Knowledge Graphs in Digital Humanities: Research Progress and Future Trends (Postprint)

Authors: Zhu Liya, ZHANG Jun, sonorous, Luo Shaohui, Landau

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

Abstract

[Purpose / Significance] To conduct a systematic review of knowledge graph research in the digital humanities field, aiming to provide possible future research directions and open research topics.

[Method / Process] Taking relevant literature published in domestic and international conferences and journals as the research object, and adopting a comprehensive induction method, this study systematically reviews the theoretical and practical development of knowledge graphs in the digital humanities field. It elaborates on the relevant concepts of knowledge graphs in digital humanities, and based on current research hotspots, reveals research trends from three aspects: data resource construction, key construction technologies, and platform intelligent applications, while also providing an outlook on future research trends.

[Result / Conclusion] It summarizes the future development trends of digital humanities knowledge graph research, indicating that future development will exhibit trends of multi-source data integration, multi-modal knowledge fusion, and interdisciplinary application.

Full Text

Knowledge Graph in the Field of Digital Humanities: Research Progress and Future Trends

Zhu Liya[1], Zhang Jun[1], Hong Liang[1], Luo Shaohui[2], Lan Du[2]
[1]School of Information Management, Wuhan University, Wuhan 430072
[2]Surveying, Mapping and Geographic Information Institute of Nanning, Nanning 530022

Abstract

[**Purpose/Significance**] This paper conducts a systematic review of knowledge graph research in the field of digital humanities, aiming to identify future research directions and open research topics. [**Method/Process**] Taking relevant literature published in domestic and international conferences and journals as the research object, this study employs a comprehensive induction method to systematically 梳理 the theoretical and practical development of knowledge graphs in digital humanities. It elaborates on the relevant concepts of knowledge graphs in this field and, based on current research hotspots, reveals research trends from three aspects: data resource construction, key construction technologies, and intelligent platform applications, while also providing an outlook on future research directions. [**Result/Conclusion**] The paper summarizes future development trends in digital humanities knowledge graph research, which will demonstrate multi-source data integration, multi-modal knowledge fusion, and multi-disciplinary cross-application.

Keywords: digital humanities; knowledge graph; smart data; data resource construction; semantic mining
Classification Number: G252.8
Citation Format: Zhu Liya, Zhang Jun, Hong Liang, et al. Knowledge Graph in the Field of Digital Humanities: Research Progress and Future Trends [J/OL]. Knowledge Management Forum, 2022, 7(1): 87-100 [citation date]. http://www.kmf.ac.cn/p/277/.

---

1. Introduction

Digital Humanities (DH) originated from humanities computing in the late 1940s, which focused on researching, learning, and innovating at the intersection of computing and humanities [1]. As information technology has deepened and digital resources have proliferated, humanities computing alone has become insufficient for achieving higher-level academic discoveries. Consequently, the concept of digital humanities emerged as a new interdisciplinary research field that conducts humanities research supported by emerging technologies such as computer technology, network technology, and multimedia technology [2].

In China, how to stimulate innovation and creativity through digitalization to promote the high-quality development of cultural industries and better meet the people's growing spiritual and cultural needs has become an important topic. For example, the Ministry of Culture' s "13th Five-Year Plan for Cultural Industry Development" released in 2019 emphasized promoting innovative development of digital cultural industries, including advancing the "Culture +" and "Internet +" strategies to facilitate the application of high-tech technologies like the Internet in all aspects of the cultural industry. In 2020, the national "14th Five-Year Plan" proposed implementing a digital strategy for the cultural industry. With the arrival of the "digital intelligence era" and the rise of digital humanities, data infrastructure and digital academic environments in digital humanities research have become important aspects of digital humanities resource development and utilization.

Integrating knowledge graphs into digital humanities research can bring new methods and perspectives. On the one hand, as an advanced knowledge organization method in the artificial intelligence era, knowledge graphs can provide excellent technical support for digital humanities research to discover patterns and connections previously invisible in textual resources. On the other hand, as an expression form of smart data, knowledge graphs provide a foundation for the mining and analysis of digital resources. Large-scale knowledge graph construction can improve the efficiency of building intelligent digital humanities systems and provide professional and intelligent knowledge services for researchers in the field and others who want to understand the humanities. However, although there are many research achievements in digital humanities knowledge graphs, they are relatively scattered and lack a systematic framework. Therefore, this paper conducts in-depth research on knowledge graphs in the digital humanities field and integrates relevant research findings.

2. Conceptual Analysis and Literature Collection in Digital Humanities Knowledge Graphs

In the library and digital humanities fields, the concept of knowledge graphs is deeply rooted in knowledge organization systems [3]. Digital humanities knowledge graphs aim to use this advanced knowledge organization method to integrate originally dispersed and heterogeneous massive data to meet the research needs of domain scholars and achieve intelligent knowledge services. Compared with general knowledge graphs, digital humanities knowledge graphs have the following characteristics:

First, regarding data, researchers have recognized the limitations of traditional resource utilization and development models and have begun to consciously transform ordinary digital resources in the digital humanities field into intelligent resources. The form has gradually evolved from databases with only search functions to intelligent platforms with reasoning and analysis functions, making full use of new information technologies to deeply mine knowledge.

Second, digital humanities knowledge graphs are based on scholar-oriented research needs. Their purpose differs from general knowledge graphs in that they are not designed to cover a wide range of knowledge for comprehensive retrieval, but rather to build a more comprehensive knowledge system on the basis of broad knowledge coverage to support intelligent domain knowledge service platforms.

Finally, digital humanities knowledge graphs involve relatively broad domains, and the construction process must fully consider the influence of different research fields. For example, when constructing a Tang poetry knowledge graph, Zhou Lina et al. [4] proposed that since Tang poetry knowledge involves three major fields—poetics, philology, and historiography—analyzing existing unresolved problems in these three fields can comprehensively identify the construction requirements for Tang poetry knowledge graphs. Therefore, there are many differences between digital humanities knowledge graphs and general knowledge graphs in construction methods, particularly in ontology construction, knowledge extraction, and knowledge reasoning.

2.1 Literature Sources

(1) Retrieval Scope. The literature for this study was primarily obtained through domestic and foreign databases. Considering the novelty of the research, literature from 2010 to 2021 was selected. Domestic literature was sourced from CNKI, focusing on core academic journals in library and information science such as *Journal of Library Science in China*, *Journal of the China Society for Scientific and Technical Information*, and *Data Analysis and Knowledge Discovery*. Foreign literature was sourced from WOS, Elsevier, EBSCO, and Springer databases, focusing on core academic journals in the Information Science & Library Science field, such as *MIS Quarterly*, *Journal of Information Technology*, and *International Journal of Information Management*.

