International Journal of Science and Engineering Applications
Volume 12-Issue 02, 89 – 91, 2023, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1202.1032

Intellectual Property Information Intelligent System Based on Content Data Mining and ID4 Algorithm

Qiong Jiang
School of Law
Guilin University of Electronic Technology
Guilin, Guangxi,541004, China

Abstract:This paper studies the intellectual property information system based on content data mining and ID4 algorithm. First, data mining technology is applied to patent information analysis, such as using clustering algorithm to mine patent text, and using association rules to mine patent inventors. The system can analyze the structure of DOCDB patent data files, extract relevant patent information, and store the processed data in the database. The experimental results show that the system can efficiently process patent data and effectively improve the automation level of patent preprocessing. Use support vector machines, naive Bayes, and radial basis neural networks to classify and test patent samples.

Keywords: Intellectual Property, Information Intelligent System, Content Data Mining, ID4 Algorithm

1. INTRODUCTION

In recent decades, the amount of patent information has increased dramatically, and its multiplication cycle has been shortening. At present, there are more than 50 million patent documents [1] in the world, and the total number of patent documents published by various countries exceeds 1.5 million each year [2]. Patent information has also become an inexhaustible treasure trove of technical literature and knowledge. As we all know, patent information is the crystallization of human wisdom, records the achievements and trajectories of human society's inventions and creations [3], contains the most important It has strong fault tolerance, divides large files into many small files, and automatically copies and saves each small file as a copy. The user can customize the value, and the default is. [4], and is the most comprehensive and latest technology in the world. Source of intelligence. In today's technological revolution, the original text analysis method and simple statistical analysis method in the patent information analysis method have also become the past [5]. It is replaced by advanced computer technology. In-depth analysis of hidden laws in patent data provides reliable decision-making basis and intelligence guarantee for technological innovation and enterprise development [6].

In this case, a patent analysis technology based on data mining came into being. Data mining is a multi-disciplinary field that integrates artificial intelligence, machine learning, statistics, knowledge engineering [7], database technology, information retrieval and other new technological research results. Its application range is very wide. In addition to the above external environmental factors, my country's internal environment also prompts [8] companies to This method enhances the fault tolerance of the system and ensures the integrity of the data. Users can also access the target data nearby, reducing the data access delay to a certain extent. of economic structure and building an innovative country [9]. It has been vigorously promoting the progress of intellectual property work from the level of national policies and systems [10]. For enterprises, it is necessary not only to protect their own intellectual property rights, but also to master key technologies and develop products with independent intellectual property rights. For governments at all levels, it is not only necessary to provide enterprises [12] with a good innovation environment, but also to guide enterprises to conduct independent research and development. path of. Based on the above comprehensive factors of the domestic and foreign environment [13], the importance of intellectual property rights is self-evident, and the analysis and research of patent information has become an important aspect of the implementation of intellectual property rights [14]. In recent years, the number of domestic patent applications has increased year by year, and the demand for patent research has also continued to be strong. In order to meet the needs of enterprises and governments for patent information analysis [15], many domestic research institutes are engaged in research in this area, and have achieved some theoretical results [16]. At the same time, many domestic software companies have also launched their own patent analysis software to analyze actual cases for users. Provide tool support [17]. Data mining is a multi-field and multi-disciplinary interdisciplinary, and its development is affected by multiple disciplines [18].

These include database systems, machine learning, statistical technology, and information science, and even biology, neural networks, etc. In addition, it is also affected by the data mining method used [19]. Due to the multi-field and multi-intersection of data mining, data mining technology can use related technologies of other disciplines, such as knowledge representation [20], neural networks, high-performance computing, inductive logic programming, and fuzzy or rough set theory. In addition, The number of physical machines will also have different effects on the amount of computation. With the increase of physical machines, cloud nodes and edge nodes have different costs, as shown in Figure 5-15. recognition technology, spatial data analysis technology [21], information collection and retrieval technology, pattern recognition analysis technology, image processing technology, signal analysis technology, visualization [22] technology, and technology. Bioinformatics technology and other fields. Traditional patent analysis methods mainly use original text analysis methods and simple data statistics. Faced with a large amount of patent document data, not only the workload is heavy, but the application of patent documents only stays on the surface [23].

International Journal of Science and Engineering Applications
Volume 12-Issue 02, 89 – 91, 2023, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1202.1032

With the development of computer technology, the ability of computers to process massive amounts of data has become stronger and stronger, and their applications in information processing have become more and more extensive. At the 1st International Joint Artificial Intelligence Academic Conference [24] held in January 2009, the term data mining, also known as knowledge discovery, was proposed for the first time, and its research focus is slowly changing from the research of discovery methods to the research of system application technology. Moreover, in the recent development, more and more attention has been paid to the combined application of multiple discovery methods and technologies, and the trend of mutual penetration between multiple disciplines has become more and more obvious.

