South African Journal
of Information Management
[DoE accredited]

Internet News Vol.7(1) March 2005

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.

Imaging the news

10x10 is a futuristic Flash-based news interface that presents an ever-changing grid of pictures in the news, each linked to a set of news stories linked by a common topic. Scanning such a grid of pictures – a visual representation of news activity around the world – can be more intuitive than reading headlines.

Every hour, 10x10 scans the RSS feeds of several leading international news sources, such as BBC World, Reuters and the New York Times, and performs an elaborate process of weighted linguistic analysis on the text contained in their top news stories. After this process, conclusions are automatically drawn about the hour's most important words. The top 100 words are chosen, along with 100 corresponding images, culled from the source news stories. At the end of each day, month, and year, 10x10 looks back through its archives to conclude the top 100 words for the given time period. In this way, a constantly evolving record of the world is formed, based on prominent world events, without any human input.

10x10 is designed to be simple, intuitive and easy to use. When 10x10 is opened, a grid of the top 100 world images for that hour are presented, ranked in order of importance, reading left to right, top to bottom. Along the right edge of the screen are listed the corresponding top 100 words, one for each image. Moving the mouse around the images reveals which words match which images and moving the mouse up and down the word list highlights the corresponding images. Click any word or image to zoom in and see the news headlines behind the word. Click the headline links to read the original news stories. Click the zoomed image a second time to see the image full screen. Users can move through adjacent hours, and also browse through past hours, days, months and years. Find out more and use 10x10 yourself at http://www.tenbyten.org/10x10.html.

Searching the content

If 10x10 is different, how about this? A San Francisco-based start-up is rethinking the way search results are retrieved and organized. First launched in July 2004, blinkx takes a different approach to searching – rather than using keywords, blinkx returns results as a user is viewing a Web page, working on a document or reading an e-mail. Utilizing blinkx, users

are no longer limited to Boolean keyword search. Instead, blinkx automatically and implicitly conducts searches based on the content being viewed by each individual user whether it is on the Web, in the news or on the desktop. In the newly released Version 2.0 of its search tool, Blinkx Inc. introduced a feature called 'smart folders', where users can automatically populate a folder with Web and local content (e.g. pdfs, PowerPoint files, Excel and Word documents) based on the context of the documents in the folder or a description of the folder's purpose.

Blinkx, available as a free download from http://www.blinkx.com, installs a mini tool bar across Windows applications and Web browsers and an application for pairing hard-drive searches of e-mails and documents with Web, news, Weblog, shopping and video search. The tool bar includes six icons, where search results are continuously returned based on the context of what is on a user's screen. With its introduction of smart folders, blinkx has added another way to access search results. Users can set up smart folders either by adding documents and files about a given subject into the folder or entering a string of text in its setup. By analysing the text of the files, the smart folders then create shortcuts to relevant content.

Make your own concept map

The Institute for the Interdisciplinary Study of Human & Machine Cognition (IHMC) was established in 1990 as an interdisciplinary research unit of the University of West Florida. Since that time, IHMC has investigated a broad range of topics related to understanding cognition in both humans and machines with a particular emphasis on building computational tools to leverage and amplify human cognitive and perceptual capacities. One such tool enables the building up of concept maps – which are similar to mental maps. The CmapTools program empowers users to construct, navigate, share and criticize knowledge models represented as concept maps. It allows users to construct Cmaps on their personal computers, share them on servers (CmapServers) anywhere on the Internet, link their Cmaps to other Cmaps on servers, automatically create Web pages of their concept maps on servers, edit their maps synchronously (at the same time) with other users on the Internet, and search the Web for information relevant to a concept map.

The CmapTools client is free for use by anybody, whether its use is commercial or non-commercial. In particular, schools and universities are encouraged to download it and install it in as many computers as desired, and students and teachers may make copies of it and install it at home. (Commercial companies that install their own CmapServer do need to get a separate license for a CmapTools client that will talk to the commercial version of the CmapServer.) More information, as well as concept maps at http://cmap.ihmc.us/.

Exploring the modern world

Two Web sites bring the user closer to scientific achievements and discoveries. An ambitious Web-based pilot project called *Origins* explores the origins of matter, the universe, earth and even life itself. The goal was to use the tools of interactive media, such as live video Webcasts and interactive Web sites, to illuminate the human endeavour behind scientific research. Other goals were to use a museum setting to cover contemporary science, and to reach an audience of millions. The project was organized around virtual field trips to eight scientific observatories where significant, ongoing research is conducted at locations that are generally off-limits to the public, including a rain forest research station in Belize, underneath Antarctica’s sea ice, and an underground particle accelerator. Using the Web, the project enabled the audience to look over the shoulders of scientists at work in laboratories and in the field.

For each location, a content-rich Web site was created with interactive elements, video clips, articles and images, providing context for a series of live Webcasts. Each Web site included pieces on related topics – from particle physics to penguin research – and introduced the ideas that drive research, the tools that make discoveries possible and the scientists and support staff that conduct the work. The observatories covered include CERN in Switzerland, Cold Spring Harbor in New York, the Space Telescope Science Institute in Washington, Las Cuevas Research Station in Belize and the Natural History Museum in London, a radio telescope in Aricebo, Puerto Rico, and Antarctica. Discover the origins for yourself at http://www.exploratorium.edu/origins/ and then back up to the main Exploratorium page for more rich content.

Another site, Making the Modern World, brings you powerful stories about science and invention from the eighteenth century to today. It explains the development and the global spread of modern industrial society and its effects on all our lives. The site expands upon the permanent landmark gallery at the Science Museum, using the Web and dynamic multimedia techniques to go far beyond what a static exhibition can do. Making the Modern World provides 82 animated scenes that explore historical episodes, explain landmark discoveries and take you to special places like the Apollo cockpit and James Watt's workshop. Twenty-five story timelines run from the scientific Enlightenment of the eighteenth century through the age of coal and steam and brings us right up to today's world of science and scepticism. From the Rocket locomotive to DNA, this virtual exhibition displays over 116 amazing 'firsts' in the Science Museum that mark some of the crucial steps in science and invention. Nineteen learning modules link closely to the stories and are aimed at supporting students in their studies and offering a deeper understanding of a wide variety of subjects to all interested users. Eight dynamic guided tours help you follow stories and paths linked across time, while users can also e xplore over 400 everyday life objects used at home, work and play. Make the most of the modern world at http://www.makingthemodernworld.org.uk/.

Google's contribution to learning

As the two sites above show, much of scholarly research is learning what others have discovered and building on it. Google is also trying to help in this respect and has recently launched the beta version of a service that they hope will help the knowledge acquisition and use process. Google Scholar (http://scholar.google.com/) is a free service that helps users search scholarly literature such as peer-reviewed papers, theses, books, preprints, abstracts and technical reports from all broad areas of research. Google Scholar can be used to find articles from a wide variety of academic publishers, professional societies, preprint repositories and universities, as well as scholarly articles available across the Web. Just as with Google Web Search, Google Scholar orders search results by how relevant they are to the query, so the most useful references should appear at the top of the page. This relevance ranking takes into account the full text of each article as well as the article's author, the publication in which the article appeared and how often it has been cited in scholarly literature. Google Scholar also automatically analyses and extracts citations and presents them as separate results, even if the documents they refer to are not online. This means that search results may include citations of older works and seminal articles that appear only in books or other offline publications.

But not content with that, Google has decided to try and put as much human knowledge online as possible by digitizing the collections of revered libraries so that every Google user can search them instantly. Google Print was launched in October 2004 to make the gigantic world of books not only more discoverable, but also more searchable. Many books in the world are out of print and by working with libraries as well as publishers, Google will provide access to millions of books – many of them unique. Google is beginning by

scanning book collections belonging to the Universities of Michigan, Harvard, Stanford and Oxford, as well as the New York Public Library, so that they become more searchable. But it is also calling on publishers large and small to participate in the Google Print programme by adding their books to Google's search results. Advantages cited include increasing a book's visibility at no cost, attracting new readers and boosting book sales worldwide, driving qualified traffic to the publisher's Web site, and earning new revenue from Google contextual ads.

Google will scan books sent to them by authors, publishers and libraries and add them to their index. Clicking on one of a publisher's titles in the Google search results will lead users to the page from the book on which the search terms appear, as well as publisher information and links leading to online booksellers. Publisher content is secure and users will only be able to view a limited portion of book. Google Print lets you look at whole pages for free books and excerpts from copyrighted works. Find out more at https://print.google.com/publisher/.

How common can you get?

You are probably all familiar with the Word Count tool in Microsoft Word which tells you the number of pages, words and characters in a document. Well, taking this concept (and name) to new levels is WordCount – an artistic experiment in the way we use language. WordCount presents the 86800 most frequently used English words, ranked in order of commonness. Each word is scaled to reflect its frequency relative to the words that precede and follow it, giving a visual barometer of relevance. The larger the word, the more we use it. The smaller the word, the more uncommon it is.

WordCount data currently come from the British National Corpus (BNC), a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent an accurate cross-section of current English usage. WordCount includes all words that occur at least twice in the BNC. In the future, WordCount will be modified to track word usage within any desired text, Web site and, eventually, the entire Internet.

WordCount was designed with a minimalist aesthetic, to let the information speak for itself. The goal is for the user to feel embedded in the language, sifting through words like an archaeologist through sand, awaiting the unexpected find – for instance, observing closely ranked words can tell us a great deal about our culture. Amuse yourself in your spare time by going to http://www.wordcount.org/main.php.

Getting spam? The why not attack the spammers?

It seemed like a nice idea to combat spammers – bombard spam Web sites with data to bump up their bandwidth bills. At the beginning of last December, Lycos Europe released a screensaver that bombarded spam Web sites with data to try to increase the cost of running such sites. Analysis showed that, in some cases, within days of release, spam Web sites were being completely overwhelmed by the traffic being directed their way. The Lycos plan also came under fire for encouraging vigilantism.

Lycos Europe's 'Make love not spam' campaign was intended as a way for users to fight back against the avalanche of junk mail messages coming their way. Participants were encouraged to download the Lycos screensaver which, when their PC was idle, would then send lots of data traffic to Web sites that peddle the goods and services mentioned in spam messages. Lycos said the idea was to get the spam sites running at 95% capacity and generate big bandwidth bills for the spammers behind the sites.

The screensaver was downloaded more than 100000 times in the week since it was launched. But monitoring firm Netcraft, which analysed response times for three of the sites, found that the campaign was being too successful. Two of the sites being bombarded by data were completely knocked offline. One other site was responding to requests only intermittently as it struggled to cope with the traffic the screensaver was pointing its way.

The downing of the sites could dent Lycos claims that what it was doing did not amount to a distributed denial of service attack (DDoS). In such attacks, thousands of computers bombard sites with data in an attempt to overwhelm them. Finnish anti-virus firm F-Secure advised against using the screensaver in case of legal problems. Currently laws in many countries do not explicitly outlaw DDoS attacks but many nations are re-drafting computer use laws to make them specific offences. Read the story at http://news.bbc.co.uk/1/hi/technology/4061375.stm.

Americans use the Web for travel

Some 23% of Americans, or 68 million unique visitors, flocked to online travel destinations during November 2004, making up 46% of all active Web surfers, according to Nielsen Netratings. In comparison, there were 60 million Web users who visited online travel sites in November 2003, growing 13% year-over-year. The volume of traffic and annual growth underscored the Internet's role in helping consumers make travel preparations for the holiday season. In November, MapQuest ranked number one in online travel destinations with 31 million unique visitors, while Expedia followed with 14 million visitors. Travelocity drew 11 million online surfers, while Orbitz.com and Southwest Airlines rounded out the top five online travel destination sites with 11 million and seven million unique visitors, respectively.

Online consumers spent almost $920 million on travel during November 2004 – an increase of 11% spent in November 2003. Much of the online travel spending stemmed from lower to upper middle class households. Those with household incomes of $50,000 to $74,999 comprised 28%, or the largest group of online travel visitors, during November 2004, while those with household incomes of $25,000 to $49,999 made up 20% of online travel visitors. Research shows that those living in household incomes of $75,000 or more tend to favour brand-specific travel Web sites over any other category of travel sites, while those earning $75,000 or less tend to visit multi-category travel sites such as Expedia and Travelocity where one can do comparison shopping. Read the complete Web travel analysis and comment at http://www.nielsen-netratings.com/pr/pr_041216.pdf.

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.

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 the Centre for Research in Web-based Applications, Rand Afrikaans University