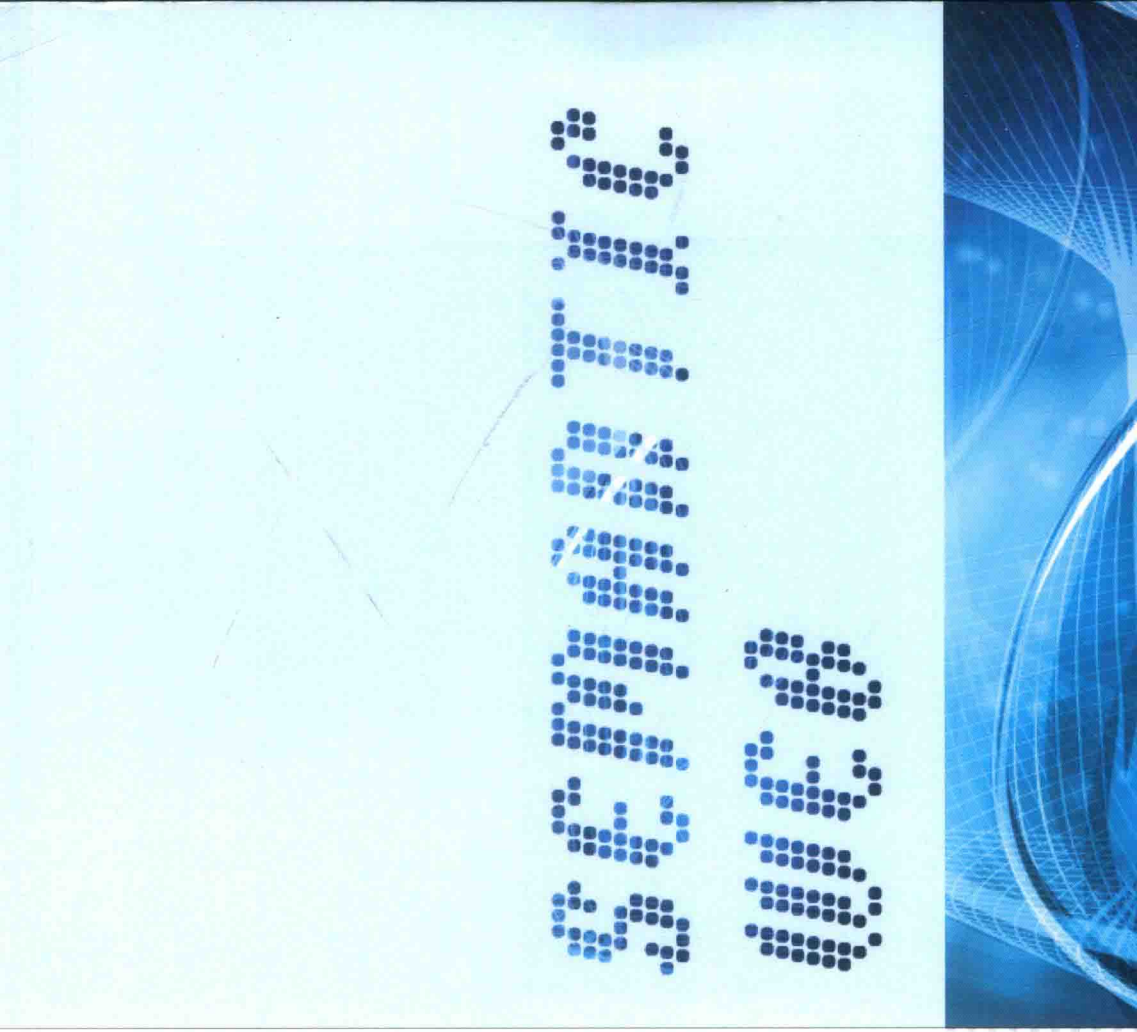


基于语义网的个性化网络学习服务

Semantic Web

吴笛 著

WUHAN UNIVERSITY PRESS
武汉大学出版社

中国博士后科学基金第58批资助项目“大数据环境下基于本体关联的学习资源个性化推荐研究”（编号：2015M580661）成果

基于语义网的个性化网络学习服务

Semantic Web

吴笛 著

安徽师范大学图书馆 藏书

WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

基于语义网的个性化网络学习服务/吴笛著. —武汉：武汉大学出版社，2017，4

ISBN 978-7-307-19230-0

Ⅰ. 基… Ⅱ. 吴… Ⅲ. 网络教学—研究 Ⅳ. G434

中国版本图书馆 CIP 数据核字(2017)第 079112 号

责任编辑：林 莉 责任校对：李孟潇 整体设计：马 佳

出版发行：**武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件：cbs22@whu.edu.cn 网址：www.wdp.com.cn)

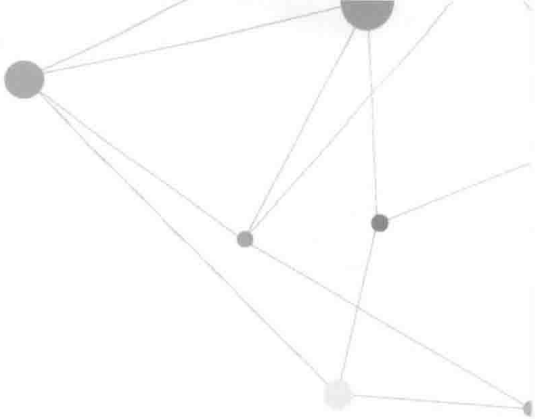
印刷：虎彩印艺股份有限公司

开本：720×1000 1/16 印张：11 字数：152 千字 插页：1

版次：2017 年 4 月第 1 版 2017 年 4 月第 1 次印刷

ISBN 978-7-307-19230-0 定价：35.00 元

版权所有，不得翻印；凡购我社的图书，如有质量问题，请与当地图书销售部门联系调换。



作者简介

吴笛 1984出生，男，湖北潜江人。讲师，博士后，毕业于武汉大学计算机学院，现任教于武汉大学教育科学研究院，主要从事多媒体技术和教育技术等方面研究。曾参与国家863计划引导项目“电子学习系统平台和教育资源库研究与开发”研究，在国内外刊物发表论文十余篇。

前　言

随着移动终端的普及和在线学习平台的盛行，各种灵活的教育方式帮助求学者根据自己在不同的地点根据自身需求，利用碎片化时间进行各种不同形式的学习。本书以语义网、数据挖掘、信息检索、个性化推荐等新一代信息技术为基础，促进构建感知化、个性化、智能化的现代网络教育技术体系，力图提高在线学习者对于网络学习方式的灵活性与适应性、主动性相结合的自我导向学习能力。

本书分为语义网、语义标注、知识本体、知识表示、关联数据、个性化服务、情境学习、推荐技术、总结与展望九个章节，通过研究基于语义网的学习资源语义分析技术为个性化学习服务提供异构的学习资源管理方式，书中整理和提炼了多种能够有效标注网络学习资源所含的语义信息方法，介绍了如何研究网上用户的学习行为，挖掘其知识背景、学习兴趣、学习风格、社会关系等信息，综合运用数据挖掘、关联分析、个性化推荐、逻辑推理等算法建立了一个高效、高内聚、低耦合的个性化学习服务技术体系，能够为学习者提供个性化的知识导航、内容推荐等服务。

本书力求深入浅出、阐述清楚，为从事教育技术、语义检索、知识管理的科技人员提供参考。书中所设计的个性化网络学习服务体系目的在于研究和探索如何从技术角度挖掘学生的自主学习能力，培养学生的主观能动性，这种学习方式、学习工具、学习理念的革新，能够迅速准确地为每个学习者提供合适的知识，并高效全面地完成学习，提高学习

此为试读，需要完整PDF请访问：www.ertongbook.com

效率和学习效果，使学习者更注重学习过程中的主动参与和协作建构。采用语义网和数据挖掘等技术开拓网络学习资源共享和个性化服务是一项有理论意义和应用前景的新探索，对促进语义网架构下的电子学习系统研究具有重要的现实意义和实用价值。本书是中国博士后科学基金第58批资助“大数据环境下基于本体关联的学习资源个性化推荐研究”（编号：2015M580661）成果。

