

# ONTOLOGY-BASED AGRICULTURAL KNOWLEDGE ACQUISITION AND APPLICATION

Nengfu Xie<sup>\*</sup>, Wensheng Wang, Yong Yang

*Agricultural Information Institute, The Chinese Academy of Agricultural Sciences, Beijing, China, 100081*

*\* Corresponding author, Address: Agricultural Information Institute, No. 12 Zhongguancun South St., Haidian District Beijing 100081, P. R. China, Tel: +86-10-68919819, Fax: +86-10-68919820, Email: nf.xieg@caas.net.cn*

**Abstract:** Agricultural knowledge is a special kind of domain knowledge, and is a significant basis for agricultural knowledge-based intelligent information service system. This paper presents main results of our ongoing project of Agricultural Knowledge Processing and Applications (AKPA): 1) An agriculture-specific ontology, 2) A method for agricultural knowledge acquisition and representation, 3) An experiment of An Intelligent Agricultural knowledge-based knowledge service system. At last, we will conclude the paper.

**Keywords:** Ontology; Agricultural knowledge; knowledge Service System

## 1. INTRODUCTION

Agricultural knowledge is a special kind of domain knowledge, and is a significant basis for An Intelligent Agricultural knowledge-based knowledge service system, such as agricultural instructional systems, agricultural knowledge processing systems and agricultural expert systems.

As the starting point of our project AKPA, we have chosen several influential sources of agricultural knowledge and consulted a few agricultural specialists. The knowledge sources includes: 1) the Agriculture Volume of

the Encyclopedia of China (Cai, 1996), 2) A Dictionary of Agriculture (Com et al., 1998), 3) Illustrated Handbook of Food Crops, Economic Crops and Herbal Plants. Based on the knowledge sources and the advice from our consultants, we started to develop an ontology of agricultural objects, called AgriOnto.

In AgriOnto, each category has a frame-based representation, and each category has a list of slots (attributes or relations) for describing its instances. Categories can be inter-related in several semantic relationships; some are general, such as is-a and have-part(s), and some are more specific to agriculture, such as IsVariantOf(x), immune-disease(x), Harm(x).

In addition to offering a terminology for describing its instances, AgriOnto plays other two significant roles. Firstly, attributes and relationships in a category are knowledge 'place-holders' of the instances of the category, and they are to be filled in during the knowledge acquisition for the instances. Secondly, axioms in an ontology can be used both in knowledge inference and knowledge verification during the knowledge acquisition procedure. When acquired knowledge violates such axioms, the knowledge engineer is alarmed to identify and solve the problems.

We designed a semi-automated method for agricultural knowledge acquisition sources (Cao et al., 2002). In practice, it is a valid and feasible method by building ontology-based base. In essence, the work is primarily divided into the following steps: Firstly, to build AgriOnto hierarchy. Secondly, to formalize the text knowledge on the basis of AgriOnto. Thirdly, to compile and check the knowledge. Fourthly, to built knowledge-based service systems. The primary step is knowledge acquisition, that is to say, knowledge formalization depicted in figure 1.

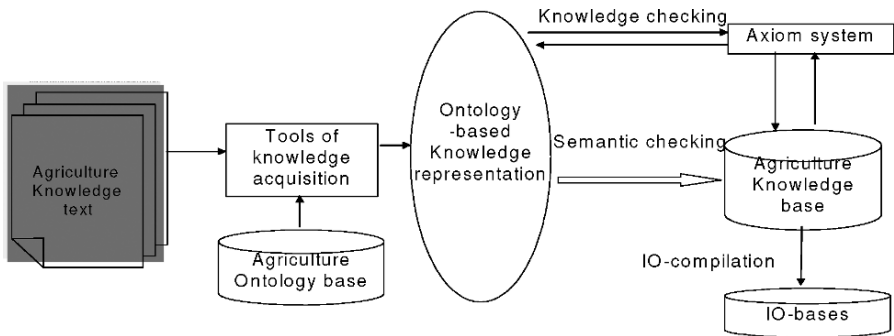


Figure 1. The flow of Ontology-based knowledge acquisitions

The paper is organized as follows. Section 2 presents how to build the hierarchy of agriculture knowledge and the overall description of agriculture-specific ontology. Section 3 gives the method of agriculture knowledge acquisition. Section 4 exemplifies the application of agriculture knowledge. Finally, Section 5 concludes the paper.

