

OTCSS 关联数据服务的研究与实现^{*}

黄华军 曾新红 林伟明

(深圳大学图书馆 NKOS 研究室 深圳 518060)

【摘要】分析关联数据基本原则及技术体系,以《中国分类主题词表》为基础,概述中文叙词表本体共建共享系统 OTCSS 的关联数据应用程序的实现,详细介绍其服务功能,并与 LCSH Linked Data 进行比较。

【关键词】关联数据 中文叙词表本体共建共享系统 中国分类主题词表

【分类号】G250.73 TP393.4

Research and Implementation about Linked Data Service of OTCSS

Huang Huajun Zeng Xinhong Lin Weiming

(NKOS Research Office of Shenzhen University Library, Shenzhen University, Shenzhen 518060, China)

【Abstract】This article analyzes the basic principles and technology stack of linked data. It provides an overview of implementation about linked data service of OTCSS based on Chinese Classified Thesaurus, and introduces the service functions of linked data in detail, further makes comparison with linked data realized based on LCSH.

【Keywords】Linked data OTCSS CCT

1 引言

Linked Data 是一个较新的研究课题,近年来逐渐引起学术界和工业界的广泛关注。国外在这方面的研究和应用稍早,已在许多领域崭露头角并取得不少成果,例如瑞典国家图书馆 LIBRIS 国家书目关联数据^[1]、美国国会图书馆关联数据服务^[2]以及英国 BBC 网站的关联数据服务^[3]等。自 2008 年以来,国内有关 Linked Data 的研究逐渐热门起来,但目前公开可获取的研究资料大部分属于关联数据报告、综述及应用展望论文,较少涉及技术实践问题,还未见公开发布的 Linked Data 服务。

本文以《中国分类主题词表》(Chinese Classified Thesaurus, CCT)为实验源数据,根据 Linked Data 基本原则和相关技术,在深圳大学图书馆 NKOS 研究室相关研究成果及国家图书馆知识组织标准规范草案^[4,5]的基础上,构建和发布了 CCT 的 Linked Data 实验性服务。

本文的主要目的在于:

- (1)以国外已有的 NKOS 关联数据应用为参考,提出图书馆的传统中文 KOS Linked Data 化的一种实现途径;
- (2)验证在版权机构许可下或帮助版权机构构建和发布它们的关联数据的可行性;
- (3)尝试在国内开放实验性的关联数据服务,并留有链接不同机构的关联数据和不同来源数据集的空间;
- (4)为国内相对封闭的图书馆数据实现从文献 Web 向数据 Web 的转变提供实验性的探索和实验数据,以更好地发挥图书馆在网络时代传播知识的职能。

收稿日期:2012-06-19

收修改稿日期:2012-08-15

^{*} 本文系深圳大学人文社会科学基金项目“图书馆关联数据服务的实现与应用研究”(项目编号:12QNCG11)、广东省哲学社会科学“十一五”规划基金项目“中文知识组织系统的形式化语义描述标准体系研究”(项目编号:GD10CTS02)和国家社会科学基金项目“中文知识组织系统的形式化语义描述标准体系研究”(项目编号:12BTQ045)的研究成果之一。

2 关联数据简介

2.1 关联数据原则

2006 年, Berners - Lee^[6] 提出关联数据这一概念时, 详细阐述了关联数据的 4 个基本原则:

(1) 使用 URI 作为事物的标识名称 (Use URIs as names for things)。

(2) 使用 HTTP URI 使人们可以访问这些标识名称 (Use HTTP URIs so that people can look up those names)。

(3) 当有人访问某个标识名称时, 使用 RDF、SPARQL 等标准, 提供有用的信息 (When someone looks up a URI, provide useful information, using the standards (RDF *, SPARQL))。

(4) 尽可能提供相关的 URI 链接 (即属性资源描述), 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)。

从以上原则可以看出, Linked Data 使用 HTTP URIs 对资源进行标识, HTTP 协议对资源进行检索和传输, RDF 对资源进行描述, 因此 HTTP、URI 和 RDF 一起构成了 Linked Data 的基础, 下面分别对此进行简要介绍。

2.2 关联数据技术体系

(1) 资源与资源标识

资源: 在发布关联数据之前, 首先需要确定所发布的资源是什么。任何事物, 只要被认为是具有发布意义的, 或者有被引用的价值, 都可以称为资源, 如文档、图片、馆藏书目、叙词概念、分类号等。

资源标识: 任何资源都可用一个 HTTP URI 来标

识。之所以使用 HTTP URI 来标识, 是希望数据能够通过 HTTP 协议传输, 真正实现基于 Web 的访问与互联。

URI 规定用 ASCII 字符作为资源标识符的符号, 这不利于非英语国家和地区的使用者。因此, 本文采用引入了 Unicode 字符的资源标识符, 即国际化资源标识符 (Internationalized Resource Identifiers, IRI)^[7], 它是 URI 的扩展, 包括了所有语种的文字字符, OWL2 推荐标准已采用 IRI 作为资源名称的唯一标识^[8]。显然中文知识组织系统也适合使用 IRI 作为资源标识符。

(2) 资源描述与关联

资源描述框架^[9] (Resource Description Framework, RDF) 是基于主体 - 谓词 - 客体三元组结构的机器可理解可处理的模型。主体表示需要描述的资源, 谓词表示主体的某个属性或关系, 客体表示属性或关系的值。主体、谓词都需要用 HTTP URI 来表示, 客体可用 HTTP URI 标识另一资源, 也可以是字符串表示的纯文本。主体可看作是类资源, 谓词可看作是类资源的属性资源, 客体是类资源或文字型资源。根据 RDF 模型中客体的种类, 可以将三元组分为两类^[10]: 文字型三元组, 表示类资源与文字型资源的关联; RDF Links 型三元组, 表示类资源之间的关联。

RDF Links 表现为三元组的形式, 由三个 URI 引用组成, 属性资源成为不同资源之间实现关联的桥梁, 因此根据 Linked Data 的原则 (4), 如需使人们发现更多资源, 应尽可能多地提供属性资源的描述, 即提供更多 RDF Links, 让更多的资源或数据关联起来。RDF Links 的例子如图 1 所示:

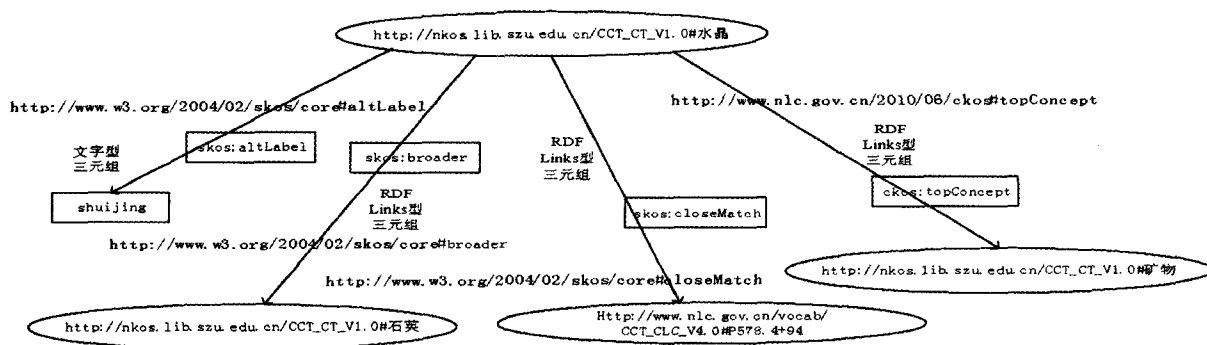


图 1 RDF Links 资源描述

(3) 资源检索

人们可通过 HTTP/URI 机制,直接检索获得资源(Thing)的信息,实现一种 Web 上的富链接机制,即将超文本链接(文件之间的链接)转变为超数据链接(资源之间的链接)。

HTTP 通信协议允许通过 URIs 使数据从 Web 的服务器传送到浏览器,因此 HTTP 成为 Linked Data 一种简单而通用的检索机制,该方式不仅可以检索网络上的数字资源,还可以检索那些自身无法通过 HTTP 在网络上传输的实体的描述^[11,12]。一般可通过以下三种方式从网络获取 Linked Data:提供 HTTP URIs 参引;使用 SPARQL 本体查询语言;提供多种共享格式的下^[10,13]。

3 LCSH 的 Linked Data 服务介绍

《美国国会图书馆标题表》(Library of Congress Subject Headings, LCSH)是目前世界上使用时间最长、使用范围最广、规模和影响最大的一部综合性标题表。LCSH 含有 26.5 万条规范记录,传统上以 MARC21 格式进行发布。2006 年国会图书馆开始探索 LCSH/MARC 向 SKOS 的转换,目前已成功地在 Web 上发布了 LCSH/SKOS 版本,提供 Linked Data 服务、SKOS 版本下载服务、SPARQL 查询服务等^[14]。

LCSH 的 Linked Data 发布网址为 <http://id.loc.gov/authorities/> (后改为 <http://id.loc.gov/>), 其 Linked Data 提供的服务主要有:

(1) 使用 HTTP URI 作为对象的唯一标识(如“<http://id.loc.gov/authorities/sh85054367#concept>”唯一标识标目“German language -- Old High German, 750 - 1050—Etymology”),通过 HTTP URI 可以定位到具体的对象,如图 2 所示:

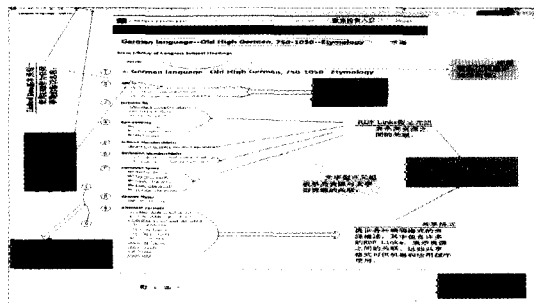


图 2 LCSH Linked Data 服务

(2) 通过查询对象的 URI,可以提供很多有意义的信息,如该标目的款目信息(Alternate Labels、Broader Terms、Narrower Terms、Related Terms、Created、Modified、Editorial Notes、Sources 等)以及 URI、Type、Instance Of、Alternate Formats 等信息;

(3) 在 Alternate Formats 中提供 RDF/XML、N - Triples 和 JSON 等多种格式的下^[10,13]或在线浏览;

(4) 提供到其他词表中相似概念的关联(Similar concepts from other vocabularies);

(5) 提供可视化展示和用户建议功能。

LCSH Linked Data 服务是目前国际上词表系统关联数据应用的代表。同为词表,传统中文叙词表(如 CCT)与 LCSH 可提供信息的项目是类似的,因此本研究刻意模仿了 LCSH Linked Data 服务的界面,以便更容易被业内人士所接受。但中文叙词表和 LCSH 毕竟是两种词表,它们的结构和服务对象都有所不同,因此本文所采用的 URI 方案、提供的下载格式^[15]的表示方法和种类以及检索结果的展示等都是从我国的具体情况出发而设计的。本研究在 OTCSS 原有实现技术的基础上,遵循关联数据创建的 4 个原则,实现了 OTCSS Linked Data 服务。

4 OTCSS Linked Data 的实现

OTCSS Linked Data 服务的具体实现方法包括以下 4 个方面:

(1) 准备数据源:本文的关联数据实验数据源以目前中国图书馆界应用最为广泛的综合性词表《中国分类主题词表》为例。中文叙词表本体(OntoThesaurus)和中文叙词表本体共建共享系统(OTCSS)^[15]是国家社会科学基金项目“基于本体和知识集成实现中文叙词表的升级、共享和动态完善”(项目编号:05CTQ001)的研究成果,OntoThesaurus 是在中文叙词表的基础上通过引入本体相关理论和技术构建而成,在其基础上已实现了较为完备的 OTCSS,功能包括:从中文叙词表到 OntoThesaurus 的自动转换功能、一致性检测推理功能、网络术语服务功能(供人使用的 OntoThesaurus - TS 和供应用程序使用的 Web Service OntoThesaurus - API),以及全面深入的网络共建功能(用户界面 + 修订专家界面)^[15,16]。本文的 Linked Data 源直接从已有的研究成果 OntoThesaurus 中获取。

(2)URI 方案:为中文叙词表本体所有叙词概念款目中的主体、谓词和客体分配唯一的 URI(文字型三元组的客体除外),这里采用引入了 Unicode 字符的 URI 即 IRI 来描述包含中文字符的资源,已通过实验证明其可行性。

URI 的解决方案有两种,它们都可以满足标识真实世界任何实体对象甚至抽象概念或不存在的事物的要求,即 303 URIs 和 Hash URIs,使用哪一种可根据用户的实际需求做出选择,文献[17]详细介绍了这两者的具体情况。

具体词表的 URI 表示方案应由版权机构来制定,笔者可以帮助词表的版权机构制定并实现 URI 方案。作为一项实验性的研究,这里以《中国分类主题词表》(一版)为例,暂时采用“http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#叙词概念”作为其叙词概念的 URI 方案,即 Hash URIs 方案,例如叙词“中越关系”的 URI 是“http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系”。

(3)关联实现:Linked Data 的关键在于 RDF Links,价值在于支持结构化数据的语义互联,通过“链”实现在不同资源之间的导航。对于中文叙词表本体所有叙词概念款目中的每个三元组,通过 RDF Links 即表示谓词的 URI,将表示主体的 URI 指向表示客体的 URI 或文字,如图 1 所示,可点击主体的关联对象链接查询客体详情实现数据的发现。目前主流的浏览器(例如 IE)对于包含#号的 URI 的处理仅仅是为了使用锚进行页面内的定位,还无法直接实现 Linked Data 资源导航,虽已有个别浏览器(例如 Firefox)通过开发插件的方式可以实现,但需要针对不同的浏览器开发相应的插件兼容所有浏览器,这增加了实现的难度和复杂度。鉴于上述情况,本文采用 URL 地址重写技术实现 Linked Data 资源导航,即 Linked Data 中所有对象通过 URL 地址重写链接到其具体内容。

(4)Linked Data 发布:为了拥有良好的用户体验,帮助非业内人士使用,本文发布的 Linked Data 服务不仅提供供人浏览的 HTML 展示页面,还提供供机器和应用程序使用的各种数据模型(即 9 种共享格式)。

5 OTCSS Linked Data 服务的功能及具体使用方法

5.1 OTCSS Linked Data 服务的功能

OTCSS Linked Data 服务提供的功能如图 3 所示。

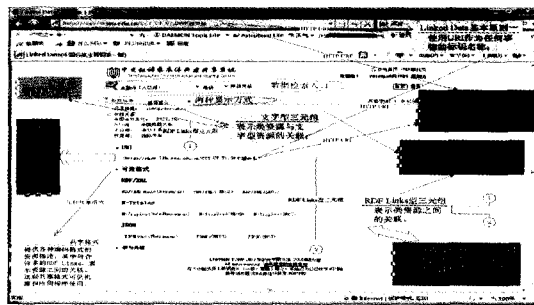


图 3 OTCSS Linked Data 服务

(1)为中文叙词表本体的所有叙词概念(即所要发布的资源)提供 HTTP URI,以此作为叙词概念的唯一标识,通过此 HTTP URI 可以定位到具体的叙词概念;

(2)提供地址栏和检索框两种方式输入 URI 或叙词概念的名称查询关联数据,叙词款目的展示提供专业显示和通俗显示两种方式;

(3)通过查询对象的 URI,可以提供该对象的很多有意义的信息(关联对象),如叙词款目信息(拼音、中图法分类号、入口词、英译名、上位词、下位词、相关词、族首词等)以及 URI、可选格式等信息,并可通过关联对象的 RDF Links 查询其详情;

(4)在可选格式中提供 RDF/XML、N-Triples 和 JSON 格式(这三种格式又分为 OntoThesaurus、CNKOS、SKOS 三种不同描述格式,共 9 种格式)的下载或在线浏览;

(5)未来可提供到其他词表相似概念的关联(如到中图法分类号对应类目的映射链接);

(6)可视化展示和一些其他属性的关联正在建设中。提交用户建议等共建功能则通过 OntoThesaurus - TS^[15]的界面进行。

比较 LCSH Linked Data 和 OTCSS Linked Data,前者选择的是 303 URIs 方案,后者选择的 Hash URIs 方案,两者提供的共享格式表示方法和种类有所不同。此外,后者的检索结果界面保留了传统的叙词款目格式,并扩展了通俗显示格式,以满足业内外人士的使用需求。LCSH Linked Data 为网络资源利用 LCSH 建立资源之间的直接联系,使主题数据机读化、Web 化提供了一个很好的基础^[18]。希望 OTCSS Linked Data 方案也能促进我国的传统知识组织系统达到同样的目标。

5.2 OTCSS Linked Data 使用方法

实验系统于 2011 年 1 月实现并发布,登录地址: http://nkos.lib.szu.edu.cn/CCT_CT_V1.0。网络用户无需注册登录就可以直接检索中文叙词表本体中的所有叙词概念。

(1) 用户可以通过两种方式查询叙词概念:

① 在图 3 所示界面的检索框中输入任意叙词概念, 点击检索按钮, 出现该叙词概念的所有相关信息, 并且地址栏中的 URI 也相应变化。

② 从 URL 地址栏中输入叙词概念的 URI, 例如输入“http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系”后回车(需刷新页面), 出现如图 3 所示界面, 其中检索输入框中也自动填入“中越关系”。

(2) 用户可以根据喜好选择通俗或者专业两种方式来显示叙词款目;

(3) 点击页面中叙词款目中的超链接, 如“中外关系”, 页面的所有信息也相应地变化, URL 地址栏中出现“http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中外关系”, 检索输入框中自动填入“中外关系”;

(4) 页面中的 URI 显示的是叙词概念的 URI, 即地址栏 URL 中和检索框中输入的是叙词概念的入口词;

(5) 用户可以点击页面的 9 种格式超链接来下载叙词款目信息语义描述的不同文本。

5.3 共享格式的下载

LCSH Linked Data 服务规范标目的可选格式提供 RDF/XML、N-Triples、JSON 等多种共享格式。本文根据中文叙词表本体的特点和已有的 NKOS 语义描述研究成果, 将这三种格式又分别细分为三种格式, 即针对 RDF/XML、N-Triples、JSON 格式, 将它们分别通过 OntoThesaurus、CNKOS、SKOS 三种语义描述格式来表示(这三种格式在指定条件下可以实现相互之间的转换)。

SKOS(简单知识组织系统)^[19,20]为知识组织系统在 Web 上的共享和链接提供了一个通用的数据模型。SKOS Reference 20090818 已成为 W3C 的正式推荐标准。SKOS 的定义有较大的弹性, 是基于 RDF 的非严格意义上的形式化语言, 以容纳结构较为松散的 KOS。其数据模型与 OWL Full 兼容, 有较好的表达性, 但与 OWL DL 不兼容, 不具备完备的推理性能。

本研究依据的是国家图书馆制定的“受控表语义描述规范”^[4], 其中对中文叙词表(主题词表)的 SKOS 描述方法进行了具体的规定。只采用 SKOS 来描述中

文叙词表会丢失一些语义信息。

CNKOS^[4,5]是国家图书馆正在制定的一种面向国内传统知识组织系统(如主题词表、分类法等受控表)的语义描述规范格式, 它在 SKOS 基础上扩展了 CKOS 词汇, 以期实现我国传统知识组织系统的全描述。该规范已通过国家图书馆组织的终审验收。本实验研究的其中一个目标就是验证其 URI 方案和语义描述方法的可行性, 为将来的支持系统实现提供更多的经验。

OntoThesaurus 设计用来表示结构规范的中文叙词表(主题词表), 以实现其本体化升级和在语义 Web 环境中的共享应用和网络化共建, 可视为 OWL 在中文叙词表领域的一个应用子集。简化后也可用于中文规范档、专业分类表(Taxonomy)等 KOS 类型的语义描述。采用 OWL DL, 可实现完备的推理^[16]。其命名域为 <http://nkos.lib.szu.edu.cn/2010/10/ont/>, 缩写为“ont”。

OTCSS Linked Data 服务采用以上三种语义描述格式, 前两种依据的是国家图书馆知识组织标准规范 D009-3^[4,5]和 D009-4 报告, 后者依据的是中文叙词表本体 OntoThesaurus 词汇表^[21]。它们的形式化程度由低到高, 可以满足不同的语义描述需求, 在指定的条件下也可以实现相互之间的转换。

RDF/XML 是 W3C 推荐使用的 RDF 的 XML 序列(Serialization), 术语注册中心一般要求以 RDF/XML 文件格式提交所要注册的术语集合。图 4 至图 6 分别对应 RDF/XML 的三种格式, 即 RDF/XML(OntoThesaurus)、RDF/XML(CNKOS)、RDF/XML(SKOS)。这些格式都是机器可理解的。

```
<?xml version="1.0" encoding="GBK"?>
<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdfs="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vad="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:ont="http://nkos.lib.szu.edu.cn:8080/2010/10/ont#"
  xml:base="http://nkos.lib.szu.edu.cn:8080/CCT_CT_V1.0#">
  <ont:Concept rdf:ID="中越关系">
    <ont:CLC>D622.5②</ont:CLC>
    <ont:pinyin>zhongyueguanxi</ont:pinyin>
    <ont:topConcept rdf:resource="http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#国际关系"/>
    <ont:broader rdf:resource="http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中外关系"/>
    <ont:hasTerm rdf:resource="http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中国新建关系"/>
  </ont:Concept>
</rdf:RDF>
```

图 4 RDF/XML(OntoThesaurus)

N-Triples^[22]是 W3C 开发的一种面向行的 RDF 序列句法。它展示三元组三个组成部分的完整 URI, 即“<主体 URI> <谓词 URI> <客体 URI>”, 禁止

```
<?xml version="1.0" encoding="GBK">
<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:log="http://www.nlc.gov.cn/2010/06/skos#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:ontc="http://nkos.lib.szu.edu.cn:8080/2010/10/ontc#"
  xml:base="http://nkos.lib.szu.edu.cn:8080/CCT_CT_V1.0#">
  <skos:Concept rdf:ID="中越关系">
  <skos:prefLabel xml:lang="zh">中越关系</skos:prefLabel>
  <skos:closeMatch rdf:resource="http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3"/>
  <ont:CLC>822.3</ont:CLC>
  <skos:altLabel xml:lang="zh-pinyin">zhongyueguanxi</skos:altLabel>
  <skos:topConceptOf rdf:resource="http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#国际关系"/>
  <skos:broaderOf rdf:resource="http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中外关系"/>
  <skos:altLabel xml:lang="zh">中国南越关系</skos:altLabel>
  </skos:Concept>
</rdf:RDF>
```

图 5 RDF/XML(CNKOS)

```
<?xml version="1.0" encoding="GBK">
<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xml:base="http://nkos.lib.szu.edu.cn:8080/CCT_CT_V1.0#">
  <skos:Concept rdf:ID="中越关系">
  <skos:prefLabel xml:lang="zh">中越关系</skos:prefLabel>
  <skos:closeMatch rdf:resource="http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3"/>
  <skos:altLabel xml:lang="zh-pinyin">zhongyueguanxi</skos:altLabel>
  <skos:broaderOf rdf:resource="http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#国际关系"/>
  <skos:altLabel xml:lang="zh">中国南越关系</skos:altLabel>
  </skos:Concept>
</rdf:RDF>
```

图 6 RDF/XML(SKOS)

相对 URI 的引用。如果客体是文字,则表示为用引号括起来的字符串。

图 7 至图 9 分别对应 N-Triples 的三种格式,即 N-Triples (OntoThesaurus)、N-Triples (CNKOS)、N-Triples (SKOS)。

```
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://nkos.lib.szu.edu.cn/2010/10/ontc#Concept>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:closeMatch> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:altLabel> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:broaderOf> <http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#国际关系>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:altLabel> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:altLabel> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
```

图 7 N-Triples (OntoThesaurus)

```
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2004/02/skos/core#Concept>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:closeMatch> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:altLabel> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:broaderOf> <http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#国际关系>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:altLabel> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:altLabel> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
```

图 8 N-Triples (CNKOS)

```
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2004/02/skos/core#Concept>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:closeMatch> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:altLabel> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:broaderOf> <http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#国际关系>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:altLabel> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
<http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系> <http://www.w3.org/2004/02/skos/core#skos:altLabel> <http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3>.
```

图 9 N-Triples (SKOS)

JSON (JavaScript Object Notation) 是一种轻量级的数据交换格式,易于人阅读和编写,也易于机器解析和生成。JSON 采用完全独立于语言的文本格式,但是也使用了类似于 C 语言家族的习惯(包括 C、C++、C#、Java、JavaScript、Perl、Python 等)。这些特性使 JSON 成为理想的数据交换语言^[23]。

图 10 至图 12 分别对应 JSON 的三种格式,即 JSON (OntoThesaurus)、JSON (CNKOS)、JSON (SKOS)。

```
{
  "ont:Concept": "中越关系",
  "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系.json_ont",
  "ont:CLC": "822.3",
  "ont:pinyin": "zhongyueguanxi",
  "ont:topConcept": "国际关系",
  "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#国际关系.json_ont",
  "ont:Concept": "国际关系"
},
{
  "skos:broader": {
    "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中外关系.json_ont",
    "ont:Concept": "中外关系"
  },
  "skos:altLabel": {
    "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中国南越关系.json_ont",
    "ont:Concept": "中国南越关系"
  }
}
```

图 10 JSON (OntoThesaurus)

```
{
  "skos:Concept": "中越关系",
  "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系.json_cnkos",
  "skos:prefLabel": {
    "xml:lang": "zh": "中越关系",
    "xml:lang": "zh-pinyin": "zhongyueguanxi"
  },
  "skos:closeMatch": {
    "uri": "http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3"
  },
  "skos:topConcept": {
    "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#国际关系.json_cnkos",
    "skos:Concept": "国际关系"
  },
  "skos:broader": {
    "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中外关系.json_cnkos",
    "skos:Concept": "中外关系"
  },
  "skos:altLabel": {
    "xml:lang": "zh": {
      "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中国南越关系.json_cnkos",
      "skos:Concept": "中国南越关系"
    }
  }
}
```

图 11 JSON (CNKOS)

```
{
  "skos:Concept": "中越关系",
  "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中越关系.json_skos",
  "skos:prefLabel": {
    "xml:lang": "zh": "中越关系",
    "xml:lang": "zh-pinyin": "zhongyueguanxi"
  },
  "skos:closeMatch": {
    "uri": "http://www.nlc.gov.cn/vocab/CCT_CLC_4.0#822.3"
  },
  "skos:broader": {
    "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中外关系.json_skos",
    "skos:Concept": "中外关系"
  },
  "skos:altLabel": {
    "xml:lang": "zh": {
      "uri": "http://nkos.lib.szu.edu.cn/CCT_CT_V1.0#中国南越关系.json_skos",
      "skos:Concept": "中国南越关系"
    }
  }
}
```

图 12 JSON (SKOS)

6 结语

本文以在《中国分类主题词表》一版数据基础上建立的 CCT1_OTCSS 为例,构建和发布了 Linked Data 服务,详细说明可以参见深圳大学图书馆 NKOS 研究

室网站 2011 年 6 月刊出的“OTCSS Linked Data 服务”^[24],并欢迎登录网站^[25]进行实时测试。这是一项实验性的研究,URI 暂定,笔者希望通过实践来验证 URI 方案和语义描述规范的可行性。

目前已有的实践证明,这套 Linked Data 服务解决方案是可行的,可以投入实用。目前国内已有的 130 余部中文叙词表均可依此方法完成相应的部署。

随着研究的深入,笔者将开发更多的关联数据服务功能。如叙词概念的其他一些属性的关联及可视化、分类法关联数据服务、馆藏书目关联数据服务以及馆藏书目关联数据到传统知识组织系统(如《中国分类主题词表》、《中图法》等)的关联。

- [1] National Library of Sweden. LIBRIS [EB/OL]. [2012 - 04 - 11]. <http://libris.kb.se/>.
- [2] Library of Congress. LC Linked Data Service [EB/OL]. [2012 - 06 - 19]. <http://id.loc.gov/>.
- [3] BBC. BBC Website [EB/OL]. [2012 - 04 - 18]. <http://www.bbc.co.uk/>.
- [4] 曾新红. 受控表语义描述规范(D009-3) [R]. 国家图书馆, 2011. (Zeng Xinhong. Semantic Description Specifications for Chinese Controlled Vocabularies (D009 - 3) [R]. National Library of China, 2011.)
- [5] 曾新红. 中文知识组织系统形式化语义描述标准体系研究(一)——扩展 SKOS 实现传统受控词表全描述 [J]. 中国图书馆学报, 2012, 38 (3): 57 - 68. (Zeng Xinhong. Formal Semantic Description Standard Family for Chinese KOS (1): Realizing Full Description for Chinese Controlled Vocabularies by Expanding SKOS [J]. *Journal of Library Science in China*, 2012, 38 (3): 57 - 68.)
- [6] Berners - Lee T. Linked Data [EB/OL]. [2011 - 12 - 11]. <http://www.w3.org/DesignIssues/LinkedData.html>.
- [7] Duerst M, Suignard M, et al. Internationalized Resource Identifiers (IRIs): Draft - duerst - iri - 11 [EB/OL]. [2012 - 06 - 05]. <http://www.w3.org/International/iri - edit/draft - duerst - iri - 11.txt>.
- [8] W3C. OWL2 Web Ontology Language Quick Reference Guide; W3C Recommendation 27 October 2009 [EB/OL]. [2012 - 06 - 15]. <http://www.w3.org/TR/2009/REC - owl2 - quick - reference - 20091027/>.
- [9] W3C. RDF Primer; W3C Recommendation 10 February 2004 [EB/OL]. [2011 - 11 - 13]. <http://www.w3.org/TR/rdf - primer/>.

- [10] Bizer C, Cyganiak R, Heath T. How to Publish Linked Data on the Web [EB/OL]. [2011 - 11 - 07]. <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/>.
- [11] Bizer C, Heath T, Berners - Lee T. Linked Data - The Story So Far [J]. *International Journal On Semantic Web and Information Systems*, 2009, 5 (3): 1 - 22.
- [12] 潘有能, 张悦. 关联数据研究与应用进展 [J]. 情报科学, 2011, 29 (1): 124 - 130. (Pan Youneng, Zhang Yue. Research and Application Progress on Linked Data [J]. *Information Science*, 2011, 29 (1): 124 - 130.)
- [13] 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. (Bai Haiyan. Linked Data and DBpedia Case Analysis [J]. *New Technology of Library and Information Service*, 2010 (3): 33 - 39.)
- [14] Summers E, Isaac A, Redding C, 等. LCSH, SKOS 和关联数据 [J]. 现代图书情报技术, 2009 (3): 8 - 14. (Summers E, Isaac A, Redding C, et al. LCSH, SKOS and Linked Data [J]. *New Technology of Library and Information Service*, 2009 (3): 8 - 14.)
- [15] 深圳大学图书馆 NKOS 研究室. CCT1_OTCSS [EB/OL]. [2011 - 06 - 13]. <http://nkos.lib.szu.edu.cn:8080/ThesaurusProjectForCCTWL/login.jsp>. (NKOS Research Office of Shenzhen University Library. CCT1_OTCSS [EB/OL]. [2011 - 06 - 13]. <http://nkos.lib.szu.edu.cn:8080/ThesaurusProjectForCCTWL/login.jsp>.)
- [16] 曾新红. 中文叙词表本体——叙词表与本体的融合 [J]. 现代图书情报技术, 2009 (1): 34 - 43. (Zeng Xinhong. OntoThesaurus (Chinese - Thesaurus - Ontology) —— An Integration of Thesaurus and Ontology [J]. *New Technology of Library and Information Service*, 2009 (1): 34 - 43.)
- [17] W3C. Cool URIs for the Semantic Web; W3C Interest Group Note 03 December 2008 [EB/OL]. [2012 - 01 - 06]. <http://www.w3.org/TR/cooluris/>.
- [18] 曾蕾. 关联的图书馆数据 [EB/OL]. [2012 - 03 - 29]. <http://wenku.baidu.com/view/9b26ca136edb6f1aff001f13.html>. (Zeng Lei. Linked Library Data [EB/OL]. [2012 - 03 - 29]. <http://wenku.baidu.com/view/9b26ca136edb6f1aff001f13.html>.)
- [19] W3C. SKOS Simple Knowledge Organization System Primer; W3C Working Group Note 18 August 2009 [EB/OL]. [2010 - 02 - 23]. <http://www.w3.org/TR/2009/NOTE - skos - primer - 20090818/>.
- [20] W3C. SKOS Simple Knowledge Organization System Reference; W3C Recommendation 18 August 2009 [EB/OL]. [2010 - 02 - 23]. <http://www.w3.org/TR/2009/REC - skos - reference - 20090818/>.

- [21] 曾新红. 中文叙词表本体 OntoThesaurus 词汇表 [EB/OL]. [2010-11-11]. <http://nkos.lib.szu.edu.cn/2010/10/ont/>. (Zeng Xinhong. OntoThesaurus Vocabulary [EB/OL]. [2010-11-11]. <http://nkos.lib.szu.edu.cn/2010/10/ont/>.)
- [22] Ogbuji U. Thinking XML:N-Triples 简介 [EB/OL]. [2011-05-05]. <http://www.ibm.com/developerworks/cn/xml/rd/part17/index.html>. (Ogbuji U. Thinking XML:Introduction to N-Triples [EB/OL]. [2011-05-05]. <http://www.ibm.com/developerworks/cn/xml/rd/part17/index.html>.)
- [23] 介绍 JSON [EB/OL]. [2011-05-05]. <http://www.json.org/json-zh.html>. (Introduction to JSON [EB/OL]. [2011-05-05]. <http://www.json.org/json-zh.html>.)
- [24] 深圳大学图书馆 NKOS 研究室. OTCSS Linked Data 服务 [EB/OL]. [2011-06-13]. http://nkos.lib.szu.edu.cn/level_2/service/linkedata.jsp. (NKOS Research Office of Shenzhen University Library. OTCSS Linked Data Service [EB/OL]. [2011-06-13]. http://nkos.lib.szu.edu.cn/level_2/service/linkedata.jsp.)
- [25] 深圳大学图书馆 NKOS 研究室. OTCSS Linked Data 服务网站 [EB/OL]. [2011-01-30]. http://nkos.lib.szu.edu.cn/CCT_CT_V1.0. (NKOS Research Office of Shenzhen University Library. OTCSS Linked Data Service Website [EB/OL]. http://nkos.lib.szu.edu.cn/CCT_CT_V1.0.)
- (作者 E-mail: zengxh@szu.edu.cn)

G. Sayeed Choudhury 获得 2012 年图书馆与信息技术研究奖

G. Sayeed Choudhury 获得 2012 年图书馆与信息技术研究奖 (Frederick G. Kilgour 奖)。该奖项由 OCLC 和美国图书馆协会 (ALA) 图书馆的信息技术分会 (LITA) 共同发起,表彰与信息技术发展相关的研究,尤其是对信息的出版、存储、检索和传播,或者对信息、数据操作和管理程序有积极、实质性影响的工作。

奖项评委会将奖项授予 Choudhury 先生,表彰其在数据管理领域的前沿研究,通过美国国家科学基金会支持的数据保护工程,扩展了研究图书馆服务于研究人员及其所在机构的服务能力;同时也表彰其在图书馆事业领域的持续影响力,将图书馆学原理应用在数字资料的管理和保存。

Choudhury 先生是约翰霍普金斯大学研究数据管理学院副院长、Sheridan 图书馆数据研究和管理中心主任。他也是约翰霍普金斯大学的数据密集型工程科学院 (IDIES) 运营主任,同时还是美国国家科学院研究数据与信息委员会 (BRDI) 和美国密西根大学校际政治及社会研究联盟 (ICPSR) 委员会成员、美国图书馆与信息资源研究理事会高级会员。他是美国国家科学基金会、美国博物馆与图书馆服务学院、安得鲁梅隆基金会和微软研究院共同资助的项目的主要研究者,是数据保存的主要研究者。

Choudhury 说:“我非常荣幸能获得此殊荣。在我的整个职业生涯中,一直很幸运能与 Sheridan 图书馆及约翰霍普金斯大学所有图书馆的这么多优秀的同事共事。而且,我很高兴也很荣幸能与整个领域的同事合作,所以希望能与他们共享这份荣誉,并期待能有更多的机会共同协作,进一步推进图书馆事业的发展。”

2012 年 Frederick G. Kilgour 奖评奖委员会成员有: Nancy Roderer, 约翰霍普金斯大学 (主席); Michael Witt, 普渡大学 (前任主席); Jason Battles, 阿拉巴马大学; Cecilia Schmitz, 奥本大学; David King, LITA 委员会联络人; Roy Tennant, OCLC 联络人。

(编译自: <http://www.oclc.org/news/releases/2012/201220.htm>)

(本刊讯)