

# Literature Analysis of Vocational Education Group based on CNKI Data

Yue Yan\*, Juan Zhang, Taizhi Lv

School of Information Engineering, Jiangsu Maritime Institute, Nanjing 211170, China

\*292676187@qq.com

## Abstract

**The vocational education group is a school running mode with distinctive characteristics of the times. Facing the fierce competition trend of market economy, the establishment of vocational education group has become an inevitable choice for the development of vocational education. In order to study the research status of vocational education groups and reveal the development status and context of vocational education groups, literature analysis based on CNKI (Chinese National Knowledge Infrastructure) data is carried out. It crawls the literature information with the vocational education group as the key word from the database of CNKI by Python crawler technology. The information of the literature is stored in the MySQL database and displayed through the front-end and back-end separation technology.**

## Keywords

**Vocational Education Group; Big Data Technology; Spring Boot; Vue; Literature Analysis; Data Visualization.**

## 1. Introduction

Vocational education group is an educational group organized by vocational colleges, industrial enterprises and other organizations to realize resource sharing, complementary advantages and cooperative development [1-2]. It is an important model for China to accelerate the reform of vocational education running mechanism and promote the open and sharing of high-quality resources in recent years. The characteristics of the operation of the vocational education group is to carry out various activities of the group with enterprises, vocational colleges and research institutes as the center, and guarantee the normal operation of the activities of the vocational education group through the coordination, guidance and other functions of the government and industry organizations, forming a pattern of guidance and guarantee around the three main bodies with enterprises, vocational colleges and research institutes as the center, government and industry organizations as the assistance.

Literature data analysis is to describe, evaluate and predict the current situation and development trend of a research problem by using various characteristic parameters of the literature and using mathematical and statistical methods [3]. Data analysis of literature characteristics is a quantitative evaluation method with cost and efficiency advantages, accuracy and objectivity. CNKI is the most comprehensive dynamic resource system and the most advanced knowledge service platform and digital learning platform in China. CNKI contains multiple databases, covering science and engineering, social science, electronic information technology, agriculture, medicine and other disciplines [4].

Based on CNKI literature data, this paper uses big data technology to realize statistics and analysis of literature features. It reveals the research status and development trend of vocational education groups.

## 2. Literature Data Analysis Process

The data analysis process includes three steps: data crawling, back-end statistics and front-end visualization. Based on data crawler technology, CNKI literature data is acquired to the MySQL database. It then realize data statistics through back-end technology, and provide data interface. Finally, It displays the results through data visualization technology.

### 2.1. Data Crawling

The data source of the study is from CNKI, with the vocational education group as keyword. The information acquired includes the title of the paper, author, journal name, publication time, citation, download, abstract, keyword, fund, classification number, page number, and information. Data crawling is implemented based on Request and BeautifulSoup components of Python. The process is as follows:

Step 1: The CNKI database is queried with the vocational education group as keyword.

Step 2: It parses the web page by using BeautifulSoup and lxml parsing library. On the paper list page, the table item is found where the list is located, then all the tr items under the table item is traversed. The title, author, journal name, cited times, download times and link address of the detailed page is obtained by the tr item. The literature information is stored in the Redis cache.

Step 3: It jumps to the next page until all pages are crawled.

Step 4: It extracts the basic literature information from the cache in turn, crawl the literature details according to the hyperlink, including the abstract, keywords, index, page number, etc., and store the literature information in the MySQL database.

### 2.2. Data Statistics and Back-end Services

This research implements data statistics based on Spring Boot and SQL statement. It provides data interface services in JSON format. The back-end uses Spring Boot to implement the interface, MyBatis to implement the mapping between database and Java, Druid to implement database connection pool, and PageHelper component to implement record paging. The introduction of annotations in Spring Boot integrates a large number of frameworks, simplifies the configuration and development process of Web frameworks, and avoids version conflicts caused by third-party dependencies on packages [5]. MyBatis is a Java-based persistence layer framework that supports object mapping. It has built-in JDBC and only focuses on SQL [6]. Through the Spring Boot framework, realize the data interaction between the front and back ends, and apply the three-layer architecture model of Controller, Service and Dao.

### 2.3. Data Visualization

This research adopts Vue as front-end framework and Vue-cli to quickly create projects. The front-end layout adopts ElementUI component-based development. Vue is a progressive JavaScript framework based on MVVM (Model-View-ViewModel) mode, which can quickly build front-end applications [7]. The front-end sends the request based on the Axios asynchronous request, and the back-end returns JSON data, which is bound to the Model and synchronized to the ECharts visualization chart through the ViewModel. ECharts is a JavaScript based data visualization library, which can display nonlinear data in a graphical way [8]. It is easier for researchers of vocational education group to mine the hidden laws in the literature.

