# Current Status and Implications of "Digital-Oriented" Research in China's Digital Humanities from the Perspective of Data and Technology Application: Postprint

**Authors:** Shi Jing, Li Yang

**Date:** 2023-04-01T16:02:58+00:00

## Abstract

[Objective/Significance] In recent years, the domestic digital humanities field has seen a surge of achievements, gradually transitioning from a cognitive exploration stage to a digital exploration stage. From micro perspectives of "data resources" and "digital technology", this study employs quantitative analysis to depict the true full picture of the collision between "digital" and "humanities", with the aim of gaining an intuitive and practical understanding and reflection on "digital-type" research in digital humanities. [Methods/Process] Utilizing methods such as bibliometrics, qualitative analysis, and multidimensional scaling analysis, this study conducts an in-depth analysis of domestic digital humanities-related research from dimensions including research landscape distribution, interdisciplinary characteristics, "digital" application status, and main research types, and offers relevant reflections. [Results/Conclusions] From the perspective of research landscape, digital humanities research is evenly distributed with a hundred schools of thought contending, possessing broad space across multiple disciplines; from the perspective of data foundation, a few large-scale thematic databases and researcher self-organized data constitute the main data sources; from the perspective of research types, the three categories of resource construction-oriented, tool-driven, and scenario integration-oriented research each have their own focus, with scenario integration-oriented research achieving a better balance between "digital" and "humanities". In the future, it is necessary to actively build relevant data infrastructure, cultivate scholars' digital academic capabilities, and promote the transformation of the "humanities" role to support the paradigm shift in digital humanities research.

## Full Text

# The Present Situation and Enlightenment of "Digital-Type" Research in Digital Humanities in China from the Perspective of Data and Technology Application

**Shi Jing, Li Yang**
School of Information Management, Nanjing University, Nanjing 210023

**Abstract:** [Purpose/Significance] In recent years, numerous research findings have emerged in the field of digital humanities in China, gradually transitioning from a cognitive exploration stage to a digital exploration stage. From the micro perspectives of "data resources" and "digital technology," this study employs quantitative analysis to depict the actual landscape of the collision between "digital" and "humanities," aiming to provide an intuitive and practical understanding of "digital-type" research in digital humanities. [Method/Process] Using bibliometrics, qualitative analysis, and multidimensional scaling analysis, this paper conducts an in-depth analysis of domestic digital humanities research from dimensions including research institution distribution, interdisciplinary characteristics, "digital" utilization, and main research types, and proposes relevant reflections. [Result/Conclusion] From the perspective of research institutions, digital humanities research is evenly distributed with diverse schools of thought, demonstrating broad space across multiple disciplines. From the perspective of data foundations, a few large-scale thematic databases and researcher self-collected data constitute the primary data sources. From the perspective of research types, resource construction research, tool-driven research, and scenario-integrated research each have distinct focuses, with scenario-integrated research achieving a better balance between "digital" and "humanities." In the future, it is necessary to actively build relevant data infrastructure, cultivate scholars' digital academic capabilities, and promote the transformation of the "humanities" role to support paradigm shifts in digital humanities research.

**Keywords:** digital humanities; interdisciplinary; data resources; new technology; humanistic value

Digital humanities (DH), originating from humanities computing, is an interdisciplinary academic field where computer or digital technology intersects with humanities disciplines, and can also be defined as a new approach to humanities scholarship. Its distinctive feature lies in cultivating a bidirectional relationship between humanities and digital technology: using "digital" to answer "humanities" questions while simultaneously using "digital" to pose new "humanities" questions. Digital humanities has attracted widespread attention in China. As early as 2003, the article "Frontier Exploration of the Informationization Process—A Summary of the 'Digitalization and 21st Century Humanistic Spirit' Seminar" cited the arguments of renowned Chinese philosopher Professor Chen Zhiliang, first using the term "digital humanities" [2]. In 2016, Peking University hosted the inaugural "Digital Humanities Forum," after which the digital

humanities wave erupted domestically, generating a large volume of research primarily focused on basic theories of digital humanities, the transformation of humanities scholarship driven by technology, changes in academic culture triggered by new collaboration models, and infrastructure construction for digital humanities research [3]. In 2018, "Digital Humanities Research from a Big Data Perspective" was selected as one of China's top ten academic hotspots of the year, with scholars from multiple disciplines—including history, literature, linguistics, library and information science, and art history—increasing their research efforts in digital humanities.

