

# Research on the Application of News Material Mining Technology in the Information Age Post-print

**Authors:** Li Yunxia, Li Wei

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

## Abstract

[Objective] The advent of the information age has positively facilitated the development of the news industry. The industry has gradually shifted from traditional news collection models to Internet-era models. In this context, the rapid and accurate acquisition of required news materials has become a critical issue facing the news sector.

[Method] This article analyzes the application scenarios and advantages of news material mining technology in the news industry against the backdrop of the information age, dissects its key technologies, and discusses specific applications from five aspects: information acquisition, material processing, media monitoring, public opinion analysis, and event reporting.

[Results] Utilizing relevant technologies to mine and process massive information enables the rapid and accurate acquisition of required news materials, thereby improving work efficiency.

[Conclusion] With continuous technological progress and innovation, news material mining technology will see the development and realization of more application scenarios, effectively promoting the healthier and faster development of China's media industry.

## Full Text

### ChinaXiv Cooperative Journal

#### Research on the Application of News Material Mining Technology in the Information Age

**Li Yunxia, Li Wei**

*(Hua' an County Media Convergence Center, Zhangjiakou, Hebei 076150)*

## Abstract

The advent of the information age has significantly facilitated the development of the news industry, which has gradually transitioned from traditional news gathering models to Internet-era models. Against this backdrop, how to quickly and accurately obtain required news materials has become a critical challenge facing the news industry. This paper examines the application scenarios and advantages of news material mining technology in journalism within the context of the information age, analyzes its key technologies, and discusses specific applications across five dimensions: information acquisition, material processing, media monitoring, public opinion analysis, and event reporting. By leveraging relevant technologies to mine and process massive information volumes, news organizations can rapidly and accurately obtain needed materials, thereby enhancing work efficiency. As technology continues to advance and innovate, news material mining technology will see broader application scenarios developed and implemented, effectively promoting healthier and more rapid development of China's media industry.

**Keywords:** information age; news material; mining technology; application research; news data

**CLC Number:** G2

**Document Code:** A

**Article ID:** 1671-0134(2023)05-064-04

**DOI:** 10.19483/j.cnki.11-4653/n.2023.05.014

**Citation Format:** Li Yunxia, Li Wei. Research on the Application of News Material Mining Technology in the Information Age[J]. China Media Technology, 2023(05): 64-67.

With the continuous development of society, the media industry has also undergone constant evolution. In the new media environment, traditional media forms have gradually been replaced, and people now have increasingly diverse channels for accessing information, facing ever-growing choices. This poses higher demands on the media industry. How to quickly and accurately obtain needed information from massive data volumes has become a critical challenge. Therefore, to rapidly and accurately acquire required news materials, relevant technologies must be employed to process vast amounts of information.

## 1. Application Scenarios and Advantages of News Material Mining Technology in Journalism

News material mining technology refers to the process of extracting and processing news materials to ultimately obtain important information that reflects current social conditions and significance. In the information age, the Internet has become one of the primary channels for people to access information. News materials encompass all text, audio, image, and other data within a specific field that can reflect objective facts and development conditions in that domain.

## 1.1 Application Scenarios

**Discovering Newsworthy Content from Massive Information:** With the rapid development of network technology, information dissemination has become increasingly fast, and news events spread more quickly than ever. To ensure news reporting is rapid, comprehensive, and accurate, it is necessary to screen and extract newsworthy materials from vast amounts of information. Traditional media requires editorial staff to manually organize and filter news events to identify valuable content. Data mining technology can assist editors in finding newsworthy content and extracting it efficiently, providing media organizations with valuable news materials.

**Mining Content Already Reported by Other Media:** Mining content already reported by other media can help journalists discover not only unreported news and events but also identify relevant events and figures that have been covered by other media but may have been overlooked. During interviews, journalists encounter various people and situations. Through analysis of interview subjects, news material mining technology can provide journalists with valuable potential interviewees.

## 1.2 Advantages

News material mining technology offers significant advantages in news reporting, helping journalists better grasp the pulse of events and improving work efficiency and accuracy. Specifically: (1) It enables comprehensive acquisition of news information published across various media platforms, including text, images, and video, unrestricted by time or space. (2) It leverages natural language processing and machine learning to rapidly analyze and classify large volumes of news materials, enhancing work efficiency. (3) It tracks event development by mining and analyzing news materials, enabling timely understanding of latest developments and stakeholder positions. (4) It assists journalists in conducting in-depth analysis of news events to accurately identify underlying causes and influencing factors. (5) By analyzing specific elements within news materials, it discovers newsworthy information including hot topics, social discussions, and human interest stories, positively contributing to news reporting innovation and development. (6) It reduces error rates by quickly screening and classifying large information volumes, thereby minimizing journalist oversights and errors.

