

“When librarians rule the world: the skills and tools we need to thrive in a Web 2.0 world”

Presentation by Mary Ellen Bates, Bates Information Services, www.BatesInfo.com, held on 29th February 2008, Zentral- und Landesbibliothek Berlin

Compiled by Anke Bruns¹ and Susanne Schaper²

¹MPDL, Munich / GWDG, Göttingen, anke.bruns@mpdl.mpg.de

²MPI for terrestrial Microbiology, Marburg, schapers@mpi-marburg.mpg.de

As Mary Ellen Bates' presentation was only available on paper and not on the web, we have put together much of her information in this report, using her own structure and headings.

The URLs mentioned were functional on 7 April 2008.

Content:

“When librarians rule the world: the skills and tools we need to thrive in a Web 2.0 world”	1
I. Hidden features of the major search engines.....	1
A. Google's secrets	1
B. Yahoo's secrets	2
C. Live.com's Secrets	3
II. Specialized and lesser-known search engines	4
III. Creating your own search engine	5
IV. Searching Web 2.0	5
V. Becoming an Info Pro 2.0	7
VI. Adding value to our information	8
VII. Summary	9

I. Hidden features of the major search engines

Mary Ellen Bates explained how the well-known search engines and some of their less well-known features can be used to operate more sophisticated searches on the web:

A. Google's secrets

Google's experimental search

Adding `view:timeline` or `view:info` to the search query shows results in a timeline, with dates highlighted, with images featured oder highlighted on a map, e.g.:

<http://www.google.com/experimental/>

SearchMash

Unbranded Google site; results sorted by web page, images, blogs, videos, Wikipedia. Easy search-within-site feature, sorted by domain name (<http://www.searchmash.com/>)

Date search

`daterange:nnnnnnnnnnn`, must use Julian calendar (Converting tool : <http://aa.usno.navy.mil/data/docs/JulianDate.php>); never entirely reliable

Date limiting

"Advanced Search screen, => **pull down menu**: last 24 hours, last week.

More information on Google date limiting: <http://www.jobmachine.net/google-time-frame-searching>

More date limiting

Can do "last X days/weeks/years": Add `&as_qdr=XN` to the search results page URL (X=d, w or y, N=number of d, w or y to limit with), e.g.: `&as_qdr=w5` (the results to pages spidered in the last 5 weeks)

Timeline search on Google News Archive Search

"View timeline" button for viewing diagram showing how many articles on a search term there were at different times; can be used to show trends (<http://news.google.com/archivesearch>)

Search range of numbers with “.” operator added to search query

Syntax is number.number, .number or number.. (e.g. 525.., or ..100, 1999..2001 – no spaces!) Can be used for year ranges. Also used for price ranges for products, but recognizes only \$ for currency, not €, £ etc. so maybe not so useful for German market ;-)

Firefox fix for Google

some nice features for customization. (<http://customizegoogle.com>)

Simply Google

(http://www.usabilityviews.com/simple_google.htm): structured overview site of all Google's features and sites, downloads, blogs etc. Interesting not only for librarians but maybe also for their users.

B. Yahoo's secrets

Mindset feature

Search results can be modified with slide switch to determine whether intention is rather to shop or to research – not well known but very nice! Apparently this has been developed by a Yahoo employee who was himself annoyed about not being able to choose. (<http://mindset.research.yahoo.com>)

"Feature:"ing

looks for page with that feature, e.g. `feature:index` (just home pages), `feature:table` (pages with a link to an .xls file), `feature:image` (pages with images in them - but "not" an image search), `feature:audio`, `feature:video` (same principle as `feature:image`)

Regional limits

e.g. `region:europa` (Europe incl. UK), possible also: `africa`, `asia`, `centralamerica`, `downunder`, `mediterranean`, `mideast`, `northamerica`, `southamerica`, `southeastasia`. Based on top level domain (.de, .uk etc) so misses .com, .net, .org domains.

Customized shortcuts

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

`!wiki` => searches Wikipedia entry, e.g. `!wiki helmut schmidt`


`!audio` – audio search;

`!news` – news search

Create your own shortcuts (Users with a Yahoo account):

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

`!aiip` – the AIP directory => For users with a Yahoo account

 To use Open Shortcuts, please [sign in](#) first. For information on Open Shortcuts, please type `!help` in the Search Box.

