语义网技术体系

瞿裕忠　胡 伟　程 龚　编著

科学出版社

语义网技术体系

瞿裕忠　胡　伟　程　龚　编著

科学出版社

北　京

内 容 简 介

本书简要回顾万维网的发展历史及体系结构，系统介绍语义网的基本理念及技术体系，包括RDF数据、Web本体、语义网推理技术和RDF数据查询技术等方面的基本概念和前沿研究；详细阐述语义网应用中的基础技术，包括语义网搜索技术、语义网本体匹配技术和语义网浏览技术，并介绍作者在语义网搜索、语义网本体匹配和语义网浏览方面的研究成果。

本书适合于语义网及相关领域的研究人员、语义网应用开发者以及想要深入了解语义网技术体系的读者。本书也可作为信息技术类学科的研究生与高年级本科生相关课程的参考用书。

图书在版编目(CIP)数据

语义网技术体系/瞿裕忠，胡伟，程龚编著. —北京：科学出版社，2015

ISBN 978-7-03-042213-2

Ⅰ.①语… ②瞿… ③程… Ⅲ.语义网络-技术体系-研究
Ⅳ.①TP18

中国版本图书馆CIP数据核字(2014)第243219号

责任编辑：陈岭啸 惠 雪/责任校对：胡小洁
责任印制：赵 博/封面设计：许 瑞

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

三河市骏杰印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2015年10月第 一 版 开本：720×1000 1/16
2015年10月第一次印刷 印张：14 1/4
字数：283 500

定价：59.00元

（如有印装质量问题，我社负责调换）

前　言

自 1990 年创建以来，万维网已然成为人类有史以来最庞大的信息系统，并改变着人类社会的诸多方面。万维网空前成功的背后应该有诸多因素，我们认为系统的开放性、易用性和易扩展性是其中的重要因素。然而，万维网最初追求的是一个互相链接的超文本文件系统，这些文件可以通过“浏览器”来查看。这意味着，人们可以方便地浏览和理解万维网上的信息。随着万维网的逐步成熟，人们希望机器能够理解和集成万维网上的数据，从而能更好地为人类服务。为此，在万维网创始人蒂姆·伯纳斯-李(Tim Berners-Lee)的倡导下，万维网联盟(W3C)于 2001 年建立语义网行动计划。

语义网是 W3C 进一步发展万维网的一个愿景，它提供了一个公共框架，使得数据的共享和复用可以跨越应用系统、企业和社区的边界。W3C 提出了资源描述框架(RDF)，并鼓励人们采用那些基于 RDF 数据模型的数据格式来构建或发布数据，这样机器或者计算机程序能够按照 RDF 数据模型来理解和集成有关数据。在构建或生成 RDF 数据时，通常需要使用某个领域的词汇表，即一组类和属性。W3C 并没有规定数据提供者使用何种词汇表来描述资源或事物，而是希望人们按照 RDFS 或 OWL 等本体语言来定义或描述他们的词汇表，这样机器能够按照本体语言的语义模型来理解这些词汇表，从而可以更好地理解采用这些词汇表的 RDF 数据。当然，对于采用相同词汇表的多个 RDF 数据源，其集成会更加容易。

经过十年多的努力，语义网的理论基础已经奠定，W3C 有关语义网的技术规范也逐步得到了完善。同时，链接数据(linked data)指导原则已经成为在万维网上发布 RDF 数据的基本准则，一个基于 RDF 数据模型的数据之网(Web of data)正在快速成长。特别地，DBpedia 是一个从维基百科中提取的 RDF 数据集，它已经成为数据之网的枢纽，越来越多的语义网数据链接到它，覆盖领域包括地理信息、公司、人员、电影、音乐、基因、药品、书籍和科学出版物等。以 DBpedia 为核心的链接开放数据(LOD)推动了数据之网的快速增长，也拉开了语义网应用的序幕。事实上，信息领域的业界已经开始倡导知识库的构建及其在搜索引擎中的应用，比如 Google 正在大力推行“知识图谱”。可以预见，基于语义网技术的数据共享和利用将成为语义网技术及应用的发展方向。

