



冯志勇 李文杰 李晓红 编著

本体论工程及其应用

清华大学出版社

Ontology

本体论工程

——连接信息系统和哲学之间的纽带



TP3-05

7

2007

冯志勇 李文杰 李晓红 编著

本体论工程及其应用

清华大学出版社
北京

内 容 简 介

本书是一本讲述本体论工程及其应用技术的专业书籍,在设计和构思上力争为读者提供全面、系统且翔实的内容,对本体、本体论、语义网及其相关技术和应用领域都进行了详细的讲解。

该书通过深入剖析本体论理念在信息系统中所表现出的优越品质,对本体理念等进行了介绍和描述,希望能够为信息系统中本体论的建设提供初步的理论和实践的指导。

全书由本体理念概述、本体的形式化描述、本体描述语言、本体的实现和运算、本体论工程方法学、领域本体知识库、本体论的应用、本体编辑工具 8 章组成。最后还给出本书中专业术语的英汉对照表。

本书可作为计算机、通信专业的高年级本科生和研究生的专业课教材或参考书,还可作为相关领域有一定研究基础的专业人员的参考书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

本体论工程及其应用/冯志勇,李文杰,李晓红编著. —北京:清华大学出版社,2007.5
ISBN 978-7-302-14627-8

I. 本… II. ①冯… ②李… ③李… III. 计算机科学—教材 IV. TP3

中国版本图书馆 CIP 数据核字(2007)第 015885 号

责任编辑:张瑞庆 孙建春

责任校对:梁 毅

责任印制:李红英

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

c-service@tup.tsinghua.edu.cn

社 总 机:010-62770175

邮购热线:010-62786544

投稿咨询:010-62772015

客户服务:010-62776969

印 刷 者:北京四季青印刷厂

装 订 者:三河市李旗庄少明装订厂

经 销:全国新华书店

开 本:185×230 印 张:11 字 数:231 千字

版 次:2007 年 5 月第 1 版 印 次:2007 年 5 月第 1 次印刷

印 数:1~3000

定 价:22.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:010-62770177 转 3103 产品编号:024084-01

PREFACE

Across the Great Wall we can reach every corner in the world

——越过长城,走向世界(CNNIC,2003)

本体理念的渊源可以追溯到对世界的本质进行描述的哲学领域。在信息系统领域,本体论目前主要用于知识的表示、组织和管理。本质上,本体提供的知识共享方式可适用于任意基于知识的系统,因此本体论理念不仅在人工智能和知识工程领域引起研究人员的兴趣,而且在信息系统的很多领域都正在得到广泛应用,例如智能信息集成、知识管理、自然语言处理、信息检索系统、电子商务,以及传统的软件工程等方面。

尤其最近,本体论在万维网上的应用直接导致了语义网的诞生。创建语义网的主要目的是解决 Web 信息共享中的语义问题,这给 Web 的未来发展带来了勃勃生机和无限遐想。语义网得以研发的主要原因之一就是当前对基于网络的知识交换和知识共享存在着巨大的需求与期望,因此,随着各种已有的和将要研发的 Web 应用的逐渐增多,人们不仅希望不同的系统能够共享某个共同公认的知识库或者词汇(术语)集合,而且希望提供统一的领域模型,这样将会极大地促进和实现不同信息系统之间数据或知识的交换、共享和复用。

因此,支持数据、信息与知识的交换、共享和复用已成为当今的信息系统要迫切面临的任务之一。本体的本质就是领域知识的共享和复用,此外,标准且形式化的领域本体能够为信息系统之间的高层互操作提供很好的工具,即本体的最大贡献在于规范某个或多个特定领域的概念和术语,为其在该领域或领域之间的实际应用提供便利,同时,形式化的描述可成为信息系统中可复用和共享的组件。在这样一系列理念的支持下,本体论将会大大提高人们对知识的搜索、积累和共享的效率,从而使真正意义上的知识共享和知识复用成为现实。

尽管本体论是信息系统领域中引入的新理念,但它和现有的方法与技术有着密切的联系。本体可以将传统的面向对象方法中对象知识的概念及其之间的关系约束进行

较为精确的形式化定义。在本体论工程和软件工程学科之间存在着很多潜在有用的相似性,而且,软件工程学科和本体研究者之间很少讨论的很多内容都会有助于高级本体论工程的开发,例如设计模式。此外,软件体系和软件过程也都能归并入本体论工程框架中,因为在此框架内,软件体系可以采用静态本体论进行表示,而软件过程可以采用动态本体论进行描述。

