

# Seek and ye shall find?



**Kerry Webb**

kerry.webb@alianet.alia.org.au

*This site provides a search facility over weblogs and RSS feeds, to track what people are writing and talking about...*

**W**illiam A Woods is a scientist with Sun Microsystems, and a researcher in natural language processing and knowledge representation. He has written a thoughtful paper summarising many of the issues in helping systems determine just what people are looking for. He comes to the conclusion that an understanding of the meaning of a particular document (or set of text) is fundamental aspect to the search for better retrieval methods. See it at <http://research.sun.com/projects/knowledge/WoodsACMQueueApril2004.pdf>.

## **Beyond Google**

We know that Google does not provide all the answers, and I suspect that many Google users are also coming to that conclusion. If you need some guidance in additional resources, here is an article from *PC World* that gives a few pointers to other sites and services that might provide the answer that you are looking for: <http://www.pcworld.com/resource/printable/article/0,aid,114725,00.asp>. Topics that are covered include technical terminology, problems with home appliances, lists of companies and (most importantly) recipes.

## **Now who was that?**

If you are a fan of good movies or bad movies or obscure television shows, you will remember those moments when a character appears and you wonder where you have seen her before. Well, you can try <http://www.what-a-character.com/> which has thumbnails, biographies and filmographies of hundreds of lesser-known stars. And the compiler has a nice turn of phrase — Jack Elam is described as one-eyed and multi-faceted.

## **New from W3C**

The World Wide Web Consortium (W3C) continues to do great work in developing new systems which we are going to use way into the future. They have been running a series of workshops around the country showcasing their research. At the Canberra session, I heard Max Froumentin talking about how they are putting the standards in place to bring about consistent and useful systems to integrate voice into web-based applications. His talk 'Giving voice to the web' is at <http://www.w3.org/2004/Talks/04-voice/>. Others to speak were Jane Hunter from DSTC on the Semantic Web, and that giant of his profession, Dean Jackson who spoke about a whole raft of issues — SVC, RSS and lots of other good stuff.

## **A garden is a lovely thing**

One site that Dean recommended is CSS Zen Garden at <http://www.csszengarden.com/>. Its purpose is to demonstrate what you can achieve in design through the use of cascading style sheets. The results

are interesting and often quite beautiful, and designers are invited to take the basic HTML of the page and apply their own styles to it. For anybody who has wondered what the fuss about CSS is, this site is worth checking out.

## **Your listening post on the Live Web**

Another of Dean's recommendations was Technorati at <http://www.technorati.com/>. This site provides a search facility over weblogs and RSS feeds, to track what people are writing and talking about. He said that through technology like this, US Democratic Party candidate Howard Dean was able to get instant feedback on his speeches and tailor them accordingly. Now, this is not necessarily a good thing, but it does show where we may be heading, if we are not careful.

## **And even more from W3C**

If you have ever built a site with meaningful links, you would know the anguish of checking back later, only to find that the links have been lost or changed. And you are bound to understand how tedious it is to keep checking the links to make sure that they are still there. The bad news is that there is no complete solution to the problem, but the folks at W3C have been working on it. The W3C Link Checker at <http://validator.w3.org/checklink/> will give you a report on broken links (including those in linked documents, down to a specified level) as well as letting you know about where pages are being redirected.

## **Folding stuff**

Paper Toys is a site with dozens of cut-out models of 3-D objects, including cars, oil-rigs and the Sydney Opera House. They give you full instructions on folding and mounting the finished work. It is at <http://www.papertoys.com/>. And when you have finished that, you can move onto architectural origami, which is the art form that is the basis of pop-up books and cards. Here is a good place to start: [http://dmoz.org/Arts/Crafts/Origami/Origamic\\_Architecture/](http://dmoz.org/Arts/Crafts/Origami/Origamic_Architecture/).

## **No more heavy reading**

It is important to write for your audience: many web workers do not. There are tests that you can use to check how readable your pages are, but it can be cumbersome to run these: adding up the words, sentences, syllables and then doing the maths. Here is a site that does it all for you: <http://www.readability.info/>. It gives you explanations of the indexes and compares the score for common websites. I ran it over the April column and it looks as though I am writing for people with a tertiary education, which sounds about right. But if you are setting up a site with a wide audience, that might not be an appropriate level. The *New York Times*, for instance, is typically directed to

a Year 9 level of reading.

#### **Making it plain**

And on the subject of comprehensibility, have a look at the Plain English Campaign site at <http://www.plainenglish.co.uk/>. One thing that I liked was the very clear way that they have set out their home page, explaining briefly what the campaign is about, and also telling you what you will find in the various sections of the site. Too often, we assume that people will understand that 'About us' is about the organisation, or could it be about the site?

#### **A sound investment**

The British Library is about to embark on a digitisation program for around 4000 hours of sound recordings, to be made freely-available to higher education. It will include classical and popular music, broadcast radio, oral history, and field and location recordings of traditional music. Examples of the holdings at <http://www.bl.uk/collections/sound-archive/nsa.html> include a live recording of Paul Robeson in *Othello*, Florence Nightingale speaking (one of the earliest sound recordings), and Arthur Conan Doyle talking about the genesis of Sherlock Holmes.

#### **Exposing net censorship**

Academics from Harvard Law School, the University of Cambridge and the University of Toronto have begun monitoring worldwide internet censorship and surveillance. The OpenNet Initiative at <http://www.opennetinitiative.net/> will be using a combination of electronic tools and human investigation to get a better picture of what is going on under the surface of the internet, to check on the 'watchers' to measure their practices against general principles of human rights.

#### **For all of the people all of the time**

The Library of Congress has collected more than 20 000 documents about Abraham Lincoln as part of its American Memory project at <http://memory.loc.gov/ammem/alhtml/malhome.html>. There is an extensive display on Lincoln's assassination, complete with a summary, a timeline and a gallery of ephemera relating to the tragic event. And if you are more interested in the bright side of life, the other special presentation is on the Emancipation Proclamation, his document that turned a war into a crusade for freedom.

#### **How to cite?**

A common question in this cyber-era is 'how do I cite electronic resources?' Well, there are quite a variety of places where you can find this information (including several books!), but they have been collected in the one location at <http://www.cyberbee.com/citing.html>. One of the

links is to <http://www.easybib.com/> where a wizard will ask you to enter the URL, page title, author and date accessed, and will then present you with the appropriate citation. Impressive.

#### **Brewster to the rescue**

The Computers, Freedom and Privacy Conference (CFP) is an annual conference that has been running for fourteen years. It is one of the most important venues for discussing matters relating what we can and should be doing in the online world. The online reporting of this year's conference was achieved through blogs (what else?) and can be seen at <http://cfp2004.org/blogs/>. Brewster Kahle (creator of WAIS, and more recently, the Internet Archive) was a keynote speaker, and he asked the big questions: 'Can we? May we? Will we?' The answer to the first is easy — yes we can, and it is getting cheaper all the time. The problems start in the 'may', with rights becoming more and more important. I could not find a copy of his speech, but here is some discussion arising from it — <http://www.corante.com/copyfight/archives/003351.html>.

#### **Web citation index**

ISI Web of Knowledge and NEC Citeseer have announced that they are joining forces to create a Web Citation Index, a comprehensive, multi-disciplinary citation index for web-based scholarly resources. Thomson ISI and NEC will run a pilot project this year, with full access planned to be available early in 2005. It will cover journals and archives, and index journal articles, preprints, conference proceedings, technical reports. Some (but not all) of the results will be available to users free of charge. It is interesting that they have ruled out providing any archiving activity, instead concentrating on providing access tools.

#### **More online books**

Perhaps you are looking to fill in some time before your next appointment or while you are waiting for a response from a slow site somewhere? You could try Read Print, a free online library at <http://readprint.com/>. Think of it as *Bartleby* without the ads, or something a little bit flashier than Gutenberg. The text is very well set out (in a serif format) and there is a brief biography for each author. Another nice touch is a short collection of articles on English

*Brewster Kahle  
...asked the big  
questions: 'Can we?  
May we? Will we?'*

## Libraries Alive!

Library and information services consultancy. Fresh, pragmatic, effective. Tactical planning, technology management, service improvement. Proving library value. Information? Please contact Sherrey Quinn or Ian McCallum on (02) 6257 9177 [www.librariesalive.com.au](http://www.librariesalive.com.au)