

迈向语义网

基于本体的 知识管理

张晓星 陈华明 译

TOWARDS THE SEMANTIC WEB
Ontology-driven Knowledge Management

(英) John Davies

(奥) Dieter Fensel

编著

(荷) Frank van Harmelen

 机械工业出版社
CHINA MACHINE PRESS



迈向语义网

基于本体的知识管理

(英)John Davies

(奥)Dieter Fensel 编著

(荷)Frank van Harmelen

张晓星 陈华明 译



机械工业出版社

本书主要探讨如何利用本体技术和知识管理方法,来促进语义网的发展。首先介绍语义网和知识管理的基本概念、相关工具及其架构,简要描述可扩展置标语言(XML)、资源描述框架(RDF)和资源描述框架模型(RDFS)的基本特征,以及本体语言OIL和DAML+OIL的相关知识,接着分析面向应用的本体开发和本体管理的方法论。本书还对语义网技术进行了探讨,说明基于本体的知识管理的应用。书中还以虚拟组织EnerSearch为例,证明通过知识管理实现语义网的有效性。最后对语义网的未来发展进行了展望。

John Davies, Dieter Fensel, Frank van Harmelen / Towards the Semantic Web: Ontology-driven Knowledge Management.

ISBN: 0-470-84867-7

Copyright © 2003 John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England

All rights reserved. Authorized translation from the English language edition published by John Wiley & Sons, Ltd.

本书中文简体字版由英国 John Wiley & Sons, Ltd. 授权机械工业出版社在全球独家出版发行。版权所有。

北京市版权局著作权合同登记号: 图字: 01-2004-3143

图书在版编目(CIP)数据

迈向语义网 基于本体的知识管理/(英)戴维斯(Davies, J.)等编著; 张晓星, 陈华明译. —北京: 机械工业出版社, 2006.4

书名原文: Towards the Semantic Web: Ontology-driven Knowledge Management
ISBN 7-111-18763-6

I. 迈... II. ①戴...②张...③陈... III. 语义网络—基本知识
IV. TP18

中国版本图书馆CIP数据核字(2006)第025322号

机械工业出版社(北京市百万庄大街22号 邮政编码 100037)

责任编辑: 李馨馨

责任印制: 李妍

北京中兴印刷有限公司印刷

2006年5月第1版·第1次印刷

1000mm×1400mm B5·7.375印张·300千字

0001—3000册

定价: 34.00元

凡购本图书,如有缺页、倒页、脱页,由本社发行部调换

本社购书热线电话:(010) 68326294

编辑热线 (010) 88379739

封面无防伪标均为盗版

译者说明

语义网、本体和知识管理是近两年来业界十分关注的话题。为了给该领域的研究人员或对该领域感兴趣的人士提供一本完整、权威的论著,经英国 John Wiley & Sons, Ltd. 出版公司的授权,我们对《Towards the Semantic Web: Ontology-driven Knowledge Management》一书进行了翻译,以飨读者。

本书主要探讨如何利用本体技术和知识管理方法,来促进语义网的发展。第 1 章简要介绍语义网和知识管理的基本概念、相关工具及其架构。第 2 章简要介绍了可扩展置标语言(XML)、资源描述框架(RDF)和资源描述框架模型(RDFS)的一些基本特征,并对两种用于语义网的本体语言进行了阐述,即 OIL 和 DAML + OIL。第 3 章介绍本体知识管理的有关方法论,提出了应用知识本体开发的 5 个阶段,包括可行性研究、准备阶段、细化阶段、评估阶段,以及维护和改进阶段。第 4 章讨论本体的管理问题,涉及储存、排列、维护等方面,并对本体的变化过程进行跟踪,随时更新本体,以保证本体概念的时效性和有效性。第 5 章至第 11 章讨论语义网技术,包括 Sesame、OntoBuilder、OntoEdit、QuizRDF、Apectacle、OntoShare 等。第 12 章介绍瑞士寿险集团的两个案例,用以说明基于本体的知识管理的实际应用。第 13 章以 EnerSearch 这一虚拟组织为例,证明通过知识管理实现语义网的有效性。第 14 章和第 15 章对语义网的未来发展进行了展望。