本书作者长期从事语义网领域的研究工作，希望通过本书系统地介绍语义网的基本理念及技术体系，详细阐述语义网应用中的基础技术，并介绍作者在语义

网搜索、本体匹配和语义网浏览等方面的研究成果。本书第 2、3、7 章由胡伟副教授负责撰写，第 5、6 章由程龚副教授撰写，其余各章由瞿裕忠教授主笔。作者希望这本书能够成为语义网及相关领域研发人员的基础读物或参考书，也能够成为信息技术类研究生或高年级本科生相关课程的参考用书。因作者能力和写作时间所限，书中内容难免有不足之处，恳请读者指正。关于本书的最新情况，请访问：http://ws.nju.edu.cn/swbook。

本书的出版得到了国家自然科学基金专项基金项目“面向大数据的媒体内容分析与关联语义挖掘研究”(项目编号：61223003)和国家自然科学基金面上项目“语义 Web 浏览方法与技术的研究”(项目编号：61170068)的资助，也得到了科学出版社的大力支持，在此表示感谢！本书中介绍的部分研究成果得益于研究组中博士生和硕士生的辛勤工作，在此一并致谢！

作　者

2015 年 6 月

南京大学

目　录

第1章 绪 论

本章首先回顾万维网的发展历史及体系结构，简述万维网应用开发技术；接着，概述语义网及其技术体系，并简要介绍语义网应用现状；最后说明本书内容框架是如何组织的。

1.1 万维网简介

1990年，蒂姆·伯纳斯-李(Tim Berners-Lee)在欧洲粒子物理实验室(CERN)成功地实现了万维网(World Wide Web，WWW或Web)的一个原型系统。按照蒂姆的最初设想，万维网是一个包含互相链接的超文本文件的系统，这些文件可以通过互联网(Internet)访问。1993年，美国国家超级计算应用中心(NCSA)发布了一个称为“Mosaic”的万维网浏览器，它是第一个能够在同一窗口中显示图像和文本的浏览器。同年，CERN宣布万维网技术可以被任何人自由地使用，无需付费。1994年，万维网联盟(World Wide Web Consortium，W3C)宣告成立，它致力于开发高品质的技术标准，以引领万维网充分发挥其潜力。从此，万维网很快发展成为人类有史以来最庞大的信息系统，并改变着人类社会的方方面面。

在万维网出现之前，历史上最著名的超文本/超媒体系统为道格拉斯·恩格尔巴特(Douglas Engelbart)于1968年演示的在线系统(oN-Line System，NLS)。该系统首次向公众展现了计算机鼠标、超媒体和屏幕上的视频会议等多项崭新技术。而超文本(hypertext)和超媒体(hypermedia)这两个技术名词是由西奥多·纳尔逊(Theodor Nelson)于1963年定义的。在20世纪60年代，随着计算机文字处理系统的发展，诞生了多种标记语言，包括国际商业机器公司(IBM)的通用标记语言(generalized markup language，GML)，它的基本思想是把文档的内容结构与样式分开，推崇描述型标记，提倡标记的严格性和使用的灵活性。标准的通用标记语言(standard generalized markup language，SGML)正是以GML为蓝本制定的，于1986年成为国际标准化组织的一个标准(ISO 8879:1986)。也正是在20世纪80年代，以TCP/IP(传输控制协议/互联网协议)为基础的互联网逐步形成并快速发展起来。

在发明万维网时，蒂姆·伯纳斯-李的出发点是将超文本嫁接到互联网上，并提出一个互相链接的超文本文件系统的设想，这些文件可以通过“浏览器”来查看，为此，他设计了超文本标记语言(hypertext markup language，HTML)用来

书写万维网中的文件。事实上，HTML 是一个基于 SGML 的标记语言，可提供有限种标记，支持超链接并注重文本的呈现效果。为了统一地标识万维网中的文件，蒂姆又提出了通用文件标识符(universal document identifier，UDI)。后来，UDI 演变为统一资源定位符(uniform resource locator，URL)，逐渐地 URL 被统一资源标识符(uniform resource identifier，URI)所替代。URI 可以用来标识任何需要标识的资源。由于 URI 被限制在 ASCII 字符集的一个子集，因此国际化资源标识符(internationalized resource identifier，IRI)被提出来。IRI 可以容纳通用字符集(ISO/IEC 10646)，其中包括汉字、韩文和斯拉夫字母等。目前，绝大部分的 IRI 仍然是 URI，因此本书是以 URI 替代 IRI 来阐述相关技术。

为了让互联网用户能够按照一种规范的方式访问万维网中的文件，蒂姆及其合作者设计了超文本传输协议(hypertext transfer protocol，HTTP)。HTTP 是一种建立在 TCP 之上的应用层协议，是一种请求/响应式的协议。通常，一个 Web 站点(简称网站)是指互联网上的某个计算机系统，实现了 HTTP 服务端接口，并提供诸多网页，包括 HTML 文件以及其他可访问的文件，供互联网用户使用 Web 浏览器(简称浏览器)来查看。而一个浏览器不仅要实现 HTTP 客户端接口，而且能够对接收到的网页按照其格式呈现出来。至 1990 年年底，蒂姆·伯纳斯-李在 CERN 实现了世界上第一个浏览器(称为“WorldWideWeb”，后来改名为“Nextus”)和第一个 HTTP 服务器软件(称为“CERN httpd”)，并开发了世界上第一个网站(网址：http://info.cern.ch)，这也标志着万维网的诞生。

作为互联网上的一个分散式信息系统，万维网具有跨平台和开放等特性，并具备优异的易用性和易扩展性，这使得万维网很快发展成为人类有史以来最庞大的超媒体信息系统。作为互联网上最具影响的一种应用，万维网也反过来拉动了互联网的发展壮大。紧接着，万维网的商业化应用掀起了互联网经济的浪潮。

关于万维网的原创设计及终极命运，建议读者阅读文献(Berners-Lee & Fischetti，2000)。该文献能够帮助大家理解万维网的本质，充分利用万维网发挥效用。在该文献中，万维网创始人蒂姆·伯纳斯-李不仅指出了在万维网上找到商业和社会力量之间理想平衡的需要，而且还对万维网当前状态提出了一些批评意见。最后，蒂姆·伯纳斯-李就万维网的未来给出了他自己的计划，并呼吁程序员、计算机制造商以及社会组织积极支持和参与，使其成为现实。

1.2 Web 应用开发技术

根据万维网体系结构(Jacobs & Walsh，2004)，万维网的基础技术包括 URI、HTML 和 HTTP。其中，HTTP 规范了浏览器和 Web 服务器之间的交互行为，这样万维网用户只需一个浏览器(比如 Mozilla Firefox)就能自如地浏览有

关网站的内容，而网站建设者在安装某个 Web 服务器(比如 Apache HTTP server)之后就可以轻松地部署一个站点，包括配置有关文件的 URI(俗称“网址”)，当然，网页的制作通常会借助某个 HTML 文档写作工具来完成。

早期，万维网上的内容大部分是静态的网页。为了生成能够反映用户输入的动态网页，通用网关接口(CGI)技术就诞生了。它是 Web 服务器与外部应用之间的交互接口，也打开了 Web 应用的大门。Web 应用通常是指在计算机网络上可以使用浏览器访问到的应用系统，通常会使用 HTTP、HTML 和 URI 等基本的 Web 技术以及数据库管理技术。后来，以 Java Servlet 为代表的 Web 应用服务器技术逐渐兴起。与此同时，以 JavaScript 脚本语言为代表的客户端技术为 Web 应用的用户界面和人机交互带来了动态性。特别地，Ajax(asynchronous JavaScript and XML)技术能够进一步增强 Web 应用的互动能力。

关于可扩展标记语言(extensible markup language，XML)，它可以粗略地看作 SGML(标准通用标记语言)的一个子集。SGML 的基本思想是把文档的内容结构与样式分开，推崇描述型标记，提倡标记的严格性和使用的灵活性。需要指出的是，SGML 是一个元语言，可以用来定义特定的标记语言。事实上，HTML 就是一个用 SGML 定义的标记语言。作为一个特定的标记语言，HTML 只提供有限种标记，且注重于文本的呈现效果，难以满足万维网上数据表示和交换的发展要求。XML 比 SGML 简洁很多，并继承了 SGML 的大部分优点，也是一个元语言，可以用来定义应用领域中所需的标记。目前，XML 已成为万维网上数据表示和交换的一个重要语言。

伴随着 Web 应用的发展，Web 应用之间的互操作问题逐渐显现。为解决这一问题，Web 服务(Web service)及相关技术被提出。Web 服务是这样的一个软件系统，它有一个用 WSDL(Web services description language)描述的接口，以便在互联网上被其他程序或者 Web 服务所调用。通常，一个 Web 应用可以被包装成一个 Web 服务，而服务的调用大多采用基于 HTTP 之上的 XML 消息格式。随着 Web 服务逐渐增多，Web 服务的发现及组合技术逐渐受到关注。

万维网的快速发展带来了信息爆炸，如何帮助用户快速地找到他们所需的信息则成为一个重要问题。很自然地，众多 Web 搜索引擎相继出现，而基于关键词的全文检索很快成为一种常规的搜索模式。Web 搜索引擎的基础技术主要包括页面获取、索引和排序等，在这些基本技术日益成熟的同时，Web 搜索领域的新技术层出不穷，比如 Web 广告模型、查询扩展和语义搜索等。另外，Web 上大量的数据来自背后的关系数据库，这就引发了对于深网(deep Web)数据的探究，从而也推动了包括 Web 信息抽取和数据挖掘在内的 Web 数据管理技术的深入发展。

1.3 语义网简介

经典的万维网是一个互相链接的超媒体文件系统，这些文件(文本、图像或视频)是供人们直接浏览的，而计算机却难以理解这些文件中的内容，从而就难以复用和集成万维网中的数据来提供更有用的信息服务。为此，W3C 于 2001 年开始建立语义网(Semantic Web)行动计划(http://www.w3.org/2001/sw/)，共同开发一套技术规范，使得符合语义网技术规范的数据容易地被计算机所理解，让不同的应用之间能够更方便地共享和复用彼此的数据。也就是说，语义网是 W3C 进一步发展万维网的一个愿景，它提供这样的一个公共框架，使得数据的共享和复用可以跨越应用系统、企业和社区的边界；而在传统万维网上只有文档的交换和共享。图 1-1 是 W3C 给出的语义网技术栈。

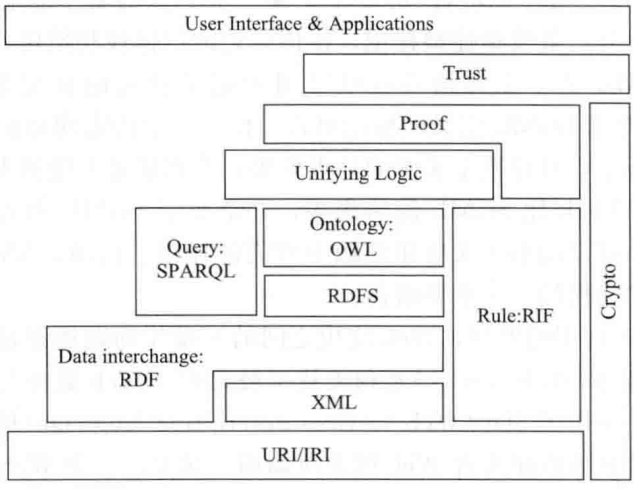

图 1-1 语义网技术栈

(图片来源：http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/)

语义网以资源描述框架(resource description framework，RDF)作为基石。RDF 是一个公共的数据模型，它以 RDF 三元组(RDF triple)作为基本的数据单元来描述资源的类型和属性，而一个 RDF 三元组由主语(subject)、谓语(predicate)和宾语(object)3 部分组成。其中，URI(统一资源标识符)可以出现在三元组中的任何位置，而空白节点(blank node)不能作为谓语，字面量(literal)只能作为宾语出现。URI 用来标识任何需要标识的资源，包括信息资源(比如一个网页)、现实世界中的事物(比如一本书)或者人们在社会实践中形成的概念(比如书和作者)等；空白节点只能作为局部的资源标识，不具备 URI 的全局标识能力；

字面量通常用来表示基本类型的数据，如字符串、整数和实数等。

本体(ontology)在语义网中扮演着重要的角色。在哲学领域，本体论主要探讨事物的基本特征及其分类体系。在人工智能及信息技术领域，本体论的概念被用在知识表示上，按照 Gruber(1993)的定义，一个本体是一个共享概念模型的显式的形式化规约。在语义网中，RDF 数据中使用到的类型和属性也需要给出一个明确的形式化规约，只有这样，应用程序才能理解数据的含义。通常，某个应用领域中一组相关的类和属性(统称术语)称为一个词汇表(vocabulary)。W3C 在语义网技术体系中采用本体来规约词汇表。作为一个规约，本体需要通过某种语言表达。为此，W3C 开发了 RDF 词汇描述语言 RDF Schema 和 Web 本体语言 OWL。这样，各个组织或机构可以使用 RDF Schema 或 OWL 表示各自领域的本体，并发布在万维网上以共享。领域本体中的类型和属性可以用来描述相应领域中的事物及其联系，形成 RDF 数据。基于共享本体的 RDF 数据不仅能够实现语义的共享，而且使推理成为可能。事实上，RDF Schema 和 OWL 均定义了若干推理规则。例如，如果已知一个对象的类是“人”，又已知“人”是“动物”的一个子类，那么通过推理规则可以得知该对象也是一个“动物”，尽管这个事实可能在原始的数据中并未出现。运用推理技术使得信息提供者不必对所有信息全部罗列出来，应用程序可以根据现有的数据和推理规则自动派生出蕴含的信息，这对检测数据的一致性、保证查询的完整性和提升搜索的召回率等具有重要意义。关于语义网中的本体论，建议读者阅读文献(Horrocks，2008)。

