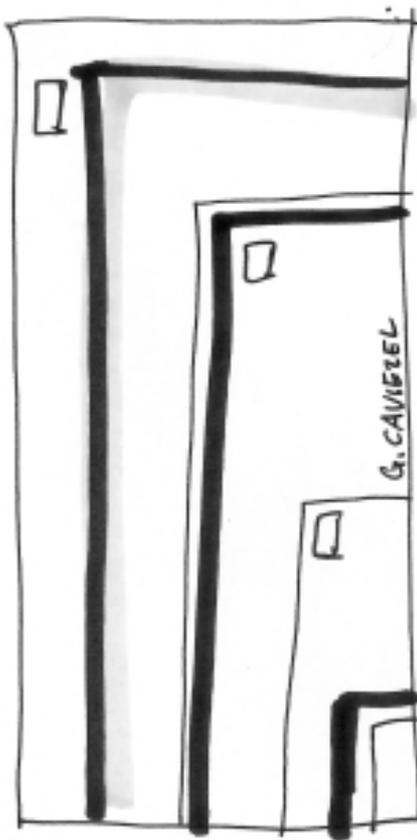


Strumenti e strategie per la ricerca di informazioni www

Una rassegna per tipologie

di Riccardo Ridi

Il World wide web, col suo formato costitutivo HTML, non costituisce certo la totalità di quanto è disponibile in rete (ci sono anche le mailing list, i newsgroup, i server FTP, gli OPAC e i database interrogabili via Telnet, i libri e i periodici elettronici in formato PDF, i file audio e video, gli emergenti formati SGML e XML ecc.), ma sicuramente costituisce la parte di Internet cresciuta più rapidamente negli ultimi anni (c'è chi stima che abbia superato all'inizio del 1998 i trecento milioni di pagine, nel settembre 1999 gli ottocento milioni e che abbia ormai abbondantemente superato la boa del miliardo di pagine). Inoltre la maggior parte delle risorse "non-www" disponibili in rete difficilmente rinuncia a crearsi almeno una "finestra sul web", ovvero una pagina HTML con una breve descrizione della risorsa stessa e un link che permetta al browser di attivare automaticamente il software necessario per raggiungerla. Riuscire a trovare informazioni disponibili sul web, insomma, tende sempre più a coincidere col recuperare informazioni *tout court* disponibili via Internet. Esistono numerose ti-



pologie di strumenti creati, mantenuti e messi a disposizione gratuitamente, spesso grazie ai proventi pubblicitari, per facilitare la ricerca in rete.

Motori di ricerca

Indici web per parola, detti anche "motori di ricerca", "search engine", "robot", "spider" o "worm". Grazie ad appositi programmi raccolgono automaticamente in grandi archivi centinaia di milioni di pagine web e permettono di individuare al loro interno tutte le occorrenze di uno o più termini digitati dal ricercatore in una apposita maschera. Talvolta i motori ricercano anche tra i messaggi di numerosi newsgroup. I risultati trovati, purtroppo, non sono sempre pertinenti, anche se questi strumenti vengono continuamente perfezionati per consentire interrogazioni e metodi di ordinamento sempre più sofisticati. Il problema consiste nella difficoltà di far individuare automaticamente a un programma, senza l'intervento di catalogatori umani, omonimie, sinonimie e più in generale il contenuto semantico delle pagine web.

Antenati. Veri e propri progenitori degli attuali motori di ricerca web sono stati VERONICA (Very Easy Rodent-Oriented Net-wide Index to Computerized Archives) e Archie, due fra i più antichi strumenti di ricerca disponibili in rete, oggi in gran parte abbandonati. Con un client Archie (oppure collegandosi via Telnet a uno dei suoi residui server) si possono ricercare in tutti i siti FTP mondiali i file che contengono nel proprio nome una determinata sequenza di caratteri; con ciascuna VERONICA (cfr. ad esempio quella ancora mantenuta dall'Università di Pisa <gopher://serra.unipi.it:2347/7>) si possono invece ricercare (anche con gli operatori booleani AND, OR e NOT) tutte le parole contenute nei menu dei non molti Gopher sopravvissuti alla diffusione dei web.

Principali. I maggiori motori di ricerca internazionali sono attual-

mente AltaVista <<http://www.altavista.com>> e HotBot <<http://www.hotbot.com>>, entrambi dotati di numerose opzioni per raffinare la ricerca (operatori booleani, “frase esatta”, *case sensitivity*, troncamenti, ricerca per campi, limitazioni, ecc.), con comandi testuali nel primo caso e grafici nel secondo, FAST <<http://www.alltheweb.com>>, col più vasto archivio di pagine ma con poche opzioni di ricerca, Google <<http://www.google.com>>, che punta molto sull’individuazione dei siti più rilevanti mediante l’analisi citazionale dei link in entrata e in uscita, e Northern Light <<http://www.northernlight.com>>, che permette di cercare anche decine di migliaia di articoli (denominati “special collection”) provenienti da periodici e libri a stampa (il cui testo integrale è visualizzabile a pagamento) e dotato di un innovativo metodo di ordinamento dei risultati in base a criteri variabili in base alle caratteristiche comuni ai siti rintracciati. Fra i motori di ricerca specializzati nel “dragare” il webspace italiano si possono citare invece Arianna <<http://arianna.iol.it>> e Il trovatore <<http://www.iltrovatore.it>>.

Pagine simili. Alcuni motori, come Excite <<http://www.excite.com>>, Infoseek <<http://www.infoseek.com>>, Lycos <<http://www.lycos.com>> e lo stesso Google permettono di rintracciare (con una ricerca che viene talvolta definita impropriamente “per concetto” ma che andrebbe chiamata “per somiglianza”) ulteriori pagine web contenenti una certa percentuale di parole uguali a quelle delle pagine già trovate, nel tentativo di recuperare parte del contenuto semantico sfuggito con la ricerca per parole.

Relevance ranking. Quando si effettua una ricerca su centinaia di milioni di documenti con tecniche di interrogazione non particolar-

mente sofisticate come quelle in dotazione ai motori di ricerca web, ci si può facilmente aspettare di ottenere una quantità esorbitante di risultati, non sempre facilmente “raffinabili” con ulteriori interrogazioni. Diventa allora di importanza capitale il criterio utilizzato dal motore per effettuare l’ordinamento (*ranking*) dei risultati. Il criterio più diffuso è quello che i motori stessi amano chiamare “ordinamento in base alla rilevanza” (*relevance ranking*), e che in realtà cela un complesso e mutevole¹ mix (tenuto rigorosamente segreto per motivi commerciali) di criteri puramente quantitativi e statistici, fra cui i seguenti:

- 1) frequenza – numero di occorrenze dei termini utilizzati per la ricerca all’interno della singola pagina recuperata;
- 2) densità – rapporto fra il numero di occorrenze dei termini all’interno della pagina e il totale delle parole contenute nella pagina;
- 3) rarità – rapporto inverso fra il numero di occorrenze dei termini utilizzati per la ricerca all’interno del database del motore e il totale delle parole contenute nel database stesso. Alcuni motori dispongono anche di una lista di *stopwords* che vengono ignorate;
- 4) compresenza – presenza, nella pagina, di più d’uno dei termini richiesti o addirittura di tutti. In questo modo l’operatore booleano AND contribuisce, in una certa misura, al risultato finale anche se nella ricerca iniziale, come spesso avviene, è impostato di default l’OR;²
- 5) prossimità – presenza, nella pagina, di più d’uno dei termini richiesti – o addirittura di tutti – nella medesima frase o comunque vicini fra loro;³
- 6) priorità – ordine in cui sono stati immessi i termini nel form di ricerca;
- 7) posizione – presenza dei termini in particolari zone della pagina.⁴

Nella sezione visibile BODY vengono tipicamente “premiati” le prime parole e quelle contenute nei tag H e HREF (ovvero nei vari titoli, *heading*, e nei link). La sezione nascosta HEAD è complessivamente “premiata” più di quella BODY. In particolare vengono considerati molto “pesanti” il tag TITLE e, in subordine, quelli KEYWORD e DESCRIPTION;

8) aggiornamento – a parità di requisiti, alcuni motori privilegiano le pagine più recenti;

9) popolarità – numero dei link “in entrata”⁵ provenienti da altre pagine contenute nel database del motore e numero delle selezioni della pagina effettuate dagli utenti a partire dalle schermate di risposta del motore. Si tratta di un criterio di recente introduzione e ancora poco utilizzato;

10) tariffazione – AltaVista ha recentemente annunciato che sarà possibile, pagando, garantire una posizione di rilievo alle proprie pagine nella schermata di risposta. C’è chi sospetta che tale criterio sia già utilizzato da vari motori senza dichiarazioni esplicite, almeno per quanto riguarda pagine relative a prodotti propri o di aziende collegate.

Metadata. Fondamentale è l’importanza, ai fini del recupero dell’informazione disponibile in rete, della presenza di appropriati “metadata” e “metatag”⁶ nelle pagine web, che contribuirebbero in maniera determinante a ridurre quell’*information overload* di cui molti si lamentano. Esistono a tale proposito vari progetti, fra cui il più noto è lo standard Dublin Core (cfr. le FAQ <<http://purl.oclc.org/dc/education/>> e la traduzione italiana <<http://www.aib.it/aib/lis/std/t9503.htm>>), molto apprezzato negli ambienti bibliotecari. Attualmente però i principali motori di ricerca prendono in considerazione una serie di “metatag” molto meno ➤

s sofisticati ma ben più diffusi, fra cui i seguenti:

1) AUTHOR – indica l'autore della pagina, in genere identificato con l'"htmlizzatore" piuttosto che con chi detiene la paternità intellettuale del suo contenuto;

2) COPYRIGHT – dichiarazione esplicita e sintetica sui diritti relativi al documento, senza link a formulazioni più articolate;

3) GENERATOR – indica il software utilizzato per creare il file HTML. Molti di tali software inseriscono automaticamente il proprio nome nel file stesso senza neppure "avvertire" l'utilizzatore. Evidenti le implicazioni pubblicitarie;

4) ROBOT – indica ai software di ricerca dei motori se e come tenere conto della pagina. Non tutti i robot sono però rispettosi delle indicazioni ricevute;

5) TITLE – il metatag più importante. Non dovrebbe mai mancare in una pagina ben costruita e dovrebbe includere alcune parole particolarmente significative per individuare la risorsa a cui si riferi-

sce. È il metatag più utilizzato e più "pesante" per i motori di ricerca, che fra l'altro lo visualizzano in testa alle "schede" che si ottengono effettuando una interrogazione. Inoltre il TITLE, che potrebbe essere definito come qualcosa a metà strada fra il soggetto e il titolo uniforme, appare anche nei bookmark e in testa alla finestra del browser. Alcuni motori,⁷ infine, permettono di limitare la ricerca alle sole parole contenute in questo metatag;

6) DESCRIPTION – breve descrizione del contenuto della pagina, ovvero un mini-abstract che molti motori visualizzano nel corpo delle loro "schede". In caso di assenza di questo metatag il motore si arrangia da solo, visualizzando in genere le prime parole presenti nella parte visibile al browser della pagina (BODY), con risultati talvolta efficaci ma più spesso inutilizzabili;

7) KEYWORD – ulteriori parole chiave (oltre a quelle contenute in TITLE e DESCRIPTION) ritenute particolarmente significative dall'autore della pagina per caratterizzarne il contenuto.