在翻译过程中,我们尽可能与原文的风格保持一致,但由于原书每章的作者都不尽相同,语言和用语也稍有差别,加之一些新概念,中文翻译还存在不尽人意之处。我们恳请广大读者在阅读本书的过程中,提出宝贵意见和建议。根据国内读者的阅读习惯,我们把原书的“索引”(Index)改成了“词汇表”,以方便读者的理解。

本书在翻译和出版过程中得到了蒋一民博士、施雪霖博士、邢春晓博士、张勇博士以及金富生、吕坤和彭勃老师的协助和支持,译者在此向他们表示衷心的感谢。此外,阎艳女士对本书的出版作了很多的联系和沟通工作,译者在此一并表示感谢。

译者

2006 年 4 月

前 言

还是这句话,知识就是力量!

J. Hendler 马里兰大学

30 多年前,ACM 图灵奖得主 Ed Feigenbaum 曾以“知识就是力量”为题,预言了商业计算革命。在这个口号的指导下,Feigenbaum 提出的特定域专家系统引起了计算界的关注。目前,由于该系统使用符合编码规则的特定域知识而产生了强大功能,所以被配置在各类报税软件中,并作为功能模块嵌入到某一全球畅销的软件产品中,估计有超过三分之二的“财富 500 强”公司在使用该系统。该专家系统包括各种税法规则、单词的拼写规则,以及支配市场运作的特定商业规则。在所有这些系统中,这种专门用途的知识就是力量之所在。

然而,在过去的 10 年里,一个新的课题逐渐演变成我们今天所知的语义网研究的一部分。我们也可以把这个方法叫做“知识就是力量”,但它的含义却完全不同。Feigenbaum 将力量想象成击打的力量,即把知识等同于电流的力量。就像流经电缆的电流一样,新力量流经互联网的路由器,而非来自为特定用途精心策划的汇集知识库的力量。

Tim Berners-Lee 提出了语义网的概念,他发明了网络并创造了“语义网”这个词。但他并不是第一位或唯一一个认识到语义网的力量的人。一些不受传统人工智能知识表现方法约束的研究人员,甚至在网络出现之前就谈到了“知识服务器”、“语义引擎”、“本体管理系统”,以及其他处理普遍知识的方法。但是,随着 Berners-Lee 的万维网的影响不断扩大,用于这些普遍知识的配置工具也愈加清晰。而且,用于网络的人工智能技术也在为语义网功能的增强提供知识管理技术。

因此,语义网的力量不仅来自于人工智能领域研发的知识技术,而且也来自于网络开发者研发的电力网格技术。在资源描述框架(RDF)和可扩展置标语言(XML)的基础上,新的语义网语言把功能强大的人工智能概念与已经改变了世界的网络基础设施很好地联系起来。现在,网络已经接入到全世界的几乎每台电脑,并能承载人工智能领域的知识。

要使人造智能(AI)实验室中的研究成果发展成为互联网标准,最重要的工作是要开展网络本体领域的研究,这一点已经变得越来越清晰了。从 20 世纪 90 年代中期至末期,数项重要项目的研究成果已经证明,可以将机读本体与网络资源连接起来。这些课题引起了政府对该领域的高度重视。在美国国防部高级研究计划

局(DARPA)和 IST 项目的资金支持下,语义网在规模、容量和影响方面呈跳跃式发展。目前,将知识嵌入到网络的方法已经实现了标准化,产业界也开始高度关注这一新的发展趋势。负责惠普公司软件开发的技术总监 Richard Hayes-Roth 说:“我们希望语义网的发展成为一场如同当初网络本身一样的大革命”(《商业周刊》,2002 年 2 月)。