2. THE PROPOSED METHODOLOGY

2.1 The Content Data Mining

Data mining technology appeared in the late 1980s, mainly for business applications. After more than 20 years of development, the research focus has gradually shifted from discovery methods to system applications, focusing on the integration of multiple technologies and the interpenetration of multiple disciplines to tap the intelligence value of information. This feature makes it have a wide range of application prospects in deep-level patent information analysis, but because of its short research time, there is no mature theory at present. The preprocessing process of patent text information is basically the same as the preprocessing process of the text collection in the general Chinese text mining process. It has to go through the five steps of data cleaning, Chinese automatic word segmentation, feature item extraction, feature item weight calculation, and vector space model representation.

Since the cleaning of patent text data is mainly based on the user's analysis topic, the patent text information retrieved from the patent data source is filtered, and the patent data that is not related to the analysis topic is removed. This process is generally manual operation and involves a lot of subjectivity. Factors and specific circumstances. Here, a vector space model is used to represent the patent text after data cleaning. To represent the text as a vector, the text must be segmented first, and then the feature items that can represent the text content are extracted from the segmentation result, and finally a certain method is used to the text feature items are weighted so that a text is expressed as a vector. The following will introduce these five steps. These five steps are the core steps for the content mining of patent information, and they are also steps that require automatic computer processing. The processing effect will directly affect the accuracy of the patent information content mining results.

The main function of data mining refers to the process of using data mining related technologies to find specific valuable data patterns. Generally speaking, data mining tasks can be divided into description tasks and prediction tasks. Descriptive tasks find general characteristics of data in existing databases. The predictive task is to infer, discover and predict the development trend of the data based on the current data analysis. Association rule mining can find out the association or correlation between itemsets and itemsets in the original data set. With the collection and storage of more and more data, more and more researchers related to data mining technology are showing interest in discovering the correlation between data sets from existing databases. The main reason for the low cost of edge computing is that the multi-container technology in the software has more image files, so even if the number of physical machines increases, the cost will not change or even decrease.

2.2 The ID4 Algorithm

There is no doubt that time information is an essential component of video signals when performing various video processing. Time information also plays an irreplaceable role in the perception of external things by the HVS system. Therefore, this paper proposes a method for solving the JND value in the temporal domain based on video motion, by solving the difference of the internal frames of the reconstructed video sequence as the characteristic parameter in the temporal domain of the video signal.

The original LUCENE was developed with the java scripting language as the development language. Due to the wide application of the .NET platform, a ported version of LUCENE, LUCENE.NET, came into being. It is not a complete full-text search engine, but the architecture of a full-text search engine, providing a complete query engine and indexing engine. Developers can implement full-text search functions based on LUCENE.NET.

In addition, achievements and methods in other fields are also introduced, such as expert systems [8], artificial intelligence, web data mining and other advanced technologies, which have improved the efficiency of the question answering system and expanded the research direction of the question answering system. View the interactive messages in the gateway through the Docker in the gateway and the energy management platform in the cloud.

2.3 The Intellectual Property Information Intelligent System

The so-called patent information analysis is collecting patent information from patent documents, processing, sorting and analyzing the patent information through scientific methods, and finally forming a collection of scientific labor of patent information and strategies. The essence of patent information analysis is to conduct directional selection and scientific abstract research on the text content of patent information, patent citations, patent quantity, etc., to study their interrelationships, and to dig out the truth hidden in them, so as to make specific technologies. Trend forecast, follow-up research on competitors, etc. Automatic creation of DOCDB patent database. The database structure for recording DOCDB patent data is very large, with a total of 294 fields. Manually creating this database will be very cumbersome. Therefore, it is necessary to implement flexible automatic creation of database tables.

DOCDB patent data analysis and import. Analyze the input patent data in XML format, extract valid information from the data, and perform data preprocessing on it. ③Types of DOCDB patents are screened and entered. The user can specify the patent data to be stored in the database according to the type code of the patent (Kind-Code). ④Import data in batches. The system can not only process a single patent document, but also process a batch of data in a centralized manner. ⑤Data storage. Store the preprocessed data information into the existing database. In the text clustering method, because neural networks have the advantages of high tolerance to noisy data and low error rate, the application in data mining classification is getting more and more attention. Especially in text clustering, most use SOM (Self-Organizing Feature Map) neural network clustering algorithm, that is, self-organizing feature map neural network, which is also called Kohonen network. The network is a self-organizing,

self-learning, and clustering neural network composed of fully connected neuron arrays.

3. CONCLUSION

On the basis of drawing lessons from data mining technology and ideas, this article carried out related work on the important preprocessing links. Using patent data in the European Patent Office document management database as the data source, the content and structural characteristics were analyzed and compared with the related database structure was designed, and the preprocessing method of this kind of patent data was proposed. Finally, the patent data preprocessing system of the European Patent Office document management database was designed and realized. Through experimental verification, the system can effectively process this type of patent data, resulting in a unified, easy-to-analyze historical database of Germany and Japan.

4. REFERENCES

[1]Jiang Mingliang, Cao Liang, Zheng Jianhua, et al. Patent information platform design based on data mining technology[J]. Computer Knowledge and Technology, 2018(2):138-140.

[2] Jiang Mingliang, Cao Liang, Zheng Jianhua, et al. Patent information platform design based on data mining technology [J]. Computer Knowledge and Technology, 2018, 014(023): 138-140.

[3] Kang Na, Yu Qi, Li Lin, et al. Research on the Library Smart Service System Based on Data Mining[J]. 2021(2019-2):1-3.

[4] Zheng Huizhong, Guo Jihua, Wang Ruimin. Research on the Strategy of Smart Library Construction Based on Big Data Mining and Knowledge Discovery[J]. Information Recording Materials, 2021, 22(2): 2.

[5] Yang Juan. The systematic setting of text and data mining fair use exception norms[J]. Library Forum, 2020, 40(4):10.

[6] Ma Li, Zhou Xiaohu. Research on the Evaluation Mechanism of Innovation and Entrepreneurship Team Management Based on Data Mining Classification[J]. Modern Electronic Technology, 2019, 42(11): 4.

[7] Miao Lei. Application of Web Data Mining Technology in Information Management [J]. Heilongjiang Science, 2021, 12(18): 2.

[8] Xia Chuanxin. Research on the construction of a smart people's livelihood network system based on data mining technology [D]. University of International Business and Economics, 2018.

[9] Xue Jian. Research on the Information Management of University Books Based on Data Mining Technology[J]. Information Technology, 2019.

[10] Gan Ying, Li Yinsuo, Gao Jing. Design and Implementation of Data Mining Function of Cultural Industry Based on Internet [J]. Digital User, 2018.

[11] Huang Xinwen. An integrated forest fire prevention information system based on big data mining:, CN108346252A[P]. 2018.

[12] Tang Hanqin, Shen Lei. Construction of a benefit evaluation model of university education informationization based on cloud model data mining algorithm[J]. Modern Electronic Technology, 2020, 43(13): 4.

[13] Zhao Yunhua, Zhang Jing. Design and implementation of patent data preprocessing system based on data mining[J]. 2021(2011-12):1851-1855.

[14] Liao Shubing, Xing Li. Building an information system for college sports management based on data mining [J]. 2021(2011-11): 37-40.

[15] Sun Hanxiao. Design of user comment data mining system based on Bayesian classification algorithm [J]. Modern Computer, 2021(4): 4.

[16] Jiang Runfa. Research on Online Recommendation Algorithm of University Library Based on Data Mining[J]. Information and Computer, 2020, 32(23):3.

[17] Zhang Fenfen, Xu Xinhua. Application of Apriori Algorithm in University Course Data Mining[J]. Information Technology and Informatization, 2020(11): 4.

[18] Yi Ruiqiang. Coal mine intelligent precision mining system based on data mining technology[J]. Shaanxi Coal, 2020, 39(4):7.

[19] Zhang Yihua. Personalized Service of University Library Based on Big Data Mining and Decision Analysis System[J]. Electronic Technology and Software Engineering, 2018(7):1.

[20] Chen Chen. The construction of a smart library based on big data mining and knowledge discovery[J]. 2021(2017-8):85-91.

[21] Li Xiaofeng, Chen Hao. Research on Intelligent Analysis Strategy of Logistics Information Based on Dynamic Data Mining[J]. Computer Fan, 2018(4):1.

[22] Fan Hengyu. Research and practice of micro-credit method based on data mining technology[J]. Information and Computer, 2021, 33(4):3.

[23] Li Juntao. Research on Hybrid Intrusion Detection Based on Data Mining[J]. Guangdong Communication Technology, 2021, 41(7): 4.

[24] Tang Peipei, Wu Minghui. Exploration and practical research on "data mining technology" course construction based on Kaggle competition data [J]. Industry and Information Education, 2021.