KNOWLEDGE SHARING OVER P2P KNOWLEDGE NETWORKS: A PEER ONTOLOGY AND SEMANTIC OVERLAY DRIVEN APPROACH

SYED SIBTE RAZA ABIDI[†]
Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia, B3H 1W5, Canada

XIAOLIN PANG
Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia, B3H 1W5, Canada

Effective knowledge sharing between communities of knowledge workers impacts the productivity, innovation and competitiveness of organizations. The emergence of Peer-to-peer (P2P) technology has provided some solutions to knowledge sharing amongst peers. In this paper we present a sophisticated an *Agent-based Semantic Knowledge network*—ASKnet—that features a unique combination of mobile agents, semantic web and semantic overlay networks to realize an effective knowledge sharing environment. ASKnet enables the sharing of personalized knowledge repositories—typically documents thematically organized by a knowledge worker—of individual knowledge workers with the entire community. The premise here is that if we trust the intellectual ability of the knowledge worker then his/her personalized knowledge repository can be regarded as a source of *best-quality*, *relevant* and *validated* knowledge for that domain/organization.

1 Introduction

The prevailing 'knowledge age' or 'knowledge economy' places a premium on the collective knowledge owned and managed by organizations. Knowledge is perceived as a commodity and its flow across the organization—in order to effectuate innovation, competitive advantage, organizational learning and improved productivity—is deemed as an important factor in the sustainability of the so-called knowledge economy (Zack 1999). In fact, Nonaka argues that "in an economy where the only certainty is uncertainty, the one sure source of lasting competitive advantage is knowledge" (Nonaka 1998). For that matter, organizations nowadays are formulating well-defined knowledge management policies that extend beyond the traditional knowledge management activities involving the acquisition and storage of in-house knowledge. The current knowledge management themes focus on the pragmatic effects of knowledge sharing and re-use by knowledge workers (Liebowitz 2000). Our research investigates the formulation of technology-mediated knowledge management solutions for effective knowledge sharing within a community of knowledge workers.

Knowledge sharing entails both knowledge creation and knowledge reuse; in fact these two activities are not orthogonal, as new knowledge builds on the re-use of existing knowledge. Knowledge sharing involves three main activities: (i) location of relevant

[†] Work supported by the Canadian NSERC grant # 262072 - 03

explicit knowledge; (ii) selection of relevant/significant knowledge; and (iii) application of the knowledge in a particular context.

From a knowledge sharing perspective, the notion of a community of practice (Wenger and Snyder 2000), or as we say *community of knowledge workers*, is of significant relevance. In principle, community of practice is an informal group of workers that are bound together by a shared interest, knowledge or enterprise—essentially peers involved in the execution of a common objective. Advances in communication technology has allowed the members of community of practice to be geographically dispersed yet virtually accessible—Internet mediated discussion forums are a case in point. Through collaboration a community of practice not only fosters knowledge sharing and community-wide learning, but it also leads to the 'creation' of new knowledge and 'validation' of existing knowledge/viewpoints/practices/beliefs. Hence, it is fair to state that a community of knowledge workers in itself is a knowledge resource.

In a knowledge sharing parlance, our work focuses on enabling the sharing of private knowledge repositories—i.e. the private collection of specialized documents thematically organized by a knowledge worker—of individual knowledge workers within the entire community. The premise here is that knowledge workers whilst discharging their duties progressively collect and maintain a private knowledge repository that comprises knowledge-rich documents (i.e. research papers, reports, electronic books, notes, guidelines, learning modules, presentations etc) spanning multiple subjects of interest and collected from respectable sources. If we trust the intellectual ability of the knowledge worker then his/her private knowledge repository can be regarded as a source of *high-quality*, *specialized* and *validated* knowledge. Given the escalation in the volume of dispensable knowledge and the proliferation of knowledge dissemination web-sites there is always a question mark on the quality and validity of the available knowledge content. Furthermore there are noted problems of identification, selection and evaluation of the available knowledge (typically over the WWW). In this scenario, we argue that leveraging the private knowledge repository of a 'trusted' knowledge worker's (provided the said knowledge worker allows access to it) can alleviate the problem of knowledge identification, selection and validation; in fact it could serve as the initial point of search for high-quality knowledge. We foresee that an interconnected and ubiquitous collection of private knowledge repositories of a community of knowledge workers can be regarded as *validated* domain knowledge—akin to the intellectual capital of an organization (Liebowitz 2000). We believe that a network of private knowledge repositories can not only serve as a meta-knowledge resource but also act as a vital collaborative learning and innovation medium for the entire community of knowledge workers.

In this paper we focus on the networking aspect of knowledge sharing within a community of knowledge workers. We present a knowledge sharing framework that interconnects the private knowledge repositories of participating knowledge workers using peer-to-peer (P2P) communication technology. P2P technology is being used for sharing digital resources and services between participating peers—i.e. computers, wireless devices and PDAs. However, the existing P2P based knowledge sharing initiatives exhibit certain limitations, such as (a) coarse granularity of knowledge sharing (i.e. only file-sharing) is provided without considering the content of the file; (b) Peer's neighbors are pre-defined and cannot be changed dynamically according to the relativity of shared knowledge on different nodes; (c) no support for semantic interoperability between the peer's indexing of documents; and (d) inability to scale-up a to larger set of documents. In our P2P knowledge sharing solution we:

- Incorporate notions of semantic web with the P2P knowledge network in an attempt to overcome the heterogeneity present in the different private knowledge resources available within the knowledge network. The use of two ontologies—a domain ontology defining the domain and a peer ontology defining the characteristics of peers—allows the conceptualization of both knowledge content and peers. It may be noted that current knowledge sharing over P2P networks do not adequately account for semantic relationships amongst resources. In our case, we use the semantic web technology to overcome this limitation and in turn provide content-based knowledge search and a mechanism for establishing agreements between knowledge searching mobile agents.
- Integrate mobile agent technology within the P2P based knowledge network to facilitate the automatic search for relevant knowledge across the knowledge network. Our mobile agents are designed to perform various knowledge search and collection tasks on behalf of users (Wang et al 2001).
- Formulate a *Semantic Addressable Overlay* network to support the automatic identification of relevant peers (across the P2P knowledge network) based on their profiles and on the content of their knowledge repositories.

Figure 1 illustrates a P2P knowledge network from a peer's point of view.

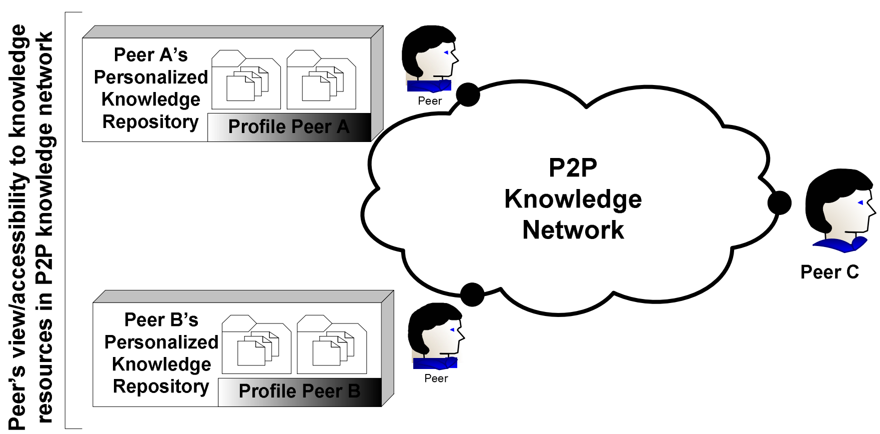

Figure 1. A P2P based knowledge network

2 Background

P2P technology is rapidly emerging as a potential solution for sharing of digital content over the Internet. P2P systems can be classified into two distinct categories according to their architectures. Category 1 is the “pure” P2P architecture (as shown in Figure 1), which consists of peers that have equal responsibility and capability, and may join or leave the network at any time, thus eliminating the need for a centralized server. Peers dynamically discover other peers on the network and interact with each other by sending and receiving messages. Category 2 is the “server-mediated” P2P architecture—also termed as hybrid systems. In this architecture, the central server is responsible for peer discovery and information search as it maintains a registry of shared information about peers. For instance, if peer A is looking for a file it sends a request to the central server. The central server replies to peer A with a set of peers who possess this requested file. Subsequently, Peer A directly interacts with the candidate peers to retrieve the file. In short, P2P technology realizes a de-centralized computer network architecture between ‘willing’ computers.

