

# Application of Big Data Technology in Product Quality and Safety Risk Information Monitoring

Chenyi Zhang\*, Fang Liu, Leiming Li

Shandong Institute of Product Quality Inspection, Jinan 250102, China.

---

**Abstract:** Product quality and safety risk monitoring with “risk management” as the core thought has been gradually introduced into the market supervision system. This paper discusses the application method of risk information collection and analysis in product quality risk monitoring through the application practice of big data technology in quality monitoring. Through the emotional analysis and judgment of the article, the quality risk signal is intelligently excavated to provide technical means and monitoring methods for product quality risk monitoring.

**Keywords:** Big data; Product quality and safety; Risk monitoring; System

---

## Introduction

With the establishment of the State Administration for Market Regulation, the pattern of “big market, high quality, and strict supervision” has initially taken shape. Under the new development pattern, we are accelerating the construction of a new quality and safety supervision system based on safety assessment, with classification supervision as the starting point, credit supervision as the support, and intelligent supervision as the adaptation<sup>[1]</sup>. As a result, the work form of product quality and safety monitoring is becoming increasingly complex.

Product quality and safety risk monitoring is a forward-looking and proactive work, with the main task of monitoring and evaluating industry-specific issues and potential hazards that affect product safety. Monitoring and processing information on product quality issues has clearly become an effective means of product quality risk monitoring<sup>[2]</sup>. With the rapid development of market environment and social informatization, online public opinion intelligence has gradually become the main platform for information exchange among numerous consumers and organizations. By the end of December 2020, the number of internet users in China reached 989 million, with an increase of 85.4 million compared to March 2020, and the internet penetration rate reached 70.4%, with an increase of 5.9 percentage points compared to March 2020<sup>[3]</sup>. The trend of mainstream internet media is unstoppable. By utilizing technologies such as big data and conducting effective information collection and analysis, a weather forecast for product quality and safety can be formed. Early detection, assessment, warning, and disposal of quality and safety risks are becoming increasingly important.

Based on the current status of product quality and safety, this paper solves the following five problems for the daily supervision of the quality inspection system through big data analysis and judgment:

- » Product quality risk monitoring in various industries;
- » Targeted product sampling;
- » Immediate recall of defective products;
- » Enterprise credit rating evaluation;
- » Quality inspection work such as instant exposure affecting poor quality products.

The main objective of this paper is to enhance the monitoring and response capabilities of market supervision towards enterprise product quality public opinion, and effectively increase the tracking of product quality and safety risks through online public opinion supervision. The second is to timely expose the negative impact of enterprise products through real-time public opinion, and shape a clean and standardized product safety supervision environment. The third is to timely discover and recall defective products with serious problems through industry public opinion search, strictly enforce the company’s safety management regulations, and protect consumer rights. The fourth is to enhance communication with the public through reasonable public opinion exchange, shape the image of market supervision, and create a good public opinion atmosphere for quality and safety.

# 1. Construction of Intelligence Data Center

## 1.1 Data Center Facilities

The entire application system of the data center will be implemented on the J2EE platform and can run on any operating system and middleware servers that support the J2EE standard. At the same time, the system adopts a persistence framework based on J2EE to implement database access operations, ensuring that the system can run on any mainstream database system.

The NOC (Network Operation Center) network monitoring center in the data center can provide continuous monitoring from 7\*24, regularly scanning customer servers on the network to determine whether the specified service port can be accessed. When the specified port service cannot be accessed, the monitoring alarm system will automatically notify the technical experts in the data center in various ways based on the predetermined parameters and fault level. The system administrator can perform fault recovery operations for the client server in the first time, including restarting the server, restarting the specified service, changing the IP address, etc.

## 1.2 Data Center Collection Capability

Hundreds of servers distributed in data collection points across the country adopt a distributed collection method, continuously collecting and processing data 7\*24 hours a day. The total number of data sources collected is about 50000, including news, forums, blogs, Weibo, foreign websites, and flat media electronic newspapers. The daily collection total exceeds 3 million, and Weibo collects more than 10 million pieces of data. At the same time, hundreds of sets of information analysis tools are used to perform article cleaning, deduplication, classification, and other operations on the collected and stored information, achieving sub second level retrieval speed on G-level datasets. The average data index update time is less than 0.02/s per record (4Kb per record), and the index space expansion rate of the full-text retrieval database is less than 0.5. The powerful data center centralized processing ensures the real-time nature of the collected data and the accuracy of filtering. Through the collaborative work of data center servers, invalid and interfering information is automatically filtered, and effective information is automatically pushed to the end user database, making it easy for users to query or access data on private cloud platforms.

