Web Services based Framework for Knowledge Management in Peer-to-Peer Environment

XU Zhengchuan    ZHANG Chenghong    DONG Liang
Department of Information Management and Information System
School of Management, Fudan University
zcxu@fudan.edu.cn; chhzhang@fudan.edu.cn; 0049006@fudan.edu.cn

Abstract

The problem of Knowledge Management (KM) is eminent in the P2P area. This paper presents a general framework of knowledge management based on web services in peer-to-peer (P2P) environment. P2P knowledge management yields significant advantages when combined and applied as web services for KM. By adopting web services techniques and distributed approaches, we partition the knowledge network into numerous knowledge communities self-adaptively, and provide a mechanism for knowledge management in these knowledge communities with our framework. Our framework implements the integration of KM functions and autonomy of knowledge peers in knowledge network, and at the same time, the cost of building and maintaining KM systems is reduced.

Keywords: Knowledge Management, Peer to Peer, Web Services, Knowledge Community

1. Introduction

In recent years, knowledge has become the most precious property of any company. Knowledge Management (KM) plays the key role in upgrading the competitiveness of the company. The classical myth of effective Knowledge management in organization is becoming increasingly critical in the dynamically changing world of business. Research on KM concerns the management aspect including organizational learning, personal management, cultural, etc. (Vadim, D., et al., 1998), and the technical aspect including models, support tools and environments (Zhuge, H. (b) 2002). This paper focuses on the technical aspect and the application background of the network organization.

In a network organization, all personnel and resource are connected together through Information System. In such network environment, all subjects offering knowledge can be regarded as a equivalent entity (Peer), a node in the knowledge network of the organization. It is a difficult problem to manage knowledge effectively under this environment. In this paper, we proposed a basic framework based on web services techniques to solve this problem.

Now, with the rapid emergence of the e-business, Web application develops from localization to globalization, from B2C (Business-to-Customer) to B2B (Business-to-Business), from concentrated type to distributed type. As a new developing web application mode, web services is a brand-new distributed computing model. It is an effective mechanism of the data and information integration on Web.

Besides, P2P networks are distributed systems whose nodes (peers) have equal roles and equal capabilities in exchanging information and services directly with each other. P2P technology eliminates concentrated server, makes resource and service directly exchanged between nodes. Besides, any node can freely join and leave this P2P system. (Aoying Z., Bo
L, 2004). The distribution feature of P2P provides a better environment for the developing Web services. (Qing, W., 2003)

Nowadays, the notion of peer has been extended. A peer can be computation, data, software, agents, and even people (Zhuge, H. (b) 2002). So a peer can be regarded as an integrated mechanism that enables the sharing of various kinds of resources. In an organization, everyone offering knowledge can be regarded as a Knowledge Peer. The whole knowledge resource of the company forms a P2P knowledge environment. In this environment, how to effectively manage the knowledge in the organization and let each Knowledge Peer can freely release, acquire and share knowledge becomes a problem to be solved urgently. Web services technology provides the technological foundation to solve this problem. This paper has proposed a basic framework to solve this problem.

There are a few people who present the solution to KM in P2P environment. In their applications, they only treat the knowledge holder as a peer, and only consider the knowledge acquisition and knowledge retrieval. Unlike those approaches, we adopt the Web services techniques for not only acquiring and retrieving knowledge but knowledge publishing and knowledge querying.

We use a web service as an agent for the knowledge peer, and aim at completing the processes of the knowledge management of the peer, which is from publishing to finding knowledge, then to binding it. Messages are encoded in XML formats (i.e. SOAP messages), and peers can negotiate essential information (e.g. schemas) about shared knowledge before exchanging, which provides more flexibility than common P2P.

The instability of peers, which refers to the freedom for peers to establish and release the connection from the network, results in the uncertainty of providing services. It presents new challenges to the Web services applications. We present a mechanism for notifying service failures and finding backups.

If the service broker is neglected from the web service applications, it may be difficult for a service requestor to find the precise location of the service provider at the large scale of the Web by flooding only. Considering the traditional steps of Web services and the advantages of P2P technology, we employ a distributed approach to implement the service brokers without affecting the autonomy of peers.

The remainder of this paper is organized as follows: Section 2 introduces the related work, Section 3 presents some basic conceptions. Section 4 gives the overview of our framework, and Section 5 proposes the mechanism for interaction among Knowledge Peers through web services. Finally, future research issues are discussed in Section 6.

2. related work
Oscar etc. introduce the concept on P2P Knowledge Management, and presents the architecture of peer-to-peer (P2P) management of knowledge artifacts based on a distributed ontology. They call the knowledge owner as knowledge peer. They think the main processes of Knowledge Management only contain knowledge acquisition and knowledge retrieval. They use ontology to capture implicit knowledge of the knowledge workers and to associate it with knowledge artifacts for classification, search, and browsing purposes. (Oscar 2003)
Sato etc. explain a method of describing the context using XML Linking Language (XLink). They also propose an architecture for sharing the context among users using peer-to-peer technology to help users better understand how to reuse existing contents on computer networks.

Castano etc. propose a general framework, called HELIOS, conceived for supporting dynamic ontology-based knowledge sharing and evolution in P2P systems. The knowledge sharing and evolution processes in HELIOS are based on peer ontology, describing the knowledge of each peer, and on interactions among peers, allowing information search and knowledge acquisition/extension, according to predefined query models and semantic techniques for ontology matching. (Castano etc. 2003)

In Data Mapping, an approach to determine and handle mappings among heterogeneous data sources in a P2P framework is described. Each node determines semantic mappings among instances of its entities, and takes care of mapping consistency interacting with domain experts. Relations are shared with other peers, that run a comparison and search algorithm to create new relations between received mappings and their own data schemas. Results will be distributed again to progressively increase the knowledge of each community member. (Kementsietsidis, etc. 2003)

The InfoQuilt system developed at LSDIS Lab supports heterogeneous information sharing in a distributed framework. Each node describes its own information using an ontological language (DAML+OIL). InfoQuilt has a P2P hybrid architecture: each peer queries the directory node and the inter-domain ontology residing at the node to derive the location of the peers storing semantically related information in the system. Data is addressed with a direct connection to the identified peers. (Arumugam, etc., 2002)

Another related topic with P2P KM is Knowledge Grid, Zhuge proposes a knowledge grid model for sharing and managing globally distributed knowledge resources. The model organizes knowledge in a three-dimensional knowledge space, and provides a knowledge grid operation language, KGOL. Internet users can use the KGOL to create their knowledge grids, to put knowledge to them, to edit knowledge, to partially or wholly open their grids to all or some particular grids, and to get the required knowledge from the open knowledge of all the knowledge grids. (Zhuge, H., 2002, a)

3. Background
In this section we introduce some related background knowledge as follows:

3.1 Web Services
As their name implies, Web services are defined by a set of core standards that are based on existing Internet and Worldwide Web standards. The three primary web services standards are the Simple Object Access Protocol (SOAP), web services Description Language (WSDL), and Universal Description, Discovery and Integration (UDDI). SOAP is a messaging framework that utilizes the HTTP as its default transport mechanism to enable communication between client and server application programs. SOAP messages are created using the XML which provides a flexible method for creating structured, self-describing messages. WSDL is an XML-based language that is used to describe both the business and technical characteristics of web services and how to interact with them.
There are three major roles within the web service architecture, Service provider, Service requestor, Service registry. Figure 1 shows the major web service roles and how they interact with each other.

![Web services Architecture](image)

3.2 Peer-to-Peer (P2P)

Peer-to-Peer (P2P) is a set of protocols, a computing model, and a design philosophy for distributed, decentralized, self-organizing systems. In other words, P2P is a set of methodologies and technologies to allow two or more computers to collaborate in a network of equals (peers), without central coordination (Barkai, D., 2001). The goal of P2P is from sharing files via sharing data then to sharing services across the participants (peers) (Serge, A., et al., 2001). It seems to be a promising solution to the problems existing in centralized systems. In a P2P system, a large number of peers (computers) interconnected together share their resources (e.g. data, services). A peer can become a provider and requestor in the meanwhile. Due to the autonomy and the independence, P2P systems have more availability, scalability, and extensibility than centralized systems.

So far, many applications have taken the advantage of P2P technology, they have shown the vitality of P2P technology. As a kind of special distributed system, the characteristic of P2P network framework makes it obviously different from the system framework used extensively at present.

The central aspect of P2P is the management of a large number of peers, in a usually unstable environment (where peer appear and disappear continuously) without a central coordination. The main potentiality of P2P is the ability of exploiting idle resources (computing), facilitating the exchange of information (information discovery and content distribution). The most interesting features of P2P are scalability, self-configuration, autonomic management, dynamic resource discovery and fault tolerance.

3.3 P2P Knowledge Management

"Knowledge, particularly as manifested in the creation of new products and services, has become a prominent source of wealth creation and sustainable competitive advantage”(Cole, E. 1998)

Knowledge management concerns innovating, spreading, sharing, and using of knowledge. (Zhuge, H. (b), 2002) A classical KM approach is managed by a centralized model, since centralization brings advantages in terms of scope, control and organization. However,
centralized KM systems remain issues on ad-hoc or loosely coupled systems, specialist file sharing, specialist indexing and deployment (Oscar 2003). P2P computing model allows separating the concepts of authoring information and publishing that same information which allows for decentralized application design. The notion of P2P-based distributed application design also enables decentralized and distributed KM systems thus achieving a self-organized management.

Every knowledge owner can be regarded as a knowledge peer, the knowledge owners connecting with each other by Internet make up a P2P Knowledge environment. Then we can manage knowledge in P2P style thus residing on a node in the P2P information system and acting both as client and server (i.e. servant). Each node can directly communicate with other nodes (i.e. building a P2P network) for sharing knowledge. The problem of P2P KM is becoming eminent, we provide a framework for this problem as follow.

4. Overview of framework
This section provides an overview of the web services based framework for knowledge management in Peer-to-Peer Environment. There are some similarities between the Web services and P2P environment. Therefore, it is a relatively ideal scheme to set up Web services on P2P computing platform. It can utilize the advantage of P2P itself to realize the autonomy of resources and integration of services high-efficiently. (Serge Abiteboul, Omar Benjelloun, 2001)

We put forward a knowledge management frame in the P2P environment based on the Web Services. Its system framework is in Figure 2.

In our framework for KM, the ontology (Gruber 1993) is used to capture implicit knowledge of the knowledge workers and to associate it with knowledge peer for search, representation, publish, storage, and sharing purposes.
There are two parts in our framework. The first part is a Knowledge agent, which is a web service, used as a agent for the Knowledge Peer. The knowledge agent allows knowledge to be communicated independently of its holder. The other part is the knowledge base, all fact and rules are stored in it, fact is stored by ontology.

The knowledge agent has three layers. The first layer has two models for knowledge searching (K-searching) and publishing (K-publishing). Knowledge peers submit their requests for knowledge described with ontology by K-searching model, then the model searches the knowledge network to acquire the wanted knowledge. The K-publishing model publishes the knowledge owned by the knowledge peer, then all other knowledge peers can retrieve and use it.

The second layer has four models for knowledge sharing, indexing, reuse and integration respectively. K-sharing model enables knowledge peers to share background information to enhance intellectual creative work and knowledge management. K-indexing model classifies the knowledge and creates the catalog for it. K-reuse model concerns the reuse of problem-solving methods with different domain knowledge. K-integration model integrates the existing knowledge into a single solution for all knowledge peers.

The third layer contacts with knowledge base, it takes charge of knowledge storage, generation and transportation. The K-storage model stores explicit knowledge in knowledge base, and the K-generation model completes mapping from knowledge into ontology or other explicit representations, structured or semi-structured or unstructured. K-transportation model achieves the transformation between different knowledge with different format.

The knowledge base part is consisted of a rule base, a fact base and a KM engine. The rule base contains all reasoning rules existing in the knowledge of this peer, and all the fact is stored in the fact base with ontology or case methods. The KM engine manages meta data and the two bases. All queries for the knowledge base will be submitted to KM engine, it searches the knowledge base and returns the results to users.

In fact, the knowledge agent is a web service, we must assign the service provider, service requestor and service broker as described in section 3.1. Some knowledge agents will be designated as knowledge coordinator, there is a public space in this knowledge coordinator, knowledge peers put all public knowledge there, then every peer in this network can access it.

When a knowledge peer wants to provide his/her knowledge to other peers, he/she will publish the knowledge by K-publishing model. K-publishing model will put the knowledge into a public space. If he/she wants to search knowledge on some topic, he/she will submit the searching task to K-searching model, which will search all public space and shared knowledge to find the wanted knowledge. Every knowledge peer has a shared area with authentication, only other knowledge peers who have access rights can retrieve the knowledge in shared area, the K-sharing model takes charge of this.

Generally speaking, the service broker is centralized. But this characteristic leads to many drawbacks, so we must provide knowledge management mechanisms for knowledge agents in P2P environment.

5. Mechanisms of KM in P2P environment
For the interactions among knowledge management entities, the web services in concentrated
environment will meet a lot of defects, when a lot of services request are submitted to the only knowledge coordinator. Then the knowledge coordinator become the bottlenecks of knowledge management, meanwhile many other knowledge peers are idle. And when knowledge coordinator breaks down, the whole P2P KM system will decline.

In P2P environment, we can adopt distributed approaches to solve the problems above. First we partition the global knowledge network into numerous Knowledge Communities self-adaptively. That is if a knowledge peer can get the knowledge locally, it is unnecessary to visit other communities. Every knowledge community has a coordinator, which acts as public knowledge service broker and registry center in this community. Then the information storm will be avoided, because the search of the related knowledge is limited in this community or among the coordinators in this network, the quantity of the communication data is greatly reduced.

At the same time, some knowledge agents are assigned as centric peers, where the knowledge’s catalogues are stored, the catalogues contain the meta information of classified relevant knowledge. For example, all the product design information is put in the product design catalogue, and all material information is put in material catalogue. All these catalogues lie in centric peers, and their locations are stored in the coordinator. Figure 3 shows the architecture of Knowledge Communities for P2P KM. Its mechanism is as follows:

When a peer wants to join in a Knowledge Community, it must submit the registry request to the coordinator (registry center) in this community, the coordinator records its relevant information (such as the status of the peer, the knowledge catalog of the peer, and the location, etc.).

Figure 3  Architecture of Knowledge Communities for P2P KM
But when a peer wants to leave this community, it need not send message to the coordinator. The coordinator communicates with all peers in the knowledge community periodically to update each peer’s information and sense whether the peer has already leave this community, if the peer did leave, the coordinator automatically deletes registry and related information from registry center.

Every coordinator has a public space, and every knowledge peer has a shared area. When a peer wants to provide knowledge in public, it needs to publish the knowledge in the public space of the coordinator. And if it only shares the knowledge with the peers in a group, it can put the knowledge in its shared area, only the authorized peers can access the knowledge in the shared area. All members in this group maintain a peer list in their local space.

When a peer wants to search some knowledge, firstly it will access the shared areas where it is authorized. If it hasn’t got what it wants to get, it will submit the query to the coordinator, the coordinator will search the target in its public space, then search the catalogues in centric peers by the location information in the coordinator. If the coordinator can’t satisfy the querying peer, it will communicate with the coordinators in neighbor knowledge communities to search the knowledge needed. All message passed among knowledge peers is SOAP message in XML format.

The single-point failure of some key peers (such as, coordinators, community centric peers) will lead to decline of the whole system. In order to avoid these, every key peer must have a corresponding backup. The backup will synchronize with the key peers periodically (per hour, day or week, etc). When the key peer fails, the backup becomes the new corresponding key peer automatically and appoints a new backup peer for itself again.

By adopting distributed approaches, our web services based framework reduces the bottleneck of the service brokers in web services technology, and moreover, it improves the utilization of knowledge resource connected by network. The centric peers containing meta-knowledge increase the efficiency of knowledge searching.

Now we give an example to show our framework how to work in real-world knowledge management applications.

We assume the Fudan University is a knowledge network, every researcher constructs a knowledge agent, every agent is a knowledge peer, all peers are connected with other each to constitute a knowledge network. A researcher stores his/her knowledge into his/her agent, and by the K-Indexing model and K-Integration model in the agent, the researcher can organize his/her knowledge conveniently.

All researchers in Fudan University are always affiliated to a department, the researchers in a certain department have alike research interest, their agents make up of a knowledge community. Every department appoints a knowledge coordinator, which records related information of all knowledge peers in this department, and maintains a public sharing area and centric peers list.

All researcher can put the knowledge which they want to share with others, into the public sharing area by K-Publishing models in their knowledge peers. All other researchers in Fudan University can access the knowledge in public sharing area freely. These researcher still can open their partly knowledge to the members in the same team, the K-sharing models maintain
a access control list for the team members. These researchers can search the wanted knowledge by K-Searching model.

Every department manages knowledge by its coordinator, the knowledge communication among all departments in Fudan University is dependent on it too. The coordinator contacts with every peer in the department every one hour to determine whether the peer left this knowledge community or not, at the same time every peer can leave the knowledge network at ease. By this framework and mechanism, the knowledge sharing in Fudan University is implemented.

6. Conclusion
In this paper, we present an web services based framework for P2P knowledge management and address the mechanism for knowledge peers joining, searching, and publishing knowledge in P2P environment. The framework facilitates achieving autonomy and self-organized management.

By partitioning the knowledge network into numerous knowledge communities self-adaptively, and assigning knowledge coordinator, centric peers, our framework solve the problem that it is difficult to search target for lack of centric nodes in P2P environment. The communication cost is reduced, and the efficiency of searching knowledge is improved greatly. By designating backup for the key peers, such as, coordinator and centric peers etc., we guarantee that the web services based system for KM is robust.

We are incrementally developing prototypes applied within a case study conducted in some large corporations, such as, baoshan iron & steel CO., LTD. etc. Based on the prototype and case study results, we will address particular research issues in further work, such as, investigating the P2P management of knowledge peers incorporating richer data (e.g. relational, semi-structured, XML, meta-data), further addressing issues concerning the P2P management of distributed ontology (e.g. merging, combining, semantic mismatch, searching), and investigating security issues such as trust, privacy, confidentiality, and authenticity in the distributed peer-to-peer knowledge environment. Knowledge distribution is another further issue to be studied too.

References


Qing, W., and Yang Y., Junmei Z., and Aoying Z., “Peer-Serv: A Framework of Web Services in Peer-to-Peer Environment,” *In Proceedings of the 4th International Conference on Web-Age Information Management (WAIM), 2003 LNCS.*


Tsur, S., “Are web services the next revolution in e-commerce?,” *In Proceedings of International Conference on Very Large Databases (VLDB), 2001.*

