Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 6, No. 3, 2023

Python-based Epidemic Data Visualization System

Na Xue*, Fan Yao, Zehao Xie

North China University of Science and Technology, Langfang 065201, China
* Corresponding author

Abstract: In order for people to timely understand the relevant dynamics of the epidemic and do their own protection, the epidemic data visualization system has been developed and designed. Based on Python language and Echarts visualization technology, the obtained epidemy-related data are visually classified and displayed, so that the data is more clearly displayed in front of users, providing users with a relatively comprehensive channel to obtain epidemic information.

Keywords: Python crawler, ECharts,Visualization.

1. Introduction

Since December 2019, cases of pneumonia of unknown etiology have been found one after another, and subsequently diagnosed as COVID-19. After people are infected with the coronavirus, it can cause a series of serious diseases such as MERS or SARS. In the past history, the existence of a new coronavirus has not been found. Because of its fast spreading speed and strong infectivity, it has been widely concerned by various countries. Therefore, attention to the spread of pneumonia is particularly important today.

Science and technology are developing rapidly today, big data technology and visualization technology are also applied in various fields. The visualization of data provides great convenience for people to focus on a certain field or information. At the same time, the visualization of data is not limited to any form, ICONS, tables, images, etc., can become the carrier of displaying data, so that people can intuitively perceive the trend and regular changes of data, and analyze data for users. Researchers are also using data visualization to dig deeper. For example, Jin Sichen et al. proposed a visual analysis system for infectious diseases, which can directly analyze the space-time pattern of infectious diseases and interactively explore the correlation and similarity between different diseases and regions. Zhu Junjiang and others developed and designed a visualization software to provide a supporting platform for analyzing data on the ocean floor.

Today, when the novel coronavirus pneumonia is still spreading, this paper introduces how to crawl the epidemy-related data we need, how to visualize the data, and finally show the current situation of the epidemic. Through this platform, you can understand the real-time dynamics of the novel coronavirus pneumonia, understand the current situation, and remind people to always do their own protection.

2. System Overview

In order to show epidemy-related data and information to users clearly and succinctly, the system uses Python to crawl epidemy-related data, such as the number of infected people and epidemy-related news, etc., and combines Echarts data visualization technology with statistical charts to build an epidemic visualization platform.

(1) System architecture

The system design adopts a four-layer architecture, as shown in Figure 1.

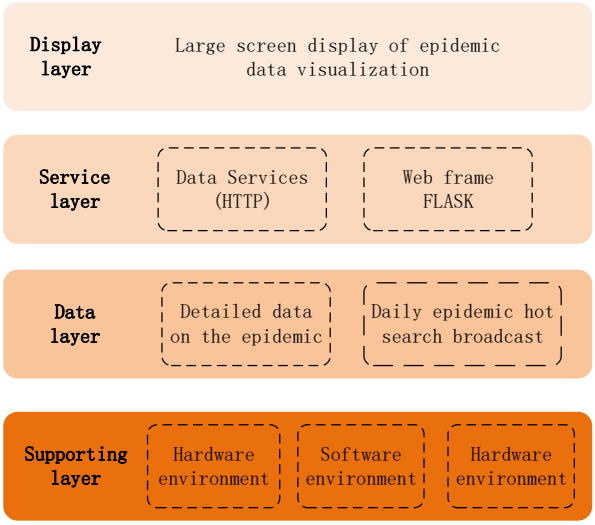

Figure 1. System architecture diagram.

The bottom layer is the support layer, which has the basic hardware environment, software environment and network environment. Data layer for the use of python language to climb the epidemic data information and daily epidemic hot search broadcast, stored in the database. Flask, a lightweight Web framework written in Python, is selected for the service layer based on the requirements of the epidemic visualization platform. Flask is flexible, simple, and expandable, which greatly improves the efficiency of the computer. Using the jinja2 engine, the back end responds to the front end request and passes the data to the front end in JSON format. The top layer is the display layer, with the support of Echarts to realize the visual display effect of the epidemic.

3. Data Processing

According to certain rules, the program or script that automatically crawls the World Wide Web information is called web crawler. We use Python crawlers to extract epidemy-related data, such as the number of new daily confirmed cases across the country and the number of deaths. The data of this project mainly comes from Tencent News and Baidu News.

(1) Data source

The epidemic list module, epidemic visualization map module, cumulative trend module, new trend module, and TOP5 data of new confirmed cases are all from Tencent News, and the data of today's epidemic report is from Baidu News (the data is available until March 30, 2022).

Information such as parameters, device, and access time requested by a user is obtained using Headers. On the JupyterNotebook, you get the URL of the link you want to crawl, res gets the response, and Json converts the string into a dictionary. Obtain the corresponding data through the for loop, and change the time format to prevent data insertion into the database error. The data acquisition flow chart is shown in Figure 2.

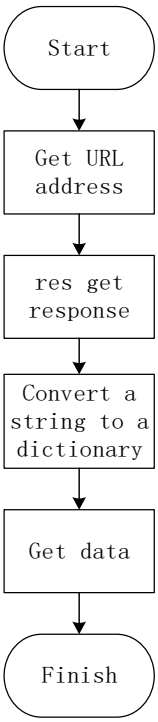

Figure 2. Get a data flow chart

Baidu pages use dynamic rendering technology, in order to simulate the real browser visit behavior, using the selenium plug-in as the crawling tool of choice. Selenium is an open source Web application testing tool that runs directly in a browser to simulate real user access behavior. It is worth mentioning that before using Selenium, users need to download the corresponding browser driver. Then, we find the Xpath path to the information label we want to crawl, and then, following the previous procedure, we crawl the information.

(2) Data table structure

The Python crawler was used to climb the data published on Tencent News and Baidu News every day, and the epidemic data of various provinces and cities and the hottest news every day were climbed.

The data is stored in three separate tables. They are the history table, details table, and hostsearch table.

The history table stores daily epidemic data, the details table stores daily epidemic details, and the hostsearch table stores daily epidemic hot searches. The format of the final crawled data is shown in the following table.

Table 1. History data table structure

Field name	type	Instructions
ds	datetime	date
confirm	int(11)	Cumulative diagnosis
confirm_add	int(11)	New confirmed on the day
suspect	int(11)	Residual suspicion
suspect_add	int(11)	New suspected
heal	int(11)	Cumulative cure
heal_add	int(11)	Day added cure
dead	int(11)	Cumulative death
dead_add	int(11)	Daily deaths

Table 2. Details data table structure

Field name	type	Instructions
id	int(11)	Major key
update_time	datetime	Data was last updated
province	varchar(50)	province
city	varchar(50)	city
confirm	int(11)	Cumulative diagnosis
cofirm_add	int(11)	Newly confirmed diagnosis
heal	int(11)	Cumulative cure
dead	int(11)	Cumulative death

Table 3. Hostsearch data table structure

Field name	type	Instructions
id	int(11)	
dt	datetime	
content	varchar(255)	

(3) Data storage

After obtaining the data, we should carry out the subsequent interaction between MySQL and Python to store the crawled data. First of all, obtain the link to the database on the JupyterNotebook. After linking to the database, insert the crawled data into the database.

4. Data Visualization

Data visualization is to mine and transform relatively complex and miscellaneous big data, extract structured data content, and display it in a visual way that is easier for users to understand, and clearly express the information and rules inherent in the data. The visualization of data provides a strong guarantee for people to explore the correlation between data and explore the potential laws of data. This paper uses ECharts to visualize the crawled data. ECharts is an open source visualization library implemented in JavaScript, which has the ZRender library underneath, supports many chart types, and can display dynamic data.

Epidemic data show

Retrieve the data from the database and query it using the select statement in the sql statement. The cumulative number of confirmed cases, the remaining number of suspected cases, the cumulative number of cured cases and the cumulative number of deaths were obtained.

Epidemic map implementation

Map option was copied in PyCharm, corresponding js was imported, and the confirmed number of each province and city was obtained by select statement. Since we have to

update the data several times, we need to get the most recent set of data with the timestamp.

The rest of the cumulative trend chart is realized

Copy the line chart option in PyCharm and get the data through the select statement. After data is obtained, the cumulative trend chart and new trend chart are displayed normally. The bar chart option was copied in Pycharm to obtain the data, and the top 5 confirmed new cases were displayed normally.

Copy the wordcloud graph option in PyCharm and import wordcloud.min.js. Then obtain the data, use jieba to obtain the keyword, after obtaining the data, today's real-time epidemic report will be displayed normally.

5. Conclusion

In this paper, the epidemic related data was obtained by Python and stored in the database to realize the interaction between Python and MySQL. Flask was used to build a Web platform, and Ajax requests were used to realize the interaction between the front and back ends. The data could be updated every 12 hours to realize real-time broadcast of epidemic data. Finally, the data was visualized through Echarts, and the current epidemic situation was clearly displayed to users, providing a reference method for other data capture and display in the future.

References

[1] ZHU Junjiang, OU Xiaolin, Yang Yue, Chen Ruixue, ZHANG Shaoyu, JIA Zhongjia, Zhang Shengsheng, Wang Pengcheng, Li Sanzhong, LIU Yongjiang, JIA Yonggang. Design and analysis of ocean floor data visualization and mapping [J]. Earth Science Frontiers,2022,29(05):255-264.(in Chinese)

[2] Jin Sichen, Tao Yubo, Yan Yuyu, et al. Infectious disease model analysis based on multi-dimensional spatiotemporal data visualization [J]. Journal of Computer-Aided Design and Graphics,2019,31(02):241-255. (in Chinese)

[3] Amjady N. Short-term hourly load forecasting using time series modeling with peak load estimation capability [J]. IEEE Transactions on Power Systems, 2001, 16(4): 798-805.

[4] Tencent News. Real-time update: the latest trends of the new champions league pneumonia outbreak [EB/OL]. (2021-05-10). https://news.qq.com/zt2020/page/feiyan.htm#/global.

[5] Baidu News. Real-time update: the latest trends of the new champions league pneumonia outbreak [EB/OL]. (2021-05-10). https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_pc_3#tab1.

[6] Jiang Wen, Liu Likang. Automated testing of Web Software based on Selenium [J]. Computer Technology and Development,2018,28(09):47-52+58. (in Chinese)

[7] SUN Yuanbo, Wen Zhiyi, Xu Ruige, Luo Fei, LI Ge. Overview of novel coronavirus pneumonia data visualization design [J]. Packaging Engineering, 2019,41(08):51-62.(in Chinese)

[8] CUI Peng. Application of ECharts in Data Visualization [J]. Software Engineering,2019.6: 42-46.(in Chinese)