## **2. AGRICULTURAL ONTOLOGY**

In general, ontology is an explicit specification of conceptualization (Chaudhri et al., 1997; Cao et al., 2002; Gu et al., 2003). Nevertheless, the term ontology has been controversial in current AI practice, and so far no formal definition exists. In our work, we have selected to use the term of domain-specific ontology (DSO). In practical term, developing AgriOnto includes three steps:

- Building a domain-specific knowledge hierarchy
- Defining slots of the categories and representing axioms
- Knowledge acquisition, this is to say, filling in the value for slots of instances.

### **2.1 Agriculture knowledge hierarchy**

AgriOnto indicates formal definition of agriculture and their relation (Fig. 2). The definition and relations form an integrated hierarchy of agriculture. With labor object as the center of agriculture hierarchy, we divide agriculture knowledge into seven taxa: labor object, production process, production technology, agriculture engineering, agriculture branch, and agriculture environment and agriculture regulation. The labor object as the center of agriculture knowledge hierarchy aims at facilitating the people, who want to know the labor object knowledge, to access to the related knowledge of other taxa.

Agriculture branches are divided into farm branch, forest branch, herd branch, sideline branch and fishery branch. Agriculture engineering is classified as agriculture machine, soil and water conservation, land and land utilization, rural building and structure, agricultural product and by product processing and rural energy resources by agriculture production requirement. Agriculture environment and agriculture regulation are classified by agriculture branch characteristic.

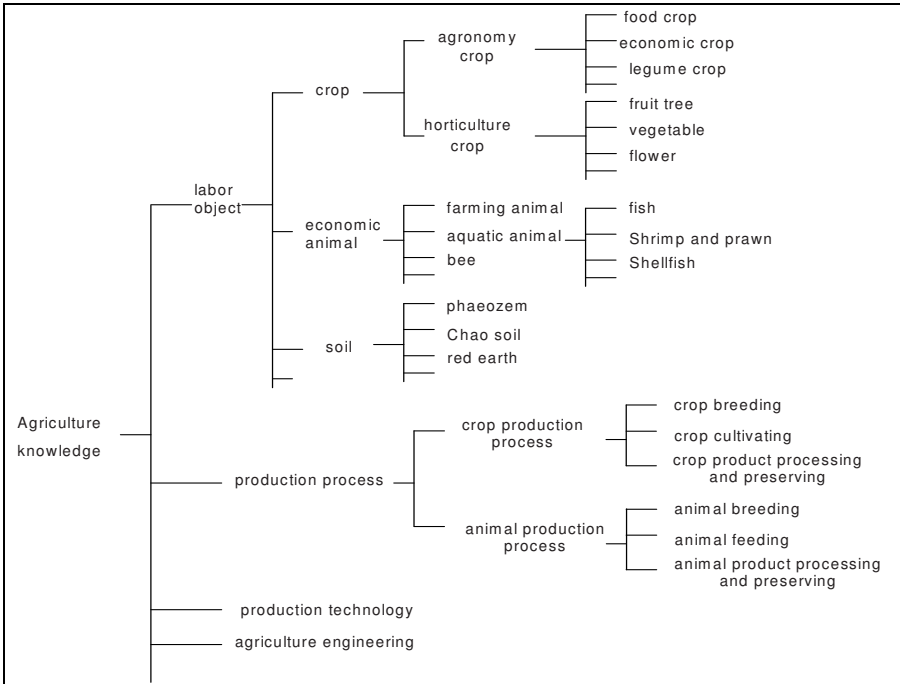


Figure 2. This shows some parts of agriculture knowledge hierarchy

To represent AgriOnto, we have designed a formal language that description the category definition and axiom. The syntax of the language is given in (Gu et al., 2003; Zhang et al., 2002). In practice, to some degree we have classified the attributes of a category when we classify the categories of agriculture knowledge, but this is not yet describe the natural character of a category and need be divided into deeper subclasses so that a concept can be clearly represented.