正如关系型数据库(relational database，RDB)需要标准化的结构化查询语言(structured query language，SQL)来规范数据查询的表达，RDF 数据也需要相应的数据存取规范，称作 SPARQL 协议与 RDF 查询语言(SPARQL protocol and RDF query language，SPARQL)。SPARQL 查询语言提供了描述跨越数据源的 RDF 数据查询的能力。SPARQL 查询的基本组成单元是三元组模式(triple pattern)，三元组模式与 RDF 三元组类似，区别在于其主语、谓语、宾语位置均可设置为变量。三元组模式通过合取、析取等方式组合成为图模式(graph pattern)，能够匹配到图模式的 RDF 数据便作为查询结果返回。作为 SPARQL 查询语言的补充，SPARQL 协议描述了一种包装 SPARQL 查询服务的通用方式，并提供了一个基于 HTTP 和简单对象访问协议(simple object access protocol，SOAP)的实现。

2004 年，RDF 模型论语义以及 OWL 的语义和抽象语法等 12 个技术规范正式发布；2008 年，W3C 发布了 RDF 数据存取的技术规范 SPARQL。这标志着语义网的数据模型、本体语言和数据存取的技术基础已经奠定。这些技术帮助我们在万维网上编织可复用的数据，并使得这些数据能够共享、复用和集成。需要指出的是，基于 HTTP 的 URI 具有网络可访问性，RDF 数据具有易集成性，使

用共享本体的 RDF 数据有利于计算机来理解、复用和集成这些数据。有关语义网的进一步认识与思考，建议读者阅读文献(Shadbolt, et al., 2006)。

随着语义网技术规范的日趋完善和研究的不断深入，随之而来的是语义网上 RDF 数据量的快速增长。特别地，蒂姆·伯纳斯-李倡导的链接数据(linked data)理念以及相应的链接开放数据(linking open data, LOD)项目赢得了包括欧洲和美国政府的广泛支持，已经汇集了数百亿的 RDF 三元组，这些数据涉及地理信息、社交网络、政府数据和生物医学等众多领域(Bizer, et al., 2009a)。语义网搜索系统 Falcons(Cheng & Qu, 2009)就已经采集到了数十亿条 RDF 三元组。可以预见，语义网技术正在带来一个巨大的"数据之网"(Web of data)，这必将带来语义网应用的繁荣。

1.4 语义网应用

众所周知，维基百科(Wikipedia)是人类合作编辑的一个自由的百科全书，也是人类分享知识的平台，然而，由于其知识一般采用文本表示，使得机器难以理解，不利于人们深入分享和利用这些知识。为此，欧洲的 DBpedia 项目(Bizer, et al., 2009b)从维基百科中提取结构化信息，转化为 RDF 数据，并遵循链接数据指导原则发布到万维网上。这个 RDF 数据集称为 DBpedia。目前，DBpedia 已经成为数据之网的枢纽，越来越多的语义网数据链接到它，覆盖领域包括地理信息、公司、人员、电影、音乐、基因、药品、书籍和科学出版物等。随着语义网技术的深入发展，DBpedia 有了许多创新而有趣的应用。例如，地图集成、分面式搜索和关系查询等。另外，对于协作的知识工程，也可以直接采用语义网技术，在文本编辑时进行相应的标注来表示该文本所关联对象的属性。事实上，OntoWiki(Auer, et al., 2006)在内容编写时提供了丰富的视图，比如列表、地图、日历等，能够根据背后的 RDF 知识库动态地提供某个已有对象可用的属性及其取值，方便用户采用 RDF 三元组的方式进行编辑。这些 RDF 数据可以用来进行查询，比如根据属性值进行分面式浏览。

