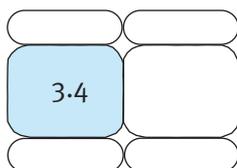




L'ERA DEL MOBILE COMPUTING

Antonella Carbonaro
Luigi Colazzo
Andrea Molinari



Partendo da un'analisi del settore *mobile computing*, viene presentata una schematizzazione dei suoi elementi qualificanti, cioè le tecnologie abilitanti, le reti di telecomunicazione, gli apparati mobili e le applicazioni. Il contributo analizza poi con maggior dettaglio gli aspetti innovativi che riguardano gli utilizzatori di apparati mobili, fornendo alcuni esempi di ambiti applicativi che mostrano come il mondo del *mobile computing* stia cambiando il rapporto tra utenti finali e tecnologie della computazione.

1. LO SCENARIO

Con l'espressione *mobile computing* si intende qualsiasi tipo di elaborazione effettuata da un dispositivo elettronico in movimento nello spazio fisico. Si noti che nell'idea di eseguire processi di elaborazione su apparati mobili sono implicite due aspetti. Il primo riguarda la mobilità di coloro che usano il dispositivo. Il secondo riguarda il dispositivo stesso che potrebbe agire in modo autonomo e indipendentemente dalla mobilità del suo proprietario, esattamente come avviene se dimentichiamo su un treno il telefono cellulare. I semplici processi di elaborazione che consentono al cellulare di comunicare con la rete durante il viaggio in treno avvengono anche quando siamo seduti davanti all'incaricato dell'ufficio oggetti smarriti. Più seriamente, robot che possano muoversi autonomamente nello spazio fisico rappresentano esempi di elaborazione in movimento. Recentemente, il *mobile computing* si spinge oltre, fino a far parlare alcuni studiosi di "*pervasive computing*" o "*ubiquitous computing*", un passo ulteriore verso la totale libertà di cal-

colo rispetto allo spazio fisico [1]. La rapida evoluzione della tecnologia hardware ha inglobato dispositivi computerizzati miniaturizzati nei prodotti più disparati, avvicinando sempre più la visione di *disappearing technology* proposta da Weiser [2].

Storicamente, l'idea del *mobile computing* comincia a prendere forma agli inizi degli anni '80, con l'affermazione sul mercato dei personal computer. Rispetto alle macchine di quel periodo, il PC era enormemente meno ingombrante ed il suo scarso ingombro ne permetteva il trasporto in modo relativamente semplice, anche se le intenzioni dei costruttori non erano certo quelle di farne un apparato mobile, quanto uno strumento appunto personale. Questa caratteristica di potenziale portabilità di elaborazione venne comunque quasi subito utilizzata per la produzione di computer cosiddetti trasportabili, ovvero computer costruttivamente progettati (scheda madre, disco fisso, lettori, schede di espansione ecc.) per il trasporto. Nel 1981 la Osborne Computer mise in commercio Osborne 1, un computer trasportabile, con

un monitor da 5 pollici, 64 kBytes di memoria, due unità a dischetti da 5.25 pollici. Successivamente, si è assistito ad un fiorire di apparati sempre più piccoli e sempre più leggeri, adatti all'utilizzo in movimento, commercialmente definiti con appellativi che ricordano comunque l'aspetto mobile (*laptop*, *notebook*, *subnotebook*, *tabletPC* ecc.).

Il concetto di *mobile computing* comprende ormai fasce di prodotti molto ampie: una macchina fotografica digitale, un lettore MP3, una unità diagnostica portatile, un navigatore satellitare, un telefono cellulare, una apparecchiatura cardiologica portatile ecc., sono tutti esempi di dispositivi che incorporano o sono di fatto dei computer, disponendo tutti di un processore, di memoria, di dispositivi di input/output. Inoltre, la mobilità deriva dalla disponibilità (al contrario dei tradizionali computer da tavolo) di una fonte autonoma di energia che garantisce autonomia rispetto alla rete elettrica. Questa diffusione degli apparati di elaborazione in prodotti di consumo è solo agli inizi: potenzialmente ogni apparato di una certa complessità potrebbe essere equipaggiato di hardware e di software adeguati al compito e alle circostanze. Se un tale apparato fosse per una qualche ragione mobile, esso rientrerebbe a pieno titolo nell'universo del *mobile computing*.

Un altro elemento che caratterizza il *mobile computing* è la connettività. La capacità di compiere elaborazioni mentre l'apparato è in movimento rende possibile la comunicazione con altri dispositivi, sia fissi che in movimento nello spazio. L'avvento di Internet, lo sviluppo delle telecomunicazioni, la convergenza in atto tra multimedialità, reti e calcolo, e (recentemente) l'avvento della tecnologia *wireless* (senza fili), tutti questi fattori hanno determinato una capacità di connessione dei dispositivi mobili senza il legame fisico rappresentato dalle tecnologie guidate, ovvero i cavi. Ormai, *mobile computing* assume anche il significato di *always connected*, sempre connesso. La connettività senza fili è ovviamente un patrimonio dell'uomo moderno dai tempi della TV e della radio, dove la semplice antenna bastava a ricevere il segnale. La telefonia ha poi raggiunto il terreno della mobilità, grazie alla diffusione delle reti cellulari. In questi ultimi tempi i dispositivi di

calcolo hanno cominciato a poter essere utilizzati in condizioni "nomadiche" (una persona, un dispositivo, tanti luoghi), superando così il vincolo del passato costituito dalla necessità di un cavo di comunicazione per connettersi alla rete. Fatta eccezione per tecniche sofisticate e non certo di massa, l'elaborazione dei computer è sempre stata condizionata, per la sua trasmissione, dalla disponibilità di mezzi guidati (cavi coassiali, dop-pini, fibre ottiche). Ora la libertà dal *wired* (cablato), ovvero dai cavi di collegamento, crea una attenzione verso il settore *wireless* forse eccessiva, con adozioni di queste tecnologie spesso scorrette e soprattutto poco sicure, come dimostrano i frequenti casi di installazioni *wireless* non criptate che danno spazio a fenomeni come il *wardriving*, ovvero quelle attività che consistono nell'intercettare reti *wireless*, in automobile o a piedi con un computer portatile, per sfruttare il collegamento o per catturare dati riservati [3].

Già oggi è ben chiaro che anche nel settore *wireless*, sia pur con uno/due ordini di grandezza in meno rispetto al settore della connettività fissa, si assiste ad una impetuosa crescita dell'offerta di larghezza di banda. La rivoluzione del *wireless* prepara il possibile lancio di standard come *MobileFi* (*Mobile Fidelity*), secondo lo standard IEEE 802.20, oppure di *WiMax*, basato sullo standard 802.16 approvato nel giugno del 2004.

MobileFi è un sistema radiomobile basato interamente su protocollo IP (*Internet Protocol*), offre una velocità di trasmissione di almeno 1 Mbit al secondo per utente a piena mobilità (fino a 250 km/h). *MobileFi* si espande verso le *Wide Area Networks* (WAN) ed è proposto per soddisfare i bisogni del mercato "automotive" (rete tra i navigatori delle automobili per il controllo del traffico) in concorrenza con i sistemi cellulari a larga banda, *UMTS* e loro estensioni, per aumentare la capacità il sistema *Hsdpa* (*High Speed Downlink Packet Access*) incrementa la capacità del canale downlink UMTS da 2 a 10 Mbit/s e oltre, fino a 30 Mbit/s. *MobileFi* completa verso l'alta mobilità e le lunghe distanze la famiglia di tecnologie secondo lo standard IEEE 802, che comprende anche le PAN (*Personal Area Networks*), e cioè le reti a breve distanza (qualche metro) che con-

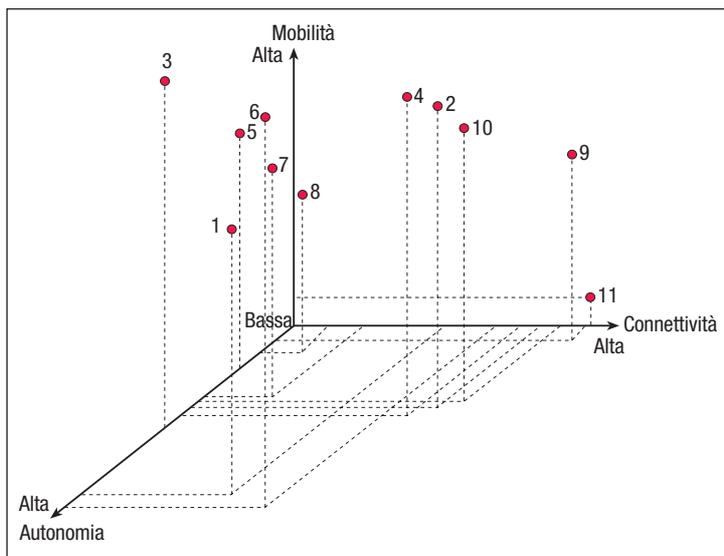


FIGURA 1

Lo spazio tecnologico del mobile computing

	Mobilità	Connettività	Autonomia
1 Robot	Media	Media	Alta
2 Palmari	Alta	Alta	Media
3 Lettore MP3	Alta	Nulla	Media
4 Telefono Cellulare	Alta	Alta	Media
5 Macchina fotografica digitale	Alta	Nulla	Bassa
6 GPS (Global Positioning System)	Alta	Alta	Alta
7 Apparecchiature Medicali	Media	Bassa	Media
8 Strumentazione per la diagnostica	Media	Bassa	Bassa
9 Notebook	Alta	Alta	Bassa
10 TabletPC	Alta	Alta	Media
11 PC	Bassa	Alta	Nulla

TABELLA 1
Dispositivi mobili principali rispetto alle tre variabili Mobilità, Autonomia, Connettività

sentono l'interazione tra gli oggetti intelligenti che circondano le persone (standard *BlueTooth*, il rivoluzionario *UltraWideBand* (UWB) e l'innovativo *ZigBee*). *WiMAX* (*Worldwide Interoperability for Microwave Access*) rappresenta un complemento dell'attuale tecnologia *WiFi*, e può essere collegato alle altre reti locali (LAN) e alle reti WAN al fine di fornire incanalamenti comuni. *WiMAX* viene considerata come una tecnologia di rete di area metropolitana (MAN, *Me-*

tropolitan Area Network) *wireless*, finalizzata alla connessione ad Internet dei punti di accesso ad alta velocità, ponendosi sostanzialmente come concorrente delle connessioni DSL che oggi coprono il cosiddetto "ultimo miglio". Infatti, lo standard IEEE 802.16 raggiunge una estensione di area di servizio lineare fino a 50 km, anche in presenza di mancanza di contatto visivo. Questo costituirà elemento fondamentale per raggiungere aree non altrimenti cablabili con il collegamento a banda larga o alternativo al rame e alla fibra ottica. *WiMAX* supporta velocità di trasmissione molto elevate rispetto agli standard attuali, raggiungendo i 70 Mbits al secondo, secondo la proposta dello standard.

Volendo schematizzare questo universo di dispositivi piuttosto articolato, potremmo individuare tre dimensioni degli apparati mobili:

- la prima dimensione riguarda la *mobilità* dell'apparato di elaborazione, ovvero la possibilità dell'apparato di effettuare elaborazioni mentre è in movimento nello spazio fisico;
- la seconda riguarda l'*autonomia*, ovvero la possibilità che l'apparato possa essere autonomo rispetto alla rete elettrica;
- la terza dimensione riguarda la *connettività*, ovvero la capacità dell'apparato mobile di connettersi con altri apparati sia fissi che in movimento.

Queste tre dimensioni possono essere rappresentate in un diagramma in cui disporre le tecnologie e gli apparati su tre assi cartesiani: mobilità, autonomia e connettività. Nella figura 1 sono rappresentati alcuni fra i dispositivi mobili maggiormente conosciuti, rispetto alle tre variabili indicate. Ogni variabile assume valori compresi fra "assente" e "pieno", secondo valutazioni indicative rappresentate nella tabella 1.

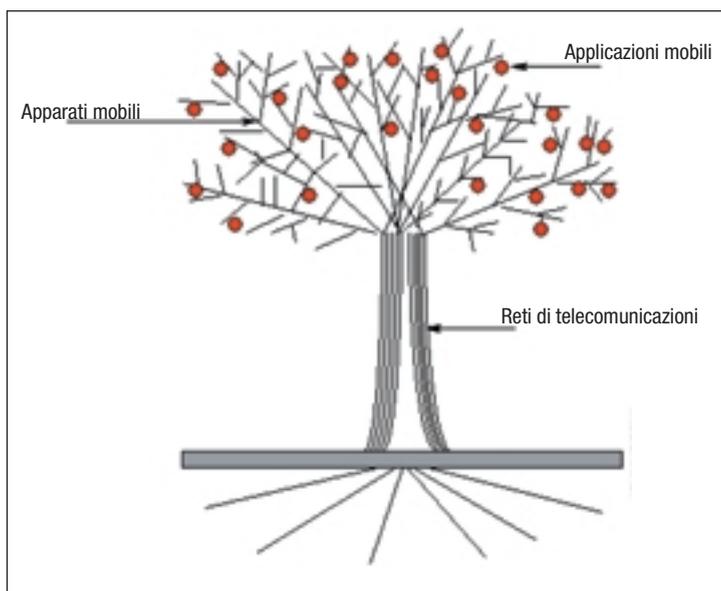
L'espressione *mobile computing* contemplerebbe, a ben vedere, solo l'elemento mobilità, non necessariamente la connettività e la comunicazione con altri dispositivi. Va quindi considerato che le tecnologie mobili devono molte delle proprie potenzialità e del successo odierno non solo alla "nomadicità" con esse consentite, quanto alla capacità di connettersi a reti di telecomunicazioni, e quindi di essere primariamente degli strumenti di comunicazione. Questa circostanza comporta una fitta rete di apparati fissi, come ad

esempio i rilevatori geostazionari, che usano tecnologie mobili (*GSM*, *GPS*), oppure le celle e i ripetitori per le applicazioni di telefonia mobile. Tutte queste attrezzature, senza le quali non ci sarebbe connettività e quindi solo una relativa mobilità per i riceventi, sono saldamente ancorate alle loro posizioni. Ne consegue che un'esaustiva descrizione del campo dovrebbe comprendere anche una descrizione esauriente di questi aspetti del mobile, ovvero delle tecnologie su cui possono contare gli apparati mobili per poter essere connessi durante il movimento. Nella discussione sul mobile bisognerà quindi considerare differenti livelli di argomentazione. L'universo mobile è costituito da un insieme di tecnologie tra loro integrate in maniera sinergica.

Il livello più basso è quello delle *tecnologie abilitanti*, ovvero quelle tecnologie di base che consentono, per esempio, la miniaturizzazione degli apparati (più piccolo vuol dire più facile da trasportare), l'autonomia rispetto alle fonti energetiche fisse, la disponibilità di schermi sufficientemente luminosi ecc..

Il livello successivo è quello delle *reti di telecomunicazioni*: questo livello contiene sia la relazione degli apparati mobili con la rete internet, sia la possibile relazione degli apparati mobili con una pluralità di reti private. Inoltre, esiste una pletera di tecnologie accessorie ormai consolidate nel mondo delle telecomunicazioni, di cui il mondo mobile (viste le sue limitazioni oggettive) trae beneficio: per esempio, la possibilità di trasmettere e ricevere dati di differenti natura (multimedialità), l'ottimizzazione della larghezza di banda attraverso gli algoritmi di compressione, la gestione della continuità del servizio a fronte della mobilità ecc..

Il terzo livello di indagine riguarda la struttura tecnologica degli *apparati mobili* che, come si è detto, travalica di molto il concetto di computer. Infine, abbiamo il livello *delle applicazioni*, un campo strettamente legato all'evoluzione tecnologica dei dispositivi mobili e contemporaneamente all'evoluzione dei sistemi informativi in generale. Va detto che il campo delle applicazioni mobile sta crescendo molto, in direzioni assolutamente variegata. Da quanto emerge dallo studio pubblicato di recente dal Primo Osservatorio sul *mobile business* in Italia [4], il settore delle applicazioni



mobile ha ormai abbracciato un po' tutte le aree di applicazioni nei sistemi informativi, dalla gestione della forza vendita basata su reti cellulari alle applicazioni di *mobile-to-mobile* legate al tele-controllo, al tele-rilevamento e alla tele-lettura, dal monitoraggio della gestione del magazzino attraverso dispositivi mobili al mondo delle applicazioni legate alla nuova tecnologia RFDI (*Radio Frequency Identification*), dalle applicazioni CRM (*Customer Relationship Management*) all'idea del supporto al lavoro quotidiano dell'utente nell'ambito del proprio ufficio (*wireless office*), dalla gestione delle flotte di automezzi ai rapporti con la catena dei fornitori. Altrettanto interessante risulta essere l'ormai raggiunta piena integrazione tra le applicazioni mobile e le applicazioni tradizionalmente facenti parte del sistema informativo. La possibilità di poter visionare la posta elettronica sul proprio computer o su un cellulare, connettendosi allo stesso server di posta, oppure quella di poter utilizzare un applicativo gestionale dal proprio posto di lavoro o da un palmare, rappresentano chiari esempi di questa convergenza e di questa avvenuta integrazione. Nella figura 2 è mostrata una rappresentazione ideografica del campo.

È quindi evidente la complessità, la pervasività e l'ineluttabilità dei cambiamenti introdotti nella nostra era dal concetto di "mobile". Non è semplice prevedere gli effetti di questi cambiamenti sulla nostra società e sul nostro

FIGURA 2

Lo spazio semantico del mobile computing

modo di vivere, in quanto la capacità di elaborare in movimento si sposa con una diffusione dei processi di elaborazione in apparati sempre nuovi. L'universo potenzialmente risultante è quindi enormemente differente da tutto ciò a cui siamo, in quanto specie, abituati. Può essere utile qui riflettere brevemente sulla dimensione storica del fenomeno.

La capacità distintiva degli umani di rappresentare il proprio ambiente esterno e interno è cominciata 2.000.000 di anni fa (ancor prima della nascita dell'*homo sapiens sapiens*) con le modificazioni biologiche che hanno portato allo sviluppo della laringe (possibilità di fonazione articolata) e della corteccia cerebrale. L'acquisizione del linguaggio, così come oggi lo intendiamo, è di circa 300.000 anni fa, e segna la prima tappa epocale della civilizzazione. È la nascita della cultura: per la prima volta le conoscenze apprese possono essere accumulate e trasmesse. Ma le culture orali producono tecnicamente un certo tipo di pensiero, che potremmo definire mitico/eroico, dalla forma che il pensiero deve assumere per poter ricordare e quindi trasmettere le rappresentazioni del mondo. All'incirca nel 3000 A.C. la nascita della scrittura segna un nuovo cambiamento epocale. Ora le rappresentazioni del mondo possono essere fissate su un supporto e trasmesse da una generazione ad un'altra senza che esista la necessità di un contatto diretto tra gli autori della rappresentazione e i loro lettori. Anche questa innovazione tecnologica ha mutato radicalmente il pensiero, che smette di essere mitico/eroico per diventare un pensiero logico/razionale. Un modo diverso di comprendere il mondo è ora possibile; ad esempio la scrittura rende possibile la nascita della matematica (impensabile in una società senza scrittura). Per quanto altre innovazioni tecnologiche abbiano contrassegnato la storia successiva (stampa, radio, televisione) nessuna di esse è dal punto di vista della civilizzazione, una tappa epocale. Certamente ognuna di esse ha avuto effetti importanti sullo sviluppo dei fatti del mondo, ma non ne ha mutato radicalmente il presupposto di fondo. Bisognerà aspettare la seconda metà del 1900 per la successiva tappa epocale, ovvero la nascita della elaborazione automatica. Ora le rappresentazioni sono elaborabili su un processore

artificiale oltre che raccontate, scritte, duplicate, fotografate, filmate, riprese mediante strumenti artificiali/naturali. È questa capacità di elaborazione che permette anche la comunicazione tra artefatti e quindi l'universo della comunicazione elettronica che cominciamo ad intravedere oggi. A noi che osserviamo il fenomeno essendone coinvolti come utenti, consumatori di tecnologie, comuni cittadini o esperti delle sue diverse sfaccettature, può essere utile considerare da questa prospettiva il fenomeno delle TIC (*Tecnologie dell'Informazione e della Comunicazione*). Ritornando infatti al *mobile computing*, esso appare in questa prospettiva come il dispiegarsi delle potenzialità insite nella elaborazione su processori artificiali. Tutto ciò avrà effetti profondi sul modo di pensare della specie anche se per ora non possiamo prevedere appieno la natura di questi effetti.

Da quanto detto, risulta chiaro come l'universo del *mobile computing* in tutta la sua estensione semantica è difficilmente sintetizzabile in un contributo singolo: lo spazio necessario supererebbe di gran lunga quello disponibile. Affronteremo quindi questa rassegna limitandoci agli aspetti innovativi che riguardano gli utilizzatori di apparati mobili, ovvero le conseguenze che il mondo mobile ha sugli utenti finali, con alcuni esempi di ambiti applicativi relativi all'uso di questi apparati. Pur con questa restrizione, l'estensione del lavoro rimarrebbe comunque molto ampia. Concentreremo quindi la nostra attenzione su una porzione di tali applicazioni, ovvero quelle che a nostro giudizio rappresentano esempi significativi di come il mondo del *mobile computing* stia cambiando il rapporto tra gli utenti finali e le tecnologie della computazione.

Il lavoro è organizzato come segue: il paragrafo 2 presenta una (necessariamente) breve rassegna delle applicazioni mobile. Il paragrafo 3 prende in considerazione le caratteristiche degli agenti software che nel campo del *mobile computing* contribuiscono all'elaborazione di informazioni *context sensitive*. Il paragrafo 4 presenta un settore in grande evidenza per il mondo mobile, ovvero il *mobile learning*, ampliando la nostra rassegna sugli ambiti applicativi *mobile* ritenuti interessanti per questa trattazione. Il paragrafo 5 presenta come approcci per la localizzazione

e personalizzazione delle informazioni basati su tecniche di profilatura dell'utente e rappresentazione ontologica del dominio applicativo possano essere utilizzati nel contesto del *mobile computing* per agevolare gli utenti nella ricerca di documenti. Alcune considerazioni finali concludono il lavoro.

2. L'UNIVERSO DELLE APPLICAZIONI MOBILI

Telefoni cellulari, computer portatili e Web stanno rapidamente diventando il sistema più utilizzato per essere sempre in stretto contatto con le idee e le attività che contano per noi. L'hardware e il software che si nascondono dietro questi processi non diventeranno mai invisibili, ma saranno sempre meno intrusivi, consentendo di focalizzare la nostra attenzione sui contenuti informativi veicolati. Ci stiamo servendo di questi nuovi strumenti di calcolo portatili per interagire con gli altri in modi che nessuno aveva previsto e non sarà facile separarci da tali strumenti. Al di là dell'aspetto puramente tecnologico, il successo del *mobile computing* è strettamente legato alla disponibilità di applicazioni per gli utenti finali. È noto, infatti, come il carburante per lo sviluppo di una nuova tecnologia sia rappresentato dal miglioramento delle condizioni di lavoro/vita/svago che l'utente trae dall'utilizzo di quelle tecnologie. La disponibilità di applicazioni da svolgere in mobilità o, anche solo in condizioni nomadiche costituisce quindi un fattore cruciale per questo settore: di recente, una serie di migrazione tecnologiche ha consentito di migrare applicazioni tipicamente riservate al mondo dei computer verso gli apparati mobili. In generale, la progettazione e lo sviluppo di sistemi intelligenti nelle applicazioni informatiche odierne prevede la possibilità di fornire parte o tutti i servizi previsti anche su piattaforme mobili. Il *mobile computing* apre spazi di sviluppo praticamente in tutti i campi dove i computer possono essere utilizzati. È quindi del tutto prevedibile che in futuro qualunque sistema informatico prevederà come default la possibilità di interagire con utenti in movimento.

La possibilità di elaborare mentre i sistemi sono in movimento influenza maggiormente alcuni campi applicativi rispetto ad altri. A

ben vedere però, questo è dovuto soltanto ad una immaturità tecnologica: in linea di principio, tutti i sistemi informatici hanno degli utenti e questi ultimi hanno interesse a spostarsi nello spazio fisico senza rinunciare a interagire con sistemi di cui sono appunto utenti. Inoltre, molti oggetti di uso comune (macchine fotografiche, stampanti, cellulari ecc.) sono di fatto apparati di elaborazione. Tali apparati possono comunicare direttamente tra loro senza la mediazione di un computer general purpose. Esempi già disponibili sul mercato sono le macchine fotografiche digitali che si connettono direttamente alla stampante, oppure i pagamenti effettuati con il telefono cellulare verso distributori automatici di vario genere. Se consideriamo inoltre apparati con un alto grado di autonomia (robot capaci di muoversi autonomamente in uno spazio e di connettersi a sistemi di comunicazione), l'universo delle applicazioni ipotizzabili diviene talmente vasto da rendere difficile prevederne l'evoluzione, se non nelle sue linee essenziali. Queste prospettive vanno quindi ben oltre il classico slogan *anywhere, anytime, anyplace*. In linea di principio, agenti hardware o software potrebbero essere ovunque, liberi di muoversi nello spazio e di effettuare operazioni di elaborazione per nostro conto. In effetti, le sonde semoventi che negli anni scorsi sono state fatte atterrare su pianeti del sistema solare sono esempi di questa classe di robot. *Anywhere, anytime, anyplace* risulta quindi uno slogan riduttivo, perché non solo gli utenti possono muoversi nello spazio teoricamente (almeno per ora) anche i loro agenti robotizzati potrebbero perlustrare lo spazio in modo parallelo. Questo ragionamento porta a fare luce su un altro settore in grande espansione nel mondo del *mobile computing*, ovvero quello degli "agenti", di cui parleremo nella prossima sezione. Già oggi, per esempio, possiamo equipaggiare la nostra abitazione di campagna e il nostro appartamento in città con telecamera, e controllare il tutto dal nostro cellulare. Sicché mentre l'utente è in movimento, può controllare la situazione in due spazi fisici remoti. In altri termini, possiamo avere contemporaneamente la percezione di tre luoghi fisici diversi di cui due virtualizzati.

Anche la “domotica”, quindi, viene ad avere molti punti di contatto con il mondo del *mobile computing*: si veda, per esempio, [5] per una rassegna delle architetture *home-oriented* che consentono mobilità rispetto alla locazione fisica di dispositivi, adattatori, servizi e componenti chiave.

Come ogni innovazione tecnologica che estenda la capacità di comunicare e percepire, la mobilità continuerà ad avere in misura progressivamente crescente, effetti culturali, sociali, economici e giuridici difficilmente prevedibili. Un esempio concreto è rappresentato da un settore apparentemente lontano dal mondo tecnologico. Le questioni giuridiche connesse alla mobilità sono balzate alla ribalta quando ci si è chiesti, ai sensi della nuova normativa sul trattamento dei dati personali (D.lg. 196/2003), quali fossero le conseguenze relative all’uso delle telecamere dei cellulari in termini di *privacy*. Non è quindi difficile immaginare quali e quante questioni giuridiche si genererebbero se un robot che si comporti come un *avatar* (personaggio sintetico di mondi virtuali) andasse a spasso per una città trasmettendo al proprio utente immagini e suoni di vario genere. Citando ancora esempi di problemi legati alla disponibilità di informazioni sullo spazio e alla mobilità delle persone, si pensi alle recenti vicende legate al motore di ricerca Google™ e al suo servizio Google Earth. Il *mobile computing* e gli agenti consentono quindi di configurare ipotesi paradossali in cui il soggetto blocca la propria mobilità a favore di agenti software che suppliscono alla sua staticità con la *propria* mobilità, rendendo, di fatto, l’utente mobile solo virtualmente.

Analizzando il recente mercato del *mobile computing*, i campi applicativi maggiormente esplorati riguardano alcune aree quali:

- *mobile commerce*, area questa vastissima che comprende sia gli aspetti tecnologici, sia la creazione di modelli di business e prodotti specifici per il mobile-commerce;
- servizi di *location-based* con conseguente personalizzazione del servizio;
- *front-end* mobili per applicazioni aziendali tradizionali;
- *mobile health*, ovvero le applicazioni mobile applicate alla medicina e a tutte le sue componenti;

□ servizi legati all’apprendimento mediato da computer in situazioni nomadiche (*mobile learning*).

Sull’onda del miraggio dell’*e-commerce*, a fine anni ‘90 si cominciò a parlare di *mobile commerce*, prospettiva particolarmente interessante vista la crisi dell’*e-commerce* dovuta principalmente alla mancanza di dispositivi trasportabili e sicuri per effettuare micropagamenti. In questo caso, l’apparecchio mobile era rappresentato dal telefono cellulare, il quale poteva fornire con le sue capacità di calcolo adeguata potenza per gli algoritmi di crittografia e per la gestione della comunicazione sicura. Altrettanto esplorato il *mobile banking*, con un interesse prevalentemente monodirezionale, in consultazione di estratti conto, posizioni contabili oppure come strumento di ricezione di SMS a fronte di eventi sul proprio portafoglio. Ancora poco esplorata, in questo settore, la possibilità di passare ad effettuare operazioni o disposizioni attraverso l’apparecchio mobile.

I servizi di localizzazione e personalizzazione delle informazioni utilizzano tecniche di profilatura dell’utente assieme ad una rappresentazione ontologica del dominio applicativo; essi rispondono all’esigenza di agevolare gli utenti nella ricerca (anche *online*) di documenti, sia riducendo i tempi delle ricerche, sia razionalizzando i criteri della ricerca in relazione agli interessi degli utenti. Le tecnologie utilizzate per implementare tali servizi devono essere adeguate a gestire sia una struttura di attività sempre più distribuita logicamente e fisicamente, sia una tecnologia che diventa sempre più pervasiva e onnipresente nell’ambiente in cui gli utenti operano. Fra le tendenze più importanti, e forse inattese, nel campo delle applicazioni facciamo riferimento all’avvento del Web come piattaforma per il software sociale, e le relative estensioni nel mondo mobile. La lista delle applicazioni sociali è già impressionante. I primi esempi riguardano la messaggistica testuale verso telefoni e cercapersone, per arrivare ai servizi telefonici basati sulla commutazione di pacchetto, incluse le nuove reti *Voice Over IP* (VOIP), oppure all’area della gestione sociale della conoscenza, rappresentata in parte da siti di interconnettività sociale come

Friendster (<http://www.friendster.com/>), LinkedIn (<https://www.linkedin.com/>) o Ecademy (<http://www.ecademy.com/>). Si assiste cioè alla trasformazione dell'originario Web costituito da documenti statici in una raccolta di pagine che rappresentano interfacce aperte su piattaforme applicative. Questi servizi stanno moltiplicandosi così in fretta perché possono essere realizzati con l'aiuto di strumenti di programmazione e linguaggi condivisi e standardizzati, sviluppati per lo più nell'ambito della comunità dell'*open source*. Il settore degli agenti software intelligenti ha trovato nuova linfa proprio nel *mobile computing*: avendo un dispositivo con limitate capacità di calcolo e di connessione, l'idea di un agente al quale affidare una ricerca ed ottenere la segnalazione dei risultati sul dispositivo mobile è un campo di sperimentazione piuttosto affermato.

Anche per quanto riguarda i software tradizionali di comunicazione, le estensioni nel mobile sono ormai presenti in tutti i prodotti: gli ambienti di sviluppo dispongono di specifici *framework / template* per lo sviluppo di applicazioni per dispositivi mobili, oppure prevedono la trasportabilità delle applicazioni *desktop* verso mondi *mobile*. I requisiti precedentemente individuati come importanti nei campi applicativi esplorati (località, percezione, personalizzazione, comunicazione ecc.) sono tipicamente soddisfatti da modelli basati sugli "agenti". Il prossimo paragrafo analizza caratteristiche, limiti e possibilità di tale approccio.

L'ambito applicativo gestionale, ovvero quello legato alla gestione dei sistemi informativi aziendali, rappresenta il settore dove si è assistito al maggior sviluppo di applicazioni nel mondo mobile. In questo ambito, sostanzialmente si assiste alla creazione di estensioni delle parti più significative del sistema per l'utenza mobile in modo tale che gli utenti lontani dalla sede aziendale possano accedere al sistema informativo. Volendo presentare alcune tra le applicazioni di maggior successo che vedono il dispositivo mobile come strumento di accesso, possiamo elencare le seguenti macro-categorie applicative:

a. gestione posta elettronica, agenda, contatti, ovvero il dispositivo mobile come *per-*

sonal organizer connesso con il sistema informativo aziendale;

b. accesso al sistema di contabilità / ERP (*Enterprise Resource Planning*) / CRM (*Customer Relationship Management*) aziendale;

c. gestione ordini e ciclo passivo da remoto, specialmente per rappresentanti, agenti di commercio ecc.;

d. gestione vendite dettaglio / magazzino / ristorazione: particolarmente in evidenza il settore dell'impiego dei dispositivi mobili come terminali per la raccolta delle ordinazioni nei locali al pubblico;

e. controllo remoto apparecchiature, dalla diagnostica alla videosorveglianza.

L'ambito applicativo mobile legato alla medicina è principalmente strutturato da un lato sulla mobilità del medico, considerandolo come utente che necessita (al pari di altri utenti) di accedere alle informazioni contenute nel proprio sistema informativo mentre si trova lontano da esso (per esempio, consultazione di cartelle cliniche), dall'altro ci si interessa di mobilità in merito a tutto il problema della diagnostica remotizzata, o comunque effettuata con apparecchiature non fisse. Per certi versi, la mobilità diagnostica rappresenta (con le dovute proporzioni) un passaggio epocale che ha qualche similitudine con quello affine della portabilità dei computer. Immaginare di effettuare diagnosi computerizzate "sul campo", e quindi di portare le proprie apparecchiature medicali elettroniche con sé invece che dover portare il paziente presso la struttura dove l'apparecchiatura si trova, ha punti di contatto con la possibilità per l'utente di portarsi appresso il computer, invece che doversi muovere verso il luogo dove il computer si trova.

Per concludere questa rapida rassegna di alcuni dei settori applicativi mobile emergenti nell'ambito della ricerca applicativa, va detto che l'attenzione fino ad oggi è stata maggiormente dedicata agli aspetti prettamente tecnologici, legati all'hardware di dispositivo e di rete che consente la mobilità. Analizzando le riviste e le conferenze del settore, si nota una grande attenzione allo sviluppo delle infrastrutture di rete e di gestione delle reti mobili, allo sviluppo di protocolli, alla sicurezza, alla velocizzazione e miniaturizzazione dei dispositivi (si pensi all'attenzione dedicata dagli RFID e dalle

nanotecnologie), ai sensori, alle reti e ai relativi protocolli, alla gestione degli errori, alle reti veicolari, alle tecnologie per la fornitura di elettricità, alle tecnologie per i dispositivi di input/output, ai problemi di routing ecc..

Una attenzione minore è stata invece dedicata ai servizi (o almeno non sembra emergere la consapevolezza che essi vadano per certi versi riprogettati), al software, alle piattaforme di sviluppo, ma soprattutto ai modelli di sviluppo nel settore investito dall'ondata mobile. Citiamo due esempi significativi:

□ Nell'ambito del *mobile commerce*, i dispositivi sono tecnologicamente pronti per effettuare transazioni economiche, ma manca un'analisi raffinata sulla psicologia del consumatore, che ancora è restio a virtualizzare il proprio rapporto con il bene acquistato e/o con le conseguenti modalità di pagamento. In questo caso, le carenze sembrano essere prevalentemente strutturali del sistema economico/finanziario, legato alla sicurezza necessaria per effettuare transazioni economiche da dispositivi mobili.

□ Nell'ambito dell'*e-learning*, grande attenzione è tuttora dedicata all'idea di *mobile learning*, ma poco si è ancora fatto per capire se questo modello di apprendimento può essere efficace, e in quali forme. In questo caso, neanche la tecnologia è del tutto pronta, soprattutto in termini di dispositivi di input/output. Ripareremo di questo tema nel paragrafo 4.

3. IL PARADIGMA AD AGENTI PER IL SUPPORTO AL MOBILE COMPUTING

In questi ultimi anni numerosi sono gli esempi applicativi pratici dell'uso di un nuovo paradigma di sviluppo software denominato *Programmazione Orientata agli Agenti*. Tale paradigma nasce dalla fusione di alcuni concetti dell'intelligenza artificiale con la tecnologia degli oggetti distribuiti. Il paradigma ad agenti prevede la realizzazione di un'applicazione software come collezione di componenti, gli agenti appunto, autonome, proattive e comunicative. Le caratteristiche principali che permettono di descrivere un agente sono: la sua architettura, i ruoli che esso ricopre e le ontologie che esso conosce.

□ L'architettura determina il modo in cui l'agente pianifica le proprie strategie; essa è strettamente correlata al grado di autonomia, di reattività e di proattività dell'agente. Il modello architetturale di una applicazione realizzata con tecnologia ad agenti è intrinsecamente *peer-to-peer*, in quanto ogni agente è potenzialmente in grado di iniziare una comunicazione con ogni altro agente del sistema.

□ I ruoli determinano il tipo di interazione possibile fra gli agenti; determinano allora quali messaggi l'agente è in grado di ricevere e quali può trasmettere, nonché l'ordine con cui gestirli.

□ Infine, le ontologie servono a strutturare e rappresentare la conoscenza di ogni agente. In un contesto multi-agente è necessario gestire in maniera opportuna le possibili diverse ontologie, sia per il funzionamento del singolo agente che per lo scambio di informazioni fra gli agenti.

Individuare classi di agenti (cioè agenti che ricoprono gli stessi ruoli, hanno la stessa architettura e condividono le stesse ontologie) può essere di aiuto per la corretta analisi, ed il successivo sviluppo, di un'architettura ad agenti. I recenti domini applicativi in cui la tecnologia degli agenti mobili ha mostrato la sua efficacia vanno dalla gestione integrata di reti, di sistemi e di servizi, al supporto al *mobile computing*, dal commercio elettronico, al reperimento intelligente e pianificato di informazioni sul Web. La caratteristiche del paradigma ad agenti risultano infatti tali da fornire soluzioni appropriate in conseguenza della capacità degli agenti di migrare localmente verso la risorsa su cui operare e della loro asincronicità rispetto all'utente. Si può quindi dimostrare utile per realizzare interazioni remote robuste su reti non affidabili come quelle wireless e per ridurre il traffico sulla rete rispetto a una modalità rigida di tipo client/server. Analizziamo alcuni casi in cui un modello ad agenti mobili può essere efficientemente adottato.

□ Un'applicazione client/server in cui il client ha un numero limitato di interazioni con l'utente, mentre deve eseguire un numero elevato di interazioni con il server. In questo caso, il client potrebbe essere modellato come un *agente mobile* che si sposta nel nodo server, esegue localmente le interazioni e restituisce il risulta-

to, riducendo il carico della rete e quindi la latenza (le interazioni diventano locali).

□ Un'applicazione che deve periodicamente monitorare lo stato di un insieme di risorse remote. L'applicazione potrebbe essere un *agente mobile* che si sposta direttamente sui nodi dove si trovano le risorse e le interroga localmente restituendo gli eventi di interesse al gestore. La ricerca delle informazioni sul Web ricade in questa tipologia di applicazione; gli agenti potrebbero per esempio monitorare la pubblicazione on line di articoli scientifici in maniera asincrona ed autonoma alla ricerca di quelli di possibile interesse per l'utente. Inoltre, anziché esplorare Internet da un unico sito centralizzato, gli agenti potrebbero muoversi attivamente attraverso i vari siti www, riducendo il traffico di rete e gestendone la ripartizione del carico computazionale.

□ Un'applicazione per la gestione di una segreteria automatica che interroghi ciclicamente i membri di un gruppo di lavoro per fissare la data di un incontro.

Nello scenario degli agenti mobili non devono essere sottovalutati gli aspetti legati all'interoperabilità e quelli relativi alla sicurezza. Le specifiche per l'interoperabilità tra agenti realizzati da produttori diversi e con tecnologie diverse sono gestite a livello internazionale da FIPA (*Foundation for Intelligent Physical Agents*) (<http://www.fipa.org>), che definisce un linguaggio di comunicazione tra gli agenti detto ACL (*Agent Communication Language*) e le modalità di interazione fra gli agenti, senza entrare nello specifico della struttura interna di un agente.

I numerosi vantaggi descritti possono essere ottenuti, per esempio, nei casi applicativi sopra citati, campi in cui la tecnologia ad agenti mobili si preannuncia efficace; molto lavoro deve però ancora essere fatto per verificare queste possibilità in altri ambienti operativi. Uno dei maggiori ostacoli all'ampia diffusione del paradigma ad agenti è legato agli aspetti della sicurezza concernenti gli sviluppatori, gli amministratori e le informazioni in oggetto. Molti meccanismi di sicurezza sono stati proposti per affrontare i rischi di tipo agente/agente, agente/piattaforma e piattaforma/agente. Spesso però introducono vincoli nelle performance che possono anche annullare i benefici dell'utilizzo di agenti mobili. In particolare, un agente

software migra da un host ad un altro trasferendo codice, dati e informazioni relative al suo stato che gli permettono di recuperare l'esecuzione dal punto in cui essa era prima di lasciare l'host precedente. Durante la sua navigazione nella rete, l'agente è esposto ad attacchi e può essere fonte di attacchi verso altri. Alcuni fra i problemi che devono essere risolti riguardano la protezione dell'host verso agenti ostili e la protezione dell'agente verso host ostili [6]. Meccanismi tipo quelli di controllo degli accessi, protezione tramite password, e sand boxes permettono di risolvere il primo problema; il secondo appare più difficile. L'ambiente di esecuzione (*host*) ha generalmente il controllo pieno sul programma che si sta eseguendo, quindi per proteggere un agente mobile da *host* malintenzionati si deve ricorrere a *tamper-proof hardware*. Per esempio, Yee propone un approccio che usa coprocessore sicuro che esegue le elaborazioni critiche e memorizza le informazioni critiche in un registro sicuro. Altri approcci sono presenti [7].

4. MOBILE COMPUTING PER E-LEARNING

La diffusione di dispositivi mobili nella società e nel mercato è ormai un dato di fatto che comporta una serie di conseguenze pratiche sulle organizzazioni e sugli individui assai rilevante. Anche il mondo della formazione, nelle sue varie espressioni, non è esente da queste conseguenze: molti individui, operatori di settore, ricercatori e organizzazioni stanno considerando i dispositivi mobili come strumenti potenzialmente utili per l'apprendimento e l'addestramento. I vantaggi di questo approccio alla formazione sono ovvi, per quanto abbiamo detto fino ad ora. La "edu-ubiquità", come potremmo definirla, rappresenta un aspetto allettante, soprattutto nel settore della formazione situata (*situated-learning*) e, in generale, nei contesti lavorativi. Contemporaneamente, l'applicazione di queste tecnologie all'apprendimento non è esente da aspetti problematici.

Questo campo di applicazione specifico per il *mobile computing* è ormai conosciuto come *mobile learning* [8]. La definizione di *mobile learning* è piuttosto semplice: la specializzazione dell'*e-learning* che si occupa del-

l'utilizzo di apparecchiature mobili (telefoni cellulari, *smartphones*, *Personal Digital Assistant* (PDA), *TabletPC* ecc.) nell'ambito dell'attività didattica. Ally [9] definisce come *mobile learning*, o *m-learning*, la distribuzione di materiale didattico elettronico su apparecchiature mobili per consentirne l'accesso da qualsiasi luogo in qualsiasi momento (*anytime-anywhere*). Va comunque tenuto conto che una generalizzazione completa non è (ancora) possibile: il panorama tecnologico dei dispositivi mobili è ancora troppo eterogeneo e in continua evoluzione, per poter configurare una visione unica della disciplina. Passiamo infatti da dispositivi cellulari dotati solamente di interfaccia WAP (*Wireless Application Protocol*, ovvero lo standard di comunicazione che permette la visualizzazione di contenuti multimediali su dispositivi portatili), a telefoni cellulari con accesso ad Internet, fino a PDA o a sub-notebook, senza contare la quantità di dispositivi ibridi che coniugano aspetti di *entertainment* con telefonia e computer.

Il *mobile learning* ha visto, come settore scientifico e come ormai accade da tempo per le discipline legate alle evoluzioni della tecnologia, l'avvento sul mercato delle prime soluzioni riguardanti specialmente i dispositivi palmari ancora *prima* che il mondo scientifico si fosse occupato a fondo della tematica, soprattutto in merito agli aspetti di fruizione di contenuti didattici "mobili" e ai modelli di apprendimento legati a questo settore. Abbiamo assistito inizialmente ad una diffusione di prodotti formativi per i *Personal Digital Assistants* (PDA), nella forma prevalente di materiali didattici semplicemente adattati dalla rispettiva versione per PC, e dall'altro allo studio di aspetti legati principalmente all'elettronica e alle telecomunicazioni. Sui due problemi (a nostro giudizio) centrali, ovvero contenuti specifici per il mondo mobile e modelli di apprendimento "mobile", esistono ancora alcuni punti interrogativi, ed una certa carenza di esperienze e analisi da parte del mondo scientifico. È probabile che pochi si siano ancora avventurati in sperimentazioni approfondite e originali, intendendo con questo termine esperienze che non siano derivazioni di ambienti didattici computerizzati provenienti dalle tradizio-

nali esperienze e-learning, *Computer Based Training* (CBT), *Web-Based Training* (WBT) e via dicendo. Questo ritardo può avere una serie di motivazioni:

- i costi di sperimentazione, legati alla disponibilità presso l'utenza di dispositivi adeguati alla fruizione di contenuti didattici mobili;
- i costi di connessione, particolarmente elevati a tutt'oggi per contenuti multimediali (specialmente se visti dalla prospettiva di uno studente);
- la presenza di oggettive limitazioni hardware, non ancora completamente risolte. Principalmente, come vedremo, trattasi di disponibilità di strumenti di input e di output adeguati. Si può comunque notare che la ricerca nell'ambito del *mobile learning* è particolarmente estesa, anche se molto orientata agli aspetti hardware: per esempio, l'*Institute of Electrical and Electronics Engineers* (IEEE, l'organismo americano che emette le specifiche IEEE) pubblica una rivista prestigiosa solo in questo ambito, *IEEE Transaction on Mobile Computing*. Troviamo lavori che spaziano dagli ambiti legati alle tecnologie mobili e alle periferiche hardware, fino alle teorie dell'apprendimento mediato da strumentazioni mobili passando per il mondo delle telecomunicazioni, la miniaturizzazione e gli aspetti legati ad onde, frequenze e campi. Per dare una idea della vastità dei temi, riportiamo nella tabella 2 la fusione degli argomenti principali di tre recenti conferenze mondiali (*World conference on m-learning -MLearn 2004, 2005* e la prossima 2006) sul tema del *mobile learning*. Come si può notare, il settore riflette praticamente tutti i temi tipici dell'*e-learning*, con l'aggiunta di alcuni aspetti maggiormente legati al rapporto tra tecnologia mobile e apprendimento.

Il *mobile learning*, come si faceva cenno, è però affetto da alcuni problemi strutturali, legati a doppio filo a quanto in realtà tutto il settore del *mobile computing* presenta, ma che l'apprendimento legato all'uso di questi dispositivi per certi versi amplifica. Ne presentiamo di seguito una sintesi:

- Per elaborare elementi multimediali, molto efficaci nella formazione mobile, è necessaria potenza di elaborazione: dopo fasi in cui le applicazioni multimediali erano limitate da questo fattore, la situazione sta rapidamente

ML 2004	ML 2005	ML 2006
○ Accessibility and mobility	• From pilot projects to mainstream implementation: strategies for the large-scale deployment of m-Learning	• Building and implementing m-Learning strategies in educational institutions, corporations and government
○ Ambient intelligence	• Mobile technology to support open and distance learning (ODL)	• Case studies in m-Learning
○ Ambient learning	• M-Learning and mobile technology applications in teaching and learning	• Cost effective management of m-Learning processes
○ Intermittently connected learners	• Mobile technology for student and learning support	• Creating interactive and collaborative m-Learning environments
○ Blended learning	• Mobile technology for educator and teaching support	• Emerging hardware and software for m-Learning
○ Contextual learning	• Assessment techniques and practices in m-Learning	• Evaluation in m-Learning
○ Informal and lifelong learning	• Design and development of learning material for m-Learning	• Implementing m-Learning
○ Development of content/teaching materials	• Informal and lifelong learning with the aid of mobile technology	• Informal learning using mobile devices
○ Development of international standards	• Challenges for m-Learning in developing countries	• Instructional design for m-Learning
○ Innovation in concept, design and pedagogy	• Building and implementing m-Learning Strategies in Educational Institutions, Companies and Public Sector Organisations	• Intelligent agents for learning
○ Learning management systems	• Effective and efficient management of m-Learning processes	• Learning objects and metadata for m-Learning
○ Researching/evaluating mobile learning	• M-Learning management systems (mLMSs)	• Mass personalization and socialization
○ Situated collaborative learning	• Creating interactive and collaborative m-Learning environments	• M-Learning and mTraining
○ Supporting and engaging the mobile learner	• Future trends in m-Learning technology, including the impact of emerging technologies	• m-Learning in developing countries
○ Architecture of Context Aware Learning Technology Systems	• Emerging hardware and software for m-Learning	• m-Learning management systems (mLMSs)
○ Mobile spaces for learning communities	• Mass individualisation	• m-Learning: the next generation
○ Learning methodologies	• Improving the quality of m-Learning through evaluation; including mobile assessment and certification	• m-Learning: theory and pedagogy
○ Instructional design theories	• M-Learning as a tool for social change	• Open and distance learning (ODL) using mobile devices
○ Ad hoc networks for organisational learning	• Ambient intelligence and ambient learning	• Quality in m-Learning
○ Learning objects and metadata for mobile learning	• Innovation in learning theory and pedagogy	• Social implications of m-Learning
○ Mobile learning applications	• Learning objects and metadata for m-Learning	• Student support in m-Learning

TABELLA 2

Temi delle conferenze mondiali su m-learning

0

cambiando. Ormai la convergenza tra telefonia, contenuti, informatica e telecomunicazioni trova il suo punto d'incontro proprio in questo settore, ed è per questo che troviamo grandi investimenti da parte dei costruttori hardware verso i prodotti mobili. La soluzione di questi problemi passa anche per la miniaturizzazione, forse l'aspetto più evidente del progresso tecnologico in atto. È comunque chiaro che la miniaturizzazione nel settore del *mobile learning* non può spingersi molto in là, pena incidere sull'ergonomia dell'apparecchio mobile. È indubbio che potremmo avere dispositivi mobili più piccoli (si pensi alle nano-tecnologie, agli RFID ecc.), ma oltre certi limiti ergonomici non è possibile spingersi.

1

0

□ Un altro aspetto cruciale delle tecnologie mobili è la durata delle batterie. Questo elemento rappresenta la contropartita principale della miniaturizzazione: è importante una riduzione consistente di potenza assorbita dal dispositivo, e quindi una maggiore durata delle batterie. Le migliori pile all'idruro di nickel-metallo, agli ioni di litio e al polimero di ioni di litio riescono a garantire una autonomia ad un computer portatile per un massimo di otto-dieci ore, e ad un telefonino per cinque ore di conversazione continua. In questo settore, notizie incoraggianti vengono comunque dal campo della chimica, dove la continua evoluzione nelle capacità di accumulo della carica elettrica fa intravedere un futuro più appetibile per la mobilità. Le pile di combustibile compatte potrebbero quadruplicare i tempi citati, ma non verranno distribuite in massa prima del 2010.

□ Le limitazioni hardware relative ai dispositivi di I/O. Questo tema rappresenta forse il freno più importante per l'evoluzione del mercato, soprattutto proprio nel settore del *mobile learning*. Le limitazioni riguardano sia i dispositivi di input che di output.

1

0

Per il versante dell'input, alla cronica e ovvia carenza di spazio per tastiere *full-size* si sono accompagnate nel tempo soluzioni di vario tipo. Innanzitutto, soluzioni software come la compilazione intelligente di parole (come lo standard T9), oppure la visualizzazione di tastiere sullo schermo da utilizzare con penne o stilo. Abbiamo poi una notevole inventiva nel campo dei dispositivi di input esterni, che vanno dalla tradizionale tastiera *full-size* (ma-

gari pieghevole), alle più recenti tastiere olografiche nelle quali, come in un normale ologramma, l'immagine di una tastiera reale viene incisa su una pellicola di plastica. Quando la pellicola è attraversata dalla luce, l'immagine della tastiera viene proiettata verso l'esterno, dando l'illusione di avere di fronte un oggetto tridimensionale e quindi "reale". Nell'ambito di soluzioni già disponibili su larga scala, troviamo le tastiere "virtuali" create da semplici proiezioni laser in luce rossa su un piano. Tutte queste soluzioni presentano aspetti positivi e negativi piuttosto evidenti, dall'ingombro del trasporto alla non sempre perfetta applicabilità delle soluzioni di proiezione in presenza di utenti mobili per la necessità di avere a disposizione un piano su cui proiettare la tastiera. Qualche successo stanno riscuotendo dispositivi che integrano nativamente tastiere *full-qwerty*, quindi tutti i tasti necessari già disponibili senza aggiunte, estensioni o emulazioni, anche se questa certo non è una soluzione soddisfacente per chi ha necessità di usare a lungo una tastiera. Forse l'input vocale potrà costituire una ragionevole alternativa ai tradizionali strumenti di input, ma questo è legato alla potenza di calcolo del dispositivo stesso e allo sviluppo di algoritmi di *voice recognition* compatibili con le caratteristiche dei dispositivi mobili (ambienti dal rumore di fondo intrinsecamente non controllabile).

Per il versante dell'output, la disponibilità di schermi piccoli rappresenta una grossa limitazione alla fruizione dei contenuti, ma qui la situazione sta cambiando: il mercato comincia ad accettare dispositivi poco ingombranti, o perlomeno la miniaturizzazione dell'apparato mobile si è fermata di fronte a limiti oggettivi. In apparecchiature mobili attuali, cominciano a diffondersi schermi molto luminosi, retroilluminati, con contrasti elevati per visibilità in situazioni di luce difficili, che raggiungono la risoluzione di 640x480 pixel, ovvero la risoluzione VGA che tanto ha contribuito all'avvio delle applicazioni grafiche nel mondo PC. Una risoluzione VGA su uno schermo di pochi pollici rappresenta un ottimo compromesso tra un moderno schermo di computer e gli attuali schermi dei telefoni cellulari in quali è praticamente impossibile la fruizione di contenuti multimediali. Si può

quindi affermare che le capacità differenti e condizionanti una applicazione estesa del *mobile learning* derivano dalle diverse disponibilità di schermi, e dalla capacità di utilizzo di strumenti di input come tastiere o stilo, con relativi programmi di riconoscimento della scrittura a mano. L'ambito dell'apprendimento su dispositivi mobili è infatti molto legato alla qualità e alle performance di questi dispositivi di input / output.

□ La progettazione dell'interfaccia uomo/macchina. Quest'ultima è ancora scadevole. Ciò è dovuto ad un mix tra esigenze legate alla telefonia e la disponibilità di un dispositivo con caratteristiche e software tipici di un computer. Il rapporto tra l'anima "telefono" e quella "computer" di un dispositivo mobile è ancora tutto da chiarire, e gli apparecchi ibridi (come le varie *Phone Edition* di certi PDA) non sempre brillano di usabilità per l'utenza finale: spesso una delle due anime è privilegiata rispetto all'altra. La progettazione dell'interfaccia risente di altri limiti importanti, legati al punto precedentemente affrontato di disponibilità o meno di certe risoluzioni per gli schermi da un lato, e di dispositivi di input adeguati dall'altro. Con schermi VGA e tastiere *full-qwerty*, la progettazione dell'interfaccia è completamente diversa rispetto all'attuale panorama offerto dagli apparecchi mobili.

□ I problemi relativi alla connessione alla rete. Si tratta di problemi che riguardano i costi e le modalità di tariffazione, la garanzia di qualità e continuità di servizio (*Quality Of Service*, QOS), la gestione delle infrastrutture, la larghezza di banda disponibile, la relativa copertura ecc.. Molto si sta facendo per la parte più tecnologica di questa classe di problemi, ma altrettanta strada resta da fare, soprattutto se vista in un'ottica di *mobile learning*: si pensi solo al versante delle tariffe per la formazione di utenti mobili che volessero fruire di contenuti multimediali. È evidente che, ad oggi, questo tipo di servizio è quasi utopistico non tanto per questioni tecnologiche (o per lo meno, non come causa primaria), ma principalmente per problemi di sostenibilità dei relativi costi. Una istituzione accademica, per esempio, potrebbe attuare delle sperimentazioni di questi servizi su piccoli numeri di utenti, ma nel momento in cui

si dovesse estendere la sperimentazione all'intera utenza dell'Ateneo, le modalità di tariffazione risulterebbero insostenibili per i singoli utenti o per l'Ateneo stesso (qualora si facesse carico dell'intero traffico di rete necessario per erogare contenuti multimediali su mobile). Esistono limitate sperimentazioni in questo senso, spesso però legate a finanziamenti provenienti da progetti europei o a sponsorizzazioni non estendibili su scala nazionale. Se poi passiamo alla formazione primaria e secondaria, la situazione attuale del mercato delle telecomunicazioni mobili non permette di immaginare a breve una diffusione del *mobile learning* in questi ambiti. Un altro problema significativo nel *mobile learning* riguarda la produzione di materiali didattici specificamente progettati e costruiti per questi dispositivi. Da molti studiosi viene rilevata una carenza cronica di progettazione dei cosiddetti *learning objects* specifici per l'ambito mobile [10]. I formatori tendenzialmente dovrebbero progettare materiali didattici non più soltanto per i tradizionali computer, ma adeguare la loro progettazione assecondando o adattando la propria didattica alle oggettive limitazioni di tali dispositivi. La progettazione di materiali didattici per dispositivi mobili è quindi solo in parte assimilabile alle tecniche note nel mondo della formazione a distanza o assistita/mediata da computer. Per i motivi già citati legati ai dispositivi di input / output, alle condizioni ambientali, ai fattori di tempo e di copertura di connessione, la progettazione e realizzazione di materiali didattici mobili dovrebbe seguire le specificità di questa disciplina. Si pensi soltanto alla produzione di testi, alla scelta delle modalità di esposizione dei concetti, alla frammentazione dei contenuti nei *chunk* di un prodotto didattico multimediale. I testi, i ritmi della lezione, gli spazi, le immagini, la frammentazione dei contenuti, in generale i contenuti stessi andrebbero quindi pensati specificamente per il mondo mobile, in modo diverso da quanto accade per l'*e-learning* tradizionale. È ancora frequente l'errore da parte dei docenti di "riciclare" su dispositivi mobili il materiale didattico preparato per schermi normali. Oltre a ciò, normalmente i materiali didattici sono sviluppati in *chunks* che mal si adattano al *mobile learning*.

Un altro filone di ricerca applicata in merito alla questione dei materiali didattici mobile si è orientato alla creazione di strumenti di conversione/ adattamento di materiali didattici pensati per dispositivi computerizzati standard verso dispositivi mobili. Il discorso si fa ovviamente diverso nel caso in cui il *learning object* sia in formato web/puro testo e se sia strutturato in un qualche standard (per esempio, *Sharable Content Object Reference Model* (SCORM)) [11], oppure (come spesso accade) se il contenuto didattico originale sia in un formato proprietario (PowerPoint, PDF, Tex ecc.). La strada seguita è comunque quella di una comunicazione client/server tra dispositivo e piattaforma di gestione dell'*e-learning* (*Learning Management System*), dove il device informa il server in merito alle proprie caratteristiche hardware. In questo senso, esistono alcune esperienze riguardo il *Resource Description Framework* (RDF), in particolare il *Composite Capabilities/Preferences Profile* [12], che è promosso dal W3C come standard di comunicazione per poter scambiare caratteristiche delle apparecchiature e preferenze utente. Gaedke et al. [13] hanno proposto un approccio per automatizzare la conversione del contenuto Web in formato mobile, mentre Castellanos et al. [14] propongono uno schema generale di comunicazione per consentire a dispositivi come i PDA e i telefoni cellulari di scaricare risorse da librerie digitali (*Digital Libraries*, DL). Per quanto riguarda contenuti in formato proprietario, viene proposto [15] un sistema adattativo per i contenuti in grado di decidere la versione ottimale di contenuto per la presentazione e la migliore strategia per derivare quella versione, attraverso un prototipo per l'adattamento di documentazione in formato PDF. Sicuramente, la strada dell'integrazione tra contenuti sviluppati per il mondo desktop e dispositivi mobili passa per il linguaggio XML, ora che anche i dispositivi mobili iniziano a disporre di potenze di calcolo tali da elaborare in tempi ragionevoli lunghi *stream* testuali. In particolare, per i contenuti multimediali [16], l'integrazione con SMIL [17], uno standard XML *content-oriented* ormai giunto alla versione 2.0, consente di prevedere sviluppi interessanti, soprattutto in considera-

zione della futura conversione dei formati proprietari (Microsoft Office innanzitutto) verso il mondo degli standard aperti basati su XML. Rimane comunque evidente la necessità, per i dispositivi mobili, di:

- aumentare la potenza di calcolo al fine di garantire performance adeguate alla visione di contenuti multimediali,
- avere in dotazione tecnologie *wireless* ad alta velocità, al fine di scaricare velocemente tali contenuti che diventano sempre più ingombranti,
- disporre di sufficiente memoria per l'archiviazione di questi materiali.

Esistono molte sperimentazioni con gli SMS [18, 19, 20] con la consultazione di pagine *web mobile*, oppure con l'integrazione tra messaggistica disponibile sul dispositivo mobile e sistema centrale.

Un altro settore dove le tecnologie mobile e sistemi di *e-learning* si incontrano è sul fronte della valutazione: quiz e test si prestano bene alle limitazioni dei dispositivi mobili e soprattutto alla loro caratteristica principale, ovvero *anytime anywhere*. La possibilità per il discente di autovalutarsi nelle situazioni spazio-temporali più disparate rappresenta un elemento di sicuro impatto per questi dispositivi.

Esistono anche strette connessioni tra il tema dell'apprendere in condizioni nomadiche e la localizzazione del soggetto attraverso il *Global Positioning System* (GPS) o altre tecniche satellitari o *wireless*. Questo aspetto, definito *situated learning*, ha delle prospettive molto interessanti soprattutto per il mondo lavorativo. L'esempio tipico riguarda il lavoratore posto di fronte ad un intervento di assistenza di un macchinario, quando necessita di un apprendimento "in loco" in quanto non dispone di tutte le informazioni necessarie per riconoscere il contesto, il cliente o il macchinario, e la disponibilità di un server da cui scaricare materiale informativo atto a risolvere la situazione. Se si vuole estendere il concetto, è evidente che anche le sempre più frequenti esperienze di visite turistiche guidate con dispositivi mobili *context-aware* può rappresentare una forma di apprendimento estendibile a molti contesti.

Per concludere questa panoramica, non possiamo dimenticare come esistano problemi

specifici legati ai modelli formativi mobili, ovvero come cambia la didattica e l'apprendimento a causa dell'utilizzo di questi strumenti, e alle modalità fisiche con cui i soggetti (docenti e discenti) erogano e acquisiscono formazione [21]. Si apprende usando dispositivi mobili? Quale tipo di apprendimento si verifica? Esistono modelli di apprendimento diversi? [22, 23, 24]. Sicuramente le tecnologie mobili costituiscono una grande occasione per lo sviluppo di nuovi scenari formativi, e i problemi legati specificamente agli aspetti hardware e alle prestazioni saranno sicuramente superati nel prossimo futuro. Parte degli interrogativi rimangono invece sulla definizione del modello cognitivo legato al *mobile learning* da un lato, e a certi aspetti connessi con il bisogno "antropologico" di comunicazione degli esseri umani legato alla formazione a distanza in generale, ma sicuramente acuito dalla mobilità del fruitore del servizio.

5. LOCALIZZAZIONE E PRESENTAZIONE DI INFORMAZIONI NEL MOBILE COMPUTING

Nell'area dei sistemi per il reperimento e l'accesso all'informazione riveste un ruolo strategico l'analisi, la progettazione e la realizzazione di strumenti innovativi ed efficaci per l'accesso e la distribuzione personalizzata dell'informazione attraverso le reti mobili. Mediante tali strumenti è possibile portare all'attenzione degli utenti informazioni selezionate, il cui interesse è assicurato da processi automatici in grado di valutare quando e quanto un'informazione corrisponde agli effettivi bisogni informativi dell'utente, minimizzando così la quantità di informazione inutile cui ciascun utente viene esposto. Quando poi il dispositivo utilizzato è di tipo mobile, le limitazioni della rete e dei dispositivi giocano un ruolo importante. In particolare, in base anche alle caratteristiche del dispositivo, serve adattare opportunamente l'interfaccia, a livello grafico e di contenuti, con corrispondente cambiamento della quantità di informazione presentata. Proveremo a fare una rapida sintesi delle applicazioni che si stanno sviluppando in questa direzione, con la consapevolezza che tale sintesi sarà certamente riduttiva e limitata.

La georeferenziazione del soggetto, dell'apparato mobile o di entrambi, ovvero la localizzazione del soggetto o del dispositivo apre prospettive di sviluppo di applicazioni molto interessanti, anche se solleva questioni in merito alla privacy dell'individuo e al rapporto tra informazione georeferenziata e sicurezza / trattamento dell'informazione stessa. Volendo analizzare la localizzazione per quanto di positivo propone, possiamo partire dagli ambiti legati all'*e-government*, ovvero al rapporto tra cittadino e amministrazione pubblica mediato attraverso le ICT. Si assiste con sempre maggiore frequenza ad esperienze di *mobile-government* (*m-government*), ovvero servizi che vengono erogati dalle amministrazioni pubbliche ai cittadini attraverso strumenti di *mobile computing*.

Tra i tanti esempi, citiamo il portale *mobile-city.org*, creato dalla città di Brema, per fornire servizi ed applicazioni mobili a cittadini, piccole-medie imprese e sanità pubblica. Lo scopo è quello di sperimentare vari modelli di business e di servizi su dispositivi mobili (PDA e telefoni cellulari). Sempre in questo ambito, troviamo varie esperienze di *m-government* orientate all'acquisto di biglietti per fruire di servizi pubblici di vario genere attraverso il cellulare, come per esempio quanto avviato dal comune di Francoforte o da Vienna e altre città austriache. In Olanda, la localizzazione e il mondo del *m-government* si incontrano nell'ambito delle piattaforme per i servizi per la Polizia, dove georeferenziazione e localizzazione del singolo agente attraverso *smartphone* consente l'accesso ai servizi di consultazione di mappe, gestione e trasmissione dei verbali, multe, rilievi ecc.. Va certamente ricordata una delle esperienze pilota in questo ambito, ovvero il progetto NOMAD (<http://www.projectnomad.org.uk/>), progetto per lo sviluppo dei servizi di *m-government* il cui obiettivo è quello di intervenire sulla produttività dei dipendenti pubblici che si trovano in situazioni di mobilità, cercando di incrementare la loro efficienza. Con il motto *Efficiency on the Move*, troviamo sperimentazioni sui senior manager pubblici (gestione posta elettronica su dispositivi mobili), l'invio da parte di pubblici ufficiali di dati ed immagini relativi ad incidenti ed infrazioni stradali, il supporto agli ispettori pubblici, una serie di iniziative che puntano alla flessibilità

del rapporto di lavoro attraverso il telelavoro. Nelle Filippine, Paese che ha detenuto di recente il record mondiale di utilizzo di SMS e di penetrazione dei dispositivi mobile rispetto ad Internet (http://www.ndu.edu/irmc/elearning/newletters/newletters_pdf/itto6o3.pdf), sono stati di recente avviati progetti di *m-government* che spaziano dalle esperienze di *citizen participation* (segnalazione via SMS di qualsiasi problema di ordine pubblico o di infrazione; servizio che ha avuto successo al punto che nei primi sei mesi di sperimentazione sono state ricevute in media ben 29.000 messaggi giornalieri), al progetto che mira a ridurre drasticamente, attraverso i dispositivi mobili, i tempi di riscossione e pagamento di tasse, tributi, dazi doganali ecc.. Esistono poi, anche in Italia, esperienze di *mobile government* che, sfruttando la localizzazione e la mobilità, consentono di fornire servizi di utilità pubblica di *alerting* (esempio, in presenza di calamità), informazioni meteorologiche e sulle condizioni del traffico, *reminder* di scadenze fiscali e tributarie, avvisi di entrata in vigore di normative ecc..

Grande successo stanno ottenendo tutti i dispositivi basati sul sistema satellitare GPS. Dai palmari ai sistemi antifurto, da impianti singoli per i PC all'abbinamento con sistemi di navigazione stradale da usare sui veicoli, i dispositivi GPS stanno aggiungendo alla mobilità, intesa come libertà di movimento nello spazio, un nuovo elemento solo apparentemente banale, ovvero la precisione di localizzazione nello spazio del dispositivo e del soggetto qualora egli lo detenga. Elemento chiave di queste applicazioni risultano essere le mappe digitalizzate, ormai dotate di livelli di precisione interessanti, e di interi team di operatori che provvedono al loro aggiornamento, qualora il servizio di georeferenziazione si traduca in guide stradali. Quindi non solo il soggetto è mobile, ma sa anche dove si trova esattamente, dove sta andando e dove può andare. L'arricchimento dei navigatori satellitari con cartine georeferenziate sempre più precise e aggiornate, con caratteristiche sempre più sofisticate e ricche di informazioni, crea una forte attenzione del mercato (consumer, prevalentemente) verso questo settore, e di conseguenza un fiorire di applicazioni nei settori più disparati.

Esistono o sono in fase avanzata di sperimentazione (per esempio, in Provincia di Trento) sistemi di trasporto pubblico in cui l'acquisto dei biglietti avviene via internet e cellulari. Il biglietto viene inviato sul cellulare del passeggero ed il pagamento viene ottenuto attraverso un addebito diretto allo stesso prezzo del sistema cartaceo. Il mezzo di trasporto è costantemente georeferenziato attraverso connessioni GPS: la sua posizione viene sfruttata dal sistema informativo di controllo del traffico nel trasporto pubblico, che comunica in tempo reale le caratteristiche, la posizione, i tempi di arrivo del mezzo ai cittadini che attendono alla fermata. Inoltre, grazie alla mobilità e alla connettività del mezzo di trasporto, informazioni sia di pagamenti effettuati, sia di diagnostica del mezzo, sia comunicazioni vocali tra personale a bordo e centro di controllo garantiscono, grazie alle tecnologie *mobile*, una totale integrazione tra mezzo in movimento e sistema informativo.

La gestione delle flotte in tempo reale grazie alle tecnologie *mobile* rappresenta una preziosa opportunità non solo per il mondo del trasporto pubblico, ma anche e soprattutto nel mondo della logistica, dove è importante effettuare il tracking esatto, per esempio, del pacco spedito, del cargo in navigazione, del prodotto attualmente in viaggio, del container nel porto ecc.. Le esperienze delle grandi compagnie di autotrasporto, oppure i sistemi di tracciamento del pacco adottato dai corrieri internazionali, sono solo alcuni esempi del potenziale che questo settore potrà esprimere nell'immediato futuro. Si consideri, pur nelle polemiche nate dalla loro adozione, anche il potenziale di strumenti di identificazione molto compatti e facilmente installabili come gli RFID. Queste tecnologie, grazie alla loro ridotta dimensione e quindi alla loro mobilità, aprono scenari di identificazione/autenticazione/autorizzazione molto interessanti: si va dalle Smart Card ai problemi di *ticketing*, alla etichettatura di prodotti o animali, alla gestione dell'identificazione dei capi nelle lavanderie, alla gestione della catena di produzione (dove è il prodotto che si identifica rispetto al nastro trasportatore), a strumenti di sicurezza legati a dispositivi anti-rapina, all'identificazione di veicoli per il pagamento di

pedaggi, fino alla gestione degli accessi delle persone agli spazi.

Un'altra classe di applicazioni correlata alla localizzazione e alla presentazione di informazioni all'utente è rappresentata dalle applicazioni gestionali nell'ambito medico. L'azienda sanitaria può implementare un sistema informativo per la gestione dei dati clinici che utilizza dispositivi di comunicazione mobile all'interno di un processo di cura dei pazienti non localizzati in ospedale. Anche i medici stessi, nella mobilità ristretta di un reparto ospedaliero, possono trarre vantaggio dai dispositivi mobili per la consultazione del sistema informativo, la scheda del paziente che hanno di fronte, i dati degli esami effettuati in tempo reale. Da *mobile health* (o *m-health*) a *mobile care*, ovvero personale infermieristico e paramedico (in ospedale o presso l'abitazione del paziente) in grado di accedere attraverso dispositivo mobile a tutte le informazioni cliniche, acquisendo on-line i dati del paziente. Le tecnologie mobile in questo settore si possono spingere fino alla creazione di una piattaforma distribuita di *mobile-procurement*, per accelerare i tempi di approvvigionamento dei farmaci.

Molto apprezzato e diffuso, anche se con dispositivi non sempre tecnologicamente avanzati, risulta essere l'ambito delle applicazioni turistiche correlate con il mondo del *mobile computing*. In questo contesto va precisata la possibilità di implementare nuovi servizi per i dispositivi mobili che integrano un chip per la ricezione del GPS, cioè quei servizi in cui la localizzazione automatica della posizione dell'utente può essere utile per privilegiare la presentazione di contenuti correlati spazialmente e/o logicamente al luogo in cui ci si trova. Per esempio, potremmo leggere la recensione di un ristorante mentre gli passiamo accanto, o visitare un monumento leggendo fotografie e commenti lasciati da altri e prelevati dalla Rete.

Esistono ormai molte esperienze di guide multimediali evolute, utilizzate nell'ambito dei musei, esposizioni, fiere ecc., dove il servizio di localizzazione dell'utente si rapporta con la presentazione di informazioni relative al luogo dove l'utente si trova fisicamente. Di fronte ad un quadro, per esempio, riceverò sul dispositivo mobile una scheda relativa al-

l'opera, commenti sulla stessa di diversi critici, dettagli, biografia dell'autore ecc.. L'ulteriore possibile integrazione di mappe territoriali permetterebbe la generazione di percorsi e di itinerari personalizzati sia in base alla posizione in cui si trova l'utente, che alle sue caratteristiche.

Sempre nell'ambito turistico, troviamo applicazioni che forniscono informazioni (tipicamente sulla disponibilità ricettiva) dell'area in cui si trovano utenti georeferenziati (sistemazioni, eventi, attrazioni ecc.), filtrando i contenuti sulla base delle caratteristiche dell'utente, del dispositivo utilizzato e del contesto di interazione.

L'idea di utilizzare un approccio semantico per la rappresentazione del dominio applicativo su dispositivi mobili anche nel caso di informazioni georeferenziate è frutto di studi e di investimenti di risorse da parte di una larga fetta della comunità scientifica ed imprenditoriale. Se pensiamo al world wide web, da tempo ci si è mossi per adeguare il mondo delle tecnologie a disposizione degli sviluppatori verso il *mobile computing*. Il W3C ha da qualche anno illustrato l'architettura su cui si dovrebbe basare questa nuova generazione del web. Essenzialmente sono stati individuati tre livelli ai quali corrispondono tre nuovi standard indicati come "*Recommendation*": XML, RDF e OWL.

XML mantiene tutte le potenzialità espositive dell'HTML ma consente di definire nuovi *tag* con cui delineare la semantica delle risorse contenute all'interno della pagina.

Il *Resource Description Framework* (RDF) è la struttura ideata e proposta dal W3C per la descrizione delle informazioni all'interno dei metadati. RDF si basa sul linguaggio XML e fornisce un modello per la creazione di dichiarazioni *risorsa-proprietà-valore* con cui definire la risorsa in maniera che un software apposito possa riconoscerla. RDF, inoltre, essendo essenzialmente codice XML, può essere incorporato direttamente all'interno di qualsiasi risorsa; XMP di Adobe™ è, ad esempio, uno strumento che consente di aggiungere metadati basati su RDF alla maggior parte dei formati di file supportati dalle applicazioni che Adobe sviluppa. I valori delle proprietà possono, inoltre, essere a loro volta risorse, creando così meccanismi di de-

finizioni a cascata che agevolano il riuso del codice e l'integrazione dei dati. Le proprietà e le risorse utilizzate all'interno di RDF devono essere univoche per essere comprese senza ambiguità da applicazioni software. I meccanismi appena illustrati permettono però agli autori una forte personalizzazione della semantica; da un lato questo è necessario per mantenere la scalabilità e l'indipendenza del web, dall'altro rappresenta però un problema per quanto concerne l'interoperabilità dei sistemi. Manca una base di conoscenza comune a cui le applicazioni software possano fare riferimento. In questo contesto si inserisce OWL e più in generale a livello delle ontologie. All'interno di un'ontologia vengono definiti i concetti e le relazioni che intercorrono fra essi, appartenenti ad un determinato dominio, modellandoli in maniera *machine-understandable*. Per garantire l'unicità del significato di una risorsa si rende quindi necessario fare riferimento, all'interno dei metadati, ad una o più ontologie. Per standardizzare anche questa caratteristica del semantic web è stato proposto dal W3C OWL. OWL permette di definire, istanziare ed estendere ontologie ottimizzate per il web fornendo tutta una serie di *features* e di vincoli.

Se opportunamente arricchiti con strumenti e meccanismi, gli agenti software sono in grado di interoperare fra loro utilizzando gli aspetti semantici dei dati, cioè di produrre e consumare servizi ed informazioni semanticamente annotate. Lo scenario in cui ci si trova a dover lavorare per integrare i sistemi multiagenti con le tecniche del Semantic Web è spesso caratterizzato da differenti tecniche di modellazione della conoscenza del dominio e da differenti necessità. Da una parte, i più recenti sviluppi nei linguaggi standard legati alle ontologie, dall'altra, la popolarità di Java per lo sviluppo di sistemi multiagente portano alla necessità di avere una rappresentazione delle informazioni più in linea con il modello object-oriented. Per esempio, in [25] viene proposto un approccio alla gestione delle ontologie a due livelli: il primo, incluso in ogni singolo agente, fornisce un supporto distribuito per importare ontologie OWL come una gerarchia di classi *object-oriented*, e uno o più agenti dedicati, chiamati *ontology servers*, che forniscono

una gestione più espressiva ed efficace, poiché centralizzata, delle ontologie condivise agli agenti del sistema.

Recentemente, l'integrazione dei sistemi multiagente e delle tecniche legate al web semantico trovano applicazione negli ambienti intelligenti per l'apprendimento; un esempio interessante si trova in [26]. Uno dei vantaggi, derivante dal disporre di agenti in grado di comprendere il significato delle nostre richieste, è quello della selezione delle informazioni. In effetti, il problema principale dell'organizzazione attuale delle informazioni non è quello del reperimento di materiale bensì della sua selezione. Il Web Semantico offre un supporto agli agenti autonomi per identificare il dominio ontologico di riferimento, in modo da poter filtrare tutte le informazioni non inerenti a tale dominio. I motori di ricerca tradizionali trarrebbero da questa organizzazione enormi vantaggi, in quanto potrebbero indicizzare le risorse in base al loro reale contenuto, e non più sulla base di parole chiave le cui problematiche e limiti sono già stati descritti. Tutte le risorse che ora non sono direttamente reperibili, in quanto costituenti il Web profondo, potrebbero finalmente comunicare tra loro, creando una vera e propria interconnessione tra risorse molteplici; offrendo la possibilità agli utenti di navigare non tra le risorse (le foglie) ma tra i concetti (i rami).

Tramite l'uso degli agenti è possibile velocemente e facilmente fare accesso a specifici dati non strutturati da intranet e siti web, anche protetti da password, inserendo informazioni al nostro posto, ricercando termini, cliccando su link, così come automaticamente visitare siti e restituire risultati in real time. Gli utenti possono per esempio personalizzare la sottoscrizione ad un sito di *news* scegliendo quanto spesso ricevere i dati e tramite quale mezzo. Le informazioni possono essere posizionate in fogli elettronici o database, oppure pubblicate tipo *newsletter*. I dati possono essere spediti organizzando liste di posta elettronica.

Estrarre informazioni da un documento significa dover eseguire sul testo una quantità di operazioni atte alla individuazione, selezione e "disambiguazione" dei termini. Alcune delle parole contenute all'interno di un documento

possono essere considerate come non indispensabili; per dare un senso a quello che si legge si possono eliminare alcune forme grammaticali, come ad esempio articoli e preposizioni. Per esempio, si possono prendere in considerazione solo i nomi ed i verbi, termini che più di tutti gli altri determinano il senso di uno scritto. Gli strumenti per il *Natural Language Processing*, le ontologie e le tecniche di *word sense disambiguation* opportunamente integrate permettono di costruire sistemi ad agenti orientati alla localizzazione e personalizzazione nel reperimento di informazioni.

La necessità di determinare il grado di similarità semantica fra due concetti è un problema che ha occupato un numero sempre maggiore di ricercatori in linguistica computazionale. Questo tipo di misurazioni vengono utilizzate in una numerosa quantità di applicazioni, dal *text summarization* all'*information extraction and retrieval*. Ricerche recenti prendono in considerazione due tipi di misurazioni per il calcolo della similarità fra due concetti: la *semantic distance* in cui si calcola la distanza fra due concetti all'interno di una rete lessicale e la *semantic relatedness*, un approccio più generale perché fra due entità vengono considerate non solo la loro somiglianza di significato ma anche altri tipi di relazioni come la meronimia o l'antonimia. In [27] vengono proposte tecniche ed algoritmi per il calcolo della similarità ed il modello utilizzato per la rappresentazione dei documenti che considera i concetti espressi e non i semplici termini contenuti in una risorsa testuale e le relazioni che intercorrono fra essi.

6. CONCLUSIONI E DIREZIONI FUTURE

In questo lavoro abbiamo cercato di dimostrare come il *mobile computing* modifichi radicalmente, in virtù della capacità di eseguire elaborazioni su apparati in movimento, la percezione utente dei processi di elaborazione e di comunicazione. Il concetto di "mobile" è davvero molto vasto e abbraccia ambiti tra loro diversi ma spesso accomunati, a volte impropriamente: si può parlare infatti di "nomadicità", ovvero la possibilità di collegarsi da fermo con un terminale alla rete nelle aree coperte da segnale opportuno (come per esem-

pio, nell'ambito del WiFi o del futuro WiMAX), ma si può estendere il concetto anche alla "portabilità" (impiego di un terminale ma nell'ambito di soggetti fermi o in movimento a bassa velocità, come i pedoni). Infine, il concetto di "mobilità" nella sua accezione più specifica si aggancia alla vera mobilità del soggetto, indipendentemente dalla velocità con cui il soggetto si sposta (almeno nei limiti attuali della tecnologia, ovvero 250 km/h).

Nell'articolo sono state evidenziate le molteplici applicazioni che si aprono nell'ambito dei sistemi informativi, ma soprattutto nell'ambito del quotidiano vivere dei soggetti e del loro modo di interagire/comunicare con altri soggetti. Ormai gli utenti sono molto attenti alle evoluzioni in questo settore, spesso però soltanto qualora esistano vantaggi evidenti non solo per il proprio lavoro, ma anche e soprattutto per la propria vita quotidiana: si pensi solo alle implicazioni del *mobile computing* sui servizi sanitari, sull'istruzione, sul contatto multimediale con persone lontane, sul sistema dei pagamenti di beni/servizi e via dicendo, da cui ben si comprende come alcuni studiosi abbiano addirittura coniato il termine *mobile life* proprio ad indicare la pervasività delle tecnologie mobili nel nostro vivere quotidiano.

