Knowledge Graph Construction and Analysis in Digital Humanities: An Empirical Postprint Based on Keywords and Citation Contexts from WoS Literature

Authors: Xu Xin, Chen Luyao, Yang Jiaying

Date: 2023-07-26T00:00:00+00:00

Abstract

[Objective/Significance] Citations serve as the nexus between citing literature and cited works, reflecting the intellectual borrowing and validation of subsequent researchers. Building upon traditional bibliographic keyword networks, this study innovatively employs citation context keywords as research material. The constructed knowledge graph not only reveals deep-level thematic information within documents but also reflects the knowledge process by which audiences subjectively select and utilize literature. [Method/Process] With digital humanities selected as the research domain, this study obtains three document corpora and two citation text corpora, constructing two undirected keyword co-occurrence networks and two directed literature citation-based keyword networks. Through co-occurrence networks, it observes the absorption and diffusion of knowledge within the digital humanities field; through citation keyword networks, it observes the formation and transformation of digital humanities. [Results/Conclusion] The study reveals the research priorities, core domains, and core technologies of digital humanities, providing reference and guidance for future research in the digital humanities field from the audience' s perspective.

Full Text

Preamble

Knowledge Mapping in Digital Humanities Research: An Empirical Study Based on Keywords and Citation Contexts in WoS Literature

Xu Xin, Chen Luyao, Yang Jiaying

Department of Information Management, School of Economics and Management, East China Normal University, Shanghai 200241

Abstract

[**Purpose/Significance**] Citations serve as the link between citing and cited literature, reflecting the intellectual borrowing and affirmation of subsequent researchers. Building upon traditional keyword networks derived from bibliographic records, this study innovatively incorporates keywords from citation contexts as research material. The resulting knowledge map not only reveals deeper thematic information within the literature but also reflects the knowledge-based process by which audiences actively select and utilize documents. [**Method/Process**] Focusing on digital humanities as the research domain, we collected three literature sets and two citation text sets to construct two undirected keyword co-occurrence networks and two directed keyword networks based on citation relationships. Through co-occurrence networks, we observed knowledge absorption and diffusion in digital humanities; through citation keyword networks, we examined the formation and transformation of the field. [**Result/Conclusion**] The study identifies key research priorities, core domains, and foundational technologies in digital humanities, providing audience-based insights for future research in the field.

Classification Number: G253
Keywords: knowledge mapping, digital humanities, citation context, keyword network, visualization

2 Literature Review

In library and information science, knowledge mapping is defined as a domain knowledge map that displays the development process and structural relationships of scientific knowledge, facilitating knowledge discovery. It represents a network-based visualization method that maps abstract scientific information into spatial structures and graphics under a given theme. Knowledge networks enable knowledge creation and transfer, particularly for domain-specific networks that reflect internal knowledge flow and dissemination. However, traditional knowledge network research primarily relies on author-subjective information such as titles, abstracts, and keywords, reflecting knowledge transmitted from the creator' s perspective. This approach cannot capture the actual information obtained from the audience' s viewpoint. Research on citation texts—through which audiences actively read, filter, select, and utilize literature—can uncover latent knowledge difficult to detect in explicit information, thereby fully leveraging the value and function of citation content.

The predecessor of digital humanities was humanities computing, which emphasized the application of computer science to traditional humanities disciplines. With the advent of the information age and the proliferation of digital technologies, humanities scholars have increasingly adopted digital thinking to solve

humanities problems, gradually shifting the research focus from narrow computational thinking to the rich implications of digitization, and from methodological innovation to the humanities content itself, thus giving rise to the concept of digital humanities. Digital humanities development is inseparable from its historical context and has brought qualitative transformations to humanities research. S. Schreibman et al. reviewed early digital humanities research history, covering archaeology, art, literature, music, performance, and multimedia. Research content continues to expand, extending from humanities computing to various digital technologies including data storage, data organization, and visual analysis. For example, D. Cooper extracted and encoded place names or spatial-related textual content for map-based presentation, while U. Hinrichs constructed collections of objects, word clouds, symbols, and timelines from science fiction novels for visualization. Librarians and information scientists serve as core contributors to digital humanities, contributing to teaching services, user engagement, talent cultivation, and resource development.

