

Beginning and Development of Digital Library in National Library of China

Sun Chengjian
Researcher and Deputy Director
National Library of China

Zhen Xihui
Chief Engineer
Wenjin IT Research Center
National Library of China

Abstract:

Digital library is a resource system of digital information by using modern high and new technology. It will become the managerial model of information resources for the next-generation Internet. This Paper summarized the research project since 1996 and the digital library experimental project by National Library of China.

Keywords:

Digital library, National Library of China, SGML

With the rapid development of modern high and new technologies, Internet regarding computer technology and network communication technology rose almost overnight. Modern society began to take pace into an information age after experiencing agricultural society and industrial society. At the same time, digital library appeared, representing a revolution and renovation in the history of library.

1. Background

Digitization and great development of Internet led to globalization of economy, commerce and information dissemination, bringing deep effects on all levels of society. Mass information gushes into Internet. However, it becomes more difficult to search information on Internet because much information is in chaos. There is an urgent need to

create a new model to manage information resources. This is the background of putting forward the concept of digital library.

In a technical sense, digital library is a resource system of digital information by using modern high and new technology. It will become the managerial model of information resources for the next-generation Internet. It will make the scattered information on Internet in order.

Digital library is based on unified standards and criterions. Distributed mass repositories support it. It is dealt with intelligent retrieval technology and managed by e-business. In addition, information will be transferred to a large number of households and families via broadband high-speed networks. It involves the whole process of producing, processing, storing, searching, transferring, protecting, using, documenting and deleting of digital information resources. It would become difficult for an individual library to do it without cooperation between libraries, museums, art galleries and archives home and abroad.

2. Research and Development of Digital Library in National Library of China

National Library of China attaches great importance to the global research and development of digital

library. Since 1995, some libraries were been specially assigned for following up the scent of the development of digital libraries in the world, and made relational technical research and experiments.

2.1 Project for Experimental Digital Library

In early 1996, National Library of China submitted "Project for Experimental Digital Library" to the Ministry of Culture.

The project began with the image data pool of Chinese Doctoral theses by using client/server model. It uses the bibliographic data server to manage the index and inquiry of data, and the image data server to manage the digitized information. The scanning images adopt 300-dpi resolution ratio, and compress and decompress them according to CCITT Group 4 Standards. The images are stored in JukeBox. It realizes retrieval by establishing multi-level indexes and multiple repository connection and is able to provide service via networks.

2.2 Research and Development Project for Feature-based Multimedia Information Search System

This is a co-operative project by National Library of China and the Computing Technology Institute of the Chinese Academy of Sciences. It has passed the technical authentication.

The project mainly makes research on feature-based image information search, brings about the search according to the textures of images, colors and figures, the search of full-text Chinese search and the research of information memory management methods, realizing cross-platform search at the customer end.

2.3 Application of SGML in Library

The Ministry of Culture appointed this project in 1997. It was undertaken by Wenjin IT Research Center of National Library of China and jointly developed with the Computer Institute of Peking

University. SGML (Standard Generalized Markup Language) is a standard for information processing issued by the International Standard Organization in October 1986. At present, it has more than twelve relative international standards. Since 1994, the State Technology Supervision Bureau has carried out three relative national standards, which were at first used in the field of information processing in publications.

Currently, libraries in China generally provide readers with OPAC (Online Public Access Catalogue) system. Readers can find books by searching literature's title, author, call number and/or subject. With the development of Internet, it is far from enough just to search bibliographies. We decided to present information in SGML format at first in the processing of Chinese book catalogues, with the goal in mastering SGML and adopting SGML-supported search engine for fast response to users' variant requests, and to improve the performance of the OPAC system. More importantly, our potential aim is to meet new demands under network environment, making a sound foundation for the repository construction of digital library.

In the past two years, National Library of China and its Wenjin IT Research Center, with the cooperation of the Computer Institute of Peking University and referring to the usage of SGML in other countries, have developed an SGML-based library application system. Users now can access to data in the OPAC system via Internet browsers.

National Library of China began to provide readers with this system in January 1998, and has good effects. The system was approved as a regular service system in July 1999.

2.4 Project for China's Experimental Digital Library

In May 1996 National Library of China put forward the Project for China's Experimental Digital Library, and submitted it to the State Planning and Development Committee. This project was

appointed in 1997 as one of the state essential technology projects. National Library of China leads it. Shanghai Library, Liaoning Library, Nanjing Library, Zhongshan Library and Shenzhen Library joined in this project. The project includes several databases with overall and uniform technology. These databases are: Image Database of Chinese Ancient and Rare Books, Image Database of Chinese Doctoral Theses, Database of Historical Atlas and Pictures, Global Text Database of Sun Yat-sen's Literature, Literature Database of Shenzhen Special Economic Zone, Multimedia Database of National and Overseas Tourism, Literature Database of Nanjing Government under the Republic of China (1911-1949), Graphic Catalogue Database of North-eastern China, and Literature for International Digital Library.

