Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 6753671, 10 pages
https://doi.org/10.1155/2022/6753671

Research Article

Dynamic Web Page Graphic Design Method for Internet Big Data Information System

Dan Tao

Department of Arts, Huaide College of Changzhou University, Taizhou 214500, Jingjiang, China

Correspondence should be addressed to Dan Tao; anewtd@smail.cczu.edu.cn

Received 14 May 2022; Revised 8 July 2022; Accepted 11 July 2022; Published 5 August 2022

Academic Editor: Zaoli Yang

With the rapid development of the Internet, the page design technology of PC search engine tends to be mature. The content displayed in front of users not only pays attention to the practicability of information, but also takes into account the beauty of page layout. However, the web page design of the Internet mobile terminal is slightly backward, so we study the multimodal method based on the feature of large data density to extract the text, and use the K-means clustering algorithm to classify and recognize the text. The results show that the comprehensive evaluation of the design search engine on multiple platforms is higher than 0.89; In fuzzy keyword retrieval, the average accuracy of dynamic k-value clustering algorithm is 87.5%, while the average accuracy of traditional K-means clustering algorithm is 78.5%. Finally, in terms of user evaluation, the satisfaction of search pages increased by 5%–10%. Experiments show that the optimized algorithm and page design not only improve the accuracy and applicability in function, but also optimize the layout of text and pictures on the page.

1. Introduction

Search engine refers to a system that collects information from the Internet according to certain strategies and specific computer programs, provides retrieval services for users after organizing and processing the information, and displays the relevant information retrieved by users to users. Search engines include full-text index, catalog index, meta search engine, vertical search engine, collective search engine, portal search engine, and free link list. Baidu and Google are representatives of search engines. The background of search engines is that in the early stage of Internet development, there are relatively few websites, and it is relatively easy to find news. However, with the rapid development of news technology, especially the rapid popularization of Internet applications, there are more and more websites, and the number of global Internet pages increases by tens of millions every day. To find the required materials in the vast network news is like looking for a needle in a haystack. At this time, search websites to meet everyone's news retrieval needs came into being. With the increasing popularity of informatization in China, the Internet has connected the production and life of the whole society. The search engine integrating a large amount of network information is also widely used in various retrieval systems because of its simple use. The value of network data is recognized by the majority of people. People's demand for security information promotes the progress of network information extraction methods and network security defense means. For example, in the defense of offensive URLs, using search engines to study a malicious URL detection method based on the difference between normal and abnormal URL lengths (search volume) can prevent abnormal URLs from stealing user information [1]. As for the acquisition method of educational network resources, the query-related ranking model of specific curriculum or education level is constructed by using domain knowledge, which can effectively improve the quality of received educational resources [2].

The main method of this research is to improve the text image display function and layout of search engine by constructing a multimodal fusion text extraction and dynamic K-means optimization clustering algorithm. The density score is constructed by using the density

characteristics of text and the symbol density characteristics of text. At the same time, the average value of K-means algorithm is optimized based on the text-clustering threshold. Finally, the basic text of search engine is established through the two to display the web page layout. Through the optimization and joint use of the above optimization algorithms, the accuracy of text and picture extraction is improved, and the interference information of fuzzy keyword search is reduced.

The innovation of the research is to optimize the text extraction algorithm and text-clustering algorithm of search engine web pages at the same time, and connect the functionality of text display with the layout optimization of web page design. In the text extraction, multimodal is used to construct the density score of text, which avoids the difficulty of extraction caused by the different labels of text and picture. At the same time, the fixed K-means is dynamically optimized to make it more suitable for the diversified clustering of big data information.

The second part briefly describes the application of K-means theory and the experience of researchers in different industries in recent years. The second part is divided into four parts. The third part proposes a multimodal text extraction method based on the text density feature and studies the optimization process of K-means algorithm for text clustering. The last part is the simulation test of the search engine, which uses selenium automatic processing tool to verify the effectiveness of the algorithm optimization and applies the research optimization search engine to the specific library official account collection retrieval program to investigate the service satisfaction.

2. Related Works

In recent years, the rapid development of the Internet has promoted the continuous updating of web technology, and the functions of web pages are becoming more and more rich. The page design of web pages by scholars is developing toward the trend of high interactivity and high visual communication requirements. Liu et al. combined multimodal web page evaluation to obtain an overall feature vector, decomposed the complex subjective evaluation task into multimodal subtasks, and obtained more specific and accurate feature vectors by merging subtask evaluation. Experiments show that this method can solve the complex problem of subjective evaluation task of web design. The proposed model has good correlation with the subjective measurement of visitors and provides a strong reference for the aesthetic evaluation of web pages [3]. Based on digital technology and open collaboration network, Marinello et al. constructed a model for defining the main functions of digital enterprise to enterprise (B2B) platform. Experiments show that the quality-function deployment method and house of quality tools of the model can promote the development of new business cooperation between small- and medium-sized enterprises (SMEs) [4]. Ichindelean et al. explored the elements, structure, and design to improve the usability of web pages based on eye tracking technology and found that the center and right side of web pages were the most concerned according to the feedback of eyeball instrument, fixation position, sequence and duration, and saccade and revisit of the same elements [5]. Marco et al. designed a web-based application to predict probe binding specificity from genome-scale sequence alignment information by using supervised machine learning. The results show that the model realizes the function expansion of intelligent and easy-to-use image interface and simulation pipeline framework [6]. Murugudu et al. proposed and introduced a novel and efficient two-stage deep learning data crawler framework. The first stage uses search engines to collect accurate and highly relevant links, and the second stage uses adaptive site ranking to explore fast and relevant website links within the site. Experiments show that this method is suitable for single-query and multiquery forms with adaptive weight characteristics [7]. Eiter et al. designed a semantic rich dynamic map (LDM) based on Web RN and spatial stream database technology. Experiments show that the model can infer new information and provide expressive query function on Web stream, which can provide users with richer dynamic information of geographical environment [8]. Hu et al. proposed a guided convolution collaborative multimodal machine learning model (Grad CAM) to diagnose and explain mental diseases and brain function. Experiments show that the model has stable performance in cognitive function group classification and potential biological mechanism discovery [9].

