eLucidate Vol. 3 Issue 4, July 2006

ISSN: 1742-5921

Meeting Report: Reaching beyond Bibliographic Referencing

Report of a UK E-information group one-day seminar held at Kings College, London, 1st June 2006

As I'm part of a small team that delivers EndNote bibliographic software training to staff and students at the University of Manchester, this course caught my eye – just in case there were some new developments that we needed to be aware of. Though only about 20 people arrived on the day, most (though not all) were from University libraries.

Tracy Kent from the University of Birmingham set the scene with her opening remarks, outlining some of the challenges that librarians and information workers face currently; these include:

- Over-use of, or over-reliance on Google

- Frustrations caused when linking software takes users to a reference in a journal that their organisation doesn't have a subscription to

- Capturing and organising the references that users do manage to find.

Plus, of course, how librarians/information workers manage the processes above.

Ben Lund, from Nature Publishing Group, took the first session, with a presentation on a comparatively new phenomenon, "social bookmarking". Until quite recently, most people will have been passive users of the Web, finding useful web sites, adding them to their own personal bookmarks and leaving it at that. Social bookmarking goes a step further – there are a few web sites such as Delicious (http://del.icio.us) and Connotea (www.connotea.com) which allow researchers to upload a URL of a site that they found useful, adding some keywords (the hip word is "tags") to allow for subject searching, which then allows others to see someone else's list of bookmarks. The tags can also be shared and even though most of them are going to be natural language, ultimately a thesaurus of the most popular words will be developed.

The so-called "social web" is about to become more advanced, with more websites allowing sharing of photos (www.flickr.com), personal goals (www.43things.com) and events (http://upcoming.org/) – which could become "Time Out" on the web. Some of these sites even encourage a scoring system, to advertise those links seen as "important". Surprisingly, most of these web services are free, just requiring registration. Tim Berners-Lee must be very relieved.

From the point of view of bibliographic referencing, "citeulike" (www.citeulike.org) is an interesting development, allowing users to capture their own records for full-text online articles, letting them add their own notes and allowing the transfer of records to an Endnote library.

Robert Bley, from Ex-Libris, then gave a comprehensive session on the usefulness of openURL resolvers – one of the "invisible" services provided by libraries which most users don't realise is there. The OpenURL resolver allows library holdings of electronic

ISSN: 1742-5921

resources to be dovetailed with databases, allowing users to click on a database link and go straight (most of the time) to a full-text online article. The key point here for librarians is that effective use of both databases and e-journals is achieved, vital when we are all spending fortunes on e-resources – and when all the evidence still shows that students still prefer to use a search engine to find information, with Google being the engine of choice. In Higher education, students are still not searching effectively for information, despite the best efforts of library staff.

A key development has been the "digital object identifier", a numeric code which allows for unique identification of an electronic record. This has allowed more deeplinking, allowing users to go directly to an article and not just a journal home page. The only subject area holding out on this is law, which seems strangely slow to take it up.

Grant Young, from TASI (Technical Advisory Service for Images) had the graveyard slot, directly following lunch, with a session on finding, making and managing digital images. TASI is funded by JISC, but is made use of by the non-HE sector as well.

Finding images can be a tricky proposition – and Google (heretical notion!) is not necessarily the best search engine. Copyright and quality issues are a further problem. At the John Rylands University library in Manchester we're gradually building our own image database using Luna – but if you can't do that there are several image search engines, with pro's and con's for each. Although Google and Yahoo are bigger, Picsearch shows more of the results than either of its competitors. It is also rumoured that Microsoft are developing an image search facility. One of the strange things about searching for images is that most people have to use keywords to find them……..

Managing your own collection of images throws up rafts of issues which need to be decided. The questions that need to be resolved include the functionality required from the system, access to a decent IT infrastructure and the degree of interoperability required with other systems. Cost is always a bugbear and copyright is usually lurking about somewhere..

Yoshimasa Tsuruoka, from the National Centre for Text-mining (NaCTeM – based at the University of Manchester), gave a highly technical and complex session on text-mining, most of which went right over my head. Text mining is essentially a tool which allows more precision in searching for information than combinations of keywords allow. One of the activities that NaCTeM undertook was the analysis of the entire PubMed database to determine which concepts are involved. Once the concepts are discovered, information access can go further than "simple" indexing terms. For example, a search using natural language (eg, what causes breast cancer?) will return much more precise results. I ought to apologise for a gross over-simplification of Dr Tsuruoka's presentation – but I don't think I was the only person mystified… Computer performance at text analysis is still hovering at about 93-94% accuracy, but is improving slowly.

Tracy Kent rounded the day off with an update on the key features now to be found in reference software. Reference software allows users to build their own personal knowledge base of the resources they need – and there's no distinction between print and electronic resources when you need to cite them in a paper. The simple reformatting of references into different styles is a key feature. At Manchester both EndNote and Reference manager are supported, but there are several other packages available, with

ISSN: 1742-5921

an increasing number using the web as a platform. Packages which may be new to some include Mayan, which uses a novel approach using three open panes in a window, showing the original bibliographic reference, the keywords searched for and where those keywords occur in the body of the article. A new version of RefViz was recently launched, which uses both text analysis and a visualisation tool (which is based on keywords).

In her conclusion, Tracy summed up the concepts involved in the other presentations, showing how they were improving the opportunities for researchers to find and manage the information they needed.

I enjoyed the day – much of it was set at the right level (for me, anyway), though the mathematical content of text-mining was a minor problem. Thanks are due to Tracy for managing the time and ensuring everything went smoothly, not forgetting all the hard work put in by Christine from UkeiG.

Martin Snelling, John Rylands University Library, Manchester