

Information Retrieval Method of Natural Resources Data based on Hash Algorithm

Qian Li*

College of Computer and Information Engineering,
Guangxi Vocational Normal University, Nanning 530007, China

Abstract—In order to improve the ability of searching and identifying information in natural resources data, this paper puts forward a method of searching information in natural resources data based on hash algorithm. Through the data center technology, the problems of information source positioning, data directory organization, data semantic definition and expression, and data entity relationship construction in natural resource data center are solved. Combined with the distribution of resource data stream, the information structure reorganization and data encryption in natural resource data center are realized by using hash algorithm, and the parameters of information quality control model in natural resource data center are established. Through natural resource data governance and semantic reconstruction, the characteristics detection and redundancy arrangement of information data in natural resource data center are realized by standardizing data collection rules, and they are stored in the intermediate database. Through data governance rules, the information in the natural resources data is structured and managed, and stored in the publishing library. Through all kinds of data processing tools, all kinds of data are processed, cleaned and reconstructed, and through Hash algorithm and data aggregation processing, information detection in natural resources data is realized. The simulation results show that the precision rate of natural resource data retrieval by this method is high.

Keywords—Hash algorithm; natural resource data; information structure reorganization; search; data encryption

I. INTRODUCTION

Due to the fact that natural resource data has not yet formed a complete and clear hierarchical directory and corresponding data items, and has not been marked according to business scenarios. This affects the business department's grasp of the actual availability of data results, which can easily lead to data duplication and low reuse rates of results. Comprehensive data mining analysis is not sufficient to effectively meet the needs of decision support [1-4]. The application development of the upper layer is "strongly coupled" to the data platform of the lower layer. The data service provision of "standard components" is to customize their own data service platform from top to bottom, cooperating with different upper layer applications, thereby disrupting the normal formation of the basic data directory service system. Any changes in the underlying data environment will directly affect the normal use of other business applications. Such as the entry and exit of data, i.e. the definition of data lineage and data flow. Therefore, the research on automatic retrieval of data resources is of great significance [5].

Currently, there are many studies related to automatic retrieval methods. Study [6] proposes a mental health management model for college students based on wireless sensor networks, which uses wireless sensor systems to complete the statistics of college students' mental health data, thereby achieving automatic retrieval of information. Research [7] proposes a clustering analysis algorithm in the analysis of college students' mental health education, according to which the relevant data of college students' mental health education can be calculated, Provide a foundation for automatic information retrieval.

To solve the problems in the above methods, this paper proposes a natural resource data information retrieval method based on hash algorithm. This article uses data center technology to solve problems such as information source localization, data catalog organization, data semantic definition and expression, and relationship construction between data entities in natural resource data. Combining the distribution of resource data streams, hash algorithms are used to restructure the information structure and encrypt data in the natural resource data center, establish information quality control model parameters for the natural resource data center, organize and manage the data in the natural resource data center through data governance rules, and store them in the release library. Through various data processing tools, various data are processed [8], cleaned up, and reconstructed. Through hash algorithms and data aggregation processing, information detection in natural resource data is achieved. Experimental results show that this method has excellent performance in improving the information retrieval ability of natural resource data.

II. INFORMATION STORAGE STRUCTURE MODEL OF NATURAL RESOURCE DATA CENTER AND ANALYSIS OF BIG DATA CHARACTERISTICS

A. Information Storage Structure Model of Natural Resources Data Center

In order to realize the system design of information retrieval in natural resource data based on hash algorithm, a block-by-block control model of information retrieval in natural resource data is constructed by combining bidirectional reference control and fuzzy retrieval, and the information processing terminal of information retrieval in natural resource data is established by combining semantic similarity fusion and database natural language fast retrieval method [9-11]. The semantic detection and data comprehensive management of information retrieval in natural resource data are realized by

adopting expert system identification method, and the overall structure of information retrieval in natural resource data is obtained (see Fig. 1)

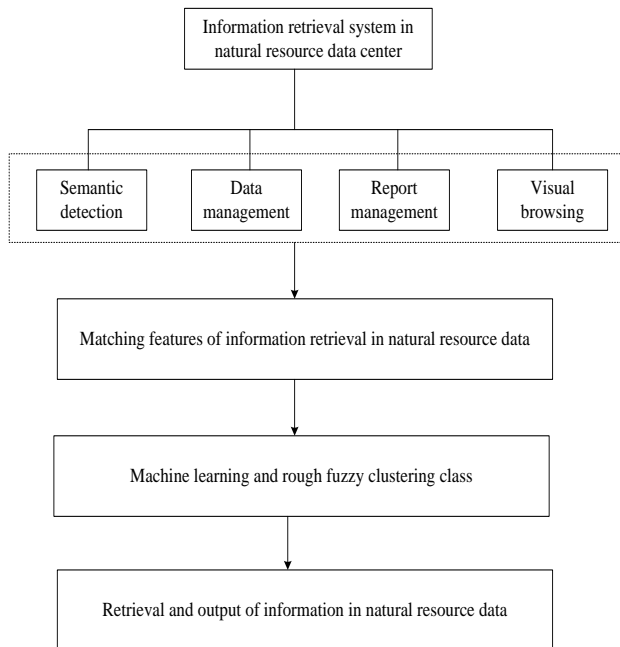


Fig. 1. The overall structure of information retrieval in natural resources data.

The overall structure of information retrieval in natural resource data mainly includes semantic detection module, data management module, report management module and visual management module. Using the embedded B/S framework method, the program control of the information retrieval system in natural resources data is carried out, and the optimized storage structure model of information retrieval in natural resources data is constructed by combining the dictionary ordering storage mechanism [12]. The distribution of storage nodes of information in natural resources data is shown in Fig. 2.

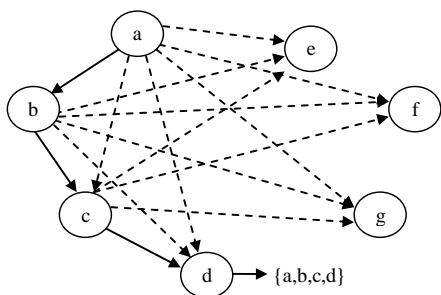


Fig. 2. Storage nodes of station information in natural resources data.

According to the storage node model of information in natural resources data in Fig. 2, based on data storage and distributed clustering, the information retrieval in natural resources data is divided into five layers. They are:

1) *Data source layer.* All kinds of heterogeneous databases are compatible and used as data sources. In the process of data

warehousing, they are divided into intermediate database and publishing database. Organize the source data by standardizing the data collection rules and store it in the intermediate database; through data governance rules, the data is structured and managed, and stored in the publishing library. This system provides complete and effective data support for the whole data center.

2) *Data resource layer.* Through all kinds of data processing tools, all kinds of data are processed, cleaned, reconstructed, etc., and finally the data formats and data structures required by data service, application and tools are formed.

3) *Platform service layer.* Provide unified data scheduling service, support all kinds of services, applications and tools, and manage all information of service layer in the form of service bus.

4) *Application resource layer.* Provide various data-related applications, such as data query, visual analysis, label management, knowledge map, knowledge search, etc. At the same time, according to the relevant service specifications, the secondary development interface is provided to lay the foundation for future expansion.

5) *Portal layer.* According to the user's rights and responsibilities, it provides quick and humanized operation mode, and provides corresponding data management tools and applications for people with different rights. The management end also provides simple and effective management mode correspondingly [13].

According to the above-mentioned rule set distribution and the optimized design of the retrieval system, the structural model of the information retrieval system in natural resources data is constructed [14].

B. Data Fusion of Information Retrieval in Natural Resources Data

The dictionary storage mechanism is used to construct the optimized distribution structure model of the information in natural resources data [15]. Combined with the analysis of the storage structure of the information in natural resources data, the distribution order of the information in natural resources data is $u_i \in R^m$ through fuzzy matching and hashing algorithm. Similarly, in the conversation component, the conversation protocol of the information retrieval in natural resources data is constructed, and the distribution set of related features of the information retrieval in natural resources data is obtained under the guidance of the retrieval mode as follows:

$$M^x = \{m_{i_1, i_2, \dots, i_{n+1}} = \frac{\theta_{i_1, i_2, \dots, i_{n+1}}(x)}{L - n}, i_k \in B\} \quad (1)$$

Wherein, $m_{i_1, i_2, \dots, i_{n+1}}$ is the dimension of fuzzy comprehensive clustering, L is the sample length of information retrieval in natural resources data, and n is the sampling sample sequence. The matching model of information retrieval in natural resources data is constructed by using Observer coprocessor, and the element combination

parameters of information retrieval in natural resources data are obtained by semantic degree analysis in file x_t , when:

$$P(x_t | x_{t-1}, x_{t-2}, \dots, x_1) = P(x_t | x_{t-1}, x_{t-2}, \dots, x_{t-n}) \quad (2)$$

Where, x_t is the length of hash algorithm and x_{t-n} is the regression distribution parameter. In the storage node S set of station information in natural resource data, the edge feature distribution set is satisfied. The semantic correlation dimension $x_t \in B$ of natural resource data is defined. Based on the fusion of large data sets, the feature quantity $\theta_{i_1, i_2, \dots, i_{n+1}}(x)$ of natural resource data association rules about q_i vector combination is obtained. By using the method of identifying natural resource data information, the association of natural resource data information is carried out in the cluster center. Combined with the fuzzy comprehensive decision method, the retrieval control of natural resource data information is carried out, and the association rule items are obtained as $d_i = (d_{i1}, d_{i2}, \dots, d_{in})$. The dense subgraph of natural resource data information is.

$$W(q_i, d_{ij}) = \frac{g(q_i, d_{ij}) \times \log_2[f(d_{ij}) + 1]}{\sum_{j=1}^n \{g(q_i, d_{ij}) \times \log_2[f(d_{ij}) + 1]\}} \quad (3)$$

Wherein, $g(q_i, d_{ij})$ is the nearest neighbor feature distribution set, and $f(d_{ij})$ is the joint autocorrelation matching set of station information in natural resources data, thus realizing the data fusion of station information retrieval in natural resources data.

III. OPTIMIZATION OF INFORMATION RETRIEVAL ALGORITHM IN NATURAL RESOURCE DATA

A. Characteristic Clustering of Information Retrieval in Natural Resources Data

The association rule set of information distribution in natural resources data is constructed, and the control time of information retrieval in natural resources data is S by using hash algorithm. By using multi-table connection and semantic matching, the optimal feature solution set of information in natural resources data is obtained:

$$\xi = 2\rho_{\max} \lambda_{\max}(Q_i) N \delta^2 K T \quad (4)$$

Let ρ_{\max} be the correlation attribute set of information retrieval in natural resource data, λ_{\max} is the fuzzy matching coefficient, Q_i is the correlation rule coefficient, K is the detection statistical feature quantity, and T is the semantic adjacent parameter, and the attribute value of the distribution set a_i in natural resource data is $\{c_1, c_2, \dots, c_k\}$. Through mining the correlation attribute of information retrieval in natural resource data, the table structure attribute retrieval method is adopted, and the adaptive feature matching of information in natural resource data is adopted by cloud computing, so as to construct the node retrieval of information in natural resource data.

$$x_k = f_k(x_{k-1}, u_{k-1}) + w_{k-1} \quad (5)$$

$$z_k = h_k(x_k, u_k) + v_k \quad (6)$$

Wherein, f_k is the spectrum parameter of information distribution in natural resource data, which is the width of semantic sample, u_{k-1} is the clustering center of information in natural resource data, w_{k-1} is the iteration number of hash algorithm, and v_k represents the two-way reference matching component of information retrieval in natural resource data. A clustering model of information in natural resource data is constructed by two-dimensional semantic fusion, which is denoted as $l \in [0, L-1]$, $\tau_0 < \tau_1 < \dots < \tau_{L-1}$. The correlation distribution features of information in natural resources data are α_i and SD, where SD, through data matching degree query, obtains the joint autocorrelation distribution set of information retrieval in natural resources data which satisfies $c_j T_c < T_f$, τ_i indicates that the parameter set of information in natural resources data is, carries out fuzzy clustering on $\vec{s} = (s_1, s_2, \dots, s_n)^T$ unknown information components to obtain the matching set of source knowledge points of information retrieval in natural resources data, and adopts joint linear correlation fusion to obtain the linear hash algorithm model of information retrieval in natural resources data as follows:

$$\vec{x} = \vec{H}\vec{s} = \sum_{j=1}^n \vec{h}_j s_j, j = 1, 2, \dots, n \quad (7)$$

Wherein, $\vec{H} = [h_1, h_2, \dots, h_n]$ is the association rule information of information retrieval in $m \times n$ -order natural resource data, \vec{h}_j is the dimension of second-order hash algorithm, and s_j is the statistical distribution set. The characteristic clustering model of information retrieval in natural resource data is:

$$\begin{bmatrix} x_1(t) \\ \dots \\ x_m(t) \end{bmatrix} = \begin{bmatrix} h_{11} & \dots & h_{1n} \\ \dots & \dots & \dots \\ h_{m1} & \dots & h_{mn} \end{bmatrix} \begin{bmatrix} s_1(t) \\ \dots \\ s_2(t) \end{bmatrix} \quad (8)$$

Wherein, h_{11} and h_{1n} respectively represent the ambiguity from one hash algorithm to n times of learning, h_{m1} and h_{mn} represent the convergence coefficient of the corresponding clustering center, and $s_1(t)$ and $s_2(t)$ represent the joint autocorrelation coefficient of information in natural resources data. According to the above analysis, the clustering analysis of information retrieval in natural resources data is realized by using hash algorithm and rough set clustering.

B. Retrieval and Output of Information in Natural Resources Data

The hash algorithm is adopted to realize iterative fusion and adaptive control in the process of information retrieval in natural resources data, and the fuzzy parameter distribution

domain of information retrieval in natural resources data is constructed as follows:

$$D_{node}(i) = \frac{N_{node}^r(i)}{N_{node}} \quad 1 \leq i \leq N_{node} \quad (9)$$

Wherein, N_{node} is the dimension of information retrieval nodes in natural resources data, and $N_{node}^r(i)$ is the detection statistical component corresponding to information retrieval nodes in natural resources data [16-18].

Combined with the correlation of information retrieval in natural resources data[19-23], the fuzzy decision model of information retrieval in natural resources data is constructed, and the Transport/Session transmission protocol and session management protocol of information retrieval in natural resources data are designed[23-26]. The output of the optimized retrieval model is as follows:

$$\max T$$

$$s.t. \sum_{(i,j) \in E} \hat{g}_{ij}^c - \sum_{j \in V} \sum_{(j,i) \in E} \hat{g}_{ji}^c = T\sigma_i^c, \quad \forall i, c \quad (10)$$

$$\left(\sum_{c \in C} \sum_{(i,j) \in E} e_i^c \hat{g}_{ij}^c + \sum_{c \in C} \sum_{j \in V} \sum_{(j,i) \in E} e_i^c \hat{g}_{ji}^c \right) T \leq E_i, \quad \forall i \quad (11)$$

Wherein, \hat{g}_{ji}^c is the input joint parameter of information retrieval in natural resource data, T is the sampling sample width of information in natural resource data, and σ_i^c is the ambiguity detection coefficient. The ambiguity detection is carried out on the bidirectional index channel of information in natural resource data. According to the above analysis, the association rule set of information distribution in natural resource data is constructed, and the iterative fusion and adaptive control in the process of information retrieval in natural resource data are realized by using hash algorithm.

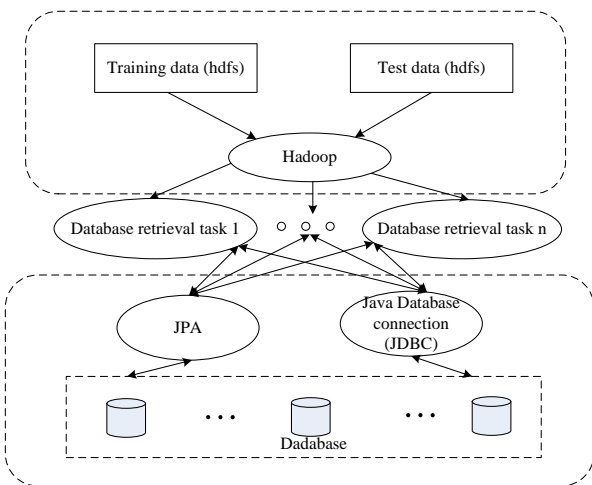


Fig. 3. Implementation process of information retrieval in natural resources data center.

According to Fig. 3, the training data (hdfs) and test data (hdfs) are input to Hadoop Company and transported to the database retrieval task through Hadoop Company 1; the database retrieval task, so as to complete the information retrieval of the natural resource data center.

IV. SIMULATION TEST

A. Experimental Environment Settings

In the test platform, according to the overall structure, the middle platform includes seven modules: catalogue information system, data resource management, electronic license management, interface management, system management, system monitoring and system tools, and supports specific functions such as metadata management, catalogue management, atlas management, automatic generation of electronic licenses, resource statistics and application monitoring. The Natural Resources Department has more than 66.93 million data records, covering 23 business offices; there are about 22 million data records in other departments and bureaus, covering 19 industries, sharing 163.9TB of aggregated spatial data. The distribution set of statistical features of central information retrieval in natural resources data is 1,206, and 29 categories of thematic data and 1,600 element layers are aggregated. See Table I for statistical characteristics of station information distribution in natural resource data.

TABLE I. STATISTICAL CHARACTERISTIC INFORMATION OF STATION INFORMATION DISTRIBUTION IN NATURAL RESOURCES DATA

Middle station information sample	Test set similarity	Sample set regression parameter	Training set ambiguity
Sample1	0.289	0.462	0.479
Sample2	0.266	0.441	0.458
Sample3	0.288	0.482	0.361
Sample4	0.270	0.506	0.451
Sample5	0.203	0.576	0.484
Sample6	0.245	0.414	0.312
Sample7	0.293	0.438	0.383
Sample8	0.265	0.434	0.301
Sample9	0.218	0.585	0.369
Sample10	0.240	0.550	0.364

B. Discussion of the Experimental Methods and the Results

According to the parameter distribution in Table I, search the information of natural resources data, and store it in the form of key values to assist label management and quick resource positioning. At the same time, the index of conditional retrieval label is added to support full-text retrieval. On the basis of the category system of natural resources data labels, perfect data query and guidance functions are established to provide support for data sharing and data trading. Query entities, labels, relationships and portraits that support management, add functions such as data type inference, condition retrieval, paging query, expression query, etc., introduce the intelligent guidance concept, combine hash algorithm, and mark feature points to analyze the reliability of

information retrieval in the middle station. According to the sample test, the sample feature distribution of information in the middle station of natural resources data is shown in Fig. 4.

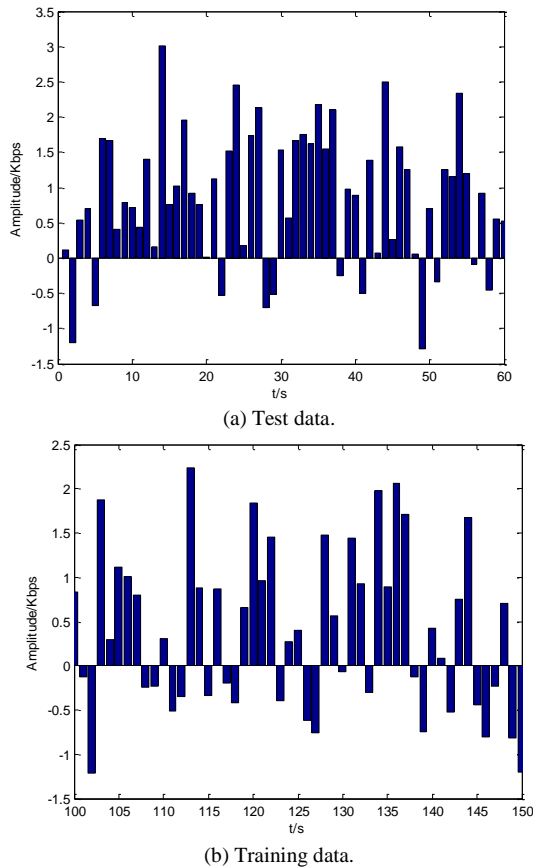


Fig. 4. Distribution of sample characteristics of station information in natural resources.

According to the sample distribution structure of natural resources information in Fig. 4, the information of natural resources data is searched and the confidence level distribution is shown in Fig. 5.

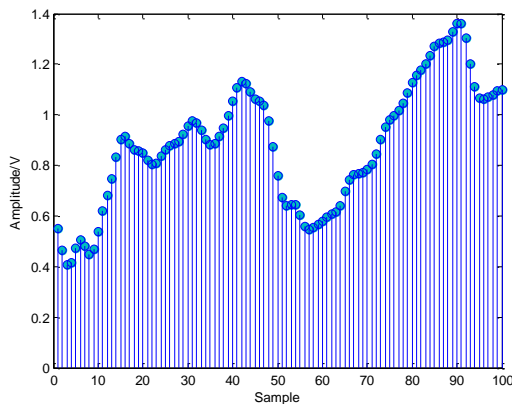


Fig. 5. Confidence level of information retrieval in natural resources.

According to the distribution characteristics of frequency domain, a knowledge map of the relationship among data

clouds, networks, systems, events and data resources is established, and the use of resources is monitored by distribution and exchange. The visual display mode is adopted to display the overall situation of resources, the real-time status of data sharing, and the ranking of resource applications, to make statistics and measurement of data usage, and to carry out early warning and monitoring of important indicators. It is concluded that the clustering of the information retrieval in natural resources data by this method is good, and the recall rate of different methods is tested. The comparison results are shown in Fig. 6, and the analysis of Fig. 6 shows that the recall rate of the information retrieval in natural resources data by this method is high.

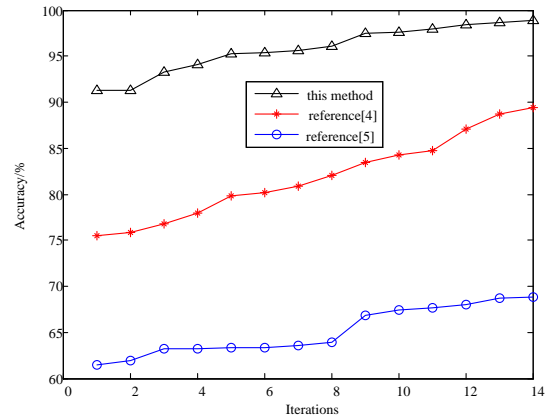


Fig. 6. Comparison of retrieval performance.

From the analysis of Fig. 6, it is known that the method in this paper has a high recall rate and a good retrieval performance. At the same time, establish data inflow and outflow ledger, assist resource managers to master and understand the weak links of individual work, and promote the resource construction of the platform. According to the data security level, the data will be managed hierarchically and desensitized, and the abnormal behavior, data access, operation of data resources and metadata of users in the data center will be monitored, counted and risk analyzed to ensure the security of government data of natural resources.

V. CONCLUSION

Systematic management of natural resources data is a long-term process, which needs to be carried out from methodology, standards and technical realization. The construction of data platform can open up the links of data acquisition, data processing, data service, etc. through the double model of business and data, which can further improve the way of data sharing and exchange, and form an open, flexible and extensible unified natural resources data management mode. At present, there are many factors that restrict the governance of government data of natural resources based on the technology of data center. First, the industry lacks norms and standards for the construction of data center, and the specific form, sharing mode and service standard of data center for the future development and construction of natural resources are not defined. Second, the data return mechanism has not been improved. Limited by network, security, technology, etc., the

data generated and output by the same level, higher and lower levels are not effectively shared. For example, local governments use superior systems to report data, and in the process of reporting, a large number of real, effective and structured data have been sorted out and produced.

However, in actual work, due to the lack of data return mechanism and sharing channels, these reported data have been managed and used by the superior departments. If local departments need to use or connect them to their own systems, there will be a second job. Next, this study will further upgrade and optimize the data middle station on the basis of in-depth exploration of the types, nature, collection cycle and collection mode of business data and continuous improvement of data management and service mechanism. At the same time, on the basis of data resource service convergence, make full use of big data analysis technology to explore the construction of knowledge map of natural resources, find out the relationship between entities, better analyze the problems in natural resource management, and provide practical and valuable reference for administrative decision-making.

REFERENCES

- [1] Z. Zhang, Z. Wang, Z. Cui, "Air Traffic Flow Pattern Recognition and Analysis in Terminal Area Based on the Geodesic Distance," *Mobile Networks and Applications*, 2022, 44(5): 10-15.
- [2] Z. Chen, "A Traffic Flow Forecasting Method Regarding Traffic Network as an Digraph," *International Journal of Pattern Recognition and Artificial Intelligence*, 2021, 35(15): 67-72.
- [3] O. Maria, "Directional assessment of traffic flow extremes," *Transportation Research Part B*, 2021, 150(5): 353-369.
- [4] J. Aniel, "Convection indicator for pre-tactical air traffic flow management using neural networks," *Machine Learning with Applications*, 2021, 45(3): 43-48.
- [5] A. Somya, V. Durgesh, "Impact of vehicular traffic stream on pedestrian crossing behavior at an uncontrolled mid-block section," *Transportation Research Interdisciplinary Perspectives*, 2021, 9(3): 67-72.
- [6] P. Wang, D. Li, X. Meng, "The Management Mode of College Students' Mental Health Based on Wireless Sensor Network," *International Journal of Frontiers in Sociology*, 2022, 40(2): 45-49.
- [7] W. Zheng, "Cluster Analysis Algorithm in the Analysis of College Students' Mental Health Education," *Applied bionics and biomechanics*, 2022, 63(6): 19-21.
- [8] L. Yi, "Analysis on the Innovative Strategies of College Students' Mental Health Education under the Background of We-Media," *International Journal of Education and Teaching Research*, 2021, 2(3): 32-38.
- [9] K. Tamar, "Enhancing Racial/Ethnic Equity in College Student Mental Health Through Innovative Screening and Treatment," *Administration and policy in mental health*, 2021, 49(2): 1-16.
- [10] X. L. ZHANG, X. LI, Y. T. YANG, "Analysis of bi-directional reranking model for Uyghur-Chinese neural machine translation," *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2020, 56(1): 31-38.
- [11] Z. HANG, P. YAO, J. WANG, "Analytical modeling of surface roughness in precision grinding of particle reinforced metal matrix composites considering nanomechanical response of material," *International Journal of Mechanical Sciences*, 2019, 157-158: 243-253.
- [12] Z. ZHANG, P. YAO, J. WANG, "Analyze modeling of surface roughness in precision grinding of particle reinforced metal matrix composites considering nanomechanical response of material," *International Journal of Mechanical Sciences*, 2019, 157-158: 243-253.
- [13] Z. WANG, T. LIN, X. HE, "Fabrication and properties of the TiC reinforced high-strength steel matrix composite," *International Journal of Refractory Metals and Hard Materials*, 2016, 58: 14-21.
- [14] Z. H. LV, R. R. LOU, A. K. SINGH, "AI empowered communication systems for intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(7): 4579-4587.
- [15] Y. Zhao, X. Hui, "Research on search engine optimization of field condition monitoring data for transmission and distribution equipment," *Electronic Design Engineering*, 2021, 29(12): 29-32+38.
- [16] G. Liu, and H. Zhuang, "Evaluation model of multimedia-aided teaching effect of physical education course based on random forest algorithm," *Journal of Intelligent Systems*, 2022, 31(1): 555-567.
- [17] Liu Lifang, Ma Yuanyuan. A cross-modal information retrieval method based on multi-perspective symmetric non-negative matrix factorization [J]. *Journal of Shandong University (Science edition)*, 2022, 57 (07): 65-72 + 84.
- [18] Feng Jiao, Lu Changyu. Cross-media retrieval method based on the residual attention network [J]. *Computer Science*, 2021, 48 (S1): 122-126.
- [19] Wang Star, Yu Limei, Chen Ji. The Chinese word segmentation method of convolutional neural network fused with word root information [J]. *Small microcomputer system*, 2022, 43 (02): 271-277.
- [20] Liu Xing, Yang Lu, Hao Fanchang. Finger vein image retrieval method based on multiple feature fusion [J / OL]. *Journal of Shandong University (Engineering Edition)*: 1-9 [2023-03-29].
- [21] Wu Renbiao, Liu Yang, Jia Yunfei, Liu Shining, Qiao Han. Civil aviation key passenger risk assessment method based on improved XGBoost [J]. *Journal of Safety and Environment*, 2023, 23 (03): 651-658.
- [22] Qiu Ming language, Gan Shu. Experimental analysis of remote sensing information extraction method for multi-level rule classification [J]. *Journal of Guizhou University (Natural Science Edition)*, 2023, 40 (02): 67-73.
- [23] Chen Xiang, Yu Chi, Yang Guang, et al. A Bash code annotation generation method based on dual information retrieval [J]. *Journal of Software*, 2023, 34 (03): 1310-1329.
- [24] Jin Ru, Li Ying, Xu Yu, etc. Study on disaster-induced similar typhoon retrieval method based on the optimal empowerment of path and environmental field [J / OL]. *Water conservancy and hydropower technology (in Chinese and English)*: 1-17 [2023-03-29].
- [25] Miao Zhuang, Zhao Xinxin, Li Yang, Wang Jiabao, Zhang Rui. Depth-supervised hash image retrieval method based on Swin Transformer [J / OL]. *Journal of Hunan University (Natural Science Edition)*.
- [26] Gong Yun. English tamper-proof retrieval method based on perception hash [J]. *Automation and Instrumentation*, 2023 (02): 137-141 + 145.