ISSN : 0974 - 7435 Volume 10 Issue 18

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(18), 2014 [10819-10824]

Research on the prototype of intelligent forestry retrieval system based on ontology

Fengnan Zhang
Shaanxi Xueqian Normal University, Xi'an, 710100, (CHINA)

ABSTRACT

The similarity between the thesaurus and the ontology lays in their functions and features. In regard to agriculture, the establishment of intelligent retrieval prototype should be based on forestry thesaurus and then shifts to ontology, so as to get more efficient retrieval of forestry terminology with higher retrieval speed and accuracy. In this paper, however, the focus would be the construction of intelligent forestry ontology and its intelligent retrieval module. On this basis, to establish an efficient intelligent retrieval prototype with clear elements, the study would combine the management and service of ontology, together with the acquisition of browser information. To further pinpoint the methods and elements of intelligent forestry retrieval prototype meeting higher standards and to provide solid theoretical foundation for its optimization, the last research would be about the document predisposition and other factors. This paper's research method, demonstrating compacted process and clear scientific property, lays reliable basis both practically and ideologically for further study.

KEYWORDS

Forestry ontology; Thesaurus; Prototype of intelligent forestry retrieval; Construction research.

© Trade Science Inc.

INTRODUCTION

In term of the construction of the prototype of intelligent forestry retrieval, ontology, being the basic element, plays a vital role in the process of the transition from the thesaurus to itself. Therefore, the paper would focus on the construction of intelligent ontology, its system frame and the document predisposition. Established with those concerns, the accuracy of for the retrieval of forestry terminology would be improved on stable ground.

CONSTRUCTION OF INTELLIGENT FORESTRY ONTOLOGY

The similarity between the thesaurus and the ontology is that they both are effective organizations and expressions of scientific knowledge, expounding on the relations of words. The thesaurus, however, essentially about the inner connection of descriptors, focuses on the scientific description toward one certain field, and is usually compiled by the experts in this field. It is the shift of relative knowledge into ontology that optimizes its construction, where the thesaurus' scientific property comes from. This transition process, fully utilizing the knowledge system, contributes to the system's inheritance and development. Also established with this idea, the construction of forestry ontology further makes use of the knowledge system[1]. Compiled by the document information center of China's Academy of Forestry, the forestry thesaurus, consisting 64638 descriptors, has become an important reference for the writing of books and pieces about forestry. Combining forestry thesaurus with construction of the intelligent forestry retrieval module, the paper realizes the shift from knowledge available in the thesaurus ontology to ontology, thus the construction of the intelligent forestry retrieval module based on established ontology would be achieved.

System's function and its basic frame

The major function of this system is to serve the forestry ontology's intelligent retrieval and navigation. The retrieval means the display of similar words of imported items. The intelligent navigation means the matching of superordinate and hyponym by the inner relations between words. Figure 1 illustrates clearly the construction process of this system.

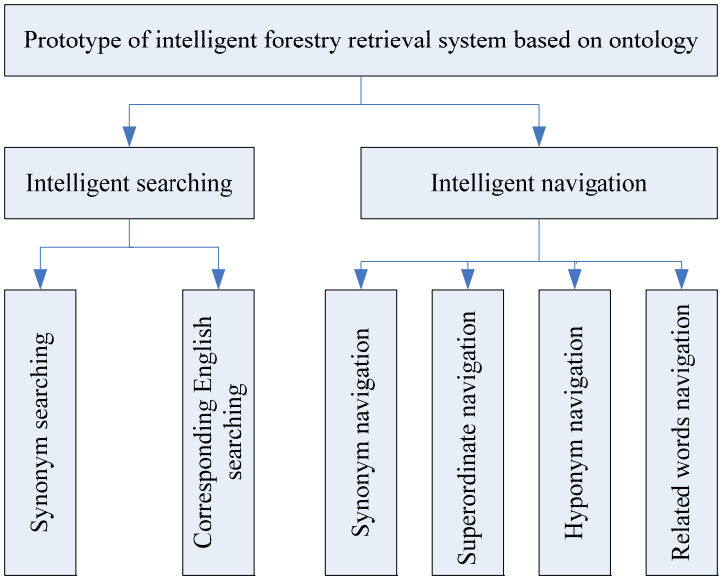

Figure 1 : Prototype of intelligent retrieval function modules

Inferences based on forestry ontology

Now there are many tools for calculation methods of ontology, such as Jena, dial-up software and others. But since the system's construction in this paper is consisted of relatively simpler inference procedures, those tools were not applied. Major techniques used in this study to assist its inference methods are standard object model and XPath information retrieval, which analyzes forestry OWL texts. Below are some of the OWL fragments from shifted forestry thesaurus.

Provided that the retrieved words and the respective synonyms, the relation of superordinate and hyponym are defined, the call of this sub-function can return it to the synonym retrieval list. For the English items and the related descriptors had already been shifted in forestry thesaurus, the retrieval can directly extract English items. The intelligent optimization of the semantic field based on this can compact the near-synonyms with this system and the same goes for the integration of superordinate and hyponym with their own features.

```
<owl:Class rdf:ID="c_7213_Glycine max>
<rdfs:label xml:lang="en">Glycine max</rdfs:label>
<rdfs:label xml:lang="zh">Glycine max</rdfs:label>
<owl:equivalentClass><owl:Class rdf:ID="c_36431_Petits pois_"/></owl:equivalentClass>
<owl:equivalentClass><owl:Class rdf:ID="c_19842_Rhizobium sojae"/></owl:equivalentClass>
<rdfs:subClassof><owl:Class rdf:about="#c_10128_Legume crop"/></rdfs:subClassof>
<rdfs:subClassof>
<owl:Restriction>
<owl:onProperty>
<owl:SymmetricProperty rdf:about="#r_90"/>
</owl:onProperty>
<owl:someValuesFrom>
<owl:Class rdf:ID="c_10176_Soyabean_oil"/>
</owl:someValuesFrom>
</owl:Restriction>
</rdfs:subClassof>
</owl:Class>
```

The prototype of intelligent forestry retrieval system based on ontology

The construction of this system lacks the respective database for it is a prototype system. The main retrieving method is to use Google's searching add-ons to conduct the test through free data interface. By searching Google's data with related codes and back-feeding, the searching results can use the calling program[3]. The feedback data is well determined and contains six formats including Web and Book. The procedures are illustrated in Figure 2.

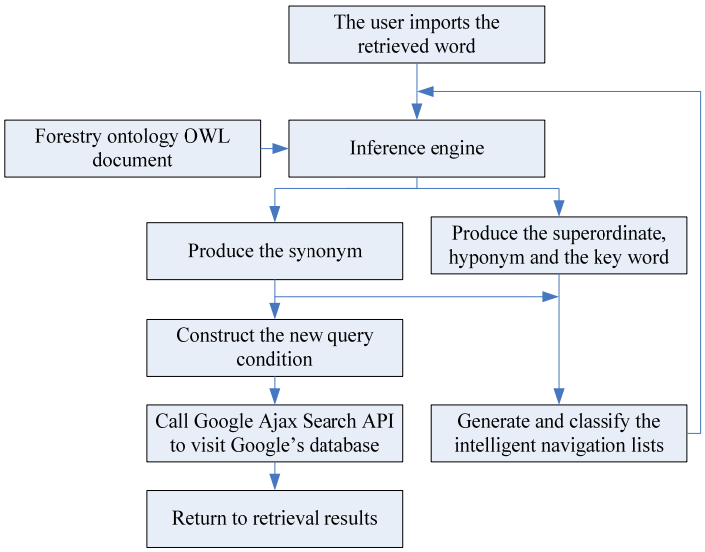

Figure 2 : Retrieval procedures of the system

The procedures clearly are: the user imports the retrieval target, and the inference engine will save the OWL documents into the forestry ontology according to the user's demands and infer the documents. Then Google AJAX would call on the database and conduct the API search, so the retrieval is completed. On the other level, when retrieving the items, the inference engine would also retrieve the relevant words of the related lexical category, including the superordinate and hyponym. By doing this, a group of dynamic hyper-links would be produced and the intelligent navigation lists would be generated. When the user call on the hyper-link to retrieve new words, the text-box would automatically update the needed retrieval items and with a new turn of retrieval be done, the navigation list would also be updated.

SYSTEM FRAME

In term of the retrieval-only system, this paper combines its construction with the semantic module and creates a respective intelligent retrieval module. (illustrated in Figure 3). The construction is focused on seven modules: the management and service of ontology, the inquiring and processing, and the information collection and its related modules.