In terms of research achievements in digital humanities, numerous studies employ scientometric methods to summarize the field. Domestically, scholars primarily use tools like CiteSpace to construct network maps from keywords, institutions, authors, and journals in digital humanities literature, identifying hot topics and analyzing evolutionary paths. This approach has gained international recognition. However, such studies typically focus on a single relationship type and rely on explicit domain knowledge without excavating deeper latent knowledge from the audience perspective. Current research only attends to explicit information transmitted by creators while neglecting potential information embedded in citation texts that citing parties extract through reading and refinement. Therefore, digital humanities urgently requires new approaches to systematically review and grasp its historical research trajectory.

Scientific knowledge mapping represents the concept of scientometric knowledge graphs. In knowledge network prediction, studying citation networks, collaboration networks, or bipartite networks formed by separated sets enables the identification of connections among knowledge units. Scientific knowledge mapping is widely applied to the production, presentation, and dissemination of scientific information and knowledge. At the word level, scholars primarily utilize author-assigned keywords for co-occurrence, clustering, or coupling analysis to identify research frontiers and hotspots. Song Yanhui et al. employed author keyword coupling and compared it with conventional document-based methods, finding that keyword-based approaches yield more intuitive and clear research content. Keywords not only associate more closely with content but also reflect knowledge absorption and diffusion. Zhang Lingling et al. distilled thematic keywords to trace knowledge diffusion directions and contexts. Luo Shuangling et al. argued that citation keyword aggregation forms community structures, referring to high-frequency title keyword networks as "thematic communities."

Citation represents an active behavior by audiences after reading original literature. Citation texts are created by citers to summarize cited literature themes,

with citation contexts as extensions that reveal deeper cited content, thus possessing significant research value. Citation contexts represent the borrowing and affirmation of cited achievements by subsequent researchers, establishing a cognitive relationship between citing and cited documents. Y. Liu et al. demonstrated that both article word usage and citation word usage reflect knowledge integration influences within frameworks of knowledge fusion and diffusion. Currently, few studies analyze citation contexts through network approaches. L. Bornmann et al. conducted co-occurrence network analysis on citation contexts of works by the scientometrician E. Garfield, finding that citation contexts better reflect cited literature content than citing article titles and abstracts.

Under the open innovation paradigm, the proliferation of network mapping tools has promoted continuous attention to and expansion of knowledge mapping research routes. Chen Chaomei' s Java-based CiteSpace software is particularly renowned in China. Using this tool, Xiao Ming constructed network maps from CiteSpace literature keywords, institutions, authors, and journals, though analysis remained at the explicit information level. Scholars domestically and internationally primarily employ knowledge network methods to trace research trajectories and detect hotspots in various domains, especially emerging fields, with digital humanities being a significant application area.

3 Research Design

3.1 Data Acquisition and Definition

This study uses the Web of Science Core Collection as the primary data source, constructing three literature sets (core literature set, reference literature set, citation text set) and obtaining full texts to establish two citation text sets (core literature citation text set, citing literature citation text set). During full-text acquisition, resources from various full-text databases and internet search engines were also consulted to ensure data completeness.

The core literature set is defined as 768 documents retrieved using "digital humanities" as a keyword in the topic field (as of January 2018). The citing literature set comprises 1,100 citing records corresponding to the core literature, downloadable through WoS cited reference links. The reference literature set includes reference information from the 20 most highly cited core literature documents, totaling 956 records.

After acquiring the three literature sets, we extracted original author-assigned keywords from each document and downloaded full-text papers based on each set' s records. The citation text set includes the core literature citation text set (citation texts where core literature cites references) and the citing literature citation text set (citation texts where citing literature cites core literature). We extracted relevant citation context texts through reference lists and citation identifier information.

3.2 Identification of Citation Context Keywords

Citation context is fundamental to citation content analysis, and its identification and application have become research hotspots. In A. Bader' s citation research, different citation window lengths were tested: 10 words, 30 words, 50 words, and entire citation sentences (using the complete sentence containing the citation identifier, bounded by punctuation) before and after the citation identifier. Results demonstrated that citation contexts of 50 words best represent cited literature content, corroborating earlier findings by S. Bradshaw. Therefore, this study selects citation context texts containing 50 words around the citation identifier.

