Topic Identification and Evolution Analysis of Papers on China' s Literature Resource Security System: Postprint

Authors: Tian Xuran, Ma Xiaoxiao, Li Yuhai

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose / Significance] This study analyzes the evolutionary path of research themes on China's literature resource guarantee system to provide references for its reconstruction. [Method / Process] The LDA model is primarily employed to cluster the collected literature. First, based on time period divisions, a keyword co-occurrence network is constructed to explore the interactive relationships among various themes. Subsequently, similarity calculations are used to determine the internal evolutionary paths of each theme, and Sankey diagrams are drawn to visually present the evolutionary results. [Results / Conclusion] The study reveals that relevant themes on China' s literature resource guarantee system had essentially emerged by around 2000. The themes mainly comprise two major categories: resource perspective and institutional perspective, which are significantly influenced by computer technology and national policies. Corresponding countermeasures and recommendations are proposed for these two major categories of themes.

Full Text

Identification and Evolution Analysis of Literature Themes in China' s Literature Resource Guarantee System

Tian Xuran[1], Ma Xiaoxiao[1], Li Yuhai[1,2]

[1]School of Information Management, Central China Normal University, Wuhan 430079
[2]China Library Innovation and Development Research Center, Central China Normal University, Wuhan 430079

Abstract: [Purpose/significance] By analyzing the evolutionary path of themes in China' s literature resource guarantee system, this paper provides

references for reconstructing the national literature resource guarantee system. [Method/process] This study primarily employs the LDA model to cluster collected literature. First, time periods serve as the basis for division to construct a co-occurrence network of topic terms, exploring interactions among various themes. Then, similarity calculations determine the evolutionary paths within each theme, with Sankey diagrams visualizing the results. [Result/conclusion] The study finds that relevant themes of China' s literature resource guarantee system had basically all emerged by around 2000. The themes mainly fall into two categories: resource perspective and institutional perspective, heavily influenced by computer technology and national policy. Corresponding countermeasures and suggestions are provided for these two major theme categories.

Keywords: literature resource guarantee system; LDA; thematic identification; thematic evolution

1 Introduction

As the world enters the 5G era (the fifth-generation mobile communication standard, also known as fifth-generation mobile communication technology), network data transmission speeds will be significantly faster than in the previous 4G era, enabling more rapid and efficient application of 5G technology across multiple domains. Both traditional literature resources and new digital literature resources continue to accumulate, and through the overall construction of literature information resources, a literature information resource system that can effectively guarantee social literature needs within a certain scope—the literature resource guarantee system—has been established [1].

In this era that should be characterized by interconnectivity, foreign database providers, relying on their abundant literature resources, have raised prices arbitrarily, attempting to continue monopolizing resources and triggering dissatisfaction among domestic librarians. A report released by the European University Association (EUA) shows that transaction costs between academic institutions, libraries, and publishers such as the American Chemical Society (ACS), Elsevier, Wiley, Springer Nature, and Taylor & Francis are increasing at an annual rate of 3.6%. Literature resources, especially scientific and technological literature resources, represent the latest frontier research findings. Abandoning the purchase of relevant resources would result in losing international research competitiveness, while continuing to comply with these coercive terms would maintain a passive situation. Therefore, following a series of challenges represented by Mr. Cheng Huanwen' s "Ten Questions to Data Providers!!!" , reconstructing China' s literature resource guarantee system has become an urgent priority.

Currently, some foreign universities have responded to the threat of monopolistic pricing by data providers through open access publishing. In February 2019, after terminating its agreement with Elsevier, the University of California signed the first open access publishing agreement in U.S. history with Cambridge Uni-

versity Press in April of the same year. However, China has yet to develop a complete literature resource guarantee system solution that can effectively respond to price increases. The road to reconstructing the literature resource guarantee system is long, and grasping it requires a macro-level understanding of past research. By reviewing previous studies and understanding what research themes and work contents exist under the framework of the literature resource guarantee system, we can provide guidance for its reconstruction, change the passive development situation in recent years where progress has been driven by computer technology, and thus build a self-sufficient literature resource guarantee system capable of competing with domestic and foreign database providers, using history as a guide and emerging technology as a driving force.

2.1 Introduction to the LDA Model

Previous research has mostly adopted the Term Frequency-Inverse Document Frequency (TF-IDF) method, representing the document collection as a matrix with documents as rows and words as columns, where values are proportional to a word' s frequency in a specific document and inversely proportional to its frequency across multiple documents. TF-IDF tends to produce sparse matrices, representing documents from a word frequency perspective rather than semantic representation [3], and it can also underestimate words that frequently appear in a class but represent that class' s theme [4], requiring continuous adjustment of TF-IDF parameters to meet practical needs [5]. Therefore, this paper uses the LDA (Latent Dirichlet Allocation) topic model as the foundation for thematic identification of literature related to the literature resource guarantee system over the years. LDA is a three-layer Bayesian probability model based on "document-topic-word." The specific joint probability formula is:

P w z formula (1)

where represents the topic distribution, α is the prior distribution parameter of the topic distribution (i.e., Dirichlet distribution), β is the prior distribution parameter of the keyword distribution, z represents the theme generated by the model, w represents the keyword finally generated by the model, N represents the number of words in the document, and M represents the number of documents. The three-layer probability model is shown in Figure 1 [Figure 1: see original paper].

2.2 Research Design

To explore the research themes of past literature on the literature resource guarantee system, it is necessary to study the evolutionary trends of relevant published literature themes. Since a paper' s keywords may represent research questions, methods, or objects, analyzing themes based solely on keywords is not conducive to thematic identification [2]. Current research mostly uses thematic models to mine themes and explore thematic evolution. The simplest method is Term Frequency-Inverse Document Frequency (TF-IDF). Few scholars have pre-

viously conducted thematic evolution analysis on literature related to China' s literature resource guarantee system. This paper primarily uses the LDA model for thematic identification of relevant literature, realizing the application of the LDA model in this domain. Furthermore, it constructs keyword co-occurrence networks and thematic evolution Sankey diagrams to conduct evolutionary analysis from both macro-level quantitative and micro-level temporal perspectives. The thematic identification process is divided into four modules, as shown in Figure 2 [Figure 2: see original paper].

(1) Database selection and data collection. The research object of this paper is the thematic evolution and development of domestic literature on the literature resource guarantee system, so the CNKI journal full-text database is selected as the data source. The search terms are "文献资源保障体系" (literature resource guarantee system), "文献信息资源保障体系" (literature information resource guarantee system), "文献保障体系" (literature guarantee system), and "文献资源保障" (literature resource guarantee), with "OR" as the logical connector. The search year includes all relevant literature before 2021. After manually removing some irrelevant literature, a total of 1,429 relevant documents were collected, with titles, keywords, and abstracts compiled as the data source.

(2) Data preprocessing. Based on the collected data, the titles, keywords, and abstracts of all literature were merged and treated as long texts representing each document. Then, the Python jieba segmentation tool package was used for Chinese word segmentation. To improve segmentation effectiveness, a user-defined dictionary was needed. Based on multiple segmentation experiments, proper nouns such as "文献资源" (literature resources), "双一流" (Double First-Class), and "大数据" (big data) were added to the custom dictionary to improve segmentation validity. A stop word list was also added during segmentation, and this paper used the commonly used Chinese stop word list—the Harbin Institute of Technology stop word list. Finally, the segmentation results were saved as data for LDA model construction.

(3) LDA modeling. Before using the LDA model for thematic identification, the optimal number of topics must be calculated. This paper uses the K-means algorithm in Python' s scikit-learn toolkit, calculating the Distortions (within-cluster sum of squared errors) and Silhouette coefficients of the segmented texts to determine the optimal cluster number k, where smaller Distortions coefficients are better and larger Silhouette coefficients are better. The results are shown in Figure 3 [Figure 3: see original paper] and Figure 4 [Figure 4: see original paper].

As shown in Figures 3 and 4, considering both the Distortions and Silhouette coefficients comprehensively, selecting 12 clusters is more appropriate. In subsequent LDA modeling, the number of thematic clusters was set to 12, with α and β maintaining the default values in the Python library. Since titles, abstracts, and keywords were all treated as a single long text, their respective weights were

considered.

To explore the evolutionary paths of different topics in LDA model clustering results, in addition to using the natural attribute of publication time, it is necessary to calculate text similarity and set a certain threshold to determine text topics with high correlation, thereby identifying them as evolutionary relationships to form thematic evolution paths. This paper uses the cosine similarity method to measure the similarity of texts within the same cluster across different years, thereby determining evolutionary paths between themes. The cosine similarity calculation formula is:

cos( ) similarity || || * formula (2)

The cosine similarity value ranges between (0, 1), with larger values indicating greater relevance between two texts. Since abstract content is relatively long, to avoid excessive data sparsity, the similarity threshold was set at 0.1, meaning that when the similarity between two texts exceeds 0.1, they can be considered to have an evolutionary relationship. Combined with the publication year of the texts, thematic evolution paths were plotted.

(4) Thematic result analysis. Based on the quantity distribution of literature related to the literature resource guarantee system and the thematic clustering results of the LDA model, further in-depth analysis was conducted. On the one hand, changes in the quantity of relevant literature on China' s literature resource guarantee system were explored from a macro-level quantitative perspective; on the other hand, analysis was conducted from a micro-level thematic perspective.

3.1 Analysis of LDA Model Thematic Identification Results

Based on the number of relevant literature published over the years, a yearly line chart was drawn (Figure 5 [Figure 5: see original paper]). In 1983, China promulgated the "National Standard of the People's Republic of China - General Rules for Bibliographic Description" (GB3792.1-83), which defined "literature" as "all carriers that record knowledge."This concept provided a relatively unified definition of the connotation and extension of "literature" in academia. Gradually, the terms "books" and "collections" were also encompassed under the definition of "literature." According to the search results, relevant literature on the literature resource guarantee system has been published successively since 1984, accompanied by discussions on the transformation of library functions: shifting from collection to utilization of library collections to better guarantee literature resource needs for all sectors of society. Around 2000, the number of relevant literature began to rise rapidly, while after 2010, it began to show a downward trend.

Table 1 shows the LDA thematic model identification results. From the terms under the same theme, words with higher probability and thematic significance were selected, and corresponding theme labels were summarized to represent

each theme. As shown in Table 1, themes such as information resource evaluation, literature resource construction, literature resource sharing, and digital libraries were identified. Based on the yearly changes in relevant literature publication numbers, the literature data was divided into three periods: Period I (1984-1999), Period II (2000-2010), and Period III (2011-2020). CiteSpace visualization software was used to draw keyword co-occurrence networks for relevant literature (Figures 6 [Figure 6: see original paper] - 8 [Figure 8: see original paper]), showing connections between relevant thematic terms.

Table 1 LDA Model Thematic Identification Terms

Topic 1: Information Resource Evaluation	Topic 2: Characteristic Literature Resources	Topic 3: Library and Information Institutions
Institutional Setup	Business Reorganization	Grassroots Libraries
Library Consortia	Library Building Construction	

Topic 4: Literature Resource Sharing	Topic 5: Literature Resource Guarantee System	Topic 6: University Libraries
Co-construction and Sharing	Interlibrary Loan	Literature Resource Layout
Resource Layout	Resource Co-construction	Literature Resource Guarantee
CALIS	BALIS	University Libraries
Key Disciplines	Digital Libraries	Digital Literature

Topic 7: Digital Resources	Topic 8: Literature Resource Construction	Topic 9: Document Delivery
Library Automation	Metadata	Electronic Libraries
Smart Libraries	Scientific and Technical Literature Resources	Standard Literature
Agricultural Literature	Sports Literature	Medical Literature

Topic 10: Literature Collection	Topic 11: Knowledge Services	Topic 12: Literature Organization
Original Text Delivery	Journal Subscription	Knowledge Services
Reference Consultation	Knowledge Economy	Legal Protection

Topic 10: Literature Collection	Topic 11: Knowledge Services	Topic 12: Literature Organization
Literature Publishing		

Period I (1984-1999) themes mainly focused on Topic 3 (Library and Information Institutions), Topic 4 (Literature Resource Sharing), Topic 6 (University Libraries), Topic 8 (Literature Resource Construction), Topic 10 (Literature Collection), and Topic 12 (Literature Organization). The research content primarily concerned the functions of traditional libraries and other library and information institutions, as well as prospects and planning for the new century.

Literature resources in the 20th century were mainly paper-based, but digital literature resources gradually developed, corresponding to the development and changes in databases. In 1986, the National Marine Administration Intelligence Institute first introduced foreign CD-ROM databases, after which university libraries and intelligence institutions also introduced CD-ROMs for subject retrieval, including Selective Dissemination of Information (SDI), retrospective searching, specialized services, achievement novelty searches, and patent examination novelty searches [7]. In 1992, the "Chinese Scientific and Technical Journal Article Database" (CB ISTIC/CEPC Periodicals ChinaBase) CD-ROM version, launched by the Chongqing Branch Database Research Center of the China Institute of Scientific and Technical Information, became the first Chinese data CD-ROM in mainland China. In January 1997, the "China Academic Journals (CD-ROM)" was officially issued regularly, representing China' s first large-scale integrated academic journal full-text database, which the library community regarded as a milestone marking China' s entry into the digital library era.

However, the development speed of the Internet far exceeded that of CD-ROM databases. After the World Bank' s 1998 World Development Report proposed the concept of National Knowledge Infrastructure (NKI), Wang Mingliang proposed building the China National Knowledge Infrastructure (CNKI) in March 1999. Chongqing VIP Information Co., Ltd. established the VIP Information Network in 2000. Wanfang Data Corporation launched the first domestic information product, the "Chinese Enterprises, Companies, and Products Database," in the early 1990s. Thus, CNKI, VIP, and Wanfang gradually became the three mainstream domestic data service platforms.

Period II (2000-2010) and Period III (2011-2020) themes mainly focused on Topic 8 (Literature Resource Construction), Topic 4 (Literature Resource Sharing), Topic 6 (University Libraries), and Topic 5 (Literature Resource Guarantee System). The main challenges of the 21st century involve constructing literature information resources oriented toward the new era and various social fields, gradually establishing a literature resource guarantee system.

The literature resource guarantee system is a social system integrating literature

collection, storage, revelation, delivery, utilization, and many other functions [8]. Regarding the operational model of the entire literature resource guarantee system, Xiao Ximing believes that a hierarchical structure control method is the correct choice for constructing China' s literature resource guarantee system model. While adopting the hierarchical structure control of large systems as the basic framework, it should absorb advantages of other control methods to build a three-level network structure model consisting of regional (provincial, municipal, autonomous region) level, regional (administrative region) level, and national-level literature resource networks [8]. Sun Ruiying proposed adding the establishment of an international-level guarantee system on this basis [9].

Topic 5, the Literature Resource Guarantee System, represents important practice in literature resource guarantee system research. As the earliest literature information resource guarantee system launched in China, the China Academic Library & Information System (CALIS) was formally established in 1998. As part of the public service system of the "211 Project," CALIS provided support and guarantee for the development of key disciplines in universities, covering various disciplinary fields such as agricultural literature, legal literature, business literature, medical literature, and sports literature, and including different types of literature such as standard literature, scientific and technical literature, and foreign language literature. In addition to CALIS, multiple literature resource guarantee systems were subsequently developed, such as the Beijing Academic Library & Information System (BALIS) [10], Jiangsu Academic Library & Information System (JALIS) [11], China Academic Social Sciences and Humanities Library (CASHL) [12], National Science and Technology Library (NSTL) [13], and Chinese Science Digital Library (CSDL) [14].

Entering Period III (2011-2020), after the State Council issued the "Overall Plan for Promoting the Construction of World-Class Universities and First-Class Disciplines"(referred to as the "Double First-Class"construction) in 2015 [15], the "211 Project"construction gradually shifted to "Double First-Class"construction, and the functions of CALIS and other literature resource guarantee systems changed accordingly, mainly providing guarantees for the literature information needs of universities' "first-class disciplines."

3.2 Thematic Evolution Analysis

In addition to the coarse-grained division of literature themes into three periods for keyword co-occurrence analysis, this paper also draws Sankey diagrams based on cosine similarity combined with publication time to further explore the evolutionary paths of literature resource guarantee system themes. The Sankey diagram, also known as the Sankey energy distribution diagram, originated from the "Steam Engine Energy Efficiency Diagram" in 1898. In Sankey diagrams, objects are represented by element blocks, and the direction and connections of energy flow between objects are represented by connecting lines. In this paper, element blocks represent research themes, connecting lines represent evolutionary relationships between themes, and the year when a theme first appeared is

marked in parentheses after the theme element block. To make the thematic evolution paths clearer, theme words that appear repeatedly in subsequent evolution paths are no longer displayed.

As shown in Figure 9 [Figure 9: see original paper], the main research content of the literature resource guarantee system formed relatively early, with relevant literature themes basically covered by around 2000. Subsequent discussions focus on updates and improvements in content and methods for different construction fields in the new century and new environment. The following sections mainly analyze the evolution within each theme.

(1) Topic 4, Literature Resource Sharing, is the main research theme of the literature resource guarantee system. In 1973, the International Federation of Library Associations and Institutions (IFLA) first proposed the concept of UAP (Universal Availability of Publications), or "international library resource sharing." After this concept was introduced to China, it received high praise from domestic scholars. Yuan Zheng pointed out that achieving literature resource sharing would help university libraries break away from the "self-sufficient" natural economic situation, alleviating problems such as insufficient book funds, lack of management personnel, limited cataloging capabilities, and saturated collection space [16]. Interlibrary loan is an important approach and means to achieve literature resource sharing. The earliest proposal of interlibrary loan in China was around 1939, when the Ministry of Education of the Republic of China issued the "Revised Library Regulations" and "Library Work Outline," which mentioned "interlibrary loan and mailing" [17], representing the beginning of literature resource sharing to some extent. However, limited by the domestic context at that time, literature resource sharing failed to develop well. By the end of the 20th century, due to the development of information technology, many scholars began advocating for library automation [18-19] and digital libraries [20-21], and interlibrary loan of physical books gradually evolved into online resource sharing with broader significance, generating a series of research themes closely related to Internet technology development, such as cataloging standardization, network construction, and information resource networking. During this period, some scholars also proposed building a literature resource sharing service network center or adopting a paid sharing model [22], but subsequent related research results were limited.

In August 2006, Google CEO Eric Schmidt first proposed the concept of "Cloud Computing" at the Search Engine Conference (SESSanJose2006) [23]. Cloud computing is essentially a completely new concept of network application, where users can obtain resources on the "cloud" anytime according to their needs [24]. This concept brought new ideas for literature resource sharing models.

As a major agricultural country, the sharing of agricultural literature information resources is also of strategic significance for national development. The ultimate goal of national agricultural literature resource co-construction and shar-

ing is to provide resource support and services for national agricultural teaching, research, production, and management, thereby promoting breakthrough development in Chinese agriculture [25]. To improve the economic and social development level of western regions and consolidate national defense, the State Council established the Western Region Development Leading Group in January 2000. Previously, scholars had proposed corresponding measures for agricultural literature resource co-construction and sharing in western regions [26]. The establishment of relevant literature guarantee systems would also help libraries provide suggestions for western development and guarantee literature resources for western regions [27]. With technological development, the China Academy of Agricultural Sciences Information Service Network has now been launched, providing a series of information services to build a smart agricultural science collaboration platform. In September and October 2013, Chinese President Xi Jinping proposed the "New Silk Road Economic Belt" and "21st Century Maritime Silk Road" cooperation initiatives, opening doors for China' s western regions and coastal areas. Properly organizing and guaranteeing relevant historical literature not only helps apply for World Cultural Heritage status and promote the construction of characteristic literature resources but also strengthens research on regional cultures of neighboring countries and even alleviates disputes over border or coastal islands to some extent [28-29].

(2) Topics 9 (Document Delivery), 10 (Literature Collection), 11 (Information Services), and 12 (Literature Organization) are all more fine-grained themes that evolved from Topic 8 (Literature Resource Construction). Document delivery and collection are basic functions of libraries. University library collections are important guarantees for the construction of key university disciplines. Due to different economic development situations and quantities of local characteristic literature in different provinces and cities, literature resource layout varies. Before purchasing and collecting relevant literature, sufficient research on literature resource layout must be conducted, followed by procurement and collection of needed literature resources that are lacking.

In addition to focusing on the evolution and development of database technology, literature collection and delivery also involve multiple aspects such as multiple backups and timely migration, open description methods, simulation environments and environmental encapsulation, data recovery and data archaeology, technical frameworks and overall solutions, and standardization technologies [30].

To better collect and deliver literature resources, effective description is necessary. Metadata is an important part of information description in literature resource organization. Metadata not only plays an important role in digital resource cataloging but is also key technology enabling libraries to move toward automation. MARC (Machine-Readable Catalogue) and Dublin Core are two relatively mature metadata standards widely recognized in the library and information science community. In 1965, MARC (later called MARC I) devel-

oped by the U.S. Library of Congress represented preliminary achievements in machine-readable cataloging. MARC II was introduced in 1968 through UK-U.S. cooperation. In 1991, Chinese authorities compiled the "China Machine-Readable Cataloging Communication Format"(CNMARC) based on UNIMARC with specific fields added, which has been revised multiple times. In 1995, OCLC and NCSA jointly held the first Dublin Core conference, finally establishing a core set containing 15 core elements. Due to certain limitations of MARC in granularity, language, and extensibility, the Library of Congress proposed the Bibliographic Framework Model (Bibframe) in May 2011, striving to greatly integrate existing bibliographic resources, though how it adapts to the Chinese cataloging environment remains to be studied in depth.

The rapid development of the Internet environment in the new era will continue to drive innovation and reform in literature information resource organization methods and concepts. Literature information resource organization will rapidly develop toward interdisciplinary integration, intelligent semantic organization, and integrated information methods [31].

The ultimate purpose of literature resource construction remains serving users and readers. In May 1995, Comrade Jiang Zemin proposed implementing the strategy of "revitalizing the country through science and education"in his speech at the National Science and Technology Conference. This prompted libraries to move from information services to knowledge services, boosting the implementation of the science and education revitalization strategy through knowledge services. Knowledge service refers to the process of extracting and transmitting knowledge from various explicit and implicit information resources according to people' s needs [32]. Knowledge service is an advanced stage of information service based on literature information resource construction. To fully develop knowledge services, it is necessary to deeply mine users' knowledge needs, use intelligent means to add value to explicit knowledge, and enable the transmission and reception of implicit knowledge, thereby providing personalized information services [33]. These all require improvements in knowledge mining, knowledge organization, knowledge development, and knowledge service personnel quality [34]. Meanwhile, online information service providers and large publishers have gradually begun semantic web application experiments, giving rise to semantic publishing, a new publishing form that transforms literature resources from isolated, static knowledge packages into knowledge tools embedded in an interconnected and interactive knowledge system [35]. Semantic publishing helps users discover or verify new knowledge on one hand, and enables publishing institutions to obtain new profit returns and profit space on the other. In the long run, knowledge services and semantic publishing will remain a fiercely competitive area in the information service field.

(3) Topic 1: Information Resource Evaluation. At the macro level, constructing a literature resource guarantee system requires high-level macro-control institutions, corresponding policies, regulations, and standards, as well as support from all sectors of society [36]. At the micro level, to ensure

better construction of the literature resource guarantee system, relevant subjects and objects need to be evaluated and assessed. Suo Chuanjun et al. divided evaluation subjects into individual and institutional categories, where individual evaluators refer to experts and scholars from different fields, and evaluation institutions include business institutions, service institutions, and academic institutions [37]. Different evaluation systems are then formed according to different evaluation objects, such as journal evaluation systems, collection literature resource evaluation systems, and digital literature resource evaluation systems. An Yueying constructed a two-level collection resource evaluation system, where first-level indicators include resource content, retrieval system, economy, and storage system, and second-level indicators include collection resource guarantee capability, authority, timeliness, standardization, retrieval function, retrieval effectiveness, usability, cost, usage, storage system efficiency, and security [38]. Ma Haiqun et al. constructed a set of network information resource evaluation indicators containing 16 indicators from four major aspects: information source content, information source organization, information source performance, and other indicators [39]. Journal evaluation system indicators include but are not limited to total downloads, impact factor, 5-year impact factor, cited impact factor, average citation count, Web immediacy download rate, immediacy index, comprehensive total citations, citable literature volume, citing journals, cited journals, and various other indicators [40].

Meanwhile, part of university discipline evaluation also involves assessing relevant discipline literature resources. Typically, Chinese university departments can independently purchase Chinese and foreign literature according to research and teaching needs. This independent procurement method is highly targeted and specialized [41]. However, some university library literature resource purchases, while meeting the reading needs of teachers and students, also consider the development of the institution' s key disciplines to build characteristic collections and key discipline collections. With the implementation of projects such as "Double First-Class," university mergers and multi-campus operations have emerged during the educational process, causing university libraries to face changes in political and ideological work, institutional setup and personnel management, standardization of rules and regulations, network and software updates, fund management, resource sharing, and collection layout [42-43]. This series of university and discipline reforms will ultimately require supporting optimization of relevant literature resource services.

Overall, the various themes of the literature resource guarantee system are intertwined and develop together. Their thematic evolution roughly aligns with the four development stages of library automation and trends in information technology development. The first stage is the development of integrated library automation management systems, the second stage is the phase of global, integrated electronic literature information services online [44], the third stage is the digital library stage, and the fourth stage is the smart library stage. During

this period, information technology has continuously updated and iterated, such as the rise and fall of CD-ROMs, from local area networks to the Internet, the gradual replacement of MARC by the Bibframe model, and the vigorous development of new technologies such as cloud computing, big data, and the Internet of Things. The national level has also issued different policy documents, from the "211 Project" to "Double First-Class" disciplines, from "Western Development" to the "Belt and Road Initiative," and from the "Golden" projects in the Ninth Five-Year Plan to the "Cyber Space Community with a Shared Future" in the Fourteenth Five-Year Plan, all continuously promoting the development and evolution of China's literature resource guarantee system. Overall, the evolution of China' s literature resource guarantee system is a process in which, through the joint efforts of library and information science, computer science, and many other academic communities, and relying on advanced information technology, continuous contributions have been made to China' s literature guarantee cause.

4 Conclusion and Discussion

This paper mainly conducts thematic identification based on the LDA topic model, realizing the application of the LDA topic model in the field of literature resource guarantee systems. In the thematic evolution paths of the literature resource guarantee system, rich and diverse themes have been formed. Around 2000, the main research themes of the literature resource guarantee system were basically included, which can be mainly divided into two major categories: the literature resource level and the institutional level. From the resource perspective, research content includes literature resource types and collection, literature resource organization and construction, and literature resource service and sharing. From the institutional perspective, university libraries have always been the main research object. With the development and construction of the literature resource guarantee system, national-level institutions and local grassroots institutions have developed rapidly. These two major categories are also influenced by macro policies such as "Western Development," "Double First-Class" discipline construction, and the "Belt and Road Initiative," as well as the cross-impact of emerging technologies such as "big data," "cloud computing," and "data mining."

To reconstruct the literature resource guarantee system, corresponding adjustments also need to be made from both resource and institutional perspectives. From the resource perspective, in the literature resource type and collection stage, some university literature resources lean toward key and strong disciplines while neglecting the procurement of materials for weak disciplines, ethnic minority areas, and non-English foreign language literature. However, from a system optimization perspective, "weak disciplines" and diversified development should also be considered to drive the improvement of disciplines in all universities. In the literature resource organization and construction stage, digital construction is the general direction of literature resource construction, with metadata construction as its core, achieving multi-channel metadata integra-

tion, multi-type metadata mapping, and multi-level metadata identification to establish an intelligent literature organization platform with knowledge association functions. In the literature resource service and sharing stage, while establishing a new digital resource open ecology model using open data, open access, and open publishing as means, user needs and values should be fully emphasized. User portraits should be constructed by capturing and analyzing users' subjective characteristics, behavioral data, preference data, and feedback to provide fine-grained personalized resource services for users.

From the institutional perspective, although various specialized institutions have been developed over the years, there exist problems such as redundancy, serious fragmentation, duplicated responsibilities, and lack of unified management to varying degrees. To establish a management mechanism for the literature resource guarantee system with unified coordination and departmental linkage, it is first necessary to establish a cross-system, cross-disciplinary, and cross-departmental unified national permanent institution at the central level to oversee the command and coordination of relevant literature resources. Second, a national expert committee and specific administrative agencies should be established to be responsible for technical guidance and implementation. Finally, relying on various societies and local libraries, regional alliances and grassroots organizations among various systems should be established to achieve classified, graded, and division-of-labor guarantees [45], thereby building a nationwide literature resource guarantee system spanning various fields.

This study has certain limitations: In LDA model clustering, some clusters contain relatively few documents, which cannot fully reflect the literature evolution patterns. When conducting evolutionary path analysis, theme words that appear again in subsequent periods are no longer used as objects for analysis and display, which to some extent ignores more microscopic evolutionary changes in the thematic evolution process. Future research will continue to explore more detailed aspects of thematic evolution in the literature resource guarantee system and discuss the mechanisms of evolution and identification and prediction of future new themes.

References

[1] Wang Cuiping, Yang Peichao. Outline of National Literature Information Resource Guarantee System Construction[J]. Library Science Research, 2000(2): 15-17,14.

[2] Liu Ziqiang, Wang Xiaoyue, Bai Rujiang. Research on Multi-dimensional Visual Analysis Method of Discipline Theme Evolution—Taking Big Data Research in China' s Library and Information Science Field as an Example[J]. Journal of Library Science in China, 2016, 42(6): 67-84.

[3] Fan Yunman, Ma Jianxia. Review of LDA-based Emerging Topic Detection Technology[J]. New Technology of Library and Information Service, 2012(12): 58-65.

[4] Zhang Yufang, Peng Shiming, Lü Jia. Improvement and Application of TFIDF Method Based on Text Classification[J]. Computer Engineering, 2006(19): 76-78.

[5] Shi Congying, Xu Chaojun, Yang Xiaojiang. Review of TFIDF Algorithm Research[J]. Computer Applications, 2009, 29(S1): 167-170,180.

[6] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation[J]. Journal of machine learning research, 2003, 3: 993-1022.

[7] Xia Xu, Zeng Haibiao. Current Situation and Countermeasures of CD-ROM Database Introduction, Development, and Utilization[J]. Library Journal, 1996(3): 32-35.

[8] Xiao Ximing. Outline of China' s Literature Resource Guarantee System[J]. Library, 1996(6): 8-12.

[9] Sun Ruiying. Concept of Establishing National Literature Information Guarantee System[J]. Information Science, 2002(7): 680-683.

[10] Li Xiao. Analysis of Development Status of Beijing Academic Literature Resource Guarantee System[J]. Agricultural Library and Information Science Journal, 2014, 26(3): 16-20.

[11] Yang Yonghou. Review and Reflection on Construction of Jiangsu Provincial Academic Literature Guarantee System[J]. Journal of Academic Libraries, 2002(1): 72-74,92.

[12] Li Duo. Current Situation and Development of CASHL Network Service System[J]. Journal of Academic Libraries, 2005(3): 16-18.

[13] Yuan Haibo, Meng Liansheng. Practice of Information Resource Co-construction and Sharing Under Network Environment—With Description of Construction and Development of National Science and Technology Library and Documentation Center[J]. Journal of the China Society for Scientific and Technical Information, 2002(1): 57-62.

[14] Zhang Xiaolin. Chinese Science Digital Library and Its Construction Progress[J]. Bulletin of Chinese Academy of Sciences, 2005(4): 344-346,343.

[15] Zhu Lili. Discussion on University Library Literature Resource Construction Strategy Under "Double First-Class" Construction Background[J]. Library and Information Science Journal, 2018, 3(8): 7-11.

[16] Yuan Zheng. Practical Significance of Achieving Resource Sharing for University Libraries[J]. University Library Communication, 1987(4): 10-13,19.

[17] Cui Muyue, Dai Genxing. On Interlibrary Loan[J]. Henan Library Science Journal, 1990(3): 54-57.

[18] Xia Xu. A Major Trend in Resource Sharing Development—Integration of CD-ROM Technology and Communication Technology[J]. Journal of Academic Libraries, 1995(5): 40-42.

[19] Huang Jiannian. MARC Data and Libraries[J]. Journal of Tianjin Library Science, 1997(4): 45-48.

[20] Cao Zuohua. On Networking, Digitization and University Library Collection Construction Strategy[J]. Library Development, 1999(3): 23-25.

[21] Wang Yuanru, Ning Shenghong. Digital Library and Literature Information Resource Co-construction and Sharing[J]. Modern Information, 2000(6): 15-16.

[22] Guo Ye. Discussion on Creating a Paid Literature Resource Sharing System[J]. Journal of Ningde Teachers College (Philosophy and Social Sciences Edition), 2004(2): 82-83,86.

[23] Xu Ziming, Tian Yangfeng. Development History and Application of Cloud Computing[J]. Information Recording Materials, 2018, 19(8): 66-67.

[24] Luo Xiaohui. Discussion on Development of Cloud Computing[J]. Electronics World, 2019(8): 193-194.

[25] Wan Zhangqi. Discussion on Co-construction and Sharing of National Agricultural Literature Resources[J]. Agricultural Library and Information Science Journal, 1997(1): 21-22.

[26] Wang Ziyu. Concept on Construction of Agricultural Library Literature Resource Sharing in Five Northwestern Provinces[J]. Gansu Science and Technology, 1998(6): 2-3.

[27] Huang Quancai. Strategies for Library Participation in Western Development[J]. Library World, 2001(4): 1-6.

[28] Chen Binqiang. Construction of Maritime Silk Road Literature Resource Guarantee System[J]. Library Construction, 2015(5): 88-92.

[29] Zhou Chun, Feng Caifen, Ma Cichang. Demand and Guarantee of China Peripheral Regional Research Literature—Taking Sun Yat-sen University as an Example[J]. Journal of Academic Libraries, 2016, 34(5): 73-77,83.

[30] Wang Wei. Research on Long-term Preservation Technology of Digital Resources[J]. Information Science, 2012, 30(11): 1751-1754.

[31] Wei Min. Information Organization 4.0: Transformation Process and Future Vision[J]. Journal of the National Library of China, 2018, 27(1): 78-85.

[32] Tian Hongmei. Discussion on Library Moving from Information Service to Knowledge Service[J]. Information Science, 2003(4): 312-314.

[33] Yi Ming, Wang Xuedong, Deng Weihua. Research on Social Tagging Network Analysis and Personalized Information Service Based on Social Network Analysis[J]. Journal of Library Science in China, 2010, 36(2): 107-114.

[34] Zhao Ping, Ma Jiangbao. On Library Knowledge Service and Its Realization[J]. Library Science Research, 2005(8): 32-35.

[35] Wei Rui, Chu Jingli. Academic Book Open Access and University Library Publishing Services in the United States[J]. Journal of Academic Libraries, 2014, 32(3): 17-22.

[36] Wang Tao, Xiao Ximing. Construction of Literature Resource Guarantee System Under New Information Environment[J]. Library and Information, 1999(1): 33-36.

[37] Suo Chuanjun, Wu Qilin. Research Progress on Network Information Resource Evaluation at Home and Abroad[J]. New Technology of Library and Information Service, 2006(8): 55-59,93.

[38] An Yueying. Digital Collection Evaluation Indicator System Based on Analytic Hierarchy Process[J]. Library, 2008(4): 82-84.

[39] Ma Haiqun, Lü Hong. Research on Construction of Network Information Resource Evaluation Indicator System and Its Dynamic Fuzzy Evaluation Model[J]. Information Science, 2011, 29(2): 166-171.

[40] Chen Xiaoshan, Chen Guofu, Zhang Rui. Research on Structural Relationship of Journal Evaluation Indicators Based on Factor Analysis and SEM Model[J]. Information Science, 2016, 34(10): 61-64,71.

[41] Tang Dinghai. Preliminary Exploration of Department Self-procurement Literature Management[J]. Library Construction, 2009(1): 52-54.

[42] Li Jiaqing. Problems and Countermeasures Faced by Merged University Libraries[J]. Journal of Academic Libraries, 2001(S1): 10-12,30.

[43] Yang Shuyu. Discussion on How to Establish a Cost-effective Library with Multiple Campuses[J]. Information and Documentation Services, 2001(2): 54-56.

[44] Yang Zongying, Zheng Qiaoying, Xia Peinong. New Stage of Library Automation Development[J]. Journal of Academic Libraries, 1997(3): 1-5.

[45] Zhu Ze, Li Yuhai, Wang Changjue, et al. Reconstruction Road: Development and Future of China' s Digital Resource Guarantee System—Review of "2021 National Digital Resource Guarantee System Reconstruction Academic Seminar" [J]. Digital Library Forum, 2021(6): 30-35.

Author Contribution Statement

Tian Xuran: Conducted data analysis, paper writing and revision.
Ma Xiaoxiao: Conducted data analysis, paper writing and revision.
Li Yuhai: Responsible for paper topic selection, proposed the paper framework, and conducted paper revision and finalization.

Note: Figure translations are in progress. See original paper for figures.