

Theory and Engineering Application of Ontology

本体理论与工程应用

徐享忠 汤再江 杨建东 孙俊峰 编著



国防工业出版社
National Defense Industry Press

本体理论与工程应用

徐享忠 汤再江 杨建东 孙俊峰 编著

国防工业出版社

·北京·

内 容 简 介

本书重点阐述本体的工程应用,以确定性本体为主、概率本体为辅,并根据需要引入相应的本体理论。全书分为本体理论、确定性本体及工程应用、概率本体及工程应用和附录等四部分,贯穿 Web 本体语言(OWL 2)、语义网规则语言(SWRL)、确定性本体建模工具 protégé 5.2.0、概率本体建模工具 UnBBayes - MEBN 4.22.18 的使用。

本书可作为信息系统互操作、系统仿真、知识工程等相关专业本科生和硕士研究生的教材或教学参考书,也可供有关工程技术人员自学和参考。

图书在版编目(CIP)数据

本体理论与工程应用 / 徐享忠等编著. —北京:
国防工业出版社, 2019. 10
ISBN 978 - 7 - 118 - 11960 - 2

I. ①本… II. ①徐… III. ①指挥信息系统 - 研究
IV. ①E94

中国版本图书馆 CIP 数据核字(2019)第 221183 号

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

三河市德鑫印刷有限公司印刷

新华书店经售

*

开本 710×1000 1/16 印张 17 字数 314 千字

2019 年 10 月第 1 版第 1 次印刷 印数 1—2000 册 定价 78.00 元

(本书如有印装错误,我社负责调换)

