

Analysis on Digital Library Construction in the Big Data Era

Gemin Li¹, Xiaoguo Liu² and Jing Tian³

¹Jilin Agricultural University Library, Changchun, China

²Jilin Agricultural University, Changchun, China

³Changchun Vocational Institute Of Technology, Information Technology Academy

Keywords: Big data; Digital library; Construction; Resource sharing; Analysis

Abstract. As an important carrier of knowledge-based economy, digital library is an important part of China's information infrastructure and has become one of the important symbols to evaluate a country's information capability. Therefore, construction of digital library is forward-looking and strategic.

Introduction

In this era with rapid development of information and technology, the collection and dissemination means by the traditional library has undergone great changes means, and digital library has become a creative tool to expand the human ability. Under the impetus and support of modern information technology, to realize the digitization of library document resources and the networking of service means is not only a revolution in the information field, but also a new stage of the development of the library industry.

Overview of Digital Library

The origins of digital library research can be traced back to the origin of electronic library. In 1948, American mathematician Wiener first proposed the theory of electronic library, in 1978, the famous American library scientist Lancaster described the prospect of electronic library, in 1979, Harry (British) proposed the concept of "virtual library", while the formal concept of digital library was proposed by Woolf (member of the US National Science Foundation) in his international cooperation white paper until 1988. With the increasing development and maturity of computer science and information technology, digital library have made great progress in the field of theory and practice. In China, the digital library started in the middle 1990s and got relatively rapid development.

Many famous experts and scholars of libraries and literature information disciplines at home and abroad agree that, the digital library is the advanced stage of library automation, but their definition are not the same. The US National Digital Library Federation defines that, a digital library is an organization with relevant resources including professionals, which selects, organizes and provides intelligent access, translation, dissemination, integrity maintenance and permanence of digital resources, so that these digital resources can be quickly and economically used by specific users or groups[1]. The Information Infrastructure Technology and Applications Group of US government noted in its report that, digital library provides critical information management technologies for the national information infrastructure and its primary information base and resource pool. In other words, the digital library is the core of the national information infrastructure. Digital library refers to that all the work processes of library are based on the computer, and library resources are digitalized. Digital library is a large system with distributed, large-scale and organized databases and knowledge bases, in which the users or user groups can make conforming access to the databases and knowledge bases within the system to obtain the final information they need. At present, there is no a complete definition for the digital library at home and abroad, but there is basically an agreement on the understanding even with minor differences.

Digital Library Has Four Basic Characteristics: (1) Fully realization of computer networking on library management;

(2) Library resources digitalization, the organization of digital information and provision of digital information services;

(3) Link to other digital libraries through the network and provision of online search and consulting services;

(4) Resource sharing. The main task of the digital library is firstly to digitize the existing physical library resources, that is, to digitize the traditional literature information such as books and audio-visual materials in the library so that the users can use the library's information through the network in any place, and secondly to realize ordering and collection of the existing digital resources, that is, to effectively organize the current digital information resources on the network, so as to efficiently meet the user's needs and better share humans' knowledge resources[2].

Problems in the Construction of Digital Library in China

One-Sided Understanding on the Digital Library Concept and System Structure. From the start to 2001, the theory research and construction of China's digital library basically stayed in the theoretical exploration and experimental exploration stage. After 2011, China entered a large-scale practical stage and the digital library has got real development. Especially in recent years, the voice of digital library in China gets higher and higher, but some people have a vague understanding of the digital library concept and a simple one-sided understanding that the digital library is only digitalization of library materials, so they simplify the digital library, a highly systematic project with a wide range to the scanning of books and reference materials. Taking Shanghai Library as an example, in 2010, the procurement expenditure of digital resources accounts for 7.91% in that of the total literature information resources, and it increased in the succeeding years and reached 30.01% by 2015, which shows the growth of digital resources. The librarians should strive to understand and grasp the development trend of digital resources, and actively meet and lead the digital reading, especially the demand for mobile reading. In fact, the digital library should be a national digital cultural platform and digital education platform, namely, a national digital resource center. It can be seen that, the digital library is not only a digitalized library, but also a communication media of culture, a network business platform for cultural products, the basis for the organization, development and utilization of national digital resources, a network culture center and the distribution center of network culture. Digital library construction is to distribute the digital library to the area with needs, and to make the digital library enter tens of thousands of households.

To Be Improved Quality of Staff and Users. (1) The quality of the staff is the key factor to ensure the normal development and operation of digital library. It not only requires the staff to master the traditional library science, information science theory and technology, but also pay attention to the learning and training of the basic theory and practical skills of modern technology, and in particular, to improve the application of computer technology and network, in order to qualify themselves in the work including hardware management and network maintenance in digital library. As the Internet contains a wealth of foreign language information, the staff must also have a high level of foreign language ability to complete the network services in the new era, and be proficient in intellectual property (copyright) protection, network security knowledge and maintenance knowledge.

(2) The digital library environment requires higher requirements for the overall quality of the users. Highly educated users can be proficient to obtain the information they need through the network, but some users are far away from such requirements since they can not master the computer, do not have sensitive information awareness and can not conduct information retrieval by computer. Therefore, it is difficult for these users to get access to in the network environment freely.

Technical Issues in the Digital Library Construction. (1) Literature storage and compression technology. Digital library faces a large amount and many types of data, so how to preserve and manage such data is one of the core design tasks.

(2) Classification, indexing and retrieval technology. There is no uniform standard for the classification, indexing and retrieval of digital library. There is a large amount data in the ocean of

knowledge. It is like searching for a needle in a haystack for the users to obtain the necessary information, thus information retrieval and application is a big problem.

(3) Server terminal scheduling technology. The speed of server terminal is too low since the export bandwidth is too narrow.

(4) Choice of tools and platforms. This is the biggest challenge that digital library faces, which includes the overall structure standards, hardware and software technology, information entry tools, search tools, knowledge mining tools, etc[3].

Network Information Security. The primary enemy of network information security is computer virus, and the second is that due to the encryption and anonymity of network, it is difficult to find the intruders. The hackers' attack on information resources is not only because of curiosity and self-challenge, and programming experts can tamper system control parameters. The third is espionage activities with theft behavior, which replicate or spread confidential information by spy. The fourth is intentionally damage, destruction of the information processing function of the system or damage of electronic communication devices. These problems often result in data loss, theft, distortion or damage, thus affecting the security of online information services [4].

Copyright issues have not been Properly Addressed. Copyright is the basic problem of digital library construction, and also the main problem which restricts the development of digital library. The core of the digital library is to establish a variety of resources, collect, process and clearing up the copyright works. It is not doubt that to digitize the works and make it into an electronic version is to make duplication. To digitalize the published works involves the right of authorship, modification, protection of the integrity of works, use and receive remuneration. For unpublished works, it also involves the copyright. At present, there is a phenomenon that some people ignore the copyright law and blindly digitize the literature. On the other hand, there is also the phenomenon that relevant provisions about copyright treatment are not sound, which result in various difficult problems in the construction of the digital library.

Development Countermeasures of Digital Library in China

In-Depth Analysis of Digital Library Content, And Strengthening the Training of Staff and Users' Quality. Building a digital library is not a simple task which can be completed in a short time. It needs in-depth understanding of digital library theory to turn the theory into reality and successfully build a digital library entity. The participants of digital library construction must be clear that digital library construction is not just digitalization of library materials, but a process to solve the problems which combines user, content, technology and economy, society and legal environment, and a process to make standard specifications and regulations on resource description, marking, querying, exchange and use. In this project, it is required to make good preparation on technology, capital, talent, system, etc.

Strengthening the Training of Staff and Users' Quality. Digital library is a big project which integrates high-tech and culture. In the process to promote China's digital library, it is needed to continue discovering and cultivating the compound type of informatization talents with both technical knowledge and profound cultural awareness. The staff training needs to be broadened in the means, namely, to adopt various ways and means, and rational planning must be developed. The planning should include short-term, medium-term and long-term goals that can be adjusted during the training to realize optimization. The staff training can be based on the resources of the unit itself, but it also needs to introduce new resources. The first is the training based on local resources: take long-term uninterrupted training by using the unit's environmental and human resources. Fully tap the staff's potential, effectively play the team strength, guide the weak ones by strong ones, make mutual discussion and learn from each other. The second is to going out for further studies, training, etc., to broaden horizons: short-sighted should be avoided on the digital library construction. The unit should strengthen the training capital investment, and it should not stop training and put barriers for digital library construction due to financial problems. The staff's quality can only be comprehensively improved by going out, receiving new theories and participating new practice activities. The third is to invite experts to make on-site training and guidance: The digital library is a relatively new issue,

and it has problems on the theoretical study and substantialization, so timely collection of expert authority information is a necessary way to strengthen the staff's understanding of digital library and improve their quality [5].

Digital library is different from the traditional library, and it has different requirements on the quality of the users, therefore, it is an issue which must be solved by digital library builders to find effective ways and means to train users and improve users' quality in practice. At present, the favorable situation is the popularization of computer science and the quality training of information users in colleges and universities. Therefore, the main problem of user training is to improve their awareness of information quality and make them understand the necessity of high information quality in current society.

Countermeasures to Copyright Issues. (1) To reasonably use the rights set forth in the copyright law and try to use them to the best.

(2) To promote the further improvement of copyright legislation. The operation of digital library and online activities can not be conducted until after the completion of the law, and the current law should be adjusted to adapt to social development. The operation of digital library will also promote also copyright legislation to make it more perfect during the completion of its own goals.

(3) To take appropriate measures to control the copyright management, such as the establishment of online charging system, in which the users need to pay appropriate fees for some digital resources, to pay the author's copyright fees.

(4) To protect the copyright from the prospect of technology. That is, the use of advanced computer technology can effectively prevent piracy and illegal copying [6]. It will not be a far away future for the humans to fully share the world's various types of information resources.

Conclusions

The digital resource revealing platform should realize the analysis of the user's knowledge utilization history and the tracking of reading behavior as much as possible, to capture the reader's interests and hobbies and effectively provide the precise offering of the resources, thus setting up a all-round dynamic knowledge service model from collection, excavation, integration to revealing, promotion and guide, in order to maximize the application value of knowledge resources.

References

- [1] Mary Auckland. Reskilling for research: an investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers [M]. London: Research Libraries UK, 2012.
- [2] Corral, S. Roles and responsibilities: Libraries, librarians and data [A]. Graham Pryor. Managing Research Data[C]. London: Facet Publishing, 2012.
- [3] Virginia Jobs. Posting Details [EB/OL]. [2012-11-25] .https://jobs.agencies.virginia.gov/applicants/jsp/shared/position/JobDetails_css.jsp?postingId=705575.
- [4] Sugimoto CR, Ding Ying, Thelwall M. Library and information science in the big data era: Funding, projects, and future [a panel proposal [EB/OL]. [2015-10-20]. http://www.Ischool.Drexel.edu/faculty/mkhoo/docs/12_asist_panel_description.pdf.
- [5] Qiu Judy, Jha S, Luckow A, et al. Towards Hpc-abds: An initial high-performance big data stack[J]. ACM Transactions on Embedded Computing Systems, 2014, 1(1):1-17.
- [6] Chen Ming, Mao Siwen, Liu Yunhao. Big data: A survey [J]. Mobile Networks and Applications, 2014, 19(2):171-209.
- [7] Chen CLP, Zhang C Y. Data-intensive applications, challenges, techniques and technologies: A survey on big data[J]. Information Sciences, 2014, 275(10):314-347.

- [8] James Manyika et.al. Big data: The next frontier for in-novation, competition, and productivity [EB/OL].[2014-12-17]. [http: //www.mckinsey.com/insights/Mgi/research/technology_and_innovation_big_data_the_next_frontier_for_innovation](http://www.mckinsey.com/insights/Mgi/research/technology_and_innovation_big_data_the_next_frontier_for_innovation).
- [9] Liz Lyon. The Informatics Transform: Re-Engineering Libraries for the Data Decade [J]. *The International Journal of Digital Curation*,2012,(1): 126-138.
- [10] LinkedIn Corpration. Daureen Neddill [EB/OL].[2012-11-25]. [http ://www.linkedin.com/pub/daureen-nesdill/10/6aa/384](http://www.linkedin.com/pub/daureen-nesdill/10/6aa/384).
- [11] InsideHPC.JOB BOARD [EB/OL].[2012-11-25] .[http : //insidehpc.jobamatic.com/a/jbb/job/details/761910](http://insidehpc.jobamatic.com/a/jbb/job/details/761910).
- [12] Zoom Information Incorporated. Chris Kollen [EB/OL].[2012-11-25].<http://www.zoominfo.com/search/profile/person?personId=46360201&targetid=profile>.
- [13] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, The art of writing a scientific article, *J. Sci. Commun.* 163 (2000) 51-59.