Programming the Semantic Web

语义Web编程

Toby Segaran,
Colin Evans, Jamie Taylor 著
胡鹤 译

O'REILLY®
机械工业出版社
China Machine Press
HZ BOOKS
华章IT

语义 Web 编程
Programming the Semantic Web

[美] 托比·塞加兰 (Toby Segaran)
科林·埃文斯 (Colin Evans)
杰米·泰勒 (Jamie Taylor) 著
胡鹤 译

Beijing · Boston · Farnham · Sebastopol · Tokyo O'REILLY®

O'Reilly Media, Inc. 授权机械工业出版社出版

机械工业出版社

图书在版编目（CIP）数据

语义 Web 编程 /（美）托比·塞加兰（Toby Segaran）等著；胡鹤译 . —北京：机械工业出版社，2019.1

（O'Reilly 精品图书系列）

书名原文：Programming the Semantic Web

ISBN 978-7-111-61587-3

I. 语… II. ①托… ②胡… III. 语义网络－网络编程－指南 IV. TP18-62

中国版本图书馆 CIP 数据核字（2018）第 281670 号

北京市版权局著作权合同登记

图字：01-2017-8671 号

书　　名 / 语义 Web 编程

书　　号 / ISBN 978-7-111-61587-3

责任编辑 / 陈佳媛

封面设计 / 张健，Karen Montogomery

出版发行 / 机械工业出版社

地　　址 / 北京市西城区百万庄大街 22 号（邮政编码 100037）

印　　刷 / 北京市荣盛彩色印刷有限公司

开　　本 / 178 毫米 ×233 毫米　16 开本　17.25 印张

版　　次 / 2019 年 1 月第 1 版　2019 年 1 月第 1 次印刷

定　　价 / 79.00 元（册）

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：(010)88379426；88361066

购书热线：(010)68326294；88379649；68995259

投稿热线：(010)88379604

读者信箱：hzit@hzbook.com

译者序

我很荣幸成为这本书的译者，本书是O'Reilly公司出版的第一本面向语义Web编程的经典教材。作者Toby Segaran是畅销书《Programming Collective Intelligence》的作者，他创办了Incellico生物技术软件公司。另两位作者Colin Evans和Jamie Taylor分别就职于Metaweb公司和Freebase公司，是机器学习和语义分析方面的专家。该书是一本适合语义Web入门实践的优秀指南，有利于对语义Web研究感兴趣的初学者快速掌握相关的编程基础。

Web的发明者国际W3C主席Tim Berners-Lee首次提出了语义Web概念："语义Web并不是一个孤立的Web，而是对当前Web的扩展，语义Web上的信息具有定义良好的含义，使得计算机之间以及人类能够更好地彼此合作。"在语义Web的愿景中，人们会用带有标准语义的方式来清晰地描述所提供的各种资源内容，从而为互联网上的各种资源建立完善的语义描述。在这种语义描述基础上，进一步构建标准化的语义推理、证明以及信任体系，为各种智能代理提供可靠的基础设施和标准化的交互模式。语义Web中的计算机能够通过智能代理，自动化地在海量的互联网资源中准确定位到所需要的信息，从而将目前分散孤立的Web信息融合集成为一个巨大的全球数据库，革命性地改变人们利用Web的方式。

自从1998年Tim Berners-Lee提出语义Web设想到现在，已经有二十年的时间。这二十年间语义Web的发展道路并不平坦，距离理想中的语义Web愿景还有很长的路要走。近年来，随着知识图谱系统的流行，语义关联数据越来越丰富，语义Web的相关研究进入快速发展期，吸引了很多新的研究者投入该领域研究，本书正是在这样的背景下翻译完成的。值得注意的是，书中示例所使用的网站*www.semprog.com*已经停止服务了，读者可以从*https://resources.oreilly.com/examples/9780596153823/*获取本书的源代码进行练习。

在翻译本书的过程中，译者虽然尽最大努力尊重原义，尽可能避免产生歧义，但由于才疏学浅，难免存在翻译不当之处，敬请广大读者批评指正。

本书能得以顺利出版，要感谢华章公司的大力支持，尤其是王春华和陈佳媛两位编辑给予了无私的帮助，在此一并表示衷心感谢！

译者

2018 年 11 月

序言

几年前，Tim Berners-Lee（万维网发明人）认为，当人们不再问“为什么？”而是开始问“如何做？”时，我们就会知道语义 Web 将走向成功——整个过程就像许多年前的万维网一样。有了这本书，我终于可以舒服地说，我们已经经过了这个转折点。本书是关于“如何做”的书——它为开发语义 Web 的程序员提供了当下所需的工具！