First, considering the knowledge connection between different domain-specific and the redundancy of NKI base (Cao et al., 1998; Cao et al., 2002), we share many botanical concepts by inherit relations in botanical ontology. In addition to this inheriting knowledge, we must analyze the particular character of crop. The crop knowledge has a close relation with other subject knowledge such as botany, zoology, physics, etc. Firstly, crop knowledge can be look as one part of botany, so it contains the attribute category such as heredity category, form and structure category, distributing category and so on. From agricultural view, there are crop growth category such as plant-disease and insect-pest category, crop environment category. Each attribute category also contains many slot definition like concept category. Figure 3

shows crop attribute hierarchical system. crop-distribute category inherit biological-distribute category and then contains all slots in biological-distribute category (Figure 4).

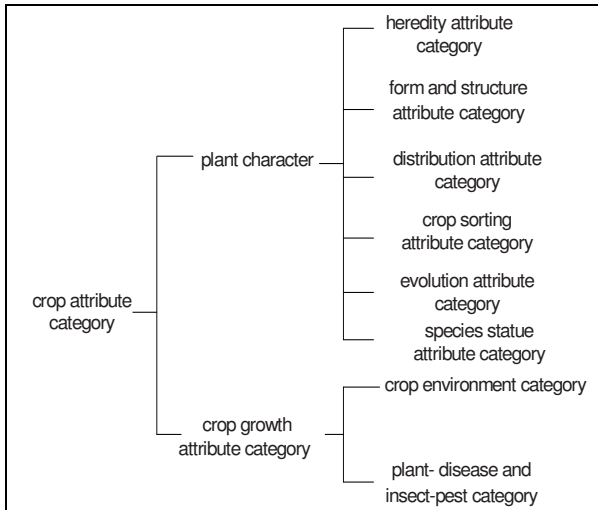


Figure 3. The hierarchical structure of crop- crop attribute categories

<pre> defcategory Biological-distribute-attributes {     attribute: original-pro-area         : type StringArray     attribute: producing-area         : type String Array     attribute: distributing-area-in-China         : type StringArray     attribute: foreign-distributing-area         : type StringArray     attribute: original-pro-area         : type StringArray     ... }         </pre>	<pre> defcategory Crop-distribute-attribute {     IsSubClassOf: Biological-distribute-attributes     attribute: planted-area         : type StringArray     attribute: cultivated-area         : type String Array     attribute: origin-of-plant-area         : type StringArray     attribute: most-production-area         : type StringArray     attribute: wild-area         : type StringArray     ... }         </pre>
--	---

Figure 4. This figures shows biological-distribute and crop-distribute-attribute categories and their relations

## 2.2 Slots of category

In agriculture domain, Addition to attribute, relation is an most important definition in category. It primarily describes the relationship of entity concept. It indicates a proposition or an assertion so we usually use verb to represent it, but sometimes we also use noun to represent it. Relation can connect knowledge between concepts so that we can get the related

knowledge when we look for part of concept knowledge. An important point is that a word as relation is also as attribute, which we call it attribute-relation definition. In table 2, it shows some common relations in the current AgriOnto. Cultivate-technology and breed-technology relation also represent a plant attribute.

In order to describe relation and attribute correctly, in other words, to implement the semantic integration, in category we define many facets such as time, confidence, basis, cause, etc. facet gives additional explanation for relation and attribute. Facet plays important role in maintaining the integration of frame knowledge. For example,  $x$ -Identical-to:  $y$  {facet: length} indicates  $x$  is identical to  $y$  in length, so get an integrated semantic representation.

Table 1. Distribute attributes

Attribute	Meaning
original-pro-area	The original area where plant lived
producing-area	Area that plant used to live
planted-area	The area of planted crop
cultivated-area	The area of cultivated crop
origin-of-plant-area	The area where plant was planted first
most-production-area	The area where there are most production of plant
foreign-distributing-area	The foreign area where plant live
wild-area	A wild plant area

Table 2. Some common agriculture relations

Relation	Meaning
cultivate-technology	The technology that a crop usually is cultivated
breed-technology	The technology that a crop usually is bred
Insect-pests	Some insect pests that destroy the plant, Such as budworm, ladybird beetle and so on
plant-diseases	Some diseases that destroy a part, an organ, or a system of plant, such as from various causes, such as wheat ergot, potato mop-top virus and so on.
immune-disease	Some diseases that a plant immunize.
IsVariantOf	It indicates that a species is a variant of another one.

### 2.3 Axiom of category

We have been building a very large agriculture knowledge base from several knowledge sources (Cai, 1996; Com, 1998). In our practice, we find that it is extremely important to ensure that the agriculture knowledge stored in the knowledge base is accurate and consistent. For each slot defined in a category, we have to specify one or more axioms to constrain their interpretation. These constraints are actually integral components of our categories. We have summarized a list of agriculture-specific axioms both for identifying inconsistency and inaccuracy in the acquired knowledge and for reasoning with the acquired knowledge. They form a first-order axiomatic system, and are an integral part of our whole ontology of crop. When a piece of crop knowledge is stored into the knowledge base, it is first checked by these axioms. If one of the axioms is violated, relevant information is reported to a knowledge engineer.

### **3. KNOWLEDGE ACQUISITION FROM TEXT: ONTOLOGY-DRIVEN METHODS**

In recent years, knowledge acquisition from text has received much attention (Zhang et al., 2002). A key reason is that majority of the knowledge of a domain are presented in domain texts and documents.

In this paper, we utilize two methods to acquire agriculture knowledge from free-structured text. The first KAT system is a frame language for knowledge engineers to formalize text, together with the frame compiler mentioned above in OKEE. After the text is formalized, a frame compiler compiles frames into IO-models based on relevant categories. Although this method is not natural, most of our project knowledge engineers choose the frame language (NKI-FL) in formalizing domain knowledge. The second system is OMKE system which is an ontology-mediated knowledge extractor. The input to this system is semi-structured text (Cao et al., 2002). By semi-structured text, we mean that the syntax of the text is relatively fixed and thus can be easily summarized manually. Experiment in AgriOnto shows it can extract 50,000 Chinese characters per minute, this is, about 40 pages of A4 size per minute.

### **4. AN INTELLIGENT AGRICULTURAL KNOWLEDGE-BASED KNOWLEDGE SERVICE SYSTEM**

In knowledge engineering, ontology is used to share and reuse knowledge and as standard for communication between computer and man. We develop some ontology-based applications on basis of knowledge base. In following, we introduce An Intelligent Agricultural knowledge-based knowledge service system: a Web-based consultant system.

A Web-based consultant system is a platform building on AgriOnto base. It can provide a user with what he wants to know quickly and correctly when he put logical query into its human knowledge interface (Feng et al., 2002). A query may have many expression forms, but they represent the same meaning. The consultant system will answer the query by relating the query's intension with corresponding ontology. For example the two queries “菜豆的栽培技术是什么?” (what's wheat's cultivated technology?) and “菜豆怎样栽培?” (how to cultivate wheat?) can be look as to ask the same question that discuss wheat's cultivated technology, so the answer must be same (Fig. 5). In our system, it provides two means for user to communicate with computer: the first means is keyboard communication. The second is voice communication, that is to say, a user query the computer by



Figure 5. It is Web-based consultant system interface. The “**询问**” (ask) edit box is used to put the query. The “**反馈**” (feedback) edit box is used receive the answer the system will give when the “**提交**” (submit) button is pressed

microphone and the computer will give the answer in voice. But in voice communication, there exists noise and then causes many ambiguous semantic sentences which it is difficult or impossible for the computer to answer them. On basis of AgriOnto base, we restore these sentences to their original before the computer processes them.

## 5. CONCLUSIONS

This paper presents ontology-based acquisition of agriculture knowledge. In practice, ontology-based knowledge facilitates the sharing and the application of agriculture knowledge. At last we give an example of the application of ontology-base knowledge. With further study of AgriOnto, much future work immediately suggests itself:

- A most acquisition of knowledge is still not automatic for natural language complexity. The rate of acquiring knowledge is very low. the ongoing automatic acquisition tool is still more work.

- We have done some work about knowledge analysis (Xie, 2007). The automatic evaluation algorithm of knowledge is necessary including insistency checking, incompleteness analysis and knowledge redundancy removing.

- Further research on intelligent agricultural knowledge-based service systems.



## ACKNOWLEDGEMENTS

This work is supported by Special fund project (2007211) for Basic Science Research Business Fee, AII, the Chinese Academy of Agricultural Sciences: Research on Grid-based Massive agriculture information technology. Thanks to Prof Cao of ICT, CAS for experiment platform.

## REFERENCES

- Cai SL 1996. The Volume of Agriculture in the Chinese Encyclopedia, China Encyclopedia Publishing House, 1996.
- Cao CG 1998. National Knowledge Infrastructure. A Strategic Research Direction in the 21st Century, Computer World, 1-3, 1998.
- Cao CG 2002. Progress in the Development of National Knowledge Infrastructure. Journal of Computer Science & Technology, Vol. 17 (No. 5) 2002, 523-534.
- Chaudhri VK, Farquhar A. et al. 1997. The generic frame protocol 2.0. SRI International Technical Report, 1997.
- Com 1998. The Editorial Committee of Agriculture: An Agricultural Dictionary, China Agriculture Press, 1998.
- Feng QZ, Cao CG, Si J, Zheng Y 2002. A Uniform Human Knowledge Interface to the Multi-Domain Knowledge Bases in the National Knowledge Infrastructure. the 22nd SGAI International Conference on Knowledge Based Systems and Applied Artificial Intelligence. Applications and Innovations in Intelligent Systems. pp. 163-176. 2002.
- Gu F, Cao CG, Sui YF 2003. A Domain-Specific Ontology of Botany. Journal of Computer Science and Technology, 2003.
- Lenat DB 1995. "Cyc: A Large-Scale Investment in Knowledge Infrastructure." Communications of the ACM 38(11): 33-38, 1995.
- Liu XH 1994: Automatic Inference based on Resolution Principle, 1994.
- Lu RQ, Cao CG 1990. Towards knowledge acquisition from domain books. In: Wielinga, B, Gaines, B., Schreiber, G., Vansomeren, M. (eds.): Current Trends in Knowledge Acquisition. Amsterdam: IOS Press, pp. 289-301, 1990.
- Natalya F. Noy, Mark A 1999. Musen. SMART: Automated Support for Ontology Merging and Alignment, Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management, Banff, Canada, July 1999.
- Si JX, Cao CG, et al. 2002. An Environment for Multi-domain Ontology Development and Knowledge Acquisition, in Proc. First International Conference, EDCIS 2002, LNCS 2480, Springer-Verlag, Berlin, Germany, 104-116, 2002.
- Tian W, Gu T, Cao CG 2002. Designing a Top-Level Ontology of Human Beings: A Multi-Perspective Approach, Journal of Computer Science & Technology, Vol. 17 (No. 5) 2002, 636-656
- Xie NF 2007. Agricultural Knowledge inconsistency Research. Agriculture Network Information, pp. 11-13, 2007.
- Zhang CX, Cao CG et al. 2002: A Domain-Specific Formal Ontology for Archaeological Knowledge Sharing and Reusing, in Proc. 4th International Conference PAKM 2002, LNAI 2569, Springer-Verlag, Berlin, Germany, 213-225, 2002.