SWAP - A Semantics-Based Peer-to-Peer System Demonstration

Marc Ehrig[1], **Peter Haase**[1], **Steffen Staab**[1], **Christoph Tempich**[1]
[1]Institute AIFB, University of Karlsruhe, D-76128 Karlsruhe
{ehrig,haase,staab,tempich}@aifb.uni-karlsruhe.de

1 Introduction

In today's knowledge-based economy, the competitiveness of enterprizes and the quality of work life are directly tied to the ability to effectively create and share knowledge both within and across organizations. The literature about knowledge management distinguishes two main aspects to enable knowledge sharing, *i.e.* organizational and technical aspects. In this paper the SWAP-environment[1] is presented which enables the exchange of knowledge on the basis of a peer-to-peer system [Susarla *et al.*, 2003; Bonifacio *et al.*, 2002], *viz.* it is focusing on technical aspects.

Ontologies Within SWAP ontologies are used for knowledge representation. They have shown to be the right answer to knowledge structuring and modelling by providing a formal conceptualization of a particular domain that is shared by a group of people in an organization [D. OLeary, 1998]. However, knowledge management systems (KMSs) based on centralized ontologies need a long development phase and are difficult to maintain.

Peer-to-Peer From a technological point of view **peer-to-peer** (P2P) solutions are particularly well suited for knowledge sharing, because they make it possible for different participants (organizations, individuals, or departments) to maintain their own knowledge structure while exchanging information. However, today's P2P solutions are extremely limited (they mostly rely on keyword search) and not appropriate for the high requirements of a KMSs.

Emergent Semantics In a P2P setting one cannot assume that each peer uses one single ontology or that an ontology development process has been pursued for each peer. Emergent Semantics [Maedche, 2002] builds on lightweight (e.g. a file structure with files as instances) and/or heavyweight ontologies that different participants have created. It considers the overlap between simple ontology definitions in order to build shared ontologies. As new semantical structures emerge from known structures, knowledge management can occur in a distributed fashion without overhead through central administration.

In the remainder a scenario will illustrate a concrete application of the described system from which requirements can be drawn. In section 3 the architecture of SWAP is introduced, section 4 reviews related work while the last section concludes this paper.

[1]http://swap.semanticweb.org/

2 Scenario and Requirements

In the scenario a virtual organization in the tourism domain is described. The virtual organization comprises public authorities, hotels and event organizers. The public authorities require the number of guests visiting the country to plan for example public transport and waste management. Event organizers can customize their offerings according to the number of visitors and the age. Hotels can publish this information to make the stay more pleasant. Today the exchange of this kind of information is time consuming, unpunctual and error prone, although it is often available in electronic form at every level. However, the different organizations have diverse objectives and therefore use different conceptualizations of their domains.

Requirements The different organizations can be seen as one or many independently operating nodes within a "knowledge" network. A node must be designed to meet the following requirements that arise from the task of sharing information from the external sources with other peers: (1) Integation of multiple sources of information, which can lead to information heterogeneity and inconsistencies. (2) Mostly uniform treatment of internal and external sources. (3) Multiple views on available information. (4) Support for query answering and routing. (5) Distribution of information within the network.

3 The SWAP environment

The SWAP environment is a generic infrastructure which was designed to meet the requirements on a knowledge node. The overall architecture of a SWAP node is described in [Ehrig *et al.*, 2003]. We will now briefly present the individual components.

Storing Knowledge All knowledge of the peer is stored as an RDF(S) model [Brickley *et al.*, 1998] in a knowledge repository. A peer may share knowledge from different sources like file systems, email systems or databases. For each external source a separate ontology is defined which captures the properties of the external source (e.g. for the file system concepts like "file" or "folder" were modelled[2]).

No central domain ontologies are defined. However, a metadata model[3] is used to cope with (1) the knowledge integration task (knowledge items can be traced to its origin), (2) information heterogeneity (two peers name the same

[2]see http://swap.semanticweb.org/2003/01/swap-common
[3]see http://swap.semanticweb.org/2003/01/swap-peer

concepts differently and vice versa), (3) security (Who is allowed to see the knowledge?) and (4) caching (augment network efficiency). This is described in more detail in [Broekstra *et al.*, 2003].

Sharing Knowledge In order to create the knowledge, a set of methods for constructing the repository is provided. Methods have been developed to create repository content mostly automatically from the above mentioned local information sources. The information items like a file remain in the original store. Just the meta information for the file is stored in the repository.

In addition to the possibility to query other peers manually an informer component is provisioned which actively searches for knowledge of other peers and advertises its own knowledge to others.

