

# A saper cercare bene...

Ci sono oltre tre miliardi di pagine, su Internet.

E crescono a un ritmo impressionante.

Ci sono un sacco di informazioni utili, su Internet.

E ci sono un sacco di modi per trovarle, queste informazioni.

Cioè, ci sono un sacco di motori di ricerca e di directory.

Diamo un'occhiata a cosa (e quali) sono.

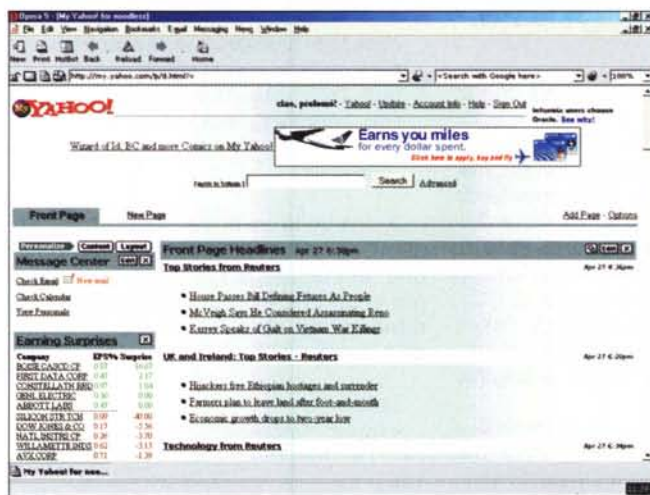
Prima parte

di Raffaello De Masi

A volerle cercare, le cose, le si trovano sempre. Basta volerlo! Secondo me esiste un meccanismo extrasensoriale, praticamente sconosciuto, che permette alle cose di avverarsi più o meno rapidamente secondo il nostro grado di desiderio personale. In pratica, gli eventi, in base a forti desideri personali, posso essere guidati e raggiungere l'obiettivo da noi desiderato o, viceversa, allontanarne se da noi paventato.

Un esame all'università particolarmente temuto, una valutazione cui dobbiamo essere sottoposti, perfino una malattia o una situazione di disagio, secondo questa mia personale filosofia, sembrano subire l'influsso del nostro stato d'animo, dirigendosi secondo i nostri desideri. Solo con le ragazze, per qualche arcano motivo, non funziona, o forse sarà perché, come si dice dalle mie parti "o' mobile è quello che è"; fatto sta che Samantha De Grenet, Carrie Otis e Anna Galiena non riesco a includerle nella rubrica del mio telefonino.

Ci deve essere nella mia ricerca filosofica, certe volte, mi è parso di esserci arrivato molto vicino - una specie di "motore di ricerca" nostro personale che ci porta a rag-



giungere i risultati desiderati in funzione delle richieste che noi, inconsciamente, gli poniamo. Purtroppo non sono riuscito a metterlo a punto ancora bene, ma non mancherò di tenervi aggiornati sugli sviluppi di tale dottrina filosofica, al cui confronto le teorie di Hegel sono bigliettiini dei cioccolatini; ma nel frattempo, visto che la filosofia non porta pane a casa, mi tocca parlare invece di

motori più terra terra, quelli di ricerca su Internet. E, nell'ottica della specializzazione adombrata nella premessa, di motori, per così dire, da "super ricerca".

## Quante ore spese appresso a una ricerca!

Cerca e troverai, intenzioni nobilissime ma forse un tantino ottimistiche nell'era della ricerca digitale. Purtroppo il Web è troppo grande - 3 miliardi di pagine che crescono a un ritmo giornaliero inimmaginabile - e nessuno può dire che la propria ricerca sia più esauriente e più completa di quella eseguita da un altro, o che un motore sia capace di fornire risultati più affidabili. E ancora nessuno



può dire che l'informazione che noi cerchiamo non sia nascosta in un oscuro server ai margini della galassia di Internet. In altre parole, una ricerca per essere coronata da successo ha bisogno di tempo, tentativi, ed attrezzi del mestiere precisi e finalizzati, e purtroppo non ci sono scorciatoie da poter utilizzare.

Ci sono due metodi principali per mettere ordine nel caos. Il primo si basa sui motori di ricerca che a loro volta si basano su robot software, detti anche

spider, è ancora quello animatore. Cercare in base a parole chiave ha buone possibilità di restituire siti rilevanti; ciononostante i diversi motori di ricerca si differenziano per una serie di

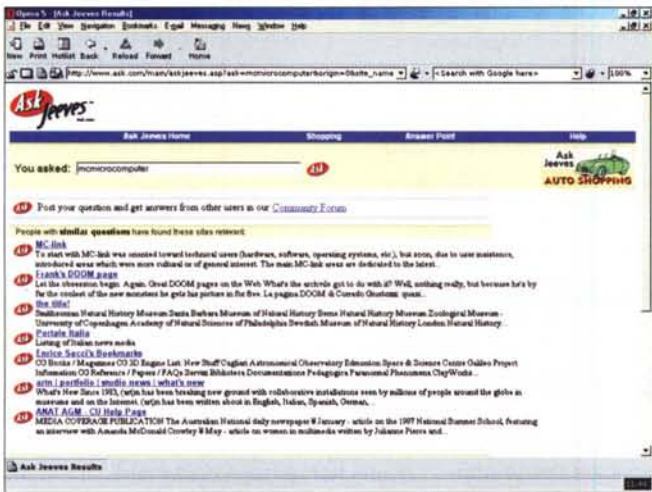
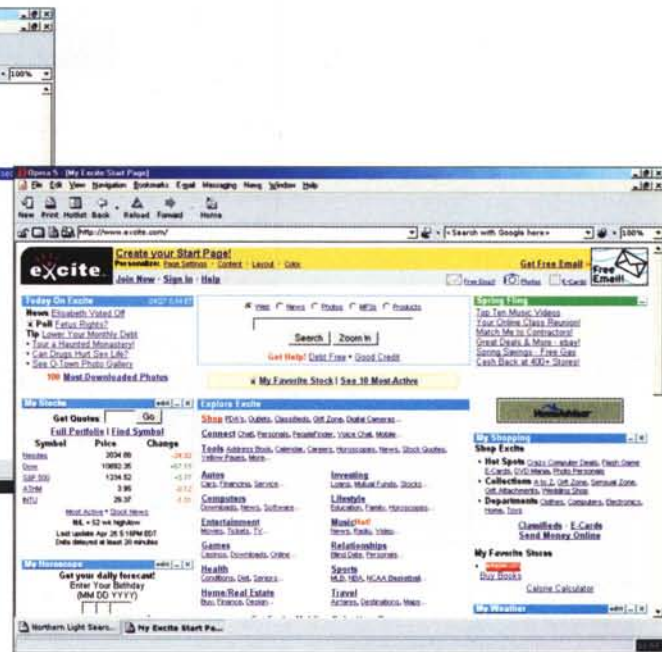
variazioni sul tema. Alcuni spider analizzano i meta tag, mentre altri li ignorano completamente, certi esplorano il Web più velocemente di altri, ed taluni ancora analizzano e strutturano i siti in base ad algoritmi filtro, talvolta originali, che permettono di assimilare tra di loro pagine somiglianti. Il maggiore vantaggio in questa tecnica è

che è relativamente facile ingannare questi spider introducendo parole chiave ingannevoli nel sito stesso (è nota la tecnica, molto usata in passato, di introdurre nelle pagine principali, anche centinaia di volte una parola chiave, scritta in caratteri bianchi piccolissimi su fondo bianco, in modo da essere invisibili all'utente, ma da essere catturate dallo spider).

Una tecnica alternativa è l'approccio attraverso directory, strutture in cui i siti Web sono catalogati sulla base del contenuto, e indicizzati, in una struttura ad albero inserita in un data base, in categorie, sottocategorie, e sotto-sottocategorie all'infinito. Si tratta di un assetto monolitico difficile da maneggiare e riorganizzare - si immagini ad esempio di voler strutturare in questo modo una gigantesca collezione musicale o una biblioteca in base ai generi - ma questa struttura (un esempio classico è il motore Yahoo) ha un grande vantaggio; è possibile scorrere la directory per quel che interessa e, se ci si accorge che la ricerca si sta allontanando da quel che cerchiamo, ci sono buone possibilità che l'argomento non possa essere trovato.

## Le categorie dei motori di ricerca

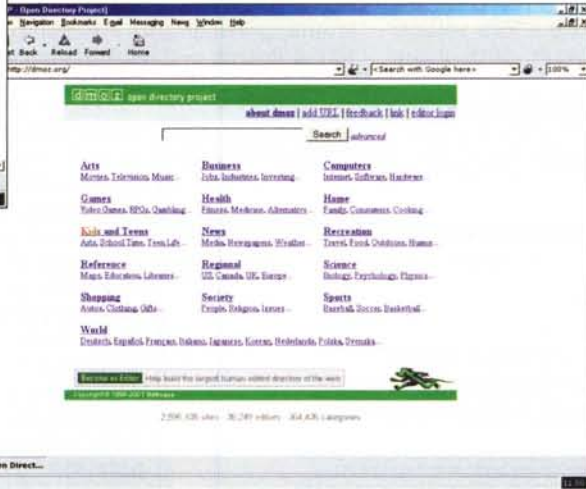
Allora, come organizzare questo articolo, in modo da poter dire qualcosa di utile sui mezzi che Internet mette a disposizione per cercare, e trovare con la minore fatica, argomenti, fatti, informa-



spider, che vanno in giro per il Web, per così dire, e, in teoria, analizzano ogni parola di ogni pagina, le conservano e strutturano da queste un indice di ricerca. Questa proposizione, del tutto teorica, oggi non ha possibilità di sopravvivere; nessun motore di ricerca è completamente comprensivo di tutto quello pubblicato sul Web ma il principio, ancorché ridotto nelle sue funzio-



senza alcun dubbio, **AltaVista** (<http://www.altavista.com>), uno dei più potenti e popolari, dotato di uno spider raffinato ed originale. Si è trasformato negli



zioni del tutto generiche, o magari assolutamente puntuali? Ecco, la struttura più adatta per questo articolo pare quella di trattare, per paragrafi successivi, tool, motori di ricerca, directory e quant'altro in base a una loro "specializzazione". Andremo quindi dai motori del tutto generali alle directory, ai motori regionali, a quelli specialistici, a quelli dedicati a discipline, età, hobby particolari. Insomma una specie di vademecum per chi desidera avere a portata di mano, in quel momento, il mezzo giusto al momento giusto per poter eseguire una ricerca che dia risultati di un certo significato e, soprattutto, eviti come spesso accade, di trovarci di fronte ad alcune centinaia di risultati, vale a dire, di fronte a nessuna conclusione efficace.

È ovvio che la precedenza, nella descrizione di questi tool, va data ai motori di ricerca veri e propri. Cosa dire di questi mattoni fondamentali del Web? Solo che, come tutti gli attrezzi fondamentali di qualunque mestiere, sono buoni a fare tante cose, ma i risultati che danno sono assolutamente generici. Quindi affidarsi a loro significa sapere fin dall'inizio che si può andare incontro a risultati tanto numerosi quanto inutili, e che per migliorare le proprie possibilità di successo occorre raffinare sempre di più la sintassi del seme di ricerca e provare, provare tante volte.

Il re dei re dei motori di ricerca è,

anni in un portale fastidiosamente affollato, e il disegno della pagina home non aiuta certo ad aumentarne la leggibilità. Veloce nelle ricerche, dotato di un database molto esteso e articolato, sempre ben aggiornato sull'evoluzione dei link, è dotato anche di tool aggiuntivi piuttosto interessanti tra cui un servizio di traduzione in linea cui, successivamente, anche altri motori hanno fatto riferimento.

Motore storico è anche **Excite** (<http://www.excite.com>), che ha diverse volte rinnovato il suo look. Oggi è un sito moderno, probabilmente un po' troppo affollato, e ha mutuato da Yahoo! la struttura pluricolonnare personalizzabile. Offre numerosi servizi integrati, peraltro necessari a questo tipo di motori per poter sopravvivere agli attacchi dei portali e delle directory, più accattivanti ed appetibili.

**Google**, (<http://www.google.com>) è l'enfant terrible di questa categoria. As-

surto alla ribalta recentemente per le sue notevoli qualità, ha uno spider che utilizza una tecnica del tutto inusuale, in quanto stabilisce l'importanza di un sito valutando e "contando" quanti altri siti "importanti" sono collegati ad essa. Sebbene questa tecnica possa essere considerata, sotto molti versi, significativa e sfuggibile ai trucchi ideati dai programmatori per essere riconosciuti dai motori e salire nella gerarchia dei "ritrovamenti", ha come contropartita il difetto che i nuovi siti incontrano molta difficoltà ad essere analizzati e valutati e, ancora, i siti specialistici con buona probabilità rischiano di non essere catalogati del tutto. Ciononostante si tratta di un tool di gran valore, ben costruito e dot-

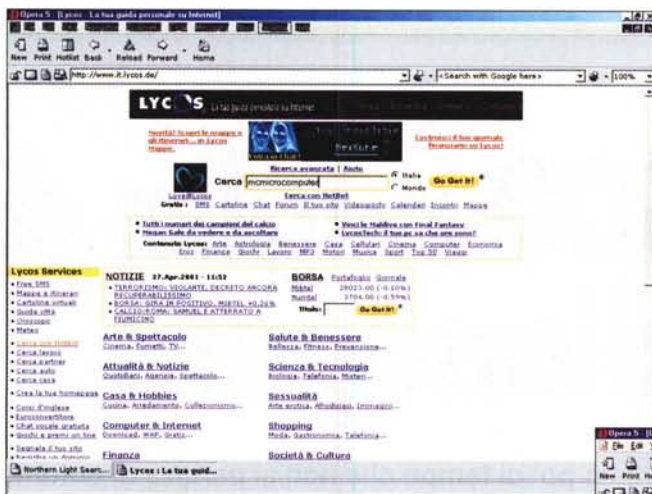
to di una serie di algoritmi complementari che cancellano pagine vecchie e riferimenti a collegamenti non più resistenti, con un database di circa un miliardo di pagine e, soprattutto, a una velocità assolutamente sorprendente.

**All the Web** (<http://www.alltheweb.com>) è il sito vetrina della tecnologia messa a punto dalla compagnia norvegese FAST (ricorsivamente, Fast Search &

Tracking). Si tratta di un motore dotato di una home page pulita, ordinata, e può essere utilizzato sia in maniera semplice sia attraverso il setup di numerose funzioni avanzate, che consentono di creare ricerche molto flessibili. Si tratta di un motore degno di nota e meritevole di attenzione.

**Raging Search** (letteralmente ricerca violenta, furiosa), <http://www. raging-search.altavista.com>, si autodefinisce il principale "sito di ricerca" del Web. Semplicissimo nell'interfaccia, "powered by Altavista" per quel che attiene il database e lo spider, adotta propri algoritmi per la ricerca di link perduti o siti abbandonati. Veloce (ma non quanto Google), non offre, nell'ottica della filosofia della massima semplicità, alcun servizio aggiuntivo, ma anche nessun banner pubblicitario. Prendere o lasciare!

**Oingo**, <http://www.oingo.com>, utiliz-



za un algoritmo esclusivo, chiamato "Oingo Ontology", attraverso cui cerca di interpretare il significato del seme di ricerca per fornire i risultati più intelligenti possibile. La storia dell'intelligenza, fin troppo sbandierata nel mondo Web,

potrebbe sembrare il solito specchio delle allodole, invece abbiamo potuto constatare, da alcune prove, che i risultati sono molto più precisi di quanto ci si potesse aspettare, e, comunque, al di sopra della media per quel che attiene ai puri motori di ricerca.

**AskJeeves**, <http://www.askjeeves.com>, o, semplicemente <http://www.ask.com>, da sempre ha pubblicizzato la possibilità di poter porre una domanda in linguaggio parlato (inglese), come, ad esempio, "What does Chicago have to offer in the arts?" o "Who is the winner of 1985 Pulitzer Prize?". In effetti, c'è ben poco di intelligente; dall'altra parte, il motore elimina solamente dalla frase le parole più comuni, come articoli, verbi, preposizioni, lasciando solo le stringhe più significative. Insomma, ben poco in più rispetto agli altri, anche se mascherato sotto belle maniere e frasi eleganti (I risultati sono offerti con la frase "People with similar questions have found these sites relevant: "). Insomma, un saper vendere la propria merce!

**SimpleSearch**, <http://www.simplesearch.com>, è costruito in maniera originale e, soprattutto, rispettosa dell'utente. In default si giunge a una finestra di ricerca spoglia, essenziale, che fornisce risultati veloci anche se, come sempre, ridondanti. Cliccando il bottone "Directory based search", in alternativa, si accede a un servizio parallelo basato su una directory; il risultato è di avere due forme di servizio tagliate in base alle preferenze dell'utente. Praticamente nulli i servizi complementari, e manca quasi ogni forma di banning.

## E passiamo alle directory

L'altra grande famiglia delle pagine di ricerca, le directory, sono prestigiosamente rappresentate da alcune firme di grande credito. Sebbene meno numero-

se dei motori, le directory godono di una vastissima popolarità, grazie sovente anche agli innumerevoli servizi accessori offerti.

**Yahoo!**, <http://www.yahoo.com> è il leader mondiale del settore, e sovente funge anche da motore semplice, visto che offre risultati sia in un senso sia nell'altro. E' troppo noto per dover essere qui descritto, e dispone di versioni localizzate in base a quasi tutte le lingue del mondo. E' il punto di riferimento e la pagina di apertura di gran parte della comunità mondiale Web, e offre una infinità di servizi collaterali di grandissima qualità. E' stato il primo sito ad offrire la possibilità di personalizzare i contenuti delle sue pagine.

**Lycos**, <http://www.lycos.com>, inizialmente partito come semplice motore e successivamente trasformatosi in portale con servizi di directory, ha abbandonato da tempo la gestione personale del servizio basandosi, per i servizi di motore, su FAST (AllTheWeb, v. paragrafo sui motori) e, per quelli di directory, su Open Directory (v. succ.). Preciso e puntuale, offre numerosi servizi aggiuntivi.

**About**, <http://www.about.com>, si autodefinisce "L'Internet umano"; quasi del tutto privo di banner e pubblicità, con un'interfaccia moderna e pulita, permette ricerche anche basate sul semplice spoglio delle directory interne, sempre presenti in ordine alfabetico. I servizi offerti, tramite partner, sono buoni, soprattutto per quelli di advertising.

**Open Directory Project**, <http://www.dmoz.com>, ci fa sognare che forse la libertà individuale non è ancora morta, e si presenta, per questo, con lo slogan "humans do it better!". I titolari del sito proclamano di aver creato il servizio senza dover sottostare ad alcuna restrizione esterna, e la mancanza di strip pubblicitarie e di banner rutilanti conferma l'impressione iniziale. I risultati sono attendibili e la struttura ad albero

non inutilmente ramificata. I servizi aggiuntivi sono essenziali, ma curati giornalmente.

**Northern Light**, <http://www.northernlight.com>, si auto-proclama animato dal desiderio di indicare e classificare l'umana conoscenza.



za in uno standard unico e consistente, e di renderlo disponibile a chiunque, nel mondo, attraverso una sola finestra di ricerca. Il sito ha una struttura raffinata ed elegante ed alcuni servizi esclusivi, alcuni dei quali a pagamento. Del tutto originale il servizio di "Geosearch", limitato agli USA e al Canada, per cui, in una pagina, si può inserire un argomento-oggetto di ricerca e stabilire geograficamente il raggio entro cui questa cosa, prevedibilmente un oggetto, può essere trovata (il servizio è fornito da Vicinity Inc.).

**HotBot**, <http://www.hotbot.com>, è da tempo passato sotto la proprietà di Lycos, che però ne ha mantenuto l'originaria struttura indipendente. Anche qui la struttura è simile a quella vista nelle più comuni directory, con categorie non eccessivamente e inutilmente differenziate e servizi aggiuntivi ridotti all'essenziale.

## Conclusioni

Abbiamo appena sfiorato l'argomento, che offre ben ampio margine di discussione e confronto, sempre più specializzato. Nella prossima puntata avremo modo di vedere tool e mezzi più specialistici, per ricerche più efficaci e legate a particolari esigenze. A presto!

ME