Y!Q

(<http://yq.search.yahoo.com>): Contextual searching, big search box where entire paragraphs fit in. Suggests alternate phrases and "key concepts". Good if one has not yet a very clear concept of what to search. It finds rather "fuzzy matches" than exact results like a Boolean search. Works the better the more words are put into search box. Help page is said to be not very comprehensible.

SearchAssist

on Search Results page (SRP) look for an unobtrusive little arrow (see screenshot below); clicked on, it displays similar and related concepts, search terms to be "added" to the original search.



C. Live.com's Secrets

Some functions on Live.com Search by Microsoft (<http://search.live.com>) are hidden in the search results page:

Bug or feature?

Auto-correcting of typos: unlike Google, Live.com doesn't ask "Did you mean...?" but instead just goes ahead. Of course to many this might seem rather a bug than a feature. Can be avoided by enclosing the search term/phrase in quotation marks (").

"More"-Link

e.g. academic (Wissenschaftlich):



"Prefer" option

Add `prefer:word` to search, word being one of the search terms. Results containing word are ranked higher, but results aren't limited to these.

Synonym suggestion tool

Intended to help search engine marketers identify similar words, "Keyword Group detection". May be helpful to get some ideas what to search if one search term doesn't produce good results.

(<http://adlab.microsoft.com/contextSim/default.aspx>)

Limit to within the body

`inbody:search term` limits search to pages where word is only found within the body of the page, not in its title, metatags, anchor tags and so on.

Limit search to certain top level domains

Similar to Yahoo's "Regional limits", again, this misses all .com, .org etc.
(http://search.msn.com/docs/help.aspx?t=SEARCH_REF_Codes.htm)

Search for pages that have a link to a RSS or an ATOM feed

hasfeed:search-term

"Has a link to [file type]"

contains:file-type looks for pages that contain a link to a certain file type, e.g.
contains:doc or contains:pdf (as opposed to looking for .doc or .pdf pages).

Live Search Images

looks for: faces (filter:face), portraits (filter:portrait), black and white images (filter:bw). Not perfect but can be quite useful. E.g. the black and white filter if looking for historical photographs (<http://search.live.com/images>)

II. Specialized and lesser-known search engines

These search tools picked by Mary Ellen Bates for the workshop present some interesting approaches to searching, meta-searching, viewing and clustering results. Some of them are beta versions though. When checking them we found that not all of them work stably. Moreover, some of them obviously have only a relatively small coverage. Perhaps some will become better as they go along?

Specialized

- ❖ **OmniMedicalSearch** (<http://omnimedicalsearch.com>)
Different "flavours" of medical search – e.g. for patients (WEB), for medical professionals (MEDPRO), health-related news, images. Metasearch engine of authoritative search engines. Also medical portal: Acronym finder, conditions & diseases. List of medical search engines:
<http://omnimedicalsearch.com/databases.html>.

Search interfaces

- ❖ **Gigablast** (<http://gigablast.com>)
Search engine (interesting search possibilities: ranking tools, see help file for syntax) plus web directory
- ❖ **Exalead** (<http://www.exalead.de/>)
Search for "spelllike" or "soundlike" if one doesn't know how to spell search term. Near-Operator (e.g. for finding quotations), search for regular expressions
- ❖ **Intelways** (<http://intelways.com>)
Easy comparison of search results from many different finding tools by clicking on them. Useful if you don't know where to start. No merging of results as in metasearch engines.
- ❖ **SRCHR:** (<http://SRCHR.com>)
web page can be configured with different search tools, sites to search can be added. Creates RSS feeds for search results, mouseover for thumbnails...

Presentation of results

- ❖ **Factbites** (<http://factbites.com>)
shows search terms on resulting web pages in their context by presenting full sentences containing search term. Unfortunately no phrase searching or advanced search is possible. Shows only up to 30 results.

Clustering

- ❖ **Clusty** (<http://clusty.com>)
Metasearch engine by vivisimo (<http://vivisimo.com>); clustering of results; "Sources" tab (SRP, on the

left) shows which search engines the results come from . Build a topic cloud for your own website:
<http://cloud.clusty.com>.

