Network Community Oriented Information Sharing System for Databases

Yuka EGUSA†, Masao TAKAKU†, Hidehiro ISHIZUKA‡
† Graduate School of Information and Media Studies,
University of Library and Information Science
‡ Institute of Library and Information Science, University of Tsukuba

Abstract

We have developed a network community-oriented information sharing system for databases. Today, it is difficult to find out which databases are available and to select appropriate databases which satisfy users' needs, because various databases are available. We classified information about databases into two categories. One is the fundamental information sufficient for users to select databases depending on their own needs. We designed 21 items as fundamental information and developed the system such that users are able to browse and add fundamental information. The second type of information is commentary information, which is comprised of, for example, users' know-how and grouping information from various viewpoints. In our system users are able to browse and add commentary information. We have used 3,614 records from databases in Japan, derived from Database Daicho Soran, as a sample data set for our system. The system allows users to share information about databases, easily discover suitable databases and use them.

Keywords: information sharing, fundamental information about databases, network community

1 Introduction

Many more databases are available than in the past, because information resources, including databases accessed without restrictions of time or a place, have been increasing with the spread of the Internet and many of them are offered freely. Moreover, users' behaviors have been varied because of the variety in the database users. Before the spread of the WWW, most users of databases were specialists, but today most of the users are end users of various jobs and ages. Also it is necessary to use a number of databases often in order to satisfy the demands of a search, because the demands vary and the number of records included in one database is limited. For example, traditional search engines have limits specifying that dynamically generated pages not be covered[1], though they are intended for the whole Web. Consequently it is difficult to find out which databases are available, to select appropriate databases and to make full use of databases.

These problems can be solved by building community to share fundamental information about databases. We designed items required to describe fundamental information about databases and constructed a system where users were able to input information and browse it. In this system, users can also share their own know-how about databases.

2 Related works

Our approach to finding, selecting and using databases is one in which users are able to input and use information about databases with each other. A similar approach to this problem cannot be found. In this section, we describe related works related to this problem.

Database directories have been created to summarize information about databases; e.g. Gale Directory of Databases[2]; Database Daicho Soran[3] (In Japan); database directory[4] by National Institute of Informatics (In Japan); and Dnavi[5] by National Diet Library of Japan, which is limited to databases on the WWW. These directories are compiled from reports or inquiries by the respective organizations. The directories supply users with know fundamental information about databases. However, the scope of the information covered by these directories is limited to commercial databases and/or CD-ROM based databases. The frequency at which they are updated is 1-2 times per year. Another problem is that these directories cover fundamental information only.

There are also books for beginners or for specific fields which summarize typical uses of databases[6]. The scope of information covered by these books is limited, because these books are created for various purposes from the authors' viewpoints. Additionally, information in these books does not follow updates of the databases.

There are some studies that help users to select databases by using an expert system[7][8]. These studies are modeled on specialists' strategies for

database selection. They extract information from some database directories and/or search specialists' knowledge. In contrast with our study, these focus only on databases which are already known by an operator of an expert system. In these studies, discovering a new database depends on the operator.

WAIS[9] is a system which has a feature to provide information about databases. WAIS has a special database, which is called directory-of-servers. A directory-of-servers stores fundamental information about each distributed database: location, keyword, description, etc. Its approach is similar to ours in sharing fundamental information about databases. However, only a database provider is able to add information about a database in WAIS. It does not have a feature allowing casual users to add information. While our system covers all kinds of databases, a directory-of-servers covers databases based on a WAIS protocol only and does not have a feature to handle users' know-how.

Meng et al.[10] classified approaches of database selection in metasearch engines into three categories: rough representative approach; statistical representative approach; and learning-based approach. Many approaches have been proposed to tackle the database selection problem[11][12]. The database selection in the context of metasearch engine research means that a system or a user selects one or more appropriate databases from a limited set of databases, which a metasearch engine handles, while our approach covers all kinds of databases including unknown databases in a distributed environment. In short, our approach is different from the metasearch engine approach from the viewpoint of discovering a database.

Iwasawa et al.[13] developed a system to organize and share users' know-how of reference service in a library. Their approach is similar to ours for sharing reference tools including databases used in Q&As, but their system focuses on sharing know-how and does not have a feature for reusing fundamental information.

There are some studies on designing and sharing collection metadata[14][15]. Collection metadata describes library collections, museum collections, digital archives, and search services such as Yahoo! and Altavista. Collection metadata describes databases as well. The collection metadata approach is similar to ours in the sharing of fundamental information about collections. But collection metadata does not cover users' know-how.

3 Sharing fundamental information

Using our system allows users to share various information about databases with each other. A user of our system is able to find suitable databases for his/her information needs through the shared information, to understand how to use the databases, and organize that information from his/her point of view.

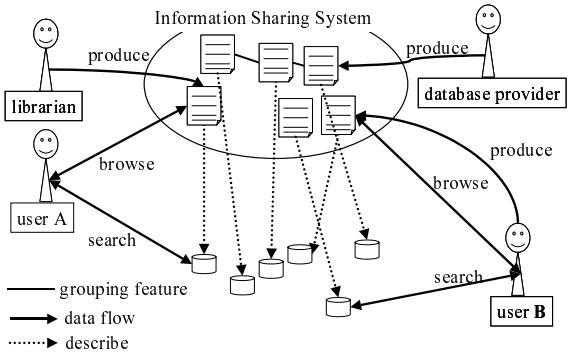

Figure 1: Overview of sharing fundamental information

Figure 1 shows an overview of our system for sharing fundamental information. Using our system allows users to cooperatively input and share information about databases. For example, if a librarian wants to provide a list of available databases in his/her library, he/she can extract only the desired information and make the list. If a provider of databases gives out fundamental information about their databases, users of the databases can acquire that accurate information more quickly, and the provider can benefit as if it were an advertisement. Furthermore, because all users of our system can add new information, this information becomes available to all users.

4 Fundamental information about databases

Fundamental information which is contained commonly in most databases gives users sufficient information about the database, such as the name of the database, the name of the system, the database producer's name, the area covered by the database, the number of records in the database, the language of the records, the subjects of the database, and the terms of use. We designed 21 items as fundamental information.

