



Tales from the Terminal Room November 2006, Issue No. 74

Editor: Karen Blakeman

Published by: RBA Information Services

ISSN 1467-338X

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

Tales from the Terminal Room (TFTTR) is a monthly newsletter, with the exception of July and August, and November and December, which are published as single issues. TFTTR includes reviews and comparisons of information sources; updates to the RBA Web site Business Sources and other useful resources; dealing with technical and access problems on the Net; and news of RBA's training courses and publications.

Tales from the Terminal Room can be delivered via email as plain text or as a PDF with active links. You can join the distribution list by going to <http://www.rba.co.uk/tftr/> and filling in the form. You will be sent an email asking you to confirm that you want to be added to the list. TFTTR is also available as an RSS feed. The URL for the feed is <http://www.rba.co.uk/rss/tftr.xml>

In this issue:

- Search Tools
 - Windows Live new (ish) search features
 - Undocumented Yahoo commands
 - Unclustering results
- Assessing the Quality of Information – Top Tips
- Web exceeds one million web sites
- Browser shares October 2006
- Online Information 2006
- Conference and seminar presentations available
- Information Resources
 - Blogs and RSS for competitive intelligence
 - Passport launched for OurProperty and PetrolPrices.com
- Gizmo of the month
 - Firefox add-on – Resurrect Pages
- Meetings and Workshops
 - Online Information 2006, London
 - Advanced Internet Search Strategies, Manchester UK

Search Tools

Windows Live new(ish) search features

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

The image search in Windows Live is not exactly new but I have only just spotted the significance of the scroll bar on the right hand side of the results screen. Scroll down and you'll quickly see that Live lets you go through all of your results on one single page. Much better than the usual default 10 or 20 displayed in the other search engines. Also, if you search for a person a list of related people appears on the right hand side of the screen.

The Video search option seems to come and go from day to day. When It is there and working it comes up with what I consider to be sensible results, but then I am usually looking for business related files such as TV interviews or podcasts on a company's annual results from CEOs. There are occasions, though, when it keels over completely. Just remember that it is still in beta.

The linkfromdomain command definitely is new. This gives you a list of links to other pages from your specified domain. For example linkfromdomain:rba.co.uk. This can be useful if you have a trusted source of information and want to see what other sites they recommend or link to. It can also be used as a means of assessing the bias of a site so that you can see the diversity, or lack of it, in the pages that are 'cited'.

Undocumented Yahoo commands

In his pre Internet Librarian workshop on power searching, Greg Notess highlighted some of the undocumented Yahoo commands that the search engine inherited when it acquired the Inktomi database. Former Hotbot users may remember some of them. For example the region command enables you to restrict your search by continent or region, for example region:europe. The list that Greg gave included:

region:africa
region:asia
region:centralamerica
region:downunder [This really does pick up Australian and New Zealand sites. Honest!]
region:europe
region:mediterranean
region:mideast
region:northamerica
region:southamerica
region:southasia

Another command is originurlextension. If you have ever tried to limit your search in Yahoo by filetype you will have discovered that you have to use the box and drop down list on the advanced search page. The filetype command only finds pages that contain links to documents in the specified format; your search terms will appear in the web page but not necessarily in the document itself. The equivalent filetype command in Yahoo is originurlextension, for example originurlextension:pdf .

These and other Yahoo commands can be found on Greg Notess's site at

<http://www.searchengineshowdown.com/features/yahoo/>

Unclustering results

Another command that Greg Notess came up with at his ILI pre-conference workshop in October enables you to 'uncluster' pages from the same site in Google results. Normally Google shows you just two pages from the same site regardless of how many contain your search terms. You have to click on the 'More results from....' link to see them all. To uncluster the results, run your search as normal and then add &filter=0 to the end of the URL of the search page.

In Yahoo you can uncluster results by adding &dups=1 to the results page URL.

Assessing the Quality of Information: Top Tips

Or: Paranoia 'r' us

This is a list of Top 10 Tips that the participants of Assessing the Quality of Information compiled at the end of a workshop held at TFPL in London on 31st October 2006. On a scale of 1 to 10, most of the delegates started out with a paranoia level of around 7 or 8. By the time they had worked through half the exercises a couple of them had increased that to 25-30! Paranoia had eased off slightly by the end of the day and at least they had a toolkit at their finger tips that they could use to help evaluate and assess the quality and validity of information.

