

Www: Internet come un ipertesto

Con World wide web (Www) la rete è a portata di mouse

di Alberto Salarelli

Lo sviluppo degli strumenti di navigazione in Internet fin dalla nascita della rete ha sempre avuto come obiettivo prioritario l'incremento di potenza spesso a discapito della facilità d'uso. La creazione del World wide web è stata, da questo punto di vista, una vera e propria rivoluzione copernicana, molto simile a quella rappresentata dall'avvento dell'ambiente operativo Apple-Macintosh sul mercato dei personal computer: l'abbattimento degli ostacoli tecnici attraverso la strutturazione in modalità ipertestuale dell'informazione disponibile in Internet ha avvicinato la rete a portata di mouse in modo da facilitare all'utente non solo l'accesso alle informazioni già presenti sulla rete stessa, ma anche la possibilità di contribuire di persona alla creazione di nuove risorse informative. Il Web è effettivamente la più completa interfaccia utente finora realizzata per la navigazione in Internet allo scopo di fornire "un accesso universale a un ampio universo di documenti".¹

Cosa abbia significato questa svolta nella tecnologia telematica è ben rappresentato dalle statistiche del traffico di rete su Nsfnet ovvero il

backbone, la spina dorsale di Internet in territorio statunitense. Come si può osservare in Fig. 1 (nella pagina accanto) il volume di traffico veicolato su Internet per mezzo del Web ha avuto una crescita continua ed esplosiva passando dagli 8.000 Mb dell'aprile 1993 (0,1 per cento del traffico totale) agli oltre 3.500.000 Mb dell'aprile 1995 (26 per cento), divenendo così il più diffuso strumento di navigazione della rete.²

Agli inizi di aprile, ricevendo all'Università di Padova una laurea honoris causa, lo psicologo Donald Norman affermava nella tradizionale lectio: "Virtuale come opposto a reale, significa una comunità di persone che si incontrano e interagiscono nella ragnatela intangibile di uno spazio informatico invisibile, nel quale la gente, l'informazione e l'interazione possono raggiungere forme e strutture arbitrarie".³ I dati ci dimostrano che questa ragnatela informatica non è più un sogno, ma una realtà: di nome, "World wide web" significa proprio "una ragnatela vasta come il mondo", e di fatto. È certamente difficile dire quanti, ma certamente molti, oggi identificano Internet con il Web. E invece così non è.

L'idea

L'idea del Web è nata in Europa. Più precisamente al Cern, il laboratorio europeo di fisica delle particelle di Ginevra dove, alla fine degli anni Ottanta, si iniziò a concepire la possibilità di organizzare le informazioni disponibili in rete mediante la creazione di un grande ipertesto costituito da nodi (i documenti) e da legami di collegamento fra i medesimi. Si trattava di ideare una sorta di diaframma che celasse all'utente sia le differenze costituite dai differenti protocolli di collegamento relativi ad ogni strumento di navigazione, sia le diverse modalità di approccio ai vari sistemi operativi degli host collegati in linea. L'utente avrebbe avuto la possibilità di saltare da un documento all'altro senza preoccuparsi del come fare, senza ricordarsi complicate procedure d'accesso o indirizzi di macchine. Tim Berners-Lee e Robert Cailliau erano gli autori del progetto originario, nella loro proposta presentata all'approvazione del Cern si legge: "Le incompatibilità delle attuali piattaforme e degli strumenti di navigazione rendono impossibile l'accesso all'informazione esistente attraverso un'interfaccia comune, ciò provoca perdite di tempo, frustrazione e risposte obsolete in fase di recupero di dati. Potenzialmente si può ricavare un grande beneficio dall'integrazione dei vari sistemi in un'unica modalità che consenta all'utente di seguire dei collegamenti puntando direttamente da un'informazione all'altra".⁴

Il progetto Web dunque, a differenza del Gopher, non si proponeva la creazione di nuove fonti di informazione strutturata, bensì aveva come scopo la creazione di una interfaccia per poter rendere più facilmente consultabile alla comunità degli studiosi di fisica ciò che già era disponibile sui database residenti al Cern. Le fasi

del lavoro erano essenzialmente due: la prima consisteva nel creare uno standard di comunicazione in grado di formare l'ossatura dell'ipertesto, la seconda prevedeva la creazione di un particolare tipo di software (denominato *browser*) in grado di interpretare e rendere leggibili le informazioni veicolate sulla rete.

