

# 国外术语注册与术语服务综述\*

欧石燕

**摘 要** 本文对国外上最具代表性的三个术语注册与术语服务系统(FAO 注册系统与术语服务, OCLC 术语服务, Open Metadata Registry)和四个相关研究项目(STAR, HILT, ADL 地名表与叙词表协议, PoolParty)进行了详细的调研分析,在此基础上对我国术语注册与术语服务系统的开发与构建提出建议,主要包括:采用新兴的语义网和关联数据技术开发兼具术语注册与术语服务功能的完整系统,提供人与机器对注册词表元数据和词表内容的访问,采用 RESTful Web 服务构建术语服务,支持以关联数据方式发布词表内容等。图 5。表 4。参考文献 30。

**关键词** 术语注册 术语服务 语义网技术 关联数据

**分类号** G250.76

## A Review of Foreign Terminology Registries and Terminology Services

Ou Shiyan

**ABSTRACT** This paper gives a thorough review of three most representative terminology registries and terminology services (FAO VEST Registry & VocBench, OCLC Terminology Services, Open Metadata Registry) and four related projects (STAR, HILT, ADL Gazetteer & Thesaurus Protocol, PoolParty). And on this basis it puts forward the suggestions to the development of domestic terminology registries and terminology services, including developing an integrated system which has dual functionalities of terminology registry and terminology services using the burgeoning Semantic Web and Linked Data technologies, providing both human and machine access to the metadata and content of registered controlled vocabularies, building terminology services with RESTful web services, and enabling the publication of controlled vocabularies as Linked Data. 5 figs. 4 tabs. 30 refs.

**KEY WORDS** Terminology registries. Terminology services. Semantic Web technologies. Linked Data.

### 0 引言

术语表、分类表、叙词表、本体等各类词表(也称为知识组织系统)在信息资源描述、组织、管理、发现等方面的强大功能已经得到图书情报界和其他相关领域的广泛认可。为了促进对这些知识组织资源的有效利用,需要对它们进

行组织与管理。因此,早期一些机构(如图书馆)常常创建和维护许多词表列表以供内部使用,譬如欧盟发布的 Thesaurus Guide 列出了至少以一种欧洲语言表示的约 700 个词表<sup>[1]</sup>。1996 年以来,国外陆续出现了一些以电子格式发布的在线词表列表,如加拿大英属哥伦比亚大学图书情报学院创建的词表索引<sup>①</sup>,德国 Max Planck 数字图书馆的 Koch 创建和维护的词表列

\* 本文系国家社科基金项目“基于 SOA 架构的术语注册和服务系统构建与应用研究”(编号:11BT0023)的研究成果之一。

通讯作者:欧石燕,Email:oushiyan@nju.edu.cn

① <http://www.slais.ubc.ca/resources/indexing/database1.htm>

表<sup>①</sup>,但是它们中的大多数都没有得到持续的维护与扩展,有些甚至已经消失。

1998年,在美国计算机学会(ACM)数字图书馆会议的专题讨论会上,网络知识组织系统(NKOS<sup>②</sup>)的研究人员开始对词表注册进行讨论,探讨如何对叙词表等知识组织系统提供交互式的网络访问和使用<sup>[2]</sup>。前期工作主要集中在词表元数据方案的设计上,Hill领导的任务小组首先提出了一个以词表注册为目的的非常详尽的元数据方案,被称之为NKOS Registry,该方案包含了选择词表时所需的大部分信息。1998年在NKOS网站上发布了NKOS Registry的第二版<sup>[3]</sup>。2001年,来自OCLC的Vizine-Goetz创建了一个更为正式的文档(即数据元素的参考文档),仿照DC元数据元素,采用来自ISO/IEC 11179标准的一组属性来定义词表元数据方案中的每个元素,这就是NKOS Registry第三版的草案<sup>[4]</sup>。

在2003—2004年间,都柏林核心元数据启动计划(DCMI)在开发其元数据注册系统时开发并测试了“词表编码体系(Vocabulary Encoding Schemes)”注册系统,用以描述和命名能够在元数据记录中使用的词表,这可看作是最早的词表注册系统,但该系统因种种原因最终没有得以持续。目前DCMI的注册系统总称为都柏林核心元数据注册系统(Dublin Core Metadata Registry)<sup>③</sup>,提供关于DCMI元数据术语及其相关词表的最新权威信息。这个注册系统中包含的词表数量非常少,仅限于DC元数据中使用的词表。

2005年以来,术语注册与术语服务成为各种元数据会议和知识组织系统会议的热门主题。2005年都柏林核心元数据(DC)大会的NKOS特别论坛将词表注册作为其中的一个议题<sup>[5]</sup>,Heery做了关于元数据和词表注册的报告<sup>[6]</sup>。2006和2007年欧洲NKOS研讨会<sup>④</sup>也讨论了对注册系统的需求<sup>[7-8]</sup>。2008年欧洲NKOS研讨会分析并讨论了术语注册与术语服务的需求和应用以及用于词表注册的元数据<sup>[9]</sup>。2008年DC大会的NKOS特别论坛讨论了不同类型注册系统的异同点<sup>[10]</sup>。2009年美国NKOS研讨会和欧洲NKOS研讨会也将术语注册与术语服务作为会议的主要议题之一<sup>[11-12]</sup>。

在上述背景下,国外许多机构开展了术语注册与术语服务的研究与开发工作,既有不少实用系统问世,如FAO VEST Registry与VocBench<sup>⑤</sup>、OCLC术语服务<sup>⑥</sup>、Open Metadata Registry<sup>⑦</sup>,也有大量的研究项目进行,如HILT项目<sup>⑧</sup>、STAR项目<sup>⑨</sup>等,这些系统和研究项目对于构建我国面向中文知识组织系统的术语注册与术语服务有着非常重要的借鉴意义和参考价值。因此,本文对国外著名的术语注册与术语服务系统及相关研究项目进行调研分析,从而为我国术语注册与术语服务的开发研制提供建议和参考。

## 1 术语注册与术语服务概述

所谓注册(Registry),是指权威的、集中控制

① <http://www.mpdll.mpg.de/staff/tkoch/publ/koslist.html>

② Networked Knowledge Organization Systems/Services/Structures, 见 <http://nkos.slis.kent.edu>

③ <http://dcmi.kc.tsukuba.ac.jp/dcregistry>

④ 欧洲NKOS研讨会属于欧洲数字图书馆国际会议(ECDL)的一部分。

⑤ <http://aims.fao.org/vest-registry> 和 <http://aims.fao.org/tools/vocbench-2>

⑥ <http://tspilot.oclc.org/resources/index.html>

⑦ <http://metadataregistry.org>

⑧ 全称 High Level Thesaurus Project, 英国 RSLP 和 JISP 基金资助的一个研究项目。

⑨ 全称 Semantic Technologies for Archaeological Resources, 英国 AHRC 基金资助的一个研究项目。

的信息存储<sup>[13]</sup>。集中控制有两种情况:一是集中存储,即将信息集中存储在一个服务器中,从而进行集中管理和访问,这种也被称为存储库(Repository);二是分布式存储,虽然信息是分散存储在不同的服务器中,但进行集中管理并提供统一访问。本文所述的注册系统涵盖了上述两种情况。注册系统通过对注册资源进行集中发布、描述和管理并提供统一访问,能够有效地促进信息交换和知识共享。注册系统对注册资源提供的管理与访问分为两个层次:一个是元数据层面,对注册资源的特征和属性进行描述,供用户浏览和查询注册资源;另一个是内容层面,提供对注册资源内容的操作与访问。注册系统对注册资源可只提供元数据层面的管理与访问,也可同时提供上述两个层面的管理与访问。常见的信息注册系统有以下几种。

- 元数据注册:提供对元数据规范及规范中的元数据术语(元数据元素和元数据记录中需使用的规范词汇)的注册,譬如都柏林核心元数据注册系统。
- 术语注册:提供对各类受控词表以及词表中术语的注册,因此也被称为词表注册。术语注册常常伴随术语服务功能,从而支持机器对机器的词表访问,譬如 OCLC 术语服务。
- 本体注册:提供对本体及本体中定义词汇(类和属性)的注册,譬如美国国家自然科学基金项目海洋元数据互操作(Marine Metadata Interoperability)中研发的 MMI 本体注册和存储库(MMI Ontology Registry & Repository)<sup>①</sup>。
- 服务注册:提供对 Web 服务的注册,通常采用网络服务描述语言 WSDL 对 Web 服务进行描述,譬如英国联合信息系统委员会(JISC)的信息环境服务注册(Information Environment Service Registry)<sup>[14]</sup>。术语服务既可单独在服

务注册系统中进行注册,也可在术语注册中列出。

- 馆藏注册:提供对信息资源集合的注册,譬如 Cornell 大学图书馆的数字馆藏注册(Registry of Digital Collection)<sup>②</sup>。
- 数据标准注册:对各种数据标准(如数据词典、数据模型、方案和代码集等)进行注册,譬如美国儿童支持执行办公室面向儿童支持群体的数据标准注册(Data Standards Registry)<sup>③</sup>。

有些注册系统同时支持多种类型资源的注册,如 Open Metadata Registry 和 FAO VEST Registry 同时提供对元数据元素集和受控词表的注册,这类系统属于综合注册系统。本文主要对术语注册系统和提供术语注册功能的综合注册系统进行调研分析。

术语注册是指对各种受控词表提供权威的、集中控制的存储,从而促进词表的发现、重用、管理、标准化和互操作。一个术语注册系统能够列出、描述、识别、指明在信息系统和信息服务中可用的词表集合,并且提供图形化界面和术语服务以供用户访问和使用词表内容(指词表成员术语、概念及其相互关系)<sup>[15]</sup>。术语服务是指对词表元数据和词表内容进行浏览、查询、应用的各种 Web 服务的统称<sup>[16]</sup>。术语服务通过 Web 应用程序接口(API)支持机器对词表及其内容的访问和调用,是在网络环境下对词表进行应用的重要途径。术语注册和术语服务两者相辅相成,前者是后者的前提和保证,后者是前者的目的和应用。术语注册和术语服务系统是各种知识组织系统共建共享的重要平台,是网络知识组织系统由理论走向实际应用的关键环节,也是一个国家或领域内重要的信息基础设施。

① <http://mmisw.org/orr/#b>

② <http://rdc.library.cornell.edu/search/index.php?mode=browse&type=Collection>

③ <http://www.acf.hhs.gov/cse/dsr>

## 2 代表性术语注册与术语服务系统

目前,能够在线访问并提供服务的代表性术语注册与术语服务系统有:FAO 注册系统与术语服务、OCLC 术语服务和开放元数据注册 (Open Metadata Registry)。

### 2.1 FAO 注册系统与术语服务

联合国粮农组织 (FAO) 下属的农业信息管理标准 (AIMS) 团队创建并维护着大量农业领域的信息资源,其中包括各种农业类受控词表。为了促进对这些词表资源的利用,FAO 开发和构建了针对受控词表的注册系统和 Web 服务,提供对词表的浏览与检索以及对词表内容的访问与使用。FAO 的术语注册和术语服务分为两部分:一个是综合性农业信息管理注册系统 VEST Registry,提供对受控词表、元数据元素集和信息管理工具三种资源的注册;另一个是多语言词表编辑与管理工具 VocBench,主要针对 AGROVOC 词表提供编辑、管理与术语服务。

#### 2.1.1 VEST Registry

FAO 早期构建了一个受控词表注册系统,称作 FAO KOS Registry,存储了 90 多个与农业和农业管理相关的叙词表、规范文档、标题表、本体等。2011 年,FAO 将 KOS 注册系统升级为综合性的 VEST 注册系统,不仅包括受控词表,还包括元数据元素集和信息管理工具。VEST Registry 提供了一个基于 Web 的用户界面,使得用户能够通过该界面检索和分类浏览各种注册资源。该注册系统实质上是一个目录系统,只支持对注册资源元数据的检索与浏览,无法支持对注册资源内容的访问,尤其是对词表内容的访问。此外,VEST Registry 中注册的词表是事先存在的,该注册系统不支持用户注册并上传新词表,因此不能算是一个真正意义上的术语注册系统。

#### 2.1.2 VocBench

AGROVOC 多语言农业词表是农业领域的

核心词表,因此 FAO 重点开发了针对该词表的编辑、管理与访问工具。

1996 年,FAO 采用 OWL 语言定义 AGROVOC 词表的语义化描述模型,力图以更精炼和更准确的方式描述叙词表中的语义和词汇关系,从而在农业领域建立一个多语种的概念库 (即 Concept Server)<sup>[17]</sup>。2010 年,在 W3C 正式推出 SKOS 语言之后,FAO 又改用 SKOS 和 SKOS-XL 语言对 AGROVOC 词表重新进行语义化描述<sup>[18]</sup>。为了实现对传统格式 AGROVOC 词表的语义化转换,FAO 采用 JAVA 语言开发了一个基于 Web 的多语言本体概念构建和维护工具,称作 AGROVOC Concept Server Workbench,以帮助分布在全球各地的 AGROVOC 词表维护者合作构建农业领域的多语言本体和词汇系统<sup>[18]</sup>。Workbench 具有词表编辑管理、术语查找、词表验证、词表内容统计、词表输出和用户管理等功能<sup>[19]</sup>。目前,Workbench 已从单一的 AGROVOC 词表编辑工具发展成为一个管理基于 SKOS 的叙词表、规范列表、地名表等多种类型词表的多语言编辑和 workflow 工具 VocBench。该工具相当于一个术语注册系统,提供词表内容层面的管理与访问。

通过 VocBench 编辑和维护的 AGROVOC 词表已被 FAO 发布为关联数据,并且与 LCSH、DDC 等多个词表建立了关联,成为关联开放数据 (LOD) 云的一部分。AGROVOC 的关联数据版采用 SKOS 和 SKOS-XL 语言表示为 RDF 数据,存储在 Allegrograph 三元组存储器中,可通过 SPARQL 终端对词表内容进行查询,并通过关联数据前端 Pubby 将词表中概念的 URI 地址转换为 HTTP 协议可解引用的 HTML 文档。

针对终端用户,FAO 对 AGROVOC 词表提供了基于 Web 的浏览与查询、SPARQL 查询和术语服务三种访问方式,支持对 AGROVOC 词表内容的访问。通过 Web 界面,用户可查找词表中与输入字符串相匹配的术语,浏览词表中某个术语的详细信息及其所处的概念层级,还可以某种序列化格式下载整个词表。通过 AllegroGraph 的

SPARQL 查询终端,用户可自行构建 SPARQL 查询以检索词表中的特定内容,虽然这给用户提供了极大的灵活性,但编写 SPARQL 查询对于普通用户来说并不是件容易的事情。因此,FAO 还提供了 18 个基于 SOAP 协议的术语服务以支持机器对词表内容的访问。这些术语服务在后台通过 SPARQL 查询终端来实现,但相应的 SPARQL 查询已由开发者编写完成,并以 Web 服务的方式提供。

## 2.2 OCLC 术语服务

OCLC 术语服务是 OCLC Research 在 2004 至 2008 年间开展的一项研究项目,基于 10 个受控词表<sup>①</sup>构建了一个实验型术语服务系统。该系统提供的功能有:①浏览某个词表的描述(即元数据信息);②搜索某个指定词表中的术语(即概念或标题);③浏览某个术语与其他术语的关系(含等价关系、层次关系和网状关系等);④通过 URI 标识符检索某个术语。

在 OCLC 术语服务实验系统中,词表元数据(即词表的描述)和词表内容(即词表中的术语)均采用 MARC 21 规范数据格式(Marc 21 Format for Authority Data in XML)进行表示和存储,为了便于检索,还为词表元数据和词表中的术语建立了索引。整个系统采用 SRU(Search/Retrieve via URL)检索协议和 CQL(Common Query Language)查询语言实现对词表元数据和词表内容的检索。SRU 是针对 Web 的信息检索协议,采用 REST 风格的 Web 服务架构,通过 URI 参数方式提交检索请求,用于查询网络上的索引或数据库并以 XML 流返回检索结果<sup>[20]</sup>。CQL 是一种面向信息检索系统(如 Web 索引、图书馆编目)的形式化查询语言,不仅具有直观、人类可读可写的优点,还具有较强的表达能力,能用于构建复杂的查询。在 OCLC 术语服务中,检索结

果可以 HTML、MARC XML、MADS、SKOS 和 Zthes 多种不同的表示形式输出,用户可从中自行选择输出形式。

OCLC 术语服务可以有多种应用,譬如:①提供用于社会标签的术语源;②在搜索(如搜索引擎)中对查询进行精炼;③为检索词提供上下文关系或背景;④对元数据中的名称和主题词进行验证;⑤为跨数据库检索提供便利。OCLC 术语服务在元数据创建和查询词扩展中已经有了一些实践性应用。在 OCLC 术语服务项目的早期,开发者实现了一种通过 Microsoft Office 2003 中的 Microsoft Research Task Pane<sup>②</sup>调用 OCLC 术语服务的方式,通过该方式,可在 OCLC Connexion 编目工具中使用术语服务<sup>[21]</sup>,譬如,通过术语服务自动获取其他词表(如 LCSH)中与“love sorties”一词相映射的词并拷贝到编目工具的相应字段(见图 1)<sup>[22]</sup>。此外,美国印第安纳大学数字图书馆项目利用 OCLC 术语服务提供了一个查询扩展功能,譬如,通过术语服务自动获取“temples”一词的下位词并将其作为查询扩展词推荐给用户(见图 2)<sup>[23]</sup>。

## 2.3 Open Metadata Registry

Open Metadata Registry,原名 NSDL Registry,源于 2005 年美国自然科学基金资助的一个大型数字图书馆研究项目“美国自然科学数字图书馆(NSDL)”。在该项目中设计和开发了大量元数据元素集(metadata schemas)、受控词表(schemes)和应用规范(application profiles),为了在项目成员中共享这些方案(schemas)和概念体系(schemes),NSDL 开发了一个综合性注册系统,对上述元数据方案、受控词表和应用规范提供统一的存储与管理,以支持它们的发现、重用、创建、管理、标准化和

① OCLC 术语服务实验系统支持的 10 个受控词表是 LCSH、MESH、FAST、GSAFD、LCSHAC、LCTGM、GMPPC、META、LCGFT 和 BISAC。

② Microsoft Research Task Pane 为 IE 和 MS Office 2003 用户提供了一种访问松耦合的第三方服务的简便方式。



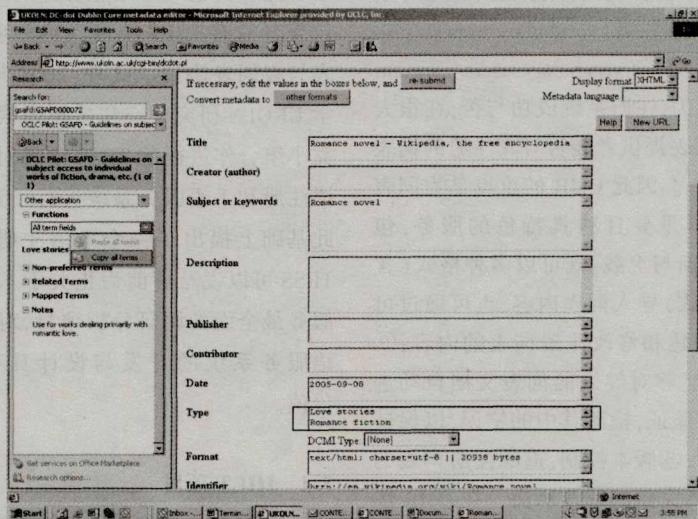


图1 在 OCLC Connexion 编目工具中使用术语服务<sup>[22]</sup>

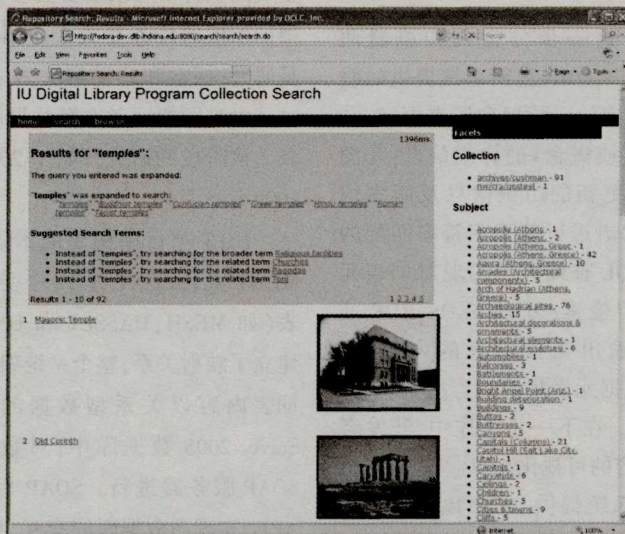


图2 印第安纳大学数字图书馆项目中利用 OCLC 术语服务实现的查询词扩展功能<sup>[23]</sup>

互操作。当时正值 SKOS 语言的草案出台,因此,NSDL Registry 也成为第一个事实上采用 SKOS 作为核心标准的注册系统。在 NSDL 项目结束后,项目的开发者将其作为一个完全免费

的注册系统向公众开放,并于 2010 年 7 月将其改名为 Open Metadata Registry (OMR)。目前,在 OMR 系统中共注册有 316 个受控词表和 83 个元数据元素集。

OMR 注册系统的开发者将系统的服务对象分为两大类:词表提供者(vocabulary owners/providers)和词表消费者(vocabulary users/consumers)<sup>[24]</sup>。他们认为,注册系统成功与否,在很大程度上取决于词表提供者能否对其注册的词表进行持续维护<sup>[24]</sup>。因此 OMR 的重点是为词表提供者提供非常重要且独具特色的服务,包括<sup>[24]</sup>:①词表注册与上载:既可以多种格式(含非 RDF/XML 格式)导入词表内容,也可通过可视化编辑界面创建和修改注册词表的内容;②词表自动验证:能够对导入的词表文档自动进行句法和完整性验证,检测其中的错误,以保证注册词表的质量;③版本控制:追踪注册词表的变化情况,保留其完整的修改历史。在 OMR 注册系统中,注册词表以 RDF 格式表示和存储,其中受控词表采用 SKOS 语言描述,元数据元素集采用 RDFS 语言描述,均采用 SPARQL 查询实现对词表内容的检索。

词表消费者可进一步分为人和机器两种<sup>[24]</sup>。对于人类用户,OMR 注册系统提供的服务有:①浏览注册词表的元数据和词表内容;②浏览词表成员(术语或概念)的详细信息;③浏览注册词表修改或更新的历史信息及版本信息;④对词表中的术语进行基于字符串匹配的检索;⑤提供 SPARQL 查询界面实现对词表元数据和词表内容的检索;⑥以 SKOS RDF 或 XML Schema 格式输出受控词表的内容,以 RDFS 格式输出元数据元素集的内容;⑦提供词表更新的自动通知。在下一步工作中,开发者还计划提供词表内容的可视化浏览<sup>[25]</sup>。对于机器用户,OMR 注册系统提供 REST 风格的 Web 服务,支持机器对机器的词表访问。总体来说,OMR 注册系统提供了非常强大的术语注册功能,但是其面向机器的术语服务功能还比较薄弱。

### 3 相关研究项目

除了在线的术语注册与术语服务系统,许

多研究项目也对术语服务进行了研究探索,比较著名的有 HILT 项目、STAR 项目、ADL 地名表与叙词表协议。此外,2008 年,英国巴斯大学 UKOLN 研究中心和格拉摩根大学超媒体研究小组合作进行的 TRSS 项目对当时主要的术语注册和术语服务系统进行了调研分析,并在此基础上提出了一个术语注册推荐模型<sup>[15]</sup>。TRSS 可以说是目前为止关于术语注册与术语服务最全面的调研分析项目,对术语注册与术语服务系统的开发与设计具有重要的参考价值。

#### 3.1 HILT 项目

HILT 项目是由英国斯特斯克莱德大学信息资源局数字图书馆研究中心和爱丁堡大学国家学术数据中心开展的一系列研究项目的总称,由英国联合信息系统委员会(JISC)资助。HILT 第四阶段的任务是关于术语服务,其主要任务是研究、探索和开发解决多学科信息环境中交叉搜索难题的试点方案并提供各种术语检索工具。该阶段的研究开始于 2007 年 4 月,完成于 2009 年 5 月。

HILT 项目中构建的术语服务基于 LCSH、MESH 和 HASSET 等 12 个受控词表,且部分词表(如 MESH、HASSET 和 UNESCO 等)与 DDC 建立了映射关系,整个系统架构如图 3 所示<sup>[26]</sup>。词表内容以关系型数据的形式存储在 SQL Server 2005 数据库中,对数据库的访问通过 SOAP 服务器进行。SOAP 服务器提供了七个 APIs,实现与数据库的交互,检索结果被包装为 SKOS 格式<sup>[26]</sup>。这七个 SOAP 服务采用一组简单易用的 PHP 类创建和调用。虽然在 HILT 项目中也测试了 REST 方式的 Web 服务,但是开发者认为 SOAP 方式在现阶段足够使用且更加稳定<sup>[26]</sup>。最后,SRU/SRW 服务器通过 SRU/SRW 协议访问 SOAP 服务器提供的 APIs,并以 SKOS 格式返回检索结果<sup>[26]</sup>。



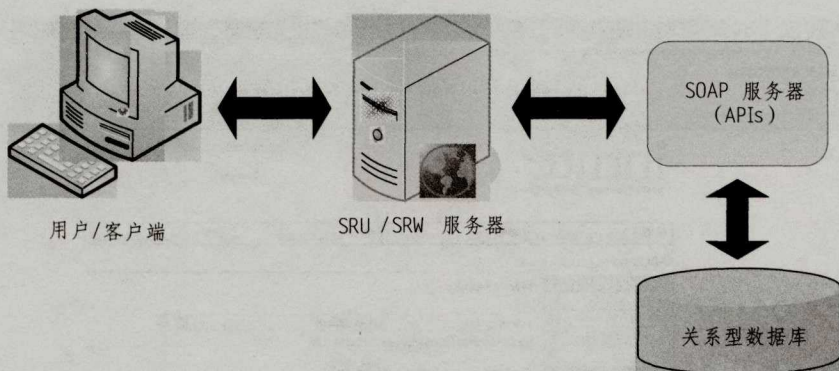


图3 HILT 术语服务的实现机制<sup>[26]</sup>

HILT 项目的合作伙伴,包括 Intute<sup>①</sup>, The Depot at EDINA<sup>②</sup>, Scottish Collection Network (SCONE)<sup>③</sup>等,将该项目中开发的术语服务嵌入到各自系统中,演示术语服务在实际应用中的功能与作用。譬如,在 Intute 中,通过术语服务从 MESH 和 HASSET 词表中获取与“rocket”相关的术语以及相匹配的 DDC 分类号,并返回到用户界面,以帮助用户扩展检索词(见图4)<sup>[26]</sup>。但非常遗憾,这些演示系统目前都已经无法在线访问。

### 3.2 STAR 项目

STAR 是由英国 AHRC 研究基金会资助的研究项目,由英国格拉摩根大学、英国文物委员会和丹麦图书馆学和情报学皇家学院合作研究。该项目历时三年(2007—2010),主要目的是探索语义术语工具对于扩展和提升数字考古资源可访问性的潜在作用。

在该项目中,提供服务的词表是关于英国遗产的受控词表(含叙词表和术语表),整个

系统的架构如图5所示<sup>[27]</sup>。词表内容采用 SKOS RDF 格式进行表示,通过 SemWeb<sup>④</sup> 开源类库以 RDF 三元组的形式存储在底层的 MySQL 数据库中,中间层采用 SemWeb 支持的 SPARQL 查询实现对词表数据的检索,上层通过 SOAP 风格的 Web 服务提供远程应用系统访问词表内容的 API 接口。STAR 项目中一共实现了七个术语服务,提供术语查找、浏览和概念扩展等功能。STAR 项目中开发的 Web 服务客户端有两种:一种是基于 Web 浏览器的微件(Web Widget),另一种是基于 Windows 的应用程序。

### 3.3 ADL 地名表与叙词表协议

ADL Gazetteer 和 Thesaurus Protocol 是 2001 至 2002 年间在美国亚利山大数字图书馆项目中开发的访问分布式地名表和叙词表资源的两个协议,是基于 XML 和 HTTP 协议的轻量级无状态协议。目前该项目已结束,这两个协议的最后修改日期是 2009 年 9 月。

① Intute 是英国的一个社会科学信息门户网站。

② EDINA 是为英国教育和科研提供共享信息服务和信息基础设施的国家数字中心,Depot 是其中的一个项目,提供对英国科研人员的学术论文和其他成果的免费在线收藏和存储。

③ SCONE 是提供对英国苏格兰地区图书馆、博物馆和档案馆馆藏进行描述和搜索服务的网站。

④ SemWeb 是采用 C# 语言为 Microsoft's .NET 1.1/2.0 编写的语义网类库。



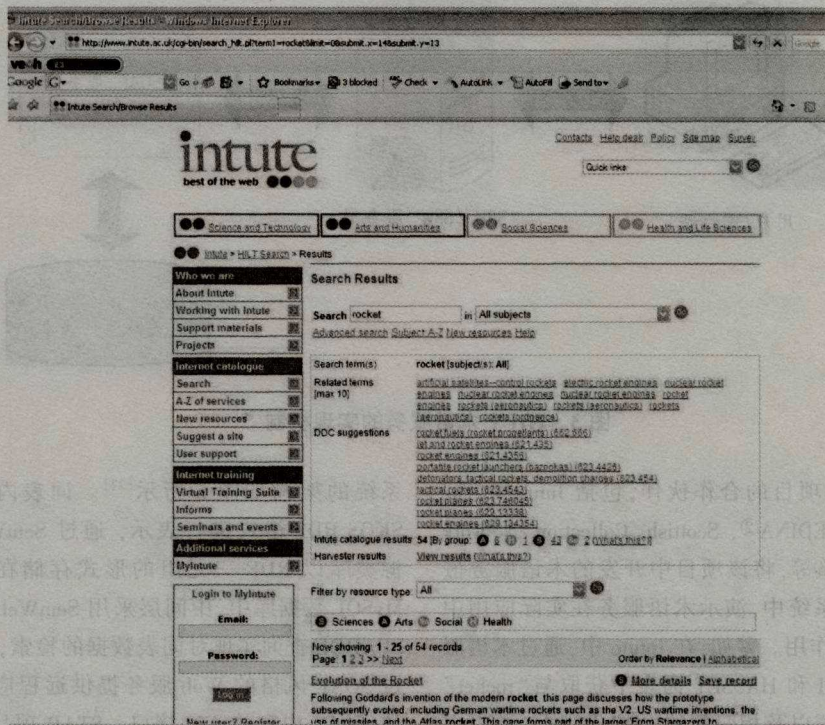


图4 在 Intute 通过术语服务获取“rocket”的相关术语和 DDC 分类号<sup>[26]</sup>

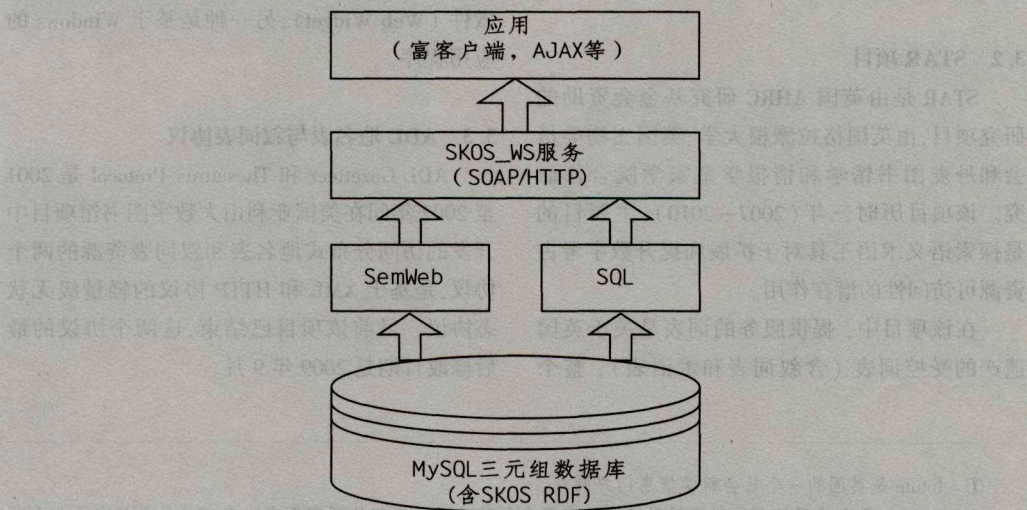


图5 STAR 项目中基于 SKOS 的术语服务架构<sup>[27]</sup>

ADL 地名表协议首先半形式化地定义了一个简单的地名表抽象模型,给出了地名表的通用结构和内容构成。该模型是定义地名表术语服务、报告格式和查询语言的基础。ADL 叙词表协议共提供了三个 SOAP 方式的术语服务,实现对地名表属性和地名表中地名的查询<sup>[28]</sup>。

ADL 叙词表协议遵循 ANSI/NISO Z39.19-1993 标准<sup>①</sup>构建了一个叙词表模型,自定义了八种 XML 格式,分别用于描述叙词表中的术语和属性以及术语的组织与排列。该协议用于机器对叙词表的访问和使用,提供了五个独立的、无状态的 SOAP 方式的术语服务,用于实现对叙词表内容的查询与浏览,但不支持叙词表的创建、维护、共享与映射等操作<sup>[29]</sup>。

### 3.4 PoolParty

PoolParty 是奥地利的 Semantic Web 公司开发的一个基于 Web 的叙词表管理工具,具有术语注册与术语服务的功能。PoolParty 采用 JAVA 语言基于语义网标准与技术进行构建,采用 SKOS 语言表示叙词表内容,采用语义网工具包 Sesame 的 Storage & Inference Layer (SAIL) API 对词表 SKOS RDF 数据进行存储与推理,采用关联数据技术发布词表,采用 SPARQL 查询访问词表内容<sup>[30]</sup>。

PoolParty 有基本、高级和企业三个版本;基本版提供了基于 Web 的可视化编辑环境,支持基于 SKOS 的叙词表的创建与编辑;高级版增加了关联数据前端,能够将叙词表中的概念在网络上发布为关联数据,并且与其他开放数据集(如 DBPedia、WordNet、LCSH、GeoNames)中的资源建立关联关系,从而成为关联开放数据(LOD)云的一部分,同时提供 SPARQL 终端供用户访问词表内容;企业版增加了文本挖掘功能,能够从大规模文档集中自动抽取短语或命名实体,并基于叙词表中的概念对文档进行标

注或利用文档中的新词扩充叙词表中的词汇,然后将结构化或非结构化信息集成到一个强大的智能搜索索引或 RDF 三元组存储器(如 Virtuoso 或 OWLIM)中。PoolParty 支持以 SKOS 格式(含 RDF/XML、N-Triples、Turtle、N3、TriX、Trig 序列化格式)或 Zthes 格式表示的外界词表的导入,采用 PoolParty 创建的 SKOS 词表也能够以各种 RDF 序列化格式导出。此外,PoolParty 还能够对基于 SKOS 的叙词表的质量自动进行检测,检测内容包括句法检测、SKOS 完整性检测、关联数据检测等几个方面。

PoolParty 还发布了一组 SKOS API,提供通过 RESTful Web 服务方式对叙词表元数据和内容的访问,使得机器和不熟悉 SPARQL 查询语言的用户也能够方便地访问叙词表。通过这些 Web 服务 API,能够实现基于叙词表的各种语义网应用,譬如:文档标注标签的自动推荐,输入标签时的自动拼写提示,相似文档推荐,语义搜索等<sup>[30]</sup>。

## 4 分析比较

在本节中,笔者从总体功能、数据格式、术语服务架构模式、术语服务种类与功能四个方面对所调研的三个在线系统(FAO, OCLC, OMR)和四个相关项目(HILT, STAR, ADL, PoolParty)进行分析比较。

### 4.1 总体功能

所调研的七个系统/项目的总体功能如表 1 所示。上述系统中,只有 OMR (Open Metadada Registry) 和 PoolParty 同时提供术语注册与术语服务功能。PoolParty 本质上是一个软件工具而非信息系统,但是基于该工具能够很方便地实现一个术语注册系统并提供术语服务。PoolParty 的注册功能主要在于叙词表的创

① ANSI/NISO Z39.19 - 1993: Guidelines for the Construction, Format, and Management of Monolingual Thesauri, 一个关于多语言词表的创建、格式和管理的标准。

建与编辑,而叙词表及其术语的浏览与检索功能则比较薄弱。OMR 则是一个真正的术语注册与术语服务系统。一方面,提供词表的注册、上载、格式转换、编辑、维护功能;另一方面,通过 Web 界面向人类用户,通过 Web 服务向机器用户,提供词表及其术语的浏览与检索功能。OMR 的术语注册功能可以说是非常强大,但是术语服务功能则不够成熟,是在后期才增加的,实现的功能比较简单。除此之外,其他系统都只针对有限、固定的词表提供术语服务,而无术语注册功能。虽然 FAO 有 VEST

注册,且 VocBench 支持对 AGROVOC 词表的在线编辑,但其术语服务只是针对 AGROVOC 一个词表。通过术语注册,一方面能够支持新词表的不断注册和上载,基于更多的词表提供术语服务,保证术语服务的可持续发展;另一方面,能够对词表进行更新和维护,保证提供服务的词表的权威性。因此,术语注册与术语服务实际上是系统密不可分两个部分,前者着重于词表的管理与维护,为后者提供支持与保证,后者着重于基于注册词表提供词汇服务,是前者的重要应用。

表 1 所调研系统/项目的总体功能

系统/项目名称	开发时间	当前状态	支持的 词表数量	术语注册	术语服务
FAO	2011 年至今	在线服务,持续更新中	单一受控词表	AGROVOC 词表 在线编辑	有
OCLC	2004—2008	在线服务	10 个受控词表	无	有
OMR	2005—2009 (NSDL) 2010 年至今	在线服务	300 多个受控 词表	词表导入、创建、 编辑、维护、管理、 错误检测	有
HILT	2007—2009	停止服务	12 个受控词表	无	有
STAR	2007—2010	停止服务	多个英国遗产叙 词表和术语表	无	有
ADL	2001—2002	无在线服务,停止 更新	多个地名词表和 叙词表	无	有
PoolParty	2009 至今	产品,持续更新中	一个或多个叙 词表	叙词表的导入、创 建与编辑、质量 控制	有

上述系统中,除了 FAO 术语服务是专为 AGROVOC 一个词表开发的,其他都是面向任何词表或者某一类型的众多词表开发的,这也是术语注册与术语服务的主流。支持的词表数量越多,术语服务的功能会越强大。在多词表环境中,还可在词表间建立映射关系,提供基于映射的术语服务。譬如,在 HILT 术语服务中,将词表中的术语与 DDC 分类号进行映射,提供通过类号查找对应术语或者通过术语查找对应类号的服务。

#### 4.2 数据格式

所调研的七个系统/项目的数据输入、存储和输出格式如表 2 所示。早期的术语注册与术语服务系统一般采用传统的 MARC XML 格式或者其他 XML 格式表示并存储词表内容,如 OCLC、ADL 和 HILT。随着词表语义化表示语言 SKOS 的出现与发展,尤其是 2009 年 8 月 SKOS 成为 W3C 的推荐标准以来,术语注册与术语服务基本上都采用 SKOS 作为词表的表示格式,采

用 RDF 三元组存储器进行存储,采用 SPARQL SKOS 格式的系统(如 OCLC 和 HILT),也提供了查询语言查询词表 SKOS RDF 数据,如 FAO、检索结果的 SKOS 格式输出。OMR、STAR 和 PoolParty。即使是底层不支持

表 2 所调研系统/项目中使用的数据格式

系统/项目名称	词表输入格式	数据存储格式	词表输出格式	查询语言
FAO	SKOS	RDF 三元组存储	SKOS RDF	SPARQL
OCLC	MARC XML	MARC XML (关系型数据库存储)	HTML, MARC XML, MADS,SKOS 和 Zthes	CQL
OMR	多种格式,包括非 RDF/XML 格式	SKOS	SKOS RDF/XML 格式, XML Schema 格式	SPARQL
HILT	XML	关系型数据库	SKOS	SQL
STAR	SKOS	基于关系型数据库的 RDF 三元组存储	未知	SPARQL,SQL
ADL	自定义 XML 格式	自定义 XML 格式	自定义 XML 格式	未知
PoolParty	SKOS,Zthes	基于文件系统的 RDF 三元组存储	SKOS RDF	SPARQL

4.3 术语服务架构模式

术语服务的实质是一种 Web 服务,由拥有词表和术语资源的服务器通过网络向应用程序提供各种词汇服务。所调研的七个系统/项目中术语服务的架构模式如表 3 所示,主要使用了两种 Web 服务架构模式:SOAP 模式和 REST 模式。SOAP 是一种比较成熟的 Web 服务架构模式,安全性较高,同时也比较复杂。与之相

比,REST 模式的 Web 服务则比较简洁,在效率和易用性上也要优于 SOAP 模式,但是安全性较差,因此比较适用于对效率要求高而对安全性要求不高的场合。从调研情况来看,早期的术语服务(如 HILT、STAR 和 ADL)采用 SOAP 模式较多,但后期的术语服务(如 OCLC 和 OMR)则多采用了 REST 模式,这与 Web 服务开发中越来越多地采用 REST 架构的情况相一致。

表 3 所调研系统/项目中术语服务的架构模式

系统/项目名称	Web 服务模式	协议	是否有客户端应用 DEMO
FAO	SOAP	SOAP,HTTP	无
OCLC	REST	SRU	查询扩展,元数据创建
OMR	REST	HTTP GET/POST	无
HILT	SOAP	SRW	拼写建议,获取与术语相对应的 DDC 分类号,获取相关术语
STAR	SOAP	SOAP,HTTP	无
ADL	SOAP	SOAP,HTTP	无
PoolParty	REST	HTTP GET/POST	文档标注标签推荐,语义搜索,相似文档推荐



除了 SOAP 和 REST 协议,采用非 RDF 关系型数据库存储格式的两个系统(即 OCLC 和 HILT)还采用了 SRU/SRW 协议。SRU/SRW 是一对针对 Web 的信息检索协议,用于查询网络上的索引或数据库并返回检索结果。它们利用 Web 服务的架构实现了 Z39.50 的一些基本服务,是下一代 Z39.50 协议 ZING 的核心组成部分。SRW 是 SOAP 模式的 Web 服务,使用 HTTP 与 SOAP 的无状态通信,采用 XML 作为信息传输编码,也可以单纯使用 URL 传递查询请求,用 WSDL 来定义 Z39.50 传输的格式信息,检索结果也以 XML 格式输出。而 SRU 则是 REST 模式的 Web 服务,只能通过 URL 参数方式提交检索请求,不支持完整的 SOAP 消息包(只支持 SOAP 消息包中的内容序列)。在 SRU 和 SRW 中采用 CQL 查询语言描述查询。OCLC 和 HILT

因为采用关系型数据库的数据存储格式,因此采用 SRU/SRW 协议支持异构数据间的互操作。对于采用统一的 SKOS RDF 数据格式的系统,无需采用 SRU/SRW 协议。

#### 4.4 术语服务种类与功能

在所调研的七个系统/项目中,FAO、STAR、HILT 和 PoolParty 都列出具体的术语服务,OCLC 虽然没有给出具体的术语服务,但是提供对术语服务功能的描述,只有 OMR 中没有关于术语服务的具体信息。笔者根据 OMR 可视化界面中提供的浏览与检索功能对 OMR 可能实现的术语服务功能进行推测,因为可视化界面中提供的功能可以非常容易地转换为 Web 服务方式提供给机器使用。对术语服务种类与功能的调研结果如表 4 所示。

表 4 所调研系统/项目中术语服务的种类与功能

术语服务名称	服务数量	具体功能
FAO 基于 AGRO-VOC 词表的术语服务	18	<p>getConceptByKeyword: 获取与输入字符串相匹配的概念及其详细描述<sup>(2)</sup></p> <p>getConceptByKeyword2: 获取与输入字符串相匹配的概念<sup>(2)</sup></p> <p>searchByModeLangScopeXML: 搜索与输入字符串相匹配的特定语种与类型的概念及其详细信息<sup>(2)</sup></p> <p>simpleSearchByMode2: 搜索与输入字符串相匹配的术语<sup>(2)</sup></p> <p>getConceptInfoByTermcode: 通过术语代码获取概念信息</p> <p>getConceptInfoByURI: 通过 URI 标识符获取概念的详细描述<sup>(3)</sup></p> <p>getDefinitions: 通过术语代码获取该术语指定语种的定义信息</p> <p>getAllLabelsByTermcode2: 通过术语代码获取该术语所有语种的标签</p> <p>getTermByLanguage: 通过术语代码获取该术语指定语种的标签</p> <p>getURIByTermAndLangXML: 获取与输入字符串相匹配的概念的 URI 标识符</p> <p>getFullAuthority: 获取规范文档的 URI 标识符及其详细描述</p> <p>getConceptByURI: 通过 URI 标识符获取概念及其详细描述<sup>(3)</sup></p> <p>getConceptByRelationshipValue: 获取具有某个特定属性值的概念</p> <p>getlatestUpdates: 获取词表在指定期限内的修改更新记录</p> <p>getTermcodeByTermAndLangXML: 获取输入术语的术语代码</p> <p>getTermExpansion: 获取输入术语的同义词<sup>(6)</sup></p> <p>getReleaseDate: 获取数据集的发布日期</p> <p>getWebServicesVersion: 获取术语服务的版本</p>

续表

术语服务名称	服务数量	具体功能
OCLC 术语服务	未知	浏览词表的详细描述(即元数据信息) <sup>(5)</sup> 搜索词表中的术语,即概念或标题 <sup>(2)</sup> 浏览术语与其他术语的关系,包括等价关系、层次化关系和网络状关系 <sup>(1)</sup> 通过 URI 标识符检索术语 <sup>(3)</sup>
OMR 术语服务	未知	跨词表检索与输入字符串相匹配的概念 <sup>(2)</sup> 浏览词表的详细描述(即元数据信息) <sup>(5)</sup> 浏览词表中的所有概念 <sup>(4)</sup> 浏览词表中概念的详细信息(如标签、上下位关系、相关关系) <sup>(1)</sup>
HILT 术语服务	7	get_filtered_set:在一个或多个词表中检索与输入字符串相匹配的术语以及它的其它标签与上下位概念和相关概念 <sup>(1,2)</sup> get_ddc_records:在 DDC 或者与 DDC 相映射的词表中检索与指定术语相对应的 DDC 分类号 get_non_ddc_records:在与 DDC 相映射的词表中检索与指定 DDC 类号相对应的术语 get_all_records:合并 get_ddc_records 和 get_non_ddc_records 的结果 get_collections:在 JISC 资源集中检索具有指定 DDC 类号的资源 get_spelling_suggestions:从 HILT 术语数据库的索引中获取指定术语的拼写建议 get_wordnet_suggestions:获取 WordNet 中对指定术语的描述
STAR 术语服务	7	GetTopmostConcepts:获取指定词表中的顶级概念 <sup>(4)</sup> GetConceptScheme:获取指定词表中的概念体系 <sup>(7)</sup> GetConceptSchemes:获取术语服务支持的概念体系 GetConcept:获取指定 URI 标识符的概念 <sup>(3)</sup> GetAllConceptRelatives:获取与输入概念直接相关的概念 <sup>(1)</sup> GetKeywordMatch:获取与输入字符串相匹配的概念 <sup>(2)</sup> ExpandConcept:在指定的语义相似度范围内对指定概念进行扩展 <sup>(6)</sup>
ADL 术语服务	5	get-properties/get-capabilities:获取叙词表/地名表的详细描述(元数据信息) <sup>(5)</sup> download:获取叙词表/地名表中的所有术语 <sup>(4)</sup> query:在叙词表/地名表中查询与输入字符串相匹配的术语 <sup>(2)</sup> get-broader:在叙词表中获取输入术语的上位词的层次结构 <sup>(1)</sup> get-narrower:在叙词表中获取输入术语的下位词的层次结构 <sup>(1)</sup>
PoolParty 术语服务	8	getProjects:获取所有叙词表及其详细描述(即元数据) <sup>(5)</sup> getConceptSchemes:获取叙词表中的概念体系 <sup>(7)</sup> getTopConcepts:获取叙词表中的顶级概念 <sup>(4)</sup> getConcept:获取指定 URI 标识符的一个概念的详细描述 <sup>(3)</sup> getConcepts:获取指定 URI 标识符的多个概念的详细描述 <sup>(3)</sup> getBroaderConcept:获取输入概念的上位概念 <sup>(1)</sup> getNarrowerConcept:获取输入概念的下位概念 <sup>(1)</sup> getRelatedConcept:获取输入概念的相关概念 <sup>(1)</sup>

注:具体功能中各项服务后的上标指的是所属服务种类

下面对七种认可度较高的术语服务进行详细描述,认可度按照提供该项服务的系统/项目的数量来衡量(括号中为提供该服务的系统/项目名称)。

(1)获取与输入术语具有某种语义关系的概念(FAO+OCLC+OMR+HILT+STAR+ADL+PoolParty):系统输出术语的上位概念、下位概念或相关概念。

(2)查询与输入关键词相匹配的术语或概念(FAO+OCLC+OMR+HILT+STAR+ADL):用户输入一个关键词(即文本字符串),系统返回与之相匹配的术语或概念,或者同时包含其详细描述。此处需要说明两点:(a)匹配方式可以有多种,如精确匹配、包含、前方一致、后方一致;(b)如果输出结果为概念,匹配的可以是概念的任一标签,如首选标签、可选标签、隐藏标签。

(3)通过 URI 标识符获取某个概念及其详细描述(FAO+OCLC+STAR+PoolParty):用户输入一个 URI 标识符,系统返回以该标识符命名的概念或者同时包含其详细描述。

(4)获取指定词表中的概念(OMR+STAR+ADL+PoolParty):获取某个指定词表中的顶级概念或者所有概念。该服务适用于词表中概念较少的情况。

(5)获取指定词表的详细描述(OCLC+OMR+ADL+PoolParty):获取某个指定词表的详情,即该词表的元数据信息。

(6)对术语进行扩展(FAO+STAR):系统输出与输入术语语义相似的术语或概念,如同义词和语义相似的词。

(7)获取指定词表中的概念体系(STAR+PoolParty):获取某个指定词表中所包含的概念体系。

## 5 结论与建议

术语注册与术语服务在信息检索、信息浏览、信息发现、自动翻译、语义推理、编目和元数据创建、知识组织等许多领域都有着非常重要

的应用。开展术语注册和术语服务方面的研究,构建适用于我国知识组织系统的术语注册与术语服务系统,是十分必要和迫切的。通过对代表性术语注册与术语服务系统及相关研究项目进行调研分析,笔者针对我国术语注册与术语服务系统的开发与构建提出如下建议:

- 构建兼具术语注册与术语服务功能的完整系统,从而能够对术语服务提供更多、更好的词表支持;
- 采用新兴的语义网和关联数据技术开发术语注册与术语服务系统,采用 SKOS RDF 作为词表的表示格式,采用 RDF 三元组存储器对词表内容数据进行存储,采用 SPARQL 查询访问词表内容;
- 术语注册功能主要在于词表元数据注册,词表内容上载与导入,词表自动验证与审核,词表更新,词表版本控制,词表权限控制,词表元数据与词表内容浏览与检索;
- 既提供对注册词表元数据的访问,也提供对词表内容(即词表中的概念、术语及其相互间关系)的访问;
- 支持以多种序列化格式(如 RDF/XML、Turtle、N-Triples 等)上载和输出词表内容;
- 提供词表自动验证审核机制,从而保证注册词表的权威性与正确性;
- 既提供基于 Web 的可视化用户界面,也提供针对机器(计算机程序)的术语服务 API;
- 既提供 SPARQL 查询接口,供专业用户灵活定制查询内容,又提供特定功能的术语服务,方便机器的访问和非专业用户的使用;
- 采用 RESTful Web 服务构建术语服务;
- 尽量针对多个词表构建术语服务,词表之间最好建立映射关系,并提供基于映射关系的术语服务;
- 提供所支持的术语服务的元数据,促进术语服务的发现与使用;
- 支持以关联数据方式发布词表内容,使其成为关联数据(LOD)云的一部分;
- 至少提供七种基础术语服务,包括获取语义

相关的概念,查询术语信息,获取词表中的概念,术语扩展,获取词表的元数据信息,通过URI标识符获取概念,获取词表中的概念体系。

通过建立术语注册和术语服务机制,可以

加强对增长迅速、类型多样、内容复杂、来源不同的各类词表的维护和管理,并可直接通过网络为各类应用提供方便、快捷、强大的术语支持,让各类词表在网络环境下发挥更大的效益和价值。