Nell'ambito di questo panorama variegato, ci siamo concentrati sugli aspetti legati a due tra le molte applicazioni che attraggono l'attenzione degli operatori e degli utenti per il prossimo futuro: da un lato la possibilità di utilizzare i dispositivi mobili per attività legate all'apprendimento (*mobile learning*), e dall'altro alle applicazioni legate alla localizzazione e alla presentazione contestualizzata di informazioni. Sono due tra i molti esempi di direzioni future nello sviluppo di questo settore, molto legato alla tecnologia e ai suoi avanzamenti, ma altrettanto condizionato dalla reale utilità di queste tecnologie nell'ambito lavorativo, personale, sociale.

Riferimenti

- [1] Satyanarayanan M.: Pervasive computing: vision and challenges. *IEEE Personal Communications*, Vol. 8, n. 4, April 2001, p. 10-17.
- [2] Weiser M.: *The computer for the 21st century*. Scientific American, September 1991.

- [3] Geier J.: *Wireless LANs Implementing High Performance IEEE 802.11 Networks* (2-nd ed.). Indiana: Sams Publishing, 2002.
- [4] *Mobile Business: un'opportunità per un sistema di imprese in profonda trasformazione*. Primo Osservatorio sul Mobile Business in Italia, 2005, Assinform - School of Management Politecnico di Milano.
- [5] Lucenius J., Suomalainen J., Ventola P.: *Implementing Mobile Access to Heterogeneous Home Environment, HOIT 2003*. The Networked Home and the Home of the Future Conference, April 6-8, 2003, Newport Beach, CA (USA).
- [6] Sander T., Tschudin C.: *Protecting Mobile Agents Against Malicious Hosts*. In: Vigna G., editor, *Mobile Agent Security*, 1988, p. 44-60. Springer-Verlag: Heidelberg, Germany.
- [7] Lee H., Foss J., Harrison S.: *The Use of Encrypted Functions for Mobile Agent Security*. Proceedings of the 37-th Hawaii International Conference on System Sciences, 2004.
- [8] Waycott J.: *An Investigation into the Use of Mobile Computing Devices as Tools for Supporting Learning and Workplace Activities*. 5-th Human Centred Technology Postgraduate Workshop (HCT-2001), Brighton, UK, September 2001, available online at <http://www.cogs.susx.ac.uk/lab/hct/hctw2001/papers/waycott.pdf>
- [9] Ally M.: *Using learning theories to design instruction for mobile learning devices*. Proceedings of the Mobile Learning 2004 International Conference, Rome, 2004.
- [10] Sharples M., Corlett D., Westmancott, O.: The Design and Implementation of a Mobile Learning Resource. *Personal Ubiquitous Comput.* Vol. 6, n. 3, (Jan. 2002), p. 220-234. DOI= <http://dx.doi.org/10.1007/s007790200021>
- [11] Wang T., Chang H., Sie Y., Chan K., Tzou M., Shih T. K.: *Reading SCORM compliant multimedia courses using heterogeneous pervasive devices*. In Proceedings of the 13-th Annual ACM international Conference on Multimedia (Hilton, Singapore, November 06 - 11, 2005). MULTIMEDIA '05. ACM Press, New York, NY, 806-807. DOI= <http://doi.acm.org/10.1145/1101149.1101325>
- [12] CC/PP: At <http://www.w3.org/Mobile/CCPP/>, 25 June 2005.
- [13] Gaedke M., Beigl M., Gellersen H-W, Segor C.: *Web Content Delivery to Heterogeneous Mobile Platforms*. University of Karlsruhe, At <http://www.teco.edu/~gaedke/paper/1998-lncs1552.pdf>, June 12, 2005.
- [14] Castellanos N., Sanchez J.A.: PoPS: Mobile Access to Digital Library Resources. *JCDL*, Vol. 01, 2003, p. 184.
- [15] Lum W-Y & F.C.M. Lau: *A Context-Aware Decision Engine for Content Adaptation*. IEEE Pervasive Computing, July-September 2002, p. 41-49.
- [16] Rodrigo Arias., Grete Pasch, Matthias Reichenbach: *XML saves the day: Porting a Rich-Media Collection to a Mobile Platform in Three Weeks Flat*. M-learn 2005, 4-th World Conference on Mobile Learning, 25 - 28 October 2005, Cape Town, South Africa.
- [17] Ayars Jeff, et al.: editors. 2005. Synchronized Multimedia Integration Language (SMIL 2.0) - [Second Edition]. Available at: www.w3.org/TR/smil20/
- [18] Garner I., Francis J., Wales K.: *An Evaluation of the Implementation of a Short Messaging System (SMS) to Support Undergraduate Students*. Proceedings of the European Workshop on Mobile and Contextual Learning, p. 15-18, Birmingham, UK, June 2002.
- [19] Stone A., Briggs J., Smith C.: *SMS and Interactivity - Some Results from the Field, and its Implications on Effective Uses of Mobile Technologies in Education*. Proc. of IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE 2002), p. 147-151, Växjö, Sweden, August 2002
- [20] Colazzo L., Molinari A., Ronchetti M.: *Integrating Mobile Devices in a multi-platform Learning Management System using web services*. In Proc ED-MEDIA 2005-World Conference on Educational Multimedia, Hypermedia and Telecommunications, Montreal, Canada, 29 jun-2 jul 2005.
- [21] Divitini M., Haugalokken O.K., Norevik P.: *Improving communication through mobile technologies: Which possibilities?*. Proc. of IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE 2002), p. 86-90, Växjö, Sweden, August 2002.
- [22] Kukulka-Hulme A.: *Cognitive, Ergonomic and Affective Aspects of PDA Use for Learning*. Proceedings of the European Workshop on Mobile and Contextual Learning, p. 32-33, Birmingham, UK, June 2002.
- [23] Andronico A., Carbonaro A., Colazzo L., Molinari A.: *Personalization Services for Learning Management Systems in Mobile Settings*. Special issue on the Theme "Adaptivity in Web and Mobile Learning Services", International Journal of Continuing Engineering Education and Lifelong Learning, Vol. 14 - Issue 4/5 - 2004, p. 353-369 (ISSN 1560-4624).
- [24] Traxler J.: *Evaluating m-learning*. Proceedings of the European Workshop on Mobile and Contextual Learning, p. 63-64, Birmingham, UK, June 2002
- [25] Tomaiuolo M., Turci P., Bergenti F., Poggi A.: *An Ontology Support for Semantic Aware Agents*. 7-th International Workshop on Agent-Oriented Information Systems (AOIS), The Netherlands, 2005.

- [26] Huang W., Eze E., Webster D.: *Integrating Semantics of Multi-Media Resources and Processes in e-Learning*. To appear in ACM/Springer Journal on Multimedia Systems, Special Issue on Educational Multimedia Systems, 2005.
- [27] Carbonaro A., Ferrini R.: *Considering semantic abilities to improve a Web-Based Distance Learning System*. ACM International Workshop on Combining Intelligent and Adaptive Hypermedia Methods/Techniques in Web-based Education Systems, 2005.

ANTONELLA CARBONARO è professore associato di Informatica presso il Dipartimento di Scienze dell'Informazione dell'Università di Bologna. Nel 1992 ottiene la laurea in Informatica e nel 1997 termina il dottorato di ricerca su Sistemi Artificiali Intelligenti presso la Facoltà di Ingegneria dell'Università di Ancona. Dal 1997 al 1999 riceve una borsa di ricerca sul tema Intelligenza Artificiale. I suoi interessi di ricerca riguardano la personalizzazione ed i servizi basati sul contenuto nel campo del data e knowledge mining e degli ambienti di apprendimento, ad esempio, per recuperare, filtrare e classificare automaticamente contenuti digitali.
E-mail: carbonar@csr.unibo.it

LUIGI COLAZZO è professore di Database Management e Multimedia alla Facoltà di Economia dell'Università di Trento. I suoi interessi scientifici e di ricerca sono nell'area dei sistemi per il recupero dell'informazione, delle architetture software per il Data Base Knowledge Discovery e dei sistemi per il supporto alla cooperazione (CSCW). Dal 1989 il suo lavoro è orientato all'uso di sistemi Web ipermediali per il supporto alle attività didattiche. E' membro di diverse associazioni quali IEEE, ACM, AACE, e AICA.
E-mail colazzo@cs.unitn.it

ANDREA MOLINARI è consulente aziendale e docente universitario a contratto nell'ambito dell'Information & Communication Technology. Svolge attività didattica presso la Facoltà di Ingegneria e la Facoltà di Economia dell'Università di Trento, e presso la facoltà di Computer Science della Libera Università di Bolzano. Si occupa di ricerca nell'ambito delle tecnologie informatiche avanzate applicate ai sistemi informativi aziendali, in particolare nel settore delle applicazioni Internet e dell'e-learning. Dal 1988 si occupa di multimedialità e ipermedia, con varie collaborazioni a livello nazionale ed internazionale. Relatore in varie conferenze e tavole rotonde sul tema, ha approfondito maggiormente gli aspetti legati all'utilizzo delle nuove tecnologie nell'ambito dell'e-learning. Dal 2002 si occupa di sperimentazioni nell'ambito del mobile learning, in particolare occupandosi di far evolvere un tradizionale Learning Management System verso un sistema per la gestione di comunità virtuali di apprendimento.
E-mail amolinar@cs.unitn.it

LA COMUNITÀ VIRTUALE DEGLI ESAMINATORI ECDL



I processi di apprendimento, indispensabili per lo sviluppo di nuove competenze, hanno forti legami con i contesti dove tali competenze vengono spese. Le tecnologie mettono a disposizione diverse opportunità per condividere questi processi e per costruire ambienti dove vengano valorizzate le esperienze condotte dai singoli e siano messe a sistema per massimizzarne l'impatto sull'intera organizzazione: le comunità virtuali professionali. A titolo di esempio, viene descritta l'esperienza del progetto ECDL.

**Daniela Maria Ricco
Giuseppe Giliberto
Giuliano Russo**

1. GLI ARTEFATTI TECNOLOGICI

In una prospettiva socio-culturale, le tecnologie si presentano come artefatti culturali, che mediano le nostre interazioni cognitive e sociali con il mondo che ci circonda.

In questa prospettiva non si parla più tanto di interazione uomo-computer, quanto dei nuovi contesti creati dalle tecnologie dell'informazione e della comunicazione, per permettere alle persone di cooperare e di comunicare fra loro [8].

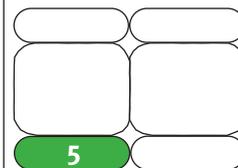
La presenza di tali artefatti ed il loro uso nella vita quotidiana cambiano in modo sostanziale il nostro modo di lavorare. Infatti l'uso di strumenti e le modalità di esecuzione dei compiti si influenzano a vicenda. Carroll e Campbell [3] hanno messo in evidenza il carattere circolare della relazione artefatti-compiti (*task-artifact cycle*): le persone svolgono determinati compiti, con maggiore o minore soddisfazione, usando determinati strumenti. I nuovi strumenti, una volta adottati, alterano i compiti per cui vennero progettati e modificano le situazioni in cui i compiti venivano svolti in precedenza. In questo modo lo strumento nuovo

crea un nuovo compito e una nuova situazione sociale in cui si svolge il compito. Questi ultimi generano poi a loro volta il bisogno di ulteriori miglioramenti che sarebbero possibili grazie ad una nuova famiglia di artefatti da inventare per svolgere meglio il nuovo compito e così via. Questo processo circolare innesca l'incessante innovazione tecnologica caratteristica della moderna cultura occidentale.

Lo strumento modifica non solo la pratica, ma anche le abilità, le competenze professionali, il modo di pensare che accompagnava l'operazione precedente. Esso altera la struttura so-

Necessariamente infatti queste cose si apprendono insieme, e insieme la menzogna e la verità dell'intera sostanza, con molta applicazione e molto tempo, come ho detto in principio; quando questi singoli elementi (nomi, definizioni, vista ed altre sensazioni) vengono con sforzo sfregati gli uni contro gli altri e sottoposti a confutazione in dispute benevole e in scambi di domande e risposte fatti senza animosità, allora, a chi compie ogni sforzo consentito alle possibilità umane, riluce d'un tratto intorno a ciascun problema comprensione e intuizione.

Platone, Lettera VII, 344



In una comunità virtuale – a differenza di una comunità reale - si realizza, grazie al supporto della tecnologia, la possibilità di mantenere un corpo documentale di riferimento, un sistema per la gestione della memoria di gruppo, un *repository* della conoscenza, uno spazio di condivisione di esperienze, storie, strumenti, soluzioni, procedure ecc..

3. COME INDURRE LO SVILUPPO DI UNA COMUNITÀ VIRTUALE PROFESSIONALE

Una delle finalità delle comunità virtuali professionali è quella di contribuire a trasformare la conoscenza dal livello individuale a quello organizzativo e ad agevolare, mediante modalità particolari, l'emergere della conoscenza esplicita quale patrimonio a cui tutta l'organizzazione può attingere.

Organizzare comunità virtuali affinché queste possano effettivamente giocare un ruolo nella nuova economia della conoscenza significa gestire processi comunicativi che sappiano equilibrare esigenze di partecipazione e qualità dell'informazione offerta [9].

La strutturazione di un circuito comunitario di accumulazione del sapere costituisce un'operazione complessa. L'animazione di forme di partecipazione periferica, la validazione delle informazioni, la loro organizzazione e diffusione in supporti coerenti sono parti di un processo che non può essere improvvisato. La "coltivazione" di comunità virtuali richiede quindi la definizione di metodologie e strumenti rivolti a sostenere le potenzialità e le prestazioni delle dinamiche di apprendimento che caratterizzano la comunità. In particolare risulta di cruciale importanza la definizione dei seguenti aspetti:

□ il "confezionamento" della conoscenza costitutiva della libreria virtuale della comunità. Questa attività richiede uno sforzo molto maggiore rispetto alla semplice traduzione nella forma digitale di contenuti ospitati in media tradizionali, presupponendo prima di tutto la conoscenza delle modalità con cui i partecipanti alla comunità coprono i propri deficit cognitivi incrociando le due variabili chiave: il tempo (urgenza) e la profondità (approfondimento).

□ Il processo di validazione della conoscenza generata all'interno.

La varietà cognitiva creata dal sistema di dialoghi a geometria variabile caratterizzante la comunità deve assoggettarsi ad un processo di selezione, rivolto a individuare soltanto la conoscenza meritevole di diventare patrimonio comune. Tale processo di validazione si deve snodare lungo una serie di passi attraverso i quali la validità dell'innovazione verrà socializzata tra i componenti della comunità e capitalizzata da una figura chiave: il *Knowledge Manager*.

□ I ruoli chiave all'interno della stessa che sovrintendono e gestiscono la comunità nel suo complesso.

La necessità di garantire una dinamica continua di apprendimento e di crescita impone l'identificazione in seno alla comunità virtuale di altri ruoli, oltre a quello già delineato del *knowledge manager*. Nel loro insieme, e in senso lato, tutte queste figure predefinite hanno il compito di integrare a sistema la parte statica della comunità (il *repository* della conoscenza) con la parte dinamica (dimensione dialogica) per garantirne il successo: sostenere il processo di apprendimento della comunità.

Al *Community Leader* spetta un importante compito di legittimazione delle nuove tecnologie come veicolo dei dialoghi e di accesso alle informazioni all'interno della comunità; il *Facilitatore* agirà come propulsore dell'attività dialogica tra i componenti della comunità, cercando di sollecitarli e coinvolgerli nella discussione e garantendo ai nuovi utenti un accesso facile e amichevole; al *System Administrator* spetta il compito di gestire tutta la parte tecnologica del sistema e i *Guest Speaker* rappresentano importanti occasioni di crescita e di aggiornamento per tutta la comunità.

Quali misure organizzative, metodologiche, tecnologiche possono contribuire al successo di una comunità virtuale?

Prendendo spunto da quelle che sono le caratteristiche di una comunità di pratica [15] possiamo definire alcuni elementi utili nell'approccio metodologico utilizzato per la progettazione di una comunità d'apprendimento. In questo caso, le esigenze possono essere organizzate in alcune categorie:

□ la comunicazione, rappresentata nella rete di impegni generati dagli atti linguistici che i

instaurano all'interno della comunità virtuale. Negli ultimi anni sono stati elaborati modelli valutativi di impianto olistico, che intendono avventurarsi oltre il dominio di una valutazione puramente comportamentista e quantitativa. Ecco in sintesi due sperimentazioni significative.

Presso l'Istituto Tecnologie Didattiche del Consiglio Nazionale delle Ricerche di Genova è stata sviluppata una metodologia di valutazione della qualità dei processi d'apprendimento collaborativo in rete che è stata sperimentata nel corso di formazione per docenti d'Educazione Ambientale EuMEDEA [9], condotto interamente a distanza.

Il processo valutativo si è basato sull'esame di tutti i messaggi prodotti nel processo di formazione in rete, all'interno di moduli didattici articolati in cinque fasi secondo la scansione di un'attività di *cooperative learning*.

I parametri della qualità che hanno guidato l'analisi valutativa sono stati:

- la congruenza contestuale, ossia la collocazione in un'area adeguata;
- la congruenza formale, ossia formattazione, lunghezza;
- l'appropriatezza dei codici, ossia di tutti gli elementi che permettono di cogliere l'aspetto metacomunicativo del messaggio (come il colore, il carattere, i puntini di sospensione);
- la costruzione di nuova conoscenza, rispetto agli obiettivi del contesto;
- il contributo all'interazione, rispetto alla funzione di stimolo e animazione del messaggio;
- la correttezza dei contenuti, rispetto alle consegne richieste;
- la tempestività delle risposte e delle consegne.

Il valore della qualità globale del corso in rete EuMEDEA è stato ricavato dalla relazione tra i valori della qualità dell'interazione, dei contenuti e dell'apprendimento.

Un'ulteriore sperimentazione significativa dal punto di vista metodologico è quella che è stata condotta a partire dalle discussioni di una comunità denominata "CKBG" (*Cooperative Knowledge Building Group*), utilizzando un web forum ideato da un gruppo di ricerca canadese allo scopo di avviare una riflessione sulle modalità di comunicazione e di costruzione della conoscenza all'interno del forum stesso [10].

Sono state predisposte a tal scopo due tipologie di schede d'analisi:

- una scheda della discussione (definita "sequenze") per raccogliere informazioni sul numero di partecipanti, dei messaggi (definiti "note"), sull'autore del messaggio che ha originato la discussione stessa;
- una scheda del messaggio, che viene descritto rispetto al suo contenuto e alle sue caratteristiche linguistiche.

Tale analisi ha condotto ad individuare due principali funzioni dei messaggi: una funzione cognitiva e una relazionale.

La funzione cognitiva è stata valutata attraverso alcuni indicatori di contenuto, quali la presenza di nuovi elementi e la problematizzazione di quelli proposti. Si tratta di un indicatore significativo del raggiungimento di uno degli scopi principali di una comunità virtuale di apprendimento.

La funzione relazionale è stata riferita ad indicatori significativi nella direzione della costruzione di un'identità collettiva, ossia di condivisione degli stessi significati e conoscenze.

6. INTRODUZIONE AL CASO DI STUDIO

6.1. Il progetto ECDL

AICA opera dal 1997 come ente responsabile per l'Italia dell'ECDL (*European Computer Driving Licence*), comunemente nota come Patente Europea del Computer, che certifica - mediante prove di esame pratiche - le competenze informatiche di base. Attualmente vengono effettuati circa un milione di esami all'anno.

Pur mantenendo la responsabilità e il controllo della certificazione mediante appositi *audit*, AICA delega attività importanti del processo di certificazione a sedi di esami esterne, dette Test Center, attualmente più di 2.700 sul territorio italiano. Per omogeneizzare la valutazione degli esami e ottimizzarne i tempi, facilitando così i Test Center, AICA ha sviluppato un sistema di Qualità definito attraverso alcune procedure secondo il modello ISO9000 ed un apposito strumento informatico per la gestione degli esami denominato ATES (*Automatic Testing and Evaluation System*).

Ogni Test Center deve disporre di almeno due Esaminatori che operano alle sue dirette dipendenze e sono dunque persone esterne ri-

petto ad AICA, con la quale non hanno un rapporto contrattuale diretto, benché svolgano un ruolo assai rilevante nella gestione operativa delle sessioni di esami per conto e secondo le regole di AICA.

Nel processo di erogazione degli esami AICA delega infatti agli Esaminatori attività di particolare importanza ai fini dell'imparzialità, oggettività e trasparenza della certificazione; ad esempio il riconoscimento dei candidati, la sorveglianza della sessione, la registrazione degli esami, la gestione di eventuali reclami sull'esito degli stessi. Pur non rientrando in senso stretto nell'organizzazione di AICA, gli Esaminatori possono dunque essere considerati parte integrante del suo sistema di certificazione.

La formazione, la motivazione, il coinvolgimento degli Esaminatori e l'efficacia della comunicazione sono stati individuati da AICA quali fattori chiave per il miglioramento della Qualità e pertanto si è ritenuto strategico curarne la professionalità e istituire con essi un canale di comunicazione diretto attraverso un ambiente Web integrato di servizi di *e-learning*, comunità virtuali e archivi condivisi di documenti.

La scelta di tecnologie Web-based per la formazione e la comunicazione è stata motivata dall'alto numero di destinatari (più di 5.000 potenziali partecipanti), dalla loro distribuzione sul territorio nazionale e dalla loro accertata competenza nell'utilizzo degli strumenti di ICT. Si sono riconosciuti nella categoria degli Esaminatori alcuni tratti distintivi di una potenziale comunità di pratiche, su cui si è fatto leva per la costituzione di una comunità virtuale professionale. Per questo la scelta della tecnologia si è orientata ad un ambiente di *virtual community*, comprendente ambienti di discussione, strumenti di comunicazione e archivi condivisi.

Gli Esaminatori accedono all'ambiente di community previa la fruizione di un corso on line, che ha per oggetto le procedure AICA di qualità rilevanti per gli esami ECDL e le funzionalità del sistema ATES, il software di erogazione e valutazione automatica degli esami.

Una volta acquisite, tramite il corso, le competenze richieste per esercitare il proprio ruolo, gli Esaminatori diventano a tutti gli effetti membri della comunità virtuale professionale, sviluppando il dialogo (condividendo codici in-

terpretativi e conoscenze) con i colleghi e ponendo quesiti ad AICA, che garantisce un servizio di risposte dedicato. Tale comunità, attualmente costituita da di più di 4.000 iscritti, è una delle più ampie comunità professionali virtuali esistenti in Italia.

Il progetto è iniziato nell'autunno del 2002. Il primo corso on line è stato erogato agli Esaminatori nel dicembre del 2003.

A due anni dall'avvio del progetto si propone, in questo intervento, una riflessione sui risultati ottenuti finora relativamente alla comunità Esaminatori ECDL Core, al fine di definire strategie di sviluppo finalizzate a raggiungere traguardi via via più ambiziosi, individuando ulteriori figure professionali in base ai nuovi sviluppi dei programmi ECDL, EUCIP e del Sistema di Qualità, nonché sviluppando nuovi servizi basati su contributi anche provenienti dalle comunità virtuali stesse.

L'analisi effettuata concernente alcune caratteristiche dei messaggi presenti nel forum può offrire una panoramica delle dinamiche comunitarie, attraverso la quale approfondire le dimensioni cognitivo-relazionali della comunità. Per valutare l'efficacia dell'attività nel suo complesso si dovranno inoltre raccogliere e interpretare non solo informazioni provenienti dall'interno della comunità ma anche dall'andamento delle attività di certificazione e degli *audit* di qualità.

I risultati di questo studio rappresentano un orientamento per sviluppare a regime un'attività valutativa permanente in itinere.

6.2. La tecnologia

La piattaforma atta a soddisfare le esigenze sopra espresse è stata individuata nel prodotto Ghol3, resa disponibile dalla società Skill On Line su Internet in modalità ASP, ed è stata personalizzata per mettere a disposizione:

- un ambiente di *e-learning* utilizzato per la pubblicazione dei contenuti e per l'attivazione dei corsi con tutoring remoto;
- un ambiente di *virtual community* orientato al dialogo tra i partecipanti con funzionalità di gestione dei messaggi e dei profili degli utenti;
- un ambiente di *virtual repository*, dove i partecipanti possono organizzare archivi di documenti condivisi;
- una *bacheca* per la gestione della messaggistica della comunità virtuale;

□ un ambiente di *reporting* per il monitoraggio di tutte le attività realizzate negli ambienti descritti precedentemente.

Per supportare le attività on line AICA ha istituito una redazione - composta da progettisti e esperti nei diversi temi affrontati, oltre che da personale di segreteria - che svolge attività di progettazione e realizzazione di percorsi didattici, tutorship, animazione, elaborazione di risposte puntuali alle sollecitazioni degli esaminatori e produzione di elaborati e strumenti di lavoro ad uso dei partecipanti.

7. I MESSAGGI E I CRITERI DI ANALISI DELLA COMUNITÀ

I dati relativi ai messaggi del forum, che è stato replicato integralmente in un database per consentire anche in momenti successivi ulteriori operazioni di elaborazione, si basano su parametri in parte ricavati direttamente dalla piattaforma (argomento, data-ora di scrittura, nominativo del mittente, nominativi dei lettori) e in parte dall'esame del contenuto dei messaggi (contatori, classificazione dei partecipanti e delle motivazioni, orientamento della comunicazione).

Occorre rilevare che non è stato finora possibile registrare l'andamento temporale relativo alla lettura dei messaggi all'interno della comunità, poiché l'unico dato rilevabile in ordine alle letture di ciascun messaggio è costituito dai nominativi dei lettori. I dati sono stati rilevati una prima volta il 9 maggio 2005, poco più di un anno dopo l'apertura del forum e, successivamente, il 21 febbraio 2006. È stato così possibile effettuare un confronto tra le rilevazioni relative alle due finestre temporali.

8. I PARTECIPANTI DELLA COMUNITÀ

Alla comunità partecipano, oltre agli Esaminatori che hanno completato i corsi on line, gli Ispettori ECDL (Auditor della Qualità) che ne hanno fatto richiesta e i membri della redazione il cui incarico è di rispondere ai quesiti posti dagli esaminatori, moderare e stimolare il dibattito, predisporre elenchi delle domande più frequenti (FAQ). Nella tabella 1

sono indicati come "Attivi" quanti abbiano almeno letto un messaggio.

9. GLI INTERVENTI NEL FORUM

I messaggi registrati nel periodo d'osservazione sono stati globalmente 1.285, suddivisi per autori come indicato nella tabella 2.

Alcune considerazioni risultano preliminari alla comprensione dell'analisi dei dati raccolti:

□ per quanto riguarda gli Esaminatori, non disponendo al momento di informazioni sulla data del loro ingresso nella comunità, sono precluse per ora alcune ulteriori analisi, come ad esempio l'andamento della loro partecipazione (fatta eccezione per il raffronto tra le due rilevazioni sopra citate);

□ la presenza degli Ispettori non è stata rilevante sotto il profilo quantitativo (ha aderito all'iniziativa meno del 50% del corpo ispettivo) e decisamente scarsa la loro partecipazione ai dialoghi della comunità forse a causa di una possibile difficoltà a coniugare gli aspetti di controllo, relativi al loro ruolo professionale, con quelli di mediazione e facilitazione propri della dinamica comunitaria ed inoltre l'ambiente di community non prevedeva alcun servizio orientato a tale categoria;

□ la presenza dei membri della redazione (gli esperti incaricati di fornire risposte ai vari argomenti, il responsabile della Funzione Qualità di AICA e gli amministratori della piattaforma della Comunità Virtuale) è strettamente correlata al numero di quesiti posti dagli Esaminatori, poiché è in tutti i casi prevista una loro risposta, anche quando il chiarimento sia stato già fornito da un altro membro della comunità.

Categoria	Iscritti	Attivi
Esaminatori	4.193	2.013
Ispettori	39	20
Redazione	15	15

TABELLA 1

Riepilogo popolazione costituente la comunità

Categoria	Mittenti	Messaggi
Esaminatore	230	845
Redazione	9	438
Ispettore	2	2

TABELLA 2

Messaggi registrati all'interno della comunità

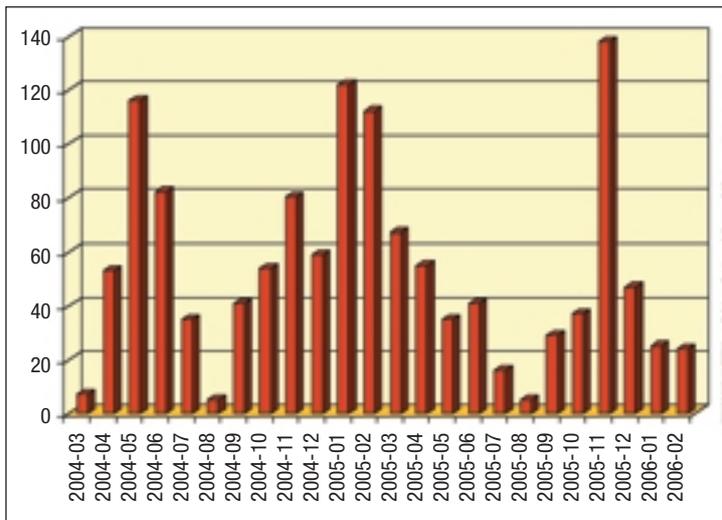


FIGURA 1

Rappresentazione dei messaggi inviati

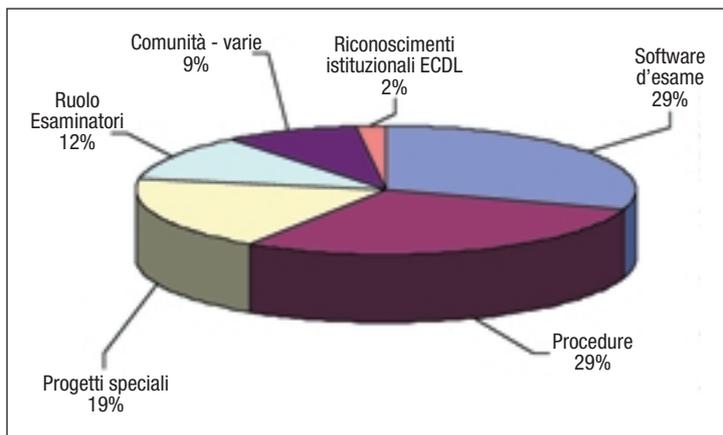


FIGURA 2

Distribuzione per argomento dei messaggi inviati

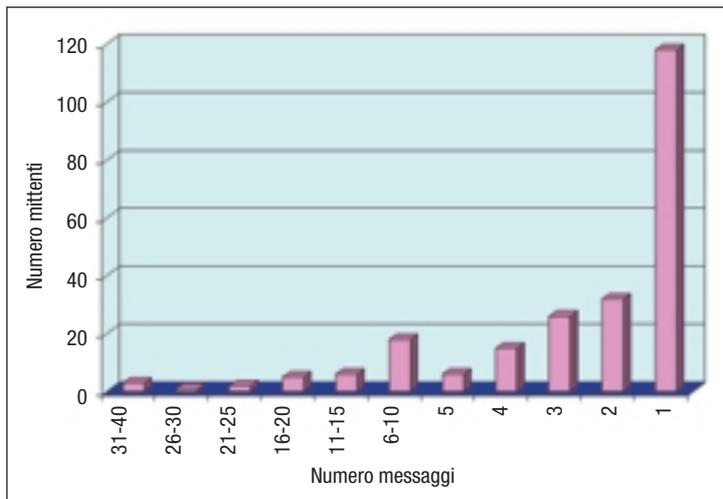


FIGURA 3

Rappresentazione del numero messaggi inviati dagli utenti per numero di utenti

9.1. Distribuzione temporale

L'andamento temporale dell'invio dei messaggi individua tre significativi picchi (Figura 1): il primo in corrispondenza dell'apertura della comunità (maggio - giugno 2004); il secondo in occasione dell'entrata in vigore delle modifiche alle procedure e al software d'esame nei primi due mesi del 2005; il terzo nell'autunno 2005, in concomitanza con l'annuncio dell'anticipo della transizione al nuovo software d'esame ATLAS.

La presenza di picchi nella produzione dei messaggi si correla con la necessità di ottenere chiarimenti relativi alle novità introdotte nel lavoro degli Esaminatori. Tale motivazione scema nell'arco di un paio di mesi: il fenomeno è attribuibile alla possibilità di leggere risposte già fornite o di consultare le FAQ predisposte dalla redazione.

9.2. Distribuzione per argomento

Gli argomenti sui quali si è concentrata la maggior parte degli interventi riguardano:

- il software d'esame;
- le procedure d'esame;
- i progetti speciali;
- il ruolo degli esaminatori.

La ripartizione dei messaggi per argomento (Figura 2) evidenzia una preponderanza dei principali temi connessi con l'operatività delle sessioni d'esame:

- il software, che consente di registrare i dati anagrafici dei candidati, di predisporre il calendario delle sessioni, di effettuare i test e di registrare i risultati;
- le norme procedurali che definiscono i requisiti dei test center e le modalità di organizzazione e tenuta delle sessioni d'esame.

10. PROFILO DEGLI ESAMINATORI

Analizzando la quantità di messaggi inviata da ciascuno degli Esaminatori, si evidenzia che un ristretto numero di loro costituisce il vero e proprio nucleo trainante, con un elevato numero di messaggi e notevole propensione ad intervenire nel dibattito, esprimendo opinioni ed interloquendo con i colleghi. La maggior parte di coloro che intervengono si limita ad un solo messaggio, tipicamente per ottenere informazioni (Figura 3).

10.1. Orientamento della comunicazione

Un aspetto significativo per valutare il senso d'identità collettiva può essere ricavato dall'orientamento della comunicazione da parte degli Esaminatori. Sono stati a tal fine conteggiati i messaggi degli Esaminatori rivolti ai membri dello stesso gruppo di riferimento ponendoli in relazione con i messaggi totali, che includono anche quelli rivolti agli esperti di AICA (Figura 4). Si è cercato inoltre di approfondire l'analisi, sia disaggregando lo stesso fenomeno per argomento, sia valutando il suo andamento nel tempo.

Su 838 messaggi inviati dagli esaminatori, 431 (51%) sono stati rivolti alla comunità e 407 agli esperti di AICA; la deviazione standard è del 14%.

La percentuale di messaggi rivolti alla comunità si mantiene intorno alla media nei momenti di massima produzione di messaggi: maggio-giugno 2004 (entrata in funzione della piattaforma), gennaio-febbraio 2005 (cambiamenti nelle procedure), novembre 2005 (introduzione del nuovo software d'esame). Nell'arco di un trimestre, le oscillazioni vanno da un minimo del 46% ad un massimo del 62%.

L'analisi per argomento (Figura 5) ha evidenziato valori intorno al 35% per quanto riguarda due argomenti che attengono più specificamente ad aspetti normativi (Procedure e Ruolo degli Esaminatori), per i quali la comunicazione è essenzialmente rivolta agli esperti AICA, mentre sugli altri temi è stata maggiore la propensione a comunicare con i colleghi.

10.2. Funzione relazionale dei messaggi

L'analisi degli aspetti relazionali dei messaggi è stata estesa con l'analisi delle "deissi", ossia dei riferimenti impliciti di un testo a situazioni spaziali, temporali e personali.

Per quanto riguarda le deissi personali sono stati analizzati separatamente:

- i riferimenti impliciti ad un interlocutore singolo;
- i riferimenti impliciti alla comunità;
- i riferimenti alla propria identità.

L'analisi è stata realizzata integrando due metodologie:

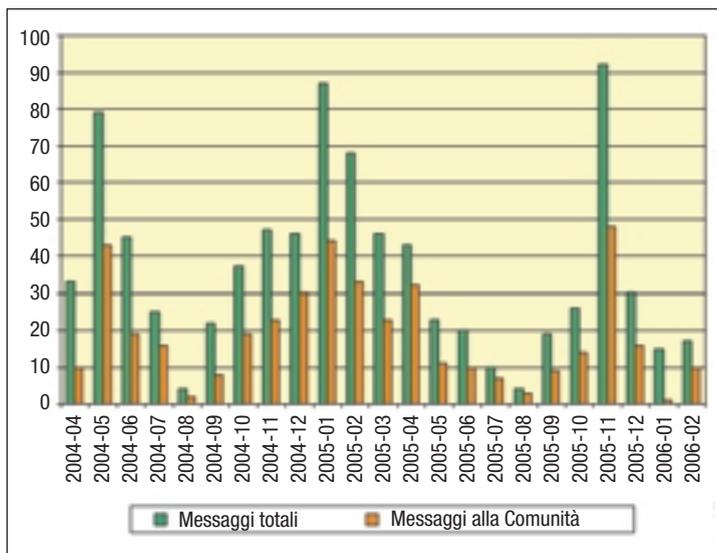


FIGURA 4

Riproduzione dei messaggi inviati: dati globali

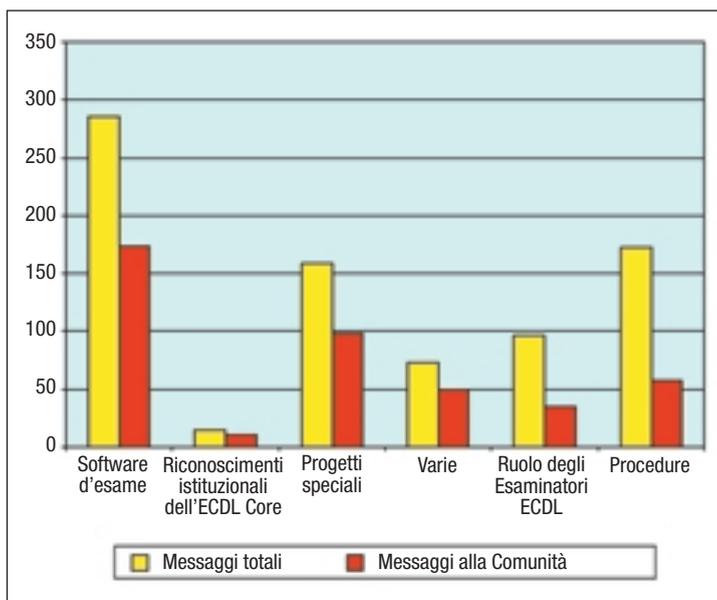


FIGURA 5

Riepilogo dei messaggi degli esaminatori suddivisi per argomento

□ *Analisi automatica*, ottenuta con una serie di interrogazioni sul testo del messaggio, tramite il predicato "like" che ha consentito di individuare la maggior parte dei messaggi contenenti deissi. La validità degli algoritmi è stata in precedenza verificata su un campione di 44 messaggi.

□ *Esame del testo*, effettuato su tutti i messaggi che con l'analisi automatica risultavano privi di deissi, specialmente per l'uso di predicati verbali (per i quali l'analisi automatica risultava di ardua applicabilità).

Globalmente, i messaggi contenenti almeno

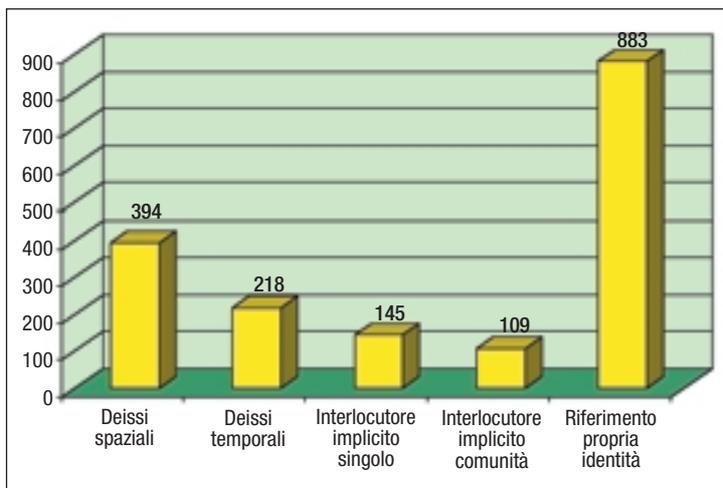


FIGURA 6

Rappresentazione dei messaggi contenenti deissi

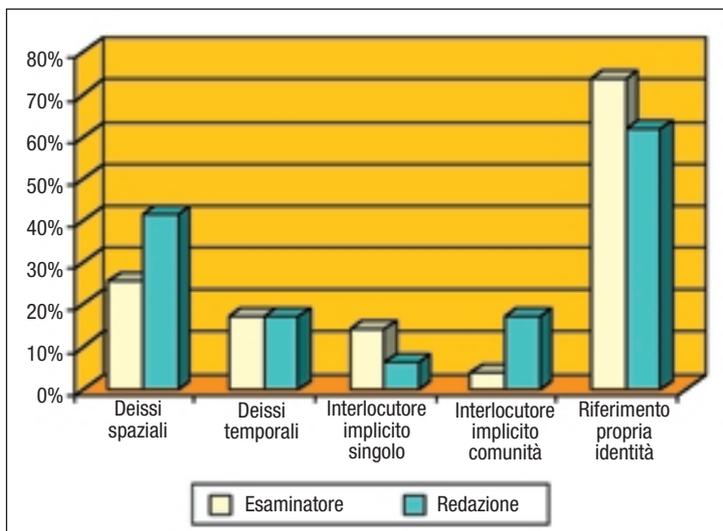


FIGURA 7

Analisi della percentuale di messaggi contenenti deissi

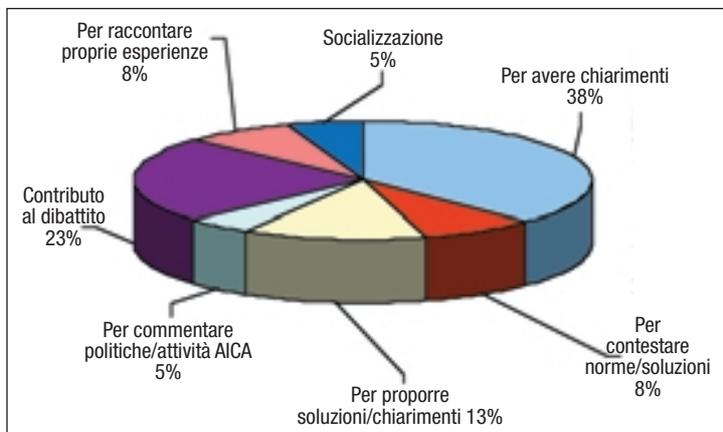


FIGURA 8

Rappresentazione dei motivi di intervento nel forum degli esaminatori

uno dei tipi di deissi considerate è di 1.157 su 1.283 messaggi esaminati (90%).

I messaggi sono sintetizzati nei grafici delle figure 6 e 7. Come si vede, gli esaminatori sono più propensi dei membri della redazione a far riferimento alla propria identità e a rivolgersi implicitamente ad un singolo interlocutore, mentre fanno meno uso di deissi spaziali e riferimenti impliciti alla comunità.

10.3. Analisi delle motivazioni

È stata definita una griglia per rilevare le motivazioni degli interventi degli Esaminatori, che sono così stati classificati analizzando il contenuto del messaggio. È stato possibile individuare alcune caratteristiche del profilo dell'Esaminatore per quanto riguarda le sue esigenze informative e le sue conoscenze (Figura 8).

La comunità è percepita come uno strumento per avere chiarimenti convalidando l'esigenza iniziale per la quale in fase di progettazione si è pensato allo sviluppo di una community in rete.

Emerge inoltre una propensione alla socializzazione delle proprie esperienze.

10.4. Letture nel Forum

Nel periodo d'osservazione sono state effettuate 138.508 letture da parte di 2.034 lettori, con una media di 68 messaggi letti da ciascuno. Una prima separazione dei dati per categoria di lettori è riportata nella tabella 3.

Un'analisi più dettagliata dei lettori per numero di messaggi letti contenuta nel grafico di figura 9, mostra che la maggior parte degli esaminatori legge pochi messaggi, limitandosi in genere alle risposte ai propri interventi (i dati non comprendono le letture

Tipo lettore	n. lettori	Letture effettuate	Media messaggi letti
Esaminatore	2.000	125.302	63
Ispettore	20	4.593	230
Redazione	14	8.613	615

TABELLA 3

Letture effettuate per tipologia di lettore

effettuate dagli esperti della redazione, i quali sono ovviamente tenuti a leggere tutti i messaggi).

11. RAFFRONTI CON I DATI DEL 2005

Alcuni raffronti possono essere fatti con le analoghe analisi effettuate nel maggio 2005, relativamente agli esaminatori (Tabella 4). Dal raffronto tra la percentuale d'aumento degli iscritti e quelle degli altri parametri emerge un significativo incremento dell'utilizzo del forum, non solo per quanto riguarda la presenza (lettori e mittenti), ma specialmente per il livello di partecipazione (letture e messaggi inviati), che è cresciuto ad un tasso più che doppio.

Significativo è, inoltre, il raffronto tra i dati relativi alle motivazioni degli interventi, esposti nella tabella 5.

Le maggiori percentuali d'aumento relative alla contestazione di norme/soluzioni e al commento sulle politiche di AICA è probabilmente da ricondurre alla transizione al nuovo sistema automatico d'esame ATLAS. Occorre qui ricordare che per tale transizione, nell'autunno 2005, i tempi comunicati risultavano particolarmente critici ed hanno quindi generato una preoccupazione diffusa.

In questa circostanza d'emergenza, come dimostrano i dati quantitativi e l'analisi delle motivazioni, la comunità virtuale professionale, oltre a rappresentare un'immediata cassa di risonanza di dubbi e malumori, ha svolto un prezioso compito di trasmissione delle informazioni in entrambi i sensi, integrando gli altri canali comunicativi predisposti per la circostanza, anche grazie alla partecipazione diretta dei responsabili di AICA per lo specifico settore.

Il dato relativo alla socializzazione, che pur essendo di segno positivo è inferiore agli altri incrementi, è da considerarsi naturale rispetto ad un buon sviluppo della comunità che non richiede più un'azione esplicita per legittimare l'appartenenza alla comunità stessa.

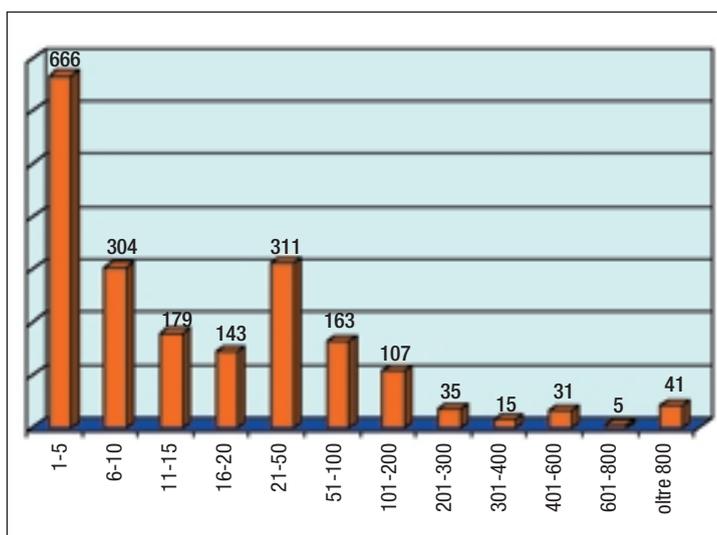


FIGURA 9

Distribuzione esaminatori per messaggi letti

	Maggio 2005	Febbraio 2006	Variazione %
Iscritti	3.480	4.193	+17
Attivi	1.599	2.013	+21
Mittenti	173	230	+33
Messaggi inviati	581	845	+45
Lettori	1.599	2.000	+25
Letture	82.703	125.302	+52
Media messaggi letti	52	63	+21

TABELLA 4

Confronto dati quantitativi maggio 2005 – febbraio 2006

	Avere chiarimenti	Contestare norme e soluzioni	Proporre soluzioni e chiarimenti	Commenti a politiche e attività AICA	Contributi al dibattito	Raccontare proprie esperienze	Socializzazione	Totali
Maggio 2005	223	37	79	25	131	49	37	581
Febbraio 2006	315	67	109	40	195	69	44	839
Variazione %	+41	+81	+38	+60	+49	+41	+19	+44

TABELLA 5

Confronto dati inerenti le motivazioni (maggio 2005 – febbraio 2006)

12. IL GRADIMENTO

Per acquisire il punto di vista degli Esaminatori e sviluppare futuri servizi condivisi da tutta la comunità, nel 2005 è stato somministrato un questionario on line di gradimento, riferito agli attuali servizi e a proposte per nuovi servizi del forum ECDL. Hanno risposto circa 1.200 esaminatori cioè il 40 % degli iscritti. L'attuale servizio "L'esperto risponde", rilevante all'interno della comunità, ha ricevuto un ottimo riscontro da parte del nostro campione d'Esaminatori (Figura 10).

Si è tra l'altro proposta la pubblicazione di un periodico on line che rappresenti l'organo ufficiale di comunicazione della comunità, il cui titolo (*E-saminando*) è stato preferito dalla maggior parte del campione.

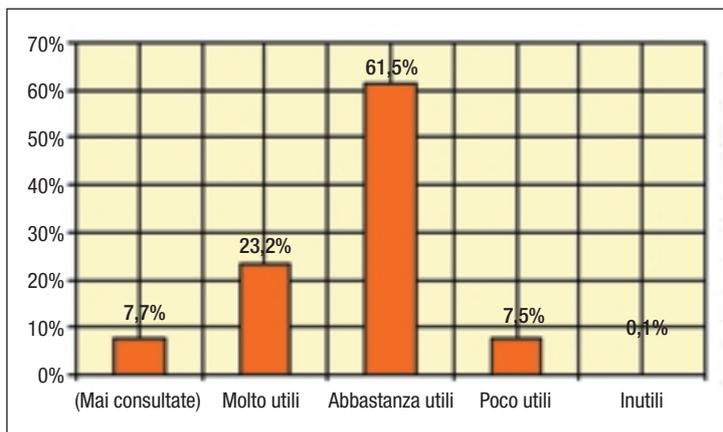


FIGURA 10

Rappresentazione gradimento del servizio "L'esperto risponde"

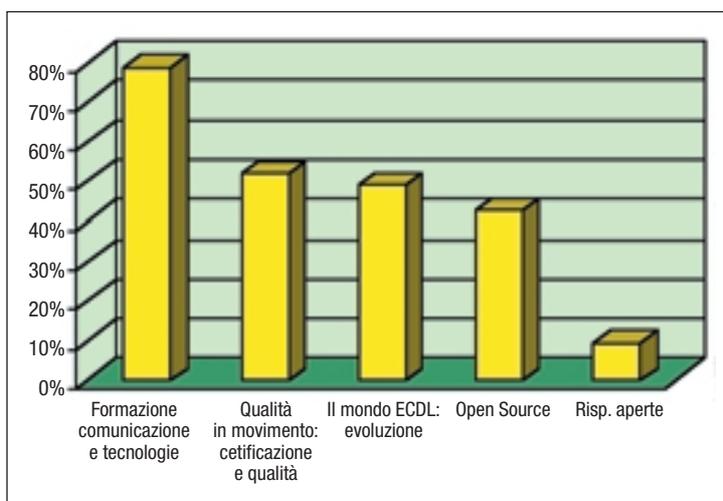


FIGURA 11

Risposte relative agli argomenti da inserire nelle rubriche di interesse (scelta multipla)

Alla domanda relativa alla scelta degli argomenti ritenuti più utili ed interessanti dagli Esaminatori, sono emerse le indicazioni che hanno costituito gli argomenti centrali del notiziario (Figura 11).

Dai risultati del questionario si registra inoltre che il 96% degli esaminatori consiglierebbe ad altri colleghi la partecipazione alla comunità, in quanto considerata utile per lo svolgimento delle attività.

13. CONCLUSIONI

La categoria degli Esaminatori ha saputo cogliere la possibilità di riconoscersi in una comunità di pratiche che condivide competenze, obiettivi/problematiche, linguaggi grazie all'uso degli strumenti tecnologici che hanno permesso di superare i limiti derivati dalla non contiguità fisica, elemento ostativo alla condivisione sistematica delle proprie esperienze.

L'appartenenza a una comunità virtuale professionale ha consentito agli Esaminatori di rafforzare la propria identità di ruolo e nel contempo di sentirsi attori di un processo che va al di là dell'inserimento in una organizzazione, ma che si concretizza in un'attività funzionale all'obiettivo comune.

I risultati della comunità possono essere evidenziati in due direzioni: la prima riguarda la costruzione di nuove pratiche e la loro diffusione da parte dei membri, che sono riusciti, attraverso un confronto sulle attività quotidiane, a sviluppare modalità operative più efficaci a beneficio dell'intero sistema; la seconda è relativa all'attivazione di processi di miglioramento degli strumenti (procedure, sistema software) attuati dai responsabili del progetto ECDL come sintesi dei contributi emersi dall'attività degli Esaminatori.

La soluzione di *e-learning* sviluppata da AICA sembra dimostrarsi efficace almeno per quella parte significativa degli Esaminatori che partecipano alle attività on line, mentre la sensibile differenza tra il numero di messaggi pubblicati ed il numero delle loro letture conferma la presenza di un gruppo cospicuo di partecipanti "silenti", che seguono l'attività senza parteciparvi direttamente.

Dai dati raccolti si rileva che la diffusione della conoscenza degli Esaminatori è in particolare focalizzata sui temi delle procedure e del software d'esame, connessa con la qualità dei servizi di certificazione offerti. Ulteriore conferma viene espressa dal Responsabile della Qualità di AICA, il quale ha rilevato una maggiore diffusione della conoscenza sulle procedure da adottare; questo elemento risulta significativo in quanto indica il raggiungimento dell'obiettivo primario del progetto, confermando l'adeguatezza del modello utilizzato.

La dimostrazione dell'utilità dell'impianto metodologico è inoltre rilevabile dalla tipologia di documenti presenti nell'archivio condiviso, in particolare le procedure d'esame e le FAQ. Durante il periodo in cui si è svolta l'attività della comunità, infatti, si è verificata la necessità di apportare miglioramenti alle procedure d'esame che sono stati negoziati all'interno di gruppi specifici di esaminatori e che rappresentano un risultato condiviso. Questo risultato è stato determinato dalla valorizzazione delle esperienze maturate dai singoli impegnati sul campo e da un processo esplicito di selezione e confezionamento della nuova conoscenza generata messo in atto dal gruppo di progetto, ottenendo un diffuso consenso alle novità introdotte e facilitandone l'applicazione.

Il raffronto tra i dati del 2005 e 2006 esplicita un trend positivo di crescita e consolidamento della comunità, soprattutto effettuando un'analisi di tipo qualitativo in quanto crescono i messaggi che contestano norme e soluzioni, cresce significativamente il commento alle politiche adottate da AICA e i contributi al dibattito.

Il progetto contempla, da un lato, l'organizzazione di incontri in presenza che permettono il rafforzamento dell'identità collettiva e, dall'altro, uno sviluppo di dinamiche di scambio che consentono alla comunità di evitare il rischio di auto-referenzialità. In questa direzione se ne sono già svolti due in occasione di convegni o eventi istituzionali di AICA.

Tra gli obiettivi condivisi delle prossime fasi del progetto si prevede l'implementazione di un sistema di formazione continua attraverso il puntuale aggiornamento del mate-

riale didattico sulle procedure d'esame e sul software. Tale sistema sarà strutturato in differenti percorsi formativi rivolti sia ai nuovi esaminatori e ispettori, sia a coloro che da anni operano nel circuito, che risponda alle esigenze di aggiornamento legate agli sviluppi dei sistemi di certificazione di AICA.

Ringraziamenti

La comunità virtuale qui presentata è stata realizzata nel quadro delle attività a sostegno della qualità del programma ECDL coordinate da Paolo Fezzi, responsabile Qualità di AICA.

Bibliografia

- [1] Benigno V., Trentin G.: *La valutazione secondo il modello Polaris*. In Trentin G. (a cura di) *Telematica e formazione a distanza, il caso Polaris*, Franco Angeli, Milano, 1999.
- [2] Bocconi S., Minorò V., Sarti L.: *Valutazione della qualità nella formazione in rete*. TD, *Tecnologie Didattiche*, n. 16, 1999.
- [3] Carroll J.M., Campbell R.L.: *Artifacts a psychological theories: The case of human-computer interaction*. In "Behaviour and Information Technology", 1989.
- [4] Fulk J., Steinfield C. (a cura di): *Organizations and communications technology*. Newbury Park, CA, Sage, 1990.
- [5] Goodman P.S. Sproull, L.S., (a cura di): *Technology and organizations*. San Francisco, CA, Jossey-Bass, 1990.
- [6] Lave J., Wenger E.: *Situated Learning Legitimate Peripheral Participation*. Cambridge University Press, Cambridge, 1991.
- [7] Manca S.: *Multimedialità, comunicazione ed apprendimento: una rivisitazione dei rapporti tra parola, immagine ed azione*. TD, *Tecnologie Didattiche*, n. 24, 2001.
- [8] Mantovani G.: *Comunicazione e identità. Dalle situazioni quotidiane agli ambienti virtuali*. Il Mulino, Bologna, 1995.
- [9] Minorò V., Briano R., Persico D., Sarti L., Trentin G.: *EuMEDEA: a European online course to train European teachers in Environmental Education*. In Davies G. (ed.) *Teleteaching '98, Distance Learning, Training and Education*, Proc. Of the XV IFIP World Computer Congress, Vienna and Budapest, 31/8-4/9/1998, p. 725-736.
- [10] Spadaro P., Logorio M.B.: *Come analizzare il discorso nei forum: come e di cosa si parla*. http://www.formare.ericson.it/archivio/maggio_04/2spadaro.html

- 0
-
- [11] Sproull L., Kiesler S.: *Connections – New ways of working in the networked organizations*. Cambridge, MA.:The MIT Press, 1991.
- [12] Suchman L.: *Plan and situated actions: the problem of humane-machine communication*. Cambridge, Cambridge University Press, 1987.
- [13] Varisco B.M.: *Tecnologie Didattiche, Apprendimento e Valutazione*. In Persico D. (a cura di) Atti del Convegno TED, Genova, 12-14 febbraio 2001.
- [14] Zucchermaglio C.: *Vygotsky in azienda*. Carocci, Roma, 1996.
- [15] Wenger E.C.: *Communities of practice. Learning, meaning and identity*. Cambridge University Press, New York, 1998.
- 1
- 0

GIUSEPPE GILIBERTO è consulente di e-learning presso organizzazioni pubbliche e private tra cui CNIPA, CSI Piemonte e AICA. Si occupa di e-learning dal 1996 nella progettazione e realizzazione di piattaforme tecnologiche, di materiali didattici e di sistemi di formazione e delle loro applicazioni. Ha seguito il progetto della comunità virtuale ECDL come progettista e referente tecnico-scientifico.

E-mail: gibba@gibba.it

DANIELA MARIA RICCO è laureata in psicologia indirizzo lavoro ed organizzazioni, si è occupata di progetti di formazione presso grandi aziende pubbliche e private sia nelle fasi progettuali che in quelle erogative. Ha pubblicato In "Conversazioni virtuali" M. Bonaiuto, Guerrini e Associati, alcuni studi sulla comunicazione interna aziendale supportata dalle nuove tecnologie.

E-mail: daniela.ricco@gibba.it

GIULIANO RUSSO è laureato in Ingegneria Chimica alla Sapienza di Roma, ha lavorato fino al 2000 presso strutture informatiche della Pubblica Amministrazione. Si è occupato di sviluppo applicativo, progettazione di Banche Dati, e Project Management, prima come tecnico, infine con incarichi direzionali. Ha svolto in proprio attività di consulenza per la P.A. e aziende informatiche. Dal 2001 svolge per conto di AICA il ruolo di ispettore ECDL. Coordina il servizio ispettivo per il Lazio e partecipa come membro della redazione, alla Comunità Virtuale degli Esaminatori ECDL. Ha recentemente completato l'iter per Valutatore di Sistemi di Gestione per la Qualità.

E-mail: ispettorerm1@ecdlauditing.it

1

0

IL POLIEDRICO MONDO DELL'INFORMAZIONE



Giuseppe O. Longo

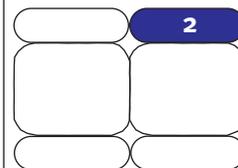
L'odierna "società dell'informazione" è solcata da una quantità crescente di dati elaborati e scambiati. Alla base di questa espansione vi è una tecnologia sempre più efficiente ed economica. Accanto agli aspetti tecnici, sono importantissimi gli aspetti concettuali del fenomeno, che comprendono la natura dell'informazione, il suo carattere contestuale e relazionale, i suoi aspetti semantici e pragmatici. Si delinea l'esistenza di un vero e proprio universo informazionale, che comprende la teoria di Shannon, ma che si estende ben al di là della formalizzazione matematica.

1. INTRODUZIONE

L'informazione sta alla base di ogni attività umana. Anzi tutta l'evoluzione, prima biologica e poi sociale e culturale, è stata contrassegnata da scambi d'informazione sempre più intensi e diffusi. Tuttavia solo verso la metà del Novecento si è cominciato a indagare questi fenomeni con strumenti non soltanto filosofici e qualitativi ma anche matematici. Questa impostazione, dovuta soprattutto a Shannon, è molto utile sotto il profilo ingegneristico, ma la sua utilità deriva da una notevole semplificazione: il *significato* dei messaggi e il loro valore *pragmatico* sono del tutto ignorati, perché l'unico scopo è quello di riprodurre presso l'utente ogni messaggio generato dalla sorgente in modo che sia distinguibile dagli altri possibili messaggi. La teoria matematica dell'informazione s'inquadra in una teoria assai più vasta, in cui sono importanti anche gli aspetti semantici e pragmatici dell'informazione, i suoi rapporti con i dialoganti e con il contesto comunicativo. Di questa teoria più ampia si vuol fornire qui una rassegna, per quanto somma-

ria. Si parte con l'illustrazione di alcune delle caratteristiche basilari dell'informazione, cominciando (paragrafo 2) dalle differenze più importanti tra il mondo della materia e il mondo della comunicazione. Rilievo è dato al passaggio dalla realtà (il territorio) alla sua rappresentazione (la mappa), che comporta una riduzione di informazione (paragrafo 3). Si sottolinea poi il rapporto essenziale tra l'informazione e il suo supporto (paragrafo 4). Questi concetti sono ripresi ed esemplificati nel paragrafo 5, dedicato alla simulazione, che può illuminare i rapporti e le differenze tra il mondo dei simboli e il mondo della materia, differenze riprese e approfondite nel paragrafo 6, che riguarda le nozioni di contesto e di complessità. Una delle caratteristiche più importanti del mondo dell'informazione è la ridondanza, e il paragrafo 7 ne illustra il rapporto con il significato e con il caso. Un breve cenno alla teoria matematica dell'informazione (paragrafo 8) fornisce un esempio molto particola-

La mappa non è il territorio
Alfred Korzybski



re, ma importantissimo per le applicazioni, di modello comunicazionale, mentre la prospettiva si allarga nel paragrafo 9, dove i concetti precedenti sono immersi nel contesto socioculturale, di cui costituiscono potenti fattori di trasformazione.

2. CREATURA E PLEROMA

Il Novecento è stato caratterizzato da un'attenzione esplicita per i fenomeni della comunicazione, che fino a quel momento erano rimasti un po' ai margini: l'evoluzione tecnica e le esigenze belliche diedero grande impulso allo studio quantitativo dell'informazione. Nel 1948, quando Claude Shannon [10] formulò la teoria matematica della comunicazione e Norbert Wiener (Figura 1) fondò la cibernetica [11], le discipline attinenti all'informazione, al significato, alla comunicazione, alla struttura, alla relazione e all'ordine avevano uno statuto piuttosto vago rispetto alla solidità delle teorie fisiche e chimiche. Ma il tempo perduto fu recuperato in fretta e tutto il settore ricevette una sistemazione formale.

Nonostante le limitazioni dovute al carattere quantitativo della teoria, l'incontro tra la matematica, l'ingegneria, la nascente informatica e il mondo della comunicazione fu quanto mai fecondo di risultati. In pochi anni nacquero e si svilupparono discipline come la teoria dell'informazione, la teoria dei controlli e dei servomeccanismi e la cibernetica, cui

si aggiunse più tardi, verso la fine degli anni cinquanta, l'intelligenza artificiale. Inoltre, dispiegandosi in tutta la sua ricchezza, il mondo della comunicazione cominciò ad agire da centro di aggregazione anche per discipline molto lontane dal nucleo tecnico-formale, come la psicologia, la sociologia, la filosofia e la linguistica.

Da una parte il processo comunicativo ricevette una formulazione matematica, dall'altra si venne delineando una sorta di "teoria generale dell'informazione": fu cioè riconosciuto che accanto al dominio della fisica - il mondo delle forze, delle masse e delle quantità di moto - esiste il dominio della forma, delle relazioni, delle differenze, dell'organizzazione e del significato. Le leggi che vigono in questo dominio sono spesso diverse da quelle della fisica e talora sorprendenti.

Seguendo Gregory Bateson, cui dobbiamo l'analisi più penetrante del concetto di informazione, chiameremo *Pleroma* il mondo della materia e delle forze e *Creatura* il mondo dell'informazione e della struttura. Nel Pleroma regna l'opacità pesante e indifferenziata della materia, mentre nella Creatura l'attività organizzatrice dell'uomo identifica e separa le cose, assegna i nomi e introduce leggi e distinzioni. Mentre nel Pleroma ogni cosa rappresenta soltanto sé stessa, nella Creatura ogni cosa può, in seguito a un accordo, rappresentare ogni altra cosa e divenire pertanto un *simbolo*. I codici simbolici, aprendo le prospettive amplissime del significato, consentono ogni sorta di gioco linguistico. In genere il significato di una cosa non le è intrinseco, ma le è conferito dall'attività comunicativa dell'uomo. In questo senso si giustifica il termine *Creatura*, perché è l'attività conoscitiva e linguistica che crea questo mondo [1, 9].

Nella Creatura non vi sono leggi di conservazione per l'informazione: poiché gli osservatori (viventi) posseggono fonti energetiche e supporti materiali propri, essi sono in grado di replicare presso di sé l'informazione ricevuta da una fonte esterna, come fa chi a lezione prende appunti su un taccuino: quindi l'informazione *si moltiplica* per il numero degli osservatori o utenti (e non si divide come la materia o l'energia) (Figura 2). L'assenza di

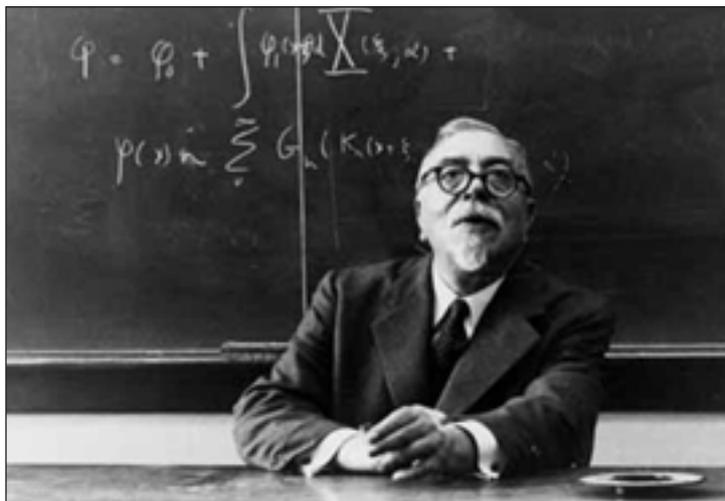


FIGURA 1
Norbert Wiener

informazione può essere informazione: una risposta non data può scatenare una reazione anche vivace. Per esempio la risposta *non data* a un'ingiunzione dell'ufficio delle imposte può provocare l'arrivo di una multa. L'ufficio confronta l'alternativa "non risposta" con l'alternativa "risposta" e agisce sulla base della differenza tra le due. Dunque ciò che conta nella Creatura sono le *differenze*. L'informazione sta nelle differenze e l'unità d'informazione può essere definita come *la più piccola differenza capace di generare una differenza*. I canali di comunicazione sono i supporti che trasferiscono le differenze, opportunamente trasformate e codificate. Una differenza è sempre tra un A e un B, tra un prima e un dopo, tra un qui e un altrove, ma non si può ubicare, perché non ha le caratteristiche spazio-temporali degli oggetti materiali. E il mondo della differenza e della comunicazione, la Creatura, è in sostanza il mondo degli esseri viventi, in particolare dell'uomo. Grazie al loro metabolismo, gli organismi accumulano energia e possono rilasciare questa energia in risposta a uno stimolo anche minimo (o alla perdurante assenza di uno stimolo). Basta una parola per scatenare una rissa, basta un lungo silenzio per provocare un'aggressione.

Un esempio, forse un po' crudele, che Bateson porta per illustrare quanto le leggi della Creatura siano diverse da quelle del Pleroma riguarda il diverso comportamento del sasso e del cane in seguito a un calcio: il sasso (che appartiene solo al Pleroma) compie una parabola, atterra e lì sta. Se la pedata è abbastanza vigorosa, anche il cane (che appartiene al Pleroma e alla Creatura) può compiere una parabola secondo le leggi della fisica, ma di solito questo aspetto del suo comportamento non è interessante. Interessante è ciò che il cane fa dopo essere atterrato: per esempio può decidere di rivoltarsi contro chi l'ha maltrattato e tentare di azzannarlo. Ma l'energia che il cane mette in giuoco in questo attacco non è quella della pedata, bensì quella "collaterale" del suo metabolismo, e quest'energia viene attivata non dal calcio in sé bensì dal dolore provato e dal trauma "relazionale" subito dal cane. Tanto è vero che, invece di attaccare l'uomo, il cane può anche decidere di fuggire, mentre il sasso non ha scelta: il determinismo della rela-

zione causa-effetto vale solo per il sasso, il cui comportamento è invariabile e, almeno in linea di principio, prevedibile, mentre il comportamento del cane dipende da fattori di natura sistemica e contestuale difficili da valutare (la sua storia precedente, il suo rapporto con l'uomo che gli dà il calcio, il numero di calci che ha preso in vita sua e così via) [1] (Figura 3).



FIGURA 2

La lezione di anatomia del dottor Tulp di Rembrandt



FIGURA 3

Cave canem

Un alieno assiste a una conferenza

Entrai in una vasta sala illuminata fiocamente, dove molti Terrastranieri stavano seduti in comode poltrone, tutte rivolte nella stessa direzione. Guardando meglio, mi resi conto che il pavimento della sala era lievemente inclinato e convergeva, per così dire, verso un punto centrale sfarzosamente illuminato, dove stava in piedi un Terrastraniero alto e imponente. Immaginai che potesse essere il Sindaco della Città, o addirittura il Re, visto quanto era distinto e soverchiante fra tutti. Quando la confusione che mi aveva colto all'entrare in sala si dileguò e potei distinguere meglio i particolari, ebbi l'amara sorpresa di vedere e udire che il maestoso Terrastraniero che ritenevo fosse il Re emetteva dal suo apparato masticatorio strani suoni gutturali, gridi e squittii e brontolii, come se soffrisse di un molesto dolore fisico. Ma, con mio grande stupore, nessuno moveva un dito per aiutarlo o assisterlo: anzi, al contrario, mirando le facce dei Terrastranieri presenti, vidi che alcuni ghignavano, altri con ogni evidenza si divertivano, altri ancora sorridevano beatamente come se il grande dolore del loro Monarca fosse per loro causa di diletto. Questo fu per me il segno di una perfida crudeltà ed ebbi conferma dei tanti racconti che avevo udito sulla malvagità della popolazione di Terrastrana.

Insomma nella Creatura valgono principi di funzionamento diversi da quelli del Pleroma. Nel Pleroma non c'è informazione (Pleroma significa "pieno", dunque "privo di differenze"), ma esso è la matrice di tutte le differenze e la sorgente di tutte le informazioni. La struttura del Pleroma, in sé, non riesce a spiegare (in modo adeguato) molti degli eventi che interessano l'uomo in quanto essere informazionale, sociale, culturale. Per esempio una conferenza può essere descritta in molti modi e a molti livelli diversi. È possibile (in linea di principio) darne una descrizione in termini di meccanica quantistica oppure in termini di chimica molecolare, ma queste descrizioni hanno un interesse limitato e non colgono l'essenza dell'evento "conferenza". Ciò vale in quasi tutte le situazioni rilevanti per gli esseri umani, le quali possono essere descritte, spiegate e comprese *adeguatamente* solo in termini di informazione, relazione, significato e comunicazione. La descrizione ordinaria di una conferenza, nel linguaggio comune, è adeguata e facile da comprendere grazie alla lunga *esperienza* che l'osservatore ha delle conferenze. Un Alieno che non avesse esperienza di confe-

renze ne darebbe una descrizione stravagante, che non ne coglierebbe (quella che per noi è) l'essenza (*vedi riquadro*).

Quando si parla di "essenza di un evento" o di "descrizione adeguata" si fa sempre riferimento implicito a un osservatore o destinatario: dunque l'informazione è *relativa* a un osservatore. La stessa energia sonora modulata associata a una data parola può causare una risposta amichevole, uno sguardo stupito o uno scatto d'ira a seconda della storia precedente dell'ascoltatore, della sua relazione col parlante, del suo stato d'animo, dell'interesse che ha per il contenuto della frase, della lingua cui appartiene la frase e del contesto in cui essa è pronunciata. *L'informazione sta nell'orecchio dell'ascoltatore più che nella bocca del parlante.*

3. MAPPA E TERRITORIO

Torniamo all'aforisma di Korzybski (Figura 4) citato in esergo: "la mappa non è il territorio". La mappa appartiene al mondo della Creatura: dal Pleroma, cioè dal territorio, ogni osservatore può ricavare la sua mappa particolare, che dipende dai suoi interessi ed è diversa da ogni altra mappa. Per esempio, se il territorio è una regione geografica chi è interessato agli aspetti agricoli ne costruirà una mappa diversa da chi è interessato agli aspetti minerari, o politici, o idrologici. Tutte queste mappe si riferiscono allo stesso territorio, sono tra loro compatibili ma diverse, e nessuna fornisce una descrizione esauriente della regione. La regione in sé è muta, siamo noi che la facciamo "parlare": la costruzione delle varie mappe, cioè la costruzione della Creatura, rende esplicite le informazioni che nel territorio (Pleroma) sono implicite. Anche un libro è muto finché non c'è un lettore che lo fa vivere. In questo senso l'attività dell'osservatore è costruttiva, poiché la Creatura nasce dall'interazione conoscitiva tra osservatore e Pleroma. Il Pleroma è la matrice e la Creatura è il deposito di tutte le conoscenze (Figura 5). Noi di continuo trasformiamo il vasto territorio del mondo pleromatico nella complicata mappa creaturale dei *segni*. I segni hanno una loro struttura, una loro ecologia, nella quale siamo immersi a tal punto da dimenti-

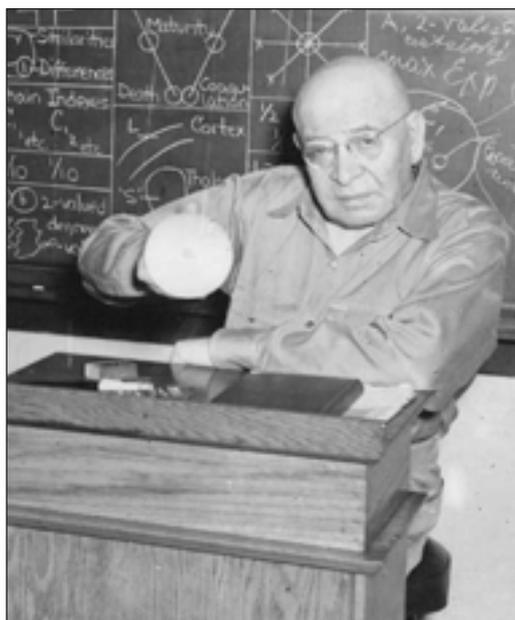


FIGURA 4

Alfred Korzybski

care talora la realtà che li ha originati grazie al nostro assiduo lavoro, e ci domandiamo il senso di tutti quei segni. Per fortuna il mondo è sempre là e possiamo intraprenderne nuove interpretazioni, ricavandone nuove mappe e nuovi segni (*vedi riquadro*).

Schematizzando molto, per ogni osservatore l'informazione può assumere tre aspetti:

1. aspetto *sintattico* (rilevamento delle differenze tra le parti del messaggio o tra messaggi successivi);
2. aspetto *semantico* (confronto delle differenze rilevate con altre, registrate in precedenza, per ricavarne il *significato*);
3. aspetto *pragmatico* (impiego delle differenze rilevate e interpretate per il conseguimento dei *fini*).

Ciascuno di questi tre aspetti presuppone il precedente: l'informazione può essere impiegata per agire solo quando la si sia interpretata; e può essere interpretata solo quando la si sia rilevata. Ma in molte circostanze pratiche sono tantissime, potenzialmente infinite, le differenze che si possono cogliere in un oggetto o in un fenomeno, cioè in un territorio. Quali fra tutte queste differenze rileverà effettivamente il nostro osservatore, trasformandole da potenziali in attuali? Come illustra l'esempio delle carte geografiche, la scelta dipende dal suo interesse, quindi dai fini che ha



FIGURA 5

America sive novi orbis (A. Ortelius)

I segni per Calvino

Situato nella zona esterna della Via Lattea, il Sole impiega circa 200 milioni di anni a compiere una rivoluzione completa della Galassia.

Esatto, quel tempo là ci si impiega, mica meno - disse Qfwfq, - io una volta passando feci un segno in un punto dello spazio, apposta per poterlo ritrovare duecento milioni d'anni dopo, quando saremmo ripassati di lì al prossimo giro. Un segno come? È difficile da dire perché se vi si dice segno voi pensate subito a un qualcosa che si distingue da un qualcosa, e lì non c'era niente che si distinguesse da niente; voi pensate subito a un segno marcato con qualche arnese oppure con le mani, che poi l'arnese o le mani si tolgono e il segno invece resta, ma a quel tempo arnesi non ce n'erano ancora, e nemmeno mani, o denti, o nasi, tutte cose che si ebbero poi in seguito, ma molto tempo dopo. La forma da dare al segno, voi dite non è un problema perché, qualsiasi forma abbia, un segno basta serve da segno, cioè sia diverso oppure uguale ad altri segni: anche qui voi fate presto a parlare, ma io a quell'epoca non avevo esempi a cui rifarmi per dire lo faccio uguale o lo faccio diverso, cose da copiare non ce n'erano, e neppure una linea, retta o curva che fosse, si sapeva cos'era, o un punto, o una sporgenza o rientranza. Avevo l'intenzione di fare un segno, questo sì, ossia avevo l'intenzione di considerare segno una qualsiasi cosa che mi venisse fatto di fare, quindi avendo io, in quel punto dello spazio e non in un altro, fatto qualcosa intendendo di fare un segno, risultò che ci avevo fatto un segno davvero.

(Italo Calvino, *Un segno nello spazio*, *Le Cosmicomiche*, Einaudi, Torino, 1965).

in quel momento: ecco allora che il cerchio si chiude, perché il rilevamento delle differenze (aspetto sintattico) è guidato dai fini dell'osservatore (aspetto pragmatico) (Figura 6). L'attività conoscitiva dell'uomo, cioè la costruzione delle mappe, è sempre guidata dai bisogni, dagli interessi e dai fini dell'individuo, è limitata dai suoi mezzi e vincolata alle sue capacità. Le mappe vengono poi usate per conseguire gli scopi nell'ambiente reale (o per simularne il conseguimento in un ambien-

te mentale). Azione e conoscenza sono dunque inseparabili: ogni azione è fonte di conoscenza e ogni conoscenza richiede un'azione [2]. Le conoscenze acquisite mediante un'azione costituiscono la base per ulteriori azioni che apportano nuove conoscenze [7].

4. INFORMAZIONE E SUPPORTO

Non esiste informazione senza supporto: le differenze debbono "incarnarsi" in materia o in energia. Ma a tutta prima sembra che il supporto, benché indispensabile, non sia essenziale e che si possa trasferire l'informazione da un supporto all'altro senza perdite. Se l'informazione consiste nelle differenze o modulazioni del supporto, basta introdurre nel secondo supporto differenze o modulazioni isomorfe alle differenze o modulazioni pre-

senti nel primo: questa è l'operazione di *codifica*. A prima vista ciò sembra possibile senza difficoltà, specie quando si abbia a che fare con i codici discreti o digitali. Ma se si analizza meglio il procedimento di codifica, si scopre che esiste sempre un livello di osservazione fine al quale il trasferimento non è più fedele. Insomma non si può separare l'informazione dal supporto originale se non a patto di impoverirla e distorcerla: un concerto per violino non può essere eseguito col trombone se non con gravi distorsioni. Sarà comunque l'osservatore a dichiararsi soddisfatto o insoddisfatto del grado di fedeltà della codifica (nel caso digitale l'osservatore si accontenta di una codifica anche grossolana, purché conservi la *distinguibilità* tra i messaggi).

L'impossibilità di una codifica del tutto fedele deriva dal fatto che come non esiste informazione senza supporto, così non esiste supporto senza informazione: ogni supporto possiede un'informazione intrinseca, originaria, costituita dalle differenze che contiene prima di ogni processo di codifica e che non sono frutto di intenzionalità. Quando poi si introducono nel supporto le informazioni intenzionali, che tendono a modularlo, esse interagiscono con le informazioni preesistenti e ne vengono più o meno distorte: la pesantezza e la riottosità della materia si oppongono all'operazione di codifica. Ciò conferma l'intimo legame tra informazione e supporto. Facciamo un esempio: scrivere un messaggio su un supporto significa introdurre delle modulazioni o differenze, ma il risultato dell'operazione dipende dalla natura del supporto: la pietra (Figura 7) presenta una resistenza maggiore della carta, scrivere su un foglio già scritto può essere complicato, la tastiera del computer è più "leggera" della stilografica. Ogni supporto quindi condiziona l'esito della scrittura e incide profondamente non solo sul modo ma anche su ciò che si scrive, tanto da giustificare il detto di McLuhan "il mezzo è il messaggio". I paleografi sono abituati a interpretare le abbreviazioni cui si ricorreva per la difficoltà di scolpire il marmo, ma anche nei messaggi sui cellulari si riscontra un continuo ricorso alle abbreviazioni. Chi ha praticato la scrittura letteraria a mano e poi è passato alla videoscrittura ha percepito un cambiamento che si riflette perfino sullo *stile* della narrazione [9].

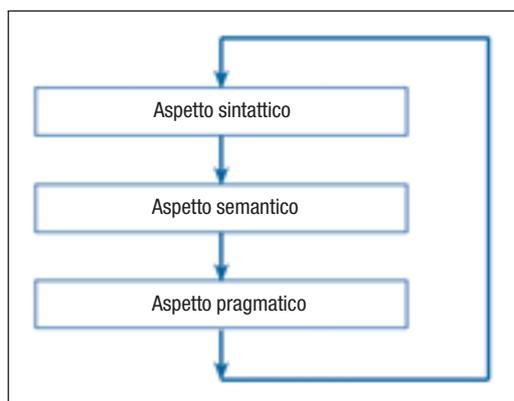


FIGURA 6
I diversi aspetti dell'informazione

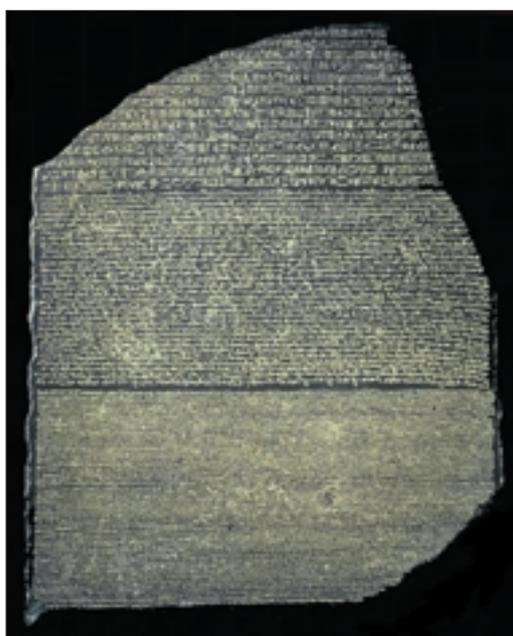


FIGURA 7
Stele di Rosetta

5. LA SIMULAZIONE

Riprendiamo ora i concetti precedenti, in particolare il rapporto tra informazione e supporto, sotto un'angolazione diversa, fornita dalla simulazione. La simulazione appartiene al mondo dell'informazione e non della materia e costituisce un esempio importante e diffuso di passaggio dal territorio alla mappa, cioè da un fenomeno a un suo modello. Questo passaggio comporta come sempre una perdita d'informazione, ma in molti casi la simulazione è non solo utile ma addirittura indispensabile.

Per gli esseri umani la *simulazione* costituisce uno strumento di notevole valore economico e di sopravvivenza, perché ci evita i rischi e gli sprechi legati all'attuazione pratica. Prima di intraprendere un'azione concreta, di solito la simuliamo servendoci della nostra mente, o di altri strumenti che della mente costituiscono un potenziamento o un prolungamento. Possiamo così analizzare i possibili effetti dell'azione e decidere se compierla, correggerla o rinunciarvi.

Il mondo dell'informazione è caratterizzato dalla possibilità di stipulare codici *arbitrari*: come abbiamo detto, una cosa può, per convenzione, significare qualsiasi altra cosa; ma la simulazione va al di là di questa codifica arbitraria e convenzionale, poiché si fonda su una somiglianza, almeno parziale, e istituisce tra le due "cose", quella simulata, diciamo il fenomeno, e quella simulante, diciamo il modello, una corrispondenza molto stretta almeno a qualche livello di descrizione. Se la corrispondenza si verifica a tutti i livelli (nei limiti della precisione adottata), non si parla più di simulazione, bensì di "riproduzione". Per esempio nel caso di un cervello umano e di un calcolatore elettronico che effettuino un'operazione aritmetica, la corrispondenza si ha a livello dei passaggi aritmetici, ma non a livello strutturale né a livello funzionale fine, poiché a questi livelli non c'è somiglianza tra i neuroni e le loro attività da una parte e i circuiti e le loro attività dall'altra.

Per giudicare l'adeguatezza di una simulazione non ci si basa su una corrispondenza totale, bensì su una corrispondenza parziale di esiti e di effetti osservabili, adottando

una prospettiva che è tipica del comportamentismo. Con riferimento all'intelligenza artificiale, il famoso criterio proposto da Turing (Figura 8) nel 1950 per dichiarare intelligente una macchina si basa appunto su una simulazione di natura comportamentistica. Mediante telescrivente, un esaminatore pone domande a una persona e a una macchina e, ancora mediante telescrivente, ne riceve le risposte. Entrambi gli esaminati si sforzano di persuadere l'esaminatore di essere umani e, sulla sola base delle risposte ricevute, l'esaminatore deve stabilire quale dei due è davvero l'uomo. Se la macchina riesce a ingannare l'esaminatore per un tempo stabilito, è dichiarata intelligente. La macchina deve compiere in questo caso una simulazione più complessa e difficile di quella relativa all'esecuzione di un'operazione aritmetica, ma non si richiede certo che somigli all'uomo in tutto e per tutto: basta che manifesti un comportamento verbale umano [5].

A proposito del problema fondamentale dell'intelligenza artificiale, cioè se la mente sia simulabile e trasferibile in un altro supporto, possiamo arrischiare questa proposta: se la mente sta tutta nel mondo informazionale, come afferma il funzionalismo, una sua simulazione almeno a qualche livello significativo è possibile; se sta anche nel mondo fisico, come molti ritengono, la cosa è più ardua, poiché anche la materia di cui è fatto il supporto della mente è rilevante. Si



FIGURA 8
Alan Turing

Due esempi di simulazione

Consideriamo due esempi: le simulazioni al computer di un matematico e di una mucca. Mentre la mucca simulata non può essere munta e il latte simulato che essa produce non può essere bevuto (se non da un contadino simulato, che simulerebbe di berlo), nel caso del matematico simulato le dimostrazioni simulate che egli produce sono in tutto e per tutto equivalenti alle dimostrazioni eseguite da un matematico vero.

Che differenza c'è tra latte e dimostrazioni? Le dimostrazioni appartengono (quasi) per intero alla Creatura, mentre il latte appartiene (quasi) per intero al Pleroma e non è possibile simulare con l'informazione gli oggetti fisici. Questa impossibilità risulta più evidente se si adotta un criterio di distinzione basato sugli *effetti* che le cose e le loro simulazioni hanno sul mondo reale (il nostro mondo): nel caso del latte gli effetti sono molto diversi, mentre nel caso della dimostrazione gli effetti sono identici. Tenendo presente la distinzione tra informazione e supporto, possiamo anche dire che per il latte il supporto (cioè gli atomi e le molecole che lo compongono) è essenziale: non si può modificare l'identità degli atomi e delle molecole, poiché la configurazione, le relazioni reciproche e i legami chimici, che ne costituiscono la parte strutturale o informazionale, non sono sufficienti a darci il latte. Se gli atomi di carbonio vengono sostituiti da atomi di silicio, pur supponendo di conservare tutte le relazioni tra gli atomi, non si ottiene più il latte ma un sasso (forse liquido). Per quanto riguarda la dimostrazione, invece, il supporto, benché indispensabile, è inessenziale: quello che conta sono le relazioni e le differenze, cioè le informazioni, che possono essere riprodotte anche nel computer.

Si rifletta sui casi seguenti:

- Un intervento chirurgico; un documentario filmato sull'intervento; un intervento fittizio in un film; o in un'esercitazione didattica in cui si usi un paziente vero, un cadavere, un robot per anestesisti; il resoconto scientifico scritto di un intervento; la parola "intervento".
- La dimostrazione di un teorema eseguita alla lavagna da un matematico; la stessa dimostrazione eseguita a teatro (oppure mimata) da un attore; una successione di segni senza senso tracciati a teatro su una lavagna da un attore che finge di eseguire la dimostrazione; la stessa dimostrazione svolta da un computer, da un sordomuto, da un profano che abbia imparato a memoria i passaggi ma non capisca ciò che scrive (Figura 9).

6. L'EPISTEMOLOGIA INFORMATIVALE

L'epistemologia o gnoseologia è, grosso modo, l'insieme dei procedimenti che seguiamo per acquisire le conoscenze, cioè per costruire la Creatura. Sul concetto di informazione intesa come differenza che genera una differenza si basa un'epistemologia piuttosto diversa da quella tradizionale delle scienze della natura, in particolare della fisica. Semplificando molto, mentre la fisica raggiunge i propri risultati grazie a una semplificazione che consiste nel sopprimere il contesto e nel considerare solo sistemi e fenomeni isolati, nell'epistemologia informativa *il contesto è fondamentale*: non vi sono fenomeni, eventi, comunicazioni, accadimenti, trasformazioni che non siano essenzialmente inseriti in un contesto, nel senso che solo dal contesto essi ricevono il loro significato e solo in base ad esso possono essere descritti e spiegati. In altre parole: mentre la fisica ha ottenuto e in buona parte ancora ottiene i suoi cospicui risultati attraverso la pratica del riduzionismo, cioè grazie all'eliminazione di tutti i legami che a priori connettono il fenomeno o il sistema considerato al contesto più ampio, e in particolare i legami tra il fenomeno e l'osservatore, nel campo dell'informazione e della comunicazione questi nessi non possono essere recisi perché costituiscono le *relazioni* che definiscono il fenomeno per quello che è.



FIGURA 9

Estrazione della pietra della follia di Hieronymus Bosch

osservi che nessun fisico confonde mai una gassia col suo modello matematico, mentre in intelligenza artificiale mente e programma vengono confusi: tanto forte è la tendenza a considerare la mente un'entità solo informativa (vedi riquadro).

Che nell'ambito della comunicazione il contesto sia importante, anzi ineliminabile, è noto a tutti dall'esperienza della lettura: ogni lettera è inserita in una sillaba, ogni sillaba in una parola, ogni parola in una frase, e così via; e ciascun elemento riceve il proprio significato da tutti i contesti più ampi, cioè dagli elementi di livello superiore di cui è parte. Come si vede, in genere esistono più contesti, di ampiezza diversa, contenuti l'uno nell'altro. D'altra parte se la parola riceve il suo significato dalla frase, la frase a sua volta lo riceve dalle parole che la compongono: la nozione di *significato contestuale* non è lineare o unidirezionale, bensì circolare.

La scoperta di questa circolarità è di enorme importanza, perché storicamente la prima impressione che l'uomo ha avuto nei confronti del mondo è che esso funzioni in base a *catene causali lineari*, che hanno origine in un certo punto (spesso arbitrario, o meglio scelto in base a finalità pratiche) e finiscono chissà dove nella vastità dei fenomeni. Per contro, la descrizione o spiegazione contestuale, imperniata sul concetto d'informazione, si basa su circoli o anelli di retroazione (o *feedback*) [1, 11].

Come ho detto, il Pleroma è l'oscura matrice potenziale di tutte le differenze, ma non contiene differenze: è il mondo di base, è la realtà in sé, che non potremo mai conoscere direttamente perché i nostri sensi e il nostro intelletto la sottopongono immediatamente a una radicale rielaborazione. Il Pleroma è il fondo primordiale di una realtà inarrivabile perché ogni volta che tentiamo di coglierla in sé la *filtriamo* con le nostre operazioni mentali e con i nostri sensi, e il risultato di questa elaborazione (che è in gran parte *inconsapevole*) è il mondo così come noi lo percepiamo. In questo senso, l'epistemologia basata sull'informazione non è né *idealista* (cioè non afferma che il mondo è una nostra creazione mentale) né *realista* (cioè non afferma che esiste un mondo esterno che viene rispecchiato passivamente dalla nostra mente): è invece *costruttivista*. La realtà percepita è il risultato di una continua interazione costruttiva tra noi e il mondo in sé [1].

Nell'ambito dell'epistemologia informazionale non esiste un'*unica descrizione* di un fenomeno: per spiegarlo, o anche solo per descriverlo,

sono utili, anzi necessarie, più descrizioni, che non sempre sono riconducibili l'una all'altra e che a volte possono anche essere tra loro contrastanti. Svanisce dunque il pregiudizio (o la speranza) che vi sia *una descrizione vera* del mondo, che la fisica dovrebbe via via svelare. Ma anche la fisica ha un po' attenuato la sua pretesa di offrire, prima o poi, l'*unica* descrizione vera del mondo e contempla descrizioni multiple dello stesso fenomeno. C'è quindi una sorta di convergenza tra l'epistemologia tradizionale e quella informazionale, e il ponte è costituito dalla meccanica quantistica, per la quale il mondo (o il fenomeno o il sistema che stiamo indagando) risponde in maniera diversa a seconda del punto di vista che adottiamo. A seconda delle domande che gli poniamo, lo "stesso" fenomeno si presenta sotto forma ondulatoria o corpuscolare. La meccanica quantistica ci obbliga ormai ad accettare la stretta dipendenza del fenomeno osservato dall'osservatore e dalle sue domande. Anche nel dominio della fisica si va dunque rivelando una profonda *struttura comunicativa* che la fisica tradizionale non aveva scorto o aveva cercato di escludere [6].

Un'altra caratteristica importante dell'epistemologia dell'informazione è che mentre nel Pleroma basta il rilevamento di una differenza per innescare un'azione (il rilevamento di una temperatura diversa da quella di riferimento fa scattare il termostato e la caldaia si accende o si spegne), nella Creatura alla base dell'azione, oltre al rilevamento, c'è quasi sempre anche l'*interpretazione* delle differenze. La Creatura è infatti il mondo simbolico e segnico, il mondo dei codici, e i codici vanno interpretati.

7. LA RIDONDANZA

La *ridondanza*, è una delle caratteristiche più importanti dei fenomeni comunicazionali, come testimonia qualunque testo, parlato o scritto, in qualunque lingua naturale, ma più in generale come dimostra il nostro rapporto col mondo. Della ridondanza si può dare una definizione molto generale. Presentiamo a un osservatore una configurazione (una figura, uno scritto, una successione numerica...) in parte nascosta, in modo che l'osservatore possa osservarne solo una porzione. Se basandosi sulla parte visibile l'osservatore può

fare inferenze sulla parte nascosta, cioè se può risalire alla configurazione totale con esito migliore di quello puramente aleatorio, allora la parte visibile contiene informazioni su quella nascosta e si dice che nel suo complesso la configurazione è *ridondante*. Quando poi la parte nascosta viene mostrata, l'osservatore ne ricava una quantità d'informazione (o di "sorpresa") minore di quanto ne avrebbe ricavato se prima non avesse avuto accesso alla parte scoperta. Al limite, se la parte scoperta consente di risalire in modo univoco e completo a quella nascosta, questa, una volta svelata, non fornisce alcuna informazione che l'osservatore già non possedeva (non gli provoca nessuna sorpresa, non gli elimina nessuna incertezza). La presenza di ridondanza e la sua quantità dipendono ancora una volta dall'osservatore, dalle sue capacità di rilevazione e di associazione, dalla sua storia, dal suo addestramento e così via.

Dal punto di vista comunicativo in genere, e in particolare dal punto di vista tecnico, le implicazioni della ridondanza sono enormi. Per esempio la ridondanza conferisce *robustezza* a un testo scritto o a un'enunciazione orale, cioè consente di ricostruirli anche quando ne venga distrutta o distorta una parte, perché la parte superstita può supplire a quella perduta. Per esempio un testo italiano scritto contiene ridondanza: infatti eliminandone alcune lettere (non troppe!), in particolare le vocali, l'univocità dell'interpretazione non ne risente. Ciò significa che le consonanti da sole bastano alla ricostruzione (alla comprensione) del testo.

Quindi la ridondanza si oppone validamente ai disturbi e alle interferenze nella comunicazione, e infatti i codici che si usano per correggere gli errori di trasmissione nei canali per le telecomunicazioni sono basati sull'introduzione di ridondanza nei messaggi trasmessi [3, 7, 10].

La ridondanza è strettamente connessa con il *significato*, anzi si può arrivare a dire con Bateson (Figura 10) che ridondanza e significato sono *sinonimi*: una figura contiene ridondanza quando da una sua parte possiamo risalire con buona probabilità al tutto, cioè quando osservandone quella parte possiamo dire "ho capito che cosa rappresenta la figura, ne ho colto il significato!". Se una figura è priva di ridondanza, ogni sua parte ci fornisce una quantità d'informazione che non dipende dalle parti già osservate: non la si può "capire", è una figura aleatoria, come la successione binaria di teste e croci che si ricava lanciando una moneta non truccata. Osservando la parte iniziale della successione non siamo in grado di fare sulla parte successiva previsioni migliori di quelle che potremmo fare senza osservare il tratto iniziale: ad ogni lancio testa e croce sono equiprobabili e questo è l'unico dato che possediamo per quanto lunga sia la successione che abbiamo potuto osservare fino a quel momento. Non c'è modo di "capire" una successione binaria aleatoria, priva di ridondanza e di significato, mentre si può capire una successione strutturata, cioè ridondante. Una volta capita la successione, si può anche comprimerla, cioè riassumerla in una formula (vedi riquadro) [1, 4, 8].

L'informazione algoritmica

Consideriamo una successione binaria. Si definisce contenuto d'informazione algoritmica della successione la lunghezza del più breve di tutti i programmi di calcolo capaci di fornire la successione e di arrestarsi. Questa definizione, pur essendo precisa e coerente, non ci consente di stabilire il contenuto d'informazione algoritmica di una successione, perché qualunque programma noi troviamo capace di riprodurre la successione e di arrestarsi, nessuno ci assicura che non ne esista un altro più breve che abbia le stesse proprietà. Anzi, può darsi che l'algoritmo generatore più breve non sarà mai scoperto. Questa stranezza non ci impedisce di fare delle considerazioni di carattere generale. Date due successioni: $S = 01010...01$, che è una ripetizione (lunghissima) della coppia 01, e T , che invece riproduce una serie altrettanto lunga di lanci di una moneta perfetta, è evidente che esiste un programma brevissimo che genera la prima (per esempio "scrivi '01' n volte") mentre la seconda è *incomprimibile*, cioè non presenta regolarità che si possano sfruttare per scrivere un programma più breve di T stessa. Non sempre le regolarità di una successione sono evidenti: per esempio il numero π (il cui programma generatore è semplicissimo: "rapporto tra una circonferenza e il suo diametro") scritto in binario "somiglia molto" a una successione aleatoria. Quindi è possibile valutare il livello di casualità di una successione binaria (o anche decimale). Una successione infinita può benissimo essere compressa: la successione 1, 3, 5, 7... può essere riassunta dalla formula $2n + 1$ (dove n assume tutti i valori interi non negativi); la successione 1, 2, 5, 10, 17... è riassunta da $n^2 + 1$; la serie di Fibonacci 0, 1, 1, 2, 3, 5, 8, 13... è generata dalla formula ricorsiva $a(n) = a(n-1) + a(n-2)$, e così via. Altre successioni hanno formule generative più complicate ma sempre finite, altre non ne hanno affatto (e allora sono casuali). Ciò dimostra che una successione infinita può fornire una quantità *finita* d'informazione perché contiene delle "regolarità", cioè ridondanza. Si noti che i tre puntini posti alla fine delle porzioni esplicite delle successioni stanno a indicare che la successione continua in modo "regolare", altrimenti la formula generatrice può essere molto complicata o non esistere affatto.

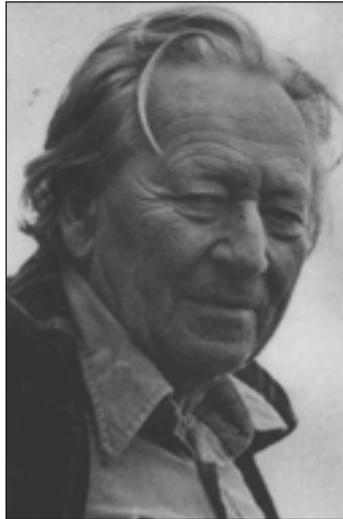


FIGURA 10
Gregory Bateson



FIGURA 11
Claude Elwood
Shannon

8. LA TEORIA MATEMATICA DELL'INFORMAZIONE

La teoria matematica dell'informazione formulata da Shannon nel 1948 [10] presenta un interessante paradosso: da una parte essa è figlia del suo tempo e segue i principi formali e riduzionistici delle discipline quantitative, riconducendo lo sfaccettato concetto di informazione nell'alveo della formalizzazione riduzionistica. Dall'altra questo concetto è per sua natura contestuale e relazionale e si oppone a qualunque forma di riduzionismo. Il desiderio di fornire dell'informazione, della comunicazione, del significato e così via, una rappresentazione quantitativa, atomistica e acontestuale s'inquadra nella grande operazione "imperialistica" della fisica matematica, che ha tentato e in parte ancora tenta di ricondurre ai propri metodi e protocolli tutto lo scibile umano. Che l'informazione sia per sua natura refrattaria a questa colonizzazione è dimostrato dal fallimento di ogni tentativo di costruire una teoria matematica dell'informazione *semantica*.

La teoria di Shannon (Figura 11) riguarda solo l'informazione *sintattica*, basata sulla scelta tra più alternative distinte, ciascuna a priori non certa e non impossibile, e quindi non coglie tutta l'essenza dell'informazione, che sta anche nel significato e nel valore pragmatico che essa ha per gli interlocutori, nelle modifiche che introduce nella loro relazione, negli aspetti *metacomunicativi* che sempre accompagnano gli atti linguistici umani (verbal e non verbal): di questi aspetti fanno parte i

metamessaggi, cioè quei "messaggi sui messaggi" che determinano la "natura" dei messaggi scambiati. Un esempio di metamessaggio è il seguente: "il messaggio che ti sto inviando è un gioco (oppure un comando, un resoconto di un fatto accaduto o un'anticipazione di un fatto che accadrà, una finzione, una narrazione)". I metamessaggi non sono necessariamente verbali, possono essere veicolati dal tono di voce, dall'atteggiamento, dall'espressione del viso, dal contesto. Per esempio la frase: "adesso ti uccido" pronunciata sul palcoscenico durante una recita è qualificata di "finzione" dal contesto teatrale; pronunciata da un delinquente nel corso di una rapina, lo stesso messaggio "adesso ti uccido" ha valore letterale o costituisce una minaccia. Siccome possono essere qualificati in svariati modi, i messaggi non hanno dunque valore assoluto [1].

La comunicazione mediata dalla tecnologia è in genere molto più elementare di quella che si svolge tra umani e spesso sono carenti le dimensioni semantica e pragmatica, così come mancano i metamessaggi e in genere gli aspetti metacomunicativi, che nella comunicazione umana sono importantissimi e vengono convogliati da canali paralleli a quello principale (espressione, tono di voce e così via). Questa semplificazione si riflette nell'uso di linguaggi artificiali, molto più semplici delle lingue naturali, meno flessibili e meno espressivi, ma nello stesso tempo meno ambigui. Questi linguaggi sono *acontestuali*, cioè non sono qualificati da metamessaggi, e il loro

aspetto semantico e pragmatico è univocamente ricondotto al livello sintattico. Questa loro povertà presenta tuttavia il grande vantaggio di renderli “comprensibili” al computer. Questi linguaggi artificiali si prestano dunque all’elaborazione automatica [5].

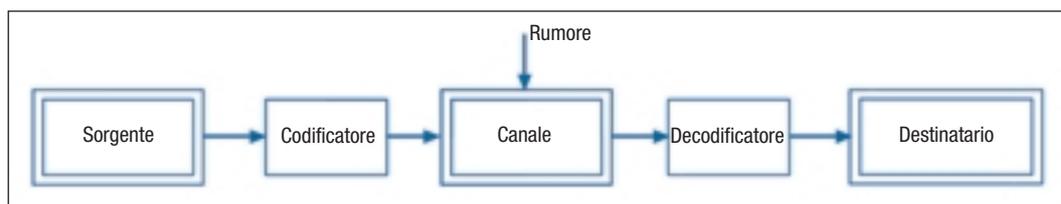
La teoria sintattica di Shannon coglie solo l’aspetto differenziale, ma lo coglie in maniera comunque più ricca di quanto non faccia di solito la teoria fisica. Quando due interlocutori usano per comunicare un alfabeto costituito da un numero finito di simboli, diciamo $(0,1,2,\dots,n)$, è importante che essi non solo si accordino sul significato dei simboli (0 significa “c’è il sole”, 1 significa “piove”, 2 “tira vento”, ..., n “il cielo è nuvoloso”), ma sappiano anche *quanti* sono i simboli che stanno usando e sappiano *distinguerli* tutti l’uno dall’altro. Se queste condizioni non sono soddisfatte, quando uno dei due interlocutori dice “0”, l’altro non sa se il messaggio era “0 e non 1” oppure “0 e non 1 e non 2”... oppure “0 e non 1 e non 2 ... e non n ”: queste situazioni sono tutte tra loro diverse e la quantità d’informazione che lo “stesso” messaggio “0” fornisce è diversa da caso a caso (cresce al crescere di n).

Questa è una differenza essenziale tra la spiegazione fisico-matematica e la spiegazione informazionale. Infatti la prima è di tipo positivo: se osservo un fenomeno F non lo metto (e in genere non riesco a metterlo) in contrapposizione con tutti gli altri fenomeni F', F'' ... che *avrei potuto* osservare. Per contro la spiegazione informazionale è di tipo *negativo* (o meglio insieme positivo e negativo), perché quando ricevo 0 lo metto in contrapposizione con *tutti* gli altri messaggi $1,2,\dots,n$ che *avrei potuto* ricevere e che non ho ricevuto. Esistono poi anche informazioni di livello superiore: se mi aspetto di poter ricevere solo due messaggi, 0 e 1, e invece ricevo il messaggio 2, non solo ricevo l’informazione “2 e non 0 e non 1”, ma ricevo anche l’informazione “meta” (cioè di ordine superiore) che non ho a che fare con un codice o alfabeto binario, bensì almeno ternario.

La teoria matematica dell’informazione fornisce una descrizione formale di concetti e nozioni che nel linguaggio comune sono usati in modo vago e non sistematico. Partendo da una misura della quantità d’informazione, la teoria costruisce modelli matematici per le *sorgenti* d’informazione e per i *canali* di trasmissione. Nel linguaggio comune informazione è sinonimo di “notizia” o di “messaggio”. In senso più tecnico, l’informazione è fornita in risposta a una domanda, esplicita o implicita, posta da un soggetto sull’esito di un’esperienza aleatoria di suo interesse attuale o potenziale. Grazie all’informazione ricevuta, il destinatario riduce, al limite annulla, l’incertezza sul risultato dell’esperienza. Poiché l’incertezza scaturisce dalla natura probabilistica dell’esperienza e dei suoi possibili risultati, il formalismo della teoria dell’informazione si basa sul calcolo delle probabilità.

L’informazione generata dalla *sorgente* S deve giungere al *destinatario* U e a questo scopo si usa un mezzo trasmissivo o *canale* C che li collega. C dovrebbe replicare presso U i dati generati da S , ma spesso sul canale agiscono dei segnali spuri (rumore), che alterano l’informazione in modo più o meno rilevante, sicché U riceve una versione distorta dell’uscita di S . Per migliorare le prestazioni del sistema di trasmissione, cioè per ridurre la frequenza degli errori di interpretazione da parte di U , si possono adottare due tattiche: la prima, tipica delle telecomunicazioni tradizionali, è un’impostazione “di forza” e consiste nel migliorare la qualità dei componenti, in particolare del canale C . Si tratta di un metodo costoso (il canale è di solito il componente più importante del sistema) e non sempre applicabile (per esempio nelle trasmissioni da satellite a terra il canale è lo spazio libero e non si può modificare). L’alternativa adottata da Shannon è un’impostazione “di finezza” e consiste nell’elaborare il flusso dei dati, inserendo tra S e C un *codificatore* Cod e tra C e U un *decodificatore* Dec (Figura 12). Si tratta di due elaboratori specializ-

FIGURA 12
Schema a blocchi di un sistema di comunicazione





zati, progettati per adattare nel modo migliore l'uno all'altro i tre blocchi *S*, *M* e *U* del sistema. Insomma, invece di intervenire sui blocchi *S*, *C*, *U* modificandoli, si interviene sui dati con un'opportuna elaborazione [3, 7, 10].

L'idea di Shannon si è dimostrata geniale. L'elaborazione da lui suggerita consiste essenzialmente nell'aggiungere alla sequenza dei simboli d'informazione che rappresentano il messaggio da trasmettere altri simboli, detti *di controllo*. La sequenza ridondante così costruita dal Cod sopporta meglio gli effetti del rumore e il Dec può ricavarne i simboli d'informazione con una fedeltà, in media, superiore. Il prezzo da pagare per questo miglioramento delle prestazioni del sistema è duplice:

1. una maggior complessità del sistema, che si manifesta con la presenza dei dispositivi Cod e Dec,
2. un rallentamento della trasmissione dell'informazione dovuto alla presenza dei simboli ridondanti di controllo.

Ma in certe condizioni abbastanza generali si dimostra che le prestazioni del sistema si possono migliorare in misura sostanziale senza che il rallentamento sia eccessivo. Tutto dipende dal rapporto tra la quantità media d'informazione generata dalla sorgente nell'unità di tempo (*entropia*) e la quantità media d'informazione che il canale può trasmettere senza ambiguità nell'unità di tempo (*capacità*). Se la capacità è maggiore dell'entropia, si possono costruire Cod e Dec in modo da conseguire un miglioramento sostanziale della fedeltà senza rallentare troppo il flusso dell'informazione.

C'è da osservare che i dati inviati sul canale non sono quasi mai quelli direttamente generati dalla sorgente, ma sono il risultato di una pre-elaborazione, detta *codifica di sorgente*, che consiste in una *compressione*: dalla successione dei messaggi generati dalla sorgente si elimina per quanto possibile la ridondanza ed è a questa successione compressa o purificata che si aggiunge poi la ridondanza utile per la protezione dagli errori di trasmissione. Queste due operazioni sembrano annullarsi a vicenda: perché non affidare la protezione dei messaggi alla ridondanza in essi già presente quando escono dalla sorgente? Il fatto è che di solito la ridondanza originaria non è adatta, poiché è tipica

delle lingue naturali (o delle immagini o dei suoni), mentre sui canali è utile avere una ridondanza *strutturata*, capace di adattarsi ai linguaggi artificiali che vi si impiegano. Il blocco codificatore ha dunque più funzioni: esegue la compressione (codifica di sorgente), traduce la successione compressa in un linguaggio artificiale, per esempio in una sequenza binaria, e in questa inietta la ridondanza protettiva (codifica di canale).

Non entriamo in ulteriori particolari della teoria di Shannon, che ci porterebbero troppo lontano. Diciamo solo che molti dei risultati più importanti della teoria dell'informazione sono di carattere esistenziale e non costruttivo; questa limitazione importante ha spinto molti ricercatori a occuparsi della costruzione effettiva di algoritmi di codifica e decodifica, dando corpo alla teoria (e alla pratica) dei codici (di sorgente e di canale), che costituisce oggi un florido capitolo della teoria dell'informazione, ormai autonomo rispetto al nucleo originario della teoria di Shannon.

9. ASPETTI SOCIOCULTURALI DELL'INFORMAZIONE

Il rapidissimo sviluppo della tecnologia informazionale, la sua facilità d'uso, la diminuzione dei costi, la potenza crescente e la diffusione capillare dei dispositivi stanno provocando una serie di importanti conseguenze quantitative e qualitative. Intanto si assiste a una vera e propria esplosione dei dati elaborati e scambiati. Inoltre la virtualità e la simulazione prendono il posto di realtà più concrete, la visione dei rapporti umani si modifica e la cultura stessa subisce profonde trasformazioni. La società assume forme aperte, flessibili e molteplici, in cui agiscono aggregazioni e disaggregazioni aleatorie e fuggevoli che prendono il posto delle robuste e durevoli strutturazioni tipiche del passato. I valori vengono sottoposti a una revisione rapida e continua. La cultura si frammenta in una complessa e multiforme ecologia che tende ad appiattire i concetti e a spezzettare le idee in un mosaico dove quasi tutte le tessere sono interessanti ma nessuna è fondamentale. La moltiplicazione delle scelte e la facilità di reperimento e di riproduzione allargano a dismisura le possibilità ma invitano anche alla



FIGURA 13
Il Prestigiatore
 di Hieronimus
 Bosch

dispersione, alla superficialità, al consumo. Spesso il curioso sostituisce l'importante e, di fronte all'ampliarsi delle opzioni, la scelta è affidata al caso e all'invasione dei persuasori occulti. L'impossibilità di dilatare il tempo (unica risorsa veramente limitata) porta a comportamenti nevrotici, ossessivi, frettolosi ed effimeri.

La flessibilità della tecnologia informazionale la rende diversa dalle altre, nel senso che, potendo essere impiegata indifferentemente in ingegneria meccanica, in fisica, in letteratura e in medicina, essa si presenta come una *meta-tecnologia*, guscio strutturato ma vuoto, suscettibile di ricevere un contenuto di volta in volta diverso. Tecnologia metodologica e catalizzante, da applicare alle altre per accelerarne le fasi progettuali e di sviluppo o per migliorarne gli aspetti di conduzione e di controllo. Questo carattere "meta" della tecnologia dell'informazione le consente di applicarsi anche a sé stessa, in un circuito retroattivo che esalta la rapidità dello sviluppo, la complessità dei dispositivi e la potenza dei linguaggi, in una spirale di cui non è facile prevedere gli esiti. Il calcolatore serve non solo a progettare macchine idrauliche e a gestire le banche di dati, ma anche a disegnare i microcircuiti e a scrivere i programmi che faranno funzionare altri calcolatori più potenti e veloci.

Inoltre le tecnologie informatiche, telematiche e di intelligenza artificiale interagiscono con una massa umana pensante e alfabetiz-

zata in forte espansione: si tratta di una *retroazione positiva* (cioè accrescitiva), i cui effetti a lungo termine non è facile prevedere. A prescindere dal tipo e dalla qualità della massa pensante che si sta sviluppando in questo gioco interattivo, prima o poi dovranno manifestarsi fenomeni di saturazione, se non di regressione, quantitativa e forse qualitativa. Oppure si presenterà un salto qualitativo verso livelli superiori di creatività e cognizione non solo nei settori scientifici, ma anche in quelli artistici, architettonici, musicali, figurativi, sui quali lo strumento informatico esercita un'influenza ancora difficile da valutare.

L'adozione delle tecnologie dell'informazione provoca crisi ricorrenti nel delicato equilibrio tra libertà personali e sicurezza. L'informatica sembra promettere a tutti l'accesso a tutte le informazioni, ma questa trasparenza illimitata è un'utopia, e non soltanto perché la divulgazione di certi dati riservati può ledere gli interessi economici o la privacy di una persona o di un gruppo, ma anche perché una circolazione completa e senza barriere dell'informazione può portare a patologie comunicative spontanee o indotte (tramite i virus) e rende comunque troppo facili le interferenze informatiche. La casistica di queste interferenze contempla furti, azioni dimostrative o burlesche, atti vandalici che portano alla distruzione di archivi e banche di dati, nonché pratiche invasive intraprese per pura e semplice curiosità (Figura 13).

Esistono agenzie specializzate nella raccolta di numeri telefonici, indirizzi, date di nascita e così via. Questi dati frammentari vengono integrati per ricavarne il profilo di clienti o acquirenti potenziali e venduti poi a catene di negozi, che se ne servono per incrementare le vendite con campagne pubblicitarie mirate e personalizzate. È evidente che queste informazioni potrebbero essere usate anche per scopi meno innocui, precludendo a inquietanti forme di controllo e di condizionamento sociale o addirittura a estremi di totalitarismo poliziesco.

Oggi chi fornisce i propri dati personali, per esempio al medico curante, corre un rischio reale: grazie all'interconnessione, i dati potrebbero cadere preda di pirati informatici, produttori di farmaci e venditori di servizi, polizia, investigatori privati e datori di lavoro.

O semplicemente di ladri. Tutti pronti a sfruttare le informazioni per controllare, ricattare, imporsi, pedinare, vendere, rubare, cioè per sfruttare la caratteristica più evidente e delicata dei dati personali, quella di essere merce di grande valore. Viviamo nella società dell'informazione, e questo significa in primo luogo che le informazioni non sono più un bene d'uso, ma un bene di scambio: mai come oggi è stato vero che "sapere è potere". Intorno all'informazione ruotano capitali cospicui e si aggrega un progressivo orientamento della società verso la richiesta di beni immateriali. Di conseguenza l'informazione, al pari dei beni più tradizionali, diviene oggetto di una diffusa attività criminosa alla quale la legislazione vigente nei vari Paesi si oppone a fatica.

Bastano questi pochi cenni per capire quanto il mondo dell'informazione, che l'informatica onnipotente sta attuando, sia lontano dal mondo della materia nel quale si sono formati in gran parte le nostre tradizioni filosofiche, scientifiche, giuridiche e sociali. La natura impalpabile dell'informazione sembra rendere superati molti concetti che nel mondo della materia hanno una provata solidità, ma poiché l'informazione interagisce con la materia, non vi è motivo per non accordare la stessa solidità a quei concetti anche sotto il profilo dell'informazione, con gli adattamenti necessari.

In un mondo sempre più virtuale e sempre più invaso da riproduzioni e da simulazioni, se è ancora difficile attribuire alle simulazioni tutto lo spessore e tutta l'efficacia che esse hanno, è altrettanto difficile distinguere la realtà dalla sua imitazione. Come è facile cadere nella tentazione di negare concretezza

all'informazione, per esempio considerando il furto dei dati meno grave del furto dei beni materiali, così è facile identificare la simulazione con la realtà: il concetto stesso di realtà, insomma, va sottoposto a un'attenta analisi critica [9].

Bibliografia

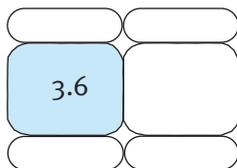
- [1] Bateson Gregory: *Verso un'ecologia della mente*. Adelphi, Milano, 1977, II edizione accresciuta, Adelphi, Milano, 2000.
- [2] Brillouin Léon: *Science and Information Theory*. Academic Press, New York, 1956.
- [3] Cover Thomas, Joy A. Thomas: *Elements of Information Theory*. Wiley & Sons, Ltd., Hoboken, 2006.
- [4] Dapor M.: *L'intelligenza della vita*. Springer, Milano, 2002.
- [5] Filippazzi, Franco: *Il computer tra fantasia e realtà: La sfida dell'intelligenza*. Tessere, CUEN, 1996.
- [6] Ghirardi GianCarlo, Francesco de Stefano: *Il mondo quantistico: una realtà ambigua*. In *Ambiguità*, a cura di Longo G. O., Magris C., Moretti e Vitali, Bergamo, 1966.
- [7] Longo Giuseppe O.: *Teoria dell'Informazione*. Boringhieri, Torino, 1980.
- [8] Longo Giuseppe O.: *Introduzione*. In *Ambiguità*, a cura di Longo G. O., Magris C., Moretti e Vitali, Bergamo, 1996.
- [9] Longo Giuseppe O.: *Il nuovo golem: come il computer modifica la nostra cultura*. Laterza, Roma-Bari, 1998.
- [10] Shannon Claude E.: A Mathematical Theory of Communication, *Bell System Tech. J.*, n. 27, 1948.
- [11] Wiener Norbert: *Cybernetics, or Control and Communication in the Animal and the Machine*. Hermann et C.ie, Paris, The MIT Press, Cambridge, Mass., Wiley and Sons, New York, 1948.

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, s'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.
longo@univ.trieste.it



PATTERN RECOGNITION TECNICHE E APPLICAZIONI

Alessio Malizia



Molti bambini, già in tenera età, sono in grado di riconoscere chiaramente lettere e numeri, scritti in corsivo o a macchina. Possiamo pensare che questa attività di riconoscimento, nota come *Pattern Recognition*, sia semplice finché non tentiamo di automatizzarla, ovvero di far riconoscere automaticamente simboli e caratteri ad un computer. L'obiettivo di questo articolo è di sintetizzare i principali metodi di *Pattern Recognition* e presentare le linee della ricerca nel settore.

1. INTRODUZIONE

Il riconoscimento automatico di oggetti (*pattern*) e la loro descrizione, classificazione e raggruppamento (*clustering*) sono argomenti importanti in una grossa varietà di problemi sia nell'area ingegneristica che scientifica, quali: la biologia, la psicologia, la medicina, il marketing, la visione artificiale, l'intelligenza artificiale fino alla fusione di dati (meglio conosciuta come *data fusion*).

La domanda che sorge spontanea nel descrivere la *pattern recognition*, è: cos'è esattamente un *pattern*?

Watanabe [18] descrive un *pattern* come "l'opposto del caos; ovvero un'entità, generalmente definita a cui si può dare un nome". Per esempio, un *pattern* può essere l'immagine di una impronta digitale, una parola scritta in corsivo, l'immagine di un volto, o un segnale acustico.

Dato un *pattern*, il suo riconoscimento (detto anche classificazione) può essere principalmente effettuato in due modi [18]:

1. supervisionato, in cui il *pattern* in input viene identificato come membro di una clas-

se tra quelle predefinite, ovvero viene classificato in maniera supervisionata avendo l'utente o il progettista definito a priori le classi di interesse;

2. non supervisionato, in cui il *pattern* in input viene assegnato ad una classe sconosciuta a priori, ovvero i *pattern* vengono raggruppati nei cosiddetti *cluster* (tecniche di *clustering*).

Il problema della *pattern recognition* quindi viene posto nella forma di classificazione o identificazione delle categorie di appartenenza, dove le classi o categorie possono essere sia definite dal progettista del sistema (nei metodi supervisionati), sia apprese dalle similarità tra i *pattern* (nei metodi non supervisionati).

I sistemi di *pattern recognition* creano, chiaramente, nuovi scenari applicativi sia a livello accademico che industriale. Si pensi, per esempio, ai vantaggi per un processo industriale che, impiegando il riconoscimento automatico di oggetti, può beneficiare di un controllo qualità automatizzato. Sempre nell'ambito industriale, l'impiego di algoritmi di

riconoscimento automatico, porta ad arricchire le funzionalità dei robot impiegati, come ad esempio la possibilità di muoversi in determinati ambienti riconoscendo automaticamente gli ostacoli. Altre applicazioni di interesse, soprattutto ai giorni nostri, sono quelle di video sorveglianza automatizzata. Pensiamo, ad esempio, ad una stazione ferroviaria o ad un aeroporto ed a come, vista la mole di persone in transito, un riconoscimento automatico di oggetti incustoditi possa essere cruciale per le funzionalità di sorveglianza. Il riconoscimento di volti, è un'altra area di applicazione che consente di aumentare notevolmente la sicurezza sia in ambito pubblico che aziendale. Un ulteriore scenario, mol-

to interessante, riguarda la cosiddetta "Realtà Aumentata" dove informazioni reali percepite da utenti umani vengono arricchite da informazioni sull'ambiente estratte automaticamente da un sistema automatizzato. Un esempio tipico è quello militare con la possibilità di avere visori che includono tracciamento automatico dei soggetti individuati e stima della distanza dall'osservatore. Già da questi primi esempi è chiaro come l'uso di tecniche di riconoscimento automatico possa notevolmente migliorare le attuali applicazioni tecnologiche in diversi campi, dal settore scientifico a quello sociale ed industriale. La tabella 1 mostra, in sintesi i principali domini applicativi dei sistemi di riconoscimento

Dominio	Applicazioni	Pattern	Classi
Analisi di documenti	Ricerca su Internet e Semantic Web	Documenti testuali / corpus di documenti digitali o digitalizzati	Categorie semantiche (esempio, affari, sport, scienza ecc.)
Analisi di immagini di documenti	Strumenti di supporto per i diversamente abili, digitalizzazione automatica di documenti	Immagini di documenti (esempio, acquisite da scanner)	Riconoscimento dei caratteri (OCR), parole, regioni di testo
Automazione Industriale	Verifica delle schede di circuiti, verifica del software	Immagini all'infrarosso o spettroscopiche, e moduli software	Rilevamento di componenti hardware e software difettose
Data Mining	Ricerca di pattern significativi in insiemi di dati/documenti	Punti nello spazio multidimensionale / vettori multidimensionali	Gruppi omogenei e ben separati (cluster)
Bioinformatica	Analisi sequenze DNA	Sequenze DNA/Proteine	Tipi di geni/proteine
Information Retrieval su basi dati multimediali	Ricerca su internet o in reti aziendali di immagini ed audio/video	VideoClip ed immagini	Generi video, tipi di immagini o tipologia di inquadrature (ed. nello sport: goal, rigore, fuorigioco ecc.)
Riconoscimento BioMetrico	Identificazione personale attraverso rilevazione dati biometrici	Immagini del volto, dell'iride, dell'impronta digitale	Gestione controllata accessi
Riconoscimento del parlato	Dettati automatici, inserimento dati per persone diversamente abili, interfaccia audio per l'uso di telefonia	Onde sonore della voce	Parole pronunciate, suoni modulati
Sensor Data Fusion	Riconoscimento della persona e della posizione geografica attraverso diversi sensori	Informazioni provenienti da sensori (esempio, accesso mediante RFID - identificazione a radio frequenza)	Posizione, risorse disponibili, contesto dell'utente

TABELLA 1

Applicazioni della pattern recognition

automatico evidenziando anche le classi di oggetti/entità coinvolte nei suddetti processi. Le principali aree di applicazione della Pattern Recognition sono:

- Automazione dei processi industriali.
 - Riconoscimento di oggetti, ispezioni visuali, controlli di qualità.
 - Navigazione e coordinazione di robot.
- Sorveglianza e tracking.
 - Traffico, sicurezza.
- Interazione uomo-computer.
 - Rilevamento e riconoscimento di volti.
 - Assistenza ai disabili.
- Realtà virtuale e visualizzazione.
 - Ricostruzione di scene e modelli 3D.
 - Fotometria.
- Applicazioni spaziali e militari.

