

关联数据

万维网上的结构化数据

Linked Data
Structured data on the Web

大卫·伍德 (David Wood)

玛莎·扎伊德曼 (Marsha Zaidman)

卢克·鲁思 (Luke Ruth)

迈克尔·豪森布拉斯 (Michael Hausenblas)

著

蒋楠 译



关联数据

万维网上的结构化数据

Linked Data

Structured data on the Web

大卫·伍德 (David Wood)

玛莎·扎伊德曼 (Marsha Zaidman)

卢克·鲁思 (Luke Ruth) 著

迈克尔·豪森布拉斯 (Michael Hausenblas)

蒋楠 译

人民邮电出版社

北京

图书在版编目(CIP)数据

关联数据：万维网上的结构化数据 / (美) 大卫·伍德 (David Wood) 等著；蒋楠译. — 北京：人民邮电出版社，2018.1

ISBN 978-7-115-47264-9

I. ①关… II. ①大… ②蒋… III. ①数据管理
IV. ①TP274

中国版本图书馆CIP数据核字(2017)第286206号

版权声明

Original English language edition, entitled *Linked Data: Structured data on the Web* by David Wood, Marsha Zaidman, Luke Ruth and Michael Hausenblas, published by Manning Publications Co., 20 Baldwin Road, Shelter Island, NY 11964. Copyright © 2014 by Manning Publications Co.

Simplified Chinese-language edition copyright © 2017 by Posts & Telecom Press. All rights reserved.

本书中文简体字版由 **Manning Publications Co.** 授权人民邮电出版社独家出版。未经出版者书面许可，不得以任何方式复制或抄袭本书内容。

版权所有，侵权必究。

-
- ◆ 著 大卫·伍德 (David Wood)
玛莎·扎伊德曼 (Marsha Zaidman)
卢克·鲁思 (Luke Ruth)
迈克尔·豪森布拉斯 (Michael Hausenblas)
- 译 蒋楠
- 责任编辑 傅道坤
- 责任印制 焦志炜
- ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京鑫正大印刷有限公司印刷
- ◆ 开本：800×1000 1/16
印张：17
字数：372千字 2018年1月第1版
印数：1—2000册 2018年1月北京第1次印刷

著作权合同登记号 图字：01-2013-8993号

定价：69.00元

读者服务热线：(010)81055410 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京东工商广登字20170147号

内容提要

关联数据（linked data）是在万维网上表示和链接结构化数据的一系列技术，最早在 2007 年 5 月提出，旨在构建一张计算机能够理解的语义数据网络，而不仅仅是人能读懂的文档网络，以便在此之上构建更智能的应用。

本书分为 4 个部分，第 1 部分主要介绍了关联数据的基础知识、RDF（资源描述框架）数据模型，以及表示关联数据的通用标准序列化格式，旨在引导读者识别并使用万维网上的关联数据；第 2 部分重点讨论了开发和发布关联数据所用的技术，以及聚合数据所用的高级搜索技术；第 3 部分则讨论了如何使用 RDFa（属性中的资源描述框架）对网页进行 SEO、RDF 数据库与传统的关系数据库的区别、在万维网上共享用户数据集和项目的最佳方式，以及对语义网搜索结果中包含的项目和数据集进行优化；第 4 部分则将之前的内容进行了汇总，使用一个开源的关联数据应用服务器开发一个复杂的应用程序，并总结了从准备到发布关联数据的全过程。

本书适合具备 HTML、URI、HTTP 等基本的 Web 技术基础，并且想要了解、使用和发布关联数据的应用程序开发人员阅读。

序

本书正是一本我们所需要的关联数据技术指南。对于如何在万维网上使用和发布结构化数据，本书作了有益的介绍。

关联数据体现了我对万维网的最初设想，它是未来万维网的重要组成部分。万维网作为一种超链接文档的集合，虽然适合人类用户使用，却无法有效地用作数据。

实际上，大部分万维网都是数据驱动的，数据隐藏在服务器内部的文件中。1994 年，我在第一届国际万维网大会（International Conference of World Wide Web）总结发言的幻灯片中指出，文档在描述人物和事物（比如房屋产权地契）时，并未采用便于处理的方式捕获数据（实际所有权）。随着万维网的发展，其数据驱动的特质越来越明显。遗憾的是，无法将变化和隐藏的数据展示给用户。而发布遵循关联数据标准的数据有助于人类阅读和机器处理，从而让之前隐藏的数据流得以显现。

