
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202304.00295

Review of Research Progress on the Digitization of Ancient Chinese Books (2009-2019): Postprint

Authors: Li Mingjie, Zhang Xianke, Chen Mengshi

Date: 2023-04-01T16:15:51+00:00

Abstract

[Purpose/Significance] To review the research history and current status of digitalization of ancient books in order to explore its development trends. [Method/Process] Co-occurrence mapping analysis of authors and keywords in digitalization of ancient books research over the past decade was conducted using CiteSpace software; the main research fields and topic distribution of journal articles and National Social Science Fund projects were analyzed; and the primary functions of existing digitalization products for ancient books were examined and evaluated. [Results/Conclusion] Three major overall characteristics and five main problems of digitalization of ancient books research in China over the past decade are summarized; it is pointed out that driven by academic inheritance, user demands, and information technology development, orientation towards knowledge services represents the inevitable trend for future digitalization of ancient books research.

Full Text

Review on the Research Progress of Ancient Books Digitization (2009-2019)

Li Mingjie, Zhang Xianke, Chen Mengshi

School of Information Management, Wuhan University, Wuhan 430072

Abstract:

[Purpose/Significance] This paper systematically investigates the research status of ancient books digitization in China to explore its development trends. [Method/Process] Using CiteSpace software, we conducted co-occurrence mapping analysis of authors and keywords in ancient books digitization research over the past decade. We analyzed the main research fields and thematic distribution of journal papers and national social science fund projects, and examined

the primary functions of existing ancient books digitization products through investigation and hands-on experience. [Result/Conclusion] We summarize three general characteristics and five major problems in China's ancient books digitization research in recent ten years, and point out that under the drives of academic inheritance, user demand, and information technology development, knowledge service-oriented digitization represents the inevitable future trend.

Keywords: ancient books digitization; knowledge service; academic inheritance; user demand; technology development

Classification Number: G255.1

DOI: 10.13266/j.issn.0252-3116.2020.06.015

Since the late 1970s, research and practice in ancient books digitization have undergone nearly 40 years of development. From initial text input and index compilation to large-scale construction of various types of ancient books databases (such as bibliographic, full-text, and image databases), and today to deep processing of digital ancient texts using computer information processing technologies—including automatic punctuation, automatic collation, automatic annotation, semantic indexing, text clustering, and knowledge mining—people's understanding of ancient books digitization has continuously evolved. Initially, ancient books digitization was understood simply as converting textual information from paper ancient books into digital form for long-term preservation and utilization, then leveraging computer technology to develop convenient and efficient full-text retrieval functions. With the development of digital information technology, especially the rise of the “digital humanities” concept at the beginning of this century, people discovered that converted ancient books digital texts could serve as data for further processing and organization, enabling the development of new knowledge products through a series of data processing techniques. The concept of ancient books digitization has thus advanced toward a deeper “knowledge domain.” As ancient books digitization enters a new stage, the massive data following digital text conversion has become an unavoidable subject for in-depth knowledge processing for ancient books collation and researchers. Therefore, this paper aims to summarize ancient books digitization research over the past decade, analyze its general characteristics, existing problems, and internal driving forces, and explore future development directions for ancient books digitization in combination with the functional requirements of humanities and social sciences research, providing references for ancient books collation and related humanities and social sciences research.

2. Overview of Ancient Books Digitization Research (2009-2019)

2.1 Basic Situation of Relevant Journal Papers

Using the CNKI database with “journal” as the source and “subject” as the field, we conducted a search with primary terms including “ancient books,” “classics,” “rare books,” “ancient texts,” “ancient literature,” “ancient books collation,”

and “ancient books protection,” and secondary terms including “digitization,” “electronization,” “computer,” and “database.” The search used logical OR for both sets of terms, covered 2009-2019, included synonym expansion, and was conducted on July 1, 2019, yielding 1,053 results. After screening and elimination, 759 relevant results were obtained. We then used CiteSpace software for quantitative and visual analysis, producing author collaboration network and keyword co-occurrence diagrams, shown in [Figure 1: see original paper] and [Figure 2: see original paper].

The author collaboration network reflects important contributors and connections among researchers in the field. In the network diagram, node size represents the number of publications by an author. [Figure 1: see original paper] shows that authors in ancient books digitization research are relatively dispersed, indicating that researchers in this field over the past decade come from complex backgrounds with diverse research directions. In terms of publication volume, scholars such as Ge Huaidong, Mao Jianjun, and Zhang Wenliang rank at the top. Some researchers conduct academic exploration through collaboration, such as Li Bing, Fu Yongchi, and Wang Rui, who all come from the same research institution and focus primarily on traditional Chinese medicine ancient books. The dispersed author network reveals two issues: first, researchers come from different disciplinary backgrounds, confirming the interdisciplinary nature of current ancient books digitization research involving library and information science, literature and history, and computer science, yet cooperation between different types of research institutions remains limited and the interdisciplinary nature needs strengthening; second, researchers are striving to explore new topics, resulting in relatively dispersed research themes that have not yet formed a stable and complete research system.

