bookmarks, though, share them with others. The continuing value of Furl is not only in its personal web features but in the shared knowledgebase and recommendation system of the social bookmarking network.

Note that another social bookmarking service that has gotten some good reviews lately is *del.icio.us* (yes, that's the URL: http://del.icio.us). It offers some nice features if you are interested in seeing what other people have bookmarked, including an RSS feed of bookmarks added to del.icio.us on a specific topic. However, as Joshua Schachter, the creator of del.icio.us concedes, "This system is pre-pre-alpha; many features have yet to be added. Additionally, many, many bugs remain. Please be careful." (That sounds like those signs you see in gift stores: you break it, you buy it!)

My primary objection to del.icio.us is that there is no straightforward way to keep your bookmarks private. The strong default is to share your bookmarks and their associated "tags" with other subscribers. While I understand that this is the way that a social bookmarking system grows, to virtually require that all bookmarks be public reduces the system's usefulness for many researchers. If Schachter adds the ability to set the user default to private, rather than public, bookmarks, and if *Del.icio.us* gets past pre-pre-alpha, it would be a real competitor to Furl.

An RSS feed for my Search Tip of the Month is at http://www.batesinfo.com/tip.rss

What's New in Search Tools or "Alas Poor Google..."

This article is based on presentations given by Karen Blakeman at the ISG Meeting "To Google or not to Google" on 23rd November 2004, Cambridge UK and the Information Masterclass at Online 2004, Tuesday 30th November.

This review of what has been happening over the last year looks at Google's new features, the alternatives to Google (and yes there are plenty!), and desktop search (a really hot topic at present).

What has Google been up to? http://www.google.com/

Apart from its stock exchange listing, Google seemed to go into hibernation during the summer of 2004. There was little in the way of new features and it looked as though its database was not being significantly updated. Then, in response to the launch of Microsoft's new beta search engine, Google doubled its database to 8 billion pages and a new numeric range search was added to the Advanced Search page. It did not stop there, though. Google then launched Google Print, Google Scholar, Google

Libraries, Google Desktop Search and Google Suggests to name but a few.

Google Print http://print.google.com/

Google Print is aimed at publishers both large and small. Books are supplied by authors and publishers as hard copy and are scanned in by Google. Type in a search in the format 'books about.....' and Google lists those that match at the top of the results list. It appears that there is a maximum of three books per search but that could just reflect the number of books that have been supplied by publishers. Google displays an image of the page in the book that mentions your terms the most, and you can view a maximum of two pages either side. There are also links to bibliographic information, to book stores or the publishers own site where you can buy the book.

The Google Print is programme is now being expanded to include selected books from the libraries of Harvard, Stanford, the University of Michigan, and the University of Oxford as well as The New York Public Library. Google Print only works on the Google.com site and not on any of the country versions.

Google Scholar

Google Scholar (http://scholar.google.com/) enables you to search for "scholarly literature including peer-reviewed papers, theses, books, pre-prints, abstracts and technical reports from all broad areas of research." A wide range of academic publishers are covered such as professional societies, pre-print repositories and universities, as well as scholarly articles available across the web. There is no source list so it is down to guesswork and experience in working out what is covered.

The advanced search options are limited to author, journal and date of publication. The author search is a bit hit and miss and varies depending on the format in which you type the name. 'K H Blakeman' and 'KH Blakeman' (without the space between the K and H) give different results.

Google Scholar automatically analyses and extracts citations. This means that your results may include citations of older works, books or other off-line publications. The results are listed by relevance, which is a disappointment for those of us accustomed to sorting this type of literature search by date, author etc. Google Scholar is still in beta so there is hope that at least a date sort will be added.

