Semantic Analysis Technology

Meeting of the International Society for Knowledge Organization (ISKO), 3 November 2008, University College, London

This well-attended meeting (over 100 attendees) considered the topical issue of semantic analysis, more specifically the use of semantic-based tools to improve searching. The problem with meetings that have "semantic" in the title is that they can stray in one of two directions. Either they move to advanced concepts that are all too often incomprehensible, thereby unintentionally demonstrating that "semantic" equates with "unintelligible", or they turn into a sales pitch stating bluntly that semantic analysis has been used to create a result via some clever jiggery pokery, which is not explained, but which is available for sale. This meeting was remarkable fro the low proportion it contained of either.

It was clear from the session that semantic analysis has progressed considerably over the past five years or so, and can demonstrate immediate value on implementation (in other words, it isn't just a leap in the dark). It demonstrated as well the fundamental challenge of semantic analysis: better searching requires better metadata. HTML is crippled by its lack of good metadata, and Google's strengths and limitations are largely down to what can be managed using only HTML.

If more metadata is the answer, do you add that metadata automatically or by hand? There were presentations here both by technologists, who wanted to do things automatically, and who were satisfied with the results despite the errors, and the information professionals, who saw the errors as proof that the process needed to be done by hand. The best solution probably lies somewhere in the middle.

In addition, the event combined presentations by vendors (Italian-based Expert System, UK-based Smart Logic), as well as reports from practical projects (at the Financial Times and the BBC). I found the reports of semantic analysis in practice the most interesting presentations, perhaps because I don't need convincing of the need to improve the quality of searching, I just need evidence that someone has worked out how to do it better.

For me, the most exciting presentation was by Rob Lee of Rattle Central (www.rattlecentral.com). He described a project by which automatic links were added from BBC news stories to articles on Wikipedia, via DBpedia, (http://dbpedia.org), a structured version of Wikipedia that is generated from Wikipedia entirely automatically, and which provides a set of machine-readable links to much of the content in Wikipedia. If anyone wants to be convinced of the power of the Semantic Web, here is a good place to start.

Michael Upshall