1. Check **who is behind the domain name** of a web site using <http://www.allwhois.com/>. The contact details sometimes just give the ISP or service that organised the domain name for the web site owner but at least it is a starting point if you need to contact the owner to discuss any issues about the content. If someone really wishes to hide, they can use an agent to do the registration for them and in that case there is little one can do to track down the real owner. Note that you can only find out who owns a domain name; you cannot take a person's or company's name and find out which domain names they own.
2. Try the **Wayback Machine** (Internet Archive) (<http://www.archive.org/>) for tracking down pages or sites that have disappeared. Type in the web site URL or the URL of the document/page you have 'lost'. This can pick up pages no longer cached by the search engines (see number 3 below). This trick is not guaranteed: some sites have asked to be removed from the archive or have designed their pages so that they automatically refresh to the most recent page. This can also be a useful tool for reviewing how a company presented itself on the web in the past and how organisations have evolved, both of which can be useful components of assessing quality.
3. Look at the search engine **cached copies** of pages for more recent past pages. This is especially useful if the current web page that you found via Google et al does not seem to resemble your search strategy in any way. The cached copy is the copy that the search engine has in its index and it will also highlight your search terms within the page.
4. Use **links to and from the site or page** to find pages that are similar to a known quality page (pages of similar content tend to link to one another), or to see what other people saying about the page in terms of quality and the authority and of those that link to it. Use Windows Live (<http://www.live.com/>) . For pages that link in to your known or 'suspect' page use the link and linkdomain commands.

Link will find pages that link to an individual page, for example:
link:www.rba.co.uk/sources/stats.htm

Linkdomain will find pages that link to anywhere within a web site, for example:
linkdomain:rba.co.uk

To find out what page a site links to (can give you an idea of bias, political stance, ideology etc) use linkfromdomain, for example: linkfromdomain:rba.co.uk

5. Use **'hoax busting' sites** for if you are suspicious about a site or a 'well known and accepted fact'. Examples are:
<http://www.snopes.com/>
<http://hoaxbusters.ciac.org/>
<http://www.vmyths.com/> (concentrates on virus myths and hoaxes)
<http://www.regrettheerror.com/>
 6. If relevant and appropriate **double check information** and data with other independent sources (not always possible and you may find yourself going round on circles chasing sources that quote each other!)
 7. Use the search engine **advanced options to focus your search**. For example the domain and site command or box to limit your search to, for example, UK government sites (gov.uk), academic sites (.ac.uk, .edu etc), a known trusted site.
 8. Use **different search tools** and their features to give you results that are prioritised in a different order or for suggestions on alternative search strategies:
Yahoo – <http://search.yahoo.co.uk/> – for results sorted in a different order from Google
AltheWeb Livesearch – <http://livesearch.alltheweb.com/> – for results that change as you type and suggestions for alternative search terms
Ask – <http://www.ask.co.uk/> - for ways of narrowing down or broadening your search
Exalead – <http://www.exalead.com/> - for its unique advanced search commands and related terms
Windows Live – <http://www.live.com/> – for its link, linkdomain and linkfromdomain commands
- Think about using different types of resources for example reference sources, video/audio, blogs and RSS feeds (yes, there are some good ones around!). Have a look at Trovando (<http://www.trovando.it/>) for some starting points. And don't forget evaluated listing such as Intute (<http://www.intute.ac.uk/>) and, for business, Alacrawiki (<http://www.alacrawiki.com/>).
9. If you are looking for up date to market research etc. use market research content aggregators to identify who is publishing on a topic and **go direct to the publisher**. Individual publishers do not always give their full catalogue to the aggregators, may embargo their information for weeks or months, and may have more up to date information on their web site. You can also sometimes get a better deal by going direct to the publisher.
 10. **Dates**. Compared with structured databases, proper and accurate date searching is almost impossible with Google et al. A web page is assigned a date by the web server when it is loaded or reloaded onto the web site. It is not when the information was gathered or written. The web server date is the one that the search engines look at when you use the date option in the advanced search. Neither should you automatically trust the date that so often appears at the bottom of a page. It may be accurate and reflect the date of the content, but pages can be set-up to incorporate the date the page was loaded or reloaded onto the site, the date when minor changes are made, or even today's date :- (If the date is not obvious from the content, contact the author.

