Ontology-based Intelligent Retrieval System for Soil Knowledge

ZHAO MING, ZHAO QINGLING, TIAN DONG, QIAN PING, ZHANG XIAOSHUAN*
College of Information and Electrical Engineering
China Agricultural University, Beijing, 100083
No.17 Qinghua Donglu, Haidian District, Beijing 100083
P. R. CHINA
*Corresponding Author: zhxshuan@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.

The soil knowledge system as example and native XML (eXtensible Markup Language) Database---- Tamino as information navigation have been taken for the retrieval system based on ontology theory of agriculture' s characters and combining with the major of soil and agricultural chemistry. According to the demands inputted by users, this system will display related information by tree and understand users' demands through clicks, primarily realize Web's intelligent searching. The design and implement process of the intelligent retrieval system,XML and JSP (Java ServerPages) technology were introduced in detail. The application of system can be spread for other shared information resources retrieval, providing efficient and relevant services for users.

In this paper, the research goal is to deeply understand the basis of the theory of ontology, the

Chinese Academy of Agricultural Sciences, Science and Technology Information Center database and the literature of soil in areas related to the 18,522 data as an example, the knowledge involved in the soil system, and try to establish good professional knowledge systems embedded in the soil retrieval system, developed a test based on the ontology of the application system, Intelligent Web Search.

1.1 Conception of Ontology

An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an Ontology is a systematic account of Existence. For AI systems, what "exists" is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, we can describe the ontology of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms. Formally, an ontology is the statement of a logical theory.

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. In AI or knowledge-shared field, the conceptual aggregation can be understand by machine, so it’ll play a better role in many ways such as intelligent information integration, knowledge management, information retrieval, resource shared, electronic business.

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. A perfect ontology can provide relation of structural subject and concept such as relation among superclass, subclass and instance, property value and other relation that rely on used express language. Compare to subject word it not only contains relation but also defines normatively and applies clearly these relations.

Due to the nature of ontology, it contains machine-recognized definitions of concept and can do semantic inference on specific term, using semantic information for guiding the query answering process. That is to say, application based on ontology is able to understand what user means and precisely locate to relevant information by semantic inference of keywords or questions requested by users. 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, speed ability of retrieval system.

2 Organization of soil classification knowledge system

2.1 Analysis on the information about the original literature database

Data is the core of the database, data processing is a database management in the work of the most important, it is bound to become a knowledge organization system of the soil an important part. The original literature for soil classification in the database information to the Chinese classification of library materials as a standard, combined with statistical analysis of words and key words for extraction, cleansing, standardization, and integration, the realization of conceptualization, standardization, and ultimately the formation of the field soil domain ontology

This study is related to soil science and agriculture as the foundation of knowledge based on

ontology categories are Domain ontology (the area of Ontology): It provides the concept of this area, as well as the vocabulary of mutual relations between them.

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 list in table 1.

table 1 The sample of record of the original literature database

Record number	Classify number	Literature title	Key words	Publishing time
560825	S157.2	Mechanism and Application for the control of slope soil erosion of techinques for three dimension of vegetation network	water and soil conservation ; vegetation ; three dimension of vegetation network ; techinques for the control of slope soil erosion ; mechanism for the control of slope soil erosion	2007

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.

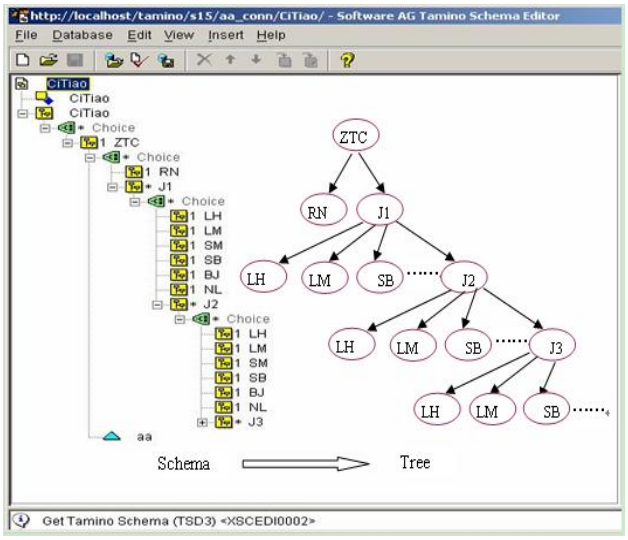

Figure 1 Soil knowledge system structure and corresponding tree

This Schema in figure 1 contains three main elements that schema should be have:

Collection: aa_conn equal to any database of RDBMS

Doctype: Citiao equal to any table or view of RDBMS

Element: Citiao, ZTC, J1, J2, J3, LH and so on equal to fields of RDBMS

So far as the every Schema document, the value of collection and doctype are unique, but the value of element is one to more. The Schema not only embody inherit relation but also process complicated hiberarchy data, which is the most

difference between Tamino and RDBMS as well as the reason of using it to display result intuitively by tree structure.

Schema document for each term, collection and doctype are only of value, and the element value is a multiple, but there is a level of structure, can be well reflected in his son, brother and so inherit the relationship between and be able to handle the complexity of multi-level structure, which is a relational database Tamino and the biggest difference is it to choose an intuitive tree structure to show results.

Seen from the keyword-class cross table, the class number corresponding to each keyword is unique and distributed in different fields of the science and class layers. No grade and logic relations between them were set up. In order to utilize the data efficiently, some regulations were made to organize and maintain the data. Then, in the navigation information system, each keyword is in response to a group of classes and subclasses. It means the relation between each keyword with its relevant class information is one to many and occurs at multiple layers.

The classification information is organized with several first-order classes and their subclasses. The relation among the layers is referred to the actual classification table as the standard. It demands to explain that the first-order class is not the first layer in the classification table, it is the tallest layer that the keyword involved. Therefore, user does not need to begin the navigation from the first layer of the classification table instead of starting from the layer containing data. The first three classes contain the most information. Since user would lose patience if they have to make too many clicks, the depth of class is limited to three.

CiTiao_aa.xml is XML document of soil knowledge system that taking defined Schema as templet. Here display part data.

```
<?xml version=”1.0”encoding=”gb2312”?>
<CiTiao>
  <ZTC aa=”paddy field”>
  <RN>160</RN>
  <J1>
    <LH>S344</LH>
    <LM>cultivation system （faming system）
and way </LM>
    <J2>
      <LH>S344.1</LH><LM>rotation
system<LM>
       <J3> <LH>S344.17</LH>
<LM> rotation system of paddy field and glebe
<LM>
       </J3>
    <J2>
  <J1>
    <LH>X53</LH><LM>soil pollution and
prevention and cure </LM>
    </J3>
  ……
</ZTC>
</CiTiao>
```

Table 2: Soil knowledge system structure element attribute setting

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

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. Otherwise, if that information were stored in

the relational database, query must be fulfilled by specialized interfaces and arithmetic, adding to complexity, lowering efficiency and performance. In contrast to relational database, Tamino conforms to tree structure. Navigation information is stored in Software AG Tamino 3.1, and 56 thousands original literatures are stored in Microsoft SQL Server 2000.

The thoughts on the design of the soil intelligent retrieval system are as follows:

(1)Gathering data from relevant sources, normalizing those data according Domain ontology, storing them in Tamino navigation information database in the format of XML;

(2)Advanced dynamic web page technologies are used to fetch user's requests from retrieval interface and display results matching from information navigation database;

Results is represented by a DOM object, which will be intuition and straightforward to the browser by XML parser and XSLT API.

3.1 Three levels in the retrieval system

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

Web browser: Classification tree structure and navigation information will be displayed here. Users are requested to input their requests as well as specify some classification in which they are interested, and query results are also browsed here;

Web server: All application logic are encapsulated here: it is responsible for dynamic linking to database, SQL and XQL query, XML parse, DOM tree traverse, data's representation form built using XSLT;

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 Three levels in the retrieval system

Server Software: Microsoft Windows 2000 server.
Client Software: Windows 98 or later, IE 5.0 or Netscape 6.0 or later.
Web server: Microsoft IIS.
Languages: JSP, JAVA, XML, XSLT, HTML

XML and JAVA/JSP are all web development application language of international standard. The process of develop program had been simplified because XML separated data storage and representation form and JSP made operation logic and page display apart. "Write one, run anywhere" makes Java program language spanning platform and XML is unrelated with platform, which would extend and apply this system well. So the system use XMLDBMS/RDBMS+XML+JAVA/JSP to construct "Browser/Server" three levels system structure [8,9].

3.3 Intelligent retrieval model of the retrieval system

Figure 2 is Flow chart of intelligent retrieval model

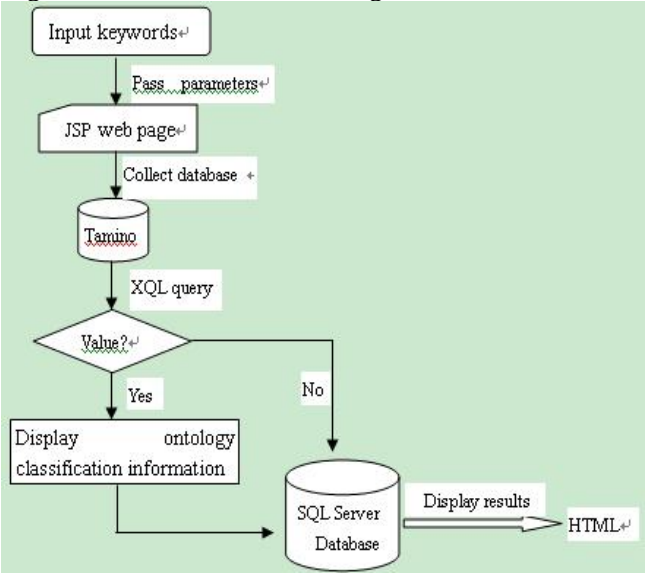

Figure 2 Flow chart of intelligent retrieval model

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.

3.4 Merits of the retrieval system

Conventional keyword search-based full-text search of more mature technology, in government, business, education and other fields have achieved remarkable successes. However, with the development of Internet, network exponentially increase the amount of information of the situation, the retrieval rate and retrieve relevant information (recall rate and precision) of the full-text search technology challenges.

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

3.4.1 Advanced in some degree

Conventional search after search terms in the input, usually by the server-side processing, direct to the background database query, the user in order to get accurate information you want, the need for a large number of results pages to choose, that is, to continue to send requests and server-side interact, which obviously increased the number of clicks users to spend time at the same time. Search in Intelligent Retrieval System, the first presented to the user and the search terms entered by the classification of knowledge related to the tree, the user can clearly see not only related to the relationship between the level of classified information, but can be any flexibility to select the classification tree identified information, Successive clicks through the user, the system can be a preliminary analysis of the smart user's needs, their choice was not until well after the retrieval to the underlying database, so that with a reduction of the direct interaction between the server for Web-based search applications appear particularly important.

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. In addition, Tamino, the native XML database, is used to store navigation information with advanced technology, JSP/java/XML, in web implement.

3.4.2 Practical

The soil Domain ontology information is stored in Tamino intelligent information retrieval system navigation database, the original SQL Server database all the preservation of documentary records, the navigation information database is the database structure on the original literature can be described as a search engine, The characteristics of the system applied to other areas of the future, the existing data retrieval database to provide the possibility of a great deal.

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.

3.4.3 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.

After inputting term in the browser, relevant classification information will be appeared. User can select interesting class and documents will be searched in database. In this way search in original literature database is speed up and effective, and user will not need to seek for desirable information through navigating in many relevant web pages, interactions with remote server are also reduced.

3.4.4 Flexible and efficient information retrieval

Soil knowledge ontology retrieval system is conceived by adopting "B/S" architecture in which web server servers as the intermediate between client and server rather than "C/S" architecture in which client is connected to server directly.

User is classified as manager and registered member, different user is permitted to relevant

privileges. Anyone who wants to retrieval in the system must logging on.

In HTML, both the tag semantics and the tag set are fixed. In contrast to HTML, XML specifies neither semantics nor a tag set. All of the semantics of an XML document will be defined either by the applications or by style sheets. Many standard markup languages have been made in different industries; unification of standard in respective domain will facilitate data transmitting over Internet and makes data available to others. XML is self-description and extensible, easily to describe elements as tree structure and inherited relationships between elements.

Java server pages language is easy to extend with extensible tags support that permits programmer to define tags freely by extending JSP syntax. With standard tags, it will be easy to access and instantiate components that are reusable and platform-independent.

The assumption that the user needs a search for: laterite and the relationship between air pollution may be in the smart retrieval system input box, enter the "laterite" search term, in the event of a classification of information click on the "environmental science, labor protection science" will be "laterite" and the type of its "X" to the front of the same group with 54 records have been inquiries, after the professionals to determine which 10 do not meet the conditions of their precision calculated to 81.5%. Correspondingly, the conventional search of "red" and "environment" phase, investigation was recorded as 37, of which nine are not eligible for precision 75.7%. As a result of more than 56 database records, not by hand to determine which red soil and air pollution related to the relationship between the number of records, it will be the result of two search methods is a database collection that contains the total number of related topics, first of two search way to compare the results of detection, in which conditions in line with the record of five in a repetition (as shown in Figure 3-15), we know the system missed 23 of the record, the recall rate was 44 / (44 +23) = 65.8%. Conventional search missed 39 of the record, recall rate was 41.8%. Method to count and so the secondary sub-node "X5", three-node "X51", respectively, and "laterite" group to carry out inquiries with the situation in Table 2.

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.

Table 2 Result comparatively of intelligent retrieval and conventional retrieval

retrieval mode	retrieval formula	number of record retrieval	number of record accord with term	Recall ratio (%)	Precision ratio (%)
intelligent retrieval	"laterite" & "X"	54	44	**65.8**	**81.5**
conventional retrieval	"laterite" & "environment"	37	28	41.8	75.7
intelligent retrieval	"laterite" & "X5"	42	35	**76.1**	**83.3**
conventional retrieval	"laterite" & " air pollution "	29	17	40.0	58.6
intelligent retrieval	"laterite" & "X51"	18	17	**85.0**	**94.4**
conventional retrieval	"laterite" & " acid rain " or " methane "	18	16	80.0	88.9

In short, the systems take full account of people's way of thinking, in the organization of information, it shows the level of the structure and improve the efficiency of the user's click and the concept ontology to describe the relationship between the application of network access, especially in the user's query goal is not to be too determined, the better realization of the intelligent analysis of user needs and the role of search results.

4 Implement of the Retrieval System

4.1 Construction of the System

Organize the knowledge of soil, only to apply to the actual retrieval system in order to reflect the thinking of the establishment of ontology-based Domain ontology of the advantages of soil, soil knowledge systems Intelligent Retrieval System of the basic design idea is as follows:

1) Information sources to gather relevant data, with reference to the Domain ontology has been established to standardize data collection and information stored in XML format navigation database.

2) The realization of dynamic pages with user search interface extraction query request, and navigation from the database of information to meet the conditions of matching the data result set

3) Result set data is DOM trees, the use of XML parser and XSLT technologies to intuitive, simple forms presented to the user.

JSP/JAVA+XML+NXDB/RDB were applied in the system development. The system includes presentation layer (Web browser), logic layer (Web application server) and data layer (backbone database).

In the representation layer, JSP procedure is responsible for the interactivity with the user, include to display the classification tree and classification navigation information, user's input, selection and search results.

In the logic layer, JAVA procedure wraps all of the business logic, such as establishing the dynamic connection with the database, submitting the query from SQL and XQL, parsing XML and presenting the results with XSLT, etc.

There are two database management systems in the data layer, Software AG's Tamino 3.1 and Microsoft SQL Server 2000. The navigation information is saved in the Tamino database, the agricultural scientific literature information are kept in the SQL Server database. In fact, those data could also be stored in Tamino database to simplify the development. However, the navigation information and literature information are stored separately and will benefit to the development of the NXDB and relation database respectively. It is easier for the reuse and promotion of the system in the future.

4.2 The Main Functions of the System

The main functions of the system include user registration, login and management as well as browse search and intelligent search.

In browse-based search, considering the network application, the system displays the actual classification table in tree form dynamically. The user could input class name to position the query in the classification table, click the class name to conduct retrieval on the literature database, or expand or compass the actual classification table to expand or compass retrieval score. Using actual classification table could increase the efficiency of the retrieval comparing to using classification method directly.

In intelligent search, response to the search word, the system prompts the classes and their hierarchical relations. The user selects what wanted from the prompt. In this way, the user not only knows the distribution of the information concerned, but also finds desired information from any node. If the navigation information database did not contain this search word, the system would do the retrieval in the database directly and show the results.

5 Soil knowledge systems Intelligent Retrieval System Problems

The system is based on the soil knowledge systems, the introduction of the concept of domain ontology, following the establishment of agriculture-related services ontology system to carry out the guiding ideology. Category was expressed for agriculture as the foundation of knowledge is the soil in the field of disciplines, and it also describes the concept and soil-related, provided that these concepts and relationships between them. When users search, the search terms according to the author shows that the classification information associated with the user further options, step by step to identify, close to the user's needs, and ultimately to help users find the information you need.

The application described in this article retrieval system, only the realization of a preliminary search of intelligent, due to development time constraints and the limitations of research, there are still looking forward to resolve several issues:

1) Incomplete data sources: the system is based on the soil knowledge systems technology from the Chinese Academy of Agricultural Sciences of the Library Information Center data, emphasis on the professional literature published in the publication of resources, and soil and other related information at different levels of resources relatively small , such as the Internet, real-time news on the soil, practical technology, the basis of knowledge and policies and regulations, etc., should be included.

2) Less normative data: ontology knowledge system in the soil of the building process, because of the normative data resulted in poor data inefficient, time-consuming. Indexing by staff, recording a result of the negligence of personnel data to the knowledge of an unreasonable system brought unnecessary troubles, so in the current intelligent retrieval system there are still some need for human intervention process, the future system should be to achieve computer to automatically determine the source of classified data, build a knowledge ontology, and then embedded applications. It should be noted that this system has been carried out on samples pre-treatment data to ensure that the normative data.

3) Complex Boolean search logic: the system in the future to achieve the user can enter complex Boolean-type search, Domain ontology shows that the classification of information, and then direct users to query.

6 Conclusion

Domain ontology building process, including specification, conceptualization, standardization, implementation, maintenance, each stage must have completed work on the existing knowledge in the field of soil science for analysis, extraction, restructuring and achieving soil science and agriculture as the foundation of the academic language norms, services and management, on the basis of the application through the retrieval system to the Internet. The system is based on the theory of ontology is not only intelligent retrieval system to improve the recall rate and precision, and the application of Web-based media through the Internet to provide users with more services, the sharing of resources.

"Browser / Server" three-tier architecture, using open standards, non-specific to ensure that its application versatility and cross-platform nature; it is only with the client browser can be common, maintenance and upgrade work is done on the server side, and not the client of any change, they are greatly reduced development and maintenance costs; the three parts of the system relatively independent of each module, when one part of the module changes, other modules are not impact, the system becomes very easy to improve, and can be used by different manufacturers of the products to form a better system performance; system in the client and database server add another layer between the Web server, so that the two are no longer directly connected to customers machine can not directly manipulate the database, effectively prevent the illegal invasion of users to protect the security of the system.

Native XML database NXDB (Native XML DataBases) Tamino is a specialized XML format for document management and data access to the database operation, it also features a combination of the general database, such as support services, concurrency control, security mechanisms, such as the secondary development interface (criminal Chunxiao, Li Lei, Computer World), the only difference in its internal storage model based on the XML document tree structure, rather than the relational model. Involved in the system of soil classification information data such as relational database, use the table of many interrelated to that used NXDB Tamino can easily be achieved, and good performance of the multilevel nature of the data structure, data between father and son, brothers, relations and inheritance relations, to support the labels and the path-based operation, to support the most up-to-date technical standards for XML.

XML to the Web-based application system to give a powerful functionality and flexibility, so it gives users and developers have brought many benefits.

Database information is to uniquely identify XML, so the software necessary to understand each and every search a database is how to build and follow a uniform standard, so that retrieval becomes easy and fast. Data organization, established, XML is sent to the client browser to use. Can also be sent to other applications, or middle-tier server object for further processing, XML and HTML, script, common object model for flexible combination of the three-tier Web application software provides the necessary technology.

XML logo as a result of the data and display the complete separation of, developers in the structure of the data description of nested procedures in order to demonstrate how to display data, which makes the client computer with the user interaction between the reduction of as much as possible, but also reducing the volume of data exchange server and browser response time, greatly enhance the performance of the server upgrade.

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.

The research is just the beginning of the study on ontology and its application in agriculture. It was supported by the national key project of science and technology, especially regarding the intelligent retrieval of agricultural information.

Acknowledgement

This research is supported by the Hi-tech development plan (863 projects under Grand No. 2006AA10Z239). The authors are grateful for the

anonymous reviewers who made constructive comments.

References:

[1] J C Y, Li Qing. WebReader, A Mechanism for Automating the Search and collecting Information from the World Wide Web[C]. Proceedings of the First International Conference on Web Information Systems Engineering, 2000; 2:47~54.

[2] Wan Jie, Teng Zhiyang, Application of Ontology in Content-based Information Retrieval. Computer Engineering, 2003;29(4):122～123.

[3] Li Jing. Some problems discussion on Ontology for knowledge oganization, Collection information of Renda, 2003,4.

[4] Liao Minghong, Ontology and Information Retrieval, Computer Engineering, 2000;26(2): 56-58.

[5] Ming Zhong Cai Shubin,Li Shixian, Ontology construction based on ontology inheritance, WSEAS Transactions on Computers, v 5, n 5, p 975-982, May 2006

[6] Deng ZhiHong, TangShiwei, Yang Dongqing, Semantic -Oriented Integration The role of Ontology in Web Information Integration, Computer Applications, 2002,22(1): 15~17.

[7] Qian Ping,Su xiaolu, Self-evaluation and check report of special subject for National Key Science and Technology Plan, 2003.

[8] Yang, Sheng-Yuan. How does ontology help web information management processing, WSEAS Transactions on Computers, v 5, n 9, p 1843-1850, September 2006.

[9] Hong xi-jun, Tian Yong-tao, Li Cong-xin. Dynamic information issue technique based on JSP, Computer Engineering and Design, 2002, Vol.23, No.2, pp.29-32

[10] Adda Mehdi, Lei Wu, Feng Yi. Personalization based on domain ontology, WSEAS Transactions on Computers, v 5, n 3, p 598-603, March 2006