Yao et al. explored the best architecture of multimodal fusion based on neural architecture search and proposed a relational graph prediction NAS (rgnas) method to make up for the deficiency of tag architecture. Experiments show that this method makes up for the deficiency of tag architecture in improving the accuracy of predictor, and realizes a reasonable trade-off between the accuracy of multimodal fusion and the complexity of search time [10]. Sotgiu et al. designed a multimodal artificial finger sensor based on flexible polyimide-based capacitive tactile sensor array. Experiments show that the sensor represents a new capacitive tactile sensor device with a resolution of sub millimeter of human fingertip sensitivity [11]. Kia et al. used decision-level (later stage) and feature-level (early stage) fusion methods to integrate emotional cross modal information and proposed a new context aware multimodal emotion analysis framework. Experiments show that the context integration of multimodal features such as text, sound, and vision provides better performance (91.39%) compared with unimodal features (89.24%) [12]. Sheikhhosseini et al. provided an algorithm combining weighted K-means clustering analysis and particle swarm optimization. The number of clusters was determined by Davies bouldin's measure (DB) and Chou Su Lai's measure (CS) validity index, which was used to analyze seismic event data and automatically identify the global optimal clusters. The results showed that the seven types of models based on DB effectiveness index identification were suitable for the considered seismic catalogue [13]. Y et al. proposed an anti-interference technology of SAW sensor based on K-means algorithm to distinguish saw response and sinusoidal interference. The experimental results show that this method can clearly distinguish saw

response and co-frequency interference signal and effectively improve the stability of SAW sensor [14]. Huang et al. proposed a novel and effective viscosity detection method combined with K-means clustering and moving window. Experiments show that this method can not only provide viscosity band estimation but also detect serious valve viscosity or accidental valve closure [15]. Chen et al. integrates big data information fusion and K-means clustering algorithm to realize the clustering and integration of English teaching ability index parameters, compile the corresponding English teaching resource allocation scheme and realize the evaluation of English teaching ability. The results show that using this method to evaluate English teaching ability has better information fusion and analysis ability and improves the accuracy of teaching ability evaluation and the efficiency of teaching resource application [16].

To sum up, multimodal fusion algorithm and K-means clustering algorithm have been widely used in medicine, automatic robot, education, geological survey, and other fields. The research direction of web design is mainly based on the function and practicability of the program, while it is relatively less in the beautiful layout, comfort, and functional logic. Therefore, based on the text extraction and text-clustering algorithm of search engine, this paper optimizes the page layout of search engine, hoping to give consideration to the function optimization and page layout optimization of dynamic web pages.

3. Dynamic Web Page Graphic Design Method of Search Engine for Mobile Internet

3.1. Multimodal Text Extraction Method Based on Data Information Density Feature. Once the web content is published to the web server, the content of each static web page is saved on the web server, whether or not there is user access. That is to say, a static web page is a file actually saved on the server. Each web page is an independent file with relatively stable content and poor interactivity. Based on database technology, dynamic web pages can greatly reduce the workload of website maintenance. Websites using dynamic web page technology can achieve more functions, such as user registration, user login, and online survey. Dynamic web pages are not actually web page files that exist independently on the server. Only when the user requests, the server will return a complete web page. The question mark of dynamic web pages has some problems for search engine retrieval. Therefore, when using dynamic websites to use search engines, certain technologies are needed to meet the requirements of search engines.

Static web page refers to the web page with fixed page content after HTML code generation and the page content of its corresponding dynamic web page will change with time. On search engine web pages, keyword completion, user registration and login and other functions depend on the interaction between dynamic web page and background data [17]. Dynamic web pages use asynchronous JavaScript and XML (Ajax) technology to load data. This technology can provide local refresh function for web pages without HTML code change, so as to feed back the data to users faster. The dynamic web page application design of search engine is shown in Figure 1.

As can be seen from Figure 1, Internet search engine is essentially a large database, which covers the data information of multiple platforms, so as to provide users with a variety of information retrieval services. However, on the mobile terminal, the keyword based search method has some defects, such as keyword ambiguity, incomplete feedback content, poor image, and video effect of dynamic web pages [18]. The representation of text and the selection of its feature items are a basic problem in text mining and information retrieval. It quantifies the feature words extracted from the text to represent the text information. Transform them from a unstructured original text into structured information that can be recognized and processed by the computer, that is, scientifically abstract the text and establish its mathematical model to describe and replace the text so that the computer can recognize the text through the calculation and operation of this model. At the same time, the traditional text extraction method is based on the source code structure of web pages, while the codes of multiple platforms covered by search engines are different, which leads to problems such as difficult image extraction and heavy workload of search engines [19]. Therefore, the research uses the multimodal deep learning method to extract the text data of the platform. First, the source code of the platform web page is converted into the form of string to obtain the web page object node, and then the label node irrelevant to the text is removed to obtain the main content of the web page. At the same time, each node in the web page body code is traversed, and the node is processed by hash function. Finally, the text density of the node is calculated. The calculation formula is shown in formula (1):

