Research on Knowledge Base Construction of Agricultural Ontology Based on HNC Theory

Hao Xinning, Xie Nengfu, . Sunwei, Zhong Xiaochun, Zhang Xuefu

Research on Knowledge Base Construction of Agricultural Ontology based on HNC Theory

Hao Xinning, Xie Nengfu, Sunwei, Zhong Xiaochun[a,*], Zhang Xuefu

Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing, 10081

[a]zhongxiaochun@caas.cn

Abstract. With the development of agricultural research and production, agricultural information and data processing technology have become increasingly demanding. In recent years, organizing and expressing semantic knowledge based on the form of ontology has become the focus of research in artificial intelligence field. And the construction of ontology-based knowledge base is the basic condition. HNC theory has great advantages in processing of Chinese character. However, complexity of the system has restricted its development. This paper summarized the existing research results, took rice production (cultivation) and processing as an example, proposed an idea and methods for the knowledge base construction of agricultural ontology based on HNC theory.

Keywords: agriculture ontology, HNC theory, knowledge base

1. INTRODUCTION

The rapid development of modern agriculture puts forward new and higher requirements for research on agricultural information. How to make a better use of modern information technology to innovate ideas and methods in the field of agricultural information science, and how to continuously promote research to enhance the performance of agricultural information have become an inescapable mission for research scientists in this area.

The rapid development of Internet, cloud computing, big data and other information technologies, and the continuous improvement of a variety of advanced mathematical theory spawn more IT research methods that are more convenient. Meanwhile, with the continuous advances in information science and technology, data management and information management have been difficult to fulfill modern demand for high-quality information. As a result, knowledge management has become one of the latest ways of information management. The main objective of knowledge management is to build domain knowledge in the form of knowledge networks, knowledge base and so on. This can reveal the relationship between semantic knowledge. Thus it is possible to form a more complete, proven and accurate knowledge platform and to support knowledge service, effective management and use of knowledge on the basis [1].

Nowadays, the mainstream technology of classification, retrieval, processing and filtering of agriculture information is still traditional key technology. However, this literal information retrieval system can handle only the literal meaning without carrying out more in-depth analysis, which leads to the lack of semantic understanding. Although it can make effective semantic analysis in artificial way to filter, process, classify and analyze information, the processing capacity and efficiency are insufficient to deal with the vast amounts of data generated daily. Therefore, only the thematic retrieval way, through the establishment of ontology knowledge base to describe relationship between domain

concepts in an accurate, clear and standard way and use abundant examples to fully describe the conceptual system of domain knowledge, can effectively improve the matching degree of retrieving objects and then make the best use of agricultural information resources in a fast way to achieve the knowledge management of agricultural information.

2. Basic technical concepts

2.1 HNC Theory

Hierarchical Network of Concept (HNC) is a theoretical system for the entire natural language understanding and processing, which is founded by Professor Huang Zengyang, Institute of Acoustics. The theory attributes semantic processing to conceptual representation. It uses primitive symbols and their combinations that highlight the relevance between concepts to represent conceptual semantic connotation. It represents the semantic content explicitly in natural language by mapping the natural language symbol system to HNC concept notation system. HNC concept notation system can describe lexical semantic, concept type and composition of the statement. The system can complete statement analysis of natural language through the concept analysis of statement and words [2]. HNC provides the basis for the computer to grasp the semantics by using HNC1, HNC2, HNC3 and HNC4 to digitize the word, sentence, sentence group and chapter. It contains not only the word level of knowledge, but also statement and context level of knowledge. The key point is to use HNC theory to process natural language and establish the mapping relationship between natural language and the concept space [3,4].

2.2 Ontology

Ontology is originated from philosophy area and it is a relative concept of Epistemology. Epistemology focuses on subjective perception, while ontology researches objective existence. So far, academic circles have not yet formed a unanimous conclusion on the definition of ontology.

Ontology uses its accurate, accepted concept definition to achieve specificity and semantic disambiguation of conceptual and terminology, and also completes the standardization and uniformity of reality concept definition. So it can guarantee the consistency between Human-Computer Interaction and machines and maximum the degree of realization of semantic disambiguation. The relationships between concepts, properties, instances and other components can achieve logical reasoning of knowledge concept as well as the description of knowledge conceptual system in the field. The relationships can effectively improve knowledge dissemination, sharing and retrieval efficiency in the aspects of semantics and pragmatics. It can also provide knowledge services under the big Data era background.

2.3 Ontology-based Knowledge Base

As a special form of knowledge, ontology aims to describe the facts and common vocabulary that are long-standing and unanimously approved by the knowledge workers. Knowledge base is aimed at describing the particular state of things that are related to facts and vocabulary, as well as the cognition state of knowledge workers. From another perspective, ontology refers to information that is unrelated to the particular state of things, while knowledge base refers to information that is related. From the

point of structure to analyze, ontology provides a set of terms and concepts to describe a field, while knowledge base use these terms to express the fact in this area.

In natural language processing, semantic analysis requires the support of huge knowledge base. Since the 1980s, a number of semantic knowledge base have been developed in China and abroad, such as the famouse WordNet, HowNet, Beijing University of CCD, etc. All of these can be called ontology-based knowledge base from the perspective of knowledge representation.

Ontology mainly contains ontology concept layer and ontology instance level. Ontology Concept layer is a definition on the concept within a specific range and the relationship between them, which mainly aims to reveal the rich relationships between the concepts. Ontology instance level refers to instances corresponding to a concept. It is a concrete expression of certain concept. Numerous and full instances will greatly enhance the capacity and application value of ontology knowledge base. Ontology knowledge base can illustrate domain concepts and their relationships in an accurate, clear and standard way and fully describe the conceptual system of knowledge in the field. So it is good at semantic description, logical reasoning, and revealing the hidden relationship between the concepts. As a result, it can effectively solve the unhandled problems of information retrieval and information sharing [5].

Ontology-based knowledge base can not only reveal the framework of the domain knowledge, but also help domain ontology function in concept unification, standardization, knowledge retrieval, knowledge sharing and other applications. Unlike the traditional information retrieval techniques, Ontology-based knowledge base will accurately locate the required knowledge; profoundly reveal the meaning of semantic information. And with the continued expansion of the boundaries of the domain ontology, the boundary between some areas becomes fuzzy and interdisciplinary emerges. As a result, the scope of knowledge domain that ontology knowledge base covers will continue to extend to make it play a greater role. Paying more attention to information semantic meaning analysis will be the future trend of information retrieval and information world development of association data and Semantic Web.

Different from the symbolic representation of HNC theory, ontology-based knowledge base can clearly express hierarchy and relationship between concepts to facilitate people's understanding and application. Meanwhile, ontology using formal description language can also be directly applied to natural language processing. If the concept expression in HNC theory can be represented by the general form of ontology, the accuracy of the human-computer interaction can be effectively improved [6].

3. Knowledge Base Construction of Agricultural Ontology based on HNC Theory

3.1 Ontology Construction Method

Ontology construction method mainly starts with the domain concepts and their relationship. With more suitable and mature ontology construction methods to guide the building process, we can effectively guarantee the consistency of domain ontology construction and also make ontology construction standardized and modularized. Duties, tasks and requirements in each part of the building process are clearly defined. There are no uniform, generally accepted principles of ontology construction [7-9]. In the ontology community, five principles proposed by Gruber are the most widespread:

(1) Clarity: Ontology provides concepts involved with clear, authoritative, accurate and standardized description. The definition of concept should be combined with specific areas and professional background. It should be objective, independent, authoritative and described formally with logical

axiom.

(2) Consistency: Logic rules of the ontology should be strict and rigorous. The definition of concept inferred by logical axiom should be correct.

(3) Extensibility: The design of ontology should go with the changes and development in the field. And it can be adjusted and extended continuously.

(4) Smallest coding error: Conceptualization should specify the level of knowledge, not varying according to different symbolic coding.

(5) Minimum ontology commitment: The establishment of the ontology needs to satisfy specific knowledge sharing needs. If the coverage of the ontology is too large, it tends to lead to low specificity and ambiguous concepts, which will result in losing the characteristics of ontology itself.

Currently, the mainstreams of ontology construction method are as follows.

(1) TOVE: It is a relatively new method of ontology, which regards demand problems and Completeness Theorem as the considerations of ontology construction. However, it lacks documented process description and specific steps description of ontology construction.

(2) Skeletal Methodology: The method provides a framework and guidelines for construction in each stage, requires a documented process and gives the steps of ontology assessment. Therefore, it is full of reference value. However, it lacks specific methods and techniques, and only provides guidelines for the development of corporate ontology.

(3) METHONTOLOGY: The method provides a description of the steps of ontology assessment and is suitable for the development of large ontology programs. It first proposes the concept of "writing specification" and also details the ontology construction tools, concept sources and concept extraction method. This promotes the standardization and normalization of ontology construction. However, it doesn't provide any ontology assessment method, so it can't evaluate the quality of ontology construction.

(4) Cyclic Acquisition Process: The main contribution of it is a new method using a cyclic structure for ontology acquisition. However, it does not provide details of guidance and technical explanation for specific method of ontology acquisition. Therefore, it is difficult to put the method into practice.

(5) "Seven Steps" by Noy and McGuinness: Because of its comprehensiveness and systematisms, it is favored by many people, including us. Combined with Protege software, the main steps are as follows:

1) Determine the field and scope of ontology;

2) Consider reusing existing ontology;

3) List important terms in ontology;

4) Define the class and its hierarchy;

5) Define the property of the class;

6) Define the restrictions of properties;

7) Create an instance.

3.2 Knowledge Base Construction of Agricultural Ontology based on HNC Theory

Compare with traditional English words, agricultural vocabulary is significantly professional, it is necessary to establish a new vocabulary base. Taking rice as an example based on HNC theory, this paper establishes an agricultural ontology knowledge base and realizes the establishment, management and update of knowledge base by Protege software. Since the terms in the HNC Theory based agricultural ontology knowledge base are clearly defined, accurately expressed and unified in the

conceptual level, it will not return duplicate or irrelevant results when retrieving. The architecture is shown in Figure 1.
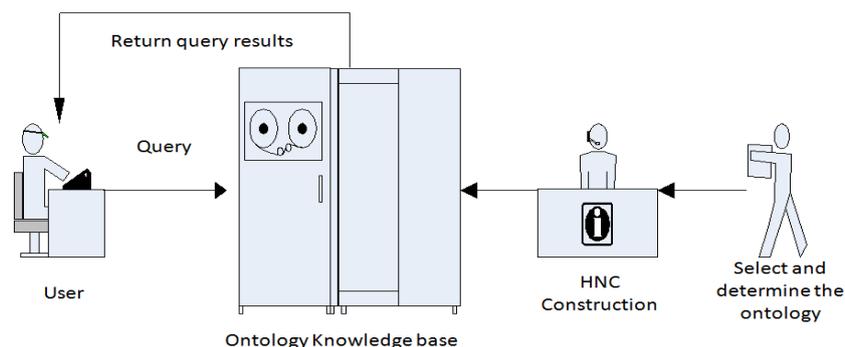

Figure.1. The structure of ontology knowledge base

HNC theory digitizes words, sentences, sentence groups and discourses, which provides a basis for computer to grasp the semantics. The key point is to establish a mapping relationship between natural language and concept space by HNC codes. The main techniques include Concept Matrix, Link expression, composite structure and so on. The building model is shown in Figure 2.
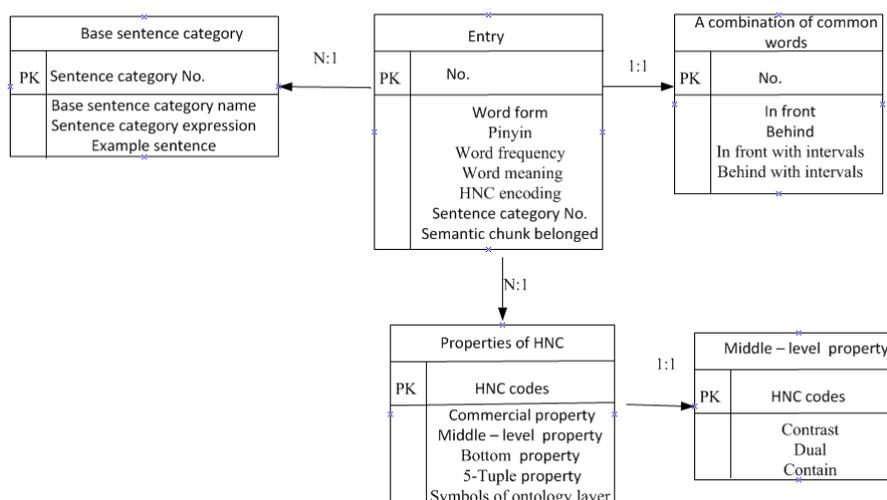

Figure.2. HNC based ontology knowledge base model

The ontology knowledge base shows the semantic description of words in six fields: word form, concept category, HNC symbol, semantic category belonged, relevant semantic category and synonymous appellation. The features are as follows. First, the words included are different from common words when considering the word form. The words may contain the form of non-Chinese characters, such as numbers, letters, etc. Second, for the same thing, the corresponding appellation may be more than one. Third, it focuses on the deep semantic relationship between words.

