面向云出版的语义关键技术

田萍芳 著

WUHAN UNIVERSITY PRESS
武汉大学出版社

面向云出版的语义关键技术

田萍芳 著

WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

面向云出版的语义关键技术/田萍芳著. —武汉:武汉大学出版社,2015.4
ISBN 978-7-307-15548-0

Ⅰ.面… Ⅱ.田… Ⅲ.语义网络—应用—出版工作—研究
Ⅳ.G230.7

中国版本图书馆 CIP 数据核字(2015)第 066607 号

责任编辑:辛 凯 责任校对:汪欣怡 版式设计: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:湖北恒泰印务有限公司
开本:787×1092 1/16 印张:7.25 字数:170 千字 插页:2
版次:2015 年 4 月第 1 版 2015 年 4 月第 1 次印刷
ISBN 978-7-307-15548-0 定价:25.00 元

前　言

博客、微博/Twitter 等社交媒体及移动互联网络的发展，不仅改变了人们的阅读习惯，而且也正引发出版领域的一场新的技术革命。云出版是目前日益受到关注的发展方向之一，它被国外的出版从业者认为是出版领域科技创新的集大成者。云出版是以云计算及虚拟化技术为基础，以出版业按需付费为商业模式，具备弹性扩展、动态分配和资源共享等特点的数字内容出版、交易及管理模式。作为一种新兴技术和商业模式，出版云将加速出版业信息产业和信息基础设施的服务化进程，催生大量新型出版业信息服务，带动出版产业格局的整体变革。

以语义技术为代表的知识管理技术和大数据存储与分析技术将是云出版依赖的基础性技术，这方面的典型代表是 BBC 的“动态语义发布技术”，它采用基于 RDF 的语义网技术和基于 NoSQL 的大数据存储与分析技术，改变了传统的新闻报道模式，并于 2010 年足球世界杯的报道中进行了有益的尝试，它采用一种基于语义的动态发布框架，精选 700 多个网页中有关运动员的报道，动态输出“自动化元数据驱动网页”，提高用户的访问体验；在 2012 年伦敦奥运会的报道中，该项技术得到了进一步增强，已经能够自动生成每一名运动员(超过 1 万名)、200 个运动队的相关新闻报道。国内出版界也开始进行了相关有益尝试，武汉大学和中华书局联合进行了二十四史的知识图谱构建的尝试，舟山图书馆和华东理工大学联合构建了海洋主题图书馆的知识图谱。

本书将我们近年来在基于语义的云出版领域的相关工作作了介绍，重点介绍面向云出版的语义网关键技术，包括支持出版云的知识图谱构建，海量出版资源表示、存储及获取等方面的技术。本书主要分为 7 个章节，第 1、2 章介绍了语义网的基础知识及知识图谱的基础知识，在此基础上介绍了一种出版物的垂直领域知识图谱构建方法；由于教育类出版资源在整个出版领域中占有非常重要的比重，而且教育类资源的专业性和准确性要求更高，因此本书在第 3 章重点介绍了教育类出版资源知识图谱构建的基本方法；本书第 4 章至第 6 章重点介绍了海量出版资源表示、存储与查询优化方面的相关技术；第 7 章在第 2 章构建的出版物知识图谱的基础上介绍了一种出版资源推荐新技术。

本书的出版得到了武汉科技大学计算机学院各位领导的支持和帮助，也离不开实验室研究生祝中华、邹玉薇、陈凤娇、徐芳芳、熊力、范玉玲、王凯东、彭彬、彭燊、许磊、康恒、田雨晴、龚宇、董豪等人的支持和帮助，他们为本书的内容贡献了相关的素材和实验数据与评估结果。本书的研究内容也得到了国家自然科学基金(编号：60803160，61100133，61272110)、国家社会科学基金重大计划(编号：11&ZD189)、湖北省自然科学基金计划(编号：2013CFB334)、湖北省教育厅科研项目(编号：Q20101110，D2009110)、武汉市科技局关键技术攻关计划(编号：2013010602010216)、湖北省高等学校优秀中青

