

Ontology Molecule Theory-based Information Integrated Service for Agricultural Risk Management

Qin Pan

College of Economics Management, Huazhong Agricultural University, Wuhan430070, China
Email: panpan8170@163.com

Chunnian Liu and Dehui Yang

School of Information Engineering, Nanchang University, Nanchang330031, China
Email: pan81706@163.com

Abstract—Currently Chinese agriculture is facing a serious challenge of the natural risk, market risk and information risk, etc. At the same time, agricultural risk management is under the constraints of concept, capital investment, information sharing, and legal conditions. From the perspective of the information integrated services, this paper made efforts to provide some information and strategy support for the agricultural risk management. By applying a new theory in the field of the ontology and knowledge management, namely, the ontology molecule theory, this paper discussed the principles and process of developing domain ontology of the agricultural risks. An information integrated service model for the agricultural risk management, which consisted of six modules, was designed on the basis of the ontology molecule theory. The paper emphasized on discussing the function modules and the implementation of the integrated service model.

Index Terms—Ontology, Ontology Molecule, Information Integrated Service, Agricultural Risk Management

I. INTRODUCTION

In information society, the available information around people is often heterogeneous and distributed. In order to provide effective and convenient information sharing services to the users, we need to utilize information integration technology to integrate the information from various data sources and return the information to the users with a holistic manner. The purpose of information integration is to shield the heterogeneity of the underlying data sources to provide users with a holistic view. The traditional methods of information integration, such as materialized way and virtual view, can solve system-level, syntax-level and structure-level heterogeneity, but not semantic-level heterogeneity [1]. Since the ontology theory was proposed and widely used, researchers try to apply the ontology technology in information integration, which

has been proved to be effective in solving the problem of the problem of semantic heterogeneity. Ontology provides complete semantic models, meaning in the specified domain all related entities, attributes and base knowledge among entities have sharing and reuse characteristics which could be used to solve the problems of common sharing and communication [2][3]. Using the ontology-based information integration in knowledge management area, however, cannot solve the management of the dynamic knowledge and relative knowledge. For this issue, the ontology molecule theory can be used to successfully manage and control dynamic knowledge[4].

This paper apply the ontology-based information integrated services in the area of agricultural risk management, aiming to provide farmers and other users of agricultural risk information with effective and convenient information services to reduce the negative effect of agricultural risk. The organization of this paper is as follows: Section I introduces the ontology molecule theory and the process of building the agricultural risk ontology. A model of ontology molecule theory-based information integrated service in agricultural risk management, which consists of six main modules, is proposed in Section II. Then, in section III we briefly introduce the implementation of the model. Section IV concludes the paper and presents some future works for research.

II. THE ONTOLOGY MOLECULE THEORY AND THE CONSTRUCTION OF AGRICULTURAL RISK ONTOLOGY

A. The Ontology Molecule Theory

Ontology and ontology reasoning mechanism are core technologies of Semantic Web. With the development of the Semantic Web, ontology is widely used in various fields. Ontology can solve problems of knowledge organization, knowledge integration, et. However it encounters some problems in managing dynamic knowledge and dynamic information resources. The ontology molecule theory, which is based on the ontology

Manuscript received January 1, 2010; revised June 1, 2010; accepted July 1, 2010.

This work was supported by The National Science Foundation of China under Grant 70873046. Chunnian Liu is corresponding author

theory, description logic, graph theory and other theories, is a new theory used to implement the organization and management of dynamic knowledge[5].

Similar to the composition of the physical molecule, an ontology molecule consists of two parts, namely the nucleus and the ion [6]. For example, the "domain property" in a domain ontology, which is unchanging, can be seen as the nucleus of the ontology molecule. If the nucleus of an ontology molecule has changed, which indicates that the fundamental property of the domain ontology changes, then the original ontology could not exist. On the basis of the ontology's basic element (ontology instance, ontology triple), the ontology molecule, which is a balance between the ontology's basic element and the ontology base, can extend and deepen the existing ontology theory, and make the relatively coarse-grained knowledge management possible. By using the ontology molecule theory, the dynamic information resources in various areas can be managed and controlled. Besides, ontology reuse, which can be achieved in line with the molecule theory, involves building a new ontology through maximizing the adoption of pre-used ontology or ontology components [7]. When core concept of ontology does not change, we can reuse ontology by adding a variable part, namely the ion of the ontology molecule.

B. The Construction of the Agricultural Risk Domain Ontology

Gruber proposed five principles of ontology construction in 1995 as the follows: clarity and objectivity, integrity, consistency, maximum one-way scalability, the least constraint [8]. These principles have been widely applied to the ontology construction of various areas. Based on the above principles, we can conclude the following four main methods of building the domain ontology, including hand-built ontology by experts, conversion of existing thesauri, automatically building ontology from the database, and semi-automatic construction from the document content. Although at present the main methods of ontology construction is still hand-built, yet as many machine learning methods are gradually applied to the construction of ontology, semi-automatic and automatic building ontology have become possible, which can speed up the construction process of ontology and also save a lot of manpower and time.

In the management of the agricultural risks, we can build domain ontology to describe the agricultural risks. This ontology serves the objective of making the service system understand users' query more comprehensively and carry on query processing and document indexing on the semantic level. The process of building the agricultural risk ontology is as follows.

a. Analyze the information needs of the agricultural risk management, and establish the basic framework of the agricultural risk ontology according to needs analysis and guidance of experts in the field. In general, production, price and input risks were perceived as important sources of risk by the farmers [9].

b. Define concepts of domain ontology and properties of concepts. Key words were extracted from previous research literatures and other data sources to form concepts of ontology. Core concept in the field is considered nucleus of an ontology molecule, while relatively unimportant concepts as ion of an ontology molecule. For example, in domain ontology of agricultural risks, "risk" is nucleus of the ontology, which is unchanging. The properties of agricultural risks including main elements of risk, risk value, risk levels, risk probability, and so on [10]. Agricultural risk concepts are acquired by the following methods: professional dictionaries in agriculture, agricultural thesaurus and indexing records from history research papers, etc.

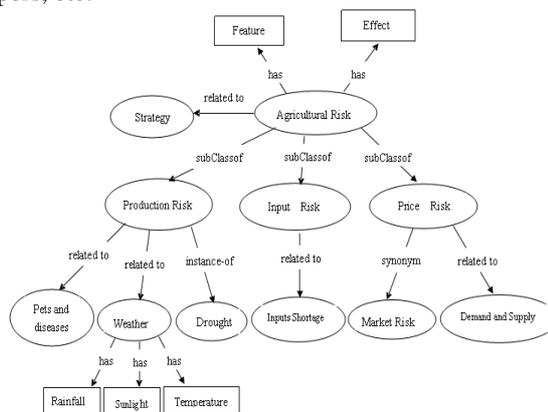


Figure 1. Frame of the agricultural risk ontology

c. Determine the relationship between the concepts and get a set of the relations between those concepts. there were four basic relations between the concepts: is-a (inheritance), part-of (part and overall), instance-of (example and concepts) and attribute-of (attribute relations). For instance, the relationship between the agricultural risk and the price risk is "subclass of".

d. Identify constraints and define the instances. In area of agricultural risks, constraints of concepts consist of time constraint, mutual exclusion constraint and mutual inclusion constraint. Defining corresponding instance of concept is similar to definition of the object of a class. For example, an instance of production risk is drought. The basic framework of agricultural risk ontology is shown in Figure 1.

e. Encode on the domain ontology and make it formal and standardized. In this paper, we use protégé [12] as the tool of ontology construction, and RDF as the description language of the ontology. At last, the domain ontology is encoded and formalized, and then taken out in the standardized language.

III. A MODEL OF THE ONTOLOGY MOLECULETHEORY-BASED INFORMATION INTEGRATED SERVICE FOR THE AGRICULTURAL RISK MANAGEMENT

According to the agricultural risk ontology built in the section I, in order to forecast the probability of the agricultural risks and the possible effect, based on agricultural risk ontology and ontology molecule theory,

this paper established a model of the information integrated services for the agricultural risk management. The architecture of the model is shown in Figure 2. Six function modules are included in the architecture.

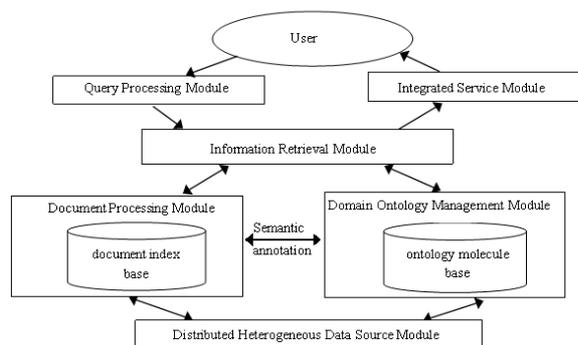


Figure 2. Architecture of the model of the ontology molecule theory-based information integrated service for the agricultural risk management

A. Query processing module

The main function of query processing module is to analyze and expand users' query, and then have query information stored into eigenvector arrays. In this module, the first operation is word segment. According the core glossary from domain ontology, key words are extracted and stored in the eigenvector arrays from the user input string. Then we can expand the words of users' query in the semantic level, mainly including synonym expanding, upstream expanding and downstream expanding. With query expansion, the user is guided to formulate queries which enable useful results to be obtained[13]. The expanding results are stored into different eigenvector arrays, which can be used to prepare for the information retrieval module.

B. Distributed heterogeneous data source module

The distributed heterogeneous data source module consists of various information resources, such as government policy on agriculture, demand and supply information of agri-food, risk prevention strategies etc, which are main data foundation of the whole framework of agricultural risk information integrated service. Meanwhile, the data sources mostly include database, web data, multimedia data, historical literature, XML documents, and so on. In the framework, the distributed heterogeneous data source module provides the data base for the ontology management module and document processing module.

C. Document processing module

The architecture of the document processing module is shown in Figure 3.

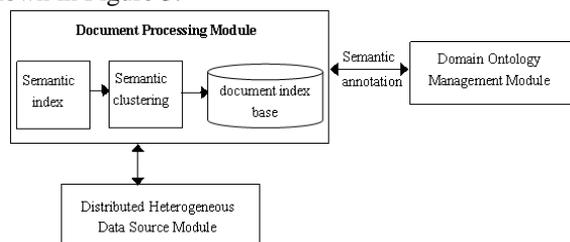


Figure 3. Document processing module

The document processing module deals with indexing and storage of document resources. Useful information, which is converted into entity information with certain structure, is extracted from raw data and non-structural document. This module mainly includes the following operations: semantic index, semantic clustering, and semantic annotation. Firstly, document is indexed in semantic level. Key words are extracted from data source documents based on core glossary of domain ontology and then indexed documents are stored for future use. Semantic clustering is made next. The goal of document clustering is to carry on cluster of different agricultural risk items. In addition to semantic index and clustering, semantic annotation[14] is carried on corresponds to core glossary of ontology to recognize the indexed documents.

D. Domain ontology management module

The structure of domain ontology management module is shown in Figure 4.

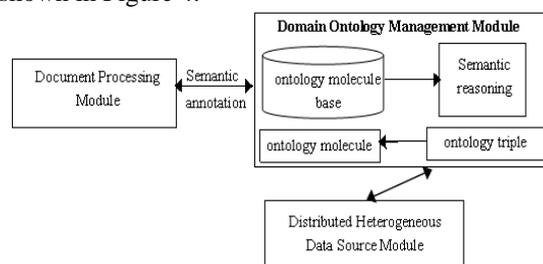


Figure 4. Domain ontology management module

The main function of domain ontology management module is to carry on ontology management, such as construction of ontology triples, ontology molecules and ontology molecule base, specification description, semantic reasoning, and concepts mapping etc..

Based on a variety of information resources in the distributed heterogeneous data source module, a class can be used to describe a particular type of thing and an instance to a certain thing. Then ontology triples are used to indicate concepts and relations among them. Ontology instances and relationships can be represented as nodes and edges. On the basis of the ontology triples, we also define a number of ontology molecules, which can be intuitively represented as aggregation of nodes and edges [15]. In accordance with the standards of W3C Semantic Web, using RDF/OWL as the ontology language, protégé as the ontology construction tools and OMProtegePlugin as the ontology molecule construction tool [16], we can build ontology molecule base. The ontology molecule base contains all content of the static triples in the ontology base, and also the dynamic knowledge, which cannot be expressed by the ontology base, including dynamic triples, triple identifiers and triple dimensions [17]. Ontology reasoning, which is a process of error detection and reasoning to the domain knowledge and domain relations, can check the consistency of ontology and support the semantic expansion for users' query, namely providing the users with a unified interface.

E. Information retrieval module

The architecture of information retrieval module is shown in Figure 5.