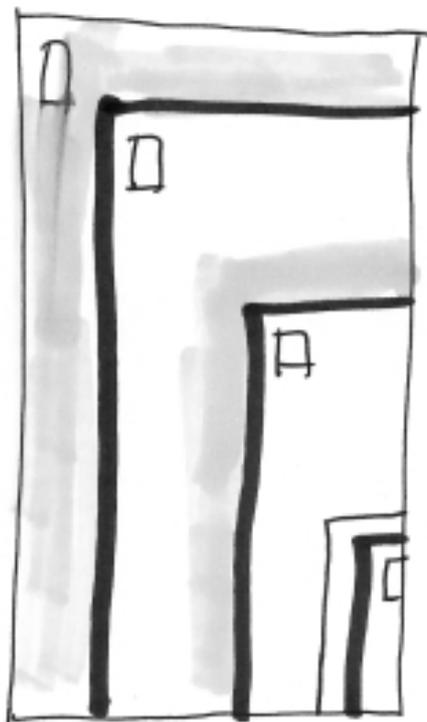
Altre tipologie di strumenti

Indici web per argomento. Detti anche "directory", "subject tree" o "pick list", questi indici non sono costruiti automaticamente come quelli dei search engine, ma vengono costruiti da personale specializzato che seleziona e indicizza le pagine web in base al soggetto, scartando quelle meno interessanti e talvolta aggiungendo un breve riassunto e una valutazione. Il lavoro umano garantisce un filtro qualitativo e diminuisce il "rumore" rispetto alle ricerche effettuate coi motori, ma per problemi di costi gli archivi sono di gran lunga più piccoli di quelli usati dai motori. Pochi di questi indici adottano sistemi formalizzati di classificazione

bibliotecaria (CDD, CDU, LC), mentre la maggior parte utilizza una propria gerarchia di classi annidate piuttosto approssimativa dal punto di vista del rigore concettuale. Alcuni, come Yahoo!, permettono una ricerca per parola all'interno del proprio archivio di "istituzioni", una possibilità che gli utenti meno esperti confondono spesso con la ricerca full-text nell'intero webspace tipica dei search engine. Alcuni indici per argomento, quali ad esempio Argus clearinghouse, New Athenaeum e WWW virtual library, vengono definiti "distribuiti", perché sono costituiti dalla collaborazione tra diversi siti indipendenti, ciascuno dedicato a un determinato argomento, e quindi potrebbero anche essere definiti dei "repertori di virtual reference desk". Fra le principali directory internazionali si possono citare:

- About.com <<http://www.about.com>>;
- AlphaSearch: gateway to the academic web <http://www.calvin.edu/Lib_Resources/as>;
- Argus clearinghouse <<http://www.clearinghouse.net>>;
- Google directory <<http://directory.google.com>>;
- Infomine: scholarly Internet resource collection <<http://lib-www.ucr.edu>>;
- Librarians' index to the Internet <<http://sunsite.Berkeley.EDU/InternetIndex/index.html>>;
- New Athenaeum: the official Internet resource metaguide <<http://members.spree.com/athenaeum/mguide1.htm#000>>;
- WWW virtual library <<http://www.w3.org>>;
- Yahoo! <<http://www.yahoo.com>>.

Indici web per parola limitati a una disciplina. Costituiscono degli ibridi fra le due categorie precedenti (search engine e directory), delle quali cercano di unire i pregi. Applicano la potenza "cieca" dei



motori di ricerca esclusivamente a un gruppo di siti dedicati a una particolare disciplina o argomento e indicizzati da personale specializzato. Tra questi si possono citare Argos <<http://argos.evansville.edu>>, che si definisce un “limited area search engine of the ancient and medieval Internet”, e Hippias <<http://hippias.evansville.edu/>>, per una “limited area search of philosophy on the Internet”. Numerosi di tali strumenti sono raggiungibili attraverso la Search engine alliance <<http://searchenginealliance.com>>, “the confederation of expert search engines & web directories”. Un interessante esperimento di progressivo accorpamento cooperativo di più siti settoriali al fine di permetterne una interrogazione per parola cumulativa è condotto dal Resource discovery network <<http://www.rdn.ac.uk>>.

Indici specializzati in particolari tipologie di documenti o utenti. Sono directory (solo talvolta dotate anche di funzioni di search engine) limitate esclusivamente a documenti sonori (ad esempio Audiofind <<http://www.audiofind.com>> e Listen.com <<http://www.listen.com>>) o grafici (ad esempio Filez <<http://www.filez.com/zhub.shtml>>, Image surfer <<http://isurf.interpix.com>> o WebSEEk <<http://disney.ctr.columbia.edu/WebSEEk/>>) o adatti a particolari tipologie di utilizzatori (come i bambini nel caso di Bess <<http://www.bess.com>>, Sussidiario <<http://www.sussidiario.it>> e Yahoooligans! <<http://www.yahoooligans.com/>>) o comunque con caratteristiche peculiari.

Metaindici. Permettono l'accesso (talvolta anche l'interrogazione, non sempre contemporanea) a un certo numero di indici e repertori primari, come quelli elencati finora. In alcuni casi è possibile scegliere se impostare la ricerca diret-

tamente dalla pagina del metaindice oppure se collegarsi prima a quella dell'indice primario. Vengono chiamati anche “unified search engine”, e possono rivelarsi utili, tra l'altro, per testare le differenze fra i vari indici. Alcuni metaindici permettono di immettere una sola volta i termini di ricerca, lasciando al software il compito di ripetere l'interrogazione su tutti gli indici selezionati e di produrre una risposta cumulativa; il risultato è “sporco”, cieco rispetto alle peculiarità dei vari archivi e delle relative tecniche di interrogazione, ma la rapidità e la potenza di questi strumenti li rende comunque interessanti.