The project plans to develop and complete a set of performance technologies of digital library keeping the track of international standards. The project plans to:

- Establish SGML-based object description and editing system;
- Organize digital object repositories with open architecture and establish general access protocols;
- Establish uniform metadata scheme and metadata set;
- Make experiment to establish a handle system for the distributed digital object with different sources;
- Design and realize convenient user's interface of network and practical interface of system management;

At the same time, this project also plans to finish the report "the Concept and Development Research of Digital Library", making comprehensive survey and study from theory to practice and from technology to application, and analyzing the concept of digital

library, the major relational technologies, tendency of foreign countries, vista of application and development, effects and influence on society and economic results in an all-round way.

Now the participated libraries and companies have already made some progress in relational research and development.

2.5 Knowledge Network—Systematic Project of Digital Library

It belongs to the State 863 Project of 1998. It will be completed with the cooperation of National Library of China and Beijing Shuguang Tianyan IT Corporation Limited.

This project aims at designing and developing the basic level architecture of digital library. It will be built on Internet environment. Its architectural structure includes several distributed digital repositories. Artificial intelligence technology will be adopted focusing on rapid inquiry of several cross-repositories.

The system puts emphasis on the practical functions of the next-generation Internet and integrates several high-tech achievements of the State 863 Projects, including intelligent agent software and protection of intellectual properties. In addition, we will accept and use some works of other countries.

2.6 Research Topic on Group of Digital Libraries in Zhongguancun High-tech Park

Zhongguancun in Beijing boasts China's largest hub of talents and intellectual resources, taking advantage of foreign experiences and expertise. By combining its rich intellectual resources and high-tech industries, Beijing starts another round in its campaign to develop a knowledge-based economy, turning the abundant intellectual resources into productive forces. Zhongguancun's rapid development is significant to Beijing's ongoing restructuring of its industries, as well as technological innovations of the whole country.

Officials in charge of Zhongguancun high-tech park holds that the construction of digital library is a basic project for information digitization in this area. National Library of China, situated in the high-tech park, takes main responsibility for this project. In July 1999, two projects approved.

The first project: *Research Topic on Digital Library Group in Zhongguancun High-tech Park* discusses the methods and applicable standards of the digital library construction by taking research and investigation of the current resources. Moreover, in the end it must put forward a preliminary and feasible plan. National Library of China is collaborating with Peking University, Tsinghua University and Chinese Academy of Sciences in the project.

The other project: *Development and Strategy Research of Digital Library* focuses on researching the development of digital libraries in the world, making comprehensive analysis of digital library's definition, major relational technologies, new development trends, vista of application and development, effects and influence on society and economic results. It will also make suggestions in view of global development trends and characteristics for libraries in China.

These two projects are a bridge to small-scale experiment of China's digital libraries. They are expected to be finished by December 1999.

3. Experimental System of Chinese Digital Library in National Library of China

National Library of China has completed the experimental environment and experimental system by end of March 1999.

3.1 Preliminary construction of the experimental system of digital library

We adopted UNIX as the main platform in accordance with large capacity and parallel. The

system is composed of five SUN UNIX servers, several NT servers and PC workstations, taking shape of a distributed experimental environment via high-speed networks.

3.2 Experimental system of Chinese digital library

It includes five multimedia repositories. They are Thousand Poems by Various Poets, Chinese Ancient Buildings, the Forbidden City, the Aquarium World and Universe. It has information with 5GB. Two of the repositories are positioned in the Branch Library 10 kilometers away.

The system is being run on the LAN with 1,000MB-bit, realizing real transfer.

The Thousand Poems Repository digitizes the colorful drawings of the Ming Dynasty, adding watermarks on them by using IBM's technology. There are pictures, notes and reading with music to 36 poems.

The Chinese Ancient Architecture Repository reflects palaces, temples, pavilions, grottoes and the Great Wall with pictures and explications.

The Forbidden City Repository gives us the glory and magnificence of the buildings and its art curiosa with color pictures and explications.

The Aquarium World Repository selects many audio and video materials, arousing people's interest in seas and oceans.

The Universe Repository aims at popularizing basic knowledge of universe, with pictures and explications.

According to the demands of the construction, these repositories are connected to the national bibliographic database and some encyclopedias, realizing multi-repository connection, cross-repositories search and cross-region and cross-platform search.

Through development of the experimental system, we have gained many experience in designing and processing of multimedia digital resources, the extraction of metadata and generation of metadata database, the construction of object database, the selection of handle system and cross-repositories search. We have developed software on the processing, markup and management of data, software on the loading and management of metadata and object database, interface software with the capability of supporting SGML, distributed database search, search software and the connection of metadata and object database, and software capable of generating dynamic pages that users can use Web browsers to browse.

After the completion of the experimental system, leaders from the State Council and the Ministry of Culture and some experts from various libraries came to National Library of China for a visit. They thought highly of it.

By several years of hard work, National Library of China has made great progress in the research and development of digital libraries. Nevertheless, there is still a long way to go to catch up with the developed countries. We will widen our cooperation with companies and institutes both home and abroad to quicken the construction of digital libraries of China.