作为分布式异构环境中的知识体系,本体论为知识的语义理解、计算机理解,以及人机交互提供了应用框架和实现途径。虽然本体论在人工智能和知识表示中已经被广泛的讨论与实践,但目前我国对本体论理念的很多研究和应用都还处于不成熟的探索阶段,还仅仅局限在概念的界定和初步的理论研究层面上。此外,与传统的软件工程相比,目前的本体论开发方法学还未能成为“科学”或“工程过程”的完整方法论。因此,本书在深入剖析本体论理念在信息系统中所表现出的优越品质的基础上,对本体理念进行详细的介绍和描述,希望能够为信息系统中本体论的建设提供初步的理论和实践的指导。

由于本体论理念及其相关技术的优越特性,国内外对其研究的兴趣日益浓厚,使得其在很多领域都得到广泛的研究和应用,但是与此相关的资讯过于分散且没有系统化,所以相关领域的专业研发人员迫切需要相应的专业书籍对其提供理论和实践上的指导与帮助。因此,通过阅读本书可以从中获悉比较全面具体的信息,缩短搜集相关资讯的时间,节省研发人员的人物财力,为了解和掌握本体论理念及其相关技术提供便利。此外,本书的编写也会在一定程度上为本体论及其技术全面推广和应用,以及未来语义网的实现做出应有的贡献。

本书在设计和构思上力争为读者提供更加全面、系统且翔实的内容,把现有的资料按照特定的次序进行归纳与整理,对本体、本体论、语义网及其相关技术和应用领域都进行了详细的讲解,使得本书在内容安排和组织上有条不紊。尽管各个章节之间有着必然的关联,但是每部分的内容却都是独立成章,这符合循序渐进的学习过程,也可针对特定问题参阅相应的章节。

本书面向于计算机、通信专业的高年级本科生、研究生和有一定研究基础的专业人员。

本书的结构

本书的内容安排如下:

第1章简单概述本体理念。给出本体从哲学领域引入到信息系统领域后定义的演变,并采用不同的标准对本体论进行分类,描述本体论工程与软件工程等学科的相似之处,同时介绍本体的功能、特点,及其与普通术语的区别。

第2章对本体的形式化进行详细的数学描述。介绍本体形式化的三个定义,并给

出本体内的代数系统中五种最基本的建模基元,以及代数系统中概念之间四种最基本的关系。

第3章介绍描述本体的语言。在分析XML、RDF和描述逻辑的基础上,简要介绍六种本体描述语言,其中着重讲解面向语义网的OWL语言及其基本功能、元建模体系与子语言,并比较不同的本体描述语言的功能和特点。

第4章对本体理念的实现与运算进行描述,即分别介绍语义网和本体知识库的操作。在概述语义网理念的基础上,借助于语义网的体系架构和知识后勤理念及其相关技术来讲解系统知识的层次化体系架构,描述与语义网相关的工具与技术,简述当前知识库系统的建模架构,并通过数据集成和知识融合来对本体知识库的操作进行描述。

第5章概述本体论工程的方法学。详细描述本体论的开发过程的七个步骤,对五种本体论工程方法学的特点进行对比,并介绍本体管理阶段的主要需求,同时简要叙述本体论开发过程中的命名惯例。

第6章对领域本体知识库的构建进行概述。采用本体理念及其技术详细地描述应用领域,以便建立不同抽象与粒度级别的领域本体知识库,并把领域本体表示成有向无环图或词汇概念图,介绍术语之间四种主要的语义关系,给出构造领域本体的五条规则,同时描述构造新的领域本体的不同方式,提出困扰领域本体建模的两个问题,讲解集成并复用现有知识基的两种方式,知识基之间共享知识需要满足的两种规则,以及建立知识共享的使能技术所需的三个构件,简单介绍领域本体的应用案例。

