

Motori di Ricerca

Introduzione

Come abbiamo detto nella parte introduttiva di questo modulo, il *Web* può essere visto come un immenso documento ipertestuale multimediale distribuito. Per accedere ad una pagina (o a una qualsiasi altra risorsa *Internet* (file multimediali, documenti in vari formati digitali, eccetera) pubblicata sul *Web* occorre conoscere il suo indirizzo (url). Nella maggior parte dei casi però l'utente *Web* non conosce l'url della risorsa *Internet* a cui vuole accedere. Solitamente il navigatore ha a disposizione soltanto informazioni generiche circa l'argomento della sua ricerca.

Alla fine dell'anno 2001 il *Web* conteneva circa 4 miliardi di documenti (pagine *Web*). In media, ogni giorno, vengono pubblicate sul *Web* 7 milioni di nuove pagine.

Immaginatevi di essere uno studente universitario che vuole utilizzare *Internet* per reperire informazioni sull'argomento trattato nella sua tesi di laurea, oppure immaginatevi di voler trovare l'offerta più conveniente per un mutuo prima casa, o ancora di voler acquistare un biglietto aereo per le vostre vacanze. Come è possibile scovare l'indirizzo delle pagine che contengono queste informazioni? Questo è in sintesi il problema che affronteremo in questo approfondimento.

Cosa è un motore di ricerca

Risulta evidente dalle considerazioni appena fatte la necessità di disporre di uno strumento automatico che ci metta in grado di ricercare informazioni in *Internet* in modo veloce, accurato ed economico. Uno strumento che svolga la stessa funzione dell'elenco telefonico quando dobbiamo cercare un numero di telefono o di un catalogo di libri quando dobbiamo cercare un volume in una biblioteca. Nel caso di *Internet* la situazione è però molto più complessa in quanto:

- *Internet* contiene un numero gigantesco di documenti che si trovano fisicamente distribuiti su tutto il pianeta;
- ogni giorno moltissimi documenti cessano di esistere e moltissimi altri vengono pubblicati per la prima volta;
- le risorse in *Internet* sono di molti tipi e hanno una struttura disomogenea;
- non esiste una chiave sola di ricerca come accade nel caso dell'elenco telefonico.

Attualmente gli strumenti di ricerca che meglio soddisfano i requisiti di velocità accuratezza ed economicità sono i cosiddetti motori di ricerca.

In estrema sintesi un motore di ricerca è un sito *Internet* che, partendo dalla descrizione dell'argomento oggetto della ricerca, è in grado di restituire un elenco (possibilmente vuoto se la ricerca non va a buon fine) di pagine (e relativi url) che hanno a che fare con quell'argomento.

- Come può un sito *Internet* ricercare informazioni sul *Web*?
- Come è possibile descrivere ciò che vogliamo ricercare?
- Come si fa a stabilire che una certa pagina *Web* è interessante per la nostra ricerca?

Queste sono solo alcune delle domande che dovrebbero nascere spontanee dalla definizione che abbiamo dato di motore di ricerca e alle quali cercheremo di rispondere in questo approfondimento.

Motori di ricerca e directory

Gli strumenti di ricerca disponibili su *Internet* sono sostanzialmente di due tipi: **directory** e motori di ricerca veri e propri (a volte chiamati anche *spider*).

Spesso viene usato impropriamente il termine motore di ricerca anche per le *directory*. In realtà *directory* e motori di ricerca sono due concetti distinti. Nella tabella riportata qui di seguito elenchiamo alcune caratteristiche basilari delle *directory* e dei motori di ricerca.

Motori veri e propri (<i>spiders</i>)	Directory
Automatici	Risultato di lavoro umano
Indicizzano url	Classificano interi siti
Mirano a indicizzare tutto il Web	Mirano a diventare guide alle migliori risorse Web
Si basano su algoritmi matematici	Si basano su giudizi di valore qualitativi
Interrogabili attraverso combinazioni di parole chiave	Percorribili per categorie e sotto-categorie

La confusione tra *directory* e motori nasce dal fatto che quasi tutti i motori ospitano una *directory* (Altavista ospita *Looksmart*, Excite ospita *Excite's Guide* ...) e quasi tutte le *directory* sono interrogabili per navigazione ma anche per parole chiave (anche se meno efficacemente).

Per chiarire la differenza tra *directory* e motori veri e propri eseguiamo una ricerca utilizzando prima una *directory* e poi un motore di ricerca.

Esempio di ricerca su directory

Supponiamo di voler ricercare informazioni sulla squadra di calcio dell'Inter. Utilizziamo la *directory* di Altavista (Figura 1) (<http://it.altavista.com/dir/>).

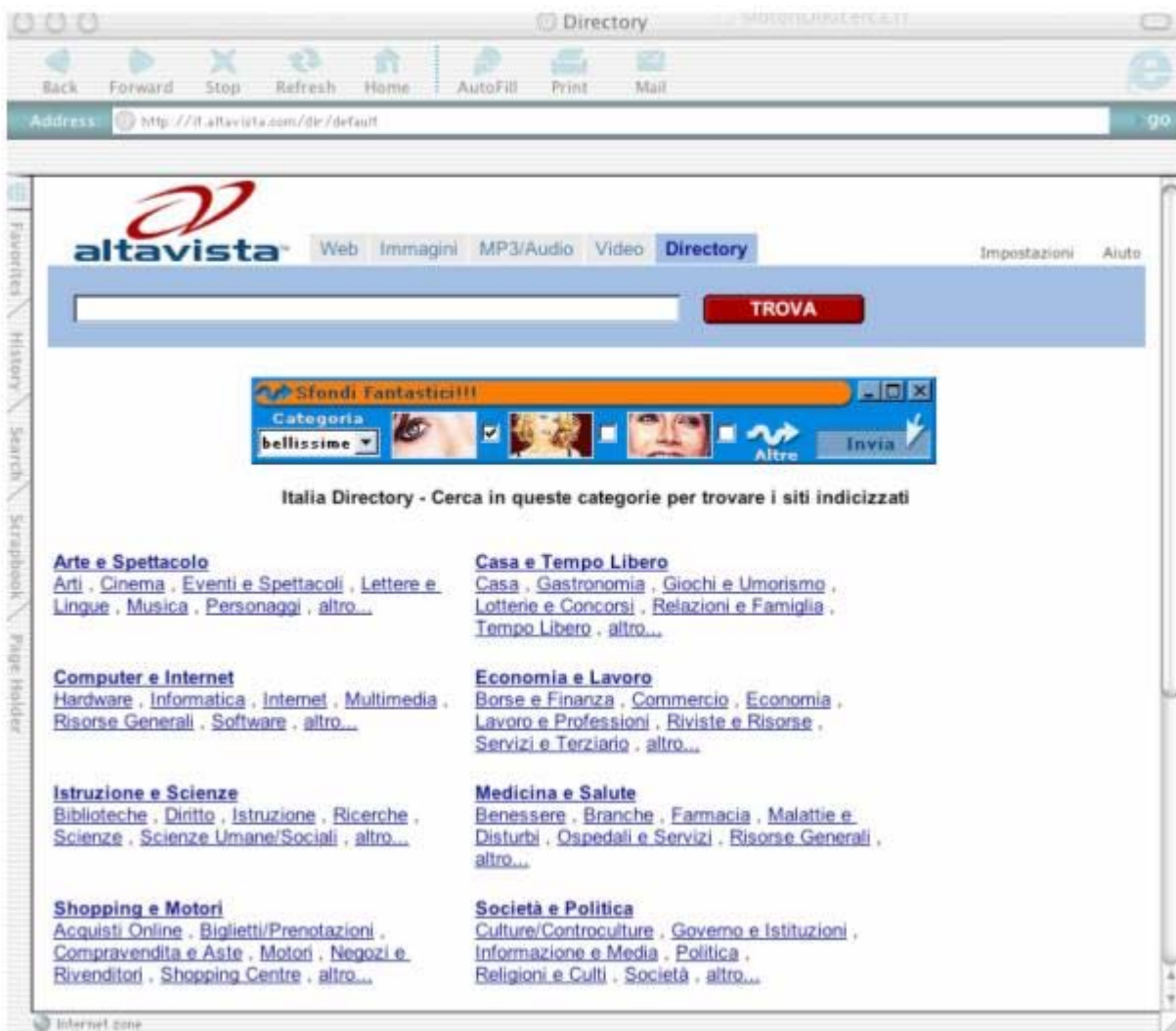


Figura 1. Pagina Iniziale della *directory* Altavista.

Successivamente, tra le varie categorie elencate, clicchiamo su Sport e *Fitness* ottenendo una nuova pagina parte della quale è raffigurata in Figura 2.

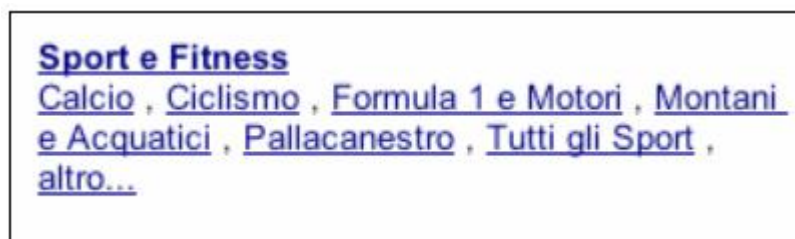


Figura 2. Pagina iniziale della *directory* >Sport e *Fitness*.

Poi clicchiamo tra le varie sotto-categorie di Sport e *Fitness* su Calcio ottenendo una pagina parte della quale è raffigurata in Figura 3.



Figura 3. Pagina iniziale della *directory* >Sport e *Fitness*>Calcio.

Clicchiamo adesso su Squadre....



Figura 4. Pagina iniziale della *directory* >Sport e *Fitness*>Calcio>Squadre

...e poi su Serie A...



Figura 5. Pagina iniziale della *directory* >Sport e *Fitness*>Calcio>Squadre>Serie A

...e infine su Inter ottenendo una serie di *link* (risorse con relativo url) mostrati in Figura 6.

