Advanced Science and Technology Letters
Vol.134 (MAS 2016), pp.1-7
http://dx.doi.org/10.14257/astl.2016.134.01

A Java based Mobile Network Search Engine Design Research

[1]Guang Zheng, [2]Yiran Wang

[1]School of Computer Science and Technology, Zhoukou Normal University, Henan Zhoukou, 466001, China
zhengguang@zknu.edu.cn
[2]College of Network Engineering, Zhoukou Normal University, Henan Zhoukou, 466001, China
286206308@qq.com

Abstract. With the rapid development of Internet technology, people are increasingly apparent to the requirement of information filtering.This makes search engine technology to more accurate, more quickly, more convenient in the direction of development.In the early stage of the market, on the basis of demand analysis, the application of the vertical search engine thought of more specialized in mobile network, using the Java language to realize its function in all kinds of booking. The engine can satisfy the specific requirements of users for mobile Internet search and more practical value.

Keywords: Search engine; Java; The Lucene index

1 Introduction

In the 21st century, with the scale and sustained development of the Internet have multiplied the information resources. People's demand to the Internet from an initial conversation, listening to music, and other entertainment activities also slowly into a more practical food, clothing, shelter, transportation, medical, job search, shopping and so on all aspects of life. Due to the vertical search using artificial intelligent strategies such as classification, professional knowledge, so it better than the above mentioned three generation of search engines will be more effective and accurate, we will be the perfect vertical search engine called the fourth generation of search engine [1-2].

2 Related works

2.1 Vertical search engine

Vertical search engine is a relatively large amount of information of general search engine, the query is not accurate, such as depth of the new search engine service

ISSN: 2287-1233 ASTL

mode, by targeting a specific area, a particular group or a particular demand have provided valuable information and related services. Compared with huge amounts of information disordering of general search engine, vertical search engine is more focused, specific and thorough. Vertical search engine to provide users with not hundreds or even thousands of relevant web pages, but the range is very narrow, highly targeted specific information. Compared with the general web search engine, electronics and other industries of search engine technology, no matter from the orientation, content, users, or market strategy, etc., there are some differences. Electronics industry, in other words, the search engine, users demand more clearly, the search range is small, the corresponding information should be more detailed, which must be able to provide differentiated services to the customer. According to the specific industry requirements and different theme, appear this kind of at a new search technology for service pattern search service system, is for the vertical search engine [3-4].

2.2 Java technology is introduced

Java is a can write cross-platform applications of object oriented programming language, is by the Sim company was launched in May 1995.The Java programming language and Java platform is divided into three versions, namely the Java SE (standard), Java EE (enterprise) and Java ME (small).Java technology has the remarkable versatility, high efficiency, portability and security platform, widely used in personal PC, data center, game consoles, science super computers, mobile phones and the Internet, at the same time has the world's largest professional developer community. In the global cloud computing and mobile network each industry environment, Java is more has the significant advantages and broad prospects [5].

JSP page logic and Web page design and display of separation, support reusable component-based design, make the development of a web-based application quickly and easily. And, in fact, at the bottom of the container JSP page at run time will be compiled into a Servlet for processing again, finally to feedback the information to the user in the browser to see [6-8].

3 Based on the Lucene index and search

Lucene is a completely open source full-text retrieval tool kit. Lucene is to use Java development in the initial stage. But because of its powerful function, gradually has been translated into many languages. The Java Lucene is a high-performance full-text retrieval toolkit, it USES the index structure is inverted file. Indexing is an important step of search engines work. In this paper with the aid of Lucene kit has two of the most important concepts in the Document and Field domain logic file. They correspond to the Document in the Lucene classes and class Field. The meaning of the Document for the Document, it represents a kind of logical file. Lucene itself cannot be indexed to the physical file, but to recognize and deal with the Document type of Document. In some cases can be corresponding to each Document with a physical

file, use a Document as a substitute for a physical file; And more, the Document has nothing to do with physical file, it is as a collection of data sources, to provide the original to Lucene to index text content. Lucene from Document related data source, and corresponding processing according to the configuration properties. For example, when a Document with a physical file corresponding to the up, can extract a variety of data sources, such as file name, file content, file creation time, modified time, etc. As shown in figure 1, can also be extracted from different physical files to the data source, in the same Document.

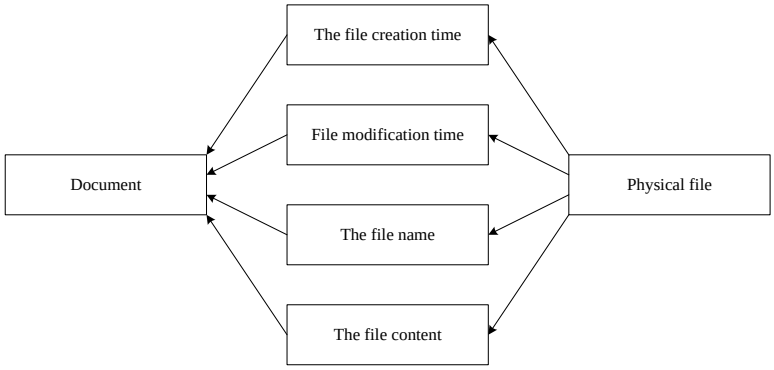

Fig. 1. Document and multiple data sources

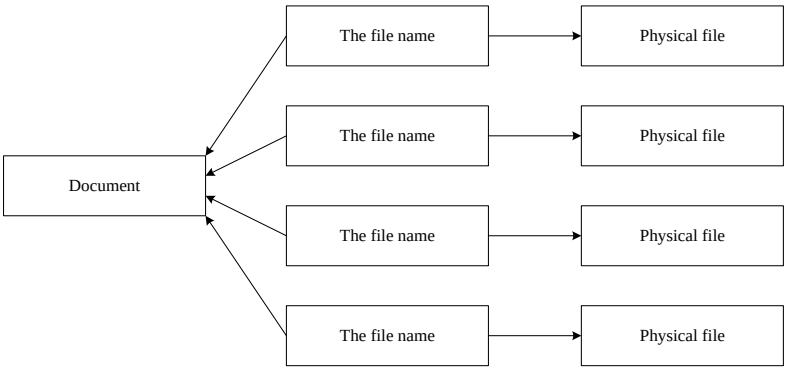

Fig. 2. Document and multiple files of data sources

As shown in figure 2 will all of the files in a directory file name in a Document, let the Lucene index for it. Because the Document is only responsible for collecting data and, therefore, cannot use physical file to build a Document, a text, Numbers, and even some links are building a Document as a data source. As long as they are added to the Document object, Lucene can build indexes for these data sources, and users can find them [9-10].

4 Vertical search engine design based on Java

4.1 The basic functional requirements

In the domestic each big mobile phone portal and the needs of customers on the basis of careful analysis, intends to build a vertical search engine function of mobile phone information query system, providing customers with a search according to all kinds of mobile information platform. Allows customers to easily get what they pay attention to details of one mobile phone. Specifically, the information service system should possess the following two important functions.

Cell phone information automatic acquisition is one of the core functions of website, main purpose is to use the current vertical search engine technology, selection, access to all kinds of mobile phone information related web pages, and put those pages downloaded to the local image stored. And then analyze the web structure, the use of structured information extraction technology, extract the structural information of cell phone information, such as mobile phone brand, model, price, origin, time to market, mobile phone shape, the size of the home screen, the home screen material, main function, etc., deposited in the database, for users to query and retrieve [11].

4.2 The overall structure of the system

This mobile phone product vertical search engine USES Eclipse3.1 + MySQL5.0 + Tomcat5.5 development environment. In addition, this system has good portability, as long as to master the principle, can be in very small changes developed in the field of any other vertical search engine system. For example can easily build up to the notebook computer, MP3, digital cameras and other vertical search engine system. Figure 3 is the system running flow chart:

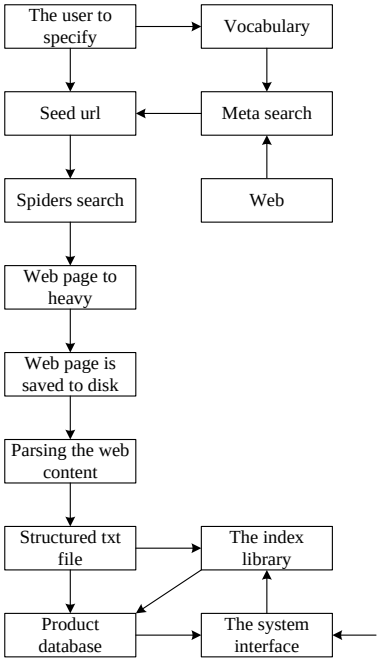

Fig. 3. Flow chart of system running

For the scraping of the information of the mobile web part, in order to make information has the certain extent and comprehensiveness, in principle by meta search engine to obtain the best seeds site, but this article focuses on the construction of a vertical search engine, in order to easy to switch to the user specified directly seed sites to crawl for Heritrix spider program starting site. In fact, as long as we choose several domestic famous mobile website, the information on the covers almost all common cell phone information [12-15].

4.3 The establishment of the index and the database design

The preparation is complete, this part is relatively simple. Mainly is to write programs to read each product under the specified file path corresponding text files, each corresponding field data into the database, and then returns the actual order date Id to Lucene index, thus the database record Id and match them with the Lucene index. Insert data to database and indexed is synchronous. When indexing, should first define the Lucene Document format, the product all the information encapsulated into a logical Document. Index of the content should be less as far as possible, as long as can satisfy user retrieval. This article focuses on a kind of vertical search engines, so strive to specific application data of the simple and clear. System is realized by entering a product brand or product model to retrieve the relevant details of the product function, USES only a data table to store all the product information. The database table structures of the Products are shown in table 1 below:

Table 1. Database table structure design

Field name	Data type	Meaning
Id	Int	A primary key
Category	Varchar(128)	Classification
Name	Varchar(128)	Model
Type	Varchar(128)	Detailed parameters
Content	Varchar(5000)	Product Summary
URL	Varchar(1024)	Product information index page
Imageurl	Varchar(1024)	Product pictures stored path

5 Conclusion

Through and general search engines and comparative analysis on the key technology in information service, points out that the vertical search engine prospects and development space. Introduces the fulltext retrieval tools currently the most widely used Lucene source and frame structure, and its index and search functionality provided by the inverted index is very important in the technology. Proposed to build a product idea and realization method of the vertical search engine, the system is introduced in the system design process, and regulation system to complete the function, and analyzes its feasibility and its technical route.

References

1. Wang, C. Y., Yu-Fu, L. I.: Study of Information Filtering Technology in a Vertical Search Engine [J]. Information Science, 2014.
2. Wang-Wei.: The Design and Implementation of Computer Vertical Search Engine[C]// Measuring Technology and Mechatronics Automation (ICMTMA), 2015 Seventh International Conference on. IEEE, 2015.
3. Wang-Wei. The Design and Implementation of Computer Vertical Search Engine[C]// Measuring Technology and Mechatronics Automation (ICMTMA), 2015 Seventh International Conference on. IEEE, 2015.
4. Zhang, QY., Hui-Hui, YU., Chen, YY., Construction of Chinese word segmentation dictionary based on agricultural vertical search engine [J]. Guangdong Agricultural Sciences, 2015.
5. Qian, Y.J., Cao, B.X.: An Improved Ranking Algorithms Based on Vertical Search Engine [J]. Electronic Technology, 2015.
6. Wei, R., Zhenqiang, W.U.: Design and implementation of vertical search engine for education video resources [J]. 计算机工程与应用, 2014.
7. Zhou, Q.Y., Jing-Feng, H. U., Li-Li, H.E.: Study and Application of Vertical Search Engine Based on the Depth Mining of Enterprises [J]. Computer Programming Skills & Maintenance, 2014.

8. Yue, H.G., Zhang, L., Meng FJ.: Research and Implementation of a Vertical Search Engine in the Financial Domain[J]. International Journal of u- and e- Service, Science and Technology, 2014, 7.
9. Yue, S., Wanlong, L.I., Wang, L.: Database Full-Text Retrieval Based on Lucene Index [J]. Journal of Jilin University, 2014, 52(05):995-1000.
10. Mathew, A.B., Pattnaik, P., Madhu Kumar, S. D.: Efficient information retrieval using Lucene, LIndex and HIndex in Hadoop[C]// Computer Systems and Applications (AICCSA), 2014 IEEE/ACS 11th International Conference on. IEEE, 2014.
11. Diesendruck, L., Kooper, R., Marini, L., Using Lucene to index and search the digitized 1940 US Census [J]. Concurrency & Computation Practice & Experience, 2014, 26(13):2167–2177.
12. Shi, X., Wang, Z.: An Optimized Full-Text Retrieval System Based on Lucene in Oracle Database[C]// Enterprise Systems Conference (ES), 2014. IEEE, 2014:61-65.
13. Shi, X., Wang, Z.: An Optimized Full-Text Retrieval System Based on Lucene in Oracle Database[C]// Enterprise Systems Conference (ES), 2014. IEEE, 2014:61-65.
14. Wang, X.R., Zheng, Q.H., Fa-Mei, H.E.: Research and implementation of desktop search engine based on Tika and Lucene [J]. Computer Engineering & Design, 2014.
15. Guo, Y., Lu, Y.: Research and Application of Full-text Retrieval Technology for Document Based on Lucene [J]. Microcomputer Applications, 2014.