Google Suggests

Google Suggests can be found on the Google Labs page (http://labs.google.com/). Simply start typing in your search and Google comes up with a list of suggestions for completing your strategy together with the number of results each will give. My search on gin

and tonic came up with 'gin and tonic recipe', 'gin and the devil', and 'gin and tonic movie'! (See figure 1)

Figure 1: Google Suggests

Limit on number of search terms increased

Google has at long last increased the limit on the number of search terms from 10 to 32. As many commentators have said, it is long overdue and makes it a lot easier to carry out more sophisticated searches and searches for quotations.

Alternatives to Google

So what is wrong with Google? For many searches - nothing. Google comes up with the goods most of the time for the majority of users but there are times when one just cannot find relevant information. This can be because Google, for a variety of reasons, has not covered the pages that are relevant but more often the most relevant information is buried way down near the bottom of your list of results. Different search tools rank the results of the same strategy differently and what is number 22,736 in Google's list of pages may be number 5 in Yahoo's.

Try comparing the performance of different search engines in Thumbshots Ranking (http://ranking.thumbshots.com/). This covers Google, Yahoo, AltaVista, MSN, AltheWeb, Teoma and Wisenut, and you can compare two at a time. Type your search strategy into one of the search boxes and then select the tools you want to compare. Thumbshots compares the first 100 pages found by the search engines and represents them as two lines of circles. The pages that are common to both search engines are represented as solid blue circles and linked by a blue line. (see figure 2)

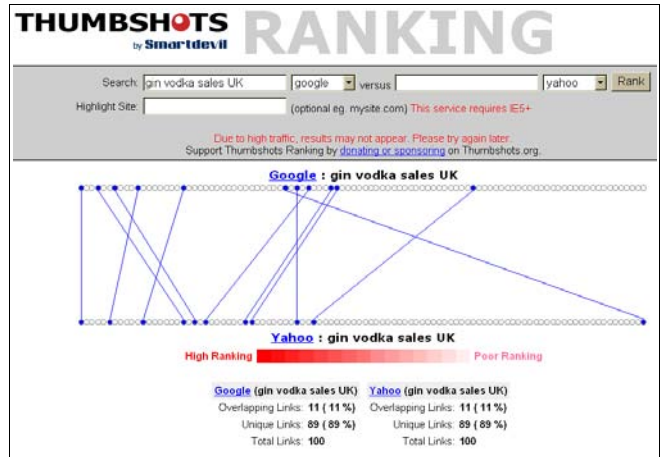Unfortunately, Thumbshots Ranking only works in Internet Explorer.

Figure 2: Thumbshots Ranking

Yahoo http://www.yahoo.com/ or http://www.yahoo.co.uk/

Over the last 18 months Yahoo has bought up Inktomi, Overture, AlltheWeb and AltaVista. At the start of 2004 it dumped Google as its partner for web search and launched its own web database. Yahoo has not revealed the size of its database but it is certainly giving Google a run for its money and it does sometimes gives better results than Google.

How do Google and Yahoo compare?

1. The advanced search features of Yahoo and Google are very similar.

2. Yahoo has the advantage that it supports all three Boolean operators (AND, OR, NOT) and parentheses for nested searches. Google only supports OR and you cannot carry out sophisticated nested Boolean searches.

3. Google has a link: command that finds pages that link to your specified URL, but it will only find links that match your URL exactly and shows only a small selection of linking pages. Yahoo has two commands. 'Link:' finds pages that links to a specific URL, for example *link:http://www.rba.co.uk/search/* (you must include the http://), whilst 'linkdomain:' finds pages containing links to any page on the specified domain, for example *linkdomain:rba.co.uk* (do not include the http:// with this command). The link command is a useful way of finding similar types of pages, the assumption being that pages that link to another are often similar in content.

4. Google indexes the first 100K of a page: Yahoo indexes up to 500K. This may not sound significant when searching plain old html web pages but can make a difference when looking for file formats such as PDFs, which can be very large. Some

search gurus have spotted that Google sometimes drops the 100K limit but it varies from day to day and results are not reproducible. This could be something that Google is working on.

5. As well as the usual PDF and Microsoft Office file formats Yahoo allows you to restrict your search to RSS/XML formats. Perfect if you are searching for news feeds on a particular subject.

