

CSSCI 本体概念模型的构建和描述 *

王 昊 苏新宁

摘 要 现有的数据结构导致 CSSCI 信息服务产生了检索方式简单化和隐含知识获取难两大问题。通过对 CSSCI 数据结构和内容特点的分析,提出基于本体面向对象的知识组织方式来重构 CSSCI 学术资源知识结构的解决方案,并试图通过建立概念层次结构和设置概念属性来构建 CSSCI 本体概念模型,用于表达该领域的知识组织。可以使用 Protégé 工具以 OWL 编码方式和图形可视化方式对概念模型进行合理描述。图 5。表 1。参考文献 7。

关键词 CSSCI 本体概念模型 知识组织 本体构建 本体描述 多元关系分解

分类号 G250.76

ABSTRACT The existing data structure has caused two problems; one is that the retrieval mode is too simple, and the other is that implicit knowledge is difficult to obtain. On the basis of the analysis on characteristics of the CSSCI data structure and content, the authors propose the solution of changing knowledge structure of CSSCI academic resource based on object oriented knowledge organization mode of ontology, and try to construct conceptualization of CSSCI ontology by building concept hierarchy and setting concept attributes, used to express knowledge organization in this field. They point out that the CSSCI ontology conceptualization could be rationally described by Protégé with OWL coding method and graphical visualization method. 5 figs. 1 tab. 7 refs.

KEY WORDS CSSCI. Conceptualization. Knowledge organization. Ontology construction. Ontology description. N-ary relations decomposition.

CLASS NUMBER G250.76

1 引言

中国社会科学引文索引(Chinese Social Science Citation Index,简称 CSSCI)自诞生以来,以其规范、权威的信息检索和引文分析服务得到了一致认同。然而就现有的 CSSCI 信息服务来看,存在着两大问题:①检索方式简单化。仅能查询来源文献和被引文献,其他学术资源均围绕两者展示。②隐含知识获取难。学术资源间关联揭示不充分,研究热点表达不明确,学科发展趋势难以把握。

为了进一步完善 CSSCI 学术资源服务,需要改变其原有的数据结构;随着语义网(Semantic Web)^[1]概念的提出,具有语义描述能力的本

体机制获得了广泛的关注。为此,笔者提出基于本体来组织 CSSCI 学术资源的方案,试图借助本体对知识的有效组织,从本质上改变 CSSCI 现有的数据结构,建立基于本体的 CSSCI 学术资源网络模型,以新的语义检索服务和引文分析模式代替原有的服务模式,解决用户需要更完善的知识服务和现有 CSSCI 提供学术资源服务相对落后之间的矛盾。

CSSCI 本体(以下称 CSSCI_Onto)的构建从总体上可以分为两个部分,即知识组织和知识描述。知识组织是指建立知识结构框架,明确知识领域中的概念、概念关系以及概念间约束(公理),即概念模型的建立;知识描述则是基于知识组织所定义的知识框架,标注领域中具体的对象、对象间关联和约束。本文重点探讨 CSSCI_

* 本文系“江苏省研究生培养创新工程——科技创新计划”人文社科项目“基于本体的学术资源网络模型研究”(项目编号:CX07B-252r)研究成果之一,并得到了南京大学人才引进科研启动基金的资助。

Onto 构建的知识组织过程和方法,试图建立能够用于指导 CSSCI 数据语义标注的本体概念模型。

2 CSSCI_Onto 的建模过程

CSSCI 从全国 4000 多种期刊中遴选出 400—500 种质量较高、影响较大、编辑出版比较规范的中国大陆出版的人文社会科学精品期刊作为来源期刊,收录刊载论文及其相关的关键词、作者、机构、期刊、学科、被引文献、时间等重要的学术资源。

2.1 CSSCI 来源数据的基本情况

从数据结构来看,CSSCI 来源数据以关系数据库的形式存在,共包含 6 个关系:①来源文献表:收录 CSSCI 选用刊所载的期刊论文的基本信息;②来源作者表:收录来源文献的作者信息,以文献编号与来源文献表建立关联;③被引文献表:收录来源文献的引用文献信息,以文献编号与来源文献表建立关联;④期刊载文表:收录来源期刊各期的载文情况,以期刊号与来源文献表建立关联;⑤来源刊表:收录了 CSSCI 来源刊的基本情况,以期刊代码与期刊载文表建立关联;⑥字典表:收录了在上述各表中使用的类型代码及其中文释义,包括引文语种、来源文献类型、引文类型、基金类别、机构类型、地区代码等。

从数据内容来看,CSSCI 庞大的数据量包含了丰富的学术资源实例知识。本文收集了 2000—2006 年共 7 年的实例数据作为研究对象,其中包括近 56 万个主题实例,50 余万篇来源文献,160 余万篇被引文献,涉及近 31 万名作者。在这些实例集合中,存在一些隐含关联在原始数据中没有直接体现的信息,需要进一步挖掘和探索。如某一作者的关联作者、某一具体来源文献的相似文献、某一机构的同被引机构、某一年度的发文引文量等。

2.2 CSSCI_Onto 构建模型

根据来源数据特点以及学术资源本体的构建目标,本文建立如图 1 所示的 CSSCI_Onto 构建模型。该模型共有 6 个子模块,分别是概念抽取、数据整理、本体实例化、CSSCI 本体库、本体评价和本体应用。从总体上确立了本体构建的基本思想:首先从 CSSCI 元数据中抽取学术资源概念,建立概念模型,用于指导学术资源实例及其关系的抽取;由数据整理和本体实例化子模块在算法库的支持下共同完成实例的语义标注;本体概念模型和实例库一起进入本体库进行存储和描述;本体库一方面接受领域专家的正确性和合理性评价,另一方面提供给用户使用,根据实际应用情况进行有效性评价;在评价中发现问题,并对概念模型进行调整,进而完善本体。

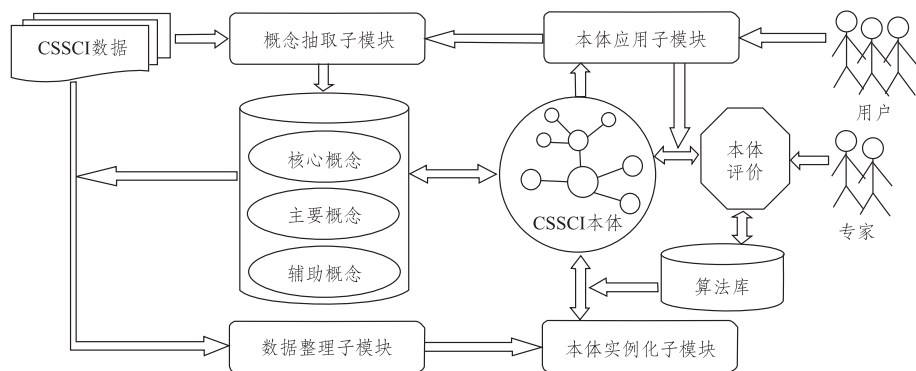


图1 CSSCI 学术资源本体构建模型

基于上述的本体构建思想,可以将 CSSCI_Onto 的构建过程分为 6 个步骤:

(1) CSSCI 概念层次结构的建立。包括:①概念的抽取和扩展;②概念层次结构的生成。

(2)定义概念属性。仅仅通过概念不足以确切地描述一个领域,还需要定义属性来表示概念间的语义联系。因此,概念属性的定义过程实际上是关联概念的探测过程,同时也规定了实例之间可能存在的关系。

(3)概念实例化。也称为语义标注,即根据概念模型从原始数据中获得实例,并为每个实例属性设置属性值,建立实例间关联。CSSCI 庞大数据量的特点决定了其概念实例化是本体构建的难点,也是本体是否能够被有效应用的关键。

(4)本体存储和展示。本体库由概念库和实例库两部分构成:①对于概念模型,可以使用 Protégé-OWL 进行系统规划,以 OWL 语言描述,其层次结构可使用 OWLViz、TGVizTab 等图形化插件展示;②对于学术实例,可以采用自动生成 OWL 文件或关系型数据库等形式进行存储,并以列表或图形方式展示实例间的关系。

(5)本体评价。笔者认为可以通过用户和

领域专家从评价准则和实践应用两个方面进行:①根据领域专家的专业知识,按照一定的评价原则(如 Gruber 提出的 5 大准则^[2]等)进行本体评价;②由用户在本体应用实践过程中发现问题,完善本体。

(6)本体应用。主要包括基于本体的知识检索服务和引文分析服务。

上述 6 个步骤并不采用“瀑布”模式,而是基于“循环式增量迭代”模式(见图 2)实际展开,即在下一个步骤开始之后如果发现问题,及时返回到上一步骤中进行修改,前后步骤之间存在迭代(overlap);整个开发过程“循环往复”,通过领域专家和用户的评价和实践应用发现问题,继而进行本体再开发。其中,前两个步骤即建立概念层次结构和定义概念属性属于领域知识组织的任务范畴,它们的完成标志着 CSSCI_Onto 概念模型的实现。

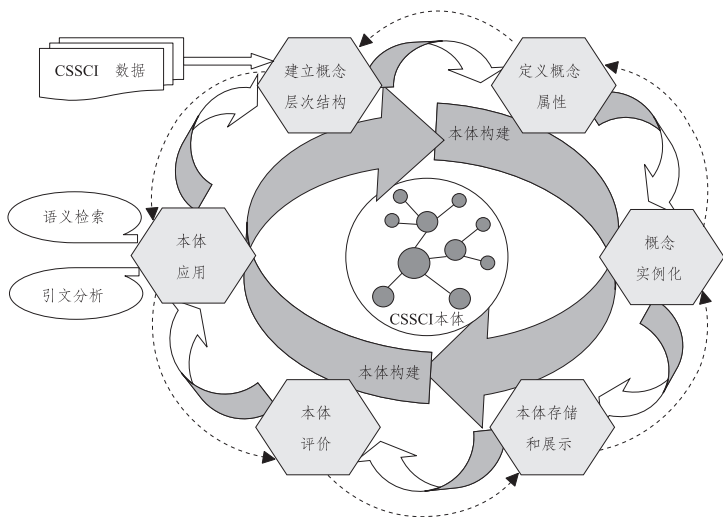


图 2 基于“循环式增量迭代”模式的 CSSCI 知识本体的构建过程

3 CSSCI_Onto 概念层次结构的建立

本体是对概念与概念关系的明确描述。概念是本体中的基本元素,只有明确了概念的定义和划分,才能在此基础上通过属性设置建立

概念之间的关系。

3.1 CSSCI 主要学术概念的抽取

目前建立领域本体概念模型的方法主要有 4 种^[3]:①自顶向下(Top-Down)方法。存在两种理解:一是鉴别出领域中所有概括性抽象概念,然

后逐步细化,由抽象概念派生出具体概念;二是以现有的上层本体(通用或领域本体)为基础分离出领域或应用本体^[4],如基于 Sensus 本体的剪枝等。②自底向上(Bottom-Up)方法。先确定领域中大多数具体性概念,然后通过二义性处理、归纳、概念聚类等手段从具体概念中泛化(generalize)出综合性的抽象概念。③混合(Hybrid)方法。结合上述两种过程,首先定义顶层的综合性抽象概念和底层的特殊性具体概念,然后分别从细化和泛化两个方向扩展,最终关联到同一中间层概念。④由里而外(Inside-Out)方法。亦称为核心扩展(Middle-Out),也是结合①和②,但和混合方法不同的是,其扩展的起点是中间层概念,即首先确定领域中的核心概念,然后泛化和细化出其它概念,以扩展本体概念模型。

上述4种方法在实际本体构建过程中都有广泛采用,但是适用环境不同。方法①适用于原来就存在本体概念模型或抽象概念容易获取的情况;方法②适用于具体概念容易获得,但是所属类别不清或概念间关系不明的情况,可以采用层次聚类或形式概念分析等自动化技术进行概念泛化;方法③是在①和②的条件都足够成熟时,为了加快本体概念模型的设计或使本体构建始终保持一个合理的方向而采用的一种方法,然而在实际本体开发中,这种方法的条件往往很难得到满足;方法④适用于领域的核心概念容易获得的情况下。具体采用何种方法主要依赖于领域知识自身的特点(数据环境)以及本体开发人员对领域知识结构的理解程度。

在 CSSCI_Onto 概念模型的构建中,没有现成可利用的上层本体存在,现有资源以关系型数据库的形式存在,具有明确的元数据结构,主要概念作为关系的主关键字或外部关键字。根据领域数据的这一特点,本文决定采用核心扩展方法来构建 CSSCI_Onto 概念模型。

考察 CSSCI 元数据,在领域专家的帮助下,经过慎重的识别、分析和统计,笔者确定主题、文献、期刊、学科、作者、单位、时间等7个概念作为主要概念,并将其中的主题定为核心概念。主要概念是 CSSCI_Onto 的描述重点,其本身具有重要的实际应用意义,满足无二义性、互不相

交的特点,CSSCI 中的其他知识都是用于描述主要概念而存在的。

3.2 CSSCI 概念集合的扩展

明确了 CSSCI_Onto 的主要概念之后,即可对本体雏形进行扩展。扩展的方向取决于主要概念在概念层次结构中所处的位置,这就需要明确 CSSCI_Onto 概念层次结构中概念关系的组织方式。一般来说,领域中的知识组织存在两种方案:

(1)使用部分-整体关系(part-of)组织概念层次结构,即认为下位类是上位类的部分,例如“市”是“省”的下位类。基于这种观点,根据对象范围,可以把整个领域分为来源文献和引用文献两大类,其他概念作为两者的组成部分。例如来源文献包括主题、作者、期刊、机构、学科等部分,机构又分为高校、研究所、社科院等。这是一种面向结构的知识组织方式。

(2)使用子-父类关系(kind-of)组织概念层次结构,即认为下位类是上位类的子类,例如“省”和“市”都是“地区”类的子类。基于这种观点,根据对象类型,可以把整个领域分为主要概念和辅助概念,再从这些概念中派生下位类,上下位类之间是继承关系(IS_A)。这是一种面向对象的知识组织方式。

本文选择方案(2),原因在于:①面向对象的知识组织方式以其封装性、继承性等特点更符合人们的思维方式,而且其稳定性和可扩展性已经在软件工程领域得到了实践的检验;②概念继承的思想可以避免冗余类的产生,例如在来源文献和引用文献类中,都存在年代类,按照方案(1),年代类将被定义两次。

基于子-父类关系的概念层次结构,本文将主要概念置于层次结构的顶层,于是 CSSCI_Onto 核心扩展的方向主要有两个:

(1)同级扩展

对 CSSCI 元数据再进行分析、总结和抽取,结合 CSSCI_Onto 的应用目的和领域专家的意见,可以获得领域中其他具有应用价值的概念作为辅助概念,和主要概念一起作为顶层概念,完成对整个 CSSCI 领域的知识覆盖。在 CSSCI_

Onto 中存在包括地区、基金、项目、类型、出版社等在内的 5 大辅助概念。这些概念一方面可以作为提供信息检索服务的辅助元素,例如检索某年度国家社科基金发表的所有论文等等;另一方面,也是引文分析的重要对象,例如统计某地区各年度的发文、引文情况,可以分析获得该地区学术活动的发展变化等。

(2) 向下扩展

根据已经定义好的抽象父类,自顶向下逐步细化,扩展出下位子类。概念细化实际上是一个概念分类或预分类的过程,包括:①根据数据收录情况进行细化。CSSCI 中包含来源文献和被引文献相关信息,可以根据来源或被引关系将概念进一步分类,如文献可细化为来源文献和被引文献两个子类,类似的还有期刊、作者等;对于时间概念,根据数据的颗粒度可以将发文时间精确到

年期,而引文时间和统计时间以年代为准。②根据实际应用情况细化。例如,对于来源期刊概念,为了讨论单期刊期间的相互关联,需要将期刊对象以“期”作为讨论单位。因此,根据颗粒度的不同,可以从来源期刊中派生出来源单刊(以期为单位)和来源种刊(以种类为单位)。又如,根据不同类型引文对象的被引次数统计具有现实的引文分析意义,可以对引用文献进行细化等。③根据概念模型可扩展的需要细化概念。类型概念是为了概括 CSSCI 中所有具有有限编码的分类准则而定义的,在现有的数据基础上派生出机构类型、文献类型、语种类型等子概念。随着分类准则的增加,可以从类型中继续派生子类,实现概念模型的可扩展。

基于学术概念的抽取和扩展,本文建立如图 3 所示的 CSSCI_Onto 三层概念层次结构。

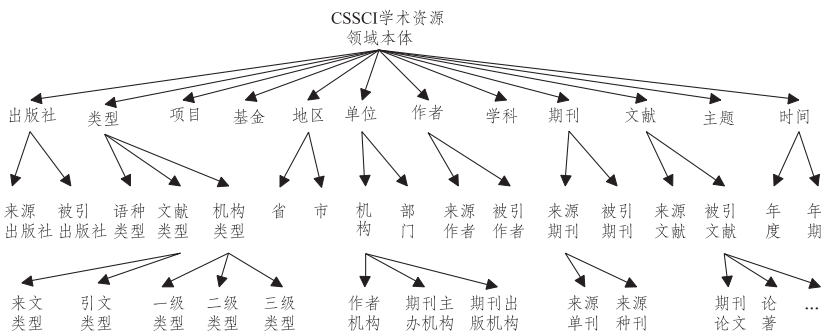


图 3 CSSCI 学术资源本体的三层概念模型

4 CSSCI_Onto 概念属性的设置

概念层次结构是本体骨架,而其血肉则是以属性形式存在的概念间关系(或约束)。结合来源数据,参考领域专家意见,可以分别为图 3 所示的各种概念设置属性,建立概念间关系,约束概念间行为,并最终生成概念模型,为实例属性值的设定和实例间关系的计算奠定基础。针对 CSSCI 概念的具体特征,笔者认为它们的属性可以分为两大类:

(1) 数值属性 (Datatype Property)

用于描述概念的自身状态和结构等信息的

属性,其取值仅与概念实例自身本质相关,其他对象的变化以及周围环境因素等不会对其产生影响,如“主题名称”、“文献篇名”等。

(2) 对象属性 (Object Property)

以某一概念的对象作为属性值,用于描述实例间关系,其取值随着对象间关系的变化或环境状态的影响而发生改变。例如主题的“来源文献”属性就是一个典型的对象属性,揭示了主题与来源文献概念之间的语义关联。

在 CSSCI_Onto 概念模型中,对象属性根据属性值来源可以分为:①同类对象属性。如主题概念的上下位、共现和交叉主题等属性的取值均为主题,用以描述同类概念之间的关联。

关于学科地图、知识地图等的绘制,就是对同类概念之间关联的语义描述;Astrova^[5]在从关系数据库中进行学习本体的研究中,也提出了通过对关系数据库元组的分析来获得概念间“继承”关系,指的也是建立同类概念间的关系。在CSSCI中存在着丰富的同类对象潜在关系需要挖掘。②异类对象属性。以其他类型对象作为概念属性值的属性,例如主题的来源种刊、学科、来源作者等属性分别以期刊、学科和作者等对象作为属性值,揭示的是主题对象和期刊、学科、作者等对象的相互关系。异类概念属性通常用于描述本体实例之间的非分类语义关系。

CSSCI_Onto 中学术概念属性设置的主要依据包括:①原始数据结构中对象之间的相互依赖。将原来关系数据库中的对象依赖直接移植到本体中,例如来源文献概念的文献作者、所属期刊、所属学科等属性直接来自原始元数据中的关系字段。②根据实际应用的需要。在CSSCI分析中,经常根据同类概念之间的关联,或探讨相互合作的程度,或向用户提示关联入口,或分析学科交叉的可能性和必然性等。例如研究作者间关联可以揭示作者的社会关系,构建作者知识地图等。此外,由于引文分析的需要,可以在概念中设置统计属性,例如发文类型的“年度发文量”、“年度引文量”等属性,揭示研究论文、综述、评论等文献类型的年度分布情况等。③考虑概念层次结构中的概念继承。父类的属性通过继承被所有子类所共享,因此,子概念的通用属性可以在父概念中设置,以减少属性冗余。例如被引期刊论文、被引论著等概念除了来源不同(前者具有“所属期刊”属性,而后者具有“出版社”属性),其他属性均一致,可以将两者相同的属性设置在其父概念被引文献上,通过继承它们可以获得这些属性。④领域专家的意见。构建本体概念模型的目的是为了指导领域的知识标注,因此,概念的抽取以及概念属性的设置都应该以领域知识为背景,以应用为前提,时刻接受领域专家的监督和意见。

基于上述依据,本文为概念模型中的各种概念设置了数值属性、同类对象属性和异类对象属性。以主题概念为例,笔者认为其属性包

括:主题名称、上位(hyperonymy)主题、下位(hyponymy)主题、共现(co-occurrence)主题、交叉(intersection)主题、来源文献、来源单刊、来源种刊、关联学科、来源文献类型、来源作者、部门、机构、机构类型、省、市、发文年度、引文年度、项目、基金、引证期刊文献、引证论著、引用期刊、引文出版社、引文语种、引文类型、引文作者等。限于篇幅,其他概念的属性不再一一列出。

概念属性的设置决定了整个本体的使用价值和知识体系,具有非常重要的意义。本文认为概念属性的定义并不是一蹴而就的,需要根据实际使用的情况以及属性值获取的可能性来进行不断的修正和完善,这也是一个循环往复和增量迭代的过程。

5 CSSCI_Onto 概念模型的描述

概念模型设计完成之后,可以使用标准语言(如OWL)对其进行形式化表达,方便用户重用、修改和完善其所表达的知识框架;也可以采用可视化方式(如星型图)对其进行图形化展示,便于领域知识框架的揭示。本文借助现有的本体概念模型描述工具Protégé^[6],实现对CSSCI_Onto元模型的编码和图形化描述。

5.1 概念模型的OWL编码描述

在Protégé中输入概念、属性和约束,可以自动生成相应的OWL代码,借此可以将CSSCI_Onto概念模型转化为OWL文档形式存储。Protégé提供了建立概念间继承关系的机制,对于描述概念间关系的属性,则需要分情况讨论:①数值属性的属性值通常为单值,可以使用3元组<概念,属性,属性值>来描述,例如<主题,主题名称,值>。②对象属性揭示概念间关系,其中处于核心地位的概念称为主体,属性概念称为客体,主客体关系有时是二元的,可以通过3元组<主体,属性,客体>描述,如<主题,下位主题,主题>(<竞争情报系统,下位主题,企业竞争情报系统>)。然而,主客体关系在更多的时候是多元(N-ary)的,例如主题概念的“交叉主题”属性,其属性值不仅需要指出具体的客

体,还需要设定主客体关联度,客体概念和主客体关联度一起构成属性值,此时就需要使用多元组描述主客体关系: < 主体,属性,客体,主客体关联度 > (如 < 语义网,交叉主题,RDF,0.12975 >)。使用多元组描述的对象属性又称为复合属性,在使用 OWL 编码时需要对其进行 N-ary 关系分解^[7]。

N-ary 关系分解的方法主要有 3 种:①引入属性类。当实例之间不仅存在关联,还需要明确描述关联的确定性(certainty)、强度(strength)等时,可以将属性值整体定义为一个属性类,各个子属性值则成为该属性类的属性。例如可将 < 主体,属性,客体,主客体关联度 > 分解为 3 个 3 元组 < 主体,属性,属性类 >、< 属性类,名称,名称 > 和 < 属性类,关联度,关联度值 >。②定义关系类。在主体不明确的情况下,把主客体看作一个整体(关系类),主客体都是关系类的

属性值。例如上述 4 元组可被分解为 < 关系类,属性,主体 >、< 关系类,属性,客体 > 和 < 关系类,属性,主客体关联度 >。③引入顺序类分解排序语义。当属性值是有秩序的对象集合时,可以在主客体之间引入若干顺序类(相互之间通过排序属性形成链结构),把 N-ary 顺序关联转化为了了一系列顺序类之间的 2-ary 关联。例如 < 来源文献(主体),来源作者(属性),来源作者 1(客体),来源作者 2(客体),... >,可以分解为 < 来源文献,第一作者(属性),第一作者类 >、< 第一作者类,作者名称(属性),名称 >、< 第一作者类,下一作者(属性),第二作者类 >、.....等等 2-ary 关联。

通过 N-ary 关系分解,可以将 CSSCI_Onto 概念模型中的所有关联转化为 3 元组形式,进而实现 OWL 编码。图 4 为采用 OWL 描述的 CSSCI_Onto 概念模型片段。

<pre> <owl:Class rdf:ID="cited_literature"> <rdfs:subClassOf> <owl:Class rdf:ID="literature"/> </rdfs:subClassOf> </owl:Class> <owl:ObjectProperty rdf:ID="article_citing"> <rdfs:range rdf:resource="#cited_literature"/> <owl:versionInfo rdf:datatype="http://www.w3.org/2001/XMLSchema#string" >1.0 </owl:versionInfo> <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string" >描述来源文献概念的被引文献的基本情况 </rdfs:comment> <rdfs:domain rdf:resource="#source_literature"/> <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">引证文献 </rdfs:label> </owl:ObjectProperty></pre>	<pre> <owl:Class rdf:ID="city"> <owl:disjointWith> <owl:Class rdf:ID="province"/> </owl:disjointWith> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:TransitiveProperty rdf:ID="part_of"/> </owl:onProperty> <owl:allValuesFrom> <owl:Class rdf:about="#province"/> </owl:allValuesFrom> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Class rdf:ID="area"/> </rdfs:subClassOf> </owl:Class></pre>
---	--

图 4 基于 OWL 的 CSSCI_Onto 概念模型片段

5.2 概念模型的可视化描述

基于 OWL 能够实现对本体概念模型的完

美描述,然而形成的文档类似计算机程序,非专业人员很难读懂,因此可以借助图形化工具将 OWL 文档可视化,使普通用户也能够清楚了解

本体的总体框架及其概念之间的语义关系,同时也为本体开发人员修改或完善本体概念模型提供可分析和可检查的条件。

Protégé 是一款基于插件的本体编辑工具,可以通过附加插件扩展其功能。目前,已经出现了多种用于 Protégé 的可视化插件,其中常用的有 OWLViz、Jamabalaya、OntoViz、TGViz、OntoSphere 3D 等。各种插件具有不同的表现能力,

表 1 列出了上述 5 种工具提供可视化功能的比较情况。从表中可以看出,TGViz 和 OntoSphere 3D 的可视化功能比较齐全,是 2D 和 3D 可视化工具的典型(图 5 为使用 TGViz 描述的呈星型状的 CSSCI_On to 概念模型);OWLViz 相对来说功能较少,但是也相对直观;Jamabalaya 和 TGViz 支持本体检索,这是目前本体可视化的发展趋势,便于快速定位。

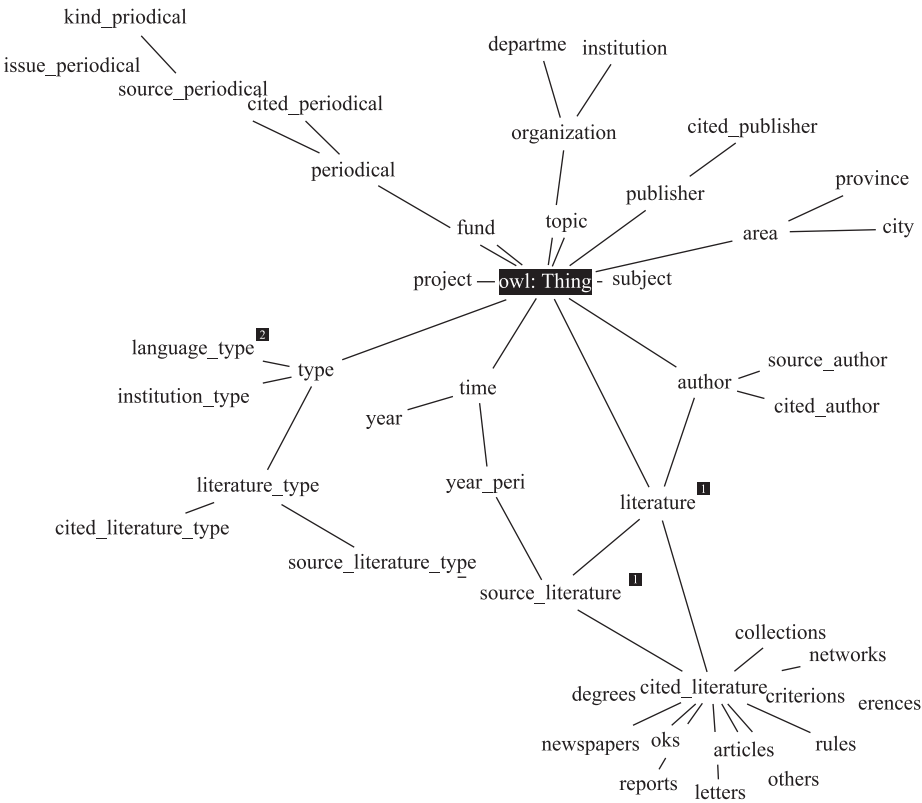


图 5 基于 TGViz 的 CSSCI_On to 概念模型

表 1 5 种 Protégé 可视化插件的功能比较

功 能 \ 插 件	OWLViz	Jamabalaya	Ontoviz	TGViz	OntoSphere 3D
层次结构树的可视化	√		√		√
双曲线(hyperbolic)树的可视化			√	√	
可视化的操作		√	√	√	√
图形化的选择		√	√	√	√
深度控制	√			√	
高亮显示	√	√	√	√	√

续表

功 能 \ 插 件	OWL Viz	Jamabalaya	Ontoviz	TGViz	OntoSphere 3D
类层次结构 (IS_A) 的可视化	√	√	√	√	√
实例的可视化			√	√	√
属性的可视化		√		√	√
本体图形导航		√		√	√
字符串检索		√		√	

除上述介绍的插件外,可以集成在 Protégé 中的本体可视化工具还有很多,例如 ezOwl、OntoBase、Cloud Views 等等,它们都以不同的形式试图增强本体的可理解性和可编辑性。

6 结束语

改变 CSCI 信息服务(包括信息检索和引文分析等服务)的现状,改善服务质量,关键在于调整其原有的知识结构。本文提出了使用本体面向对象的知识组织方式来重构 CSCI 学术资源的解决方案。根据本体构建的 6 步循环增量迭代法,笔者认为,CSCI_Onto 概念模型的建立包括概念层次结构的架构和概念属性的设置两个过程。本体概念模型确立了领域知识的组织结构,是本体实例化和应用的基础。

CSCI 是数字图书馆建设中的重要学术资源。类似的,本体面向对象的知识组织方式也可以应用于数字图书馆的其他对象的知识组织中,即根据本文提出的方法,采用本体来统一组织图书馆中的所有资源,包括图书、期刊、人员等,将各类资源置于一个巨大的本体概念模型中,从而实现对各种对象的语义描述,充分揭示各类对象之间的潜在关联。

CSCI_Onto 的特点在于其丰富的实例知识以及隐含的实例间关联。因此,本体概念模型的建立仅仅是 CSCI_Onto 本体构建的一部分(知识组织),应用的主体还在于基于概念模型实现 CSCI 大规模数据的语义标注(知识描述)和本体评价。这也是笔者需要进一步探讨的问题。

参考文献:

[1] A. Maedche, S. Staab. Ontology learning for the semantic web [J]. IEEE INTELLIGENT SYSTEMS, 2001, 16(2):72 - 79.

[2] T. R. Gruber. Towards principles for the design of ontologies used for knowledge sharing [J]. International Journal of Human Computer Studies, 1995, 43: 907 - 928.

[3] M. Uschold, M. King. Towards a methodology for building ontologies [C]. In Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI - 95, Montreal, Canada, 1995.

[4] B. Swartout, P. Ramesh, K. Knight, T. Russ. Toward distributed use of large-scale ontologies [C]. Symposium on Ontological Engineering of AAAI, Stanford (California), Mars, 1997.

[5] Astrova. Reverse engineering of relational database to ontologies [C]. Proc. of the ESWC 2004. Heidelberg: Springer-Verlag, 2004: 327 - 341.

[6] [2008-10-25]. <http://Protege.stanford.edu>.

[7] N. Noy, A. Rector. Defining N-ary relations on the semantic web [OL]. [2008-10-25]. <http://www.w3.org/TR/2006/NOTE-swpb-n-aryRelations-20060412/>.

王 昊 南京大学信息管理系讲师,博士。通讯地址:江苏省南京市汉口路 22 号南京大学信息管理系鼓楼校区。邮编 210093。

苏新宁 南京大学信息管理系教授,博士生导师。通讯地址同上。

(收稿日期:2008-10-28;修回日期:2008-12-07)