Construction of Characteristic Thematic Database of Record Room in Colleges and Universities -- Discussion on the Construction of Chinese Language Thematic Database

XIUHUA HAN*, HONGWEN YU
Shandong Agricultural and Engineering University, Jinan, Shandong, 250100, China

ABSTRACT: The combination of Internet communication technology and multimedia technology has led to the development and application of digital information technology. The construction of thematic database in the record room of universities in China aims to meet the new teaching tasks and the needs of the development of information technology. According to the survey results, most universities in China have carried out the basic construction of thematic databases. However, how to combine database construction with computer big data processing has rarely been studied. In the existing databases, automated data mining and learning ability are poor, and automatic archiving can't be carried out. In this paper, taking the subject of Chinese language as the research object, the construction of related databases was discussed and studied. The basic framework of machine learning resource database based on intelligent data mining was proposed, and the evaluation criteria for computer server side of resource services in thematic databases was discussed. Based on the literature and information characteristics of Chinese language major, the system platform of thematic database was constructed.

INTRODUCTION

In recent years, the double development of computer information technology and Internet communication technology has led to the development of digital information resources. Libraries and record rooms in universities are the storage places of important documents, as well as the main bases of various scientific researches. In order to ensure that record room can provide scientific research and teaching services for universities, it is necessary to make effective use of modern information technology and various digital resources [1]. Through the storage capacity of the original information resources, more and more university record room has constructed more databases of special topics. The construction of the thematic database can not only reflect the intelligence learning ability of the digital information system in the university, but also the perfection of the construction of digital information transfer system in the era of Internet plus.

Which is the only way to realize the information sharing and development of the network [2]. In the early 90s of last century, China put forward the goal of building the 211 Engineering University, and the colleges and universities needed literature guarantee system based on scientific research documents. The construction of university characteristic thematic database in China became the basis for later scientific research and teaching. From the establishment of CALIS to the initial stage of operation, the navigation window and its characteristic thematic database were designed. Later, during the period of 12th Five-Year and 13th Five-Year, China's investment in the construction of special databases in university libraries or record room was very great. In 11th Five-Year, the first batch of key disciplines and thematic databases was initially built, and the database resources of academic degrees in universities were increasing day by day [3]. Driven by the project, other universities in China have begun to attach importance to the database construction of basic subject resources, and a wave of database building has begun nationwide [4]. The Ministry of education has issued the database construction of university libraries and record room in China, which requires to include the characteristic construction of
university database into the key management work during 12th Five-Year and 13th Five-Year.

In view of the construction of special databases in university libraries and record room, many scholars have done research and practice, and have put forward many excellent methods of resource construction. Ana at present, the research on the construction of thematic databases has also achieved certain achievement. However, from the results of the investigation and analysis, the computer data mining technology and learning ability of the database are not mature, which is still in the process of continuous improvement [5]. Although in some databases there are issues such as lack of specificity, lack of depth of expertise and lack of professional unified planning, after theoretical research and in-depth study of related technologies, some existing problems have been solved preliminarily. Most of the research results are based on theoretical research, but there is less analysis on technical solutions. Moreover, there is a lack of research on the resource integration, quality control and some system evaluation of the special database of the record room in the university. In the evaluation of database, some evaluation methods and related contents have also been put forward. However, there is no broad consensus on the basic methods of evaluation [6]. Computer cloud computing and basic used metadata and UML technology are applied in the construction of database, but the future development level of the database is still the focus of future research. The construction of characteristic thematic databases of record room in universities is the storage and invocation of document information in the digital environment, which is an important core content of information resources construction. In this paper, the construction of characteristic thematic databases of the record room in universities was deeply studied and discussed, and the ability of the system to drive intelligent learning and store and sort file was analyzed. Then a certain intelligent service evaluation system was established, and the actual problems were analyzed and summarized. The purpose of this paper is to provide data support for the establishment of characteristic artificial intelligence database of record room in universities. And the functional design of the big data database involved in the Internet should be improved in terms of quality and construction on the basis of existing research.

2. The theory and current situation of the construction of characteristic artificial intelligence database of record room in universities

2.1 Types and construction significance of intelligent database construction

At present, both the library and the record room have proposed the concept of the special information database of information organizations. Although there are some differences in address, thematic resources and characteristic resources that are available on the website for the public are all of the same attributes. Library and record room are important framework and core structure of social information, bearing the important responsibility for the development, utilization and transmission of literature resources. Special database is also an important task for the development of digital information construction of record room and libraries in universities [7]. Relying on the special database, the abundant network resources in the archives are the deep exploration and revelation of the informatization of a specialized field for the information needs of users, which can deal with standardized and digitalized information data more systematically, and meet the requirement of more and more personalized information database [8]. It can be seen that the artificial intelligence database of the record room in the university is built up on the basis of the existing digital and informational resources, which can fully display the basic construction of the discipline, and provide more perfect professional characteristics according to the standardization process. In the construction of database, it is necessary to enrich the professional construction of university network resources, in which, some of the storage materials are ancient books and western featured publications which are relatively rare and difficult to fully understand and utilize [9]. The characteristic thematic database of colleges and universities is an independent interdisciplinary subject database established around the subject scope, based on the advantages of the resource of the university, and combined with the characteristics of the school running. In databases, navigation and reference systems are typically included.

The development of the artificial intelligence database of the record room in the university is also the lifeline of the school's continuous development of discipline construction, which reflects the discipline characteristics and the advantages of running a school, and can help teachers and students understand the development trend of individual subjects in some fields. As can be seen in figure 1, the dynamic details of the internationalization developments save a lot of valuable time [10]. It is an important manifestation of the superiority discipline to be able to consult foreign academic documents and expert professors' research results in the archives and
to look up famous annuals and speeches [11]. The construction of characteristic thematic databases has become an important standard for the construction of information resources in colleges and universities. The record room in universities should carry out some construction of specialized information bases on the basis of the school's educational characteristics, especially the thematic databases with more urgent research [12]. The standard establishment of such databases should be carried out on the premise of achieving information sharing, so as to reflect standardized platform construction standards. In the early stage of data collection, data mining should be carried out on the basis of the original manual acquisition. Intelligent computing ability is needed here as a technical support to establish a standardized input function of the system, and to achieve the acquisition of basic data by artificial intelligence, thus to prominent discipline characteristic and special topic characteristic, and display scientific research specialized superiority [13]. The construction of a wide range of databases sometimes does not reflect depth, and the characteristics of relevant industries authoritative data is rarely reflected. These problems will be the problems that should be solved urgently in the construction of record room database in the future.

2.2 Overview of the development of intelligent thematic databases

After investigating the characteristic thematic database of libraries and record room in universities and colleges of China, the results show that the construction of record room in the 211 Engineering University is very extensive, and also covers the establishment of school's professional characteristic thematic database. School education will also adopt files and content stored in the database. Most of the documents and data in general collections are related to humanities and geography, which are difficult to obtain in conventional channels [14]. Through the basic literature reading survey, Northeast Normal University and Jiangnan University and other liberal arts schools have more professional database construction, and the number of building databases is shown in table 1, constructing more than 15 databases of specialized features database, accounting for about 6%. As a result, thematic databases require relatively high levels of funding and quality. Therefore, the basic construction of the database can provide better support for the construction of digital information in the record room of key universities in China [15]. In the construction database, the number of database construction for specialized subject is less than half of the total database construction, indicating that colleges and universities are still weak in the construction of characteristic thematic databases for record room [16]. The database retrieval methods of record room in the university need to provide accurate file names, classification and language. Vague retrieval methods can be used to retrieve a part of the retrieval vocabulary. However, Boolean logic calculation method can carry out more complex retrieval, and complete the calculation of attributes, so as to achieve fuzzy search, but accurate search for information.

| Table 1 Quantitative statistics of university libraries with characteristic thematic databases |
|-----------------------------------------------|-----|-----|-----|-----|-----|
| Characteristic library quantity              | 5   | 6~10| 11~15| >15 | Total|
| Library quantity                             | 48  | 39  | 12  | 6   | 105  |
| Percent                                       | 45.7%| 37.1%| 11.42%| 5.7%| 100% |

In view of the main distribution form of the intelligent special features database, the general idea of the database construction of the university's professional features is mentioned in the literature. Table 2 presents the main retrieval theme titles of the national university record room's thematic feature database. As can be seen from table 2, the database construction of the record room in our universities is based on the strengths of the school, and the network resource navigation and collection characteristics are relatively few links in database construction. There are a large number of special subject databases in universities, and scientific research, academic degrees and periodicals are the main goals of construction [17]. The characteristics and specialties of the school reflect the characteristics of the school. Therefore, the record room in universities needs to pay attention to the advantages of the discipline and
carries specialized and characteristic construction of the main database. For example, Jiao tong University is suitable for building the subject database of the transportation record room, so as to give full play to the strengths of the dominant discipline, and the Agricultural University is suitable for the construction of agricultural record room database [18]. At present, the domestic special database has entered a full range of development from a single literature search [19]. Aiming at the construction of special subject database, both of the construction of characteristic thematic database and the construction of informatization are relatively little. In order to satisfy the user's demand for the original text, it is necessary to reveal the text and depth of data more deeply, and provide more personalized services for users [20]. The construction of specialized subject databases in Chinese language needs to set up detailed editions, types, the publication age of literary documents, and the sources and uses of ancient books and other information, so as to give brief directions for use, and help users understand how to use Chinese language subject databases.

### Table 2 Subject statistical table of the subject database of record room in Universities

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number</th>
<th>Percent</th>
<th>Theme</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection</td>
<td>85</td>
<td>14%</td>
<td>Region</td>
<td>54</td>
<td>22%</td>
</tr>
<tr>
<td>University features</td>
<td>174</td>
<td>28%</td>
<td>Network teaching</td>
<td>21</td>
<td>26%</td>
</tr>
<tr>
<td>Discipline characteristic</td>
<td>163</td>
<td>26%</td>
<td>Other</td>
<td>23</td>
<td>21%</td>
</tr>
</tbody>
</table>

3. **Theoretical techniques of characteristic thematic organizations and the construction method of database platform**

3.1 **Resource organizing method of characteristic topic information**

In the process and method of artificial collection of digital information with characteristic subjects, the thematic information needed for books, paper and documents is manually transferred into the computer. The user retrieves the relevant content of the document through the computer, and when the information is useful, the user reads through the digital delivery process. The process is shown in figure 2.

![Fig.2 Artificial collection of thematic information](image)

The process of organizing the topic needs the establishment of the database, tagging the useful elements, indexing the index results and saving the result data. In the process of building data, a comparison of content features is required for the included field type, and the internal and external comparisons are also needed through the title or structure of the article, so as to establish the field length value control for the database, as shown in figure 3.

![Fig.3 Subject information element data manual indexing process](image)

After the organization of information, the computer changes the database features thematic information into metadata information, and the type of information file is XML. The data access interface is advantageous for querying data. Records of information retrieval topics are downloaded directly, and the data is transformed into corresponding XML files by mapping. After rewriting, the header file is converted to the XML file type. Then in encoding, the file is to be defined, and the form in the database stores the underlying files and contents. Figure 4 is a flow chart (XML file type) that generates metadata in bulk.

![Fig.4 Flow chart](image)
The principle of retrieving thematic databases is based on the architecture. The architecture of this research is B/S three-layer architecture, which are browser layer, service layer and database. Cross searches of multiple databases can be accomplished through the flow of thematic information among different layers. As shown in Figure 5, the browser layer is the process of information interaction and understanding between people and machines. Through the calculation system, the customer carries on the logical coding as well as the user demand acceptance. And the request made to the service layer of the second layer is transferred to the customer through the HTTP protocol. The service layer implements the demand of browser transmission through data processing method. Database, the third layers, is to complete the storage and control of data, and deal with the results of the upper layer.

By using ASP dynamic service technology, the script of server is designed from the script environment of server, and the dynamic service program is formed. The browser applies the file to the server, and then the server calls the script for the private file, and returns it to the browser and leaves a certain response. Script running is not the operation situation after browsing, which needs to generate a standard page (HTML type) and passes it to the browser. The advantages of using ASP files are that the files are simple to create, and there are many types of browsers supported by objects that can be fused with many files, as shown in figure 6.

When ASP calls the associated components, the server accesses the database by using ODBC. The generated files are in the middle layer structure, and the files passed to the browser are XML types, thus forming the display of the dynamic web pages of the XML intermediate document. The method of combining XNL with ASP is adopted, and the interface database schematic diagram is shown in figure 7.
3.2 A scientific evaluation method for characteristic thematic databases

Hierarchical analysis can be used as the basis for evaluating the resource service of the database. The decision problem is broken up into different components according to their characteristics, as shown in figure 8.

![Hierarchical structure diagram of decision problem](image)

The standard expression for the consistency of the calculations is:

\[
CI = \frac{\lambda_{\text{max}} - n}{n - 1}
\]

In view of the established database model, the relevant experts are invited to evaluate the index system, and the main purpose of the study is to give feedback to experts on indicators. The correlation matrix of judgment is established, and the weight of evaluation index is determined by theoretical analysis and calculation, and the weight is calculated as:

The independent row elements of the weights are computed:

\[
M_i = \prod_{j=1}^{n} a_{ij} \quad (i = 1, 2, ..., n) \tag{2}
\]

In addition, after the product is solved, the root mean square of the Mi is solved, and the calculation method is:

\[
\overline{W_i} = \sqrt{M_i} \tag{3}
\]

After normalization, the equation is:

\[
W_i' = \frac{\overline{W_i}}{\sum_{j=1}^{n} \overline{W_j}} \tag{4}
\]

The maximum eigenvalue of the matrix is obtained by calculation, and the solution equation after verification is:

\[
\lambda_{\text{max}} = \frac{\sum_{i=1}^{n} (AW)_i}{nW_i} \tag{5}
\]

According to the above solution, combined with the hierarchical analysis of professional transfer verification, the results are determined. By understanding the index evaluation method of characteristic thematic database in universities, the weights of evaluation indexes are calculated, so as to determine whether the calculation process is reasonable.

4. The realization and discussion of the thematic database platform based on Chinese language major

4.1 Database operation environment and platform implementation results

The encapsulation of the thematic database resources of Chinese language majors is based on the standard data structure of resources. Content packages require a simple encapsulation method. Moreover, in the content package, the list of contents and the form of the exchange file are summarized. XML language can't reflect its own value in the absence of the use of its own, especially the attribute characteristics of its resources for database construction. The resource features of Chinese language combine with XML features, and form the corresponding side information of construction attributes. A separate packet can represent the learning content reused by a group. Certain independent documents must be decomposed and also independent. The file packs are individually compressed to form DVD or CD files and are distributed to the public by using the Internet or multimedia, and the file type published in the media network is XDR or XSD.

The list of resource contents for Chinese language majors is to meet scalability within the prescribed scope, and some of the lists can be learned and used independently of own needs. The basic content of the composition is to be closely connected and can't be disturbed by external interference, and the resources and organizational forms of the file can only be described by using the same list. After the content is assembled, a higher level of content list is formed.

![Form a higher level content list](image)

The evaluation of Chinese language resources is carried out in terms of teaching, technology, images, and text resources based on the above improvements. And the management analysis and theoretical calculation transmitted through resources ensure that the prepared resources are needed by the user. The resource structure of Chinese language major must satisfy the basic frame structure. According to the content classification, the data resources are
standardized and transformed into standard packages which can be reused, so as to expound the basic concepts of culture and content of Chinese language major, and to better establish the encapsulation model. XML technology is used to seal the contents of Chinese language subject resources, so as to complete the framework of seal preparation.

![Fig.10](image1.jpg)

**Fig.10 XML technology seals up the Thematic Resources of Chinese language**

The organization and construction of Chinese language resources need to ensure the authenticity of content and culture, and to ensure the accurate and effective integration of resources. A standardized work system and standardized resource collection methods is established. And the content formulation of Chinese language resources needs to meet the requirements of packaging content. Moreover, the supply of resources, and the construction methods of digital libraries and record room require constantly learning. After meeting the above requirements, the construction methods of the special feature database are combined with the requirements of the construction of the Chinese special subject database, and finally the publishing interface of the database is shown in figure 11.

![Fig.11](image2.jpg)

**Fig.11 Database publishing interface**

4.2 Discussion

The construction of artificial intelligence collection database of the characteristic record room in the university effectively combines with the information acquisition technology of machine learning, and forms the characteristic thematic database of university archives of artificial intelligence based on Internet big data. The establishment of intelligent special database requires the classification of archival data of professional disciplines, while pure manpower can't complete more than a certain order of classification of data files. Therefore, the retrieval technology and the publishing technology of database need the process of machine learning, so as to optimize the publishing model of intelligent database. The collection of information about artificial intelligence is the basic issues that should be addressed first in technical topics. For the documentation of the paper category, the key topics should be manually collected in the file room, and then burned into the disk and stored in the computer. By means of effective computer calculation and automatic acquisition of information with learning machines, the strong technology of XML in the actual data storage application were extracted, so as to effectively play XML's big data learning and organizational skills. Through the database retrieval function, the rationalization analysis was carried out, and the issue of database was mainly studied and analyzed. And the dynamic publication of data for artificial intelligence files was elaborated and explained in detail. The database construction platform was closely related to the requirements of the overall system. Therefore, taking the Chinese language major as an example, the platform of the special database was designed, and the metadata basic structure of XML was given. And in the construction of Chinese artificial intelligence database, the design idea was used, and the design goal was achieved with good running state. In the use of a Chinese language thematic database, data should be updated in a timely manner, so as to consistent with the structure and content of the original data, and to complete data management and routine maintenance.

5. Conclusion
The construction of characteristic thematic databases has always been an important part of information construction in colleges and universities, especially the construction of characteristic thematic databases in record room. In the research, the typical problems of the domestic characteristic thematic database construction were summarized, and the data classification and storage publication of database information which were not completed in the domestic special databases were found out. Therefore, in view of the above problems, the storage and classification technology of intelligent information digitization was studied in this paper. The construction method of artificial intelligence database was given, and the acquisition technology framework of intelligent thematic information was constructed. And the search method of text file data mining with high similarity was introduced. Then combined with related theoretical analysis, the publishing technology of intelligent thematic data was obtained. In the construction phase of intelligent database, the framework of basic database platform was designed, and a special information organization model of metadata based on data learning XML technology was constructed. Finally, by using the above technical framework and organization model, a special database of artificial intelligence data acquisition and storage based on Chinese language major was constructed, and the platform construction of the thematic database was also completed. The research results of this topic will be of guiding significance for the construction of intelligent database of record room in specialty discipline. Through the construction of intelligent special subject database, the efficiency of consulting data in record room will be greatly improved, and the resource acquisition process will be simplified, which can not only save time and effort, but also can promote the progress and development of the Chinese language field.

7. REFERENCES