Nel 1992, dopo un periodo di sperimentazione interna al Cern, il Web venne ufficialmente presentato al pubblico e il browser reso disponibile via Ftp.

La successiva svolta nel Web Project porta il nome di Marc Andreessen: l'inventore di Mosaic.

Una volta costituita la rete ipertestuale si trattava di migliorare le possibilità degli strumenti ad essa connessi: "i primi client Www (i

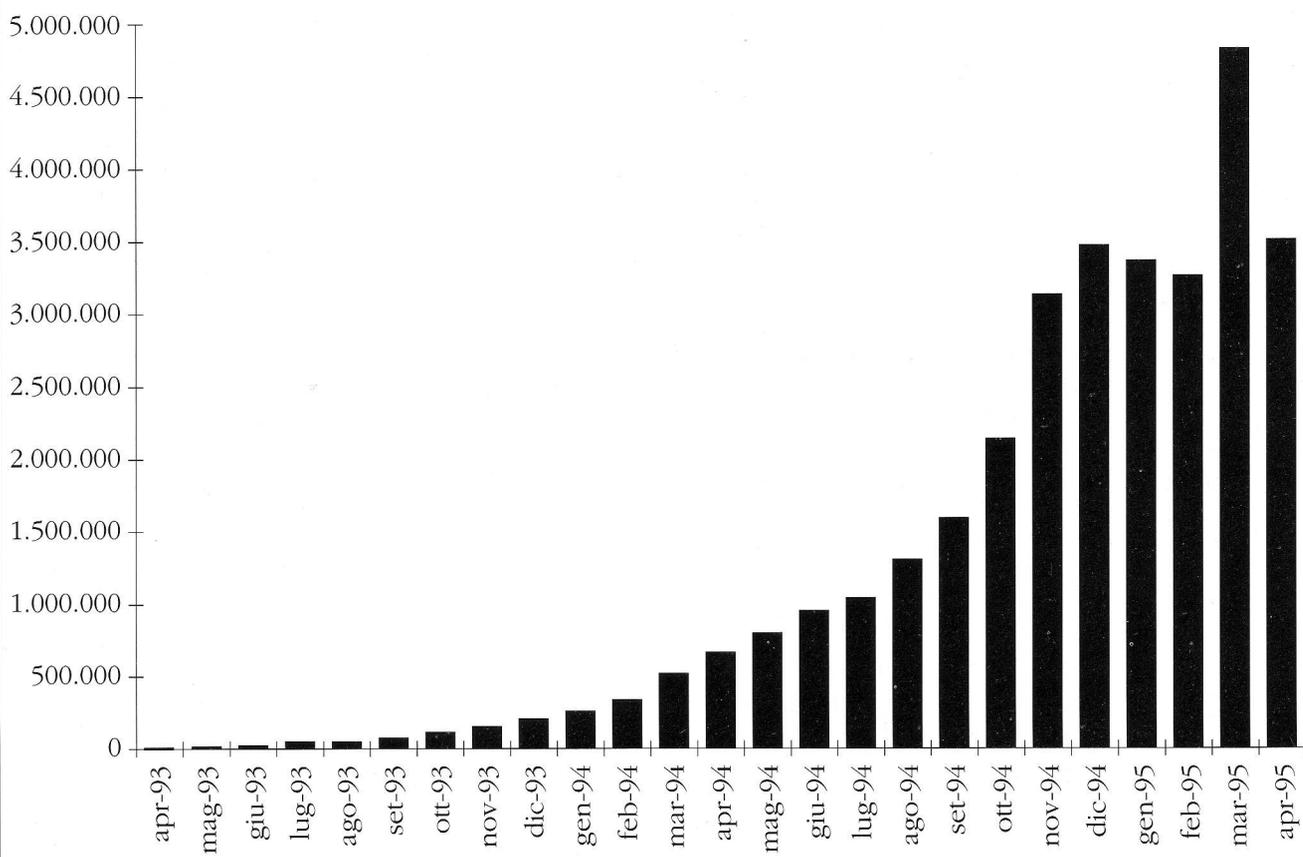
primi lettori) erano browser di tipo testuale, capaci di seguire gli hyperlink (ipercollegamenti) da un documento all'altro spesso localizzato su un sito remoto".⁵ L'interesse di Andreessen, giovane ricercatore presso il National Center for Supercomputing Applications (Ncsa) dell'Università di Champaign, Illinois, era volto all'implementazione di un browser flessibile e multifunzionale per la lettura dei documenti disponibili sul Web, uno strumento in grado non solo di visualizzare testi, ma anche di riprodurre immagini, filmati e suoni. La prima versione di Mosaic fu resa disponibile nel 1993: è grazie a questo software che Internet è arrivato a portata di tutti tramite il "point-and-click". Ed è sempre grazie ad Andreessen che il Web