Tra i campi più innovativi nell'applicazione del riconoscimento di *pattern*, c'è ad esempio l'*affective computing*, identificato da Picard [18], che introduce la possibilità da parte del sistema di riconoscere e esprimere emozioni, al fine di rispondere in maniera intelligente alle espressioni umane.

In genere, un sistema di *pattern recognition*, riceve in ingresso la descrizione di un oggetto, ovvero un insieme di misure che lo caratterizzano (*feature*) e sulla base di queste "decide" a quale classe l'oggetto appartiene. La decisione della classe di appartenenza ha un costo associato all'assegnazione dell'oggetto alla classe sbagliata e l'obiettivo con cui realizzare un buon sistema di *pattern recognition* è quello di minimizzare il costo di classificazione.

Comunque, la caratteristica comune a tutti questi sistemi e domini applicativi è che le misure caratteristiche (*feature*) non sono suggerite o identificate dagli esperti di dominio, ma vengono estratte ed ottimizzate da procedure che elaborano dati; questo avviene perché in genere abbiamo migliaia di misure possibili anche per un singolo *pattern*.

La crescente disponibilità di risorse per il calcolo automatico non solo consente di elaborare contemporaneamente grosse moli di dati, ma permette anche l'analisi e la classificazione di grandi insiemi di *pattern*. Allo stesso tempo, la domanda di applicazioni di *pattern recognition* cresce continuamente grazie alla disponibilità di grosse basi di dati e del calo dei costi dei supporti di memorizzazione. Tut-

tavia, in molte delle applicazioni di *pattern recognition* risulta evidente che non esiste un approccio "ottimale" e "generale" alla classificazione e che la scelta di integrare diversi metodi e approcci è la migliore da perseguire. La progettazione di un sistema di *pattern recognition* richiede essenzialmente di affrontare i seguenti aspetti:

1. Acquisizione e pre-elaborazione (e normalizzazione) dei dati.
2. Rappresentazione e classificazione dei dati o *pattern*.
3. Decisione e classificazione.

Il dominio di applicazione, in pratica, influenza enormemente la scelta dei sensori (nel senso generale del termine), delle tecniche di pre-elaborazione e di normalizzazione dei dati, della rappresentazione degli stessi e del modello decisionale di classificazione.

È noto che per i problemi di riconoscimento ben definiti e quindi condizionati opportunamente (esempio: piccole variazioni nei valori appartenenti alla stessa classe e grosse variazioni in quelli appartenenti a classi diverse) esistono soluzioni efficienti e buone strategie decisionali di classificazione. Per esempio, usare approcci orientati all'apprendimento, basandosi su un insieme di dati per l'addestramento automatico del sistema (*training set*) consente la necessaria flessibilità del sistema, che addestrato con dati diversi, può reagire a situazioni diverse con un margine di errore controllabile.

I quattro principali approcci alla *pattern recognition* [9] sono:

1. approcci basati su modello (*Template Matching*);
2. approcci con classificazione statistica (*Statistical Classification*);
3. approcci sintattici o strutturali (*Syntactic or Structural Matching*);
4. approcci basati su reti neurali (*Neural Networks*).

Questi approcci non sono necessariamente separati ed indipendenti, infatti a volte lo stesso metodo può essere usato nei diversi approcci con diverse interpretazioni. Esistono anche tentativi di progettazione e realizzazione di sistemi ibridi [5] che introducono e integrano più modelli all'interno dello stesso sistema. Nei paragrafi successivi verranno descritti e analizzati i quattro approcci menzionati al

Approccio	Rappresentazione	Riconoscimento	Criterio di Classificazione
Basato su Modello (<i>Template Matching</i>)	Dati campionati, pixel, curve introduzione di distanze	Correlazione statistica,	Basato su stima dell'Errore di Classificazione
Classificazione Statistica (<i>Statistical Classification</i>)	Misure o Caratteristiche (Feature)	Funzioni discriminanti	Basato su stima dell'Errore di Classificazione
Sintattico o Strutturale (<i>Syntactic or Structural Recognition</i>)	Primitive	Regole, Grammatiche	Basato su stima dell'Errore di Accettazione
Reti Neurali (<i>Neural Network</i>)	Dati campionati, pixel, Caratteristiche (Feature)	Funzioni sinaptiche e di rete	Basato su stima dell'Errore quadratico medio

TABELLA 2

Approcci e modelli per i sistemi di pattern recognition

fine di evidenziarne non solo le caratteristiche ma valutarne i vantaggi e gli svantaggi alla luce dei diversi domini applicativi. Nella tabella 2 vengono presentati i diversi approcci con una descrizione sintetica delle loro caratteristiche.

2. SISTEMI DI PATTERN RECOGNITION BASATI SUL MODELLO

Uno dei primi e più semplici degli approcci alla pattern recognition è quello basato sul modello, il cosiddetto *template matching*. I sistemi di *pattern recognition* basati sul modello utilizzano la cosiddetta tecnica del "matching"; il *matching* è definito come l'operazione che viene usata per determinare la similarità tra due entità dello stesso tipo: punti, curve o forme ad esempio. Nell'approccio basato su modello, quest'ultimo (il modello) è tipicamente un prototipo del *pattern* od oggetto da riconoscere e classificare, tipicamente una forma bidimensionale. L'oggetto da riconoscere (il *pattern*) viene confrontato con tutti i modelli memorizzati nel sistema considerando i cambiamenti di scala e le possibili invarianze per rotazione o traslazione. Naturalmente ci saranno casi in cui sarà utile l'invarianza di scala, cioè ci interessa riconoscere l'oggetto semplicemente dalla forma, mentre in altri casi sarà necessario distinguere non solo rispetto alla forma ma anche rispetto alla dimensione. Nella figura 1 è riportato un esempio di sistema basato su modello [1] dove si mostra un *pattern*, l'aereo in input (A), e l'estrazione della forma (B). In

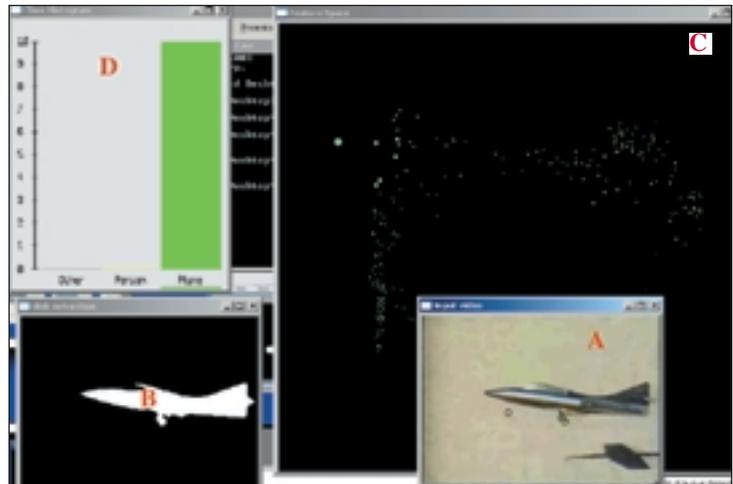


FIGURA 1

Esempio di sistema di pattern recognition basato su modello: A pattern in input (aereo), B estrazione forma, C base dati di addestramento (training set), D istogramma di classificazione

figura è presente anche una finestra sullo sfondo (C) con dei punti rappresentanti i valori delle caratteristiche per i modelli di aereo presenti nella base dati; i punti di dimensione più grande rappresentano i modelli che sono più simili all'aereo in input (risultato della fase di *matching*). L'istogramma di colore verde (D), mostra come l'oggetto sia classificato "aereo" dal sistema, perché la maggior parte (tutti in questo caso) dei punti della base dati di oggetti nel suo intorno erano stati classificati (in fase di addestramento) come aerei (colore verde).

La misura di similarità utilizzata nei sistemi basati su modello, è solitamente una correlazione statistica e può essere ottimizzata apprendendo i parametri dal *training set*,

ovvero da un insieme di esempi di oggetti forniti in input al sistema. I sistemi basati su modello sono computazionalmente inefficienti, richiedono molte risorse di calcolo per confrontare il modello del pattern in input con tutti gli altri presenti nella base dati, ma grazie alla continua crescita della potenza di calcolo questo approccio viene sempre più utilizzato soprattutto in ambito industriale. Per come abbiamo definito il modello fino ad ora si dovrebbe parlare di sistemi basati su modelli rigidi, ovvero sistemi dove il modello è descritto con parametri ed insiemi di valori fissati. Questo tipo di sistemi tende a non essere efficiente in presenza di distorsioni dovute all'acquisizione dell'oggetto (per esempio: effetti di errori sulle immagini), cambiamento dei punti di vista, o grosse variazioni di valori delle caratteristiche all'interno della stessa classe (*inter-class variance*). In questi ultimi casi si tende ad utilizzare tecniche di modelli deformabili o elastici [8] al fine di consentire un confronto di similarità (*matching*) tra modelli le cui deformazioni non possono essere rappresentate esplicitamente a priori. Per intenderci, questi sistemi sono utili quando non possiamo modellare a priori tutte le possibili deformazioni che un oggetto possa avere, ed allora si agisce creando un modello dell'oggetto e studiando quali agenti esterni lo possono deformare (modellando quin-

di le possibili deformazioni) a volte simulando anche forze fisiche.

3. SISTEMI DI PATTERN RECOGNITION CON CLASSIFICAZIONE STATISTICA

Nell'approccio statistico ogni pattern è rappresentato in termini di un vettore di lunghezza d , dove d indica il numero di misure caratteristiche o feature dello stesso e viene rappresentato come un punto nello spazio d -dimensionale delle caratteristiche (*feature o vector space*). L'obiettivo, quindi, è quello di selezionare un insieme di feature che consenta a vettori appartenenti a categorie differenti di occupare spazi compatti in regioni disgiunte dello spazio d -dimensionale delle feature. Nell'esempio di figura 2 [3] è rappresentato uno spazio con due caratteristiche, la dimensione e la luminosità; queste due caratteristiche sono usate in questo esempio per distinguere immagini di pesci, più esattamente tra due classi di pesci: i salmoni e le spigole. Si può notare come i valori per ogni pesce nello spazio delle caratteristiche si distribuiscono in due regioni, legate alla dimensione ed alla luminosità della livrea dei pesci.

L'efficacia della rappresentazione spaziale delle feature (*feature set*) è rappresentata da come i pattern delle diverse classi sono ben separati nello spazio. Dato un insieme di addestramento (esempi di pattern) per ogni classe, l'obiettivo è quello di stabilire delle regioni di decisione nello spazio delle feature che consenta di separare i pattern appartenenti a classi diverse; nell'esempio di figura 2 la regione di decisione è data dalla retta che separa (con un piccolo margine di errore) le due classi di pesci.

Nell'approccio statistico, le regioni di decisione sono determinate dalle distribuzioni di probabilità per ogni pattern di appartenere ad una delle classi; le distribuzioni vanno quindi specificate (anche parametricamente) oppure apprese mediante addestramento del sistema (*machine learning*) [2, 4].

Uno degli approcci più utilizzati è quello dell'analisi dei discriminanti per la classificazione:

□ prima di tutto si sceglie, sulla base dell'esperienza una forma parametrica per la re-

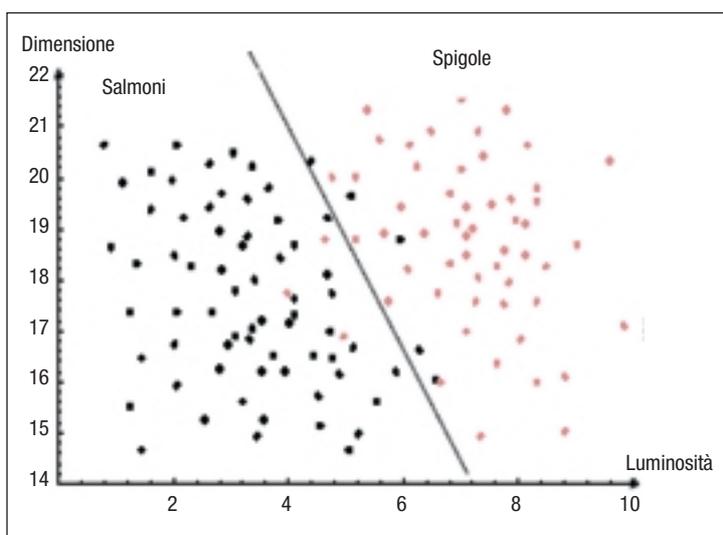


FIGURA 2
Spazio bidimensionale delle caratteristiche per due classi di pesci, i salmoni e le spigole

gione di decisione, per esempio lineare (figura 2) o quadratica;

□ successivamente si sceglie, sulla base dell'insieme di addestramento, la migliore regione di decisione ai fini della classificazione.

In genere si usano metodi statistici per la stima degli errori per determinare le regioni di decisione. Il motivo di questa scelta si basa sul fatto che la migliore regione di decisione è quella che minimizza l'errore nell'assegnare il pattern alla classe sbagliata.

Uno degli altri principali approcci alla costruzione delle regioni di decisione, l'approccio diretto, nasce dalla cosiddetta filosofia di Vapnik [17]. Questa si basa sull'idea che se si possiede un insieme ridotto di informazioni sul problema che si sta affrontando, è meglio cercare di risolvere il problema direttamente, senza affrontare generalizzazioni dello stesso come passo intermedio. È possibile, infatti, che in questo caso si posseggano informazioni sufficienti solo per una *soluzione diretta* dello stesso, ma che le stesse siano insufficienti per una soluzione più generale.

4. SISTEMI DI PATTERN RECOGNITION CON APPROCCI SINTATTICI O STRUTTURALI

In molti problemi di *pattern recognition* che coinvolgono oggetti complessi, una soluzione può essere quella di considerare una prospettiva gerarchica, dove i pattern vengono visti come composti da subpattern più semplici [6, 14].

I *pattern elementari* o *subpattern* sono chiamati *primitivi* e i pattern complessi da riconoscere sono rappresentati sotto forma di insiemi e relazioni tra primitivi. Nei sistemi di *pattern recognition* sintattici si crea un'analogia formale fra la struttura dei pattern complessi e la sintassi di un linguaggio. I pattern vengono visti come frasi di un linguaggio, mentre le primitive vengono rappresentate come l'alfabeto del linguaggio, e le frasi quindi sono generate a partire da una grammatica per questo linguaggio. Quindi un pattern viene riconosciuto o classificato se la rispettiva frase appartiene al linguaggio specificato per quella classe ed è quindi accettata dalla grammatica specificata. Quindi un grosso insieme di pattern

complessi può essere descritto da un semplice insieme di primitive e regole grammaticali. La grammatica per ogni classe di pattern può essere inferita o appresa dall'insieme di pattern di addestramento o di esempio per ognuna delle classi di interesse.

La *pattern recognition* strutturale (o sintattica) è molto interessante perché, oltre alla classificazione dei pattern, fornisce anche una descrizione di come essi siano costruiti a partire dalle primitive, da qui l'uso della parola strutturale.

Questo approccio è molto usato in situazioni in cui la struttura dei pattern può essere catturata da insiemi di regole, come per le forme d'onda degli elettrocardiogrammi, le *texture* o trame (rappresentazioni fotorealistiche applicate ad un modello sintetico d'immagine) e l'analisi dei contorni [6].

L'implementazione di un approccio sintattico alla *pattern recognition* è in generale piuttosto complessa e questo è dovuto per esempio al problema della **segmentazione delle immagini** al fine di estrarre le primitive ed all'inferenza per estrarre la grammatica dall'insieme di pattern di addestramento. Negli anni ottanta, Fu [6] introdusse la nozione di grammatica con attributi che unifica i concetti di *pattern recognition* sintattico con quello basato su classificazione sta-

Segmentazione di immagini

La segmentazione è un partizionamento di immagini digitali in più regioni (insiemi di pixel) in base ad un certo criterio stabilito. L'obiettivo della segmentazione è tipicamente quello di individuare gli oggetti di interesse che possano essere estratti dall'immagine originale. Tipicamente la maggior parte degli algoritmi di segmentazione sono troppo semplici per risolvere completamente questo problema; la segmentazione può essere vista anche come un tipico problema di visione artificiale (o computer vision).

Un semplice ma tipico esempio può essere quello di usare una soglia fissa per immagini a toni di grigio. Fissato un valore di grigio soglia t : ogni pixel p è assegnato ad una tra le due classi (o regioni), P_0 o P_1 , a seconda se $L(p)$, che è il valore di grigio (tipicamente tra 0 corrispondente al nero, 255 al bianco, e la gamma di valori intermedi alle tonalità di grigio) è tale che $L(p) < t$ o $L(p) \geq t$.



Immagine originale



Immagine dopo l'operazione di soglia

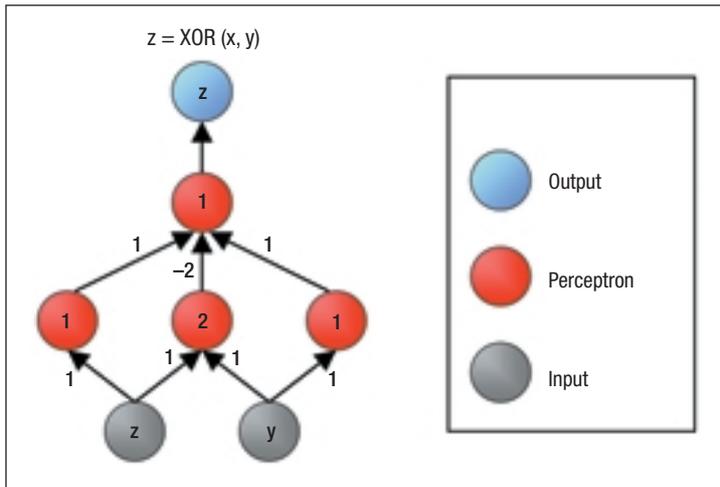


FIGURA 3
Una rete neurale del tipo multi-layer perceptron a due livelli in grado di calcolare la funzione XOR di due variabili

tistica. In effetti, un problema dell'approccio sintattico è che può generare una esplosione di combinazioni da investigare per inferire dalle primitive e dalle loro relazioni la grammatica conseguente; per cui necessita di grandi insiemi di pattern per l'apprendimento e quindi grosse risorse di elaborazione dati [15].

5. SISTEMI DI PATTERN RECOGNITION BASATI SU RETI NEURALI

Le reti neurali possono essere considerate come un sistema computazionale fortemente parallelo costituito da un grande numero di processori estremamente semplici ma intensamente interconnessi fra loro.

I modelli basati su reti neurali cercano di utilizzare principi di organizzazione dei dati, come: addestramento, generalizzazione, tolleranza ai guasti e distribuzione dei dati, in una rete di grafi pesati; i nodi sono rappresentati da neuroni artificiali e gli archi diretti (pesati) rappresentano le connessioni tra l'output di un livello di neuroni e l'input del successivo livello, come è mostrato in figura 3 [19].

La caratteristica principale delle reti neurali è la loro capacità di addestrarsi, mediante procedure iterative, e gestire relazioni non lineari fra l'input e l'output, adattandosi ai pattern di esempio.

Le reti neurali vengono divise in famiglie a seconda della loro struttura e del tipo di connessioni tra i neuroni artificiali; la famiglia di reti neurali più utilizzata per i sistemi di *pattern recognition*, e quindi per la classificazione [10], è quella basata sulle cosiddette *feed-forward network*, che includono tra gli altri i *Multi-Layer Perceptron* (MLP) e le *Radial-Basis Functions* (RBF). Queste reti sono organizzate per livelli e hanno connessioni unidirezionali tra i neuroni appartenenti ai differenti livelli.

Un altro insieme, molto usato, di reti neurali è costituito dalle *Self-Organizing Map* (SOM), o anche reti di Kohonen [11], che sono principalmente usate per il raggruppamento automatico di pattern, chiamato *data cluster*.

Il processo di addestramento nelle reti neurali, consiste nell'aggiornare i pesi di connessione sugli archi dei neuroni artificiali al fine di migliorare la precisione dei risultati della rete a seconda del compito assegnato, di classificazione o di raggruppamento (*clustering*).

Il crescente utilizzo dei modelli per la *pattern recognition* basati su reti neurali è dovuto principalmente alla caratteristica per cui sembra che le reti neurali siano indipendenti dal dominio applicativo del sistema di *pattern recognition*, in contrasto con quanto invece avviene per i sistemi basati su modello o con approccio sintattico/strutturale. In più, per le reti neurali sono disponibili algoritmi efficienti per l'apprendimento da insiemi di esempio dei pattern da classificare.

In definitiva le reti neurali forniscono un insieme di algoritmi non lineari per l'estrazione delle feature dai pattern (mediante l'utilizzo del livello nascosto o *hidden layer*) e per la classificazione dei pattern (per esempio i *multi-layer perceptron*).

In più, gli algoritmi di estrazione delle feature e di classificazione dei pattern possono essere direttamente applicati (matematicamente) su realizzazioni hardware delle reti neurali ai fini di incrementarne l'efficienza.

Diversamente da quello che sembra, molti dei più noti modelli di reti neurali sono implicitamente equivalenti o molto simili ai metodi di *pattern recognition* basati su modelli statistici, come si può vedere dallo schema di un generico modello statistico

presentato in figura 4. Ma indipendentemente dalle similitudini, le reti neurali offrono molti vantaggi ai progettisti di sistemi di *pattern recognition*, quali: un approccio unificato alla selezione ed estrazione delle feature, ed una strategia generale e flessibile per trovare soluzioni ai problemi di classificazione in cui le regioni di decisione siano non lineari.

6. SISTEMI DI RICONOSCIMENTO AUTOMATICO PER BASI DATI MULTIMEDIALI

Approfondiremo, ora, i sistemi di riconoscimento basati su immagini [13] (che utilizzano, in genere, un mix delle tecniche presentate), visto che al giorno d'oggi viene prodotta una sempre maggiore quantità di documenti multimediali (immagini, filmati ecc.), per lo più memorizzati in forma digitale. L'enorme diffusione dei dati multimediali in rete negli ultimi anni ha portato alla crescita di basi dati sia di immagini che di sequenze video. Tuttavia, l'uso efficace di immagini in vari domini applicativi è ostacolato dalla difficoltà di classificare e gestire questi dati in modo efficiente. Sebbene la "navigazione" di sequenze di immagini sia un tema importante in molte applicazioni nelle quali si richiede all'utente di selezionare poche sequenze tra molte, o quando deve trovare un punto particolare all'interno di una sequenza, gli strumenti a disposizione per eseguire tali attività rimangono piuttosto primitivi.

Per progettare un sistema per l'interrogazione di basi dati di immagini è necessario stabilire:

1. quale rappresentazione adottare per descrivere il contenuto delle immagini,
2. in quale modo organizzare le informazioni estratte per poi memorizzarle,
3. che tipo di strategie utilizzare per il recupero e la visualizzazione delle immagini memorizzate.

Nella selezione del tipo di rappresentazione delle immagini, si deve tentare di trovare un insieme di caratteristiche che descrivano nella maniera più completa possibile le immagini presenti nella base dati. Molte delle tecniche adottate per rappresentare il contenuto delle immagini si basano sulla conoscenza, cioè utilizzano informazioni note a priori sul

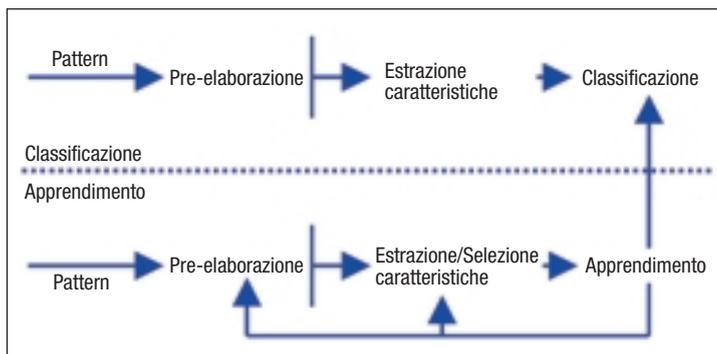


FIGURA 4

Schema generale di un sistema di pattern recognition basato su classificazione statistica. Come si può vedere dalle fasi rappresentate in figura, lo schema descrive anche in generale i sistemi basati su reti neurali

dominio specifico a cui è rivolta l'applicazione, non facilmente utilizzabili per il recupero di immagini in basi dati generiche. In genere gli oggetti possono essere caratterizzati mediante: la loro forma (contorno, baricentro ecc.), le regioni omogenee (segmentazione), o caratteristiche geometriche di vario tipo (esempio, scheletro – tecnica che consiste nel rimuovere ripetutamente i pixel del contorno fino a ridurre l'oggetto a uno "scheletro" largo un pixel) [12].

La memorizzazione del contenuto delle immagini è organizzata mediante indici per rendere più efficiente la successiva fase di recupero. La capacità di recupero delle informazioni in una base dati multimediale è strettamente legata anche al linguaggio per le interrogazioni fornito con essa. Esistono svariate forme di interrogazione su immagini, come ad esempio le interrogazioni tramite identificatori, le interrogazioni testuali, le interrogazioni basate su regioni (viene scelta una regione dell'immagine e tutte quelle che la intersecano vengono selezionate) ecc..

Infine per ottenere i risultati (le immagini), gli indici delle immagini memorizzate sono confrontati con i parametri estratti dalla descrizione dell'immagine fornita in input dall'utente. La rappresentazione visuale dei risultati, quindi, consente all'utente di valutare i risultati ottenuti e di variare gli intervalli entro i quali devono essere compresi i parametri legati ai modelli degli oggetti recuperati, in modo da ottenere risultati più o meno precisi a seconda del grado di accuratezza desiderato.

Per quanto riguarda i video, la descrizione delle entità "sequenze video" dipende dal contesto del filmato in esame e dall'ambito cognitivo in cui viene utilizzata. Per questi motivi, la fase di indicizzazione dei documenti deve sfruttare il contenuto della sequenza video in base alle caratteristiche tipiche del dominio applicativo. Per esempio, può essere interessante effettuare ricerche per soggetto (singolo attore, particolare scenografia), tipo di inquadratura (primo piano, quadro d'insieme), tipo di azione (entrata in scena, uscita dalla scena, coro d'insieme).

Per facilitare la navigazione e il reperimento di sequenze video digitali, è utile descrivere i filmati in termini di dati strutturati e riconoscere le interruzioni logiche al loro interno. Tale tecnica è detta *shot boundary detection* e il suo compito è quello di segmentare lunghe sequenze, in modo automatico, sfruttando una analisi basata sul contenuto. Il rilevamento e la classificazione di interruzioni di scene è un primo passo nel processo di annotazione automatica di sequenze video digitali. Una volta eseguita la segmentazione, il video originale può essere pensato come suddiviso in porzioni internamente uniformi e ininterrotte, sulle quali è possibile effettuare ricerche più specifiche.

7. CONCLUSIONI E FUTURI SVILUPPI

Nella prefazione del libro *Pattern Recognition* del 1974, Vapnik e Chervonenkis scrissero (l'originale è in russo): "...Per costruire la teoria della *pattern recognition*, prima di tutto è necessario formalizzare completamente il problema del riconoscimento automatico di oggetti (...) e questo risultato sembra essere difficile da raggiungere."

Come abbiamo visto nei paragrafi precedenti le tecniche di *pattern recognition* sono molto dipendenti dal dominio, ed anche le reti neurali che sembrano un approccio più generale in realtà dipendono dall'insieme di apprendimento (training set) e dai parametri degli algoritmi di back-propagation usati nell'addestramento (per esempio: il learning rate).

Essenzialmente, i differenti punti di vista sulla formulazione del problema della *pattern recognition* sono rappresentati dalla risposta

ad una domanda: esistono principi generali adeguati per descrivere classi di pattern di diversa natura, oppure lo sviluppo di un linguaggio di descrizione dei pattern è un problema che va affrontato dai singoli esperti per ogni dominio in ogni singolo caso?

Se la prima parte della domanda fosse vera, allora la scoperta di questi principi dovrebbe essere la principale direzione della ricerca nella *pattern recognition*. [7].

Altrimenti la *pattern recognition* si ridurrebbe a minimizzare l'errore nell'assegnare il pattern alla classe sbagliata, o *risk minimization*; questo implicherebbe la definizione di regioni di decisione specifiche per ogni classe di pattern e diventerebbe un problema di statistica applicata.

Finora le risposte alla domanda fondamentale non sono state trovate, e questo è il motivo per cui formulare un modello di *pattern recognition* è ancora un problema arduo. La maggior parte dei ricercatori ha comunque adottato il secondo punto di vista e la teoria della *pattern recognition* è attualmente legata strettamente alla teoria della *risk minimization* con una classe di regioni di decisioni dipendenti dal dominio.

Alla fine degli anni cinquanta, i primi ricercatori sulla *pattern recognition* cercavano di formulare nuovi modelli e non di ridurre i problemi a schemi matematici o statistici noti. Infatti in quest'ottica quello che attualmente accade, e cioè di ridurre i problemi di *pattern recognition* a quelli di *risk minimization*, ha sollevato non poco disappunto nelle attese dei ricercatori. A quell'epoca però molti dividevano la sensazione che si potesse trovare qualcosa di completamente nuovo per modellare i problemi di *pattern recognition*.

Attualmente, dopo più di quarant'anni, è chiaramente naturale chiedersi se l'importanza della modellazione formale nella rappresentazione dei pattern è stata pienamente compresa. Evidentemente la risposta è no, poiché l'atteggiamento di riutilizzare schemi matematici noti continua a prevalere.

Naturalmente va detto che non c'è stato un grandissimo sforzo in questo senso per arrivare ad un formalismo unificante. Negli anni '70 ed '80, per esempio, i nuovi approcci sintattici e strutturali alla *pattern recognition* diedero un ruolo centrale alla rappresenta-

zione non numerica dei pattern, e ci si aspettava che il futuro del campo fosse connesso con l'integrazione degli approcci strutturali con quelli classici, basati sui *vector spaces* (spazi di vettori di *feature*). Ma questo non è ancora avvenuto, o meglio non sembra sufficientemente unificante.

Siamo in realtà ad una svolta cruciale nel campo della *pattern recognition* in quest'ultimo periodo; infatti, c'è la necessità di produrre inquadramenti formali che generalizzino l'approccio in modo da evidenziare la natura intrinsecamente "intelligente" dell'attività di *pattern recognition*. Basandoci su questo vanno definiti i concetti chiave su cui poggiare queste infrastrutture formali o framework. Come delineato nel workshop satellite dell'International Conference on Pattern recognition del 2004, dal titolo *Pattern representation and the future of pattern recognition*, va compreso che l'investimento di ingenti risorse nello sviluppo di nuove tecniche basate sul vecchio approccio (riduzione a statistica applicata) è irrazionale comparato a quello necessario per costruire un impianto formale adeguato. Infatti, è da sottolineare il fatto che la Scienza è orientata al futuro e quindi alla scoperta e costruzione degli scenari generali o "big picture" e non alla risoluzione indipendente di problemi, per quanto essi siano complicati.

Gli sviluppi futuri, quindi, si stanno orientando verso la *inductive class representation* che, descritta in poche parole, è un approccio sia generativo che induttivo alla pattern recognition. Il fatto di essere un approccio generativo implica che ha la capacità di "generare" oggetti dalle classi basandosi sulla rappresentazione formale delle stesse. L'induzione nell'approccio è intesa come la possibilità che la rappresentazione delle classi possa essere dedotta da un piccolo insieme di esempi (*training set*) e che tale rappresentazione sia stabile rispetto ai vari tipi di "rumore" ed errori presenti a seconda della rappresentazione scelta.

I ricercatori propongono attualmente un approccio chiamato ETS, *Evolving Transformation System*. L'ETS, differentemente dai classici modelli computazionali (per esempio, linguaggi e grammatiche formali) evidenzia la natura evolutiva delle classi di oggetti e

quindi della loro rappresentazione. Queste strutture matematiche permettono di rappresentare dinamicamente, durante il processo di apprendimento, la struttura composta degli oggetti all'interno di un ambiente induttivo o evolutivo. In altre parole, il modello è basato su un nuovo tipo di struttura per la rappresentazione (induttiva) degli oggetti, chiamata *struct* che è definita come un elemento di uno spazio generalizzato di misure, su cui la misura di similarità (tra oggetti) è rappresentata da un insieme (ottimo) finito di trasformazioni strutturali pesate (per esempio, come le operazioni di inserimento sensibili al contesto) costruite durante il processo di apprendimento.

Si può infine notare che questa scelta sta aprendo realmente nuove strade per la ricerca dei prossimi anni, poiché il concetto di rappresentazione evolutiva di classe non può essere integralmente introdotto nei tipici formalismi per via delle loro limitazioni. Per esempio nel formalismo numerico, la rappresentazione di una classe non è generativo, mentre nei linguaggi formali (comprese le grammatiche su grafi), il concetto di classe non è induttivo; questo discende dal fatto che quest'ultimo è un formalismo computazionale (logico) e non basato sulla rappresentazione.

Per concludere segnaliamo la presenza, sia in Italia che nel mondo, di molti centri di ricerca, sia accademici che industriali, fortemente orientati alle applicazioni di riconoscimento automatico. Tra i principali attori in ambito accademico internazionale possiamo citare: il Machine Intelligence Laboratory dell'Università di Cambridge (UK), il Computer Vision Lab dell'Università Linköping (Svezia), il centro di ricerca LORIA (Francia), l'Intelligent Systems Lab di Amsterdam (Olanda), il MIT Media Lab (USA), il Rensselaer Polytechnic Institute (USA), il Computer Vision Lab dell'Università del Maryland (USA), il gruppo di Pattern Recognition and Image Processing dell'Università del Michigan, MSU (USA), il Pattern Recognition Lab della Lehigh University (USA).

A livello industriale i principali centri di ricerca sono: il PARC (Palo Alto Research Center, USA), il Vision Technology Group della Microsoft (Cambridge, UK), l'Exploratory Computer Vision Group di IBM (USA).

Nel panorama nazionale sono presenti molti gruppi accademici di fama internazionale distribuiti su tutto il territorio dal nord al sud, isole comprese. Esistono, inoltre, numerose riviste internazionali dedicate a queste tematiche, tra le principali menzioniamo: *The Journal of the Pattern Recognition Society* edito dalla Elsevier, e *Transaction on Pattern Analysis and Machine Intelligence* edito dalla IEEE. Le conferenze di riferimento sono l'IC-PR, ovvero *International Conference on Pattern Recognition* che si svolge ogni due anni a livello internazionale; mentre a livello nazionale (si svolge sempre in Italia ma con respiro internazionale) è ormai ben conosciuta l'ICIAP, *International Conference on Image Analysis and Processing*, che include molte tematiche di Pattern Recognition.

Bibliografia.

- [1] Biagiola L.: *Riconoscimento automatico di soggetti in movimento*. Tesi di Laurea, Dipartimento di Informatica, Università "La Sapienza" di Roma, Italia, A.A. 2002-2003.
- [2] Devroye L., Györfi L., Lugosi G.: *A Probabilistic theory of Pattern Recognition*. Berlin: Springer-Verlag, 1996.
- [3] Duda R.O., Hart P., Stork D.G.: *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons, 2000.
- [4] Duda R.O., Hart P.: *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons, 1973.
- [5] Fu K.S.: A step toward unification of Syntactic and Statistical Pattern Recognition. *IEEE Transaction on Pattern Recognition and Machine Intelligence*, Vol.5, n. 2, 1983, p. 200-205.
- [6] Fu K.S.: *Syntactic Pattern Recognition and Applications*. Englewood Cliffs, N.J., Prentice-Hall, 1982.
- [7] Lev Goldfarb, David Gay, Oleg Golubitsky, Dmitry Korkin: *What is a structural representation? Second variation*. Faculty of Computer Science, U.N.B., Technical Report TR04-165, March 2004.
- [8] Grenander U.: *General Pattern Theory*. Oxford University Press, 1993.
- [9] Jain A., Duin R., Mao J.: Statistical Pattern Recognition: A Review. *IEEE Transaction on Pattern Recognition and Machine Intelligence*, Vol. 22, n. 1, 2000, p. 4-37.
- [10] Jain A., Mao J., Mohiuddin M.K.: *Artificial Neural Networks: A Tutorial*. Computer, Mar. 1996, p. 31-44.
- [11] Kohonen T.: *Self Organizing Maps*. Springer Series in Information Sciences, Vol. 30, Berlin, 1995.
- [12] El-Kwae E.A., Kabuka M.R.: Efficient content-based indexing of large image databases. *ACM Trans. Inf. Syst.*, Vol. 18, n. 2, Apr. 2000, p. 171-210.
- [13] Ngu A.H., Sheng Q., Huynh D., Lei R.: Combining multi-visual features for efficient indexing in a large image database. *The VLDB Journal*, Vol. 9, n. 4, Apr. 2001, p. 279-293.
- [14] Pavlidis T.: *Structural Pattern Recognition*. New York, Springer-Verlag, 1977.
- [15] Perlovsky L.I.: Conundrum of Combinatorial Complexity. *IEEE Transaction on Pattern Recognition and Machine Intelligence*, Vol. 20, n. 6, 1998, p. 666-670.
- [16] Picard R.: *Affective Computing*. MIT Press, 1997.
- [17] Vapnik V.N.: *Statistical Learning Theory*. New York: John Wiley & Sons, 1998.
- [18] Watanabe S.: *Pattern Recognition: Human and Mechanical*. New York: Wiley, 1985.
- [19] Wikipedia - the free encyclopedia, <http://www.wikipedia.org>.

ALESSIO MALIZIA si è laureato in Scienze dell'Informazione presso l'Università "La Sapienza" di Roma nel 1997, dove attualmente svolge la sua attività di ricerca. La sua ricerca comprende il riconoscimento automatico delle forme, le biblioteche digitali, e la visione artificiale. Attualmente lavora su Interazione uomo-macchina, modelli collaborativi, ed interazione multimodale. Nella sua passata esperienza ha lavorato per IBM e Silicon Graphics nell'ambito delle applicazioni multimediali. È stato visiting researcher presso lo XEROX PARC (Palo Alto Research Center) dove ha lavorato nel gruppo di Human-Document Interaction.
E-mail: malizia@di.uniroma1.it



ICT E INNOVAZIONE D'IMPRESA

Casi di successo

Rubrica a cura di

Roberto Bellini, Chiara Francalanci

La rubrica *ICT e Innovazione d'Impresa* vuole promuovere la diffusione di una maggiore sensibilità sul contributo che le tecnologie ICT possono fornire a livello di innovazione di prodotto, di innovazione di processo e di innovazione di management. La rubrica è dedicata all'analisi e all'approfondimento sistematico di singoli casi in cui l'innovazione ICT ha avuto un ruolo critico rispetto al successo nel business, se si tratta di un'impresa, o al miglioramento radicale del livello di servizio e di diffusione di servizi, se si tratta di una organizzazione pubblica.

Il caso HotelSolutions

Stefano Modafferi

1. INTRODUZIONE

Uno degli scenari più maturi per l'applicazione di strumenti ICT è quello del turismo. I portali web che se ne occupano sono innumerevoli, ma anche i singoli operatori ormai si dotano di strumenti per essere raggiungibili attraverso i nuovi canali di comunicazione (internet su tutti). La maturità degli strumenti ICT usati per il B2C nel turismo è ormai elevata, restano invece ampi spazi per sviluppare strumenti ICT per il B2B nel turismo. Con il proliferare dei portali turistici online si sta infatti sviluppando un livello intermedio fra le aziende che gestiscono i portali e gli operatori del settore turistico.

Questo livello intermedio diventa il canale di comunicazione fra l'operatore e quello che è ormai il più diffuso venditore del suo prodotto. La realtà presentata in questo articolo è quella delle prenotazioni alberghiere. Da una parte internet e i portali online hanno radicalmente cambiato il modo con cui un utente prenota un albergo, dall'altra strumenti di analisi dei dati sofisticati consentono di fare stime e previsioni dell'occupazione che consentono di massimizzare il profitto. In questo scenario le teorie sviluppate intorno agli anni '90 sullo **Yield Management** [3] sono state rivisitate e reinterpretate in maniera proficua per tenere conto dei nuovi

YIELD MANAGEMENT PROCESS

I processi chiamati Yield Management, anche conosciuti come revenue process, sono i processi attraverso i quali è possibile capire, anticipare e reagire al comportamento del cliente/consumatore per massimizzare il proprio guadagno [3].

Nel settore alberghiero lo Yield Management è una disciplina economica nelle quali la durata del soggiorno, il controllo dei prezzi, la tipologia di camera, la partizione del tipo di clientela in fasce di mercato vengono combinati con analisi statistiche al fine di espandere il mercato coperto dall'azienda ed incrementare il guadagno per camera.

Gli Yield Management process sono di solito composti da una parte automatica utilizzata per aggregare un insieme di dati proveniente da svariate fonti e da una parte manuale o semiautomatica che costituisce l'analisi di questi dati aggregati.

Nonostante la parte automatica dei sistemi di Yield management sia fondamentale e molta letteratura si concentri su tale parte, questi processi non sono un mero sistema automatico o una serie di tecniche da applicare pedissequamente. In realtà, infatti, uno Yield Management system è caratterizzato da una "filosofia del business" ed una "metodologia applicativa" che vanno al di là degli specifici strumenti utilizzati per aggregare dati [2].

Esistono diverse euristiche per il calcolo della previsione. Sono sostanzialmente legate al dominio applicativo e alla dimensione del soggetto da analizzare. [1] presenta un'analisi di diversi algoritmi.

scenari e dei nuovi mezzi tecnici a disposizione. Lo Yield Management anche se può sembrare una disciplina molto vicina al data mining, presenta alcuni aspetti di analisi dei dati che vengono estratti automaticamente che lo avvicinano molto al Knowledge Management. In questa accezione il Knowledge Management non è tanto una serie di processi utilizzati per codificare e rendere condivisibile la conoscenza, quanto un processo che grazie ad un'interpretazione esperta dei dati trasforma l'informazione che essi presentano in conoscenza utilizzabile per produrre servizi ad alto valore aggiunto.

Il presente articolo è dedicato alla presentazione di una giovane azienda, la HotelSolutions S.R.L. che ha deciso di porre il Knowledge Management alla base del suo business dedicandosi alla produzione di pacchetti software e servizi per il mondo alberghiero. La società è stata fondata nel Marzo del 2005. I soci fondatori P. Catoni, M. Beolchini, V. Maggio e U. Piero hanno notevole esperienza lavorativa nel settore alberghiero e nell'e-commerce.

L'attività primaria dell'azienda è quella di costruire modelli previsionali per l'occupazione alberghiera con un orizzonte temporale di sei mesi. Tali modelli vengono utilizzati per massimizzare il ricavo attraverso opportune politiche di prezzo. Accanto a questa attività l'azienda fornisce supporto generale alla gestione automatica delle prenotazioni on-line sia sul sito dell'hotel che soprattutto sui portali Internet che oggi costituiscono un irrinunciabile strumento da mettere a disposizione dei clienti per poter essere raggiungibili.

La filosofia della HotelSolutions è quella di offrire alle aziende alberghiere soluzioni il più possibile flessibili. Per questo motivo vengono proposti sia prodotti che servizi. Nel primo caso vengono venduti dei pacchetti software, ma tutta l'operatività è lasciata all'albergo, nel secondo, invece, viene fornito un servizio per cui tutta l'operatività è gestita dalla stessa HotelSolutions.

2. LA REALTÀ ALBERGHIERA ITALIANA E L'IDEA HOTEL SOLUTIONS

Come spiegato anche nel riquadro, i benefici di una politica dei prezzi sono stati largamente studiati negli anni '90. Queste politiche sono state prima applicate al settore del trasporto aereo e successivamente a quello alberghiero. L'incremento esponenziale dei portali per il turismo, ha portato alla

necessità di istituire ruoli che si occupassero specificamente della gestione dei prezzi di un hotel, sia da un punto di vista strategico, cioè decidere concretamente il prezzo, che da un punto di vista pratico, cioè caricare i prezzi e le disponibilità delle camere sul maggior numero possibile di *Global Distribution Service* (GDS), cioè sui portali web e su altri strumenti specifici utilizzati per esempio dalle compagnie aeree per associare l'offerta degli alberghi al volo. È nata, o meglio si è affermata, quindi la figura del *Revenue Manager*.

La realtà delle aziende alberghiere a livello mondiale è caratterizzata soprattutto da grandissime e grandi catene di alberghi. Le dimensioni di queste catene sono tali che ciascuna può permettersi uno o più revenue manager, decidere le proprie politiche di prezzo e sviluppare ad hoc sistemi per la loro gestione.

La realtà alberghiera italiana è invece profondamente diversa ed è caratterizzata da piccole e medie imprese a conduzione per lo più familiare che di solito si tramanda da alcune generazioni. In questo scenario, se da una parte già solo far comprendere i benefici di politiche di pricing avanzate non è semplice, dall'altra il settore si rivela fertile per aziende che offrano l'outsourcing delle attività di revenue management. Sfruttando l'outsourcing la piccola impresa con un costo contenuto ottiene l'indispensabile presenza sui principali portali Web, una politica di pricing basata su criteri scientifici e risparmia il costo di un operatore che gestisca il tutto.

L'idea di base avuta dai fondatori della HotelSolutions è stata proprio quella di creare un'azienda la cui attività principale fosse l'outsourcing del revenue management e del livello operativo di gestione.

Come detto l'idea di business si rivolge ad un settore particolarmente sviluppato in Italia, ma ancora abbastanza digiuno di ICT. Al di là delle grandi catene alberghiere, esistono, infatti, poche esperienze di revenue management avanzato. Anche gli alberghi che dicono di farlo, di solito, applicano soluzioni *ad hoc*, che sono acquistate come prodotto all'estero o difettano di un sufficiente *background* che gli conferisca una certa validità scientifica.

3. COME DECIDERE IL PREZZO?

Il *background* di base del *know how* dell'azienda è costituito dalle teorie di Yield Management che sono un insieme di tecniche di previsione, mo-

delli di ottimizzazione e procedure con le quali è possibile determinare quali tariffe conviene applicare nell'ottica di massimizzare i guadagni. Il modo più semplice per determinare il costo è quello di utilizzare una stima basata sul budget, semplicemente considerando il numero di stanze occupate quel dato giorno l'anno precedente e ripartendo i costi ed il guadagno atteso su tale numero. È chiaro che tale sistema, per altro ancora molto usato nel mondo alberghiero, non è flessibile e può addirittura essere controproducente. Minimi miglioramenti possono essere apportati dall'esperienza dell'albergatore che analizza un periodo di tempo più lungo e il contesto (presenza di fiere, eventi speciali ecc. ecc.) e tara le stime effettuate. Un cambio di prospettiva radicale è dato alla considerazione che gli alberghi, con buona approssimazione, lavorano a "costo fisso". Ciò vuol dire che il numero di clienti non incide su questa voce ed il prezzo è l'unico elemento che può essere variato per massimizzare l'incasso. Ma il prezzo è proporzionale al numero dei clienti e quindi, da un'analisi previsionale dell'occupazione, è possibile ricavare una stima del ricavo per ogni singola giornata. Questa stima è funzione della previsione (legata al prezzo) e dell'occupazione effettiva (le prenotazioni già effettuate). Si può quindi costruire in un certo arco di tempo (di solito a par-

tire da 60 giorni prima) una curva che rappresenta la domanda ed il prezzo può essere utilizzato come leva perchè la questa curva porti al massimo del rendimento. Rispetto ai sistemi di Yield Management tradizionali la soluzione HotelSolutions prevede una semplificazione dei dati utilizzati come input ed un loro reperimento totalmente automatico, ma anche l'analisi dei competitor per tarare il prezzo. Concretamente quindi il sistema della HotelSolutions si compone di una parte automatica dedicata al recupero dei dati di occupazione corrente e di costruzione della curva della domanda e da una parte di interpretazione di questi dati che viene al momento effettuata manualmente ed è il passaggio cruciale verso un sistema che è molto più di Knowledge Management che di Data Mining. Il sistema non è infatti in grado di utilizzare euristiche non matematiche (per esempio, a Milano durante la "Fiera della plastica" qualunque sia il prezzo applicato l'hotel sarà fully booked) nè di specifiche considerazioni sui competitor (se il 90% dei competitor ha una certa tariffa è probabilmente inutile che l'hotel ne applichi una sensibilmente minore). Per questo motivo è necessaria l'interpretazione dei dati da parte di un soggetto umano che, con il suo intervento, trasforma l'informazione automatica in conoscenza. Come mostrato nella figura 1, i dati relativi all'occupazione vengono recuperati in maniera auto-

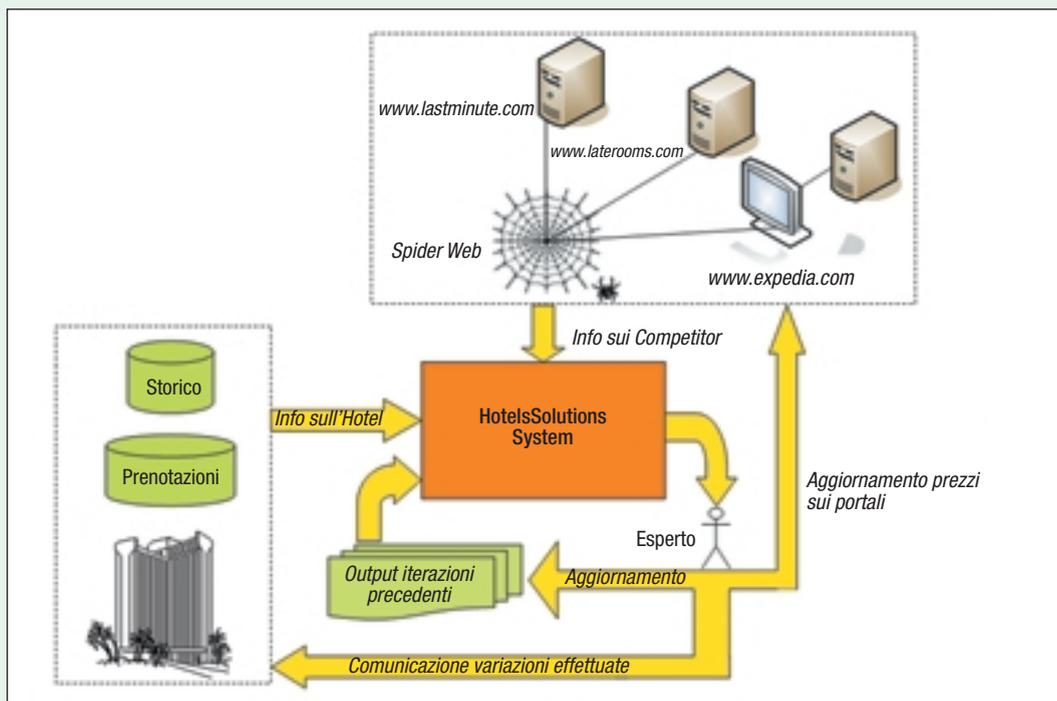


FIGURA1
Descrizione
del sistema
HotelSolutions

matica attraverso un **Web Crawler** progettato appositamente. Questo spider è in grado sia di realizzare accessi diretti (dove possibile grazie a specifici accordi) alle basi di dati sia di recuperare informazioni simulando in tutto e per tutto il comportamento umano. Tali dati costituiscono una parte dell'input al sistema informatico. L'input è quindi completato dalle informazioni sull'Hotel oggetto dell'analisi riguardanti i prezzi attuali e le tipologie delle camere, la loro eventuale occupazione, lo storico dell'occupazione nei successivi 60 giorni e la stima dei costi.

L'output del sistema è costituito da una serie di informazioni riguardanti i successivi 60 giorni; per ciascun giorno è riportato il prezzo attualmente praticato, il prezzo praticato dai competitor ed un eventuale proposta di adeguamento del prezzo.

4. DATI ECONOMICI, PARTNER COMMERCIALI E CONCORRENTI

La politica della HotelSolutions nella scelta dei clienti è improntata alla acquisizione di un solo cliente di una determinata tipologia (di solito legata al numero di stelle) per ciascuna zona. Le zone sono definite in base al numero totale di alberghi che vi sono ed in base alla dimensione della domanda. Attualmente HotelSolutions S.R.L. of-

fre i suoi prodotti a circa 50 hotel ed i suoi servizi a circa 15 sparsi su tutto il territorio nazionale, con una discreta concentrazione a Roma e Milano.

Per quanto riguarda il dato economico del primo anno (2005-2006), caratterizzato da molti investimenti, la HotelSolutions ha avuto un fatturato pari a €100.000 con una perdita complessiva di € 7.000. La previsione per l'anno 2006-2007 è invece di un fatturato di € 220.000 con un utile prima delle tasse pari a € 10.000.

4.1. Partner Commerciali

Nell'analizzare un'azienda presente da poco tempo sul mercato così flessibile e variegato come quello delle soluzioni ICT per aziende, è importante capire il tipo di relazione che si instaura con altri soggetti operanti nello stesso settore o in settori affini. La scelta operata da HotelSolutions è quella di interagire con partner dalle dimensioni aziendali significative e leader nei rispettivi settori. Questa scelta è dettata dalla fiducia nei propri mezzi e nella propria idea imprenditoriale che porta a privilegiare la qualità del rapporto rispetto al rischio di restare schiacciati dalle maggiori dimensioni.

Di particolare rilievo sono le collaborazioni con: □ "TravelClick" (www.TravelClick.net) che è una società americana leader nel settore di servizi alberghieri online.

WEB CRAWLER

Un Web crawler (anche conosciuto come Web spider) è un programma che passa in rassegna il World Wide Web in un modo metodico e automatizzato [4].

Un web crawler è un tipo di bot (programma o script che automatizza delle operazioni tipicamente eseguite da utenti umani). In generale, il web crawler prende in input una lista di URL da visitare. Durante l'analisi di un URL, identifica tutti gli hyperlink presenti in questa pagina e li aggiunge alla lista di URL da visitare. I web crawler possono rallentare notevolmente la visita di un sito ai normali utenti e, per questo motivo, operano di solito nelle ore in cui i server sono più scarichi e sono obbligati ad obbedire a specifici protocolli coi quali i proprietari di Web site possono indicare quali pagine non dovrebbero essere analizzate.

Particolari Web Crawler sono quelli che vengono utilizzati per effettuare query automatiche sui database associati a portali Web. Questi Web Crawler non sono interessati a ricostruire la struttura del Web, ma piuttosto a simulare le ricerche che un utente umano potrebbe fare.

Esempio tipico sono gli spider utilizzati per controllare i prezzi effettuati da ogni compagnia aerea sulle rotte di competenza o i prezzi di ciascun albergo esposto su un portale di prenotazioni on-line. Esistono due approcci diversi per l'utilizzo di questi strumenti: uno collaborativo ed uno indipendente.

Nel primo approccio il gestore dell'oggetto di analisi del Web Spider collabora ed in qualche modo favorisce la sua operazione preparando i dati di interesse per lo spider ed inviandoglieli quando vengono richiesti. Questo metodo riduce sensibilmente il carico di lavoro del server del gestore ma presuppone che questi abbia un effettivo interesse a collaborare ed a fornire i propri dati in maniera sistematica a strumenti di analisi.

Il secondo approccio, definibile come indipendente, viene di solito utilizzato quando il gestore non manifesti interesse a collaborare per fornire dati (comunque pubblici) in maniera sistematica. In questi casi il Web Spider simula in tutto e per tutto l'azione umana eseguendo tutte le query necessarie per recuperare i dati di suo interesse. Il costo di questa indipendenza è pagato in parte dal gestore del server che viene sovraccaricato ed in parte dal sistema di analisi che impiega molto più tempo per ottenere i dati.

□ “Merlin Wizard” (www.MerlinWizard.it) che è una società che offre alle imprese supporto, consulenza e gli applicativi necessari per operare sul mercato sempre più globalizzato e con gli strumenti della internet economy.

□ “PrideGroup” (www.Pride.it) che è un gruppo attivo da oltre quindici anni nel mercato dell’Information & Communication Technology. Si può notare come i soggetti con cui HotelSolutions collabora abbiano notevole esperienza e siano leader nel proprio settore.

4.2. Aziende Concorrenti

Anche se, come spiegato precedentemente, la realtà alberghiera italiana sembra particolarmente appetibile per aziende come la HotelSolutions, in realtà al momento non esistono reali competitor sul territorio nazionale. Le aziende che offrono al momento servizi di outsourcing si limitano sostanzialmente a due tipologie: vendita di software per l’aggiornamento dei portali web o vendita di servizi di aggiornamento dei portali web. In entrambi i casi manca del tutto l’analisi della domanda e il supporto alla politica del pricing. Spesso l’utile dell’azienda che offre questi servizi è legato ad una percentuale sulle vendite effettuate attraverso i portali gestiti.

5. PANORAMICA DI PRODOTTI E SERVIZI FORNITI DA HOTEL SOLUTIONS

Le soluzioni proposte alle aziende alberghiere dalla HotelSolutions sono molto varie e vanno dal semplice supporto per l’immissione di prezzi sui portali Web, all’analisi completa della domanda per la determinazione del prezzo da applicare. I prodotti sono:

□ iHotelier è un sistema di prenotazione centralizzato (CRS) web based. Attraverso questo sistema l’hotel può gestire le proprie tariffe e disponibilità su tutti i canali di distribuzione. Con iHotelier è possibile quindi gestire da una sola piattaforma e contemporaneamente:

i) Tariffe negoziate, corporate e preferenziali,

ii) Offerte speciali e pacchetti,

iii) Tariffe gruppi,

iv) il proprio Call center e, attraverso lo specifico strumento OneScreen, le prenotazioni attraverso il proprio sito internet.

Questo è il prodotto di base e non fornisce alcun valore aggiunto specifico, limitandosi ad unifica-

re su un’unica piattaforma di accesso una serie di servizi che l’hotel offre.

□ Hotelligence è una suite composta da Hotelview ed HotelRate che sono due prodotti software sviluppati in collaborazione con la PrideGroup che supportano l’inserimento delle tariffe sui principali portali e di visualizzare le tariffe con cui sugli stessi portali, si presentano i competitor. Il focus di questo prodotto è quello di consentire una gestione unificata di una serie di portali, ciascuno dei quali ha sue specifiche regole per l’inserimento dei dati. Sono presenti sul mercato molti altri prodotti analoghi.

□ Hotelligence è lo strumento proposto da HotelSolutions, in collaborazione con TravelClick, che aiuta a migliorare le performance sui GDS. Analizzando la produzione GDS dei concorrenti nei mesi precedenti e confrontandola con quella dell’hotel, comparando le loro tariffe di vendita e quali sono le agenzie che li prenotano, è possibile avere accesso ad informazioni fondamentali per migliorare le vendite sui canali GDS. Con questo prodotto il focus si sposta verso l’analisi dei competitor. Nello scenario globalizzato conoscere i propri competitor è fondamentale e quindi questo strumento è particolarmente apprezzato.

□ FuturePACE è lo strumento proposto da HotelSolutions, in collaborazione con TravelClick, che insieme a Hotelligence aiuta a migliorare le performance sui GDS. È lo strumento di previsione dell’occupazione che, comparando la stima di occupazione futura sui GDS con quella dei concorrenti dà la possibilità di definire le migliori tariffe adattandole a quelle proposte dal mercato di riferimento allo scopo di massimizzare la produttività del canale GDS. Questo è il prodotto di punta che sintetizza tutti i diversi aspetti dell’offerta dei prodotti HotelSolutions.

È importante notare come i prodotti per il pricing descritti sopra siano semiautomatici e l’ultimo step (cioè la decisione effettiva del prezzo da applicare) spetti comunque all’utente umano. Per questo motivo un aspetto qualificante dei servizi offerti, che sostanzialmente si basano sui prodotti descritti, è l’esperienza dell’operatore HotelSolutions.

Il servizio che rappresenta il Core Business è quindi la gestione in outsourcing della distribuzione elettronica. Sostanzialmente per un determinato albergo la HotelSolutions si incarica di utilizzare il prodotto FuturePACE comunican-

do giornalmente all'hotel tutte le operazioni effettuate.

Sono poi disponibili altri servizi che vanno dalla consulenza, alla formazione del personale, alla progettazione e sviluppo del sito internet dell'albergo.

Particolarmente simpatico è un servizio chiamato MysteryGuest che è utilizzabile per verificare la professionalità e la competenza dello staff alberghiero. Sostanzialmente viene effettuato un soggiorno anonimo presso l'hotel ed al termine, attraverso un ampio rapporto finale scritto viene presentato il risultato con un raffronto con le aziende concorrenti, suggerendo anche le possibili azioni di miglioramento.

6. CONCLUSIONI

L'idea di base avuta dai fondatori della HotelSolutions è stata quella di creare un'azienda la cui attività principale fosse l'outsourcing del revenue management e del livello operativo di gestione. Questa idea si inquadra nel settore turistico/alberghiero dove, a fronte di un enorme sviluppo degli strumenti ICT per il B2C, il B2B presenta

ancora ampi spazi di crescita. In particolare l'idea imprenditoriale di HotelSolutions risulta particolarmente valida nella realtà alberghiera italiana che è caratterizzata in larga parte da piccole e medie imprese a conduzione familiare piuttosto che da grandi catene come avviene invece nel mercato americano ed europeo.

L'azienda non si limita a vendere prodotti ICT e a fornire consulenze, ma ha scelto anche di offrire servizi completi di revenue management per le aziende alberghiere. Considerando anche la giovane età, i dati economici ed i solidi partner commerciali che ha, la rendono una realtà interessante e da seguire nel panorama delle aziende ICT italiane.

Bibliografia

- [1] *A comparative revenue analysis of hotel yield management heuristics*. Decision Science, 1999.
- [2] Yield management: System or program? 1999.
- [3] http://en.wikipedia.org/wiki/yield_management_system. 2006.
- [4] http://it.wikipedia.org/wiki/web_crawler. 2006.

ROBERTO BELLINI è docente nel corso di Concept Dvelopment & Industrial Engineering presso il Politecnico di Milano e docente di Marketing e Gestione della Relazione con il cliente nell'ambito del MIP, con una focalizzazione sulla innovazione nelle reti di imprese.

Presiede la Sezione AICA di Milano ed è responsabile per AICA del progetto eCCO-eCertification and Competence Tool, sviluppato con Federcomin e Fondazione Politecnico.

E-mail: roberto.bellini@polimi.it

CHIARA FRANCALANCI è professore associato di Sistemi Informativi al Politecnico di Milano. Ha scritto numerosi articoli sulla progettazione e sul valore economico delle tecnologie informatiche, svolto attività di ricerca e consulenza nel settore finanziario e manifatturiero sia in Italia sia presso la Harvard Business School ed è editor del Journal of Information Technology.

E-mail: francala@elet.polimi.it

STEFANO MODAFFERI è laureato in Ingegneria Elettronica presso l'Università Mediterranea di Reggio Calabria. Al momento è dottorando in Ingegneria dell'Informazione presso il Politecnico di Milano. La sua area di interesse è quella dell'ICT, in particolare si occupa degli aspetti legati al Knowledge Management e all'introduzione dei Web-Service nella gestione dei Business Process.

E-mail: modafferi@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 gestione dell'accessibilità di siti e applicazioni web

Giorgio Brajnik, David D'Agostini, Antonio Piva

1. INTRODUZIONE

Con il passare degli anni il numero di utenti che si connettono ad internet è notevolmente aumentato e, grazie al progresso tecnologico che ha caratterizzato l'informatica fin dai suoi albori, l'utilizzo delle tecnologie ha coinvolto anche i soggetti disabili. Proprio per questo motivo si sono realizzati standard internazionali che permettessero agli sviluppatori di realizzare siti web fruibili anche da soggetti che presentassero forme più o meno gravi di disabilità. Basti pensare al lavoro svolto dal World Wide Web Consortium e dalla *Web Accessibility Initiative* (WAI)¹ che nel 1999 ha dato alla luce le Linee Guida per l'Accessibilità ai Contenuti Web².

Si è quindi resa necessaria, anche a livello nazionale, l'elaborazione di strategie operative volte al raggiungimento di tali obiettivi, ponendo delle regole che ricalcassero gli standard internazionali in modo da garantire una compatibilità con il continuo evolversi delle tecnologie e tenendo conto delle diverse tipologie di disabilità. Questi motivi hanno ispirato l'approvazione della legge 9 gennaio 2004 n. 4 "Disposizioni per favorire

l'accesso dei soggetti disabili agli strumenti informatici"³, detta anche Legge Stanca, che si prefigge come finalità il riconoscimento e la tutela del diritto di ogni persona ad accedere a tutte le fonti di informazione e ai relativi servizi, ivi compresi quelli che si articolano attraverso gli strumenti informatici e telematici. Viene peraltro tutelato e garantito, in particolare, il diritto di accesso ai servizi informatici e telematici della pubblica amministrazione e ai servizi di pubblica utilità da parte delle persone disabili, in ottemperanza al principio di uguaglianza sancito dall'articolo 3 della Costituzione. I soggetti erogatori⁴ sono, quindi, obbligati a stipulare contratti che contengano tra le clausole il rispetto dei requisiti tecnici, definiti dalla normativa anche regolamentare, per la realizzazione di nuovi siti Internet.

Con il successivo Decreto del Presidente della Repubblica n. 75 del 1° marzo 2005, in attuazione a previsto dall'articolo 10 della legge sopra citata, vengono stabiliti i "criteri e principi operativi e organizzativi generali dell'accessibilità". Tale testo normativo, in buona sostanza, detta le regole dell'accessibilità ai siti web, tra cui le modalità con cui può essere richiesta la valutazione.

¹ Il sito ufficiale del World Wide Web Consortium è reperibile all'indirizzo www.w3.org, quello della Web Accessibility Initiative (WAI) è reperibile all'indirizzo www.w3.org/WAI.

² La versione delle Web Content Accessibility Guidelines è reperibile all'indirizzo www.aib.it/aib/cwai/WAI-trad.html.

³ Pubblicata in G.U. n.13 del 17 gennaio 2004 e reperibile all'indirizzo www.pubbliaccesso.gov.it/normative/legge_20040109_n4.htm.

⁴ I soggetti che devono sottostare alla legge 4/2004 sono: Pubbliche Amministrazioni, enti pubblici economici, aziende private concessionarie di servizi pubblici, aziende municipalizzate regionali, enti di assistenza e riabilitazione pubblici, alle aziende di trasporto e telecomunicazioni a prevalente partecipazione di capitale pubblico e alle aziende appaltatrici di sistemi informatici.

Riquadro 1 - Definizioni

- **Accessibilità:** capacità dei sistemi informatici, nelle forme e nei limiti consentiti dalle conoscenze tecnologiche, di erogare servizi e fornire informazioni fruibili, senza discriminazioni, anche da parte di coloro che a causa di disabilità necessitano di tecnologie assistive o configurazioni particolari.
- **Fruibilità:** la caratteristica dei servizi di rispondere a criteri di facilità e semplicità d'uso, di efficienza, di rispondenza alle esigenze dell'utente, di gradevolezza e di soddisfazione nell'uso del prodotto.
- **Verifica tecnica:** valutazione condotta da esperti, anche con strumenti informatici, sulla base di parametri tecnici.
- **Verifica soggettiva:** valutazione articolata su più livelli di qualità ed effettuata con l'intervento del destinatario, anche disabile, dei servizi sulla base di considerazioni empiriche.

Sia la legge che il decreto di attuazione, definiscono i concetti di accessibilità, fruibilità, verifica tecnica e verifica soggettiva, indicati nel Riquadro 1.

La verifica tecnica si riferisce all'accessibilità dei contenuti dei siti che vengono indicati nel Decreto Ministeriale dell'8 luglio 2005, riguardante i requisiti tecnici e i diversi livelli per l'accessibilità agli strumenti informatici.

La verifica tecnica è composta di 22 requisiti (definiti nell'allegato A di tale decreto) che traggono spunto dalle WCAG 1.0⁵ e dalla normativa americana di riferimento⁶. In particolare vengono illustrate le metodologie per la verifica della conformità agli standard che si articolano nelle seguenti attività:

- valutazione della correttezza del codice della pagina web con strumenti automatici di validazione;
- verifica dell'uso corretto di tali linguaggi;
- esame delle pagine con vari browser diversi fra loro, compresi quelli testuali;
- verifica delle modalità grafiche di presentazione del contenuto, con l'utilizzo due algoritmi per il controllo della differenza di luminosità e di colore;
- redazione del rapporto con indicate la conformità e le non-conformità, compresa la non applicabilità dei requisiti sulla pagina in esame.

I requisiti della verifica tecnica vengono applicati a tutti i soggetti erogatori, indicati in nota 4, compresi i casi in cui le informazioni e i servizi

sono erogati mediante tecnologie Internet, sia che essi vengano utilizzati all'interno di reti Intranet o Extranet, che per mezzo di supporti ottici (CD-ROM, DVD) quando il computer dell'utente non è connesso a una rete.

All'allegato B del menzionato decreto ministeriale, sono indicate le metodologie e i criteri di valutazione per la verifica soggettiva; quest'ultima si riferisce alla fruibilità⁷ delle informazioni e dei servizi offerti, tenendo in debito conto la facilità e semplicità di utilizzo, nonché l'efficienza e l'efficacia nell'impiego oltre che la soddisfazione dell'utente finale.

La Presidenza del Consiglio dei Ministri – Dipartimento per l'innovazione e le tecnologie⁸, avvalendosi del CNIPA⁹, si impegna a effettuare il monitoraggio dell'attuazione della legge, a vigilare sul suo rispetto da parte delle amministrazioni statali, a indicare i soggetti che si sono distinti per l'impegno nel perseguire le finalità indicate oltre che a promuovere iniziative e programmi per migliorare e diffondere le tecnologie assistive¹⁰ e l'accessibilità.

Inoltre, in caso di inosservanza della legge Stanca da parte dei soggetti tenuti al suo rispetto¹¹, per il responsabile dell'accessibilità¹² si prospetterebbero conseguenze dirigenziali e disciplinari che possono arrivare nei casi più gravi fino al mancato rinnovo dell'incarico, ai sensi degli articoli 21 e 55 del decreto legislativo 30 marzo 2001, n.165.

⁵ Si veda la nota 2.

⁶ Si fa riferimento al paragrafo 1194.22 della Section 508 del Rehabilitation Act (www.section508.gov).

⁷ Definita dal Decreto come caratteristica dei servizi di rispondere a criteri di facilità e semplicità d'uso, di efficienza, di rispondenza alle esigenze dell'utente, di gradevolezza e di soddisfazione nell'uso del prodotto.

⁸ Il sito del ministero per l'innovazione è reperibile all'indirizzo www.innovazione.gov.

⁹ Centro Nazionale per l'Informatica nella Pubblica Amministrazione www.cnipa.gov.it.

¹⁰ *Tecnologie assistive*: ai sensi dell'art. 2 della legge n.4 del 2004, gli strumenti e le soluzioni tecniche, hardware e software, che permettono alla persona disabile, superando o riducendo le condizioni di svantaggio, di accedere ai servizi erogati dai sistemi informatici.

¹¹ Si vedano in nota 4.

¹² Introdotto all'articolo 9 comma 1 del d.p.r. 75 del 1 marzo 2005 e che solitamente coincide con il dirigente responsabile dei servizi informativi.

Con questa normativa si cerca così di abbattere barriere dell'informazione paragonabili a quelle architettoniche, cercando di garantire (anche attraverso il sito di riferimento www.pubblicaccesso.gov.it dedicato all'accessibilità e alla normativa italiana che lo regolamenta) il diritto di accedere alle tecnologie informatiche senza discriminazione di sorta.

2. L'ACCESSIBILITÀ: UN FATTORE DI QUALITÀ DIFFICILE DA CARATTERIZZARE

L'accessibilità è un fattore di qualità che va ad incidere sia su aspetti tecnici del sito (per esempio, la correttezza del codice delle pagine), sia su aspetti legati all'interazione con gli utenti. Soprattutto per quanto riguarda questo secondo aspetto, l'accessibilità è assimilabile alla fruibilità, all'usabilità ed alla qualità in uso del sito. Nonostante sia possibile poter far riferimento a dei requisiti tecnici relativi alla codifica dell'interfaccia utente di un sito (si pensi alla codifica appropriata delle tabelle usate per presentare dati tabellari), ciò non deve far credere che il loro mero soddisfacimento alla lettera di requisiti tecnici possa garantire una buona accessibilità. Una definizione che cattura bene il concetto comunemente inteso di accessibilità, e allo stesso tempo lo differenzia da quello di usabilità/qualità in uso è derivata da quella data da due studiosi statunitensi nel 2003, Slatin e Rush¹³: un sito è accessibile se esso permette lo stesso grado di efficacia e sicurezza per utenti disabili e utenti non disabili. Sono due i pregi fondamentali di questa definizione: da una parte essa mette a fuoco due parametri legati alle prestazioni degli utenti, ossia l'efficacia e la sicurezza; la prima viene normalmente intesa come la capacità da parte dell'utente del sito di raggiungere determinati obiettivi in maniera accurata (precisa, senza errori) e completa, la seconda fa riferimento invece all'assenza di danni fisici (esempio, attacco epilettico) o immateriali (esempio, economici) causati dall'uso del sito. Dall'altro lato viene fornito un livello di riferimento per tali parametri che devono essere almeno equivalenti a quelli garantiti a utenti non disabili. La conseguenza è che un sito accessibile non discrimina, a livello di possibilità di ot-

tenere delle informazioni o di svolgere dei compiti, tra i due tipi di utenti.

L'importanza di una tale definizione deriva dall'esame delle modalità tipiche di utilizzo del web da parte di utenti disabili. Esse dipendono dal tipo di disabilità, che potremmo classificare nei seguenti gruppi:

□ *non-vedenti*: per poter usare il web tali persone usano delle applicazioni, installate sul proprio computer, che pronunciano ad alta voce il contenuto testuale delle finestre presenti sullo schermo, i lettori di schermo. Tali programmi, denominati lettori di schermo, vengono realizzati in maniera da poter sfruttare l'architettura dell'accessibilità offerta dalla piattaforma utilizzata (per esempio, per i sistemi Microsoft esiste una libreria "Microsoft Active Accessibility SDK" che gli sviluppatori delle interfacce utente grafiche devono utilizzare; similmente per applicazioni Java e per applicazioni GNU/Linux sotto GNOME o KDE). Questi software, pilotati dall'utente mediante la tastiera, trasformano stringhe di caratteri in fonemi pronunciati da voci molto spesso sintetiche e permettono di spostarsi nella pagina, di compilare campi di testi ed altri controlli delle form, leggere la descrizione di un'immagine, estrarre dalla pagina la lista dei link, la lista dei titoli di sezione (in html identificati dai tag H1 ... H6) o dei frame e di saltare a quelli di interesse. Talvolta i non vedenti usano anche le barre Braille, dispositivi di uscita che producono su una schiera di matrici di aghi (tipicamente 40 matrici 6 × 2 o 8 × 2) la rappresentazione in celle Braille di una sequenza di caratteri. Passandoci sopra il polpastrello delle dita il non vedente rileva la configurazione di ciascuna delle matrici e quindi legge i corrispondenti caratteri. La barra Braille è direttamente controllata dal lettore di schermo, dato che spesso entrambi i dispositivi (uno hardware e uno software) vengono impiegati congiuntamente;

□ *ipo-vedenti*: persone con difetti di vista che ne limitano l'acuità, il campo visivo e la percezione dei colori. Lo strumento più frequentemente usato è l'ingranditore di schermo, un'applicazione software che ingrandisce, fino a 36 volte, una porzione dello schermo, ne può modificare lo schema dei colori (per esempio, invertendo la polarità e fornendo quindi uno sfondo scuro ed il materiale di primo piano chiaro) o il livello di contrasto, e può variare la rappresentazione del puntatore del mouse. Con questo tipo di programmi l'utente ingrandisce lo schermo del fattore desiderato, e può spo-

¹³ Slatin e Rush, Maximum Accessibility: making your web site more usable for everyone, Addison Wesley.

stare questa “lente virtuale” mano a mano che deve esaminare le parti rilevanti della finestra;

□ **disabili motori:** persone con difficoltà di movimento e di controllo degli arti superiori, del busto, del capo e/o delle mani (altri tipi di disabilità motorie sono ininfluenti per quanto riguarda l'uso del calcolatore). Tali persone possono usare tastiere, mouse, joystick, trackball particolari: ingranditi, coi tasti distanziati tra loro, con “scudi” che permettono di appoggiarci di peso la mano sopra e di guidare il movimento delle dita in modo da colpire esattamente il tasto desiderato (Figura 1);

□ **sordi o ipo-udenti:** sono impossibilitati a fruire del materiale audio disponibile in un sito web (per esempio, il podcasting) e non dispongono di alcuna “protesi” specifica per l'uso del calcolatore;

□ **disabili cognitivi:** persone con disfunzioni cognitive legate a processi di comprensione, di memorizzazione, di concentrazione, di lettura e/o scrittura. Anche in questo caso non esiste alcuna “protesi” per l'uso del calcolatore.

A seconda del tipo di disabilità (che può essere più di una e a vari gradi), un generico sito web può presentare vari tipi di barriere.

Per esempio, per un utente di un lettore di schermo il fatto che un'immagine sia cliccabile, ma priva di alcuna descrizione testuale (ad es. un pulsante grafico con disegnato, in pixel, la scritta “Acquista” e privo dell'attributo ALT previsto dal linguaggio HTML) risulta essere una barriera critica in quanto il lettore di schermo pronuncerà il nome del file che contiene l'immagine (continuando il precedente esempio,

esso potrebbe essere “`immag/a.gif`”) rendendo impossibile la comprensione del pulsante¹⁴. Un'altra possibile barriera è l'assenza dei cosiddetti *skip-links*; essi consentono di muoversi con il solo utilizzo della tastiera (per esempio, `<shift-tab>` per muoversi all'indietro, `<enter>` per seguire un link, i tasti spazio e l'uso delle frecce per compilare le form) all'interno di una pagina¹⁵.

Bisogna notare inoltre che, in conseguenza delle modalità con cui il sito web viene fruito (per esempio, mediante resa audio del contenuto testuale e comandi inviati da tastiera, nel caso di utilizzo di un lettore di schermo), i parametri legati alla produttività e alla soddisfazione possono cambiare radicalmente. Vari studi sperimentali hanno verificato che il numero di errori commessi e il tempo richiesto per lo svolgimento di determinati compiti nel caso di utenti di lettori di schermo e di ingranditori di schermo crescono di 3-5 volte, e misure di efficacia calano di altrettanto.

3. IL PROBLEMA DELLA GESTIONE DELL'ACCESSIBILITÀ

Il vero problema è la gestione dell'accessibilità in maniera sostenibile, cioè sia tecnicamente valida (in modo da riuscire a fornire un sito accessibile – almeno in parte) che economicamente fattibile.

Il problema è complicato dal fatto che il sito non è un'entità statica, ma può cambiare anche a distanza di pochi minuti (per esempio, per un sito di notizie). I cambiamenti possono essere localizzati e limitati (per esempio, un nuovo comunicato) oppure in certi casi possono riguardare una parte cospicua delle pagine (per esempio, uno scoop, una nuova campagna promozionale, un nuovo servizio offerto).

Talvolta i cambiamenti sono associati all'utilizzo di altri media (per esempio, podcasting, materiale multimediale) o altre tecnologie realizzative (per

FIGURA 1
Joystick con scudo in plexiglass (Ufficio H della Comunità Piergiorgio di Udine www.piergiorgio.org)



¹⁴ Si veda il sito Web Quality www.dimi.uniud.it/wq per una descrizione dettagliata delle possibili barriere.

¹⁵ Una soluzione frequentemente adottata è quella di includere un link posizionato prima del materiale da saltare che punti ad un'ancora posizionata alla fine del materiale da saltare. Può essere implementato nel seguente modo:
.... `Salta al contenuto` altro materiale ...
`` qui inizia il contenuto
In questo modo è possibile fare Tab fino a raggiungere il link “Salta al contenuto” e attivandolo ci si posiziona subito prima del contenuto stesso.

esempio, AJAX¹⁶) che vanno ad impattare sull'interfaccia utente e quindi sulle potenziali barriere. Inoltre sono varie le persone che partecipano di prima mano alle modifiche. Per esempio in un sito universitario, le modifiche di pagine del sito possono venir svolte da persone del team responsabile per il web, ma anche da membri dell'ufficio relazioni pubbliche (tipicamente giornalisti), da segretarie di istituti, da docenti, ricercatori o membri di gruppi di ricerca (non necessariamente informatici). Una conseguenza è di non poter uniformare il livello di qualità delle pagine. Anche nel caso di personale tecnico specializzato, le conoscenze legate all'usabilità, all'accessibilità e agli strumenti di realizzazione dell'interfaccia utente (HTML, CSS, JavaScript in primo luogo) possono essere lacunose o errate. Quindi l'accessibilità pone seri problemi in merito all'organizzare dei processi di realizzazione, gestione, manutenzione del sito per poter mantenere determinati livelli di accessibilità.

4. POLITICHE E PROCESSI AZIENDALI PER L'ACCESSIBILITÀ

La soluzione è la stessa che si adotterebbe per quanto riguarda la gestione della qualità in generale: occorre in primo luogo capire cosa si vuole ottenere, perché, e come ottenerlo e in seguito verificare periodicamente il livello cui si è giunti. In concreto, per quanto riguarda l'accessibilità, un approccio ingegneristico prevede che sulla base di un'analisi costi-benefici si definisca una politica aziendale per l'accessibilità. Come tutte le attività legate al controllo di qualità, anche l'accessibilità ha i suoi costi: quello dello svolgimento di verifiche volte a identificare le barriere; della loro diagnosi e la loro rimozione; della dotazione di infrastrutture tecniche e organizzative (alcuni processi di sviluppo devono cambiare, mettendo in primo piano il ruolo dell'accessibilità e dell'usabilità) e della formazione e aggiornamento del personale. D'altro canto però l'accessibilità porta indiscutibili benefici: oltre a evitare di escludere utenti disabili (che numericamente costituiscono una rilevante porzione dell'audience) e al di là dell'ottemperanza di norme giuridiche, la realizzazione di un sito accessibile è diventata una

questione di immagine e un marchio di professionalità. Inoltre risulta più facilmente modificabile e maggiormente compatibile con le tecnologie usate dagli utenti (per esempio, per l'utilizzo del web mediante cellulari o palmari, o mediante portali vocali) ed è interoperabile con altre tecnologie (per esempio, molti dei suggerimenti che i tecnici di Google danno al fine di migliorare il posizionamento del sito nel motore di ricerca coincidono con una parte dei requisiti di accessibilità; pagine scritte in XHTML sono facilmente elaborabili da agenti software specializzati).