Keywords represent the core themes and main content of articles. Through keyword co-occurrence analysis, we can identify research themes in the field. In the keyword co-occurrence diagram, node size represents keyword frequency—more frequent keywords correspond to larger nodes. Lines between nodes represent co-occurrence relationships, with line thickness indicating co-occurrence strength. [Figure 2: see original paper] marks keywords with high co-occurrence frequencies with square frames and special keywords with elliptical frames. Based on [Figure 2: see original paper] and relevant literature, we identify six main research areas in ancient books digitization over the past decade:

- (1) **Research on the concept and current status of ancient books digitization.** This category includes a relatively large number of papers, indicating that the academic community pays considerable attention to basic theoretical issues while emphasizing integration with practice. Discussion on the concept of ancient books digitization began in the mid-to-late 1990s and continues to this day. People’s understanding of ancient books digitization has deepened with its development, roughly experiencing three stages: storage medium conversion, text organization and retrieval, and data processing and knowledge service. Research on the current status of

ancient books digitization mostly involves summarization and analysis of processes, problems, countermeasures, and trends, with representative papers including “Current Status and Significance of Chinese Ancient Books Digitization” [1], “Review on the Progress and Research of Ancient Books Digitization in China” [2], “Problems and Countermeasures of Ancient Books Digitization in Mainland China” [3], and “Dilemmas and Response Strategies of Ancient Books Digitization” [4].

- (2) **Ancient books digitization research related to ancient books preservation.** This research views ancient books digitization as an effective means of regenerative preservation, with representative papers including “Strategies and Suggestions for Ancient Books Preservation and Development” [5], “Fidelity Issues in Ancient Books Digitization” [6], “Research on the Protection and Development of Historical and Cultural Resources under Digital Survival—Centered on Shaanxi” [7], and “Exploration of Long-term Digital Preservation Mechanisms for Ancient Books in Libraries in the Digital Humanities Era” [8].
- (3) **Research on ancient books digital resource construction related to library operations.** This mainly involves development of various types of ancient books databases, integration of ancient books digital resources, and collaborative construction and sharing of ancient books digital resources. As the main institutions for ancient books collection, management, development, and utilization, libraries’ research on ancient books digitization focuses on digital resource construction, with representative papers such as “Research on Standards and Evaluation of Ancient Books Bibliographic Databases” [9], “Discussion on Construction Standards and Norms of Ancient Books Bibliographic Databases in China” [10], “Exploration of Development Trends of Ancient Books Digital Resource Databases” [11], “Ancient Books Digitization and Collaborative Construction and Sharing” [12], “Research on Integration of Library Ancient Books Digital Resources” [13], and “Research on Aggregation of Ancient Books Resources” [14]. Additionally, construction of ancient books version databases and thematic knowledge bases has attracted scholars’ attention, such as “Construction of Chinese Mongolian Ancient Books Version Database Based on Unicode” [15] and “On the Construction of Ancient Books Version Databases” [16] and “Development Status of Traditional Chinese Medicine Ancient Books Knowledge Base and Construction of Ancient Books Health Preservation Knowledge Base” [17].
- (4) **Digitization research on specialized categories of ancient books.** This involves digitization of traditional Chinese medicine ancient books, minority nationality ancient books, local documents, genealogies, local gazetteers, and rubbings, such as “Research on Digital Collation Plans for Traditional Chinese Medicine Ancient Books” [18], “Research and Implementation of Digital Genealogy” [19], “Digitization and Practice of Transcribed Local Gazetteer Ancient Books” [20], “Digitization of Stone Rub-

bings at the National Library” [21], “Research on Digital Co-construction and Sharing Guarantee System for Mongolian Resources” [22], and “Construction and Platform Design of Traditional Chinese Medicine Ancient Books Digitization” [23]. Such research mostly proceeds from the perspective of literature preservation and cultural inheritance, with few breakthroughs in digitization technology itself.

- (5) **Research on collaboration mechanisms for ancient books digitization.** Issues such as duplicate topics and lack of planning have long troubled ancient books digitization. Many scholars have realized that national-level mechanism innovation is necessary to solve these problems. Representative papers include “Reflections on Current Stage Ancient Books Digitization—On the Importance of National Action in Ancient Books Digitization” [24], “Construction of Coordinated Mechanisms for Ancient Books Digitization Work” [25], “On the National Macro Policy System for Ancient Books Digitization” [26], and “Preliminary Exploration of the Main Components and Collaboration Mechanisms of Chinese Ancient Books Digitization” [27].
- (6) **Research combining ancient books digitization with digital humanities,** including linked data, knowledge mining, Geographic Information Systems (GIS), and knowledge organization. In the past five years, such papers have been increasing annually, indicating that ancient books digitization research has entered a new stage oriented toward knowledge service. Representative papers include “Application Research of Named Entity Recognition in Gazetteer Content Mining” [28], “Connotation of ‘Digital Humanities’ and Deep Development of Ancient Books Digitization” [29], “‘Digital Humanities’ and Historical Geography Informatization Research” [30], “Research on Chinese Ancient Books Geographic Information System Based on GIS” [31], “Large-scale Ancient Books Text Visualization Analysis and Mining for Digital Humanities Research” [32], “Digitization, Datafication, and Text Mining of Local Historical Documents: Taking the Chinese Local Historical Documents Database as an Example” [33], “Structured Knowledge Representation of Commentary Literature Based on Ontology and XML” [34], and “Research on Open Application of Chinese Historical Geographic Linked Data in Library Digital Humanities Projects” [35]. In traditional Chinese medicine ancient books digitization, remarkable achievements have been made in knowledge mining and reorganization, such as “Exploration of Building Traditional Chinese Medicine Ancient Books Knowledge Base Based on Ontology” [36], “Research on Knowledge Representation of Traditional Chinese Medicine Ancient Books Prescriptions Based on Knowledge Units” [37], “Research on Construction of Semantic Network of Knowledge Concepts in Warm Disease Ancient Books” [38], and “Construction of Traditional Chinese Medicine Academic Inheritance Context Based on Ontology” [39].

