W H E R E

Unless you use advanced search techniques, you will continue to wade through reams of digital garbage rather than zoom in on a dozen or so relevant sites

There is a nifty site on the Web (www.webcrawler.com/Search Ticker.html), which lets you peek at all the searches happening at Webcrawler. We spent some time there, and quickly realised how poorly people tend to frame their search queries. Search strings ranged from the monosyllabic `love` to the verbose `what will I get my husband for our anniversary.`

Curious to know what results the hopeful searchers got? `Love` was worth 98,018 hits; and among the top five was a link for the fans of Jennifer Love Hewitt. If getting 20,221 hits was any consolation to the anxious wife, it was short-lived. Out of the top ten gift ideas she received, one was a suggestion to attend the 25th anniversary of UNIX, and another offered the company of Laura Grafe, a certain lady residing in upstate New York.

www.chip-india.com

Something is definitely amiss here! Clearly, quite a few of us are not searching intelligently.

Seeking information

`Searching the Net` is not a particularly new phenomenon. Tools for searching were around even before the World Wide Web became popular. One of the earliest of these was Archie (a variation of `archive`). Archie was a system that indexed FTP (File Transfer Protocol) sites. FTP, which allows files to be transferred across the Internet, is still widely used.

Another early bloodhound was gopher, a menu-based system that enabled access to files in a way not drastically different from the way we browse the Web today. Gopherspace has two search engines: Veronica (Very Easy Rodent-Oriented Netwide Index to Computerised Archives) and Jughead (Jonzy`s Universal Gopher Hierarchy Excavation And Display). These acronyms may seem contrived, but just try constructing acronyms to go with Archie!

Today, searching is easier and more intuitive than ever before. Search engines help us quickly find what we are looking for. Some search sites come with their own directories, popular among them being Yahoo and Excite. In this case, a directory is a carefully classified collection of links. Their power lies in the fact that selection and classification of links is done by humans thereby ensuring a high quality of search results. Their weakness is obvious: the Web is too vast for a roomful of people to handle.

How do search engines work? They contain three main elements: a spider, an index and sifting software. The spider scours the Web by visiting pages, usually once or twice a month, to monitor changes. The Web pages are then added to an index that as the name implies, attempts to catalogue the pages the spider visited. The last step is the sifting (sorting by relevance) required to present the user with the final results. There is no unique and optimal design of these three components, and search engines differ in their implementation. For most ordinary searches, a little knowledge of some basic search engine math will be sufficient.

↓Be specific: The first rule of any intelligent searching is to be as specific as you can. Search for `sonia gandhi`, not just `sonia`. Use lower-case searches `they will search in upper case`, but the converse is not true.

↓Increase the scope: To make sure that you find pages that have all the words you enter, use the `+` symbol. For example, if you want to find sites relating to the movie Star Wars Phantom Menace, `+star +wars` would get you more hits than you`d care for. A more refined search would be: `+star +wars +phantom +menace`

↓Narrow the search: Use the `-` symbol to find pages that have one word on them but not another word. For example, if you want to find information about Clinton but don`t wish to read about the `Lewinsky scandal`, you could try: `clinton -lewinsky`

POWER UP YOUR SEARCHES

	HOW TO	ENGINES THAT SUPPORT
To Match Any Term	Auto	AltaVista, Excite, GoTo, Infoseek, Netscape, Yahoo
	By Menu	HotBot, Lycos, MSN Search
	Special cases	Northern Light (use OR); Google (impossible)
To Match All Terms	Auto	Google, Hotbot, Lycos, MSN Search, Northern Light
	`+` option	Rest all of the sites
For Title Search	Use title:	AltaVista, HotBot, Infoseek, MSN Search, Northern Light
	Special cases	Lycos (via menu), Yahoo (t:)
	Not possible	Excite, Google, Netscape, WebCrawler
For Site Search	Domain:	HotBot, MSN Search
	Other	AltaVista (host:), Infoseek (site:), Lycos (via menu)
	Not possible	Excite, Google, Looksmart, Northern Light,

.Enclose in quotes: Enclosing your query within double quotation marks ensures that the search engine looks for the entire phrase. When we searched for Star Wars, we would have been better off entering star wars rather than +star +wars.

Many search engines allow you to search within the HTML title of the Web page. This is the text that appears within the title tag of a document and usually reflects the content of the document.

Another powerful feature is the ability to control what sites are included in, or excluded from, a search. Too often, you will find unwanted results. Imagine you are looking for Windows 98 bugs but do not want to refer to the Microsoft site, then try a search string (refer to the table for the exact syntax):
  windows   98   +bugs
  host:Microsoft.com

Boolean searching

Boolean methods, named after math-ematician George Boole, have been traditionally used to search databas-es.. With the advent of friendlier searching techniques like natural lan-guage searching, their utility might be a bit diluted. Nevertheless, for power and exactness, nothing beats a Boolean search. It might take you a lit-tle time to get used to all the ANDs and ORs but you will soon realise that it is a wise investment of your time.

.The Boolean AND command is used to specify that all search terms be present on the pages that are list-ed in response to a query. For exam-ple, if you are looking for references to Vajpayee in the Kargil context, use: vajpayee AND kargil

.OR allows any of the specified search terms to be present on the Web pages listed in the results. Let s suppose you are searching for Gandhi, and want to accommodate the peculiar spelling habits of some Americans who spell it as Ghandi , then the command would be: gandhi OR ghandi

.By now, you can probably guess what the NOT command is used for. Searching for vajpayee outside the context of Indian elections might look like this: vajpayee NOT election

.The NEAR command is helpful to specify how close the terms should be to each other. You use a com-mand like this: vol-cano NEAR lava.

The real power of Boolean searching lies in the ability to combine Boolean terms to form complex queries. Let s say you use Photoshop and Paintshop Pro, two popular image soft-ware packages, and would like to find out about image filters pertaining to any one of them. Here is what a preliminary query would look like in Boolean terms:
  filter AND (photo-shop OR paintshop)

Metasearching

Research has revealed that search engines usually have a low overlap with one another. Metasearch engines significantly increase your chances of finding the page you want. These search engines allow for a single search query to be applied to multiple search engines, in the hope that what a single search engine can do, a dozen can do better.

The power of a metasearch lies in the multitude of search engines avail-able at its disposal, but this is also its main weakness. Every search engine has its own way of sorting through results, and there is no easy way of compiling such a disparate array of search results. Still, if your search field is returning fewer results than you want, you should definitely try metasearching.

Among metasearch engines, Metacrawler (www.metacrawler.

com) gets top honours. It collates pages from a slew of sites, besides offering the effective MiniCrawler, which performs searches in a dis-creet window on your desktop.

For the rookie, there is Ask Jeeves (www.askjeeves.com), a natural lan-guage search engine. If you are look-ing for power and don t mind investing some time in customisa-tion, go for SavvySearch (www.savvy search.com) and watch its masterful integration of results.

ROMIT MITRA

SOME USEFUL SEARCH SITES	
All the Web www.alltheweb.com	Blazingly fast and very comprehen-sive. You have to try it to believe it!
AltaVista www.altavista.com	Reliable and still one of the best. Has a very good image search
BigFoot www.bigfoot.com	Looking for a friend, or just about anyone. BigFoot delivers
Deja.com www.deja.com	Helps in Internet purchases. Its motto is discuss, decide and pur-chase
Excite www.excite.com	A medium-sized index, Excite throws in non-Web material like company information and sports scores
FindLaw www.findlaw.com	As the name suggests, use FindLaw to look for legal information
Google still www.google.com	Fast and effective even though it s a beta. Makes use of link popularity to rank its results
HotBot www.hotbot.com	Its Inktomi search engine has been ported to many search sites
Go / Infoseek www.go.com	Basically a portal site, but well known for its quality searches
Khoj www.khoj.com	The Great Indian Search Engine , does a fair job
Lycos Lycos www.lycos.com	One of the oldest search engines, shifted to the Yahoo directory model. Not very effective.
MetaCrawler www.metacrawler.com	The most popular metasearch engine, and rightfully so
MP3 Lycos mp3.lycos.com	Your one-stop to free mp3 music. Over 500,000 searchable files
Microsoft Network www.msn.com	Powered by HotBot s Inktomi, MSN politely offers you access to other search engines
Netcenter engine. www.netcenter.com Altavista	Netscape s proprietary search Secondary results come from
GoogleNorthern Light www.northernlight.com	Has the largest index of any search engine around.
OneLook www.onelook.com	The most comprehensive dictionary on the Web