Specifically, when extracting 50-word windows around citation identifiers, we followed these rules: (1) Considering document structure, citation context extraction must remain within the same paragraph. (2) Each citation context can contain only one citation identifier. When a citation sentence contains multiple identifiers, except when identifiers appear simultaneously or are connected by "and," the citation context length must be shortened by sentence boundary segmentation. (3) For multiple citations of the same document, multiple citation context texts are retained. Using the citation identifier as the unique identifier, one citation context is extracted per identifier occurrence.

After extracting citation context texts, we identified keywords using the following method: First, all citation contexts were treated as a whole for LDA (Latent Dirichlet Allocation) topic identification. LDA is a document topic generation model, also known as a three-layer Bayesian probability model comprising word, topic, and document layers. By calling Python' s sklearn module and setting parameters to ignore common words appearing in 50% of context documents (max_{df}=0.5), we identified the top 10 topics and their top 10 keywords to understand primary research themes and directions in citation texts. Second, all citation texts were segmented and frequency-statistics were performed. Using original keywords from the core literature set, we built a custom dictionary and manually compiled stopword and synonym replacement lists. Based on these three lists, we processed the raw frequency statistics to generate a word weight table, removing stopwords, merging synonyms, and increasing custom word weights. Finally, using this weight table, we extracted up to 5 highest-weight words from each citation text as citation context keywords.

3.3 Research Scheme

Knowledge networks help clarify relationships among documents by extracting information and knowledge units to understand knowledge structure and evolution. Existing word co-occurrence networks primarily reflect research themes and inter-theme relationships to identify current hotspots and topic clusters. Most scientific citation networks use documents or authors as nodes and citation relationships as edges to construct networks among cited literature. Through such networks, we can understand scientific knowledge dissemination

and flow, discover inheritance-development or transformation-innovation relationships, and study the trajectory and structure of domain knowledge development.

Keywords provide more intuitive revelation of document themes and represent condensed document content. Drawing on the above methods, we constructed keyword citation relationship networks to reflect how the literature context constructed by one keyword cites another keyword' s context. We define this as a keyword network based on document citation, abbreviated as citation keyword network.

Specifically, we examine digital humanities from two perspectives: First, constructing keyword co-occurrence networks to understand knowledge structure; second, constructing citation context keyword networks and cited literature keyword networks based on citation relationships to understand knowledge context.

In keyword co-occurrence networks, we built knowledge absorption and diffusion networks for digital humanities core literature keywords through merging keyword sets. Merging means connecting identical keywords from different sets while preserving word co-occurrence relationships. The knowledge absorption network merges the reference literature set, core literature citation text set, and core literature set keyword co-occurrence networks. The knowledge diffusion network merges the core literature set, citing literature citation text set, and citing literature set keyword co-occurrence networks. This transformation of dynamic knowledge evolution into static networks enables analysis of core knowledge absorption and diffusion in digital humanities.

In citation-based keyword networks, we consider relationships between core and citing literature, and between highly cited core literature and references. In the relationship between core and citing literature sets, cited keywords are original keywords from the core literature set while citing keywords are citation text keywords from citing literature. In the relationship between highly cited core literature and reference sets, cited keywords are original keywords from the reference literature set while citing keywords are citation text keywords from core literature citing references.

4 Knowledge Networks Based on Core Literature Keywords

4.1 Knowledge Absorption Network

Keywords in the core literature set represent core knowledge in the current digital humanities field, which can be considered as derived from citing and transforming core knowledge in the reference literature set. Based on this, merging keywords from the reference literature set, core literature citation text set, and core literature set through co-occurrence relationships allows comprehensive observation of knowledge absorption processes and outcomes for digital humanities core literature keywords.

[Figure 1: see original paper] shows the merged keyword co-occurrence network, containing 2,968 distinct keywords forming 13,069 co-occurrence pairs. The network reveals that digital humanities knowledge absorption and formation cluster around two core keyword groups, with some independent small networks at the periphery.

In the core network, the digital humanities concept shows strongest associations with humanities disciplines, education, and digital history, indicating these are important disciplines in the key knowledge absorption process. It also shows strongest associations with technical approaches like text mining, data visualization, and linked data, suggesting these technologies exert the greatest influence. Additionally, concepts such as humanities computing, GIS, distant reading, digital libraries, social media, and big data are highly relevant, indicating these related knowledge areas influence digital humanities knowledge absorption and formation. At the core network' s edge, two small branches exist: one is a social network knowledge structure formed by Twitter and microblogging; the other is a process-oriented knowledge structure formed by digitization, text, libraries, and innovation.

Beyond the core network, several independent small networks exist: (1) A small network of literature, art, history, and maps reflecting interdisciplinary historical research based on mapping technology; (2) A small network of topic models, topics, and trees reflecting tree-based topic models as a relatively important technical method in early digital humanities; (3) A strong relationship between cultural heritage and serious games, highly independent in the overall network, representing a small branch dedicated to displaying and disseminating cultural heritage through serious game models; (4) A relationship between journals and citations, primarily for journal evaluation based on citation data and metrics.

The h-strength of the merged knowledge absorption network is 48, meaning at least 48 connections in the network have strength $ $48. Refining the network through h-strength yields the h-subnet shown in [Figure 2: see original paper], including connections with strength $ $48 and their associated nodes. The h-subnet forms a core network centered on digital humanities concepts, with other independent small networks. In this core network, library-related knowledge is particularly prominent, with concepts like embedded librarianship, digital libraries, academic libraries, close reading, distant reading, and open access all present, indicating library-based research receives considerable attention in current digital humanities. A small network branch extends from humanities disciplines, digitization, technology, innovation, data, and tools, showing the intrinsic connection between digital humanities and humanities disciplines. On the social media branch, Twitter remains the most studied platform, with altmetrics emerging as a new concept influencing the scientometrics field.

4.2 Knowledge Diffusion Network

Core knowledge from the core literature set influences subsequent knowledge formation during dissemination, representing a knowledge diffusion process. Merging keywords from the core literature set, citing literature citation text set, and citing literature set through co-occurrence relationships enables comprehensive observation of knowledge diffusion processes and outcomes for digital humanities core literature keywords.

[Figure 3: see original paper] shows the merged keyword co-occurrence network, containing 3,790 distinct keywords forming 60,366 co-occurrence pairs. The network reveals that digital humanities knowledge diffusion and dissemination primarily radiate from one keyword cluster, with some small knowledge networks emerging.

The h-strength of the knowledge diffusion network is 42, meaning at least 42 connections have strength $ $42. Refining through h-strength yields the h-subnet shown in [Figure 4: see original paper]. In this subnet, digital humanities core knowledge diffusion centers on the digital humanities concept, forming a radial pattern. Library-related knowledge is particularly prominent, with concepts like embedded librarianship, digital libraries, academic libraries, close reading, distant reading, and open access displayed. A small network branch comprises humanities disciplines, digitization, technology, innovation, data, and tools, indicating the intrinsic connection between digital humanities and humanities disciplines. On the social media branch, Twitter remains the most studied platform, with altmetrics emerging as a new concept.

4.3 Comparison Between Knowledge Absorption and Diffusion Networks

Comparing quantitative characteristics of both networks, Table 1 shows that the knowledge diffusion network is larger than the knowledge absorption network in overall scale, including node count, connection count, and network density. This indicates digital humanities knowledge diffusion distributes knowledge points more widely with stronger overall interconnections. In h-subnets, the knowledge absorption network contains more high-strength knowledge nodes.

Comparing h-subnet content, [Figure 5: see original paper] shows the node sets of both networks have substantial intersections. These represent key knowledge in digital humanities absorption and diffusion processes. Beyond the digital humanities concept, humanities computing and GIS-related knowledge are closely related to digital humanities. Library-related knowledge includes digital libraries, archives, and distant reading. Discipline-related knowledge includes history, digital history, humanities, and education. Data digitization knowledge includes digitization, data visualization, and big data. Cultural heritage and Twitter/social media are research foci. Ontology, linked data, corpus linguistics, and text mining are key technologies.

Differences show that knowledge absorption focuses more on mapping technology, tree-based topic models, and implicit information, while knowledge diffusion emphasizes information infrastructure, information literacy, embedded librarianship, altmetrics, open access, close reading, sustainability, and disciplinary differences.

5 Citation-Based Digital Humanities Knowledge Networks

5.1 Citation Keyword Network Between Highly Cited Core Literature and References

When core literature cites references, citation text keywords cite reference original keywords. If core literature A cites reference B (original keywords: b1, b2, b3) and citation context C is identified in A' s full text (citation text keywords: c1, c2, c3), then words c1, c2, c3 each cite words b1, b2, b3, forming 9 citation pairs: (c1,b1)(c1,b2)(c1,b3)(c2,b1)(c2,b2)(c2,b3)(c3,b1)(c3,b2)(c3,b3). After counting and merging identical pairs, the values represent relationship strength.

