Construction of a Diversified Product System Based on a Public Opinion Data Middle Platform: A Case Study of Southern Public Opinion (Postprint)

Authors: Wu Xian, Xiao Zhuoming, Hong Dan

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

Abstract

As the public opinion market continues to expand, how to flexibly respond to variable and diverse user requirements and rapidly generate diversified product services has become a critical issue. This paper, grounded in public opinion application scenarios, constructs an intelligent data middle platform through standard specification definitions and service encapsulation and orchestration, which bridges technology and leads business operations while enabling rapid connection and extraction. This platform efficiently satisfies front-end data analysis and product service needs, guiding the development of public opinion business toward deeper levels.

Full Text

Building a Diversified Product System Based on a Public Opinion Data Middle Platform: A Case Study of Southern Public Opinion

Abstract: As the public opinion market continues to expand, a critical challenge emerges: how to flexibly respond to variable and diverse user needs while rapidly generating diversified product services. This paper, grounded in public opinion application scenarios, constructs an intelligent data middle platform through standardized specification definitions and service encapsulation orchestration. This platform bridges technology and business, enabling rapid connection and extraction to efficiently meet front-end data analysis and product service needs, thereby guiding the deep development of public opinion business.

Keywords: public opinion; data middle platform; data modeling
Classification Number: TP393

Document Code: A
Article ID: 1671-0134(2019)01-119-03
DOI: 10.19483/j.cnki.11-4653/n.2019.01.033

By Wu Xian, Xiao Zhuoming, Hong Dan

In recent years, traditional media have continuously sought paths for convergence and transformation, expanding "Media+" services to create greater value for users. To build a new pattern of public opinion guidance, an increasing number of traditional media have integrated brand resources, political and economic resources, and information resources to enter the public opinion service domain. As the public opinion market expands, diverse needs for public opinion products and services have emerged among government and enterprise users, between provincial-level government users and district/county grassroots government users, and even among local government users and functional department users. The same user also demonstrates highly variable needs for public opinion management across different contexts. When these variables accumulate to a certain volume, customized development for each user becomes prohibitively expensive while product efficiency declines. This paper draws from the actual business development of Southern Public Opinion, learning and practicing Alibaba' s "large middle platform, small front-end" concept to introduce the operational thinking of a public opinion data middle platform. This approach supports the rapid generation of diversified product applications, creates a comprehensive package of public opinion product services, and seizes development opportunities in the public opinion market through "technology cost reduction, application efficiency improvement, and business enablement."

1. Challenges and Difficulties

Facing complex public opinion application scenarios and breaking through traditional system architectures, building a public opinion data middle platform that addresses diverse and changing user needs presents numerous technical challenges.

Challenge One: Comprehensive Data Collection and Storage. Driven by demand, how to real-time collect and introduce multi-channel (websites, forums, blogs, apps, Weibo, WeChat public accounts, radio and television) and multi-form (internal business systems, internet collection, third-party exchange) data to build a multi-source, massive, and dynamic foundational data pool poses significant challenges.

Challenge Two: Standardized Data Architecture and Development. How to construct a layered and horizontally decoupled data structure that achieves architectural uniformity, reliability, and flexibility through standardized data formats and standardized interaction interfaces to rapidly support upper-layer data applications and services is a technical issue worthy of exploration.

Challenge Three: Cross-Domain Data Integration and Knowledge Accumulation. How to establish a fusion model that achieves cross-domain public opinion data integration through multi-dimensional modeling, while excavating public opinion data from individual tagging to global indexing, deeply extracting data value, and achieving knowledge precipitation for common applications, is key to providing foundational capabilities for public opinion business support.

Challenge Four: Data Encapsulation Application and Service Opening. The scaled development of data is about providing service-oriented capabilities. How to encapsulate services according to application requirements, open them to external service users through diversified product forms, achieve rapid sharing of data value, and bridge the last mile to service users represents the ultimate goal of building a public opinion data middle platform.

2. Technical Architecture and Key Technologies

The concept of the data middle platform was first proposed by Alibaba as "building a standardized, comprehensively connectable and extractable, intelligent data processing platform." Its construction goal is to efficiently meet front-end data analysis and application needs. To address the complex variability of public opinion service demands, Southern Public Opinion has designed and built a public opinion data middle platform from actual business operations, aiming to achieve rapid deployment of customized and personalized public opinion products. The overall architecture and key technologies are described in [Figure 1: see original paper].