(2) Retrieval Keywords. Domestic databases used "digital humanities" and "knowledge graph" as search terms, while foreign databases used "digital humanities" and "knowledge graph," searching through titles and subjects. The results were then screened, deduplicated, and corrected, removing literature with low relevance to the topic. Considering the limitations of using only these two keywords, which could not deeply reflect the specific research content of knowledge graphs in digital humanities, additional search terms such as "smart data," "ontology," "knowledge extraction," and "linked data" were selected to explore specific applications of knowledge graphs in digital humanities research, ensuring that the retrieval results could comprehensively cover representative research achievements in the field. The retrieval results were again screened, deduplicated, and corrected, ultimately yielding 131 domestic literature and 187 foreign literature pieces as the initial sample.

2.2 Research Hotspots Overview

Overall, research on knowledge graphs in digital humanities presents multidisciplinary and arts-sciences integration characteristics, covering history, philology, computer science, management, library science, and other disciplines. It organically integrates technology and culture that were easily separated in past research, leveraging rich data resources and mature practical systems from other disciplines to provide strong foundational support for digital humanities knowledge graph research, greatly enriching the research content in this field and significantly advancing the intelligent research system of digital humanities. The main research hotspots focus on three aspects:

(1) Data Resource Construction in Digital Humanities. This type of research is the starting point for digital humanities knowledge graph studies both domestically and internationally, mainly exploring various data resource constructions related to the digital humanities field, including ancient books, documents, images, videos, audio, and other structured, semi-structured, and unstructured data sources. F. Kaplan [5] treats big data research in digital humanities as a structured research field and proposes a division of three concentric research domains. Based on this foundation, domestic and foreign scholars have conducted in-depth research on data resource classification, characteristics, and digitalization methods in digital humanities, such as Dong Zheng' e et al. [6], who investigated digital humanities literature resources according to the characteristics of digital humanities. Data resource construction, as a fundamental step in building digital humanities knowledge graphs, can provide data source support for them.

(2) Knowledge Graph Construction Technology in Digital Humanities. This type of research is the focus of digital humanities knowledge graph studies. It utilizes various digital humanities data sources, addresses the characteristics of digital humanities data, and studies ontology construction, knowledge extraction, and disambiguation to solve the fusion of different knowledge graphs and cross-language entity alignment problems. Although domestic research started relatively late, it has initiated many targeted studies based on China' s cultural characteristics, such as the SinoPedia platform built by Chen Tao et al. [7], which uses RDF triples to assign unique URIs to currently public domain encyclopedia concept terms for resource persistence, contributing to the standardization and promotion of Chinese knowledge graphs and Chinese domain ontologies.

(3) Intelligent Application Platforms for Digital Humanities Knowledge Graphs. This research represents the inevitable path for the development of digital humanities knowledge graph studies, mainly focusing on the application of linked data technology in digital humanities to support large-scale, reusable digital humanities research, such as R. Hoekstra et al. [8], who introduced the ecological cycle of digital humanities data management projects. Using linked data technology in digital humanities enables researchers to publish and use data in flexible ways. Additionally, it emphasizes the reconstruction of data through reorganization, transforming it into "smart data" that can support domain research, forming a global knowledge network, and providing open-source and shared intelligent knowledge services for the public, researchers, and research institutions [9].

Based on the research hotspots identified through the above literature collection, the following sections will discuss in detail three aspects: data resource construction, knowledge graph construction technology, and intelligent service platforms in the digital humanities field.

3. Data Resource Construction in Digital Humanities

Digital humanities resource construction requires three stages, as shown in Figure 1 [Figure 1: see original paper]. The first stage is dataset construction, aimed at achieving electronic documentation and storing it in databases and other forms [10]. The second stage involves converting structured, semi-structured, and unstructured data into RDF-structured data to achieve syntactic-level unification. The final step is to associate different data sources through ontology fusion and resource linking to achieve distributed fusion of resources and semantic-level unification.

3.1 Achieving Digitalization of Domain Resources

Dataset construction is located at the foundational stage of the digital humanities application process. GLAMs (Galleries, Libraries, Archives, and Museums) have significant advantages in data accumulation, making them the main institutions for dataset construction, digitizing paper materials and organizing them. Digital humanities data is primarily in text form, but also includes multi-source data formats such as images, audio, video, and 3D data. Different data resource formats require different construction technologies, which will be analyzed below.

(1) Textual Materials. Textual materials include local classical texts, books, handwriting, genealogical materials, etc. These materials need to record and preserve the appearance, structure, and content of original documents through imaging technology. This process mainly utilizes Charge-Coupled Device (CCD) imaging technology and Complementary Metal Oxide Semiconductor (CMOS) image sensor technology for resource collection, which needs to be combined with Optical Character Recognition (OCR) to convert images into computer-recognizable ASCII codes and then into text resources, while requiring machine learning to accomplish recognition tasks. For example, M. Kestemont et al. [11] focused on studying medieval Latin manuscripts, using convolutional neural networks to recognize manuscripts and explaining the feasibility of automatic classification.

(2) Images. Images include maps, paintings, murals, etc. Their electronic methods are similar to textual materials, primarily using OCR and machine learning technologies for scanning and recognition tasks. For instance, S. A. Oliveira et al. [12] focused on Napoleonic cadastral maps from the early 19th century in the Veneto region, proposing the first fully automatic system that can automatically segment and interpret these maps. The system uses machine vision algorithms to extract geometric shapes from each fragment and further classifies, reads, and interprets handwritten labels.

(3) 3D Data. 3D data includes cultural relics, vessels, sculptures, etc. The digitization of 3D data utilizes photography, digital scanning, and editing technologies to digitally store information or reconstruct 3D digital models, finally using relevant software for digital restoration [13]. 3D scanning technology can record the most authentic and comprehensive morphological characteristics of cultural relics according to needs. Today, 3D scanning technology is increasingly applied in cultural relic protection, making the display and retrieval of cultural relics more digital. Meanwhile, this technology also facilitates cultural relic research, sharing, and dissemination. Foreign countries started earlier in this area with many influential projects, while domestic research, despite starting later, has also achieved many effective results. A famous project is Stanford University' s "Michelangelo Project," which conducted 3D scanning of world-famous sculptures for digital preservation.