Possono essere distinti in tre sottocategorie:

– *metaindici in senso stretto.* In questo caso un'unica maschera di ricerca permette l'interrogazione cumulativa di vari indici contemporaneamente. I risultati talvolta sono ordinati in base all'indice di provenienza, con eventuali ripetizioni, mentre a volte vengono “schiacciati”, eliminando le ripetizioni, e vengono ordinati in base alla supposta rilevanza rispetto alla richiesta oppure ad altri criteri. Cfr., ad esempio, Inference Find <<http://www.ifind.com>> e MetaCrawler <<http://www.metacrawler.com>>;

– *multiindici.* Si tratta di pagine con numerosi form per la ricerca su vari indici, interrogabili però solo uno alla volta. Cfr., ad esempio, All in one <<http://www.allonesearch.com/all1www.html#WWW>> e WebTaxi <<http://www.webtaxi.com/>>;

– *indici di indici.* Sono semplici liste di link a indici. Più che veri e propri metaindici sono dei repertori di indici, qualche volta ampiamente commentati e arricchiti di tabelle comparative, recensioni, confronti, segnalazioni di novità, suggerimenti per trovare i siti altrui e far trovare il proprio come nel



caso della Guida completa ai motori di ricerca <<http://www.motoridiricerca.it/>>, Search engine watch <<http://www.searchenginewatch.com/>> e Search engine showdown <<http://www.notess.com/search/>>, curato dall'ex bibliotecario Greg Notess (di cui si veda, in particolare, l'utilissima Search engine features chart <<http://www.notess.com/search/features/>>). A livello metarepertoriale va segnalata la pagina Guide per la ricerca in Internet <<http://www.aib.it/aib/lis/motori.htm>> curata da Mariateresa Pesenti, una ricca raccolta commentata di guide, bollettini, gruppi di discussione, bibliografie e raccolte di link sugli strumenti per la ricerca in Internet.

Virtual reference desk (VRD). Fondamentali metarisorse dalle molteplici denominazioni, fra cui “virtual library”, “electronic reference desk”, “gateway”, “trailblazer page”, “metapage”, “homepage” e ➤

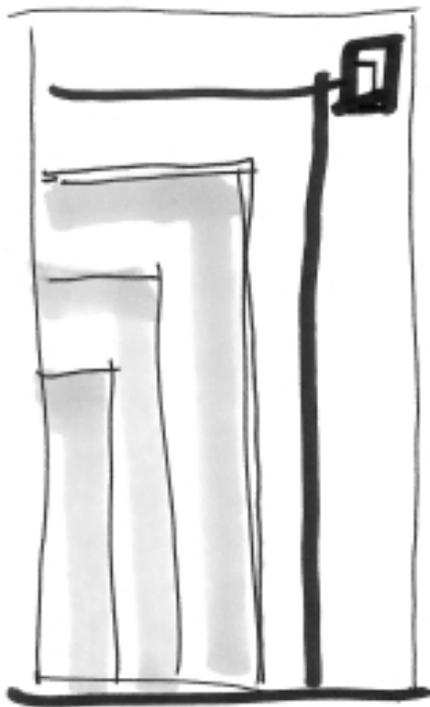
“hub”. Proprio come le sale di consultazione delle biblioteche “reali” questi reference desk virtuali raccolgono, ordinano e talvolta valutano e commentano le principali fonti informative e i più utili strumenti di ricerca disponibili in rete, relativamente a una determinata disciplina o argomento (virtual reference desk specializzati) o a Internet in generale (virtual reference desk generali). Talvolta includono anche (o si presentano come) guide alla ricerca più discorsive, che spiegano metodi e trucchi per la ricerca in rete. In ogni caso si contraddistinguono per la forte selettività delle scelte operate. I VRD costituiscono l’alfa e l’omega di molte delle più efficaci ricerche di informazioni che si possono svolgere in rete, che iniziano orientandosi a partire da un buon VRD generale per poi utilizzare vari strumenti e strategie fino ad individuare un particolare VRD specializzato in grado di soddisfare il bisogno informativo iniziale e, al tempo stesso, eventuali successive necessità di approfondimento o aggiornamento.

Come esempi di virtual reference desk generali, che talvolta (se molto analitici) sconfinano nella categoria degli indici web per argomento, e che spesso ereditano dagli enti che li gestiscono (spesso biblioteche pubbliche o universitarie) una spiccata propensione per il *quick reference* rivolto a una particolare tipologia di utenza (cittadini di una determinata regione, studenti di una determinata facoltà ecc.), si possono ricordare:

- DeskRef <<http://ansernet.rcls.org/deskref>>, a cura del Ramapo Catskill Library System;
- Eureka <<http://biblioteca.goldnet.it/zweb/Eureka/home.htm>>, a cura di Renzo Brun del Servizio bibliotecario provinciale vicentino;
- Guida pratica alle risorse web <[http://www.biblioteca.cologno-](http://www.biblioteca.cologno-monzone.mi.it/servizi_di_informa-)

- [zione/siti.htm](http://www.biblioteca.cologno-monzone.mi.it/servizi_di_informazione/siti.htm)>, a cura della Biblioteca civica di Cologno Monzese;
- Internet per gli umanisti <<http://lettere1.lett.unitn.it/lavori/carl.htm>>, a cura di Carlo Favale dell’Università di Trento;
- Ready reference collection <<http://www.ipl.org/ref/RR>> della Internet Public Library;
- Riccardo Ridi’s bookmarks <<http://www.burioni.it/forum/ridi/home.htm>>, orientati verso il mondo delle biblioteche;
- Virtual reference desk <<http://www.iue.it/LIB/external/external.html>>, dell’Istituto universitario europeo di Fiesole;
- Virtual reference desk per le biblioteche pubbliche <<http://www.cultura.regione.toscana.it/bibl/ref/index.htm>>, curato da Il Palinese per la Regione Toscana.

