

基于 Ontology 及 Semantic Web 技术的教育资源库应用框架

陈伯栋, 贾积有, 王爱华

(北京大学 教育学院 教育技术系, 北京, 100871)

摘要: Ontology 和 Semantic Web 技术是人工智能领域的重要研究对象, 并在众多领域有着越来越重要的应用。随着教育资源库建设的持续升温, 缺乏共享和重用机制的问题逐渐凸现, 造成了大范围的资源重复建设, 同时也产生了资源获取壁垒, 不利于资源的使用。本文提出了一种基于 Ontology 和 Semantic Web 技术的教育资源库解决方案的框架, 详细介绍了该框架的核心组成部分: 中间层、网络服务层和元数据表示, 以及具体实现技术, 以使分散的局域教育资源库相互整合, 从而达到可获取、可共享和可重用的目的。

关键词: 语义网; 本体; 教育资源库; 学习对象; 元数据

中图分类号: G434 **文献标识码:** A

一、前言

随着网络技术的进一步发展以及新的应用需求的不断涌现, 语义网 (Semantic Web) 逐渐成为 WWW 领域的研究热点以及智能化网络服务和应用开发中的关键技术之一。^[1]语义网能够在机器理解的基础上提供更加丰富的应用功能, 实现语义层面的运算。它赋予网络信息和知识以更明确、更完备的语义, 使计算机能理解网络信息和知识, 实现网络数据处理和网络服务的智能化。语义网在网络服务、基于代理的分布式计算、基于语义的搜索引擎、基于语义的数字图书馆等领域有着越来越重要的应用。同时, 随着应用研究的不断深入, 作为语义网研究重点之一的本体 (Ontology) 技术已经趋于成熟, 语义的元数据描述方法也形成了标准, 语义网的发展开始进入实际应用阶段, 在数字图书馆和数字博物馆等应用中, 基于语义的信息表示和搜索都有了初步探索和应用。

教育资源库建设是教育信息化的重要环节, 为了全面推动新课程改革和国家教育资源体系的建设, 全国教育科学规划领导小组、中国教育技术协会设立了“国家级教育资源库建设与应用研究项目 (NERRP)”作为新课程改革的配套研究课题和重点建设项目。^[2]在应用需求和相关政策的推动下, 各层次的教育资源库建设迅速展开。尽管一些学者指出, 在教学资源库持续发展的问题上管理因素的重要性远远超过技术因素, 但目前教育资源库面临的技术问题同样令人担忧。目前资源库建设各自为战、缺乏共享机制的问题十分突出, 制约着资源库发挥其应

有功效。国外的资源库建设经验表明^{[3][4]}, 有效的学习对象 (Learning Object) 共享机制、国家范围内资源服务器的互连与互操作、技术标准与规范的开发和推行等是决定教育资源库建设成效的关键环节, 国内外众多学者也一直在寻求有效的方法来实现这些目标。

本文针对目前教育资源库建设中存在的问题及应用需要, 把语义网中的关键技术引入到教育资源库建设中, 提出了初步的基于语义网技术的教育资源库建设框架, 从基于语义的元数据描述、教育资源库本体模型、资源库间的模式匹配等几个教育资源库建设中的重要方面入手, 探讨了该框架的核心理念和关键技术环节。

二、研究目标

(一) 基于语义的元数据描述

目前国际上已经有了几套应用较为广泛的学习对象元数据描述规范, 如 IEEE LOM、Dublin Core。在我国, 全国信息技术标准化技术委员会也根据本地化需要推出了我国的元数据描述规范。但是, 在技术实现上, 当前教育资源的元数据描述通常都依赖于 XML 描述, 描述方法停留在基本信息层面, 还没有达到语义层面, 因此无法进行语义上的操作。

W3C 发布了 RDF (Rich Description Framework) 语言来对知识进行语义编码, 借助于 XML 语法, RDF 通过一系列的三元组 (Triplet) 构建应用所需的语义网络。本研究的目的之一就是总结出一套合适的学习对象的 RDF 描述方法。

(二) 教育资源本体模型

在语义网中, 本体 (Ontology) 是网络应用和智

能软件代理文档语义的重要表述方式,一个本体定义了用于描述和表示一个领域知识的术语,包含了计算机可理解的该领域内的基本概念和概念间的相互关系。^[9]本体能被用于提供面向机器的语义,实现更高度的自动化和支持组织和教学法层面的支持。^[6]

为了让本体实现网络环境中的语义整合,我们需要对网络本体语言进行标准化。2004年2月,W3C 标准化了 OWL 语言,作为网络资源的语义描述语言。本研究希望通过对教育资源领域的统一的本体定义,为教育资源在语义层面的共享提供可能,以实现不同资源库之间的互操作。

(三) 基于语义的搜索

目前典型的搜索引擎都基于关键字,它们的查准率和查全率往往受到多义词和同义词的影响。通过此类搜索引擎,用户可能搜索不到任何数据,或搜到的结果相关度低,或者由于术语集(Terminology)的不同造成使用关键词无法获得相应结果。

语义网使用本体来表达网络资源的语义,基于语义的搜索都是在一定的本体环境中执行的,能够消除多义词或同义词的影响,从而提高搜索结果的查准率和查全率。

三、系统框架

为了实现教育资源技术上的可获取和可重用,实现分散的教育资源库间的互操作,实现资源的基于语义的搜索,优化资源生命周期管理机制,在参考其他应用实例的基础上,本文设计了一个教育资源的系统框架,以提供整合的、全局化的概念和本体表述机制,和对资源的基于语义的表述方法,整个框架的结构如右侧图1所示,本框架分为三个部分。

表现层:提供系统的用户接口,能实现通信协议和程序语言的互操作,同时为不同的用户群体(包括学校、科研机构和普通个人用户)提供不同的视图和管理机制。

中间层(或称语义网层):展示在全局数据和局部数据之间建立的通用的教育资源的概念模式。接收用户请求,参照本体模型进行语义转译,通过查询元数据库定位被请求的资源所在的局域资源库,之后把请求转发给局域资源库。当接到返回的资源时,再根据请求的内容整合请求结果,发送给用户。

数据层:包括分布式的网络服务和分散部署的网络资源库,两者之间通过一定的模式匹配原则实现映射和同步。

下面将就框架的三个最核心的部分展开讨论。

(一) 中间层(语义网层, Semantic Web Layer)

中间层是系统的关键部分,由三个基本模块构

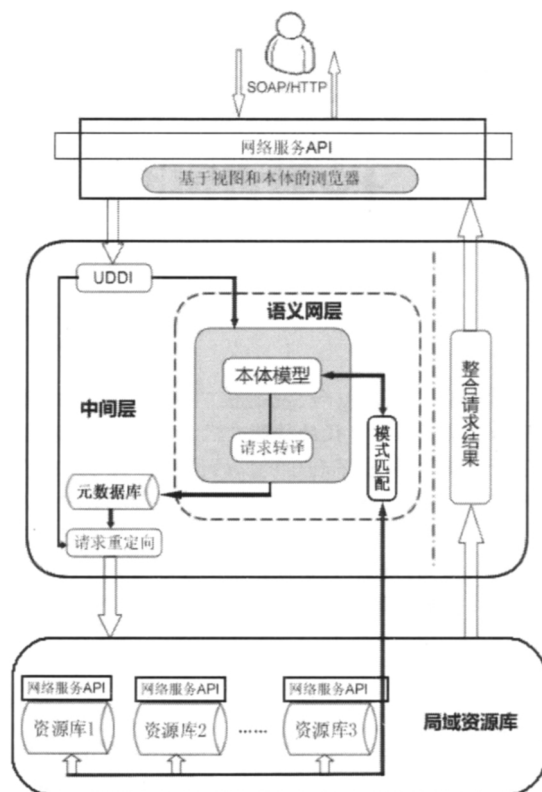


图1 教育资源库应用框架

成:本体模型模块、模式匹配模块和用户请求处理模块,图2说明了一个用户请求是如何经由这些模块得到处理的。

当用户发出基于本体模型的一个请求时,系统

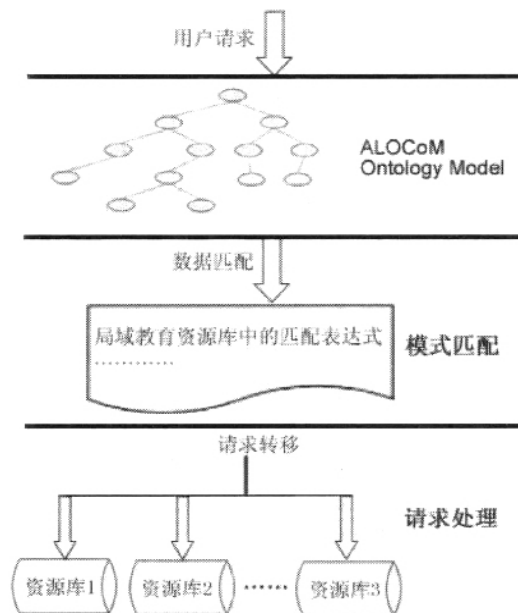


图2 用户请求在语义层的数据流动

将根据存储在模式匹配模块的模式匹配规则,请求处理模块把用户请求进行转译、重定向,最后,处理后的

用户请求在分散部署的资源库中执行并返回结果。

1. 本体及学习对象领域的参照模型

“学习对象本体”扮演的是“语义粘着剂(Semantic Glue)”的作用,来粘合不同的教育资源库的学习对象,以达到跨资源库的教育资源共享,去除资源库的应用壁垒。

ALOCoM 本体是一个通用的内容模型,它定义了一个针对内容对象和内容对象组件的框架。它通过定义表示不同种类的学习对象的组件及其结构的概念,使得可以把单个的学习对象析构成若干部分,继而可以将这些部分组装成新的学习对象。因此,我们可以利用 ALOCoM 本体框架作为我们教育资源库建设的本体模型的基础,在此基础上进行必要的扩充,形成完整的本体框架。

关于如何确定整个应用框架的本体模型,可以参考图 3 所示的模式和方法^[7]。

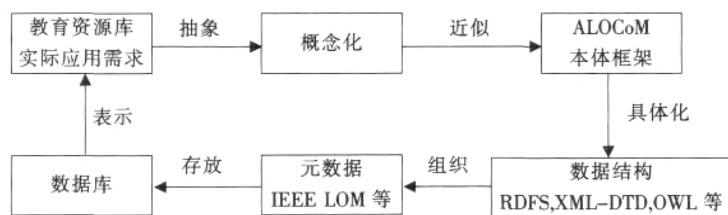


图 3 资源库本体模型的建立方法

要建立资源库本体模型,首先要从教育资源库应用需求出发抽象出问题的概念模型,然后依照 ALOCoM 本体框架对概念模型进行近似,再根据相关的技术规范(如 RDFS, XML-DTD, OWL)完成概念模型在数据结构层次的代表,最后参照成熟的标准组织成元数据存储于数据库中。

2. 模式匹配(Schemata Mapping)

在语义世界里,匹配(Mapping)指的是一个语义集合里的元素和另一个集合里的元素建立映射关系的过程。在教育资源库建设中,匹配更确切地是指在分散的局部资源库的本体或模式与中心本体模型之间建立起映射关系。当然,如果局部资源库本来就使用了相同的本体模型,那么在整合分散的资源库时就可以跳过模式匹配的步骤了。但是,当局部资源库设计和开发的本体都只反映了对领域各自的观点,或者使用了不同的描述语言和格式时,本体的不匹配是不可避免的。那么,跨本体的资源共享就常常被概念和关系的语义多样性所阻碍。

目前,分散的局部资源库已经建立起来了,再去

遵照新设计的本体重新建设是不可行的。但另一方面,在资源库间统一认识又显得十分必要。因此,我们可以从已有的资源库中提取出各自的参照模型建立起全局的本体,然后提供从全局到局部的映射。图 4 描述了怎样根据不同的教育资源库中的本体组建全局本体。

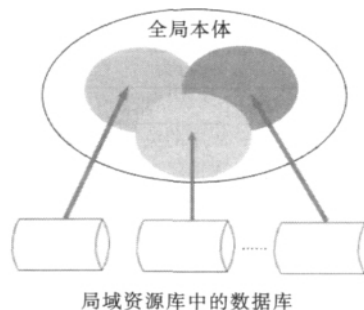


图 4 中心资源库与局部资源库间的本体匹配

在加拿大的 EduSource 项目中存在类似的可参考的应用实例,该项目采用 IMS 的数字资源库互操作(IMS Digital Repositories Interoperability)规范,实现了采用不同通信协议,不同元数据框架、规范和标准的资源库的搜索和资源获取。

3. 用户请求处理

在本文设计的教育资源库中,用户通过自然语句或普通的限定字段的组合搜索方式进行资源搜索。学习对象属于不同的局部资源库,它们在地域上相互分离。不同的资源库通常建立在不同的计算机平台、不同的数据格式和不同的编程语言上,它们提供各自的资源查询和获取机制。因此,当发生用户请求时,中央服务器需要通过一定的机制定位能提供服务的局部资源库,然后将请求转发给分散部署的局部服务器。在具体的系统设计中,我们可以参照一些数学化的方法^[8]。

(二) 网络服务层(Web Services Layer)

在资源库建设方式多种多样的情况下,为了能在资源库之间共享资源、有效整合分散资源库的教学资源,在与用户交互的表现层笔者选取 Web Services 技术来整合分散的资源库。Web Services 是在现有的各种异构平台的基础上构建一个通用的与平台无关、语言无关的技术层,各种不同平台之上的应用就依靠这个技术层来实施彼此间的连接和集成。^[9]

在整个资源库中,每一个局部资源库作为服务的提供者,只需要将自己提供的服务和已有功能发

参见 <http://ariadne.cskuleuven.be/alogom/>

布到全局系统的 UDDI (Universal Description Discovery and Integration)上,由 UDDI 来负责服务的注册和发送即可。所有的信息都通过 SOAP(Simple Object Access Protocol)在整个系统中发送。下表描述了一个用户通过一个浏览器搜索和访问资源的基本过程。

用户访问资源的过程

步骤	行为
1	用户通过浏览器发送一个搜索请求。
2.1	如果搜索是针对局部资源库的关键词或内容搜索,UDDI 接受请求并把在服务注册数据库中的搜索结果发送给命令转发模块来定位到能提供该服务的局部资源库。
2.2	如果搜索是基于概念的,它将被传递给语义层(中间层),在语义层中完成本体映射和请求的公式化转化,之后发送给命令转发模块来定位到能提供该服务的局部资源库。
3	每一个被定位到的局部资源库执行被调用到的 Web Services,然后把结果返回到结果重组模块。
4	结果重组模块重组所有从局部资源库返回的请求结果,然后发送给用户。

(三)元数据表示

要让资源库能提供基于语义的搜索和获取功能,学习对象的元数据描述应建立在语义层次上,W3C 发布的 RDF 描述语言可以帮助建立基于语义的元数据描述。RDF 采用的是语义的方法,而非 XML 的单纯的结构化方式。现在我们已经能够找到 RDF 编辑器来创建 RDF 三元组,因此,为学习对象构建语义元数据是可行的。

在本文的资源库的元数据设计上,XML 被用来以机器可理解的方式描述一个学习对象的结构,也用于描述学习对象的元数据。RDF 的引入可以让学习对象与元数据规范和其他相关信息以一种更灵活的方式建立联系,从而有利于学习对象的发现和交换^[10]。Ontology 规定了领域内用来标记学习对象的概念和术语的规范,通过采用共同的本体,不同的资源库系统对于特定的学习对象可以达成本义上的共识。

但是,要在所有学习资源上建立一套 RDF 编码的元数据,还需要 RDF Schema。在实际应用中,IMS 为 LOM 提供了一个 XML 绑定和 RDF 绑定。但是要建立或找到一个合适的 RDF Schema 却不容易,有待进一步探索。

此外,要在实际应用中体现语义元数据的意义和优势,需要基于语义的查询语言的支持。RDQL 语言可以被用来构造复杂的语义查询来准确地获取用户需要的学习对象的信息,已经在很多 RDF 系统中被用来有效地提取信息。图 5 给出了一条 RDQL 查询语句,用来查询用 "Castilian"写的"文学"类的"中等难度"的课程^[11]:

```
SELECT ?a,
WHERE (?a, <rdf:type>, learn-ont:Course),
      (?a, <lam-edu:difficulty>, lam-edu:MediumDifficulty),
      (?a, <dc:subject>, "LITERATURE"),
      (?a, <lam-edu:language>, learn-ont-idiom:Castilian),
```

图 5 RDQL 查询语句示例

四、技术实现

系统的技术实现基于 Web Services 的理念,为了提高用户客户端的易操作性,系统采用 B/S 模式。考虑到分散的资源库可能基于不同的平台系统,因此采用具备跨平台特性的基于 Java 的开发框架。系统的技术框架如图 6 所示。

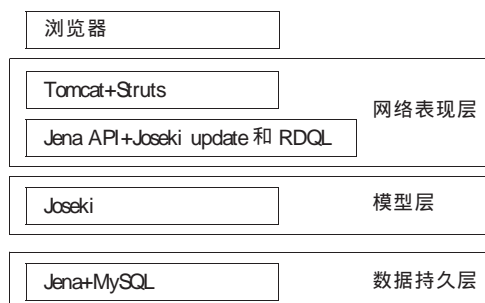


图 6 系统技术实现框架

语义层上,中心服务器的本体模型、映射规则、查询公式化规则等信息通过 Jena 数据持久化框架存储在 MySQL 数据库中。Jena 是用于开发语义网应用的 Java 框架,在查询语言方面采用 RDQL 查询语言,它为 RDF、RDFS、OWL 和 SPARQL 提供了一个程序环境,还包含了一个基于规则的推理引擎。

在与分散的资源库的映射上,我们采用 Joseki RDF server。

在整体框架上,我们采用 Struts 框架来实现 Model-View-Controller 的网络逻辑分离。

五、结论

本文从教育资源库开发的现状与共享需求出发,讨论了将语义网技术引入到教育资源库建设中的框架和关键问题。教育信息化的建设经历了“无路”“有路无车”“有车无货”的基本发展阶段,随着各级教育资源库项目的启动与运作,我们的“货”逐步丰富了起来,而面临的问题随之演变为资源的共享和重用机制的匮乏上。本文提出的基于语义网技术的应用框架采用 Web Services 的思想和框架,配合 Ontology 和 Semantic Web 技术,实现教育资源的有效共享和重用。框架的核心是语义层的设计,其中包含了本体模型的设计、匹配模式和请求处理过程的设计。技术实现方面采用 XML、RDF、RDFS、OWL、

基于学校信息化管理的智能校园卡的防伪设计与实现

朱 鹏¹, 江林升², 冯亦¹

(1. 南京师范大学, 江苏 南京 210097; 2. 南京森林公安高等专科学校, 江苏 南京 210046)

摘要: 信息技术的广泛应用, 为学校管理的信息化提供了新方法和新途径。智能校园卡作为学校信息化管理的重要工具之一, 其防伪的重要性也凸现出来。本文提出了一种基于数字水印与数字签名技术的智能校园卡的防伪技术, 以满足现代智能校园卡的防伪要求。

关键词: 信息化; 智能校园卡; 数字水印; 防伪

中图分类号: G434 **文献标识码:** A

随着信息技术在教学中的应用和渗透, 教育迅速走向网络化和数字化, 同时也使学校管理走上了信息化的道路, 催生了一系列的变革。学校管理信息化的最终目的是提高教育教学工作效率, 有效地利用信息技术来改变传统的学校教学组织形式和管理方式, 综合利用网络、数据库、多媒体等先进信息技术, 将教学、管理工作的信息数据数字化、标准化、集成化, 对需要传递的技术数据有效地实现共享, 提高信息资源的再利用性。现代智能型校园卡正是在学校信息化管理大力推行的过程中产生的。

一、智能校园卡的防伪需求

随着各个学校管理信息化的不断发展, 校园卡

的功能也越来越完善, 在传统校园卡的基础上, 现代校园卡集成了多项功能, 成为名副其实的智能卡。这种智能校园卡适应了现代学校管理信息化的要求, 集成了学生身份识别、图书借阅、门禁出入、考勤考绩、学籍学分、医疗信息等学校信息化管理所需要的功能。在与银行卡配合使用后, 还可以在校内校外消费, 在银行营业网点、ATM 机进行存、取款等。这种应用一方面方便了学生的使用和学校的管理, 但同时也使造假者有利可图。例如, 校园卡可作为学生考试时身份证明的工具, 部分学生为了找人替考, 不惜重金制作假的校园卡, 这就严重妨碍了学校的管理工作。假的校园卡的流通使用, 更会造成学校数据库系统数据紊乱, 造成很大的经济损失, 因此智能校园

ALOCOM、Jena、Jbaki、Struts 等技术的 Java 集成解决方案。随着语义网技术及人工智能的发展, 教育资源库建设领域的技术和应用框架必然进一步走向成熟和完善。

参考文献:

- [1] 龚洪泉等. Semantic Web 研究综述[J]. 计算机应用与软件, 2005, 22(2): 1—7.
- [2] 全国教育科学规划领导小组、中国教育技术协会. 基础教育资源库建设及其在课程与信息技术整合中的应用研究[DB/OL]. <http://gdhdgjzx.com/xiazai/sywk.doc>.
- [3] Clifford A. Lynch, Jan K. Lippincott. Institutional Repository Deployment in the United States as of Early 2005[DB/OL]. <http://www.dlib.org/dlib/september05/lynch/09lynch.html>.
- [4] Rory McGreal, etc. EduSource: Canada's learning object repository network[DB/OL]. <http://klatu-dev.pc.athabasca.ca:8080/dspace/handle/2149/226>.
- [5][11] Juan M. Santos, etc. Standardization. Design of a Semantic Web-based Brokerage Architecture for the E-learning Domain[DB/OL]. <http://ieeexplore.ieee.org/iel5/10731/33854/01612291.pdf>.
- [6] Miguel- Angel Scilia, etc. A semantic lifecycle approach to learning

object repositories[DB/OL]. <http://ieeexplore.ieee.org/iel5/10175/32498/01517672.pdf?arnumber=1517672>.

- [7] Martin Doerr, etc. Metadata and the CIDOC CRM- a Solution for Semantic Interoperability [DB/OL]. <http://www.cidoc.net>.
- [8] HONG- ZHE LIU, etc. An ontology- based architecture for distributed digital museum[DB/OL]. <http://ieeexplore.ieee.org/iel5/10231/32623/01526913.pdf?arnumber=1526913&isnumber=32623>.
- [9] 钱巧能. Ontology 在 Semantic Web Services 中的应用综述[J]. 计算机与数字工程, 2005, 33(11): 126—129.
- [10] Yolaine Bourda. The semantic Web for learning resources[DB/OL]. <http://ieeexplore.ieee.org/iel5/8621/27318/01215104.pdf>.

收稿日期: 2007 年 7 月 9 日

责任编辑: 马小强