A Brief Analysis of the Application of Big Data in Book Publishing (Postprint)

Authors: Fan Chunlong

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

Abstract

In an era of rapid scientific advancement, big data has emerged as a principal indicator of epochal transformation. Excelling at self-positioning its developmental trajectory and serving as one of the mainstream technologies of the new era, the convergence between big data and the publishing industry is currently engendering substantial information processing challenges. Under the deluge of big data, publishing houses are manifesting disruptive advancements, and building upon international experiences in big data-assisted book publishing, are reconfiguring and realigning national publishing strategies for the future. To further harness the extant advantages of big data, this paper conducts a comprehensive analysis of the publishing operational models presently employed in China' s book publishing industry, grounded in its practical realities, and proffers pertinent recommendations from an industry-level vantage point, thereby aiming to provide reference value for the prospective digital and informatized development of publishers.

Full Text

A Brief Analysis of Big Data Application in Book Publishing

Abstract: In today' s rapidly evolving scientific landscape, big data has become a primary hallmark of the era' s transformation. As one of the mainstream technologies of the new era, big data is adept at positioning its own development direction, yet the interface between big data and the publishing industry now generates numerous information processing contradictions. Under the torrent of big data, publishers have demonstrated disruptive progress, drawing upon foreign experiences of big data-assisted book publishing to replan and readjust domestic future publishing strategies. To further leverage the existing advantages of big data, this article provides a comprehensive interpretation of current publishing operation models adopted by the domestic book publishing industry

based on actual conditions, and proposes relevant recommendations from an industry perspective, aiming to provide reference value for the future digital and information-based development of publishers.

Keywords: big data; book publishing; application
CLC Number: G239.2
Document Code: A
Article ID: 1671-0134(2018)08-115-02
DOI: 10.19483/j.cnki.11-4653/n.2018.08.047
Author: Fan Chunlong

1.1 Effective Integration of Big Data and Book Publishing

Since 2009, a new generation of big data has become a key carrier in this industry sector due to its strong integration capabilities, and its influence has expanded to the book publishing industry. As the big data era arrives, some traditional publishing industries have fallen into confusion due to a lack of comprehensive understanding of new things. The enterprise that adapted to emerging technologies most quickly was Google. Following the "big data"trend, Google established a "word frequency statistics tool" on top of the massive Google Books database, forming a new application system. Vocabulary from 15 million books in global collections has been incorporated into this system. Web users can input up to five different words on the webpage and observe the frequency of these words across different years' data through the Google Books Ngram Viewer. Unlike the small data era, the big data era breaks the traditional fixed mindset that a series of information resources could be obtained from a small amount of data, emphasizing instead the observation and research of data itself.

1.2 Disadvantages of Big Data Application in Book Publishing

With the comprehensive popularization of electronic books, traditional book publishers have faced tremendous pressure and impact. Even timely measures such as price reductions still cannot resist the "devouring" effect of the internet. The formation and development of big data not only satisfy readers' rich reading demands but also bring novel and personalized reading experiences. From a historical perspective, the publishing industry has undertaken a significant mission in the evolution of human civilization. However, with the continuous changes in ideological trends, the traditional publishing industry faces squeezes and challenges from different aspects, such as digital publishing, online reading, and knowledge sharing.

To change this situation, a batch of domestic e-commerce enterprises have made a series of attempts around big data publishing. However, the CEO of domestic Dangdang.com stated: "The combination of publishers' Internet Plus and book big data may render the value of publishers zero in the future," citing multiple application disadvantages as follows:

First, facing doubts from numerous publishers and network operators, Dangdang.com' s statistics show that its reading app generates 45 million comments daily. For example, on platforms like Toutiao, after authors finish writing, they no longer need editors' preliminary review to determine content direction. Instead, they allow readers to comment directly, filtering and recommending books to readers.

Second, after purchasing books online, readers can share with other interested readers in other online communities. Although this achieves the sharing of book copyrights, it simultaneously breaks through copyright protection defenses.

1.3 Future of the Book Publishing Industry Under Big Data

For the digital publishing industry, digitalization plays a crucial role in its stable operation. Under the big data era background, digitalization has become an inevitable trend, and data worldwide may achieve sharing, with this goal' s realization depending to some extent on the development level of internet technology. In recent years, online book sales scales of Dangdang.com, Amazon, JD.com, Xinhua Online, Taobao, and others have shown obvious growth trends. While expanding the overall market size, they have also significantly diverted traffic from offline bookstores. Therefore, it is necessary to accelerate the effective integration of the book publishing industry with internet technology and fully utilize internet technology to analyze obtained shared data and individually read data.

2. Considerations for Big Data Application in Book Publishing Houses

2.1 Appropriate Use of Big Data for Book Planning

The application of big data technology in book publishing planning has extremely strict requirements. Under the current social background, data mining provides relatively limited beneficial value for topic selection planning, a disadvantage particularly prominent in some highly specialized book fields, with the gap continuously widening. Some book publishers, constrained by their personnel configuration and funding, apply big data processing technology to "small data" processing tasks to achieve good book planning results.