本书介绍的语义 Web 方法非常适合于要积极利用这些新 Web 技术的程序员社区。十多年前，像我这样的研究人员开始接触语义 Web 背后的一些想法，从 1999 年到 2005 年，大量的研究经费投入到这个领域。来自研究人员的“噪声”有时会掩盖这样一个事实，即该研究领域的实用技术并不是（高端的）火箭科学。事实上，本书中所介绍的技术已经非常成熟，现在已成为 Web 开发人员工具包中的重要组成部分。

在 2000 年和 2001 年，有关语义 Web 的文章开始出现在 Web 的内容空间中。2005 年左右，我们不仅看到一些小公司开始参与这个领域，还看到了一些像 Oracle 这样的大公司也在拥抱这项技术。2006 年年底，John Markoff 在《纽约时报》发表了一篇关于“Web 3.0”的文章，越来越多的开发人员开始认真研究语义 Web，并开始喜欢他们所看到的内容。这个开发者社区帮助创建了相关的工具和技术，因此在 2009 年，我们开始看到此领域真正起飞。语义 Web 和各种相关技术用途的文章几乎每天都在出现。

美国政府正在使用语义 Web 技术来提高政府数据的透明度。谷歌公司和雅虎公司正在从 Web 文档中收集和处理嵌入式 RDFa，而微软公司最近在基于语言的 Web 应用程序中讨论了它所做的一些语义工作。Web 3.0 应用程序正在吸引各种各样的用户，正是这些用户使早期的 Web 2.0 应用程序受到公众关注，而一些你可能还没听说过的创新初创公司正在探索如何将语义技术加入到日益广泛的 Web 应用程序中。

然而，所有这些令人兴奋的进展都遇到了明显的困难。现在有更多的人在问“怎么做？”，但是由于这项技术刚刚出现，很少有人知道如何回答这个问题。像我这样早期的语义 Web 传播者已经非常善于向包括数据库管理员、政府雇员、实业家和学者在内的众多人士解释语义 Web 的发展愿景，但最近提出的问题越来越难以解决。当一家财富 500 强

公司的首席技术官问我为什么要关注这类技术时，我迫不及待地想要回答。但是，当他的开发人员问我如何为某些嵌入式 RDFa 中表达的谓词找到最适当的宾语时，或者 SPARQL 查询的 OPTIONAL 子句中的 BNode 绑定如何工作时，我知道这些问题很快会超出我的能力范围之外。然而，随着本书的出版，我现在可以指着它说："答案就在那里"。一直以来缺少从程序员的角度讲解语义 Web 工作的文献，如今这个漏洞终于被填补了。

这本书还解决了另一个重要的需求。鉴于语义网络"多层蛋糕"（参见第 11 章）的顶部仍然处于研究阶段，所以存在很多混淆。一方面，像"关联数据"和"Web 3.0"这样的术语被用来描述当今 Web 应用程序所需要的立即可用和快速扩展的技术；另一方面，人们还在探索将为下一代语义 Web 提供支持的"语义 Web 2.0"开发。本书为读者提供了一个简单的方法，让读者能够区分"当前的实际"和"天上的馅饼"。

最后，我喜欢这本书的另一个原因是：它包含了我经常称为"懂点语义学走得更远"的哲学。[注1] 在 Web 上，开发人员不需要成为哲学家、AI 研究员或逻辑学家来理解如何让语义 Web 正常工作。然而，弄清楚究竟需要多少知识才能胜任工作是一个真正的挑战。在本书中，Toby、Jamie 和 Colin 将向你展示"刚好够用的 RDF"（第 4 章）和"刚好够用的 OWL"（第 6 章），让你这个程序员准备就绪并开始工作。

简而言之，语义 Web 技术就在这里，相关工具已经准备就绪，本书将告诉你如何使它为你工作。你还在等什么？ Web 的未来就在你的指尖。

Jim Hendler

纽约州，奥尔巴尼

2009 年 3 月

注 1：http://www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html

目录

第四部分 后记

前言

与生物有机体一样，计算机运行在复杂的、相互关联的环境中，系统中的每个部分都会影响其他很多部分。类似于捕食者与被捕食者的关系，应用程序和它们消费的数据往往遵循着共同进化的路径。应用程序中的累积更改最终需要修改其操作的数据结构。与之相反，当向数据源增加内容时，表达附加信息的结构通常会迫使应用程序做相应修改。不幸的是，由于涉及很大的工作量，这种连锁变革往往会阻碍应用程序和数据源的改进。