3.3 A case study-- rice production (cultivation) and processing

The construction of agricultural ontology knowledge base is a huge project. This research applied rice field as an example to briefly introduce the construction of ontology knowledge base. The overall framework is shown in Figure 3.
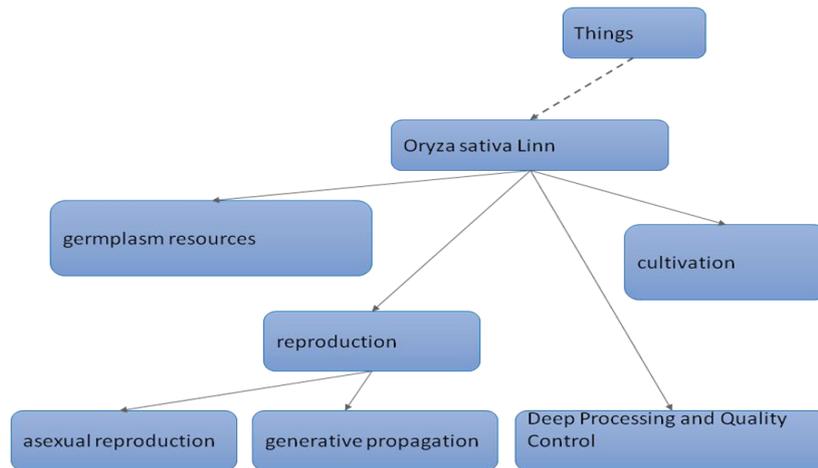
Figure.3. Overall framework of rice production (cultivation) and processing

The structure of ontology base contains 4 second-level directories and 10 third-level directories, Figure 4 is shown an example page of object properties in data processing, this experiment cumulatively includes more than 700 rice ontology terms, 108 object type attributes and 23 data type attributes to organize rice ontology knowledge base. In every single object properties, there are 4 different parts, every single part was to make a connection to an object property.
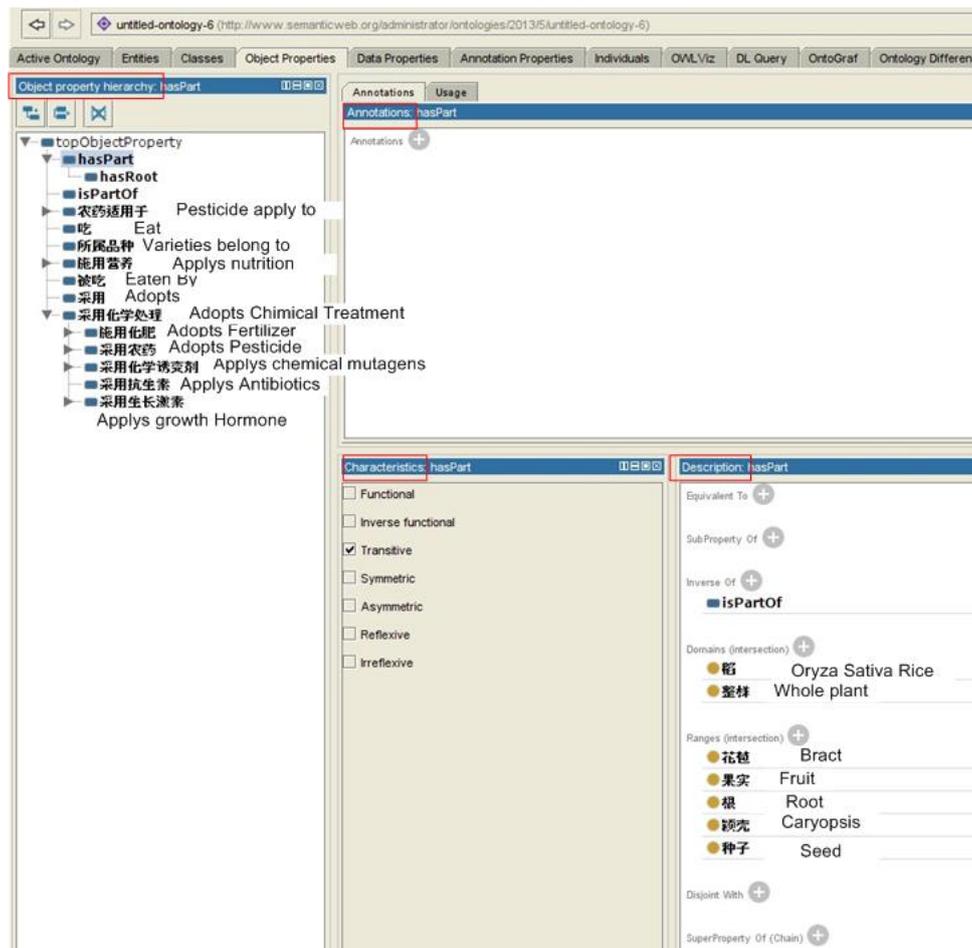

Figure.4. An example page of object properties in data processing

The final structure proposed rice production (cultivation) and processing ontology, including field cultivation techniques, rice harvesting, storage, processing and product, rice production machinery, rice varieties. The final result is shown in Figure 5.
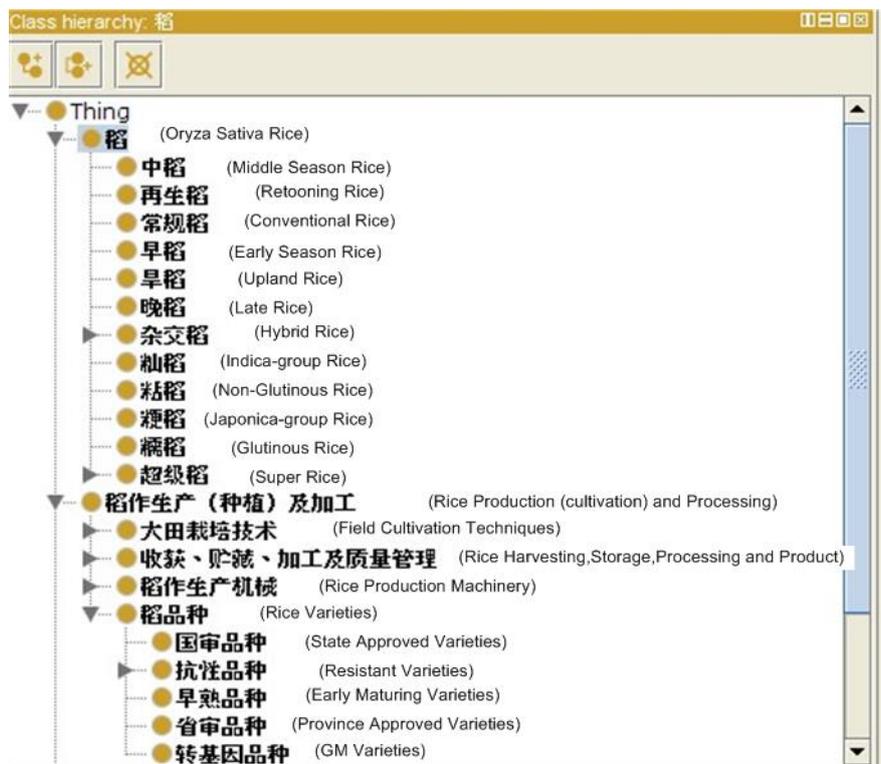

Figure.5. Rice ontology knowledge base

4. Conclusion and Outlook

By combining HNC theory with knowledge base construction, this paper proposed the vision of knowledge base construction of agricultural ontology based on HNC Theory. Taking rice production (cultivation) and processing as an example, it initially realizes intelligent query of information to effectively improve the precision and recall. However, for lack of unified systematic classification standards in the field of agricultural production, future research should pay more attention to the systematicity and rationality of classification. The design of HNC codes is relatively simple and its relevance is not sufficient. This remains to be applied in further study.

Acknowledgment

References

1. Zhou Yigang, A Review of Dynamic Knowledge Management Under Semantic Web[J].Journal of Modern Information,2014,(9).166-170

2. Huang Zengyang. HNC theory summary[J]. Journal of Chinese Information Process.1997,11(4).12-21

3. Huang Zengyang. HNC(Hierarchical Network Concepts)theory[M]. Beijing: Qinghua University Press.1998

4. Li Shanshan, Wang Yu. Retrieval of Periodical Literature Knowledge Element Based on Hierarchical Network of Concepts Theory[J]. Journal of Intelligence .2013,32(9).190-194

5. Zhang Shouxue,Yan Bo, Construction method of knowledge base based on ontology,Computer Era.2014,(4)45-46

6. Yang Jianyan. Construction and Application of Agricultural Word Knowledge base of HCN[J]. Beijing: The Second International Seminar on Computer and Computing Technologies in Agriculture " and "The Second China Development Forum of rural information".2008.10

7. Miao Chuanjiang. The research of HNC sentence category knowledge[D]. Beijing: Institute of Acoustics of Chinese Academy of Sciences PhD thesis.2001

8. JinYaohong. The language understanding technology and its application of Hierarchical Network Concepts (HNC)[J]. Journal of Yunan Normal University (Philosophy and Social Sciences). 2010,42 (4) .19-23

9. Wang Chuan, Liu Shangwang. Study on Construction of Ontology Knowledge Base for Wheat-weed[J]. College of Computer and Information Engineering, Henan Normal University.2014,42(6).138-142