Design of Network Public Opinion Information Intelligent Retrieval System Based on Wireless Network Technology

Quan-wei Sheng(✉)

Changsha Medical University, Changsha 410219, China

Abstract. Wireless sensor routing protocol affects the retrieval work of the system. The traditional intelligent retrieval system of network public opinion information is difficult to obtain the event probability due to the selected technology, which leads to the weak ability of the system to retrieve massive data. This paper designs an intelligent retrieval system of network public opinion information based on wireless network technology. In terms of hardware, Lucene search engine architecture and rs323 bus circuit are designed; in terms of software design, network public opinion information similarity calculation model is designed, wireless sensor routing protocol is set based on wireless network technology, and network public opinion information intelligent retrieval logic is established. In the experiment, the amount of public opinion information to be retrieved is 100, 5000 and 100000 respectively. In the same test period, when the amount of network public opinion information to be retrieved is large, the retrieval system based on wireless network technology can obtain the target information, while the amount of target information obtained by traditional system is far less than expected.

Keywords: Wireless network technology · Network public opinion information · Intelligence · Retrieval system

1 Introduction

In the face of the rapid growth of Internet public opinion information, some scholars refer to the content of [1] and design a Python based retrieval system. However, according to the practical application of other systems, it is found that the correspondence between the retrieval algorithms and models of most traditional systems is weak. Therefore, taking the Python based retrieval system as a reference, this paper studies the intelligent retrieval system of network public opinion information based on wireless network technology [1]. The so-called wireless network refers to the network that can realize the interconnection of various communication devices without wiring. Wireless network technology covers a wide range, including not only global voice and data networks that allow users to establish long-distance wireless connections, but also infrared and RF technologies that optimize short-distance wireless connections. Literature [2] Designing a public opinion information search system for Baidu search engine, relying

S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 417, pp. 74–86, 2022.
https://doi.org/10.1007/978-3-030-94554-1_7

on Baidu search engine, mainly researching crawler, search engine, text information mining and other technologies, should develop a complete solution for the application needs of network public opinion monitoring work. It can greatly improve the speed of searching public opinion information on Baidu website and Baidu Tieba and the accuracy of obtaining information. However, the target information acquisition volume of the system is far from expected.

Based on this, this article starts with wireless network technology and designs an intelligent retrieval system for network public opinion information that is more in line with expectations.

2 Hardware Design of Network Public Opinion Information Intelligent Retrieval System

2.1 Design Lucene Search Engine Architecture

Lucene, as a full-text search engine architecture, needs to be developed twice on the basis of Lucene if it is to be fully applied to intelligent retrieval of network public opinion information. Lucene is known to provide the analysis interface, index engine, storage management and query engine to the retrieval system. According to the above basic modules, Lucene core logic architecture is constructed, as shown in Fig. 1 below.

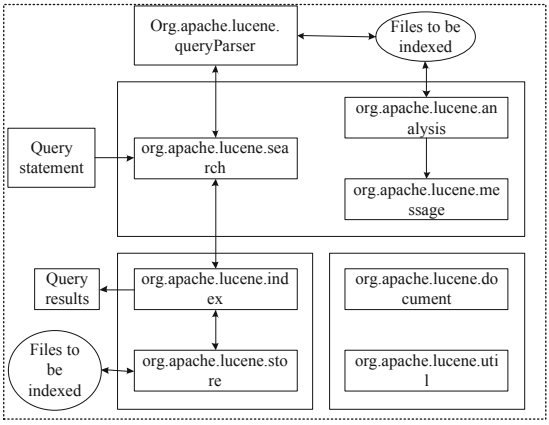


Fig. 1. Lucene search engine architecture

Through the basic components of the logical structure, we can see that there is a calling relationship between the various modules when Lucene is running. The interface module uses the parser to analyze the index text, and writes the analyzed data into the index file through the infrastructure common module. When users query, the query statements are sent to the index core module through the interface module, and the query results are read through the interface module Take the data in the index file, get the retrieval results, and execute the feedback operation [3].

2.2 Design rs323 Bus Circuit

When the system receives network public opinion information, there are synchronous transmission data and asynchronous transmission data. It needs to design rs323 bus circuit to ensure that the system can receive all kinds of information. The rs323 interface has 9 pins or 25 pins. The rs323 with 25 pins is selected in the design. Under the transmission rates of 50, 100, 600, 1200, 2400, 4800, 9600 and 19200 baud respectively, it is tested whether the rs323 allows 2500pf capacitive load. On the basis of expanding the communication distance, it is ensured that the hardware has the ability to suppress noise and other interference information. The bus design is shown in Fig. 2 below Circuit [4].

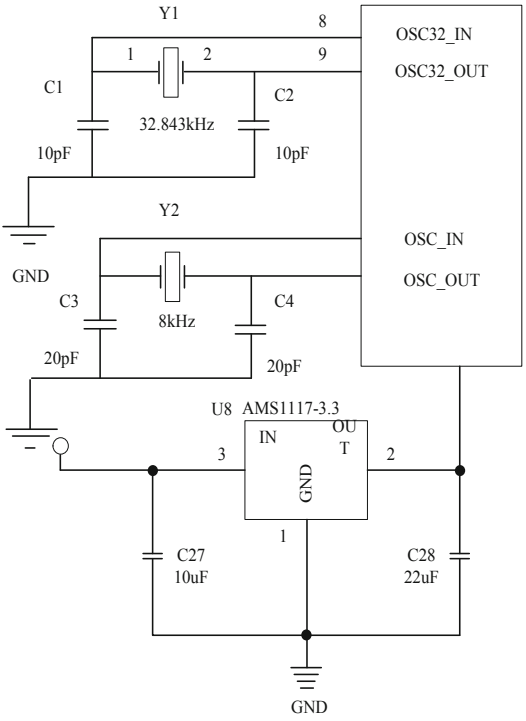


Fig. 2. Bus circuit design

Considering that the system is connected with the computer, MAX3232 chip is used as the transceiver chip of rs323 bus circuit. The MAX3232 chip has two channels, and it has a dedicated low-voltage transmitter output stage. Through two receivers and two drivers, it can realize the data receiving and sending with the data rate of 240kps, and ensure the efficiency of the retrieval system in processing massive network public opinion data. So far, the hardware design of the library and information resources integration system is completed.

3 Software Design of Network Public Opinion Information Intelligent Retrieval System Based on Wireless Network Technology

3.1 Design the Similarity Calculation Model of Public Opinion Information on the Network

When retrieving network public opinion information, we need to perform the retrieval task according to semantic similarity. The common semantic feature of two concepts is the similarity between semantics. This index can be represented by $sim(x, y)$, and x and y represent the conceptual elements between the two information. In the form of calculation, similarity is required to meet the following calculation requirements: make similarity index between, so $sim(x, y) \in [0, 1]$. When the two information are completely similar, $sim(x, y) = 1$; when the two information is completely different, then $sim(x, y) = 0$. The similarity index is symmetrical. Therefore, according to the above calculation requirements, combined with the influence of retrieval semantic distance on similarity index, according to the decreasing function relationship, assuming the degree of change between similarity index and semantic distance is $k^{-dis(x,y)}$, the influence of network public opinion information on similarity index can be described by the following formula:

$$f(x, y) = \frac{C(x) + C(y)}{|C(x) - C(y)| + 1} \tag{1}$$

In the formula, $C(x)$ and $C(y)$ represent the content level of information x and y respectively. When $f(x, y) > 1$, the range of similarity index is not met. Therefore, divide f by 2 times the depth of layer d, so that the influence of hierarchy factors on conceptual similarity is kept within the range, and a new calculation result is obtained:

$$f(x, y) = \frac{C(x) + C(y)}{2d\left[|C(x) - C(y)| + 1\right]} \tag{2}$$

According to the above calculation formula, with the increase of concept level, the similarity between them is also increasing. At the same time, with the expansion of the difference between the two concept levels, the similarity between retrieval elements is gradually decreasing [5]. Therefore, the change degree between the two elements is obtained

$$dis(x, y) = C(x) + C(y) - 2F(f(x, y)) \tag{3}$$

In the formula: $dis(x, y)$ represents the semantic distance between two concepts; $F(f(x, y))$ represents the level of the minimum classification of two concepts. By quantifying the influence of semantic distance and hierarchical factors on the similarity index, we get the following results:

$$Z_1(x, y) = \frac{k^{-dis(x,y)} \cdot q_{x,y}}{2d \cdot (dis(x, y) + 1)} \tag{4}$$