Based on word nodes and relationships, we constructed the citation network between highly cited core literature and references. [Figure 6: see original paper] shows high-strength nodes and relationships. Twitter, topic model, humanities, and digital are core nodes, indicating these are important in digital humanities highly cited knowledge formation. In direction, research follows the digital humanities path to digitize humanities; in methodology, topic identification is crucial; as a medium, Twitter is an important platform for international digital humanities research.

In the humanities-digitization cluster, digital libraries, communities, computers, archives, GIS, and journals were key early research objects. In the topic identification cluster, structures and tree algorithms were methodological foci. In the Twitter cluster, user, tool, and social information show digital humanities' attention to social platform research—primarily content-based thematic analysis and retweeting behavior studies, technical details difficult to mine through bibliographic keyword networks.

Reference set keywords supplemented by citation text keywords reveal latent paper details. For example, Twitter and topic model are reference keywords, but adding citation text keywords reveals hidden path associations like user and retweeting, further verifying that social platform digital humanities research often uses users as bridges.

5.2 Citation Keyword Network Between Core Literature and Citing Literature

From core literature, we extracted 1,220 keywords (actually 757 words/phrases). From citing text keywords, we extracted 1,220 keywords (actually 383

words/phrases), with 113 also belonging to cited literature keywords. Cited and citing words generated 5,508 citation pairs.

First, analyzing composition shows knowledge points condensed from 757 to 383 words/phrases during citation, making research foci more concentrated. Table 2 shows citation pairs with relationship strength >30. In cited words, digital humanities appears frequently, indicating explicit research around the concept resonates with scholars, focusing on data, digitization, and techniques. In citing words, Twitter appears most frequently, indicating audience perspective shows substantial valuable information about Twitter being transmitted.

High-frequency co-citation pairs include journalism, computational journalism, archives, journals, technology, GIS, data, and scholarly communication. In 5,508 pairs, 61 are self-citations (same word citing itself), representing about 16% of the 384 unique citing words, indicating at least this proportion of knowledge maintains direct transmission. Table 3 shows the top 10 self-citation pairs, with Twitter and digital humanities having the highest strength. At the macro level, knowledge around digital humanities continues to propagate under this concept; at the micro level, Twitter-specific research also spreads Twitter as a special research object.

In the citation network, node degree centrality reflects status and influence. Average degree is 3.2, indicating each word node connects with ~3 others; average weighted degree is 7.2, indicating each node has ~7 citation connections on average.

Exploring the core citation network by thresholding connection weights, [Figure 8: see original paper] shows the network with edge weights $ $20, forming two independent clusters. The left cluster centers on journalism and digital humanities, connecting both sides. Journalism primarily cites other knowledge, indicating it' s a major hotspot integrating data, technology, and culture in computational journalism, new institutionalism, ethnography, political economy, and journalism sociology. Digital humanities is primarily cited by others, showing much research explicitly explores the concept through data information, digitization, visualization, technology, and history.

The right cluster focuses on Twitter, primarily integrating information from other keywords including social networks, user studies, altmetrics, webometrics, digital communication systems, and conferences. Disciplinary difference research cites Twitter studies as representative of social platforms for exploring new academic communication and metrics.

In directed networks, out-degree counts connections from a node; in-degree counts connections to a node. In digital humanities citation networks, in-degree indicates cited keyword popularity and main transmitted knowledge; out-degree indicates citing scholars' concerns. Table 4 shows top in-degree and out-degree keywords. In-degree rankings show digital humanities concepts, history/digital history, libraries/archives, and technology (including text mining) are most widely disseminated. Out-degree rankings show scholars similarly focus on dig-

ital humanities concepts, history, and technology, but emphasize digitization, digital approaches, and differences brought by new methods.

Weighted centrality considers connection weights, reflecting transmission depth. Table 5 shows weighted in-degree and out-degree rankings. Culture, data, and Twitter show inherited depth; internet-based emerging terms like altmetrics and webometrics show deep propagation within small ranges. Academic communication and disciplinary differences also receive attention. Weighted out-degree adds focus on journalism and visualization.

References

[1] Jiao Xiaojing, Wang Lancheng. Research on concept differentiation and disciplinary positioning of knowledge mapping[J]. Library and Information Service, 2015, 59(15): 5-11.

