

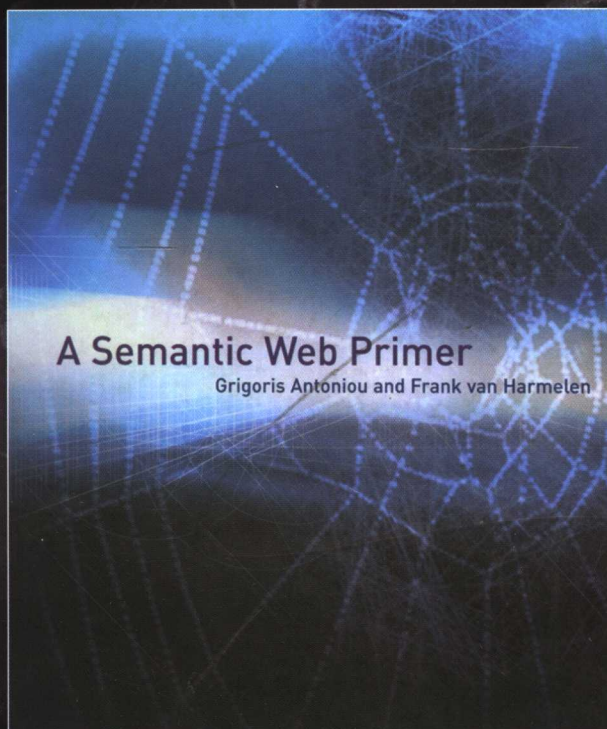
HZ BOOKS
华章教育



计 算 机 科 学 丛 书

语义网基础教程

(希) Grigoris Antoniou (荷) Frank van Harmelen 著 陈小平 等译
赫拉斯研究与技术基金会 Vrije大学 中国科学技术大学



A Semantic Web Primer



机械工业出版社
China Machine Press

计 算 机

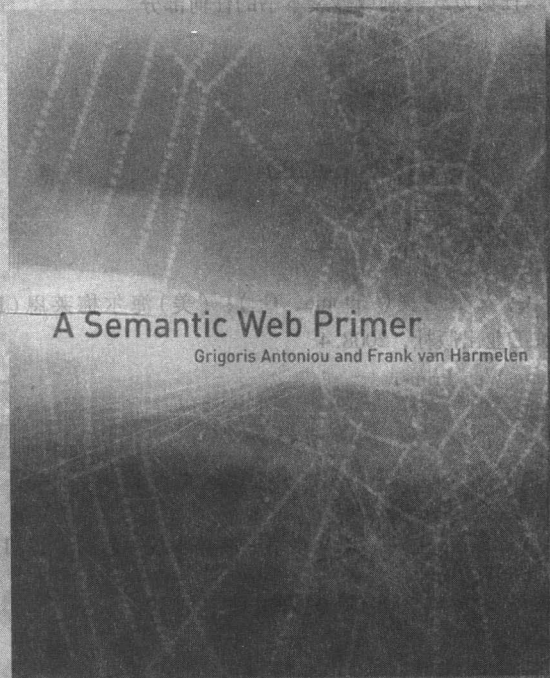
TP18/166

2008

书

语义网基础教程

(希) Grigoris Antoniou (荷) Frank van Harmelen 著 陈小平 等译
赫拉斯研究与技术基金会 Vrije 大学 中国科学技术大学



A Semantic Web Primer



机械工业出版社
China Machine Press

本书是一本语义网的入门性教科书,内容包括语义网技术概述、XML 语言、结构化文档、RDF 和 RDF Schema、网络本体语言 OWL、与语义网相关的逻辑与推理规则和本体工程等方面。本书还介绍了一些知名行业公司的语义网应用实例,极具实用参考价值。

本书主要用于大学语义网课程的教学,也是语义网技术专业人士十分难得的参考读物。本书还可作为大学计算机专业本科生教材。

Grigoris Antoniou, Frank van Harmelen: A Semantic Web Primer (ISBN 0-262-01210-3).

Authorized translation from the English language edition published by MIT Press.

Copyright © 2004 by MIT Press.

All rights reserved. No part of this publication may be reproduced or distributed in any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Chinese simplified language edition published by China Machine Press.

Copyright © 2008 by China Machine Press.

本书中文简体字版由美国麻省理工学院出版社授权机械工业出版社独家出版。未经出版者预先书面许可,不得以任何方式复制或抄袭本书的任何部分。

版权所有,侵权必究。

本书法律顾问 北京市展达律师事务所

本书版权登记号:图字:01-2006-4280

图书在版编目(CIP)数据

语义网基础教程/(美)安东尼奥(Antonio, G.);(美)海尔梅莱恩(Harmelen, F.)著;陈小平译. —北京:机械工业出版社,2008.4

书名原文:A Semantic Web Primer

(计算机科学丛书)

ISBN 978-7-111-23734-1

I. 语… II. ①安… ②陈… III. 语义网络-高等学校-教材 IV. TP18

中国版本图书馆 CIP 数据核字(2008)第 036291 号

机械工业出版社(北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑:王 玉

北京京北制版厂印刷 · 新华书店北京发行所发行

2008 年 4 月第 1 版第 1 次印刷

184mm × 260mm · 12.5 印张

标准书号:ISBN 978-7-111-23734-1

定价:32.00 元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换

本社购书热线:(010)68326294

出版者的话

文艺复兴以降，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭橥了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短、从业人员较少的现状下，美国等发达国家在其计算机科学发展的几十年间积淀的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章图文信息有限公司较早意识到“出版要为教育服务”。自1998年开始，华章公司就将工作重点放在了遴选、移译国外优秀教材上。经过几年的不懈努力，我们与Prentice Hall, Addison-Wesley, McGraw-Hill, Morgan Kaufmann等世界著名出版公司建立了良好的合作关系，从它们现有的数百种教材中甄选出Tanenbaum, Stroustrup, Kernighan, Jim Gray等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及度藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力襄助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专程为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近260个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍，为进一步推广与发展打下了坚实的基础。

随着学科建设的初步完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都步入一个新的阶段。为此，华章公司将加大引进教材的力度，除“计算机科学丛书”之外，对影印版的教材，则单独开辟出“经典原版书库”。为了保证这两套丛书的权威性，同时也为了更好地为学校和老师服务，华章公司聘请了中国科学院、北京大学、清华大

学、国防科技大学、复旦大学、上海交通大学、南京大学、浙江大学、中国科技大学、哈尔滨工业大学、西安交通大学、中国人民大学、北京航空航天大学、北京邮电大学、中山大学、解放军理工大学、郑州大学、湖北工学院、中国国家信息安全测评认证中心等国内重点大学和科研机构在计算机的各个领域的著名学者组成“专家指导委员会”，为我们提供选题意见和出版监督。

这两套丛书是响应教育部提出的使用外版教材的号召，为国内高校的计算机及相关专业的教学度身订造的。其中许多教材均已为 M. I. T., Stanford, U. C. Berkeley, C. M. U. 等世界名牌大学所采用。不仅涵盖了程序设计、数据结构、操作系统、计算机体系结构、数据库、编译原理、软件工程、图形学、通信与网络、离散数学等国内大学计算机专业普遍开设的核心课程，而且各具特色—有的出自语言设计者之手、有的历经三十年而不衰、有的已被全世界的几百所高校采用。在这些圆熟通博的名师大作的指引之下，读者必将在计算机科学的宫殿中由登堂而入室。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证，但我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。教材的出版只是我们的后续服务的起点。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方法如下：

电子邮件：hzjsj@hzbook.com

联系电话：(010)68995264

联系地址：北京市西城区百万庄南街1号

邮政编码：100037

专家指导委员会

(按姓氏笔画顺序)

尤晋元	王 珊	冯博琴	史忠植	史美林
石教英	吕 建	孙玉芳	吴世忠	吴时霖
张立昂	李伟琴	李师贤	李建中	杨冬青
邵维忠	陆丽娜	陆鑫达	陈向群	周伯生
周克定	周傲英	孟小峰	岳丽华	范 明
郑国梁	施伯乐	钟玉琢	唐世渭	袁崇义
高传善	梅 宏	程 旭	程时端	谢希仁
裘宗燕	戴 葵			

中文版前言

自 20 世纪 90 年代末发端以来，语义网研究已经成为一项全球性事业，相关的各种研究项目在欧洲和美国几乎同时出现，亚洲的研究者也表现出极大的兴趣。

从那时以来，中国学者在本体映射、推理技术、基于本体的浏览与搜索等各个方面展开了研究工作，形成了非常活跃的研究局面。通过这些工作，中国已经成为国际语义网研究领域中的一个重要成员。2006 年第一届亚洲语义网大会在北京举行。中国学者在语义网的其他国际学术会议上也占据了重要位置，最近还在一个语义匹配的竞赛中取得了领先地位。此外，中国还有一个非常活跃的讨论语义网网站(<http://bbs.w3china.org/>)。

语义网研发的一个首要目标是实现语义可共用性(semantic interoperability)。当简单的语法可共用性无法实现或难以满足应用需求时，就必须借助语义可共用性。在教学中，我们经常通过欧洲语言与汉语之间的巨大语法差异来说明不同语言背景的读者仅借助语法手段实现不同语种材料的共用是多么困难。因此，本书中文版的出版是很有意义的。这部十分出色的中译本由中国科学技术大学陈小平教授主持完成。

我们非常高兴陈小平教授和他的团队愿意承担本书的翻译工作，他们做得非常出色。我们在访问中国科学技术大学期间，与译者对翻译中遇到的问题进行了广泛、深入的讨论。我们通过回答这些问题对原著的内容做进一步澄清，但更多情况下，我们的讨论涉及原著的更深层内容，甚至远远超越原著的范围，进入对前沿问题的探讨。在翻译过程中，荷兰阿姆斯特丹 Vrije 大学的高级研究员黄智生博士也对翻译提供了专业性支持，帮助澄清了翻译中的很多技术性问题。

作为作者，我们认为他们是最好的翻译团队。我们完全相信中译本的质量和完整性；并且诚挚地希望本书中文版将为从事语义网的教与学的师生提供有益的帮助，并对进一步提升语义网的研发兴趣发挥积极作用。

Since it's inception in the late '90s, the Semantic Web programme has been an example of a truly global enterprise. Almost simultaneously, research programs arose in Europe and the US, and there was significant interest from researchers in Asia.

Since those early years, Chinese researchers have built a very active research community, with activities in many different research areas, such as ontology mapping, inference techniques, ontology-based browsing and searching, and many others.

Through these activities, China has become a significant player in Semantic Web research. China was hosting the first Asian Semantic Web Conference in Beijing in 2006, Chinese research groups have a significant presence at international Semantic Web conferences, there is a very active web-site and bulletin board on Semantic Web topics (<http://bbs.w3china.org/>), and a Chinese entry gained a leading position in recent Ontology Alignment Challenges.

A major goal of Semantic Web research and development is to enable semantic interoperability. Semantic interoperability is required when simple syntactic interoperability is not achievable, or not sufficient for a particular task. In lectures, we often use the large syntactic differences between European languages and the Chinese language to illustrate how hard is to achieve interoperability using only syntactic means. It is therefore fitting that this Chinese translation of our Semantic Web Primer has now become available. This very fine translation is from the hand of Prof. XiaoPing Chen of the University of Science and Technology (USTC) in Hefei.

We were very pleased with the offer of Prof. XiaoPing Chen of the University of Science and Technology (USTC, Hefei) to translate this volume. He and his team have done a remarkable job. During an invited visit to their group they asked many knowledgeable, detailed and probing questions about unclarities they had come across during the translation process. Sometimes, answering their questions was a matter of providing some clarification, but more often, their questions pointed at deep issues underlying the text, and often lead to technical discussions reaching far beyond the text of the Primer, and into open research issues. Prof. XiaoPing Chen and his team were professionally supported by dr. Zhisheng Huang, a senior researcher at the Vrije Universiteit Amsterdam, who helped to clarify many detailed translation issues.

We could not have wished for a better translating team. We have all confidence in the quality and integrity of the translation, and we sincerely hope that this translated edition of the Semantic Web Primer will contribute to further strengthen the interests in Semantic Web research and development, and that it will be a useful aid to students and teachers alike.

Frank van Harmelen

Grigoris Antoniou

译者序

语义网是一个有些“奇特”的新兴领域。它是现实主义与理想主义的奇妙组合。其现实主义主要表现为坚持以“易行性”(tractability)为语义网技术的首要条件,从而确保技术成果计算上的可行性,这与传统人工智能的风格不尽相同。另一方面,它又设立了很高的目标,试图通过网络信息“理解”的机械化而实现网络信息处理的根本性变革,这与万维网及其应用的现行技术路线也不一样。可以预见,语义网技术一旦市场化,必然引起万维网应用的全面改观。

语义网飞速发展导致本领域信息严重过量,从而给初学者带来极大不便。如原版前言指出的那样,本书的主旨是要解决这一难题。确实,作为一本语义网技术的导论性教材,尤其相对于国内众多读者,本书在很多方面是相当出色的——简练而不简陋、通俗而不肤浅、前沿而不学究……原著的出版对于读者来说是值得庆幸的,对于译者而言却实属不幸。如果译者的工作能对中文读者有所助益,将是不幸中之万幸!经过两年多的努力,中译本终于与读者见面了。对于翻译中的种种疏漏差错,请读者批评指正。

本书的翻译是集体合作的结果,并经过了多轮翻译和修订。分工如下:周熠(第1、8章)、杨斌(第7章)、姜杰汇、蒙云松(第6章)、宋志伟、李永喜(第4章)、李欢、吉建民(第5章)、杨方凯(第3章)、范正洁(第2章)。陈小平对所有译稿作了全面修订。

原作者 Frank van Harmelen 解答了一些疑难问题。原作者的同事、语义网专家黄智生博士为本书的翻译提供了大量帮助。对机械工业出版社的编辑所做的工作,在此一并致以谢忱!

译者

2007年6月

前言

万维网(WWW)的出现改变了人与人之间交流的方式,改变了信息传播、获取的方式和商业运作的方式。语义网(Semantic Web)包含着能够显著改进当前万维网及其应用的技术。本书旨在介绍这门发展中的新兴技术。

判断一本书的成功与否应当参照作者的写作目的。本书是一本语义网的入门性教科书,主要用于大学语义网课程的教学,同时也适用于语义网技术的自学。

既然所有关于语义网的信息都可以在网上找到,到底有没有必要编写这样一本教材呢?作者认为有这个必要,因为在网上有太多质量参差不齐的资源 and 过量的信息。这些信息有的是正确的,有的却是过时的或错误的,而且大部分讨论的是具体细节。任何初学者或想开设语义网课程的教师都面临着这些问题。本书的目的就是帮助读者解决这个难题。

一部教科书必须对所涵盖的内容有所选择,尤其是在这样一个飞速发展的领域里,更要着眼于那些直到将来仍然有价值的基础内容。当然,作者总是难以摆脱自己的偏爱。

即便对于被涵盖的内容,本书也无意成为一本包含全部细节的工具书。对于某些专题,比如 XML,已经有很多专著。由于关于语义网技术的所有定义和手册都可以在网上找到,所以没有必要为语义网这个领域专门写一本工具书。因此,本书着眼于介绍语义网的基本思想和主要技术,并为读者灵活地掌握和应用这些基础内容提供充分的细节,进而使读者有能力从其他来源自行获得所需的其余细节。在每章的最后,给出一些相关参考文献并附带必要的说明。

致谢

感谢 Jeen Broekstra、Michel Klein 和 Marta Sabou,为我们在荷兰 Vrije 大学开设基于万维网的知识表示课程提供很多第一手资料;感谢 Annette ten Teije、Zharko Aleksovski 和 Wouter Jansweijer 仔细阅读本书的初稿。感谢 Christoph Grimmer 和 Peter Koenig 校对部分书稿并帮助完成图表的生成和 LaTeX 处理。此外,感谢为本书终稿提供专业协助的 MIT 出版社员工和定义 $\LaTeX 2_{\epsilon}$ 宏命令的 Christopher Manning。

目 录

出版者的话	
专家指导委员会	
中文版前言	
译者序	
前言	
第1章 语义网概览..... 1	
1.1 万维网现状 1	
1.2 从当前万维网到语义网：实例 分析 3	
1.2.1 知识管理 3	
1.2.2 企业对个人电子商务 3	
1.2.3 企业对企业电子商务 4	
1.2.4 个人代理：未来的展望 5	
1.3 语义网技术 5	
1.3.1 显式的元数据 6	
1.3.2 本体 7	
1.3.3 逻辑 9	
1.3.4 代理 11	
1.3.5 语义网和人工智能的对比 ... 12	
1.4 语义网的逐层递进 12	
1.5 全书总览 14	
1.6 小结 14	
推荐读物 15	
第2章 XML 结构化网络文档 17	
2.1 引言 17	
2.2 XML 语言 20	
2.2.1 序言 20	
2.2.2 元素 20	
2.2.3 属性 21	
2.2.4 注释 21	
2.2.5 处理指令 22	
2.2.6 合式 XML 文档 22	
2.2.7 XML 文档的树形模型 22	
2.3 结构化 23	
2.3.1 DTD 方式 24	
2.3.2 XML Schema 28	
2.4 命名空间 33	
2.5 XML 文档的寻址和查询 35	
2.6 处理 38	
2.7 小结 43	
推荐读物 44	
练习与设计 45	
第3章 用 RDF 描述网络资源 47	
3.1 引言 47	
3.2 RDF 的基本思想 49	
3.2.1 资源 49	
3.2.2 属性 49	
3.2.3 陈述 49	
3.2.4 关于陈述的三种观点 49	
3.2.5 陈述的具体化 51	
3.2.6 数据类型 51	
3.2.7 对 RDF 的评论 52	
3.3 基于 XML 的 RDF 语法 53	
3.3.1 rdf: resource 属性 55	

3.3.2 嵌套描述	56	3.9.3 模式查询	81
3.3.3 rdf:type 元素	57	3.10 小结	82
3.3.4 简略语法	57	推荐读物	82
3.3.5 容器元素	58	练习与设计	83
3.3.6 具体化	62	第4章 网络本体语言 OWL	85
3.4 RDF Schema 的基本思想	62	4.1 引言	85
3.4.1 类和属性	62	4.1.1 本体语言需求	85
3.4.2 类层次结构和继承	63	4.1.2 RDF Schema 表达能力的 局限性	86
3.4.3 属性层次结构	64	4.1.3 OWL 与 RDF/RDFS 的 兼容性	87
3.4.4 RDF 与 RDFS 层次	64	4.1.4 三个 OWL 子语言	87
3.5 RDF Schema 语言	65	4.2 OWL 语言	89
3.5.1 核心类	66	4.2.1 语法	89
3.5.2 用于定义关系的核心属性	66	4.2.2 头部	89
3.5.3 用于约束属性的核心属性	67	4.2.3 类元素	90
3.5.4 用于具体化的常用属性	68	4.2.4 属性元素	90
3.5.5 容器类	68	4.2.5 属性约束	92
3.5.6 工具属性	68	4.2.6 特殊性质	94
3.5.7 例子: 大学	69	4.2.7 布尔组合	95
3.5.8 例子: 机动车	70	4.2.8 枚举	96
3.6 用 RDF Schema 描述 RDF 和 RDF Schema	71	4.2.9 实例	96
3.6.1 RDF	71	4.2.10 数据类型	97
3.6.2 RDF Schema	72	4.2.11 版本控制信息	98
3.7 RDF 和 RDF Schema 的公理 语义	73	4.2.12 OWL 分层	98
3.7.1 基本方法	73	4.3 例子	99
3.7.2 基本谓词	74	4.3.1 一个非洲野生生物本体	99
3.7.3 RDF	74	4.3.2 一个打印机本体	103
3.7.4 RDF Schema	76	4.4 OWL 自描述	107
3.8 RDF 和 RDFS 的直接推理系统	78	4.4.1 命名空间	107
3.9 使用 RQL 查询	79	4.4.2 元类	108
3.9.1 基本查询	80	4.4.3 类等价	108
3.9.2 使用 select-from-where 查询	80	4.4.4 类的组合	110
		4.4.5 类属性约束	110

