

电子商务与智能服务丛书

服务资源智能聚合 与推荐技术

刘平峰 等著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

电子商务与智能服务丛书

服务资源智能聚合与 推荐技术

刘平峰 聂规划 陈冬林 傅 魁 姚梦迪 著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

针对传统的商品搜索和推荐技术已不能满足客户对服务资源个性化需求的问题,本书在国家科技支撑计划、863 计划、国家自然科学基金、教育部人文社科等课题的支持下,以用户复杂需求为出发点,充分利用本体技术在解决语义歧义和能够语义推理方面的优势,以“服务资源统一描述、发现与聚合——客户复杂需求描述和获取——满足客户复杂需求的服务资源推荐”为主线,研究服务资源语义模型和语义词典、基于语义模型的服务资源智能发现与聚合、客户复杂需求的本体建模与智能获取、满足客户复杂需求的服务资源均衡推荐与组合推荐四个层次的关键问题,为海量服务资源环境下电子商务智能个性化服务提供语义基础技术和应用关键技术。

本书可供从事电子商务研究、教学或学习的高、中级人员,研究生,以及相关专业的电子商务工作者使用,可帮助读者理解满足用户复杂需求的服务资源智能聚合与个性化推荐方法和技术。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

服务资源智能聚合与推荐技术 / 刘平峰等著. —北京: 电子工业出版社, 2015.10

(电子商务与智能服务丛书)

ISBN 978-7-121-27351-3

I. ①服… II. ①刘… III. ①电子商务—商业服务—研究 IV. ①F713.36

中国版本图书馆 CIP 数据核字(2015)第 234014 号

责任编辑: 徐蔷薇 特约编辑: 刘广钦

印 刷: 北京季蜂印刷有限公司

装 订: 北京季蜂印刷有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 720×1000 1/16 印张: 17.75 字数: 280 千字

版 次: 2015 年 10 月第 1 版

印 次: 2015 年 10 月第 1 次印刷

定 价: 49.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

前 言

现代电子商务正朝着深度融合的方向发展,电子商务运营平台聚集了海量各类服务资源,面临日益激烈的电子商务竞争,深度挖掘服务资源组合的价值,智能为用户提供个性化服务已成为电子商务发展的利器。但传统的电子商务个性化服务技术(用户需求智能获取技术和满足用户需求的推荐技术)已无法满足现代电子商务发展的要求。原因在于:①用户需求已由单个服务资源发展到服务资源组合(多粒度性),且对服务资源的需求表现为功能、属性和质量上多层次的均衡需求(多层次性),用户需求的语义歧义、多粒度性和多层次性造成了用户需求的高度复杂性,而传统用户需求表达方式单纯基于评分、特征或案例;②跨平台(多源)服务资源分类标准各异和语义歧义,导致服务资源聚合困难,限制了用户多源服务资源的选择能力;③服务资源组合推荐需要解决用户复杂需求的多粒度分解、多层次均衡和无语义歧义的精确匹配问题,传统推荐技术对此无能为力。以上问题若不能有效解决将导致企业对用户的需求出现误判,电子商务的优势即精准营销无法有效实现,电子商务效率和效能降低,电子商务服务的潜力受到限制。

因此,面向现代电子商务个性化服务发展的需要,在多源海量服务资源运营环境下,如何准确表示与获取用户个性化的复杂需求,克服多源标准不一和语义歧义的困难,智能发现和聚合服务资源,推荐满足用户复杂需求的服务资源及其组合成为亟待解决的问题。本体技术使得计算机可以分辨和识别语义信息,并对其自动进行解释、交换和处理,为电子商务服务资源智能聚合与精准推荐问题的解决提供了新思路。但已有通用本体技术不能直接解决领域语义问题,现有电子商务领域本体技术相对匮乏,不足以支撑面向用户复杂需求的智能服务应用。

本书在国家自然科学基金课题“基于知识网格的电子商务智能推荐系统”(70572079)、“多源电子目录语义集成与个性化服务理论研究”(70972094)、



“基于客户视角的 Web 服务组合优化与协同管理研究”(71072077), 国家 863 重点项目子课题“基于语义的信息在线获取和智能分析技术”(2009AA043508), 国家科技支撑计划项目子课题“电子商务与现代物流系统中语义集成技术研究”(2006BAH02A08)、“数字家庭服务资源智能挖掘与推荐技术研究”(2011BAH16B02), 教育部人文社科项目“面向多粒度知识服务的 Web 信息融合研究”(10YJAZH050) 等项目的资助下, 以用户复杂需求为出发点, 充分利用本体技术在解决语义歧义和能够语义推理方面的优势, 研究面向用户复杂需求的海量服务资源智能聚合与推荐技术, 为海量服务资源环境下电子商务智能个性化服务提供语义基础技术和应用关键技术。主要研究内容与成果包括:

(1) 研究了用户复杂需求的语义建模与获取技术, 解决了用户复杂需求的准确描述、规范表达与智能获取问题。电子商务相对于传统商务的绝对优势之一就是电子商务可以智能发现用户需求并主动为之推荐个性化服务, 能否准确获取用户需求是精准营销成败的关键。传统基于服务资源评分、特征(属性)或规则的用户需求表达方式过于简单, 未考虑用户需求固有的复杂特性: 一是语义歧义性, 用户对需求目标的习惯性表达可能产生语义歧义; 二是多粒度性, 用户可能需要单个服务资源, 也可能需要有序的服务资源组合; 三是多层次性, 用户需求表现为服务资源功能、属性和质量的多层次均衡需求。本项目提出的用户复杂需求语义建模和获取技术能够准确描述和获取用户复杂需求, 通过语义解析和规范化解了用户需求的语义歧义、多粒度性和多层次性问题, 从而为电子商务的精准营销提供更为可信的依据。

(2) 研究了基于元模型的服务资源智能发现与聚合技术, 实现多源电子商务环境下海量服务资源的统一集成化管理。电子商务具有开放性和规模性特征, 跨系统的多源异构产品分类标准和商务信息让用户眼花缭乱, 多企业、跨系统、分布式的电子商务服务资源存在概念众多、标准各异、集成度低的问题; 同时电子商务业务信息中存在大量一义多词和一词多义的情形, 导致不同主体对商务信息理解不一致。以上问题限制了海量服务资源环境下用户基于服务资源功能和属性快速检索所需服务资源的能力, 也限制了企业为用户提供整合服务资源的能力, 已成为阻碍电子商务开放发展的制约因素。研究电子商务

本体元模型,建立基于语义的服务资源目录标准,基于电子商务本体元模型智能发现网上的服务资源,通过服务资源本体映射与合并、服务资源本体实例消重,实现基于元模型标准的服务资源统一聚合技术,以基于本体的匹配技术保证海量环境下服务资源检索的高精度,以消除冗余和不一致性的服务资源本体库保证海量环境下服务资源检索的高效率,提升服务资源聚合能力和效率,促进了电子商务的开放性。

(3) 研究了基于语义的服务资源智能推荐技术和服务资源组合优化推荐技术,实现推荐结果与用户个性化复杂需求的高精度匹配,解决了传统推荐技术推荐精度不足,以及现有 Web 服务资源组合推荐技术以供应商全局服务质量(Quality of Service, QoS)最优为目标导致的对用户个性化支持不足的问题,以深度挖掘服务资源组合的价值。基于语义的服务资源智能推荐技术考虑用户在服务资源功能、属性和质量上的均衡需求,通过语义多层匹配向其推荐所需服务资源,大大提高了推荐精度;服务资源组合优化推荐技术则考虑用户需求的多粒度性,将复杂需求分解至元服务资源的需求,通过基于服务属性的多层关联规则向用户推荐流行的服务资源组合,通过模型优化向用户推荐个性化的服务资源组合,实现面向用户多粒度需求的服务资源组合精准推荐,解决了现有面向供应商全局 QoS 最优的 Web 服务资源组合推荐技术对用户个性化支持不足的问题。

