

Web nascosto. Il 50 per cento dei contenuti non viene rilevato dai motori di ricerca. E oltre il 10 per cento delle informazioni sono perse ogni anno. Per una buona ricerca online bisogna consultare più fonti
(*Corriere Economia, dicembre 2006*)

Chi pensa che i motori di ricerca consentano di esplorare "in toto" il cyberspazio del web, ebbene si sbaglia. Perché *solo il 50* per cento dei contenuti Internet vengono rilevati nel corso delle ricerche. Il rimanente rimane invisibile, non raggiungibile. Ma non per questo inesistente. Dunque nella rete si trova una grande zona d'ombra inaccessibile ai vari Google, Yahoo, Msn e Altavista. Però, quando digitiamo sullo schermo una parola, il nome di una persona o una frase da cercare e abbiamo un numero limitato di recursività (qualche volta nullo), questo non significa che la risposta non esista. A rivelarlo è Albert Laszlo *Parabasi*. Autorevole professore di fisica dell'università Notre Dame nell'Indiana, nonché direttore del Center of Complex Network Research (Ccnr). Uno dei massimi teorici del web, che nel suo ultimo libro "Struttura e dinamiche della rete", cerca di dare risposta alla complessa organizzazione di Internet.



Il web dunque, si presenta con una struttura a guscio. Una specie di cipolla da sfogliare. La parte superiore facilmente raggiungibile e zone più interne difficili da penetrare. A meno che non si conoscano gli indirizzi diretti, i cosiddetti Url (Uniform resource locator). Allora sarà lo stesso browser di navigazione come Explorer, Opera o Firefox, tanto per citare i più usati, a farci navigare all'interno delle pagine.

«Ma la situazione in termine di perdita dei dati potrebbe peggiorare - spiega **Gianroberto Casaleggio**, esperto milanese di strategie della rete - perchè ogni anno almeno il 10 per cento delle informazioni svanisce nel nulla. Come inghiottito da un buco nero». Sono dati e contenuti web non più raggiungibili. I motivi sono i più disparati. Semplicemente il vecchio sito di riferimento cambia indirizzo e quello nuovo non è ancora stato rilevato dai motori. Oppure sono gli stessi database, i portali governativi e aziendali, le pagine di intrattenimento a non riaggiornare i link presenti in milioni di pagine. Collegamenti che cambiano continuamente in modo dinamico. E in questi casi ai cybernaviganti esce sullo schermo il messaggio sibillino "404 Not Found".

Così Google nel 2004, proprio per colmare le sue lacune, ha creato **Google Scholar**, un motore che effettua le ricerche tra le pubblicazioni (working-paper) universitarie e materiale accademico, come le tesi di laurea, normalmente seppellite in fondo al web.

Non si tratta però di novità. Perché per questo tipo di ricerche esistevano già "Web of Science", creato dalla Thomson Scientific, e Scopus dell'editore Elsevier. Rispetto ai quali Scholar presenta però il vantaggio di risultare gratuito. Ma nessuno di questi tre è perfetto. Una ricerca svolta a fine 2005 da quattro bibliotecari dell'Università di Yale – Nisa Bakalbasi, Kathleen Bauer, Janis Glover e Lei Wang – dimostra che non è facile stabilire quale sia la fonte migliore e che dipende: «dalla materia e dall'anno di pubblicazione dell'articolo che si sta cercando». Ergo per un'informazione completa vanno consultate tutte e tre.

Portare a galla il profondo web è la ragione di vita di un'altra realtà Internet, il progetto **OAIster**. Un gioco di parole tra OAI (Open access iniziative), un'iniziativa internazionale per il libero accesso via Internet a testi accademici, e oyster, ostrica. «...trovare le perle» è il sottotitolo di questo sito partito da un'iniziativa della biblioteca digitale dell'Università del Michigan e che raccoglie oggi l'adesione di 712 istituzioni di tutto il mondo. Per l'Italia vi partecipa, per ora, soltanto un pugno di università: Bologna, Napoli, Parma, Pisa. Ma a livello mondiale il progetto continua a ricevere adesioni, con un archivio di quasi 10 milioni di documenti. «Anche se OAIster è senza



dubbio un motore di ricerca, non siamo esattamente concorrenti di Google o Yahoo – spiega Kat Hagedorn, una delle responsabili del progetto – semmai, siamo complementari rispetto a loro. Cerchiamo quei documenti che gli altri motori di ricerca non trovano». Così OAIster è una realtà di cui Google e Yahoo non possono non tenere conto. Ecco perché entrambi verificano i suoi "metadati", cioè le informazioni relative a ciascun sito, che vengono poi indicizzate dai motori di ricerca e su cui si basano confronti e risultati finali.

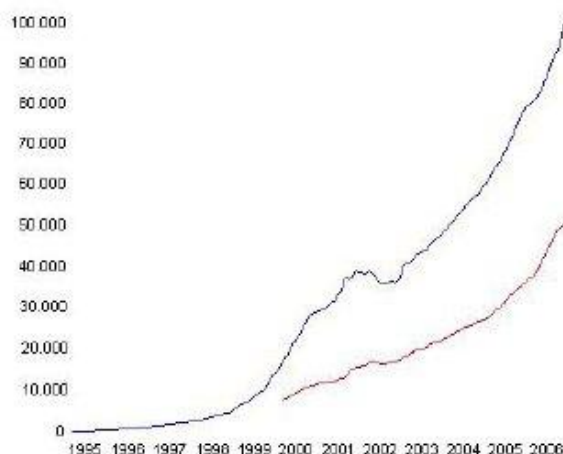
Al buon ricercatore online non rimane quindi che rassegnarsi all'uso simultaneo di più motori. «Ecco perché è bene controllarne più di uno – conclude Gianroberto Casaleggio – verificando ovviamente che la fonte di riferimento non sia la stessa, ottenuta semplicemente duplicando un unico documento». Perché i danni derivati da notizie errate, riprese in sequenza da più pagine, risultano incalcolabili. In termini di disinformazione prima, e caduta dell'immagine poi, nel caso si tratti di riferimenti personali.

Ne sa qualcosa, a proposito, John Seigenthaler. Ex collaboratore di Robert Kennedy. Wikipedia (la maggiore enciclopedia online del mondo) lo aveva erroneamente inserito tra le persone coinvolte all'assassinio. Lampante il caso. Qualcuno aveva volutamente "postato" l'informazione calunniosa. Ebbene John ha ottenuto l'azzeramento della stessa da parte dei responsabili di Wikipedia. Ma non ha potuto fare nulla nei confronti di tutti i siti che avevano preso spunto dalla pseudo-notizia "originale".

Ricerca alternativa, adesso sono di moda i "metamotori"

Il mercato globale della ricerca è per la maggioranza dominio di Google (62 per cento). Il principale concorrente, secondo i dati rilasciati da ComScore prima dell'estate 2006 è Yahoo. Con una quota del 20 per cento. A distanza Seguono Msn, Aol, Ask Jeeves, e altri semiconosciuti, ma non per questo meno efficaci. E' il caso dei "*metamotori*", così chiamati perché effettuano ricerche in rete interrogando contemporaneamente più motori di ricerca.

Siti web nel mondo – 1995-2006
numeri in migliaia



(Fonte: Netcraft 2006 - La linea rossa indica i siti che risultano effettivamente "attivi")

«I metamotori – così li definisce Wikipedia – non operano con una propria banca dati o attraverso ricerche dirette, ma con una base dati virtuale». Cioè ricevono la domanda dall'utente, la girano a diverse database e quindi raggruppano i risultati in modo omogeneo. Metacrawler, ad esempio, invia le

richieste a Google, Yahoo, Windows live search (la nuova versione di Msn), Ask.com, About.com e altri. I metamotori sono molti. Per un elenco completo è bene consultare Wikipedia alla voce "Metasearch engine", e secondo l'enciclopedia online: «non ne esiste uno uguale all'altro. Alcuni basano le loro ricerche soltanto sui motori più popolari, altri passano al setaccio anche quelli meno noti e i newsgroup».

Di particolare efficienza è il metamatore *Clusty*, sviluppato da tre programmatori della Carnegie Mellon University. La sua principale caratteristica consiste nel raggruppare i risultati in cartelle o grappoli. Cluster in inglese. Ogni cartella contiene i risultati della ricerca, catalogati per affinità con la parola indicata. Per esempio, inserendo la parola "pavese", otteniamo i seguenti cluster: Cesare Pavese (97 risultati), Oltrepo pavese (64), provincia pavese (5), Frank Pavese (3). Abbiamo così una panoramica di tutti i risultati. Non più visualizzati in un lungo elenco ma divisi per argomenti. Un sistema concreto per mettere ordine nella Babele di Internet. E aiutare i naviganti. Spesso disorientati dalle migliaia di pagine proposte.

#