

Research on the Construction of a Digital Memory Platform for Chinese Library Science Scholars Based on Knowledge Graphs: Postprint

Authors: Zhang Wenliang, Chen Chongyang, Li Xuan

Date: 2025-06-20T09:30:49+00:00

Abstract

The history of libraries is not only a history of socio-cultural development, but also a history of library scholars inheriting civilization and serving society. The construction of digital memory of library scholars, at the public level, helps enhance the general public's understanding of the development trajectory of China's library cause and important academic figures; at the professional level, it holds unique value in aspects such as the spiritual inheritance of library scholars, knowledge mining and discovery, and revealing the research fields and values of library scholars. This research takes "Pioneers of the Chinese Library Cause: Du Dingyou" as an example, focuses on library scholars as the core, and constructs an ontology model of Chinese library scholars from a spatio-temporal dimension. Combining automated methods such as named entity recognition, part-of-speech tagging, and regular expressions with manual indexing methods, entities and relationships were extracted from biographical texts, and a knowledge graph of Chinese library scholars was constructed through Neo4j. Finally, based on the ontology model of Chinese library scholars, a digital memory platform for Chinese library scholars was constructed, including six functions: personal profiles, work retrieval, chronology of major events, knowledge graph, spatio-temporal trajectory, and intelligent Q&A, which can provide strong support for the preservation and inheritance of the memory of Chinese library scholars.

Full Text

Preamble

Research on the Construction of a Digital Memory Platform for Chinese Librarians Based on Knowledge Graphs

Zhang Wenliang, Chen Chongyang, Li Xuan

School of Information Science and Technology, Northeast Normal University,

Changchun

Abstract: The history of libraries is not only a record of social and cultural development but also a testament to how librarians have preserved civilization and served society. Constructing a digital memory of librarians serves the public by enhancing understanding of the evolution of Chinese librarianship and its key academic figures. For the profession, it holds unique value in inheriting the spirit of librarianship, enabling knowledge mining and discovery, and revealing research fields and contributions of librarians.

This study takes *Pioneer of Chinese Library Career: Du Dingyou* as a case example, constructing a Chinese librarian ontology model with librarians at its core from spatial-temporal dimensions. Combining automated methods such as entity recognition, part-of-speech tagging, and regular expressions with manual annotation, entities and relationships were extracted from biographical texts to build a knowledge graph of Chinese librarians using Neo4j. Finally, based on the Chinese librarian ontology model, a digital memory platform for Chinese librarians was constructed, featuring six functions: character profiles, work retrieval, chronology of major events, knowledge graph, spatio-temporal trajectory, and intelligent Q&A. This platform provides robust support for the preservation and inheritance of Chinese librarian memory.

Keywords: Librarian; Knowledge graph; Thematic database; Digital memory; Digital humanities

CLC Number: G250.7

Document Code: A

1 Literature Review

Research on librarians has examined figures across different generations. Cheng Huanwen pioneered the “four generations of librarians” framework, dividing 20th-century Chinese librarians into four cohorts: the foundational first generation (including Shen Zurong, Hu Qingsheng, Li Xiaoyuan), the bridging second generation (including Qiu Kaiming, Wang Chongmin), the pioneering third generation (including Tong Zenggong, Guo Yixing), and the developing fourth generation (including Li Guoxin, Huo Guoqing). This classification has become widely accepted in library science circles. Early research primarily focused on thematic databases to preserve the thoughts and contributions of historical figures, helping to understand the relationship between these figures and their historical contexts. For instance, Zheng Lifan examined the first generation of Chinese librarians who studied in the United States during the Republican era. Fan Fan analyzed the group characteristics of Republican-era librarians, while Cui Ran reviewed the collective image of the first generation. Zheng Lifan also systematically investigated the background, American experiences, and contributions of the second generation of librarians who studied abroad.

With the application of digital technologies and the internet, scholars began using digital humanities techniques to digitize historical figures. Numerous historical figure databases emerged, such as the Historic Slave Trade Database and the China Biographical Database (CBDB). However, research on librarians in the digital humanities domain remains relatively weak. While librarians themselves have made significant contributions to digital humanities, their role as knowledge organization experts is often unrecognized by scholars in other fields. In constructing digital memory for librarians, librarians serve as both content experts and knowledge organization experts, while also undertaking some technical responsibilities. This process can effectively reveal librarians' research fields and knowledge organization value, helping more scholars recognize the importance of librarians in digital humanities research.

