

Enterprise Search — The New Frontier?

David Hawking

CSIRO ICT Centre, Canberra, Australia

David.Hawking@csiro.au

<http://es.cmis.csiro.au/people/Dave/>

The advent of the current generation of Web search engines around 1998 challenged the relevance of academic information retrieval research – established evaluation methodologies didn't scale and nor did they reflect the diverse purposes to which search engines are now put. Academic ranking algorithms of the time almost completely ignored the features which underpin modern web search: query-independent evidence and evidence external to the document. Unlike their commercial counterparts, academic researchers have for years been unable to access Web scale collections and their corresponding link graphs and search logs.

For all the impressive achievements of the Web search companies, great search challenges remain. Nowhere is this more so than behind the organisational firewall, where employees cry out for effective search tools to permit them to find what they need among huge accumulations of text data, heterogeneous both in type and in format, and subject to security and privacy restrictions. Worldwide, there are almost certainly hundreds of thousands of organisations whose electronic text holdings are larger than (but very different from!) the TREC ad hoc corpus. Do we as an academic community know anything about the character of these collections? Do we know how employees search? What they search for? How they judge the value of what is retrieved? Do we have effective algorithms which can deliver results tailored to the context of their search? Enterprise search is at a more manageable scale than the Web, but nonetheless presents formidable problems for academic researchers. Can academic researchers overcome them, or will the field be left to commercial companies?

The talk will outline the nature of the enterprise search domain, review the current state of research in the area, present some research results, highlight some non-standard applications of search, discuss evaluation methodologies and pose challenges.

Biography. David Hawking is the founder and chief scientist of CSIRO's enterprise search engine project (Funnelback: <http://funnelback.com>). Funnelback is a commercial product permitting effective metadata and/or content search of heterogeneous enterprise information sources including websites, email, fileshares and databases. David was a coordinator of the Web track at the international Text Retrieval Conference from 1997-2004 and has been responsible for the creation and distribution of text retrieval benchmark collections now in use at over 120 research organisations worldwide. In 2003 he was awarded an honorary doctorate from the University of Neuchatel in Switzerland for his contributions to the objective evaluation of search quality. He won the Chris Wallace award for contribution to computer science research in Australasia, for the years 2001-2003.