Analysis and Design of Medical Information Search System

Lin Ye[1,a], Hui Cao[2,b,*],Xiang Zhang[2,c] and Haojie Zhang[2, d]

[1,2]Northwest Minzu University, Key Laboratory of Chinese Language and Information Technology of Ministry of Education, Lanzhou city, Gansu province, China's

[a]574140345@qq.com,[b]147625251@qq.com,[c]1006107774@qq.com,[d]547301347@qq.com

Abstract: At present, the development of information technology is extremely rapid, the original library collection has been unable to meet the massive medical information storage, and began to gradually become a network resource storage.It's more convenient and fast for people to find their own information.This paper designs a search system based on B/S structure for medical information,combining with the present search engine features and on the current common search engine to explore the working principle of the study,analyzing and design a web-based medical information retrieval system.The system is designed to be divided into four processes:the first part is the web crawl and analysis module, that is the web crawler;second, the establishment of medical information database;third, web index database management module;last part is user interface function module.

Keywords: Medical information, Search, Spider.

1. Introduction

Nowadays, due to the continuous reform and innovation of new technologies,China in all aspects and the industry's information services have made considerable progress.Information search has gradually become an indispensable part of our daily life.Through the search engine, we can easily access the relevant information.However, when you need to retrieve a particular piece of information,you will seem powerless, can not find the results that you want,which reflects the depth of some search engine digging enough.For a specific professional search without a suitable search engine, can not reflect the professional unique vocabulary and the corresponding index.Now proposed a new search engine service model that is "vertical search engine",through a specific area, a specific population or a specific demand to provide a certain value of information and related services[1-4].Therefore, how to quickly and accurately query the information needed by the Internet on such a large amount of data has become the focus of attention.The openness and universality of the Internet, so that anyone can publish their own on the network want to release the message.Today, most users need a domestic easy to query their own symptoms of the search system,we can accurately and quickly query the relevant drug information required,keeping abreast of their physical condition, and further help users to relieve symptoms.

2. The analysis of System

2.1 The web crawl and analysis module(web crawler)

The crawling of related literature information is done by calling the method.But the document crawls content that is different from other pages,only need to enter a specific search keyword,click the specified content and then uses a width-first traversal strategy.By inserting the link in the new download page directly into the tail of the queue waiting to be crawled.It is the same as the web crawler, firstly, crawling all the pages that are linked in the starting page, and then selecting one of the links to continue crawling all the pages in the link.When the crawler program scans the current page,

it will be based on the development of a good strategy to access the list of the next hyperlink address.The web crawl module links the web page and downloads the page HTML code [5].

After the crawling is completed, use the analysis module to analyze the web code and extract the web information associated with them, such as the web site URL, the content and the hyperlinks to the page, and so on. Then put the relevant information together Web index database.For figure 1:
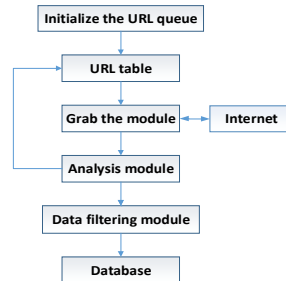

Fig.1 Web crawler structure

2.2 Establish a medical information database

The network crawler after the data for the corresponding data cleaning and data normalization,establishing corresponding data attribute table, respectively, the definition of each category of data tables, and then deposited into the medical information database.Five types of data are defined in the database,they are disease name, symptom presentation, solution, complication, and susceptible population.When the user interface has a search request,it can from the database to match the corresponding keyword to return the query results.Keywords match the standard is not the same, then the accuracy of the results returned will be different.Table 1 below shows the storage structure of the data content of the system part.

Table 1 Partial data storage structure

Attribute name	Style	Allow
Disease_name	varchar	Yes
symptoms	varchar	Yes
solution	varchar	Yes
complication	varchar	Yes
Susceptible_population	varchar	Yes

2.3 Web Index Database Management Module

This article uses the Lucene open source toolkit.It's core package to achieve the search function, highlight and other auxiliary functions by the peripheral function package to achieve.Index management package to achieve the establishment and deletion of the index.Retrieve the management package, as the name suggests, it is based on the conditions to be retrieved to retrieve the results. And maintain the data in the index database, including the data editing and deletion and other operations.Index database design novel,which uses a divide and rule of the method, because the storage capacity can improve efficiency, so the larger storage of the table is divided into a number of smaller storage table.When a certain range of search queries, you can improve the efficiency of search engines.

2.4 User interface function module

To provide users with web search request interface, according to the user's search requirements, in the local index database to quickly retrieve data in the web interface to display the results required by the user.Provide two search functions, a fuzzy search function, a precise search function.

3. System implementation

3.1 Web search and analysis function module

The main function of the function module is to crawl the web page, analyze and filter the web data, and store the data in the database.The main implementation classes are the HTML Parser class and the Analyzer class,using the HTML Parser class to parse and extract web data, and some fields use regular expressions, and the Analyzer class is used to parse the text data.

