A Preliminary Analysis of the Supporting Applications of Internet Big Data in Media Business: A Case Study of the Public Opinion Monitoring System of People' s Public Security Newspaper (Postprint)

Authors: Wei Chunguang

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

Abstract

To meet the demands of cybersecurity work under the new circumstances, network public opinion monitoring systems have emerged. These systems aim to conduct technical processing—including collection, classification, integration, and screening—of various types of online information, thereby maximizing social benefits, fostering healthy industry development, and maintaining social stability. This paper seeks to provide insights by reviewing cybersecurity-related policies, analyzing the application of network public opinion monitoring systems, and summarizing the social benefits derived from technological implementation.

Full Text

Abstract

To meet the needs of cybersecurity work under new circumstances, network public opinion monitoring systems have emerged. These systems aim to collect, classify, integrate, and filter various types of online information through technical processing to maximize social benefits, promote healthy industry development, and maintain social stability. This paper reviews relevant cybersecurity policies, analyzes the application of network public opinion monitoring systems, and summarizes the social benefits brought by technological implementation, hoping to provide inspiration for readers.

Keywords: People's Public Security Daily; public opinion monitoring; internet; big data; media convergence
CLC Number: TP393

Document Code: A
Article ID: 1671-0134(2019)06-080-03

The system targets functions including relevant information collection, thematic event analysis, social hotspot discovery, key content monitoring, statistical data analysis, public opinion briefing production, search, and management, striving to achieve comprehensive content, complete functionality, ease of use, open compatibility, and security reliability. In summary, comprehensive understanding and mastery of public opinion constitute one of the important tasks of public opinion monitoring systems.

The construction purposes of this system are: to grasp the main viewpoints and attitudes of netizens; to understand media coverage and focus areas; to automatically generate public opinion briefings for timely response to emergencies and improved work efficiency; to enable continuous tracking and analysis of specific events; and to form a unified operational service platform that serves as a business support auxiliary platform for news topic selection and editorial work.

2.1.1 Necessity of System Construction

The construction of a public opinion management system is, first and foremost, necessary to ensure long-term national stability. It facilitates the construction, utilization, and management of the internet, maintains the overall situation of reform and development, and consolidates the Party' s governance foundation. Second, establishing correct public opinion guidance is a prerequisite, which helps maintain close contact with the masses, accurately grasp social conditions and public opinion in a timely manner, effectively guide online discourse, and control the development direction of public opinion. Third, it meets the urgent need to adapt to future network public opinion management, enabling timely response to the trend of media convergence across various new network media and mobile internet platforms, and improving the accuracy of handling internet public opinion issues. Furthermore, employing high-tech means is an urgent need to enhance management capabilities, helping to adapt to the information development strategy in the new era and improve the e-government system. Finally, it meets the need for network propaganda work to keep pace with the times and innovate, helping to improve the management capabilities and ideological level of internet practitioners and give play to the industry self-discipline mechanism of online media.

2.1.2 Detailed Scientific and Technical Content

(1) Distributed Data Management Technology: The Hybase Big Data Management System is a new-generation big data management system centered on storage, retrieval, and statistics, employing an elastically scalable architecture design. It integrates multiple advanced technologies including full-text retrieval, natural language processing, index sharding, multi-replica mechanism, peer node mechanism (decentralization), column storage, and in-memory index-

ing, providing efficient management and intelligent retrieval of unstructured big data for various big data analysis applications. Its advantages include:

Flat Design: The flat architecture ensures that a single node failure does not affect the entire system' s service provision. Meanwhile, this architecture provides good scalability, allowing new nodes to be added online to expand system capacity and increase external service capabilities.

Automatic Recovery with Anomaly Awareness: When the system automatically detects that a server is in an abnormal state, it can perform self-repair. The system treats hardware anomalies as common exceptions, preventing a single node failure from rendering the entire system unavailable.

Flexible Multi-Engine Technology: The system employs a multi-engine mechanism by defining a standard engine interface. For different application requirements, different engines can be used, and users can even build their own engines to extend the system' s data processing capabilities.

Heterogeneous Data Support: The system supports unified retrieval of structured, semi-structured, and unstructured data.

Efficient Partitioned Indexing Mechanism: Based on query characteristics, the system can automatically partition and index data.

Hybrid Indexing Methods: The system provides word-based, character-based, and hybrid word-character indexing methods to meet different application scenarios' requirements for recall and precision.

In-Memory Tables: The system supports creating data tables in memory to accommodate applications with small data volumes but high requirements for query concurrency and response speed.

Column Storage: The system supports column storage to achieve efficient access to specific data columns, improving the speed of classification statistics and sorting for particular data columns.

Asynchronous Retrieval: It supports asynchronous retrieval mode to adapt to high-connection application scenarios, avoiding the problem of excessive thread resource consumption in synchronous retrieval mode.

Multi-Level, Multi-Granularity Distributed Cache: The system features both single-node retrieval cache and merged overall retrieval cache, which can greatly improve cache hit rates, reduce retrieval node pressure under high concurrency, and thus substantially enhance the system' s data retrieval capabilities under high-concurrency conditions.

Extensible Retrieval Mode: It supports root-word retrieval and English root retrieval combining algorithms and dictionaries, with an accuracy rate reaching 99.9%. It also supports extended retrieval based on synonyms and subject terms.

Hadoop Standard Compatibility: TRSHyBase seamlessly integrates with Hadoop, can fully utilize HDFS reliability, and undertake storage of large objects such

as images, audio, and video.

(2) Internet Information Collection: Real-time monitoring of massive internet data covering news, print media, forums, blogs, microblogs, WeChat, APPs, search engines, etc. The platform integrates three major functions: tracking public opinion information dissemination channels, tracing sources, and guiding public opinion control and command. Ultimately, it forms and produces distinctive routine public opinion monitoring products, including daily, monthly, and annual reports for various industries.

(3) Internet Information Intelligent Processing: For different types of public opinion content, the Hybase big data management system utilizes advanced statistical techniques and intelligent text analysis and mining technologies to achieve data filtering. The system features multi-language recognition and automatic transcoding, automatic word segmentation, automatic classification, automatic clustering, automatic hotspot discovery, similarity retrieval, article deduplication, automatic summarization, and key information extraction. It can perform basic data processing for various functions of the public opinion monitoring platform according to actual work requirements.

(4) Full-Text Retrieval Function: The system supports multiple classification-based retrieval methods including by source, time, domestic/overseas, and information source, providing intelligent analysis-based information retrieval services. Different users can retrieve relevant content based on their permissions. It supports advanced retrieval by body, title, time, author, website, etc., with average retrieval response time not exceeding 5 seconds. Additionally, the system allows user permissions to be set for retrieving relevant content.

(5) Internet Information Analysis Applications: The system implements early warning and reminders for key information, trend analysis of key events, website analysis, figure analysis, hotspot analysis, and automatic public opinion report generation. In terms of permissions, it provides complete user and permission management mechanisms to fully ensure the security of intelligence information content. Users are grouped and classified with graded permissions. The system supports permission control by classification, applicable to both users and roles, ensuring system data security and applicability. It provides multi-user login functionality with hierarchical management settings for user functional permissions, keywords, columns, topics, and information. It manages articles through topping, favoriting, hiding, inputting, editing, and reviewing, can preserve evidence of web traces, and uses probe functions to discover whether original links are valid. The system provides complete and detailed logs that can track user login and management activities. Logs can be queried by conditions, enabling detailed recording of system operation logs and application statistics for various departments and users, facilitating audit administrators in application auditing.

2.2 Innovation Points: Big Data Management

As network systems become increasingly complex—a trend in technology application and development—data volume continues to grow, with information achieving a leap from TB-level to PB-level scale, making latitude indicators for data analysis more extensive. The big data management system developed for this project can, on the one hand, achieve unified management and retrieval of structured, semi-structured, and unstructured data; on the other hand, it conforms to the technical trend of "structured processing of unstructured data and unstructured processing of structured data."

2.2.1 Information Collection Technology

The focus of this project's collection efforts is deep-level collection technology for areas rarely covered by search engine technology (oriented toward DeepWeb). With the rapid development of network application technologies, network information exhibits certain "heterogeneous" characteristics. As internet community development and Web 2.0 rise, network information using HTTP as the network transmission protocol and HTML as the display format can no longer meet development needs, and the content contained in web pages is undergoing profound changes. The internet, originally dominated by website/webpage content, has gradually evolved into a landscape where websites, microblogs, WeChat, forums (communities), and blogs coexist. Platforms such as microblogs, WeChat, forums, and blogs contain massive amounts of information and have become important sources of information on the internet. For industry search engine construction, information on these platforms has more important use value than information on ordinary websites.

The system not only performs intelligent analysis and mining of data but also, based on this, fully utilizes knowledge tags obtained through intelligent data analysis technology to fuse and process knowledge, thereby constructing knowledge graphs that enable users to query and browse knowledge entries like using an encyclopedia, as well as knowledge graphs with extensive associative relationships. The system must fully utilize metadata content obtained through text mining to create "story flow" style services that provide intelligent assistance for news production. The system must conduct dissemination analysis of communication content from perspectives such as positive/negative information, attention level, and dissemination speed to obtain dissemination effects, laying the foundation for intelligent decision-making at the newspaper.

In summary, the public opinion monitoring system plays a connecting and inclusive role, meeting both the functional requirements of system construction and revitalizing newly added massive data assets to achieve data value addition and reuse. It supports news discovery and intelligent creation at the newspaper, thereby promoting media convergence development and earnestly implementing the spirit of General Secretary Xi Jinping's important speech at the Party's news and public opinion work symposium.

2.2.2 Comprehensive Comparison with Current Domestic and International Similar Research and Technologies

Distributed Big Data Management System: Organizing and managing massive data requires a scalable storage and processing framework. Currently, highly scalable distributed cloud computing environments using inexpensive computers have attracted the attention of commercial giants such as IBM, EMC, and Microsoft, and have achieved success in companies like Google, Amazon, and Yahoo. Cloud computing environments generally include scalable file systems, concurrent processing operation primitives, and reliable data storage. Since massive data management requires entirely new computing and storage models, industry players such as Google, Yahoo, Microsoft, and IBM, as well as research institutions, have fully utilized the data storage and concurrent processing functions provided by underlying cloud computing environments to achieve massive data storage and management.

The development of distributed computing environments has provided storage and processing foundations for massive data. Major companies have begun building SOA-based massive data integration systems on distributed computing environments. Based on current progress, main problems include: metadata in current unstructured data may include general metadata such as anchor text and dates, or manually input information, with ineffective integration of information extraction and unstructured data management; distributed computing models can conveniently support keyword queries but lack good optimization for conditional data queries, with data query processing efficiency needing improvement; the Pay-as-you-go approach for massive unstructured data systems requires further support, including underlying storage support for merging, decomposing, and optimizing storage of different attributes.

2.2.3 Intelligent Text Processing Technology

Foreign research institutions have conducted text mining and information extraction research relatively early, with numerous well-known institutions including Carnegie Mellon University, University of Maryland, UC Berkeley, and IBM Corporation. Domestic research on text mining and information extraction began in the 1980s, with major institutions including Peking University, Tsinghua University, Harbin Institute of Technology, Institute of Computing Technology of the Chinese Academy of Sciences, and Microsoft Research Asia. China attaches great importance to research in this field, with the National 863 Program having organized multiple specialized evaluations of text intelligence technologies such as word segmentation, classification, summarization, keyword indexing, information extraction, and sentiment analysis. These evaluations have greatly promoted the development of related domestic technologies.

The intelligent text processing system developed for this project utilizes advanced statistical techniques and intelligent text analysis and mining technologies to filter data content for different types of public opinion. It features multi-

language recognition and automatic transcoding, automatic word segmentation, automatic classification, automatic clustering, automatic hotspot discovery, similarity retrieval, article deduplication, automatic summarization, and key information extraction, providing basic data processing for various functions of the public opinion monitoring platform.

3. Social Benefits Brought by Big Data Technology

The network has become the main mode of information transmission in China. Therefore, the network environment must be taken seriously. Only by maintaining a good network environment can the network truly play its role, better bring convenience to users, and have a positive impact on economic and cultural development. This project is an important component of comprehensively implementing the spirit of General Secretary Xi Jinping' s important speech at the Party' s news and public opinion work symposium. It is an important step in applying big data technology research results to media transformation practice and has significant social importance.

3.1 Comprehensively Implementing the Spirit of General Secretary Xi Jinping' s Important Speech at the Party' s News and Public Opinion Work Symposium

This project closely follows the spirit of General Secretary Xi Jinping' s important speech, adheres to the correct political direction and public opinion guidance, seizes historical opportunities in information development, accelerates breakthroughs in core information technologies, maintains network social security, creates a clean and righteous cyberspace, and gives full play to the leading role of information technology in economic and social development.

3.2 Conducive to Curbing the Spread of Harmful Information and Speech with Correct Public Opinion

Network public opinion can help understand public sentiment. Effective scientific screening, quantitative statistics, and analysis of online public opinion data, combined with practical experience and closely integrated with historical development and China' s national conditions for research and judgment, can have a predictive effect on tendencies and emerging issues. Through the construction of this project, we can strengthen our grasp of network public opinion trends, do a good job in public opinion collection and comprehensive analysis, and provide references for central decision-making. It can also build a "green channel" between the government and the masses.

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