年科技创新团队计划(编号：T201202)、武汉大学软件工程国家重点实验室开放基金(编号：SKLSE2012-09-07)、湖北省教育厅教研项目(编号：2011s005)的部分资助，在此一并感谢。

编　者

2015年3月

目　录

第1章　语义网基础

1.1　语义网概要

万维网的缔造者，蒂姆·伯纳斯-李(Tim Berners-Lee)于1998年提出语义网(Semantic Web)的概念。指出语义网是数据的网络(Web of Data)，它不仅仅用于人与人之间的交流，而且机器们也能够参与和帮助人与人之间的通信[1]。也就是说，语义网是在搭建一个桥梁，使人与人之间，人与互联网之间相互沟通，并且在跨应用、跨组织和跨平台上，以各自不同的形式实时的共享和复用各自产生的数据，即共用具有语义信息的元数据(metadata[2]，数据的数据)，让万物互联互理解。

2001年5月，蒂姆·伯纳斯·李在《科学美国人》(*Scientific American*)上发表的"*The Semantic Web*"一文中介绍语义网是对现有网络的一个扩展和延伸，它赋予信息良好的语义知识，便于计算机和人更好地协同工作[3]。

对于语义网来说，目前处于整个互联网的发展阶梯的第三个阶段，如图1.1所示。其中，Web 1.0是信息的网络(Web of Information)，主要是浏览信息，只有少量人员发布信息；Web 2.0是文档的网络(Web of Documents)，用于人之间的连接，有着更多的人参与

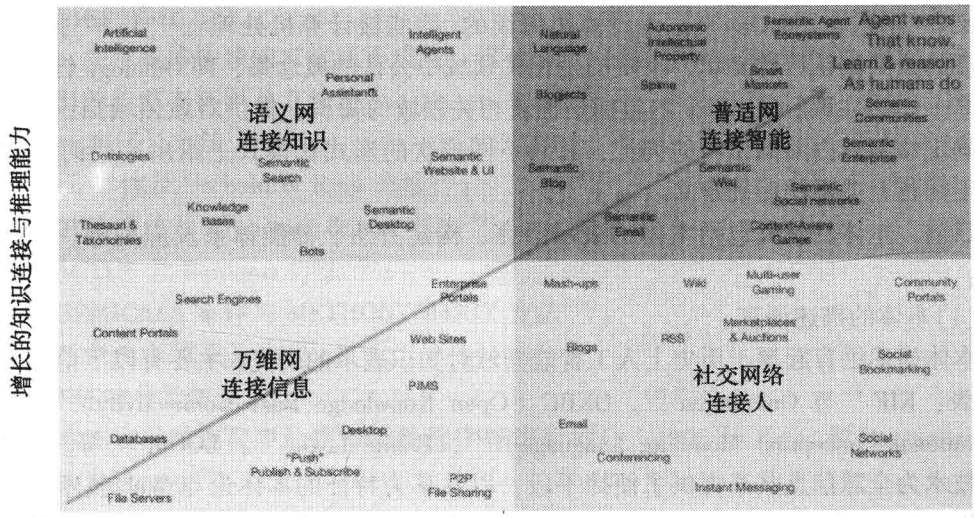

图1.1　互联网发展趋势图

到信息的发布和浏览；Web 3.0 是数据的网络(Web of Data)，将知识链接在一起，让机器参与人类之间的沟通；Web 4.0 是连接智能的网络(Web of Agents)，可以像人一样进行学习和推理。我们正处于 Web 2.0 朝着 Web 3.0 发展的一个阶段，使用语义网技术(如元数据描述语言(如 RDF/RDF 等)，本体描述语言(如 OWL 等)，开放数据链接(Linked Open Data，LOD)，语义网查询语言(SPARQL)，逻辑描述和推理技术等)将知识互联，通往 Web 2.0 的下一代人与机器互联的网络——Web 3.0。

1.2 本　　体

起源于哲学的本体论(Ontology) 近年来受到信息科学领域的广泛关注[4]，其重要性也已在许多方面表现出来并得到广泛认同。尤其最近本体论在 Web 上的应用导致了语义 Web[5]的诞生，在 W3C 的主导下有望解决 Web 信息共享时的语义问题，从而实现世界范围内的知识共享和智能信息集成。

在物质世界和精神世界中，存在着各种各样的事物。人们在长期的实践中，会逐渐形成对这些事物的认知，并形成概念、事实和规则等，这些知识形成后，往往以语言、文字或图形等方式予以表示，以满足人与人之间交流的需要，并进一步地进行知识的积累。1993 年，Gruber 给出了 Ontology 的一个最为流行的定义[6]即“Ontology 是概念模型的明确的规范说明”。后来，Borst 在此基础上，给出了 Ontology 的另一种定义[7]“Ontology 是共享概念模型的形式化规范说明”。Studer 等对上述两个定义进行了深入的研究，认为 Ontology 是共享概念模型的明确的形式化规范说明。这包含四层含义[8]：概念模型(Conceptualization)、明确(Explicit)、形式化(Formal) 和共享(Share)。“概念模型”指通过抽象出客观世界中一些现象(Phenomenon) 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。“明确”指所使用的概念及使用这些概念的约束都有明确的定义。“形式化”指 Ontology 是计算机可读的(即能被计算机处理)。“共享”指 Ontology 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 Ontology 针对的是团体而非个体的共识。Ontology 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(术语)和词汇间相互关系的明确定义。

目前，本体的研究包括本体的表示语言、构建方法、分类体系及应用等。详细介绍如下：

(1)本体的描述语言。

本体描述语言起源于历史上人工智能领域对知识表示的研究，主要有以下语言或环境为代表：KIF[9] 与 Ontolingua[10]，OKBC (Open Knowledge Base Connectivity)[11]，OCML (Operational Conceptual Modeling Language)[12]，Frame Logic[13]，LOOM[14] 等。近年来，Web 技术为全球信息共享提供了便捷手段，以共享为特征的本体论与 Web 技术结合是必然趋势。在此背景下，基于 Web 标准的本体描述语言(以下简称为“Web 本体语言”) 正成为本体论研究和应用的热点，如 SHOE (Simple HTML Ontology Extension)[15]，OML (Ontology Markup Language)[16]，XOL (XML Based Ontology Exchange Language)[17]等。

在标准方面，由 W3C 主持制定的 RDF(Resource Description Framework)[18]和 RDF Schema[19]是建立在 XML 语法上，以语义网(Semantic Networks)为理论基础，对信息资源进行语义描述的语言规范。RDF 采用“资源”(Resources)、“属性”(Properties)以及“声明”(Statements)等三元组来描述事物。RDF Schema 则做进一步扩展，采用了类似框架的方式，通过添加 rdfs：Class，rdfs：subClassOf，rdfs：subPropertyOf，rdfs：domain，rdfs：range 等原语，对类、父子类、父子属性以及属性的定义域和值域等进行定义和表达。这样，RDF (S) 成为一个能对本体进行初步描述的标准语言。

描述逻辑(Description Logics，DL)[20]是近 20 多年来人工智能领域研究和开发的一个相当重要的知识表示语言，目前正被积极应用于本体描述，或者作为其他本体描述语言的基础。最近几个主要的 Web 本体语言 CKML，OIL[21]，DAML + OIL[22]以及已成为 W3C 国际标准的 OWL (Ontology Web Language)[23]就是建立在描述逻辑的基础上。本体描述语言的演变可以用图 1.2 描述。

