KIT
Karlsruher Institut für Technologie

aifb

**Institut für Angewandte Informatik
und Formale Beschreibungsverfahren**

Graduiertenkolloquium Angewandte Informatik

User Interfaces to Semantic Knowledge Bases based on Natural Language Generation and Semantic Wikis

M.Sc. Basil Ell
AIFB

Abstract:
The core idea of the Semantic Web vision is the evolution from a Web of hyperlinked human-readable web pages to a machine-interpretable Web of Data. Since natural language is a suitable knowledge representation formalism for humans and not for machines, knowledge representation formalisms have been developed. This development naturally leads to a gap between machine-interpretable and human-understandable content. We tackle several problems related to this gap.

We analyze the human-readability of the Web of data in terms of the availability of human-readable labels of entities. Since our analysis shows that labels are missing for a significant fraction of entities we developed an approach to derive labels from SPARQL query logs.

In the context of search interfaces to semantic data a class of SPARQL query-generating systems exists where users signify their information needs in the form of keywords or (controlled) natural language questions. For the purpose of enabling a user to observe a potential discrepancy between an intended question and the system-generated query we developed an approach to SPARQL query verbalization - the meaning of a query encoded in SPARQL is conveyed to the user via English text.

The different syntaxes of RDF are not suitable for the presentation to casual users. However, information encoded in RDF can be of interest, e.g., when RDF data is returned by a search interface. We introduce a template-based approach for RDF graph verbalization. Since manual creation of these templates is tedious work we developed a language-independent method for extracting RDF verbalization templates from a parallel corpus of text and data based on distant-supervised simultaneous multi-relation learning and frequent maximal subgraph pattern mining.

In a context where semantic data is well-labeled, we explore how semantics support researchers in various stages of corpus-based analysis, such as importing research data, enriching, cleansing, exporting and sharing research data to allow for reuse and to facilitate new socio-technical interactions between researchers and libraries in a Virtual Research Environment based on semantic wiki technologies. We give an example of a concrete research practice, i.e., the qualitative and quantitative analysis of a large digital corpus of educational lexica in the field of history of education.

Termin: Freitag, 11. Juli 2014, 14.00 Uhr

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

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.)