

Webb's web

Be careful what you ask for

The Library of Congress is reporting some success in its collaboration with Flickr requesting readers to tag its photos. The pilot project, named The Commons <http://www.flickr.com/commons>, allows any Flickr user to add a descriptive metatag to one of the photos in the collection. Many of them are useful, but the project team are questioning whether all of the light-hearted tags will be worth recording. I don't know - if they're interested in getting our responses, who's to say what is worth keeping?

Open books online

Late last year, the New York Times reported on a sort of backlash by some libraries against the mass digitisation efforts by Google and Microsoft, opting instead to go with the Open Content Alliance <http://www.opencontentalliance.org>. The libraries that are going the Open route (the Boston Public Library, the Smithsonian Institution and the Boston Library Consortium among them) have to find the \$30 per book, but they regard the money well spent if it delivers the content to everybody. The Times story is at <http://www.nytimes.com/2007/10/22/technology/22library.html> (registration may be required).

There are ways of asking Google and the other search engines to take things out of their cache... but it's cumbersome and fiddly.

Do it yourself

For a while now, we've been encouraged to consider how web 2.0 concepts could be introduced in the government sphere to improve services on offer.

The problem is that government agencies are a little touchy about the accuracy, reliability and relevance of the data on their sites, so there's not been much progress. Except in the UK National Archives, where they've successfully introduced 'Your Archives', a wiki that lets anybody write and edit articles about their historical collection. So far there are over 3000 registered users and more than 1400 articles. See it at <http://yourarchives.nationalarchives.gov.uk>

Good new stuff

Courtesy of Belinda Weaver's 'What's New', here are a few really useful sites. Through JustFreeBooks <http://www.justfreebooks.info> you can explore than 400 websites for all sorts of e-books, searching by title, author or subject. Manuscriptorium http://www.manuscriptorium.com/site/eng/default_eng.asp, on the other hand, concentrates on manuscript material with links to a virtual library of digitised material. The search facility could do with some improvement, but the National Library of the Czech Republic should be commended on a fine start. Finally, Musgle <http://www.musgle.com> uses Google to search for music, which is a very neat use of technology.

Artists online

The University of NSW (with the support of the Australian Research Council) has established the Dictionary of Australian Artists Online (DAAO) as a new virtual research environment for the promotion of scholarship on Australian artists. It's been built to take online biographical submissions, commentary and corrections, and was set up with around 7,000 biographies, taken from a number of existing print publications. They publish new biographies each week from its member base of over 400 artists, scholars, commercial galleries and historians. Have a look at <http://www.daa.org.au>. I like the concept and admire the effort that went into it, but I do wonder about the need to agree to abide by their terms and conditions. The good news though is that it looks like you only have to do it once.

Keeping the search engines under control

I came across an interesting little issue this week. This organisation had a requirement to publish contracts on their website, but only for the term of the contract. When the contract expired, they removed the document and all seemed to be well. Unfortunately, the contractor found that the document had been cached by Google and their out of date pricing information was now findable and readable by all. There are ways of asking Google and the other search engines to take things out of their cache (or their index) but it's cumbersome and fiddly. It's better to think in advance about whether you want your pages and documents indexed in the first place and whether you want them retained in the cache. Find out more with Google's Webmaster Help Center <http://www.google.com/support/webmasters/bin/topic.py?topic=8459>. One other thing: don't forget that beyond the usual suspects of Google, Yahoo! and Live, the Internet Archive <http://www.archive.org> also grabs Web content for posterity.

From the horse's mouth

One thing I've noticed from watching Google's Page Creator system <http://pages.google.com> is how much people don't know about search engines. There's a belief that all you have to do is publish a page and Google and the others will pounce on it and add it to their huge indexes. The reality is somewhat different. The facts are out there on the Web but you do have to search for them. Well, they've now published a concise description that explains how their search engine works. If you need to know or need to show your boss what it's all about, go to <http://books.google.com/googlebooks/pdf/webmastertools.pdf>

Don't forget to visit my blog <http://www.alia.org.au/webbsblog>