## 2. Key Technologies for News Material Mining in the Information Age

### 2.1 News Material Preprocessing Technology

In the information age, traditional news information collection models can no longer meet industry development needs. To improve news collection quality and efficiency, relevant news material mining technologies must be applied for information preprocessing. Preprocessing technology refers to operations per-

formed on required information before collection, including data conversion, compression, and extraction. This process primarily involves organizing and screening collected data to form a user-compliant dataset. Preprocessing technology not only enhances the quality and efficiency of news material mining but also improves news collection quality and efficiency. Additionally, effective preprocessing of large data volumes ensures comprehensiveness and accuracy of obtained information, providing a solid foundation for subsequent mining work.

## 2.2 Keyword Extraction Technology

Traditional news collection required extensive manual effort for information gathering, editing, and organization, consuming considerable time and resulting in significant content duplication. As information technology and journalism continue developing, the volume of news materials requiring collection has increased substantially. For such large datasets, accurate keywords are often difficult to identify, preventing effective information extraction. Keyword extraction technology now enables staff to rapidly obtain needed information while ensuring completeness and accuracy, laying a solid foundation for subsequent work.

Keyword extraction technology automatically extracts keywords from relevant news through software and matches them with news materials to complete automatic news screening. The technology primarily comprises keyword detection systems and information management systems. Keyword detection systems automatically detect required news materials through computer software, classify them, and match them with database information to retrieve needed data. Information management systems store, manage, and control relevant data through software, automatically classifying and organizing required data and news materials. Overall, keyword extraction technology effectively improves efficiency and quality in news information retrieval and content discovery, positively impacting news material mining and editorial work.

## 2.3 Text Classification Technology

During news material mining, collected information must be preprocessed to extract information from raw data and classify it. Text classification technology is widely used in news material mining, involving preliminary categorization of collected information according to certain standards. Two key considerations must be addressed: ensuring no duplicate classified information and maintaining high classification accuracy.

## 2.4 Sentiment Analysis Technology

Sentiment analysis technology primarily extracts emotional information contained in news and further filters it through sentiment analysis to provide journalists with valuable news materials. When applying sentiment analysis technology, text must be preprocessed using word segmentation systems for noise

removal and tokenization to enable computer recognition. During preprocessing, certain principles should be followed: first, ensuring extracted emotional information represents genuine emotions, and second, ensuring extracted emotional information has not been modified or tampered with. Care must be taken to avoid extracting emotional information unrelated to or weakly correlated with news events.

Sentiment analysis technology has extensive applications in news material mining: (1) **Public opinion analysis:** Identifying and classifying emotional tendencies in news materials to monitor and evaluate public opinion on news events, helping media understand public attitudes and emotional tendencies. (2) **Event detection and tracking:** Assisting in automatically discovering and tracking events in news materials and judging emotional tendencies of event participants. By monitoring and analyzing event-related emotional information, journalists can better understand event development trends and key points. (3) **News recommendation:** Intelligently recommending news materials based on user interests and emotional preferences to enhance reading experience. For example, users with positive emotional preferences can be recommended uplifting news materials. (4) **Brand monitoring:** Automatically analyzing and monitoring brand-related news materials to understand public emotional tendencies toward brands, identify potentially negative information affecting brand image, and help enterprises respond promptly.

## 2.5 Knowledge Graph-Based News Material Mining Technology

With continuous development of Internet technology, big data technology, and related information technologies, knowledge graphs have found extensive application across various industries and fields. A knowledge graph organizes and displays entities and their relationships within a specific domain, describing entity characteristics to form an interconnected network. Knowledge graph information is authentic, valid, and practically connected to real life, demonstrating strong utility.

