

Report: Research Visit to Vrije Universiteit Amsterdam February 16th to March 11th 2005

Wolf Siberski, L3S Research Center, Hannover

Topic/Goals:

This exchange was a follow-up to Heiner Stuckenschmidts visit in Hannover from January 15 to Feb 11, 2005. The goal was to combine Semantic Web and peer-to-peer technologies to support inter-organizational information sharing. The aim of the research visit was to bring together complementary expertise in knowledge representation (Amsterdam) and information sharing infrastructures (Hannover), to identify synergies between the two fields and discuss joint work on combining these two technologies in the context of the KnowledgeWeb work plan.

Activities:

- 1) Combine ontology mapping and distributed retrieval algorithms

At L3S an approach to distributed information retrieval based on classification and content has been developed. Nowadays, keyword-based search is already in large-scale use (e.g., search engines such as Google), while full-fledged distributed reasoning on ontologies and metadata is not yet scaleable. Therefore, it seems reasonable to move step-wise from pure full-text search towards ontology-based search. Here, simple document classification based on a taxonomy has been added as search criterion. The algorithm developed at L3S allows to identify relevant documents in a P2P network of interconnected information sources, taking keywords and preferred document classification into account. Currently, it only applies to homogeneous classification contexts, i.e., all connected information sources have to use the same classification. This is unrealistic in a Semantic Web context, where heterogeneous sources are to be interconnected for information sharing. Therefore, the goal is to extend the approach with classification mapping capabilities.

Such capabilities can be provided via an ontology mapping algorithm developed at VUA which creates probabilistic mappings based on several heuristics. This fits very well in a distributed context, because no manual mapping activities are required. During the research visit, details of how to combine both approaches have been worked out. Work on a test bed has started. The document collection used consists of Medline abstracts which are already classified. Using hand-crafted mappings from (Medical ontologies), the existing classification is transformed to others, and the re-classified documents are distributed among simulated information sources. In this test bed, the combined algorithm will be evaluated. The precision and recall for this scenario will be compared to precision and recall for the homogeneous classification case, based on the initial classification.

- 2) Design dimensions of P2P Systems for ontology and (meta-) data integration

The work started in Hannover (see exchange report of Heiner Stuckenschmidts visit) has been concluded and has lead to the publication of the journal article "Heiner Stuckenschmidt, Wolf Siberski and Wolfgang Nejdl. *Combining ontologies and peer-to-peer technologies for inter-organizational knowledge management*. The Learning Organization 12(5), 2005, pp. 480-491". It also has become part of a book chapter: „Heiner Stuckenschmidt, Frank van Harmelen, Wolf Siberski and Steffen Staab. *Peer-to-peer and Semantic Web*. In Steffen Staab and Heiner Stuckenschmidt (Editors). *Semantic Web and Peer-to-Peer*, Springer, 2005.“