

Kolloquium Angewandte Informatik

Semantic-enhanced search: Finding meaning in large-scale scanned text collections

Dr. Annika Hinze
University of Waikato, New Zealand

Abstract:

This talk presents Capisco, a system for semantic-enhanced search in a digital library of full-texts. Document search in Digital Libraries typically use purely lexical analysis, which cannot address the inherent ambiguity of natural language. A semantic search approach offers the potential to overcome the shortcoming of lexical search, but even if an appropriate network of ontologies could be decided upon it would require a full semantic markup of each document. Capisco instead analyzes documents by the semantics and context of their content. The disambiguation of search queries is done interactively, to fully utilize the domain knowledge of the scholar. Our method achieves a form of semantic-enhanced search that simultaneously exploits the proven scale benefits provided by lexical indexing.

For established systems, completely replacing, or even making significant changes to the document retrieval mechanism would require major technological effort, and would most likely be disruptive. We explored ways to use the results of semantic analysis and disambiguation, while retaining an existing keyword-based search and lexicographic index. We engineer this so the output of semantic analysis (performed off-line) is suitable for import directly into existing digital library metadata and index structures, and thus incorporated without the need for architecture modifications.

More information at <http://www.cms.waikato.ac.nz/people/hinze>

Termin: Freitag, 07. Oktober 2016, 14.00 Uhr

Ort: Kaiserstr. 89, 76133 Karlsruhe
Kollegiengebäude am Kronenplatz (Geb. 05.20), 1. OG, Raum 1C-04
(Hinweise für Besucher: www.aifb.kit.edu/web/Kontakt)

Veranstalter: Institut AIFB, Forschungsgruppe Betriebliche Informationssysteme

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

A. Oberweis, H. Schmeck, R. Studer, Y. Sure-Vetter