SWAP: Ontology-based Knowledge Management with Peer-to-Peer Technology

[1]Marc Ehrig   [1]Christoph Tempich   [2]Jeen Broekstra
[2]Frank van Harmelen   [2]Marta Sabou   [2]Ronny Siebes   [1]Steffen Staab
[2]Heiner Stuckenschmidt

[1]{meh,cte,sst}@aifb.uni-karlsruhe.de
Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany
[2]{jbroeks, frankh, marta, ronny, heiner}@cs.vu.nl
Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

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. Many enterprizes have therefore spent huge amounts of money to implement centralized knowledge management systems (KMSs).

Ontologies 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 [O'L98]. However, KMSs based on centralized ontologies need a long development phase and are difficult to maintain.

[BBT02] suggest a distributed approach to Knowledge Management which better fits the true situation in organizations and the processes people are used to. From a technological point of view **peer-to-peer** (P2P) solutions are particularly well suited, 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.

The creation of ontologies is still the bottleneck for a fast development. **Emergent Semantics**[Mae02] 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.

The Semantic Web and Peer-to-Peer (SWAP) project[1] demonstrates that taking the successful technologies of the above research areas, will allow support for decentralized envi-

---

[1]funded by EU under contract No. IST-2001-34103; http://swap.semanticweb.org

ronments. Participants can maintain individual knowledge structures on their peers (PCs), while sharing knowledge in ways such that administration efforts are low, but knowledge sharing and finding is easy. This paper gives a general presentation of the project: the baseline technologies, an overview of the system, the case studies, and related work.
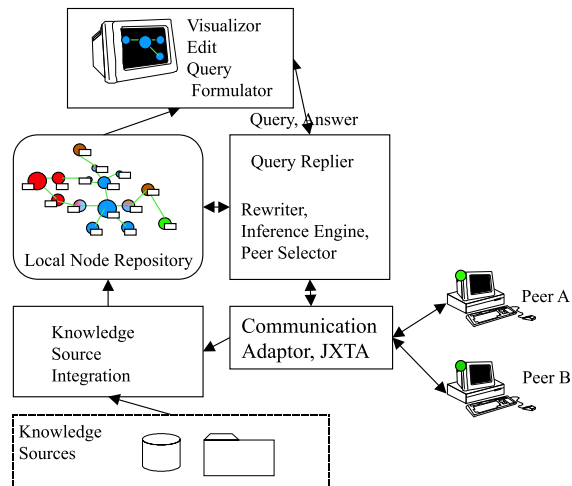
2 The SWAP System

Figure 1: Architecture

The system consists of a set of peers called "SWAP Nodes". The knowledge of a particular peer is extracted from several Knowledge Sources, then integrated and stored in the Local Node Repository LR. A user interface ensures that the user can edit/browse/query the knowledge. Queries that cannot be answered by the available knowledge are sent to the whole system. A specialized component deals with rewriting these queries and selecting the peers which are likely to know the answer. Further, we describe several major issues.

Providing Knowledge: Every participant in the SWAP system has to be enabled to provide his knowledge. In this module the user can select among existing knowledge sources from his personal computer for addition (emails, files, folders, databases, bookmarks, or whole ontologies). A special component will extract ontology-like structures from the selected sources. These sources will then be integrated automatically into the LR, which is represented in RDF(S). Merging routines will be required at this point. Content is therefore stored as an ontology, which allows high-quality processing. Besides the content we also include technical information about each entity i.e. an ID and the location (e.g. peer123//-outlook//abcdef1234). The trust in a specific information is another important measure which is saved and maintained. The process guarantees that even a novice can start quickly.

Updates and Changes: The LR is an integrated view on all known information sources, both local sources and other peers. Any change in the knowledge sources is propagated to the LR. The edit functionality is thought to be useful for advanced users and knowledge engineers. Those who are familiar with ontologies and their graphical structures can

intervene and append their own structures by adding, deleting, or renaming concepts and relations between the concepts. Changes to the LR due to editing are currently not propagated to the underlying structures, since they can provoke undesirable contradictions. Changes will also occur when answers from other peers provide information return. This information is incorporated to get an overview of the knowledge in the P2P network.