(4) Audio and Video. Audio and video data include interviews, documentaries, and other multimedia data. Digitizing audio and video involves using technology to scan, reproduce, and transcribe them to achieve digitalization. In recent years, rescue and protection of audio-visual archives have gradually become key research directions, and combining digital technology has become an inevitable trend [14]. To achieve long-term preservation and broader utilization of audio and video archives, digitalization is a feasible method [15]. Therefore, restoration is a crucial 环节 in the audio and video digitalization process. For example, the Inner Mongolia Autonomous Region Archives uses COOL EDIT PRO 2.1 and ADOBE AUDITION CC software to digitally restore audio files, first normalizing volume levels, then performing noise reduction, and finally manually intervening to remove residual noise points. For video restoration, the principle of "minimum intervention" should be adhered to, maximizing the preservation of the original evidential value of audio-visual archives while ensuring they can be "heard clearly and seen clearly" [16].

3.2 Achieving Syntactic-Level Unification of Resources

With the development of science and technology, artificial intelligence and smart data have continuously entered people' s 视野, and research on them in various industries has deepened, promoting the development of digital humanities from "interconnection" to "intelligent connection." Data resources in humanities disciplines are diverse in type, multi-sourced, massive in volume, and heterogeneous in environment. Therefore, constructing data resources in this field requires achieving unification at both syntactic and semantic levels to effectively solve problems such as data heterogeneity, entity disambiguation, and linked sharing, thereby achieving semantic enhancement and value enhancement of data.

For structured data, conversion is typically performed using RDB2RDF methods, such as using D2R tools or the R2RML mapping language [17]. EXCEL and CSV files, which also have structured data characteristics, can be converted using OpenRefine. Semi-structured data falls between structured and unstructured data and can be regarded as a form of structured data that does not conform to relational database data model structures but contains relevant markers that can be used to separate semantic elements and hierarchically structure records and fields. Therefore, it is also called self-describing structure. We can use tools such as XML2RDF or JSON2RDF to achieve the conversion of unstructured data to RDF-structured data, a process known as RDFizer implementation. Unstructured textual data requires combining Natural Language Processing (NLP) and Named Entity Recognition (NER) technologies to extract structured data before RDF conversion. For structural extraction from image and audio-video files, resource entities are first identified through object detection before conversion.

3.3 Achieving Semantic-Level Unification of Resources

After structured, semi-structured, and unstructured data resources are uniformly converted into RDF-structured data, only syntactic-level unification is achieved. To achieve semantic-level unification and distributed fusion of resources, local RDF datasets need to be linked with externally open linked data resources.

Semantic association between resources from different data sources is typically accomplished through two steps: ontology fusion and resource linking. **(1) Ontology Fusion.** Current ontology fusion research mainly focuses on finding mappings between ontologies. With the development of ontology technology, semantic matching mechanisms and mapping methods between ontology concepts, instances, and attributes can achieve similarity correspondence between the smallest elements of ontologies, thereby achieving ultimate ontology fusion [18]. Currently, there is increasing research on ontology fusion both domestically and internationally, with many mature ontology fusion systems such as PROMPT and GLUE. AnchorPROMPT [19], developed by Stanford University, is a tool for finding mappings between ontologies. It first compares concepts and then uses ontology structures to identify potentially similar ontology components, but it cannot handle complex concepts and relational ontology mappings. GLUE [20] is one of the ontology mapping generation systems based on instances, using machine learning technology to find 1:1 mappings between ontologies based on classification ontologies. M. Lamé et al. [21] proposed a new ontology alignment framework that enables cultural heritage data providers to generate well-defined and well-formed terminology.

(2) Resource Linking. Different institutions often use their own domain names to define resource URL addresses during the RDF structuring process of entity data, requiring linking operations between these resources. Tools and frameworks such as LIMES, SILK, and LDIF can be used for automated linking between different resources, primarily through machine learning and character similarity algorithms to compare resource attribute values.

4. Key Construction Technologies for Digital Humanities Knowledge Graphs

4.1 Digital Humanities Knowledge Graph Construction Framework

Both linked data and generalized knowledge graphs use nodes and edges to represent entities and relationships. This paper mainly discusses how to use linked data to explain technologies in generalized knowledge graphs. Semantic knowledge graphs represented by linked data must have entities named in RDF, and different graphs have standard SPARQL query languages, thus solving knowledge representation and web service problems.

There are many differences between digital humanities knowledge graph construction methods and general knowledge graph construction methods, particularly in ontology construction, knowledge extraction, and knowledge fusion. This section combines knowledge graph construction technologies with the characteristics of digital humanities knowledge and summarizes the digital humanities knowledge graph construction framework based on the general knowledge graph structural framework, as shown in Figure 2 [Figure 2: see original paper].

The digital humanities knowledge graph construction framework mainly includes multi-disciplinary supporting foundations, digital humanities domain data sources, and digital humanities knowledge graph construction. Based on multi-disciplinary comprehensive support, ontology construction, knowledge extraction, knowledge fusion, and knowledge reasoning are performed on massive, multi-source, and heterogeneous digital humanities domain data sources to provide smart data output for the digital humanities field. Having previously analyzed data resource construction issues, the following sections will provide in-depth analysis of key technologies in the digital humanities knowledge graph construction process.

4.2 Analysis of Key Construction Technologies

4.2.1 Ontology Construction Ontologies can be divided into general ontologies and domain ontologies based on their target scope. The former aims to establish ontology knowledge that can be widely applied in different scenarios and represents a normative description of general knowledge. The latter provides a corresponding knowledge specification description for specific domains [22].

Currently, mainstream ontology construction methods are divided into manual construction and automated or semi-automated construction based on machine learning. The former relies on domain experts' knowledge and experience, resulting in high costs and low efficiency. Meanwhile, different experts have different understandings of the same 事物, making manual construction poorly scalable. The latter refers to automatically extracting relevant terms and attribute relationships from corpora by combining natural language processing, machine learning, and other technologies within an established ontology semantic framework. This method has gradually become mainstream.

Foreign research on systematic analysis of domain ontology construction methods is relatively mature. Literature analysis shows that there are eight typical ontology construction methods abroad: IDEF5 method, skeleton method, TOVE method, METHONTOLOGY method, KACTUS engineering method, SENSUS method, seven-step method, and cyclic acquisition method [23]. In contrast, domestic research started later with relatively backward technology, requiring 借鉴 of foreign construction methods while combining new content to form new perspectives. Currently, there are two main representative ontology construction methods in China: the thesaurus-based construction method and the semi-automated and automated construction method based on ontological engineering [24].

