"十一五"国家重点图书

农业本体论研究与应用

当代农业学术专著系列丛书

钱平 郑业鲁 主编

中国农业科学技术出版社

论研究与应用

农业本体论研究与应用

钱平 郑业鲁 主编

图书在版编目（CIP）数据

农业本体论研究与应用/钱平，郑业鲁编著．—北京：中国农业科学技术出版社，2006．1

ISBN 7-80167-917-2

Ⅰ．农…　Ⅱ．①钱…②郑…　Ⅲ．计算机应用－农业－研究－中国　Ⅳ．S126

中国版本图书馆 CIP 数据核字(2006)第 008540 号

责任编辑：鱼汲胜

责任校对：马丽萍　张京红

整体设计：马　钢

出 版 者：中国农业科学技术出版社

北京市中关村南大街 12 号　邮编：100081

电话：(010)62145303　传真：68919709

发 行 者：中国农业科学技术出版社

北京市中关村南大街 12 号　邮编：100081

经 销 者：新华书店北京发行所

印 刷 者：北京佳信达艺术印刷有限公司

开　　本：880mm × 1230mm　1/16

印　　张：20.25

字　　数：460 千字

版　　次：2006 年 1 月第 1 版

印　　次：2006 年 1 月第 1 次印刷

印　　数：1 ~ 1 000 册

定　　价：116.00 元

《农业本体论研究与应用》编委会

前　言

本体论（Ontology）来自哲学领域，旨在研究客观事物存在的本质和组成。也有学者认为，本体论是关于世界某个方面或领域的一个特定的分类体系，这个体系不依赖任何特定的语言。近年来，随着信息科学的飞速发展，本体论逐渐用于知识工程和信息科学等领域之中。在知识工程领域中，本体就是通过对于概念、术语及其相互关系的规范化描述，勾画出某一领域的基本知识体系和描述语言。构建领域本体的目标是要将某个或多个特定领域的概念和术语规范化，为其在该领域或领域之间的实际应用提供便利。

从应用的角度来看，农业本体（Agriculture Ontology）可以被定义为“农业学科领域中一套得到认同的、关于概念体系的明确、正式的规范说明”。农业本体主要由农业知识中的概念、概念间的关系以及计算机可以识别的形式化描述语言组成。构建农业本体的目标是要形成对于农业信息组织结构的共同理解、认识并分析农业领域的知识，为进一步建立农业语义网络奠定基础。因此，农业本体论是关于用计算机语言规范农业知识概念表示、进行农业知识组织、开展农业知识服务的科学方法论。

网络技术和通讯技术的发展使计算机从单一的进行科学计算的设备演化成为在世界范围内进行信息交换和事务处理的网络入口。支持数据、信息和知识的交换、重用和共享成了当今计算机技术要迫切面临的任务。对知识工程、自然语言处理、信息检索系统、智能信息集成和知识管理、信息交换和软件工程等领域而言，由于本体在词法和语义上都比数据库所能表示的信息丰富得多，并且它不单单是一个存放数据的结构，而是

定位于提供一个领域严谨丰富的理论，描述事物或概念的运动和变化，因此本体是领域内重要实体、属性、过程及其相互关系形式化描述的基础。这种形式化的描述可以成为软件系统中可重用和共享的组件，从而可以实现不同系统之间的互操作和继承；也可以为知识库的构建提供一个基本的结构。

本体在很多领域都有广泛的应用前景。在信息资源整合中，本体提供了基于知识或概念的科学组织方法，深刻地揭示了概念以及概念间的相互关系，有利于进一步的知识发现。在本体支持下，知识搜索、知识积累、知识共享等知识服务的效率将大大提高，真正意义上的知识重用和知识共享也能成为现实。在知识管理系统中，本体就是一个有语义支持的词汇表，可以将领域知识的概念和相互间关系进行较为精确的定义。在企业逻辑建模中，由于本体适合表示抽象的描述，因此，可以让知识表达更加准确便捷，有助于进行更好的企业决策。

我国在本体论基础理论研究方面有一定基础，但是仅在少数单位开展，相互联系或合作非常有限；缺乏国家级专项支持用于系统的应用研究项目和实际应用的开发，系统性专著或成果较少；缺乏必要的本体论开发工具和平台，制约了应用系统的开发；总体上与国外差距较大。

需要强调指出的是，领域本体的构建需要领域专家和信息技术专家的共同努力和密切合作。本体的建立应该吸收前人长期积累的工作和经验。作为一种捷径，本体的建立可以在分类法、叙词表的基础上开展。相比较而言，叙词表采用科学术语，过于严谨，非专业人员难以掌握。分类法层次丰富、结构清晰、易于理解。国际上一般以叙词表为基础，如 FAO 的 AOS 项目以 AGROVOC 为基础。AGROVOC 叙词表目前虽然在国际上得到普遍的认同，却没有一部与之对应的在国际上有同样权威的分类法。就我国农业领域而言，中国图书分类法和衍生的农业专业分类表是普遍接受的分类体系。因此我们认为：以分类法为主、结合叙词表作为农业本体的基础，更加适合我国的实际。采用这种技术路线由于没有更多的经验可以借鉴，是一种探索和创新。

在我国，农业本体论研究正在成为农业信息领域研究中的热点，是知识工程中的新方向。农业本体论研究有助于进行农业科学的知识组织和发现，为系统地建立农业领域内的知识组织体系提供理论和方法，为进一步开展农业语义网研究和应用奠定基础。同时，为在农业信息领域开展数据挖掘、知识表示、自动标引、信息分类、智能检索、多语互译、知识发现等领域的研究和应用创造条件。促进国内外在农业本体论研究领域的合作，加强农业领域专家和 IT 专家的密切合作，将从总体上提升我国在农业本体论领域的研究水平，与国际接轨。我们长期从事农业科技文献信息检索技术和农业网络信息的分类与智能检索的研究，承担了“十五”国家攻关计划“农业信息化技术研究”中“农业信息智能检索与发布技术研究”研究课题，从事农业本

体论的研究及其在农业知识组织中的应用。该项研究2005年通过农业部组织的成果鉴定，并获中国农业科学院科技进步二等奖。同时参加过联合国粮农组织“农业本体服务计划”项目工作组会议和亚洲农业信息联盟年会，发表了论文，进行了学术交流。历时5年，从理论、方法和技术方面对农业本体论做了系统的研究，并在土壤学、花卉学领域进行了构建领域本体原型的研究，同时完成了中国农业科技文献信息智能检索系统和农业本体管理系统，在诸多农业领域得到应用。同时形成了我国第一个农业本体论的研究团队。

本书由基础篇、技术篇和应用篇组成，分别从基础理论、技术方法和实际应用三个角度介绍农业本体论的研究与应用。

基础篇：介绍农业本体论的定义与内涵、农业本体的基本要素、农业本体技术基础、农业本体相关标准、农业本体、主题词表和分类法、农业本体与农业语义网络等农业本体论的基本概念。并概述了国内外本体论的研究进展、联合国粮农组织的农业本体服务计划、我国农业本体论研究现状。

技术篇：从技术角度来说，农业本体的表示语言、构建方法和构建工具与其他领域没有本质的不同，两者都依赖计算机技术和信息技术的发展，所处理的对象都是以概念或知识元为最小单元。因此在本篇中，从通用高度比较了本体的主要表示语言，较为详细地介绍了描述逻辑语言、RDFS、DAML、DAML + OIL、CycL、OWL等表示语言；比较了本体的主要构建方法，对七步法做了详细介绍；比较了本体的主要构建工具，介绍了Protégé、OntoBroker、Ontolingua、OntoEdit；讨论了面向对象技术与构件技术、本体开发瀑布模型、本体库管理系统、构建语义网的关键技术等相关本体构建技术。

应用篇：介绍了农业本体论在农业信息组织、农业科技文献检索、家畜疾病诊断、花卉学本体模型构建中的应用以及开发基于本体论的知识管理平台等方面的实例。本篇的绝大部分内容是我们的最新研究成果。

此外，本书列出了国内外大量的参考文献、相关术语解释和相关网址，方便读者进一步了解国内外关于本体论方面的研究进展。参考文献包括本书各章节叙述中的引用文献，对引用文献不再重复列出。

中国农业科学院为支持本书的出版提供了农业科技专著出版基金。中国农业科学院农业信息研究所和广东省农业科学院情报研究所领导对本书的撰写和出版给予了大力支持。参与编写的同志们认真负责，付出了辛勤的劳动。本书是我国农业本体论研究团队的集体成果。

中国农业科学院农业信息研究所钱平研究员和广东省农业科学院情报研究所所长郑业鲁研究员担任本书的主编，中国农业科学院农业信息研究所苏晓路研究员、崔运鹏博士，广东省农业科学院情报研究所王众博士、李泽、何绮云和中国标准研究院李景博士承担了不同章节的写作。钱平、王众对本

书进行了统稿，崔运鹏对文稿进行了统排。

在过去5年的时间里，我们在农业本体论研究方面投入了巨大的精力与心血，期间遇到了不少困难，我们一方面积极争取各方面的支持，另一方面利用各种机会加强与各学术团体的交流，努力攻关，终于取得了阶段性成果。农业本体论研究作为新兴的研究与应用领域，其中亟待研究的问题还很多，需要相关部门的进一步大力支持和更多的有识之士加盟合作。我们希望我国农业本体论研究团队日益壮大，更希望我们的研究成果能为我国农业信息资源整合、开展农业知识服务作出积极的贡献。

本书通过进一步对已有成果进行总结和提升，系统介绍农业本体论的理论、方法、技术以及应用案例，有助于我国农业信息和知识的组织管理，提高我国在这方面的科技创新能力。同时，本书将是国内第一部关于农业本体论的理论与实际相结合的科学专著，反映我国在这一领域中的研究进展，及时填补了我国在这方面的空白。本书可以作为农业信息学科研究生的教材和从事农业信息组织和管理的研究人员的参考书。

联合国粮食与农业组织信息系统主管 Johannes keizer 博士为本书写了序言。进一步介绍了 AOS 计划的最新进展，不失为一篇很好的论文，为本书增加了光彩。在本书完成之际，我们对 Johannes keizer 博士先生表示感谢和敬意。

由于受时间和篇幅所限，我们无法对涉及农业本体论研究、开发和应用领域的内容和进展做出全面和详尽的介绍。本书引用了一些学者的论文或观点，出于保持全书系统性的考虑，在参考文献中一并列出，没有在引用处直接注明，希望得到各位学者的理解。本书中各种疏漏和错误难免存在，我们也期盼专家和读者给予批评指正，并期待在后续的版本中不断进行充实和改进。

Preface

The World Wide Web, since its creation, has revolutionized the way in which we publish, look for information and communicate. In the first years of its existence, it provided excellent opportunities for publishing and discovering vast amounts of previously inaccessible information. The new Web2.0 applications like Blogs, Wikis and social tagging tools have now brought an added dimension to the way people communicate with each other on the internet.

Nearly all information and communication on the web is based only on strings, which have to be interpreted by humans. Only by reading through a document or the result of a search, can people understand if this information is relevant to them or not. Computers cannot really understand the meaning of what we write and publish. As a result, search engines have difficulties in distinguishing relevant information from irrelevant information, performing language independent searches or putting a request in the relevant context.

To overcome these problems Tim Berners Lee, inventor of the Internet, proposed to go a step further in the development of a "semantic web"

"*The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users..... The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.*". — *Tim Berners Lee*

The ultimate goal is to make the information computer-processable rather than only human-processable.

