Meeting Report: Desktop Search Tools – Managing the Flight Deck

SOAS, London, 14 June 2006

Gary Horrocks, our UKeiG chairman, in his introduction to the day, highlighted the importance of the topic by pointing out that it can be easier to search for information on the whole of the World Wide Web than to find what's on your own PC! The meeting addressed search tools that could be used not only to search a personal desktop but also corporate intranets, and throughout the day consideration was given to both personal and corporate information management issues. The concept of desktop search is much wider than just finding Word documents or other kinds of saved files such as images. It encompasses e-mail folders, Web bookmarks, Web histories and even databases and corporate digital records. The new search tools available don't just look at file names, but some can search within documents, including the text in pdfs and the metadata attached to images etc.

We heard first from David Graham from MSN-UK who gave us a preview of the beta version of Windows Live, which will shortly be launched to take over from Microsoft's existing desktop search offerings. Windows Live aims to bring together searches of the desktop, intranet and the external Web and David stressed the advantage of the "familiar interface". Windows live will create an index from your desktop and intranet files and also websites of your choosing, but unlike their current product, it does not slow down your computer when you want to work on it; instead, it pauses as soon as you start to work. Add-ins will also be included to support Mozilla Thunderbird (e-mail) and Open Office for those who prefer a less than total Microsoft experience.

In a lively question time, David confirmed that Boolean logic searches would be supported, and promised to take back some suggestions made, for example that checkboxes be included in results lists, for subsequent processing.

We were scheduled to hear next about Copernic Desktop Search from Martin Bouchard, but unfortunately this speaker did not materialize. Stephen Robertson of Microsoft research UK brought forward his offering "Research for Search." This was a more academic overview of the way in which search engines work and how they can be 'taught' to work more effectively. Search engine ranking algorithms are the key to good retrieval performance, and these can be optimised in different ways. Training by feedback from evaluation of search results can be very effective, but is also expensive. The big players in Web search, such as Google, invest heavily to make their Web search engine the best. In very large companies it might be feasible to employ search engine trainers to optimise enterprise search, but that is not going to happen in the desktop search environment, so alternatives are needed, for example autotraining. This is where more

research is needed, and Stephen was confident that ongoing research, both academic and commercial, would deliver these improvements in future, to make desktop search as effective as Web search.

In the ensuing question time it was it was pointed out that, unlike Web searching, where the use of metadata has been abandoned because of spamming, desktop or enterprise search could be improved by exploiting metadata. Retrieval and ranking could also be made more effective if meaningful file names and good folder management were used. Greater weight might be given to objects meaningfully named, and contained within suitably named folders, thus contradicting the sales pitch from some desktop search vendors that you can abandon good file management practices if you use their product – a reassuring point for information professionals who favour and advocate such systematic tidiness!

In the continued absence of our missing speaker, Karen Blakeman gave us a useful introduction to the UKeiG Wiki, and encouraged us all to take a look and experiment – if not on the real thing, then on the dog's breakfast!

After a tasty lunch and the UKeiG AGM, we resumed our desktop search theme with a presentation about ISYS search software from Terry Clift, General Manager of their Global operations. ISYS is a well established player in the market, having been operating since 1988, starting with a DOS version. Their product is pretty sophisticated and is aimed mainly at the enterprise market, using intranets. Terry demonstrated the current version, 7 and gave us a sneak preview of version 8. The software is able to handle massive volumes of data and indexes discriminatorily and in the background. Searching allows advanced techniques such as Boolean and proximity searching and can search for dates no matter the format in which they appear. Fuzzy matching and possible misspelling suggestions are included as well as multi-lingual support via Unicode, and metadata can be used intelligently. The results may be ranked by user-selectable preferences and categorised into types. Hits are highlighted and one can navigate from hit to hit. The demonstration showed how synonyms could be automatically searched and were highlighted in the hits shown, e.g. a search used the word "America", but the results included "USA" etc. Different tenses of the same word could also be detected.

Finally, Karen Blakeman gave us an overview of the good, the bad, and the ugly in desktop search tools, comparing eight different offerings from, Ask, Blinkx, Copernic, Exalead, Google, ISYS, Windows, and Yahoo. After outlining the features to look for, each product was explored with reference to these needs, and some were given the thumbs down, at least in their current version, if they didn't come up to scratch on these requirements. ISYS was considered supersonic business class (using the flight deck metaphor), and great for corporate intranets, but it is probably too pricey for personal use, while the current Windows desktop search was 'budget airline'. Of the free competitors,

Copernic and Exalead seemed to be the best bet at present, but the field is very fast moving. To keep up with developments we were all encouraged to keep an eye on the UKeiG Wiki.

A panel discussion followed, which continued the lively discussion, a feature of the whole day. I certainly found it to a highly enjoyable and informative day, and the active participation of so many seemed to indicate that I was not the only one to do so. All I need now is some time to try out some of the recommendations and keep up with the wiki.

Andria McGrath, Information Specialist at King's College London