In recent years, scholars have constructed some large-scale general ontologies, such as DBPedia Ontology and YAGO. Large-scale practical domain ontologies in natural science fields have developed rapidly because their inter-concept relationships are relatively clear. Currently, influential domain ontologies include GeoNames Ontology, The Drug Ontology, UMLS SemNet, Gene Ontology, and SNOMED [25]. Unlike natural science fields, large-scale practical ontologies are relatively rare in humanities and social science fields where top-level semantic frameworks are difficult to define and conceptual relationships are more flexible [26]. Some scholars have attempted to conduct ontology construction research in related fields such as history and philosophy, such as national history ontology, Twenty-Four Histories ontology, and philosophy ontology [27-28]. Deng Jun et al. [29] constructed an oral history archive resource domain ontology model for the archival field, which helps archival scholars conduct in-depth research. Meanwhile, in drama, folklore, and other fields, some scholars have used metadata and ontology technologies for information resource description and organization [30].

In the semantic environment, domain ontology application has become an inevitability. Although current domestic construction methods are not yet perfect, automated and semi-automated construction methods will certainly be the future development trend. Further optimization of domain ontology construction will focus on the following aspects: establishing a complete evaluation mechanism, improving ontology reusability, and emphasizing ontology sharing. Meanwhile, constructing large-scale practical ontologies for digital humanities disciplines will also become an important direction for future scholar research.

4.2.2 Knowledge Extraction Knowledge extraction methods in the digital humanities field have matured, mainly divided into two approaches: rule-based methods and statistical learning-based methods.

The core of rule-based knowledge extraction lies in relationship rule definition and entity extraction on both sides of rules, where rule accuracy directly affects extracted knowledge quality. In digital humanities, rule-based methods need to consider collocation relationships and contextual contexts between words. This method has the advantages of high accuracy and simple construction. For example, Liu Youran et al. [31] proposed a rule-based statistical method for ancient Chinese sentence patterns. After labeling high-frequency characters, this method can label and count sentence patterns for unlabeled words according to set constraint rules, thereby simplifying manual statistical work in ancient Chinese research. The sentence pattern statistics accuracy can exceed 95% when constraint rules are reasonably set. However, this method also has many limitations, particularly its weak generalization ability due to strong rule specificity for texts in digital humanities. For example, Xie Minghong et al. [32] proposed identifying character relationships through fixed sentence pattern rules, but due to the diverse expression methods in Chinese texts, prediction results may be inconsistent with reality. To achieve better extraction effects, new rules must be formulated. Therefore, digital humanities researchers tend to prefer statistical machine learning-based methods.

Statistical machine learning-based methods have been increasingly applied in digital humanities. Compared with rule-based methods, statistical learning methods do not require rule construction and generally automatically learn parameters from training corpora. For example, L. L. Liu et al. [33] studied an algorithm for recognizing literary Chinese named entities for historical research using a conditional random fields-based method, which performed well in tests and extracted numerous personal and place names from *Local Gazetteers* to enrich the China Biographical Database (CBDB). Qin Heran et al. [34] used the TextRank model to extract keywords from ancient Chinese texts. Through experiments using the TextRank model to extract keywords from *Spring and Autumn Annals*, accuracy reached 84%, enabling digital humanities scholars to quickly understand historical events and the era's features. Moreover, as natural language processing technology continues to develop, this model' s application space is broad, applicable not only to ancient Chinese texts but also to modern Chinese, such as constructing automatic summarization systems.

Comprehensively, to obtain richer data to support digital humanities knowledge graph construction, data preprocessing can be performed before extraction to reduce extraction time and improve accuracy. Rule-based and statistical methods can also be combined. Since entities and relationships in digital humanities have certain characteristics, rules can be automatically generated after a small amount of manual annotation, which also helps improve the accuracy and efficiency of domain knowledge extraction.

4.2.3 Knowledge Fusion Traditional knowledge fusion problems mainly involve three aspects: knowledge fusion frameworks, knowledge fusion algorithms, and knowledge fusion applications. Knowledge fusion algorithms can be divided into two categories: knowledge fusion algorithms based on information fusion technology and knowledge fusion algorithms based on fusion rules, with most knowledge fusion frameworks built upon ontologies [35]. Many studies

借鉴 information fusion algorithms, transplanting them into knowledge fusion to construct new algorithms specifically for knowledge fusion. Three popular knowledge fusion methods for fusion decision processing are based on Bayes methods, D-S theory, and ant colony optimization algorithms. Zhou Fang et al. [36] improved result credibility and enhanced system task goal achievement capabilities through fusion processing in the knowledge management field. The latter approach uses rules for knowledge representation by finding associations between information.

In digital humanities, knowledge fusion is mainly used to establish associative relationships between entities from different sources, integrating data discovered from multiple distributed heterogeneous information sources while identifying and resolving potential ambiguities, data redundancies, and uncertainties to form new knowledge [37]. Knowledge fusion can effectively solve problems such as data heterogeneity, entity disambiguation, and linked sharing in digital humanities, achieving semantic enhancement and value enhancement of data. For example, Chen Tao et al. [38] linked converted RDF data with datasets such as the Shanghai Library Name Authority Database, VIAF, and DBPedia when building CBDBLD (CBDB Linked Data Platform), using the SILK or LIMES framework for linking. F. Frontini et al. [39] proposed an algorithm to automatically eliminate ambiguities mentioned in French literary criticism corpora, successfully combining general knowledge bases (such as DBpedia) with domain-specific knowledge bases.

4.2.4 Knowledge Reasoning Knowledge reasoning in knowledge graphs targets the incompleteness of existing facts or relationships to mine or infer unknown or implicit semantic relationships. Generally, knowledge reasoning objects can be entities, relationships, and knowledge graph structures. Currently, there are three main methods: logic rule-based knowledge reasoning, embedding representation-based knowledge reasoning, and neural network-based knowledge reasoning. As one of the core functions of knowledge graphs, knowledge reasoning provides ideas for solving the incompleteness and uncertainty of knowledge caused by the historical nature of digital humanities, though mature applications are still rare in current digital humanities projects.

Path rule-based knowledge reasoning improves computational efficiency by randomly sampling extracted relationship path features but reduces the utilization of information in knowledge graphs. Meanwhile, relationship inference models built using supervised learning methods are largely influenced by training data. In response, Liu Qiao et al. [40] proposed a bidirectional semantic hypothesis for global relationship reasoning, combining local modules for weighted merging to obtain a complete logical rule reasoning algorithm. Zhou Lina [41] proposed a domain knowledge reasoning framework for ontology construction. Through the TPO4DK model, formal reasoning rules were constructed to perform knowledge reasoning on Tang dynasty poets' schools, poetry themes, and subject attributes between poets and poetry-poet ontologies, achieving application in version authentication and forgery detection in Tang poetry philology. Lu Quan et al. [42] proposed a fuzzy ontology representation model based on OWL language, using SWRL language to represent precise and fuzzy rules to construct a reasoning model for knowledge discovery. This model can simultaneously describe precise and fuzzy knowledge, simplifying the representation and processing of fuzzy knowledge.

Meanwhile, multi-source heterogeneous data contained in digital humanities resources, particularly semantic relationships and conceptual hierarchical structures between image data resources, also drive knowledge reasoning in the field. For example, Zhou Zhi et al. [43] conducted multi-layer descriptions of image resource semantics by referring to the Eakins image semantic hierarchy model and the digital image semantic description hierarchy model proposed by Wang Xiaoguang et al. [44], achieving deep association between entities and concepts to meet the needs of knowledge reasoning.

Embedding representation-based knowledge reasoning technology has obvious advantages. By mapping implicit association information in graph structures to Euclidean space, originally difficult-to-discover relationships become apparent. Therefore, embedding representation-based reasoning is an important component of knowledge graph reasoning technology. Neural network-based knowledge reasoning fully utilizes neural networks' modeling capabilities for nonlinear complex relationships, enabling deep learning of graph structure features and semantic features to effectively predict missing relationships in graphs. Neural network methods applied to knowledge graph reasoning mainly include CNN methods, RNN methods, Graph Neural Networks (GNN) methods, DRL methods, etc. [45].

5. Intelligent Application Platforms for Digital Humanities Knowledge Graphs

5.1 Overview of Relevant Platform Projects

Against the backdrop of rapid information technology development, the ways of information acquisition, storage, and dissemination have undergone tremendous transformation, making data one of the foundations and cores of digital humanities research. Consequently, digital humanities scholars' approaches to processing research materials in the field have also changed dramatically. In traditional humanities research, scholars often focused on data collection and organization. However, due to the lack of digital technology and quality issues with original materials, data compiled by scholars was often incomplete and fragmented. After in-depth development of digital technology, although data collection in the humanities field became faster and more comprehensive, the data remained fragmented and 不利于 domain scholars' research. As digital humanities knowledge graph scales gradually expand, traditional relational databases cannot effectively manage the data. Scholars in this field often need cross-querying of multiple datasets, and associations exist between different data types such as images, texts, and audio. Discovering these connections helps advance humanities research. Therefore, current research generally adopts linked data technology (i.e., semantic knowledge graphs) to achieve data management in digital humanities. After comparing linked data technology with generalized knowledge graphs, Chen Tao et al. [46] pointed out that linked data focuses on knowledge publishing and linking. Unlike generalized knowledge graphs that emphasize "mining," linked data technology emphasizes "reasoning," i.e., displaying associations between resources. Using linked data technology can support large-scale, reusable digital humanities research [47], transforming data through reorganization into "smart data" that can support domain research and forming a global knowledge network.

Table 1 lists several typical representatives of digital humanities platform construction both domestically and internationally.

Table 1. Digital Humanities Linked Data Platform Practices

Platform Name	Developing Institution	Main Features
China Biographical Database (CBDB) Linked Data Platform	Peking University Center for Research on Ancient Chinese History, Harvard University Fairbank Center for Chinese Studies	Displays kinship and social relationships between figures, forming unique social networks to mine and present implicit relationships
Sheng Xuanhuai Archives Knowledge Base	Shanghai Library, Shanghai Institute of Scientific and Technical Information	Collects private archives on politics, economy, society, military, diplomacy, finance, trade, and education—first-hand historical materials for modern Chinese history research

Platform Name	Developing Institution	Main Features
Europeana	European Commission	Integrates representative European cultural heritage resources, providing one-stop browsing and retrieval services for dissemination and sharing of European digital cultural resources
Chinese Genealogy Knowledge Service Platform	Shanghai Library	Based on large amounts of data combined with time and space, provides panoramic visualization and statistical analysis of surnames, figures, and their relationships
Venice Time Machine	École Polytechnique Fédérale de Lausanne, University of Venice	Digitizes, transcribes, indexes, and links massive historical archives
Dunhuang Murals Thesaurus Linked Data Service Platform	Wuhan University Digital Humanities Center	Provides a controlled vocabulary for deep semantic annotation, semantic retrieval, knowledge organization, information linking, and sharing of Dunhuang murals digital resources

Platform Name	Developing Institution	Main Features
Getty Digital Museum	Getty Research Institute, USA	Publishes descriptive metadata standards and cataloging specifications, various data value standards, and data exchange standards for cultural relic digitalization construction
Chinese Traditional Architecture Digital Research Tool Project	Vanderbilt University Center for Digital Humanities	Develops digital research tools consisting of four interrelated parts: open database websites, building complexes, individual structures, and structural elements
Chinese Local Gazetteers Database	Beijing Wenjin Bookstore	Arranges according to existing administrative divisions, covering Song, Yuan, Ming, Qing, and Republic of China period block-printed, handwritten, and manuscript versions of local gazetteers
Welsh Newspapers Online	National Library of Wales	Digitizes newspaper archives, currently containing approximately 420,000 digitized newspapers from and related to Wales

It can be seen that digital humanities linked data platforms span rich fields, mainly including history, archival science, art, and literature, among which history is one of the fields with the most digital humanities platform practices. Other related fields are also inextricably linked to history, reflecting the importance that countries around the world attach to the preservation and utilization of historical and cultural resources.

5.2 Platform Characteristics Analysis

5.2.1 Strong Cross-Border Cooperation A prominent feature of advanced digital humanities linked data platforms both domestically and internationally is cross-border cooperation, which is required by the interdisciplinary nature of digital humanities. Cooperation methods can be mainly divided into two aspects: On the one hand, there is extensive cooperation between domestic and international institutions. S. Wong [48] pointed out that the collaborative nature of digital humanities is one of the core values of the field, and collaborative approaches can utilize the strengths and expertise of various institutions to produce profound impacts. For example, Europeana has over 200 cultural institutions from more than 15 countries contributing to its open datasets, including famous institutions such as the British Library in London, the Rijksmuseum in Amsterdam, and the Louvre in Paris, as well as smaller heritage organizations and libraries elsewhere in Europe [49]. Additionally, the China Biographical Database (CBDB) project, developed collaboratively by Peking University' s Center for Research on Ancient Chinese History and Harvard University' s Fairbank Center for Chinese Studies, is a classic practice of establishing databases through cooperation between domestic and foreign research centers. The platform can display various relationships between historical figures and form unique social networks to mine and present implicit relationships [50], promoting Western understanding of traditional Chinese culture while researching Chinese history.

On the other hand, there is cooperation between off-campus institutions and universities. Most digital humanities institutions are affiliated with universities, relying on university libraries for platform construction, with university libraries and archives providing data resources and talent, and enterprises and foundations providing funding. For example, the Lincoln Writings Database developed by the University of Illinois at Urbana-Champaign' s Institute for Computing in Humanities, Arts, and Social Sciences in cooperation with the Abraham Lincoln Presidential Library and Museum was led by the University of Illinois, providing talent resource support and subsequent platform management and services, while the Abraham Lincoln Presidential Library provided relevant historical resources, forming mutual support between talent resources and historical resources [51]. Establishing broad connections between academic institutions, libraries, archives, museums, enterprises, and foundations, combined with multi-disciplinary participation from humanities, social sciences, and science and engineering, facilitates resource integration and innovative utilization.

5.2.2 Strong Practice Orientation First, many digital humanities linked data platforms provide services for a series of disciplines including humanities, offering various digital academic functions such as digital imaging, digital preservation, metadata creation, data curation and management, GIS and digital mapping, and digital publishing. For example, the Europeana Digital Manuscripts Project, developed through cooperation between the Berlin School of Library and Information Science at Humboldt University of Berlin, the University of Mannheim, the Open Knowledge Foundation, and multiple institutions, constructed the DM2E dataset, providing metadata and links as well as relevant tools for displaying, processing, and integrating data, enabling digital humanities researchers and the public interested in European history and culture to directly access digitized content from various European cultural heritage institutions [52]. This reflects the characteristic of digital humanities linked data platforms serving practice and solving practical problems.

Second, these platforms emphasize the presentation and promotion of results to the public. For example, the Dunhuang Murals Thesaurus Linked Data Service Platform reduces the cognitive difficulty of thesauruses through visualization, achieving the transition from professional thesauruses to versions suitable for public use [53]. The Chinese Genealogy Knowledge Service Platform developed by Shanghai Library [54] provides panoramic visualization and statistical analysis of surnames, figures, and their relationships based on large amounts of data combined with time and space. Thus, presenting results through digital humanities linked data platforms helps promote digital humanities research and enhance the discipline' s influence, while also facilitating the extension of culture from the real world to digital space and enriching human digital civilization.

5.2.3 Prominent Data Silo Phenomenon The wisdom of digital humanities linked data platform data resources is mainly reflected in three aspects: timeliness, accessibility, and usability. Therefore, dynamic, openly linked data resources need to be formed to continuously enrich their content and forms. In recent years, increasing attention has been paid to digital humanities linked data platform construction both domestically and internationally. However, new challenges have also emerged. Wang Xiaoguang proposed the phenomenon of data disorder in digital humanities research, which is particularly serious in mainland China. He pointed out that the diversification of digital resource construction entities, with relevant research institutions such as libraries, museums, and archives investing considerable funds and human resources, has led to the emergence of countless larger "data silos," more severe than in the era of paper documents [55]. This phenomenon dilutes the efforts made by domain scholars for platform construction and may even give the impression that the lifecycle of digital humanities research is very short.

Examining the causes of data silo formation: First, as research progresses, the volume of resource data and the scope of research materials are expanding. In addition to traditional literature resources, other materials such as physical objects, images, audio, and video are included in digital humanities scholars' research scopes. The complexity of basic materials in digital humanities can easily lead to isolated research. Second, many platform managers tend to focus on planning and launching new projects, often neglecting subsequent management and maintenance of old projects [56]. Over time, original data resource formats may become incompatible with existing technologies, making old data resources unable to match new target user needs. If existing technical methods and operating environments cannot be updated in a timely manner and new projects are launched blindly, platform resources will struggle to remain fresh. How to improve the data silo phenomenon and achieve unified representation of smart data resources in digital humanities has become an important issue on the development path of intelligent knowledge service platforms in digital humanities.

6. Future Trends in Digital Humanities Knowledge Graph Research

Based on a comprehensive review of recent digital humanities knowledge graph research achievements and the current intelligent trends in digital technology, the following development trends can be observed:

(1) Multi-Source Data Integration. Long-term preservation of data is a very important basic function of digital humanities knowledge graph platforms. Compared with other fields, data in digital humanities is relatively special, containing languages, documents, paintings, music, and other forms whose dimensions transcend physically measurable ranges and depend more on semantics and syntax [57]. Research in digital humanities cannot be separated from the digitization of humanities literature, and massive data resources have extraordinary value in the digital humanities field. How to process this data and transform it into machine-understandable and processable resources is crucial. Digital humanities research cannot rely solely on previous data resources but also requires large amounts of fresh, currently being created data. Therefore, social networks and open access information can be used as information sources to make cross-regional, cross-disciplinary, and cross-national connections closer. After deep semantic annotation, structuring, formalization, and visualization processing, data can be transformed into advanced forms of smart data and promoted to more refined analysis.

(2) Multi-Modal Knowledge Fusion. Early multi-modal knowledge fusion in digital humanities mainly targeted various types of knowledge from different knowledge sources, emphasizing the diversity of knowledge origins. In the future, multi-modal knowledge fusion will further break through traditional time and space limitations, covering and expanding diverse features of different knowledge sources. Relying on the data integration capabilities of knowledge graph intelligent platforms, it will 打通 semantic resources across multiple dimensions such as text, images, entities (people, places, time periods, regions, events), providing capability support for systematic, semantic, and systematic digital humanities resource organization and research. Additionally, different interpretations of the same knowledge source constitute different dimensions and levels of digital humanities resources, better meeting the deep information needs of digital humanities research and achieving continuous intelligent knowledge services in big data environments.

(3) Multi-Disciplinary Cross-Application. Innovative research on digital humanities linked data platform construction applied to multiple disciplinary fields helps form organic and mutually complementary and verifiable integrated results, narrowing the distance between different disciplines and promoting disciplinary integration. On the one hand, the degree of disciplinary specialization continues to improve, with internal development gradually becoming refined, enabling more specific and in-depth coverage of digital humanities content. On the other hand, disciplinary integration generates new disciplines such as digital arts and digital history. Liang Chen et al. [58] pointed out that digital technology or database platforms can also be accelerators or colliders for micro-information, presenting various features, trends, and patterns during data crossing and collision. These changes gradually require domain researchers to continuously break through boundaries between different specialties, bringing new unique research paradigms to digital humanities research and further promoting the stable development of interdisciplinary studies.

