Digital Humanities Research in the Field of Information Resource Management in China: Topic Identification, Interpretive Analysis, and Trend Outlook (Postprint)

Authors: Lei Pei, Jin Yuan, Jiangfeng Liu

Date: 2026-04-01T17:39:50+00:00

Abstract

[Purpose/Significance] In recent years, breakthrough transformations in artificial intelligence technology have exerted a profound impact on digital humanities research within the field of information resource management. A deep analysis of the thematic characteristics and development trends of digital humanities research in China' s information resource management field aims to provide a reference for the expansion and disciplinary positioning of the field.

[Method/Process] Using journal articles from 2015 to 2025 indexed in China National Knowledge Infrastructure (CNKI) as the data source, the BERTopic topic model and the Gemini large language model were employed to identify eight topics from the digital humanities research literature in China' s information resource management field. These were further summarized into four aspects for interpretive analysis and trend forecasting through manual reading and content identification.

[Result/Conclusion] Digital humanities research in China's information resource management field can be categorized into four areas: digitization of literature resources and knowledge mining technology in digital humanities, digital humanities research and services in the fields of libraries and archives, digital humanities research on ancient books, and digital humanities education and disciplinary development. Future research will focus on themes such as the upgrading of digital humanities research methods and tools, the enhancement of digital humanities service levels in libraries and archives, the empowerment of ancient book revitalization through generative artificial intelligence and multimodal technologies, and the transformation of the digital humanities education system driven by interdisciplinary exchange and generative artificial intelligence.

Full Text

Preamble

Digital Humanities Research in the Field of Information Resource Management in China: Topic Identification, Interpretation, and Trend Outlook

Pei Lei, Jin Yuan, Liu Jiangfeng (School of Information Management, Nanjing University, Nanjing)

Abstract

Digital Humanities (DH) represents a significant paradigm shift in the humanities and social sciences, driven by the integration of information technology and traditional scholarship. Within the field of Information Resource Management (IRM) in China, digital humanities has emerged as a vibrant and rapidly evolving research area. This study aims to systematically identify and interpret the core research themes within Chinese IRM-based digital humanities research and provide an outlook on future trends. By employing bibliometric analysis and content analysis on a corpus of high-quality academic literature, we identify key research clusters, including digital resource construction, data-driven knowledge discovery, cultural heritage preservation, and the evolution of research paradigms. The analysis reveals that while Chinese scholars have made substantial progress in technical applications and resource digitization, there is a growing need for deeper theoretical integration and interdisciplinary collaboration. Finally, this paper proposes future research directions, emphasizing the importance of human-centric AI, ethical data governance, and the expansion of digital humanities toward broader social impact.

1. Introduction

In recent years, the rapid development of digital technology has profoundly reshaped the landscape of academic research. Digital Humanities, as an interdisciplinary field situated at the intersection of computing and the humanities, has gained significant traction globally. In China, the field of Information Resource Management (formerly Library, Information, and Archival Science) has played a foundational role in promoting digital humanities research. Leveraging its expertise in resource organization, data processing, and information services, the IRM community has become a primary driver of DH initiatives, ranging from the digitization of ancient texts to the development of sophisticated spatial-temporal analysis platforms.

Despite the proliferation of research outputs, there is a need for a comprehensive synthesis of the current state of the field. Understanding the thematic structure and evolutionary trajectory of DH research within the Chinese IRM context is crucial for identifying research gaps and guiding future scholarly efforts. This

paper seeks to address this need by providing a systematic identification and interpretation of research topics, followed by a discussion of emerging trends.

2. Research Design and Methodology

To ensure a rigorous and objective analysis, this study adopts a multi-methodological approach combining quantitative bibliometrics with qualitative content analysis.

2.1 Data Collection The primary data source for this study is the China National Knowledge Infrastructure (CNKI).

Abstract

Introduction

[Purpose/Significance] In recent years, breakthrough transformations in artificial intelligence (AI) technology have exerted a profound impact on digital humanities research within the field of Information Resource Management (IRM). These advancements have not only revolutionized the methods of data processing and analysis but have also expanded the boundaries of inquiry across various humanistic disciplines.

[Method/Process] This study systematically examines the evolution and current state of AI-driven digital humanities research. By synthesizing existing literature and case studies, we analyze how deep learning, natural language processing, and computer vision are being integrated into the lifecycle of information resource management—ranging from the digitization of archival materials to the semantic enrichment of cultural heritage data.

[Result/Conclusion] The integration of AI into digital humanities facilitates a shift from simple digitization to intelligent knowledge discovery. However, this technological shift also brings forth challenges regarding algorithmic transparency, data ethics, and the preservation of humanistic interpretation. We conclude that future research in IRM must balance technical innovation with critical humanistic perspectives to ensure the sustainable development of digital scholarship.

Abstract

[Purpose/Significance] In recent years, breakthrough transformations in artificial intelligence have exerted a profound influence on digital humanities research within the field of Information Resource Management (IRM). This study conducts an in-depth analysis of the thematic characteristics and developmental trends of digital humanities research in China' s IRM field, aiming to provide a reference for the expansion and disciplinary positioning of the field. [Method/Process] Using journal articles from the China National Knowledge

Infrastructure (CNKI) as the data source, this study employs the BERTopic topic model and the Gemini large language model to identify eight themes from the literature. Through manual review and content identification, these themes are categorized into four dimensions for interpretive analysis and trend forecasting. [Results/Conclusion] Digital humanities research in China' s IRM field can be summarized into four key areas: the digitization of literature resources and knowledge mining techniques; digital humanities research and services in the fields of librarianship and archives; digital humanities research on ancient texts; and the development of digital humanities education and the discipline itself. Future research is expected to focus on upgrading digital humanities methodologies and tools, enhancing the service levels of libraries and archives, empowering the revitalization of ancient texts through generative AI and multimodal technologies, and transforming the digital humanities education system driven by interdisciplinary exchange and generative AI.

Future research directions will specifically emphasize the integration of multimodal technologies to empower the revitalization of ancient texts, the promotion of interdisciplinary collaboration, and the systemic reform of digital humanities education driven by generative artificial intelligence.

Keywords

Abstract

In the context of the digital humanities, the management of information resources has evolved from simple digitization to the deep mining of semantic knowledge. This paper explores the application of the BERTopic model—a next-generation topic modeling technique—within the field of information resource management. By leveraging transformer-based embeddings and class-based TF-IDF (c-TF-IDF), BERTopic allows for the extraction of dense, interpretable clusters from large-scale unstructured textual data. This study demonstrates how BERTopic overcomes the limitations of traditional Latent Dirichlet Allocation (LDA) by preserving semantic context and providing dynamic topic modeling capabilities, thereby offering a more robust framework for analyzing historical archives, literary corpora, and academic databases.

1. Introduction

The rapid development of digital humanities has led to an explosion of heterogeneous information resources. Effective information resource management (IRM) now requires sophisticated tools to organize, categorize, and extract meaning from vast quantities of text. Traditional methods of topic modeling, most notably Latent Dirichlet Allocation (LDA), have long served as the standard for discovering latent themes. However, LDA often struggles with short texts and fails to account for the semantic relationships between words, as it relies on a "bag-of-words" approach.

Recent advancements in natural language processing (NLP), particularly the emergence of pre-trained language models like BERT (Bidirectional Encoder Representations from Transformers), have revolutionized how machines understand human language. BERTopic, a model built upon these architectures, offers a promising alternative for researchers in the digital humanities. This paper examines the technical workflow of BERTopic and its specific utility in managing and interpreting digital cultural heritage and academic information resources.

