

Referent: David Milne
Titel: Knowledge-based Information Retrieval with Wikipedia

Abstract

In knowledge-based information retrieval, search engines consult external sources of knowledge--ontologies, taxonomies, thesauri, glossaries, gazeteers--to help process the documents they encounter and the requests they receive. The idea is old, obvious, and compelling--but results have been singularly unimpressive. The best performing and most widely used search systems are still those that deal in lexical character patterns without using any structured knowledge to understand them.

Wikipedia is changing all that. This open, constantly evolving encyclopedia represents a vast pool of topics, translations and semantic relations. It is arguably the largest knowledge base humanity has ever seen. At last we have a resource that is (or may be) sufficiently broad, deep, and timely to be applicable to open-domain information retrieval. However, it brings its own challenges. Wikipedia's haphazard and only partially machine-readable structure bears little resemblance to the carefully crafted knowledge bases that have been used to assist information retrieval in the past.

This talk will discuss Wikipedia's promises and shortcomings, and describe ongoing investigations of how best to apply it to organizing and retrieving information.