

IL COMPUTER GIOCA A SCACCHI

La storia delle macchine che giocano a scacchi è ancor più antica di quella del computer. Infatti la prima macchina scacchistica venne esibita nel 1770 a Vienna: era il Turco, un famoso automa che giocava con l'inganno, perché celava abilmente una persona tra i suoi ingranaggi. In questo articolo viene delineata la storia più recente delle macchine che giocano a scacchi e i principali aspetti di ricerca ad esse correlati.

1. INTRODUZIONE

Per valutare l'efficacia di un dispositivo hardware ne analizziamo le caratteristiche tecniche: possiamo per esempio misurare la velocità di calcolo di un processore, la dimensione di una memoria centrale o periferica, la risoluzione di un monitor, la banda di comunicazione disponibile su una connessione. Ma come si valuta l'efficacia di un software? Per esempio, come possiamo confrontare due sistemi operativi? Il confronto tra Windows e Linux di solito avviene su basi ideologiche, dunque fortemente soggettive. E comunque, come si dimostra che la versione 2.0 di un certo programma è migliore della 1.0? Non stiamo parlando di contare gli errori; supponiamo che entrambe le versioni siano prive di errori: come si confronta la loro "efficacia"?

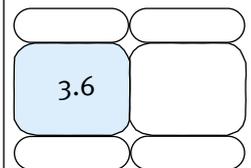
Nel caso dei programmi che giocano a scacchi, la risposta è semplice: tra due programmi, è migliore quello che batte l'altro. In figura 1, tratta da [11], vediamo un grafico che mostra l'evoluzione di questi programmi. In ascissa ci sono tre scale correlate che descri-

vono rispettivamente la profondità media di analisi dell'albero di gioco (in *ply*, ossia semiosse), il numero medio di posizioni analizzate per secondo, i MIPS offerti dall'hardware. In ordinata abbiamo la scala Elo, che descrive la forza di un giocatore di Scacchi. La scala prende il nome da A. Elo, uno studioso di statistica, che tra il 1960 ed il 1970 mise a punto un sistema di valutazione della forza di gioco che è tutt'oggi usato in tutto il mondo. Per esempio, il campione del mondo "vale" di solito circa 2800 punti, un principiante vale 800 punti, mentre la curva di Gauss che descrive la "popolazione scacchistica" mondiale ha una mediana intorno ai 1400. Le variazioni di punteggio avvengono in seguito al conteggio delle vittorie e delle sconfitte in tornei ufficiali. La figura mostra le date di nascita e la forza relativa dei principali programmi di gioco, e culmina con Deep Blue, una macchina creata da IBM nei laboratori di Yorktown appositamente per battere il campione del mondo.

In effetti, Domenica 11 maggio 1997 è una data che gli storici dell'informatica segnano



Paolo Ciancarini



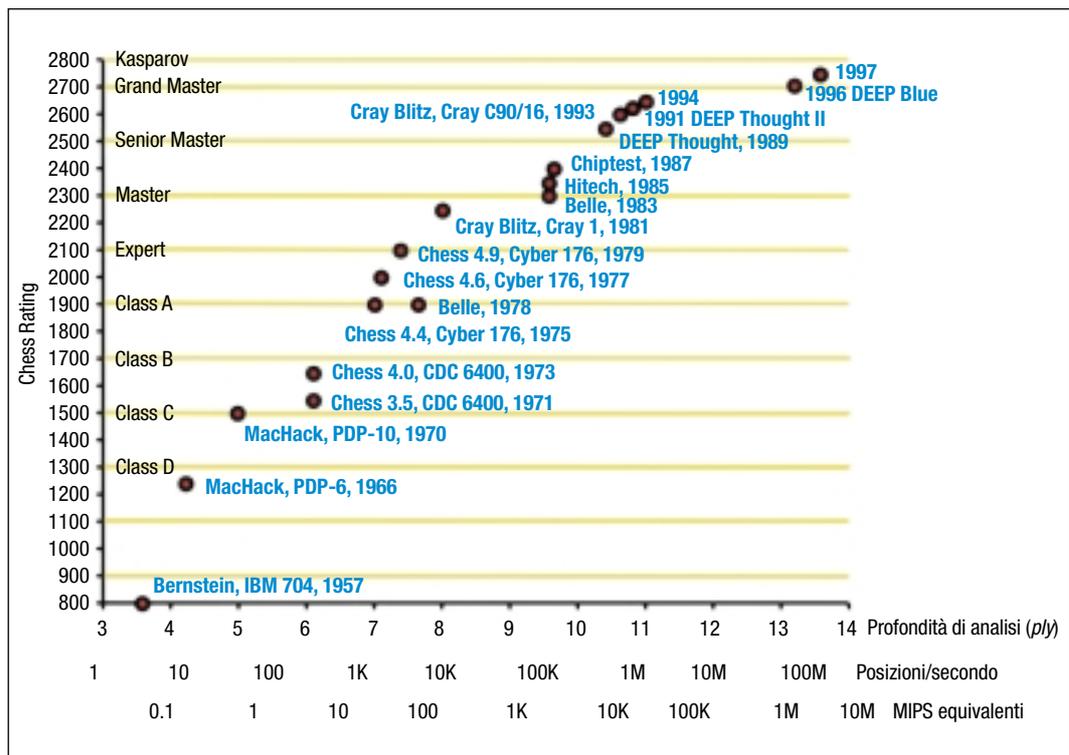


FIGURA 1
Il progresso
dei programmi
di gioco

con un sassolino bianco. Quel giorno Deep Blue ha sconfitto il campione del mondo di scacchi, il russo Garry Kasparov, dopo un match di sei partite (due vinte, una persa e tre patte). L'evento è stato seguito e commentato in diretta da decine di migliaia di persone connesse a Internet [10]. Tale impresa va considerata come uno dei grandi successi scientifici di questo secolo, cui hanno contribuito in varia misura sin dalla fine degli anni '40 parecchi studiosi, come Alan Turing, Claude Shannon, Herbert Simon, John McCarthy e Ken Thompson, solo per citarne alcuni.

Per quali ragioni si costruiscono macchine capaci di giocare? La giustificazione scientifica di tali sforzi è stata espressa quasi 50 anni fa in un articolo di Newell, Shaw e Simon:

“Gli Scacchi sono il gioco intellettuale per eccellenza. Senza far uso di strumenti casuali (come i dadi o la roulette), che inquinerebbero la contesa, due intelletti vengono contrapposti in una situazione così complessa che nessuno dei due può sperare di comprenderla completamente, ma sufficientemente analizzabile di modo che ciascuno dei due può sperare di sconfiggere l'altro. Il gioco è tanto profondo e sottile che ha per-

messo la nascita di giocatori professionisti, ed ha sopportato senza esaurirsi 200 anni di partite e di studi analitici intensivi. Tali caratteristiche rendono gli Scacchi un'arena naturale per i tentativi di meccanizzazione. *Se si potesse sviluppare un giocatore artificiale vincente, si potrebbe affermare di aver penetrato il nucleo dell'attività intellettuale umana.*”

Gli Scacchi permettono di studiare le procedure astratte che usano gli esseri umani quando prendono decisioni. Se riusciamo a far giocare bene un computer, possiamo sperare di insegnargli a ragionare “come noi”, e forse meglio, grazie alla grande velocità con cui la macchina elabora immense quantità di dati. Alcuni problemi pratici che avrebbero tratto giovamento dallo studio del progetto di giocatori artificiali vennero elencati da Shannon in [15]:

- macchine capaci di progettare componenti elettronici;
- macchine capaci di sostituire le centrali telefoniche elettromeccaniche;
- macchine capaci di tradurre frasi da una lingua in un'altra;
- macchine capaci di decisioni strategiche in campo militare o economico;

I macchine capaci di orchestrare una melodia;
 I macchine capaci di deduzione logica.
 È davvero mirabile osservare come quasi tutti gli obiettivi proposti da Shannon siano stati raggiunti, e questo grazie anche al contributo delle ricerche sui giocatori artificiali.

2. PICCOLA STORIA

Già all'inizio degli anni '50 Shannon [15] e Turing [17] avevano descritto algoritmi per giocare a scacchi, ma non avevamo macchine adatte a realizzarli. Verso il 1955 A. Newell e H. Simon della *Carnegie Mellon University* iniziarono la progettazione del programma CP-1 (*Chess Player 1*). Il loro scopo era quello di costruire una macchina "intelligente", capace di dimostrare teoremi matematici o di progettare sistemi complessi. Newell e Simon decisero che gli scacchi fornivano un buon terreno di prova per progettare una "macchina intelligente". Nella sua autobiografia Simon racconta di come una mattina del gennaio del 1956 entrò in classe annunciando ai suoi studenti: "Durante le vacanze di Natale io e Newell abbiamo progettato una macchina che pensa" [16]. Negli stessi giorni scrive:

"... Allen Newell e io abbiamo fatto progressi sostanziali sulla macchina che gioca a scacchi, solo che per il momento non sa giocare a scacchi ma solo cercare e scoprire dimostrazioni di teoremi di logica simbolica ...".

Il programma che avevano progettato si chiamava *Logic Theorist* ed era scritto in IPL-2, un linguaggio di programmazione appositamente sviluppato. Il programma di scacchi venne effettivamente scritto solo qualche anno dopo, in collaborazione con C. Shaw, per cui venne chiamato *NSS*. Il programma di gioco era abbastanza sofisticato, ma l'uso di un linguaggio sperimentale di programmazione penalizzò talmente le prestazioni della macchina che occorreva un'ora per fare una mossa. A causa della sua lentezza, il programma venne usato per giocare una sola partita, ma comunque entusiasmo i suoi autori che ottimisticamente profetizzarono che entro 10 anni il Campione del Mondo sarebbe stato un giocatore artificiale.

Un altro programma famoso fu scritto al MIT di Boston da A. Bernstein e altri alla fine degli anni '50. Il programma di Bernstein usava una strategia che Shannon definì di tipo B: sceglieva in una posizione le sette mosse migliori usando alcuni principi generali, analizzava per ciascuna mossa le sette migliori risposte, e ripeteva il procedimento per ciascuna delle 49 posizioni così ottenute. Il programma veniva eseguito da un elaboratore IBM 704 ed impiegava circa otto minuti per analizzare $7^4 = 2401$ varianti prima di effettuare una mossa. L'esame di una singola posizione comportava l'effettuazione di 7.000 operazioni di macchina, che realizzavano alcune semplici valutazioni strategiche: bilancio del materiale, misura della mobilità dei pezzi, analisi del controllo dello spazio e valutazione del grado di sicurezza del Re. Il programma giocava come un principiante, ma venne enormemente pubblicizzato negli Stati Uniti, contribuendo a far conoscere al grosso pubblico scopi e prospettive dell'Intelligenza Artificiale.

Il MIT per diversi anni fu il più importante centro di ricerca sul gioco artificiale. Il principale ricercatore impegnato nel tentativo di costruire macchine intelligenti era J. McCarthy. Egli aveva ottenuto la disponibilità di uno dei primi mainframe IBM, un modello 704, che i suoi studenti programmavano usando un nuovo linguaggio, il LISP, inventato da McCarthy stesso. L'interesse di McCarthy nei programmi di scacchi era puramente teorico, ma alcuni suoi studenti si entusiasmarono del progetto e lo portarono avanti. Nel 1962 A. Kotok, uno di tali studenti, terminò di scrivere un programma di scacchi e lo presentò come tesi di laurea. Per ragioni di efficienza il programma era scritto in FORTRAN e assembler. Il programma di Kotok venne ulteriormente sviluppato da McCarthy che si era trasferito a Stanford, in California, e nel 1967 divenne il rappresentante americano in una famosa sfida contro una macchina sovietica. I sovietici avevano iniziato le loro ricerche da alcuni anni, e si avvalevano dell'aiuto di alcuni esponenti della loro grande scuola scacchistica. Era la prima volta che le ricerche sovietiche venivano allo scoperto, e il risultato fece sensazione: l'URSS vinse per 3-1 quella sfida, con due vittorie e due patte. L'incontro fu

particolarmente interessante dal punto di vista scientifico, perché i due programmi usavano strategie di ricerca opposte. Il programma sovietico usava una strategia di Shannon di tipo A (o “forza bruta”): esplorava tutte le varianti possibili fino alla profondità di 5 semimosse. Il programma americano usava una ricerca di tipo B: non tutte le varianti possibili venivano esplorate, ma solo quelle definite “plausibili” da una funzione euristica.

Negli anni della corsa alla Luna, una sconfitta tecnologica di questo genere venne vista negli USA come una vera e propria onta nazionale. Il risultato dell’incontro stimolò nuove ricerche, e R. Greenblatt costruì poco dopo una nuova macchina. Il programma si chiamava MacHack, usava un Digital PDP-6 del MIT ed ebbe l’onore di essere il primo giocatore artificiale a debuttare in un torneo a Boston nel 1967. La prestazione di MacHack al torneo fu

pessima. Tuttavia, col suo debutto agonistico la lunga corsa verso il Campionato del Mondo era davvero cominciata: infatti al programma fu permesso di iscriversi alla federazione scacchistica americana (USCF), stabilendo così un precedente. A parte il risvolto commerciale, è importante che partecipando a tornei ufficiali un programma ottiene un punteggio Elo. MacHack raggiunse una valutazione Elo pari a 1500 punti: il livello di un dilettante di media forza. MacHack non riuscì mai a giocare bene, ma quando fu mostrato ad una conferenza che si tenne ad Edimburgo nel 1968 ebbe il merito di attirare l’attenzione di un giocatore scozzese, D. Levy, che era anche uno studioso di Intelligenza Artificiale. MacHack aveva scatenato l’euforia dei congressisti presenti, alimentando rosee speranze per il futuro. Fu allora che Levy, colpito dall’entusiasmo ingiustificato che avevano acceso le mediocri

Tecniche di Intelligenza Artificiale applicata agli scacchi

Nei programmi di scacchi sono state messe alla prova parecchie tecnologie tipiche dell’Intelligenza Artificiale.

Rappresentazione della conoscenza. La conoscenza scacchistica è di vario tipo. Una forma di conoscenza *dichiarativa* riguarda la rappresentazione dei pezzi e della scacchiera, e poi la formazione di concetti in base ai principi (mettere al sicuro il Re) e la conseguente definizione di piani e tattiche di gioco (arrocca, per mettere al sicuro il Re). Una forma di conoscenza *procedurale* riguarda invece il modo in cui si calcolano le conseguenze di una serie di scambi di pezzi (conviene sempre iniziare gli scambi dal pezzo di valore minore). Per rappresentare questi tipi di conoscenze sono state usate varie tecniche, in particolare per la conoscenza dichiarativa la logica e i linguaggi logici come Prolog, mentre la conoscenza procedurale è stata rappresentata mediante euristiche, di solito all’interno della funzione di valutazione.

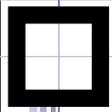
Esplorazione dello spazio degli stati. Esistono diversi algoritmi di esplorazione dello spazio degli stati di una partita a scacchi. Sono stati infatti inventati raffinamenti algoritmici che si comportano mediamente meglio di Alfabetà, ed altri se ne continuano ad inventare.

Machine Learning. L’apprendimento automatico viene usato per aumentare le conoscenze dei programmi specie nella fase di apertura. I cosiddetti libri di aperture vengono da tempo creati in modo automatico, a partire da partite giocate da Grandi Maestri. È interessante notare che ci sono stati molti tentativi di costruire programmi che apprendono giocando contro se stessi, ma sono tutti falliti (nessuno ha ottenuto un giocatore efficace usando questa tecnica).

Pattern recognition. Una tecnica importante in alcuni programmi è il riconoscimento di schemi di gioco, di solito allo scopo di applicare una strategia che in altre partite ha dato buona prova. Il riconoscimento avviene definendo prima degli schemi (in inglese: pattern, o chunks) associati a una o più euristiche da applicare quando viene rilevato lo schema nella posizione in analisi.

Algoritmi genetici. Questa tecnica permette di far evolvere i programmi di gioco, raffinandone le funzioni di valutazione. Vengono create alcune varianti della funzione di valutazione, e poi le si fanno “competere” usando un insieme di posizioni di test. Le funzioni migliori (quelle che risolvono meglio le posizioni di test) vengono poi leggermente alterate in modo casuale, ripetendo il ciclo. Alla lunga (dopo migliaia di cicli) si ottiene una sorta di “selezione naturale” evolutiva delle funzioni di valutazione. Gli algoritmi genetici hanno dato buona prova nel miglioramento delle funzioni di valutazione.

Reti neurali. Esistono programmi le cui funzioni di valutazione sono reti neurali. Tali programmi, giocando o osservando le partite di altri, sono in grado di sviluppare una “teoria subsimbolica” (cioè non basata su conoscenza dichiarativa) degli scacchi e di metterla alla prova. Questa tecnologia non ha dato buoni risultati nel caso degli scacchi.



prestazioni delle macchine che aveva esaminato, scommise che nessun programma sarebbe riuscito a batterlo nei 10 anni seguenti. Sfidare Levy divenne un punto d'onore per tutti i principali ricercatori. La scommessa venne accettata da due ricercatori presenti alla conferenza, J. McCarthy e D. Michie. Nel corso degli anni la scommessa venne ribadita ed altri ricercatori scommisero contro Levy; nel 1976 valeva in totale 1.250 sterline, che Levy vinse facilmente battendo nel 1978 il miglior programma dell'epoca.

La sfida di Levy ebbe il merito di spronare i membri della comunità scientifica, che decisero di incontrarsi annualmente per misurare i progressi conseguiti. Nel 1970 a New York iniziò infatti una competizione riservata ai giocatori artificiali che continua ogni anno ancora oggi, raccogliendo sempre una partecipazione molto qualificata: il torneo dell'*Association for Computing Machinery (ACM)*, in seguito denominato torneo *NACC (North American Computer Championship)*. Nel 1974, a Stoccolma, venne introdotta su suggerimento di Levy una nuova manifestazione ufficiale: il Campionato del Mondo per giocatori artificiali. La novità del torneo di Stoccolma fu la partecipazione di un programma sovietico, *Kaissa*, diretto discendente del programma che aveva vinto la sfida USA-URSS del 1967. *Kaissa* era stato scritto da M. Donsky e V. Arlazarov, dell'Istituto per le Scienze Sistemiche di Mosca. Ebbe l'onore di diventare il primo giocatore artificiale a fregiarsi ufficialmente del titolo di Campione del Mondo, vincendo tutte le partite. Questa prima affermazione però non venne seguita da altri successi. Infatti tre anni dopo, a Toronto, *Kaissa* non riuscì a ripetersi e giunse solo secondo dopo il programma americano *Chess 4.6*. Andò anche peggio nel 1980, nel torneo di Campionato del Mondo svoltosi a Linz: *Kaissa* giunse solo sesto, perché l'hardware impiegato era diventato obsoleto rispetto alle macchine più moderne a disposizione dei suoi competitori. La sconfitta fu dovuta soprattutto al fatto che il tempo di macchina in URSS era costosissimo, a causa della scarsità di calcolatori, quindi gli autori del programma non avevano potuto migliorarlo quanto i suoi concorrenti. *Kaissa* sparì del tutto dalle competizioni, e da allora nessun

programma sovietico ha più partecipato ad alcuna competizione ad alto livello.

3. LA NUOVA GENERAZIONE: MACCHINE BASATE SU HW SPECIALE

In quegli stessi anni venivano gettate le basi teoriche e tecnologiche per la costruzione della nuova generazione di programmi di gioco, che cominciò a dominare i tornei all'inizio degli anni '80. Le novità tecnologiche più importanti della fine degli anni '70 furono due: l'introduzione del microprocessore, ed i grandi miglioramenti delle tecniche di progettazione di hardware specializzato (VLSI). L'invenzione del microprocessore mise alla portata di tutte le borse un calcolatore personale; nel 1974 esistevano già i primi microprocessori per hobbysti. Nel 1977 venne commercializzata la prima scacchiera elettronica (Figura 2): si chiamava *Chess Challenger*, era stata progettata da R. Nelson e venne prodotta dalla *Fidelity International*. Costava poche centinaia di dollari e venne venduta in decine di migliaia di esemplari. Per comprendere il significato economico di questa invenzione, basterà riportare alcune cifre. I sedici computer partecipanti al campionato mondiale di Toronto nel 1977 erano tutti grossi calcolatori; venne calcolato che il loro costo totale era di 40 milioni di dollari, e un'ora di gioco costava 10.000 dollari. Al



FIGURA 2
Fidelity Chess Challenger, la prima scacchiera elettronica

0

campionato del 2004 hanno partecipato invece normali personal computer che costano al più qualche migliaio di dollari.

L'altra grande novità che rivoluzionò profondamente la ricerca sui giocatori artificiali fu l'introduzione di hardware specializzato per la generazione di mosse. Mentre i primi giocatori artificiali utilizzavano calcolatori "normali", ovvero di uso generale, ad un certo punto si cominciò a costruire hardware progettato specificatamente per giocare. Nel 1977 uno studente di Berkeley, O. Babaoglu (oggi docente dell'Università di Bologna) progettò un circuito capace di funzionare come generatore di mosse. Poco dopo venne progettata nei Laboratori Bell da J. Condon e K. Thompson una nuova macchina, chiamata Belle. Nel 1972 i due ricercatori avevano scritto per uno dei primi sistemi Unix un programma che si era comportato senza infamia e senza lode ai vari tornei cui aveva partecipato. Dopo la pubblicazione della tesi di Babaoglu i due ricercatori decisero che valeva la pena di sviluppare una macchina concepita appositamente per gli scacchi, in cui lo sviluppo delle varianti veniva calcolato direttamente da alcuni circuiti in hardware.

Belle conseguì rapidamente in tornei ufficiali della USCF il grado di Maestro di scacchi (2200 punti Elo). Una prima pietra miliare nella storia del gioco artificiale era stata raggiunta.

I successi di Belle, e la simultanea nascita di un ricco mercato per le macchine commerciali, diedero finalmente una certa credibilità sportiva alle ricerche sul gioco artificiale. Tra il 1982 ed il 1985 vennero organizzate a Milano da M. Somalvico e B. Pernici tre conferenze sui giocatori artificiali intitolate *L'Intelligenza Artificiale ed il Gioco degli scacchi*. A Londra venne creata la *International Computer Chess Association* (ICCA), un'associazione di scienziati che si interessano di informatica scacchistica. Oggi ICCA ha cambiato nome in ICGA (*International Computer Games Association*) e si occupa di organizzare manifestazioni agonistiche per giocatori artificiali. È a cura dell'ICGA l'organizzazione delle conferenze *Computer Games* e *Advances in Computer Games*.

All'inizio degli anni '80 i progettisti di giocatori artificiali cominciarono ad esplorare le possibilità di una nuova tecnologia, quella

del *calcolo parallelo*. L'idea del calcolo parallelo è ingannevolmente semplice: due processori sono meglio di uno perché possono effettuare una doppia quantità di calcoli. In realtà esistono grandi problemi di coordinamento delle operazioni: in molti casi non si ottiene nessun guadagno usando più calcolatori invece di uno solo, perché essi perdono molto tempo a cercare di mettersi d'accordo! Nel 1983 a New York il Campionato del Mondo fu vinto da Cray Blitz, un programma che si avvantaggiò della potenza del calcolatore parallelo Cray-1, che era allora il sistema di elaborazione più potente del mondo. Cray Blitz fu il secondo giocatore artificiale capace di guadagnarsi il titolo di Maestro, e dominò la scena per la prima metà degli anni '80.

Sia Belle che Blitz usavano programmi basati sul concetto di *espansione cieca dell'albero di gioco* (strategia A): gli anni '70 erano stati dominati da programmi basati su espansione euristica, ma ormai l'hardware era diventato così veloce da permettere lunghe analisi per forza bruta. Nel 1985 però si affacciò sulla scena Hitech, un programma di H. Berliner, che si basava ancora su euristiche. Sin dal suo apparire Hitech guadagnò una valutazione record di 2220 punti Elo Usa, e subito sconfisse Cray Blitz. L'anno successivo, nel 1986, Hitech aveva già 2360 punti Elo. Per qualche tempo sembrò che l'approccio "intelligente" scelto da Berliner fosse quello giusto, lento ma sicuro verso il Campionato del Mondo assoluto. E invece nel 1987 apparve una nuova macchina basata sul principio dell'analisi per forza bruta, Chiptest. Era stata progettata da uno studente cinese, FH. Hsu, in collaborazione con uno studente di Berliner, M. Campbell. L'anno successivo Hsu e Campbell battezzarono il nuovo prototipo di ChipTest con un nuovo nome: Deep Thought, dal nome di un personaggio di un libro di fantascienza. Analizzando 700.000 posizioni al secondo grazie al suo generatore hardware, Deep Thought era in grado di analizzare tutte le varianti possibili ad una profondità media di 10 semimosse in apertura (con un libro di aperture estremamente ridotto) e 9 semimosse nel mediogioco.

Nel 1990 sia Hsu che Campbell vennero assunti dai laboratori di ricerca di IBM a Yorktown, e lì svilupparono la macchina che oggi conosciamo col nome di Deep Blue.

4. ANALISI DEL PROBLEMA

Analizziamo brevemente i dati del problema, e vediamo come è stato “risolto” dai progettisti di Deep Blue. Nel gioco degli scacchi:

■ ci sono due avversari che alternano le mosse e conoscono in ogni istante le stesse informazioni, cioè hanno *informazione completa* sullo “stato del gioco”;

■ ad ogni turno di gioco le mosse ammesse dalle regole sono in numero *limitato* e ben definite;

■ ogni partita *termina* con la vittoria di uno dei due giocatori, oppure in parità.

Nella posizione iniziale del gioco degli scacchi si hanno 20 mosse possibili per il Bianco. Per ogni mossa del Bianco il Nero può rispondere in 20 modi diversi, per un totale di 400 posizioni possibili dopo sole due semimosse. Per quanto gli scacchi siano un gioco finito, e quindi in teoria completamente analizzabile, siccome il numero medio di mosse in ogni posizione è circa 33 e una partita dura mediamente 40 mosse, ovvero 80 semimosse, un programma che volesse “risolvere” il gioco dovrebbe esplorare 33^{80} partite, un numero di 120 cifre! Immaginando un computer talmente veloce da poter analizzare un miliardo di mosse al secondo (un compito leggermente al di sopra della tecnologia attuale), servirebbero circa 10^{105} anni per analizzare tutte le partite possibili. E se ci si limita ad analizzare le posizioni, queste sono di meno, circa 10^{43} , ma certo è impensabile analizzarle tutte per poi memorizzarne l’esito.

La soluzione universalmente adottata nei programmi di gioco è di fermare la generazione delle mosse ad un certo livello di profondità dell’albero, proporzionalmente alla potenza di calcolo disponibile, di applicare un’euristica sulle foglie dell’albero e poi di valutare i nodi interni mediante l’**algoritmo minimax** (o meglio mediante la sua ottimizzazione chiamata algoritmo *alfabeta*) [3].

Un esempio del funzionamento di Minimax è descritto nelle figure 3 e 4. L’algoritmo inizia valutando con la funzione di valutazione tutte le posizioni del livello più profondo dell’albero. A questo punto l’algoritmo “trasmette” verso l’alto i valori delle posizioni successive usando la regola minimax.

Per esempio, nella figura 3 la posizione d, in

Algoritmo minimax

Tutte le possibili evoluzioni di una partita, a partire da una certa mossa, vengono rappresentate con un albero i cui nodi sono caratterizzati da un certo peso. L’elaboratore viene programmato per ricercare la miglior mossa possibile all’interno di una determinata strategia. Una possibile strategia è quella realizzata con gli algoritmi minimax: ogni mossa viene scelta in modo da ridurre al minimo il massimo dei vantaggi che l’avversario potrà conseguire. L’algoritmo *alfabeta* consente di velocizzare quello minimax, “potando” l’albero dei rami che risulterebbe inutile esaminare.

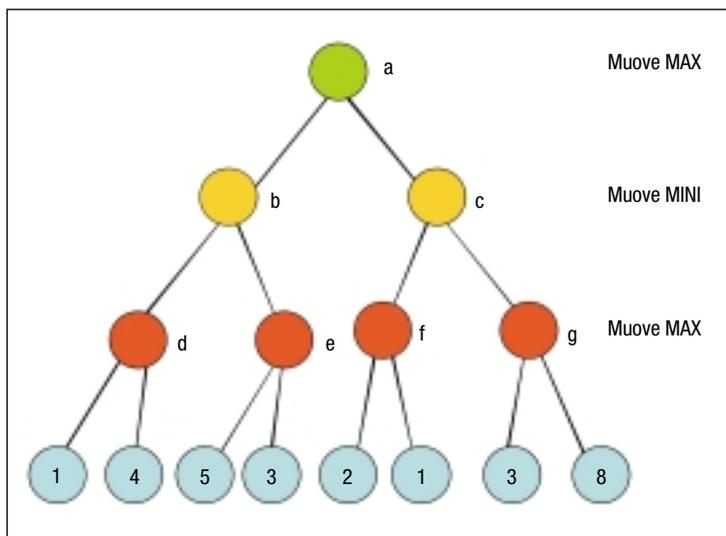


FIGURA 3

Situazione iniziale dell’algoritmo Minimax

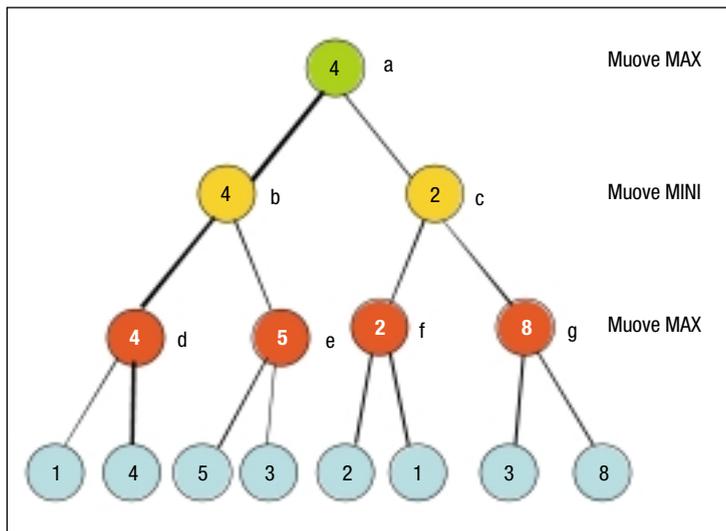


FIGURA 4

Risultato di Minimax; le linee in evidenza indicano la variante principale (a-b-d)

cui deve muovere MAX, ottiene il valore 4, che è il massimo tra le valutazioni delle posizioni sottostanti. La posizione e ottiene il valore 5 per lo stesso motivo. La posizione b, in cui deve muovere MIN, ottiene invece il valore 4, che è il minimo tra i valori delle sue posi-

zioni sottostanti. La situazione finale, in cui la radice ha ottenuto un valore, è schematizzata nella figura 4.

La variante principale è quindi composta dalle mosse che producono in successione le posizioni a-b-d, e la mossa da giocare è quella che porta da a in b.

L'ipotesi grazie alla quale questo metodo funziona è che più profondo è l'albero da valutare, più è precisa la valutazione della mossa da giocare. Dal punto di vista teorico questa ipotesi non è mai stata dimostrata, ed anzi alcuni studi tenderebbero addirittura a dimostrare l'ipotesi contraria, che cioè più profondo è l'albero, meno precisa risulta la valutazione. In effetti in un albero molto grosso le posizioni del livello finale saranno molto diverse tra loro, e quindi sarà difficile paragonarle. D'altra parte i risultati empirici confermano che un programma X che usa uno hardware più veloce Y batte sistematicamente lo stesso pro-

gramma che usa uno hardware meno veloce Z, perché esplora più approfonditamente ed estesamente l'albero di gioco.

A questo proposito K.Thompson effettuò un famoso esperimento con la sua macchina Belle. Fece giocare un torneo tra sei diverse versioni; l'unica differenza tra i programmi era la profondità di analisi dell'albero di gioco, che variava tra 4 e 9 semimosse. Ogni incontro comprendeva 20 partite. La tabella 1 riassume i risultati del torneo.

Come si vede, un incremento di profondità nell'analisi pari a una semimossa comporta in media un incremento nella forza di gioco di quasi 200 punti Elo. Si confronti questa tabella con quella ottenuta da un diverso esperimento di J. Schaeffer [13], che non variava la profondità di analisi del programma, ma solo la sua "conoscenza", cioè l'importanza della funzione di valutazione. Schaeffer effettuò una serie di esperimenti per valutare il bilanciamento relativo dei vari fattori componenti una funzione di valutazione. Egli usò il suo programma Phoenix per giocare un torneo cui partecipavano varie versioni del programma stesso, che differivano per la funzione di valutazione utilizzata. La versione di riferimento (che chiameremo versione 1), tentava semplicemente di massimizzare il materiale. Le altre erano via via arricchite come segue:

	P4	P5	P6	P7	P8	P9	Elo
P4	-	5	0.5	0	0	0	1235
P5	15	-	3.5	3	0.5	0	1570
P6	19.5	16.5	-	4	1.5	1.5	1826
P7	20	17	16	-	5	4	2031
P8	20	19.5	18.5	15	-	5.5	2208
P9	20	20	18.5	16	14.5	-	2328

TABELLA 1

Incremento di forza in funzione della profondità di analisi

versione 1: solo materiale	1110
versione 2: vers. 1 + mobilità e spazio	1420
versione 3: vers. 2 + controllo del centro	1530
versione 4: vers. 3 + struttura pedonale	1600
versione 5: vers. 4 + valutazione mosse	1630
versione 6: vers. 5 + sicurezza del Re	1750
versione 7: vers. 6 + case deboli e P passati	1760
versione 8: vers. 7 + pianificazione	1780

TABELLA 2

Confronto Elo tra diverse funzioni di valutazione

versione 2: mobilità e spazio: mosse pseudo-legali + controllo delle case nella metà avversaria della scacchiera.

versione 3: controllo del (grande) centro;

versione 4: struttura pedonale: penalizzazioni per pedoni isolati o doppiati;

versione 5: valutazione incrementale di mosse: catture di pedone verso il centro, torre in settima, sviluppo di un pezzo, arrocco;

versione 6: sicurezza del Re;

versione 7: case deboli e pedoni passati;

versione 8: pianificatore strategico.

Le otto versioni del programma giocarono lungamente tra di loro, ed i risultati sono riassunti nella tabella 2.

Confrontando i due esperimenti si deduce che si guadagna molto di più migliorando la velocità di analisi di un programma che non la sua conoscenza specifica della strategia scacchistica.

Infatti i grandi successi dei giocatori artificiali si sono verificati a partire dagli anni '80, quando i progressi delle tecnologie VLSI hanno permesso il progetto di hardware specializzato per la generazione di mosse [5].

Deep Blue aveva un'architettura parallela: un modello SP/2 con 24 processori, che controllano 512 processori specializzati per la generazione delle mosse. Ciascun processore specializzato analizza circa 1 milione di posiz/s, e complessivamente i 512 processori arrivano a circa 200 milioni di posiz/s utili. Questo permetteva a Deep Blue di esplorare alberi profondi 12/14 semimosse in circa 3 min, il tempo medio per mossa di torneo.

5. LO STATO DELL'ARTE

Un'importante misura dell'efficienza di un algoritmo parallelo è lo *speed-up* (accelerazione). Questo parametro è definito come il rapporto tra il tempo di esecuzione di una implementazione sequenziale *efficiente* di un algoritmo e quello di una sua versione parallela. L'obiettivo ideale è di ottenere valori per lo speedup che aumentino linearmente con il numero di processori usati. Purtroppo questo risultato è difficile da ottenere per gli algoritmi paralleli di ricerca su alberi di gioco. Esistono infatti diversi motivi di degrado delle prestazioni di questa classe di algoritmi che vengono indicati con il termine *overhead* (degradazione) [14].

L'*overhead in tempo* (OT) esprime una misura quantitativa della degradazione totale dell'algoritmo: è la perdita percentuale di speedup confrontata con lo *speed-up* ideale. È espressa come:

$$OT = (\text{tempo con } n \text{ CPU}) \cdot \frac{n}{\text{tempo con } 1 \text{ CPU}}$$

L'*overhead* totale è in realtà composto da più componenti:

□ *overhead di ricerca* (OR): in ambiente sequenziale tutte le informazioni ricavate fino ad un dato punto della ricerca sono disponibili per effettuare decisioni quali il taglio di un sottoalbero.

In ambiente distribuito le stesse informazioni possono essere disperse su macchine diverse e quindi non efficacemente utilizzabili: ciò

può portare ad una esplorazione non necessaria di una porzione dell'albero di gioco. La crescita delle dimensioni dell'albero è chiamata *overhead* di ricerca. Tale forma di degrado può essere approssimata osservando che la dimensione dell'albero esplorato è proporzionale al tempo di ricerca. OR è precisamente definito dalla relazione:

$$OR = \frac{\text{nodi visitati da } n \text{ CPU}}{\text{nodi visitati da } 1 \text{ CPU}} - 1$$

□ *overhead di comunicazione* (OC): è il carico aggiuntivo che un programma parallelo deve sopportare quando è impiegato un tempo non trascurabile per la comunicazione dei messaggi fra i processi. Questo costo può essere limitato in fase di programmazione scegliendo opportunamente le dimensioni e la frequenza dei messaggi. Una stima di OC è data dal prodotto fra il numero di messaggi inviati e il costo medio di un messaggio.

□ *overhead di sincronizzazione* (OS): è il costo che occorre quando alcuni dei processori sono inattivi. In teoria tutti i processori dovrebbero essere occupati nello svolgere lavoro utile per tutto il tempo di esecuzione. Nella realtà questo comportamento ideale viene meno quando un processo deve sincronizzarsi con un altro determinando attesa attiva. Potrebbe accadere per esempio che un processo P voglia agire su un valore condiviso in mutua esclusione, mentre un altro processo Q sta già operando su tale valore. Il processo P rimarrà bloccato (e quindi inattivo) fintanto che Q non abbia completato la sua operazione. Tale *overhead* può presentarsi anche quando un processo è in attesa che certi risultati, senza i quali non può continuare, gli siano forniti da un altro.

L'*overhead* complessivo OT è funzione delle tre forme di *overhead* elencate:

$$OT = f(OR, OC, OS)$$

Per massimizzare le prestazioni di un algoritmo parallelo è necessario minimizzare i vari tipi di degradazione. Purtroppo essi non sono mutuamente indipendenti e lo sforzo nella minimizzazione di uno di essi può risultare nell'aggravio di un altro. Per esempio la riduzione dell'*overhead* di ricerca richiede nor-

Sistema	Posizioni/sec	HW
Ananse	6.000	Intel 486/66
MChess	10.000	Pentium 90
Amy	10.000	Sun Sparc10
Arthur	20.000	Sun Sparc20
Cray Blitz	750.000	Cray C90
*Socrates	1.000.000	CM5 / 512
DeepThought 2	4.000.000	IBM/6000
Deep Blue (1997)	200.000.000	IBM SP/2 (24 processori) + 512 chip speciali
Fritz, Junior, Shredder (2004)	3-5 milioni	Quadriprocessore su PC di ultima generazione

TABELLA 3
Numero di posizioni per secondo analizzate da alcuni sistemi

malmente un aumento del numero delle comunicazioni.

Esistono comunque diverse fonti di parallelismo nella ricerca di alberi di gioco le quali hanno originato approcci completamente differenti al processo di parallelizzazione della ricerca sequenziale:

- parallelismo nella valutazione statica dei nodi terminali;
- generazione parallela delle mosse;
- decomposizione dell'albero di gioco.

5.1. Parallelismo nella funzione di valutazione statica

Una considerevole porzione del tempo di ricerca è spesa nella valutazione statica dei nodi terminali. La qualità delle scelte strategiche operate durante il gioco è fortemente legata alla affidabilità e quindi alla complessità della funzione di valutazione. Spesso si sceglie di praticare una valutazione statica semplice e grossolana preferendo impiegare il tempo di ricerca nel condurre esplorazioni più in profondità nell'albero di gioco. Il tempo dedicato alla valutazione dei nodi terminali può tuttavia essere ridotto distribuendo tale operazione su più processori, ciascuno dedicato a valutare differenti termini della funzione di valutazione. I risultati del lavoro così partizionato saranno riuniti per originare un unico valore per la posizione esaminata.