In order to be able to reuse data, the fundamental information is described as Extensible Markup Language (XML) data. Figure 2 is an example of actual data, which is derived from Gale Directory of Databases[2]. We have used 3,614 records from databases in Japan, derived from Database Daicho Soran[3], as a sample data set for our system.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE database_metadata SYSTEM
"db.dtd">
<database_metadata>
<id>9394</id>
<source_id></source_id>
<created_date>2004-01-09T18:42:42
</created_date>
<update_date>2004-01-09T18:46:55
</update_date>
<userid>sakura</userid>
<dbname>ERIC</dbname>
<system></system>
<condition>RIE archival file, $900; annual
subscription with monthly updates, $900.
</condition>
<format/>
<contributor>U.S. Department of Education:
Office of Educational Research and
Improvement(OERI)</contributor>
<description>Contains more than 1 million
records from both the journal and report
literature in the field of education and
education-related areas.</description>
<subject>education</subject>
<type/>
<lang>English</lang>
<period>1966 to date</period>
<total></total>
<interval>Monthly</interval>
<interval_num></interval_num>
<region>United States</region>
<category/>
<access></access>
</database_metadata>
```

Figure 2: Example of data

5 Commentary information about databases

Another type of information in databases is the commentary information about the databases. This information is a kind of know-how, such as tips for search techniques, grouping information from various points of view, and so on. Because the commentary information is generally heterogeneous, it is difficult to reuse it for any other purposes. However, this information in itself is often useful, and sharing it is useful as well. Nevertheless, if a user knows only the commentary information for a database and does not know the fundamental information for the database, he/she cannot access the database. Because of that, it is necessary for users to know the fundamental information as well as the commentary information. In other words, it is effective to refer the commentary information and the fundamental information complementarily.

In our system, users can browse and input both the commentary information and the fundamental information. Specifically, the grouping feature in our system allows users to bring together similar databases on a subject. The grouping information allows users to browse the fundamental information from various viewpoints. Furthermore, users are able to communicate with each other, since the commenting feature in our system allows any comments, tips and Q&A for each database.

6 System features

Our system has the following features: browsing, search (field specific or full-text), user authentication, input and updating of fundamental information, comment, and grouping.

6.1 Browsing

The browsing feature mainly consists of brief representation and full representation of the fundamental information and the other information, and shows these to users. The brief representation is displayed in groups of 20 records. It shows IDs, names, descriptions, and subjects for each database. Figure 3 shows an example of a brief representation of the fundamental information, which is the top page of the system and sorted by the last date modified. A full representation displays all the fundamental information for a database, comments on the database, and the grouping information for the database. In the full representation of the database, keywords and subjects are hyperlinks to brief representations of records which have the same keywords or subjects. Figure 4 shows an example of a full representation of the fundamental information.

6.2 Adding and updating the information

There are three methods for users to add or update the fundamental information in the system: newly added input, updated input, and imported input. A newly added record is created from a scratch. An updated input means that a user updates an existing record. The user can modify a part of the record and register it into the system. An imported record means that a user exports an existing record and inputs it into the system as a new record.

6.3 Comment

The commenting feature allows users post a comment on the fundamental information for each database. This feature is similar to “BBS” feature. By using this feature, users can discuss and share the commentary information about databases.

6.4 Grouping

The grouping feature allows users to relate and organize multiple databases around each other. In our system, a group consists of four components: the name of the group, a description of the group, a set of databases which belong to the group, and the ID of the user who has created the group. A user can independently create a group and bring databases together to the group. Since a set of grouped databases is linked from the full representation of the database

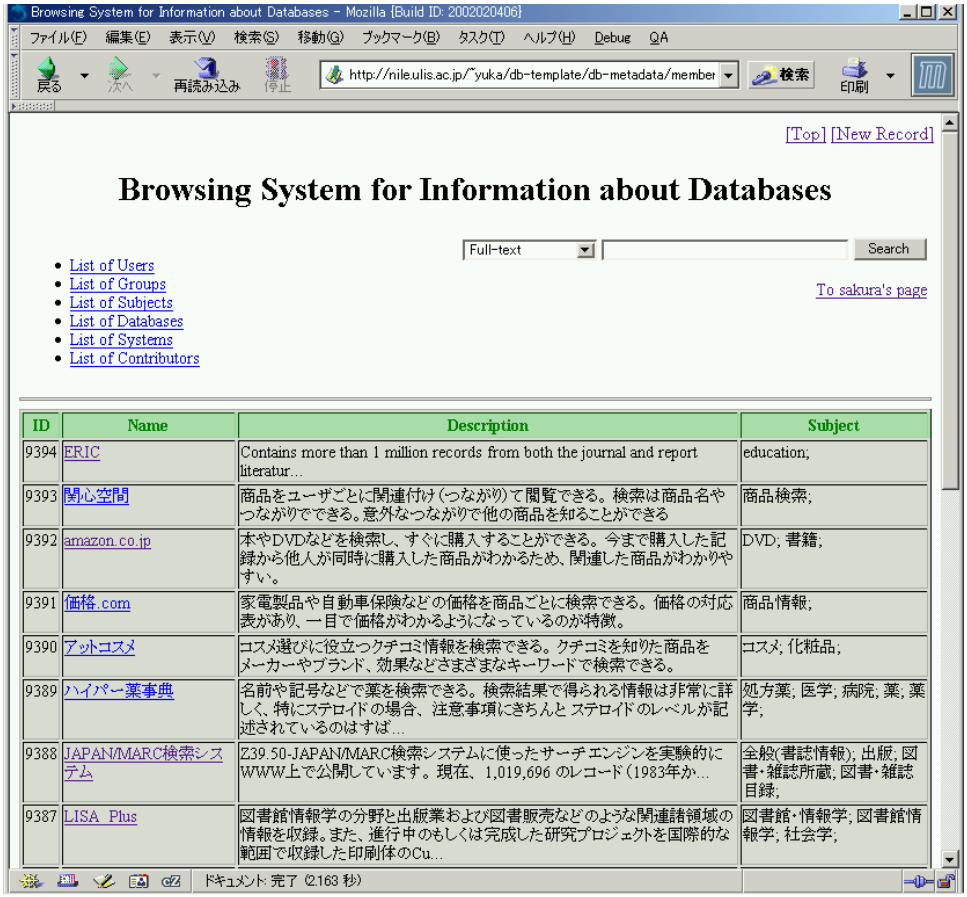

Figure 3: Brief representation of the fundamental information

regardless of the person who creates the database and the group, users can browse and share the organized view of databases with each other.

Figure 5 shows an example of a user's page. In this page, the user can add a new group, edit the grouping information and browse it.

6.5 User authentication

The user authentication feature allows our system to identify a user and get the user's permission by using a login ID and a password. In our system, a user who does not have a login ID is a guest user and a user who has a login ID is a login user. Guest users are able to browse all the information but do not edit the added information. Only login users are able to add fundamental information and commentary information and to update them. Only the user who creates them can update the fundamental information or grouping information. Since each user has a different permission to edit a record, the user interface of the system changes accordingly. For example, if a login user visits the fundamental information which is created by him/her, a link to edit the information appears.

6.6 Development environment of the system

Our system is constructed under the framework of Common Gateway Interface (CGI) on the Web. Our system stores the fundamental information as an XML format, builds the indices from it for the field search, and stores the commentary information as a plain text format. When users search or browse in the system, CGI programs of the system access these data files and display them properly. When the system displays a brief or full representation of the fundamental information, the system converts the XML file into an HTML format by using XSLT stylesheets[16] and returns it to the browser.

7 Discussion

This section describes related works and possible usages of the system.

7.1 Related works

This section describes related works from three perspectives: community-oriented systems, shared cataloging system and OAI systems.

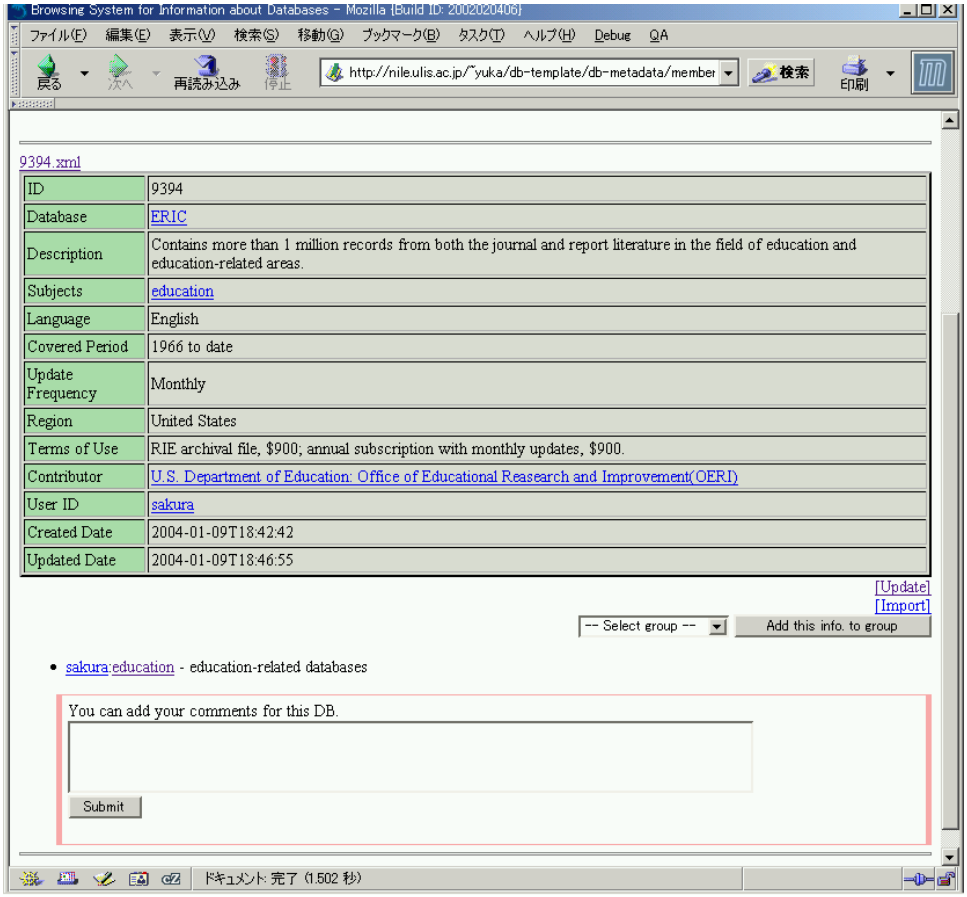

Figure 4: Full representation of the fundamental information

7.1.1 Community-oriented systems

There are several network community oriented information sharing systems. Community sites such as Amazon.com[17] is one such system. Community sites are the sites in which the fundamental information about special topic is stored and users discuss the information. These sites are similar to our system in terms of sharing the commentary information, but such sites which can add the fundamental information and can reuse it are not found. Another example of a community-oriented system is Wiki[18], which has capabilities for users to create communities in a free way. Wiki serves as a site in which visitors freely browse, create and edit pages. But Wiki has no capabilities for reusing the information in it unlike our system has.

7.1.2 Cooperative (shared) cataloging system

A shared cataloging system like OCLC[19] is another example of an information sharing system like ours. Shared cataloging system helps a participating library make catalogs of books and share them with each other. If one library makes a catalog (fundamental information on a book) by using this cataloging system, any other libraries can import it into their own library system and avoid duplicating the cataloging of the book. The concept of sharing the fundamental information intended for books is widely employed.

We extended this concept for databases to which share fundamental information. In this case, databases are different from books. Since books are physically fixed media, once the fundamental information for a book is inputted, it is almost never modified. On the other hand, there is a tendency to add new data to a database, update data and migrate the system for that database. Also, databases are offered in many formats such as through networks or on CD-ROM. For these reasons, it is difficult to identify a database.

In our system, the following approaches are taken in order to solve the problems described above. Users do not identify a database and can make multiple records for one database. Instead, users can organize databases as a group if they seem to be the same database. If there is imported fundamental information from a database, the source of the database is displayed.

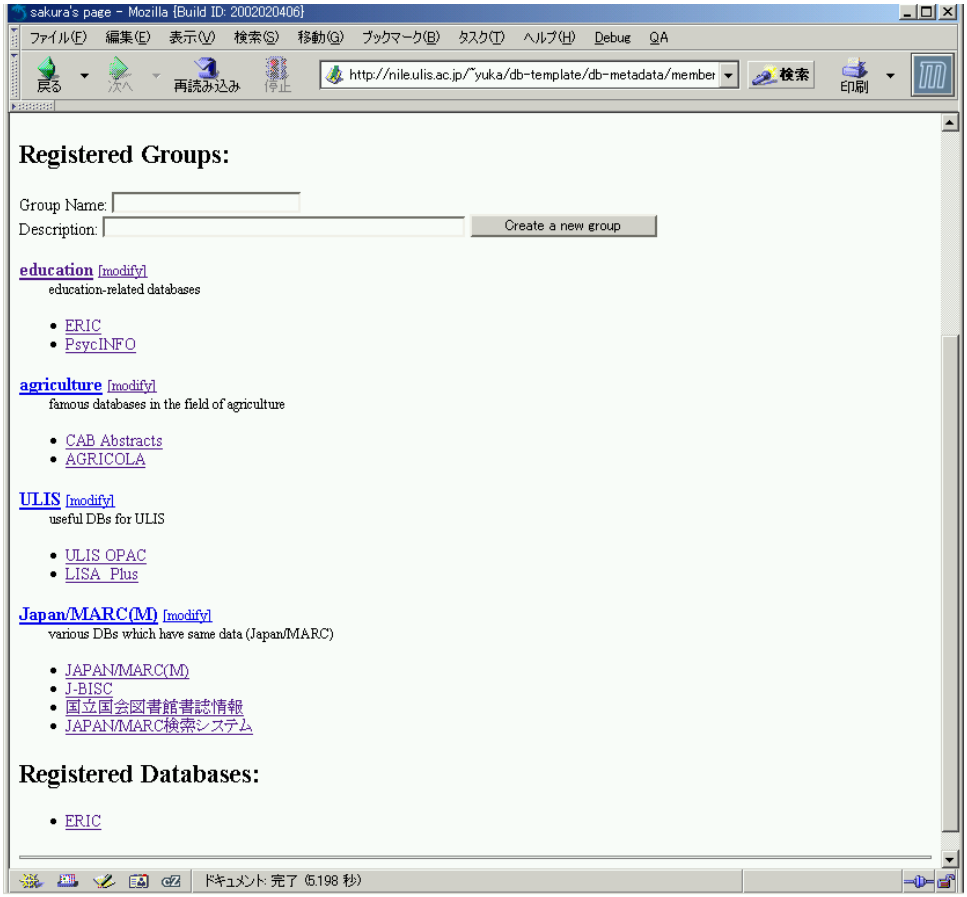

Figure 5: User's Page

7.1.3 OAI-based systems

The approach of OAI[20] protocol-based systems is to enable access to Web-accessible material through interoperable repositories for metadata sharing and aims for a low-barrier interoperability solution to access across heterogeneous repositories. Its approach is similar to our system in terms of sharing information, but our approach is to share information including that which is not yet digitized, which each individual user has, while OAI's approach aims at sharing already digitized repositories.

7.2 Possible usages of the system

When some individuals and/or organizations share information about databases, there are various ways of managing it. For example, there are some restrictions on permission for user's login, some rules for adding new information, etc. Since the policies of the operators of the systems vary, there is no single best way of management. Each organization which runs a system has to decide on its own policy. While our system employs a simple policy which allows casual users to edit records freely, the system can support various policies without any modification. In this section, we describe things related to types of system management. First, we describe policies for the system which should be considered. Next, we describe capabilities of our system which support those policies.

Our system policy is seen from two viewpoints. First, there is a trade-off between quality and quantity of information in creation of information. There are a lot of factors in the trade-off; e.g. limiting users who are able to add information to the system; creating guidelines for adding information; checking the information added; creating a controlled vocabulary. In this paper we do not discuss the full combination of factors, but we do think that adequately setting a level for the factors and combining the factors to control the trade-off well will result in better information sharing. Next, from the viewpoint of use of information limiting users in various ways would also be effective. This is discussed below.

We describe capabilities of our system in relation to the factors discussed above. In order to increase the quality of information limiting users who are familiar with databases would be expected, or in order to increase the quantity of information granting anyone, including guest users, to add information would be expected. In our system, anyone including guest

users can browse fundamental information and login users can browse and add fundamental information. Furthermore, even if there is a request to restrict access to local users of their own enterprise or a request to restrict access to users familiar with databases, our system supports those requests without any modification. Consequently the information about databases can be shared at various levels depending on the policy of the organization which runs the system.

Some guidelines for adding information can be created: e.g. defining the databases to add; or specifying each item to describe. Since there are various ways of integrating a guideline into the system, whether to customize our system or not depends on requirements of the guideline.

Checking added information means that a checker checks and/or corrects this information. In our system, a checker is able to point out another user's errors by using a commenting feature. In our system, since a checker who is not an author of the fundamental information can not modify the fundamental information directly, it is necessary to modify our system to enable a checker to modify the other user's information.

In order to maintain the secrecy of databases used in a company, or their usages, users of our system should be limited within their own companies. Because our system is developed under a framework of WWW/CGI, it supports its request without any modification, the only exception being the configuring of a setting on the WWW server where the system runs.

8 Conclusion

We have developed a network community-oriented information sharing system for databases. We designed items required to describe the fundamental information about databases and developed the system where users are able to browse and add this information. Furthermore, in the system users are able to browse and add the commentary information for databases. Consequently, users are able to share information for databases, easily discover suitable databases and use them.

Acknowledgments

We appreciate the anonymous reviewers for their valuable comments. We express our thanks to Prof. Fred Anderson for his kind advice about the English on our paper.

References

[1] Bergman, Michael K.: "The Deep Web: Surfacing Hidden Value". The Journal of Electronic Publishing, Vol.7, No.1, 2001. <http://www.press.umich.edu/jep/07-01/bergman.html> (accessed 2004-01-12).

[2] "GALE DIRECTORY OF DATABASES", Detroit, Gale Research Inc., 1336p., 2000.

[3] Ministry of Economy, Trade and Industry: "Database Daicho Soran, 2000", Japan, Database Promotion Center, 1035p., 2001. (In Japanese).

[4] National Institute of Informatics: "Database Directory", <http://www.nii.ac.jp/ir/dbmember/dbdr-j.html> (accessed 2003-12-08) (In Japanese).

[5] National Diet Library of Japan: "Dnavi", <http://dnavi.ndl.go.jp/> (accessed 2003-12-08) (In Japanese).

[6] Shulz, Hedda; Georgy, Ursula: "From CA to CAS online : databases in chemistry", New York, Springer-Verlag, 311p., 1994.

[7] Zahir, Sajjad; Chang, Chew Lik: "Online-Expert: An Expert System for Online Database Selection", Journal of the American Society for Information Science, Vol.43, No.5, pp.340-357, 1992.

[8] Morris, Anne: "Online company database selection: an evaluation of directories and CIDA (an expert system)", Journal of Information Science, Vol.20, No.4, 1994.

[9] Kahle, Brewster; Medlar, Art: "An Information System for Corporate Users: Wide Area Information Servers", Online, Vol.15, No.5, pp.56-60, 1991.

[10] Meng, Weiyi; Yu, Clement; King-lup, Liu: "Building Efficient and Effective Metasearch Engines", ACM Computing Surveys, Vol.34, No.1, pp.48-89, 2002.

[11] Manber, Udi; Bigot, Peter A.: "The Search Broker", Proceedings of the USENIX Symposium in Internet Technologies and Systems, Monterey, California, 1997-12, USENIX, pp.231-239, 1997.

[12] French, James C.; Powell, Allison L.: "Metrics for Evaluating Database Selection Techniques", World Wide Web, Vol.3, No.3, pp.153-163, 2000.

[13] Iwasawa, Mariko; Yoshida, Azumi; Ohshita, Shoji; Nakanishi, Yoko: "A Knowledge Management System for the Question & Answers at the Reference Counters of Libraries", Journal of Japan Society of Information and Knowledge, Vol.12, No.2, pp.37-48, 2002. (In Japanese).

[14] Hill, Linda L.; Janee, Greg; Dolin, Ron; Frew, James; Larsgaard, Mary: “Collection Metadata Solutions for Digital Library Applications”, Journal of the American Society for Information Science, Vol.50, No.13, pp.1169-1181, 1999.

[15] Powell, Andy; Michael Heaney; Lorcan Dempsey: “RSLP Collection Description”, D-Lib Magazine, Vol.6, No.9, 2000. <http://www.dlib.org/dlib/september00/powell/09powell.html>.

[16] Clark, James: “XSL Transformations (XSLT) Version 1.0”, World Wide Web Consortium, 1999, REC-xslt-19991116, <http://www.w3.org/TR/xslt>.

[17] “amazon.com”, <http://www.amazon.com/>（accessed 2003-12-08）.

[18] Leuf, Bo; Cunningham, Ward: “The Wiki Way: Quick Collaboration on the Web”. Addison-Wesley, 440p, 2001.

[19] OCLC Online Computer Library Center: “OCLC Online Computer Library Center”. <http://www.oclc.org/>（accessed 2003-12-08）.

[20] Liu, Xiaoming; Brody, Tim; Harnad, Stevan; Carr, Les; Maly, Kurt; Zubair, Mohammand; Nelson, Michael L.: “A Scalable Architecture for Harvest-Based Digital Libraries”, D-Lib Magazine, Vol.8, No.11, 2002. <http://www.dlib.org/dlib/november02/liu/11liu.html>.