In the formula: $q_{x,y}$ is the sum of levels $C(x) + C(y)$. At the same time, the logarithmic function is introduced and the nonlinear function is used to evaluate the semantic similarity:

$$Z_2(x, y) = \log_2\left(1 + \frac{|s(x) \cap s(y)|}{|s(x) \cup s(y)|}\right) \tag{5}$$

In order to avoid infinite values of x and y, 1 is added to the true part of logarithmic function to make its value 0. Through the above calculation process, the semantic similarity calculation model is designed, and parameters w_1 and w_2 are added to adjust the similarity index:

$$sim(x, y) = \begin{cases} w_1 Z_1 + w_2 Z_2 & x \neq y \\ 1 & x \neq y \end{cases} \tag{6}$$

According to the above calculation formula, the design of network public opinion information similarity calculation model is realized [6].

3.2 Wireless Sensor Routing Protocol Based on Wireless Network Technology

When wireless sensor networks transmit data, there are a lot of uncertainties among the cloud security situation elements, so wireless network technology is used to determine the uncertainty of cloud security situation elements. Wireless network technology uses credibility function H to calculate the uncertainty relationship between various elements of cloud security situation:

$$H(U, V) = W(U, V) - Y(U, V) \tag{7}$$

In the formula: U represents the result caused by uncertain reasons; V represents the uncertain reasons between cloud security elements; W and Y represent the growth of trust and distrust respectively. The parameter calculation equations of the two functions are as follows:

$$W(U, V) = \begin{cases} 1, if P(U) = 1 \\ \dfrac{\max\left(P\left(\frac{U}{V}\right), P(U) - P(V)\right)}{P(U)} \end{cases} \tag{8}$$

$$Y(U, V) = \begin{cases} 1, if P(U) = 0 \\ -\dfrac{\min\left(P\left(\frac{U}{V}\right), P(U) - P(V)\right)}{P(U)} \end{cases} \tag{9}$$

In the formula, $P(U)$ is the prior probability of the result; $P\left(\frac{U}{V}\right)$ is the conditional probability of X when the uncertainty occurs. When n different security events cause a security event, if the event is $n_1, n_2, \ldots, n_n$, then n credibility can be obtained by wireless network technology. The uncertainty relationship between cloud security situation elements can be measured by using the results obtained. The comprehensive trust value is calculated according to the direct trust value, indirect trust value and energy trust value. According

to the actual number of packets received and sent, the direct trust value between nodes is predicted:

$$X^{now(1)} = \mu(\omega_1 S_1 + \omega_2 S_2)^{pre} + (1-\mu)(S_1 + S_2)^{now} \tag{10}$$

In the formula, μ represents the weight of historical trust value; $(1-\mu)$ represents the weight of current trust value; ω_1 and ω_2 represent different volatile factors; S_1 and S_2 respectively table send and receive packets, accounting for the proportion of total packets. Direct trust value is used to calculate indirect trust value between different nodes. Suppose $A_n = (A_1, A_2, \ldots, A_m)$ is a set of neighbor nodes, where m is the actual number of public nodes, so the calculation result of indirect trust value is as follows:

$$X^{now(2)} = \frac{1}{n} \sum_{a \in A_n}^{n} \left(X_{ia}^{now(1)} \cdot X_{aj}^{now(1)} \right) \tag{11}$$

In the formula, $X_{ia}^{now(1)}$ and $X_{aj}^{now(1)}$ represent the direct trust values between node i and node a, node a and node j respectively. Calculate the energy trust value, the formula is:

$$E_j = \frac{S_1 E_j}{E_0} \tag{12}$$

In the formula, $S_1 E_j$ is the residual energy value; E_0 is the initial energy value. The comprehensive trust value is calculated according to the above three formulas:

$$X = H_1 X^{now(1)} + H_2 X^{now(2)} + H_3 E_j \tag{13}$$

In the formula: H_1, H_2 and H_3 are the prediction weights of wireless network technology. According to the basic definition of compressed sensing theory, if the dimension of measurement matrix is $M \times N$, then the column vector of $N \times 1$ dimension can be represented by b, then:

$$y = \Phi b = \begin{Bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{M1} & \alpha_{M2} & \cdots & \alpha_{MN} \end{Bmatrix} \begin{Bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{Bmatrix} \tag{14}$$

$$y_i = \sum_{j=1}^{N} \alpha_{ij} b_j \tag{15}$$

In the formula: Φ represents sparse matrix; α_{MN} represents transformation coefficient in $M \times N$ dimension; α_{ij} represents transformation quantity in row i and column j. By multiplying and adding N nodes b_j, M measurement results are obtained. The row elements of the sparse matrix are used as the transmission path projection of WSN, and

the matrix is divided into M sub regions:

$$\Phi = \begin{Bmatrix} \alpha_{11} & 0 & \alpha_{12} & 0 & \cdots & 0 & \alpha_{1\frac{N}{M}} & 0 \\ 0 & \alpha_{21} & 0 & 0 & \cdots & \alpha_{2\frac{N}{M}} & 0 & 0 \\ 0 & 0 & 0 & \alpha_{31} & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \alpha_{M\frac{N}{M}} \end{Bmatrix} = \begin{Bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \Phi_4 \\ \vdots \\ \Phi_M \end{Bmatrix} \tag{16}$$

According to the data in the formula, the matrix contains multiple regions and many 0 elements [7, 8]. The non-zero elements in the block matrix are multiplied by the corresponding data, and then the data is transmitted to reduce the energy consumption. According to the above process, ant colony algorithm divides cluster tree wireless sensor network into zones, and executes the routing protocol according to the established search method.

3.3 Establishing Intelligent Retrieval Logic of Network Public Opinion Information

In order to implement the search routing protocol designed above, it is necessary to formulate a verification search logic matching with it. Based on the characteristics of character type data, the verification retrieval logic was developed by stage selection. The occurrence frequency of known characters has certain deviation. For example, the occurrence probability of data in ordinary text may be greater than the occurrence probability of data. Therefore, after the index is generated, the character value is mapped to the corresponding bucket number according to the frequency hiding operation of region division, and then the order of all bucket numbers is straightened out to extract the frequency information characteristics of hidden characters [9]. According to the data obtained in the research process, the frequency order of English letters is given, as shown in Table 1 below.

Table 1. English alphabet frequency sorting table/%

Letter	Frequency	Letter	Frequency	Letter	Frequency
A	8.17	J	0.15	S	6.33
B	1.49	K	0.77	T	9.06
C	2.78	L	4.03	U	2.76
D	4.25	M	2.41	V	0.98
E	12.7	N	6.75	W	2.36
F	2.23	O	7.50	X	0.15

(*continued*)

Table 1. (*continued*)

Letter	Frequency	Letter	Frequency	Letter	Frequency
G	2.02	P	1.93	Y	1.97
H	6.10	Q	0.10	Z	0.07
I	6.97	R	5.99	–	–

In the process of hiding frequency operation, the characters are assigned to each bucket according to the frequency of characters, and then converted into the corresponding bucket number, so that users can infer the corresponding relationship between the bucket number and characters according to the number of buckets and character frequency. Combined with the letter frequency in Table 1, the basic verification process of the verification algorithm is designed. In the first step of verification, input the letter frequency and the number of buckets in Table 1, and calculate the frequency value that each bucket can hold

$$r = \frac{1}{n}(1 \pm 10\%) \tag{17}$$

In the formula: n represents the number of buckets; $(1 \pm 10\%)$ represents the floating range of splitting the same character into different buckets in the subsequent allocation process. The second step is to ensure that there is a corresponding mapping relationship between characters and bucket numbers according to the following restrictions: first, assign characters to each bucket in order until the end of allocation; second, when a character with the highest frequency cannot be placed in the remaining space in the bucket, it is proved that the bucket can end the allocation task and start to allocate the next bucket. In the third step, hash function is used to capture the random value of bucket number. The verification process of the function is controlled by the following formula:

$$hash(*) = s\varphi keymod\,1(0, 1) \tag{18}$$

In the formula: s represents the hash table size; $hash(*)$ represents the downward rounding of the expression by the multiplication hash function, and the value range of parameter φ is $(0, 1)$; $\varphi keymod\,1$ represents the fractional part of the numerical product of key multiplied by $(0, 1)$ interval [10]. In order to make the verification mode and encryption mode have a matching relationship, the default parameter $\varphi = 0.618$, through the above calculation and analysis process, the network public opinion information intelligent retrieval logic is established, and the network public opinion information intelligent retrieval system based on wireless network technology is realized.

4 Experimental Study

4.1 Purpose of the Experiment

Taking the retrieval system designed in this paper as the test object, find out the parts of the system that have conflicts, and then adjust for such problems to ensure the use

of the system in the actual work. Test whether the hardware of the system can connect stably and whether the redesigned hardware is compatible with the software. Finally, the system is applied to the actual working environment. Under different test conditions, the differences between the designed retrieval system and the traditional system are compared.

4.2 System Stability and Reliability Test

In order to ensure that the designed system has more powerful retrieval function and test the stability of the system, the test results of the system stability are shown in Fig. 3 below.

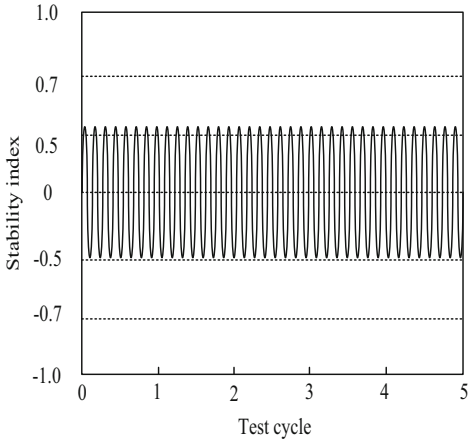


Fig. 3. System stability test results

According to the distribution state of the curve in the figure, the stability index test results of the system fluctuate evenly between −0.5–0.5, and there is no big difference between the stability indexes in each test stage. It can be seen that the designed system meets the stability test requirements. Set up system test data sets, use different data sets to record different scales of network public opinion information, test the retrieval efficiency of the system. The following Table 2 is the setting table for data sets of different sizes. The size of each small field is required to be between 2.5 and 3.0 mb.

Table 2. System test data set

Serial number	Number of files	Total dataset size
1	100	234.35 MB
2	500	1.86 GB
3	1000	3.43 GB
4	5000	17.75 GB
5	10000	38.43 GB

The design of the retrieval system as the experimental group, set up the general retrieval target of the retrieval system. The data in Table 3 below are the comparison results between the expected value and the actual retrieval time of the system.

Table 3. Comparison of retrieval time of Internet public opinion information

Number of files	Expected value	Actual value of the designed system
100	115 ms	115.7 ms
500	630 ms	663.2 ms
1000	1345 ms	1399.8 ms
5000	5750 ms	5884.1 ms
10000	13280 ms	13497.5 ms

In order to further illustrate the above comparison test results, the relationship curve between the actual value and the expected value of the system is drawn, as shown in Fig. 4 below.

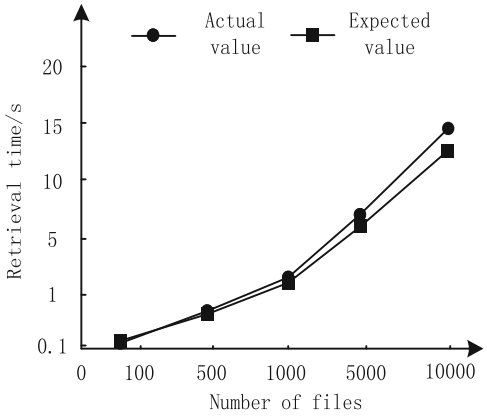


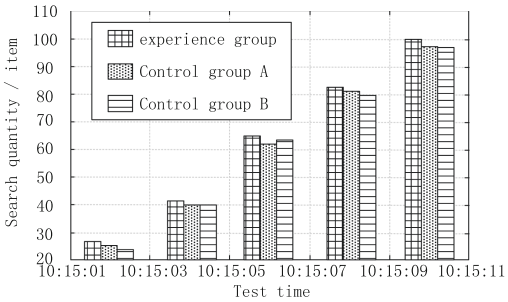
Fig. 4. Search result curve

According to the test results shown in Fig. 4, under the control of wireless network technology, the designed system achieves the retrieval efficiency which is highly similar to the expected value.

