THE RESEARCH OF VERTICAL SEARCH ENGINE FOR AGRICULTURE

Weiying Li, Yan Zhao [*] , Bo Liu, Qiang Li

College of Information Science and Technology, Agriculture University of HeBei, Baoding 071001,China

[*] *Corresponding author, Address: College of Information Science and Technology, Agriculture University of HeBei, Baoding 071001,China, Tel: +86-312-7526424, Email: leeweiying@163.com*

Abstract: Following rapid expansion of huge Agriculture information body on the Web, the efficient Agriculture information gathering on specified top becomes more and more important in search engine research. Through the statement of the developing trend of search engine and sharing agriculture information resource, this paper discusses the necessity of building search engine for agriculture information. The author clarifies the working principles of professional search engine for agriculture and finally analyses the improvement of searching technique of agriculture and proposes a model for agriculture - focused search.

Keywords: search engine, agriculture

1.   INTRODUCTION

The internet in China has enormously promoted the development of agriculture. The amount of agriculture information networks in China has reached to huge numbers, which has shown a marked efficacy in terms of agriculture information communication and technological achievements communication. With the substantial increase of the information resource in the network, the phenomenon of information overload has been paid attention. It has become a problem that needs urgent solution how to obtain the web page that contains the information the user needs in an effective and accurate way.

Using the Vertical Search Engine is the most effective way to settle this problem. The vertical search engine directed towards agriculture searches the agriculture information of the website appropriately to make it possible that the information can be searched efficiently. With the exponential increasing of the information on Internet, there is a great deal of information waste. It becomes an important side that to provide the high quality and a modest number of query result.

2.   CURRENT STATUS AND DEVELOPMENT TREND

Most of the Vertical Search Engines are lying in the phase of scientific research. The portals have appeared that faces some fields after making use of the searched result and after the professional person's processing. The research about the subject-based searching engine is getting hotter overseas, but in our country it is at the first time step.

The actual Vertical Search Engine adopts two kinds of technologies list below:

One is based upon the content which is the extension of the traditional information retrieval technologies. Its main way is to establish a word list in connection with the subject in the Search Engine. The crawler of Search Engine makes index from web according to the word list in the Search Engine. The complexity of establishing the word list is quite differently according to the different systems.

The other is based upon the analysis of interlink. Some scholars consider that the interlink age between websites is very similar as the traditional citation Index. The relationship between websites can be found out through analyzing the interlink age. The lots of Web pages can be easily classified according to the relationship of quoting because the details of websites that are referenced and references are interrelated in content.

3.   KEY PROBLEMS IN TECHNOLOGY

The Vertical Search Engine that faces agriculture has its property. The four key technologies of it are listed below.

 (1) The target-oriented, high real-time and manageable technique for webpage collecting：The Vertical Search Engine that faces agriculture has its professional requirement and target. It only collects the websites of partial source. The quantity of web pages is moderate. But the all sidedness and depth of collecting is demanding and the priority of collecting the dynamic web pages is highly required. The technologies of website collecting should

control the objective and range; support the deep collecting and the complicated dynamic web pages collecting according to needs in practice. The more target-oriented, Real-timing and more inclined of management of the collecting technology is needed, as well as the shorter cycle of information refreshing and the prompter acquiring for message.

(2) The web page analyzing technology of structured data：Because of the particularity of the Vertical Search Engine that faces agriculture, the period, source, and the other metadata are needed as well as the specifically content in web pages. For the purpose of providing preferable and more valuable searching service in the Vertical Search Engine, it demands that the author, the theme, the region, the name of the institute and product and the specialized shoptalk must be extracted.

(3) full-text index and combined retrieval technology：Because of the higher requirements in the professional information and useful value, the Vertical Search Engine that faces agriculture can support full-text searching and precise searching and can provide many different ways of sorting. In addition, the combined retrieval which is structured or unstructured should be support according to need, such as combined retrieval based on the author, the content and the classification.

(4) Intelligent zed text mining technology：The Vertical Search Engine that faces agriculture takes the structured data as the basic components. The Search Engine can furnish the more valuable service more accurately based on the integrating of structured data and full-text data. The entire structured information extracting goes through the process from analyzing to handling. In accordance with the above requirements, the Vertical Search Engine has the function of intelligent processing, such as automatic classification, automatic indexing, automatic decomposition and text mining. They are leading technology in the areas of Vertical Search and message processing.

4.   DESIGN PROPOSAL

4.1   Technology pathway

(1)According to the characteristic of agricultural information's distributing and the practical requirements of the customer and based on thorough investigation, after understanding and comparing of the significant addition in the same field, put forward the overall framework of the platform.

(2)Specify the platform further combining the object-oriented technology to determine the specific pattern of the project.

(3)In connection with the concrete issue that is faced with, such as how to increase the crawl speed, the system's resource constraints, the precision of

web page classification, the resolve of HTML file, provide effectual solutions.

(4) Reduction to practice and form a platform faced agriculture for information searching and sharing. Measure and test the multiple parameters further, improve and optimize them to form a user-friendly Search Engine, which meets the user's requirement in the response speed, recall ratio and precision ratio.

4.2   System architecture
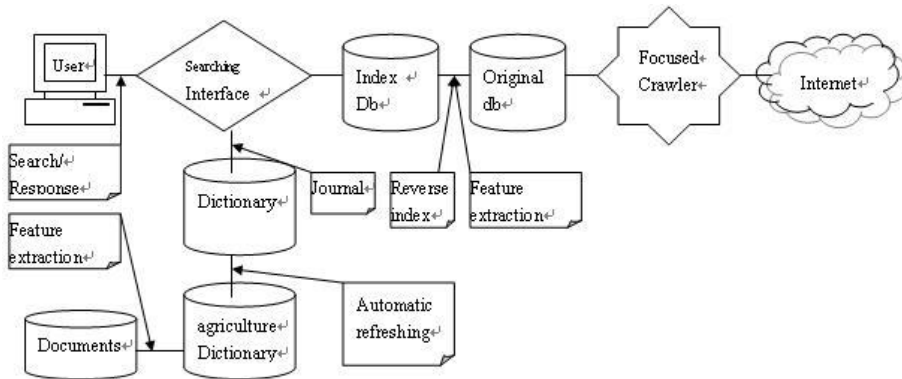
The construction drawing is list below:

Figure. 1. The construction drawing of the Vertical Search Engine that faces agriculture

4.3   Technical innovations

(1)The system has the modular construction and the coupling is relaxed. It makes reference to the characteristic of Internet search engine and takes the file manipulation as an individual module.

(2)The system is developed by JAVA Language, so some code can be reuse availably.

(3)The object-oriented ideas are drawn into during the design process which is convenient for the developer to exchange with each other. The code is reconstructed ceaselessly during the coding session, which makes it high efficiency and its readability is raised too.

(4) The XML document is used to transit information between modules when the system is integrated.

(5) Many hash tables are used to raise the system's performance.

5. CONCLUSIONS

The primary focus about search engine has varies from how to find more information to how to find the exact and useful information. The precision ratio becomes the first priority for many of the search engine. The Vertical Search Engine that faces agriculture seeks the thematic information appropriately to make the user search what they want effectively. It can affect the development of search engine deeply.

REFERENCES

Chen H. Machine learning for information retrieval: neural networks, symbolic learning, and genetic algorithms [J].J. Am. Soc. Inf Sci. . 1995.46(3): 194-216

WANG Zheng, WANG Qing, WANG Ding-wei.Design of Domain-specific Search Model for Meta-search Engine on Internet, JOURNAL OF SYSTEM SIMULATION, 2008, 20(5)