However, as research has progressed, issues such as unclear research positioning, lack of theoretical core, technology worship, and "fast-food" knowledge production have become increasingly prominent, leading some scholars to question: What has "digital" actually brought to "humanities"? What is the role and positioning of "digital" in humanities disciplines [4]? Academic and practical debates on these questions remain unresolved. It must be acknowledged that the collision between "digital" and "humanities" has opened new spaces for thinking about the prosperous development of humanities disciplines. Problems that were previously unstudiable or difficult to study in traditional humanities research can often be resolved or given new solutions with digital support.

In recent years, influenced by concepts and environments such as open data, big data, and intelligence, research exploring traditional or novel humanities questions with the aid of digital technology, data resources, and digital infrastructure has increased significantly, revealing new trends. First, China's digital humanities research is transitioning from cognitive exploration to digital exploration—that is, gradually shifting from thinking-type research primarily based on basic theories, review comments, and development orientations to "digital-type" research centered on practical problems and leveraging digital data resources. Second, in this "digital-type" research, there has been a shift from early engineering thinking that focused solely on data to research thinking that emphasizes "data serving humanities." Based on these characteristics and transformation trends, it is necessary to conduct in-depth review and analysis of relevant "digital-type" research in China's digital humanities. Therefore, unlike most "introductory" reviews, this paper focuses on the fundamental elements of digital humanities—"data" and "technology"—and limits its scope to relevant "authentic" academic research, aiming to reveal the true landscape of the collision between "digital" and "humanities," help "newcomers" gain an intuitive and practical understanding of such research, and provide reference for the development of new trends in domestic digital humanities.

## 1 Data Sources and Research Methods

### 1.1 Data Sources

Based on the research questions above, this study limited its data scope to domestic scientific literature that truly applies digital technology and resources to

explore humanities questions. Literature meeting the inclusion criteria needed to simultaneously possess the following characteristics: using humanities data resources or digital tools to conduct research; aiming to answer academic questions in humanities fields. To ensure comprehensive retrieval and precision, this paper adopted a combination of subject term retrieval and manual screening for literature selection. First, in the CNKI database, the search formula "SU = (digital humanities)" was used to retrieve articles, including all journal and conference articles. As of December 31, 2020, a total of 928 documents were obtained. Subsequently, by browsing titles and abstracts, literature that did not meet any of the three inclusion criteria was manually excluded, resulting in a final corpus of 147 documents. The annual distribution of these documents is shown in Figure 1 [Figure 1: see original paper]. It can be observed that "digital-type" research in digital humanities began to emerge around 2012; after 2017, the number of articles showed a rapid upward trend. The earliest study belonged to the relatively mature field of computational linguistics abroad: scholar Huang Mincong used the usage frequency of 55 basic Chinese characters in books to quantitatively analyze the evolution of simplified and traditional Chinese characters [5].

### 1.2 Research Methods

This paper combines bibliometrics, qualitative analysis, and multidimensional scaling analysis. First, based on bibliographic information, a basic descriptive analysis of relevant research was conducted, including research institution distribution, disciplinary distribution, and interdisciplinary characteristics. Then, based on literature content information, qualitative analysis methods were used to code and annotate the research content, providing a foundation for classified discussion of data and method usage. Finally, high-frequency keywords were used for co-word analysis to summarize main research types.

## 2 Basic Descriptive Analysis

### 2.1 Institutional and Disciplinary Distribution

Statistics on authors' affiliated secondary institutions are presented in Table 1 . The School of Information Management/Information Resource Research Center of Wuhan University is the primary research institution, having established a Digital Humanities Research Center in 2011. Since 2018, related research has gradually increased, focusing on image resource construction [6], semantic representation methods [7], knowledge fusion technology [8], and other directions, with projects such as Dunhuang murals and cultural heritage producing significant impact. The Shanghai Library ranks second, having conducted relevant research on genealogical digital humanities services since 2016 [8], and subsequently constructing multiple thematic knowledge bases including the Dong Qichang database, personal name authority database, and seal knowledge base, which have been widely utilized in academic research. Overall, institutional

distribution has the following characteristics: university-level disciplinary research units remain dominant, accounting for 70.78%; research institutions are relatively dispersed and balanced, demonstrating a "hundred schools of thought contending" situation, with secondary units such as information management schools, economics and management schools, and information science schools accounting for relatively high proportions; digital humanities provides an emerging pathway for library practitioners to participate in academic research, with related research accounting for 23.38%, fully reflecting the characteristic of close integration between theory and practice in digital humanities research.

Further analysis of the first-level disciplines associated with the journals publishing relevant articles reveals the disciplinary distribution of related research, as shown in Figure 2 [Figure 2: see original paper]. Overall, disciplinary distribution is diversified, involving 23 first-level disciplines covering all categories including humanities and social sciences, natural sciences, engineering and technology, and agricultural sciences, with obvious interdisciplinary characteristics. Among them, library, information, and archival science research accounts for the highest proportion, with main contributions in providing technologies, methods, and tools, focusing on technical methods for the entire process of information resource management, such as information retrieval in information acquisition [12], semantic representation in information description [7], metadata in information organization [13], data analysis in information mining, and visualization in information utilization [14-16]. This is followed by history, sociology, and literature, with related research increasing after 2017. Although the concept of digital humanities has existed for a long time, early digitalization of historical materials was mostly carried out in practice; after 2018, with the explosion of digital humanities, humanities scholars increasingly participated in research. In terms of quantity, historical research far exceeds sociology and literature, with "quantitative history" and "digital history" becoming important research branches. Among non-humanities and social science categories, computer science and technology research accounts for the highest proportion, with scholars from this field mainly contributing to overcoming technical difficulties in data resource construction and usage. Unlike library and information science, computer science focuses on more detailed and targeted technical issues, such as processing uncertain data [10] and multi-feature knowledge fusion [11].

### 2.2 Interdisciplinarity

Digital humanities has carried interdisciplinary genes since its inception. This section labels and statistics the disciplines involved in each study based on the first-level disciplines of authors' affiliated secondary units, with results shown in Figure 3 [Figure 3: see original paper]. Overall, interdisciplinary research exceeds single-disciplinary research, accounting for 58%; among interdisciplinary studies, most involve 2-3 disciplines.

An interdisciplinary network was constructed from interdisciplinary research, as shown in Figure 4 [Figure 4: see original paper]. In the figure, node size reflects

degree centrality: larger nodes represent disciplines that intersect with more other disciplines. The network involves 23 first-level disciplines with 96 pairs of disciplinary cross-relationships; each first-level discipline is connected to an average of 8 other disciplines, with relatively balanced distribution and overall close connections among disciplines.

Library, information, and archival science, information science and systems science, and computer science and technology, as representatives of the "digital" disciplines in digital humanities, occupy the network center, forming a "tripod" pattern. These disciplines are connected to many other disciplines with relatively strong ties, often playing a "technical support" role, though with different emphases. Library, information, and archival science has the largest number of associated disciplines, serving as the main provider of data resources and digital tools with permeating applications across multiple disciplines. Information science and systems science shows the strongest average disciplinary association intensity, particularly with library and information science and computer science. History, literature, and sociology, as the main representatives of "humanities" in digital humanities, are associated with many disciplines but mostly with low association intensity; obvious cross-fertilization among the three has produced numerous excellent studies [19-20]. The figure also shows several disciplines distributed on the network periphery, which can be regarded as expansion scenarios for digital humanities, fully demonstrating the development space of digital humanities across numerous disciplines. Among them, management and economics, though not having many associated disciplines, show relatively strong association intensity.

Table 2 Statistics on the number of cross-disciplines and average association intensity for each discipline.

## 3 Analysis and Discussion of "Digital" Application

### 3.1 Data Resource Usage

Digital humanities data consists of various digital files, including digitized humanities research objects and their metadata. Literature analysis shows that relevant research involves three main data sources: thematic databases, self-collected data, and comprehensive data sources.

**3.1.1 Thematic Databases** Thematic databases refer to databases established for specific topics. Here, "database" is a broad concept encompassing databases, datasets, or data platforms, generally characterized by focusing on specific themes, large data scale, high data quality, and open accessibility. Based on the nature of the leading construction institution, they mainly include thematic databases led by national government departments, those led by universities and research institutions, and enterprise-built thematic databases. From a topic selection perspective, most databases have relatively clear and distinctive disciplinary themes, usually highly related to the nature or research

scope of the sponsoring institution, primarily supporting long-term research of research teams. These include both relatively narrow-scope databases, such as the Dunhuang Studies Database by the Dunhuang Academy Digital Center, the Overseas Chinese Thematic Database by Jinan University, the Genglu Database by Hainan University Library [21], and the Zhang Xueliang Historical Materials Database by Northeastern University [22]; and relatively broad-scope databases such as the most widely used Modern Chinese Intellectual History Professional Database, the China Biographical Database (CBDB), and the Celebrity Manuscripts Archive.

Due to their high knowledge concentration, strong accessibility, excellent data quality, and large scale, thematic databases are favored by researchers, with 57% of studies involving them, including many in-depth excellent studies. For example, in the article "Toward a New World: The Formation and Evolution of the Concept of 'World' in Modern China from a Digital Humanities Perspective," scholars obtained modern Chinese political thought documents from the Modern Chinese Intellectual History Professional Database (1830-1930) and, based on text processing and mining, obtained high-quality and large-scale literature data to conduct in-depth analysis of the evolution process since the emergence of the "world" concept in 1899 [23]. This paper summarizes the frequently utilized thematic databases in Table 3 .

**3.1.2 Self-Collected Data** Self-collected data refers to data that researchers collect and organize around specific research questions. Generally, such data is small in scale, heavily colored by individual research, and mostly serves relatively limited research questions. For example: to depict and analyze the German collective portrait, scholars organized and mined the text of *Deutsche Biographie* [28]; to explore war cooperation and confrontation relationships among various vassal states during the Spring and Autumn period, scholars conducted information extraction and annotation on war texts in *Zuo Zhuan* to identify attackers and defenders in each war [29]. In a sense, the acquisition method of self-collected data is an extension of traditional humanities research data collection methods. Before the emergence of information technology and digital tools, scholars were also engaged in "extracting information from original humanities historical materials," though mostly through manual processing with smaller scalable data size and more complex management and organization. The emergence of digital tools has liberated humanities scholars from this heavy workload. From this perspective, regardless of whether digital humanities has brought a paradigm shift to humanities disciplines, it is beneficial for related research.

**3.1.3 Comprehensive Data Sources** Comprehensive data sources refer to openly accessible data resources not specifically built for digital humanities, characterized by massive data scale, relatively high quality, and broad thematic content. They can be divided into two categories: first, various literature databases such as CNKI, CSSCI, and Web of Science; second, network open resources such as Wikipedia, Baidu Baike, and various online communities (e.g.,

Douban, Friend of a Friend), which sometimes appear as auxiliary data sources in relevant research. For example, in the article "Seeking Jiangnan Delicacies: Constructing a Shanghai Food Atlas by Fusing Multi-Source Ancient and Modern Text Data," the authors combined local chronicle data from the Airusheng thematic database with modern recipe data from the "Xiachufang" website to construct a knowledge graph of Shanghai cuisine for analyzing the evolutionary relationships among entities [12].

Overall, the three types of data sources complement each other, jointly forming the data resource foundation for digital humanities research. Thematic databases have moderate data scale, good quality, and certain specificity, making them the most widely applied. Self-collected data has the smallest scale but the strongest specificity, serving single-point research of particular scholars but generally lacking openness. Comprehensive data sources have large volume, broad themes, timely updates, and easy access, but poor specificity, sometimes requiring combination with other data sources based on specific questions.

### 3.2 Main Research Types

Keywords are words or phrases extracted by article authors to summarize article themes and research content. This section extracted keywords from each article, obtaining 397 keywords total. To address the issue of different expressions for synonyms, manual standardization was performed: for example, "topic model" and "topic modeling" with similar meanings were standardized to "topic model"; "bibliometrics" and "bibliometric" were standardized to "bibliometrics"; "GIS" and "geographic information system" were standardized to "geographic information system." After processing, 378 keywords were obtained. Keywords appearing more than 5 times were defined as high-frequency keywords. The word "digital humanities," which is too closely related to the theme, was removed, and the remaining 71 high-frequency keywords formed the basis for co-word analysis, as listed in Table 4 .

A co-word matrix was generated from keyword co-occurrence frequencies, and a dissimilarity matrix was created based on this, calculated as follows:

`Dissimilarity_a,b = 1 - cooccur_a,b / sqrt(occur_a $\times$ occur_b)`

Where `Dissimilarity_a,b` represents the dissimilarity coefficient between keywords a and b, `cooccur_a,b` represents the co-occurrence count of keywords a and b, `occur_a` represents the frequency of keyword a, and `occur_b` represents the frequency of keyword b. SPSS was used for two-dimensional scaling analysis (ALSCAL), selecting a plane-symmetric graphic to describe the keyword research structure. Two-dimensional scaling analysis of the keyword dissimilarity matrix yielded the visualization result shown in Figure 5 [Figure 5: see original paper].

In the figure, the horizontal axis represents the "digital" dimension, with larger values indicating stronger "digital" characteristics in research, while the verti-

cal axis represents the "humanities" dimension, with larger values indicating more "humanities" factors. The results show keywords distributed across three regions, representing the knowledge structures of different research types, indicating the positions and relationships of different schools, fields, or academic communities in relevant research, which can be used to analyze research paths of various types [30].

**3.2.1 Resource Construction Research** Resource construction research focuses on data infrastructure development, dedicated to studying the production and organization of various digital files and constructing data resources independent of specific applications and research fields that can be long-term preserved, co-constructed, shared, and openly utilized, including databases, datasets, data platforms, and other forms. In the figure, this type of research includes 23 keywords, mainly distributed in the lower right quadrant. Some keywords relate to data organization and management, such as knowledge models, knowledge organization, and databases, involving specific research such as: constructing characteristic databases of Republican documents, local chronicles, and intangible cultural heritage based on special collections; organizing domain knowledge and vocabularies to build domain-specific metadata, ontologies, and thesauri; and utilizing interlibrary alliances for data resource fusion to provide data foundations for linking and integrating different thematic databases. From the distribution position, they are in the lower half of the "humanities" horizontal axis and the right half of the "digital" horizontal axis, with fewer humanities characteristics and increasingly strong technical attributes from left to right, such as text visualization, knowledge models, named entity recognition, and ancient text information processing. A few studies involve scholars' discrimination of specific humanities questions combined with data resources.

Resource construction research is the earliest and most mature type of research in digital humanities, also forming the foundation for subsequent research, which aligns with the research stage division by scholars such as Ke Ping [3]. This type of research has strong technical attributes, with researchers mainly concentrated in library, information, and archival science disciplines and GLAM institutions, while humanities scholars generally only provide assistance in domain knowledge and construction requirements. Additionally, this research features practice preceding theory, with many early studies first conducting practice before summarizing theory. However, as data resources surge, problems such as information silos and redundant construction caused by this approach have become increasingly prominent, with more scholars emphasizing the importance of theory and standard construction.

**3.2.2 Tool-Driven Research** Tool-driven research refers to applying relatively mature tools, models, or methods to new humanities scenarios, such as CiteSpace, topic models, etc., with a relatively large number of studies. This type includes 27 keywords. From the position perspective, they are mainly located on the left side of the "digital" vertical axis, with technical attributes

strengthening and humanities attributes weakening along the vertical axis downward, such as ontologies and digitization. Keywords are relatively evenly distributed across the "humanities" horizontal axis. Compared with resource construction research, tool-driven research overall has stronger "humanities" characteristics and deeper integration with humanities disciplines.

Due to differences in the technical strength of tools used and the degree of humanities thought penetration, the final appearance of such research varies. For example, research in the lower half of the horizontal axis involves keywords such as "CiteSpace," "CSSCI," and "ancient book digitization," with relatively simple technology and shallow humanities thought penetration, generally only providing problem scenarios. Research in the upper half involves keywords such as "spatial humanities," "computational narrative," and "geographic narrative," with stronger humanities thought penetration, not only providing usage scenarios for tools but also offering goal guidance for technical tool application, making research tools serve research questions. Notably, spatiotemporal humanities research in the upper distribution occupies a relatively large proportion. This research not only uses digital humanities historical materials but also involves extracting temporal and spatial location information, examining specific humanities questions across temporal and spatial dimensions, representing research achievements that combine both "digital" and "humanities" thinking. Examples include extracting main spatial orientation words from Tang poetry to explore changes in Tang poets' spatial cognition of paradise [19]; combining biographical data and geographic location data to explore the spatiotemporal distribution and evolution of the Ming Dynasty Jinshi group [16]; and examining the influence of socio-economic conditions and basic education on writers' geographic distribution from a spatial perspective [31].

**3.2.3 Scenario-Integrated Research** Scenario-integrated research is dominated by specific research contexts in humanities disciplines, comprehensively utilizing digital resources and data thinking to conduct in-depth analysis of particular humanities questions. This category involves 21 keywords. From the distribution position, it lies on the diagonal of the coordinate system, representing the type that achieves the most balanced and deeply integrated combination of "digital" and "humanities" among the three types.

In terms of content, this research uses more specialized technologies than the previous two types, such as knowledge graphs, conditional random fields, and artificial intelligence. Humanities thinking also penetrates more deeply, mostly not directly enjoying the fruits of data resource construction but involving detailed processing in data infrastructure construction, such as literature discrimination and stop-word processing. This reflects that in current digital humanities "digital-type" research, a research path of deep collision between "digital" and "humanities" has emerged. In such research, humanities scholars are not only users of data and tools but also builders of data infrastructure. This role transformation affects specific research processes, moving humanities scholars from

the application stage of data resources and digital technology to the construction and development stage. This full-process deep participation is the requirement of the times for deepening digital humanities research and represents an important future direction.

The typical characteristic of scenario-integrated research is that the entire research process is dominated by humanities questions and scenario needs, with humanities perspectives and data thinking constantly intertwined. Although early research mostly involved simple migration and application of data tools and methods in humanities fields, with "digital" outweighing "humanities," in recent years, with deepening research and improved information literacy of humanities scholars, data technology and symbolic attributes in the media convergence era have changed the logic of story organization in humanities disciplines to some extent. More humanities scholars have begun to focus on exploring humanities questions based on digital resources or using data thinking. Scholars from other disciplines such as information management are no longer merely "building databases for the sake of building databases" but are conducting data resource construction more in combination with humanities thinking and problem scenarios.

## 4 Reflections on Digital Humanities Research

### 4.1 Data Infrastructure Construction for Digital Humanities Research

Scientific research cannot be separated from data support. Massive, standardized, open, and linked data resources are also the basic "blood" of digital humanities research. Although databases supporting digital humanities research have been well improved, the following issues urgently need to be addressed: (1) Data isolation and redundant construction. Although thematic database construction for digital humanities is booming, the lack of unified norms and standards prevents interoperability and sharing. Only a few large thematic databases are most widely applied, while numerous databases remain unavailable or unopened, causing data isolation and resource waste on one hand, and on the other hand, the contradiction between difficult data linkage and retrieval and growing data demands leads to redundant construction of similar data. (2) Relatively single data sources, primarily text data. From research themes, research such as computational film and spatiotemporal humanities has emerged, increasing scholars' demand for non-text data, while currently available data resources mostly remain text-based. As digital humanities rapidly penetrates other fields, diversification of data resource types is an inevitable trend. However, how to handle existing unstructured data in various media forms remains a challenge, still dependent on future technological breakthroughs. (3) Insufficient development of digital tools. Data usage complexity often reduces scholars' enthusiasm [32], with most resource construction still focusing on data collection and knowledge organization and insufficient attention to knowledge services. Currently, applied tools in relevant research are mostly directly transferred from other disciplines, such as directly migrating CiteSpace and social network analysis to

humanities scenarios, lacking specialized tools for characteristic databases and research questions, which is also a reason for low data usability.

Solving these problems depends on coordinated cooperation among all parties, especially library and information science disciplines should actively play a role. As some scholars have proposed, library science is the foundational discipline for digital humanities emergence and should undertake the work of exploring digital humanities theory and practice [3]. In fact, many institutions are currently paying attention to digital humanities infrastructure construction, but most adopt independent development models, with occasional crowdsourcing models needing further promotion.

### 4.2 Digital Academic Capability of Digital Humanities Scholars

From the characteristics of humanities disciplines, their research domain features and individual colors are relatively obvious, with no universal routines, and specific research questions have vastly different requirements for data, tools, and platforms. For example, historical research requires using digital technology to extract and organize complex network relationships in large-scale historical documents, while philosophical research may focus more on the generation and evolution of concepts. Therefore, in digital humanities practice, all stages from data acquisition, selection, analysis to presentation require close participation and guidance of humanities scholars in specific contexts, naturally placing higher demands on their digital academic capabilities.

Current research status and trends also demonstrate this: resource construction and tool-driven research occupy relatively high proportions, both closely related to digital technology. Simply utilizing data resource construction achievements and relatively simple digital technology tools is no longer sufficient. From future trends, the role of "digital" for "humanities" is not merely simple tool application but has surpassed data resources, tools, and even methods [4]. Scenario-integrated research places even higher demands on humanities scholars' information literacy, requiring not only their participation in data resource construction and technology tool development but also comprehensive improvement in their data thinking and data literacy, enabling flexible use of data and tools according to problem contexts and research processes to achieve deep integration of "digital" and "humanities."

### 4.3 The Role Transformation of "Humanities" in Digital Humanities

As mentioned above, the digital humanities boom has sparked scholars' cold reflection. Discussions and disagreements about the relative importance of "digital" versus "humanities" have persisted. What is the positioning of humanities in digital humanities? What are the advantages and disadvantages of digital humanities for humanities development? Such questions emerge continuously. It has become almost consensus that digital humanities cannot replace traditional humanities. Digital humanities is an extension, not a substitute, for humanities

disciplines, providing humanities scholars with convenient data resources and digital tools that partially liberate them from heavy data finding and organization work. More importantly, it provides humanities scholars with broader speculative space. As digital humanities develops and scholars' digital literacy improves, they can better use digital technology to feed back into traditional humanities, truly achieving bidirectional symbiosis between "digital" and "humanities." Of course, the popularity of digital humanities and its considerable achievements may inevitably reduce humanities scholars' attention to other issues to some extent, negatively affecting the speculative nature and depth of humanities and causing the phenomenon of "emphasizing digital, neglecting humanities." Ultimately, digital humanities remains a field dominated by humanities characteristics, with the term "digital" having obvious tool attributes (though not merely tools). Tools themselves are neither good nor bad; their effects depend on how users apply them. If users have "data-only" thinking misconceptions or lack data capabilities, the results will likely be more harmful than beneficial. However, throwing out the baby with the bathwater is not advisable; the method to reduce negative impacts should be improving users' capabilities rather than abandoning tools. Moreover, the "digital" transformation of humanities disciplines is also a future trend, because fundamentally, the emergence of digital humanities does not come from the radiation of powerful disciplines like computer science and information science on humanities but from changes in the research infrastructure and academic environment of humanities disciplines themselves [34].

