

基于本体论的 农业知识管理

关键技术研究



崔运鹏 著

中国农业科学技术出版社

基于本体论的 农业知识管理

关键技术研究

崔运鹏 著

中国农业科学技术出版社

图书在版编目 (CIP) 数据

基于本体论的农业知识管理关键技术研究/崔运鹏著. —北京:
中国农业科学技术出版社, 2009. 1
ISBN 978 - 7 - 80233 - 705 - 3

I. 基… II. 崔… III. 农业经济 - 经济管理 - 信息管理 - 研究
IV. S126 F302.4

中国版本图书馆 CIP 数据核字 (2008) 第 143453 号

责任编辑 杨玉文

责任校对 贾晓红 康苗苗

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) 82109704 (发行部) (010) 82106631 (编辑室)
(010) 82109703 (读者服务部)

传 真 (010) 82106636

网 址 <http://www.castp.cn>

经 销 者 新华书店北京发行所

印 刷 者 北京华创印务有限公司

开 本 850 mm × 1 168 mm 1/32

印 张 8.75

字 数 220 千字

版 次 2009 年 1 月第 1 版 2009 年 1 月第 1 次印刷

定 价 28.00 元

版权所有 · 翻印必究

序

本书以农业知识管理为切入点,以农业知识管理的生命周期为线索,研究论述了知识的获取、组织、检索、标识、评价等关键环节涉及的技术,并针对其中一些关键环节开发了相应的计算机系统加以实现,最后作者还论述了基于本体的知识网格,是目前为止农业知识管理领域较为全面的一本论著。

本体论是近年来知识管理和信息管理领域一个比较热门的话题。本体论的本质是对问题域甚至整个世界建模,从而建立一套由计算机和人类都能够理解和处理的知识体系,并将所有的知识都纳入其组织范围,由此,计算机可以理解人类的思维方式,以人类的思维模式进行推理,换言之,让计算机能够像人一样思考。

在过去的几年中,信息资源建设成为了信息化建设的重中之重,从中央到地方的政府部门、科研机构、企业建成了大量专业数据库、信息资源库、网站、应用系统,使涉农信息资源数量迅速增长,但同时一些问题也不可避免地显露出来。由于缺乏统筹规划和相关标准,信息资源建设呈现出一种“各自为政”的局面,重复建设现象严重,信息孤岛大量存在,这些问题大大影响了信息资源的使用效率,信息资源的集成、共享和交换成为信息化建设的“瓶颈”。面对海量的信息资源,重新进行信息资源的梳理和加工几乎是不可能的。那么如何解决这一至关重要的问题呢?

答案是在现有的信息资源的基础上重新建立一个“层”,用于实现对现有信息资源的组织和梳理。这就需要找到一种机制,让计算机和人类都能够理解,并以近乎相同的方式对信息资源进行处理。这种机制就是本体论。本体论的定义是对概念化的精确描述,一个本体往往就是一个正式的词汇表,其核心作用就在于定义某一

领域或领域内专业词汇的含义及他们之间的关系。这一系列的基本概念如同构成一座大厦的基石，为交流各方提供了一个统一的认识。本体论可以实现知识和信息的规范描述和组织，就好像一张无形的网，用这些规范化的概念将信息资源实体“网罗”到一起，从而实现信息资源的共享、交换、重用及智能化的检索等功能。

从本体论目前的研究状况来看，它离实际应用还有一定的距离。主要表现在缺乏成熟的领域本体、还没有出现更深层次的应用、本体的构建也没有普遍适用的方法论等。但是，应该看到，本体论提供了一种解决问题的思路，并为信息服务向知识服务过渡以及实现真正意义上的知识服务带来了曙光，国际国内关于知识管理的研究仍在不断推进，新的技术及标准不断地涌现出来，越来越多的专家学者加入到这一研究行列中来，做出了大量具有实际意义的工作。本体论和语义网的研究既然被证明了其价值所在，就应该坚定地按照这一目标走下去，将知识管理推向新的高度，相信这一目标最终会实现。

孙素芳

北京市农林科学院农业科技信息研究所 所长

二〇〇八年十二月

内 容 提 要

从信息管理提升到知识管理是信息化发展的必然和提高信息服务质量的迫切需求。开展基于农业知识管理技术的研究,将为农业知识管理提供高效和科学的手段和技术支撑。本体论是起源于哲学领域的概念,近年来为信息技术领域使用并用于知识的组织和管理。本体的定义是“An ontology is a specification of a conceptualization”,即本体是对概念化的规范说明。本体论是关于基于本体的用计算机语言规范知识概念表示、进行知识组织、开展知识服务的科学方法论。运用本体论的理论、技术和方法进行农业知识管理,引起国内外广泛关注,成为学术研究的前沿和热点。

本研究的目的是设计基于农业本体论的完整的农业知识管理解决方案,重点研究农业知识组织、农业知识检索和农业知识评价的关键技术,同时对其他相关的关键技术进行较深入的研究。所采用的主要研究方法是,在整体方案设计阶段,通过用户需求调查,收集不同用户对农业知识管理的需求。在具体环节的关键技术研究时,结合主流技术,在本体论的基础上提出各个环节关键问题的解决思路,并在一些重点环节加以实现。

本研究在设计农业知识管理整体解决方案的基础上,针对农业知识管理的获取、组织、检索、表示、评价环节的技术做了深入细致的研究。此外,还对知识服务和知识网络进行了较为深入的研究。在农业知识获取方面,研究了目前网络信息获取的模式和方法,以及基于 RSS 的知识获取,利用 VB.NET 实现了基于 RSS 的知识获取的核心功能,论述了基于本体论和语义网的知识获取。在农业知识组织方面,研究了利用元数据、分类法和本体论实现知识组织的方式和方法。实现了基于元数据的农业知识的组织,制定了

农业科技信息核心元数据标准（ASTICM）及其扩展原则，设计实现了基于 ASTICM 的计算机著录信息管理系统。农业知识检索方面，研究了知识检索的一般模型以及基于本体的知识检索方法，实现了基于本体的农业科技文献知识的组织和智能检索平台，构建了一个基于本体的智能化猪病诊断系统。在农业知识评价方面，利用数理统计方法建立了领域本体和知识检索的评价模型，设计了一个基于层次分析法的动态指标管理系统，用于评价结果的计算和输出。在知识服务和知识网格的研究方面，比较了知识服务和信息服务的异同点，以及知识服务的模式和发展方向，论述了本体论和语义网技术与知识服务的关系；提出了一个基于本体论的 SOA 的农业知识网格架构。

本研究提出了基于本体论进行农业知识管理的思路 and 整体解决方案。对农业知识管理各环节的关键技术进行了深入的研究，实现了农业知识管理过程中利用元数据和本体进行知识组织、基于本体的知识检索和知识推理、利用数理统计方法进行知识评价等几个关键技术，并对其他环节的关键技术进行了探索和尝试。因为本体论的研究核心是提出知识组织和管理的理论、技术和方法，本研究所涉及的技术也都是国际上最新的主流技术，本研究具有前沿性和新颖性。整个研究将提升我国农业本体论的研究深度，促进其广泛应用，并与国际上相关领域的研究接轨。

ABSTRACT

To promote information management to knowledge management is a inevitable trend of the development of informationization, and it's also an urgent need for improving the quality of information service. To develop the research of agricultural knowledge management technologies will support and provide the efficient and scientific means for agricultural knowledge management. The concept of Ontology was originated from philosophy, and then it has been used in the field of information technology for knowledge organization and management in recent years. The definition of Ontology is: "An ontology is a specification of a conceptualization". Agricultural knowledge management based on the theories, technologies and methods of Ontology attracted widespread attention at home and abroad, and became the front and hotspot of academic research.

The purpose of the research is to design the complete agricultural knowledge management solution based on agricultural Ontology and study the other relevant key technologies. The emphasis of the research is about the key technologies of agricultural knowledge organization, agricultural knowledge search and agricultural knowledge evaluation, and other key technologies were studied. The method adopted in the solution design stage of the study is to obtain and collect the different users' different requirements about agricultural knowledge management through user investigation. And in the key technologies study stage, the whole thinking of the solution of every section based on main technologies and the thought of Ontology was put forward. And the technologies of several

key sections were implemented.

The acquisition, organization, searches, express of agricultural knowledge based on the whole solution were studied. Additionally, knowledge service and knowledge grid were also studied. In agricultural knowledge acquisition section, the patterns and methods of the network information acquisition, agricultural knowledge acquisition based on RSS were studied, the core function based on RSS knowledge acquisition was implemented in VB.NET language, the agricultural knowledge acquisition based on Ontology and semantic web was also discussed. In agricultural knowledge organization section, the patterns and methods of knowledge organization with metadata, taxonomy and Ontology were studied, respectively. The knowledge organization based on metadata was implemented, the core agricultural scientific and technical information metadata (ASTICM) standard and its application profile were constituted, and a registration information management system based on ASTICM was developed. In agricultural knowledge retrieval section, the common pattern of knowledge retrieval and also the retrieval based on Ontology were studied, an intellectual agricultural literature knowledge organization and search platform was developed, an intellectual pig disease diagnose system based on Ontology was developed to demonstrate the inference based on Ontology. In agricultural knowledge evaluation section, the statistical method was used to establish the evaluation models for both knowledge retrieval and knowledge evaluation. A dynamic indices management system was developed to calculate and output the results. In the field of knowledge service and knowledge grid, the differences between knowledge service and information service was compared, and the patterns and models of knowledge services were studied, an agricultural grid architecture based on SOA and Ontology was put forward.

ABSTRACT

The thought and solution based on Ontology were put forward in this research. The research covered the studies on the key technologies in every section of agricultural knowledge management, implemented the agricultural knowledge organization with metadata and ontologies, the knowledge retrieval and inference based on ontologies, and the knowledge evaluation with statistical methods independently. Also, the key technologies of other section were studied. Because the core of Ontology is to put forward the theories, technologies and methods of knowledge organization and management, the technologies involved in the research are all the latest technologies in the world, so the research is advanced and novelty. The entire research will promote the depth and wide application of the study of agricultural Ontology in China, and converge with the relevant international research.

目 录

第一章 引 言	(1)
一、问题的提出	(1)
二、研究的目的和意义	(5)
(一) 研究的目的	(5)
(二) 研究的意义	(6)
三、本研究的总体思路、研究方法与创新点	(8)
(一) 总体思路	(8)
(二) 研究方法	(8)
(三) 研究创新点	(9)
四、本研究的内容与技术路线	(10)
(一) 研究内容	(10)
(二) 研究技术路线	(11)
第二章 研究背景	(13)
一、知识管理和知识管理系统	(13)
(一) 知识管理	(13)
(二) 知识管理系统	(19)
(三) 知识管理技术	(24)
二、本体论	(26)
(一) 本体论的起源及发展	(26)
(二) 本体的构建	(28)
(三) 农业本体论研究现状	(40)
三、本体描述语言	(48)
四、基于本体的知识表示	(60)
五、语义网及相关技术	(61)

(一) 语义网的提出及发展	(61)
(二) 语义网的研究重点	(64)
六、元数据和元数据标准	(68)
(一) 元数据的概念和内涵	(68)
(二) 元数据标准的研究过程	(68)
(三) 元数据的著录	(68)
(四) 元数据的作用	(69)
(五) 国内外元数据标准的研究现状	(69)
七、相关计算机技术	(79)
(一) Microsoft. NET 框架	(79)
(二) J2EE 框架	(83)
(三) 数据库技术	(86)
(四) 软件工程	(88)
第三章 农业知识管理总体解决方案	(97)
一、知识管理的两种不同认识	(97)
二、农业知识管理解决方案的需求分析	(98)
(一) 目前农业信息管理中存在的问题	(98)
(二) 农业知识管理需求	(99)
三、农业知识管理系统总体框架	(101)
(一) 农业知识管理系统功能模块划分	(101)
(二) 农业知识管理系统结构	(109)
第四章 农业知识获取关键技术研究	(112)
一、网络信息自动获取技术研究	(112)
二、基于 RSS 的知识获取	(120)
三、基于本体和语义网的知识获取	(125)
第五章 农业知识组织关键技术研究	(128)
一、利用元数据实现农业信息的组织	(128)
二、利用分类法实现农业知识的组织	(131)
三、利用本体论实现农业知识的组织	(134)

四、农业科技信息核心元数据标准框架研究	(136)
(一) 农业科技信息核心元数据标准	(136)
(二) 农业科技信息核心元数据标准扩展原则	(140)
五、基于农业科技信息核心元数据标准的著录信息 管理系统	(145)
第六章 农业知识检索关键技术研究	(156)
一、知识检索的一般模型	(156)
(一) 一般的信息检索模型和基于知识的信息 检索模型	(156)
(二) 知识检索模型	(157)
二、基于本体的知识检索	(159)
三、基于本体的农业科技文献信息智能检索平台	(161)
(一) 检索策略	(162)
(二) 检索平台的结构	(164)
(三) 平台的功能	(165)
(四) 与常规检索的比较	(171)
四、基于本体的猪病诊断	(173)
第七章 农业知识表示关键技术研究	(184)
一、知识表示方法研究	(184)
二、基于主题图的知识表示	(188)
(一) 主题图介绍	(188)
(二) 主题图的细节	(191)
三、主题图工具	(204)
第八章 农业知识服务关键技术研究	(207)
一、知识服务与信息服务的比较研究	(207)
二、知识服务的模式及其发展趋势	(209)
三、本体论、语义网与知识服务	(213)
第九章 农业知识评价关键技术研究	(217)
一、领域本体的评价	(217)

二、知识检索评价	(223)
三、基于层次分析法的知识评价模型的计算机实现	(225)
第十章 农业知识网格关键技术研究	(233)
一、知识网格技术初探	(233)
二、本体论与农业知识网格	(239)
第十一章 讨论和展望	(242)
一、关于本研究的讨论	(242)
二、关于后续研究的展望	(246)
参考文献	(248)
致谢	(262)
英文缩略表	(264)

第一章 引言

一、问题的提出

在人类发展进化的过程中，人们坚持不懈地通过自己的努力改造着自然，改造着世界。在这个过程中，人类不断地通过各种可能的途径获取知识、积累知识、验证知识，并最终把知识用来改造我们赖以生存的自然，力图使自然与人达到和谐统一。因此，可以毫不夸张地说，人类社会发展的历史，就是人类获取知识、积累知识、验证知识、运用知识的历史。

由于近代工业革命和科技革命的兴起，人类改造自然的能力达到了前所未有的高度，同时，人类所积累的知识也在以难以置信的速度飞速增长。社会分工的细化，导致学科发展向深度和广度两个方向发展，这样，人类积累知识的速度越来越快，积累的数量越来越多。

知识的载体是信息，知识的积累也是通过信息的积累来体现的。要充分了解知识管理，必须了解知识与信息的关系。

外国学者马克·波拉特（Marc Porat）认为“信息就是经过组织与沟通的数据（马克·波拉特，1989）”。如果从信息构成要素分析，信息是由信息载体、信息符号和编码以及信息内容构成的，它既不是物质也不是能量，而是依附于自然界客观事物而存在的。例如文献，它的物理存在形式，即纸介质或光、电介质都是信息赖以存在的外部形式，也就是我们说的载体。载体用以表达特定含义的符号及其组合就是信息的符号和编码。文献所阐述的观点就是信息内容，它以抽象的方式依托载体和符号及其组合而成。

丹尼尔·贝尔（Daniel Bell）认为（Daniel，1998）：“知识是

一组对事实或想法经过组织的陈述，呈现出一个合理的判断或一项实验结果，这些判断和实验结果是以某种有系统的形式经由某种通讯媒体传给其他人。”也就是说，知识是人们在复杂的社会生产过程中，对客观事物长期观察、思考的结果，是对客观事物的抽象反映，是一种意识的东西，只有人们的大脑才能产生、识别和利用它，是人类认识世界的成果或结晶。它借助于一定的语言形式或物化为某种劳动产品的形式，可交流和传递，形成了人类共同的精神财富。

知识产生于信息，是经过系统化加工的信息，但与信息相比，知识呈现出两个明显不同于信息的本质特征。第一，知识只能为人类所拥有，最终只能是人来使用知识，知识与人是密不可分的，而信息可以存在于多种信息载体当中。第二，知识是处于发展变化当中的。它只有在知识劳动者对信息的应用中体现，并可能在交流中发生裂变、聚变从而创造出新知识（柴永红，2004）。史蒂芬·埃布尔伦（Stephen Abram）把“知识的创造及使用过程看作为从数据转变为信息、信息转变为知识及知识驱动与支持人类的行为和决策的一个连续过程的流程（Stephen Abram，1997）”。而信息是静态的，它不需要人的交流与应用也能存在。可见，信息是可编码的知识（即显性知识），是知识的有形体现（通过以书刊、文章或学习等与使用者有关联的产品出现），其价值在于它能够及时满足使用者的特殊要求。知识是信息以个人的角色、学习行为以及经验为内涵。当然，二者在适当条件下可以互相转化，即信息可通过学习、认知、过滤、评估转化为知识，知识被编码后转化为信息，这些信息又成为知识再产生的基础。二者的关系也符合唯物辩证法的对立统一规律。他们既互相区别，又互相联系，并在一定条件下互相转化。

知识总量的迅速增长导致了知识获取问题的出现。在浩如烟海的知识当中，人们如何获取自己所需的知识呢？众所周知，知识的主要载体是信息，那么怎样找到所需要的信息并把它转化为知识

呢?人们迫切地需要这样一种工具,帮助人们快速得到所需的信息并在最短时间内把它转化为可以储存在自己大脑里的知识。

信息的管理及服务是人们走出的第一步。

伴随着互联网的出现,为满足人们信息检索的需求,互联网上出现了很多提供信息检索服务的网站。典型的搜索引擎有 Google (<http://www.google.com>)、百度 (<http://www.baidu.com>)、雅虎 (www.yahoo.com)、爱问 (<http://iask.com>)、北京大学天网搜索引擎 (<http://e.pku.edu.cn/>), MSN 中文搜索 (<http://beta.search.msn.com.cn>), 现在发展为微软的 Live 搜索门户 (<http://www.live.com>) 等。这些网站都提供基于关键字的检索,用户通过输入一个或多个主题词(多个主题词可以用某种逻辑关系连接起来以实现更加复杂的检索),来得到用户所需的信息结果集。信息结果集中列出相关信息的标题、地址,并以超级链接的方式提供给用户,用户可以点击超级链接以查看具体的信息内容。

这种基于关键字的检索在互联网发展的初期发挥了巨大的作用。而随着用户信息使用水平的提高,新的问题出现了:一方面,人们不能通过基于关键字的搜索引擎精确地查到自己所需的信息;另一方面,人们也不能通过基于关键字的搜索引擎查到所有所需的信息及其相关信息。

上述两个问题分别称为“查准率不高”和“查全率不高”。

查准率 (precision ratio) 是衡量某一类文献检索系统的信号噪声比的一种指标。它的数值等于 w/m , 式中 w 是用户鉴别检出的 m 篇文献时,认为实际满足要求的文献篇数。这一指标最初是 1956 年由 J. W. 佩里、A. 肯特等人提出的 (<http://www.wiki.cn/wiki/%E6%9F%A5%E5%87%86%E7%8E%87>)。

查全率 (recall ratio) 衡量某一情报检索系统从特定文献集合中检出相关文献成功度的一项指标。它的数值等于 w/x , 式中 w 为用户鉴别检出的 m 篇文献时,认为实际满足要求的文献篇数, x 为特定检索系统中所包括的全部 n 篇文献中实际与某一课题相关