(4) 研究了电子商务语义基础技术,解决了电子商务领域术语级、概念级和实例级三级语义计算问题,为电子商务服务智能化提供共性的语义技术支撑。本体内在的语义表达、映射与推理能力能够消除电子商务环境中一义多词和一词多义导致的语义歧义,使得计算机能自动地准确理解用户需求信息和业务信息,并进行智能推理,主动为用户提供个性化服务,但基于语义的映射与自动推理必须以本体和语义词典为基础,必须解决中文电子商务语义词典构建,以及电子商务语义本体的构建和管理问题。因此,需要研究中文环境下电子商务语义词典构建、中文本体映射与集成技术、中文本体管理技术,为电子商务应用提供术语级、概念级和实例级的语义描述与计算功能,为电子商务智能服务提供共性技术支撑。

本书由刘平峰副教授主笔,负责全书写作大纲的拟定并定稿。撰写工作分



工如下：第1章由姚梦迪博士生执笔，第2章由聂规划教授执笔，第3、6、7章由刘平峰副教授执笔，第4章由傅魁副教授执笔，第5章由陈冬林教授执笔。

在前期的研究过程和本书的撰写过程中，我们引用了国内外同行专家的相关研究成果，同时得到了有关院校的同行和相关部门的大力支持与帮助，特别是武汉理工大学电子商务专业的老师和研究生们，他们对本书的撰写提出了许多宝贵的意见与建议，在此一并致谢。

目 录

第 1 章 国内外相关理论研究现状	1
1.1 电子商务语义本体	1
1.1.1 电子商务语义词典	2
1.1.2 本体及本体集成	3
1.2 用户兴趣建模与获取	13
1.2.1 基于评分的用户模型	13
1.2.2 基于特征的用户模型	14
1.2.3 基于规则的用户模型	15
1.3 服务资源发现与聚合	16
1.3.1 服务资源发现	16
1.3.2 服务资源聚合	20
1.4 服务资源推荐技术	21
1.4.1 传统推荐算法研究	22
1.4.2 混合推荐技术	30
参考文献	32
第 2 章 电子商务语义基础技术	42
2.1 电子商务中文语义词典构建	42
2.1.1 中文语义词典系统体系结构	42
2.1.2 中文语义词典的词典结构	43
2.1.3 中文语义词典语义相似度计算	45
2.2 中文本体语义映射与集成技术	53
2.2.1 电子商务语义映射与集成系统框架模型	53
2.2.2 基于本体的语义查询流程	56
2.2.3 语义集成查询关键技术	59



2.3 中文本体管理技术	69
参考文献	72
第3章 用户需求语义建模与智能获取技术	74
3.1 用户需求语义建模	74
3.1.1 用户需求本体模型	74
3.1.2 用户概要本体模型	76
3.2 用户需求语义解析与规范化	79
3.3 用户偏好获取技术	81
3.3.1 基于综合反馈的用户偏好获取技术	82
3.3.2 基于最大熵模型的用户偏好获取技术	98
3.3.3 基于粗糙集的用户偏好获取技术	100
3.3.4 基于兴趣图谱的用户偏好获取技术	105
参考文献	123
第4章 服务资源智能发现方法	126
4.1 服务资源的语义建模	126
4.1.1 服务资源语义建模需求	126
4.1.2 面向用户的服务资源目录本体建模层次结构	129
4.1.3 面向用户的服务资源目录本体建模元语	130
4.1.4 服务资源目录本体元模型的 OWL DL 描述	132
4.2 服务资源目录本体自学习方法	133
4.2.1 基本原理	133
4.2.2 电子商务 Web 页面预处理	134
4.2.3 基于目录网站层次的目录本体概念获取	139
4.2.4 基于语义和关联规则的概念关系学习	142
4.2.5 基于模式匹配和在线统计的属性识别	146
4.3 服务资源实例发现方法	153
4.3.1 基本思想	153
4.3.2 Web 网页中个体知识表示特征	155

4.3.3 基于本体的网页主题概念和个体知识表示特征的判定	157
参考文献	167
第 5 章 服务资源智能聚合技术	171
5.1 服务资源本体实例消重机制	171
5.1.1 实例名称语义相似度计算	171
5.1.2 实例属性语义相似度计算	173
5.1.3 实例关系语义相似度计算	174
5.1.4 实例综合语义相似度计算	175
5.1.5 实例消重过程	175
5.1.6 实例消重算法举例	175
5.2 服务资源智能聚合系统关键技术	178
5.2.1 系统设计原理	178
5.2.2 服装行业本体建模	181
5.2.3 数据库模型设计	193
5.2.4 基于倒排模型的语义索引	196
5.2.5 语义信息检索	209
5.3 服务资源智能聚合原型系统实现	214
参考文献	216
第 6 章 服务资源智能推荐技术	217
6.1 基于语义的服务资源智能推荐技术	217
6.1.1 基于商品属性隐性评分的协同过滤算法	218
6.1.2 基于模糊聚类和资源平滑的协同过滤算法	222
6.1.3 基于本体的 B2B 电子商务 MAS 模型及商品匹配算法	224
6.1.4 基于本体的语义相似性的协同过滤算法	229
6.2 基于适合度的服务资源智能推荐技术	231
6.2.1 FQoS 适合度匹配算法	233
6.2.2 QoS 适合度匹配算法	240
6.3 基于兴趣度的服务资源智能推荐方法	244
参考文献	246



第 7 章 服务资源组合与推荐技术	251
7.1 基于服务属性多层关联规则的服务资源组合智能推荐技术	251
7.1.1 服务组合方法相关基础理论	251
7.1.2 基于服务属性多层关联规则的服务资源组合推荐基本 原理	254
7.1.3 基于服务属性多层关联规则的服务组合推荐案例研究	257
7.2 基于模型优化的服务资源组合智能推荐方法	259
7.2.1 组合服务的 QoS 计算相关基础理论	259
7.2.2 基于模型优化的服务资源组合智能推荐基本原理	262
7.2.3 基于模型优化的服务组合推荐案例研究	266
参考文献	268

第 1 章 国内外相关理论研究现状

1.1 电子商务语义本体

电子商务作为一种新型的商务模式在近二十年来得到了迅猛发展,用户或企业在网上即可搜索或购买到自己所需的信息或商品,企业间电子商务进行了广泛的集成以支持有效供应链管理。B2B 电子商务的标准经历了从 EDI 到 XML 的变化,现阶段较有影响的基于 XML 的电子商务标准包括非营利组织发起的 RosettaNet、由联合国贸易促进和电子商务中心 (UN/CEFACT) 和 OASIS 维护的 ebXML 标准,以及中科院电子商务中心制定的 cnXML 标准等,这些标准的共同特点是基于 XML 向企业提供一套模块化的规范,使得世界上任何一个地方的公司,无论其规模大小都能用标准的方法在互联网上交换商业信息,建立双方贸易关系,用通用的术语进行沟通。这些标准在克服 EDI 缺点,极大地推动企业电子商务应用的同时,也暴露出其存在的固有缺陷,它们只是提供了一个通用的架构,在此基础上行业可以定义自己的基于 XML 的电子商务应用交换文档来交换商务信息,企业间文档集成通过 XPath 与 XSL 结合来实现对 XML 文档的转换,这种直接映射的方法存在规则复杂、程序维护成本高和映射重用性差等不足。同时,对企业商务过程进行硬性建模不能适应企业商务流程变化的灵活性,缺乏信息配置和集成的柔性,不能适应电子商务的开放性和灵活性,并且 XML 本身是基于语构的,机器并不能理解其中的含义,从而限制了企业间事务处理的自动化。

语义本体技术的发展使得其在电子商务领域的应用成为一大研究热点,该技术能有效解决电子商务存在的挑战:数据源是异构的,描述词汇表不统一,如何从大量存在的知识中发现蕴含的知识。语义使得利用机器来自动处理大量的电子商务应用成为可能,其关键基础技术包括语义词典、本体工程、基于本体的企业信息与应用集成等方面。



1.1.1 电子商务语义词典

人与人之间沟通、交流所使用的语言，称为自然语言，相对于直接用二进制代码指令表达的计算机语言，自然语言是人类智慧的结晶，也是人工智能研究中最为困难的问题之一。

自然语言具有特有的语法、语义和语用知识等，自 20 世纪 60 年代美国的 ALPAC 报告提出之后，随着语义网络、语义领域和语义框架等语义理论的提出，人们对语义的研究也越来越多，但语义分析依旧是自然语言理解的难点之一。为了使计算机能够像人类一样理解自然语言中包含的信息，代替人类进行大规模的文本语义分析和理解，必须构建大规模的语义资源，计算机通过学习其中的知识理解自然语言中所包含信息的意义。

语义词典的任务是建立某种语言的词汇语义分类体系，其直接目标是为自然语义处理系统提供语义知识的支持。在语义词典当中应当设置语义处理机制，利用一定量的语义信息进行多义词的同义判别和歧义排除，以增强词汇语义理解的准确性。语义词典按照使用者不同分为两类，一类是供计算机使用的语义词典，如 WordNet 和 HowNet 等；另一类是供人用的语义词典，如我们常用的新华字典等。

我国著名机器翻译专家董振东先生创建的双语语义词典 HowNet 是一个以汉语和英语所代表的概念为描述对象、以揭示概念与概念之间，以及概念所具有的属性之间的关系为基本内容的常识知识库，智能化是其主要特点。语义词典和语义原语的分类树是 HowNet 的基础文件，每个词的概念和对其描述形成其中的一条记录。HowNet 的核心思想在于：世界上的一切事物都在特定的时间和空间内不停地运动和变化着，它们通常从一种状态变化到另一种状态，这种变化通常通过其属性的改变体现出来。HowNet 语义词典包含的 16 种概念语义关系如表 1-1 所示。

表 1-1 HowNet 语义词典的概念语义关系

编 号	概念语义关系	编 号	概念语义关系
1	上下位关系	9	物品-事件关系
2	同义关系	10	工具-事件关系



(续表)

编 号	概念语义关系	编 号	概念语义关系
3	反义关系	11	场所-事件关系
4	对义关系	12	时间-事件关系
5	部分-整体关系	13	值-属性关系
6	属性-宿主关系	14	实体-值关系
7	材料-成品关系	15	角色-事件关系
8	主体-事件关系	16	相关关系

电子商务语义词典是智能电子商务的语义基础, 现有以 WordNet 为代表的英文通用语义词典系统和以“现代汉语语义词典 (SKCC)”“知网 HowNet”为代表的中文通用语义词典系统, 收录的服务资源概念、属性和关系很少涉及电子商务专业术语的语义解析, 不适用于电子商务服务资源动态变化环境下智能服务的需要。综观国内外语义词典的相关研究成果, 国际上现有的通用语义词典 WordNet 虽然提供了大量术语间的语义关系, 并提供接口, 具有可操作性, 却无法支持中文环境下的语义计算; 国内现有的语义词典 HowNet 虽然提供了中文术语间的语义关系, 但它是通用词典, 没有提供服务资源的相关术语, 并且未提供接口, 可扩展性和可操作性受到限制, 无法适应中文环境下服务资源推荐语义的计算要求。

1.1.2 本体及本体集成

1. 本体的概念

本体这个术语来自哲学, 是研究存在的本质的问题, 但近年来本体被越来越多地应用到计算机界, 包括语义网、知识的表示等一系列应用, 并在人工智能、计算机语言及数据库理论中扮演着越来越重要的角色。

在人工智能界, 1991 年 Neches 等人最早给出 Ontology 在信息科学中的定义: “给出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这些词汇外延规则的定义”。1993 年, Gruber 给出了 Ontology 的一个最为广泛流传的定义: “explicit formal specifications of the terms in the domain and relations among them”, 即本体是对某一领域中术语及术语之间关系的规范而明确的说明。本体作为领域概念及概念之间关系的规范化描述, 这种描述是



规范的、明确的、形式化的，共享的。

(1) 概念模型 (Conceptualization): 指通过抽象客观世界中一些现象的相关概念而得到的模型。概念模型所表现的含义独立于具体的环境状态。

(2) 明确 (Explicit): 指所使用的概念及使用这些概念的约束都有明确的定义。

(3) 形式化 (Formal): 指本体是计算机可读的 (能被计算机处理)。

(4) 共享 (Share): 指本体中体现的是共同认可的知识，反映的是相关领域中公认的概念集，即本体针对的是团体而非个体的共识。

在实现上，本体论是概念化的详细说明，一个本体往往就是一个正式的词汇表，其核心作用就在于定义某一领域内专业词汇的语义及它们之间的关系。这一系列的基本概念如同一座大厦的基石，为交流各方提供了一个统一的认识。在这一系列概念的支持下，知识的搜索、积累和共享的效率将大大提高，真正意义上的知识重用和共享也成为可能。

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和词汇间相互关系的明确定义。总体而言，尽管本体定义有很多不同的方式，但从内容来看，本体所包含的一些基本要素，如概念、概念之间的特化/泛化关系，以及概念之间的其他关系等，已经被大多数人认可；从内涵来看，不同研究者对于本体的认识也是统一的，都把本体当作领域内部不同主体 (包括人、机器、软件系统等) 之间进行交流的一种语义基础，即由本体提供一种明确定义的共识。

本体是一份正式定义名词之间关系的文档或文件。一般 Web 上的 Ontology 包括分类和一套推理规则。分类用于定义对象的类别及其之间的关系；推理规则则提供进一步的功能，完成语义网的关键目标，即“机器可理解”。本体的最终目标是“精确地表示那些隐含 (或不明确的) 信息”。

基于上述本体定义，描述本体的基本建模元语包括如下 5 个。

(1) 类 (Classes) 或概念 (Concepts): 指任何事务，如工作描述、功能、行为、策略和推理过程。从语义上讲，它表示的是对象的集合，其定义一般采用框架 (Frame) 结构，包括概念的名称、与其他概念之间的关系的集合，以



及用自然语言对概念的描述。

(2) 关系 (Relations): 指领域中概念之间的交互作用, 形式上定义为 n 维笛卡儿积的子集: $R: C_1 \times C_2 \times \cdots \times C_n$, 如子类关系 (subclass-of)。在语义上关系对应于对象元组的集合。

(3) 函数 (Functions): 是一类特殊的关系。该关系的前 $n-1$ 个元素可以唯一决定第 n 个元素。形式化的定义为 $F: C_1 \times C_2 \times \cdots \times C_{n-1} \rightarrow C_n$ 。如 mother-of 就是一个函数, mother-of(x, y) 表示 y 是 x 的母亲。

(4) 公理 (Axioms): 代表永真断言, 如概念乙属于概念甲的范围。

(5) 实例 (Instances): 代表元素, 从语义上讲实例表示的就是对象。

从语义上讲, 概念间主要有 4 种基本关系, 如表 1-2 所示。在实际建模过程中, 概念之间的关系不限于表 1-2 中列出的 4 种基本关系, 可以根据领域的具体情况定义相应的关系。

表 1-2 本体概念之间的 4 种基本关系

关 系 名	关 系 描 述
kind of 或 is a	表达概念之间的继承关系, 类似于面向对象中的父类与子类之间的关系, 如家用计算机是计算机的一种
attribute-of	表达某个概念是另一个概念的属性, 如“价格”是桌子的一个属性
part-of	表达概念之间部分与整体的关系, 如 CPU 是计算机的组成部分
instance-of	表达概念的实例与概念之间的关系, 类似于面向对象中的对象和类之间的关系

2. 本体的描述

本体在 Web 上广泛运用需要一个重要的先决条件, 即需要一种描述本体并使得本体之间能够进行信息交换的语言标准。目前的主流 Web 本体语言有 RDF(S) 和 RDF, OWL 定义了机器可处理的数据模型, 即通过三元组作为基本的模型原语, 用 XML 语法表示, 由 3 种对象类型组成: URL 命名的资源、用于描述资源某一特性或关系的属性, 以及一个命名属性加上某资源的该属性下的取值构成一个 RDF 语句。

RDF Schema 是描述 RDF 资源属性和类的词汇表, 规范了机制以约束类及关系的组合和这些约束冲突的检测。RDF(S) 仅仅是简单、轻量的本体语言,



其对知识/语义的描述能力不足。因此,在 RDF(S)基础上,W3C 于 2004 年推出了标准 Web 本体语言 (Web Ontology Language, OWL), OWL 是 RDF(S) 的扩展,通过提供更多具有形式化语义的词汇,具有比 RDF(S)更强的 Web 内容表达能力和语义推理能力。

W3C 总结了几种本体语言的开发经验,于 2004 年 2 月正式推出 OWL, OWL 是语义网发展过程的里程碑,它经过了广泛的讨论,并得到了一致的认可。OWL 是为了语义 Web 而设计的,它在 RDF 的基础上增加了描述类和属性的词汇。OWL 在表示和推理方面比 XML、RDF 等强,因此,基于 OWL 的知识表示成为研究和应用的热点。

OWL 是语义网活动的一个组成部分,这项工作的目的是通过对增加关于那些描述或提供网络内容的资源的信息,从而使网络资源能够更容易地被那些自动进程访问。由于语义网络固有的分布性,OWL 必须允许信息能够从分布的信息源收集起来。其中,允许本体间相互联系,包括明确导入其他本体的信息,能够部分实现这样的功能。

另外,OWL 提出了一个开放世界的假设。也就是说,对资源的描述并不局限于一个简单的文件或范围内。例如,类 C_1 本来是由本体 O_1 定义出来的,然而,它也可以是由其他的本体扩展出来的。对 C_1 进行这样的假设的结果是单调的。新的信息不能否定之前的信息。新的信息可以是和旧的信息矛盾的,但是事实和推导只能被增加而不能被删减。当设计一个本体的时候,设计者必须考虑到这种矛盾的可能性。幸运的是,本体设计工具可以帮助我们探测到这样的情况。

为了能写出一个能被唯一翻译的而且能被软件(代理)使用的本体,我们要求 OWL 有一个语法和正规的语义。OWL 是 RDF 的一个词汇扩充。在 OWL 网络本体语言语义和简明语法中,有 OWL 的语义定义。OWL 提供了 3 种表达能力递增的子语言,以分别用于特定的实现者和用户团体。

(1) OWL Lite: 用于提供给那些只需要一个分类层次和简单约束的用户。例如,虽然 OWL Lite 支持基数限制,但只允许基数为 0 或 1。提供支持 OWL Lite 的工具应该比支持表达能力更强的其他 OWL 语言更简单,并且从辞典 (thesauri) 和分类系统 (taxonomy) 转换到 OWL Lite 更为迅速。