Pertanto il primo passo nella definizione della politica di accessibilità deve essere l'identificazione dell'obiettivo da perseguire: perché preoccuparsi dell'accessibilità e cosa aspettarsi da essa, in termini di indicatori economici (per esempio, ampliamento dell'audience). Una volta identificato l'obiettivo, sarà possibile sia scegliere una il livello di accessibilità (per esempio, stessa efficacia per tutti gli utenti), sia capire come misurarla. La misurazione del livello di accessibilità è un ulteriore aspetto critico, che spesso viene dato per scontato. Ne è una dimostrazione il gran numero di siti web italiani che si professano accessibili solo in virtù del fatto che qualche strumento di verifica non ha segnalato alcun problema. La misurazione deve ovviamente essere valida, ossia deve permettere di decidere se il sito è accessibile o meno sia per quanto riguarda gli aspetti tecnici (verifica tecnica) che per quanto riguarda gli aspetti legati alla fruibilità (verifica soggettiva) ed all'efficacia. In particolare per quest'ultimo aspetto, la misurazione sarà centrata soprattutto sulla capacità degli utenti di arrivare in fondo ai compiti soddisfacendo le proprie esigenze. La misurazione potrà essere basata su test sperimentali con utenti (user testing), su valutazioni soggettive autonome (subjective assessments), su analisi delle barriere svolte a tavolino da esperti (barrier walkthrough)¹⁷. Considerando il concetto di fruibilità (come definito dalla legge Stanca) allora la misurazione dovrà includere anche le proprietà di produttività e soddisfazione, che dovranno essere rese esplicite per mezzo di misurazioni basate sui metodi elencati in precedenza.

¹⁶ AJAX: Asynchronous Javascript and XML; informazioni in <http://it.wikipedia.org/wiki/AJAX>

¹⁷ Si veda http://www.dimi.uniud.it/giorgio/publications_flat.html#linzo6 per una discussione di vari metodi di misurazione dell'accessibilità del web.

CONCLUSIONI

Come si può notare non si è posto l'accento sui metodi per la verifica di conformità basata su linee guida (come per esempio, l'elenco dei 22 requisiti tecnici previsti dalla normativa Stanca, o le linee guida WCAG 1.0 e quelli previsti dalla nuova bozza WCAG 2.0) in quanto la verifica delle eventuali violazioni di linee guida in un sito di per sé non garantisce che si ottenga una misura efficace dell'accessibilità.

L'unico modo per rendere operativo e utile un tale metodo consiste nel derivare, dalle linee guida ufficiali, alcune *best practice* aziendali che siano state adattate alle particolarità del sito e la cui violazione comporta un calo di accessibilità. Per esempio, solo chi conosce bene il sito è in grado di decidere quali immagini hanno puramente uno scopo decorativo (per le quali il testo alternativo potrebbe essere nullo) e quali invece comunicano delle informazioni che valga la pena rendere anche in maniera testuale.

Un altro aspetto legato all'implementazione dell'obiettivo (oltre la misurazione) è la definizione del perimetro dell'intervento. Come per la pulizia di casa propria, occorre decidere a che livello di igiene si vuole portare ciascuno dei locali di casa e il livello non è lo stesso per tutti i locali, dato che esso dipende da chi e come utilizza il locale. Anche per un sito web è necessario decidere a che livello di accessibilità bisogna portare le varie sezioni del sito e non è detto che il livello sia il medesimo. Una ragione per differenziarlo è la complessità: se una parte del sito è basata su materiale multimediale o animazioni, o applicazioni altamente interattive in AJAX, allo stato attuale della tecnologia sarà molto difficile renderla accessibile. È importante accorgersene subito e decidere in maniera specifica come affrontare questo problema.

Infine è necessario stabilire il processo di implementazione dell'accessibilità: bisogna decidere innanzitutto quali sono i soggetti che devono procedere all'applicazione dei metodi scelti per la misurazione dell'accessibilità, nonché la frequenza, il momento e gli strumenti da utilizzare. Sarà necessario provvedere alla realizzazione di un'adeguata infrastruttura tecnica, per esempio, dotando le persone di strumenti di tracciamento

dei difetti (*bug tracking*, per esempio, Bugzilla) e di strumenti di test dell'accessibilità¹⁸, nonché umana (con la messa in opera di processi di monitoraggio, diagnosi e soluzione, verifica, ottenimento di feedback da parte dell'utenza e di risposta, di disseminazione delle conoscenze allo scopo di prevenire l'immissione di difetti, di formazione ecc.).

Una ricaduta positiva di questo tipo di organizzazione è la sua diretta applicabilità ad altri fattori di qualità di difficile caratterizzazione, come per esempio, l'usabilità o la credibilità. Anche in questi casi è auspicabile la definizione di una politica aziendale, l'identificazione dei metodi più appropriati per la sua misurazione e l'instaurazione dei processi che implementano tale politica giorno per giorno: in questo modo, con un metodo fortemente incentrato sull'ingegneria gestionale, sarà possibile attuare l'accessibilità e la qualità in maniera sostenibile.

GIORGIO BRAJNIK è ricercatore presso la Facoltà di Scienze Matematiche Fisiche e Naturali dell'Università di Udine, dove insegna e fa ricerca nella progettazione orientata all'utente di siti e applicazioni web. Dal 1999 si occupa specificatamente di usabilità e accessibilità del web, e in special modo in metodi e strumenti di supporto per chi sviluppa siti. È stato consulente scientifico di Usablenet Inc., un'azienda statunitense che produce e vende strumenti di verifica dell'accessibilità. In tale veste ha collaborato direttamente con personaggi e aziende di spicco del settore, tra i quali Jim Thatcher, Bob Regan, Jakob Nielsen, Macromedia, IBM.
E-mail: giorgio@dimi.uniud.it

ANTONIO PIVA laureato in Scienze dell'Informazione, Membro del Consiglio Nazionale e Presidente della commissione di informatica giuridica dell'ALSI (*Associazione Nazionale Laureati in Scienze dell'Informazione ed Informatica*).

Docente a contratto di diritto dell'informatica all'Università di Udine.

Consulente sistemi informatici e Governo Elettronico nella PA locale, valutatore di sistemi di qualità ISO9000 e ispettore AICA ECDL base ed advanced.
E-mail: antonio_piva@libero.it

DAVID D'AGOSTINI avvocato, ha conseguito il master in informatica giuridica e diritto delle nuove tecnologie, fornisce consulenza e assistenza giudiziale e stragiudiziale in materia di *software*, *privacy* e sicurezza, contratti informatici, *e-commerce*, nomi a dominio, computer crimes, firma digitale. Ha rapporti di partnership con società del settore ITC nel Triveneto. Collabora all'attività di ricerca scientifica dell'Università di Udine e di associazioni culturali.
E-mail: david.dagostini@adriacom.it

¹⁸ Si veda la pagina del W3C www.w3.org/WAI/ER/tools/complete per un elenco esaustivo.

DENTRO LA SCATOLA

Rubrica a cura di

Fabio A. Schreiber

Dopo aver affrontato negli scorsi anni due argomenti fondanti dell'Informatica – il modo di codificare l'informazione digitale e la concreta possibilità di risolvere problemi mediante gli elaboratori elettronici – con questa terza serie andiamo ad esplorare “*Come parlano i calcolatori*”. La teoria dei linguaggi e la creazione di linguaggi di programmazione hanno accompagnato di pari passo l'evolversi delle architetture di calcolo e di gestione dei dati, permettendo lo sviluppo di applicazioni sempre più complesse, svincolando il programmatore dall'architettura dei sistemi e consentendogli quindi di concentrarsi sull'essenza del problema da risolvere.

Lo sviluppo dell'Informatica distribuita ha comportato la nascita, accanto ai linguaggi per l'interazione tra programmatore e calcolatore, anche di linguaggi per far parlare i calcolatori tra di loro – i protocolli di comunicazione. Inoltre, la necessità di garantire la sicurezza e la privacy delle comunicazioni, ha spinto allo sviluppo di tecniche per “non farsi capire” da terzi, di qui l'applicazione diffusa della crittografia.

Di questo e di altro parleranno le monografie quest'anno, come sempre affidate alla penna (dovrei dire tastiera!) di autori che uniscono una grande autorevolezza scientifica e professionale ad una notevole capacità divulgativa.



Linguaggi per Basi di Dati

Letizia Tanca

1. INTRODUZIONE

L'informatica può essere definita come la scienza che si occupa della *rappresentazione* e della *gestione* dell'informazione per mezzo di macchine digitali. Il problema della rappresentazione dell'informazione riguarda lo studio di metodi di descrizione appropriati all'elaborazione da parte di una macchina digitale; consiste perciò nell'*astrarre* i concetti importanti per la realtà applicativa di interesse da quelli trascurabili. Il problema della gestione riguarda invece l'uso e la trasformazione dell'informazione in maniera funzionale al problema applicativo da risolvere.

Ancora ai tempi di Turing, non era chiaro se fosse più opportuno utilizzare un modello di calcolo che distinguesse in maniera netta questi due aspetti; infatti, nella computazione basata sulla macchina di Von Neumann, che è diventata il modello standard di elaboratore, vi è una netta distinzione tra il ruolo dei programmi (preposti alla *gestione* dell'informazione) e quello dei dati (il cui ruolo è *rappresentarla*).

La macchina concreta proposta da Turing, in al-

ternativa, prevedeva che i programmi potessero agire anche su parti del programma stesso, e così, eventualmente, automodificarsi; questo approccio è presente nella disciplina dell'Intelligenza Artificiale, i cui tipici linguaggi di programmazione hanno un paradigma che permette al programma di modificare se stesso. A parte questa disciplina, però, nell'informatica l'approccio “a-la-Von Neumann” è rimasto quello più comune, e ha dato luogo anche alla tradizione di mantenere in memoria centrale (o virtuale) il(i) programma(i) in elaborazione, mentre i dati, che di solito costituiscono un insieme piuttosto voluminoso, restano in memoria di massa, da “richiamare” all'occorrenza.

Anche per questo motivo le comunità degli esperti di rappresentazione dell'informazione e di gestione della stessa sono in qualche modo distinte: i primi si sono concentrati sulla ottimizzazione della organizzazione e dell'accesso a memoria di massa, per le caratteristiche di persistenza e di maggior quantità di spazio disponibile di quest'ultima.

L'articolo è organizzato come segue: dopo una breve carrellata sulle nozioni preliminari a pro-

posito delle basi di dati, introdurremo i modelli dei dati, e in particolare il modello relazionale; questa trattazione preliminare è necessaria, perché per parlare di linguaggi per basi di dati non si può prescindere dalle definizioni più importanti di questa disciplina. Ci concentreremo quindi sui linguaggi di interrogazione per le basi di dati relazionali, che costituiscono oggi il 90% dei sistemi in uso, e infine faremo anche un breve riferimento ai sistemi e ai linguaggi di interrogazione orientati agli oggetti.

Lo sforzo dei ricercatori e dei progettisti di linguaggi di interrogazione continua comunque a concentrarsi anche sullo studio di proprietà matematiche (logiche o algebriche) dei linguaggi, che richiede l'uso di formalismi appropriati. Per questo motivo sono stati da lungo tempo introdotti i cosiddetti linguaggi *formali* di interrogazione, cui sarà specificamente dedicato un altro articolo di questa serie.

2. NOZIONI PRELIMINARI

Nei sistemi informatici (e non solo in essi), le *informazioni* vengono rappresentate attraverso i *dati*. Le *informazioni* sono le notizie, gli elementi che consentono di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere. I dati sono ciò che è immediatamente presente alla conoscenza, prima di ogni elaborazione; in pratica, nell'informatica sono costituiti da simboli che debbono essere elaborati. Per esempio, il numero 30 è un dato, ma costituisce informazione solo se inquadrato in un contesto di riferimento: "Mario ha 30 anni" oppure "Giovanni ha ottenuto 30 all'esame di Informatica".

Una *base di dati* è una collezione di dati, utilizzati per rappresentare le informazioni di interesse per una o più applicazioni di una organizzazione. Una tale collezione di dati è però particolarmente efficace se se ne mette in evidenza la capacità informativa, organizzandola secondo una struttura, detta *modello dei dati*, mantenuta e gestita da un apposito sistema. Un modello dei dati è un insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne la dinamica; i componenti fondamentali di un modello dei dati sono perciò i suoi meccanismi di strutturazione (o costruttori di tipo). Per esempio, il modello relazionale prevede il costruttore *relazione* (tabella), che permette di strutturare la base di dati in insiemi di record

che hanno tutti la stessa forma. Tale struttura realizza, di fatto, una organizzazione *logica* dei dati, che permette di *trasformare i dati stessi in informazioni*. Con riferimento all'esempio citato sopra, il dato "30", ritrovato nel record relativo all'esame di Informatica dello studente Giovanni, nel campo VOTO, fornisce l'informazione che Giovanni ha ottenuto 30 all'esame.

Un Sistema di Gestione di Basi di Dati (*DataBase Management System*, DBMS) ha il compito di:

- mantenere i dati strutturati secondo un modello dei dati;
 - condividere i dati tra più applicazioni: poiché le stesse informazioni, eventualmente diversamente organizzate, possono essere utilizzate da applicazioni diverse, è utile mettere i relativi dati a fattor comune, per evitare ridondanze;
 - controllare la concorrenza di vari utenti sugli stessi dati: la base di dati è una risorsa condivisa, ma per motivi di efficienza occorre elevare il più possibile il livello di concorrenza; ciò vuol dire che il controllo di concorrenza deve scendere a un livello di granularità molto basso, eventualmente del singolo record;
 - mettere a disposizione meccanismi di definizione della privatezza dei dati e di limitazioni all'accesso (autorizzazioni);
 - garantire l'affidabilità sia delle operazioni che vengono svolte sui dati, sia del dispositivo di memorizzazione;
 - ultimo compito, ma tra i più importanti: permettere l'accesso ai dati in lettura e scrittura. A tale scopo, il primo passo consiste nel ritrovare i dati mediante opportuni linguaggi, detti di *interrogazione*. Come vedremo, interrogare una base di dati mediante questi linguaggi permette di accedere in modo diretto ai dati *basandosi sulle loro proprietà*. Su questo aspetto ci concentreremo in questo articolo.
- Come già detto, un *modello dei dati* è una collezione di concetti che possono essere usati per rappresentare la realtà: oltre all'esempio delle tabelle, possiamo a questo scopo pensare ad alberi, grafi ecc.. I modelli costituiscono quindi una strutturazione semplificata della realtà, che ne accoglie aspetti specifici e aiuta a comprenderla meglio.

Lo *schema dei dati*, invece, serve per rappresentare una specifica parte della realtà di interesse, mediante un modello dei dati. Per esempio, un modello ad albero si presta in modo naturale a rappresentare l'albero genealogico o il

tracciato di un fiume; un modello a grafo ben si presta a rappresentare il sistema fluviale o il sistema stradale di un Paese, mediante il modello relazionale organizzo in tabelle gli studenti con i relativi esami e voti.

A questo punto, avendo deciso qual è il modello dei dati più opportuno, ed avendo deciso come modellare la realtà di interesse progettando uno schema, i dati veri e propri, cioè il contenuto informativo, costituiscono l'istanza della base di dati, cioè una collezione di valori che *rispetta la struttura dettata dallo schema*. Per esempio, avendo deciso di rappresentare mediante un albero la genealogia delle famiglie, la famiglia Tanca e la famiglia Rossi saranno istanze di alberi genealogici. La rete viaria della Germania come è nella cartina edita il giorno 22 febbraio 2006 sarà un'istanza dello schema di rete viaria basato sul modello a grafo, mentre i dati sugli studenti e gli esami del Politecnico di Milano, in questo preciso istante, sono un'istanza della base di dati relazionale rappresentante studenti ed esami di cui sopra. Possiamo anche dire che, mentre lo schema descrive i dati a livello *intensionale*, fornendone, cioè, le *proprietà strutturali*, l'istanza ne esibisce l'aspetto *estensionale*, e cioè *costituisce l'informazione stessa*.

In un DBMS i dati si possono vedere, o meglio descrivere, a livelli diversi. Questa possibilità permette di liberare da una parte il progettista della base di dati, dall'altra il programmatore delle applicazioni, dalla preoccupazione di come i dati sono fisicamente organizzati in memoria di massa, concentrandosi sul loro significato piuttosto che sulla loro organizzazione fisica.

Il più alto livello di astrazione è il livello *esterno*, o *delle applicazioni*. A questo livello, i programmi applicativi vedono i dati organizzati secondo il modello dei dati del DBMS prescelto (o anche, eventualmente, secondo modelli diversi), ma con una visione personalizzata, detta *vista*, che permette di trascurare dati non di pertinenza di quell'applicazione, riorganizzando i dati di pertinenza in maniera appropriata e allo stesso tempo realizzando il controllo della privatezza delle informazioni. Il livello *logico*, immediatamente sottostante, descrive i dati, sempre in accordo col modello dei dati del DBMS, ma secondo una visione comune, che prevede l'organizzazione dei dati stessi in uno schema unico. Infine, il livello *fisico* descrive il modo in cui i dati sono organizzati nel supporto fisico di memo-

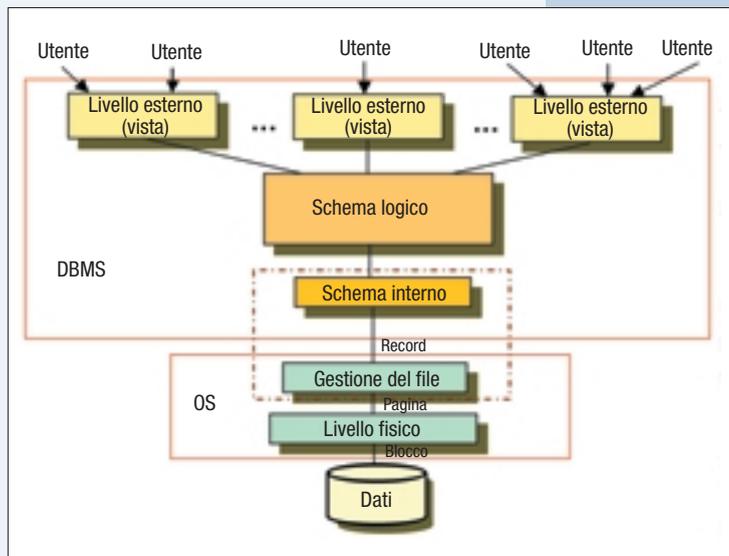


FIGURA 1
Architettura ANSI/SPARC delle moderne basi di dati

rizzazione, tipicamente il disco magnetico, che però è oggi affiancato anche da altre tecnologie, come i dischi ottici o le memorie flash. La figura 1 mostra la cosiddetta Architettura ANSI/SPARC delle moderne basi di dati. Il sistema di gestione della base di dati si fa carico di mantenere lo schema logico dei dati indipendente dalle strutture fisiche di memorizzazione, realizzando dunque una proprietà fondamentale delle moderne basi di dati, la cosiddetta *indipendenza dei dati*. L'accesso avviene infatti solo tramite il livello esterno (le cosiddette *viste d'utente*) che può, nel caso di un sistema informativo semplice, con esigenze poco diversificate, coincidere con il livello logico. Il DBMS garantisce il disaccoppiamento tra il livello esterno e quello logico (*indipendenza logica*), facendo in modo che aggiunte o modifiche alle viste non richiedano modifiche al livello logico e viceversa, e tra il livello logico e quello fisico (*indipendenza fisica*), facendo in modo che le modifiche allo schema logico lascino inalterata l'organizzazione fisica dei dati e viceversa. Coerentemente con quanto detto, i DBMS prevedono due tipi principali di modelli dei dati: i *modelli fisici*, che servono per organizzare i dati in memoria di massa tenendo conto della struttura a blocchi dei dischi, e i *modelli logici*, che vengono usati dai programmi e dagli umani per riferirsi ai dati descrivendoli mediante uno schema logico, sia al livello logico che al livello esterno (o *delle viste d'utente*).

I modelli fisici sono utilizzati nei DBMS esistenti, ma ad essi fa riferimento solo il DBMS per realizzare le strutture fisiche di memorizzazione (*schema interno*); esempi di tali modelli sono strutture indicizzate come alberi bilanciati (B-tree, B+-tree), file sequenziali, file ordinati, file organizzati con strutture hash ecc.. I modelli logici sono pure utilizzati nei DBMS esistenti, ma ad essi fanno riferimento i programmi, i progettisti e gli utenti finali per vedere i dati organizzati in accordo con il loro significato.

3. IL MODELLO RELAZIONALE DEI DATI

I tre modelli logici dei dati tradizionali sono il modello *gerarchico*, quello *reticolare* e quello *relazionale*. I modelli gerarchico e reticolare sono basati, rispettivamente, su strutture ad albero e su strutture a grafo. Sono modelli in realtà abbastanza vicini alle strutture fisiche di memorizzazione, poiché in essi i record contengono al loro interno dei puntatori che fanno da riferimento ad altri record ad essi collegati. Per tornare all'esempio degli studenti, dei corsi e degli esami, in una base di dati gerarchica o reticolare ogni record relativo a un esame conterrebbe dei puntatori, rispettivamente, al record del corso e a quello dello studente. Si realizza perciò una struttura logica "sovrapposta" a quella fisica, con una evidente perdita di indipendenza dalla struttura di memorizzazione.

Il modello relazionale è più astratto e meglio si presta a realizzare tale proprietà, in quanto in tale modello i record non possono contenere puntatori ad altri record, ma i riferimenti fra dati in record appartenenti a strutture (relazioni) diverse sono rappresentati per mezzo dei valori dei dati stessi. Per esempio, nel caso relazionale il record dell'esame contiene, oltre al voto e alla data, anche il codice del corso e la matricola dello studente, che implementano perciò un riferimento basato su valori.

Negli ultimi 15 anni, sulla scia dei linguaggi di programmazione orientati agli oggetti, è stato studiato anche il modello a oggetti. Il modello a oggetti è un ibrido tra queste due concezioni, in quanto, da una parte, si colloca a un alto livello di astrazione: il concetto di classe e quello di oggetto fanno infatti riferimento a quello di tipo di dato astratto, un'astrazione, appunto, che permette di riunire in sé la definizione della struttu-

ra di dati e quella dei sottoprogrammi che implementano le operazioni. D'altra parte, l'oggetto ha un suo identificatore che permette ai programmi e agli altri oggetti di far riferimento ad esso, e che in un certo senso svolge il ruolo del puntatore esplicito dei modelli reticolare e gerarchico. In questo articolo ci concentreremo sui *linguaggi per basi di dati relazionali*, che costituiscono ormai l'80% dei sistemi di gestione di basi di dati esistenti. A tale scopo, introdurremo brevemente il modello relazionale dei dati.

Il modello relazionale fu proposto nel 1970 da Edgar F. Codd, in un articolo fondamentale¹, per favorire l'indipendenza dei dati, ma reso disponibile come modello logico in DBMS reali nel 1981; non è facile infatti implementare l'indipendenza mantenendo i necessari livelli di efficienza e affidabilità. Esso si basa sul concetto matematico di *relazione*.

In matematica, si dice *relazione* un qualsiasi sottoinsieme del prodotto cartesiano di uno o più insiemi, detti *domini*. Il numero di tali insiemi è detto *grado*, o *arità*, della relazione, mentre il numero dei suoi elementi è la sua *cardinalità*. Consideriamo per esempio tre insiemi: l'insieme M di tutte le possibili matricole di studenti, composto da stringhe alfanumeriche di lunghezza 6, i cui primi due caratteri sono alfabetici e i restanti quattro sono numerici; l'insieme C di tutti i possibili codici di corsi, composto da stringhe numeriche di lunghezza 5, e l'insieme V dei voti possibili, cioè i numeri compresi tra 18 e 31 (che rappresenterà il 30 e lode). Una relazione matematica su $M \times C \times V$ sarà un qualsiasi sottoinsieme di questo prodotto cartesiano, e avrà grado 3. I suoi elementi saranno le terne *ordinate* che al primo posto hanno una matricola, al secondo un codice corso e al terzo un valore compreso tra 18 e 31. Un esempio di relazione è presente in tabella 1.

AB3145	56747	28
AC1455	44567	24
BA3768	98709	30
AG5676	88076	19

TABELLA 1

¹ E. F. Codd: A Relational Model of Data for Large Shared Data Banks. *Commun. ACM*, Vol. 13, n. 6, 1970, p. 377-387.

Nel modello relazionale si apportano alcune variazioni a questo concetto fondamentale. Innanzitutto, una relazione *avrà sempre un numero finito di elementi*. Inoltre, mentre le n -uple di una relazione matematica sono ordinate, cioè l' i -esimo valore di ciascuna proviene dall' i -esimo dominio, nel modello relazionale dei dati si preferisce identificare i domini mediante dei *nomi*, detti *attributi*, che permettono di riferirsi ai valori presenti in una n -pla senza dover conoscere la loro posizione, che diventa irrilevante. Nel modello relazionale, le n -ple così ottenute si chiamano *tuple*. A questo punto, una relazione è un insieme di tuple, e può essere appropriatamente visualizzata come una tabella dove:

1. i valori di ciascuna colonna sono fra loro omogenei (cioè appartengono allo stesso dominio),
2. le righe (tuple) sono diverse fra loro (ricordiamo che si tratta di un insieme),
3. le intestazioni delle colonne sono diverse tra loro, e perciò i campi sono distinguibili mediante il loro nome e non la loro posizione. Sempre pensando al nostro esempio, le due relazioni presenti in tabella 2 e in tabella 3 sono perfettamente equivalenti.

Si vede ora che, come già abbiamo avuto occasione di dire, il modello relazionale è *basato su valori*, cioè i riferimenti fra dati in relazioni diverse sono rappresentati per mezzo di valori dei domini che compaiono nelle tuple. In tal modo si realizza la desiderata indipendenza dalle strutture fisiche, che possono cambiare dinamicamente, si rappresenta solo ciò che è rilevante dal punto di vista dell'applicazione (dell'utente), i dati sono portabili più facilmente da un sistema ad un altro, e i riferimenti sono bidirezionali, contrariamente al caso dei puntatori, che sono direzionali. Ovviamente, a livello fisico le strutture di memorizzazione sfruttano puntatori per quanto necessario all'implementazione efficiente delle operazioni, in modo completamente trasparente all'utente.

Data una relazione R sui domini D_1, \dots, D_n , lo schema di R è dato dal nome R seguito dall'insieme dei suoi attributi A_1, \dots, A_n , e si indica con $R(A_1, \dots, A_n)$. Un insieme di schemi di relazione con nomi diversi:

$R = \{R_1(X_1), \dots, R_n(X_n)\}$ (ove X_i è il vettore degli attributi della relazione R_i) costituisce lo schema di una base di dati. Un'istanza di relazione su uno schema $R(X)$ è un insieme di tuple su X , mentre un'istanza di base di dati su uno sche-

Codice corso	Matricola	Voto
56747	AB3145	28
44567	AC1455	24
98709	BA3768	30
88076	AG5676	19

TABELLA 2

Matricola	Codice corso	Voto
AB3145	56747	28
AC1455	44567	24
BA3768	98709	30
AG5676	88076	19

TABELLA 3

ma $R = \{R_1(X_1), \dots, R_m(X_m)\}$ è un insieme di m istanze di relazioni sugli elementi di R .

Un altro concetto molto importante è quello di *chiave di una relazione*: si tratta di un insieme (minimale) di attributi che identifica univocamente le tuple di quella relazione. Una piccolissima base di dati relazionale, contenente informazioni su studenti, corsi ed esami, potrebbe avere il seguente schema, dove gli attributi sottolineati costituiscono le chiavi:

CORSI (CodiceCorso, NomeCorso, Docente, Semestre, N_Crediti)

STUDENTI (MatricolaStudente, Nome, Cognome, Indirizzo, CorsoDiLaurea, Anno)

ESAMI (MatricolaStudente, CodiceCorso, Voto)

4. LINGUAGGI DI INTERROGAZIONE PER BASI DI DATI RELAZIONALI

Come si è detto in precedenza, uno dei compiti più importanti del DBMS è permettere ai suoi utenti, siano essi umani o programmi, di ritrovare i dati mediante i *linguaggi di interrogazione*, che consentono di accedere in modo diretto ai dati *basandosi sulle loro proprietà*.

Supponiamo di essere interessati a trovare i nomi e cognomi di tutti gli studenti che hanno ottenuto almeno 28 all'esame di Fondamenti di Informatica. Piuttosto che cercare nella tabella CORSI il codice di Fondamenti di Informatica, e poi scandire i blocchi del file ESAMI, alla ricerca dei voti maggiori di o uguali di 28 che corrispondono al codice trovato, e poi ancora cercare nella tabella

STUDENTI i nomi e cognomi corrispondenti alle matricole rintracciate, un utente certamente preferisce semplicemente enunciare, in un opportuno linguaggio: “voglio trovare tutti gli studenti che hanno ottenuto almeno 28 all’esame di Fondamenti di Informatica”, così *dichiarando* quale proprietà i dati richiesti debbano soddisfare.

Oltre che per interrogare, si ha bisogno di interagire con la base di dati per altre due ragioni fondamentali. Da una parte, dobbiamo essere in grado di creare, modificare ed eventualmente cancellare *componenti dello schema*, dall’altra, invece, dobbiamo essere in grado di inserire, modificare e cancellare *tuple*, cioè *l’istanza* della base di dati. Qualunque sia il modello dei dati sottostante, i linguaggi per basi di dati sono dunque tradizionalmente composti da due ingredienti fondamentali: il *Linguaggio di Definizione dei Dati (Data Definition Language, DDL)* e il *Linguaggio di Manipolazione dei Dati (Data Manipulation Language, DML)*, operanti, rispettivamente, sullo schema e sull’istanza della base di dati. Il linguaggio di interrogazione, che anch’esso opera sull’istanza, è parte del DML.

Il linguaggio di interrogazione costituisce però la parte più importante di questo complesso, poiché l’attività di interrogare è di gran lunga la più frequente, e richiede di realizzare i meccanismi per *accedere* ai dati, necessari anche a tutte le operazioni di modifica. Questa è quindi la parte che richiede un maggiore sforzo da parte dei progettisti del linguaggio, sia dal punto di vista dell’ottimizzazione, sia per concepire linguaggi la cui semantica sia ben chiara all’utente.

Un linguaggio di interrogazione *dipende dal modello logico dei dati*: si intuisce infatti che, per quanto dichiarativa, la struttura della interrogazione debba far riferimento alla percezione che l’utente ha della struttura dei dati. Per i modelli logici basati su riferimenti, come quello gerarchico, il reticolare, o il modello a oggetti, l’utente dovrà comunque, nello scrivere una interrogazione, far riferimento ai legami tra gli oggetti; viceversa, in un modello basato su valori, come quello relazionale, poiché tali legami sono espressi attraverso i dati stessi, essi vengono introdotti come condizioni sui dati, analogamente alle proprietà che i dati debbono soddisfare. Per esempio, la specifica dell’interrogazione “trovare tutti gli studenti che hanno ottenuto almeno 28 all’esame di Fondamenti di Informatica” conterrà anche l’indicazione che il codice dell’esame cercato nella tabella ESAMI dovrà

essere uguale al codice del corso di Fondamenti di Informatica trovato nella tabella CORSI.

Lo sforzo dei ricercatori e dei progettisti di linguaggi di interrogazione si è da lungo tempo concentrato sul concepire linguaggi *dichiarativi*, linguaggi cioè che mascherano i dettagli dei meccanismi di esecuzione permettendo all’utente di trascurare l’aspetto implementativo e concentrarsi sulla natura dei dati ricercati. Per questo motivo, al lavoro dei progettisti di DBMS si è associato quello di generazioni di ricercatori, che hanno escogitato linguaggi *formali* di interrogazione, che, pur mantenendo l’equivalenza con i costrutti dei linguaggi adottati dai DBMS commerciali, si lasciassero studiare in termini di proprietà matematiche (logiche o algebriche), prestandosi a ragionamenti non ambigui. A questi linguaggi sarà specificamente dedicato un altro articolo di questa serie.

Si noti che *il risultato di qualsiasi interrogazione a una base di dati relazionale è sempre una relazione*, cioè un insieme di tuple che soddisfa le proprietà specificate.

Tra i linguaggi commerciali per basi di dati, ne sono stati concepiti di due categorie fondamentali. La prima comprende i *linguaggi testuali*, che specificano l’interrogazione in termini di stringhe di testo, analogamente alla maggior parte dei linguaggi di programmazione. Esempio principe di questa categoria è il linguaggio relazionale SQL (*Structured Query Language*), definito presso i laboratori IBM di San Jose nella sua prima versione negli anni ’70 per il primo prototipo di DBMS relazionale, *System R*.

Un esempio di specifica dell’interrogazione di prima in SQL è il seguente:

```
SELECT Nome, Cognome
FROM STUDENTI,ESAMI,CORSI
WHERE STUDENTI.Matricola=ESAMI.Matricola AND
ESAMI.CodiceCorso=CORSI.CodiceCorso AND
CORSI.NomeCorso="Fondamenti di Informatica"
and ESAMI.Voto>=28.
```

SQL è un linguaggio molto potente, in grado di esprimere interrogazioni molto sofisticate ed anche alcune forme elementari di aggregazione, come la media, la somma, e funzioni semplici come il minimo e il massimo di un insieme. Bisogna però stare attenti a non confondere il potere espressivo dei linguaggi di interrogazione per basi di dati con quello dei linguaggi di program-

mazione. Da questi ultimi, infatti, si richiede un potere espressivo pari a quello della macchina di Turing, poiché mediante un tale linguaggio occorre poter esprimere tutte le funzioni computabili. Da un linguaggio di interrogazione, invece, si richiede di essere in grado di esprimere *tutte le possibili condizioni* sui dati stessi, e perciò il confronto di espressività si fa di solito con la *logica del primo ordine*. Senza scendere in dettagli formali, per i quali si rinvia il lettore interessato ai libri di testo di basi di dati elencati in bibliografia, basti sapere che, nel caso dei linguaggi classici per basi di dati relazionali, ci si accontenta di un potere espressivo inferiore (per esempio non è possibile esprimere cicli), e per le computazioni si ricorre all'*immersione* delle interrogazioni in programmi tradizionali. Nasce così il concetto di linguaggio "embedded", espressione che indica il fatto che l'interrogazione viene inserita all'interno di un programma scritto in un linguaggio di programmazione "ospite", come ad esempio C o Java. Ciò accade quando l'interrogazione non viene effettuata da un utente in modo interattivo, ma è inserita in un programma applicativo che interagisce con la base di dati.

Una categoria di linguaggi commerciali interessante, ma utile solo in determinati casi, è quella dei linguaggi *grafici o visuali*, che forniscono interfacce "amichevoli", basate su una visualizzazione della forma che la tabella risultato dell'interrogazione dovrebbe avere. Un esempio di tali linguaggi è *Query By Example* (QBE), la cui più famosa realizzazione commerciale è Microsoft Access[®]. Come si può intuire, oltre al difetto di non poter essere immersi in linguaggi di programmazione, questi linguaggi hanno il problema di essere, sì, molto facili da usare nel caso di interrogazioni relativamente semplici, ma di avere una semantica poco intuitiva nel caso di interrogazioni complesse, e pertanto di essere in generale proni ad errori.

Nonostante la potenza dei linguaggi di interrogazione relativamente alle operazioni di specifica di dati da cercare, molti ricercatori non sono comunque soddisfatti del loro potere espressivo. Il motivo principale di ciò è un'anomalia che insorge nell'uso di questi linguaggi nei programmi applicativi, quando vengono immersi nei linguaggi di programmazione. Si è detto che la risposta a qualsiasi interrogazione relazionale è una relazione, il che si esprime dicendo che i linguaggi di interrogazione relazionali sono *orientati agli in-*

siemi; i linguaggi di programmazione tradizionali gestiscono però i record uno alla volta (sono, come si dice, *orientati alle tuple*), cioè un programma interagente con la base di dati non "si aspetta" che la risposta alla interrogazione sia un insieme, ma un semplice valore, o un record.

Supponiamo ad esempio di voler inserire la nostra interrogazione in un programma applicativo che stampi l'elenco degli "Studenti Bravi", definendo come bravi coloro che hanno ottenuto almeno 28 all'esame di Fondamenti di Informatica. Il programma non è in grado di ricevere e stampare direttamente l'intero set di studenti che soddisfano la proprietà richiesta, ma dovrebbe contenere un ciclo che li stampa uno per volta; questo problema viene chiamato *disaccoppiamento di impedenza* (*impedance mismatch*).

La soluzione comunemente adottata è l'uso dei *cursori*, un concetto che fa parte da tempo dell'SQL standard. Un cursore accede al risultato di una interrogazione in modo "set-oriented", e restituisce le tuple al programma una per volta. Il frammento di programma seguente svolge l'operazione richiesta.

...

```
exec sql begin declare section;
char Nome[20], Cognome[20];
exec sql end declare section;
exec sql declare StudentiBravi cursor for
```

```
SELECT Nome, Cognome
FROM STUDENTI,ESAMI,CORSI
WHERE STUDENTI.Matricola=ESAMI.Matricola AND
ESAMI.CodiceCorso=CORSI.CodiceCorso AND
CORSI.NomeCorso="Fondamenti di Informatica" AND
ESAMI.Voto>=28
```

```
exec sql open StudentiBravi;
exec sql fetch StudentiBravi into :Nome,:Cognome;
printf("StudentiBravi");
while (sqlca.sqlcode == 0)
{
printf("Nome Studente: %s %s ",Nome,Cognome);
exec sql fetch StudentiBravi into :Nome,:Cognome;
}
exec sql close StudentiBravi;
}
...
```

Il disaccoppiamento di impedenza è un problema interessante, la cui soluzione pratica non

soddisfa molti ricercatori, e che ha spinto la comunità delle basi di dati a svolgere molta ricerca. Alcuni tentativi di soluzione saranno discussi nell'articolo sui linguaggi formali per basi di dati; invece, per quanto riguarda i linguaggi commerciali, una possibile soluzione è costituita dai linguaggi e i DBMS orientati agli oggetti. Questi sistemi offrono alle aree applicative più avanzate, servite talvolta in modo insoddisfacente dai sistemi tradizionali, vantaggi indubbi, sia dal lato della ricchezza di strutture messe a disposizione dal modello dei dati, sia da quello della capacità semantica di modellare dati e comportamenti in maniera compatta. Per esempio, nei DBMS orientati agli oggetti il disaccoppiamento di impedenza viene superato dal fatto che il paradigma del linguaggio di programmazione è lo stesso del linguaggio di interrogazione, perciò o sono entrambi orientati alle tuple o entrambi orientati agli insiemi. D'altra parte, mentre negli ultimi anni i linguaggi di programmazione orientati agli oggetti, in particolare Java, si sono diffusi nell'uso comune, proprio a causa della loro estrema flessibilità e capacità di modellazione, i sistemi di gestione di basi di dati orientate agli oggetti non hanno acquistato lo stesso peso nel mondo dell'informatica. Essi presentano infatti una difficoltà di implementazione che supera di gran lunga quella di un linguaggio di programmazione: in aggiunta alla complessità del linguaggio stesso, un sistema di gestione di basi di dati deve gestire le informazioni in memoria persistente, e quindi risolvere gli usuali problemi di efficienza, consistenza, di condivisione di dati, di sicurezza, di privacy, di eventuale distribuzione dei dati su più siti ecc.. Anche le difficoltà teoriche non sono banali: si pensi per esempio al concetto di incapsulamento nei linguaggi di programmazione orientati agli oggetti, e a come

questo contrasti proprio con il principio fondante i linguaggi di interrogazione, che è quello di poter reperire tutte le informazioni necessarie in modo diretto, basandosi su proprietà logiche, senza passare per uno specifico metodo. Un altro punto teorico un po' critico è costituito dal ritorno, nei linguaggi a oggetti, del collegamento tra concetti basato su riferimenti, che contraddice l'approccio dichiarativo così caro agli specialisti delle basi di dati.

5. CONCLUSIONI

Per concludere, vorremmo far notare l'importanza della comprensione, per tutti gli informatici, dei concetti fondamentali delle basi di dati e dei loro linguaggi. Questo mondo, infatti, mentre da una parte condivide molti principi di quello dei linguaggi di programmazione, spesso se ne discosta per le necessità specifiche dell'obiettivo che si propone: si pensi alla necessità di linguaggi dichiarativi, strettamente legata all'esigenza di mantenere l'indipendenza fisica, o alle necessità di ottimizzazione affatto diverse da quelle di ottimizzazione dei programmi.

Bibliografia

Elencare i riferimenti ai concetti, modelli e linguaggi introdotti comporterebbe una lista infinita; indichiamo qui soltanto alcuni testi classici sulle basi di dati, dai quali il lettore potrà desumere ulteriori riferimenti bibliografici.

Atzeni P., Ceri S., Paraboschi S., Torlone R.: *Basi di dati: modelli e linguaggi di interrogazione*. 2^a edizione, Mc Graw Hill Italia, 2006.

Elmasri R., Navathe S.: *Sistemi di basi di dati. Fondamenti*. 4^a edizione. Pearson/Addison Wesley, 2004.

LETIZIA TANCA è Professore Ordinario di Basi di Dati presso il Politecnico di Milano, e presidente del Consiglio di Corso di Studi di Ingegneria Informatica del Politecnico di Milano, campus Milano Leonardo. 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 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, le basi di dati per piccoli dispositivi.
E-mail: tanca@elet.polimi.it