Research on the Application of Big Data in Newspaper Industry

Liu Ruoxin
Advertising Academy
Communication University of China
Beijing, China
duoduocuc@163.com

Wang Xuan
Media Museum
Communication University of China
Beijing, China
wenyujun@cuc.edu.cn

Song Kang
National Broadcast Media Language
Resources Monitoring & Research
Center
Communication University of China
Beijing, China
k.song@cuc.edu.cn

Abstract—**Big data is another subversive technological revolution in the IT industry after cloud computing and the Internet of things. At present, many native Internet enterprises have made a lot of achievements in some aspects by using big data technology. For example,ByteDance,a news application, uses big data technology to accurately portray users to achieve personalized news push; Tencent and Alibaba use the same technology to analyze the audience and package new advertising products called "programmed advertising".In 2009, publishing companies like The Guardian, The Economist, and the New York Times, and other internationally renowned publications began to explore big data news, and made significant achievements in using digital technology to expand the newspaper industry. In China, the newspaper industry tries to improve the effect of media communication by strengthening the creation of a comprehensive database and exploring the integration and development of new media. This paper will sort out the process of information development in the newspaper industry and the application of big data and new technology in news creation.**

Keywords—big data, newspaper industry, informationization, personalization

I. INTRODUCTION

In the face of massive information, traditional expansion has been unable to meet the development needs of the newspaper industry.Therefore, it began to build two network platforms: news websites and electronic newspaper editions, to open up new space for information production and management.In terms of mobile construction, the mobile news application has become the basic product of the traditional newspaper industry. Almost every newspaper group has built its own mobile application. Although Sohu News, Baidu News, ByteDance and other commercial portal news apps and aggregation news platforms continue to lead he information market, the performance of some mainstream media is still excellent.

II. RESOURCES OF BIG DATA IN THE NEWSPAPER INDUSTRY

At present, the newspapers have completed the basic information construction and the level of informatization construction is relatively small compared with ByteDance, Yidian, and other commercial companies.Informatization construction has helped the newspaper industry accumulate data sources, making the newspaper industry possess comprehensive data resources such as content data, user data, and Internet data.From the perspective of data sources, newspaper big data resources can be divided into internal data and external data.

A. Internal Data

The internal data of the newspaper industry mainly refers to the content data accumulated during the production of news by the newspaper industry, as well as user data collected from websites and mobile phone applications.The newspaper industry produces a wealth of news information every day. In addition to the news information released, there is also a large amount of news material. Taking the People's Daily as an example, it has accumulated a wealth of digital resources through various means such as text, pictures, cartoons, audio and video, forums, blogs, mobile phones, RSS, and live broadcasts. [1]

With the development of network technology, the newspaper industry can collect user data through various data collection and analysis tools, such as overall website visits, visit duration and column visits, clicks on certain articles and interactions with comments, Weibo retweets, WeChat interactions, daily active users, monthly active users, reading preferences, positions, behaviors, and many others. User information collection is an important part of information construction. People's daily has built big data user behavior collection platforms to collect user data from various channels.In 2015, this platform collected about 3 billion pieces of data per day from daily news application, website, Weibo, WeChat public account fans, and various social media and Weibo WeChat collections.

B. External Data

Based on the internal data of newspapers, the media have used information capture technology to collect information from mainstream domestic and foreign media, news websites, Weibo, WeChat, mobile applications, etc., and established Internet information databases.The collected content include picture links, video description information, and so on.

By the end of 2015, the People's Daily News had collected a total of 6,100 news media, including 3,000 domestic news websites, 1,000 foreign news websites, 2000 government industry websites, and 100 news apps; a total of 1,200 collection forums, including 1,000 domestic forums, and 200 foreign communities.

III. APPLICATION OF BIG DATA IN NEWSPAPER INDUSTRY

On the basis of rich data analysis, the newspaper industry mainly applies the results of data analysis to news topic selection,audience portraits,personalized recommendations and public opinion monitoring.

A. Big Data's Analysis Helping to Select the Topic of News

One of the important functions of big data technology for news editing is to find valuable news clues. Using big data analysis, the produced news can not only meet the needs of the audience, but also find directions for news editing. [3]

The application of big data can improve the efficiency of editing and production. For example, the use of big data technology analysis can achieve effective prediction of hot news and grasp global public opinion trends. News editors use a simple operation to automatically collect content from different channels such as websites, forums, Weibo, WeChat, and mobile applications, and then analyze multiple results such as hot topics, geographical distribution, relevant people, relevant institutions, and Internet users , and then gain a deeper understanding of hot topics and event propagation trends.This technology has changed the mode of relying only on the experience of editors in the traditional news production process and lacking accurate data support.

B. Audience portraits Help with Accurate Distribution

Establishing user portraits, grasping user behavior habits, and providing targeted accurate content and personalized services to users based on user characteristics have become indispensable and important capabilities in the development of Internet services and mobile applications.

Application of big data technology can collect user behavior data or access logs on newspaper websites and clients. Based on these data, the behavior characteristics of users can be analyzed.

C. Personalized Recommendation

ByteDance is a successful case of domestic personalized news recommendation. It provides users with personalized news push services in accordance with individual information needs and reading interests. The great success of ByteDance has made journalistic personalization technology attract the attention of journalists. On January 4, 2017, the People's Daily client added a big data intelligent recommendation function to achieve the effect of publishing personalized news.However, personalized recommendations for newspaper websites and mobile applications are different from commercial media. In addition to selecting accurate information based on the characteristics of the audience, the newspaper industry needs to focus on the transmission of values and create a personalized recommendation with a sense of news responsibility.

IV. PUBLIC OPINION MONITORING

The moral principle of Chinese journalists is adhering to the right direction of public opinion and improving the ability to guide public opinion, so, guiding the public opinion is an important responsibility of the media. Big data technology has changed the methods of social public opinion governance. With the support of big data technology, the monitoring and early warning of public opinion has been automated. The newspaper industry uses big data technology to achieve two things. On the one hand, it can monitor the Internet information dynamics in real time, provide materials for news topic selection, and on the other hand, it can carry out public opinion monitoring services. For example, the People 's Daily Network has launched the "people 's public opinion" channel. Dahe.com established the Dahe Public Opinion Research Institute to help the government collect information efficiently and provide scientific and correct data support for public opinion guidance.

V. DEVELOPMENT TREND OF NEWSPAPER INDUSTRY BIG DATA

Emerging media such as Weibo and WeChat have redefined the way people access information. Changes in the content and form of information dissemination place new demands on the development of the media and stimulate further technological innovation. The author believes that in the future, the application of big data in the newspaper industry will have the following trends.

A. Intelligent content material extraction

Editors usually use the title search method to select the materials needed for the manuscript, but because many data are unstructured data, they cannot be obtained directly through the search method, which causes difficulties in obtaining resources.By using big data technology, you can intelligently recommend materials and improve the efficiency of editing and obtaining materials such as pictures, audio, and video clips.

B. Promoting the Overall User Profile

At present, most websites and apps are collecting user behavior data and sketching user portraits. However, based on the data of its own platform, it does not form a complete portrait. In the future, it is necessary to make user portraits more accurate by expanding data sources.

C. Establishing the Evaluation System of Big Data in the Newspaper Industry

The traditional assessment management method of the newspaper industry is the use of specific publication volume of reporters as the evaluation index. In the new media era, the publication volume cannot objectively reflect the quality of work of reporters and editors. Through big data technology, the introduction of retweets, reads, reprinted media, comments, and likes to build comprehensive evaluation indicators is the future development of assessment management.

[1] Hao Guannan. Adhering to the Policy of "All Media" and Improving the Communicative Ability of Party Newspapers [D]. Beijing: Capital University of Economics and Trade, 2011.

[2] Zhou Linghua. A Brief Analysis of the New Trend of News Acquisition and Editing under the Background of Big Data [J].Journal Research Guide, 2018,9(04): 150.

[3] Lu Zhouli. Research on Reshaping User Connection in Media Integration of People's Daily [J]. News Communication, 2017 (04): 116-117