Soil Knowledge Intelligent Retrieval System Based on Ontology

ZHANG XIAOSHUAN, ZHAO QINGLING, TIAN DONG, ZHAO MING*
College of Information and Electrical Engineering, China Agricultural University, Beijing, 100083

*Corresponding author, Email: zhaoming@cau.edu.cn

Abstract: - With the development and popularization of Internet, The research focuses on how to get the requirement quickly and exactly from a large number of information. Using ontology provides a new intelligent searching method based on Web. In this paper, According to ontology theory of agriculture's characters and combining with the major of soil and agricultural chemistry, the retrieval system took the soil knowledge system as example, took native XML(eXtensible Markup Language)Database--Tamino as information navigation database. According the demands input by users, this system will display related information by tree and understand user's demands through clicks, primarily realize Web's intellective searching. This article still introduces the design and implement process of the intellective retrieval system, XML and JSP(Java Server Pages) technology in detail. The system application can be spread for other shared information resources retrieval, providing efficient and relevant services for users.

Key-Words: - Ontology, soil Knowledge system, intelligent retrieval

1 Introduction

With the popularization of Internet/Intranet, there is a large number of information in network. How to get the real time information is always an important problem in the field of information retrieval. Now search engine based on matching of key words or retrieval of subject sort (such as Google, Yahoo, et al.) [1]. In generally, users have to spend much time in filtering the useless information. In the other words, the more data is in the Internet, the garbage is will be found. The major problem is that the engine cannot understand what information users really want and what data means. Using the search engine, users have to take much time to get over irrespective information, because they had got a lot of link that have nothing to do with their requirement. At the same time, there are different expression methods for the same concept between user' and network, users usually can’t receive the useful information. So content expression of concept, that is semantic should be lead into retrieval. Then retrieval evolved the matching of content from key words so as to overcome all kinds of drawbacks from matching of the only expression method [2]. Ontology plays an important role in the intelligent course of the retrieval. Since ontology contains level structure of concepts and logical inference, it has been applied widely in the area of knowledge-based retrieval systems [3]. Taking ontology as theory guide, using scientific soil knowledge system and seeking a new searching method based on web, this article discusses the application of ontology of agricultural character in intelligent retrieval system.

1.1 Conception of Ontology

Ontology is playing more and more important role in computer science. However so far it is a difficult to define exactly ontology in computer field. Cruber from Stanford University defined ontology as “ontology describes accurately conceptualization” that had got approved. The final aim of ontology is expressing accurately undefined information, which can be reused or shared by software system [4].

Ontology is a conceptual model that describes the concepts and the relationships among the concepts. In AI field, many definitions have been given to the term ontology. At present, the definition of ontology accepted widely is “an explicit formal specification of a shared conceptualization [5].

1.2 Ontology character

Ontology is not common conceptual aggregation. It contains not only a complete set of specification of conception but also the relationships among the concepts, which embody immanent structure relation of knowledge. Ontology is concept abstracted from terms with an eye to define concept and express relation among concept. It expounds correctly mapping from terms to concept. Ontology gets semantics by comparison among logical structures of concept, results in improvement of
performance in effectiveness and accuracy. It turns out to be better performance than Thesaurus in application areas. It achieved high recall ratio and precision ratio [6].

This preliminary research involved the soil branch of basic sciences in agriculture. It should be classified as Domain ontology: it consists of concepts and the relationships among the concepts in this field. It can not only be the theoretic basis of soil science, but also improve reuse, reliability, normality and speed ability of retrieval system.

2 Organization of soil classification knowledge system

2.1 Analysis on the information about the original literature database

There are total 56 thousands records in the original literature database, among which 18,522 records fall under soil classification. Each record consists of five fields: Record number, Classify number, Literature title, Key words, Publishing time.

Following Chinese National Classification Standards on books and information, the soil domain ontology is built on the basis of conceptualization and normalization by extracting, cleaning; standardization, integrating on classification labels and keywords from literature database. Statistical analysis methods are also used. According to the results of survey and analysis on classification information, incorporating with characteristic of web information retrieval [7]:

1) In the cases only one record is assigned to some class, which may also contain very important information on some special domain, some measures should be taken in building ontology;

2) To those classes that contain many records, after sorted by record numbers, they will be classified according to taxonomy, enables users to browse those classes that have most records firstly.

3) After sorted according to classification rules, those classes that contain no records and only one class will be cancelled and merged into upper classes.

4) If there are few records in classes below level 3, these classes will be cancelled and records will be assigned to upper level class.

After processing classes with above rules, the soil domain ontology that consists of keywords and relevant numbers of records and class was obtained. In order to embody the superiority of soil domain ontology found based on ontology idea, it’ll be applied to real retrieval system—soil knowledge intelligent retrieval system.

2.2 Foundation of soil knowledge database

Tamino is the first database that using native and standard XML form to process data storage and reading. It realizes integral XML database system and it is Web server of HTTP structure.

Soil knowledge system took ontology as theoretic guidance and based on statistical and agricultural classification standard. Figure 1 is system structure and corresponding tree. System’s element attribute setting detailedly in table 1.

