Meetings Reports

Metasearching - Better Searching?

**British Computer Society Electronic Publishing Specialist Group one-day seminar
Said Business School, Oxford, Thursday 22 July 2004**

This was a lively meeting that provided good insights on a key topic for journal and for reference publishers. 'Metasearching' is the use of tools to provide better search and retrieval, across multiple databases, services, platforms, protocols, and vendors. From the rapidly increasing number of vendor solutions on the market, it is clearly a growth sector.

Andy Powell of UKOLN provided a clear overview of metasearching, based on the JISC-funded work UKOLN carried, out, for example towards the JISC Information Environment, a set of standards for seamless access across multiple resources. He pointed out the problems of searching across incompatible Web resources: they have different user interfaces; everything is in HTML, so it is difficult to merge, copy and paste the information you have retrieved, and you don't necessarily get access to the resource itself. Metasearching has two approaches: cross-searching, which is real-time searching across several databases, and harvesting, batch processing of pages into a local database. The latter provides a quicker result for the user. The JISC Research Discovery Network is a nightly harvest of several smaller gateways. Searching across databases typically uses the Z39.50 protocol; harvesting is typically done using the OAI (Open Archives Initiative) protocol. Metasearching websites involves the wonderfully named 'web scraping' of information from the site, which, as its name suggests, is an imperfect solution. As soon as the web page is updated, the scraping may no longer work.

Increasingly, solutions are appearing to the problem that provide better standards for metadata, for example the NISO Metasearch Initiative, and repositories of metadata such as the eGov metadata standard and the LOM metadata profile.

Other presentations showed how these approaches are being implemented in specific sectors. John Davidson of Sentient Learning described how Sentient tools can be used to make a university reading list shared across all the stakeholders, including the library, students, and local bookshops, thereby enabling better use of existing resources. James Culling of ExLibris described widely used tools such as SFX, an Open-URL compliant link server, providing context-sensitive linking for journals, and MetaLib, providing access to an institution's e-resources, whether local or remote.

Hilary Ollerenshaw described a recent initiative from the North Bristol NHS Trust. Entitled Knowledge4Health, it is a portal of internal and external resources, providing a single point of access for evidence- and Trust-based patient information, including local, regional, and national services.

Chris Knowles of Magus Research described a rather different problem: searching across an organised, structural and managed body of information. Merchant banks, for example, frequently want to search across several structured but incompatible resources. He described tools to facilitate such a search from a single search screen; a similar approach is now being implemented for law firms.

Finally, Martin Kelly of the Institute of Physics Publishing talked about initiatives to improve searching across large-scale scientific databases. These included using a proprietary search tool, Vivisimo, which groups search results into clusters, and compared it to Verity. Vivisimo worked well with unstructured data, while Verity was better where a taxonomy already exists.

**Michael Upshall
Consultant advising publishers
on electronic product development and
dissemination
michael@consultmu.co.uk**

Book Reviews

**Metadata for information management and retrieval
David Haynes. London: Facet Publishing, 2004
ISBN 1-85604-489-0 xiv, 186pp. £39.95**

It's over eight years since I last worked full-time as a librarian. Though I do still return to the fray occasionally, temping or doing contract work, my primary occupation now is as a freelance indexer – mostly of books or other print materials, but I've also ventured into web site indexing, dabbled in indexing of other electronic texts, and generally tried to stay on the track of emerging ideas and new developments in ways of getting information from electronic resources.

Anyone who has followed a similar trail will know that it's fairly easy to pick up a basic understanding of metadata as a tool for resource description and