

# Research on Distributed Knowledge Base System Architecture for Knowledge Sharing of Virtual Organization

Ruzhi Xu<sup>1,2</sup> Peiguang Lin<sup>1</sup> Cheng Liu<sup>1</sup>

<sup>1</sup>School of Computer & Information Engineering, Shandong University of Finance, Jinan 250014, P. R. China

<sup>2</sup>School of Economics & Management, Tsinghua University, Beijing 100084, P. R. China

## Abstract

In this paper, an agent-based architecture is proposed to integrate distributed heterogeneous Knowledge Base. BBS is used as the public information exchange platform to implement an asynchronous uncoupling communication mechanism. A group of agents is deployed to every Knowledge Base node to complete users' retrieval requests through their autonomous behaviors and collaboration on the BBS. A unified extendable external interface is provided through which Knowledge information can be shared. Every kind of enterprise Knowledge Base can be integrated into the system while its physical structure and geographical distribution do not need any modification. The inter-connection among Knowledge Base using the agent-based distributed Knowledge Base architecture will bring great positive impact to the growing of Knowledge Base Systems.

**Keywords:** Agent, Knowledge Base (KB), Virtual Organization, Knowledge Sharing

## 1. Introduction

Economic globalization and global networks have become the trend of world development. Especially with the

development of information technology and the innovation of management theory, modern enterprises need to go beyond the boundaries of traditional companies to achieve for a rapid and effective integration of resources on a global scale, thus can produce high-quality products that meet the needs of the market or provide services the customer need in a very short period of time. In this case, a new organizational model - virtual organization came into being, and increasingly been respected and widely used by business and government organizations. It has become the 21st century's most important organization and modus operandi.

While the essence of virtual organization is a member unit of the dynamic alliance, aimed at collaboration with member organizations to achieve common goals, the core is the integration of resources and knowledge exchange and cooperation is related to knowledge sharing. Only through effective knowledge-sharing between members it can achieve organizational synergy and innovation in knowledge together to create a more excellent products and services. In reality, virtual organization members have motivation to share knowledge but need to ensure their proprietary and security. It is necessary to meet the needs of member organizations of the knowledge-sharing, but also to

achieve cross-organizational and cross-cultural knowledge sharing, and to meet the needs of agile virtual organization dynamically.

Compared to a single organization, knowledge-sharing in virtual organization faces the barriers in cross-organizational knowledge search and knowledge transfer. It needs to address the distribution and heterogeneity of knowledge, knowledge proprietary and information asymmetry problems. It also needs to provide a kind of effective methods and means for rapid discovery and dynamic integration of virtual organizations knowledge resources.

The subject starts from the analysis of knowledge-sharing features and knowledge-sharing process of virtual organizations, then goes into dynamic integration of distributed heterogeneous knowledge base to carry out in-depth study methods. The study involves modern management, information retrieval theory and technology in many fields, belonging to the application of cutting-edge basic research in the field of knowledge management. It helps to enrich and develop the theory of virtual organization and knowledge-sharing, promoting a theoretical significance and wide application value in virtual organization and knowledge management applications.

The theory of functional dependencies [10~12] for relational data-bases can not be directly applied in XML documents as there are significant differences in their structures: relational model are flat while XML schemas are nested. For XML functional dependencies, there are two major approaches to define in XML research community. The first approach is based on paths in XML document, such as Refs. [13-18]. Unfortunately, they do not deal with the tree-structured situation proposed in this paper. The second approach is based on sub-graph or sub-

tree in XML documents, such as Ref.[19], but it does not deal with tree-structured situation with some constraint conditions proposed in the paper. Ref. [20] deals with XML functional dependencies with constraint condition, but without specifying what kind of constraint they allowed.

In this paper, we give the definition of XML functional dependencies based on constraint tree. The definition proposed in our paper overcomes the shortcomings of previous definitions in the following aspects: (1) It captures the tree-structured characteristics of XML documents in structure. (2) It considers a more general situation of XML functional dependencies based on sub-trees with some constraint conditions in XML documents. More discussions can be found in Sub-section 3.2.