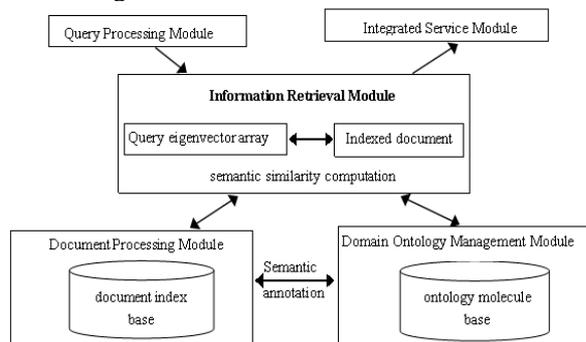


Figure 5. Information retrieval module

The main function of information retrieval module is combining corresponding data and giving an initial query result to integrated service module. One of the most important processes is to carry on semantic similarity computation between eigenvector array that takes out from query processing module and indexed document that takes out from document processing module. Before this, we can first compare every token in eigenvector arrays to speed up information retrieval process. Meanwhile, the building of the domain ontology can provides the specification description for the similarity computation. Then, we need to check whether the user query is the projection of some roles of the objects satisfying some certain constraints. The duplicate objects are eliminated for union, while the common ones are taken into account. Finally, the correlated result is returned to the integrated service module as the answer to the initial query.

F. Integrated service module

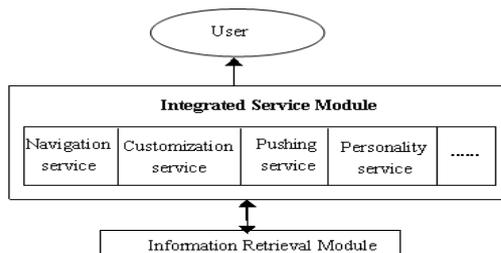


Figure 6. Integrated service module

According to level of information integration, information integrated service mode can be divided into two kinds: mode based on shared information and mode based on knowledge integration. The information integrated service based on built domain ontology is the key to solve problems of information integration in the syntax level [18]. Information sharing is the basis of the information integrated services, but not the ultimate goal. The ultimate goal of information integrated services is knowledge integration, which means providing users with comprehensive, decision-making and consulting information services according to effective integration and analysis of the distributed information. Knowledge integration can solve information integration issues in semantic and even pragmatic level.

Through the semantic integration and the pragmatic integration of information, the users of agricultural risk information, such as farmers and agricultural technicians, can conveniently receive the right query answer and effective information services. The information integrated services combine a variety of information service, such as navigation service, user customization service, pushing service, intelligent agent service, and personality service etc, so as to meet the user's diverse and individual needs. In this way, agricultural risk data and information can be used more efficiently, and the efforts of farmers to deal with agricultural risks can be improved effectively.

IV. IMPLEMENTATION OF THE MODEL

On the basis of the agricultural risk domain ontology, through the above proposed model, we can carry on various operations, such as semantic analysis and ontology reasoning, to analyze the query information submitted by the farmers or agricultural technicians. Then, related results were returned to the users in the form of information integrated services. The implementation process mainly includes the following steps:

Firstly, word segment and semantic expansion are carried on according to user query strings and domain ontology. For example, if the user wants to query "market risk strategy", we can divide the query into "market risk" and "risk strategy". Then based on semantic relations of concepts, such as synonymy relation, upstream relation, and downstream relation, we can find more related tokens: "price risk", "agricultural risk" and "demand and supply" etc. The expansion tokens are stored into query eigenvector arrays.

Secondly, semantic similarity computation between every segment in the query eigenvector arrays and the concept from indexed document base is carried on. Thus relevant documents are found.

Thirdly, we can integrate corresponding data and documents and give query answer to the user in an integrated service way. Besides, we can adjust service type and query results according to users' feedback.

V. SUMMARY AND OUTLOOK

As a new theory of dynamic knowledge management, the ontology molecule theory has its unique advantages when it is applied to ontology construction and information integrated services. First of all, the user's query can be effectively and comprehensively understood and analyzed in the semantic level. Then, a large number of distributed and heterogeneous information resources can be stored as ontology molecules in ontology molecule base to support the information integrated services for the agricultural risk management. This paper, by applying ontology molecule theory, first constructed agricultural risk ontology, and then established a model of information integrated service for agricultural risk management. This model can provide rich standardized expression, integration and retrieval for related entries of agricultural risk resources in semantic level. Our future

research is on basis of a more in-depth study of ontology molecule theory. And we develop an information service system for agricultural risk management to predict agricultural risks more accurately and reduce the loss of farmers caused by risks more effectively.

REFERENCES

- [1] Zhai J, Jiang JT, Yu Y, et al. Ontology-based Integrated Information Platform for Digital City [C].. In: 4th International Conference on Wireless Communications, Networking and Mobile Computing, Dalian, 2008,pp.1-4
- [2] Yang SY Developing an ontology-supported information integration and recommendation system for scholars[J]. *Expert Systems with Applications* (27) 2010, pp.1-15
- [3] He, Q ,Tang, Y , Huang, Y. Workflow model based on ontology[J]. *Journal of Computational Information Systems*. 2009,3 (3),pp.1153-1158
- [4] Dong H, Jiang Y, Gao J, et al. Research on Digital Library Based Ontology Evolution and Knowledge Management (I)—Ontology Molecule Theory[J]. *Journal of the China Society for Scientific and Technical Information*2009,28(3),pp. 323-330
- [5] Dong H, Jiang Y, Gao J, et al. Research on Digital Library Based Ontology Evolution and Knowledge Management (I)—Ontology Molecule Theory[J]. *Journal of the China Society for Scientific and Technical Information*2009,28(3),pp.323-330.
- [6] Dong H, Jiang Y, Gao J, et al. Research on Digital Library Based Ontology Evolution and Knowledge Management (I)—Ontology Molecule Theory[J]. *Journal of the China Society for Scientific and Technical Information*2009,28(3),pp.323-330.
- [7] Lonsdale D et al. Reusing ontologies and language components for ontology generation[J]. *Data & Knowledge Engineering* 2010,69,pp.318–330.
- [8] Li J Study on the Theory and Practice of Ontology and Ontology-based Agricultural Document Retrieval System—Floricultural Ontology Modeling. PhD Thesis, Center of Document and Information, University of the Chinese Academy of Sciences,2004
- [9] Meuwissen MPM, Huirne RBM, Hardaker JB Risk and risk management: an empirical analysis of Dutch livestock farmers[J]. *Livestock Production Science*2001,69,pp.43-53
- [10] Chen L Research on Design and Application of Domain Ontology of Construction Project Risk Management[J]. *Construction Economy* 2007,10,pp.99-101
- [11] Hu CP, Zhao Y An Ontology-based Framework for Knowledge Service in Digital Library [C].. In: 3rd International Conference on Wireless Communications, Networking and Mobile Computing, Shanghai, 2007,pp.5345-5348
- [12] Gennari JH, Musen MA, et al. (2003). The evolution of protégé: And environment for knowledge-based systems development[J]. *International Journal of Human-Computer Studies* (58), pp. 89–123
- [13] Bhogal J, Macfarlane A, Smith P A review of ontology based query expansion [C].. *Information Processing and Management* 2007,pp.866–886
- [14] Zhu HM, Tian QJ, Liang YQ, et al. Domain Ontology Component-based Semantic Information Integration [C].. In: 1st International Workshop on Education Technology and Computer Science, Wuhan, 2009,pp.101-104
- [15] Dong H, Jiang Y, Zeng J, et al. Research on Digital Library Based Ontology Evolution and Knowledge Management (III)—The Description of Dynamic Knowledge[J]. *Journal of the China Society for Scientific and Technical Information*2009,pp.643-650
- [16] Dong H, Jiang Y, Gao J, et al. Research on Digital Library Based Ontology Evolution and Knowledge Management (II)—Dynamic Knowledge Organization[J]. *Journal of the China Society for Scientific and Technical Information*2009,pp.483-491
- [17] Dong H, Jiang Y, Gao J, et al. Research on Digital Library Based Ontology Evolution and Knowledge Management (II)—Dynamic Knowledge Organization[J]. *Journal of the China Society for Scientific and Technical Information*2009,pp.483-491
- [18] Buitelaar P et al. Ontology-based information extraction and integration from heterogeneous data sources[J]. *Int. J. Human-Computer Studies*2008,pp.759-788



Qin Pan is 35 years old at present and is the Ph.D Candidate of agricultural economic management. His main research interests are involved in the application of computer and information technology in the field of agricultural risk management.