzando da [12] possiamo riprendere queste definizioni:
Conoscenza esplicita: espressa in modo simbolico, facilmente comunicabile attraverso il linguaggio parlato o scritto, quindi tramite documenti, testi, manuali, lezioni, corsi.
Conoscenza implicita (*o tacita*): conoscenza nascosta nei pattern concettuali di un individuo, spesso usata in modo automatico, difficilmente traducibile in informazioni e quindi difficilmente trasmissibile con la comunica-

Alcune definizioni di Ontologia nell'ambito della rappresentazione della conoscenza

An Ontology is an explicit specification of a conceptualization.
T.R. Gruber, *A translation approach to portable ontology specification*, Knowledge Acquisition, 5, 1993.

An Ontology defines the basic terms and relations comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary.
R. Neches, R.E. Fikes, T. Finin, T.R. Gruber, R. Patil, T. Senator, W.R. Swartout, *Enabling technologies for knowledge sharing*, AI Magazine, 12, 3, 1991.

An Ontology provides the means for describing explicitly the conceptualization behind the knowledge represented in a knowledge base.
S. Bernaras, I. Laresgoiti, J. Corera, *Building and reusing ontologies for electrical network application*, Proceedings 12th European Conference on Artificial Intelligence, 1996.

An Ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e., its ontological commitment to a particular conceptualisation of the world. The intended models of a logical language using such vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualisation) by approximating these intended models.
N. Guarino (ed.), *Formal ontologies in information systems*, IOS Press, 1998.

zione, ma, piuttosto acquisibile attraverso l'imitazione, è strettamente correlata con la pratica, l'esperienza e l'azione.
Osservazione, insegnamento pratico, tradizione, esperienza in comune, sono i mezzi, con cui si trasmette la conoscenza tacita. Un principiante possiede usualmente molta conoscenza esplicita acquisita con lo studio e scarsa conoscenza tacita che non essendo trasferibile con la comunicazione è trasmissibile solo con l'esperienza diretta che solo un esperto possiede.
Questo tipo di analisi è oggi molto enfatizzata nel Knowledge Management, ma le sue radici teoriche e applicative si trovano proprio nell'area operativa dei Sistemi Esperti. Avendo a che vedere con conoscenze specialistiche, chi operava in questo settore era perfettamente consapevole del ruolo fondamentale (e del suo valore) della conoscenza tacita, non esplicitamente formalizzata nei protocolli operativi, posseduta dagli esperti e necessaria alla soluzione di problemi chiave nell'operatività. L'area dei Sistemi Esperti ha enfatizzato gran parte della sua attività nell'acquisizione, nella rappresentazione e nel trattamento computazionale di questo tipo di conoscenza, spesso anche a scapito della sua coniugazione con i fattori fondamentali racchiusi in quella che viene definita come conoscenza esplicita (formalmente strutturata e organizzata nelle procedure e nei protocolli operativi). L'*Ingegneria della Conoscenza* si è infatti sviluppata proprio per l'analisi di questo tipo di competenze, sviluppando strumenti metodologici e computazionali per l'estrazione e la rappresentazione della conoscenza.
Per le nuove generazioni di sistemi di Knowledge Management che devono trattare anche con il *Core Knowledge* (oltre ad altre forme in cui la conoscenza di organizza ed evolve) è necessaria un'armonizzazione bilanciata sulla rappresentazione di ambedue questi aspetti della conoscenza.
Il filone di studio che ha sviluppato l'aspetto della conoscenza esplicita, come conoscenza strutturata e condivisa da una comunità che possiede un ricco patrimonio lessicale e concettuale comune è noto con il termine "rappresentazione basata su ontologie" (si veda le definizioni nel riquadro) [20, 19]. Il termine ha radici nella filosofia greca classica, e una parte della metafisica ne ha determinato le varie declinazioni di significato. L'interpretazione moderna di una parte degli studi filosofici permette di scegliere questo approccio per almeno tre motivazioni generali:
- ❑ permettere il riuso di domini di conoscenza strutturata e fruibile e testata (in poche parole, resa esplicita), al fine di rendere più efficace la ricerca di soluzioni nuove e di introdurre standard che garantiscano l'interoperabilità;
- ❑ specificare assunzioni esplicite sul dominio e, contemporaneamente, facilitare i cambiamenti sulle assunzioni di dominio e rendere fruibile la comprensione e l'aggiornamento dei dati esistenti;
- ❑ permettere l'accordo sul significato del vocabolario utilizzato per la conoscenza condivisa.

Le fasi di concettualizzazione e rappresentazione presuppongono la definizione formale di un dominio di conoscenza (definizione di classi, proprietà, assiomi, individui) per giungere a una descrizione strutturata gerarchicamente dei concetti fondanti del dominio stesso e delle loro proprietà che trovi il consenso di diversi attori interessati a condividerla e utilizzarla. Strumenti atti a questo sono già utilizzabili e a supporto della fase di acquisizione e strutturazione (vedi il caso di "*Protégé: Onto-*

logy Editor and Knowledge Acquisition System" http://protege.stanford.edu/).
I metodi di acquisizione, strutturazione e rappresentazione delle ontologie hanno solo in parte attinto alle metodologie dell'Ingegneria della Conoscenza, che pur fornendo strumenti stabilizzati, fa ancora parte del corredo formativo dei programmi di Intelligenza Artificiale. Nella tradizione classica dell'ingegneria del software, questa parte riguarda la fase di analisi e stesura di specifiche, e le più recenti riflessioni sul tema dell'integrazione tra Ingegneria della Conoscenza (nella sua specifica fase di acquisizione) e metodi di analisi producono modelli che evolveranno verso metodologie per lo sviluppo di complessi sistemi per il trattamento dell'informazione e della conoscenza. In questo modo, il ruolo dell'ingegnere della conoscenza potrà fruire di strumenti di supporto che siano in grado di sgravare la verticalità della fase di acquisizione di competenze da un unico esperto, ma di fruire di tutti gli aspetti di crescita e di *problem solving* condiviso che caratterizzano le organizzazioni ad alto impatto innovativo.
Infine, va menzionato che gli aspetti intrinsecamente tassonomici che molte delle rappresentazioni ontologiche possiedono ci riportano ancora a ricordare i Sistemi Esperti della generazione di Internist (un sistema Esperto dedicato alla diagnostica nella medicina interna, che ha un'esplicita rappresentazione tassonomica di conoscenze strutturate in apparati e organi a cui associare sintomatologie [16]) e alla sua struttura concettuale, precursore dell'esigenza di incorporare conoscenze strutturalmente articolate e dinamiche. Se invece ci focalizziamo sugli strumenti di rappresentazione, ci basta solo accennare al ruolo che i *frame-systems* hanno storicamente giocato nell'evoluzione dei paradigmi di programmazione e nell'articolare le rappresentazioni basate su ontologie.

2.1. Nuovi ruoli per l'Ingegneria della Conoscenza

I Sistemi Esperti si sono distinti per il massiccio impiego di strutture per la rappresentazione e il trattamento della conoscenza basate su Regole (costrutti della forma associativa "se-allora" che permettono di rappresentare conoscenze non strutturate, e che direttamente permettono delle connessioni inferenziali tra premesse <<se>> e conclusioni <<allora>>, senza passare attraverso l'esplicita struttura concettuale che le presuppone), il cui scopo era quello di puntare direttamente sulla possibilità di incorporare in un modello computazionale le *conoscenze implicite*, possedute dagli esperti di un settore molto verticalizzato.
Il passaggio dalla descrizione di un metodo applicato di *problem solving* attuato dagli esperti alla sua implementazione con questo metodo di rappresentazione, è completamente a carico dell'abilità dell'Ingegnere della Conoscenza.È con questo termine che nell'era dei Sistemi Esperti si è delineata una figura professionale specializzata in tecniche di acquisizione ed elicitazione della conoscenza. Il maggior contributo dell'area dell'Ingegneria della Conoscenza va a carico dello sviluppo della consapevolezza del bisogno di nuovi strumenti di analisi e di stesura di specifiche funzionali per nuove classi di sistemi per il trattamento e la gestione della conoscenza. Oggi, questa figura professionale, se di qualità, è "*border line*" con le discipline dell'organizzazione, aggregando competenze non solo di natura informatica, ma anche rivolte alla conoscenza delle strutture organizzative e del ruolo dell'utente finale nel disegno di complessi sistemi di Knowledge Management [21, 12]. Questo vale non solo per la realizzazione di sistemi basati su regole, ma per tutto l'ambito della creazione di sistemi informativi ad alto impatto innovativo.
Il merito principale dell'Ingegneria della Conoscenza è quello di mirare all'individuazione (anche all'interno di ambienti organizzativi) delle conoscenze di maggior valore (*Core Knowledge*), quelle che implicitamente supportano gran parte dell'attività di *problem solving*, specialmente nei casi in cui il loro maggior valore sta nelle capacità di produrre innovazioni.
Infatti, uno dei principi fondamentali su cui poggia l'Ingegneria della Conoscenza e che ne fonda le pratiche può rozzamente essere sintetizzato in "*non esiste conoscenza di valore separata dall'esperienza*". Le metodologie che si sono articolate in quest'area individuano questo processo: acquisizione, modellazione, rappresentazione, trattamento e gestione automatiz-

zati di conoscenze ritenute di grande valore intrinseco e strategico, tradizionalmente, in un ambito produttivo. L'integrazione dei risultati di questo processo (per esempio un Sistema Esperto che modella un aspetto fondamentale del *Core Knowledge*) con tutte le altre componenti che oggi caratterizzano i sistemi informativi è oggi scontata, come è scontata la possibilità di accedere a questo tipo di conoscenza per via telematica: lo sviluppo delle tecnologie d'integrazione tra sistemi eterogenei non ha lasciato in disparte quelle di sviluppo di Sistemi Esperti, e quindi il fantasma dello "stand-alone" è fugato (citiamo il caso della piattaforma JESS - http://herzberg.ca.sandia.gov/jess - per lo sviluppo di sistemi basati su regole che offre molte facilitazioni d'integrazione).
Se, da un lato, le facilitazioni tecnologiche agevolano la possibilità di integrare oggi diverse fonti di provenienza di dati, informazioni e conoscenze opportunamente strutturate, dall'altro, per consentire l'evoluzione di sistemi basati sulla conoscenza verso la fruizione comune e la condivisione, è stato necessario un passaggio teorico che tenesse in considerazione luoghi organizzativi e sociali in cui la conoscenza (in particolare quella *core*) opera e si sviluppa: questo passaggio era necessario per ovviare a quello che è stato precedentemente menzionato come il difetto dell'"isolamento" tipico dei Sistemi Esperti.

3. DAL SINGOLO ALLA COMUNITÀ: LO SFONDO TEORICO

Nati in un momento storico in cui ancora le potenzialità telematiche erano in luce, la filosofia di fondo che ha alimentato la nascita e lo sviluppo applicativo dei Sistemi Esperti era quelle delle soluzioni "uniche": ogni Sistema Esperto rispondeva a una specifica esigenza, e conteneva conoscenze altamente specialistiche per supportare o addirittura sostituire l'esperto umano. Quando vi erano delle esigenze di dominio che richiedevano una distribuzione delle soluzioni proposte dal sistema, si trattava di "distribuire" applicazioni sostitutive dell'esperto nei vari contesti. L'"isolamento" di un Sistema Esperto, che conteneva conoscenze altamente specializzate, rappresentava la garanzia della qualità delle soluzioni che proponeva. Ovviamente, questo pregio si trasforma in difetto quando la consapevolezza della natura sociale della conoscenza in un'organizzazione diventa perno delle possibilità di crescita di un'impresa. Quando tale consapevolezza ha iniziato a stimolare chi operava nell'area dei Sistemi Esperti (siamo ormai verso la fine degli anni '80), le prime esperienze hanno riguardato soluzioni tecnologiche di estrema avanguardia per il periodo. Sono state sviluppate architetture a *blackboard*, e si sono buttate le basi per un approccio di *Distributed Artificial Intelligence* [11], immediato precursore delle tematiche che oggi riguardano i modelli computazionali e le tecnologie *Agent-based* [10].
Gli sforzi che, sul livello più tecnologico, venivano compiuti nei tardi anni dell'era dei Sistemi Esperti per renderli fruibili nel contesto sociale e produttivo in cui dovevano operare, sarebbero stati premiati solo con la loro coniugazione con la rivoluzione di Internet e con lo sviluppo di modelli teorici dei contesti sociali in cui la conoscenza viene prodotta e utilizzata.
Per il secondo di questi aspetti, lo sviluppo di modelli concettuali e computazionali per un approccio distribuito alle fonti della conoscenza, e le basi teoriche per le più attuali applicazioni di Knowledge Management ci sono pervenuti proprio dall'ambito delle scienze dell'organizzazione aziendale e dai modelli sociali che in questo ambito sono stati sviluppati (Figura 1).
Tra i vari contributi che provengono da questo settore, quello che più ha influenzato (e tuttora influenza nelle sue varie più moderne declinazioni) a riconsiderare le potenzialità dell'approccio al trattamento della conoscenza dei Sistemi Esperti (e che ne ha ridisegnato i contorni) è quello sintetizzato da Wenger [23] e che riguarda la nozione di *Comunità di Pratica* (CoP), sistemi sociali, orizzontali e informali, immersi nell'esperienza, che si riconoscono in una forma di lavoro e di apprendimento collettivo finalizzato alla soluzione di problemi.
In una prospettiva di Knowledge Management, le *comunità di pratica* (con la loro caratterizzazione formale e dinamica) rappresentano un sostrato teorico di natura sociologica di grande utilità per il disegno di soluzioni computazionali (dall'architettura fino all'interfaccia uomo-macchina) a supporto della gestione della conoscenza. Si tratta infatti di gruppi

di persone che comunicano e interagiscono in modo strutturato o informale in un'organizzazione, per condividere le loro conoscenze attraverso un processo sociale di soluzione di problemi e di apprendimento reciproco. I membri una CoP sono caratterizzati dall'impegno in una qualsiasi attività, dalla forte coesione sociale che li unisce e dalla condivisione di una "cultura" specifica. Sono infatti sistemi sociali, orizzontali e informali, immersi nell'esperienza, che si riconoscono in una forma di lavoro e di apprendimento collettivo finalizzato alla soluzione di problemi. Sono numerosi gli esempi: un'equipe medica, il controllo di una centrale energetica, un gruppo di archeologi, e via di questo passo.

Wenger ha individuato tre dimensioni fondamentali che fanno delle CoP una struttura reticolare complessa di conoscenze (che si forma ed evolve nel tempo) parallela alla struttura funzionale che caratterizza ogni organizzazione produttiva, quale che sia la sua geometria. Vediamole in sintesi:

❑ **Iniziativa comune (*Joint Enterprise*)** - tale dimensione, assunta dalla relazione fra pratica e comunità, riguarda la formazione di un'immagine condivisa dei problemi e delle alternative percorribili; la creazione di obiettivi e priorità negoziate fra i membri e lo sviluppo di una comune consapevolezza dei compiti.

❑ **Impegno alla reciprocità (*Mutual Engagement*)** - nel momento in cui gli individui accedono e lavorano interconnessi alla dimensione comunitaria, essi stessi e le loro pratiche assumono un riconoscimento come parte di qualcosa di più grande (la comunità), che a sua volta è riconosciuto come parte di qualcosa di ancora più grande, l'azienda o l'ente che l'accoglie. Lungo questa dimensione, gli individui interagiscono e condividono l'*esperienza* che è propria di ciascuno, allo scopo di alimentare l'apprendimento collettivo.

❑ **Repertorio condiviso (*Shared Repertoire*)** - in questo senso, la relazione fra pratica e comunità genera un insieme di conoscenze, strumenti, artefatti e processi di sviluppo, che veicolano il sapere collettivo e custodiscono la memoria della comunità stessa.

Questa sintesi prende le sue mosse da lavori svolti da alcuni studiosi [5, 14] focalizzati sugli aspetti dell'apprendimento. Questi studi, che hanno fortemente influenzato gli aspetti teorici che sono alla base dei modelli di Knowledge Management, hanno indotto uno spostamento dell'attenzione dai *processi cognitivi* che avvengono nell'attività di problem solving di un individuo (un esperto, per esempio, visto in una prospettiva cognitivista) all'aspetto eminentemente *pratico* della conoscenza.

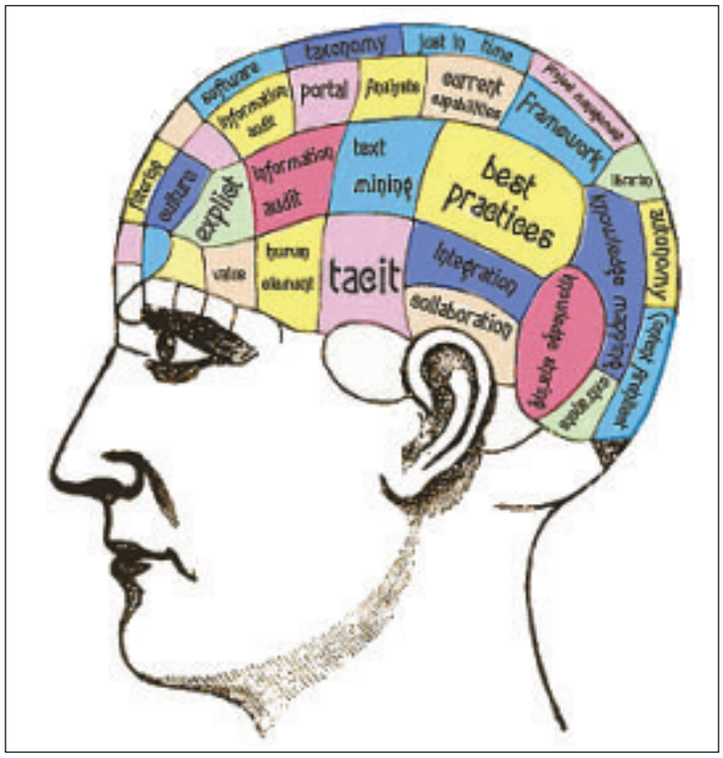


FIGURA 1

Un'immagine evocativa del complesso di conoscenze per la collaborazione che un esperto ha per coniugare le sue competenze in un ambito organizzativo (http://rmitvnim2007b.wikidot.com/km)

Per questi autori, *la conoscenza nasce dalla pratica e da essa non può essere separata*: che, nella prospettiva "filosofica" dei Sistemi Esperti, rimane quella importante da rappresentare. Non si tratta più di stimolare la formazione delle rappresentazioni individuali, fornendo strumenti simbolici e di tecniche di manipolazione all'attività di problem solving, quanto piuttosto, elaborare delle situazioni pratiche nelle quali immergere l'individuo, in modo che elabori - agendo - le sue rappresentazioni in vista della soluzione di compiti o problemi concreti (*teorie dell'esperienza situata*).

L'idea centrale è che la nozione di *pratica* debba essere intesa come un'attività che investe completamente l'individuo, fatta di elementi taciti e espliciti, di pensiero e azione, di saper

fare e di sapere situati in un contesto comunitario. Così intesa, la pratica sociale genera un senso *condiviso* che è completamente riconducibile al processo continuo di *negoziazioni* dei significati fra gli attori del contesto. Tale senso condiviso non corrisponde alla percezione del singolo, né ad una realtà autoreferenziale, ma è il frutto di una coproduzione tra i partecipanti, che lo influenzano e ne sono a loro volta influenzati. Si osservi che è proprio in questo patrimonio culturale condiviso che la comunità trova *legittimità*. Quando gli individui entrano a far parte come membri della comunità, fanno loro questo patrimonio comune e lo interiorizzano come parte della loro identità *in quanto* membri della comunità.
La gestione parzialmente automatizzata di questo modello sociale su cui imperniare soluzioni computazionali richiede la messa in azione di differenti tipi di tecnologie, opportunamente integrate in architetture complesse e dinamiche: si tratta delle principali componenti che, in diverse configurazioni, compongono i sistemi di Knowledge Management, e che, a seconda della prevalenza di qualche aspetto considerato fondamentale per l'accrescimento del valore della conoscenza, si estende su un ampio ventaglio, che va dal groupware al *data mining*, dall'*information retrieval*, fino alle tecnologie di *knowledge representation*.
Per quanto riguarda l'esperienza proveniente dall'ambito dei Sistemi Esperti e dei modelli di *knowledge representation* che nel loro ambito sono stati sviluppati, le configurazioni architetturali che supportano sistemi di Knowledge Management su cui meglio si innestano, riguardano quelle che danno prevalenza alla valorizzazione del *Core Knowledge* (Paragrafo 1.1), cioè quelle competenze possedute da comunità delegate al processo di creazione di soluzioni produttive di natura strategica e innovativa, e alla rappresentazione di conoscenze non sempre contenute in modo strutturato nella documentazione o formalizzate in protocolli espliciti, ma riconosciute come fondamentali nella dinamica di successo di un'impresa nell'era della *Knowledge Economy*.
In uno scenario organizzativo come quello prospettato dalla valorizzazione delle conoscenze condivise nelle Comunità di Pratica, sono da mettere in risalto tre fattori:
1. il patrimonio concettuale e lessicale condiviso;
2. il valore della pratica e dell'esperienza;
3. le storie esperite e la loro narrazione.
Il maggior contributo che i Sistemi Esperti forniscono riguardo al primo fattore è quello che deriva dai temi di rappresentazione della conoscenza non solo basata su regole euristiche (che comunque rimane una grande componente dei Sistemi Esperti), ma sulla rappresentazione di conoscenze strutturate in maniera tassonomica e ontologica. Solo per citare un esempio tra i vari possibili, ricordiamo il già citato esempio di Internist [16], in cui l'esigenza di fissare un repertorio concettuale condiviso era affrontata mediante l'uso di strumenti di rappresentazione concettuale di tipo classificatorio. Se teniamo conto che la maggioranza dei sistemi odierni per il Knowledge Management adotta un approccio *ontologico* alla rappresentazione della conoscenza [9, 19, 20], possiamo vedere in Internist (e in altre analoghe esperienze minori) un sistema "paradigmatico" precursore di questo modo di creare sistemi di supporto alla tesaurizzazione e gestione della conoscenza di valore, con tutte le lacune dovute al contesto storico.

4. CONOSCENZA EURISTICA ED ESPERIENZIALE: DALLE REGOLE AL CASE BASED REASONING

Tra i modelli per la rappresentazione e il trattamento della conoscenza provenienti dall'area dei Sistemi Esperti che ancora oggi offre soluzioni metodologiche e computazionali preziose, non possiamo non menzionare quella del *Case-Based Reasoning* (CBR - ragionamento basato su casi) [15], che con successo si è coniugata con quella parte del Knowledge Management che, piuttosto che sulla gestione di documenti e dei dati e le conoscenze in essi contenuti [1], pone la sua attenzione sul Core Knowledge, sulle potenzialità del sapere condiviso e su un altro fattore caratterizzante gli ambienti organizzativi in cui si formano e crescono comunità di pratica. Ci riferiamo al valore della *conoscenza narrativa*, fattore di coesione sociale per la soluzione di nuovi problemi, a partire da casi già precedentemente risolti.
Il presupposto epistemologico su cui si fonda

l'approccio del CBR afferma l'impiego di capacità inferenziali basate sul *ragionamento per analogia*, piuttosto che per deduzione-induzione-abduzione, come avviene in generale per i sistemi basati sulla conoscenza: data un'analogia su problemi, si ricava un'analogia tra soluzioni, con opportuni adattamenti (problemi simili hanno soluzioni simili), Quindi, soluzioni adottate in passato possono essere recuperate e adattate opportunamente per risolvere nuove situazioni critiche. Si tratta di un processo che usa le esperienze passate per comprendere le nuove situazioni, per suggerire un modo adatto a risolvere i nuovi problemi e per criticare le soluzioni proposte.
Dal punto di vista computazionale, i sistemi CBR si strutturano in un ciclo (Figura 2) rappresentato da quattro fasi distinte che coinvolgono diversi moduli che compongono il sistema stesso:

- *retrive*: fase di ricerca per analogia tra un problema corrente e una base di casi opportunamente memorizzata, al fine di recuperare il (o i) caso passato più simile;
- *reuse*: fase di eventuale riutilizzo delle soluzioni dei casi recuperati;
- *revise*: fase di revisione del caso, che consiste nell'adattamento della soluzione ritrovata al problema corrente da risolvere;
- *retain*: l'archiviazione del nuovo caso risolto nella memoria dei casi, in cui la situazione corrente, con la sua nuova soluzione, viene memorizzata per un recupero futuro.

Questo approccio ha trovato immediatamente un riscontro nelle applicazioni di Knowledge Management: sul fronte della ricerca, nell'ambito dell'AAAI (*Association for the Advancement of Artificial Intelligence, già American Association for Artificial Intelligence*) già dal '99 si è formato un gruppo (KM/CBR) dedicato ai sistemi di gestione della conoscenza basati su architetture evolute e distribuite di CBR. Per citare solo qualche applicazione di successo di questo approccio nell'ambito produttivo, ricordiamo il caso dell'industria farmaceutica, in cui Susan Craw e il suo gruppo dell'Università di Aberdeen, in collaborazione con AstraZeneca, ha applicato un sistema di CBR integrato con un modulo di sistema esperto per la creazione di nuovi prodotti farmaceutici [8]. Un analogo esempio è quello sviluppato su commessa Pirelli al Laboratorio di Intelligenza Artificiale del Dipartimento di Informatica, Sistemistica e Comunicazione dell'Università di Milano-Bicocca nell'ambito della formulazione chimica di mescole battistrada per pneumatici dedicati alla competizione sportiva [2], e infine l'uso di tecniche di *Learned Lessons* (una più moderna interpretazione del *Case Based Reasoning*) per lo sviluppo di applicazioni distribuite di supporto all'ambito militare, ad opera del Dipartimento di Intelligenza Artificiale della Marina Militare Americana (*Naval Research Lab*), ad opera di David Aha e del suo gruppo [22].
Va rimarcato il ruolo che le tecniche *rule-based* dei Sistemi Esperti possono giocare per l'implementazione del modulo di adattamento della soluzione in un sistema di *Case-based Reasoning*. È qui che sta la chiave, quando questo è possibile, per catturare l'esperienza pratica e narrativa che tesaurizza la conoscenza tacita degli esperti (possessori di Core Competence), che anche in questa forma fonda le

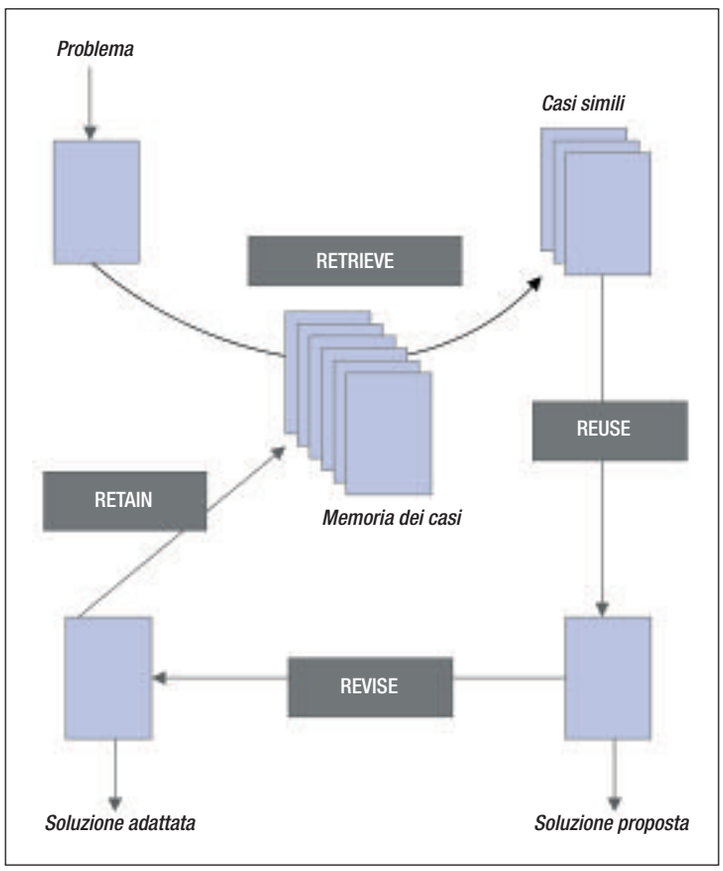


FIGURA 2
Il ciclo di un sistema di Case-based Reasoning

caratteristiche comunitarie di un'organizzazione, permettendo di contribuire in modo sostanziale alla messa in azione delle capacità innovative di una realtà produttiva. La gestione della conoscenza e il trattamento di conoscenza strutturata mediante casi è possibile grazie alla natura intrinseca del patrimonio narrativo delle comunità di pratica (ruolo dello *storytelling* [4]) e dal ruolo fondamentale dell'applicazione di una nuova soluzione come adattamento innovativo di soluzioni già esperite in precedenza: questo fattore implica in modo diretto la possibilità di valorizzare la traccia delle esperienze vissute nella prassi operativa accumulata nel tempo. Ogni sistema di Knowledge Management che voglia incorporare anche questo aspetto della conoscenza può far tesoro dell'esperienza dei sistemi che trattano il ragionamento basato su casi.

5. UNA NOTA CONCLUSIVA

Le considerazioni qui riportate, in modo molto sintetico, tracciano un percorso che, a partire dalle esperienze di progettazione e creazione di Sistemi Esperti, conduce al trattamento e alla gestione della conoscenza (Knowledge Management) che parte dall'esigenza dell'esplicita rappresentazione di conoscenze esperte e specialistiche (*Core Knowledge*) per il progresso innovativo di un'impresa.
Chiunque si sia impegnato o abbia seguito anche lateralmente le esperienze progettuali e applicative dei Sistemi Esperti può trovare un codice interpretativo per leggere le nuove frontiere del Knowledge Management che il ritmo del progresso impone nella direzione della Società della Conoscenza.
Il tema è assai vasto e, fortunatamente, caratterizzato da continue riflessioni che provengono da ambiti applicativi. Inoltre, i contributi multidisciplinari che contraddistinguono i tratti del Knowledge Management, permettono oggi di abbattere confini (spesso ideologici) di settore, in modo che la soluzione a problemi di automazione sempre più complessi armonizzi i contributi provenienti da differenti territori del sapere. In quest'ottica, ricordiamo il progetto PKM360° (http://labs.unicatt.it/pkm360/), che offre un forum di discussione multidisciplinare in cui questi temi di discussione trovano spazio e riscontro.

Bibliografia

[1] Althoff K.D., Weber R.O.: Knowledge Management in Case-based Reasoning. The *Knowledge Engineering Review*, n. 20, 2005.

[2] Bandini S., Manzoni S.: *Modeling Core Knowledge and Pratices in a Computational Approach to Innovation Process*. In:. Magnani L, Nersessian N.J., (eds.), "Model-Based Reasoning: Science, Technology, Values", Kluwer, New York, 2002.

[3] Bandini S., Colombo E., Vizzari G.: *The Role of Knowledge Artifacts in Knowledge Maintenance*. Proceedings 2nd Indian International Conference on Artificial Intelligence (IICAI 2005), Pune, 2005.

[4] Bandini S., Petraglia F., Sartori F.: *Social and Domain Knowledge Management through the Integration of Storytelling and Case Based Reasoning*. In: Sicilia M.A., et al. (eds.), "Advances in Metadata and Semantics Research", Springer Verlag, 2008.

[5] Brown J.S., Collins A., Duguid P.: *Situated Cognition and the Culture of Learning*. Educational Researcher, n. 18, 1989.

[7] Cobos R., Esquivel J.A., Alaman J.A.X.: *IT Tools for Knowledge Management: a Study of the Current Situation*. The European Online Magazine for the IT Professional (www.upgrade-cepis.org), Vol. III, n. 1, 2002.

[8] Craw S., Wiratunga N., Rowe R.: *Case-based Design for Tablet Formulation*. Proceedings 4th European Workshop on Case-Based Reasoning, Dublin, 1998.

[9] Davies J., Fensel D., Van Harmelen F.: *Towards the Semantic Web: Ontology-driven Knowledge Management*. J. Wiley & Sons, 2003.

[10] Ferber J.: *Multi–Agent Systems*. Addison–Wesley, 1999.

[11] Gasser L., Huhns M.N.: *Distributed Artificial Intelligence*. Morgan Kaufmann, Los Altos, CA, 1990.

[12] Guida G., Berini G.: *Ingegneria della conoscenza: strumenti per innovare e per competere*. EGEA Ed., Milano, 2000.

[13] Hayes-Roth F., Waterman D.A., Lenat D.B.: *Building Expert Systems*. Addison-Wesley, Reading, MA, 1983.

[14] Kirshner D., Whitson J.A.: *Situated Cognition: Social, Semiotic, and Psychological Perspective*. Lawrence Erlbaum, NJ, 1997.

[15] Kolodner J.: *Case-Based Reasoning*. Morgan Kaufmann, San Mateo (CA), 1993.

[16] Miller R.A., Pople H.E., Myers J.D.: Internist-1, an Experimental Computer-based Diagnostic Consultant for General Internal Medicine. The *New England Journal of Medicine*, Vol. 307, n. 8, 1982.

[17] Nonaka I., Takeuchi H.: T*he Knowledge-Creating Company: How Japanese Companies Crea-*

te the Dynamics of Innovation. Oxford University Press, 1995.

[18] Prahalad C.K., Hamel G.: T*he Core Competence of the Corporation*. In. "Strategic Learning in Knowledge Economy: Individual, Collective and Organizational Learning Process", Cross R.L., Israelit S.B., (eds.), Butterworth Heinemann, Boston, 2000.

[19] Sowa J.: *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Pacific Grove, CA 2000.

[20] Staab S., Studer R., (eds.): *Handbook on Ontologies, International Handbooks on Information Systems*. Springer Verlag, 2004.

[21] Studer R., Benjamins V.R., Fensel D.: Knowledge Engineering: Principles and Methods. *Data & Knowledge Engineering*, n. 25, 1998.

[22] Weber R.O., Aha D.W.: Intelligent Delivery of Military Lessons Learned. *Decision Support Systems*, Vol. 34, n. 3, 2003.

[23] Wenger E.: *Community of Practice: Learning, meaning and identity*. Cambridge University Press, Cambridge, MA, 1998.

STEFANIA BANDINI, Professore Ordinario di Informatica presso il Dipartimento di Informatica, Sistemistica e Comunicazione dell'Università di Milano-Bicocca, dove dirige il Dottorato di Ricerca in Informatica. Direttore del Centro di Ricerca "Complex Systems & Artificial Intelligence" dello stesso Ateneo. La sua attività di ricerca riguarda due settori principali: Intelligenza Artificiale (Case-based Reasoning, Knowledge Management) e Sistemi Dinamici Discreti (modelli e sistemi basati su Automi Cellulari e Situated Agents). In questi settori partecipa in ruoli di responsabilità a numerosi progetti con enti pubblici e privati.
E-mail: stefania.bandini@csai.disco.unimib.it

LA VISIONE ARTIFICIALE PER LA VIDEO SORVEGLIANZA

Rita Cucchiara

Gli studi di Visione Artificiale e Pattern Recognition propongono soluzioni sempre più sofisticate per essere applicate alla Video Sorveglianza, un settore emergente nell'ICT. I sistemi di Video Sorveglianza, infatti, non richiedono solo un'efficiente supporto hardware per l'acquisizione e la gestione dei video, ma anche tool software con algoritmi efficaci per l'analisi dei dati in tempo reale. In questo articolo si presentano gli aspetti della Video Sorveglianza e delle tecniche di Visione Artificiale necessarie per diverse applicazioni. Infine vengono illustrati alcuni progetti di ricerca italiani.

3.7

1. VIDEO SORVEGLIANZA

La *Video Sorveglianza* è quel settore dell'ICT che si occupa dell'acquisizione, elaborazione e gestione in tempo reale di video provenienti da telecamere installate in luoghi pubblici e privati per molteplici scopi, comunque correlati al concetto di *sicurezza*. Nei Paesi anglosassoni si parla di "*security* & *safety*", coinvolgendo quindi sia la sfera pubblica, per la lotta alla criminalità e la prevenzione di reati, sia la sfera privata per garantire l'incolumità e la tranquillità personale.
Rientrano nella grande categoria di sistemi di Video Sorveglianza tutti gli apparati installati nel contesto urbano (piazze, strade e monumenti), nei luoghi affollati (stazioni, metropolitane, aeroporti, stadi), nelle banche, nelle abitazioni private e nei siti industriali, nonché i sistemi per il monitoraggio del traffico (veicolare, aereo, navale) e quelli per il controllo ambientale (incendi, alluvioni, frane).
È indubbio che la Video Sorveglianza sia per il mondo ICT una delle sfide applicative più interessanti degli ultimi anni. La sua progettazione coinvolge l'*hardware* per i sensori e i sistemi di acquisizione, l'architettura dei sistemi di calcolo, l'infrastruttura di comunicazione *wired* e *wireless*, il *middleware* e soprattutto la sfera algoritmica e cognitiva di competenza della *Visione Artificiale* e della *Pattern Recognition*.
Sebbene le soluzioni architetturali e hardware siano ormai ben consolidate, moltissimi problemi rimangono aperti nell'analisi automatica dei video, dove la ricerca scientifica sta investendo al massimo.
La **Visione Artificiale** (*Computer Vision*) fino a pochi anni fa era bollata come un vezzo accademico, una disciplina di nicchia al più assimilata a una sotto-disciplina dell'*Intelligenza Artificiale* (IA). Con l'IA, la Visione Artificiale condivide molti dei meccanismi inferenziali e strumenti per la rappresentazione della conoscenza. Tuttavia, a differenza dell'IA, la computazione oltre che del livello simbolico si occupa del basso livello, spesso chiamato livello "iconico". In altre parole, i processi di *Computer Vision* non possono prescindere dall'essenza dell'input che è fatta di sequenze di

pixel, le cui informazioni visuali sono per natura parziali e corrotte da rumore sintattico (a livello di segnale) e semantico (i cosiddetti "distrattori" nella scena).
La Visione Artificiale si avvale fortemente della **Pattern Recognition**, la disciplina che studia tecniche, spesso statistiche, di classificazione e riconoscimento di forme a partire da dati misurati o calcolati, siano essi visuali, audio, testuali, multimediali in genere o numerici. Ora la Visione Artificiale e la Pattern Recognition sono ben affermate nel mondo dell'ICT e sono materie di studio in molti corsi di laurea di ingegneria informatica e di informatica. Due ottimi testi su tali discipline sono [1, 2].
Questo articolo vuole introdurre i fondamenti della Video Sorveglianza basata sulla Visione Artificiale, mostrando le esigenze hardware e software per i sistemi della prossima generazione. Tali sistemi sfrutteranno a pieno tecniche e algoritmi per la percezione, l'apprendimento automatico (*machine learning*), modelli di attenzione e di ragionamento su immagini (*cognitive vision*), tecniche di *clustering* e classificazione (*pattern classification*) e tecniche inferenziali proprie dell'IA, per fornire allarmi in modo automatico, comprendere la scena e il comportamento di persone e oggetti di interesse, e annotare i dati visuali.
Video Analytics è il termine che sta emergendo per denotare l'insieme di queste tecniche, sia per i sistemi in tempo reale, sia per i sistemi *off-line* di analisi di dati multimediali legati alla sicurezza: in questo caso la Video Sorveglianza si estende a discipline limitrofe quali il *content-based video retrival* per il recupero per similarità di dati visuali o il *video mining* per la ricerca di associazioni e correlazioni e l'estrazione di conoscenza dal video.
Si apre la strada ad applicazioni innovative per il supporto all'indagine forense o all'intelligence, e per mercati emergenti nel campo dell'ICT quali l'identificazione di comportamenti in ambienti sociali, pubblici e privati, supermercati o musei. Alcuni esempi di possibili applicazioni sono riportati nella figura 1.

FIGURA 1

Applicazioni di video sorveglianza. Dall'alto a sinistra: tracking di persone davanti a vetrine (video di PETS'2002); individuazione di veicoli fermi; rilevazione di pacchi abbandonati (PETS'2006); riconoscimento della testa e del volto da telecamera PTZ. Per informazioni si veda http://imagelab.ing.unimore.it

2. VIDEO SORVEGLIANZA E L'OPINIONE PUBBLICA

Tutti ora parlano di Video Sorveglianza, sui giornali, TV e Web, spesso intendendo in forma molto riduttiva solo le centinaia di telecamere installate nelle nostre città, nelle strade e negli uffici collegati a centri di controllo dove le immagini dovrebbero essere osservate con continuità da occhi attenti di addetti alla sicurezza. Ovviamente ciò non accade e questi sistemi istallati, le cui immagini non sono controllate né da operatori umani né dal calcolatore, hanno un effetto preventivo e reattivo di fatto assai limitato. Le telecamere hanno comunque una funzione deterrente e possono fornire un validissimo aiuto come supporto alle indagini forensi a posteriori.

Ciò nonostante, l'interesse e l'aspettativa dell'opinione pubblica in Italia è decisamente altalenante: da un forte contrasto si è passati al plauso incondizionato e viceversa. Dieci anni fa, mentre negli Stati Uniti tali sistemi si stavano capillarmente distribuendo, in Italia spesso si assolveva la carenza di infrastrutture e denaro pubblico con una legittimazione del garantismo ad oltranza supportato da un dibattito serrato sui potenziali problemi di privacy. Basta ricordare l'abuso di citazioni al grande fratello del "1984" di Orwell e a bui scenari di occulto controllo dei cittadini che presupporrebbero una rete di intelligence e una tecnologia di distribuzione in tempo reale dell'informazione fino a pochi anni fa impensabile.

Negli ultimi anni, invece, e soprattutto negli ultimi mesi, la diffusa sensazione di insicurezza nei luoghi pubblici e nei centri urbani sembra aver fatto dimenticare il problema della privacy e si richiedono a gran voce nuove e capillari installazioni di telecamere come placebo ad ogni illegalità. Sui giornali locali non manca ogni giorno una richiesta di nuove telecamere per la sicurezza dei cittadini, a cui si associa la richiesta dalle forze dell'ordine delle telecamere per il controllo del traffico e delle infrazioni stradali (semafori, velocità ZTL).

Le telecamere istallate in Italia solo dagli enti pubblici sono migliaia, centinaia a Roma e a Milano e numerose anche nelle città di provincia se si pensa che nel 2008 il record sembra essere di Reggio Emilia con una telecamera ogni 350 abitanti.

Si diceva però opinione altalenante: infatti proprio recentemente è stata enfatizzata sui giornali italiani una notizia sull'inutilità dei sistemi di video sorveglianza.

In un'intervista al *The Guardian* del maggio 2008 Mick Neville, ispettore di Scotland Yard, ha bollato la Video Sorveglianza come un "*utter fiasco*", affermando che l'Inghilterra a fronte dell'enorme diffusione degli apparati (4,2 milioni di telecamere con un rapporto 1 a 14 tra telecamere ed abitanti) e dei miliardi di sterline impiegati, ha sistemi decisamente inefficaci; ha stimato solo in un 3% il miglioramento nella lotta alla criminalità grazie alle telecamere.

Il problema è quello prima richiamato: le forze dell'ordine non possono materialmente osservare in tempo reale centinaia di migliaia di video e l'efficacia di tali sistemi nel momento in cui il fatto si verifica e' di fatto nulla. L'ispettore al termine dell'intervista però assolve i sistemi del futuro e ipotizza come unica soluzione plausibile l'affidarsi al software di analisi di immagini per elaborare in tempo reale i video e fornire allarmi.

Da anni la comunità scientifica internazionale sta lavorando sulla Visione Artificiale per trovare soluzioni all'interpretazione della scena in modo automatico. I primi progetti di ampio respiro provengono dagli Stati Uniti e sono stati finanziati dal DARPA alla metà degli anni '90, come i progetti VSAM per l'inseguimento di persone e MugSpot per il riconoscimento del volto. Piano piano anche nel mondo commerciale la proposta di tools di "*Intelligent Surveillance*" o "*smart Surveillance*" o "*Surveillance with Video Analytics*" diventa maggiore. Con l'avvento di queste tecnologie l'ente inglese di analisi di sicurezza *IMS Research* ipotizza una crescita del 65% annuo dal 2004 ad oggi e predice un valore del mercato di 215 milioni di sterline nel 2009 solo per la parte software di *video analytics*.

3. ARCHITETTURA DI UN SISTEMA DI VIDEO SORVEGLIANZA

I sistemi di Video Sorveglianza non possono più essere considerati solo il supporto elettronico per l'acquisizione e memorizzazione di video. Essi invece hanno un'architettura complessa, sia hardware che software.

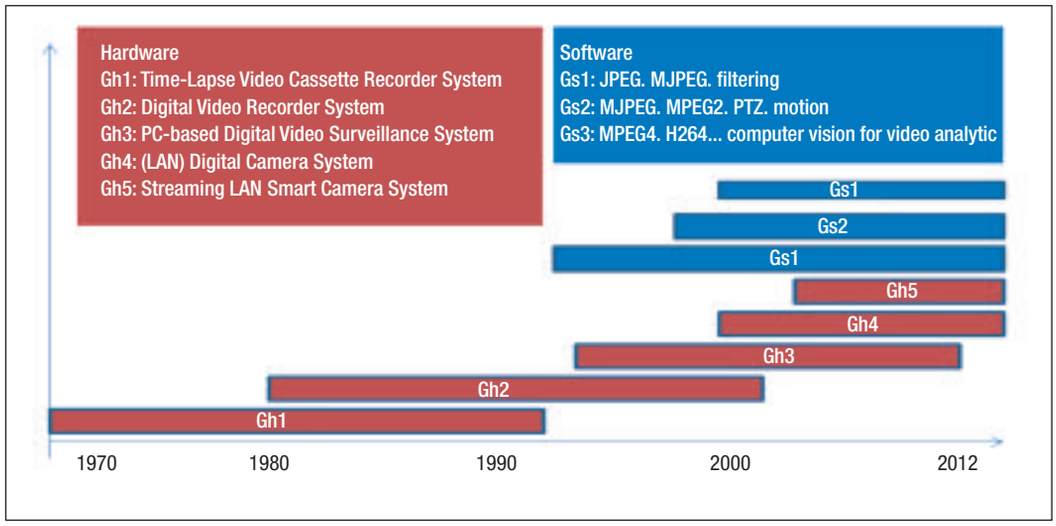


FIGURA 2
Evoluzione delle tecnologie hardware e software per sistemi di video sorveglianza

Per quel che riguarda la parte hardware, si è assistito ad una grande evoluzione negli ultimi cinquant'anni. Il primo sistema pubblicizzato risale al 1969, istallato al Muncipal Building del City Hall di New York: in quel tempo i sistemi di prima generazione (Gh1 nella Figura 2) erano costituiti da telecamere CCTV analogiche collegate a VCR (*Video Cassette Recorder*) e monitor. Questa tecnologia è rimasta in auge per più di vent'anni ed è ancora in uso, per la falsa credenza che le video cassette prelevate fisicamente solo da personale autorizzato siano più sicure di altri supporti digitali. La seconda generazione (Gh2 nella Figura 2) ha visto l'impiego di tecnologia digitale nella registrazione: le telecamere analogiche CCTV, solitamente a livello di grigio, venivano collegate a video *switch* o *multiplexer* e trasferite a DVR (*Digital Video Recorder*) per salvare qualche immagine selezionata, solitamente in JPEG. Il video *switch* permetteva quindi sia una visualizzazione analogica sia una registrazione digitale. Questi sono tuttora i sistemi a basso costo in uso in sedi dove non è richiesta un'interazione con il servizio di sorveglianza (per esempio, banche).
Una grande trasformazione si è verificata quando le tecnologie informatiche hanno soppiantato quelle puramente elettroniche, impiegando il calcolatore (Gh3 nella Figura 2), sostituendo al *multiplexer* e al DVR, un PC con *video capture card* capace di comprimere immagini e video su hard disk. L'avere un PC anche industriale permetteva inoltre un possibile collegamento remoto nonché una sostituzione dei monitor CCTV con monitor digitali. Sistemi di questo tipo furono impiegati massivamente: un primo esempio risale all'istallazione sempre a New York attorno all'area del Word Trade Center dopo il primo attacco di febbraio del 1993. Purtroppo inutilmente. I sistemi di questo tipo, tutt'oggi assai venduti hanno tipicamente 1-16 telecamere collegate, compressione in modo *intra-frame only* (in MJPEG) e con risoluzione e *frame-rate* bassi dipendenti dal numero di telecamere collegate e dalle schede di *frame-grabbing*. Tipicamente questi sistemi visualizzano una sola o poche telecamere per volta a risoluzione completa e a colori mentre registrano al 25% della risoluzione (esempio, formato CIF 320 × 240) a 25 o 30 frame al secondo complessivamente per tutte le telecamere. Questi sistemi sono ottimali per la visualizzazione di piccole istallazioni private (negozi, uffici); spesso però sono impiegati anche in grandi centri come nei centri di polizia, ma hanno una qualità ben poco accettabile per l'analisi forense, dando a disposizione pochi frame (esempio, 3 al secondo con 10 telecamere) e a una risoluzione così bassa che spesso rende impossibile l'identificazione degli individui o delle targhe.
I sistemi più recenti (Gh4 nella Figura 2) adottano telecamere digitali, collegabili anche alla LAN con tecnologia IP e con processori RISC o con FPGA (*Field Programmable Gate Array*) a bordo che permettono un forte risparmio nell'infrastruttura. Queste telecamere hanno spesso la possibilità di essere controllate in re-

moto e brandeggiate in pan, tilt o zoom (PTZ) e sono spesso impiegate per larghe distanze come ad esempio negli stadi. Le telecamere sono collegate a uno *switch ethernet* che le collega a centri di controllo e di registrazione remoti. In larghe istallazioni urbane, "foreste di telecamere" sono collegate in fibra ottica a centri di acquisizione con una topologia a stella da cui i dati sono trasferiti a centri di controllo.
Infine le ultime soluzioni hardware (generazione Gh5) impiegano *smart cameras* con RISC e/o DSP a bordo per permettere una miglior compressione video anche in streaming (si parla di *streaming lan cameras*) con tecnologie MPEG2, MPEG4 o H264; esse riassumono tutte le capacità precedenti, come il controllo PTZ, necessitano di PC per la decompressione video e visualizzazione o per la memorizzazione. L'uso di IP e di streaming naturalmente implica tutti i problemi di sicurezza legati ad internet e quindi non sempre vengono installate come dovrebbero, se non in intranet o in reti sicure.
L'impiego di processori, sia a bordo delle telecamere, sia nel sistema di raccolta dei flussi video ha aggiunto la possibilità di un'elaborazione software delle immagini *on-the-fly* lasciando campo aperto all'uso di tecniche più o meno sofisticate di Visione Artificiale.

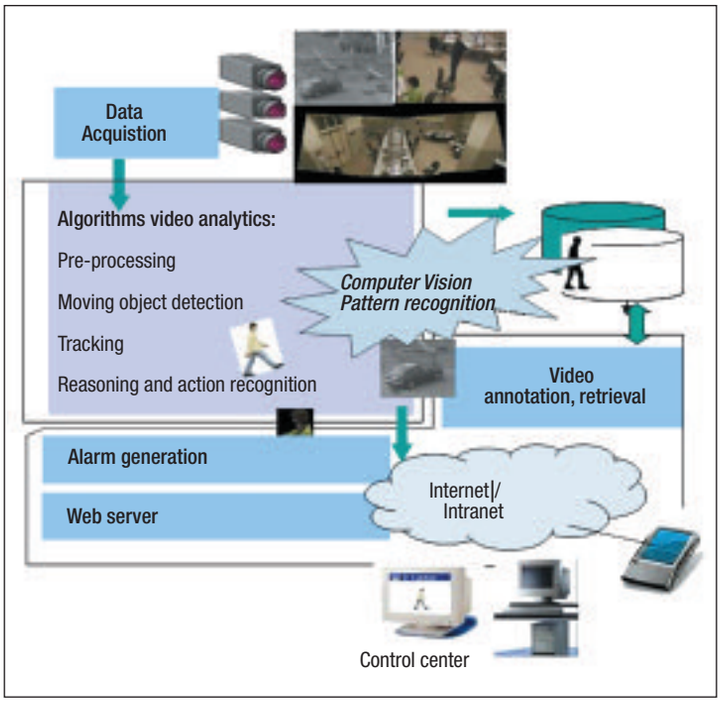


FIGURA 3
Architettura di un sistema di videosorveglianza

Anche per il software si può parlare di diverse generazioni commerciali.
Nella prima generazione (GS1), con tecnologie hardware dalla terza generazione in poi, l'impiego di software era assai limitato, solo a qualche filtraggio digitale per il rumore e alla compressione.
Nella seconda generazione (Gs2) che comprende la quasi totalità di sistemi sul mercato, vengono adottati alcuni semplici algoritmi di elaborazioni di immagini per una buona compressione video e per il "*motion detection*", di cui se ne parlerà nella prossima sezione.
La nuova generazione (Gs3) che si sta affacciando ora sul mercato ma che probabilmente lo saturerà per il prossimo quinquennio, adotterà in modo massivo la Visione Artificiale, impiegando *tools* per il tracciamento di oggetti in movimento, il riconoscimento di azioni di interesse, la registrazione e la trasmissione selettiva anche su *mobile*, e nuove interfacce multimediali per l'annotazione e il *retrieval*. Nella figura 3 è riportato uno schema completo dove la parte centrale e' basata su algoritmi e tecniche di Visione Artificiale e Pattern Recognition. I dati così elaborati generano allarmi automatici anche via web e sono raccolti in un DBMS accessibili anche remotamente e a posteriori. Ora molti colossi aziendali offrono soluzioni di questo tipo come IBM con il sistema S3 [3] o Siemens [4], ma anche molte piccole aziende si stanno muovendo: tra esse citiamo la Wavelet Technology Italia[1] per il sistema realizzato a Reggio Emilia. Stanno inoltre entrando sul mercato *software house* che si occupano solo della parte software, impiegando hardware general-purpose. Molte di queste aziende sono nate come spin off universitari, come l'Object Video[2] americana ora leader mondiale nel settore.

4. VISIONE ARTIFICIALE E PATTERN RECOGNITION PER LA VIDEO SORVEGLIANZA

La Visione Artificiale è quella disciplina che si occupa di dare al calcolatore capacità percettive e visive simili a quelle biologiche, almeno in applicazioni specifiche. Si avvale forte-

[1] http://www.wti.it.
[2] http://www.objectvideo.com.

mente di competenze di geometria, di ottica e di calcolo numerico per definire modelli computazionali al fine di riconoscere il mondo 3D a partire da immagini 2D. È strettamente legata alla *Pattern Recognition*, che si occupa di tecniche di *data analysis*, soprattutto statistiche ma anche sintattiche e strutturali (o *graph-based*) per ragionare sui dati estratti dalle immagini, avendo come fine soprattutto la classificazione supervisionata o non supervisionata. Entrambe le discipline hanno specifiche tecniche e modelli per la Video Sorveglianza, per elaborare i video e riconoscere la scena in tempo reale, e, con vincoli computazionali più rilassati, per analizzare *off-line* i dati relativi alla sicurezza.
Non è possibile fare una trattazione completa di tutte le tecniche impiegate, ma in questo contesto si vogliono piuttosto ricordare i passi fondamentali che un sistema di Video Sorveglianza di nuova generazione deve prevedere.

