Quo vadis information retrieval research

Kal Järvelin

The talk begins by an analysis of the tradition of IR research based on the TREC approach. The structure of IR experiments, so characteristic for IR research, is discussed in detail, with IR effectiveness as the dominating dependent variable. The strengths of the approach in research and in practice are acknowledged. The talk then moves on to pointing out mounting challenges to IR (evaluation) and IR system development:

- Sometimes there are no articulated information needs preceding information access. The searcher first needs to find a focus, then develop questions. Is it possible to support the searcher?

- Sometimes there are neither unique questions / queries nor unique right answers. The searchers are individual and inconsistent throughout. Is it possible to support the searcher?

- Practical IR is often a process, not a single shot at the database, while IR evaluation is by-and-large based on one-query sessions. IR processes have rarely been sufficiently described in recent times. Therefore they cannot be understood, properly supported, nor evaluated.

- Practical IR is rarely performed in vacuum at the center of the universe. Rather it is highly integrated with the other components of the searcher's information environment. This is unfortunately not reflected in IR evaluation.

- Relevance is dynamic and multidimensional while measured in the opposite way as stable, topical and binary. Simple measures like searchers' clicks are increasingly taken as indications of relevance while they are insufficient and do not reliably predict relevance.

- Searchers' task performance is independent of IR system effectiveness. People cope with clearly degraded retrieval systems as well as with better ones. Just their behavior is changed. Are we investing our research efforts optimally?

All this suggests that, in addition to search engine development, IR (experimentation) might be deserving of other serious foci. These studies might not almost exclusively, and certainly not primarily, focus on search engine effectiveness as the paramount dependent variable. The talk finishes by discussing a cognitive approach to IR as a way for developing material theories on information access with more varied designs of dependent and independent variables. It is the author's view that, by lifting one's eyes from the search engine effectiveness fixation, it is readily understood that 80% of the IR terrain is unmapped, even from the CS viewpoint. Or, if no alternative approaches matter, we might close up shop.

Kal Järvelin is an Academy Professor at the Academy of Finland, working at the Dept. of Information Studies, University of Tampere. He holds a PhD in Information Studies (1987) from the same university. Kal's research covers information seeking and retrieval, database management, and structured documents; and linguistic and conceptual methods in IR. He has authored over 200 scholarly publications and supervised fourteen doctoral dissertations. Kal has served the ACM SIGIR Conferences as a program committee member (1992-2005), Conference Chair (2002) and Program Co-Chair (2004, 2006). He is an Associate Editor of Information Processing and Management (USA).