

# 教育资源本体 OntoER 创建研究 ——基于行业技术规范的本体创建案例

薛小平<sup>1</sup>, 耿 壮<sup>1</sup>, 薛元韬<sup>2</sup>

(1 深圳信息技术学院, 广东 深圳 518029; 2 墨尔本大学, 墨尔本 3183)

**摘 要:** 以 W3C 发布的 Web Ontology Language (OWL) 为本体语言标准, 以教育部教育信息化技术标准委员会发布的《教育资源建设技术规范》为知识领域研究如何创建教育资源本体——OntoER. 研究结果可作为《教育资源语义 Web》的技术内核, 为研制下一代互联网——语义 Web 突破技术瓶颈, 使得在一个知识领域率先创建语义 Web 成为可能. 通篇采用描述逻辑数学工具, 对研究对象、研制过程及结论进行形式化表述与抽象, 从中得出一些规律性的结论.

**关键词:** 元数据; 本体; T-box 公理集; A-box 公理集; 描述逻辑

**中图分类号:** TP391.13; TP311.12

**文献标识码:** A

**文章编号:** 1000-7180(2011)01-0099-05

## Research on Educational Resources Ontology Constructing

XUE Xiao-ping<sup>1</sup>, GENG Zhuang<sup>1</sup>, XUE Yuan-tao<sup>2</sup>

(1 Shenzhen Institute of Information Technology, Shenzhen 518029, China;

2 University of Melbourne, Melbourne 3183, Australia)

**Abstract:** The approach of this project is how to construct the Ontology about Educational Resource (OntoER) with Web Ontology Language (OWL) released by W3C as technology standards and Technical Specification for Educational Resource Construction (TSERC) granted by National Ministry of Education (NME) as knowledge domain. The production about this approach would be of the technical core of Educational Resource Semantic Web. The technical bottle-neck about Next Generation Internet would be broken through with the approach. The dream would be realized for people to construct the Semantic Web on one knowledge domain. The mathematic tool of Description Logic has been introduced. The research objects as well as approach and conclusions have been formally presented, so some commonly conclusion be elicited.

**Key words:** metadata; ontology; T-box axiom set; A-box axiom set; description logics

### 1 引言

把网格应用于高校教育资源管理的问题研究中, 就能把地理上分布的教育资源进行一体化组织与处理, 可以有效解决高校信息资源分散性与需求集成性的矛盾, 有力地促进高校信息资源的共享和使用<sup>[1]</sup>.

建立某一知识领域的本体实际上就是创建该领域的知识库(knowledge base). 在描述逻辑系统中, 一个知识库系统就是领域知识的集合, 它由两个子

集构成, 一个子集是 T-box, 另一个是 A-box; T-box (terminology box), 是一个概念性知识公理的集合, 而 A-box(assertion box)断言集则是领域的实例性知识的集合. 所谓建立行业资源本体就是建立行业领域的 T-box 和 A-box 集合.

### 2 教育资源技术规范简介及 UML 建模

**定义 1** 不继承任何概念的概念为初始概念 (RootConcept). 在特定的 Ontology 中初始概念是惟一的.

**定义 2** Ontology 中概念和概念之间的关系组成 Ontology 继承关系图  $G(Vconcept, Einberit-relatic)$ , 其中  $Vconcept$  是 Ontology 中所有概念的节点集合,  $Einberit-relatic$  对应于 Ontology 中继承关系的有向边集合.

**定义 3** 继承关系图  $G$  中, 节点  $concept_i$  到  $concept_j$  的有向边的序列为  $concept_i$  的一种继承关系路径, 路径长度  $L^k$  为第  $K$  条路径上的边数<sup>[2]</sup>.

由《Learning Object Metadata》(LOM)<sup>[3]</sup>和《教育资源建设技术规范信息模型》<sup>[4]</sup>形成的教育资源行业规范是一个体量相当庞大的技术文档.

### 2.1 教育资源规范中的聚合类

对于《教育资源建设技术规范信息模型》中的初始概念 (RootConcept) 用“教育资源库”表示. 它由 LOM 规范中的“必须元数据”、“可选元数据”、“学习对象类”组成. 其中, 必须元数据类规定了在 Web 页面中呈献学习对象类时, 所应提供的一些基本元数据和通用概念的数据模式, 并为教育资源属性提出标注标准、评价标准及管理标准. 根据 LOM 规范, 它由如图 1 所示的六个子类组成, 它们彼此是不相交的 (disjoint). 用 UML 中聚集关联来建模, 如图 1 所示.

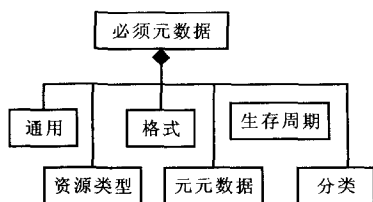


图 1 必须元数据的构成

### 2.2 教育资源规范中的继承—泛化类

教育资源中的学习对象类是一个抽象类, 它本身并没有学习对象实例. 这个抽象类所包含的语义信息由必须元数据类提供, 并被试题库等七个子类所继承, 七个子类彼此是不相交的 (disjoint), 如图 2 所示.

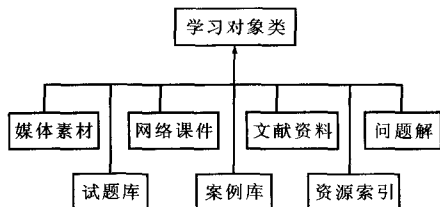


图 2 继承和泛化关联

### 2.3 教育资源规范中的 ER 模式

为了建立《教育资源语义 Web》中的实体—关系 (ER) 模型, 这里将以试题库资源的建模为例进行分析, 如图 3 所示.

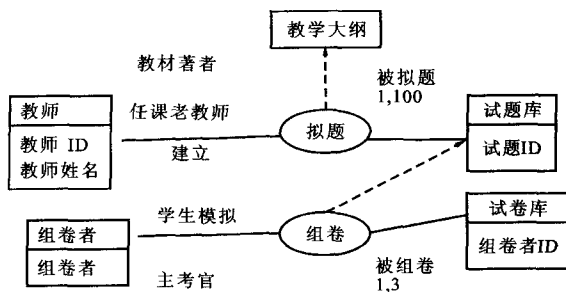


图 3 试题库资源的 ER 建模

至此, 可对教育资源 UML 模型小结如下: ① 聚合类型; ② 继承—泛化类型; ③ 实体—关联; 类之间的连接叫关联 (association), 这种关联一般用一个动词来描述. 动词两端的类在关联中扮演着某种角色. 和实体类一样, 关联也有自己的属性和操作. 此时的关联叫做关联类 (association class).

### 3 ER 模式的语法与语义

**定义 4** 一个 ER 模式是一个五元组  $Schema = (lang, isa, att, rel, card)$ , 其中:

①  $lang$  是一个有限的词语表:  $lang = E \cup A \cup R \cup U \cup D$ , 即  $lang$  是实体 (entity)、属性 (attribute)、关系 (relationship)、角色 (role)、域符号 (domain) 等集合的并. 这五个子集彼此是不相交 (disjoint) 的;

②  $isa \subseteq E \times E$  是一个广义笛卡尔积, 它是一个二元关系, 用于表示两个实体间的  $isa$  联系;

③  $att$  是一个映射函数, 它将  $E$  集合中的实体映射为属性  $A$  和符号域  $D$  的标示关系,  $att(E) = [\dots, A; D, \dots]$ . 进一步地, 对  $E$  中的每个实体,  $A$  在基本域  $B(D)$  中有唯一取值, 则称属性  $A$  是  $E$  的唯一属性;

④  $rel$  也是一个映射函数, 它将关系  $R$  集合的实体映射为角色  $U$  和实体  $E$  的标示关系,  $att(R) = [\dots, U; E, \dots]$ .

⑤  $card$  是一个关于  $E, R, U$  集合间函数,  $card(E, R, U) = [mincard, maxcard]$ . 它表示一个关联实体  $e$  和关系  $r$  特定角色的基约束 (cardinality constraints).

### 4 OWL 本体的描述逻辑表示

正如前述引言中所说, 在描述逻辑系统中, 一个 OWL 本体, 就是相关领域的知识库系统, 它由两个子集构成, 一个子集是 T-box, 另一个是 A-box;

T-box 是一个概念性知识及公理的集合, 在 OWL 规范中<sup>[5]</sup>, 概念用类表示, 也即, T-box 是一个

类及相关公理的集合.由此,可对 OWL 本体中的 T-box 做如下定义:

**定义 5** 一个 OWL 本体的 T-box 是一个二元组,  $T\text{-box} = (id, axiom)$ , 其中:

①  $id = CID \cup DRID \cup OPID \cup DPID$ , 它是类的有限标识符 (identifier) 集合. CID 是类标识符、DRID 是数据值域标识符、OPID 是对象属性标识符、DPID 是数据属性标识符;

② axiom 是 T-box 中的有限的公理序列, 其中包含类公理和属性公理.

- 基本类公理 (Primitive class axioms):  $(C_N \subseteq C_1 \cup C_2 \cup \dots \cup C_h)$ ,

表明一个类名包含于若干个子类术语中;

- 类定义公理 (Class definition axiom):  $(C_N \doteq C)$ ,

表明了一个类名和一个类术语 (原子类) 间的等价性.

- 不交公理:

$$C_1 \subseteq \neg (C_2 \cup C_3 \cup \dots \cup C_n)$$

$$C_2 \subseteq \neg (C_3 \cup C_4 \cup \dots \cup C_n)$$

...

$$C_{n-1} \subseteq C_n$$

表明在不交公理限定下, 本体中任一个类 ( $C_i$ ) 的成员个体不可能是同组中其它类的成员.

- 存在量词约束 (Existential Restrictions):  $(C_1 \subseteq \exists \text{ hasP. } C_2)$

表明, 做为  $C_1$  类, 至少有一个 hasP 属性, 其值 (filler) 是  $C_2$  的成员;

- 全称量词约束 (Universal Restrictions):  $(C_1 \subseteq \forall \text{ hasP. } C_2)$

表明, 做为  $C_1$  类, 若有一个 hasP 属性, 则其值 (filler) 必是  $C_2$  成员. 所以, 对于 hasP 属性, 全称量词  $\forall$  并不保证约束关系一定存在. 但若存, 则约束值 (filler) 一定是  $C_2$  的成员.

- 终止公理 (Closure Axioms): 终止公理由两部分子句组成, 第一部分子句是由存在量词形成的若干个约束 ( $\exists \text{ hasP. } C_2$ ), 第二部分子句是全称量词约束 ( $\forall \text{ hasP. } C_3$ ), 其中的  $C_3$  是存在量词 ( $\exists \text{ hasP. } C_2$ ) 中所有属性值 ( $C_2$ ) 的交集. 有了终止公理, 可防止推理机 Racer<sup>[6]</sup> 对被约束类的语义做无边界演绎.

- 基数约束:

$$C_1 \subseteq \text{hasP} \geq nR$$

$$C_1 \subseteq \text{hasP} \leq nR$$

$$C_1 \subseteq \text{hasP} = nR$$

表明被约束类  $C_1$  具有的 hasP 属性个数的量值.

由此可归纳出, 构建一个知识领域的 T-box 就是要构建出其中两个子集, 其一是标识符子集, 其二是公理子集. 标识符子集由类标识符等四个下级集合组成; 公理集合由基本类公理等七个公理组成.

**定义 6** 一个 A-box 包含知识领域中所有个体的断言 (assertion), 它是一个二元组,  $A\text{-box} = (i, \text{assertion})$ , 其中:

① 个体名集  $i$  是 A-box 所有个体的署名, 且要求个体名集  $I$  与类名集  $C$  及角色名集  $R$  是不相交的;

② 断言 (assertion) 是 A-box 中所有个体、角色、的声明的集合, 共有四种类型的断言:

- 个体断言: 声明某一个体名 (IN) 是一个相关类 ( $C$ ) 的实例;

- 角色断言: 声明一个与个体 1 ( $IN_1$ ) 相关角色  $R$  的填充值是个体 2 ( $IN_2$ );

- 属性断言: 声明一个对象名 (ON) 是一个与个体名 (IN) 相关角色  $R$  的填充值;

- 约束: 声明了领域中对象间的关系;

## 5 从教育资源 UML 模型到教育资源本体翻译

以下分别给出聚合类型、继承-泛化类型、实体-关联 (ER) 类型模型转换为 OWL 本体的规则.

### 5.1 聚合类的翻译

**规则 1** 将 UML 模型中的聚合类翻译为 OWL 本体规则可归纳为, 聚合类中父-子类转换后仍为相应父子关系, 类的名称在不引起混淆的条件下可同名使用; 每个类的属性翻译为量词约束和终止公理; 但父类不能有量词约束和终止公理; 对于行业规范中大量存在的枚举型属性值翻译为覆盖公理.

必须元数据  $\subseteq$  (通用  $\cup$  生存周期  $\cup$  元元数据  $\cup$  格式  $\cup$  资源类型  $\cup$  分类)

通用  $\subseteq \exists \text{ has 标题. (xsd : string) cardinality}$  (1)

(1)

通用  $\subseteq \exists \text{ has 关键字. (xsd : string)}$  (3)

通用  $\subseteq \exists \text{ has 描述. (xsd : string)}$  (4)

通用  $\subseteq \forall \text{ has 标题. (xsd : string) } \cup ( \forall \text{ has 关键字. (xsd : string) } ) \cup ( \forall \text{ has 描述. (xsd : string) } )$  (5)

.....

格式  $\subseteq \exists$  has 位置. (xsd : string) (6)

格式  $\subseteq \forall$  has 位置. (xsd : string) (7)

.....

资源类型  $\subseteq \exists$  has 交互类型. (交互值分区) cardinality(1) (8)

资源类型  $\subseteq \exists$  has 终端用户. (用户值分区) cardinality(1) (9)

资源类型  $\subseteq \forall$  has 交互类型. (交互值分区)  $\cup$  ( $\forall$  has 终端用户. (用户值分区)) (10)

## 5.2 继承—泛化类的翻译

规则 2 继承—泛化类中的父类必须有相应的量词约束和终止公理,但不可以有实例(个体),除此之外,结论 1 中的规则都适用于继承—泛化类。

## 5.3 ER 模型到 OWL 本体的翻译算法<sup>[7-8]</sup>

一个 ER 模式与一个 OWL 本体在主要方面存在着对应关系,基于以上形式化表述,将一个 ER 模式翻译成一个 OWL 本体的形式化方法由下列算法 1 给出。

算法 1 从 ER 模式到 OWL 本体的翻译算法:

输入: ER 模式 Schema = (lang, isa, att, rel, card)

输出: 由翻译函数  $\varphi$  定义的 OWL 本体 Ontology =  $\varphi$ (Schema) = (ID, axiom), 其中, ID =  $\varphi$ (lang), axiom =  $\varphi$ (isa, attr, rel, card)。

值得指出的是这个翻译算法应是在语义保持的条件下进行的,关于算法的语义保持性证明可参考许卓明先生的《从 ER 模式到 OWL DL 本体的语义保持翻译》。

### 5.3.1 从有限词语表(lang)到标识符集的翻译

ID =  $\varphi$ (lang), 各标识符中的片断标识符在不引起混淆的情况下可与 lang 同名使用,具体规定如下:

- 对每个实体符号  $e \in E$ , 创建一个类标识符  $\varphi(E) \in \text{CID}$ ,  $\varphi(E)$  的片断标识符可与  $e$  同名;

- 对每个关系符号  $r \in R$ , 创建一个类标识符  $\varphi(R) \in \text{CID}$ ,  $\varphi(R)$  的片断标识符可与  $r$  同名;

- 对每个属性符号  $a \in A$ , 创建一个数据类型属性标识符  $\varphi(A) \in \text{DPID}$ ,  $\varphi(A)$  的片断标识符可与  $a$  同名;

- 对每个域符号  $d_1 \in D$ , 创建一个预定义的 XML Schema 数据值域标识符集及一个值分区类标识符集:  $\varphi(D) \in \text{DRID}$ ;

- 对每个角色符号  $u \in U$ , 创建一对互逆对象属

性的标识符  $\varphi(U) \in \text{OPID}$  及  $V = \text{inv-}\varphi(U) \in \text{OPID}$ ,  $\varphi(U)$  的片断标识符可与  $u$  同名;

5.3.2 从 ER 模式构造子元素到 OWL 本体公理的翻译: axiom =  $\varphi$ (isa, att, rel, card)

规则 3 根据有关 T-box 定义, OWL 本体的公理包含有: 属性公理和类公理. 而属性公理分为数据属性公理和对象属性公理; 类公理包含有: 通用概念包含公理、存在量词公理、全称量词公理、终止公理、基数公理等。

- 创建包含性实体公理: 对于每对实体  $e_1, e_2 \in E$  且  $e_1$  is-a  $e_2$ , 可创建包含类公理:

$$\text{Class}(\varphi(e_1)) \text{ isSubclassOf } \text{Class}(\varphi(e_2)) \quad (11)$$

这个公理体现了 T-box 中的通用包含公理。

- 创建实体类的语义约束公理及终止公理: 对于每个实体  $e \in E$ , 且  $\text{att}(E) = [a_1 : d_1, \dots]$ , 创建类公理:

$$\text{Class}(\varphi(e_1)) \subseteq \exists \text{ has } A_1. D_1 \dots \exists \text{ has } A_h. D_h; \quad (12)$$

并创建类终止公理:

$$\text{Class}(\varphi(e_1)) \subseteq \forall \text{ has } A_1. D_1 \dots \exists \text{ has } A_h. D_h; \quad (13)$$

还可创建用对象属性描述类公理:

$$\text{Class}(\varphi(e)) \subseteq \text{has } V. D, \geq (m), \leq (n) \quad (14)$$

这个类公理体现了 T-box 中的基数约束公理。

- 创建关系类公理: 对于每个关系  $r \in R$ , 且  $\text{att}(R) = [u_1 : e_1, \dots, u_h : e_h]$ , 创建类公理:

$$\text{Class}(\varphi(R)) \subseteq \exists \text{ has } \varphi(U_1). \varphi(E_1) \dots \exists \text{ has } \varphi(U_h). \varphi(E_h) \quad (15)$$

并创建终止公理:

$$\text{Class}(\varphi(R)) \subseteq \forall \text{ has } \varphi(U_1). \varphi(E_1) \dots \forall \text{ has } \varphi(U_h). \varphi(E_h) \quad (16)$$

这个类公理是对 UML 关联类所作的终止公理约束。

- 创建数据属性公理: 对于每个实体  $e \in E$ , 且  $\text{att}(E) = [a_1 : d_1, \dots]$ , For  $i = 1, 2, \dots, h$  DO

创建数据属性公理(仅当  $A_i$  的性质具有唯一性时, Functional 选项有效):

$$\text{DatatypeProperty}(\varphi(A_i) \text{ domain}(\varphi(E)) \text{ range}(\varphi(D_i)) [\text{Functional}]) \quad (17)$$

- 创建对象属性公理( $V_i = \varphi(U_i)$ ): 对于每个关系  $r \in R$ , 且  $\text{att}(R) = [u_1 : e_1, \dots, u_h : e_h]$ , For  $i = 1, 2, \dots, h$  DO

可创建属性公理如下:

$$\text{ObjectProperty}(V_i \text{ domain}(\varphi(E_i)) \text{ range}(\varphi$$

(R))) (18)

及  $\text{inv\_Vi} = \text{inverseOf } \varphi(\text{Ui})$  (19)

· 创建覆盖公理:覆盖公理由两部分组成,一部分是被覆盖类(A),另一部分是覆盖类( $C_1, C_2, \dots, C_h$ ),当  $C_1, C_2, \dots, C_h$  互不相交时,它们满足:

$A \subseteq C_1 \cup C_2 \cup \dots \cup C_h$  (20)

表明 A 被  $C_1, C_2, \dots, C_h$  的并集所覆盖.它意味着,凡是 A 中的一个成员,必是  $C_1, C_2, \dots, C_h$  中某一类的一个成员.在 T-box 中,可将必要元数据和可选元数据类中的枚举类属性值设为一个值分区(ValuePartition),其中值分区中的各元素是互不相交的.将此值分区指向一个元数据类的属性时,语义 Web 就得知该类的属性值仅在此值分区中取得.

#### 5.4 翻译清单(略)

### 6 测试结果及分析

以下对所建本体进行语义搜索(Semantic Searching)测试.语义搜索主要是指本体中类的描述逻辑查询(DL Query).例如,搜索本体中“物理学科且有前需”语义的资源.可用如下等价语句定义一个匿名类:

LO and hasDiscipline some 物理学科  
and hasPrerequisite some 有前需

在调用 RacerPro 推理机后,得到语义搜索结果如图 4 所示.

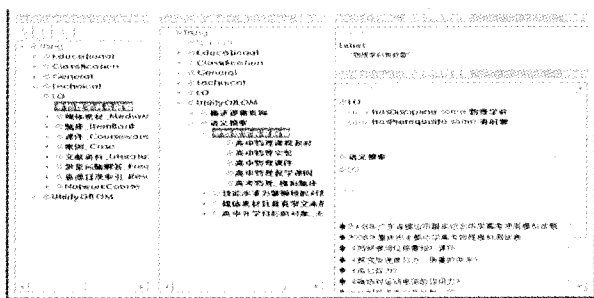


图4 物理学科且有前需的语义搜索

在这个视图中,可看到 RacerPro 推理机的运算结果:“物理学科的资源”类包含了 5 个子类——语义查询的结果;从图 4 中可看出,推理机不仅按“物理学科有前需”的语义查询出该语义所蕴含的子类,而且也查询出了满足该语义的所有个体类成员,在《教育资源本体》中共有 13 个个体满足该语义搜索.

这个实验结果说明了两个重要结论:① 定义类断言类层次视图中没有包含任何子类,但这仅是说明定义类没有显性包含;② 在推理类层次视图中,

才可看到此定义类隐含了五个子类,这个隐含结果只有在推理机操作后才显现出来.而这一点正是本项目研究的目的——建立一个本体,以支持《教育资源语义网》进行语义浏览和语义搜索.

### 7 结束语

本文探讨了如何以行业信息化技术规范为知识领域创建建行业本体,主要结论如下:① 大部分行业技术规范都可被抽象为 UML 模型中的聚合类、抽象—泛化类、ER 实体关系模型;② 聚合类模型可套用规则 1 中式(1)~式(10)的步骤转化为 OWL 本体类;③ 抽象—泛化类也可套用规则 1 的步骤转化为 OWL 本体类,但需要对该模型中的父类做出相应语义约束;④ 对于 ER 实体关系模型可套用规则 3 中式(11)~(20)的步骤转化为 OWL 本体;⑤ 将①~④所做结果合在一起,就构成了描述逻辑系统中 T-box 集合.

### 参考文献:

- [1] 赵桂云. 基于网格的高校教育资源管理研究[J]. 微计算机信息, 2009(3): 179—181.
- [2] 杨保明, 刘晓东, 姚兰, 等. 基于本体论的农业知识的 OWL 描述[J]. 微电子学与计算机, 2007, 24(5): 64—66.
- [3] IEEE. 1484. 12. 1—2002 Draft standard for learning object metadata[S/OL]. [2009—12—12]. [http://1tsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_v1\\_Final\\_Draft.pdf](http://1tsc.ieee.org/wg12/files/LOM_1484_12_v1_Final_Draft.pdf).
- [4] 余胜泉, 朱凌云. 教育资源建设技术规范《征求意见稿》[EB/OL]. [2009—12—10]. <http://www.etc.edu.cn/zhu-yewenzhang/jiao-yu.doc>.
- [5] W3C. OWL Web Ontology Language Guide W3C Recommendation[S/OL]. [2009—12—08]. <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>.
- [6] Volker Haarslev, Ralf Moller Racer. A Core Inference Engine for the Semantic Web[EB/OL]. [2009—12—11]. [http://www.franz.com/agraph/racer/Racer\\_whitepaper.pdf](http://www.franz.com/agraph/racer/Racer_whitepaper.pdf).
- [7] 许卓明, 董逸生, 陆阳. 从 ER 模式到 OWL DL 本体的语义保持翻译[J]. 计算机学报, 2006(10): 60—70.
- [8] Tim Berners Lee. Relational Databases on the Semantic Web[EB/OL]. [2009—12—05]. <http://w3.org/Design/Issues/RDB-RDF.html>.

### 作者简介:

薛小平 男, (1955—), 硕士, 副教授. 研究方向为知识工程、语义 Web 方面.