Digital Trekking

Volume 1 No 1— ISSN 1444-6391

Googling down the river, on a Sunday afternoon

'Google is to its competitors as a laser is to a blunt stick.'
—**Time Digital**, May 2000

I want to introduce you to a search engine that is highly applicable to research, yet seems to be relatively little known by university faculty and students. I refer to *Google*, or to give it its URL www.Google.com.

What is so fantastic about Google? In a nutshell, it goes beyond traditional library catalog indexing, and gives you more relevant pages faster than any other search engine on the Web. Even more significantly, it adds citation searching to your Internet toolkit.

Traditional search strategy

The traditional search engines (such as *Altavista*, *Lycos*, *Infoseek*, *HotBot*, *Yahoo*, etc) have web-crawlers that regularly traverse the web and automatically index the pages they find depending on their content, storing the results in a catalog on their server.

When you submit a query to one of these search engines, it finds pages that contain the terms you specified. If you find between 1 and 10 pages you heave a sigh of relief and hope you haven't missed something critical by badly phrasing your query. If the query returns hundreds, thousands or millions of pages, you are basically lost and must rephrase the query. What you really want is probably lost amongst a sea of irrelevancies.

Each search engine indexes pages slightly differently. Some assume that repeated use of a word is important, so some commercial web sites include invisible text the same colour as the background to try to convince the web-crawler the words are really important while not annoying you, saying something like 'book book book…' or 'sex sexual sexy…'. As a consequence, since this issue is on the Web, it will be indexed as being relevant to 'book' and 'sex'.

Some search engines assume titles are more important than the text, or that words that occur early in the page are more important than later ones.

Other search engines rate keywords in <meta> tags highly. Any <meta> data is invisible to you, but will be familiar since it is similar to subject indexing in the library catalog or in the front matter of a book. There are standard sets of <meta> tags such as the *Dublin Core* (see for example www.tased.edu.au/tasonline/metadata/) which writers of pages can use to include other library-like cataloguing data, but usually don't. You will not be surprised to learn that some web authors abuse this facility too in an effort to get to the top of the search lists.

The problem is that this is a traditional paper-based library approach simply translated to the Internet. Just like searching the library catalog, there is a premium on phrasing your query the 'right way' and if you don't you may miss something significant. Alternatively and even more likely you may be overwhelmed by a heap of rubbish as web pages jostle for your attention. There is also a requirement on page authors to do proper cataloguing.

In reality, the typical Internet search problem is not that of missing an important page, but of stopping yourself from being overwhelmed by junk.

Google's search strategy

Google, initially called *BackRub*, was developed by Larry Page and Sergey Brin, a pair of Stanford University graduate students, who then went out into the world to make their millions before completing their PhDs. Although it uses similar indexing

methods to select pages, it uses a unique method to rank the order in which it presents them to you. In brief, it puts the most popular pages (popular in the sense that other people rate them highly and therefore are prepared to give them a link on their page) at the top. In library-talk, it gives you the pages most often cited by someone else. It does this by analysing the pages that its webcrawler finds for links to other pages, and storing this information in its database.

The chances are high that the pages you are given are really important, or are especially well-written, or are those Internet hub sites you are really looking for. They are so confident that one of the search options they give you 'I'm feeling lucky' is a gamble that gives you only one, highest ranked, site!

So, if the topic you are looking for has a moderate number of entries on the Web, Google is likely to get you to the important places faster than anything else.

If you've read so far you will want to try it out. Stop. Fire up your browser, type in www.Google.com and go play. Here are some suggested sample queries drawn from my recent surfing and teaching experience, but by all means try some of your own.

Alexander the Great
echidna REM sleep
hypothermia in water

If you don't want to take my advice at this point, here is a sample search. Let's try to find out what a *coolibah tree* is (as mentioned in the Australian song *Waltzing Matilda*). I searched Google for *Coolibah tree* and here are the first few results.

Google results **1-10** of about **337** for **coolibah tree**. Search took **0.05** seconds.
Category :
Reference > Dictionaries > Language > W > Warlpiri

#64- **Coolibah Tree**
... **Coolibah Tree** Eucalyptus microtheca Evergreen, height to 60',...
www.unlv.edu/Landscape/plant64.htm - Show matches (Cache) - 2k - Similar pages

Talk: The **Coolibah Tree** - N M A
...The **Coolibah Tree** Preface. This is an account of what it's...
www.newmodelarmy.org/talk_4.htm - Show matches (Cache) - 73k - Similar pages

CSIRO Australia - Media Release 98/181, **coolibah tree**
...Waltzing Matilda, is at risk. The **coolibah**, like many trees that adorn...
...image is of a swaggie under a gum **tree** beside a billabong. Those gum...
www.csiro.au/news/mediarel/mr1998/mr98181.html - Show matches (Cache) - 10k - Similar pages

New Age, Meditation, Spirituality, Cosmology, Soul, Spirit, Jollyswagman 2
...NATIONAL ANTHEM! Glossary **Coolibah Tree** ONCE A JOLLY SWAGMAN...
...BILLABONG UNDER THE SHADE OF A **COOLIBAH TREE**, AND HE SANG AS HE...
www.jollyswagman.com/The%20Quest/opening.htm - Show matches (Cache) - 16k - Similar pages

Waltzing Matilda
...billabongs Under the shade of a **Coolibah tree** And he sang as he...
...billabong Under the shade of a **Coolibah tree** And he sang as he sat...
www.virtualaustralia.com.au/general/waltzing_matilda.htm - Show matches (Cache) - 27k - Similar pages

GAMILARAAY FINDERLIST
...thawuma-li cook, to guwiya-gi **coolibah tree** gulubaa copulate, to...
...**tree**, brigalow burrii **tree**, **coolibah** gulubaa **tree**,...
www3.aa.tufs.ac.jp/~austin/GAMFL.HTML - Show matches (Cache) - 54k - Similar pages

soc.culture.australian FAQ (Part 6 of 6) (monthly posting)
...billabong, Under the shade of a **coolibah tree**, And he sang as he...
...Billabong, Under the shade of a **Coolibah tree**; And he sang as he...
www.faqs.org/faqs/australian-faq/part6/ - Show matches (Cache) - 101k - Similar pages

M & L Forum | Burke & Wills - April 11th - 12th
...had left message, carved on a **coolibah tree**: DIG 3 FT. N.W....
...Cooper's Creek, no carving on a **coolibah tree**, to say where we...
www.mcb.co.uk/mlf/current/burke/1997/apr11b.htm - Show matches (Cache) - 9k - Similar pages

A Little Bit Of Austaliana
...billabong Under the shade of a **Coolibah tree**, And he sang as he...
...billabongs, Under the shade of a **Coolibah tree**, And he sang as he...
www.chariot.net.au/~kday/australiansongs.html - Show matches (Cache) - 17k - Similar pages

More data

I hope you were impressed. While I expected the song *Waltzing Matilda* to be mentioned several times, the first result was spot-on and there were at least two other sites I wanted to explore further.

So Google uses citation information that traditional search engines ignore or don't elevate to the same importance. Google's site actually contains pages that explain what I just said in much more depth and more accurately. I glossed over the details.

You won't find fancy search strategies on Google. It pins its faith on the veracity of the hyperlinks. Hyperlinks don't lie. Web page owners only put links in their pages if they think they are worth something. Therefore whether you use capitals or not in the query is ignored, as is whether the page uses them. No alternate plurals are formed and neither are common misspellings or synonyms, so you may have to do two or more searches (say for Australia and Australian, or echidna and echidnas). And it only finds pages that have *all* the words (a simple AND function).

Google is a simple, powerful tool, which lets you be in control, and which works superbly. It is not one of those confusing applications that demand you take a course before you can use it well.

If you like Google, there are some simple instructions on their Web site for adding a Google button to your browser's toolbar. It uses *Javascript*, so you should be using a reasonably up-to-date browser, at least Explorer 4 or Netscape 4. Don't forget to tell your friends about Google. Fortunately the URL is easy to remember.

Where's the catch?

Where is Google less useful? Well if the topic is so specialized that there are only two or three pages to be found, it won't be any better than any of the other search engines. In this case you may be best advised to use a meta-search engine that merges the results from a large set of other engines.

At the other end of the spectrum, if you are trying to research topics like 'sex practices' or 'bookselling' then you are likely to find that commercial vendors of these services are far more popular than your academic interest.

Trying to see how hard it was to find my own home website I was overwhelmed by people offering merchandise for *sale* based on *Arthur the Aardvark* (a children's TV cartoon character).

Citation searching

While Google's search accuracy makes it worth knowing, it has an even more significant feature to offer. Since it has all those citation links in its database, it will allow you to search them backwards. So if you find a page you think is important (a paper perhaps, or a hub site) you can ask what were the sites that had links to it. The chances are that some of them are relevant too. In the case of journal article repositories, you may get follow-on work, or related work.

To do this, you type a special kind of query into the search box: the characters 'link:' followed by the URL of the page you want Google to use as the base for your citation search. For example, the query link:www.comp.utas.edu.au will show all the sites that Google knows have links to the School of Computing home page in the University of Tasmania. Similarly the query link:http://www.ips.gov.au/asfc/ gives links to the 134 pages that have links to the *Australian Space Forecast Centre*.

Importance of citations

I find it quite impossible to over-emphasize the importance of citation searching.

Traditional references at the end of academic papers take you backward in time (any paper referenced was always written earlier), but citation searching allows you to search forward in time. Once you find that really important seminal paper, the citations give you those people who read it and thought it important enough to put it in as a reference. These papers tell you what was wrong (or right) with it, and who used or verified the data subsequently. Once you can search links in both directions, you can explore throughout the journal literature starting from any significant seed article.

Citation searching from traditional paper journals was always possible through Citation Indexes, and more recently through electronic equivalents such as the BIDS service to UK universities (and commercially). However, academics and research students don't yet seem to have discovered how important citation searching is. To use an analogy, using references without citations is like rowing a boat without being able to look over your shoulder.

What Google does is extend the concept to the Web, and make it available free. As an e-journal, *Digital Trekking* itself has no traditional references. Instead, every reference appears as a hyperlink taking the reader directly to the source. I guess I would put a traditional reference in if the topic were only available in paper.

This emphasizes the importance of every academic in every university in the world putting all their papers online as soon as they send off the submission to a journal, with hyperlinked references. You can probably read or download this issue of *Digital Trekking*, for example, on your subscribing university or college website complete with active links. Otherwise it is also on my personal website http://arthur.sale.tripod.com/

Popularity polls

Citation searching also allows you to do some crude popularity poll comparisons on websites. For example, when I looked last, there were 6040 links from other pages to the Tasmanian Government site www.tas.gov.au and 8540 links to the University of Tasmania's home page www.utas.edu.au. Why don't you have a look at your favourite university or college home page and compare it with similar ones?

Summary

Google is not the answer to every search. But it is so far ahead of the competition in quality of results and citation capability that everyone in a university should know about it. It uses priceless information that the other search engines discard or undervalue.

Tip—Web Sheets

Many of you are familiar with Microsoft™ Excel™, but do you know how to put Excel spreadsheets on the Web, and how they can be protected against accidental disaster?

Actually putting a spreadsheet on the Web is easy. You just arrange for your workbook (saved on your PC perhaps as Sample.xls) to be stored as a file on the Web server. Then you arrange for a relative hyperlink in a Web page to that Sample.xls file. Anyone clicking on this link will cause the spreadsheet to be opened up in a window on their machine, provided they have a compatible version of Excel installed. If this does not mean anything to you, it will to your computer assistant. Be assured that if you want to put an Excel spreadsheet on the Web you can, very easily.

Why would you want to put a spreadsheet on the Web? Well, perhaps you have a blank form with formulas all set up into which a student can enter experimental data and get accurately computed consequences. You might be teaching economics, statistics, physics, ceramics or engineering to take a few examples. In each case you want the student to only change certain cell values, because if your formulas get corrupted the exercise will be meaningless—worse, it will be misleading and possibly dangerous. So how do you prepare your spreadsheet for publication?

The first thing to do is to determine which cells you want to be changeable. Sit down and think about it carefully. Then select all these cells (hold down the **Ctrl** key on a PC while you select them). Go into the **Format/Cells** menu, select the **Patterns** tab, and set all the cells to be an friendly pale pastel color, such as aqua. Close the dialog box. Now the user will know what they can change. If you want, you can do it the other way around and put the background of all the unchangeable cells to an unattractive color like light grey.

Now to actually set up the protection. Make sure you have the changeable cells selected (they may still be) and go into the **Format/Cells** menu again. This time select the **Protection** tab, and make sure the **Locked** check box is *unselected*. Close this menu. Now pull down the **Tools/Protection** menu and select **Protect Sheet** or **Protect Workbook** as you prefer. Use a password if you want to be totally sure that the viewers can't change your spreadsheet. Save the file now.

Check that you've done this all correctly by trying to change one of your locked cells. Excel should object. Try to change the unprotected cells. That should be ok. Do a complete dummy run to make sure that all the cells you intended to be changeable actually are.

Congratulations. You're on the way to putting some interactive content on the Web instead of just adding to those billions of static pages which have the behavioural attributes of corpses.

Here, for example, is a link to a page of mine: arthur.sale.tripod.com/id104.htm. The resources show people firing a glass kiln how to set up a firing schedule. The spreadsheets compute times from the settings of the kiln controller, and the forecast time of the peak temperature, when you may wish to manually intervene.

Editorial

An issue of *Digital Trekking* will be published every two months, normally in January, March, May, July, September and November.

The next issue will look at electronic strategies to aid those looking for traditional book resources for their courses, their research, or their study.

The policy of the publication is to provide professional and relevant advice on operating in the digital world to the teaching, learning and research members of a university or college. In other words to academics and teachers, their support staff, and students and scholars.

In achieving this target, emphasis will be placed on selective and valuable advice, so that most readers gain significant value from each issue. I recognize that most members of universities, colleges and schools have insufficient time to keep up with the technology and benefit from professional direction.

My background? I have been a Professor of Computer Science, Chair of Electrical Engineering and Computer Science, Chair of the University of Tasmania's Academic Senate, and a Vice-President of the University. As VP, my divisions won two Australian National awards for Student Service Excellence, and we were runner-up in the third year. I also served as National Vice-President of the Australian Computer Society. I have given seminars on five continents and travelled widely. I know the kinds of advice that universities, colleges and schools need in these times of rapid change. Now working as an independent IT consultant, I live in beautiful Tasmania.

Suggestions for topics or comments are welcome, addressed to me at ahjs@ozemail.com.au.