

中文知识组织系统形式化语义描述标准体系研究(二) ——分类法共享服务系统 CLSS 研究与实现*

黄华军 曾新红 林伟明 陈若韵

摘 要 本文旨在为国内外分类法拥有者快速实现现有分类法的 Web 版、Web Service 术语服务、Linked Data 化和自动分类等网络共享服务提供参考和支持。以《中国图书馆分类法》第四版为例,采用 CNKOS 对分类法进行语义化描述,使用 Lucene 全文检索引擎、ExtJs 插件、Axis SOAP 引擎和 URL Rewrite 等关键技术实现了分类法共享服务(CLSS)原型系统。经验证明所有功能模块运行和调用都符合预期效果。实践证明该套解决方案可投入使用,而且其他分类法类型的中文知识组织系统均可依此方法快速完成相应的部署。但是,对于部分复杂的自动分类需求,如要获得更为准确的分类号还需更深入的研究或人工辅助。图 5。表 4。参考文献 21。

关键词 分类法 分类法共享服务系统 网络服务 关联数据 自动分类 《中国图书馆分类法》 知识组织系统 CNKOS

分类号 G254.1 G254.29 TP393.4

Formal Semantic Description Standard Family for Chinese KOS (II): Research and Implementation of Classification Scheme Sharing Service System

HUANG Huajun , ZENG Xinhong , LIN Weiming & CHEN Ruoyun

ABSTRACT

In the context of globalization of network information, it is an urgent need to consider networked development and sharing for classification schemes that have already had widespread users in the library field. Since 2000, the international mainstream classification schemes, e. g. DDC, LCC and UDC, have had their Web versions published one after another, while the Web version for CLC (Chinese Library Classification) was also released in 2011. However, there is still a lack of an integrated, reusable and Semantic Web oriented solution for classification schemes and classification-like knowledge organization systems. This paper aims to provide a reference and support for users in the rapid implementation of networked sharing services of the existing classification schemes, including their Web version, Web Service, Linked Data service and automatic classification. It also provides a promising solution that can meet the requirements of both human and machine use.

* 本文系国家社科基金项目“中文知识组织系统形式化语义描述标准体系研究”(编号:12BTQ045)和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及应用研究”(编号:12QNCG11)的研究成果之一。(This article is an outcome of the project “Research of the formal semantic description standard family for Chinese KOS” (No. 12BTQ045) supported by National Social Science Foundation of China and the project “Research of implementation and application of Linked Data service in library” (No. 12QNCG11) supported by Humanities and Social Science Foundation of Shenzhen University.)

通信作者: 曾新红 ,Email: zengxh@szu.edu.cn(Correspondence should be addressed to ZENG Xinhong ,Email: zengxh@szu.edu.cn)

By using computer programs to automatically analyze the source files of the classification schemes in the MARC or HTML format, this paper adopts CNKOS (Common CNKOS Semantic Description Standard) to produce the semantic description of the classification schemes, and employs key technologies, such as the Lucene full-text retrieval engine, the ExtJs plug-in, the Axis SOAP engine, and the URL Rewrite plug-in, to realize classification scheme sharing service and to build a prototype system (known as CLSS). It is verified that, in the case of CLC v4.0, all operations and calls of its functions and modules are in line with the expected results. Constructed as a unified support system for the classification-like Chinese knowledge organization systems, CLSS offers all functionalities required in classification schemes, categorization schemes or standard code tables in CNKOS. It is proven that this solution can be put into practical use, and can facilitate the quick deployment of other classification-like Chinese knowledge organization systems.

The semantic description of classification scheme enables it to offer knowledge service in the Semantic Web environment; the construction of a Web version of classification scheme can promote its networked development and application; the web services provided by classification scheme can facilitate the interoperability among different systems or applications; the implemented Linked Data service of classification scheme can link various resources both inside and outside the library, and allow new resource discovery and access services for users, thereby can greatly improve the knowledge spreading role of the library in the internet era; the automatic classification realized by simulating the artificial classification process can offer effective automatic or computer-aided classification service.

CNKOS is an extension of SKOS. It reveals explicitly a lot of semantics within the classification schemes that used to be implicit. On this basis, this paper adopts international advanced open-source technologies, takes the representative domestic CLC as an example, and constructs the CLSS prototype system that integrates a Web version, automatic classification, Linked Data service and Web Service. The successful implementation of the system verifies the scientificity and practicality of CNKOS. It plays a good demonstrating and promoting role and provides a reusable solution for other domestic classification schemes and classification-like knowledge organization systems to quickly realize their Semantic Web style sharing services. Moreover, the solution can serve as a reference for similar international knowledge organization systems in realizing their Semantic Web style sharing services. For some complex automatic classification requirements, further research or human intervention is required to obtain more accurate results. 5 figs. 4 tabs. 21 refs.

KEY WORDS

Classification scheme. CLSS. Web Service. Linked Data. Automatic classification. CLC. KOS. CNKOS.

0 引言

在网络信息全球化背景下,图书馆界也要考虑拥有广泛用户基础的分类法的网络化发展和共享,并注重自动化和社会化等应用实践,才能在网络信息资源组织中继续发挥重要作用。分类法共享服务旨在提供一种富有生命力,可同时满足人、机器和应用程序需求的解决方案。

本文基于国内传统知识组织系统的特点和发展需要,尝试分类法类型(Classification-like)的中文网络知识组织系统的构建和应用实践,主要介绍分类法共享服务系统(Classification

Scheme Sharing Service System, CLSS)的设计与实现方法,并为利用该系统快速实现国内外现有分类法的语义化描述和网络共享服务提供具体实践经验。CLSS系统是为分类法类型中文知识组织系统开发的统一支持系统,以CNKOS格式^[1-2]进行语义描述的分类法、分类表或标准代码表等可以具备该系统提供的所有功能。

本文的主要目的在于:①快速实现分类法的语义化描述,以适应其在网络环境中提供知识服务的要求;②构建分类法的Web版,促进其网络化发展与应用;③提供分类法的Web Service术语服务,便于不同系统或应用程序之间的互操作;④实现分类法的关联数据(Linked

Data) 化,加快图书馆界构建数据 Web 的步伐;
⑤模拟人工分类过程实现自动分类,尝试开放实验性的自动或辅助分类服务。

1 分类法共享服务系统简介

1.1 分类法与语义化描述的分类法

分类法是图书文献标引和检索的重要工具,作为一个完整的、严谨的知识分类体系,在图书馆界起着无可替代的作用。目前,世界流行的分类法主要有《国际十进分类法》(UDC)、《杜威十进分类法》(DDC)和《美国国会图书分类法》(LCC)等,随着时代的进步,其内容和形式都在不断完善。《中国图书馆分类法》(Chi-

nese Library Classification, CLC)是一部大型综合性分类法,是当今国内图书馆使用最广泛、影响最大的分类法体系,业界简称为《中图法》。

语义化描述的分类法是在分类法的基础上通过引入本体相关理论和技术构建而成的,采用语义 Web 描述语言进行描述。CNKOS^[1-2]是国家图书馆组织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 CK-OS 词汇,以期实现我国传统知识组织系统的全描述。该规范是国家图书馆知识组织标准规范的组成部分。分类法的 CNKOS 本体的构建是其网络化发展和应用的有效途径,采用的具体语义描述规范如表 1 所示。

表 1 分类法的 CNKOS 描述采用的 SKOS 和 CKOS 词汇表①

URI	定义或说明	URI	定义或说明
skos: Concept	概念(类)	ckos: referenceClassEntry	有参考类目(属性)
skos: ConceptScheme	概念体系(类)	ckos: Auxiliary	附表(类)
skos: inScheme	属于概念体系(属性)	ckos: topConcept	顶级概念是(属性)
skos: hasTopConcept	有顶级概念(属性)	ckos: relatedClassNote	类目参照注释(属性)
skos: prefLabel	首选标签(属性)	ckos: stopNote	停用注释(属性)
skos: notation	标记或符号(属性)	ckos: useClassEntry	改入类目(属性)
skos: note	通用注释(属性)	ckos: altClassEntryOf	是……的交替类目(属性)
skos: scopeNote	范畴注释(属性)	ckos: altClassNote	交替类目注释(属性)
skos: historyNote	历史注释(属性)	ckos: classEntryType	类目类型(属性)
skos: related	相关概念(属性)	ckos: classGuideNote	相关类目注释(属性)
skos: mappingRelation	映射关系(属性)	ckos: combineNote	复分注释(属性)
skos: closeMatch	近似映射(属性)	ckos: combineFrom	有复分表(属性)
skos: relatedMatch	相关映射(属性)	ckos: notationSpan	类号范围(属性)
skos: broader	直接上位概念(属性)	ckos: notationCommon	共有类号(属性)
skos: narrower	直接下位概念(属性)	ckos: notationBegin	起始类号(属性)
ckos: nonCloseMatch	非近似映射(属性)	ckos: notationEnd	终止类号(属性)

① 此表列出的词汇是本项目中目前应用到的,有关分类法类型中文知识组织系统语义化描述词汇的详细信息见参考文献[1]。

本文以《中图法》第四版为例,以 CLCMARC 格式数据或 HTML 格式电子版数据为数据源,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的基础上,通过源数据与 SKOS 或 CKOS 词汇的对应关系,使用计算机程序自动将其转换成 CNKOS 本体,即实现了《中图法》的形式化语义表示。具体转换方案见 2.2 节。

1.2 分类法共享服务系统

1.2.1 国内外研究现状

自从 20 世纪末世界各主流分类法陆续推出电子版以来,分类法版权所有人经过持续努力,分别于 21 世纪初推出各自的网络版。例如,DDC 版权所有人 OCLC 于 2000 年率先开放 Web 版服务 Web Dewey^[3],紧随其后,2001 年,另外两部国际著名的分类法 LCC 和 UDC 也先后发布了网络版的 Classification Web^[4]和 UDC On-line^[5]。曹树金等^[6]对较早的 DDC、LCC 和 UDC 网络版给予了综合评析,对三者提供的主要功能做了详细的比较。