2.2 Basic Situation of Relevant National Social Science Fund Projects

Using the National Social Science Fund Project Database [40], we set the time range to 2009-2018 and limited the discipline category to “Library, Information and Documentation Science,” retrieving 54 projects related to ancient books digitization. In terms of project numbers, the overall trend is increasing year by year, with only 1 project in 2009 (the minimum) and 13 projects in 2017 (the maximum). In terms of project types, Western region projects are the most numerous with 23 projects (42.59%), followed by general projects with 21 (38.89%), youth projects with 9 (16.67%), and major projects with 1 (1.85%). Thematic distribution can be roughly divided into the following categories:

- (1) **Rescue collation and protection of minority nationality ancient books and traditional Chinese medicine ancient books.** This category has the largest number with 21 projects, typical examples include “Research on the Value of Buyi Nationality Pictographic Ancient Yue Books and Its Resource Rescue, Construction and Development Utilization” (10XTQ010), “Research on the Value of Mongolian Nationality Documents and Its Document Resource Construction” (10XTQ012), “Research on Yunnan Dai Medicine Ancient Books Literature Collation and Basic Database Construction” (11CTQ041), “Research on Salar Nationality Document Database Construction” (11XTQ011), “Research on Sichuan Southern School Tibetan Medicine Ancient Books Literature Rescue Collation and Database” (13XTQ012), “Research on Collection, Collation and Digitization of Tibetan Medicine Ancient Books Literature in Qinghai Region” (16BTQ051), “Research on Collection, Collation and Digitization of Dong Medicine Ancient Books Literature in Hunan-Guizhou-Guangxi Border Area” (17CTQ017), and “Research on Digitization and Collation of Existing Ningxia Old Gazetteers” (18XTQ007).
- (2) **Digital projects related to oral history, special archives, and intangible cultural heritage protection.** Strictly speaking, these projects do not belong to ancient books digitization but fall within the category of historical documents, so they are counted together. This category includes 15 projects, such as “Research on the Value of Yunnan Minority Oral Documents and Construction of Their Digital Voice Database” (10CTQ015), “Investigation and Thematic Database Construction of Bai Nationality Oral Literature Intangible Cultural Heritage” (12CTQ018), “Integration and Research of Confucius Mansion Archives Documents Through Dynasties and Construction of Full-text Database” (13&ZD108), “Compilation, Collation and Digital Construction of Yi Affairs Chinese Memorial Archives in Qing Dynasty” (16BTQ045), “Research on Digital Protection and Inheritance of Endangered Oral Poetics—Qinghai Salar Nationality ‘Yu’er’ ” (17XTQ011), and “Research on Inheritance and Digital Protection of Endangered Intangible Cultural Heritage—Qinghai Tu Nationality ‘Daola’ ” (18BTQ035). Since 2016, the National Social Science Fund has significantly increased support for oral

history and intangible cultural heritage protection in minority regions.

- (3) **Thematic research or specialized ancient books digitization projects.** This category includes 14 projects. In terms of thematic research, some explore theoretical issues of ancient books digitization, such as “Research on the Value Realization of Academic Tradition of Literature Collation in Ancient Books Digitization” (17BTQ009) and “Research on Digital Humanities Development and Application Mode of Library Ancient Books Documents” (17XTQ003); some explore industrial development issues, such as “Research on Taiwan Region Digital Archives Industry Chain Construction” (17CTQ033); some focus on construction of digitization platforms for certain ancient books, such as “Research on Vietnam and South China Sea Place Name Verification and Historical Geography Information Platform Construction” (18BTQ048); and some explore specific technical issues, such as “Research on Mongolian Character Shape Standardization Issues for Digital Publishing” (17XTQ001), “Research on Automatic Construction Method of Chinese Traditional Culture Knowledge Expression System Based on Classics” (18BTQ063), and “Research on Huizhou Documents Classification and Metadata Standard Design” (12CTQ007).
- (4) **Digitization research on single ancient books.** This category includes 4 projects, such as “Research on Semanticization of Chu Ci Documents” (10BTQ031), “Research on Vocabulary Measurement Based on ‘Bon Religion Tripitaka · Tengyur’ ” (12CTQ020), “Research on Compilation and Digital Application of ‘Qin Dynasty Calligraphy Dictionary’ ” (12XTQ009), and “Research on Place Name Verification and Geographic Information System Platform Construction of ‘Reading History and Geographic Records’ ” (17BTQ038).

