Overview of the Tianji Public Opinion System (Postprint)

Authors: Zhai Lidong, Zhihua Yu, Ding Guodong

Date: 2017-03-10T00:00:00+00:00

Abstract

With the increasingly profound development of the Internet, it has become one of the primary media for information dissemination in contemporary society. Online public opinion forms rapidly and exerts a tremendous influence on society; thus, strengthening the monitoring and analysis of online public opinion warrants high attention from all sectors of society. The Tianji Public Opinion Monitoring System is specifically designed to meet the work requirements and characteristics of online public opinion monitoring. It not only employs professional search engine technology but also integrates more intelligent data mining techniques, establishing a daily-cycle online public opinion monitoring platform complemented by weekly or monthly-based public opinion analysis reports, thereby providing a convenient, scientific, and operable platform for public opinion work.

Full Text

Preamble

Vol.8 No.4
Information Technology Letters

Overview of the Tianji Public Opinion System
Li-Dong Zhai, Zhi-Hua Yu, Guo-Dong Ding

Abstract

As the Internet continues to penetrate deeply into society, the network has become one of the primary media for information dissemination. Online public opinion forms rapidly and exerts tremendous social influence, making the strengthening of online public opinion monitoring and analysis a matter that warrants serious attention from all sectors of society. The Tianji Public Opinion

Monitoring System is specifically designed for the operational requirements and characteristics of online public opinion monitoring. It employs not only specialized search engine technology but also more intelligent data mining techniques to establish a daily-cycle network public opinion monitoring platform, supplemented by weekly or monthly public opinion analysis reports, thereby providing a convenient, scientific, and actionable platform for public opinion work.

Keywords: Public Opinion; Information Retrieval; Classification and Clustering; Data Mining

1 Introduction

Online public opinion refers to the socio-political attitudes, beliefs, and values held by the populace regarding public issues and social administrators, as expressed within a certain social space on the Internet surrounding the occurrence, development, and evolution of social events. It represents the aggregate expression of beliefs, attitudes, opinions, and emotions by large numbers of people concerning various social phenomena and issues.

The influence of online public opinion is increasingly significant, permeating every level of society from government decision-making to social, political, economic, cultural, and daily life. However, alongside this trend come various biased comments and false information that harm national security and social stability. Correctly guiding online public opinion has become a major challenge in maintaining social stability.

The second part of this paper provides an overview of the current state of online public opinion; the third part analyzes the requirements for online public opinion monitoring; the fourth part introduces the Tianji Public Opinion System solution in a comprehensive yet accessible manner; and finally, we summarize the technical features and practical implementation of the Tianji Public Opinion System.

2 Current Status and Characteristics of Online Public Opinion

Online public opinion represents an emerging form of public discourse that possesses unique characteristics distinct from traditional media. First, it offers anonymity and virtuality, making effective supervision and regulation difficult, as speakers generally need not consider legal accountability. Second, the authenticity of online information is hard to verify. Many netizens rarely question the veracity of information nor engage in mature critical thinking, instead expressing opinions based on subjective assumptions with strong emotional overtones, which can lead to the concealment of truth and distortion of discourse. Third, online public opinion spreads rapidly, reaches wide audiences, and exhibits strong aggregation effects, but is difficult to control. When defamatory messages are pasted across the network, authorities can only passively delete them, making

prevention nearly impossible.

Due to these characteristics, online public opinion can produce tremendous guiding effects. On the positive side: (1) It facilitates the full expression of genuine public sentiment, enabling the government to understand the people' s voice promptly and accurately. The virtual nature of online space emboldens individuals to express their true thoughts, particularly in online forums where netizens can discuss current affairs and engage in debates. (2) It provides a basis for scientific government decision-making, helps expand public participation rights, and enhances citizens' sense of ownership. Overall, the emergence of any form of online public opinion reflects netizens' concern for the nation' s future and social public affairs, as well as their growing awareness of participation and civic responsibility. In the information age, online public opinion has gradually become an important channel for the government to listen to and understand the populace. (3) Through the guidance of opinion leaders, it can dilute some emotionally charged and extreme remarks, helping to alleviate increasingly tense social contradictions. Online forums host a large number of "opinion leaders" with strong writing skills, deep analytical abilities, and unique insights whose statements subtly influence other netizens' views. Therefore, healthy and rational discourse guided by them can control negative impacts with 事半功倍 (multiplier) effects. (4) It promotes supervision of government officials and increases social transparency, helping to constrain unhealthy practices. The universality and uncontrollability of netizen participation make online public opinion ubiquitous, creating a "heavenly net" of mass supervision that places the operation of public power under the sunlight, thereby promoting increased social transparency and government information disclosure.

However, online public opinion also brings negative effects: (1) Netizens' emotional tendencies lead to frequent occurrences of online violence. Due to the concealed identity of speakers and the lack of rules and effective supervision, some individuals use the "virtual" online world as a space to vent negative emotions when facing difficulties and problems. (2) The absence of online "gatekeepers" leads to the proliferation of false information, interfering with netizens' correct judgment and even disrupting normal social order. (3) Malicious guidance by "opinion leaders" triggers a "butterfly effect." Some reactionary and separatist forces appear on the network in various guises, creating and exploiting online rumors to incite radical emotions, making online public opinion exceptionally complex.

3 Requirements for Online Public Opinion Monitoring

The characteristics of online public opinion determine that public opinion information work is extremely time-sensitive, with opinion dynamics often changing on an hourly basis. To compile public opinion data, one must browse and search through massive amounts of online information, including news reports, related comments, online forums, blogs, etc., to extract public opinion information related to events and then analyze its temporal and spatial distribution.

To improve the timeliness of public opinion work, modern network technology must be fully utilized for timely and effective information collection, processing, analysis, feedback, and decision-making, which requires strong technical support.

From a business requirements perspective, online public opinion monitoring includes two modes: daily monitoring and emergency monitoring. Daily monitoring refers to the continuous, uninterrupted conduct of online public opinion monitoring as a routine departmental function to keep abreast of the direction, characteristics, and trends of online discourse. The significance of daily monitoring lies in constantly understanding the dynamics and direction of online public opinion, enabling timely measures to guide daily public opinion and provide decision-making support for relevant departments when negative or major false public opinion signals are detected.

Emergency monitoring refers to the monitoring of relevant online public opinion when mass emergencies occur. Such events involve many variable factors, complex internal relationships, and unpredictable development trends, with vast and complicated information that greatly increases the difficulty of information judgment and decision-making. Moreover, because conflicting parties in emergencies are often in opposition, normal information communication channels may be affected or obstructed, creating opportunities for various "hearsay" to fill the information vacuum. Such incidents are highly sudden, have significant social impact, and allow decision-makers little time to think. Without timely and accurate access to the latest information for analysis and processing, the consequences can be extremely serious. Therefore, when emergencies arise, a comprehensive public opinion monitoring mechanism and timely, effective collection and analysis of public opinion information to fully grasp all information closely related to the event are extremely important.

By establishing mature mechanisms for daily online public opinion monitoring and emergency response, negative public opinion can be guided and controlled at its nascent stage. When major emergencies occur, various forces can be mobilized and integrated within a short time to form synergy for crisis response. After crisis events, effective evaluation can be conducted, including assessment of the crisis situation, measures taken, projections for the next phase, and summaries, reflections, and recommendations on the previous response, thereby continuously improving online public opinion monitoring and crisis response capabilities.

Online public opinion monitoring must be conducted among the massive amount of information on the Internet, with extremely high timeliness requirements that make manual browsing methods inadequate for collecting and processing the vast sea of online information. This necessitates support from modern information network technology, particularly automated computer software systems.

Among relevant technical tools, search engines provide general information retrieval services that can offer some assistance to public opinion work. However,

their main limitations are inherent restrictions on the timeliness and depth of data collection, making it difficult to accurately meet user needs. Valuable public opinion clues are often buried in large numbers of search results, and search engines do not provide statistical analysis of information, failing to meet the in-depth analysis requirements of public opinion monitoring in a timely manner. Additionally, search engine results are subject to intervention by regulatory authorities, policies, and commercial interests, often making desired results unsearchable and creating blind spots in public opinion monitoring. Vertical search engines, such as news search, forum search, lifestyle search, financial search, image search, etc., although relatively timely in information updates, still cannot provide statistical correlation analysis of public opinion information, nor can they support multi-user collaborative work and resource sharing. Daily monitoring is already heavily burdened, and when emergencies occur, it becomes even more difficult to integrate various forces to form synergy and respond to online crises.

4 Tianji Public Opinion System Solution

[Figure 1: see original paper]. Basic Architecture of the Tianji Public Opinion System

The Tianji Public Opinion Monitoring System is specifically designed for the operational requirements and characteristics of online public opinion monitoring. It employs not only specialized search engine technology but also more intelligent data mining techniques to establish a daily-cycle network public opinion monitoring platform, supplemented by weekly or monthly public opinion analysis reports, thereby providing a convenient, scientific, and actionable platform for public opinion work.

The Tianji Public Opinion System utilizes technologies including information collection, information extraction, Chinese word segmentation, full-text retrieval, automatic deduplication, association analysis, keyword extraction, automatic summarization, classification and clustering, and personalized information proactive push. It can intelligently retrieve, process, and analyze massive amounts of information, providing users with services such as hot topic monitoring, personalized information push, and precise full-text retrieval. The system delivers objective, complete, and accurate public opinion monitoring reports that facilitate data and chart-based interpretation of public opinion, reducing subjective judgment errors.

The Tianji Public Opinion System adopts the three-layer software architecture shown in Figure 1, which can meet the following requirements of public opinion monitoring:

1. **Improving Efficiency and Expanding the Scope of Public Opinion Monitoring**: Manual monitoring of online information is not only slow and inefficient, requiring substantial human and material resources, but also cannot comprehensively cover all hot topics and subjects of concern.

By leveraging the Tianji System' s information collection technology, the entire network and thousands of designated channels can be monitored 24/7 without repetition or omission, undoubtedly increasing the quantity and scope of public opinion monitoring while improving quality.

2. **Real-Time and Accurate Public Opinion Monitoring**: Through 24/7 directional monitoring, any latest information on the network can be promptly discovered and identified. Through judgment and analysis by the monitoring knowledge base, alarm messages are generated immediately upon detecting issues and are pushed to management departments in real time, enabling them to grasp the outbreak points and development trends of public opinion promptly and truly serving as a "network supervisor."
3. **Calm Response to Emergencies**: When encountering emergencies and major topics, the system automatically tracks the original posting address, analyzes statistical trends and dissemination paths, and can continuously monitor post-deletion status and active online users 24/7. It analyzes multiple viewpoints and opinions from netizens, thereby effectively grasping the public opinion situation and generating synergy among various forces through authority linkage.
4. **Flexible Generation of Evidence Databases and Public Opinion Reports**: The system can permanently preserve key information by generating evidence databases; it can produce daily briefings, monthly special reports, and rapid reports on important matters. It can also automatically generate hot topic public opinion reports focused on social 舆论 hotspots. The reports are richly illustrated with charts and provide extensive statistical analysis and netizen viewpoint analysis, offering significant assistance for decision-making support. After task completion, disposal result reports can be generated to reflect work effectiveness and efficiency.

With the assistance of the Tianji Public Opinion System, monitoring the entire network can be achieved by focusing only on a small amount of intelligently pushed information, making public opinion work both efficient and engaging.

[Figure 2: see original paper]. Tianji Public Opinion System Monitoring Public Opinion Information

5 Features of the Tianji Public Opinion System

The Tianji Network Public Opinion Monitoring System integrates more than ten years of research achievements from the Institute of Computing Technology, Chinese Academy of Sciences, in the field of deep content mining technology, incorporating multiple patented technologies. Its predecessor, the Internet Intelligent Search and Mining System, won the National Science and Technology Progress First Prize. Its multi-document summarization, web and blog expert search, and information filtering technologies have successively won first place in the Text Retrieval Conference (TREC), a world-class competition in infor-

mation intelligent processing. Its Chinese word segmentation technology ranks first in domestic and international public evaluations and is recognized internationally as the top brand in Chinese language analysis. Its topic detection and tracking technology won second place globally in the international Topic Detection and Tracking (TDT) evaluation, among other honors. These advanced technologies bring four major advantages to the Tianji Public Opinion System: comprehensiveness, efficiency, accuracy, and depth, as embodied in the following four aspects:

1. **Comprehensive Information Acquisition**: Information acquisition refers to the high-speed and accurate collection of data from the network and extraction of relevant metadata. The Tianji Network Public Opinion Monitoring System can quickly and efficiently acquire web pages, supporting various complex page formats including Javascript, automatic web encoding recognition and conversion, cookie-based state detection, and anti- "anti-refresh" collection mechanisms. It can directionally collect and extract complex information content from news, forums, blogs, and other sources. It also supports meta-search-based thematic collection, capturing only user-interested content from major search engines with comprehensive information and timely updates.

2. **Natural Language Understanding**: The goal of natural language understanding is to enable computers to truly comprehend language semantics in various complex types of information, providing a credible knowledge basis for further in-depth mining. The Tianji Public Opinion System embeds a powerful Chinese lexical analyzer ICTCLAS and integrates efficient body text and keyword extraction technology that can remove noise from web pages and automatically calculate representative keywords. Its text classification and clustering algorithms are fast and accurate, and its document summarization can automatically analyze document content to provide concise and accurate text summaries.

3. **Intelligent Information Search**: Intelligent information search provides more intelligent, specialized, and humanized information search services. The Tianji Public Opinion System employs the distributed full-text retrieval system I3Search, embedding the latest research results in query understanding to automatically mine potential semantic associations. Its kernel is carefully designed as a perfect combination of high scalability and high performance. The system supports efficient indexing of text, numeric, date, string, and other data types; supports rich query languages; and supports both 32-bit and 64-bit hardware platforms with mainstream operating systems including Windows and Linux. It achieves indexing speeds up to 9MB/s, supports online indexing, and realizes millisecond-level query response.

4. **Comprehensive Public Opinion Mining**: This provides one-stop, all-around monitoring and browsing of monitored information sources. Comprehensive public opinion mining addresses the growing needs for public

opinion monitoring, competitive intelligence, and crisis public relations on the Internet. It performs comprehensive mining and analysis of information collected from the Internet, including forums, blogs, news, search engines, news comments, follow-up posts, images, audio, and video, to achieve automatic topic discovery and all-around tracking and tracing. It provides temporal and spatial distribution and trend analysis, intelligent analysis of the tendency of articles and comments, deep mining of relationships among network objects, and pushes valuable public opinion information and statistical reports, providing public opinion monitoring and crisis public relations response services.

The Tianji Network Public Opinion Monitoring System architecture offers excellent scalability and can be flexibly customized according to requirements. It has achieved good application results among users in government, finance, education, and other industries. The Tianji System has been widely applied in key departments such as the Ministry of Industry and Information Technology, the State Administration of Radio, Film and Television, and the China Securities Regulatory Commission (CSRC), playing a practical role. Notably, the network information monitoring system built for the CSRC has significantly advanced its public opinion monitoring work, providing effective support for maintaining capital market stability and protecting the interests of small and medium investors, and winning the second prize of the 2009 Securities and Futures Industry Science and Technology Award.

References

[1] China Internet Network Information Center. The 24th Statistical Report on Internet Development in China (July 2009) [OL].

[2] Lai-Hua Wang. Public Opinion Research Concepts—Theory, Methods, and Realistic Hotspots [M]. Tianjin: Tianjin Academy of Social Sciences Press, 2003.

[3] Public Opinion Information Bureau, Publicity Department of the CPC Central Committee. Introduction to Public Opinion Information Work [M]. Beijing: Xuexi Press, 2006.

[4] Ling-Di Lou, Xiao-Bin Zhou. Analysis of the Operational Status of Online Public Opinion [J]. Theory and Practice of Special Zones, 2009.2.

Author Biographies

Li-Dong Zhai: Assistant Researcher, Network Key Laboratory, Institute of Computing Technology, Chinese Academy of Sciences.

Zhi-Hua Yu: Deputy Director and Senior Engineer (Associate Researcher Level), Network Key Laboratory, Institute of Computing Technology, Chinese Academy of Sciences. Email: yzh@ict.ac.cn.

Guo-Dong Ding: Senior Engineer (Associate Researcher Level), Network Key Laboratory, Institute of Computing Technology, Chinese Academy of Sciences.

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