2.1 Public Opinion Data Collection: Intelligent Collection and Storage of Comprehensive Data

The comprehensive data intelligent collection platform primarily connects to data forms including internet data collection, cooperative complementary data, exclusive offline media sources, and internal editorial business data. Internet data employs distributed crawlers, intelligent collection scheduling, adaptive collection strategies, data collection proxies, and automatic login verification technologies to flexibly configure collection rules, crawling depth, scanning frequency, and other collection strategies, achieving unified collection management of various channel data sources. Relying on distributed architecture, multi-point load balancing, and adaptive bandwidth design ensures real-time collection efficiency, stability, and data integrity.

Distributed computing architecture enables rapid identification and information crawling of massive-scale data, applying different crawling strategies to different information types to achieve automation of internet information crawling. The system employs distributed multi-threaded concurrent instruction execution architecture, incremental real-time indexing, intelligent word segmentation, and other technologies, achieving high collection and data management efficiency.

It enables concurrent crawling of multiple websites simultaneously, distributed concurrent multi-point processing of a single task, and multi-point load balancing effects. This prevents sending excessive access requests to the same website within a short period, improving the efficiency and performance of large-scale data collection. Using IP proxy pools and API imitation mechanisms, the system performs IP rotation collection on high-frequency updated data, effectively preventing website restrictions on system IPs while intelligently and proactively reducing collection frequency to decrease the possibility of IP blocking. Intelligent scheduling of distributed collection effectively improves data collection stability.

Self-collected internet data, cooperative complementary data, offline source data, and editorial business data enter the data analysis layer through standardized data interfaces with unified formats, breaking down data silos, solving the multi-source heterogeneous problem of public opinion data, reducing siloed collaboration, and ensuring the diversity and integrity of public opinion data. This provides fundamental support for personalized and customized public opinion products.

2.2 Public Opinion Data Analysis: Data Standardization and Accessibility

Utilizing infinite-level priority nested matching rules with AND (+), OR (|), and NOT (-) operators, and based on efficient indexing and sorting algorithms, the multi-dimensional retrieval and keyword parsing system supports complex combinations of various index conditions to maximize satisfaction of diverse data application encapsulation requirements. Through automatic push scripts, retrieval results are intelligently pushed to facilitate efficient sharing of public opinion data, providing powerful data support for further public opinion business and extended business development, achieving one-time tracking with multi-terminal usage. The push mechanism uses XML Schema specifications as the standard format for data exchange, shielding differences between heterogeneous data sources. Data formats adopt XML/JSON for convenient invocation and strong adaptability.

The system automatically identifies and filters junk information, cleaning advertisements, irrelevant images, hyperlinks, dynamic Flash, and other useless information. It uses intelligent parsing to automatically extract effective information elements such as titles, timestamps, sources, authors, and main text. Through a content deduplication engine, the system performs deduplication and merging based on semantic analysis of data content, automatically identifying duplicate articles to achieve automatic deduplication and consolidation. Distributed storage clusters store processed standardized public opinion data, snapshots, and indexes, achieving integrated management of structured and unstructured data resources. The analyzed and standardized public opinion data provides a calling foundation for public opinion service applications, responding to basic public opinion business needs through service interfaces.

2.3 Public Opinion Data Modeling: Multi-Dimensional Tagging and Indexing of Data

Deep mining of massive public opinion data utilizes intelligent parsing and matching of keyword regular expressions to extract event-related information, performing multi-dimensional analysis on hotspot event information. The system processes event information through word segmentation, sentiment analysis, heat analysis, high-frequency word extraction, correlation analysis, and data statistics. Combined with intelligent operations such as automatic summarization, classification, and clustering, it analyzes event development trends, sensitivity indices, geographic distribution, propagation paths, key figures, positive/negative tendencies, and netizen viewpoints, enabling in-depth analysis of event essential causes and forming foundational data for modeling.

