

Come orientarsi tra i motori di ricerca

Una panoramica sugli strumenti di recupero delle informazioni in Internet

di Giovanna Grifoni

La crescita esponenziale delle risorse informative, che ogni giorno vanno ad aggiungersi a quelle già esistenti nel vasto mondo di Internet, è un fenomeno di così vaste dimensioni che il rischio di perdersi e soprattutto di perdere del tempo nel tentativo di rintracciare ciò che serve sta diventando una realtà con cui bisogna fare sempre più i conti. Alla soluzione del problema, ma forse sarebbe meglio dire al tentativo di soluzione, hanno provveduto i cosiddetti motori di ricerca, coetanei della rete, strumenti creati allo scopo appunto di facilitare il reperimento delle informazioni e di mettere ordine in un universo disordinato di notizie. Ma se l'antico adagio "chi cerca trova" è vero anche per chi si avventura tra le "ragnatele" di Internet, qualche perplessità sorge quando si vanno ad analizzare i risultati della ricerca, soprattutto per quanto riguarda la pertinenza e l'efficacia delle risposte restituite. Come sanno, infatti, tutti coloro che hanno avuto l'occasione di sperimentare i motori di ricerca non sempre è facile capire come funzionano e soprattutto quali sono le principali differenze tra l'uno e

l'altro e quale deve essere la migliore strategia di ricerca da adottare nei vari casi.

Negli ultimi tempi si è assistito inoltre ad un proliferare di questi strumenti, cosicché alle due principali categorie finora conosciute che distinguevano i motori di ricerca (*search engines*) dalle *directory* o liste di soggetti (Yahoo! è il più noto)¹ se ne sono aggiunte altre definibili nella categoria degli indici specializzati per materia e dei *meta-search engines*, tanto da rendere necessaria un'ulteriore esigenza di ordine. Preliminarmente è opportuno ricordare che per motore di ricerca s'intende qualsiasi sistema informatico che permette il recupero delle informazioni contenute in un database ed il termine è precedente all'avvento di Internet; è questo il caso, per esempio, dei sistemi di ricerca per il recupero delle informazioni dei produttori di cd-rom. Ma qui ci vogliamo occupare soltanto dei motori di ricerca, gratuiti, sviluppati per la ricerca delle informazioni in Internet.

Pur rimanendo attuale la distinzione fra *search engine* e *directory* la tendenza odierna è rivolta a costruire sistemi misti che in aggiunta

ai potenti software, che viaggiano sulla rete a caccia di nuovi documenti web ed informazioni in generale, affianchino liste di soggetti, create e riviste da un apposito staff di persone.

Anche la natura delle informazioni tende sempre più ad assumere caratteristiche aggiuntive, non più legate soltanto al mondo dei servizi di rete finora conosciuti. Interazioni giornalistiche con tanto di direttori responsabili, capi servizio e redattori collaborano alla messa a punto di nuovi servizi d'informazione di tipo giornalistico.

Basta dare un'occhiata alle versioni recenti di Yahoo!, Excite, Lycos, Infoseek. La trasformazione in corso è sempre più orientata a far sì che da sistemi per la ricerca delle informazioni in rete si passi a sistemi di fornitura diretta delle informazioni, con conseguenti aumenti dell'aspetto pubblicitario e quindi inevitabilmente economico dell'intera faccenda. Yahoo! per esempio, è oggi una società che ha fatturato nell'anno precedente 19 milioni di dollari. Il servizio "Today's News" è il risultato dell'accordo siglato nel 1995 tra Yahoo! e l'agenzia di informazioni Reuter. È così che Yahoo! è diventata una delle società per azioni più quotate in Borsa. Così sempre più spesso i "vecchi" motori di ricerca si stanno trasformando anche in agenzie giornalistiche con tanto di redazioni (v. Excite e Lycos) che vanno a caccia di informazioni in rete.

Tenteremo ora di distinguere le principali caratteristiche di questi strumenti per il recupero delle informazioni fornendo una panoramica delle problematiche più comuni legate al loro uso e del dibattito in corso su questo argomento, rinviando, per quanto concerne una attenta analisi delle singole procedure di funzionamento, ai numerosi articoli in linea e non che sono apparsi sull'argomento.²

Qualsiasi tentativo di distinzione

deve dunque tener conto delle seguenti caratteristiche:

— tipologia: indici o liste di soggetti, o entrambi;

— modalità di aggiornamento del database: automatico o "umano" o da parte dell'utente;

— dimensioni del database: numero di documenti web o altro indicizzati;

— ambito di servizi indicizzati: web, usenet, ftp, gopher ecc.;