## 3. Analysis of Literature Statistics Results

### 3.1. Analysis of Basic Literature Information

The number of published papers is one of the important indicators to measure the degree of concern in a research field. The annual analysis of the number of published papers can reflect

the change of the degree of concern in the research field. As shown in Figure 1, from the perspective of the literature number issued in the 22 years since 2000, the number of literatures issued in 2016 was 216, the highest. The number of literatures issued in 2017-2022 showed a downward trend. In today's era of rapid development of vocational education, this does not match the social focus of vocational education groups.

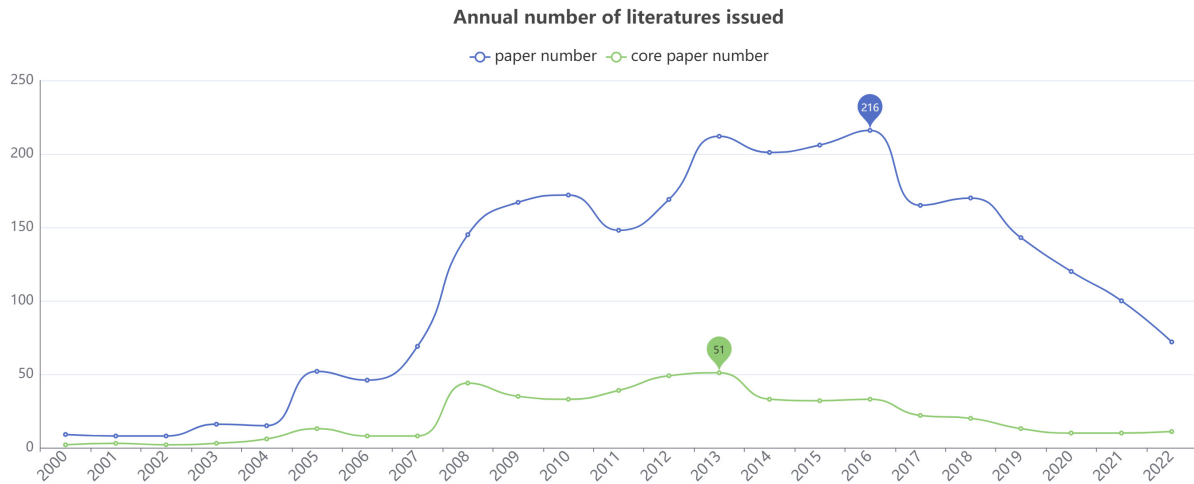


Figure 1. Annual number of literatures issued

Statistical analysis of published journals is conducive to understanding the direction of the journal and helping contributors to submit papers. According to statistics, 1833 papers were published in 515 journals. 138 papers were published in Vocational and Technical Education which published the most papers about vocational education group. Figure 2 shows the statistical information of the top 10 journals.

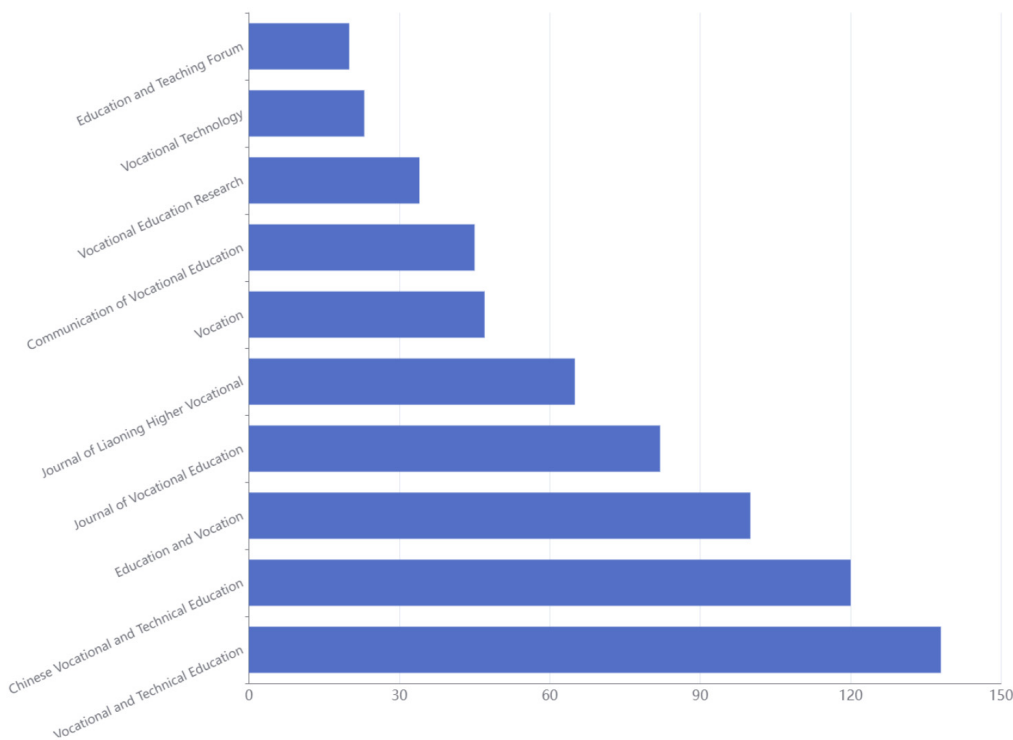
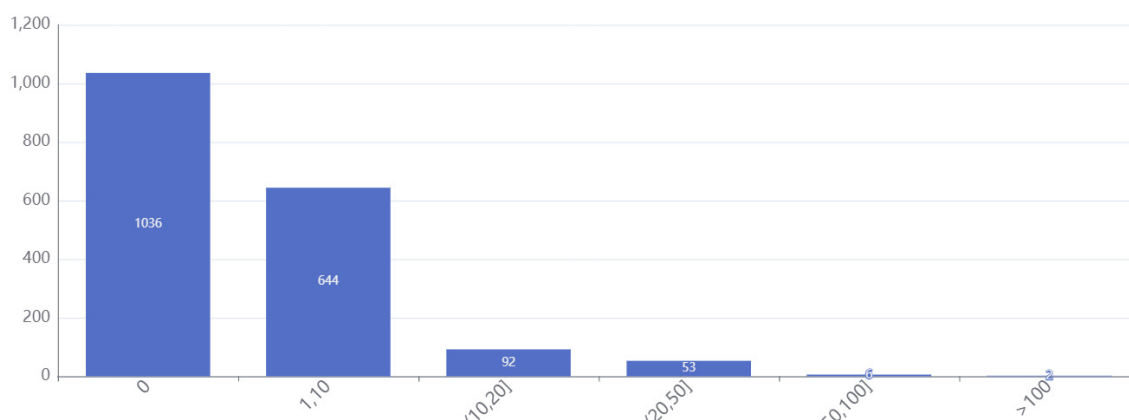


Figure 2. Literature distribution from different journals



papers have zero citations, indicating that many of the research contents are not understood and recognized by most scholars.



**Figure 4.** Citation frequency statistics of literatures

## 4. Conclusion

To better operate the vocational education group, it is necessary to take charge of the current situation and development trend of the vocational education group. The research of vocational education group based on CNKI data is helpful to excavate the core laws in the literature of vocational education group, and help vocational institute managers to better operate the vocational education group. The system realizes the mining of CNKI literature data through data crawling, data statistics, data interface, data visualization and other technologies. The research has completed the basic data statistics function. In the future, more models need to be built, and the deep learning algorithm needs to be used to mine more information, such as the current hot issues of vocational education groups, the main factors affecting the operation of vocational education groups, etc.

## Acknowledgments

This work was financially supported by the funding of the 13th Five-Year Plan Project of Education Science in Jiangsu Province(B-a/2018/03/22), the fundamental computer education teaching research project (2021-AFCEC-285) of the association of fundamental computing education in chinese universities.

## References

- [1] Zhang, Ying, Jinsi Wang, and Bo Chen. "A Game Analysis on the Government Function in the Process of Vocational Education Group School-running." 2013 Conference on Education Technology and Management Science (ICETMS 2013). Atlantis Press, 2013: 418-421.
- [2] Lu, Chunmei, and Hengshi Huang. "Research on Operation Mechanism of Regional Vocational Education Group under Perspective of Secondary and Higher Vocational Education Connection." 4th International Conference on Management Science, Education Technology, Arts, Social Science and Economics 2016. Atlantis Press, 2016: 811-814.
- [3] Donthu, Naveen, et al. "How to conduct a bibliometric analysis: An overview and guidelines." *Journal of Business Research* 133 (2021): 285-296.
- [4] Tong, Zheng, et al. "Quality of randomized controlled trials of new generation antidepressants and antipsychotics identified in the China National Knowledge Infrastructure (CNKI): a literature and telephone interview study." *BMC medical research methodology* 18.1 (2018): 1-11.

- [5] Lv, Taizhi, Zhiyang Song, and Chenyong He. "Research on ship dynamic map based on AIS." 2021 International Symposium on Computer Technology and Information Science (ISCTIS). IEEE, 2021; 276-279.
- [6] Zou, Yujuan, Peiyi Tang, and Taizhi Lv. "Design and implementation of ship shore power data analysis system based on Doris data warehouse." 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). IEEE, 2022: 367-370.
- [7] Zhang, Juan, and Taizhi Lv. "Design and implementation of ship data analysis platform." *Frontiers in Computing and Intelligent Systems* 2.1 (2022): 20-22.
- [8] Lv, Taizhi, Jun Zhang, and Chenyong He. "Research on posts analysis based on data process automation." 2021 2nd International Symposium on Computer Engineering and Intelligent Communications (ISCEIC). IEEE, 2021: 156-159.
- [9] Price, Derek de Solla. "A general theory of bibliometric and other cumulative advantage processes." *Journal of the American society for Information science* 27.5 (1976): 292-306.