

Graduiertenkolloquium Angewandte Informatik

Exploiting Social Semantics for Multilingual Information Retrieval

Dipl.-Inform. Philipp Sorg
Institut AIFB

Information Retrieval (IR) deals with delivering relevant information items given the specific information needs of users, which is an every day problem that concerns almost everybody in our society. In this talk, we present research results on the problem of multilingual IR (MLIR), which defines retrieval scenarios that cross language borders. MLIR is a real-world problem which we motivate using different application scenarios, for example search systems having users with reading skills in several languages or expert retrieval.

As the main topic of this talk, we consider how user-generated content that is assembled by different popular Web portals can be exploited for MLIR. These portals, prominent examples are Wikipedia or Yahoo! Answers, are built from the contributions of millions of users. We define the knowledge that can be derived from such portals as Social Semantics. Further, we identify important features of Social Semantics, namely the support of multiple languages, the broad coverage of topics and the ability to adapt to new topics. Based on these features, we argue that Social Semantics can be exploited as background knowledge to support multilingual retrieval systems.

Our main contribution is the integration of Social Semantics into multilingual retrieval models. Thereby, we present Cross-lingual Explicit Semantic Analysis, a semantic document representation that is based on interlingual concepts exploited from Wikipedia. Further, we propose a mixture language model that integrates different sources of evidence, including the knowledge encoded in the category structure of Yahoo! Answers.

Termin: Dienstag, 14. Juni 2011, 11:00 Uhr

Ort: Englerstraße 11, 76131 Karlsruhe
Kollegiengebäude am Ehrenhof (Geb. 11.40), 2. OG, Raum 253
(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 Schreck, Detlef Seese, Wolffried Stucky, Rudi Studer (Org.), Stefan Tai