Chiaramente lo speedup massimo ottenibile è limitato dalla scomponibilità e modularità della funzione di valutazione implementata. Deep Blue fa uso di un hardware specifico per l'esecuzione della valutazione statica. In particolare esso consiste di due componenti denominati rispettivamente valutatore veloce e lento:

□ l'hardware per la valutazione veloce aggiorna i termini della funzione di valutazione che possono essere calcolati in modo incrementale, cioè a mano a mano che vengono eseguite o retratte le mosse. Tale aggiornamento avviene in parallelo alle altre fasi di esplorazione interna dell'albero; quando sarà raggiunto un nodo terminale parte dei suoi termini saranno quindi già valutati.

□ i restanti termini della funzione di valutazione sono calcolati dal valutatore lento quando la ricerca arriva ad un nodo terminale. Il loro calcolo è completato molto rapidamente poiché la risorsa di elaborazione ad esso dedicata utilizza un insieme di tabelle indirizzate via hardware. Ogni tabella contiene informazioni riguardanti proprietà salienti di una posizione aggiornate durante la ricerca.

5.2. Generazione in parallelo delle mosse

La generazione delle mosse è un'altra operazione molto frequente nella ricerca su un albero di gioco ed in generale incide pesantemente sul tempo globale di esecuzione. Negli scacchi la velocità di generazione delle mosse è il fattore che limita le dimensioni della ricerca poiché le regole che governano il movimento dei pezzi sono piuttosto complicate. Un accorgimento desiderabile è dunque la generazione in parallelo di tutte le mosse relative ad un nodo.

Ogni colonna della scacchiera potrebbe essere assegnata ad un diverso processore, che genererà tutte le mosse per tutti i pezzi disposti sulla propria colonna. Poiché ogni colonna può contenere più pezzi di un'altra è necessaria una forma di bilanciamento del carico, cioè un riassetto dei lavori per mantenere tutti i processori attivi. In questa tecnica un motivo di degrado delle prestazioni è dovuto al fatto che per molti nodi non è necessario generare tutte le mosse (a causa di tagli) e così parte del lavoro andrà perduto. Il metodo descritto si presta per una sua im-

plementazione in hardware, cioè che utilizzi componenti architetturali progettati *ad hoc*. HITECH [5] utilizzava hardware specifico per la generazione in parallelo delle mosse. In particolare veniva impiegata una matrice 8 × 8 di processori, uno per ogni casa della scacchiera: ciascun processore calcolava in parallelo agli altri le mosse legali per il pezzo che occupa la casa cui è associato (ammesso che essa non sia vuota). Il generatore implementava inoltre l'ordinamento delle mosse: quella suggerita è la mossa con maggiore priorità; la priorità di una mossa è una stima del suo valore ed è attribuita autonomamente dal processore che l'ha generata. Ogni processore deposita la priorità della sua migliore mossa su un bus accessibile da tutti gli altri: quando il processore con la mossa a più alta priorità riconosce che nessun altro dispone di mossa migliore, esso presenta la sua mossa al modulo dedicato al controllo della ricerca.

5.3. Decomposizione dell'albero di gioco

Il metodo di decomposizione divide l'albero in sottoalberi e stabilisce che sottoalberi diversi siano esplorati da processori in genera-

le distinti. Tale approccio al parallelismo non prevede, in linea di principio, limitazioni per lo *speed-up*. In [4] sono confrontate le prestazioni di diversi algoritmi distribuiti di decomposizione dell'albero di gioco. La prospettiva di prestazioni elevate su architetture specializzate ha fatto sì che i maggiori sforzi nella ricerca siano stati concentrati nella direzione tracciata da questa classe di algoritmi paralleli. In effetti, questo è l'approccio scelto da Deep Blue [1].

5.4. I programmi commerciali

Quelle che abbiamo delineato sono le linee di ricerca più note, in quanto pubblicate su riviste scientifiche dagli autori dei programmi. Esistono però centinaia di programmi di vari autori e di diversa forza, alcuni dei quali hanno valore commerciale e di cui nulla si sa circa il funzionamento interno. I principali tra questi programmi si sfidano annualmente in un Campionato del Mondo organizzato da ICGA. Nel 2004 il Campionato si è svolto presso l'Università Bar-Ilan di TelAviv, con i risultati mostrati nella tabella 4.

Questa tabella è interessante perché sembra

Nome	Nazione	Hardware	Punti (su 11 incontri)
Junior	Israele	4proc 2.2GHz, Proliant HP	9
Shredder	Germania	4proc AMD 2.0 GHz, Transtec	8.5
Diep	Olanda	4proc AMD 2.0 GHz	7.5
Fritz	Olanda	4proc AMD 2.0 GHz, Transtec	7
Crafty	USA	4proc AMD 2.0 GHz	7
Jonny	Germania	AMD64 3200+	6.5
ParSOS	Germania	AMD64 3200+	6
Falcon	Israele	AMD64 3200+	6
ISiChess	Germania	AMD64 3400+	6
Deep Sjeng	Belgio	AMD64 3400+	5.5
Woodpusher	UK	Pentium 4 2.8 GHz	3
Movei	Israele	Pentium 4 2.8 GHz	3
The Crazy Bishop	Francia	Pentium 4 2.8 GHz	2
FIBChess	Spagna	Pentium 4 2.8 GHz	0

TABELLA 4

La classifica finale del Campionato del Mondo 2004

mostrare che l'elemento dominante anche con software diversi sia l'hardware.

I programmi commerciali sono così "ottimizzati" che ormai battono regolarmente i migliori giocatori umani. Per esempio nello scorso settembre 2004 all'IRST di Trento il migliore giocatore italiano, Michele Godena, ha sfidato il programma campione del mondo 2004, Deep Junior, creato in Israele da Bushinsky e Ban (Figura 5). La partita, vinta dal programma, è la seguente:

[Event "Uomo vs Computer"]
[Site "Povo - Trento ITA"]
[Date "2004.09.27"]
[Round "1"]
[White "Deep Junior"]
[Black "Godena, Michele"]
[Result "1-0"]
[PlyCount "89"]

1. e4 e5 2. Nf3 Nc6 3. Bb5 a6 4. Ba4 Nf6 5. o-o b5
6. Bb3 Bc5 7. a4 Bb7 8. d3 b4 9. c3 d6 10. a5 h6
11. d4 Ba7 12. Be3 o-o 13. Nbd2 bxc3 14. bxc3
exd4 15. cxd4 Nb4 16. Qb1 Rb8 17. Re1 d5 18. e5
Nd7 19. e6 fxe6 20. Bxh6 Qf6 21. Bg5 Qf7 22. Bh4
Rfe8 23. Bg3 Nf8 24. Ba4 Bc6 25. Ne5 Qf6 26.
Ndf3 Bxa4 27. Rxa4 Nc6 28. Qd3 Nxe5 29. Bxe5
Qf5 30. Qxa6 Rb1 31. Ra1 Rxa1 32. Rxa1 Rb8 33.
Rf1 Rb1 34. Qxa7 Rxf1+ 35. Kxf1 Qb1+ 36. Ke2
Qb5+ 37. Ke3 Qb3+ 38. Kf4 Qc2 39. h4 Qxf2 40.
Qb8 Qxg2 41. a6 Qe2+ 42. Qxc7 Ng6+ 43. Kg3
Nxe5 44. Qd8+ Kh7 45. dxe5 1-0



FIGURA 5

Da sinistra: Shay Bushinsky e Amir Ban, autori di Deep Junior

6. CONCLUSIONI

La ricerca sul gioco artificiale è ritenuta tuttora da alcuni studiosi di importanza capitale per lo sviluppo della teoria e delle tecnologie della Intelligenza Artificiale. Secondo Donald Michie:

“la costruzione di giocatori artificiali di scacchi, una ricerca di tipo tecnologico ma che ha una portata che va ben al di là della pura tecnologia, è oggi la ricerca scientifica più importante del mondo. Possiamo confrontarla con le ricerche fatte da T.Morgan a New York durante la Prima Guerra Mondiale sulla *Drosophila*. Quelle ricerche ebbero un impatto fondamentale sulle basi teoriche della genetica moderna. Ci accorgiamo oggi delle conseguenze industriali dell'ingegneria genetica, che è figlia delle ricerche di Morgan. Stiamo oggi usando l'analisi scientifica del gioco degli scacchi come studio preliminare per l'ingegneria della conoscenza”.

Tuttavia, anche se l'Intelligenza Artificiale può guadagnare ancora molto dall'approfondimento di ricerche sui giochi, non c'è dubbio che il risultato di Deep Blue va considerato in primo luogo un successo della ricerca sul calcolo parallelo.

Il successo di Deep Blue è irripetibile, perché la macchina è stata smantellata. Negli ultimi anni ci sono stati altri incontri tra i più forti giocatori del mondo e i migliori programmi commerciali che giravano su normali personal computer. La maggior parte di questi incontri si è conclusa in parità. Dunque diciamo che al momento la superiorità delle macchine sugli umani nel gioco degli scacchi non è affatto netta. Certo però ormai sono poche decine gli umani che riescono a resistere alla forza di gioco delle macchine. Per esempio i migliori giocatori dell'Internet Chess Club sono da anni i computer: al momento domina il programma Shredder.

Occorre domandarsi come si orienterà la ricerca in futuro. Esistono molti giochi più complessi degli scacchi, ad esempio il Go o gli scacchi cinesi. Noi crediamo che siano i giochi ad informazione incompleta ad *n* giocatori quelli che debbono essere studiati in futuro: un esempio è il Risiko. Questi giochi creano situazioni molto complesse che sono poco studiate.

MINIGLOSSARIO SCACCHISTICO

ALBERO DI GIOCO

In un *gioco ad informazione completa*, è la rappresentazione astratta dello sviluppo di tutte le varianti possibili, ciascuna calcolata fino ad una posizione terminale. Negli scacchi l'albero del gioco è completamente sviluppabile in teoria ma non in pratica, a causa dell'*esplosione combinatoria* delle varianti.

EFFETTO ORIZZONTE

Difetto intrinseco di un algoritmo di valutazione basato su una visita parziale dell'albero di gioco. Si rimedia in parte continuando ad espandere l'albero fino ad ottenere solamente posizioni quiescenti.

ELO, PUNTEGGIO

Sistema di misura e classificazione della forza dei giocatori di scacchi. Introdotto dal prof. A. Elo negli Stati Uniti durante gli anni '60, venne ufficialmente adottato dalla FIDE nel 1970. Al punteggio Elo, che varia col variare dei risultati conseguiti in tornei ufficiali, è legato il conseguimento dei titoli nazionali e internazionali, secondo la seguente tabella indicativa: Maestro = 2200; Maestro Fide = 2300; Maestro Internazionale = 2400; Grande Maestro = 2500. Il sistema Elo locale di una nazione può differire da quello internazionale. In particolare, il sistema Elo della federazione americana USCF è leggermente inflazionato rispetto a quello internazionale. Per questo motivo i punteggi Elo dichiarati dai costruttori di scacchiere elettroniche sono solitamente riferiti all'Elo USCF.

OBBIETTIVO

Situazione di gioco da raggiungere mediante un piano strategico.

PIANO

Definizione dell'obiettivo e dei mezzi per raggiungerlo.

PLY

Vedi Semimossa.

POSIZIONE QUIESCENTE

Posizione in cui non sono possibili scacchi o prese. Una posizione quiescente viene valutata dalla funzione di valutazione senza ulteriore sviluppo dell'albero di gioco.

POSIZIONE TERMINALE

Nel gioco degli Scacchi una posizione è terminale se uno dei due giocatori ha dato scacco matto all'avversario, oppure se non è più possibile il matto, e quindi la partita è patta (esempio: Re solo contro Re solo).

POSIZIONE TURBOLENTA

Posizione non quiescente. Tipicamente, una posizione in cui sono possibili scacchi o prese.

SEMIMOSSA

Movimento di un pezzo di uno dei due giocatori. Due semimosse consecutive, una per ciascun giocatore, costituiscono una mossa nel senso tradizionale.

STRATEGIA

Definizione ed esecuzione di un piano di gioco a lungo termine allo scopo di conseguire un certo obiettivo. Nei commenti ad una partita la descrizione di una strategia è spesso generica e non dettagliata.

TATTICA

Definizione ed esecuzione di piani di gioco a breve termine, solitamente subordinati alla strategia. Un piano tattico va elaborato dettagliatamente.

VARIANTE PRINCIPALE

Risultato ottenuto dalla valutazione minimax di una posizione. Contiene la mossa da giocare ed il seguito più probabile previsto dalla macchina.

ZUGZWANG

Una posizione di zugzwang è una posizione in cui la parte che deve muovere non può che peggiorare le proprie possibilità, mentre invece se potesse "passare" la mossa non comprometterebbe nulla.

Bibliografia

- [1] Anantharaman T., Campbell M., Hsu F.: Singular Extensions: Adding Selectivity to Brute Force Searching. *Artificial Intelligence*, Vol. 43, 1990, p. 99-110.
- [2] Berliner HJ., Goetsch G., Campbell M., Ebeling C.: Measuring the Performance Potential of Chess Programs. *Artificial Intelligence*, Vol. 43, 1990, p. 7-20.
- [3] Ciancarini P.: *I giocatori artificiali di scacchi*. Mursia, 1992.
- [4] Ciancarini P.: Distributed Searches: a Basis for Comparison. *Journal of the International Computer Chess Association*, Vol. 17:4, 1994, p. 194-206.

- [5] Ebeling C.: *All the Right Moves: A VLSI Architecture for Chess*. MIT Press, 1987.
- [6] Finkel RA., Fishburn J.: Parallelism in Alpha-Beta Search. *Artificial Intelligence*, Vol. 19, 1982, p. 89-106.
- [7] Hsu FH.: *Behind Deep Blue*. Princeton University Press, 2002.
- [8] Marsland TA., Campbell M.: Parallel Search of Strongly Ordered Game Trees. *ACM Computing Surveys*, Vol. 14:4, 1982, p. 533-551.
- [9] Marsland TA., Popowich F.: Parallel Game Tree Search. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 4:7, 1985, p. 442-452.
- [10] McGrew T.: Collaborative Intelligence. *IEEE Internet Computing*, Vol. 1:3, 1997, p. 38-42.
- [11] Moravec H.: When will computer hardware match the human brain?. *Journal of Evolution and Technology*, Vol. 1, 1998 (anche in Moravec H., *ROBOT: Mere Machine to Transcendent Mind*, Oxford Univ. Press, 1998).
- [12] Newborn M.: *Deep Blue: An Artificial Intelligence Milestone*. Springer, 2003.
- [13] Schaeffer J.: *Experiments in Search and Knowledge*. PhD Thesis, University of Alberta, 1986.
- [14] Schaeffer J.: Distributed Game-Tree Searching. *Journal of Parallel and Distributed Computing*, Vol. 6, 1989, p. 90-114.
- [15] Shannon CE.: Programming a Computer for Playing Chess. *Philosophical Magazine*, Vol. 41:7, 1950, p. 256-275.
- [16] Simon H.A.: *Models of My Life*. Basic Books, 1991.
- [17] Turing AM.: Digital Computers Applied to Games. In *Faster Than Thought* (Ed. Bowden BV.), Pitman 1953, p. 286-295.
- [18] Siti web: www.research.ibm.com/deepblue - sito di IBM su Deep Blue.
www.icga.org - sito dell'associazione internazionale di giochi e computer.
www.chessclub.com - internet Chess Club, dove molti giocatori sono computer

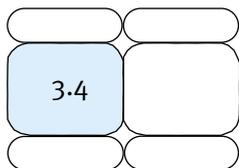
PAOLO CIANCARINI è Ordinario di Ingegneria del Software all'Università di Bologna. È Candidato Maestro di Scacchi, anche se non gioca più in torneo da parecchi anni. È uno degli organizzatori del Campionato del Mondo di Scacchi per computer che si terrà nel 2006 a Torino nell'ambito delle Olimpiadi degli Scacchi. È socio e consigliere di AICA, nonché membro del Comitato Scientifico di Mondo Digitale.
 ciancarini@cs.unibo.it



VERSO UNA RETE TUTTA OTTICA

Gastone Bonaventura

L'articolo riporta in sintesi i problemi relativi all'evoluzione della rete di trasporto dall'attuale rete SDH verso una Rete Tutta Ottica (AON). Il primo passo, in corso di realizzazione in questi ultimi tempi, sarà costituito dalla Rete di Trasporto Ottica (OTN), arricchita da un'architettura di Rete Ottica Commutata Automaticamente (ASON). Nel testo sono anche descritte le maggiori problematiche oggi ancora aperte, la cui soluzione condiziona il successivo passaggio ad una Rete Tutta Ottica.



1. INTRODUZIONE

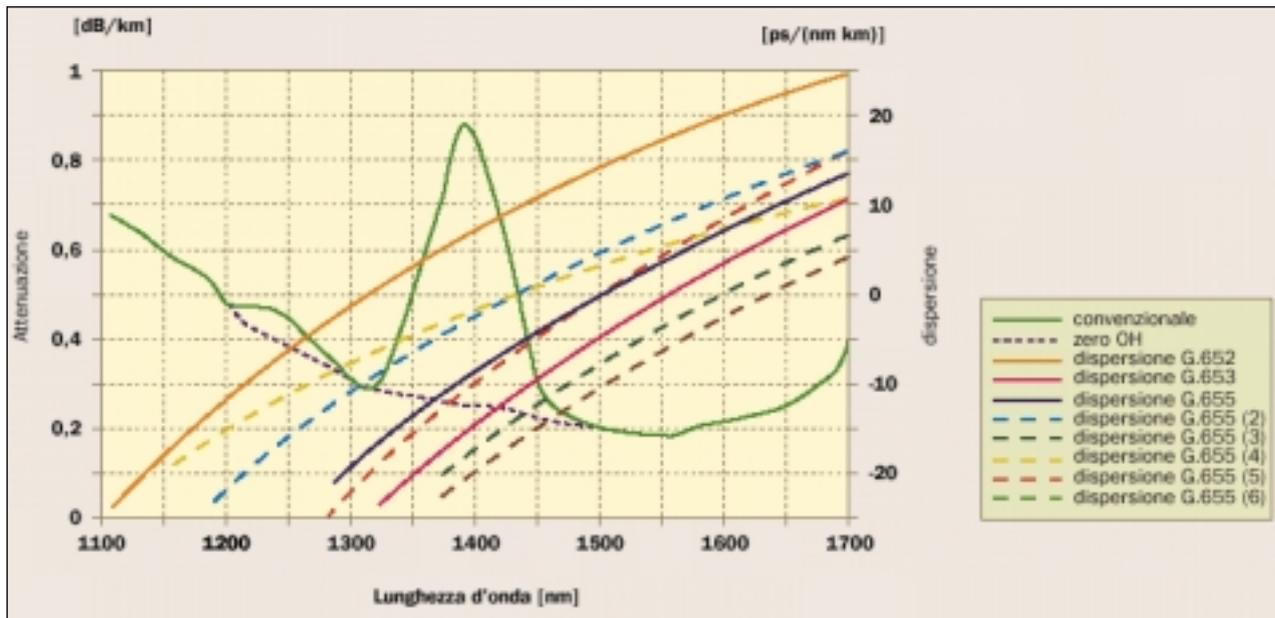
Una rete di trasporto è costituita da una serie di nodi, nei quali si effettua il trattamento dei segnali cliente (come la permutazione, l'estrazione e la reintroduzione dei segnali, la multiplazione ecc.) e da una serie di collegamenti tra i nodi. Da un punto di vista generale, la rete di trasporto si considera divisa in tre parti principali: reti di accesso, reti metropolitane, rete dorsale (*backbone*). Le risorse trasmissive di una rete di trasporto sono adatte a trasportare indifferentemente tutti i diversi tipi di segnali cliente (voce, musica, immagini, dati e video).

Con l'avvento della liberalizzazione delle reti e dei servizi di telecomunicazioni, anche la rete di trasporto di un operatore si trova ora ad operare in regime di competitività con quelle di altri operatori. La sua realizzazione e la successiva gestione devono essere perciò finalizzate a ben precisi obiettivi di costo, di qualità e di flessibilità. Tra questi obiettivi ha assunto di recente una particolare importanza la *disponibilità*, in conseguenza del ruolo sempre più incisivo svolto dalle teleco-

municazioni nello svolgimento di un gran numero di attività economiche. Gli operatori offrono attualmente la capacità trasmissiva con diversi obiettivi di disponibilità (e naturalmente di costo), generalmente divisi in classi che arrivano fino al 99,999% (corrispondente ad un'indisponibilità del servizio di 5 minuti l'anno). È evidente che questi obiettivi pongono alcuni vincoli sia alla struttura della rete (in termini di configurazioni e di ridondanze) sia alle caratteristiche e alla tecnologia degli apparati e dei portanti.

La tecnologia ottica costituisce oggi l'elemento fondamentale per la realizzazione dei collegamenti nelle reti di trasporto, con l'eccezione delle reti di accesso dove rimane predominante l'impiego dei tradizionali conduttori metallici.

In questo articolo viene esaminata la prevedibile estensione dell'impiego delle tecnologie ottiche nelle reti metropolitane e nella rete dorsale in un'evoluzione verso una Rete Tutta Ottica, l'AON (*All Optical Network*). La diffusione delle tecnologie ottiche nella rete di accesso non rientra negli obiettivi di questo articolo.



2. LE FIBRE OTTICHE

Nell'ambito delle reti metropolitane e delle reti dorsali la percentuale dei cavi ottici sul totale dei portanti impiegati in Italia supera di molto il 90%, in quanto, dalla seconda metà degli anni Ottanta, tutti i nuovi collegamenti sono stati realizzati in fibra ottica. I vantaggi delle fibre ottiche sui conduttori metallici in termini di prestazioni, ingombro, peso, costo sono, infatti, tali che, non appena la tecnologia è stata sufficientemente matura, il loro impiego nelle reti di telecomunicazioni si è immediatamente generalizzato. Le fibre ottiche impiegate nelle reti di trasporto sono tutte di tipo singolo-modo, costituite da un nucleo e da un mantello concentrici di silice variamente drogata per realizzare il desiderato profilo dell'indice di rifrazione. Il diametro del mantello è di 125 μm , diametro che, con il rivestimento protettivo primario, diventa di 250 μm [1]. I segnali che si propagano lungo le fibre ottiche sono soggetti a degradazioni causate principalmente da: attenuazione, dispersione cromatica, dispersione di polarizzazione, effetti legati alle non linearità.

2.1. Attenuazione

L'attenuazione subita da un segnale ottico nella sua propagazione lungo una fibra ottica è dovuta alla combinazione di diversi effetti

(diffusione di Rayleigh, dovuta a fluttuazioni di densità a livello molecolare della silice vetrosa; assorbimento molecolare, massimo nel medio infrarosso; assorbimento atomico, massimo nell'ultra-violetto; presenza di impurezze nella silice della fibra, principalmente dovuta a presenza di ioni ossidrilici ecc.).

L'attenuazione varia con la lunghezza d'onda del segnale trasmesso, come mostrato in figura 1, e linearmente con la lunghezza della fibra. Nella figura 1 sono mostrati due andamenti dell'attenuazione: il primo con un piccolo intorno a 1390 nm per la presenza nella fibra di ossidrilici OH^- . Il secondo, relativo a fibre di recente produzione nelle quali l'ossidrilico OH^- è stato eliminato, presenta invece una attenuazione regolarmente decrescente con il crescere della lunghezza d'onda fino a circa 1700 nm. La figura mostra ancora che il campo di lunghezza d'onda più attraente, dal punto di vista dell'attenuazione, è nel vicino infrarosso ed è compreso tra 1260 nm e 1675 nm. Il principale organismo di normativa in materia, l'ITU-T (*International Telecommunication Union – Telecommunication Standardization Sector*) ha suddiviso questo intervallo spettrale complessivo in sei bande (vedi riquadro).

La banda oggi di gran lunga più utilizzata è la banda C (1530-1565 nm) perchè è quella dove si verificano le minori attenuazioni (0,20-0,22 dB/km) e nella quale si ha anche la disponibili-

FIGURA 1

Attenuazione e dispersione cromatica delle fibre ottiche
(Fonte: [6])

Gamma delle lunghezze d'onda

La figura 1 mostra che la gamma di lunghezze d'onda nella quale l'attenuazione delle fibre è sufficientemente bassa ($<0,4$ dB/km) è molto estesa (oltre 200 nm). Ai fini di utilizzare una comune terminologia L'ITU-T [7] ha suddiviso l'intervallo di lunghezze d'onda nelle sei bande seguenti:

Banda	Nome	Campo
O-Band	Original	1260-1360 nm
E-Band	Extended	1360-1460 nm
S-Band	Short wavelength	1460-1530 nm
C-Band	Conventional	1530-1565 nm
L-Band	Long wavelength	1565-1625 nm
U-Band	Ultralong wavelength	1625-1675 nm

A questo campo complessivo di lunghezze d'onda corrisponde una banda di frequenza di circa 50 mila GHz (50 THz).

lità di una vasta gamma di componenti ottici. Maggiori dettagli sull'impiego delle varie bande di lunghezza d'onda sono riportati nel paragrafo 3 che esamina i sistemi di trasmissione. Infine l'andamento dell'attenuazione in funzione della lunghezza d'onda mostrato in figura 1 è praticamente lo stesso per tutti i diversi tipi di fibra impiegati nelle reti di trasporto in quanto è legato essenzialmente al tipo di materiale impiegato (la silice) e dipende molto poco dalle caratteristiche dimensionali e geometriche della fibra stessa.

2.2. Dispersione cromatica

Un secondo tipo di degradazione che subisce un segnale che si propaga in una fibra è dovuto al fatto che la velocità di propagazione varia con la lunghezza d'onda. Considerato che le sorgenti ottiche (laser) non sono in genere monocromatiche, le varie lunghezze d'onda che compongono ogni impulso immesso in fibra, in cui è codificato il segnale di trasmissione, arrivano al punto di ricezione in tempi differenti provocando un allargamento della forma dell'impulso trasmesso. Si ha, quindi, un'interferenza intersimbolica che, ove superi certi limiti, può provocare degli errori nella fase di identificazione alla ricezione degli impulsi o in quella di rigenerazione. L'entità della dispersione cromatica dipende quindi dalla larghezza spettrale delle sorgenti ottiche e dalle

caratteristiche di dispersione delle singole fibre. Tali caratteristiche variano fortemente a seconda del progetto della fibra stessa perchè esse risultano dalla combinazione di un effetto di dispersione del materiale (dovuta alla dipendenza spettrale dell'indice di rifrazione della silice costituente la fibra) e di un effetto di guida (dovuto alle caratteristiche ottico-geometriche della fibra medesima). In conseguenza di questa situazione l'ITU-T ha standardizzato diversi tipi di fibre, ottimizzato ciascuno per un impiego specifico.

La figura 1 mostra l'andamento della dispersione cromatica in funzione della lunghezza d'onda per alcune tipologie di fibre, indicate con il numero della corrispondente Raccomandazione dell'ITU-T [2]. Appare chiaro che l'andamento per le singole fibre è molto simile, ma il punto di zero è posto a lunghezze d'onda diverse, modellando soprattutto l'andamento dell'indice di rifrazione del nucleo, in funzione dei diversi tipi di applicazione (lunghezza d'onda di lavoro, impiego dei sistemi nelle brevi, medie o grandi distanze, sistemi a singolo canale ovvero multicanale ecc.) cui la fibra è destinata. La dispersione cromatica cresce linearmente con la lunghezza del collegamento ed è un fenomeno di natura deterministica per cui può essere compensata con adatti dispositivi prima che raggiunga valori critici.

2.3. Dispersione di polarizzazione

La dispersione di polarizzazione è legata alla diversa velocità di propagazione delle due componenti ortogonali del campo elettromagnetico trasmesso lungo la fibra. L'effetto è ancora quello di un allargamento dell'impulso di potenza ottico che dà origine a interferenza intersimbolica. L'entità della dispersione dipende da molti fattori (alcuni variabili nel tempo): il metodo di fabbricazione delle fibre, la struttura del cavo ottico, il tipo di posa, la temperatura ecc..

La dispersione di polarizzazione è quindi un fenomeno di tipo statistico e il problema della sua compensazione è più complesso e non così efficace come quello della dispersione cromatica. Questo tipo di dispersione si accumula con la radice quadrata della lunghezza del collegamento ed i suoi valori tipici non superano generalmente limiti compresi tra 0,2 e 0,5 ps/km⁻².

2.4. Non-linearità

In generale l'indice di rifrazione del materiale costituente la fibra non dipende dall'intensità del campo elettromagnetico che lo attraversa. Ma per intensità particolarmente elevate, quali quelle rese disponibili da molti amplificatori ottici, l'indice di rifrazione può mostrare una dipendenza dall'intensità. La non-linearità può provocare diversi tipi di degradazione del segnale trasmesso. Il più temibile è generalmente l'interazione a quattro fotoni, più nota come FWM (*Four Wave Mixing*), che dà origine ad alcune repliche del segnale stesso a lunghezze d'onda diverse. Questi segnali spuri possono andare a cadere in corrispondenza di altri canali provocando rumore e/o diafonia. Naturalmente questo tipo di degradazione ha un impatto solamente nel caso di sistemi che trasmettono più canali ottici sulla stessa fibra (sistemi WDM). Gli effetti delle degradazioni dovute alla non linearità delle fibre sono tanto maggiori quanto più bassa è la dispersione cromatica nella gamma di lunghezze d'onda di lavoro.

Le caratteristiche geometriche, dimensionali e ottiche delle fibre possono essere scelte in modo da ridurre l'entità di una delle degradazioni di cui sopra (spesso però a scapito delle altre) in funzione della loro utilizzazione.

La fibra G.652, che è decisamente la più diffusa (costituisce, infatti, oltre l'80% delle fibre posate nel mondo), presenta una dispersione cromatica relativamente alta nella banda 1530-1565 nm, ma ha effetti non lineari molto ridotti. La fibra G.655 è caratterizzata da una dispersione cromatica sufficientemente elevata nella stessa banda per limitare gli effetti delle non linearità, ma abbastanza bassa per non rendere eccessivamente complesso il

problema della compensazione. La fibra G.656, che costituisce lo standard più recente, è invece caratterizzata da una dispersione cromatica relativamente bassa in tutta la banda 1460-1625 nm (bande S + C + L).

3. I SISTEMI OTTICI DI TRASMISSIONE

I sistemi ottici di trasmissione comprendono gli apparati necessari per trasmettere sulle fibre ottiche i segnali clienti. Essi possono essere suddivisi in due grandi categorie: i sistemi a canale singolo e quelli multicanale.

3.1. I sistemi a canale singolo

I sistemi a canale singolo, sono chiamati anche TDM (*Time Division Multiplexing*) in quanto i segnali trasmessi sono multiplati a divisione di tempo; essi trasmettono un solo canale ottico su una coppia di fibre, che è necessaria perchè i segnali cliente sono bidirezionali. I sistemi ottici a canale singolo sono costituiti, in linea di principio (Figura 2), da un terminale trasmittente (dove si effettua, in particolare, la conversione elettro-ottica del segnale da trasmettere), da un terminale ricevente (dove si effettua, tra l'altro, la conversione ottico-elettrica del segnale ricevuto) e da un numero di amplificatori ottici che è funzione della lunghezza del collegamento da realizzare. In pratica, l'impiego dei sistemi TDM è limitato a collegamenti sufficientemente corti da non richiedere amplificatori intermedi. Il costo dell'inserzione in linea di amplificatori intermedi giustifica in genere l'installazione di un sistema WDM atto ad utilizzare gli amplificatori di linea con più canali. Il campo di impiego dei sistemi ottici TDM va

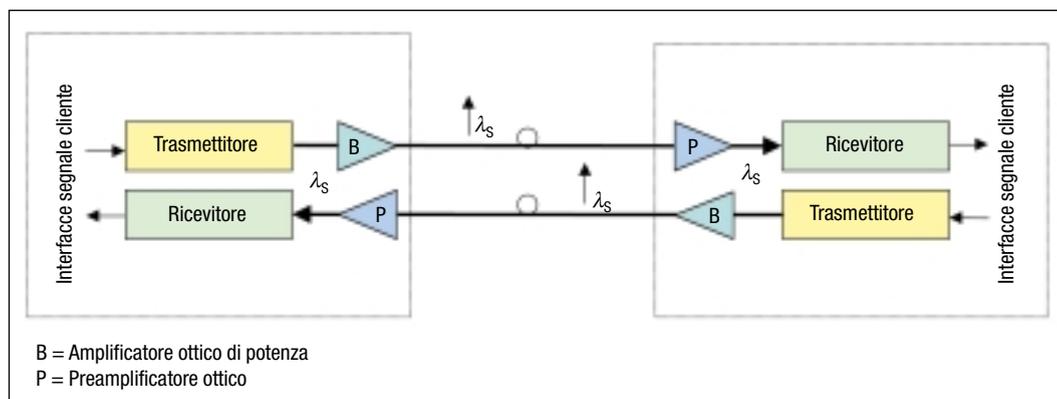


FIGURA 2

Schema di un sistema ottico TDM

quindi dai raccordi all'interno delle centrali ai collegamenti periferici nelle reti metropolitane. Le Raccomandazioni ITU-T [3] prevedono lunghezze fino a 160 km per sistemi con capacità fino a 2,5 Gbit/s e lunghezze fino a 120 km per sistemi con capacità fino a 10 Gbit/s. La tecnologia ottica attuale consente tuttavia di realizzare anche sistemi TDM senza amplificazione intermedia dell'ordine di 250-300 km per applicazioni particolari, come, per esempio, i collegamenti sottomarini.

3.2. I sistemi WDM

I sistemi multi-canale, indicati generalmente come WDM (*Wavelength Division Multiplexing*), trasmettono invece molti canali a diversa lunghezza d'onda su una coppia di fibre ottiche. In questi sistemi i terminali trasmettenti, le cui uscite sono evidentemente a lunghezze d'onda differenti, sono seguiti da un moltiplicatore ottico che ha il compito di affiancare tutti i canali da trasmettere sulla stessa fibra (Figura 3). Un demoltiplicatore ottico con funzioni complementari a quelle del moltiplicatore riporta in ricezione tutti i canali alla stessa lunghezza d'onda. La lunghezza tipica delle sezioni di amplificazione è di 80 km. La rigenerazione elettrica fatta nel terminale ricevente annulla gli effetti di tutte le degradazioni (legate alle fibre e agli apparati) introdotte lungo il collegamento.

A seconda delle applicazioni, la spaziatura tra i canali in termini di lunghezza d'onda (o, equivalentemente, di frequenza) è diversa. Da questo punto di vista i sistemi WDM si dividono in due grandi categorie: i sistemi

DWDM (Dense WDM) con spaziature tra i canali dell'ordine dei 100 GHz in frequenza (circa 0,8 nm in lunghezza d'onda) e i sistemi CWDM (Coarse WDM) con spaziatura tra i canali di 20 nm in lunghezza d'onda (circa 1600 GHz in frequenza).

I sistemi DWDM, stante la piccola spaziatura in frequenza tra i canali, hanno un'elevata capacità in termini di numero di canali trasmessi sulla stessa coppia di fibre; campo tipico di applicazione per questi sistemi è costituito oggi dalle reti dorsali. I sistemi più diffusi consentono la trasmissione di 40 canali a 10 Gbit/s spazati di 100 GHz nella banda C (1530-1565 nm), su distanze di oltre 600 km senza rigenerazione elettrica intermedia [4]. Sono disponibili sul mercato anche soluzioni più avanzate con 80 e 160 canali a 10 Gbit/s, con spaziature inferiori a 100 GHz e allocazione nelle bande C + L, che possono coprire senza rigenerazione intermedia distanze superiori a 1000 km [4]. In applicazioni particolari, come i collegamenti transatlantici, si sono superate anche distanze di 8 mila km senza rigenerazione intermedia; questo risultato è stato reso possibile soprattutto dalla particolarità dell'ambiente di posa e dall'adozione di sezioni di amplificazione relativamente corte (50 km). L'impiego di soluzioni così avanzate accresce l'importanza di alcuni aspetti del progetto e della gestione dei sistemi DWDM, quali, per esempio quelli di seguito riportati:

a. al crescere della distanza tra due punti di rigenerazione elettrica consecutivi, il problema del cumulo degli effetti delle degradazioni dovute alle fibre ottiche e agli apparati (soprat-

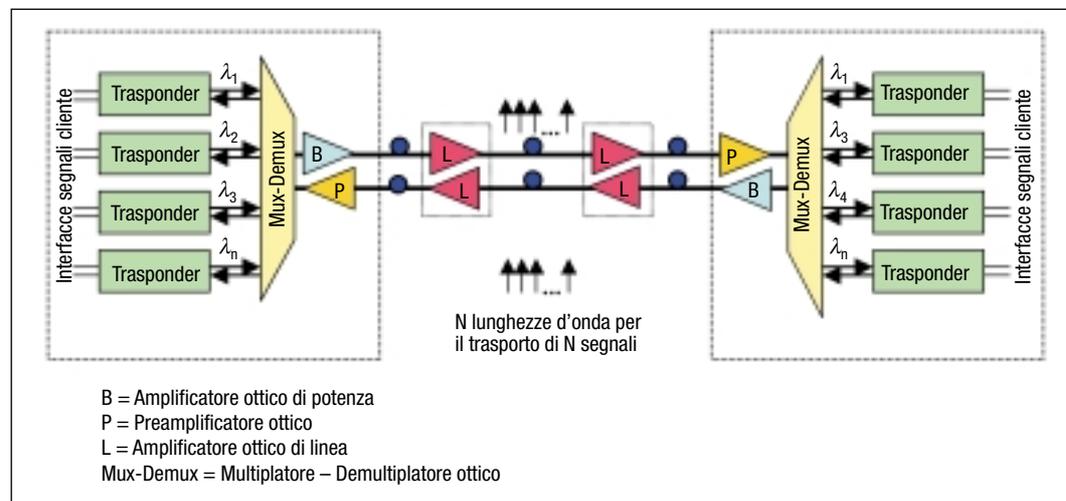


FIGURA 3
Schema di un sistema ottico WDM

tutto rumore introdotto dagli amplificatori ottici) diventa sempre più difficile da gestire. Sono quindi necessarie tecniche sempre più raffinate di compensazione delle dispersioni, di equalizzazione, di regolazione dei livelli ecc.;

b. sistemi DWDM dei tipi sopra indicati danno la possibilità di trasmettere su una coppia di fibre capacità superiori a 1 Tbit/s (1 Tbit/s = 1000 Gbit/s). Considerato che i cavi ottici contengono sempre alcune decine di fibre (l'ultima rete di cavi ottici posata da Telecom Italia, la T-Bone, ha 48 coppie di fibre per cavo) si vede come la capacità trasmessa su un cavo ottico può arrivare a diverse decine di Tbit/s. A fronte di capacità di quest'ordine di grandezza trasmissibili su un solo cavo, gli effetti di possibili guasti avrebbero conseguenze disastrose sul traffico su esse instradato, è perciò necessario adottare strutture di rete ridondate (ad anello ovvero a maglia) con opportune capacità di scorta per effettuare, con tempestività, il reinstradamento dei flussi interessati dal guasto.

I sistemi CWDM, stante l'elevata spaziatura tra i canali, sono caratterizzati da un limitato numero di canali (non superiore a 18 con spaziatura di 20 nm) [5] e dal fatto che essi occupano un'estesa gamma di lunghezze d'onda che va da 1260 a 1610 nm (banda O + E + S + C + L). La frequenza di cifra massima oggi prevista dalla normativa ITU-T per tali sistemi è di 2,5 Gbit/s. Le distanze massime sono dell'ordine di 80 km, pari cioè ad una tratta senza amplificazione. Non sono infatti ancora disponibili amplificatori per l'intera gamma 1260 -1610 nm.

