CIT. Journal of Computing and Information Technology, Vol. 32, No. 4, December 2024, 265–276
doi: 10.20532/cit.2024.1005865 

# Big Data Related Patent Retrieval System Based on Filtering Rules

Xuanrong Zhao

Sichuan Top IT Vocational Institute, Sichuan, China

In the information age, patents are an important carrier of scientific research achievements. How to protect patents and effectively transmit them has become an important development measure in the current information age. A big data-related patent retrieval technology based on filtering rules is proposed to address issues such as poor keyword and phrase retrieval positioning in patent analysis. This new technology combines multiple filtering and retrieval methods and builds a data storage and transmission system. The model achieved the best performance when the threshold was set to 100. The frequency of using the training set before and after keyword filtering increased by 10 and the frequency of using the test set before and after keyword filtering increased by 16. The Euclidean distance of the research method decreased by 0.883 compared to other methods. The mean value increased by 0.1611 compared to other methods. The cosine value increased by 0.4300 compared to other methods. Therefore, the new method has a better filtering effect on patent keywords compared to other methods. This has a good guiding effect on the retrieval of big data-related patents.

*ACM CCS (2012) Classification:* Information systems → Information retrieval → Retrieval tasks and goals → Document filtering

*Keywords*: patents, big data, filtering rules, retrieval

## 1. Introduction

When information technology develops, data grows exponentially. The generation and accumulation of massive data have made the processing and analysis of big data (BD) a hot research topic in academia and industry [1]. BD technology mainly manages, processes, and analyzes data effectively, enabling more useful data to quickly and accurately find relevant patent information from massive databases. Traditional patent retrieval methods often appear inadequate in the face of large amounts of data and complex retrieval needs. Therefore, more convenient and efficient BD retrieval technology has become a focus of current research [2]. As an important carrier of technological innovation, patents have significant importance in terms of retrieval and analysis in science, technological research and development, market competition, and other fields [3]. Traditional patent retrieval methods mainly rely on keyword matching and classification retrieval, which suffer from low retrieval efficiency and low accuracy of results. As BD technology develops, the magnitude and complexity of patent data continues to increase. Traditional methods are no longer able to solve problems such as patent search [4]. Filtering rules can perform preliminary screening of patent information in the BD environment, filtering out irrelevant or low correlation data, thereby improving the efficiency and accuracy of retrieval. Based on this, a new patent retrieval method based on filtering rules is proposed and developed in this work. This new technology innovatively combines the advantages of multiple methods to enhance the search ability for patent keywords. The use of Django as the overall framework for data processing can improve the data processing ability for filtered patent keywords.

## 2. Related Work

Patent search is a technology that uses artificial intelligence to analyze and screen patent keywords and patent data. In recent years, more and more different patent retrieval techniques have been applied to patent analysis and application. Björkqvist *et al.* [5] proposed a graph-based patent search engine to address the challenges of existing technology retrieval in patent writing and invalidation claims. This engine simulated the work of professional examiners through graph neural networks and novelty citation data from the patent office. Graph-based methods were an effective approach in patent retrieval, significantly improving retrieval efficiency and accuracy. Van Rijn *et al.* [6] proposed a patent situation analysis method to enhance their understanding of the utility of patent information, which was used to outline the technical field and predict regulatory gaps. Patent situation analysis showed significant effectiveness in revealing scientific exploration, application status, and potential regulatory gaps. This helped biotechnology researchers to effectively retrieve and analyze relevant information, promoting the rapid launch of innovative products. X. Han *et al.* [7] proposed a technology opportunity analysis method based on patent and trademark data to link technology opportunities with commercial applications. By constructing a file-keyword matrix and correlation network, potential technological opportunities were identified based on the existing technological foundation of the enterprise. The new method effectively identified technological opportunities in undeveloped business areas, providing validation for case studies of 3D printing. W. Du *et al.* [8] proposed a patent transaction recommendation method based on a knowledge-aware attention bidirectional long short-term memory network to reduce the technology search cost for patent buyers. The new method captured sequential patterns in the company's historical records and designed attention mechanisms for different technological interests. The new method outperformed the baseline in recommending patent transactions. The attention visualization of randomly selected companies demonstrated the effectiveness of recommendations. F. Yu *et al.* [9] proposed a cross-domain technology recommendation method based on patent analysis to address the difficulty for designers to generate radical concepts in new product development. This method summarized product functions through patent feature words. This method combined subject object structural similarity and functional similarity to quickly extract cross domain technologies and conducted objective evaluations. The new method helped to break through design rigidity, providing radical concepts for new product development, which was successfully applied in peeling equipment projects. K. Song *et al.* [10] proposed a patent digitization system framework that combined neutral language processing and machine learning techniques. This method aimed to improve the efficiency of technology trading and research and development cooperation in industrial enterprises. This framework included patent recommendation, transferability assessment, and detection models. This system could effectively improve the efficiency of technology transactions, assist industrial enterprises in identifying relevant patents and positioning research teams.

In summary, for patent retrieval technology, more and more experts and scholars are using different artificial intelligence technologies to study patent applications and data retrieval. However, most of these methods can only be applied to specific fields. These methods also have limitations in terms of keyword filtering and extraction effectiveness. Therefore, the study aims to improve the retrieval ability of patent keywords and proposes a new patent filtering data storage system by combining different filtering methods.

