A Scheme for cross-searching of Multimedia Databases : Toward sharing and joint Creation of cultural Resources

2.3 A Scheme for cross-searching of Multimedia Databases: Toward sharing and joint Creation of cultural Resources

Masatoshi Kubo
National Museum of Ethnology

2.3.1 Introduction

A huge number of web pages are now open to the general public, and it is now possible to access not only textual, but multimedia databases as well. Thus, the possibilities of widespread use of myriad information resources are increasing rapidly. The potential for new ideas and discoveries that exceed the traditional boundaries of disciplines is increasing due to cross-searching of databases used only in individual, specialized fields in the past. The understanding of different cultures is now an increasingly important theme of the twenty-first century. Therefore, it is imperative to establish cross-searching methods which will allow the opening of information on cultures, namely cultural resources, to the general public and sharing of them. If this goal can be achieved, it will contribute not only to the understanding of different cultures, but lead to discovery and accumulation of human wisdom as well.

The Minpaku Multimedia Information Retrieval system (MMIR), a system for cross-searching of the series of multimedia databases built within the Museum to the present, is already being developed at the National Museum of Ethnology. However, this system is limited to use within the Museum, and the indexing items and search method of the database are unique to the Museum. Therefore, this method cannot be applied to cross-searching of the Internet. This paper attempts to exposit the technical issues necessary for cross-searching of the myriad databases available on the Internet.

2.3.2 Common Formats for Databases

A number of studies on schemes that allow cross-searching of databases have been undertaken. These schemes can be broadly divided into the common indexing method and metadata method.

The former method attempts to provide common indexes through standardization of the indexing items, and it is an example being studied by the Japanese Council of

Science Museums. Natural science museums tend to have highly common indexing items due to the nature of the field, and that strength has been capitalized on to promote standardization. However, human science museums have been establishing items in accordance with the specialized nature of archeology, history, folklore, ethnology, and other fields, and establishment of common items is difficult. Although there have been many proposals to the present, many of these proposals have ended in unproductive arguments, and there are few successful examples.

A function which prepares standardized metadata for data items and maps the items of databases into standardized metadata has been proposed to solve this problem. This approach is the metadata method, and Dublin Core (DC) is the representative metadata. This approach establishes a total of 15 items as the minimum common elements for efficient retrieval of the different information resources on the Web. These 15 items are classified into; (1) those relating to the contents such as title, subject, description, type, source, relation, coverage, (2) those relating to the intellectual property such as creator, publisher, contributor, rights, and (3) those relating to the instantiation such as date, format, identifier, language. Although it is possible to expand this DC by various specialized areas, there has not been adequate consideration of how to map the different expanded items into standardized metadata. Nevertheless, if agreement could be achieved within specialized areas, common items could be established among each area. In short, it is possible to realize a DC divided by common universal portions for all databases and specialized portions common to field-specific databases.

Additionally, a thesaurus function is also required in the mapper which maps the values of database items into the DC. Not only must differences in terms between languages and specialized fields be assimilated, but historical dictionaries for conversion of the calendars of different cultures, gazetteers reflecting historical changes, and other functions are also required in thesauri. There are already many examples of thesauri open to the public on the Web.

2.3.3 Common Format for Search System

The Z39.50 has been proposed as a common standard of the search protocol in the On-line Public Access Catalogues (OPAC) of libraries. In the past, OPAC searches employed different search methods that varied according to the databases of libraries. Therefore, OPAC searches were difficult for users when they tried to access various OPACs. Consideration of this problem led to the development of Z39.50, an international standard which defined the search queries, search results, charges, authentication, and other functions required by information retrieval systems. It is a server-client system independent from the software and hardware of database systems. Unlike the Web, intermediate conditions are saved during searches, and it has other

distinctive characteristics. Therefore, searches can be performed using the same method even for different systems and databases if databases (servers) and searchers (clients) have interfaces compatible with this standard. Information retrieval systems employing this standard are diffused in Europe and the United States, and it is extensively employed for mutual use of the OPAC of libraries. However, there are still few examples of this in Japan. The distinguishing characteristic of Z39.50 also allows simultaneous cross-searching of the Internet. Therefore, many experiments in cross-searching are already being conducted.

2.3.4 A Scheme for sharing of Information Resources

As a method for sharing information resources dispersed on the Internet, the following two methods have long been considered; (1) building of central servers in which information resources are concentrated, and (2) building of clearing houses in which locations, data access methods, and other metadata are concentrated. The National Institute of Informatics in Japan and other information centers fulfill the aims of method (1), and GDM also can be said to strive for this objective. However, the loads of central servers are extremely high under this method, updated results of individual databases are not immediately reflected, and there are other deficiencies. In method (2), individual organizations with databases open to the public on the Internet register metadata in the clearinghouses. Users first search a clearinghouse to learn which types of databases are offered on the Internet. Searches based on this method are also quite common. The Electronic Cultural Atlas Initiative led by the University of California is a prime example. However, this method does not necessarily target cross-searching, and there is the demerit of huge burdens for clearinghouses.

Unlike methods that entail major responsibilities such as the central sites outlined above, a method that allows users to directly access distributed databases is cross-searching method that combine the international standards of DC and Z39.50 described in parts 2 and 3 respectively. Users of client systems compatible with this standard can essentially access all databases in the same manner as using a single database. The experimental unified database system for Japanese classical literature led by the National Institute of Japanese Literature also employs this strategy.

2.3.5 Open Forum-type Database Systems

However, not only can users easily access data, but users will be able to participate in database generation and joint efforts from the user side if there is a function for adding comments or data to accessed data records. In the field of ethnology in recent years, it has been necessary to ensure collaboration between ethnologists and informants to eliminate the asymmetry between ethnologists who describe and site people who are

described. There is huge significance in realizing this collaboration in ethnographical databases. I would like to propose addition of a data-writing function in Z39.50 to make this possible. Of course, authentication of the writer, protection of the intellectual property rights of contributors, and other functions will also be necessary.

The following effects will be achieved through realization of this kind of open forum-type database system.

(1)Movement from Exclusive Control by Specialists to Sharing and Joint Creation
    Now, as the haughty claim that only ethnological researchers have specialized knowledge is being refuted, site researchers, site participants, researchers in other fields, and non-specialists are achieving reciprocal results in collective forums for joint efforts and the sharing of knowledge.

(2)Turnover of Research Ethics
    We are in transition from the past method in which researchers unilaterally engaged in cultural expropriation from sites to research method that protect the intellectual property rights of sites. A trend from unilateral expropriation to sharing with sites, joint information collection, and formation of common intellectual properties of humankind will be established.

(3)Upgrading of Information through Sharing Information collection and processing is being advanced by "point-of-field method," which is a site information processing method based on joint cooperation with site participants. This method will stimulate the discovery capabilities of users. Additionally, information in the past ethnographies with low accuracy will be upgraded by feedback of information from the present sites.

(4)Cultural Restitution to Sites Sharing of information collected in the past with sites will contribute to the restoration, inheritance, and revival of lost site cultures. In these times, rapid changes are occurring in existing cultures along with the progress of globalization, and groups working to restore cultures that are being lost are currently increasing as a method of representing and asserting one's own cultural identities. Thus, this will emerge as a major channel for returning research results to the related sites.

(5)Discovery of Human Wisdom through Constructing Knowledge-bases
    If the data mining method is applied to ethnographies collected from those open forum-type databases, which are information treasuries of the actual world, then there is also the possibility of building ontology of human wisdom as a conceptual model of human knowledge.

(6)The Effects of Fusion of the Human Sciences and Natural Sciences

Ethnographies perceived only as ethnological research materials include information that contain the seeds for solving environmental, developmental, peace, and other global problems. The sharing of ethnographic information by people in the fields of social science and natural science will become a bridge for academic fusion of the human sciences and natural sciences.

This open forum-type of database system is a vehicle for joint creation of an encyclopedia of human wisdom, and application to educational fields is also expected through development of educational materials. This system will also contribute to the cooperation between museums and art galleries as well. The idea of allowing writing by users has already been incorporated into the experimental GDM system. Open forum-type database systems that have further developed this concept in GDM have an orientation of supporting the sharing and joint creation of databases, and will unquestionably become a model of functions required to the documentation and information centers that have to meet the social demand in the near future.