Research on Construction Method of Agricultural Domain Ontology

Guoxia Yang [1, a], Ziyu Liu[2], Xiaomin Shen[2]

[1]Institute of Information Science & Engineerning, Hebei University of Science and Technology, ShiJiaZhuang HeBei 050018, China

[2]School of Economics and Management, Hebei University of Science and Technology, ShiJiaZhuang HeBei 050018, China

[a]email: 1012451379@qq.com

Abstract. Based on the two major methods for the construction of domain ontology, that is, ontology engineering and thesaurus-based ontology construction, this paper puts forward a construction methodology of agricultural domain ontology based on thesaurus. This paper details all parts of the methodology. Under the guidance of this methodology, we build agricultural domain ontology.

Introduction

At present, the retrieval method which based on the keyword play an important role in searching and utilizing the resources of information, but it is poverty in recall ratio and precision ratio. For example, when people search relevant information in the field of agricultural and input the search term "soybean", the search results include a great deal of information about "soybean", while "soybeans" is "soybean", it is not reflected in the search results. Secondly, when searching information, people might don't know the exactly search term, while just know that has a certain relationship between the words, such as when inputting "soybean", people don't want to find information about "soybean", they want to find "legume crops" which range is broader than the concept of "soybean" or " beans " which range is smaller than the concept of "soybean", or it is " soybean oil" which is related to "soybean" and other aspects of the information. As a modeling tool of conceptual model, ontology can describe information system in semantic and knowledge level and provide a new approach to resolve these problems which are the frontier and hotspot of research in the field of information [1]. Therefore, the purpose of this paper is to construct the agricultural domain ontology and improve the recall ratio and precision ratio of the information resources in the field of agriculture.

According to the degree of dependence of the field, ontology can be divided into general ontology, domain ontology, task ontology and application ontology [2]. The understanding of the field was defined by domain ontology explicitly, it can effectively organize the knowledge in the field and make the knowledge better sharing and reusing. At present, there are mainly two ways to acquire domain ontology. One way is exploring and discovering the basic concept of the industry from existing industry database through knowledge getting technology, and then confirming and establishing the association of concept by the field experts. Another way is converting the traditional knowledge organization system, such as classification and thesaurus, to the initial core ontology. At present, the focus of attention is the second method. Agriculture has its own vocabulary which is "Agricultural Thesaurus", so it is a comparatively scientific method of constructing agricultural domain ontology based on traditional classification/thesaurus.

The Thought of the Construction of Agricultural Domain Ontology

The domain ontology is a model which used to describe the concepts and the relationship between domain concepts, and it includes relatively complete knowledge and the abundant semantic relation in a field. Agricultural domain ontology is mainly consisted of the concept of agricultural knowledge, the relationship between the concepts, and the formal description language that the

computer can identify. When constructing of agricultural domain ontology, firstly, determine the key concepts and the relationship between these concepts through the system arrangement of domain knowledge. Secondly, determine the instances and the axiom of the concepts.

Thesaurus has many Similarities to ontology, such as they are both used to describe and organize discipline knowledge, and they are both contain the words (the concept, class) and the relationship between the words (concept, class) etc.[3-5]. Generally, thesaurus includes relatively complete descriptors and the relationship between descriptors in a discipline filed, and it is accomplished by many experts in the field. Transforming the existing system of knowledge in thesaurus into the ontology, it can accelerate the ontology building process, and increasing the scientific of ontology. At the same time, it is the scientific use of existing knowledge and keeps the continuity and inheritance [6-10]. "Agricultural Thesaurus" contains a total of 64638 Thesaurus. Over ten years of practice showed that "Agricultural Thesaurus" has become the industry standard of Chinese Agricultural.

In our construction of agricultural ontology, we take the "Agricultural Thesaurus" as the mainstay reference, and also with reference to "China agricultural professional classification", "Chinese Library Classification (Fourth Edition)", "Agronomy Introduction", and "animal science theory" etc.

Constructing Method of the Agricultural Domain Ontology
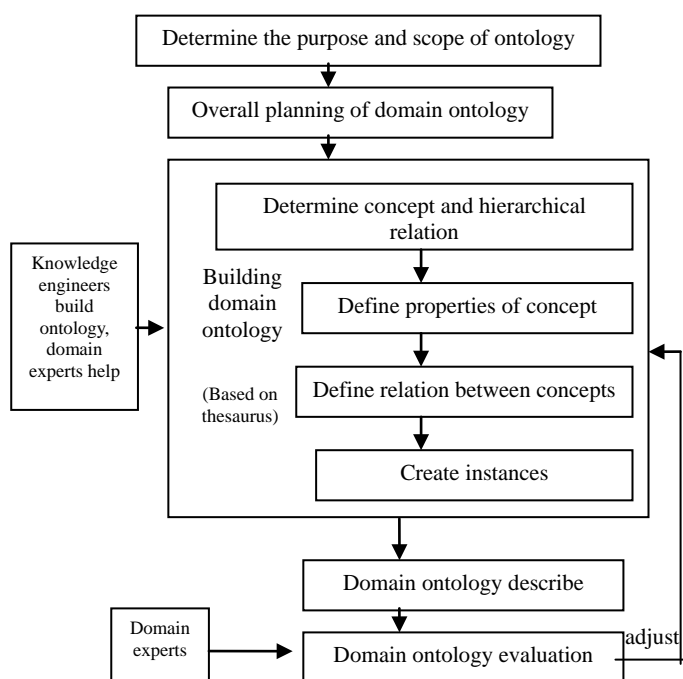

Fig.1. The flow of agriculture domain ontology construction based on thesaurus

With reference to the existing method of ontology construction, this paper proposed a method of constructing agricultural domain ontology which based on thesaurus, as shown in Fig. 1. The method is consisted of the following 5 steps.

(1) Determine the purpose and scope of ontology

The main purpose of construction of agricultural domain ontology prototype at this stage is semantic organization and semantic retrieval for the agricultural knowledge. It is a very difficult work to establish complete or perfect domain ontology, may even be a huge project. Therefore, specifying the application purpose of domain ontology which has important significance to limit its scope, enhance its pertinence, reduce the construction difficulty and shorten the construction time.

(2) Overall planning of domain ontology

The main task of this step includes determining the goal of ontology and concept scope, determining the ontology evaluation standard and some related principles, selecting the ontology development tool and the ontology language etc. We select protégé as the ontology development tool, and choose OWL as ontology description language.

(3) Building domain ontology

Based on agricultural thesaurus, the knowledge engineer build domain ontology, and it also need the professional staff's help. This step includes concept selection, construction of the relationship between concepts, defining properties of concept and creating instances.

(4) Domain ontology describe

You can use natural language, framework, semantic network to represent ontology, but usually often use ontology language. Usually, developers take advantage of existing ontology development

tools to assist the ontology construction, and it can be directly converted into the corresponding ontology representation language format. This paper use OWL to express agricultural domain ontology.

(5)Domain ontology evaluation

We can invited the agricultural experts to evaluate the ontology prototype, mainly including the concept, attributes, and the clarity and accuracy of the relationship, and according to their views return to the third step to modify. Finally, the agricultural domain ontology prototype was constructed basically after repeated modifications.

Construction of Agricultural Domain Ontology

A complete ontology is consisted of five basic elements, including concept, relation, function, axiom and instance. The ontology can be expressed as the following form: O={C, R, F, A, I}.

It includes the concepts, properties, relations and instances in agricultural domain ontology model.

(1) Choice concept

The first class concept main comes from the words of "Agricultural Thesaurus". The second class concept is the classification of the first class concept, and if the first class concept has two or more division methods, we often choice the most general division method as the second class concept. If the second class concept can still be classified, then divide it down, and until no classification.

For example, according to "Agricultural Thesaurus", the agricultural knowledge is divided into three categories: the object of agricultural production, agricultural means of production and agricultural production process [7]. So the first class of the agricultural domain ontology respectively "production object", "means of production" and "production process". The second class of "production object" is "crop", "economic animal" and "soil". The third class of "crops" is "the crop", "horticultural crops". The fourth class of the "the crop" is "food crops", "economic crops", "legume crops", "tubers" and "forage crop".

(2) Determine the property

The property is the description of a concept, and it can be used to illustrate the characteristic of concepts and the feature of some instances. For example, the property of "legume crops" has "name", "the main production area", "alias", "production cycle" etc..

(3) Determine the instance

An instance is the individuality of a concept, and it inherits all property belongs to the concept and has the corresponding attribute values. For example, the instances of "legume crops" have "soy", "green beans", "bean" etc.

(4) Determine the relation

There are two types of relations in ontology. One is hierarchical relationship including the hyponymy relation and instance relation. The other is non hierarchical relationship, and it can be described by adding the special property of concept to connect two concepts. Non hierarchical relationship need to be determined by domain experts. For example, add the special property of "cultivation" to the concept of "breeding technology", which can describe the non hierarchical relationship between the concepts of "breeding technology" and "legume crops". Therefore, we can get the relation: breeding technology < cultivation> legume crops.

This paper chooses OWL Lite that is setting by the W3C as the description language of ontology. OWL Lite not only can accurately depict the semantics, but also can be supported by efficient reasoning system. The tools of ontology construction is Protégé which developed by Stanford University in USA.

The following is a segment of "soybean" in agricultural domain ontology:

```
<owl:Class rdf:ID="soybean">
<owl:equivalentClass><owl:Class rdf:ID="green beans"/></owl:equivalentClass>
<owl:equivalentClass><owl:Class rdf:ID="soybean"/></owl:equivalentClass>
<rdfs:subClassOf><owl:Class rdf:about="#legume crops"/></rdfs:subClassOf>
```

</owl:Class>

Conclusion

In this paper, we put forward a construction methodology of agricultural domain ontology based on thesaurus. Under the guidance of this methodology, we build agricultural domain ontology, and use OWL language to describe it.

On the basis of agricultural domain ontology, we use the Eclipse as a development environment, and use the Jena as the inference engine, initially realized the ontology query and related literature semantic annotation of the agriculture domain. Through extend the user's query word, we realize the semantic expansion retrieval for agricultural literature. Improvement the quality of agricultural domain ontology, reasoning retrieval and the ontology evolution is our study in the next step.

Acknowledgement

References

[1] KLEIN M, BERNSTEIN A. Searching services on the semantic: Web using process ontologies[C]. Proceedings of the International Semantic Web Working Symposium (SWWS), Amster2, dam: IOS Press, 2001, 159-172.

[2] Ali, W., Khan, S., 2008. Ontology driven query expansion in data integration[C]. In: Proceedings of the 2008 Fourth International Conference on Semantics, Knowledge and Grid, IEEE Computer Society, 2008, 57-63.

[3] M Fernandez-Lopez, A Gomez-Perez, A Pazos-Sierra, et al. Building a chemical ontology using methontology and the ontology design environment[C]. IEEE Intelligent Systems and Their Applications, 1999, 14(1): 37-46.

[4] K Knight, et al. Filling knowledge gaps in a broad-coverage machine translation system[C]. In: The 14th Int'1 Joint Conf on Artificial Intelligence, Montreal, Canada, 1995: 1390-1396.

[5] Li Jing, Meng Liansheng. Comparison of seven approaches in constructing ontology[J]. New Technology of Library and Information Service, 2004,112(7): 17-22.

[6] Eman Jayven. Owl Exports From a Full Thesaurus[J]. Bulletin of the American Society for Information Science and Technoloy, 2005,32(1): 22-26.

[7] Mao Jun. Research on RDF-based thesaurus[J]. Journal of the China Society for Scientific and Technical Information, 2003,22(2): 163-168.

[8] Asanee Kawtraku1. Automatic Term Relationship Cleaning and Refinement for AGROVOC[OL].[2006-07-14]. http://ftp.fao.org/docrep/fao/008/af240e/af240e00.pdf.

[9] Qin jian, Paling Stephen. Converting a controlled vocabulary into an ontology: the case of GEM. Information research,2001,6(2).[2008-11-20]. http://informationr.net/ir/6-2/paper94.html.

[10] B.J.Wiefinga, A.Th.Schreiber, J.Wielemaker, J.A.C.Sandberg. From Thesaurus to Ontology. In: Proceedings of the 1st international conference on knowledge capture, 2001: 194-201. [2008-11-20]. http://www.cs.vu.nl/guus/papers/Wielinga01a.pdf.