Research on the Application of the Vertical Search Engine based on Ontology in the Library Resources

Zhang Dongwei*
Information & Control Engineering Faculty
Shenyang Jianzhu University
Shenyang, china
e-mail: arnold0110@sina.com
* Corresponding Author

Kan Fenglong
Information & Control Engineering Faculty
Shenyang Jianzhu University
Shenyang, china

Mao Yongming
Information & Control Engineering Faculty
Shenyang Jianzhu University
Shenyang, china

Qi Ning
Library
Shenyang Jianzhu University
Shenyang, china

Wang Xin
School of Information & Control Engineering
Shenyang Jianzhu University
Shenyang, China

Wang Changtao
Information & Control Engineering Faculty
Shenyang Jianzhu University
Shenyang, china

Chen Nan
Information & Control Engineering Faculty
Shenyang Jianzhu University
Shenyang, china

Wang Bin
Northeastern University at Qinhuangdao
Qinhuangdao, China

Abstract—In order to solve the problems of the traditional search engines in terms of accuracy, the paper puts forward the concept of vertical search engine based on ontology, and using it in the library resources. The paper firstly describes the development of vertical search engines situation, characteristics and technical principles. Then we analyze the ontology technology, advantages and in the role of vertical search engines. We propose the basic model of vertical search engines based on ontology in the library resources. And on this basis, we give the ontology-based information acquisition strategy, ontology-based information extraction strategy, ontology-based retrieval and its implementation algorithm. Finally, we utilize the ontology in the vertical search engine application, and the experimental effect is very ideal.

Keywords—Ontology; Vertical Search; Library Resources; Search Enginge; Sphere

I. INTRODUCTION

With the rapid development of network technology, the internet has become the mainly channel to obtain the information. Number of the books on the internet resources is growing rapidly. Traditional search engine has good performance in the number of accessing information, but the quality of the information is not very ideal. Compared with the general search engine, vertical search for a particular field, a particular crowd or a specific needs to provide a certain value of information and related services, and ontology is the abstraction and description of the concept of domain knowledge. Based on ontology, semantic search engine is the developing direction of the future search engine. Search engine design based on ontology's ultimate goal is to make computer has the intelligence of people. The current goal is to make the computer return results more targeted and accuracy[1].

II. VERTICAL SEARCH ENGINE

Traditional search engines have a large amount of information, however the query is not accurate, depth and other issues. Now people put forward a new kind of search engine service mode by targeting specific area, specific vertical crowd, particular demand have provided valuable information and related services. This search engine is vertical search engine[2].

A. The characteristics of the search engine

Vertical search engine is designed for querying specific information in the field of information search tools[3]. Vertical search engine has the following features。

1) Vertical search engine contains field range is small, and the amount of information is relatively small, so that by the method of expert classification indexing to organize the information collected, it can further improved the quality of information. So we can set up a high quality, complete vertical information included we can real-time update of index database.

2) Vertical search engine only involves one or a few specific areas, so that we can reduce the vocabulary and expressions of the probability of polysemy, and we can

take advantage of the captioned table to regulate and control, so as to improve the recall of information search.

3) Vertical search engine's accuracy is high. It can hire professionals to deal with the user feature retrieval. It can provide online consultation and online tutorials, so as to improve the clarity and precision of the query, and the query result accuracy is greatly increased.

The size of the database and index database is small. So it can reduce the corresponding query time, also it can be used in a complex query syntax, to improve the accuracy of the user's query.

B. *Principle of vertical search engine web crawler*

Vertical search engine compared with the traditional search engine on the web page collection technology are quite different. Because the vertical search engine and subject is closely related, so the algorithm and the working process of the crawler is relatively complex[4].

Vertical search engine need according to certain web filtering analysis algorithm to deal with the theme of links, and retain the useful links to release the URL of the queue. It will then according to certain search strategy choice to grab the next web URL from the queue, and repeat the process, until it reaches system a certain conditions when it is stop. All spiders crawling web pages will be storage system, certain analysis, filtering, and indexing, so that after the query and retrieval. For professional web crawler, the process of the results on the analysis to the following process may also give feedback and guidance.

III.   THE ONTOLOGY TECHNOLOGY

A. *Definition of the ontology*

In the field of computer application, ontology is the conceptual object clear and description. From the perspective of the definition of conceptual object, the term should be a domain ontology semantic network including the basic information. Ontology is the domain concepts clear Shared formalized description of the model. To describe things of the concept is abstract, concept can have their own attributes. The concept of ontology is objective existence as well as the relationship between concepts[5].

B.   *Use ontology to describe the world*

Ontology can explain the conceptual object clear and description. In general ontology using the following main concepts to describe the content.

1) Classes. Classes is any transaction, such as job description, function, behavior, strategies and reasoning process. In the semantically speaking, it represents a collection of objects, the definition generally adopts frame structure, including the name of the concept, the relationship between other concepts of collection, as well as the description of the concept in nature language.

2) Relations. Relations is the interaction between concepts in the field, formally defined as a subset of the set: $R{:}C1 \times C2 \times ... \times Cn$ 。 On the semantic relationship between corresponding to the object of collection.

3) Function. Function is a special kind of relationship. The formal definition for $F{:}C1 \times C2 \times ... \times Cn\text{-}1 \rightarrow Cn$ 。

4) Instances. Instances represent elements that the semantic representation is object[6].

From the objective point, the description of the ontology is not relevant to the specific task and purpose, but there is always a certain task background. Ontology knowledge described for the choice is associated with a specific task. It is also a different top ontology [7].

C. *The advantages of ontology*

Given the nature of the ontology itself and its application research to explore the ontology construction is the core of the search engine has the following advantages [8].

1) The unit of the ontology: Ontology can provide a clear definition to things, and unified to clarity and formalized description, which makes the system is a unified description for the same thing. In different systems which has different data can be applied in standard.

2) Shared ontology: Terminology in the field of ontology is a certain set and the terms of the relationship between set. It is set up by the domain experts with sharing agreements made, which is handy for various applications in the field of system call multiple data. It also reduce the individual system constructing ontology of duplication. It has established ontology which can merge to form a large-scale ontology in many areas, so that the system involves the expansion of the field.

3) Ontology semantic: Ontology of formal properties make it has a machine readable, which can help computer to realize data exchange on the meaning, rather than the exchange of data to itself, and it is similar to human communication.

The relationship between the ontology: Ontology can describe rich metadata, such as a concept, attribute and relation of data. Take advantage of these, the relationship between the ontology can establish connections between things, and this connection has a subclass and inheritance of instance, so that will bring helps to mining association relationship.

IV.   THE VERTICAL SEARCH ENGINE OF THE LIBRARY RESOURCES BASED ON ONTOLOGY

A.   *Overall architecture of the system*

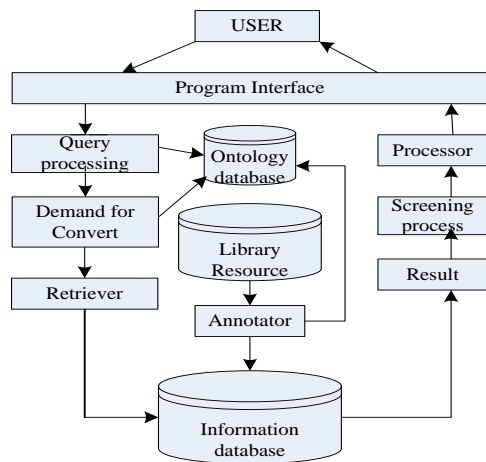The system overall implementation framework is shown in figure 1.

Figure 1. vertical search engine system architecture

1) Ontology is the core of this system work, respectively be queried the preprocessor, requirements, converter and annotators calls. Library resources information database is according to the classification of the concept of ontology to establish a repository of digital library. The framework consists of four modules in the system, which is query preprocessing, demand for converter, retrieval and annotators.

2) Annotators. On the basis of ontology for comes in various forms of library resources information organization, will library resources information into the computer database, in order to search and other operations on it.

3) Query. Preprocessor transform the user query requests into ontology, so that achieve a certain degree of semantic understanding. So that the user submits the natural language query conditions of information into computer can understand.

4) Demand for converter. Combined with query processing results of the preprocessor and the user's choice of search scope and other constraints, the user can submit query conditions into direct search keywords.

5) Retrieval. It will be processed by the keyword search, and it has set up a good library resources information database indexing characters matching.

B. Set Topic Crawler Design Based On Ontology

Because the vertical search requires a high degree of professionalism, so we must improve the universal crawler for library resource oriented topic crawler. We put forward a kind of correlation to determine the theme of the strategy based on ontology, using ontology to the concepts and relationships between concepts in the field of clear definition to improve the decision accuracy[9].

Web crawling circulation beginning page grab, constantly from network information related to the theme, and that will display the results to the user. At the same time, the user can also modify the control parameters to control the scraping of the page. It requires analysis of the topic filtering and links provided by ontology knowledge sharing ontology evaluation as the basis. Another cycle is ontology cycle, from sharing ontology to topic filtering.

User can manage of the ontology with the help of domain experts. It is designed as shown in figure 2.
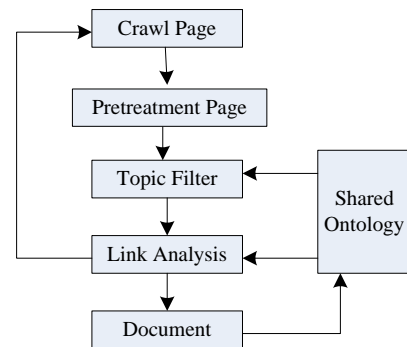

Figure 2. crawler frame of the Specify topic

The user managed the ontology with the help of domain experts. At first the experts should clear the field of shared concepts and the relationship between concepts. Through the relationship they construct the domain ontology. According to the high frequency of the concept of ontology is updated and maintenance.