On-To-Knowledge 项目对这场即将到来的革命做出了重要贡献。该项目由许多在本体领域出类拔萃的欧洲研究人员参与,本书许多论述也来自该项目的研究成果。本书探讨了在网络上进行知识开发和知识利用的新途径,并为在这个令人振奋的新兴领域中开展高水平研究开创了一个先例。因此,本书描述了最新的工作进展,展示了最新研究方法对网络知识管理所产生的巨大力量。

简而言之,我们已经看到这一天的到来,即对特定领域应用进行的精心知识封装,将会被无所不在的、由知识源链接而成的大型、分布式知识网络所取代。在机读本体的帮助下,网络里的所有数据库、网络服务和文档均可以产生这种力量,同时激发新一波各种应用的涌现。本书中所描述的项目预示了这场即将到来的革命,它们将成为这场“知识就是力量”革命的先导。

James Hendler

马里兰大学

致 谢

感谢 Cath McCarney 为本书排版所做的巨大贡献。

第 3 章:感谢 Hans-Peter Schnurr、Hans Akkermans 和 Karlsruhe 大学 AIFB 研究所的同事们。

第 8 章:本书作者感谢 Dirk Wenke、Siggi Handschuh 和 Alexander Mädche 为 OntoEdit 本体工程环境的实施所做的大部分工作;感谢 Jürgen Angele 和 Steffen Staab 为本书所做的录入工作。

第 10 章:感谢 Nick Kings 为 OntoShare 系统的设计和开发所做的贡献。

第 13 章:本书作者感谢以前的同事 Bernd Novotny 和 Martin Staudt,他们为本书两个案例的早期阶段作了大量工作。

本书部分得到了欧盟研究项目 OnToKnowledge(IST-1999-10132)和瑞士教育与科学联邦办公室(项目号:BBW 99.0174)的支持。感谢 Vincent Obozinski、Wolfram Brandes、Rober Meersman 和 Nicola Guarino 对 On-To-Knowledge 提出的建设性反馈意见。

目 录

译者说明

前言

致谢

第 1 章 概述	1
1.1 语义网与知识管理	2
1.2 本体的作用	3
1.3 基于语义网的知识管理架构	4
1.3.1 知识获取	4
1.3.2 知识表达	4
1.3.3 知识维护	4
1.3.4 知识使用	4
1.4 基于语义网的知识管理工具	5
1.4.1 知识获取	5
1.4.2 知识表达	6
1.4.3 知识维护	6
1.4.4 知识使用	7
第 2 章 OIL 和 DAML+OIL:用于语义网的本体语言	8
2.1 概述	8
2.2 语义网的语言金字塔	8
2.2.1 用于数据交换的 XML	8
2.2.2 用于声明的 RDF	9
2.2.3 用于简单本体的 RDF Schema	11
2.3 OIL 的设计原理	11
2.3.1 基于框架的系统	12
2.3.2 描述逻辑	13
2.3.3 Web 标准:XML 和 RDF	13
2.4 OIL 语言的构建	13
2.5 不同的句法形式	16
2.6 语言分层	18
2.7 语义	20
2.8 从 OIL 到 DAML+OIL	21

2.8.1 与 RDFS 集成	21
2.8.2 对个体的处理	23
2.8.3 DAML + OIL 数据类型	23
2.9 经验和未来的发展	24
第 3 章 基于本体的知识管理的方法论	26
3.1 概述	26
3.2 可行性研究	27
3.3 准备阶段	30
3.4 细化阶段	32
3.5 评估阶段	33
3.6 维护和改进阶段	33
3.7 相关工作	33
3.7.1 框架方法论	34
3.7.2 KACTUS	34
3.7.3 方法论	35
3.7.4 本体分析的规范工具	35
3.8 结论	36
第 4 章 本体管理:本体的存储、调整和维护	37
4.1 对本体管理的要求	37
4.2 本体的调整	38
4.2.1 本体调整的必要性	38
4.2.2 注释性 XML 文档的调整	38
4.2.3 元本体的映射	39
4.2.4 OIL 中的映射	41
4.3 对本体变更的支持	42
4.3.1 本体不断变更	42
4.3.2 本体变更所涉及的问题	43
4.3.3 对本体变更的管理	45
4.4 本体的组织	48
4.4.1 Sesame 要求	48
4.4.2 本体存储系统的功能	49
4.4.3 当前的存储系统	50
4.4.4 对存储系统的要求	51
4.5 小结	53
第 5 章 Sesame:存储与查询 RDF 以及 RDF Schema 的通用结构	55
5.1 需要 RDFS 查询语言	55
5.1.1 句法层次的查询	56

