A Clustering Observation of Multiple Theories in Digital Publishing in China from 2010-2022 --Bibliometric Analysis Based on Knowledge Graphs

Yiwei Wu

College of Communication and Art Design, University of Shanghai for Science and Technology,Shanghai 200093, China.

Abstract: With the arrival of the 5G era, the publishing industry has ushered in unprecedented changes. The industry is continuously deconstructed and reshaped within the impact of the Internet wave. In the current era, grasping the advantages of the Internet and leveraging digital media are new opportunities to promote the high-quality development of the publishing industry. In this study, a literature visualization analysis of 4365 research papers on digital publishing from 2010-2022 is conducted by drawing a knowledge map. The purpose of this study is to systematically understand the development history and academic theoretical system of the digital publishing industry and to analyze the internal logic between multiple disciplines and theories vertically.

Keywords: Digital Publishing; Publishing Industry; New Media; Communication Science

Introduction

The Internet, digital media, mobile, and artificial intelligence are all powerful driving forces for the development of China's digital publishing industry, and the academic community has launched a large number of studies on the development laws, internal logic, and integration development methods of the digital publishing industry. This study visualizes 4365 papers from 2010 to 2022 by means of literature visualization to show the characteristics of multiple transformations of research hotspots in the past ten years. It is hoped that this will deepen the understanding of the theoretical system of diversity in digital publishing and lay the foundation for further promoting digital publishing-related research in the future.

1. Research Design

1.1 Data source

The China Knowledge Network database (CNKI) was selected for this study. To ensure the authority of the data source, the journal type was therefore selected as CSSCI (Chinese Social Sciences Citation Index), and the search criteria were set in the CNKI database: (TS=(digital publication) OR TI=(digital publication) AND ((yearBetween ('2010', '2022')) AND (CSSCI journals='Y')). A total of 4630 documents were retrieved, and after manual screening of each document, 265 papers such as information, conference announcements, duplicate documents, and unrelated topics were excluded, and a total of 4365 documents were retained as data analysis texts.

1.2 Research Methodology

This study summarizes the evolution trend of annual publication volume, industry development trend, main theoretical systems, characteristics of research hotspot transformation, and main interdisciplinary disciplines involved in China's digital publishing industry from 2010 to 2022 by drawing a knowledge map. Knowledge mapping refers to the study of a certain knowledge field as the research object, and shows the image of the development process and structural relationship of knowledge in the field through data analysis and

literature analysis. [1] The study uses Citespace visualization and analysis software to visualize the literature selected in the field of digital publishing. Citespace software is currently one of the mainstream tools for knowledge mapping and visualization and analysis, which can observe the development of a field efficiently and accurately. [2] This multifaceted and dynamic analysis is more appropriate for the field of digital publishing, which is currently being given high priority.

1.3 Analysis process

Firstly, the selected 4365 analyzed texts were pre-processed by exporting the literature in Refworks format from the China Knowledge Network database and then formatted by Citespace, so that the data could be effectively recognized by the software in the analysis process. Secondly, the time slice is selected as 1 during the analysis, so that the research content of each year can be systematically sorted out and its transformation characteristics can be observed as a whole. The g-index algorithm is selected as the data selection criterion to filter the text and ensure the objectivity of the research results. Based on this, a cluster analysis is conducted to carefully observe and interpret the research flow and evolution trends.

2. Research results

ClusterID	Label (LSI)	Label (LLR)
0	Digital Publishing; Industry Integration; Information Construction; Reader Resources; Augmented Reality | Business Model; Development Path; Traditional Publishing Units; Open Access	Digital Publishing (469.43, 1.0E-4); Business Models (103.31, 1.0E-4); Profitability Models (34.8, 1.0E-4); Editing (28.99, 1.0E-4); Convergence (23.25, 1.0E-4)
1	digital publishing; springer; cultural communication; international academic forum; calis | digital publishing industry; private book industry; green publishing; eco-friendly paper; textbook recycling	Publishing (115.67, 1.0E-4); Publishers (76.16, 1.0E-4); Digital Publishing (51.21, 1.0E-4); E-Books (43.02, 1.0E-4); Digital Age (40.39, 1.0E-4)
2	Digital Publishing; Knowledge Services; Ecosystem; Emerging Publishing; Technology Journals | Traditional Publishing; Artificial Intelligence; Digital Transformation; Smart Publishing; Smart Media Era	Convergence Development (141.74, 1.0E-4); Traditional Publishing (97.57, 1.0E-4); Knowledge Services (88.04, 1.0E-4); Transformation and Upgrading (68.43, 1.0E-4); Publishing Transformation (54.36, 1.0E-4)
3	Science and technology journals; Development strategy; scielo; Multilingual publishing; Open publishing | Digital publishing; Profit model; Science and technology journals; Knowledge services; Development strategy	Science and Technology Journals (125.5, 1.0E-4); Publishing (73.51, 1.0E-4); Transformation (59.31, 1.0E-4); Strategy (54.21, 1.0E-4); Dilemma (44.98, 1.0E-4)
4	Academic Journals; Academic Publishing; Mobile Publishing; Academic Communication; Mobile Internet Environment | Digital Publishing; Industry Transformation; Crowdfunding Model; App Mobile Client; Academic Library	Scholarly Journals (123.19, 1.0E-4); Scholarly Publishing (88.27, 1.0E-4); Digital (70.45, 1.0E-4); Open Access (42.51, 1.0E-4); Cloud Publishing (38.84, 1.0E-4)
5	Talent Cultivation; Digital Publishing; cssci; Professional Content Construction; Development Trend | Research Hotspots; Co-term Analysis; Internet Copyright; Cluster Analysis; Social Network Analysis	Talent Cultivation (53.79, 1.0E-4); Going Global (46.63, 1.0E-4); Publishing (40.24, 1.0E-4); Research Hotspots (40.17, 1.0E-4); China Publishing (39.57, 1.0E-4)
6	Digital Publishing; Book Editing; Career Relationship Model; Digital Publishing Industry; Main Structure | Publishing Industry; Children's Publishing; Children's Book Publishing; Professional Books; Industry Model	Publishing Industry (62.81, 1.0E-4); Digital Reading (52.37, 1.0E-4); Reading for All (45.09, 1.0E-4); Internet+ (38.05, 1.0E-4); Industry Chain (27.86, 1.0E-4)
7	Digital Publishing; Artificial Intelligence; Information Society; Technology Journals; Digital Economy |Copyright Protection; Fine Art Works; Convergent Publishing; Journal Conversion; Editorial Thinking	Copyright Protection (92.47, 1.0E-4); Blockchain (62.95, 1.0E-4); Copyright (44.46, 1.0E-4); Big Data (40.89, 1.0E-4); Convergent Publishing (28.77, 1.0E-4)
8	digital publishing; composite publishing; publishing brand; publishing concept; core competence | media convergence; paper publishing; paper reading; cssci; Chinese online	Media convergence (88.8, 1.0E-4); Mobile publishing (49.78, 1.0E-4); Omni-media (41.2, 1.0E-4); Chinese online (20.29, 1.0E-4); Publishing brands (18.89, 1.0E-4)
9	Publishing Group; Capital Operation; Digital Publishing; Transformation and Restructuring; Development Method | Cultural Industry; Publishing Enterprise; Diversified Business Model; Practice Path; Evaluation System	Publishing groups (40.96, 1.0E-4); Cultural industries (38, 1.0E-4); Literature measurement (32.01, 1.0E-4); Capital operations (27.69, 1.0E-4); Practice paths (9.19, 0.005)

Table 1: Keyword clustering and log-likelihood rate

A total of nine clusters were obtained by plotting keyword co-occurrence mapping and keyword timeline mapping, in the order of #0 digital publishing, #1 publishing industry, #2 convergence development, #3 scientific and technical journals, #4 academic journals, #5 talent cultivation, #6 publishing industry, #7 copyright protection, #8 media convergence, and #9 publishing groups.(Table 1) When conducting such analysis, two main indicators, Modularity Q and Weighted Mean Silhouette, are used to judge the clustering analysis results, Q>0.3 means significant knowledge mapping structure; S>0.7 indicates credible clustering. [3] In this study, Modularity Q=0.355 and Weighted Mean Silhouette=0.727. The value of the log-likelihood rate directly indicates the representativeness of the

