

Reflections on Information Retrieval: the 2016 Strix Annual Lecture

Gary Horrocks, editor - eLucidate

gary.horrocks@gmail.com

“I will no doubt be accused of elitism if I expressed my long-held view that the processes of information management and retrieval can never be simplified to a point where they be conducted by half-wits...Finding useful information is an intelligent process requiring intelligent people because at the end of the day only the intelligent can recognise what is useful.”

Tony Kent in a letter to Jan Wyllie, 1991, excerpted from the Tony Kent Strix Award 2016 booklet

In 2015 Peter Ingwersen, Professor Emeritus at the Royal School of Library and Information Science, Copenhagen University, Denmark, was presented with the prestigious Tony Kent Strix award at a ceremony in London in recognition of his outstanding contribution to the field of information retrieval. It was both a surprise and an honour for him, and he marked the award with his 2016 Annual Strix Lecture “Context in Interactive Information Retrieval” on Monday 31st October. [[Abstract and video available.](#)] “There were many supporting letters from outstanding IR researchers worldwide. I was very grateful for their acknowledgement of my contributions to the field over three decades of research.”

The 2015 presentation took place as part of the Enterprise Search Europe Conference held at the Olympia Conference Centre in Kensington. “The last time I’d participated in a similar conference was in 1993, when I had the honour of receiving UKeIG’s Jason Farradane Award. At the time the IT explosion was just beginning. The personal computer was becoming a major player, but the Internet was not a hot topic and no smart phones existed, only ordinary cell phones. In 1993 the information profession was, unknowingly, coming to the end of the ‘intermediary’ era, because end-users were increasing their direct online access to (scientific) information. Within five years the global Internet and Web were with us. Boolean logic, operators and set combinations were still in use, but becoming less dominant. With Web searching, set combinations all but disappeared.”

Ingwersen argues that, while value-added bibliographic databases and commercial (often costly) discovery tools still exist as part of the deep “hidden web”, the rise of free and accessible-to-all search engines means personalised value addition is now taking place in real time during the search process using relevance ranking algorithms, spelling correction and auto-completion query suggestion and reformulation, for example. He articulates a spectrum of relevance ranking from algorithmic search engine generated results to the highly subjective social and “emotional” relevance based on perceived pertinence and

utility, highlighting an IR model which includes elements of relevance feedback where the assessment of search results is fed back into a new iteration of the search.

While ranking has its uses it is still inherently flawed and Ingwersen, playing devil's advocate, reflects that the peril of disintermediation still haunts the profession in an age of budget cuts and austerity. "We all know that more than 60% of the 'good stuff' is hidden and never found. The so-called 'deep or invisible web.' But who cares? There is always something found that is 'relevant' or useful. And this fact might be the big 'killer' of the information profession in the future, because politicians, institutional and business managers can be very short sighted. As long as you can find 'something' using a search engine, why bother to maintain a library collection or an in-house information system?"

Professor Stephen Robertson opened the 2016 Strix Annual Lecture with a thought provoking presentation, "Search: Then and Now." He reflected on the pre- and post-web worlds and the forces that have shaped search over the years, taking us back fifty years to library card catalogues, classification schemes and subject headings to printed indexes (back-of-the-book, phone directories dictionaries, encyclopaedias) and scientific abstract journals. He discussed human assigned indexing languages: thesauri, faceted classification and controlled vocabularies. It was fascinating to recall the early days of search engines in the 1990s: Jonathon Fletcher's [JumpStation](#) crawled and searched titles and headings only, but had no ranking capability. [WebCrawler](#) was the first full text index, [Lycos](#) - the first big commercial endeavour and [AltaVista](#) came on the scene offering ranking and natural language search. He argued that ranking and natural language searching both pre-date the Web, and that research and initiatives in these areas paved the way for modern web search engines.

Ranking research has been prevalent since the 1960s, and started to percolate into early web search engines in the 1990s using a variety of data including word matching, frequency, field information (title and abstract) and anchor text, for example. Natural language query formulation has been researched since the late 1970s, mainly associated with Boolean operators and machine extraction of terms. The Okapi text retrieval systems project, based at City University London in the early 1980s, was a great example of research into a best match ranking system. When Google came along in the late 1990s it implemented all the existing ideas (crawling, anchor-text gathering, word indexing, ranking) throwing [PageRank](#) into the mix.

Twenty years on, Google still strides like a colossus across the information landscape, changing search forever. Robertson articulated the state of play today. "Search is easy. We can search using a few natural language words. Search is almost instantaneous. If it doesn't produce what we want, we can just try again. We can search over a vast range of different types of information, for a vast range of different types of need. We don't need to be able to spell. We hardly need to think." I can visualise this statement as a potential exam question for today's LIS students.

Discuss ☺