Exploring the Service Model of Digital Archives Based on Knowledge Graphs (Postprint)

Authors: Xiong Huixiang, Wuyue Yan

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

Abstract

[Purpose/Significance] In response to current limitations such as insufficient intelligence in digital archival service quality and monotonous service content, this study proposes the construction of a comprehensive architecture for digital archival knowledge graphs, aiming to achieve statistical analysis of digital archival big data, integration and optimization of digital archival resources, and enhancement of overall digital archival service standards. [Method/Process] Through literature collection and analysis of service deficiencies in digital archives, experiential investigation of web service levels of digital archives across various provinces and municipalities, and compilation of shortcomings in current service models of different archives, this study refines the architectural process of digital archival knowledge graphs, ultimately presented in flowchart form. [Results/Conclusion] Knowledge graphs can transform diverse types of digital archival texts into computer-comprehensible data, thereby enhancing intelligent computer recognition capabilities. Furthermore, the dynamic temporality and specificity inherent in knowledge graphs can improve archival integration according to temporal nodes, while updating user data based on user browsing and retrieval traces to elevate service quality. Additionally, strengthening cooperation and exchange among digital archives achieves resource integration effects, providing valuable references for optimizing digital archival services.

Full Text

Preamble

Title: Research on Digital Archives Service Mode Based on Knowledge Graph
Authors: Xiong Huixiang, Yan Wuyue
Affiliation: School of Information Management, Central China Normal University, Wuhan 430079

Abstract:
[Purpose/Significance] Aiming to address the shortcomings of current digital archives services, such as insufficient intelligence and limited service content, this paper proposes constructing a comprehensive knowledge graph architecture for digital archives to achieve statistical analysis of digital archive big data, integration and optimization of digital archive resources, and overall improvement of digital archive service quality. [Method/Process] Through literature review, we analyzed service deficiencies in digital archives, conducted experiential investigations of web service levels across provincial and municipal digital archives, summarized limitations of existing service models, and refined the architectural process of digital archive knowledge graphs, ultimately presenting the framework in flowchart form. [Result/Conclusion] Knowledge graphs can transform multiple types of digital archive texts into computer-understandable data, enhancing intelligent recognition capabilities. The dynamic temporality and targeted nature of knowledge graphs can improve archive integration according to time nodes, update user data based on browsing and retrieval traces to enhance service quality, and strengthen cooperation and communication between digital archives to achieve resource integration, providing valuable references for optimizing digital archive services.

Keywords: knowledge graph; digital archives; archive intelligence service

Funding: This work is supported by the National Social Science Foundation Annual Project "Research on Online Academic Resource Mining and Recommendation Integrating Knowledge Graphs and Deep Learning" (Project No.: 19BTQ005).

Author Biographies: Xiong Huixiang (ORCID: 0000-0001-9956-3396), Professor, Ph.D., Doctoral Supervisor; Yan Wuyue (ORCID: 0000-0003-1652-7926), Master's Student, Corresponding Author, E-mail: 348101761@qq.com.

Received: May 13, 2021
Published: August 19, 2021

---

1 Introduction

The outstanding performance of semantic technologies has demonstrated their convenience and irreplaceable advantages across many domains, prompting the archives community to envision new possibilities for the future development of digital archives. Simultaneously, with the evolution of the Internet, interactions between people and between archives and users have become increasingly frequent, making the integration of archival services with semantic networks inevitable. Currently, the semantic organization of digital archive resources remains in its preliminary stages, and how to semantically organize these resources to provide more precise services has become an urgent practical challenge for archival service departments. Addressing the fragmentation, complexity, and insufficient intelligence of current digital archive services, this paper proposes integrating knowledge graph technology with digital archive services. As an emerging semantic processing model, knowledge graphs can connect entities, mine and display inter-entity relationships, extract core terms from different archives, integrate archival knowledge, promote the convergence of digital archive resources, and enhance user experience. This enables users to input a single archival knowledge point and obtain other related archival knowledge through the knowledge graph, with the graph automatically expanding relevant content, reducing retrieval steps, improving retrieval efficiency, and enriching retrieval results.

2 Related Research

A knowledge graph is a semantic network depicting relationships between entities and represents a primary manifestation of knowledge engineering—an important research field in artificial intelligence. Currently accessible knowledge base resources include foreign platforms such as Freebase, Wikidata, DBpedia, and YAGO, as well as domestic resources like the Chinese Conceptual Graph CN-Probase published by Fudan University. The knowledge graph architecture constructed in this paper is not a generalized universal architecture but a domain-specific knowledge graph for archives. Unlike general knowledge graphs, domain-specific knowledge graphs can rapidly construct knowledge bases using domain-specific knowledge, as seen in medical knowledge graphs, geographic knowledge graphs, military knowledge graphs, and agricultural knowledge graphs. Current research on knowledge graphs primarily focuses on exploring metadata semantic interoperability and mapping relationships based on existing metadata standards (EAD, Dublin Core, etc.). For instance, Yang Qianya introduced semantic ontology concepts into enterprise archival data applications to implement archival data semantic analysis processes and built a connected electronic archives knowledge graph system. Lei Jie et al. constructed a computer-understandable semantic model for scientific research archives knowledge graphs based on Protégé and OWL technologies. Shu Zhongmei proposed an archival spatio-temporal ontology model and archival data extraction framework under the current digital humanities development background, constructing archival linked data knowledge graphs displayed in visual form. B. S. Balaji et al. employed semantic parsing for document description and constructed a cloud service recommendation system.

In digital archive services, users have traditionally been familiar with passive service models involving passive retrieval and viewing, resulting in isolated relationships between archival files and making access difficult and operations cumbersome. Currently, most domestic scholars address this problem by transforming service models and constructing resource platforms. For example, Cao Ling et al. studied the digital archive information service models of eight Ivy League universities in the United States, analyzing them across three modules—operation mode, service objects, and service methods—to propose optimization recommendations for China's archival services. Wang Wenqiang analyzed the advantages and disadvantages of smart services and digital archives to transform enterprise archival information service models from "information services" to "knowledge services" and "smart services." Lian Zhiying classified digital archive information users and constructed user models based on analysis of user needs and behaviors, building digital archive information resources and service platforms to provide corresponding personalized digital archive information services. Internationally, research primarily combines digital archives with cultural heritage preservation and historical applications, while also exploring digital archive construction. For instance, T. Hauswedell et al. examined the complex interplay of institutional, knowledge, economic, technical, practical, and social factors through a series of semi-structured interviews with public and private providers of major newspaper digitization programs, arguing that emerging understandings of rarely highlighted factors fundamentally shape the depth and scope of digital cultural heritage archives and should be considered for their potential benefits in future archival development. C. H. Marcondes analyzed the feasibility and challenges of using linked data technologies in archives, libraries, and museums.

In summary, current digital archive services largely remain focused on constructing service content and frameworks, concentrating on discussions of resources, data, and service models, with limited practical research. It is evident that systematic management and intelligent service provision in provincial and municipal digital archives remain incomplete. As China rapidly advances into the smart era, scholars have fully recognized that semantic network technologies significantly impact digital archive service efficiency and quality. The integration and systematization offered by knowledge graphs can effectively address the fragmentation of digital archives and provide targeted, comprehensive services through organized recommendations. Furthermore, considering the diverse roles of digital archive services across industries, this paper examines the unified characteristics of digital archives, connecting digital archive entities through semantic associations and employing visual charts to display inter-entity relationships, thereby constructing knowledge graphs to advance digital archive service model upgrades.

3.1 Limitations of Existing Service Models

Archive services refer to the process by which archival institutions leverage their collection advantages to guide users in utilizing archives and obtaining archival information. Two key measures of service model quality are the service quality of archival personnel and the intelligence level of archival institutions. Regarding personnel quality, traditional archive service models are predominantly passive. Under conditions of low overall informatization, most enterprise or government archival staff have unclear understanding of archival services, with their functions limited to managing and preserving paper archives and maintaining information security. More practically significant tasks such as providing archival retrieval services are not within their responsibilities, leaving users to solve the problem of finding needed archives from massive databases themselves.

Regarding institutional intelligence, the rapid development of "Internet + Big Data" in the new era has propelled archive services toward active and intelligent models. Service models that connect computer equipment with various Internet terminals to provide diverse archival information or products through technical means have gradually become mainstream, with digital archives and smart archives being the most representative. However, current digital archive construction remains in its infancy, offering extremely limited personalized services, low intelligence levels, and complex operations that make integrated service delivery difficult. Based on experiential investigations of existing provincial and municipal archive websites, we found that while most provide open archive query services, they fundamentally rely on autonomous user queries with high requirements for query keywords and numerous redirect links, making the query process complex. For example, when querying an archive on the official website of Hubei Provincial Archives, the system prompts users to visit the archival information network for queries without providing corresponding links. When users find the archival information network link themselves, clicking the results directly redirects back to the Hubei Provincial Archives homepage. These examples demonstrate that current archive websites suffer from limited service levels, low integration, and operational complexity.

3.2 Feasibility of Knowledge Graph-based Digital Archive Service Model

With the development of semantic networks and neural networks, related technologies such as knowledge graphs, graph databases, and deep learning have gradually penetrated knowledge management research, making archival data semantic description and association, knowledge fusion, and information visualization possible. Meanwhile, driven by smart city construction, archive services have begun emphasizing user needs and shifting service models toward user-centered approaches. However, digital archives and smart archives still exhibit deficiencies in processing archival data and mobilizing archival resources. Knowledge graph-based digital archive services can accurately extract the core of digital archive resources through semantic association and ontology refinement, presenting them to users through concise and clear visualizations.

First, archival data quality is relatively high. Compared to massive and complex Internet big data, archival data represents authentic event records and data preservation that have already undergone screening, classification, and organization during accessioning, with corresponding standardized formats and requirements for different digital archives, providing a solid foundation for basic archival data processing. Second, the systematic nature of knowledge graphs can effectively help organize fragmented archives. Using tools such as graph mapping and wrappers, digital archive structures can be cleaned, transformed, and integrated based on deep processing needs, turning them into computer-understandable nodes. On this basis, entities are identified, connected, and analyzed to establish semantic associations between archival entities, forming a digital archive knowledge graph. Finally, the output mode of knowledge graphs can significantly enhance user experience. By visually presenting relevant archives retrieved by users, knowledge graphs can more clearly and conveniently display connections between archives, while their dynamic update characteristics can timely modify user preferences, making services more intelligent.

4 Knowledge Graph-based Digital Archive Service Model Architecture

Knowledge graph construction typically follows two patterns: top-down and bottom-up. Based on the strong domain specificity of archives, this paper adopts a bottom-up construction approach, which involves iterative updates from information extraction to knowledge fusion, knowledge processing, and finally knowledge updating, following a pattern from data analysis to service output. The ultimate goal of integrating knowledge graphs with digital archive service models is to optimize service delivery. Therefore, based on knowledge graph theory, ontology theory, intelligent agent technology, cloud computing technology, and personalized recommendation technology, the digital archive knowledge graph service platform is structured into four layers: (1) Knowledge Extraction Layer, which primarily collects and structurally classifies external digital archive information resources in different formats, transforming them into computer-recognizable and understandable data and standardizing formats for entry into the graph construction layer; (2) Knowledge Graph Construction Layer, which performs archival knowledge extraction and ontology construction, then forms a digital archive resource knowledge base through entity recognition and ontology alignment; (3) Knowledge Graph Update Layer, which maintains graph real-time performance by collecting new archives and user browsing/query traces, thereby improving service precision and graph adaptability; and (4) Knowledge Graph Application Layer, where the complete knowledge graph construction process integrates and organizes different digital archive resources from digital archives, forming an archival knowledge graph that outputs to the application level. Through processing generated data from digital archive information sources, the structured association data produced is used for deep learning algorithm training to obtain judgment models that can solve specific scenario problems, thereby generating value through service forms including dynamic graph updates based on association rule algorithms, personalized user preference recommendations based on clustering algorithms, intelligent digital archive search based on classification and prediction algorithms, and digital archive resource sharing based on overall optimization. The specific architecture is shown in Figure 1 [Figure 1: see original paper].

4.1 Knowledge Extraction Layer

The knowledge extraction layer primarily includes data integration and information extraction. The data integration component organizes and selects collected digital archive data from different sources, including data gathered from internal business systems across various industries, compiled from archival domain encyclopedias, and received from external business systems, encompassing digital archive data in different forms and structures. Archival data includes records with archival characteristics collected by various archival institutions at all levels, including archival resources in various data formats such as digital and multimedia archives, as well as data generated during archival management and utilization processes like archival website browsing records, platform logs, query service data, and archival statistics. Daily office documents such as Word, PPT, Excel, and PDF files constitute unstructured digital archive resources. To address the current inconsistency in digital archive resource structuring, we propose using natural language processing (NLP) and wrapper tools to transform these unstructured resources. This involves unified adjustment, cleaning, and correction of resources, including using dictionary-based, statistical, and rule-based methods for word segmentation of archival data, followed by cleaning and filtering based on segmentation results such as removing stop words, eliminating single characters, and performing part-of-speech tagging. Additionally, models like Word2vec can reduce archival word representations from one-hot encoder forms to shorter word vectors through word embedding, enabling computers to better understand and analyze data and discover semantic relationships between entities. The information extraction component includes entity extraction, relation extraction, and attribute extraction. Archival entity extraction employs rule-based and dictionary-based methods, statistical machine learning, and open-domain approaches to identify and extract entities from digital archive resources. Archival relation extraction uses supervised learning or distant supervision learning to extract relationships between entities, resolving semantic linking issues among archival resource entities. Archival attribute extraction involves extracting certain features and properties of archival resource entities, which can also be viewed as a nominal relationship between entities and attribute values. The specific process is shown in Figure 2 [Figure 2: see original paper].

4.2 Knowledge Graph Construction Layer

The knowledge graph construction layer is divided into knowledge fusion and knowledge processing. After processing large amounts of unstructured archival information, we can obtain information about entities, relations, and attributes, but these results may contain substantial errors and redundant information with unclear relationships, lacking hierarchy and logic. Therefore, knowledge fusion is required for cleaning and integration. Knowledge fusion includes entity linking, which connects extracted entities to knowledge bases, primarily through entity disambiguation and coreference resolution. Entity disambiguation resolves ambiguity when entities share the same name, primarily using clustering methods. Coreference resolution addresses cases where multiple references correspond to one entity object, with relatively mature foreign research and models such as entity similarity models and context similarity models effectively solving this problem. Knowledge merging incorporates third-party data into the database, including merging external knowledge bases and relational databases.

Through knowledge fusion, we obtain a series of basic factual expressions, but facts do not equal knowledge, necessitating the knowledge processing stage. Knowledge processing primarily includes three components: ontology construction, knowledge reasoning, and quality assessment. After knowledge fusion, identified entities become standardized knowledge with associated attribute relationships and are incorporated into the knowledge base. Due to current technological limitations, knowledge elements extracted through information extraction may still contain errors, requiring quality assessment before constructing a complete archival knowledge graph. By quantifying knowledge credibility and discarding low-confidence knowledge, we ensure data quality within the archival knowledge graph. Simultaneously, starting from existing entity relationships, computers can discover new entity relationships through reasoning, further improving and updating the archival knowledge graph network. Digital archive resource ontologies represent high-quality knowledge representation models that are organized. In knowledge graph construction, ontologies primarily serve to control graph quality, not necessarily participating in the graph establishment process but acting as a reference resource library for data quality assessment, playing an auxiliary role in archival knowledge graph construction while helping optimize knowledge graph updates.

