Proceedings of I-KNOW '05
Graz, Austria, June 29 - July 1, 2005

Intelligent Community Lifecycle Support

Stephan Bloehdorn*, Peter Haase*, Mark Hefke, York Sure* and Christoph Tempich***
*Institute AIFB, University of Karlsruhe, D-76128 Karlsruhe, Germany
{bloehdorn,haase,sure,tempich}@aifb.uni-karlsruhe.de
**Forschungszentrum Informatik, D-76128 Karlsruhe, Germany
{hefke}@fzi.de

Abstract: Knowledge sharing in *communities* has attracted much attention in the field of knowledge management in research and practice. In this paper we outline a view where the community lifecycle is supported at different stages. The central component of our framework is the community ontology SWRC+COIN that describes the typical structure of communities. We exemplarily show how communities in the academic domain can be detected automatically by means of analyzing information flow in a bibliographic Peer-to-Peer system and how the instantiated community knowledge base can be exploited to support cooperative work in the communities.

Key Words: Communities of Practice, Knowledge Management, Peer-to-Peer, Ontologies, Metadata, Semantic Portals
Category: H.3.3,H.5.3

1 Introduction

Knowledge sharing in *communities* has attracted much attention in the field of knowledge management in research and practice [Davenport and Prusak, 2000, Wenger and Snyder, 2000]. Communities are groups within or across organizations who share a common set of information needs or problems. Communities help participating members and their organizations to reduce transactional costs for knowledge acquisition, to foster organizational learning and innovation, and to create common strategic visions. Although communities are inherently fuzzy, informal and self-organizing, practice shows that communities typically require explicit centralized organizational and technical measures that accompany their initialization and operation. Specifically, this *"Paradox of Management"* [Wenger and Snyder, 2000] asks for supporting activities in the two phases of development:

1. **Community Identification:** identifying potential communities and initiating their formal implementation,
2. **Community Infrastructure:** providing an environment that fosters community operation and evolution,

Research on communities has directed a lot of attention on organizational aspects of community development while *information technology* has typically played the role of the junior partner with a limited focus.

In this paper we outline a view where the community lifecycle is supported at different stages by means of information technology: (1) As a central component of our framework we describe the community ontology SWRC+COIN, that models the key concepts by means of which communities, their members and internal structures can be described; (2) we show how communities can be detected by analyzing the communication patterns in a semantic Peer-to-Peer system using a matching conceptual schema and (3) how the instantiated community knowledge base can be used to support running community activities.

The remainder of this paper is organized as follows. As a motivation, we describe a usecase from the academic domain in Section 2. We present the community ontology SWRC+COIN in detail in Section 3. Section 4 exemplarily shows the detection and knowledge base instantiation step based on the bibliographic Peer-to-Peer system Bibster. We outline potential applications of the resulting community knowledge base in different settings in Section 4. After reviewing related work we conclude in Section 6.

2 Application Scenario: Supporting Academic Communities with Bibster

The following use case exemplifies the application scenarios for the presented infrastructure in an academic domain. One of the main objectives of researchers is to advance on the state-of-the-art in there respective research areas and to publish their advances. The analysis of existing work and referencing it is therefore a researchers affairs. Collecting complete bibliographic entries for related work, however, can be a time-consuming task in particular for recent research. Furthermore, it is often difficult for new entrants to identify the most senior researchers in a given field, or to pick the most comprehensive work describing the current state of affairs.

The award-winning semantics-based Peer-to-Peer application Bibster[1] tries to overcome these limitations. It aims at researchers who want to benefit from sharing bibliographic metadata [Haase et al., 2004]. Bibster allows users to share bibliographic metadata among peers in an infrastructure that is built on exploiting the semantic descriptions of the shared knowledge items, targeting the first challenge. The community ontology presented in section 3 models the concepts to target the second challenge. Analyzing the shared bibliographic entries, user queries and their respective results, we are able to populate the community ontology with information about the peers' expertise, interests, and the flow

[1] `http://bibster.semanticweb.org/`

of information in the network. This can be done in a completely decentralized manner based only on the local knowledge of the peers.

In our case study we were, however, confronted with the typical cold start problem of decentralized systems, i.e. the challenge to motivate users to contribute content in the early stage. Therefore, the infrastructure presented in this paper combines the advantages of both solutions.

A centralized portal based on the VICO infrastructure provides access to the more established publications and is an entry point for the Peer-to-Peer system. The latest publications are available via the Peer-to-Peer network. Communities are detected automatically from data gathered from the Peer-to-Peer system. The central system collects this data from the participating peers in regular time intervals. With intelligent analyzing techniques communities can be identified. For each community a new instance of the community portal is created using only the information relevant for the specific community. In the portal we thus present information about researchers working in similar areas, emerging topics of interests in the community, etc.

3 The SWRC+COIN Community Ontology

The central component of our framework is the SWRC+COIN community ontology, which describes the structure and key entities of communities and their corresponding interrelations. Specific communities instantiate and may also extend this common ontology locally. In general, the community ontology has to address the following purposes: (1) It acts as the backbone for integrating information from different information sources (e.g. different peers) as well as their mutual relations, (2) it acts as a reference for communication, e.g. about roles in the community, during complementary organizational activities, (3) it allows to issue complex queries on the community knowledge base.

We have based our community ontology on the SWRC ontology[2] which - although initially phrased *Semantic Web Research Community Ontology* - generically models key entities in academia, e.g. people, organizations, projects and research topics. Consequently, SWRC+COIN extends SWRC by additionally modelling community specific extensions.

Figure 1 shows a view on the main entities of the SWRC+COIN ontology, which we briefly explain in the following: The *Community* concept is the central entity of the SWRC+COIN ontology. The community is formed by its people, represented by the concept *Person*. In addition to the direct membership relation (*hasMember*), people can play roles in communities, expressed via the *CommunityRole* concept. Possible roles include *CommunityFounder*, *CommunityManager*

[2] Both, SWRC and SWRC+COIN ontologies are available at `http://ontoware.org/projects/swrc/`.

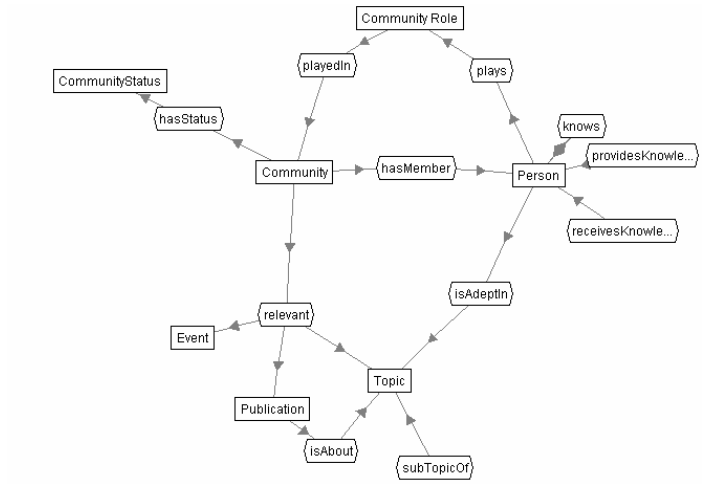

Figure 1: A view on the main entities of the SWRC+COIN ontology

and *CommunityExpert*. With the *CommunityStatus* we are able to capture the status of the community in its life cycle, i.e. whether it is currently in formation, in operation, or closed. Further we model *Events*, *Publications* and *Topics* that are *relevant* to the community. A final important aspect is the knowledge flow in the community, which is modelled with the relations *receivesKnowledgeFrom* and *providesKnowledgeTo*.

All the conceptualizations made in the ontology are carefully designed to be extendible locally by the community itself. For example, different communities might like to introduce new specific roles, which can be done easily by subclassing the *CommunityRole* concept.

4 Discovering Communities in Peer-to-Peer Communication

Successful communities emerge from existing interaction of their prospective members in a social network. Peer-to-Peer networks naturally embody this paradigm: *peers* in the network correspond to prospective community members and interact by sharing *knowledge items* like data records or unstructured textual information. We argue that the main ingredients of communities – their domain and members – emerge naturally from analyzing these communication patterns.

During the *Community Identification Phase* our framework therefore aims at analyzing the communication patterns in the Peer-to-Peer network. By differen-

tiating the flow of knowledge items in the network with regard to their content, e.g. by filtering metadata information or by clustering textual knowledge items, different *thematic* network topologies emerge. By applying techniques from network analysis [Girvan and Newman, 2002] on these topologies or clustering users based on their communication profiles [Hartigan, 1975], the system is able to detect community structures in the network.