关联数据或许不像超文本万维网那样广为人知，但它能在很大程度上提高商业、科研等各个领域的工作效率。与万维网上其他形式的数据相比，机器可以更有效地读取、跟踪与组合关联数据。

一直以来，机器都被定位于从技术层面协助人们进行交流；如今，机器开始积极参与到这种交流之中。藉由关联数据，机器在人们的日常生活中发挥了越来越大的作用。

最近几年来，关联数据的应用已进入成熟期。在过去两年中，Google 宣布推出知识图谱服务，并在 Gmail 中采用 JSON-LD 序列化格式，还为 Schema.org 贡献了大量通用术语；IBM 宣布将 DB2 数据库升级为关联数据服务器；Facebook 则通过图谱 API 向外界公开关联数据。其他大型企业和政府机构也纷纷跟进。为了向刚刚涉足这个领域的程序员介绍关联数据开发，我们需要一本与本书类似的图书。本书将解答读者关心的问题——即便它无法提供所有问题的答案。对于希望了解、使用和发布关联数据的开发人员而言，本书是一本很好的教程。

我和本书作者 Dave Wood 认识已有十年，他开始为万维网联盟工作时我们就已相识。之后，我们共同从事一项研究。从 20 世纪 90 年代末以来，Dave 一直在为开发语义网和关联数据框架而不懈努力。作为一名开发人员，Dave 很好地展示了他的工作能力。

关联数据并非一个全新的概念。1989年，我在欧洲核子研究中心（CERN）工作时，撰写的万维网原始建议^①中就包括带有语义的超链接。建议中提到，“我们需要的系统类似于一个由圆圈和箭头组成的图表，这些圆圈和箭头可以代表任何内容”。这是我当时的愿景。实际上，我在1980年开发的Enquire程序已能捕获图谱中各种事物之间的关系。如今，通过增加计算机可以处理的含义，关联数据正在让这一愿景逐步成为现实。

众所周知，在基本的超文本万维网中，所有箭头都指向同一件事：“这里的信息很有趣！”关联数据支持使用箭头表示可以通过URI命名的任何事物，从而扩展了“文档万维网”的范畴。超链接可以获得所需的语义，并会在这个过程中变得更加有用。

功能强大的关联数据网是对超文本链接文档网的有益补充。之所以“链接”，是因为网页的价值在很大程度上取决于所链接的内容，以及网页中信息的内在价值。关联数据的语义网同样如此，可能更甚。数据本身的价值已然不菲，与其他数据的链接让这种价值再次提升。

我认为，无论国籍、语言、经济动因或兴趣如何，万维网的发展应遵循为所有人服务的宗旨。关联数据只是发展浪潮中的一分子，它并非结束，而是另一个开始。仍有大量工作等着我们去做，欢迎大家投身于下一代万维网的发展之中！

Tim Berners-Lee^②

万维网联盟主席

2016年度图灵奖得主

麻省理工学院工程系教授

南安普顿大学电子与计算机科学系教授

① 这份名为《关于信息管理的建议》（*Information Management: A Proposal*）涉及Enquire，并提出了一个更为精巧的信息管理系统，后者基于嵌入到可读文本的链接。原文参见<https://www.w3.org/History/1989/proposal.html>。
——译者注

② Tim Berners-Lee，1955—，英国计算机科学家，万维网发明者，被称为“万维网之父”，曾入选《时代》周刊“20世纪最重要的100位人物”。Tim放弃了申请万维网的专利，将这项技术无偿推向全世界，极大地改变了人类的生活方式。2017年4月，Tim因“发明万维网、第一个Web浏览器以及支持万维网扩展的基础协议和算法”而获得2016年度图灵奖。图灵奖是计算机领域的最高荣誉，有“计算机界的诺贝尔奖”之称。——译者注