da rete ipertestuale è effettivamente divenuta anche rete multimediale: da hypertext a hypermedia. "L'ipermediale è l'estensione naturale dell'ipertesto. In ambiente ipermediale è possibile effettuare collegamenti con documenti che contengono, oltre al testo: grafica, fotografie, messaggi audio e video. L'ipermedialità rende i documenti vivi, e il personal computer diventa uno strumento che può generare molto più interesse ed impatto della radio o della televisione."⁶

Come funziona

Il Web dunque non è Internet, bensì è un sistema per navigare su Internet, e cioè sulle reti interconnesse che utilizzano il comune ➤

Fig. 1 - Traffico generato da Www lungo la dorsale Nsfnet

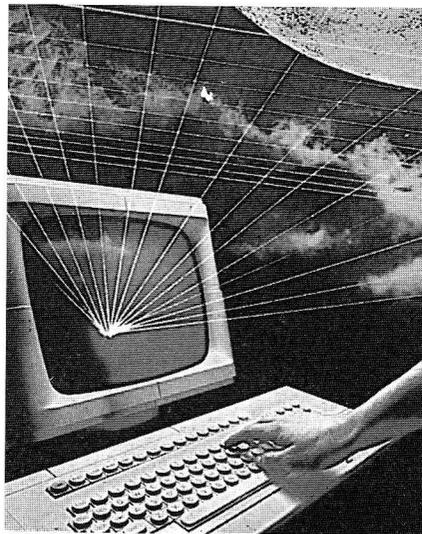


protocollo di comunicazione Tcp/Ip. Ma non solo: il Web è anche un modo per organizzarne le risorse disponibili in rete e per rendere accessibili nuove fonti di informazione. In sostanza "Il World Wide Web è una convergenza di risorse computerizzate per collegare informazioni disperse attraverso Internet in un modo da renderle facilmente accessibili".⁷

La struttura del Web è basata sul modello server-client: il server è un programma che rende una macchina in grado di rispondere alle chiamate di altri computer connessi in rete, computer che per poter interpretare le richieste dell'utente, per poterle trasmettere al server e per saper leggere le risposte necessitano a loro volta di applicazioni client, altresì denominate *browser*.

Il protocollo utilizzato dal server e dal client per poter comunicare è l'Http (Hyper text transfert protocol), protocollo che è indipendente dai sistemi operativi utilizzati dalle varie macchine interconnesse in rete. Le applicazioni che permettono la telecomunicazione "devono essere in grado di riconoscere e utilizzare l'Http per poter spedire e ricevere documenti ipermediali. Per questa ragione i Web server sono spesso denominati Http server".⁸ Quando con il nostro personal computer tentiamo di collegarci con il server della Casa Bianca, il browser che risiede sulla nostra macchina prima avverte il computer del presidente Clinton che vogliamo entrare in collegamento (il messaggio solitamente visualizzato è "Connecting to Http server"), se si riesce a stabilire la connessione ("Host contacted, waiting for reply") è il computer del presidente che spedisce al nostro terminale ("Transferring data") l'ormai celebre messaggio di benvenuto.

Html è l'acronimo per Hyper text markup language. È grazie a questo linguaggio comune di formatta-



zione che tutti i documenti presenti sulla rete possono essere consultati dai singoli computer degli utenti, indipendentemente — lo ribadiamo ancora una volta — dai sistemi operativi e dai software utilizzati per la telecomunicazione. Quando accedo ai documenti gestiti da un qualsiasi server il mio browser è in grado di visualizzare immagini e testo nella giusta sequenza proprio perché sa riconoscere le etichette di riconoscimento delle singole parti del documento, etichette che costituiscono la sostanza del linguaggio Html. Compilare un documento in formato Html significa dare la possibilità, a tutti coloro che vi accederanno, di poterlo consultare in modo corretto e ordinato. Il linguaggio Html, almeno nelle sue funzioni di base, è estremamente semplice e pratico (è infatti sufficiente un qualsiasi editor di testo per poter compilare un documento) ma il suo ruolo è essenziale nell'economia del Web: è l'Html che dice "come deve essere formattato il testo, quale immagine inviare, dove collegarsi se la voce è attiva".⁹ Le voci attive, evidenziate mediante sottolineatura o tramite differente colorazione rispetto agli altri caratteri del documento ipertestuale, consentono il collegamento con altri documenti

che possono risiedere sul medesimo host, ma che possono anche trovarsi su computer distanti migliaia di chilometri. "Html utilizza gli Uniform resource locators (Urls) per indicare collegamenti ipermediali e collegamenti, all'interno dei documenti, a risorse disponibili in rete. Tramite un Url è in pratica possibile identificare qualsiasi file o risorsa disponibile in Internet".¹⁰ Il protocollo di passaggio da un documento a un altro è dunque l'Url: Uniform resource locator. La sintassi di un qualsiasi Url è schematizzabile in questo modo: [metodo di accesso]://[indirizzo dell'host]/[directory e file]. Per "metodo di accesso" si intende il modo in cui è organizzata l'informazione che andiamo cercando. Se stiamo aprendo un collegamento con un Web server la prima parte dell'Url inevitabilmente sarà Http: per esempio l'Url di accesso alla Casa Bianca è "http://www.whitehouse.gov".

Questo Url ci indica che il computer www.whitehouse.gov è un Web server, in grado di trasferire documenti ipertestuali. Ma un Url può anche permettere il collegamento con server che utilizzano differenti sistemi di interscambio di dati come telnet, ftp, gopher, wais ed anche le news. Tramite il Web è quindi possibile saltare con estrema facilità da un'applicazione ad un'altra, con l'unica avvertenza di indicare, nella prima parte dell'Url, quale sistema intendiamo utilizzare. Ecco qualche esempio relativo al mondo delle biblioteche:

- telnet://locis.loc.gov (Library of Congress);
- ftp://ftp.eff.org/pub/CAF/library (documentazione dell'American Library Association);
- gopher://libgopher.yale.edu (fornisce una serie di indirizzi di opac in tutto il mondo);
- wais://NPR.library (Library of National Performing Review documents);

— news://soc.libraries.talk (gruppo di discussione sulle biblioteche).

Navigare con il Web

La crescita imponente e rapidissima delle risorse disponibili tramite il Web (l'incremento annuo percentuale dei nuovi server che si inseriscono nel Web è circa del 3.000 per cento)¹¹ pone un problema ben noto ai bibliotecari: non si dà informazione quando le fonti sono disorganizzate, quando manca la mediazione catalografica, quando alle domande poste dall'utente non si è in grado di rispondere per la mancanza di efficaci strumenti di reference.