I sistemi CWDM trovano quindi applicazioni ideali nelle reti metropolitane. La realizzazione di sistemi CWDM a 16-18 canali richiede inoltre la disponibilità di fibre caratterizzate dall'assenza del picco di attenuazione dovuto all'ossidrilico OH⁻, fibre che oggi, come si è detto, sono però ancora poco diffuse. Le applicazioni attualmente più comuni sono quindi limitate a 8 canali nella banda 1460 -1610 nm. A fronte di queste limitazioni i sistemi CWDM hanno però il vantaggio di costare molto meno dei sistemi DWDM, perchè l'ampia spaziatura tra i canali consente l'impiego di componenti più a buon mercato, quali sorgenti laser non controllate in temperatura e filtri meno sofisticati.

3.3. Criteri di progetto dei collegamenti ottici

L'obiettivo di progetto per un collegamento ottico compreso tra due punti di rigenerazione elettrica è quello di garantire un tasso di errore, *BER* (*Bit Error Rate*), migliore di 1×10^{-12} . L'ammontare complessivo delle degradazioni che si accumulano lungo il collegamento deve essere perciò mantenuto entro limiti ben precisi in modo da rispettare questo obiettivo. I parametri degli apparati e delle fibre da considerare sono, come si è visto, assai numerosi; il progetto non è quindi semplice e, naturalmente, diventa sempre più complesso al crescere della lunghezza del collegamento, della frequenza di cifra dei canali, del numero dei canali trasmessi su una stessa coppia di fibre [6].

A titolo indicativo si danno qui di seguito alcuni criteri suggeriti dall'ITU-T [7] per il progetto di un collegamento ottico a 10 Gbit/s per un singolo canale. Nell'esempio è in particolare considerato un sistema a 10 Gbit/s con una potenza immessa in fibra all'uscita del trasmettitore di 0 dBm e una potenza accettata all'ingresso del ricevitore atta a soddisfare l'obiettivo di *BER* di cui sopra, pari a -24 dBm. Entrambi i valori di potenza sono considerati a fine vita e comprendono quindi i margini di deterioramento degli apparati stessi. Sono quindi disponibili 24 dB per compensare le degradazioni introdotte nel collegamento (Figura 4).

Sempre in [7] l'attenuazione delle fibre a 1550 nm è assunta essere 0,27 dB/km, comprendendo in questo valore l'attenuazione della fibra (0,22 dB/km), quella dei giunti, quella dei connettori terminali e il margine di manutenzione del cavo, atto a consentirne il

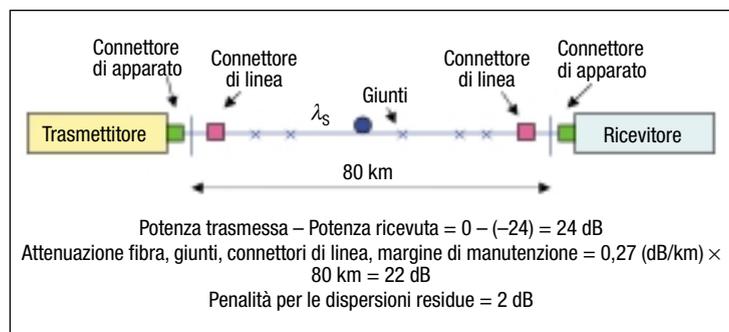


FIGURA 4

Esempio di budget di potenza per un sistema TDM a 10 Gbit/s

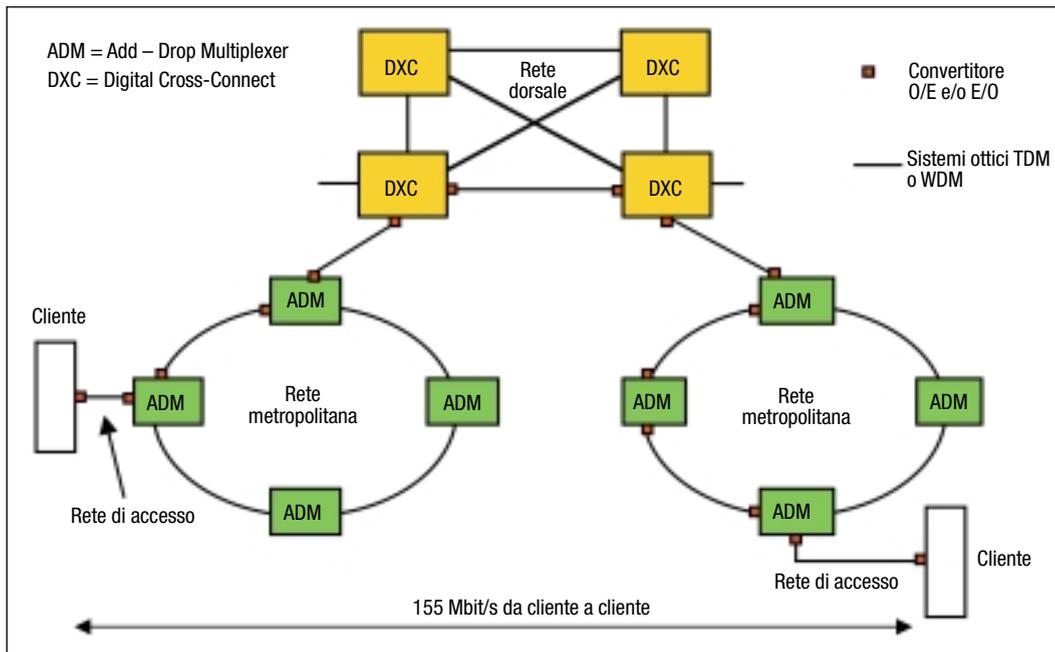


FIGURA 5
 Schema di un generico collegamento in una rete SDH

Un secondo punto di debolezza è relativo, in particolare, alla gerarchia di multiploazione SDH progettata per un'utilizzazione prevalente con segnali telefonici. La rete SDH presenta quindi alcuni limiti in una visione delle telecomunicazioni sempre più orientata al trasporto dei dati. In particolare, i dati devono essere sempre mappati nei *payload* SDH per essere trasportati sui sistemi ottici. Non è infatti possibile una trasmissione diretta al livello ottico, in quanto tutte le funzionalità di gestione sono concentrate negli apparati SDH. Gli attuali sistemi ottici di trasmissione non hanno alcuna funzione di controllo delle prestazioni, ma hanno solo alcune funzioni di allarme (mancanza della potenza ottica). In particolare, per esempio, i codici di linea NRZ (*Non Return to Zero*) e RZ (*Return to Zero*) non presentano alcuna ridondanza che consenta di effettuare una valutazione, anche se grossolana, delle prestazioni di errore al livello ottico, proprio perchè sono stati sviluppati per operare in sinergia con la SDH. L'ultima generazione di apparati SDH, arricchita di nuove prestazioni per essere più aperta al mondo dei dati (funzionalità di *Generic Frame Procedure* e *Virtual Concatenation*), mantiene però la necessità di dovere in ogni caso utilizzare lo strato SDH e sostenere il costo dei relativi apparati.

5. OTN: LA RETE DI TRASPORTO OTTICA

Con l'OTN (*Optical Transport Network*) si persegue l'obiettivo di eliminare i punti di debolezza della rete SDH indicati al paragrafo precedente e, allo stesso tempo, di fare un ulteriore passo verso una Rete Tutta Ottica. Le caratteristiche della OTN sono ormai ben definite in una serie di Raccomandazioni ITU-T [8]. Una OTN è composta da un insieme di elementi di rete ottici ONE (*Optical Network Element*), connessi da collegamenti in fibra ottica, atti a fornire funzionalità di trasporto, multiploazione, instradamento, gestione, supervisione di canali ottici che trasportano segnali cliente, con un trattamento eseguito *predominantly* in regime fotonico. Il termine *predominantly* chiarisce che la OTN non è una rete tutta ottica. I punti qualificanti della OTN sviluppati nel seguito sono: l'architettura, la gerarchia dei canali ottici, la loro trama e una certa trasparenza ottica [9]. In analogia con la rete SDH, l'architettura della rete OTN è strutturata su tre livelli: canale ottico, sezione di multiploazione dei canali ottici, sezione di trasmissione (tra terminali e/o tra amplificatori ottici adiacenti). Ciascuno di questi tre livelli ha precisi obiettivi nel quadro generale del controllo e della gestione. Per il trasporto di segnali cliente l'OTN definisce tre tipologie di contenitori: la prima

(ODU-1) adatta a trasportare segnali cliente con frequenza di cifra fino a 2,5 Gbit/s; la seconda (ODU-2) per segnali da 2,5 a 10 Gbit/s; la terza (ODU-3) per segnali da 10 a 40 Gbit/s. Tali tipologie di contenitori costituiscono anche gli elementi di una gerarchia ottica OTH (*Optical Transport Hierarchy*) che consente il trasporto di quattro ODU-1 in un ODU-2 e di quattro ODU-2 in un ODU-3, nonché diverse altre combinazioni di canali. Segnali cliente di ogni tipo (IP, ATM, Ethernet, STM-N) possono essere mappati in questi contenitori: all'ingresso della rete è previsto che i trasponder, oltre ad assegnare al canale ottico la giusta lunghezza d'onda, provvedano anche ad inserire i singoli segnali nei canali ottici (a seconda, naturalmente, della capacità che essi hanno, in modo da ottenere il massimo rendimento) e a completare con cifre di riempimento la parte della trama che rimane libera. Questa caratteristica della OTN garantisce un'elevata trasparenza ai segnali cliente all'interno del canale ottico prescelto. La trama di ciascuno dei tre tipi di canale è costituita da un *payload*, nel quale sono mappati i dati, e da un *overhead*, dove sono contenute tutte le informazioni necessarie per la gestione delle configurazioni e degli allarmi, nonché per il controllo delle prestazioni. Con canali ottici OTN così strutturati (*payload + overhead*) si supera il vincolo dell'impiego del livello SDH, in quanto ciascun canale ottico è dotato di tutte le risorse necessarie per la propria gestione. I segnali cliente dati possono essere quindi immessi e trasportati sui canali ottici senza dover essere prima mappati nelle trame SDH. D'altra parte i canali cliente SDH possono trovare posto nei canali ottici al pari dei segnali dati. Una caratteristica della OTN è la trasparenza ottica, cioè la capacità della rete di trasportare segnali ottici su lunghe distanze e attraverso i suoi nodi senza necessità di conversioni O/E e E/O. Questa trasparenza non è assoluta per via di diverse limitazioni

(che rendono conto della parola *predominantly* posta nella definizione) dovute sia al già citato accumulo delle degradazioni nella trasmissione ottica sia al fatto che non tutte le funzionalità di rete possono al momento essere svolte al livello ottico. La tecnologia rende infatti disponibili oggi apparati per la derivazione e per l'introduzione di canali ottici OADM (*Optical Add Drop Insert Multiplexer*), ma per i ripartitori ottici PXC (*Photonic Cross-Connect*) la situazione è più articolata. I PXC attualmente sul mercato hanno infatti ancora caratteristiche insufficienti per un loro estensivo impiego in rete sotto l'aspetto dell'affidabilità, dell'ingombro, del consumo e soprattutto, del costo. Dal punto di vista della trasparenza ottica, l'OTN si presenta quindi come costituita da un insieme di isole ottiche trasparenti (Figura 6) di dimensioni limitate (100-200 km), all'interno delle quali i canali ottici sono sempre mantenuti al livello ottico, collegate tra loro da rigeneratori 3R (per eliminare le degradazioni accumulate) e/o da permutatori per canali ottici, che effettuano tuttavia le permutazioni al livello elettrico.

Un notevole arricchimento dell'OTN è dato dalla ASON (*Automatic Switched Optical Network*) [10]. Un'architettura ASON è basata sulla realizzazione di un piano di controllo in grado di svolgere alcune funzioni tipiche dei sistemi di gestione centralizzati e posto tra i vari elementi di rete ottici della OTN e il piano di gestione (Figura 7). A questo scopo ogni ONE dovrebbe essere dotato di un OCC (*Optical Connection Control*) in grado di interconnettersi con gli altri OCC e con il sistema di gestione centralizzato e di impartire comandi a uno o più ONE. Si vogliono così rendere più semplici, più diretti e più rapidi i processi di attivazione, di cessazione, di modifica di istruzione di ciascun canale ottico, facendo svolgere le operazioni direttamente dagli organi di controllo presenti negli ONE dove sono

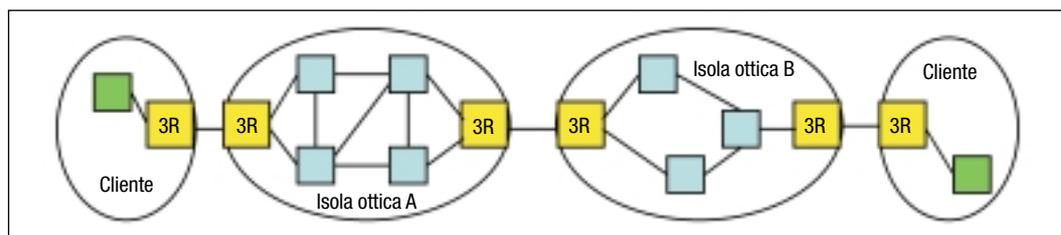


FIGURA 6
La trasparenza ottica nella OTN

raccolte e continuamente aggiornate tutte le informazioni necessarie.

Un'architettura ASON con intelligenza distribuita presenta numerosi vantaggi rispetto ad un'architettura tradizionale con un'intelligenza concentrata nel sistema di gestione. I tempi per l'esecuzione delle operazioni di configurazione si riducono a qualche decina di millisecondi o a qualche secondo a fronte di qualche decina di minuti necessari con il sistema centralizzato. Inoltre l'ASON consente non solo di migliorare la qualità dei servizi di trasporto offerti (in termini di tempi di attivazione e di reistrazione in caso di guasti), ma anche di offrire nuovi servizi come la *Bandwith on Demand* e la *Optical Virtual Private Network*. I clienti possono con il primo servizio attivare e rimuovere servizi di connettività in tempi brevi e con tariffazione a tempo e con il secondo effettuare autonomamente le operazioni di configurazione dei flussi trasmissivi della propria rete. Le Raccomandazioni ITU-T per la ASON sono ormai praticamente complete [11], anche se le prime sperimentazioni in campo iniziate di recente (in Italia sono state avviate da 8-10 mesi) sono ancora basate su prodotti di un singolo produttore (gli ONE e il Piano di Controllo sono dello stesso fornitore e non sono possibili interconnessioni tra Piani di Controllo di fornitori diversi)

6. AON: LA RETE TUTTA OTTICA

La Rete Tutta Ottica costituisce per i gestori l'obiettivo finale dell'evoluzione della rete di trasporto, poichè essa consente di eliminare quasi completamente i rigeneratori elettrici posti all'interno della rete di trasporto e permette di avere una completa trasparenza ai servizi. L'evoluzione della OTN verso la AON dovrebbe avvenire con una graduale estensione della superficie delle isole ottiche fino al costituirsi di un'unica isola ottica (Figura 8) nella quale il se-

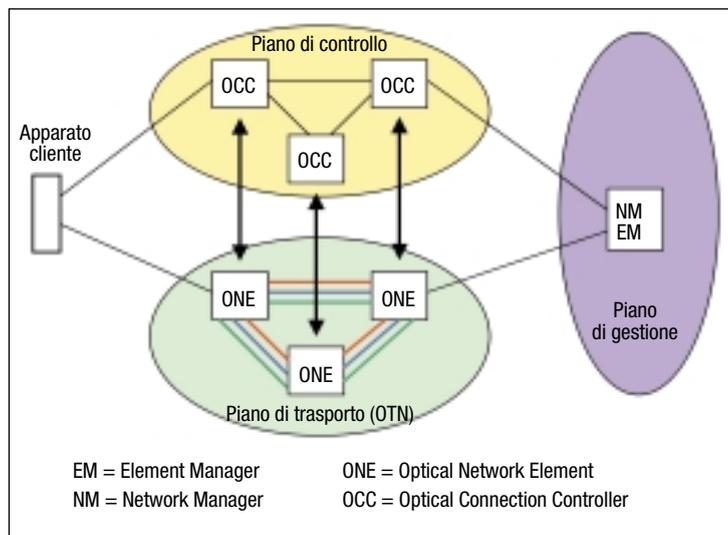


FIGURA 7
Architettura di una ASON

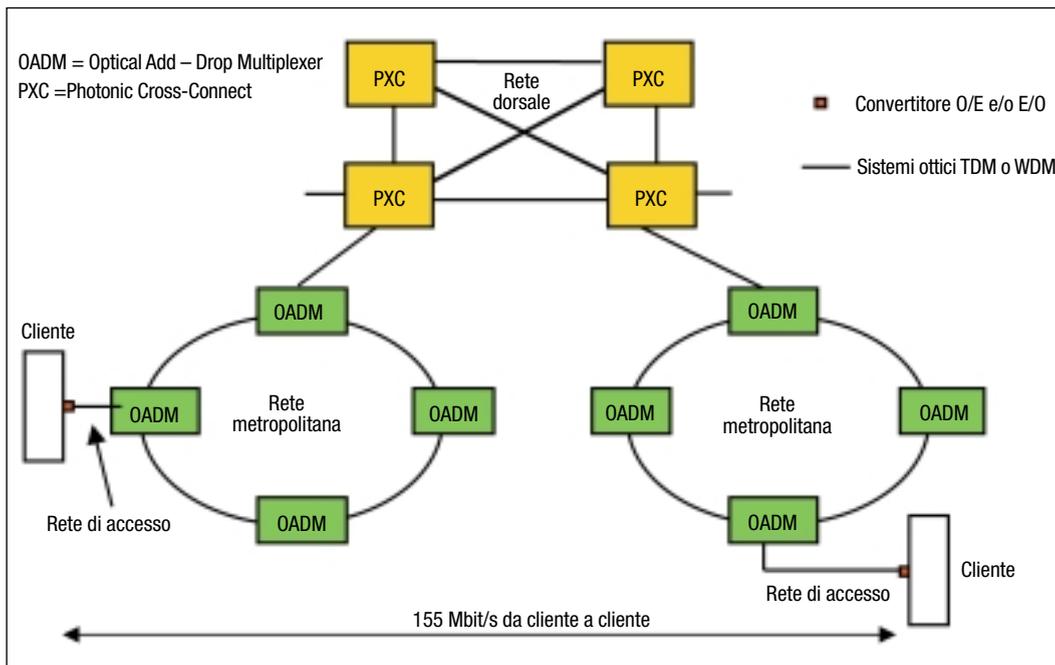


FIGURA 8
Schema di un generico collegamento in una rete AON

gnale cliente subisce una conversione elettro-ottica nel punto di origine (sede A del cliente) e una conversione ottico/elettrica nel punto di destinazione (sede B del cliente). Questa evoluzione sarà possibile a mano a mano che saranno disponibili gli elementi di rete che operano al livello ottico finora mancanti (PXC, convertitori di lunghezza d'onda, rigeneratori ottici ecc.) e che, al contempo, saranno ottimizzate le modalità di compensazione delle degradazioni che si accumulano al livello ottico (equalizzazione, regolazione ecc.). In realtà la realizzazione di una AON non è solamente legata alla disponibilità di apparati e di componenti più evoluti, ma anche ad altri vincoli posti dalla operatività e dalla gestione della rete di trasporto [12, 13].

Allo stato attuale una rigenerazione al livello elettrico sembra essere sempre necessaria nel punto di interfaccia tra la rete del cliente e la rete dell'operatore allo scopo di mantenere ben distinte e separate le responsabilità a fronte del rispetto del SLA (*Service Level Agreement*). Uno o due punti di rigenerazione dovrebbero rimanere ancora presenti tra i diversi livelli di rete (accesso, rete metropolitana, connessioni dorsali) per effettuare operazioni di raggruppamento (*grooming*) dei segnali cliente di piccola media / capacità in flussi di capacità superiore, per avere un'elevata percentuale di utilizzazione dei sistemi di trasmissione.

Un altro punto critico di una rete tutta ottica riguarda la difficoltà di mantenere sotto controllo la qualità dei canali ottici [14]. La misura diretta delle prestazioni di errore, i cui obiettivi sono definiti nei SLA, è infatti possibile solo al livello elettrico. Le misure effettuate al livello ottico (*optical monitoring*) come la potenza trasmessa e ricevuta dai singoli canali e dai segnali multiplati, il rapporto segnale / rumore ottico e la deviazione di frequenza dei canali, mentre consentono di effettuare una corretta gestione dei guasti, danno solo un'"indicazione" di come stanno andando le prestazioni di errore e non una misura delle prestazioni stesse.

Per rendersi conto della entità del problema basti pensare ad un canale ottico Venezia-Napoli nel quale si accumulerebbero le degradazioni di 800 km di cavo ottico, di sedici amplificatori ottici bidirezionali e di quat-

tro permutatori ottici. Ove si registrasse all'interfaccia tra la rete e il cliente finale una degradazione, per esempio, del tasso di secondi con errore (senza un'interruzione del canale ottico), le "indicazioni" ottenute dai parametri ottici controllati sarebbero quasi sempre tardive (rispetto al momento nel quale il cliente avverte la degradazione) e spesso insufficienti a individuare con tempestività e precisione le cause e il punto del degrado. Un miglioramento di questa situazione si verificherà con il diffondersi dei punti di misura dei parametri di qualità ottici in tutti i punti critici degli apparati e con l'acquisizione di una vasta messe di dati e di correlazioni. Tuttavia, allo stato attuale, una soluzione soddisfacente del problema non sembra vicina e quindi anche in una rete tutta ottica sarà necessario disporre di rigeneratori elettrici sui quali terminare provvisoriamente in più punti successivi il canale degradato, per localizzare il punto in cui si origina la degradazione del segnale.

A fronte delle criticità insite nelle nuove tecnologie, delle esigenze di raggruppamento di canali, delle carenze della supervisione al livello ottico, alcuni gestori cominciano a chiedersi se quello di una Rete Tutta Ottica sia veramente l'obiettivo da raggiungere o se invece sia solo un obiettivo a cui tendere. A questa riconsiderazione degli obiettivi non è estranea la comparsa sul mercato di nuovi trasponder realizzati con tecnologia SFP (*Small Form Factor Pluggable*) che hanno già oggi dimensioni, consumi e costi inferiori a quelli tradizionali, e che lasciano intravedere la possibilità di essere realizzati in nuove versioni di costo ben inferiore. L'impiego di questi trasponder potrebbe infatti ridurre di molto l'onere complessivo legato alla presenza di punti di conversione ottico-elettrico-ottico.

7. CONCLUSIONI

Le fibre ottiche e i sistemi di trasmissione ottici, a venti anni dalla loro prima introduzione nella rete di trasporto, hanno praticamente sostituito i conduttori in rame nelle reti metropolitane e nelle reti dorsali in tutte le applicazioni (collegamenti terrestri e sottomarini di ogni tipo e di ogni capacità). Tuttavia, l'attuale

rete di trasporto SDH è rimasta sostanzialmente elettrica perchè, anche se i nodi sono collegati con sistemi ottici, tutte le operazioni che sono effettuate sui segnali nei nodi stessi sono ancora svolte al livello elettrico.

Solo di recente sono state standardizzate architettura, funzionalità e apparati per una *Rete di Trasporto Ottica* (OTN) in grado di costituire e gestire canali ottici trasparenti all'interno di isole ottiche di limitata dimensione e le prime realizzazioni di questa rete sono ora in corso. Di pari passo con la OTN sta procedendo lo sviluppo di una architettura di rete ottica intelligente (ASON) che ne costituisce un arricchimento realizzando un Piano di Controllo atto ad effettuare in modo automatico le funzioni di gestione delle configurazioni dei canali ottici.

La strada da percorrere per arrivare alla realizzazione di una Rete Tutta Ottica (AON) non sembra nè breve, nè facile, in quanto mancano ancora alcuni elementi essenziali (i permutatori fotonici) e allo stesso tempo perchè devono essere risolte numerose problematiche di gestione e, più in particolare, quelle legate ai limiti intrinseci del monitoraggio ottico. Tuttavia i vantaggi che le società manifatturiere e i gestori di rete vedono all'orizzonte fanno ritenere che il cammino verso questo obiettivo procederà ancora speditamente, salvo l'effettuazione di periodiche verifiche per definire quale sia veramente il punto di arrivo più conveniente.

In questa rapida analisi dell'evoluzione della rete di trasporto verso una soluzione tutta ottica si è focalizzata più volte l'attenzione sugli standard e sui problemi di gestione, che spesso sono un pò trascurati nella letteratura tecnica, ma che, a parere dello scrivente,

più di altri decretano il successo o il fallimento di una tecnologia.

Bibliografia

- [1] Montalti F.: Fibre ottiche per sistemi DWDM. *Notiziario Tecnico Telecom Italia*, n. 3, dicembre 2001.
- [2] *Fibre ottiche*: Raccomandazioni ITU-T G.652, G.653, G.654, G.655, G.656.
- [3] *Sistemi ottici TDM*: Raccomandazioni ITU-T G.957, G.691, G.693.
- [4] *Sistemi DWDM*: Raccomandazioni ITU-T G.694.1, G.692, G.696.1.
- [5] *Sistemi CWDM*: Raccomandazioni ITU-T G.694.2, G.695.
- [6] Aureli G., Pagnan P.: I sistemi DWDM: problematiche trasmissive e loro impatto sul progetto dei collegamenti. *Notiziario Tecnico Telecom Italia*, n. 2, dicembre 2000.
- [7] *Optical system design and engineering considerations*. ITU-T Supplemento n. 39.
- [8] *OTN*: Raccomandazioni ITU-T G.872, G.709.
- [9] Aureli G., Pagnan P.: La Rete di Trasporto Ottica (OTN): stato dell'arte e prospettive evolutive. *Notiziario Tecnico Telecom Italia*, n. 1, dicembre 2001.
- [10] Pagnan P.: La gestione del backbone di trasporto ottico. *Notiziario Tecnico Telecom Italia*, n. 2, dicembre 2002.
- [11] *ASON*: Raccomandazioni ITU-T G.8080, G.7712 / G.7717.
- [12] Bonaventura G., Pagnan P., Augusto S.: *La Rete Tutta Ottica: problemi aperti*. Fotonica 2003.
- [13] Bonaventura G., Pagnan P., Augusto S.: *Conditions for developing All Optical Islands as a part of a Cost-Effective and Future-proof Migration Strategy to an All Optical Network*. IIR Conference, Londra, aprile 2002.
- [14] Bonaventura G., Pagnan P., Farina T.: *Optical monitoring for DWDM systems*. IIR Conference, Cannes, giugno 2004.

GASTONE BONAVENTURA ha speso la vita professionale operando con diversi gestori di telecomunicazioni (Azienda di Stato per i Servizi Telefonici, Italcable, SIP, STET, Telecom Italia).

Per conto di Telecom Italia, è oggi Vice-Presidente della Commissione 15 (Reti Ottiche di Trasporto) dell'ITU-T e Presidente del Working Party "Tecnologia Ottica". È anche docente di "Impianti di Telecomunicazioni" nel Corso post-lauream per ingegneri tenuto dal Ministero delle Comunicazioni.

gastone.bonaventura@telecomitalia.it

L'EVOLUZIONE DELLE MEMORIE PERIFERICHE

L'avvento delle LAN ha favorito lo sviluppo di tecnologie come NAS e SAN con le quali, attraverso diversi strati hardware e software, possono essere interconnessi più dispositivi per formare un unico complesso di memoria. Si ottengono così sistemi altamente affidabili, con elevate prestazioni, facilmente gestibili e a costi ragionevoli. Per la tecnologia di registrazione magnetica continua il processo di miniaturizzazione, mentre, in prospettiva, si profilano nuove tecnologie, come le memorie olografiche.

1. LA CRESCITA DELLE INFORMAZIONI

Secundo uno studio della *School of Information Management and Systems* della University of California di Berkeley [1], dal 1999 al 2002 la produzione mondiale di informazioni è cresciuta del 30%.

Le nuove informazioni prodotte nel corso del solo anno 2002 sono equivalenti a mezzo milione di nuove biblioteche, ognuna contenente una versione digitalizzata dell'intera *Library of Congress* degli USA.

Stampa, pellicola, carta, supporti magnetici e ottici contengono oggi circa 5 Ebyte, ossia cinque miliardi di miliardi di byte, di nuova informazione, di cui il 92% immagazzinato su supporto magnetico, il 7% su pellicola, lo 0.002% su supporto ottico e, infine, lo 0.015% su supporto cartaceo.

Analizzando con attenzione i dati della ricerca si può individuare la reale natura di una simile esplosione di informazioni. Mentre i flussi e gli aggregati tradizionali di dati crescono a tassi contenuti quella che letteralmente esplose è la nuova informazione memorizzata

sui supporti informatici. In sostanza è il settore dell'informatica il principale attore di questa crescita di nuova informazione dal momento che la televisione e la radio crescono ormai in modo asintotico, i giornali a loro volta aumentano solo grazie al contributo dei paesi emergenti e, infine, la carta cresce solo per la produzione di documenti d'ufficio.

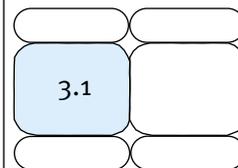
La ricerca indica anche un decremento significativo, a partire dal 1999, nell'uso della pellicola in ambito fotografico a fronte di un rapido sviluppo sul mercato delle fotocamere digitali. Lo stesso fenomeno sta avvenendo nel cinema, sempre più incline a sfruttare la tecnologia digitale, anche per i minori costi di realizzazione dei film.

L'unico segmento che per ora sembra rimanere fedele alla pellicola tradizionale è quello medico, in cui lo sviluppo di radiografie registra una crescita del 16%.

La crescita di nuove informazioni immagazzinate su supporto magnetico, a partire dal 1999, è stata pari all'80%. Grazie ai costi decrescenti e alla crescente varietà di standard e di dimensioni, i dispositivi di memo-



Ernesto Hofmann



ria ad accesso diretto, ossia i dischi magnetici, sono diventati il segmento a più rapida crescita.

I dispositivi ottici rappresentano il supporto più utilizzato per la distribuzione di software, dati, film e musica, come anche di altri tipi di informazioni digitali.

Il declino nella produzione e vendita dei CD tradizionali è stato controbilanciato dalla crescente diffusione dei CD riscrivibili e soprattutto dei DVD, che hanno conquistato quote di mercato ad un ritmo ancor più veloce delle tecnologie precedenti.

Tuttavia, c'è un nuovo fenomeno che darà probabilmente un ulteriore impulso alla continua crescita dell'informazione digitale.

Infatti, in un futuro abbastanza prossimo, qualunque entità avrà valore sarà interconnessa con qualche altra entità in un'infrastruttura che ormai sempre più spesso viene denominata l'"Internet delle cose".

Una prima evidenza di questa straordinaria trasformazione si avrà in tempi abbastanza brevi nell'ambito dell'industria della distribuzione.

I cosiddetti RFID tag (identificatori a radiofrequenza) possono, in modo analogo ai codici a barre, essere apposti o inseriti in oggetti per essere successivamente letti o aggiornati da un sistema informativo, senza intervento diretto di una persona, per mezzo di un'antenna opportunamente posizionata.

Per identificare univocamente i prodotti lungo la catena di distribuzione occorre peraltro un'infrastruttura standardizzata. Ciò viene reso possibile dall'adozione dell'*Electronic Product Code* (EPC), che è in sostanza un'evoluzione dell'*Universal Product Code* utilizzato nel codice a barre.

L'EPC è un codice costituito da un'etichetta virtuale che identifica il costruttore, la classe di prodotto e il numero di serie. In funzione del tipo di tag l'EPC può riconoscere fino a 268 milioni di costruttori diversi, ciascuno con 16 milioni di prodotti. Ogni prodotto, a sua volta, può contenere fino a 68 miliardi di unità. Siamo quindi nell'ordine di migliaia di miliardi di unità di prodotto identificabili nell'ambito dell'infrastruttura EPC.

Se si moltiplicano tali numeri per le quantità dei singoli prodotti, diventa intuibile quanto grande possa diventare l'ulteriore informa-

zione digitale generata da un'infrastruttura RFID operativa su larga scala.

2. DAL COMPUTER ALLE UNITÀ PERIFERICHE

La storia dei supporti per la memorizzazione dei dati informatici ha avuto inizio con la nascita stessa del computer, sia all'interno del computer (memoria centrale) sia all'esterno, per mezzo di opportuni dispositivi elettromeccanici collegati col computer. Il primo supporto di memoria esterna è stato la carta. I programmi e i dati erano registrati sotto forma di aree intatte (zero) o perforate (uno) su opportune schede di cartoncino, divise in ottanta colonne, del formato di un dollaro. Ma già intorno al 1950 è avvenuta una prima grande rivoluzione con l'introduzione dei nastri magnetici, già utilizzati per le applicazioni audio. Uno dei maggiori vincoli strutturali dei nastri era la lettura sequenziale che imponeva di scorrere meccanicamente il nastro stesso per raggiungere l'informazione desiderata, con evidente degrado delle prestazioni nell'accesso alle informazioni. Tuttavia i nastri magnetici rappresentano comunque un'importante tecnologia per l'archiviazione delle informazioni, soprattutto nelle medie e grandi imprese. Il vero "breakthrough" tecnologico nella memorizzazione delle informazioni digitali su supporto esterno al computer è stato però la creazione nel 1955 del cosiddetto *Hard Disk Drive*, ovvero HDD, spesso anche denominato DASD (*Direct Access Storage Device*). I progettisti avevano compreso, infatti, che fosse possibile realizzare una testina di lettura elettromagnetica, sospesa sopra un piatto di alluminio rotante ad alta velocità, in grado di leggere le informazioni che scorrevano rapidamente sotto la testina stessa. Il punto chiave della scoperta era che la testina non doveva essere a contatto con il piatto, che poteva ruotare ad alta velocità e consentire quindi di raggiungere l'informazione desiderata "quasi direttamente", a differenza dei nastri [2]. Il primo HDD prodotto è stato il 305 RAMAC (*Random Access Method of Accounting and Control*) dell'IBM, annunciato il 13 settembre del 1956. Tale dispositivo poteva memorizzare 5 milioni di caratteri (circa cinque megabyte, ma un

“carattere” era costituito allora da sette bit su 50 grandi dischi, ciascuno di 24 pollici (circa 60 cm) di diametro. La densità di registrazione era di circa 2000 bit/pollice² (ossia 310 bit/cm²). I dispositivi a dischi attuali hanno densità dell'ordine di miliardi di bit per centimetro quadrato. La velocità di trasferimento dei dati era comunque già molto elevata: 8.800 byte/s.

Negli anni successivi la tecnologia è progressivamente migliorata, sia nella densità di registrazione, sia nella capacità di memorizzazione complessiva, sia infine nei tempi di accesso alle singole unità di registrazione (i cosiddetti *record*). Nel 1962 ancora l'IBM annuncia il modello 1301 *Advanced Disk File*. L'aspetto tecnico più significativo di questo nuovo dispositivo era la possibilità di inserire un vero e proprio “cuscinetto di aria” tra testina di lettura e piatto rotante tale che la distanza tra i due si poteva ridurre da 800 a 250 μ m (micropollici), ossia da circa due centesimi di millimetro a circa sei millesimi di millimetro.

Nel 1973, infine, l'IBM annunciava il *Disk Drive* modello 3340, che è oggi considerato il progenitore dei moderni dischi. Questo dispositivo era costituito da due dischi separati, uno permanente e uno rimovibile, ciascuno con una capacità di memorizzazione di 30 Mbyte. Per questo motivo fu colloquialmente denominato 30-30; e poiché anche un modello del famoso fucile Winchester era denominato 30-30, il disco venne da allora chiamato anch'esso Winchester. Utilizzando evolute tecniche di isolamento dalle impurità dell'aria esterna e nuove tecniche micromeccaniche fu possibile ridurre “l'altezza di volo” delle testine a soli 12 μ m (micropollici), ossia a circa tre decimillesimi di millimetro.

3. LE MEMORIE DI MASSA E IL RUOLO DEGLI HDD

L'insieme dei dispositivi esterni di memoria, genericamente denominato memorie di massa, si compone approssimativamente di tre grandi famiglie di dispositivi.

■ **Nastri magnetici:** sono soprattutto utilizzati nei centri medio-grandi di elaborazione per creare periodicamente copie (*backup*) del contenuto degli HDD, in modo da poter recuperare i dati in caso di malfunzionamenti di

varia natura. Le operazioni di backup sono però abbastanza lente (anche alcune ore) e vengono in generale effettuate di notte.

I nastri magnetici sui quali vengono trascritte le informazioni hanno una vita media di 10-30 anni (mentre i microfilm si suppone che possano durare anche fino a 300 anni). In realtà per i nastri, come per altri supporti di memorizzazione dei dati, si è constatato che i dispositivi di lettura/scrittura si sono evoluti molto prima che i nastri stessi si deteriorassero. Anche nell'elettronica per uso domestico si assiste alla continua scomparsa di supporti come dischi 78 giri, 33 giri, nastri VHS, DAT, e altri ancora, a fronte di nuove tecnologie.

■ **Dischi ottici:** sono dischi su cui la memorizzazione dei dati avviene “bruciando” con un laser una sezione di superficie, che da lucida diventa opaca. In generale i dati scritti su di un disco ottico non possono più essere cancellati. Esistono tuttavia dei dischi particolari (CD-DVD riscrivibili) che consentono la cancellazione e la riscrittura per un numero comunque limitato di volte (ad ogni cancellazione la superficie tende a deteriorarsi sempre di più finché non diventa inutilizzabile). I DVD (*Digital Versatile Disk*) sono apparentemente molto simili ai CD, ma possono contenere fino a 25 volte la capacità di un normale CD e vengono usati soprattutto per memorizzare i film in formato digitale. I dischi ottici come CD, DVD, DVD/R, DVD/RW, hanno finora utilizzato per leggere (e scrivere) una tecnologia laser con una frequenza nello spettro del rosso. Tuttavia sono già pronti nuovi dischi ottici denominati Blu-ray proprio perché utilizzano il laser nella frequenza blu-violetto. I nuovi dispositivi di lettura potranno anche leggere dischi CD e DVD tradizionali. Il beneficio di utilizzare una lunghezza d'onda di 405 nm invece di 650 nm consente di scrivere un maggior numero di informazioni per unità di superficie. Nella tabella 1 vengono riportati i parametri più significativi delle tre tecnologie CD, DVD e Blu-ray.

Poiché il disco viene letto da un laser, non c'è contatto fisico tra supporto e lettore, e dunque la frequenza e la quantità delle riproduzioni non incide sulla qualità del contenuto. Una volta registrati su DVD, i dati non potranno più essere cancellati accidentalmente né subire alterazioni dannose. La vita media dei CD e

Tecnologia	CD	DVD	Blu-ray
Laser	780 nm	650 nm	405 nm
Numero di facce utili	1	1/2	1/2
Capacità massima	750 Mbyte	9.4 o 17 Gbyte	da 25 a 120 Gbyte

TABELLA 1

DVD di tipo R (sola lettura) si colloca tra 25 e 250 anni, con un valore medio di circa 100 anni. I DVD-RW hanno invece una vita più breve, dell'ordine di 25-100 anni con un valore medio di 50. Sono in atto continue ricerche per migliorare la resistenza dei materiali, per evitare l'ossidazione delle parti di superficie danneggiata da agenti come urti, sostanze chimiche e anche sbalzi termici. È stato osservato che mantenere CD e DVD costantemente al di sotto di 30° e al disotto dell'80% di umidità può raddoppiarne la vita media. I costruttori di dispositivi di lettura/scrittura sono peraltro attenti anche ad un altro parametro molto importante che riguarda il "drive", ossia il supporto rotante che sostiene il CD-DVD e che, proprio per avere organi meccanici in movimento, ha una vita media più breve. Un tipico valore che viene dichiarato per l'unità "drive" è un MTBF di 100.000 h di power on. Si vedrà meglio più avanti come debba intendersi il parametro MTBF (*Mean Time Between Failures*) che da solo può indurre in errori di valutazione. Essendo un anno costituito di poco meno di 8.800 h si potrebbe immaginare un utilizzo continuo di oltre dieci anni. In realtà è più realistico pensare ad un utilizzo a pieno tempo di un anno, che è in genere proprio la garanzia che viene fornita. I dispositivi di memoria ottica sono di particolare interesse. Se, infatti, si ignora la possibilità del disco ottico di essere rimosso dal suo supporto, il disco ottico sembrerebbe in ogni aspetto inferiore al disco magnetico. Ma se si considerano applicazioni che richiedono la distribuzione di programmi o dati si comprende allora come questa tecnologia possa diventare attraente. Essa quindi mantiene una notevole quota di mercato per la distribuzione di audio preregistrati e soprattutto per la distribuzione di film. La memorizzazione ottica resta tuttavia più costosa della memorizzazione su dischi magnetici in caso di grandi volumi e le sue prestazioni (unite al costo unitario) per l'accesso al singolo elemento di

informazioni (record) non la mettono in grado di competere con le memorie magnetiche.

■ **Dischi magnetici:** sono dischi sui quali la memorizzazione dei dati avviene magnetizzando la superficie, tramite un'apposita testina di lettura/scrittura. Sullo stesso disco i dati possono essere scritti, cancellati e riscritti per un numero indefinito di volte senza logorarne la superficie di registrazione.

Le memorie di tipo HDD (*Hard Disk Drive*) sono di gran lunga i dispositivi più importanti nell'attuale gerarchia di memoria della moderna informatica, come viene messo in evidenza dalla crescente quantità di investimenti in questa tecnologia da parte di imprese di ogni dimensione. Le caratteristiche più importanti degli HDD sono il costo per Mbyte, la velocità di trasferimento e il tempo di accesso alla singola informazione (al *record*).

Poiché il costo di questo tipo di memoria è tuttora decisamente inferiore a quello delle memorie statiche (RAM) su microchip, gli utenti sono disposti ad accettarne anche gli aspetti meno attraenti, come un tempo di accesso meno veloce rispetto alle memorie statiche centrali, un più elevato consumo di corrente elettrica, il rumore di fondo tipico delle unità meccaniche (dischi che ruotano), e infine l'intrinseca minore affidabilità dei componenti elettromeccanici rispetto a quelli puramente elettronici, come appunto i chip di memoria. D'altro canto gli HDD sono una tecnologia "non volatile", ossia non richiedono corrente elettrica per mantenere nel tempo l'informazione registrata, e ciò costituisce un elemento di particolare importanza nell'economia della conservazione dei dati di un'impresa.

4. LA STRUTTURA DEGLI HDD

L'HDD è costituito da una serie di "piatti", disposti l'uno sopra l'altro e fissati allo stesso perno centrale che li fa ruotare ad altissima

velocità (fino a 10.000 giri il minuto e oltre). I singoli bit (0 e 1) che compongono l'informazione digitale vengono memorizzati nello speciale materiale magnetico (sotto forma delle due direzioni del campo magnetico) di cui è costituita la superficie del piatto.

Le operazioni di lettura e di scrittura sono eseguite da una testina elettromagnetica che è sospesa sulla superficie magnetica del piatto in rotazione (generalmente realizzato in alluminio), senza mai toccarla così da non rovinare lo strato magnetico di registrazione. Ciascun piatto dispone di una testina per faccia. Le testine sono tutte fissate sul medesimo supporto, e si muovono quindi tutte insieme. L'involucro di un disco è realizzato in due parti che sono, rispettivamente, il basamento che ospita l'elettronica e tutte le parti meccaniche, e il coperchio che concorre alla dissipazione del calore.

Attraverso una speciale valvola dotata di opportuni filtri, l'aria all'interno dell'unità viene mantenuta priva di polvere e la valvola stessa fa sì che la pressione interna sia in equilibrio con quella esterna.

L'organizzazione logica delle due superfici di registrazione di ciascun piatto consiste in una struttura costituita da cerchi concentrici ("tracce") che hanno origine nel perno centrale del piatto e settori derivati dall'intersezione delle tracce con i raggi del piatto stesso.

La densità areale indica la quantità di bit che possono essere memorizzati su di una specifica unità di area e viene espressa di solito come bit per pollice quadrato (BPSI: *Bits Per Square Inch*).

Essendo una misura bidimensionale, la densità areale è il prodotto di due unità di misura unidimensionali che sono:

■ **la densità delle tracce**, che è una misura di quanto siano disposte densamente le tracce concentriche attraverso i raggi dei piatti, ossia quante tracce possono essere disposte concentricamente lungo un pollice di raggio. Tuttavia non potranno essere utilizzate tracce troppo vicine al perno centrale di rotazione, così come non potranno essere usate tracce troppo vicine alla circonferenza del piatto. Se, per esempio, il raggio del disco è di 1,87 pollici se ne potranno ragionevolmente utilizzare 1,2 pollici. Quindi, se lo spazio utile del disco è costituito di 22.000 tracce si

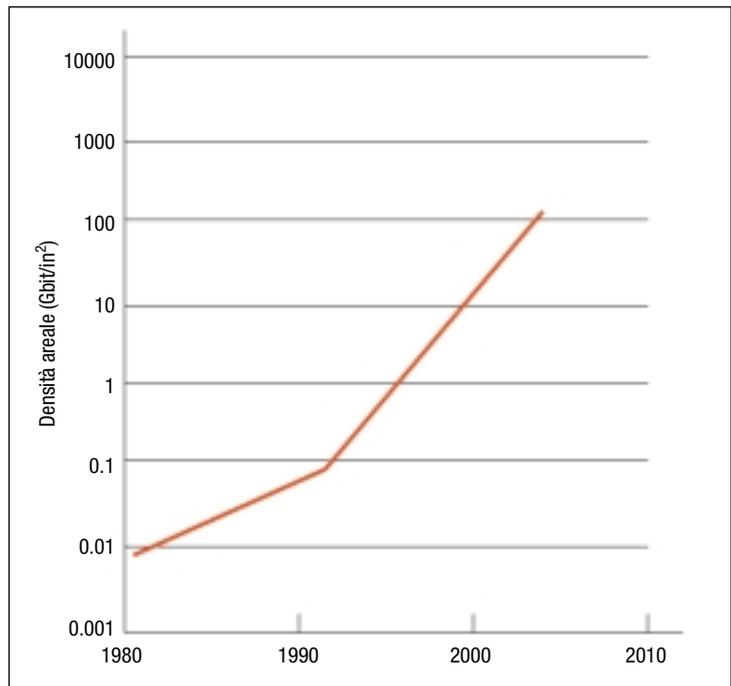


FIGURA 1

Aumento della densità di registrazione degli HDD negli ultimi 20 anni

può concludere che ci saranno circa 18.333 tracce per pollice (22.000 : 1,2).

■ **la densità lineare di registrazione**, che è una misura di quanto densamente siano registrati i bit lungo una traccia. Se lungo un pollice di traccia è possibile registrare 300.000 bit allora la densità BPI sarà appunto 300.000. Poiché le tracce sono disposte concentricamente, esse sono di lunghezza diversa e quindi non ogni traccia verrà scritta con la stessa densità. I costruttori indicano di solito quale sia la massima densità possibile di registrazione.

Moltiplicando quindi i due fattori (densità delle tracce e densità di registrazione per traccia) si ottiene la densità areale. Nell'esempio considerato, con i valori di 18.333 tracce per pollice e 300.000 BPI, si ottiene una densità areale di 5.500.000.000 bit per pollice quadrato, ossia 5.5 Gbits/in². I dischi più recenti hanno valori superiori di circa un ordine di grandezza, mentre i primi HDD dei Personal Computer avevano densità areali molto basse, dell'ordine di 0.004 Gbits/in² (Figura 1).

Sui dischi non è possibile accedere ad una singola informazione, occorre leggere o scrivere un intero settore. Per esempio, per modificare un singolo byte, è necessario leggere il settore che lo contiene, modificare il byte in

memoria e riscrivere il settore. I tempi di accesso all'informazione sono tipicamente dell'ordine di alcuni millisecondi. L'accesso ai dischi è in parte "diretto" (posizionamento della testina sulla traccia), in parte "sequenziale" (lettura/scrittura del settore).

Due parametri caratteristici dell'HDD sono, rispettivamente, il tempo di posizionamento (*seek time*) e il tempo di latenza (*latency*). Il primo è il tempo necessario perché la testina raggiunga la traccia su cui si trova l'informazione da leggere/scrivere, ed è proporzionale sia al numero di tracce attraversate sia alla velocità (meccanica) di spostamento. Il secondo è invece il tempo impiegato perché, nella rotazione del disco, il settore cercato giunga in corrispondenza della testina di lettura ed è inversamente proporzionale alla velocità di rotazione del disco.

Il cosiddetto "tempo di accesso" è la somma del tempo di posizionamento e della latenza.

5. L'EVOLUZIONE DEGLI HDD

Se il numero di transistor disponibili per unità di area raddoppia, secondo la legge di Moore, circa ogni due anni, il numero di bit memorizzati per unità di area HDD raddoppia circa ogni anno.

La densità di registrazione dei bit degli HDD è

così migliorata, in pochi decenni, di ben sette ordini di grandezza e ciò ha consentito di ridurre fortemente anche lo spazio complessivo occupato dagli HDD.

Ai dispositivi HDD sono stati poi aggiunti nel tempo, ulteriori strati sia di hardware che di software.

I primi HDD erano direttamente controllati dal computer cui erano connessi; ma già a partire dal 1965 apparvero nuovi dispositivi, denominati unità di controllo, che si interponavano tra il computer e gli HDD e offrivano funzionalità più evolute.

Il vantaggio principale di tali unità era costituito dal fatto che i comandi di lettura/scrittura provenienti dal computer venivano gestiti dalle unità di controllo così che gli HDD fossero in grado di operare in modo asincrono rispetto alle attività dell'unità centrale. Quest'ultima poteva perciò eseguire altre istruzioni (di altri programmi), dando luogo al meccanismo della multiprogrammazione che, a sua volta, avrebbe richiesto più raffinati sistemi operativi.

Nel corso del tempo cresceva l'enfasi sulla disponibilità dei dati e le unità di controllo cominciavano anche a supportare percorsi multipli per lo scambio di comandi, interruzioni e dati con l'unità centrale.

Poco alla volta, l'insieme degli HDD di un computer è diventato un'unica entità logico-fisica, ossia un vero e proprio sottosistema di memoria organizzato non solo per memorizzare i dati ma anche per garantirne l'affidabilità, ossia la continua disponibilità e la gestibilità.

Il costo unitario degli HDD è diminuito dal 1980 di cinque ordini di grandezza, mentre il costo dell'intero sottosistema di memoria durante lo stesso periodo è diminuito di 2,5 ordini di grandezza.

La più rapida caduta del costo degli HDD indica che il prezzo dei soli HDD è diventato nel tempo una frazione progressivamente meno importante nell'intero costo del sottosistema di memoria.

La riduzione di costo dei sottosistemi di memoria digitale ha fatto sì che dal 1996 la memorizzazione digitale sia diventata meno costosa della memorizzazione su carta, e dal 1998 anche più economica della tecnologia utilizzata in radiologia per memorizzare su film immagini mediche (Figura 2).

Dal 1990 la tecnologia di memoria ha supera-

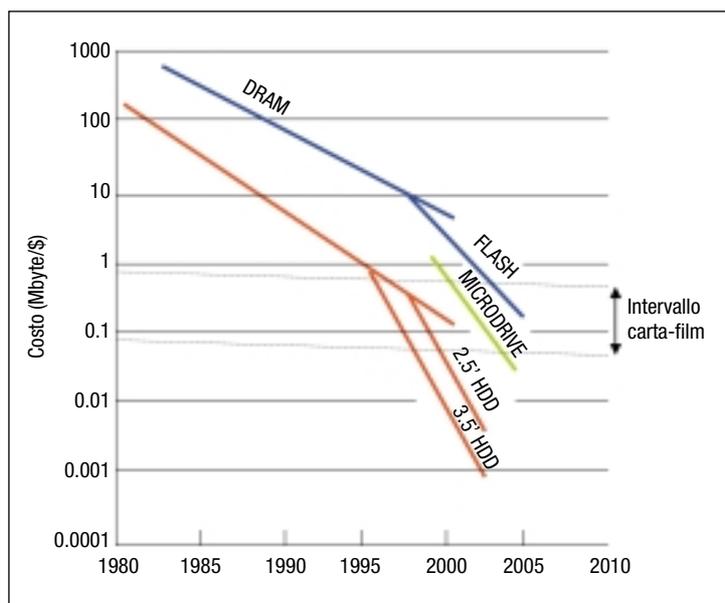


FIGURA 2

La diminuzione del costo degli HDD ha reso questa tecnologia più economica della registrazione su carta e su film

to in termini di costo/prestazioni sia le tecnologie di processore sia quelle di comunicazione (Figura 3). La disponibilità di simili tecnologie di memorizzazione ha quindi fortemente influenzato le architetture informative che vediamo oggi.

Si è così assistito al diffondersi di molteplici copie degli stessi dati, per esempio con memorie di transito, ossia *cache* di vario livello, così come al diffondersi di sistemi di memorizzazione vicini all'utente per evitare i ritardi di rete.

A partire dal 1990 la disponibilità di HDD di minori dimensioni (5,25 pollici e quindi 3,5 pollici) ha favorito una nuova filosofia organizzativa nei sottosistemi di memoria: utilizzare molteplici "piccoli" HDD invece di un minor numero di HDD più massicci.

La tecnologia RAID (*Redundant Array of Independent Disks*), che è alla base di tale evoluzione, permette di effettuare in parallelo diverse operazioni verso molteplici dispositivi HDD, consentendo di ottenere più elevati livelli di affidabilità e di prestazioni rispetto a quelle offerte da una singola unità HDD più grande.

Detto in termini molto semplificati, si può immaginare come gruppi di informazioni possano essere distribuiti in parallelo su veri e propri grappoli di piccoli HDD. L'indisponibilità di uno di questi può essere superata attraverso tecniche di ridondanza che consentono di ricostruire abbastanza facilmente la parte di informazione danneggiata. La tecnologia RAID è stata inizialmente utilizzata in sistemi di minor costo, ma in seguito ha cominciato ad essere utilizzata anche su sistemi di grandi dimensioni, proprio per le intrinseche funzioni di ridondanza (non c'era un "single point of failure") e di possibilità di duplicazione dei dati (copie di backup), funzionalità che consentivano di aumentare notevolmente la disponibilità complessiva delle informazioni.

Il valore per l'utente del sottosistema di memoria si è così progressivamente evoluto verso funzionalità più avanzate e verso l'integrazione di tali funzionalità nel sistema stesso. Attualmente, il costo della gestione del sottosistema di memoria è la parte maggiore del costo complessivo del sistema stesso. Ciò significa che il valore reale per l'utente del sottosistema di memoria risiede nella capacità di quest'ultimo di estendere le sue funziona-

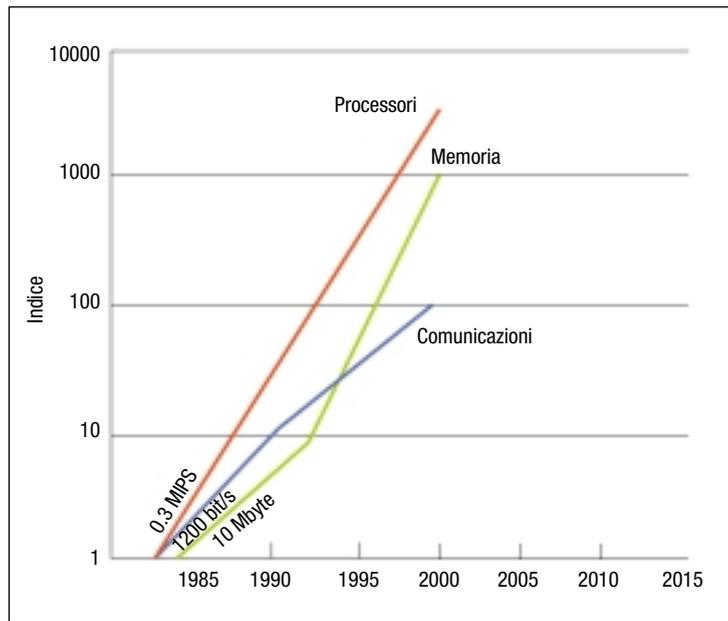


FIGURA 3

Miglioramento del rapporto prezzo-prestazioni delle tecnologie di microprocessore, registrazione e comunicazione

lità oltre a quanto possa fornire il solo dispositivo HDD, e specificamente nella capacità di ridurre i costi di gestione e di fornire una più elevata disponibilità dei dati (per esempio con copie di backup e servizi di replica).

6. LA DISPONIBILITÀ DEI DATI (DATA AVAILABILITY)

Il "disaster recovery", ossia il ripristino delle informazioni digitali a fronte di malfunzionamenti di varia natura, è diventato un requisito fondamentale per tutti i sistemi informatici e ha fortemente influenzato l'architettura dei sottosistemi di memoria. Negli attuali hard disk l'interfaccia testina di lettura-pista da leggere è il punto nel quale più spesso si presentano malfunzionamenti dovuti ad eventi come instabilità termo-meccaniche, urti, danneggiamenti del drive ecc. ed il parametro più comunemente utilizzato per indicare la disponibilità del drive tra due malfunzionamenti è il MTBF (*Mean Time Between Failures*). Intorno al 1990 il MTBF degli HDD era di circa 200.000 h, ma a seguito dei continui miglioramenti tecnologici il valore attuale è pari a circa 1.200.000 h. Tuttavia, il MTBF, considerato a sè, si presta a fraintendimenti, infatti, per essere interpretato correttamente il MTBF deve essere utilizzato

insieme ad un altro parametro, la cosiddetta "useful service life", ossia la vita media utile, che è il tempo tipico superato il quale possono presentarsi guasti dovuti al consumo dei componenti elettro-meccanici. I costruttori di HDD dichiarano attualmente una vita media utile di circa cinque anni. Il MTBF si applica solo a grandi quantità di dispositivi, mentre non dice nulla riguardo alla singola unità. Se per esempio, il MTBF di uno specifico modello di HDD è pari a 500.000 h e la vita media utile è di cinque anni (un anno è equivalente a 8.760 h), purché ogni cinque anni quell'unità venga sostituita essa potrà essere utilizzata per 57 anni, in media, prima che presenti un guasto ($500.000 : 8.760 = 57$).

Per proteggere i dati sono quindi necessarie ulteriori possibilità, che possono essere costituite da soluzioni *ad hoc* che consentano di creare manualmente occasionali copie di salvataggio su nastro, ma anche da vere e proprie tecnologie hw/sw che creino automaticamente le copie di salvataggio dei dati. Possono essere eseguite ciclicamente copie "virtuali" dei dati che verranno poi effettivamente trascritte su supporti fisici se reputato necessario, inoltre può anche essere utilizzata una tecnica di copiatura continua (*mirroring*), con la quale vengono costantemente mantenute due copie degli stessi dati, una primaria (locale) e una secondaria (in un altro sito). Quest'ultima metodologia è in generale sincrona, ossia la copia del sistema secondario deve essere scritta prima che l'operazione sul sistema locale venga considerata conclusa. Benché le operazioni sincrone siano desiderabili dal punto di vista dell'affidabilità dei dati, esse hanno tuttavia validità pratica solo nell'ambito di opportune distanze (dell'ordine del centinaio di km) tra il centro primario e quello secondario.

La metodologia di backup più comune è stata per molto tempo quella di effettuare una copia su nastro, o nel caso di personal computer su CD, dei dati ritenuti importanti.

Tuttavia, quando le memorie periferiche (soprattutto gli HDD) hanno iniziato ad essere interconnesse in sistemi di backup basati su LAN, questi ultimi hanno progressivamente rimpiazzato le metodologie manuali e sempre più spesso hanno consentito di eseguire copie automatiche di backup da HDD ad HDD.

I sistemi di backup non sono, in generale, sistemi semplici perché devono trattare molti tipi diversi di dati, a loro volta di differente importanza ed utilizzati da un'ampia varietà di utenti. Spesso, poi, i livelli di sicurezza necessari per tali dati possono essere al di là delle intrinseche possibilità tecnologiche dei dispositivi sui quali vengono memorizzati. Benché il puro costo della memoria HDD stia diminuendo, gli altri tipi di dispositivi di memorizzazione, come nastri e dischi ottici, rimangono importanti e occorre quindi un sistema che possa intelligentemente gestire i vari livelli tecnologici di memoria.

7. DALLE LAN ALLE SAN

L'emergere delle tecnologie LAN (*Local Area Network*), a basso costo, ha determinato la tendenza più significativa dei sottosistemi di memoria a partire dal 1988. In molti uffici i PC sono stati interconnessi in rete ed è emerso il modello *client-server*.

Tuttavia, mentre molte applicazioni sono migrate dai centri di elaborazione ai PC, i dati hanno spesso mantenuto rilevanza aziendale, piuttosto che individuale, e hanno quindi avuto bisogno di essere protetti e condivisi tra gli utenti.

Il PC era notoriamente inaffidabile e non gestito come i grandi computer, quindi, per ottenere la condivisione dei dati tra utenti diversi occorre la disponibilità di sistemi di memorizzazione locale (per esempio in un ufficio) condivisibili, a loro volta, da molteplici PC. Questi sistemi di memorizzazione erano più importanti (*mission critical*) dei singoli PC client, ed era naturale quindi che diventassero i primi candidati per tecnologie come *file-serving*, RAID, e sistemi di backup basati su LAN.

Il software utilizzato per l'interconnessione (*networking*) dei PC era spesso il Novell NetWare, così come altri software di PC vendor. Allo stesso tempo si affermava anche lo UNIX, sia nelle cosiddette workstation, che erano sostanzialmente dei PC molto più potenti, sia nei server.

L'ampia disponibilità di protocolli di condivisione dei dati come l'*NFS (Network File System)* dava un ulteriore impulso all'utilizzo di simili server.

Il passo successivo è stato l'avvento dei siste-

mi NAS (*Network-Attached Storage*) che offrivano le capacità di Network File System in un'unica macchina, la quale poteva gestire vari protocolli software come il CIFS (*Common Internet File System*), l'HTTP (*HyperText Transfer Protocol*), e l'FTP (*File Transfer Protocol*). I sistemi NAS erano semplici da installare perché venivano già forniti con funzioni di utilità e funzioni di gestione.

Allo stesso tempo, i responsabili dei sistemi informativi cercavano di "riconquistare il controllo" del patrimonio informatico distribuito tra i vari uffici e dipartimenti. Le funzioni di gestione dei dati venivano perciò progressivamente rimosse dai sistemi client e venivano create funzionalità globali di gestione per "mapping, sharing, and free-space management" di tutti i dispositivi di memoria connessi.

Il centro elaborazione dati ha così progressivamente riacquisito grande importanza e si è compreso che sarebbe stato molto utile se i molteplici server che ne facevano parte avessero potuto accedere ad un sottosistema di memoria complessivo, e non ciascuno a gruppi separati di dispositivi periferici di memoria.

Sono quindi nate le SAN (*Storage Area Network*)

che, potendo essere progettate e sviluppate indipendentemente dai server, permettevano la condivisione (*pooling*) delle risorse di memoria (configurate staticamente) e offrivano una maggiore efficienza e minori rischi per gli utenti. Gli investimenti nelle SAN non dipendevano da un particolare server o da una particolare scelta di software. La tecnologia SAN ha introdotto nuove possibilità nella semplificazione delle connessioni, nella capacità di crescita (*scalability*), nei costi e nelle tecniche di gestione delle memorie di massa. A sua volta, la fibra ottica è diventata la tecnologia predominante di interconnessione. Ragionando in termini essenzialmente intuitivi, si possono distinguere le NAS dalle SAN tenendo presente che le prime lavorano, quasi tradizionalmente, con file e protocolli di accesso ai file (CIFS, HTTP, FTP), mentre le seconde utilizzano propri blocchi di dati che possono essere memorizzati, ove necessario, nella gerarchia dei dispositivi di cui la SAN è costituita (Figura 4 e Figura 5).

La progressiva diffusione delle LAN Ethernet e del TCP/IP ha anche influenzato l'evoluzione delle SAN. Infatti, l'unificazione del networking con il TCP/IP e con Ethernet ha consentito una migliore standardizzazione

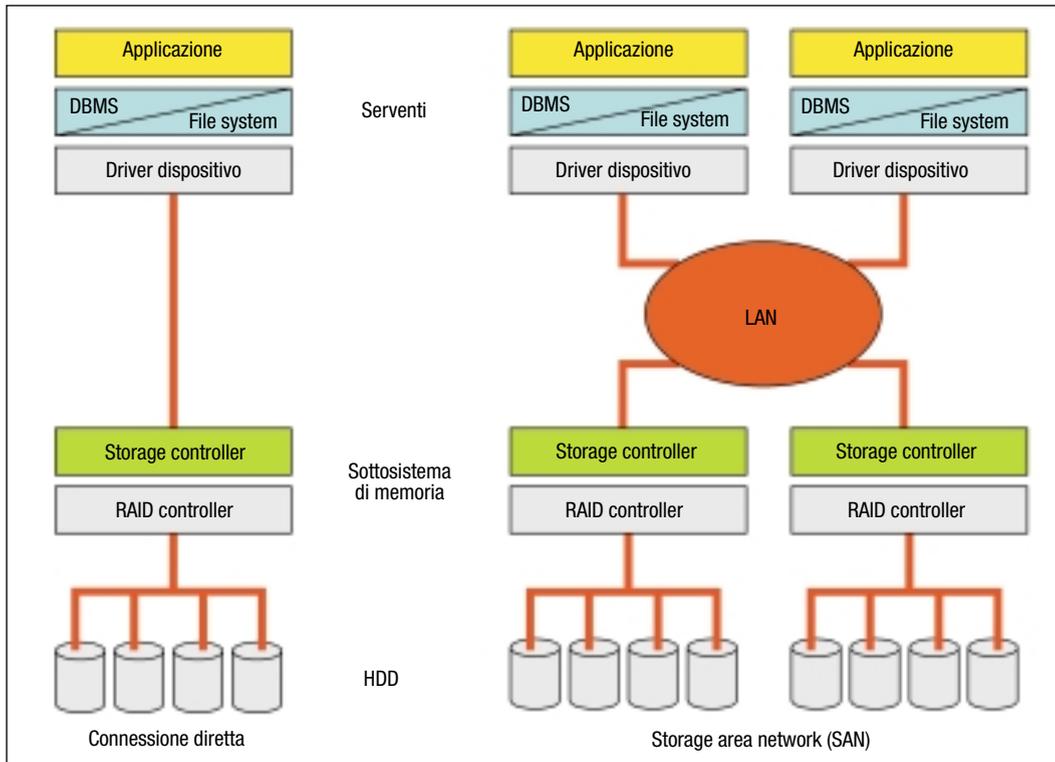
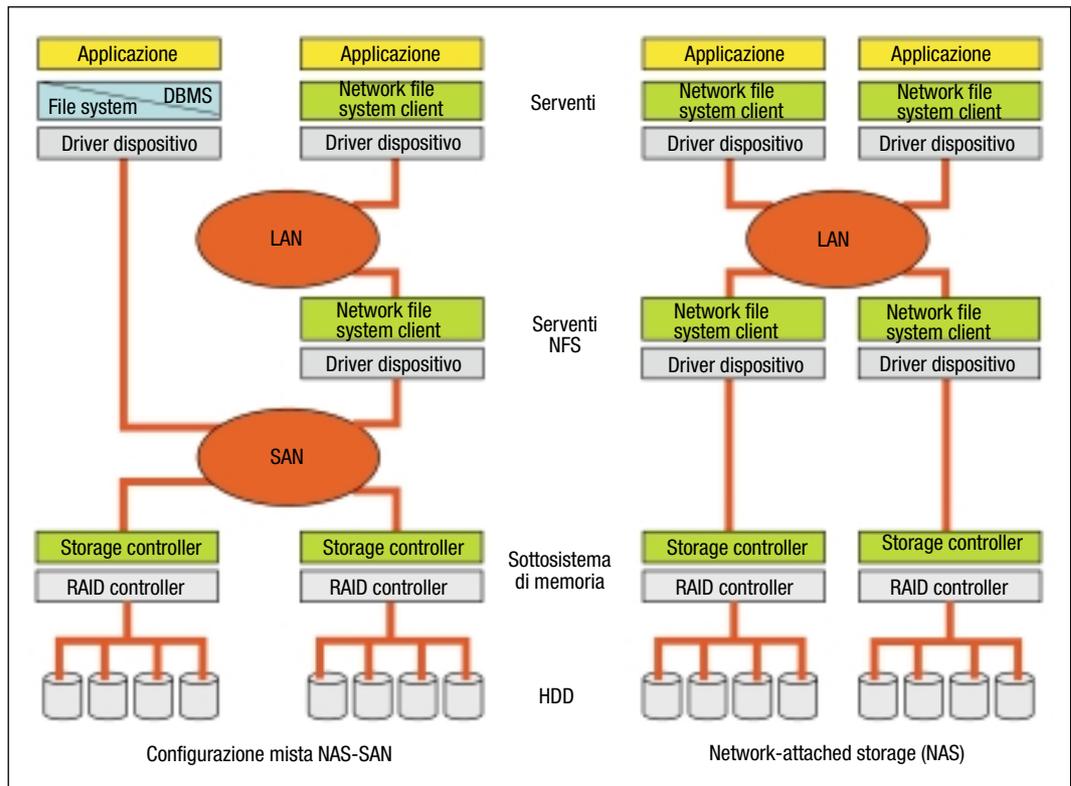


FIGURA 4
Trasmissione dei dati tra computer e HDD nel modello ad accesso diretto e in quello SAN

FIGURA 5
Trasmissione dei dati tra computer e HDD nel modello NAS e in quello misto SAN-NAS



dei sottosistemi di memoria, un crescente ritmo di innovazione e più bassi costi sia nell'hardware sia nel software sia, infine, nel supporto ai dispositivi di memoria.

È stato così possibile ridurre i costi di gestione, poiché gli operatori devono conoscere un solo tipo di rete. Inoltre, nell'ambito dei computer più piccoli, lo standard iSCSI (*Internet Small Computer System Interface*) è stato introdotto per consentire al protocollo SCSI di essere trasportato attraverso la rete TCP/IP. L'architettura delle SAN, che prevede una sostanziale condivisione di dati e spazi disponibili tra computer diversi, è in un certo senso antitetica all'originale progetto dei computer della terza e della quarta generazione sui quali si sono evoluti i maggiori sistemi operativi e i relativi sottosistemi di gestione dei dati.

L'obiettivo di progetto di tali computer prevedeva che questi fossero i soli ad accedere ai dati memorizzati sui dispositivi periferici e non erano stati progettati per condividere dati o spazio su disco con altri computer.

Le SAN hanno aggirato il problema togliendo ai computer il controllo esclusivo delle memorie di massa accessibili creando un'infra-

struttura hardware-software globale in grado di gestire il "mapping", lo "sharing" e la gestione dello spazio libero di memoria su tutti i dispositivi contemporaneamente connessi ai diversi computer del centro di elaborazione dati.

Oltre alla semplificazione nella configurazione e nella gestione dello spazio libero, nuove possibilità sono via via venute ad aggiungersi, come la condivisione dei dati, il cosiddetto "non disruptive change" e l'automazione di molte funzioni di gestione.

Poiché la gestione dei cosiddetti metadati è sotto il controllo di un "metadata server" nuove capacità vengono rese possibili e viene introdotto un più ampio spettro di servizi che semplifica la gestione e migliora la disponibilità dei dati. Per esempio politiche associate a specifici gruppi di dati determinano come questi ultimi debbano essere gestiti in termini di prestazioni, disponibilità, sicurezza e costi [3, 4].

C'è inoltre l'esigenza di estendere geograficamente l'utilizzo dell'architettura SAN. Soluzioni SAN per mezzo della fibra ottica possono ora estendersi per oltre 100 km per creare delle SAN "metropolitane" utilizzando tecno-

logie trasmissive come la DWDM (*Dense Wavelength Division Multiplexing*); ma è possibile anche un'estensione geografica più ampia che, per mezzo del protocollo IP e la fibra ottica, utilizzi la tecnologia WAN per creare una vera e propria "global storage network", come mostrato in maniera molto schematica nella figura 6.

Con la crescita delle SAN e la sempre più ampia diffusione del protocollo IP, i sistemi di memoria basati su reti IP, quali LAN, MAN e WAN, si sono ormai affermati come l'avanguardia dello sviluppo tecnico.

8. UNA ROADMAP TECNOLOGICA

Per i prossimi anni non è previsto alcun rallentamento nella velocità di evoluzione della tecnologia. I problemi attuali sembrano però di natura più complessa rispetto a quelli di qualche anno fa. Essi riguardano la termodinamica dell'energia memorizzata nei bit magnetici, le difficoltà connesse con la distanza testina-piatto, che ormai sono dell'ordine di grandezza di un diametro atomico, e l'intrinseca velocità di commutazione dei materiali magnetici. L'ulteriore miniaturizzazione delle aree di registrazione, ossia il cosiddetto *basic scaling*, è più o meno simile allo *scaling* di qualunque soluzione tridimensionale del campo magnetico (Figura 7). Se le proprietà magnetiche dei materiali restano costanti, la configurazione del campo e le grandezze correlate restano invariate, anche se tutte le dimensioni vengono ridotte di un fattore s , fintanto che anche la corrente venga ridotta di un fattore s .

Pertanto se si vuole incrementare la densità lineare (ossia i bit per traccia) di un fattore 2, e quindi la densità areale di un fattore 4, occorre fare lo *scaling* di tutte le dimensioni per due, lasciare intatta la velocità di rotazione e raddoppiare la velocità di trasferimento dei dati. Se i materiali mantengono le stesse proprietà nelle nuove dimensioni tutto funziona come prima.

Tuttavia, ci sono motivi per i quali nel mondo reale questo semplice *scaling* non funziona. Il primo è che aumentare la velocità di trasferimento dei dati in proporzione alla densità lineare può essere al di là delle possibilità costruttive.

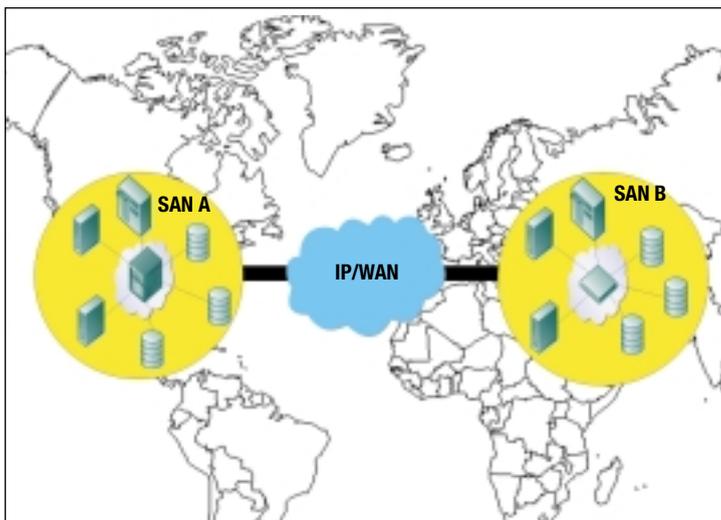


FIGURA 6

Il protocollo IP e la fibra ottica consentono alle SAN di infrangere la barriera della distanza

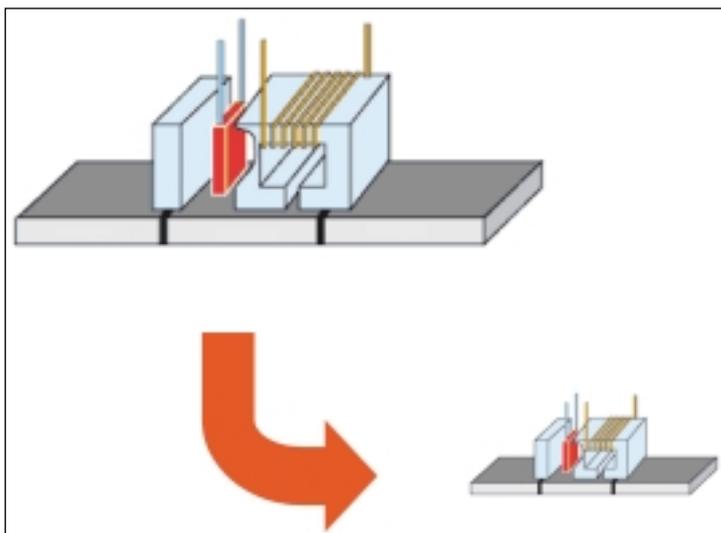


FIGURA 7

Le tre dimensioni in gioco nel processo di miniaturizzazione delle testine di registrazione degli HDD

La seconda ragione è che la forte competizione tra vari costruttori per creare HDD ad alte prestazioni ha costantemente prodotto un incremento di velocità nella rotazione dei dischi (per ridurre la latenza) e ciò ha finito con l'aggravare il problema perché la lettura di un maggior numero di dati a più alta velocità diventa molto più complessa.

C'è poi da considerare che un segnale induttivo di lettura (*readback*) diminuisce con lo *scaling* e il disturbo elettronico cresce con il traffico dei dati (*bandwidth*), tanto che il rap-

ventato, sempre più, l'insieme delle informazioni riguardanti i processi aziendali, siano esse applicazioni software o database di varia natura.

Il sottosistema di memoria è una parte del complessivo sistema informatico, ma svolge ormai un ruolo fondamentale affinché un'organizzazione possa rispondere rapidamente, flessibilmente e a costi molto più bassi alle nuove sfide: la maggior parte delle imprese investirà progressivamente più nella tecnologia "storage" che in quella "server".

Le memorie di massa, peraltro, non sono più un semplice aggregato di dispositivi hardware, ma sono diventate una complessiva entità multifunzionale nella quale molte importanti funzioni sono realizzate con la tecnologia software.

Dal punto di vista della pura tecnologia hardware, soprattutto per quanto riguarda i dispositivi HDD, cominciano a comparire sostanziali limitazioni fisiche che frenano la miniaturizzazione (*scaling*) della registrazione magnetica. È probabile quindi che il ritmo di aumento della densità areale, e quindi la riduzione del costo per bit, possa cominciare a rallentare nei prossimi dieci anni.

Ciononostante non si vedono ancora tecnologie che possano in breve tempo sostituire le attuali tecniche di registrazione magnetica, che continuano quindi ad occupare stabilmente una solida nicchia del mercato IT.

Al momento si prevede che al massimo si potrà ottenere una densificazione di due ordini di grandezza (x 100) rispetto all'attuale tecnologia di registrazione magnetica. Entro dieci anni, peraltro, ulteriori tecnologie potranno forse offrire una reale alternativa.

Molto promettenti sembrano le cosiddette *memorie olografiche*. I recenti progressi indicano, infatti, la possibilità di costruire memorie nelle quali un'intera pagina di informazioni può essere simultaneamente memorizzata attraverso meccanismi di interferenza ottica in materiali fotosensitivi, incrociando due raggi laser coerenti all'interno del materiale. In questo caso la grande capacità di memorizzazione discende dal fatto che la memoria olografica è di tipo volumetrico e non di superficie [6]. La luce proveniente da una fonte laser coerente viene divisa in due fasci (beam), che sono rispetti-

vamente quello di segnale, che trasferisce i dati, e quello di riferimento. Quando i due fasci di luce si intersecano sulla superficie del supporto si crea un ologramma, ossia un disegno formato da zone chiare e scure che possono quindi rappresentare i bit 0 e 1. L'ologramma è quindi costituito da una specie di reticolo ottico che, enormemente ingrandito, potrebbe sembrare la superficie di una scacchiera che può contenere ben più di 64 caselle. I dati sono quindi organizzati in tali "scacchiere" di bit. L'ologramma viene registrato, a seguito di una reazione chimica, in un supporto sensibile alla luce (come accade in una pellicola fotografica) nel punto dove si intersecano i fasci di riferimento e di segnale. Modificando l'angolo e la lunghezza d'onda del fascio di riferimento, o la posizione del supporto, diventa possibile registrare ologrammi differenti nello stesso volume. Se il materiale stesso viene colpito da uno dei due fasci che in precedenza aveva costruito l'ologramma si può ricostruire l'informazione originaria del secondo fascio e quindi leggere le informazioni memorizzate.

Alcuni costruttori come la InPhase Technologies e la Japan's Optware hanno annunciato un dettagliato piano di evoluzione per la commercializzazione di questa tecnologia. La Inphase Technologies ha recentemente dichiarato che nel 2006 sarà disponibile un disco da 300 Gbyte e nel 2009 da 1.6 Tbyte, mentre la Optware, che ha denominato HVD (*Holographic Versatile Disk*) il nuovo supporto, afferma che entro un anno saranno disponibili HVD da 100-200 Gbyte per arrivare in altri due o tre anni a 3.9 Tbyte. Supporti come gli HVD, poiché non sono in rotazione, non devono leggere sequenzialmente bit dopo bit e possono trasferire in parallelo fino a 10 kbyte, con un evidente grande miglioramento nelle prestazioni di diverse applicazioni, quali soprattutto quelle mediche, geospaziali, di business e multimediali.

Ma c'è un'ulteriore possibilità, ossia quella di utilizzare particolari polimeri per memorizzare in piccoli spazi enormi quantità di dati. I polimeri sono materiali organici flessibili e robusti (come per esempio i giubbotti antiproiettile) costituiti da enormi ca-

tene di molecole identiche. Nel 1970 si è scoperto che tali materiali da isolanti possono anche diventare conduttori e possono quindi essere utilizzati per applicazioni elettroniche. Negli ultimi anni sono state esplorate diverse soluzioni tecnologiche tra le quali la cosiddetta *Thin Film Memory* nella quale un substrato di un opportuno materiale (come il silicio) può essere rivestito con uno strato estremamente sottile (thin film) di polimero. Il thin film è inserito tra due gruppi ortogonali di microfilari (elettrodi) che si incrociano in molteplici punti. Ogni punto di intersezione tra due microfilari rappresenta una cella di memoria che contiene un bit. Differenti voltaggi applicati opportunamente ai fili possono consentire di scrivere o di leggere. Il meccanismo non è poi tanto diverso da quello di una memoria a transistor realizzata su silicio. Gli elettrodi vengono, per così dire, stampati direttamente sul film di polimero. Far crescere le dimensioni della memoria vuol quindi dire sovrapporre un nuovo strato sul precedente, e così via.

Ogni strato di polimero ha una sezione verticale di meno di un decimillesimo di millimetro ed è facile da depositare. Su di una scheda delle dimensioni di una carta di credito diventa quindi possibile costruire una memoria con la capacità di quasi mezzo milione di CD o cinquantamila DVD. In questo momento il maggiore sforzo industriale è proprio nel cercare di mantenere le linee di produzione di silicio, dal wafer al chip, per cercare di adattarle a questo ulteriore processo costruttivo in modo da ridurre gli investimenti necessari.

Negli ultimi anni si è così riusciti a costruire memorie a polimeri di tipo ROM, e ora si sta cercando di ampliare il processo di costruzione alle memorie di tipo RAM.

Un'ultima tecnologia molto complessa, che, di fatto, è una vera e propria nanotecnologia, la cosiddetta *probe-based storage technology*, si avvale di tecniche basate sul microscopio a forza atomica (AFM) e sul microscopio ad effetto tunnel (STM) per raggiungere densità di registrazione delle informazioni dell'ordine di 1 Tbit, ossia mille miliardi di bit, per pollice quadrato [7]; ma sembra ancora abbastanza lontana.

Si può infine osservare come Internet possa diventare esso stesso una gigantesca memoria utilizzabile da miliardi di utenti.

L'università di Berkeley ha avviato nel 2002 un progetto, *OceanStore*, per creare un sistema in grado di memorizzare in maniera organica le informazioni distribuite su Internet. Per mezzo di *OceanStore* i dati dovrebbero diventare accessibili a tutti, simultaneamente, e in qualunque momento. Inoltre i dati non dovrebbero mai andare perduti o distrutti e al tempo stesso restare inaccessibili a chi non è autorizzato. Il sistema organizza gli archivi e li fraziona in blocchi di dati che possono essere continuamente trasferiti e moltiplicati su altri nodi della rete. Un unico blocco può essere sufficiente per ricostruire l'intero archivio.

Di conseguenza se anche un nodo della rete cessa di funzionare, lo specifico archivio non va perduto. *OceanStore* memorizza quindi dati in maniera promiscua: ogni server può creare una replica locale di qualsivoglia aggregato di dati e queste copie locali consentono anche migliori prestazioni poiché riducono il traffico complessivo di rete localizzandolo dove necessario. *OceanStore* assegna a ciascun blocco un codice identificativo univoco, denominato *globally unique identifier*.

Quando un utente necessita di un determinato archivio chiede al proprio nodo di ritrovare l'insieme dei blocchi più vicini e di ricostruire l'archivio stesso. Quest'ultimo non è accessibile quindi che agli utenti che possiedono le chiavi giuste. Unitamente ad opportuni sistemi di crittografia questa tecnica dovrebbe garantire la sicurezza dei dati memorizzati su Internet.

Alcuni componenti di *OceanStore* sono già funzionanti mentre l'intero prototipo è ancora in fase di sviluppo; ma quando e se dovesse completarsi realizzerebbe forse quanto Borges aveva detto già nel 1941 nel suo racconto "La biblioteca di Babele":

"Quando si proclamò che la Biblioteca comprendeva tutti i libri, la prima impressione fu di straordinaria felicità. Tutti gli uomini si sentirono padroni di un tesoro intatto e segreto. Non v'era problema personale o mondiale la cui eloquente soluzione non esistesse: in un qualche esagono. L'universo era

giustificato, l'universo attingeva bruscamente le dimensioni illimitate delle speranze".

Bibliografia

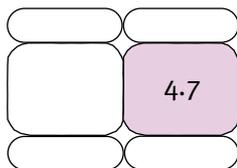
- [1] Lyman Peter, Varian Hal R.: *How Much Information 2003?* <http://www.sims.berkeley.edu/research/projects/how-much-info-2003/>
- [2] *A Brief History of the Hard Disk Drive.* <http://www.storagereview.com/guide2000/ref/hdd/hist.html>
- [3] Chirillo John, Blaul Scott, Chirillo John, Blaul Scott: *Storage Security: Protecting, SANs, NAS and DAS Wiley.* 1 edition, December 20, 2002.
- [4] Schulz Greg: *Resilient Storage Network: Designing Flexible Scalable Data Infrastructures.* Digital Press, April 7, 2004.
- [5] Tarnopolsky Giora J.: *Areal density growth.* www.thic.org/pdf/Mar03/tarnotek.gtarnopolsky.030304.pdf
- [6] Ashley J., Bernal M.-P., Burr G. W., Coufal H., Guenther H.: *Holographic data storage.* <http://www.research.ibm.com/journal/rd/443/ashley.html>
- [7] Eleftheriou E., et al: *A Nanotechnology-based Approach to Data Storage.* Proc. 29-th VLDB Conf., Berlin, Germany, Sep. 2003. http://www.zurich.ibm.com/pdf/sys/storage/ELE_VLDB2003.pdf

ERNESTO HOFMANN laureato in fisica presso l'Università di Roma, ha lavorato oltre 35 anni per la IBM in Italia, in Francia e negli USA, nei maggiori laboratori di sviluppo dei grandi computer. È autore di molteplici pubblicazioni sia di carattere tecnico sia divulgative, nonché di svariati articoli e interviste. È attualmente docente di "Economia e gestione delle imprese internazionali" presso l'Università Cattolica del Sacro Cuore di Milano. ernesto_hofmann@mac.com



ICT PER I BENI CULTURALI ESEMPI DI APPLICAZIONE

Paolo Paolini
Nicoletta Di Blas
Francesca Alonzo



Per ICT s'intende, comunemente, quell'insieme di tecnologie che provengono dallo sviluppo dell'informatica (intesa in senso ampio) e dalle reti di telecomunicazione, con Internet naturalmente al centro della scena. Lo scopo di questo articolo è mostrare come le ICT abbiano supportato e in qualche caso rivoluzionato (o stiano per rivoluzionare), il settore dei Beni Culturali. Come esempi si utilizzeranno alcuni progetti che il laboratorio HOC¹ del Politecnico di Milano, da tempo attivo nel settore, ha realizzato nel corso di un decennio.

1. INTRODUZIONE

All'interno di questo articolo verranno considerati i "beni culturali" in un'accezione piuttosto ampia (che gli Anglosassoni chiamano più convenientemente *Cultural Heritage*). Per "beni culturali" s'intendono, infatti, l'insieme di artefatti in senso lato (oggetti, pitture, sculture, edifici) ma anche tradizioni, costumi, che in qualche modo costituiscono il nostro "patrimonio culturale", tramandatoci da coloro che ci hanno preceduto. Concretamente, per quanto riguarda l'Italia, possiamo dire che si tratta di parchi archeologici, musei, oggetti di valore storico-artistico, edifici rilevanti, monumenti, chiese, quadri, mobili, arredi, oggetti di culto, tessuti, ed anche città, strutture urbanistiche, tradizioni popolari, tra-

dizioni alimentari (oggi di gran moda), tradizioni culinarie, tradizioni religiose ecc.. Come premessa generale diciamo che non ci occuperemo di arte e cultura contemporanee, in rapporto alle tecnologie: ci porterebbero su un terreno molto diverso da quello "classico", che considera la cultura tradizionale.

La prima cosa che vogliamo proporre alla riflessione del lettore è che il termine "culturale", applicato a quanto si produce oggi, non si riferisce alla specifica natura fisica dell'oggetto, né al suo valore "estetico" (gli oggetti d'arte, in senso stretto, sono solo una frazione dell'immenso patrimonio culturale italiano). Per "culturale" s'intende che l'oggetto d'interesse (sia esso un edificio, un vaso greco, o una ricetta di cucina):

- ha una storia (nota o meno);
- questa storia ci viene da lontano (il tempo lo ha in qualche modo "nobilitato");
- questa storia si correla in qualche modo con noi, alla società odierna, e ha quindi un qualche valore (questo per non far coincidere il concetto di bene culturale con qualsiasi cosa che sia semplicemente molto vecchia);

¹ HOC, *Hypermedia Open Center* (<http://hoc.elet.polimi.it>) è un laboratorio multidisciplinare, creato all'interno del DEI (Dipartimento di Elettronica e Informazione) del Politecnico di Milano. HOC è specializzato nello sviluppo di progetti e metodologie per applicazioni multimediali, interattive, basate sul Web e altri canali.

Esiste una forma di consenso (sebbene non necessariamente universale) sul valore dell'oggetto.

Come si distingue un sasso qualsiasi (di nessun interesse) dal sasso (di grande valore culturale) con cui Davide uccise Golia? Innanzitutto, perché c'è una storia legata a quel sasso; in secondo luogo, perché la storia del giovinetto che sconfigge un gigante è intrigante di per sé; in terzo luogo, perché è una storia di molti secoli fa²; infine, perché, sia attraverso la tradizione religiosa sia attraverso "l'immaginario collettivo", questa storia è sentita ancora come significativa per noi e per i nostri valori nella società attuale. Quindi un sasso qualsiasi non c'interessa, come bene culturale, mentre il sasso di David (se fosse rintracciato) sarebbe bene accetto in qualsiasi museo del mondo.

Con una sintesi estremamente ardita possiamo, quindi, elencare i principali usi dell'ICT per i beni culturali. La classificazione che usiamo nel seguito si basa sulla "destinazione d'uso", piuttosto che sulle tecnologie utilizzate. Per questo motivo alcune tecnologie verranno citate più volte in punti diversi, in quanto utilizzate per scopi differenti.

In relazione alla finalità, le principali applicazioni delle tecnologie ICT al settore dei Beni Culturali possono essere classificate come segue:

- **Gestione:** applicazioni che consentono una migliore (più efficace e a minor costo) gestione del patrimonio culturale.
- **Studio e ricerca:** applicazioni che favoriscono le attività di studio (a livello universitario e superiore) e di ricerca inerenti ai beni culturali.
- **Diagnosi:** applicazioni che consentono (di aiutare) la diagnostica dello stato di conservazione (o degrado) dei beni culturali.
- **Restauro:** applicazioni che consentono (di aiutare) le attività di restauro di beni culturali.

□ **Tutela:** applicazioni che consentono di tutelare il patrimonio, sia contro atti criminosi sia rispetto a calamità naturali.

□ **Comunicazione-divulgazione:** applicazioni che consentono di "comunicare" al grande pubblico la rilevanza del nostro patrimonio culturale, sia a scopo divulgativo generale che per favorire il turismo.

□ **Formazione:** applicazioni rivolte ai giovani studenti (scuola media superiore o livello universitario) per aiutarli negli studi.

□ **Fruizione:** applicazioni che consentono di fruire in modo più efficace del patrimonio culturale.

La classificazione sopra indicata non ha alcuna pretesa di scientificità, ma solo lo scopo di organizzare l'esposizione dei contenuti di questo articolo. È evidente, inoltre, che diverse attività sono tra loro correlate: un buon inventario di beni culturali di una regione (utile per la gestione), per esempio, può diventare la base per un'attività di tutela. Tuttavia, bisogna dire molto chiaramente che non esistono nessi causali tra le varie applicazioni ed è, quindi, falso che un'applicazione ne implichi automaticamente un'altra (come spesso si tende a far credere in Italia): l'inventario, per restare all'esempio di cui sopra, non assicura di per sé una efficace tutela del patrimonio e, a volte, nemmeno la favorisce³.

2. GESTIONE DEI BENI E DEL PATRIMONIO CULTURALE⁴

Ogni soggetto che si trovi ad essere responsabile di beni culturali deve preoccuparsi della loro "gestione": termine generico che vuol dire conoscere quali sono i vari beni, le loro caratteristiche generali, la loro collocazione, il loro stato di manutenzione ecc. I soggetti che necessitano di queste applicazioni sono

² Un lettore malizioso potrebbe domandarsi se gli oggetti de "Il Grande Fratello" fanno parte del nostro Patrimonio Culturale. Certamente hanno una storia che li caratterizza e una forte relazione con la società odierna. L'unico "difetto", che ci consente di astenerci dal metterli sotto tutela come oggetti culturali, è che non ci provengono dal passato. Ma tra qualche secolo, chissà? Questa possibilità è meno paradossale di quanto sembri: agli inizi degli anni '60 i film di Totò e di Franco Franchi e Ciccio Ingrassia erano considerati dalla cultura ufficiale espressioni deteriori del gusto popolare, oggi sono considerati parte della cultura italiana e circolano nei cinema d'essai.

³ Si è verificato più volte, infatti, che l'elenco dei beni per un'attività (la gestione del patrimonio, ad esempio) non fosse affatto utile per un'altra attività (la tutela del patrimonio, per esempio). Per questa ragione ogni attività tende a fare il proprio "inventario", sia per mancanza di coordinamento, che per diverse esigenze sulle informazioni necessarie e sul modo di organizzarle.

⁴ Queste applicazioni, per i musei, vengono chiamate nel mondo anglosassone *collection management*.

le sovrintendenze⁵, le regioni, le province, i musei, le diocesi, le banche e tutti i privati proprietari di beni culturali rilevanti.

Sostanzialmente, si tratta di applicazioni di Basi di Dati: tutti gli immobili e gli oggetti considerati devono essere “inventariati” in modo opportuno e queste schede di inventario devono essere organizzate in modo utile. Questa operazione, mutuata dalle applicazioni aziendali, risulta (per i beni culturali) in realtà spesso più complessa di quanto sembri: immaginiamo un parco archeologico, per esempio. Quali sono gli oggetti da inventariare? Gli edifici? Le singole pareti, qualora ospitino un affresco particolarmente significativo? Gli oggetti vanno presi singolarmente o vanno considerati i loro raggruppamenti? Si pensi, per esempio, ai diversi oggetti che compongono un corredo funerario! Bisogna tener presente il fatto che in un parco archeologico tipico possono trovarsi diverse migliaia o decine di migliaia di “oggetti”, da un tempio ad un frammento d’anfora. Analogo ragionamento può essere fatto per le chiese, che contengono miriadi di “oggetti” di potenziale interesse. Se l’inventario riguarda un museo d’arte, è abbastanza facile identificare gli oggetti da inventariare; se, invece, si tratta di altre realtà, l’operazione risulta evidentemente assai più complessa.

In ogni caso la domanda fondamentale è: a cosa serve veramente l’inventario dei beni culturali, vale a dire la cosiddetta “catalogazione”? Un obiettivo gestionale “classico” è quello di poter controllare “il magazzino”: sapere quello che “ci dovrebbe essere” e avere la possibilità di verificarne la presenza effettiva.

Un secondo obiettivo è quello di fungere da supporto all’attività di studio e ricerca: lo studioso, grazie all’inventario, può sapere “cosa c’è e dove si trova”.

Un terzo obiettivo è quello di supporto alle attività di tutela: l’inventario consente di pianificare le azioni di tutela, o di sapere cosa manca, per esempio in caso di furto o alluvione.

Un quarto obiettivo, recentemente venuto in auge, è quello di poter dare un “valore” al patrimonio culturale, quasi fosse – ancora – il contenuto di un classico magazzino aziendale. Dando un potenziale valore economico a ciascun oggetto inventariato, si può arrivare a stimare il valore potenziale del nostro patrimonio culturale⁶!

Se l’analogia con la gestione aziendale sembra facile (tutte le aziende, con un minimo di organizzazione interna, fanno l’inventario del loro magazzino), la realtà del settore dei beni culturali è ben diversa: la maggior parte del patrimonio culturale italiano non è inventariato, o inventariato male, con modelli di rappresentazione dei dati antiquati. A questo si deve aggiungere che gli inventari esistenti non vengono quasi mai usati, nella realtà quotidiana: sono considerati necessari adempimenti burocratici, ma di scarsa se non nulla utilità operativa. La cosiddetta valorizzazione economica poi, lodevole nelle intenzioni, ha incontrato tali difficoltà oggettive (per esempio, con quali criteri attribuire dei valori), tanto da restare, di fatto, lettera morta.

C’è stato poi, negli ultimi decenni, il tentativo di standardizzare le informazioni dell’inventario: il MiBAC, attraverso i suoi organi preposti, ha definito quali devono essere i dati dell’inventario (dati diversi a seconda del tipo di bene). Il problema è che alle difficoltà oggettive dell’operazione si sono aggiunte difficoltà di altra natura: una impostazione burocratica e centralistica, la confusione tra inventario e altre applicazioni (scientifiche e di tutela, per esempio) ecc.. Pertanto, l’obiettivo di avere un inventario “ragionevole” del Patrimonio Culturale Italiano pubblico è ancora lontano, malgrado l’ingente sforzo economico degli ultimi decenni. I milioni di “schede di catalogazione” prodotte, e tuttora in produzione, al costo di decine di milioni di Euro, giacciono tristemente in archivi che non vengono consultati da nessuno. Le schede contengono dati eterogenei e, in genere, poco utili a tutti. L’affidabilità

⁵ Per i lettori meno esperti sull’argomento si ricorda che sul territorio nazionale il Ministero dei Beni e Attività Culturali, MiBAC, è articolato in “sovrintendenze”, secondo criteri sia geografici (regioni, province, città...) sia di settore (per esempio, archeologia, architettura...).

⁶ Verrebbe da osservare che, con stretta analogia con i magazzini aziendali e il loro uso nei bilanci, è possibile “giocare” sui valori attribuiti, modificando (verso il basso o l’alto, a seconda delle esigenze) il valore del patrimonio nazionale.

dei contenuti delle schede è spesso dubbia, se non quasi nulla⁷. La conclusione è che sarebbe meglio, prima di proseguire nell'opera e investire massicciamente le – già scarse – risorse disponibili in nuovi sforzi di catalogazione, rivedere seriamente i requisiti e gli obiettivi della stessa, reimpostarla adeguatamente e solo allora, eventualmente, riprendere.

Un cenno merita un argomento di cui oggi molto si parla: la “geo-referenziazione”, vale a dire, l'uso di tecniche moderne per identificare con precisione le coordinate sulla crosta terrestre del patrimonio culturale. Se alcune attività, come per esempio la tutela e lo studio/ricerca, si possono avvantaggiare della geo-referenziazione (per altro molto costosa), il pensare che questa sia la soluzione della sostanziale mancanza di un inventario usabile non è un'ipotesi credibile; sarebbe, forse, più auspicabile una riflessione seria su come gestire meglio il patrimonio, piuttosto che sperare in una nuova tecnologia come soluzione di tutti i mali (soprattutto di quelli di origine organizzativa).

3. STUDIO E RICERCA

In questa categoria possiamo annoverare varie applicazioni:

- **Catalogazione:** applicazioni di basi di dati, estensione della inventariazione, con dati utili per gli specialisti del settore e le loro attività di ricerca.
- **Cartografia:** applicazioni di mappe cartografiche usate per studiare il territorio.
- **Analisi:** varie applicazioni che consentono di studiare in profondità edifici e oggetti, ad esempio per verificare le tecniche utilizzate, i materiali, lo stato di conservazione ecc..
- **Ricostruzioni virtuali:** ricostruzioni di edifici o ambienti non più esistenti; queste applicazioni sono di interesse anche per la comunicazione.

□ **Altro:** altre applicazioni descritte in seguito.

3.1. Catalogazione

Ai dati necessari per inventariare (da un punto di vista amministrativo) un bene culturale, bisogna aggiungere ulteriori dati utili per le attività di studio e ricerca. Quali siano veramente i dati utili non è certo chiaro: come già detto, dietro l'impulso del Ministero dei Beni Culturali, si sono create milioni di “schede di catalogazione”. Queste schede sono spesso farraginose (in base agli standard emanati), di difficile o addirittura impossibile consultazione e gestite con tecnologie svariate, alcune delle quali, oggi, obsolete. Inoltre, per giustificare lo sforzo di catalogazione, si è spesso creata confusione con l'inventariazione: non è chiaro se lo scopo sia creare un inventario oppure offrire uno strumento di lavoro ai ricercatori. Fonte di confusione è la struttura stessa delle schede “ufficiali” di catalogazione⁸, che mescolano vari tipi di dati (per esempio, di inventario, di descrizione fisica, di valutazione critica, di interpretazione ecc.) e hanno una struttura spesso inadeguata allo scopo⁹. Una ulteriore fonte di confusione è la pretesa implicita che la catalogazione sia la base “sicuramente necessaria e forse anche sufficiente” per l'attività di comunicazione e valorizzazione. La realtà è ben diversa: lo sforzo di catalogazione fin qui messo in atto non ha prodotto quasi nessun risultato apprezzabile ai fini della comunicazione e della valorizzazione. I milioni di schede create sono difficilmente accessibili alla comunità di studiosi per i quali, in teoria, sono state create. I contenuti delle schede sono spesso dettagli di dubbia utilità; le schede sono state create principalmente da personale giovane e inesperto, con risultati qualitativamente scarsi; la natura essenzialmen-

⁷ Molte delle catalogazioni sono state realizzate da precari poco addestrati, lavoratori socialmente utili e simili; pochissime catalogazioni sono state seriamente revisionate e controllate, per verificare la qualità e attendibilità dei contenuti.

⁸ Dovute all'attività dell'ICCD (Istituto Centrale del Catalogo e della Documentazione) organo del ministero che ha definito i cosiddetti “standard di catalogazione” per vari tipi di Beni Culturali.

⁹ Un problema gestito molto male è, per esempio, la modellazione della situazione in cui ci sia un aggregato (per esempio, un corredo funerario) e dei singoli oggetti contenuti in esso: le attuali schede obbligano a duplicare una serie di dati. Un altro caso gestito male dalle schede (di natura gerarchica) sono i riferimenti incrociati tra beni (per esempio, tra una tomba e il relativo corredo funerario).

te testuale delle schede, con poche immagini, le rende ancora meno utilizzabili.

3.2. Cartografia

La cartografia dovrebbe essere di grande aiuto, soprattutto per l'archeologia o per indagini che avessero bisogno di analizzare il territorio (per esempio, nel rilevare i tipi di attività umane che hanno "antropizzato" in qualche modo il territorio, tracce di antiche viabilità ecc.). Per le attività scientifiche (non legate cioè alla tutela e alla prevenzione di disastri naturali) le banche dati cartografiche sono ancora poco utilizzate nel settore dei Beni Culturali. Per esempio, esiste da diversi anni una tecnologia innovativa (quella legata alla cartografia satellitare) sia su spettro visibile che su spettro diverso da quello visibile. In teoria, questa tecnologia dovrebbe essere di grande interesse per l'archeologia: è possibile, infatti, rilevare, con precisione fino a un metro, fenomeni del terreno (quali la presenza di rovine sepolte, per esempio) non facilmente percepibili nello spettro visivo. Le tecnologie cartografiche, con la sola eccezione della mappatura delle zone a rischio idrogeologico, sono ad oggi molto poco usate. Ciononostante la georeferenziazione (sopra descritta) è considerata da molti uno sviluppo di grande interesse, spesso in modo poco "fideistico".

Le ICT di questo settore sono di tipo specialistico, riconducibili alle metodiche dei GIS (*Geographical Information System*). I GIS sono molto sviluppati nel settore amministrativo, soprattutto su scala regionale, ma tuttora poco usati per i Beni Culturali, per una varietà di motivi (sia culturali che economici).

3.3. Analisi

Vari tipi di analisi (per esempio, raggi X, laser ecc.) possono essere utilizzati non solo per la diagnostica ma anche per rilevare aspetti importanti (di varie stesure dei livelli di pittura su di quadro, ad esempio). Banche dati di immagini non tradizionali, ottenibili con varie tecnologie, sono ancora poco sviluppate, mentre potrebbero essere di grande ausilio agli studiosi. Essenzialmente, ad oggi, ci sono episodi interessanti di uso di queste tecnologie avanzate, ma una scarsa diffusione e generalizzazione.

3.4. Ricostruzioni virtuali

Le ricostruzioni virtuali, a due dimensioni o a tre dimensioni, consentono di riprodurre artefatti: edifici, pareti, vie, spazi urbani, oggetti, oggi non più disponibili o gravemente deteriorati o comunque non più corrispondenti all'originale. Le ricostruzioni virtuali possono essere importanti per discutere ipotesi scientifiche su edifici, ambienti oggi distrutti. Nonostante le aspettative e le promesse di qualche anno fa, le ricostruzioni virtuali non sono diventate uno strumento di lavoro per studiosi e addetti ai lavori (contrariamente a quanto succede in altri settori quali la medicina e la biologia, per esempio). Oggi si può dire che le ricostruzioni virtuali sono confinate, e nemmeno con molto successo, all'ambito della divulgazione. Per le attività di studio e ricerca sono un settore di fatto non più di interesse sostanziale: il perché sussista questa sia la situazione non è chiaro.

3.5. Altro

Ci sono svariate applicazioni delle tecnologie ICT che riguardano ambiti specifici di utilizzo. Citiamo, per esempio, la "carta del rischio", vale a dire la possibilità di conoscere, in modo preventivo o per situazioni di emergenza, la situazione di rischio del nostro patrimonio culturale e la sua dislocazione sul territorio. Altre applicazioni possibili possono essere quelle riconducibili a integrazioni di banche dati provenienti da varie fonti (o almeno integrazione delle banche dati che insistono sulla stessa porzione del territorio). Sebbene, da tempo, si discuta circa questo tipo di applicazioni, non esistono ancora realizzazioni concrete: si resta a progetti su carta, studi di fattibilità o, nella migliore delle ipotesi, a prototipi dimostrativi che non hanno la possibilità di diventare veri e propri strumenti di lavoro.

4. DIAGNOSI

Il problema della diagnostica applicata ai beni culturali, in generale, è quello di riuscire ad ottenere informazioni dettagliate e accurate sullo stato di un bene (oggetto, scultura, affresco, edificio ecc.) senza danneggiare in alcun modo (o il meno possibile) il bene stesso. Non è, quindi, possibile, come si fa per altri beni, "saggiare" l'intonaco togliendone un

pezzo, asportare del materiale ecc.. Qualsiasi prelievo fisico (si tratti di edifici, statue, oggetti, dipinti ecc.), in generale, è da escludersi. Le tecniche utilizzate, in funzione del tipo di bene e del tipo di danno da rilevare, possono basarsi sui raggi-X (rilevando, per esempio, la diversa "trasparenza" dei vari materiali), sui raggi laser (rilevando, per esempio, i diversi andamenti di una superficie), sulle onde sonore (intercettando, per esempio, le diverse vibrazioni emesse, in reazione a uno stimolo sonoro, dalle varie parti di una superficie) ecc.. Le tecnologie informatiche e delle telecomunicazioni, in genere, non hanno un ruolo da protagoniste in queste attività di diagnostica. Possono servire a "filtrare" in vario modo il segnale per aumentarne la "leggibilità", a memorizzare in una base di dati i vari rilievi effettuati, e poco più. Le tecnologie veramente rilevanti sono altre.

Una possibilità, potenzialmente rilevante, è quella del tele-rilevamento ai fini diagnostici: come per altre misure "ambientali" dei sensori (opportunosamente dislocati sul territorio e/o in edifici) potrebbero inviare i loro dati in modo continuo a delle centrali informatizzate, che potrebbero sintetizzarli per renderli più leggibili agli esperti. Di queste applicazioni per i beni culturali si parla, ma per ora non si vedono veri e propri risultati, sia per il costo di simili operazioni, sia per la difficoltà (sopra citata) di non essere "invasivi" (come spesso i sensori sono), sia per la complessità organizzativa (in Italia, capita spesso che varie porzioni del territorio siano di pertinenza di amministrazioni diverse, che gli edifici siano di pertinenza di soggetti diversi, che i beni mobili siano di pertinenza di altri ancora).

5. RESTAURO

Restaurare un bene culturale vuol dire modificare il suo stato, o ai fini di una migliore conservazione, o per ripristinare uno stato che in qualche modo si ritiene più corretto (si veda di seguito la discussione sui tipi di restauro). Come per la diagnosi, e ancor più in questo caso, le tecnologie dell'informazione non sono protagoniste. In genere, al di là dell'utilità per la gestione dei dati e l'accesso (anche remoto) agli stessi, informatica e telecomunicazioni hanno poco da dare alle tecniche di restauro.

Le tecniche di comunicazione (sotto descritte) potrebbero giocare un ruolo importante nel dirimere alcune annose questioni in merito al restauro. Spesso capita di discutere su quale debba essere il fine del restauro: *preservare* la situazione attuale nel modo migliore possibile (per esempio, limitandosi a cercare di "conservare" lo stato attuale di un intonaco); *ripristinare* la situazione ottimale (per esempio, "ravvivando" i colori del dipinto sull'intonaco); *ricostruire* la situazione originale (per esempio, ricostruendo le parti di muro crollate e colmando le lacune dell'intonaco). La disputa estrema è tra "comunicatori" (che si spingono, a volte, fino a suggerire la ricostruzione di parti non più esistenti o seriamente danneggiate) e "filologi" (che sostengono che i beni vanno conservati, nel miglior modo possibile, nello stato in cui si trovano), con netta prevalenza, in Italia almeno, dei secondi sui primi. Le tecnologie informatiche potrebbero aiutare in due modi: consentendo, mediante simulazioni, di valutare la qualità "estetica" (e scientifica) del restauro, prima di compierlo; oppure, e più credibilmente, mediante ricostruzioni virtuali, soddisfacendo coloro che vorrebbero, in qualche modo, rivedere l'aspetto originale (di un edificio, di un dipinto, di un oggetto...), lasciando ai "filologi" campo libero. Il restauro "filologico" dei beni culturali, accompagnato da adeguate ricostruzioni virtuali, potrebbe essere una via perseguibile, con soddisfazione di tutti.

6. TUTELA

La tutela del patrimonio dei Beni Culturali in Italia è, naturalmente, una priorità, data la ricchezza del patrimonio stesso; si tratta nello stesso tempo di un'attività difficile, data l'estrema frammentazione del patrimonio e la sua distribuzione sul territorio. Bisogna tener conto del fatto che il patrimonio dei Beni Culturali, nel nostro Paese, si trova non solo nei posti noti (musei, palazzi presidiati, parchi archeologici nazionali) ma anche presso una miriade di realtà poco o per nulla presidiate (chiese, palazzi privati, piccole collezioni e piccoli musei ecc.).

Per quanto riguarda la tutela contro le calamità naturali, la possibilità principale offerta dall'informatica è "la carta del rischio", vale a

nimi. La cultura non è legata alle caratteristiche fisiche degli oggetti, ma alla loro capacità di suscitare emozioni, istituire nessi, risvegliare curiosità. Comunicare equivale, dunque, a fare cultura e fare cultura equivale a comunicare. Per rendere importante un bene culturale agli occhi di un pubblico (grande o piccolo che sia) è importante comunicarne gli aspetti salienti e quindi, per l'analogia sopra stabilita, fare cultura. Comunicare bene è dunque il modo per valorizzare i beni culturali agli occhi del pubblico, e questo vuol dire rendere importanti i beni culturali agli occhi della società. Questo favorirebbe tra l'altro un maggior flusso di risorse economiche verso il mondo dei beni culturali, provenienti sia dalla pubblica amministrazione (in base al consenso popolare) che da sponsor privati (in base all'interesse del grande pubblico) o da mecenati (per promuovere il proprio nome o il ricordo di sé) ecc..

“Valorizzare” non equivale a “monetizzare” in modo diretto e immediato i beni culturali (come schemi semplicistici farebbero intendere), ma piuttosto aumentare le risorse economiche, in seguito ad una corretta comunicazione (che, lo ripetiamo, vuol dire fare cultura).

La comunicazione di cui parliamo, naturalmente, non coincide con una volgare e scorretta divulgazione; bisogna stare attenti d'altro lato che la pretesa di scientificità (a volte purtroppo un paravento che nasconde l'incapacità di comunicare efficacemente) non blocchi ogni sforzo di fare cultura in maniera adeguata ai diversi tipi di pubblico. Le istituzioni culturali del Nord Europa e del Nord America, considerando se stesse come “luoghi di diffusione della cultura”, hanno sviluppato e messo in campo efficaci strategie di comunicazione, con le quali raggiungono ampi strati della popolazione: le famiglie, le scuole, i giovani ecc.. Le conseguenze di questo atteggiamento sono molteplici: le istituzioni culturali sono

immensamente popolari, sono una presenza attiva e vivace nella società, sono capaci di attrarre grandi risorse economiche¹¹, riescono a fornire “servizi” a costi molto contenuti, e spesso non costano nulla alla pubblica amministrazione. Il paragone con le istituzioni culturali in Italia è troppo facile (in senso negativo) per poter essere sottaciuto.

La comunicazione, verso la società nel suo complesso, è la grande assente tra gli obiettivi principali delle istituzioni culturali in Italia: per verificarlo, è sufficiente guardare gli organigrammi delle istituzioni culturali, il modo che hanno di allocare le risorse, i loro programmi attuali e futuri, il modo che hanno di parlare al pubblico¹². Si tenta di compensare questa assenza di comunicazione con eventi spettacolari, che non fanno cultura e non comunicano nulla se non mondanità; oppure si sognano fantomatiche comunicazioni multimediali planetarie che dovrebbero produrre altrettanti fantomatici ricavi (editoriali o turistici) immediati.

Questa lunga premessa serve a chiarire come le possibilità di comunicazione mediante ICT siano realtà vive in gran parte del mondo occidentale e ancora un'aspettativa (troppe volte delusa) in Italia.

Esaminiamo, quindi, le possibilità di comunicazione, offerte dalle tecnologie:

■ **Applicazioni multimediali interattive:** sono applicazioni, disponibili su CD-ROM (ora DVD-ROM), che rendono accessibile il patrimonio culturale. Di gran moda a cavallo tra gli anni '80 e '90, sono state ora in gran parte soppiantate dal Web. L'attuale mercato è costituito soprattutto da applicazioni per le scuole e, solo in minima parte, da prodotti editoriali per adulti. Restano, tuttavia, nella storia dell'editoria multimediale, alcuni prodotti mirabili, per qualità e profondità. La qualità dei siti Web non ha ancora raggiunto la qualità dei migliori CD-ROM. La caratteristi-

¹¹ Una cosa sconvolgente per un europeo è costatare come i grandi musei americani siano gestiti da fondazioni private che si autofinanziano con donazioni, sponsorizzazioni, iniziative culturali e (in misura minima) contributi statali. Naturalmente questo è anche legato alla “detassazione” di spese per la cultura: ma questo è un tema che ci porterebbe lontano.

¹² Il lettore può controllare, ad esempio, quale materiale (brochure, cartelloni, schede ecc.) sia offerto al visitatore di un tipico museo italiano, o la “qualità comunicativa” dei cartellini associati a quadri e oggetti esposti. Esempi brillanti (o esilaranti) sono diciture quali “oinokoe”, “figura maschile”, “vaso di bucchero”, ripetuti all'infinito, con impatto culturale sul visitatore che si può immaginare.

ca negativa di queste applicazioni è che necessitano di una catena distributiva per renderle accessibili al pubblico; la caratteristica positiva è che sono fonte diretta di reddito (anche se sono stati pochissimi i titoli che hanno prodotto veri e propri guadagni). Il mercato rimasto è quello delle grandissime tirature, a basso costo, diffuse in edicola: non sempre queste edizioni risultano, tuttavia, soddisfacenti sia per i contenuti sia per le soluzioni multimediali che offrono.

■ **Siti Web:** oggi sono il maggior veicolo per la diffusione di applicazioni multimediali interattive, nell'ambito dei beni culturali. Sono immediatamente accessibili da tutto il mondo e non hanno costi di distribuzione. La loro facilità di realizzazione ne ha favorito la diffusione, ma anche prodotto risultati di livello molto discontinuo come qualità. Se alcuni siti sono di elevata qualità, molti sono invece di basso livello¹³ e sostanzialmente inutili, se non controproducenti. La caratteristica positiva è che non necessitano di una catena distributiva; la caratteristica negativa è l'impossibilità di procurare ricavi diretti, dato che tradizionalmente l'accesso ai siti Web è gratuito.

Là dove la comunicazione mediante Internet è diffusa (America settentrionale soprattutto e anche Europa settentrionale), i siti Web hanno percorso un'evoluzione tipica: all'inizio, siti che presentavano l'istituzione ("siti vetrina"); in seguito, un'introduzione dettagliata alle collezioni permanenti (quasi una riproduzione delle classiche esposizioni museali); attualmente, descrizione delle mostre ed eventi temporanei (anche a scopo promozionale, per attrarre visitatori), mostre virtuali (utilizzando le tecnologie per eventi disponibili solo via Web), giochi educativi e applicazioni per le scuole (una delle attività più diffuse oggi). Grazie a questi sviluppi le istituzioni riescono a raggiungere un ampio pubblico e a svolgere il loro ruolo di diffusori di cultura. Internet, oggi, è uno straordinario "museo virtuale", accessibile a tutti; i maggiori musei del mondo, ma anche, e soprat-

tutto, i piccoli musei, sono una presenza attiva e stimolante.

Per quanto riguarda l'Italia, la presenza dei Beni Culturali su Internet è a dir poco imbarazzante: l'Italia è virtualmente assente (salvo poche eccezioni) e molto spesso la presenza è purtroppo di bassa qualità. La ragione non è un deficit professionale, quanto, piuttosto, un atteggiamento negativo di fondo del mondo culturale italiano rispetto alla comunicazione, di cui abbiamo già parlato.

■ **Applicazioni per palmari e apparecchi mobili:** queste applicazioni rappresentano la nuova frontiera nel settore, oggi "embrionalmente" disponibili, ma sempre più efficaci nel prossimo futuro. Si tratta dell'evoluzione degli strumenti precedenti: i siti Web, le audioguide e le *brochure* distribuite all'interno dei musei o dei parchi archeologici. Da un lato, consentono approfondimenti e ricerche (come i siti web), dall'altro, guidano il visitatore (come le brochure); offrono inoltre commenti sonori a supporto delle visite (come le audioguide). In aggiunta, queste applicazioni offrono giochi educativi, chat tra visitatori all'interno del museo (o parco), informazioni pratiche, acquisti *on-line* ecc..

■ **Ricostruzioni Virtuali:** sono state già descritte tra le applicazioni a possibile supporto dell'attività scientifica. In realtà, l'uso di tecnologia (soprattutto grafica 3D) per ricostruzioni virtuali si è dimostrato valido soprattutto per la comunicazione e l'intrattenimento. Anche in questo ambito, superato l'entusiasmo iniziale (in cui va riconosciuto che l'Italia ha prodotto alcuni degli episodi più notevoli), il numero di applicazioni realizzate è in notevole diminuzione: forse agli alti costi, necessari per produrre applicazioni di qualità, non corrispondono ricavi equivalenti, mentre applicazioni di bassa qualità non soddisfano più le aspettative del pubblico.

■ **Giochi interattivi:** sono applicazioni d'intrattenimento, con soggetto culturale. La meccanica delle applicazioni è quella dei "classici giochi": puzzle, giochi di abilità, percorsi ad ostacoli ecc.. Al di là del fatto che l'argomento di riferimento sono i beni culturali, i giochi hanno caratteristiche di puro intrattenimento. Che si tratti di vera e propria comunicazione culturale è discutibile, nel senso che non vengono trasmessi veri e propri valori culturali.

¹³ Prodotti di bassa qualità editoriale erano meno comuni per i CD-ROM, dato che i costi di produzione e distribuzione assicuravano almeno i livelli minimi di professionalità.

L'attrazione su giovani e ragazzi è certa; che il loro uso serva ad aumentare la diffusione della cultura è ancora da dimostrare. Certo è che le istituzioni culturali nordamericane sono molto impegnate in questo settore, mentre le istituzioni italiane sono quasi del tutto assenti.

Applicazioni educative: sono applicazioni sempre rivolte ai giovani, ma più tradizionali. Sono applicazioni interattive, in genere più con grafica che multimediali, basate su "racconti" semplici (sullo schema classico dei libri per ragazzi) con soggetto di natura culturale. Possono essere usate (per citare alcuni esempi recenti), per spiegare come giocavano a palla gli Atzechi, come andò la scoperta dell'America, come erano strutturati i canti popolari Messicani ecc.. Sono tra le più divertenti applicazioni oggi in circolazione, ma non hanno (ancora) una diffusione tale da farli considerare fenomeni di massa.

8. FORMAZIONE

Queste applicazioni fungono da supporto alle attività di studio. Ancora una volta stiamo parlando della realtà nordamericana piuttosto che di quella italiana. Si tratta di applicazioni di *e-Learning*, specializzate per il settore dei beni culturali. Materiali educativi, documenti e, soprattutto, immagini, vengono messi a disposizione degli studenti. Il consorzio AMICO, per esempio, mette a disposizione di studenti delle università consorziate immagini di opere d'arte di elevato livello, fornite da alcuni principali musei mondiali; va rilevato che al consorzio non partecipa alcun museo italiano.

9. FRUIZIONE

Queste applicazioni sono ancora in fase di sviluppo iniziale. In generale, la tecnologia non può essere troppo visibile all'interno d'istituzioni culturali. I visitatori di musei d'arte o archeologici non amano mescolare gli oggetti in mostra con apparecchi tecnologici. Per questo motivo all'interno delle istituzioni culturali si usano ancora i tradizionali pannelli cartacei e mai i pannelli elettronici.

La situazione sta mutando per la disponibilità dei "palmari": calcolatori sufficientemente piccoli e leggeri da essere portati in mano.

Come si è già detto questi apparecchi (che combinano le funzioni di approfondimento, audio-guida, e altro ancora) cominciano a diventare una realtà all'interno di musei e parchi archeologici. Hanno sicuramente grande potenzialità nell'offrire al visitatore una esperienza più ricca e valida: che questo succeda realmente è ancora da dimostrare ma fortemente plausibile.

10. I PROGETTI DEL LABORATORIO HOC

Per dare concretezza a quanto discusso fino a questo punto, verranno presentati, in questo paragrafo, alcuni progetti realizzati dal laboratorio HOC, presso il Politecnico di Milano. Si tratta di progetti adatti a fornire una esemplificazione concreta dei concetti sopra esposti (oltre che una testimonianza dello sforzo di un'istituzione italiana per "fare cultura" nel senso che si è detto).

Presenteremo alcuni progetti di applicazioni per musei¹⁴ e per la RAI (Radio Televisione Italiana); un progetto per l'integrazioni di Banche Dati di beni culturali (DICE); un progetto con finalità educativa altamente innovativo (SEE) e, infine, una metodologia che consente l'accesso ai siti Web (relativi al patrimonio culturale) ai non vedenti (WED).

10.1. Applicazioni museali

È il lontano 1996 quando il laboratorio HOC crea una delle prime visite "virtuali" a una sala di un museo: l'esperienza, svolta per e condivisa con i Musei Civici di Pavia, riguardò i preziosi e rari oggetti dell'Arte orafa longobarda (Figura 1). Questa applicazione multimediale, su CD-ROM, si proponeva un duplice obiettivo: offrire al pubblico la possibilità di visitare virtualmente l'allestimento di una sala dei musei che, in realtà, sarebbe stata accessibile fisicamente solo dopo molti mesi, e la possibilità di consultare in seguito, sul posto, le informazioni relative agli oggetti d'arte. Un approccio in-

¹⁴ Musei Civici di Pavia, Museo Appiani-Lopez di Ceresio, museo Poldi Pezzoli di Milano, Museo della Scienza e della Tecnica di Milano, Museo Paolo Giovio di Como, il Museo Archeologico di Milano, la Triennale di Milano e il Museo Civico di Scienze Naturali di Milano.



FIGURA 1

Arte Orafa Longobarda (CD-ROM)

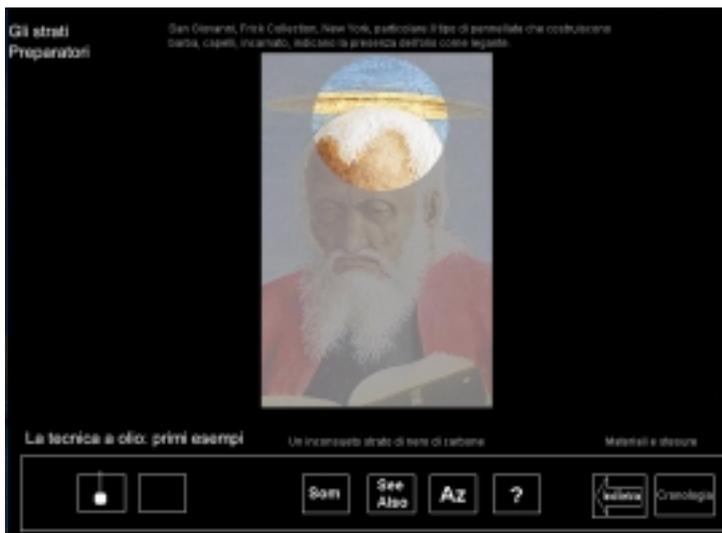


FIGURA 2

Esempio di animazione per tecnica di restauro (Museo Poldi Pezzoli)

teressante, che entusiasmo i curatori dei musei civici e il pubblico, perché apriva la prospettiva nuova della creazione di mostre virtuali in anteprima e punti informativi all'interno del museo stesso. Tali punti informativi erano utilizzati come strumenti propedeutici alla visita e organizzati in percorsi "preconfezionati" (*guided tour*) che l'utente poteva scegliere in base ai propri interessi, seguendo tutte le tappe oppure interrompendo il percorso per proseguire in maniera tradizionale. Il risultato di questa esperienza fu la creazione una sorta di "ipermuseo", un moderno ambiente virtuale navigabile con continuità, a dispetto della frammentarietà delle testimonianze.

Nell'ambito della "multimedialità museale" le iniziative HOC hanno visto la realizzazione di un CD-ROM dedicato al museo etnografico Appiani-Lopez di Porto Ceresio. Il CD-ROM presentava una collezione di manufatti d'arte religiosa, con l'obiettivo non solo di catalogare ma di presentare questi oggetti, esaltandone l'aspetto etnografico. Il supporto multimediale restituì, in pieno, l'espressione della ricca cultura popolare e ne conservò lo spirito, guidando l'utente in un corretto approccio all'argomento. Nonostante il contenuto potesse sembrare insolito, almeno per un prodotto multimediale, il lavoro si rivelò interessante, dimostrando che proprio con le nuove tecnologie era possibile esaltare aspetti "trasversali" al contenuto vero e proprio.

Alla fine degli anni '90, HOC affrontò una sfida difficile: la realizzazione del punto informativo dedicato al "Polittico Agostiniano" per il museo Poldi Pezzoli di Milano, un progetto complesso e di grande successo. La presentazione multimediale del restauro del Polittico di Piero della Francesca, in particolare della tela raffigurante "San Nicola", dimostrò che attraverso le nuove tecnologie si possono "vedere" e presentare aspetti di un'opera inimmaginabili. All'ingresso del museo venne preparato il punto informativo multimediale, attraverso il quale il pubblico poteva documentarsi sulla storia del ritrovamento della preziosa tela, sulle fasi di analisi e attribuzione della tela a Piero della Francesca, sulle modalità artistiche e scientifiche del restauro, sulle tecniche adoperate, sui materiali utilizzati e sulle difficoltà incontrate durante tutto il periodo del restauro (Figura 2). Il successo di questa esperienza consistette nel raggiungimento di due obiettivi: rendere una presentazione multimediale (sintetica per sua "natura") significativa e interessante; restituire in forma multimediale un'infinita mole di contenuti di alto valore scientifico. Molti "artifici", tecnologici, animazioni, effetti televisivi, come ad esempio le sfumature e l'uso dello zoom sui dettagli dell'opera, permisero di presentare in maniera immediata contenuti ostici ai più.

Un altro progetto multimediale per la divulgazione di una materia complessa come la filosofia venne realizzato da HOC per la RAI. Il CD-ROM "Le rotte della filosofia", ideato da

Renato Parascandolo e pubblicato da Paravia, raccoglieva parte dell'imponente lavoro dell'Enciclopedia Multimediale degli Studi Filosofici. Le numerose interviste a filosofi, tutt'ora viventi e conservate negli archivi della Rai, rappresentavano il cuore dell'applicazione multimediale, arricchita da altri contenuti: brani antologici delle opere filosofiche (classiche e non), biografie degli intervistati, presentazione dei periodi storici a cui i pensatori appartenevano. "Le rotte della filosofia" era un'opera realizzata per gli studenti delle scuole superiori, con l'obiettivo di rendere più "interattivo" e più critico lo studio della filosofia. Lo studente poteva, infatti, costruirsi delle lezioni su misura attingendo al ricco materiale presente nell'applicazione. Uno degli esperimenti "culturali-tecnologici" più "azzardati" ha visto protagonista il Museo della Scienza e della Tecnica di Milano. Nell'estate del 2000, HOC realizzò la prima "visita virtuale cooperativa" in ambienti 3D, dedicata alle macchine di Leonardo da Vinci (www.museoscienza.org/leonardo; Figura 3), un'applicazione che permetteva ai "visitatori" presenti contemporaneamente nell'ambiente 3D di azionare le macchine, scambiarsi opinioni in tempo reale, visitare il mondo virtuale "attraverso gli occhi" della guida, senza dover navigare in modo attivo. Il progetto ha ricevuto molti premi, tra cui il riconoscimento americano alla conferenza "Museums & Web" del 2000 (premio come miglior "exhibition on-line"), seguito da una dimostrazione (a New Orleans) ad un pubblico di 400 operatori del settore, provenienti da 30 Paesi diversi. Dopo il successo americano, il museo ha organizzato numerose visite guidate virtuali, con insegnanti e studenti delle scuole medie inferiori, interessati all'approccio ludico dell'applicazione che consentiva di insegnare principi della fisica in maniera semplice e immediata. Storia recente sono gli esperimenti culturali, frutto del fortunato sodalizio con il Museo della Scienza e della Tecnica, che hanno dato alla luce un altro impegnativo progetto: la visita cooperativa alla "Città ideale di Leonardo" (www.museoscienza.org/idealcity). Si tratta della rappresentazione di un luogo che, in realtà, non è mai esistito: una città ideale, costruita unendo le idee architettoniche, urbanistiche, tecniche e artistiche di

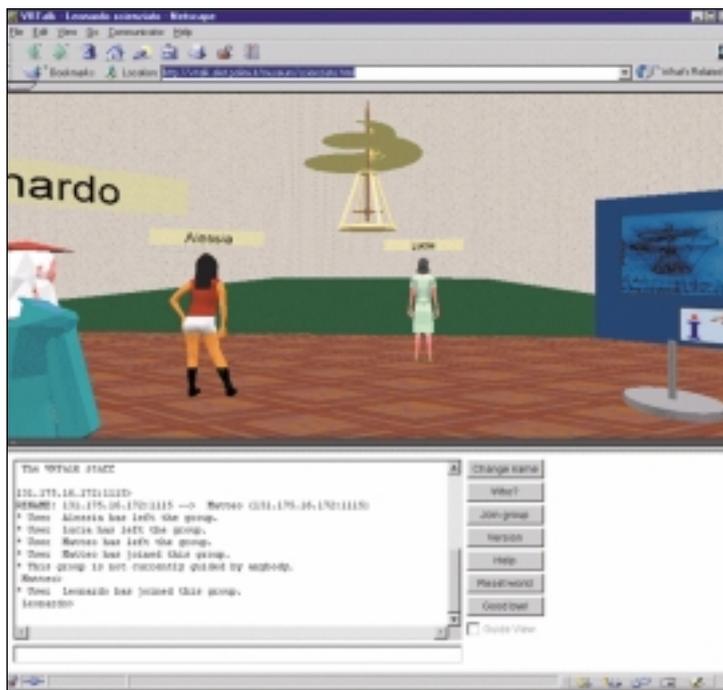


FIGURA 3

"Le macchine di Leonardo": la stanza della vite aerea (Museo della Scienza e della Tecnica)

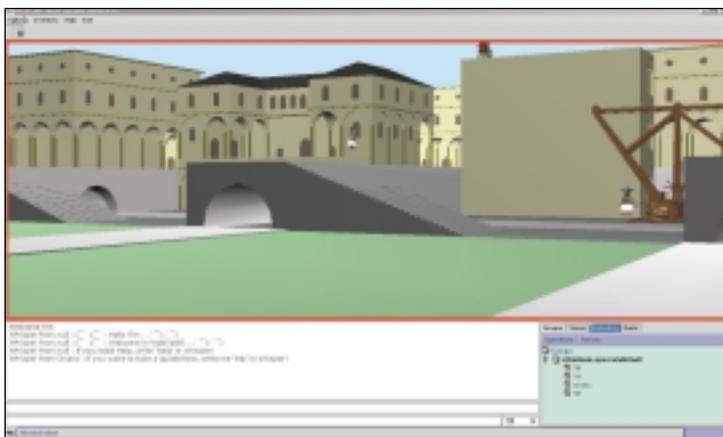


FIGURA 4

La città ideale di Leonardo ricostruita in 3D (Museo Scienza e Tecnica)

che, urbanistiche, tecniche e artistiche di Leonardo. Ispirandosi a un plastico degli anni '50, custodito al museo, HOC ha realizzato geometrie tridimensionali nelle quali è possibile incontrarsi e interagire muovendosi tra i canali, le strade sotterranee, gli edifici, le torri di guardia e i lunghi colonnati della città immaginata dal grande architetto (Figure 4 e 5). Numerosi progetti dedicati ai musei hanno visto anche l'uso della realtà virtuale fotografica, come i tour realizzati per il Museo Paolo Giovio di Como, il Museo Archeologico di Milano (Figura 6), il Museo Civico di

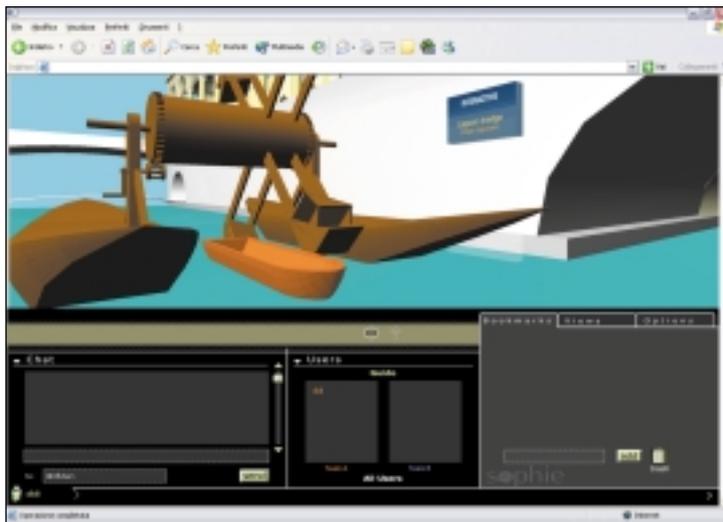


FIGURA 5

Una macchina di Leonardo immersa nel canale della città ideale (Museo Scienza e Tecnica)



FIGURA 6

Una vetrina del Museo Archeologico di Milano

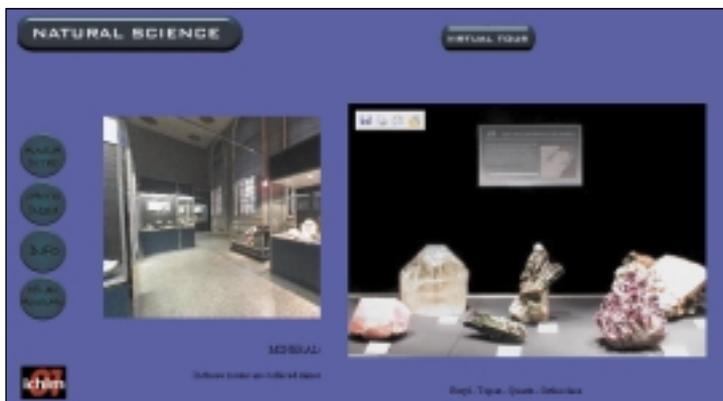


FIGURA 7

Una vetrina minerali del Museo Civico di Storia Naturale

Storia Naturale di Milano (Figura 7), il Museo della Scienza e della Tecnica e la Triennale (<http://hoc.elet.polimi.it/micromusei>; <http://hoc.elet.polimi.it/giovio>). La realtà virtuale fotografica è una tecnologia abbastanza “giovane” che permette di “entrare” in una foto, di guardarsi attorno osservando il mondo che ci circonda in tutti i suoi dettagli. Punto d’incontro tra fotografia professionale e le nuove tecnologie multimediali, la realtà virtuale fotografica trasforma l’immagine 2D in un’esperienza immersiva, unendo grafica 3D e componenti interattive. L’utente ha, perciò, la possibilità di muoversi in un volume di realtà sferico con il semplice utilizzo del mouse, senza bisogno di indossare scomodi occhiali, caschi o guanti. Per realizzare queste affascinanti navigazioni virtuali, tra le tecnologie scelte da HOC c’è IPIX™. IPIX™ permette di realizzare, a costi notevolmente contenuti e con modalità più semplici che con altre tecnologie, un’immagine “sferica” navigabile, a partire da due immagini (due emisferi dell’ambiente) scattate con una particolare lente (*fisheye*). Con questa tecnologia fu realizzato il primo tour all’interno del Museo “Paolo Giovio” di Como: esplorando con il *mouse* le foto si ha l’impressione di visitare le stanze e grazie a *hyperlink*, creati sulle teche, è possibile aprire un’ulteriore finestra del *browser* e “mettere il naso” sulla vetrina per vedere gli oggetti conservati. Quest’applicazione, come quelle che seguirono (il Museo Archeologico di Milano, il Museo Civico di Storia Naturale di Milano, il Museo della Scienza e della Tecnica e la Triennale di Milano), si sono rivelate molto utili per la visualizzazione dei dettagli architettonici di un museo come, per esempio, i pavimenti, spesso maiolicati, e i soffitti, ricchi di decorazioni. L’utilizzo di questa tecnologia potrebbe essere molto utile per la conservazione di opere d’arte che il tempo oppure disastri naturali potrebbero far sparire, cose che non riusciremo più a vedere se non sotto questa forma. A tale proposito, anche il tour, dedicato alla mostra temporanea del “Design Italiano” alla Triennale, aprì un’altra importante prospettiva: rendere “eterno” le mostre temporanee. Chi avesse perso questa mostra del 2001, può ancora visitarla on-line!

10.2. DICE: Integrazione di diverse fonti di informazione sul patrimonio culturale

L'Italia ospita una delle più grandi collezioni di beni culturali al mondo, sparsi sul territorio nazionale e appartenenti ai più diversi soggetti (il Ministero per i Beni e le Attività Culturali, le Regioni, le Province, le città, la Chiesa Cattolica, i musei, privati cittadini...). Le fonti d'informazioni sono altrettanto variegate: dai cataloghi e basi di dati di coloro che custodiscono/ possiedono i beni si va alle produzioni scientifiche delle università, all'editoria del settore artistico, alle produzioni legate al turismo, ecc.. Questa situazione rende complesso raccogliere informazioni su un determinato bene o "tema": uno studioso, un operatore turistico, un editore, deve spendere tempo e denaro per rintracciare e consultare le diverse fonti.

DICE (*Distributed Infrastructure for Cultural Heritage*) è un progetto, co-finanziato dal MIUR (Ministero della Università e Ricerca scientifica), che coinvolge diverse aziende (EDA, eWorks, Marconet, Paleopolis, Colonnese), e università di spicco (Politecnico di Milano e CEFRIEL-Politecnico per la tecnologia e Forma, Scuola Normale di Pisa, per gli aspetti culturali), con il coordinamento scientifico di HOC. Obiettivo di DICE è dimostrare che è possibile integrare diverse fonti d'informazioni creando un ambiente di lavoro efficace per utenti professionali, "mediatori culturali" verso il pubblico (ricercatori, promotori di eventi culturali, curatori museali, operatori turistici ecc.) che necessitano d'informazioni per costruire "prodotti culturali" (guide, cataloghi, monografie ecc.).

DICE si compone di una infrastruttura tecnologica, una piattaforma *software*, un modello organizzativo e un modello culturale.

La base tecnologica è un'infrastruttura *peer-to-peer*. Le informazioni sono rese disponibili a tutti, direttamente, ma ciascun possessore mantiene il controllo sulle proprie. L'utente di muove così in un "universo informativo", con i diversi elementi interconnessi in una rete. L'accesso si svolge secondo quattro modalità:

- Itinerari: sequenze di "schede" predefinite secondo un tema (per esempio, "gli animali negli affreschi romani").

- Ricerche (interrogazioni alla base di dati) basate su profili semantici assegnati alle schede informative.

- Cartelle di lavoro personali, vale a dire *folder* costruiti dall'utente.

DICE fornisce un modello diverso per il contenuto (descritto con schemi XML) e una diversa classificazione semantica per ciascun profilo utente.

Dal punto di vista organizzativo, una comunità DICE si basa su un gruppo di soggetti, fornitori d'informazioni, che si accordano su un "modello culturale" comune, vale a dire la strutturazione delle informazioni e dei profili semantici. Uno schema XML viene utilizzato per descrivere la struttura logica di ciascuna fonte di informazione e i dizionari semantici (comuni alle varie fonti). Algoritmi "fuzzy", basati su attributi di classificazione, sono usati per creare i *link* che sono alla base della navigazione.

Per dimostrare la validità del suo approccio, DICE ha implementato nel marzo 2004 un "dimostratore" con le seguenti caratteristiche:

- Aree culturali: "Archeologia in Campania" e "Ceramica in Campania" (Figura 8).

- Fonti informative: sovrintendenze, musei, ricercatori, editori, per un totale di 20 fonti per ciascun'area culturale.

- Utenti: ricercatori, archeologi, editori, scrittori, promotori turistici ecc..

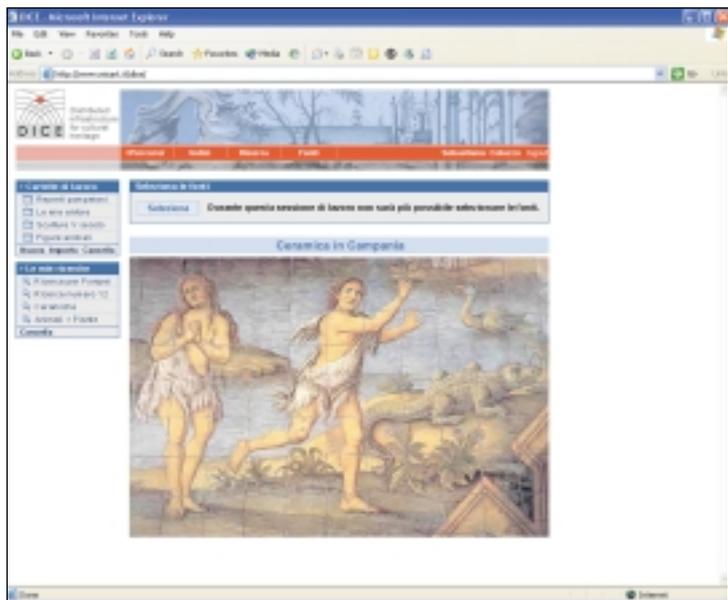


FIGURA 8

Pagina di accesso all'area "Ceramica in Campania" (progetto DICE)

Il dimostratore, che contiene più di 3.000 schede informative, serve da campo di prova per una versione più ampia (con diverse migliaia di schede). Inoltre, è in corso di definizione un accordo con il MiBAC per estendere il dimostratore (utilizzando le varie banche dati sulla Campania) e utilizzarlo come base per realizzare una “sezione dimostrativa” del Portale della Cultura Italiana (per il quale il Politecnico ha realizzato uno studio di fattibilità).

10.3. SEE: un ambiente di edutainment per la scuola

SEE (*Shrine Educational Experience*) è un’entusiasmante esperienza di didattica e intrattenimento (“edutainment”: “education and entertainment”): attività di studio e ricerca tradizionali sono intercalate da cooperazione in un mondo 3D condiviso in Internet, nel quale studenti di scuole e paesi diversi s’incontrano virtualmente. SEE è sta-

to progettato in vista di 4 obiettivi didattici:

- fornire contenuti culturali rilevanti sui rotoli del Mar Morto¹⁵ e argomenti ad essi correlati,
- favorire scambi interculturali,
- offrire interazione e coinvolgimento come stimolo per attività tradizionali di studio,
- incoraggiare l’uso di Tecnologie Innovative per la didattica.

Ogni “Esperienza” coinvolge 4 classi di studenti tra i 12 e i 19 anni d’età, di paesi diversi, e dura 6 settimane: nell’arco di questo periodo gli studenti svolgono diverse attività che li conducono da una non esistente o scarsa conoscenza del rotoli del Mar Morto, attraverso l’esame approfondito di alcuni temi particolari, a una comprensione profonda dei nessi tra una cultura antica di 2.000 anni e il mondo attuale. L’aspetto innovativo di SEE risiede nei 4 incontri on-line che si svolgono in un ambiente virtuale tridimensionale (che rappresenta una diretta evoluzione di quello realizzato a suo tempo per il “Leonardo Virtuale”): 2 studenti per classe, rappresentati da “avatar” (rappresentazioni grafiche degli utenti), “entrano” nel mondo virtuale per incontrare le altre classi e una “guida” (Figura 9).

I contenuti – scaricabili in formato stampabile dal sito SEE (www.seequmran.it) – consistono in interviste a esperti mondiali dei manoscritti, di letteratura antica ecc.. A differenza dei libri di scuola, le interviste offrono un punto di vista critico e sfaccettato sulla ricerca di livello accademico, ma allo stesso tempo in un formato gradevole e diretto.

Gli studenti sono organizzati in due squadre, in competizione tra loro: questa è la parte più coinvolgente dell’esperienza e uno stimolo potente allo studio dei contenuti. Per quanto l’abilità “fisica” sia richiesta per vincere, nessun punto viene assegnato senza una risposta soddisfacente ai quiz culturali. In questo modo, anche gli studenti meno inclini allo studio, desiderosi di far bella figura nei giochi, studiano la loro parte e supportano la squadra (Figura 10 e 11).



FIGURA 9
“Avatar” nel mondo 3D



FIGURA 10
Un “avatar” si cimenta nei “giochi olimpici”

¹⁵ Manoscritti databili tra il II sec. a.C. e il 68 d.C., scritti probabilmente da una comunità di Ebrei Esseni che viveva nel deserto di Giuda presso il sito archeologico di Khirbet Qumran. Rappresentano la più antica versione di libri della Bibbia, una fonte unica per comprendere le radici della civiltà occidentale.

La prima fase sperimentale di SEE (primavera 2003) ha coinvolto 36 classi da Italia e Israele, con circa 1000 studenti, e ha dimostrato lo straordinario valore educativo del progetto. Una nuova sperimentazione, con classi da Italia, Israele, Belgio e Germania, si è svolta nella primavera del 2004, confermando entusiasmo ed efficacia.

10.4. Il progetto WED: accessibilità al Web “come un dialogo”

Il Web si è rivelato uno dei mezzi di comunicazione più efficaci e al contempo economici per diffondere il patrimonio culturale e raggiungere un ampio pubblico; tuttavia, la sua natura “visiva” esclude di fatto una fascia di utenti socialmente molto rilevante: i disabili alla vista. Gli utenti non vedenti accedono al Web facendo uso di *screenreader*, vale a dire, un software che interpreta il codice HTML e lo legge “ad alta voce”. Non tutti i siti, però, hanno le caratteristiche necessarie per “essere letti” dagli *screenreader*: a questo riguardo, il consorzio W3C ha emanato, nel maggio 1999, una serie di linee guida per l’accessibilità, la cui seconda versione è attualmente in corso di preparazione (www.w3.org/TR/2003/WD-WCAG20-20030624). Tali linee guida hanno lo scopo di aiutare a costruire siti effettivamente accessibili ai non vedenti. È risultato presto chiaro però che la rispondenza alle norme W3C non è assolutamente sufficiente a garantire un uso efficiente e soddisfacente del Web ai non vedenti.

Il progetto WED (*Web as Dialogue*), svolto in cooperazione dal Politecnico di Milano e l’Università della Svizzera Italiana, affronta il tema dell’accessibilità. L’obiettivo è quello di costruire nuovi principi di *design* a partire dal paragone tra dialoghi umani e dialoghi “uomo-macchina”, per spostare l’interazione dal canale visivo a quello orale e creare una sorta di “dialogo” tra il sito e l’utente.

Attualmente, i dialoghi con il Web sono assai poco “naturali”: gli *screenreader* adottano una strategia di lettura dall’alto a sinistra a in basso a destra che rende l’interazione con il sito lenta e faticosa.

Il lettore può rendersene conto provando a leggere un quotidiano con la stessa strate-



FIGURA 11

Gli “avatar” riflettono sulla risposta al primo quiz

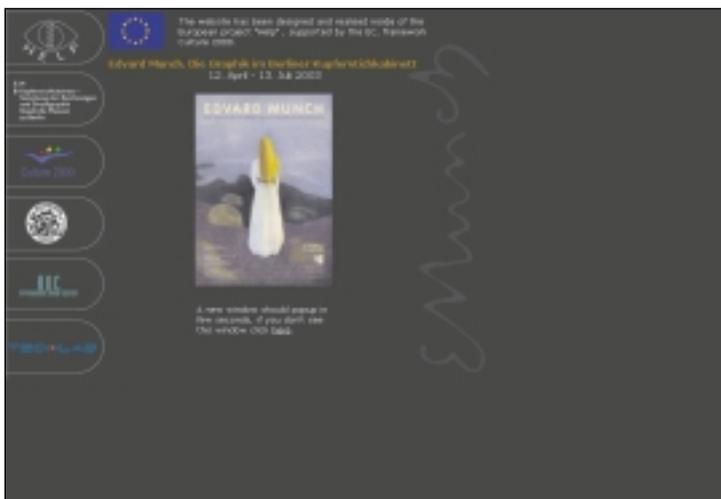


FIGURA 12

Home page del sito di Munch

gia: quanto tempo occorre prima di raggiungere un argomento rilevante? O perché sia chiaro quali sono i principali argomenti trattati?

Allo scopo di mettere alla prova i primi risultati della ricerca, il gruppo WED ha realizzato, nell’ambito del progetto europeo HELP, il sito per una mostra di stampe di Munch che si è svolta ai Musei di Stato di Berlino nella primavera del 2003 (www.munchundberlin.org) (Figura 12).

Nel sito di Munch si trovano diverse funzioni speciali rivolte a migliorarne l’accessibilità; una delle più importanti è il “*page schema*”. Lo *screenreader* legge anzitutto lo schema della pagina, vale a dire, una sorta d’indice delle sezioni di contenuto principali. In questo modo, l’utente può accedere direttamente alla sezione cui è interessato. Il *page schema* non compare visualizzato nella pagina, ma viene solo letto dallo *screenreader* (Figura 13).



FIGURA 13 **11. CONCLUSIONI**
 Template di pagina
 dal sito di Munch

Malgrado le numerose critiche alla situazione italiana, le conclusioni sono improntate a un cauto ma convinto ottimismo:

■ il patrimonio italiano di Beni Culturali è tra i più rilevanti al mondo;

■ la preparazione scientifica degli “addetti ai lavori” italiani è spesso notevole;

■ c’è una crescente attenzione per un uso efficace dell’ICT nel mondo dei Beni Culturali;

■ in molti corsi (di Master soprattutto) si trovano dei giovani brillanti laureati in discipline umanistiche che si interessano alle tecnologie e al loro uso;

■ è crescente l’attenzione della “politica” per il problema.

È sicuramente necessario tuttavia reperire maggiori risorse economiche “dal mercato” e quindi stimolare un dibattito più consapevole e coraggioso.

Questo lavoro vuole appunto costituire uno stimolo (in qualche caso una provocazione) a una discussione più incisiva, seria, libera dalle eccessive pastoie della “tradizione” ma anche meno legata alle mode e alle fughe in avanti.

Bibliografia

Relativa alla comunicazione dei beni culturali attraverso le nuove tecnologie:

[1] Bearman D., Tran J.: (Eds.) *Museums and the Web 2004*. Selected Papers from an international conference. Pittsburgh: Archives & Museum Informatics, 2004.

[2] *International Cultural Heritage Informatics Meeting*. Proceedings from ichimo4. CD ROM. Archives & Museum Informatics, 2004.

[3] *International Cultural Heritage Informatics Meeting*. Proceedings from ichimo3. CD ROM. Archives & Museum Informatics, 2003.

[4] Bearman D., Tran J.: (Eds.) *Museums and the Web 2003*. Selected Papers. Pittsburgh: Archives & Museum Informatics, 2003.

[5] Bearman D., Tran J.: (Eds.) *Museums and the Web 2002*. Selected Papers. Pittsburgh: Archives & Museum Informatics, 2002.

[6] Bearman D., Garzotto F.: (Eds.) *International Cultural Heritage Informatics Meeting*. Proceedings from ichim01. Pittsburgh: Archives & Museum Informatics, 2001.

[7] Bearman D., Tran J.: (Eds.) *Museums and the Web 2001*. Selected Papers. Pittsburgh: Archives & Museum Informatics, 2001.

Relativa ai progetti presentati in questo articolo:

Progetti Leonardo Virtuale e La città ideale

[8] Barbieri T., Paolini P.: *Reconstructing Leonardo’s Ideal City - From Handwritten codexes to WebTalk-IIa: A 3D Collaborative virtual environment system*, in proceedings ACM VAST 2001, Athens, 2001.

[9] Barbieri T., Paolini P.: *Cooperation Metaphors for Virtual Museums*, in Bearman D. & Trant J. (Eds), (2001) *Museums and the Web 2004*. Selected Papers from an International Conference, Archives & Museum Informatics, Seattle, U.S.A.

[10] Barbieri T.: *Networked Virtual Environments for the Web: The WebTalk-I and WebTalk-II Architectures*. In: Proceedings IEEE for Computer Multimedia, Expo 2000 (ICME), New York, USA, July 2000.

[11] Paolini P., Barbieri T., et al.: *Visiting a Museum Together: how to share a visit to a virtual world*, in Bearman D., Trant J. (eds), (1999) *Museums*

and the Web 2004. Selected Papers from an International Conference, Archives, Museum Informatics, New Orleans, U.S.A..

- [12] Paolini P., Barbieri T., et al.: Visiting a Museum Together: how to share a visit to a virtual world. In *Journal of The American Society for Information Science*, Wiley&Sons, October 1999.

Progetto DICE

- [13] Colazzo S., Perrone V.: *Integrating Distributed Heterogeneous Information Sources for Cultural Heritage: the DICE approach*. In Proceedings of the Third International Workshop on Presenting and Exploring Heritage on the Web – PEH '04, co-located with DEXA 2004, 30 August – 3 September 2004, Zaragoza, Spain.

Progetto SEE

- [14] Di Blas N., Gobbo E., Paolini P.: *Immersive 3D and Cultural Heritage: Suggestion or Realism?* In Bearman D., Trant J. (Eds), (2004) *Museums and the Web 2004*. Selected Papers from an International Conference, Archives & Museum Informatics, Vancouver, British Columbia, U.S.A. (accepted paper).
- [15] Gobbo E., Paolini P.: *What We Have Learned By Using Cultural Heritage, Technology Based, Educational Environment: SEE (Shrine Educational Experience)*. ICHIM International Cultural Heritage Informatics Meeting 2004 Proceedings, Berlin.
- [16] Di Blas N., Paolini P., Poggi C.: *A Virtual Museum where Students can Learn*. In R. Subramaniam (ed.) *E-learning and Virtual Science Centers*, Idea Group Inc., U.S.A. (i.c.s.).
- [17] Di Blas N., Paolini P., Poggi C.: *Learning by Playing. An Edutainment 3D Environment for*

Schools, in Proceedings of ED-MEDIA 2004. World Conference on Educational Multimedia, Hypermedia, Telecommunications, June 21-26, 2004; Lugano, Switzerland.

- [18] Di Blas N., Paolini P., Poggi C.: *Shared 3D Internet environments for education: usability, educational, psychological and cognitive issues*. In J. Jacko & C. Stephanidis (eds) *Human - Computer Interaction: Theory and Practice*. Volume I of the Proceedings of HCI International 2003, LEA 2003 Lawrence Erlbaum Associates, Mahwah, New Jersey.

- [19] Di Blas N., Hazan S., Paolini P.: *Edutainment in 3D virtual worlds*. The SEE experience, in Bearman D. & Trant J. (Eds), (2003) *Museums and the Web 2003*. Selected Papers from an International Conference, Archives & Museum Informatics, Charlotte, South Carolina, U.S.A..

Progetto WED

- [20] Di Blas N., Paolini P., Speroni M.: *Web Accessibility for Blind Users. Towards Advanced Guidelines*. Proceedings of the 8-th ERCIM UI4ALL Workshop, June 27-28, Vienna, Austria, 2003.
- [21] Di Blas N., Paolini P., Speroni M., Capodiecchi A.: *Enhancing accessibility for visually impaired users: the Munch's exhibition*. In Bearman D. & Trant J. (Eds), (2004) *Museums and the Web 2004*. Selected Papers from an International Conference, Archives & Museum Informatics, Arlington, Washington, U.S.A.
- [22] Di Blas N., Paolini P.: "There And Back Again": What Happens To Phoric Elements in a "Web Dialogue". *Journal of Document Design*, Vol. 4, n. 3, 2003, p. 194-206.

PAOLO PAOLINI è professore ordinario al Politecnico di Milano e docente all'Università della Svizzera Italiana di Lugano. È coordinatore scientifico di NET-LAB, una rete di laboratori impegnati nella ricerca su comunicazione avanzata e nuove tecnologie, specialmente nell'ambito dei beni culturali. NET-LAB comprende il laboratorio HOC (Politecnico di Milano, campus di Milano e Como), TEC-LAB (Università della Svizzera Italiana) e SET-LAB (Università di Lecce).
paolini@elet.polimi.it

FRANCESCA ALONZO è socio fondatore e responsabile di Sophie srl, un'azienda incubata presso l'Acceleratore del Politecnico di Milano, che sviluppa progetti di comunicazione multimediale. È stata consulente presso l'HOC (Hypermedia Open Center) del Dipartimento di Elettronica e Informazione - Politecnico di Milano, collaborando alla progettazione di applicazioni ipermediali culturali e di siti collaborativi per il Web.
alonzo@sophie.it

NICOLETTA DI BLAS insegna Teoria della Comunicazione presso il Polo di Como del Politecnico di Milano. È laureata in Lettere Classiche e ha ottenuto un dottorato in Linguistica all'Università Cattolica di Milano. Attualmente la sua attività di ricerca, all'interno del laboratorio HOC del Politecnico di Milano, si concentra su tematiche di linguistica, usabilità e applicazioni con finalità didattiche, specialmente nel campo dei beni culturali.
diblas@elet.polimi.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 TEKNO

Roberto Bellini¹

1. INTRODUZIONE

TEkNO MP è una piccola azienda del settore meccanico, con sede nel Varesotto, specializzata nella produzione per conto terzi e su commessa di verricelli e, successivamente, di piccoli sistemi a trazione elettrica (fino a 2 tonnellate) basati su motoriduttori; a seguito dell'introduzione sul mercato della nuova famiglia di prodotti, decide di introdurre anche una innovazione di processo con lo sviluppo di applicazioni ICT e con la certificazione di qualità. A fine 2004 TEKNO fattura circa 10,3 mni di euro con 70 dipendenti e nel corso del 2005 sta ulteriormente incrementando fatturato e margine. Il successo viene imputato oltre che alla nuova linea di prodotti, alla conduzione a buon fine di un progetto di reingegnerizzazione dei processi di produzione ottenuto con la introduzione della certificazione di qualità e di un sistema integrato di pianificazione e gestione della produzione supportato da una applicazione ICT avanzata; il caso approfondisce in particolare questa ultima innovazione, che permette alla TEKNO di distinguersi per la capacità di adeguarsi alle necessità della clientela in termini di volu-

mi e qualità della produzione dei prodotti finali, pur lavorando su commessa.

2. STORIA DELLA TEKNO MP

TEKNO MP nasce circa 25 anni fa da due giovani e intraprendenti operai, Danilo Molla e Giovanni Pirola, che dopo una breve esperienza di lavoro come dipendenti, decidono di mettersi in proprio come artigiani, specializzati in lavorazioni sofisticate da tornio e fresa; l'azienda parte con 5 dipendenti, inclusi i proprietari, e dopo qualche anno inizia a produrre, come terzista di una multinazionale, ingranaggi per verricelli. L'azienda cliente è la Harken, numero uno al mondo come produttore di verricelli Winch (Figura 1), che equi-



FIGURA 1
Verricello Winch

¹ Hanno collaborato Luca Peron e Antonello Leonello, studenti del Politecnico di Milano.

paggia oggi anche una delle barche della Coppa America. TEKNO consolida il suo rapporto con la Harken passando dalla produzione dei 16 ingranaggi del verricello al prodotto finito Winch e nel corso degli anni ne diventa il fornitore unico: verso la metà degli anni '90, TEKNO si è consolidata con un fatturato di 2,3 mni di euro e 25 addetti. Comunque già dall'inizio della loro attività imprenditoriale, gli imprenditori Molla e Pirola avevano messo a fuoco una strategia che prevedeva lo sviluppo di un secondo tipo di prodotto, che fosse sviluppabile a partire dalle stesse competenze e che permettesse all'a-

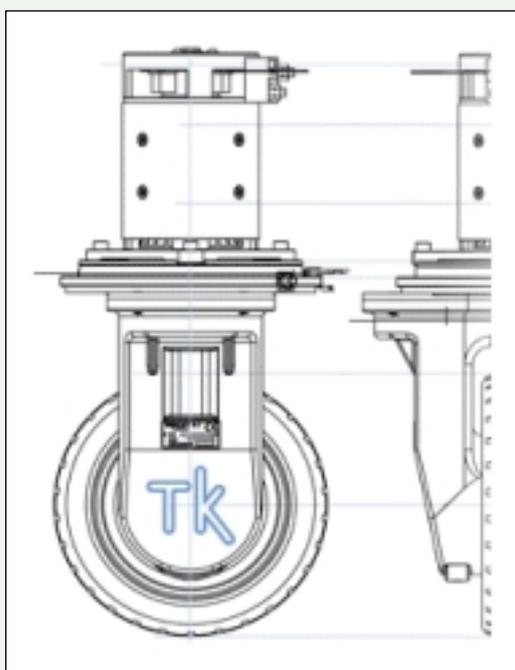


FIGURA 2
Sistema di trazione

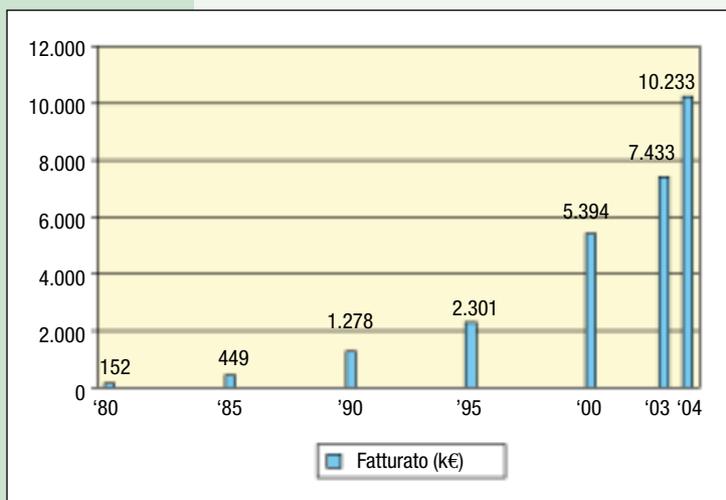


FIGURA 3
Andamento del fatturato TEKNO

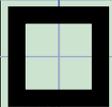
zienda di stare sul mercato con propri prodotti e un certo numero di clienti qualificati. Una volta raggiunto il consolidamento sul primo prodotto, nel 1997 la TEKNO decide quindi di entrare sul mercato con la sua nuova linea di motoriduttori che equipaggiano "sistemi di trazione elettrici"; tali tipi di sistemi si vanno diffondendo sempre di più ed oggi sono il componente fondamentale di carrelli elevatori, di macchine per la pulizia industriale, di piccole piattaforme aeree, di golf-car, di veicoli a 4 ruote per disabili ecc. (Figura 2).

L'innovazione di prodotto costringe l'azienda alla trasformazione da terzista mono-cliente ad azienda con un proprio portafoglio prodotti che deve soddisfare un certo numero di clienti, in linea di massima di medio- grandi dimensioni, nazionali e internazionali, fra i quali si possono annoverare Fiat OM Carrelli Elevatori, Linde, Diversey Lever, Stöcklin, Schabmüller, Atech, Gansow ecc..

Negli ultimi 4 anni il fatturato della TEKNO si impenna: l'incremento del fatturato 2004 sul 2003 è del 37% (Figura 3); si sviluppa anche l'area produttiva, in un primo tempo su oltre 3000 mq di superficie coperta, e successivamente nel corso del 2004 fino a 5000 mq, per far posto alle nuove macchine di produzione a controllo numerico (CNC).

I fattori principali che hanno determinato il successo vengono considerati i seguenti:

1. i motoriduttori non sono un nuovo prodotto per il mercato; l'innovazione introdotta da TEKNO è costituita dal diverso approccio alla clientela di questo tipo di prodotti: mentre i concorrenti (le "grandi", come vengono chiamati i principali concorrenti) sviluppano sistemi di trazione come prodotti standard, Pirola e Molla scelgono un approccio di fornitura su commessa: basandosi sulle loro competenze, offrono la co-progettazione su specifica del cliente e poi la relativa produzione su commessa a prezzi comunque competitivi rispetto ai prodotti standard;
2. il mantenimento dell'ormai tradizionale lavoro di terzista per la Harken, che contribuisce a finanziare l'innovazione di processo;
3. lo sviluppo, una volta che il nuovo prodotto ha dimostrato di sapersi conquistare e mantenere una nicchia di mercato, di un progetto di reingegnerizzazione dei processi di produzione, sia per la linea winch che per quella dei mo-



toriduttori, in cui il fulcro è costituito dal nuovo sistema di pianificazione.

3. ORGANIZZAZIONE E PROCESSI TEKNO

Una volta colti i bisogni del mercato (spazio nell'offerta di "motoriduttori progettati e realizzati su commessa) e costruito un primo portafoglio di clienti business, TEKNO si dedica ad analizzare il proprio modo di operare rispetto al modo in cui si esprimono e si potranno esprimere le richieste del mercato; obiettivo di tale analisi è quello di capire se e in qual modo l'azienda potrà essere capace di rispondere con successo. Dall'analisi che l'azienda condusse su se stessa in termini molto crudi, emerse la seguente diagnosi:

□ **Pianificazione:** la pianificazione della produzione veniva sviluppata in modo molto artigianale rispetto ai bisogni del mercato; anche se i clienti acquisiti non manifestavano particolari malcontenti, TEKNO riconobbe che le proprie prestazioni in fase di acquisizione dell'ordine in termini di risposte fornite ai clienti erano incerte, comunque approssimate, in ritardo e inefficienti; sostanzialmente la previsione di vendita annuale contrattata con il cliente, veniva ribaltata sulla produzione che a sua volta si impegnava con gli ordinativi di componenti commerciali e grezzi da lavorare presso i propri fornitori.

□ **Produzione:** una volta acquisito l'ordine, la produzione riusciva a fatica a rispondere alle tempistiche di consegna con cui l'azienda si era impegnata con i clienti; inoltre il disordine e la confusione del reparto produzione andavano aumentando con la necessità di produrre in parallelo due linee di prodotti, tutti su commessa. Infatti, la mensilizzazione degli acquisti e della produzione dei semilavorati fatta a inizio anno non permetteva di rispondere in termini ottimali alle successive variazioni richieste dai clienti stessi in funzione delle rispettive esigenze di mercato, pur mantenendo l'impegno sui volumi totali ordinati: a causa delle richieste di variazione nei tempi di consegna da parte di ciascun cliente, si venivano a creare a livello mensile sia problemi di *stock out* che problemi di eccesso di materie prime, semilavorati e componenti. Emerse con forza che per mantenere e sviluppare un ruolo sul mercato sarebbe stato neces-

sario introdurre una serie di innovazioni di processo che permettessero

- di migliorare il livello di flessibilità della produzione;
- di ridurre i costi ottimizzando i processi;
- di aumentare la qualità della produzione e migliorare la soddisfazione dei clienti.

Dall'analisi del flusso dei materiali e delle lavorazioni emerse ancora che i lead time medi di consegna, mano a mano che si riducevano per la pressione dei clienti, erano diventati inferiori ai lead time di produzione; per poter rispondere a questa esigenza TEKNO modificò in modo significativo il flusso dei materiali e delle lavorazioni introducendo una parziale modalità push (per previsioni); la soluzione per poter operare in parte in modalità push e in parte in modalità pull fu quella di inserire, all'interno del flusso produttivo, un **magazzino semilavorati** che permettesse di disaccoppiare le attività di produzione da quelle di assemblaggio.

Furono poi identificati altri due tipi di interventi:

- l'introduzione della certificazione di qualità Vision 2000;
- l'installazione di un sistema di supporto informatico per la pianificazione e gestione della produzione, che viene analizzata nel capitolo 4.

3.1. Come si presenta l'azienda dopo la certificazione

Il flusso dei materiali e delle lavorazioni si presenta articolato come segue:

a. ricevimento della fornitura

□ TEKNO ha sviluppato uno stretto e continuativo rapporto con circa un centinaio di Fornitori (di cui 25 terzisti) per l'approvvigionamento di semilavorati grezzi o di componenti commerciali;

□ da ciascuno di questi vengono ricevuti i pezzi ordinati a suo tempo; l'automazione del processo di Pianificazione Strategica della produzione permette da una parte di fornire le indicazioni sui tempi di consegna dei lotti di prodotto commissionati dai clienti e dall'altra di alimentare il processo di pianificazione operativa per il lancio ottimale degli ordini ai fornitori tenendo conto sia del tempo di fornitura del terzista che di quello di produzione interna dei semilavorati che alimentano il magazzino semilavorati.

b. ciclo di lavorazioni

Le lavorazioni che si svolgono in azienda sono sia per asportazione di truciolo sia per dentatura e rettifica ingranaggi e vengono effettuate

mediante torni, frese, dentatrici e rettifiche evolventi. Mediante queste lavorazioni si ottengono a partire dal grezzo componenti finiti che vengono passati al magazzino semilavorati, da cui vengono poi ripresi per la fase di assemblaggio; tali lavorazioni vengono anche eseguite tramite le macchine a controllo numerico CNC introdotte in azienda a fine 2004.

Le distinte di produzione si articolano su due livelli:

- un grezzo di partenza;
- più fasi interne ed esterne che formano il Ciclo per ottenere il FINITO (1.300 fasi in anagrafica di cui 1.000 interne e 300 esterne).

Il numero di componenti che possono entrare a far parte del Prodotto Finito sono circa 1.500 di cui:

- 700 acquistati, di tipo commerciale (viti, bulloni, motori, freni, cuscinetti....);
- 500 di produzione interna (ingranaggi, alberi, coperchi,....) chiamati FINITI;

- 300 di grezzi commerciali per la lavorazione (Stampati, Fusioni, Barre).

c. *assemblaggio manuale*

Ciascuna delle due linee di prodotto (winch e motoriduttori) viene assemblata su una linea separata nell'ambito del reparto assemblaggio; l'assemblaggio viene alimentato dal magazzino semilavorati, opportunamente rifornito in base alla Pianificazione Operativa; i montaggi durano circa due giorni e più codici di prodotto vengono montati contemporaneamente con frequenze mediamente mensili (Figura 4).

3.2. Volumi di produzione e dotazioni tecnologiche

Ogni mese vengono lanciati in produzione circa 120 codici di FINITI, fra winch e motoriduttori; le distinte base delle due linee (winch e riduttori) non hanno componenti comuni mentre all'interno della stessa linea di prodotto i componenti comuni tra i vari codici sono frequenti. TEKNO ha adottato, infatti, la tecnica della progettazione modulare per abbattere i costi; con questa tecnica si assemblano tra loro dei moduli standard già progettati e conosciuti dall'azienda per ottenere il prodotto finale che deve rispondere alle esigenze funzionali richieste dal cliente.

La dotazione tecnologica per reparto è molto consistente:

- Reparto Tornitura-2 Centri di lavoro comprende ciascuno 8 Torni a CNC;
- Reparto Dentatura e Rettifica comprendente 9 dentatrici, 2 rettifiche denti, 1 rettifica diametri esterni;
- Reparto Montaggio Riduttori comprendente 6 banchi di montaggio, per altrettanti posti di lavoro;
- Reparto Montaggio WINCH comprendente 6 banchi di montaggio, per altrettanti posti di lavoro.

3.3. La distribuzione delle responsabilità

Dal punto di vista della distribuzione delle responsabilità, i proprietari Molla e Pirola svolgono di fatto il triplice ruolo di **Imprenditori, Direzione Generale e Direzione Operativa**, coprendo quindi sia le responsabilità sulla strategia che quelle operative.

Un ruolo particolare viene riconosciuto al sig. Sartori, **responsabile di Processi, Qualità e**

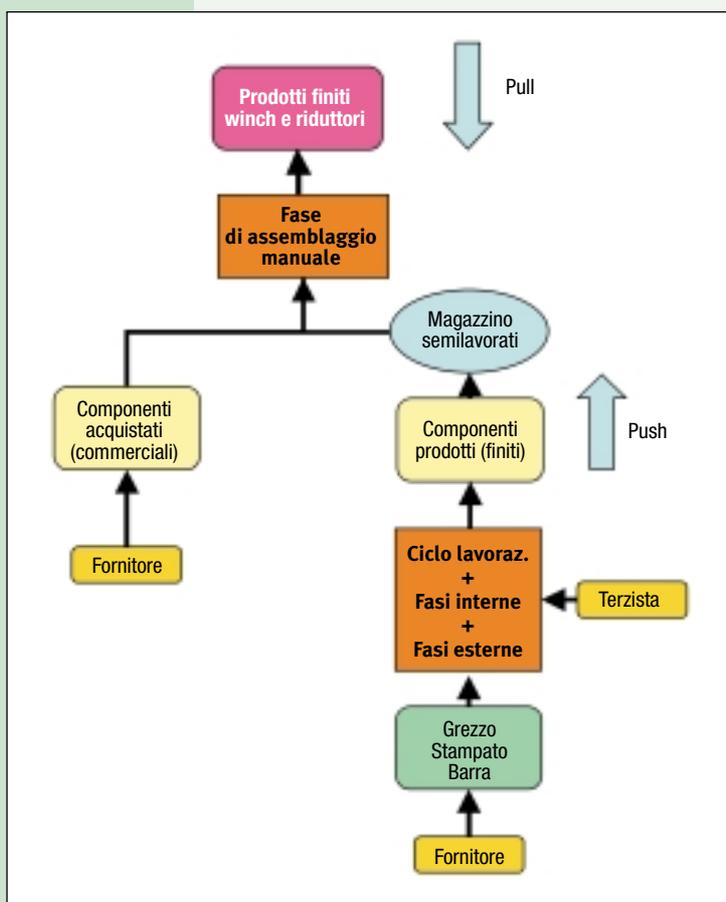


FIGURA 4

Flussi dei materiali e delle lavorazioni con l'inserimento del magazzino semilavorati

ICT; Sartori ha la responsabilità della ottimizzazione operativa del funzionamento aziendale e in particolare di sviluppare i piani di miglioramento e di farli accettare dall'organizzazione; in questo ruolo ha gestito con successo il progetto di introduzione dell'ERP e del sistema di supporto alla pianificazione.

Dal punto di vista operativo il sig. Pirola ha la duplice responsabilità di **Direttore tecnico e vendite**; la decisione di concentrare in un'unica figura queste due competenze deriva dal riconoscere che i prodotti trattati dal TEKNO hanno un alto contenuto tecnologico; chi deve trattare coi clienti non deve soltanto avere capacità relazionali e abilità commerciali, ma deve anche possedere un bagaglio tecnico adeguato per comprendere i bisogni del cliente e, eventualmente, consigliarlo e fornirgli assistenza. La funzione impegna 4 addetti, oltre a Pirola; questi d'altra parte ricopre anche il ruolo di responsabile dell'amministrazione, con il supporto di 2 figure operative.

Il sig. Molla, come **Direttore di produzione**, è responsabile non solo della pianificazione, produzione ed assemblaggio ma anche degli acquisti, della logistica e del controllo qualità, svolgendo un ruolo sostanzialmente di responsabile delle operations. La funzione produzione coinvolge il maggior numero di responsabili intermedi (10) ed il maggior numero di addetti, dedicati alla produzione.

Il totale degli addetti a fine 2004 è di 70 unità.

4. STRUTTURA FONDAMENTALE DEL SISTEMA INFORMATIVO AZIENDALE

Il sistema informatico di TEKNO consiste in:

- hardware HP e Software Microsoft, per le elaborazioni dati di ufficio e la gestione della posta elettronica;
- rete di fibre ottiche per la connessione di 25 PC e 2 Server per la gestione dei data base aziendali.

Tutti i processi sono gestiti dalle seguenti applicazioni software:

- Enterprise software;
- Production planning through MRP II;
- Production scheduling;
- MES (*Manufactory Execution System*);
- Documents management through Microsoft Office;

- Projects Management through Microsoft Project.

La più recente innovazione è consistita nell'introduzione in TEKNO di due applicazioni informatiche, integrate fra di loro, a supporto delle attività di pianificazione e programmazione della produzione:

- Sistema gestionale ERP (*Enterprise Resource Planning*);
- APS (*Advanced Planning and Scheduling*).

Il **Sistema ERP** è quello di Microarea su SqlServer e gestisce: amministrazione, ciclo attivo e passivo, anagrafiche di magazzino, movimenti di magazzino, ordine di produzione, DDT c/lav, anagrafiche distinte e cicli, consuntivazione produzione, stampe ordini di produzione.

Il **Sistema APS**, integrato con il sistema ERP, permette di sfruttare al meglio le base dati di quest'ultimo e le sue procedure di aggiornamento, per supportare le seguenti funzionalità:

- Stesura del piano di montaggio dei prodotti finiti (modulo MRP II - MPS);
- Esplosione dei fabbisogni di materiali (modulo MRP II - MRP);
- Schedulazione degli ordini di produzione (modulo SCHEDULATORE);
- Fattibilità del piano di montaggio (modulo SCHEDULATORE);
- Raccolta dati di avanzamento (modulo MES).

In figura 5 vengono sintetizzate le funzionalità che permettono di ottenere i risultati della Pianificazione Strategica della produzione e di quella Operativa.

Le funzionalità e i dati messi a disposizione dal sistema ERP e dal sistema APS supportano le 2 fasi di pianificazione cruciali per l'ottimizzazione dei cicli di produzione:

1. la pianificazione strategica;
2. la pianificazione operativa.

La Pianificazione Strategica, a partire dalle previsioni di vendita, permette di valutare la capacità di produzione rispetto al carico delle risorse e fornisce l'indicazione dei tempi di consegna al committente del lotto.

La Pianificazione Operativa permette di lanciare gli ordini ai fornitori e di programmare l'attività di produzione in TEKNO dei semilavorati in modo che siano disponibili al momento giusto per l'assemblaggio nel Magazzino Semilavorati.

L'intera funzionalità del network è garantita da un contratto d'assistenza 24 h al giorno

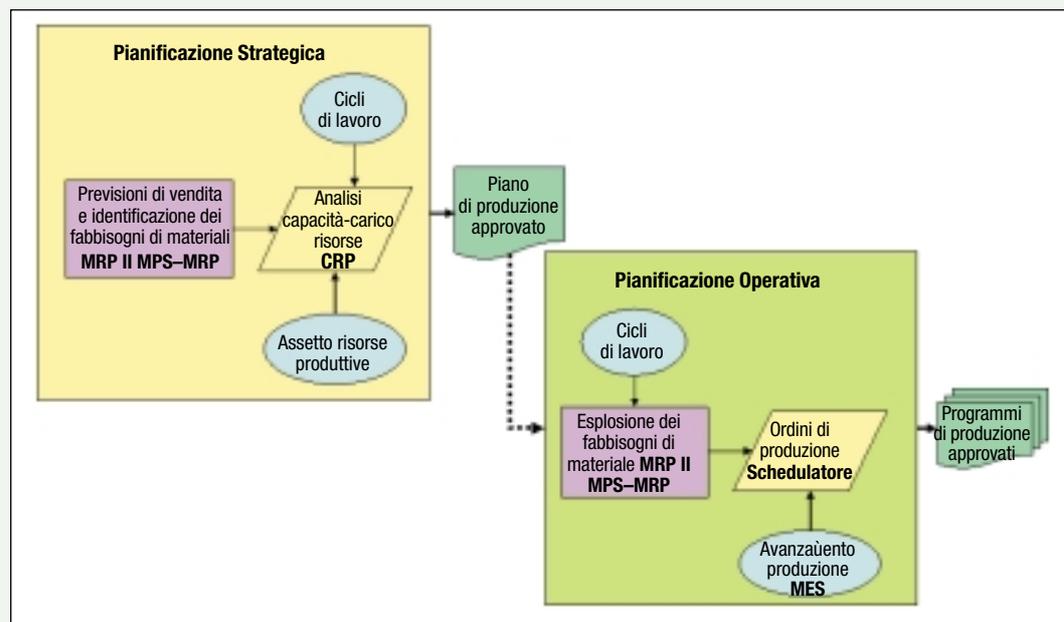


FIGURA 5
Flusso
pianificazione della
produzione TEKNO

Prestazione	Obiettivo	Risultato	Note
Tempo medio di attraversamento (gg), considerato come indicatore di flessibilità	- 10 gg	- 12 gg	Prima dell'introduzione dell'innovazione il ritardo medio di consegna era pari a 6 giorni; per tale motivo si è stabilito di ridurre il Lead Time di attraversamento (unica causa di ritardi) di almeno 10 giorni per azzerare il ritardo, aumentare la flessibilità ed essere in grado di rispondere ad ulteriori riduzioni.
Ritardo medio di consegna (gg)	- 6 gg	- 6 gg	A fronte della riduzione dei Lead Time di attraversamento, TEKNO è riuscita a rispettare le date di consegne portando il ritardo medio prossimo a zero.
Riduzione Scorte, considerato anche come indicatore di costo	Quanto possibile	- 20%	Il vecchio modo di operare di TEKNO aveva comportato un forte incremento del livello delle scorte di magazzino; grazie all'innovazione nella pianificazione e all'introduzione del Magazzino Semilavorati, è stato possibile ridurre in modo sostanziale le scorte dei semilavorati grezzi e dei componenti commerciali.
Resi cliente, considerato anche come indicatore di qualità	3-2,5/1000	2/1000	Complessivamente è cresciuta la qualità dei prodotti finiti come testimonia la riduzione dei resi da parte dei clienti.

TABELLA 1
Risultati

per 365 giorni l'anno; i costi di gestione del sistema informativo aziendale si possono considerare trascurabili non solo rispetto al fatturato ma anche rispetto ai costi di produzione dei semilavorati e a quelli di assemblaggio del prodotto finito per il cliente: l'incidenza dei costi di gestione delle tecnologie ICT, calcolando il contributo sia dei costi esterni che di quelli interni, non supera rispettivamente lo 0,17% del fatturato nel 2003 e lo 0,13% nel 2004.

5. RISULTATI DELLA INNOVAZIONE INTRODOTTA, ECONOMICI E GESTIONALI

La tabella 1 indica i risultati raggiunti. Dall'osservazione della tabella risulta evidente come si siano raggiunti tutti gli obiettivi specifici della innovazione di processo. Un risultato assai importante si riscontra poi analizzando il fatturato pro capite (Figura 6); questo indicatore cresce con regolarità fino al 2002, si mantiene stabile nel 2003, quando il

sistema integrato ERP+ASP non è ancora in funzione, mentre si incrementa decisamente nel 2004 con l'entrata in funzione del nuovo sistema di pianificazione e gestione della produzione. Dati alla mano, TEKNO è riuscita ad aumentare la produzione del 27% contro un incremento degli addetti del 8%.

Infine una nota sugli investimenti: l'analisi dell'incidenza degli investimenti in ICT sul totale è pari all'8% nel 2001, al 21% nel 2002, mentre scende al 6% nel 2003 e praticamente sparisce (2%) nel 2004; in quattro anni l'incidenza degli investimenti in ICT è stata pari al 10,5% degli investimenti totali, per un valore assoluto di 345.000 €. Tutto sommato poca cosa!!!

Tuttavia l'efficacia dell'investimento in ICT è stata molto rilevante dal punto di vista sia operativo che della capacità di sostenere il business (Figura 7).

6. CONCLUSIONI

Il caso TEKNO è molto significativo dal punto di vista dell'importanza che può assumere un investimento in ICT, anche se di ammontare relativamente limitato, rispetto all'efficacia del sostegno al business attuale e futuro; ci sono tre elementi in particolare che vanno rimarcati:

1. prima viene il prodotto: se non ci fosse stata l'introduzione della nuova linea di motoriduttori, non ci sarebbe stata una crescita così significativa nel business; durante il periodo di lancio sul mercato della nuova linea di prodotti non c'è stato né tempo né spazio per dedicarsi alla ottimizzazione dei processi; tale situazione va considerata assolutamente fisiologica (cioè normale) come dimostra l'analisi di centinaia di casi di innovazione di prodotto;

2. solo quando la fase di innovazione di prodotto e il rapporto con i nuovi clienti si è stabilizzato, si possono affrontare i problemi di razionalizzazione produttiva; la razionalizzazione produttiva richiede un insieme di interventi specifici che toccano i processi operativi, le tecnologie produttive, gli spazi di lavoro e le risorse manageriali e professionali critiche e, naturalmente, anche le tecnologie di gestione, come sono i sistemi ERP e APS; la certificazione della qualità nel caso TEKNO è stata utilizzata correttamente in modo sostanziale per attivare una profonda revisione dei processi produttivi;

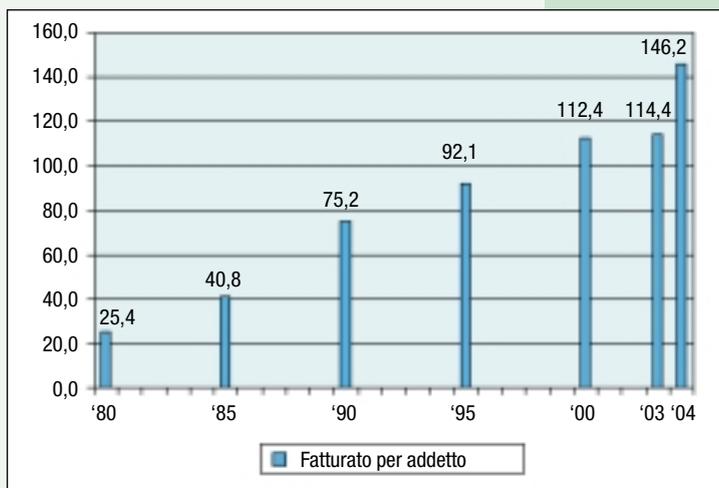


FIGURA 6
Andamento del fatturato per addetto TEKNO

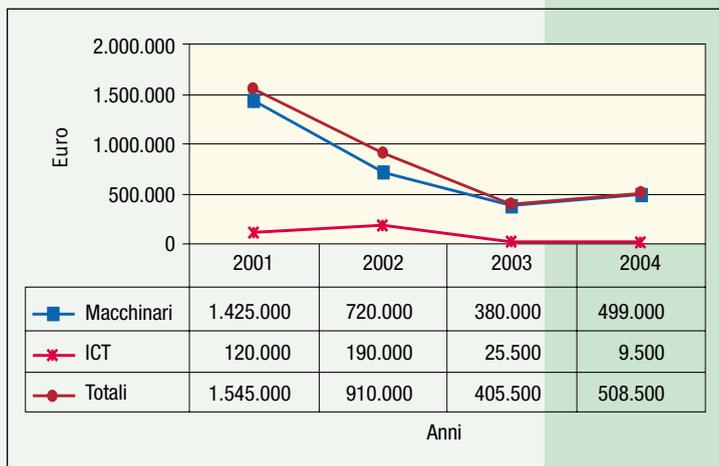


FIGURA 7
Andamento investimenti in TEKNO

3. le tecnologie di gestione ICT svolgono un ruolo molto importante quando vanno a toccare alcuni processi critici che sono poi quelli che permettono di pianificare e controllare l'operatività; nel caso TEKNO in particolare questo processo critico viene individuato nel processo di pianificazione della produzione, in cui con l'uso della tecnologia diventa possibile integrare e mantenere l'allineamento fra l'assunzione dell'impegno nei confronti dei clienti da una parte, in termini di volumi di fornitura e tempi di consegna, e la gestione degli ordini ai fornitori e il lancio e l'avanzamento delle commesse nella fase operativa della produzione dall'altra; il sistema è sufficientemente potente per reggere adeguatamente anche nei casi in cui sia commercialmente op-

portuno, da parte di TEKNO, soddisfare il cliente che, dopo avere accettato i tempi di consegna proposti da TEKNO in fase di ordinativo della commessa, richiede successivamente ulteriori variazioni nel mix dei prodotti ordinati e nei tempi di consegna.

La domanda finale dell'informatico sul contributo alla innovazione di processo introdotta in azienda è la seguente: "senza l'introduzione

del sistema di supporto alla pianificazione e gestione della produzione, si sarebbe potuto ottenere un incremento di redditività pari al 27,8% del fatturato procapite fra il 2003 e il 2004?" La risposta è molto semplice: "No"; ma attenzione, non ci sarebbe neanche senza l'incremento del fatturato del 37,8%, dovuto alla nuova impostazione strategica sviluppata dagli imprenditori!!

ROBERTO BELLINI è docente di Gestione e Marketing dell'Innovazione 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 Osservatorio delle Competenze e delle Certificazioni Informatiche, sviluppato con Federcomin e Fondazione Politecnico.
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.
francala@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.



L'archiviazione ottica dei documenti

1. INTRODUZIONE

Appare evidente come l'innovazione tecnologica applicata alla gestione dei flussi documentali possa comportare una serie di notevoli vantaggi per imprese, professionisti e pubbliche amministrazioni.

In particolare il processo di archiviazione e conservazione del *workflow* in forma digitale permette di garantire nel tempo l'integrità e la reperibilità dei documenti, consentendo in ultima istanza una maggiore efficacia, nonché considerevoli risparmi di spesa.

Se, inoltre, si aggiunge la possibilità di condividere le medesime informazioni in tempo reale e a distanza, oppure di permettere l'accesso esclusivamente alle persone autorizzate (in ottemperanza alla normativa sul trattamento dei dati personali), ben si comprende la tendenza - soprattutto in molte realtà aziendali e pubbliche amministrazioni - di ridurre progressivamente l'impiego del supporto cartaceo a beneficio degli strumenti informatici.

Al fine di portare a compimento questa rivoluzione digitale, vale a dire per poter eliminare la carta, è necessario garantire al trasferimento degli archivi cartacei su supporti informatici un valore giuridico che soddisfi i vigenti obblighi normativi inerenti alla conservazione documentale (si pensi per esempio a quanto previsto in materia fiscale).

Per tale ragione, a partire dai primi anni '90, non sono mancati gli interventi legislativi che, regolamentando i procedimenti di archiviazione e conservazione dei documenti informatici, hanno sgombrato il campo dagli ostacoli di na-

tura legale alla diffusione su larga scala di tali tecnologie [riquadro 1].

2. IL QUADRO NORMATIVO

Sotto il profilo normativo, premesso che i concetti richiamati si fondano sul riconoscimento del documento informatico e della sottoscrizione elettronica (cfr. questa rubrica nel numero di Marzo 2005 di Mondo Digitale), non meraviglia che la riproduzione e la conservazione di documenti siano attualmente trattate nello stesso D.P.R. 445/00 (Testo Unico in materia di documentazione amministrativa) che disciplina la firma digitale.

Già a partire dal 1993 la legge n. 537 aveva previsto che gli obblighi di conservazione e di esibizione dei documenti per finalità amministrative e probatorie potessero essere soddisfatti se realizzati mediante supporto ottico, purché le procedure adottate fossero conformi a regole tecniche dettate dall'*Autorità per l'Informatica nella Pubblica Amministrazione* (AIPA, ora sostituita dal *Centro Nazionale per l'Informatica*

Riquadro 1

Vantaggi dell'archiviazione elettronica

- Riduzione volumetrica
- Facilità e rapidità di consultazione
- Risparmio sui costi vivi
- Conservazione dei documenti
- Condivisione dei documenti
- Accesso riservato ai documenti
- Gestione integrata (Protocollo informatico, gestionali, fax, posta elettronica)

nella *Pubblica Amministrazione* – CNIPA) che si era già adeguata con due deliberazioni, la n. 15/94 e la n. 24/98.

In seguito, la materia è stata aggiornata e riordinata dal legislatore con il Testo Unico del 28 dicembre 2000 n. 445, che all'art. 6 attribuisce alle pubbliche amministrazioni e ai privati la facoltà di sostituire i documenti dei propri archivi, nonché le scritture contabili, la corrispondenza e tutti gli altri atti di cui è prescritta la conservazione, con la loro riproduzione su supporto fotografico oppure ottico o con altro mezzo idoneo a garantire la conformità dei documenti agli originali.

La norma in questione demanda ancora una volta all'Autorità per l'informatica nella pubblica amministrazione il compito di stabilire le regole tecniche per l'archiviazione su supporto ottico, dettate quindi dall'AIPA con la deliberazione n. 42 del 13 dicembre 2001.

Tale provvedimento è stato successivamente sostituito dalla deliberazione del CNIPA n. 11 del 19 febbraio 2004 con la precipua finalità di adeguare le disposizioni alle esigenze dettate dall'evoluzione tecnologica.

3. ARCHIVIAZIONE ELETTRONICA E CONSERVAZIONE SOSTITUTIVA

In primo luogo appare opportuno distinguere il processo di conservazione sostitutiva da quello dell'archiviazione elettronica, che può essere propedeutico al primo, ma non è obbligatorio. Per la semplice archiviazione elettronica dei documenti non sono previste particolari modalità operative: l'adempimento viene lasciato all'iniziativa del soggetto interessato, il quale potrà utilizzare qualsiasi tipo di supporto di memorizzazione, stabilendo un metodo per l'acquisizione del documento, la sua classificazione e l'attribuzione di un codice di identificazione univoco, in modo da consentirne un accesso facile e razionale.

Il processo di conservazione sostitutiva (solitamente successivo all'eventuale archiviazione elettronica) risulta finalizzato a rendere un documento non deteriorabile e, quindi, disponibile nel tempo in tutta la sua integrità e autenticità. A tale scopo le regole tecniche suddividono i documenti in due grandi categorie: i documenti informatici (per esempio i file non solo di testo, ma anche video e audio) e i documenti analogici (carta, cassette VHS, altri nastri magnetici ecc.).

Il processo di conservazione delle due categorie è sostanzialmente identico, tranne per la necessità di digitalizzare il documento analogico: in buona sostanza, per quanto concerne un documento cartaceo, si tratta di ricavarne l'immagine utilizzando uno scanner (e possibilmente un software OCR¹): il CNIPA non impone un particolare formato, tuttavia si consiglia il PDF per la sua diffusione universale combinata con la capacità di preservare tutte le font, la formattazione, i colori e le immagini, indipendentemente dall'applicazione e dalla piattaforma utilizzate. La deliberazione autorizza l'impiego di qualsiasi tipo di supporto di memorizzazione che consenta la registrazione mediante la tecnologia laser, vale a dire dischi ottici, magneto-ottici, CD-R e DVD.

Inoltre, tenuto conto di quanto previsto nel già citato Testo Unico circa la possibilità d'impiego ai fini della conservazione di altro mezzo idoneo a garantire la conformità dei documenti agli originali e in considerazione dell'evoluzione tecnologica nel frattempo avvenuta, la deliberazione consente di utilizzare anche altri supporti di memorizzazione (oltre quelli a tecnologia laser), sempre che non ostino particolari motivazioni e, comunque, nel rispetto delle stesse regole tecniche.

La conformità dell'immagine trasposta su supporto di memorizzazione rispetto al documento d'origine dev'essere garantita dal "responsabile della conservazione" mediante l'apposizione della propria firma digitale.

La conservazione sostitutiva di documenti cartacei da un lato richiede il mantenimento dell'ordine "fisico" (per esempio cronologico), dall'altro rende possibili molteplici aggregazioni diverse (per esempio oggetto, cliente, prestazione ecc.); appare, quindi, necessaria in via preliminare un'operazione di "cartellinatura" che consiste nella preparazione dei documenti cartacei, già raccolti fisicamente in un fascicolo, attraverso procedure definite dall'utente del sistema. Successivamente vengono approntati strumenti di consultazione ed eventuali codificazioni aggiuntive in modo da rilevare collegamenti fra gli atti e i documenti riprodotti (per campi di aggregazione).

¹ La sigla OCR sta per *Optical Character Recognition* (riconoscimento automatico dei caratteri) e indica i programmi in grado di convertire l'immagine digitalizzata di un documento in testo.

Gli art. 3 e 4 della deliberazione CNIPA prevedono che il processo di conservazione si realizzi attraverso la memorizzazione dei documenti ed eventualmente anche delle loro impronte² sul supporto ottico, con le formalità dell'apposizione del riferimento temporale e, infine, della firma digitale sull'insieme di documenti destinati alla conservazione (ovvero, su un'evidenza informatica contenente una o più impronte dei documenti).

L'apposizione del riferimento temporale prima della firma digitale trova motivazione nella necessità di fissare ad una data certa il contenuto dei documenti memorizzati e conservati, realizzando il cosiddetto "file di chiusura".

Per i soli documenti informatici originati da "documenti analogici originali unici" è prevista anche l'apposizione del riferimento temporale e della firma digitale da parte di un pubblico ufficiale (il notaio, ovvero il dirigente competente nella pubblica amministrazione), il quale attesti la conformità di quanto conservato al documento d'origine.

Solamente al termine del processo di archiviazione sostitutiva appena illustrato, il documento analogico d'origine del quale sia obbligatoria la conservazione potrà essere distrutto [riquadro 2].

4. IL RESPONSABILE DEL PROCEDIMENTO

Secondo quanto stabilito dalla deliberazione CNIPA il responsabile della conservazione:

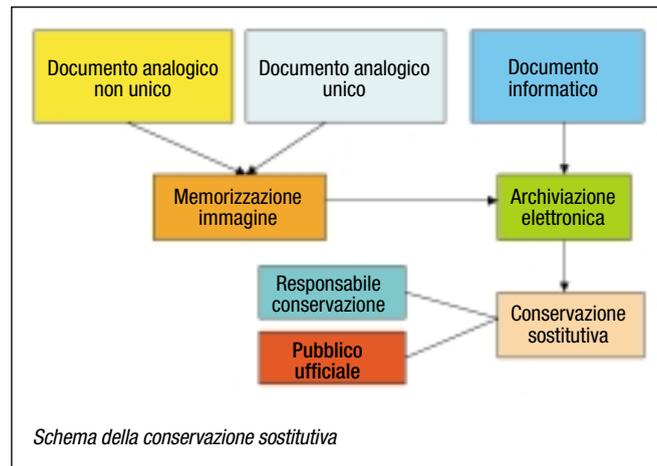
- a. definisce le caratteristiche e i requisiti del sistema di conservazione in funzione della tipologia dei documenti (analogici o informatici) da conservare, della quale tiene evidenza; organizza di conseguenza il contenuto dei supporti ottici e gestisce le procedure di sicurezza e di tracciabilità che ne garantiscono la corretta conservazione, anche per consentire l'esibizione di ciascun documento conservato;
- b. archivia e rende disponibili, con l'impiego di procedure elaborative, relativamente a ogni

² Per "impronta" si intende una sequenza di bit, di lunghezza predefinita, generata mediante l'applicazione di una funzione di *hash* tale che risulti di fatto impossibile, a partire dall'impronta risalire al file e, altresì, risulti di fatto impossibile determinare una coppia di sequenze di simboli binari per le quali la funzione generi impronte uguali.

Riquadro 2

Conservazione documento cartaceo

1. Memorizzazione dell'immagine (scanner);
 2. apposizione del riferimento temporale;
 3. firma digitale del responsabile;
 4. apposizione del riferimento temporale
 5. firma digitale del notaio
- } solo per gli originali unici



supporto di memorizzazione utilizzato, le seguenti informazioni:

1. descrizione del contenuto dell'insieme dei documenti;
2. estremi identificativi del responsabile della conservazione;
3. estremi identificativi delle persone eventualmente delegate dal responsabile della conservazione, con l'indicazione dei compiti alle stesse assegnati;
4. indicazione delle copie di sicurezza;
- c. mantiene e rende accessibile un archivio del software dei programmi in gestione nelle eventuali diverse versioni;
- d. verifica la corretta funzionalità del sistema e dei programmi in gestione;
- e. adotta le misure necessarie per la sicurezza fisica e logica del sistema preposto al processo di conservazione sostitutiva e delle copie di sicurezza dei supporti di memorizzazione;
- f. richiede la presenza del pubblico ufficiale nei casi in cui sia previsto il suo intervento, assicurando allo stesso l'assistenza e le risorse necessarie per l'espletamento delle sue attività;
- g. definisce e documenta le procedure di sicurezza da rispettare per l'apposizione del riferimento temporale;

0

h. verifica periodicamente, con cadenza non superiore a cinque anni, l'effettiva leggibilità dei documenti conservati provvedendo, se necessario, al riversamento diretto o sostitutivo del contenuto dei supporti.

Il responsabile del procedimento di conservazione sostitutiva può delegare, in tutto o in parte, lo svolgimento delle proprie attività a una o più persone che, per competenza ed esperienza, garantiscano la corretta esecuzione delle operazioni delegate.

Si tenga infine presente che, per espressa previsione della deliberazione, il procedimento di conservazione sostitutiva può essere affidato, anche solo parzialmente, ad altri soggetti (il cosiddetto *outsourcing*).

5. APPLICAZIONI FISCALI

Quasi contemporaneamente alla deliberazione del CNIPA, il 23 gennaio 2004 veniva emesso il decreto del Ministro dell'Economia e delle Finanze n. 27 avente a oggetto le modalità di assolvimento degli obblighi fiscali relativi ai documenti informatici e alla loro riproduzione in diversi tipi di supporto.

Fino a quel momento il Ministero aveva tollerato l'archiviazione elettronica esclusivamente a patto che fossero comunque conservati i documenti cartacei. Il D.M. 23 gennaio 2004 estende finalmente l'applicazione delle regole sulla conservazione sostitutiva sopra richiamate anche alle scritture contabili e ai documenti informatici rilevanti ai fini delle disposizioni tributarie.

Questi ultimi devono avere la forma di documenti statici non modificabili, ossia il loro contenuto dev'essere immutabile nel tempo e non alterabile durante le fasi di accesso e di conservazione; a tal fine il documento informatico non deve contenere macroistruzioni o codice eseguibile, tali da attivare funzionalità che possano modificare i dati rappresentati.

Anche in questo caso viene richiesta l'apposizione del riferimento temporale e della sottoscrizione elettronica, al fine di garantirne l'attestazione della data, l'autenticità e l'integrità dell'atto.

È ammessa la memorizzazione su qualsiasi supporto di cui sia garantita la leggibilità nel tempo, purché sia assicurato l'ordine cronologico e non vi sia soluzione di continuità per ciascun periodo d'imposta; inoltre, devono essere consentite le funzioni di ricerca e di estrazione

delle informazioni dagli archivi informatici in relazione al cognome, al nome, alla denominazione, al codice fiscale, alla partita Iva, alla data o associazioni logiche di questi ultimi.

Per quanto concerne l'esibizione in caso di verifiche, controlli o ispezioni, il documento viene reso leggibile a video e, a richiesta, disponibile su supporto cartaceo e informatico presso il luogo di conservazione delle scritture contabili; inoltre viene riconosciuta anche l'esibizione per via telematica secondo le modalità stabilite dalle competenti Agenzie fiscali.

Il processo di conservazione dev'essere effettuato con cadenza almeno quindicinale per le fatture e annuale per i restanti documenti di rilievo fiscale.

Per quanto concerne l'assolvimento dell'imposta di bollo sui documenti informatici, è stato previsto il versamento presso banche e poste utilizzando il modello F23.

Ulteriore vincolo previsto dal decreto ministeriale consiste nell'obbligo di trasmettere (entro il mese successivo alla scadenza dei termini per la presentazione della dichiarazione dei redditi, dell'IRAP e dell'IVA), alle competenti Agenzie fiscali, l'impronta dell'archivio informatico oggetto della conservazione, la relativa sottoscrizione elettronica e la marca temporale.

6. FATTURA ELETTRONICA

Il 28 febbraio 2004 è stato pubblicato in Gazzetta Ufficiale il D. Lgs. n. 52 in recepimento della direttiva comunitaria 2001/115/CE. Il decreto modifica e armonizza le modalità di fatturazione in materia di IVA, regolando, tra l'altro, in modo specifico la trasmissione delle fatture per via elettronica e la loro conservazione.

Per la prima volta nel nostro ordinamento (ma anche negli altri Paesi dell'Unione Europea) viene ammessa la possibilità di gestire un documento contabile in forma integralmente immateriale, senza la necessità di procedere alla sua stampa.

La possibilità di emettere la fattura elettronica è subordinata all'osservanza delle seguenti condizioni:

a. il documento non deve contenere macroistruzioni, né codice eseguibile: in tale modo il suo contenuto non viene alterato né durante le fasi di accesso, né durante la conservazione dello stesso;

b. è necessaria l'apposizione del riferimento temporale, ossia di un'informazione contenente la data e l'ora di trasmissione del documento;

c. è richiesto il consenso del destinatario della fattura;

d. devono essere garantiti l'autenticità dell'origine del documento e l'integrità del contenuto con l'apposizione della firma elettronica qualificata dell'emittente; in alternativa possono essere impiegati, in luogo della firma elettronica, gli idonei sistemi EDI (*Electronic Data Interchange*) di trasmissione basati sullo scambio elettronico di dati.

La trasmissione di una fattura elettronica vincola colui che la riceve ad effettuare l'archiviazione nella medesima forma, senza la necessità

di stamparla su carta; in mancanza del consenso l'emittente può comunque procedere alla formazione elettronica della fattura che, però, dovrà essere consegnata o spedita unicamente in forma cartacea.

Il soggetto ricevente è tenuto ad assicurare - per finalità di controllo - l'accesso automatizzato all'archivio e a consentire che tutti i documenti e i dati in esso contenuti siano stampabili e trasferibili su altro supporto informatico.

Le disposizioni del d.lgs. 52/04 andranno a integrarsi con le regole in materia di archiviazione sostitutiva sopra menzionate, come fossero tasselli del medesimo mosaico nel quale l'innovazione passa attraverso la digitalizzazione delle informazioni.

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.

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.

david.dagostini@adriacom.it



DENTRO LA SCATOLA

Rubrica a cura di

Fabio A. Schreiber

Il Consiglio Scientifico della rivista ha pensato di attuare un'iniziativa culturalmente utile presentando in ogni numero di Mondo Digitale un argomento fondante per l'Informatica e le sue applicazioni; in tal modo, anche il lettore curioso, ma frettoloso, potrà rendersi conto di che cosa sta "dentro la scatola". È infatti diffusa la sensazione che lo sviluppo formidabile assunto dal settore e di conseguenza il grande numero di persone di diverse estrazioni culturali che - a vario titolo - si occupano dei calcolatori elettronici e del loro mondo, abbiano nascosto dietro una cortina di nebbia i concetti basilari che lo hanno reso possibile.

Il tema scelto per il 2004 è stato: "**Perché gli anglofoni lo chiamano computer**, ovvero: **introduzione alle aritmetiche digitali**". Per il 2005 il filo conduttore della serie sarà: "**Ma ce la farà veramente?**, ovvero: **introduzione alla complessità computazionale e alla indecidibilità**" e il suo intento è di guidare il lettore attraverso gli argomenti fondanti dell'Informatica e alle loro implicazioni pratiche e filosofiche. La realizzazione degli articoli è affidata ad autori che uniscono una grande autorevolezza scientifica e professionale a una notevole capacità divulgativa.

Potenza e limiti del calcolo automatico: la complessità di calcolo degli algoritmi

Angelo Morzenti

1. INTRODUZIONE

Questo è il terzo articolo di una serie di quattro destinata a illustrare gli aspetti più concettuali e teorici dell'informatica. I primi due si sono focalizzati sulla nozione di algoritmo come formalizzazione del concetto di calcolo automatico e sulla possibilità di risolvere problemi in modo algoritmico. Qui l'attenzione è rivolta esclusivamente ai problemi risolvibili algoritmicamente, per valutare l'efficienza degli algoritmi, intesa come stima della loro capacità di fare uso in maniera ottimale delle risorse di calcolo. Nel fare ciò, cercheremo di essere precisi, adottando metodi quantitativi di valutazione basati su strumenti matematici e di prescindere da dettagli inessenziali, mettendo a fuoco gli aspetti più concettuali degli algoritmi e dei problemi da essi risolti.

Illustreremo concetti e metodi relativi a due dei problemi più rilevanti e maggiormente studiati: quello di ricercare un elemento in un insieme dato e quello di ordinare un insieme di elementi.

2. OGGETTO E METODI DELL'ANALISI DI COMPLESSITÀ

Avendo definito l'efficienza di un algoritmo come la sua capacità di fare un uso il più possibile limitato delle risorse di calcolo, la prima domanda da porsi è: di quali risorse siamo interessati a valutare l'utilizzo. La risposta non è univoca e dipende dallo scopo e dal contesto in cui l'analisi viene svolta. Tra i fattori da considerare ci sono, per esempio, il tempo impiegato per completare l'esecuzione dell'algoritmo, la quantità di memoria, centrale o esterna, richiesta per memorizzare dati e risultati, i dispositivi di interfaccia utilizzati, i canali di comunicazione e le relative larghezze di banda nel caso che il calcolatore comunichi con altri a esso collegati. Nel breve spazio di questo articolo ci concentreremo sulla valutazione del tempo di calcolo, che assume grande rilevanza sia pratica che concettuale. Nel seguito, quando parleremo di "costo" o di "complessità di calcolo" di un'operazione o dell'esecuzione di un algoritmo, intenderemo perciò riferirci al tempo impiegato. L'analisi dello spazio di memoria utilizzato da un algoritmo diventa rilevante per gli algoritmi

che utilizzano complesse strutture dati, che in questo ambito per semplicità e brevità non consideriamo.

Come valutare il tempo di calcolo di un algoritmo? Un semplice e immediato approccio di tipo empirico porterebbe a codificarlo in un linguaggio di programmazione, effettuare una serie di esecuzioni su un determinato calcolatore fornendogli valori di ingresso opportunamente selezionati, cronometrare (tramite opportuni strumenti hardware e software) i tempi di esecuzione per poi compilare tabelle, calcolare statistiche. Questo approccio, se pur del tutto sensato e corretto in certi contesti (tipicamente quando si vogliono effettuare delle operazioni di *benchmarking* per calcolatori), è insoddisfacente ai nostri scopi, perché il tempo di esecuzione di un algoritmo su un determinato calcolatore dipende da svariati fattori di natura contingente che fanno perdere generalità ai risultati ottenuti, per esempio dalla velocità della CPU, delle interfacce, dei bus e delle periferiche, dalla qualità del codice generato dal compilatore (se l'algoritmo è codificato in un linguaggio di alto livello), dalle eventuali interferenze con altri programmi in corso di esecuzione (nel caso che il calcolatore sia multiprogrammato). In realtà noi siamo interessati a studiare gli algoritmi in sé, prescindendo il più possibile dagli aspetti relativi alla piattaforma di calcolo sulla quale vengono eseguiti. Il primo passo per rendere le nostre valutazioni indipendenti dalla piattaforma è quello di assumere un modello di macchina calcolatrice programmabile il più semplice possibile, al quale tutte le altre (o un sottoinsieme molto significativo) siano riconducibili e adottare un linguaggio per codificare gli algoritmi similmente universale e rappresentativo. Per evidenti ragioni di spazio non possiamo dare una definizione precisa e rigorosa della piattaforma di calcolo e del linguaggio di programmazione cui faremo riferimento nel seguito: facciamo appello all'esperienza e all'intuito del lettore, e lo rimandiamo ai testi citati in bibliografia.

Assumiamo un classico modello di macchina RAM (*Random Access Machine*), cioè di una macchina:

- che può accedere con un costo uniforme a tutte le celle della sua memoria;
- le cui istruzioni vengono eseguite una alla volta senza alcun grado di parallelismo;
- che possiede nel suo repertorio le usuali

operazioni aritmetiche (+, -, *, /, arrotondamenti a interi), logiche (connettivi preposizionali di congiunzione, disgiunzione, negazione) e istruzioni per spostare dati dalla memoria alle unità di elaborazione e viceversa;

- che memorizza nelle celle della memoria numeri interi o con parte decimale, caratteri, simboli rappresentanti elementi di insiemi finiti;
- che NON possiede operazioni dissimili da quelle elementari di un comune calcolatore (per esempio, supponiamo di *non* poter calcolare con una singola operazione un'arbitraria potenza di un numero o la versione ordinata di un insieme di elementi);
- che NON fa uso di una gestione gerarchica della memoria, utilizzata nei calcolatori moderni per ottimizzare le prestazioni (memorie cache, memoria virtuale): tali sistemi ne rendono il comportamento assai complesso, difficile da analizzare e poco predicibile.

Il lettore ha sicuramente notato la grande differenza tra la macchina RAM e il modello della macchina di Turing adottato nei precedenti articoli per formalizzare la nozione di algoritmo. La macchina RAM è equivalente alla macchina di Turing per quanto attiene alle questioni di calcolabilità; per gli aspetti riguardanti la complessità di calcolo degli algoritmi la macchina RAM è molto più vicina ai calcolatori elettronici impiegati nella realtà per l'elaborazione automatica dell'informazione.

Il linguaggio in cui esprimiamo gli algoritmi ha le seguenti caratteristiche:

- è ispirato ai più comuni linguaggi di programmazione (C, Pascal, Java...) e contiene le istruzioni della programmazione strutturata (istruzioni iterative e condizionali); nelle istruzioni composte usiamo l'incolonnamento per indicare le sequenze di sotto-istruzioni; l'operatore di assegnamento viene indicato con una freccia verso sinistra " \leftarrow " (quando l'istruzione $v \leftarrow e$ viene eseguita la variabile v assume il valore dell'espressione e); i commenti iniziano con una doppia sbarra "//" e terminano alla fine della riga;
- gli algoritmi sono scritti sotto forma di procedure con parametri, l'istruzione *restituisci* può essere usata per terminare l'esecuzione restituendo esplicitamente un risultato;
- le variabili rappresentano un dato semplice in una cella di memoria, con l'unica eccezione dei vettori, che rappresentano un insieme di dati elementari memorizzati in celle consecutive: **A**[1..n]

indica un vettore di n elementi, $A[i]$ indica l' i -esimo elemento, e $\text{lung}(A)$ la dimensione del vettore. Come primo esempio su cui illustrare il metodo di analisi della complessità di calcolo degli algoritmi, consideriamo il problema della ricerca di un elemento in un insieme prefissato, memorizzato in un vettore, e l'algoritmo della cosiddetta ricerca sequenziale che confronta in successione, dal primo all'ultimo, gli elementi del vettore con quello ricercato, fino a quando lo trova (e allora restituisce la posizione nel vettore) oppure oltrepassa la fine del vettore (allora restituisce 0 in segno di insuccesso della ricerca). Il riquadro 1 riporta l'algoritmo; una prima colonna separata indica, mediante costanti simboliche, il costo attribuito a ogni singola operazione e una seconda colonna il numero di volte in cui ogni operazione viene ripetuta.

Riquadro 1

Algoritmo di ricerca sequenziale

Ricerca Sequenziale (A, elem)	Costo	# esecuzioni
$i \leftarrow 1$	c_1	1
fintantoche $i \leq \text{lung}(A)$ ripeti	c_2	K , con $1 \leq K \leq \text{lung}(A) + 1$
se $A[i] = \text{elem}$	c_3	H , con $1 \leq H \leq \text{lung}(A)$
allora restituisci i	c_4	0 oppure 1
altrimenti $i \leftarrow i + 1$	c_5	J , con $0 \leq J \leq \text{lung}(A)$
restituisci 0	c_6	0 oppure 1

Il tempo T di esecuzione dell'algoritmo è dato dalla somma dei tempi delle singole istruzioni, e si può esprimere diversamente a seconda che l'elemento cercato sia presente o assente. Nel primo caso, supponendo che l'elemento cercato si trovi in posizione P ,

$$T = c_1 + P \cdot (c_2 + c_3) + (P - 1) \cdot c_5 + c_4$$

mentre nel secondo (elemento non presente)

$$T = c_1 + (n + 1) \cdot c_2 + n \cdot (c_3 + c_5) + c_6$$

Il caso a costo minimo (chiamato caso ottimo) è quello in cui l'elemento cercato si trova in prima posizione: in questo caso $T = c_1 + c_2 + c_3 + c_4$; il caso in cui l'elemento non è presente è anche quello in cui viene eseguito il maggior numero di operazioni elementari, e per questo è detto caso pessimo.

Di solito il caso ottimo ha un interesse limitato, perché non è frequente né rappresentativo. Il ca-

so pessimo è invece considerato con grande attenzione quando si vuole adottare un approccio cautelativo e prudentiale tipico delle attività progettuali in cui, per esempio nell'ingegneria civile o meccanica, si utilizzano ipotesi di massimo carico per dimensionare i manufatti progettati. Ciò è particolarmente opportuno nella progettazione dei sistemi informatici che presentano delle criticità (si pensi ad apparecchiature medicali, o a sistemi di controllo dei mezzi di trasporto) in cui è essenziale poter contare sul fatto che una certa componente che esegue un algoritmo termini producendo il risultato richiesto entro un tempo prefissato. Per questi motivi nel seguito concentreremo le nostre valutazioni di complessità sul caso pessimo. Per i sistemi, meno critici, che hanno al più l'esigenza di massimizzare le prestazioni a fronte di un numero molto alto di esecuzioni con dati diversi (per esempio sistemi transazionali in ambito amministrativo), i casi ottimo e pessimo non sono rappresentativi delle prestazioni complessive, che vengono valutate con metodi di tipo statistico, a partire da appropriate ipotesi sulla distribuzione dei dati. La valutazione del caso medio è spesso più complessa e laboriosa di quelli ottimo e pessimo: per la ricerca sequenziale, supponendo che il dato ricercato sia presente e che con la stessa probabilità si trovi in una qualsiasi delle n posizioni nel vettore, il costo medio richiede di valutare la media di tutti i valori della posizione P , ottenendo perciò:

$$T = c_1 + \frac{n}{2} \cdot (c_2 + c_3) + \frac{n-1}{2} \cdot c_5 + c_4$$

Possiamo constatare, nei casi pessimo e medio appena esaminati, che la complessità temporale dipende da n , la dimensione dei dati in ingresso. Ciò è tipico di tutti i problemi non banali, perciò per un algoritmo che abbia dei dati in ingresso di dimensione n il tempo di esecuzione viene indicato, evidenziando la dipendenza dal parametro, come una funzione $T(n)$.

Inoltre nel considerare la funzione $T(n)$ si cerca di darne una caratterizzazione il più possibile sintetica e generale, che permetta di prescindere dai fattori relativi alla piattaforma di calcolo su cui viene eseguito l'algoritmo (e quindi dalle costanti c_1, \dots, c_6 nel nostro esempio) e dall'influenza che possono avere sull'andamento della funzione le parti marginali dell'algoritmo, per esempio quelle che vengono eseguite un numero fisso di volte. Si astrae da questi fattori

mediante lo strumento matematico della valutazione dell'andamento asintotico delle funzioni, che stima la loro velocità di crescita prescindendo dalle costanti di proporzionalità e considerando valori della variabile indipendente n sufficientemente elevati da escludere i termini di ordine inferiore. Per esempio, per la complessità della ricerca sequenziale si dice che $T(n)$ è $\Theta(n)$, intendendo che cresce allo stesso modo della funzione $f(n) = n$. In generale, quando la funzione $T(n)$ è composta da più termini e contiene costanti moltiplicative, si considera il termine dominante (quello che cresce più velocemente con la variabile n) e si ignorano le costanti moltiplicative; così facendo si ottiene la più semplice funzione $f(n)$ che ha lo stesso andamento asintotico di $T(n)$, e si dice che $T(n)$ è $\Theta(f(n))$.

La notazione Θ è un potente strumento di astrazione: due algoritmi A_1 e A_2 che hanno complessità $T_1(n)$ e $T_2(n)$ che siano entrambe $\Theta(f(n))$ per qualche funzione $f(n)$ sono considerate equivalenti per quanto riguarda la complessità temporale.

3. RICERCA E ORDINAMENTO

Ricordiamo ora che la nozione di algoritmo è ben distinta da quella di problema: per ogni problema si possono trovare più algoritmi che lo risolvono e che differiscono per vari aspetti, tra cui la complessità di calcolo. Illustriamo questo fatto su due problemi classici dell'informatica, tra i meglio studiati a causa della loro grande rilevanza teorica e applicativa: quello della ricerca (per il quale si è già visto l'algoritmo di ricerca sequenziale) e quello dell'ordinamento (che, dato un insieme di elementi disposti in sequenza in un ordine qualsiasi, consiste nel disporli in senso crescente). È istruttivo considerare un algoritmo alternativo alla ricerca sequenziale che, sotto l'ipotesi che l'insieme degli elementi in cui effettuare la ricerca siano disposti in ordine crescente, permette di svolgere la ricerca in modo molto più efficiente. Si tratta della ricerca binaria, codificata nella procedura mostrata nel riquadro 2. La ricerca viene svolta con riferimento a un segmento di vettore individuato dai due indici *inf* e *sup*; ad ogni iterazione l'elemento ricercato *elem* viene confrontato con quello presente nel punto medio del segmento considerato: se è uguale la ri-

Riquadro 2

Algoritmo di ricerca binaria

Ricerca Binaria (A , elem)

inf \leftarrow 1

sup \leftarrow lung(A)

med \leftarrow (inf + sup) / 2

fintantoche (inf \leq sup) e (elem \neq $A[\text{med}]$) ripeti

se elem $>$ $A[i]$

allora inf \leftarrow med + 1 //scarta la prima metà del segmento di vettore

altrimenti sup \leftarrow med - 1 //scarta la seconda metà del segmento

//di vettore

se inf \leq sup //se inf \leq sup allora elem = $A[\text{med}]$

allora restituisci med

altrimenti restituisci 0

cerca termina con successo, se è maggiore, la ricerca può proseguire nella seconda parte del segmento (perché essendo il vettore ordinato, *elem* è maggiore anche di tutti gli elementi della prima parte del segmento), altrimenti, prosegue nella prima metà.

Valutiamo questo algoritmo in modo informale e sintetico, tenendo presente che siamo interessati alla complessità asintotica nel caso pessimo. Basta quindi contare il numero di volte in cui viene ripetuto il corpo del ciclo: poiché il numero degli elementi del segmento sotto esame viene dimezzato a ogni ripetizione, il numero massimo di ripetizioni è pari al numero di volte per cui occorre dividere un numero n per 2 per arrivare a 0, e questo valore è strettamente legato al logaritmo in base 2 di n , quindi $T(n)$ è $\Theta(\log n)$. È noto che la funzione $\log n$ cresce molto lentamente al crescere di n (aumenta di 1 al raddoppiare di n ; per n pari a circa un miliardo $\log n$ vale circa 30, per n pari a circa mille miliardi vale circa 40...). Effettuare una ricerca in un insieme di dimensioni rilevanti è molto più efficiente se l'insieme è ordinato, come ben sanno tutti i genitori che cercano di (far) tenere in ordine la camera dei propri figli, e tutti coloro che hanno mai provato a immaginare quanto sarebbe utile una guida telefonica con gli abbonati riportati in un ordine casuale.

L'algoritmo di ricerca binaria ci dà l'opportunità di constatare un fatto che vale per tutti gli algoritmi di ricerca e anche per quelli di ordinamento: la valutazione della complessità asintotica può essere svolta conteggiando il numero di confronti tra elementi, trascurando le altre operazioni (valutazione e assegnamento di variabili, gestione delle istruzioni iterative...). La ragio-

ne di ciò è duplice. In primo luogo, nel caso in cui gli elementi da confrontare abbiano una complessa struttura (e.g., siano numeri elevati, polinomi, vettori di numeri o stringhe) allora l'operazione di confronto diventa più onerosa e il suo costo prevalente rispetto a quello delle altre; la seconda, e più importante ragione, particolarmente evidente nel caso della ricerca binaria, è che il numero delle altre operazioni è proporzionale al numero dei confronti e quindi la loro inclusione nel conteggio non porterebbe a risultati diversi per quanto riguarda la valutazione asintotica della complessità.

Non c'è bisogno di enfatizzare la rilevanza dal punto di vista pratico dell'operazione di ricerca di un elemento in un insieme: si pensi anche solo alla miriade di applicazioni informatiche di tipo amministrativo e gestionale che ne fanno uso. Alla luce del relevantissimo miglioramento nell'efficienza ottenibile dalla ricerca binaria rispetto a quella sequenziale grazie all'ipotesi di disposizione ordinata degli elementi nel vettore **A**, l'operazione di ordinamento assume anch'essa grande importanza e infatti questo problema è stato studiato a fondo, portando alla formulazione di un gran numero di algoritmi; ne analizziamo ora due tra i più significativi: l'ordinamento per inserzione e quello per fusione.

L'operato dell'ordinamento per inserzione, riportato nel riquadro 3, è simile a quello di un giocatore che, all'inizio di una partita, raccoglie le carte dal banco e le dispone ordinatamente in un mazzo che tiene in mano. Se ha già raccolto il 3, il 4 e il 7 e li tiene in mano in questo ordine, e poi raccoglie il 6, lo infila tra il 4 e il 7.

L'istruzione "per tutti *i*" scandisce gli elemen-

ti del vettore dal secondo all'ultimo e l'istruzione *shintantoché*, per ogni elemento in posizione *j*, con $2 \leq j \leq \text{lung}(A)$, lo porta nella posizione che gli compete nel segmento iniziale di vettore, lungo $j-1$, già ordinata nelle passate precedenti. A ogni ripetizione del ciclo *shintantoché*, la lunghezza del segmento iniziale ordinato si allunga di 1, quindi alla fine del processo tutti gli elementi sono disposti in ordine crescente. Lo sforzo complessivo per portare l'elemento $A[j]$ nella posizione giusta dipende dal suo valore relativamente a quello degli elementi che occupano le posizioni precedenti da 1 a $j-1$: se $A[j]$ è minore di tutti lo sforzo è massimo, proporzionale a j (il valore P_j è pari a j , perché occorre confrontare $a[j]$ con tutti gli elementi che lo precedono); all'estremo opposto, se $A[j]$ è maggiore di tutti gli $A[i]$, $1 \leq i < j$, poiché questi sono ordinati in senso crescente basterà confrontare $A[j]$ con $A[j-1]$ per concludere che $A[j]$ si trova già in posizione corretta e non deve essere spostato. Abbiamo quindi individuato il caso ottimo e quello pessimo. Il caso ottimo è quello in cui il vettore è già ordinato, il parametro P_j nella sommatoria vale 1 e $T(n)$ è $\Theta(n)$, cioè la funzione di complessità è lineare. Il caso pessimo è quello con $A[j]$ minore di tutti gli elementi che lo precedono, cioè il vettore è ordinato in senso decrescente, il parametro P_j vale sempre j , poiché la sommatoria dei primi n numeri interi vale $n \cdot (n-1)/2$, $T(n)$ è $\Theta(n^2)$. L'ordinamento per inserzione è quindi quadratico nel caso pessimo.

Esaminiamo ora il secondo algoritmo di ordinamento, quello per fusione. L'algoritmo è molto intuitivo nella sua formulazione più astratta, ma ben più complesso dei precedenti se ne consideriamo tutti i dettagli: ne diamo quindi una descrizione di massima e uno schema di codice incompleto, omettendo alcuni dettagli ininfluenti ai fini della valutazione di complessità.

L'idea alla base dell'algoritmo, che ne determina anche il nome, è descritta nel riquadro 4: il vettore da ordinare viene diviso in due parti, le parti ottenute vengono ordinate (*ricorsivamente*, cioè applicando a esse lo stesso algoritmo) e poi fuse ottenendo la versione ordinata del vettore di partenza. La procedura di ordinamento per fusione, riportata nel riquadro 5, ha quindi due parametri aggiuntivi, gli indici *inf* e *sup* che

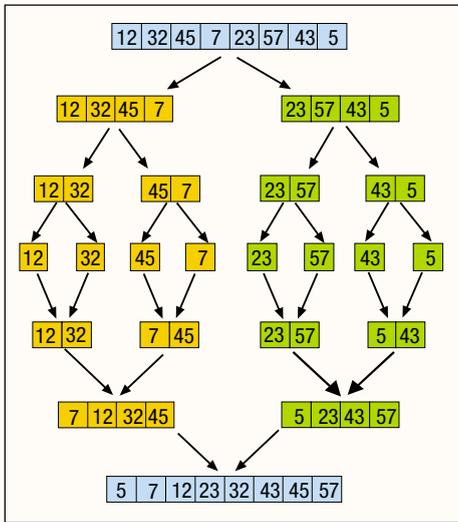
Riquadro 3

Ordinamento per inserzione

Ordina per Inserzione (A)	Costo	# esecuzioni
per tutti <i>i j</i> da 2 a $\text{lung}(A)$ ripeti	c_1	n
elem $\leftarrow A[j]$	c_2	$n-1$
//inserisci $A[j]$ nella porzione		
//ordinata $A[1..j-1]$ del vettore		
<i>i</i> $\leftarrow j-1$	c_3	$n-1$
shintantoché (<i>i</i> > 0) e (elem < $A[i]$) ripeti	c_4	$\sum_{j=2}^n P_j$
$A[i+1] \leftarrow A[i]$	c_5	$\sum_{j=2}^n (P_j-1)$
<i>i</i> $\leftarrow i-1$	c_6	$\sum_{j=2}^n (P_j-1)$
$A[i+1] \leftarrow \text{elem}$	c_7	$n-1$

Riquadro 4

Ordinamento per fusione: simulazione grafica



(come nella ricerca binaria) delimitano la porzione del vettore alla quale viene applicata.

Per valutare la complessità dell'ordinamento per fusione osserviamo che la divisione in due del segmento da ordinare non effettua alcuna operazione sul vettore (consiste solo nel calcolo della posizione mediana) e quindi ha un costo costante, indipendente dal numero di elementi del segmento di vettore da ordinare. Il grosso del lavoro viene quindi svolto nell'operazione di fusione; per brevità e semplicità non codifichiamo la procedura *fondi*, ma, osservando il riquadro 4, è facile comprendere come si svolge: gli elementi da inserire nel segmento di vettore risultato della fusione vengono presi in sequenza, dai più piccoli ai più grandi (cioè andando da sinistra a destra) da una o dall'altra delle due porzioni da fondere, sfruttando il fatto che queste sono già (ricorsivamente) ordinate. A ogni inserzione si sceglie il minore tra il primo elemento di una porzione e il primo dell'altra, lo si preleva dalla porzione in cui si trova e lo si immette nella porzione di vettore ordinato in corso di costruzione, in posizione successiva a quella degli elementi immessi nei passi precedenti (cioè alla loro destra), in modo che anch'essa risulti ordinata in senso crescente. Sulla base di questa, pur informale, descrizione possiamo facilmente concludere che la complessità dell'operazione di fusione è proporzionale al numero complessivo degli elementi del-

Riquadro 5

Ordinamento per fusione: codice

Ordina per Fusione (A, inf, sup)

se $inf < sup$

allora $med \leftarrow (inf + sup) / 2$

Ordina per Fusione (A, inf, med) //ordina la metà inferiore
//del segmento

Ordina per Fusione (A, med + 1, sup) //ordina la metà superiore
//del segmento

fondi (A, inf, med, sup)

le due porzioni che vengono fuse, cioè è $\Theta(k)$, dove $k = sup - inf + 1$; l'algoritmo *fondi* ha perciò complessità $T(n)$, è $\Theta(n)$ ovvero è lineare. Per completare la valutazione della complessità facciamo riferimento al riquadro 4 e "tiriamo le somme" considerando tutte le operazioni di fusione effettuate globalmente dall'algoritmo, prescindendo dal loro ordine, evidentemente irrilevante. Considerando le sezioni orizzontali del riquadro 4 che descrivono le operazioni di fusione, notiamo che in ognuna di esse il numero complessivo di elementi coinvolti nell'operazione di fusione è sempre lo stesso, cioè n , il numero complessivo di elementi del vettore da ordinare (nelle sezioni alte del riquadro 4 si fondono molti segmenti piccoli, precisamente 2^i segmenti con $n/2^i$ elementi ognuno, mentre nelle sezioni basse si fondono pochi segmenti più grandi, ma il numero totale di elementi è sempre n). Quante sono le sezioni orizzontali del riquadro 4? Possiamo stabilirlo con un ragionamento simile a quello fatto per valutare la complessità della ricerca binaria: ogni volta un segmento viene diviso in due, quindi il numero delle sezioni orizzontali è $\log n$. In conclusione il costo complessivo dell'algoritmo di ordinamento per fusione è proporzionale a n (il costo delle fusioni complessivamente effettuate in corrispondenza a una delle sezioni orizzontali del riquadro 4) moltiplicato per $\log n$ (il numero delle sezioni orizzontali), quindi la complessità $T(n)$ è $\Theta(n \cdot \log n)$.

4. SI PUÒ FARE DI MEGLIO? PROBLEMI APERTI E CHIUSI

Abbiamo visto sui due esempi della ricerca e dell'ordinamento che, come nella realtà quotidiana ci sono più modi per risolvere un problema, anche in informatica si possono trovare di-

versi algoritmi, che differiscono per logica di funzionamento e per complessità di calcolo. È chiaro che gli algoritmi più facili e immediati da concepire e formulare non hanno molta probabilità di essere efficienti: per ottenere i risultati migliori sono necessari una profonda comprensione della natura del problema e fantasia per inventare soluzioni originali e non banali. La fiducia nell'ingegno e nell'inventiva umani ci può far credere che, con sufficienti sforzi, per qualsiasi problema possano essere trovate soluzioni sempre più efficienti. Lo scenario alternativo a quello appena prospettato è quello in cui per un determinato problema, esista una certa soglia di efficienza algoritmica invalicabile e quindi, una volta trovato un algoritmo che appartenga a questa classe di complessità asintotica, per quanti sforzi si facciano per ideare nuovi algoritmi, non si riesca a oltrepassare questa barriera, perché i nuovi algoritmi portano soltanto miglioramenti marginali (per esempio, abbassano le costanti moltiplicative o migliorano le prestazioni per valori limitati delle dimensioni dei dati). Ebbene, purtroppo questo secondo scenario è quello corrispondente alla realtà. Per esempio, per il problema dell'ordinamento, è possibile dimostrare che un qualsiasi algoritmo che si basi su operazioni di confronto tra gli elementi da ordinare ha necessariamente complessità asintotica di classe $\Theta(n \cdot \log n)$ o peggiore. In altri termini, $n \cdot \log n$ costituisce un *limite inferiore* alla complessità del problema dell'ordinamento; sappiamo inoltre che questo limite viene raggiunto dall'algoritmo, relativamente semplice, di ordinamento per fusione. Similmente, si dimostra che qualsiasi algoritmo di ricerca basato su operazioni di confronto deve avere complessità almeno logaritmica, come l'algoritmo di ricerca binaria.

Se, per un determinato problema, si riesce a trovare una classe di complessità che ne costituisce un limite inferiore e allo stesso tempo è noto un algoritmo con complessità apparte-

nente a tale classe, si dice che tale problema è *chiuso*. Il problema dell'ordinamento è dunque chiuso. Oltre all'algoritmo di ordinamento per fusione, sono noti in letteratura e utilizzati nelle applicazioni industriali dell'informatica, altri algoritmi, quali il *quicksort* e l'*heapsort*, che sono migliori dell'ordinamento per fusione sotto certi aspetti (per esempio, la complessità nel caso medio, o l'uso della memoria) ma gli sono equivalenti dal punto di vista della complessità asintotica per il caso pessimo.

Fornire un limite inferiore alla complessità di un problema è solitamente difficile, a causa della generalità e dell'astrattezza del procedimento dimostrativo, che deve riuscire a prescindere da ogni particolare algoritmo risolutivo di tale problema e allo stesso tempo esser valido per tutti i possibili algoritmi (inclusi anche quelli non ancora inventati), modellando gli aspetti essenziali che li accomunano; tali dimostrazioni si basano spesso su ragionamenti di natura combinatoria, e utilizzano risultati della matematica non elementare.

Bibliografia

Gli argomenti trattati in questo articolo sono tipicamente oggetto dei corsi universitari di algoritmi e strutture dati. Tra gli innumerevoli testi a carattere divulgativo, didattico o scientifico, segnaliamo in particolare i due seguenti. Il testo di Udi Manber [1] costituisce un'eccellente introduzione agli algoritmi, alle tecniche per la loro progettazione e analisi, molto godibile e avvincente nella lettura. L'opera di Cormen, Leiserson, Rivest e Stein [2] fornisce una trattazione sistematica, di carattere enciclopedico e approfondito.

- [1] Udi Manber: *Introduction to Algorithms: a Creative approach*. Addison-Wesley, Reading, MA, 1989.
- [2] Cormen Thomas H., Leiserson Charles E., Rivest Ronald R., Stein Clifford: *Introduction to Algorithms*. 2-nd Edition, The MIT Press, Cambridge, MA, 2001 (disponibile anche l'edizione in italiano da McGraw-Hill Italia, 2005).

ANGELO MORZENTI consegue il dottorato di ricerca in Ingegneria Elettronica dell'Informazione e dei Sistemi al Politecnico di Milano nel 1988. Dal 1990 è ricercatore e dal 2001 professore ordinario al Politecnico di Milano. I suoi interessi di ricerca sono rivolti ai linguaggi, i metodi e gli strumenti per la specifica, l'analisi, la verifica e la progettazione di sistemi informatici ad alta criticità, embedded e in tempo reale. Su questi temi Angelo Morzenti pubblica numerosi lavori sulle più autorevoli riviste internazionali e conduce progetti di ricerca nazionali e internazionali.
morzenti@elet.polimi.it