第7章介绍本体论及其相关技术在信息系统领域中的广泛应用。对本体论在分布式信息系统、人工智能、软件系统、多主体系统和计算机辅助工艺设计系统等领域的应用进行表述,并详细描述本体论理念在信息检索系统中所表现出的优越特性,不但可以弥补现有搜索引擎的缺陷,而且还为用户提供更加个性化且智能化的信息检索服务。

第8章对本体编辑工具的特点进行概述,从设计目标、功能特性和框架结构等方面介绍八种编辑本体的工具,同时从系统评估框架的通用性、协作性和本体论工程等角度对本体编辑工具进行对比。

致谢

本书在编写过程中参阅并借鉴了国内外大量的文献资料和网络资源,特向作者表示最诚挚的谢意。清华大学出版社对本书的出版给予了大力的支持,蒋宗礼教授仔细审校了本书的初稿,并提出许多宝贵的修改意见和建议,在此一并表示衷心的感谢。

由于本书的专业性强,内容的涉及面广,加上时间紧迫以及作者的编写水平有限,因此书中难免有错误或不当之处,敬请同行和读者朋友们批评指正。

编者

2007年1月

读者意见反馈

亲爱的读者：

感谢您一直以来对清华版计算机教材的支持和爱护。为了今后为您提供更优秀的教材，请您抽出宝贵的时间来填写下面的意见反馈表，以便我们更好地对本教材做进一步改进。同时如果您在使用本教材的过程中遇到了什么问题，或者有什么好的建议，也请您来信告诉我们。

地址：北京市海淀区双清路学研大厦 A 座 602 计算机与信息分社营销室 收

邮编：100084

电子邮件：jsjic@tup.tsinghua.edu.cn

电话：010-62770175-4608/4409

邮购电话：010-62786544

教材名称：本体论工程及其应用

ISBN: 978-7-302-14627-8

个人资料

姓名：_____ 年龄：_____ 所在院校/专业：_____

文化程度：_____ 通信地址：_____

联系电话：_____ 电子信箱：_____

您使用本书是作为：☐指定教材 ☐选用教材 ☐辅导教材 ☐自学教材

您对本书封面设计的满意度：

☐很满意 ☐满意 ☐一般 ☐不满意 改进建议_____

您对本书印刷质量的满意度：

☐很满意 ☐满意 ☐一般 ☐不满意 改进建议_____

您对本书的总体满意度：

从语言质量角度看 ☐很满意 ☐满意 ☐一般 ☐不满意

从科技含量角度看 ☐很满意 ☐满意 ☐一般 ☐不满意

本书最令您满意的是：

☐指导明确 ☐内容充实 ☐讲解详尽 ☐实例丰富

您认为本书在哪些地方应进行修改？（可附页）

您希望本书在哪些方面进行改进？（可附页）

电子教案支持

敬爱的教师：

为了配合本课程的教学需要，本教材配有配套的电子教案（素材），有需求的教师可以与我们联系，我们将向使用本教材进行教学的教师免费赠送电子教案（素材），希望有助于教学活动的开展。相关信息请拨打电话 010-62776969 或发送电子邮件至 jsjic@tup.tsinghua.edu.cn 咨询，也可以到清华大学出版社主页（<http://www.tup.com.cn> 或 <http://www.tup.tsinghua.edu.cn>）上查询。

目 录

CONTENTS

第 1 章 本体理念概述	1
1.1 本体的概念	1
1.2 本体论的分类	3
1.2.1 概念分类理论	3
1.2.2 本体论分类规则	4
1.3 本体论工程的相似性	6
1.3.1 软件工程学科	6
1.3.2 其他学科	10
1.4 本体的功能、特点及区别	11
1.4.1 本体的功能	11
1.4.2 本体的特点	12
1.4.3 本体与一般术语的区别	13
1.5 本章小结	15
思考题	15
第 2 章 本体的形式化描述	16
2.1 本体的形式化定义	16
2.1.1 OWA 形式化定义	16
2.1.2 Guarino 形式化定义	17
2.1.3 KAON 形式化定义	18
2.2 本体内的代数系统	20
2.2.1 建模基元	20
2.2.2 概念之间的关系	20

2.3 本章小结.....	21
思考题	21
第3章 本体描述语言	23
3.1 本体语言概述.....	23
3.2 XML 和 RDF	24
3.2.1 XML 简述	24
3.2.2 RDF 简述	26
3.2.3 XML 与 RDF 的关系和区别	29
3.2.4 两者的不足之处	31
3.3 描述逻辑.....	31
3.3.1 描述逻辑简述	33
3.3.2 描述逻辑的语义表达式	34
3.4 本体描述语言的种类.....	35
3.4.1 简单 HTML 本体扩展	36
3.4.2 本体标记语言	36
3.4.3 基于 XML 的本体交换语言	37
3.4.4 本体交互语言	37
3.4.5 DAML 语言的新版本——DAML+OIL 语言	39
3.4.6 Web 本体语言	39
3.5 本体描述语言之间的比较.....	45
3.6 本章小结.....	46
思考题	46
第4章 本体的实现和运算	48
4.1 语义网.....	48
4.1.1 语义连续体	50
4.1.2 语义网的体系架构	52
4.1.3 语义网应用系统开发工具 Jena	55
4.1.4 语义网服务标记语言 OWL-S	55
4.1.5 语义网查询语言 OWL-QL	57
4.1.6 语义网挖掘	58
4.1.7 语义网和本体的比较	60

4.2	本体知识库的操作	60
4.2.1	知识库系统的建模架构	61
4.2.2	数据集成	62
4.2.3	知识融合	64
4.3	本章小结	68
	思考题	68
第5章	本体论工程方法学	70
5.1	本体论开发概述	70
5.2	本体论开发步骤	71
5.2.1	确定本体的领域和范围	72
5.2.2	考虑现有本体的复用	73
5.2.3	枚举本体的重要术语	73
5.2.4	定义类和类层次	73
5.2.5	定义类的特性	76
5.2.6	定义属性的约束	76
5.2.7	生成实例	77
5.3	本体论工程方法学	78
5.3.1	骨架法	78
5.3.2	评估法	79
5.3.3	Bernaras 法	79
5.3.4	METHONTOLOGY 法	80
5.3.5	SENSUS 法	80
5.4	各工程方法学之间的比较	81
5.4.1	IEEE 1074-1995 标准	81
5.4.2	方法学之间的对比	81
5.5	本体管理需求	83
5.6	命名规则	83
5.6.1	大小写和分隔符	84
5.6.2	单复数	84
5.6.3	前后缀	84
5.6.4	其他规则	84
5.7	本章小结	85

思考题	85
第 6 章 领域本体知识库	86
6.1 领域本体知识库概述	86
6.2 领域本体知识库的构建	88
6.2.1 领域本体的表示	88
6.2.2 领域本体的构造	90
6.2.3 领域本体的建模	92
6.2.4 领域本体的复用	93
6.2.5 领域本体的应用	95
6.3 本章小结	96
思考题	97
第 7 章 本体论的应用	98
7.1 信息系统领域	98
7.1.1 分布式信息系统领域	100
7.1.2 人工智能领域	101
7.1.3 软件领域	102
7.1.4 多主体系统领域	103
7.1.5 计算机辅助工艺设计系统领域	106
7.2 智能信息检索系统	113
7.2.1 智能信息检索系统的目标	115
7.2.2 基于搜索引擎的信息检索系统	116
7.2.3 基于本体论的智能信息检索系统	117
7.3 基于领域本体的个性化智能信息检索系统	119
7.3.1 个性化智能信息检索系统的体系架构	120
7.3.2 信息处理中心	121
7.3.3 个性化偏好库	122
7.3.4 本体索引中心	124
7.3.5 算法库	125
7.3.6 本体转换中心	126
7.3.7 领域本体知识库	129
7.3.8 个性化智能信息检索系统的实现	129

7.4 本章小结	132
思考题.....	132
第8章 本体编辑工具	134
8.1 本体编辑工具概述	134
8.2 本体编辑工具的类型	135
8.2.1 JOE	135
8.2.2 KADS22	135
8.2.3 OILed	136
8.2.4 OntoEdit	136
8.2.5 Ontolingua	136
8.2.6 Protégé 2000	136
8.2.7 WebODE	137
8.2.8 WebOnto	138
8.3 本体编辑工具之间的比较	138
8.3.1 通用性.....	138
8.3.2 协作性.....	138
8.3.3 本体论工程.....	139
8.4 本章小结	140
思考题.....	140
附录 专业术语英汉对照表	141
参考文献	145

第 1 章

本体理念概述

1.1 本体的概念

本体(ontology)的概念最初起源于哲学领域,可以追溯到公元前古希腊哲学家亚里士多德(公元前 384—公元前 322)尝试对世界上的事物分类。在哲学中把 ontology 定义为“对世界上客观存在物的系统地描述,即存在论”(Gruber,1993);牛津(Oxford)英语词典把 ontology 解释成“存在的科学或研究”;依据韦伯斯特(Webster)词典,ontology 是“与存在的本质和联系相关的形而上学分支,或与存在的本质或存在物的种类有关的特殊理论”。当不同的理论学家提出本体的不同建议,或者不同的知识领域谈论各种本体建议时,应该使用本体的复数形式,以便表示总的本体集合,即本体论(ontologies)(Chandrasekaran,1999)。

信息系统和哲学之间的关系好像永远是两个不同的国度,原因是每个都有其自己的语言和文化。事实上,两者各自的研究方向是相互正交的。但今天,哲学的分支——本体论可以充当连接信息系统和哲学之间的桥梁纽带,尽管本体论在信息系统中的作用与哲学中的作用完全不同(Zuniga,2001)。信息系统需要推理世界模型,因此研究者在程序中利用术语“本体”来描述目标世界的信息。

信息系统本体论是表述特殊知识领域的形式化语言;而哲学本体论解释世界某些领域不依赖于任何特定语言的特殊分类系统,尽管运用语言的概念机制作为描述手段,但却既不可约也不等同于语言或形式体系。与信息系统本体论相似,哲学本体论确实解释研究领域的知识和概念框架,主要目的是预先忠实地描述,即寻求真理。无论存在着何种区别,哲学本体论仍能对概念化的框架以及信息系统本体论的开发做出一定的贡献,最大的贡献就是发现研究领域中的某些事实,即领域的本性、范围、边界和独特性等(Zuniga,2001)。

近十多年来,尽管本体论的相关研究日益成熟,也已经远远超过了哲学领域的范畴,并和信息技术、知识工程及人工智能(Artificial Intelligence, AI)都有着密切的关联,但在信息系统领域中,“本体究竟是什么”仍然是个有争议的话题之一。

1991年美国Stanford大学的Gruber和Neches等人最早把本体定义为“构成相关领域词汇的基本术语和关系,以及利用这些术语和关系构成的规定该词汇外延的规则”(Neches, 1991)。

1993年Gruber采用概念化的形式定义 $\langle D, R \rangle$ 结构(Genesereth, 1987),把本体解释成“共享概念化的明确的形式化规范”(Gruber, 1993),其中D是领域,R是D中相关的关系集合,因此该定义能够很好地表现出本体的本质特性。在此定义中,“共享(shared)”反映了本体捕获同感知识的理念,即不是限定到单个的某些人,而是一组人共同接受的知识;“概念化(conceptualization)”指的是世界中某些现象的抽象模型,辨识这些现象的相关概念;“形式(formal)”意思是机器应该可以理解本体,形式化具有不同的程度;“明确(explicit)”就是要清晰地定义所有概念的类型以及概念之间的关系约束。

为了澄清信息系统领域对本体概念的歧义,1995年意大利Padova大学的Guarino等人对不同概念解释进行深入分析,给出基本得到领域认同的本体概念,即“某些方面概念化的明确解释或表示”(Guarino, 1995)。此定义不是最终的标准定义,却符合大多数普通的标准用法,对信息系统具有理论指导意义。

1998年Guarino认为Gruber的本体定义仅指出了领域中普通的数学关系,即仅反映事物特殊状态的外延关系,却没有清楚地区分开本体和概念化(Guarino, 1998)。他认为如果要明确独立于事物状态的关系的意思,就需要引入统称为概念关系的内涵关系。基于此,把本体定义为“解释形式化词汇的指定意思的逻辑理论”,也就是世界的特殊概念化,这里的概念化指的是领域空间定义的一组概念关系,包含着领域空间中对象之间所有可能关系的意思解释。