4.1. Pre-processing

In una fase iniziale tecniche di *image processing* sono sempre impiegate per il miglioramento delle immagini digitali. La teoria dell'*image processing* è ben consolidata. Per un testo di riferimento si veda per esempio, [5]. Anche in base ai sistemi di acquisizione, sono necessarie operazioni di filtraggio, di *noise reduction*, di *image restoration* e di *quality enhanchement*. Si usano tecniche di *camera motion compensation* per eliminare le piccole fluttuazioni delle telecamere. Con telecamere PTZ si applicano algoritmi di registrazione per la creazione di mosaici (*mosaicing*), come riferimento dello sfondo (Figura 4).
Nel caso di sistemi con più telecamere e con campi di vista (Field of View, FoV) sovrapposti, si opera la calibrazione manuale o automatica per ottenere la ricostruzione 3D. Si può anche sfruttare una calibrazione parziale per costruire l'omografia e premettere a tutte le telecamere di sovrapporre la loro visuale. Nella figura 5, per esempio, si vede l'output di un sistema con due telecamere con FoV sovrapposte e con la trasformazione omografica del piano immagine di una telecamera nel piano dell'altra. La corrispondenza è ulteriormente garantita dalle linee epipolari che passano per il punto più alto della testa [5].

4.2. Moving object detection

Il passo iniziale consiste nell'estrazione degli oggetti di interesse, che di solito sono, o sono stati in frame passati, in movimento. Sono perciò necessarie tecniche di analisi del moto: o viene stimato il moto di ogni punto dell'immagine, poi raggruppato in cluster omogenei per estrarne forme con la stessa direzione di moto, o viene direttamente effettuata la segmentazione di gruppi di pixel adiacenti caratterizzati da movimento. Nel primo caso si utilizzano tecniche di stima del flusso ottico (*optical flow*) su ogni punto o su punti di interesse quali angoli o speciali *feature* come i SIFT (*Scale Invaraint Feature Tracking*) [6]. Questi metodi più complessi si impiegano solitamente in caso di telecamere mobili dove altre tecniche non sono possibili. In caso invece di telecamere fisse, si usano metodi differenziali di *change detection*, che impiegando differenze tra frame o solitamente differenze con frame di riferimento (*background subtraction*) estraggono tutti i punti che sono diversi dallo sfondo e quindi in movimento. In questo contesto c'è una fervente attività di ricerca internazionale per studiare le tecniche più efficienti e reattive per modelli di *background* che tengano conto non solo degli oggetti in movimento ma anche del fat-

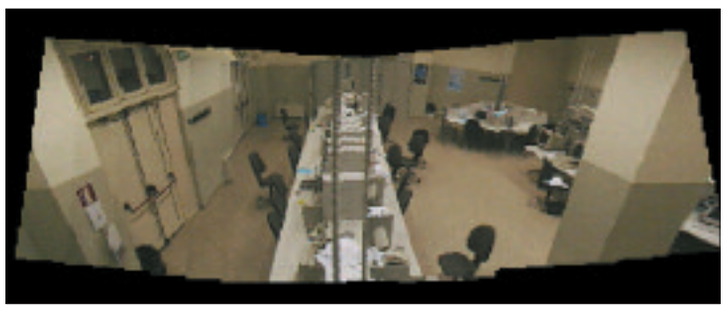


FIGURA 4
Mosaico creato da PTZ indoor

FIGURA 5
Omografia con due telecamere e rette epipolari

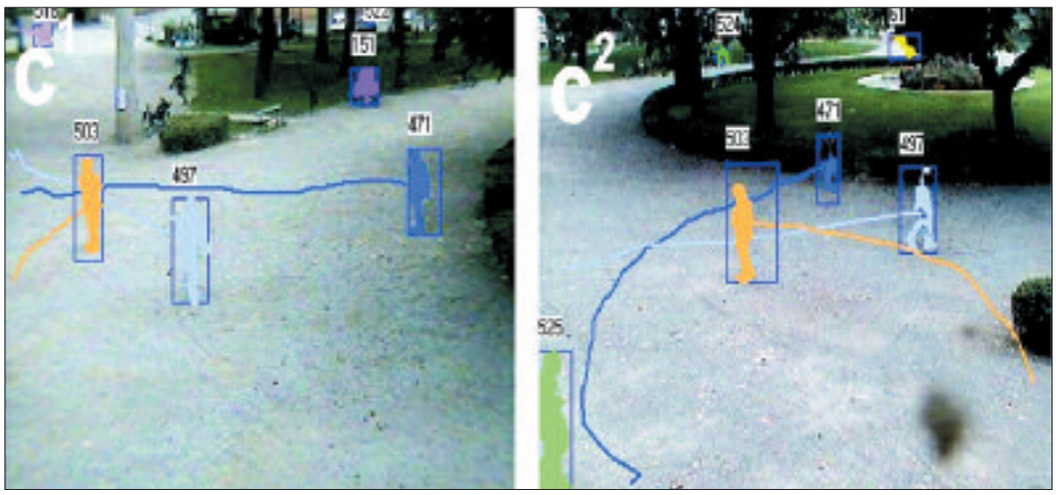


FIGURA 6
Esempio di tracking di persone in un parco (Progetto Piano Telematico ER LAICA- Comune di Reggio Emilia)

to che lo sfondo cambia dinamicamente (ore del giorno, condizioni atmosferiche ecc.). L'approccio più famoso anche se computazionalmente pesante modella ogni pixel con una *Mixture of Gaussians* [7]. I sistemi semplici che propongono solo la "*motion detection*" si fermano qui e rilevano se punti possono dichiararsi in modo avendo cambiato la loro intensità frame dopo frame. Se il numero di tali punti è alto allora il sistema suppone la presenza di movimento e inizia la registrazione. Invece questo non è che il primo passo a cui devono eseguire altri processi di *labeling* per raggruppare punti adiacenti che potenzialmente fanno parte dell'oggetto in movimento (i blob disegnati nella Figura 1), tecniche di rimozione di artefatti come ombre [8], riflessi e rumore e tecniche di classificazione degli oggetti estratti, in base all'applicazione.

4.3. Tracking

Con la segmentazione si ottiene la coerenza spaziale tra i dati di ogni frame: si estraggono così gli oggetti di interesse, che devono essere inseguiti nel tempo (*tracking*) per mantenere anche la coerenza temporale.
Si studiano tecniche sofisticate per il tracking di più oggetti in movimento interagenti, che si occludono gli uni con gli altri o che possono temporaneamente sparire nella scena. I semplici *Kalman Filters* presuppongono un modello di moto lineare che può essere adottato solo in alcuni contesti come il controllo del traffico veicolare, ma che non sono adeguato per oggetti deformabili e con moto variabile quali le persone. In questo caso si usano modelli statistici discriminativi, come quelli *appearance-based* e generativi, come il famoso algoritmo di *particle filtering* in cui si tengono in vita più stime (particelle) dello stato dell'oggetto poi messe in corrispondenza con l'osservazione ad ogni frame.
Un'interessante *survey sull'object tracking* è quello di Shah *et al*. [10]. L'output del tracking è l'informazione completa dell'oggetto, della sua forma e della traiettoria frame per frame; in caso di più telecamere viene fatto anche il *consistent labeling* per mantenere la stessa identità di ogni individuo se pur acquisito da telecamere diverse. Questo è assai importante nei nuovi sistemi dove si vuole tenere traccia degli oggetti osservati che si muovono in ampie scene coperte da più telecamere. Nella figura 6 si vede un esempio di tracking di persone nel Parco del Popolo di Reggio Emilia, uno dei prototipi del progetto LAICA (piano telematico Emilia Romagna) sviluppato dall'ImageLab di Modena [11].

4.4. Reasoning and Action Recognition

Avendo a disposizione i dati della scena e degli oggetti in movimento, possono iniziare diversi passi di *reasoning* e di riconoscimento che dipendono naturalmente dall'applicazione. Iniziano processi di riconoscimento di eventi di interesse e di interazioni tra oggetti, quali il rilevamento di oggetti abbandonati. In questo caso si impiegano spesso metodi spesso basati su grafi spaziali e temporali che cercano oggetti che cambiano il loro sta-

to da moto a fermo e che sono correlati con altri oggetti (persone) in movimento.
L'attività di ricerca ora più attiva nel settore è quella dell'*action analysis*, che studia le azioni delle persone, le traiettorie e le modifiche di forma, per monitorarne attività o riconoscere il comportamento (*bheavior analysis*). Anche in questo ambito si usano spesso grafi probabilistici quali gli HMM (*Hidden Markov Model*), tecniche di ragionamento con regole e *parser* sintattici. Un buon *survey sull'human motion analysis* è [12].
La ricerca in questo settore sta producendo diversi prototipi interessanti e si stanno diffondendo anche i primi esempi commerciali. Questo è certo uno dei settori dove si stanno concentrando maggiormente gli investimenti pubblici e privati. La ricerca si diffonde anche per il trasmettersi del sapere sull'argomento: molte librerie software sono disposizione anche per chi si affaccia per la prima volta alla materia come i toolbox di Matlab e le OpenCV[3], librerie supportate da Intel ma totalmente open source che si arricchiscono di giorno in giorno.
Nella ricerca è però fondamentale il confronto, e questo è uno degli aspetti più critici nel settore. Mentre in altri ambiti i *benchmark* di riferimento sono ormai ben diffusi (esempio, FERET per l'analisi del volto, TRECVID per il video retrieval) nella Video Sorveglianza ci sono tanti progetti poco coordinati. Tra essi si citano i *benchmark* PETS[4], CAVIAR[5] e il recente VISOR[6] che è un *open repository* di video di sorveglianza, con relativa ontologia, annotazione in XML e strumenti per la visualizzazione e l'annotazione manuale basati sui tools VIPER dell'Univ. of Maryland (Figura 7).

5. LA RICERCA IN ITALIA

La Visione Artificiale per la Video Sorveglianza è un campo di ricerca dove l'Italia ha un ruolo di grande rilievo nella scena mondiale. Ne sono prova i tanti convegni internazionali organizzati da ricercatori italiani negli ultimi anni, i tanti progetti di ricerca nazionali ed internazionali. Alcune di queste attività di ricerca sono state presentate a VISIT2008[7], organizzato dal GIRPR[8] che ha creato uno specifico TC in Video Sorveglianza.
Ricercatori del GIRPR attualmente sono coinvolti in due PRIN (*Progetti di Rilevante Interesse Nazionale*): il primo coordinato dall'U-

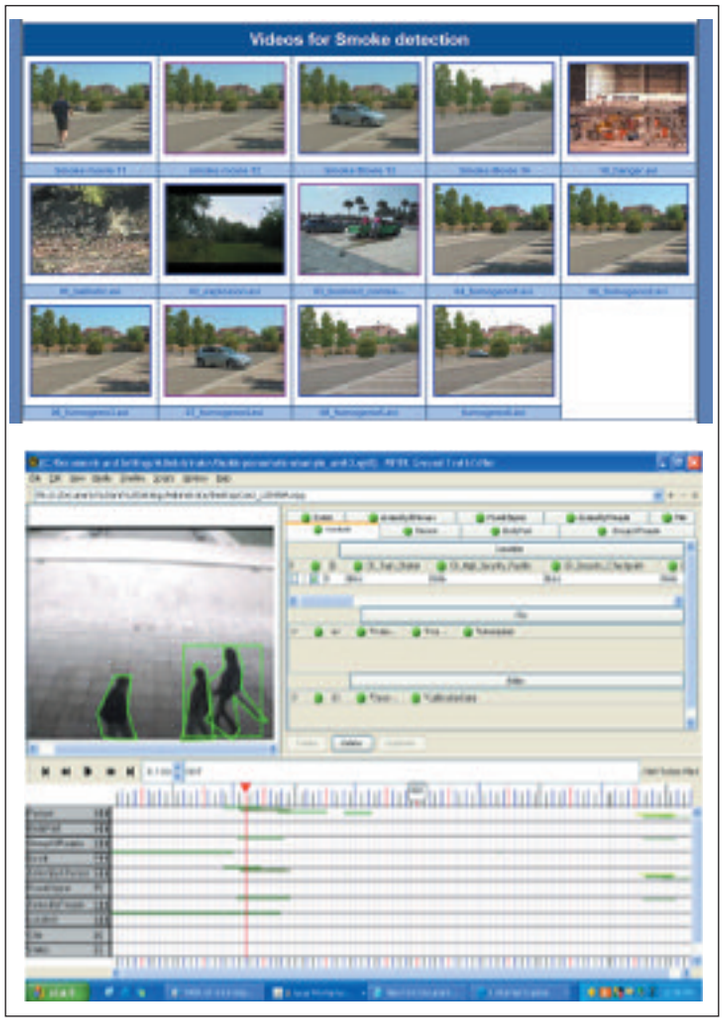


FIGURA 7
In alto: VISOR - repository di video con annotazioni XML e MPEG7. In basso: esempio di annotazione con il tool VIPER

[3] OPENCV: http://opencvlibrary.sourceforge.net/
[4] PETS: http://www.cvg.cs.rdg.ac.uk/slides/pets.html
[5] CAVIAR: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/
[6] VISOR: http://imagelab.ing.unimore.it/visor
[7] VISIT2008: http://imagelab.ing.unimore.it\visit2008
[8] Il Gruppo Italiano di Ricercatori in Pattern Recognition (GIRPR) è il capitolo italiano della International Association in Pattern Recognition (IAPR). Raccoglie ricercatori accademici ed industriali che si occupano di pattern recognition, computer vision e trattamento di dati multimediali in Italia. Per conoscere conferenze, eventi, attività e iscriversi al GIRPR, consultare: GIRPR: http://www-girpr.dis.unina.it/IAPR: http://www.iapr.org/

niversità Udine, assieme a Roma e Pavia[9] si occupa di Ambient Intelligence ed in particolare di gestione, riconfigurazione ed analisi di prestazioni di reti di sensori e telecamere. Il secondo[10] coordinato da Modena con le Università di Firenze e Palermo e con la Fondazione Kessler di Trento e' invece focalizzato su algoritmi innovativi di visione per la Video Sorveglianza; si vogliono proporre nuovi paradigmi per sistemi di sicurezza compatibili con i vincoli di privacy. Da una parte si vuole garantire la privacy degli individui, ad esempio eliminando in tempo reale dai video i caratteri identificativi biometrici quali i volti, dall'altra si vogliono generare allarmi automatici per eventi di interesse a partire da piattaforme di nuove generazione con telecamere fisse, mobili, PTZ e reti di sensori.

I gruppi di ricerca coinvolti in questi progetti PRIN sono attivi da anni nel settore. Per quel che riguarda l'Università di Udine ricordiamo inoltre il progetto Europeo HamLET (*Hazardous Material Location and Tracking*) e il progetto NATO AMSS-CT (A*dvanced Multisensory Surveillance Systems for Combating Terrorism*). Anche l'Università di Modena e Reggio Emilia coordina un progetto NATO (programma *Science for Peace*) chiamato BESAFE (*Behavioral lEarning in Surveilled Areas with Feature Extraction*) in collaborazione con la Hebrew University di Israele e collabora con l'Università di Firenze nel progetto Europeo VIDI-VIDEO per il citato portale VISOR. Esistono molti altri gruppi attivi in Italia: presso il DIEE di Cagliari si lavora sull'integrazione di sistemi di video sorveglianza e di autenticazione biometrica (fingerprint) anche in collaborazione con i Ra.C.I.S. dell'arma dei Carabinieri. Presso i laboratori IMM e ISSIA del CNR di Lecce e Bari rispettivamente si studia l'impiego di nuove telecamere (*range sensor*) per l'analisi delle immagini di profondità.

Il gruppo TeV della fondazione Kessler di Trento da anni sviluppa sistemi di Video Sorveglianza: si ricorda il progetto pilota finanziato dal Ministero dei Trasporti SCOCA per l'analisi del traffico veicolare in incroci cittadini.

IL VIPLAB di Verona sta lavorando nel progetto Europeo SAMURAI, per la coordinazione uomo-macchina nella gestione di reti di telecamere. Il Laboratorio LABVIS di Bologna infine da anni svolge attività di ricerca per la video sorveglianza e l'augmented reality nell'esplorazione di luoghi pubblici e musei.

Il Gruppo di Telecomunicazioni di Genova è attivo da anni nella videosorveglianza con risultati di eccellenza internazionali. Tra i diversi progetto si ricorda il VICOM (*Virtual Immersive Communication*), un grande progetto triennale finanziato dal MIUR che oltre alle università italiane ha coinvolto il CNIT, il CNR e l'ISCTI.

Pur ben lungi dall'essere esaustivo, quest'elenco ha voluto mostrare il ricco e dinamico panorama di ricerca italiano nel settore, coordinato dal GIRPR e in stretto legame con la ricerca internazionale nell'ambito dell'IAPR e nell'IEEE.

La ricerca in Video Sorveglianza è estremamente fruttuosa in termini accademici, sia per progetti, sia per le rilevanti pubblicazioni internazionali e sta guidando la ricerca applicata in stretta collaborazione con aziende italiane e internazionali. Certi che quest'attività sarà trainante nel settore ICT nel prossimo futuro, non possiamo far altro che augurarci un ulteriore coinvolgimento nazionale degli enti di ricerca e governativi, per rafforzare un ambito scientifico che davvero può dirsi un fiore all'occhiello della ricerca italiana ed internazionale.

[9] "Ambient intelligence:event analysis, sensor reconfiguration and multimodal interfaces".

[10] "FREE SURF Free surveillance in a privacy respectful way".

Bibliografia

[1] Forsyth D.A., Ponce J.: *Computer Vision: A Modern Approach*. Prentice Hall 2002.

[2] Bishop C.: *Pattern Recognition and Machine Learning*. Springer 2006.

[3] Shu C., Hampapur A., et al.: *IBM smart surveillance system (S3): a open and extensible framework for event based surveillance*. IEEE Conf. AVSS, 15-16 Sept. 2005, p. 318-323.

[4] Ramesh V.: *Real-time vision at Siemens Corporate Research*. IEEE Conf. AVSS 2005, 15-16 Sept. 2005, p. 300-305.

[5] Gonzalez R., Woods R.: *Digital Image Processing*. Prentice Hall, 2002.

[6] Calderara S., Cucchiara R., Prati A.: *Bayesian-competitive Consistent Labeling for People Surveillance*. In: IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 30, n. 2, 2008, p. 354-360.

[7] Lowe D.G.: Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, Vol. 60, n. 2, 2004, p. 91-110.

[8] Stauffer C., Grimson W.: Learning patterns of activity using real-time tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, n. 8, 2000, p.747–757.

[9] Prati A., Mikic I., Trivedi M.M., Cucchiara R.: Detecting Moving Shadows: Algorithms and Evaluation. In: *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, n. 7, July, 2003, p. 918-923.

[10] Yilmaz A., Javed O., Shah M.: Object tracking: A survey. *ACM Computing Survey*, Vol. 38, n. 4, Dec. 2006.

[11] Cucchiara R., Prati A., Vezzani R.: *Ambient Intelligence for Security in Public Parks: the LAICA Project*. In: Proc. of IEE Intern. l Symp. on Imaging for Crime Detection and Prevention, London, UK, 2005.

[13] Moeslund T.B., Hilton A., Krüger V.: A survey of advances In vision-based human motion capture and analysis. In: *Computer Vision and Image Understanding*, Vol. 104, n. 2, 2006, p 90-126.

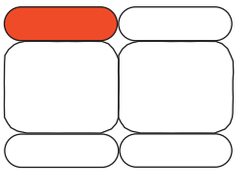
[14] Cuntoor N.P., Yegnanarayana B., Chellappa R.: Activity Modeling Using Event Probability Sequences. *IEEE Trans.on Image Processing*, Vol. 17, n. 4, 2008, p.594-607.

RITA CUCCHIARA è Professore Ordinario di Sistemi di Elaborazione presso la Facoltà di Ingegneria di Modena; è docente di Calcolatori Elettronici e di Visione Artificiale. È Vice Preside della Facoltà e coordinatore del corso di "Computer Engineering and Science" della Phd School in ICT. Coordina il Laboratorio ImageLab che svolge attività di ricerca in Visione Artificiale, Pattern Recognition e Multimedia, per la video sorveglianza, la sicurezza, la visione robotica e per sistemi multimediali per Digital library di immagini e video in diversi progetti nazionali ed internazionali. È autore di più di 150 lavori su riviste e congressi internazionali. Per riferimenti alla ricerca e alla pubblicazioni si veda: http://imagelab.ing.unimore.it
E-mail: rita.cucchiara@unimore.it

TOPI, FINESTRE, ICONE E TENDINE
LE RADICI DELL'INTERFACCIA GRAFICA

Silvio Hénin
Massimo Zaninelli

L'interfaccia grafica, che oggi tutti usiamo per interagire con il nostro calcolatore personale, è il risultato di un'evoluzione iniziata nei primi anni della guerra fredda e continuata grazie a cospicui finanziamenti governativi, alla creatività di brillanti ricercatori e alle esperienze accademiche e imprenditoriali svolte al MIT, allo Stanford Research Institute e allo Xerox PARC. Pur nell'eterogeneità dei fini perseguiti e dei mezzi usati, è possibile tracciare un'ideale *fil rouge* che collega le diverse tappe di questo sviluppo.



1. INTRODUZIONE

Quarant'anni fa, il 9 dicembre 1968, un migliaio di persone affollava la Brooks Hall del Convention Center di San Francisco. Davanti al pubblico si ergeva uno schermo di sei metri che, quando si accese, mostrò in primo piano il viso del relatore incastonato tra scritte, simboli e schemi. Nelle successive due ore i partecipanti assistettero a qualcosa di mai visto prima: le righe di testo sullo schermo venivano evidenziate, modificate, cancellate, reinserite; le strutture ad albero dei dati si espandevano e si comprimevano, da alcune parole evidenziate si "saltava" ad altre informazioni ad esse logicamente collegate. I primi piani mostravano anche le mani del relatore, intente a manipolare una tastiera e altri strani oggetti, tramite i quali sembrava comandare gli eventi. Al termine vi fu un attimo di silenzio, seguito da un'ovazione di alcuni minuti. Paul Saffo, dello *Institute for the Future* di Palo Alto, commentò: "*...fu come se un UFO fosse atterrato sul prato della Casa Bianca*". La presentazione, costata quasi 15.000 $, passò nella mitologia informatica come "l*a Madre di tutte le Demo*"[1]. Per capire il suo effetto dirompente basta ricordare che quasi tutti i calcolatori dell'epoca operavano ancora in *batch-mode*[2] ed era quasi impossibile interagire con essi 'in tempo reale' [1, 2, 3].
L'evento si era svolto nel corso della *Fall Joint Computer Conference*, il relatore si chiamava Douglas Engelbart, ciò che aveva mostrato era lo NLS (*oN Line System*[3]) e uno degli strani oggetti nelle mani di Engelbart era il primo *mouse* [1, 2, 3]. La storica presentazione fu

[1] Il termine gergale anglosassone "demo" è un'abbreviazione di "demonstration" nel senso di dimostrazione, azione dimostrativa, presentazione pubblica.

[2] Il *batch-processing*, o *elaborazione in lotti*, era la precedura utilizzata comunemente nei calcolatori fino agli anni '70; i vari "compiti", e i relativi dati di ingresso, venivano introdotti nel computer che poi li eseguiva secondo una sequenza pre-assegnata. Non era possibile nessuna interruzione dei processi, né alcuna modifica dei dati, fino a quando questi non erano terminati [23].

[3] L'acronimo è imperfetto, ma occorreva distinguere tra *On-* (NLS) e *Off-Line Information System* (FLS).

uno dei segnali di quella "rivoluzione informatica" che avrebbe caratterizzato i successivi vent'anni, resa possibile dalla miniaturizzazione, dalla drammatica riduzione dei costi della tecnologia e dall'aumento della velocità di calcolo, previste da Gordon E. Moore pochi anni prima[4]. Più ancora dello sviluppo tecnologico, furono determinanti il cambio di "destinazione d'uso" dei computer – dalla pura esecuzione di calcoli matematici alla più generale elaborazione di informazioni (messaggi, testi, immagini, suoni) – e la realizzazione di un sistema di interazione uomo-computer più semplice e diretto, comprensibile da chiunque.

2. MIT, WHIRLWIND E SAGE

Nei primi anni '50 vi erano, in tutto il mondo, solo una manciata di calcolatori elettronici, dedicati quasi esclusivamente ai calcoli per la fisica, l'ingegneria e la statistica. Il clima dell'epoca è ben rappresentato dall'aneddoto secondo cui il presidente di IBM, Thomas Watson Sr., avrebbe affermato nel 1943: "*credo che al mondo ci sia mercato forse per cinque computer*[5]". Presto però la guerra fredda contribuì a smentire la presunta dichiarazione, accelerando drasticamente lo sviluppo e la diffusione delle macchine per il calcolo automatico. Nello scenario di intenso riarmo e di reciproco sospetto tra le superpotenze si intraprese lo sviluppo di SAGE (*Semi-Automatic Ground Environment*): un sistema di difesa aerea che connetteva la rete di avvistamento radar con la guida dei caccia intercettori, mediante un calcolatore elettronico digitale, un *super-computer* a valvole battezzato Whirlwind (riquadro a p. 55). Si trattava della prima macchina elettronica digitale in grado di operare in "tempo reale", che elaborava i dati man mano che questi le arrivavano dalle diverse fonti (radar, avvistatori, centri di comando) e trasmetteva istantaneamente alla difesa aerea la presunta rotta dei bombardieri nemici. I due padri di Whirlwind, Jay Forrester e Robert Everett, previdero che l'inevitabile intervento umano avrebbe dovuto essere semplice ed immediato e utilizzarono perciò un terminale video che riproduceva la traccia radar e ne visualizzava le coordinate calcolate dal computer. Per l'identificazione del bersaglio Everett realizzò il primo organo di *input* semplice ed ergonomico, la *light-gun*, una specie di pistola fotoelettrica che rilevava la traccia luminosa sullo schermo e ne inviava i dati al calcolatore [4].

3. ARPA, IPTO E J. C. R. LICKLIDER

Il 4 ottobre 1957 l'Unione Sovietica lanciava in orbita il primo satellite artificiale, lo Sputnik, assestando un duro colpo alla presunta superiorità tecnologica degli Stati Uniti. Il Presidente Dwight D. (Ike) Eisenhower e il Dipartimento della Difesa corsero ai ripari varando nuovi progetti ed investendo enormi somme di denaro per sostenerli. Per rimediare all'annosa e deleteria competizione tra le tre Forze Armate fu fondata l'*Advanced Research Project Agency* (ARPA), con il compito di stimolare e finanziare progetti di ricerca per la difesa. Una sezione dell'ARPA, l'*Information Processing Technology Office* (IPTO) si concentrò sullo sviluppo delle tecnologie informatiche e il suo primo direttore fu Joseph Carl Robnett Licklider [5].
Nel 1960, ben prima che la tecnologia necessaria fosse disponibile, Licklider aveva pub-

[4] G. E. Moore, "*Cramming more components onto integrated circuits*", Electronics Magazine, 38 (8), 19 aprile 1965. Secondo la cosiddetta "legge di Moore", il numero dei componenti di un circuito integrato sarebbe raddoppiato ogni anno e il costo per componente sarebbe diminuito in proporzione. Questa regola empirica, più volte reintepretata e talvolta travisata, si è dimostrata sostanzialmente valida fino ad oggi.

[5] La frase è probabilmente apocrifa, pubblicata per la prima volta in: C. Morgan "*Facts and fallacies: a book of definitive mistakes and misguided predictions*", Londra, Corgi, 1984. Alcuni la attribuiscono ad altri 'padri' dell'informatica, come l'inglese Douglas Hartree o l'americano Howard Aiken, il che lascia comunque intendere che si trattava di un'opinione piuttosto diffusa. Per una trattazione vedi http://en.wikipedia.org/wiki/Thomas_J._Watson o http://baris.typepad.com/venture_capitalist/-2007/11/the-tj-watson-p.html.

blicato "*Man–Computer Symbiosis*", un articolo in cui delineava un nuovo paradigma d'interazione tra l'uomo e il calcolatore e una rivoluzionaria destinazione d'uso di questo [5, 6, 7]. Alcuni tra gli obiettivi enunciati restano di grande attualità: "*...fare in modo che i computer stimolino la formulazione dei pensieri ... mettere uomini e computer nelle condizioni di cooperare nel compiere scelte complesse senza dipendere rigidamente da programmi predeterminati;* ... [creare tra uomini e computer] *un rapporto "simbiotico" in cui l'uomo stabilisce gli obiettivi, formula le ipotesi, fissa i criteri ed esegue le valutazioni; il computer svolge i compiti di routine necessari a prendere le decisioni nell'ambito tecnico e scientifico*". Nella visione di Licklider, il computer doveva diventare una specie di "collega artificiale" o "collaboratore interattivo". "*I computer odierni sono progettati primariamente per risolvere problemi pre-formulati o elaborare dati secondo rigide e predeterminate procedure* – proseguiva Licklider, riferendosi al *batch-processing* – *Se invece si potesse creare una relazione simbiotica tra l'uomo e una macchina veloce, capace di cercare le informazioni ed elaborarle ...*" [7]. Licklider non poteva certo immaginare che un costoso calcolatore dell'epoca potesse essere dedicato ad un solo utilizzatore, che lui identificava ancora come un ingegnere o uno scienziato. Peraltro, il direttore dell'IPTO non diede contributi personali alla realizzazione di questa sua visione, ma, come afferma Alan J. Perlis "*Quello che fece fu di tradurre in azione, in molte sedi diverse, l'immaginazione e l'energia di persone che vedevano come il computer, di per sé, richiedeva un dinamismo ... indipendente dalle applicazioni per le quali era stato inventato*" [8].

4. TIME SHARING E MINICOMPUTER

Ormai la tecnologia necessaria a rendere il computer "personale e interattivo" non era più così remota. Due novità tecnologiche, a volte in competizione, a volte in sinergia tra loro, erano apparse già nella prima metà degli anni '60: lo *time-sharing* e i *minicomputer*. Lo time-sharing permetteva di condividere le costose risorse di un grande calcolatore tra più utenti simultaneamente, grazie ad un sistema operativo che poteva assegnare segmenti di tempo ad ogni diverso compito o utente. La minore velocità d'azione dell'utente umano rispetto a quella di elaborazione del calcolatore fa sì che l'utilizzatore abbia la sensazione che l'intera macchina stia lavorando solo per lui. I primi esperimenti di time-sharing iniziarono al MIT sotto l'impulso di John McCarthy, realizzatore del *Compatible Time Sharing System* (CTSS, del 1961); seguirono il JOSS alla RAND (1963)[6] e il progetto GENIE all'Università di Berkeley (1964) [2, 23]. Licklider era un convinto sostenitore dello time-sharing e vi profuse molte delle risorse disponibili all'IPTO [5]. La seconda novità fu la nascita dei minicomputer[7], calcolatori abbastanza piccoli e di prezzo abbastanza contenuto da poter essere installati in qualunque istituto di ricerca o piccola azienda. I primi minicomputer furono prodotti dalla *Digital Equipment Corporation* (DEC) quali il PDP-8 del 1964, venduto a soli 16.000 $[8]. Anche sui minicomputer, per ammortizzare ulteriormente i costi, furono implementati sistemi di *time-sharing* per servire più utenti.
L'uso "personale" del computer da parte di non-addetti rendeva però necessario lo studio di nuovi paradigmi d'interazione con l'utente. Le telescriventi avevano già permesso l'interazione in tempo reale sostituendo così le schede perforate, ma erano lente, rumorose e limitate ai soli caratteri disponibili sulla tastiera. Gli schermi a tubo catodico (CRT) inaugurati con Whirlwind erano promettenti, ma ancora molto costosi. Un'interfaccia grafica che utilizzasse organi dedicati, come le *light-pen* (eredi della *light-gun* di SAGE), i *joystick*, le tavolette grafiche e i *touch-screen*, erano ancora un pri-

[6] La RAND (Research and Development), è un'agenzia di ricerca fondata nel 1946 con il finanziamento dall'Aeronautica USA. Da non confondere con la Rand Kardex Co. fondata nel 1898, incorporata nella Remington-Rand (1927), che confluirà nella Sperry-Rand (1955) e infine nella Unisys (1986).

[7] Il prefisso 'mini' ha forse un legame culturale con due oggetti simbolo dell'epoca: la Mini Morris (1959) e la minigonna (1965).

[8] PDP è acronimo di 'Programmed Data Processor', si noti l'eliminazione dei termini *computer* e *calculator*, proprio a sottolineare la nuova e più generale destinazione d'uso.

vilegio di pochi addetti; il *mouse* non era ancora stato inventato. Ciò nonostante, alcune interfacce utente meritano di essere ricordate per la loro relativa diffusione e per la rilevanza storica: il LINC, progettato nel 1964 da Wes Clark al Lincoln Lab (MIT) per l'elaborazione di dati bio-medici, usava un terminale grafico su cui rappresentare in contemporanea testi, tabelle e grafici; il PROMIS (*PRoblem Oriented Medical Information System*) dell'Università del Vermont, del 1967, usato per registrare dati clinici in ambiente ospedaliero e dotato di un terminale grafico con *touch-screen*; il sistema PLATO, nato all'inizio degli anni '60 all'Università dell'Illinois, usato come supporto alla didattica, che fu nel 1972 il primo a sfruttare schermi piatti al plasma; infine il GRAIL della RAND, del 1969, che usava uno schermo CRT ed una tavoletta grafica per il disegno dei diagrammi di flusso [2, 8].

4.1. SRI, Engelbart e NLS

Nei decenni 1960 – 1970, a sud di San Francisco, nella Silicon Valley, cominciarono a sorgere le nuove imprese che assumeranno una posizione dominante nell'evoluzione degli strumenti informatici: Intel, Apple, Oracle, per citarne solo alcune. Forti legami umani e culturali univano queste imprese ai centri universitari dedicati allo studio dei calcolatori, come l'Università della California a Berkeley, a Los Angeles (UCLA), a San Diego (UCSD), oltre ad istituti di ricerca *no-profit*, come la RAND e lo *Stanford Research Institute* (SRI). Nel 1962 allo SRI approdò Douglas Engelbart, sognatore col fermo proposito di affrontare il problema della "gestione della complessità", e che immaginava "*Un sistema integrato e comprensivo che colleghi gli aspetti tecnici e sociali della tecnologia del calcolatore personale* [sic]" [2, 10]. A quanto lui stesso ricorda, Engelbart era stato fortemente influenzato, oltre che da Licklider, anche dalla lettura di un articolo di Vennavar Bush, "*As we may think*"[9] in cui si prefigurava un congegno elettromeccanico (MEMEX: *MEMory EXpander*), che avrebbe permesso ad un ricercatore di gestire la crescente mole di pubblicazioni scientifiche, annotandole e creando "percorsi" tra le diverse fonti. [2, 11] Engelbart battezzò il suo progetto "*Augmenting Human Itellect*", prefigurando, diversamente dal "collega artificiale" di Licklider, la funzione del computer come quella di un "potenziatore" dell'intelletto umano, una specie di 'protesi della mente' e preconizzando una "co-evoluzione" di uomo e computer. Con i finanziamenti concessi, prima dall'Aeronautica e poi dall'IPTO, Engelbart fondò, presso lo SRI, il suo *Augmenting Research Center* (ARC) [2, 8, 9]. L'interfaccia immaginata da Engelbart non puntava tanto a semplificare l'interazione tra uomo e computer: l'odierno concetto di *user-friendliness* era per lui un falso obiettivo e il suo prototipo di "utente" era fondamentalmente quello di un informatico-programmatore. Coerentemente, allo ARC si perseguì l'efficienza dell'interazione uomo-computer, piuttosto che la facilità d'uso. Oltre alla normale tastiera, una tastierina aggiuntiva (il *chord-keyset* con cinque tasti per produrre una trentina di comandi), un interruttore "a ginocchio" e il mouse richiedevano all'utente di coordinare i movimenti di dieci dita e una gamba, oltre a memorizzare un lungo elenco di comandi, fino a 5.000 [2, 3]. Il risultato fu che "*era impossibile insegnare agli utenti come usare lo NLS*" [9].

Dopo la presentazione del 1968, allo ARC iniziò un triste declino. Gli obiettivi e le priorità divennero sempre più vaghi e la comunicazione tra Engelbart e i suoi collaboratori sempre più difficile. Nel 1975 il braccio destro di Engelbart, William English, lasciò lo SRI, seguito presto da molti altri valenti ricercatori. L'anno successivo lo ARC venne venduto alla privata Tymshare Corporation e praticamente smise di esistere [2, 9]. Nonostante il fallimento, l'influenza dello NLS sui successivi sviluppi dell'interfaccia utente fu certamente fondamentale.

4.2. Xerox PARC e Alto

Non pochi dei transfughi dall'ARC approdarono ad un nuovissimo centro ricerche sorto a Palo Alto. Nel 1968 la Xerox, produttore di fotocopiatrici di cui aveva praticamente il monopolio, pensò di lanciarsi nel settore dei calcolatori elettronici, con l'ambizioso ma

[9] Vannevar Bush, consigliere scientifico di F. D. Roosevelt, prevedeva che "U*n giorno potremo estrarre* [click off] *informazioni da una macchina con la stessa sicurezza con cui ora usiamo un registratore di cassa*" [11].

non ben definito obiettivo di "*controllare l'architettura dell'informazione*". Nacque così il *Palo Alto Research Center* (*Xerox PARC*) che diede un contributo determinante alla nascita della prima vera interfaccia grafica [12, 22] e Robert Taylor ne fu nominato direttore tecnico. Già a capo dello IPTO, Taylor aveva dato grande impulso allo sviluppo di ARPAnet[10], dimostrandosi abilissimo reclutatore di persone capaci e creative [12]. Grazie a lui giunsero al PARC ricercatori dal MIT, dall'Università dello Utah, da Berkeley, oltre a una decina di persone dall'ex-gruppo di Engelbart. Questi ultimi tentarono di trasferire al PARC idee e tecnologie nate dalle precedenti esperienze, concentrandosi soprattutto sul progetto di time-sharing POLOS (*Parc On Line Office System*). Altri, come Butler Lampson, Charles P. Thacker e lo stesso Taylor, puntarono sullo sviluppo di un piccolo computer veramente "personale" [9, 12, 22].

A rinforzare il gruppo di Palo Alto giunse anche un giovane idealista: Alan Kay. Le sue precedenti esperienze e l'influenza della scuola pedagogica di Jean Piaget lo avevano portato ad immaginare un calcolatore piccolo, portatile e di bassissimo costo (meno di 500 $), tanto facile da usare da potere essere affidato ad un bambino. [12, 13, 22] Conscio che la tecnologia necessaria non era ancora disponibile, ma convinto che presto lo sarebbe stata, Kay sviluppò ulteriormente la sua idea al PARC, battezzandola "Dynabook" e immaginando che l'interazione con il computer avvenisse manipolando oggetti virtuali sullo schermo, senza alcun comando testuale. Dynabook avrebbe dovuto poter essere "*posseduta da chiunque ... per gestire tutte le* [sue] *necessità di informazione*" e operare in modo che "*non ci fosse alcuna pausa discernibile tra la causa e l'effetto*", quindi non un time-sharing, ma un calcolatore "*dedicato completamente all'utente*". Questi non era più il *knowledge-worker* di Licklider e di Engelbart, ma la segretaria, il commerciante, lo studente [2, 8, 9, 12, 22].

Le idee di Alan Kay ottennero il consenso di numerosi colleghi del PARC e il progetto fu sviluppato da Lampson in un memorandum del 1972, intitolato "*Why Alto*". Le difficoltà tecniche di conciliare costo, dimensioni e prestazioni non furono poche e la loro soluzione condusse a numerose innovazioni software e hardware. Il risultato finale fu Alto, il primo computer per uso personale con interfaccia grafica che già comprendeva molte delle caratteristiche a cui oggi siamo abituati, quali le finestre sovrapponibili, le icone, i *menu* e l'uso del mouse [riquadro a p. 56]. Alto, il cui costo risultò ben più elevato del sogno di Alan Kay, non fu mai commercializzato; il suo successore, lo Xerox Star, fu messo in commercio nel 1981 a 16.595 $. Fu comunque un insuccesso; nonostante l'innovativa interfaccia utente e le notevoli caratteristiche, era una macchina anacronistica: computer del costo di un migliaio di dollari erano già sul mercato e proprio nel 1981 fu lanciato il PC IBM, che non aveva neppure lontanamente le potenzialità di Star, ma grazie al basso costo ed alla fama del produttore conquistò il mercato e divenne uno standard *de facto* [12, 14, 22].

4.3. Il Personal Computer

Poco lontano da Stanford e da Palo Alto, qualcosa di importante era infatti avvenuto, quasi all'insaputa dei principali attori dell'informatica professionale, accademica e commerciale. Nel 1971 la Intel aveva lanciato il primo processore a 4 bit "in un solo chip", il "4004" progettato dall'italiano Federico Faggin e da Ted Hoff e destinato alle prime piccole calcolatrici elettroniche, seguito nell'anno successivo dal più potente "8008" a 8 bit. Il microprocessore trovò il suo primo successo nell'ambiente degli hobbisti (oggi diremmo degli *hackers*): una variegata popolazione di radioamatori, appassionati di alta fedeltà e modellisti ferroviari, che rincorrevano l'ultimo *gadget* elettronico di moda e si riunivano nella Silicon Valley allo Homebrew Computer Club (Club del computer

[11] ARPAnet fu la prima rete a commutazione di pacchetto per la connessione tra calcolatori che permetteva lo scambio di informazioni tra macchine con diversi sistemi operativi e linguaggi. Entrata in funzione nel 1969, ARPAnet collegava inizialmente numerosi computer dedicati alla ricerca per progetti militari. Dalla tecnologia ARPAnet si sviluppò l'odierna Internet. Per la storia di ARPAnet ed Internet vedi, per esempio, J. Abbate, "*Inventing the Internet*", Inside Technology, 1999, oppure K. Hafner, "*When the Wizards stay up late: the Origins of the Internet*", Simon & Schuster, 1998.

fatto-in-casa). In questo curioso ambiente nacque e si diffuse lo storico ALTAIR 8800, il primo *microcomputer* alla portata praticamente di chiunque[11]. Subito sorsero numerose imprese che ne commercializzavano copie più o meno identiche, già assemblate o in scatola di montaggio, ma il mercato degli hobbisti fu presto saturo e la maggior parte di queste imprese sparì in un paio d'anni. L'idea fu però ripresa da Steve Jobs e Steve Wozniak che idearono una macchina completa, dotata di tutti gli organi necessari per un utilizzo immediato - salvo il software applicativo: l'Apple II (1977), venduto al prezzo di 1.295 $[12], dotato di una tastiera incorporata, un registratore a cassette come memoria di massa e che usava un comune televisore come display. Due anni dopo viene lanciato Visicalc, il primo foglio di calcolo, e le vendite di Apple esplosero [1, 14, 16, 17].
Jobs e Wozniak capirono presto che, per allargare il loro mercato, occorreva uscire dal settore amatoriale e puntare sui professionisti e sulle piccole imprese, fino ad allora esclusi dal mondo dell'informatica. Una delle strade percorribili era proprio il miglioramento dell'interfaccia utente, così da aggirare lo scoglio dell'addestramento. I due Steve fecero proprio, forse inconsciamente, l'imperativo di Ted Nelson, che nel suo "*Computer Lib / The Dream Machine*" così pontificava in quegli anni: "*Potete e dovete imparare* [ad usare] *il computer ora*!" [15]. Nel 1978 nacquero i progetti Lisa e Macintosh. Nel 1983 Lisa[13] fu indubitabilmente il primo "personal" commerciale dotato di interfaccia grafica, ma ancora troppo costoso[14], lento e con una limitata dotazione di software; l'anno dopo fu soppiantato da Macintosh, meno caro, più veloce e con una discreta scorta di applicativi, e fu un clamoroso successo [1, 14].
Molto inchiostro è stato versato sul presunto plagio di Apple ai danni di Xerox, ipotesi ispirata dalla visita che Steve Jobs fece al PARC nel dicembre 1979, durante la quale esaminò non solo le funzionalità di Alto, ma anche le soluzioni tecniche adottate. Alcuni reduci dal PARC furono effettivamente arruolati da Apple, ma non sembra che abbiano avuto ruoli fondamentali in Lisa e Macintosh. Steve Wozniak riassume così la questione: "*Steve Jobs sollevò il caso con i dirigenti di Xerox, sostenendo che essi avevano una grande tecnologia, ma che Apple l'aveva resa disponibile tanto da cambiare il mondo*" [16]. A sua volta Apple, nel 1988, denunciò Microsoft per presunta incorporazione dell'aspetto visivo dell'interfaccia di Macintosh in quella di Windows.
Il resto è cronaca. L'odierna GUI (*Graphic User Interface*) e il paradigma WIMP (*Windows, Icon, Mouse, Pointer*) che la caratterizza sono più o meno stabili da quasi vent'anni e rappresentano ormai uno standard *de facto* [18]. Muovere il mouse e schiacciarne il tasto produce un effetto prevedibile per analogia, qualunque sia l'applicativo che stiamo eseguendo; l'illusione della scrivania, gli effetti "taglia-e-incolla" e 'trascina e lascia' sono intuitivi, non tanto intrinsecamente, ma perché ormai comuni ai diversi ambienti di lavoro. In qualche modo la simbiosi uomo-computer prefigurata da Licklider sembra sia stata realizzata, anche se in modi e per fini diversi da quelli previsti in origine.

4.4. Una riflessione

Contrariamente a quanto popolarmente tramandato, la diffusione dell'informatica personale e il principale prerequisito per il suo successo, un'interfaccia utente diretta, semplice e intuitiva, non nacquero 'eroicamente' dal nulla in un "garage di casa", ma si plasmarono grazie ad un'intricata rete di influenze culturali, sviluppatasi nel corso di quasi 25 anni nei laboratori di Università e Istituti di ricerca, spesso generosamente finanziati dal governo USA per finalità più o meno diretta-

[11] Al prezzo base di 621 $, lo ALTAIR non era dotato di memoria di massa, né di organi di ingresso e uscita; lo si doveva collegare ad una telescrivente, molto più costosa del computer stesso.

[12] Preceduto nel 1976 dallo sperimentale Apple I.

[13] Il nome "Lisa" era ufficialmente l'acronimo di Local Integrated Software Architecture, ma era anche il nome della figlia di Steve Jobs. Il nome 'Macintosh' origina da quello di una qualità di mele (McIntosh Red), molto popolare negli Stati Uniti, in ovvia relazione col marchio Apple.

[14] Il prezzo di Lisa era 9.950 $. Il primo "Mac" costava solo 2.495 $, grazie anche alle sue intelligenti soluzioni tecniche.

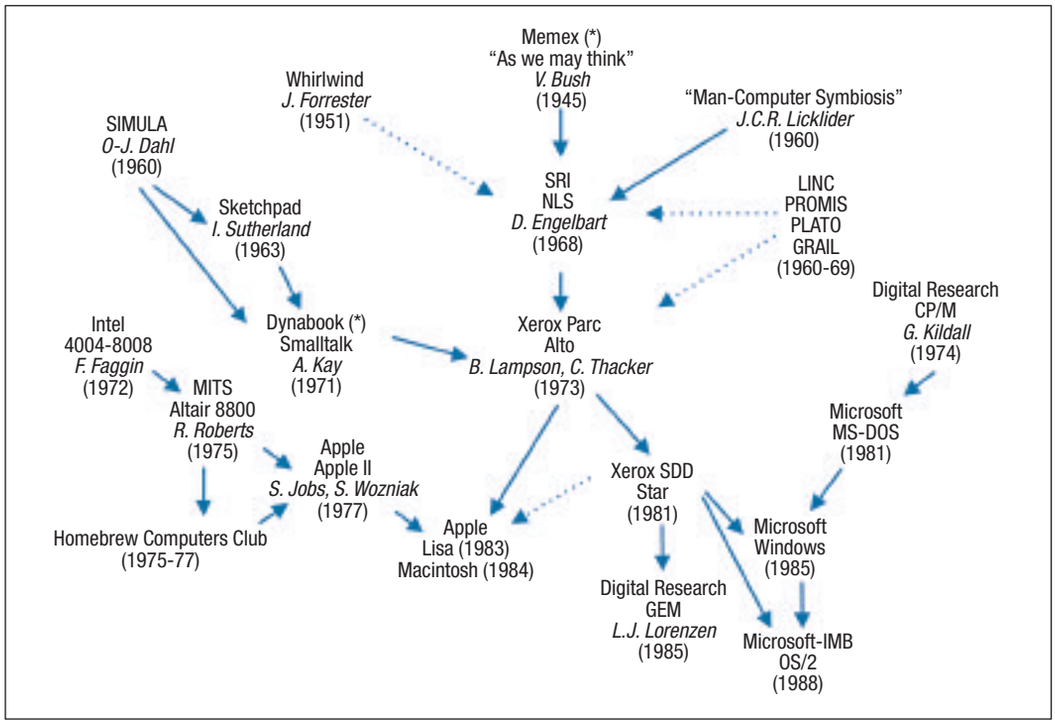


FIGURA1
Albero genealogico della GUI. L'asterisco indica gli strumenti mai realizzati, le frecce punteggiate indicano le influenze probabili, ma non documentate

mente correlate alla Difesa [19 e Figura 1]. Ciò, fra l'altro, dimostra l'infondatezza del mito secondo cui la preminenza degli Stati Uniti nell'informatica dipenderebbe solo dalla capacità imprenditoriale privata, senza alcun sostegno o ruolo da parte degli enti pubblici, mito a volte utilizzato per spiegare i limitati risultati europei nel settore. In realtà, i grandi attori dell'informatica commerciale per lungo tempo non percepirono le potenzialità di un mercato diverso dal mondo dei loro clienti istituzionali – come nelle grandi aziende "storiche" – o non seppero sfruttare le loro stesse innovazioni – come in Xerox. Personaggi come Wozniak, Jobs e Gates, seppero invece capire come la miniaturizzazione e il calo dei costi dell'hardware, unitamente alla GUI, agli applicativi pronti all'uso e ai collegamenti in rete, avrebbero permesso la diffusione globale del personal computer, ma senza le conoscenze e le esperienze accumulate negli anni precedenti, altrove e per fini diversi, difficilmente avrebbero raggiunto i loro risultati.

Bibliografia

[1] Freiberger P., Swaine M.: *Fire in the Valley. The Making of the Personal Computer*. McGraw-Hill, 1984.

[2] Bardini T.: *Bootstrapping: Douglas Engelbart, Coevolution, and the Origins of Personal Computing*. Stanford University Press, 2000.

[3] MouseSite: http://sloan.stanford.edu/MouseSite/1968Demo.html

[4] Redmond K.C.: *From Whirlwind to MITRE. The R&D Story of the SAGE Air Defense Computer*. MIT Press, 2000.

[5] Waldrop M.M.: *The Dream Machine. J.C.R. Licklider and the Revolution that Made Computing Personal*. Viking, 2001.

[6] Tomasello T.K.: *Content Analysis of Citations to J. C. R. Licklider's "Man-Computer Symbiosis, 1960 - 2001: Diffusing The Intergalactic Network"*. Ph.D. Thesis, Florida State University, 2004, http://etd.lib.fsu.edu/theses/available/etd-04102004-214127/unrestricted/tkt_dissertation.pdf.

[7] Licklider J.C.R.: *Man-Computer Symbiosis*. IRE Transactions on Human Factors in Electronics. Vol. HFE-1, Marzo 1960, http://groups.csail.mit.edu/medg/people/-psz/Licklider.html.

[8] Goldberg A., (a cura di): *A History of Personal Workstations*. ACM Press, 1988.

[9] Markoff J.: *What the Dormouse said. How the 60s Counterculture Shaped the Personal Computer Industry*. Viking, 2005.

[10] Engelbart D.C.: *Augmenting human intellect: a conceptual framework*. Stanford Research Institute, Project n. 3578, October 1962.

[11] Vannevar Bush: *As We May Think*. The Atlantic

Monthly, luglio 1945, http://www.theatlantic.-com/doc/194507/bush.

[12] Smith D.K., Alexander R.C.: *Fumbling the future. How Xerox invented, then ignored, the first personal computer*. To Excel, 1999.

[13] Shasha D., Lazare C.: *Out of their minds*. Copernicus, 1995, p. 38-50.

[14] Allan R.A.: *A History of the Personal Computer*. Allan Pub,ishing, 2001.

[15] Nelson T.: *Computer Lib /The Dream Machine*. Hugo's Book Service, 1975.

[16] Wozniak S., Smith G.: *iWoz. How I invented the personal computer, co-founded Apple, and have fun doing it*. W.W. Norton & Co., 2006.

[17] Zaninelli M.: Macchine pensanti crescono. *Le Scienze*, 2002, p. 405.

[18] Wichary M.: Guidebook. *Graphical User Interface Gallery*. http://www.guidebookgallery.org/.

[19] Committee on Innovations in Computing and Communications, National Research Council: *Funding A Revolution. Government Support for Computing Research*. National Academy Press, Washington, D.C. 1999, http://www.nap.edu/-readingroom/books/far/contents.html.

[20] Wadlow T.A.: *The Xerox Alto Computer*. Byte, n. 9/1981, p. 58-68.

[21] Lampson B.W., Taft E.: *Alto User's Handbook*. Xerox Parc, 1979.

[22] Hiltzik M.A.: *Dealers of Lightning. Xerox Parc and the Dawn of the Computer Age*. Harper, 1999.

[23] Ancilotti P., Boari M.: Evoluzione dei sistemi operativi. 1ª Parte: Dai sistemi batch a Linux. *Mondo Digitale*, n. 25, marzo 2008, p. 48-62.

Whirlind e SAGE

Nel corso della II Guerra Mondiale, il Laboratorio Servomeccanismi del Massachusetts Institute of Technology si impegnò a realizzare un simulatore di volo per la Marina degli Stati Uniti. Il progetto fu battezzato Whirlwind (Turbine) e fu affidato a Jay Forrester e Robert Everett. Si tentò inizialmente di utilizzare la tecnologia analogica, ma questa soluzione non avrebbe garantito precisione e versatilità sufficienti. Nel 1946, sulla scorta dell'esperienza di ENIAC – il primo calcolatore elettronico – si pensò di adottare la tecnologia digitale con cui si sarebbe potuto ottenere uno strumento facilmente programmabile per simulare le molte situazioni che un pilota avrebbe potuto incontrare. Diversamente dagli altri calcolatori digitali elettronici dell'epoca, un simulatore di volo avrebbe dovuto però operare 'in tempo reale', accettando ed elaborando immediatamente le informazioni generate dalle azioni del pilota per poter simulare la conseguente reazione dell'aereo; l'interattività in tempo reale e la velocità di risposta erano quindi requisiti indispensabili. Tre eventi storici cambiarono però il destino del progetto Whirlwind: il blocco di Berlino (1948), la prima atomica sovietica (1949) e la Guerra di Corea (1950). La paura di un attacco nucleare spinse gli Stati Uniti a riprendere l'enorme sforzo scientifico-industriale per la difesa che era stato sospeso nel 1946 e il fisico George Valley del MIT istituì un comitato tecnico per lo studio di un sistema destinato ad automatizzare la sorveglianza radar del Nord America. Risultato dello studio fu il progetto di un sistema di difesa aerea, il SAGE (*Semi-Automatic Ground Environment*) che doveva integrare l'acquisizione dei bersagli con la guida dei caccia intercettori. L'elaborazione dei dati e il coordinamento degli interventi doveva essere affidato ad un computer molto veloce e capace di elaborazione in tempo reale; la scelta cadde su Whirlwind. Venne così assicurato il completamento di un progetto che gli stessi militari stavano invece per abbandonare a causa della continua lievitazione dei tempi e dei costi.
Nel 1947 Forrester e Everett avevano completato una prima versione di Whirlwind, una macchina costituita da 5500 valvole che operava in parallelo su parole di 16 bit, ottenendo una velocità teorica 16 volte superiore ai computer coevi che operavano in modo seriale, un bit alla volta. Questo non era abbastanza e negli anni successivi la velocità fu aumentata di quattro volte, grazie anche all'introduzione delle nuove memorie a nuclei magnetici (2000 parole di 16 bit), ideati dallo stesso Forrester; Whirlwind eseguiva 50.000 operazioni al secondo, in tal modo riusciva a elaborare in tempo reale le coordinate inviategli da 70 radar, seguendo contemporaneamente fino a 1000 obiettivi diversi, diventando così il computer più veloce dell'epoca. La sua versione definitiva fu terminata nel 1953; nel settembre dello stesso anno si mise a punto l'intero sistema di integrazione tra le stazioni radar, Whirlwind e le basi degli intercettori tramite collegamenti telefonici digitali e lo si sperimentò a Cape Cod, Massachusetts. I risultati ottenuti nel *Cape Cod Test* permisero di passare alla fase operativa, affidando alla IBM il ruolo di *Prime Contractor* nella costruzione della versione industriale, denominata AN/FSQ-7 (*Army-Navy Fixed Special eQuipment*). La IBM, per inciso, ne trasse ingenti profitti e soprattutto un prezioso bagaglio di know-how innovativo (*hardware, software, project management*).

segue

Gli AN/FSQ-7 furono certamente i più grandi calcolatori mai realizzati, ogni macchina utilizzava 55.000 valvole, occupava 2.000 m^2, pesava 275 t e necessitava di 3 MW di potenza elettrica. Per garantire la sicurezza di funzionamento al 99% – funzionamento che doveva essere ininterrotto, 24 h al giorno per tutti i giorni dell'anno – ogni centrale SAGE ne utilizzava due, uno operativo e uno di riserva in *stand-by*. La parte più interessante ed innovativa di Whirlwind era il pannello di controllo che permetteva all'operatore di interagire con il sistema, visualizzare le mappe radar ricevute, identificare i bersagli, rilevare i dati di posizione e rotta, aggiornati continuamente dal computer, per inviarli ai caccia o ai missili antiaerei. L'addetto interagiva con il computer tramite una *light gun*, un dispositivo di puntamento ottico simile ad una pistola, che veniva puntato sulla traccia radar del bersaglio sul grande schermo a raggi catodici della consolle. Whirlwind acquisiva la posizione della light gun, eseguiva i calcoli, ne presentava i risultati sullo schermo e li inviava in tempo reale agli intercettori per guidarli verso il bersaglio. La consolle incorporava anche un accendino elettrico e un posacenere: *user-friendliness* di altri tempi!

La prima Divisione SAGE fu operativa nel 1959 e nel 1963 esistevano già 22 Centri di Settore e tre Centri di Combattimento. L'intero sistema costò tra 8 e 10 milioni di dollari del 1964, ma una volta terminato, era già obsoleto: i bombardieri strategici avevano lasciato il posto ai missili balistici intercontinentali. Lo Whirlwind del MIT rimase al servizio di SAGE fino al 1959; nel 1973 fu recuperato da un deposito di rottami ed ora è in mostra al Computer History Museum di Mountain View, California.
(http://www.computerhistory.org/about/)

Consolle di comando dello AN/FSQ-7. Si noti la light-gun nella mano dell'operatore

Xerox Alto: il primo Personal Computer?

Alto fu il prodotto dell'ingegno e della dedizione di Chuck Thacker e Butler Lampson dello Xerox Parc, basato sull'originale idea di Alan Kay di costruire un calcolatore piccolo, portatile e poco costoso, ma abbastanza potente e con buone capacità grafiche, destinato all'utilizzo individuale.
Alto poteva essere sistemato in una comune scrivania. Il processore a 16 bit era costituito da parecchi circuiti integrati, quindi non era ancora un "*computer-su-un-chip*", e aveva una velocità di 400.000 istruzioni/secondo. La memoria principale era di 128 kbyte, inclusa la memoria grafica *bitmap*, espandibile a 512 kbyte. Un *hard-disk* da 3 Mbyte completava la macchina. Il sistema operativo permetteva l'esecuzione di 16 attività concorrenti. La grafica era ottenuta rappresentando ogni pixel del video come 1 bit di memoria; quello che oggi è la norma, all'epoca poteva sembrare un esagerato spreco, ma era fondamentale per garantire la velocità di visualizzazione necessaria all'interfaccia grafica.
Peculiarità di Alto era lo schermo col lato maggiore in verticale che simulava la disposizione di una comune pagina a stampa. Il video presentava un'immagine in bianco e nero di 606 per 808 pixel in verticale (60 righe di 90 caratteri). Il mouse aveva tre bottoni con cui si comandavano molte delle applicazioni di Alto, senza ricorrere necessariamente alla tastiera.
Alto era collegabile in rete ad altre macchine simili grazie alla tecnologia Ethernet a 800 kbyte/s. L'architettura prevista per un ufficio medio era composta da più workstations Alto connesse a *file-server* per i documenti condivisi e un *print-server* collegato ad una stampante laser. Anche la rete Ethernet e la stampante laser erano state inventate e sviluppate in Xerox.
La dotazione di applicativi comprendeva: Bravo, un text-editor con il quale si poteva lavorare contemporaneamente su diversi documenti, aperti in differenti finestre; Draw, un programma per creare e modificare immagini costituite da linee, curve e testo; Sil, un programma grafico vettoriale; Markup, un editor grafico

segue

bitmap; Laurel, un programma per la posta elettronica; Trek e Mazewar, due giochi multiutente, usabili in rete e controllati dal mouse. Alto poteva essere programmato in Smalltalk e Lisp, due linguaggi di alto livello orientati agli oggetti. Con Smalltalk si potevano costruire finestre sovrapponibili, posizionabili e dimensionabili.
Il *software* disponibile, lo schermo verticale, la condivisibilità di documenti in rete e la disponibilità di una stampante di alta qualità, dimostrano che la macchina era indirizzata prevalentemente all'editoria personale, un settore costituito da segretarie, editori, scrittori, pubblicisti ecc.. La mancanza di un foglio di calcolo e di un database inibiva ad Alto il mercato commerciale, che avrebbe necessitato di strumenti per la contabilità e l'archiviazione di dati.
Alto non fu mai messo in commercio, ma il costo stimato sarebbe stato di 32.000 $. Ne furono prodotti almeno 1500, quasi tutti utilizzati in Xerox; solo alcuni esemplari furono distribuiti all'Università di Stanford, alla Carnegie-Mellon ed al MIT, diventando lo standard rispetto al quale furono misurate altre macchine dell'epoca [12, 20, 21, 22].

Lo Xerox Alto. La tastiera è estraibile. Si noti il mouse a tre tasti

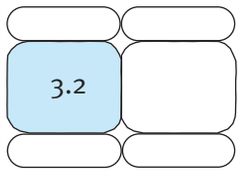
SILVIO HÉNIN - Diplomato in Elettronica e laureato in Scienze Biologiche, ha lavorato per un decennio all'Università di Milano per poi passare alla farmaceutica Roche dove ha implementato sistemi computerizzati per l'information management e retrieval. Fondatore di un'associazione di information scientists (GIDIF-RBM), è stato membro del direttivo dell'Ass. Italiana Documentazione Avanzata (AIDA) e del Comitato per la Documentazione di Federchimica. Cultore della storia della tecnologia, particolarmente di quella del calcolo automatico, collabora con riviste del settore e con il Museo della Scienza e della Tecnica 'Leonardo da Vinci' di Milano.
E-mail: silvio.henin@fastwebnet.it

MASSIMO ZANINELLI - Maturità scientifica e laurea in storia contemporanea, si dedica da anni alla storia della scienza e, in particolare, a quella dell'informatica. Pubblicista, ha collaborato con Il Giornale, con Rete A e, più di recente, con Le Scienze; su quest'ultima ha pubblicato, tra l'altro, vari articoli di storia del calcolo automatico. Ha lavorato con responsabilità crescenti nelle funzioni di comunicazione di Roche, Ketchum, Iveco e attualmente è responsabile della comunicazione corporate presso Chiesi Farmaceutici. In tutte queste imprese ha contribuito a progettare e realizzare sistemi informativi 'web based' per la comunicazione interna ed esterna.
E-mail: massimo.zaninelli@gmail.com

PROGETTAZIONE DEI DATI CON L'USO DEL CONTESTO

Cristiana Bolchini
Giorgio Orsi
Elisa Quintarelli
Fabio A. Schreiber
Letizia Tanca

Gli utenti dei sistemi informativi e di Internet si trovano oggi ad affrontare la necessità di contenere il "rumore informativo" in modo da non essere subissati e confusi da una grande quantità di dati, spesso poco legati al contesto di fruizione. L'attività di adattare i dati alle necessità delle applicazioni può essere assimilata al problema di confezionare un vestito su misura: da qui il nome Data Tailoring. In questo articolo si mostra come le "forbici" usate per questa operazione possano essere rappresentate dal contesto, inteso come un insieme di variabili che caratterizzano l'ambiente specifico nel quale l'applicazione si trova ad operare.



1. INTRODUZIONE

Fin dall'inizio degli anni '80, i sistemi informativi di molte organizzazioni si sono estesi ed evoluti per rispondere alle esigenze di un mercato che richiede sempre più la presenza di un'*impresa allargata* comprendente diversi partner. Anche tecnologicamente, sono state introdotte funzionalità di *Knowledge Management* per permettere una facile condivisione dell'informazione tra i diversi membri dell'organizzazione; l'enorme fonte di informazioni costituita da Internet ha richiesto l'introduzione di strumenti quali XML, ontologie e servizi web, che devono essere utilizzati e gestiti insieme a quelli di altre fonti di dati più tradizionali. Questo aumento di disponibilità, se non opportunamente controllato, porta ad un sovraccarico di dati riducendo drammaticamente i benefici di un ricco sistema informativo. Tuttavia, distinguere l'informazione utile dal rumore, cioè da tutta quella non rilevante allo specifico problema applicativo, non è una questione di facile soluzione; una stessa "unità di informazione" (un oggetto, un'entità) può essere considerata in modo diverso – anche dallo stesso utente – in posti e situazioni diverse; in una parola, in un *contesto* diverso [7, 8, 9].

Si noti che lo scenario appena delineato suggerisce anche un problema ortogonale, altrettanto pressante, che è quello dell'integrazione di dati provenienti da sorgenti eterogenee; esso tuttavia costituisce un argomento di ricerca di per sé assai vasto [3, 10] che – seppur strettamente correlato a quanto esporremo nel seguito – non verrà affrontato in questo lavoro.

Un'ulteriore spinta verso la focalizzazione delle informazioni deriva dall'introduzione di dispositivi mobili; in questo caso la risposta a una stessa interrogazione può cambiare a seconda della localizzazione dell'utente e della situazione ambientale istantanea. Al fine di evitare interruzioni del lavoro dovute alla connessione intermittente alla rete e la ridotta capacità di memorizzazione, tipica dei dispositivi palmari e dei telefoni cellulari, si ha la necessità di mantenere una copia locale della quantità minima di dati in grado di soddisfare le esigenze legate al contesto corrente.

La nozione di contesto, all'inizio emersa nell'ambito di discipline diverse quali la psicologia e la filosofia, sta assumendo oggi notevole importanza anche nelle attività collegate all'informatica. Nel senso comune, il contesto è visto come un insieme di variabili che possono interessare un attore e influenzarne le azioni: un cambiamento del contesto può infatti causare un cambiamento nella rappresentazione mentale che l'attore si crea della realtà, anche se quest'ultima non è cambiata. La parola stessa, derivata dal latino *cum* (con o assieme) e *texere* (tessere), descrive il contesto non solo come un profilo, ma anche come un processo *attivo relativo a come gli uomini utilizzano la loro esperienza all'interno del proprio ambiente per fornirle un significato*.

1.1. Data tailoring basato sul contesto

Scopo di questo lavoro è mostrare come si possa usare il contesto per definire delle viste (*data views*) sui dati di grandi sistemi informativi, ritagliando solamente informazioni rilevanti per una determinata applicazione, in modo da ottenerne un sottoinsieme personalizzato. Da un punto di vista concettuale, le necessità dell'utente di un sistema informativo possono dipendere da due diversi aspetti: il *dominio applicativo*, che rappresenta la realtà in esame, e l'*ambiente di lavoro*, in altre parole, il contesto.
Nel mondo dei sistemi informativi e delle metodologie di progettazione, sono ben noti i classici meccanismi di astrazione attraverso i quali, dall'analisi della realtà di interesse, si genera una rappresentazione concettuale della stessa, e quindi del *dominio applicativo*. L'astrazione che rappresenta il "punto di vista" (*viewpoint abstraction* [11]), e quindi, in buona sostanza, il *contesto di fruizione* dei dati, ha invece ricevuto scarsa attenzione dalla comunità dei ricercatori. In particolare, la strategia di progettazione mista, basata sulla produzione delle viste d'utente e sulla loro successiva integrazione, vede la produzione delle viste come prioritaria all'esistenza della base di dati. Oggi però occorre tener conto della necessità di individuare, e opportunamente configurare, viste su masse di dati già esistenti, delle quali sia eventualmente noto lo schema. In tal caso, l'individuazione del punto di vista non è necessariamente prioritaria alla costruzione della base di dati, e costituisce un'attività a sé, parallela o addirittura successiva alla progettazione della stessa.
Coerentemente, anche l'attuale generazione di sistemi commerciali per la gestione di basi di dati (DBMS), basati su SQL, non permette una definizione ed una gestione esplicita delle informazioni di contesto, ma offre solamente il meccanismo delle *viste di utente* (view) per determinare la porzione di dati gestibile da ciascun utente. Nel seguito, mostreremo come la nostra metodologia sfrutti queste caratteristiche dei DBMS per introdurre la dipendenza dal contesto senza richiedere alcun cambiamento al processo di progettazione delle sorgenti dei dati e neppure richiedere funzionalità particolari al DBMS. Nel paragrafo 4 viene mostrata l'architettura di un sistema dipendente dal contesto che realizza l'approccio metodologico proposto.
Dato uno scenario applicativo, la nostra *metodologia guidata dal contesto* aiuta il progettista nella fase di identificazione dei vari contesti possibili e nella fase successiva di definizione delle viste sull'intero insieme di dati, che risultano rilevanti per ciascun contesto. Tre sono i componenti fondamentali della metodologia:
1. un ***modello di contesto*** che cattura tutti gli aspetti – le cosiddette *dimensioni* – che permettono una rappresentazione implicita dei possibili contesti applicativi;
2. una ***strategia*** che identifica, per ciascuna dimensione in modo indipendente da tutte le altre, una porzione rilevante dell'intero schema dei dati - le cosiddette *viste parziali*;
3. un ***insieme di operatori*** che combinano tra loro le viste parziali in modo da derivare la vista definitiva, associata a ciascun contesto.

1.2. La nozione di contesto nella letteratura

Recentemente, sono stati proposti modelli, a vari livelli di sofisticazione, che permettono di realizzare applicazioni sensibili al contesto (*context-aware applications*). Possiamo elencare i diversi significati che, in essi, sono stati attribuiti alla parola "contesto":
- *Orientato alla presentazione*: il contesto è percepito come la possibilità del sistema di adattare la presentazione dei contenuti a diversi canali informativi o a diversi dispositivi. Questi modelli sono spesso rigidi in quanto

progettati per specifiche esigenze applicative e si basano su un insieme di variabili di presentazione perfettamente noto [16].

❑ *Orientato alla località*: la caratteristica di questi modelli consiste nel permettere di gestire coordinate spaziali e temporali con grande precisione e flessibilità [18].

❑ *Centrato sull'utente*: questi modelli mettono a fuoco l'attività dell'utente: diventa importante la storia e l'evoluzione del suo contesto, e vengono fornite alcune funzioni di ragionamento contestuale; se disponibili, si possono usare funzioni di apprendimento automatico per inferire l'attività dell'utente dalle rilevazioni di sensori [14].

❑ *Basato sul concetto di comunità*: è un approccio interessante che considera il contesto come un insieme di variabili condivise da un gruppo di utenti (*peer*), per i quali costituisce un presupposto alla cooperazione. Diversamente dagli approcci precedenti, si raggiunge una definizione di contesto comune in modo cooperativo [13].

L'uso del contesto per i sistemi informativi e per le basi di dati è stato introdotto in [4, 5, 11, 12, 15, 17, 19], ma è lungi dall'essere stato esplorato approfonditamente. Solo alcune di queste proposte sono orientate al *data tailoring*, mentre per altre lo scopo è di presentare i dati in una prospettiva diversa da contesto a contesto. Senza pretesa di completezza, ci dilungheremo un momento anche su questi altri modelli.

In accordo con l'approccio presentato in [11], Motschnig-Pitrik e Mylpoulos [12] vedono un "information base" come decomposto in sottoinsiemi non disgiunti, che rappresentano i contesti. Un contesto è caratterizzato dal suo contenuto, e d'altra parte i contesti stessi sono oggetti del sistema informativo che, insieme con gli oggetti elementari, costituiscono l'insieme delle *information units*. In questo modo, i contesti vengono considerati alla stregua di qualunque altro elemento del sistema informativo, e il contenimento può essere ricorsivo. Mediante questa astrazione vengono supportati:

- la denominazione e la rappresentazione, basate sul contesto, di tutte le entità concettuali del sistema informativo;
- l'esecuzione di transazioni relativizzata al contesto;
- operazioni di costruzione, manipolazione di contesti e propagazione dei cambiamenti tra contesti contigui.

Per definire un contesto, occorre quindi definire: il suo contenuto, i nomi locali degli oggetti del contesto (o lessico), le regole di autorizzazione relative a diversi utenti e diverse transazioni e, infine, il modo in cui si propagano i cambiamenti tra un contesto e l'altro.

Una tipica applicazione di questo modello, suggerita dagli autori, è lo scenario di cooperazione, dove soggetti diversi utilizzano contesti sia personali sia condivisi per aggiornare in modo cooperativo dei manufatti, come articoli o moduli software. La proposta non si basa sull'ipotesi di un singolo schema globale del database; piuttosto, vi sono schemi diversi, non necessariamente disgiunti, per i diversi contesti. La relativizzazione al contesto di tutte le operazioni risolve il problema dell'inconsistenza, che è molto rilevante nel mondo dei sistemi informativi cooperativi e integrati.

L'approccio di [19] è molto simile al precedente, ma ha il vantaggio di aver introdotto la non-unicità dei nomi per gli oggetti. In questo modo, il modello è in grado di gestire omonimie, sinonimie e oggetti anonimi, risolvendo i conflitti di nome, tipici degli ambienti integrati. Anche in questo lavoro si prevedono primitive di manipolazione dei contesti: in particolare, sono molto ben formalizzate la creazione di contesto e le operazioni insiemistiche sui contesti, tenendo proprio conto della necessità di gestire i conflitti di nome.

In [15] gli autori propongono una visione contestuale di una base di dati relazionale. Diversamente dal nostro approccio, la loro soluzione non suddivide la base di dati in porzioni, bensì valorizza diversamente alcuni attributi, ottenendo varianti (*facets*) diverse per la stessa tupla, in contesti differenti. Un attributo potrebbe addirittura non esistere in qualche contesto. Esempi di applicazione di questo modello sono: un prodotto le cui specifiche cambiano in base al Paese in cui viene commercializzato, oppure una pagina Web che deve essere visualizzata diversamente a seconda del dispositivo. Anche qui un contesto (o *mondo*) è specificato assegnando valori a un opportuno insieme di dimensioni. Una tabella contestuale è quindi una sequenza di tabelle relazionali normali, una per

ogni mondo possibile, alla stregua del modello delle basi di dati temporali, dove però agli istanti di tempo si sostituiscono i vari contesti di validità delle tabelle.
Simile, ma non uguale, è l'approccio di [17]: qui il modello dei dati è analogo a quello dei sistemi OLAP, ma le dimensioni del cubo non sono quelle scelte per l'analisi dei dati, bensì le dimensioni da valorizzare nella definizione del contesto. Gli autori propongono un modello di contesto da associare alle preferenze dell'utente, il quale, all'interno di ogni contesto, può privilegiare determinati valori dei dati. I parametri (dimensioni) del contesto possono essere statici o dinamici: per esempio, il tempo atmosferico, in una certa applicazione, può essere un parametro statico perché il suo valore viene esplicitamente aggiornato, mentre la locazione geografica viene computata in maniera automatica, e quindi varia senza necessità di aggiornamento esplicito. Analogamente al caso OLAP, in questo modello le dimensioni vengono considerate in gerarchia, per esempio la dimensione spaziale potrebbe prevedere i valori ***città, regione, stato***.
Le dimensioni del contesto proposte in questo modello sono abbastanza simili alle nostre; poiché però qui lo scopo è gestire le preferenze dell'utente, la ricerca è orientata a definire in maniera operazionale il rapporto tra contesti e preferenze, e non a ritagliare i dati su misura. Per una rassegna più articolata su questi argomenti si veda [2].

1.3. Lo scenario applicativo

Tra i diversi scenari e casi di studio presi in considerazione per lo sviluppo del modello e della metodologia, ci si riferisce in questo lavoro al mondo delle agenzie immobiliari, ed in particolare ad una rete di franchising in cui sono presenti diverse agenzie sparse sul territorio, che fanno riferimento ad un'unica rete commerciale. Ciascuna agenzia dispone di un insieme di agenti, ognuno dei quali segue i clienti nel vendere o affittare il loro immobile (sia residenziale sia commerciale) o nel trovare l'immobile che fa per loro. Il sistema informativo deve quindi gestire tutte le informazioni relative a agenti, clienti ed immobili, oltre alle agende degli agenti per quanto concerne le visite che essi organizzano per far visionare gli immobili.
L'accesso a queste informazioni avviene dai terminali dell'ufficio oppure dai dispositivi palmari e smartphone, quando gli agenti effettuano le visite presso gli immobili volendo avere a disposizione non solo la propria agenda degli appuntamenti, ma anche tutte e sole le informazioni relative agli immobili oggetto della visita.
Come abbiamo già detto, la mobilità introduce un nuovo aspetto nella fruizione dei dati: se da una parte, il recente progresso tecnologico ha reso sempre più capaci i dispositivi mobili personali (palmari, smartphone ecc.), dotandoli di buone capacità di elaborazione e memorizzazione, e offrendo diverse alternative per quanto riguarda la connettività, dall'altra, i dispositivi sono ancora di dimensioni limitate (display piccolo e interazione a menù quantomeno lunga), e il costo dell'accesso ai dati è ancora piuttosto alto. Questi aspetti spingono alla realizzazione di applicazioni e servizi in grado di fornire esattamente le informazioni cui l'utente è interessato, per semplificarne l'accesso.
Tornando al nostro esempio, oltre che tramite dispositivo mobile, gli stessi dati relativi agli immobili sono anche visibili dal sito web della rete commerciale, per offrire un altro punto di accesso a potenziali *clienti* (che quindi possono semplicemente consultare i dati pubblici degli immobili) e ai clienti delle agenzie (che quindi avranno accesso ad un insieme più ricco di informazioni).
In questo scenario applicativo, dunque, le figure dei possibili utenti (o ruoli) individuate sono le seguenti: gli ***agenti*** delle agenzie immobiliari, ciascuno con il proprio portafoglio di immobili e di clienti, i ***clienti*** che hanno dato il mandato all'agenzia per vendere o affittare l'immobile di loro proprietà, insieme a quelli che desiderano acquistare o affittare un immobile, i ***visitatori*** del sito web, che si distinguono dai precedenti ***clienti*** in quanto caratterizzati da uno stato di anonimato nei confronti dell'agenzia, il ***manager*** dell'agenzia ed infine l'amministratore delegato (***CEO***) della rete commerciale, interessato principalmente alla produttività dei propri agenti ed al portafoglio degli immobili.
È possibile distinguere macro *aree di interesse* dei dati cui possono voler accedere i vari tipi di utenti: dati relativi agli ***immobili*** (con

Agenzia(***agenziaID***, indirizzo, centroOperativo, nome, orario)
Dipendente(***dipID***, nome, cognome, dataNascita, tel, email, indirizzo, ***managerID***, agenziaID)
Immobile(***immobileID***, clienteID, descrizione, metri2, mappaID)
Multimedia(***contenutoID***, tipo, ampiezza, altezza, dato)
ImmobileMultimedia(***immobileID***, ***contenutoID***)
Cliente(***clienteID***, nome, cognome, indirizzo, CAP, tel, email, categoria)
RichiestaCliente(***ricID***, ***acquirenteID***, budget, AcquistoAffitto)
Agenda(***data***, ***ora***, ***agenteID***, immobileID, clienteID)
Visita(***immobileID***, ***data***, ***agenteID***, ***clienteID***, durata)
Affitto(***immobileID***, ***clienteID***, dataIn, agenteID, rateo, condizioni, durata)
Vendita(***immobileID***, ***proprietarioID***, ***acquirenteID***, ***data***, agenteID, prezzoIniziale, prezzoFinale, condizioni, percAgenzia)

TABELLA 1
Schema logico della base di dati del caso considerato

la possibilità di raffinare la selezione in base, per esempio, alla categoria commerciale o residenziale, al fatto che siano in vendita oppure in affitto), dati relativi ai ***clienti*** (di rilievo per gli agenti e i manager delle agenzie), informazioni relative ai propri ***agenti*** o più in generale alle proprie ***agenzie***.
La selezione dei dati di interesse è ulteriormente influenzata – in determinati frangenti, quale per esempio la consultazione delle informazioni degli immobili – dalla posizione: tale localizzazione può essere sia determinata in modo automatico mediante strumenti in grado di rilevare le coordinate geografiche, sia esplicitamente dichiarata dall'utente. L'ambito applicativo in esame, infatti, mostra l'esigenza di poter disporre di dati filtrati rispetto alla propria posizione. Per esempio per un agente che si trova a dover reperire i dati degli immobili da lui gestiti entro un certo raggio rispetto alla posizione in cui si trova, durante i sopralluoghi. D'altra parte, è possibile voler selezionare i dati in base ad una posizione in cui ci si verrà a trovare in seguito per poter scaricare in anticipo le informazioni sul proprio dispositivo. Quindi la dimensione spaziale, come quella temporale, contribuisce alla nozione di contesto in due modi diversi: mediante una determinazione automatica di posizione (o istante di tempo), oppure mediante una richiesta esplicita da parte dell'utente. Infine, come nella maggior parte delle situazioni in cui la mobilità gioca un ruolo importante, il tipo di dispositivo (e quindi l'interfaccia, il canale ecc.) mediante il quale si accede alle informazioni ne influenza la ricchezza e il livello di dettaglio, prima ancora della modalità di presentazione che dovrà essere adottata. Dunque, anche nell'esempio proposto, l'interfaccia utilizzata determina un tipo di filtraggio dei dati. Parte dello schema logico della base di dati di questa applicazione è riportato nella tabella 1.
Nel seguito del lavoro verranno presentati il nostro modello di contesto e la sua utilizzazione per il ritaglio dei dati, insieme con alcuni esempi di applicazioni.

2. IL MODELLO DI CONTESTO

Il modello di contesto, chiamato *Context Dimension Tree* (CDT) [4] (riquadro), è stato definito con lo scopo di filtrare i dati relativi ad una particolare applicazione, in base alle reali necessità dell'utente. Esso è una rappresentazione ad albero dell'insieme delle possibili prospettive e dei relativi valori, uti-

La struttura di un Context Dimension Tree
In un Context Dimension Tree vi sono tre tipi di nodi (si veda la Figura 1):
- la *radice*: rappresenta l'insieme dei contesti descritti dall'intero Context Dimension Tree;
- i *nodi dimensione*: sono graficamente rappresentati come nodi neri e descrivono le dimensioni (per esempio, ***interest-topic***), o le sotto-dimensioni (per esempio, ***prezzo***, ***categoria*** ecc.), utili per il ritaglio dei dati;
- i *nodi concetto*: sono rappresentati come nodi bianchi e identificano i valori della dimensione alla quale sono collegati (per esempio, i valori della dimensione ***role*** sono i concetti *CEO*, *manager*, e *agente*).