The rest of the paper is organized as follows. Some notations are given in section 2 as preliminary work. The definition of XML functional dependencies is given in Section 3. The XML expression of XML functional dependencies and their relationship with XML keys are also given in Section 3. Section 4 concludes the paper and points out the directions of future work.

## 2. Related work

Domestic and foreign scholars have conducted some deep research on the establishment of specialized knowledge-sharing system. The Institute AIFB of University of Karlsruhe in Germany developed ontology-based knowledge management system OKMS under the project Onto logging, using ontology mapping ontology to establish the integration between distributed heterogeneous mechanism [1]. Institute of the ISI in The University of Southern California has developed a knowledge acquisition environment for the

establishment of the field based on ontology technology to provide non-programmers with knowledge acquisition tools to build a knowledge base [2]. On this basis, ISI has also launched TRELIS project to study the WEB-based semantic knowledge asset management, and use information to achieve the semantic annotation from the original source of information and other knowledge fragments to generate new knowledge fragments [3][4]. Based on initiative from the U.S. Defense Advanced Research Projects Agency (DARPA), Stanford University, MIT and other institutions participate in "knowledge-sharing action" HPKB project [5], jointly developed a large-scale knowledge base that can be reused by distributed multiple users around the world, which contains information about the comprehensive knowledge of the field simultaneously.

These studies provide a solid theoretical foundation and technical support for the settlement of distribution of knowledge-sharing environment, particularly the use of ontology technology to solve the heterogeneous knowledge resources and knowledge sharing, but the dynamics of virtual organization agility require a virtual organizational knowledge sharing system must be a dynamic, easy to reconstruct, easy to expand. QI Er-shi, ZHENG. Xiaodong have proposed a Web-based virtual enterprise knowledge management system model and built a virtual enterprise knowledge management platform [6], but the virtual organizational knowledge management platform is based on a common knowledge base system that does not explicitly give a virtual organization Knowledge of dynamic integration of various business methods. YANG Bo, XU Shenghua have proposed a virtual enterprise based on multi-Agent

Knowledge Management System Framework [7], and solved knowledge integration between the members of the knowledge management system, but this integration is based on the leader of an alliance who has set up a centralized base. The system's load balancing and system maintainability need to be improved. SHEN Jie, designed a distributed knowledge management model based on Multi-Agent Systems, but the integration of heterogeneous knowledge resources depends mainly on the direct communication between the Agents, which affected the efficiency of information retrieval and the security of knowledge base [8]. ZHANG Chun-Xiao has studied the distributed knowledge-based systems based on P2P technologies and presents a distributed query mechanism based on deductive reasoning. The query can be concurrently performed in a few nodes to optimize the computing load balancing and increase knowledge query efficiency [9]. The research results seems a good choice only from the knowledge query effect in theory, but the exchange of visits between the corporate knowledge bases affects the knowledge proprietary as well as safety and it does not take into account the characteristics of the members of virtual organizations dynamic and cooperate completed a target, so it is difficult to achieve the coordination of knowledge-sharing behavior.

Related research on digital libraries and knowledge network technology has a great meaning for the study of this topic on the distributed knowledge in the virtual organization environment [10] [11]. However, digital libraries and other information retrieval technology emphasis more on the control of individual search servers, while the two main problems are the division of literature collection and source selection search. As for the division of literature,

because the background knowledge sharing in virtual organization has been determined by various companies to establish their own knowledge base, forming a natural division of the state, we can not change this split situation. For source selection in the search, since we have to focus on protecting the security of corporate knowledge base, so take the initiative to retrieve the behavior is issued by enterprises, rather than the Chinese and control server to instructions, that is, removed from the external control and access to enterprise within the possibilities. And a simple distributed information retrieval technology does not take into account the characteristics of virtual organizations, knowledge sharing, and the same lack of organizational goals for the completion of the virtual knowledge-sharing behavior of the various member organizations of effective coordination mechanisms.

