

# 基于GWT的开放知识组织 集成系统设计\* 与实现\*

杨锐<sup>1</sup>, 刘毅<sup>1</sup>, 宋文<sup>2</sup>, 周子健<sup>1</sup>, 张士男<sup>2</sup>

(1. 中国科学院武汉文献情报中心, 武汉 430071; 2. 中国科学院文献情报中心, 北京 100190)

**摘要:** 针对网络环境下开放知识组织体系的应用研究, 提出以GWT开源技术框架为基础构建开放知识组织集成系统的整体设计方案, 然后进行开放知识组织引擎构建以及开放知识组织体系的导入、浏览、检索和编辑等主要功能实现。系统通过映射形成统一CAS词表, 较好的实现了不同开放知识组织体系的有效融合, 为第三方系统知识服务提供了有力支撑。

**关键词:** 知识组织体系; 开放知识组织引擎; GWT

**中图分类号:** G250

**DOI:** 10.3772/j.issn.1673-2286.2015.05.004

## 1 引言

随着互联网技术的快速发展, 传统知识组织体系正在向数字化、系统化的集成方向靠拢。在数字资源环境下可通过对传统知识组织体系的高效集成实现对网络信息资源的深度挖掘和分析。传统知识组织体系包括各种类型的综合或专业的知识组织体系, 例如用复杂关系结构表示的本体、用等级结构表示的分类体系和主题词表, 以及其它术语表、规范文档、受控词表等<sup>[1]</sup>。在网络环境下通过对传统知识组织体系使用映射、重组和融合等方法能够快速有效地将它们有机集成整合起来应对网络环境下知识多样化、精准化的实际需求。

## 2 系统定位分析

目前知识组织体系的有效整合和利用一直受到国内外学术界的重视, 其中国外在对多种类型的分类法、主题词表、受控词表和术语表集成等方面实施了很多

积极有效的研究工作, 且一些研究成果在实践应用中取得较好的效果。目前欧美等国家在该领域的研究一直处于全球领先地位, 许多关键技术已经通过研究阶段, 迈向社会化和商业化的应用阶段。例如欧盟信息社会技术资助的Renardus项目其宗旨是通过统一界面为欧洲高等教育的教学和科研人员提供有选择、高质量的因特网资源信息。Renardus项目以DDC分类法为基础, 在各种知识组织体系映射和DC资源类型映射的基础上, 建立集成的欧洲学术主题门户信息集成访问服务, 改进了对欧洲现有学术信息的访问方式<sup>[2]</sup>。美国国会图书馆和OCLC分别以国会图书馆分类法和杜威分类法为基础, 尝试对各种信息资源自动进行学科主题分类, 支持对大规模信息对象的主题搜索和按学科主题进行组织的发现机制<sup>[3,4]</sup>。

结合国内目前还没有较为成熟完善的知识组织体系集成系统现状以及中科院文献情报服务不断向知识组织和知识发现的新型科研知识环境转变, 中科院文献情报中心提出在数字化网络化综合科技信息环

\* 本研究得到国家科技支撑计划“科技知识组织体系共享服务平台建设”项目子课题“科技知识组织体系 (STKOS) 的开放查询和推理接口建设” (编号: 2011BAH10B03-5) 资助。

境中构建知识组织体系集成系统 (Open Knowledge Organization System, OpenKOS) 的建设思路, 利用开放知识组织引擎支持知识服务和知识创造力聚合的科研知识环境。整个系统定位包括:

(1) 支持基于OpenKOS的专业领域知识环境、综合科技信息资源和公共文献资源系统的资源组织和集成揭示。利用系统的开放知识组织引擎允许用户根据知识环境建设需要, 动态更新公共知识组织体系。

(2) 利用系统提供的开放接口, 允许第三方系统动态调用公共知识组织体系, 支持针对开放知识组织体系结构的呈现和资源内容的浏览, 支持第三方系统对数字知识资源的集成检索和发现。

(3) 允许用户构建独立的知识组织体系, 通过输入外部广泛应用的综合知识组织体系, 利用映射、关联和合并等方式形成专业领域公共知识组织体系。

(4) 系统能够提供开放的知识组织体系裁切、选择和抽取工具, 用户可以使用这些工具, 在公共知识组织体系中选择符合其应用需要的知识组织体系子集, 通过标准格式输出。

### 3 系统技术路线

结合以上国内外研究现状, 分析系统技术路线, 主要包括系统技术架构、开放知识组织引擎和开放知识组织体系等方面。

#### 3.1 系统技术架构

整个OpenKOS系统基于Google Web Toolkit (GWT) 技术架构。GWT允许开发者使用Java编程语言快速构建和维护复杂的高性能JavaScript前端应用程序, 能够降低系统的开发难度。系统开发涉及的软件和工具包括: Java、Eclipse、MySQL和WebProtege。系统使用Google Web Toolkit为整个应用系统构建

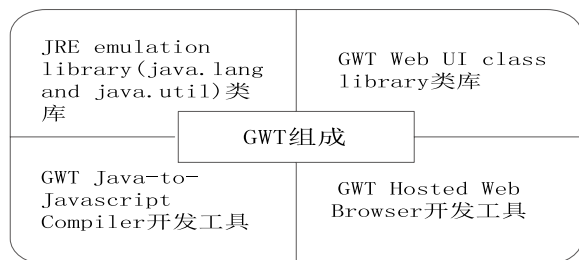


图1 GWT框架组成结构

Widget, 通过Java语言编写Ajax前端, 然后由GWT交叉编译到优化的JavaScript中, 通过JavaScript自动在所有主流浏览器上运行<sup>[5]</sup>。GWT主要由四个部分组成, 包括: Java-to-JavaScript Compiler、Hosted Web浏览器和两套Java类库, 如图1所示。

#### 3.2 开放知识组织引擎

在OpenKOS系统中利用GWT的交互式接口构建开放知识组织引擎实现客户端与服务器的通信。通过开放知识组织引擎进行系统页面加载, 加载过程不向服务器请求页面内容, 所有的转换都在客户端进行, 同时数据向服务器提交或者从服务器获取, 开放知识组织引擎负责整个系统的功能调用、执行以及结果的反馈, 通过这种方式保证数据或者服务的异步处理。开放知识组织引擎在服务器端负责处理数据的对象在系统中定义为Service, 每个Service由三种类型的类组成, 分别是开放知识组织引擎服务方法定义接口 (Service), 开放知识组织引擎异步调用接口 (ServiceAsync) 和开放知识组织引擎服务方法实现类ServiceImpl<sup>[5]</sup>。开放知识组织引擎的核心类各部分之间的关系如图2示。

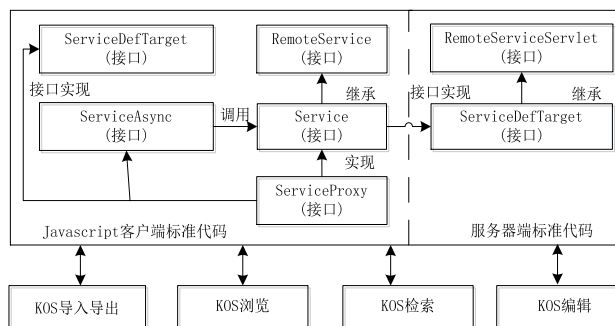


图2 OpenKOS引擎核心类设计

#### 3.3 开放知识组织体系

根据OpenKOS系统的定位, 整个系统围绕开放知识组织体系集成展开, 开放知识组织体系类型包括: 受控词表 (主题词表、叙词表、术语表)、分类法和本体, 采用通用知识描述语言描述。在OpenKOS系统中各种类型的受控词表 (主题词表、叙词表、术语表) 和分类法采用符合SKOS规范的RDF语言, 本体概念模型采用标准的OWL/XML或者RDF/XML格式, 字符集编码采用UNICODE, 语义规则表达采用SWRL语言<sup>[6]</sup>。

开放知识组织体系在OpenKOS系统中的任何一个概念、属性、词汇都被看作为一个Web资源,分配唯一的URI,通过URI进行系统标识并实现多个开放知识组织体系之间的映射,各种类型的开放知识组织体系都有一个DDC分类号,均要与DDC进行关联映射,并组织在DDC的分类体系框架下<sup>[7,8]</sup>。OpenKOS系统中的术语按照Berners-Lee的关联数据四原则发布,术语和概念通过URI访问,通过URI被其它系统引用<sup>[9,10]</sup>。开放知识组织体系结构如图3所示。

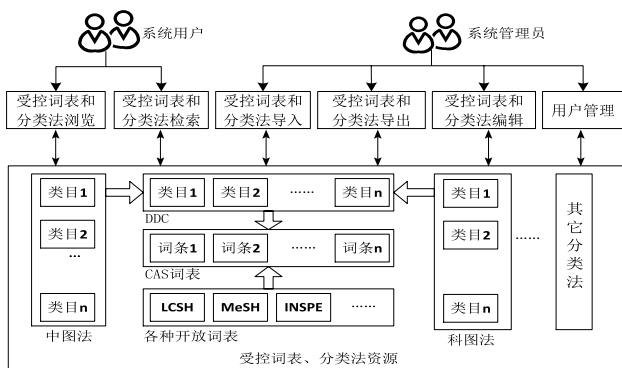


图3 开放知识组织体系结构

- ① 浏览主要采用等级浏览和字顺浏览。对于分类法采用等级浏览的方式,而对主题词表,主要按照字母排序浏览,通过词条间的关系显示与其关联的词条。
- ② 检索主要针对主题词检索。系统在数据库中进行匹配,查找与该主题词相关的其它关键词,检索结果为关键词列表以及相关信息的。
- ③ 导入主要是将OpenKOS受控词表和分类法按照RDF节点进行解析,获取相关信息,按照设计的关系数据库结构存入相关数据表。
- ④ 导出主要是将系统数据库存储的词表、分类法以XML格式,严格按照SKOS-CORE标准规范导出为文件形式。
- ⑤ 编辑主要是对OpenKOS受控词表及其它分类法与DDC分类法之间建立映射关联关系的操作。

## 4 系统整体设计

OpenKOS系统整体设计围绕开放知识组织引擎利用映射关联等方式将开放知识组织体系进行有机集成来完成。OpenKOS系统实现对开放知识组织体系的导入、浏览、检索和编辑等操作并提供标准格式接口和工具供第三方系统获取。OpenKOS系统总体架构如图4所示。

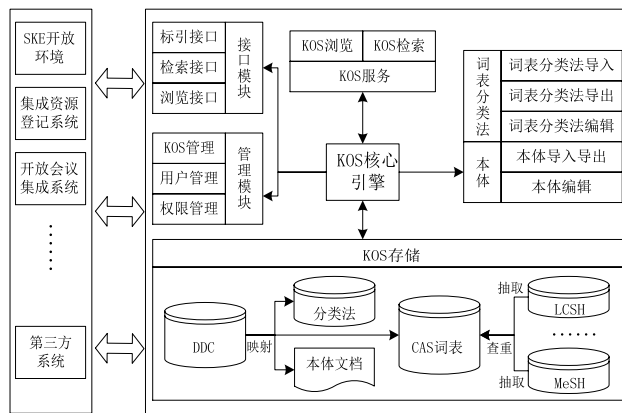


图4 OpenKOS系统总体架构

OpenKOS系统主要包括以下五个模块<sup>[10,11]</sup>:

### (1) 开放知识组织引擎模块

主要负责各个功能模块之间调用和任务分发,其中包括负责系统中各个功能模块与各种类型的开放知识组织体系(包括分类法、叙词表、本体)存储之间的通信,同时支持其它各个模块通过核心引擎模块对开放知识组织体系的调用操作。

### (2) 开放知识组织操作模块

支持注册用户针对各种类型的开放知识组织体系进行操作,主要分为两个子模块,受控词表、分类法操作模块和本体操作模块:

**受控词表、分类法操作模块:**支持注册用户对各种受控词表(主题词表、叙词表、术语表)和分类法进行操作,包括导入、导出和编辑。该部分在设计上采用关系数据库模式,用户通过网页浏览器对受控词表和分类法进行操作。

**本体操作模块:**支持注册用户对各种本体模型进行操作,包括本体导入、编辑和导出。不同于目前流行的本体编辑管理工具,该模块通过Web方式为注册用户提供服务,用户无需安装任何软件,即可利用浏览器对本体模型进行操作。

### (3) 开放知识组织服务模块

该模块是OpenKOS系统的对外展示模块,用户通过服务模块,可以方便快捷的了解OpenKOS系统内涵盖的所有开放知识组织体系及其概念结构。主要提供的服务包括针对开放知识组织体系的浏览和检索服务。

### (4) 开放知识组织管理模块

提供系统管理员对开放知识组织体系管理和用户管理的入口,用户包括人和机器,主要针对不同用户对

核心引擎的操作进行权限分配。同时为每种开放知识组织体系分配使用权限。系统中提供的各种开放知识组织体系操作都受限于用户角色与权限管理, 根据不同的权限进行不同级别的操作, 利用核心引擎模块调用管理模块获取用户角色, 判断用户以及其对各模块操作的权限。

#### (5) 开放知识组织接口模块

利用Web Service接口提供第三方系统调用。主要用于将开放知识组织服务模块的相关功能通过标准协议进行封装, 第三方系统通过这些协议调用相关的功能或者访问相关的数据以实现资源共享。接口支持中科院SKE环境、集成登记系统、开放会议系统以及第三方系统对开放知识组织体系的调用, OpenKOS系统通过接口功能为第三方系统提供服务, 主要包括标引接口、浏览接口、检索接口。

## 5 系统功能实现

OpenKOS系统按照中科院研究领域和信息资源状况, 重点集成了数学、物理、化学、天文、地球科学等各类开放知识组织体系共64个。系统主要功能包括开放知识组织体系的导入、浏览、检索和编辑等。

### 5.1 开放知识组织体系导入

开放知识组织体系导入主要包括受控词表、分类法和本体模型两个方面。

#### (1) 受控词表、分类法导入

受控词表和分类法的导入操作是将以文件形式存储的词表和分类法按照规范的数据库结构进行存储。用户通过操作界面, 填写基本描述信息并上传要导入的文件。导入操作流程如图5所示。

“词表、分类法基本描述信息”这一步要为每一个词表或者分类法指定一个DDC大类。在导入文件前从RDF读取基本信息, 如名称、交替名称、编制机构、类型、执行导入的用户名、导入日期和BaseURL等, 将这部分信息作为词表配置信息, 并生成词表ID。导入文件时需要判断导入的是分类法还是词表。如果导入的是词表文件, 在词表导入操作的同时, 需要往CAS词表中添加新加词条信息(只对叙词进行操作)。首先从词表文件中抽取其中的叙词, 将叙词与CAS词表中已存在的词条进行比较, 未发现重复词条则往CAS词表中写入叙词信

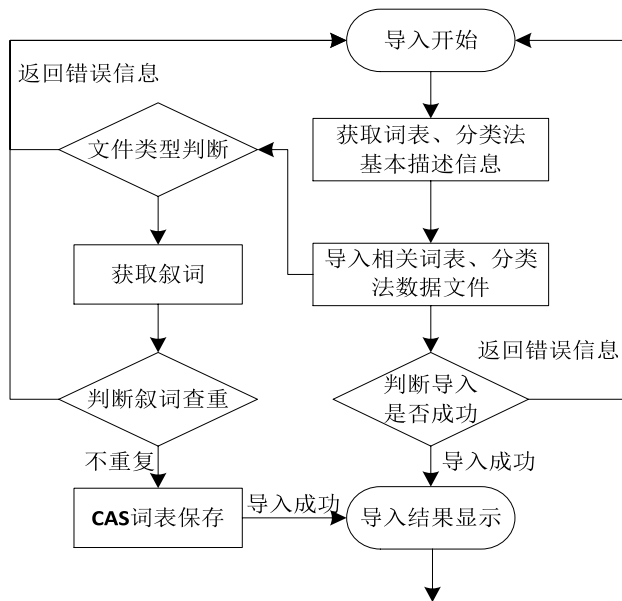


图5 受控词表和分类法导入流程

息, 发现重复词条则抛弃。如果导入的是分类法, 则需要将分类法中的顶层类或者前几层类存入到分类法索引表中。

#### (2) 本体模型导入

本体模型导入格式采用符合WebProtege规范的OWL或者RDF语言。用户通过Web操作界面填写开放知识组织体系元数据并上传本体文件。导入过程输入参数包括本体名称、交替名称、编制机构、类型、命名空间URI、命名空间前缀、所属DDC分类(提供方便的界面供用户选择)、执行输入用户名和输入日期。系统加载时将用户在页面输入的描述信息写入WebProtege提供的MetaProject(系统本体)中, 并且将文件上传到相关文件目录中。OpenKOS导入本体模型后显示所有的本体名称和描述信息, 点击本体名称能够浏览本体的类、属性、实例并进行编辑, 如图6所示。

### 5.2 开放知识组织体系浏览

OpenKOS对系统中所有的开放知识组织体系提供元数据信息以及概念内容两种浏览模式。对于元数据信息浏览模式细分为字顺浏览、分类浏览、类型浏览:

① 字顺浏览: 首先区分中英文, 英文按首字母顺序排列; 中文按拼音字母顺序排列。

② 分类浏览: 按照DDC结构浏览, 每种开放知识组织体系根据与DDC的映射关系进行浏览。

③ 类型浏览: 根据开放知识组织体系的不同类



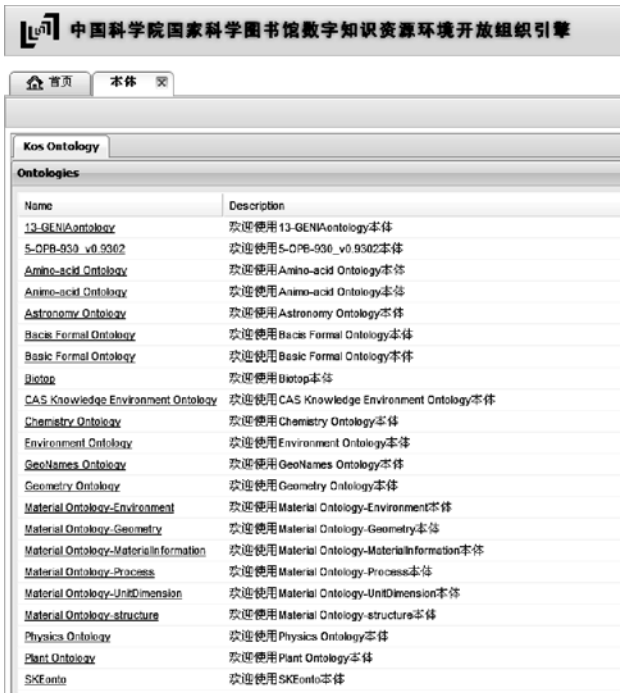


图6 本体模型导入浏览效果

型进行组织，分为三大类，分别是叙词表、分类法、本体。

OpenKOS系统提供的概念内容浏览模式实现对指定的开放知识组织体系内所有概念（包括叙词表、分类法、本体）的分级浏览，其浏览模式为树型浏览和可视化浏览：

① 树型浏览：按照开放知识组织体系中概念的等级组织分层浏览，以树型结构呈现。

② 可视化浏览：利用可视化组件对开放知识组织体系中的概念进行图形化呈现。

整个开放知识组织体系浏览模块的功能流程图如图7所示。

用户提交浏览请求，直接进入开放知识组织核心引擎中进行处理分析，并根据请求的目标进行Servlet分配调用。上图中元数据信息浏览模式通过Servlet进入

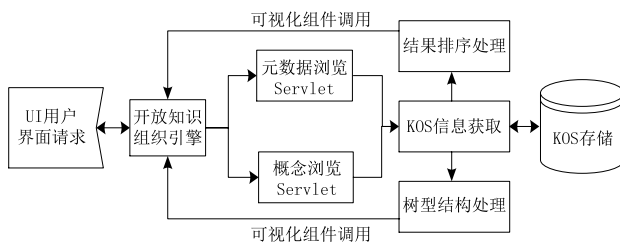


图7 开放知识组织体系浏览流程

具体的功能模块程序，与数据库进行交互并将结果进行排序处理，最终将结果返回开放知识组织核心引擎；概念内容浏览模式通过Servlet进入具体的功能模块程序，与数据库交互并进行结果的树型结构处理与可视化组件调用，最终将处理好的结果返回开放知识组织核心引擎。两种浏览模式最终由开放知识组织核心引擎利用GWT框架向客户端返回结果呈现。

OpenKOS系统的开放知识组织体系浏览模块利用GWT框架的Tab和Portlet功能进行各类开放知识组织体系的操作。整个OpenKOS系统界面按Web2.0风格设计，浏览效果如图8所示。

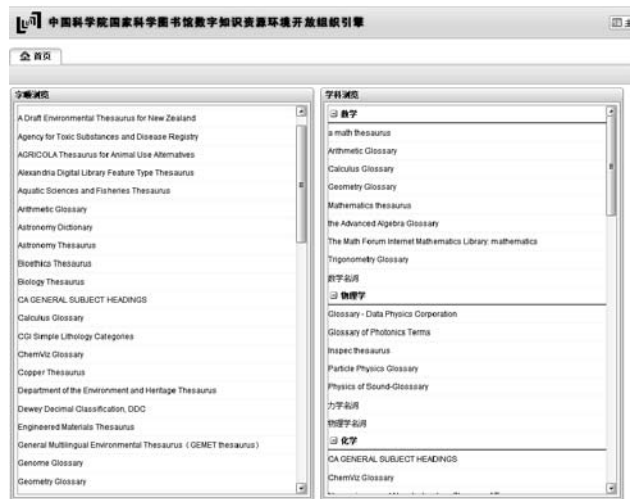


图8 开放知识组织体系浏览效果

### 5.3 开放知识组织体系检索

开放知识组织体系检索主要是针对OpenKOS系统中概念的检索，检索范围和检索类型可以由用户自己设定，检索结果显示按树型结构或可视化方式并且按开放知识组织体系进行分面展示。开放知识组织体系检索模块的功能流程图如图9所示。

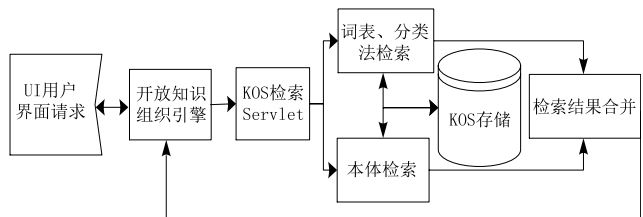


图9 开放知识组织体系检索流程

用户提交概念检索请求后，请求直接通过开放知识组织核心引擎进行处理分析，并根据请求的目标进行

Servlet分配调用。在上图中用户输入检索词,通过核心引擎调用相应的检索Servlet,进入后台具体功能模块程序。首先分析检索对象是否为词表、分类法或本体,根据检索对象分别进入词表、分类法检索和本体检索模块,将得到的检索结果进行合并处理,并送回开放知识组织核心引擎。最后核心引擎利用GWT规范向客户端返回结果呈现。检索效果如图10所示。



图10 开放知识组织体系检索效果

#### 5.4 开放知识组织体系编辑

开放知识组织体系编辑主要是进行CAS词表及其它分类法与DDC分类法之间的映射关联编辑操作。系统最终提供分类法以及CAS词表供用户选择,如果用户选择编辑CAS词表,按检索字顺列表方式列出CAS词表中的词条,并在词条右边列出与该词条关联的DDC类目,用户可添加或删除与该词条关联的DDC类目。如果用户选择编辑分类法,在用户选定要编辑的分类法后,用树状结构或支持检索的字顺列表列出分类法中的类目,在类目右边列出与该分类法关联的DDC类目,用户可添加或删除与该类目关联的DDC类目。DDC类目的选择可以采用树状列表、查询或者直接输入DDC分类号的方式进行。CAS词表编辑效果如图11所示。

### 6 系统接口应用

OpenKOS系统通过WebService接口对外提供服务,使第三方系统能够利用模块化集成的方式灵活地



图11 开放知识组织体系编辑效果

将开放知识组织体系的标引、分类、检索和查询等功能嵌入到自己的应用中。OpenKOS系统的WebService接口采用开源框架Apache CXF提供的WebService功能<sup>[12]</sup>,其开放接口主要提供以下功能:

① 第三方系统使用OpenKOS标引接口对资源进行标引。OpenKOS标引接口利用分类法或主题词表支持第三方系统的知识组织。标引过程除提供浏览方式外,还提供检索方式让用户查找类目或主题词。第三方系统在使用OpenKOS接口进行标引以后,需要设计更为合理的获取信息策略来显示更多资源信息。

② 第三方系统可通过OpenKOS浏览接口提供的分类体系来浏览自身的资源,这需要其中的资源已经按OpenKOS系统中的分类法进行了分类。OpenKOS系统根据用户提交请求返回知识组织体系的概念,在第三方系统按树形结构或者可视化方式显示。由于资源按照分类法进行分类,用户可以查看任意分类所属的资源。

③ 第三方系统通过OpenKOS检索接口实现对概念和主题的整体帮助其增加查全率,扩展或收缩检索范围。检索接口支持关键词的扩展检索。将用户提交的检索词映射到多个入口词并返回,用户可以用全部入口词进行检索,增加查全率。同时用户检索词可以对应到用户专用词表,将返回用户检索词对应的叙词上下位词和相关词,提示用户可用这些词缩小或扩展检索。