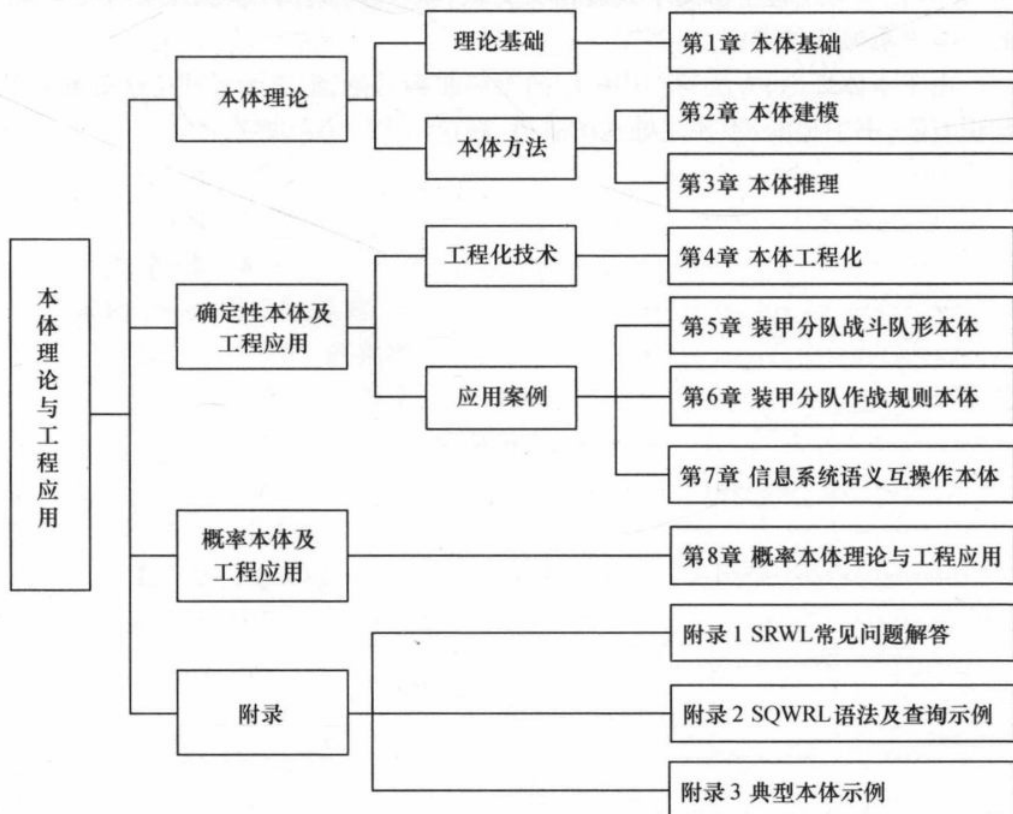
国防书店:(010)88540777
发行传真:(010)88540755

发行邮购:(010)88540776
发行业务:(010)88540717

前 言

本体能够为领域术语显式地赋予语义,为信息智能处理奠定了良好基础,从而在语义网、知识库、信息系统互操作、数字图书馆情报检索等领域得到了非常广泛的应用。为便于理解,本书重点阐述本体的工程应用技术及相关应用案例,并根据需要引入相应的理论与工具,而不过多阐述形式化语言、描述逻辑、非单调推理等抽象内容;同时,不同于已有文献大多关注本体建模的现状,本书在应用案例中着重关注本体的推理,以充分发挥本体技术的优势。

全书总体上按照“本体理论—本体工程应用”的体系编写,分为本体理论、确定性本体及工程应用、概率本体及工程应用和附录等四部分:



全书共8章,第1章以语义网整体框架为统揽,辨析了本体的基本概念,阐

述了本体描述语言、语义网规则语言、本体查询语言等基础内容,重点阐述本体描述语言 OWL 2 的新特性;第 2 章阐述本体建模的一般原则、一般过程、主要方法及主流工具,重点阐述改进的“七步建模法”及确定性本体建模工具 protégé 5.2.0 的使用;第 3 章阐述描述逻辑、开放世界假设、本体推理的特点、常用推理机及典型的推理案例;第 4 章重点从改进本体模型的质量和提本体推理的效率两个方面,阐述了 OWL 本体中完整性约束、实体批量录入、本体重构、本体导入、加速本体推理等相关工程化技术,并介绍了 OWLAPI、SWRLRuleEngineAPI、SQWRLQueryAPI 的二次开发;第 5~7 章为确定性本体建模的 3 个工程应用案例,均涵盖本体建模及推理两大方面;第 8 章阐述概率本体及概率本体建模周期理论,并运用概率本体建模平台 UnBBayes - MEBN 4.22.18,给出了军用车辆识别多实体贝叶斯网络(MEBN)建模实例;3 个附录则分别给出了语义网规则语言(SWRL)的常见问题解答、本体查询语言(SQWRL)语法及查询示例、典型本体示例(Manchester OWL 语法格式),是资料性附录。

本书是军队科研项目(项目编号:9140A04030214JB35059)研究成果的总结,作者在编写过程中参阅了大量相关文献,吸收了同行们辛勤劳动的成果,在此一并表示诚挚的感谢。

由于本体理论、方法与工程应用的发展非常迅速,加之作者理论基础和专业知识有限,书中疏漏、不妥之处在所难免,敬请广大读者批评指正。

徐享忠

2018 年 10 月

目 录

第 1 章 本体基础	1
1.1 语义网概述	1
1.1.1 语义网的概念	1
1.1.2 语义网整体框架	3
1.2 本体的概念及类型	7
1.2.1 本体的概念	7
1.2.2 本体的类型	9
1.3 本体描述语言	13
1.3.1 RDF/RDFS	14
1.3.2 DAML + OIL	16
1.3.3 OWL	16
1.3.4 OWL 2	18
1.3.5 KIF	28
1.4 语义网规则语言	28
1.4.1 SWRL 的架构	29
1.4.2 SWRL 的原子	30
1.4.3 SWRL 与 OWL 的关系	31
1.5 本体查询语言	32
1.5.1 SPARQL	32
1.5.2 SQWRL	35
1.5.3 DL 查询	38
1.6 本章小结	38
第 2 章 本体建模	39
2.1 本体建模的一般原则	39