Two additional general points were made in conclusion:

- it is important to build up your own personal collection of sites, relevant to your sector and applications, and that you have already quality assessed and trust

- errors and misleading information are not new and pre-date the Internet era. Nothing has changed in that mistakes and bias in the media - whatever form - are a fact of life. What has changed is that **everyone** now has the opportunity to become involved in creating and perpetuating myths and mis-information, which means that we have to wade through so much more rubbish and spend more time separating the gold from the dross.
-

Web exceeds 100 million 'web sites'

<http://news.netcraft.com/>

Just when you thought you had got on top of information overload, Netcraft releases their latest survey revealing that there are now more than 100 million domain/host names out there. This is an increase of 3.5 million over last month, and the accompanying graph suggests that there is no slow down in peeps wanting to set up their own domain on the Net. The report is confusing because the text refers to web sites but the graph is labelled Domains August 1995 - November 2006 and the relevant line is labelled host names. Domain names and web sites are NOT synonymous. A single web site can be set up so that several domain names point to it - until recently the UKeIG web site had 9 domain names. Also many domain names are bought but not used for an active web site.

A little more reassuring and manageable is the line on the graph that shows active domains, which is getting close to 51 million. Compare that with Netcraft's first survey in August 1995 which identified 18,957 hosts. No wonder we can't find anything on the net. There is too much information - my brain is beginning to hurt !

Browser Shares - October 2006

The UK eInformation Group (UKeIG) have blogged their usually monthly analysis of which browsers people are using to access the UKeIG web site and blog. The figures are based on 34,294 page views in October 2006 and it comes as no surprise that IE 6 is still in the lead but with only 58% usage. Firefox 1.x comes in at number 2 with 24% and Firefox 2.x already has a share of 7%. So that makes Firefox's share a whopping 31%.

UKeIG regularly reports high Firefox percentages, which is not really that surprising. UKeIG actively promotes the Firefox via its factsheets, on the wiki and in its workshops, and many visitors to the site are likely to be tech savvy readers who have control over what software they can install or are in organisations that give freedom of choice re browsers.

Other browsers included Netscape 5.* (3%), IE7 (2%), IE 5 (2%), Konqueror, Mozilla 1, Opera 9 and Safari 2.

Online Information 2006

Online Information 2006 starts on November 28th and runs until Thursday 30th November, at Olympia in London. There is still time to book a place at the conference and although the early-bird discount no longer applies you may be eligible for a 15% discount if you belong to one of the 'partner' organisations or associations such as CILIP or UKeIG. Further details are on the Online Information web site at <http://www.online-information.co.uk/ol06/conference.html>

If you can't make it to the conference, there are free seminars and master classes taking place in the exhibition hall. Full programme and details are at <http://www.online-information.co.uk/ol06/seminars.html>

Conference and seminar presentations available

Many of the presentations made at Internet Librarian International 2006, London, are now available on the ILI web site (<http://www.internet-librarian.com/presentations.shtml>). If you want to 'flesh out the bones', try and track down blog postings on the event. Tom Roper's blog (<http://tomroper.typepad.com/>) is a good place to start. He has neatly summarised my presentation on out-Googleing Google (http://tomroper.typepad.com/tr/2006/10/internet_librar_12.html), so worth reading it along side my PowerPoint slides which are available on the ILI webs site or at <http://www.rba.co.uk/ili/index.html> For other blogger accounts try Technorati (<http://www.technorati.com/>) and use the tag ili2006.

If you are interested in seeing photos of the event go to Flickr (<http://www.flickr.com/>) and use the tag ili2006.

This was yet another great event from Infoday with plenty of good practical tips and case studies. This year there must have been at least 350 people there and the key note sessions were packed solid with standing room only for late arrivals! The rest of the program was split into parallel tracks so you could **almost** guarantee a seat. At the rate this conference is expanding, next year ILI will have to take over the entire hotel :-). If you have never been, watch the web and feeds for news of next year's event.

CILIP's East of England Information Services Group put on an excellent day covering the whole range of '2G' internet services. The Let's Get Wiki'd event included Library 2.0 and Web 2.0 (Dave Pattern from the University of Huddersfield), Mashups (Richard Wallis, Talis), Social Bookmarking and Connotea (Joanna Scott, Nature), and wikis and the semantic web (Nick Kings, BT). I finished off the day with a session on assessing the quality of collaboratively collected information. The presentations will soon be made available via the ISG web site (<http://www.cilip.org.uk/specialinterestgroups/bysubject/information/services/regions/ee>). In the meantime, mine is at <http://www.rba.co.uk/isgeoe/index.html>

Information Resources

Blogs and RSS for competitive intelligence

http://www.digimind.com/en/download/White_Paper_Blogs-RSS_EN_2006.pdf

There is increasing awareness that blogs and RSS are an important part of competitive intelligence. This document, published by Digimind (<http://www.digimind.com/>), is an excellent overview and introduction to how they can be used in competitive intelligence, and includes an extensive list of references and further reading. Highly recommended if you are wondering what all the fuss is about and whether or not you should leap onto the bandwagon.

