Semantic Matching in the SWAP Project

KR & R Group, Vrije Universiteit Amsterdam

1 Semantic Web and P2P

The current state-of-the-art in Knowledge Management solutions still focuses on one or a relatively small number of highly centralized knowledge repositories with ontologies as the conceptual backbone for knowledge brokering. As it turns out, this assumption is very restrictive, because, (i), it creates major bottlenecks and entails significant administrative overheads, especially when it comes to scaling up to large and complex problems; (ii), it does not lend itself to easy maintenance and the dynamic updates often required to reflect changing user needs, dynamic enterprise processes or new market conditions. In contrast Peer-to-Peer computing (P2P) offers the promise of lifting many of these limitations. At the same time, today's P2P solutions support only limited update, search and retrieval functionality, e.g. search in Napster is restricted to string matches involving just two fields: "artist" and "track". These flaws however make current P2P systems unsuitable for knowledge sharing purposes. The SWAP project aims at a P2P based knowledge management system that integrates the advantages of Semantic Web-based knowledge management technology developed in successful IST projects like On-To-Knowledge, KnowNet, or Comma. SWAP aims at benefits of a P2P based system that show just by installing the client software, viz. immediate automatic access to knowledge stored at peers. Of course, explicitly modelled ontologies may increase the benefits brought by any knowledge management solution, because they may improve the accuracy of knowledge access and sharing. SWAP solutions, however, may produce benefits even with near zero investment - in contrast to conventional knowledge management systems that need an extensive and expensive set-up phase. Conventional knowledge management repositories will still appear as just another, powerful peer in the network. Hence, a combined Semantic Web and P2P solution may always outperform the sophisticated, but conventional centralized system.

2 Semantic Matching

Information sharing in semantics-based P2P systems relies on the existence of mappings between the semantic models of different peers. The distributed and dynamic nature of P2P systems makes it unattractive to spend effort on creating fixed mappings manually. Instead, there is a need for automatic or semi-automatic matching algorithms that establish a connection between the semantic models of different peers on the fly. In this context, research in the SWAP project will be focussed on the following question: Are automatic matching methods powerful enough to support P2P information sharing in a practical setting ? In order to answer this question, the following research activities will be carried out by the academic partners in the SWAP project:

Assessment of Matching Methods A number of different approaches for matching semantic models have been proposed. These methods differ in the way they identify correspondences between models including, manual matching, lexical matching, structural matching, semantic matching and similarity-based matching. As a first step we will define a set of benchmark matching problems and compare the results of these different methods wrt. these problems. It is easy to see that one matching approach is not always better than another, but that certain matching approaches are better suited for certain matching problems. Based on the results of the evaluation of different matching approaches we will try to formulate guidelines for selecting a certain matching approach to solve a specific matching problem. For this purpose, we first have to identify relevant characteristics of the matching problem and relevant features of the matching method as well as their interaction.

Optimization of Matching Methods Initial experiments with real life data has shown that many matching algorithms are inherently complex and often fail to scale up to realistic scenarios. If we want to use matching in a practical setting, we have to make sure that the response time is still acceptable for the user. In this context, we will investigate the use of approximate matching methods instead of exact ones. In particular, out goal is to find the right trade-off between the run-time behavior of a matching algorithm and the accuracy of the matching result. For this purpose we will carry out experiments with different approximations of the same matching algorithm and compare the results wrt. matching accuracy and runtime. Based on the experiences gained in this experiments, we plan to develop efficient algorithms for semantic matching and implement them in an optimized way.