



关联数据及其应用现状综述

沈志宏^{1,2,3} 张晓林¹

¹(中国科学院国家科学图书馆 北京 100190)

²(中国科学院计算机网络信息中心 北京 100190)

³(中国科学院研究生院 北京 100049)

【摘要】自主、异类、异构和分布的海量数据间的语义关联一直是知识组织和知识发现的核心问题,关联数据(Linked Data)技术的出现为此提供一种轻型、渐增化、可伸缩和可扩展的动态机制。基于对大量文献的梳理和分析,描述 Linked Data 的原则与技术体系,介绍 Linked Data 发布、Link 浏览器、Linked Data 搜索引擎、Link 更新维护等核心技术,简要介绍若干典型应用,提出在具体领域实施 Linked Data 应用的关键挑战。

【关键词】关联数据 关联开放数据 语义网 数据网络

【分类号】TP393

Linked Data and Its Applications: An Overview

Shen Zhihong^{1,2,3} Zhang Xiaolin¹

¹(National Science Library, Chinese Academy of Sciences, Beijing 100190, China)

²(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

³(Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

【Abstract】Semantic linking of multi genre, multi typed, autonomous, and distributed data is always in the center of knowledge organization and discovery, and the concept of linked data provides a light-weight, incremental, scalable and extensible mechanism. Based on systematic review of the field, the paper describes the four rules and the basic technical framework of linked data, presents the key techniques enabling linked data publishing, linking parsing and browsing, linked data search engines, link update and maintenance. A few typical applications are given and key challenges in applying linked data in a practical domain are also explored.

【Keywords】Linked data Linking open data Semantic Web Web of data

1 问题的提出

互联网日益丰富的海量信息资源具有异类(即具有不同内容类型和体裁,例如小说、戏剧、论文、专著、数据记录、Twitter 消息、财务报表、博客文章等)、异构(即具有不同的数据格式及相应的语义规则)和分布的特点,在这样的环境下,如何根据信息内容中的知识对象、知识关系连接和集成相互关联的内容,从而支持基于网络的知识组织和知识发现,一直是互联网信息服务希望解决的问题,也是互联网信息革命需要跨越的核心门槛之一。

对此,人们从多个角度进行了努力。除了元搜索技术^[1]、集成数据平台^[2-4]外,WebService 技术^[5]为网络信

图 1 展示了 LOD 社区发布的最新的 LOD 数据云图,其中收录了很多知名的数据集,如 DBpedia、DBLP Bibliography、GeoNames、Revyu、Riese、UMBEL、Sensorpe-

dia、FOAF、DOAP、OpenPSI、MusicBrainz 等,这些数据集涉及地理、生命科学、医药、出版、媒体、社会网络等领域。表 1 列出了一些常见的 LOD 数据集。

表 1 LOD 数据集列表

数据集名称	数据内容	元数据格式	网址
Bio2RDF	27 个生物、基因与医疗数据集	BioPAX ChEBI Ontology Ensembl Ontology KEGG Ontology UniProt	http://bio2rdf.wiki.sourceforge.net/
DBLP Berlin	计算机科学论文与作者编目,包含 80 万篇文章,40 万作者	Dublin Core FOAF Vocabulary	http://www4.wiwiw.fu-berlin.de/dblp/
DBpedia	Linked Data 化的 Wikipedia		http://dbpedia.org/
DrugBank	5 000 个美国 FDA 批准的小分子和生物技术药物	DrugBank Vocabulary	http://www4.wiwiw.fu-berlin.de/drugbank/
FOAF	人群(Friend Of A Friend)	FOAF Vocabulary	http://www.foaf-project.org/
GeoNames	地名(8 亿条)	GeoNames Ontology	http://www.geonames.org/
Linked MDB	电影		http://www.linkedmdb.org/
MusicBrainz	艺术家相册	FOAF Vocabulary	http://dbtune.org/musicbrainz/
OpenCyc	Linked Data 化的 OpenCyc	Open Cyc Ontology	http://sw.opencyc.org/
RDF Book Mashup	图书编目数据,来源于 Amazon、Google、Yahoo	Dublin Core GoodRelations Ontology	http://www4.wiwiw.fu-berlin.de/bizer/bookmashup/
UMBEL	2 万个主题概念,来源于 Wikipedia	UMBEL Ontology	http://www.umbel.org/
YAGO	200 万条人员、组织、城市等实体	YAGO Ontology	http://www.mpi-inf.mpg.de/yago-na-ga/yago/

实际上,Linked Data 并不仅仅适用于开放数据(Open Data)。除 LOD 之外,由 Zepheira Team 主导的连接企业数据(Linking Enterprise Data,LED)^[16]则重点关注企业的数据。与开放数据相比,企业数据更关注一些额外的限制条件,如法律法规遵从情况以及企业优势等。

3 Linked Data 原则与技术体系

不同于 Document Web,Linked Data 将互联网上任一信息内容或其子内容看成是一个可采用标准方法规范描述和调用的知识对象,通过创建和发布关于各类知识对象及其与各类其他知识对象之间关系的规范化描述信息,通过建立基于知识内容的检索以及基于知识关系的分析关联机制,Linked Data 便可支持在特定信息环境下对不同知识对象的关联发现。

Linked Data 制订了关于内容对象的描述原则:

(1)使用 URI 来标识事物(Use URIs as names for things);

(2)使用 HTTP URI 使人们可以访问到这些标识(Use HTTP URIs so that people can look up those

names);

(3)当有人访问到标识时,提供有用的信息(When someone looks up a name, provide useful information);

(4)尽可能提供关联的 URI,以使人们可以发现更多的事物(Include links to other URIs so that they can discover more things)。

前两条原则分别建立规范化命名机制和调用内容对象的机制,第三条原则要求用结构化、规范化方式来描述内容对象,第四条原则要求建立内容对象与其他内容对象的关联,以支持从内容对象出发对相关内容的关联检索。这些原则并没有对内容对象的内部组织机制、系统调用接口、关联解析机制等提出具体要求,因此人们可以使用多种方式来实现关联检索,这使得 Linked Data 成为一种普适的、轻量的、低成本的数据关联机制。

这组原则及其实现要求奠定了 Linked Data 的整体技术体系。本文认为,这个体系包括 4 个部分,如图 2 所示。

(1)“Linked Data 数据内容组织”可以是任意数据组织形式下的具体数据内容,但要求能用 URI 予以标



图2 Linked Data 技术体系

识、能通过 HTTP 调用,例如 <http://qinghailake.csd.cn/res/river/buha>。这个 URI 可用来指代青海湖基础数据集中的布哈河,但由于这个 HTTP URI 本身只是一个抽象的概念,Linked Data 发布时常常需要采用内容协商^[17]实现从这个 URI 到实际数据记录的映射,以返回实际的内容(HTML 网页、图像、用 RDF/XML 表示的文本数据等)。

(2)“Linked Data 创建与发布”是实现 Linked Data 的关键环节,需满足以下三个要求:

①用规范形式(通常是 RDF)来描述内容对象的内部结构及其内含内容的语义(例如一本专著的标题、作者、出版社、主题、章节题名、内含图表、内含数据集、内含软件等),这个描述的深度取决于内容对象本身的内容深度和所依据元数据格式(Metadata Formats)的丰富程度,描述的结果是基于元数据格式转换的 RDF 记录。

②在上述描述信息里建立这个内容对象与其他内容对象的关联描述(Includes Links to Other URIs),与其他内容对象不同的关联关系则需要根据不同的关联种类来分析和确定,这属于整个 Linked Data 体系中最富有挑战性的工作。只有建立了这种关联描述,所创建和发布的才是 Linked Data。

③以某种系统机制存储所建立的 RDF 描述信息,并支持以公共方法进行检索和解析。

(3)“Linked Data 浏览与检索”要求能用标准方法检索和解析 RDF 信息(通常采用 SPARQL),并能挖掘和利用 RDF 描述中丰富的结构与关系信息来支持对其他内容对象的基于知识点和知识关系的检索,及在此基础上对复杂关联的多个内容对象的结构与关系表达。人们往往通过支持 Linked Data 的浏览器来实现检索与关联。在结构与关系比较复杂的情况下(例如 UMLS Relations Network 所支持的复杂关系),这种检索与关联会面临许多挑战。

(4)“Linked Data 互联与维护”是采用自动或半自动的方法,创建不同数据集数据之间的关联,并在源内

容对象和目标内容对象发生变化时保持关联信息的准确及时。在数据量很大的情况下,自动的关联发现算法很有必要,这一般可通过配置化的规则来生成。一旦某些数据发生变化,其他的数据集有必要及时增加、删除或者修改与之的连接。这通常需要自动的同步机制与同步语言来实现,在满足多个数据源间消息及时传递的同时,还要保证消息的通信不会造成太大的性能上的降低。

4 Linked Data 创建与发布

在 Linked Data 出现之前,人们已经通过各种途径积累了大量的数据资源,如何快速地将现有的数据资源以 Linked Data 的形式发布,成为急需解决的问题。

选择发布 Linked Data 的技术方案需要考虑三个问题^[17]:待发布的数据量的大小(How much data do you want to serve?)、数据的更新频率(How often does your data change?)、以及数据的存储方式(How is your data currently stored?)。如果数据量很小(几百条 RDF 三元组或者更少),可以采用静态的 RDF 文件(静态发布)。如果数据量很大,则需要将它们放进 RDF 库中,并选择 Pubby^[18]等服务器作为 Linked Data 服务的前端。如果数据的更新频率很大,就需要引入更新机制,或在请求数据时再根据原始数据在线生成(On-the-fly Translation) RDF。

对于那些以数据库方式存储的数据内容,可以使用类似于 D2RQ 平台^[19]这样的转换工具(类似的工具还有 Virtuoso Universal Server^[20]、Triplify^[21]等)来发布。很多数据集(如 DrugBank、Linked MDB、MusicBrainz 等)都采用了 D2RQ 作为发布平台。D2RQ 平台由 D2R 引擎、D2R 服务器以及 D2RQ 映射语言组成,如图 3 所示。

D2R 引擎将用户输入的 SPARQL 转换成 SQL 提交给关系型数据库管理系统,并将 SQL 查询结果封装成 RDF 三元组。D2R 服务器基于 D2R 引擎之上提供了 Web 化界面(包括 HTML、RDF 和 SPARQL 界面),通过界面,浏览器用户可以查看数据网页,Linked Data 客户程序可以获取 RDF 数据,SPARQL 客户程序则可以通过 SPARQL 查询获取结果。D2RQ 映射语言^[22]支持从关系型数据库到 RDF 数据转换规则的编写,通过该语言,用户可以定义如何将关系型数据库中的表、

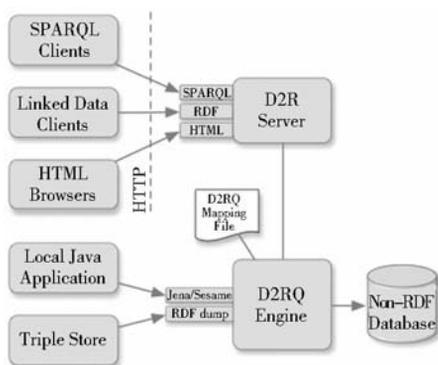


图 3 D2RQ 平台体系结构^[19]

列、行、列值、外键值映射成 RDF 数据中的类、属性、资源、文本、RDF 链接。手动编写映射文件比较麻烦，因此 D2RQ 提供了 Generate - Mapping 脚本支持映射文件的自动生成，用户可以在生成的映射文件(n3 格式)的基础上再进行修改。

如果数据内容可以通过程序化的 API 或者 Web 化的服务接口(如 OAI - PMH 协议)得到，就需要开发一些包装器完成服务到 Linked Data 接口的转换，如：OAI2LOD Server^[23]。对于其他一些通用格式，如：Microsoft Excel、CSV 或者 BibTeX，也有一些 RDFizer 可供选择，如：JPEG2RDF^[24]、MARC/MODS2RDF^[25]、Email2RDF^[26]等。有时候，还需要构造更为复杂的发布框架，以 RDF Book Mashup^[27]为例，它通过远程调用 Amazon API 和 Google Base API 获取图书信息，并包装成 Linked Data 格式。RDF Book Mashup 的技术架构如图 4 所示：

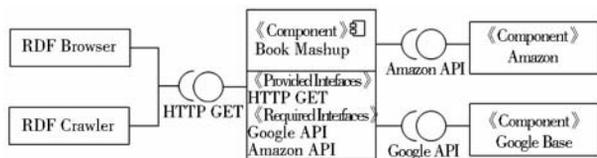


图 4 Book Mashup 数据发布框架^[27]

除以上类型的数据和服务外，还有人在研究如何发布原始数据文件。如文献[28]提出 TripFS，研究如何将文件系统以 Linked Data 接口暴露出来。文献[29]介绍 IM (Interlinking Multimedia)，它将多媒体项的片段发布成 Linked Data。文献[30]则研究如何将实时的流数据发布出来。

在 Linked Data 创建与发布的过程中，还存在一些挑战，如：词表的定义和选择、复杂数据(如化学结构

式)的结构化(RDF 化)，以及数据集内数据关联的自动生成等。

5 Linked Data 浏览与检索

当人们纷纷将数据集发布成 Linked Data 之后，就需要合适的浏览器来浏览这些数据，并通过各数据之间的 RDF 链接指引用户在不同的数据源之间导航。常见的 Linked Data 浏览器如表 2 所示：

表 2 Linked Data 浏览器列表

浏览器	开发者	网址
Tabulator RDF Browser	Tim Berners - Lee 等	http://www.w3.org/2005/ajar/tab/
Disco Hyperdata Browser	Freie University Berlin	http://sites.wiwiss.fu-berlin.de/suhl/bizer/ng4j/disco/
OpenLink Data Web Browser	OpenLink	http://demo.openlinksw.com/DAV/JS/rdfbrowser/index.html/
Objectviewer	SemanticWebCentral	http://objectviewer.semweb-central.org/
Marbles		http://marbles.sourceforge.net/

与传统意义上的浏览器不同，Linked Data 浏览器大多数运行在服务器端，为用户提供基于 Web 浏览器的浏览界面，只有少数才提供客户端浏览器插件。此外，还有一些 RDF 浏览器，如：Longwell、MSpace、Facet、BrowseRDF、RDFGravity 以及 IsaViz。但是由于它们不能实现跨数据源的语义网浏览，所以还不能称之为 Linked Data 浏览器。

除浏览器外，人们也可以借助搜索引擎技术来检索所需要的数据。常见的 Linked Data 搜索引擎如表 3 所示：

表 3 Linked Data 搜索引擎列表

搜索引擎	开发者	收录 RDF 文档数	网址
Falcons	IWS China	700 万	http://iws.seu.edu.cn/services/falcons/
Sindice	DERI Ireland	大于 2 000 万	http://www.sindice.com/
Watson	KMi, UK		http://watson.kmi.open.ac.uk/WatsonWUI/
Semantic Web Search Engine (SWSE)	DERI Ireland		http://www.swse.org/
Swoogle	UMBC USA	230 万	http://swoogle.umbc.edu/

Linked Data 浏览与检索技术往往可以结合起来，从而给用户提供更加智能的数据浏览服务。Marbles 是几款 Linked Data 浏览器中比较有特色的一款。Marbles 采用了 Fresnel^[31]与样式单，通过对 RDF 内容的选择、排序及格式化，从而为机器可读(Machine - Readable

ble)的 RDF 数据产生比较友好的人读(Human - Readable)界面,而在这方面,其他浏览器生成浏览界面往往都比较单调。Marbles 浏览器会同时加载来自于多个数据源的数据,并集成在一个页面中显示。在加载指定 URI 的 RDF 资源数据的同时,Marbles 会从 Sindice 与 Falcons 中加载关于该资源的关联信息,并从 Revyu 加载关于该数据的评论,并最终在同一个界面中展示给用户。图 5 是在 Marbles 中浏览 Tim Berners - Lee 信息的情形,其中不同颜色的彩球代表来源于不同的数据源。



图 5 在 Marbles 浏览器中查看 Tim Berners - Lee 的信息

在实现 Linked Data 浏览与检索过程中,还有一些细节问题需要解决,如 RDF 数据的可视化、浏览轨迹的跟踪、海量索引数据的存储与快速检索、数据访问统计与用法评价、检索结果的排名等。

6 Linked Data 互联与维护

Linked Data 鼓励人们在发布 RDF 数据的同时,发布数据之间的 RDF 链接,以帮助人们发现更多的知识。Linked Data 的互联技术研究已成为近年来的研究热点。在 LDOW2010 (WWW2010 Workshop on Linked Data On the Web)中,数据互联是会议的一大专题,其他专题包括 Linked Data 发布、基础设施与架构、Linked Data 应用等。另外,在 COLD2010 (International Workshop on Consuming Linked Data)发布的几大研究热点中,居于首位的就是 Linked Data 的互联算法,其次是溯源与信任、数据集动态、用户界面、分布式查询、评估。

以 MP3 音乐数据为例,可以通过某首音乐的作曲

者姓名,生成指向 MusicBrainz 中关于该作曲者的描述记录的链接,如果再做更深的关联计算,甚至可以根据该音乐的标题,生成指向 DBpedia/YAGO 中引用到该音乐的电影的链接。比较实际的例子如 Linked MDB^[32],它实现了与其他 LOD 数据集如 DBpedia/YAGO、GeoNames、FlickrWrapper、RDF Book Mashup、MusicBrainz、Revyu.com 的互联。Linked MDB 包含实体 233 103 项,指向其他 LOD 数据的关联数为 162 199 项,这种关联如图 6 所示:



图 6 Linked MDB 与其他 LOD 数据集之间的关联^[32]

有一部分研究关注于如何通过一些自动和半自动的方法来发现数据之间的关联,Raimond 等结合音乐数据集^[33],介绍了自动关联的几种途径:基于实体的文本映射(简单文本查找、扩展文本查找)及基于图相似度的映射。基于规则的互联框架 Silk^[34]则允许用户制定 Silk - LSL (Silk Link Specification Language) 规则文件,Silk 框架通过这些规则自动生成不同数据集之间的实例级的链接。Silk 不仅能够生成数据之间的 owl:sameAs 关联,也可以生成其他类型的关联,如:DBpedia 电影与 Linked MDB 导演之间的 dbpedia:director 关联。Silk 主要通过给定的两个数据集中数据的属性相似度来计算它们之间的关联关系。Silk - LSL 支持的相似度算法包括 Jaro 距离、Jaro - Winkler、Levenshtein、Q - grams 文本相似性计算、文本等价性、数值距离、日期距离等算法。

如果数据集实现了互联,一旦某个数据集的内容发生更改,就需要适当的机制及时通知关联的变更。Haslhofer 等提出 DSNotify^[35],用来检测和修复 LOD 数据集之间的断链。PTSW (Ping The Semantic Web)^[36]提供了针对 RDF 数据的在线归档服务,为网络爬虫及其他的软件提供最近更新的 RDF 数据。一旦某个数据源的 RDF 数据发生更改,它就会通过 PTSW 的服务接口来通知该内容的更改,从而保证归档 RDF 数据的一致性。另外,WOD - LMP 协议 (Web Of Data Link Maintenance Protocol)^[37]针对数据网络中的链接同步问题,定义了发送链接至目标、获取目标变化列表 (Change List)、将变化发布给目标三种用例,并详细定义不同消息的具体格式。

在 Linked Data 互联与维护过程中,还有一些其他细节问题需要解决,如 RDF 数据的溯源、基于第三方数据集 (背景数据集) 的互联、数据集关联度的计量、RDF Link 的断链检测问题等。

7 Linked Data 应用

与前面介绍的 Linked Data 通用技术相比,Linked Data 应用专指在不同领域和场景中消费和操作 Linked Data 的应用。关于“Linked Data 应用”这一术语概念,DERI (Digital Enterprise Research Institute) Linked Data 研究中心 (Linked Data Research Centre, LiDRC) 在“Linked Data 应用 - Linked Data 使用的起源与挑战”技术报告中给出了两种不同的解释:

(1)指 Linked Data 在不同领域(包括生物学、统计学、软件工程、多媒体等)中的应用。

(2)指基于 Linked Data 之上构建 Web 应用,这类 Web 应用被称为“由 Linked Data 驱动的 Web 应用 (Linked Data - driven Web Applications)”,主要关注如何在 Web 上消费和操作 Linked Data。

本文将综合采用这两种含义。

Linked Data 已经开始应用到数据融合、语义标注、集成式问答系统、事件管理等多个场合。DBpedia Mobile^[38]是一款用在移动环境中的 Linked Data 应用。DBpedia 实现了 Wikipedia 的 Linked Data 化,描述了其中 218 万多个实体,其中包含了近 300 000 条的地理位置。DBpedia Mobile 充分利用这些位置数据,通过 GPS 信号,结合 DBpedia 与其他 LOD 数据集 (如:

GeoNames、US Census、CIA Factbook 与 EuroStat) 的关联,自动发现当前位置附近的相关内容,并通过 Linked Data 浏览器提供地图化的界面,如图 7 所示:



图 7 DBpedia Mobile 地图化界面^[38]

与此类似,Data.gov 提供了一种基于 Linked Data 的服务,实现针对英国领土内地理实体的导航和获取^[39]。Zhao 等基于 LOD 数据集 Richard Cyganiak 设计了专家发现系统 (Expert Search and Profiling Systems)^[40],结果表明,较之传统的、封闭的、非结构化的数据应用,基于 LOD 的系统能够实现数据与假说之间的解耦,而不受平台的限制。Latif 等利用 Journal of Universal Computer Science (JUCS) 和 DBpedia,实现从开放数字杂志中来发现和构建作者的简介信息^[41],并提供了 CAF - SIAL 用户界面程序。通过 CAF - SIAL,用户可以获取到某一作者的介绍、照片、个人信息、职业信息、学术信息、发表作品等综合性视图,如图 8 所示:

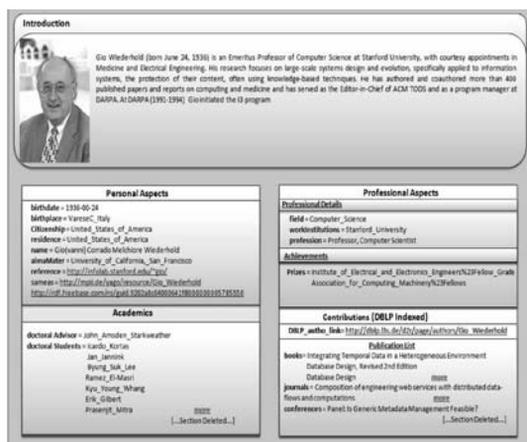


图 8 在 CAF - SIAL 中查看作者的简介信息^[41]

可以看出,随着 Linked Data 理念的传播,Linked Data 已经深入到人们生活中的各个领域,不同来源的数据也因此面临着更广阔的应用场景。

8 结 语

Linked Data 的提出,有效地解决了自主、异类、异构和分布的海量数据的知识发现问题。Linked Data 提供了一套低成本的标准数据访问机制,因其非集中式的特点,采用 Linked Data 机制可以有效地规避一些复杂的数据权益纷争问题,有利于数据的健康发展和信息的繁荣。另一方面,各种完全自治的“数据孤岛”通过 Linked Data 连接起来,形成一个更为全面的浩瀚的知识库,为更智能的知识发现的实现提供条件,同时为上层数据应用(如数据集成检索、数据融合)提供丰富的数据源。

作为一种折中的语义化知识组织与发现机制,Linked Data 主要还是侧重于为不同的知识服务系统之间的互操作定义一套统一的访问接口和 RDF 数据模型,对上层的概念模型以及知识的表达能力尚存不足(这些工作主要还是留给了内容提供商)。如果要完成对内容的检索和浏览,常常需要采取透明的方式,如提交一个 SPARQL 查询或者借助于语义搜索引擎 Robots.txt 和 Sitemap.xml 规范来完成。Linked Data 在用户身份认证、数据访问控制方面没有做更多的标准化工作(目前还没有标准的方法可推荐使用),因此数据访问过程中的统一认证和权限控制还必须由应用服务器来实现,这也大大限制了不同系统之间的互操作性。由于 Linked Data 是分布式的,多个站点之间往往需要一套良好、高效的数据同步机制。除此之外,目前还缺乏一个高效的基于 Linked Data 的分布式查询和按需 Mashup 平台,这方面仍需要继续投入。

Linked Data 目前还是一套标准化的、实用可行的机制。以中国科学院为例,中国科学院积累了丰富的数字资源,包括文献、数据、教育、科普资料等,存在自主、异类、异构和分布的特征,因此基于 Linked Data 来实现内容关联的知识发现在理论上具有可行性。然而,由于机制和数据内容的问题,这样的操作仍存在着巨大的挑战。在现有资源的 Linked Data 化(即 Linked Data 的发布)过程中,需要充分考虑因不同的服务方式(如:文献资源通过文献检索系统,科学数据资源采用数据库管理系统)、不同的分类体系、不同的数据内容、以及不同领域中采用不同的本体及词表所带来的复杂性。在 Linked Data 数据互联的过程中,由于同名异

义、同名异义、同体异构等情况,数据分面复杂、类型多样,再加上数据资源的大数据量,给实体识别与自动互联算法的准确性、高效性提出了较高的要求。除此之外,由于科研活动一直是一个动态的过程,知识在整个科研活动中按照生命周期循环流动,Linked Data 内容的组织与表达需要充分考虑数据及其关联的实时性,解决数据的一致性问题。严谨的科研过程同样注重数据的历史溯源,Linked Data 的内容表达要充分考虑到数据的溯源信息,允许用户通过这些信息发现到更有价值的内容。这些问题的解决,对数字科研环境下的基于 Linked Data 机制的知识发现具有更为实际的意义。

参考文献:

- [1] 王芳, 张晓林. 元搜索技术: 原理与应用[J]. 现代图书情报技术, 1998(6):18-21.
- [2] Serials Solution Summon Web - Scale - Discovery [EB/OL]. [2010-07-06]. <http://www.serialssolutions.com/summon>.
- [3] Ex Libris Primo Central [EB/OL]. [2010-02-14]. <http://www.exlibrisgroup.com/category/PrimoCentral>.
- [4] NLM Gateway Service [EB/OL]. [2010-07-06]. <http://gateway.nlm.nih.gov>.
- [5] Web Services Architecture [EB/OL]. [2010-02-11]. <http://www.w3.org/TR/ws-arch/>.
- [6] Semantic Web [EB/OL]. [2010-09-23]. <http://www.w3.org/standards/semanticweb/>.
- [7] Mashup (Web Application Hybrid) [EB/OL]. [2010-09-23]. [http://en.wikipedia.org/wiki/Mashup_\(web_application_hybrid\)](http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid)).
- [8] Berners-Lee T. Linked Data - Design Issues [EB/OL]. [2009-02-18]. <http://www.w3.org/DesignIssues/LinkedData.html>.
- [9] SweoIG/TaskForces/CommunityProjects/LinkingOpenData [EB/OL]. [2010-07-06]. <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.
- [10] 刘炜. 关联数据: 意义及其实现 [EB/OL]. [2010-09-16]. <http://www.kevenlw.name/archives/1435>.
- [11] 刘炜. 数据的万维网 (The Web of Data) [EB/OL]. [2010-09-16]. <http://www.kevenlw.name/archives/1185>.
- [12] 黄永文. 关联数据在图书馆中的应用研究综述 [J]. 现代图书情报技术, 2010(5):1-7.
- [13] 黄永文. 关联数据驱动的 Web 应用研究 [J]. 图书馆杂志, 2010(7):55-59.

- [14] 白海燕. 基于关联数据的书目组织深度序化初探[C]. 见: 2010 图书馆前沿技术论坛, 上海. 2010.
- [15] Open Science Data [EB/OL]. [2010 - 11 - 01]. http://en.wikipedia.org/wiki/Open_Data.
- [16] Linking Enterprise Data (LED) [EB/OL]. [2010 - 11 - 01]. <http://zephaira.com/community/LED/>.
- [17] Bizer C, Cyganiak R, Heath T. How to Publish Linked Data on the Web [EB/OL]. [2008 - 07 - 17]. <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/>.
- [18] Pubby - A Linked Data Frontend for SPARQL Endpoints [EB/OL]. [2008 - 07 - 17]. <http://www4.wiwiw.fu-berlin.de/pubby/>.
- [19] The D2RQ Platform v0.7 - Treating Non - RDF Relational Databases as Virtual RDF Graphs [EB/OL]. [2010 - 11 - 01]. <http://www4.wiwiw.fu-berlin.de/bizer/d2rq/spec/>.
- [20] OpenLink Software [EB/OL]. [2010 - 09 - 11]. <http://virtuoso.openlinksw.com>.
- [21] Tripify.org: Overview [EB/OL]. [2010 - 09 - 12]. <http://tripify.org/>.
- [22] D2RQ 映射语言的 Schema 定义 [EB/OL]. [2010 - 09 - 12]. <http://www4.wiwiw.fu-berlin.de/bizer/d2rq/0.1/>.
- [23] Haslhofer B, Schandl B. The OAI2LOD Server: Exposing OAI - PMH Metadata as Linked Data [C]. In: *Proceedings of International Workshop on Linked Data on the Web (LDOW2008)*, Beijing, China. 2008.
- [24] JPEG RDFizer - SIMILE [EB/OL]. [2010 - 09 - 12]. http://simile.mit.edu/wiki/JPEG_RDFizer.
- [25] MARC/MODS RDFizer - SIMILE [EB/OL]. [2010 - 09 - 12]. http://simile.mit.edu/wiki/MARC/MODS_RDFizer.
- [26] Email RDFizer - SIMILE [EB/OL]. [2010 - 09 - 12]. http://simile.mit.edu/wiki/Email_RDFizer.
- [27] RDF Book Mashup - Serving RDF Descriptions of Your Books [EB/OL]. [2010 - 08 - 15]. <http://www4.wiwiw.fu-berlin.de/bizer/bookmashup/index.html>.
- [28] Schandl B, Popitsch N. Lifting File Systems into the Linked Data Cloud with TripFS [C]. In: *Proceedings of Linked Data on the Web (LDOW2010)*, Raleigh, North Carolina, USA. 2010.
- [29] Hausenblas M, Troncy R. Interlinking Multimedia - How to Apply Linked Data Principles to Multimedia Fragments [C]. In: *Proceedings of International Workshop on Linked Data on the Web (LDOW2009)*, Madrid, Spain. 2009.
- [30] Barbieri D F, Valle E D. A Proposal for Publishing Data Streams as Linked Data [C]. In: *Proceedings of International Workshop on Linked Data on the Web (LDOW2010)*, Raleigh, North Carolina. 2010.
- [31] Fresnel - Display Vocabulary for RDF [EB/OL]. [2010 - 09 - 12]. <http://www.w3.org/2005/04/fresnel-info>.
- [32] Hassanzadeh O, Consens M. Linked Movie Data Base [C]. In: *Proceedings of LDOW2009*, Madrid, Spain. 2009.
- [33] Raimond Y, Sutton C, Sandler M. Automatic Interlinking of Music Datasets on the Semantic Web [C]. In: *Proceedings of LDOW 2008*, Beijing, China. 2008.
- [34] Volz J, Bizer C, Gaedke M, et al. Silk - A Link Discovery Framework for the Web of Data [C]. In: *Proceedings of LDOW 2009*, Madrid, Spain. 2009.
- [35] Haslhofer B, Popitsch N. DSNotify - Detecting and Fixing Broken Links in Linked Data Sets [C]. In: *Proceedings of the 20th International Workshop on Database*. 2009.
- [36] Web of Data Link Maintenance Protocol - Maintaining Links Between Changing Linked Data Sources [EB/OL]. [2010 - 08 - 15]. <http://www4.wiwiw.fu-berlin.de/bizer/silk/wodlmp/>.
- [37] Ping the Semantic Web. com - Share Your RDF Data with the World! [EB/OL]. [2010 - 08 - 15]. <http://pingthesemantic-web.com/>.
- [38] Becker C, Bizer C. DBpedia Mobile: A Location Enabled Linked Data Browser [C]. In: *Proceedings of LDOW2008*, Beijing, China. 2008.
- [39] Sheridan J, Tennison J. Linking UK Government Data [C]. In: *Proceedings of LDOW2010*, Raleigh, North Carolina. 2010.
- [40] Zhao J, Klyne G, Shotton D. Provenance and Linked Data in Biological Data Webs [C]. In: *Proceedings of LDOW2008*, Beijing, China. 2010.
- [41] Latif A, Afzal M T, Helic D, et al. Discovery and Construction of Authors' Profile from Linked Data (A Case Study for Open Digital Journal) [C]. In: *Proceedings of LDOW2010*, Raleigh, USA. 2010.

(作者 E-mail: shenzhihong@mail.las.ac.cn)