VRD specializzati. Numerosi VRD italiani e stranieri specializzati nel settore della biblioteconomia sono elencati in AIB-Web il web dell’Associazione italiana biblioteche <<http://www.aib.it>>, all’indirizzo



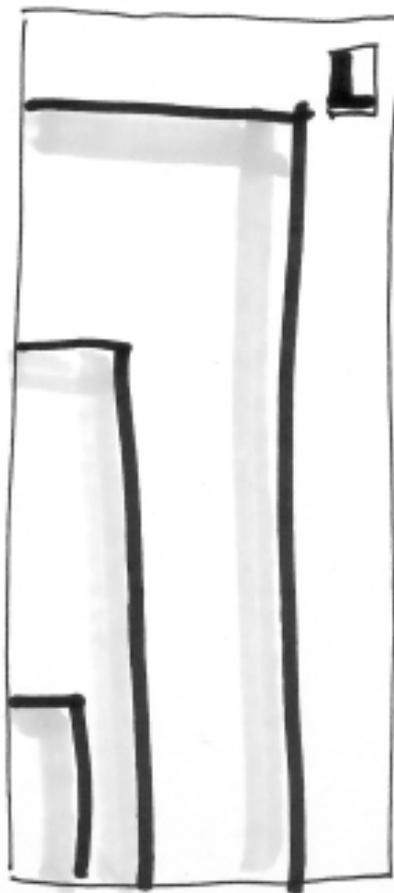
<<http://www.aib.it/aib/lis/vrd.htm>>. Altri due ottimi esempi di VRD specializzati sono la Rassegna degli strumenti informatici per lo studio dell’antichità classica <<http://www.economia.unibo.it/dipartim/stoant/rassegna1/intro.html>> curato da Alessandro Cristofori presso l’Università di Bologna e Documentazione di fonte pubblica in rete (DFP) <<http://www.aib.it/aib/commiss/pubuff/guida.htm>>, coordinata da Fernando Venturini presso AIB-Web. Il repertorio Virtual library italiane <http://www.cilea.it/Virtual_Library/test/bvita.htm> raccoglie numerosi VRD specializzati prodotti in Italia; il curatore Fabio Valenziano spiega che

vi segnaliamo i siti italiani più importanti nei quali esperti di varie discipline raccolgono, selezionano, localizzano e organizzano le risorse informative più interessanti, disponibili non solo localmente ma nell’intera Internet, nei singoli settori. In ciascuna raccolta tematica troverete link a risorse informative sia italiane che straniere. Lo scopo delle segnalazioni è duplice: da un lato indirizzare il ricercatore o lo studioso verso i siti che agevoleranno la ricerca delle risorse informative, dall’altro stimolare gli enti intenzionati a creare nuove homepage tematiche, a non duplicare inutilmente gli sforzi e valutare l’opportunità di attivare collaborazioni con gli enti citati.

Indici di risposte. Si autodefiniscono “Internet search service” o “cybrarian” e pretenderebbero di “comprendere” domande effettuate in linguaggio naturale. I più noti sono Ask Jeeves <<http://www.askjeeves.com>> e Infoplease <<http://www.infoplease.com>>, che in realtà sono solo un magazzino di possibili risposte predisposte in anticipo, scelte fra quelle più probabili (o, per motivazioni pubblicitarie uguali e contrarie a quelle che dettano tale selezione, fra quelle decisamente più improbabili) e dotate di una serie di link a siti pertinenti preselezionati da staff umano. I

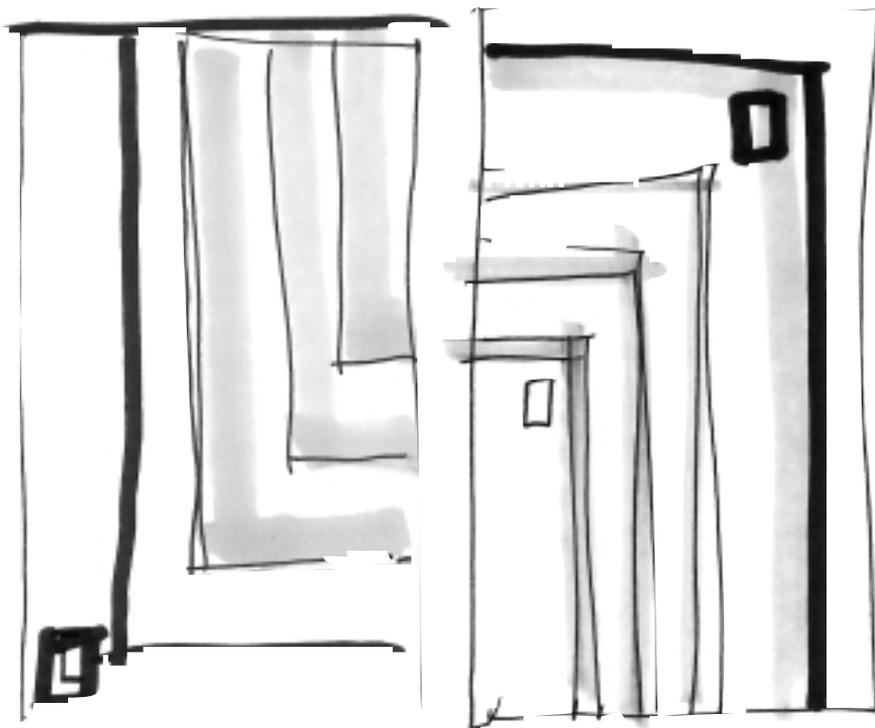
termini di ricerca immessi dall'utente vengono utilizzati da un normale motore di ricerca per parola, con la differenza che stavolta l'archivio da cui si estraggono i risultati non è costituito da una buona fetta delle pagine web mondiali ma solo da alcune migliaia di risposte preconfezionate. Le possibilità di ottenere una risposta pertinente ed esauriente sono piuttosto scarse, e legate più alla capacità di formulare quesiti abbastanza ovvi da parte dell'interrogante che a quelle di acutezza nella comprensione e di perspicacia nella ricerca da parte dell'interrogato. Per non lasciare mai l'utente a mani vuote, se nessuna delle risposte precotte ha proprio niente a che vedere con la domanda, i termini utilizzati vengono "sparati" contro una manciata dei più popolari motori di ricerca per parola, pur di ottenere uno straccio di risultato purchessia. Da non confondere con autentici servizi di reference online "umano" effettuati (con tempi, ovviamente, non istantanei) da *information broker* professionisti oppure da bibliotecari volontari, come nel caso di Ask a Librarian <<http://www.earl.org.uk/ask/>>.