❖ **Carrot2** (<http://carrot2.org>)

Open Source search results clustering engine, choose between different clustering algorithms and different search engines (PubMed being one of them). Result display with tabs for topics, sources and sites – bit like Clusty.

Web directories

❖ **Library internet information** (<http://www.lii.org>)

Directory of internet sites manually selected by librarians

❖ **Open directory project** (<http://dmoz.de/>)

Another human-edited directory, maintained by a community of volunteers.

To find yet more search engines: (<http://altsearchengines.com>)

III. Creating your own search engine

In this part of the workshop we got an idea of how one can configure one's own search engine meeting special needs and preferences. Thus for example a search engine for a library's website may be created (perhaps this is not so important for MPS which has its own Google Search Appliance (<http://find.mpg.de/>) that can be used and customized by MPIs).

There are quite a few of these services available for free. Keep in mind, though, that nothing is really for free – whatever uses the providers may get out of their “free” offer some day or other...

❖ **What's a custom search engine?**

Users' search queries are filtered according to previously defined characteristics and/or limited to a choice of domains or sites. Relevance ranking may be tweaked.

❖ **Examples of Customizable search engines are:** Google Custom search engine (<http://www.google.com/coop/cse/>), Yahoo Search Builder (<http://builder.search.yahoo.com>), Swicki (<http://www.eurekster.com/>), Rollyo (<http://rollyo.com/>), Live.com (<http://search.live.com/macros>)

IV. Searching Web 2.0

Here Bates showed us yet more search tools to be discovered, these being especially designed to search wikis, blogs, social networks and so on:

❖ **WikiIndex** (<http://wikiindex.com>)

A wiki presenting other wikis (hierarchical directory). Best way to find relevant wikis seems to go through categories.

❖ **Wikiseek** (<http://wikiseek.com>)

Searches Wikipedia pages and sites referenced within Wikipedia. Its index, however, seems not really up to date, as it contains versions of pages older than 25 december 2007. Also searches only the English Wikipedia.

❖ **Wikia** (<http://wikia.com>)

Jimmy Wales's new project, a platform to create one's own wiki for free. Plus an open source search engine (<http://alpha.search.wikia.com>) whose quality of search results admittedly is “low” at the moment. People are supposed to contribute feedbacks on the pages found by Wikia search, thus improving the quality.

- ❖ **Technorati** (<http://technorati.com>)
One of the best-known blog search tools so far. Can search for: blogs about X, blog posts tagged with X, blog posts that mention X, videos tagged with X. Filter by “authority”: How many other blogs point to this blog. Advanced search: (<http://technorati.com/search?advanced>).
- ❖ **Google Blog Search** (<http://blogsearch.google.com>)
Searches anything with a feed. Indexes *only* feed content. Use inpostauthor: to find what someone’s saying, to judge authority. Advanced search: http://blogsearch.google.com/blogsearch/advanced_blog_search.
- ❖ **Serph.com** (<http://serph.com>)
Searches blog search engines and online social media, e.g. Technorati, Google Blog Saearch, Flickr, Newsvine, YouTube, del.icio.us, Yahoo Answers, Bloglines.
- ❖ **50Matches** (<http://50matches.com>)
“Social powered search” – crawls web sites that were bookmarked in social media like *del.icio.us*, *digg*, *reddit*. Maximum of 50 results: good for limiting to ‘top’ sites.
- ❖ **Yahoo Pipes** (<http://pipes.yahoo.com>)
Tool to “aggregate, manipulate, and mashup content from around the web”. Users can create their own pipes by combining feeds into one feed, sorting, filtering them etc. Examples of pipes: Library pipe (<http://pipes.yahoo.com/pipes/pipe.info?id=qsFtxSZm3BGW7mvB8ivLAg>) and pipe on library technology: <http://pipes.yahoo.com/pipes/pipe.info?id=vgu00tRv3BGq3okECB2yXQ>.
- ❖ **SearchTheTail.com** (<http://searchthetail.com>)
In Google search engine clusters results not belonging to the „Top 100“ using obscure, less popular terms (=“the long tail“). Doesn’t like phrase searches with “”
- ❖ **Track Discussion Forums**
Boardtracker (<http://boardtracker.com>), Boardreader (<http://boardreader.com>)
Search forums and message boards. Useful when looking for information on very specialized topics and “weirdo stuff”
- ❖ **Search podcasts**
PodcastAlley (<http://podcastalley.com>), <http://pluggd.tv>
- ❖ **Podcast Chunking**
Podscope (<http://podscope.com>), Everyzing (<http://everyzing.com>): Not only metadata but search within podcasts (speech recognition!); searches transcripts of podcasts.
- ❖ **Social Network Search**
LinkedIn (<http://linkedin.com>), YoName (<http://yosome.com>) – Look for people (search by name, company, position...) who are members of a social network like MySpace, LinkedIn, YouTube, Twitter etc. In order to see all information usually you need an account yourself – give away information in exchange if you want to use others’ information – fair enough but remember you’re making yourself known not only to people around the world but also to the provider’s data base.
- ❖ **NationMaster** (<http://nationmaster.com>)
Gets data by data mining from WHO, World Bank, CIA World Factbook, World Resources Institute, UNESCO, OECD and other sources and uses them to build up statistics and correlations. **Very good** for getting first impressions on political topics on an international scale.
- ❖ **Authoratory** (<http://authoratory.com>)
A search engine for finding experts, the content of which is produced by data mining from PubMed. Useful for finding information on scholarly authors. Authoratory community membership is free for members of educational or government organisations. Only with personal login full information on individual authors is accessible. Ranking of authors (by no. of articles, highest funding or “top

keywords”) can be seen without login.

- ❖ **TouchGraph.com** (<http://touchgraph.com>)
Data visualisation tool, finds relationships among URLs by using Google’s “Similar pages” function, finds also related books in Amazon and connections in Facebook. Have to get used to visualisation, as inexperienced users we couldn’t yet make much of it.
- ❖ **Snap Search** (<http://www.snap.com>)
Visual search results, snapshot of live page can be seen in the results. Based on Gigablast. Firefox plugin available to add preview option to other search tools like Google, Amazon, to blogs, news sites and so on.
- ❖ **Quintura** (<http://quintura.com>)
On-the-fly clustering of results, based on Yahoo’s search engine. Mousing over a term re-executes results. Tabs for results in video, images and Amazon.

V. Becoming an Info Pro 2.0

In the next two parts Bates explains her concept of how information professionals (Info Pros) can use web 2.0 features for the benefit of both their clients and themselves. To become an Info Pro 2.0, according to her you should use interactive and collaborative tools and features where users contribute content and transparency is the key – in short: become an active user of Web 2.0.

- ❖ **Wiki that!**
 - Use internal wikis to team with groups within your organisation,
 - Build wiki of industry and internal acronyms
 - Create a virtual reference desk (which highlights your activities)
 - Share resources among Info Pros
- ❖ **Build RSS feeds**
develop customized RSS feeds for clients (e.g. with Yahoo Pipes (see Chapter IV)) – to monitor community blogs, wikis &c.
- ❖ **Online Catalogs 2.0**
 - Interactive navigation through OPAC or other information collection: data visualisation tools, track page views.
 - Allow tagging of OPAC records: Identify new taxonomy terms, identify useful content, navigation via tag cloud
 - see for inspiration e.g. LibraryThing (<http://www.librarything.com/>, <http://www.librarything.de/>): Collaborative tool to catalogue, review, suggest and share information about books
- ❖ **Blogs for everyone**
Lead blog initiative within your organisation. E.g.: <http://blogs.sun.com> – SUN Corp. sets up blogs for all employees. Separate internal and external blogs, the public can identify experts, code of behaviour for blog posts
- ❖ **For your mobile clients**
Consider Instant Messages reference service, delivery of information to mobile devices, post deliverables on server in mobile-friendly format. (S.a. http://en.wikipedia.org/wiki/Digital_reference_services)
- ❖ **Expanding internal knowledge**
Wikis, internal blogs and podcasts, surface “hidden” experts within organisation, build an internal podcasting lab, employees do brain dump of workshop, conference, meeting...
- ❖ **Life-long learning (be proactive rather than (re-)active)**
Create portal to (course) podcasts (Examples: Univ. of Bath, <http://www.bath.ac.uk/podcast>), Harvard:

<http://athome.harvard.edu>, Google's series of author interview podcasts on Youtube:
<http://youtube.com/atgoogletalks>).

❖ **Stay in touch**

Do podcasts of: local events, weekly news your users can use, presentations given in the library or within your organisation, summaries of conferences others have attended.

VI. Adding value to our information

Mary Ellen Bates explains that it is not enough just to retrieve information on the web and send it on to the information professional's clients. In order to be of real use for them, the Info Pro should edit and enrich the information retrieved according to his or her clients' specific information needs and present it in a user-friendly form:

❖ **Beyond "Rip and Ship"**

- Before delivering information, answer at the minimum (in your head) the following questions:

* Why was an information resource used?

* What is interesting about it?

- Write a cover memo

- Highlight the good stuff

- Create table of contents

- Surface your sources

- Less is more!

- "Brand" your results (use a single template/layout)

❖ **Write article summaries (Provide information about information)**

to help your clients identify interesting resources more quickly. Extract the key paragraph and bibliographic citation; include hyperlink from summary to article.

❖ **Distill your results**

... e.g. in a table

❖ **Create graphics**

... to visualise information for your clients, e.g. using a table calculation program like excel (pull down insert menu, select chart; select format and label options)

❖ **Use fee-based services' feature (and put it in a nice format)**

- Dialog: DialogLink

- DataStar: ReportAid, WebCharts

- Factiva: Newsletter

- LexisNexis: Publisher

❖ **Data mining**

- Use the special output features of the online services; Dialog has richest menu: RANK, SORT, REPORT, user-defined output.

- Use Dialog's RANK to identify trends, experts

- Use Factiva's "Discovery Pane" (e.g. for statistics)

- LexisNexis does clusters

❖ **Tips and strategies**

- Look for "information-dense" sources – charts, graphs, tables, analysis.

- Look for sources your patron doesn't have (or know about) – telephone research, public records, invisible web content, podcasts, wikis, encyclopedias

- Use BlogPulse's Trendsearch (<http://www.blogpulse.com/trend>) to demonstrate change in blogging over time, e.g. on "Global Warming"

- Google Trends (<http://www.google.com/trends>) watches search word trends over time, also maps news articles on those topics, explains WHY there are spikes

- <http://furl.net> lets you generate annotated webliographies, creates also snapshots of a webpage at a given moment – so the information is kept even though web page may be changed.

❖ **Technorati “authority”**

Shows ranking according to how many blogs link to a website assuming that the more blogs link to it the more “authority” it has; see also: <http://technorati.com/weblog/2007/05/354.html>

❖ **Furl’d bibliography**

see: (<http://www.furl.net/>)

❖ **In your deliverable**

Include a three-question evaluation form:

- Did we answer your information need?
- Is there any additional information you need?
- If you could change one thing about the library, what would it be?

❖ **More branding**

- Develop a single mail signature for all library staff which you can update monthly to promote a different product/service
- Put your Unique Selling Proposition in your tag line, e.g.: “Answering impossible questions since 1998”

❖ **What is your unique value?**

Bates encouraged us to think about (and tell people) what qualities we as Info Pros bring to our organisation, defining our job in a way to make clear we’re doing something special: Instead of saying: ‘I’m a stone cutter’ or ‘I’m doing the best job of stone masonry in the country’, we should say: ‘I am building a cathedral’!

So: **Don’t let the urgent have priority over the important!**

VII. Summary

Mary Ellen Bates, in her enthusiasm for online information services and tools, does not mention one important drawback of many of the services mentioned, namely that the more you use collaborative tools on the web, the more you expose (more or less personal) information about yourself and your habits not only to other people around the world but also to the data bases of the service providers without ever knowing what they are going to do with your data (e.g. sell them to advertising companies,...or worse!). This is of course only true for services “out there” and not for a wiki or a blog installed locally in your institution.

A huge variety of interesting search tools was presented. Among our favorites are Nationmaster, Authoratory... There are many hints as to how search interfaces and result sets can be made more effectively and user-friendly. So we do think that these tips may be useful for any librarian within the MPS.

Not especially Web 2.0 related but yet useful were her remarks on the information professional’s work in general, probably one can’t be reminded often enough of these tips...

Acknowledgement

The critical reading by Urte Brauckmann, MPI for the History of Science is gratefully acknowledged.