

网格环境下教育资源元数据模型

The Framework of Education Resource Metadata Based on Grid

陈阳 汪千松 王健峰 (安徽工程科技学院 现代教育技术中心 安徽 芜湖 241000)

摘要: 网格是实现教育资源共享、管理和提供信息服务的热点技术。针对现存教育资源存在的信息孤岛现象,分析教育资源的元数据功能和标准,对资源描述方法进行了比较,提出了一个基于 LOM 的 RDF 模型,并应用网格资源描述框架 RDF 对 LOM 进行了形式化描述,为教育信息的共享提供了进一步研究的思路和指导。

关键词: 网格 教育资源 元数据

1 引言

随着教育信息化的发展,网络教育资源呈现出海量庞杂、分布松散、类型繁多、动态增长等特征,教育资源的管理存在着信息孤岛、搜索困难、共享困难等问题,而网格是把地理位置上分散的资源实现集成的一种基础设施,其出现为解决上述问题提供了一个有利的契机。

对于教育资源,广义上包括有学习对象、学习记录、学籍管理、教育资料等相关的信息,狭义上包括有教学课件、试题、教学资料、案例、文献及教学中涉及到的学习对象等^[1]。网格环境中存在各种动态的教育资源,它们在地理上分散,又可以动态地加入或离开不同的区域管理组织。如何使网格应用程序方便地使用各种资源是必须要解决的问题。由于教育信息化的标准众多,在网格环境下,许多不同的网格应用需要获取和共享相同的资源元数据。因此,为了实现网格环境下教育资源的共享,应有一种通用的语法对教育资源进行描述并建立相应的模型,这样在网格中便可方便地进行资源元数据的管理。

本文通过对一些教育资源的元数据标准进行分析,应用基于学习对象的元数据 LOM(Learning Object Metadata) 标准作为网格环境下教育资源描述实现的标准,应用网格资源描述框架 RDF(Resource Description Framework)形式,建立基于 LOM 的 RDF 模型,并对 LOM 进行形式化描述,为网

格环境下教育资源的统一描述提供了一个参考范例,并为课题的后续研究提供了依据。

2 教育资源元数据及其标准

元数据(Metadata)一词最早出现在 1988 年美国国家航空航天局(NASA)发布的《目录交换格式》(DFI)手册中,是为了解决网络资源无序化而提出来的。目前国际上关于元数据的定义有很多种。一般认为比较确切的说法是国际图联提出的“元数据是关于数据的数据”^[2]。它专门用来描述数据的特征和属性,其目的是提供比查看内容本身更多的有用信息。它使得用户可以根据关键字、所有者等已知的属性值来精确地定位和获取知识,学习对象元数据是学习对象技术的关键之一。

目前常用的学习对象元数据标准包括有:IMS 的 Learning Resource Metadata(学习资源元数据规范),IEEE 的 LTSC(Learning Technology Standards Committee,学习技术标准委员会)的 LOM(Learning Object Metadata:学习对象元数据)、OCLC(Online Computer Library Center)Dublin Core 的 DC 元数据标准等。其中,LTSC 是 1996 年 IEEE 成立的学习技术标准化委员会,致力于学习技术标准的制定工作。几年以来,LTSC 已经制定出几十个与学习技术相关的标准草案,它们涉及到学习者相关标准、学习环境相关标准、学习资源相关标准、平台

基金项目:安徽工程科技学院青年基金(2006YQ004)

收稿时间:2009-04-24

相关标准等^[3]。在本文中，我们应用基于学习对象的元数据标准 LOM 作为网格环境下教育资源描述实现的标准。

3 教育资源的基本结构

网格教育资源结构如图所示，共包括三大部分，分别为严格遵守的必需数据元素、作为参考的并对每类资源都适用的可选通用数据元素和针对资源特色属性的扩展数据元素^[4]。

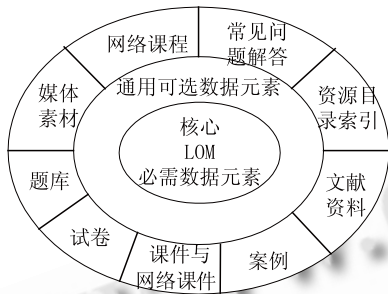


图 1 教育资源建设技术规范基本结构

必需数据元素(LOM 核心集)：这类数据元素与学习对象元数据规范(LOM)中的必需数据元素一致。它是任何类型资源都必须具备的属性标注。

可选数据元素(通用可选集) 这类数据元素是从学习对象元数据规范(LOM)的可选数据元素中抽取出了与教育资源密切相关、并对各类教育资源都适用的属性集合。可根据用户需求和开发者自身的工作过程作为参考属性有选择的使用，如果本规范没有推荐的属性取值，要求与学习对象元数据规范(LOM)的取值相一致。

扩展数据元素(分类扩展集) 这类数据元素根据九类资源(媒体素材、试题、试卷、课件、文献资料、案例、常见问题解答、资源目录索引和网络课程)各自的特点，从 LOM 模型的可选集中选取与某类资源密切相关的属性。并补充了一些基本的、必要的特殊资源分类属性。

4 教育资源元数据模型

4.1 基于 RDF 的教育资源描述方法

RDF 是万维网联盟 W3C 提出的一个描述资源的标准，用于表达资源的元数据。一个 RDF 文件包含多个资源描述，而一个资源描述是由多个语句构成，每个语句又是由资源、属性、属性值构成的三元组，这些三元组构成了 RDF 模型图。RDF 语句可以描述特定

资源或两个资源之间的关系。RDF 模型由三部分组成：Resource、Property 和 Statement。

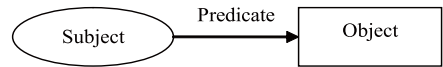


图 2 RDF 模型三元组示意图

三元组的形式为：{Predicate,Subject,Object }，图中椭圆表示资源，属性值以矩形表示，属性以箭头表示。

4.2 RDF Schema

RDF 通过类和属性来陈述资源，而这些特定的类和属性并没有在 RDF 中预定义这时就需要使用 RDF Schema 来定义所需的类和属性。这些类和属性构成了词汇表。

RDF 预定义的词汇表只能提供由 Resources，Properties 和 Statements 三种对象组成的基本模型。在进行资源描述时这些词汇是远不够用的。为此 RDF 定义了标准 Schema，在其中规定了描述类、属性及其语义的机制，使用户可以定制 RDF 词汇表。有了这些词汇表就可以对各种资源进行 RDF 描述^[5]。



图 3 教育资源 RDF Schema 层次模型

RDF Schema 中规定，类是任何具有 rdf:type 属性、并且该属性的值为 rdfs:Class 的资源。通过 RDFSchema 中的资源(rdfs:class, rdfs:Resource)和属性(rdf:type, rdfs:subclassOf)来表示。描述类的属性，是用 RDF 类 rdf:Property 以及 RDF Schema 属性 rdfs:domain (定义域), rdfs:range(值域), rdfs:subPropertyOf(子属性)等来描述的。

4.3 基于 LOM 的 RDF 元数据模型

LOM 规范所建立的数据模型是学习对象属性的一个最小集合，它只对学习对象的一些公用属性进行描述。而且对于某些数据元素，LOM 规范直接给出了推

荐的取值集合(称为词汇表),以获得更好的数据互换性。

LOM 规范为学习对象的属性描述提供了一个语义模型,用于支持信息的交换。但它并不涉及任何的实现问题,包括编码绑定、API 绑定、用户界面以及学习对象如何使用等等。这样做的目的在于:语义模型与绑定的分离,语义的互操作性。

为了有效地描述网格中的教育资源,我们把教育资源及其相应属性抽象为网格资源元数据,并采用 RDF Schema 描述资源的元数据集,构成描述资源的词汇表。

表 1 资源数据元素属性词汇表

字段名	RDF 属性	属性描述	数据类型
标题	Title	教育资源的名称	Char(1000)
目录项	Catalog_Entry	素材的唯一标识码; 入库时自动生成	Structure
语言	Language	教育资源所使用的语言	Char(100)
描述	Description	对教育资源内容的文本描述	Char(2000)
关键字	Keywords	描述教育资源的关键字	Char(1000)
版本	Version	教育资源的版本	Char(50)
贡献者	Contribute	教育资源的创建者	Structure
元数据方案	Metadata Scheme	用于创建元数据实例的规范的名称和版本	Char(30)
元数据语言	Metadata_Language	元数据实例所使用的语言	Char(100)
格式	Format	教育资源的技术格式	Char(500)
学习资源类型	Learning_Resource_Type	教育资源用于教育教学中的具体类型	Vocabulary
分类目的	Classification_Purpose	对教育资源进行学科分类	Vocabulary
分类描述	Classification_Description	该教育资源所属的学科	Char(2000)

4.4 基于 LOM 的 RDF 形式化描述

LOM 的 RDF 绑定中每个 LOM 的分类都拥有自己的命名空间,另外还定义了一个基本命名空间来包含公共的结构,下面给出了这些命名空间的缩写:

LOM={General,Lifecycle,Metametadata,Technical,Educational,Rights,Relation,Annotation,Classification}

其中各元素的定义如下:

通用信息:

General={Identifier,Title,Catalogentry,Lang

uage,Description,Keywords,Coverage,Structure,Aggregation}

描述性信息: Metadata={Identifier, Catalog entry, Contribute, Metadata schema, Language} 教育信息:

Education={Interactivity_type, Learning_resource_type, Interactivity_level, Semantic_density, Intended_enduser_role, Context, Typical_age_range, Difficulty, Typical_time, Description, Language }

价值信息: Rights={Cost, Copyright and the restrictions, Description}

分类信息: Classification={Purpose, Tax o path, Description, Keyword}

5 结语

作为下一代互联网的关键技术,网格技术可以消除信息孤岛和知识孤岛,实现信息资源和知识资源的共享,是目前对资源共享问题的较好解决方案。本文研究教育资源的元数据功能和标准,应用基于学习对象的元数据标准 LOM 作为网格环境下教育资源描述实现的标准。并对网格资源描述方法进行了比较,提出基于 LOM 的 RDF 模型,应用网格资源描述框架 RDF 形式对 LOM 进行了形式化描述,为网格环境下教育资源的统一描述研究提供有益的借鉴。进一步的工作是研究教育资源的感知模型和互操作模型、提高教育资源网络服务质量的方法、教育资源发现策略等。

参考文献

- 1 Foster I, Kesselman C. The Grid: Blueprint for a New Computing Infrastructure(Second Edition). Morgan Kaufmann, San Francisco, 2002.
- 2 Peng WH, Yang ZK, Liu QT, Zhao CL. Design and development of learning resource management system based on CELTS. Journal of Computational Information Systems, 2006,2(1):27 - 35.
- 3 McClelland, Marilyn. Metadata Standards for educational resources.Computer, 2003,36(11):107 - 109.
- 4 全国信息技术标准化技术委员会教育技术分技术委员会. <http://www.celtsc.edu.cn>
- 5 朱莹.基于分类元数据的网格资源描述和发现研究[硕士学位论文].南京:南京工业大学,2006.35 - 38.