## 参考文献

- [1] Commission of the European Communities, EUROBrokerS. Thesaurus guide: analytical directory of selected vocabularies for information retrieval, 1992 (2<sup>nd</sup> version) [R]. Luxembourg: Office for Official Publications of the European Communities, 1993.
- [2] Workshop on application of terminology and classification tools for digital collection development and network-based search [EB/OL]. [2014-06-10]. [http://nkos.slis.kent.edu/workshop\\_slides](http://nkos.slis.kent.edu/workshop_slides).
- [3] Networked knowledge organization systems (NKOS) registry help document [EB/OL]. [2014-06-10]. <http://nkos.slis.kent.edu/registry2.htm>.
- [4] Networked knowledge organization systems (NKOS) registry reference document for data elements [EB/OL]. [2014-06-10]. <http://nkos.slis.kent.edu/registry3.htm>.
- [5] Hodge G, Koch T, Tudhope D. NKOS special session at DC-2005 [EB/OL]. [2014-06-10]. <http://www.ukoln.ac.uk/nkos/nkos2005/NKOSatDCMI05.html>.
- [6] Heery R. (Metadata and) vocabulary registries [EB/OL]. [2014-06-10]. [http://www.ukoln.ac.uk/ukoln/staff/r.heery/presentations/ppt-2000-html/heery-nkos-2005\\_files/v3\\_document.html](http://www.ukoln.ac.uk/ukoln/staff/r.heery/presentations/ppt-2000-html/heery-nkos-2005_files/v3_document.html).
- [7] The 5<sup>th</sup> European networked knowledge organization systems (NKOS) workshop [EB/OL]. [2014-06-10]. <http://www.ukoln.ac.uk/nkos/nkos2006>.
- [8] The 6<sup>th</sup> European networked knowledge organization systems (NKOS) workshop [EB/OL]. [2014-06-10]. <https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2007>.
- [9] The 7<sup>th</sup> European networked knowledge organization systems (NKOS) workshop [EB/OL]. [2014-06-13]. <https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2008>.
- [10] Preliminary agenda of DC 2008, NKOS special session [EB/OL]. [2014-06-13]. [http://colab.mpg.de/mediawiki/NKOSevents\\_2008\\_preparation](http://colab.mpg.de/mediawiki/NKOSevents_2008_preparation).
- [11] Knowledge organization systems: managing to the future (a joint CENDI/NKOS workshop) [EB/OL]. [2014-06-13]. <http://nkos.slis.kent.edu/2009workshop/NKOS-CENDI2009.htm>.
- [12] The 8<sup>th</sup> European networked knowledge organization systems (NKOS) workshop [EB/OL]. [2014-06-13]. <https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2009>.
- [13] Hass H, Brown A. Web services glossary: W3C working group note 11 February 2004 [EB/OL]. [2014-08-10]. <http://www.w3.org/TR/ws-gloss>.
- [14] Johnston P. Collection description focus briefing paper 3: the JISC information environment service registry [R/OL]. [2014-06-13]. <http://www.ukoln.ac.uk/cd-focus/briefings/bp3/bp3.pdf>.
- [15] Golub K, Tudhope D. Terminology registry scoping study (TRSS): final report [R/OL]. UK: Joint Information Systems Committee (JISC), 2009 [2011-01-28]. <http://www.jisc.ac.uk/media/documents/programmes/sharedservices/trss-report-final.pdf>.
- [16] Tudhope D, Koch T, Heery R. Terminology services and technology: JISC state of the art review [R/OL]. UK: Joint Information Systems Committee (JISC), 2006 [2011-01-28]. [http://www.jisc.ac.uk/Terminology\\_Serv](http://www.jisc.ac.uk/Terminology_Serv)



- ices\_and\_Technology\_Review\_Sep\_06.
- [17] Lauser B, Sini M. From AGROVOC to the agricultural ontology service / concept server: an OWL model for creating ontologies in the agricultural domain [C/OL]. Proceedings of the 2006 International Conference on Dublin Core and Metadata Applications. Singapore: Dublin Core Metadata Initiative, 2006: 76-88[2014-06-10]. <http://dcpapers.dublincore.org/pubs/article/download/841/837>.
- [18] Rajbhandari S, Keizer J. The AGROVOC concept scheme - a walkthrough [J/OL]. Journal of Integrative Agriculture, 2012, 11(5): 694-699[2014-07-28]. <http://www.fao.org/docrep/015/an910e/an910e00.pdf>.
- [19] Yongyuth P, Thamvijit D, Suksangri T, et al. The AGROVOC concept server workbench: a collaborative tool for managing multilingual knowledge [C/OL]. Proceedings of the 12<sup>th</sup> World Congress of the International Association of Agricultural Information Specialists, 2008[2014-06-10]. <http://eprints.relis.org/15616>.
- [20] Morgan E L. An introduction to the search/retrieve URL service (SRU) [J/OL]. Ariadne, 2004 (40) [2014-05-05]. <http://www.ariadne.ac.uk/issue40/morgan>.
- [21] Vizine-Goetz D, Houghton A, Childress E. Web services for controlled vocabularies [J/OL]. Bulletin of the American Society for Information Science and Technology, 2006, 32(5): 9-12 [2014-05-05]. [http://www.asis.org/Bulletin/Jun-06/vizine-goetz\\_houghton\\_childress.html](http://www.asis.org/Bulletin/Jun-06/vizine-goetz_houghton_childress.html).
- [22] OCLC terminologies pilot Connexion browser getting started [EB/OL]. [2014-05-05]. [https://www.oclc.org/content/dam/oclc/productworks/terminologiespilot/metadata\\_editors.doc](https://www.oclc.org/content/dam/oclc/productworks/terminologiespilot/metadata_editors.doc).
- [23] Vizine-Goetz D. Terminology services [EB/OL]. [2014-05-05]. <http://www.oclc.org/content/dam/research/presentations/vizine-goetz/cendi-nkos-isko.ppt>.
- [24] Hillmann D, Sutton S A, Phipps J, et al. A metadata registry from vocabularies up: the NSDL registry project [C/OL]. Proceedings of the 2006 International Conference on Dublin Core and Metadata Applications. Singapore: Dublin Core Metadata Initiative, 2006: 65-75 [2014-06-20]. <http://arxiv.org/ftp/cs/papers/0605/0605111.pdf>.
- [25] Phipps J, Hillmann D. The open metadata registry: an update [J/OL]. Bulletin of the American Society for Information Science and Technology, 2011, 37(4): 35-37[2014-06-20] [http://www.asis.org/Bulletin/Apr-11/AprMay11\\_Phipps\\_Hillmann.pdf](http://www.asis.org/Bulletin/Apr-11/AprMay11_Phipps_Hillmann.pdf).
- [26] Nicholson D, McCulloch E, Joseph A. JISC final report: HILT IV and embedding extension [R/OL]. [2012-04-30]. <http://strathprints.strath.ac.uk/14046/1/strathprints014046.pdf>.
- [27] Binding C, Tudhope D. Terminology web services [J/OL]. Knowledge Organization, 2010, 37(4): 287-298 [2014-06-20]. [http://hypermedia.research.southwales.ac.uk/media/files/documents/2010-12-16/cbinding\\_dtudhope\\_KO\\_rev7.pdf](http://hypermedia.research.southwales.ac.uk/media/files/documents/2010-12-16/cbinding_dtudhope_KO_rev7.pdf).
- [28] Janee G, Hill L. The ADL gazetteer protocol [EB/OL]. [2014-06-20]. <http://legacy.alexandria.ucsb.edu/~gjanee/gazetteer/specification.html>.
- [29] Janee G, Ikeda S, Hill L L. The ADL thesaurus protocol [EB/OL]. [2014-06-20]. <http://legacy.alexandria.ucsb.edu/thesaurus/protocol/specification.html>.
- [30] PoolParty thesaurus manager user guide [EB/OL]. [2014-06-10]. <https://grips.semantic-web.at/download/attachments/21890292/PPT-UserGuide.pdf?version=7&modificationDate=1327409781000&api=v2>.

欧石燕 南京大学信息管理学院教授, 博士生导师。

通讯地址: 江苏省南京市栖霞区仙林大道 163 号南京大学仙林校区信息管理学院。

邮编: 210046。

(收稿日期: 2014-06-27)