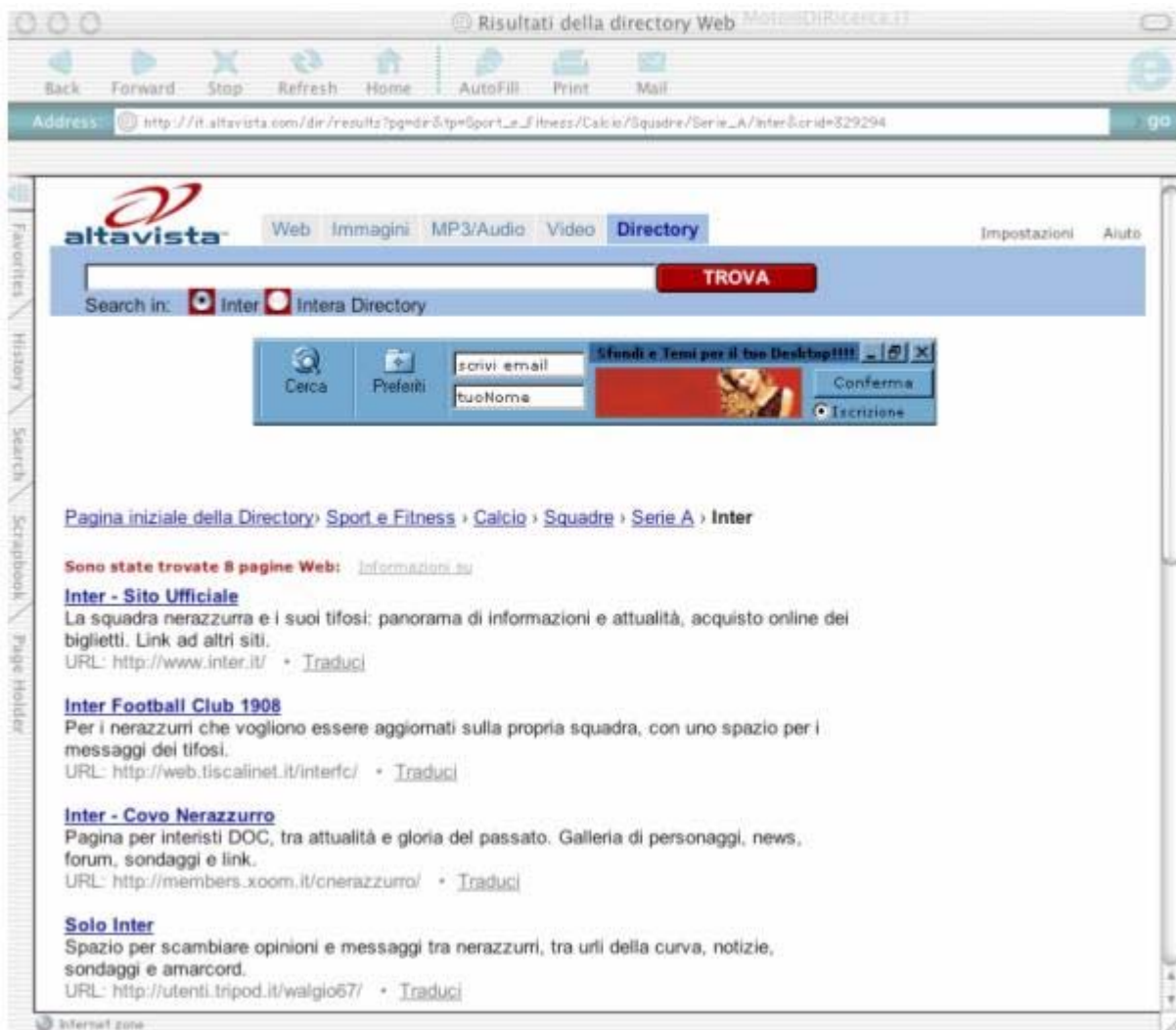


Figura 6. Risultato della ricerca ...

Esempio di ricerca su motore di ricerca

Cerchiamo ancora informazioni sull'Inter, però adesso utilizziamo un motore di ricerca vero e proprio. Per non far torto a nessuno questa volta usiamo il motore di ricerca *google* mostrato in Figura 7. (<http://www.google.it/>).

Per far ciò inseriamo la parola chiave *inter* e vediamo cosa ci restituisce *Google*...

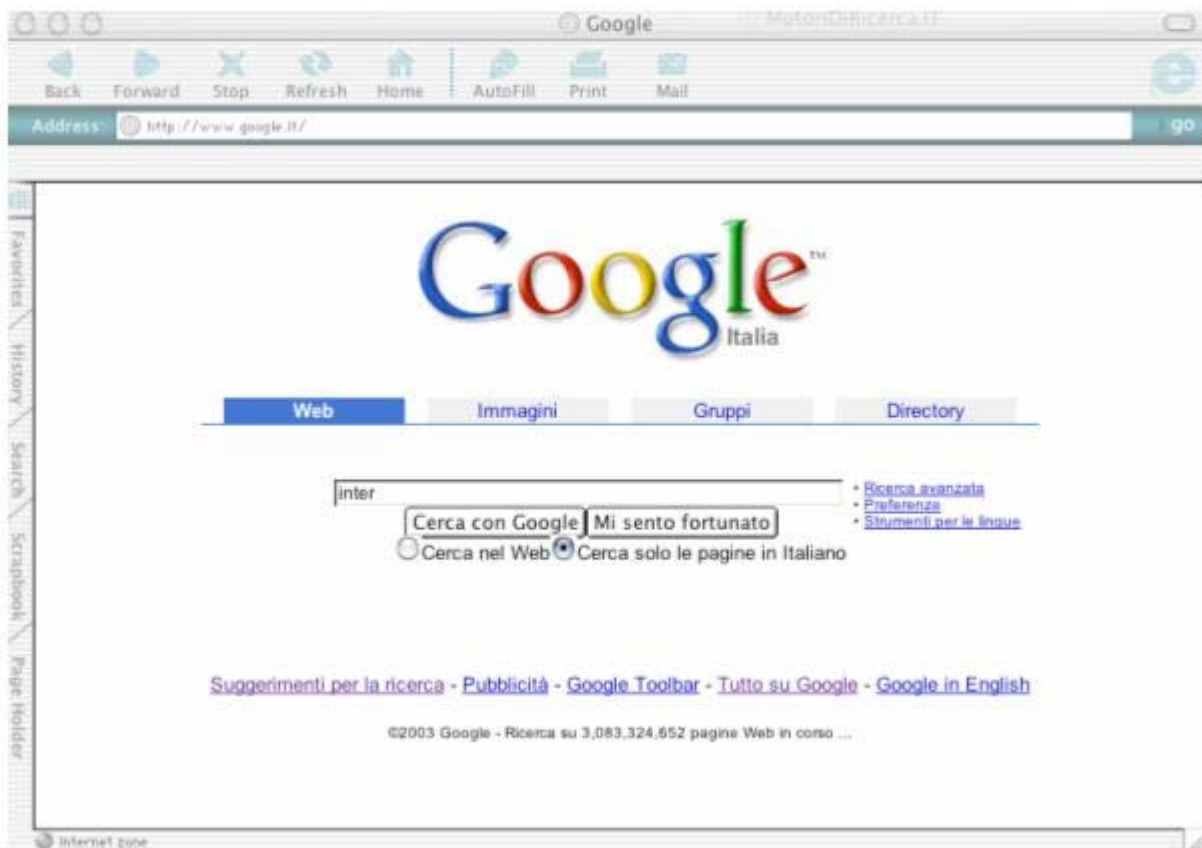


Figura 7. Pagina iniziale di *google*.

Otteniamo una lista (Figura 8) di 81800 *link* contenenti il termine *inter*. Siamo stati fortunati perché *Google* ci ha presentato i risultati in modo tale che tra i primi *link* proposti ci fossero proprio quelli riguardanti la squadra di calcio *Inter*. Non succede sempre così purtroppo...

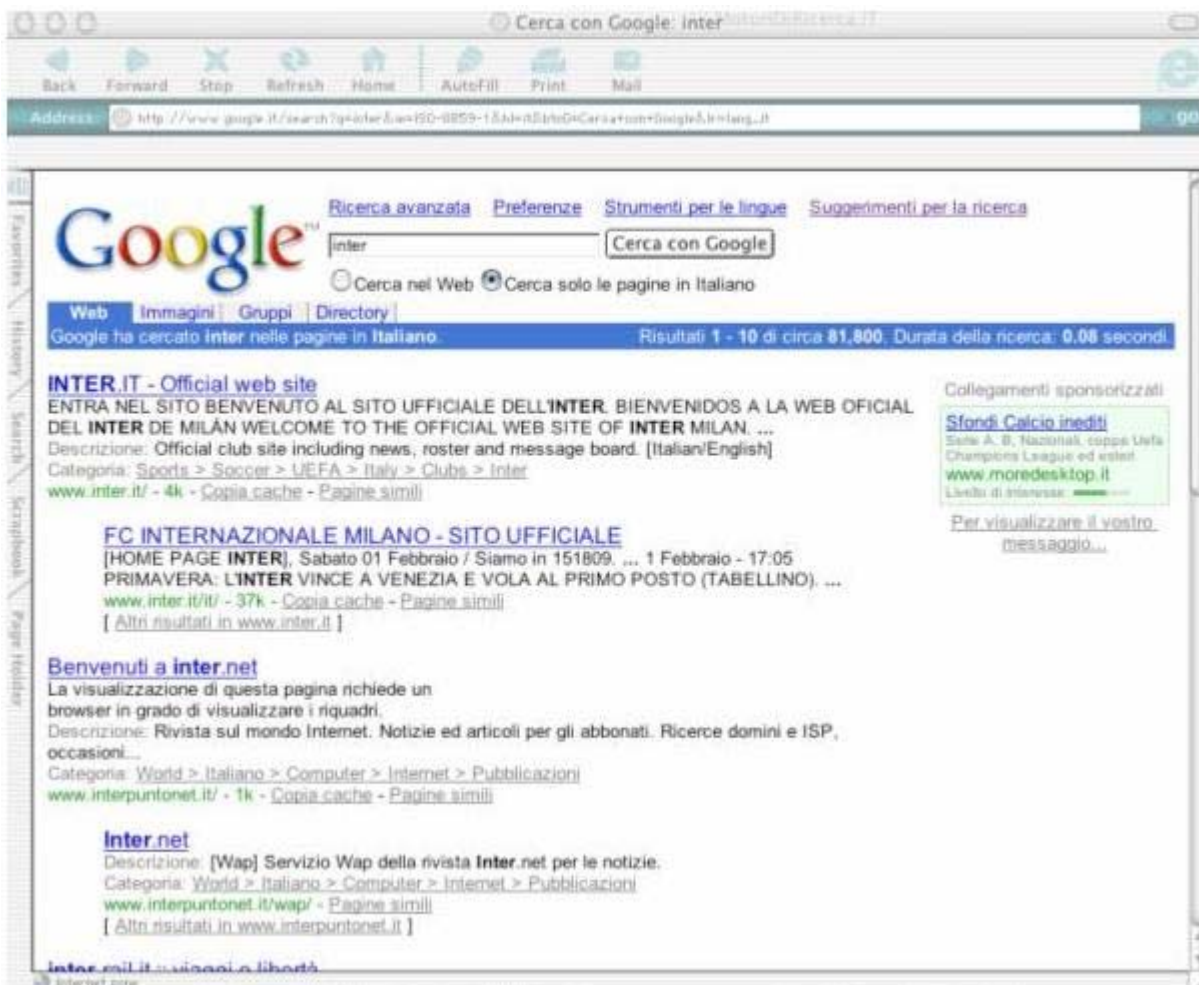


Figura 8. Risultato della ricerca con parola chiave inter

Se scendiamo nella lista troviamo molti altri *link* contenenti la parola inter, ma non inerenti la squadra di calcio come ad esempio il *link* al Centro Inter-Bibliotecario dell'Università di Bologna (Figura 9).



Figura 9. Altri *link* non inerenti alla società di calcio Internazionale FC

Tipicamente, i siti di ricerca più importanti mettono a disposizione sia *directory* che motori di ricerca. La scelta tra i due non è sempre facile e dipende sostanzialmente dal tipo di risorsa che vogliamo ricercare e da quanto dettagliata è l'informazione che abbiamo a disposizione. In prima approssimazione, se l'informazione è molto dettagliata probabilmente converrà usare un motore di ricerca. Se l'informazione è generica, usando motori di ricerca rischiamo di trovare migliaia di *link* molti dei quali totalmente scorrelati da ciò che stiamo cercando.

Qui di seguito elenchiamo alcuni motori di ricerca (italiani e internazionali) e per ognuno di essi indichiamo quali motori di ricerca e quale sistema di *directory* vengono usati.

Nome	Motore	Directory
<i>Yahoo.com</i>	<i>Google</i>	proprietario
<i>Google</i>	proprietario	DMOZ
<i>Altavista.com</i>	proprietario	<i>Looksmart</i>
<i>Looksmart</i>	<i>Inktomi</i>	proprietario
<i>InWind</i>	Arianna	DMOZ
Virgilio	<i>Google</i>	proprietario
<i>KataWeb</i>	<i>Inktomi</i>	proprietario

Ad esempio *Google* utilizza un motore di ricerca proprietario (da lui stesso inventato) e una *directory* basata sul sistema DMOZ, mentre Virgilio utilizza il motore di ricerca fornito da *Google* e una *directory* proprietaria.

Motori di ricerca

Un motore di ricerca si basa su diversi componenti:

- un programma che interroga periodicamente il *Web* per scaricare le pagine e catalogarle.
- Un *database* delle pagine catalogate.
- Una interfaccia di interrogazione.

Programma

Il programma che cataloga le pagine *Web* è il cuore del motore di ricerca. È continuamente attivo e non fa altro che navigare su tutto il *Web*, scaricare pagine, estrarre le informazioni dalle pagine scaricate che ritiene utili per la loro classificazione e inserire queste informazioni nel *database*. Questo procedimento è molto complesso, richiede molto tempo ed è cruciale per tutte le altre fasi della ricerca.

Database

Il *database* contiene l'informazione estratta dalle pagine *Web* esaminate dal programma. L'informazione contenuta nel *database* è organizzata in modo tale da permettere di ricercare in modo efficiente tutte le pagine che soddisfano i parametri di ricerca specificati dall'utente tramite l'interfaccia.

Interfaccia

Quando un utente si collega a un motore di ricerca interagisce con la sua interfaccia per definire i parametri della ricerca. Solitamente la ricerca avviene sulla base di una combinazione di parole chiave inserite dall'utente.

Ricerche per parole chiave

La ricerca più semplice è quella che consiste nel fornire al motore una sola parola chiave. Nell'esempio che abbiamo svolto all'inizio di questo approfondimento abbiamo fornito una sola parola chiave e la ricerca è andata a buon fine. Questo perché la parola chiave fornita specificava in modo preciso ciò che stavamo cercando. Non sempre è così, anzi. Nella maggioranza dei casi infatti una parola può essere utilizzata con diversi significati in diversi contesti.

Per questa ragione le interfacce di interrogazione dei motori di ricerca permettono di inserire più parole chiave legate tra di loro da un certo numero di operatori. Gli operatori più comunemente usati sono:

- PrimaParola *Or* SecondaParola: vengono ricercate le pagine che contengono almeno una tra PrimaParola e SecondaParola.
- PrimaParola *And* SecondaParola: vengono ricercate le pagine che contengono sia PrimaParola che SecondaParola.
- PrimaParola *But Not* SecondaParola: vengono ricercate le pagine che contengono PrimaParola e non contengono SecondaParola.
- PrimaParola SecondaParola: vengono ricercate le pagine che contengono la frase PrimaParola SecondaParola.
- PrimaParola *Near* SecondaParola: vengono ricercate le pagine che contengono PrimaParola e non troppo distante SecondaParola. È una restrizione dell'operatore *And*.
- PrimaParola*: vengono ricercate le pagine che contengono parole che iniziano con PrimaParola.

Directory

Una *directory* è un sito *Internet* che vi permette di selezionare progressivamente attraverso il meccanismo gerarchico delle categorie e delle sotto-categorie i *link* alle risorse che vi interessano.

All'inizio della ricerca l'utente si trova di fronte ad un elenco di categorie molto generiche quali ad esempio politica, sport, tecnologia, eccetera. Selezionando una di queste categorie l'utente implicitamente elimina dalla ricerca tutti i siti che non appartengono alla categoria selezionata.

Procedendo con questo sistema di raffinamento progressivo l'utente arriva abbastanza velocemente a determinare una categoria specifica contenente *link* di suo interesse.

Questo tipo di ricerca è particolarmente utile quando l'utente non ha molto chiaro dall'inizio cosa sta cercando e ha bisogno di rendersi conto di che tipo di informazione può trovare sulla rete circa l'argomento di suo interesse. Spesso l'utente ha solo voglia di girellare su *Internet* senza un preciso percorso in mente. Anche in questo caso il meccanismo della ricerca per categorie può rivelarsi utile.

È cruciale in una *directory* l'organizzazione delle categorie e delle sotto categorie. Una buona *directory* deve mettere in grado l'utente di capire velocemente ad ogni passo quale categoria selezionare e al tempo stesso deve permettere di raggiungere la categoria finale con il minor numero possibile di scelte.

Solitamente le *directory* sono aggiornate a mano.

Alcuni motori e directory

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

Yahoo! è la *directory* più grande del *Web*. Non è propriamente un motore di ricerca (anche se fornisce tale servizio appoggiandosi allo *spider* di *Google*), ma un catalogo di siti entro il quale si può navigare per categorie raffinando progressivamente la scelta degli argomenti proposti. È stato fondato nel 1994 presso la *Stanford University* da *David Filo* e *Jerry Yang* quando erano ancora studenti di ingegneria elettronica.

Il loro scopo iniziale era quello di classificare siti a cui erano particolarmente interessati. Adesso *Yahoo!* è un motore popolarissimo in diretta competizione con *Altavista*. È stato il primo motore di ricerca ad essere pubblicizzato al di fuori di riviste specializzate di informatica al pari di qualsiasi altro prodotto commerciale.

Yahoo! ha fatto della ricerca per categorie il suo punto di forza. L'albero delle categorie proposto da *Yahoo!* è molto più esteso e meglio organizzato che altrove. *Yahoo!* è lo strumento di ricerca ideale per tutti coloro che iniziano una ricerca senza avere un'idea chiara di cosa stanno cercando e delle parole chiave che potrebbero descrivere ciò a cui sono interessati. È sicuramente lo strumento di ricerca più usato da utenti non esperti.

Il catalogo dei siti di *Yahoo!* viene continuamente aggiornato aggiungendo siti proposti dagli stessi autori. I siti proposti vengono analizzati, valutati ed eventualmente inseriti nel catalogo. Per proporre un sito all'attenzione dei *surfer* di *Yahoo!* (gli esperti che valutano se inserire o meno il sito nel catalogo) basta andare sulla *home page* di *Yahoo!* e cliccare su *Proponi un sito con Yahoo! Express* (servizio a pagamento).

<http://www.google.com/>

Google è una delle aziende con la tecnologia di ricerca più interessante. Sviluppato all'università di *Stanford* da *Sergey Brin* e *Larry Page*, due giovani di 26 anni laureati in scienze informatiche, ha debuttato *on-line* a fine settembre 1999 dopo oltre un anno di test.

Google al contrario di *Yahoo!* ha fatto della ricerca per parole chiave il suo punto di forza. La caratteristica di *Google*, è quella di selezionare i risultati di ricerca valutando l'importanza di ogni pagina *Web* con metodi matematici, in base ad un controllo di oltre 500 milioni di variabili e di 2 miliardi di termini. Questa tecnologia, chiamata *PageRank* (brevettata da *Google*), controlla non solo il contenuto della pagina *Web*, ma verifica anche altri eventuali siti che hanno un *link* verso la pagina: in base alla quantità ed al tipo di *link*, la pagina riceve una valutazione più o meno alta.

Altre caratteristiche di *Google* sono la sua velocità di esecuzione, con un tempo medio dichiarato per ogni ricerca di 0,29 secondi, e la grandezza del suo archivio che ha oltre 3 miliardi di pagine.

La convinzione di fornire risultati rilevanti, è confermata dal bottone *I'm feeling lucky* (mi sento fortunato), attraverso il quale l'utente può essere direttamente portato sul primo sito che *Google* avrebbe indicato a fronte dei termini indicati.

L'interfaccia di *Google* è volutamente essenziale e priva di pubblicità.

<http://www.virgilio.it/>

Virgilio ha tutti i numeri per risultare una delle più importanti *Guide Internet*. Censisce circa 100.000 siti italiani con l'aggiunta di 150 nuovi siti ogni giorno, ed offre una vastissima gamma di

servizi funzionali e di orientamento, tali da considerarlo come un vero punto di partenza per ogni ricerca sulla rete, e non solo in Italia.

Il cuore di Virgilio è la sua *directory*, ossia un elenco di siti suddiviso in base alla loro tipologia, e strutturato in macro categorie a loro volta distinte in gruppi più analitici.

La registrazione dei siti è effettuata manualmente da uno staff interno a Virgilio (servizio a pagamento) e ciò permette una classificazione efficiente e coerente con la suddivisione in categorie. Sono censiti solo i siti in lingua italiana, con l'unica esclusione di quelli i cui contenuti sono ritenuti non idonei, ad esempio quelli pornografici, violenti, eccetera.

<http://www.overture.com/>

Overture è nata a fine 1997 a *Pasadena* in California ed è arrivata *on-line* il 21 febbraio 1998. Il suo fondatore, *Bill Gross*, già creatore nel 1988 del programma *Lotus Magellan*. Il CEO è *Jeffrey Brown*, in passato cofondatore di *City Search*.

Overture è sempre stato un motore di ricerca piuttosto originale rispetto ai più noti siti del settore, sia per caratteristiche che per strategia.

Il modello di *business* di *Overture* è quello di far pagare il posizionamento in testa ai risultati delle ricerche in base alle visite effettivamente generate. Ad esempio, un'azienda che vende vino *on-line* può fare in modo che il suo sito figuri in testa alle ricerche effettuate dagli utenti con i termini vino, *wine*, vini italiani, eccetera. Per ottenere le posizioni di testa, l'azienda si impegnerà a pagare a *Overture* una cifra per ogni click effettuato, in base ad una specie di asta. Se infatti un produttore di vini concorrente volesse figurare in posizione predominante, dovrebbe impegnarsi a pagare una cifra superiore.

I risultati delle ricerche forniti da *Overture* sono quindi piuttosto rilevanti, evidenziando siti di aziende certamente affini con le richieste fatte dall'utente, il quale peraltro ha ben evidenziate le posizioni a pagamento e l'ammontare che il destinatario è disposto a pagare a *Overture* per la sua visita.

La qualità delle risposte, è anche assicurata dall'utilizzo della tecnologia *Inktomi* per la gestione dei risultati di ricerca successivi a quelli a pagamento. Inoltre *Overture* non sembra voler mettere i vestiti del portal come hanno fatto tutti i principali motori di ricerca, per cui non vengono presentati altri contenuti sul sito se non quelli riguardanti le ricerche effettuate dall'utente.

Oltre ad implementare per primo il sistema di posizionamento a pagamento, *Overture* è stato anche il primo motore di ricerca ad introdurre un *Affiliate Program*, ossia un accordo con altri siti, i quali possono esporre un box di ricerca di *Overture* ricavandone una commissione per ogni ricerca effettuata dai loro utenti.

Motori tematici

I motori di ricerca di cui abbiamo parlato finora sono detti motori generalisti. Ciò significa che forniscono informazioni su risorse *Internet* di qualsiasi tipo e su qualsiasi argomento.

Esistono però anche strumenti di ricerca tematici o finalizzati che sono specializzati nelle ricerche di risorse *Internet* su determinati argomenti.

Ci sono motori ad esempio specializzati nella ricerca di *software*, di *file* musicali (ad esempio *file* in formato mp3), di annunci immobiliari, di offerte di lavoro e di molto altro ancora.

Questi motori solitamente riescono a fornire risultati più accurati sullo specifico tema trattato rispetto ai motori generalisti.

Un Esempio: *Tucows*

Tucows può essere considerato a tutti gli effetti un motore di ricerca per scovare nella rete il *software* che vogliamo. *Tucows* è l'acronimo di *The Ultimate Collection Of Windows Software*, ma è anche un gioco di parole, ovvero è l'equivalente della pronuncia inglese di *two cows* che significa due mucche; ecco il motivo dell'icona raffigurante per l'appunto due mucche simbolo ufficiale del sito <http://www.tucows.com/>, vedi Figura 10.

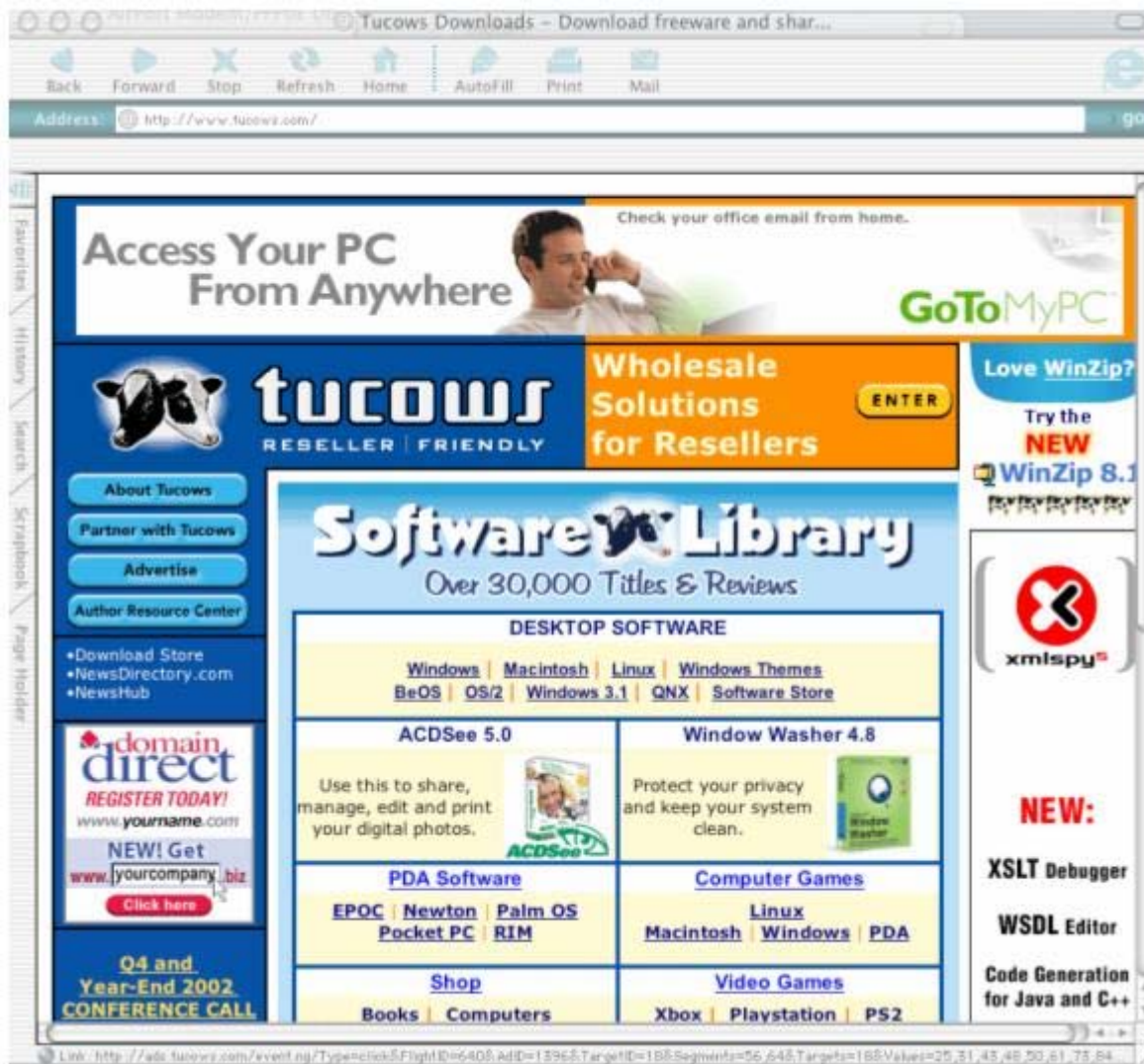


Figura 10. Pagina iniziale del sito *Tucows*.

In questo sito è possibile ricercare *software* sia *freeware* che a pagamento per ogni piattaforma (tra le principali citiamo *unix*, *linux*, *windows* e *macintosh*).

Meta-Motori

Infine, senza voler scendere in dettaglio vorremmo menzionare l'esistenza dei meta-motori. Un meta-motore non è altro che un sito che si appoggia sui servizi forniti da un certo numero di motori di ricerca. L'utente usa l'interfaccia del meta-motore per impostare i suoi parametri di ricerca e riceve in *output* dal meta-motore il risultato ottenuto mediante la consultazione di più motori di

ricerca.

Questo modo di ricercare informazioni può risultare più efficace rispetto alla consultazione di un singolo motore di ricerca nei casi in cui la nostra ricerca sia particolarmente difficile da portare a termine. I singoli motori di ricerca infatti coprono ciascuno non più del 35-40% di *Internet*. Usare più motori di ricerca contemporaneamente può quindi risultare utile.

Url di alcuni meta-motori:

- <http://www.thebighub.com/>
- <http://www.cerca.com/>
- <http://www.dogpile.com/>
- <http://www.highway61.com/>
- <http://www.metacrawler.com/>

Kartoo (<http://www.kartoo.com/>)

Kartoo è un meta-motore. La caratteristica principale di *Kartoo* è quella di rappresentare graficamente i risultati della ricerca sotto forma di una o più mappe (Figura 11). Inserite le parole chiave, *Kartoo* effettua la ricerca sui principali motori di ricerca in circolazione, mette insieme i risultati, e li riveste graficamente affinché possano essere utilizzati in modo semplice e intuitivo dall'utente.



Figura 11. Risultato della ricerca con parola chiave batistuta sul motore di ricerca *Kartoo*.

Il risultato della ricerca è visualizzato sotto forma di mappa di collegamenti.

La mappa restituita da *Kartoo* è composta da diverse sezioni.

Nella sezione circolare di Figura 11 abbiamo il risultato vero e proprio della ricerca composto da un

insieme di cerchi (nodi) di diverse dimensioni e da un insieme di collegamenti (*topics*) che collegano un cerchio ad un altro.

Ogni cerchio rappresenta un sito *Internet* e la sua dimensione è proporzionale alla suo grado di importanza all'interno della ricerca. Cerchi più grandi rappresentano pagine molto significative all'interno delle quali è più probabile trovare ciò che stiamo cercando. I collegamenti tra un nodo e un altro sono etichettati con una o più parole chiave che possono essere aggiunte alla ricerca (cliccandoci sopra) in fasi successive. La ricerca avviene quindi per raffinamenti successivi ottenuti inserendo o escludendo una sequenza di parole chiave (suggerite dalla mappa o inserite dall'utente).

Inoltre i siti più significativi vengono visualizzati al centro della mappa;

In ogni momento della ricerca è possibile **inserire il proprio sito nei motori di ricerca**.

Inserire il proprio sito nei motori di ricerca

Finora abbiamo trattato il problema di ricercare l'url di risorse in *Internet* fornendone una descrizione più o meno dettagliata attraverso vari meccanismi (parole chiave, categorie di appartenenza). Supponiamo adesso invece di voler affrontare il problema inverso e cioè come far sì che il maggior numero di utenti possibile raggiunga il nostro sito *Internet* per effetto di una ricerca su motori o *directory*.

Questa problematica diventa di fondamentale importanza nel momento in cui, ad esempio, si intraprenda una attività di commercio elettronico e si voglia quindi pubblicizzare il nostro prodotto:

- Come è possibile inserire il nostro sito in una *directory*?
- Come è possibile far sì che il nostro sito appaia nelle prime posizioni nella lista di url ottenuti da un determinato motore di ricerca?

Immaginatevi di voler ad esempio vendere via *Internet* un prodotto tipico della vostra zona geografica di provenienza. Come è possibile pubblicizzare la vostra iniziativa? Ovvero far conoscere a più persone possibili la vostra offerta commerciale?

Il problema non è nuovo. Chiunque voglia entrare nel mondo del commercio ha bisogno di pubblicità. Su *Internet* sono disponibili però strumenti specifici per acquisire notorietà. Rimangono validi gli strumenti pubblicitari classici: televisione, radio, giornali, manifesti, ma a questi se ne affianca uno nuovo come già avrete immaginato, ovvero *Internet* stessa tramite l'utilizzo dei motori e delle *directory*.

Non vogliamo entrare nei dettagli di questa problematica ampia e complessa, ma desideriamo comunque chiarire alcuni concetti di base:

- La prima cosa da dire è che un sito dotato di grafica eccezionale e di una struttura ben organizzata e completa, ma non inserito in nessun motore di ricerca rischia di non riscuotere il successo che merita!
- I motori di ricerca veri e propri navigano in *Internet* autonomamente cercando nuovi siti da catalogare senza bisogno che nessuno suggerisca loro di inserire un certo sito nel *database*. Le *directory* invece, al contrario dei motori di ricerca, svolgono una esplorazione del *Web* limitata in quanto tale attività viene svolta tipicamente da esseri umani (*surfer*) senza l'impiego di strumenti informatici automatizzati. Il metodo più efficace quindi per inserire un sito in una *directory* è senza dubbio quello di proporre il proprio sito per la catalogazione invece di sperare che esso venga notato e catalogato dalla *directory*.
- Spesso registrare un sito in una *directory* costa denaro, soprattutto se la *directory* in questione è una delle più importanti e diffuse.

- Quando si effettua una registrazione presso un motore o una *directory* occorre fornire informazioni sul sito che vogliamo registrare. Solitamente viene richiesto all'autore del sito di fornire parole chiave oppure di scegliere una o più categorie nelle quali inserire il sito. Lo stesso linguaggio HTML (linguaggio per scrivere pagine *Web*) mette a disposizione costrutti sintattici specifici per poter inserire le parole chiave in base alle quali poi la ricerca verrà effettuata.
- Alcuni motori (esempio *Google*) oltre ad associare ai siti le relative parole chiave operano anche utilizzando il cosiddetto *ranking*. Il *ranking*, come abbiamo già accennato, è un meccanismo di valutazione dei siti in base alla loro importanza alla loro visibilità e al numero di visite ricevute. I siti con un *ranking* più alto saranno visualizzati nelle ricerche in posizioni migliori.

Web nascosto o dinamico

Non tutte le pagine che visualizziamo sul nostro *browser* sono catalogate nei *database* dei motori di ricerche. Alcune di esse infatti non esistono! Per essere più precisi esistono solo per un certo lasso di tempo (il tempo durante il quale sono visualizzate sul vostro *browser*) solo perché la vostra interazione con un sito *Web* le ha generate.

Facciamo un esempio per chiarire ciò che abbiamo appena detto. Supponiamo di collegarci ad un sito che gestisce la compravendita di autovetture nuove ed usate. Il sito vi permetterà ad esempio di specificare marca e modello dell'auto che state cercando. Fatto questo il sito genererà dinamicamente una pagina (il suo codice html) contenente le offerte che fanno al caso vostro. Quella pagina è frutto di una interazione tra noi e il sito su cui stiamo navigando e viene generata ad-hoc da un programma che gira sul *server* dove il sito è pubblicato. Quella pagina ovviamente non è catalogata in nessun motore di ricerca.