From construction to providing infrastructure support for digital humanities research, digital humanities knowledge graph research has experienced continuous development and transformation to adapt to the transition from traditional literature resources to smart data resources in the "digital intelligence era." Currently, digital humanities knowledge graphs can already provide good knowledge discovery and reasoning functions, support description and fusion of multiple types of digital humanities resources, and meet the needs of long-term cultural preservation and collaborative construction. This paper takes relevant literature published in domestic and international conferences and journals as the research object, investigating three aspects: data resource construction, knowledge graph construction, and intelligent service platforms in digital humanities. It recognizes that digital humanities knowledge graph research can provide unified and normative methodological references for digital construction of domain resources, provide infrastructure for digital humanities research, and better achieve the transformation and upgrading of smart data resources. In this process, new opportunities and challenges continue to emerge, and knowledge graphs, as an advanced knowledge organization method in the artificial intelligence era, can fully play their role as knowledge fusion intermediaries, providing continuous momentum for development in the "digital intelligence era" and guidance and direction for China' s future digital humanities development path.

References

[1] Li Qihu, Yin Li, Zhang Quan. Humanities Computing in the Information

Age [J]. Science, 2015, 67(1): 35-39, 4.
[2] Liu Wei, Ye Ying. Discussion on the Technical System and Theoretical Structure of Digital Humanities [J]. Journal of Library Science in China, 2017, 43(5): 32-41.
[3] Haslhofer B, Isaac A, Simon R. Knowledge Graphs in the Libraries and Digital Humanities Domain [J]. arXiv preprint, 2018, arXiv:1803.03198.
[4] Zhou Lina, Hong Liang, Gao Ziyang. Construction of Tang Poetry Knowledge Graph and Design of Its Intelligent Knowledge Service [J]. Library and Information Service, 2019, 63(2): 24-33.
[5] Kaplan F. A Map for Big Data Research in Digital Humanities [J]. Frontiers in Digital Humanities, 2015, 2(1): 1.
[6] Dong Zheng' e, Chen Huilan. Investigation and Development Countermeasures of Digital Humanities Resources [J]. Information and Documentation Services, 2015(5): 103-109.
[7] Chen Tao, Liu Wei, Zhu Qinghua. Research on the Construction of SinoPedia: A Chinese Encyclopedia Concept Term Service Platform [J]. Journal of Library Science in China, 2018, 44(4): 4-18.
[8] Hoekstra R, Merono-Peñuela A, Dentler K, et al. An Ecosystem for Linked Humanities Data [C]// European Semantic Web Conference. Cham: Springer, 2016: 425-440.
[9] Zeng M L. Smart Data for Digital Humanities [J]. Journal of Data and Information Science, 2017, 2(1): 1-12.
[10] Wang Jun, Zhang Liyuan. Research on International Digital Humanities Progress [J]. Digital Humanities, 2020(1): 1-23.
[11] Kestemont M, Stutzmann D. Script Identification in Medieval Latin Manuscripts Using Convolutional Neural Networks [C]// Premiere Annual Conference of the International Alliance of Digital Humanities Organizations. Montreal: McGill University, 2017.
[12] Oliveira S A, Kaplan F, Di Lenardo I. Machine Vision Algorithms on Cadaster Plans [C]// Premiere Annual Conference of the International Alliance of Digital Humanities Organizations. Montreal: McGill University, 2017.
[13] Zhang Hui, Wang Dongmei. Research on Digital Protection of Tang Dynasty Sculptures Based on 3D Scanning Technology [J]. Art and Design (Theory), 2016, 2(4): 91-93.
[14] Liu Jiangxia. Research on Quality Control of Analog Audio-Video Archives Digitization [J]. Archives Science Study, 2018(1): 101-106.
[15] Qian Wanli. Digital Processing of Traditional Audio-Visual Archives [J]. Archives and Construction, 2007(8): 22-24.
[16] Luo Yongjun, Bi Xiaoran, Hao Yang. Research on Rescue Technology for Precious Audio-Video Archives of Inner Mongolia National Culture [J]. Heilongjiang Archives, 2020(5): 43-45.
[17] R2RML: RDB to RDF Mapping Language [EB/OL]. [2021-07-23]. https://www.w3.org/2001/sw/rdb2rdf/r2rml/.
[18] Xiong Shun, Liu Pingzhi, Su Zongyi, et al. Research on Geographic Information Ontology Fusion Method Based on Semantic Matching Mapping [J]. Surveying and Mapping Science and Engineering, 2017(1).

[19] Noy N F, Musen M A. The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping [J]. International Journal of Human-Computer Studies, 2003, 59(6): 983-1024.
[20] Doan A H, Madhavan J, Dhamankar R, et al. Learning to Match Ontologies on the Semantic Web [J]. The VLDB Journal, 2003, 12(4): 303-319.
[21] Lamé M, Pittet P, Ponchio F, et al. Heterotoki: Non-Structured and Heterogeneous Terminology Alignment for Digital Humanities Data Producers [C]// Open Data and Ontologies for Cultural Heritage. Rome: Antonella Poggi, 2015.
[22] Ren Feiliang, Shen Jikun, Sun Binbin, et al. Survey on Domain Ontology Construction Technology from Text [J]. Chinese Journal of Computers, 2019, 42(3): 654-676.
[23] Shang Xinli. Comparative Analysis of Foreign Ontology Construction Methods [J]. Library and Information Service, 2012, 56(4): 116-119.
[24] Yue Lixin, Liu Wenyun. Comparative Study of Domain Ontology Construction Methods at Home and Abroad [J]. Information Studies: Theory & Application, 2016, 39(8): 119-125.
[25] Wimalasuriya D C, Dou D. Ontology-Based Information Extraction: An Introduction and a Survey of Current Approaches [J]. Journal of Information Science, 2010, 36(3): 306-323.
[26] He Lin, Chen Yaling, Sun Kedi. Research on Knowledge Ontology Construction Technology for Pre-Qin Classics [J]. Library and Information Service, 2020, 64(7): 13-19.
[27] Wang Ying, Zhang Zhixiong, Sun Hui, et al. Research on Semantic Revelation and Organization of National History Knowledge [J]. Journal of Library Science in China, 2015, 41(4): 55-64.
[28] Thakker D, Karanasios S, Blanchard E, et al. Ontology for Cultural Variations in Interpersonal Communication: Building on Theoretical Models and Crowdsourced Knowledge [J]. Journal of the Association for Information Science and Technology, 2017, 68(6): 1491-1507.
[29] Deng Jun, Wang Ruan. Knowledge Organization and Association Analysis of Oral History Archive Resources [J]. Information and Documentation Services, 2021, 42(5): 58-67.
[30] Zhou Yaolin, Zhao Yue, Sun Jingqiong. Research Path of Intangible Cultural Heritage Information Resource Organization and Retrieval—Investigation and Design Based on Ontology Methods [J]. Journal of Intelligence, 2017, 36(8): 166-174.
[31] Liu Youran, Long Dan. Design and Implementation of a Rule-Based Statistical Method for Ancient Chinese Sentence Patterns [C]// Proceedings of the 15th Chinese Lexical Semantics Workshop. Beijing: Foreign Language Teaching and Research Press, 2014: 428-433.
[32] Xie Minghong, Ran Qiang, Wang Hongbin. Chinese Character Relationship Extraction Method Based on Synonym Forest and Rules [J/OL]. Computer Engineering & Science, 2021, 43(9): 1660-1667.
[33] Liu C L, Huang C K, Wang H, et al. Mining Local Gazetteers of Literary Chinese with CRF and Pattern-Based Methods for Biographical Information

in Chinese History [C]// Proceedings of 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, 2015: 2459-2468.
[34] Qin Heran, Wang Dongbo. Application of Keyword Extraction for Pre-Qin Ancient Chinese in Digital Humanities—Taking "Spring and Autumn Annals" as an Example [J]. Library Journal, 2020, 39(11): 97-105.
[35] Tang Xiaobo, Zhu Juan. Review of Key Issues in Knowledge Fusion Under Big Data Environment [J]. Library Journal, 2017, 36(7): 10-16.
[36] Zhou Fang, Liu Yuzhan, Han Liyan. Research on Knowledge Fusion Method Based on Fuzzy Set Theory [J]. Journal of Beijing Institute of Technology (Social Sciences Edition), 2013, 15(3): 67-73.
[37] Gao Jinsong, Liang Yanqi. Research on Knowledge Fusion Model in Linked Data Environment [J]. Information Science, 2016, 34(2): 50-54.
[38] Chen Tao, Liu Wei, Shan Rongrong, et al. Research on Application of Knowledge Graphs in Digital Humanities [J]. Journal of Library Science in China, 2019, 45(6): 34-49.
[39] Frontini F, Brando C, Ganaschia J G. Semantic Web Based Named Entity Linking for Digital Humanities and Heritage Texts [C]// Proceedings of First International Workshop Semantic Web for Scientific Heritage at the 12th ESWC 2015 Conference. Portorož: Fabien Gandon, 2015: 77-88.
[40] Liu Qiao, Han Minghao, Jiang Yiyi, et al. Relation Reasoning Algorithm Based on Double-Layer Random Walk [J]. Chinese Journal of Computers, 2017, 40(6): 1275-1290.
[41] Zhou Lina. Research on Tang Poetry Ontology Construction and Intelligent Application for Domain Knowledge Service [D]. Wuhan: Wuhan University, 2020.
[42] Lu Quan, Liu Ting, Zhang Liangtao, et al. Research on Fuzzy Ontology Fusion and Reasoning Model for Knowledge Discovery [J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(4): 333-344.
[43] Zhou Zhi, Jiang Lin. Research on Knowledge Organization Model Construction for Digital Humanities Image Resources [J]. Research on Library Science, 2021(8): 66-72, 65.
[44] Wang Xiaoguang, Jiang Yanyu, Zhang Lu. Empirical Research on Semantic Description Hierarchy Model for Dunhuang Mural Images [J]. Library and Information Service, 2015, 59(19): 122-129.
[45] Tian Ling, Zhang Jinchuan, Zhang Jinhao, et al. Survey on Knowledge Graphs—Representation, Construction, Reasoning, and Knowledge Hypergraph Theory [J]. Journal of Computer Applications, 2021, 41(8): 2161-2186.
[46] Chen Tao, Liu Wei, Shan Rongrong, et al. Research on Application of Knowledge Graphs in Digital Humanities [J]. Journal of Library Science in China, 2019, 45(6): 34-49.
[47] Hoekstra R, Merono-Peñuela A, Dentler K, et al. An Ecosystem for Linked Humanities Data [C]// Proceedings of European Semantic Web Conference. Cham: Springer, 2016: 425-440.
[48] Shun Han Rebekah W. Digital Humanities: What Can Libraries Offer? [J]. Libraries and the Academy, 2016, 16(4): 669-690.
[49] Isaac A, Haslhofer B. Europeana Linked Open Data–data.europeana.eu [J].

Semantic Web, 2013, 4(3): 291-297.
[50] Tsui L H, Wang H. Harvesting Big Biographical Data for Chinese History: The China Biographical Database (CBDB) [J]. Journal of Chinese History, 2020, 4(2): 505-511.
[51] Institute for Computing in Humanities, Arts, and Social Sciences [EB/OL]. [2021-12-15]. http://chass.illinois.edu/.
[52] Baierer K, Dröge E, Eckert K, et al. DM2E: A Linked Data Source of Digitised Manuscripts for the Digital Humanities [J]. Semantic Web, 2017, 8(5): 733-745.
[53] Wang Xiaoguang, Hou Xilong, Cheng Hanghang, et al. Construction and Linked Data Publishing of Dunhuang Murals Thesaurus [J]. Journal of Library Science in China, 2020, 46(4): 69-84.
[54] Xia Cuijuan, Liu Wei, Chen Tao, et al. Development Practice of Genealogy Linked Data Service Platform [J]. Journal of Library Science in China, 2016, 42(3): 27-38.
[55] Wang Xiaoguang. Digital Humanities and Smart Data [J]. Shanghai University Library and Information Work Research, 2018, 28(2): 25, 24.
[56] Reed A. Managing an Established Digital Humanities Project: Principles and Practices from the Twentieth Year of the William Blake Archive [J]. Virginia Tech, 2014, 8(1): 1-17.
[57] Schöch C. Big? Smart? Clean? Messy? Data in the Humanities [J]. Journal of Digital Humanities, 2013, 2(3): 1-8.
[58] Liang Chen, Li Zhongqing. From Micro Data to Macro History: Digital History as a Bridge [J]. China Social Science Review, 2021(2): 84-92, 159.

Author Contributions

Zhu Liya: Participated in framework formulation, collected and organized materials, wrote and revised the paper;
Zhang Jun: Collected and organized materials, wrote and revised the paper;
Hong Liang: Proposed the paper topic and research framework, guided paper writing;
Luo Shaohui: Proposed writing ideas for some sections;
Lan Du: Proposed writing ideas for some sections.

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