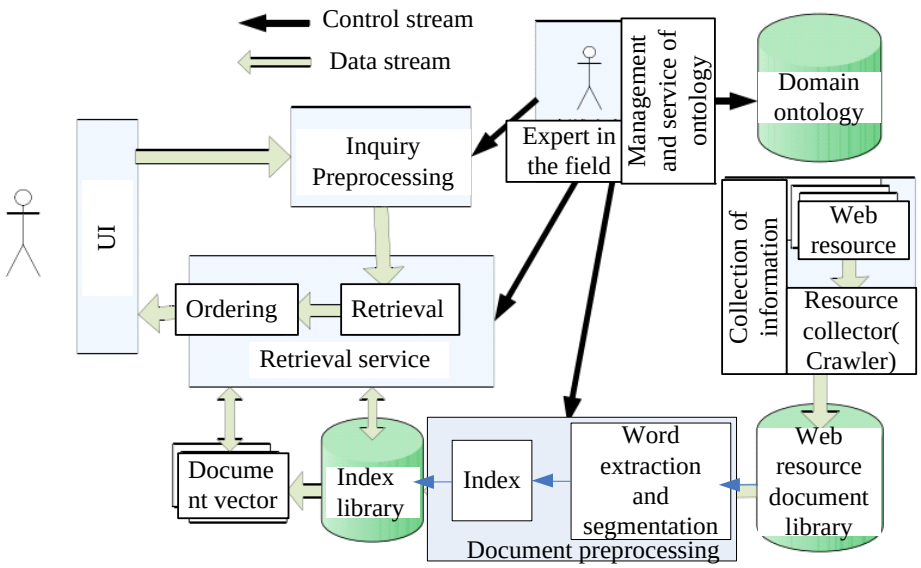

Figure 3 : The prototype of web intelligent retrieval system frame based on ontology

As is shown in Figure 3, this prototype system is differentiated from traditional models. The difference mainly lies in the retrieval service and the processing of related documents. The system in Figure 3 concerns the service module and related ontology management module in various degrees. These modules, however, are based on the forestry ontology. The following discussion will focus on these modules[4].

Management and service of the ontology

The basis to realize intelligent retrieval is to establish the ontology concept system, namely to construct the ontology. With the formal description of the constructed ontology, the correlation matrix of concepts will be established to serve the other modules in the system. The ontology, being the supporting platform of intelligent retrieval's technical process, plays an important role in understanding the user's retrieving purpose and improves the retrieving efficiency. Being the foundation and support of the whole system, this module has the following functions:

With the assistance from the experts in the field and certain construction methods and tools, the module establishes the ontology in the field and stores it in given forms, such as documents, database and others.

To update the established ontology. For example, to add, delete or rewrite the type, case and the property. Such functions' realization is currently manual. But the interfaces of them are kept in the codes for the later realization.

Extracting the stored ontology and analyzing its type, case and relation.

Acquisition of web information

The main retrieval targets in this system's construction are the Web information on the browser. Its collection and gathering depends mainly on the information collection module. The Web information on the browser are dynamic and open and its organization form are also diversified. These features are not optimal for the information collection, so relevant information gathering tools are needed. The information in these tools should be gathered into the browser's information documents before this step is started, so as to provide solid ground for the next step's documents and their processing.

The commonly applied program for the browser's information collection is the crawler, a typical Web program. Constructed by the socket programming, this clustered frame can download the relevant addresses and visit these Webs. The design of such information collection and gathering software should meet the following requires.

First, it should be configurable enough and can capture the initial configurations of the Webs, so as to ensure its depth and breadth.

Second, it can distinguish new connections as soon as possible to enable some simple analysis of downloaded contents thus the network connections are efficiently extracted and added to the Web list.

Third, it can store the data from the Web with efficiency and contribute to the analysis of these data.

Last, it should be able to conduct the real-time-update and effective inspection of the downloaded Webs.

DOCUMENT PREPROCESSING

In the construction of this system, the document preprocessing is the key step for the preparation of the retrieval. The quality of document preprocessing determines that of the retrieval[5]. In this process, the extraction of vectors is an important criterion to determine whether this retrieval system can conduct the semantic retrieval. Figure 4 illustrates the procedures of the document preprocessing.

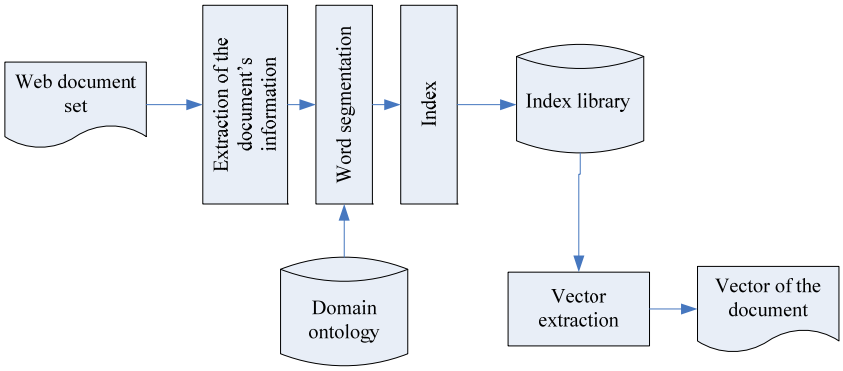

Figure 4 : Illustration of document preprocessing

Information extraction

To ensure the value of the extracted contents, the work should be done in the preprocessing is to extract the content elements and the structured information in the Web. The specific processes are the following[6].