# 2. Construction of Product Quality Problem Intelligence Analysis Software

## 2.1 Data Collection

### 2.1.1 Tieba and Forum Collection

The system fully supports forum and post bar information collection, supporting two working modes, one-time collection and tracking collection. One time collection is similar to news, where each post is not tracked after collection, and only newly added main posts are accessed for the next collection. The advantage of this strategy is high performance, and the main content is generally not missed because important content is usually published in the main post.

The tracking and collection of information, in addition to collecting new main posts, will further examine whether there are any new replies collected within a certain period of time. If there are further new replies, they will be separately captured. This ensures comprehensive information without any omissions. But for large forums, it takes a long time because there may be many historical posts, and each time it needs to traverse all the posts, which takes a long time and occupies a lot of storage space.

The collection supports SSO. Large forums in China may span multiple servers, and the system supports SSO access between multiple forums. Although this feature is not obvious to users, it can avoid duplicate login and optimize performance when accessing forums.

The overall configuration is simple and fast, convenient for users to maintain, supports visual configuration, and is easy and convenient for users to add new forums.

### 2.1.2 Weibo Collection

To use a unique JavaScript scripting engine to parse, which are supported by the three major platforms of NetEase, Sina, and Tencent, and support dynamic password login.

Supporting the collection of information on Weibo index pages, as well as specific personnel. The frequency of information collection on Weibo can be set separately to adapt to the rapid changes in Weibo.

Suiting Weibo image collection, if images are also published, they can be collected together with text.

### 2.1.3 Collection of Blogs and News Comments

Blogs and news comments can be crawled (including Ajax comments from Sina, NetEase, etc.), and can also be continuously tracked like forums.

The system also supports one-time collection and tracking collection, which only captures web page snapshots at the time of collection. Tracking collection can continuously track changes in news comments for collection, and the system intelligently recognizes the pagination of the comment system, reducing the workload of configuration.

The News Review Department provides an important parameter for evaluating the popularity of articles through full-text search or when viewing specific articles. By collecting news comments, the system has a constantly changing parameter, which greatly improves the accuracy of article popularity evaluation.

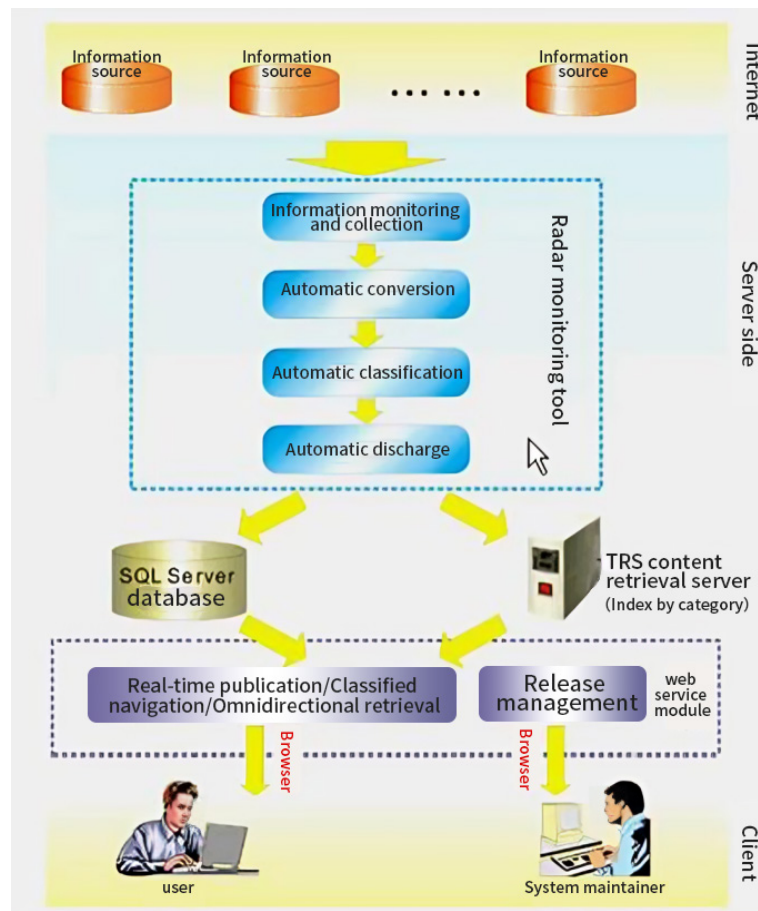


Figure 3-1 Construction diagram of data collection system

## 2.2 Data Storage

After collecting web pages, the main content and intelligent summary can be extracted. Based on this, one or more pieces of information can be selected and exported in batches to Word, Excel, text files, and other files, which can be better utilized by information processing personnel.