Indici personalizzabili e agenti di ricerca. Ancora in gran parte sperimentali, questi strumenti mirano a rintracciare autonomamente tutte le risorse di interesse per l'utilizzatore, sulla base della preventiva definizione di un accurato "profilo di ricerca" e, nei casi più sofisticati ma anche più imprevedibili, di decisioni "autonome" dei cosiddetti "agenti intelligenti". Talvolta utilizzano tecnologie di tipo push, che "spingono" automaticamente le informazioni sul desktop dell'utente. Per un esempio piuttosto semplice, che si limita a permettere una riorganizzazione personalizzata dell'ordinamento delle informazioni proposte, si veda My Yahoo! <<http://www.my.yahoo.com>>.



Portali. Molto utilizzati dagli utenti meno esperti, e poco adatti invece per professionisti della ricerca in Internet, si candidano a costituire il sito di riferimento dei navigatori non solo per la ricerca di informazioni ma anche per ogni altra attività (comunicazione, giochi, acquisti in linea, prenotazione di servizi ecc.) effettuabile in rete. Quasi sempre presenti una directory molto orientata sulle necessità della vita quotidiana, un motore di ricerca in proprio o mutuato dai maggiori e un coacervo dei più svariati servizi: notiziari, quotazioni di borsa, indirizzi e-mail e spazio web gratuiti, oroscopi, stradari, chat e forum, messaggistica SMS, previsioni del tempo ecc. Si vedano, ad esempio, in Italia, CiaoWeb <<http://www.ciaoweb.it>>, Infinito <<http://www.infinito.it>>, Jumpy <<http://www.jumpy.it>>, KataWeb <[\[www.kataweb.it\]\(http://www.kataweb.it\)>, SuperEva <<http://www.supereva.it>> e Virgilio <<http://www.virgilio.it>>; ma anche giganti di altre categorie come Altavista e Yahoo! si stanno rapidamente "portalizzando" sulla spinta delle pressioni pubblicitarie. C'è chi li distingue in portali "orizzontali" \(cioè generali\) e "verticali" \(cioè specializzati, come ad esempio Alice.it <<http://www.alice.it>>, dedicato al mondo del libro e dell'editoria\).](http://</p>
</div>
<div data-bbox=)

Catalogazioni bibliotecarie. Numerose sono le sperimentazioni effettuate da singole biblioteche (ad esempio, in Italia, la Biblioteca Angelo Monteverdi dell'Università di Roma "La Sapienza" <<http://bam.let.uniroma1.it/inrete/inrete.htm>>), ma i maggiori progetti del genere a livello mondiale sono l'inglese BUBL Link <<http://bubl.ac.uk/link>> e l'americano InterCat, promosso da OCLC, ALA e dalla Library of Congress, che ha esteso il formato USMARC e le regole di catalogazione angloamericane per consentire a un vasto gruppo di bibliotecari di tutto il mondo di creare delle vere e proprie "schede elettroniche" riguardanti oltre 92.000 risorse disponibili in rete, complete di autore, titolo, data, soggetto LC, classificazione DDC e abstract. I risultati del progetto InterCat, ormai concluso, sono interrogabili gratuitamente in rete, in un archivio dedicato <<http://orc.rsch.oclc.org:6990>> o <<http://www.oclc.org:6990/>>, oppure tramite il più vasto archivio di NetFirst <<http://www.oclc.org/oclc/netfirst>>, reso accessibile a pagamento da OCLC nell'ambito del servizio FirstSearch. NetFirst è in continua crescita e oggi comprende più di 120.000 schede; un particolare valore di questo progetto è dato dal fatto che le schede di NetFirst, grazie al loro formato, possono essere prelevate e inserite nel catalogo di un OPAC disponibile via web. Da luglio 2000 sarà disponibile an- ➤



che il nuovo progetto di catalogazione promosso da OCLC CORC (Cooperative Online Resource Catalog) <<http://www.oclc.org/oclc/corc/faq/faq.htm>>.

Strategie

La ricerca di informazioni in Internet sta ormai diventando una vera e propria professione, o almeno una specializzazione consolidata nel mondo bibliotecario e documentario. È impossibile ridurre a poche pillole di saggezza strategie che richiedono un ampio bagaglio professionale tradizionale unito a una vasta esperienza di rete. Qui schematizziamo quindi soltanto alcune indicazioni di massima su come condurre una ricerca e sull'ordine con il quale interrogare le diverse metarisorse esistenti:

- se si desidera solo un primo orientamento, per prepararsi a eseguire in seguito ricerche più specifiche, la cosa migliore per cominciare è la visita a un virtual refe-

rence desk generale, a uno di biblioteconomia e a uno specializzato, scelto in base alla tipologia della propria biblioteca;

- se invece si conosce già quale tipo di informazioni si desidera (ad esempio quando si cerca l'indirizzo di una biblioteca), occorre affidarsi a un repertorio specializzato, o eventualmente a un virtual reference desk generale o specializzato per scoprire se esistono repertori di quell'ambito (ad esempio repertori di biblioteche);

- quando si cercano informazioni di ogni tipo su un certo tema, conviene invece partire da un indice per argomento: con un po' di fortuna si possono trovare così una o più risorse soddisfacenti; con molta fortuna, si potrebbero trovare addirittura uno o più virtual reference desk specializzati da usare come guida per le ricerche successive;

- nel caso in cui le strategie precedenti non abbiano avuto buon esito, si può provare a interrogare un indice per parola, utilizzando prima le tecniche più elementari ed

eventualmente seguendo poi le istruzioni per quelle più complesse e potenti: tra i risultati si dovrebbe riuscire a individuare qualche risorsa specifica o addirittura un virtual reference desk;

- nel caso si cerchi qualcosa di molto specifico, infine, ben individuabile con una o più parole (un autore minore, un particolare software, un ente ecc.) conviene provare subito a usare un indice per parola o addirittura un metaindice;