5.1.2 结构层次的查询	57
5.1.3 语义层次的查询	58
5.2 Sesame 架构	59
5.2.1 RQL 查询模块	60
5.2.2 Admin 模块	61
5.2.3 RDF 输出模块	62
5.3 SAIL API	62
5.4 各种经验	64
5.4.1 应用:上升到知识	64
5.4.2 实践中的 RDFS	65
5.4.3 PostgreSQL 和 SAIL	65
5.4.4 MySQL	67
5.5 未来工作	68
5.5.1 对交易撤销的支持	68
5.5.2 对版本变化的支持	68
5.5.3 功能模块的添加和扩展	68
5.5.4 DAML + OIL 支持	69
5.6 结论	69
第 6 章 OntoBuilder:语义网本体的生成	70
6.1 概述	70
6.1.1 OntoBuilder 及其与 CORPORA 系统的关系	71
6.1.2 OntoExtract	71
6.1.3 OntoWrapper 和 TableAnalyser	73
6.2 网络读取	75
6.2.1 互联网上的语义	75
6.2.2 文档自然语言文本检索的问题	76
6.2.3 文档处理	76
6.2.4 标准化	77
6.2.5 语句多样化	77
6.2.6 文档类别归类	78
6.2.7 写作风格	78
6.2.8 布局问题	78
6.3 信息抽取	79
6.3.1 内容驱动与目标驱动	79
6.3.2 语言分析的水平	80
6.3.3 CognIT 的远景	81
6.4 自然语言文档的知识生成	82

6.4.1	句法与语义	82
6.4.2	语义结构的生成	83
6.4.3	文本资源的本体生成	83
6.4.4	可视化与导航	84
6.5	利用自动文本抽取创建网络资源本体的有关问题	85
第7章	OntoBuild: 本体合作工程	88
7.1	概述	88
7.2	准备阶段	89
7.3	改进阶段	92
7.3.1	交易管理	94
7.3.2	概念层级子树的锁定	95
7.3.3	概念锁定	96
7.4	评估阶段	97
7.4.1	典型查询的分析	97
7.4.2	错误的避免与定位	98
7.4.3	能力问题的利用	98
7.4.4	协作评估	98
7.5	相关工作	98
7.6	小结	99
第8章	QuizRDF: 语义网的搜索工具	101
8.1	概述	101
8.2	本体标引	103
8.3	本体搜索	105
8.4	备选数据模型	107
8.4.1	新模型中的标引	108
8.4.2	新模型中的搜索	108
8.5	未来工作	108
8.5.1	技术的改进	108
8.5.2	评估	109
8.6	小结	109
第9章	Spectacle	111
9.1	概述	111
9.2	Spectacle 内容表现平台	111
9.3	Spectacle 的架构	112
9.4	基于本体的映射方法论	113
9.4.1	信息实体	114

9.4.2 本体映射	114
9.4.3 实体描述	114
9.4.4 导航说明	115
9.4.5 导航描述	115
9.4.6 浏览	116
9.4.7 用户说明	116
9.5 基于本体的信息可视化	117
9.5.1 分析	117
9.5.2 查询	120
9.5.3 导航	121
9.6 小结	122
第 10 章 OntoShare:知识共享系统中的本体演变	123
10.1 概述	123
10.2 OntoShare 中的知识共享与检索	124
10.2.1 OntoShare 中的知识共享	124
10.2.2 本体表示	125
10.2.3 OntoShare 中的显性知识检索	127
10.3 变化本体的创建	129
10.4 专家定位与隐性知识	130
10.5 社会技术问题	132
10.5.1 隐性知识与显性知识流	132
10.5.2 虚拟社区	132
10.6 评估和未来工作	134
10.7 小结	135
第 11 章 本体中间件与推理	136
11.1 本体中间件:特征与架构	136
11.1.1 On-to-Knowledge 架构中的位置关系	137
11.1.2 术语	138
11.2 变化的跟踪、版本变化与元信息	139
11.2.1 相关工作	139
11.2.2 有关要求	140
11.3 RDF(S)知识仓库的版本变化模型	141
11.3.1 穿越对等状态的历史	142
11.3.2 作为知识仓库标识状态的版本	142
11.3.3 实施方法	143
11.3.4 元信息	144
11.4 DAML + OIL 的示例推理	145

11.4.1 推理服务	147
11.4.2 DAML+OIL 推理器的功能接口	148
第 12 章 基于本体的知识管理:瑞士寿险公司案例研究	149
12.1 概述	149
12.2 技能管理	149
12.2.1 什么是技能管理	149
12.2.2 SkiM:瑞士寿险公司的技能管理	151
12.2.3 SkiM 的架构	152
12.2.4 基于本体方法的 SkiM	153
12.2.5 查询工具	156
12.2.6 评估与展望	157
12.3 从文本中自动抽取“轻量级本体”	158
12.3.1 动机	158
12.3.2 本体的自动抽取	158
12.3.3 查询本体的部署	160
12.3.4 评估与展望	162
12.4 结论	164
第 13 章 语义网工具在虚拟组织中的试验	165
13.1 概述	165
13.2 作为虚拟组织的 EnerSearch 工业研究协会	165
13.3 使用语义网方法的有效性	167
13.4 语义网现场实验的设计考虑	168
13.4.1 不同的信息模式	168
13.4.2 不同的目标用户群	169
13.4.3 不同的个体认知风格	169
13.4.4 要验证的假设	172
13.5 在虚拟组织中搭建实验系统	172
13.5.1 选择目标测试用户	173
13.5.2 测试工具	173
13.5.3 测试的任务及其组织	173
13.5.4 实验程序	174
13.5.5 确定搜集哪些数据	175
13.5.6 评估矩阵和尺度	175
13.6 语义网实验的技术和系统	176
13.6.1 系统设计	176
13.6.2 本体工程,总体,注释	177
13.7 基于本体的信息检索	178

13.7.1 本体和语义分布图	178
13.7.2 基于语义的信息检索	180
13.8 经验教训	182
第 14 章 展望未来:开发用于知识管理的 P2P 技术和语义网技术	184
14.1 概述	184
14.2 现代知识管理的展望	185
14.2.1 知识整合	185
14.2.2 知识分类	185
14.2.3 内容认知	186
14.2.4 个性化	186
14.2.5 知识门户建设	186
14.2.6 社区实践	187
14.2.7 P2P 计算及其在知识管理中的应用	187
14.2.8 虚拟组织及其影响	188
14.2.9 eLearning 系统	188
14.2.10 知识网格	188
14.2.11 智力资本评估	188
14.3 本体一览:动态的含义网络	189
14.3.1 本体或如何避免悖论	189
14.3.2 空间异构:作为含义网络的本体	190
14.3.3 适时开发:应用中的本体	191
14.4 P2P、本体和知识	191
14.4.1 单独使用 P2P 和本体的缺陷	192
14.4.2 P2P 与本体整合所面临的挑战	193
14.5 结论	196
14.5.1 用于知识管理的 P2P	196
14.5.2 用于本体的 P2P	196
14.5.3 用于 P2P 和知识管理的本体	197
14.5.4 社区构建	197
第 15 章 结论:本体驱动的知识管理——走向语义网	198
词汇表	199
参考文献	210

