ISSN: 2688-8653

Literature Analysis on the Research of the Three Education Reform

Jun Zhang*, Taizhi Lv

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

* Corresponding Author: 1052871890@qq.com

Abstract

The "Three Education Reform" is a crucial driving force for the development of higher vocational education in China. Its goal is to adapt to the demands of socio-economic development, cultivating a greater number of high-quality, high-skilled talents. Traditional teaching methods, materials, and pedagogies may no longer be suitable for the requirements of modern society for skill-oriented talents. Through the "Three Education Reform", students can receive a higher quality of education that better meets societal demands. To study the current research status of the "Three Education Reform" and reveal its development status and context, this article utilizes big data technology. Using Python's web scraping techniques, literature information with "Three Education Reform" as the keyword is extracted from the CNKI (China National Knowledge Infrastructure) database. The literature information is stored in a MySQL database and is presented using a frontend-backend separation technique.

Keywords

Literature Analysis, CNKI, Three Education Reform, Data Visualization, Web Crawling Technology.

1. Introduction

With the rapid development of China's economy, the demand for high-quality skilled talents is also increasing. Traditional education models can hardly meet society's needs for skills and knowledge. The development of new-generation information technologies such as big data, large language models, blockchain, and deep learning provides new platforms and methods for education, requiring higher education to adapt accordingly [1-2]. The "Three Education Reform" refers to the reform in the fields of "teaching, textbooks, and pedagogy". Its core goal is to enhance the quality of higher education and better adapt to the rapid developmental demands of society, technology, and the economy. The "Three Education Reform" aids higher education in aligning more closely with the socio-economic development needs, fostering talents more in line with market requirements. By reforming teaching methods, materials, and pedagogy, education can be made more targeted, thus improving students' learning outcomes. Under the new educational paradigm, greater emphasis is placed on cultivating students' innovative thinking and practical abilities. Integrating new-generation IT and teaching methods can make education more modern and international. The "Three Education Reform" is a long-term and intricate process that requires continuous research, practice, and refinement. It's foreseeable that the "Three Education Reform" will have a profound impact on the development of higher education in China and talent cultivation.

Literature feature data analysis refers to employing statistical and data analysis methods to study the content, structure, authors, publication time, citation times, and other features of literature [3]. It's a cost-effective and efficient quantitative evaluation method with accuracy and objectivity. Through statistical analysis of the number of documents, topics, and keywords,

ISSN: 2688-8653

research hotspots, frontiers, and trends in a particular discipline or field can be revealed. By classifying and analyzing literature topics and keywords, a knowledge system and structure of a certain field can be constructed, further exploring its evolution process and pattern. Data analysis of literature features assists in devising more effective retrieval strategies, enhancing literature search accuracy and efficiency. It offers a macroscopic approach to understand and study academic research, helps unveil the intrinsic rules of academic research, promotes academic exchange, enhances research efficiency, and supports related decision-making. CNKI (China National Knowledge Infrastructure) is China's largest academic information repository, offering rich literature resources like academic papers, master's and doctoral theses, journals, newspapers, conferences, yearbooks, statistical data, etc.

This paper utilizes the literature resources provided by CNKI, applying data analysis methods to study the content, authors, publication time, citation times, etc., revealing the current research status and development trends of the “Three Education Reform”.

2. Literature data analysis process

To achieve automated analysis of literature data, this paper constructs a framework based on web crawling technology to collect data from CNKI, analyze, store, and visualize the data. This framework includes a data collection layer, a data computation layer, a data storage layer, and a data visualization layer. Figure 1 shows the system architecture.

