



Internet applications, sites, trends and happenings

David Raitt
david.raitt@esa.int

This column aims to draw your attention to various interesting Web sites which I have come across and which might appeal to you, and to keep you up to date with news and views on Internet trends, developments and statistics. It offers essentially a personal selection rather than comprehensive coverage.

Blogs in space

According to an item in Space Daily (<http://www.spacedaily.com/news/internet-05zrz.html>), MindComet has launched a Web site sending bloggers where no blog has gone before: deep space. BlogInSpace.com is a free service for bloggers allowing them to submit their blog feeds for transmission into space. The site will aggregate blog content into transmission packages and beam the content up, Scotty, via a powerful earth-based satellite broadcast. Bloggers are thus being given a unique opportunity to send a piece of their lives into space to potentially connect with extraterrestrials in a new, meaningful way!

A previous MindComet service, BlogStar Network, was launched earlier this year and met with incredible interest from bloggers and advertisers alike. MindComet is expecting similar results with the launch of BlogInSpace. Fasten your seat belts and take the trip to BlogInSpace.com for free registration and more information.

Have Google will travel and see the world

In my last column I talked about Google Earth and the images you could see. Now, I have come across Google GlobeTrotting (a site not affiliated with Google) but which contains imagery retrieved via Google Earth. The site allows users to post satellite images of pictures of interesting or unusual buildings, landscapes, artworks, events, transportation – the list of categories is endless. Want to see some murder sites? How about a helicopter departing Camp Babylon? Or a concert taking place in Golden Gate Park? Or maybe the Allied Domecq Quick Service Restaurants Headquarters which is (or was, perhaps) under construction at the time the satellite photo was taken. Always wanted a bird's eye view of Muizenberg Pavillion or Strandfontein Sewage Works? Then look no further – there are currently over half a million map views. The site also has a tour facility, discussion forum and full details on how to capture an image from Google Maps, create a thumbnail and post to the Web site.

Check it out at <http://www.googleglobetrotting.com/>.

Future of search

I attended the World-Wide Web conference in Cape Town a few weeks ago and, on the plane on the way back, I was reading the 5 September 2005 issue of *Time*. There was an interesting piece on the frontier of search (p. 28–30). The article discussed some of the new features coming our way – from 3-D pictures of your house to news tailored to your 'clickstream'. The problem is simply that there are an estimated 10 billion pages of text on the Web (and that does not include the invisible ones) and Americans alone conducted nearly five billion searches in the month of July – an increase of 22% over 2004. 'But', the magazine asks, 'who needs 14.120.000 results in response to a simple question?' People do not want a list – they want an answer. [Rhetorical question for those old enough to remember: Did we not have the same kind of problem way back in the mid-sixties when we were starting out with all these online computer systems, search languages (not engines) and databases?]

By 2010, search engine advertising will be a \$22 billion industry worldwide, apparently, and so the battle is on for the next generation of search – smarter, more tailored to individual needs, embracing video and music and being accessible from any device with a chip. Google has 37% of all queries, while Yahoo has 31%, with MSN at 16%. But all are aggressively pursuing new ideas – Google Earth or Google Talk for instance, or Yahoo! muscling in on Google's announcement to scan books in libraries. But others are pushing for a slice of the pie too.

Singingfish (<http://search.singingfish.com/sfw/about.html>) can search AOL's video library of 15000 titles, plus millions more over the Web, by looking for their titles and metadata. The mobile phone service 4INFO gives local information on theatres, restaurants, news headlines etc. by text or the Internet on your cell phone. Yahoo! does the same. Mobot has developed technology that maps features in a picture taken with a cell phone camera and matches it to a database of images. Yahoo! has 61% of the mobile Web market, with 15 products, including search, and has developed a technique that simplifies Web pages for small mobile screens. Blinkx.TV can also track down video clips in a few seconds because it is tracking, listening to and watching over 15000 podcasts and videoblogs so users can find them easily (<http://www.blinkx.tv/>).

Text searches are all very well – but the next step is semantic searching – looking for meaning, not just matching key words. At the University of Washington, language analysis programs are used to power KnowItAll which scans documents for facts. The domain-independent system was developed to address the problem of accumulating large collections of facts. It treats the Web as a large corpus of text and extracts massive amounts of information from it in an autonomous, scalable manner (<http://www.cs.washington.edu/research/knowitall/>).

Another fast growing search technique is tagging whereby users label Web sites with descriptive tags, thus building a taxonomy or network of knowledge. Quick to spot the trend, Yahoo! recently bought Flickr, a photo Web site organized with a communal tagging model (<http://www.flickr.com/>). Flickr allows users to add tags (keywords or labels) to a photo to make it easier to find later. The possibility is also there to allow others to add tags to your photos. The site gives advice on attaching tags and also maintains a list of the most tags and those added within the past 24 hours.

Not to be outdone, though, by everyone else, especially Google Earth, Yahoo! and A9.com (which sent trucks round 22 US cities with digital cameras linked to laptops to photograph every street – with 35 million pictures so far overlaid on maps), Microsoft is combining approaches from the air with its Virtual Earth project. Flying planes over cities, the aim is to have views from all directions so users can circle buildings on-screen – a bit like a virtual reality experience! What the system does in effect is to combine mapping and local search

to give the answers to search questions in a geographical context (<http://virtualearth.msn.com/>).

Yahoo! to scan books too

Some nine months ago, Google announced a programme to create searchable digital copies of entire collections at leading research libraries. Not only has this plan run into a series of challenges, but an unusual alliance of corporations, non-profit groups and universities, led by Yahoo! has just announced an ambitious plan to digitize hundreds of thousands of books over the next several years and put them on the Internet, with the full text accessible to anyone, according to an article in the *New York Times* (<http://www.nytimes.com/2005/10/03/business/03yahoo.html?ex=1129003200&en=bd0f025f254fac4d&ei=5070&emc=eta1>).

The new project, called the Open Content Alliance, has the wide-ranging goal of digitizing historical works of fiction along with specialized technical papers. In addition to Yahoo, its members include the Internet Archive, the University of California and the University of Toronto, as well as the National Archive in England. The digitization of print materials has been a continual effort on the part of various research libraries for the last several years, but the potential power of the new collaboration lies in the collective ability of many institutions to compare and cross-reference materials.

The announcement ups the stakes in the battle between Yahoo! and Google over index size – the number of documents that can be found in a search engine's database. However, the new project's approach differs from Google's in several ways. Once a book has been digitized, Yahoo will integrate the content into its index and provide an engine for the group's Web site (<http://www.opencontentalliance.org>). In a departure from Google's approach, the Open Content Alliance will also make the books accessible to any search engine, including Google's. Under Google's programme, a digitized book would show up only through a Google search. And by focusing at first on works that are in the public domain – such as thousands of volumes of early American fiction – the group is sidestepping the tricky question of copyright violation.

A few weeks ago, a group that includes the Authors Guild, which represents several thousand writers, filed a lawsuit against Google. The suit contends that the company's digitization programme, Google Print, is engaged in copyright infringement because although only text fragments are displayed, a book must be digitized in its entirety to make it searchable. In August, Google put on hold its plan to scan copyrighted books, to give authors and other copyright holders the opportunity to opt out of the programme. But see another article in the *New York Times* which sympathises with Google (<http://www.nytimes.com/2005/09/28/opinion/28oreilly.html?ex=1129003200&en=d71c5b314ff6f197&ei=5070&emc=eta1>).

Yahoo! did not disclose the overall budget for the project, although its own contribution has been estimated at between \$300 000 and \$500 000 for the first year. Hewlett-Packard and Adobe Systems are contributing equipment to the project, and the Internet Archive will do the actual digitizing and archiving of the books. The Internet Archive has set up shop at the University of Toronto and has scanned some 2 000 volumes at a cost of about 10 cents a page.

About the author

Dr David Raitt is senior technology transfer officer with the European Space Agency in the Netherlands. His work involves finding applications for space technologies in non-space sectors, particularly those useful for improving everyday life. An information scientist by

education and training, David is also editor of *The Electronic Library* and chairman of the Internet Librarian International conferences.

Disclaimer

Articles published in SAJIM are the opinions of the authors and do not necessarily reflect the opinion of the Editor, Board, Publisher, Webmaster or the Rand Afrikaans University. The user hereby waives any claim he/she/they may have or acquire against the publisher, its suppliers, licensees and sub licensees and indemnifies all said persons from any claims, lawsuits, proceedings, costs, special, incidental, consequential or indirect damages, including damages for loss of profits, loss of business or downtime arising out of or relating to the user's use of the Website.



ISSN 1560-683X

Published by [InterWord Communications](#) for Department of Information and Knowledge Management,
University of Johannesburg