From resource construction and tool-driven to scenario-integrated research, the collision between "digital" and "humanities" is continuously deepening. This process will inevitably bring evolution to various research paths. Humanities scholars need to increase their participation in different types of research, which not only benefits the development of these two types of research but can also indirectly elevate integrated digital humanities research to new heights through improved data resource quality, better applicability of technical tools, and enhanced scholars' digital academic capabilities. Therefore, humanities scholars should maintain critical awareness and vigilance while embracing new opportunities and challenges, seeking balance between humanistic spirit and data thinking.

## 5 Conclusion

This paper takes "digital-type" research in digital humanities as its object, exploring the intersection and integration of "digital" and "humanities" in digital humanities research. From research overview to "digital" application and main research types, and then to deep reflection, we have reached some conclusions and insights: (1) Library and information science, history and literature, and computer science are the main research institutions, gradually penetrating into sociology, economics, and other fields, demonstrating strong interdisciplinarity. There is no obvious "dominance by one school" phenomenon; instead, distribu-

tion is balanced with diverse schools of thought. Various disciplines and institutions have broad space in related research, so utilizing their own characteristics and advantages to build disciplinary discourse in digital humanities currently presents a "window of opportunity." (2) A few large thematic databases and researcher self-collected data are the main data sources, while most thematic databases have low utilization rates. On one hand, this is due to poor data quality and low accessibility; on the other hand, it is also limited by researchers' data literacy. In the future, attempts can be made to conduct top-down reorganization, integration, and reuse of various "idle" resources through institutional or disciplinary alliances, while enhancing user awareness and paying attention to researchers' opinions to make thematic databases truly serve academic research. (3) Related research mainly falls into three categories: resource construction, tool-driven, and scenario-integrated research. Scenario-integrated research emerged relatively late with smaller quantity, but it achieves better balance and deep integration between "digital" and "humanities." From the binary scale analysis results, its boundary with tool-driven research is gradually approaching. With accumulated research and deepening studies, scenario-integrated research is expected to become the future trend of tool-driven research and influence the direction of resource construction research. From the researcher perspective, thinking transformation is needed, with scenario awareness as the guide, cleverly mining data and appropriately using tools to seize opportunities in the new trend and improve the quality of digital humanities research.

Of course, this paper also has limitations: Using only literature as the data source may not cover all relevant research, especially thematic databases and industry reports established by social institutions. There may be some subjectivity in literature selection and coding annotation, which may cause some bias in results. Therefore, future research could consider expanding to multi-source data analysis and further increasing reviews of foreign literature to conduct comparative analysis of domestic and international digital humanities development, providing references for digital humanities and related disciplinary development.

## References

[1] KIRSCHENBAUM M G. What is digital humanities and what's it doing in English departments?[J]. ADE bulletin, 2010(120): 55-61.

[2] Li Zhenxi, Jian Wu. Frontier exploration of the informationization process—"Digitalization and 21st century humanistic spirit" seminar summary[J]. Exploration and Free Views, 2003(2): 34-36.

[3] Ke Ping, Gong Ping. Analysis of the evolution path and hot fields of digital humanities research[J]. Journal of Library Science in China, 2016, 42(6): 13-30.

[4] Wang Lihua, Liu Wei, Liu Shengying. Prospects for the theoretical trend of digital humanities[J]. Journal of Library Science in China, 2020, 46(3): 17-23.

[5] Huang Mincong. Quantitative analysis of the evolution of simplified and traditional Chinese characters based on humanities computing[J]. Science & Technology Vision, 2012(18): 64-67.

[6] Zeng Ziming, Zhou Zhi. Research on image semantic description model for digital humanities[J]. Information Studies: Theory & Application, 2018, 41(1): 116-121.

[7] Song Ningyuan, Wang Xiaoguang. Research on semantic structured representation methods for narrative texts based on plot ontology[J]. Journal of Library Science in China, 2020, 46(2): 96-113.

[8] 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.

[9] Tong Yin. Research on semantic knowledge model construction and linked data—Dong Qichang digital humanities data mapping report[J]. Public Communication of Science & Technology, 2020, 12(5): 135-137.

[10] Yang Jiaying, Deng Luxiang, Xu Xin. Seeking Jiangnan delicacies: Constructing a Shanghai food atlas by fusing multi-source ancient and modern text data[J]. Library Tribune, 2020, 40(10): 24-33.

[11] Wang Xiaoguang, Hou Xilong, Cheng Hanghang, et al. Construction and linked data publishing of Dunhuang mural thesaurus[J]. Journal of Library Science in China, 2020, 46(4): 69-84.

[12] Ouyang Jian. Large-scale ancient book text visualization analysis and mining for digital humanities research[J]. Journal of Library Science in China, 2016, 42(2): 66-80.

[13] Yang Haici, Wang Jun. Construction and visualization of Song Dynasty academic apprenticeship knowledge graph[J]. Data Analysis and Knowledge Discovery, 2019, 3(6): 109-116.

[14] Deng Jun, Sun Shaodan, Wang Ruan, et al. Spatiotemporal network structure analysis of the Ming Dynasty imperial examination Jinshi group from a digital humanities perspective[J]. Library and Information Service, 2020, 64(17): 4-17.

[15] Guo Linfei, Liu Guangzhong. Research on uncertain data processing technology based on Neo4j[J]. Computer Technology and Development, 2020, 30(1): 25-31.

[16] Liang Jiwen, Jiang Chuan, Wang Dongbo. Research on Chinese-English sentence alignment of pre-Qin classics based on multi-feature fusion[J]. Data Analysis and Knowledge Discovery, 2020, 4(9): 123-132.

[17] Qiu Weiyun, Yan Cheng. Heart in paradise, gazing at the world: Digital humanities research on spatial orientation in mid-Tang poetry[J]. Journal of

Southwest Minzu University (Humanities and Social Sciences), 2020, 41(8): 161-171.

[18] Qiu Weiyun, Jin Guantao, Liu Qingfeng, et al. Digital humanities research on the formation of modern Chinese equality concepts: Centered on newspapers and periodicals[J]. Journal of Tsinghua University (Philosophy and Social Sciences), 2018, 33(6): 1-33.

[19] Li Wenhua, Chen Hong, Li Dongrui. Comprehensive research on the South China Sea Genglu Bu from a digital humanities perspective[J]. Journal of Academic Libraries, 2020, 38(2): 91-98.

[20] Yao Tianhong, Chen Yanmei, Liu Ge, et al. Semantic knowledge organization of digital humanities historical resources based on CIDOC-CRM: Taking Zhang Xueliang historical materials as an example[J]. Journal of Library Science, 2019, 41(7): 35-43.

[21] Qiu Weiyun, Zheng Wenhui. Toward a new world: The formation and evolution of the concept of "world" in modern China from a digital humanities perspective[J]. Nanjing University Journal (Philosophy, Humanities and Social Sciences), 2020, 57(5): 88-106.

[22] Wang Tao. Collective portrait depiction and type analysis of Deutsche Biographie under the digital humanities framework[J]. Historical Research, 2018(5): 148-166.

[23] Shanghai Library Personal Name Authority Database[EB/OL]. [2021-02-03]. http://names.library.sh.cn/mrgf/service/person/list.

[24] Han Dian Ancient Books Corpus[EB/OL]. [2021-02-03]. http://gj.zdic.net/.

[25] National Index to Chinese Newspapers and Periodicals Database[EB/OL]. [2021-02-03]. https://www.cnbksy.com/search/advance.

[26] UNIVERSITY H, SINICA A, UNIVERSITY P. China Biographical Database (CBDB)[EB/OL]. [2021-02-03]. https://projects.iq.harvard.edu/cbdb.

[27] Fan Wenjie, Li Zhongkai, Huang Shuiqing. Quantitative and visualization research on wars in Zuo Zhuan based on social network analysis[J]. Library and Information Service, 2020, 64(6): 90-99.

[28] Zhang Qin, Ma Feicheng. Foreign knowledge management research paradigms: A co-word analysis approach[J]. Journal of Management Sciences in China, 2007(6): 65-75.

[29] Peng Xia, Liu Min, Yang Li, et al. Geographic distribution and causes of Ming-Qing female writers from a spatial perspective: Taking Songjiang Prefecture as an example[J]. Journal of Library and Information Science in Agriculture, 2020, 32(9): 31-38.

[30] Zhao Xueqin, Mo Changlei, Li Tian'e, et al. Factors influencing humanities scholars' acceptance intention of digital humanities: An investigation centered

on historians[J]. Library Tribune, 2020, 40(4): 40-49.

[31] Chen Haizhu, Bao Ping. Digital humanities as an innovative paradigm for humanities research: A literature review based on the application and penetration of digital humanities in historical research[J]. Library Theory and Practice, 2017(3): 11-16.

**Author Contributions:** Shi Jing was primarily responsible for data analysis and content writing; Li Yang was primarily responsible for logical organization and content revision.

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