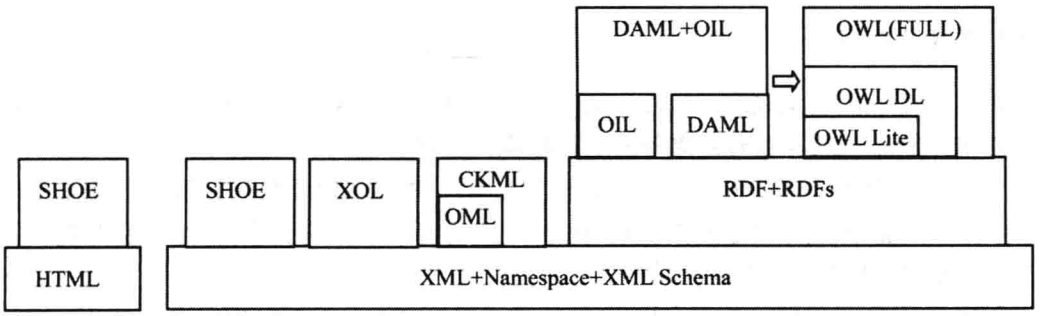

图 1.2 本体描述语言的层次

(2)本体的构建方法。

目前，关于本体构建的方法还不成熟，没有一套完整的统一的方法论。下面列举一些已经开发出的典型本体以及方法论：

① Cyc 本体及方法[24]；

② 企业本体及 Uschold & King 方法[5]；

③ TOVE 本体及 Grüninger & Fox 方法[5]；

④ KACTUS 及 Bernaras 方法[25]；

⑤ CHEMICALS 本体与 METHONTOLOGY 方法[26]；

⑥ SENSUS 本体及方法[27]。

(3)本体的分类体系。

根据本体表示详细程序及领域依赖程序等进行划分，1999 年 Perez 和 Benjamins 归纳出了 10 种 Ontologies[28]：

知识表示 Ontologies

普通 Ontologies

顶级 Ontologies

元(核心)Ontologies
领域 Ontologies
语言 Ontologies
任务 Ontologies
领域-任务 Ontologies
方法 Ontologies
应用 Ontologies
(4)本体的应用。

经过多年的研究，人们已经逐渐将本体论应用到人及组织之间的交流，系统之间的互操作，软件工程等领域。

1.3　基于 RDF(S)的描述

1.3.1　RDF(S)简介

RDF(Resource Describing Framework)[18][19]是 W3C 于 1999 年颁布的一个因特网建议。它的功能是利用当前存在着的多种元数据标准来描述各种网络资源，形成人机可读，并可以由计算机自动处理的文件。RDF 的目标是建立一个供多种元数据标准共存的框架。在这个框架中，能够充分利用各种元数据的优势，“并能够进行基于 Web 的数据交换和再利用”。因此，RDF 的关键是框架结构。

RDF 框架由三个部分组成：RDF Data Model、RDFSchema 和 RDF Syntax。DataModel 形成对资源的形式描述；Schema 定义描述资源时需要的属性类及其意义、特性；Syntax 则把形式描述通过其宿主语言 XML 转换成机器可以理解和处理的文件。

资源：所有在 Web 上被命名、具有 URI(Unified Resource Identifier，统一资源描述符)的东西。比如，网页、XML 文档中的元素等资源可以是 Web 的一个页面，也可以是不能通过 Web 直接访问的对象，如一本书。

属性：属性用于描述资源的特定方面，特征，属性和关系。每个属性具有特定的含义，定义其允许值，可描述的资源类型，其他属性的关系。

声明：一个特定的资源加上该资源命名的属性及属性的值构成一个 RDF 声明。声明的这三个独立部分分别称为主体、谓词和客体，声明的客体可以是另一个资源或者文字。

1.3.2　RDF 的语法特点

RDF 资料模型只是一个抽象与概念的框架，要真的能够承载或交换元数据，需要通过具体的语法。RDF 以 XML 作为编码与传输的语法，此外，RDF 也需要透过 XML 的名称空间(Namespace)来指定宣告属性(Property)词汇的模式(Schema)。RDF 规格提供了两种 XML 语法来对 RDF 资料模型进行编码，第一种称为序列语法(Serialization syntax)，是以正规的方式来表达完整的 RDF 资料模型；第二种称为简略语法(Abbreviated Syntax)，是以较精简的方式来表达 RDF 资料模型的一部分，理想的状况是希望 RDF 解释器(Interpreter)

