

Information Retrieval with Query Hypergraphs

Michael Bendersky*

Department of Computer Science
University of Massachusetts Amherst
Amherst, MA
bemike@cs.umass.edu

July, 2012

Current information retrieval models are optimized for retrieval with short keyword queries. In contrast, in this dissertation we focus on longer, verbose queries with more complex structure that are becoming more common in both mobile and web search. To this end, we propose an expressive query representation formalism based on query hypergraphs.

Unlike the existing query representations, query hypergraphs model the dependencies between arbitrary concepts in the query, rather than dependencies between single query terms. Query hypergraphs are parameterized by importance weights, which are assigned to concepts and concept dependencies in the query hypergraph, based on their contribution to the overall retrieval effectiveness.

Query hypergraphs are not limited to modeling the explicit query structure. Accordingly, we develop two methods for query expansion using query hypergraphs. In these methods, the expansion concepts in the query hypergraph may come either from the retrieval corpus alone or from a combination of multiple information sources such as Wikipedia or the anchor text extracted from a large-scale web corpus.

We empirically demonstrate that query hypergraphs are consistently and significantly more effective than many of the current state-of-the-art retrieval methods, as demonstrated by the experiments on newswire and web corpora. Query hypergraphs improve the retrieval performance for all query types, and, in particular, they exhibit the highest effectiveness gains for verbose queries.

Available online at <http://ciir.cs.umass.edu/~bemike/pubs/phdthesis.pdf>

*Current affiliation: Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA.