Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

Research of Medical Intelligent Diagnosis System Based on Elasticsearch

Yun Hu[1], Libao Xing[2], Xintang Liu[2], Zuojian Zhou[1*], Guokai Han[2*], Hui Li[2*], Chunting Wang[3]
[1]School of Information Technology, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China
[2]School of Computer Engineering, Jiangsu Ocean University, Lianyungang, Jiangsu, China
[3]Kangda College of Nanjing Medical University, Lianyungang, Jiangsu, China
*Corresponding Author.

Abstract:

To address the problems that the experience of ambulance accompanying doctors is insufficient, and that hospitals have difficulty in obtaining diagnostic information of patients' diseases before the ambulance arrives at the hospital, and thus cannot make emergency preparations in advance. A medical intelligent diagnosis system with Elasticsearch as the core, combined with Spring Boot open source technology framework, python web crawler and other technologies has been studied and implemented. Extensive tests show that the system has a high accuracy of diagnosis, up to 83%; The diagnostic results are complete, including various types of information such as clinical examinations, commonly used drugs and treatments; and the average response time for diagnosis is about 21.27ms. Consequently, the system has a certain auxiliary function for the diagnosis of emergency patients.

Keywords*: Diagnosis, Elasticsearch, Search engine, Springboot, Spider.*

I. INTRODUCTION

At present, doctors' diagnosis of diseases is still in the traditional experience stage. Doctors' diagnosis of patients mainly depends on clinical diagnostic indicators and various examination results. At the same time, acute diseases have a sharp onset, rapid changes in the condition, severe symptoms, and the best rescue time for most acute diseases is very short, for example, the best rescue time for angina pectoris is about 4 minutes; The best rescue time of intracerebral hemorrhage is about 4 to 6 minutes; The best rescue time for airway obstruction is about 5 to 10 minutes and so on. Missing the best treatment opportunity and misdiagnosis of diseases may cause the emergency patients to suffer unbearable consequences. Therefore, increasing the research and development investment of medical intelligent diagnosis system to assist ambulance doctors in diagnosis is of great significance to ensure human health.

Expert system is a system that simulates the decision-making process, reasoning and judgment of human experts according to the knowledge and rules provided by human experts in specific fields. Therefore, it is mainly used to solve complex professional problems. However, there are still some difficult problems in most expert systems.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

Firstly, the cost of acquiring knowledge and rules by expert system is high. In order to build an effective expert system, developers and human experts often need to spend a lot of time and energy discussing and deciding whether the rules in specific fields applied by the expert system are applicable. The knowledge that expert system needs to store is very difficult to define clearly. It is also difficult for human experts to list and accurately express all the rules and truly effective knowledge applied in their work. For these reasons, knowledge acquisition has become one of the biggest bottlenecks in the construction of expert system.

Secondly, the expert system does not have the function of memory storage because of its characteristics of rules stored based on knowledge base. For example, for a medical intelligent diagnosis expert system, when inputting the symptoms of a new patient, the system will call a large number of rules for complex reasoning, and finally make the corresponding diagnosis. If other patients with the same symptoms need to be diagnosed later, the expert system will still repeat the same and complex reasoning. This leads to the excessive consumption of system resources and the low efficiency of system diagnosis, which is absolutely not suitable for the disease diagnosis of emergency patients. More seriously, the expert system can not remember and correct the mistakes it has made. Therefore, the same diagnosis process will still make the same diagnosis mistakes. Only by consuming a lot of resources and artificially modifying the rules can we avoid the recurrence of mistakes [1].

Thirdly, the robustness of expert systems is poor. The knowledge base of expert system is limited by its own stored rules. For a problem with no rules as a basis for inference and no alternative solution, the expert system cannot give any effective conclusions. Therefore, the expert system is fragile to a great extent.

Finally, due to the high complexity of the professional artificial intelligence expert system, its development often needs advanced development tools and high-level developers, resulting in high development costs, low openness and flexibility of expert system, which greatly affects the universality of its application.

In view of the above problems, starting from the data source and technology selection, and based on natural language processing technology, this paper studies and implements a medical intelligent diagnosis system which takes ES as the core to assist ambulance doctors in diagnosing patients' diseases. Compared with the traditional expert system, this system has higher efficiency and better robustness, and enhances the generalization and universality of disease diagnosis for emergency patients.

II. THEORETICAL BACKGROUND

2.1 Elasticsearch

ES is a search and analysis engine developed by Java, which has the characteristics of distributed, high expansion and high real-time. It has strong advantages in storing massive data, full-text retrieval and

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

analysis. ES takes Lucene as the core, but uses restful API to hide the complexity of Lucene. ES has good scalability and can be extended to hundreds of servers to process Pb level structured or unstructured data.

2.1.1 Data format

Compare ES storage data with MySQL storage data, as shown in Fig 1.

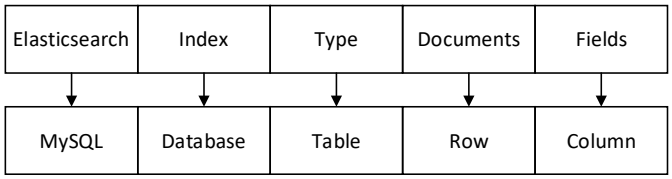

Fig 1: Comparison between es and mysql

2.1.2 Routing

A primary partition in an ES cluster is a route. ES uses the hash algorithm to store the document in the corresponding slice, so as to realize the horizontal segmentation of data. The formula is shown in equation 1.

$$shard\_num = hash(\_routing)\%num\_primary\_shards \quad (1)$$

In the formula, _routing refers to the value of the document, which defaults to the document id value. It also supports developers to customize its value in actual development, such as the document numberIn the formula, _routing refers to the value of the document, which defaults to the document id value. It also supports developers to customize its value in actual development, such as the document number and so on [2].

2.1.3 Indexing process

ES uses inverted index to store data. ES first uses the specified word splitter to segment the text, and then stores the document number, word frequency, position, offset and other information of the segmented keyword into the index database. In the index library, type is the logical partition of the index, and one type is a set of document objects in JSON format that store multiple identical fields [2]. The structure is shown in Fig 2.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

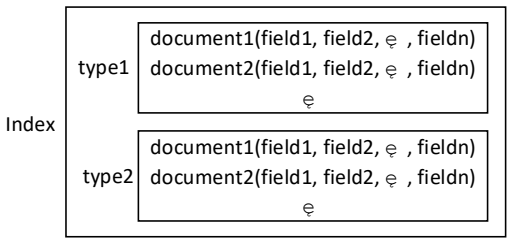

Fig 2: Es index structure

2.1.4 Search process

Any node in the ES cluster knows the storage location of each document, and can directly forward the request to the node storing the required document, and collect and return the finally retrieved document. Retrieval is divided into query stage and retrieval stage [3].

First, the client sends a request to the coordination node. Then it enters the retrieval and query stage. The coordination node broadcasts the request to all slices of the index and creates a priority queue. The requested fragments are retrieved locally and a priority queue with the size of from + size is created, and then the document ID and document similarity score in the queue are returned to the coordination node. Then, the coordination node receives and stores it in its priority queue, and performs global sorting to generate the final retrieval list.

After entering the retrieval phase, the coordination node sends a GET request to the relevant fragments according to the final search result list. After receiving the request, the relevant fragment retrieves, enriches and returns the request document. Finally, it is received by the coordination node and returned to the client. The search process is shown in Figure 3.

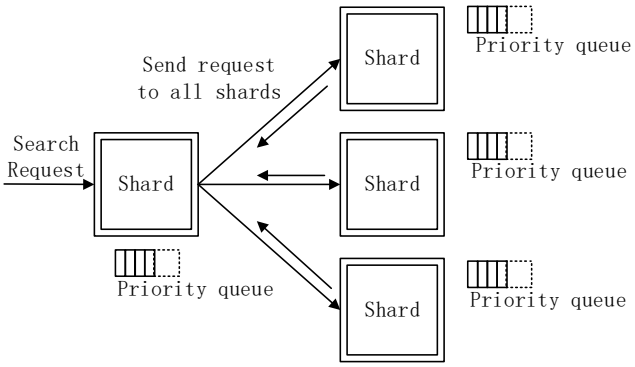

Fig 3: Search process

2.1.5 Scoring mechanism

The similarity score between documents and search keywords determines the ranking order of search results of ES. With the decrease of similarity score, documents are lower and lower in the ranking of search results. Of course, the ES scoring criteria can also be customized according to the actual production

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

environment needs and other factors. The default text similarity algorithm of ES is TF-IDF algorithm [2], and its calculation formula is shown in equation 2.

$$F(d) = coord_d(q) * queryNorm(q) * \sum_{t\ in\ q} (idf_d(t)^2 * boost_d(t) * norm_d(t)) \quad (2)$$

Where: $coord_d(q)$ indicates the number of words in query statement q in document d; queryNorm(q) represents the scale factor, which is a fixed coefficient; $tf_d(t)$ indicates the frequency of word t in query statement q in document d; $idf_d(t)$ indicates the anti-document frequency, that is, the total number of documents / the number of documents that hit the word t in the query statement q; $boost_d(t)$ represents the weight of word t in query statement q in document d; $norm_d(t)$ represents the length factor of document d, The fewer words the document contains, the larger it will be.

2.2 Elasticsearch-Analysis-Ik (IK)

In the case of repeated comparison of several Chinese word segmentation devices [4], the system uses the widely used IK word segmentation device to segment Chinese words for the symptom description of ambulance emergency patients.

When storing data, word segmentation is required first. Word segmentation refers to the separation of continuous character sequences according to certain rules and the re combination of word sequences with independent idiom meaning. This rule is determined by the built-in word breaker mechanism. Different word breakers correspond to different rules.

As the medical intelligent diagnosis system is mainly for Chinese users, but the default standard word splitter of ES can not effectively segment Chinese text, and will only split Chinese text into Chinese characters one by one. Therefore, this system adopts a Chinese word splitter developed by Chinese people, IK [5]. Of course, developers can also create custom analyzers. The word segmentation process of IK is shown in Fig 4.

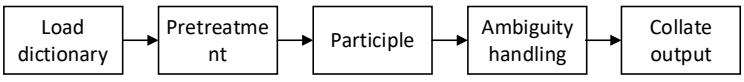

Fig 4: Text word segmentation process

IK provides two different word segmentation strategies for developers to choose, namely ik_Smart and ik_ max_ word. IK uses the forward maximum matching algorithm in natural language processing for word segmentation. The algorithm flow is shown in Fig 5.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

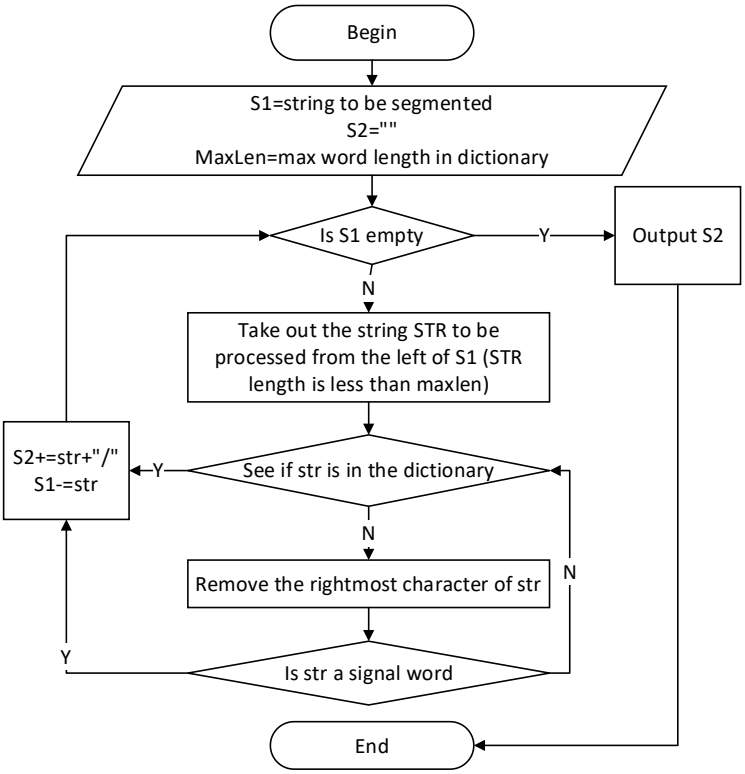

Fig 5: Maximum matching algorithm

III. RESEARCH AND IMPLEMENTATION

3.1 Research

3.1.1 TRIE and FST

TRIE is also called prefix tree and dictionary tree. Prefix tree is mainly used in large-scale string statistics, sorting and saving. The prefix tree uses space for time to reduce the number of invalid string comparisons as much as possible through the public prefix of the string, so as to reduce the time overhead. To find a string in the prefix tree is to compare the characters along the path by the root node of the prefix tree. It can be seen that the time complexity of the prefix tree is related to the length of the searched string, which is O(n). And the search operation can be carried out without entering a complete string in the prefix tree. The search in the prefix tree can not only find all strings starting with a prefix, but also find strings that have modified one or more characters. In addition, the prefix tree storing Chinese can use the hash table to store the child nodes of the prefix tree node to improve efficiency. In view of the strong advantages of prefix tree, search engines generally use prefix tree for text word frequency statistics.

FST is Finite State Transducers, which is widely used in the field of natural language processing. FST has the dictionary function similar to the map set in Java language, but its search efficiency is only

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

O(n), which is only equal to the length of the search keyword.

The TRIE and FST constructed with the words Mon, Tues and Thurs are shown in Fig 6.

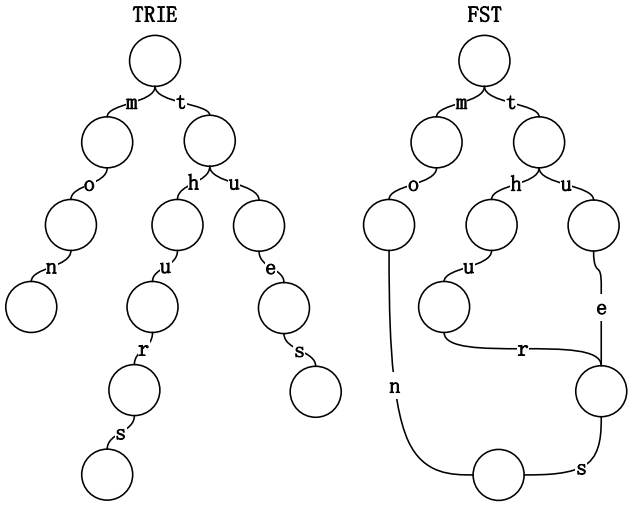

Fig 6: Comparison of prefix tree and fst

3.1.2 TF-IDF

The TF-IDF method is used to optimize the degree of similarity of words in the text. If a word appears frequently in one document and low in other documents of the corpus, it has a high degree of importance and representativeness for this article.

TF is word frequency. After word frequency statistics, the distribution of word frequency contained in a document can be observed, which can more intuitively describe the linguistic characteristics of the text. The formula is shown in equation 3.

$$\mathrm{tf}_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{3}$$

However, the general text contains a large number of characteristic words irrelevant to the main content of the text, such as prepositions, exclamations, conjunctions, punctuation marks and so on. Therefore, in terms of text classification, only using TF will greatly reduce the accuracy of document classification. In addition, if a word has a high TF value in all documents, it is not suitable as a keyword. Therefore, in order to solve the above problems and improve the accuracy of text classification, it is often necessary to comprehensively consider the word frequency and inverse document frequency.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

IDF is the inverse document frequency, which can well describe the uniqueness of a word in all documents. The formula is shown in equation 4.

$$idf_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \tag{4}$$

According to the above description, the following TF-IDF calculation formula. The formula is shown in equation 5.

$$TF - IDF = TF * IDF \tag{5}$$

The ability of a word to distinguish different texts increases with the increase of its TF-IDF value. By calculating TF-IDF value, select text words and keywords, and establish word vector. Finally, judge text similarity through cosine similarity. The formula is shown in equation 6.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|b\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{6}$$

3.1.3 Inverted index

The first element of a search engine is to provide faster and more efficient search when finding documents that meet the search criteria. Inverted index is the key core of search engine. Compared with forward index, inverted index first stores the data, then associate the key word with its corresponding document, and saves the relationship to the inverted index table. When querying, first segment the query content, then match the keywords in the inverted index table, and finally retrieve the documents corresponding to the keywords. The inverted index is shown in Fig 7.

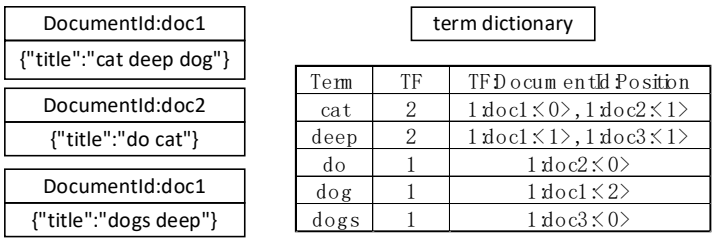

Fig 7: Schematic diagram of inverted index structure

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

3.2 Implementation

3.2.1 Data acquisition and preprocessing

Because there is a lot of structured information in 39 Health Network, this paper first collected the disease data of 39 Health Network. The data acquisition process is shown in Fig 8.

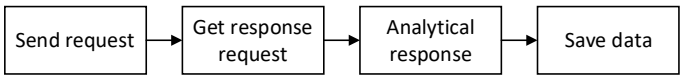

Fig 8: Data acquisition process

Although 39 Health Network has made a perfect summary of most diseases, there is still a lot of noise in the disease data of 39 Health Network. Therefore, the system only retains the disease data of emergency department required by the knowledge base through the data screening mechanism. Then clean the screened disease data. Data cleaning is one of the important prerequisites to ensure that NLP model can work correctly. The data cleaning work carried out by the system mainly completes the following items: deleting duplicate items of disease data, deleting wrong items of disease data, deleting redundant text data, checking and processing missing values of disease data, etc [6].

Finally, 411 emergency department disease data are used as the knowledge base to construct the data set.

It can be expected that with the continuous development of the system, the increase of the amount of data and the improvement of the medical knowledge base, the overall diagnosis effect of the intelligent diagnosis system will be steadily improved.

3.2.2 System diagnosis flow

The goal of this system is to create a medical intelligent diagnosis system that allows ambulance doctors to input the symptoms, disease location, vital signs and other information of emergency patients in the form of natural language text, and then the system will diagnose according to the input information and give the corresponding diagnosis information. The diagnosis process of the system is shown in Fig 9.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

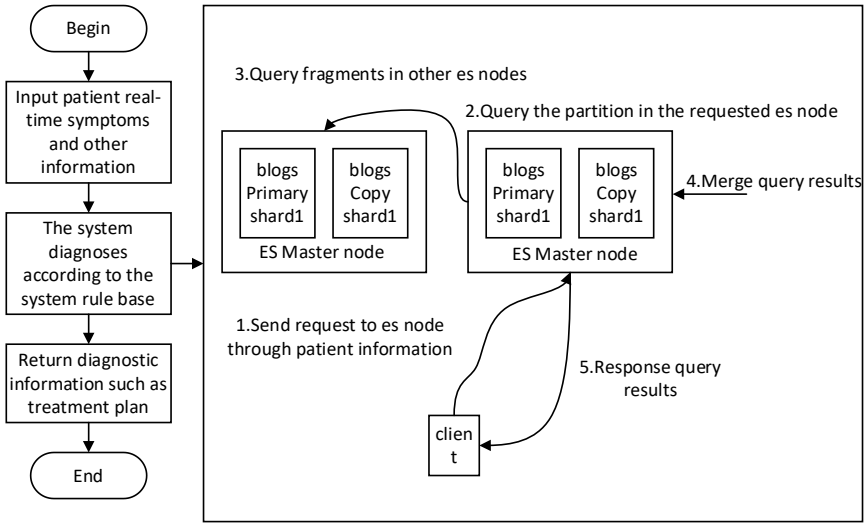

Fig 9: System diagnosis flow chart

Because the intelligent diagnosis system automatically diagnoses diseases, the diagnosis accuracy must not reach 100%. Therefore, in the auxiliary diagnosis report output by the system, the diseases most likely to be suffered by a specified number of emergency patients can be output according to the needs of users. The doctors and nurses accompanying the ambulance can refer to the diagnosis results given by the intelligent diagnosis system to further make disease diagnosis and corresponding first-aid measures, and then send the final diagnosis report to the hospital. The hospital can receive the diagnosis report of the first-aid patient before the ambulance arrives at the hospital, so as to inform the doctors of relevant departments in advance to participate in the treatment of the first-aid patient, register the first-aid patient, prepare relevant physical examination, drugs, etc, To some extent, it reduces the time for emergency patients to receive effective treatment.

IV. EXPERIMENT

4.1 Test Description

The experimental running environment is windows10 operating system, and the experimental tool is postman. The experimental data are mainly from the textbook surgical nursing and the description of acute diseases and their clinical manifestations related to 39 Health Network. 50 pieces of data are extracted for the test and compared with the correct diagnostic results. In addition, the experiment takes the system performance, system robustness and system diagnosis accuracy as the evaluation indexes.

4.2 System Performance Test

The system throughput test is shown in Fig 10 and Fig 11.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

