Scientific and Social Research, 2025, Volume 7, Issue 5
https://ojs.bbwpublisher.com/index.php/SSR
Online ISSN: 2981-9946
Print ISSN: 2661-4332

Data from China: Analysis of Research Hotspots in the Field of Digital Therapeutics in the Chinese Context (1989–2021)

Jike Bai[1]*, Oriana Luc[2], Shaobing Liao[1], Qige Qi[3], Jing Jin[4]*, Xuan Liu[5]*

[1]Dongguan Seventh People's Hospital (Dongguan Mental Health Centre), Dongguan 523000, China
[2]Saint John of NB, Saint John E2K 5G4, Canada
[3]The Third People's Hospital of Longgang District Shenzhen, Shenzhen 518000, China
[4]Innovation Center of Social Technology for Aging of the National Innovation Center par Excellence (NICE), Shanghai, China
[5]NERVTEX Co., Ltd., Seoul, Republic of Korea

**Authors to whom correspondence should be addressed.*

Abstract: *Objective:* China is globally leading in digitalization. However, due to language barriers, there are few international studies on digital medicine based on the Chinese context. This paper focuses on Chinese data to explore the distribution and characteristics of research hotspots in the field of digital therapy in China. *Methods:* By quantifying the published literature on digital therapy from 1989 to 2021 in the authoritative Chinese database CNKI, keyword maps were drawn to sort out specific research directions and hotspot distributions in the field of digital therapy in China, and to predict future research trends and challenges. The literature was retrieved from the CNKI database, exported, and converted into CiteSpace 5.7.R2 software for solving six quantitative indicators, clustering, and map drawing. Under the guidance and constraints of the results, the development context, research directions, core keywords, and hotspots in the field of digital therapy were analyzed. *Results:* A total of 458 articles were included. The clustering was credible ($Q = 0.8379 > 0.3$; mean $S = 0.9447 > 0.6$), and 19 significant research directions were obtained ($S > 0.6$). Four high-frequency core keywords ($F > 33$) and eight high-centrality core keywords ($C > 28$) were identified; there were no prominent nodes. *Conclusion:* In the field of digital therapy represented by Chinese literature, there are four mainstream research directions: exploration of digital medical theoretical concepts, multi-field integration of digital medical extensions, digital applications in specific disease scenarios, and infrastructure and policy support. However, from the perspective of literature centrality and mutation indicators, a solid theoretical core and landmark technological breakthroughs have not yet been formed, highlighting the huge potential and challenges of future research and application in this field in the Chinese context.

Keywords: Digital therapeutics; CiteSpace; Keyword co-occurrence; Knowledge graph; Hotspot

Online publication: June 6, 2025

1. Introduction

With the rapid development of the internet and artificial intelligence technologies in China, various industries are quickly advancing toward digitalization and intelligent transformation, with the healthcare sector being especially prominent. Against this backdrop, Digital Therapeutics (DTx) has emerged. Its advantages include low cost, minimal physician time, and support for patient self-management of health data, which has driven its rapid development globally. DTx is an innovative form of treatment that is based on digital technology and grounded in evidence-based medicine. The Digital Therapeutics Alliance defines it as: delivering evidence-based therapeutic interventions via high-quality software to prevent, manage, or treat medical disorders and diseases [1]. Through smart devices, digital platforms, and software programs, DTx provides real-time monitoring and personalized intervention, demonstrating significant application value in areas such as chronic disease, addiction disorders, rehabilitation, and mental health [2–4]. In recent years, the field of digital health has developed rapidly worldwide. As a key component of software as a medical device (SaMD), DTx not only complements traditional healthcare but also drives the digital transformation of healthcare services [5–6].

Globally, the U.S., Europe, and some Asian countries are leading in the regulation and practical application of DTx. For example, the U.S. FDA has included DTx in the regulatory framework for SaMD, and some products have entered commercialization and clinical practice. However, promotion is still challenged by insufficient technical standards and clinical validation [7]. Commercialization is also affected by funding, data privacy, and cross-border regulatory coordination [8]. In China, research and application in the DTx field began later but have grown rapidly in recent years, driven by policy support and technological progress. Early signs of success have been seen in areas like hypertension, non-alcoholic fatty liver disease (NAFLD), and addiction disorders [9–11]. However, current Chinese research mainly focuses on application in specific domains, lacking a systematic overview of overall development and research hot spots [12].

Given China's leading position in global digitalization and the scarcity of international research based on Chinese-language digital health due to language barriers, this study uses bibliometric methods to analyze Chinese-language literature on DTx from 1989 to 2021 using CiteSpace software. By constructing knowledge graphs, the study aims to comprehensively reveal the current research landscape and development trajectory of DTx in China, explore the evolution of research hot spots, and predict future trends. The findings aim to support theoretical research, policy-making, and practical implementation of DTx in China, while offering valuable references for the global digital health field.

2. Materials and methods

2.1. Study subjects

The keywords from literature in the Digital Therapeutics (DTx) field were designated as the study subjects.

2.2. Data collection

All Chinese literature related to Digital Therapeutics (DTx) was retrieved from CNKI, with the retrieval date set to May 21, 2021. The search strategy employed was: SU = (digital health) OR (digital therapeutics) OR (digital medicine), covering the time span from the inception of the database to 2021. The search results were reviewed by two researchers who independently examined the titles and abstracts of the literature and excluded irrelevant documents. In cases of differing opinions, consensus was achieved through discussion. If a conclusion could not be reached, a third researcher was consulted for a final decision. The bibliometric analysis utilizing CiteSpace allows for a certain degree of tolerance for incorrectly included literature (as indicated by unrelated nodes in the spectrum), thereby emphasizing the importance of recall over precision; nonetheless, this study ensured precision through manual screening while maintaining a high recall rate.

2.3. Data conversion

The CiteSpace data conversion function was used to reformat the CNKI data for analysis. The fields converted included keywords, authors, publication years, and additional relevant information.

2.4. Parameter settings and data preprocessing

Due to CiteSpace software constraints, the earliest analysis start year was set to 1921. Based on recommendations from CiteSpace developers, prior research, and preliminary observations of the retrieved literature, parameters were configured as follows: Time slice = 1 year; Node Type = Keyword; Edge weight algorithm = Cosine; g-index (k = 25) per time slice [13–18]. The software processed data for field generation, time partitioning, and pruning.

2.5. Keyword map generation and metric evaluation

The program was executed, and the AliasList function was used to merge nodes with similar or identical semantics. The resulting keyword maps and a series of metric indicators were obtained, including keyword co-occurrence maps, keyword co-occurrence maps for each slice, and keyword network clustering maps. The measurement indicators included: (1) node frequency (frequency, F); (2) node centrality (centrality, C); (3) node burstiness (burstiness, B); (4) modularity (modularity, Q); (5) silhouette clarity (silhouette, S); and (6) results from three representative class label extraction algorithms (tf*idf, LLR, MI) [14, 17, 20–22]. Indicators "(1)" and "(2)" yielded ranked keyword groups, used to elucidate the developmental trajectory of the field, through an integrated analysis of results from these two measurement methods; indicator "(3)" provided a ranked set of keywords aimed at identifying hot topics in the field, with a significance threshold set at B>10, referencing prior studies [23–25]; indicators "(4)" and "(5)" produced results ranging from 0 to 1 and -1 to 1, respectively, to evaluate clustering effectiveness, with values closer to 1 indicating stronger clustering quality. A clustering effect was deemed favorable and reliable when Q>0.3 and S>0.6; indicator "(6)" generated representative labels for each cluster, aiding in the analysis of category characteristics. Since the three algorithms are based on different assumptions, they provide three distinct sets of class labels, which, when analyzed together, yield superior insights. The three authors evaluated and discussed the DTx literature published from 1989 to 2021, culminating in a comprehensive summary based on the directions and constraints outlined by each map and measurement indicator. Following the generation of each indicator, additional functionalities of CiteSpace were employed for result refinement, including: (1) Utilizing the CitationHistory function to identify nodes with high burstiness, particularly those experiencing sudden increases; (2) After clustering, employing the Timeline function to convert the map into a timeline view, which identifies the time span of various literature.

2.6. Map interpretation

The upper left corner of the map indicates the software version information, the computation time of the results, time slice information (Slice length), the number of network nodes (N), the number of edges (E), and the network density, among other details. The colors of nodes and edges correspond to the graphical illustration above (from blue to orange), representing the years in which the corresponding keywords were used, with red indicating nodes that do not correspond to any specific year, suggesting high burstiness. The size of the node label text is positively correlated with frequency—larger text indicates a higher occurrence frequency of the keyword. The thickness of the edges reflects the co-occurrence frequency of the keywords; thicker lines denote closer associations. Utilizing two methods for measuring node importance will yield different maps. In the centrality measurement map, nodes are uniformly displayed in yellow, with a larger diameter indicating greater importance; in the frequency measurement map, a larger node indicates higher significance, and the rings surrounding the nodes represent the number of papers published in each respective year (the specific years are indicated in the graphic). A wider ring in year A reflects a higher occurrence frequency for that year.

Upon clustering, nodes that fall into the same category form a unit, each represented by a distinct color for differentiation, with category names indicated by "#" followed by the class label (the three algorithms will automatically identify three different class labels, with the MI algorithm possibly recognizing multiple labels), displayed in red.

Once the clustering graph is converted into the Timeline view, documents from the same cluster are arranged on a single horizontal line, with the document dates positioned at the top of the view, progressing from left to right as time advances.

3. Results

3.1. Data extraction results

A total of 1,218 articles were obtained from the literature search. After dual independent screening, 825 irrelevant articles were excluded. Ultimately, 458 articles were included in the analysis.

3.2. Data preprocessing

After setting the parameters in CiteSpace for preprocessing, 33 time partitions were generated. The spatial/node/edge configurations for each partition are detailed in **Table 1**.

Table 1. Configurations for the space/nodes/links graph

Year	Selection criteria for nodes	Total	Number of nodes	Number of edges/total
1989	top100%	32	32	57/57
1990	top100%	27	27	47/47
1991	top100%	27	27	47/47
1992	top100%	27	27	47/47
1993	top100%	27	27	47/47
1994	top100%	27	27	47/47
1995	top100%	27	27	47/47
1996	top100%	29	29	48/48
1997	top100%	27	27	47/47
1998	top100%	27	27	47/47
1999	top100%	27	27	47/47
2000	top100%	34	34	68/68
2001	top100%	29	29	50/50

Table 1 (Continued)

Year	Selection criteria for nodes	Total	Number of nodes	Number of edges/total
2002	top100%	27	27	47/47
2003	top100%	52	52	135/135
2004	top100%	53	53	89/89
2005	top100%	52	52	102/102
2006	top100%	77	77	178/178
2007	top100%	63	63	111/111
2008	top100%	73	73	199/199
2009	top100%	88	88	220/220
2010	top100%	67	67	178/178
2011	top100%	99	99	226/226
2012	top100%	70	70	118/118
2013	top100%	52	52	93/93
2014	top100%	83	83	171/171
2015	top100%	71	71	146/146
2016	top100%	53	53	88/88
2017	top100%	67	67	114/114
2018	top100%	88	88	164/164
2019	top100%	206	206	520/520
2020	top100%	296	296	705/705
2021	top100%	143	143	429/435

3.3. Node clustering, spectra, and associated metrics

After clustering, a total of 23 categories were identified (#0 to #22). Categories #18, #19, #20, and #22 did not yield keywords during the processing with CiteSpace and will be manually reviewed and categorized into similar classes. As shown in **Figure 1**, the overall clustering performance is satisfactory, with a high degree of reliability ($Q = 0.8379 > 0.3$; $meanS = 0.9447 > 0.6$).

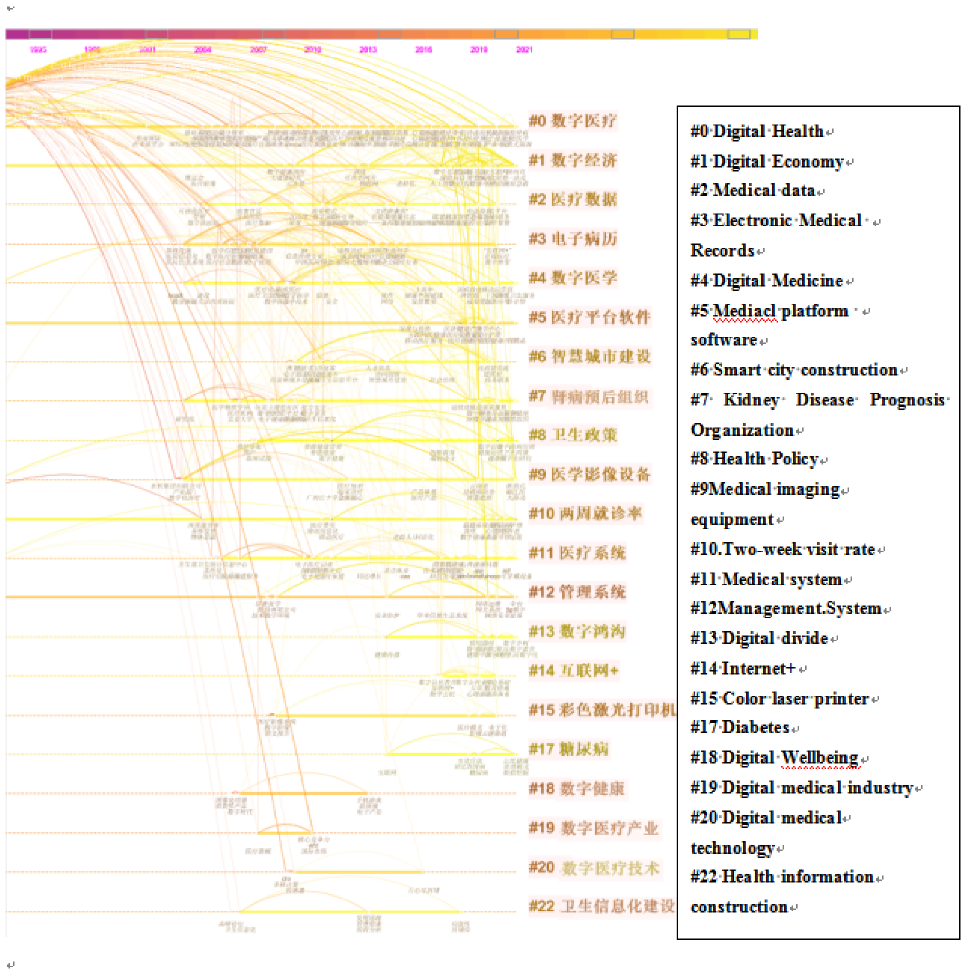

Figure 1. Timeline view

3.4. Node weight measurement results and graphs

The frequency and centrality measurement results are summarized in **Table 2**. A total of 8 nodes exhibit a weight greater than 28 (C > 28), and 4 nodes exceed a weight of 33 (F > 33), with no significant nodes identified.

Table 2. Top weighted keywords detected by the pair of metrics (only those ineffective clusters are listed)

C	Keywords	F	Keywords
304	Digital healthcare**	172	Digital healthcare**
65	Artificial intelligence	35	Chronic disease management
42	Digital economy	35	Health
38	Health management	34	Mdt
34	Electronic medical records	33	Life expectancy prediction
32	Cloud computing	33	Machine learning
30	Epidemic prevention and control	33	Online health information search
29	Industrial park	33	Offline medical treatment behavior
28	Health care	33	Doppler diagnostic instrument
28	Smart healthcare	33	Probabilistic analysis

Note: *emergent nodes; **overlapping nodes

The top ten projects ranked by SIGMA are as follows: digital healthcare, artificial intelligence, digital economy, health management, electronic medical records, cloud computing, epidemic prevention and control, industrial parks, medical health, and smart healthcare. However, the value for each of these is zero. This indicates that a stable hotspot has yet to emerge in the field of digital healthcare. Additionally, no projects have been identified using BURSTS as an indicator, suggesting a lack of emergence in this research area and indicating that there have not been any significant trends or fluctuations over a period of time.

4. Discussion

Bibliometrics posits that scientific knowledge encompasses two fundamental components: the intrinsic "knowledge structure elements", which establish a stable framework within a domain, and the "research bridging elements", which create research hotspots and promote innovation. In keyword co-occurrence networks, clustering is employed to identify the former, while high-degree nodes are utilized to detect the latter. Understanding both aspects allows for a clearer comprehension of several critical questions: What are the historical developments of domain hotspots? How long has each hotspot persisted? Are they interconnected or mediated by other hotspots? What are the recent directional changes in these hotspots? Internationally, certain advancements have been made in this area of research. For instance, Erten analyzed the ACM conference proceedings in the field of computer science to map temporal trends and track the evolution and emergence of "hotspots", revealing that terms like "design" and "system" have consistently maintained their relevance, while phrases such as "automated design" and "database" have surfaced intermittently, effectively forecasting the trajectory of domain development. As articulated by Chen Chaomei, the developer of CiteSpace, the purpose of scientific knowledge mapping is to provide researchers with highly quantitative "intuitive cues," guiding and constraining literature analysis to mitigate perceptual errors. Scholars should analyze the results generated by CiteSpace in conjunction with their understanding of the field and, when necessary, consult original texts for deeper insights.

In recent years, digital therapeutics (DTx) as an emerging field has garnered increasing attention. Particularly propelled by advancements in large language models and technologies like ChatGPT, the application domains of digital therapeutics are continuously expanding, giving rise to numerous novel application scenarios. These technologies play a crucial role in language translation services for cross-border healthcare and exhibit immense potential in managing areas such as cognitive and mental health disorders, chronic disease management, physical rehabilitation, visual rehabilitation, and drug computation.

4.1. Specific research directions in the DTx field

Based on the clustering analysis of keyword co-occurrence networks, this study has identified 19 high-confidence knowledge structure units that represent specific research directions within the DTx field. These directions not only reveal the core framework of the domain but also provide references for emerging research trends. In terms of theoretical concepts related to digital healthcare, the clustering analysis indicates that concepts such as digital healthcare, digital medicine, Internet+, and digital health constitute the foundational areas of DTx research [26]. The discussions predominantly revolve around the transformation of digital healthcare models in China, the role of intelligent health management, and the evolution of hospital digitization [27]. For instance, with the rapid development of information technology, hospitals are transitioning from basic information systems toward intelligent and integrated solutions, necessitating more systematic theoretical support [28].

In the application domain of digital therapies, the treatment of cognitive and mental health disorders emerges as one of the most active areas of research. The introduction of large language models has made personalized mental health interventions feasible, leading to more intelligent treatment plans. Additionally, the application of digital technologies such as virtual simulations, digital robotics, and medical imaging is increasingly prevalent in clinical healthcare, driving precision and efficiency in diagnosis and treatment. Studies indicate that "Internet+" has profoundly integrated into the healthcare sector, leading to significant transformations in medical service delivery and health management models.

At the macro level of digital healthcare, research topics such as the digital economy, smart city development, digital divide, health policies, and health information systems concern the impact of the digital economy on healthcare and health information issues in smart city initiatives. For example, local governments need to optimize the hardware and software environment to facilitate the integration of the digital economy and healthcare services.

