

Digital Resources of CADAL: Construction, Utilization and Promotion

Qing Li ^a and Ling Huang ^b

University Library, Huazhong University of Science and Technology, Luoyu Road 1037, Wuhan 430074, China

^aliqing@mail.hust.edu.cn, ^btrs@mail.hust.edu.cn

Abstract. Construction of digital library is the product of currently vigorous development of computer science and quick popularity of network information technology. In order to create a large-scale digital library, Chinese government invested a key national project called as CADAL in 2003. After two building stages, CADAL had been successfully constructed. At present, nearly 2.5 million digitized books and other information resources are available in CADAL. This paper attempts to introduce the general situation and resource utilization of CADAL. To achieve the goal of CADAL project which is to realize the co-construction and sharing of global digital resources, this paper proposes several strategies for construction of CADAL, such as speeding the link among all members of university libraries and the CADAL central system, improving the website function, and strengthening the propagation of CADAL resources.

Keywords: CADAL; Digital resources; Utilization; Global sharing.

1. Introduction

With the development of computer science and popularity of network information technology, digital library is being vigorously developed. In this background, a key national investment project namely CADAL was started in China, in order to create a large-scale digital library which includes various domestic and abroad resources. So far, CADAL project have been built for two stages.

Initially, the full name of CADAL was China-America Digital Academic Library, which was commonly called as the first building stage of CADAL [1, 2]. This project was built from 2003 to 2006, for which Chinese government invested 70 million RMB and USA government invested 2 million dollars. Zhejiang University and graduate school of Chinese Academy of Sciences became the leading institutions of CADAL project while other 16 academic libraries of universities including Peking University, Tsinghua University, Huazhong University of Science and Technology and so on participated in the project construction. After the first stage of CADAL project, one million and twenty-three thousand digital books and large quantities of multimedia resources had been built. This provided a solid foundation for the construction and service of digital library [3].

In the second building stage of CADAL, the full name of the project was changed to China Academic Digital Associative Library [4]. This building stage spanned four years from 2009 to 2012, for which 150 million RMB was invested. The new CADAL project management team was centered at Zhejiang University, who was responsible for the implementation of all applications. Meanwhile, 85 academic libraries of universities and scientific research institutions participated in the construction. The goal of this project was to complete the handling of additional 1.5 million digital resources based on the first building stage of CADAL. Moreover, a distributed data center and service system would be established, in order to ensure data security and provide global services. The total amount of various digital resources of CADAL has reached nearly 2.5 million items to date [5].

2. Types, Quantity and Digitalization Process of CADAL Resources

At present, all kinds of digital resources of CADAL project including liberal arts, science, engineering, agriculture, medicine and miscellaneous information are integrated in a center website of CADAL (<http://www.cadal.cn>). This website provides a strong support for the teaching and scientific researches of participated institutions. Moreover, it is open access to the whole world, which can promote the cultural exchange between China and foreign countries. According to the statistical data

newly advertised by CADAL center, nearly 2.5 million items of digital resources are available in portal website of CADAL. These digital resources are mainly from the characteristic collection of libraries in domestic and foreign institutions. Especially some characteristic Chinese resources such as Chinese Calligraphy [6, 7] and traditional Chinese medicine documents [8] are available in CADAL system, which may be difficult to be found in other database.

It needs to be mentioned that all CADAL electronic resources are currently stored using the format of Djvu [9]. In this format, picture files can be browsed and downloaded quickly. Compared with other picture compression formats, the compression of Djvu mode is more effective. Thus, the files stored in the format of Djvu have smaller file sizes but higher capacities. Although this format may be not well known to all users, these files can be conveniently browsed as long as adobe flash player, one of the most widely used software, is installed in the users' computers. These CADAL electronic resources are the cooperation results of all participating institutions. In different building stages of CADAL, each participating institution was arranged to contribute to various types of resources.

3. Retrieval and Utilization of Digital Resources

According to the original objective of the project, all digital resources of CADAL would be shared by the participating institutions. Therefore, the website of CADAL is now accessible to all participating institutions within IP range to provide the service of digital resources. These participating institutions include 64 Chinese academic institutions (Zhejiang University, Tsinghua University, Fudan University, Huazhong University of Science and Technology, etc.), 5 American academic institutions (Carnegie Mellon University, UIUC, IA, etc.) and 16 Indian academic institutions (India Academy of Sciences, Anna university, Goa university, etc.) [4]. The service of CADAL digital resources is provided by campus networks. The CADAL system can control the reading permission to users by checking whether their IP addresses are within the IP range of participating institutions. All digital resources of CADAL are free for participating institutions within IP range. If they are not accessible, it is convenient to connect with CADAL center.

When users browse the digital resources of CADAL, they are requested to log in the website at <http://www.cadal.cn>. Therefore, the users should register for the first time. Afterwards, they can log in the website by providing their personal account and password. Because the CADAL system provides the lending service of digital books, the authentication to users' IDs is required. All non-registered users can only browse the homepage to carry out simple information retrieval, but cannot obtain the lending service.

4. How to Effectively Share and Utilize CADAL Resources?

Chinese government had paid a heavy price for collecting dispersive but precious paper resources and converting them into electronic resources. Thus, a treasure of massive digital resources is formed. If these resources are not effectively utilized, they would be like a deeply buried treasure without price, and large quantities of manpower, material resources and funds are wasted. Therefore, it is necessary to discuss how to effectively share and utilize CADAL resources. The framework of personalized cross-media retrieval in CADAL digital library has been proposed and the details of the algorithms used in the personalized cross-media retrieval have been introduced [10]. Moreover, a recommender system for CADAL digital library using a ranking-oriented collaborative filtering approach based on users' access logs has been presented [11]. Based on the utilization status of CADAL, we propose the following three strategies to further promote the effectiveness of utilization of CADAL resources.

4.1 Speeding up Linking among All Members of University Libraries and CADAL Central System.

After the second building stage of CADAL passed the examination in May 2012, CADAL 4.0 was released online in October 2013 to provide free service of digital resources. Although the CADAL system is continuously upgraded and improved and new system version provides better and better experience for users, the utilization efficiency of CADAL is not high because many users are not

familiar with CADAL resources and their utilization modes. To overcome this problem, the CADAL center specially develops an information retrieval system based on an open technology framework namely OPEN API, which can provides service for each CADAL member and their terminal users by realizing a friendly sharing environment of CADAL digital resources across the resources, platforms and systems [12]. At present, the bibliographic management systems of participating institutions in China can be represented by the INNOPAC system of Library of Tsinghua University, the Huiwen OPAC system of Library of Xiamen University, and the ALEPH system of Library of Northeast Normal University. Although those three participating institutions had linked successfully with OPEN API and desirable effect was exhibited, most of participating institutions have not taken an active action. For this purpose, the CADAL center held a training meeting for CADAL service and application system on 19-20, November 2015. This meeting provided a special training of link with OPEN API system for each participating institution. It was required that each participating institution should speed up the link with the resources of CADAL center and effectively utilize the digital resources of CADAL. Moreover, it was expected that more and more participating institutions could complete the system link as soon as possible and the users could experience the convenience of one-stop from CADAL digital resources.

4.2 Improving the Website Function.

The frequency of users browsing a website correlates with many factors: (1) whether the resources are plentiful and useful; (2) whether the interface is friendly; (3) whether the retrieval function is perfect. For CADAL system, the digital resources are abundant, almost like a large-scale library. Moreover, the interface of CADAL is also very friendly, and global users can log in the website and utilize the resources without obstacles. However, the retrieval page of CADAL is too simple. There is only one blurry search box without any limited search function. When users input one word to search the resources, a lot of words appear on the page. They are fussy about the problem that they cannot obtain desirable resources with one step. Therefore, we suggest that CADAL center can modify the simple search page in the website, and add the search items of topic, title, author, publishing house, keywords, full paper, and so on. At the same time, users can increase or reduce the number of search box according to the necessity. Thus, they would more accurately and quickly find their desired literatures. Moreover, the CADAL central website can add advanced search item using Boolean logic algorithms to satisfy the advanced requirements of users.

4.3 Strengthening the Propagation of CADAL Resources and Realizing the Global Sharing.

The statistical results till November 2015 indicated that the visiting IP addresses were from 1130 Chinese universities. Although the visiting IP address range looks wide, the frequently utilized institutions actually focus on several state key universities such as Sun Yat-Sen University, Renmin University of China, Nanjing University and Zhejiang University. The most visiting number of oversea universities is in USA; however, it is only 1/7 of visiting university number in China. This should be due to the fact that many institutions or people do not know this project except the participating institutions. Therefore, it is necessary to strengthen the propagation of CADAL website, CADAL project and CADAL digital resources either in domestic or foreign institutions. Moreover, CADAL can be propagated on various famous websites so that more people know CADAL project and utilize CADAL resources. Eventually, more institutions and organizations voluntarily participate in the CADAL project and upload their precious characteristic literatures. Thus, the goal of CADAL project for co-construction and sharing of global resources can be realized.

Even for the books without copyright, the intensity of opening should be remarkably increased. Since the purpose of the government paying a heavy price for merging various digital resources is to bring convenience for scientific research and culture spreading, why are many obstacles put to prohibit the users' utilization? The users' difficulty in browsing a website would lead to a result that users spurn this website and acquire their desirable resources from other ways. Thus, less and less people would utilize CADAL resources, leading to the waste of resources.

5. Summary

The goal of CADAL project is to construct a huge number of multidisciplinary, multicategory and multilingual digital resources. Since the construction of digital resources of CADAL project began more than ten years ago, some initial success has been achieved. It is expected that during the third building stage of CADAL in the near future, continuously strengthening the propagation of CADAL project and attracting more Chinese and foreign institutions to participate in the construction should be carried out. Moreover, the search page of CADAL should be more personalized and convenient. The browsing of digital resources in the website of CADAL should be more expeditious. The CADAL project will integrate various service resources based on previously built service system by employing cloud computing to further construct service sharing cloud with oversea digital library, share user resources with different institutions and provide information service for various organizations. We believe that according to continuous construction of CADAL project, the co-construction and sharing of global resources as well as culture integration will be achieved.

References

- [1] 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.
- [2] X. Shen, J. Zong, S. Han, J. Zhou, On the Development of China Academic Digital Library&Information System, *Document, Information & Knowledge* (2) (2007) 30-36.
- [3] X. Shen, Z. Zheng, S. Han, C. Shen, A review of the major projects constituting the China academic digital library, *The Electronic Library* 26 (2008) 39-54.
- [4] L. Bie, L. Xiao, Q. Zhu, X. Yao, Construction and sharing of digital resources of academic libraries of China, *Digital Library Forum* (2) (2011) 52-58.
- [5] W. Xiong, Construction of digital library in universities: a case of CADAL project, *University Education* (3) (2015) 165-166.
- [6] X. Xiao, An overview of construction of CADAL data center, *Journal of Library Science* (7) (2013) 45-47.
- [7] Y. Xia, J. Wu, P. Gao, Y. Lin, T. Mao, Ontology-based model for Chinese calligraphy synthesis, *Computer Graphics Forum* 32 (2013) 11-20.
- [8] P. Gao, J. Wu, Y. Lin, Y. Xia, T. Mao, Fast Chinese calligraphic character recognition with large-scale data, *Multimedia Tools and Applications* 74 (2015) 7221-7238.
- [9] W. Wang, Y. Zhang, B. Wei, Y. Li, An ontology-based domain modeling framework for knowledge service in digital library, *Advances in Intelligent Systems and Computing* 278 (2014) 373-383.
- [10]C. Huang, Digital library based on DjVu format, *Journal of Information* (5) (2004) 50-54.
- [11]Y. Zhang, J. Wu, Y. Zhuang, Personalized multimedia retrieval in CADAL digital library, in: *Advances in Multimedia Information Processing - PCM 2008*, 9th Pacific RIM Conference on Multimedia in Taiwan, 2008, Springer-Verlag Berlin, pp. 703-712.
- [12]C. Yang, B. Wei, J. Wu, Y. Zhang, L. Zhang, CARES: A ranking-oriented CADAL recommender system, In: *JCDL 09: Proceedings of the 2009 ACM/IEEE Joint Conference on Digital Libraries* in Austin, TX, 2009, ASSOC Computing Machinery, New York, pp. 203-211.
- [13]H. Chen, H. Zhu, A Summarization on the China-America Digital Academic Library, *Journal of Academic Libraries* (1) (2005) 3-6.