## 3. Research Methodology

### 3.1. Big Data Patent Retrieval Method Based on Filtering Rules

In practical applications, when searching for patents, patent owners often abstract the information of patents. This makes the entire BD patent search process more difficult. To this end, filtering rules are used to filter patent keywords in patent search. The C-value method is utilized to screen patent phrases. Equation (1) is a phrase selection process [11].

$$C_x = \begin{cases} \log|x| F_x \\ \log|x| \left( F_x - \frac{\sum_{b \in T_x} F_x}{N(T_x)} \right) \end{cases} \quad (1)$$

In equation (1), $C_x$ is the patent keyword phrase obtained through screening. $x$ is the phrase character to be filtered. $F_x$ is the frequency at which the entire phrase appears in the vocabulary. $T_x$ is a set of filtered phrase characters. $N(T_x)$ is the total number of filtered phrases. Keywords and their phrases are filtered through C-value. Word2vec is utilized to complete the transformation of phrases and vectors. To improve the data statistics and analysis capabilities of the entire retrieval technology, the Term Frequency–Inverse Document Frequency (TF-IDF) is utilized as the data weighting method. Equation (2) is the TF-IDF regular phrase's weight [12].

$$tfidf = tf_{xy} * xdf_x \quad (2)$$

In equation (2), $tf_{xy}$ represents the occurrences of the phrase $x$. $idf$ is the number of times a document is retrieved in reverse. Equation (3) is the probability of a phrase [13].

$$tf_{xy} = \frac{n_{xy}}{\sum_{x=1}^{k} n_{ky}} \quad (3)$$

In equation (3), $n_{xy}$ represents the total probability of the occurrence of $x$ in the corresponding corpus. $k$ is the total number of texts in the corpus. $\sum_{x=1}^{k} n_{ky}$ is the sum of the total probabilities of all phrases appearing in the total corpus document. The times a document can be retrieved in reverse can be represented by equation (4).

$$idf_x = \log \frac{|D|}{|y; t_x \in d_y|} \quad (4)$$

In equation (4), $|D|$ represents the total number of retrieved texts in the total phrase library. $|y; t_x \in d_y|$ is the quantity of texts for all phrases. In practical operation, if the phrase does not appear in any text, that is, if the total text amount of the current phrase is 0, it will result in the variable of the denominator in equation (4) being 0. Therefore, to avoid this situation, when using filtering rules for BD patent search, the denominator in equation (4) is replaced, so that the total denominators increase by a new text amount on the original basis. When filtering patent phrases, it is assumed that the phrases appearing in the text exceed the current preset number. The phrase expression at this point can be represented by equation (5) [14].

$$\begin{cases} p = \{ p_r \mid f(p_y) \in \sigma \quad \text{and} \quad p_r \in RBF\text{-set} \} \\ C = \{ p_x \mid p_x \in p \quad \text{and} \quad p_x \in RBF\text{-set} \} \end{cases} \quad (5)$$

In equation (5), $p$ refers to the occurrences of phrase sets in the text not less than the set threshold. $RBF$-set is a text keyword obtained without filtering rules. $p_r$ refers to the total number of phrases that do not use filtering rules. $p_x$ refers to a set of phrases that do not use filtering rules. $C$ refers to the quantity of phrases in a phrase that exceeds a preset threshold. If the phrase situation after using filtering rules is better than that without using filtering rules, then the current filter used is considered as a usable filter. Equation (6) is the situation where the phrase is better after use than before use [15].

$$\text{number}(x) = \sum_{x=1}^{n} w_x * fc(p_x) \quad (6)$$

In equation (6), number($x$) is the cumulative phrase frequency of the first phrase appearing in unfiltered phrases. $w_x$ is unfiltered phrases' weight. $fc(p_x)$ is the number of times a character appears in the phrase text. The weight is represented by equation (7).

$$w_x = \begin{cases} 0.1, p_x \in C \\ 0, p_x \notin C \end{cases} \quad (7)$$

In equation (7), when the set of phrases that do not use filtering rules belongs to phrases that exceed the threshold, the phrase weight is 0.1. When the set of phrases that do not use filtering rules does not belong to phrases that exceed the threshold, the phrase weight is 0. At this point, the frequency difference between the phrases before and after filtering is represented by equation (8) [16].

$$D(x, \alpha) = \text{number}(x) - \text{number}(x + \Delta\alpha), \quad \text{where } x \in \{10, 10 + \Delta\alpha + 2\Delta\alpha, \ldots\} \quad (8)$$

In equation (8), $D(x, a)$ represents the difference in the occurrences between the $x$-th and $(x+\Delta x)$-th phrases. By using filtering rules, different patent phrases can be retrieved and analyzed. More critical phrase idiomatic corpora can be extracted from them. Figure 1 shows the BD patent retrieval under filtering rules.

From Figure 1, when conducting BD patent search, the process should first analyze the patents that need to be searched, obtain corresponding patent data, and obtain keywords and phrases that have not been searched. Furthermore, the phrase data are filtered using the above filtering rules. A more comprehensive filter is designed to set filter thresholds and parameters. Secondly, the filtered data are subjected to noise filtering to extract more useful data information. Finally, keyword and phrase data associated with "BD" are obtained based on relevant data.

## 3.2. Design of a Big Data Patent Retrieval System

Due to the use of filtering rules to retrieve patent phrases or keyword data, there is usually a situation of data redundancy. Therefore, in the analysis of BD patent search phrase data, the system should preprocess and analyze the data [17]. Figure 2 shows the BD patent retrieval system.

In Figure 2, the patent retrieval system consists of three main modules, namely data processing, keyword retrieval, and patent data integration. The data processing module is mainly responsible for processing the retrieved data information and removing the filtered and more complex keyword data. The keyword search module is mainly utilized to search for keywords and phrases in patents, mainly using filtering rules for retrieval. The patent data integration module

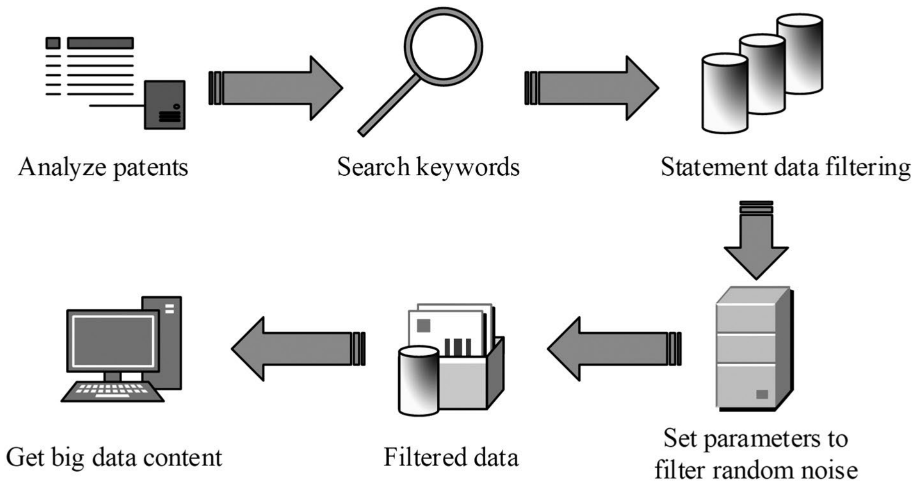



*Figure 1.* Big data patent retrieval process under filtering rules.

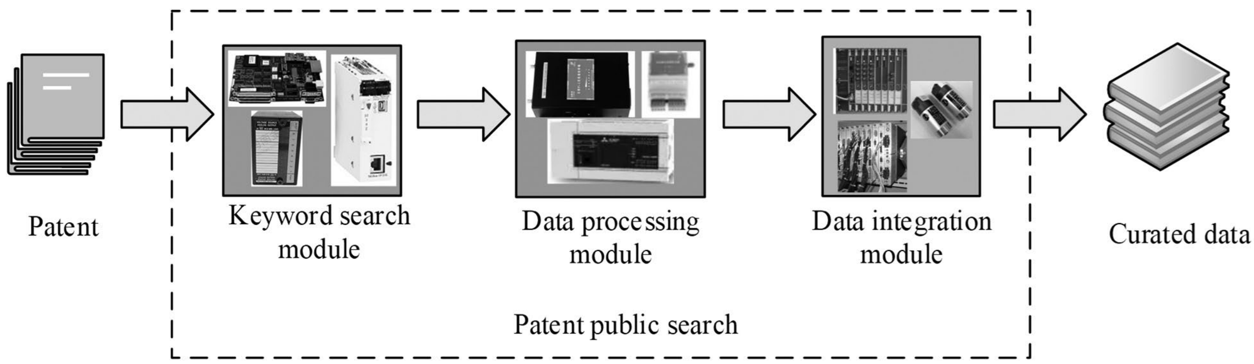



*Figure 2.* The big data patent retrieval system.

mainly integrates the resources of the retrieved and filtered data and integrates the keywords quality. Django was used as the overall framework system for patent data retrieval. Figure 3 shows the Django framework processing.

In Figure 3, when using filtering rules for BD patent search, the system will issue a search request. After receiving a retrieval request from the system, the request will be sent to the intermediate file section, where it will determine whether to return the retrieval situation. If so, the request is directly mapped to enable it to be executed directly. If not, the request is first returned to the system data layer to determine if there are any data anomalies in the current request. If there is a data anomaly, the model's central component is reached, and it is determined whether to return the intermediate file section. If not, it is determined whether to directly return the intermediate file section. If it cannot be directly returned, the function is mapped to the intermediate file and the instruction is executed by executing the view. Figure 4 shows the retrieval process of the keyword retrieval module.

In Figure 4, when conducting patent keyword search, it should first obtain candidate BD patent data information for these keywords. Based on this data information, suitable BD patent keywords are concatenated. Based on the concatenated patent keywords, a list of patent candidates is recommended. Meanwhile, filtering and searching are performed by generating search formulas during recommendation. TF-IDF is utilized during retrieval. A secondary search evaluation is conducted based on the search characteristics of the patent. Finally, it is determined whether the search results need to be filtered. If necessary, screening is carried out. If not needed, the search results are directly viewed. In data integration, patent data is sourced from BD sources. Therefore, the patent data retrieved throughout the integration cannot be directly stored. For this purpose, the study utilizes an IK tokenizer to store patent data. Figure 5 shows the patent data storage process.

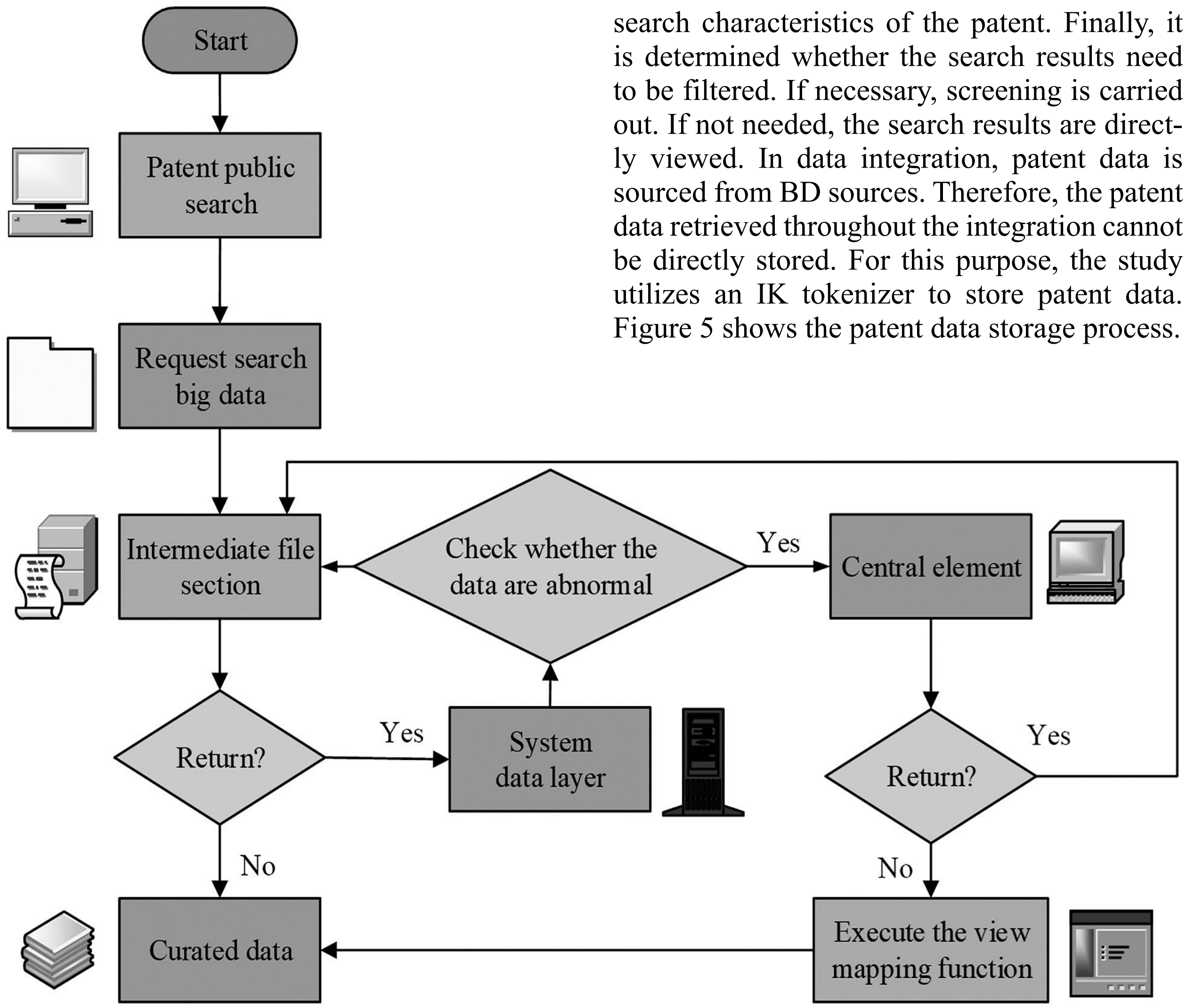



*Figure 3.* Django framework processing.

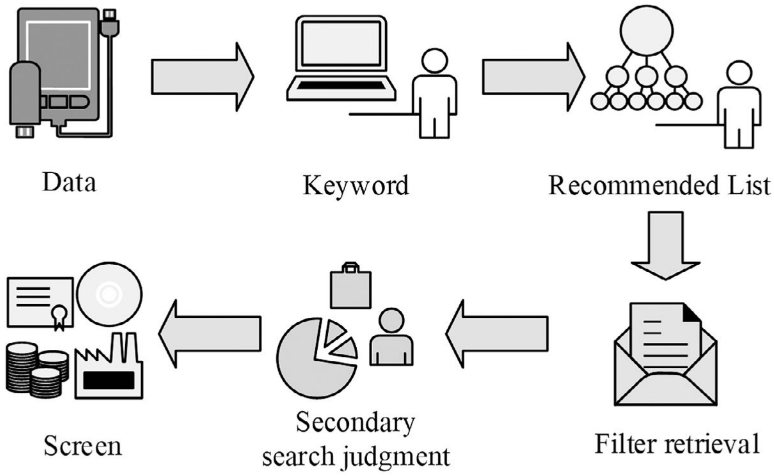



*Figure 4.* Keyword retrieval module retrieval process.

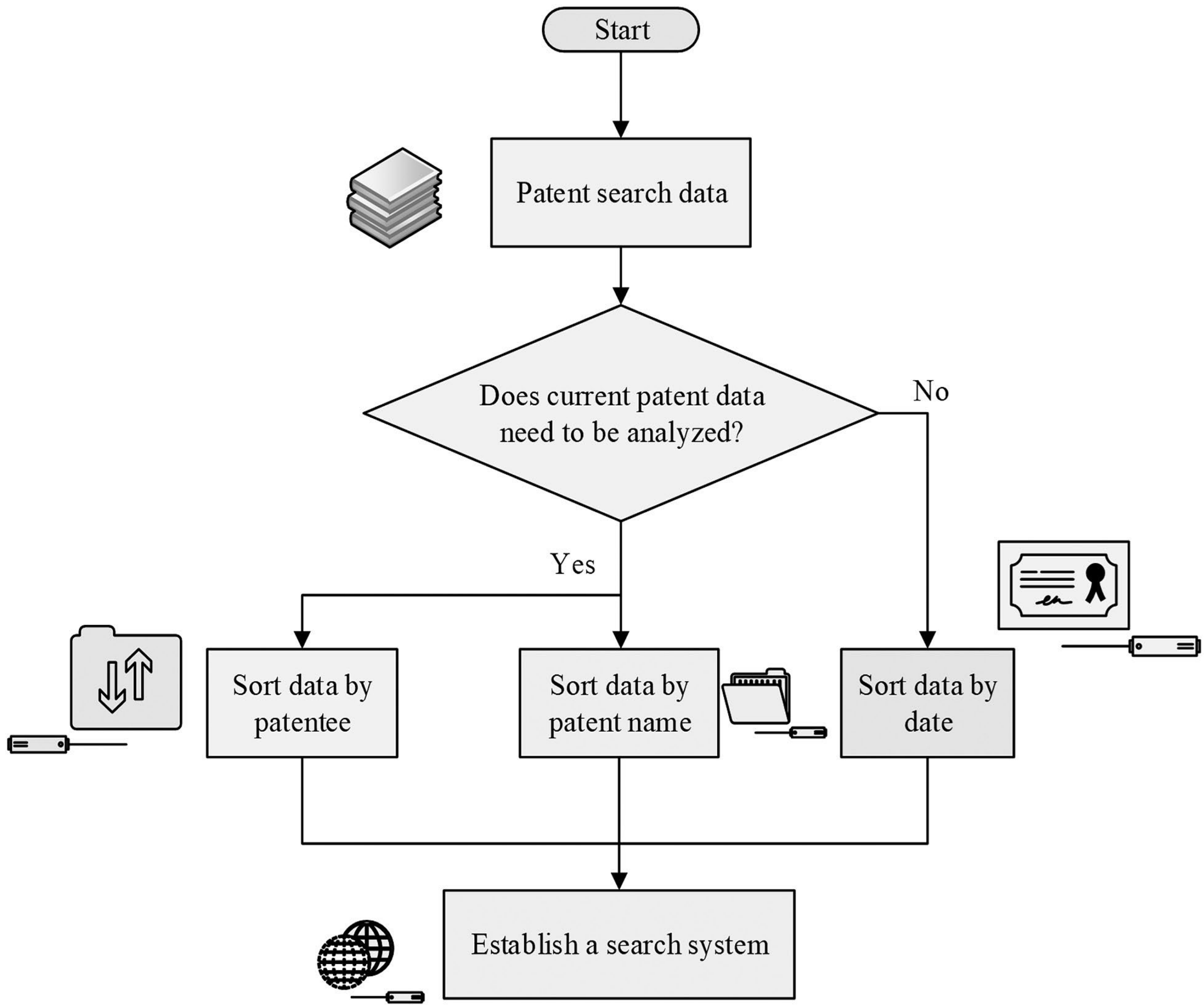



*Figure 5.* Patent data storage process.

In Figure 5, after obtaining patent search data, the data information is first judged. It is determined whether the current patent data needs to be analyzed. If necessary, classification is made based on data such as patent owners and patent names. If not required, these data are directly divided by date. Then, a comprehensive system search engine and database are built for the partitioned data. To test the correlation changes of patents through the system, the Euclidean distance, mean, and cosine values are calculated. Equation (9) is the Euclidean distance [18].

$$OD_k = \sum_{y=1}^{n} (v_y^{(k)} - v_y^{(\theta)})^2 \tag{9}$$

In equation (9), $OD_k$ refers to the distance between two keyword vectors. $n$ is the vector length. $v_y{}^{(k)}$ is a similarity vector ($c_1$, $c_2$, $c_3$). $v_y{}^{(\theta)}$ is the basic vector size. Equation (10) is the mean value.

$$M_k = \frac{1}{n}\sum_{x=1}^{n} v_x,\ k = 1,2,3,4 \tag{10}$$

In equation (10), $M_k$ refers to the vector's mean value. $v_x$ is the vector's element number. $n$ is the filtered phrases number. Equation (11) is the cosine similarity [19].

$$C_k = \frac{\sum_{x=1}^{n} (v_x^{\theta 1} v_x^{\theta 2})}{\sqrt{\sum_{x=1}^{n} (v_x^{\theta 1})^2}\sqrt{\sum_{x=1}^{n} (v_x^{\theta 2})^2}} \tag{11}$$

In equation (11), $C_k$ is the cosine value of the calculated base vector and similarity vector, with values ranging from −1 to 1. The closer to 1, the higher the model's similarity.

# 4. Results and Discussion

## 4.1. Verification of Model Filtering Effect

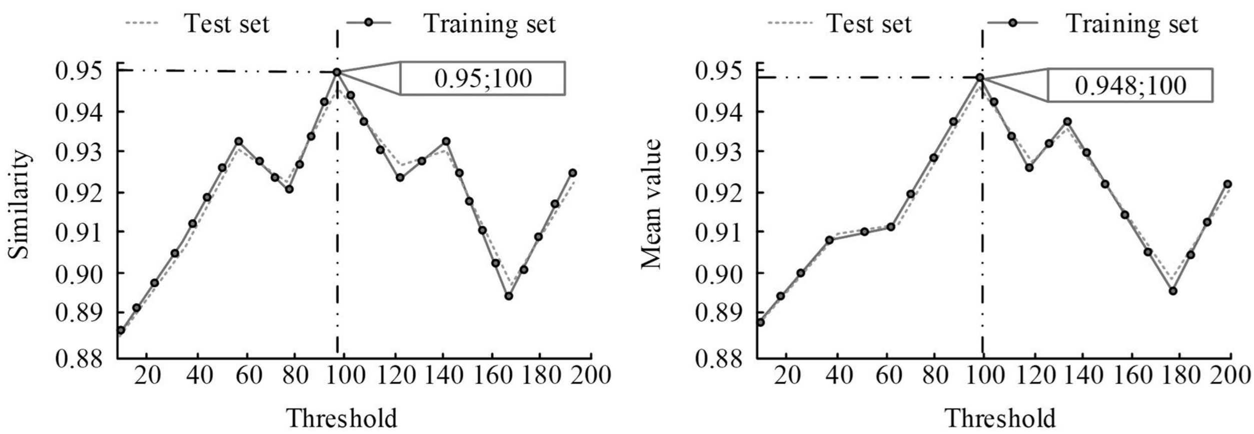
The training set includes encyclopedia style entries and content, article titles and abstracts, and article metadata to better train BD patent keywords. Chinese Wikipedia contains millions of articles. The total text may reach billions of characters. The training and test sets were divided in a 4:1 ratio. According to the formula in the previous chapter, the parameter $\Delta\alpha = 10$ was calculated. The filtered keywords were set to 200. The Chinese Wikipedia dataset was utilized as the training set. The threshold relationship was compared with all phrase sets and filtered keyword sets in the document as presented in Table 1.

From Table 1, when filtering patent keywords and phrases, the model threshold increased, and the keywords of its patents decreased. When the threshold was set to 10, the model filtered out 3354 patent keywords. When the threshold was set to 20, the model filtered out 2684 patent keywords. When the threshold increased from 10 to 20, the filtered patent keywords decreased by 670. This indicated that the filtered keywords increased as the threshold increased. Chinese Wikipedia was utilized for patent keyword training. Figure 6 shows the data variation graph obtained.

The larger the model's cosine value and mean, the better the similarity of the current keywords and the better the filtering effect. From Figure 6 (a), when the threshold was 100, this model achieved the optimal cosine value of 0.95 in the local range. At this point, the better the similarity of keywords, the more obvious the filtering effect on keywords. The curve changes in the test set during cosine analysis were basically consistent with those in the training set, both reaching the optimal cosine value at a threshold of 100. From Figure 6 (b), the model's mean reached its highest value of 0.948 at a threshold of 100. Therefore, when filtering data, the threshold needed to be set to 100. At this point, the keyword filtering ability and filtering ability of the entire model reached the optimal state. The occurrences of keywords before and after filtering were calculated based on changes in parameters and the keywords number, as shown in Figure 7.

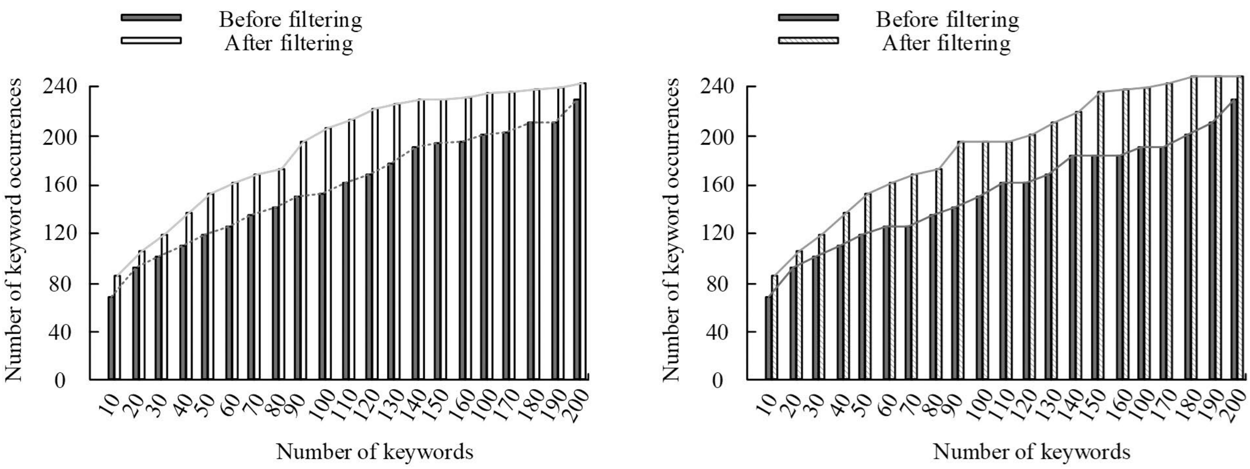
*Table 1*. Comparison of changes in patent keywords and phrases after filtering.

| Threshold | Total patent phrase collections | Number of filtered keywords | Threshold | Total patent phrase collections | Number of filtered keywords |
|---|---|---|---|---|---|
| 10 | 14982 | 3354 | 110 | 1956 | 987 |
| 20 | 9024 | 2684 | 120 | 1758 | 968 |
| 30 | 6564 | 2157 | 130 | 1648 | 935 |
| 40 | 5135 | 2013 | 140 | 1523 | 885 |
| 50 | 4248 | 1684 | 150 | 1425 | 845 |
| 60 | 3536 | 1602 | 160 | 1310 | 803 |
| 70 | 3015 | 1351 | 170 | 1212 | 726 |
| 80 | 2684 | 1245 | 180 | 1048 | 712 |
| 90 | 2354 | 1214 | 190 | 1035 | 698 |
| 100 | 2154 | 1084 | 200 | 1011 | 667 |




*(a)* Comparison of cosine similarity. *(b)* Mean comparison.

*Figure 6.* Comparison of cosine similarity and mean values in the dataset.




*(a)* Training set. *(b)* Test set.

*Figure 7.* Changes in keyword frequency before and after filtering.

From Figure 7 (a), during evaluation on the training set, the frequency of keywords appearing before and after model filtering increased with the increase of keywords. When the keywords were set at 200, the occurrences before model filtering reached 226. At this point, the filtered keywords appeared 236 times. Therefore, the filtered keywords significantly increased, with an overall increase of 10. From Figure 7 (b), during evaluation on the test set, there was a significant increase in the frequency of keyword occurrences before and after model filtering. Similarly to the training set, the frequency of keywords appearing in the test set increased with the increase of keywords. When the model keywords were set at 200, the filtered keyword frequency reached 240. Compared to the 224 times before filtering, the frequency of keyword appearances increased by 16. Therefore, filtering rules effectively increased the frequency of keyword occurrence, indicating that the research model effectively filtered and screened keywords.

## 4.2. System Performance Test Results

Different filtering and data extraction methods were compared and tested to test the actual system performance. TF-IDF, Word2vec, C-value, and SegPhrase were compared and tested with the filtering rules used in this study. The cosine value, mean, and Euclidean distance were utilized as comparative parameter data. The smaller the Euclidean distance, the better the data effect. Also, the larger the cosine value and mean, the better the data effect. Table 2 shows the results obtained.

*Table 2.* Comparison results of parameters for different filtering methods.

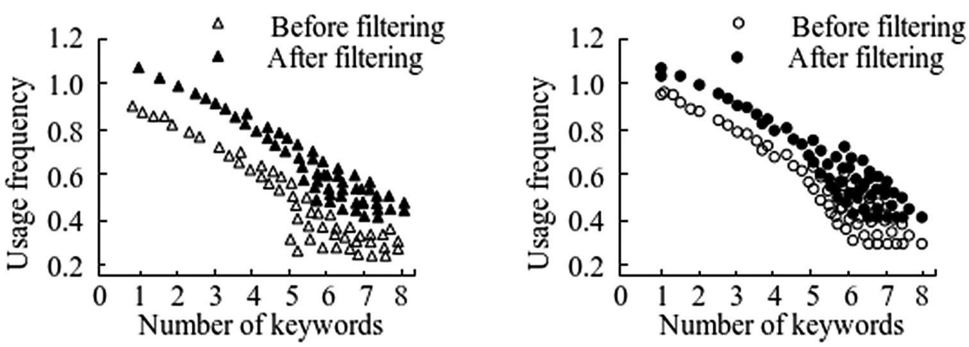
| Model | TF-IDF | Word2vec | C-value | SegPhrase | Filter-refinement |
|---|---|---|---|---|---|
| Euclidean distance | 3.6851 | 3.6845 | 4.3652 | 4.0358 | 3.1528 |
| Mean value | 0.3658 | 0.4582 | 0.4368 | 0.4869 | 0.5269 |
| Cosine value | 1.2684 | 0.9236 | 1.3524 | 1.0365 | 1.3536 |

From Table 2, the filtering rule used in different data parameters comparison showed good testing performance. The minimum Euclidean distance value of the filtering rule was 3.1528, which was 0.883 lower than the maximum value of 4.0358 among several models. This research method's mean value was 0.5269, which was 0.1611 higher than the TF-IDF with the lowest mean. This research model's highest cosine value was 1.3536, which was an increase of 0.4300 compared to Word2vec, which had the lowest cosine value. Therefore, this research method had good filtering performance when considering the effects of different filtering parameters, which might be due to the addition of more data filtering methods. To test the practical application effect of the model, the actual application of different patent keywords after filtering was tested. A comparative analysis was conducted on different types of BD articles, as shown in Figure 8.

From Figure 8 (a), when comparing the filtering effects of keywords before and after journal filtering, the frequency of keyword usage decreased with the increase of keywords. When a certain quantity was reached, the total times tended to be relatively stable. This might be due to the limitation of the number of keywords on this model's filtering effect. The frequency of keyword usage after system filtering was higher, with a significant improvement compared to the frequency before filtering. From Figure 8 (b), the frequency of keyword usage in the journal also decreased with the increase of keywords. However, the usage number after filtering varied less compared to the number before filtering. This might be due to the poor system performance in extracting keywords from journals.

## 5. Conclusion

A new method for filtering patent keywords is proposed in this study, as most authors have unclear descriptions of patent keywords in patent analysis. Multiple data filtering methods were combined, and data processing and storage modules were utilized. Then, a BD patent data processing system based on filtering rules was built. When filtering patent keywords and phrases, threshold settings could affect the filtering effect. When the thresholds were 10 and 20, the filtered keywords decreased by 670. In keyword similarity comparison, when the threshold was set to 100, the model performance reached its best, and the cosine value was optimal at 0.95. The highest mean value also appeared at a threshold of 100, which was 0.948. The frequency of keywords in the training set increased by 10 before and after filtering. The frequency of keywords in the test set increased by 16 before and after filtering. In the comparison of different filtering models, the research method's lowest Euclidean distance was 3.1528, which was 0.883 lower than other methods. This research method's mean and cosine values showed good results among several methods. The highest mean value was 0.5269, an increase of 0.1611 compared to other methods. The highest cosine value was 1.3536, an increase of 0.4300 compared to other methods. Among different article keyword filtering




*(a)* Patent. *(b)* Journal.

*Figure 8.* Comparison of filtering effects before and after different types of articles.

effects, patent keyword filtering had a better filtering effect. Therefore, using filtering rules for patent keyword analysis and filtering has a good effect. Although some achievements have been made in the research, there are still some shortcomings. Subsequent research will explore how to improve the keyword filtering performance of different types of articles. In the future, research will also be conducted on how to reduce keyword duplication during filtering.

## References


[1] M. Suzgun *et al.*, "The Harvard USPTO Patent Dataset: A Large-Scale, Well-Structured, and Multi-Purpose Corpus of Patent Applications", *Advances in Neural Information Processing Systems*, vol. 36, pp. 57908–57946, 2023.