Moreover, the issue of the digital divide faced by the elderly population represents another research hotspot. The phenomenon whereby older adults are marginalized in the digital economy not only affects their quality of life but also poses challenges to health equity. Therefore, investigating the relationship between health disparities among the elderly and the digital divide is crucial for achieving health equity. In the context of disease-related digital healthcare, disease management represents one of the core application scenarios within the DTx field, encompassing aspects such as medical data, renal prognosis organizations, two-week consultation rates, and diabetes management.

Research demonstrates that the level of information technology in modern hospitals directly influences service quality and patient satisfaction. For instance, diabetes management systems based on big data not only enhance patient adherence but also optimize physicians' clinical decision-making. Furthermore, there has been a significant upward trend in the research of digital therapy applications related to renal diseases and chronic disease management, reflecting the value and potential of digital healthcare in specific disease areas.

Finally, regarding the digital healthcare infrastructure, the clustering also includes nodes such as electronic medical records, cloud computing, and smart healthcare, highlighting the importance of the digital healthcare infrastructure. Electronic medical records, serving as the core tool for information management, combined with cloud computing and smart healthcare, enhance the efficiency of medical data storage, analysis, and sharing. For instance, the introduction of cloud technology provides efficient computing and large-scale storage support for electronic medical records, while smart healthcare optimizes the allocation of medical resources through IoT technologies. Looking ahead, AI-based smart healthcare is anticipated to further enhance diagnostic and treatment efficiencies.

4.2. Distribution and evolution of research hotspots in digital therapeutics

Research indicates that the field of digital therapeutics has yet to establish stable hotspots. High-frequency keywords such as "digital health" and "artificial intelligence" have seen significant increases in usage in recent years; however, their overall impact is insufficient to indicate the formation of a long-term trend. Furthermore, there has been no emergence of keywords that focus on burstness as a core metric, suggesting that the updating speed of hotspots within the field is relatively slow, exhibiting a more balanced development trend.

Nevertheless, it can be observed from the time-zone diagram that in recent years, the application direction of digital health has gradually shifted from a macro framework (such as the digital economy) to specific disease management (such as diabetes and kidney disease), as well as infrastructure development (such as electronic medical records and cloud computing). This trend reflects the transformation pathway from theoretical exploration to practical application in digital health, highlighting the interaction between technological development and societal needs.

5. Conclusion

This study conducted a systematic bibliometric analysis of Chinese literature from 1989 to 2021, utilizing CiteSpace to create a co-occurrence map of keywords, thereby revealing the research hotspots, development trajectory, and future trends in the field of digital therapeutics. The findings indicate that research in digital therapeutics primarily focuses on four areas: exploration of theoretical concepts, interdisciplinary integration, specific disease applications, and infrastructure and policy support. Despite the increasing volume of research in recent years, a solid theoretical foundation and landmark technological breakthroughs have yet to emerge, indicating that this field remains in its early stages of development, with significant potential and challenges ahead.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Zeng Q, Gao XY, Bai SZ, 2022, Theory and Practice of Smart Health Management. Chinese Journal of Health Management, 16(1): 3–6.

[2] Miao BY, Arneson D, Wang M, et al., 2022, Open Challenges in Developing Digital Therapeutics in the United States. PLOS Digit Health, 1(1): e0000008. https://doi.org/10.1371/journal.pdig.0000008

[3] Hong JS, Wasden C, Han DH, 2021, Introduction of Digital Therapeutics. Computer Methods and Programs in Biomedicine, 2021(209): 106319.

[4] Li RXG, Zhao X, 2023, Application of Digital Therapeutics in Asthma Management. Chinese Journal of Respiratory Diseases, 11(3): 344–348.

[5] Wang C, Lee C, Shin H, 2023, Digital Therapeutics from Bench to Bedside. NPJ Digital Medicine, 6(1): 38. https://doi.org/10.1038/s41746-023-00777-z

[6] Furstenau D, Gersch M, Schreiter S, 2023, Digital Therapeutics: Apps on Prescription. Business & Information Systems Engineering, 65(3): 349–360. https://doi.org/10.1007/s12599-023-00804-z

[7] Miao BY, Arneson D, Wang M, et al., 2022, Open Challenges in Developing Digital Therapeutics in the United States. PLOS Digital Health, 1(9): e0000027. https://doi.org/10.1371/journal.pdig.0000027

[8] Refolo P, Sacchini D, Spagnolo AG, 2022, Ethics of Digital Therapeutics (DTx). European Review for Medical and Pharmacological Sciences, 26(9): 3221–3230. https://doi.org/10.26355/eurrev_202209_29741

[9] Chinese Association of Drug Abuse Prevention and Treatment, 2023, Expert Consensus on Digital Therapeutics for Addictive Disorders. Zhonghua Medical Journal, 103(35): 2757–2764.

[10] He JY, Zhao LM, Liu K, 2023, Assessment of Digital Therapeutics for Self-management in Chinese Hypertensive Patients. Chinese Medical Journal, 136(5), 387–392.

[11] Sun C, Fan J, 2022, Application of Digital Therapeutics in the Treatment of Nonalcoholic Fatty Liver Disease. Clinical Hepatology, 5(2): 123–129.

[12] Zeng Q, Gao XY, Bai SZ, 2022, The Theory and Practice of Intelligent Health Management. Chinese Journal of Health Management, 16(1): 1–8.

[13] Chen CM, 2006, CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. JASIST, 57(3): 359–377.

[14] Chen CM, Chen Y, Horowitz M, 2009, Towards an Explanatory and Computational Theory of Scientific Discovery. Journal of Informetrics, 3(3): 191–209.

[15] Chen CM, Song IY, Yuan XJ, 2008, The Thematic and Citation Landscape of Data and Knowledge Engineering (1985–2007). Data & Knowledge Engineering, 6(7): 234–259.

[16] Song YH, Yang SL, 2014, Research Progress in Library Science, Information Science and Archival Science from an International Perspective: Information Visualization Analysis Based on CiteSpace. Library Forum, 6(3): 1–13.

[17] Fang Y, 2015, Visualizing the Structure and the Evolving of Digital Medicine: A Scientometrics Review. Scientometrics, 105(1): 5–21.

[18] Jaccard P, 1901, Etude Comparative de la Distribution Florale dans Une Portion des Alpes et des Jura. Bulletin del la Societe Vaudoise des Sciences Naturelles, 37(5): 547–579.

[19] Newman MEJ, 2006, Modularity and Community Structure in Networks. PNAS, 103(23): 8577–8582.

[20] Salton G, Yang CS, Wong AA, 1975, Vector Space Model for Information Retrieval. Communications of the ACM, 18(11): 613–620.

[21] Dunning T, 1993, Accurate Methods for the Statistics of Surprise and Coincidence. Computational Linguistics, 19(1): 61–74.

[22] Osareh F, 1996, Bibliometrics, Citation Analysis and Co-Citation Analysis: A Review of Literature. Libri, 46(3): 10–21.

[23] Zhao RY, Wang M, 2011, Visual Analysis of Traditional Chinese Medicine Research from an International Perspective. Journal of Medical Informatics, 2011, 7(2): 36–41.

[24] Song YH, Yang SL, 2014, Research Progress in Library Science, Information Science and Archival Science from an International Perspective: Information Visualization Analysis Based on CiteSpace. Library Forum, 6(3): 1–13.

[25] Fang Y, 2015, Visualizing the Structure and the Evolving of Digital Medicine: A Scientometrics Review. Scientometrics, 105(1): 5–21.

[26] Guo Q, 2011, Intelligent Health Management. Health Research, 31(2): 81–85.

[27] Xu PH, Zong WH, Zhou Z, et al., 2015, Investigation and Analysis of the Current Status of Digital Medical Construction in Public Hospitals in my Country. Journal of Chinese Health Information Management, 2015(2): 124–130.

[28] Chen JX, 2014, Preface to "Moving towards Digital Medical Care — Reconstructing the Theoretical System of Digital Hospitals". China Information Industry (e-Medical), 2014(7): 64.