$$TD_i = \frac{N_i - LN_i}{NG_i - LNG_i}. \tag{1}$$

In formula (1), i is the node object, TD_i is the text density of the node, N_i is the number of string words of the node, LN_i is the number of linked string words of the node, NG_i is the number of node labels, and LNG_i is the number of connected labels of the node. After the text density is obtained, the symbol density of the node is calculated, and the calculation formula is shown in formula (2):

$$SD_i = \frac{N_i - LN_i}{S_i + 1}. \tag{2}$$

In equation (2), SD_i is the symbol density of the calculated node i, and S_i is the number of punctuation symbols counted after traversing the string. Because the text characters in the node occupy a small part of useless information, and there are some useless symbols in the number of symbols, the symbol density of the node is calculated by using the ratio of the number of string words to the number of symbols, which can avoid the interference of unknown amount of noise in text extraction. After the text density and symbol density of the node are obtained, the comprehensive score of the node is calculated, and the calculation formula is shown in formula (3):

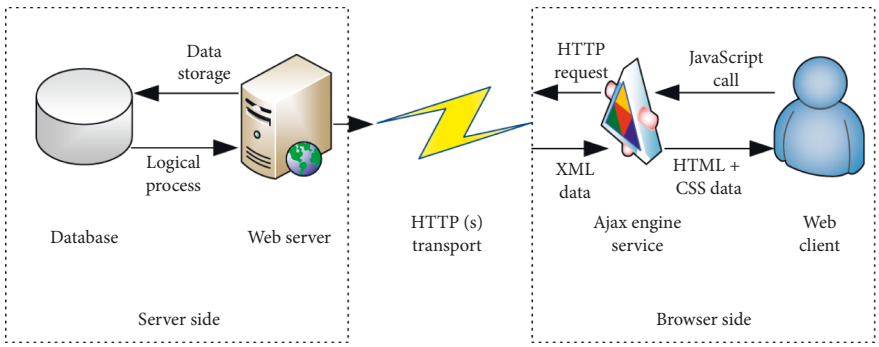

Figure 1: Dynamic web application of search engine.

$$TSD_i = TD_i \times SD_i, \tag{3}$$

TSD_i in equation (3) is the comprehensive score of nodes. After calculating the comprehensive scores of all nodes of the web page, sort according to the size, extract the node text with the highest score as the text content, and record the text label. For the extraction of pictures and videos in the text, it is necessary to calculate the label path from the text node to the picture node, extract the picture with the shortest path calculated as the text picture, and set the threshold $l_{\max} = 5$ to exclude the pictures in the nontext part of the web page. After extracting the text body, evaluate the algorithm, calculate, and analyze the accuracy of text extraction and the proportion of the correct text extracted by the algorithm in the manually marked text. The specific expression is shown in formula (4):

$$\begin{cases} A_i = \dfrac{DT_i \cap RT_i}{DT_i} \\ R_i = \dfrac{DT_i \cap RT_i}{RT_i}. \end{cases} \tag{4}$$

In formula (4), A_i represents the text extraction accuracy of web page i, R_i represents the text extraction recall rate of web page i, DT and RT represent the text collection extracted by the algorithm, and the manually marked text collection respectively. Finally, the comprehensive evaluation of the performance of the web page text extraction algorithm is calculated through the accuracy and recall rate, and its calculation formula is shown in formula (5):

$$S_i = \frac{2 \times DT_i \times RT_i}{DT_i + RT_i}. \tag{5}$$

3.2. Research on Text Clustering Algorithm Based on Optimized K-Means. Clustering algorithm realizes text category aggregation by analyzing the correlation between text data. In terms of keyword retrieval of search engine, the meaning of keywords retrieved by users partially coincides with other text contents, resulting in a large gap between the extracted content and the required content [20]. Therefore, K-means clustering algorithm is introduced. Assuming that in the mixed text dataset $T = \{t_1, t_2, \ldots, t_n\}$, the similarity measures adopted by the text dataset are aggregated into a subsets, and each subset represents a category, the clustering function formula to judge the clustering effect is shown in formula (6):

$$f_a = \sum_{j=1}^{a} \sum_{k=1}^{n_j} \left\| t_k - m_j \right\|^2. \tag{6}$$

In equation (6), j is the center of the text data subset, f_a is the function of the data sample and the collection center, and m_j is the mean value of the samples in the class. Its calculation formula is shown in .equation (7):

$$m_j = \frac{1}{n} \sum_{j=1}^{n_j} t_j. \tag{7}$$

In equation (7), n is the number of samples of text data. The smaller the sum of squares of errors calculated by the clustering function, the better the clustering effect. However, due to the large amount of text data and various text types in the Internet, the initial K value cannot be determined. Therefore, the dynamic K value is designed based on text data and the K-means clustering algorithm is improved. First, preprocess the text, eliminate irrelevant contents in the text, and form a text vector. The formula for calculating word frequency is shown in equation (8):

$$w(i, j) = \log \frac{n_i}{n_j}. \tag{8}$$

In equation (8), $w(i, j)$ represents the frequency of the word i in the text j, n_i represents the number of words i, n_j represents the total number of words in the text, and the inverse text frequency of the word is calculated. The calculation formula is shown in equation (9):

$$I_i = \log \frac{N}{N_i}. \tag{9}$$

In equation (9), I_i refers to the inverse text frequency of the word i, indicating the association weight between a word and the text subject, N represents the number of samples in the text data sample set, N_i represents the number of texts containing the word i, and the larger the I_i value, the wider the universality of the word i and the smaller the association with the text subject. According to the word frequency and

inverse text frequency, the feature vector of the text is obtained, and its calculation formula is shown in equation (10):

$$TV(i, j) = w(i, j) \times I_i. \tag{10}$$

In equation (10), $TV(i, j)$ represents the feature vector of the text j. After calculating the feature vector of the text, the similarity of the text can be calculated according to the feature vector, and the similarity calculation formula is shown in equation (11):

$$sim(j, k) = \frac{\sum_{i=1}^{n} j_i k_i}{\sqrt{\sum_{i=1}^{n} j_i^2} \sqrt{\sum_{i=1}^{n} k_i^2}}. \tag{11}$$

