Research on Efficient Retrieval Method of Key Information of Discipline Inspection and Supervision Based on Digital Technology

Tiancheng Chen[1], Yilin Cai[2*]

124417877@qq.com[1], 1027144905@qq.com[2*]

Shenzhen Power Supply Bureau Co.,Ltd., Shenzhen, Guangdong, 518000, China

Abstract. Influenced by the characteristics of discipline inspection and supervision information itself, it is difficult to guarantee the correlation between the corresponding retrieval results and the retrieval objectives. Therefore, this paper puts forward the research on the efficient retrieval method of key information of discipline inspection and supervision based on digital technology. After comprehensively analyzing its specific attributes from four aspects: discipline inspection and supervision workflow information, discipline inspection and supervision object type information, discipline inspection and supervision content information and discipline inspection and supervision result information, the LightLDA technology in digital technology is introduced in the specific retrieval method design process, and the information in the retrieval likelihood model of discipline inspection and supervision information is fused with LDA by Wei and Croft in the LightLDA technology, and the information with the greatest likelihood to the retrieval information is taken as the retrieval output result. In the test results, the nDCG value is always stable above 40.0%,recall rates are consistently above 90%, which has obvious advantages compared with the control group.

Keywords: Attribute characteristics;Digital technology; Efficient retrieval;Information fusion;Key information of discipline inspection and supervision; LightLDA technology; Likelihood model;

1.   Introduction

At this stage, the development of discipline inspection and supervision information retrieval has achieved certain results. First of all, discipline inspection and supervision organs at all levels have established a relatively complete discipline inspection and supervision information system, which makes information retrieval more convenient and efficient. Secondly, with the continuous progress of information technology, the accuracy and comprehensiveness of information retrieval have been greatly improved [1-2]. In addition, relevant laws and regulations have also been improved, providing a clearer legal basis for information retrieval. However, despite this, there is still some room for optimization in the practical application of discipline inspection and supervision information retrieval, mainly in the aspect that the intelligence of information retrieval needs to be further improved. The existing discipline inspection and supervision information system still has some manual intervention in the process of information retrieval, and has not fully realized intelligent retrieval. This may lead to low retrieval efficiency, and even the deviation of retrieval results [3-4]. Therefore, the

future development should pay more attention to improving the intelligence of information retrieval and introduce more advanced artificial intelligence technologies, such as natural language processing and machine learning, in order to improve the accuracy and efficiency of retrieval. Secondly, information sharing and integration still need to be further strengthened. Although the discipline inspection and supervision organs have established a relatively perfect information system, the degree of information sharing and integration among different regions and departments is still insufficient [5-7] . In the future, we should further promote information sharing and integration between different systems and departments, break the phenomenon of information islands and improve the comprehensiveness and integrity of information. In addition, the particularity of discipline inspection and supervision determines that the security of its information system is very important. However, the existing information systems still have certain risks [8-10] in information security, such as data leakage and network attacks. Therefore, in the future, we should pay more attention to improving the security of information systems and adopt more advanced security technologies, such as data encryption and authentication, to ensure the security and stability of information [11-12].

Combined with the above analysis, this paper puts forward the research on efficient retrieval method of key information of discipline inspection and supervision based on digital technology, and analyzes and verifies the application value of the designed retrieval method through experimental testing. With the help of the research and design of this paper, I hope it can bring reference to the further development of discipline inspection and supervision information retrieval.

2. Design of efficient retrieval method for key information of discipline inspection and supervision

2.1 Discipline Inspection and Supervision Information Attribute Analysis

For discipline inspection and supervision information, its content is comprehensive. Because the discipline inspection and supervision work involves all aspects of enterprise operation, including employee management, business process, assessment and evaluation, etc. the content of discipline inspection and supervision information also covers these aspects, which is one of the main reasons for its extensiveness and comprehensiveness. Secondly, the discipline inspection and supervision information is obviously dynamic. With the development of enterprises and the change of operation, the content of discipline inspection and supervision information will also change accordingly. In view of this, in order to retrieve the key information of discipline inspection and supervision more accurately, this paper first comprehensively analyzes its specific attribute characteristics from four aspects: discipline inspection and supervision work flow information, discipline inspection and supervision object type information, discipline inspection and supervision content information and discipline inspection and supervision result information.

Table 1. Information Attributes of Discipline Inspection and Supervision.

Information category	Category	Content	Attribute characteristics
Disciplinary inspection and supervision workflow information	Reflects the overall planning and implementation process of disciplinary inspection and supervision work.	Procedures for handling cases, investigating cases, meeting erroneous facts, writing investigation reports, and handling cases.	The specific content of each link, including the person in charge, time, location, method, steps, approval procedures, etc.
Disciplinary inspection supervision object type information	The type of supervised object involved.	Office organization, staff, and monitoring objects, as well as the characteristics and behavioral manifestations of these objects.	The identity, position, job nature, scope of responsibilities, social relationships, etc. of the target person.
Disciplinary inspection and supervision content information	Specific content related to disciplinary inspection and supervision work	Supervision of corporate culture and clean governance construction, anti-corruption struggle, supervision and inspection, etc.	Specific work scope, work content, work focus, methods and measures, etc.
Information on disciplinary inspection and supervision results	Reflecting the effectiveness and results of disciplinary inspection and supervision work	Discovering problems, investigating evidence, handling results, etc.	Classification of discovered problems, areas and stages of occurrence, severity and scope of impact of problems, handling methods and results, etc.

Combined with the analysis results of the attribute characteristics of discipline inspection and supervision information in Table 1, they together constitute a complete system of discipline inspection and supervision information. Based on the reasonable integration and utilization of all kinds of discipline inspection and supervision information characteristics, this paper carries out research on improving the quality and efficiency of discipline inspection and supervision retrieval.

2.2 Discipline Inspection and Supervision Information Attribute Analysis

Combined with the above analysis results of the attribute characteristics of discipline inspection and supervision information, this paper introduces the LightLDA technology in the digital technology in the specific retrieval method design process.

Firstly, the likelihood model of information retrieval for discipline inspection and supervision is constructed, which can be expressed as equation (1)

$$P(Q|D) = \prod P(w|D) \tag{1}$$

Among them, $P(Q|D)$ represents the probability of generating a query by the document model under the assumption of the bag-of-words model, Q represents retrieval information, D represents all discipline inspection and supervision information, $P(w|D)$ represents the maximum likelihood parameter and w represents the term in the retrieved information Q.

The formula for $P(w|D)$ is

$$P(w|D) = 1 - \frac{N}{N + \mu} \qquad (2)$$

Where, N represents the total number of terms in discipline inspection supervision information, and μ represents prior parameters.

According to the principle of the above formula, Wei and Croft in LightLDA technology are used to calculate the merged discipline inspection supervision information as

$$P(w|D) = 1 - \frac{N}{N + \mu} \qquad (3)$$

Where, $P_{LLBDM}(w|D)$ represents the discipline inspection and supervision information after fusion, λ represents the linear combination coefficient, and $P_{ML}(w|D)$ represents the discipline inspection and supervision information in the subject form obtained by LightLDA training.

On this basis, the specific retrieval method can be expressed as follows

$$X = \max \sum P_{LLBDM}(w|D) \rightarrow x \qquad (4)$$

Among them, X represents the retrieved discipline inspection and supervision information, x represents the retrieval request information.

According to the way shown above, the key information of discipline inspection and supervision can be retrieved efficiently, and the quality and effect of retrieval can be guaranteed to the maximum extent.

3. Application testing

3.1 Test data preparation

When analyzing the performance of the efficient retrieval method of key information of discipline inspection and supervision designed in this paper, an enterprise with a certain scale and business complexity is selected as the test object, which has many departments, business segments and employees and has certain discipline inspection and supervision needs. The composition of discipline inspection and supervision information participating in the test is analyzed, which is mainly divided into four parts as shown in Table 2.

Table 2. Composition of discipline inspection and supervision information tested.

Number	Information type	Information composition	Information Attributes
1	employee information	Including employee's personal information, work history, family background, social relationships, etc.	It can help disciplinary inspection and supervision institutions understand the ideological status, conduct, and performance of employees in their work.