译者序

从 2010 年 8 月起,全球最大的图书馆之一——大英图书馆开始通过关联数据规范发布书目信息,并向研究人员免费开放全部馆藏书目。借由英国国家书目关联数据平台(BNB Linked Data Platform),公众可以自由访问采用关联开放数据规范发布的英国国家书目信息。而在大英图书馆之前,瑞典皇家图书馆、匈牙利国家图书馆、德国国家图书馆已相继将书目数据发布为关联数据。

2016 年 10 月,英国教育和研究领域数字解决方案提供商 Jisc 提出了国家书目知识库(NBK)的设想,旨在满足高等教育社区的学习和研究需求,并对实体资源与数字资源的获取方式进行改进,以实现构建英国国家数字图书馆的愿景。这个为期两年的项目于 2017 年初正式启动,Jisc、大英图书馆以及其他组织密切合作,共同致力于这一有可能在未来英国信息基础设施中扮演关键角色的国家级项目。在 2012 年《大英百科全书》停止印刷版发行,全面转向数字化之后,摆脱实体存在的图书馆或许将是未来的趋势。

在图情领域,关联数据和大数据之间也逐渐呈现出一种“你中有我,我中有你”的态势。

随着 RDF 数据量的迅速增长,不少关联数据应用开始寻求大数据解决方案。由于关系数据库并非存储大规模 RDF 数据的最佳方式,采用 NoSQL 这样的非关系数据库逐渐成为当前的研究热点。这种利用大数据解决方案实现的关联数据应用,不妨称之为“大”关联数据应用。

而相当一部分大数据应用也开始采用 RDF 数据模型对信息进行描述和编码,使数据包含可以被机器识别的语义,不仅能丰富大数据的语义,也让数据具备更好的互操作性。这种引入关联数据技术的大数据应用,不妨称之为“关联的”大数据应用。

无论“大”关联数据应用还是“关联的”大数据应用,都是为了实现数据的有效管理。虽然基于 NoSQL 的 RDF 存储和查询研究目前尚处于初级阶段,但是关联数据与大数据的结合或许将为图情领域带来一场技术变革。

作为第一种可行的语义网实现方式,关联数据在智能化搜索中也得到了应用。从 2012 年 5 月 Google 发布知识图谱(Knowledge Graph)服务以来,各大搜索引擎相继跟进,纷纷推出了自己的知识图谱类产品。借由知识图谱,搜索引擎将围绕搜索请求产生的零散知识内容聚合在一起,以提供条理性更强的信息,以使用户发现意想不到的知识。

以 Tim Berners-Lee 作为关键字在 Google 中进行查询,除常规的搜索结果外,页面右侧将显示 Tim Berners-Lee 的个人信息(见图 1)。而对于“特朗普女儿的身高”(Trump's daughter's husband's height)这种相对复杂的查询,知识图谱同样可以直接给出答案(见图 2)。



图 1



图 2

百度在 2013 年推出的知识图谱类产品“知心”,是这家互联网巨擘在构建知识集群方面的一种尝试。如果能在垂直领域提供更为精细的搜索结果,使用户可以在搜索框内一次性完成所有搜索需求,就有助于搜索引擎为广告客户提供定制化服务,驱动在线营销从粗放型向精细型转变。而实现“搜索即答案”,也是下一代搜索引擎追求的目标之一。

在加州长滩举行的 TED2009 上,万维网之父 Tim Berners-Lee 表示,关联数据有助于摆脱信息孤岛。他以研究老年痴呆症的科研团队为例,阐述了关联数据在探索跨学科问题中的应用。团队在一个数据库中建立了基因图组,在另一个数据库中建立了蛋白质数据,并将基因图组和蛋白质数据发布为关联数据。科学家们或许希望了解,哪些蛋白质参与了信号转导并与锥体神经元相关?如果在 Google 上搜索这个问题,返回的 22.3 万个结果并无太大意义,因为之前没有人问过类似的问题。不过,如果检索关联数据,则可以获得 32 个精准的结果,每个结果都是与特征相关的蛋白质。Tim Berners-Lee 强调,破除信息孤岛对于解决人类面临的重大问题至关重要。

关联数据并非只和大企业有关,它关乎我们每个人的生活,本书对此作了详细论述。这本面向 Web 开发人员的教程保持了 Manning Publications 一贯的高水准,有助于初涉该领域的开发人员了解关联数据的前因后果。Manning Publications 以出版技术类图书著称,其“服饰书”系列与 O'Reilly Media 的“动物书”系列备受开发人员的推崇。

非常感谢人民邮电出版社的傅道坤编辑给予译者翻译本书的机会,他为本书中文版面世所作的努力以及专业和严谨的态度,给译者留下了深刻印象。虽然译者尽力而为,但水平有限,疏漏之处在所难免。恳请读者不吝赐教,提出宝贵的意见和建议。译者的联系方式: milesjiang314@gmail.com。

蒋楠

2017 年 12 月

前言

我们热爱万维网，也欣赏它的发展方式。20 世纪 90 年代初，万维网还只是一个将文档链接在一起的简单网络；如今，它已成为全球信息的框架。显然，如何表达万维网上的数据是下一步需要解决的问题，但这个问题并不简单。

本书几名作者接触万维网的方式各不相同，但因为关联数据而走到一起。David 最初是一名程序员，之后投身商界；Marsha 是一名教育工作者；Luke 是一名学生。Marsha 和 David 接触计算机时，穿孔卡与纸带还是主要的输入设备。万维网本质上是对 1 和 0 的一种抽象，自面世以来深受人们的青睐。

David 于 1993 年进入 DEC，在加利福尼亚州富有传奇色彩的 Western Research Lab 工作，并从那时起开始接触万维网。这段经历让 David 大开眼界。他接触的第一个大型网站展示了梵蒂冈数千件艺术品的照片。另一个网站则显示了 DEC 的在研项目列表，这份列表与所有研究人员的 Web 服务器相连，便于研究人员获取每个项目的详细文档。David 被这个网站深深吸引，这是他遇到的最有意思的项目网站：只要能链接到数据库和电子表格，就能链接到相应的文档。

Marsha 同样在很早的时候就已接触万维网。那时候，Gopher 还是主要的信息检索工具^①，Web 浏览器还需要依靠终端来工作。Marsha 紧跟万维网快速发展的步伐，为培养新一代计算机人才尽心竭力。在漫长的职业生涯中，Marsha 见证了电子表格和数据库的诞生，以及它们为决策过程所带来的惊人变化，这激发了人们将数据迁移到万维网的兴趣。

在关联开放数据项目启动时，Marsha 介绍 David 前往玛丽华盛顿大学任教。2011 年，Luke 参加了 David 为美国本科生开设的第一门关联数据课程，并在 David 的指导下进行独立的研究和实习。他最终被 David 聘用，参与到关联数据项目中。

Luke 和 David 是 Callimachus 项目的积极贡献者。Callimachus 是一个开源的关联数据平台（详见第 9 章），我们利用它为美国政府机构、制药公司、出版社、医疗保健企业等领域的客户开发应用程序。这些项目都涉及关联数据的创建、操作和使用。

① 一种在互联网上分发、搜索和检索文档的 TCP/IP 协议，由美国明尼苏达大学开发。在万维网出现之前，Gopher 是最主要的信息检索工具，但目前已很少有人使用。——译者注

前言

本书作者决定为 Web 开发人员编写一本介绍关联数据的图书，因为这方面的资料相当匮乏。我们研究了各种关联数据规范，并阅读了相关的学术论文。尽管也可以找到其他介绍关联数据的图书（David 参与了其中两本书的编辑工作），但没有一本书专门面向开发人员。在写作过程中，我们将一线开发经验和教学经验相结合，致力于打造一本实用的教程，希望读者能认可作者的努力。

本书作者有幸与万维网联盟合作，这个松散的国际性组织致力于研究和规范万维网上的数据应用。希望读者能阅读本书，并与我们一起为万维网的发展添砖加瓦。对于万维网的未来，我们拭目以待。

致谢

作者谨向最初的关联开放数据项目团队成员表示感谢，本书引用了不少他们的工作成果。作者对 Michael Stephens、Jeff Bleiel、Ozren Harlovic、Maureen Spencer、Mary Piergies、Linda Recktenwald、Elizabeth Martin、Janet Vail 以及 Manning Publications 的其他工作人员表示感谢，他们的辛勤工作让本书得以付梓。

作者还要对以下审校人员表示感谢，他们在多个审校阶段阅读了书稿，并提出许多宝贵的意见和建议：Alain Buferne、Artur Nowak、Craig Taverner、Cristofer Weber、Curt Tilmes、Daniel Ayers、Gary Ewan Park、Glenn McDonald、Innes Fisher、Luka Rajjevic、M. Edward Borasky、Michael Brunnbauer、Michael Pendleton、Michael Piscatello、Mike Westaway、Owen Stephens、Paulo Schreiner、Philip Poots、Robert Crowther、Ron Sher、Thomas Baker、Thomas Gängler 以及 Thomas Horton。

特别感谢本书的技术校对 Zachary Whitley，他在本书付印前仔细审定了最终书稿。同时感谢 Tim Berners-Lee 为本书作序。

Author Online 论坛、公共 LOD 邮件列表以及万维网联盟 RDF 工作组的专业人士提出的意见和建议，极大提高了本书的质量。衷心感谢参与 Manning 抢先体验计划（Manning Early Access Program, MEAP）的读者，他们在 Author Online 反馈的意见对最终书稿的质量产生了很大影响。最后，感谢美国剑桥、纽约、华盛顿、北弗吉尼亚州、马里兰州中部等地的语义网 Meet-up 的组织者给予作者介绍本书的机会。

David 对 Bernadette 表示感谢，她的支持让 David 能心无旁骛地投入到每个项目中。David 还要感谢本书的其他作者，正是由于大家的共同努力，本书才得以面世。

Marsha 对她的丈夫 Steven 表示衷心感谢，他的信任和鼓励让 Marsha 满怀信心开始一段新的征程。Marsha 特别感谢 David 的邀请，并相信她可以将之前的教学经验应用到本书写作中。感谢两位合著者 Luke 和 David，与他们的合作是一段难忘而有益的经历。

Luke 感谢 David 和 Marsha 邀请他参与本书的编写工作。无论技术层面还是社会层面，Luke 都从两位合著者身上受益良多。Luke 还要感谢父母 Rick 和 Tania 使他认识到教育和尝试新事物的重要性，以及妻子 Laura 始终如一的支持。

关于本书

关联数据（Linked Data）是在万维网上表示和连接结构化数据的一系列技术。本书将向读者介绍如何访问、创建并使用关联数据。关联数据的一个神奇之处在于，它很容易就能与其他关联数据进行组合，从而构成新的知识。

关联数据让万维网成为一个全球性数据库，我们称之为数据网（Web of Data）。开发人员可以利用 SPARQL 查询语言同时查询多个信息源的关联数据，并动态合并查询结果，这是传统数据管理技术很难或根本无法做到的。书中的示例取自公共信息源，不过所介绍的技术很容易就能用于私有数据。读者或许对书中使用的某些资源不太了解，但它们都不难从万维网上找到。如果在实际中遇到这些资源，读者不妨认真研究一下。如果书中的截图和引用的 URL 与读者在浏览这些网站时所看到的实际内容不一致，我们提前对此表示歉意。万维网始终处于快速变化之中，任何印刷品都无法绝对精确地反映所有变化。但我们承诺，所有截图和 URL 在本书付印时都是正确的。

藉由关联数据技术，我们更容易和他人共享数据。理论上说，可以采用关联数据描述任何内容。万维网上的关联数据可以被发现、共享并与其他用户的数据进行合并。与传统的数据管理系统不同，关联数据将信息从专有容器（proprietary container）中释放出来，任何人都能使用这些信息。与其他数据一样，关联数据的质量和效用由数据使用者负责评估。人们只信任可靠的数据源。

本书读者对象

对于希望了解、使用和发布关联数据的应用程序开发人员而言，本书值得一读。本书假定读者对 HTML、URI、HTTP 等基本的 Web 技术已有所了解。本书将介绍关联数据并在各种背景下讨论其应用，并论述指导关联数据使用的 4 项原则。此外，本书还将讨论如何在万维网上查找、使用和发布关联数据，并通过复杂性逐步增加的一些实际应用加以说明。

路线图

本书总共 11 章，分为 4 个部分，并包括两个附录和一个词汇表。

第 1 部分“关联数据网”将介绍关联数据的基础知识，论述 RDF（资源描述框架）数据模型，并讨论表示关联数据的通用标准序列化格式。这部分内容将引导读者识别并使用万维网上的关联数据。

- 第 1 章将介绍关联数据，在各种背景下讨论其应用，论述指导关联数据使用的 4 个原则，并通过一个应用程序展示关联数据的使用。
- 第 2 章将介绍 RDF 以及它与关联数据的关系。我们将论述 RDF 数据模型，并讨论实际中可能用到的一些关联数据重要概念。在这一章最后，我们将讨论文件类型和 Web 服务器遇到的常见问题，并给出解决这些问题的方法。
- 第 3 章将讨论万维网的分布式特性，并介绍数据和文档相互连接的方法。读者将了解文档网（Web of Documents）和数据网之间的关系，以及如何在万维网上查找并使用关联数据。

第 2 部分“关联数据进阶”将重点讨论开发和发布关联数据所用的技术，并介绍聚合数据所用的高级搜索技术。我们利用 SPARQL 查询语言来搜索相关的关联数据数据集，并将搜索结果加以聚合。

- 第 4 章将介绍如何利用 FOAF 词表和 Relationship 词表在万维网上创建、链接与发布关联数据。
- 第 5 章将介绍 RDF 所用的 SPARQL 查询语言，后者可以像查询数据库一样查询数据网——尽管数据网是一个非常庞大的、由大量分布式数据集构成的数据库。

第 3 部分“关联数据实战”将讨论如何使用 RDFa（属性中的资源描述框架）对网页进行 SEO（搜索引擎优化）。我们将介绍 RDF 数据库，并讨论它与传统的关系数据库之间的区别。我们还将介绍在万维网上共享用户数据集和项目的最佳方式，并对语义网搜索结果中包含的项目和数据集进行优化。

- 第 6 章将介绍如何利用 RDFa 强化 HTML 网页，从而获得更准确的搜索结果。我们将介绍面向业务的 GoodRelations 词表以及其他使用 schema.org 的技术。
- 第 7 章将介绍 RDF 数据库，并讨论 RDF 数据库较之关系数据库的差异和优势。一般来说，集成 RDF 格式的信息相对不难，但用户所需的信息通常存储在非 RDF 数据中，需要进行转换以便处理。这一章将介绍如何将非 RDF 数据转换为 RDF 格式，以便集成到其他应用程序中。
- 第 8 章将介绍如何描述新创建的关联数据，并将其链接到更大的关联数据系统中。我们将讨论并应用 DOAP 词表（用于描述项目）、VoID（用于描述数据集）以及语义站点地图（用于描述网站中的关联数据产品）。此外，这一章还将介绍在 LOD 云上发布数据时需要遵循的规则。

第 4 部分“归纳与整合”将把之前讨论的所有知识点串联在一起。本章将使用开源的关联数据应用服务器开发一个复杂的应用程序，并总结从准备到发布关联数据的全过程。

- 第 9 章将介绍 Callimachus，这是一个开源的关联数据应用服务器。我们将讨论 Callimachus 的基本用法以及利用 RDF 数据生成网页，并展示如何通过 Callimachus 构建应用程序。
- 第 10 章将总结从准备到公开发布关联数据的全过程，并对构建 URI、自定义词表等容易

忽视的环节进行说明。

- 第 11 章将讨论语义网目前的发展状态，以及关联数据在其中所扮演的角色。我们将介绍几个有趣的关联数据应用程序，并尝试对语义网和关联数据今后的发展方向进行预测。

正文后的附录将提供一些补充信息。

- 附录 A 将介绍如何设置本书所用工具的开发环境。
- 附录 B 将介绍常用的 SPARQL 查询结果格式。
- 词汇表将列出并定义本书所用的术语。

本书用法

为充分利用本书，建议读者按顺序阅读每个章节，下载和执行示例应用程序，并尝试修改这些示例以加强对概念的理解。如果程序中需要使用某些软件，我们将提供相应资源的链接。希望本书能为读者了解、使用和发布万维网上的关联数据打下坚实的基础。

目录

第1部分 关联数据网

第1章 关联数据简介 2

- 1.1 关联数据定义 3
- 1.2 关联数据并非万能 4
- 1.3 关联数据实战 5
 - 1.3.1 释放数据 6
 - 1.3.2 关联数据在 Google 富摘要和 Facebook “点赞” 中的应用 6
 - 1.3.3 关联数据拯救了 BBC 7
- 1.4 关联数据原则 9
 - 1.4.1 第1原则：使用 URI 命名事物 10
 - 1.4.2 第2原则：使用 HTTP URI 以便于用户查找事物名称 10
 - 1.4.3 第3原则：在用户查找 URI 时提供有用的信息 11
 - 1.4.4 第4原则：包含指向其他 URI 的链接 12
- 1.5 关联开放数据 (LOD) 项目 12
- 1.6 数据描述 13
- 1.7 RDF：关联数据所用的数据模型 16
- 1.8 关联数据应用程序剖析 18
 - 1.8.1 获取设施的关联数据 19

- 1.8.2 通过关联数据创建 UI 21
- 1.9 小结 24

第2章 RDF：关联数据所用的数据模型 25

- 2.1 关联数据原则让 RDF 得以扩展 26
- 2.2 RDF 数据模型 30
 - 2.2.1 三元组 31
 - 2.2.2 空节点 32
 - 2.2.3 类 33
 - 2.2.4 类型字面量 34
- 2.3 RDF 词表 35
 - 2.3.1 通用词表 36
 - 2.3.2 自定义词表 39
- 2.4 关联数据所用的 RDF 格式 40
 - 2.4.1 Turtle：人类可读的 RDF 41
 - 2.4.2 RDF/XML：企业所用的 RDF 44
 - 2.4.3 RDFa：嵌入 HTML 网页的 RDF 46
 - 2.4.4 JSON-LD：JavaScript 开发者所用的 RDF 49
- 2.5 与 Web 服务器和关联数据发布有关的问题 52
- 2.6 文件类型与 Web 服务器 54
- 2.7 对 Apache 服务器的控制有时如何处理 55

2.8 关联数据平台 56

2.9 小结 56

第3章 使用关联数据 57

3.1 像万维网一样思考 57

3.2 如何使用关联数据 58

3.3 查找分布式关联数据的工具 60

3.3.1 Sindice 60

3.3.2 SameAs.org 61

3.3.3 Data Hub 62

3.4 聚合关联数据 63

3.4.1 聚合已知数据集中的关联数据 63

3.4.2 使用浏览器插件获取网页中的关联数据和 RDF 67

3.5 关联数据网的抓取与数据的聚合 69

3.5.1 使用 Python 抓取关联数据网 69

3.5.2 利用聚合后的 RDF 输出 HTML 72

3.6 小结 72

第2部分 关联数据进阶

第4章 利用 FOAF 创建关联数据 74

4.1 创建个人 FOAF 配置文件 75

4.1.1 FOAF 词表简介 76

4.1.2 方法 I: 手动创建基本的 FOAF 配置文件 77

4.1.3 改进基本的 FOAF 配置文件 78

4.1.4 方法 II: 自动生成 FOAF 配置文件 80

4.2 为 FOAF 配置文件添加更多内容 83

4.3 发布 FOAF 配置文件 85

4.4 FOAF 配置文件的可视化 86

4.5 应用程序: 采用自定义词表链接 RDF 文档 87

4.5.1 创建愿望清单词表 87

4.5.2 创建、发布并链接愿望清单文档 88

4.5.3 为愿望清单文档添加内容 89

4.5.4 小书签程序初探 91

4.6 小结 92

第5章 SPARQL: 查询关联数据网 93

5.1 典型 SPARQL 查询概述 94

5.2 采用 SPARQL 查询扁平 RDF 文件 95

5.2.1 查询单个 RDF 文件 95

5.2.2 查询多个 RDF 文件 98

5.2.3 查询万维网上的 RDF 文件 100

5.3 查询 SPARQL 端点 100

5.4 SPARQL 查询类型 102

5.4.1 SELECT 查询 102

5.4.2 ASK 查询 104

5.4.3 DESCRIBE 查询 105

5.4.4 CONSTRUCT 查询 105

5.4.5 SPARQL 1.1 Update 106

5.5 SPARQL 结果格式 (XML 与 JSON) 107

5.6 利用 SPARQL 查询创建网页 108

5.6.1 创建 SPARQL 查询 109

5.6.2 创建 HTML 页面 110

5.6.3 创建 JavaScript 表格 111

5.6.4 创建 JavaScript 地图 112

5.7 小结 115

第3部分 关联数据实战

第6章 强化搜索引擎的结果 118

6.1 通过嵌入 RDFa 以强化 HTML 119

- 6.1.1 利用 FOAF 词表添加 RDFa 标记 122
- 6.1.2 在 HTML span 属性中使用 RDFa 125
- 6.1.3 从包含 FOAF 的 HTML 文档中提取关联数据 126
- 6.2 采用 GoodRelations 词表嵌入 RDFa 127
 - 6.2.1 GoodRelations 概述 127
 - 6.2.2 利用 GoodRelations 强化嵌入 RDFa 的 HTML 130
 - 6.2.3 对选择 RDFa GoodRelations 的进一步观察 136
 - 6.2.4 从包含 GoodRelations 的 HTML 文档中提取关联数据 138
- 6.3 采用 Schema.org 词表嵌入 RDFa 141
 - 6.3.1 Schema.org 概述 141
 - 6.3.2 通过 Schema.org 强化使用 RDFa Lite 的 HTML 143
 - 6.3.3 对利用 Schema.org 选择 RDFa Lite 的进一步观察 145
 - 6.3.4 从包含 Schema.org 的 HTML 文档中提取关联数据 147
- 6.4 选择 Schema.org 还是 GoodRelations 148
- 6.5 从 HTML 中提取 RDFa 并执行 SPARQL 查询 149
- 6.6 小结 150

第 7 章 RDF 数据库基础 151

- 7.1 RDF 数据库分类 151
 - 7.1.1 RDF 数据库的选择 153
 - 7.1.2 RDF 数据库与关系数据库的比较 153
 - 7.1.3 RDF 数据库的优点 158
- 7.2 将电子表格数据转换为 RDF 159
 - 7.2.1 将 MS Excel 转换为 RDF 的简单示例 159
 - 7.2.2 将 MS Excel 转换为关联数据 161
 - 7.2.3 选择 RDF 转换工具 163

- 7.3 应用程序：在 RDF 数据库中收集关联数据 163
 - 7.3.1 过程概述 163
 - 7.3.2 利用 Python 聚合数据源 164
 - 7.3.3 理解输出 167
- 7.4 小结 169

第 8 章 数据集 170

- 8.1 DOAP 词表 171
 - 8.1.1 创建 DOAP 文件 172
 - 8.1.2 使用 DOAP 词表 174
- 8.2 利用 VoID 记录数据集 178
 - 8.2.1 VoID 概述 178
 - 8.2.2 准备 VoID 文件 179
- 8.3 站点地图 181
 - 8.3.1 非语义站点地图 182
 - 8.3.2 语义站点地图 183
 - 8.3.3 启用站点发现 185
- 8.4 链接到其他用户的数据 187
- 8.5 示例：利用 owl:sameAs 实现数据集之间的互联 193
- 8.6 加入 Data Hub 195
- 8.7 从 DBpedia 请求指向用户数据集的出站链接 197
- 8.8 小结 198

第 4 部分 归纳与整合

第 9 章 Callimachus: 关联数据管理系统 200

- 9.1 Callimachus 入门 202
- 9.2 使用 RDF 类创建网页 202
 - 9.2.1 为 Callimachus 添加数据 203
 - 9.2.2 向 Callimachus 通告 OWL 类 204
 - 9.2.3 将 Callimachus 视图模板与用户的类相互关联 205