2. Methodology: The BERTopic Framework

The BERTopic model follows a modular pipeline designed to transform raw text into coherent, hierarchical topics. Unlike traditional probabilistic models, BERTopic utilizes a clustering-based approach supported by high-dimensional vector representations.

2.1 Document Embeddings

The first stage involves converting documents into numerical vectors. By using pre-trained models such as Sentence-BERT (SBERT), BERTopic captures the contextual meaning of sentences. This ensures that words with similar meanings are positioned closely in the vector space, even if they do not share the same character strings.

2.2 Dimensionality Reduction and Clustering

Because transformer embeddings are high-dimensional, BERTopic employs UMAP (Uniform Manifold Approximation and Projection) to reduce dimensions while preserving the local and global structure of the data. Subsequently, HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) is used for clustering.

Keywords

Abstract

In the context of the digital age, the intersection of Digital Humanities and Information Resource Management has become a significant area of scholarly inquiry. This study utilizes the BERTopic modeling approach to analyze the evolving landscape of these fields. By leveraging advanced machine learning techniques, we aim to identify key thematic trends and structural shifts in the literature. Our findings provide insights into how computational methods are reshaping traditional humanistic research and information management practices.

1. Introduction

The rapid development of digital technologies has fundamentally transformed the way information is managed and how humanities research is conducted. Information Resource Management (IRM) and Digital Humanities (DH) are increasingly overlapping, creating new paradigms for data processing, archiving, and knowledge discovery. This paper explores the thematic evolution of these disciplines, categorized under the Chinese Library Classification codes TP391 (Information Processing) and related fields.

2. Methodology

To analyze the vast corpus of academic literature, we employ BERTopic, a state-of-the-art topic modeling technique that utilizes bidirectional encoder representations from transformers (BERT) and class-based TF-IDF (c-TF-IDF). Unlike traditional Latent Dirichlet Allocation (LDA) models, BERTopic creates dense clusters of documents, allowing for more nuanced and contextually aware topic discovery.

2.1 Data Collection and Preprocessing

The dataset consists of academic papers indexed under Digital Humanities and Information Resource Management. We focused on metadata including titles, abstracts, and keywords. Preprocessing involved the removal of noise and the standardization of technical terminology to ensure the accuracy of the embedding process.

2.2 Model Implementation

The BERTopic workflow follows several key stages: 1. **Embedding**: Converting text into numerical vectors using a pre-trained transformer model. 2. **Dimensionality Reduction**: Utilizing UMAP (Uniform Manifold Approximation and Projection) to manage high-dimensional data. 3. **Clustering**: Applying HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) to identify distinct thematic groups. 4. **Topic Representation**: Using c-TF-IDF to extract the most descriptive words for each cluster.

3. Results and Analysis

Our analysis revealed several dominant themes within the intersection of DH and IRM. These include the digitization of cultural heritage, the application of linked data in libraries, and the use of deep learning for text analysis.

3.1 Thematic Evolution

The results indicate a shift from basic···

Introduction

Digital Humanities (DH) traces its origins to the "Humanities Computing" initiatives proposed by the Italian scholar Roberto Busa in the 1940s. At the beginning of the 21st century, alongside the digital revolution and the rapid advancement of information technology, the discipline of Digital Humanities was formally established. In recent years, DH research has increasingly become a focal point across various academic fields, spanning literature, history, linguistics, computer science, and information resource management. Digital methods and information technologies have broadened the horizons of humanities research and facilitated the integration of information technology with humanistic disciplines. Currently, as an intersectional field of computational science, digital technology, and the humanities, the fundamental definition of Digital Humanities has yet to reach a consensus within domestic or international academic circles.

At the 2009 Modern Language Association (MLA) annual convention, the concept of Digital Humanities gained significant attention within the North American academic community and was subsequently introduced to China by Wang Xiaoguang. He posits that Digital Humanities, as an interdisciplinary research field, achieves the deep application of modern computing and network technologies within traditional humanistic education and research. The Association of Research Libraries (ARL) suggests that Digital Humanities is a research direction that utilizes computer technology to explore the characteristics of the humanities, representing an emerging field of study. Digital Humanities research in China started relatively late and has been driven by modern information technology, characterized by distinct cross-domain collaboration [**?**, **?**]. However, research has largely focused on drawing from foreign achievements, and a complete research system has yet to be fully formed. Bibliometric results indicate that DH research within the field of Information Resource Management (IRM) in China has developed rapidly; the volume of publications shifted from linear, slow growth between 2012 and 2016 to exponential growth after 2017. The primary research entities are concentrated in institutions such as Nanjing University, Renmin University of China, Wuhan University, Shanghai Library, and Nanjing Agricultural University, forming a collaborative network of core authors [**?**, **?**]. Some scholars, through analysis of literature classified under "Library and Information Science and Digital Libraries" and "Archives and Museums" in the China National Knowledge Infrastructure (CNKI), have found that research in this field primarily focuses on the application of DH technology, DH disciplinary construction, the relationship between DH and library development, and DH education [**?**]. This indicates that the IRM discipline has become a vital frontier for DH research in China. In recent years, breakthroughs in Artificial Intelligence (AI) have injected new vitality into DH research within the IRM field. Regarding the digital preservation and development of ancient texts, IRM scholars have actively applied advanced AI technologies to conduct theoretical and practical explorations. For instance, the SikuBERT model and

its series of applications developed by Wang Dongbo' s team at Nanjing Agricultural University (scholars in the field of Library and Information Science) based on the *Siku Quanshu* [**?**] provide technical platform support for the knowledge organization and application of ancient book resources. With the emergence of large language models such as ChatGPT, IRM scholars have begun to explore the innovative application of Artificial Intelligence Generated Content (AIGC) technologies in library, archive, and museum settings, as the revolutionary impact of GPT technology is reshaping the paradigms of DH research. In 2022, China' s primary discipline "Library, Information and Archive Management" was officially renamed "Information Resource Management (IRM)," and "Digital Humanities" was subsequently listed as a secondary discipline under IRM. In this context, Digital Humanities is not merely an expansion of methodology but a core academic growth point for the IRM discipline during its digital and intelligent transformation. The time frame selected for the empirical research in this paper is 2015–2025, which represents the "golden decade" of China' s IRM discipline as it moves from "basic digital construction" toward "deep digital and intelligent empowerment." Systematically reviewing the research themes and developmental trajectory of China' s IRM discipline over these ten years is of great significance for clarifying disciplinary boundaries and forecasting future development trends.

1.1 Data Acquisition and Preprocessing

To comprehensively review the development of digital humanities research within the field of information resource management in China, this study utilizes journal articles indexed in the China National Knowledge Infrastructure (CNKI) over the past decade as the source for empirical research data. The initial search was conducted using "digital humanities," "humanities computing," and "computational humanities" as subject terms. These terms were combined using the "OR" logic. To ensure the quality and relevance of the data, the source categories were restricted to "Peking University Core" and "CSSCI" (Chinese Social Sciences Citation Index) journals. Furthermore, the disciplinary classification was limited to "Library, Information, and Digital Libraries" and "Archives and Museums," covering a time frame from January 2015 to November 2025.

Following the initial retrieval, the results underwent a manual screening process. This involved the exclusion of conference proceedings, newspaper articles, literature reviews, duplicate records, and irrelevant documents. This refinement process resulted in a final dataset of approximately 1,100 valid literature records. Finally, using the Jieba library, a custom dictionary was established to perform word segmentation on the literature abstracts, utilizing the stop-word list provided by the Harbin Institute of Technology.

The data preprocessing of the literature was completed by removing stop words and Arabic numerals, followed by part-of-speech tagging based on the PKU standard. Non-central words, such as adverbs, numerals, and locative particles,

were filtered out to ensure the quality of the text data.

1.2 Topic Identification Model and Analysis Framework

Topic models are generally characterized by their unsupervised nature, efficiency in processing massive text datasets, and ability to automatically identify latent themes. Since Deerwester et al. proposed Latent Semantic Analysis (LSA) based on Singular Value Decomposition (SVD) in 1990, topic modeling has undergone an evolutionary process from Probabilistic Latent Semantic Analysis (PLSA) to Latent Dirichlet Allocation (LDA).

As the field has evolved, the limitations of traditional topic models have become increasingly apparent. In contrast, the BERTopic model effectively overcomes the constraints of traditional models regarding short text processing and semantic coherence. By utilizing Class-based Term Frequency-Inverse Document Frequency (c-TF-IDF) and Transformer embeddings, BERTopic significantly enhances the semantic aggregation of topic extraction. Given these advantages, this paper selects the BERTopic model in combination with the Gemini Large Language Model (LLM) to identify and label topics within the digital humanities literature in the field of Information Resource Management (IRM) in China. This approach aims to provide an in-depth interpretation of hot topics and a forward-looking analysis of future research directions, serving as a reference for digital humanities research and practice within the Chinese IRM domain.

This study employs the BERTopic model, integrating text summarization techniques and the c-TF-IDF weighting strategy, to create dense clusters through Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). By implementing BERTopic, we achieve precise matching between keywords and core article content, circumventing the limitations of probabilistic generative topic modeling. This ensures that topic interpretations are clear while preserving key terminology. Furthermore, the Gemini LLM is utilized to assist in topic naming. The specific operational methodology is as follows: first, the Sentence-BERT (SBERT) model is applied to perform word embedding on the preprocessed text, using neighboring words in the context to determine the corresponding vectors for input vocabulary. Second, the UMAP algorithm is used to reduce the dimensionality of the embeddings, and the density-based clustering technique HDBSCAN is employed to optimize the clustering process, with the "min_{{{cluster}}_{{size}}}" parameter being specifically defined. Third, the c-TF-IDF model is used to extract topic terms. Finally, the Gemini-1.5-pro-preview model is used to name the topics, facilitating a deep analysis of the thematic distribution characteristics of digital humanities research in China' s IRM field in recent years. The process is illustrated in

.

Figure 1

Figure 1: Figure 1

2.1 Research Topic Identification

The BERTopic topic model was employed to conduct topic modeling on the digital humanities research literature within the field of Information Resource Management (IRM) in China. Given the relatively large sample size, the parameters were fine-tuned as follows: `n_{neighbors}` was set to 15, `n_{components}` to 5, and `min_{{{cluster}}_{{size}}}` to 20. The `ngram_{range}` was set to (1, 4), while all other parameters were maintained at their default values. Ultimately, eight relevant research topics were extracted. These topics were then named using Google' s Gemini large language model (specifically the Gemini-3-pro-preview version).

To ensure the scientific rigor of the research process, the specific operational steps were optimized. First, parameter tuning was conducted by setting the `min_{{{cluster}}_{{size}}}` to 20. Given the sample size of 1,100 papers, this setting was designed to filter out long-tail fragmented noise and precisely extract the core issues most representative of the discipline. Second, an academic validation process was applied to the naming generated by the large language model. A closed-loop strategy of "machine-automated naming + manual review" was adopted to overcome potential semantic drift inherent in large language models during topic labeling. Through an in-depth reading of the core documents associated with each topic, experts performed academic standardization and corrections on the Gemini-generated names, ensuring that the identification results were deeply aligned with the context of the IRM discipline. The results indicate that digital humanities research in China' s IRM field focuses on eight primary themes. To further analyze the content of these themes, we visualized the top nine characteristic words for each of the eight topics, as shown in [FIGURE:2].

As illustrated by the visualization of topic characteristic word distributions in [FIGURE:2], core terms such as "digital humanities," "research," and "archives" appear repeatedly across multiple topics, reflecting the inherent characteristics of topic modeling. The contexts formed by these terms when combined with specific keywords under different topics are unique; these differences in semantic combinations effectively reveal the deep structures and internal logic within the texts. Based on the initial clustering results of the BERTopic model, combined with the topic naming from Gemini-3-pro-preview and manual verification, this study identified eight specific research themes in the field of digital humanities within China' s IRM domain (corresponding to Topic 0 through Topic 7 in [FIGURE:2]). Topic 0 represents "Archival Research from the Perspective of Digital Humanities" ; Topic 1 represents "Digital Humanities and Ancient Text Research" ; Topic 2 represents "Research and Development of Digital Humanities" ; Topic 3 represents "Research on Digital Humanities Services in Chinese

Academic Libraries" ; Topic 4 represents "Implications of Digital Services in U.S. Academic Libraries" ; Topic 5 represents "Research on International Digital Humanities Projects" ; Topic 6 represents "Research on the Development of Red Archival Resources" ; and Topic 7 represents "Knowledge Graph Construction and Digital Resources."

2.2 Topic Interpretation and Analysis

Considering that the results generated by large language models may be influenced by biases in the training data, we conducted an in-depth reading of the literature across eight specific research topics. By combining keyword scores with "document-topic" distributions, we performed manual optimization of the initial clustering results. The manual optimization process primarily involved splitting or merging specific topics: topics with an excessively broad scope were subdivided, while those with closely related content were consolidated. Based on this refinement, and considering the homogeneity of research objects and the contextual relevance of technical applications, we categorized the digital humanities research topics within the field of Information Resource Management in China into four primary dimensions: digitization of documentary resources and knowledge mining technologies in digital humanities; digital humanities research and services in the fields of librarianship and archives; digital humanities research on ancient texts; and digital humanities education and disciplinary development.

2.2.1 Digitization of Documentary Resources and Knowledge Mining Technologies in Digital Humanities

(1) Digitization of Documentary Resources Research in digital humanities regarding the digitization of documentary resources primarily revolves around two aspects. First, in terms of the digital preservation and development of documentary resources, information technologies such as digital photography, holographic imaging, Optical Character Recognition (OCR), 3D scanning, knowledge modeling, Virtual Reality (VR), Augmented Reality (AR), and 3D animation play a vital role. Research focuses on the application of these digitization technologies; for instance, leveraging digital technology to empower the development and utilization of historical documentary resources through the construction of "specialized historical documentary resource digital sharing platforms," or employing natural language processing and image recognition to solve the challenges of digitizing Ming and Qing dynasty archives, thereby accelerating their datafication process. Second, regarding the impact of cultural and tourism integration on the utilization of documentary resources, studies suggest that high-quality documentary resources serve as the cradle for cultivating premium cultural and creative Intellectual Property (IP) and form the cornerstone for constructing panoramic presentation models for archival resources. For example, Xiao Hongqiang deeply investigated the core el-

ements of high-quality cultural and creative IPs for Qiaopi archives and proposed corresponding cultivation strategies; Chen Hui et al. explored practical schemes for creating a panoramic presentation model in Honghu City using Red archival resources within the context of all-for-one tourism.

(2) Knowledge Mining and Visualization of Documentary Resources The application of computer technology in the knowledge mining of documentary resources is a prominent feature of digital humanities research. Relevant studies mainly involve deep learning technology, semantic web technology, and visualization technology. Deep learning and semantic web technologies enable digital knowledge mining of documentary resources, facilitating operations such as word segmentation, annotation, and data extraction for large-scale corpora, and are frequently used for sentiment analysis. For example, Li enhanced the ability of the mainstream BBiC (BERT-BiLSTM-CRF) model to represent the semantics of ancient texts by integrating structure-characteristic semantic enhancement technology, thereby improving the accuracy of punctuation and sentence segmentation in ancient books; Huang Zijing et al. utilized the BERT model to demonstrate the complexity of sentiment distribution in *The Diaries of John Rabe*. Regarding the construction of semantic web ontologies, Zhang Yunzhong et al. built an ontological model of the sacrificial system of the Sanxingdui ancient city, corroborating the application value of ontology technology in humanities research; Liu Jing et al. implemented ontology modeling and instance filling with the help of ontology editors to achieve fine-grained analysis and three-dimensional correlation of knowledge units in art works themed on the history of the Communist Party of China. Semantic web linked data can achieve the semantic organization and knowledge aggregation of related resources; for instance, Cheng Jiejing et al. introduced linked data to reconstruct the organizational system of historical and cultural resources, explaining entity attributes and determining ontological models.

Visualization technologies primarily include the visualization of mapping knowledge domains and Geographic Information System (GIS) visualization. Zhang Qiang et al. integrated knowledge reconstruction, static correlation of knowledge graphs, and GIS trajectory display to deeply analyze and visualize the spatio-temporal emotional journey of the Southern Song Dynasty poet Xin Qiji; Yang Lujia combined GIS technology with mathematical statistical methods to reveal the regional differentiation characteristics of cultural relics of private library buildings in China.

2.2.2 Digital Humanities Research and Services in the Fields of Librarianship and Archives

(1) Library Digital Humanities Services Currently, libraries are undergoing a major transformation from digital libraries to smart libraries. The rapid development of artificial intelligence technology directly empowers

libraries to achieve virtual resource reconstruction, service model innovation, and support for global academic activities. Digital humanities services represent a significant growth point for library service innovation; libraries must serve not only as providers of literature and digital resources for digital humanities outcomes but also as producers and creators of such outcomes. Dong Hongli et al. [?] utilized a SWOT-PEST analysis matrix to examine the internal and external environments of university libraries providing digital humanities services and proposed corresponding strategies. Liu Yingli [?] summarized the characteristics of digital humanities services in 20 American university libraries, highlighting the importance of conceptual reinforcement, system construction, and cross-border cooperation for improving service quality. Regarding the digitalization of university libraries, scholars have analyzed the current status of digital humanities practices in leading foreign university libraries and proposed recommendations such as improving digitalization policies, forming professional teams, strengthening librarian training, innovating service models, and emphasizing platform construction [?]. In terms of the construction and utilization of special collections in university libraries, Wen Yi [?] suggested establishing a "co-construction and sharing" large-scale collection system by formulating relevant policies. Furthermore, technologies such as natural language processing, spatiotemporal visualization, linked data, and the International Image Interoperability Framework (IIIF) have been employed to create user-centered digital special collection revelation platforms [?].

(2) Archival Research and Archival Knowledge Services from a Digital Humanities Perspective The national cultural digitalization strategy and the advancement of digital humanities research are driving a paradigm shift in archival science from "resource management" to "knowledge services." The knowledge representation of historical archives is moving from basic digitalization toward deep semanticization and knowledge-based structures [?]. Regarding the deep mining of archival value, the intervention of digital humanities technical tools has become a key research direction. For example, Li Yue [?] empowered the development of archives for Beijing as a historical and cultural city through digital technology, creating the spatial resource platform "Beijing Imprints." Fan Xuejun [?] summarized the theoretical development and core technical systems of historical archive text mining and further explored the process integration and paradigm innovation of AIGC (Artificial Intelligence Generated Content) technology in the field of text mining.

In the realm of archival knowledge services, the field is transitioning from traditional resource provision to a multi-dimensional, intelligent, and visualized knowledge support system. Li Feipeng [?] proposed the construction of a user-centered, multi-modal archival knowledge service system characterized by collaborative co-construction among diverse stakeholders. Liu Shijin et al. [?] explored strategies for constructing archival digital humanities knowledge service mod-

els oriented toward user needs, proposing relevant implementation paths and guarantee mechanisms. Li Chunyan [**?**] compared the interpretive narratives of museums with the structural narratives of archives, proposing a theoretical framework that integrates values, narratives, and mechanisms to build cross-institutional collaborative development. From the perspective of culture and tourism integration, Xu Xinxin [**?**] suggested that digital means should be used to optimize the inheritance paths of archival cultural heritage.

2.2.3 Digital Humanities Research on Ancient Books

Digital humanities research on ancient books primarily revolves around the digitization of these texts and the subsequent development and utilization of their digital resources. As documentary cultural heritage that preserves the traces of human production and daily life, ancient books face numerous challenges regarding long-term preservation due to the limitations of traditional recording media and transmission methods.

Consequently, "ancient book digitization" projects are regarded as fundamental engineering efforts for protecting and inheriting the integrity of ancient content. In recent years, scholars across various fields have actively promoted the integration of digital humanities with the digitization of ancient books, forming a new research trajectory and technical framework. Digitization serves as the top-level design for the revitalized utilization of these texts. Scholars have conducted in-depth explorations into the logical development and implementation paths of this field, reflecting on the long-term vision for the preservation of heritage and digital publishing. For instance, Li Ming posits that the exploration of ancient book digitization in China is a progressive process moving from digitization to datafication, then to knowledge-based organization, and finally to value realization. Li Shiyu et al. constructed a digitization path by integrating the digital humanities technical system with the theory of five primitives, suggesting that domestic research should adhere to Chinese characteristics, disciplinary integration, resource interoperability, standardized improvement, technical deepening, and platform construction. Zhang Ning argues that in the new era, the digital publishing of ancient books should adopt a roadmap driven by data, grounded in talent, and empowered by revitalization.

Regarding the knowledge mining of digitized ancient book resources, reuse is currently achieved primarily through the construction of knowledge bases, which facilitates the extraction of deep semantic meaning from texts via model building. For example, establishing a corpus database based on citation calculations from the *Chunqiu Zuozhuan Zhengyi* provides foundational data for analyzing the characteristics of cited works. Similarly, constructing a database centered on biographical knowledge of Western Han Dynasty Confucian classics scholars found in the *Hanshu Yiwenzhi* can reconstruct their academic lineages. Furthermore, the development of ancient text topic classification models—integrating Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and multi-head attention mechanisms—enables more precise identification of themes

within ancient texts [**?**].

2.2.4 Digital Humanities Education and Disciplinary Development

(1) Digital Humanities Education Digital humanities education refers to an educational approach that promotes digital learning and educational activities through the intersection and integration of digital technologies and the humanities. It aims to enhance the digital literacy and innovative capabilities of individuals and communities within the humanities domain. With the widespread application of digital humanities methods and tools, alongside the continuous advancement of higher education reform and the development of the "New Liberal Arts" in China, the importance of digital humanities has become increasingly prominent. Digital humanities not only broaden the horizons of archival science education but also catalyze the integration of archival science with other disciplines. For instance, some scholars have drawn upon the experiences of renowned British universities in curriculum design and interdisciplinary practice to explore methods for combining digital humanities with archival science education models and disciplinary construction. Other researchers have proposed a theoretical framework for collaborative education through the integration of industry and education, embedding digital humanities into archival science to promote an organic synthesis of the two fields. Addressing the challenges faced by public libraries in digital humanities education, scholars have suggested a series of countermeasures, including raising awareness of digital humanities education, developing digital humanities learning resource centers, and establishing multi-level service objectives. Simultaneously, the role definition and competency analysis of digital humanities librarians have become research focal points. For example:

It has been pointed out that archival digital humanities librarians must perform multiple roles, such as data managers, resource integrators, and interdisciplinary academic communicators, thereby driving a shift toward proactive and comprehensive service models. Ma Pengyun compared the professional differences between digital humanities librarians in archives and those in libraries, noting that professional capabilities should be enhanced by cultivating interdisciplinary thinking and improving both humanistic and technical literacy. Furthermore, some scholars have provided enlightening suggestions for the development of digital humanities education in China based on a systematic review of educational practices in Europe, the United States, and Japan. These suggestions include strengthening top-level design, optimizing curriculum systems, improving evaluation frameworks, and building international academic communities [**?**].

(2) Development of Digital Humanities within the Information Resource Management Discipline Group The national "14th Five-Year Plan" explicitly proposes embracing the digital era and accelerating the construction of a digital economy, society, and government. This strategy aims to drive

transformations in production methods, lifestyles, and governance through comprehensive digitalization. This strategic context has profoundly reshaped the research ecology, working environment, and career development paths within the discipline of Information Resource Management (IRM), prompting the academic community to reflect deeply on the developmental direction of digital humanities. Zhang Xiaolin et al. systematically analyzed the inherent compatibility of the transformation from Library, Information, and Archival Management toward digital humanities, emphasizing that the core values of the discipline must be maintained during this transition while offering targeted policy recommendations. Cao Ruzhong et al. argued that through deep integration and innovation with data science, IRM can cultivate unique disciplinary characteristics and discourse systems, discover new growth points, and promote high-quality disciplinary development. Yang Wen et al. and Zhang Weidong et al. contended that the digital transformation of archival science following the renaming of the first-level discipline is an inevitable trend for self-renewal and innovative development. They emphasized the urgent need to accurately position construction goals, optimize curriculum systems, and strengthen the development of teaching materials to achieve a dynamic upgrade of the disciplinary architecture. Huang Shuiqing et al. further elucidated the theoretical connotations of "Computational Humanities." By aligning with national strategic needs, the context of disciplinary renaming, and the opportunities presented by the "New Liberal Arts" construction, they analyzed the mission of the discipline and systematically outlined the framework of computational humanities using sub-disciplines such as computational linguistics, computational history, computational philology, and computational literature as examples.

In summary, driven by the dual forces of the national digitalization strategy and the "New Liberal Arts" initiative, the discipline of Information Resource Management is undergoing a profound transformation from traditional paradigms toward digital humanities or computational humanities. This transition is not only an inevitable choice empowered by technology but also an innovative practice through which the discipline actively adapts to the changes of the era, reconstructs its knowledge system, and expands its boundaries. It lays the foundation for constructing a new ecosystem for Information Resource Management with Chinese characteristics.

3 Trend Outlook

In the evolutionary trajectory spanning 2015 to 2025, digital humanities research within China's Information Resource Management (IRM) field has undergone a paradigm shift from "resource digitization" to "knowledge intelligence." The research focus has transitioned from early efforts in linked data standardization and digital resource organization toward deep semantic understanding and intelligent services driven by generative artificial intelligence in recent years. This

technological transformation is not merely an isolated iteration of tools; rather, it profoundly aligns with the functional transformation of the IRM discipline under the "New Liberal Arts" framework—moving from simple "documentary preservation" toward the higher-dimensional goals of "cultural heritage revitalization" and "knowledge reconstruction."

Furthermore, multimodal technologies, represented by GPT-4, have formed a powerful "technological singularity" at the end of this decade-long timeline. These advancements not only empower digital academic services in libraries and archives but also demonstrate the potential to reshape disciplinary boundaries within ancient book conservation and the construction of educational systems. Based on the thematic identification and analysis results discussed above, it is foreseeable that a new era of digital humanities research in IRM is beginning—one characterized by the deep coupling of technological foundations, disciplinary functions, and social values. This new era primarily manifests through the following developmental trends:

3.1 Upgrading of Digital Humanities Research Methods and Tools

Digital technology empowers the advancement of research methods and tools within the humanities, catalyzing a revolution in scientific research paradigms. Computer information technology, particularly the rise of generative artificial intelligence and deep learning models, has provided an innovative driving force and a transformative upgrade of tools for digital humanities research.

The advent of technologies such as GPT-4 and PaLM-E (Pathways Language Model with Embodied) has brought about significant changes in intelligent information processing, specifically in areas such as information understanding and analysis, retrieval and recommendation, and interaction and presentation. In contrast, traditional digital humanities research primarily relies on basic descriptive statistical methods to provide preliminary accounts of underlying data, which makes it difficult to reveal deep-seated correlations. At the current stage, deep neural network models demonstrate numerous advantages in data fitting, analysis, mining, and prediction. Examples include the application of Neural-Prophet and Long Short-Term Memory (LSTM) models in time-series forecasting, the effectiveness of Convolutional Neural Networks (CNN) in text sequence and image processing, and the breakthroughs of large language models such as BERT and ChatGPT in the field of natural language processing. Generative AI and deep learning models provide digital humanities research with powerful and flexible methodological tools, greatly enriching research techniques and analytical perspectives while driving the in-depth development of the field. Future research must address the challenges that digital humanities may encounter during data collection and processing, including potential risks to data privacy and security. By further analyzing and exploring corresponding solutions, scholars can promote more profound and comprehensive digital humanities research within the field of information resource management.

3.2 Enhancement of Digital Humanities Service Levels in Libraries and Archives

Currently, the development of machine learning technologies, particularly generative artificial intelligence, has significantly enhanced the efficiency of text analysis and image recognition. These advancements have optimized user interaction experiences and elevated the digital humanities services provided by libraries and archives to a new level. For instance, multimodal pre-trained models, represented by GPT-4, have demonstrated keen insight into user needs and preferences through their exceptional capabilities in semantic search and intelligent information processing, enabling the delivery of smart and personalized services. As technology progresses and user demands evolve, libraries and archives will continue to explore the utilization of various advanced technologies to improve the quality of digital humanities services and satisfy user requirements.

As the primary institutions for the centralized storage and management of national documentary and information resources, libraries and archives serve as the fundamental guarantee for cultural heritage, academic research, talent cultivation, and social services. Enhancing the level of related digital humanities services will undoubtedly become a critical research topic in the field of information resource management. Future research will seek innovative pathways to improve the digital humanities service capabilities of libraries and archives by synthesizing advanced practical experiences and successful case studies.

3.3 Generative AI and Multimodal Technologies Empowering Ancient Book Digitization

Ancient texts serve as vital carriers of cultural heritage, embodying profound historical and cultural significance. The digitization of these documents not only preserves cultural legacies but also revolutionizes the methods of knowledge acquisition, thereby enhancing the accessibility and utilization of ancient literature. Currently, the emergence of generative artificial intelligence (AI) and multimodal technologies has injected new vitality into the digitization of ancient texts, significantly improving both the quality and efficiency of digital conversion processes.

For instance, multimodal large models can perform in-depth analysis of document images to extract deep-seated, multi-dimensional data. By transforming ancient texts into carrier-independent data formats, these technologies broaden the channels for knowledge dissemination and satisfy diverse user needs. In the realm of natural language understanding for ancient texts, generative large language models, represented by ChatGPT, can automatically translate and annotate classical prose. This assistance deepens content comprehension and accelerates the overall research process.

Regarding the construction and utilization of digital resource platforms for ancient texts, intelligent agent programs can conduct comprehensive analyses

of submitted information, automatically retrieving and optimizing application tools. This enables highly efficient digital searching and allows users to manage massive amounts of information with ease by simplifying complex workflows. Consequently, by leveraging generative AI technology, it is possible to construct intelligent digital resource platforms that integrate various forms of ancient text data. Furthermore, the application of AI agent technology can provide personalized services and recommendations, ultimately enhancing the convenience of studying and applying ancient literature.

3.4 Interdisciplinary Exchange and Generative AI Driving Transformation in Digital Humanities Education Systems

In recent years, digital humanities (DH) education has increasingly become a focal point of academic attention, with numerous universities establishing digital humanities research centers or launching specialized degree programs. Within the discipline of Information Resource Management in China, sub-disciplines such as "Archival Science," "Digital Humanities," and "Ancient Books Preservation and Philology" have been explicitly defined. However, digital humanities education in China remains in its nascent stages, and the educational framework has yet to be fully matured. Consequently, there is a pressing need for an in-depth exploration of corresponding educational models in the future.

Future research should focus on pedagogical methods, curriculum systems, course content, and evaluation frameworks, thereby establishing a solid foundation for cultivating high-quality professionals in the field of digital humanities. Among these components, interdisciplinary collaboration will be one of the primary focal points for future research in Digital Humanities (DH) education and disciplinary development. Such collaborative efforts facilitate the integration of methodologies and technologies across diverse fields, thereby fostering research innovation. For instance, computational methods are now widely applied within the realms of literature, history, and philosophy to enable large-scale data analysis, giving rise to entirely new research paradigms. Furthermore, interdisciplinary academic exchange within the DH field contributes to the cultivation of professionals equipped with diverse perspectives and skill sets. Future research should emphasize the innovation of pedagogical approaches and the development of students' interdisciplinary thinking within this context.

Simultaneously, the rapid advancement of generative artificial intelligence has presented a significant opportunity for the transformation of Digital Humanities education. Since 2023, ChatGPT Plus has demonstrated unprecedented capabilities in natural language processing, particularly in text generation and programming code synthesis. Consequently, DH teaching methodologies may undergo profound changes; foundational courses could potentially be replaced by ChatGPT-like applications, shifting the role of the instructor toward the cultivation of students' creativity, critical thinking, and problem-solving abilities. In summary, future research directions will likely center on interdisciplinary collaboration, the impact of generative AI on DH education, and the comprehensive

restructuring of the Digital Humanities educational framework.

4 Conclusion

This study utilizes the BERTopic topic model and the Gemini large language model (LLM) to systematically review the literature on digital humanities within the field of Information Resource Management (IRM) in China from 2015 to 2025. The research identifies four core research clusters: digitization of documentary resources and knowledge mining techniques in digital humanities; digital humanities research and services in the fields of library and archival science; digital humanities research on ancient texts; and digital humanities education and disciplinary development. Digital humanities research in China' s IRM field is currently undergoing a paradigm shift from basic resource organization to deep "data-intelligence"empowerment. The focus of research has transitioned from early concerns with linked data standardization and the construction of static resource repositories toward deep semantic understanding, multimodal interaction, and intelligent knowledge services driven by generative artificial intelligence (AI). In the fields of ancient text revitalization and interdisciplinary education, the intervention of AI technology is reshaping disciplinary boundaries.

The contributions of this paper are primarily reflected in two aspects. First, it offers methodological innovation by constructing a closed-loop analysis framework of "BERTopic clustering + LLM naming + manual review." This approach effectively overcomes the fragmentation issues in semantic understanding inherent in traditional topic models, thereby enhancing the precision of topic identification and its alignment with academic discourse. Second, it expands the theoretical perspective by systematically outlining the digital and intelligent transformation path of China' s IRM discipline following its official renaming. This provides empirical evidence for future disciplinary positioning and paradigmatic evolution. Despite these efforts to provide a comprehensive perspective, some limitations remain. First, the data source is primarily based on CNKI and does not yet cover international journals or practical reports, which may lead to a bias regarding international frontier trends and industry applications. Second, while LLMs demonstrate high efficiency in topic naming, algorithmic bias may still exist in defining the subtle nuances of specific professional terminology, necessitating expert manual intervention. Finally, due to the limited timeframe of the data, the fundamental impact of generative AI on digital humanities research is still in the early stages of observation, and its long-term transformative effects require continuous tracking. Future research should further integrate multi-source heterogeneous data, deeply explore the construction standards, ethical rules, and social values of vertical large models for digital humanities generated by AI, and build a more mature autonomous knowledge system for digital humanities within the IRM field that possesses Chinese characteristics.

[REFERENCES] [1] Sun Hui. Analysis and Reflection on the Research Framework of Digital Humanities [J] . Information Studies: Theory & Practice, 2018,

41(7): 7-13. [2] Wu Jinglin, Wang Jiaying, Shan Jing. Visual Analysis of Domestic Digital Humanities Research Based on CiteSpace [J] . Jiangsu Science & Technology Information, 2023, 40(9): 19-24. [3] Huang Shuiqing. Humanities Computing and Digital Humanities: Concepts, Problems, Paradigms and Key Links [J] . Library Development, 2019(5): 68-78. [4] Yang Hetian. Visual Analysis of Digital Humanities Research in the Field of Information Resource Management in China Based on CiteSpace [J] . China High-Tech, 2023(16): 80-81, 86. [5] Wang Xiaoguang. The Emergence, Development and Frontier of "Digital Humanities" [C] //National University Social Science Research Management Association. Innovation and Development of Philosophy and Social Sciences. Wuhan: Wuhan University Press, 2010: 207-221. [7] Ping Shuo, Li Shiyu, Zou Dingjie. A Review of Domestic Digital Humanities Research Based on Bibliometrics [J] . Library Journal, 2022, 44(6). [8] Liu Yang. Identification of Content and Characteristics of Digital Humanities Research in the Field of Information Resource Management in China [J] . Information Science, 2023, 41(11). [9] Pei Lei, Jin Yuan, Liu Jiangfeng. Topic Identification and Trend Outlook of Digital Humanities Research in the Field of Information Resource Management in China [J] . Journal of Guangdong Open University, 2024, 33(1): 94-102. [10] Zhang Xu, Hong Yixuan, You Jian. Research Hotspots and Trends of Digital Humanities in China [J] . Library Work and Study, 2023(10): 69-76. [11] Wu Jinglin, Wang Jiaying, Shan Jing. Visual Analysis of Domestic Digital Humanities Research Based on CiteSpace [J] . Jiangsu Science & Technology Information, 2023, 40(9): 19-24. [12] Shi Haiyan, Niu Guoyi. Current Status and Prospects of Digital Humanities Research in the Field of Library, Information and Archives in China [J] . Hebei Library Journal, 2023, 36(2). [13] Wang Dongbo, Liu Chang, Zhu Zihe, et al. SikuBERT and SikuRoBERTa: Construction and Application of Pre-trained Models for Siku Quanshu in Digital Humanities [J] . Library Tribune, 2022, 42(6): 31-43. [14] Liu Chang, Wang Dongbo, Hu Haotian, et al. Research on Automatic Word Segmentation of Ancient Texts Integrating External Features for Digital Humanities—Taking SikuBERT Pre-trained Model as an Example [J] . Library Tribune, 2022, 42(6): 44-54. [15] Geng Yundong, Zhang Yiqin, Liu Huan, et al. Research on Automatic Part-of-Speech Tagging of Ancient Chinese Texts for Digital Humanities—Taking SikuBERT Pre-trained Model as an Example [J] . Library Tribune, 2022, 42(6): 55-63. [16] Xu Runhua, Wang Dongbo, Liu Huan, et al. Research on Automatic Summarization of Zizhi Tongjian for Ancient Book Digital Humanities—Taking SikuBERT Pre-trained Model as an Example [J] . Library Tribune, 2022, 42(12): 129-137. [17] Liu Jiangfeng, Feng Yutong, Wang Dongbo, et al. Research on SikuBERT-enhanced Entity Recognition of Historical Records from the Perspective of Digital Humanities [J] . Library Tribune, 2022, 42(10): 61-72. [18] Sun Wenlong, Zhang Yiqin, Wang Fanming, et al. Research on Semantic Vocabulary Extraction of Ancient Texts for Digital Humanities—Taking SikuBERT Pre-trained Model as an Example [J] . Library Tribune, 2022, 42(10): 31-41. [19] Liu Jiangfeng, Liu Chufei, Qi Yue, et al. Practical Exploration of AIGC Assisting Digital Humanities Research: Research on Ancient Poetry Generation Driven by SikuGPT [J] . Information Studies: Theory & Practice, 2023, 46(5):

23-31. [20] Ethan-yt/guwenbert. [21] Lü Ruijuan, Zhang Jingbei, Yan Dan, et al. Review of Innovative Development of AIGC and GLAM—Based on the Frontier Academic Forum "Generating the Future: AIGC and GLAM Innovative Development"[J]. Journal of Agricultural Library and Information Science, 2023, 35(5): 27-36. [22] Wang Jingjing, Hong Yun, Ye Ying. Discussion on Reshaping Digital Humanities by GPT-type Technology Application [J] . Information Studies: Theory & Practice, 2023, 46(6). [23] Cao Lina, Tang Xijin. Analysis of BBS Topic Evolution Trend Based on Topic Model [J] . Journal of Management Sciences in China, 2014, 17(11). [24] Nie Yaqing, Wu Tingzhang, Wang Ruojia, et al. Research on Topic Mining and Development Evolution of Health Informatics Based on BERTopic Model [J] . Information Science, 2024, 42(4): 98-110, 118. [25] Grootendorst M. BERTopic: neural topic modeling with a class-based TFIDF procedure [J] . arXiv preprint arXiv: 220305794, 2022. [26] Xu Hanqing, Teng Guangqing. Opportunities and Challenges: Library Topic Text Mining in AI Environment Based on BERTopic [J] . Information Science, 2024, 42(4): 119-128. [27] Zhang Junai, Liu Haijun. Application of ChatGPT in Middle School Geography Teaching: Scenarios, Limitations and Breakthrough Strategies [J] . Geography Teaching, 2024(5): 16-20. [28] Chen Haiyu, Wang Cong, Chen Yu, et al. Exploration of the Construction of Digital Sharing Platforms for Special Historical Literature Resources from the Perspective of Digital Humanities—Taking the Historical Literature of Southwest Associated University as an Example [J] . Digital Library Forum, 2024, 20(4): 22-32. [29] Tian Chengbin, Wang Ning. Construction and Development of Ming and Qing Archival Information Resources from the Perspective of Digital Humanities [J] . Archives Management, 2023(2): 81-83. [30] Xiao Hongqiang. Cultivation Strategy of Archival Cultural and Creative IP Driven by Digital Humanities and Cultural-Tourism Integration—Taking Qiaopi Archival Development as an Example [J] . Archives Management, 2023(4): 73-75. [31] Chen Hui, Liu Sainan, Liao Xueqin. Exploration of Panoramic Presentation Mode of Red Archival Resources under the Background of All-for-one Tourism—Taking the Development and Utilization of Red Archives in Honghu City as an Example [J] . Archives Science Study, 2023(4): 108-117. [32] Li Peiqi, Wang Hao, Ren Qiutong, et al. Research on Semantic-enhanced Ancient Book Punctuation Recognition Method Integrating Structural Characteristics [J] . Journal of the China Society for Scientific and Technical Information, 2023, 42(2): 150-163. [33] Huang Zijing, Qiu Yuqian, Shen Tong, et al. Sentiment Recognition and Analysis of The Diaries of John Rabe from the Perspective of Digital Humanities [J] . Library Tribune, 2023, 43(3): 54-63. [34] Zhang Yunzhong, Qiu Jing. Construction and Scenario Application of Sacrificial Map of Sanxingdui Ancient City for Digital Humanities Research [J] . Library Journal, 2024, 43(2): 73-83. [35] Liu Jing, Wang Meng, Liu Guofeng, et al. Research on Knowledge Graph Construction of Red Cultural Resources from the Perspective of Digital Humanities [J] . Library Journal, 2023, 42(7): 102-109, 125. [36] Cheng Jiejing, Wang Puyu. Research on Relational Organization of Biographical Historical Materials in Ancient Books—Taking the Group of Western Han Confucian Scholars in Hanshu Yiwenzhi as an Example [J] . Library Tribune, 2023, 43(3): 64-74. [37] Zhang Qiang, Gao Jingsong, Long

Jiaqing, et al. Visualization Research on Spatio-temporal Emotional Trajectory of Poets Based on Knowledge Reconstruction—Taking Xin Qiji as an Example [J] . Journal of the China Society for Scientific and Technical Information, 2023, 42(6): 729-739. [38] Yang Lujia. Influencing Factors of Regional Differentiation of Cultural Relics of Private Library Buildings in China for Digital Humanities [J] . Library Tribune, 2024, 44(2): 79-91. [39] An Jie. Exploration of Digital Humanities Practice in Columbia University Library [J] . Library and Information Service, 2018, 62(8): 48-53. [40] Yu Yaxiu, Li Xin. Digital Humanities Service Practice in American University Libraries—Taking UCLA as an Example [J] . Library Tribune, 2018, 38(9): 30-37. [41] Dong Hongli, Yu Xingshang. Research on Digital Humanities Service Strategy of University Libraries from the Perspective of SWOT-PEST [J] . Library Theory and Practice, 2023(1): 69-76. [42] Liu Yingli. Practice and Enlightenment of Digital Humanities Services in American University Libraries [J] . Library Work and Study, 2023(2): 56-63. [43] Dong Mengyuan, Tang Shukun. Digital Humanities Practice and Enlightenment of Bodleian Library, University of Oxford [J] . Digital Library Forum, 2023, 19(9): 64-70. [44] Wu Jianzhong. From Digital Library to Smart Library: Opportunities, Challenges and Innovations [J] . Library Journal, 2021, 40(12): 4-11. [45] Wang Chunying, Zhu Kunhao, Li Chunqiu, et al. Practice and Enlightenment of Cultural Heritage Digitization in American University Libraries [J] . Library, 2023(6): 64-74. [46] Zhao Fang. Current Status and Enlightenment of Digital Humanities Librarians in American University Libraries [J] . Library, 2024(1): 53-59. [47] Wen Yi. Analysis of Special Collection Resource Construction Strategy of University Libraries under the Background of Digital Humanities [J] . Library Work and Study, 2023(1). [48] Zhang Yi, Zhao Chenming, Chen Dan. Practice and Reflection on Special Collection Resource Construction of University Libraries in Digital Humanities—Taking the Construction of Full-text Special Collection Database of Modern Chinese Translations as an Example [J] . Journal of the National Library of China, 2023, 32(1): 68-78. [49] Zhao Shuo. Knowledge Representation of Historical Archival Resources from the Perspective of Digital Humanities [J] . Shanxi Archives, 2025(9): 112-114. [50] Li Yue. Analysis of Archival Development Path for the Protection of Beijing as a Historical and Cultural City from the Perspective of Digital Humanities [J] . Beijing Archives, 2023(4): 40-42. [51] Fan Xuejun. Application of Digital Humanities Tools in Text Mining of Historical Archives [J] . Shanxi Archives, 2025(11): 110-112. [52] Li Feipeng. Memory Reconstruction and Value Release Strategy of Multimodal Archives in the Digital Age [J] . Shanxi Archives, 2025(10): 157-160. [53] Liu Shijin, Tian Weifu. Innovation of Archival Digital Humanities Knowledge Service Mode Oriented to User Needs under the Support of Intelligent Technology [J] . Shanxi Archives, 2025(9): 131-133. [54] Li Chunyan. Experience Mutual Learning and Fusion Innovation of Digital Exhibitions in Museums and Archives [J] . Shanxi Archives, 2025(9): 152-154. [55] Xu Xinxin. Protection and Inheritance of Archival Cultural Heritage from the Perspective of Digital Humanities Tourism [J] . Shanxi Archives, 2025(2): 80-83. [56] Li Ming. Evolutionary Characteristics and Internal Logic Analysis of Ancient Book Digitization Research in China Driven by Digital Humanities [J] .

Library Theory and Practice, 2023(4): 121-127. [57]Li Shiyu, Zhang Xiangxian, Shen Wang, et al. Analysis of Current Research Status and Path Construction of Ancient Book Digitization at Home and Abroad [J] . Modern Information, 2023, 43(11): 4-20. [58] Zhang Ning. Target Evolution, Turning and Approach of Ancient Book Digital Publishing in China [J] . Publishing Science, 2023, 31(4): 5-16. [59] Sun Yan, Liu Liu, Wang Dongbo. Computational Humanities Research on Cited Books in Chunqiu Zuozhuan Zhengyi [J] . Library and Information Service, 2023, 67(2). [60] Wu Shuai, Yang Xiuzhang, He Lin. Research on Topic Recommendation of Ancient Book Texts Based on Multi-view Fusion DJ-TextRCNN [J] . Journal of the China Society for Scientific and Technical Information, 2024, 43(1): 61-75. [61] Wu Shuai, Yang Xiuzhang, He Lin, et al. Research on Topic Classification of Ancient Book Texts Based on SWPF2vec and DJ-TextRCNN[J]. Journal of the China Society for Scientific and Technical Information, 2024, 43(5): 601-615. [62] Burdick A, Drucker J, Lunenfeld P, et al. Digital humanities [M] . Cambridge: MIT Press, 2012. [63] Li Mingxia, Yang Xinxin, Wang Luyao, et al. Progress of Digital Humanities Education in the UK and Its Enlightenment to the Development of Archival Discipline [J] . Archives Management, 2023(2): 118-123. [64] Kong Rong. Theoretical Framework and Practical Path of Collaborative Education through Integration of Industry and Education in Digital Humanities Embedded Archival Science [J] . Archives Management, 2023(6): 112-114, 117. [65] Yang Yandi. Analysis of Digital Humanities Education Strategy in Public Libraries [J] . Library Work and Study, 2023(S1): 91-94. [66] Jiao Jun. Analysis of Competency Transformation Based on Role Positioning of Archival Digital Humanities Librarians [J] . Archives Management, 2023(3): 89-90, 94. [67] Ma Pengyun. Comparative Analysis of Professional Competency of Digital Humanities Librarians in Libraries and Archives [J] . Archives Management, 2023(3): 86-88. [68] Jing Xiaoyue. Digital Humanities Education from a Global Perspective: Practical Enlightenment from Europe and America [J] . Shanxi Archives, 2024(5): 100-102. [69] Shi Liangliang. Research on Experience and Enlightenment of Digital Humanities Education in Japan [J] . Shanxi Archives, 2024(5): 96-99. [70] Zhang Xiaolin, Cao Ruzhong, Xiang Hong, et al. Theoretical Fit and Logical Approach of the Transformation and Development of Library, Information and Archives Discipline toward Digital Humanities under the Background of New Liberal Arts Construction [J] . Library, 2023(5): 15-20, 33. [71] Cao Ruzhong, Guo Hua, Xiong Hongjun, et al. Research on the Positioning of Digital Transformation and Development of Library, Information and Archives Discipline under the Background of New Liberal Arts Construction [J] . Library Theory and Practice, 2023(5): 17-23. [72] Yang Wen, Yao Jing. Digital Transformation of Archival Discipline Construction and Talent Cultivation—Reflections Based on the Renaming of the First-level Discipline of Library, Information and Archive Management to Information Resource Management [J] . Library and Information Service, 2023, 67(1): 99-107. [73] Zhang Weidong, Chen Xipeng, Yang Sihan. Expansion and Deepening of Archival Academic Research Topics under the First-level Discipline of Information Resource Management [J] . Library and Information Service, 2023, 67(1): 90-98. [74] Huang Shuiqing, Liu Liu, Wang

Dongbo. Connotation, System and Opportunities of Computational Humanities Discipline [J] . Library and Information, 2023(1): 1-11, 153, 145.

Topic Identification, Interpretation Analysis and Trend Outlook of Digital Humanities Research in the Field of Information Resource Management in China
Pei Lei, Jin Yuan, Liu Jiangfeng (School of Information Management, Nanjing University, Nanjing 210023, China)

Abstract

Purpose/Significance In recent years, breakthroughs in artificial intelligence technology have brought profound impacts on digital humanities research in the field of information resource management. Conducting an in-depth analysis of the thematic characteristics and development trends of digital humanities research in China' s information resource management field aims to provide references for the expansion and disciplinary positioning of this domain.

Method/Process Using journal articles indexed in the China National Knowledge Infrastructure (CNKI) from 2015 to 2025 as the data source, the BERTopic topic model and the Gemini large language model were employed to identify eight topics from the digital humanities research literature in China' s information resource management field. Through manual reading and content analysis, these topics were categorized into four aspects for interpretation and trend forecasting.

Result/Conclusion Digital humanities research in China' s information resource management field can be summarized into four aspects: the digitization of literature resources and knowledge mining technologies in digital humanities, digital humanities research and services in libraries and archives, digital humanities research on ancient texts, and digital humanities education and disciplinary development. Future research will focus on the upgrading of digital humanities research methods and tools, the enhancement of digital humanities services in libraries and archives, the revitalization of ancient texts through generative AI and multimodal technologies, and interdisciplinary exchanges along with the transformation of digital humanities education systems driven by generative AI.

Keywords

Abstract

In the context of the digital humanities, this study explores the application of advanced topic modeling techniques within the field of Information Resource Management. By leveraging the BERTopic framework, we aim to enhance the precision and interpretability of topic identification in large-scale scholarly corpora. The integration of deep learning and machine learning methodologies allows for a more nuanced understanding of evolving research trends and thematic structures. This research contributes to the methodological toolkit of

Figure 1

Figure 2: Figure 1

digital humanities by providing a robust approach to automated content analysis.

1. Introduction

The rapid expansion of digital scholarship has necessitated the development of more sophisticated tools for information resource management. Traditional methods of topic modeling, such as Latent Dirichlet Allocation (LDA), often struggle with the semantic complexities and contextual nuances inherent in academic discourse. As the field of digital humanities matures, there is an increasing demand for computational techniques that can bridge the gap between quantitative data processing and qualitative interpretative analysis.

2. Methodology

2.1 BERTopic Framework

This study utilizes BERTopic, a neural topic modeling technique that leverages BERT (Bidirectional Encoder Representations from Transformers) embeddings and a class-based TF-IDF (c-TF-IDF) procedure. Unlike traditional probabilistic models, BERTopic creates dense clusters of documents, allowing for the identification of coherent topics while maintaining the semantic integrity of the original text.

The process involves three primary stages: 1. **Embedding**: Transforming document text into high-dimensional vector representations using pre-trained transformer models. 2. **Dimensionality Reduction and Clustering**: Utilizing UMAP (Uniform Manifold Approximation and Projection) to reduce vector space and HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) to identify dense clusters. 3. **Topic Representation**: Applying c-TF-IDF to extract the most representative keywords for each cluster, thereby defining the topic.

2.2 Data Collection and Preprocessing

The dataset comprises a comprehensive collection of academic publications within the domain of Information Resource Management. Preprocessing steps include the removal of noise, such as metadata artifacts and non-textual elements, to ensure the quality of the input for the embedding model.

3. Results and Analysis

Through the application of BERTopic, we identified several core thematic clusters that define the current landscape of Information Resource Management. The model demonstrated superior performance in capturing fine-grained topics compared to baseline LDA models.

As shown in , the identified···