— finalità del recupero: specifica o generale;

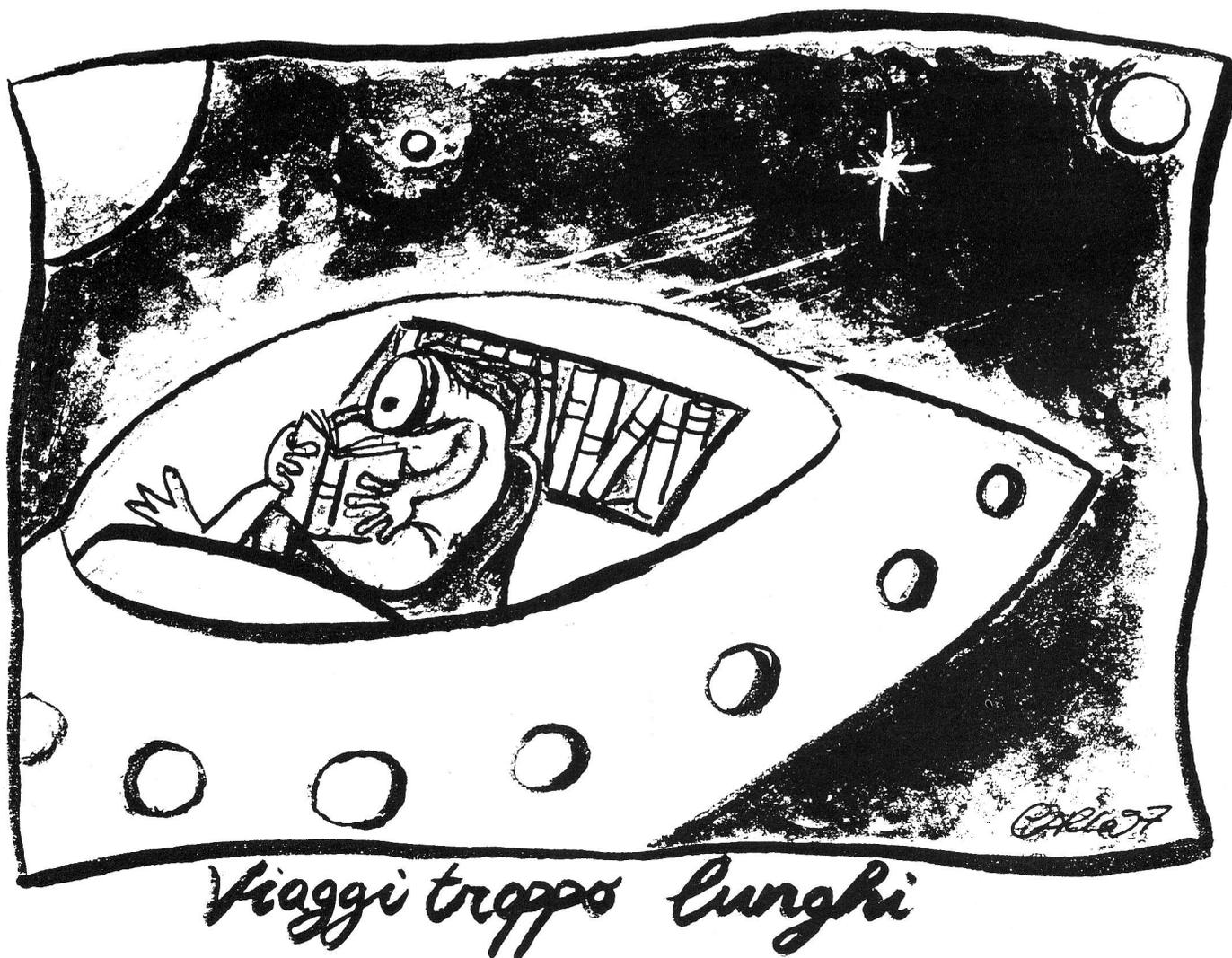
— caratteristiche della ricerca: semplice e avanzata, personalizzazioni.

Rispetto al primo punto la maggiore distinzione finora esistente è stata tra strumenti di ricerca che utilizzano indici e quelli che utilizza-

no *directory*, i primi sono generati dai cosiddetti *web crawler* o *spider* o *robot*, programmi software che recuperano le pagine HTML, analizzando ed indicizzando il loro contenuto, in modo automatico. Le *directory* sono invece elenchi ipertestuali di siti, organizzati in modo gerarchico in categorie e sottocategorie, che per passaggi successivi conducono ai siti contenenti le informazioni cercate, in una struttura che va dal termine più generale a quello più specifico. Entrambi portano alla fine ad un collegamento, ma in modi e fasi diverse.

Nel primo caso il risultato è più immediato, sempre che l'interrogazione sia stata formulata correttamente, conoscendo l'esatta sintassi del-

la *query* e l'ambito delle informazioni contenute nel database, se cioè sono raccolte le informazioni relative ai siti web, usenet, ftp, gopher; nel secondo, grazie al fatto che si possono scorrere delle liste di soggetti espressi in linguaggio naturale e che la compilazione di tali liste è il frutto di una selezione operata da menti umane, il risultato è magari meno immediato, ma in molti casi più preciso e più facilmente raggiungibile anche dai non esperti. A questo proposito tra i più famosi motori di ricerca esistenti in Internet si possono definire indici web a pieno titolo: Alta Vista, HotBot, Open Text Index e World Wide Web Worm. Motori di ricerca misti, in grado cioè di svolgere ➤



CARLO MANTOVANI

Viaggi troppo lunghi

EXCITE (<http://www.excite.com>)

Vanta a detta della società produttrice oltre 50 milioni di pagine a testo pieno indicizzato. Risulta particolarmente efficace per effettuare ricerche generali su web. Originariamente chiamato Architext, Excite è stato sviluppato da un gruppo di lavoro alla Stanford University nel 1993. Dalla fine di novembre Excite è diventato ufficialmente lo strumento di ricerca sul sito di AOL (American Online) il più popolare servizio Internet online degli Stati Uniti. L'efficacia delle sue risposte è dovuta anche al fatto che la compagnia che lo produce spinge i vari amministratori di web ad includere un collegamento ad Excite nelle loro pagine offrendo in cambio la possibilità di scaricare il software di ricerca Excite per usarlo direttamente dai loro siti.

Va bene per: siti web, usenet.

Ricerca semplice: nelle ricerche di base in linguaggio naturale restituisce coincidenze legate non solo alle parole chiave utilizzate, ma anche ai concetti riferiti ad esse, grazie al fatto che il suo database comprende anche recensioni operate da uno staff, le ricerche con più parole chiave restituiscono pagine contenenti almeno un termine. È utile quindi, quando non si è particolarmente sicuri dell'esattezza del termine che si sta cercando.

Ricerca avanzata: pur non essendo presente una interfaccia grafica apposita è possibile inserendo manualmente gli operatori booleani AND, OR, NOT e AND NOT. Volendo si possono sostituire AND e AND NOT con un segno + e un segno - davanti alla parola chiave. (Esempio, McCartney + Starr - Lennon è uguale a McCartney AND Starr AND NOT Lennon). Per cercare un nome proprio insieme ad altre parole mettere il nome con le iniziali maiuscole tra virgolette. L'opzione Sort by Site permette di trovare altri siti contenenti informazioni simili.

ricerche sia come indici che come *directory* sono quelli di Infoseek, Excite, Lycos e WebCrawler. Ad un'altra specie appartengono poi le nuove versioni di Yahoo! e di Magellano Internet Guide che sono essenzialmente *directory*, che però abbinano, come nel caso di Magellano, un database di 15 milioni di indirizzi non ordinati in categorie o come, ed è il caso di Yahoo!, la possibilità di dirottare su Alta Vista la ricerca quando il primo esito è negativo. Al gruppo degli ultimi nati appartengono poi i cosiddetti *meta-search engines*, motori di ricerca che ne attivano, contemporaneamente o in successione, altri, tra i più noti: MetaCrawler, All-in-one, SavvySearch.

I vantaggi di una ricerca condotta con questi metamotori sono evidenti, meno evidenti, ma ugualmente reali, sono gli svantaggi soprattutto per quanto riguarda la quantità di risposte irrilevanti recuperate, dovute al problema del diverso modo che ogni motore di ricerca ha per aggiornare e selezio-

nare le notizie nel proprio database. In pratica questi metamotori non hanno un proprio database ma si affidano a quello di altri, forniscono cioè un'interfaccia di ricerca per sottoporre una interrogazione ad altri motori di ricerca e per ottenere questo si affidano a due diversi modi di accesso.