The system supports saving snapshots of collected information, and supports saving data on multiple servers. Data access adopts a distributed access method. Through the support of distributed storage, the system has greater horizontal scalability, thus supporting the storage

of data for longer periods of time.

### 2.3 Data Analysis

The information collected from literature needs further processing in order to achieve multiple applications. The system needs to seamlessly integrate functions such as denoising, automatic classification, automatic deduplication, internal code conversion, index semantic calculation, and automatic summarization of information within the system, achieving high integration of automatic processing.

### 2.4 Emotional Analysis

The sentiment analysis model is a backend operation model for information processing, which is developed separately based on the mining function of text mining systems. It is an analysis model for scoring and assigning sentiment values to articles for public opinion information. This model divides news, forum, blog information, electronic newspaper articles, microblog articles, WeChat official account news and other media information into words, matches the emotional value of the emotional words in the article with the emotional lexicon, and then calculates the emotional tendency of the article through the scoring model.

Assuming that there are a total of X emotional words in the article, which appear in Y positions, each position appearing N times, then  $F(C_i)$  represents the basic emotional strength of emotional word i,  $F(S_{ij})$  represents the weight of emotional word i appearing in position j, and  $F(T_{ijt})$  represents the weight of emotional word i appearing t times in position j. The values of the weights for each segment are shown in the table below. Finally, the sentiment values are weighted based on dimensions such as intensity, position, and frequency to obtain the final sentiment value of the entire article. The formula is as follows:

Table 1 Values of Weights for Each Segmentation

|              |          |           |                     |                      |     |
|--------------|----------|-----------|---------------------|----------------------|-----|
| Weigh (1-10) | C1       | C2        | C3                  | .....                |     |
| Position     | Title    | First row | First paragraph     | First two paragraphs | End |
| Weigh        | 10       | 8         | 5                   | 4                    | 3   |
| Frequency    | One time | Two times | Three or more times |                      |     |
| Weigh        | 3        | 5         | 10                  |                      |     |

Table 2 Sample Table of Corresponding Weights for Product Quality Words

| Positive                           | Weigh | Negative                | Weigh |
|------------------------------------|-------|-------------------------|-------|
| Excellent quality                  | 7     | Crudely made            | -2    |
| High quality products              | 7     | Irresponsibility        | -2    |
| high-grade, precision and advanced | 7     | Quality Risk            | -2    |
| standard specification             | 7     | Problem Product         | -2    |
| Focus on quality                   | 7     | The lack of supervision | -2    |
| Finely crafted                     | 7     | Gap                     | -2    |
| Craftsmanship spirit               | 6     | Serious problem         | -2    |
| Industry leading                   | 6     | Backward technology     | -2    |
| Strict regulatory measures         | 6     | Work errors             | -2    |
| Responsible                        | 6     | Jellyfish workshops     | -2    |
| Good techniques                    | 6     | Negligence              | -2    |
| ...                                | ...   | ...                     | ...   |

### 2.5 Theme Determination and Rule Determination in Sentiment Analysis

In the process of sentiment analysis of internet information in the product quality testing industry, it is necessary to determine the topic vocabulary of sentences in the article according to the expression rules of Chinese semantics, and objectively judge the emotional tendency of specific articles through the expression of sentence structures.

During the analysis process, the positively correlated thematic vocabulary and sentence structure classification vocabulary are summarized as follows:

Table 3 Example of organizing thematic vocabulary in the quality inspection industry

| Key words                         | Word types             | Key words                | Word types                        |
|-----------------------------------|------------------------|--------------------------|-----------------------------------|
| Regulatory agencies               | Regulatory agencies    | Supervision department   | Regulatory agencies               |
| Enterprises                       | Regulated institutions | manufacturing enterprise | Regulated institutions            |
| Stores                            | Regulated institutions | Circulation enterprises  | Regulated institutions            |
| E-commerce                        | Regulated institutions | Circulation enterprises  | Regulated institutions            |
| Market Supervision Administration | Regulatory agencies    | Supervision department   | Regulatory agencies               |
| Food and Drug Bureau              | Regulatory agencies    | Supervision department   | Regulatory agencies               |
| Consumers' association            | Regulatory agencies    | Supervision department   | Regulatory agencies               |
| Quality Inspection Institute      | Regulatory agencies    | Sales enterprise         | Inspection technical organization |
| Testing organization              | Regulatory agencies    | person responsible       | Inspection technical organization |
| ...                               | ...                    | ...                      | ...                               |