Ontology research for historical figures has progressed through three stages: basic information ontology, relationship ontology, and event ontology. Basic information ontologies like FOAF describe fundamental attributes such as name and birthdate. Relationship ontologies define connections between similar figures, such as the British King's College Digital Humanities Department's study of Syrian biographies. Event ontologies, centered on events with time and location dimensions, describe dynamic life events. Shen Xueying et al. constructed an ontology model for ancient Chinese literati based on life events. Si Li et al. built a biographical ontology model for librarian Peng Feizhang, which included 11 classes, 24 data properties, and 11 object properties. These models provide valuable references for constructing librarian ontologies.

2 Value of Constructing Digital Memory for Chinese Librarians

The history of libraries is not only a record of social and cultural development but also a history of librarians inheriting civilization and serving society. The establishment and development of library undertakings in China have depended on the diligent work and dedicated contributions of generations of librarians. Constructing digital memory for librarians holds significant value for the public, the profession, and library science education.

For the public, digital memory helps break down misconceptions about library work and enhances understanding of libraries' value and role. For the library community, it facilitates the inheritance of professional spirit and enables knowledge mining and discovery. The knowledge mining value manifests in two aspects: mining and discovering knowledge related to librarians themselves, and mining and discovering library science knowledge through librarian biographies. Based on entity relationships in the knowledge graph, knowledge reasoning can be performed to complete missing relationships and discover latent knowledge, thereby revealing hidden connections between different pieces of knowledge about librarians. For library science students, digital memory can vividly and accessibly help them understand the development of Chinese library science, fostering a sense of mission, honor, and enthusiasm for the profession.

3 Construction and Visualization of Chinese Librarian Ontology Model

Continuous biographical materials for librarians are scarce, making research on individual librarians challenging. This study takes Mr. Du Dingyou, a representative figure in Chinese library science, as an example. Only two biographical works exist: *Du Dingyou and Chinese Library Science* by Wang Zizhou and *Pioneer of Chinese Library Career* by Yang Hengping, both primarily based on collected works or internal materials analyzing Du's academic thoughts and achievements with limited attention to his life details.

From spatial-temporal dimensions and with librarians at the core, this study constructs a Chinese librarian ontology model. The model reuses existing ontologies to enhance semantic interoperability and development efficiency, primarily referencing the Chinese Literati Life Ontology Model and the Biographical Materials Ontology Model. The core classes are determined as: Person (including librarians and related figures), Event (including creative events, work events, educational events, political events, life events, and migration events), Organization, Work, Time, Place, and Period.

The Person class describes librarians and related figures. The Event class is the core, with creative events referring to writing and publishing works; work events including employment at libraries or universities; educational events covering both learning and teaching; political events involving government departments; life events describing daily interactions and major life events; and migration events referring to changes in residence. The Organization class includes institutions involved in librarians' experiences. The Work class focuses on librarians' academic writings to support analysis of their thoughts. The Time and Place classes describe temporal and spatial characteristics. The Period class divides librarians' lives into stages to reveal evolution in their academic thought.

Based on the reused ontologies and research requirements, the study designed 16 data properties and 14 object properties. Data properties include event specific time and location, while object properties describe relationships between persons (social and kinship), between persons and organizations (study and employment relationships), between persons and works (creation relationships), and between persons and events (participation relationships).

Using this ontology model, the study processed relevant corpora through a combination of automated methods and manual annotation. Python's jieba library performed word segmentation using Harbin Institute of Technology's stopword list. A user dictionary was constructed by calculating co-occurrence frequencies of bigrams and combining pairs exceeding 1% frequency. The NLPPIR-Parser tool provided entity recognition and part-of-speech tagging, supplemented by manual annotation. Rule-based methods extracted entities and relationships, such as extracting content within book title marks and time-expressing phrases like "at the age of X." Event instances were manually annotated to ensure continuity and completeness of biographical information.

4 Construction of Chinese Librarian Digital Memory Platform

The platform aims to integrate materials from various generations of librarians, construct a knowledge graph, and achieve visualization of librarian knowledge to support preservation and inheritance. Platform requirements include knowledge visualization (character profiles, knowledge graph, spatio-temporal trajectory) and knowledge interaction (intelligent Q&A).

The character profile module presents librarians' photos and biographical sketches to enhance understanding of basic information and major achievements. The work retrieval module displays various works (monographs, journal articles, innovative physical objects) with free search functionality. The chronology module implements temporal association analysis through a thematic river and major events timeline. The thematic river reveals active periods and focus areas, while the timeline contextualizes librarians' lives within broader historical backgrounds. For Du Dingyou, the thematic river shows his difficult early education (1898-1921), productive work period after returning from library science studies in the Philippines (1922-1936), decreased creative output during the Japanese invasion when he focused on protecting Sun Yat-sen University library collections (1937-1949), and resumed creative activity during the stable post-1949 period.

The spatio-temporal trajectory module enables spatial association analysis by examining geographical distribution of events and travel routes. Guangdong Province was crucial in Du's career, providing a base for his library practice and document preservation efforts during the war. The trajectory vividly illustrates his first book relocation journey to avoid conflict, traveling by boat through Sanshui, Zhaoqing, and Nanjiangkou to reach Sun Yat-sen University's new campus in Luoding.

The knowledge graph module supports relationship association analysis by integrating dispersed knowledge into an interconnected network. It reveals relationships between persons, persons and organizations, and persons and events. For Du Dingyou's study period in the Philippines (1918-1921), the knowledge graph shows his study at the University of the Philippines, where he earned bachelor's degrees in literature, library science, and education, along with a secondary school teaching certificate, laying a solid foundation for his lifelong dedication to librarianship.

The intelligent Q&A module parses user questions and selects appropriate response methods using the knowledge graph as a question-answering knowledge base, effectively covering diverse user inquiries about librarians and library work.

5 Recommendations for Chinese Librarian Digital Memory Platform Construction

Strengthen Continuous Archival Materials Construction: Continuous materials cover librarians' entire lives from birth to death and include as many important deeds as possible. Current research relies heavily on fragmentary materials, with few comprehensive studies. Construction requires systematic collection of published materials, including other scholars' writings about librarians and librarians' own works. Non-public materials should be obtained through contacts with institutions where librarians worked and through oral histories from students and descendants. These diverse sources should be integrated to form continuous biographical records.

Integrate Multimodal Resources: Multimodal resources include text, images, audio, and video. Integration establishes connections between different modalities to aggregate resources about the same entity, providing users with multiple perspectives. For instance, *Pioneer of Chinese Library Career* provides only vague descriptions of Sun Yat-sen University's relocation to Yunnan, while video materials offer more intuitive and detailed accounts that can fill such gaps. As digital technology advances, platforms must increasingly support multimodal resource integration.

Emphasize Digital Technology Application: Digital technology is the core driver for platform development and continuous improvement. Natural language processing and knowledge organization technologies should be used to organize materials, extract entities and relationships, and improve knowledge graph construction efficiency. Large language models can support intelligent Q&A functions to meet personalized user needs and evoke emotional resonance with librarians' life stories.

6 Conclusion and Outlook

Taking Mr. Du Dingyou, a well-documented representative figure, as a case study, this research constructed a Chinese librarian ontology model from spatial-temporal dimensions with librarians as the core. The model includes 16 data properties and 14 object properties. Combining automated methods (entity recognition, part-of-speech tagging, regular expressions) with manual annotation, entities and relationships were extracted from biographical texts to build a Chinese librarian knowledge graph using Neo4j. Based on this knowledge graph, a digital memory platform for Chinese librarians was constructed, featuring six functional modules that provide robust support for constructing and disseminating digital memory of Chinese librarians.

Limitations include reliance on a single biographical work (*Pioneer of Chinese Library Career*) and insufficient research depth. The current platform still falls short of the envisioned goal. Future work should systematically collect and analyze more multimodal materials, including oral histories, letters, diaries, and

audio-visual resources from multiple librarians. More advanced natural language processing and knowledge graph technologies should be employed to refine the ontology model and incorporate evaluations of librarians by others. Additionally, VR and AR technologies could provide immersive interactive experiences to deepen user engagement with librarians' life stories.

Abstract

The history of libraries records both social-cultural development and librarians' contributions to preserving civilization and serving society. This study constructs a digital memory platform for Chinese librarians based on knowledge graphs. Taking *Pioneer of Chinese Library Career: Du Dingyou* as a case study, we built a Chinese librarian ontology model from spatial-temporal dimensions with librarians at the core. Combining automated methods (entity recognition, part-of-speech tagging, regular expressions) with manual annotation, we extracted entities and relationships from biographical texts to construct a knowledge graph using Neo4j. The resulting digital memory platform includes six functions: character profiles, work retrieval, chronology, knowledge graph, spatio-temporal trajectory, and intelligent Q&A, providing robust support for preserving and inheriting the legacy of Chinese librarians.

Keywords: Librarians; Knowledge graph; Thematic database; Digital memory; Digital humanities

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

Source: ChinaXiv –Machine translation. Verify with original.