目　录

此为试读，需要完整PDF请访问：www.ertongbook.com

第1章 语 义 网

1.1 概述

大约从20世纪50年代中期开始，社会发展已经进入了信息化时代，其代表性象征为“计算机”，工业生产力以信息技术为主体，是信息产生价值的时代。计算机互联网的发展促进了世界范围内信息的交流，使人们在获取信息和共享信息时更为便捷，让个人也具备了传统的大众传播工具才具有的传播空间和能力，但也带来了数据爆炸的问题，即人们很难从海量信息中寻找出自己真正想要得到的信息和数据资源，根本原因则是计算机无法真正理解数据信息所隐藏的内涵和关联，无法对关联内容进行有效的梳理。因此，万维网之父 Tim-Burners Lee 提出了语义网以解决上述问题，并且语义网是在现有网络上的一个延伸而不是一个独立的网络，语义网是在万维网基础上进行的扩展[1]。

1.2 语义网的特点

(1)处理对象

语义网的使用，既可以满足人们在上网时对信息的直观理解，也能

使计算机可以“解读”其语义信息。直观地讲，它依托万维网拥有的基础技术特点，同时融合新的技术使计算机在被使用时可以自己“解读”人为操作页面内的语义信息，为人们信息检索提供帮助，在通过内部处理信息后提供智能化的服务。因此，语义网不仅帮助个人可以理解其语义信息，还可使计算机能够理解。

(2)组织方式

语义网是在人理解信息和计算机理解信息基础上建立的一个枢纽，因此在组织它的理解信息时，既要对表面意思有一定理解也要兼顾其关联信息并相互连接，再依靠对信息内容本体的概念理解来关联网上相关资源，以健全网络信息共享资源形成知识数据库。

(3)构建目的

语义网不仅满足网络环境下对资源和信息的共享、获取，还能专门针对日益突出的“信息过载”等网络问题，其主要特点是通过技术手段对繁琐的信息进行理解并添加上相应的语义信息，使计算机理解其含义并帮助人们进行信息检索，这样便促进了网络应用更加人性化、智能化。

(4)处理过程

语义网是通过技术手段对信息加以相应的语义理解，使在利用计算机搜索时可以智能化的区分人们所需要的信息以及资源，在不同的用户使用时可以根据其个性化的需求梳理出他所需的相关资源和信息，所以人们在运用语义引擎搜索信息时可以根据需求个性化显示信息和资源。

1.3 体系结构

Tim Berners-Lee 于 2001 年设计并提出了语义网体系结构[2]，如图 1-1 所示，整个体系结构被分为七层，从底层开始每一层的功能逐渐向下进行扩展。

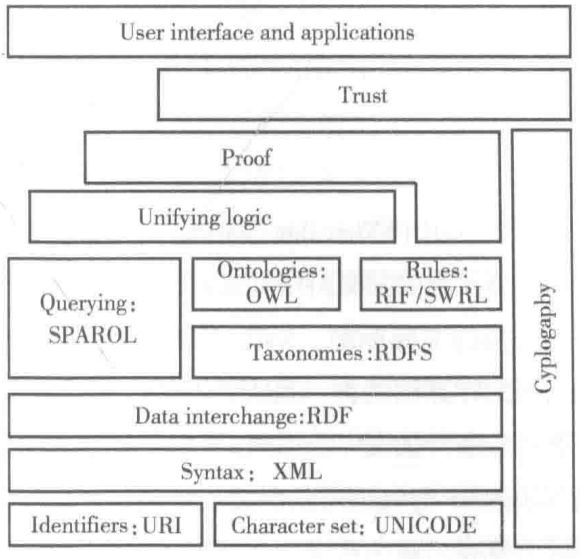


图 1-1 语义网体系结构

1.3.1 URI 和 Unicode

URI(Uniform Resource Identifier)是整个框架体系的底层基础部分，它是来标志、识别网络中的概念资源和物理资源的基础环节。Unicode作为一种字符集，也被称作为单一码、统一码或万国码，用以支持三种编码(UTF-8、UTF-16、UTF-32)之间的相互转换。众所周知，不同国家和地区使用的语言、文字集是不完全一样的，所以在互联网的数据交换中，往往会因为不同的字符集而出现乱码的状况，特别是在不同系统平台进行操作时对这种现象的处理十分困难。