Views: A view is a pre-defined perspective of the knowledge. It can be generated out of selected structures (e.g. only email folders, peer123 perspective) or the entire LR. They are implemented using different visualization techniques (topic hierarchies, thematic maps) [FSvH02]. The user can easily browse through this graphical knowledge representation.

Query generation, distribution, and answering: 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. Eventually, it is sent to the internal inference engine which tries to solve the request. If the inference engine can not get an answer from the local repository, it splits the query and distributes the sub-queries in the P2P network. 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. While the peer communicates with others, information about the network is stored which can then be used for finding better paths when having an own query.

File sharing: If the answer received from the SWAP system is a file title or an email subject the user might like to get the entire file from the answering peer. The request is sent back to the responding peer with all meta information from the answer and the peer solves this file request with a separate module.

3   Case studies

First, in the Knowledge Management Case Study we are going to investigate the work of investment analysts at **Dresdner Bank**, one of the largest German banks. The environment for this case study is fairly closed in terms of topics and technology, and comprises only a defined and limited number of participants. The case study surveys their use of current central knowledge management technology. The analysts store big amounts of knowledge in various formats on their personal computers. The SWAP P2P based solution will be installed and test-driven in order to come up with a realistic estimation of the work that might be saved (or wasted) by the P2P system.

Second, in the virtual enterprize case study, contractor **IBIT** (Fundació de les Illes Balears per a la Innovació Tecnológica) will use P2P technology to enable knowledge sharing between different small and medium-sized enterprizes that all have a stake in sustainable tourism. This topic is of high importance to the local economy of the Balearic Islands, and it involves a large number of very different peers (local government, tour operators, hotel operators, university researches etc.), and involves very heterogenous knowledge sources. Also, the peers are geographically dispersed, they are very different in nature, and they

will be intermittently connected to the network.

4 Related Work and Summary

Edutella[ANS02] 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 easily access them. The InfoQuilt System[ASA01] 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. A set of InfoSleuth[FPNB99] agents collaborate at a semantic level to execute information gathering and analysis tasks, where the underlying information sources can be diverse both in their structure and content, as in SWAP. LARKS[SWKL02] is an agent capability description language that enables advertising, requesting and matching over the Internet. The aim of EDAMOK[BBT02] 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.

SWAP however has a different approach. The combination of Semantic Web and P2P is highly innovative with prospective benefits to the individualization of work views as well as to the facilitation of knowledge sharing between peers. SWAP takes on the challenges brought up by this novel combination of ontologies and P2P computing such that knowledge finding and sharing is effectively possible. It also considers how Emergent Semantics knowledge is constructed from the distributed environment. The technology is implemented into a system of which the architecture has been shown. The case studies from different companies for this specific problem define the application.

References

[ANS02]   B. Ahlborn, W. Nejdl, and W. Siberski. OAI-P2P: A Peer-to-Peer Network for Open Archives. In *Workshop on Distributed Computing Architectures for Digital Libraries - ICPP2002*, 2002.

[ASA01]   M. Arumugam, A. Sheth, and I. B. 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.

[BBT02]   M. Bonifacio, P. Bouquet, and P. Traverso. Enabling Distributed Knowledge Management: Managerial and Technological Implications. *Novatica and Informatik/Informatique*, III(1), 2002.

[FPNB99]   J. Fowler, B. Perry, M. H. Nodine, and B. Bargmeyer. Agent-Based Semantic Interoperability in InfoSleuth. *SIGMOD Record*, 28(1):60–67, 1999.

[FSvH02]   C. Fluit, M. Sabou, and F. van Harmelen. Ontology-based Information Visualisation. In Vladimir Geroimenko, editor, *Visualising the Semantic Web*. Springer Verlag, 2002.

[Mae02]   A. Maedche. Emergent Semantics for Ontologies. *IEEE Intelligent Systems*, 2002.

[O'L98]   D. O'Leary. Using AI in knowledge management: Knowledge bases and ontologies. *IEEE Intelligent Systems*, May 1998.

[SWKL02]   K. Sycara, S. Widoff, M. Klusch, and J. Lu. *LARKS: Dynamic Matchmaking Among Heterogeneous Software Agents in Cyberspace*, chapter 5. Kluwer, 2002.