Analysis of Big Data Applications in Book Publishing: Post-Print Edition

Authors: Wang Xuegang

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

Abstract

Against the backdrop of the new era, big data has gradually emerged as one of the pivotal symbols of epochal transformation. As an emerging mainstream technology, big data has achieved effective integration with the book publishing industry. Publishing houses have made tremendous progress supported by big data, and have fully drawn upon successful international experiences of big data applications in the book publishing sector to formulate rational strategies for China' s future publishing endeavors. To further harness the potential of big data in book publishing, this paper examines the industrial development of big data in the publishing domain, conducts an in-depth analysis of its applications, presents brief case studies, and discusses relevant considerations for its implementation, with the aim of providing valuable insights for China' s book publishing industry.

Full Text

Preamble

Big Data Applications in Book Publishing

(Shandong Polytechnic Vocational College, Jining, Shandong 272100)

Abstract: In the new era, big data has gradually emerged as a significant symbol of transformative change. As one of the mainstream emerging technologies, big data has achieved effective integration with the book publishing industry. Supported by big data, publishing houses have made substantial progress, drawing upon successful international experiences to formulate rational strategic plans for China' s future publishing endeavors. To further harness the potential of big data in book publishing, this paper examines the industrial development of big data applications in publishing, conducts an in-depth analysis of its practical implementations, presents case studies, and discusses key

considerations, aiming to provide valuable insights for China' s book publishing industry.

Keywords: big data; industrial development; book publishing; application cases; considerations

CLC Number: G230.7
Document Code: A
Article ID: 1671-0134(2022)04-125-03
DOI: 10.19483/j.cnki.11-4653/n.2022.04.038
Citation Format: Wang Xuegang. Analysis of Big Data Applications in Book Publishing [J]. China Media Technology, 2022(04): 125-127.

1. Big Data and the Evolution of Publishing Formats

The widespread application of digital technology in book publishing has fundamentally transformed traditional publishing models. Network publishing, enabled by internet technology, has developed rapidly and matured, making digital publishing the new norm. Digital publications are no longer simple reproductions of print content; both their production methods and content delivery mechanisms have undergone significant changes. Publishing has expanded beyond traditional paper formats to encompass comprehensive multimedia presentations across paper and digital platforms. In the era of big data, book publishing has evolved into a diversified industry with expanded development space. This transformation extends beyond format changes to encompass the entire publishing workflow, which is becoming increasingly data-driven through the integration of information technologies that reshape and restructure traditional processes. For instance, self-publishing and crowdfunding models have introduced substantial innovations to conventional publishing practices. Furthermore, big data has facilitated diversified publishing dissemination while enhancing its targeting and accuracy. Network-based books, though virtual products, can leverage internet propagation to achieve precise market positioning.

2.1 Topic Selection and Planning

In the topic selection and planning phase, editors must collect comprehensive information about various books before completing manuscript organization and design. The application of big data in this phase manifests in three primary aspects. First, during information collection, big data enables thorough and specific exploration of potential topic directions. Second, in public opinion analysis, big data technologies assess the potential social impact of selected topics by constructing semantic network diagrams to compare topic merits and evaluate market prospects. Third, during manuscript organization, big data facilitates clustering and simulation analysis of author identities, tag information, and

their published works [1].

2.2 Editing Process

The editing phase comprises both the editing procedure and methodology, which have evolved from traditional manual transcription to typewriter printing, and now to simplified digital processes using various office software for visualization and enhanced effects. Big data leverages its advantages to transform manuscript processing from paper-based review to electronic document handling. Manuscripts can be transmitted via the internet, reviewed online, and processed using computer software [2]. Additionally, big data has streamlined plagiarism detection by enabling rapid comparison against comprehensive databases containing diverse literature. Editors can quickly complete plagiarism checks by comparing submissions against these databases [3]. In the big data era, data effectively reflects societal progress and development, enabling efficient database management through filtering outdated documents and retracting fraudulent publications.

2.3 Production

Both overstocking and stockouts significantly impact publishers' profitability. Excess inventory leads to resource waste, while stockouts result in missed revenue opportunities relative to market demand. Applying big data technology in the production phase effectively addresses these challenges. Publishers can leverage big data to accurately forecast sales figures, potential social impact, and popularity before production, providing crucial data support that reduces resource waste and enhances economic benefits.

2.4 Marketing

Book sales represent a critical link in publishing and the primary channel for generating economic and social value. Integrating big data technology in marketing helps publishers conserve resources and maximize profits [4]. Cultural products differ substantially from other commodities, as requirements vary significantly across regions, age groups, professions, and genders. Publishers can employ big data to analyze markets by examining related industries and products, gaining comprehensive insights into specific book categories' market performance. Regional variations in living standards and purchasing power create different demands and preferences for books. Publishers should analyze these regional and categorical differences to optimize production and reduce costs. Targeted promotion based on market analysis can effectively expand economic returns.

2.5 Post-Sales Service

In the new era, innovative service models contribute to the stable development of the publishing industry. Reforming post-sales services can provide readers with

superior experiences, expand information dissemination channels, and enhance both social and economic benefits. The publishing industry should integrate big data technology to build diversified information transmission methods, prioritize reader needs, and conduct research on reader demands to accurately identify market requirements and deliver targeted post-sales services. This approach promotes innovative development in editorial work and establishes a comprehensive publishing system spanning from topic selection to post-sales support. In practice, editors must emphasize the integration of traditional editing with new media platforms, expand post-sales services across multiple dimensions, create diversified and three-dimensional service models, enhance reader satisfaction, and leverage post-sales service as a crucial pillar of marketing strategy to build brand effects and strengthen market competitiveness, thereby achieving strategic development goals and creating greater social and economic value.

3.1 International Case Studies

International applications of big data in publishing fall into two main categories: first, publishing enterprises using big data as an analytical tool for topic selection and marketing; second, specialized big data companies providing high-quality services to the publishing industry through data analysis reports.

A notable case is Scholastic' s *The 39 Clues* series, which combined online and card games. The publisher used big data to identify the most popular characters and plotlines from the games to inform book planning. Coliloquy' s e-books feature plotlines derived from readers' choices about characters and story directions, allowing readers to determine narrative outcomes. Additionally, many emerging digital publishers conduct big data analysis of website data to write topical stories based on trending subjects, thereby attracting readers.

Professional big data companies offering outsourced services include American firm Hiptype, which developed an analytics tool for e-books and markets itself as "Google Analytics for e-books." This tool integrates rich e-book data for specialized big data analysis, profiling reader characteristics including age, income, and location. It tracks conversion rates from free chapters to purchases, completion rates, average reading progress, and favorite chapters, providing publishers with clear data analysis. Another example is Netflix, which offers topic selection assistance through big data analysis of massive user bases, helping publishers understand reader behavior patterns—such as which demographics prefer reading fantasy novels on tablets or smartphones at night, or which plotlines readers revisit most frequently. These cases demonstrate that international publishing industries have maturely integrated big data applications, offering valuable lessons for China' s sustainable development.

3.2 Domestic Case Studies

JD.com has long analyzed the purchasing power of tens of millions of consumers, identifying online user consumption patterns. For instance, readers who pur-

chase *One Hundred Years of Solitude* mostly also buy *A Calm Life is Not Lonely*, while many fitness enthusiasts purchase *General History of China*. Using big data to analyze web browsing and purchasing habits, JD.com found that 60% of these users were males aged 20-40. Based on this data, JD.com collaborated with publishers to release *David Beckham*, promoting it as a top book for male readers through big data-driven marketing, making it the flagship launch for "JD Publishing."

Many small and medium-sized publishers have also leveraged big data for marketing innovation with notable success. One publisher analyzed book sales data from the 11th to 12th Five-Year Plan periods, sourced from Dangdang and Taobao, examining pricing and reviews to identify patterns and develop solutions. This publisher conducted self-assessment of strengths and weaknesses, formulated a big data strategy for topic selection, and applied big data across topic planning, production, and sales, optimizing traditional publishing workflows. Big data excels at analyzing user behavior, social data, geographic location, and consumption patterns. Publishers applying these analytics have achieved excellent results, such as using big data to identify popular science topics within professional fields for topic selection reference, or predicting print runs and controlling inventory through sales forecasting.

4.1 Appropriate Application of Big Data in Book Planning

Applying big data technology in book planning requires contextual appropriateness. Current data mining has significant limitations for topic selection, particularly substantial bias in specialized publishing fields. Some publishers, constrained by staffing and budgets, conduct big data analysis using existing "small data" —a reasonable approach. While most Chinese publishers actively claim to use big data technology, they are not handling truly "big" datasets, as data volumes remain insufficient. This stems from underdeveloped big data governance models in specialized publishing, with data acquisition and storage handled independently by individual publishers. Variations in data investment, technical staffing, and organizational priorities make it difficult to establish a unified publishing data center quickly. Although industry standards for big data are under development and provide a favorable environment, standards for data service platforms, analysis applications, and data trading remain inadequate, hindering widespread adoption [1].

4.2 Outsourcing for Key Marketing and Topic Selection

Technical limitations persist in applying big data to publishing topic selection. During topic planning, massive content, reader, and platform data largely remain fragmented across branch offices rather than centrally controlled by publishers. As digital publishing becomes normalized, its advantages become more apparent, though rapid digitization of traditional content constitutes a significant portion of output value [5]. Big data demands high professional expertise

and involves substantial workloads, making it difficult for publishers to operate independently with existing human resources. Consequently, outsourcing is common. Analyzing content captured from online and new media platforms requires advanced technical skills, leading publishers to prefer outsourcing over in-house talent development. Since retailers control extensive reader channels and new media companies dominate reading platforms and social networks, publishers possess limited data foundations, making it challenging to establish comprehensive big data capabilities. This has spawned an outsourcing service industry focused on data trading and collaborative mining, helping publishers optimize big data application levels.

4.3 Emphasizing Positive Value Orientation in Topic Selection

Some specialized publishers face authenticity issues in big data applications. While the publishing industry' s overall digitalization level requires improvement, books differ from other products in practice, requiring publishers to consider their social value and forward-looking guidance—ensuring books correctly guide readers toward truth, goodness, and beauty [6]. For example, science and technology books may lead humanity toward higher civilization through practical application, while humanities books resonate with readers through content that reflects and transcends daily life. When evaluating book quality, publishers must not only consider popularity but also address readers' actual needs and whether books positively guide readers and align with mainstream social values. Additionally, publishers should maintain focus on reading content itself, recognizing that big data serves only as a theoretical guide rather than the primary work of publishing.

References

[1] Sun Hanqing, Zhang Jiani. Analysis of Big Data Applications in Book Publishing [J]. Publishing Wide Angle, 2021(16): 37-39.
[2] Zhu Jianli. Discussion on Big Data Technology Applications in Book Publishing [J]. Science and Technology Communication, 2020(15): 128-129.
[3] Liu Yaping. Innovative Applications of Big Data in Book Publishing [J]. Research on Communication Capacity, 2020(2): 128+130.
[4] Zhang Hongjie. Exploration of Book Publishing Marketing Innovation Under Big Data Background [J]. Media Forum, 2019(20): 142+144.
[5] Pan Yuhan. Analysis of Digital Publishing Status in the Big Data Era [J]. Science and Technology Communication, 2020(1): 133-134.
[6] Zhou Huan. Applications of Big Data Technology in the Book Publishing Industry [J]. Industrial Technology Innovation, 2019(27): 73-74.

Author Profile: Wang Xuegang (1976-), male, from Jining, Shandong, Assistant Librarian, research focus: library management.
(Responsible Editor: Yang Hu)