但是这两种本体定义都没有涉及跨学科,况且Gruber的本体定义太含糊,而Guarino的本体定义对推理原理又模棱两可。这就需要在“形式化”和“逻辑理论”之间进行折中,因此信息系统本体论应该是“特定的形式化语言产生的清晰公理理论”(Zuniga, 2001),本体论的粒度越细则含有的公理也就越多。该本体论至少用于一个特殊且实际的应用,能描绘特定对象领域的结构,还能解释研究领域系统中使用的形式化词汇或协议的指定意思。

过去的十几年中,在信息系统中已经出现了本体的很多定义,术语“本体”大多分成两种意思(Chandrasekaran, 1999)。第一,本体是表示性词汇,经常指定到某些领域或主题。简单来讲,不是把词汇当成本体,而是获取词汇中术语的概念化。特别强调的

是,概念化是语言无关的,而本体是语言相关的,即应该符合特定的形式化语言。第二,本体有时指的是使用表示性词汇来描述某些领域的知识体,特别是用来描述领域的共识知识。换句话说,表示性词汇提供描述某些领域的事实的一套术语,而使用词汇的知识体是领域的事实集合。

本体的目标是捕获相关领域的知识,提供对该领域知识的共同理解,确定该领域内共同认可的词汇,并从不同层次的形式化模式上给出这些词汇(术语)和词汇之间相互关系的明确定义。

尽管信息系统中各领域之间对术语“本体”的理论解释还存在着很多矛盾和问题,但是本体论已成为信息系统中专业语言的必要组成部分,并在信息交换时起到至关重要的作用,因此信息系统的研究者在多数情况下已经基本认同这种歧义状况的存在,并用其表示系统中隐含或不明确的信息,以便使知识的共享和复用成为可能。

对本体概念的认识,可归纳为以下六点(王英林,2001):

- 本体是对某一领域概念化的表达;
- 概念是现实对象在某一或某些属性空间上的投影;
- 投影规则可能非常复杂,可能涉及多次投影或其他转换;
- 对同一领域的概念化有某些共同点,但概念化可能有所差异;
- 任意本体均不可能包括现实对象的全部属性,只能限定到所研究的领域范围内;
- 一个本体的声明转换到另一本体的声明不一定可逆。

1.2 本体论的分类

1.2.1 概念分类理论

本体的理论研究除了本体代数外,还包括有概念以及概念之间的分类。具有代表性的是 Guarino 等人对概念的分类所做的深入且细致的研究(Guarino,1994,1995,1998)。他们首先从一般的意义上分析概念、概念的特性和概念之间的关系,随后对概念进行分类,并提出一套具有指导作用的概念分类理论,其基础是概念的元特性。

Guarino 认为概念之间的差别不仅体现在概念的定义上,同时也体现在概念的某些特性上。从这些特性出发,归纳出概念最基本的特性——元特性,从而用公式给出元特性的严格的形式化定义。在此基础上,又讨论元特性之间的关系和约束,最终把研究成果作为概念分类的基本理论依据,并提出一套完整的概念分类体系结构。

概念的元特性主要包括以下七个最基本的特性(Guarino,1994):

- 持久性,对某个概念而言,存在某些实例会永远属于该概念;
- 非持久性,对某个概念而言,存在某些实例不会永远属于该概念;
- 反持久性,对概念的任意一个实例而言,这个实例不会永远属于该概念;
- 半持久性,就是非持久性和反持久性的差集;
- 载体标识性,各个实例之间的差别可以通过概念来区分;
- 支持标识性,各个实例之间可以相互区分;
- 外部依赖性,概念对其他概念的某种依赖关系。

概念分类理论是以概念的元特性为出发点,按照一定的分类规则,把具有相同元特性组合的概念进行归类,进而得出一般意义上的概念分类体系。基于该理论,Guarino等人又提出本体驱动的建模方法(Guarino,1998),以便在理论上为本体建模提供通用的模式。

1.2.2 本体论分类规则

依据本体论包含的内容可分成经典本体论和混合本体论。经典本体论只包括概念,例如概念分类,每个声明表示概念之间的关系;混合本体论还包括本体的关系和事件。

依赖于所采用的语言,按照表示和描述的形式化程度的不同,可以把本体论分为完全非形式化的本体论、半非形式化的本体论、半形式化的本体论和严格形式化的本体论(Guarino,1998)。形式化程度越高,越有利于计算机进行自动处理。尽管可以采用多种不同的表示形式,但一般都包含术语的词汇表和词汇意思的某些解释,即概念的定义和概念之间的关系,以及概念之间的关系所满足的公理,从而共同在领域中设定一个结构,限定对术语的可能解释。

