

Ricerche in rete

World Wide Web è la risorsa Internet probabilmente più nota, e i suoi ritmi di espansione sono esponenziali. Le pagine informative immesse in rete riguardano gli argomenti più vari, e provengono da fornitori di informazione di natura assai eterogenea: dalle università alle industrie private (grandi e piccole), dai centri di ricerca ai negozi, dalle imprese editoriali ai partiti politici. Vi sono poi le numerosissime 'home page' personali del popolo di Internet.

Chi svolge una ricerca in rete si trova dunque davanti un duplice problema: *reperire* l'informazione cercata e *valutare* la sua correttezza, completezza, imparzialità.

Il secondo compito, assai delicato, dipende in parte dall'esperienza; un consiglio generale - una volta trovata una pagina informativa che reputiamo interessante - è quello di risalire sempre alla home page del sito che la ospita (su molte pagine sono disponibili apposite icone attive - altrimenti si può provare ad 'accorciare' progressivamente l'indirizzo nella barra delle **URL**, salendo di livello in livello nella struttura gerarchica del sito). In questo modo potremo in genere reperire informazioni su chi ha immesso in rete quella particolare pagina, in quale contesto e a quali fini.



Quanto al primo problema - quello di 'scoprire' le pagine esistenti che si occupano di un determinato argomento - una buona partenza è in genere rappresentata dagli **strumenti di ricerca** disponibili in rete. Vi sono *due tipi* di risorse che è bene conoscere e che, innanzitutto, occorre saper distinguere (molto spesso queste due diverse tipologie di risorse vengano mescolate e confuse, anche da parte di 'esperti' della rete): i **motori di ricerca per termini** e gli **indici sistematici**.

Sottolineiamo però l'importanza di un *terzo tipo* di ricerca, del quale è assai più difficile fornire un inquadramento generale: la *navigazione libera* attraverso pagine di segnalazioni di risorse specifiche.

È infatti quasi una norma di 'netiquette' che chi rende disponibili informazioni su un determinato argomento, fornisca anche una lista di link alle principali altre risorse esistenti in rete al riguardo.

Questo tipo di liste ragionate va naturalmente esso stesso cercato e trovato, cosa che in genere viene fatta usando indici alfabetici o cataloghi sistematici di risorse secondo le modalità sopra delineate. Una volta però che abbiamo individuato una di queste pagine-miniera di link specifici, potrà essere produttivo proseguire la nostra ricerca attraverso di essa.

Le risorse in tal modo segnalate presentano infatti due importanti caratteristiche: sono state scelte in maniera esplicita e ragionata, e la scelta è presumibilmente opera di una persona che conosce bene il settore in questione. Abbiamo¹ trovato comodo caratterizzare con l'espressione *navigazione orizzontale* questa terza modalità di ricerca su Web.

¹Vedi il libro in rete "Internet 2004" all'indirizzo

http://www.liberliber.it/biblioteca/c/calvo/internet_2004/html/01_indice.htm che illustra la realtà di Internet : non innovazione tecnologica ma soprattutto fenomeno sociale ed economico.

Motori di ricerca

Un motore di ricerca è uno strumento interattivo capace di aiutare l'utente a recuperare le informazioni di interesse all'interno del World Wide Web.

I **motori di ricerca per termini** permettono di ricercare parole o combinazioni di parole in un archivio indicizzato di documenti in formato digitale. In molti casi è possibile combinare le parole fornite utilizzando operatori booleani: attenzione, però, perché la sintassi corretta per utilizzare AND, OR, NOT e gli eventuali altri operatori disponibili varia da un motore di ricerca all'altro.

I motori di ricerca sono costituiti da database che contengono riferimenti a centinaia di pagine Web.

Tali database vengono costruiti mediante strumenti automatici, detti *crawlers* o *robots* o *spiders*, che navigano perennemente da sito a sito per raccogliere le informazioni e memorizzarle al loro interno.

Gli utenti possono interagire con il database mediante un'**interfaccia** all'interno della quale è possibile inserire delle parole chiave.

Successivamente il motore interroga il database, e come risultato mostra all'utente una lista di risorse che soddisfano la ricerca. Poiché i motori di ricerca utilizzano l'[ipertesto](#), gli utenti possono collegarsi direttamente alle risorse di interesse restituite come risultato.



Riassumendo un motore di ricerca consiste di tre componenti:

- un programma che naviga perennemente in Internet, di pagina in pagina, collezionando dati e associazioni. Questi software vengono chiamati *robots*, *spiders* o *crawlers*;
- un database creato dai robots contenente i dati raccolti (Internet Index) in modo tale da velocizzare l'accesso alle pagine di interesse per l'utente;
- un agente di ricerca (*Search Engine*) che è un software capace di accettare delle parole chiave inserite dall'utente e di interrogare il database costruito dai robots. Il *Search Engine* invia naturalmente all'utente i risultati dell'interrogazione in formato ipertestuale.

La ricerca attraverso un indice per termini è molto comoda nel caso di nomi propri, o nel caso in cui le informazioni che vogliamo trovare si lascino caratterizzare attraverso termini molto specifici. Occorre tuttavia tenere presente che si tratta di una ricerca meccanica: il programma utilizzato non farà altro che cercare i termini da noi forniti all'interno di un immenso indice alfabetico in suo possesso - indice tenuto aggiornato da un 'demone' software che si muove continuamente lungo la rete, seguendo ogni link incontrato e indicizzando tutte le pagine percorse - e fornirci le corrispondenze trovate.

L'intelligenza della ricerca dipende dunque in gran parte dalla scelta delle parole usate come parametri, anche se quasi tutti i motori di ricerca hanno la capacità di 'pesare' i risultati in base a elementi quali il numero di occorrenze della parola, l'occorrenza in zone significative del documento come i titoli o i link, l'*autorevolezza di chi ha consultato il documento* e così via.

Ciò significa che *se abbiamo scelto bene i nostri termini di ricerca*, riceveremo un elenco di pagine che avrà alte possibilità di iniziare da quelle per noi più significative.

Va ricordato, inoltre, che per quanto estesa la base di indicizzazione di un motore di ricerca per termini copre solo una parte delle pagine realmente disponibili in rete. I dati effettivi sono molto difficili da stimare. Una ricerca del servizio Search Engine Watch (archiviata all'indirizzo <http://searchenginewatch.com/2156481>) mostrava che nessun motore di ricerca arrivava a coprire più del 25% del numero complessivo di pagine in rete, valutato nel febbraio 1999 attorno agli 800 milioni. ... ma più ci si avvicina a oggi, più queste stime sono incerte.

Google indicizzava circa un miliardo e mezzo di pagine a metà 2002, e a maggio 2003 dichiara di indicizzare oltre tre miliardi di pagine Web. È probabile che nessun motore di ricerca arrivi a coprire più del 30-35% del numero complessivo di pagine in rete: possiamo dunque pensare che, a metà 2003, il numero complessivo di pagine Web sia a grandi linee compreso fra i 7 e i 10 miliardi. E ricordate che il Web non è fatto solo di pagine HTML: ci sono documenti di ogni genere, immagini, file audio...



Sono numerosi i motori di ricerca disponibili oggi sul WWW. Nonostante tutti i motori di ricerca siano progettati per svolgere il medesimo compito, ognuno di essi realizza tale compito secondo modalità proprie conducendo spesso a risultati sorprendentemente diversi per uno stesso tipo di interrogazione. I fattori che influenzano i risultati sono la dimensione del database, la frequenza di aggiornamento e la capacità di ricerca. I motori di ricerca differiscono tra loro anche in base alla loro velocità di ricerca, al progetto dell'interfaccia di interrogazione del database, al modo di mostrare i risultati ed alla quantità di help in linea che offrono.

Un'altra modalità per compiere ricerche in rete prevede l'uso di **indici strutturati** (ad esempio il più famoso [Yahoo!](#) o Virgilio) che, pur essendo meno esaustivi, sono però più facili da usare ed evitano il rischio di fornire troppe informazioni

Al contrario della ricerca alfabetica, la ricerca sistematica avviene su cataloghi *ragionati* di risorse, suddivisi per settori e organizzati gerarchicamente: in genere la base dati è più ristretta (non saremo sicuri di trovare direttamente tutte, o anche solo la maggioranza delle pagine che ci interessano), ma la valutazione della pertinenza o meno di una determinata informazione non sarà più meccanica, bensì risultato di una decisione umana, e l'informazione stessa sarà inserita all'interno di una struttura di classificazione. Naturalmente, in questi casi i principi utilizzati per costruire l'impianto sistematico della banca dati sono fondamentali. Un catalogo ragionato di questo tipo si basa in genere su una sorta di 'albero delle scienze', da percorrere partendo da categorizzazioni più generali per arrivare via via a categorizzazioni più specifiche. Ed è importante che questo percorso di 'discesa al particolare' avvenga attraverso itinerari intuitivi e coerenti - compito naturalmente tutt'altro che facile.

Gli indici sistematici di risorse sono strutturalmente più simili, in realtà, a grafi complessi che ad alberi: ad uno stesso nodo si può arrivare attraverso percorsi alternativi, tutti egualmente validi. Dal punto di vista dell'utente, invece, ciò comporta semplicemente che - a meno di non andarla a cercare sotto categorie palesemente innaturali - trovare una determinata risorsa informativa sarà assai facile: se ben compilato, l'indice sembrerà 'adattarsi' alle nostre scelte di categorizzazione.

La differenza tra questi due tipi di ricerca è sostanziale, nonostante sia invalso l'uso - concettualmente fuorviante - di utilizzare per entrambi il termine 'motore di ricerca', e la confusione sia accresciuta dal fatto che molti indici sistematici, come Yahoo!, permettono l'accesso *anche* a un motore di ricerca per termini, e viceversa.