- bisogna ricordarsi che talvolta può essere utile anche un approccio "archivistico" alla ricerca delle informazioni, mirando alla fonte, ovvero all'ente che emana o comunque è interessato ai documenti cercati: con un po' di fortuna costruire indirizzi del tipo <www.nome-ente-esteso.it> o <www.sigla-ente.com> può portare direttamente alla meta;

- quando un link "va a vuoto", si può provare ad "accorciare l'URL" eliminandone il segmento successivo all'ultimo slash (/), risalendo progressivamente verso l'indirizzo dell'host, finché non si trovano indizi per rintracciare la nuova collocazione della pagina scomparsa;

- utilizzare tutti gli strumenti e le strategie analizzati fin qui, ma senza mai dimenticare che il web è un ipertesto in cui è sempre possibile anche navigare "a vista" seguendo i link che paiono più promettenti. ■

Bibliografia

ACKERMANN, Ernest – ACKERMANN, Karen [1999] *The information specialist's guide to searching and researching on the Internet and the World wide web*, Wilsonville (Or.), ABF content.

BASILI, Carla [1995] *La ricerca "per soggetto" dell'informazione in Internet*, "Biblioteche oggi", XIII, 6, p. 40-47.

BASILI, Carla [1998] *La biblioteca in rete. Strategie e servizi nella società dell'informazione*, Milano, Editrice Bibliografica.

BASH, Reva [1998] *Ricerche online for*

- dummies*, traduzione di Deborah Tagliacozzo, Tullio Policastro, Paolo Morpurgo e Bruno Piatti, Milano, Apogeo (*Researching online for dummies*, Foster city Cal., IDG, 1998).
- BOWEN, James [1999] *Conducting research on the Internet*, in *Encyclopedia of library and information science*, New York, Dekker, 1968- , vol. 64 (supplement 27), p. 62-68.
- BUTLER, John A. [1998] *Cybersearch. Research techniques in the electronic age*, New York, Penguin books.
- CALVO, Marco – CIOTTI, Fabio – RONCAGLIA, Gino – ZELA, Marco A. [1999] *Internet 2000. Manuale per l'uso della rete*, Roma-Bari, Laterza, oppure <<http://www.laterza.it/internet/>>, p. 216-268.
- CAMAGNI, Valeria [1999] *I web portal italiani*, "PC professionale", 104, p. 435-445.
- CASEY, Carol [1998] *Web rings. An alternative to search engines*, "College & research libraries", LIX, 10, p. 761-763.
- CASEY, Carol [1999] *An analytical index to the Internet: dreams of utopia*, "College & research libraries", LX, 6, p. 586-.
- CHOWDHURY, G.G. [1999] *The Internet and information retrieval research: a brief review*, "Journal of documentation", LV, 2, p. 209-225.
- COHEN, Alan [1999] *Web portal e siti di ricerca: parti da qua*, "PC professionale", 104, p. 408-433.
- COHEN, Laura [1999] *Quick reference guide to search engine syntax*, updated 26 October, <<http://www.albany.edu/tree-tops/docs.library/internet/syntax.html>>
- COOKE, Alison [1999] *A guide to finding quality information on the Internet. Selection and evaluation strategies*, London, Library association publishing.
- CRUPI, Gianfranco – PARLANTI, Donatella – SIMONE, Carla [1999] *Un progetto di catalogazione delle risorse elettroniche: Easycat.ER*, Novembre, <<http://www.burioni.it/forum/crupi-er.htm>>.
- ELLIS, David – FORD, Nigel – FURNER Jonathan [1998] *In search of the unknown user: indexing, hypertext and the world wide web*, "Journal of documentation", LIV, 1, p. 28-47.
- GILSTER, Paul [1996] *Finding it on the Internet. The Internet navigator's guide to search tools and techniques*, revised and expanded second edition, New York, Wiley.
- GLOSSBRENNER, Alfred – GLOSSBRENNER, Emily [1999] *Search engines for the world wide web*, 2nd ed., Berkeley (Cal.), Peachpit press.
- GRASSI, Andrea [1999] *10 motori di ricerca*, "PC magazine", XVI, 160, p. 78-104.
- GRIFONI, Giovanna [1997] *Come orientarsi tra i motori di ricerca. Una panoramica sugli strumenti di recupero delle informazioni in Internet*, "Biblioteche oggi", XV, 5, p. 10-16, oppure <<http://www.burioni.it/forum/grif-mot.htm>>.
- HASKIN, David [1997] *Power search. Web search engines are smarter and stronger than ever. We test six of the best*, "Internet world", VIII, 12, p. 78-92.
- HOCK, Randolph E. [1997] *Sizing up Hotbot: evaluating one web search engine's capabilities*, "Online", XXI, 6, p. 24-33.
- HOCK, Randolph E. [1999] *The extreme searcher's guide to web search engines. A handbook for the serious searcher*, foreword by Paula Berin Stein, Medford (NJ), Information today.
- LESNICK, Leslie L. – MOORE, Ralph E. [1997] *Agenti di ricerca*, Milano, McGraw-Hill libri Italia (*Creating cool intelligent agents for the Net*, Foster city Cal., IDG, 1997), con allegato un cdrom di programmi per Windows.
- LUNATI, Gabriele [1996] *Strumenti per la ricerca ed il recupero dell'informazione su Internet*, <<http://www.burioni.it/forum/motori/motori.htm>>, poi, col pretitolo *Bollicine di champagne...*, anche "Biblioteche oggi", XV (1997), 5, p. 18-29.
- MAFFEI, Lucia [1996] *Il filo di Arianna: trovare informazione di qualità in Internet*, "AIDA informazioni", XIV, 2, p. 26-31.
- METITIERI, Fabio – RIDI, Riccardo [1998] *Ricerche bibliografiche in Internet. Strumenti e strategie di ricerca, OPAC e biblioteche virtuali*, Milano, Apogeo.
- MILSTEAD, Jessica – FELDMAN, Susan [1999] *Metadata: cataloging by another name... [include Metadata projects and standards]*, "Online", XXIII, 1, p. 24-41, oppure <<http://www.onlineinc.com/onlinemag/OL1999/milstead1.html>>.
- MORVILLE, Peter S. – ROSENFELD, Louis – JANES, Joseph [1996] *The Internet searcher's handbook. Locating information, people & software*, New York, Neal-Schuman.
- NOTESS, Greg R. [1997] *Internet search techniques and strategies*, "Online", XXI, 4, p. 63-66.
- NOTESS, Greg R. [1999a] *Special search engine section*, edited by Greg R. Notess, "Online", XXIII, 3, p. 20-90.
- NOTESS, Greg R. [1999b] *On-the-fly search engine analysis*, "Online", XXI, 3, p. 63-66.
- ODER, Norman [1998] *Cataloging the Net: can we do it ?*, "Library journal", CXXIII, 16, p. 47-51.
- PANTÒ, Eleonora – PETRUCCO, Corrado [1998] *Internet per la didattica. Dialogare a scuola col mondo*, Milano, Apogeo.
- PASTERIS, Vittorio [1996] *Internet per chi studia. Orientarsi, documentarsi, preparare la tesi*, Milano, Apogeo.
- PC MAGAZINE [1997] *Catching sites. Internet search engines*, "PC magazine", VI, 2, p. 109-153.
- PC PROFESSIONALE [1997] *Chi cerca trova... ma non sempre. Internet: motori di ricerca*, "PC professionale", 69, p. 271-300.
- PC PROFESSIONALE [1998] *Una guida completa alle ricerche in rete. 36 siti che vi aiutano a utilizzare le risorse del Web*, "PC professionale", 83, p. 322-363.
- RIDI, Riccardo [1996] *Internet in biblioteca*, Milano, Editrice Bibliografica.
- RIDI, Riccardo [1998] *Dal canone alla rete: il ruolo del bibliotecario nell'organizzazione del sapere digitale*, "Biblioteche oggi", XVI, 5, p. 12-19, oppure in: *Bibliotecario nel 2000. Come cambia la professione nell'era digitale*, Milano, 12-13 Marzo 1998, a cura di Ornella Foglieni, Milano, Editrice Bibliografica, 1999, p. 62-76, oppure <<http://www.fahrenheit.org/fahrenheit/Calendario/Ridi.htm>>.
- RIDI, Riccardo [1999a] *Vittime del fuoco amico. Mito e realtà delle interfacce amichevoli*, "Biblioteche oggi", XVII (1999), n. 5, p. 12-17.
- RIDI, Riccardo [1999b] *Metadata e meta-tag: l'indicizzatore a metà strada fra l'autore e il lettore*, relazione tenuta al Convegno internazionale "The digital library. Challenges and solutions for the new millennium", Bologna, 17-18 Giugno 1999, testo aggiornato e link controllati il 29 Settembre 1999 per la pubblicazione in AIB WEB, <<http://www.aib.it/aib/commiss/cnur/dltridi.htm>>.
- SALARELLI, Alberto [1996] *Ricerca* ➤