Nel suo ultimo volume Howard Rheingold, uno dei più acuti studiosi delle modalità di trasmissione di informazione tramite le reti telematiche afferma: "Il problema caratteristico dell'era dell'informazione, in particolare per gli studenti e gli studiosi che passano il loro tempo immersi in una marea di dati, è che c'è troppa informazione disponibile e pochi filtri per vagliare le risorse che sono utili e interessanti per ciascuno di noi".¹² Di fronte a questa situazione, la conoscenza e la capacità di utilizzo dei siti che consentono di pianificare la propria navigazione attraverso l'oceano ipertestuale del Web diventano di fondamentale importanza, soprattutto per coloro che, fra gli scopi della propria professione, hanno anche il compito di fornire in tempi brevi informazioni precise ed aggiornate. Una ricerca attraverso il Web può essere organizzata secondo tre direttrici.

La ricerca per soggetto

Alcuni siti offrono la possibilità di avere un primo panorama delle ri-

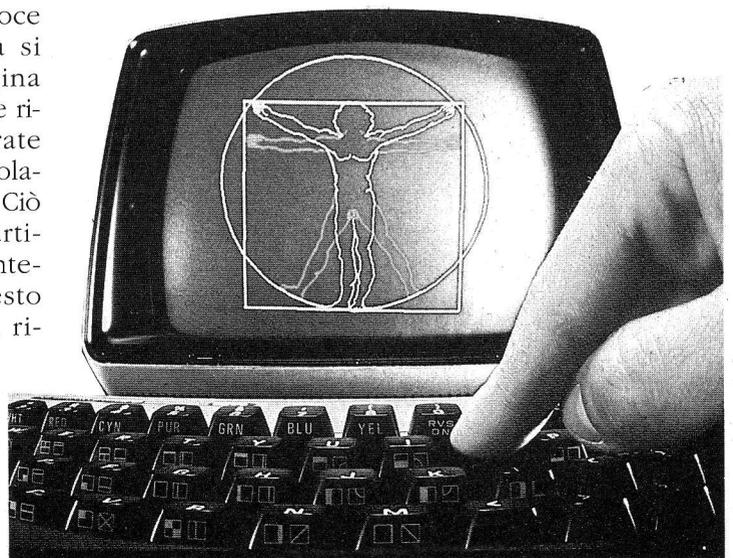
sorse disponibili riguardo ad un particolare argomento grazie a documenti ipertestuali in forma di cataloghi: si identifica la voce che interessa, cliccando su di essa si passa alla lista delle risorse che trattano dell'argomento in questione, ed infine ci si collega all'host desiderato con un altro colpo di mouse.

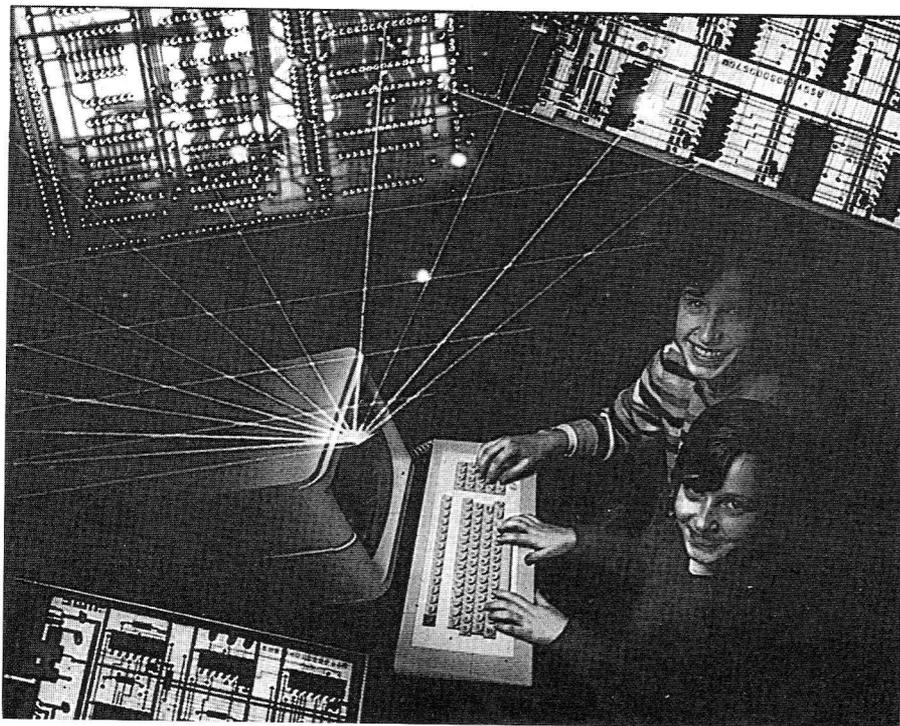
I cataloghi sono gestiti da organizzazioni scientifiche o commerciali (ma la consultazione è comunque gratuita) e sono aperti alla collaborazione di chi li utilizza per ciò che concerne l'ampliamento e l'aggiornamento delle liste di risorse disponibili. (Di alcuni degli strumenti presentati in questo e nel prossimo paragrafo viene data segnalazione nella rubrica "Navigando per Internet" che appare in questo stesso numero della rivista, ndr).

The Www Virtual Library (<http://www.w3.org/hypertext/DataSources/bySubject/Overview.html>) è il catalogo per soggetti gestito dal World Wide Web Consortium che comprende il Cern (la casa natale del Web), l'Institut National de Recherche en Informatique et en Automatique (Inria) e il Massachusetts Institute of Technology (Mit). Il catalogo è ordinato alfabeticamente per soggetto, cliccando sulla voce che interessa si apre la pagina che contiene le risorse desiderate su quel particolare argomento. Ciò che rende particolarmente interessante questo strumento di ricerca è l'idea di catalogazione partecipata su cui si fonda. Infatti gran parte

delle singole pagine che raccolgono le risorse disponibili per ogni soggetto sono curate da istituti e associazioni locati in tutto il mondo, ognuno responsabile per l'aggiornamento della propria voce, mentre al Www Consortium compete il coordinamento generale e il mantenimento dell'indice. Così, ad esempio, la pagina relativa alla medicina è a cura degli Harvard Biological Laboratories, mentre quella relativa alla filosofia è di competenza del Department of Philosophy dell'Università di Bristol. Grazie all'iniziativa di Marco Pirovano del Cnr esiste nella Www Virtual Library una sezione che illustra le risorse disponibili tramite i Web server italiani, l'Url per consultare direttamente questo catalogo di impronta nazionale è <http://www.mi.cnr.it:80/IGST/>.

The Whole Internet Catalog (<http://src.doc.ic.ac.uk/gnn/wic/index.html>) è curato dalla O'Reilly & Associates Inc., si tratta di uno dei siti più noti per chi naviga Internet, grazie anche alla pubblicità che gli deriva dalla sua versione a stampa compresa nel volume *Internet* di Ed Krol (pubblicato in Italia da Jackson Libri). Il catalogo è ordinato per classi e sottoclassi, ma offre anche la possibilità ►





di consultare l'indice alfabetico dei soggetti e l'indice alfabetico delle risorse segnalate.

EINet Galaxy (<http://www.einet.net/galaxy.html>): è un catalogo organizzato in modo analogo al World Internet Catalog, ed anch'esso è messo da un privato — in questo caso la EINet, società di servizi telematici — a disposizione gratuita del pubblico di Internet. Particolarmente ricche le pagine relative alla "Reference and interdisciplinary information".

Yahoo (<http://www.yahoo.com>) è un catalogo nato per iniziativa di David Filo e Jerry Yang. Originariamente ospitato su un host dell'università di Stanford, oggi Yahoo è divenuto un'impresa a se stante, parabola del resto comune a tante iniziative cresciute nel Web. È strutturato secondo menu gerarchici: classi, sottoclassi, fino ad arrivare alla voce più specifica che contiene la lista delle risorse disponibili, talora accompagnate da una breve descrizione. Si tratta di un catalogo estremamente ampio anche se orientato per lo più

verso il mondo commerciale e associazionistico statunitense.

La ricerca per parole chiave

The Www Virtual Library, EINet Galaxy e Yahoo, oltre a fornire informazioni ordinate per soggetti in modo alfabetico o classificato, consentono anche di scandagliare la loro lista di risorse mediante una ricerca basata su parole chiave. In un apposito modulo elettronico si inserisce la stringa di ricerca e poi è il computer che si occupa di recuperare e visualizzare gli Url che interessano.

Per chi è interessato a ricerche di questo tipo, maggiormente specifiche rispetto al recupero di informazioni correlate che deriva dall'uso di cataloghi *subject-oriented*, esistono però strumenti molto più efficaci, gli *spider*,¹³ che non limitano la loro ricerca ad un solo indice, ma che compiono analisi "trasversali" fra le varie collezioni di risorse disponibili sul Web, un modo di lavorare analogo, per certi a-

spetti, ad Archie o a Veronica.

Il recupero dell'informazione tramite gli spider avviene in modo del tutto automatico, una volta inserita la stringa di ricerca il computer analizza le collezioni di Url che sono a sua conoscenza e, di conseguenza, fornisce all'utente i risultati. Non bisogna considerare gli spider come strumenti in grado di recuperare ogni tipo di informazione: la crescita del Web è talmente rapida e incontrollata che non può esistere un unico database sempre aggiornato e puntuale. Eppure, proprio perché sono in grado di analizzare una pluralità di cataloghi contemporaneamente, gli spider "esistenti sono ottimi punti di partenza per reperire risorse nel Web".¹⁴

The World Wide Web Worm (<http://www.cs.colorado.edu/home/mcbryan/WWW.html>), creato da Oliver McBryan, ha vinto nel 1994 il premio "Best of the Web", come miglior strumento di ausilio alla navigazione telematica. Oltre alla stringa di ricerca, il Wwww chiede di delimitare il campo di indagine (Url, titolo dei documenti, Url citati nei documenti...). Le ricerche possono essere effettuate secondo due modalità: una semplificata (basata sull'utilizzo di keyword e degli operatori booleani "and" e "or"), l'altra più complessa in grado di supportare le variabili del comando *egrep* di Unix.

Lycos (<http://www.lycos.com>): con oltre 5 milioni di risorse ove poter effettuare la ricerca, Lycos è uno degli spider più potenti oggi a disposizione di chi ricerca informazioni sul Web. Non è necessario che i termini di ricerca siano esatti, infatti, analogamente a quanto avviene con una ricerca Wais, "i risultati della ricerca sono presentati all'utente in ordine di rilevanza decrescente nei confronti della stringa di ricerca applicata".¹⁵

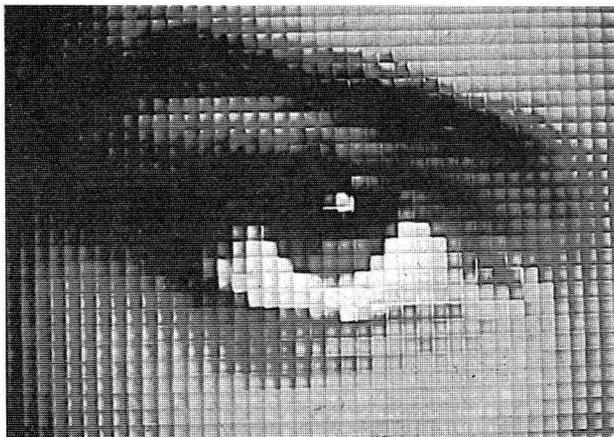
WebCrawler (<http://www.webcrawler.com>), gestito dalla Ame-

rica Online è uno spider analogo a Lycos con una differenza, fondamentale, che le ricerche vengono condotte anche full-text, all'interno dei singoli documenti. Sono inoltre indicizzati non solo documenti ipertestuali (Http), ma anche risorse disponibili tramite Gopher e Ftp.

La ricerca per localizzazione geografica dei server

Forse può apparire strano nel mondo telematico, che tende ad annullare le distanze e a superare le barriere nazionali, un percorso di ricerca orientato geograficamente. Eppure a volte si tratta della strada più proficua per cogliere informazioni di carattere pratico. Voglio fare un viaggio a Londra? Se scorro l'elenco dei server localizzati nella capitale britannica posso immediatamente rendermi conto di quante notizie mi vengono messe a disposizione: alberghi, teatri...

Oppure, altra situazione, cerco il numero di telefono o l'indirizzo di e-mail del mio collega docente all'Università di Virginia? Nulla di più semplice che puntare direttamente il mio mouse sulla mappa dello stato in corrispondenza di Charlottesville per avere immediatamente accesso al server collegato alla segreteria dell'ateneo.



L'elenco su base geografica dei Web server di tutto il mondo è mantenuto presso il World Wide Web Consortium all'Url <http://www.w3.org/hypertext/DataSources/WWW/Servers.html>.

Se siamo interessati unicamente all'ambito europeo allora l'Url giusta è <http://wings.buffalo.edu/world/europe.html>.

Ed infine la mappa dei Web server italiani è disponibile, grazie al Cnr, presso <http://www.mi.cnr.it/NIR-IT/NIR-map.html>.

Conclusioni

La crescita rapidissima del Web rende sempre più necessari i filtri in grado di far interagire l'universo telematico con l'utente: le competenze professionali dei bibliotecari si evidenziano non solo nella capacità di reperire informazioni ma anche nell'organizzazione delle medesime. Mediare l'informazione significa non solo saper utilizzare gli strumenti di reference che altri mettono a disposizione, significa anche approntarne di nuovi. La sfida che il Web, inteso come insieme di tecnologia e di persone che la utilizzano per rendere disponibili al pubblico risorse di ogni tipo, lancia ai bibliotecari è estremamente stimolante dal momento che, come afferma Neil Randall, questa ragnatela telematica "è fra

le entità tecnologiche più rapidamente adottate in un secolo che ne ha viste molteplici: comprendere il Web può essere cruciale per interpretare il secolo venturo".¹⁶ Un mezzo di comunicazione basato sui concetti di ipertestualità e multimedialità richiede dei cataloghi adeguati a cui l'uten-

te che naviga possa guardare come a punti di servizio in grado di poterlo indirizzare sulla giusta rotta: i bibliotecari hanno la grande possibilità di poter contribuire, strutturando risorse adeguate, a rendere il Web più percorribile. ■

Note

¹ Cfr. K. HUGHES, *Entering the World Wide Web: A Guide to Cyberspace*, <http://www.hcc.hawaii.edu/guide/www.guide.html> (la traduzione di questo e dei successivi brani in inglese è di Alberto Salarelli).

² Le statistiche dell'Nsfnet sono disponibili presso <http://www.cc.gatech.edu/gvu/stats/NSF>

³ D. NORMAN, *Due o tre cose che so del futuro*, "La Repubblica", 2/3/95, p. 31.

⁴ T. BERNERS-LEE - R. CAILLIAU, *World Wide Web: Proposal for a Hyper Text Project*, <http://www.w3.org/hypertext/WWW/Proposal.html>

⁵ R. AYRE - K. REICHARD, *In giro per il mondo con World-Wide-Web*, "Pc professionale", 47, marzo 1995, p. 241.

⁶ BILL EAGER - MARY ANN PIKE, *Using the World Wide Web and Mosaic*, Indianapolis, Que, 1995, p. 24-25

⁷ JOHN DECEMBER - NEIL RANDALL, *The World Wide Web unleashed*, Indianapolis, Sams publishing, 1994, p. 7.

⁸ K. HUGHES, *op. cit.*

⁹ S. PILLON, *WWW: rivoluzione o chimera?*, "Mcmicrocomputer", 146, dicembre 1994, p. 259.

¹⁰ K. HUGHES, *op. cit.*

¹¹ B. EAGER - M.A. PIKE, *op. cit.*, p. 21.

¹² H. RHEINGOLD, *The virtual community*, London, Minerva, 1995, p. 57.

¹³ "A spider is a term for a class of software programs that wander through the Web and collect information about what is found there" J. DECEMBER - N. RANDALL, *op. cit.*, p. 386.

¹⁴ *Ibidem.*

¹⁵ C. BASILI - C. PETTENATI, *La biblioteca virtuale*, Milano, Editrice Bibliografica, 1994, p. 107.

¹⁶ J. DECEMBER - N. RANDALL, *op. cit.*, p. 5.

Alberto Salarelli
e-mail: bibltec@ipruniv.cce.unipr.it