

Shih, W. C., Yang, C. T., & Tseng, S. S. (2006). An interlibrary cooperation framework for digital libraries using P2P technology. In C. Khoo, D. Singh & A.S. Chaudhry (Eds.), *Proceedings of the Asia-Pacific Conference on Library & Information Education & Practice 2006 (A-LIEP 2006), Singapore, 3-6 April 2006* (pp. 156-159). Singapore: School of Communication & Information, Nanyang Technological University.

AN INTERLIBRARY COOPERATION FRAMEWORK FOR DIGITAL LIBRARIES USING P2P TECHNOLOGY

WEN-CHUNG SHIH

*National Chi Nan University Library,
Puli, Nantou Hsien, 54561, Taiwan (R.O.C.)
Department of Computer and Information Science, National Chiao Tung University,
Hsinchu, 30010, Taiwan (R.O.C.)
E-mail: wjshih@ncnu.edu.tw*

CHAO-TUNG YANG

*Department of Computer Science and Information Engineering, Tunghai University,
Taichung, 40704, Taiwan (R.O.C.)
E-mail: ctyang@thu.edu.tw*

SHIAN-SHYONG TSENG

*Department of Information Science and Applications, Asia University,
Taichung, 41354, Taiwan (R.O.C.)
E-mail: sstseng@asia.edu.tw*

Abstract. InterLibrary Cooperation is an important activity of libraries. However, the emerging digital library architecture has not explicitly supported this requirement. In this paper, we propose an InterLibrary Cooperation Framework for digital libraries using P2P technology. An application of this framework to Faculty Publication Sharing System is presented. Besides, a reputation model based on data mining is utilized to provide libraries with incentives to join this framework.

Introduction

Traditionally, libraries have been the most important source of knowledge for people, and managers of libraries have constantly been seeking every way to promote their services. As Internet rises and knowledge explodes, the concept of digital libraries has been extensively accepted, and a lot of work has been devoted to this field (Fox & Marchionini, 1998; Borbinha et al., 2004; Maly et al., 2005; Lagoze, 2002; El-Bayoumi & Charlong, 2003).

Basically, it is impossible for a single library to collect all the books in the world and do all the work by itself. Therefore, InterLibrary Cooperation (ILC) is a necessary activity for libraries to satisfy the information requests of their users. In brief, ILC means that two or more libraries cooperate to share their resources, which might be books, journals, catalogs, etc.

While most of previous work deals with issues of one single digital library, not many researchers investigate interconnection of many digital libraries. In (Yang et al., 2004), a framework of Global Digital Library Grid was proposed, and this platform was claimed as a scalable way to interconnect libraries. However, it is built on Data Grid technology, which is designed for data-incentive computing. Basically, this framework is not suitable for documents in libraries, especially for requirements of InterLibrary Cooperation.

In recent years, Peer-to-Peer (P2P) technology has been extensively adopted in file sharing systems. Furthermore, P2P based digital library architectures have been proposed. In (Ding, 2004), meta-data integration issues are discussed. In (Ma & Aygun, 2005), a P2P based digital library is proposed to deal with multimedia collection. In (Walkerline & Rayson, 2004), a digital library prototype is built by P2P based technology.

To the best of our knowledge, existing digital library architectures have not taken ILC requirements in serious consideration. This is the primary motivation for us to propose an ILC framework for digital libraries. In this paper, a hybrid framework combining Client/Server model and P2P networking is proposed to support ILC for digital libraries. This framework has the following advantages. First, it has advantages of P2P systems, such as scalability, cost-effectiveness, etc. Second, it also has advantages of Client/Server systems, such as efficiency, manageability, etc. The key point of elaborate inte-

gration is a two-layered structure. The upper layer is a P2P network interconnecting participating libraries, and the lower layer is a Client/Server system in each library.

For information sharing among digital libraries, OAI (Open Archive Initiative) (Lagoze & Sompel, 2001) is commonly used. However, our approach is different from the OAI approach. OAI is focused on metadata access and sharing, while our framework involves metadata and documents sharing. In fact, OAI can be adopted for metadata sharing in our framework.

Our main contribution in this paper is to propose a new ILC framework for digital libraries, which investigates issues about interconnection of digital libraries, and ILC support for digital libraries. Surprisingly, this important issue has not been seriously considered in the previous literature. Finally, a reputation model based on data mining is utilized to provide libraries with incentives to join this framework.

The rest of this paper is organized as follows. The following presents our new framework. The last section concludes this paper.

Our ILC Framework for Digital Libraries

In this section, the assumptions of our framework are described. Then, the architecture of our framework is presented. Finally, we will explain the reputation mechanism.

1. Terminologies, Assumptions and Scenario

The terminologies in our framework are defined as follows.

- Document: a digital object, which is the target of archiving, publishing and delivering in our framework. For example, A document might be a thesis or dissertation.
- Peer Server (PS): the server of a participating library, which responds to users and is in charge of P2P document sharing.

Our framework is based on some assumptions, which are described as follows.

- Documents in this framework are contributed by users of Peer Libraries.
- Once a document is contributed by a user, this document can be delivered without any copyright constraints.

The concept of our proposal can be captured quickly by a scenario. Professor Chang, a member of University Library A (ULA), needs a paper whose author is Professor Lee, a member of University Library B (ULB). At first, Chang connects to local database of ULA, and do not find this paper. Then, he requests the peer server of ULA to help him. When the peer server of ULA is requested by a user, the ILC process is triggered. Through the P2P mechanism, the targeted paper is found in ULB, and is delivered to local database of ULA. Finally, Chang is notified by the peer server of ULA, and downloads this document from local database of ULA.

2. Architecture and Services

The architecture can be organized as a two-layered structure, as shown in Figure 1. The upper layer is a peer-to-peer based network, where each peer is a peer server of a digital library. Meanwhile, the lower layer is composed of users or clients. These users are grouped according to their membership. For example, the dotted line for Library 1 surrounds a peer server and its users. Basically, the users interact with their peer server in Client/Server model.

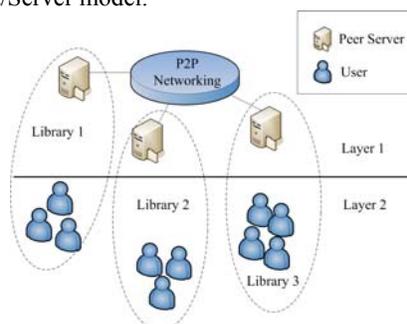


Figure 1. A two-layered structure of ILC for Digital Libraries

In our framework, each library has at least one peer server, which can be seen as a gateway of local users and remote libraries. On the one hand, the PS responds to its users at layer two as mentioned above. On the other hand, the PS interacts with other PS's at layer one.

Each PS is associated with a local database and a cache database. The former contains documents of its own archives, while the latter caches documents downloaded from remote libraries.

To implement this framework, the steps below can be followed by libraries.

1. Each library who intends to join this framework has to setup a peer server. Besides, the PS must be able to operate in the local Client/Server model and the remote P2P network.
2. Each library must collect documents from local users for sharing with other libraries.
3. Depending on the P2P protocol used by the framework, now the library can join this framework via its PS.

3. Reputation Mechanism Based on Data Mining

If some participant in this framework only retrieves documents from remote servers but never provides its own documents to others, this member is referred to as a "Free Rider" (Adar & Fuberman, 2000). Basically, our framework intends to encourage participants to contribute its own documents for others to retrieve, thus accumulating a huge P2P document databases. Therefore, serious free riding will stop this framework to grow.

The free riding problem is a well-known issue in P2P networks, and many solutions are proposed (Kamvar et al., 2003). In this paper, we propose a data mining approach to model the reputation system, which gives an incentive for libraries to join this framework and to contribute more its own documents. The idea of this reputation mechanism is that the level of service a participant can get depends on the historical reputation of the participant. If the participant has been enthusiastic to contribute, it will easily get documents from other peers. In contrast, it will be difficult for the participant to retrieve documents from others if it used to be a free rider.

The reputation system is maintained by a third-party peer. The components of this model are described as follows.

- Transaction Database: when the Peer Server completes a user request, it records this transaction log in the Transaction Database. For example, the Transaction Database can be designed as a relational database. The attributes in a transaction log includes Requesting Peer, Responding Peer, Data, Time, Category_1, Category_2, etc. In our framework, Category_1 might mean Computer Network; Category_2 might mean e-Commerce. There can be hundred or thousand of categories depending on the principle of classification for documents.
- Data Mining Engine: this engine periodically execute data mining tools on the Transaction Database. Whenever meaningful knowledge is found, it is stored in the Reputation Rule Base in the form of rules. For example, Association Rule Mining can be used to find association rules in the transaction log. Traditional reputation models rely on simple historical statistics of transaction logs to determine the trust and reputation of other peers. It is imperative to utilize deeper knowledge, such as association rules, to make good decisions.
- Reputation Rule Base: the rules acquired from the Transaction Database are stored in the Reputation Rule Base. When Peer Server is requested by users, the PS try to find another appropriate Peer Server according to the Reputation Rule Base.

Conclusion

We have proposed a P2P-based framework of ILC for digital libraries. This framework is scalable, cost-effective, and attractive. The future work is to implement a test bed prototype, and to investigate several issues, such as mechanisms of searching, routing and reputation. Besides, incorporation of personalized user interfaces into this framework will be considered in the near future.

Acknowledgments

This work was supported in part by National Chi Nan University Library and Ministry of Education of the Republic of China under the project "2006 Striving for Excellence in Education Program (SEEP)."

References

- Fox, E.A., & Marchionini, G. (1998). Toward a worldwide digital library. *Communications of the ACM*, 41(4), 29-32.
- Borbinha, J., Freire N. & Neves J. (2004). BND: the architecture of a national digital library. *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital Libraries*, 21-22.
- Maly, K., Zubair, M., Chilukamarri, V. & Kothari, P. (2005). Grid Based Federated Digital Library. *Proceedings of the 2th conference on Computing Frontiers*, 97-105.
- Lagoze, C. (2002). Core services in the architecture of the national science digital library (NSDL). *Proceedings of the 2th ACM/IEEE-CS joint conference on Digital Libraries*, 201-209.
- El-Bayoumi, J., & Charlong, L. (2003). The University of New Brunswick's pilot for an electronic theses and dissertation program. *Proceedings of the 31st annual ACM SIGUCCS conference on User Services*, 240-246.
- Adar, E., & Fuberman, B.A. (2000). Free Riding on Gnutella. *First Monday*, 5(10).
- Ding, H. (2004). Towards the Metadata Integration Issues in Peer-to-Peer Based Digital Libraries. *GCC 2004, LNCS 3251*, 851-854.
- Kamvar, S., Schlosser, M., & Garcia-Molina, H. (2003). EigenRep: Reputation Management in P2P Networks. *Proceedings of World-Wide Web Conferenc.*
- Ma, Y., & Aygun R.S. (2005). Peer-to-Peer Based Multimedia Digital Library. *Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory*, 130-134.
- Walkerline, J., & Rayson P. (2004). P2P-4-DL: Digital Library over Peer-to-Peer. *Proceedings of the Fourth International Conference on Peer-to-Peer Computing (P2P'04)*.
- Yang, C.T., Ho, H.C., & Yang, I.H. (2004). Implementation of Global Digital Library Grid Platform Using Data Grid Technology. *Proceedings of TANET 2004, Taiwan*.
- Lagoze, C., & Sompel, V. (2001). The Open Archives Initiative: building a low-barrier interoperability framework. *Proceedings of the 1th ACM/IEEE-CS joint conference on Digital Libraries*.