e valutazione delle informazioni in *World wide web*, "Archivi & computer", VI, 5, p. 413-430.

SELTZER, Richard – RAY, Eric J. – RAY, Deborah [1998] *The AltaVista search revolution*, 2nd ed. Berkeley, Osborne/McGraw-Hill.

SHA, Vianne Tang [1995] *Cataloguing Internet resources: the library approach*, "The electronic library", XIII, 5, p. 467-476.

SCHLEIN, Alan M. [1999] *Find it online. The complete guide to online research*, edited by James R. Flowers jr., Shirley Kwan Kisaichi, Peter Weber, Tempe (Az), Facts on demand press.

SONNENREICH, Wes – MACINTA, Tim [1998] *Web developer.com guide to search engines*, New York, Wiley.

WOODWARD, Jeannette [1996] *Cataloguing and classifying information resources on the Internet*, "Annual review of information science and technology", XXXI, p. 189-220.

ZORN, Peggy – EMANOIL, Mary – MARSHALL, Lucy – PANEK, Mary [1999] *Finding needles in the bay-*

stack: mining meets the Web, "Online", XXI, 3, p. 30-51.

Note

Questo articolo è presente in versione elettronica su Forum Burioni <<http://www.burioni.it/forum/ridi-mot.htm>>.

¹ Mutevole sia fra motore e motore, sia nel corso della storia di ogni singolo motore, per venire sempre più incontro alle aspettative degli utenti, che tendono a valutare i motori più dall'ordinamento dei risultati che dalle dimensioni e dall'aggiornamento dei relativi database o dall'effettiva individuazione di pagine realmente rilevanti.

² Resta in genere salva la possibilità per gli utenti di forzare il motore alla ricerca in AND con l'uso di particolari operatori (spesso il "+").

³ Alcuni motori sono dotati di liste (non liberamente consultabili) delle coppie o dei gruppi di termini che vengono più spesso utilizzati dagli

utenti nella medesima interrogazione (talvolta però solo se uniti fra loro con le virgolette in una ricerca della *exact phrase*). Tale tipo di ricerca "esatta" viene in tali casi effettuata automaticamente anche se l'utente, ignaro della "memoria" del motore, sta utilizzando i termini "gemellati" slegati fra loro.

⁴ O addirittura, come fa Google, "fuori" dalla pagina, prendendo in esame anche le parole da cui partono link diretti verso di essa e quelle circostanti.

⁵ Google "premia" anche i link "in uscita".

⁶ I metadata sono "dati sui dati", ovvero informazioni, generalmente strutturate e scandite in campi, relative a documenti primari "a testo pieno" (full-text), che ne permettono una più efficiente organizzazione e recupero. I "metatag" sono quei tag (marcatori) del linguaggio HTML che possono venir inseriti, con funzione di metadata, nella parte nascosta HEAD del file a cui si riferiscono.

⁷ Attualmente almeno Altavista, Hotbot, Infoseek, Northern Light e Lycos Pro Search.