目前OpenKOS系统提供的开放接口部分设计思路在中科院文献情报中心的科技知识组织体系(STKOS)开放检索和推理接口项目中得到应用,同时在中科院机构知识库平台进行了示范应用,取得了不错的应用效果。该系统已在主题词标引、检索聚类等方面

应用了OpenKOS系统的相关接口,通过接口返回的相关词能够按照一定的规则对中科院机构知识库平台的主题词进行扩展,最终通过扩展标引后的记录可以在被检索时达到更好的检索效果。

## 7 结语

面对中国科学院新的知识服务环境,利用开放知识组织集成系统构建不同类型开放知识组织体系的检索、浏览、关联、导航、下载的开放式共享服务,是对科技知识组织体系(STKOS)开放检索和推理接口项目建设前期准备的一次有力探索。该系统通过深度开发和利用GWT实现了第三方开放知识组织体系的集成,后续通过对开放知识组织体系内容的不断丰富和完善,将使其成为支持国内各类信息机构和科研机构的信息基础设施,更好地为上层各类应用提供有力的知识服务支撑。

## 参考文献

- [1] 孙坦,刘峥.面向外文科技文献信息的知识组织体系建设思路[J].图书与情报,2013,33(1):2-7.
- [2] Renardus-Project Details[EB/OL]. [2014-11-05]. <http://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/renardus/>.
- [3] 李育嫦.传统知识组织系统的重构及其在网络环境下的应用[J].情报杂志,2011,30(7):114-117.
- [4] 胡滨,吴雯娜.国内外知识组织系统互操作模式及方法研究[J].情报科学,2012,30(9):1292-1297.
- [5] GWT Overview[EB/OL]. [2014-11-05]. <http://www.gwtproject.org/overview.html>.
- [6] SKOS Simple Knowledge Organization System - Home Page[EB/OL]. [2014-11-05]. <http://www.w3.org/2004/02/skos/>.
- [7] 张洁.《杜威十进分类法》的应用研究进展[J].图书情报工作,2013,57(18):110-114.
- [8] 贾君枝,郝倩倩.DDC与《中图法》组合类目映射探讨[J].中国图书馆学报,2012,55(4):63-67.
- [9] 曲建峰,李芳,张轶华,等.知识组织系统自动映射规则研究与实现—以《杜威十进分类法》和《中国图书馆分类法》为例[J].现代图书情报技术,2012,32(10):83-87.
- [10] 宋文,刘毅,张旺强,等.基于开放引擎的知识组织服务[J].图书情报工作,2012,53(18):99-103.
- [11] Zoghalmi K, Kerherve B, Gerbe O. Using a SKOS Engine to Create, Share and Transfer Terminology Data Sets[C]//Signal-Image Technology and Internet-Based Systems (SITIS), 2011 Seventh International Conference on. IEEE, 2011: 46-53.
- [12] 张旺强,祝忠明,卢利农,等.机构知识库集成OpenKOS主题标引与检索聚类服务的实现及应用[J].现代图书情报技术,2012,53(3):1-7.

## 作者简介

杨锐,男,1977年生,硕士,中科院武汉文献情报中心副研究员,研究方向:知识组织、信息挖掘、语义网技术,E-mail: yangr@mail.whlib.ac.cn.

## Design and Implementation of Open Knowledge Organization Integrated System Based-on GWT

YANG Rui<sup>1</sup>, LIU Yi<sup>1</sup>, SONG Wen<sup>2</sup>, ZHOU ZiJian<sup>1</sup>, ZHANG ShiNan<sup>2</sup>

(1. Wuhan Library, Chinese Academy of Sciences, Wuhan 430071, China; 2. Library, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: According to the application and research on open knowledge organization system under the network environment, the first to put forward the overall design scheme to construct integrated system based on open source GWT technology framework, and then, achieving the main function of open knowledge organization engine construction and import, browse, retrieval and edit of open knowledge organization system. The system generated form a unified CAS vocabulary by mapping, better realizes the effective integration of different open knowledge organization system, provides a strong support for the third party systems of knowledge service.

Keywords: Knowledge Organization System; Open Knowledge Organization Engine; GWT

(收稿日期: 2015-04-22; 编辑: 王立学)