Le dimensioni figlie del nodo radice rappresentano le cosiddette *dimensioni di contesto*, ossia le principali prospettive utili per la descrizione dei possibili contesti dell'utente nello scenario applicativo scelto. Ogni concetto (nodo bianco) dell'albero può essere ulteriormente analizzato con un livello di dettaglio crescente, aggiungendo al suo sottoalbero altre sotto-dimensioni di analisi che raffinano il concetto stesso.

li alla descrizione dei diversi aspetti del contesto nel quale un utente si colloca; un esempio di CDT, relativo al nostro esempio applicativo, è riportato nella parte superiore della figura 1.
Si noti che nella rappresentazione gerarchica dei contesti così definita, ogni generazione è formata da nodi dello stesso tipo (dimensioni o concetti) e ogni nodo specifica in modo più dettagliato gli aspetti descritti dai suoi antenati. Si noti che alcune sotto-dimensioni non hanno figli che sono nodi concetto (per esempio, *prezzo*). In tal caso, il numero di diversi valori per tali dimensioni sarebbe elevato: per non indicare esplicitamente l'insieme di tutti i possibili valori, si introduce un nodo parametro che verrà istanziato al momento dell'esecuzione con il reale valore assunto dalla dimensione nel contesto corrente dell'utente. Nel caso del nodo *prezzo*, quindi il parametro potrà assumere come valore uno dei prezzi degli immobili. I parametri possono essere aggiunti anche a nodi concetto (nodi bianchi) quando si desidera che l'istanza del relativo concetto venga filtrata al momento dell'esecuzione in base ad un opportuno valore: per esempio, il parametro ***$aid***, identificatore del ruolo agente.
Nel nostro modello, *gli elementi di contesto* rappresentano l'attualizzazione delle dimensioni di contesto e corrispondono a nodi di un Context Dimension Tree che possono essere: nodi concetto senza parametro (per esempio, *clienti*), nodi concetto con parametro (per esempio, agente($aid)) o nodi dimensione con parametro (per esempio, *n_locali*(***$lid***)).
Un contesto è formalizzato come una congiunzione di proposizioni sempre relative agli elementi di contesto che lo descrivono, ognuno dei quali è espresso mediante la proposizione ***dim_name = value*** dove ***dim_name*** rappresenta il nome di una (sotto-)dimensione e ***value*** il valore per quella dimensione. Tale valore quindi può essere un nodo concetto oppure un nodo concetto filtrato dal valore del suo parametro, o ancora il valore di un parametro di un nodo sotto-dimensione foglia. Un esempio di possibile contesto estratto dal Context Di-

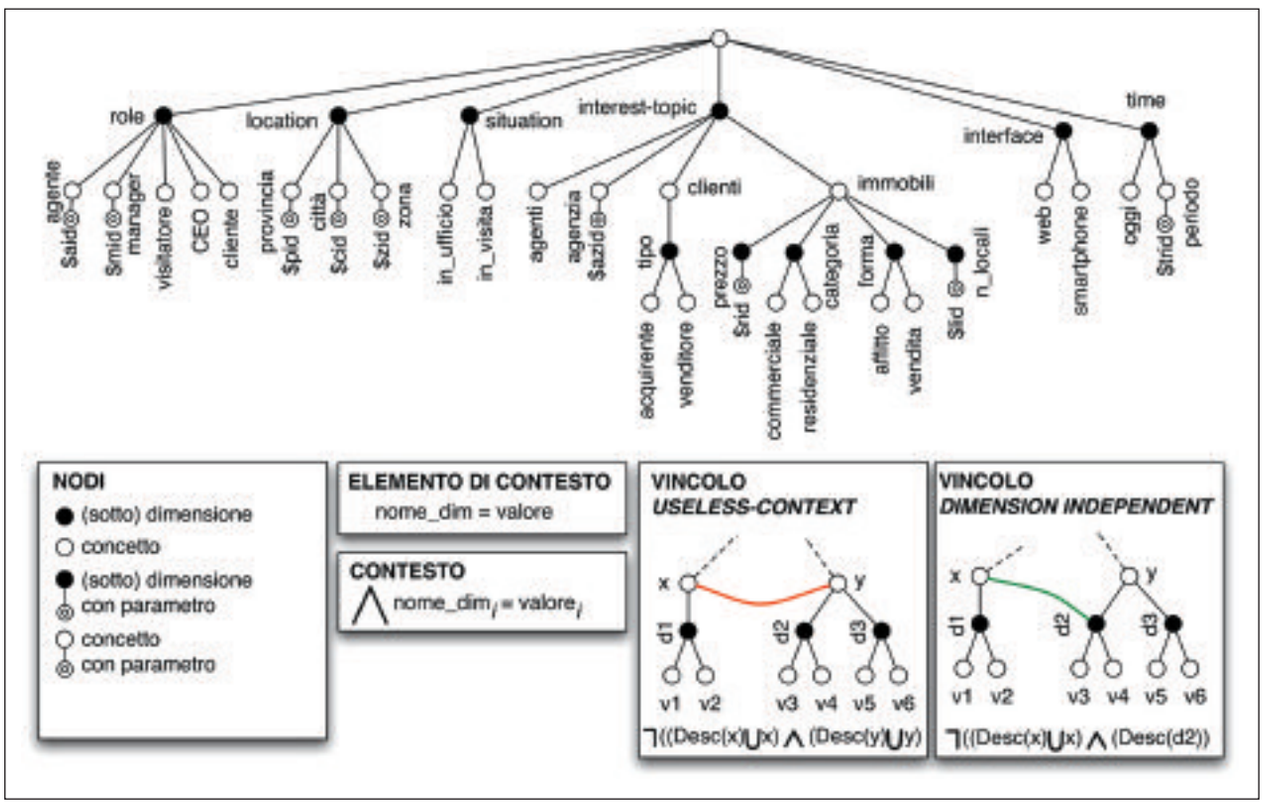


FIGURA 1
Il modello Context Dimension Tree

mension Tree di figura 1 è il seguente:

$$(\text{role} = \text{manager}(\$mid)) \wedge (\text{interest-topic} = \text{agenzia}(\$azid)) \wedge (\text{situation} = \text{in_ufficio}) \quad (1)$$

Tale contesto è relativo ad un manager – identificato al momento dell'esecuzione istanziando il suo identificatore ***$mid*** – che vuole consultare, durante il suo lavoro d'ufficio, le informazioni relative alle agenzie delle quali è responsabile.
La profondità del Context Dimension Tree per un'applicazione dipende dal livello di dettaglio utilizzato per definire i diversi aspetti che contribuiscono alla selezione delle informazioni disponibili; alcune dimensioni di contesto si prestano maggiormente ad essere rappresentate con un sotto-albero a vari livelli di dettaglio, come l'***interest-topic***, mentre altre hanno una rappresentazione più piatta (si veda per esempio, la dimensione ***situation***). È compito del progettista scegliere il livello di dettaglio adeguato nella specifica del Context Dimension Tree, e quindi dei contesti, in quanto più specifici sono i contesti più selettiva sarà l'interrogazione per selezionare le informazioni di interesse per ogni contesto. Alcune dimensioni di contesto da noi proposte sono utilizzabili nella maggior parte delle applicazioni, ma non tutte sono sempre necessarie e altre potrebbero essere aggiunte a quelle elencate in questo lavoro a seconda delle reali necessità. La tabella 2 riporta le

Dimensioni di contesto	Significato	Esempio Agenzia Immobiliare
Role	Gli utenti che utilizzano l'applicazione di interesse.	"*CEO*", "*manager*", "*agente*".
Interest Topic	Gli interessi dei potenziali utenti dell'applicazione.	"*agenzia*", ossia le informazioni su una particolare agenzia utili per il CEO; "*agenti*", informazioni sugli agenti utili per il CEO o per il manager di quell'agenzia; "clienti", informazioni su venditori e acquirenti che possono essere utilizzate dall'agente o dal manager; "*immobili*", tutte le informazioni riguardanti gli immobili da vendere o affittare. Quest'ultimo argomento può essere ulteriormente decomposto rispetto altri criteri: per esempio, *immobili commerciali/residenziali* o proprietà in *vendita/affitto*.
Situation	Fasi in cui l'applicazione verrà usata.	L'utente consulterà le informazioni di interesse durante il normale lavoro d'ufficio, "*in ufficio*", oppure nella situazione "*in visita*", quando, ad esempio, un agente è in visita ad immobili con potenziali clienti.
Time	Informazioni temporali. Il tempo può essere espresso in modo relativo o assoluto con livelli di granularità differenti.	Abbiamo scelto una visione relativa del tempo, permettendo due scelte di analisi possibili: il giorno corrente, o un intervallo di tempo variabile centrato sull'istante di tempocorrente.
Space	Informazioni spaziali, normalmente riferite alla posizione dove l'utente è situato. Lo spazio puòessere relativo o assoluto e anche la sua granularità a può variare.	"*qui*" o "*questa città*" sono valori spaziali relativi, mentre "*Roma*" è un valore assoluto.
Interface	Informazioni sul canale, o tipo di presentazione, per trasmettere le informazioni.	In alcuni casi le informazioni devono essere consultate da utenti umani, mentre in altre situazioni le informazioni vanno gestite da dispositivi elettronici che richiedono solo poche informazioni descritte in formati specifici.

TABELLA 2
Dimensioni del Context Dimension Tree: descrizioni e uso

principali dimensioni che abbiamo individuato, il loro significato intuitivo ed alcuni esempi relativi al nostro scenario applicativo.
Inoltre, gli aspetti di privatezza dei dati sono importanti in moltissime applicazioni: a tale scopo al Context Dimension Tree può essere aggiunta la dimensione *ownership*, che non è direttamente legata al ritaglio dei dati, ma può essere usata per definire i diritti di accesso, per esempio con istruzioni di *grant* e *revoke* definite sulle viste finali, quando le sorgenti informative sono relazionali. I diritti associati ai valori della dimensione *ownership* servono quindi per "oscurare" le parti sensibili delle viste finali, ottenute combinando le viste parziali associate agli altri elementi di contesto. Poiché la funzione di tale dimensione – pur potendo contribuire alla riduzione dei dati – è diversa dalle altre, non verrà approfondita in questo lavoro.
A partire dal Context Dimension Tree di un'applicazione, i contesti vengono generati prendendo un insieme (anche vuoto) di elementi per ogni dimensione, in modo tale che i vincoli del modello vengano rispettati. L'ipotesi che un elemento di contesto descriva concetti che comprendono quelli specificati nei suoi discendenti, e che i nodi concetto fratelli siano mutuamente esclusivi, porta a definire vincoli di parentela che impongono che, scelto un concetto, non vengano selezionati nel contesto altri nodi concetto presenti nel sottoalbero la cui radice è il genitore del nodo scelto, e gli ascendenti di tale nodo.
Inoltre, poiché non tutte le combinazioni di valori delle dimensioni sono significative per l'applicazione scelta, è opportuno eliminare i contesti non significativi con l'introduzione di ulteriori *vincoli*. I principali vincoli considerati nell'analisi (indicati mediante linee colorate nella Figura 1) sono:
- *useless contexts*: vengono introdotti per eliminare le combinazioni di valori di due o più dimensioni che non hanno senso per l'applicazione. Per esempio, un vincolo di questo tipo può impedire di generare i contesti nei quali il ruolo CEO compaia con la situazione "*in visita*";
- *dimension independent*: vengono introdotti quando i valori di una dimensione non influiscono sulla vista finale prodotta per i contesti che presentano un altro (o altri) valore per un'altra dimensione. Nel nostro scenario, la dimensione ***location*** non influisce sui dati associati ai contesti relativi al ruolo CEO, in quanto questo ruolo richiede dati generici contabili che non contengono informazioni spaziali.

Nella figura 1 viene indicata la formalizzazione generica dei due tipi di vincoli.

Viste su una base di dati

Una delle motivazioni principali per la nascita dei sistemi di gestione di basi di dati è stata l'esigenza di disaccoppiare il modello generale dei dati dell'azienda dalla visione che degli stessi dati hanno i singoli programmi applicativi. Il modello ANSI-SPARC definisce pertanto tre distinti livelli di rappresentazione: lo *schema logico* al livello delle strutture dei dati, lo *schema fisico*, che definisce le strutture dei dati e le relative operazioni in termini delle strutture e delle primitive del file system, e lo *schema esterno* o *vista* sullo schema logico, al livello dell'applicazione.
Nella sua pratica realizzazione, quale si ritrova nei moderni sistemi per la gestione di basi di dati, una vista è costituita da una particolare forma di interrogazione, che estrae dallo schema logico tutte e sole le informazioni che interessano un determinato programma (o insieme di programmi) applicativo, nella forma a questo più consona. Poiché nel corso della vita di una base di dati lo schema logico e i suoi contenuti sono soggetti a modifiche, si preferisce mantenere le viste in forma *virtuale*: ciò significa che l'interrogazione che realizza la vista viene eseguita ogni volta che viene richiesta da un'interrogazione dell'utente applicativo, consentendo così una grande flessibilità nell'evoluzione degli schemi al prezzo di un piccolo allungamento nel tempo di risposta. Tuttavia, qualora lo schema logico sia stabile e sia bassa la volatilità dei dati o quando considerazioni di efficienza o di affidabilità siano prevalenti, è possibile eseguire in anticipo una volta per tutte la vista e memorizzare permanentemente i risultati dell'interrogazione in tabelle, sulle quali poi agiscono le interrogazioni degli utenti applicativi; in tal caso si parla di *vista materializzata*.

3. LA METODOLOGIA PER IL RITAGLIO DEI DATI DIPENDENTE DAL CONTESTO

Dopo aver progettato il Context Dimension Tree per l'applicazione di interesse, e aver ricavato da tale rappresentazione grafica tutti i contesti significativi, il progettista dovrà associare ad ognuno di essi la porzione di dati rilevanti.
In tale fase, partendo da una rappresentazione integrata dei dati contenuti nelle sorgenti informative disponibili, per esempio uno schema relazionale globale o un'ontologia, il progettista dovrà selezionare sullo schema globale, per ogni contesto, lo schema dei dati ritenuti rilevanti, espresso mediante un'interrogazione (vista) sullo schema globale. A partire dalla vista relativa ad ogni contesto, i dati verranno successivamente estratti dalle sorgenti informative in modo automatico per il contesto corrente dell'utente (riquadro).
Poiché il numero di contesti per un'applicazione reale è solitamente elevato, nonostante il relati-

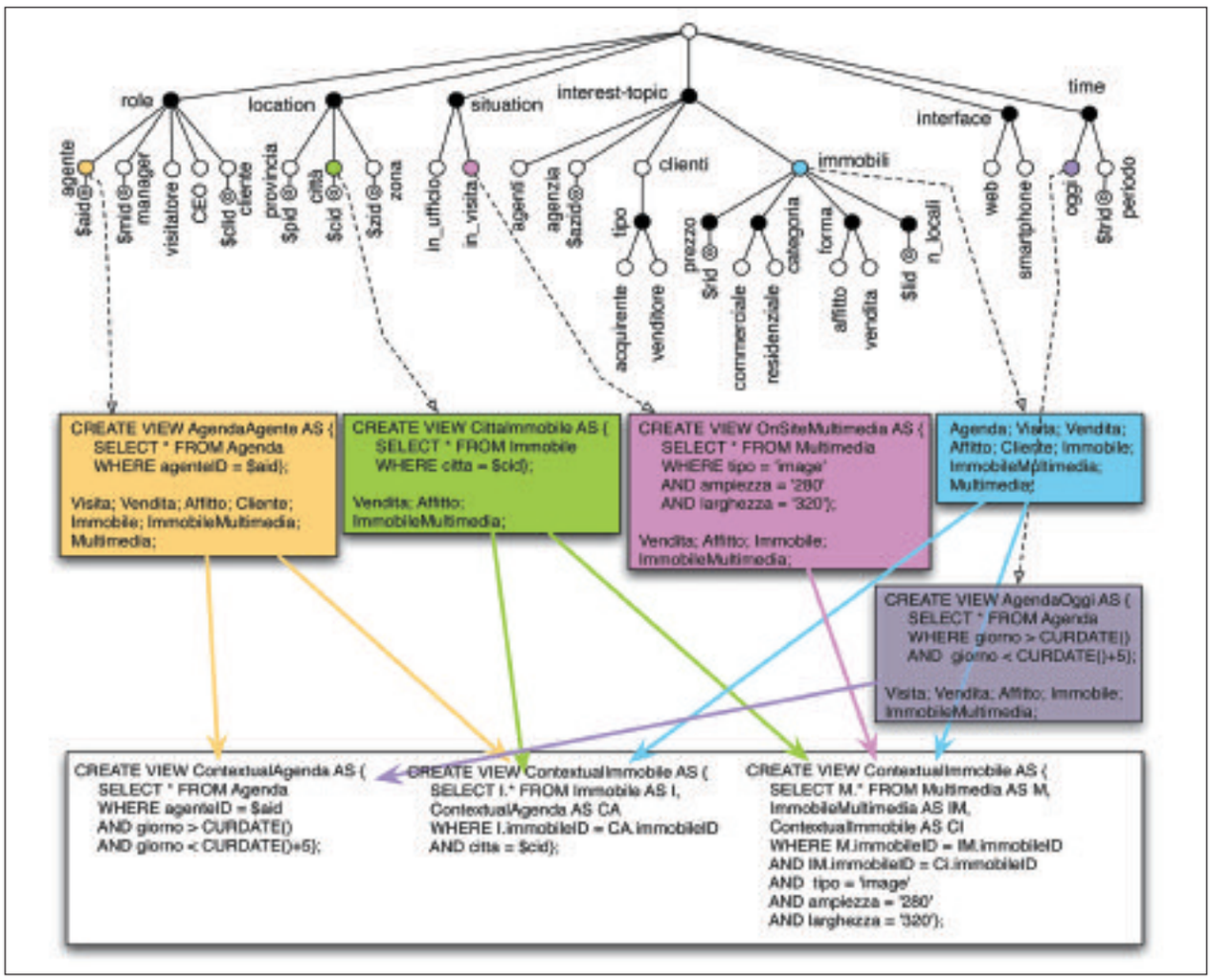


FIGURA 2

Viste associate ai singoli nodi dell'albero e derivazione per composizione della vista associata ad un contesto

vo Context Dimension Tree sia abbastanza limitato e alcune combinazioni non significative vengano eliminate con l'introduzione dei vincoli[1], l'associazione tra i contesti e le informazioni a disposizione richiede un intervento oneroso da parte del progettista. Per rendere semi-automatica tale procedura, la metodologia che viene descritta in questa sezione prevede di associare una *vista parziale* ad ogni elemento di contesto. Mediante operatori appositi, si può quindi, in modo automatico, ottenere un'approssimazione della vista relativa ad uno specifico contesto C, come combinazione delle viste parziali associate agli elementi di contesto presenti in C. Un ulteriore intervento del progettista potrà eventualmente perfezionare la vista così ottenuta. Si consideri, per esempio, il contesto relativo ad un agente interessato alle informazioni sugli immobili, mentre è in visita con possibili clienti in una specifica città:

$$(\text{role} = \text{agente}(\$aid)) \wedge (\text{interest-topic} = \text{immobili}) \wedge (\text{situation} = \text{in visita}) \wedge (\text{time} = \text{oggi}(\text{getdate}())) \wedge (\text{location} = \text{città}(\$cid)) \quad (2)$$

Il parametro ***$aid***, che identifica lo specifico agente, verrà istanziato al momento dell'esecuzione, come pure la città.
Nella figura 2 vengono mostrate, sotto il Context Dimension Tree dell'applicazione considerata, le viste SQL associate agli elementi

[1] Nel nostro esempio applicativo i contesti sono circa 1400; di questi, circa 400 vengono eliminati con l'introduzione di vincoli.

Contesto	Informazioni rilevanti
role=agente($aid) ^ interest_topic=immobili ^ situation=in_visita ^ time=oggi ^ location=città($cid)	Informazioni di uso quotidiano per un agente: informazioni sugli immobili da visitare (inclusi i dettagli degli immobili e le informazioni sulle aree circostanti), informazioni multimediali compatibili con il dispositivo mobile dell'agente, i suoi appuntamenti, le informazioni sui proprietari degli immobili e sui clienti.
role=CEO ^ interest_topic=vendita ^ time=periodo($trid)) ^ location=città($cid)	Dati aggregati relativi alle vendite: per esempio, importi medi e somme delle vendite, commissioni medie pagate sulle vendite, aggregate per intervalli di tempo e zona geografica.
role=manager($mid) ^ interest_topic=agenti ^ agende situation=in_ ufficio ^ time=periodo($trid)	Informazioni per assegnare gli immobili agli agenti: degli agenti, dati anagrafici degli agenti, immobili da assegnare e richieste dei clienti.
role=agente($aid) ^ interest_topic=clienti ^ situation=in_ ufficio ^ location=città($cid)	Informazioni per porre in contatto acquirenti e venditori: informazioni sui clienti, richieste non evase e dettagli sulle offerte.

TABELLA 3
Esempi di contesto e relative informazioni

presenti nel contesto (2); le frecce tratteggiate e colorate associano ogni elemento di contesto con la rispettiva vista parziale. L'agente sarà interessato alle informazioni relative a:

i. immobili presenti nell'area in cui lui effettua la visita con i clienti;

ii. informazioni multimediali relative agli immobili e compatibili con il dispositivo mobile in sua dotazione;

iii. gli appuntamenti fissati in un intervallo di 5 giorni, necessari per poter organizzare eventuali ulteriori impegni lavorativi.

Altri esempi di contesti per lo scenario applicativo scelto sono illustrati nella tabella 3, con una descrizione informale delle informazioni rilevanti per ogni contesto.

Le viste parziali vengono associate ad ogni elemento *n* del Context Dimension Tree, a partire dalla radice dell'albero, secondo una *politica* denominata *massimale*, che prevede di selezionare *tutte le informazioni ritenute di interesse per n*. Per esempio, i dati potenzialmente rilevanti per l'elemento ***role = agente($aid)***, come mostrato nella figura 2, saranno tutti gli appuntamenti dell'agente considerato, le informazioni sia multimediali che testuali relative agli immobili e alle relative caratteristiche, i dati anagrafici dei clienti dell'agenzia, i dettagli relativi alle visite e alle operazioni di acquisto o affitto di immobili. Inoltre, poiché il Context Dimension Tree nasce per fornire una rappresentazione gerarchica dei valori e delle dimensioni, procedendo dalla radice verso le foglie dell'albero i concetti vengono raffinati, e quindi il progettista dovrà selezionare le viste parziali in modo che la vista associata ad un nodo *n sia contenuta nella vista dei suoi antenati*.

Per esempio, la vista per l'elemento ***interest-topic = affitto*** deve essere contenuta in quella di ***interest-topic = immobili***; in particolare, dovrà contenere le sole informazioni e operazioni relative ad immobili in affitto.

Nella parte bassa della figura 2 vengono mostrate le viste SQL finali relative al contesto (2): le frecce colorate e i colori nelle viste stesse indicano i contributi delle viste parziali nella formazione di quelle finali[2].

L'operatore insiemistico usato per combinare le viste parziali associate agli elementi di contesto è la *doppia intersezione*, che estende l'operazione di ***INTERSECT*** di SQL a insiemi di relazioni. Intuitivamente, la doppia intersezione tra due insiemi di relazioni A e B, corrisponde, per ogni coppia di relazioni *RA* e *RB* (in A e B rispettivamente) che hanno in comune una parte di schema, all'applicazione dell'INTERSECT al massimo sottoschema comune. Le re-

2 In [3] gli operatori per combinare le viste parziali sono estensioni di operatori dell'algebra relazionale, banalmente traducibili nei corrispondenti operatori SQL. In questo lavoro mostreremo solo questi ultimi.

lazioni che non hanno sotto-schemi comuni non contribuiscono al risultato della doppia intersezione. Nell'esempio riportato nella figura 1 la relazione Agenda compare nelle viste parziali di tutti gli elementi presenti nel contesto (2). Nella vista SQL associata a tale contesto avremo pertanto la relazione Agenda, che verrà però filtrata per l'applicazione dell'intersezione dell'algebra relazionale, dalle condizioni imposte per gli elementi ***role = agente($aid)*** e ***time = "oggi"***. Nelle viste parziali di questi due elementi le informazioni sugli appuntamenti selezionano rispettivamente le tuple relative ai soli impegni dell'agente ***$aid***, in un intervallo di cinque giorni.
Si noti che la definizione della vista finale associata ad un contesto dipende:
i. dalla precisione e dal livello di dettaglio con cui sono state definite dal progettista le viste parziali e;
ii. dagli effetti dell'applicazione dell'operatore di doppia intersezione alle viste parziali.
Pertanto il progettista può ritenere opportuno modificare manualmente tale vista, ricavata automaticamente, aggiungendo o escludendo manualmente informazioni, ma tale intervento sarà sicuramente meno oneroso della definizione manuale di tutte le viste definitive associate ad ogni contesto significativo.

4. IL SISTEMA CONTEXT-ADDICT

Nel paragrafo precedenti abbiamo visto come la metodologia proposta permetta di realizzare sistemi per l'accesso "intelligente" ai dati, che sfruttino le informazioni proprie del dominio applicativo e della situazione corrente dell'utente, per ridurre opportunamente i dati prima di interrogarli o memorizzarli su un dispositivo client. Il sistema che abbiamo progettato si chiama Context-ADDICT, e la sua architettura è costituita da (Figura 3): le sorgenti di informazione, il *Context-ADDICT Server* (CAS), il server di Integrazione (IS) e i dispositivi client. Infatti, nel caso più generale si prevede che il sistema informativo a disposizione possa essere composto da diverse sorgenti informative, che vanno anche opportunamente integrate.

4.1. Le sorgenti informative

Le sorgenti informative comprendono tutti i sistemi che contengono le informazioni effettivamente disponibili nei diversi contesti in cui l'applicazione può operare. È possibile classificare queste sorgenti in base a criteri quali:
❑ *Il modello dei dati* - definisce se le informazioni sono rappresentate in modo strutturato (relazionale), semi-strutturato (XML o RDF) o non-strutturato (testo o HTML).
❑ *Stabilità* - rappresenta la volatilità della sorgente in termini di disponibilità durante le interrogazioni. La stabilità può essere relativa anche ai dati ovvero alla frequenza con cui i dati della sorgente vengono aggiornati.
❑ *Collaboratività* - è una misura di quanto una sorgente rende i propri dati accessibili a sistemi di terze parti. Ricordiamo infatti che la semantica dei dati è fornita dallo schema. Quanto più una sorgente rende disponibili i propri schemi in un formato direttamente elaborabile, tanto più sarà considerata collaborativa. Nel nostro contesto, la collaboratività è intesa come la disponibilità da parte della sorgente a fornire le meta-informazioni necessarie per rendere i dati accessibili in senso più generale.
❑ *Fiducia* (*trust*) - definisce quanto le risposte alle interrogazioni devono essere considerate affidabili. Tale misura comprende sia la possibilità che i dati siano errati o non aggiornati, sia il fatto che alcuni sistemi, gestiscano dati di sintesi o approssimati.
❑ *Tecnologia* - la tecnologia influenza le capacità di calcolo, i tempi di risposta, l'espressività delle interrogazioni e le modalità di co-

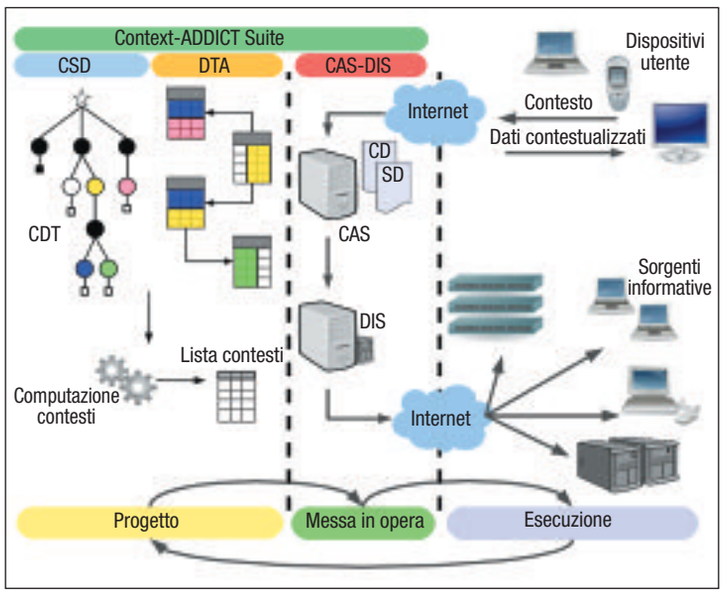


FIGURA 3
Architettura di un sistema Context-ADDICT

municazione. Tali caratteristiche influenzano la generazione del piano di esecuzione delle interrogazioni e devono dunque essere considerate.
Indipendentemente dalla collaboratività delle sorgenti, per poter generare un ritaglio coerente, occorre analizzarne gli schemi e produrre una rappresentazione degli stessi usando un modello dei dati comune, che permetta il confronto e la mappatura, con quelli delle altre sorgenti. Il modello scelto deve essere sufficientemente espressivo per rappresentare gli altri modelli dei dati e soprattutto i vincoli che devono sussistere fra le sorgenti ed il dominio modellato.

4.2. Il Context-ADDICT server

Il *Context-ADDICT Server* (CAS) rappresenta il nucleo dell'intera architettura e realizza i processi di *ritaglio dei dati* (context-aware data tailoring) e di esecuzione delle *interrogazioni contestualizzate* (context-aware query answering). Si tratta di un servizio che puó essere situato su un opportuno server applicativo come Apache Tomcat [20], JBoss [22] o IBM Websphere [21] e fornisce il supporto alla distribuzione delle interrogazioni mediante un dizionario dei contesti (*Context Dictionary* - CD). Ogni elemento del CD è una tripla (*c*, *s*, *q*) dove *c* è una specifica di contesto, *s* una sorgente informativa e *q* l'insieme di interrogazioni nel linguaggio nativo della sorgente che permette il ritaglio del contenuto (*formal query*). Il dizionario contiene anche le credenziali di accesso alla sorgente e una sua descrizione mediante meta-informazioni (*Source Descriptors* - SD). Le informazioni contenute nei SDs sono utilizzate in fase di formulazione del piano delle interrogazioni sia in un'ottica di ottimizzazione che di manipolazione delle stesse[3]. A fronte di una richiesta di informazioni di contesto da parte di un dispositivo client, il CAS:

1. individua l'elemento nel dizionario corrispondente alla specifica di contesto ricevuta;
2. per ogni sorgente coinvolta, recupera la definizione dell'interrogazione e ne istanzia le variabili con i valori contenuti nella richiesta (*actual query*);
3. recupera le credenziali di accesso alla sorgente e vi si connette;
4. nel caso la connessione sia avvenuta con successo, esegue l'interrogazione concreta e restituisce i valori al dispositivo interrogante.

Un esempio di questo processo è riportato nella figura 4.
Il CAS si avvale anche di un sistema di archiviazione di supporto (*support storage*) nel

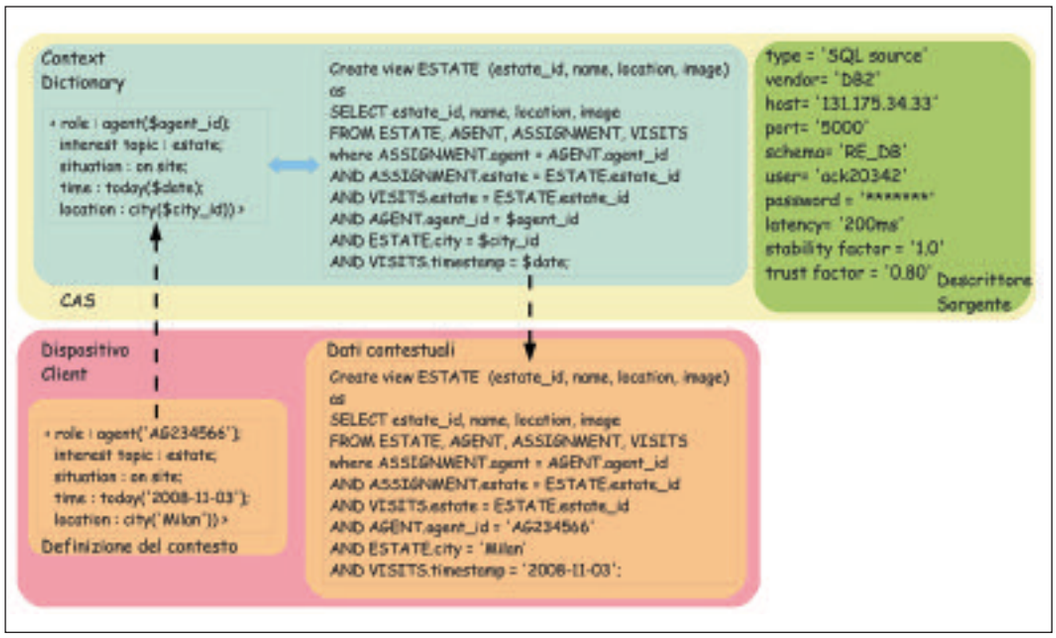


FIGURA 4
Esempio di interrogazione contestuale

[3] Per manipolazione si intende l'arricchimento di un'interrogazione con informazioni di contesto non espressamente indicate nell'interrogazione originale.

caso in cui sia più conveniente memorizzare localmente alcune informazioni piuttosto che effettuare una richiesta on-line ad una sorgente. Consideriamo per esempio il caso degli orari dei musei in un dominio applicativo di natura turistica; se la sorgente non è collaborativa, risulta necessario estrarre tali informazioni dalla pagina web del museo o dei punti di informazione turistici usando dei wrapper per pagine web. L'analisi delle pagine web mediante algoritmi di elaborazione del linguaggio naturale è un processo oneroso e non è conveniente, dal nostro punto di vista, effettuare tali operazioni ad ogni richiesta. Inoltre, informazioni cosiffatte sono in genere molto stabili (per esempio variano al cambiare della stagione o dell'anno) e possono essere soggette ad un'archiviazione locale (*local caching*) efficiente. All'atto dell'esecuzione di un'interrogazione che richieda queste informazioni, il CAS farà dunque riferimento alla memoria locale invece che alla sorgente originale. In quest'ottica, l'indice di stabilità della sorgente è uno strumento utile per definire con quale frequenza occorra aggiornare i dati memorizzati localmente.

4.3. Server di integrazione

Data la natura fortemente eterogenea delle sorgenti informative, non è sempre possibile determinare a priori se le informazioni contenute in esse appartengano ad insiemi disgiunti, ovvero vi è la possibilità che alcune sorgenti si riferiscano alla stessa porzione di dominio e quindi contengano informazioni comuni. La descrizione del processo di riconciliazione e integrazione di tali informazioni va oltre lo scopo di questo lavoro; il progetto Context-ADDICT prevede comunque un opportuno sistema di integrazione dei dati (*Data Integration System* - DIS) residente su un server di integrazione[4].

4.4. Dispositivi client

I client delle applicazioni dipendenti dal contesto che vengono realizzate con la metodologia Context-ADDICT possono risiedere su un qualsiasi dispositivo fisso o mobile. Ad ogni dispositivo saranno associate una o più specifiche di contesto, le cui variabili saranno opportunamente istanziate con i valori attuali al momento della richiesta al CAS. La gestione delle informazioni di contesto risulta quasi del tutto trasparente, fatta esclusione per le ovvie interazioni iniziali in fase di configurazione delle applicazioni client. Per i dispositivi mobili, le informazioni riguardanti la localizzazione spazio-temporale possono essere recuperate rispettivamente dal clock del sistema e da un dispositivo GPS[5]. Dimensioni come il ruolo e gli argomenti di interesse possono essere recuperate dal profilo utente o dall'analisi dei log delle interrogazioni effettuate, mentre i valori delle dimensioni più tecnologiche come l'interfaccia di comunicazione e il formato dei dati sono automaticamente determinate dalle tecnologie disponibili nel dispositivo (OBEX, WiFi, UMTS).

4.5. Il software di modellazione CADD

CADD (*Context-aware Data Designer* [6]) è il software di supporto alla modellazione di sistemi dipendenti dal contesto sviluppato al Politecnico di Milano sempre nell'ambito del progetto Context-ADDICT. CADD (Figura 5) supporta il progettista nella definizione del CDT (*Context Space Designer* - CSD), nell'associazione di ogni nodo del CDT alle porzioni di interesse degli schemi delle sorgenti informative (*Data Tailoring Assistant* - DTA) e nella fase di messa in opera del sistema (CAS Deployer - CASD).

4.5.1. Progetto del modello di contesto: CSD

Il progettista del CDT è guidato nella sua costruzione da un'interfaccia grafica intuitiva che ne permette la creazione in maniera visuale (Figura 6).
Come descritto in precedenza, è possibile esprimere dei vincoli fra i valori delle dimensioni del modello di contesto per eliminare a priori le combinazioni prive di significato (per esempio, un cliente che visita una sede aziendale, non sarà interessato alle informazioni sulla logistica in ingresso). I vincoli so-

[4] Un esempio di sistema commerciale per l'integrazione dei dati è IBM Websphere Information Integrator.

[5] Nokia prevede che entro il 2010 tutti i dispositivi cellulari e smartphones di fascia medio alta saranno dotati di dispositivo GPS integrato.

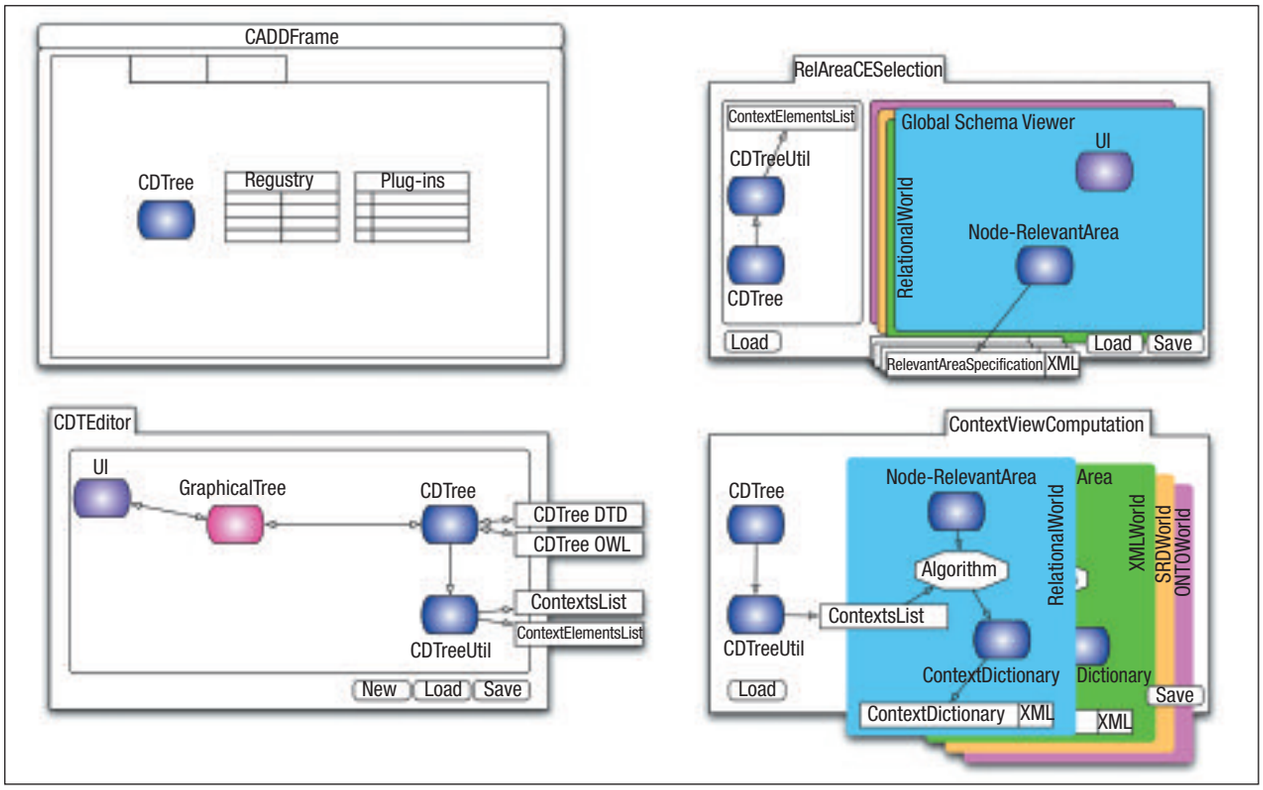


FIGURA 5

Architettura dello strumento CADD

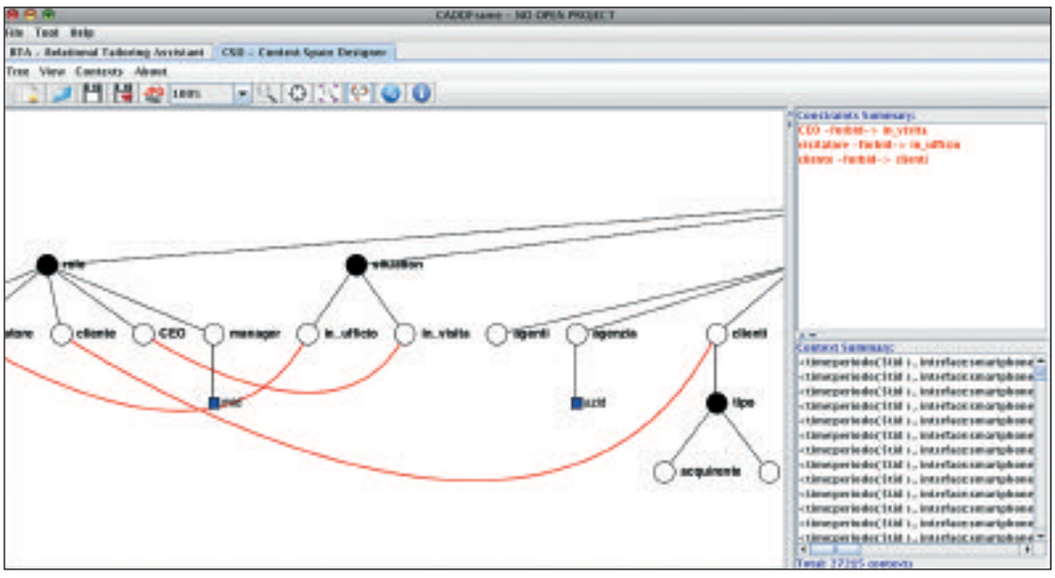


FIGURA 6

Il Context Space Designer - CSD

no espressi mediante formule logiche che predicano sulle dimensioni e sui loro valori e che possono essere create graficamente. Una volta completato il modello di contesto, CADD permette la generazione automatica di tutti i contesti validi (ovvero i contesti che rispettano i vincoli di progetto), ai quali possono essere associate le aree rilevanti degli schemi delle sorgenti.

4.5.2. Definizione delle aree rilevanti

La definizione delle aree rilevanti (*Relevant Areas*) degli schemi delle sorgenti avviene per mezzo del DTA. Il progettista associa gra-

ficamente ad ogni nodo del modello di contesto la porzione rilevante dello schema delle sorgenti; questa associazione si traduce poi in una vista sui dati della sorgente codificata nel linguaggio nativo della stessa (SQL, XQuery o SPARQL).
Come descritto precedentemente, una specifica di contesto è il risultato della combinazione di più dimensioni e quindi di più nodi del modello di contesto. Dato un contesto (cioè un insieme di nodi del CDT) CADD combina automaticamente le viste sulla sorgente dati in modo da ottenere una *vista contestuale*, determinata cioè da una specifica di contesto mediante combinazione delle viste associate ad ogni singolo nodo. Le associazioni (*contesto-vista*) tra contesti validi e interrogazione di ritaglio sono poi memorizzate nel dizionario dei contesti (Contexts Dictionary).

4.5.3. Messa in opera del sistema

CADD supporta anche la messa in opera (deployment) automatica del CAS su un server applicativo come Apache Tomcat. In una fase successiva all'installazione, è possibile modificare le viste contestuali con il DTA e aggiornare il dizionario dei contesti, in modo trasparente all'utente applicativo. È opportuno notare come l'architettura Context-ADDICT si appoggi su tecnologie esistenti senza richiedere modifiche ai sistemi di memorizzazione pre-esistenti come i database relazionali o XML. L'esecuzione delle interrogazioni contestuali atte a ritagliare le informazioni sulle sorgenti prescinde dalla natura delle stesse che rimangono meri esecutori delle interrogazioni. La complessità della federazione del sistema rimane all'interno dello strumento di modellazione ed è semplificata dall'interfaccia grafica del CADD.

5. CONCLUSIONI E SVILUPPI FUTURI

La nozione di contesto sta assumendo un ruolo sempre più rilevante nello sviluppo dei moderni sistemi informativi. Il sistema presentato in questo lavoro richiede la progettazione preventiva del modello di contesto, seguendo una precisa metodologia, mentre il valore delle variabili relative al contesto stesso viene stabilito dinamicamente in fase di esecuzione. Tale modo di procedere è reso ancora più efficace dalla disponibilità e diffusione di sensori di vario tipo e natura (dall'RFID al Telepass, al GPS ed altri ancora), che consentono di rilevare i valori direttamente dal mondo fisico circostante. Si rendono così possibili funzionalità di apprendimento, di analisi storiche e di ragionamento sui contesti [1], che costituiscono un importante passo verso la possibilità di modificare a tempo di esecuzione il modello stesso del contesto, definito per una certa applicazione. Questa é la sfida con la quale ci si dovrà confrontare nei prossimi anni.

Bibliografia

[1] AA.VV. IEEE Electronic Proc. of 5th Workshop on Context Modeling and Reasoning (CoMoRea'08), p. 406-452, Hong Kong, 2008.

[2] Bolchini C., Curino C.A., Quintarelli E., Schreiber F.A., Tanca L.: A data-oriented survey of context models. *ACM SIGMOD Record*, Vol. 36, n. 4, 2007, p. 19-26.

[3] Bolchini C., Curino C.A., Schreiber F.A., Tanca L.: *Context integration for mobile data tailoring. Electronic Proc.* of 7th Int. Conf. on Mobile Data Management (MDM'06), Nara (Japan), 2006, p. 1-8.

[4] Bolchini C., Quintarelli E., Rossato R.: *Relational data tailoring through view composition*. In: Proc. Intl. Conf. on Conceptual Modeling (ER'2007), Springer LNCS, Vol. 4801, 2007, p. 149-164.

[5] Bolchini C., Schreiber F.A., Tanca L.: A methodology for a Very Small Data Base design. *Information Systems*, Vol. 32, n. 1, 2007, p. 61-82.

[6] Bolchini C., Curino C., Orsi G., Quintarelli E., Schreiber F.A., Tanca L.: *CADD: a tool for context modeling and data tailoring*. Proc. IEEE Intl. Conf. on Mobile Data Management (MDM'07), Mannheim (Germany), 2007, p. 221-223.

[7] Chalmers M.: A historical view of context. *Computer Supported Cooperative Work*, Vol. 13, n. 3, 2004, p. 223-247.

[8] Coutaz J., Crowley J.L., Dobson S., Garlan D.: Context is key. *Communications of the ACM*, Vol. 48, n. 3, 2005, p. 49-53.

[9] Dey A.K.: Understanding and using context. *Personal Ubiquitous Computing*, Vol. 5, n. 1, 2001, p. 4-7.

[10] Lenzerini M.: Data Integration: *A Theoretical Perspective*. Proc. Of ACM PODS, 2002, p. 233-246.

[11] Motschnig-Pitrik R.: A Generic Framework for the Modeling of Contexts and its Applications. *Data & Knowledge Engineering*, 2000, p.145-180.

[12] Mylopoulos J., Motschnig-Pitrik R.: *Partitioning Information Bases with Contexts*. Proc. of the 3rd Intl Conf. on Cooperative Information Systems, 1995, p. 44-54.

[13] Ouksel A.M.: In-context peer-to-peer information filtering on the web. *SIGMOD Record*, Vol. 32, n. 3, 2003, p. 65-70.

[14] Petrelli D., Not E., Strapparava C., Stock O., Zancanaro M.: *Modeling context is like taking pictures*. In: Proc. of the Workshop "The What, Who, Where, When, Why and How of Context-Awareness". In: Proc. of CHI2000 - The future is here, 2000.

[15] Roussos Y., Stavrakas Y., Pavlaki V.: *Towards a Context-Aware Relational Model*. In: Contextual Representation and Reasoning Workshop (CRR'05), held in conjunction with CONTEXT'05, Paris, 2005.

[16] Sridharan H., Mani A., Sundaram H., Brungart J., Birchfield D.: *Context-Aware Dynamic Presentation Synthesis for Exploratory Multimodal Environments*. Proc. of IEEE Intl Conf. on Multimedia and Expo (ICME'05), 2005, p. 1014-1017.

[17] Stefanidis K., Pitoura E., Vassiliadis P.: *Modeling and storing context-aware preferences*. In: Proc. of 10th East European. Conf. on Advances in Databases and Information Systems (ADBIS'06), Springer LNCS, Vol. 4152, 2006, p. 124-140.

[18] Strimpakou M., Roussaki I., Anagnostou M.E.: *A context ontology for pervasive service provision*. In: 20th Intl. Conf. on Advanced Information Networking and Applications (AINA'06), 2006, p. 775-779.

[19] Theodorakis M., Analyti A., Constantopoulos P., Spyratos N.: A theory of contexts in information bases. *Information Systems*, Vol. 27, n. 3, 2002, p. 151-191.

[20] The Apache Software Foundation: Apache tomcat. http://tomcat.apache.org/.

[21] The International Business Machines Corporation: Ibm websphere application server, http://www-306.ibm.com/software/webservers/appserv/was/.

[22] The JBoss Community. Jboss application server, http://labs.jboss.com/jbossas/.

Cristiana Bolchini è Professore Associato presso il Politecnico di Milano. Laureata in Ingegneria Elettronica, indirizzo Automatica, nel Febbraio nel 1993, ha conseguito il Dottorato in Ingegneria Informatica e Automatica (IX ciclo) nello stesso ateneo. Gli interessi di ricerca riguardano sia problematiche di progetto di dispositivi digitali, con particolare attenzione agli aspetti di affidabilità, sia metodologie di progetto, ritaglio e gestione di dati.
E-mail: bolchini@elet.polimi.it

Giorgio Orsi è Dottorando di Ricerca presso il Dipartimento di Elettronica e Informazione del Politecnico di Milano. Nell'ottobre del 2006 ha ottenuto la laurea specialistica in Ingegneria Informatica presso il Politecnico di Milano. Il tema di ricerca maggiore riguarda lo studio di sistemi di integrazione di dati eterogenei e dinamici basati su ontologie, con particolare enfasi sulla gestione e l'evoluzione delle dipendenze dei dati.
E-mail: orsi@elet.polimi.it

Elisa Quintarelli è Ricercatrice presso il Dipartimento di Elettronica e Informazione del Politecnico di Milano, dove insegna Informatica A e Progetto di Basi di Dati. Gli interessi di ricerca attuali riguardano l'utilizzo di linguaggi di interrogazione e tecniche di datamining per documenti XML, la definizione di metodologie di progettazione di basi di dati per dispositivi portabili e di piccole dimensioni. Su questi e precedenti argomenti di ricerca è co-autrice di articoli, pubblicati su atti di conferenze e riviste internazionali. È inoltre autrice del libro "*Model-Checking Based Data Retrieval: an application to semistructured and temporal data*" pubblicato da Springer Verlag. Partecipa ai comitati di programma di workshop internazionali ed è revisore per riviste internazionali.
E-mail: quintarelli@elet.polimi.it

Fabio A. Schreiber è Professore Ordinario di Basi di Dati e di Technologies for Information Systems e vice Preside della Facoltà di Ingegneria dell'Informazione presso il Politecnico di Milano. I suoi interessi di ricerca riguardano i Sistemi di Elaborazione Distribuiti e i Sistemi Informativi; su questi argomenti ha partecipato a vari progetti di ricerca nazionali e internazionali. Ha fatto parte del Comitato Metropolitano per la Pubblica Amministrazione presso la Prefettura di Milano come consulente per i Sistemi Informativi nella P.A. Attualmente si occupa di problematiche relative ai Sistemi Informativi pervasivi, dipendenti dal contesto e comprendenti dispositivi molto piccoli (PDA, telefoni cellulari, sensori, Rfid ecc.). È autore di numerosi articoli scientifici ed è associate editor di Data & Knowledge Engineering.
E-mail: schreibe@elet.polimi.it

Letizia Tanca è Professore Ordinario di Basi di Dati e di Technologies for Information Systems presso il Politecnico di Milano. È autrice di diverse pubblicazioni internazionali sulle basi di dati e sulla teoria delle basi di dati, e del libro "*Logic Programming and Databases*", scritto con S. Ceri e G. Gottlob. Ha partecipato a vari progetti di ricerca nazionali e internazionali. I suoi interessi di ricerca riguardano tutta la teoria delle basi di dati, in particolare le basi di dati deduttive, attive e orientate agli oggetti, i linguaggi a grafi per basi di dati, la rappresentazione e l'interrogazione di informazione semistrutturata. Gli interessi più recenti vertono sul progetto di Basi di Dati sensibili al contesto, con particolare riferimento alle basi di dati per piccoli dispositivi mobili, e sulle basi formali e gli strumenti per il web semantico.
E-mail: tanca@elet.polimi.it

ICT E DIRITTO

Rubrica a cura di

Antonio Piva, David D'Agostini

Scopo di questa rubrica è di illustrare al lettore, in brevi articoli, le tematiche giuridiche più significative del settore ICT: dalla tutela del *domain name* al *copyright* nella rete, dalle licenze software alla *privacy* nell'era digitale. Ogni numero tratterà un argomento, inquadrandolo nel contesto normativo e focalizzandone gli aspetti di informatica giuridica.

La sicurezza delle informazioni e le Norme ISO 27000

Attilio Rampazzo, Antonio Piva, David D'Agostini

1. Introduzione

Il 5 aprile scorso è entrata in vigore la Legge 18 marzo 2008 n. 48, con la quale l'Italia ratifica la Convenzione del Consiglio d'Europa sulla criminalità informatica firmata a Budapest il 23 novembre 2001; tale accordo internazionale riguardante i crimini commessi attraverso internet o altre reti informatiche, rappresentò il primo tentativo di realizzare una politica comune fra gli Stati membri, attraverso l'adozione di una legislazione appropriata, che consenta di combattere il crimine informatico in maniera coordinata.
La legge citata introduce importanti modifiche al Codice penale, al Codice di procedura penale, al d.lgs. 231/2001 sulla responsabilità amministrativa delle persone giuridiche e al d.lgs. 196/2003 sul trattamento dei dati personali. Si aggiunge così un nuovo tassello alla legislazione applicabile in materia di sicurezza delle informazioni già corposa ed articolata.
Il quadro normativo è in costante e rapida evoluzione, in funzione ovviamente della crescente importanza che il bene "informazione" sta assumendo non solo in ambito economico, ma anche in ambito sociale e politico. Di conseguenza per le aziende, le pubbliche amministrazioni e gli enti "*no profit*" diventa sempre più complesso gestire in modo efficiente, efficace e soprattutto economico la crescente mole di informazioni in appropriata sicurezza e rispettando la normativa applicabile. Termini come Information *Security Compliance* ma meglio *Information Security* Governance sono all'ordine del giorno.
L'International *Organization for Standardization* (ISO), sempre attenta alle evoluzioni, da alcuni anni ha costituito un gruppo che sta emanando una serie di norme nell'ambito della Sicurezza delle Informazioni che tendono ad aiutare le organizzazioni a ben cimentarsi nell'*Information Security Governance* e a migliorare la fiducia nelle relazioni tra azienda e azienda (B2B - *Business to Business*) o tra azienda e cliente (B2C - *Business to Consumer*). Pertanto da alcuni anni sono aumentate in maniera significativa le disposizioni, seppur sempre volontarie, che prospettano una nuova visione di Gestione dell'*Information Tecnology* come elemento dominante del *Business* dell'organizzazione. La capostipite di queste norme è la ISO/IEC 27001:2005 - Tecnologia delle informazioni - Tecniche di sicurezza - Sistemi di gestione della sicurezza delle informazioni - Requisiti.