In equation (11), $sim(j, k)$ represents the similarity of j, k two texts, and j_i, k_i represents the feature vectors of the two texts, respectively. Set the initial mean $k = 2$, assuming that the total number of combined texts is N and the category label is $C = \{c_1, \ldots, c_k\}$, the expected description required for the classification of text data set is as follows:

$$P(n_1, n_2, \ldots n_k) = -\sum_{i=1}^{x} \frac{n_i}{N} * \log_2\left(\frac{n_i}{N}\right). \tag{12}$$

In equation (12), P is the classification expectation, n_i is the text quantity of the text category c_i. After clustering, calculate the text mean point of each text cluster as the new center, repeat clustering until the cluster center is fixed, and then calculate the average error sum of the category. The calculation model is shown in equation (13):

$$E = \frac{\sum_{i=1}^{m} (s_i - Ms)}{m}. \tag{13}$$

In equation (13), E is the sum of the average error of the category in which the text i is located, m is the number of texts in the cluster in which the text i is located, s_i is the similarity between the text i and the center of the cluster, and Ms is the average similarity of each cluster. When the threshold value e is set, the dynamic change expression of K value is shown in equation (14):

$$k = \begin{cases} k + 1, E > e \\ k, E \leq e. \end{cases} \tag{14}$$

Equation (14) is the change condition of the algorithm mean. When the category average error sum is greater than the threshold e, reset the initial mean value, repeat cluster division and similarity calculation in the cluster until the category average error sum is less than the set threshold e. Finally, the clustering accuracy is calculated according to the accuracy formula. The specific formula is shown in formula (15):

$$ACC = \frac{C \cap R}{C}. \tag{15}$$

In equation (15), C represents the correct web page text set obtained by clustering algorithm, and R represents the correct web page text set manually marked. According to the above text extraction method and text clustering algorithm, the search engine page of the mobile terminal is designed. The specific module of the system is shown in Figure 2.

In Figure 2, the system of dynamic web page graphic design method of mobile search engine mainly includes retrieval text extraction module, retrieval text clustering module, and text information client display module. The text extraction module calculates the text density and symbol density of data information by relying on the multimodal text extraction method and extracts useful information according to the maximum value of text density score; The text clustering method uses the k-means algorithm to design a dynamic k-value optimized k-means algorithm to improve the text clustering accuracy of big data. Finally, the user information display module is embodied in the web page graphic design of the search engine. The multimodal extraction method is used to optimize the information extraction of pictures and videos. At the same time, the k-means algorithm is used to optimize the information classification layout, so as to bring users a good retrieval service experience.

As an extension of human-machine interface design (HM), the user interface design of the website is an evolutionary evolution of the interaction between human and machine, and continues to develop with the development of computer and network technology. It uses text, image, video, audio, and other technologies as means to transform between virtual and reality, and provides end users with a multimedia, multidirectional information world. The effective organization and display of each information block is one of the design problems faced by user interface designers. Once visual designers try to separate many types of information, it is bound to generate more information blocks on the user interface, and the layout will become more complex.

4. Research on the Application of Search Engine Dynamic Web Design System

4.1. Algorithm System Performance Experiment of Search Engine Web Page Design. In order to verify the performance of multimodal text extraction algorithm and dynamic K-means optimization text clustering algorithm, selenium automatic processing tool is used to simulate the operation of browser and carry out the simulation experiment of text extraction. In the study, microblog, Taobao, headlines, and auto home were selected from four website types: social networking, shopping, news, and industry. 200 web pages were randomly selected from each platform to test the accuracy of text extraction algorithm. The multimodal text extraction method (MTE), feature extraction method based on text density (TDE), and feature extraction method based on label path (LPE) were compared. The specific experimental results are shown in Table 1.

In Table 1, "ACC" is the text extraction accuracy, "re" is the recall rate, "C" is the comprehensive evaluation. It can be seen from the table that the extraction algorithm based on text density feature performs better in accuracy on the website platform with more text, and the comprehensive evaluation of TDE algorithm in headlines and car home is 0.94 and 0.96, respectively. The extraction algorithm based on tag path feature has higher accuracy in websites with more pictures and videos. The comprehensive evaluation of

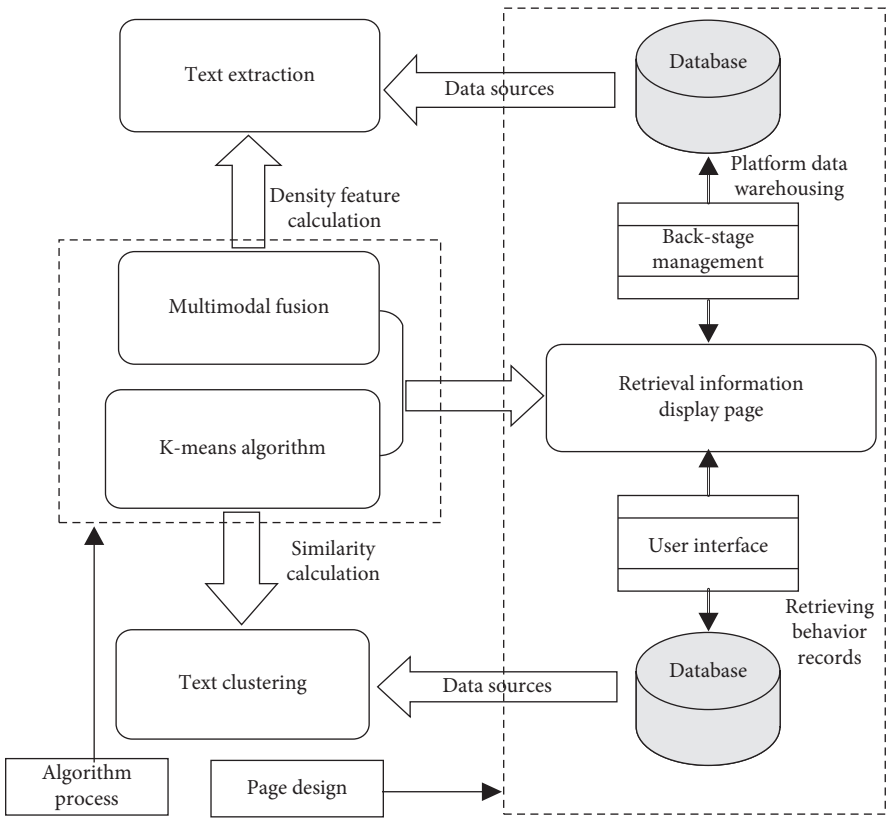

Figure 2: Web graphic design system module of mobile search engine.

Table 1: Text extraction and evaluation of different website platforms.

Data sources	MTE			TDE			LPE		
	Acc	Re	C	Acc	Re	C	Acc	Re	C
Weibo	0.91	0.90	0.93	0.92	0.89	0.89	0.87	0.91	0.90
TaoBao	0.87	0.87	0.89	0.88	0.86	0.87	0.91	0.94	0.96
Toutiao	0.91	0.89	0.91	0.94	0.91	0.94	0.89	0.92	0.91
Autocar	0.92	0.93	0.95	0.93	0.95	0.96	0.86	0.87	0.89

LPE in microwave and Taobao text extraction is 0.90 and 0.96. The multimodal feature text extraction method used in the study is more stable, and the comprehensive evaluation of text extraction on the four platforms is higher than 0.89. It is proved that the multimodal feature extraction method has wider adaptability and can accurately extract text on websites with high proportion of pictures and text. At the same time, among the 200 web pages of the four platforms, the accuracy of multimodal text extraction algorithm in image extraction is shown in Figure 3.

It can be seen from Figure 3 that the text extraction method of multimodal features used in the study is stable, and the comprehensive evaluation on the image extraction of the four platforms is higher than 0.89. The accuracy of image extraction is in the range of 0.87–0.94, the recall rate of image extraction is in the range of 0.87–0.93, and the comprehensive evaluation is in the range of 0.89–0.96. Experiments show that the text extraction method of multimodal features has superior image extraction

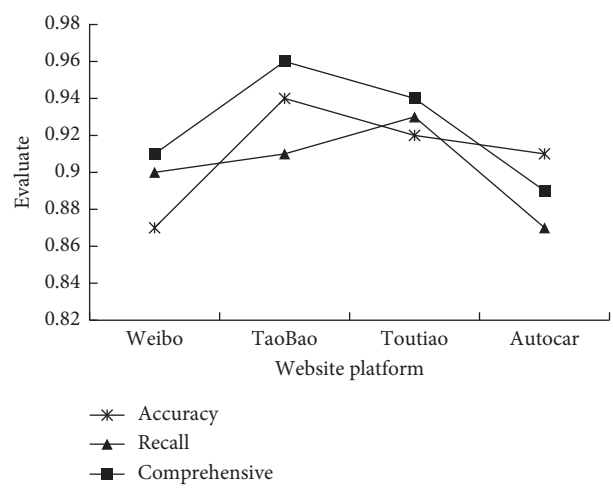

Figure 3: Evaluation of image extraction accuracy of MTE on different platforms.

performance and wide applicability. After verifying the accuracy and performance of the text extraction method, selenium automatic processing tool is used to simulate the retrieval of eight Keywords: "microblog," "social networking," "Taobao," "shopping," "headline," "news," "car home," and "industry." 200 data are extracted from each key word, and the traditional K-means algorithm and dynamic mean optimization algorithm are used to cluster the data samples. If the fixed k value of traditional K-means algorithm is 2, the comparison of clustering accuracy is shown in Figure 4.

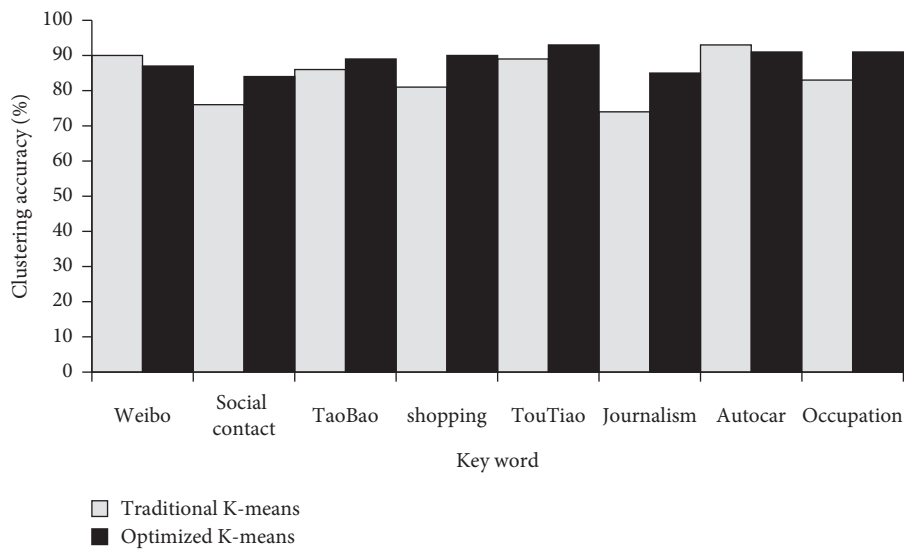

Figure 4: Optimize the clustering accuracy of K-means in keyword retrieval data.

In Figure 4, the optimized clustering algorithm of dynamic k-value in the retrieval data of 8 keywords, only two retrieval-clustering effects are lower than the traditional K-means clustering algorithm. On the whole, the average accuracy of dynamic k-value clustering is 90%, which is much higher than the K-means clustering algorithm (89.5%). The average accuracy of dynamic k-value clustering algorithm (87.5%) is higher than that of traditional K-means clustering algorithm (78.5%). Experiments show that the clustering algorithm with optimized K value is more stable in the clustering effect of retrieved data, and has higher precision performance in keywords with diverse meanings. Take 4 fuzzy keywords as data sample categories and number them 1–4, and randomly select 50 search results as data samples for classification test, then the classification results are shown in Figure 5.

In Figure 5, among the 50 test samples, the traditional clustering algorithm has 16 misjudgments, and the recognition accuracy is only 68%, while the optimized clustering algorithm has only two misjudgments, and the accuracy is 90%. The data show that the optimized clustering algorithm has higher accuracy in judging the category of fuzzy keyword retrieval data samples. By clustering the intersecting data sets existing in the fuzzy clustering algorithm, the results obtained are obviously better than the traditional clustering methods, but the clustering results for the data sets containing noise are not ideal. Compared with the traditional clustering algorithm, the convergence speed of fuzzy clustering algorithm is usually slow. Fuzzy clustering analysis is widely used in data mining, pattern recognition, machine learning, decision support, and other fields.

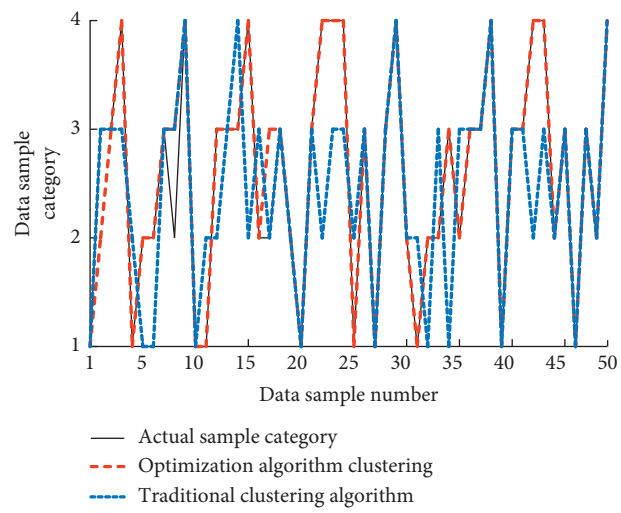

Figure 5: Fuzzy keyword algorithm classification test results.

4.2. Research on the Application Value of Search Engine Web Pages. The emergence of search engine integrates many web resources on the Internet and provides information navigation and information query services, which make the value of information widely recognized by netizens and manufacturers. However, the success of many professional websites and industry websites independent of the Internet just proves that the pattern of the Internet should be multifaceted. The nature of general search engine determines that it cannot meet the precise information needs of special fields and special people. The diversification of market demand determines the value of search engine. In order to verify the practical application effect of search engine, the research chooses to load the system design into the collection retrieval service program of the official account of a college library to test the retrieval accuracy and to investigate the users' evaluation of the book retrieval pages and service effect. 50 data records are randomly selected from the retrieval historical data of the system as test samples. The four information of book title, author, publishing house, and request number of book retrieval results are clustered and numbered as 1–4 in turn. The specific error is shown in Figure 6 by comparing the error between the information

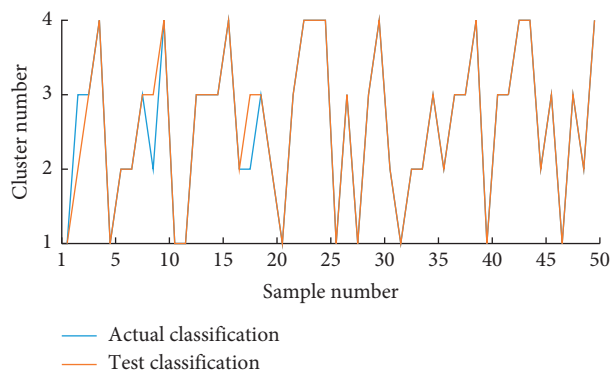

Figure 6: Clustering error of book retrieval in search engine.

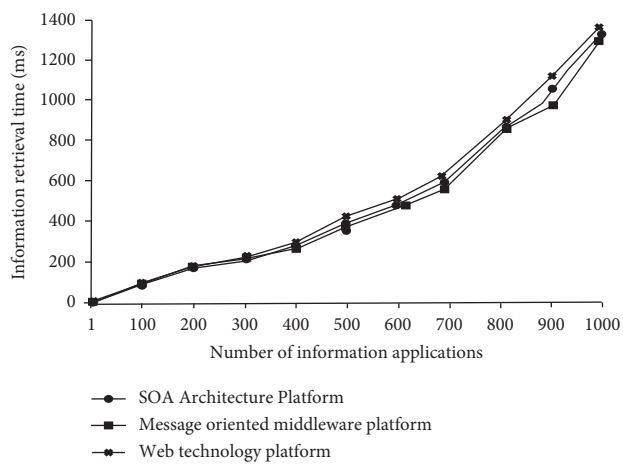

Figure 7: This paper studies the application effect of the algorithm in information retrieval under different technical platforms.

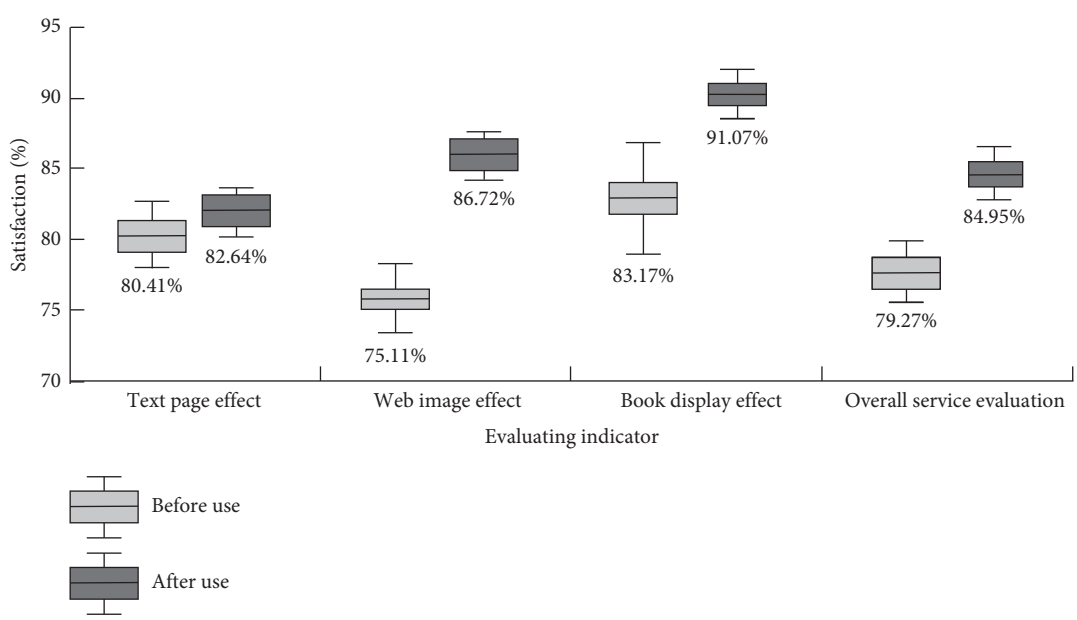

Figure 8: Average user satisfaction of book retrieval system.

clustering of retrieved books and the clustering of actual data.

In Figure 6, the error of the algorithm studied and applied in book retrieval is small. Among the 50 groups of retrieval sample data, there are only three errors, and the accuracy rate is as high as 94%, and the error is concentrated on the information clustering of authors and publishers. There are no clustering errors in other book titles and search numbers, which prove that the multimodal text feature extraction and dynamic K-means clustering algorithm have high accuracy and performance in book retrieval. Internet information covers a wide range of content levels, which make the designed dynamic web pages have a certain speed delay when processing a large amount of information data, which will also be affected by the equipment technology platform. The proposed algorithm is applied to different data platforms, and the information data processing time is counted. The results are shown in Figure 7.

It can be seen from Figure 7 that the difference of information retrieval time of the algorithm used in the research under SOA architecture platform, information middleware platform, and web technology platform is small, and the difference of time consumed is no more than 5%. With the increase of the amount of information data, the operation efficiency of the algorithm used in the research on Web page data processing is on the rise. The results show that the web page information retrieval designed by the algorithm can better adapt to different hardware technology platforms. Finally, investigate the user satisfaction of the library collection retrieval system, as shown in Figure 8.

In Figure 8, the new system feeds back the user retrieval service effect according to the user rating behavior. The average service satisfaction of users is 84.95%, which is 5.68% higher than that of the old system, indicating that users generally recognize the display effect of the optimization system on the web page. In terms of text display effect, the user satisfaction of the new system is $82.64\% \pm 1.86\%$, and that of the old system is $80.41\% \pm 0.79\%$, indicating that the pure text display effect of the retrieval web page is more beautiful and concise. On the picture display effect and book display interface, the average user satisfaction increased by 11.6% and 7.9%, indicating that the book page data of the new system is complete and the clustering is accurate.