With regards to knowledge sharing, a general framework for ontology-based knowledge sharing in P2P systems, called *Helios*, was proposed by Castano et al (2003a). Most significantly, Helios features a peer ontology (Castono et al 2003b) that characterizes the different types and attributes of peers sharing knowledge. Interactions amongst peers, knowledge search and knowledge acquisition/extension are supported by pre-defined query models and semantic techniques for ontology matching.

Subsequently, by combing a communication infrastructure layer (called Hermes) with Helios’s knowledge infrastructure layer an enhanced knowledge sharing framework, called H3, was presented by Castano et al (2003c). H3 proposes to build an overlay network among peers in which each peer maintains a peer ontology describing its knowledge of the network. For query routing, the topology of the overlay network mirrors the semantic neighborhood of the peers given by the semantic relationships among the ontologies they own. This approach allows peers to dynamically join communities of interest and to share their knowledge without regard to physical network neighborhood constraints imposed by P2P networks—in this case two peers are neighbors if they possess similar or equivalent semantic concepts (Castano 2003d; Yang and Garcia-Molina 2003).

Nejdl et al (2002) introduced Super-Peer Networks to perform effective content retrieval. In Super-Peer Networks, peers are clustered with respect to their interests, and for each cluster a Super-Peer node is designated, acting as a centralized server for queries in a cluster. Moreover, Super Peers are also connected to each other to create an overlay network.

3 Knowledge Sharing over P2P Networks

To facilitate knowledge sharing amongst a community of knowledge workers we present the design of an *Agent-based Semantic Knowledge network* (or ASKnet) that creates and manages a P2P knowledge network for sharing the private knowledge repositories of a community of knowledge workers. ASKnet is an extension to existing P2P knowledge networks in that it features a novel *Semantic Overlay Network* (or SON) to characterize the semantic relationships between the knowledge available within the P2P knowledge network as shown in figure 2. Each peer node in ASKnet consists of several components: agent management, query management, knowledge management (comprising ontology management and matching management), discovery management (including peer locating and peer routing management), as shown in figure 2.

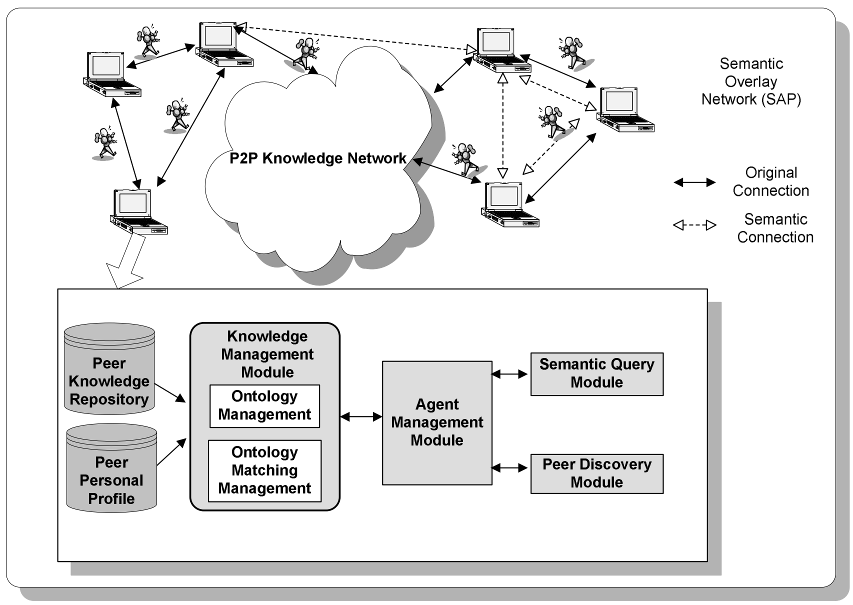

Figure 3: Functional architecture of ASKnet

We discuss ASKnet from two different views—i.e. internal and external view. The internal view refers each peer as an individual and independent unit. In this case, the corresponding knowledge management module is discussed. The external view considers peer's external operations (i.e. the interaction among peers) within the P2P knowledge network. For external view, the function of discovery management, query management, agent management will be discussed.

3.1 *Peer Ontology*

For knowledge sharing amongst peers we provide a peer ontology that semantically describes each peer and his/her knowledge contents. The peer ontology comprises two components: (1) a *peer personal profile* that characterizes each peer on ASKnet and (2) a *peer knowledge repository profile* that describes the content of each peer's personalized knowledge repository. Figure 3 shows a template for the peer personal profile.

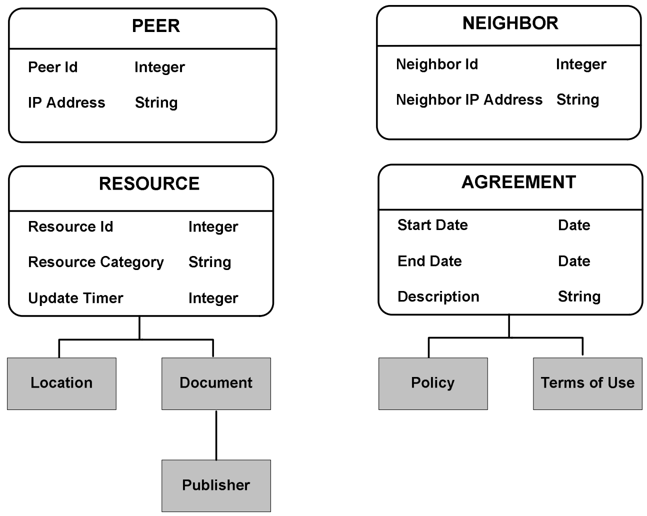

Figure 3. Template of a peer personal profile

The provision of a peer ontology enables peer-to-peer interoperability at different levels of abstraction in terms of both peer matching and subsequent knowledge sharing (Noy and McGuinness 2001). In practice, for each peer we create a peer defining ontology (using a common peer ontology template) that is made available to all other peers as an external view. We argue that by using a peer ontology we achieve the following advantages: (i) ability to share common understanding of the structure of knowledge amongst both peers and software agents; (ii) re-use of domain knowledge; (iii) making domain assumption explicit; (iv) separation of domain knowledge from the operational knowledge; and (v) clear characterization of domain knowledge.

3.2 *Peer-Specific Knowledge Management*

Within ASKnet the knowledge management activities associated with each peer mainly constitutes ontology management and ontology matching management.

Ontology Management Module performs the definitions and maintenance of the peer ontology. This module presents each peer's profile for an external view; provides its resource for sharing; maintains a list of neighbors; manages access constraints such as access and privacy policy.

Ontology Matching Management Module involves the comparison of a concept within a knowledge query against the peer ontology of other peers in order to match semantically related concepts within another peer's personal profile.

Ontology matching is achieved by a variety of methods that are classified along the lines of instance vs. schema; element vs. structure matching; language vs. constraint; matching cardinality; auxiliary information and so on (Erhard and Bernstein 2001). We note that existing ontology matching methods do not support instance-level matching. In our solution, ontology matching is achieved via both structure level and instance level matching.

3.3 *Semantic Addressable Overlay Network (SON)*

We have developed a *Semantic Addressable Overlay Network* (SON) to support semantically-driven knowledge search across ASKnet. SAP characterizes the peer neighborhood in a semantic way to ensure that peers with similar interest are grouped together according to their personal profiles.

Our approach builds on the notion of Content Addressable Network (CAN) proposed by Ratnasamy et al (2001) that dynamically partitions a virtual d-dimensional Cartesian space into zones and assigns each zone to a peer node. In a CAN the space is used for logically representing the index of shared documents (points). Routing is conducted from one zone to another in the Cartesian space. When a node joins the network, it randomly selects a point in the Cartesian space and migrates to the zone that contains the point, and splits the zone with its current peers.

We have developed and incorporated a SON that models a virtual d-dimensional Cartesian space to logically store the index of each peer's knowledge, where each index is positioned as a point in the space. The indices of peer knowledge are created by a hash function, which is denoted as (key, value) pairs, and each node stores a chunk (called a zone) of the entire hash table. In addition, a node keeps information about a small number of "adjacent" zones in the table. The semantic vector of each peer's ontology is generated as the key. To store a pair (K, V), key K is deterministically mapped onto a point P in the coordinate space. The corresponding (key,value) pair is then stored at the node that owns the zone within which the point P lies. To retrieve an entry corresponding to key K, any node can apply the same deterministic hash function to map K onto point P and then retrieve the corresponding value from the point P. If the point P is not owned by the requesting node or its immediate neighbors, the request is routed from one zone to another until it reaches the node in whose zone P lies.

Functionally speaking, SON mirrors the semantic neighborhood of the peers by providing the semantic relationships among the ontologies they own. SAP creates indices for peer ontology description and distributes them to different peer, which is replicated by several of its neighbors in case of failure by one peer. The main advantage of SON is that it provides the peer neighborhood in a semantic way, which means peers with similar interest are grouped together according to their indices. Furthermore, SAP ensures that isolated peer nodes do not exist and different communities are 'somehow' connected to support atypical queries.

In conventional P2P scenarios, each peer's neighborhood is statically predefined and doesn't change during run time. Whereas, two nodes are treated as neighbors if they possess similar peer knowledge in our approach. And each node's neighbor list can be changed dynamically according to the semantic measuring among peers.

Figure 4a depicts a peer’s neighbors in a conventional way. Peer M is a node that request for a certain information and peer M initially has three directly connected peers - Peers A, C and N. However, only peer B and peer D contain information that match peer M's current query. Peer M will forward its request to peer A and peer C to reach peer B and D. Whereas, in our approach, Peer M can send request directly to peer B and D since they have been as its neighbours according to our established semantic overlay network, which is shown in Figure 4b.

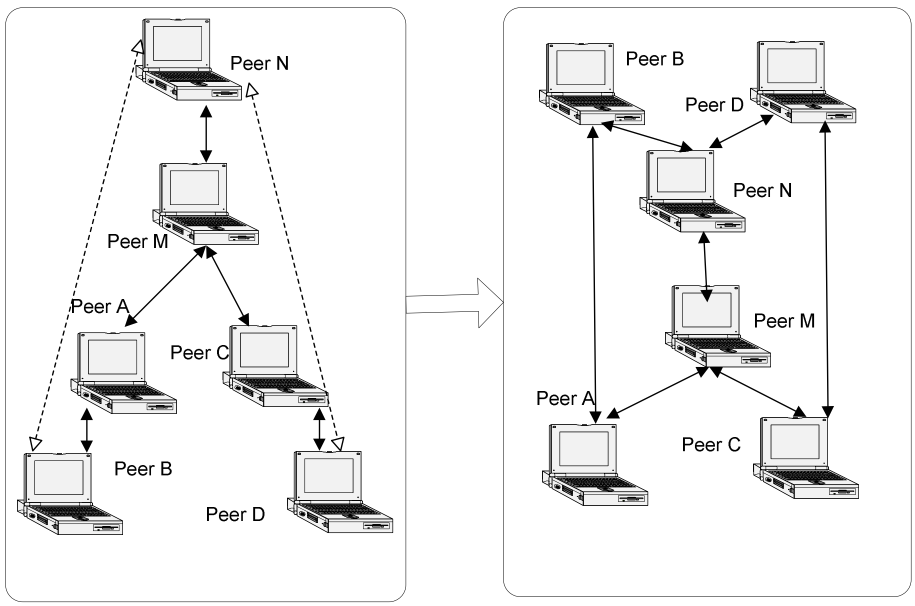

Figure 5a: Conventional Peer’s neighbors

Figure 5b: Peer’s semantic neighbors

The benefits of SON construction are truly reflected in the functionalities of the agent, peer discovery and query management modules that are discussed below.

3.4 *Agent Management Module*

The *agent management module* is responsible for the generation and dispatching of mobile agents to retrieve knowledge from their peer’s private knowledge repositories. When a peer searches for a knowledge item (say a research paper), a pool of mobile agents on behalf of the asking peer are dispatched to its neighbors—the neighborhood of the asking peer is determined by SON. It may be noted that, in the absence of SON, the agents may need to search over the entire knowledge network. In our approach, the SON keeps an index of each peer’s ontology description which helps in focusing the search in a smaller, yet potentially most relevant, region of the knowledge network. Each agent performs semantic matching by comparing with the other peers’ ontology to find an accurate match. After the mobile agent finishes the matching task, it takes back the desired results to the asking peer for further action.

3.5 *Peer Discovery Module*

A P2P network can be evaluated by the quality of peer discovery, which refers to peers that each node of the network selects (Kaplan and Duchon 1998).In semantic-based P2P applications, peer discovery are more desired on the conceptual approximation between one peer with other peers in the network. In conventional P2P networks, peer discovery is conducted in a centralized manner by the use of a central index server or the request of peer discovery is flooded from peer neighbors to neighbors (Kaplan and Duchon 1998). Obviously, these approaches are not scalable and need to be improved.

Our peer discovery approach seeks semantic routing/search (as human do), by using the SON, by implementing a decentralized non-flooding P2P knowledge discovery model: no central index server is provided and peers are grouped with respect to their community of interest based on the SON. Note that SON reflects the similarity approximation between different peers over ASKnet. In practice, when a knowledge-item is needed, the request is initially routed to a group of decentralized peers possessing similar interest as the asking peer, as opposed to broadcasting the request from one peer to another.

3.6 *Semantic Query Module*

The *semantic query module* is used by a peer to find specific knowledge-items within the P2P knowledge network. The semantic query module is based on the SON as shown in figure 5. Each peer's ontology is positioned as a point in the semantic space. The distance among points denotes the degree of peer interest similarity. The closer among points are, the higher similar of peers' interest and the vice versa. For instance, the points *A* and *B* are close; we say that their interest is similar. Each query can also be positioned in the semantic space. *q* refers as a query point in this figure. To match peer ontology relevant to a query, the searching can be performed in a small region centered at the query point since the relevance of peer ontology outside the region is a relatively low. Therefore, the search space for the query is effectively limited without affecting the accuracy of results.

In ASKnet the pSearch approach is adopted to perform the query. The search is conducted within a region which has a radius of r centred at a query point. We elucidate the process in the below discussion:

1. When a node *A* joins the network, it dispatches a pool of agents to its neighbors to update its peer ontology and its neighbors generates the semantic vector V_n to use it as a key to store the index of A.
2. When node A looks for a particular resource, a pool of mobile agents are created and dispatched to *A*'s neighbors to send its request. When one of its neighbour, say node *B*, receives the query, it generates semantic vector v_q of the query and routes the query in the overlay network using v_q as the key.

3. Upon reaching the destination, the query is flooded to nodes within a radius *r*, determined by the similarity threshold or the number of wanted resources specified by the user.
4. All nodes that receive the query do a local similarity search and agents respond the best matching resource back to the user. Some kind of matching measure is performed to locate the target concepts.

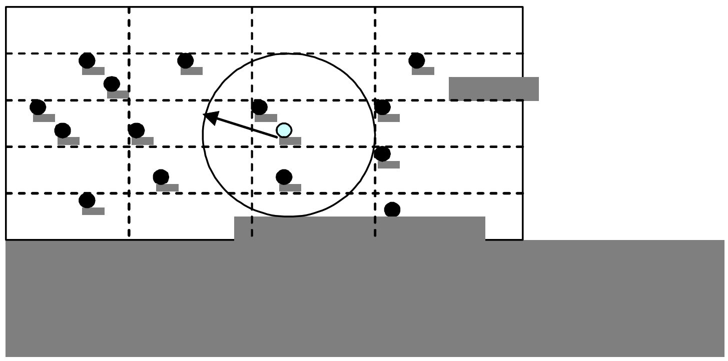

Figure 6: Semantic searching region for a semantic query

4 Concluding Remarks

Knowledge sharing networks are necessary and important in international business, R&D, management and innovation activities. However, organizations and more so knowledge workers are facing problems in identifying, selecting, operationalizing and evaluating validated knowledge resources. The emergence of P2P technology has provided some solutions to knowledge sharing amongst a community of knowledge workers.

For effective knowledge management it is therefore contingent that the knowledge sharing environment deals with the dynamic needs of knowledge workers and adapts the peer configuration accordingly. This brings to relief the need for dynamic knowledge networks that self-configure, based on the semantic content of the knowledge and the profiles of the available knowledge workers, to locate and deliver best quality and validated knowledge.

This concept paper outlines a research program for next-generation knowledge sharing vis-à-vis the incorporation of sophisticated knowledge management methods—i.e. ontologies, semantic web and mobile agents—to design an agent-based semantic knowledge network in a P2P environment. In this concept paper, some limitations on existing P2P application are investigated and the corresponding solutions are proposed: firstly, the semantic web technology as a novel paradigm is combined into P2P network, which overcome the heterogeneity of resources on the network by the use of peer ontology for knowledge conceptualization. Secondly, since mobile agent has the capability of performing operations at remote sites, mobile agent technology is integrated into P2P systems to facilitate semantic search. Thirdly, a semantic overlay network is built which provides efficient peer discovery and query processing.

In closing we will like to point out that our research aims to increase knowledge-workers' effectiveness by allowing them to share their knowledge—vis-à-vis private knowledge repositories of documents—and learn through collaboration in virtual communities.

References

Castano, S., Ferrara, A., Montanelli, S., Zucchelli, D. (2003a) "Helios: A General Framework for Ontology-Based Knowledge Sharing and Evolution in P2P Systems". *Proceeding of the 14th International Workshop on Database and Expert Systems Applications (DEXA'03).*

Castano, S., Ferrara, A., Montanelli, S. (2003b) "H-MATCH: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems." *Proc. of the 1st Int. Workshop on Semantic Web and Databases (SWDB) at VLDB 2003*

Castano, S., Ferrara, A., Montanelli, S. (2003c) "Ontologies and Matching Techniques for Peer-based Knowledge Sharing". *Proc. of the 15th Conference on Advanced Information Systems Engineering (CAISE 2003).*

Castano, S., Ferrara, A., Montanelli, S., Pagani, E., Rossi, G.P. (2003d) "Ontology addressable contents in P2P networks." *Proc. of WWW'03 1st SemPGRID Workshop,* Budapest.

Erhard, R., Bernstein, P.A. (2001) 'A survey of approaches to automatic schema matching". *VLDB Journal*, 10, 334-350

Liebowitz, J. (2000) "Building organizational intelligence: A knowledge management primer". Boca Raton:CRC Press.

Kaplan, B., Duchon, D. (1998) "Combining qualitative and quantitative methods in information systems research: A case study". *Management Information Systems Quarterly*, 12, 571-586.

Nejdl et al. (2002) "EDUTELLA: a P2P networking infrastructure based on RDF". *Proc. of the International World Wide Web Conference 2002 (WWW2002)*, Honolulu, Hawaii.

Nonaka, I. (1998) 'The Knowledge-Creating Company". *Harvard Business Review on Knowledge Management*, Harvard Business School Publishing, Boston.

Noy, N., McGuinness, D. (2001) "Ontologies Development 101: A Guide to Creating your First Ontology". *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and SMI-2001-0880.*

Ratnasamy, S., Francis, P., Handley, M., Karp, R., Shenker, S. (2001) "A scalable content-addressable network". *ACM SIGCOMM'01*.

Wang, Y., Tan, K.L., Ren, J., Pang, X.L. (2001) "An agent-mediated, secure and efficient Internet marketplace", *Proc. of the 4th International Conference on Electronic Commerce Research (ICECR-4)*, Dallas.

Wenger, E.C., Snyder, W.M. (2000) "Communities of practice: The organizational frontier". *Harvard Business Review,* 78, 139-145

Yang, B., Garcia-Molina, H. (2003) "Designing a super-peer network'. *Proc. of the International Conference on Data Engineering*, Bangalore, India.

Zack, M.H. (1999) "Developing a Knowledge Strategy". *California Management Review*, 41, 125-146.