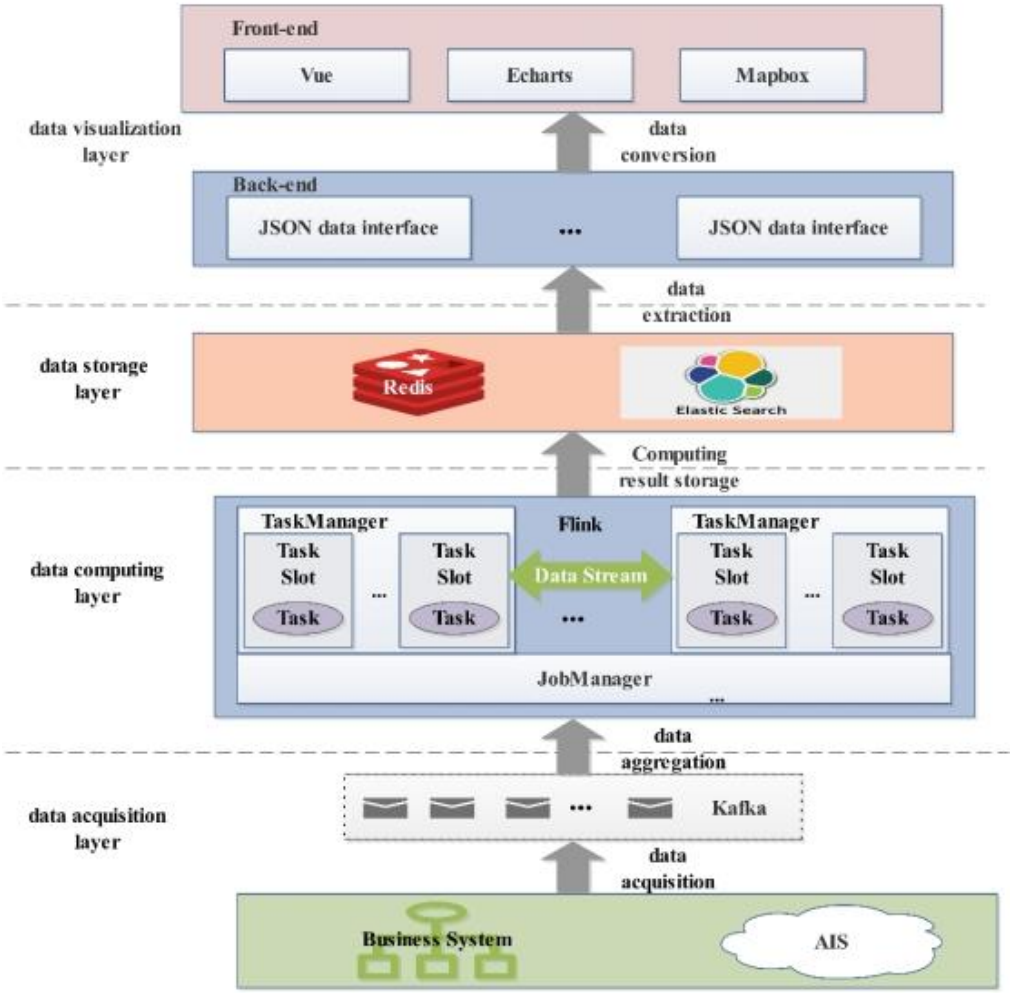

Figure 1: The system architecture

ISSN: 2688-8653

2.1. Data crawling

The data source of the study is literature with the title “Three Education Reform” collected from China Knowledge Network (CNKI). The collected information includes paper title, author, journal name, publication time, number of citations, number of downloads, abstract, keywords, funds, classification number, number of pages, inclusion status, and other information. Data crawling is accomplished based on the Request and Beautiful Soup components. The specific process is as follows:

Step 1: Use the Request library to search the CNKI database with “Three Education Reform” as the keyword, setting the request header to simulate browser behavior.

Step 2: Parse the document using the BeautifulSoup library. Use Beautiful Soup to parse the HTML content of the list page obtained, and get the root DOM object of this document. By calling the find method of the root document object, find the table item where the list is located, then traverse all the tr items under this table item, and obtain the paper title, author, journal name, number of citations, number of downloads, and the link address of the detailed page through the tr item. The literature information is stored in the Redis cache, including the name, author, link address, etc., of the literature.

