

A Cloud Computing-based Television Program Opinion Monitoring and Analysis System

Dingguo Yu

School of New Media, Zhejiang University of Media and Communications, Hangzhou, China
Email: zjydg@163.com

Huxiong Li, Zhiwen Hu

School of New Media, Zhejiang University of Media and Communications, Hangzhou, China
Email: jsj_lhx@126.com, sunneyhu@gmail.com

Abstract—Today, many television viewers are also experienced internet users. People often post reviews of television programs on social networks and have created influential and powerful opinions. To strengthen the monitoring and analysis of these opinions, this paper developed a cloud computing-based internet public opinion monitoring and analysis system for television programs. The purpose of this novel system is to assist the government and television stations with program monitoring in order to provide a decision-making basis for improvement. This paper introduced the new cloud computing-based monitoring and analysis system, including the necessary cloud computing technology, system framework, functional structure, and system workflow.

Index Terms—Television Program, Internet Public Opinion, Cloud Computing

I. INTRODUCTION

This internet public opinion monitoring and analysis service for television programs monitors and analyzes the comments and opinions of internet users to relevant television programs through the monitoring and tracking of websites (including portal websites, industry websites, forums, blogs, and microblogs, etc.). It further investigates and analyzes the viewer base and the social impact of certain television programs, and relays this information to the regulatory and decision-making departments of government institutions and television stations for references.

Due to the development and popularity of the internet, especially the wide spread of social networks, and the fact that most television viewers are also experienced internet users, the discussions and comments regarding television programs through social media such as blogs and microblogs, have developed into a strong influence on television program opinion, which has also further impacted the public opinion. Currently, there are numerous television programs with uneven qualities. A certain number of programs, especially some comments have large influences on the ideology of the viewers. Therefore, it is essential to develop an internet public opinion monitoring and analysis system for television

programs in order to launch the monitoring and analysis of public opinion on television programs, assist the government and television stations with the regulations of television programs, understand the viewer base and social impact of television programs, find problem, divert public opinions, and provide evidences for the decision-making of government and the regulatory and improvement of television programs.

II. RELATED WORK

The key of the television program opinion monitoring technology is topic detection and tracking (TDT), which is an algorithm that finds relevant information from vast amounts of data streams [1,2]. The technology has been well-developed and primarily includes two categories: clustering algorithms [3-7] and repeating string matching [8-10].

Currently, there are no monitoring and analysis systems designed for domestic television programs opinions. There are only some commonly available internet public opinion monitoring and analysis systems which are difficult to use for monitoring and analysis of the opinions specifically directed at television programs. A few commonly used systems include: Goonie internet public opinion monitoring system [11], Junquan internet public opinion monitoring system [12], TRS internet public opinion management system [13], and Founder internet public opinion and information monitoring and analysis system [14].

III. SYSTEM DESIGN

A. Cloud Computing Technology

The cloud computing technology for this television program public opinion monitoring and analysis system primarily involves the following three aspects:

(1) Uses cloud storage technology for program opinion resource data management in order to meet the need for large amounts of data.

(2) Adopts cloud computing technology architecture with a centralized television program opinion monitoring center to uniformly perform information collection and public opinion monitoring and analysis services.

Monitoring and analysis application terminals can be PCs, tablets, or smart phones, which shows the flexibility of the system.

(3) The television program opinion monitoring and analysis is task-driven. Each television program is a single monitoring task, and the system allocates different monitored keywords information to different programs. Targeted monitoring and analysis can be carried out for

various programs and different monitoring and analysis reports will be given.

B. The Design of System Framework and Functional Structure

The framework and functional structure for the cloud computing based television program opinion monitoring and analysis system is shown in Figure 1.

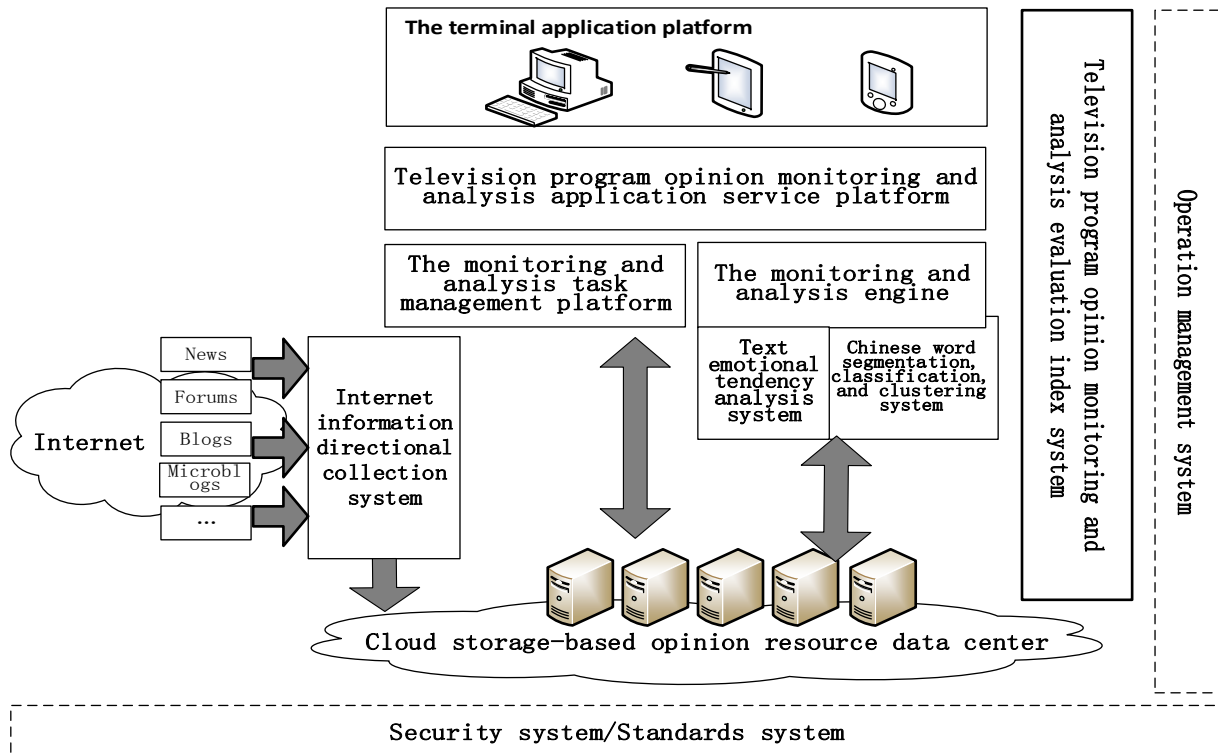


Figure 1. The design of system framework and functional structure

The main functional modules of the system include:

(1) Cloud-storage-based opinion resource data center

The data center primarily stores and manages the television program opinions corpus data. The system adopts a distributed cluster data storage management, which realizes the management technology of big data.

(2) Internet information directional collection system

The internet information directional collection system is the front and foundation of the cloud computing-based television program opinion monitoring and analysis system. This system automatically collects and classifies the information of designated websites and interprets the collected information. Specifically, it performs smart partitioning, and duplicate and noise removal for web pages. This system then identifies the page structure, transforms information into structured data, and stores them to the data center for the follow-up function modules. Additionally, the system will regularly update the collected information in order to perform information tracking.

Directional collection: Limits information collection objects through system configuration. If the collection

sites are set to be “Sina.com” and “Sohu.com”, the system will only collect information from web pages of the two portal websites.

Smart interpretation: Automatically partitions the collected web page files, removes noise, identifies web page structure, extracts the main body of the web pages and other key information (such as titles, article sources, time of publication, and authors.), and converts it into structured data.

Regular updates: Performs regular scans (for example, every 24 hours) on the collected data and examines whether the contents of the collecting sites have been updated. If updated, the original web page data will be substituted with the new content.

Classified collection: In order to increase the efficiency of the collection and the accuracy of the interpretation, the collected objects are classified. The collection and interpretation program can also be adjusted for targeted collection and interpretation. Currently, there are three resources from which to collect: news and blogs, forums, and microblogs.

(3) Television program opinion monitoring and analysis engine

This engine primarily realizes the analysis of opinion corpus data that are generated by an internet information directional collection system and stored in the data center. It's main operations include Chinese word segmentation, classification, and clustering of corpus information in the data center. It also analyzes the emotional tendencies of the data. The engine is a Lucene-based full-text search system that can be used by television opinion monitoring and analysis application service platform for searching, statistics, and invoking.

(4) The opinion monitoring and analysis task management system

The opinion monitoring and analysis task management system primarily manages and maintains the monitor tasks submitted by users, and configures and modifies the corresponding monitor indices. Its main functions include:

User management: Primarily manages users who manage the monitoring and analysis tasks, and maintains user information. Its major operations include user information maintenance and password management.

Monitoring and analysis task management: Maintains and manages relevant information of all television program opinion monitoring and analysis tasks. Its major operations include maintenance of task information and monitor keywords which are some key field information

need to be monitored and tracked for some monitoring and analysis task.

(5) Television program opinion monitoring and analysis application service platform

The application service platform performs statistical analysis of the data generated by the television program opinion monitoring and analysis engine module and offers statistical analysis data for the use of various television opinion monitoring and analysis application terminals. The statistical analysis links to the opinion corpus data files that have already been indexed by the monitored keywords. The analysis results are provided to all kinds of monitoring and analysis application terminals (such as web sites and mobile applications) in the form of service.

(6) Television programs opinion monitoring and analysis terminal application platform

The terminal application platform is the collection of various television program opinion monitoring and analysis terminal application programs, including web sites and IOS or Android-based mobile applications.

C. System Workflow Design

The workflow of all function modules are shown in Figure 2.

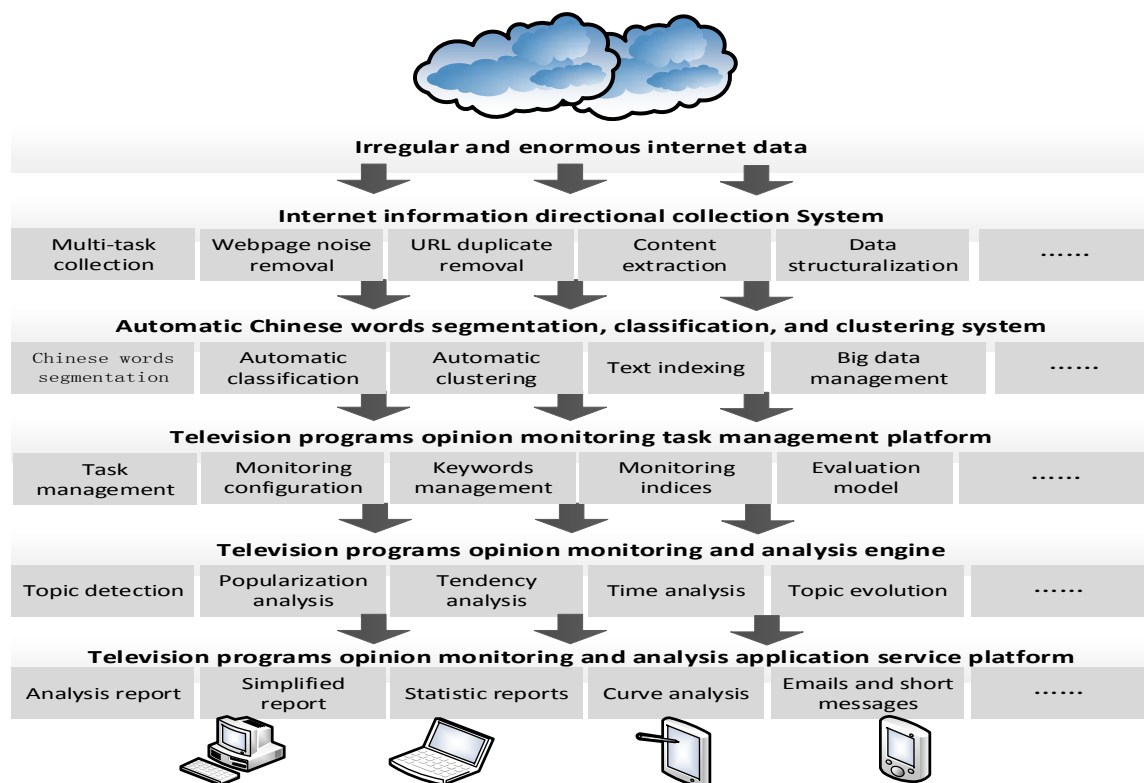


Figure 2. System workflow chart

IV. SYSTEM OPERATIONAL DEPLOYMENT

The operational deployment of the cloud computing-based television program opinion monitoring and analysis system is shown in Figure 3.

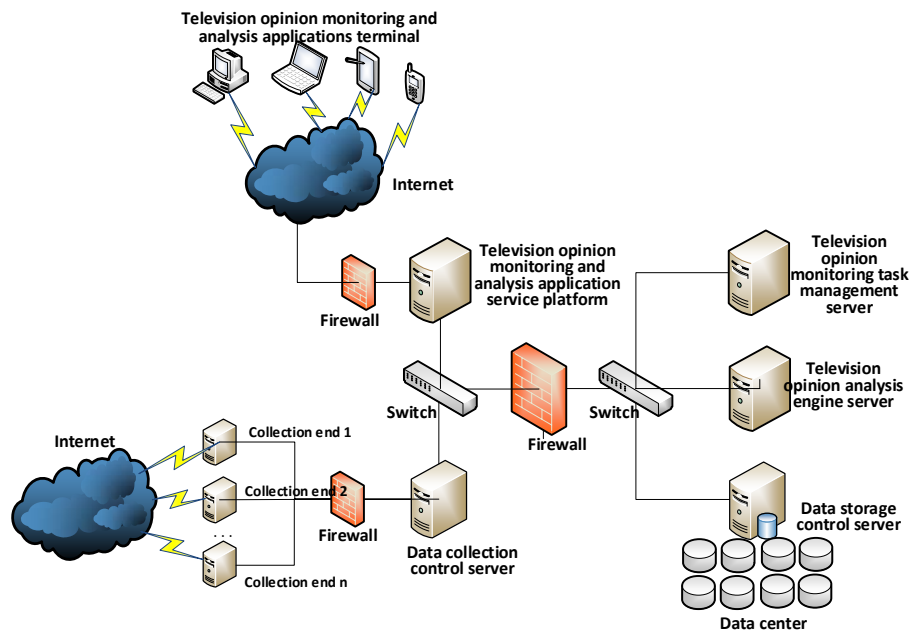


Figure 3. System operational deployment chart

The entire system deployment includes the following key parts:

(1) Internet information collection server (group)

The server (group) operates the internet information directional collection system. It is composed of a data collection control server and several data collection ends. The data collection control server uniformly manages and monitors the collection ends. The system realizes classified multi-task webpage information collection and content interpretation. Each collection end undertakes a collection task where the multi-thread webpage information collection and content interpretation is performed for each designated collection site to convert irregular webpage information into structured data and store them to the specified data center.

(2) Data storage server (group)

The server (group) is a data center composed by a data storage control server and several data storages locations. It stores the corpus data of television opinions. The data storage capacity of the data center is 20T at the time of the preliminary design.

(3) Television opinion analysis engine server

The server operates the television program opinion monitoring and analysis engine program. It analyzes the opinion corpus data that are generated by the internet information directional collection system and stored in the data center, its main function including Chinese word segmentation, classification, and clustering of corpus information in the data center. It also analyzes the emotional tendencies of the data. The server also constructs index files according to task-specified

monitored keywords for retrieval, census, and invoking by the television opinion monitoring and analysis application service platform.

(4) Television opinion monitoring task management server

The server operates the television opinion monitoring task management program which manages and maintains the monitoring task information submitted by users.

(5) Television opinion monitoring and analysis application service providing server

The server operates the television opinion monitoring and analysis application server platform which performs statistical analysis on the data generated by the television program opinion monitoring and analysis engine module and produces statistical analysis data in the form of service for the use of various television opinions monitoring and analysis application terminals.

V. CONCLUSIONS

In order to enhance the monitoring and analysis of relevant television program opinions on the internet, a cloud computing-based television programs monitoring and analysis system was developed and a 20T data center was established for the regular collection and update of relevant data from approximately 30 portal websites and news websites, 30 forum sites, and Sina microblogs. It can be seen from the testing that the intended result was achieved. Using a 100M campus network, the average speed of information collection and interpretation for news websites on a single PC (one collection terminal) is

47 pages per second. The interpretation accuracy reached 94.8%.

ACKNOWLEDGMENT

This work was supported by the Program for Zhejiang Leading Team of Science and Technology Innovation under Grant No. 2011R50019, the science and technology project of state administration of press, publication, Radio, film and television of China under Grant No. 201309, and the international cooperation project of Science Technology Department of Zhejiang Province under Grant No. 2012C24019.

REFERENCES

- [1] James Allan, "Introduction to Topic Detection and Tracking," Topic Detection and Tracking, The Information Retrieval Series, Vol.12, pp.1-16, 2002.
- [2] Yu Hong, Yu Zhang, and Ting Liu, et al., "Topic Detection and Tracking Review," Journal of Chinese Information Processing, Vol.21(6), pp.71-87, 2007.
- [3] Chunshan Li, Yunming Ye, and Xiaofeng Zhang, et al., "Clustering Based Topic Events Detection on Text Stream," Intelligent Information and Database Systems Lecture Notes in Computer Science, Vol.8397, pp. 42-52, 2014.
- [4] Shu-Wei Liu, and Hsien-Tsung Chang, "A Topic Detection and Tracking System with TF-Density," Recent Progress in Data Engineering and Internet Technology Lecture Notes in Electrical Engineering, Vol.156, pp.115-120, 2013.
- [5] Maximilian Walther, Michael Kaiser, "Geo-spatial Event Detection in the Twitter Stream," Advances in Information Retrieval Lecture Notes in Computer Science, Vol.7814, pp.356-367, 2013.
- [6] Jianfang Wang, Xiao Jia, Longbo Zhang, "Identifying and Evaluating the Internet Opinion Leader Community Through k-clique Clustering," Journal of Computers, Vol.8(9), pp.2284-2289, 2013.
- [7] Yongping Du, Changqing Yao, "Performance Evaluation of the Cyberspace Public Opinion Detection and Tracking," Journal of Computers, Vol.7(5), pp. 1284-1288, 2012.
- [8] ZENG Yi-ling, XU Hong-bo, "Research on Internet Hotspot Information Detection," Journal on Communications, Vol.28(12), pp.141-146, 2007.
- [9] Xiaoming Zhang, Zhoujun Li, "Automatic Topic Detection with an Incremental Clustering Algorithm," Web Information Systems and Mining Lecture Notes in Computer Science, Vol.6318, pp.344-351, 2010.
- [10] He Qi, Chang Kuiyu, and Lim Ee-Peng, et al., "Keep It Simple with Time: A Reexamination of Probabilistic Topic Detection Models," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.32(10), pp.1795-1808, 2010.
- [11] Goonie Internet public opinion monitoring system, <http://www.goonie.cn/products>, May 13, 2013.
- [12] Junquan internet public opinion monitoring system, <http://www.54yuqing.com>, March 17, 2014.
- [13] TRS internet public opinion management system, <http://www.trs.com.cn/product>, March 15, 2014.
- [14] The solution of Founder internet public opinion and information monitoring and analysis system, <http://wenku.baidu.com>, June 14, 2010.



Dingguo Yu is currently an associate professor in the Zhejiang University of Media and Communications, China. He received his M.S. and Ph.D. degrees from Tongji University, China in 2005 and 2011 respectively. His research interests are in the fields of cyber security, mobile computing and internet public opinion, etc.



Huxiong Li was born in Hubei, China. He received his Ph.D. degrees in 2009 respectively in Northwestern Polytechnical University. He is currently an associate professor at Zhejiang University of Media and Communications in Hangzhou, China. His current research Software engineering, Web application development, etc.



Zhiwen Hu was born in Hubei, China. He received his B.S., M.S., and Ph.D. degrees in 1996, 2003, and 2006 respectively in China University of Geosciences, Hefei University of Technology and Chinese Academy of Sciences. He is currently an associate professor at Zhejiang University of Media and Communications in Hangzhou, China. His current research interests reside in information fusion, intelligent control, etc.