4.3 Knowledge Graph Update Layer

The knowledge graph update layer includes user needs updating and archival resource updating modules, maintaining graph real-time performance by collecting and summarizing new archives and user web browsing/query traces, thereby improving service precision and enhancing graph adaptability. In terms of user needs updating, different user behavior data automatically drives ontology construction for archival resources through data-driven approaches, which are then revised and confirmed through quality assessment methods combined with manual review. After ontology construction, the knowledge base continuously revises data transactions and feedback while collecting user traces of knowledge base usage for continuous adjustment and updates. By incorporating temporal dimensions and using time-series analysis and graph similarity techniques to analyze graph structure changes and trends over time, we can capture critical information and construct dynamic temporal knowledge graphs. In terms of archival resource updating, knowledge reasoning components in knowledge graphs are employed, such as Tableaux operations that can check ontology satisfiability while testing ontologies through instances, logic programming rewriting that can customize rules for specific scenarios to implement user-defined reasoning processes, first-order query rewriting that can efficiently combine data sources in different formats while associating different query languages, and production rules that can control system execution for better forward reasoning through established mechanisms.

4.4 Knowledge Graph Application Layer

The knowledge graph application layer refers to the complete knowledge graph construction process that integrates and organizes different digital archive resources from digital archives, forming an archival knowledge graph that outputs to the application level. By processing generated data from digital archive information sources, the structured association data produced is used for deep learning algorithm training to obtain judgment models that can solve specific scenario problems, thereby generating value through service forms including dynamic graph updates based on association rule algorithms, personalized user preference recommendations based on clustering algorithms, intelligent digital archive search based on classification and prediction algorithms, and digital archive resource sharing based on overall optimization.

5 Advantages and Applications of Knowledge Graph-based Digital Archive Service Model

The construction of the knowledge graph-based digital archive service model architecture is consistently based on providing users with more effective, convenient, and intelligent services. Through data organization of digital archives and integration with knowledge graph construction processes using key technologies such as natural language processing, entity recognition, ontology construction, and relation extraction, the digital archive knowledge graph architecture can support intelligent management of digital archives, providing dynamic archival graph intelligent updates, automatic analysis of user preferences, intelligent search based on digital archive content, and digital archive resource sharing.

5.1 Dynamic Graph Updates

As Internet technology continues to develop, archival services have gradually achieved digitization, but current digital archive services remain superficial, focusing primarily on storage and management issues with limited discussion on optimizing service content. The construction of knowledge graphs based on digital archives can facilitate more convenient decomposition and classification of archival content through associations between digital archive ontologies, and through graph relationship 梳理, enable relevant archival content to generate connections that can be completely presented to users when they retrieve one item. Unlike digital library management, digital archive management must consider privacy, requiring measurement of privacy and publicity when new digital archives are accessioned, resulting in more complicated procedures. Knowledge graph construction based on semantic associations can enable automatic classification by identifying ontologies and creating links when digital archive texts are accessioned. Dynamic updates mean that graph construction is not static but continuously changes and improves with additions and deletions of archives, which can be enhanced by incorporating dimensions such as time and space to construct temporal or spatial axes that enrich graph content across different contexts. Dynamic graph updates can effectively improve archival organization efficiency, facilitate regulation of different archive types, and provide users with the latest content. Currently, engineering archives represent a key focus of digital archive collection and management, with projects like the Hong Kong-Zhuhai-Macao Bridge featuring long durations, large volumes, and multiple departments. The dynamic update characteristics of knowledge graphs can effectively meet engineering archival accessioning needs through continuous recording, summarization, and automatic classification, providing more convenient services.

5.2 User Preference-based Personalized Recommendations

The "National Archives Development '13th Five-Year' Plan Outline" explicitly states the need to "improve public archival service capabilities and enhance the recognition and user satisfaction of archival services." Faced with current complex and disordered archival resources, users must not only refine retrieval statements themselves but also handle complicated retrieval procedures, increasing retrieval time and reducing user interest, thereby diminishing retrieval demand. Meanwhile, when providing personalized services through knowledge graphs, user needs change in real-time with user behavior and browsed archival content. The variability and diversity of user needs, combined with the diversity of digital archive content and types, increase the difficulty of predicting user needs for personalized service functions. Therefore, inferring user archival preferences from dynamic behavior feedback is crucial for improving digital archive services. Based on prior knowledge processing transformations of archival data, natural language processing and machine learning methods are used for concept extraction of entities in archival resources, determination of classes and their hierarchical systems, determination of object properties and data properties of classes, and ontology evaluation, selecting multiple deep learning models such as BERT and LSTM. The prediction system can measure user preferences not only based on page dwell time and browsing speed but also identify user natural language through standardized processing into computer-recognizable language, enabling computers to better identify user preferences and increasing the accuracy of digital archive services. While predicting user needs, the system can pave the way for future services based on collected user behavior. For example, if users are interested in current news and archival anecdotes, the knowledge graph will push real-time anecdotes and related content after identifying keywords through semantic associations. If users are interested in historical archives and domain expert research, corresponding academic resources will be pushed.

5.3 Intelligent Search for Digital Archives

When using traditional archival retrieval, users must input accurate archival keywords to search titles or content to retrieve relevant information. This service model requires high precision in retrieval language, sometimes resulting in situations where users cannot find desired archival resources due to inability to accurately express keywords. The knowledge graph-based digital archive service model can automatically transform natural language used in user searches into computer-understandable language through natural language processing (NLP) technology, while mapping to different entity or attribute levels in the digital archive knowledge graph. By combining inter-entity relationships, corresponding digital archival information resources are pushed to provide systematic archival knowledge for user browsing. When users search with natural language, precise feedback of corresponding content improves both precision and recall rates while providing content association services that uniformly and coherently display related content, reducing user query frequency and steps and improving retrieval efficiency. "Archival latent needs are needs that have not been awakened or recognized, mainly including the needs of archival latent users and the unexpressed real needs of archival actual users." This intelligent content search can help digital archives deepen connections with users, attract users through innovative services of integrated content pushing, and help archives gain more users with excavated latent needs, thereby improving service quality.

5.4 Digital Archive Resource Sharing

Knowledge graph-based digital archive services aggregate resources from different digital archives, integrating them through cloud storage and cloud computing technologies to establish digital archive knowledge graphs that facilitate mutual exchange between resources, serving users conveniently while building bridges for communication between major archives. To meet the current environment of interconnected information resources, archival resource interaction has become particularly important. The knowledge graph-based digital archive service model can collect resources from different digital archives, establish semantic connections through ontology identification, and automatically categorize them into different industries based on archival domain keywords for convenient search and utilization. The assistance of knowledge graph models makes resource sharing between different digital archives highly feasible. In terms of user services, resource integration can bring richer archival content, achieving the goal of reducing retrieval steps while enriching retrieval content. Although much information can be retrieved accurately through the Internet, archival retrieval is limited by regional resources due to the privacy and individuality of archives. The convenience brought by knowledge graph resource sharing can effectively solve regional information silos and deepen innovative cooperation between archives through integration, jointly improving service quality, enabling archival users to truly understand the rich core of the archival world, expanding the archival audience, and raising public awareness and utilization value of archives. Meanwhile, digital archives can also establish strategic alliances with different institutions according to specific needs when conducting resource sharing services. When selecting partners, archives can choose digital archival institutions with expertise in the same field to strengthen collection volume in that field and demonstrate their own characteristics, or select institutions with expertise in different fields to enhance their own comprehensiveness. Through cooperation with different archives and the communication and connectivity brought by knowledge graphs, digital archive service institutions can better promote resource interaction, provide new communication environments for digital archive services, and drive digital archive service upgrades.

6 Conclusion

In the information age, people's information behaviors and methods are undergoing unprecedented changes. The rapid development of semantic web, big data, and deep learning technologies provides technical guarantees for the accuracy and intelligence of digital archive services. This paper proposes a knowledge graph-based digital archive service model that fully utilizes new smart technologies aligned with contemporary Internet technology development trends to address current challenges in archival services. Targeting problems such as difficult retrieval language recognition, complex retrieval procedures, and limited retrieval content in existing digital archive services, this study creates a knowledge graph architecture, proposes solutions, and continuously promotes the development of archival services and the integration of archival business with emerging technologies. However, this research only discusses digital archive services and still faces issues of remaining theoretical and difficult to implement. Future work will strive to advance the practical construction of digital archive knowledge graphs and further improve this concept by collecting real user feedback to achieve deeper enhancement and optimization of digital archive services.

References

[1] Wang Dianhua, Qian Tao, Qian Lixin, et al. Research on knowledge graph construction methods for archives[J]. Journal of Hubei University of Science and Technology, 2020, 40(1):127-130.

[2] Lei Jie, Li Sijing, Zhao Ruixue, et al. Research on knowledge graph construction and application for scientific research archive management[J]. Digital Library Forum, 2020(5):8-15.

[3] Yang Qianya. Research on data mining and intelligent utilization of China Unicom electronic archives[J]. Archives Science Study, 2018(6):105-109.

[4] Lei Jie, Zhao Ruixue, Li Sijing, et al. Research on construction of big data management system for scientific research archives driven by knowledge graph[J]. Digital Library Forum, 2020(2):19-27.

[5] Shu Zhongmei. Research on archival knowledge graph construction under digital humanities background[J]. Shanxi Archives, 2020(2):53-60.

[6] BALAJI B S, KARTHIKEYAN N K, KUMAR R. Fuzzy service conceptual ontology system for cloud service recommendation[J]. Computers & Electrical Engineering, 2018(69):435-446.

[7] Cao Ling, Wang Rong, Yan Xianglin. Analysis and reference of digital archive information service models of Ivy League universities in the United States[J]. Digital and Microfilm Imaging, 2013(3):30-33.

[8] Wang Wenqiang. Analysis of enterprise archival smart service model based on digital archives[J]. Archives of Machinery, Electronics, Weapons and Ships, 2019(4):76-78.

[9] Lian Zhiying. Construction of personalized digital archive information service model based on user needs[J]. Archives Science Bulletin, 2013(5):49-53.

[10] HAUSWEDELL T, NYHAN J, BEALS M H, et al. Of global reach yet of situated contexts: an examination of the implicit and explicit selection criteria that shape digital archives of historical newspapers[J]. Archival Science, 2020, 20(2):139-165.

[11] MARCONDES C H. Interoperability between digital collections in archives, libraries and museums: potentialities of linked open data technologies[J]. Ciência da Informação, 2016, 21(2):61-83.

[12] Zhang Weidong, Wang Ping. Research on personalized service model driven by archival user needs[J]. Archives Science Bulletin, 2007(2):82-86.

[13] Liu Qiao, Li Yang, Duan Hong, et al. Survey on knowledge graph construction technology[J]. Journal of Computer Research and Development, 2016, 53(3):582-600.

[14] Zhao Yue. Prospects of archival datafication in the big data era: significance and dilemmas[J]. Archives Science Study, 2019(5):52-60.

[15] Li Chao. Research on knowledge graph construction for video investigation[D]. Beijing: People's Public Security University of China, 2019:1-25.

[16] Yang Jing. Research on archival latent users[D]. Hefei: Anhui University, 2013:9.

[17] Zhao Hongyu. How to improve archival exchange between archival institutions[J]. Lantai World, 2020(S1):12.

Author Contributions: Xiong Huixiang provided research guidance; Yan Wuyue wrote and revised the paper.

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