Step 3: Find the link address of the next page through the root document, request the next page through the Request library, and jump to Step 2 until all pages are crawled.

Step 4: Extract basic literature information from the cache in turn, crawl literature details based on the hyperlink of detailed literature information, including abstract, keywords, index, number of pages, etc. The literature information is stored in the MySQL database.

2.2. Data cleaning and processing

To ensure the quality of the data and the accuracy of the analysis, it is necessary to clean the literature data obtained from China National Knowledge Infrastructure (CNKI). The main cleaning operations include deduplication, handling missing values, format standardization, and anomaly handling.

Deduplication: Due to possible repeated scraping of some literature during the data extraction process, the initial step involves making sure each piece of literature is unique. Typically, deduplication can be done based on the literature's title, authors, and publication date.

Handling Missing Values: Examine the dataset for missing values. Depending on their significance and the reason for the missing data, decide whether to fill in or delete them. For critical information like the literature title or authors, if missing, you might need to discard that record. For non-critical data, such as download counts or citation counts, filling in with zeros might be appropriate.

Format Standardization: Ensure that all date, number, and text data follow a standardized format. For instance, convert dates into a consistent “YYYY-MM-DD” format.

Anomaly Handling: For numeric data, such as citation counts, detect and treat any anomalies or outliers.

Text Cleaning: For text data, such as literature titles and abstracts, remove any unnecessary spaces, punctuation, and special characters.

The statistical and analytical evaluation of literature feature data helps scholars stay informed of academic progress, grasp research hotspots and trends, identify gaps and blind spots in research, promote academic exchanges, improve research efficiency, and provide support for related decision-making. Pandas is an open-source data analysis library that offers high-performance, user-friendly data structures and data operation tools for data cleaning and analysis [5]. This study relies on the Pandas library for the statistical evaluation and analysis of literature data. The results of the statistics and analysis are then stored in a MySQL database.

ISSN: 2688-8653

2.3. Data visualization

This paper employs a frontend-backend separation model to realize data visualization. Front-end and back-end separation is a modern Web application architecture in which the frontend and backend are developed and deployed as two distinct entities. The frontend is responsible for the user interface and experience, while the backend deals with data processing, business logic, and interactions with the database.

The back-end is developed using Spring Boot and MyBatis to retrieve analyzed data and provides services in the JSON data format. The backend employs Spring Boot to implement the interface and uses MyBatis to achieve the mapping between database records and Java objects [6]. Additionally, it uses Druid for database connection pooling. MyBatis is a Java-based persistence framework that supports object mapping. It has built-in JDBC, which simplifies database operations. The data interaction between the frontend and backend is facilitated through the Spring Boot framework, adopting the three-layer architecture model, which consists of Controller, Service, and Dao layers.

The front-end is built using the Vue framework. The Vue-cli tool is used to swiftly create projects, and the layout of the frontend is developed using Element UI components. The entire page layout is realized by installing the Bootstrap dependency. Vue is a progressive JavaScript framework based on the MVVM (Model-View-ViewModel) pattern, allowing for the rapid construction of frontend applications [7]. The front-end sends asynchronous requests using Axios. Upon receipt of the JSON data from the backend, this data is bound to the Model and synchronized with the ECharts visualization through the ViewModel. ECharts is a data visualization library based on JavaScript [8]. It can present non-linear data graphically, making it easier for researchers to discern patterns hidden within the literature.

3. Analysis of literature statistics results

3.1. Analysis of basic literature information

The number of publications is one of the significant indicators to measure the level of attention a research field receives [9]. An annual analysis of publication numbers can reflect the changing interest in that research domain. As shown in Figure 2, research on the “Three Education Reform” began in 2019, and its research trend has been rising year by year.

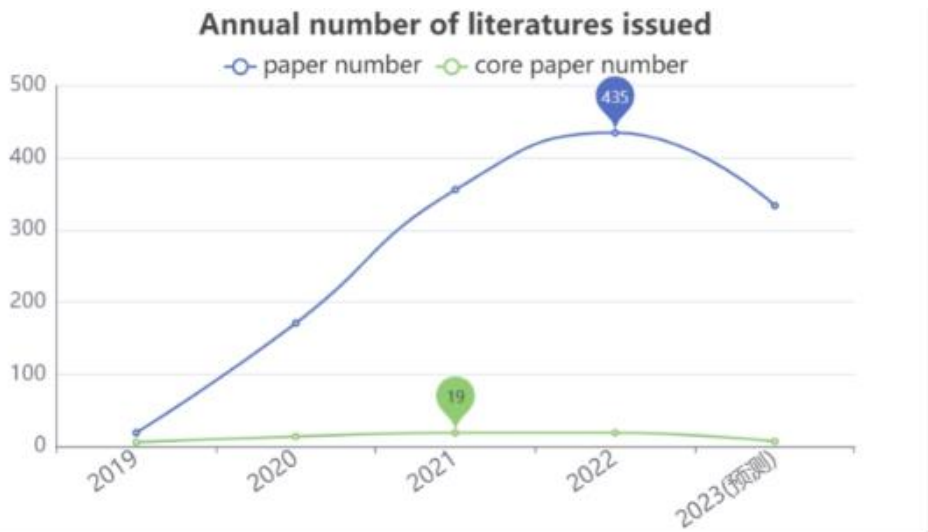

Figure 2: The curve of the annual number of literatures issued

ISSN: 2688-8653

By 2022, the number of publications on “Three Education Reform” research increased to 435, indicating the growing interest and heat around the “Three Education Reform”. This has attracted numerous scholars to delve into research on the “Three Education Reform”.

Statistical analysis of journals in which articles are published helps researchers better understand the publishing trends in an academic field and select appropriate journals for manuscript submission. As can be seen from Figure 3, 1,190 articles are spread across 408 journals, indicating that the journals publishing on this topic or field are quite dispersed. Although publications on the “Three Education Reform” are scattered across various journals, some journals have a higher publication count, such as "Modern Vocational Education," which has published 55 articles. By examining the top 10 journals by publication volume, potential authors can get a clearer idea of which journals hold more influence or are more popular in the field of “Three Education Reform” research.

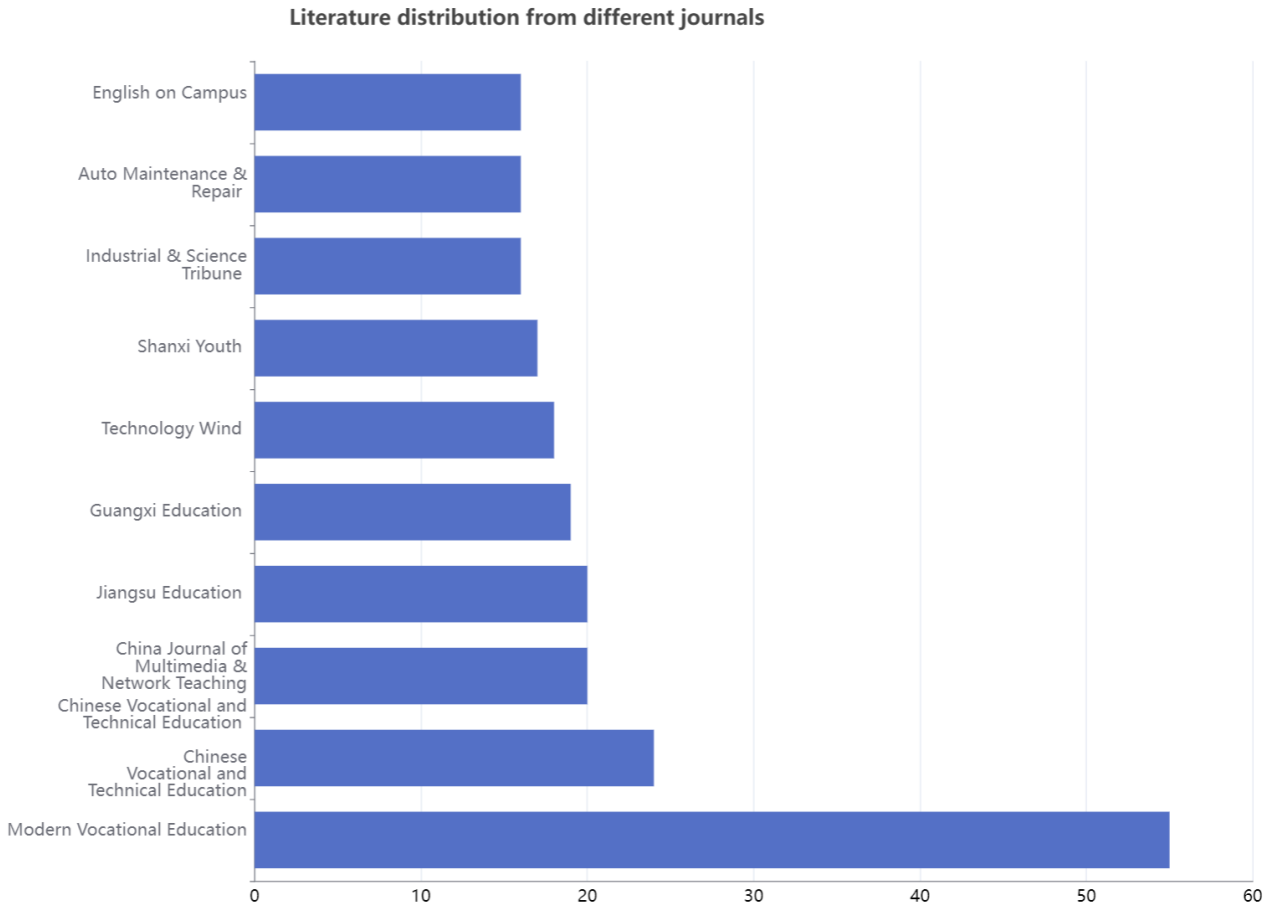

Figure 3: The histogram of the literature distribution from different journals

Conducting a statistical analysis of the institutions to which the authors belong is beneficial to understanding the identity of the authors and grasping the distribution of research capabilities. Figure 4 shows the histogram of the literature distribution from first author. According to statistics, the first authors of 1,190 articles come from 816 institutions, showing that research capabilities are quite dispersed. Looking at the number of publications from each institution, there are 10 institutions that have published 6 or more papers. These institutions have strong potential and strength in the field of “Three Education Reform”research.

ISSN: 2688-8653

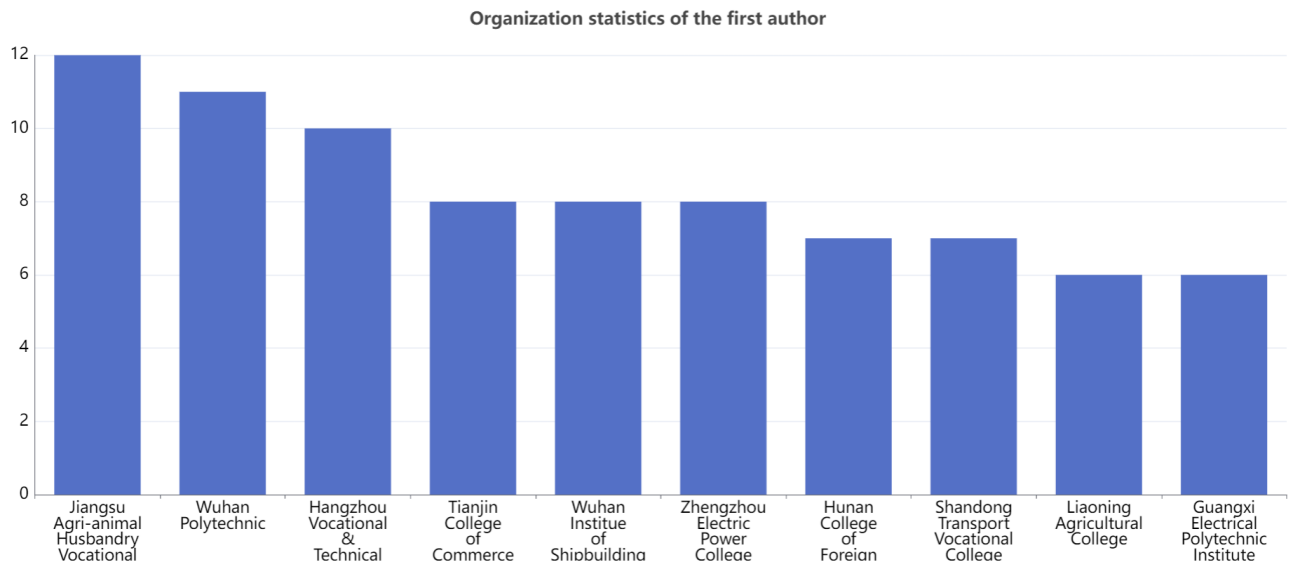

Figure 4: The histogram of the organization distribution from first author

The number of papers published by each researcher (the statistics of the first author) is related to the measurement of core authors in the research field of "Three Education Reform". A small number of core authors in a research field often promote academic innovation and discipline development. The first scholar published 9 papers, and the second scholar published 4 papers. According to the internationally popular Price core author calculation formula, the core schoars is calculated [10].

$$M \approx 0.749 \times \sqrt{N_{max}} \quad (1)$$

where Nmax is number of papers published by the largest number of authors. After calculation, the number of papers selected as the core scholar of the vocational education group research is 3. According to the statistical results, the minimum publication count for core authors in "Three Education Reform" research is 3 papers. There are 9 researchers who have published three or more papers, with a total of 40 articles, accounting for 3.3% of the total number of papers. This is far below Price's proposition that the publication volume of core authors is approximately 50% of the total publication volume of all authors. Based on these statistics, a fully-formed core group of authors for "Three Education Reform" research has yet to emerge. The research on "Three Education Reform"is still growing and lacks a stable group of highly productive core authors to continuously push the development of the field. Given the importance and impact of "Three Education Reform"research, this field will attract more researchers in the future. Over time, a core group of authors will gradually form, propelling the research on "Three Education Reform"to greater depths and breadth.

3.2. Keyword analysis

Word clouds offer us an intuitive tool that enables a quick glance at the most frequently appearing keywords in a dataset [11]. These keywords reflect the focal points and trends of research. Keywords are vital components of academic papers. With just a few words, word clouds can represent the theme of a document, encapsulating its content features and research domain. Analyzing high-frequency keywords can unveil the hotspots of a discipline. Figure 5 presents the keyword word cloud for the "Three Education Reform" literature data.

ISSN: 2688-8653

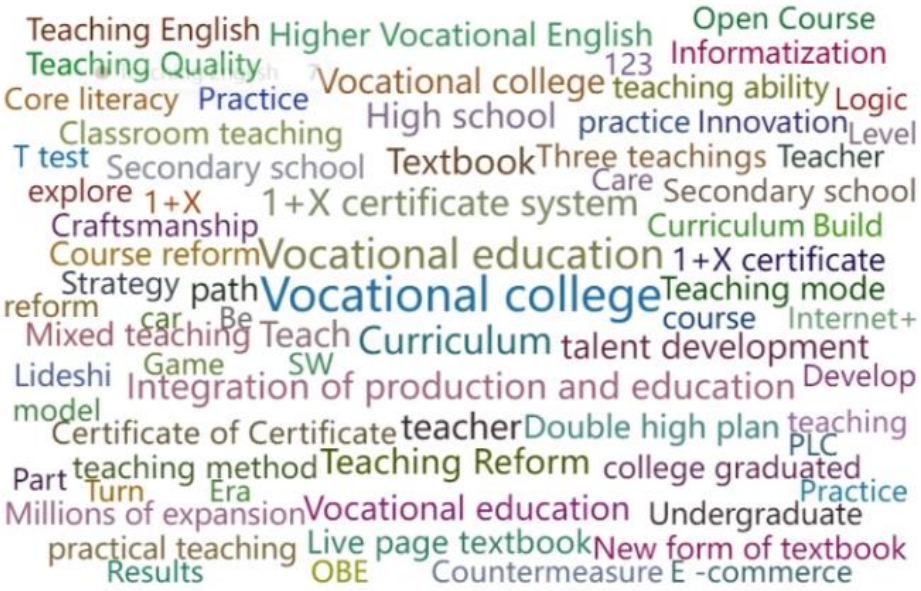

Figure 5: The word cloud diagram about keywords

From Figure 5, it can be seen that "1+X", "blended learning", "textbooks", and "vocational colleges" are focal topics within the “Three Education Reform” research.

3.3. Citation analysis

The citation frequency is an essential metric to measure the influence of an academic document. When a paper is cited many times, it indicates that the document has made a significant impact in its research field or provides key information in a domain. Citation frequency can reflect the influence and importance of a research topic or a specific author in the academic realm. As shown in Figure 6, five papers have been cited more than 100 times, suggesting that these documents might be pivotal in the “Three Education Reform” research field, offering new perspectives, methodologies, or data. The statistics also indicate that over 70% of the documents have a citation count of zero.

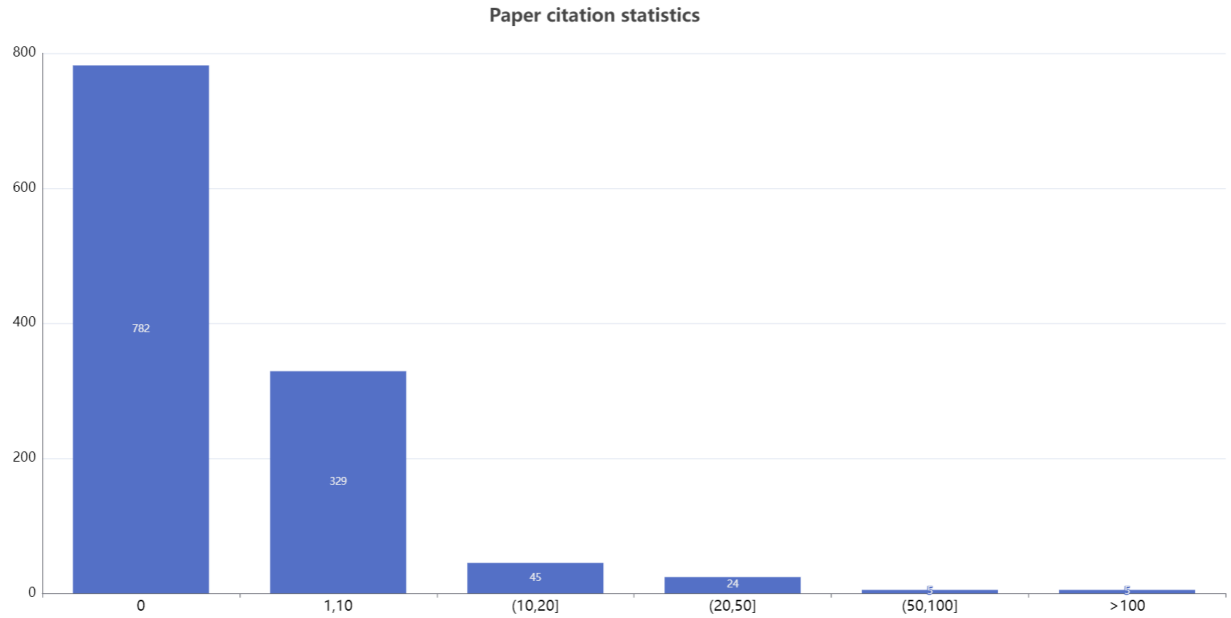

Figure 6: The histogram of the citation frequency statistics of literatures

ISSN: 2688-8653

Figure 6 reveals that a vast number of papers in the “Three Education Reform” research have not garnered widespread attention and recognition. Several reasons could account for this:

Quality issues: Some research might not be of high caliber, with unreliable methods or conclusions.