### 3. Architecture of the Distributed Knowledge Base

The principle of the distributed Knowledge Base proposed in this paper is to deploy several agents on each enterprise Knowledge Base node and complete users' retrieval through the cooperation of these agents. The advantage is that the geographical distribution and physical implementation of the enterprise Knowledge Base do not need any modification. Only several agents are deployed on each enterprise Knowledge Base node, including Release Agent, Retrieval Agent and Require Agent. These agents share a public bulletin board system comprising of Release Area, Retrieval Area and Feedback Area and fulfill users' requirements via their autonomous behaviors and collaboration.

In the system, every enterprise could send its Knowledge to the Release Area

through Release Agent. When an enterprise needs Knowledge, it can search the Release Area or send a retrieval request to the Retrieval Area through its Require Agent. In the second situation, other enterprises' Retrieval Agents can get the retrieval request, search their local Knowledge Base, and send the Knowledge found to the Feedback Area. Then the enterprise that sends the retrieval request can get the Knowledge it required from the Feedback Area through its Require Agent. To solve the heterogeneity problem among enterprise Knowledge Base, a virtual central Knowledge Base is introduced and its Knowledge specification scheme is used as the reference scheme. Each agent has a Knowledge specification translator that can implement the translation between the enterprise Knowledge specification scheme and the central Knowledge specification scheme automatically.

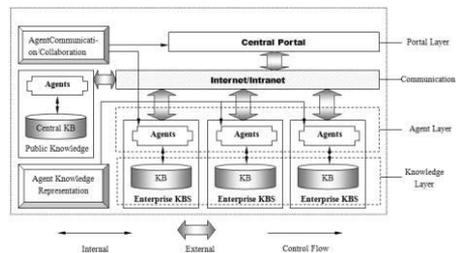


Fig. 1: Architecture of the Agent Based Distributed Knowledge Base.

Figure 1 shows the architecture of the Agent Based Distributed Knowledge Base (ABDKB). The enterprise Knowledge Base geographically distributed on different places compose the distributed Knowledge Base sub-system. The central Knowledge Base is the central portal of the whole system logically.

In Figure 1, three enterprises Knowledge Base are used to represent all the enterprise Knowledge Base in the system. In each enterprise Knowledge Base, "Agents" represents the agent group deployed in the enterprise, which provide the services of the distributed Knowledge Base. The ABDKB architecture has four layers:

(1) Portal layer: The portal layer provides a unified entrance to the system. Every user submits retrieval request and gets retrieval result through this entrance. Furthermore, this layer also provides some services that can improve the usability of the system, for example, the network topological figure and the retrieval request guide.

(2) Communication layer: The Central Portal is connected to all the enterprise Knowledge Base through Internet/Intranet (mainly Internet); therefore it can provide end users with the ability to retrieve these distributed Knowledge Base.

(3) Agent layer: As the core of the ABDKB, the agent layer is composed of the agents deployed on every member Knowledge Base. These agents provide the services such as submitting retrieval request, getting retrieval request, interacting with the local Knowledge Base, returning retrieval result, getting retrieval result, and etc. Because the differences in the network communication mechanism, the collaboration mechanism, and the formats of retrieval request/result among various local repositories become transparent to end users because these agents are autonomous and intelligent, they can perform necessary translation automatically. Therefore, the heterogeneity problem is well solved.

(4) Knowledge layer: every local Knowledge Base is designed and managed by its owner enterprise; therefore enterprises' legacy Knowledge Base can be used in the ABDKB. Theoretically this layer is able to contain unlimited numbers of local enterprise repositories (only three are drawn in Figure 4). An enterprise Knowledge Base can join the ABDKB if only it deploys appropriate agents and establishes communication link to the central portal.

In this architecture, the communication and collaboration mechanism between agents is a problem that must be well considered in the design of the agent structure and the central portal. The communication and collaboration mechanism covers the message transfer between agent and agent, information exchange between agent and the central portal, and the agent behaviors in the entire process from request submission, request handle, to result return. The brief execution procedure is:

1) Enterprise deploys appropriate agents on its local Knowledge Base, performs the registration on the central portal, and establishes the communication link to the central portal.

2) Enterprise submits the information of its sharable Knowledge to the central Knowledge Base through its release agent.

3) Personal user submits retrieval request via the interface of the central portal. When the needed Knowledge can not be found in enterprise's own Knowledge Base, the enterprise could attempt to obtain the Knowledge from outside. Under this situation, the administrator of the enterprise Knowledge Base submits the retrieval request, which then is sent to the central portal by the require agent. This

function is used to improve the method that enterprise gets needed Knowledge from outside.

4) The retrieval agent of every enterprise accesses the central portal periodically. It gets retrieval requests of other enterprises and searches its local Knowledge Base.

5) The result agent returns the retrieval result to the central portal after the local search finishes.

6) The central portal implies a classification and a sort order to the retrieval result according to its corresponding retrieval request.

7) Personal user views the retrieval results on the central portal and selects the needed Knowledge. With the information about the Knowledge Base embedded in the retrieval result, the user then contacts the relevant enterprise to obtain the Knowledge.

8) The result agent of the enterprise accesses the central portal periodically. It gets the retrieval results relevant to its retrieval request and stored them locally to wait for the administrator to view. The handling process of the administrator to the retrieval results is as same as the personal user.

The inside structure of the enterprise Knowledge Base is shown in Figure 2. The Knowledge entities and their relevant information are stored in the local Knowledge Base. A management system is responsible for managing and maintaining the local Knowledge Base.

#### 4. Collaboration Mechanism between Users and Knowledge Base

Considering that every enterprise Knowledge Base is independent in con-

text physically/logically and doesn't need to communicate with other enterprise Knowledge Base directly, a platform based on the Bulletin Board System (BBS) is used as the media to collaborate the behaviors of all agents on the central portal.

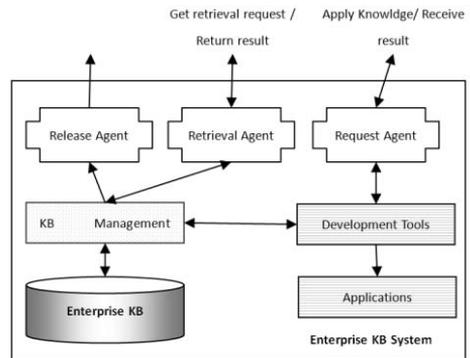


Fig.2: The Structure of the Enterprise Knowledge Base System.

The BBS is a place for information storage where sender can put messages and receiver can get messages. Every procedure can read interesting items and handle them according to its own status independently. The put and get operation do not need to be synchronized, and therefore it is a non-time-coupling collaboration model. It is suitable for the characteristics of the autonomous intelligent agent. On the other side, every member in collaboration must know the message patterns; therefore the BBS is a space-coupling information space.

In the ABDKB architecture, numbers of enterprise Knowledge Base are connected to the Internet/Intranet. The BBS on the central portal is the collaborator of the behaviors for integration. Users interact with the BBS by sending retrieval requests to it and getting retrieval results from it. There are certain numbers of agents running on each enterprise Knowledge Base. They

submit retrieval requests to the BBS, get retrieval requests from the BBS, search the local repository, return retrieval results to the BBS, and get retrieval results from the BBS. According to the above requirements; a much detail structure of the ABDKB is show in Figure 3. In such a design, every agent

may access the server, get retrieval requests and handle them on its own computer when its processor is idle. Therefore, the work load is distributed to the whole network, rather than on the server only. Work load distribution is a key problem in distributed environment.

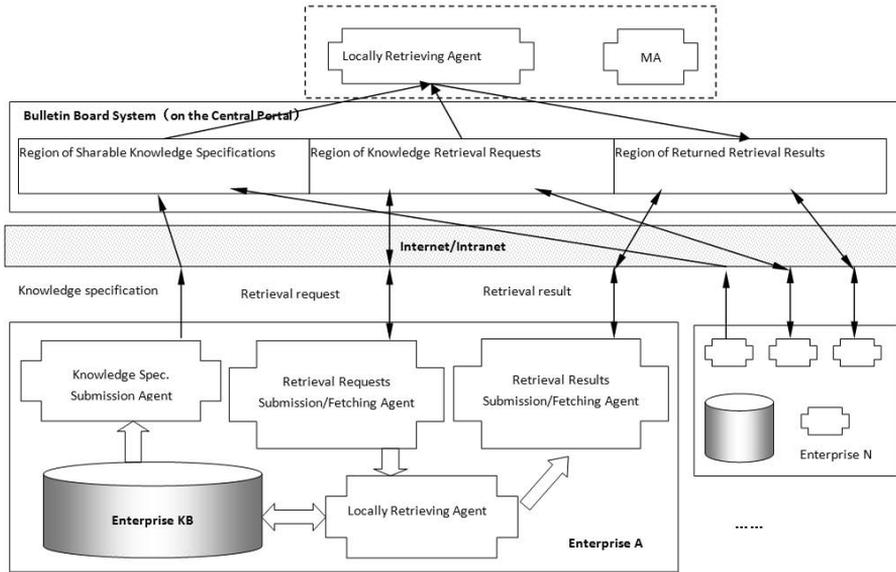


Fig.3: Detail Structure of the BBS based Collaboration Mechanism.

Figure 4 shows an information-flow-centered view of the interaction between the member Knowledge Base in the system. In the figure, arrow (1) represents a release agent sends Knowledge information to the BBS. Arrow (2) represents an enterprise sends a retrieval request to the BBS when it needs to retrieve a Knowledge. Arrow (3) represents other repositories get retrieval requests periodically. Arrow (4) represents a repository returns the retrieval result to the BBS after local retrieval. Arrow (5) represents the enterprise that submits the retrieval request accesses the BBS again after a period of time. If it finds some results relevant to

its request, it will get the results and store them locally. Figure 4 shows the situation that enterprise A provides services to enterprise B.

In fact, the position of all enterprise Knowledge Bases are equal and peer-to-peer in the whole architecture. Any enterprise Knowledge Base can be request submitter or service provider. Usually the central Knowledge Base is a territorial repository established by some public organization. It is a server and is responsible for searching the public Knowledge information according to the retrieval requests.

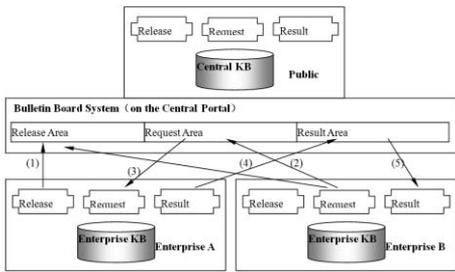


Fig.4: The Information Flow on the BBS.

### 5. Implementation

The architecture and its implementation described in this paper have been applied in the constructions of Qilusoft Knowledge Base. The prototype system involves several enterprise repositories connecting to the Internet, employs a computer server as the request server, and establishes an integrated system. It can perform the retrieval requests from users in distributed hosts and combine the returned results into a unified set, finally publish to the users. Figure 5 shows the integrated view of the system observed from the central server's monitor tool.

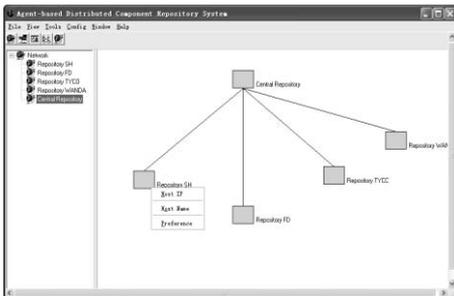


Fig.5: An integrated view of the system.

As shown in Figure 6, we provide a wizard to guide users constructing their retrieval request. The final result is formatted into a concept-dividing tree and appended to the server's request queue.

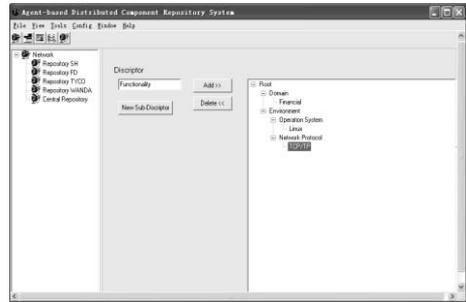


Fig.6: Request construction wizard.

### 6. Conclusion

In this paper, an effective approach to build distributed knowledge base is proposed, which can dynamically integrate all the members of virtual organizations knowledge resources. Using this method, we can maintain the existing physical distribution, realize the enterprise knowledge base mechanism left unchanged, and protect the proprietary nature of the enterprise knowledge as well as knowledge base system security under the premise of a virtual organization with the dynamics of the members of the organization to achieve the virtual members of the knowledge base for integration and knowledge sharing. It has important theoretical and practical significance to promote knowledge sharing for the Virtual Organizations.

### Acknowledgement

This work is supported by the National Natural Science Foundation of China under Grant No.60603070 (Research on Construction Method for Emergency driven Virtual Organization); the Provincial Natural Science Foundation under Grant No.2006BS01007; the Graduate Innovation Education Research Project of Shandong Province, China under Grant No. SDYY09088.

References

- [1] Maedehe A , Motik B , Stojnaovic L, Sutder R, Volz R.. Ontologies for enterprises knowledge management. IEEE Intelligent Systems.2003 , March/April, pp.26-33
- [2] Blythe J. Integrating Expectations to Support End Users to Acquire Procedural Knowledge , *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence IJCAI-2001* , Seattle , WA , August 2001.
- [3] Kim J. , Gil.Y. Knowledge Analysis on Process Models , *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI.2001)* , Seattle , WA , August 2001
- [4] Gil Y., Ranatkar V. TRELIS: An Interactive Tool for Capturing Information Analysis and Decision Making , *International Conference , the EKAW2002*, Siguenza, Spain, October 1-4, 2002.
- [5] Ramesh S.Patil, et al, The DARPA Knowledge Sharing Effort: Progress Report[R]. [http://www.isi.edu/isd/KRS\\_harin/](http://www.isi.edu/isd/KRS_harin/).2004-01-13
- [6] Er-shi QI , Xiao-Dong ZHENG, Yi-Song ZHENG, LI Gang. Web-based Virtual Enterprise Knowledge Management System [J]. *Industrial Engineering*, 2006, Vol.9 (1) :70-74.
- [7] YANG Bo, XU Shenghua. Based on Multi-Agent Virtual Enterprise Knowledge Management System Framework. *Intelligence magazine*, 2009 (10) :107-110
- [8] SHEN Jie, LUO Jian-li. Based on Multi-Agent System for Distributed Knowledge Management. *Engineering Theory and Practice*, 2006, No. 01
- [9] Chun-Xiao ZHANG. Distributed P2P-based Knowledge Base System [D]. Jinan University, a master's degree thesis, 2008.
- [10] Joshua W. Hui, Ajay Sood, Mahadevan Subramanian, Parag V. Tijare. Exploiting Database Technologies for Digital Libraries. IBM Almaden Research Center, March 2001.
- [11] Katherine J. Don, David Bainbridge, and Ian H. Witten. The Design of Greenstone 3: An Agent Based Dynamic Digital Library. University of Waikato, August,2003.