近年来,国内基于计算机网络技术对构建分类法网络版乃至本体版的知识组织系统进行了大量研究^[7-11],一致认为分类法应朝着网络化方向发展。但是,分类法的网络化实践应用却较少,更不用说提供成熟的知识服务了。为了弥补这些不足,借鉴国外分类法的先进成果

和应用经验,国家图书馆于 2011 年 12 月首次在互联网上发布了《中图法》(第五版)的 Web 版^[12],测试期从发布之初的 6 个月延长至 2012 年 9 月 1 日。刘丽静^[13]详细解析了《中图法》(第五版)Web 版的功能,并提出个人的建议和展望。相比 DDC、LCC 和 UDC 的 Web 版,CLC Web 版未提供分类号的自动构建服务。

分类法 Web 版的发布只是其走向网络化的第一步,更广泛的共享还需提供 Web Service 术语服务、Linked Data 服务和自动分类服务等,这些功能对于使传统知识组织系统走出相对封闭的图书馆,实现 Web 服务集成和共享,以及链接图书馆内外的信息资源尤为重要。

本文研究的 CLSS 系统是集分类法 Web 版、Web Service 术语服务、关联数据和自动分类等功能于一体的网络共享平台。本文提出的解决方案针对所有分类法类型的中文 KOS,希望为国内外相关团体和个人构建各类分类法类型中文 KOS 的网络化应用提供参考和支持。

1.2.2 分类法共享服务系统总体结构

CLSS 先将已有的分类法(这里以《中图法》第四版为例)转换为 CNKOS 本体,再使用 Lucene 全文检索引擎工具包构建索引,通过基于 J2EE 平台的 B/S 模式实现网络化共享服务,包括提供分类法的 Web 版、Web Service 术语服务、关联数据和自动分类等网络共享服务。系统总体结构如图 1 所示。

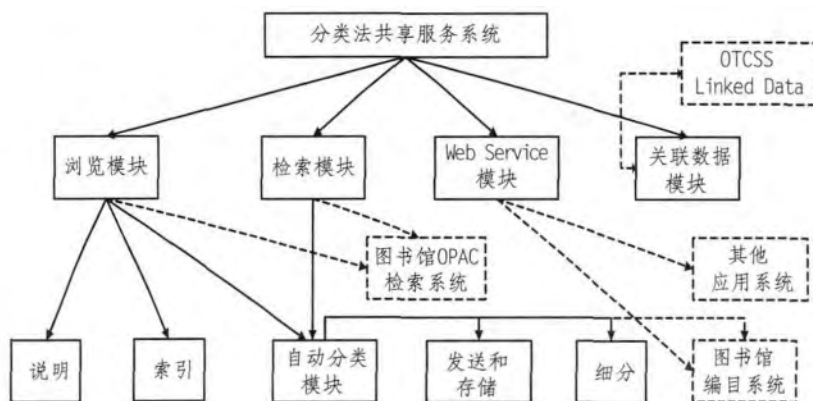


图 1 分类法共享服务系统 (CLSS) 总体结构图

2 分类法的 Web 版

所谓分类法的 Web 版,主要是指在 Web 上浏览、检索、获取甚至更新分类法版本^[6]。目前,不管是 DDC、LCC 和 UDC 还是 CLC 提供的 Web 版服务都需要付费,一般情况下,网络用户通过申请可以免费获得一段时间的试用期。CLSS 的解决方案面向所有分类法类型的中文 KOS,拟提供面向公众的免费服务。本文以《中国图法》第四版为例,简介其功能。

2.1 CLSS 的 CLC Web 版简介

2.1.1 分类法浏览和检索

CLSS 的 CLC Web 版涵盖的内容非常丰富,包括主表、专类复分表、通用复分表,既有印刷版中类目的静态内容和排版格式,也有全部类目的关联数据表示。以关联数据形式表示分类法类目的详细信息是本文描述的 CLC Web 版的一大亮点,提供与《中国分类主题词表》(简称《中

分表》)二版的各种映射关系,集中体现了分类主题一体化思想。另外,分类法类目注释中的类号、参照类号和各种相关类号都提供超链接服务,拥有便利的类号指引和操作,突破分类法传统的“线性”资源组织方式,实现网状的浏览和检索路径。浏览显示方式多样,既可以“浏览”显示整部分类法,又可以从各大类和附表的索引进入浏览每个部分。

在分类法网络版中,人性化的界面,简单直观,用户通过浏览可以清楚地了解分类法的类目层级关系。Web 版提供了多样化的检索途径,检索方式包括任意途径、类号、类名和注释四种,匹配方式包括任意一致、前方一致和精确一致三种,检索结果按系统给出的排序规则以清单形式列出。此外,Web 版还提供了一般关键词到正式主题词的映射,尽量避免漏检,保证查全率,提高分类标引的准确率。在“浏览”和“检索”中都可以点击相关类目显示其详细信息和操作界面(如开放类号或类名进行 OPAC 检索等诸多功能,见图 2)。

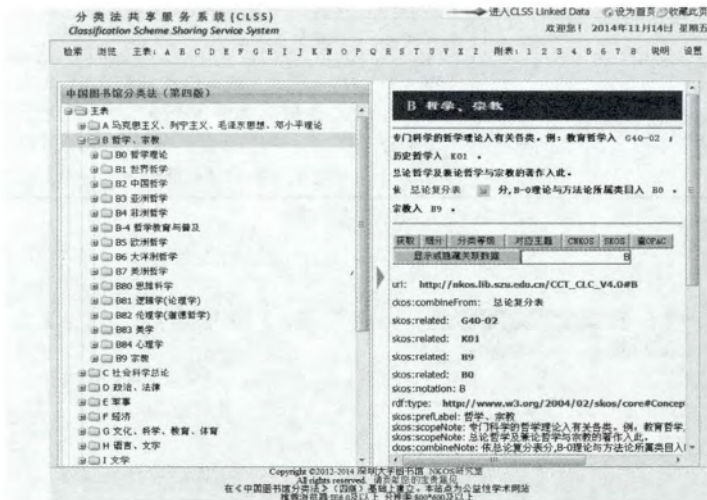


图 2 CLSS 的 CLC 第四版网络版示意图

2.1.2 分类标引

本文所述的 CLC Web 版提供标引服务。在“检索”界面输入关键词进行检索,得到检索结

果列表,从中选定某一类目,或在“浏览”界面展开分类法选定某一类目,右边都会出现该类目的详细信息和操作界面,默认标引为选中类号

或细分生成的组配类号,点击“获取”按钮,实现分类号的自动发送和存储。

把关键词与分类号的对应关系信息复制到剪贴板中,便可辅助计算机编目系统进行分类标引工作,并可以避免编目人员的录入错误。此外,计算机编目系统还可直接调用 CLSS 提供的 Web Service 术语服务 API 实现自动标引。在分类标引具体实践过程中,可以根据系统需求定制所要发送数据的格式,提高标引效率,减轻工作人员负担。把关键词与分类号的对应关系信息存储在数据库中,能够使日后的分类标引工作更加快捷。

2.1.3 类号自动构建服务

类号的自动构建是用户在系统界面上构造新的分类号,即模拟编目人员利用分类法中已有类目的号码,按照一定规则组合复合类号,用以表达类表中没有设置的专指类目,以提高对

新主题的描述能力^[14],这是分类法 Web 版编制中的一个难点。关于类号的自动构建,本文实现了两种情况:利用通用复分表的细分,以及利用专类复分表的细分。

本文的类号自动构建服务可以利用计算机程序自动计算主表和复分表组配后的类号,以表 2 的两条记录为例,自动构建的操作步骤如图 3 所示,计算机程序根据细分规则自动计算出准确的组配类号,通过调用分类法的 Web Service 术语服务中的自动分类 API 将得到相同的结果。DDC、LCC 和 UDC 的 Web 版都提供类号构造工具,目前国家图书馆的《中图法》第五版 Web 版还不能直接计算组配类号,毕竟与国际主流分类法 Web 版相比,起步较晚,但也不排除日后为用户提供从已有分类号中构建新的分类号的可能性。

表 2 细分操作随机测试数据

题名	主分关键词	细分关键词	细分规则	组配类号
维吾尔族民间工艺美术制品	民间工艺美术制品	维吾尔族	民间工艺美术制品→TS938,主表未注明依中国民族表分,但需要进行民族复分时,在中国民族子目维吾尔族类号“15”前加中国民族号“2”,即“215”。	TS938 “215”
影响的焦虑:一种诗歌理论(美)	各国文学	美国诗歌	各国文学→I3/7,该书为美国现代诗歌评论,故先依世界地区表分,再依专类复分表分。细分后的分类号为 I712.072,其中 712 为世界地区表中美国的复分号,072 为专类复分表中诗歌评论的复分号。	I712.072



图 3 CLSS 的类号自动构建示意图

2.2 分类法 Web 版的具体实现

本文的 CLC Web 版,具体实现方法和技术由以下几个方面构成。

(1) 准备数据源:在本研究室相关研究成果及国家图书馆知识组织标准规范^[1-2]的基础上,利用计算机程序将《中图法》第四版 HTML 格式电子档或 CLCMARC 格式数据自动转换成 CNKOS 格式的 RDF 本体文件。这里以 CLC-MARC 格式数据语义化为例,首先从 CLC 的 MARC 格式数据中提取每条类目的 MARC 数据记录,将每条记录转化成字节数组。依据 MARC 数据的特点,依次获取头标区和目次区,以目次区的每个目次项提供的每个数据字段的起始位置和长度,循环提取数据区各数据字段信息,通

过子字段标识提取数据字段中各子字段信息,防止信息丢失,最后根据 CLCMARC 格式与 SKOS 或 CKOS 词汇的对应关系生成每个类目及类目之间的 CNKOS 语义数据。由于源数据中包含《中分表》二版的信息,因此自动生成的 CNKOS 本体中存在 CLC 和 CCT 的关联,即实现分类主题一体化。

从 CLCMARC 格式到 SKOS 或 CKOS 词汇的转换需要复杂的判断和处理,以 250 字段为例,具体转换规则见表 3。在转换过程中需要充分考虑的问题还有:通用复分表和专类复分表的 URI 解决方案;特殊符号的表示和处理,如“/”、“+”和“-”等;防止关联字段之间的语义信息丢失及根据已有信息补齐相关语义数据等。

表 3 CLCMARC 250 字段与 CNKOS 词汇的对应转换规则

字段或子字段标识	语义规范	相关说明
250 类号字段	skos: Concept	• \$ a 构成概念的 URI 的标识。
\$ d 类级	skos: hasTopConcept (\$ d = 01)	• \$ d 的值为“01”时,该类为顶级概念。
\$ a 单一类号或起始类号的起号	skos: notation(当存在 \$ c 时,转为 ckos: notationCommon 或 ckos: notationBegin)	• \$ a 为单一类号(即不存在 \$ c)时,转为该类目概念的 skos: notation 的值(客体)。
\$ c 起始类号的尾号	与 \$ a 合并生成 ckos: notationSpan, 去掉前面相同部分(若为指示性类目,同时转为 ckos: notationEnd)	
\$ 9 上位类号(MARCXML)	skos: broader 的 skos: notation(CLCMARC 无)	• 如果没有上位类号 \$ 9, 只记录 \$ h 上位类号,最后一个 \$ h 应是上位类目概念的 skos: prefLabel, 而类名没有唯一性,需通过计算机程序定位到 skos: broader, 即上位类号。
\$ h 类号对应的上位层次类名	skos: broader 的 skos: prefLabel(最后一个 \$ h)	\$ k 也同样有此问题。
\$ k 概略类号的上位层次类名	ckos: referenceClassEntry 的 skos: prefLabel	• \$ j 转为 skos: prefLabel 的值。
\$ j 类号直接对应的(类名)	skos: prefLabel	• 有 \$ z 的是通用复分表中的类目。
\$ z 附表标识	skos: inScheme	附表名称(skos: prefLabel) 及其 URI、附表中概念的子类类型转换前由支持系统预先确定。转换时根据 \$ z 的值生成相应的 skos: inScheme 描述语句。
fb1 一、总论复分表	当存在 \$ c 时,与 notationSpan 相关的描述按以下方式生成:将 \$ a 和 \$ c 合并为 ckos: notationSpan, 合成时中间加“/”并去除 \$ c 前面与 \$ a 相同的大类号,如 B21/26。当 100 \$ a/11 = a(使用类目)时,将 \$ a 和 \$ c 前面共同的部分转为: ckos: notationCommon。当 100 \$ a/10 = d(指示性类目)时,将 \$ a 转为 ckos: notationBegin, 将 \$ c 转为 ckos: notationEnd。	
fb2 二、世界地区表		
fb3 三、中国地区表		
fb4 四、国际时代表		
fb5 五、中国时代表		
fb6 六、世界种族与民族表		
fb7 七、中国民族表		
fb8 八、通用时间、地点表		

(2) 数据处理: CLC 本体文件采用以属性为中心的 RDF 三元组思想组织后,以 Lucene 全文

检索引擎构建其三元组索引数据库,具体实现方法见参考文献[15];从 Web 界面人工添加或

由计算机程序自动批量导入一般常用关键词到正式主题词的映射。

(3) 开发和实验环境: 本文所有服务均采用 Eclipse 和 Tomcat 开发环境, 其中, 分类法界面主要采用带有 jstree 和 ExtJs 插件的 jsp 页面实现, 主要业务功能使用 Struts 框架并以 Java 编写完成, 系统相关配置和参数信息、分类标引数据等使用 MS SQL Server 数据库存储。后文讲述的 Web Service 和 Linked Data 服务分别采用 Axis SOAP 引擎和 URL Rewrite 插件进行开发。

(4) 除了分类查重外, 全部检索功能均使用 Lucene 的全文检索实现, 具体实现方法见参考文献[15]。

CLSS 的 CLC Web 版为公众提供了一个免费的开放式共享服务平台, 界面简单直观, 用户体验友好, 相较电子版而言, 无需安装软件, 不用存储设备且免于携带, 其他分类法均可依此方案快速完成相应的部署。

3 分类法的 Web Service 术语服务

分类法的 Web Service 术语服务是指对分类法进行浏览、查询和应用的各種 Web 服务的统称。Web Service 术语服务通过 Web 应用程序接口(API) 支持机器对词表的访问和调用, 是在网络环境下对词表进行应用的重要途径^[16]。

为进一步提供面向机器、系统或应用程序

的 Web Service 术语服务, 我们开发了一系列 Web Service API, 按需提供弹性服务。本文实现的分类法 Web Service 术语服务主要用于满足机器、系统或应用程序的需要, 未来可以通过术语服务注册的方式集中提供各种分类法本体的 API, 供用户自由选择调用, 系统中列出、描述并指明在 Web 服务中可利用的分类法的集合, 每种分类法通过 URI 进行识别, 从而促进分类法的发现、重用和互操作。

根据当前分类法网络应用的实际情况, 本文提供的 Web Service API 集合在 NKOS 网站^①中列出, 通过 Web Service 调用的标准方法, 所有应用都可以在不同的应用平台上实现^②。本节通过对深圳大学图书馆部分无需进行细分的图书文献进行实验, 调用分类法的 Web Service 术语服务中的自动分类 API, 基本上都能得到预期的分类效果。对于需要细分的情况, 笔者随机抽取表 4 所示两本图书文献进行验证, 得到的分类结果如图 4 所示。实践证明本文实现的分类法 Web Service 术语服务可支持计算机编目系统为已有关键词标引的图书文献、期刊、学位论文或其他信息资源自动赋予分类号。图书馆信息管理系统、数据库检索系统、搜索引擎、标签系统等, 无论基于什么平台开发, 采用何种开发语言, 都可以使用该套 API 提供的服务, 实现基于分类法的智能检索、分类标引、知识链接和知识服务。

表 4 图书文献自动标引随机测试数据

登录号	题名/著者	主分关键词	细分关键词	组配类号
A0645023	美国专利与科技纵横/马秀山编著	各国专利	美国	G306.771.2
WX001348/55	宋元方志丛刊/中华书局编辑部编	各代总志	宋	K290.44

注: 示例的细分规则为: 美国专利→各国专利文献概况→G306.7(类目下注释: 依世界地区表分, 美国: 712, 分类号: G306.771.2); 宋元方志→各代总志→K290.1/.7(类目下注释: 依中国时代表分, 宋: 44, 组配后的分类号: K290.44)。

① http://nkos.lib.szu.edu.cn/level_2/service/webservice.jsp。

② 为测试系统功能和获取更多的需求信息, 现提供实验性开放服务, 访问地址: <http://nkos.lib.szu.edu.cn:8080/ProjectForCLC/services/CLCService/getClassEntryType?schemeURI={0}&conceptURI={1}>。具体实现技术和调用方法见参考文献[17]。

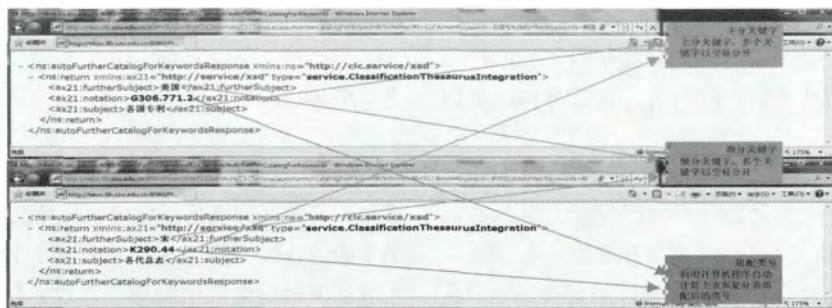


图4 调用 CLSS 的 Web Service API 实现自动分类示意图

4 分类法的 Linked Data 服务

随着信息资源向网络环境的迁移,数据 Web 逐渐成为热门话题,为满足用户随时进行信息消费的需求,分类法走出相对封闭的图书馆成为数据 Web 的一部分显得格外重要。2011 年 1 月,深圳大学图书馆 NKOS 研究室尝试发布了 OTCSS Linked Data 实验性服务^[18]。通过近三年多来的实践证明,我们提出的解决方案^[19]为图书馆传统中文 KOS Linked Data 化提供了一

种有效途径,初步验证了国家图书馆知识组织标准规范的科学性,以及帮助版权机构构建和发布其关联数据的可行性。为进一步验证其成熟度和可靠性,本文继续采纳该解决方案实现分类法的 Linked Data 服务,并将 Linked Data 化的 CLC 和 CCT 互联。目前,与馆藏书目的相互关联正在建设中。

根据分类法 CNKOS 本体的需求和特点,遵循关联数据创建四原则,本文实现了 CLC 的 Linked Data 服务,主要提供以下功能(见图 5)。

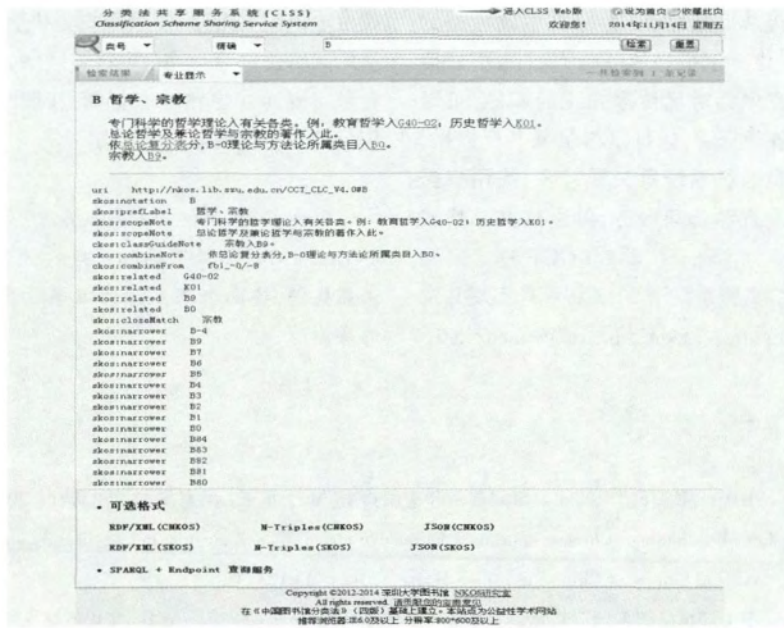


图5 分类法 Linked Data 服务示意图

(1) 为分类法本体的所有类目(即所要发布的资源)提供 HTTP URI, 以此作为类目的唯一标识, 通过此 HTTP URI, 可以定位具体的类目信息。

(2) 提供地址栏和检索框两种方式输入 URI 或类号(类名)查询关联数据, 类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, 可以提供该对象很多有意义的信息(关联对象), 如类目信息(类名、类号、映射主题、各种注释、上位类、下位类、相关类等)以及 URI、可选下载格式等信息, 并可通过关联对象的 RDF Links 查询其详情; 提供到其他词表相似概念的关联(如到《中分表》二版对应叙词款目的映射链接), 实现《中图法》和叙词表的相互关联。

(4) 在可选格式中提供 CNKOS 和 SKOS 的三种描述格式(RDF/XML、N-Triples 和 JSON)的下载或在线浏览。

(5) 进一步提供 SPARQL Endpoint 查询服务, 方便用户定制 CLC 或 CCT 中所需的机器可处理的数据格式。

分类法的关联数据可以为网络资源建立资源之间的直接或间接联系, 为分类主题数据的机读化、Web 化奠定很好的基础^[20]。为更好地发挥图书馆在网络时代传播知识的职能, 可以链接图书馆各种资源, 如可以与馆藏书目数据、其他传统知识组织系统等关联起来, 为用户提供新的资源发现和访问服务, 可通过共享格式和 SPARQL 查询实现与机器或应用程序之间的交互式服务。实验系统于 2013 年 4 月实现并发布, 登录地址: http://nkos.lib.szu.edu.cn/CCT_

CLC_V4.0。

5 结语

我们以《中图法》第四版为例, 在与叙词表建立相互关联的前提下实现了分类法的网络化共享服务, 这是一项实验性的研究, 目前已有的实践证明我们提供的服务可以投入使用, 欢迎登录网站^[21]进行实时测试。

CLSS 系统是为分类法类型中文知识组织系统开发的统一支持系统, 该系统的实现验证了 CNKOS 语义描述方案^[1-2]的科学性和有效性, 以 CNKOS 格式进行语义描述的分类法、分类表或标准代码表等可以拥有该系统提供的所有功能。

随着研究的深入, 我们将继续改进系统服务, 加强网站的建设, 在系统实际应用中加大与网络用户的交流, 链接更多的 Web 资源(如馆藏书目信息等), 或者可提供与国际主流分类法的各种映射等, 加强不同系统或应用程序间的共享和集成。

我们应该顺应网络信息时代的要求, 让分类法走出图书馆, 为广大用户提供更便捷、更先进的网络化共享服务, 并利用网络用户的集体智慧实现其共建和动态更新, 让图书馆界的传统分类法焕发出新的生命力。

致谢: 本文所使用的《中图法》第四版 B 大类 CLCMARC 数据及其相关格式说明由国家图书馆提供, 特向卜书庆老师及其团队表示衷心的感谢!

参考文献

- [1] 王军, 卜书庆. 网络环境下的知识组织规范和应用指南[M]. 北京: 国家图书馆出版社, 2012: 71-123.
(Wang Jun, Bu Shuqing. Chinese networked knowledge organization standards and application guidelines[M]. Beijing: National Library of China Publishing House, 2012: 71-123.)
- [2] 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(3): 57-67. (Zeng Xinhong. Formal semantic description standard family for

- Chinese KOS (I): realizing full description for Chinese controlled vocabularies by expanding SKOS [J]. *Journal of Library Science in China*, 2012(3): 57 – 67.)
- [3] OCLC. Dewey[®] services [EB/OL]. [2014 – 02 – 23]. <http://www.oclc.org/dewey>.
- [4] LCC. Classification Web [EB/OL]. [2014 – 02 – 26]. <http://classificationweb.net>.
- [5] UDC. BSI UDC web online [EB/OL]. [2014 – 02 – 26]. <http://www.udc-online.com>.
- [6] 曹树金, 颜丽君, 汪东波. DDC、LCC、UDC 网络版评析 [J]. *中国图书馆学报*, 2002, 28(6): 102 – 107. (Cao Shujin, Yan Lijun, Wang Dongbo. Common on the network versions of DDC, LCC, UDC [J]. *Journal OF Library Science in China*, 2002, 28(6): 61 – 65.)
- [7] 张忠秋. 近年来我国分类表与主题词表研究与实践进展综述 [J]. *图书馆工作与研究*, 2012(11): 102 – 107. (Zhang Zhongqiu. Progress of researches and practices on classification and thesaurus in recent years in China [J]. *Library Work and Study*, 2012(11): 102 – 107.)
- [8] 黄晓倩. 网络环境下分类法改进初探 [J]. *图书馆建设*, 2000(2): 40 – 42. (Huang Xiaolian. Research on the improvement of the classification under the internet environment [J]. *Library Development*, 2000(2): 40 – 42.)
- [9] 柳晓春, 左少凝. 传统图书馆分类法的网络适应性研究 [J]. *情报科学*, 2001, 19(12): 1269 – 1271. (Liu Xiaochun, Zuo Shaoning. Research on suitability of using library classification schemes for internet resources [J]. *Information Science*, 2001, 19(12): 1269 – 1271.)
- [10] 潘宇光. 《中图法》的网络化趋势 [J]. *图书馆学刊*, 2009(12): 92 – 94. (Pan Yuguang. *Chinese Library Classification network* [J]. *Journal of Library Science*, 2009(12): 92 – 94.)
- [11] 何琳, 侯汉清. 《中国图书馆分类法》在网络环境中的适应性改造研究 [J]. *图书情报工作*, 2010, 54(19): 6 – 128. (He Lin, Hou Hanqing. Research on the adaptation reform of *Chinese Library Classification* for network environment [J]. *Library and Information Service*, 2010, 54(19): 6 – 128.)
- [12] 国家图书馆. 《中图法》电子版 [EB/OL]. [2014 – 01 – 16]. <http://clc.nlc.gov.cn/ztfdzweb.jsp>. (National Library of China. Electronic version of *Chinese Library Classification* [EB/OL]. [2014 – 01 – 16]. <http://clc.nlc.gov.cn/ztfdzweb.jsp>.)
- [13] 刘丽静. 《中国图书馆分类法》(第5版) Web 版功能解析 [J]. *图书馆建设*, 2012(9): 33 – 38. (Liu Lijing. Functional analysis of *Chinese Library Classification* (5th edition) web edition [J]. *Library Development*, 2012(9): 33 – 38.)
- [14] 国家图书馆《中国图书馆分类法》编辑委员会. 《中国图书馆分类法》第五版使用手册 [M]. 北京: 国家图书馆出版社, 2012: 93. (*Chinese Library Classification* Editorial Board of National Library of China. *Chinese Library Classification fifth edition user guide* [M]. Beijing: National Library of China Press, 2012: 93.)
- [15] 曾新红, 黄华军, 林伟明. 超大型中文叙词表本体的检索与推理研究 [J]. *现代图书情报技术*, 2010 (7/8): 58 – 65. (Zeng Xinhong, Huang Huajun, Lin Weiming. Research on retrieval and reasoning of ultra – large – scale OntoThesaurus [J]. *New Technology of Library and Information Service*, 2010(7/8): 58 – 65.)
- [16] 欧石燕, 虞为. 网络知识组织系统注册与术语服务规范(D009 – 1) [R]. 国家图书馆, 2011. (Ou Shiyan, Yu Wei. NKOS system register and terminology service specifications(D009 – 1) [R]. National Library of China, 2011.)
- [17] 林伟明, 曾新红. OntoThesaurus Web Service API 及其应用研究 [J]. *图书情报工作*, 2010, 54(2): 119 – 122. (Lin Weiming, Zeng Xinhong. OntoThesaurus Web service API and its application [J]. *Library and Information Service*, 2010, 54(2): 119 – 122.)
- [18] 深圳大学图书馆 NKOS 研究室. OTCSS Linked Data 服务 [EB/OL]. [2011 – 01 – 30]. http://nkos.lib.szu.edu.cn/CCT_CT_V2.0. (NKOS Research Office of Shenzhen University Library. OTCSS linked

- data service[EB/OL]. [2011-01-30]. http://nkos.lib.szu.edu.cn/CCT_CT_V2.0.)
- [19] 黄华军, 曾新红, 林伟明. OTCSS 关联数据服务的研究与实现[J]. 现代图书情报技术, 2012(7/8): 40-47. (Huang Huajun, Zeng Xinhong, Lin Weiming. Research and implementation about linked data service of OTCSS[J]. New Technology of Library and Information Service, 2012(7/8): 40-47.)
- [20] 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>.)
- [21] 深圳大学图书馆 NKOS 研究室. 分类法共享服务系统[EB/OL]. [2014-03-05]. <http://nkos.lib.szu.edu.cn:8080/ProjectForCLC>. (NKOS Research Office of Shenzhen University Library. Classification scheme sharing service system[EB/OL]. [2014-03-05]. <http://nkos.lib.szu.edu.cn:8080/ProjectForCLC>.)

黄华军 深圳大学图书馆馆员。通信地址: 广东省深圳市南山区深圳大学图书馆。邮编: 518060。

曾新红 深圳大学图书馆研究馆员, NKOS 研究室主任。通信地址同上。

林伟明 深圳大学图书馆馆员。通信地址同上。

陈若韵 深圳大学图书馆副研究馆员。通信地址同上。

(收稿日期: 2014-10-09)