[2] Z. Qiu and Z. Wang, "What is Your Next Invention?—A Framework of Mining Technological Development Rules and Assisting in Designing New Technologies Based on BERT as well as Patent Citations", *Computers in Industry*, vol. 145, no. 11, pp. 103829–103830, 2023. http://dx.doi.org/10.1016/j.compind.2022.103829

[3] Z. Liu *et al.*, "From Technology Opportunities to Ideas Generation via Cross-cutting Patent Analysis: Application of Generative Topographic Map** and Link Prediction", *Technological Forecasting and Social Change*, vol. 192, no. 6, pp. 122565–122566, 2023. http://dx.doi.org/10.1016/j.techfore.2023.122565

[4] Z. Hu *et al.*, "Evaluation and Identification of Potential High-value Patents in the Field of Integrated Circuits Using a Multidimensional Patent Indicators Pre-screening Strategy and Machine Learning Approaches", *Journal of Informetrics*, vol. 17, no. 2, pp. 101406–101407, 2023. http://dx.doi.org/10.1016/j.joi.2023.101406

[5] S. Björkqvist and J. Kallio, "Building a Graph-based Patent Search Engine", in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, vol. 19, no. 6, pp. 3300–3304. http://dx.doi.org/10.1145/3539618.3591842

[6] T. van Rijn and J. K. Timmis, "Patent Landscape Analysis—Contributing to the Identification of Technology Trends and Informing Research and Innovation Funding Policy", *Microbial Biotechnology*, vol. 16, no. 4, pp. 683–696, 2023. http://dx.doi.org/10.1111/1751-7915.14201

[7] X. Han *et al.*, "Discovering Technology Opportunities Based on the Linkage Between Technology and Business Areas: Matching Patents and Trademarks", *Technology Analysis & Strategic Management*, vol. 35, no. 10, pp. 1324–1340, 2023. http://dx.doi.org/10.1080/09537325.2021.2003773

[8] W. Du *et al.*, "Sequential Patent Trading Recommendation Using Knowledge-Aware Attentional Bidirectional Long Short-Term Memory Network (KBiLSTM)", *Journal of Information Science*, vol. 49, no. 3, pp. 814–830, 2023. http://dx.doi.org/10.1177/01655515211023937

[9] F. Yu *et al.*, "An Approach for Radical Innovative Design Based on Cross-Domain Technology Mining in Patents", *International Journal of Production Research*, vol. 61, no. 21, pp. 7502–7523, 2023. http://dx.doi.org/10.1080/00207543.2022.2151659

[10] K. Song *et al.*, "A Digital Analysis System of Patents Integrating Natural Language Processing and Machine Learning", *Technology Analysis & Strategic Management*, vol. 36, no. 3, pp. 440–456, 2024. http://dx.doi.org/10.1080/09537325.2022.2035349

[11] J. Wang *et al.*, "Tracking and Predicting Technological Knowledge Interactions Between Artificial Intelligence and Wind Power: Multimethod Patent Analysis", *Advanced Engineering Informatics*, vol. 58, no. 10, pp. 102177–102178, 2023. http://dx.doi.org/10.1016/j.aei.2023.102177

[12] A. Subramanian *et al.*, "Automated Patent Extraction Powers Generative Modeling in Focused Chemical Spaces", *Digital Discovery*, vol. 2, no. 4, pp. 1006–1015, 2023. https://dx.doi.org/10.1039/D3DD00041A

[13] H. Jiang *et al.*, "Deep Learning for Predicting Patent Application Outcome: The Fusion of Text and Network Embeddings", *Journal of Informetrics*, vol. 17, no. 2, pp. 101402–101403, 2023. http://dx.doi.org/10.2139/ssrn.4122110

[14] C. Jiang *et al.*, "Mining Semantic Features in Patent Text for Financial Distress Prediction", *Technological Forecasting and Social Change*, vol. 190, no. 1, pp. 122450–122451, 2023. http://dx.doi.org/10.1016/j.techfore.2023.122450

[15] H. Chen and W. Deng, "Interpretable Patent Recommendation with Knowledge Graph and Deep Learning", *Scientific Reports*, vol. 13, no. 1, pp. 2586–2587, 2023. http://dx.doi.org/10.1038/s41598-023-28766-y

[16] J. Choi *et al.*, "Exploring a Technology Ecology for Technology Opportunity Discovery: A Link Prediction Approach Using Heterogeneous Knowledge Graphs", *Technological Forecasting and Social Change*, vol. 186, no. 6, pp. 122161–122162, 2023. http://dx.doi.org/10.1016/j.techfore.2022.122161

[17] M. Liu *et al.*, "Comparison of Administrative and Regulatory Green Technologies Development Between China and the US Based on Patent Analysis", *Data Science and Management*, vol. 6, no. 1, pp. 34–45, 2023. https://doi.org/10.1016/j.dsm.2023.01.001

[18] V. J. Schmitt *et al.*, "Assessment of Patentability by Means of Semantic Patent Analysis–A Mathematical-Logical Approach", *World Patent Information*, vol. 73, no. 6, pp. 102182–102183, 2023. http://dx.doi.org/10.1016/j.wpi.2023.102182

[19] P. Shao *et al.*, "An Integrated Method to Acquire Technological Evolution Potential to Stimulate Innovative Product Design", *Mathematics*, vol. 11, no. 3, pp. 619–620, 2023. http://dx.doi.org/10.3390/math11030619

*Received:* July 2024
*Revised:* August 2024
*Accepted:* September 2024

*Contact address*:
Xuanrong Zhao
Sichuan Top IT Vocational Institute
Sichuan
China
e-mail: zhaoxuanrong@scetop.cn

Xuanrong Zhao is a lecturer with a BSc degree in computer science and technology, and an M.A. degree in MBA management from the University of North Bangkok, Thailand. Her research activities include computer applications, software development, frontend and backend website development, data mining and information processing.