We have operationalized a simple variant of this idea in the context of the application scenario described in section 2. We have collected queries issued by peers in the BIBSTER network and have constructed query profiles of the individual peers. In our experiments, the peer profiles consisted of vectors that described the distribution of queries of the respective peers over a set of topics defined in the ACM topic hierarchy. Peers were clustered into groups based on these profiles.

In our experiments we used the well known k-means cluster algorithm with the cluster parameter set to 10 clusters. While a strict evaluation of clustering methods was not the focus of our work at this stage, results appear to be promising. While typically 75 % of the peers formed one large cluster, the remaining 9 clusters were formed by small groups of peers with 1 to 15 members. These clusters which we evaluated qualitatively by comparing the work of the persons corresponding to the individual BIBSTER peers showed a high level of relatedness to certain topics. A perfect example was a cluster of 9 peers corresponding to researchers in the area of Peer-to-Peer and distributed databases.

Note that while the output of such an analysis – an initial set of community members – is already a valuable result in itself, one could go beyond this step. By exploiting metadata information, the analysis results can be used to populate the community ontology with instance records covering both, the peers and the knowledge items shared together with information on their usage by peers in the network. The result of such an analysis could also include information on prospective *roles* of the community members based on the degree of the individual commitment of peers and their connectedness.

5 Supporting the Community Infrastructure through the Community Ontology

The resulting community ontology from the discovery phase can now be integrated in further software modules of the *Community Infrastructure* in the second phase. This includes e.g. its exploitation for specific methods and algorithms on the technical level (e.g. query routing in a Peer-to-Peer network) as well as the support for organizational and methodological tasks in the community.

For the latter, we show in the following, how centralized access to the knowledge generated within the Peer-to-Peer network is provided by the semantic portal VICO (Virtual Community Portal), a community infrastructure building on

KAON Portal [Maedche and Staab, 2003]. KAON portal is a standalone framework for generating ontology-based web portals. By providing the visualization of multilingual ontologies, KAON Portal supplies easy ontology navigation and searching by the use of a standard web browser. Through embedding the discovered community ontology, the VICO-Portal presents community knowledge to the members by providing a centralized reference point. Beyond providing the above-named searching and browsing functionalities, the VICO-Portal also offers a web-based query mechanism as well as matching component in order to e.g. locate experts in a specific research topic or to identify people with similar interests. We outline these functionalities in the following.

Querying the Portal. A user is able to define queries to the portal based on the underlying community ontology. A template-based query interface is created directly from the ontology. A typical query is shown in Figure 2: The user is interested in people adept in the topic "Semantic Web". He therefore (1) selects the relevant concept ("Person") from a list of all available concepts in the ontology, (2) selects the relevant property (i.e. "is adept in topic") from the set of relevant properties for this concept, and (3) selects an instance of the target concept (i.e. "Semantic Web" as instance of the concept "Topic"). As a query result, the portal returns a list of all instances that satisfy the constraints of the query.

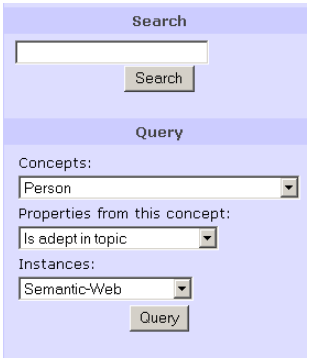

Figure 2: Query Example

Ontology-based Matching. In addition to the regular search functionality, the portal further supports ontology-based matching based on a notion of *semantic distance*. For example, in order to support the identification of (expert) people with similar interests and finally bring them together within the community, the VICO-Portal provides means for matching instances by incorporating similarity

measures that determine how closely these instances are related. These similarity measures consider various features of the instances, e.g. in the case of persons the similarity in expertise, their relationship with other persons, etc. Further, the similarity measures take background knowledge provided by the ontology (such as a topic hierarchy) into account. The final similarity, which is an aggregation of the similarities for the individual features, is used for the ranking of the matches. Figure 3 depicts results of an expert matching.

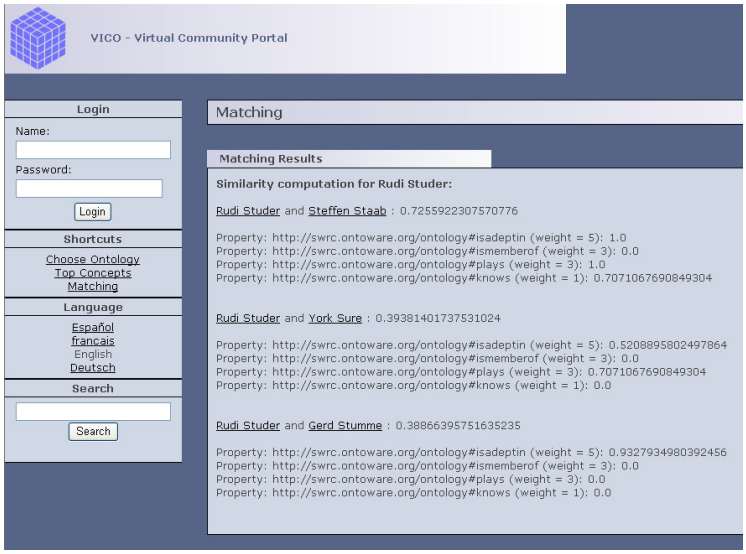

Figure 3: Matching Result

6 Conclusion and Outlook

Few endeavours have been made so far to support the community lifecycle by means of semantic technologies. Recent endeavours in looking at the semantic web from an social network analysis standpoint are reported in [Mika, 2004]. In [Alani et al., 2003], the authors describe a tool for identifying potential communities based solely on the knowledge base and its corresponding ontology by analyzing relations between instances. An ontology-based knowledge sharing system is described [Davies et al., 2004] which also serves to support WWW-based communities of practice.

In this paper we have outlined a framework for supporting the community lifecycle by means of information technology in a Peer-to-Peer environment.

We have introduced the community ontology SWRC+COIN which serves as a conceptual backbone for describing community structures. The framework we describe consisted of two stages: during the *Community Identification Phase* we have described a setting for identifying potential communities by analyzing communication patterns in a Peer-to-Peer network, and have sketched initial experiments based on the BIBSTER application scenario. For the second stage, namely the support of initiated communities through an intelligent *Community Infrastructure*, we have presented VICO, a community portal based on the Karlsruhe Ontology Management System KAON.

Future work in this direction will focus on experiments for community detection on a larger scale accompanied by qualitative *and* quantitative evaluations. Correspondingly, we aim at extending the VICO portal infrastructure through additional components.

Acknowledgements

This research was partially supported by the European Commission under contract IST-2003-506826 SEKT. The expressed content is the view of the authors but not necessarily the view of the SEKT project as a whole.

References

[Alani et al., 2003] Alani, H., Dasmahapatra, S., O'Hara, K., and Shadbolt, N. (2003). Identifying Communities of Practice through Ontology Network Analysis. *IEEE Intelligent Systems*, 18(2):18–25.

[Davenport and Prusak, 2000] Davenport, T. H. and Prusak, L. (2000). *Working Knowledge - How organisations manage what they know.* Harvard Business School Press, Boston, MA, US.

[Davies et al., 2004] Davies, J., Duke, A., and Sure, Y. (2004). OntoShare - An Ontology-based Knowledge Sharing System for virtual Communities of Practice. *Journal of Universal Computer Science*, 10(3):262–283.

[Girvan and Newman, 2002] Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 99(12):7821–7826.

[Haase et al., 2004] Haase, P., Broekstra, J., Ehrig, M., Menken, M., Mika, P., Plechawski, M., Pyszlak, P., Schnizler, B., Siebes, R., Staab, S., and Tempich, C. (2004). Bibster - A Semantics-Based Bibliographic Peer-to-Peer System. In *Proceedings of the Third International Semantic Web Conference, Hiroshima, Japan, 2004.*

[Hartigan, 1975] Hartigan, J. (1975). *Clustering Algorithms.* John Wiley and Sons, New York, NY, US.

[Maedche and Staab, 2003] Maedche, A. and Staab, S. (2003). KAON - The Karlsruhe Ontology and Semantic Web Meta Project. *Künstliche Intelligenz (KI)*, 17(3):27–31.

[Mika, 2004] Mika, P. (2004). Social Networks and the Semantic Web. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI 2004), 20-24 September 2004, Beijing, China*, pages 285–291.

[Wenger and Snyder, 2000] Wenger, E. and Snyder, W. (2000). Communities of Practice: the Organizational Frontier. *Harvard Business Review*, pages 139–145.