Contextual Insight in Search: Enabling Technologies and Applications

Aleksander Øhrn

Fast Search & Transfer ASA
Christian Frederiks plass 6
N-0102 Oslo
Norway
aleksander.ohrn@fast.no

Abstract

When the fields of efficient and scalable search, data aggregation, XML search and information extraction come together, we get a powerful and exciting mix. Taken together, these are enabling technologies that make it possible to search for information in new ways and that make new types of search applications possible. Using a state-of-the-art enterprise search platform as an example, this tutorial outlines the workings of these technologies and presents some applications of their intersection.

Search engines differ from traditional databases in several ways, yet they both address the issue of organizing information and making it retrievable. The anatomy of a modern search engine will be presented. Although search engines are typically associated with web search, some are now equipped with features usually associated with databases and structured data, e.g., the ability to do aggregation of meta data across a full result set.

Most search engines are built around a flat document model: A document is seen as a collection of typed fields, but the fields themselves have no particular structure. As such, queries tend to focus on document content, with little or no constraints on document structure. Recently, however, scalable and efficient search engines have appeared that support indexing and retrieval of complex XML. Queries that are posed to a search engine can thus combine content and structure in ways that enable extreme search precision. Furthermore, data aggregation can be restricted to take place on the level of the matching document fragments instead of on the document level, thus providing more contextually relevant statistics.

Given XML capabilities, applying text mining and information extraction techniques to the content becomes particularly interesting: By automatically detecting semantic entities, and possibly also relations between these, this information can be searched for in different ways and, e.g., enable applications that have a strong element of discovery to them.

Proceedings of the 31st VLDB Conference, Trondheim, Norway, 2005