以 DBpedia 为核心的 LOD 项目推动了链接数据的快速增长，也拉开了语义网应用的序幕。一般来说，语义网应用系统是指采用语义网技术的 Web 应用系统。正如前述，语义网技术体系是以 RDF 为基石的，因此，使用 RDF 数据的 Web 应用系统均可看做是语义网应用系统，有些应用还能够产生 RDF 数据。与一般的 Web 应用相比，语义网应用通常具备数据集成与推理的能力，从而体现出某种程度的智能化。

目前，语义网技术已经在诸多领域得到广泛应用。在数字图书馆及文化遗产领域，以芬兰和荷兰为代表的欧洲诸国率先开展语义网应用的研究与开发，有多

个应用系统取得了显著成效；在传媒领域，英国广播公司(BBC)和纽约时代公司(New York Times)较早地采用语义网技术来改进公司内部的信息集成，并提升信息服务质量；在航空航天领域，美国航空航天局(NASA)开发了多个语义网应用系统来应对智能化信息集成与检索的挑战，美国波音公司也在其信息系统中逐步采用语义网技术；在众多的社交网络服务系统中，有许多社交网站已经采用语义网技术或类似技术，以帮助用户便捷地编辑和发布各自的结构化数据，从而能够方便地集成多方的数据，为用户提供更好的服务。

值得注意的是，主流的搜索引擎正在逐步采用语义网技术或类似技术来增强用户体验。Google Rich Snippets是一种网页内容摘要技术，通过提取网页中嵌入的与查询相关的微数据(Microdata)、微格式(Microformats)和RDFa，并将它们展现在最终的摘要中，从而帮助用户了解网页的主要内容。Google Rich Snippets要求网站管理员(即发布者)对网页中的结构化数据进行语义标记，而不影响网页原有的呈现方式。然而，这些添加的标准格式的结构化数据能够被机器理解，从而能够更智能化地处理网页内容。类似于Google Rich Snippets，SearchMonkey是一个开放式的搜索平台，用来增强搜索结果的摘要。Yahoo! 提供了SearchMonkey开发工具帮助用户找到已有的数据服务，创建自己的数据服务，确定在搜索结果中如何显示额外的数据。与Google Rich Snippets相比，SearchMonkey则具有更大的灵活性。

总的来说，语义网技术可以用于从文本和多媒体数据中抽取结构化数据，形成RDF数据，或者将关系型数据转换成RDF数据，或者直接使用RDF数据，并结合本体的使用，从而赋予数据良好定义的、机器易理解的含义，使得信息集成在语法上和语义上都能够畅通进行，以提供更好的信息服务。从这个意义上来讲，对于具有信息集成和搜索等应用需求的领域，只要信息本身适合于表达成结构化数据，语义网技术就可以发挥其作用。

1.5 本书组织

本书首先介绍语义网基础技术体系，包括RDF数据、Web本体、语义网推理技术和RDF数据查询技术等；然后阐述语义网应用中的基础技术，包括语义网搜索技术、语义网本体匹配技术和语义网浏览技术等，并介绍了作者在语义网搜索、本体匹配和浏览方面的研究成果；最后对于语义网技术及应用给出总结与展望。以下是本书其余各章的概要。

第2章“RDF数据”介绍RDF数据模型及几种常见语法，阐述链接数据指导原则、链接开放数据项目以及链接数据平台，采用复杂网络分析方法研究语义网中对象的链接结构，并给出有关语义网复杂网络特性的实验结果。

第 3 章“Web 本体”介绍 RDF 模式(RDF schema，RDFS)和 OWL 等 Web 本体语言，阐述本体构建的方法学，并简要地介绍本体构建和本体维护等技术。

第 4 章“语义网推理技术”介绍 RDFS 推理和基本的描述逻辑 ALC，回顾 OWL 1 DL 的特点及其不足，详细介绍 OWL 2 DL 有关的描述逻辑及其推理方面的基本结论，介绍本体与规则的集成以及规则交互格式 RIF。

第 5 章“RDF 数据查询技术”回顾 SPARQL 查询语言的核心内容，详细介绍并比较多种 RDF 数据存储技术，讨论 SPARQL 查询处理技术，简要介绍基于规则的查询应答技术。

第 6 章“语义网搜索技术”详细阐述实体搜索技术和关联搜索技术，讨论 SPARQL 查询的生成技术，简要介绍本体搜索技术。

第 7 章“语义网本体匹配技术”介绍概念层本体匹配的定义、流程、方法和系统，并宏观分析了本体间的可匹配性；介绍语义网对象共指消解的定义、方法和系统；介绍本体与关系数据库间匹配的定义、方法和系统，并介绍 R2RML(RDB to RDF mapping language，关系型数据库到 RDF 的映射语言)和 Direct Mapping 语言。

第 8 章“语义网浏览技术”回顾典型的语义网浏览器，讨论语义网浏览器应具备的基本功能，阐述语义网浏览系统 SView 及其设计思想，并指出语义网浏览技术的发展方向。

第 9 章“总结与展望”总结语义网技术体系，展望语义网技术及应用的发展前景。

本书各章中引用的参考文献紧接在相应篇章的后面，而书中常用缩略词及中文译名则在附录给出。

参 考 文 献

Auer S, Dietzold S, Riechert T. 2006. OntoWiki—a Tool for Social, Semantic Collaboration [C]//Proceedings of the 5th International Semantic Web Conference. Berlin: Springer, 736-749.

Berners-Lee T, Fischetti M. 2000. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor[M]. New York: HarperCollins.

Bizer C, Heath T, Berners-Lee T. 2009a. Linked Data—the Story So Far[J]. International Journal on Semantic Web and Information Systems, 5(3): 1-22.

Bizer C, Lehmann J, Kobilarov G, et al. 2009b. DBpedia—a Crystallization Point for the Web of Data[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 7(3): 154-165.

Cheng G, Qu Y. 2009. Searching Linked Objects with Falcons: Approach, Implementation and Evaluation[J]. International Journal on Semantic Web and Information Systems, 5(3):

49-70.

Gruber T R. 1993. A Translation Approach to Portable Ontology Specifications[J]. Knowledge Acquisition, 5(2): 199-220.

Horrocks I. 2008. Ontologies and the Semantic Web[J]. Communications of the ACM, 51(12): 58-67.

Jacobs I, Walsh N. 2004. Architecture of the World Wide Web, Volume One[S/OL]. W3C Recommendation. http://www.w3.org/TR/webarch/ [2004-12-15].

Shadbolt N, Hall W, Berners-Lee T. 2006. The Semantic Web Revisited[J]. Intelligent Systems, IEEE, 21(3): 96-101.

第2章 RDF 数 据

本章首先介绍资源描述框架的基本概念，并给出多种常见语法；随后，阐述链接数据指导原则、链接开放数据项目以及链接数据平台；最后介绍语义网的对象链接模型，并给出有关语义网复杂网络特性的实验结果，以便人们了解语义网上对象链接的宏观结构。

2.1 RDF 数据模型

作为语义网技术体系的基石，资源描述框架(resource description framework，RDF)是 W3C 提倡的一个数据模型，用来描述万维网上的资源及其相互间的联系。

RDF 技术规范的第 1 版是由 W3C 于 1999 年发布的，主要面向的是元数据(metadata)。在本书撰写之际，W3C 发布了 RDF 1.1 技术规范(Brickley & Guha，2014)。

RDF 数据模型的核心包括资源(resource)、属性(property)、字面量(literal)以及 RDF 陈述(RDF statement)等，分别在以下各小节中介绍。

2.1.1 资源

资源可以是任何想要描述的事物，可以是具体的，也可以是抽象的。例如，资源可以是“张三”、“教授”、“计算机”、“数学”等。

在 RDF 中，每个资源通常拥有一个统一资源标识符(uniform resource identifier，URI)(Berners-Lee，et al.，2005)来标识它。URI 是一个用来标识资源的字符串，它是万维网体系结构的重要组成部分。较为常用的 URI 是用来标识万维网上域名地址的统一资源定位符(uniform resource locator，URL)，而另一个不常用的是统一资源名称(uniform resource name，URN)，比如一本书的ISBN。

URI 语法由 URI 协议名(例如“http”、“ftp”、“mailto”或“file”)、冒号，以及协议对应的内容构成。特定的协议定义了协议内容的语法和语义，而所有的协议都必须遵循一定的 URI 语法通用规则，也就是为某些专门目的保留部分特殊字符。URI 语法同时也就各种原因对协议内容加以其他的限制。例如，保证各种分层协议之间的协同性。百分号编码也为 URI 提供附加信息。尽管 URI 不一