Table 4 Summary of Word Patterns

| Useful words | Word types | Useful words | Word types |
|--------------|------------|--------------|------------|
| 虽然           | ALTHOUGH   | 况且           | BUT        |
| 尽管           | ALTHOUGH   | 如而           | BUT        |
| 虽            | ALTHOUGH   | 想不到          | BUT        |
| 虽说           | ALTHOUGH   | 可是           | BUT        |
| 虽是           | ALTHOUGH   | 但是           | BUT        |
| 固然           | ALTHOUGH   | 却            | BUT        |
| ...          | ...        | ...          | ...        |
| 经            | CP         | 来讲           | CU         |
| 纵然           | CP         | 那样           | CU         |
| 自从           | CP         | 似的           | CU         |
| 趁            | CP         | 得            | CU         |
| 沿着           | CP         | 来说           | CU         |
| 向            | CP         | 来看           | CU         |
| 像            | CP         | 来看           | CU         |
| 跟            | CP         | 一样           | CU         |
| ...          | ...        | ...          | ...        |
| 摆脱           | NOT        | 充满           | VE         |
| 别            | NOT        | 参加           | VE         |
| 不必           | NOT        | 夺得           | VE         |
| 无法           | NOT        | 发生           | VE         |
| 不            | NOT        | 倍受           | VE         |
| 无            | NOT        | 属于           | VE         |
| 莫            | NOT        | 遭遇           | VE         |
| 没有           | NOT        | 惨遭           | VE         |
| ...          | ...        | ...          | ...        |

Training and learning sentence combinations in sentiment analysis:

The combination of word types in a sentence is the foundation of generating patterns, and by annotating the results of the combination, a pattern is generated; A combination file is a file that contains a large number of combinations.

The pattern of sentence structure consists of combination and result.

The format of the sentence pattern is "Combination==>Entity 1, Entity 2, Relationship"; The combination is the generalized sentence, "Entity 1, Entity 2, Relationship".

As shown in formula (1) above, the sentiment value of the article is obtained through weighted analysis.

### 3. System Construction Planning

#### 3.1 Overall Technical Route

This system is designed based on the SOA architecture, and the software technology used complies with the J2EE technology standards. It is designed and developed using a B/S structure. The entire system has good cross platform characteristics, supports mainstream UNIX, Linux, Windows series platforms, supports mainstream middleware such as weblogic, Websphere, and Tomcat, and supports mainstream databases such as Oracle, SQL Server, and MySQL. The application layer is developed using technologies such as Java and JSP. To facilitate data exchange between systems, the system provides support for XML and WEB Services.

#### 3.2 System Deployment Planning

The system adopts a modular design, mainly composed of Internet information collection system, data receiving module, Internet information storage system, content intelligent analysis and processing module, network public opinion information portal and other functional modules. Each module can be independently deployed on different servers or centrally deployed on high-performance independent servers, and can be deployed across operating system platforms. According to business requirements and network status, and based on the design concept of multi-layer architecture, this topology diagram 4-2 is divided into internet access area, application and data service area, etc.

#### 3.3 Business Workflow

The work goal of the public opinion system is to extract content related to the unit's business from a massive amount of internet information, use computer intelligent processing systems to automatically analyze, classify, and summarize these contents based on the characteristics of the unit's business, and then combine the results with manual judgment to timely extract hot and negative public opinion information and submit it to the leadership for decision-making. And further process this information according to the leadership's instructions or the predetermined public opinion response plan. The workflow of the entire system is roughly shown in the following figure:

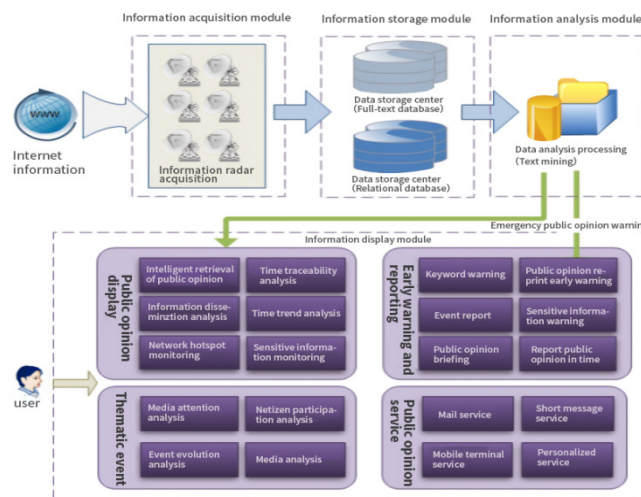


Figure 1 System flowchart

### 4. System Advantages

#### 4.1 Technical Maturity

The public opinion analysis and fortress host system involved in this plan have adopted mature advanced technology, and a large num-

ber of experiments, breakthroughs, and prototype building have been carried out on the key technologies of the system in the early stage of work, and have been well applied in the customer environment. Moreover, the hardware and software platforms selected by the system are mature products with good technical support and development prospects.

## **4.2 Reliability of Business Requirements**

We conducted detailed requirement research and business modeling to ensure that the business management mode and content (including job settings, work procedures, and requirements) comply with the current laws and regulations in China; Moreover, a modular and platformized application software that organically integrates both centralized and distributed modes can not only meet various needs of network security management, but also better solve the expansion and changes that business needs may face, with good universality and flexible expansion and upgrading capabilities.

## **4.3 Scalability**

On the premise of ensuring the unity of system interfaces and the integrity of system structure, this system advocates implementing each link of the system step by step under the premise of overall design, ensuring the design of interface unity and system structure integrity, avoiding drawbacks such as redundant construction, investment, mismatched system interfaces, and disconnected business relationships.

## **4.4 Usability**

There are not special requirements for browsing. Convenient for various operators to achieve complete or partial automation of some business operations. The system should have a consistent and user-friendly user-friendly interface, with practical operability, allowing users to quickly grasp the use of the system.

This system can achieve the function of rapid deployment. It is possible to define, design, and implement application structures and functions in the shortest possible time.

Zero client maintenance: The entire system adopts a B/S structure, and all data and applications are maintained uniformly on the server side. The user side can complete all operations as long as it supports a browser, including information collection, analysis, warning, briefing, and system management.

Easy to operate: adopting advanced system design concepts, application design follows the principle of simplicity and practicality, operators can operate with simple training, and what they see is what they get; Provide a convenient backend management system, with flexible and simple system management and page style adjustment. Management personnel do not need to write code or create complex pages to maintain and manage the system.

## **5. System Application Effects**

This plan proposes suggestions for measures to control public opinion on the external internet of the quality supervision system and efficiently operate and maintain internal IT resources. Combining with the characteristics of the quality inspection industry, the comprehensive ability of the quality inspection industry under the wave of informatization will be further improved through the public opinion analysis system, and the following construction effects will be achieved:

### **5.1 Enhancing Emergency Response Capabilities for Major Safety Incidents**

Through the massive collection, processing, and extraction of internet information through the public opinion analysis system, it helps quality inspection departments to timely, comprehensively, and accurately grasp product and industry security trends, and improve their emergency response capabilities.

## **5.2 Enhancing the Ability to Monitor Product Quality and Safety Risks in the Industry**

By utilizing core functions such as system public opinion collection, public opinion analysis, public opinion warning, public opinion search, public opinion briefing, and statistical analysis, we can track, monitor, and collect dynamic public opinion information of target websites, forums, Weibo, and blogs 24/7 online. With the vast public opinion resources, we can timely and accurately identify industry product quality and safety risks, and improve risk monitoring capabilities.

## **5.3 Improving the Speed of Exposure and Recall of Adverse Products Due to Network Impact**

Timely detection and handling of product quality issues raised by consumers, truly addressing product safety issues in their early stages.

## **5.4 More convenient solutions to public issues, creating a harmonious product safety atmosphere**

The public opinion monitoring system uses intelligent means to explore and analyze information on public response issues, enabling the product quality and safety supervision and management system to conduct in-depth data mining and timely processing of public response issues. To safeguard the immediate interests of consumers, cultivate good and smooth communication channels with the public, and create a harmonious product safety atmosphere.

## **6. Conclusion**

Text data is the carrier of information, and big data technology provides scientific means and tool support for information analysis and processing. Through emotional analysis of the article, it was found that product quality risk signals are an effective way to rely on existing big data technology to serve quality inspection work. This method can help regulatory personnel to timely explore and effectively follow up on the products of enterprises with large quantities and serious quality problems, and strengthen quality and safety sampling efforts in a targeted and efficient manner.

## **References**

- [1] An Sendong. Market Regulation Escorts High Quality Development [EB/OL], November 9, 2020, <http://www.rmlt.com.cn/2020/1109/598294.shtml>.
- [2] Gong Chunling. Analysis and Exploration of Product Quality and Safety Risk Information Monitoring [J], China Quality and Standards Introduction. 2019; (12): 62-65.
- [3] China Internet Network Information Center (CNNIC), 47th Statistical Report on the Development of Internet in China [EB/OL], 2021-02-03, [http://www.gov.cn/xinwen/2021-02/03/content\\_5584518.htm](http://www.gov.cn/xinwen/2021-02/03/content_5584518.htm).