ISSN: 2688-8645

Design and Implementation of COVID-19 Data Analysis and Visualization System

Yingying Chen[1,*], Taizhi Lv[2]

[1]School of Information Technology, Jiangsu Maritime Institute, Jiangsu Nanjing 211170, China

[2]R&D Department, Nanjing Huihai Transportation Technology Company Limited, Nanjing, China

*760846936@qq.com

Abstract

In order to facilitate the real-time analysis of COVID-19, a COVID-19 data analysis and visualization system is designed and implemented, which acuqries data from Baidu epidemic platform. It displays the epidemic dynamics, new trends, cumulative diagnosis, diagnosis distribution and other information through a big visualization screen. This system adopts the front-end and back-end separation technology, uses Scrapy to acquire epidemic data, and stores the data in the MySQL database through cleaning, conversion and other steps. The back-end obtains data based on Spring Boot, and the front-end uses Vue technology to analyze the back-end data and display the epidemic data in ECharts diversified charts. The practice under the production environment proves that the system can meet the requirements of real-time display and has high scalability.

Keywords

Spring Boot Framework; Vue; Data Visualization; Echarts; COVID-19; Data Analysis; Data Crawling.

1. Introduction

Since the first case of COVID-19 was found in December 2019, COVID-19 has ravaged the world for nearly three years, and has had a huge impact on the economy and life of people around the world. The cumulative number of infections has exceeded 600 million, and the number of deaths is close to 6.6 million [1-2]. With the development of a new generation of information technology, China's new infrastructure construction has gradually developed into a new variable to promote China's economic and social transformation and development, playing a huge role in epidemic prevention and control [3]. Big data, artificial intelligence, mobile APP and other technologies provide huge support for epidemic prevention and control, making epidemic data more complete and sufficient [4].

At present, the epidemic situation in China is relatively stable, and most cases can be controlled within a small range through management, technology, publicity and other comprehensive means. However, the epidemic situation is only controlled rather than eliminated, and there is still a risk of large-scale outbreak. In order to accurately understand the dynamic information of the epidemic situation, the epidemic data analysis and visualization system is designed and implemented. It can grasp the new trend, cumulative diagnosis, diagnosis distribution and other epidemic information in real time in a large screen way. This system collects epidemic data provided from Baidu epidemic platform, and finally stores the data in MySQL database through data cleaning, conversion and other processes. The data visualization is realized through the front-end and back-end separation technology. The back-end is based on Spring Boot and provides data interfaces according to the data layer, data service layer and business logic layer. The front-end implements visual charts based on Vue + Echarts.

ISSN: 2688-8645

2. System Design

2.1. Main Technologies

The main technologies used in this system are Spring Boot, Vue, Echarts, MySQL and Redis. Spring Boot is a framework for back-end interface development. It adopts the principle of convention over configuration to simplify various tedious Spring configurations. It is a mainstream back-end development technology [5].

ECharts is a data visualization chart library based on JavaScript, which can run smoothly on PCs and mobile devices and is compatible with most current browsers.Vue is a progressive framework for building user interfaces [6]. VUE is an implementation of MVVM. Its core is to implement DOM monitoring and data binding, which can provide drivers for complex single page applications [7].

MySQL is a relational database and a database management system with high efficiency and reliability. Its advantages lie in fast retrieval speed and convenient search.Redis (Remote Dictionary Server) is an open source log based, key value database that can also be persistent based on memory. It provides an API interface for Java language operations [8].

2.2. System Architecture

As shown in Figure 1, the system artchitecture is composed of data acquisition module, data storage module, front-end module and back-end module.

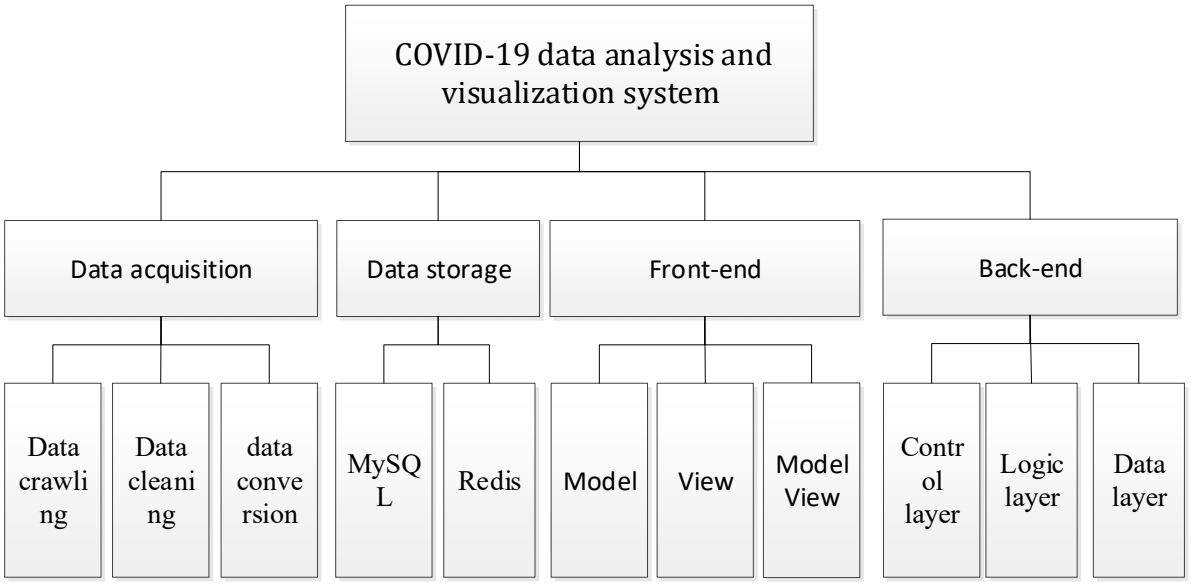

Figure 1. The system architecture

The data acquisition module realizes the process of data acquisition, cleaning and conversion. The data storage module stores epidemic data through MySQL+Redis mode. The front end mainly uses Vue framework to render the page and ECharts chart to display the data. It mainly consists of China epidemic map module, new diagnosis, cumulative diagnosis, new trend, cumulative trend, distribution of new diagnosis and domestic information modules. The back-end uses the Spring Boot framework to realize the data interface, which is divided into control layer, logic layer and data layer.

2.3. Functional Design

The contents to be displayed on the large visualization screen include diagnosis distribution, new trend, new diagnosis distribution, epidemic situation map, cumulative diagnosis distribution, cumulative trend, real-time broadcast of national consultation and other modules.

ISSN: 2688-8645

The existing diagnosis distribution and cumulative diagnosis distribution are represented by a histogram, the new trend and cumulative trend are represented by a line chart, the new diagnosis distribution is displayed by a pie chart, and the national consultation real-time broadcast is realized by an automatic roller.

The data of existing diagnosis distribution, new trend, new diagnosis distribution, epidemic situation map, cumulative diagnosis distribution and cumulative trend in the visual display are from Baidu epidemic situation platform and are crawled twice a day. The national current affairs consultation and analysis, the news about the epidemic situation obtained from China News, Sina News, Netease and other websites every 10 minutes, is clearly presented using the scroll wheel.

3. System Implementation

3.1. Data Acquisition and Cleaning

The system uses the Scrapy framework to acquire data. The acquisition process starts from the home page link and enters the directional crawl. Due to regular crawling, data duplication will occur during the crawling process. The Redis Set de duplication mechanism is introduced to compare and de duplicate each link in the Redis database to ensure that each link is crawled only once. The crawler analyzes the specific HTML page, extracts the corresponding fields, and saves them to the MySQL database.

In the data cleaning stage, the fields with null values, incomplete contents, and inconsistencies are supplemented and corrected in five steps. The default value cleaning is completed by removing and filling according to the proportion of missing values. If the number of people diagnosed is empty, 0 is used to supplement, such as removing data if the collected text is too short. Format content cleaning includes date format, keyword and other formats. Logic error cleaning is to remove unreasonable values, such as the number of infected people is not a number or too large. Correlation validation is used to validate the relationship between epidemic situations.

3.2. Back-end Implementation

The controller in Spring Boot accepts the user's access request, calls the processing mapping component to find the corresponding business controller according to the user's URL address, and the business controller obtains JSON format data through the business layer call and returns it to the browser, which visualizes the data through ECharts.

The control layer consists of the Spring Boot controller and the service controller in the implementation of this system. Register the business controller with the Spring Boot framework through the Controller annotation, and call the specified method of the corresponding controller through the URL access path. The business layer is mainly composed of model classes, data operation classes and business classes to complete business logic operations. The model layer of the system is implemented through Spring+MyBaits The configuration in the yml file enables the container to be loaded into the system, and the business processing classes are placed in the container for unified management. The corresponding entity classes, DAO files, and Mapper mapping files are automatically generated through the MyBatis Generator class. Modify the Mapper mapping file and DAO file according to business requirements, that is, realize the functions of model class and data operation class.

3.3. Front-end Implementation

The front-end implements MVVM (Model—View—View Model) architecture mode through Vue, which completes the bidirectional binding of data. Based on the Bootstrap framework, the page layout is realized through Flex. In the Flex layout, the design of main axis and cross axis

ISSN: 2688-8645

allows elements to be arranged on the page. The overall page layout is divided into several areas. In the outermost div, the class attribute is Container, the class attribute of its sub element div is set to row, and the required number of divs is placed in the sub element div. These divs differentiate different types of devices by specifying the class prefix of the class. There are four class prefixes: col xs -, col sm -, col md -, col lg, which correspond to ultra small screen, small screen device, medium screen, and wide screen device respectively.

ECharts provides pie chart, histogram, map and other graphic drawing tools. The system draws statistical reports through ECharts. The chart steps to implement ECharts are shown in Figure 3.

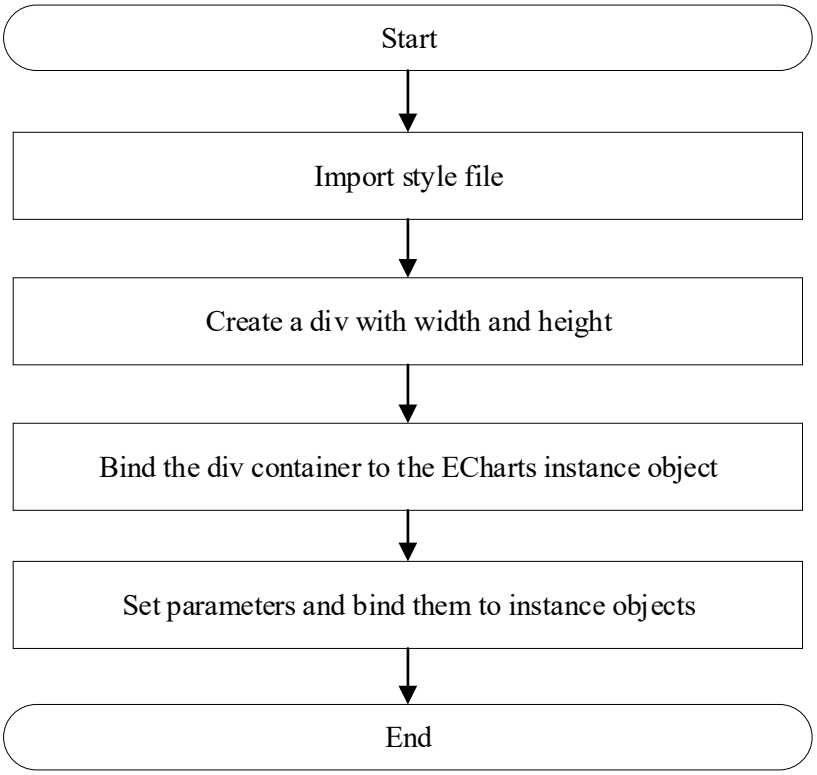

Figure 2. The visualization step

Figure 3 shows the large screen visualization. Statistics such as the number of newly confirmed persons, cumulative number of confirmed persons, and existing number of confirmed persons displayed at the top of the center. The map part reflects the current diagnosis and cumulative diagnosis values of each province through different colors. When the mouse hovers over the plate of a province, the current diagnosis and cumulative diagnosis number of that province will be displayed. The left and right histograms count the number of confirmed cases and the cumulative number of confirmed cases according to different provinces. The left and right line charts provide statistics by time, from which future trends can be predicted. The pie chart shows the proportion of provinces with newly confirmed epidemic data, from which the geographical distribution of the epidemic can be known. The consultation real-time broadcast part collects real-time epidemic information from various websites in a rolling way, so that users can grasp the epidemic information in a timely manner.

ISSN: 2688-8645

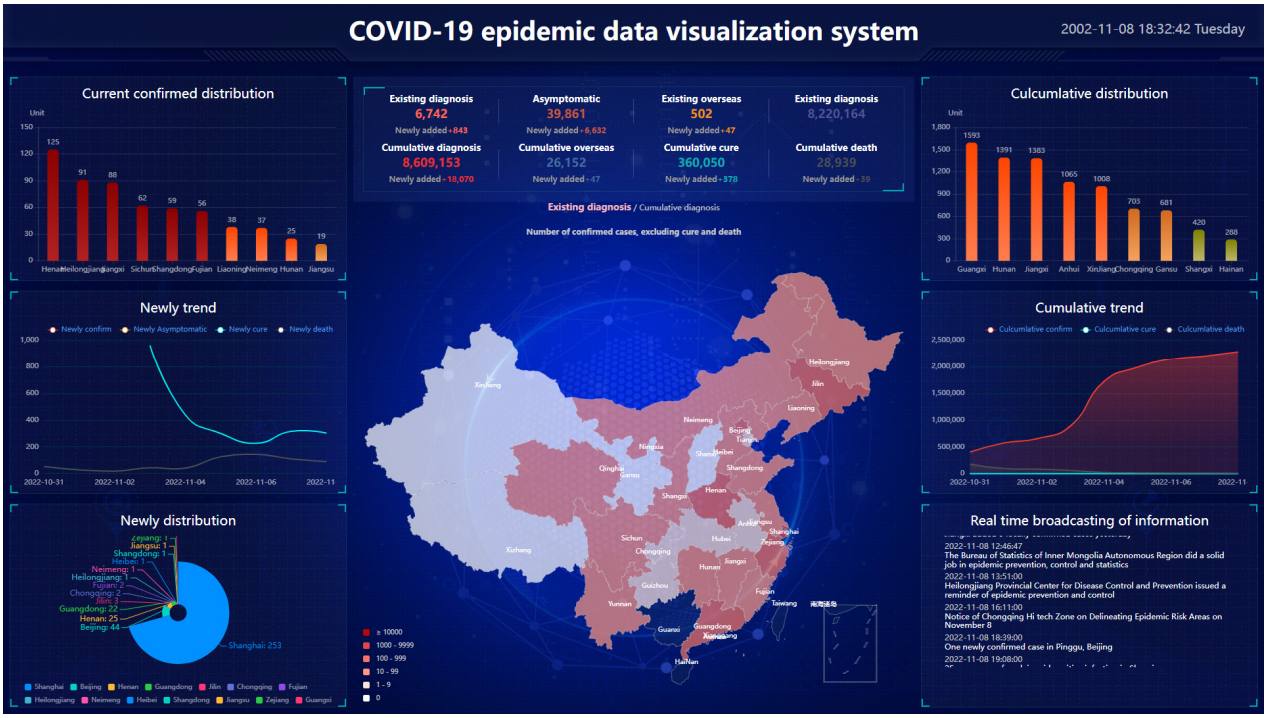

Figure 3. The visualization effection

4. Conclusion

This system adopts the front-end and back-end separation technology, and completes the epidemic data visualization system using the open source frameworks which includs SpringBoot, Vue, Echarts, Scrapy, Redis, etc. The purpose of the system is to better display the spread trend of the epidemic situation. The main visualization modules include the existing diagnosis distribution, new trend, new diagnosis distribution, epidemic map, cumulative diagnosis distribution, cumulative trend, real-time broadcast of national consultation, etc. Through real-time analysis and visualization of epidemic data, users can quickly understand the changing laws of disease conditions and epidemic trends, so that people can more intuitively understand the current epidemic situation, pay attention to the epidemic situation anytime and anywhere, and dynamically obtain epidemic information.

Acknowledgments

This work was financially supported by the Excellent Scientific and Technological innovation Team of the Jiangsu Higher Education Institutions of China (Maritime big data team), the Young Academic Leaders for QingLan Project of the Jiangsu Higher Education Institutions of China, the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (21KJB 580007).

References

[1] Fagbadebo, Omololu, and Nirmala Dorasamy. "Introduction: Governance Challenges in the Management of the COVID-19 Pandemic in Africa." Governance Challenges During the COVID-19 Pandemic in Africa. Palgrave Macmillan, Cham, 2022. 1-13.

[2] Yasmeen, Rizwana, et al. "The impact of COVID-19 on the US renewable and non-renewable energy consumption: a sectoral analysis based on quantile on quantile regression approach." Environmental Science and Pollution Research (2022): 1-16.

[3] Wang, Bingyu. "Immobility infrastructures: taking online courses and staying put amongst Chinese international students during the COVID-19." Journal of Ethnic and Migration Studies (2022): 1-19.

ISSN: 2688-8645

[4] Liang, W. N., et al. "The dynamic COVID-zero strategy on prevention and control of COVID-19 in China." Zhonghua yi xue za zhi 102.4 (2022): 239-242.

[5] Li, Zhijie. "Design of Ordering System based on Spring Boot Framework." International Core Journal of Engineering 8.5 (2022): 579-588.

[6] Li, Zhaohui, et al. "Design and Implementation of Production Visual Report Subsystem Based on Echarts." 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). IEEE, 2022: 696-701.

[7] Kentaro, Nagaki. "Browser-Based Manipulation of Virtual Objects Through MVVM Architecture with Data Binding." 2021 Ninth International Symposium on Computing and Networking (CANDAR). IEEE, 2021: 131-140.

[8] Gade, Anne N., et al. "REDIS: A value-based decision support tool for renovation of building portfolios." Building and environment 142 (2018): 107-118.