Smarter Searching

The importance of good search tools has been apparent since the early days of the Internet. There are now an estimated 100 million Web pages,

NET TIPS

and with about 200,000 added each day, it is not surprising that the main complaint among users is the difficulty of finding the documents they want. An important class of tools is the search "engine," which takes keywords as input and returns a set of Web documents. Effective use of these tools requires some knowledge of how they work.

Search engines are not directories. The latter are sites where several thousand Web sites have been categorized by human reviewers. The largest is Yahoo! (www.yahoo.com), with links to about 370,000 Web sites. Although directories often point users to quality sites, they cannot possibly catalog all of the Web. For a more thorough search, users must turn to search engines. Because these engines are not limited by the availability of human filters, they can catalog literally millions of Web documents.

Each engine consists of three distinct components: the spider, the index, and the query module. The spider (also referred to as the robot) is a program that "moves" on the Net from page to page in search of new documents. Each robot uses its own specific algorithm for finding and navigating Web pages. The harvested pages are then entered into a database on one of the search engine's computers. The database is organized by an index whose architecture is also specific for each search engine. Users search the index through a predefined query module, an interface specific to each engine.

Two concepts should be borne in mind. First, search engines do not scour the Net in real time, but rather query an index of Web pages compiled on the search engine's site. This explains why some search results point to outdated or nonexistent links. Indeed, it may take days or weeks for any particular search engine to traverse the entire Net. Second, the same query on two different search engines will not yield identical results, because the combination of proprietary robot, index, and user interface will be unique. The best strategy is to try several engines when searching for a specific Web document.

With the proliferation of commercial search engines, a new tool—the metasearch engine—has occupied a slot above them all. Metasearch engines will send a query to several other engines in parallel and return a composite report (with duplicate entries re-

moved). This can save the effort of performing several searches in series. But metaengines tend to be slower and tend to return a maximum of about only 100 results.

Because each engine has its own quirks, it is worth reading the search tips usually listed on the home page. Armed with a few search engines and perseverance, users should be able to find relevant documents in what is fast becoming the world's largest library. Some top search engines are listed in the accompanying Site Finder section, and links to additional search sites can be found at www.MedsiteNavigator.com/techsight/nettips_3.html

–Richard Peters and Robert Sikorski

Digital Mailbox:
www.sciencemag.org/dmail.cgi?53284

Search Engines

HotBot
www.hotbot.com
From the creators of *Wired*, this is a very fast search engine with an index of 53 million Web documents. *PC Magazine* selected it as their number one choice for its advanced search features. Both the Web and Usenet groups can be searched, and the query syntax allows exact matches, near matches, or boolean expressions. The search can be narrowed to specific geographic locations, calendar dates, or even media types (image, audio, java, and so on). Users can customize the interface and save the settings for future use.

SITE FINDER

Lycos
www.lycos.com
Lycos (Greek for spider) is a hybrid between a search engine (100 million pages) and a directory. LycosPro is now available for advanced searching with Java-based Power Panel relevancy controls. This engine, though, does not offer the advanced search options of something like HotBot. On the other hand, Lycos offers a human-edited directory that is divided into 18 categories, including science, education, and health. It features the top 5% Web sites reviewed by a team working for Lycos, with 100 new reviews added per week.

AltaVista
altavista.digital.com
Designed as a research project at Digital Equipment to showcase the power of their Alpha Web server, it is one of the fastest engines. It houses an index of more than 30 million pages, and lets you search the Web and Usenet groups in several languages, from Chinese to Swedish. The advanced search allows boolean operators such as "and," "or," "not," and "near" (which en-

sures that two words are within 10 words of each other).

InfoSeek
www.infoseek.com
This search engine (with 50 million documents) now has four variations: Ultrasmart offers bare bones functionality for searching the Web; Ultraseek offers advanced features such as whole-sentence searching, and word variant ("mouse" will also return documents containing "mice"), case-sensitive, and field-based searching (searches a specific field such as Title or URL on Web pages); Newscenter features the latest headlines organized by category; and Smartinfo is a collection of tools that let you search for e-mail addresses, stock quotes, phone numbers, and company profiles.

Excite
www.excite.com
Another hybrid search engine, Excite features an index of 50 million Web pages and a directory of 150,000 sites organized into 14 channels including careers and education, and health. You can also look up e-mail addresses, newsgroups, stocks, and the weather. The Powersearch option of Excite lets you search for words or phrases, but pales in comparison to those offered by the other engines. As a bonus, however, the service has just introduced free Web-based e-mail to all its users.

Metacrawler
www.metacrawler.com
This metasearch engine does not maintain a local database, but, rather, it queries in parallel those of Lycos, Infoseek, WebCrawler, Excite, AltaVista, and Yahoo! with a customized search window. Powersearch allows searches limited to a geographic location and to a word or a phrase. Although such a metaengine will ensure that you have pulled out the most popular documents from the engines and directories queried, it does not offer advanced search features.
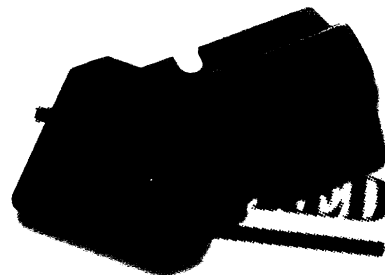
–Richard Peters and Robert Sikorski

Digital Mailbox:
www.sciencemag.org/dmail.cgi?53283

Tech.Sight is published in the third issue of each month, and appears in Science Online at www.sciencemag.org. Contributing editors: Robert Sikorski, National Cancer Institute, Bethesda, MD; Richard Peters, Harvard Medical School, Boston, MA. The editors welcome your comments by e-mail to techsight@aaas.org. Specific comments and feedback should be routed via the Web with the Digital Mailbox URLs at the end of each item.