In a sense, the application of big data technology in China' s book publishing houses has not yet reached the true "big data" standard, as the data information generated under big data has quantitative limitations. The fundamental reason is the lack of a systematic and complete big data governance model within the domestic book publishing field, with each publishing house still controlling data acquisition, exchange, repeated utilization, and storage throughout the publishing industry. Different publishers have certain differences in human resource allocation, data investment, and degree of attention, and this industry gap has become the main barrier hindering data integration in the book publishing field.

Relevant national units have systematically researched and formulated specific implementation standards for big data in the industry based on China' s current book publishing industry status. As early as 2013, they proposed the national standard for publishing industry data exchange– "China Online Information Exchange for Publications–Book Product Information Format Specification" (CNONIX)–to comprehensively achieve continuous upgrading of big data collection, analysis, storage, and application tasks in the publishing industry. Although the comprehensive implementation of this standard provides favorable conditions for improving the overall big data environment of the book publishing industry, standards for data service platforms, data analysis, and related products and services have not yet been truly established. Data mining outsourcing services based on data trading and cooperation are chaotic and still in the trial stage, making it difficult to promote and apply them on a large scale in a short time.

2.2 Recommendation to Use Outsourcing Services for Key Marketing Book Topic Planning

At the current stage, there are still technical limitations in the application of big data methods in publisher topic planning. As the main source of data for big data methods in publishing topic planning, the vast majority of data from content, readers, and platforms remains in the hands of individual branch publishers, with no publisher truly controlling it. With the rapid development of digital publishing, its powerful advantages have become increasingly prominent. Behind this development, a large portion of output value comes from the rapid digitization of traditional publishing content. However, due to the high professional requirements of big data and the enormous workload, publishers find it difficult to complete related tasks according to relevant standards under current human resource conditions, leading to frequent outsourcing phenomena.

Meanwhile, when analyzing relevant content captured from online media and new media, there are often high technical requirements. Publishers generally prefer service outsourcing when choosing between cultivating their own professional talents and service outsourcing. Sellers control the vast majority of readers' reading channels, while major social networking platforms and reading platforms are controlled by new media companies. Publishers can master relatively little data foundation and cannot form the large database required for big data mining. Although some book publishers have effectively integrated big data methods into specific topic planning, they still primarily rely on outsourcing in actual data mining processes. Therefore, a data mining outsourcing service based on data trading and cooperation has emerged, and publishing units should further optimize their application level and efficiency.

2.3 Prioritizing the Positive Guiding Role of Books in Topic Planning

In some specialized book publishers, big data technology exhibits distortion to some extent in practical application. From a purely technical perspective,

the overall digitalization level in the book publishing professional field remains relatively low. However, unlike other products and services, book commodities have obvious particularities in practice. The social value of book commodities should be considered in advance—namely, the leading nature of books' social value in guiding readers' demands toward truth, goodness, and beauty. For example, technology books listed in the "highbrow" category may lead people to another path of success through decades of practical application; humanities books are even more so, as literary works often need to be "derived from life yet higher than life" to resonate with people.

For books, determining their content quality cannot solely seek common ground with authors but should instead understand and grasp the actual reading needs of the broad author base and guide readers to be positive and upward, planning and publishing corresponding books based on adapting to mainstream social values. Furthermore, the specific application of big data in book publishing should focus on reading content related to books. Big data serves only as a theoretical guidance method and should avoid becoming a "slave" to big data technology as much as possible.

Conclusion

In summary, big data has brought new historical development opportunities to the book publishing industry. The emergence of new perspectives, new methods, and new tools means that the construction of China' s digital book publishing houses should grasp the research characteristics and patterns driven by big data, master the latest technologies, keep pace with the information age, and continuously promote innovation in modern book publishing field theories and research methods. Therefore, when using big data thinking concepts and tangible and intangible research methods to answer and solve major theoretical and practical problems facing the development of China' s publishing industry, we should also be based on the forefront of advanced science and technology, skillfully promote the deep integration of big data, artificial intelligence, and other information technologies with humanities and social sciences research, and propose original theoretical contributions from Chinese scholars regarding key issues in current big data application research in book publishing, laying a solid foundation for the construction of digital book publishing houses.

References: [1] Zhang Zhiqian, Tang Xiaoli, Wang Xiaozu. Analysis of Market Trends and Planning Suggestions for Historical Books in the Big Data Era [J]. China Editor, 2015, 16(2).

[2] Miao Jincheng, Wen Xiaoli. Opportunities and Challenges Facing Book Publishing Under the Background of Big Data [J]. Communication and Copyright, 2014, 1(9): 356-358.

[3] Xu Jian. Research on the Collection Mode of Book Topic Selection Information in the Big Data Era [D]. Anhui University, 2015.

[4] Lei Ming, Zhang Yuting, Zhang Daling, et al. Application of Big Data in On-demand Publishing of Academic Books [J]. Publishing and Distribution Research, 2015, 8(11): 281-282.

(Author' s Affiliation: Jiangxi Science and Technology Press)

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