Thanks to this advanced approach, we will be guided in our searches so that ambiguity will be minimized and the results will be more precise. We will be able to overcome language barriers, by being able to use our own language and the terminology we are familiar with, and we will be able to interrelate better information and formulate conclusions that would have taken time to be drawn or would not have even been noticed.

These new semantic technologies are allowing knowledge exchange and reuse, reasoning, interoperability, and easier data integration. They can be used everywhere, from schools to industries and in every domain, helping to bridge the linguistic and cultural barriers. The methodology for achieving this goal is the formal representation of the knowledge: giving more meaning corresponds to identifying and describing conceptual ideas and to represent them in a way that a computer can process them.

The goal of Food and Agriculture Organization① (FAO) is to combat hunger and poverty in the world. In doing so, FAO has to become a knowledge based organization that is able to catalyze the exchange of knowledge to achieve its goals. One of FAOs strategic objectives is "Improving decision-making through provision of information and assessments and fostering of knowledge management for food and agriculture", part of the STRATEGIC FRAMEWORK FOR FAO② (2000 ~ 2015). This makes a strong engagement in the area of knowledge exchange mandatory.

FAO has a normative role to play. It understands itself as an organization responsible for knowledge brokering. Knowledge and information can be brokered on vertical or horizontal lines only when common exchange standards exist. Before the explosion of the Web this problem had been mainly handled by the international Library community and its Z39.50 protocol③ for information exchange.

At FAO, we are therefore creating "ontologies" as a mechanism for the definition of concepts and their relationship. New computer languages such as Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL) have been created in order to represent logically the implicit information that could easily be understood by a human into a formalized form for computers. The already existing thesauri and other knowledge organization systems represent basic forms of ontologies and are being reused and enriched in order to create multilingual powerful ontologies. Ontologies are useful if better scoped and defined for a specific domain and application. Artificial Intelligence and Natural Language Processing techniques can be used to enrich ontologies, which in their turn can be used for knowledge discovery and extraction.

The multilingual AGROVOC thesaurus is available in the five FAO official languages which are English, French, Spanish, Chinese and Arabic. It is also available in Czech and Portuguese and Thai. Other languages such as German, Italian, Korean, Japanese, Hungarian and Slovak, are currently under translation and revision. The main role of AGROVOC is to standardize the indexing process in order to make searching simpler and more efficient, and to provide the user with the most relevant resources. Currently, it is downloaded about 5 times a month, an increase possibly attributed to its ease of applicability in multilingual web applications.

AGROVOC is the foundation that underpins the development of the Agricultural Ontology Service (AOS) project. By making use of knowledge contained in vocabulary systems and thesauri such as AGROVOC, AOS is committed to developing specialized domain-specific terminologies and concepts that will better support information management for the web environment.

---

① Food and Agriculture Organization of the United Nations http: //www. fao. org/

② Strategic Framework for FAO (2000 ~ 2015) http: //www. fao. org/docrep/X2575E/X2575E00. htm

③ The Z39.50 protocol website: http: //www. niso. org/z39. 50/z3950. html

Since Tim Berners-Lee has launched the debate about a future "Semantic Web"①, there is a growing interest in vocabularies and especially in structured controlled vocabularies among the community of Web Information Specialists. FAO has taken up the challenge by re-launching the AGROVOC thesaurus as a starting point for a basis for using ontologies in Agricultural knowledge applications. The first AOS Workshop②, which was held in 2001 in Rome, was aimed at transforming the AGROVOC Thesaurus into an Ontology server. Since then the approach has become much more phased and is directed at developing a collection of semantic tools which will serve as building blocks of the AOS project.

Since 2001, five Agricultural Ontology workshops have taken place, each increasing in the level of technical and scientific coverage of the topic. The workshops not only allow us to exchange knowledge about the semantic tools and trends but also have contributed to significant awareness of its importance to the community. Using AGROVOC, Domain Ontologies (DO) have been developed and are currently being used in some of FAO's applications③. We are currently at a junction where research meets real-world implementations. One of these real world implementations is an ontology based alert system on the depletion of fish stocks on which FAO is working with other partners, funded by a research project of the European Union.

In 2003, in collaboration with the Chinese Academy of Agricultural Science the Chinese version of AGROVOC was released. The knowledge that has been produced by China's scientists and farmers is of great value for other countries in the world, which struggle for food security. The existence of thesauri and ontologies which in Chinese will make it not only easier to retrieve Chinese knowledge for non-Chinese speakers, but also help to develop better usable tools for machine translation. Therefore, FAO and CAAS have embarked on a new project in mapping AGROVOC to the much more detailed Chinese Agricultural Thesaurus.

In 1995, the Online Computer Library Center (OCLC) organized the first workshop of the Dublin Core community with the aim of reaching a consensus on a core set of metadata elements to describe networked resources. This discussion was taken up in a meeting on agricultural standards, organised by Oneworld Europe④ in collaboration with FAO/GIL in Brussels, during autumn of 2000⑤. The meeting raised awareness amongst information providers of the new opportunities for sharing information through use of metadata standards and platform-independent formats such as XML. The participants a-

① The semantic web roadmap: http: //www. w3. org/DesignIssues/Semantic. html. The Semantic Web is comprised of the standards and tools of XML, XML Schema, RDF, RDF Schema and OWL.

② First AOS workshop website: http: //www. fao. org/agris/aos/Workshops/FirstWorkshop. htm

③ The Food Nutrition and Agriculture Journal: http: //www. fao. org/es/esn/publications/fna/index. jsp? lang = en

④ Oneworld website: http: //www. oneworld. net/

⑤ Information Management Standards for Electronic Publishing in Agriculture (Practical Workshop): http: //www. idmlinitiative. org/docs/agstandards. doc.

greed that a standard methodology was necessary and showed particular interest in the Dublin Core model.

More recently, in June 2003, UK Department for International Development (DFID) organized the "Fertile Ground" meeting, where policy makers and information service managers discussed the implications of the dispersed developments in exchange standards, tools and methodologies, thus introducing the notion of-and the need for- "coherence". Since the Brussels meeting in 2000, FAO has undertaken a number of initiatives to facilitate the standard setting process. These initiatives are under the umbrella of the project of an Agricultural Ontology Service for semantic Standards in FAO's subject areas.

Agricultural Metadata Standards Initiative (AgMES), launched in November 2000, is an attempt to promote the use of metadata through use of standardized agricultural metadata terms for the purpose of facilitating resource discovery and interoperability between richly described agricultural resources. It defines elements, qualifiers, encoding schemes and controlled lists that are necessary for the description of agricultural information resources. The use of such a standard makes it easier to integrate data from different sources allowing for creation of value-added services such as simple aggregated subject-based views, automatic news feed services etc.

Much of the activity has taken place in collaboration with internal and external partners. The AgMES Web site[①] has been continuously updated with useful information for the users. It is now furnished with official documentation, metadata schemas, metadata creation tool, glossary and Frequently Asked Questions. The Agstandards mailing[②] list provides a forum for discussions and exchange of ideas.

The AgMES initiative has received encouraging feedback internationally with a number of citations on well-known sites. Additionally, conference presentations and journal articles[③] have been written on the topic. Several applications, both within and outside FAO, have fully or partially implemented metadata schemas using AgMES elements (e. g. Open Knowledge Network, Consultative Group on International Agricultural Research, various FAO Portals).

One of the first AGMES applications was the AGRIS application profile[④]. The AGRIS Application Profile (AP) is a format that allows sharing of information across dispersed bibliographic systems. Most information systems, including AGRIS, are faced with the chronic problems of exchanging and aggregating information, starting from the differences in applications to those concerning varying database structures and cataloguing rules. The AGRIS AP is a major step towards exchanging high-

---

① Agricultural Metadata Element Set Web site: http: //ww. fao. org/aims

② Join the AgStandards Mailing list by sending an email to: http: //www. dgroups. org/groups/fao/agstandards/index. cfm? op = dsp _ join

③ http: //www. fao. org/documents/advanced _ s _ result. asp? FORM _ C = AND&SERIES = 339

④ The AGRIS Application profile: http: //www. fao. org/agris/tools/AGRIS _ AP/WhatItIs. htm quality and medium-complexity metadata in an application independent format. It provides possibilities to offer value-added services, irrespective of how the information was stored locally.

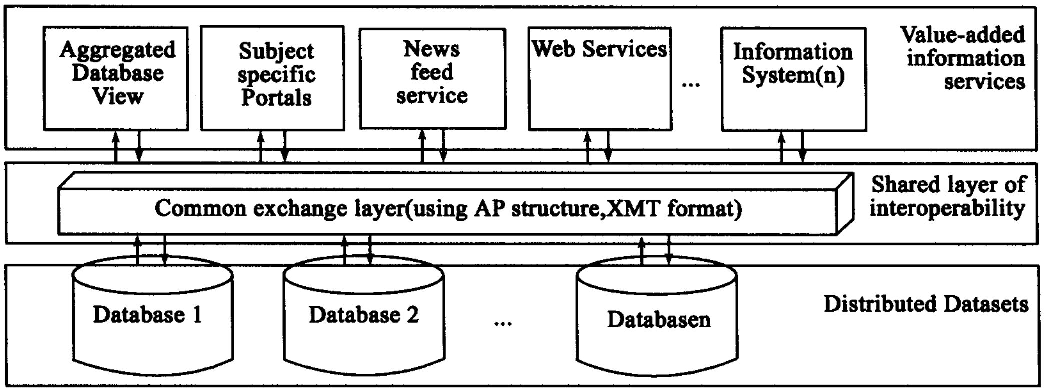

Figure 1 Interoperability, using common exchange standards, between distributed datasets allow for creation of value-added services

The result of introducing this standard can be measured quantitatively. This year, so far in the first nine periodical updates, 44105 records have been added; a significant increase from last year's total of 38532 records that were uploaded. One third of these records were submitted using the AGRIS AP XML format and WebAGRIS (as it is now available with the automatic XML-based export functionality). The most remarkable achievement of the new format is that we are now getting significant interest from some of the countries which had stopped sending in input to the AGRIS database, such as Finland and Norway.

The book brings together examples of typical applications of Agricultural Information resources management and all related aspects, including, sharing, knowledge discovery, and the support for advances semantics functions. It can serve as a valuable reference of case studies for other researchers working in the area of Agricultural Knowledge Engineering. FAO has a long history of collaboration with CAAS which began with the AGRIS network couple of decades ago while the more recent collaboration on the AOS concepts and ideas in China started in 2003. It started in the form of collaborative research in the area of agricultural terminologies and has since moved on to bigger initiatives such as the aforementioned mappings between AGROVOC and CAAS. The numerous benefits these collaborations have brought are examples of quality research and excellent collaborations.

I am very happy that FAO could work with Chinese scientists, who I am convinced, will make many important contributions to our common knowledge through improved semantic tools. I believe that the publication of the book will certainly promote research and application development using ontologies in China and the world wide promotion of the AOS concepts and what it brings specifically to the to the Agricultural Information community. The work FAO has done on development of Fisheries ontology, the Food, Nutrition and Agriculture Ontology and the early work on Food Safety Ontology are examples of some of the successes as well as the learning experiences we have had in the community.

Dr. Qian's work and his keenness to explore the new and upcoming technologies as well as his special interest in applying these concepts to the benefit of not only the Chinese agricultural community but also of the world. I would like to congratulate him on this endeavor and wish him all the best.

Last but not least, I owe thanks to my team for everything we have done in the last years here at FAO. In addition to their work in the team, I also would like to thank Margherita Sini and Gauri Salokhe for their valuable contributions to this preface.

Dr. rer. nat. Johannes Keizer

Information Systems Officer

Food and Agriculture Organization of the United Nations

目　录

第2篇 技术篇