The system categorizes data sources through intelligent tagging, logically combining data sources into arbitrary different virtual data source packages. During data retrieval, matching can be performed either within the global data source or within virtual data source packages according to different user needs, narrowing data retrieval scope, improving retrieval precision, and enhancing retrieval efficiency. This achieves flexible deployment of public opinion data retrieval and rapid response to business environment changes and business process optimization requirements.

2.4 Public Opinion Service Applications: Data Application Encapsulation and Service Opening

Based on public opinion data modeling, the system encapsulates data through virtual data source packages and keyword parsing systems. Combined with intelligent push development, it creates various public opinion service applications such as real-time alerts, periodic reports, thematic analysis, public opinion consultation, and public opinion assessment. Utilizing multiple release channels including PC terminals, mobile apps, large screens, WeChat, and SMS, it forms a public opinion service application matrix to meet comprehensive public opinion service opening needs.

3. Application Cases

Through application innovation and technological innovation, Southern Public Opinion has built a service system for rapid product generation based on the public opinion data middle platform. The following briefly elaborates on practical application cases demonstrating how the public opinion data middle platform enables business empowerment.

3.1 Social Sentiment Risk Index and Rankings

The Social Sentiment Risk Index is a distinctive product application of Southern Public Opinion based on "data precipitation and business 下沉" (deep business

integration). The product generation logic and technical implementation steps are as follows:

First, collect and aggregate historical risk events to form a connection and tagging system centered on business core objects. Extract elements from risk events including nature, level, and propagation scope, assign values and definitions to each element, and establish a social sentiment risk index calculation model. Second, within a certain period (daily, weekly, monthly, yearly), incremental public opinion data and offline historical data are synchronized and shared. Based on data standards and tagging models, data extraction is performed to feed back into the public opinion data middle platform, forming an online quantified social sentiment risk index. Third, launch the social sentiment risk index ranking product to flexibly evaluate and judge social sentiment conditions across time, geography, attribute, and other dimensions.

3.2 Multi-Dimensional Cross-Comparison and Visualization of Public Opinion

Multi-dimensional cross-comparison and visualization of public opinion is an application case of Southern Public Opinion' s "data assembly and service-oriented application." The system operation interface is shown in [Figure 2: see original paper], and its generation logic and technical implementation steps are illustrated in the figure.

First, integrate comprehensive data to unify data export and query logic, establishing a public opinion situation awareness system. This enables comprehensive public opinion monitoring within Guangdong region while rapidly discovering and displaying evolution trends for service users and emergency events. Second, by reusing public quantitative indicators and processing personalized variable indicators, the system extracts directional indicators such as leadership, development capability, execution capability, and innovation capability, while delving into refined indicators including media attention, social stability maintenance, and business environment. It establishes a user coordinate system and performs user profiling clearly and rapidly through algorithmic model matching. Third, through business application operations, the system timely responds to and visually outputs data collection analysis and indicator model effects (adapting to different presentation carriers). Through threshold settings for different indicator values, automatic alerts are achieved, rapidly completing data encapsulation and application service deployment oriented toward user needs.

The core of the public opinion data middle platform lies in capabilities such as data models, algorithm services, and data products. By building a flexible and rapidly responsive architecture, front-end product requirements are fulfilled more quickly. On one hand, it avoids redundant construction of highly reusable functions; on the other hand, all business touchpoint information can flow to the middle platform, solving data silos and forming information sharing. With the precipitation capability of the middle platform, research and development

become more flexible and business more agile. The next step will involve the phased evolution of the public opinion data middle platform, continuously forming holistic output of "technology platform + construction methodology + data products + operation services" solutions to rapidly adjust and respond to future market changes.

References: [1] Zhong Hua. The Way of Enterprise IT Architecture Transformation—Alibaba' s Middle Platform Strategic Thinking and Architecture Practice [M]. Beijing: Mechanical Industry Press, 2017.
[2] Yu Guoming, Ma Siyuan. Artificial Intelligence Enhances Network Public Opinion Analysis Capability [J]. Network Communication, 2017(2): 85-87.
[3] Ma Mei, Liu Dongsu, Li Hui. Research on Big Data-Based Network Public Opinion Analysis System Model [J]. Information Science, 2016, 36(3): 25-28.

(Author Affiliation: Southern Media Group)

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