Meeting Report: Selecting and Implementing Web and Intranet Search

CILIP Ridgmount Street, London, 22 February 2007

This course was entertainingly delivered by Martin White of Intranet Focus Limited, a very familiar figure to UKeiG members and a pillar of the information profession for the last three decades or more. He had an enthusiastic audience of 29.

Martin soon reminded us that there is no universal solution to the challenge of implementing Search. To help us work out the most appropriate solution in each case, we should bear in mind a common framework of concerns and questions to be explored. He shows this framework as a triangle, in which the apices represent:

- Information (a.k.a. content)
- Technology (for processing, storage and network management, as well as integration with other applications)
- User requirements (including definition of types of query, and usability needs).
- Study of all three aspects is a vital step in developing a specification of the search requirement.

This framework compares interestingly with the Venn diagram set out by Rosenfeld and Morville in their classic book[1] on information architecture, in which three overlapping and interlocking circles represent:

- Content
- Context
- Users

At first glance “Context” might look different from “Technology”, but the Rosenfeld/Morville model explicitly includes technology as an important element of context, alongside goals, funding, politics, culture and constraints. Plainly there’s a lot of commonality between the two models. Whether we show it with a Venn diagram or a triangle, this general framework can serve us well in a great many other information management applications.

Concerning the Information apex of his triangle, Martin advocated carrying out a full content audit before specifying the search requirement. It is helpful to know the size distribution of the documents to be indexed, the file types, the languages of the text, confidentiality restrictions, rate of expansion/updating of the collection, etc. But if the audit has revealed anomalies or inconsistencies in the content (including its metadata) or if the information architecture provides a poor structure for it, Martin warned us that it’s no good hoping a good search engine will paper over the cracks – the real problem has to be addressed first.

---

[1] Louis Rosenfeld and Peter Morville. Information Architecture for the World Wide Web. 2nd edition. O’Reilly Media, Inc: Sebastopol, USA: 2002. ISBN 0-596-00035-9 See p.23-24. [NB a third edition came out in December 2006.]

On the Technology front, Martin stressed the importance of understanding how search works, especially the building and maintenance of indexes, crucial to retrieval performance. This is not something we can leave to our IT colleagues, especially when it comes to trouble shooting, e.g. working out how it is that a document we **know** to be loaded is not showing up among the search results. Even better if understanding how it all works enables us to anticipate the problem before it occurs, and take steps perhaps to increase the spidering frequency, or to improve the consistency of meta-tagging, etc.

The “average user” is an elusive character – not only is each of us an individual, but the needs and preferences of a single user can vary from one day to the next. To overcome the difficulty of defining User requirements, Martin has found persona development a helpful technique. The idea is to build a detailed picture of a small number (say 5 or 6) of hypothetical but nonetheless typical users or ‘personas’ to be served by the search system. In a pharmaceutical company, for example, one persona might be a research chemist; another might be someone running clinical trials. Understanding the needs of the personas can help us to define user interfaces, to select a representative document set for product evaluation, to set up usability tests, etc.

With those three aspects fully understood we can build a specification of the requirements, develop a business case, and begin the process of selecting candidate software products and evaluating them. A year may easily pass before procurement ends and implementation begins. And that’s just the start – Martin warned us that implementation never ends.

During the course we had some fun picking holes in the search capabilities of websites available on the Internet. Some of them were the websites of the search engine vendors themselves, and others belonged to organisations that you might expect to be setting an example to the rest of us. “What’s new?” you may say – we have all suffered bruising experiences while trying to find things on the websites of important organisations. But things do not have to be this way. Very often there is nothing fundamentally wrong with the search engine, that cannot be rectified by putting more effort into buying all the modules we need, implementing them thoroughly, optimising the parameters and constantly monitoring performance.

Some consultants give good strategic advice while steering clear of the coal-face; others act more like contractors, getting through the legwork but contributing little to overall direction. So at this meeting it was refreshing to receive all-round guidance from one who knows the business from top to bottom. Martin’s enthusiasm inspired us all to go forth and get search working.

Stella G Dextre Clarke