在其核心上，语义技术通过使用简单的抽象模型来实现知识表示，从而将应用程序与数据分离开来。该模型释放了应用程序和数据之间的相互约束，使两者都能够独立进化。通过设计提高应用程序和数据之间的独立程度改善了数据的可移植性。任何理解相应模型的应用程序都可以处理任何使用该模型的数据源。正是这种数据可移植性构成了机器可读语义 Web 概念的基础。

当前的 Web 运行良好，因为人类是非常灵活的数据处理器。无论网页上的信息是作为表格、大纲还是多页面叙述的排列形式，我们都能够提取重要信息并用它来指导进一步的知识发现。然而，这种信息的异构性对于机器来说是无法解读的，而且网上数据的丰富表示形式只会使问题加重。如果 Web 上可用的丰富信息能够被内容提供者编码为语义数据结构，那么任何应用程序都可以访问和使用我们所依赖的丰富数据。在这个愿景中，不同来源的数据可以无缝地集成起来，从交汇融合中产生新的知识。这就是语义 Web 的愿景。

现在，应用程序是否可以利用这些丰富的数据做出任何有趣的事情，正是开发人员可以发力的地方！语义技术使开发人员可以专注于应用程序的行为而不是数据处理。当给定新的数据源时，这个系统会做什么？它如何使用改进后的数据模型？当多个数据源彼此丰富时，用户体验如何提高？将知识的利用和对底层数据的操作区分开来，可以让开发

人员专注于应用程序中带来价值的因素。

语义 Web 的愿景承诺美好，这个愿景的真正价值在于，它孕育了使数据更具可移植性和可扩展性的技术。无论你是在编写简单的混搭代码还是在维护高性能企业解决方案，本书都提供了一种标准和灵活的方法，用于集成系统和数据使之更适于未来发展。

排版约定

本书中使用以下排版约定：

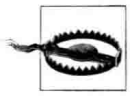
斜体（Italie）

表示 URL、电子邮件地址、文件名和文件扩展名。

等宽字体（`Constant width`）

表示程序清单，以及段落内用于引用的程序元素，如变量或函数名称、数据库、数据类型、环境变量、语句和关键字。

表示技巧、建议或一般说明。



表示警告。

使用代码示例

本书旨在帮助你完成工作。通常，你可以在你的程序和文档中随意使用本书中的代码。除非引用大量的源代码，否则无须征得我们的许可。例如，编写程序时使用本书中几个代码块是无须许可的，而销售或发行 O'Reilly 书籍中的示例 CD-ROM 需要获得许可。通过引用本书内容及示例代码来答疑解难是无须许可的，将本书中的大量示例代码加入到你的产品文档中是需要许可的。

我们赞赏，但不要求你在引用时注明出处。引用通常包括标题、作者、出版商和 ISBN。

如果你发现自己对示例代码的使用有失公允或违反了上述条款，请通过 *permissions@ oreilly.com* 与我们联系。

Safari 在线电子书

当你在喜爱的技术书籍封面上看到一个 Safari 在线电子书图标时，表示该书可通过 O'Reilly Network Safari Bookshelf 在线获取。

Safari 提供的解决方案比电子书更好。这是一个虚拟图书馆，你可以轻松搜索数以千计的高科技图书，剪切和粘贴代码示例，下载章节，并在需要最准确、最新的信息时快速找到答案。在 *http://my.safaribooksonline.com* 可免费试用。

如何联系我们

请将有关本书的评论和问题，发送给出版商：

美国：

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472

中国：

北京市西城区西直门南大街 2 号成铭大厦 C 座 807 室（100035）
奥莱利技术咨询（北京）有限公司

我们为本书设立了网页，列出勘误表、示例和其他信息。可以通过以下网址访问此页：

http://www.oreilly.com/catalog/9780596153816

要发表评论或提出有关本书的技术问题，请发送电子邮件至：

bookquestions@oreilly.com

有关我们的书籍、会议、资源中心的更多信息，以及 O'Reilly Network，请访问我们的网站：

http://www.oreilly.com

作者已经建立了一个网站作为社区资源，用于演示语义技术的实用方法。可以通过以下网址访问此网站：

http://www.semprog.com