

Top Tips for Expert Searching Inforum 2005, Prague

Karen Blakeman, RBA Information Services

New Google services

Google Mail <http://gmail.google.com/>

Free web based email service offering over 2GB of storage space. You need to be "invited" but there are plenty of people around who can send you an invitation (contact myself at karen.blakeman@rba.co.uk). **Warning:** Google Mail is tied in very closely to its other services such as My History and News Alerts. If you never log out of Google Mail you will find that your search history is being automatically recorded and news alerts that you may want under another account are hijacked.

Google Print - <http://print.google.com/>

- Books supplied to Google by publishers
- Google digitizes them
- Search on books about.... for example books about hubbert
- 3 books listed near the top of the results list
- can search within the book
- limit on number of pages that can be viewed
- Information about the book and links to book stores

Google Scholar - <http://scholar.google.com/>

- "search specifically for scholarly literature, including peer-reviewed papers, theses, books, preprints, abstracts and technical reports from all broad areas of research"
- no source list
- limited "advanced search" and author search unpredictable
- articles ranked by relevance only
- many articles are priced

Google Suggests - <http://labs.google.com/> and click on Google Suggests

Start typing in your search and Google suggests additions to your search together with the number of results.

My Search History - <http://labs.google.com/>

Once logged in to your Google Account, which can be your gmail account, Google keeps a record of all of your searches by date. Useful for revisiting searches. You can pause the monitoring and delete records of selected the searches

Google Desktop <http://desktop.google.com/>

For searching documents on your local PC. Not the best desktop search tool so far and there are implications for document management and privacy

New Google Search Features

1. Database increased to over 8 billion pages but this means there is more rubbish to sift through and that one needs to make more use of the advanced search features to get relevant results.
2. The number of terms you can have in your search has been increased from 10 to 32

Top Tips for Expert Searching: notes and resources

3. The numeric range search enables you to search for numbers within a specified range for example distances, weights, years (sometimes useful for limiting to documents published within a specific time span), prices, temperature.

Syntax is: first number followed by two full stops then the second number. There are no spaces between the full stops and the numbers. You can also include a unit of measure. For example:

toblerone 1..5 kg
DVD player \$100..200

4. Synonym search - to search for synonyms of word prefix the word with a tilde (~). For example ~aspartame will find documents that mention other artificial sweeteners such as sucrose or products that contain artificial sweeteners.

Google advanced search

Use the Advanced Search screen to refine your search and reduce the number of results:

- language
- file format (filetype:) - for example select XLS format if you are looking for spreadsheets of statistics, PDF or DOC formats if you are looking for official documents, PPT for presentations
- domain (site:) to search for particular type of organisation e.g. government, academic. Also to search large sites that have appalling internal search engines.
- Similar pages (related:) - looks for pages that similar in type and content (usually a maximum of about 30)
- pages that link to a known page (link:) - assume that pages that link to another are often similar in content. Note: Google's link command is not as good or as comprehensive as Yahoo's.

From the standard Google search box:

- define: for definitions for example define:GDP
- Use an * to stand in for a word in a phrase for example "phenol * extraction" to find documents in which phenol is separated from extraction by one word. Use additional asterisks for each word of separation, for example "phenol * * * extraction" to find phenol separated from extraction by three words.
- Use the + sign before a word to stop automatic stemming

"Google sinker"

Increase the importance of a search term in your strategy by repeating it several times. For example:

beer "market share" france germany czech
beer "market share" france germany czech czech
beer "market share" france germany czech czech czech

all give different results!

Also works in Yahoo.

Why try other search engines?

Different search tools have:

- different coverage
- different way of sorting results
- different search features
- different types of resource

Compare some of the major search engines using Thumbshots Ranking (<http://ranking.thumbshots.com>). This shows the overlap - or lack of it - in the first hundred results from two search engines at a time. The degree of overlap varies depending on the search.

Meta search tool Dogpile has produced a short paper on search engine overlap at <http://missingpieces.dogpile.com/WhitePaper.pdf>. There is also a tool at <http://missingpieces.dogpile.com/missingpiecestool.aspx> that compares Google, Yahoo and Ask Jeeves. A graphic shows how many results are in only 1, 2 or in all 3 search engines for a particular search.

