

*What-U-Seek, una strana directory dotata di ancor più strani operatori! Nonostante si proclami capace di battere tutti i record, non è più veloce di altri moderni motori.*

dai due punti, Altavista cercherà solo nella parte di documento specificata. Ad esempio battendo title: "Mario Rossi" verrà probabilmente solo visualizzata la home page di questo signore. Pratico, per chi ha le idee chiare.

## Webcrawler <http://www.webcrawler.com>

Non sempre questo metodo di analisi fornisce i migliori risultati. Per questo, nella forma avanzata di ricerca, Altavista permette di specificare le parole che occorrerà usare per la valutazione appena descritta; inoltre il motore mette a disposizione un tool molto

Sponsorizzato da America On-Line, Webcrawler è un motore di ricerca che condivide molte delle caratteristiche di Altavista. La sua proprietà più interessante è quella di accettare l'uso del linguaggio naturale; in altri termini è possibile inserire una domanda del tipo "the fastest car in the world" perché il motore filtri la frase dalle parole inutili (in gergo "the noise words", vale a dire congiunzioni e articoli) e esegua la sua ricerca. Abbiamo usato una frase in lingua inglese perché il motore, ovviamente, ri-

mento

- il numero delle ripetizioni del termine.

Questi valori sono "pesati" e a quello che raggiunge il massimo punteggio viene assegnato il valore di 1000; gli altri sono valutati in rapporto a quello.

tore mette a disposizione un tool molto raffinato, che, in mano a una persona allenata, può dare ottimi risultati; inserendo nella stringa di ricerca, all'inizio, una serie di parole chiave (come host, title, image, anchor, URL, e così via), seguite

## Alcuni consigli generali per ottenere rapidamente risultati positivi in una ricerca (o, magari, per farci trovare il più rapidamente possibile)

Ogni motore esegue le sue ricerche in maniera e con tecniche talora completamente differenti. Questo significa spesso che le diverse pagine sono indicizzate, nei database, con tecniche che possono determinare differenti risultati.

Occorre, innanzi tutto, avere ben chiaro come i motori inseriscono nei loro database la nostra pagina; qui le metodologie possono essere completamente diverse; alcuni indicizzano tutta la pagina, altri solo le prime cento parole, altri l'intero primo paragrafo. Alcuni motori lasciano al cliente la possibilità di inserire le parole chiave che, secondo lui stesso, rappresentano quelle davvero significative in fase di ricerca, altri sviluppano una statistica, altri inseriscono nel db tutti i termini significativi del sito, filtrando le parole inutili (come le congiunzioni, gli articoli, i pronomi e così via). Comunque, tenendo conto della sintassi descritta nell'articolo, ecco alcuni consigli spicci per migliorare la nostra tecnica di ricerca:

Scegliere un motore di ricerca, una directory o una libreria adatta al tipo di ricerca che stiamo svolgendo; è probabile che una ricerca, ad esempio, sui combustibili solidi sia più efficace utilizzando un motore specializzato che magari Altavista o Excite. Ricordo che una lista estesa di motori particolari è presente in Yahoo, alla base della pagina. In altri termini, chiedersi sempre se c'è necessità di cercare su tutto il Web, su un limitato numero di siti specializzati, o magari su alcuni più popolari.

Tenere sempre presente il tipo di ricerca che si sta facendo. E' assurdo cercare una particolare pubblicazione usando WWW. E' sempre il caso di chiederci se quello che si sta cercando non sia meglio reperibile in altri posti o con altre tecniche, ad esempio in

Usenet o con Gopher.

In base a quello che si è detto precedentemente, considerare se il termine, la frase o l'argomento che si sta cercando può essere presente nel primo paragrafo o nella prima parte del documento.

Usare sempre gli help che ogni motore mette a disposizione; sovente vi si aggiungono nuove caratteristiche, che restano ignorate se non le si provano per una volta.

Scegliere termini adeguati: inserire sinonimi, plurali e singolari, terminologie alternative (es. auto, automobili, autoveicoli, ecc.). Se si sta cercando un termine in lingua inglese, inserire solo il singolare tranne che nei casi irregolari (ad esempio car e non cars, ma anche mouse e mice)

Evitare di usare termini banali o comuni. Precisazione ovvia, ma non peregrina; non si può pretendere di battere "Internet" e non ritrovarsi con una ricerca inutile e fastidiosa. Ricordarsi che le percentuali di interesse che molti motori evidenziano si basano solo sulla presenza del vocabolo o della frase cercata nel documento; non si può pretendere che il motore interpreti anche il senso del documento. Solo le directory lo fanno, e neppure qui si può pretendere la sicurezza assoluta.

Fare attenzione a come il motore maneggia le maiuscole; come abbiamo visto nell'articolo, queste significano ben altro che una differenza di altezza.

Usare al meglio gli operatori multipli, per ottenere risultati il più possibile confacenti alla nostra richiesta; ricordarsi che gli operatori [+] e [-] devono essere attaccati alla parola che li segue per avere effetto.

conosce le parti inutili solo in questa lingua, ma il filtro elimina anche parole comuni; ad esempio la parola Web è anch'essa destinata alla decapitazione.

La schermata iniziale può essere customizzata, nel senso che è possibile visualizzare solo i titoli o anche un breve sommario; inoltre si può scegliere di visualizzare per ogni pagina 10, 25 o 100 titoli. Ogni titolo mostra il "confidence ranking", l'indice di interesse che qui non è un arido numero, ma una forma di larva, più o meno presente in funzione della importanza del sito. Non è dato di sapere il metro di valutazione usato, ma non pare molto di più del numero di volte che la parola o la frase compare nella pagina. Le frasi si cercano praticamente come in Altavista, e allo stesso modo si usano gli operatori booleani. L'originalità di WC sta però in due operatori più avanzati, NEAR/n e ADJ; il primo permette di stabilire l'ordine con cui devono apparire due parole; ad esempio battendo Demi NEAR/7 Moore vengono rilevate solo le occorrenze in cui il nome e il cognome, in qualunque ordine messi, sono prossimi meno di sette parole. L'altro operatore impone che una delle due parole preceda l'altra; così Demi ADJ Moore non riconoscerà l'occorrenza Moore Demi. Ancora, interessante è la sua implementazione di un catalogo di argomenti che è possibile sfogliare; in questo ambiente c'è un bottone, Spidey, che, schiacciato, esegue una ricerca, basata sui termini imposti, nell'area di preinteresse stabilita aprendo il catalogo degli argomenti.

Webcrawler è un motore moderno, con ottimi tool di ricerca, anche se il suo database non è dei più grandi. Il fatto di essere una emanazione di AOL lo rende, comunque, uno dei motori di ricerca più utilizzati.

## Lycos

<http://www.lycos.com>

È stato per lunghissimo tempo il punto di riferimento della concorrenza, per il suo database immenso e per la sua velocità. Oggi la sua fama è un po' appannata, ma non tanto da non farne il mezzo d'uso quotidiano di una gran massa d'utenti. Offre ambedue le possibilità di eseguire ricerche per parola chiave e per argomento (i soggetti di ricerca sono chiamati "directory services"), e in ambedue dimostra una velocità impareggiabile e una ricchezza di risultati ottenuti per pura forza bruta. Non

*LookSmart, il motore che ci ha fatto fare più anticamera; numerose e ben articolate le categorie, modeste le possibilità di personalizzare la ricerca.*

ha un gran numero di operatori a disposizione e l'operazione di filtraggio dei risultati è affidata solo a cinque aggettivi (loose, fair, good, close e strong), che consentono di eseguire ricerche sempre più raffinate e precise. Non è possibile indicare stringhe da escludere dalla ricerca e il termine [-] pur usato, serve solo a impedire che la stringa che lo accompagna possa influenzare il peso assegnato al documento. L'unica raffinatezza è data dalla wildcard [\$] che, aggiunta alla fine della parola, ha le funzioni di terminatore di qualsiasi lunghezza.