6. Google has a number range search.: Yahoo does not. Specify two numbers separated by two full stops with no spaces, and include a unit of measure or some other indicator of what the number range represents. For example DVD player £50..150, or toblerone 1..5 kg. Number ranges can be years, weights, prices, temperature, concentrations etc.

7. Google has a synonym search: Yahoo does not. Precede your term with a tilde (~) and Google will look for similar or related terms. There is no information on how the synonym lists have been compiled and it is not always accurate. This may not matter for more general searches but can be an issue with scientific or technical searches. For example a synonym search on the artificial sweetener ~aspartame will pick up products that contain aspartame and sucralose, which has a different structure and properties.

8. Both Yahoo and Google have an excellent current news search with similar advanced search features, including options to limit your search to one source or to sources from one country. I have found that for most of my own searches Google sometimes has better International coverage, but it does vary depending on the search. Both offer news alerts by email but Yahoo can also provide them as RSS feeds (available only in Yahoo.com at present).

MSN http://www.msn.com/

Microsoft pulled out all the stops as far as publicity was concerned when it launched this but I was very disappointed with it. It has a simple uncluttered default search screen - the current fashion - but I found that results for my test searches were poor when compared with Google and Yahoo. Advanced search features are under "Build a search" and are restricted to the link command, country/region, language, and site/domain. There is no file format command. The News search offers RSS feeds for alerts, which are provided by Moreover.

Exalead http://www.exalead.com/

Exalead was launched in October 2004. With 1 billion Web pages in its index it may seem small in comparison with the the likes of Google, who claim to have 8 billion pages, but the quality of Exalead is

excellent and there are some cracking good search features.

For starters it supports wild cards (an asterisk) that can be used to represent any number of letters after a specified string of letters. Then there is word stemming that can be set as the default under preferences, and which finds the variants of a word such as plural/singular and verb conjugations.

If you are not sure how to spell a word try phonetic search or the approximate spelling search on the Advanced Search screen. Another approach is to use the pattern matching feature. Begin and end the "pattern" with a forward slash (/) and use a full stop or period to represent single characters in the pattern. If you are not sure how many characters are in the middle of a word use the full stop followed by an asterisk (*). The help file suggests that pattern matching is ideal for solving [shouldn't that be 'cheating at'?] crossword puzzles!

Full Boolean search is supported (AND, OR, NOT) and there is a proximity command (NEAR) that will search for words within 16 terms of one another. Filetype, country and language search options are supported. RSS and blog search options are promised for the near future.

The results page is pretty nifty too. The usual list of results is augmented with a snapshot of each web page (you can switch this off if you prefer or just display the snapshots), suggestions for related terms, options to view the results by geographical location and file type (for example .doc, .pdf, .xls).

Find.com "True Business Search" http://www.find.com

Find.com was launched in June 2004 and concentrates on business search. It combines results from its own index of business web sites with those from some of the major search engines. It also finds priced reports - "premium research content" - from sources such as Thomson Gale, The Gallup Organization, Frost & Sullivan, BNET, Marketresearch.com, Datamonitor etc.

The Web option is a meta search tool and appears to cover at least Google, Yahoo, MSN, AltaVista, AlltheWeb, SmealSearch, Scrius and About. This option also includes Business Web which are business web sites identified and indexed by Find.com.

Find.com automatically searches for all of your words but supports the standard Boolean AND, OR, and NOT operators all of which must be in capital letters. The standard double quotes can be used to specify phrases.

The Advanced Search screen has options for searching the full text (the default) or titles only, and keyword search (default) or concept. The concept search takes your terms and looks for similar or related

terms. For example, type in automotive and Find.com will pick up references to automobile, motor trade, motoring etc. The drawback is that if you have more than one term in your query you cannot specify which of them is to be treated as a concept. It is all or nothing.