Yahoo!

<http://search.yahoo.com/>

- Launched at the beginning of 2004
- No indication of the size of the database
- Search features very similar to Google
- Key features
 - searches first 500K of a page (Google only 100K)
 - full Boolean search
 - link and linkdomain command better than Google
 - RSS/XML filetype search
 - News alerts available as RSS feed

Yahoo link and linkdomain commands compared with Google

Google:

link:<http://www.zefix.ch/> (68 results)

Yahoo:

link:<http://www.zefix.ch/> (1100 results) - finds pages that link to this individual page

linkdomain:www.zefix.ch (1290 results) - finds pages that link to any page on this site.

MSN

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

- Launched in autumn 2004
- 5 billion pages
- "Search Builder" = advanced search options
- No filetype search option in advanced search
- Results tend to be "consumer" oriented
- News Alerts available as RSS

Exalead

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

- 1 billion pages
- Full Boolean search
- NEAR command - within 16 words of each other
- Supports wild cards (*) in the middle and at the end of a word

Top Tips for Expert Searching: notes and resources

- Exalead Advanced Search
 - phonetic search
 - approximate spelling
 - automatic stemming
 - Pattern matching - good for solving (cheating at?) crossword puzzlesstart pattern with a forward slash, represent each missing letter with a full stop and finish pattern with a forward slash e.g. /.h.s.c..n/. use a full stop followed by an asterisk to represent one or more letters e.g. /psych.*ist/

Results can be sorted by date (newest to oldest or oldest to newest), display thumbnail of the page next to each entry, show related terms, and can be viewed by file format e.g. PDF, DOC

Kartoo

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

- Meta-search tool with graphical representation of results
- Extracts related terms from documents - very useful for identifying terminology in unfamiliar subject areas.

Unique search features of some of the major search tools

Google - numeric range search, synonym search, define command, Google Suggests

Yahoo - RSS/XML file format in advanced search

Exalead - phonetic search, approximate spelling, pattern matching, wildcards, NEAR command, related terms

Kartoo - graphical representation of results, related terms

Which search tool to use?

Synonyms and related terms - Google, Exalead, Kartoo

Wild cards, variations on words - Exalead wild card, phonetic search, approximate spelling, pattern matching

Proximity search - Exalead

Numeric range search - Google

Quick facts and reference queries

Answers.com

- "topic-based snapshot"
- 100 authoritative encyclopedias, dictionaries, glossaries and atlases

Wikipedia

<http://www.wikipedia.org/>

- free-content encyclopaedia that anyone can edit
- editors required to compile a balanced article including references to other sources
- good for quick reference and for links to other related sources

Also:

<http://acronymfinder.com/>

<http://dictionary.com/>

<http://encyclopedia.com/>

<http://brainboost.com/>

Evaluated listings

Annotated directories on a particular subject, industry or type of information

Provide access to recommended resources on a topic

Expert human assessment of resources

Examples:

- <http://www.eco5.com/> for finance and economics
- Biogate <http://biogate.lub.lu.se/> for "1000 best links in the biological sciences"
- Official Statistics on the Web <http://www.library.auckland.ac.nz/subjects/stats/offstats/>

How do you find them?

BUBL Link <http://bubl.ac.uk/>

Pinakes, a subject launchpad <http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html>

Professional or trade association web site, by personal recommendation or by chance

Meta search tools

Take your search and run it in several search engines at once

Examples:

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

<http://killerinfo.com/>

<http://vivissimo.com/>

<http://ixquick.com/>

<http://dogpile.com/>

<http://turboscout.com/> (searches one at a time)

<http://turbo10.com/> (build your own!)

RSS and blogs

What is RSS?

A way of delivering headlines and stories. Stands for Really Simple Syndication, or Rich Site Summary, or RDF Site Summary. More information at <http://www.rba.co.uk/rss/rss.htm>

You need a a program to "read" the feeds, for example <http://www.bloglines.com/> which is web based or a desktop reader such as Feedreader.com (<http://www.feedreader.com/>).

For getting started bloglines there is a tutorial at <http://tinyurl.com/ap42n> . There is a list of readers at http://en.wikipedia.org/wiki/List_of_news_aggregator/

Blogs

Blogs are online journals or diaries and vary in quality from superficial irrelevancies to extreme erudition. Blogs published by industry or subject experts are an excellent way of keeping up to date in a particular area.

Searching for RSS feeds and blogs

Yahoo Advanced Search, limit to RSS/XML file format

Also RSS and blog search tools:

<http://bloglines.com/>

<http://blogdex.net/>

<http://blogdigger.com/>

<http://daypop.com/>

<http://technorati.com/>

<http://feedster.com/>

<http://blogpulse.com/>

Desktop Search

Programs that search your PC for documents, emails, chat, IM messages, web pages etc. Useful for tracking down "lost" files but not a replacement for structured, well managed document folders.

Examples

Google Desktop - <http://desktop.google.com/>

MSN - <http://www.msn.com/>

Copernic Desktop - <http://www.copernic.com/>

Yahoo desktop - <http://desktop.yahoo.com/>

Document types supported

Varies depending on the desktop search tool - usually at least MS Office, html, text files. May not support OpenOffice, Thunderbird, Eudora, Firefox, Netscape etc.

Three useful references

1. UW E-Business Institute "Benchmark Study of Desktop Search Tools" <http://www.uwebi.org/>. Free of charge
2. Desktop Search Handbook - an Office Watch guide - <http://shop.office-watch.com/dsh/> US\$ 14.95, e-book, updates are free
3. Desktop Detectives, Davey Winder, Information World Review, May 2005, Issue 213, pp.19-21

Google Desktop Search

Range of file formats that are supported is increasing with third party plugins.

Problems: indexes documents as you view them, even password protected files unless you tell it not to; includes https files unless you tell it not to; sends back anonymous information about your searches unless you tell it not to; problems with persistent cache and indexes even after you delete the original document (remove function not easy to use).

MSN Desktop

Very Microsoft-centric and document formats support is limited

Copernic Desktop

Supports "the usual suspects" plus Firebird, Mozilla Netscape and Thunderbird

Individual results are displayed in a Quick Preview pane below the search results and search terms are highlighted. Results are categorised into groups that change depending on search type e.g. email, file, music.

Yahoo Desktop

Supports over 200 document types including Open Office and Star Office + media files. Searches inside zip files and has an excellent preview function.

Which Desktop Search?

Karen's personal choice

1. Yahoo
2. Copernic
3. MSN
4. Google

Tips on storing and organising resources

1. Organise your bookmarks/favorites into folders (and back them up!)
2. Add frequently used sites to your links tool bar. See <http://www.rba.co.uk/search/toolbar.htm> for help on how to do this.
3. Copy URLs and descriptions to your own web page or word document
4. Firefox users - install the Copy URL extension. This enables you copy the URL, page title and highlighted text of a web page in one go and to paste it into the application of your choice.

Netsnippets

<http://netsnippets.com/> - according to the web site "Capture. Organize. Share".

Stores selected text, whole web pages, files along with your comments, helps you produce reports and bibliographies and share that information with colleagues and friends.

Storing and sharing bookmarks online

There are several web based services that you can use to store, comment on and organise resources. You can keep these bookmarks private, make them totally public or share them with selected individuals. They are ideal if you are on the move a lot and are always using different computers, also for sharing project resources amongst a group of widely dispersed co-workers. The disadvantage is that such services could vanish overnight so make local backups of vital collections.

Examples:

<http://furl.net/>

<http://Spurl.net/>

<http://del.icio.us/>

<http://de.lirio.us/>

<http://connotea.org/>

Top Tips for reducing information overload - see separate document.

A copy of this document and links to related documents can be found at

<http://www.rba.co.uk/inforum/>

Contact details

Karen Blakeman

RBA Information Services

email: karen.blakeman@rba.co.uk

web: rba.co.uk

Tel: +44 118 947 2256

Fax: +44 20 8020 0253