# Design and implementation of recruitment service platform based on big data technology

*Hongxiu Duan, Wenting Ding*
Communication University of China, Nanjing, Nangjing 210000, China

**Abstract:** This paper studies the employment situation of college graduates, and provides accurate and intuitive application programs for job seekers. System based on Hadoop big data platform, through data collection, data cleaning, data analysis, data visualization and other steps, for mainstream recruitment website recruitment information and corresponding regional rental information collection and analysis, the average pay for hiring, recruitment jobs, education, work experience requirements and recruitment information, such as housing prices near, Intuitive display USES data visualization technology, using collaborative filtering recommendation algorithm for precise recommendation.

**Key words:** Big data; Recommendation algorithm; Visualization technology; Recruitment services

## 1. Introduction

Big data technology is an extremely important way to obtain the value of data, and recruiting big data can make candidates more intuitively understand the needs of the talent market. Most current recruitment platform only have basic recruitment information screening function, lack of accurate information for job seekers recommendation feature, unable to provide timely and high quality recruitment information. Besides, the recruitment website are not associated with the corresponding area of the rental housing information, job seekers need to use other software to the rent price and peripheral information query, so the system, according to a recruiter's location information gathering intelligence to recommend nearby rental information to the user. This paper proposes a recruitment service platform based on big data technology. Through data visualization, the talent demand of recruiters and rental information in the recruitment area are displayed through charts, aiming to provide intelligent employment recommendation services for the majority of job seekers, especially graduates who have just graduated from campus.

## 2. System design

The function module of this system is composed of two modules: data operation and data visualization. Data operation mainly realizes data acquisition, data cleaning and data analysis. Data acquisition mainly implements the data up deposit, timing acquisition, data analysis, data, and other functions. Data crawling mainly collects effective data for the target data website. Data analysis is mainly for web key content of returned to the client's HTML text extraction; Data storage is mainly to store the collected data into the database for calling; Timing acquisition is mainly a sustainable expansion of the system, to achieve a fixed time every day to execute the program to update the data. Data cleaning is mainly for data reading, field data cleaning, deposit, and other functions. Field cleaning are mainly for some empty value data, error value and error type, modified to make data can accord with the requirement of data calls after cleaning. Data analysis mainly realizes the requirements analysis, SQL design and other functions. Requirement analysis mainly analyzes the data requirements of back-end calls; SQL design is to write the corresponding SQL statements for data requirements to query data.

The main body of data visualization is divided into two modules: back-end call and front-end display. The back-end call includes framework construction, port writing, data testing and other functions. The framework construction is mainly based on the Spring Boot framework to build the system background. Port writing calls database data by writing SQL operations, and passes the data into the port to facilitate front-end calls. Data testing is to test the correctness of data on the port. The front-end display mainly implements the framework structures, routing configuration, user management, and other functions. The specific function modules are shown in Figure 1:

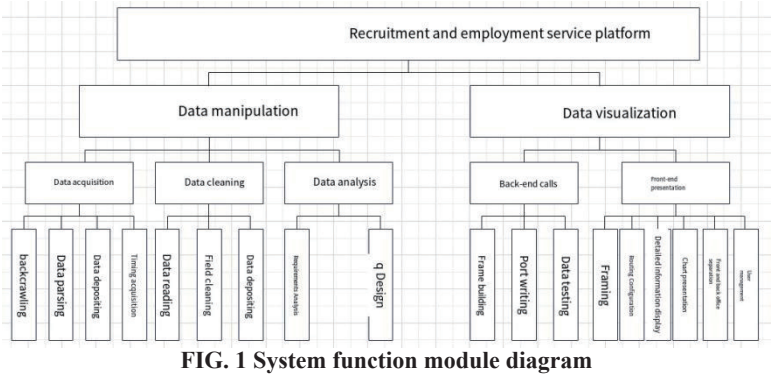


**FIG. 1 System function module diagram**

## 3. Related technologies

3.1 Data Acquisition technology - Introduction to Scrapy framework

Scrapy use Python implementation, to crawl web site data, extracting structured data written application framework. The Twisted

efficient asynchronous network framework is used to deal with network communication, which is mainly composed of scheduler, downloader, crawler, entity pipeline and Scrapy engine.

With the help of Scrapy crawler framework, job information and recruitment enterprise information are crawled from mainstream recruitment websites, including job name, salary, work experience, educational requirements, recruitment number, release time and other information, and recruitment enterprise information including enterprise name, industry type, specific location (province, region) and other information. In order to ensure the accuracy of the data, it is also necessary to remove the duplication of the data.

3.2 Data Cleaning techniques - Hadoop HDFS and Spark

Hadoop is a reliable, scalable and open source distributed computing software developed by Apache Corporation. The HDFS file distribution system is one of its core components, which is mainly used to store files and locate them through a unified namespace and directory tree. HDFS provides a distributed storage mechanism for Hadoop cluster, but also provides a powerful capacity for linear growth of mass storage.

Spark is an open source cluster computing environment similar to Hadoop, but with some differences, it enables in-memory distributed datasets, which in addition to being able to provide interactive queries, can also optimize iterative workloads. Spark is complementary to Hadoop and can run in parallel within the Hadoop file system.

3.3.Visualization -Echarts

Echarts is an open source visualization library implemented by JavaScript, which can run smoothly on PC and mobile devices and is compatible with most current browsers. The underlying layer relies on lightweight vector graphics library ZRender to provide intuitive, interactive and highly customizable data visualization charts. For example, ECharts provides regular line charts, bar charts, scatter charts, etc., as well as a variety of charts for statistics, geographic data visualization, relational data visualization, and multidimensional data visualization, and supports the mixing between graphs. In this paper, Echarts visualization technology is used to visualize the analyzed recruitment information and rental information.

3.4 Intelligent recommendation algorithm

In this system to recommend the corresponding position of job seekers, need through them to industry, and the position, on a city's search habits and users using the log information, such as related heat calculation and sorted . This system uses a user-based collaborative filtering algorithm to measure and score historical behavior data, find a number of similar users for the current user, and then recommend to the current user according to the preferences of similar users. The recommendation information will be updated dynamically according to the increase in the number of times the user uses it. The basic process of this recommendation algorithm is as follows:

(1) Extract the user's historical behavior data;

(2) data preprocessing, extracting the required data from the messy data, and dividing the training set and the test set;

(3) Get the user-job rating matrix and do relevant statistical work;

4.Train a model with a training set.

(5) use the test set to test the model indicators;

(6) According to the rules of the algorithm, the top N jobs are obtained to recommend to users.

## 4. System implementation

4.1 Data Acquisition

The platform USES Scrapy framework of the mainstream recruitment website and rental site for data collection, such as 51 job, HOME LINK net, etc., on the job title, and salary information in the website, work experience, education information, hiring, release the information such as time, as well as the recruiter's name, industry type, specific location information such as the crawl. Module is divided into client to access the head camouflage, XPath parsing data screening, deposited in the database, Settings page set up collection, timing acquisition a few more parts.

4.2 Data Cleaning

The process of data cleaning is: First, open the data to be cleaned, export the local database file to the csv format file and upload it to the server through sftp, operate HDFS on the server and upload the file to HDFS. Spark will directly access the file on HDFS, remove the first line title of the source data, and determine whether the field data of the source data meets the 10 fields. If the data is not satisfied, there is a null value in the data and it is removed as dirty data. Take the position salary as an example, you can determine whether the salary field contains "-" and ", ", if not, it will be removed as dirty data; Unify the salary format, replace all salary formats into yuan/month, rearrange the salary column into the minimum wage and the maximum wage, save the data and import it into the database.

4.3 Data Analysis

Data analysis refers to the process of analyzing processed data with appropriate analytical methods and tools to extract valuable information and form effective conclusions. Data analysis is to analyze the needs of data visualization, and write SQL statements to query the data, and extract the required data. That is, according to the field requirements of the final presented data, write the corresponding SQL statement, such as showing the number of recruitment of a position in each district of a city, need to write the query statement, if the query results meet the requirements, the SQL is retained and the query results are delivered to the back-end.

4.4 Visual Presentation

Data visualization is divided into the back-end call and front-end display two aspects, the back-end call USES the Spring framework, the Boot for MySQL database access, SQL queries the Mybatis - plus plug-in simplify the query code, and then write the corresponding

according to the results of the query interface for the front-end calls, Postman is used for back-end data detection to verify the data provided by the interface. The front-end display uses Vue framework, and the display content is divided into two modules: detailed information display and big data chart display. The database information is displayed by charts through Echarts component, which includes the average salary display in a specific region, the number of recruitment positions and other functions. The display chart types include bar charts, pie charts, radar charts, Nightingale Rose charts, etc.

5. Visualize the results of the analysis

The database information is displayed in the form of charts through the Echarts component, which includes the display of the average salary of a certain region, the number of jobs in the region, and other functions. Figure 2 shows the average salary chart after visualization according to the recruitment positions and recruitment salaries in various districts of a city. Figure 3 shows the number of jobs after visualization according to the number of jobs and recruitment in each district of the city, using a circular pie chart.

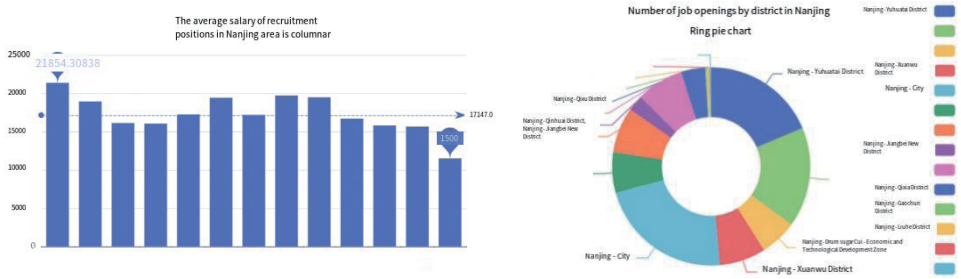


**Figure 2 a certain area average salary figure 3 urban district hiring figure number**

## Conclusion

Aiming at the numerous online recruitment information, this paper realizes the massive recruitment data processing and analysis through big data technology, and carries out direct and objective data visualization display according to the collected recruitment information and recruitment enterprise information, and provides personalized recommendation services for job seekers, including recruitment job recommendation based on collaborative filtering recommendation algorithm. And rental information recommendation according to the location of the recruitment enterprise, can help job seekers better find their favorite job. The current user object of the platform is job seekers, so the system will improve the service of recruitment enterprises in the future, realize two-way recommendation function, and provide accurate talent recommendation for enterprises.

## References:

[1] Huang Junping. College Students' Employment Management System Based on Recommendation Algorithm [J]. Information Technology and Informatization,2021,(11):6-9.

[2] Lu Jiaxin, Zhu Zhaohua, Jiang Zhilin. Research on Campus recruitment service system design under the concept of service design [J]. Industrial Design,2022,(06):103-105.

[3] Fu Mingliang, Li Gang. [3] [Fu L, Li G, et al.Research on matching between job seekers' needs and enterprises' needs based on big data analysis of recruitment websites. Journal of Statistics and Management,2021,36(08) : 91-97

[4] Huang H Z. Research on the influencing factors of recruitment APP usage based on user experience [J]. Journal of Yueyang Vocational and Technical College,2019,34(04) : 98-101.

[5] Shan Shuqian, REN Jiaxun. Design and implementation of Database principle Website Based on Spring Boot and Vue Framework [J]. Computer Knowledge and Technology.2021,17(30).

[6] Huang Jin-fan, LIANG Shao-hua, ZHANG Jia. Design and implementation of recruitment data visualization analysis system [J]. Computer Knowledge and Technology, 2012,18(18):39-41. (in Chinese)

[7] Hu Wenyu, Ying Kanghui. Research on instance level Data cleaning Technology [J]. Computer Technology and Development, 2012,32(05):22-28.

[8] TAO Jian. Research and implementation of video recommendation System Based on Spark [D]. Chongqing Normal University,2019. (in Chinese)

[9] Hong Wenxing, Wang Ning, Chen Yiwei et al. Talent Recommendation system in the era of big Data [J]. Big Data,2017,3(2) : 115-120.]

[10] LIU Jiaqi, WANG Quanquan. University Personalized Book Recommendation System Based on improved user collaborative filtering algorithm [J]. Computer and Digital Engineering,2020,48(10):2458-2461+2479.

[11] WANG Dan, DONG Lang et al. Data acquisition technology based on python+requests [J]. Science and Technology Innovation and Productivi ty.2024,45(06):142-144

[12] DU Yihang. Analysis of data cleaning technology in big data [J]. Electronic Technology, 2023,52(12) : 44-45.

[13] Xiong Ying. Big data analysis technology in the application of teaching case [J]. Integrated circuit applications. 2024, 9 (6) : 200-201.

[14] WANG Ziyi, ZHANG Chunhai et al. ECharts based data visualization analysis component design [J]. Micro computer and application, 2016, 35 (14) : 46-48 + 51.

[15] CUI Peng. Application of ECharts in Data Visualization [J]. Software Engineering, 2019,22(06) : 42-46.

**About the author:** Duan Hongxiu, Master student, Associate Professor, Research interest: Computer applications.

**Project Information:** General Program of Philosophy and Social Science Research in Colleges and Universities of Jiangsu Province 2023JSYB0635