3. Overall Characteristics and Existing Problems in Ancient Books Digitization Research

3.1 Overall Characteristics

Based on analysis of journal papers and national social science fund projects in ancient books digitization over the past decade, we identify three major characteristics:

- (1) **Regenerative protection of ancient books remains the main field of current practice and research.** From the perspective of both papers and projects, research on specialized ancient books protection and library ancient books resource construction accounts for the vast majority. In other words, China’s current ancient books digitization still focuses primarily on protecting and inheriting the content and texts of ancient books, following the 思路 of regenerative protection. This is particularly evident in ancient books with strong ethnic, regional, and cultural characteris-

tics. Among national social science fund projects, the largest number are “Western region projects,” which, beyond policy factors, is mainly because the western region has more of these three types of characteristic ancient books requiring urgent protection. Technically speaking, such research may not be particularly innovative, but from the perspective of literature preservation and cultural inheritance, its significance is profound.

- (2) **Research fields continuously expand with increasingly evident disciplinary trends.** If we compare the entire research content of ancient books digitization to a pyramid, then regenerative protection of ancient books forms its base. Above this base, new research fields continuously extend, such as basic theories of ancient books digitization, ancient books digital resource construction, ancient books digital resource management, ancient books digitization policies, and ancient books digitization technologies. From theory to practice, technology to policy, and application to management, ancient books digitization research could constitute a new discipline. In fact, the disciplinary construction of ancient books digitization has attracted scholars’ attention. For example, Mao Jianjun explored the research scope of electronic philology from the new perspectives of traditional bibliography, version studies, and textual criticism [41]; Ge Huaidong preliminarily defined the research object and disciplinary nature of ancient books digitization and proposed a basic framework consisting of digital ancient books collation studies, Chinese information processing technology, and ancient books digitization management theory [42].
- (3) **Knowledge service as the core concept in technology and philosophy.** Regardless of its development, ancient books digitization essentially belongs to the category of ancient books collation. Traditional ancient books collation provides text and semantic services for humanities and social sciences research. In the digital age, ancient books collation has entered deeper knowledge levels. At the pyramid’s apex, ancient books digitization technology and philosophy guide its development direction. Over the past decade, the combination of digital technology and humanities concepts has spawned the idea of knowledge service for humanities and social sciences research, making ancient books digitization no longer limited to providing accurate, reliable, and understandable content texts, but providing knowledge products that can assist research.

3.2 Existing Problems

Beyond research papers and projects, specific digital products also constitute part of ancient books digitization research achievements. Therefore, beyond statistical analysis of academic papers and projects, we also investigated system functions of existing ancient books digitization achievements through web surveys and hands-on experience. Combining literature research with system function investigation, we summarize the following deficiencies:

- (1) **Neglect of user demand investigation and analysis.** Ancient books digitization products ultimately serve humanities and social sciences researchers and ancient books enthusiasts. Beyond using technical means to transcribe ancient books content and protect the books themselves, ancient books digitization more importantly serves academic research. To what extent current ancient books digitization achievements can meet the needs of traditional academic research is questionable, because information technology experts lack in-depth investigation of the characteristics of humanities research and its professional knowledge information needs.
- (2) **Failure to construct complete academic norms for ancient books digitization.** Ancient books digitization is not a simple conversion of ancient books storage media, but a reasonable extension of traditional ancient books collation in the digital environment. The concepts and methods of ancient books collation, formed through long historical accumulation, adapt to traditional academic research needs and have scientifically rational aspects, such as examination and selection of base texts, textual collation, and handling of paratexts. Therefore, ancient books digitization products must be achievements of ancient books collation and must conform to academic norms of ancient books collation. However, to date, no corresponding norms have been established in text processing, technical treatment, and management processes to guarantee the academic quality of ancient books digitization.
- (3) **Insufficient borrowing from academic traditions of ancient books collation, with disciplinary cooperation needing strengthening.** China's ancient books collation has formed many excellent academic ideas through long historical accumulation, such as Confucius's "transmission without creation" principle, Sima Qian's "comprehensive view" for literature compilation, Zhi Mindu's "hundred considerations converge, ten thousand streams return to the same source" for version origin examination, Yan Zhitui's "without reading all books under heaven, one cannot arbitrarily make corrections" for textual criticism, and Zhang Xuecheng's "distinguish academic fields, examine origins and development" for bibliography. These academic ideas adapt to ancient books collation and remain important references for today's ancient books digitization practice.
- (4) **Neglect of digitization research on existing ancient books collation achievements.** In the paper literature environment, predecessors have conducted extensive ancient books collation work, leaving rich achievements including numerous prefaces, bibliographies, category sequences, abstracts, and various reference works compiled, such as official position tables, chronologies, biographies, chronological biographies, academic records, and encyclopedias. These achievements originally served to assist traditional academic research, and their digitization would achieve multiplier effects in improving service efficiency. However, existing ancient books digitization products rarely embed such collation achievements, re-

sulting in wasted academic resources.

- (5) **Limited scope of case studies.** Current ancient books digitization research has entered the data collation stage, with ontology construction, semantic association, knowledge mining, and visualization technologies widely applied. However, in terms of applied ancient books scope, they are mainly limited to traditional Chinese medicine, agriculture, and genealogy ancient books, while other distinctive categories such as historical notes, various bibliographies, encyclopedias, and gazetteers remain rare, requiring expansion in future research. Additionally, from a practical perspective, there are still not many achievements in deep development of ancient books knowledge, particularly lacking open and shared ancient books digitization platforms integrating various functions.

4. User-Oriented Knowledge Service: Discussion on Future Trends in Ancient Books Digitization Research

As mentioned above, regenerative protection of ancient books forms the base of the ancient books digitization pyramid, reflecting its current reality, while ancient books digitization technology and philosophy constitute its apex, guiding future directions. After major libraries have completed digital text conversion (i.e., regenerative protection) of rare books, unique copies, and ordinary ancient books, as well as local characteristic minority nationality ancient books, deep processing of ancient books digital resources becomes the inevitable future research topic. From the perspective of current technology and philosophy for ancient books data processing, user-oriented knowledge service represents the future trend, driven internally by academic inheritance, user demand, and information technology development.

4.1 Knowledge Service Driven by Academic Inheritance

The most significant characteristic of Chinese culture lies in its inheritance. For ancient books, on one hand, they carry knowledge content that represents achievements of academic research through generations, possessing inheritability; on the other hand, ancient books themselves are products of collation and publication by scholars through generations, and the system of literature collation ideas and methods formed through long-term practice is also part of academic inheritance. For ancient books digitization, on one hand, it must preserve the knowledge content of ancient books and inherit achievements of academic research through generations; on the other hand, it must inherit and develop the excellent academic traditions of ancient literature collation.

From the logic of knowledge production, any academic research bears the task of knowledge production and academic innovation. In advancing knowledge production, researchers' first task is literature review, which refers to systematically examining and analyzing various documents related to a research question to understand the research status in that field [43]. In other words, it is a process

of systematically identifying, locating, examining, and summarizing literature related to our research [44]. Literature review needs to establish internal academic connections between existing literature and new knowledge production: first, by clarifying existing literature to establish an academic dialogue point for new knowledge production; second, by discovering new research space through literature review to establish an academic advancement point [44]. Beyond promoting knowledge production and academic innovation, another key aspect is clarifying academic origins and promoting academic inheritance. In traditional literature environments, ancient scholars systematically organized previous documents to summarize previous scholarship, forming conventions, such as Western Han scholars Liu Xiang and others who organized government collections by “categorizing chapters and summarizing main ideas” to compile the *Bielu* (Separate Records), and Liu Xin who classified books to compile the *Qilue* (Seven Summaries), thereby systematically summarizing scholarship since the pre-Qin period. Compilation of large-scale encyclopedias (such as the *Yongle Encyclopedia*) and comprehensive series (such as the *Siku Quanshu*) also represented systematic organization and integration of scholarship of an era.

In the digital environment, modern scholars’ work of converting storage media of ancient books texts that carry ancient wisdom, parsing semantics, organizing content, and mining knowledge is essentially consistent with ancient scholars’ systematic organization of previous documents, differing only in process and method. While collating ancient books text content, the academic traditions of ancient books collation are also inherited in ancient books digitization, such as preserving original character features and various marginalia, interlinear notes, and commentaries during text transcription, reflecting the “transmission without creation” principle; integrating different types of ancient books digital resources and various functions assisting humanities research reflects the “comprehensive view” concept; and achieving automatic text clustering through semantic analysis and association, even using knowledge mining methods to reveal rich knowledge and different disciplinary themes hidden in ancient books texts, clarifying various schools of thought, exploring academic origins, and inheriting the “distinguish academic fields, examine origins and development” idea.

4.2 Knowledge Service Driven by User Demand

Reviewing the entire process from the beginning to comprehensive development of ancient books digitization research, combined with research status over the past decade, we find that serving practical work, solving real problems, and meeting users’ knowledge information needs in academic research have always been core issues. From the perspective of major ancient books collection institutions—libraries—their service models have transformed from literature to information to knowledge. Ancient books digitization has similarly progressed from digital preservation of ancient books texts, to organization and processing of text information, to mining and reorganization of knowledge within texts. This progressive process also reflects the continuous development of user needs.

To verify the driving effect of user demand on knowledge service-oriented ancient books digitization, Xu Dan [45] divided ancient books digitization product functions into six modules: “basic information browsing,” “content retrieval and filtering,” “reading display,” “auxiliary tools,” “personalized settings,” and “knowledge value-added utilization,” comprising 33 specific function points. Through offline interviews and online questionnaires, she investigated humanities and social sciences researchers’ needs for ancient books digitization product functions and functions of ancient books digitization products used by 23 large public libraries and university libraries. Using Japanese professor Noriaki Kano’s KANO model, she evaluated the relationship between each module’s functions and user satisfaction. Results show that “basic information browsing,” “content retrieval and filtering,” and “reading display” belong to basic and expected needs, while “knowledge value-added utilization” belongs to attractive needs. “Knowledge value-added utilization” functions include online dictionaries, calendar conversion, word/character frequency statistics, social network analysis, text clustering, knowledge association, content visualization, and geographic space display. If ancient books digitization products can provide these functions, they will greatly enhance user satisfaction. Currently, ancient books digitization products have basically realized functions in “basic information browsing,” “content retrieval and filtering,” and “reading display.” With technological development and research deepening, “knowledge value-added utilization” functions are gradually transforming from attractive needs to expected or basic needs. Therefore, deep knowledge mining, processing, and visualization of data after ancient books text conversion to provide knowledge services will be the focus of ancient books digitization research.

4.3 Knowledge Service Driven by Information Technology Development

As a tool, information technology plays important functions in scientific cognition and research. It not only improves research efficiency but also broadens research ideas. The combination of digital technology and traditional academic research even brings methodological transformation, with digital humanities being the best example. Ancient books digitization is both research on ancient books collation and research on information technology application. Only by grasping the development direction of information technology can we recognize future trends in ancient books digitization research.

From the development history of information technology, ancient books digitization has experienced three stages: digital conversion of ancient books texts, organization and processing of ancient books content, and discovery and visualization of ancient books knowledge. The first stage uses technologies including text input, image scanning, and OCR recognition, involving technical issues such as simplified-traditional Chinese character conversion, variant character recognition, and character set standards. The second stage uses technologies including computer-based classification and indexing, bibliography compilation, keyword

indexing, and automatic punctuation, collation, annotation, translation, and compilation, mainly meeting users' needs for original ancient books content and semantic interpretation in specific themes or disciplines. At the current stage, with further development of information technology, ancient books digitization increasingly emphasizes knowledge discovery and visualization of ancient books. This knowledge processing service using massive ancient books text data as objects breaks boundaries between disciplines and between books, mainly employing methods such as text clustering, data mining, ontology construction, social network analysis, geographic information systems, and knowledge graphs.

Currently, information technology for knowledge service is mainly applied to traditional Chinese medicine and agriculture ancient books digitization, such as “Medication Rules and Cluster Analysis of Ancient Literature on Chest Bi-Syndrome and Heart Pain” [46], “Research on Characteristics of Headache TCM Disease Nature Syndrome Elements Based on Ancient Books Literature Mining” [47], “Research on the Origin and Evolution of Infertility Disease Names Based on Traditional Chinese Medicine Ancient Books Data Mining” [48], “Analysis of Acupuncture Prescription Rules for Stroke Disease in Ancient Books Based on Data Mining Technology” [49], “Research on Construction of Spleen Visceral Manifestation Theory Knowledge System Based on Ontology” [50], “Research on Semantic Retrieval Based on Traditional Chinese Medicine Orthopedics Ancient Books Ontology” [51], “Construction and Application of Agricultural Ancient Books Ontology” [52], and “Research on Construction and Retrieval Mechanism of Agricultural Ancient Books Ontology” [53]. Additionally, in gazetteers, geography, and historical books, there are also many achievements, such as “Visualization of Product Names and Aliases in Gazetteer Ancient Books—From the Perspective of Social Network Analysis Technology” [54], “Research on Chinese Ancient Books Geographic Information System Based on GIS” [55], and “Exploration of Large-scale Ancient Books Text Application in Chinese History Quantitative Research” [56]. Other fields have considerable space for knowledge service-oriented ancient books digitization research.

Currently, at the stage of expanding ancient books knowledge value, a new trend has emerged of transforming from ancient books digital resource databases to knowledge sharing platforms. Beyond integrating large amounts of ancient books digital resources and providing browsing and retrieval functions, these knowledge sharing platforms also establish corresponding knowledge communities providing knowledge sharing and thematic discussion functions. For example, “Jihe Network” created an online knowledge community called “Ancient Books Circle,” where users can freely ask questions, speak, form groups, discuss related issues, help solve problems, and share academic notes and research insights. The “Chinese Philosophy Books Electronicization Project (CTP)” provides not only original text databases, original photocopies, internal dictionaries, and advanced search functions, but also advanced functions such as display of commentaries through dynasties, contemporary research databases, metadata systems, word analysis tables, and similar paragraph data. It also provides discussion areas for asking questions and participating in discussions, and wiki

areas for jointly editing and correcting errors in ancient books full texts [52].

Academic research on ancient books digitization guides practical activities. Through statistical analysis of research fields and thematic distribution of academic papers and national social science fund projects related to ancient books digitization over the past decade, combined with investigation and hands-on experience of existing ancient books digitization product functions, we summarize the general characteristics and existing problems of ancient books digitization research in recent ten years. From three aspects—inheritance of academic traditions, satisfaction of user needs, and promotion of technological development—we identify user-oriented knowledge service as the future trend of ancient books digitization research. While performing ancient books cultural heritage protection and digital resource construction, ancient books digitization should, according to the characteristics of humanities and social sciences research and oriented toward professional researchers' knowledge needs, develop more knowledge tools to assist humanities and social sciences research.

References

- [1] Liu Weihong. Current Status and Significance of Chinese Ancient Books Digitization[J]. *Library and Information*, 2009(4): 134-137.
- [2] Zhu Suoling, Bao Ping. Review on the Progress and Research of Ancient Books Digitization in China[J]. *Library Theory and Practice*, 2009(9): 18-21.
- [3] Gao Juan, Liu Jiazhen. Problems and Countermeasures of Ancient Books Digitization in Mainland China[J]. *Journal of Library Science in China*, 2013(4): 110-119.
- [4] Shao Zhengkun. Dilemmas and Response Strategies of Ancient Books Digitization[J]. *Library Science Research*, 2014(12): 32-34.
- [5] Liu Jiazhen, Cheng Wangao. Strategies and Suggestions for Ancient Books Preservation and Development[J]. *Journal of Library Science in China*, 2009(3): 15-20.
- [6] Ding Kan, Liu Changhua. Fidelity Issues in Ancient Books Digitization[J]. *Journal of Traditional Chinese Medicine Literature*, 2009(2): 31-33.
- [7] Zhao Dong. Research on the Protection and Development of Historical and Cultural Resources under Digital Survival—Centered on Shaanxi[D]. Jinan: Shandong University, 2014.
- [8] Chen Cheng. Exploration of Long-term Digital Preservation Mechanisms for Ancient Books in Libraries in the Digital Humanities Era[J]. *Chinese Journal of Traditional Chinese Medicine Library and Information Science*, 2019(3): 6-9.
- [9] Mao Jianjun. Research on Standards and Evaluation of Ancient Books Bibliographic Databases[J]. *Library Theory and Practice*, 2009(6): 30-33.
- [10] Zhou Linjie. Discussion on Construction Standards and Norms of Ancient Books Bibliographic Databases in China[J]. *Library Construction*, 2010(2): 47-50.
- [11] Zhao Baoying, Su Mo. Exploration of Development Trends of Ancient Books Digital Resource Databases[J]. *Library Work and Research*, 2011(2): 36-39.
- [12] Liang Aimin, Chen Lijing. Ancient Books Digitization and Collaborative Construction and Sharing[J]. *Journal of the National Library of China*, 2012(5): 108-112.
- [13] Sa Lei. Research on Integration of Library Ancient Books Digital Resources[J]. *Library*, 2014(4): 116-119.
- [14] Chen Shibing. Research on Aggre-

gation of Ancient Books Resources[D]. Taiyuan: Shanxi University, 2018. [15] Huang Feilong, Zha Yilan. Construction of Chinese Mongolian Ancient Books Version Database Based on Unicode[J]. Journal of Inner Mongolia University for Nationalities, 2009(3): 160-161. [16] Ge Huaidong. On the Construction of Ancient Books Version Databases[J]. Information Research, 2012(7): 75-77. [17] An Huan, et al. Development Status of Traditional Chinese Medicine Ancient Books Knowledge Base and Construction of Ancient Books Health Preservation Knowledge Base[J]. Western Journal of Traditional Chinese Medicine, 2016(6): 158-159. [18] Li Bing. Research on Digital Collation Plans for Traditional Chinese Medicine Ancient Books[D]. Beijing: China Academy of Chinese Medical Sciences, 2009. [19] Ke Zhou. Research and Implementation of Digital Genealogy[D]. Wuhan: Central China Normal University, 2011. [20] Hu Yitao, Song Ye. Digitization and Practice of Transcribed Local Gazetteer Ancient Books[J]. Library Theory and Practice, 2014(8): 101-103. [21] Yuan Yuhong. Digitization of Stone Rubbings at the National Library[J]. Library Theory and Practice, 2014(5): 17-18. [22] Wang Kang, Kang Jian. Research on Digital Co-construction and Sharing Guarantee System for Mongolian Resources[J]. Library and Information Service, 2015, 59(12): 39-43. [23] Cao Xia, Chang Cunku, Pei Li. Construction and Platform Design of Traditional Chinese Medicine Ancient Books Digitization[J]. Chinese Journal of Medical Library and Information Science, 2016(3): 45-47. [24] Li Yu'an. Reflections on Current Stage Ancient Books Digitization[C]//Proceedings of the 2nd International Symposium on Chinese Ancient Books Digitization. Beijing: China Intercontinental Press, 2011. [25] Chen Deyuan. Construction of Coordinated Mechanisms for Ancient Books Digitization Work[J]. Journal of Library Science in China, 2009(5): 47-51. [26] Gao Lihong. On the National Macro Policy System for Ancient Books Digitization[J]. Journal of Zhongzhou University, 2010(4): 39-41. [27] Li Mingjie, Yu Youyou. Preliminary Exploration of the Main Components and Collaboration Mechanisms of Chinese Ancient Books Digitization[J]. Library and Information, 2010(1): 34-44. [28] Zhu Suoling. Application Research of Named Entity Recognition in Gazetteer Content Mining[D]. Nanjing: Nanjing Agricultural University, 2011. [29] Fan Jia. Connotation of "Digital Humanities" and Deep Development of Ancient Books Digitization[J]. Library Science Research, 2013(3): 29-32. [30] Chen Gang. "Digital Humanities" and Historical Geography Informatization Research[J]. Nanjing Social Sciences, 2014(3): 136-142. [31] Wang Daxue, Chen Xi, Yang Guanghui. Research on Chinese Ancient Books Geographic Information System Based on GIS[J]. Fudan Journal (Natural Science), 2016(6): 684-688. [32] Ouyang Jian. Large-scale Ancient Books Text Visualization Analysis and Mining for Digital Humanities Research[J]. Journal of Library Science in China, 2016(2): 66-80. [33] Zhao Siyuan. Digitization, Datafication, and Text Mining of Local Historical Documents: Taking the Chinese Local Historical Documents Database as an Example[J]. Qing History Research, 2016(4): 26-35. [34] Ma Chuangxin, Chen Xiaochi. Structured Knowledge Representation of Commentary Literature Based on Ontology and XML[J]. Library Journal, 2017(8): 62-68. [35] Xia Cuijuan. Research on Open Application of Chinese Historical Geographic Linked Data in Library Digital

Humanities Projects[J]. *Journal of Library Science in China*, 2017(2): 40-53. [36] Sun Haishu, Fu Yongchi, Zhang Huamin, et al. Exploration of Building Traditional Chinese Medicine Ancient Books Knowledge Base Based on Ontology[J]. *Journal of Medical Informatics*, 2011(3): 64-68. [37] Ding Kan. Research on Knowledge Representation of Traditional Chinese Medicine Ancient Books Prescriptions Based on Knowledge Units[D]. Beijing: China Academy of Chinese Medical Sciences, 2012. [38] Xu Wen. Research on Construction of Semantic Network of Knowledge Concepts in Warm Disease Ancient Books[D]. Beijing: China Academy of Chinese Medical Sciences, 2015. [39] Ding Kan, Zhang Lijun. Construction of Traditional Chinese Medicine Academic Inheritance Context Based on Ontology[J]. *Journal of Traditional Chinese Medicine Literature*, 2018(5): 32-35. [40] National Office for Philosophy and Social Sciences. National Social Science Fund Project Database[DB/OL]. [2019-04-19]. <http://fz.people.com.cn/skygb/> [41] Mao Jianjun. On the Establishment of Research Scope for Electronic Philology[J]. *Library Theory and Practice*, 2010(9): 46-48, 88. [42] Ge Huaidong. Disciplinary Construction of Ancient Books Digitization[J]. *Science and Technology Education Innovation*, 2012(1): 156-157. [43] Feng Xiaotian. On Literature Review in Social Research[J]. *Journal of Central China Normal University (Humanities and Social Sciences)*, 2010(4): 40-46. [44] Zhang Zhaoshu, Gao Yuanxin. Knowledge Production and Literature Review—From Technical Guidance to Meaning Guidance[J]. *Tianjin Social Sciences*, 2019(1): 48-53. [45] Xu Dan. Investigation on Functions of Ancient Books Digitization Products Based on KANO Model for Humanities Research Needs[D]. Wuhan: Wuhan University, 2019. [46] Lu Yizhu, Li Yuanbai, Mao Jingyuan, et al. Medication Rules and Cluster Analysis of Ancient Literature on Chest Bi-Syndrome and Heart Pain[J]. *Tianjin Journal of Traditional Chinese Medicine*, 2013(10): 629-631. [47] Lei Yujiao. Research on Characteristics of Headache TCM Disease Nature Syndrome Elements Based on Ancient Books Literature Mining[D]. Zhengzhou: Henan University of Chinese Medicine, 2018. [48] Xing Yitao, Zhang Mingqiang, Wang Dingguo, et al. Research on the Origin and Evolution of Infertility Disease Names Based on Traditional Chinese Medicine Ancient Books Data Mining[J]. *Henan Traditional Chinese Medicine*, 2018(4): 603-605. [49] Chen Chuyun, Li Lixia, Li Weilin, et al. Analysis of Acupuncture Prescription Rules for Stroke Disease in Ancient Books Based on Data Mining Technology[J]. *Shanghai Journal of Acupuncture and Moxibustion*, 2017(12): 1494-1499. [50] Cui Jiapeng, Wang Caixia, Yuan Dongchao, et al. Research on Construction of Spleen Visceral Manifestation Theory Knowledge System Based on Ontology[J]. *Chinese Archives of Traditional Chinese Medicine*, 2018(2): 388-392. [51] Lin Erzhen. Research on Semantic Retrieval Based on Traditional Chinese Medicine Orthopedics Ancient Books Ontology[D]. Fuzhou: Fujian University of Traditional Chinese Medicine, 2018. [52] Cao Ling, He Lin. Construction and Application of Agricultural Ancient Books Ontology[J]. *Journal of Guangxi Normal University (Natural Science Edition)*, 2007(2): 1-4. [53] He Lin, Cao Ling. Research on Construction and Retrieval Mechanism of Agricultural Ancient Books Ontology[J]. *New Technology of Library and Information Service*, 2006(12): 37-53. [54] Li Na, Bao Ping. Visualization of

Product Names and Aliases in Gazetteer Ancient Books—From the Perspective of Social Network Analysis Technology[J]. Library Tribune, 2017(12): 108-115.
[55] Wang Daxue. Research on Chinese Ancient Books Geographic Information System Based on GIS[J]. Fudan Journal (Natural Science), 2016(6): 684-697.
[56] Ouyang Jian. Exploration of Large-scale Ancient Books Text Application in Chinese History Quantitative Research[J]. Journal of Academic Libraries, 2016(3): 5-15.

Author Contributions:

Li Mingjie: Proposed the research topic, designed the paper structure, and summarized important viewpoints;

Zhang Xianke: Collected paper materials and wrote the initial draft;

Chen Mengshi: Completed co-occurrence mapping analysis and participated in initial draft writing.

Abstract:

[Purpose/significance] In order to explore the development trend of the research on the digitization of ancient books, it makes a systematic investigation on the research status of the digitization of ancient books in China. [Method/process] By using CiteSpace software, authors and keywords of the research on the digitization of ancient books in the past decade were analyzed. Also, it analyzed the main research fields and subject distribution of papers and national social science fund projects. Authors made an investigation and had an experience on the functions of the existing digital products of ancient books. [Result/conclusion] This paper summarizes 3 general characteristics and five main problems of the research of digitization of ancient books in recent ten years, and points out that the knowledge service-oriented digitization of ancient books is the inevitable trend in the future under the drives of academic inheritance, user demand and the development of information technology.

Keywords: digitization of ancient books; knowledge service; academic inheritance; user demand; technology development

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

Source: ChinaXiv — Machine translation. Verify with original.