

基于 Ontology 的基础教育资源规划(BERP)的研究

李 丽 李吉桂

(华南师范大学计算机学院 广州 510631)

摘 要 为了解决当前数字化校园中在教育资源的分类管理、资源使用、资源的注册/发布、资源搜索等方面存在的问题与不足,研究了基于 Ontology 的基础教育资源规划(BERP)。其主要工作是提出一种新的教育资源管理中心的逻辑框架;引进本体模型,描述各类资源,在此基础上,实现资源搜索、过滤和建立资源管理中心的主要业务流程。

关键词 本体,资源本体模型,过滤,个性化管理教育资源

中图法分类号 TP393 文献标识码 A

Research of the Basic Educational Resource Planning Based on Ontology

LI Li LI Ji-gui

(School of Computer, South China Normal University, Guangzhou 510631, China)

Abstract In view of these current problems and defects in the digital campus, such as sorting management, the using of resource, the registration or publication of resource, the search for the education resource, and so on, the paper proposed to carry out a research on the Basic Educational Resource Planning based On Ontology (BERP). The paper's main contribution is to bring up the idea of the education resource management center. Besides, the paper adopted an Ontology model and described the resources. Based on them, the paper accomplished researching and identifying the education resources, and then, built the main business processes in the education resource management center.

Keywords Ontology, Education resource ontology model, Filtration, Management individually of education resource

1 引言

随着网络技术的发展,教育行业也加快了数字化的步伐,尤其是近年来基础教育对数字化校园的要求越来越高。然而,由于数字化校园中教育资源不断增多,资源类型多种多样,存储的物理位置离散地分布在校园网内,这些因素给教育资源的管理带来了很多问题。常见问题有以下几类:(1)对日益增多的资源如何进行分类管理;(2)如何解决资源存储离散分布带来的影响,如搜索、更新、使用等操作;(3)如何有效避免未经授权用户对教育资源的不适当操作;(4)如何过滤敏感资源,如教育意义下的敏感资源、带病毒的资源等;(5)如何对教育资源进行统一规划。

为解决上述问题,本文研究基础教育资源规划(BERP, Basic Educational Resource Planning),实现这样的目标:通过个性化的分类和管理,使用户通过教育资源管理中心能方便获取其所需的、高效的教育资源。可见,教育资源的管理是基础教育资源规划(BERP)的重要内容。因此,在本文研究的BERP中包含:建立教育资源管理中心、用本体(Ontology)描述教育资源管理中心的教育资源、通过资源管理中心安全使用教育资源的实现等几个主要内容。

2 基础教育资源管理中心的设计

2.1 教育资源管理中心的主要功能

- (1)对教育资源进行个性化分类,实现教育资源类别的个性化管理;
- (2)通过管理内部资源 UDDI,方便用户发布/共享/按级赋权使用资源;
- (3)通过资源匹配实现用户搜索和使用所需的教育资源;
- (4)过滤/验证资源内容是否满足教育意义下的安全性。

2.2 教育资源管理中心逻辑结构模型

教育资源管理中心主要由用户接口层、鉴别层(过滤层)、资源注册层、资源搜索层、资源存储层 5 层,3 个基本数据库:资源库、资源信息注册库和资源分类库所组成。教育资源管理中心逻辑结构模型如图 1 所示。

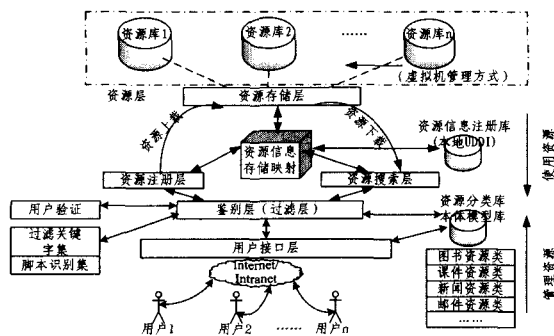


图 1 教育资源管理中心逻辑结构图

- (1)用户接口层:用户使用教育资源管理中心的唯一途

径,通过 Internet 或校园内部网,使用用户接口层递交相关的业务申请;

(2)鉴别层(过滤层):验证用户身份、过滤敏感资源及使用资源分类库中对应本体模型实现资源类别判别及过滤的功能层,接收用户提交的业务申请,从中提取业务申请中涉及的资源类别,以快速访问/查询/管理资源库中对应的资源;

(3)资源注册层:接收鉴别层(过滤层)提交的资源类别,获取用户对发布资源的规格化后的 XML 描述,写入资源信息映射表,并将资源的详细信息写入资源信息注册库中,同时向资源存储层提交存储资源任务要求,以完成用户上传/发布教育资源;

(4)资源搜索层:负责资源注册层、鉴别层(过滤层)提交的资源类别,搜索资源信息映射表,从资源信息注册库中获取所需资源的详细信息,向资源存储层提交资源获取请求;

(5)资源存储层:该层接收来自资源注册层的资源存储要求,并根据需要将资源按需存储到对应的物理位置;或者接收来自资源搜索层的资源获取请求,根据资源的详细信息到校园网络内指定的存储位置获取用户所需的资源,然后返回给用户。

3 教育资源本体模型的构建

3.1 Ontology(本体)技术

本体^[3]的概念最初起源于哲学领域。在信息系统中引入本体是为了描述目标世界的信息,从而满足系统中推理世界模型的需求。本体具有良好的概念层次结构和对逻辑推理的支持,在知识检索中有广泛应用。本文应用本体的基本概念和相关技术定义教育资源,通过教育资源本体模型来描述资源管理中心的教育资源,使资源管理中心实现管理教育资源的智能化和动态化,提高教育资源匹配过程中匹配的查准率和个性化管理能力,将资源的查找和使用过程转换为过程模型中的过程本体论与查询中所用到的过程本体论相匹配的过程,充分发挥资源管理中心的优势。

3.2 教育资源的描述

为实现教育资源管理中心的个性化资源类别管理,需提供对一个对资源进行描述的资源模型,该资源类模型也是资源本体模型的基础。有了规格化的资源模型,可以方便地按用户需求对资源进行个性化的描述,实现对教育资源的管理。在注册资源类时,可以描述该类资源的资源名、资源类别名、内容摘要、适用范围、辅助属性等相关信息。下面以教育资源管理中心常见资源——教学课件资源为例描述资源类的属性。课件资源的 E-R 图描述如图 2 所示。

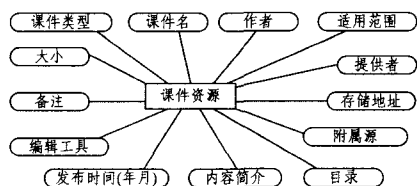


图 2 课件资源 E-R 图描述

在搜索课件资源时,主要针对“课件类型”、“适用范围”来予以匹配,并按照发布时间(从大到小)显示,实现资源的查找及使用。各类资源属性描述相似,仿照课件资源描述方式即可描述其他类别资源,限于篇幅,这里不作细述。

3.3 教育资源本体模型

从资源管理中心的逻辑结构图可知,各个业务都须将资源表示为规范的描述语言所表示的一个单位,这也是教育资源本体模型设计的主要目的。结合教育资源描述所拟定的资源模型的基本结构和 E-R 图,以本体的基本形式定义教育资源本体模型是实现个性化搜索和使用教育资源的基础和前提。

定义 1 教育资源本体模型

$Edu_RS = \{ R_Info, R_DInfo, R_AInfo, R_IInfo, R_PInfo \}$,

其中, R_Info 表示教育资源基本信息,包括资源 ID(rsID)、资源名称(rsN)、资源类型信息(rsTI)、资源大小(rsSize)、资源的作者(rsAN)等相关信息; R_DInfo 表示教育资源应用领域基本信息,包括资源的所属学科(rsSD)、资源的应用年级(rsGI)等; R_AInfo 表示教育资源的使用形式基本信息:应用类别(rsCT)、应用形式(rsAT)、应用环境描述(rsAE); R_IInfo 表示教育资源的接口信息,接口名(rsIN)、输入参数个数(rsInputN)、参数名(rsInputDN)、参数数据类型(rsInputDT)等; R_PInfo 表示教育资源的用户权限信息:用户 ID(UserID)、访问/读权限(rsRd)、删除/更新/下载等权限。

根据上述基本定义,可以构造教育资源本体模型,例如物理课程中重力试验教案的本体模型可被描述为:

重力试验课件资源 = { <高二物理试验 13>U<重力试验教学计划>U<试验教学课件文件(.ppt)>U<230K>U<张三>U...
<物理>U<中学五年级>U...
<重力试验辅导>U<教师讲授>U<物理实验室>
<Phy_Exp_13>U<1>U<物理实验编号>U<String>U...
<PhyT0623>U<T>U... }

根据教育资源本体的基本定义,可以建立与上例类似的教育资源本体。使用教育资源本体可以描述查询/搜索资源的请求或描述获取资源的请求,保证教育资源管理中心的业务正常运作。

4 BERP 中主要操作流程的实现

4.1 资源管理中心操作流程描述

根据资源管理中心结构图,在使用资源管理中心的过程中,其业务流程主要包括:

(1)注册资源类

用户注册教育资源类是在注册/发布教育资源时触发的一个过程。用户登录后,通过用户接口层,向教育资源管理中心发出“注册/发布教育资源请求”;生成“填写注册/发布教育资源的描述页面”,资源管理中心在资源类型库中搜索、获取资源类型并生成资源类型列表;用户判断待注册/发布的教育资源类是否存在于资源类型列表中;若不存在该资源类型,则填写新增属性的相关信息;资源管理中心将所有描述信息写入资源类型库中,完成用户注册教育资源类的过程。

(2)注册/发布教育资源

注册/发布教育资源是资源管理中心的 core 功能之一。用户通过用户接口层提交注册/发布教育资源的请求;资源管理中心响应请求,获取用户权限信息并生成登记资源信息页面;用户登记待注册/发布的教育资源后,资源管理中心进行

信息过滤和用户权限验证;若为教育意义下的安全资源,资源管理中心根据所需空间的大小,将该资源保存在一个按资源类划分的资源库中,同时返回存储地址和获取资源的数据接口;资源管理中心获取该信息,并写入资源信息注册库里对应的记录中,同时更新资源信息存储映射表。

(3)资源的获取

资源管理中心的作用就是为师生提供教育资源,实现教师之间、同学之间、师生之间的学术交流,提高教学质量,所以获取资源是资源管理中心的又一核心功能。用户通过用户接口层提交获取教育资源的请求;资源管理中心根据用户信息,获得其对应的权限信息,进而获取具有相应权限的资源类型,并生成资源获取申请页面;资源管理中心将用户填写完毕的资源需求描述提交至资源搜索层;通过搜索资源信息注册库,得到所需资源的相关信息;根据相关信息访问资源信息存储映射,获得资源接口地址;使用资源接口地址访问教育资源库,完成资源获取业务。

4.2 教育资源搜索的实现

作为数字化校园中为用户提供资源的教育资源管理中心,实现教育资源搜索是重要的技术之一。而教育资源搜索结果的好坏可以用查全率和查准率两个标准来衡量。查全率用以衡量搜索到的信息的多少,而查准率用以表示搜索到的信息的准确度,即是否是所期望搜索到的信息。由于本体具有良好的概念层次结构和对逻辑推理的支持,因此对于基于本体的教育资源的搜索,可以较好地兼顾查全率和查准率,进而满足用户的需求。

基于本体的教育资源搜索主要包括:用户搜索请求描述及处理和资源匹配两个步骤,其中教育资源匹配是关键的一环。在进行教育资源搜索前先对本体进行预处理,可以省去匹配时的推理过程。使用 protégé 作为建模工具,得到的 OWL 文件中包含了各教育资源的直接知识。在预处理中,使用 Jena^[5-7]的推理引擎利用 OWL 文件中的知识进行推理,得到全部知识后以 RDF 三元组的形式存储在数据库中,这样仅通过数据库查询操作就可以完成教育资源的匹配。教育资源搜索过程可以视为“教育资源实体描述”与“教育资源搜索描述”的匹配过程。具体可以描述为:

(1)教育资源管理中心通过用户接口层将用户的教育资源搜索请求提交至资源搜索层;

(2)教育资源搜索请求与教育资源的本体模型相结合,用内部规范(XML 语言描述的规范)的资源搜索请求描述用户的资源请求;

(3)提取搜索请求描述中 R_PInfo 域的用户权限信息和 R_Info 域中的资源类型信息,结合用户信息库和资源类型库以及资源信息注册库的相关信息,确定当前用户可搜索的资源描述集范围;

(4)在可搜索的资源描述集中用资源请求描述与资源描述集的资源进行匹配,求得资源与请求的相似度;

(5)结合资源管理中心所设置的相似度阈值 Sim_{min} ,筛选相应的教育资源数据集(该过程与过程(4)同步实现);

(6)返回满足条件(相似度大于相似度阈值 Sim_{min})的教育资源数据集。

4.3 过滤敏感教育资源的实现

在资源管理中心的运作流程中,无论是注册/发布资源还

是获取资源,都涉及到教育资源是否满足教育意义下的安全性,过滤敏感教育资源成为资源管理中心必须实现的技术。由于数字化校园中会结合访问控制技术、防火墙等很多过滤策略,在资源管理中心里采用一种较为简单的过滤即可,主要思想是根据教育资源源文件的文件类型使用文本过滤或 XML 脚本识别过滤。过滤算法描述如下:

(1)读取教育资源源文件信息放入待审文件集 File_Set 中,同时获取该文件的文件扩展名并放入文件类型集 FileType_Set 中;

(2)重复步骤(1),直至教育资源所有源文件均存入待审文件集 File_Set 中,同时也获得 FileType_Set;

(3)通过查阅过滤关键字集/脚本识别集(参阅图 1 教育资源管理中心逻辑结构图)获取对应的规则,最后形成该次业务的过滤规则库 Rule_Set;

(4)使用规则库 Rule_Set 审阅 File_Set 中的文件。算法结束。

资源管理中心中设置的该鉴别层(过滤层),可以更有效地保证资源管理中心中资源的安全性。其过滤的基本流程描述为:

(1)在注册/发布教育资源的过程中,首先将待注册/发布的资源源文件提交至鉴别层;

(2)鉴别层(过滤层)载入教育资源源文件集后,过滤引擎获取用户在注册/发布教育资源中填写的资源描述和资源源文件描述;

(3)过滤引擎首先根据用户的权限信息及其拥有权限的资源类型信息鉴别/过滤资源源文件类型,以保证资源类型的有效性和教育意义下的安全性;

(4)过滤引擎访问过滤关键字集和脚本识别集,获取相应资源类型的过滤规则;

(5)过滤引擎根据获得的过滤规则对待验证的教育资源进行安全性判断;

(6)验证为敏感资源时,鉴别层向用户接口层提交拒绝教育资源的反馈意见,否则相应的教育资源源文件被提交至资源存储层,并进行存储操作。

对于教育资源管理中心的设计,本文提出的利用本体的基本概念实现对教育资源的规范化,应用到对数字化校园中教育资源的存储/获取,相对于传统教育资源管理的模式是一种创新。一方面,通过规范化教育资源的描述的定义,使得对教育资源的描述具有一定的描述标准,从而实现用户对教育资源个性化的拟定;另一方面,在搜索的过程中利用基于本体的匹配兼顾了查全率和查准率的搜索指标。管理中采用资源过滤的规则,使教育资源搜索具有一定的智能化和管理的安全化。

结束语 本文从解决教育资源管理中的分类管理、资源的使用、注册/发布、搜索等问题出发,提出了教育资源管理中心的一种新结构,给出资源描述模型;分析并定义了教育资源本体模型。并将其应用于教育资源的搜索、鉴别和过滤;最后分析、实现了资源管理中心的几个主要业务流程。

参考文献

- [1] 张维明,等. 语义信息模型及应用[M]. 北京:电子工业出版社, 2002

(下转第 182 页)

- (4U) $C^*(XUY) = C^*(X) \cup C^*(Y)$;
- (5L) $X \subseteq Y \Rightarrow C^*(X) \subseteq C^*(Y)$;
- (5U) $X \subseteq Y \Rightarrow C^*(X) \subseteq C^*(Y)$;
- (6L) $C^*(C^*(X)) = C^*(X)$;
- (6U) $C^*(C^*(X)) = C^*(X)$;
- (7) $C^*(\sim X) = \sim C^*(X)$, $C^*(\sim X) = \sim C^*(X)$.

由此可知,这种覆盖粗糙集模型不同于文献[6]中的覆盖粗糙集模型,它满足 Pawlak 粗糙集的许多重要性质,但有些性质也不成立(如 $C^*(C^*(X)) = C^*(X)$, $C^*(C^*(X)) = C^*(X)$)。因此研究新的覆盖粗糙集公理化体系有广泛的现实和理论意义。

3 新覆盖粗糙集的公理体系

定理 1 中的(7)表明覆盖上下近似之间可以相互定义,一旦下近似给定,上近似也就确定了。因此我们的讨论仅针对下近似来进行。

定理 2 设 U 是非空有限论域, $f: P(U) \rightarrow P(U)$ 是一个映射。若对任意的 $X, Y \subseteq U$, f 满足:

- (1) $f(U) = U$;
- (2) $f(X) \subseteq X$;
- (3) $f(X \cap Y) = f(X) \cap f(Y)$;
- (4) $f(f(X)) = f(X)$;

则存在 U 的一个覆盖 C 使得 $f(X) = C^*(X)$ 。

为证明上述定理,我们先引入下面的引理:

引理 1 在定理 2 中,若 $f(X \cap Y) = f(X) \cap f(Y)$, 则 $X \subseteq Y \Rightarrow f(X) \subseteq f(Y)$, 即映射是递增的。

证明:由 $X \subseteq Y$ 得, $f(X) = f(X \cap Y) = f(X) \cap f(Y)$, 因此 $f(X) \subseteq f(Y)$ 。

定理 2 的证明:

设 $C = \{K \subseteq U \mid f(K) = K, K \neq \Phi\}$, 则由(1)知 C 是 U 的覆盖,对任意的 $X \subseteq U$, $C^*(X) = \{x \in U \mid \bigcap Md(x) \subseteq X\}$ 。

$\forall x \in C^*(X)$, 令 $Md(x) = \{K_1, K_2, \dots, K_m \mid K_i \in C\}$, 有 $x \in \bigcap_{i=1}^m K_i$, 又 $f(K_i) = K_i$, 可知 $x \in \bigcap_{i=1}^m f(K_i)$ 。由 $\bigcap Md(x) \subseteq X$, 引理 1 和(3)可得 $\bigcap_{i=1}^m f(K_i) = f(\bigcap_{i=1}^m K_i) = f(\bigcap Md(x)) \subseteq f(X)$, 从而 $x \in f(X)$, 即 $C^*(X) \subseteq f(X)$ 。

另一方面,任取 $X \subseteq U$, 由(4)知 $f(X) \in C$ 。若 $f(X) = \Phi$, 则显然 $f(X) \subseteq C^*(X)$ 。否则 $\forall x \in f(X)$, 有最小描述的定义知 $\bigcap Md(x) \subseteq f(X)$, 又由(2)有 $f(X) \subseteq X$, 于是得 $\bigcap Md(x) \subseteq X$, 从而 $x \in C^*(X)$, 即 $f(X) \subseteq C^*(X)$ 。

综上有 $f(X) = C^*(X)$ 。

定理 3 定理 2 给出的 4 条公理是相互独立的。

证明:设 $U = \{a, b, c\}$ 。

对任意 $X \subseteq U$, 令 $f(X) = \Phi$, 经验证,此时 f 满足(2)、(3)和(4), 但不满足(1), 故 $\{(2), (3), (4)\} \not\Rightarrow (1)$ 。

对任意 $X \subseteq U$, 令 $f(X) = U$, 经验证,此时 f 满足(1)、(3)和(4), 但不满足(2), 故 $\{(1), (3), (4)\} \not\Rightarrow (2)$ 。

令 $f(U) = U, f(\{a, b\}) = f(\{b, c\}) = f(\{a, c\}) = f(\Phi) = \Phi, f(\{a\}) = \{a\}, f(\{b\}) = \{b\}, f(\{c\}) = \{c\}$, 经验证,此时 f 满足(1)、(2)和(4), 而 $\{a\} = f(\{a\}) = f(\{a, b\} \cap \{a\}) \neq f(\{a, b\}) \cap f(\{a\}) = \Phi$ 不满足(3), 故 $\{(1), (2), (4)\} \not\Rightarrow (3)$ 。

令 $f(U) = U, f(\{a, b\}) = \{a\}, f(\{b, c\}) = \{b\}, f(\{a, c\}) = \{c\}, f(\{a\}) = f(\{b\}) = f(\{c\}) = f(\Phi) = \Phi$, 经验证,此时 f 满足(1)、(2)和(3), 而 $\Phi = f(\{c\}) = f(f(\{a, c\})) \neq f(\{a, c\}) = \{c\}$, 不满足(4), 故 $\{(1), (2), (3)\} \not\Rightarrow (4)$ 。

综上所述,定理 2 中的 4 条公理是相互独立的。

基于覆盖粗糙集的对偶原理,覆盖上近似有下面的公理体系:

定理 4 设 U 是非空有限论域, $f: P(U) \rightarrow P(U)$ 是一个映射。若对任意的 $X, Y \subseteq U$, f 满足:

- (1) $f(\Phi) = \Phi$;
- (2) $X \subseteq f(X)$;
- (3) $f(X \cup Y) = f(X) \cup f(Y)$;
- (4) $f(f(X)) = f(X)$;

则存在 U 的一个覆盖 C 使得 $f(X) = C^*(X)$, 且 4 个条件是相互独立的。

结束语 本文讨论了覆盖粗糙集的公理化和可靠性问题,给出了关于覆盖粗糙集的一组公理组,为覆盖粗糙集理论的完善提供了坚实的基础。虽然新覆盖粗糙集与 Pawlak 粗糙集有许多相同的性质,但二者的公理系统却不完全相同,原因在于新覆盖粗糙集不满足性质 $C^*(C^*(X)) = C^*(X)$, $C^*(C^*(X)) = C^*(X)$ 。因此,覆盖粗糙集的公理化体系的讨论远比 Pawlak 粗糙集的公理化体系复杂。

参考文献

- [1] Pawlak Z, Grzymala-Bausse J, Slowinski R, et al. Rough sets [J]. Communications of the ACM, 1995, 38(11): 89-95
- [2] Pawlak Z. Rough sets: Theoretical aspects of reasoning about data[M]. Dordrecht: Kluwer Academic Publishers, 1991
- [3] Lin T Y, Liu Q. Rough approximate operators; Axiomatic rough set theory[J] // Ziarko W P, ed. Rough Sets, Fuzzy Sets and Knowledge Discovery. London: Springer-verlag, 1994: 256-260
- [4] 祝峰, 何华灿. 粗集的公理化[J]. 计算机学报, 2000, 23(3): 330-333
- [5] 孙辉, 刘大有, 李文. 粗集公理组的极小化[J]. 计算机学报, 2002, 25(2): 202-209
- [6] Zakomski W. Approximations in the space (U, Π) [J]. Demonstration Math., 1983, 16: 761-769
- [7] Zhu W, Wang F Y. Reduction and axiomization of covering generalized rough sets[J]. Information Sciences, 2003, 152: 217-230
- [8] Xu W H, Zhang W X. Measuring roughness of generalized rough sets induced by a covering[J]. Fuzzy Sets and Systems, 2007, 158: 2443-2455
- [9] 基于本体语义检索技术研究[J]. 计算机工程与科学, 2008(4)
- [10] Jena 在基于 Ontology 的 TB T 文档搜索中的应用研究[J]. 中国科技信息, 2007(15)
- [11] 基于本体的资源描述和检索方法研究[J]. 计算机工程与应用, 2005(36)
- [12] 基于本体论的课件资源检索系统设计[J]. 计算机工程与设计, 2006(5)

(上接第 171 页)

- [2] 李善平, 尹奇群, 胡玉杰, 等. 本体论研究综述[J]. 计算机研究与发展, 2004(7)
- [3] 廖明宏. 本体论与信息检索[J]. 计算机工程, 2006(2)
- [4] 张钊. 基于语义的网络服务匹配机制的研究与实现[D]. 北京: 清华大学, 2005
- [5] 基于 Ontology 的语义信息检索模型研究[J]. 计算机工程与设计, 2008(11)