DOI:10.19789/j.1004-9398.2015.01.005

基于电子教材的学习资源本体建模

——以部分小学英语语法知识点为例

刘 妍 崔琳琳* 吴敏华

(首都师范大学信息工程学院,北京 100048)

摘 要

本文通过将领域本体的概念与核心技术引入到数字教材中的知识点与资源的检索中,设计一个运用推理来描述的部分语法知识点结构图,将其通过 Protégé 软件完成模型构建,形成基于 OWL 语言的语法知识点组织模型.

关键词: 领域本体; 数字教材; 学习资源; Protégé; OWL 语言

中图分类号: G434

1 引 言

电子教材是一类遵循学生阅读规律、利于组织学习活动、符合课程目标要求、按图书风格编排的电子书和电子读物[1]. 传统教学以纸质教材为主,教材的核心内容是知识框架的"浓缩";但是缺乏个性化的学习与指导,电子教材作为一种新型的教学媒介,资源丰富学习活动形式多样,教材的改革与更新也越来越得到人们的关注. 然而,随着大规模教育资源的建设,电子学习资源组织方面的问题日益凸显,学习资源、利用率低下、共享不足造成重复建设、出现资源"孤岛"等问题,这些问题严重影响了电子学习的发展,已经成为电子学习领域亟待解决的问题[2,3].

当前电子教学资源库得到了比较普遍的应用,但由于不同开发者使用不同系统工具和开发软件,相同系统间资源表示方式不同,不便于知识继承和共享,造成教学资源的重复建设,因此对于学习资源的组织与资源的描述成为日益关注的研究领域,学习资源初次利用率和重用率成为人们关注的焦点之一. 在众多的组织方法中,我们看到了各种问题和挑战. 例如面对多种形式的信息,缺乏统一的描述方式;被管理的学习资源是否具有能被理解的含义,是否能够在实现系统中被相对准确的识别,是否理解自身存储的资源内容;对学习者的需求得到把握学习者的需求,自动化、智能化推送资源;如何体现领域内知识资源的层级关系等.

随着网络技术在教育领域的发展与应用,使得学习者能够根据自己的兴趣特点、学习需求来获得丰富的资源,电子教材在满足丰富资源的同时,也出现了学习资源检索困难,结果冗余等[4]. 在英语学科上的本体应用研究也亟待解决,大量的英语学习信息和资源组要高质量的组织起来,更加自由的,个性化的针对某一学习专题或者某一课程领域,进行资源的选择和推送,更大程度上的实现共享和个性化学习,实现语义功能.

在英语数字教材系统建设的过程中,与外语教学出版社与北京市一线学校合,与一线教师交互交流,了解一线教师对学习资源的需求和利用情况. 我们发现,面对大量的资源,教师很难在短时间选取有效的、高质量的资源,分享给学生,在学生的自主学习中,当个人性化地推荐关于某些知识点的资源时,相关度与贴切度不够高. 因此,本研究提出资源组织的方式采用本体技术,希望更灵活地,有效地利用资源,高效学习.

经过大量的文献调研,在很多领域已经得到成功的应用,并且已经有成形的系统实现,集中在计算机学科中某个课程,例如计算机网络,C++语言程序设计,C 语言程序设计;法律领域;医学领域;网络

收稿日期: 2014-03-28

* 为通信作者

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

平台的资源管理[5]; 音乐学科领域等. 在教育领域的应用并不多见,具体某个学科的领域本体的构建更是较为匮乏,E-Learning 学习系统如何能够做到学习资源的尽可能共享和重用,对于学习知识的表示、组织和管理显得十分重要. 目前, 已有人将本体技术应用到 E-Learning 研究领域,但在学习知识的组织上基本是以章节为顺序来组织学习资源, 而以知识点中的相关概念来组织学习资源的研究甚少.

本体的本质特性使得基于本体的知识管理系统能够理解自身所管理的知识能够读懂使用者的需求[6],因而将基于本体的知识描述与组织引入到教育领域,是有着十分重要的意义,有助于电子教材为学习者提高个性化的学习服务,和资源展示应用. 电子教材要以学科课程标准或者当前学习主题的学习目标为焦点,根据要达成的认知目标(或情感动作等) 的不同维度,为学习者提供内容贴切难度恰当时长合适的学习资源. 为了保证英语学习资源的质量,本研究团队选择 2011 版英语课程标准中部分语法内容进行概念图和属性标注.

2　语义领域本体的构造

2.1　本体的定义

本体(ontology,本体论 Ontology) 这个术语来自于哲学,意指客观存在的一个系统的解释和说明,是客观现实的一个抽象本质. 随着人工智能的发展,本体被赋予了定义: 本体是一个为描述某个领域而按继承关系组织起来作为一个知识库的骨架的一系列术语和关系,是共享的概念模型的明确的形式化的规范说明[7]. 其中,共享(share) 指本体中反映的知识是其使用者共同认可的; 概念模型(Cerptualization) 指客观世界的现象的抽象模型,其表示的含义独立于具体的环境状态; 明确(Explicit) 指概念及它们之间的关系都被精确定义; 形式化(Formal) 指精确的数学描述,可被计算机理解与处理.

将本体应用到知识系统中,能够为系统提供一下好处[8]:

可重用性(re-usability) : 本体是领域内重要实体、属性、过程及其相互关系形式化描述的基础. 这种形式化描述可成为软件系统中可重用和共享的组件(Component) .

知识获取(knowledge acquisition) : 构造知识系统时,用已有的本体作为起点和基础来指导知识的获取,可以提高工作速度和可靠性.

可靠性(reliability) : 形式化的表达使得自动的一致性检查成为可能,从而提高了软件的可靠性.

规范性(specification) : 本体有助于确定信息系统(如知识库) 的需求和规范.

2.2　本体的建模元语

一个 OWL 本体中的是元素是与类(class) 、属性(property) 、类的实例(instance) 以及这些实例间的关系有关的. 本节给出应用这些元素所必需的语言成分.

概念(concepts) 或类(classes) : 指任何事务,如工作描述、功能、行为、策略和推理过程. 从语义上讲,它表示的是对象的集合,其定义一般采用框架结构,包括概念的名称,与其他概念之间的关系的集合,以及用自然语言对概念的描述.

关系(relations) : 领域中概念间的关系,形式上定义为 n 维笛常儿积的子集: R: $C_1 \times C_2 \times \cdots \times C_n$. 如子类关系(subclass-of) . 在语义上,关系对应于对象元组的集合.

函数(functions) : 一类特殊的关系. 该关系的前 n - 1 个元素可以唯一决定第 n 个元素. 形式化的定义为 F: C1 × C2 × … × Cn-1→C_n 如 mother-of 就是一个函数,mother-of(x,y) 表示 x 是 y 的母亲.

公理(axioms) : 代表永真断言,如概念乙属于概念甲的范围.

实例(instances) : 代表元素. 从语义上讲,实例表示的就是对象.

本体中基本的语义关系共有 4 种:

部分关系(part-of) : 表达概念之间部分与整体的关系.

所属关系(kind-of) : 表达概念之间的继承关系,类似于面向对象中的子类与父类之间的关系.

实例关系(instance-of) : 表达概念的实例与概念之间的关系,类似于面向对象中的对象和类之间的关系.

属性关系(attribute-of) : 表达某个概念是另一个概念的属性. 如“价格”是桌子的一个属性.

在实际建模过程中,概念之间的关系不限于上面列出的四种基本关系,可以根据领域的具体情况定义相应的关系.

2.3　OWL 语言作为知识表示语言

知识表示语言对表示能力,可理解性,可操作性,可扩充性的要求都很高[9,10], 并且要求能够针对不同用户群分层设计,语法简便,多语种支持等.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

Web 本体语言 OWL 是一种定义和实例化“Web 本体”的语言,具备较强的表示和推理能力. 后面文章中,我们将使用 Web 本体语言 OWL 作为知识表示语言.

根据 W3C 的推荐标准,OWL 在 web 上表达机器可理解内容的能力非常强且有效,OWL 语言遵循 XML 语法,语语法简单易用容易理解. ,目前针对于 XML 的开发维护和操作工具已经出现了很多. 而 OWL 遵循 XML 语法. 所以在 OWL 语言被指定为推荐标准之后,又出现了 Jena、OWLAPI、Protégé 等专门针对于 OWL 语言的创建与操作处理的工具. OWL 的可扩充性很好. 定义类、属性或者类的实例时,没有必要一次性地把它定义成功,因为完全可以借助于 OWL 的可扩充性在后续继续定义并完善. 如,在运用 OWL 表示高校教育资源的过程中,可以先声明一个副词类,转做别的工作,在以后需要时再为这个类添加需要的属性,如比较级等.

2.4 OWL 知识表示

在基于 OWL 的知识表示过程中,首先需要对知识资源进行面向对象的抽象,然后建立遵循 OWL 语法的抽象类,将这些类存放在一个 OWL 文档 DocType. Owl 中,然后再定义这些类的实例,也就是知识,放到另一个 OWL 文档 DocKnowledge. owl(或者根据需要存放到多个文档中). 由这些 OWL 档构成知识库. OWL 的知识表示过程如图 1 所示.

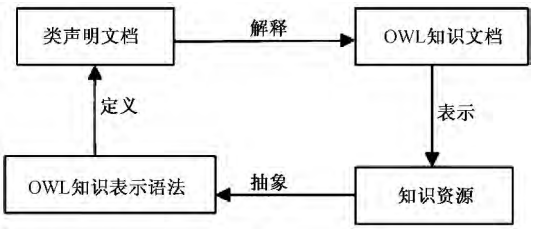


图 1 OWL 的知识表示过程

2.5 本体在课程教学领域的应用

如果把“本体结构化模型”应用到“课程教学领域知识”甚至课件,它所产生的突破,乃是把课程知识纳入到布卢姆的教育分类体系,尔后带来的知识点展示,不再是线性的,或者任意网状的,而是教育学意义上的领域结构化(这种结构化是经过基于布卢姆教育目标分类的领域设计、本体模型化、逻辑化的,超过教师手工加工)和非线性的.

除了教材固化的知识素材之外,扩展学习、自主学习需要的资源也可以这样结构化. 步骤如下:

- 建立信英语学习目标列表(相当于本体).
- 整理推送资源
- 形成英语教育的学习行为本体(相应“成就”为学习行为,见教学活动屏,以“实例”形式处理).
- 建立 OWL 模型.
- 形成领域知识本体. 由信本体持久化为数据库,按检索进行提取结构化信息,或展开可视化导航.

可深入研究: 形成教学流; 学习者建模; 本体的自主学习和调整.

3 基于课程体系的知识表示方法

3.1 部分小学英语语法的体系结构

领域本体实际上是一张机器可读的相关领域的术语(概念)及其关系的词表,它的作用是使人们对相关领域内的术语、规则达到共同认识. 要进行领域本体的构建首先要确定领域中的关键概念. 确定上位概念、较为显著的概念和常用概念,以建立核心概念集,进而确定本体中的类. 通过对“小学英语课程”教学内容的分析,我们应用 OWL 建立知识库,可以得出其知识点结构图. “小学英语课程”中知识点部分关系如图 2 所示,以“线性表”知识点为例表示了知识层次结构. “小学英语课程”课程其他各种概念及关系都采用类似方法表示. 图 2 中图标的含义可参见文献[11].

3.2 部分小学英语语法的知识表示

从技术路线角度来看,首先建立本体 Ontology,具体步骤如下.

- 确定所建立本体的领域 Domain.
- 建立这个领域内的类、子类、属性(主语、谓语、客体)、实例(个体).
- 确定类和子类的关系(is-a, is a type, type-of, is a part, part-of).
- 确定类与属性(Property)间的关系: 为每个属性建立定义域(Domain)、值域(Range)和关系约束.
- 为每个实例的属性赋值(丰富“叶节点”)建立实例(Instance),实例是上位类的子节
- 点,能够继承父类的所有约束(实际为不同类的相互制约)或属性,就是继承类的关系. 实例是实体,类的联系最后落实到实体上才有意义. 实例只有得到各种类的关系的制约,才是有意义的.

课程的知识概念经过组合、分类后,核心概念的关系利用本体编辑软件 protégé 构建本体模型,如图 3 所示,主要以下几个内容. (1) OWL Class: 在 OWL

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

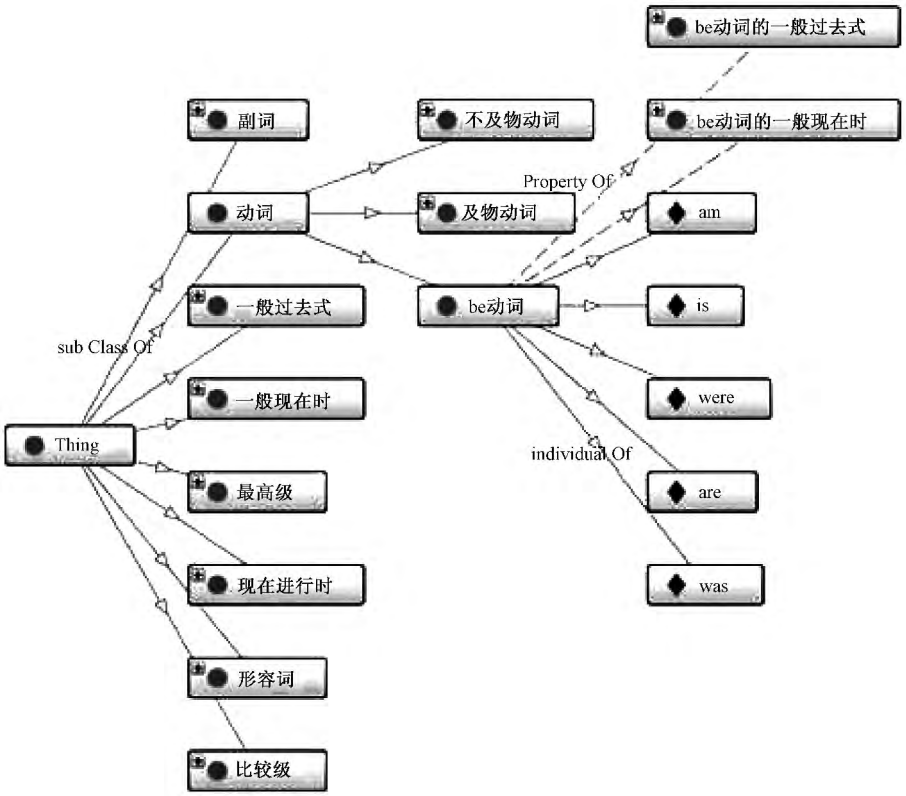


图 2　“小学英语课程”线性表知识点关系图

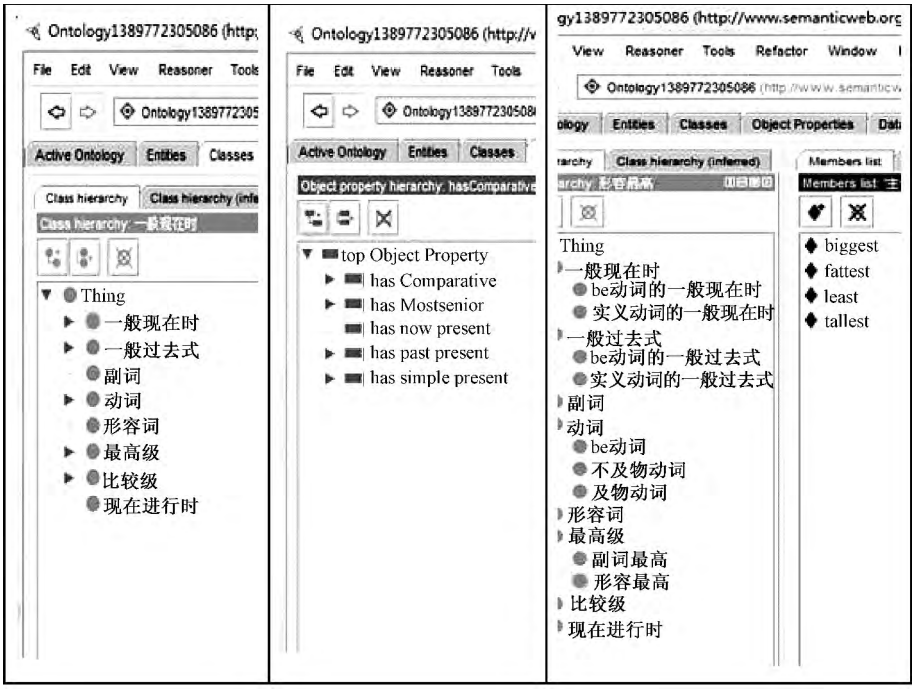


图 3　利用 protégé 软件实现“小学英语课程”知识点本体建模

Class 部件下，主要是组织在概念类之间的层次关系. 利用 OWL Class 下的 CLASS EDITOR 可以操作概念类之间更为细致的关系. (2) Properties: Properties 主要分为 3 个，Object，datatype，annotation. 通过 Object Properties 可以组织实例与实例之间的关系. 通过 datatypeProperties 可以定义实例的属性. (3) Individuals: 在 individuals 部件下，主要是定义概念类的实例. 通过 Properties 中属性来组织实例间的关系.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

图 3 中,我们定义了类 “副词”, “动词”和“形容词”;定义了类最高级子类形容词最高级下的 4 个实例;定义了“最高级”, “比较级”为对象类型属性,设定“副词”类具有“最高级”和“比较级”属性,就可以对“副词”类的个体进行最高级、比较级的描述.

2.3 数据库方式进行本体存储

按照本体存储介质的不同,本体的存储方法大致分为基于内存的存储方式、纯文本存储方式、基于数据库的存储方式和专门的管理工具方式 4 种,本文基于数据库进行存储. 该方式是将本体按照一定的策略组织在数据库中,利用现有的数据库系统对数据的操纵和管理能力来存取本体. 用关系数据库中的表和本体中的类相对应,用关系数据库中表的列与本体中类的属性相对应,对于类的多种数据类型属性可以通过数据库表中相关列的唯一性和主键设置来对应,本体类之间子类 - 超类的关系可以用数据表中主、外键的设置来对应. 他们之间数据类型的对应关系也可以用类似 xsd: integer 与数据库中 INT/INTEGER 对应关系一样一一对应起来. 利用开源工具包 jena,且配置好环境,把构建好的 OWL 本体文件持久化到数据库,并把自己的推理机制结合 SPAROL 查询语言对本体模型进行查询,存储过程如图 4 所示.

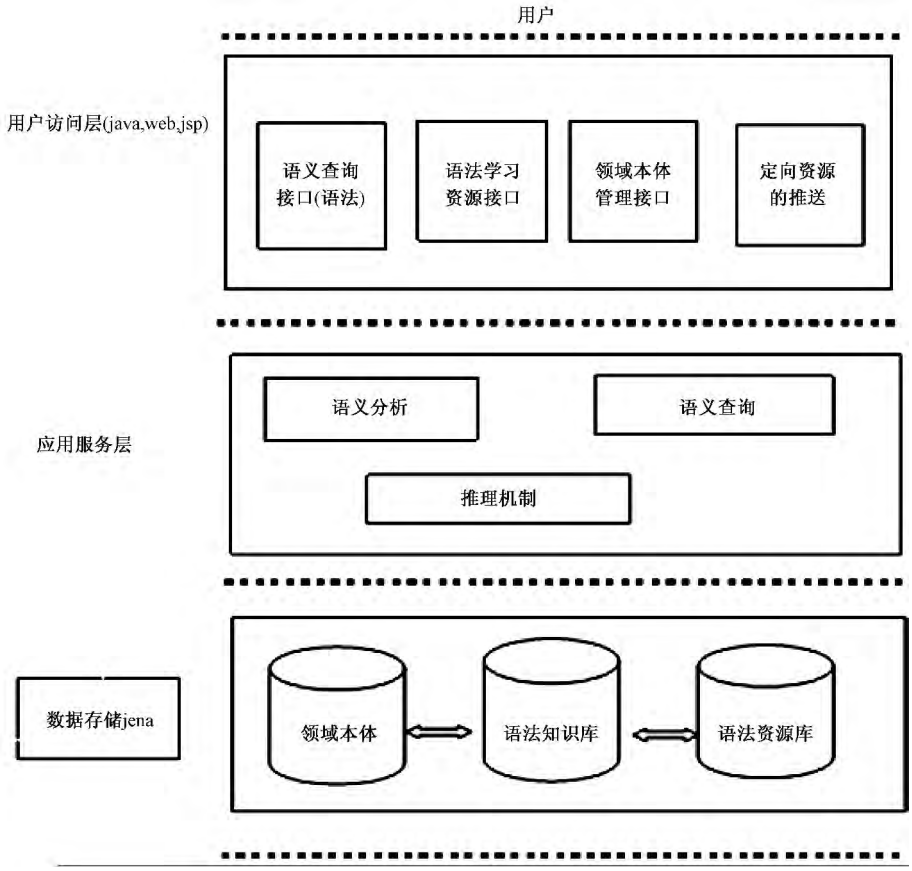


图 4 存储过程图

4 结 论

本体是一种用来描述概念以及概念之间关系的模型. 将本体应用于领域本体的建设,可以表示复杂的知识层次结构,而且本体具有允许知识共享、重用的特点,因而更便于知识的查询、验证和维护. 本研究即针对英语语法,之后有很好的推广性,可以把资源领域扩展,例如横向扩展语法变成作业、阅读等方面,纵向扩展可以是初中英语或者高中、大学英语,知识体系的拓展. 今后将领域本体应用与教育领域进行知识点的描述和组织,会提升资源的重用率和语义解析,也是之后将要研究的领域之一.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

参 考 文 献

[1] 陈桄,龚朝花,黄荣怀. 电子教材: 概念,功能与关键技术问题[J]. 开放教育研究, 2012, 18(2): 28 - 32.

[2] 刘泽文. 网络教育资源建设的现状与发展趋势[J]. 株洲师范高等专科学校学报,2005,10(5):70 - 72.

[3] 张沪寅,李鑫,陆春涛,等. LMOD: 一种基于本体描述的电子学习资源库模型[J]. 计算机应用研究,2011,28(4): 1360 - 1364.

[4] 李新成. 网络环境下的学习导航及其实现[J]. 中国远程教育, 2001 (8): 54 - 56.

[5] 谢明凤. 基于本体知识管理的远程个性化网络学习系统模型研究[J]. 中国电化教育. 2012(11) 47 - 53.

[6] 潘颖,欧启忠,肖耿毅. 面向语义的课程知识本体的构建[J]. 电化教育研究, 2007, 28(2): 19 - 21.

[7] N. Guarino and R. Poli. Formal ontology in conceptual analysis and knowledge representation. International Journal of Human and Computer Studies, 43(5/6):625 - 640, 1995.

[8] Uschold M; Gruninger M. ONTOLOGIES: Principles, Methods and Applications [J]. Knowledge Engineering Review, 1996,11(2).

[9] Deng YM, Tor SB, Britton GA. Abstracting and exploring functional design information for conceptual mechanieal product design. Engineering with Computers, 2000, 16(1):36 - 52.

[10] 年志刚,梁式,麻芳兰,等. 知识表示方法研究与应用. 计算机应用研究,2007,24(5):234 - 237.

[11] Matthew Horridge, Holger Knublauch et al. A practical guide to buildingOWL ontology using the protégé-OWL Plugin and CO-ODE tools. The University of Manchester, 2004.

E-learning Resource Modle Based on Domain Ontology: Parts of English Grammar Knowledge in Elementary School

Liu Yan　Cui Linlin*　Wu Minhua

(Information Engineering College, Capital Normal University, Beijing　100048)

Abstract

In this paper, through introducing the concept of domain ontology and the core technology to the knowledge point and resources retrieval of digital textbooks, we design a partial grammar knowledge structure chart which is described by using reasoning. We use Protégé to build model and form the grammar knowledge organization model based on the OWL.

Key words: Domain ontology, Digital textbooks, Learning resources, Protégé, OWL language.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net