2. Origine ed evoluzione della Norma ISO/IEC 27001

La realizzazione della norma risale agli inizi degli anni '90 quando il *Department of Trade and Industry* (DTI) britannico istituì un gruppo di lavoro finalizzato a fornire alle aziende una guida per la gestione della sicurezza del loro patrimonio informativo. Il gruppo pubblicò nel 1993 una rac-

colta di best practice (intitolata *Code of practice for information security management*) che costituì la base per lo standard vero e proprio pubblicato dal *British Standard Institution* (BSI) nel 1995. Questo standard, essendo opera del BSI, è stato identificato con la sigla BS 7799.
Nel 1998 fu aggiunta una seconda parte allo standard (intitolata *Specification for information security mangement systems*) che fu poi sottoposta a una revisione complessiva conclusasi con la pubblicazione, nell'aprile del 1999, di una nuova versione delle sue due parti. Nel 1999 lo standard BS7799 viene sottoposto all'approvazione dell'ISO/IEC per farlo diventare standard internazionale.
La parte prima dello standard BS7799 diventa standard internazionale ISO/IEC 17799 alla fine del 2000. La parte seconda viene sottoposta a nuova revisione nel 2002 al fine di armonizzarla rispetto agli altri Sistemi di Gestione certificati (ISO 9001, ISO 14001).Nel frattempo, il 15 luglio 2005, la ISO/IEC 17799 viene aggiornata ed implementata mentre anche la seconda parte del BS7799 viene presa in considerazione dall'ISO. Il 15 ottobre 2005 vede l'approvazione la ISO/IEC 27001 da alcune modifiche della BS 7799-2:2002 ed In luglio 2007 la ISO 17799:2005 è stata rinumerata ISO 27002.
Si dà vita così a una nuova famiglia, la ISO 27000 series che tratterà tutti gli standard sulla Gestione dei Sistemi di Sicurezza delle Informazioni.
Lo standard ISO/IEC 27001 è entrato in vigore e pubblicato in Italia in data 28/03/2006 come UNI CEI ISO/IEC 27001:2006 Tecnologia delle informazioni - Tecniche di sicurezza - Sistemi di gestione della sicurezza delle informazioni – Requisiti. Nell'autunno 2007 l'UNI pubblica la versione in lingua italiana della norma curata da UNINFO, ente di formazione federato all'UNI che si interessa di tecnologie informatiche e loro applicazioni.

3. ISO/IEC 27001:2005 e ISO/IEC 27002:2005

La differenza tra le due norme è data dai differenti obiettivi.
La Norma ISO/IEC 27002 ha l'obiettivo di fornire delle regole di buon comportamento sulla gestione della sicurezza nell'organizzazione: usa sempre il verbo *should* (dovresti) e mai *shall* (devi) nel condurre un'azione per raggiungere la conformità al requisito normativo. La ISO/IEC 27001, invece, ha l'obiettivo di specificare i requisiti per la realizzazione di un Sistema di Gestione della Sicurezza delle Informazioni: usa sempre il verbo *shall* (devi). La ISO/IEC 27001:2005 è la norma utilizzata per certificare il Sistema della Sicurezza delle Informazioni (SGSI)[1] degli Enti e delle aziende.

4. La Norma ISO/IEC 27001:2005

La Norma copre tutte le tipologie di Organizzazioni (imprese commerciali, agenzie governative, organizzazioni senza scopo di lucro). Lo sviluppo di un Sistema di Gestione per la Sicurezza delle Informazioni secondo la ISO/IEC 27001 consente tra le altre cose di gestire al meglio i requisiti della normativa applicabile allo specifico caso aziendale, contribuendo a salvaguardare l'organizzazione stessa da eventuali conseguenze negative di carattere giuridico, economico e di immagine. La ISO/IEC 27001 non pretende di essere il rimedio a tutti mali della Sicurezza delle Informazioni, ma costituisce il punto di partenza per impostare un efficace Sistema Organizzativo che comprenda tutti gli aspetti della Sicurezza e che si inserisca in un contesto di IT Governance[2].
Quando si parla di sicurezza delle informazioni si pensa subito alla sicurezza informatica: antivirus, firewall, backup, password ecc.. Si corre quindi il rischio di considerare la sicurezza delle informazioni, e le norme di riferimento quali la Norma ISO 27001, come un'attività tecnica, o come una norma "tecnica", di esclusiva competenza degli "informatici".Contrariamente, invece, la Norma ISO/IEC 27001 propone un modello organizzativo, più che uno standard tecnico: non "come" fare, ma "cosa" fare.
Per evitare pericolosi fraintendimenti, le caratteristiche in sintesi sono le seguenti:
- la ISO/IEC 27001 è uno standard gestionale (e non tecnico) nel campo del "management" e non dell'"ingegneria". I sistemi di gestione della sicurezza delle informazioni considerano infatti, oltre agli aspetti informatici, anche (e soprattutto) gli aspetti inerenti la gestione delle risorse

[1] SGSI corrispondente all'inglese ISMS – Information Security Management System.

[2] Si pensi per esempio a ITIL, ISO 20000; Cobit, CoSO - Enterprise Risk Management.

umane, la gestione dei processi organizzativi e la gestione degli spazi fisici. Paradossalmente un'organizzazione potrebbe essere dotata di sistemi informatici di ultima generazione, ma non aver sensibilizzato il proprio personale a non lasciare scritta sul post-it in bella vista la password di accesso ai sistemi oppure a non discutere per cellulare di argomenti delicati.

❑ La ISO/IEC 27001 riguarda la gestione in sicurezza di tutto il patrimonio informativo aziendale, indipendentemente dai supporti utilizzati per immagazzinare i documenti e i dati, supporti che possono essere elettronici ma anche cartacei.

❑ La ISO/IEC 27001 definisce la sicurezza delle informazioni in termini di preservazione di integrità, disponibilità e riservatezza delle stesse. Ciò premesso, la ISO/IEC 27001 ha una struttura molto agile e gli aspetti "burocratici" di sistema sono ridotti al minimo. La norma eredita i nove principi dell'OCSE (emessi in occasione della sessione del Consiglio dell'OCSE del 25 luglio 2002) nei quali si afferma che:

- i Governi hanno il compito di diffondere la cultura della sicurezza;
- tutti (anche gli utenti occasionali) devono contribuire all'attuazione ed alla gestione della sicurezza;
- rispetto dei principi etici e democratici, libera circolazione delle informazioni, protezione adeguata dei dati personali, apertura e trasparenza.

La ISO/IEC 27001 è strutturata macroscopicamente in tre blocchi di requisiti:

❑ fino al capitolo 3 - requisiti introduttivi e di spiegazione della norma (come quasi tutte le Norme ISO);

❑ dal capitolo 4 al capitolo 8 - requisiti applicativi del SGSI;

❑ appendici (o allegati) normativi e descrittivi a supporto di quanto citato nei capitoli precedenti.

I capitoli compongono la parte alta che contiene i requisiti che armonizzano la norma con gli altri schemi di certificazione (ISO 9001, ISO 14001) utilizzando il ciclo *Plan-Do-Check-Act* (o di *Deming*). Gli appendici o allegati sono tre:

❑ l'Allegato A ricopre un ruolo fondamentale nelle fasi di implementazione operativa e *audit*;

❑ l'Allegato B (informativo) contiene integrazione dei principi OECD rispetto al modello PDCA;

❑ l'Allegato C indica la corrispondenza tra ISO 9001:2000, ISO 14001:2004 e questo standard.

Essenzialmente la Norma ISO 27001 prende in considerazione, oltre agli aspetti di sistema, i seguenti aspetti inerenti la Gestione della Sicurezza delle Informazioni:

❑ identificazione del perimetro aziendale che il S.G.S.I[3]. deve proteggere. Il perimetro può essere di natura fisica (per esempio, sito o siti aziendali), di natura organizzativa (per esempio, funzioni aziendali) o di natura logica (reti informatiche e sistemi informativi);

❑ definizione di Politica e Obiettivi per la sicurezza delle informazioni, ovviamente in funzione della missione strategica aziendale e del relativo assetto organizzativo;

❑ identificazione dei beni (per esempio, relativi a sistemi informativi, sistemi informatici, sistemi di comunicazione, archivi cartacei ecc.) entro i confini del perimetro da proteggere nonché valutazione del valore dei beni stessi per l'Azienda;

❑ identificazione delle vulnerabilità inerenti i beni (per esempio, carenza di controlli, insufficiente formazione del personale, accessi fisici incustoditi, luoghi soggetti a incendi e inondazioni ecc.);

❑ identificazione delle minacce che possono sfruttare le vulnerabilità (per esempio, furti, avarie, incendi, inondazioni, picchi o interruzioni di energia elettrica, errori umani, atti dolosi ecc.);

❑ stima delle probabilità del verificarsi delle minacce individuate (eventi);

❑ stima della gravità dell'impatto delle minacce sui beni in termini di integrità, disponibilità e riservatezza delle informazioni;

❑ calcolo del rischio in funzione delle probabilità e degli impatti;

❑ scelta di come affrontare il rischio calcolato. Si può scegliere di accettarlo consapevolmente, di evitarlo, di eluderlo mediante trasferimento a soggetti terzi, per esempio, con contratti di assicurazione, oppure di ridurlo mediante applicazione di opportuni controlli;

❑ scelta ed attuazione dei controlli per diminuire il rischio. I controlli possono essere di natura organizzativa, fisica o logico-informatica.

L'allegato A della ISO/IEC 27001 si sviluppa in 11 sezioni (clausole, domini o aree di intervento), identificando per ciascuna gli obiettivi del controllo (in totale 39, ove il termine "controllo" va inteso in senso lato come "strumento di gestione") e i controlli stessi da implementare (per un totale di 133).

[3] S.G.S.I. Sistema di Gestione della Sicurezza delle Informazione traduzione del termine inglese I.S.M.S. *Information Security Management System*.

Le 11 aree di intervento:

1. Politica di sicurezza – l'obiettivo è fornire una guida e supporto alla direzione per la sicurezza delle informazioni.

2. Sicurezza Organizzativa – standard, norme e metodologie; definizione dei compiti e relative responsabilità anche quando l'elaborazione delle informazioni è affidata in outsourcing.

3. Gestione degli *asset* – l'obiettivo è garantire anche la gestione delle informazioni abbia un appropriato livello di protezione attraverso la classificazione delle informazioni stesse per una corretta valutazione del rischio associato.

4. Sicurezza delle risorse umane – l'obiettivo è ridurre i rischi di errore, furto, frode, o abuso da parte degli operatori accertandosi che gli stessi siano informati delle possibili minacce riguardanti la sicurezza delle informazioni, sostenendo così la politica sulla sicurezza della società nello svolgimento del loro lavoro.

5. Sicurezza fisica e ambientale – si vuole impedire l'accesso, il danneggiamento e l'interferenza dei non autorizzati, interni o esterni, proteggendo il flusso delle informazioni del business aziendale impendendone manomissioni o furto delle informazioni.

6. Gestione delle operazioni e delle comunicazioni – accertarsi del corretto funzionamento dei sistemi di elaborazione delle informazioni, minimizzare i rischi di guasti, mantenere l'integrità e la validità dei processi di elaborazione dell'informazione e comunicazione, garantire la salvaguardia delle informazioni in rete e la protezione delle infrastrutture a supporto.

7. Controllo degli accessi – l'obiettivo è impedire l'accesso non autorizzato e gestire un appropriato il sistema di autenticazione.

8. Acquisizione, sviluppo e mantenimento dei sistemi informativi – le attività di sviluppo e manutenzione devono essere gestite, controllate e tracciate.

9. Gestione degli incidenti di sicurezza delle informazioni – garantire che le debolezze ed i relativi eventi legati alla sicurezza siano segnalati per permettere efficaci azioni correttive.

10. Gestione della continuità operativa – obiettivo è impedire interruzioni alle attività ed ai processi critici del business aziendale dovuti a causa di eventi accidentali o fraudolenti.

11. Rispetto normativo – obiettivo è di garantire la conformità ai requisiti di legge (rispetto delle leggi civili, penali, patti contrattuali) e dei requisiti di sicurezza e delle policy: va cioè garantito il rispetto di tutto quanto può essere definito "cogente".

I controlli proposti sono una precisa descrizione a livello funzionale delle misure di protezione richieste per gestire la Sicurezza delle Informazioni. Ciononostante è importante sottolineare che l'elenco dei 133 controlli presente nell'allegato A è per esplicita ammissione non esaustivo: possono essere introdotti un numero maggiore di controlli a discrezione dell'organizzazione o possono essere eliminati controlli che non sono necessari[4].

I controlli devono essere realizzati:

- ❑ attraverso soluzioni hardware o software (di autenticazione, protezione crittografica, chiave digitale, monitoraggio software delle modifiche sul sistema ecc.) nel caso dei controlli attuabili mediante misure di sicurezza di tipo logico;
- ❑ attraverso l'installazione di sistemi anti intrusione, antifurti volumetrici e/o perimetrali, telecamere, casseforti, allarmi a distanza, armadi ignifughi, porte blindate, accesso fisico con badge magnetico, nel caso dei controlli che richiedono misure di sicurezza di tipo fisico;
- ❑ attraverso la definizione di precise procedure e la creazione di apposite strutture o cariche aziendali, per la messa in atto dei controlli di tipo procedurale (per esempio, l'istituzione del Comitato della Sicurezza per la gestione della sicurezza dell'informazione, la gestione delle risorse, la definizione dei requisiti di sicurezza da inserire nei contratti con le terze parti, le procedure di controllo degli accessi fisici, le registrazioni di manutenzione sui server, costanti audit interni sulla sicurezza ecc.); in questo caso si tratta di misure organizzative.

5. La famiglia ISO 27000 e normativa vigente

La ISO/IEC 27001 è uno standard **certificabile e auditabile** (**attraverso i requisiti**), mentre tutte le altre norme che costituiranno la famiglia ISO 27000 sono dei consigli, delle guide di supporto alla ISO/IEC 27001 quindi non sono oggetto

[4] La scelta dei controlli per ciascuna sezione di appartenenza deve essere sempre individuata attraverso un processo formale di valutazione dei rischi. L'output del processo di valutazione dei rischi è costituito dall'insieme dei controlli prescelti ed è formalizzato nella Dichiarazione di Applicabilità (S.O.A. - *Statement of Applicability*).

di certificazione ma vengono utilizzate per supportare l'attuazione di un Sistema di Gestione della Sicurezza delle Informazioni.
Attualmente la famiglia ISO 27000 è costituita da:
- ISO/IEC 27001:2005 - Sistema di Gestione per la Sicurezza delle Informazioni – Requisiti;
- ISO/IEC 27002:2005 - S.G.S.I. Consigli pratici;
- ISO/IEC 27005:2008 - S.G.S.I. Gestione del Rischi;
- ISO/IEC 27006:2007 - S.G.S.I. Linee guida per Enti di Certificazione Accreditati.

Nel corso dei prossimi anni sono previste le pubblicazioni ulteriori norme generali (tabella 1)[5] ed inoltre i vari gruppi del comitato ISO/IEC JTC 1/SC27 stanno lavorando a specifiche norme che trattano di applicazioni settoriali (sanità, finanza e banche, *automotive*, manifatturiero ecc.) o aspetti tecnici quali *incident management, network security*, IDS (tabella 2).
Quando si parla di Sicurezza delle Informazioni si pensa subito alla normativa sulla "privacy", ovvero al *Dlgs. n. 196 del 30 giugno 2003* "*Codice in materia di protezione dei dati personali*"[6]. Si corre quindi il rischio di considerare la norma ISO/IEC 27001 come un inutile "doppione " della legge sulla "privacy" e del relativo Documento Programmatico sulla Sicurezza (quest'ultimo può essere effettivamente integrato nel S.G.S.I. ma non viceversa!).
In realtà gli standard volontari di gestione (come la ISO/IEC 27001, la ISO 9001, la ISO 14001) hanno, tra le altre finalità, quella di supportare l'Azienda a gestire in modo sistematico, efficiente ed efficace i requisiti posti da specifiche norme cogenti di legge. Questo anche in considerazione del fatto che la legislazione inerente la sicurezza delle informazioni è piuttosto corposa e va ben oltre il codice sulla "privacy": si pensi per esempio al d.lgs. 8 giugno 2001 n. 231 "Responsabilità Amministrativa delle Persone Giuridiche", alla normativa in materia di tutela del diritto di autore (copyright)[7], di commercio elettronico e tutela dei consumatori, di accesso agli atti amministrativi, di reati informatici[8], di digitalizzazione della PA[9] ecc..
La normativa cogente di legge in materia di sicurezza dei dati è in genere finalizzata a tutelare gli interessi dei terzi mentre lo standard ISO 27001 è esplicitamente finalizzato a tutelare innanzitutto gli interessi dell'Organizzazione (nonché dei suoi legali rappresentanti e/o responsabili di gestione). Paradossalmente, un'Organizzazione potrebbe, per esempio, essere "tendenzialmente" in linea con la "privacy" e cessare di esistere perché i suoi dati più vitali (brevetti, specifiche di produzione ecc.) sono stati trafugati! Per non parlare dei dati di natura economico-finanziaria

• ISO/IEC 27000 S.G.S.I. - Fondamenti e vocabolario (FCD)
• ISO/IEC 27003 S.G.S.I. - Guida all'implementazione
• ISO/IEC 27004 S.G.S.I. - Misurazioni e metriche (FCD)
• ISO/IEC 27007 Linee Guida per l'Audit di un S.G.S.I. (WD)
• ISO/IEC 27008 Linee Guida per l'Audit dei Controlli

TABELLA 1
Norme della famiglia 27000 di prossima pubblicazione

• ISO/IEC 27011 Information security management guidelines for telecommunications (FDIS)
• ISO/IEC 27031 ICT readiness for business continuity
• ISO/IEC 27032 Guidelines for cybersecurity
• ISO/IEC 27033 Guidelines network security (revisione della ISO/IEC 18028-1:2006))
• ISO/IEC 27034 Guidelines for application security
• ISO/IEC 27799 Security Management in Health using ISO/IEC 27002 (pubblicata 12/6/2008)
• ISO/IEC 24762 Disaster recovery services (pubblicata 31/1/2008)

TABELLA 2
Specifiche nome settoriali o tecniche, previste o pubblicate

[5] Pur non appartenenti alla famiglia delle ISO 27000, per quanto riguarda la sicurezza informatica si ritiene doveroso menzionare anche le Norme ISO 20000 - Sistemi di Gestione dei Servizi IT - e le BS 25999 - Sistemi di Gestione di Business Continuity.

[6] Per una completa trattazione si rinvia al numero 2 del giugno 2008, al numero 2 di giugno 2007, al numero 1 di marzo 2004, all'interno della presente rubrica ICT e Diritto di Mondo Digitale.

[7] Si rinvia al numero 3 del settembre 2004, all'interno della presente rubrica ICT e Diritto di Mondo Digitale.

[8] Si rinvia al numero 2 del giugno 2004 ed al numero 4 di dicembre 2006, all'interno della presente rubrica ICT e Diritto di Mondo Digitale.

[9] Si rinvia al numero 3 del settembre 2005 ed al numero 4 di dicembre 2005, all'interno della presente rubrica ICT e Diritto di Mondo Digitale.

che, essendo relativi all'Organizzazione, non sono il fulcro centrale della tutela derivante dalla normativa sulla "privacy" (la quale si concentra soprattutto sui dati personali relativi a terzi trattati dall'Organizzazione).
Pertanto gli adempimenti cogenti inerenti la sicurezza dei dati e dei sistemi informatici (art. 31 del *Codice della privacy*) sono finalizzati a ridurre al minimo i rischi di distruzione o perdita dei medesimi, ovvero di accesso non autorizzato, di trattamento non consentito o non conforme alla finalità della raccolta, l'adozione di idonee misure di sicurezza in relazione al progresso tecnologico ed alle specifiche caratteristiche del trattamento di dati trattati, le prescrizioni sulla sicurezza dei dati e dei sistemi (*art. 34 del Codice*), la tenuta di un aggiornato documento programmatico sulla sicurezza e l'attuazione delle misure minime di sicurezza (*Allegato B del Codice della Privacy denominato "Disciplinare tecnico in materia di misure minime di sicurezza"*) devono essere effettuate al fine di prevenire le sanzioni amministrative, civili e penali derivanti dal Codice della Privacy.
L'identificazione e l'aggiornamento costante dei processi riguardanti il controllo della sicurezza fisica, logica ed organizzativa, l'analisi dei rischi per l'identificazione delle misure idonee di sicurezza, la gestione di opportune procedure ed istruzioni operative frequentemente aggiornate, il monitoraggio dei processi aziendali anche attraverso costanti verifiche interne tecnico-organizzative e *audit*, come previsto dalle norme volontarie internazionali ISO/IEC 27001[10], riguardanti i Sistemi di Gestione della Sicurezza delle Informazioni, sono utili strumenti per supportare l'azienda oltre che nella gestione la Sicurezza informatica e Privacy[11], anche e soprattutto per tutelare e garantire il business dell'organizzazione.

Organismo di Certificazione	Certificati ISO/IEC 27001:2005
TUV	85
DNV	61
RINA	13
IMQ	35
CERTIQUALITY	14
CERMET	4
Totali	212

TABELLA 3

Diffusione in Italia: situazione certificati ISO/IEC 27001 rilasciati da Organismi Italiani accreditati SINCERT
(Fonte SINCERT data : 30/6/2008)

6. La diffusione della certificazione ISO/IEC 27001

A livello internazionale le certificazioni ISO/IEC 27001 aumentano di oltre mille ogni anno con un significativo incremento a conferma del continuo interesse a questo *framework* come supporto organizzativo per la gestione della Sicurezza delle Informazioni: a fine luglio 2008 sono stati emessi 4426 certificati ISO/IEC 27001[12].
L'Italia si trova alla dodicesima posizione per certificati rilasciati e riconosciuti dall'International *Register of ISMS Certificates* ed anche se lentamente, la consapevolezza di dotarsi volontariamente, di un metodo di trattamento sicuro delle informazioni sta avanzando.
Dal sito del Sincert si evince che a fine giugno 2008 in Italia sono 212 i certificati ISO/IEC 27001:2005 rilasciati a circa un'ottantina di aziende, con un incremento di raddoppio dal 31 marzo 2006 termine del periodo di transizione della precedente norma BS 7799-2:2002 (Tabella 3). Un risultato significativo che con tutta certezza nei prossimi anni dovrebbe sicuramente ancora aumentare.

7. Conclusioni

"*La sicurezza è un processo, non un prodotto*". Questa è l'affermazione di Bruce Schneier, noto autore di libri sulla crittografia e sulla sicurezza informatica, che riassume la filosofia

[10] Ivi inclusi le citate aree di intervento, obiettivi di controlli ed i controlli stessi previsti dalla norma.

[11] Ci si potrebbe porre il dubbio su come operativamente sarebbe possibile essere a posto con le misure minime ed idonee di sicurezza imposte dal Codice della Privacy senza un'appropriata e costante gestione di tutti i processi relativi alla sicurezza come suggerito dalla norma ISO/IEC 27001.

[12] Fonte www.iso27001certificates.com version 182 July 2008; nella tabella **A** a p. 65 si riportano, in percentuale, le certificazioni rilasciate suddivise per settore.

che sta alla base dello standard ISO/IEC 27001. Non è sufficiente progettare un modello di sicurezza focalizzato sui soli aspetti tecnologici, ma è necessario tenere in considerazione anche le carenze legate all'organizzazione e alle procedure interne. Lo standard ISO/IEC 27001 attribuisce anche a questi ultimi aspetti una notevole importanza. Un Sistema di Gestione di Sicurezza delle Informazioni conforme allo standard ISO/IEC 27001 è uno strumento attraverso il quale un'organizzazione può dimostrare di essere capace di tutelare in modo globale il proprio patrimonio informativo (o quello di terzi a lei affidato). Se ne deduce che ormai dotarsi di un Sistema di Gestione della Sicurezza delle Informazioni è un'esigenza e non una "*tendenza del momento*" come ancora alcuni pensano.

Settore	Percentuale
Telecomunicazioni	25
Finanza	21
Servizi Terze Parti	17
Industria	15
Manifatture	11
Organizzazioni Governative	7
Utilities	4

TABELLA A

Certificazioni rilasciate suddivise per settore (in percentuale)

Bibliografia

[1] Quaderno AICQ n. 22 - *Gestione della Sicurezza delle Informazioni: guida alla lettura della Norma ISO 27001*.

[2] Quaderno Clusit - *Implementazione e certificazione dei sistemi di gestione per la sicurezza delle informazioni*.

[3] UNI CEI ISO/IEC 27001:2006 *Tecnologia delle informazioni -Tecniche di sicurezza - Sistemi di gestione della sicurezza delle informazioni – Requisiti*.

[4] ISO/IEC 27002:2005 *Information technology -- Security techniques -- Code of Practice for Information Security Management*.

[5] Piva A., D'Agostini D.: Sicurezza informatica e privacy nella scuola. *Mondo Digitale*, n. 2 - giugno 2008, p. 55-60. Felician L., D'Agostini D., Piva A.: 10 anni di Privacy: la protezione dei dati personali tra passato e futuro. *Mondo Digitale*, n. 2 giugno 2007, p. 65-74. Piva A., D'Agostini D.: La tutela dei dati personali nell'era digitale: il Codice sulla privacy tra vecchi e nuovi adempimenti. All'interno della rubrica ICT e Diritto di *Mondo Digitale*, n. 1 marzo 2004, p. 57-60.

ATTILIO RAMPAZZO, consulente di Sistemi Informativi e Sicurezza delle Informazioni in AlmavivA Finance Spa. Ha maturato un'esperienza più che trentennale nello sviluppo e conduzione di progetti informatici in ambito bancario e finanziario, nei quali la qualità e la sicurezza hanno ricoperto un ruolo determinante. ÈVice Presidente del Comitato AICQ "Qualità del Software e dei Servizi IT", valutatore Sistemi di Sicurezza delle Informazioni R.G.V.I. (AICQ_SICEV certificato n.3), socio AIPSI-Associazione Italiana Professionisti Sicurezza Informatica.
E-mail: attilio@rampazzo.it

ANTONIO PIVA, laureato in Scienze dell'Informazione, *Vice Presidente dell'ALSI* (Associazione Nazionale Laureati in Scienze dell'Informazione ed 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 ISO9000 ed ispettore AICA.
E-mail: antonio@piva.mobi

DAVID D'AGOSTINI avvocato, master in informatica giuridica e diritto delle nuove tecnologie, collabora all'attività di ricerca scientifica dell'Università degli studi di Udine e ha fondato l'associazione "*Centro Innovazione & Diritto*". È componente della Commissione Informatica dei Consigli dell'Ordine del Triveneto, responsabile dell'area "*Diritto & informatica*" della rivista "*Il foro friulano*", membro dell'organo di Audit Interno di Autovie Venete SpA.
E-mail: studio@avvocatodagostini.it