First, filter out the irrelevant network labels.

Then, extract in terms of the titles and the contents of the Web, and integrate these elements to acquire a structured Web page.

Last, the purification of the extracted Web information. Through ruling out the irrelevant elements like the punctuation marks, the preparation of word segmentation is done.

Content segmentation

The main components to construct a Chinese document are the Chinese phrases and relevant sentence patterns. Their concepts, however, are not clearly distinguished. But a retrieval system that lacks word segmentation function would not be able to filter out irrelevant contents, which leads to unsatisfactory retrieval results. To better the word segmentation tool's applicability and practical value, the Chinese word segmentation must be optimized to reach an omni-bearing and three-dimensional procedure. For collected information, it must be extracted in line with its targets and undertake effective word segmentation, so as to product a word set. The paper will discuss the definition of the terminology of the ontology, making the words the reference set. By this, in the processing of technical terms will occur less mistakes and the accuracy of the process will be improved.

Vocabulary index

The quality of the index will directly affect the speed of the forestry word retrieval, as well as the orientating of the Web information. Besides accelerate the retrieval speed, qualified index will also put the calculated amount into control. The construction method of the index library of the segmented words in this paper is basically the reverse index. The minimal units are expressed as the index entries, which covers the size of the document, the creation time and other information[7].

Vector extraction

The objective of vector extraction is, with the support of the ontology, by analyzing the index library, searching ontology items in each documents, construct the semantic vector with these items. The components of each vector are the weight values of respective ontology items in the document. The determination of each weight values depend on the location (title, content, anchor) of the item and other statistical information such as the TF or IDF. The data persistence of the vector determined could be either database or the disk document (shown as Figure 5).

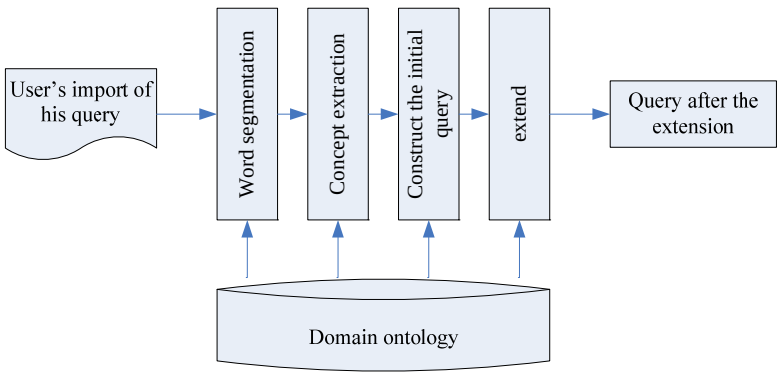

Figure 5 : Illustration of inquiry preprocessing

CONCLUSIONS

We have done the discussion of the construction procedures of the prototype of intelligent forestry retrieval system based on ontology. The research was mainly about the construction of forestry ontology and the processing of the documents. Exhibiting clearly the system frame, the method of this paper is precise and meticulous, thus ensuring the validity of the research output.

ACKNOWLEDGEMENT

The research is funded by: 1."Research of Automatic Integration of Knowledge Ontology in DM" of Science and Technology Department of Hunan province. (project number: 2014CK4010); 2. "Research of Automatic Integration of Knowledge Ontology in DM" of Education Department of Hunan province. (project number: 12C0433).

REFERENCES

[1] Hongwei Meng, Zhiping Zhang, Xiaodan Zhang; Research on intelligent information retrieval model based on domain ontology, Journal of Intelligence, **32(9)**, 180-184 **(2013)**.

[2] Xuyang Wang; Research of intelligent retrieval based on concept correlation degrees, Computer Engineering and Design, **(4)**, 1415-1419 **(2013)**.

[3] Zhijin Wang, Lina Wang, Lingling Hu; Indexing in the intellectual retrieval environment, Library Journal, **30(1)**, 16-19 **(2011)**.

[4] Ziming Zeng; An intelligent commodity information retrieval based on semantic similarity and multi-attribute decision method, New Technology of Library and Information Service, **(1)**, 22-27 **(2010)**.

[5] Sheng Xiao, Jin-zhu Hu; Study on feature item extraction method based on ontology view, Application Research of Computers, **(1)**, 42-44 **(2010)**.

[6] Haiyan Bai, Li Wang, Bing Liang, UMLS and its application in field of intelligent retrieval, New Technology of Library and Information Service, **(4)**, 1-9 **(2012)**.

[7] Guo-shun Wang; Si-ming Sun, Feng Zhang; The personalization information retrieval engineering research based on ontology, Computer Simulation, **30(7)**, 421-423 **(2013)**.