Another 'must read' paper is 'RSS: the CI Professional's Best Friend' by August Jackson, Evidence Based Research. It originally appeared in Competitive Intelligence Magazine, Vol 9 Number1, Jan-Feb 2006, pages 23- 27. A PDF of the article can be found at <http://homepage.mac.com/cornfed/RSSarticle.pdf> . August also a podcast devoted to CI at Competitive Intelligence Podcast <http://www.cipodcast.com/>

Passport launched for OurProperty and PetrolPrices.com

Fubra, who run and maintain the OurProperty (<http://www.ourproperty.co.uk/>) and PetrolPrices.com (<http://www.petrolprices.com/>) web sites, have launched the Fubra Passport. This is a single login system that will be valid for all existing and future Fubra web sites and consists of your email address and just one password. You no longer need your user name.

Fubra say that the benefits will become obvious as they launch more sites over the next few months, but no clues from them yet as to what those sites are likely to be. So far, I have been very impressed with Fubra's sites. I use regularly use OurProperty, which repackages Land Registry

data, and friends and colleagues reckon that PetrolPrices, which gives details of local petrol prices, is excellent. (We gave up our car 15 years ago as an experiment and are still managing to survive and travel with out it!). Fubra also run Compare Airport Parking (<http://www.airport-parking-shop.co.uk/>).

Gizmo of the Month

Firefox add-on – Resurrect pages

<https://addons.mozilla.org/firefox/2570/>

This is the prefect Firefox add-on for the dreaded 404 message, or if you want to see which version of a page the search engines have in their cache. You can try five page cache/mirrors in turn: CoralCDN, Google Cache, Yahoo! Cache, The Internet Archive and the MSN/Live Cache. The search engine caches usually only have copies going back days or, at most, weeks but the Internet Archive may have copies going back to 1996. If you have installed Firefox 2 you may need to use the Nightly Tester Tool add-on to persuade Firefox that Resurrect Pages is compatible.

Meetings and Workshops

Online Information 2006

Karen Blakeman will be giving two free seminars as part of the seminar and master class programme at the Online Information exhibition:

Tricks and Tips for better web search. Tuesday 28th November 2006, 11.45-12.15. Grand Hall Olympia, London, Theatre E.

What's new in search tools. Thursday 30th November 2006, 11.00 - 11.30. Grand Hall Olympia, London, Theatre E.

Details of these and the rest of the seminar programme are at <http://www.online-information.co.uk/ol06/seminars.html>

Workshop: Advanced Internet Search Strategies

Organiser: Manchester Business School

Venue: Eddie Davies Library, Manchester Business School, University of Manchester, UK

Date: Tuesday, 5th December 2006

Course presenter: Karen Blakeman

Course fee: £250 + VAT or £215 + VAT for BIS/BINN members.

URL: <http://www.mbs.ac.uk/programmes/courses-seminars/management-research.htm>

Outline: With the major search engines claiming coverage of up to and over 20 billion web pages in their databases, it is becoming increasingly difficult to locate relevant information. Google is not the only search tool and the competition is hotting up. This workshop looks at recent developments at Google and the alternatives, especially the new kids on the block. Delegates will have ample opportunity to test out advanced search techniques and to compare different search engines. There will also be sessions on blogs and RSS feeds, and help on tracking down audio, video and podcasts of news broadcasts and interviews. A significant part of the day will be taken up with practical sessions.

Contact Information

Karen Blakeman, RBA Information Services
UK Tel: 0118 947 2256
Int. Tel: +44 118 947 2256

UK Fax: 020 8020 0253
Int. Fax: +44 20 8020 0253

Address: 88 Star Road, Caversham, Berks RG4 5BE, UK

Archives

TFTR archives: <http://www.rba.co.uk/tftr/archives/>

Subscribe and Unsubscribe

To subscribe to the newsletter fill in the online registration form at <http://www.rba.co.uk/tftr/>

To unsubscribe, use the registration form at <http://www.rba.co.uk/tftr/> and check the unsubscribe radio button.

If you wish to change your subscription from the PDF to the Plain Text version, send an email to tftr@rba.co.uk with "Plain Text Version Please" in the subject line.

Privacy Statement

Subscribers' details are used only to enable distribution of the newsletter **Tales from the Terminal Room**. The subscriber list is not used for any other purpose, nor will it be disclosed by RBA Information Services or made available in any form to any other individual, organisation or company.

This publication may be copied and distributed in its entirety. Individual sections may NOT be copied or distributed in any form without the prior agreement of the publisher.

Copyright (c) 2006 Karen Blakeman. All rights reserved