C. Structured Information Extraction

Web information extraction is unstructured data to a web page according to certain requirements elicitation into structured data. It has huge difference of the vertical search engine and general search engine. Web information extraction mode of transformation is using page structure analysis and intelligent node analysis. The method of automatic extraction is the structured data. The technical is difficulty, but this way upfront development costs is high and the cycle is long. It is suitable for web library level structured data mining and search of high-end applications.

Application ontology to the structured information extraction is frontier research direction. The difficulty lies in automatic semantic annotations on the text. For a particular domain user's specific needs, based on the ontology annotations and information extraction is possible, therefore, we set up semantic tagging models for structured information extraction as shown in figure 3. In the figure the box represent file and ovals represent process. The information extraction is determined according to the user's specific needs.
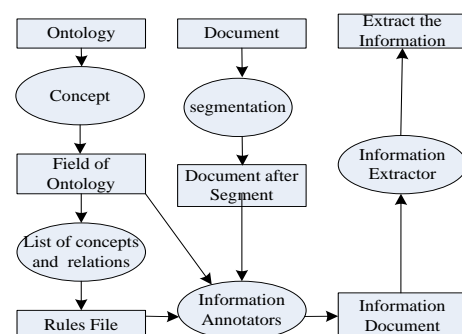

Figure 3. Information extraction based on ontology

D. Information Retrieval Based On Ontology

According to the keywords and retrieval of user input and scope, we use the relationship of the ontology and the ontology model of query, so that the RDF data model of a ternary group can be requested. So this article chooses RDQL query language for conditional ontology query to meet the demand of the user submits to reasoning.

RDQL is a language that is used to query the RDF model[10]. RDF model is a node which can be directed graph for resources or text. RDQL provide a way to define a figure model by the user, that will obtain the results fitting the pattern of defined. We believe that the concept of similarity and its semantic coincidence degree is proportional and its semantic is inversely proportional to the distance.

The concrete application steps:

step1: Calculated according to the ontology concept similarity matrix and the matrix in the file.

step2: The set of input key concept rises to the set and has nothing to do with the concept of keyword set.

step3: Set the similarity value closed, and expands the concept set.

step4: Using the expanded concept set to form the query vector. Based on vector space model of information query will find the pages satisfy certain conditions.

step5: By using the concept of similarity and some other algorithm to sort the results of the query

step6: According to the specific content of the query is analyzed, and comparing with other general search engines, finally to improve the concept of semantic similarity formula.

V. CONCLUSION

In this paper the vertical search construction strategy of library resource based ontology were studied.

Development of the vertical search based on library of the ontology includes ontology-based information collection, information extraction based on ontology, retrieval based on ontology. Retrieval based on ontology specific implementation algorithm are given, so we put forward the resources of the library of the vertical search engine based on ontology is also for the implementation of this type of system provides a new method.

REFERENCES

[1] WANG Xiao-qin, LI Shu-qin and JING Xu. Research on agriculture vertical search engine based on Nutch. Computer Engineering and Design,2014,06,2239-2243.

[2] Wang Xin,Guo Lili,Ma Lina. Cooperative Spectrum Sensing Algorithm Based on Second User Selection and Random Forest Classification[J].ICIC Express Letters, Part B: Applications,2015, 6(3): 845-850.

[3] Wang Chao, Li Shuqin and Xiao Hong. Research on Literature-based Automatic Ontology Construction Method for Agricultural Domain. Computer Applications and Software,2014,08,71-74.

[4] QI Peng, ZHANG Jun and LI Guan-yu. Design of ontology-based classification-indexing model for vertical search engine. Computer Engineering and Design,2010,23,4999-5011.

[5] Chen Lan and Jin Yuanping. Research on Ontology-based Vertica l Search Engine. Computer Applications and Software,2009,11: 129-130.

[6] Wang Xin, Huang Kuan, Gao Zhijun. The Primary Users' Signals Recognition Algorithm in Cognitive Radio Networks via KPCA a nd Random Forest[J].ICIC Express Letters,2015, 9(4):1083-1088.

[7] LV Lin-tao, CHEN Li-ping and ZHOU Hong-fang. Topic Distillat ion Algorithm for Vertical Search Engine. Computer Engineering, 2009,15,44-46.

[8] Jordi Conesa, Veda C. Storey and Vijayan Sugumaran. Improving web-query processing through semantic knowledge. Data & Kno wledge Engineering . 2007 (1)

[9] Hai-Tao Zheng,Bo-Yeong Kang and Hong-Gee Kim. An ontology -based approach to learnable focused crawling. Information Scien ces . 2008 (23)

[10] Lien F. Lai. A knowledge engineering approach to knowledge ma nagement. Information Sciences . 2007 (19)