Invited Talk

"Who's the user? Who's the researcher?"

Maarten de Rijke

Informatics Institute, University of Amsterdam

Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

mdr@science.uva.nl

Abstract

Over the past few years there has been a lot of progress in technology used for addressing monolingual or multilingual web queries in languages other than English. Nevertheless, a great deal of work still remains to be done, e.g., on the morphological analysis of non-English web queries, before the retrieval performance on English and non-English are on a par. There's another pressing issue, however, that's at least as important: we know very little about users of monolingual or multilingual (non-English) web search facilities. Who are they? What do they search for? What are their intents? At WebCLEF --- the multlingual web retrieval track run at CLEF --- these questions and concerns have led to a very explicit definition of the retrieval task, where various assumption are being recorded as part of the topic statement. In the talk I will review the choices made at WebCLEF over the past few years and detail (and motivate) the current set-up.

Another important aspect of the talk concerns the lack of user data that most academic research groups have to work with. I discuss various ways around this, one example being the use of publicly available and usable showcases and demonstrators. We (the University of Amsterdam) have run and continue to run a small number of Dutch language online search and browsing tools. At the workshop I will discuss a number of findings of this strategy, based on a brief log analysis together with both quanitative and qualitative analyses.

This talk is based on joint work with Leif Azzopardi (Glasgow), Krisztian Balog, Valentin Jijoun, Jaap Kamps (Amsterdam), and Borkur Sigurbjornsson (Barcelona).