就本质而言，全世界的书面上的语言种类很多，纷繁复杂，很难对其进行有效的管理，而对其进行编码管理不失为一种可行的解决办法。由于书面语言在计算机中转换文本会因跨语言、操作系统等原因而乱码，所以为了解决这个问题，RFC(Request For Comments)标准便为所

有国家语言字符编订了统一的二进制编码，保证了经过统一编码的字符能更方便地被语义网处理。

1.3.2 XML

可扩展标记语言 XML(eXtensible Markup Language)是由 W3C 设定的标准，是一种能够扩展的具有特殊标志的网络语言格式，它是目前互联网上进行数据交换的主要标准。XML 能够成为各种操作系统下各种应用程序之间通用的计算机语言，是因为其特点鲜明简洁、结构多变、操作简易，能够适应各种软硬件的需求。

命名空间 NS(Name Space)可以保证在不同的环境下可以使用同样的字符描述不同的事物，避免意义不同但描述相同的语句互相冲突。

DTD(Document Type Definition)文档类型定义和 XML Schema 功能类似，但 XML Schema 不仅具备有支持 XML 检验数据的功能，还能够支持各种类型的数据和命名空间 NS。命名空间 NS 运用的 URI 资源检索和 XML Schema 支持了许多种类的数据类型和信息检验功能，这是因为 XML 具有多变灵活的结构特点，这样就很适合在语法中体现不同的数据结构，进而能够将互联网中信息形式、构成和内容进行拆分，但是由于一些资源不具有语义的原因，因此需要更高层次的功能来解决对其的定义。

1.3.3 RDF

资源描述框架 RDF 是用来表征互联网上的元数据的，它是由 W3C 规定的用来表示网上资源的具有特殊标记的语言。它能够运用 XML 来完成程序之间互相运作互联网上的元数据，所以 RDF 能够运用 URI 来表示互联网上的资源。基于 RDF 可以提供一个对数据进行描述的模型，该模型所运用的语法是中立不冲突的。它的模型是一个三元组合的模

型，通过“资源—属性—属性值”这样的结构来表示需要的语言，在这个模型中资源用 URI 来表示，属性代表基础的信息，信息本身所带的属性皆具有具体的类型。RDF 仅仅对需要表达的信息框架进行了定义，但是却没有对具体的元数据进行相关的定义，因为对于不同资源来说，其所要求的元数据不一定是相同的。所以想要解决这个问题，就需要提供一个元数据的集合体，这个集合体也被叫作词汇集。在 RDF 里词汇集被看作为一种资源，能够通过用 URI 标记来表示其带有的属性和与其他词汇的关联。

W3C 制定了 RDF Schema 规范来对 RDF 的词汇集进行相关定义。RDF 表达相关信息的流程如下：首先运用 RDF Schema 支持的原语来建造被表达的资源的相关信息，然后再通过这个模式进行相关资源的表达。这样 RDF 表达的信息就拥有了特殊性，该信息的元数据就可以被计算机进行处理，同时不同应用程序也可以对该数据进行操作。

与 RDF 相关的其他标注语言还包括微格式和 GRDDL（Gleaning Resource Descriptions from Dialects of Languages），微格式（Micro Format）是一个简化的在网站上标注语义数据的语言，它通过使用一套固定的类型提供元数据，对应的属性也是固定的。GRDDL 是一种从 XHTML 与 XML 文件中解析出元数据并把它转成 RDF 图的规范。GRDDL 是把 XML 格式资料转成语义网数据的桥梁。它可以转换数据的格式，也可以再转换至更精确的应用程序。由于 GRDDL 使权威的内容能够动态转换，因而在跨越人类和机器之间的鸿沟方面有着巨大的潜力，与此同时，它为机器提供按需转换内容的机制，并且不会创建替代数据表示的永久版本。

1.3.4 Ontology

本体是共享概念明确的形式化表示。本体拥有很强的表达语义的功能，因为它可以对某一领域内的知识和各种知识之间的联系进行相关的

此为试读，需要完整PDF请访问：www.ertongbook.com