

基于OWL+SKOS的期刊本体构建与应用*

罗婷婷¹ 李娇¹ 鲜国建^{1,2} 赵瑞雪^{1,2} 寇远涛^{1,2}

(1. 中国农业科学院农业信息研究所, 北京 100081; 2. 农业部农业大数据重点实验室, 北京 100081)

摘要: 为助力以期刊实体对象为中心的多源异构数据集成、关联与融合, 本文综合应用编辑工具Protégé、建模语言OWL和SKOS建立计算机可理解和可计算的期刊本体, 主要包括8个核心类、15个对象属性和33个数据属性的设计与创建; 并通过语义中间件Jena和WebVOWL等实现实例格式转换和可视化呈现, 主要包括1.5万余种期刊实例及关系的转换及近66万个RDF三元组的创建; 最后, 将构建的期刊本体应用于标引期刊实体概念、发布开放关联数据集、实例参引解析、SPARQL查询终端和语义推理等方面, 为各类包含期刊概念的数据分析提供参考, 为用户提供期刊相关的信息查询和交互服务。

关键词: Protégé; OWL; SKOS; Jena; WebVOWL; 期刊本体

中图分类号: G203

DOI: 10.3772/j.issn.1673-2286.2018.12.007

期刊资源是各类信息服务机构收藏的重要对象, 也是各类知识服务平台开展服务的重要资源之一^[1], 期刊资源可以从多方面进行描述, 如期刊的品种、卷期、收录情况、影响因子、历史沿革、出版地等。如何以这些实体概念为主体和纽带, 更科学智能地实现多源异构期刊资源的广泛汇聚, 知识组织、语义关联和深度融合, 首要任务是将现有的期刊异构元数据集成起来, 让计算机能够自动识别和处理, 必须建立统一标准的概念体系, 即本体^[2]。所谓本体, 1998年Studer将其定义为“共享概念模型的明确的形式化规范说明”。通过规范的期刊本体构建, 可以将多源异构、动态的期刊信息资源进行形式化规范说明, 使信息资源的语义“显性”化, 从而有助于改善期刊描述信息的可扩展性, 并促进信息的智能处理^[3]。

基于以上思考, 本文在继承复用国际标准和内容的基础上, 基于本体编辑工具Protégé, 结合使用本体建模语言(OWL)和简单知识组织系统(SKOS)来描述期刊本体框架模型和实例, 并基于语义中间件Jena和WebVOWL等进行各类实例的形式化描述、RDF三元组转换和可视化呈现, 并对后续应用作了展望。

1 相关现状

目前, 国内外对本体的研究项目很多, 本体已经在人工智能、知识工程、图书情报等领域得到广泛应用, 研究成果十分丰富, 并在开放关联数据项目影响下积累了大量高质量知识库, 如Zhishi.me、Freebase、DBPedia、Yago等^[4]。其中, Zhishi.me是第一个大规模中文开放关联数据集, 拥有9个自定义本体概念, 截至2015年11月24日, 已经拥有上亿个实体; Freebase是一种遵循知识共享协议的大规模开放结构化数据集, 目前含有1 200万个实体, 每个实体赋予了能够唯一标识的ID标记; DBPedia涵盖很多领域, 是世界上较大的多领域知识本体之一, 能自动随维基百科的变化而发展, 多语种, 包含685个概念, 拥有400万个实体; Yago是覆盖面较全、数据质量较高的大规模语义知识库, 目前含有1 700万个实体和1 500万个事实^[5]。

国内外有关期刊本体的构建研究已有所涉及, 目前国际上非常成熟的常用词表有DCMI(都柏林核心元数据集)^[6]、BIBO(书目本体)^[7]、FOAF^[8]、Nature Ontologies^[9]等。其中, 由Springer Nature出版集团构

*本研究得到公益性科研院所基本科研业务费项目“农业领域实体名称规范库构建方法研究与应用”(编号: JBYW-AII-2017-26)、中国农业科学院科技创新工程项目(编号: CAAS-ASTIP-2017-AII)重点任务“多源数据融合与知识组织增值”和国家科技图书文献中心专项“文献元数据集成加工管理系统”(编号: 2018XM34)资助。

建的语义发布模型^[10]，它具有核心模型-领域模型-实例数据集三层架构，支持Nature出版平台大规模数据的组织、关联与交互，简洁实用，逻辑清晰。但Nature本体以实用为本，核心类数量不多，注重类目的有用性，而不强调全面性^[11]。2018年1月，ISSN国际中心推出新门户——The ISSN Portal^[12]，为每个用户提供免费入口来获取更精准的连续出版物信息，主要包括ISSN、ISSN-L、标题、关键标题、出版国、媒体和URL，已注册200万ISSN，每年新增6万~7万ISSN，同时ISSN数据支持多格式下载，如MARC 21、MARCXML、RDF/XML、RDF turtle、JSON等。国内相关信息机构针对特定应用目的建立了内部使用的期刊规范库（如中文名称规范联合数据库和国家图书馆规范名称数据库），其主要针对个人名称、团体名称及会议名称等进行规范，基本不涉及期刊规范。在期刊规范方面，中国科学技术信息研究所面向知识评价建立了期刊规范名称数据库；中国医学科学院医学信息研究所构建了期刊规范文档^[13]，主要针对期刊沿革情况进行规范。

总体来说，当前关于期刊本体的研究翔实程度和表现形式不统一，未能从期刊影响因子、收录情况、沿革情况等多个方面对期刊进行描述，缺乏较系统全面的便于计算机理解处理的形式化描述。随着本体构建技术的日益成熟及各类官方权威数据库的相继开放，期刊本体的构建条件已相对成熟，形成全面的期刊本体系统必将具有广阔的应用前景。

2 期刊本体构建

2.1 总体建设思路

面向以期刊的品种、卷期、收录信息、影响因子、沿革情况等有关期刊的概念实体为中心开展多源异构数据资源汇聚、关联和融合的实际需求，在参考和继承复用国际标准和内容基础上，综合应用Protégé^[14]工具、OWL和SKOS建模语言、Jena和WebVOWL等语义中间件，开展期刊本体建模、实例转换和可视化呈现，建立计算机可理解和可计算的期刊本体模型与实例库，为该本体开放、共享和互操作等应用场景提供支撑（见图1）。

2.2 基础语料资源

在构建期刊本体概念体系过程中，为保证数据的

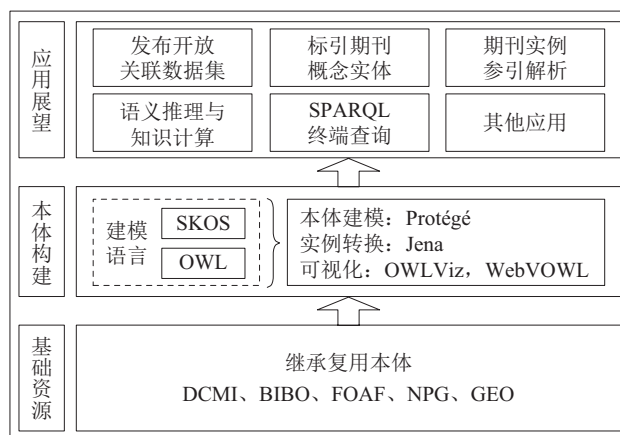


图1 期刊本体建设思路

规范性、系统性及权威性，为下一步互联互通奠定基础，部分继承复用了国际上现有通用的本体模型或描述规范，主要包括BIBO、FOAF、Nature Ontologies、DCMI。在搜集整理与期刊相关的实体对象信息时，通过网络及人工采集、加工、整理等多种途径，搜集国家农业图书馆、国家科技图书文献中心关于农业及相关领域的1.5万余种中外文期刊数据。其中中文期刊2 500多种、外文期刊1.25万余种，主要包括品种、卷期、收录情况、影响因子、历史沿革等信息，为下一步本体建模和实例转换奠定基础语料资源。

2.3 本体构建流程

目前常用的本体构建工具主要分为可视化手工构建工具和半自动化构建工具两类。根据实际的使用比较，可视化本体构建工具Protégé基本功能相对齐备，具体功能模块划分清晰，提供本体概念类、关系、属性及实例的创建，只需用户在概念层次创建领域本体模型，支持多种本体描述格式文本导出^[15]，因此笔者选择该工具来建立期刊本体模型，并采用OWL和SKOS描述与存储期刊实体。在期刊本体构建过程中引用其他本体中的概念，表1列出了相应的命名空间和前缀。

通过深入分析期刊资源的核心元数据项，本着尽可能全方位揭示期刊资源，使期刊本体更好地应用于期刊评价、论文质量排序、知识关联等应用场景，共设计与创建了8个核心类、15个对象属性和33个数据属性。下文重点阐述期刊本体中核心类、对象属性和数据属性的设计与创建，以及本体实例转换和可视化呈现等过程。

表1 期刊本体引用表

前缀	命名空间
dc	http://purl.org/dc/elements/1.1/
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
xml	http://www.w3.org/XML/1998/namespace
xsd	http://www.w3.org/2001/XMLSchema#
bibo	http://purl.org/ontology/bibo/
npg	http://ns.nature.com/terms/
skos	http://www.w3.org/2004/02/skos/core#
dcterms	http://purl.org/dc/terms/
aai_journal	http://http://linked.aginfra.cn/onts/journal#

2.3.1 核心类定义

期刊本体共定义了8个一级核心类,包括期刊品种(Journal)、期刊卷期(Issue)、影响因子(ImpactFactor)、收录情况(Holding)、出版地(PublicationPlace)、封面(img)、机构(Organization)、论文(Article)。其中,机构(Organization)类又分加工单位(ProcessingUnit)、出版商(Publisher)和主办单位(Sponsor)3个子类,Journal、Issue、Publisher继承复用了Nature本体, img继承复用了FOAF本体,还直接继承和复用了SKOS本体模型中的概念体系(ConceptScheme)、概念(Concept)等类及相关属性。若需要增加语义和逻辑严谨性,可综合应用等价于(Equivalent To)、子类(SubClass Of)和不相交(Disjoint With)等进一步约束限定类间关系(见图2)。

2.3.2 属性创建与描述

通过部分继承Nature本体模型,复用并自定义了15个一级核心对象属性,主要包括拥有影响因子(has Impact Factor)、拥有卷期(has Issue)、收录(hold)、期刊沿革(periodicals Evolution)、出版商(has Publisher)、网址(webpage)、出版地(is Published In)等(见图3)。其中4个对象属性复用了Nature本体,主要包括has Publisher属性描述期刊与出版商之间的关系, webpage属性描述与期刊相关的网址信息, has Image属性描述期刊与封面、图片之间的关系, has Issue属性描述期刊卷期; 11个对象属性为自定义,主要包括has Impact Factor属性描述期刊的影响因子等相关信息,

hold属性描述期刊被收录的情况, has Sponsor属性描述期刊主办单位的相关信息, has Issue属性描述期刊品种与卷期之间的从属关系, is Published In属性描述期刊出版地信息, periodicals Evolution属性描述期刊历史沿革信息等。此外,在期刊沿革(periodicals Evolution)下还自定义了17个子对象属性来描述期刊之间的沿革关系,主要包括继承(inherit)、部分继承(partly Inherit)、吸收(absorb)、部分吸收(partly Absorb)、合并(merge)等。

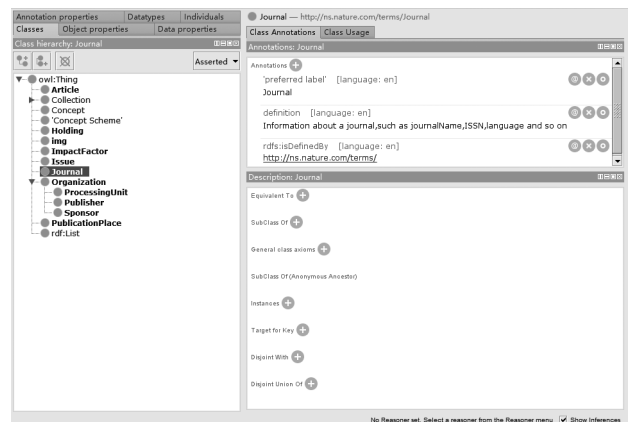


图2 期刊本体核心类定义

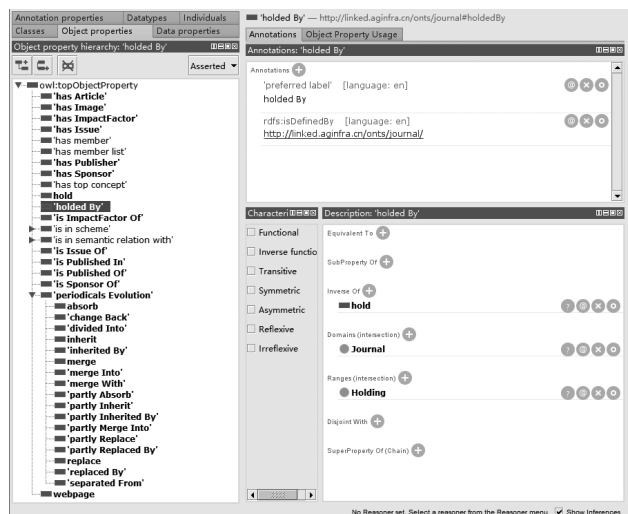


图3 期刊本体对象属性定义

数据属性用于描述类基本信息,取值为XML Schema数据类型值或者RDF Literal,期刊本体中定义了ISSN、EISSN、主题分类(dc:subject)、语种(dc:language)、收录来源(holdingSource)、收录版本(holdingVersion)、收录结束年份(holdingYearEnd)、收录开始年份(holdingYearStart)、影响因子年份(impactorYear)、影响因

子来源 (impactorSource)、期刊影响因子 (journal ImpactFactor)、出版年 (publicationYear)、出版起始年 (publicationStartYear)、出版结束年 (publicationEndYear)、出版频率 (publishingFrequency) 等33个数据属性。其中, ISSN、EISSN等复用了Nature本体, 主题分类 (dc:subject)、语种 (dc:language) 复用了DCMI描述规范, 收录来源 (holdingSource)、收录版本 (holdingVersion)、收录结束年份 (holdingYearEnd) 等为自定义的数据属性。为增强语义和逻辑严谨性, 还可对这两类属性值域和定义域的取值作更详尽的约束限定 (见图4)。

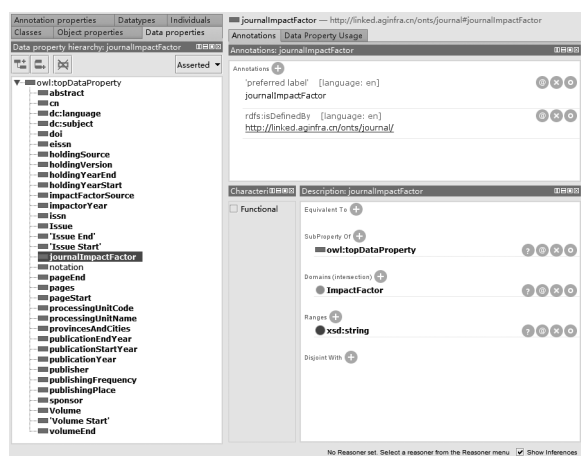


图4 期刊本体数据属性定义

2.3.3 实例创建与转换

本体实例创建与转换是指根据类和属性的约束给本体添加实例, 实现属性与资源的关联创建。尽管Protégé工具支持对象类实例的创建, 但经前面基础语料资源的搜集整理后, 形成的基础数据库体量大, 人工逐条逐项录入效率太低。为此, 基于Jena语义中间件、继承引用现有本体和本文定义的类、实例及属性进行二次开发, 对Protégé生成的OWL文件进行读取和分析, 研制期刊本体实例批量转换为RDF三元组工具。

该工具可基于Jena实现RDF/XML、Turtle、N-Triples和JSON-LD等多种格式输出。目前已应用该工具实现1.5万余种期刊实例及关系的转换, 创建了近66万个RDF三元组。以期刊“Virus Research: An International Journal of Molecular and Cellular Virology”为例, 转化后对应的部分三元组片断如图5所示。

```
<skos:Concept rdf:about="http://linked.agininfra.cn/journals/DBC7AA0D-0C36-4FFF-9DAF-80629F3F31FF">
  <skos:inScheme rdf:resource="http://linked.agininfra.cn/journals"/>
  <skos:prefLabel>Virus Research: An International Journal of Molecular and Cellular Virology</skos:prefLabel>
  <skos:altLabel>VIRUS RESEARCH</skos:altLabel>
  <skos:altLabel>VIRUS RES</skos:altLabel>
  <skos:altLabel>Virus Res.</skos:altLabel>
  <owl:sameAs rdf:resource="https://portal.issn.org/resource/ISSN/0168-1702"/>
  <np:publicationStartYear>1984</np:publicationStartYear>
  <np:geissn>0168-1702</np:geissn>
  <dc:language>eng</dc:language>
  <dc:subject>R37</dc:subject>
  <aii:journal:publisher>Elsevier</aii:journal:publisher>
  <rdf:type rdf:resource="http://ns.nature.com/terms/Journal"/>
  <aii:journal:heldBy rdf:resource="http://linked.agininfra.cn/journalholding/SCIE_2014_2014"/>
  <aii:journal:heldBy rdf:resource="http://linked.agininfra.cn/journalholding/SCIE_2015_2015"/>
  <aii:journal:heldBy rdf:resource="http://linked.agininfra.cn/journalholding/SCIE_2016_2016"/>
  <aii:journal:heldBy rdf:resource="http://linked.agininfra.cn/journalholding/SCIE_2017_2017"/>
  <aii:journal:hasImpactFactor>
    <aii:journal:ImpactFactor rdf:about="http://linked.agininfra.cn/journalimpactfactor/sci_2008_2.429">
      <aii:journal:ImpactFactor>2.429</aii:journal:ImpactFactor>
      <aii:journal:ImpactFactorYear>2008</aii:journal:ImpactFactorYear>
      <aii:journal:ImpactFactorSource>SCI</aii:journal:ImpactFactorSource>
      <skos:inScheme rdf:resource="http://linked.agininfra.cn/journals"/>
      <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    </aii:journal:ImpactFactor>
  </aii:journal:hasImpactFactor>
  <aii:journal:hasImpactFactor>
    <aii:journal:ImpactFactor rdf:about="http://linked.agininfra.cn/journalimpactfactor/sci_2009_2.563">
      <aii:journal:ImpactFactor>2.563</aii:journal:ImpactFactor>
      <aii:journal:ImpactFactorYear>2009</aii:journal:ImpactFactorYear>
      <aii:journal:ImpactFactorSource>SCI</aii:journal:ImpactFactorSource>
      <skos:inScheme rdf:resource="http://linked.agininfra.cn/journals"/>
      <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    </aii:journal:ImpactFactor>
  </aii:journal:hasImpactFactor>
  <aii:journal:hasImpactFactor>
    <aii:journal:ImpactFactor rdf:about="http://linked.agininfra.cn/journalimpactfactor/sci_2017_2.484">
      <aii:journal:ImpactFactor>2.484</aii:journal:ImpactFactor>
      <aii:journal:ImpactFactorYear>2017</aii:journal:ImpactFactorYear>
      <aii:journal:ImpactFactorSource>SCI</aii:journal:ImpactFactorSource>
      <skos:inScheme rdf:resource="http://linked.agininfra.cn/journals"/>
      <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    </aii:journal:ImpactFactor>
  </aii:journal:hasImpactFactor>
</skos:Concept>
```

图5 部分三元组片断截图

在转换过程中, 采用了OWL和SKOS相结合的方式。如描述期刊“Virus Research: An International Journal of Molecular and Cellular Virology”实例时, 遵循建立稳定、可访问、可解析的唯一标识符Http URI (Cool URI)^[16]原则, 分配了HttpURI唯一标识“http://linked.agininfra.cn/journals/DBC7AA0D-0C36-4FFF-9DAF-80629F3F31FF”, 描述所属类型时, 除指定Journals类 (http://linked.agininfra.cn/journals) 外, 还指定其属于skos:Concept, 申明其也属于SKOS:Concept的一个概念。

在描述期刊中英文等语种名称和其他各类缩写时, 也应用了skos:prefLabel和skos:altLabel标签及语种来描述, 并通过对象属性skos:inScheme属性将该实例归在期刊本体体系内 (http://linked.agininfra.cn/journals)。通过对象属性“aii:journal:heldBy”和“aii:journal:ImpactFactor”描述期刊收录情况及影响因子等信息 (见图6), 收录情况HttpURI类类似为“http://linked.agininfra.cn/journalholding/SCIE_1999_1999”, 该规则表示期刊在起始与结束年份均为1999年且被SCI收录; 影响因子HttpURI类类似为“http://linked.agininfra.cn/journalimpactfactor/sci_2003_1.719”, 表示SCI收录期刊2003年影响因子为1.719。此外, 还应用owl:same属性通过ISSN号与ISSN Portal建立关联。经过这样规范和明确的定义, 将有助于计算机理解和智能处理有关期刊等实体对象的

信息,也将为语义推理和知识计算等奠定基础。

```
<aii_journalHeldBy rdf:resource="http://linked.aginfa.cn/journalHolding/SCIE_1999_1999"/>
<aii_journalHasImpactFactor>
<aii_journalImpactFactor rdf:about="http://linked.aginfa.cn/journalImpactFactor/sci_2003_1.719">
<aii_journalJournalImpactFactor>
<aii_journalImpactYear>2003</aii_journalImpactYear>
<aii_journalImpactSource>SCI</aii_journalImpactSource>
<skos:inScheme rdf:resource="http://linked.aginfa.cn/journals"/>
<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
</aii_journalImpactFactor>
</aii_journalHasImpactFactor>
```

图6 收录情况和期刊影响因子的实例片段截图

2.3.4 本体存储

期刊本体及实例均可转化为XML/RDF三元组,采用Virtuoso来存储RDF三元组,其最重要的特征是可以作为一个图数据库支持SPARQL查询,在Virtuoso的强大支持下,我们可以高效、方便地设计和实现应用场景。

3 应用实践

经上述研究,初步构建了以期刊这一实体对象为中心的OWL及SKOS本体模型和RDF三元组实例库,本质上已成为一个国内外学术期刊的关联数据集。语义网络环境下,数据之间关联越多,数据越有价值。为此,本文研究构建的期刊本体可作为关联数据网络中的基础知识中间件,作为互联互通描述期刊、影响因子、收录情况等各类数据资源的联接点和枢纽。具体应用场景包括但不限于下列类。

(1) 本体可视化。本体可视化是指以一种直观的图形化方式展示本体,自动、动态地显示本体内概念间关系。除Protégé自带插件OWLviz可实现本体模型结构层次图形化显示外,本文应用WebVOWL^[6]进行可视化试验。WebVOWL是一个本体可视化交互的Web应用程序,并结合力导向布局图表示本体结构。在将WebVOWL本地化部署基础上,通过其提供的owl2vowl.jar转换包将实例转为符合要求的json格式文件,并上传后可实现可视化展示。

(2) 期刊实例参引解析。在语义数据网络环境下,本研究基于nginx反向代理机制与技术和virtuoso的SPARQL查询接口对接,针对期刊实体提供了一套计算机能自动理解与计算的形式化描述、解析和互操作机制,支持以HttpURI解析获取数据网络中与该期刊实例有关的所有RDF三元组,如通过“http://linked.aginfa.cn/journals/EE2BB1E3-B120-46DC-8054-4E3B3E90272B”可获取期刊的影响因子、收录情况、出版情况等多方面一体化描述信息(HTML页面见图

7),也可返回json或xml等格式以便在第三方系统集成和互操作。

p	o
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://ns.nature.com/terms/journal
http://www.w3.org/2002/07/owl#sameAs	https://portal.isn.org/resource/ISSN/0190-3187
http://www.w3.org/2004/02/skos/core#altLabel	"INT FAM PL"
http://www.w3.org/2004/02/skos/core#altLabel	"INT FAM PLAN PERSPEC"
http://www.w3.org/2004/02/skos/core#altLabel	"INT FAM PLAN PERSPECT"
http://www.w3.org/2004/02/skos/core#altLabel	"INTERNATIONAL FAMILY PLANNING PERSPECTIVES"
http://www.w3.org/2004/02/skos/core#altLabel	"Int. Fam. Plan. Perspect."
http://purl.org/dc/elements/1.1/subject	"R169"
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2004_0.791
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2005_1.0
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2007_1.342
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2002_0.681
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2009_1.429
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2001_1.035
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2003_1.617
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2006_1.538
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2008_1.649
http://linked.aginfa.cn/onts/journal#hasImpactFactor	http://linked.aginfa.cn/journalImpactFactor/sci_2010_2.118
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2001_2001
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2002_2002
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2003_2003
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2004_2004
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2005_2005
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2006_2006
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2007_2007
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2008_2008
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2009_2009
http://linked.aginfa.cn/onts/journal#heldBy	http://linked.aginfa.cn/journalHolding/SSCI_2010_2010
http://linked.aginfa.cn/onts/journal#publisher	"Alan Guttmacher Institute"

图7 期刊实例参引解析示例图

(3) SPARQL终端查询。构建遵循SPARQL1.1协议的查询终端接口,支持用户或第三方系统通过SPARQL语言查询有关期刊及各类关联关系信息,如SPARQL查询语句可查询2017年SCI收录且影响因子等于2.0的期刊HttpURI、期刊名称、影响因子等信息。

(4) 期刊概念实体标引与智能检索。类似SKOS描述叙词表及词间关系机制,可基于期刊实例不同语种的skos:prefLabel和skos:altLabel文本描述,进行各类资源中有关期刊对象实体的概念标引,标引结果是将期刊正式名称或可解析的HttpURI标识符植入各类资源,也可在检索期刊时进行中英双语和各类名称自动语义扩展检索。

4 结语

本文研究了现有的一些关于期刊本体的系统和标准,较全面地搜集整理了期刊母体各方面的数据信息,构建了一个基于OWL和SKOS的计算机可理解、可计算的期刊本体。构建的本体和个体可以作为一个以期刊概念为中心的知识中间件,如可视化、HTTP URI解析和SPARQL查询,以及支持以期刊为中心的实体的集成、关联和融合。

未来的工作主要关注本体的改进,如将期刊本体中出版商等作为对象进行描述,使得本体模型具有较好的完备性、表达性和合理性。在本体实例转换过程

中,本文完成了期刊影响因子、收录情况、期刊别名的实例转换,还未完成期刊沿革的实例转换,这些均在下一步工作中进行改进和完善。此外,将继续开拓新的应用场景。①开放链接数据集。为使本体在更广的范围内得到应用,并实现更高的应用价值,我们根据链接数据发布的原理,在LOD云上发布上层本体和RDF实例库,便于在更大范围内进行开放,以被发现和关联应用。②语义推理和知识计算。可以基于对象属性(如hold、periodicals Evolution、has ImpactFactor等)对收录类别、沿革情况、影响因子等进行推理、推荐和计算。在检索过程中实现智能检索和语义扩展,为用户提供期刊相关信息查询和交互服务。

参考文献

- [1] 王军辉,方安,任慧玲,等. 期刊规范文档建设方案与应用场景研究[J]. 数字图书馆论坛, 2015 (7): 9-13.
- [2] 王石榴,谢文亮. 数字图书馆建设中科技期刊元数据异构共享研究[J]. 图书馆工作与研究, 2015 (2): 53-56.
- [3] 徐享忠,杨建东,汤再江. 基于OWL的本体建模与推理研究[J]. 装甲兵工程学院学报, 2017, 31 (5): 81-85, 96.
- [4] 邹鼎杰. 基于关联数据的书目语义检索[J]. 数字图书馆论坛, 2018 (4): 54-58.
- [5] 朝乐门,张勇,邢春晓. DBpedia及其典型应用[J]. 现代图书情报技术, 2011 (3): 80-87.
- [6] Dublin Core Metadata Initiative [EB/OL]. [2018-10-06]. <http://dublincore.org/>.
- [7] The Bibliographic Ontology [EB/OL]. [2018-10-06]. <http://www.bibliontology.com/>.
- [8] FOAF (2000-2015+) [EB/OL]. [2018-10-06]. <http://www.foaf-project.org/>.
- [9] Nature Core Omology [EB/OL]. [2018-10-06]. <http://ns.nature.com/terms/>.
- [10] Nature Ontologies [EB/OL]. [2018-10-06]. <http://www.nature.com/ontologies/>.
- [11] 李芳,纪姗姗,周毅,等. 语义出版本体模型的构成分析——以SPAR本体和Nature本体为例[J]. 图书情报工作, 2017, 61 (24): 123-134.
- [12] TTL file extension [EB/OL]. [2018-10-06]. <http://portal.issn.org/>.
- [13] 王军辉,虞昌亮,方安,等. 期刊规范文档建设实践研究[J]. 图书馆杂志, 2017, 36 (5): 41-46, 120.
- [14] A free, open-source ontology editor and framework for building intelligent systems [EB/OL]. [2018-10-06]. <https://protege.stanford.edu/>.
- [15] MUSEN M A, TEAM T P. The Protégé Project: A Look Back and a Look Forward[J]. Ai Matters, 2015, 1 (4): 4.
- [16] Cool URIs for the Semantic Web W3C Interest Group Note 03 December 2008 [EB/OL]. [2018-10-06]. <http://www.w3.org/TR/cooluris/>.

作者简介

罗婷婷,女,1985年生,硕士,助理馆员,研究方向:知识组织、关联数据、信息管理与信息系统。

李娇,女,1989年生,硕士,助理馆员,研究方向:知识组织、关联数据、语义检索。

鲜国建,男,1982年生,博士,副研究馆员,研究方向:知识组织、关联数据、语义出版、信息管理与信息系统,通信作者,E-mail: xianguojian@caas.cn.

赵瑞雪,女,1968年生,博士,研究员,研究方向:信息管理与信息系统、信息资源管理、知识组织及数字图书馆。

寇远涛,男,1982年生,博士,副研究馆员,研究方向:数字图书馆理论与技术、信息管理与信息系统。

Construction and Application of Journal Ontology Based on OWL+SKOS

LUO TingTing¹ LI Jiao¹ XIAN GuoJian^{1,2} ZHAO RuiXue¹ KOU YuanTao^{1,2}

(1. Agricultural Information Institute of CAAS, Beijing 100081, China;

2. Key Laboratory of Agricultural Big Data, Ministry of Agriculture, Beijing 100081, China)

Abstract: In order to support the integration, correlation and fusion of periodical entity object-centered multi-source, this paper establishes a computer understandable and computable journal ontology by adopting ontology editing tool Protégé and modeling language OWL, SKOS. It mainly includes the design and creation of 8 core classes, 15 object properties and 33 data properties, and then we use semantic middlewares Jena and WebVOWL to realize instance transformation and visualization of more than 15 000 periodicals with nearly 660 000 RDF triples created. Finally, the journal ontology is used for tagging periodical entity, publishing open linked dataset, referencing country instance, SPARQL endpoint query and semantic reasoning and then can provide a comprehensive reference for all kinds of data analysis with the concept of journal, and serve users with information query and interactive.

Keywords: Protégé; OWL; SKOS; Jena; WebVOWL; Journal Ontology

(收稿日期: 2018-11-23)