Research focus: Certain papers might tackle topics that aren't current hotspots or mainstream directions in the field.

Promotion: Authors or publishing entities might not have effectively promoted these works, resulting in insufficient attention.

Language and accessibility: Some studies might be published in relatively niche journals or conferences, or language and content could limit their audience scope.

Although citation frequency is a valuable indicator, it shouldn't be the sole standard for assessing the quality of research or a document. It mostly mirrors the influence and notoriety of a paper, but it doesn't necessarily equate to the actual value or quality of the research.

4. Conclusion

The three aspects of the “Three Education Reform” complement each other and are essential for improving education quality, cultivating innovative talents, enhancing national competitiveness, promoting educational equity, and meeting diverse needs. Research on the “Three Education Reform” based on CNKI (China National Knowledge Infrastructure) data can help unearth patterns in the reform's study, assisting educators in enhancing their teaching methods. The system has achieved data mining of CNKI literature through techniques such as data crawling, data statistics, data interfaces, and data visualization. While the research has completed basic data statistics functions, there's a future need to establish more models and adopt deep learning algorithms to extract more insights, such as current hot topics in the “Three Education Reform” and the primary factors influencing the reform's progress.

Acknowledgements

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 Jiangsu Science and Technology Think Tank Youth Talent Plan Project (JSKJZK2023065), and the Excellent Teaching Team for Qing Lan Project of the Jiangsu Higher Education Institutions of China (Big Data Technology Teaching Team with Shipping Characteristic).

References

[1] Wang, Ling, Qin Yang, and Mengting Zhong. "Research on the Development and Implementation of a New Type of Loose-leaf Nursing Introduction Textbook under the Background of the" Three Education Reforms"." International Journal of Social Science and Education Research 6.6 (2023): 442-444.

[2] Zhang, Min, and Xingsheng Yu. "The construction of teaching quality evaluation system of modern apprenticeship based on big data." Journal of Physics: Conference Series. Vol. 1578. No. 1. IOP Publishing, 2020.

[3] Wang, Weihong, and Chang Lu. "Visualization analysis of big data research based on Citespace." Soft Computing 24 (2020): 8173-8186.

[4] Liu, Qing, Juan Wang, and Jiuju Zhang. "Research on national park education and interpretation based on China national knowledge infrastructure visualization analysis." 2019 6th International Conference on Information Science and Control Engineering (ICISCE). IEEE, 2019 :485-489.

[5] Lemenkova, Polina. "Processing oceanographic data by Python libraries NumPy, SciPy and Pandas." Aquatic Research 2.2 (2019): 73-91.

ISSN: 2688-8653

[6] Li, Yao Zhang, et al. "Research and application of template engine for web back-end based on MyBatis-Plus." Procedia Computer Science 166 (2020): 206-212.

[7] Li, Minghang, Jianghai Hu, and Xianwu Lin. "The Development of Web Application Front-End of Intelligent Clinic Based on Vue. js." Proceedings of 2019 Chinese Intelligent Automation Conference. Springer Singapore, 2020: 683-690.

[8] 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.

[9] Li, Junjie, et al. "Policy Analysis on Recycling of Solid Waste Resources in China—Content Analysis Method of CNKI Literature Based on NVivo." International Journal of Environmental Research and Public Health 19.13 (2022): 7919.

[10] 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.

[11] Heimerl, Florian, et al. "Word cloud explorer: Text analytics based on word clouds." 2014 47th Hawaii international conference on system sciences. IEEE, 2014: 1833-1842.