

Syntactic and semantic structure in web search queries

Rosie Jones

Traditionally, information retrieval examines the search query in isolation: a query is used to retrieve documents, and the relevance of the documents returned is evaluated in relation to that query. The query itself is assumed to consist of a bag of words, without any grammatical structure. However, queries can also be shown to exhibit grammatical structure, often consisting of telegraphic noun-phrases. In addition, users typically conduct web and other types of searches in sessions, issuing a query, examining results, and then re-issuing a modified query to improve the results. We describe the properties of real web search sessions, and show that users conduct searches for both broad and finer grained tasks, which can be both interleaved and nested. Reformulations reflect many relationships, including synonymy, hypernymy and hyponymy. We show that user search reformulations can be mined to identify related terms, and that we can identify the boundaries between tasks with greater accuracy than previous methods.

Rosie Jones is a Senior Research Scientist at Yahoo!. Her research interests include web search, geographic information retrieval, and natural language processing. She received her PhD from the School of Computer Science at Carnegie Mellon University under the supervision of Tom Mitchell. She is co-organizing the WSDM 2009 Workshop on Web Search Click Data (WSCD09). She served on the Senior PC for SIGIR in 2007 and 2008, and is a Senior Member of the ACM.