Knowledge graph-based news material mining technology represents a news material processing approach built on big data and artificial intelligence. It utilizes natural language processing, machine learning, and graph construction technologies to extract and analyze information from news materials and organize it into a knowledge graph for deep mining and understanding of news events. This technology features: (1) **Knowledge graph construction:** Building a massive knowledge graph that associates and integrates various information from news materials to achieve knowledge acquisition and understanding of news events. (2) **Semantic analysis:** Using natural language processing to analyze and process news materials, extract and understand semantic information, and identify emotions and attitudes in text. (3) **Event discovery:** Automatically discovering events and topics in news materials, classifying and integrating relevant information to achieve deep mining and understanding of events. (4) **Intelligent recommendation:** Automatically recommending relevant news materials

and events based on user interests and needs, thereby improving information acquisition efficiency. Knowledge graph-based news material mining technology will have broad application prospects in future news material processing, providing richer, more accurate, and in-depth information services for news media and users.

### 3. Applications of News Material Mining Technology in News Gathering and Writing

#### 3.1 News Information Acquisition: Obtaining Sources More Efficiently and Accurately

In the information age, people have more diversified ways to access news materials. Beyond traditional search engines, social media and online forums have become important channels. For journalists, news material mining technology enables more efficient and accurate source acquisition.

First, information can be obtained through news search engines. News search engines specialize in retrieving news information by collecting, organizing, and classifying widely available online news to provide users with efficient and accurate news search tools, such as Google News and Baidu News. These platforms allow searching by keywords, time, location, sources, and other criteria, offering comprehensive news information with fast search speeds.

Second, social media platforms based on user-generated content serve as crucial channels for real-time news. Platforms like Twitter and Weibo allow users to publish news, comments, photos, and videos anytime, providing valuable news information sources. By following experts, media organizations, and journalists in the news field, journalists can obtain more in-depth, comprehensive, and timely information.

Third, database retrieval provides access to authoritative news information. News databases specifically store news information, including historical events and latest reports. Using database retrieval technology, journalists can search by keywords, time, location, media, and other attributes to obtain authoritative news information.

Fourth, RSS subscription offers a simple way to distribute online content. Users can subscribe to news, blogs, and other website updates, receiving notifications when new content is published. This is particularly convenient for those who frequently need to check specific news sources while ensuring access to the latest information.

Through these four methods, massive amounts of news information can be acquired in a short time. With the assistance of news material mining technology, this information can be further screened, classified, and analyzed to provide more accurate, comprehensive, and in-depth materials for news reporting.

### **3.2 News Material Processing: Handling Large Volumes of Text to Extract Valuable Information**

In the big data era, news writing staff must thoroughly study big data technology. Given the large volume, complexity, and dispersion of online information, data mining technology must be fully utilized for screening, organizing, and integrating data.

First, collected news texts must be cleaned to remove useless characters, symbols, and punctuation while standardizing formats for subsequent processing. Next, natural language processing technology segments news texts into individual words for further processing. Then, keywords are extracted from segmented text using natural language processing techniques such as TF-IDF and TextRank algorithms for subsequent analysis and mining. Topic analysis or topic modeling technology classifies news texts into different themes or topics for easier processing. Machine learning techniques judge and classify emotions in news texts to identify positive, negative, or neutral sentiment tendencies. Named entity recognition technology identifies persons, locations, and organizations in news texts for further analysis. Relationship extraction technology identifies associations between different entities in news texts for subsequent analysis. Additionally, during text data processing, information must be classified and filtered according to different needs, such as through keyword filtering and text clustering to quickly and accurately screen relevant information.

### **3.3 Media Monitoring: Tracking News Information Across Platforms to Grasp Event Dynamics**

News material mining technology requires a massive database containing extensive news materials. In practice, editors must reasonably utilize this technology to comprehensively monitor news published across media platforms, then combine this with their experience to organize and analyze relevant information, providing necessary materials for subsequent writing.

First, by mining and analyzing large volumes of news texts, journalists can monitor and track event development in real time, understanding event causes, processes, results, and related information such as key participants, timing, locations, and background. This helps journalists stay informed about event progress for better decision-making and action.

Second, by analyzing emotional tendencies, attitudes, and evaluations in news content, public opinion analysis and monitoring can be conducted. For example, analyzing the reputation and image impact of a brand or public figure in society helps identify and address issues of high public concern and extensive evaluation promptly.

Finally, by monitoring news reports across major media platforms, journalists can understand focus areas, reporting angles, data sources, and attitudes or positions regarding specific events across different media. This helps enterprises

or public figures better comprehend media operation patterns and characteristics to formulate appropriate communication strategies and enhance public awareness and image.