2.3 System structure and corresponding tree

![Figure 1 Soil knowledge system structure and corresponding tree](image)
Table 1: Soil knowledge system structure element attribute setting

<table>
<thead>
<tr>
<th>Element</th>
<th>Attribute</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>CiTiao</td>
<td>complex type</td>
<td>Root element can contain any type child element</td>
</tr>
<tr>
<td>ZTC</td>
<td>complex type, contain 1~∞ child elements, &quot;a a&quot; is only sign value</td>
<td>Abstractive conceptual word according as ontology.Similarly main key of RDBMS</td>
</tr>
<tr>
<td>J1, J2, J3</td>
<td>complex type, contain 0~∞ child elements, no value</td>
<td>Corresponding classification of Chinese Agricultural Classification words</td>
</tr>
<tr>
<td>LH, LM</td>
<td>character type, one value, no any child element</td>
<td>Standard classification number and name in Chinese Agricultural Classification words</td>
</tr>
<tr>
<td>SM, SB</td>
<td>numeral type, one value, no any child element</td>
<td>Express the number of root and child node record tree. Using these data analyse data distributing and hiberarchy when constructing soil knowledge system</td>
</tr>
</tbody>
</table>

3 Application of soil Domain ontology in retrieval system

In Domain ontology of soil field, since it is a tree structure between keywords and classes, the native XML database --- Tamino, was adopted as navigation information database and deployed in the server. Navigation information is stored in Software AG Tamino 3.1, and 56 thousands original literatures are stored in Microsoft SQL Server 2000.

3.1 Three levels in the retrieval system

The retrieval system has three levels: web browser, web server and database.

Databases: Navigation information is stored in Software AG Tamino 3.1, and 56 thousands original literatures are stored in Microsoft SQL Server 2000.

As well as navigation information, 56 thousands original literatures can also be stored in Tamino in order to result in lower complexity in design. Using different databases to store our navigation information and original literatures was based on the consideration that it will be flexible for other organizations using various databases to adopt this retrieval system in their future design without merge their databases used currently into Tamino.

3.2 Intelligent retrieval model of the retrieval system

Major steps of the retrieval system can be formalized as follows:

1) Extracting keywords from user’s requests in browser, passing them as parameters to corresponding JSP web page, invoking JSP program to retrieval in navigation database, displaying those classification information in soil domain ontology that match to these keywords;

2) Classification information of soil domain ontology is displayed in the left of the web page in a view of tree structure. User can click on tree node interested, it will invoke corresponding JSP program to retrieval in original literature database using keyword and class label represented by the tree node clicked as parameters. Then, the results will be shown in the right of the web page;

3) In the cases there is no class that matches to user’s requests in navigation database, JSP program that is written to retrieval in original literature database will be invoked.

This system based on soil knowledge system which imported concept of Domain ontology and followed guidance idea of building agriculture ontology system developing correlative server. Knowledge expressed in classification aim at soil field of agricultural basic subject. It described concept related with soil and provided concept and relation among this concept. According to retrieval words that put in when user retrieve, the system will display correlative classification information. With the more selection of user, it’ll identify and approach gradually user’s demands. At last, the system will help user to find required information.

3.3 Merits of the retrieval system

Considering the factors such as web application, theoretical basis, practicability, etc. the merits of the system can be summarized as:

3.3.1 Advanced in some degree

Communication and interaction between client and server can be reduced, which is important especially in web applications. It can be explained by how the
system works: the first view for the user is the tree structure of classification relevant to keyword submitted in this browser. The user will know the ways to make future selection by click on tree node.

3.3.2 Practical
As mentioned above, navigation information can be considered as a search engine constructed upon original literature, the separate store of navigation information and original literature will make it easy for others to build similar retrieval system in their own domain.

3.3.3 Flexible and efficient information retrieval
Some search engines use keyword to search or provide classification information to user. The knowledge database in this project is constructed according to ontology idea therefore it is possible to search by combining two methods or using them respectively.

As shown in table 2, under the same condition: recall ratio is higher than normal retrieval, which are 65.8%, 76.1%, 85%; precision ratio are all above 80 percent which are also higher than normal retrieval. For example, it is assumed that user desired to search information on air pollution, only one record will be shown if “laterite” and “air pollution” were used as keywords, while specialists in air pollution are conscious of those factors such as acid decline or methane which causes air pollution. To many users who are not specialized in areas that they are search for, Boolean Search turns to be tedious and inefficient.

In conclusion, the way of user`s thinking is well considered in constructing retrieval system. Information is organized according to hierarchy and ontology, in the case that user is ambiguous about their objective it can improve click efficiency and give the evidence that intelligent analysis on user`s request is effective.

4 Conclusion
The aims of the project are to develop an experimental practical system for Web intelligent retrieval, which is really based on ontology through experimental retrieval, to adopt soil knowledge-based system conceived in early step, to demonstrate the strength and foreground in the application in agricultural knowledge mining with soil knowledge-based system based on ontology. The new method for intelligent retrieval provided the contribution to this problem, which can be used in the field of agricultural information retrieval. It provides experience and lesson for subsequent research.

References: