Graduiertenkolloquium Angewandte Informatik

Search Relevance based on the Semantic Web

M.Sc. Veli Bicer
Forschungszentrum Informatik

Abstract

The amount of data stored and shared as well as the active number of users is continuously increasing on the Web. When such a large number of people meet with such a large amount of information on the Web, finding the right information in such a quite large scale is definitely a difficult task. Remarkable progress has been made over the last decade with the help of search technologies applied on the Web that is able to collect, maintain and filter the information according to users' needs. Search relevance is a core topic in Web search and directly affects the search performance. Several proposals are made in this regard employing different techniques to represent, model and measure relevance of a particular search result to the users.

We explore the challenge of search relevance in the context of semantic search. We firstly introduce a general framework for relevance as it is understood in the context of Web search which allows us to compare different notions of relevance that exist in literature. Specifically, the notion of semantic relevance can be distinguished from the other types of relevance in Information Retrieval (IR) in terms of employing an underlying semantic model. We propose the emerging Semantic Web data on the Web which is represented in an RDF graph structure as an important candidate to become such a semantic model in a search process. However, the semantic gap between the structured representation of this type of data and the textual content is a major challenge for its applicability to Web search. For this purpose, we propose a probabilistic generative model to be trained from existing datasets of the Semantic Web data in order to bridge this semantic gap.

Based on this model, we introduce novel ranking models by extending long-studied IR-based models that have been widely applied to Web search. Specifically we present a semantic relevance model that can model and determine relevance based on the semantic data by exploiting existing datasets to improve retrieval performance. In fact using Semantic Web data as a relevance source is orthogonal to existing search paradigms and can be used in a variety of contexts. We also show how semantic relevance can be utilized to extend different types of IR-based models (i.e. generative or discriminative) to search different types of data (i.e. unstructured and structured).

Termin: Mittwoch, 21. Dezember 2011, 15:45 Uhr

Ort: Englerstraße 11, 76131 Karlsruhe
              Kollegiengebäude am Ehrenhof (Geb. 11.40), 2. OG, Raum 231
              (Hinweise für Besucher: www.aifb.kit.edu/Allgemeines/Besucher)

Veranstalter: Institut AIFB, Forschungsgruppe Wissensmanagement

Zu diesem Vortrag lädt das Institut für Angewandte Informatik und Formale Beschreibungsverfahren alle Interessierten herzlich ein.

Andreas Oberweis, Hartmut Schmeck, Detlef Seese, Wolffried Stucky, Rudi Studer (Org.), Stefan Tai