

关联数据:图书馆学研究的新领域

徐 华,曾微泊^①,王旭东^②
(南华大学 图书馆,湖南 衡阳 421001)

[摘 要] 图书馆界对关联数据的研究与应用越来越重视。关联数据将为图书馆的发展带来新的机遇,主要表现在:关联数据的核心技术为图书馆的资源集成提供保障、图书馆将自身资源发布为关联数据,提升社会价值、利用关联数据加强图书馆与教学系统间的联系。图书馆应用关联数据必须正视一些主要问题,如数据整合问题、关联数据的效用问题以及关联数据的安全性问题。

[关键词] 关联数据; 图书馆; 知识组织; 数据安全

[中图分类号] G250.76 **[文献标识码]** A **[文章编号]** 1673-0755(2011)06-0107-03

一 关联数据概述

(一)提出

2006 年,被誉为互联网之父的 Tim Berners - Lee 在《关联数据构建笔记》等一系列文章中分析了 Web 的发展和演化,提出发展数据网络的思想,而数据网络的核心和关键则是关联数据(Linked data)^[1]。自关联数据的概念提出以来,得到了广泛的认同和快速的发展。据统计,2007 年 5 月,互联网上构建并发布的关联数据大约有 5 亿个 RDF 三元组;2008 年,RDF 三元组已超过 20 亿;至 2009 年 5 月,已有超过 100 个数据集,RDF 三元组已超过 47 亿个^[2];目前数量还在不断增加。关联数据技术正在使互联网发生着深刻的变革。

(二)涵义

目前学术界对于关联数据涵义的理解尚未完全统一。一般比较认同的是维基百科中给出的定义:关联数据是语义网的主题之一,描述了通过可链接的 URI 方式来发布、分享、连接 Web 中各类资源的方法^[3]。国内学者刘炜先生认为,关联数据强调建立已有信息的语义标注和实现数据之间的关联,具有框架简洁、标准化、自助化、去中心化、低成本的特点,为构建人机理解的数据网络提供了根本性的保障,为实现语义网奠定了坚实的基础^[4-5]。

(三)基本原则和原理

关联数据的构建和实现遵循 4 个基本原则:(1)任何事物都以 URI 作为标识名称;(2)使用 Http URI 让任何人都能访问这些标识名称;(3)当有人访问某个标识名称时,能提供有用的信息;(4)尽可能提供相关的 URI,以使人们可以发现更多的事物^[6-7]。

关联数据的原理是用一种可利用分布数据集及其自主内容格式,基于标准的知识表示与检索协议、可逐步扩展的机制来实现可动态关联的知识对象网络,并支持在此基础上的知识组织和知识发现^[8]。由此可见,关联数据对于图书馆的知识组织将起到积极的作用。

二 关联数据为图书馆带来新的发展机遇

作为知识组织与处理的专门机构,图书馆的 MARC 数据、主题标目、规范记录等都可以开放成任意互联的关联数据。国外图书馆界对关联数据的研究与应用非常重视,典型的有美国国会图书馆和瑞典国家图书馆^[9]。2010 年 5 月,W3C 宣布成立图书馆关联数据孵化小组(Library Linked Data Incubator Group)^[10],以帮助图书馆建立关联数据,增强图书馆数字资源的互操作性^[11]。国内图书馆界也开始着手研究关联数据。可见,关联数据为图书馆带来新的发展机遇。

(一)关联数据的核心技术为图书馆的资源集成提供保障

众所周知,互联网中的信息是分散的。为了解决用户查询以及有效利用问题,技术人员采用将 Web 资源进行识别,并通过相互链接集成的方式,如浏览器与搜索引擎就是两种常见的与用户交互性极强的资源集成模式^[12]。有些则采取传统的超链接方式,但这些链接之间缺乏语义操作^[13]。关联数据的核心技术之一是 RDF(资源描述框架),为互联网提供了一种更好的资源集成方式,对于图书馆界也非常实用^[14]。RDF 是专门用于表达关于任何可在互联网上被标识的资源的元数据,一方面浏览器可通过 RDF 三元组描述的链接来显示图书馆的信息资源,用户在浏览某一条信息时,

[收稿日期] 2011-10-22

[基金项目] 衡阳市社科基金项目资助(编号:2010D034)

[作者简介] 徐华(1974-),女,湖南衡南人,南华大学图书馆馆员。

①南华大学图书馆副研究馆员。②南华大学电气工程学院教师。

同时还可看到来自其他图书馆的资源。另一方面搜索引擎可通过 RDF 链接对图书馆的书目、电子数据库进行搜索,并提供对图书馆集成数据的查询,查询结果以包含特定查询内容的 RDF 文件形式显示出来。因此,互联网上分散的图书馆知识资源相互关联起来,给用户使用带来极大方便^[15]。

(二)图书馆将自身资源发布为关联数据,提升社会价值

关联数据不但减轻了互联网中整合分布式异构数据源的复杂性,同时也推动了图书馆对关联数据应用的新篇章。关联数据对于图书馆在互联网上发布书目数据和规范数据具有重要意义,与网络时代的图书情报工作关系极为密切,是互联网发展到语义网时代,对网上资源和数字对象进行“编目”和“规范控制”的基础性技术,是图书馆进行信息资源发布和建设的核心技术之一^[16]。目前国际上已有一些图书馆开始采用和发布关联数据,如美国国会图书馆、瑞典国家图书馆等。在新信息化环境下,增强图书馆的社会服务能力和水平尤为迫切。知识发现对于图书馆服务而言非常重要,关联数据为图书馆的知识发现服务提供了良好的途径。关联数据将图书馆的知识和外部世界相互连接起来,增强和扩展了图书馆的知识发现平台,改善了图书馆的服务功能,提升了图书馆的社会价值^[17]。我国图书情报界应重视这一技术,尽快进行研究开发和应用推广,使我国图书馆大量的权威数据在互联网上占据一席之地。

(三)利用关联数据加强图书馆与教学系统之间的联系

利用关联数据加强图书馆与教学间的联系,国外已进行了积极探索。英国的 JISC 项目提出了分步骤实现跨高等教育机构创建关联数据的方案,从而促进机构联盟内部对教学和图书馆资料等资源的共享,通过构建教育类本体,实现基于本体的数据分析以及教育感知推理应用。资源列表管理工具的对象包括教师指定的图书、期刊文献、网页或多媒体资料等教学参考资源。目前,使用的资源管理工具大多缺乏与资源发布平台、图书馆目录、虚拟学习环境之间数据的互操作性,导致学生不能标注、评价教学参考资源,而且不支持个体根据自身特定学习任务需要增加新资源,没有反馈和互动。采用语义技术可解决现有工具存在的普遍问题,具体做法是:使用现有的本体对资源统一描述、根据关联数据原则提高数据的互操作性、用现有的模式和本体来描述相互关系、鼓励教师和学生丰富数据的语义以支持环境感知推荐功能等。利用关联数据构建起来的系统不仅实现了统一描述学习资源,而且丰富了资源的语义描述,2008 年,语义技术在英国普利茅斯大学实施,2010 年在英国大学客户群中得到推广和使用,这为国际上图书馆资源与教育教学资源之间的关联提供了成功案例^[18]。

三 图书馆应用关联数据必须正视的主要问题

(一)用于关联的图书馆数据整合问题

数据整合是将同一对象的多个数据项整合成统一的数据处理过程;从分布式的数据源检索到所需数据后,应在显示给用户之前以有效的方式作进一步整合。关联数据通常来源于分布、异构的多个数据源,用于关联的图书馆数据也一样,因此在用户浏览或作进一步处理前,应进行数据整

合^[19]。但目前大多数关联数据应用只是有分别地显示不同数据源的数据,并未作进一步整合处理,图书馆的服务功能难以充分发挥。用于关联的图书馆数据整合可分为词汇映射和数据集成两条途径,词汇映射是将多个词表或本体中的词汇通过一定的规则对应起来;数据集成则是将现实世界中某对象在数据网络中的不同含义集成为一个清晰的、稳定的表述。为了进一步整合图书馆数据,需要将来自不同数据源的术语映射到目标应用模式中,可通过使用自己的词表体系,或在现有的词表基础上,根据需要增加自定义的术语,增加其关联性。

(二)关联数据的效用问题

随着国际上图书馆将自身关联数据发布在互联网上的实践,越来越多的图书馆加入到这一行列。但遇到了棘手的问题:由于关联数据网络中的数据随时更新,过时的数据需要删减或修改,数据之间的链接也相应改变。一旦某个数据源发生变动,数据源之间的关联可能会形成断链,这会引起关联数据的应用随之也发生错误,进而导致关联数据失效^[20]。目前的一般做法是忽略这些问题,当具体的应用发现断链时再解决^[16]。当前的网络结构客观上存在一定数量的死链接,但死链接过多将使客户端发出大量的无效 HTTP 请求,从而降低数据网络的效率^[21]。为了减少应用层对断链的处理,提高数据使用效率,图书馆对数据源应确保关联关系的有效性。同时,图书馆对关联数据源还应建立监测和修正机制,以维护数据参照的完整性。

(三)关联数据的安全性问题

对于图书馆来说数据安全至关重要:书目数据、读者数据和流通数据,无论任何一项一旦丢失,后果将不堪设想。馆藏文献数字化的数据,如若丢失则更令我们图书馆人心痛不已。尽管有研究认为关联数据可依赖较完善的模式映射和链接保存策略,安全问题还是图书馆界担忧的主要问题之一^[22]。关联数据的理想目标是将目前的 Web 转化为一个全球数据库,这既为 Web 服务带来便利,同时又带来诸多安全性问题,特别是用户的个人数据可能会泄漏,网络隐私权较易受到侵犯。保护关联数据文本的隐私权不仅需要技术手段与法律方式相结合,而且也需要用户具有良好的安全防范意识。

[参考文献]

- [1] 白海燕. 关联数据及 DBPedia 实例分析[J]. 现代图书情报技术, 2010(3): 33-39.
- [2] Hausenblas M. Exploiting Linked Data for Building Web Applications [EB/OL]. [2011-09-26]. <http://sw-app.org/pub/exploit-lod-webapps-IEEEIC-preprint.pdf>.
- [3] Wikipedia. Linked data[EB/OL]. [2011-09-26]. http://en.wikipedia.org/wiki/linked_data.
- [4] 刘炜. 关联数据的意义与实现[EB/OL]. [2011-09-26]. <http://202.114.9.60/dl6/pdf/24.pdf>.
- [5] 刘炜. 几个概念: 开放数据, 关联数据, 语义 Web 和 Web 3.0[EB/OL]. [2011-10-18]. <http://www.kevenlw.name/archives/2114>.

- [6] Bemers-Lee T. Linked Data [EB/OL]. [2011-09-26]. <http://www.w3.org/designissues/LinkedData.html>.
- [7] 郭少友. OAI-PMH 元数据的关联数据化方法研究[J]. 图书情报工作, 2011(2):107-111.
- [8] 沈志宏, 张晓林. 关联数据及其应用现状综述[J]. 现代图书情报技术, 2010(11):1-9.
- [9] Bizer C, Heath T, Berners-Lee T. Linked data-the story so far [EB/OL]. [2011-10-18]. <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.
- [10] Berners-Lee, Diego Jon Phipps. Best Practice Recipes for Publishing RDF Vocabularies W3C Working Draft [EB/OL]. [2011-10-18]. <http://www.w3.org/TR/swbp-voca-pub/>.
- [11] 潘有能, 张悦. 关联数据研究与应用进展[J]. 情报科学, 2011(1):124-130.
- [12] 马费成等. 基于关联数据的网络信息资源集成[J]. 情报杂志, 2010(2):167-170.
- [13] Summery Ed. Following Your Nose to the Web of Data Information Standards Quarterly [EB/OL]. [2011-10-18]. <http://inkdroid.org/journal/following-your-nose-to-the-web-of-data>.
- [14] Ed Summers et al., 姚小乐, 刘炜译. LCSH, SKOS 和关联数据[J]. 现代图书情报技术, 2009(3):8-14.
- [15] 黄永文. 关联数据驱动的 Web 应用研究[J]. 图书馆杂志, 2010, 29(7):55-59.
- [16] 刘 炜. 关联数据:概念、技术及应用展望[J]. 大学图书馆学报, 2011(2):5-12.
- [17] 赵梓彤, 谢海先. 关联数据在网络信息管理中的应用[J]. 图书馆学研究(应用版), 2011(6):47-50+29.
- [18] 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.
- [19] 曾 蕾. 关联的图书馆数据[EB/OL]. [2011-10-18]. <http://202.114.9.60/6/26>.
- [20] Isaac, Antoine, Summers Ed. SKOS Simple Knowledge Organization System Primer [EB/OL]. [2011-10-18]. <http://www.w3.org/TR/skos-Primer/>.
- [21] Prud'hommeaux Eric, Andy Seabome. SPARQL query language for RDF [EB/OL]. [2011-10-18]. <http://www.w3.org/TR/rdf-sparql-query/>.
- [22] 白海燕. 基于关联数据的书目组织深度序化初探[C]. 2010 图书馆前沿技术论坛, 上海. 2010-08.

Linked Data: the New Research Area of Library Science

XU Hua, ZENG Wei-bo, WANG Xu-dong

(University of South China, Hengyang 421001, China)

Abstract: The library scholars pay more and more attention to linked data's research and application. Linked data will bring new chances for library's development, which are as follows: the core technology of linked data offers guarantee for library's resource integration; library issues itself resource as linked data so that prompting social value; strengthen the relations between library and teaching system by linked data. The library must envisage some main problems during linked data's applications, such as data integration, the effect of linked data and security of linked data.

Key words: linked data; library; knowledge organization; data security