[2] Li Qihu, Yin Li, Zhang Quan. Humanities computing in the information age[J]. Science, 2015, 67(1): 35-39, 4.

[3] Chen Yue, Liu Zeyuan. The quietly rising scientific knowledge mapping[J]. Studies in Science of Science, 2005, 23(2): 149-154.

[4] Zhang Bin, Ma Feicheng. Review of link prediction research in scientific knowledge networks[J]. Journal of Library Science in China, 2015, 41(3): 99-113.

[5] Zheng Yanning, Xu Xiaoyang, Liu Zhihui. Research on research frontier identification method based on keyword co-occurrence[J]. Library and Information Service, 2016, 60(4): 1-8.

[6] Wu Xiaoqiu, Lü Na. Research on hotspot analysis method based on keyword co-occurrence frequency[J]. Information Studies: Theory & Application, 2012, 35(8): 115-119.

[7] Song Yanhui, Wu Yishan. Comparative study of author bibliographic coupling analysis and author keyword coupling analysis: empirical analysis of Scientometrics[J]. Journal of Library Science in China, 2014, 40(1): 25-38.

[8] Zhang Lingling, Zhang Yu' e, Du Li. Research on knowledge diffusion in library and information science from the perspective of National Social Science Fund project outcomes[J]. Library Work and Study, 2017, 1(10): 60-67.

[9] Luo Shuangling, Zhang Wenqi, Xia Haoxiang. Disciplinary thematic evolution analysis based on semi-accumulated citation network community detection: taking "cooperation evolution" as an example[J]. Journal of Intelligence, 2017, 36(1): 100-110.

[10] Liu Yang, Cui Lei. Research on information value of citation context in document content analysis[J]. Library and Information Service, 2014, 58(6): 101-104.

[11] LIU Y, RAFOLS I, ROUSSEAU R. A framework for knowledge integration and diffusion[J]. Journal of Documentation, 2012, 68(1): 31-44.

[12] BORNMANN L, HAUNSCHILD R, HUG S E. Visualizing the context of citations referencing papers published by Eugene Garfield: a new type of keyword co-occurrence analysis[J]. Scientometrics, 2018, 114(2): 427-437.

[13] CHEN C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.

[14] Xiao Ming, Chen Jiayong, Li Guojun. Visual analysis of scientific knowledge mapping based on CiteSpace[J]. Library and Information Service, 2011, 55(6): 91-95.

[15] Fan Yunman, Ma Jianxia, Zeng Su. Research status of emerging topics based on knowledge mapping[J]. Information Science, 2013(9): 88-94.

[16] PHELPS C, HEIDL R, WADHWA A. Knowledge, networks, and knowledge networks: a review and research agenda[J]. Journal of Management, 2012, 38(4): 1115-1166.

[17] SCHREIBMAN S, SIEMENS R, UNSWORTH J. A companion to digital humanities[M]. New Jersey: John Wiley & Sons, 2008: 20-30.

[18] COOPER D, GREGORY IN. Mapping the English Lake District: a literary GIS[J]. Transactions of the Institute of British Geographers, 2015, 36(1): 89-108.

[19] HINRICHS U, FORLINI S, MOYNIHAN B. Speculative practices: utilizing InfoVis to explore untapped literary collections[J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 22(1): 429-438.

[20] WONG S H R. Digital humanities: what can libraries offer?[J]. Portal: Libraries and the Academy, 2016, 16(4): 669-690.

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

[22] Gao Shenghan, Zhao Yuxiang, Zhu Qinghua. Analysis of research progress in digital humanities at home and abroad[J]. Library Journal, 2016, 35(10): 9-18.

[23] WANG Q. Distribution features and intellectual structures of digital humanities: a bibliometric analysis[J]. Journal of Documentation, 2018, 74(1): 223-246.

[24] ALJABER B, STOKES N, BAILEY J, et al. Document clustering of scientific texts using citation contexts[J]. Information Retrieval, 2010, 13(2): 101-131.

[25] BRADSHAW S. Reference directed indexing: redeeming relevance for subject search in citation indexes[M]//Research and advanced technology for digital libraries. Berlin: Springer, 2003: 499-510.

Author Contributions:

Xu Xin: Responsible for topic selection and research design
Chen Luyao: Responsible for data collection, analysis, and initial draft
Yang Jiaying: Responsible for revisions and final manuscript

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