

基于本体的教学知识库研究与应用

郁书好¹, 郭学俊²

(1. 皖西学院 计算机系, 安徽 六安 237012;

2. 河海大学 计算机及信息工程学院, 江苏 南京 210098)

摘要: 知识库是智能教学系统的基础。由于教学知识库的描述标准不统一, 知识表示方法也不同, 所以导致教学知识难以共享和互操作。将本体引入教学领域知识库建模过程, 建立概念共享模型, 提供概念语义空间, 不仅可以解决智能教学系统中的知识共享和互操作问题, 而且易于实现基于本体的语义检索系统, 从而大大提高系统的查全率和查准率。

关键词: 本体; 知识库; 语义检索

中图分类号: G434; TP18

文献标识码: A

文章编号: 1673-629X(2007)08-0161-04

Research and Application of Teaching KB Based on Ontology

YU Shu-hao¹, GUO Xue-jun²

(1. Dept. of Computer, West of Anhui University, Lu'an 237012, China;

2. College of Computer & Information Engineering, Hohai University, Nanjing 210098, China)

Abstract: The knowledge representation of knowledge base is one of the first important problems that must be settled in intellectualized system. Because the traditional intelligence teaching system resources lacks the uniform knowledge representation framework, all the learning resources are difficult to operate and share each other in various teaching system. By introducing ontology to the modeling process of teaching KB, can erect conception of share mode and provide conception of semantic space. This not only solves the problems of sharing of knowledge and mutual operation of ITS, but also realizes semantic search based on ontologies easily. This can raise the precision and recall of information search system.

Key words: ontology; knowledge base; semantic retrieval

0 引言

基于知识的教学系统即智能教学系统(Intelligent Tutoring System, ITS)是计算机辅助教育与人工智能学科的重要研究领域,也是教育信息建设的重要组成部分。经过近40年的发展,智能教学系统已成为计算机应用于教育领域的主要形式,并正在成为21世纪人类社会数字化教育发展的必然方向。然而目前现有的智能教学系统主要存在以下问题:

(1)语义互操作不强。高度的语义互操作对于Web上不同教育应用平台之间的交流是必需的,由于教学知识库的描述标准不一,在知识的结构、类型及属性上各自定义,并且知识的表示方式不同,所以导致知识难以得到交流和共享。

(2)检索效率不高。由于教学知识内容多而广,传统的基于关键字的搜索技术不能满足需要,目前实用化的教学信息检索系统对海量信息资源的揭示效率不高、深度有限,一方面会导致检索到大量无关的学习材料,另一方面也会丢失相关的重要学习材料。

如何解决好诸如信息组织、知识表示、机器理解与人机交互等问题,对于提高智能教学中教学资源利用的效率,是非常重要和迫切的。要解决以上这些实际问题,满足智能教学系统的特点和需求,发挥其应用能力,必须使用合适的本体来描述教学知识库,并构建一个基于本体知识库的体系结构^[1]。

1 知识库及本体技术分析

1.1 知识库

知识库是合理组织的关于某一特定领域的陈述型知识和过程型知识的集合。知识库和传统数据库的区别在于它不但包含了大量的简单事实,而且包含了规则和过程型知识。知识库的知识必须以某种一致化的

收稿日期:2006-11-16

基金项目:安徽省教育厅自然科学基金项目(2006KJ097B);皖西学院自然科学基金青年研究项目(WXZQ0504)

作者简介:郁书好(1976-),男,安徽六安人,硕士研究生,主要研究本体论与语义Web;郭学俊,副教授,主要研究数据库、信息系统。

结构存储和组织,以实现计算机自动知识处理和问题求解。这就是所谓的知识的形式化表示。知识的形式化表示方法主要有:基于逻辑的知识表示、基于关系的知识表示、基于规则的知识表示、基于模型的知识表示以及基于本体的知识表示等。基于本体的知识表示方法是近年来的研究热点之一。这种观点认为:任何复杂的知识都由最基本的概念构成,这些最基本的概念称为本体,本体是基本概念的详细说明。

1.2 本体

1.2.1 本体概念

起源于哲学领域的本体,在人工智能界最早给出其定义的是 Neches 等人,他们将本体定义为“给出构成相关领域词汇的基本术语和关系,以及利用这些术语和关系构成的规定这些词汇外延的规则的定义”^[2]。后来在信息系统、知识系统等领域,越来越多的人研究本体,并给出了许多不同的定义。其中最著名并被引用最广泛的定义是 Gruber 的“本体是概念化的明确的规范说明”^[3,4]。W. N. Borst 对该定义也进行了引申:“本体是共享的概念模型的形式化的规范说明”^[5]。Fensel 定义“本体是对一个特定领域中重要概念的共享的形式化的描述”^[6]。

尽管本体的定义方式多种多样,通过研究比较,可以看出他们对本体的认识基本是统一的,都把本体当作是领域内部不同主体之间进行交流的一种语义基础,即由本体提供一种明确定义的共识,本体提供的这种共识是为机器服务的。在计算机领域讨论本体,就要讨论本体究竟是如何表达共识的,即概念的形式化问题,这需要首先分析本体的描述语言。

1.2.2 本体描述语言和建模工具

本体描述语言主要是为领域模型编写清晰的、形式化的概念描述。随着 Web 的发展,出现了基于 Web 标准的本体描述语言,如 SHOE(Simple HTML Ontology Extension),XOL(XML-based Ontology-exchange Language),RDF,RDF-S,OIL,DAML,DAML+OIL,OWL。笔者着重分析 OWL。

OWL(Web Ontology Language)是 W3C 推荐的本体描述语言的标准,它是为了在 WWW 上发布和共享本体而提供的语义标记语言。OWL 是在 DAML+OIL 的基础上发展起来的,作为 RDF(S)的扩展,目的是提供更多的元语以支持更加丰富的语义表达,并更好地支持推理。OWL 提供大量的基于描述逻辑的语义原语来描述和构建各种本体,例如类型之间的不相交性(disjointness)、基数(cardinality)、等价性、属性特征(如对称性 symmetry),以及枚举类型(enumerated classes)等。OWL 相对 XML,RDF 和 RDF Schema 拥

有更多的机制来表达语义,从而 OWL 超越了 XML,RDF 和 RDF Schema 的仅仅能够表达网上机器可读的文档内容的能力。

到目前为止,已出现了许多本体建模工具。本体建模工具(环境)可极大简化本体建立的烦琐,比较常用的六个工具进行比较分析如表 1 所示。在文中将采用 protégé 作为本体的建模工具,设计教学领域本体。

表 1 本体建模工具比较

| | |
|------------|----------------------------------|
| Ontolingua | 基于网络的本体建立工具 |
| WebOnto | 基于网络,完全以面向图形显示来实现本体的建立 |
| Protégé | 基于 Windows,也是面向图形 |
| OntoSaurus | 基于网络,与 Ontolingua 相似,但使用 Loom 语言 |
| ODE | 基于 Windows,提供面向图形与面向文档的本体建立 |
| KAD22 | 面向图形和文档,并提供推理机 |

1.3 基于本体的知识库

从知识表示角度考虑,本体和知识库有类似之处,都是对一个具体或是抽象领域中包含的知识的定义、表示和组织。其主要区别是知识库包含了特定领域抽象或特定的知识,而本体是概念层次上对概念化的清楚描述,更注重概念层次上术语及术语间关系的表述。从功能上来讲,本体和数据库有些相似,但是本体比数据库表达的知识丰富得多。首先,定义本体的语言,在词法和语义上都比数据库表示的信息丰富得多;更重要的是本体提供一个领域严谨丰富的理论,而不单单是一个存放数据的结构。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。这种形式化的描述可成为软件系统中可重用和共享的组件。

知识库系统的实现涉及到两个关键问题:知识表示和知识推理。知识的表示形式是知识库系统首要解决的问题,它应当用计算机可以“理解”的方式对知识进行表示,同时以一种人类能够理解的方式将处理结果告知用户。本体可以为知识的构建提供一个基本的结构,本体可以将领域知识的概念和相互之间的关系进行较为精确的定义,从而解决机器理解问题。在这样一系统概念的支持下知识检索、知识共享的效率将大大提高,真正意义上的知识重用和共享也能成为现实。

ITS 中领域知识库的建立,其实质是研究领域知识中的基本概念及其关系。将本体引入领域知识库的建模,可以提供共享概念模型,建立概念的语义空间,这样不仅会促进领域知识库的实体属性和术语表达的规范化,解决知识共享和重用,增强领域知识库的智能推理能力,而且将比传统数据库所使用的语言在语义、语法上更加丰富;同时在基于领域本体知识库的系统中,易于实现基于描述逻辑 DL(Description Logics)的推理查询功能。

2 基于本体的知识检索系统设计

2.1 本体知识库的设计

为研究基于本体的 ITS 设计,文中采用 protégé 建模工具设计了《数据结构》(简称 DS)课程的知识库,并用 OWL 描述,如图 1 所示。

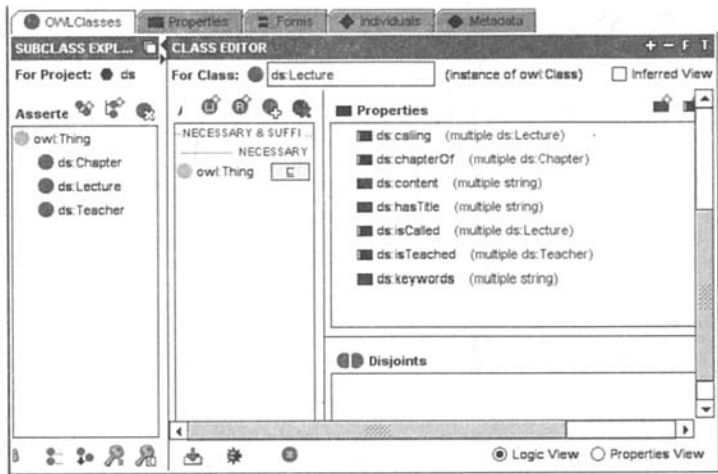


图1 实验用 DS 本体类

其中定义了三个资源对象:章节(Chapter)、讲义(Lecture)和教师(Teacher)。在此基础上,还定义了四个对象(ObjectProperty):calling, isCalled, isTeached, chapterOf。对象属性 calling 和 isCalled 是 inverse 关系,它是用来描述 Lecture 与 Lecture 之间的关系的,如线性表的删除操作与查找操作,因为要先查找到元素的位置才能删除该元素,所以这样将查找元素操作作为组件,可以为删除元素直接调用。而且也可以在相关概念介绍时用此方法,如介绍数据结构:是相互之间一种或多种特定关系的数据元素的集合。这里就用到了“数据元素”概念,而数据元素是数据的基本单位,又用到了数据的概念,等等。通过这样的调用关系,可以在检索时让系统显示出相关的内容,以便学生方便查找以前或以后用到的知识,起到举一反三之功能。

2.2 系统设计

为了能具有更好的系统可扩展性,设计的实验系统采用三层结构,即:表示层、业务逻辑层和数据层。基于本体的知识检索系统结构图如图 2 所示。

(1) 表示层。

负责实现用户和系统的交互,用户通过浏览器访问实验系统,系统的用户界面负责接收查询请求,并负责将服务端的检索结果显示给用户。由于 OWL 遵循 RDF/XML 语法,因此本层的开发主要采用 JSP + JavaBean 技术,JSP 基于 XML 语法,实现显示和编程逻辑的分离。

(2) 业务逻辑层。

本层主要负责实现知识检索系统的主要应用逻辑,它由检索引擎、语义分析组件和推理引擎三个部分构成。语义分析组件在获取检索引擎提交来的用户查询请示后,访问 DS 本体库,首先将用户的概念层次转

化为本体概念层次;接着根据 Jena OWL 推理引擎进行概念语义分析推理,对查询信息进行语义闭包求解,从而使用户的意图和应用程序理解的语义达到一致;最后检索引擎从数据层获取数据,并将检索结果返回给表示层。本层开发采用了由 HP 实验室开发提供的 Jena 2.2API 接口方法。

(3) 数据层。

数据层负责 DS 本体知识库的存储和操作服务,一方面可以创建并提炼出结构化的基于 OWL 描述的知识本体,通过数据获取组件存储在 DS 本体库中;另一方面可以由数据获取组件,将网上的相关资源通过语义标注后存储在本体库中,从而实现知识的共享和重用;本体管理组件负责对 DS 本体知识库的管理、更新和维护。

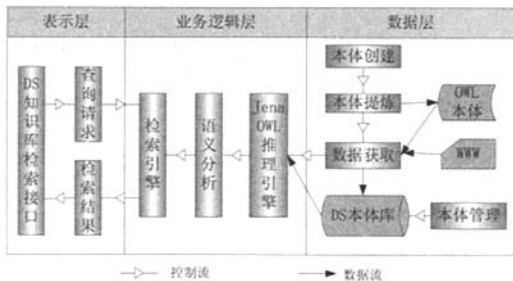


图2 基于本体的知识检索系统结构图

2.3 系统实现及分析

实验用 DS 本体知识库的属性建立了知识点的关联,使知识点之间的联系构成了网状结构。但应有个约定:知识点的产生顺序(从祖先到孩子)是树状的,中间无环,即使一个知识点同时产生于多个知识点,也只能从中选一个最有代表性的。通过可继承的子类和超类的定义,将本体知识库中的知识细分成树型的知识层次结构。避免了推理机的多重继承,也有利于知识点的 URI 的表示,因为规定了知识点 URI 为知识点树中从树根到该知识点的链,这样便于从 URI 中定位知识,也方便知识的查询和推理。然后用谓词表达知识点之间的关联,如前面的 calling(A,B)表示知识 B 是 A 的前驱,calling 是非自反、反对称和传递的。所有这些都反映在知识本体中,这是推理机的推理依据之一。

如果检索 GetElement(查找元素),与其相关联的前驱有 Data(数据)、DataElement(数据元素)、DataStructure(数据结构)和 LinearList(线性表),后继有 ListDelete(线性表删除)。检索 GetElement 知识点后得到的结果如图 3 所示。

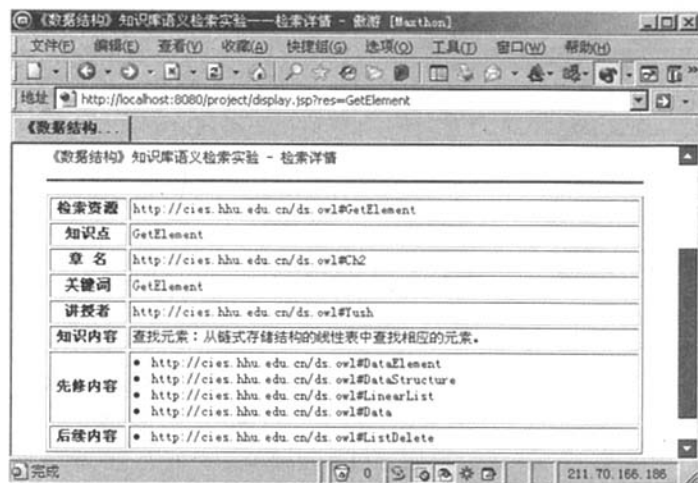


图 3 查询 GetElement 得到的检索详情

基于本体的语义检索与传统检索方式比较,语义检索是在知识(概念)层面上进行检索,刻画了概念之间的内在联系,可以发掘出那些隐含的、不明确的信息和概念,在查全查准方面可以更好地满足用户要求。如查询某章的所有内容时,与传统的基于关键字是不同的,它是通过概念之间的包含(类与子类、类与实例等)关系来进行查找的,可以有效地控制检索的范围,从而不仅可以提高查全查准率,还可以大大地提高系统的检索速度。

3 小 结

知识库是 ITS 的基础,知识库的知识表示问题是所有人工智能和智能化系统必须首先解决的问题之一。文中将本体与教学知识库建模结合起来,探讨知识共享和互操作问题,并研究了基于本体的语义检索技术。其中还有很多具体的工作要做,如本体的正确性检查,在大规模压力测试下,检索系统的性能评估等,将作为进一步工作的方向。

参考文献:

- [1] Mizoguchi R, Jacqueline. Using Ontological Engineering to Overcome Common AI - ED Problems[J]. International Journal of Artificial Intelligence in Education, 2000, 11: 107 - 121.
- [2] 邓志鸿,唐世渭. Ontology 研究综述[J]. 北京大学学报:自然科学版, 2002, 38(5): 730 - 738.
- [3] Gruber T R. A translation approach to portable ontologies[J]. Knowledge Acquisition, 1993, 5(2): 199 - 220.
- [4] Gruber T R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing[J]. International Journal of Human - computer Studies, 1995, 43: 907 - 928.
- [5] Borst W N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse [D]. Enschede: University of Twente, 1997.
- [6] Fensel D. The semantic web and its languages[J]. IEEE Computer Society, 2000, 15(6): 67 - 73.
- [7] Norrköping Visualization and Interaction Studio, Linköping University, 2000.
- [8] Yang Jing, Patro A, Huang Shiping, et al. Value and Relation Display for Interactive Exploration of High Dimensional Datasets[C]//Rundensteiner E A. Information Visualization, 2004. INFOVIS 2004. IEEE Symposium. [s. l.]: [s. n.], 2004: 73 - 80.
- [9] Wickelmaier F. Sound Quality Research Unit[C]//An Introduction to MDS, IEEE Symposium. Denmark: Aalborg University, 2003: 57 - 65.
- [10] 刘 勤,周晓峰,周洞汝. 数据可视化的研究与发展[J]. 计算机工程, 2002(23): 61 - 63.
- [11] 冯玉才,刘 嘉. 大量空间数据可视化的算法[J]. 计算机工程, 2003(29): 79 - 81.

(上接第 7 页)

利用 VaR 技术丰富的交互表达与变换工具,将各维的具体数据值采用像元进行表达,挖掘多维数据中的隐含信息与异常点,表现多维数据中各维的相关性,具有深入研究意义与价值。

参考文献:

- [1] Keim D A, Kriegel H - P. Visdb: Database exploration using multidimensional visualization[J]. IEEE Computer Graphics & Applications, 1994, 14(5): 40 - 49.
- [2] Keim D A. Designing pixel - oriented visualization techniques: Theory and applications[J]. IEEE Transactions on Visualization and Computer Graphics, 2000, 6(1): 1 - 20.
- [3] Feldt N, Pettersson H, Johansson J, et al. Tailor - made Exploratory Visualization for Statistics[R]. Sweden: NVIS -