There are also "Threshold" options that enable you to increase the number of documents that are shown from the default of 100 and to change the "minimum relevance percentage" of 10. Worryingly, nowhere on the home page or on the results page does Find.com explicitly tell you that by default it only shows you a small selection of records. This really ought to be made clearer on the home page.

There are several more advanced features but these only apply to the Business Web and Premium Research collections. You can use a question mark (?) inside a word to represent a single character, or an asterisk (*) inside or at the end of a word to represent one or more characters. If you are not sure of spelling use the "fuzzy search" feature by adding a tilde (~) at the end of a single word.

There is also a proximity option. Place your terms inside double quotes and then a tilde at the end of phrase together with a number. For example: "research development"~5 looks for the terms research and development within five words of each other.

You can boost the relevance of a term or phrase in your search by using the caret, "^", symbol together with a boost factor (a number) at the end of the term. For example: "production statistics"^4

Results are sorted by relevance. On the left hand side of the screen is an "Organize results by" column. This enables you to view your results by topic, document format (htm, pdf, ppt etc), site and source. The Topic option clusters results into subject folders and can help narrow down your search.

Answers.com http://www.answers.com/

GuruNet launched its free reference service Answers.com at the beginning of 2005 offering "the best definitions and explanations for over 1 million topics". Information on people, places, words and names is drawn from dictionaries, encyclopaedias and selected web sites. A search on Winnie the Pooh, one of my regular test searches (yes, sad isn't it?) came up with information gleaned from Who2, the New Dictionary of Cultural Literacy, and Wikipedia. For the artificial sweetener aspartame Answers.com came up with definitions from a standard and a medical dictionary, an entry from Wordnet, the entry in Wikipedia including its Chemical Abstracts Number and structure, and translations into various languages. There are links to other topics that mention your term or phrase and to searches on Google for web pages,

images and news. The blogs search option uses Technorati.

If you find you are becoming hooked on the service, as I totally am, you can download the 1-Click Answers bar. Once installed there is an additional feature that automatically runs an Answers.com search when you Alt-Click on any word or phrase. This works on any document in any program on your screen.

Turboscout http://www.turboscout.com/

This is a really neat tool that allows you to search the results of 23 search engines by typing in your search, and then clicking one by one on the engines that interest you. This may sound tedious and I could not at first see the point of it. The really useful aspect of it is that you decide which tool to search in. It could be two or three under the Web tab for example Google, Yahoo, Vivisimo and then you could switch to the Reference tab and look in Wikipedia, FindArticles or Scirus for example. See figure 3.

Figure 3: Turboscout

Desktop Search - the latest hot topic

When Microsoft announced that it was working on a tool that would combine Internet with desktop search, the competition started beavering away to beat them to it. The idea behind desktop search is that as well as, or instead of, scouring the web for information on a topic you can also search documents stored locally on your computer. These "documents" could be cached web pages, email messages, Microsoft Office documents, PDFs. In theory anything. In practice, if you use anything other non-Microsoft products for example Lotus Notes, Open Office, Pegasus Mail, Thunderbird etc. you may find you are left out in the cold.

Blinkx http://www.blinkx.com/

One of the first of the current batch of desktop search launches, Blinkx differs from the others in that it claims to eliminate the need for keywords or complex search methods and uses "self-learning algorithms to understand the context" of what you are reading on your computer screen.

When you first run Blinkx, you can tell it which directories you want it to index. The indexing can take anywhere between a few minutes to a few hours

depending on the number of files on your system. It indexes .doc, ppt, Excel, PDF, txt, html and various media files. As for email, it only supports Outlook, Outlook Express and Eudora. I had Blinkx running for 6 days. At the start, the local documents it found for me were relevant but the web results were useless. The company does emphasise that this is a "concept" as opposed to a "keyword" search tool, but there is no explanation as to how the concepts are identified and analysed. I assumed that Blinkx needed to learn more about what I regarded as relevant, but as the days passed the Internet results were no better and my local hard disk results became increasingly bizarre. I suspect that Blinkx works best with narrowly defined subjects. If, like me, you research anything and everything it loses the plot completely.

Copernic Desktop http://www.copernic.com/

Copernic is well known for its search agents that cover multiple search tools and "invisible web" resources, as well as for its web page tracking program. Installation is straightforward but the first time it starts up it can take a while to index the files on your hard disk, as the case with all of these types of tools. It automatically stops indexing when it detects significant activity but you can pause and restart it manually. By default it indexes the My Documents folder but you can customise it by adding other directories or even whole drives. Copernic Desktop can search PDF, XLS, PPT, DOC, RTF, TXT, WP and HTML files. It also indexes your search history, Internet Explorer favorites, Outlook email and contact lists. The most recent version has added Mozilla and Firefox support. A really neat feature is that under the Advanced Options you can tell it to index other file types and add file extensions that are to be indexed as text files.

Google Desktop Search http://desktop.google.com/

The beta version of Google Desktop Search (GDS) gave the impression that it was launched before it was ready in an attempt to beat Microsoft and Yahoo. It was limited to Microsoft Office formats, Outlook, Outlook Express, AOL Instant Messenger, TXT and HTML but did not index PDFs. The most recent official version has added support for PDF files, Mozilla, Firefox, Netscape and Thunderbird. There is also a plugin that indexes Open Office and Star Office file formats.

Some of the original security and confidentiality issues have been addressed but you need to be aware of their significance and know how to switch various options off. GDS can index cache secure https pages that you view, for example online bank statements: untick the relevant box under Preferences. Password protected and encrypted files can also be indexed: again untick the box under Preferences.

Another problem is that documents remain in the GDS cache on your PC even after you have deleted the original files. So it is possible for anyone using your computer to unearth those embarrassing emails and

online chats that you thought were long gone! More importantly for corporate users, though, is that this feature could conflict with document retention and management policies. There are instructions for removing documents from the cache at http://desktop.google.com/features.html#remove but it is far from straightforward and you have to know exactly which files and emails you want to delete.

MSN Desktop Search http://www.msn.com/

MSN Desktop Search is part of the MSN Toolbar Suite Beta and installs as a browser toolbar and optionally in the Windows taskbar. MSN Desktop indexes Office documents, Outlook and Outlook Express email, photos, music and email attachments. You can specify exactly what you want it to index: email and My Documents, email and all of your hard disc, or individual directories. It spots when you are active on your PC and suspends indexing quite quickly, but there is also a "snooze" button you can use to stop it immediately. The main drawback with MSN Desktop is the lack of support for non Microsoft email clients, browser and document formats.

Ask Jeeves Desktop http://sp.ask.com/docs/desktop/

Yet another search tool that assumes that everyone is using Microsoft Office, Outlook and Internet Explorer. Other files that it currently supports include images, MP3 music, video files. It also searches titles and metadata from Adobe PDF files but not the text. It does not index inside ZIP archive, Outlook attachments, contacts and one odd omission is that it does not index the content of Excel files. This a beta release, though, so additional functionality may be added later. It does take a while for it to realise when you want to use your computer and activity has increased, and there is no quick way of forcing it to cease and desist apart from shutting it down altogether.

Yahoo Desktop http://desktop.yahoo.com/

Yahoo is the latest search engine to launch a desktop search program and, in my view, it is by far the best so far. It searches the usual "popular" file formats such as PDF, Microsoft Office, html etc "plus over 200 more". A full list can be found at http://desktop.yahoo.com/filetypes. It retrieves any file on your computer and displays results as you type. The preview option is excellent and most files look exactly as they should in terms of layout. In my tests, it found all the files containing the text I specified, including files without extensions and my Open Office files. My only gripe is that when it is indexing your files, it takes a while to pause when you start using other programs.
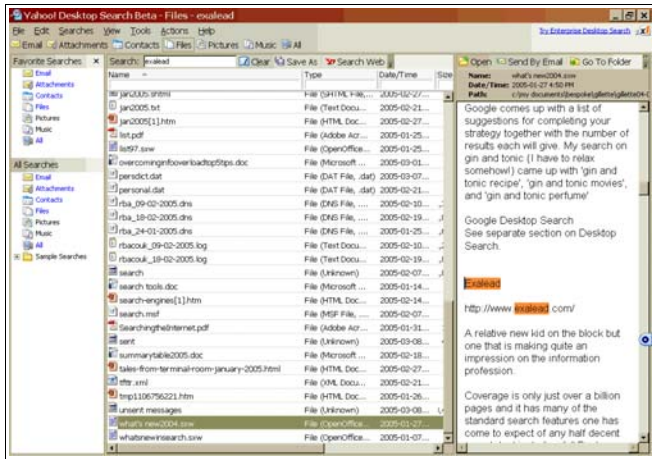
Figure 4: Yahoo Desktop

What next?

Who knows! Google will continue to be the search engine that many people go to first but Yahoo is a offers strong competition. Both continue to add features but it is a question of how relevant these are to us. For example the most recent Google launch was a weather search, but Google decided that only people in the US were interested in conditions outside! New tools will continue to be launched. Many of these will disappear without trace. Others will fail to live up to expectations, and features that promise to be "coming soon" will never materialise.

If I had to pick one to watch, it would have to be *Exalead*. It is a comparative lightweight in terms of database size but has some seriously good search options. If it continues to develop new features and launches its promised Desktop Search it could be a serious contender for Google's crown.

Karen Blakeman
RBA Information Services
karen.blakeman@rba.co.uk

Short Notes

Study shows online citations don't age well

A study conducted by two academics at Iowa State University has shown a remarkably high rate of "decay" for online citations.

Michael Bugeja, professor of journalism and communication, and Daniela Dimitrova, assistant professor of communication, looked at five prestigious communication-studies journals from 2000 to 2003 and found 1,126 footnotes that cite online resources. Of those, 373 did not work at all, a decay rate of 33 percent; of those that worked, only 424 took users to information relevant to the citation. In one of the journals in the study, 167 of 265 citations did not work.

Bugeja compared the current situation to that of

Shakespearean plays in the early days of printing, when many copies of plays were fraught with errors due to the instability of the printing medium.

Anthony T. Grafton, a professor of history at Princeton University and author of a book on footnotes, agreed that citation decay is a real and growing problem, describing the situation as "a world in which documentation and verification melt into air."

Abstract taken from Edupage 14 March 2005; the article is at Chronicle of Higher Education, 14 March 2005 (sub. req'd). See also Oxford Technology Book Predicts Growing Citation Problem

ESRC Society Today

The UK's largest funding agency for research and postgraduate training on social and economic issues - the Economic and Social Research Council (ESRC) - is soon to launch a major new website in the UK.

ESRC Society Today will offer academics, students and researchers unrivalled, free access to high quality social and economic research available, planned and in progress.

As well as bringing together all ESRC-funded research, the website will allow users to cross search material from other key online resources such as SOSIG, the UK Data Archive, the Office of National Statistics and Europa among others.

It will include early findings, full-texts and original datasets and allow users to establish online discussion fora, track down researchers in particular subject fields or find details on ESRC funding and training opportunities. The emphasis will be on accessibility and ease of use and visitors will be able to register for regular news bulletins and email alerts on their areas of interest.

ESRC Society Today - http://www.esrcsocietytoday.ac.uk - is scheduled to go live for a test period in April 2005 before being publicly launched in Summer 2005, replacing the current http://www.esrc.ac.uk and http://www.regard.ac.uk sites.

Librarians are invited to register for further information by emailing societytoday@esrc.ac.uk

Farewell

This is my final issue as Editor of *eLucidate*. Gary Horrocks (gary.horrocks@kcl.ac.uk) will be looking after the Journal whilst the Committee seeks to appoint a new Editor.

Peter Chapman