2.2	本体建模的一般过程	40
2.3	本体建模的主要方法	41
2.3.1	七步法	41
2.3.2	MCSC2O 法	44
2.3.3	骨架法	45
2.3.4	九步法	45
2.3.5	企业建模法	46
2.3.6	METHONTOLOGOY 法	47
2.3.7	KACTUS 法	47
2.4	本体模型构建方法的改进	48
2.4.1	已有本体模型构建方法的分析	48
2.4.2	改进的“七步法”	48
2.5	本体建模主流工具	51
2.5.1	Protégé 概述	52
2.5.2	Protégé 的启动配置	53
2.5.3	Protégé 的界面配置	56
2.5.4	Protégé 本体的可视化插件	59
2.5.5	Protégé 本体的 SWRL 规则插件	61
2.5.6	Protégé 本体的查询器插件	64
2.6	本章小结	72
第 3 章 本体推理		73
3.1	本体推理应用背景	73
3.2	描述逻辑	75
3.2.1	描述逻辑的理论发展	76
3.2.2	描述逻辑的基本体系	76
3.2.3	TBox	77
3.2.4	ABox	79
3.3	语义假设	79
3.3.1	开放世界假设	80
3.3.2	封闭世界假设	81
3.4	本体推理的基本原理	81
3.5	本体推理典型任务	83

3.5.1	本体推理的背景	83
3.5.2	完全信息的推理	85
3.5.3	不完全信息的推理	87
3.6	本章小结	90
第4章	本体工程化	91
4.1	OWL 本体中完整性约束	91
4.2	实体批量录入	93
4.2.1	本体维护方式的分析	93
4.2.2	批量创建类层次结构	94
4.2.3	批量创建对象属性层次结构	95
4.2.4	批量创建数据属性层次结构	95
4.2.5	电子表格到本体的映射	96
4.3	本体重构	101
4.3.1	实体重命名	102
4.3.2	公理的拆分与合并	105
4.3.3	公理的复制、移动与删除	105
4.3.4	本体的合并	106
4.4	本体导入	106
4.4.1	任务分解与团队协作	107
4.4.2	OWL 2 提供的本体导入机制	108
4.4.3	采用 Protégé Desktop 导入本体	108
4.5	加速本体推理	109
4.5.1	影响本体推理速度的主要因素分析	109
4.5.2	Protégé 加速本体推理的若干措施	110
4.6	本体二次开发	112
4.6.1	OWLAPI	113
4.6.2	SWRLRuleEngineAPI	117
4.6.3	SQWRLQueryAPI	118
4.7	本章小结	121

第 5 章 装甲分队战斗队形本体	122
5.1 装甲分队战斗队形概述	122
5.1.1 装甲分队战斗队形	122
5.1.2 装甲分队战斗队形变换实施	123
5.2 基于 OWL 的装甲分队战斗队形本体模型	124
5.2.1 创建 OWL 类层次结构	124
5.2.2 创建属性	125
5.2.3 定义战斗队形	127
5.2.4 创建个体	128
5.2.5 描述个体的约束	129
5.3 基于 SWRL 的装甲分队战斗队形变换实施	131
5.3.1 一路战斗队形到一字战斗队形变换规则	132
5.3.2 一字战斗队形到后三角队形变换规则	133
5.4 装甲分队战斗队形本体的推理	133
5.4.1 包含性检验	133
5.4.2 一致性检验	137
5.5 本章小结	141
第 6 章 装甲分队作战规则本体	142
6.1 装甲分队作战规则概述	142
6.2 装甲分队作战规则的描述	143
6.2.1 作战规则的基本结构	143
6.2.2 作战规则的主要要素	144
6.2.3 作战规则主要要素的分解	147
6.3 装甲分队作战规则的获取及典型实例	149
6.3.1 作战规则获取的一般原则	149
6.3.2 作战规则获取的主要方法	150
6.3.3 装甲分队作战规则典型实例	151
6.4 装甲分队作战规则本体模型的建立	153
6.4.1 创建 OWL 类层次结构	153
6.4.2 创建属性	154

6.4.3	创建个体	156
6.4.4	明确个体与属性的关系	157
6.4.5	基于 SWRL 的装甲分队作战规则	158
6.5	装甲分队作战规则本体的推理	162
6.6	本章小结	164
第 7 章	信息系统语义互操作本体	165
7.1	互操作的涵义	165
7.1.1	基本涵义	165
7.1.2	信息系统互操作的层次模型	166
7.1.3	仿真系统互操作	170
7.1.4	指挥控制系统互操作	171
7.1.5	指挥控制系统与作战仿真系统的互操作	172
7.2	指挥控制系统与作战仿真系统的异构性	175
7.2.1	术语体系	176
7.2.2	系统架构	177
7.2.3	信息交互模型	179
7.2.4	通信协议	181
7.2.5	数据库	182
7.2.6	小结	182
7.3	指挥控制系统与作战仿真系统语义互操作通用 技术框架设计	183
7.3.1	语义互操作的要求	183
7.3.2	面向服务的架构(SOA)的语义互操作	185
7.3.3	指挥控制系统与作战仿真系统语义互操作的 通用技术框架	187
7.4	作战任务领域本体模型	189
7.4.1	作战命令概述	189
7.4.2	采用映射方法创建作战任务领域本体模型	191
7.4.3	基于 OWL 的作战任务领域本体模型	192
7.5	装甲分队机动领域本体模型的建立	197
7.5.1	类层次结构	197
7.5.2	战斗队形本体	197

7.6	装甲分队机动领域本体的推理	199
7.6.1	任务次序的推理	200
7.6.2	编配装备的推理	203
7.6.3	战斗队形的推理	206
7.6.4	执行单位的推理	207
7.6.5	友邻关系的推理	209
7.7	本章小结	210
第8章 概率本体理论与工程应用		211
8.1	贝叶斯网络	211
8.2	多实体贝叶斯网络	212
8.3	概率本体	214
8.4	概率本体建模周期	216
8.4.1	模型需求的描述	216
8.4.2	模型的分析与设计	217
8.4.3	模型的实现	217
8.4.4	模型的测试	218
8.5	车辆识别多实体贝叶斯网络模型	219
8.5.1	建立模型结构	220
8.5.2	建立证据列表	221
8.5.3	建立车辆识别模型	223
8.5.4	进行模型测试	226
8.6	本章小结	227
附录1	SWRL 常见问题解答	228
附录2	SQWRL 语法及查询示例	240
附录3	典型本体示例(Manchester OWL 语法格式)	249
参考文献	257

第1章 本体基础

本体(Ontology)源于哲学问题,用来表示世界本源的存在。本体被较早引入计算机领域,用来表达人们对领域知识的共同理解,并以显式的方式描述研究领域概念及概念之间的关系。后来,本体又被引入到人工智能、知识工程和图书情报等领域,用来解决知识表示和共享方面的问题。现今,本体已成为新一代互联网——语义网的关键技术。本体有自己的描述语言,本体语言需要为用户提供信息描述、推理和查询等基本能力,并在表达能力和推理性能之间取得良好平衡。

1.1 语义网概述

语义网(Semantic Web)被称为第三代互联网,它以实现万维网中的信息成为机器可处理和可理解的信息为目标,并实现不同机器之间数据语义互操作。语义网不仅可以生成和发布信息与知识,而且可以进行语义校验、机器推理、形式证明,从而真正让新一代互联网形式化和语义化。

以语义网技术为代表的一系列标准规范近年来都得以确立,一系列元数据标准、本体语言和工具陆续开发了出来,一些示范性应用的成功也使人们看到了语义技术的巨大潜力。语义网成为目前计算机界最活跃的研究领域,为不同领域及领域之间的语义互操作提供了坚实的技术基础。

语义网的思想在军事领域同样产生了很大影响。早在2003年5月美国国防部签署的《国防部网络中心数据战略》中,数据的可理解性就被作为一个重要目标提出来:“用户和应用能同时在结构和语义上理解数据,并很容易确定该数据怎样用于特定需求。”

1.1.1 语义网的概念

万维网的发明者 Tim Berners - Lee、万维网联盟(World Wide Web Consortium, W3C)以及网络科学(Web Science)专业的一些语义网基础教材等从不同角度对语义网进行了定义。

定义 1.1:语义网是一个具有一致性和逻辑性数据的万维网(The Semantic Web is a consistent logical web of data.)。语义网要开发一种语言,达到以机器可处理的形式表达信息的目的。

该定义在语义网整体框架提出之前给出,其认为语义网在某种形式上类似于一个全球性的数据库,它主要关注信息背后的网络空间关联结构以及如何开发语言来表达万维网信息等问题。

Web 中充满了不完整的信息,而 Web 上信息的缺失只是意味着这些信息尚未得到显式描述。语义网的核心在于推理得出新信息的可能性,但它不支持非单调推理(Nonmonotonic Inference)。为此,语义网采用了开放世界假设(Open World Assumption,OWA)而非封闭世界假设(Close World Assumption,CWA),不遵从唯一命名假设(Unique Name Assumption,UNA)。因此,语义网与传统数据库具有本质的不同。

定义 1.2:语义网是对当前万维网的一种扩展,其中的信息被赋予明确定义的含义,使机器与人能更好地协同工作。

该定义在较高的层次上对语义网进行了抽象概括。首先,指出了“信息被赋予明确定义的含义”是语义网的基础;其次,“协同工作”体现了语义网的最终目标就是要成为协同工作的媒介。

定义 1.3:语义网应该将当下可用的结构化和半结构化的数据转换为标准的格式并发布;语义网不仅发布类层面的数据集,同时也可以发布个体层面的数据集,以及这些个体之间的关系;语义网将这些数据想要表达的语义以一种形式化的方法表示出来,进而实现机器对这些语义的自动化处理。

该定义是从语义网的设计原则角度来描述的。这些设计原则虽然可以为语义网开发人员提供技术层面的指导,但并不能作为构成语义网理论体系的基础。

定义 1.4:语义网就是一组技术和标准,用以实现机器对万维网中信息所携带语义的理解。

该定义空泛地使用语义网的相关技术来代替自身的属性,在应用技术和理论基础的认识上发生了混淆,对于理解语义网的内涵非常不利。

比较上述各种定义可以发现:这些定义都强调了语义网需要一种数据表示方法。尽管各种定义都从不同角度关注语义网,但都没有完整描述语义网各个方面的特征。

本书认为,语义网包含两个方面涵义:由机器可处理的信息所组成的抽象信息空间,以及由语义网技术所组成的研究对象。实际上,“Semantic”含有“机器可推理”的意思,而不是自然语言或人的推理,对信息来说,“Semantic”表达了对

信息能做哪些操作。

1.1.2 语义网整体框架

互联网超链接文件被认为只支持人类使用,而语义网被大众理解为一个数据网络(Web of Data)。语义网的想法是向网站创建机器可读和可理解的元数据以便实现机器检索数据、建立网站之间的连接和关系等。要将这个想法变成现实, Tim Berners - Lee 仿照开放系统互连(Open System Interaction, OSI)参考模型,提出了语义网7层整体框架(图 1 - 1),其中涉及多种方法和技术。

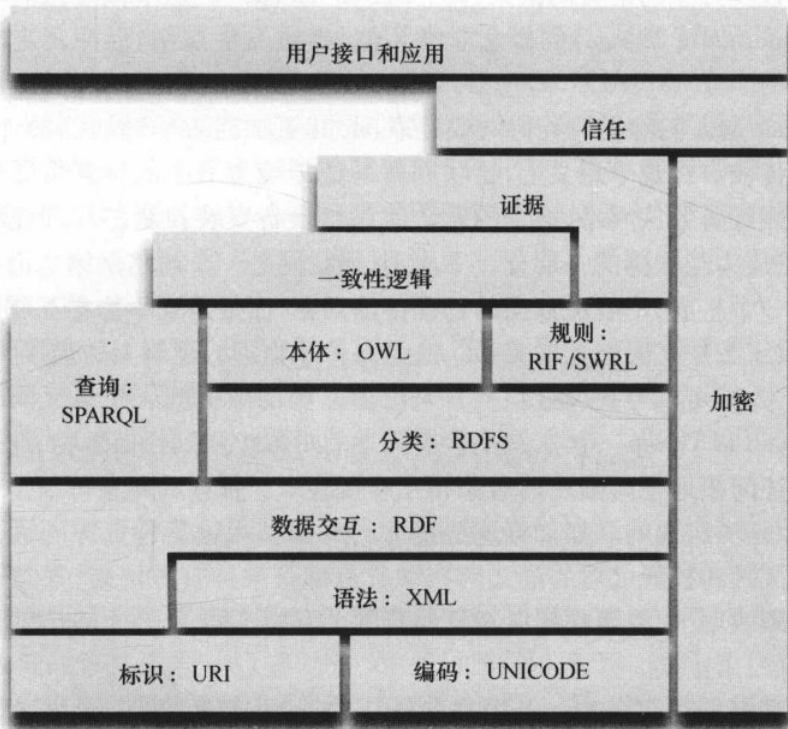


图 1 - 1 语义网整体框架

1.1.2.1 语义网标准集

语义网整体框架左下模块(加粗字体)给出了语义网所依赖的标准集。这些标准集允许语义网朝着机器支持和自动推理的新水平发展,在近年来可能实现,而在过去绝对不可能实现。另外,语义网社区内部的技术持续发展和交流也促进了在此方面达成共识和快速发展。

第一层:基础层

(1) 语义网通用交流平台的重要基础是独立于机器和厂商的规范的使用,这就允许将信号解释为符号。语义网使用统一字符编码标准(UNICODE)对网络通信提供支持,并作为最低层的构建模块之一。

(2) 第二个重要的支柱是通用资源标识符(Universal Resource Identifier, URI)的使用。网络上的每一个信息都是资源,并且可以用一种标准化的方式处理。

Tim Berners - Lee 认为,URI 不仅可以用来定位数据文件资源,理论上它可以被用来定位任何事物(包括现实世界中的各种物理事物)。在这种定位的基础上,人们就可以对这些物理事物及其相互关系进行逻辑描述,从而构建一个更加广义的包罗万象的语义网框架,即“Web of Things”。这个概念跟时下流行的物联网(Internet of Things)的概念非常相似,容易发生混淆,但两者之间是存在本质区别的。

Internet 是人们常说的英特网或互联网,其更加注重物理线路的连接,从而实现数据传输。在这个意义上,物联网就是把传统意义上的计算机通信网络扩展成为智能设备通信网络,通常的做法就是给设备安装控制芯片和互联网连接端口,使之成为互联网的一部分。本质上,物联网是一个通信网络。语义网则是万维网的一个扩展,不解决数据的物理传输问题,而是建立事物相互联系的语义描述,侧重于数据之间的互联关系。这是一种非物理的逻辑上的连接关系,通过这种连接,人们可以寻找、定位、整合自己所需要的数据集。把语义网的概念扩展成为“Web of Things”,其实是把数据描述的对象扩展到任何事物,通过一整套的语义描述间接地把所描述的对象相互连接起来。描述的对象可以是个体层面的,也可以是类层面的。如果将这些描述的对象都视为某种资源的话,万维网、初期的语义网和扩展之后的语义网在本质上都是某种资源网络,只是在语义描述机器处理效率上,以及在指向对象的范围上有所不同。

第二层:语法层

可扩展标记语言(eXtensible Markup Language, XML)为资源描述提供了基本语法,它允许在资源内部的通信,从而完成资源查找、提取等功能。这使得通过 XML Schema 来描述和限制元素的结构和内容成为可能。

第三层:数据层

(1) 资源描述框架(Resources Description Framework, RDF)允许将资源和资源之间的联系进行结构化,其作用类似于数据模型。大多数框架使用了 XML,但由于存在可选方案,这已不再是必须的。然而,只要数据需要进行一致地交换(包括针对一个或多个资源的事务的概念),就需要公共 RDF 的支持。RDF 以三元组的形式组织资源,从而提供更高层次操作的结构。

Turtle(Terse RDF Triple Language,简洁的 RDF 三元组语言)提供了一种候选的基本语法,使用了适于通用模式和数据类型的缩写语法,允许采用一种简洁的自然文本形式来描述 RDF 三元组。Turtle 是 W3C 一个团队的提案,并赢得了从业者的支持。

(2) RDF 不只是一个数据模型,而 RDF 模式(RDF Schema,RDFS)的使用可比作术语分类的定义,用以描述 RDF 资源的属性和种类。通过使用定义明确的联系类型来关联定义明确的结构并用通用术语进行标记,允许构建层次和复杂关系。RDF 采用通用的三元组,RDFS 规范了这些三元组的重要内容,并被业界所认可。

第四层:语义层

(1) Web 本体语言(Web Ontology Language,OWL)是对 RDFS 的扩展。它增加了许多标准化术语,如基数、等价、类型、枚举,来描述资源、属性和关系。如果所有的支撑资源都遵循相应的标准,OWL 则允许针对这些扩展进行推理。

(2) 资源查询并不要求资源按照 RDFS 和 OWL 格式进行编排。简单协议和 RDF 查询语言(Simple Protocol and RDF Query Language,SPARQL)支持三元组查询、三元组的交集及并集运算,以及其他操作。严格来说,SPARQL 可与 RDF、RDFS、OWL 共同使用。

(3) 规则交互格式(RIF)虽然还未标准化,但是被推荐使用。它允许在语义网内部进行约束通信,以确保各层次之间数据的一致性。一个基于资源的 OWL 替代方案是语义网规则语言(SWRL),它使用 OWL 的逻辑子集进行描述。

如果没有对可判定性及计算复杂性的问题进行恰当考虑,就不应决定使用 OWL。并不是所有的 OWL 剖面都是可判定的,并且即使它们都可判定,它们的应用也可能导致非多项式计算时间。仿真工程师应当关注 OWL DL(这是可判定的),要是能够关注为 OWL 2 所定义的剖面(OWL 2 EL、OWL 2 QL 和 OWL 2 RL)更好。关于更多信息,Hitzler 等(2009)给出了一个很好的可用语义网络工具和标准以及相应计算约束的评价。

1.1.2.2 一致性逻辑、证据和信任

除了本体层定义的术语关系和推理规则外,还需要有一个功能强大的逻辑语言来实现推理。这 3 层位于语义网整体框架的中部,也是语义表达的高级要求,目前正处于研究阶段,也有一些简单的示范性应用系统正在建设中。

第五层:逻辑推理层

逻辑推理层(Unifying Logic)提供了推理规则的描述手段,为智能服务提供基础,比如可利用分布在语义网上的各种断言或公理推理出新的知识。

第六层:逻辑验证层

逻辑验证层(Proof)通过运用逻辑推理层的规则进行逻辑推理和验证。

第七层:信任层

信任层(Trust)则负责为应用程序提供一种机制以决定是否信任给出的论证。证据语言允许服务代理在向客户代理发送断言的同时将推理路径也发送给客户代理,这样应用程序只需要包含一个普通的验证引擎就可以确定断言的真假。但是,证据语言只能根据语义网上已有的信息对断言给出逻辑证明,它并不能保证语义网上所有的信息都为“真”。因此,软件代理还需要使用数字签名和加密技术用来确保语义网信息的可信任性。

1.1.2.3 数字签名和加密

简单地说,数字签名(Digital Signature)就是一段数据加密块,机器和软件代理可以用它来无二义地验证某个信息是否由特定可信任的来源提供,它是实现语义网信任的关键技术。数字签名跨越了多层,是一种基于互联网的安全认证机制。当信息内容从一个层次传递到另一个层次时,允许使用数字签名说明内容的来源和安全性,这样接收方就可以通过数字签名鉴别其来源和安全性以决定是否接受。数字签名对于语义网及其他使用XML进行信息交换的系统非常重要。公共密钥加密算法是数字签名的基础。虽然公共密钥密码技术已存在较长时间了,但还没有真正被广泛应用。如果加上语义网各层支持,使一个团体在一定范围内可信任,就实现了信任层。这样,一些诸如电子商务等重要的应用就可以进入到语义网的实用领域中。

1.1.2.4 用户接口和应用

语义网整体框架顶部模块表示语义网的各种用户接口和应用,主要包括Web服务和典型状态传输(REST)。

(1) Web服务分为几类。早期的版本基于一个组合:使用XML定义用于交互的数据、采用简单对象访问协议(SOAP)来获取服务、使用基于XML的Web服务描述语言(WSDL)来描述服务的访问特性、使用同样基于XML的统一描述、发现与集成(UDDI)来支持发布和查询。一个常规的思路是Web服务采用WSDL描述服务的特性,并在UDDI服务器上发布。谁若是需要查找服务,可从UDDI下载服务特性的描述,并基于该描述查找合适的解决方案。如果该服务被找到并被访问,则搜索程序将使用XML提供所需数据并通过SOAP来访问该服务。缺点在于该过程从逻辑上与远程过程调用(RPC)非常接近,需要许多关于服务接口、信息交互需求等知识。

(2) 为了克服这些限制,典型状态传输(REST)方法应用于定义典型状态传输服务。REST的思路是将尽可能多的接口进行一般化,并将所需信息放入交