3.1.1 Web crawl

Place the link that has been downloaded on the page directly into the tail of the URL that is waiting to be crawled.In fact, this process is to start the page link all the pages are one by one crawl, followed by one of the selected link page, continue to crawl the link page link all the pages.The flow chart of the crawl is shown in Figure 2.
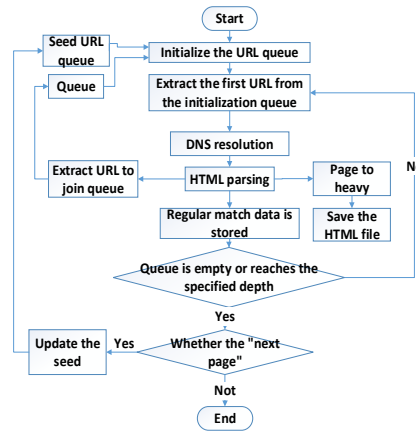

Fig. 2 Grab the flow chart

3.1.2 Web page data analysis

Get the HTML source code of the page, according to the regular expression to match the specified string, mainly to obtain the title of the page, web content and page hyperlinks [6], [8].
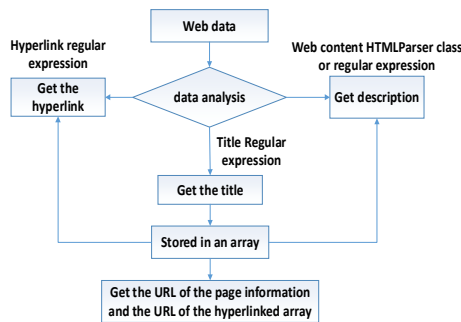

Fig. 3 Web page data flow chart

3.1.3 Data filtering and storage into the database

Crawl the page, get the title of the page, related content and hyperlinks to the page, then crawl the page hyperlinks and temporarily stored in the URL data list.While the current URL data selection page title and the original user specified string to match,if they match exactly, the keywords in the page link, the page uniform resource locator, and the page title are stored in the database,and create an index database,otherwise cleared to match the unsuccessful data, continue to analyze the data list in the next URL.

3.2 User search function module

The role of the module is based on the user to find the field and the input keywords to retrieve the database, and then return the corresponding information and display to the user.In addition users in finding information when you are prompted to enter a keyword or the possibility of a full complement of high frequency high query query hint words.When the prompt does not meet the needs of users can change a search word will be able to get the desired information.Usually it is based on the user's current query and gives the relevant tips, this method is collaborative filtering in the search on an application.In the relevant documents, the user needs to query and display the results of a similar document,if the results of the query show too much, then the user from the current results again to enter the words to be asked to narrow the scope of the query.

3.3 Database background management module

The database background management module function is to manage the index database, including the database list of data search,add, edit and delete and other basic operations.Because the amount of data in the database table is very large, creating indexes on tables helps to improve the search efficiency of users.The index is established on the basis of raw data tables supporting documentation,

the purpose is to the database of large amounts of data, it can quickly find the required data and sorting and other needs.Through it can sort the data table, quickly find that we need.There is another attribute constraint in the data table,after the establishment of the table, according to the information provided by the user to do some input data constraints,the system checks to see if these constraints are met when new data is entered.In a variety of attribute constraints and collections, the operation of the database will become very simple.

4. Conclusions

Compared with the previous printed version of the book, electronic information more convenient, so the modern medical information is mainly derived from the Internet.It is displayed in a digitized style, usually with text, images, etc,and stored on some media of the disk disk and then displayed by some electronic devices.Network resources have the characteristics of fast propagation speed, wide range and huge quantity and information update. Therefore, network information becomes an important way for medical personnel and researchers to obtain information.But some of the current medical search engine to collect information on the content is incomplete, the quality is also different, resulting in users difficult to find the information they want.So choose a suitable search engine is particularly important, while the establishment of appropriate medical information search engine for users to use far-reaching significance.In this paper, the use of open source search engine framework Lucene and with the corresponding medical information management module,analysis and design of a practical and effective search system,used for a patient through the user interface in the self-examination of the disease or drugs and provide other relevant information,allowing users to search at the same time to understand the possible complications, so that a more comprehensive disease and the corresponding drug back to the user interface,then the relevant web information output, to help people better understand their health status, good protection measures, and further help people to promote good health.

5. Acknowledgments

6. References

[1] Huilin LIU,Lai-de GUO,Lan-zhe LIU,Guang-xing WANG et al.Design and Implementation of Chinese Agriculture Theme Search Engine [J]. Journal of Zhengzhou University (Natural Science Edition), 2007,39 (2): 74-77.
[2] Xiao-lin WANG,Hong-shen LIU, et al.Design of search engine [J]. Computer Technology and Development, 2007, 17 (2): 5 - 7.
[3] Ying ZHAO,Yao XU,Jian WANG,Yingying ZHANG et al.Design and Implementation of Professional Search Engine System [J]. Microcomputer Information, 2007, 23 (2): 179-181.
[4] Design and Implementation of the Subject Search Engine in Military Medicine [J]. Journal of Information Technology, 2007, (9): 122-123.
[5] Zhi-dong SUN,Mao PAN, Zhi-xin SUN,Qiu-yan YAN et al. Design and implementation of lightweight adaptive search engine [J]. Journal of Guangxi Normal University, 2007, 25 (2): 210-213.
[6] [Prof. Fred (Friedl, J.E.F.), Yu Sheng translation. "Proficient in Regular Expressions (3rd Edition)" [M]. Electronic Industry Press, July 2007.
[7] Mike Cafarella . Doug Cutting Building Nutch .open source search a case study in writing an open source search engine [J].ACM Queue ,2006 ,2 (2) :54-61.

[8] Menczer F. complementing search engines with online Web mining agents[J].Decision Support Systems, 2006 ,35(2):195-212.