第 1 章 概 述

John Davies, Dieter Fensel 和 Frank van Harmelen

当今的万维网(World Wide Web)上有着数十亿篇文献,全球约有 3 亿多人使用着这些资料。另外数百万的网页资料在企业内部网中被检索使用。信息的数量在持续地飞速增长,这使得查找、组织、访问和维护用户所需的信息越来越困难。Berners Lee 等人在 2001 年提出了语义网(Semantic Web)的概念,它以机器可处理的元数据开发为基础,提供了强大的信息访问功能。本书,我们将特别关注语义网技术在知识管理(knowledge management)领域的新的应用。

长期以来,一个公司的价值是由它拥有的有形资产来决定的。但近年来,人们越来越清楚地认识到:在后工业时代,一个企业的成功更依赖于它的智力资产,而不是其物理资源。

在此趋势之下有着一系列的因素,比如说许多工业领域中对高技能工人的需求,对新的计算技术和电信技术的需求,对更快的革新速度以及比以往更短的产品生产周期的需求,凡此种种都引发了企业竞争方式的巨大变革:知识变成了竞争中的关键。

还有一些因素促使公司尝试更有效地管理并开发它们的智力资产。其中一个因素是日益增长的员工流动率以及更为灵活的劳动力市场,它们会导致知识遗失;另一个因素是全球化,它要求人们相互协作,跨地区、跨时区地交流知识经验。

知识管理学科就是专门应对这种挑战的。它可以被广义地定义为一个机构管理智力资产最具实效的工具、技术或过程(Davis, 2000 年 a)。这些智力资产有着多种开发途径。比如说,通过共享和复用当前最成功的实践经验,可以改进目前的业务流程,并消除重复劳动。通过搜集市场信息和产品销售情况来寻求新的商机,从而抢在商业对手之前创造和开发出新的产品和服务,并将其引入市场。

知识管理派系中常常有如下的辩论:技术只是知识管理中一个相对边缘的部分,企业文化更为重要。我们需要一个比纯技术更为广阔的视野,这种观念虽然是正确的,但它反映了一种假想,即技术和无形的企业文化是分离的。由此可见,我们要有一个全面的认识,即技术常常在文化因素中占有较大比重,而并非人们有时所想象的那样。虽然本书的重点在于为知识管理创建基于语义网的工具,但是理解这些工具能被最有效使用的文化和企业环境,同样重要。该领域内的相关工作可参见 Maxwell(2000 年)等的论述。

1.1 语义网与知识管理

企业内部网在更有效地开发显性(编码化的)和隐性(未表达出来的)知识中起着很重要的作用。对于显性知识来说,企业内部网技术使用公开的标准,为企业知识提供了一个普遍适用的接口,且费用相对较低。将纸介质的信息内容迁移到企业内部网上,对于信息更新的速度和准确性都大有好处。于是问题变成了如何在适当的时间将适当的信息传递给适当的人群。事实上,考虑显性知识的一种方式就是:它是适当的关联环境中的信息。换句话说,它是可以产生有效行为的信息。而对于隐性知识,我们可以使用基于企业内部网的工具把具有相同兴趣爱好的人联系在一起,鼓励对话,创造交流隐性知识的机会。

重要的知识常常分散在互联网及企业内部网资源的各个角落。传统搜索引擎仅能返回按重要性排序的结果列表,这些结果只为文档间的语义关系提供了极少的信息。知识工人往往要花费大量的时间用于浏览和阅读,以便找出这些文档间的关联,以及各个文档在问题域的整个结构中所处的位置。只有当知识工人开始明白种种信息之间的异同时,他们才进入了工作的精髓部分:建立关系,从而创造新的知识。

当前的知识管理系统有几个重大缺陷:

- 搜索信息:现有的基于关键字的搜索方式会检索到一些无关信息,这些信息中含有某些关键字,但其实际含义不同(一词多义现象)。如果同一个含义在某些信息中以不同于关键字的字词来表达,这些信息也会被遗漏。传统的信息检索机制把重点放在给定的查询条件(或是用户资料)和信息存储之间的关系上。另一方面,对于选中的信息间的相互关系的开发(知识本体的使用可使之更加容易),可以把原本不相关的信息集成为一个有意义的整体。这样建立起来的隐式结构可以帮助用户更加有效地利用和管理信息。
- 提取信息:目前,人工浏览和阅读对于从信息源中提取相关信息来说是必须的。因为自动智能代理不具有从文本表达中提取相应信息的常识,并且它们也不能把分散于各处的信息集中起来。
- 维护结构松散的文本资源:这是一个困难且费时的活动,当这些资源变得庞大时尤为明显。保持这些资源一致、正确且最新,需要语义的机械化表达,语义可以检测出异常情况。
- 自动文档生成:这将有助于自适应网站的形成,它们可以根据用户资料或其他适当的方面被动态修改。从半结构化的数据产生半结构化的信息,这表示需要这些信息源的语义表达,并且这样的表达要能被机器识别。

许多公司间的竞争高度依赖于他们如何开发全体的智慧和经验。目前大多数

的网络化信息都是多媒体数据,并且结构相当松散。这些网络包括互联网以及大型的公司内部网。在结构松散的表达媒介中查找和维护信息是一项具有挑战性的工作。许多公司逐渐地意识到它们的内部网是很有价值的存储全体智慧的知识库。但是随着信息容量的迅速增加,如何把这些资源组织成有用的知识就成了一个大问题。

要想把分散于 Web 中的资源集成为相互关联的信息整体,就需要用到知识管理工具。以前在信息集成方面的研究(Hearst, 1998 年)主要强调异构数据库和知识库的集成,这些库可以通过形式化语言以高度结构化的方式来表达信息。与此相对应的是,Web 中包含着大量的非结构化或半结构化的自然语言文本。

语义网可以被想象成对目前 Web 的扩展,存储于其中的文档不但能够通过 WWW 浏览器被人们阅读,而且还可以通过元信息进行注释。元信息以机器可处理的方式来定义信息(文档)的摘要。元信息的显式表达在域理论(本体)的协助下,将使网络能够提供高质量的新层次服务。它会将人类的知识编织成一个巨大的网络,并使其可被机器处理。通过机器可理解的形式访问和提供信息,多种自动服务将会帮助用户达到他们的目的。这个过程最终将创建具有丰富知识的系统以及多种专业的推理服务系统,它们几乎可以在生活的各个方面为我们提供支持,并且会像电力一样成为我们生活中的必需品。

知识本体提供了一个处理 Web 资源的异构表达的方法。一个知识本体中的域模型可以被当作一个统一的结构,利用这个结构可以给予信息一种通用的表达方式和语义。

1.2 本体的作用

本体是建立语义网技术的关键。它把人类对符号的理解同机器可处理的符号的特性交织在一起。本体论是在人工智能的研究中发展起来的,它被用于促进知识的共享和复用。从 20 世纪 90 年代初开始,本体逐渐成为了一个热门的研究课题。它被数个人工智能研究团体学习研究,这些团体的研究方向包括知识工程、自然语言处理以及知识表达。近年来,智能信息集成、合作信息系统、信息检索、电子商务以及知识管理等领域也广泛使用了本体。本体之所以受欢迎,和它所作的承诺是分不开的:提供对一个领域共享且通用的理解,并且人类和应用系统可以在该领域内实现自由沟通。同样,本体及其支持工具的使用提供了一个在大企业内大幅度改进知识管理能力的机会,而本体在该领域内的应用正是本书的主题所在。

本书描述了一个基于语义网的知识管理体系结构以及一整套用于语义信息处理的创新工具。书中也阐述了这种方法的理论基础。工具环境包含以下 3 个关键方面: