

陆建江 张亚非 编著
苗 壮 周 波

语义网原理与技术



 科学出版社
www.sciencep.com

语义网原理与技术

陆建江 张亚非
苗 壮 周 波 编著

科学出版社

北 京

内 容 简 介

随着 Web 的迅速发展,其应用范围变得更广,规模变得更大。传统的 Web 技术已经不能满足现在 Web 发展的需要。语义网汲取多个学科的研究成果,力图对 Web 上信息的表示和获取方式进行重大的改进,通过增加一些语义信息使得计算机能够参与 Web 信息的自动处理,从而更有效地利用 Web 信息。本书主要介绍语义网原理与技术,包括语义网的概念和体系结构,语义网知识表示模型——本体,语义网逻辑基础——描述逻辑,语义网表示语言,语义网查询语言,以及语义网安全。

本书可作为高等院校计算机专业高年级本科生和研究生的教材,也可作为相关领域学生的参考书。

图书在版编目(CIP)数据

语义网原理与技术 / 陆建江等编著. —北京:科学出版社,2007

ISBN 978-7-03-018694-2

I. 语… II. 陆… III. 语义网络 IV. TP18

中国版本图书馆 CIP 数据核字(2007)第 032246 号

责任编辑:张海娜 / 责任校对:鲁 素

责任印制:刘士平 / 封面设计:嘉华永盛

科学出版社出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2007年3月第 一 版 开本: B5(720×1000)

2008年4月第二次印刷 印张: 14 1/4

印数: 3 001—4 000 字数: 284 000

定价: 28.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

前 言

20 世纪下半叶,以计算机和通信技术为主导的信息技术革命极大地影响了人类的生活和生产方式。作为这场革命的里程碑,Web 从根本上改变了人类存储和交换信息的方式。从 1989 年开始,人类对 Web 的关注和频繁使用极大地促进了这项技术的发展和革新,也使得 Web 上的信息资源爆炸性地增长。经过十几年的发展,Web 已成为人类最大的数字图书馆。目前最强大的搜索引擎 Google 能检索超过 80 亿个网页的信息,而这只占整个 Web 不到 30% 的信息量。面对如此浩瀚的 Web 数据海洋,数据精确查询就像是海底捞针一样困难。简而言之,Web 已超过最初设计理念所能承载的规模,问题集中表现在:传统 Web 过于强调信息的表现,而忽略信息语义的表示,导致 Web 内容无法被机器理解,这极大地阻碍着 Web 信息的有效利用。

正是在这样的背景下,语义网的思想被提出来。语义网汲取人工智能、信息论、哲学和逻辑等学科的研究成果,力图对 Web 上信息的表示和获取方式进行重大改进,以解决目前使用 Web 时存在的问题。语义网的核心思想是通过增加一些语义信息,使得计算机能参与到自动处理 Web 信息的过程,从而有效利用 Web 信息,并为实现智能化的 Web 应用提供必要的技术基础。从 2000 年正式提出至今,语义网已成为一个新兴的学术研究热点,国内外众多科研机构都在开展语义网的相关研究。

本书涵盖语义网的主要内容,力争系统全面地介绍语义网的关键技术,并通过实例来说明,使得复杂概念容易被理解。本书包括八章:第一章概述语义网的基本概念,包括其体系结构和涉及的关键问题;第二章介绍可扩展标记语言 XML;第三章介绍资源描述框架 RDF;第四章介绍语义网知识表示模型——本体理论;第五章介绍描述逻辑,它是语义网上知识表示和推理的逻辑基础;第六章介绍本体语言 OWL;第七章介绍语义网查询语言;第八章介绍语义网的安全问题。

本书的问世是集体智慧的结晶,由陆建江、张亚非、苗壮和周波负责撰稿。另外,感谢康达周、李言辉、王进鹏、邓秀慧、徐洁和李珂玥等同学为本书的撰写付出的辛勤劳动。全书每一章的内容组织和细节都是经过多次讨论和修改才定稿,力求深入浅出,让读者轻松掌握相关的知识。尽管每一节、每一句、每篇参考文献、甚至每个标点我们都精心检查,但仍难免存在一些缺点和错误,殷切希望广大读者批评指正。

目 录

前 言

第一章 语义网导论	1
1.1 Web 的发展历史	1
1.1.1 Web 的诞生	2
1.1.2 第一代 Web	2
1.1.3 第二代 Web	2
1.1.4 第三代 Web——语义网	3
1.2 Web 的局限性	3
1.3 语义网的蓝图	4
1.3.1 语义网的诞生	4
1.3.2 语义网的思想	5
1.3.3 语义网的架构	5
1.4 语义网的研究问题	7
习题	8
第二章 可扩展标记语言 XML	9
2.1 XML 概述	9
2.1.1 标记语言的发展史	9
2.1.2 XML 与 HTML 的比较	10
2.2 XML 设计思想	11
2.2.1 持久性和自描述性	12
2.2.2 通用性和最小包含性	12
2.2.3 易修改性和可扩展性	13
2.3 XML 标记	13
2.3.1 可扩展的标记	13
2.3.2 XML 元素	14
2.3.3 XML 属性	15
2.3.4 XML 特殊标记	15
2.4 XML 名称空间	17
2.5 XML 树模型	18
2.6 XML 文档结构	20

2.6.1 文档类型定义 DTD	20
2.6.2 DTD 的不足与 XML Schema	22
2.7 XML Schema	23
2.7.1 简单类型定义	23
2.7.2 复杂类型定义	25
2.7.3 注释	26
2.8 小结	27
习题	29
第三章 资源描述框架 RDF	30
3.1 RDF 概述	30
3.1.1 RDF 的含义	30
3.1.2 RDF 的设计目的	31
3.1.3 RDF 与 XML	32
3.2 RDF 模型	33
3.2.1 基本概念	33
3.2.2 RDF 资源和词汇集	34
3.2.3 RDF 图	34
3.2.4 结构化特性与空节点	35
3.2.5 RDF 图的等价性	37
3.3 RDF/XML 语法	38
3.3.1 基本语法	38
3.3.2 简写语法	40
3.3.3 指派 URI 引用	42
3.3.4 类型文字	42
3.3.5 关于 RDF/XML 语法的争议	44
3.4 RDF 特殊表达	44
3.4.1 容器	45
3.4.2 集合	47
3.4.3 声明具体化	48
3.5 RDF Schema	49
3.5.1 简介	49
3.5.2 RDF 类	50
3.5.3 RDF 特性	51
3.5.4 RDF 特性约束	52
3.6 RDF 示例	54

3.7 RDF 语义	57
3.7.1 简单解释	57
3.7.2 rdf-解释	59
3.7.3 rdfs-解释	61
3.7.4 RDF 语义的不足	64
3.8 小结	64
习题	65
第四章 本体理论	66
4.1 本体定义	66
4.2 本体构建	69
4.2.1 本体构建的准则	69
4.2.2 构建的过程	70
4.2.3 本体编辑工具	73
4.2.4 本体生成技术	75
4.2.5 语义标注	76
4.3 典型本体介绍	80
4.3.1 分类法	80
4.3.2 Cyc	80
4.3.3 WordNet	80
4.3.4 SUMO	81
4.3.5 基因本体	81
4.4 本体映射	82
4.4.1 本体异构	82
4.4.2 解决本体异构的方法	83
4.5 本体映射分类	85
4.5.1 从映射的对象来看本体映射分类	85
4.5.2 从映射的功能来看本体映射分类	85
4.6 本体映射方法和工具	86
习题	87
第五章 描述逻辑	88
5.1 描述逻辑的发展	88
5.1.1 从网络到描述逻辑	89
5.1.2 描述逻辑发展阶段	90
5.2 描述逻辑 ALC	92
5.2.1 描述逻辑 ALC 的语法语义	92

5.2.2 描述逻辑 ALC 的知识库形式	93
5.3 描述逻辑 ALC 推理	94
5.3.1 预备知识	95
5.3.2 空 TBox 下的推理	98
5.3.3 非循环 TBox 下的推理	107
5.3.4 一般 TBox 下的推理	111
5.4 描述逻辑 ALC 的扩展	114
5.4.1 数量约束概念构造子	115
5.4.2 枚举概念构造子	120
5.4.3 关系层次	121
5.4.4 关系扩展	122
5.4.5 具体域	124
5.4.6 描述逻辑的复合扩展	129
5.4.7 描述逻辑构造子总结	130
习题	131
第六章 本体语言 OWL	132
6.1 早期本体语言	132
6.2 OWL 设计思路	134
6.2.1 描述逻辑	134
6.2.2 框架	135
6.2.3 RDF (S)	135
6.3 OWL 语法	136
6.3.1 OWL 语法概述	136
6.3.2 简单类、个体和属性	138
6.3.3 复杂类	141
6.3.4 属性的特性和约束	143
6.3.5 本体元素间的映射和版本管理	146
6.4 OWL 语言层次	148
6.5 OWL 语义	149
6.5.1 从抽象语法到描述逻辑	150
6.5.2 从抽象语法到 RDF 图	152
6.6 小结	152
习题	153
第七章 语义网查询语言	154
7.1 RDF 查询语言	154

7.1.1	RDF 查询语言 SeRQL	154
7.1.2	其他 RDF 查询语言	166
7.2	RDF 查询系统	169
7.2.1	RDF 的存储	169
7.2.2	RDF 查询的实现	170
7.3	OWL-QL	171
7.3.1	概述	171
7.3.2	查询-应答会话	172
7.3.3	查询问题	174
7.3.4	OWL-QL 语义	178
7.3.5	重复和冗余应答	180
7.3.6	OWL-QL 语法	182
	习题	184
第八章	语义网安全	185
8.1	语义网安全概论	185
8.2	XML 访问控制	186
8.2.1	XPath 简介	186
8.2.2	访问控制策略	187
8.2.3	访问控制机制	188
8.2.4	基于查询重写的 XML 访问控制方法	190
8.3	RDF 安全	199
8.3.1	RDF 加密	199
8.3.2	RDF (S) 推理控制	199
	习题	205
	参考文献	206

第一章 语义网导论

知识的记录和传播是人类文明延续和发展的根本保证。从远古的结绳记事和烽烟传信,到现代大容量高速度的存储设备和便捷的无线电、光纤通信设备,人类知识比以往任何时代都更容易被保存、组织、传播和检索。特别在信息时代的今天,Web 极大地改变了人类利用知识的形式。Web 是互联网上最重要的上层应用,它一方面作为信息的载体,以丰富的表现形式将信息及时重现;另一方面还提供各种检索和查询服务,使人们能方便获取所需信息。

尽管 Web 诞生至今只有短短的十多年,但人类对它的关注和频繁使用极大地促进了这项技术的不断发展和革新。从最初的静态 HTML 页面,到现在的动态 Web 页面;从最初对内容的简单描述,到今天将内容的结构和表示分离处理,无论 Web 的功能还是关键技术都发生了重大的变化。但到了 20 世纪 90 年代中期,随着 Web 信息量的膨胀,人类意识到缺乏自动处理 Web 海量信息的有效技术。因而人类面临一个尴尬的境地:我们创建了 Web 这个庞大的知识库,却无法有效利用它。针对这些问题,Berners-Lee 于 2000 年提出了下一代 Web 的概念——语义网。语义网汲取人工智能、哲学和逻辑等学科的研究成果,希望对 Web 上信息的表示和获取方式进行重大改进,以解决目前使用 Web 时存在的问题。语义网将改变计算机在 Web 中的角色,使得它不但可以保存和重现信息,而且能对信息进行智能化处理。语义网将是很多智能应用实现的基础。本书针对 Web 领域的最新发展目标——语义网,阐述语义网的结构,描述其主要概念和思想,探讨其中的关键技术,并介绍其相关的应用。

本章简单介绍 Web 的发展历史和语义网的出现背景,并介绍语义网的核心思想以及面临的关键问题。

1.1 Web 的发展历史

Web 是信息化时代到来的标志之一。Web 是互联网上最重要的,也是与普通用户最贴近的应用。自 1989 年诞生以来,Web 不断更新发展,无论从形式上还是功能上都发生了很大变化。回顾 Web 的发展历程,有利于了解 Web 的当前现状以及语义网提出的背景。

1.1.1 Web 的诞生

1989年,在日内瓦欧洲粒子物理实验室工作的 Berners-Lee 发明了最初的 Web。这个系统的开发基于三个方面的需求:①统一的用户界面;②具有将各种技术和文档类型协调起来的能力;③具有“普遍可读性”,就是说,在网络上任何地方的任何人,使用不同类型的计算机,都可以和别人一样轻松地查看文档。1990年10月,Berners-Lee 编写出第一个服务器和第一个客户端程序。1991年3月,Web 浏览器第一次上网使用。两个月后,该系统在欧洲粒子物理实验室中被大量使用。1991年10月,第一个支持广域信息服务搜索的网关被安装。1991年底,欧洲粒子物理实验室宣布在高能物理组中采用 Web。Web 开始了它真正的发展。

1.1.2 第一代 Web

第一代 Web 使得通过互联网浏览文档成为可能,即实现了文档表现形式与文档存储格式的分离。这时的 Web 以 HTML 语言、URL 和 HTTP 等技术为标志,以静态页面的形式来展现信息。最初的 Web 在实现上非常简单,用户只需要学习基本的 HTML 语言,并在一些网页设计工具的辅助下,就能轻松完成 Web 站点的建立。当要对页面进行更新时,只需改变相应的 HTML 文件内容。但是,如果 Web 站点上包含大量内容,维护这些内容就需要付出大量繁重的劳动。因此,第一代 Web 简单易用,但信息单调,维护困难。

第一代 Web 开创了通过互联网上交互信息的新时代,但它具有很多局限性。首先,最突出的问题是:HTML 作为一种简单的表示语言,它只能显示信息而无法表达信息的数据组织。而这一点恰恰是电子商务、智能搜索引擎等应用所必需的。其次,HTML 语言不能描述诸如矢量图形、数学公式和化学符号等特殊对象,在数据显示方面的能力也不尽如人意。再次,HTML 只是 SGML(Standard for general markup language)的一个实例化子集,可扩展性差,用户根本不能自定义有意义的标签供他人使用。最后,缺乏有效的搜索引擎来处理迅速增加的 Web 信息,这意味着用户很难方便地找到他们想要的东西。这一切都成为 Web 技术进一步发展的障碍。

1.1.3 第二代 Web

第二代 Web 以 Java script、VB script 和 XML(extensible markup language)等技术为标志。一方面,在 Java script 和 VB script 等新技术的支持下,实现了 Web 服务器的概念,它允许用户通过交互查询数据库来构建 Web 应用,并将数据库中符合要求的结果动态地生成页面,并展示给用户。这极大增强了 Web 处理大规模数据的能力。Web 由一个展示信息的平台真正变成了处理信息的平台。由

于 Web 具有统一的客户端界面——浏览器,因此大量的 Web 应用开始出现,如数字图书馆、网上购物和网上办公自动化等等。另一方面,由于可扩展标记语言 XML 的出现,使得 Web 实现了文档结构和文档表现形式的分离,即文档是存储数据的地方,但数据可以根据不同的用途有不同的表现形式。因此,第二代 Web 能更有效地表示信息。

1.1.4 第三代 Web——语义网

Web 的高速发展使得它很快成为一个庞大的知识库,但同时带来很多问题。首先,以往的 Web 技术都忽略了计算机的处理作用:计算机在其中主要扮演了展现信息的作用,而很少参与信息的处理,这使得很多智能应用无法在 Web 上实现。其次,Web 规模的快速膨胀使得整个 Web 上大部分有用信息无法得到有效利用。为解决这些矛盾,近年来提出的语义网就是要通过实现文档语义与文档结构的分离,达到使文档具有机器可处理的语义,从而解决传统 Web 无法解决的一些问题。

1.2 Web 的局限性

Web 已经成为人类获取信息和得到服务的主要渠道之一,但是 Web 并非已经尽善尽美,仍然存在很多尚待解决的问题。

(1) Web 信息无法被自动处理。当前的 Web 无论是手写的 HTML 网页,还是根据查询动态生成的网页,其目的都是供人阅读,计算机只负责显示这些信息,而没有理解和处理 Web 信息的能力。因此,在 Web 上很难开发各种智能应用。

(2) Web 信息无法被有效利用。Web 的规模不断变大,这对 Web 信息的有效利用提出了巨大挑战。基于传统技术的搜索引擎已经无法应对 Web 这个日益庞大的知识库。以最强大的搜索引擎 Google 来说,它目前能搜索 80 亿之多的 Web 页面,但这仅仅占整个 Web 规模的 25%~30%,也就是说大量的 Web 信息无法被搜索到。同时,由于计算机无法精确识别 Web 上的内容,当前搜索引擎返回的信息要么过多,要么过少,搜索结果的质量并不令人满意。

(3) Web 服务存在异构问题。在 Web 服务方面,尽管提出了 UDDI、WSDL 等技术标准,但仍然存在很多尚待解决的问题:服务发现、匹配、检索的查全率和查准率较低;服务集成仍然需要人工干预,不能完全自动化。在 CORBA、COM 和 EJB 等以跨平台互操作为目标的中间件研究和应用方面,也同样存在中间件描述、检索和自动化集成等方面的问题。无论是简单的网页信息,还是网络服务或中间件等复杂的信息和服务的聚集体,它们的主要问题都在于:描述信息的语义二义性导致机器无法自动地理解 and 处理它们。

随着 Web 上信息量的不断增加,上述的这些问题还会进一步加剧。作为一个

全球性的信息网络, Web 还远没有充分发挥它的潜能。机器不能自动地、有效地、有目的地发现、集成和复用 Web 上的各种数据, 这正是创建智能化 Web 服务的根本障碍。

1.3 语义网的蓝图

1.3.1 语义网的诞生

为克服自身的不足, Web 在其发展过程中也不断进行改进。针对 HTML 语言所暴露出来的各种局限性, W3C 于 1996 年开始开发一种超越 HTML 能力的新语言, 这就是 XML。XML 是 Web 发展中的一个重要的里程碑, 它的贡献主要在于解决了结构化和半结构化数据在语法层次上的互操作问题, 为信息语义层的互操作打下了基础, 从而大大促进了 Web 服务、电子商务的发展。但针对 Web 的改进并不局限于 XML, 因为虽然 XML 解决了语法层次的互操作问题, 但 Web 应用仍然难以实现智能化。

1997 年 12 月, Berners-Lee 在 W3C 的会议报告中指出: Web 的第一个目标是成为人们通过知识共享进行交流的媒介, 第二个目标是成为人们协同工作的媒介。他认为, 如果所有的数据都有一个公共的、定义良好的接口, Web 协作将更加方便。这次关键性的会议为语义网的正式提出打下了基础。

1998 年 9 月, Berners-Lee 首次提出了语义网的设想, 给出了它的总体框架, 并概括介绍了它的各个组成部分(Berners-Lee, 1998a), 从而拉开了语义网研究的序幕。Berners-Lee 又在一系列的文献中对语义网的思想作了补充说明(Berners-Lee, 1998b; 1998c; 1998d), 对语义网的基础和资源描述框架中的有向标记图模型, 与其他数据模型作了详细的对比。这些文献阐述了语义网与人工智能、知识表达、实体联系模型、关系数据库以及推理系统之间的联系和区别, 为接受语义网思想扫清了障碍。此后, Berners-Lee 特别指出了语义网在万维网未来发展中的重要作用(Berners-Lee, 1999)。

2000 年 12 月, Berners-Lee 在 XML2000 会议报告中首次给出了语义网功能逐层增强的层次结构图, 指明了语义网的研究框架(Berners-Lee, 2000a)。这次大会标志着语义网研究的真正全面展开。随后, 通过 *Nature* 这样的高水平学术杂志和 *Scientific American* 这样的著名科普杂志, Berners-Lee 将语义网的想法和动人前景介绍给研究者和普通大众(Berners-Lee et al., 2001a; 2001b)。

2001 年, 语义网工作研讨会在 Stanford 大学召开。2002 年 6 月, 第一届语义网国际会议 ISWC2002 在意大利召开, 该会议成为语义网领域最重要的学术会议, 代表了最近的研究成果和研究方向。与此同时, 代表万维网研究最高水平的 WWW 会议也从 2002 年起出现了针对语义网的研究论文。语义网的研究全面展

开了。

1.3.2 语义网的思想

根据 Berners-Lee 的观点,语义网并非是全新的 Web,而是对现有 Web 的扩展。它与传统 Web 的不同在于,信息的语义在语义网环境下能够很好地加以定义,并使得计算机和人类能够更好地协同工作。或者说,语义网的目标是让 Web 上的信息能够被机器理解,从而实现 Web 信息的自动处理,以适应 Web 资源的快速增长,更好地为人类服务。

W3C 是这样阐述语义网的:将机器能理解的数据发布在 Web 上正成为很多组织最优先考虑的工作,只有 Web 成为一个自动工具和人能够共享和处理数据的平台,它的潜能才可能全部发挥。对于 Web 这个范围来说,未来的程序必须能够共享和处理数据,即使它们是独立设计的。语义网是这样的一个设想:使 Web 上的数据能以一种可以被机器所理解的方式定义和联系起来,它不仅仅以显示为目的,而是为了自动集成和重用不同平台中的数据。

在语义网环境下,Web 上定义和链接的数据不仅能显示,而且可以被机器自动处理、集成和重用。只有当数据不仅可以被人而且可以被机器自动地共享和处理的时候,Web 的潜力才发挥到极致。但是,“机器可理解”并不意味着机器能够理解人类的语言,它只是说明:机器根据明确定义的数据,通过执行明确定义的操作,来解决明确定义的问题。

1.3.3 语义网的架构

根据 Berners-Lee 的设想,语义网是由一种分层的体系结构构成,如图 1-1。这是一个功能逐层增强的层次化结构,由七个层次构成(Berners-Lee,2000a)。

URI(uniform resource identifier)和 Unicode 层是标识语义网对象和统一使用国际字符集的基本手段。XML、名称空间和 XML Schema 定义了语义网上语法互操作的标准。RDF 和 RDF Schema(二者合称为 RDF(S))用来描述和定义语义网上的资源。本体层用来定义不同概念之间的关系,以支持词汇的演化。逻辑层为基于规则的系统提供一个描述公理的框架。证明层执行规则并做出相应的评估。信任层为应用程序是否信任一个给定的证明提供检测机制。数字签名和加密技术用来检测文档的改动情况,是增强 Web 安全的手段。

(1) URI 和 Unicode。URI 是 Web 的核心概念之一,它能够唯一地标识 Web 上的任意一个资源,其思想是在需要的时候通过链接引用资源,因此不需要对资源进行拷贝或集中管理。Unicode 是一种新的字符编码标准,它支持世界上所有的语言。无论在什么平台上,无论在什么程序中,无论使用什么语言,每个字符都对应于一个唯一的 Unicode 编码值。因此,它是语义网多语种支持的基础。

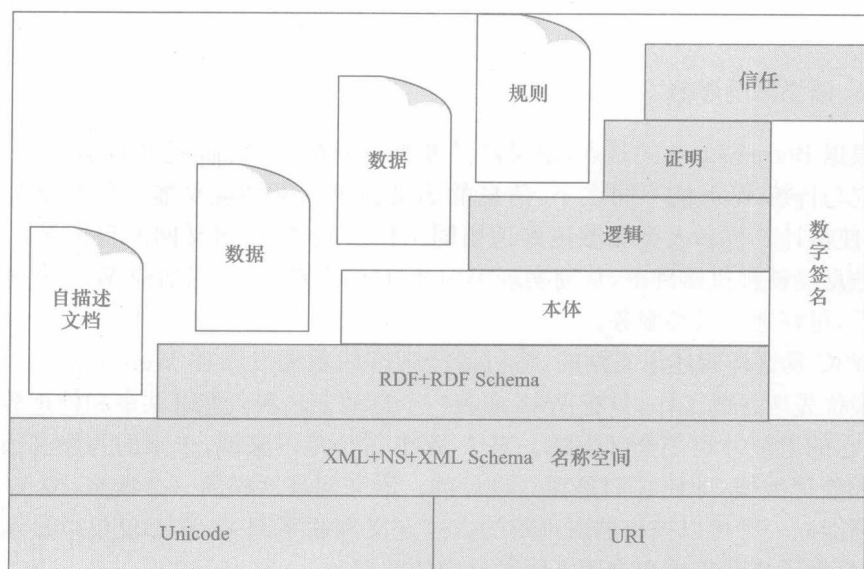


图 1-1 语义网体系结构

(2) XML、名称空间和 XML Schema。XML 提供文档结构化的语法,实现了文档结构与文档表现形式的分离,根据不同的目的同一个文档可以有不同的表现形式。XML 名称空间是名称的一个集合,用于文档元素和属性名有效性的验证,由 URI 引用来标识。XML Schema 是约束 XML 文档结构的语言。

(3) RDF(S)。XML 实现了文档结构化,但文档信息并不包含任何语义。RDF 数据模型提供简单的语义,RDF 属性可以看作是资源的属性,同时又表达了资源之间的关系,因此 RDF 数据模型对应于传统的属性二值对,又类似于 ER 图。RDF Schema 为 RDF 模型提供了一个基本的类型系统,其目的就是定义资源的属性,定义被描述资源的类,并对类和关系的可能组合进行约束,同时提供约束违例的检测机制。

(4) 本体层。虽然 RDF(S)能够定义对象的属性和类,并且还提供了类的泛化等简单语义,但它不能明确表达描述属性或类的术语的含义及术语间的关系。本体层就是要提供一个能明确的形式化语言,以准确定义术语语义及术语间关系。

(5) 逻辑、证明和信任。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。证明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,证明语言只能根据 Web 上已有的信息对断言给出逻辑证明,它并不能保证 Web 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 Web 信息的可信任性。

(6) 数字签名和加密。数字签名简单地讲就是一段数据加密块,机器和软件代理可以用它来唯一地验证某个信息是否由特定的可信任的来源提供。它是实现 Web 信任的关键技术。公共密钥加密算法是数字签名的基础。

1.4 语义网的研究问题

在目前看来,语义网涉及的主要研究问题包含如下几个方面。

(1) 语义网知识表示模型。语义网采用本体作为知识表示模型。本体形式化定义了领域内共同认可的知识,是语义网体系中的核心。因此,如何创建和管理本体是实现语义网上知识表示的基础。此外,本体的扩充、集成和演化等问题也常常需要考虑。

(2) 语义网逻辑基础。语义网的逻辑基础是描述逻辑。它是一阶逻辑的子集,这个子集具有可判定性的特点,并适合处理大规模的数据。语义网语言的表达能力直接依赖于相应的描述逻辑,推理问题也可由描述逻辑的推理算法实现。

(3) 语义网表示语言。表示语言提供了语义网中的领域信息描述的基础,是语义网的核心。语义网中需要通过表示语言来描述领域中的资源,包括分类、属性、子类关系、子属性关系、实例关系、定义域和值域等等。目前的语义网表示语言包括 XML、RDF(S)和 OWL(Web ontology language)等。

(4) 语义网查询技术。查询技术是语义网的基本技术,语义网的各种应用都离不开查询技术的支持。语义网查询的对象是以各种语义网表示语言表示的文档或知识库,因此针对各种表示语言(XML、RDF(S)和 OWL 等),发展了各自的查询语言。

(5) 语义网安全。安全主题是语义网实现的必要条件。语义网安全事务涉及语义网结构中的所有层次。安全包含每一层次的安全以及互用性的安全。高层的安全依赖底层的安全。如果在语义网的较低层次(如 XML 层)的安全性存在缺陷乃至网络安全暴露出漏洞,那么再严格的高层安全机制也可能被绕过。

语义网构想了 Web 辉煌的未来。但在语义网的实现道路上还有很多尚待解决的重要问题,而且对于这些问题研究者各抒己见,莫衷一是。并非所有的研究人员都持接受语义网的观点,甚至有些对语义网持反对观点。在他们看来,语义网所设想的目标过于远大和不切实际,其技术实现等多方面都存在着巨大的困难。尽管有这些反对或消极的看法,还是有大量的科研机构、大学中的科研人员投入到相关的研究中。因此,无论最终的结果如何,我们都相信语义网的研究会带来很多新的成果。

习 题

- 1.1 简述 Web 的发展历程。
- 1.2 传统 Web 存在哪些不足?
- 1.3 语义网的目标和思想是什么?
- 1.4 论述语义网的体系结构。