5. Conclusion

The combination of mobile Internet and information technology will promote the transformation of business model and communication technology. Therefore, the application of search engine in mobile Internet will occupy a broader future market. In order to verify the practical value and optimization performance of search engine, the simulation experiment is carried out by using selenium tool. The accuracy, recall, and comprehensive evaluation of search engine text extraction are compared. It is found that it can ensure the comprehensive evaluation of more than 0.89 in text extraction and image extraction. At the same time, the clustering effect of traditional K-means algorithm and dynamic k-value algorithm in keyword retrieval data is compared. The experiment shows that the clustering accuracy of dynamic k-means algorithm is 9% higher than that of traditional algorithm in fuzzy keyword retrieval. The research also applied the search engine to the collection retrieval system. In the error test of 50 actual samples, the accuracy rate of the traditional algorithm was only 68%, far lower than 90% of the optimization algorithm. Finally, the user's evaluation of the new retrieval system's functions and pages was investigated. The data showed that the optimized retrieval system had a 5%–11% improvement in the satisfaction of the text layout, picture clarity, and the overall display effect of books. The experimental results show that the performance of keyword extraction is more stable in terms of word layout and fuzzy search engine. The disadvantage of the experiment is that there are only 50 measured data samples, and the lack of data may affect the judgment of error accuracy. In addition, because users' retrieval has the same distribution characteristics in a period of time the research can adopt a caching strategy to cache the recent retrieval results of users so that the search engine can display the common query results of users to users faster and improve the retrieval performance of the system. These need to be discussed in future research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Jang, J. Song, and M. Kim, "A study on the detection method for malicious URLs based on a number of search results matching the Internet search engines combining the machine learning," *Journal of Electrical Engineering & Technology*, vol. 17, no. 1, pp. 617–626, 2022.

[2] A. Usta, I. S. Altingovde, R. Ozcan, and O. Ulusoy, "Learning to rank for educational search engines," *IEEE Transactions on Learning Technologies*, vol. 14, no. 2, pp. 211–225, 2021.

[3] X. Liu and Y. Jiang, "Aesthetic assessment of website design based on multimodal fusion," *Future Generation Computer Systems*, vol. 117, no. 164, pp. 433–438, 2021.

[4] S. Marinello, F. Lolli, and R. Gamberini, "An open innovation B2B web platform design: application of the QFD approach for the definition of its primary functions," *International Journal of Product Development*, vol. 25, no. 3, pp. 255–282, 2021.

[5] M. Ichindelean, M. T. Ichindelean, I. Cetin, and G. Orzan, "A comparative eye tracking study of usability—towards sustainable web design," *Sustainability*, vol. 13, 2021.

[6] P. Marco, M. Martina, B. Valeria et al., "OligoMinerApp: a web-server application for the design of genome-scale oligonucleotide in situ hybridization probes through the flexible OligoMiner environment," *Nucleic Acids Research*, vol. 48, no. W1, pp. 332–339, 2020.

[7] F. Thomaz, C. Salge, E. Karahanna et al., "Learning from the dark web: leveraging conversational agents in the era of hyper-privacy to enhance marketing," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 43–63, 2020.

[8] T. Eiter, H. Füreder, F. Kasslatter, J. X. Parreira, and P. Schneider, "Towards a semantically enriched local dynamic map," *International Journal of Intelligent Transportation Systems Research*, vol. 17, no. 1, pp. 32–48, 2019.

[9] W. Hu, X. Meng, Y. Bai et al., "Interpretable multimodal fusion networks reveal mechanisms of brain cognition," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1474–1483, 2021.

[10] X. Yao, F. Li, and Y. Zeng, "Relational structure predictive neural architecture search for multimodal fusion," *Soft Computing*, vol. 26, no. 6, pp. 2807–2818, 2022.

[11] E. Sotgiu, D. E. Aguiam, C. Calaza et al., "Surface texture detection with a new sub-mm resolution flexible tactile capacitive sensor array for multimodal artificial finger," *Journal of Microelectromechanical Systems*, vol. 29, no. 5, pp. 629–636, 2020.

[12] K. Dashtipour, M. Gogate, E. Cambria, and A. Hussain, "A novel context-aware multimodal framework for Persian sentiment analysis - ScienceDirect," *Neurocomputing*, vol. 457, pp. 377–388, 2021.

[13] Z. Sheikhhosseini, N. Mirzaei, R. Heidari, and H. Monkaresi, "Delineation of potential seismic sources using weighted K-means cluster analysis and particle swarm optimization (PSO)," *Acta Geophysica*, vol. 69, no. 6, pp. 2161–2172, 2021.

[14] Y. Fan, Y. Liu, H. Qi, F. Liu, and X. Ji, "Anti-interference technology of surface acoustic wave sensor based on K-means clustering algorithm," *IEEE Sensors Journal*, vol. 21, no. 7, pp. 8998–9007, 2021.

[15] B. Huang, D. Zheng, X. Sun, S. K. Damarla, A. Shah, and J. Amalraj, "Valve stiction detection and quantification using a K-means clustering based moving window approach," *Industrial & Engineering Chemistry Research*, vol. 60, no. 6, pp. 2563–2577, 2021.

[16] Z. Chen, "Using big data fuzzy K-means clustering and information fusion algorithm in English teaching ability evaluation," *Complexity*, vol. 2021, no. 5, pp. 1–9, Article ID 5554444, 2021.

[17] L. Kuo, T. Chang, and C. C. Lai, "Application of visual colors in dynamic web page design through affective cognition," *Multimedia Tools and Applications*, vol. 81, no. 3, pp. 4435–4454, 2022.

[18] P. Sun, B. R. Cha, K. Chung, and J. Kim, "Mobile IoT device summarizer using P2P web search engine and inherent characteristic of contents," *Peer-to-Peer Networking and Applications*, vol. 13, no. 10, pp. 684–693, 2019.

[19] L. J. Sankpal and S. H. Patil, "Rider-rank algorithm-based feature extraction for Re-ranking the webpages in the search engine," *The Computer Journal*, vol. 63, no. 10, pp. 1479–1489, 2020.

[20] D. Ghai and N. Jain, "Comparative analysis of multi-scale wavelet decomposition and k-means clustering based text extraction," *Wireless Personal Communications*, vol. 109, no. 1, pp. 455–490, 2019.