2	Business Process Information	Including the business processes, rules and regulations, work standards, etc. of the enterprise.	Discipline inspection and supervision institutions can supervise and evaluate the compliance, rationality, and efficiency of enterprise operations.
3	Assessment and evaluation information	Including assessment indicators, results, rewards and punishments for employees and departments.	Can understand the management status of the enterprise, the work performance of employees, and the performance level of the department.
4	Event processing information	Including issues, doubts, violations discovered in disciplinary inspection and supervision work, as well as corresponding handling results.	They can help disciplinary inspection and supervision agencies timely identify and correct problems in enterprise operations.

On this basis, in the process of testing, the specific information scale is set as follows: the total data volume of employee information is 20,000, the total data volume of business process information is 4,000 (the total number of business processes is 22, and each business process contains 10 steps), the total data volume is 25,000 (the assessment index of each employee is 12), and the total data volume of event handling information is 1,000 (the total number of historical events is 524).

Through the number and scale of test information set above, the performance of the information retrieval method of discipline inspection and supervision designed in this paper is tested. In order to ensure the objectivity of the final test results, the integrated information retrieval method based on random forest and the cross-modal information retrieval method based on collaborative matrix decomposition are set as the test control group respectively.

3.2 Test results and analysis

When analyzing the performance of different retrieval methods, the normalized cumulative loss gain of NDCG (Normalized Discontinued Cumulative Gain) is set as a specific evaluation index. As the normalized processing result of the discounted cumulative gain, DCG) calculated after the cumulative gain, CG) is depreciated, it can more accurately feed back the correlation between the output result of the retrieval method and the retrieval information, and the specific calculation method can be expressed as follows

$$DGG@k = \sum \frac{r(i)}{\log_2(i+1)} \tag{5}$$

$$nDGG@k = \frac{DGG@k}{IDGG_p} \tag{6}$$

Among them, $DGG@k$ represents the cumulative gain parameter of the retrieval output result TOP-k, $nDGG@k$ represents the normalized parameter of the cumulative gain of the retrieval output result TOP-k, $r(i)$ represents the cumulative gain of a single retrieval output result, n represents the cumulative quantity, $IDGG_p$ represents a normalized parameter.

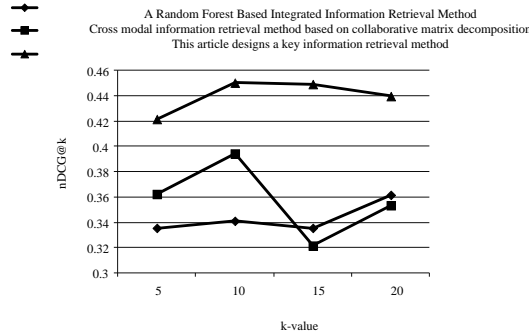On this basis, the test results of different methods are shown in Figure 1.

Fig. 1. Comparison chart of test results of different methods.

Combined with the test results shown in Figure 1, it can be seen that among the test results of three different retrieval methods, the nDCG value of the integrated information retrieval method based on random forest is relatively stable, and the overall development trend is increasing with the increase of the scale of the retrieval output results, and the corresponding maximum and minimum values are 36.17% and 33.56% respectively. Under the cross-modal information retrieval method based on collaborative matrix decomposition, the value of nDCG fluctuates obviously, and there is no linear correlation between this fluctuation and the scale of retrieval output results. The maximum value of nDCG reaches 39.44% (when the output results are 10), and the minimum value of nDCG is only 32.16% (when the output results are 15), and the corresponding fluctuation amplitude reaches 7.28%. In contrast, in the test results of the retrieval method designed in this paper, the nDCG value is always stable above 40.0%, which has obvious advantages compared with the control group.

Based on this, recall ratio W is adopted to measure the key information retrieval performance of different methods, and the calculation formula is as follows

$$W = \frac{\text{The amount of information retrieved}}{\text{Total amount of information}} \times 100\% \qquad (7)$$

The information retrieval performance of different methods is shown in Figure 2.
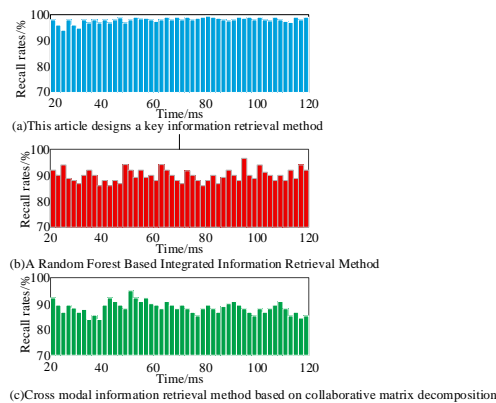

Fig. 2. Comparison of recall rates of different retrieval algorithms.

As can be seen from Figure 2, the recall rate of the proposed method is always higher than 90%. It can be seen that the proposed algorithm can effectively improve the recall rate and obtain more satisfactory and efficient retrieval results of key information for discipline inspection and supervision.

4.   Conclusion

This paper puts forward the research on the efficient retrieval method of key information of discipline inspection and supervision based on digital technology, which effectively guarantees the high correlation between the retrieval output of key information of discipline inspection and supervision and the target retrieval information. With the help of this research, I hope to provide guarantee for the discipline inspection and supervision staff to obtain the required information faster, so as to speed up the process of finding, analyzing and solving problems and improve work efficiency. Help improve the operational efficiency and management level of enterprises, and promote the innovative development of discipline inspection and supervision.

References

[1]   Deepa, N., & Pandiaraja, P. (2021). E health care data privacy preserving efficient file retrieval from the cloud service provider using attribute based file encryption. Journal of Ambient Intelligence and Humanized Computing, 12, 4877-4887.

[2]   Wang, T., Yang, Q., Shen, X., Gadekallu, T. R., Wang, W., & Dev, K. (2021). A privacy-enhanced retrieval technology for the cloud-assisted internet of things. IEEE transactions on industrial informatics, 18(7), 4981-4989.

[3]   Guo, J., Cai, Y., Fan, Y., Sun, F., Zhang, R., & Cheng, X. (2022). Semantic models for the first-stage retrieval: A comprehensive review. ACM Transactions on Information Systems (TOIS), 40(4), 1-42.

[4]   Guha, A., Samanta, D., & Islam, S. H. (2023). Iirm: Intelligent information retrieval model for structured documents by one-shot training using computer vision. Arabian Journal for Science and Engineering, 48(2), 1285-1301.

[5]   Messina, N., Amato, G., Esuli, A., Falchi, F., Gennaro, C., & Marchand-Maillet, S. (2021). Fine-grained visual textual alignment for cross-modal retrieval using transformer encoders. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 17(4), 1-23.

[6]   Zamani, H., Trippas, J. R., Dalton, J., & Radlinski, F. (2023). Conversational information seeking. Foundations and Trends® in Information Retrieval, 17(3-4), 244-456.

[7]   Pouamoun, A. N., & Kocabaş, İ. (2023). Multi-agent-based hybrid peer-to-peer system for distributed information retrieval. Journal of Information Science, 49(2), 529-543.

[8]   Gong, Y., & Cosma, G. (2023). Improving visual-semantic embeddings by learning semantically-enhanced hard negatives for cross-modal information retrieval. Pattern Recognition, 137, 109272.

[9]   Nayak, K. V., Arunalatha, J. S., Vasanthakumar, G. U., & Venugopal, K. R. (2022). Design of Deep Convolution feature extraction for multimedia information retrieval. International Journal of Intelligent Unmanned Systems, (ahead-of-print).

[10] Das, S., & Paik, J. H. (2023). Gender tagging of named entities using retrieval-assisted multi-context aggregation: An unsupervised approach. Journal of the Association for Information Science and Technology, 74(4), 461-475.

[11] Xiao, K., Li, D., Guo, P., Wang, X., & Chen, Y. (2023). Similarity matching method of power distribution system operating data based on neural information retrieval. Global Energy Interconnection, 6(1), 15-25.

[12] Zhang, F., Li, X., Zhou, X., Chan, N. W., Tan, M. L., Kung, H. T., & Shi, J. (2023). Retrieval of soil salinity based on multi-source remote sensing data and differential transformation technology. International Journal of Remote Sensing, 44(4), 1348-1368.