Seeking Knowledge Finally, the most important aspect for the user is to get answers to specific queries. Queries can be entered by clicking in the views graph or manually as text. The query itself can have various degrees of complexity from simple conjunction to recursion formulated in an RQL-related query language (SeRQL). Eventually, it is sent to the internal inference engine which tries to solve the request. In a future version the inference engine should split the query and distribute the sub-queries in the P2P network if it can not get an answer from the local repository. In this case the query first has to be rewritten in order to fit the underlying knowledge structures on other peers.
The routing is based on metainformation about their knowledge and trust figures. The other peers will answer the queries in the same fashion and finally return answers, which are put together and presented back to the user. He can then decide if he wants to add the answer to his own knowledge representation. Answers consist of statements which can also link to e.g. files. During the communication process information about the network is stored which can later be used for finding better paths for an own query.

4 Related work

Edutella [Ahlborn *et al.*, 2002] also provides a P2P infrastructure for exchanging metadata, but focuses on the education community. Lecturers can publish their notes without losing ownership and students can access them. The InfoQuilt System [Arumugam *et al.*, 2001] provides a framework for formulating complex information requests, involving multiple ontologies, and supporting a form of knowledge discovery. From the local ontologies less quoted ones eventually disappear through evolution. The aim of EDAMOK [Bonifacio *et al.*, 2002] is to develop research in information technology and software tools that support the Distributed and Autonomous Management of Knowledge, not using an ontology premise though.

5 Conclusion

In this paper the SWAP environment was presented which facilitates the sharing of knowledge in a distributed setting. The system allows for integration of knowledge from different knowledge sources and other peers. The knowledge is conceptualized by means of an ontology while different peers not necessarily share the same ontology.

In the demonstration the functionalities of the SWAP environment will be shown. In particular the extraction of lightweight ontologies from existing knowledge sources and the sharing of this knowledge with other peers will be illustrated. Furthermore, the retrieval of files based on the generated metadata and the illustration of the routing mechanism in a simulation environment is part of the demonstration.

The SWAP environment provides the infrastructure for research activities about methods to create and connect ontologies from different peers and the reliability of structures that have been extracted from information sources instead of hand-crafted knowledge structures and metadata.

Acknowledgements. Research reported in this paper has been partially financed by the European Union in the IST project SWAP (IST-2001-34103).

References

[Ahlborn *et al.*, 2002] Benjamin Ahlborn, Wolfgang Nejdl, and Wolf Siberski. OAI-P2P: A peer-to-peer network for open archives. In *Workshop on Distributed Computing Architectures for Digital Libraries - ICPP2002*, 2002.

[Arumugam *et al.*, 2001] Madhan Arumugam, Amit Sheth, and I. Budak Arpinar. Towards peer-to-peer semantic web: A distributed environment for sharing semantic knowledge on the web. Technical report, Large Scale Ditributed Information Systems Lab, University of Georgia, 2001.

[Bonifacio *et al.*, 2002] Matteo Bonifacio, Paolo Bouquet, and Paolo Traverso. Enabling distributed knowledge management: Managerial and technological implications. *Novatica and Informatik/Informatique*, III(1), 2002.

[Brickley *et al.*, 1998] D. Brickley, R. Guha, and A. Layman. Resource description framework (rdf) schema specification. Working draft, W3C, August 1998. http://www.w3c.org/TR/WD-rdf-schema.

[Broekstra *et al.*, 2003] Jeen Broekstra, Marc Ehrig, Peter Haase, Frank van Harmelen, Arjohn Kampman, Marta Sabou, Ronny Siebes, Steffen Staab, Heiner Stuckenschmidt, and Christoph Tempich. A metadata model for semantics-based peer-to-peer systems. In K. Aberer et al., editor, *Semantics in Peer-toPeer and Grid Computing - SemPGRID, Twelfth International World Wide Web Conference (WWW 2003)*, Budapest, May 20 2003.

[D. OLeary, 1998] D. OLeary. Using AI in knowledge management: Knowledge bases and ontologies. *IEEE Intelligent Systems*, 13(3):34–39, May/June 1998.

[Ehrig *et al.*, 2003] Marc Ehrig, Christoph Tempich, Jeen Broekstra, Frank van Harmelen, Marta Sabou, Ronny Siebes, Steffen Staab, and Heiner Stuckenschmidt. SWAP - ontology-based knowledge management with peer-to-peer technology. In Y. Sure and H.-P. Schnurr, editors, *Proceedings of the 1st National "Workshop Ontologie-basiertes Wissensmanagement (WOW2003)"*, 2003. To appear 2003.

[Maedche, 2002] Alexander Maedche. Emergent semantics for ontologies. *IEEE Intelligent Systems*, January/February 2002.

[Susarla *et al.*, 2003] Anjana Susarla, De Liu, and Andrew B. Whinston. *Peer-to-Peer Enterprise Knowledge Management*, chapter 39, pages 129–139. Springer, 2003.