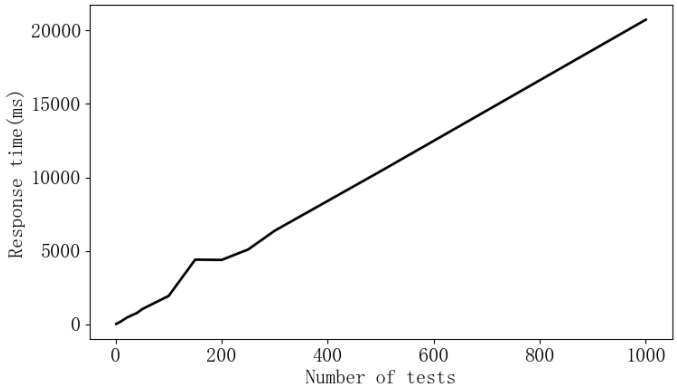

Fig 10: System performance test

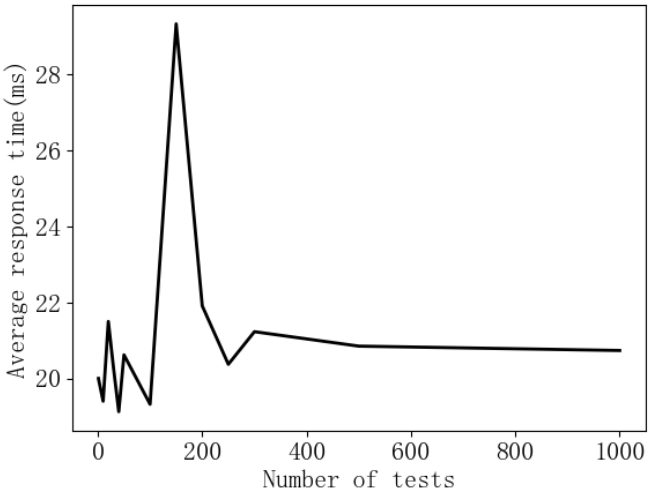

Fig 11: System performance test

After a large number of tests, the average response time of the system is 21.27ms.

4.3 Diagnostic Accuracy Test and Comparison

Due to the limitation of the content and length of the paper, the accuracy test is only one example. For example, diagnose the symptom description of "the emergency patient has nausea and vomiting, abdominal distension and reduced exhaust, accompanied by metastatic right lower abdominal pain".

According to the diagnosis results, the first three diseases with the highest probability of emergency patients are shown in TABLE I.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

TABLE I. Diagnostic result table

DISEASE	score
YERSINIA	13.082465
ACUTE APPENDICITIS	11.842348
ABDOMINAL WALL	10.407823

Note: this system can play an effective auxiliary diagnostic role. It will diagnose according to the data entered by the user and return the diagnosis results with the highest score of the first three.

According to Surgical Nursing Education [7], emergency patients have a higher probability of acute appendicitis. It can be proved that the diagnosis results of this system have a certain reference value for doctors and nurses in the ambulance to diagnose the diseases of emergency patients. The detailed description of the diagnosis result field is shown in TABLE II.

TABLE II. Diagnostic result field description table

FIELD NAME	FIELD CONTENT
DISEASE NAME	ACUTE APPENDICITIS
DISEASE ALIAS	APPENDICITIS
SITE OF ONSET	APPENDIX
INFECTIVITY	NON INFECTIOUS
MULTIPLE	ALL PEOPLE
SYMPTOM	LOWER ABDOMINAL COLIC, APPENDICEAL INFECTION, STOMACHACHE, INFLAMMATION, ABDOMINAL DISTENSION, NAUSEA AND VOMITING
COMPLICATION	PERITONITIS
REGISTRATION	GI MEDICINE SURGERY EMERGENCY DEPARTMENT
CLINICAL	DUODENAL BARIUM MEAL
COMMON DRUGS	APPENDIX XIAOYAN TABLET SENECIO TABLETS
THERAPEUTIC	SURGICAL TREATMENT AND DRUG TREATMENT

The data set is tested by method of S@n [8]. The comparison of diagnostic accuracy between the two systems is shown in TABLE III.

TABLE III. Diagnostic accuracy comparison table

	S@1	S@2	S@3
EXPERT	0.70	0.73	0.80
PAPER SYSTEM	0.66	0.75	0.83

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

According to the results in the table, when n is taken as 1, the diagnostic accuracy of the system is slightly lower than that of the traditional expert system, but when n is taken as 2 and 3, it is higher than that of the traditional expert system. In addition, the throughput, robustness and scalability of the system are better than the traditional expert system.

V. EPILOGUE

In today's society, limited by the regional economic development, the distribution of medical resources is not balanced. A doctor with perfect knowledge reserve and rich diagnosis experience often needs years of training and practice, and most of them work in hospitals in developed areas. In underdeveloped areas, the problems of lack of medical equipment, backward technology and insufficient doctor resources are more serious. In addition, more and more patients tend to go to hospitals in developed areas for medical treatment, which further exacerbates the problems of too many patients and low efficiency in hospitals in developed areas, too few patients in hospitals in underdeveloped areas, doctors' lack of practical diagnosis experience, and hospitals in underdeveloped areas can not get efficient diagnosis and treatment.

In order to popularize the achievements of the progress and development of modern computer and medical science and technology in underdeveloped areas, an intelligent diagnosis system is designed, which can not only assist doctors in diagnosis, but also provide diagnosis experience for doctors with insufficient practical diagnosis experience. It is of great significance to the country, society and individuals.

Based on the medical data of 39 Health Network, this paper establishes a medical knowledge base and realizes an intelligent medical diagnosis system, which plays a great auxiliary role in the disease diagnosis of emergency patients. The system avoids the disadvantages of the traditional expert system to a certain extent. Compared with the traditional expert system, the system has higher accuracy, flexibility and expansibility. And the system scheme design is more reasonable. After experimental verification, the diagnostic accuracy basically meets the expectation.

However, the system also has some deficiencies to be improved, including the following three points.

First of all, there are a large number of professional words and terms in the medical field, so the system will contain more unlisted words. And the system ignores the semantic information to a certain extent, which requires a large amount of data to complete these knowledge. Of course, there are ignored knowledge that cannot be completed, and ignoring the semantic information may also cause some errors.

Secondly, the system does not optimize the ES default routing hash algorithm and the sorting and scoring rules of search results.

Finally, because the system development and application are still in the initial stage, limited by the system scale and system access, considering various practical factors, the system server cluster is not

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 975-988
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

deployed. The services provided by a single server often have a certain load capacity and high concurrent processing capacity. If the performance threshold of the server is exceeded, the performance of the server will decline sharply, even the server will be down, and there will be a single point of failure, so the high availability of the system cannot be realized.

Based on the above three deficiencies, this paper puts forward the following three improvement schemes, which will be implemented in the research and development of the follow-up system.

Firstly, a perfect and professional custom thesaurus can be customized for the field of medical acute diseases to improve the efficiency and accuracy of word segmentation of medical intelligent diagnosis system, so as to further improve the performance and accuracy of medical intelligent diagnosis system.

Secondly, according to the feedback information of system users, data access frequency and other actual needs in the production environment, we can customize the routing algorithm and search result ranking and scoring rules that meet the actual needs.

Finally, according to the requirements of the production environment, the system is deployed in the server cluster rather than a single server to provide services.

ACKNOWLEDGEMENTS

This research was supported by National Natural Science Foundation of China (Grant No. 72174079, No. 1210050123, No.72101045).

REFERENCES

[1] Nie Nie. Remote intelligent medical auxiliary diagnosis system based on Internet. Xi'an University of Electronic Science and technology,2001.
[2] Hu Xin, Yao Yu, Xu Yingjie. Design and implementation of TEE case database retrieval system based on elasticsearch. Computer application,2018,38(S1):91-94.
[3] Rafał Kuć Marek Rogoziński. ElasticsearchServer development Server development. Translated by Cai Jianbin Translated by Cai Jianbin. Version 2 Beijing: Published by people's Posts and Telecommunications Publishing House, 2015.138-140
[4] Meng Bangjie, Wang zhangang Implementation and comparison of two Chinese word segmentation algorithms on cloud computing platform Network security technology and application, 2014 (12): 2
[5] IK Analysis for ElasticSearch. [2014-03-14]. https://github.com/medcl/elasticsearch-analysis-ik
[6] Ryan Mitchell. Python web crawler authoritative guide. Shenfan Xiaobao, translation. Version 2 Beijing: Published by people's Posts and Telecommunications Publishing House Shenfan Xiaobao, translation. Version 2 Beijing: Published by people's Posts and Telecommunications Publishing House, 2019.127-132
[7] Li Lezhi, the road is shallow Surgical nursing. 6th Edition Beijing: People's Health Publishing House, 2017.486-487
[8] Deng Hongli, Yang Tao, Shao Chenxi An intelligent expert system for evaluating the reliability of simulation Computer simulation, 2011,28 (8): 90-93