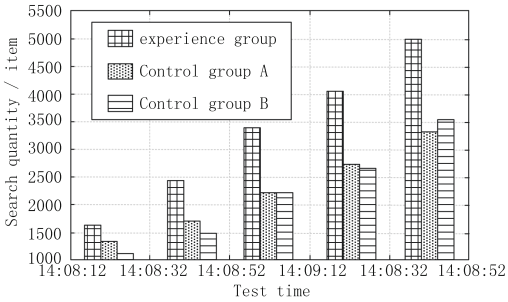
4.3 System Retrieval Ability Test

Taking the designed retrieval system as the experimental group, the traditional system 1 and the traditional system 2 designed by the traditional method as the control group A and the control group B respectively, the three groups of systems are applied to the actual network public opinion information retrieval work. The experiment set up three retrieval information quantity, which were 100, 5000 and 100000. Test the ability of the system to retrieve network public opinion information under different volumes, and the results are shown in Fig. 5 below.

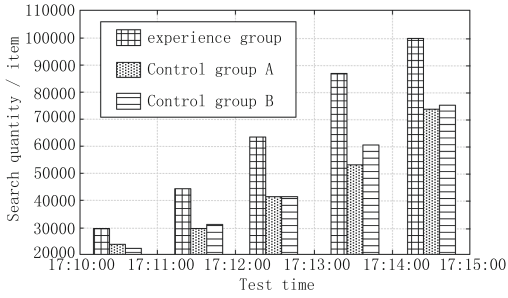
According to the test results shown in Fig. 5, when the number of network public opinion information to be retrieved is 100, the three groups of systems have better retrieval effect; when the number of network public opinion information to be retrieved is 5000, the amount of retrieval data obtained by the two traditional systems in the same test period is less than 4000, while the designed system can retrieve 5000 public opinion information; when the number of network public opinion information to be retrieved is 5000, the number of retrieval data obtained by the two traditional systems is less than 4000 When the number of public opinion information is 100000, the retrieval results of the two traditional systems drop sharply again. In the same test period, the public opinion information obtained by the two systems is less than 80000. However, the system designed in this paper obtains 100000 complete network public opinion information at the same time. According to the above test results, the system designed in this paper also has excellent data retrieval effect in the face of huge network public opinion information.



(a) Test results of 100 pieces of public opinion information



(b) Test results of 5000 pieces of public opinion information



(c) Test results of 100000 pieces of public opinion information

Fig. 5. System retrieval effect test

5 Concluding Remarks

This research is based on the shortcomings of the traditional retrieval system, through the wireless network technology to make up for the shortcomings of the traditional system, and further obtain more satisfactory retrieval results. But combined with the design of the system software, the calculation of software design is more complex, prone to system crash. In the future, we can design a set of intelligent control algorithm to realize the

operation of the software with a simpler calculation process, and provide more accurate test results for the intelligent retrieval of network public opinion information.

References

1. Xin, L.: From assessment to establishment: construction and development of cases assess catalog from a utilization perspective. J. Sichuan Univ. (Soc. Sci. Edn.) **02**, 161–173 (2020)
2. Tang, G., Zhao, C., Li, J., et al.: Research on public opinion information search system relying on Baidu search Engine. Comput. Digit. Eng. **047**(011), 2785–2790 (2019)
3. Xu, M., Zhang, X.: A small footprint keyword spotting system for voice control. J. Signal Process. **36**(06), 879–884 (2020)
4. Wu, D., Tao, S., Yan, P., et al.: Content-based instrument image retrieval system. Electron. Meas. Technol. **42**(11), 112–117 (2019)
5. Gong, P., Lyu, C., Gong, Y.: Research on object storage and retrieval system for space application data. Appl. Res. Comput. **36**(03), 833–837 (2019)
6. Li, G., Zhu, T., Liu, B., et al.: Optimization of the multi-media information retrieval system of digital library for big data. Inf. Sci. **37**(02), 115–119 (2019)
7. Yao, J., Cheng, Y.: Design and implementation of video retrieval system based on deep learning. Comput. Meas. Control **27**(06), 231–235 (2019)
8. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. Mob. Netw. Appl. (2021). https://doi.org/10.1007/s11036-020-01725-x
9. Liu, S., Pan, Z., Cheng, X.: A novel fast fractal image compression method based on distance clustering in high dimensional sphere surface. Fractals **25**(4), 1740004 (2017)
10. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. Multimedia Tools Appl. **76**(4), 5787–5802 (2014). https://doi.org/10.1007/s11042-014-2408-1