4.5 未来扩展	112	5.7.3 事实	130
4.5.1 模块与导入	112	5.7.4 规则	131
4.5.2 默认值	113	5.7.5 查询	131
4.5.3 封闭世界假设	113	5.7.6 一个 DTD	132
4.5.4 唯一名假设	113	5.7.7 RuleML 的另一种数据 模型	132
4.5.5 过程附加	113	5.8 XML 中的规则标记: 非单调 规则	134
4.5.6 属性链接规则	113	5.8.1 例子	134
4.6 小结	114	5.8.2 一个 DTD	135
推荐读物	114	5.9 小结	136
练习与设计	115	推荐读物	136
第 5 章 逻辑和推理: 规则	117	练习与设计	137
5.1 引言	117	第 6 章 应用	139
5.2 单调规则的例子: 家庭关系	119	6.1 引言	139
5.3 单调规则: 语法	120	6.2 Elsevier 的横向信息产品	139
5.3.1 规则	120	6.2.1 背景	139
5.3.2 事实	121	6.2.2 问题	139
5.3.3 逻辑程序	121	6.2.3 语义网技术的作用	140
5.3.4 目标	121	6.3 奥迪的数据整合	141
5.4 单调规则: 语义	122	6.3.1 背景	141
5.4.1 谓词逻辑语义	122	6.3.2 问题	141
5.4.2 常证据和带参证据	123	6.3.3 语义网技术的作用	141
5.5 非单调规则: 动机和语法	124	6.4 瑞士人寿保险公司的技能 寻获	143
5.5.1 非正式讨论	124	6.4.1 背景	143
5.5.2 语法定义	125	6.4.2 问题	144
5.6 非单调规则的例子: 有经纪人的 交易	126	6.4.3 语义网技术的作用	144
5.6.1 卡洛斯需求的形式化 描述	126	6.5 EnerSearch 的智囊门户网站	145
5.6.2 可选公寓的表达	127	6.5.1 背景	145
5.6.3 选择一间公寓	128	6.5.2 问题	146
5.7 XML 中的规则标记: 单调 规则	129	6.5.3 语义网技术的作用	146
5.7.1 项	129	6.6 在线学习	148
5.7.2 原子公式	130	6.6.1 背景	148

6.6.2 问题	149	7.3.1 专家知识汇编	161
6.6.3 语义网技术的作用	149	7.3.2 整合词汇表	161
6.6.4 在线学习的本体	150	7.3.3 高层本体	161
6.7 Web 服务	151	7.3.4 主题层次体系	161
6.7.1 背景	151	7.3.5 语言学资源	162
6.7.2 问题与语义网技术的 作用	151	7.3.6 本体图书馆	162
6.8 其他场景	154	7.4 使用半自动方法	162
6.8.1 伦敦警察局的多媒体收藏 索引	154	7.5 On-To-Knowledge 语义网体系 结构	164
6.8.2 戴姆勒-克莱斯勒的在线 采购	155	7.5.1 知识获取	165
6.8.3 诺基亚的设备可共用性	155	7.5.2 知识存储	165
推荐读物	156	7.5.3 知识维护	166
第 7 章 本体工程	158	7.5.4 知识使用	166
7.1 引言	158	7.5.5 技术可共用性	166
7.2 手工构建本体	158	推荐读物	167
7.2.1 确定范围	158	设计	167
7.2.2 考虑复用	159	第 8 章 结论和展望	170
7.2.3 列举术语	159	8.1 集成	170
7.2.4 定义分类	159	8.2 一些技术问题	170
7.2.5 定义属性	159	8.2.1 网络本体语言: 少就 是多	170
7.2.6 定义侧面	160	8.2.2 规则和本体	170
7.2.7 定义实例	160	8.3 未来预测	171
7.2.8 检查异常	160	附录 OWL 抽象语法	173
7.3 复用现成本体	161	索引	180

第 1 章 语义网概览

1.1 万维网现状

万维网改变了人们彼此交流的方式和商业运作的方式。发达社会正在向知识经济和知识社会转型，而万维网处于这场革命的核心位置。

这种发展使得人们对计算机的看法也发生了变化。起初，计算机仅仅用做数值计算，而现在则主要用于信息处理，典型的应用包括数据库、文档处理和游戏等等。眼下，人们对计算机关注的焦点正在经历新的转变，将其视为信息高速公路的入口。

绝大部分现有的网络内容适合于人工处理。即使是从数据库自动生成的网络内容，通常也会丢弃原有的结构信息。目前万维网的典型应用方式是，人们在网上查找和使用信息、搜索和联系其他人、浏览网上商店的目录并且填表格订购商品等等。

现有软件工具没有很好地支持这些应用。除了建立文件间联系的链接之外，最有价值和必不可少的工具是搜索引擎。

基于关键词的搜索引擎，比如 Alta Vista、Yahoo、Google 等，是使用现有万维网的主要工具。毫无疑问，假如没有这些搜索引擎，万维网不会取得现在这么大的成功。然而，搜索引擎的使用也存在一些严重的问题：

1

- 高匹配、低精度。即使搜到了主要相关页面，但它们与同时搜到的 28 758 个低相关或不相关页面混在一起，检索的效果就很差。太多和太少一样令人不满意。
- 低匹配或无匹配。有时用户得不到任何搜索结果，或者漏掉了一些重要的相关页面。虽然对于现在的搜索引擎来说，这种情况发生的频率不高，但确实会出现。
- 检索结果对词汇高度敏感。使用最初填写的关键词往往不能得到想要的结果，因为相关的文档里使用了与检索关键词不一样的术语。这当然令人不满意，因为语义上相似的查询理应返回相似的结果。
- 检索结果是单一的网页。如果所需要的信息分布在不同的文档中，则用户必须给出多个查询来收集相关的页面，然后自己提取这些页面中的相关信息并组织成一个

整体。

有趣的是, 尽管搜索引擎技术在发展, 但主要的困难还是上述几条。技术的发展速度似乎落后于网上内容量的增长速度。

此外, 即使搜索是成功的, 用户仍必须自己浏览搜索到的文档, 从中提取所需的信息, 也就是说, 对极其耗时的信息检索本身, 搜索引擎并没有提供很多支持。因此, 用信息检索(information retrieval)来描述搜索引擎为用户提供的功能, 是不确切的; 用信息定位(location finder)可能更加合适。另外, 由于现有网络搜索的结果不易直接被其他软件进一步处理, 因此搜索引擎的应用往往是孤立的。

目前, 为网络用户提供更强大支持的主要障碍在于, 网上内容的含义不是机器可解读的(machine-accessible)。当然, 有一些工具能够检索文档、把它们分割成更小的部分、检查拼写并统计词频等等。可是, 一旦牵涉到解释句子含义和提取对用户有用的信息, 现有的软件就能力有限了。举一个简单的例子。对现有技术而言, 以下两个句子的含义是难以区分的:

我是一个计算机科学的教授。

2

你不妨认为, 我是一个计算机科学的教授。

使用文本处理技术, 怎样才能改善现在的状况呢? 一种解决方案是沿用现有表达方式来表示网上内容, 并发展基于人工智能和计算语言学的一些日趋复杂的技术以解决现有难题。这条途径迄今已探索多时, 尽管取得了一些进展, 但其任务仍然显得过于困难。

另一条途径是用一种更容易被机器处理^①的表示方法来描述网上内容, 并采用智能技术来利用这种表示方法所提供的便利。我们把这个革命性的方案称为语义网(Semantic Web)运动。值得注意的是, 语义网并不是一种和现有万维网平行的新的全球化信息高速公路, 相反, 它将在现有万维网的基础上逐渐成长。

语义网运动由——万维网联盟(W3C)发起, 其倡导者正是在20世纪80年代后期发明万维网的Tim Berners-Lee。他期望通过这场运动来实现他的万维网初始蓝图, 在这个蓝图中, 信息含义所起的作用比现今万维网中重要得多。

语义网的发展得到了产业界的大力支持和各国政府的大量资助。美国政府设立了DAML(DARPA Agent Markup Language)项目, 欧盟第六研究框架计划也将语义网作为关键内容之一。

① 文献中经常使用另一个词语“机器可理解”, 我们认为不妥, 因为会产生误解。智能代理不必真正“理解”信息, 只要有效地处理信息就够了, 尽管这种处理有时让人以为机器具有理解能力。

1.2 从当前万维网到语义网：实例分析

1.2.1 知识管理

知识管理所关注的是在一个组织中获取、处理和维护知识。由于大型商业机构把内部知识视为一种能够提高生产力、创造新的价值和增强竞争力的智力资源，知识管理已成为他们的一项关键性工作，对于地理分布广泛的国际组织来说尤其如此。

[3]

目前大多数可用信息只具有弱结构组织形式，如文本、音频和视频等。从知识管理的角度来说，现有技术存在以下诸方面的局限：

- 信息搜索。公司通常依靠基于关键词的搜索引擎，这方面的局限性上文已经做了简要阐述。
- 信息抽取。需要人工浏览文档，以便从中找出相关的信息。现有的智能代理(intelligent agent)还不能令人满意地完成这项任务。
- 信息维护。现在这方面仍然存在一些问题，比如术语的不相容性和无法移除过时信息等等。
- 信息挖掘。虽然可以用数据挖掘(data mining)等手段提取隐藏在公司数据库中的新知识，但对于分布式的、弱结构化的文档集合，这个任务仍然是困难的。
- 信息视图。经常需要限制某些雇员对某些信息的浏览权限。“视图”意味着隐藏某些信息，这个概念在数据库领域里为人们所熟知，但对企业内部网(或者万维网)尚难以实现。

语义网的目的是允许更高级的知识管理系统，使得：

- 可以根据含义用概念空间组织知识；
- 具备相容性检查和新知识提取功能的自动工具支持知识维护；
- 问答系统取代关键词搜索——用人性化的方式检索、提取和表示所需知识；
- 支持跨文档的问答功能；
- 能够规定用户察看信息的指定部分(甚至文档的某些部分)的权限。

[4]

1.2.2 企业对个人电子商务

企业对个人(B2C)电子商务是万维网用户最主要的商业活动。典型场景是用户访问一个或者多个网上商店，浏览他们的报价，选择并订购产品。

理想化的情况是，用户可以收集所有(至少是主要的)网上商店的信息，包括价格、条款、说明(比如有效期)等信息，从而选择最好的产品。但是如此规模的手工浏览太费时间，