Il primo consiste nell'elencare i vari motori e fornire un modulo di ricerca per ognuno in un'unica pagina, è quello usato per esempio da All-in-One (<http://www.albany.net/allinone/>) e Beaucoup (<http://www.beaucoup.com/formengs.html>) che effettuano una ricerca sequenziale, e l'utente può aprire un motore dopo l'altro. L'altro, è il caso di SavvySearch (<http://guaraldi.cs.colostate.edu:2000/>) e MetaCrawler (<http://www.metacrawler.com/>), fornisce invece un'unica maschera di ricerca e l'interrogazione viene inviata simultaneamente a più motori. Sia All-in-One che SavvySearch sono disponibili in più lingue, compreso l'italiano. L'inconveniente maggiore rappre-

sentato dal primo caso consiste nel fatto che, essendo i moduli di ricerca (*query box*) in qualche modo standardizzati per tutti i motori elencati, molto spesso non contemplano le caratteristiche specifiche di ognuno, come ad esempio le possibilità di effettuare ricerche avanzate o di utilizzare limitazioni od estensioni, presenti invece nella pagina principale di ciascun motore di ricerca. In questo caso se si ha già una certa familiarità con la sintassi utilizzata da un determinato motore, si ottengono risultati più sicuri. Il maggiore inconveniente invece del secondo tipo di metamotori consiste nel fatto che, pur ammettendo questi l'uso di alcuni operatori booleani, non sempre la stessa opzione è parimenti supportata dai *search engine* con cui si collega il metamotore, cosicché i risultati possono non essere affidabili. In generale il problema più evidente è costituito dai limiti che un ulteriore software, quello del metamotore, i cui meccanismi di recupero dell'informazione non sono sempre chiari ed omogenei, aggiunge a quelli che già sono propri dei singoli motori di ricerca che affidano la costituzione e l'aggiornamento dei loro database a criteri diffusi e talvolta abbastanza casuali.

Per quanto riguarda invece il punto due (selezione automatica o "umana"), la principale differenza consiste nei modi in cui liste di soggetti e motori aggiornano i loro database: i primi sono compilati da uno staff di esperti, appartengono cioè alla categoria degli strumenti di ricerca in cui la compilazione e l'aggiornamento è determinata da una selezione operata da esseri umani che vantano una competenza specifica, mentre gli altri sono affidati del tutto a selezioni ed aggiornamenti automatici, operati appunto dai cosiddetti *crawler* o *spider*, che agiscono in modo diversi. Alcuni *crawler*, per esempio, danno più rilievo alle parole chiave

contenute nel titolo, nelle prime righe d'intestazione o negli elenchi di parole chiave racchiuse dai marcatori META, che a quelle inserite nel testo del documento, ma quasi tutti si concentrano su come l'informazione è strutturata, sulle convenzioni di identificazione editoriale, sulle annotazioni HTML e non sul contenuto del documento. È fuori di dubbio comunque che l'indicizzazione automatica è spesso casuale, mancano a tutti questi strumenti la precisione e l'affidabilità di un catalogo semantico risultato del lavoro di catalogatori esperti e di un thesaurus controllato, anche se chi si occupa di scienza dell'informazione sta studiando un modo per risolvere il problema.

Al tentativo di soluzione di questo problema un gruppo di lavoro guidato da Robert Wilensky, responsabile del progetto di automazione bibliotecaria, a Berkeley in California, ha pensato di provvedere con l'applicazione dell'espedito linguistico della disambiguazione, nota a tutti i bibliotecari che hanno dimestichezza con i cataloghi a soggetto. È questo il caso di tutte quelle parole che sono omografe ma non omologhe e che soltanto il grado di conoscenza e di cultura specifica di chi opera la selezione permette di distinguere assegnando i contenuti ai diversi ambiti e rendendo quindi possibile il reperimento senza dispersione di energie. Nel tentativo sopra citato un algoritmo costruisce un database di riferimento da un'analisi statistica dei vari contesti nei quali una parola viene a trovarsi e la nuova parola chiave viene associata al suo contesto di riferimento.³ Ma una sensibilità al contesto comporta molto tempo nella programmazione e comunque l'approccio non è mai in grado di raggiungere la penetrazione del giudizio di un catalogatore esperto.

A questo proposito è utile ricordare che uno dei punti nodali della qualità delle informazioni in Inter-

net riguarda proprio i criteri della valutazione. L'assenza di barriere alla proliferazione di notizie, chiunque può "pubblicare" in Internet, basta avere un accesso alla rete e un editor HTML, se ha liberalizzato la possibilità di espressione ha sicuramente comportato anche un rovescio della medaglia. Non tutte le informazioni che si trovano sono di per sé buone informazioni, non godono cioè dei benefici propri delle pubblicazioni tradizionali: il provenire cioè da una fonte autorevole e l'essere state sottoposte ad una revisione editoriale e ad una valutazione fatta da esperti della materia. Uno sforzo per comparare il tradizionale processo di pubblicazione off-line a quello on-line potrebbe descrivere il primo come il frutto di passaggi successivi, sottoposti ad un vaglio selettivo che ne assicura generalmente la validità dal punto di vista scientifico, ed il secondo invece come un pro-

cesso autolegittimante nel quale l'informazione arriva in linea da fonti quali i gruppi usenet, listserv, e pagine web.⁴

Sul tema della valutazione e sul grado di attendibilità degli strumenti di ricerca il dibattito in rete si è particolarmente infittito negli ultimi tempi tanto da registrare numerosi articoli on-line sull'argomento, ai quali conviene far riferimento per una più attenta lettura sul funzionamento e sulle caratteristiche dei casi specifici.⁵

Riguardo invece all'ampiezza dei dati indicizzati dai motori di ricerca, la quantità cioè di notizie web o usenet o altro a seconda dei casi, che generalmente ogni motore dichiara nella propria pagina iniziale, va detto che non sempre la quantità dei documenti indicizzati corrisponde a reali notizie univoche, in molti casi infatti, grazie ai numerosi trucchi messi in atto dai produttori di pagine web per ►

ALTAVISTA (<http://www.altavista.digital.com/>)

Va bene per: motore di ricerca della Digital, tra i più noti e consultati dal popolo della rete; offre un ampio database di notizie relative a documenti web ed a usenet. Recentemente la funzione ricerca è stata dotata di una opportunità aggiuntiva, particolarmente utile e non comune a molti altri motori di ricerca: ora è possibile scegliere la lingua dell'interfaccia utente tra una delle 16 lingue più conosciute d'Europa, tra cui anche l'italiano.¹ Tradotte nelle varie lingue sono pure le istruzioni in linea per le formulazioni delle domande ed in generale sui criteri per il recupero dell'informazione, anche se non sono descritte in modo molto chiaro.

Ricerca semplice: si possono utilizzare gli operatori logici anche in questa modalità.

Ricerca avanzata: se si cerca un'informazione su un qualcosa che è espresso da una sequenza di parole si consiglia di inserire all'inizio e alla fine della sequenza il segno delle virgolette (es. "public library service"); mentre nei casi della terminazione singolare o plurale di un termine, o nei casi in cui la radice di una parola può essere comune ad altre che non si vogliono escludere dalla ricerca conviene usare il segno * (es. librar* recupera library, libraries, librarian, ecc.). Dà risultati utili soltanto se è immesso dopo una sequenza di almeno tre lettere. Si consiglia inoltre di usare sempre il carattere minuscolo perché in questo modo la ricerca non è vincolata al fattore maiuscolo che invece recupera soltanto per coincidenza esatta della parola. Sono utilizzabili gli operatori AND, OR, NEAR e AND NOT, rappresentabili anche attraverso i simboli: &, |, ~, e !.

Il motore di ricerca di Altavista classifica i documenti ritrovati in modo tale che quelli contenenti il maggior numero di parole o frasi citate nella richiesta iniziale appaiano per primi nella lista. Un determinato documento ottiene un punteggio più alto di un altro nel caso in cui le parole dell'interrogazione, fatta dall'utente, compaiano nel titolo della pagina web o nelle intestazioni degli articoli delle notizie usenet o risultino vicine nel testo del documento web o siano citate molto spesso nel documento stesso.

¹ <http://www.altavista.telia.com/>

HOTBOT (<http://www.hotbot.com>)

Nato originariamente come un esperimento accademico condotto a Berkeley, Università della California, con la Inktomi, è attualmente sviluppato congiuntamente da Inktomi e Hotwired; ha, rispetto agli altri motori di ricerca, una capacità in più: quella di limitare la ricerca alle pagine web che contengono specifiche tecnologie quali JavaScript o Shockwave, file audio, immagini, documenti Acrobat (scegliere modalità Expert). Ha un database di indici a testo pieno contenente 54 milioni di documenti.

Va bene per: contenuto dei web e di usenet.

Ricerca semplice: utilizzando un apposito menù a tendina si possono utilizzare AND ("tutte le parole"), OR ("una qualsiasi delle parole") o un URL corrispondente alle parole chiave inserite.

Ricerca avanzata: oltre alla capacità, già menzionata, di ricercare file con una particolare estensione, nella modalità Expert è possibile anche ricercare un cyber luogo (cyberplace es: www.ecc.ecc) o un vero e proprio luogo geografico (geoplace, es. Europe, South Africa), o una particolare classe di dominio (es. edu).

assicurarsi una presenza, magari reiterata — il demone o la necessità della pubblicità si adatta a qualunque mezzo — delle loro pagine nel database del motore di ricerca, e dati i meccanismi automatici e dunque non controllati di recupero delle informazioni, nei risultati possono comparire molte volte di seguito le stesse informazioni.⁶ Dando per scontato che non esiste per ora all'interno della rete un motore di ricerca o in generale uno strumento di recupero delle informazioni che vada bene per qualsiasi indagine, conviene orientare la domanda di volta in volta ad una fonte o ad un'altra a seconda di ciò che si vuole trovare; ma per fare questo è necessario prima di tutto conoscere di quale settore di informazioni si alimenta il singolo database del motore di ricerca, se indicizza notizie provenienti da pagine web o da gruppi di discussione o, per esempio, se raccoglie collegamenti con siti che offrono soltanto servizi di ftp o di gopher. Questo tipo di notizia è generalmente fornita nella pagina iniziale con cui si apre il collegamento con il motore di ricerca.

In un'ottica bibliotecaria pare inevitabile chiedersi a questo punto in che misura tali strumenti siano utili per le ricerche; se si possano cioè considerare mezzi alternativi o integrativi dei tradizionali strumenti

di reference. E soprattutto quali siano i migliori motori di ricerca nell'ambito delle informazioni bibliografiche: ricerche di cataloghi di biblioteca, banche dati specializzate nel recupero dei documenti, bibliografie e testi on-line. Senz'altro i bibliotecari da ora in avanti dovranno sempre di più farsi interpreti di mediazione nel recupero delle informazioni bibliografiche tra i testi che appartengono ormai alla "letteratura" in Internet, perché i confini del mondo delle pubblicazioni in rete si stanno ampliando, pur con tutte le legittime diffidenze che si possono continuare a nutrire riguardo alla loro qualità, oltre misura. Esistono ormai numerosi articoli in rete che costituiscono a tutti gli effetti materiali di studio interessanti e quindi ricercabili da sempre maggiori utenti. E sempre più spesso chi si occupa di reference in biblioteca dovrà porsi il problema di reperire informazioni secondo il contenuto ed il formato che soddisfa le necessità di ciascun individuo.

Recentemente il dibattito si è spostato sulla questione se sia migliore ai fini della ricerca fare affidamento sui motori di ricerca, propriamente detti, e universalmente noti al pubblico di Internet o invece non sia più opportuno che ciascuna "comunità" scientifica o culturale, in base ai propri settori di interessi, si co-

struisca un proprio indice di risorse partendo da una scelta mirata e dunque non più generale. Per favorire una personalizzazione delle informazioni, per esempio, Yahoo! offre da poco tempo un nuovo servizio, chiamato "My Yahoo!" che permette di creare una versione *ad hoc*, gratuita, di apertura dell'home page del motore di ricerca su alcuni argomenti preferiti e selezionati. Per creare questa versione personalizzata basta collegarsi all'indirizzo di My Yahoo! (<http://www.my.yahoo.com>) e seguire le istruzioni che vengono offerte, selezionando le categorie d'interesse, fino ad un massimo di 30.

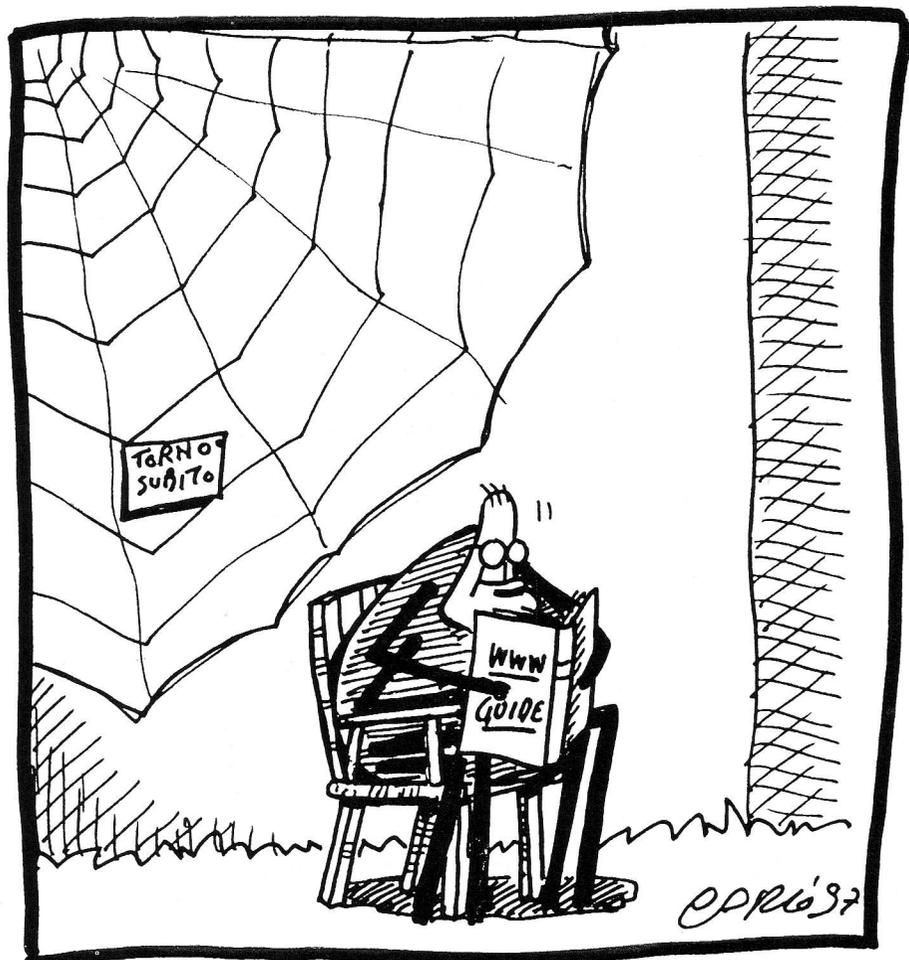
Anche la Apple Computer fornisce un servizio personalizzato che si chiama Apple Personalized Internet Launcher (<http://myhome.apple.com/home/welcome/guest>) che permette di selezionare, gratuitamente, le categorie d'interesse con le quali costruire la pagina personalizzata dando l'incarico al motore di ricerca di provvedere all'aggiornamento automatico. Un servizio è anche offerto da Your Personal Net (YPN). Tutti permettono di variare i campi d'interesse a seconda delle necessità aggiungendo o sottraendo dei soggetti. L'uso di questi servizi potrebbe risultare opportuno per esempio nella configurazione delle postazioni di lavoro dedicate al pubblico, preselezionando da parte dei bibliotecari una pagina d'apertura al motore di ricerca già impostata sui campi di ricerca che possono avere maggiore attinenza con le ricerche che più frequentemente gli utenti di quella biblioteca sono soliti fare.

Per quanto riguarda ancora la questione della qualità, quale sia cioè il motore di ricerca più adatto, quello generale o quello specifico, non si può fare a meno di notare che si stanno affermando sempre più i motori di ricerca limitati a particolari soggetti, i cosiddetti LASE (Limited Area Search Engine), più utili

per ricerche, per esempio, in ambito universitario. Sono questi motori di ricerca, più spesso gateway, costruiti da istituzioni accademiche e scientifiche che garantiscono un'accuratezza nella selezione delle informazioni sicuramente maggiore dei *search engine* commerciali che, come abbiamo già detto, si affidano per la compilazione dei loro database ai robot. Un esempio di questo tipo di motore di ricerca può essere costituito da ARGOS (<http://argos.evansville.edu/>), il database dedicato alle fonti informative in rete sul mondo antico e medievale, che funziona come un classico motore di ricerca per *query* ma è anche contemporaneamente un utilissimo gateway per altri siti web che contengono notizie controllate sulle antichità classiche e medievali, e da ARCHNET (<http://www.lib.uconn.edu.ArchNet/>), un sistema di ricerca dedicato alle informazioni sul mondo dell'archeologia, sviluppato all'Università del Connecticut, che permette la ricerca sia attraverso una lista di soggetti che attraverso l'interrogazione del motore di ricerca, o da MATHSEARCH (<http://ms.maths.usyd.edu.au:8000/MathSearch.html>), sviluppato all'Università di Sidney che è un motore di ricerca che indaga in oltre 90.000 documenti HTML dedicati alla matematica e alla statistica in lingua inglese, o SOSIG (<http://www.esrc.bris.ac.uk>) dell'Economic and Social Research Council che è un database di oltre 1.400 fonti di ricerca nelle scienze sociali interrogabile sia attraverso lo scorrimento della lista dei soggetti che attraverso l'interrogazione del motore di ricerca che fa parte dei cosiddetti SBIG (Subject Based Information Gateways). Questi sono soltanto alcuni esempi, la lista potrebbe proseguire elencando per ogni disciplina numerosi altri gateway che in generale forniscono un servizio integrato e controllato alle due principali funzioni di ricerca: lo scorrimento

delle liste e il servizio di ricerca con un motore. Al fine di migliorare i servizi di informazione sta lavorando un progetto europeo chiamato DESIRE, iniziato nel 1996 all'interno di otto paesi. L'obiettivo del progetto consiste nello sviluppare un'infrastruttura d'informazione integrata per le risorse della rete a favore della comunità accademica. Una delle articolazioni principali del lavoro consiste appunto nell'indicizzazione delle fonti di informazione finalizzata a favorire le necessità di una ricerca di alto livello. Il prodotto dovrà quindi risultare dall'unione dei due elementi finora descritti: accessi all'informazione basati su specifiche aree di soggetto attraverso cataloghi controllati di risorse e indici

automatici di fonti di informazione web su tutte le aree di soggetto. Per quanto riguarda l'ultimo punto, funzionalità della ricerca: semplice, avanzata (operatori booleani e di vicinanza), presenza di help in linea e chiarezza delle istruzioni, nonché semplicità dell'interfaccia grafica, quasi tutti i più noti motori di ricerca offrono la possibilità di ricerche semplici ed avanzate, anche se conviene sempre leggere attentamente le istruzioni in linea perché non sempre valgono le stesse regole per tutti. Per quanto riguarda l'interfaccia utente una significativa novità da segnalare all'interno del popolare AltaVista è quella della possibilità di scegliere una lingua alternativa all'inglese, tra le 16 lingue europee tra cui ►



CARLO MANTOVANI

Anche i Ragazzi
si aggiornano

anche l'italiano, per interrogare il database della Digital.

Un discorso a parte meritano poi tutti quei motori di ricerca che, mutuando l'architettura dei loro indici sulla base dei sistemi bibliotecnomici, offrono la possibilità di scorrere degli indici di argomenti organizzati secondo gli schemi delle classificazioni decimali, tra quelli nazionali è da segnalare IONIO (<http://www.ianus.cineca.it/venus/ionio/www/awww.HTM>), che fa parte dei servizi per l'accesso all'informazione del Cineca. È un motore di ricerca italiano che classifica le discipline su cui immagazzina i dati in base al sistema UDC (Universal Decimal Classification). Tra quelli stranieri, i cui indici sono strutturati in base a UDC sono da segnalare BUBL (<http://www.bu.bl.bath.ac.uk/BUBL/Tree.HTM>) e NISS (<http://www.niss.ac.uk/resource-description/new95udc.HTM>); tra quelli che invece usano la CDD (Classification Decimal Dewey) CyberDewey (<http://ivory.lm.com/>

[~mundie/DDHC/DDH.HTM](http://www.mundie/DDHC/DDH.HTM), Morton Grove Public Library Webrary (<http://www.nslsilus.org/mgkhome/orrs/webrary.HTM>); ed infine CyberStacks (<http://www.public.ia.state.edu/~CYBERSTACKS/homepage.HTM>) che utilizza la Classificazione della Library of Congress.⁷ ■

Note

¹ Inventato da due studenti di ingegneria della Stanford University, David Filo e Jerry Yang, deve il suo nome all'acronimo di "Yet Another Hierarchical Officious Oracle".

² Un'analisi esauriente in italiano è contenuta in *Internet, motori di ricerca*, "Pc Professionale", gennaio 1997, p. 271-301. Numerosi sono gli articoli online che contengono liste di motori di ricerca, suddivisi per categorie, e con valutazioni e punteggi; tra quelli odierni che ci sono sembrati più esaurienti segnaliamo: S. MITCHELL, *General Internet resource finding tools*, disponibile <http://lib-www.ucr.edu/pubs/navigato.HTM>, e T. KOCH, *Internet search service*, disponibile <http://www.ub2.lu.se/tk/demos/DO9603-meng.HTM>.

³ Cfr. G. TAUBES, *Indexing the Internet*, Disponibile <http://sci.aaas.org/aaas/computers/webindex.HTM>.

⁴ Interessante a questo proposito l'esperienza fatta nei corsi di strategie per l'informazione nelle biblioteche della Purdue University illustrate da D. SCOTT BRANDT, *Evaluating information on Internet*, "Computers in Library", may 1996, in sintesi anche <http://www.info.today.com/cilmag/may/techmans.HTM>.

⁵ Cfr., *Literature about search services*, Disponibile <http://www.ub2.lu.se/desire/radar/lit-about-search-services.HTM>; cfr. con l'appendice 3 contenuta in *Beyond surfing: tools and techniques for searching the web*, Disponibile <http://magi.com/~mmelick/it96jan.HTM>.

⁶ Esempi curiosi sui vari trucchi sono visibili collegandosi all'indirizzo <http://webreference.com/content/search/how.HTM> che contiene un interessante articolo di Bruce Grossan sul funzionamento di alcuni motori di ricerca.

⁷ Un accurato esame delle tendenze in atto in tema di servizi di ricerca è contenuto nell'articolo di T. KOCH, *Internet search service*, disponibile anche on line: <http://www.ub2.lu.se/tk/demos/DO9603-meng.HTM>.