Digital Library Forum 数字图书馆论坛
http: // www.dlf.net.cn

集成化本体管理平台的设计与实现*

□ 许德山 张运良 / 中国科学技术信息研究所 北京 100038

摘要：文章以本体管理与服务平台的建设为主线，首先分析了集成化本体管理和服务系统的应具备的各项功能，并设计了系统的整体架构。其次从本体注册、三元组检索、可视化展示和Web services等方面详细介绍平台的实现思路。文章最后对目前项目进行总结并提出了未来研究的重点方向。

关键词：本体管理，本体发布，知识服务
DOI：10.3772/j.issn.1673—2286.2013.11.004

引言

本体作为一种概念体系的形式化描述为计算机处理领域知识带来了便利，随着各种应用需求的增加，领域本体的建设也越来越多。国外研究机构开发了多种管理工具来满足不同用户的使用需求[1-3]，国内学术界也在积极地开展本体存储工具的研究[4-6]。随着研究的深入和应用的扩展，中文本体的规模和数量逐渐增多。为了整合各种中文本体资源，向用户提供有效的信息获取服务，构建集本体管理、发布、检索和服务于一体的平台系统日益迫切。本文旨在对中文本体的管理和服务进行探索，并利用开源存储工具实现一个整合服务原型系统，为后期深入研究提供参考经验。

1 本体管理平台的功能结构

早期编制的词表、分类表等信息组织系统主要供标引人员使用，其内部结构未作明确的定义，要使计算机能够自动处理词表信息，必须将其改造为更加规范的本体描述。本体管理平台旨在为已创建的各种本体资源提供一套规范化的管理方式，并通过统一的服务接口供外部用户使用。因此管理平台除了具备本体资源的存储和检索功能，还应该提供丰富的服务模式来满足不同类型用户的使用需求。整个管理平台由用户界面、服务处理和存储管理3层结构组成。其详细情况如图1所示。

（1）用户界面

为了有效地展示本体的内部结构和结点描述，管理平台提供了可视化界面，帮助用户有效地获取相关信息。可视化界面对中间层的可视化模型、SPARQL构造器、Web services接口等功能进行了包装，向用户提供可视化检索、ontology浏览以及services示范用例等便捷功能。检索界面还设置了信息提示功能，用户输入概念词汇后，系统将词汇发送到服务处理器，检索代理模块立即在知识库中查找与该词汇相关的信息，并以列表形式返回检索框，用户可以进一步选择概念的某一个侧面进行提交，以便获得更精确的检索结果。

（2）服务处理层

服务处理层主要功能是检索条件与三元组概念图的匹配。存储在后台知识库中的三元组必须读入内存形成完整的概念网络，才能实现需求与知识源的匹配。服务层接收用户界面发送的概念、限定条件和URL等信息后，将其映射为满足需求的SPARQL检索式，然后将检索式与概念模型进行图形匹配，返回命中的概念结点信息。

（3）存储和管理层

存储管理层的功能是将OWL描述的本体文件转换成统一的三元组形式，并创建知识库进行存储。同时还负责本体内容的修改、实例

* 本文系“十二五”国家科技支撑计划项目“科技知识组织体系共享服务平台建设”（编号：2011BAH10B03-2）、中国科学技术信息研究所重点工作项目“汉语科技词系统建设与应用工程”（编号：ZD2012-3-2）的研究成果之一。

的添加、多本体间的概念链接等操作。本体知识库主要由Schema、领域规则和实例组成。Schema是整个系统的概念基础，它提供了领域概念知识的类别和层次结构，并建立了各概念知识的多种联系，是定义领域规则和OWL描述实例的基础，同时也是生成查询表达式的重要依据。领域规则是在领域本体的基础上定义的，以SWRL语法描述了领域概念知识间的隐含关系，是进行推理查询的基础。作为领域概念知识的具体形式，实例是三元组检索主要的信息源，它不仅包含了直接以个体存储的领域知识，还包括经过描述逻辑推理后的隐含信息。

2 本体注册及存储

本体文件的内容描述通常与一个特定的领域有关，其内部的概念结点和关系构建了领域知识框架。由于单个的本体文件所提供的信息有限，发布平台通过识别词汇和标签形式，将多个本体中相同或相似的概念建立了语义链接，本体注册完成后，其内部的各种信息以三元组的形式存储到数据库中，形成多领域的知识库。三元组的后台存储使用Sesame工具包来完成[7]，一个新的本体文件添加后，Sesame知识库将以本体URL为命名空间生成存储文件，同时将本体内部的概念类、类间关系和属性等结构词汇在浏览页面中进行发布。用户可以查看本体页面了解其内部结构，还可以点击相应词汇通过可视化的方式了解词汇的关系和属性等描述信息。本体注册页面的效果如图2所示。

管理平台设置了3种不同的用户-管理员、注册用户和普通用户。管理员负责本体文件的添加、删除以及其他用户的授权操作。注册用户可以使用平台提供的Web services、可视化检索、知识库浏览、本体下载等功能。普通用户无需注册，但仅能使用可视化检索和知识库浏览功能。

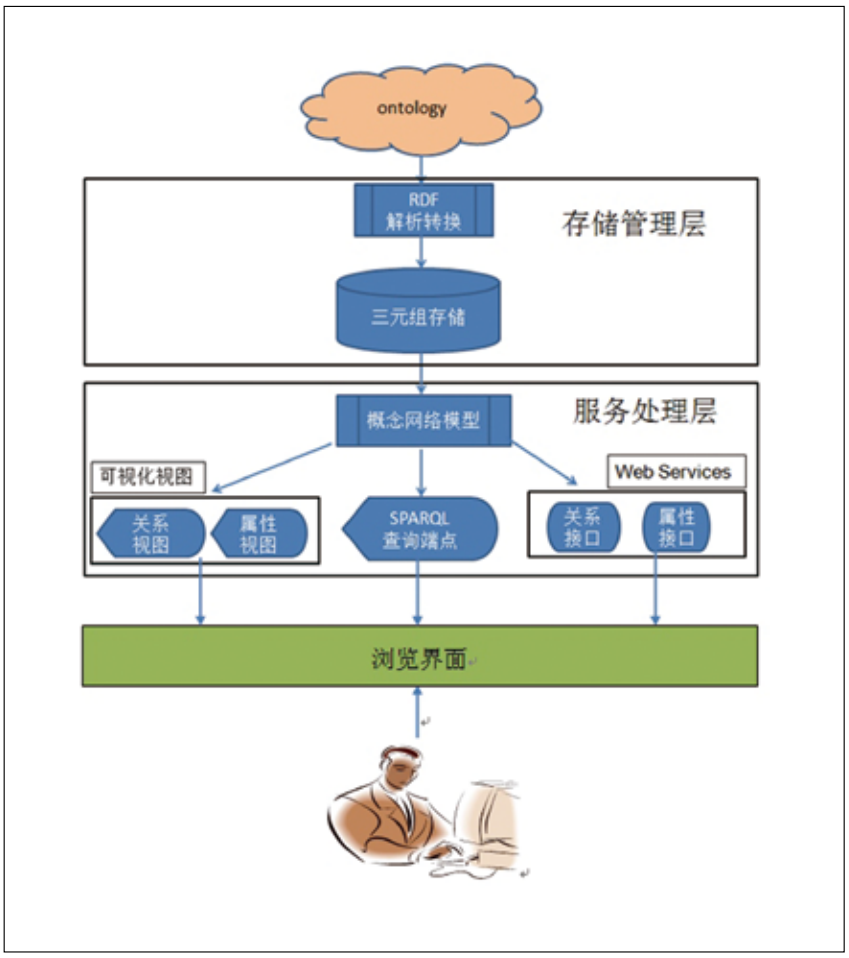

图1 本体管理平台架构图

显示所有信息　添加本体

ONTOLOGY_URI	ONTOLOGY_NAME	AUTHOR				
http://www.owl-ontologies.com/TechOntology.owl#	TechOntoloty	admin	查看	修改	删除	下载
http://www.owl-ontologies.com/Ontology1379483177.owl#	empty	admin	查看	修改	删除	下载
http://www.owl-ontologies.com/Ontology1379483178.owl#	vocgrid	admin	查看	修改	删除	下载
http://www.owl-ontologies.com/Ontology1379483179.owl#	vocgrid2	admin	查看	修改	删除	下载

首页　第1页　共1页　上一页　下一页　末页

图2 本体注册管理页面

3 检索模块

本体信息的检索是整个系统的基础和核心功能，其检索方式与资源组织形式关系密切。作为一种领域知识的整合工具，本体通过上下位概念、等同概念、参照概念等信息对重要的关键词进行语义扩展，

形成新的检索向量。检索模型首先寻找知识库中与用户输入词汇匹配的概念结点，再以此结点为结点，依次探寻与其相关的其他概念结点，直到没有新的概念可以发现。为了提高结果的精确性，检索界面提供了领域筛选功能。初次检索后，命中概念以及概念所在的本体文件会以列表的形式展示，用户进一步选择后，系统再次将概念词汇和本体URL发送给服务处理模块，Services将根据不同的需求组装相应的SPARQL语句进行检索[8]。SPARQL检索式由前缀、三元组变量及限定条件等部分组成，其语法结构与SQL语言类似，SELECT后面紧跟表示结果的检索变量，WHERE子句后则是以三元组形式表示的检索条件。为了解决同一词汇在多个领域中的使用问题，SPARQL检索式在进行三元组表示时，利用命名空间的前缀来限定元素。

SPARQL查询语言通过三元组图形模式进行匹配，三元组模式允许查询变量出现在主体、谓词或者客体的位置上。当用户从接口输入检索关键词时，输入的关键词与本体中的词汇进行相似度计算，映射为三元组中的各种元素。检索前缀由本体注册信息自动生成，当完成三元组分析和映射后，三元组列表与命名空间进行组配，形成SPARQL检索表达式。由于知识库中概念信息按一定的顺序进行排列，用户要检索的信息可能处于三元组的主语位置，也可能处于宾语位置，在映射为三元组元素的过程中，要考虑两种组配方式，避免信息的漏检。例如词汇“本体融合映射报告”可能存在于下面的事实知识中：

“kj:本体融合映射报告　kj:编写单位　kj:中信所信息技术支持中心”
“kj:知识组织与服务　kj:科研成果　kj:本体融合映射报告”

因此当用户输入上述词汇检索有关信息时，其三元组需映射为以下两种形式。

"select distinct ?s ?p where{?s ?p kj: 本体融合映射报告.}";
"select distinct ?p ?o where{kj: 本体融合映射报告 ?p ?o.}";

图3 SPARQL查询端

管理平台的SPARQL查询端提供了检索式的编辑功能，高级用户可以根据需要创建复杂的三元组模式进行知识库检索，查询结果将以列表形式返回。SPARQL查询端的设计效果如图3所示。

4 可视化展示

查询界面提供了多本体统一检索功能，用户输入待检概念后，系统会在知识库中所有的领域本体中查找相关概念，然后将命中的本体文件以列表的形式返回用户，待用户再次确认概念所属领域后，便会生成相应的结点网络模型，并通过可视化界面显示与输入概念有关的各种资源。可视化界面以输入概念为中心结点向外辐射，与其产生联系的其他概念会显示在关系视图中，而概念结点自身具有的信息则在属性视图中进行展示。若输入的概念为本体类，检索结果将返回该类的描述模型列表。若输入的为实例，则返回与该实例相关的其他概念结点及其属性信息,同时在左侧的结果面板中对概念结点进行详细说明。可视化界面使用了Ajax技术来实现结点形状、颜色渲染等动态效果，图形结点还具备放大、缩小以及拖拽功能，可以根据用户需求完成界面布局。关系视图中每个关系结点的位置取决于该节点在树中的层次，层次越深，圆环距离中心的根节点越远。视图网络的生成过程由以下7步骤组成：

（1）通过概念相似度计算和词汇扩展规范输入概念；
（2）将概念结点映射为三元组的主体和谓词部分；
（3）将本体url转换为命名空间前缀；
（4）按深度优先算法寻找主体和谓词间的关系链路；
（5）将连通链路按概念类和属性关系的搜索顺序依次排列，组合成检索三元组；
（6）从知识库中进行检索并将结果列表转换为视图模型展示给用户；
（7）根据用户的后续操作再次执行（1）-（6）步骤。

图4 可视化检索关系图

概念结点有两种组合方式：直接组合和间接组合。直接组合是指两个概念间具有明确的联系，其结构可以直接利用三元组进行表示。输入词汇和词汇间的联系分别映射为结点和关系弧，并以该输入词汇为中心元素向外辐射，形成知识地图。间接组合是指概念间无法通过一个明确的语义产生联系，但概念间存在一个关系链，可以利用其他概念作为中介，通过多个语义关系建立联系。当点击选中的外围结点时，处理程序会以当前结点作为中心结点探测与之相连的各种关系，然后通过环形的布局算法将树图转换为圆环结构，其中根节点为中心节点，关系结点分布在外围的圆环上。可视化界面效果如图4所示。

5 Web Services接口

管理平台除了提供本体信息的注册和检索功能外，还采用Web services方式为用户提供了远程调用功能。管理平台充当了服务提供者和代理两种角色，其内部操作方法通过services代理注册发布为服务接口。Services代理由action、model、sesame和services模块组成，其中action负责检索流程的控制和转发；model负责后台数据的封装和检索模型的生成；sesame模块负责知识库的初始化和连接等操作；services模块负责将用户输入的各项检索条件映射为SPARQL表达式并将其检索结果形式化。各模块间的调用关系如图5所示。

Web services接口由方法名、功能描述和wsdl文件组成，点击相应的wsdl后，页面指向服务的引用地址（例如QueryService方法的引用地址为http://168.160.18.252:8080/ontology/services/ws_findAllClass?wsdl）。用户在程序中可以直接访问服务地址生成相应的本地文件进行方法调用。接口注册和发布功能使用XFire工具包实现[9]，XFire是新一代的Java Web服务引擎，其配置简单，使用方便，易于与前台模块集成。为了方便用户使用，系统为每个Web services方法提供了示范用例，对services接口的功能和返回值类型作了说明。用户点击后，系统对相应的概念信息进行检索，并把命中信息以xml形式返回给用户。服务接口的详细信息如图6所示。

6 结语

如何有效地组织信息，并在此基础上提供有效的服务，一直是情报学领域研究的课题，其涉及的知识相当广泛，本文仅将就多本体资源的管理和使用技术进行了探讨。文章分析了一体化管理平台应具备的功能模块，进而通过各种工具实现了原型系统，对可视化检索、Web服务等前沿技术做了集成研究。为了验证集成管理平台的功能效果，笔者添加了多个科研本体资源进行了试用，为进一步实现语义检索和知识服务做了探索工作。本文实现的一体化管理服务平台有以下特点：

（1）本体注册模块具有整合

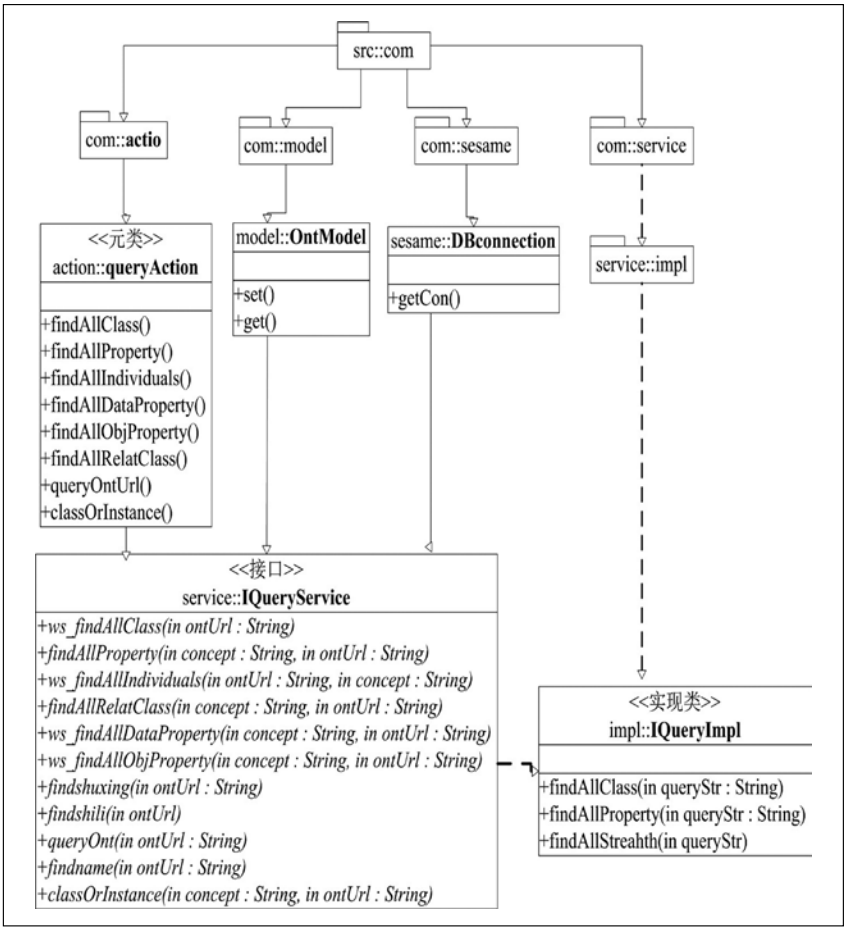

图5 模块调用关系图

功能。本体文件在存储过程中进行概念融合，将不同本体间的相关概念建立映射，检索模型采取网络扩展结构策略，以初始概念集合开始逐步扩展，提高了知识库的跨领域检索功能。

（2）可视化浏览功能。本体发布页面使用超链接与可视化模型相连，用户浏览本体结构时，可以利用可视化模型了解每个概念结点的关系连接和描述信息。

（3）提供远程使用接口。为了方便程序开发中使用知识库的检索功能，管理平台针对不同的检索需求进行了封装，并以接口的形式发布到服务页面，用户浏览相应的示范用例便可了解接口的功能和使用方法。

集成化管理系统为本体应用服务提供了支撑平台，但由于技术所限，目前的集成管理系统还不具有大规模应用的能力，后期将逐步采用分布式技术来实现多用户访问。同时也将继续完善Web服务的动态组装功能，以应对复杂多变的用户需求。

图6 Web services接口发布页面

参考文献

[1] IORDANOV B. HyperGraphDB: A Generalized Graph Database [C]// Proceedings of WAIM 2010 International Workshops, IWGD, 2010: 25-36.
[2] AllegroGraph 4.11 Introduction [EB/OL]. (2013-09-18) [2013-09-22]. http://www.franz.com/agraph/support/documentation/current/agraph-introduction.html.
[3] HARRIS S, LAMB N, SHADBOLT N. 4store: The Design and Implementation of a Clustered RDF Store [C]// 5th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS2009), 2009.
[4] 李慧颖,瞿裕忠.KREAG:基于实体三元组关联图的RDF数据关键词查询方法[J].计算机学报,2011,34(5): 825-835.
[5] 王鑫,冯志勇,杜朴风,等.Jingwei:一种分布式大规模RDF数据服务器[J].计算机研究与发展,2011,48(Z2):1-4.
[6] 袁平鹏,刘谱,张文娅,等.高可扩展的RDF数据存储系统[J].计算机研究与发展,2012(10):2131-2141.
[7] PRUD E, SEABORNE A. SPARQL Query Language for RDF [EB/OL]. (2008-01-15) [2013-09-22]. http://www.w3.org/TR/rdf-sparql-query/.
[8] User Document for Sesame 2.7 [EB/OL]. [2013-09-22]. http://openrdf.callimachus.net/sesame/2.7/docs/users.docbook?view.
[9] User Guide for XFire [EB/OL]. [2013-09-22]. http://xfire.codehaus.org/User%27s+Guide.

作者简介

许德山（1979-），男，助理研究员。研究方向：为知识组织、文本挖掘、语义Web。E-mail: xuds@istic.ac.cn
张运良（1979-），男，博士，副研究员。研究方向：为知识组织、知识工程、自然语言处理、文本自动分类。E-mail:zhangyl@istic.ac.cn

Design and Implementation of Integrated Ontology Management Platform

Xu Deshan, Zhang Yunliang / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: The work presented in this paper focuses on construction ideas of ontology management and services system. The functions that an integrated system should have are described firstly, as well as the architecture of the platform. And then, the implementation methods about various modules – ontology registration, triples retrieval, visualization and Web services, are presented in detail. Finally, the paper gives a summary about the current work and proposes the research emphasis in the future.

Keywords: Ontology management, Ontology publishing, Knowledge service

（收稿日期：2013-10-09）