

# Research on the Importance of Constructing Digital Resources of Doctoral Dissertations and Master's Theses Written in Mongolian

**Bao Meirong**

*Institute of Inner Mongolian Normal University, Hohhot, China  
jsj05547@163.com*

**Abstract:** *In this paper we focus on the importance of the construction of the characteristic resource library of Doctoral Dissertations and Master's Theses written in Mongolian of Inner Mongolia Normal University, and analyzes the opportunities and challenges brought by the construction of the characteristic resources of university libraries combined with the characteristics of the era of big data, and puts forward some countermeasures.*

**Keywords:** *Big data; Characteristic collection resources; Doctoral Dissertations and Master's Theses written in Mongolian*

## 1. Introduction

Big data technology provides unprecedented opportunities for the improvement of the construction of characteristic collection resources and the reconstruction of resource organization system in university libraries [1]. Since 2009, "Big data" has become the hottest keyword in the Internet technology industry. The emergence of the concept of big data has changed people's cognition of previous data. Data itself is of little significance. Only by analyzing and developing a large number of data then reflect the real value of data and provide more complete services for users' daily life, work and scientific research. University libraries can seize the characteristics of the era of big data and combine with their own resources to effectively enhance their own competitive advantages and improve the level of service. Especially for university libraries in minority areas, the construction of minority characteristic collection resources is the path of the future development.

## 2. About doctoral dissertations and master's theses

Mongolian is the main national language in Inner Mongolia Autonomous Region. Most colleges and universities in the region are characterized by Mongolian and Chinese bilingual teaching, which lays a foundation for the cultivation of talents with both Mongolian and Chinese knowledge. In the past 10 years, many countries in the world have made full use of the Internet and information technology to develop dissertation databases and search platforms, which are convenient for students to effectively obtain useful information, shield irrelevant information, improve learning efficiency, and provide convenient conditions for the retrieval of doctoral dissertations and master's theses especially the full text of dissertations for students. The digitization of dissertations has received wide attention from all walks of life and has become an important part and developing direction of the digitization of literature resources.

At present, domestic database of doctoral dissertations and master's theses in Chinese mainly include: database of China excellent doctoral dissertations and master's theses, Wanfang full-text database of degree papers, China Academic Library & Information System's full-text database of colleges and university, Chinese Academy of Sciences Thesis & Dissertation Database, degree papers database in national library of China, Colleges and universities establish their own dissertations database and so on. It is understood that several universities in Inner Mongolia have independent doctoral dissertations and master's theses written in Mongolian uploading system, but do not provide a unified retrieval and management platform, only as file management. This is make a large extent, constitutes a serious waste of resources and affects the way for Mongolian teachers and students to get first-hand Mongolian research materials.

In terms of dissertation management, colleges and universities archive and preserve paper documents combined with electronic documents collected through thesis submission system [2]. However, because the number of graduates increases year by year, this method has been far from meeting the needs of school information development and teachers and students.

Our school is a college with the characteristics of Mongolian and Chinese bilingual teaching and normal education. There are a lot of Mongolian graduation papers every year, but there is no unified management and retrieval platform, so we can only use Chinese keywords and topics to query information. Therefore, strictly speaking, there is no Mongolian full-text search platform. This affects researchers' efficiency and comprehensive to access the reference materials, so the development of the digital management system of Mongolian literature degree papers can well alleviate the above problems.

The research on Mongolian information retrieval in China started late with a few researchers, mainly from Inner Mongolia University and Inner Mongolia Normal University in the region. Mainly Research results in recent years include: "Research on Retrieval of Historical Mongolian Document Images", "Research on Mongolian Information Retrieval Technology for OCR'ed Documents", "Research on Mongolian Information Retrieval Model Based on Markov Random Field", "Research on Mongolian Document Data Digitization and Retrieval Technology", "Research and Implementation of Mongolian Full-text Retrieval System Based on Lucene", etc. These research results have some results on Mongolian information retrieval, but the retrieval of Mongolian literature degree papers in universities is still blank.

Inner Mongolia normal university library since opened the doctoral dissertations and master's theses Submission System from 2012 to 2022 ,Has been Submitted 1053 doctoral dissertations and master's theses, Wide range of disciplines and majors involved .such as law, education, literature, philosophy, education management, education, economics, history, rural education master, master of social work, and so on.

### **3. The opportunity of constructing characteristic resource library of Mongolian dissertation in the era of big data**

In June 2018, the National Language Resources Monitoring and Research Center for Minority Languages Mongolian Big Data Research Base was inaugurated in Inner Mongolia Normal University, aiming to strengthen international cooperation in the field of Mongolian data industry and build a "Dynamic circulation database of Chinese Minority Language Resources".

The main task of the base is to collect, process and organize the resources of major Mongolian language media (newspapers, periodicals, books, network media and audio media) into the dynamic circulation corpus of international minority languages and the corpus of bilingual education materials of people (Mongolian) and Chinese. At the same time, the research results will be submitted to the international language work department for publication, which will provide reference for the formulation and adjustment of national ethnic language policies, provide basis for the formulation of national ethnic policy, and provide services for the information retrieval of Mongolian language [3].

Doctoral dissertations and master's theses resource library written in Mongolian is the most important language resources in colleges and universities which Mongolian and Chinese bilingual teaching features. It is an important reference basis and achievement display for teachers and students to carry out scientific research. With the development of big data, users' demand for university library collection resources is increasing day by day, and higher requirements are put forward for the quality of library information service. Specifically reflected in the following aspects.

(1) From the perspective of co-construction and sharing of information resources: By analyzing the digital resources in university libraries we know the problem of miscellaneous and waste. And then demand for the co-construction and sharing of digital resources is becoming more and more significant. Therefore, starting from the construction of characteristic collection resources in colleges and universities, digital resources are no longer limited to meet the needs of teachers and students on campus, but also can make use of the advantages of colleges and universities to make resources serve the society.

(2) From the perspective of Improve the library service: To build a characteristic resource library of doctoral dissertations and master's theses resource library written in Mongolian can Change the status quo in our school that the characteristic resources of Mongolian is zero, and provides first-hand Mongolian resources for the construction of Mongolian literature and scientific research. It is convenient to search and find, and can help to obtain needed resources quickly.

(3) From the perspective of university library competition: By analyzing the characteristic resource library of university library, we can show the characteristics of university, for example subject advantage, regional culture and other. The construction of characteristic resources can improve the competitiveness and discipline services in the field of normal education to a certain extent [4-5].

#### **4. Challenges and strategies of digitized doctoral dissertations and master's theses resource library written in Mongolian in the era of big data**

##### ***4.1. The workload of encoding conversion is large***

The research on the information processing of Mongolian characters started late, one of the important reason is that the international standard encoding of Mongolian is late. The Mongolian national standard code was accepted by the ISO10646 in 1999 and included in Unicode in 2000. On the system level, standard Mongolian encoding was first supported in Vista in 2007, but the real introduction in Windows7 was in 2010 when Mongolian Baiti fonts were introduced for Unicode encoding. Before that, in the process of Mongolian information processing development, different research institutions and scholars have developed more than a dozen encoding schemes according to certain needs, but only three of them have been well promoted and have a certain user scale [6]. For example, the Menksoft encoding based on shape code developed by Menksoft Company, the founder code of Peking University Founder Company and the Sain encoding based on grapheme. But the encodings were directly incompatible with each other, could not exchange information with each other, and continued to coexist for more than a decade. This has seriously affected the development of Mongolian information processing and the construction and sharing of resources.

At present, relevant national departments have begun to guide and encourage the companies which engaged in Mongolian information processing, and the units and individuals Unified use the international standard code of Mongolian to achieve the unification of various codes, so that Mongolian information processing is more specification and standardization.

Almost all the doctoral dissertations and master's theses written in Mongolian in Inner Mongolia Normal University are written in Menksoft encoding, and input by different text editors and different input methods, so there are errors in the format and display submitted through the dissertation submission system, so they are finally submitted in PDF format. The difference of PDF version lead to some file conversion text files appear garbled and so on.

##### ***4.2. The description format of doctoral dissertations and master's theses written in Mongolian is not uniform***

In the process of digital processing of resources, if the format of resources is not uniform, it will bring a lot of trouble to batch processing, and lead to increased manual intervention, and finally affect the digital effect and efficiency. Due to the weak degree of digital informatization of degree papers written in Mongolian and the lack of uniform format requirements in schools, there are great differences in the format of papers of different majors and grades. For example, there are many corresponding terms for the term "reference" in degree papers written in Mongolian, such as "reference materials", "reference books", "reference papers" and so on. This situation obviously affects the precision and recall of information retrieval.

##### ***4.3. Users do not pay enough attention to Unicode encoding***

In recent years, the relevant national departments began to guide and encourage users to use international standard encoding as far as possible, users began to have a certain understanding of the importance of international standard encoding. But other codes and products have been on the market for a long time, and the average user is still used to input methods and editors for other codes. In Inner Mongolia Normal University, for example, almost 100 percent of the students who write their dissertations in Mongolian use Mongolian Pinyin input method and Mongolian WPS editor. Its main objective reason is that the school did not publish relevant written Mongolian literature papers related regulations, there is no limit encoding and editor, there is no format requirements, students can only in accordance with and reference have graduate students' thesis format, eventually just convert PDF document upload requirements, also can use Chinese in China National Knowledge Internet (called CNKI for short) retrieval to the Mongolian literature papers, However, it does not realize the importance

of data sharing, retrieval, secondary development and utilization. Many open source tools for Mongolian information processing (such as Lucene retrieval engine, a full-text retrieval kit developed by Apache) require that the data encoding processed must be Unicode encoding. In recent years, MenkSoft has also begun to develop character library and input method based on standard encoding, which fully shows that the technical standard of Mongolian text information processing is closer to the international standard encoding. Secondly, students do not care about the process of writing a dissertation with what input method, with what code and other problems, as long as the content of clear expression of the word correct, correct format. It does not consider the impact of different codes on the further utilization of the subsequent paper resources, such as retrieval, data sorting, dissemination and sharing. Although it is possible to retrieve Mongolian literature degree papers through the Chinese field of the CNKI doctoral thesis database, the CNKI cannot determine which of the current search results are in Chinese and which are written in Mongolian. Only by opening the original text can it be determined whether it is Mongolian literature degree papers. Moreover, only the full text in CAJ format can be provided, and the most basic copy and paste operations cannot be carried out normally, which brings a lot of inconvenience to users and affects the utilization value of the literature.

The key to solve the above problems is let teachers and students of the majority of Mongolian to pay high attention to the international standard code of Mongolian. To make them realize the importance of characteristic resources of Mongolian literature degree papers and the value of secondary utilization, activities can be carried out by setting up literature information retrieval courses for Mongolian students and conducting in-depth lectures in various colleges. The university may formulate regulations on the encoding and format of writing a Mongolian literature degree thesis and require students to do it compulsion. through the above efforts can effectively promote the implementation of international standard Mongolian encoding to a new level .

#### ***4.4. The data literacy awareness of librarians is weak***

Baidu Encyclopedia defines data literacy as a continuation and extension of the concepts of media literacy and information literacy, including at least the following five dimensions: sensitivity to data; Ability to collect data; Ability to analyze and process data; The ability to make decisions using data [7]; Critical thinking about data. The library is a site for a variety of data collections, such as the collection of books, the basic information of readers and users and borrowing amount, the library interview catalog Marc data, Mongolian literature degree papers and other characteristic collection resources. This data is a buried treasure that needs to be mined, exploited and utilized. At present, the service consciousness of most libraries is only limited to purchase books and various databases as much as possible. Librarians are not sensitive to library data. They find the data, but do not know the value of the data, so they have no consciousness to further develop and use it. Our university mainly teaches Chinese and Mongolian bilingual education. With the expansion of undergraduate and postgraduate enrollment, the Mongolian related paper collection resources in the library can no longer meet the needs of students and teachers, especially in the out-of-print books, newspapers, periodicals, dissertations and other resources. Therefore, scanning and preserving the above resources through modern digital technology to build characteristic collection resources can alleviate the problems caused by resource scarcity to a certain extent. In the era of big data, the most basic is the construction of resources. Without data resources, it is impossible to carry out deeper analysis and play the potential value of data resources. Therefore, it is very necessary to cultivate the data information literacy of librarians in the era of big data, and the construction and training of professional personnel is the key.

## **5. Conclusion**

With the rapid spread of big data, the transformation of service forms and the construction of characteristic resources have become the focus of the future development of university libraries. Especially for university libraries in minority areas, the construction of minority characteristic resources plays an important role in the protection and development of ethnic culture. Dissertation is the research results written by undergraduate students and master's and doctoral students of universities and scientific research institutions when applying for degrees at all levels. It has important reference value for the research and development of related majors. In particular, the dissertation of master degree has high academic research value and practical value, which is a very important original research record. But dissertation is a kind of grey literature, sometimes difficult to obtain and use. Especially for the papers written in ethnic languages by graduate students in universities in minority areas, it is more difficult to obtain electronic papers without a unified retrieval platform. Therefore, minority university libraries

should attach importance to the digitization of dissertations, promote the protection and inheritance of ethnic literature resources, and realize the maximum utilization of resources.

### **Acknowledgements**

This paper was supported by Bibliometric Statistical Analysis of Journal Papers of High Quality Humanities and Social Sciences Research in Western Normal Universities (No.2020ZXXM006); the Fundamental Research Funds for the Inner Mongolia Normal University (No.2022JBXC003); the Graduate students' research & Innovation fund of Inner Mongolia Normal University (No.CXJJB22006).

### **References**

- [1] Yuan Fang. *The Reform of Library Document Resource Construction Mode in the Environment of Big Data* [J]. *Library and Information Service*, 2015 (18):91-94.
- [2] Liang Chun. *The Practice and Considerations on Digitization of Academic Dissertations* [J], *Research on Library Science*, 2018(08):19-26.
- [3] *Mongolian Big Data Research base has been established in Inner Mongolia Normal University* [EB/OL]. [2018-06-15]. <https://office.imnu.edu.cn/info/1117/1007.htm>.
- [4] LIU Hua. *Discussion on the Digital Construction of University Library's Characteristic Collection Literature in the era of big data*[J]. *Journal of Library and Information Science in Agriculture*, 2016 (7):36-38.
- [5] Guo Ruifang. *Characteristic collection Construction of University library under sharing environment* [J]. *Journal of Library Science*, 2011(1):45-47.
- [6] Jin Liang, Lin Min. *On the Differences Among Three Essential Mongolian Coding Standard*[J], 2013 (2):225-227.
- [7] Wei Lijun. *Research on the Present Situation and Development Strategy of Data Literacy Education of University Library—A Case Study of Member Libraries of the Excellent 9*[J]. *Library Work and Study*, 2022(10).