### **3.4 Public Opinion Analysis: Analyzing Public Sentiment and Reactions to Understand Social Opinion Trends**

In the big data era, news material mining technology not only helps editors conduct in-depth analysis of news events but also analyzes public sentiment and reactions, enabling editors to better grasp social opinion trends. Specifically, the technology analyzes public emotions and reactions according to their varying responses to different news events. For example, when encountering frustrating or infuriating events in daily life, people exhibit various emotional expressions. In such cases, news writing staff can use news material mining technology to analyze, summarize, and categorize public emotions. Simultaneously, in actual news writing processes, journalists must master certain analytical techniques and methods while continuously learning and exploring various analytical approaches to better mine public emotions and reactions.

### **3.5 Event Reporting: Collecting, Organizing, and Analyzing Event Information to Identify More In-Depth and Comprehensive Reporting Materials**

In the big data era, data serves as an important source for news events with certain authority. Through data collection, organization, and analysis, hidden information behind events can be discovered to provide editors with more in-depth and comprehensive reporting materials.

First, using big data technology to collect news events can reveal hidden information, such as statistical analysis of participants' age and gender to identify special groups involved in the event. Additionally, analyzing timing and location information helps better understand the meaning and value behind events.

Second, using big data technology to organize and analyze news events can reveal social repercussions and public perspectives on the issue. Through deep mining of events themselves, the causes of news events and public attitudes after they generate significant social response can be identified. Analyzing this data information provides editors with more in-depth, comprehensive, and elevated reporting materials, enabling them to better grasp social development trends.

## **4. Development Trends and Future Prospects**

With continuous development of information technology, the news industry's impact on current society is substantial and will positively influence China's future development. Therefore, in contemporary society, the news industry must continuously innovate news material mining technology to meet social development needs and improve technical standards applied in journalism.

The big data era has brought explosive information growth. News material mining technology analyzes and screens these massive datasets to provide news media with more accurate, comprehensive, and rich news materials. The technology's greatest innovation lies in its ability to conduct big data analysis according to user needs while comprehensively analyzing user browsing behavior, bookmarked content, and behavioral trajectories to provide personalized services.

In future development, the technology can be applied to more fields, such as urban construction and transportation, integrating with people's daily lives to provide more convenient and efficient services.

## References

- [1] Liu Shaoqin. Radio and Television News Editing Skills and Improvement Strategies[J]. China Newspaper Industry, 2022(10): 104-105.
- [2] Zhang Jianhua. Analysis on the Deep Development and Improvement of News Materials in the Internet Context[J]. China Prefectural City Journalists, 2021(9): 34-36.
- [3] Tang Xuelian. Organization and Editing of Television News Materials from the Perspective of Media Convergence[J]. Collection, Writing and Editing, 2021(5): 44-45.
- [4] Li Xuanbing. Research and Construction of a Network Hot News System Based on Text Mining Technology[D]. Beijing: Beijing Forestry University, 2020.
- [5] Meng Wei. Research and Analysis of Information Dissemination Methods in the New Media Era[J]. Global Premiere, 2019(7): 155.
- [6] Li Xinchu. Research and Implementation of Multi-domain Network News Hotspot Mining Technology Based on Spark[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [7] Feng Xia. Research on Hot Topic Mining Technology for Network News[D]. Nanjing: Nanjing University of Science and Technology, 2019.
- [8] Zhong Zhijin, Wang Tongchen. Application of Big Data Text Mining Technology in Journalism and Communication Discipline[J]. Contemporary Communication, 2018(5): 12-18.
- [9] Luo Zheng. TRS Network Public Opinion and Social Media Mining Service Platform[J]. China Media Technology, 2011(8): 54-56.
- [10] Shang Enjie. Opportunities for the Declining "Noble" "Oriental Horizon" –A Brief Discussion on the Development of TV News Magazines Using Data Mining Technology[J]. Journalism Research Guide, 2015(14): 237+240.

## Author Biographies

**Li Yunxia** (1975-), female, from Zhangjiakou, Hebei, Editor. Research direction: news gathering and editing.

**Li Wei** (1983-), male, from Zhangjiakou, Hebei, Assistant Editor. Research

direction: news gathering and editing.

(Corresponding Editor: Zhang Xiaojing)

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

*Source: ChinaXiv –Machine translation. Verify with original.*