能够支持这两种语法，让 Metadata 的作者能自由混合使用。

1.3.3 RDF 的容器(Container)机制

我们除了描述单一的资源，有时也需要描述一群的资源，比如说，某个新闻组(News Group)可能包含了许多成员，某本书可能有许多个作者，某个软件可能有许多个下载地址。RDF 容器(Container)就是用来包装或装载一群资源的机制，RDF 定义了三种形态的容器：

- 封装(Bag)：用来包装一群没有顺序性的资源。Bag 通常用在一个属性(Property)有多个值(Value)，而这几个值的先后顺序并不重要，如通信录可能包含了许多姓名。Bag 所包含的值要在 0 个以上，也就是可以不包含值，也可以有多个重复的值。
- 顺序(Sequence)：用来包装一群有顺序性的资源。Sequence 通常用在一个 Property 有多个值，而这些值的先后顺序是重要的，如一本书如果作者在一个以上，则可能有必要区分出主要作者、次要作者。Sequence 所包含的值要在 0 个以上，也就是可以不包含值，也可以有多个重复的值。
- 选择(Alternative)：Alternative 通常用在一个 Property 有多个值可以选择，如某个软件可能提供许多个下载网址。Alternative 所包含的值要在一个以上，而第一个值是预设值。

1.3.4 RDF 模式(Schema)

综上所述，RDF 数据模型，就命名属性和值而言，为描述资源间相互关系定义了一种简单的模型。可以认为 RDF 属性是资源的属性，对应于传统的属性与其值的组合，RDF 属性同样代表了资源间的关系。因此，RDF 数据模型和实体—关系模型类似。但是，RDF 数据模型没有提供机制来说明这些属性，也没有提供机制来定义这些属性和其他资源间的关系。RDF 模式用于完成这些任务。

RDF Schema 的作用就像是一部辞典，宣布一组词汇，也就是在 RDF Statement 中可以使用的 Properties，并描述每个 Property 的意义、特性，以及 Property Value 的限制。

RDF Schema 可以是为了让人阅读的描述，也可以是机器可以处理的表示法，如果是后者，则应用程序便可以直接透过 RDF Schema 来了解每个 Property 的意义，并作自动化处理。机器可以处理的 RDF Schema 也是以 RDF 资料模型为基础的。

- 核心类

下述资源是作为 RDF 模式词汇一部分的核心类。每一个运用 RDF 模式名字空间的 RDF 模式名字空间的 RDF 模型都(隐含的)包含它们。

RDFS：Resource：由 RDF 表达式所描述的所有东西都被称为资源，并被认为是 rdfs：Resource 类的实例。

RDF：Property：表示称为属性的 RDF 资源的子集。

RDFS：Class：它对应于一般的类型或分类的概念，与面向语言中的类相似。

- 核心属性

RDF：Type：它指示资源是类的一个成员，因此具有类的成员所希望具有的所有特征。当一个资源具有这个属性，并且其值是某些特定的类时，就认为资源是指定类的一个实例。

RDFS：SubClassOf：这一个属性说明类间的一个子类/超类关系。这个属性是可以传递的。比如说，如果 A 是 B 的子类，B 是 C 的子类，那么 A 是 C 的子类。

RDFS：SubPropertyOf：这个属性是 rdfs：Property 的一个实例，它用于说明一个属性是另一个属性的特殊化。一个属性可能是零个、一个或多个属性的特殊化。

RDFS：SeeAlso：这个属性说明一个资源可能提供关于主题资源的附加信息。

RDFS：IsDefineBy：这个属性是 RDFS：SeeAlso 的子属性，指示了定义主体资源的资源。

- 约束

RDF 模式规范引入了一个 RDF 词汇来声明使用 RDF 数据中属性和类的约束。比如说，一个 RDF 模式可能描述对属性值类型的限制，让其对某一属性有效；或者描述对垒的属性的限制，对这些类这些属性是有意义的。

约束的一些例子包括：

一个属性的值应当是一个指定类的资源，这称为一个指定类的资源，这被称为一个范围(range)约束。

一个属性可能用在某一类资源上，这被称为领域(domain)约束。

下面来用一个例子来说明 RDF 的使用，如图 1.3 所示，在这个例子中我们主要使用了 subclass 属性，说明了 MotorVehicle，Truck，PassengerVehicle 之间的关系。

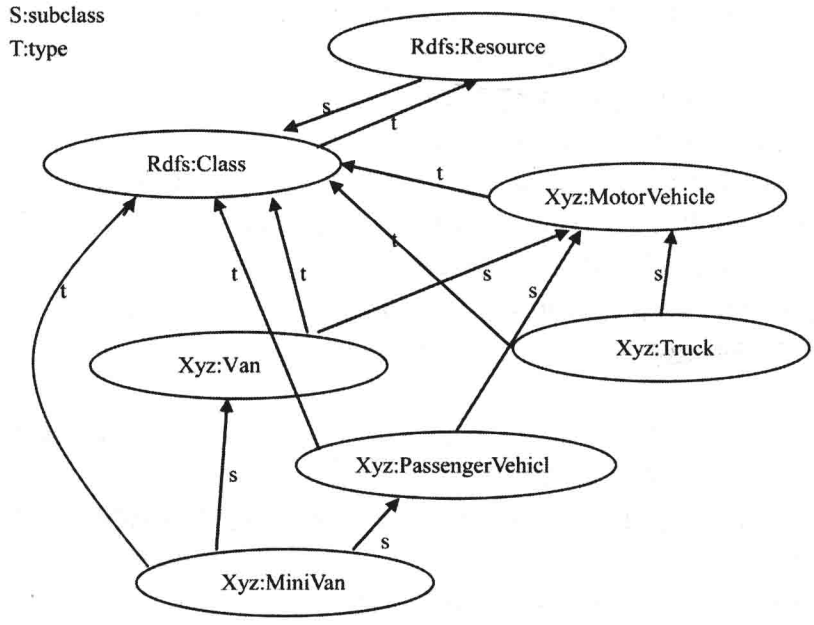

图 1.3　RDF 示例

以下是此示例的 RDF 代码：

```
rdf: RDF xml: lang="en"
    xmlns: rdf="http: //www. w3. org/1999/02/22-rdf-syntax-ns#"
    xmlns: rdfs="http: //www. w3. org/2000/01/rdf-schema#" >
<! -- Note: this RDF schema would typically be used in RDF instance data
      by referencing it with an XML namespace declaration, for example
      xmlns: xyz="http: //www. w3. org/2000/03/example/vehicles#". This allows
      us to use abbreviations such as xyz: MotorVehicle to refer
      unambiguously to the RDF class 'MotorVehicle'. -->
<rdf: Description ID="MotorVehicle" >
 <rdf: type resource="http: //www. w3. org/2000/01/rdf-schema#Class" />
 <rdfs: subClassOf
  rdf: resource="http: //www. w3. org/2000/01/rdf-schema#Resource" />
</rdf: Description>
<rdf: Description ID="PassengerVehicle" >
 <rdf: type resource="http: //www. w3. org/2000/01/rdf-schema#Class" />
 <rdfs: subClassOf rdf: resource="#MotorVehicle" />
</rdf: Description>
<rdf: Description ID="Truck" >
 <rdf: type resource="http: //www. w3. org/2000/01/rdf-schema#Class" />
 <rdfs: subClassOf rdf: resource="#MotorVehicle" />
</rdf: Description>
<rdf: Description ID="Van" >
 <rdf: type resource="http: //www. w3. org/2000/01/rdf-schema#Class" />
 <rdfs: subClassOf rdf: resource="#MotorVehicle" />
</rdf: Description>
<rdf: Description ID="MiniVan" >
 <rdf: type resource="http: //www. w3. org/2000/01/rdf-schema#Class" />
 <rdfs: subClassOf rdf: resource="#Van" />
 <rdfs: subClassOf rdf: resource="#PassengerVehicle" />
</rdf: Description>
</rdf: RDF>
```

从上面的例子我们可以看出 RDF(S)来描述 ontology 具有很多优点，它比 XML 具有更丰富的语义，从描述的结构上来看，它不仅描述了层次关系，而且还可以描述不同资源之间的关系，这样就使描述的图结构的边具有了语义，而不像 XML 描述的图那样，边是没有意义的。另外，RDF 还可以描述资源之间的约束。这种强大的功能更适合描述 ontology。

(1)简单。RDF 使用简单的资源—属性—值三元组，所以很容易控制，即使是数量很

大的时候。这个特点很重要，因为现在资源越来越多，如果用来描述资源的元数据格式太复杂，则势必会大大降低元数据的使用效率。其实从功能的角度来看，完全可以直接使用 XML 来描述资源，但 XML 结构比较复杂，允许复杂嵌套，不容易进行控制。采用 RDF 可以提高资源检索和管理的效率，从而真正发挥元数据的功用。

(2)易扩展。在使用 RDF 描述资源的时候，词汇集和资源描述是分开的，所以可以很容易扩展。例如，如果要增加描述资源的属性，则只需要在词汇集中增加相应元数据即可，而如果使用的是关系数据库，则增加新字段有可能造成大量的空间浪费。

(3)开放性。RDF 允许任何人定义自己的词汇集，并可以无缝地使用多种词汇集来描述资源，以根据需要来使用，使各尽所能。比如，在上个例子里，描述网页资源时用 Dublin Core 描述其作者属性，而在描述作者的姓名时又使用了另一个专门描述人的词汇集来描述。

(4)易交换。RDF 使用 XML 语法，可以很容易地在网络上实现数据交换。另外，RDF Schema 定义了描述词汇集的方法，可以在不同词汇集间通过指定元数据关系来实现含义理解层次上的数据交换。

(5)易综合。在 RDF 中，资源的属性是资源，属性值可以是资源，关于资源的陈述也可以是资源，都可以用 RDF 来描述。这样就可以很容易地将多个描述综合，以达到发现知识的目的。例如，在描述某书籍时指明其作者属性值是另一资源，我们就可以根据描述作者的 URI 来获得作者的信息，如毕业院校等，从而知道这本书是某一院校的毕业生写的，于是，在表面上来看，没任何关系的两者之间建立了联系，而不需要任何人工的干预。

第 2 章　构建出版资源知识图谱

2.1　知识图谱的基本概念

2.1.1　什么是知识图谱

知识图谱(Knowledge Graph)由 Google 在 2012 年提出，用于 Google 搜索引擎上面的技术，以提高 Google 搜索的质量。Google 的阿米特·辛格尔(Amit Singhal)在介绍知识图谱时说："'图'能够理解真实世界中的实体和它们的关系以及实体间的关系：是实体，而不是字符串。(a 'graph' —that understands real-world entities and their relationships and their relationships to one another: things, not strings.)"[29]知识图谱是用来描述真实世界中存在的各种实体(概念)以及实体间的关系(属性)。

对于学术界来说，知识图谱对应的概念是链接数据(Linked Data[30]或者 the Web of Linked Data)。链接数据的思想是蒂姆·伯纳斯-李于 2006 年提出来，它需要使用语义网的技术和标准来发布。链接数据需要从以下 4 个方面来理解：

(1)以机器可理解的方式(如用 RDF 语言来描述数据)发布到网络上的数据；

(2)数据含义是精确定义的；

(3)可以链接其他外部数据集；

(4)同时也能够被外部数据集关联。

概括来说，链接数据是指用于发布和关联网络上的结构化数据(RDF 数据)的集合。如果需要进一步构建链接数据，则需要满足蒂姆·伯纳斯-李在 2006 年提出来的四条基本规则，如下[30]：

(1)规则 1：使用 URIs 定义事物；

(2)规则 2：使用 HTTP URIs，因此客户端(机器或者人类阅读器)能够查找这些名称；

(3)规则 3：当查找一个 URI，需要提供有用的且可理解的信息；

(4)规则 4：链接其他数据，有利于查询到更多的数据和想要的信息。

第一条规则用于指定资源或者概念的全球唯一性。第二条规则是第一条规则的约束条件和补充说明。规则 3 进一步强调了规则 2 的作用，即当要解析一个 HTTP URIs 时，需要返回给客户端一些有用的数据信息。最后一条规则是确保链接数据的发展，只要将数据相互关联，扩大链接数据规模，数据的网络才能逐渐走入正轨。

近几年以来，随着语义网的不断发展，大量来自不同领域的资源描述框架(Resource Description Framework，RDF)数据开始被发布，开放链接数据集在不断扩大。图 2.1 是开

放数据链接云图(不同数据集之间链接关系概览)。据统计，开放链接数据集从 2011 年(294 个数据集)增长到 2014 年(1091 个数据集)，如表 2.1 所示[31]。

可见，语义网技术和标准的不断成熟，有利于链接数据(也可以说是知识图谱)在实时网络中成长，不断接近数据的网络。

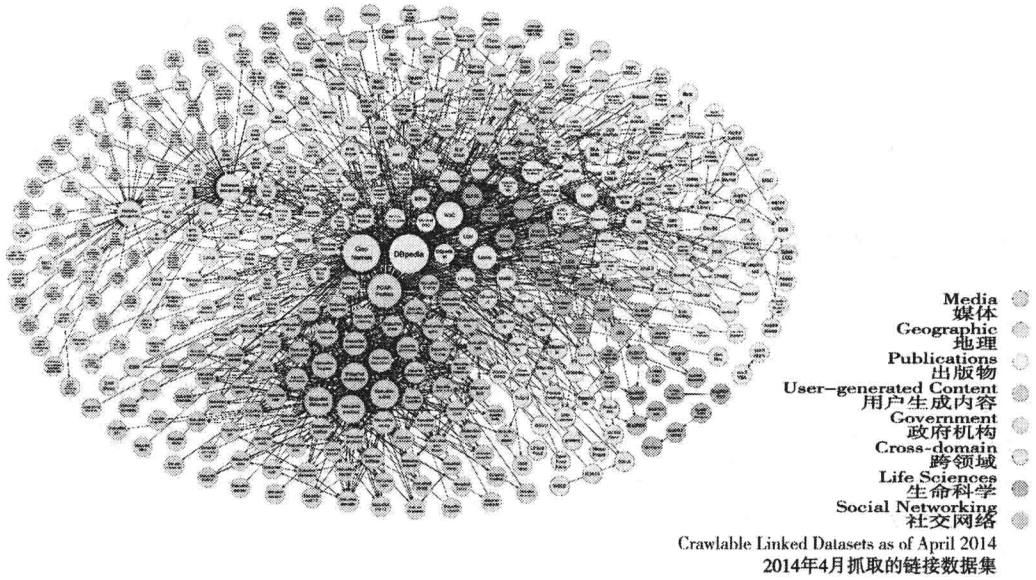

图 2.1　LOD 云图

表 2.1　　**2014 年和 2011 年数据集的数量和增长百分比比较**

Category 类　别	Dataset 2014 2014 年数据集	Percentage 百分比	Datasets 2011 2011 年数据集	Growth 增长百分比
Media 媒体	24 (-2)	2%	25	-4%
Government 政府机构	199 (-16)	18%	49	306%
Publications 出版物	138 (-42)	13%	87	59%
Geographic 地理	27 (-6)	2%	31	-13%
Life Sciences 生命科学	85 (-2)	8%	41	107%
Cross-domain 跨领域	47 (-6)	4%	41	15%
User-generated Content 用户生成内容	51 (-3)	5%	20	155%
Social Networking 社交网络	520 (-0)	48%	—	—
Total 合计	1091 (-77)		294	271%

2.1.2　知识图谱的表示方法

上一节中介绍的知识图谱是指描述现实世界中的各种实体和实体间的关系。那么用什