## Excite

<http://www.excite.com>

Excite permette di eseguire ricerche combinate per parola chiave e argomento. La seconda tipologia, chiamata "ricerca per concetto", permette di individuare documenti correlabili col seme di ricerca anche in funzione del significato del termine ricercato, e non

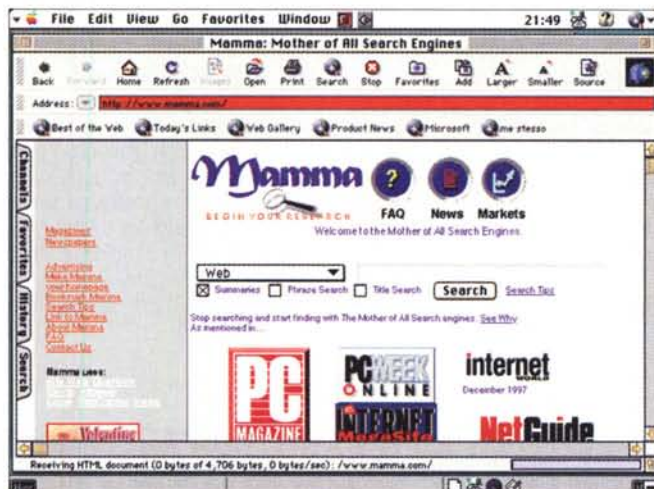
solo in base al numero delle occorrenze presenti. Non esistono veri ambienti separati per ricerca normale o avanzata, e non esiste altresì possibilità di scegliere l'aspetto (sommario/standard/avanzato) dei risultati ottenuti.

Le tecniche di impostazione della stringa o della frase di ricerca, in Excite, sono le più strane nell'ambiente dei motori; le frasi sono trattate come in Altavista, ma Excite è sensibile al numero di occorrenze della parola nella stringa di ricerca. In altri termini se battiamo Apple Apple IBM, Excite elencherà i documenti ricercati dando la precedenza a quelli che sono significativi maggiormente per Apple.

Excite supporta gli operatori booleani AND, OR, NOT e un non tanto inconsueto AND NOT (in assenza è sottinteso OR). La confidence è valutata allo stesso modo di Altavista, ma non è detto che il primo documento della lista

si veda assegnare il valore del 100%. Una opzione interessante è quella che permette di eseguire una scansione della lista e, una volta schiacciato il pulsante presente vicino all'indice di confidence, una nuova

*Di Mamma ce n'è una sola, e non ti abbandona mai, nemmeno nel cyberspazio.*



ricerca viene eseguita usando il contenuto del documento come riferimento.

## Infoseek

<http://www.infoseek.com>

Per lungo tempo motore d'elezione di Netscape, Infoseek è sufficientemente veloce e facile da usare, ma questi pregi si pagano in termini di sofisticazione. Dispone di operatori booleani ma la cosa che più infastidisce è che mostra solo i primi cento documenti trovati. Funziona

*Savvy Search, ottimo esempio di directory, ben articolato e comprensivo.*

*Un'autostrada in mezzo al mare? In WWW può esserci anche questo, e risponde al nome di Highway 61. Si può indicare al motore il tempo che si ha a disposizione, tanto noi andiamo a fare un bagno...*



sia da motore puro che da catalogo di soggetti, e permette di cercare altresì nei newsgroup Usenet, negli elenchi di indirizzi, e nelle raccolte di FAQ.

La ricerca è case-sensitive. Le frasi vanno ricercate per intero con la tecnica delle virgolette, e sono ammessi gli operatori [+ ] e [- ]. Ha una sezione di siti selezionati, e sistemando due parole tra parentesi quadre esegue la loro ricerca finché sono a una "distanza" di cento parole (il valore non è modificabile). Una volta rilevato un sito interessante, esso può essere impiegato come seme di ricerca per altre richieste, invocando il comando "Similar Pages".

## HotBot

<http://www.hotbot.com>

Autodefinitosi il primo motore di ricerca Internet capace di scalare la montagna del Web, HotBot ha una interfaccia colorata e accattivante, e alcune caratteristiche originali che lo rendono utile e gradevole. HB cerca i termini inseriti co-

me se fossero parole singole, frasi, argomenti, senza interventi da parte dell'utente; un comodo menu a tendina specifica se occorre cercare una sola o tutte le parole indicate e se esse vanno cercate nell'ordine o no. Una apposita opzione permette ulteriori affinamenti della ricerca per data, numero delle occorrenze per pagina, locazione, ecc. Ad esempio è possibile cercare pagine contenenti frammenti di codice Java, VRML, Acrobat, o immagini GIF.

L'operatore "must not" consente di eliminare un termine dalla ricerca, mentre "should" indica al motore di ricerca di dare più importanza al termine che lo segue. I veri e propri operatori booleani sono raggiungibili tramite un menu, e la ricerca è case-insensitive, tranne che caratteri maiuscoli e minuscoli siano mescolati nella stessa parola. E' possibile cercare suffissi di indirizzi (ad esempio .mil o .it), e il bottone Geoplace permette di cercare su server sistemati nei diversi continenti (in default vale "tutto il mondo"). Cercando un URL vengono visualizzati documenti che lo contengono sia nel testo che come link.

## OpenText

<http://www.opentext.com>

Non dispone di un database enorme-



me se fossero parole singole, frasi, argomenti, senza interventi da parte dell'utente; un comodo menu a tendina specifica se occorre cercare una sola o tutte le parole indicate e se esse vanno cercate nell'ordine o no. Una apposita opzione permette ulteriori affinamenti della ricerca per data,

numero delle occorrenze per pagina, locazione, ecc. Ad esempio è possibile cercare pagine contenenti frammenti di codice Java, VRML, Acrobat, o immagini GIF. mente grande, almeno comparato con quelli della concorrenza, ma è costantemente aggiornato, e supporta richieste multilinguali, inclusi set alfabetici non latini. Non è l'ideale per ricerca di risorse particolarmente specializzate, ma grazie alla sua velocità (e al fatto che forse non è poi affollato come gli altri) è un ottimo tool per ricerche di ampio respiro.

Caratteristica interessante di OpenText è la possibilità di specificare in quale parte della pagina si desidera limitare la ricerca (titolo, URL, primo paragrafo, ecc.). E' possibile anche eseguire una ricerca su una strana entità, il "Sommario", che è rappresentato da una collezione di dati, come titolo, testata della prima pagina, ancure e link principali, e una serie di parole e stringhe significative del testo.

OpenText supporta gli operatori booleani e quelli di prossimità (NEAR e FOLLOWED BY). Il primo trova una parola nell'intorno di 80 vocaboli (all'indietro e in avanti; l'intervallo non è modificabile), il secondo solo in avanti. In assenza, tra le parole inserite nel seme di ricerca è sottinteso AND; non si tiene conto delle maiuscole, e sono ignorati alcuni segni di interpunzione; singolare e plurale sono considerati valori differenti. Il foglio dei risultati ottenuti nella ricerca non può essere salvato come bookmark.

## Nlight

<http://www.nlsearch.com>

Solo qualche parola riguardo a questo motore, che per certi versi si può considerare specialistico. La finestra di ricer-

ca è piuttosto semplice; si inserisce il termine da ricercare e si schiaccia FIND. Il risultato è offerto da una videata che mostra i ritrovamenti in sei categorie separate (Information Databases, Current News, Archived News, WWW Internet Index, Desktop References e Discount Bookstore). E' sottinteso un AND implicito tra le parole, ma esiste anche l'OR e il NOT; per ciascun operatore è prevista una shortcut. In accordo con molti autori, NL è il motore in cui la lunghezza della stringa inserita influenza di poco la velocità di ricerca. Peccato che, appena ci si discosta dalla ricerca convenzionale, i servizi divengano a pagamento. Occorre però ricordare che in certi campi (come per esempio articoli e biblioteche) possiede un patrimonio impareggiabile.

## Yahoo <http://www.yahoo.com>

A rigore (vedasi l'articolo del mese scorso) Yahoo non è un motore di ricerca ma quello che, con termini complicati, viene definito un indice dei soggetti organizzato gerarchicamente. Ha subito, nel tempo, una serie innumerevole di miglioramenti e di aggiustaggi, che ne hanno fatto uno strumento di grande raffinatezza, customizzabile in maniera chirurgica, pur senza mai perdere la sua facilità, intuitività e immediatezza d'uso. La qualità delle ricerche è elevata, e probabilmente Yahoo è il migliore punto di partenza quando non si hanno le idee molto chiare su cosa cercare; in altri termini può essere considerato il migliore sito da cui, eventualmente, poi trasbordare su siti più specializzati. Basta, in questa ottica, cliccare il link "More Yahoos", presente alla base della pagina di ricerca, per avere immediato accesso a una serie di altri motori, anche essi, sovente, dotati di ulteriore gerarchia.

L'interfaccia di Yahoo è molto semplice da usare; basta inserire il vocabolo o la stringa da cercare e Y restituisce informazioni suddivise in tre gruppi principali; "categorie", vale a dire aree che corrispondono al termine in questione (ad esempio inserendo "engine" ci verranno restituite diverse categorie relative ai motori d'automobile, ai motori elettrici, ai motori di ricerca, e così via); siti finali e categorie dalle quali le parole cercate sono state indicizzate, anche se non ne rappresentavano l'argomento principale. E' possibile indicare dove ricercare (Yahoo, Usenet o indirizzi di e-

Virgilio, il primo motore di ricerca italiano (ma utilizza Infoseek).

mail), usare operatori booleani come OR o AND, specificare sottostringhe o stringhe complete, controllare il numero di risultati per pagina (10, 25, 50 o 100).

Ma la vera carta vincente di Yahoo è la sua liberalità; sebbene esso usi in default Altavista (scusate se è poco!) alla fine di ogni pagina sono elencati una serie di link a tutti i più importanti motori. Inoltre chiamando Yahoo Remote si apre una pagina secondaria che è possibile tenere a disposizione per gli usi che più ci abbisognano (ad esempio una ricerca parallela, senza perdere di vista quella principale).

## E per concludere...

Già, per concludere, si fa per dire, non possiamo non citare The Internet Sleuth e Magellan; si tratta più che altro di due servizi di ricerca, che superano il concetto implicito nei motori per gestire le ricerche in maniera un po' diversa. Il primo, ad esempio, controlla non uno, ma una serie di database, e fornisce una pagina di input in cui essi sono disponibili; le ricerche vanno fatte per approssimazioni successive (ad esempio film, poi classici, poi "Via col Vento"). Magellan, ancora, non è un motore, ma più che altro una guida in linea a Internet che contiene un elenco di siti visionati e giudicati. In altri termini non è un luogo ottimale in cui eseguire ricerche alla disperata, ma un punto di riferimento per chi desidera, eseguendo una ricerca specifica ma non molto specializzata, avere risultati affidabili e di buona qualità.

Ma molto più ampia è la messe di siti che si mette a nostra disposizione per guidarci, come Virgilio, nell'inferno tortuoso del Web; già, proprio Virgilio, il primo motore di ricerca italiano, ma occorrerà ricordare anche MetaCrawler, SavvySearch, Euroferret, EINET/Tradeware Galaxy, Aliweb, Netfirst, Ariadne, Harvest, WWW Worm (oggi divenuto Go2!), What-U-Seek, Wombat, Mam-



ma, BankDesk, Profusion, Titanic (sic!) e tanti altri, sovente utilissimi perché estremamente specializzati, sovente perfettamente inutili. Inutile darvi le coordinate per cercarli, basta schiacciare, in Yahoo, alla fine della pagina di ricerca "More Yahoos..." per averli tutti elencati e opportunamente linkati. Ancora, non dimentichiamo che molti siti offrono una griglia di motori in linea, per scegliere al volo la cilindrata che più ci interessa.

## Conclusioni

Beh, anche stavolta lo spazio si è esaurito senza accorgercene! Certo l'argomento è enorme e ci torneremo sicuramente su (ad esempio già solo i servizi aggiuntivi richiederebbero una puntata), ma lasciateci esprimere un parere e dettare un consiglio. Non scegliamo un motore di ricerca solo guardando alla grandezza del suo database. Sovente una migliore indicizzazione, la possibilità di giungere al risultato attraverso fasi successive, la disponibilità di accedere non solo ai documenti finali ma anche alle categorie, paga, alla fine, molto di più della forza pura e semplice. E' questo il motivo della gran fortuna di Yahoo; il perfetto connubio tra un potente motore (Altavista) e una directory ben costruita è un'arma sempre vincente, e il fatto che il suo sito sia visitato, mediamente, da otto milioni d'utenti al giorno (sarà vero, sarà falso, sarà la solita leggenda metropolitana del cyberspazio?) è segno del successo dell'impostazione. Non a caso le tariffe per inserire un banner nella home page di Yahoo sono cifre con molti (ma proprio molti) zeri!