

Shhhh! . . . do you want to know a secret?

By Mary Carroll and Graham Rowe

Information and information retrieval have become much more complex over the last ten years. One of the difficulties facing those who work in libraries is ensuring that users get the best information possible. In academic environments such as school libraries, this involves not only helping find the information, but also equipping users with the right attitudes and skills to maximize success and accuracy. The advent of the World Wide Web has expanded our range of resources, but our obligation to strive for quality, authority and accuracy in the information we retrieve, and to encourage this attitude in our users, remains the same. To equip ourselves and our users with the skills needed to retrieve the best information possible; we need to delve into the secrets of The Web to uncover hidden places.

Shh! . . . Secret Number One: What is a search engine?

Firstly remember a search engine is not a web browser such as Internet Explorer or Netscape. They are tools such as Google, Alta Vista and Yahoo that are used to retrieve information from the World Wide Web. Search tools include:

SUBJECT DIRECTORIES i.e. LII (Librarians' Index to Internet);

SEARCH ENGINES i.e. Google;

MULTIPLE ACCESS or META SERVICES i.e. Dogpile;

META INDEXES. i.e. metacrawler.

Search engines are the most commonly used tool to retrieve information. Search engines are

powered by 'spiders' or 'robots', which go off and find pages to index in their enormous databases for later retrieval. But how do they decide the order in which to display the pages? Each search engine does this a little differently and they keep much of what they do and how they do it secret, but there is some general information we can study to develop an understanding of this 'secret' of the web.

Shh! . . . Secret Number two: Retrieving a web page

When you type a search into a search engine you are not really searching every page on the web; what you are doing is searching an enormous index of keywords which has been created by the search engine's 'spider' (a computer program) which trawls or crawls through the web to collect these words from websites and then indexes them. When you type in a search what you are doing is looking for that keyword in the index of a particular search engine. Each search engine's software has a different way of compiling the index words, which is why you can have different results from different search engines. Not all words in a web page are indexed but the body of the text is indexed, usually excluding 'stop' words. Some search engines also search meta tags including keyword meta tags and text descriptions of an image (ALT Tags). Some web pages are indexed because the webmaster has submitted the page to be indexed while others are indexed because the 'spider' follows hypertext links within other pages and

indexes the pages it finds following these links. Some such as Google also do what is called a 'deep crawl'; they index many of the pages on a website rather than just those submitted.

The Search

Once you have you have typed in your search, the computer software then needs to perform its function and retrieve results for you. It does this in a number of ways:

- By matching your search terms to these index terms;
- Checking the structure of the search for things such as phrases (syntax);
- Compiling a list of possible results;
- Giving a weighting in terms of relevance to each page and then ranking the pages accordingly;
- Displaying this ranked result.

Shh! . . . Secret Number three: Ranking and Relevance

How do search engines rank their finds?

How then does the search engine decide on the ranking and relevance of each page to your search? Many organizations, particularly commercial ones, attempt to maximize their ranking on a search engine as this can translate into profit for them. If they can ensure that their page contains many instances of key terms then the user of the browser is more likely to find it highly ranked in their search.

In the past some organisations would include a single word thousands of times within their web page but it would be invisible as it was written white on white. The occurrence of the term so frequently meant that the relevance ranking became very high and the page was displayed high in the results list. Some search engines allow 'sponsored' sites to be listed first. These organizations pay the search engine to have their pages listed before the main search results. Not all search engines notify you that the results you are viewing are sponsored links so

be careful with your results list. Other organizations attempt to manipulate their ranking so they appear early in a search engines results. Two companies were recently given the 'Google death penalty' over such attempts at manipulation. Google gave them a relevance ranking of '0' for three days as a result. If you want to know more about this try <<http://news.bbc.co.uk/1/hi/technology/4685750.stm>>.

Early versions of search engines used page content, meta tags, bolding and keyword summaries to rank pages, however this was easily open to manipulation in the ways described previously.

The relevance, and therefore the ranking can be calculated in many different ways, including:

- If the search terms occurs in the title;
- The number of times a term is used on a whole page or at the beginning of a page (frequency);
- Proximity of the key words in relation to each other;
- Use of the term in meta tags;
- The presence of an *unusual* word in the document;
- The word(s) or term(s) are present in the title meta tags;
- The presence of word(s) in a short document may have more weight than the same word(s) in a long document.

How the major search engines rank

- Google ranks by the number of times a web page is linked to by other websites, allowing for a more accurate reflection of popularity. This is called PageRank and is a numeric value representing how important a page is on the web. Google does try and filter out some links which are attempts to manipulate the system. It still determines relevancy using the criteria above. If you want to learn more and have a technical bent try *The Anatomy of A Large-Scale Hypertextual Web Search engine* by Sergey Brin and Lawrence Page. (<<http://infolab.stanford.edu/~backrub/google.html>>).

- **ALTA VISTA** ranks sites higher that have quality content and it gives particular importance to the title, meta tags and links within a site. It values pages that contain meaningful text, and pages with a variety of links to sites with similar content. It also considers how many links there are to the page, including where the links come from, and the 'quality' ranking of those particular sites.
- **Yahoo! Search** ranks results according to their relevance to a particular query by analysing the web page text, title and description accuracy as well as its source, associated links, and other unique document characteristics. Yahoo is in fact a directory but does have a search interface.

Shh! . . . Secret Number four: Making the most out of your search engine

Google

Many of us don't use the full functionality of the search engines we use. This section is going to describe some of the lesser known functionalities of Google and Yahoo! More information can usually be found on the home pages of these services but this should be enough to get you started. Two really useful links to Google are:

<<http://www.google.com/help/basics.html>>
<<http://www.google.com/help/features.html#pdf>>.

(See table 1 (below.)

Yahoo

There are two ways to search Yahoo! You can use the basic search box on the main

Table 1: Some useful Google searches

Book Search Books.google.com	Use Google to search the full text of books. Whenever books in the Google Book Search index contain content that matches your search terms, you'll see links to those books. Click on the title and you'll see the page in that book which contains your search terms, as well as other information about the title. Many titles have only a limited preview, not the complete work.
Calculator	To use Google's built-in calculator function, simply enter the calculation you'd like done into the search box and hit the Enter key or click on the Google Search button.
Currency Conversion	To perform a currency conversion, simply enter the conversion you'd like done into the Google search box and hit "Enter" or click the Google Search button. Try: 3.50 USD in AUD
Definitions	To see a definition for a word or phrase, simply type the word "define," then a colon, and then the word(s) you want defined. Example: define:computer
I'm Feeling Lucky	The "I'm Feeling Lucky"™ button takes you directly to the first web page Google returned for your query. You will not see the other search results at all. It is not a random result but the result felt to be the most accurate
Google Scholar http://scholar.google.com	Google Scholar provides a simple way to broadly search for scholarly literature, both full text and citations. You can search many disciplines and sources: peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations.
Accessible Web Search for the Visually Impaired < http://labs.google.com/accessible >	Accessible Search is an early Google Labs product designed to identify and prioritize search results that are more easily usable by blind and vision impaired users. Accessible Search goes one step further than a normal Google search by helping you find the most accessible pages in that result set.

(Adapted from: <<http://www.google.com/help/pdf>> Look here for further details)

Yahoo! page at <http://www.yahoo.com> or the ad-free, dedicated Yahoo! Search start page at search.yahoo.com. On the Yahoo! Search results page, the links above the search box allow you to change or refine your search. Click on one of the search tab links to focus your results on images, breaking news, videos, or Yahoo! Directory listings.

Yahoo! Shortcuts

Yahoo! Shortcuts help you find what you're looking for in a flash. Shortcuts include a set of five keywords and one special character!. When you use Shortcut keywords as part of your Yahoo! Search, the information appears right on your search results page:

- For maps, the keyword is **map**. (Use with a location);
- For weather, the keyword is **weather**. (Use with a location);
- For dictionary definitions, the keyword is **define**. (Use with any English word);
- For news, the keyword is **news**. (Use with any topic).

Search Meta Words

There are special keywords you can use in Yahoo! that give you results unique to that special keyword instruction. For example if you use **site**: you find all documents within a particular domain and all its sub domains. Example: site:yahoo.com or site:google.com. You can find other meta words to use in this search engine at <http://tools.search.yahoo.com/shortcuts>.

Shh . . . Secret Number five: The Invisible or deep web

The 'visible web' is what you see in the results pages from general web search using tools such as Yahoo! and Google. The 'invisible web' is what you cannot retrieve or see in the search results and other links contained in these types of tools. According to Maureen Henninger (2003 p.147) the term 'invisible web' was first used in 1994 to describe information not indexed by general search engines. The term

'invisible web' mainly refers to a vast repository of information that search engines and directories can't retrieve directly. It is said to be 500 times bigger than the visible web. Some information not 'seen' by search engines includes:

- database-driven html pages such as web forms and web searches;
- database content (Access, Oracle, SQL Server and DB2);
- some file formats such as .zip, .exe,;
- some multimedia files, graphics, software and Portable Document Format (PDF) documents;
- ephemeral information such as newspaper articles which are not current.

Information is not found on the web because when the search engine 'spiders' meander throughout the Web, they are indexing the addresses of pages they discover, but when these software programs run into a page from the Invisible Web, they don't know quite what to do with it. These spiders can record the address, but can't tell you anything about the information the page contains, because of technical barriers and/or deliberate decisions on the part of the site owner(s), such as a university which requires a password to exclude their pages from search engine spiders.

Some useful tools to search the Invisible Web are:

- Direct Search
<http://www.freepint.com/gary/direct.htm>
- Information age consultants list of Invisible web resources at <http://www.iage.com/invisible.html>
- Resource Discovery Network
<http://www.rdn.ac.uk/>
- InfoMine <http://infomine.ucr.edu/>
- Librarians' Index <http://lii.org/>
- CompletePlanet www.completeplanet.com.

You can also use Google and other search engines to locate searchable databases by searching a subject term and the word 'database'. If the database uses the word database in its own pages, you are likely to find it in Google.

Examples for Google & Yahoo:plane crash
database:languages database:toxic chemicals
database.

Shh . . . Secret Number six: Some useful newsgroups and how to subscribe

On-line discussion lists are 'subscription' services whereby the user subscribes to a list and then will receive a record of the discussion via email that occurs on that list. As a subscriber you may also ask questions and receive answers to queries.

There are discussion lists in all areas both professional and non-professional (see the table below).

Shh . . . Secret Number seven: Some interesting and surprising web sites

Paper Craft

Templates of lots of different model animals, cars motorbikes etc <<http://www.yamaha-motor.co.jp/global/entertainment/papercraft/index.html>>.

eHOW

<www.ehow.com>. Clear instructions on how to do (just about) anything. Either select a category or search in the Google search box.

Mensa workout

<<http://mensa.org/index0.php?page=12>>. A Mensa Quiz for entertainment not to test your IQ.

Take It Apart

<<http://www.takeitapart.net>>. How to take apart many electronic devices.

Google Librarian Center

To subscribe go to the following address
<http://www.google.com/librariancenter/>

ALIA Library Technician Newsgroup

Using your normal email account send the following message with **NO SUBJECT & NO SIGNATURE**

subscribe alialibtec Firstname Lastname

to the following address

listproc@alia.org.au

You can also go to the following webpage and use the online subscription form:

<http://lists.alia.org.au/mailman/listinfo/aliaLIBTEC/>

More information about all e-lists can be found at the ALIA (Australian Library and Information Association) homepage at
<<http://alia.org.au/alianet/e-lists>>

OzTL Newsgroup (school Library newsgroup)

To receive individual posts:

1. Send an email to address OZTL_NET-request@listserv.csu.edu.au
NB: the underscore and dash in the listname.
2. In the Subject line of your message type the command 'subscribe' followed by a word that you wish to use as your OZTL_NET password, eg.
subscribe treetop3
3. Do not include anything in the body of the message.
4. You will receive a 'confirmation of subscription' message asking you to reply to the OZTL_NET-request@listserv.csu.edu.au address to confirm that you wish this subscription to proceed. Please send this reply, upon which you will receive a Welcome message.
5. Please keep the Welcome message, as it contains personal subscription data, which you may need to refer to in the future.
6. You will begin receiving individual OZTL_NET posts within 24 hours of sending your subscription request.

For further information go to <http://www.csu.edu.au/faculty/sciagr/sis/cstl/oztl_net>.

Warning Label Generator

<<http://www.warninglabelgenerator.com>>.

Create your own warning labels

LibraryThing

<<http://www.librarything.com/about.php>>.

An online community where you can catalogue your own library and check out the collections of others

Planetquest

<<http://planetquest.jpl.nasa.gov/index.cfm>>.

Images of deep space and space missions.

Aboriginal Art

<<http://www.aboriginalartonline.com/index.php>>

BookMooch

<<http://www.bookmooch.com>>. A free on-line community for exchanging used books.

Shh . . . Secret Number eight: Some odd or not so very useful secrets

The Web is full of sites which do little to enhance our useful knowledge; however, some of them can be very intriguing or entertaining. The following are websites such as this, and ease us gently into some of the more difficult secrets of the Web, but, which will entertain rather than really tell us something useful

Try these if you have a spare minute:

- <<http://typorganism.com>>. Select 'weight' (the weight of words);
- <<http://faculty.quinnipiac.edu/libraries/tballard/typoscomplete.html>>. (list of typographical errors on library catalogues). Perhaps a useful tool for teaching the problems inherent in searching electronic tools;
- <<http://www.thisdayinmusic.com/inner.php>>. Select 'Birthday number one' (find out what was number one on the charts when you were born);
- <www.wordcount.org> (word popularity and usage);
- <<http://www.shibumi.org/eoti.htm>> (end of the web).

References

AltaVista 2006 <<http://www.altavista.com>>.

BBC 2006 *Click* accessed 30/11/06 <http://news.bbc.co.uk/1/hi/programmes/click_online/default.stm>.

Brin, Sergey and Page, Lawrence *The Anatomy of a Large-Scale Hypertextual Web Search Engine* n.d. accessed 26/11/06 <<http://infolab.stanford.edu/~backrub/google.html>>.

Google 2006 <<http://www.google.com.au>>.

Henninger, M 2003 *The Hidden Web: Finding Quality Information On The Net*, UNSW Press, Sydney, NSW.

Yahoo! 2006 <<http://www.yahoo.com>>.

Mary Carroll and Graham Rowe are teachers at the Library and Cultural Unit, Victoria University.

Victorian Premier's Reading Challenge Reference Group**Thank you**

Expressions of interest in the Victorian Premier's Reading Challenge Reference Group for 2007 were called at the end of last year.

The response to this call was very positive; a large number of expressions were received from well-qualified and experienced teachers and teacher-librarians.

Unfortunately not all could be included in the final group, which was selected to ensure balance and diversity; to represent various school sectors, sizes and types across the state, and provide expertise in children's literature, literacy and books from Prep to Year 9.

Thank you to all those who took the time to express interest: your ongoing support is vital to the success of the Challenge.

Chris Finch

Project Officer

Research and Innovation Division

Department of Education and Training

www.slav.schools.net.au