按照应用领域的不同可将本体论大致分成三类(Uschold,1998a):人或组织之间达成概念共识的通讯;系统间使用本体论作为交换格式的互操作;系统工程领域,例如可复用性、知识获取、规范和可靠性等。有些领域将本体论看作是构造知识库的一种途径;另外一些将本体论视为知识库的一部分;此外还有一些领域将本体论当作与应用领域有关的交互工具和企业本体。

基于特定应用领域的规模或观点的抽象级别,可把本体论分成四种:元级本体、通用本体、领域本体和应用本体。元级本体(van Heijst,1997)是描述知识表示语言所用的基元分类的表示本体,例如开放式知识基互连(Open Knowledge Base Connectivity, OKBC)本体;通用本体,又称为核心本体,描述独立于特定问题或领域的非常通用的概念,例如空间、时间、对象、事件与行为等,几种通用本体(主要是自然语言本体)已被开发成机器可读字典(Machine-Readable Dictionaries, MRD),例如 CYC(Lenat,1995)和

WordNet(Miller,1990);领域本体通过特殊化高级本体中的术语,分别描述与通用领域或普通工作相关的词汇;应用本体描述依赖于特定领域和工作的概念,这些概念经常对应于领域实体执行某些活动时所扮演的角色,从方法的抽象模型中已开发出相应的应用本体,例如 CommonKADS(Schreiber,1994)、Generic Tasks(Bylander,1988)和 PROTÉGÉ- II (Munsen,1995)。

上面定义的本体论分类包含了与问题求解方法无关的静态知识,是构成领域层的一部分。为实现知识库系统各层次间的灵活配置,目前已提出了任务本体和方法本体的概念,它们分别描述特定任务与问题的求解方法。任务本体和方法本体本质上是从推理与问题求解角度描述领域知识的视图,它们有助于解决系统的互操作问题,即领域知识不能以与其使用方式无关的形式表示。任务本体和方法本体通过假设将领域知识与问题求解方法之间的交互明确地表达出来,充当了系统各层次之间的黏合剂,从而解决了知识库系统的复用与组件化开发中的很多关键问题。

依据目前所查阅的文献,可把本体论分成:基本的研究主题,例如哲学问题、知识表示、常识知识、通用本体库、领域本体库、工作和方法本体库等;本体的设计方法,例如 top-down、bottom-up、middle-out、分类和概念层次、内部结构与集成等;本体的应用,例如自然语言处理、知识管理、商业过程建模、智能信息检索、Internet 搜索、虚拟企业、企业供应链、仿真和建模、医学、教学、照片注释、电子商务和地理等;本体的开发,例如方法论、框架、工具、语言、对比、评估与标准化等;以及知识共享和复用,例如本体库的参与、多主体间通讯与知识库等。

按照对本体论的操作可分成以下七种类型:

- 本体编辑(Farquhar,1995;Uschold,1996b;FIPA,2000):浏览——提供浏览本体的可能性;生成——产生新本体;扩展——以不需修改现有定义的方式,基于已有词汇为特定使用来定义新术语的可能性;发布——产生本体使能访问和复用;保存——在开发过程中保存本体版本;更新——产生本体的更新拷贝。
- 本体代数(Wiederhold,1994a,1994b):交集——产生由共享实体组成的新本体子集;并集——集成本体中的所有唯一实体来产生新本体合集。
- 本体构造(Wiederhold,1994a;Chaudhri,2000;MacGregor,2003):抽取——列举和组织大型本体的领域概念以便产生接近特定领域的本体;合并——熔合两个独立开发的知识基(Knowledge Base,KB)或解决术语间名称和结构表示冲突的本体;修剪——删除给定领域不需要的概念或概念的子层次;切割——选择部分输入的本体用于新应用或新本体。
- 本体转换(Uschold,1996b;Chaudhri,2000;FIPA,2000;MacGregor,2003):术语转换——使得一种形式开发的本体可用于其他知识表示形式和不同的语言。