keyword, and the larger the value represents the larger the formation of the node. A large number of nodes were formed during the period 2010-2022, indicating the current diversified development of the digital publishing industry. The modern digital publishing industry is destined to be a comprehensive and innovative industry, and the digital publishing field needs not only the business capacity of the traditional publishing industry, but also the development of open and innovative digital business. According to the results of visual analysis, the main nodes formed are "digital reading, electronic publishing, all-media, self-publishing, new media operation, short video According to the results of the visual analysis, the main nodes are "digital reading, electronic publishing, all-media, self-publishing, new media operation, short video", etc. The basic conditions for the development of such businesses have also been studied empirically, and the main nodes are "copyright protection, personnel training, research institute, thematic publishing, resource construction, digitalization".Specifically, the clustering of "#1-#8" is summarized as follows, except for the "#0 digital publishing" node, which is significantly larger than the other nodes as a search term.

"#1 publishing industry, #6 publications, #9 publishing groups": for promoting the high-quality development of the publishing industry in the 21st century, Internet technology, big data, artificial intelligence and other emerging technologies are essential conditions. According to the current competition in China's domestic publishing industry and the international form of development of the publishing industry industry, competition tends to be fierce, and the intrinsic restructuring of the publishing industry, operation methods, and cross-regional development are facing challenges, but China's Internet economic environment and policy support make the publishing industry has the support of force.[4]

"#2 Convergence development" and "#8 Media convergence": the largest number of nodes and continuity are formed in this cluster, which includes "marketing, publishing model, development model, publishing and distribution ". For the global publishing and media industry, with the arrival of the 5G era, media convergence will be an inevitable trend for all media enterprises in the world, and accelerating the convergence and development of digital publishing and traditional publishing industry is a major issue facing China at present.[5]

"#3 scientific journals" and "#4 academic journals": among these two clusters, "traditional publishing, new media, profitability, and digitalization" are the most important nodes. This indicates that with the change of readers' reading habits and the reform of digital media technology, science and technology journals and academic journals are also facing digital transformation, covering publication forms, dissemination methods, and reading methods. Relying on the Internet, traditional journals can obtain faster dissemination speed, wider dissemination range, and higher profitability after digital transformation.[6]

"#5 Talent Cultivation": The services of digital publishing industry have changed significantly compared to traditional publishing industry. For the training of digital publishing talents need to pay attention to the following aspects, firstly, the service form of digital publishing industry produces changes, content carrier transformation, practitioners need to understand in detail the operation rules of digital platform and digital media technology; secondly, with the diversification of service objects, Internet service tends to develop personalized service solutions for users, practitioners need to master and understand user thinking orientation; finally, publishing Finally, publishing resources determine the development prospects of the digital publishing industry, and practitioners need to have a keen sense of the industry's development direction.

3. Conclusion

At present, the number of studies on China's digital publishing industry is gradually decreasing, but the development of relevant theoretical systems tends to be concentrated. In the coming years, it is necessary to continuously improve the ethics and regulations of digital publishing, accelerate the training of talents in the direction of digital publishing, promote the content integration of digital publishing industry and traditional publishing industry with the advantage of cross-disciplinary development, and produce more high-quality services to guarantee the high-quality development of digital publishing industry.

4. Research limitations

The analysis of Citespace software is based on java environment, which is limited by software version iteration and algorithm, and may have a slight gap with the real situation. In addition, for the sake of the authority of the data output, the analysis texts used in this study are all from the Chinese Social Sciences Citation Index, and the sample size source is relatively small.

References

[1] Chen Y, Chen CM, Liu ZY, Hu ZG, Wang XW. Methodological functions of CiteSpace knowledge graphs[J]. Scientology Research,2015,33(02):242-253.

[2] Hou HY, Chen CM, Liu ZY, Wang XW, Chen Y. A study on the cross-disciplinary properties of knowledge metrics[J]. Scientology Research,2010,28(03):328-332+350.

[3] Chen Y. Principles and applications of citation space analysis [M]. Beijing:Science Press, 2014.

[4] Wang GY. Development of China's publishing industry: current trends and changes [J]. Technology and Publishing, 2010(01):53-57.

[5] Liu BJ. Accelerating the integration of traditional publishing and digital publishing [J]. China Publishing, 2011(11):8-10.

[6] Cheng WH, Ren SL, Lu WR, Wang YK, Fang M. Suggestions for the transformation of China's science and technology journals from traditional publishing to digital publishing[J]. China Science and Technology Journal Research,2011,22(04):467-474.