

高等远程教育媒体素材库元数据初探

The Metadata Exploration about Media Material Database of Modern Distance Education

张玉芳 熊忠阳 吴中福

(重庆大学计算机学院 重庆400044)

Abstract For expediting the construction of education material of modern distance education, it is necessary to study the metadata about media material database. This paper first outlines the abroad researching and influencing metadata standard about the education material, then analyzes the characteristic of media material. From the resource construction point of view, analyzes advantage of repeated usable characteristic and the description information of metadata. Finally presents the metadata hierarchy model suited for the building of media material of modern distance education system. They are metadata essence set, metadata baseset and metadata detail set.

Keywords Distance education, Metadata, Media material database

1 引言

元数据是有关数据的数据,它是网络提供信息服务必不可少的工具。网络环境下信息的搜索离不开元数据的支持,目前较典型的元数据标准有:采用大型著书目录的专业化MARC;为了发现Web上资源而创建的IAFA/Whois⁺⁺, Dublin Core元数据标准;地理空间信息资源FGDC元数据标准等,这些元数据标准在初始目标及复杂性方面因应用领域的不同而存在很大的差异。

在现代远程教育系统实施过程中,随着学习对象和学习需求的增加,缺乏相关信息的元数据将制约我们发现、管理和使用这些对象的能力。为了解决这个问题,世界上很多标准化学术组织都正在致力于研究基于网络的教育资源标准化工作,并起草和制订了一些相应的规范,其中较有影响的网络教育资源规范是:在线计算机藏书中心的Dublin Core元数据标准^[1]、IMS的学习资源元数据规范^[2]、IEEE学习技术标准委员会^[3]制订的学习对象元数据模型LOM以及教育资源网关^[4]GEM等规范标准。

关于教育资源的元数据,国外的研究工作起步于九十年代,同时正处于设计、制订和完善之中,提供了相关的参考模型。在Web上由于使用元数据的应用和重点不同可能出现不同的标准,许多正在实施的计划都认为不可能出现一个通用的元数据标准^[6,7]。在我国现代远程教育资源库建设中,也不可能直接将这些标准作为我国资源建设的规范,有必要在仔细分析和研究现代远程教育资源内容的基础之上,研究和制订满足现代远程教育教学和学习需要的教育资源元数据规范。

2 目前已有的网上教育资源主要元数据标准

表1给出了目前国际上较有影响的网上教育资源元数据标准和建立的组织。

2.1 Dublin Core——电子资源内容描述模型

Dublin Core是发现电子资源的元数据元素集合,是Internet的重要组成部分。制订Dublin Core标准的最初目的是

为了规范作者对网上资源的描述,而其规范的描述又可以帮助他们发现网上资源。这一特性引起了博物馆、图书馆、政府及商业部门的关注,并渐渐地采用它作为资源形式描述的参考;与此同时Dublin Core也得到了很多国家的大力支持,现在已有网络数字图书馆以及其他各学科专家致力于研究Dublin Core的核心元素,有20多个国家参加(北美、欧洲、澳大利亚、亚洲等)Dublin Core标准的制订和研究工作。

表1 教育资源主要元数据标准

元数据标准名称	建立标准的组织
Dublin Core	Dublin Core 组织
IMS	美国高等教育协会
LOM	IEEE P1484组织
GEM	科教资源信息中心 ERIC

Dublin Core标准(简记为DC)由在线计算机藏书中心(Online Computer Library Center, OCLC)委员会组织编写及颁布,该标准几经修改,于1996年12月11日被正式确认为电子资源描述的参考规范,后来的元数据标准都是在它基础上,以它为参考模型发展起来的。风靡全球的数字图书馆通常采用DC作为资源描述标准,因此DC标准是构成数字图书馆基础结构的重要组成部分。目前DC标准聚集了图书馆界、网络和数字图书馆研究领域以及内容管理方面的专家和研究人员进行定期的研究讨论,其主要特点如下:

1)简单性:DC设计的初衷希望非专业编目人员能够像资源描述专家一样方便地使用,因此标准的很多元素其语义和图书分类卡片一样容易理解。

2)语义互操作性:在Internet环境下,由于学科领域的限制,不同的描述模型将妨碍资源的搜索能力,DC标准中满足一定规则并且容易理解的描述符号,不仅统一了数据内容的标准,而且增加了语义之间的互操作性。

3)一致性:Web空间的资源识别和发现基本上取决于有效的基础组织和设施,目前已经有20多个国家积极参与DC的建设,为其在国际范围内取得一致和统一奠定了基础。

4)可扩充性:DC中相关精心设计的模型描述不仅丰富

张玉芳 副教授,博士,主要从事远程教育和网络信息系统等领域的教学和研究工作。熊忠阳 副研究员,在职博士生,主要从事网络信息系统、数据仓库和并行技术等领域的研究和教学工作。吴中福 教授,博士生导师,主要从事网络、远程教育等领域的研究和教学工作。

了编译描述标准,而且使其在一定程度上进行扩充成为可能。

5) Web 上的元数据模块性:DC 支持一致性和独立维护性,同时支持 W3C 的资源描述框架的元数据包装。

6) 允许搜索引擎对标准域进行过滤,例如:过滤日期和作者信息。

目前 DC 标准已经被公认为是可行的网上资源的参考标准,其它后续标准要求与之兼容并能够与它进行互相转换。

2.2 IMS 元数据标准

教育系统 (Instructional Management System, IMS) 是美国高等教育协会的非盈利机构 EDUCOM (现在为 EDUCAUSE) 下的一个项目组。1997 年 IMS Project 着手于建立开放的、基于市场需要的在线学习资源标准,其中包括学习内容的元数据规范。在同一年国家标准技术协会 (National Institute for Standards and Technology, NIST) 和 IEEE P. 1484 研讨组 (现称为 IEEE LTSC——学习技术标准委员会) 开始了同样的努力。NIST 和 IMS 努力的结果,在 1998 年 IMS 和 ARIADNE (一个欧洲计划项目) 共同向 IEEE 提交了关于教育资源的元数据规范和建议,从而形成了 LOM (Learning Object Metadata, 学习对象元数据) 基本原型。

2.3 LOM 元数据标准

ISO/TC211 的 LOM 元数据标准由国际标准化组织第三工作组组织研究,项目编号为 15046-15。1996 年 2 月 9 日通过 1.0 版草案,后几经修改,于 1997 年 1 月 20 日发布 2.0 版标准 (ISO/TC211, 1997), 最新的版本^[2] (Draft 6.1) 发布于 2001 年 2 月 13 日。

IEEE LOM 基本文档定义了描述学习资源的元数据元素,它们包括:名字、定义、数据类型、字段长度;规范同样还定义了元数据的概念结构,即:元数据文档如何在语句上一致并且在行为规范上与 IEEE 一致。IEEE LOM 基本文档试图在多种实现上支持一致的元数据定义并支持它们之间的交换,但目前并不包括如何将其表示成机器可读的格式。LOM 元数据结构是一个拥有目录、数据元素和抽象数据类型的三层结构模型,顶层一级是目录,每一个目录包含中间级的数据元素,然后由抽象数据类型定义具体内容。目前 LOM 元数据拥有 9 个目录对象,分别具有多个属性元素。

LOM 标准重点在于提供方便学习对象 (远程教育中涉及到的各种对象) 的管理、定位以及评估的最小属性集合,通过数据元素和实体类型局部扩充其基本属性,其属性可能的取值有五种:强制型、必须型、可选型、条件型、不允许型。学习对象的相关属性描述包括对象类型、作者、拥有者、分布术语以及格式;如果可能的话,学习对象元数据还将涉及教学方法、教学风格或交互方式、先决条件、熟练级别及学习级别,因此对于任何学习对象可能存在多个 LOM 集合。

LOM 目的在于达到以下几个功效:

- 1) 方便学员和教员搜索、评估、获取和使用学习对象。
- 2) 通过支持学习系统的技术达到学习对象的共享和交换。
- 3) 允许学习对象以可能的方式进行合理的分解与合成。
- 4) 允许计算机代理为不同的学员自动或动态地构建个性化的学习内容。
- 5) 在开放分布式学习环境中允许多个学习对象直接联系。
- 6) 为研究者提供与各类选择数据的共享。
- 7) 定义一个简单的易于扩充的域标准以便更广范围的采

纳和应用。

2.4 GEM 元数据标准

教育资源网关 (Gateway to Education Material: GEM) 是在 DC 基础上进行扩充的,它提供快速和方便地访问教育资源的能力。GEM 由美国教育部主办,是科教资源信息中心 (Educational Resources Information Center: ERIC) 在信息技术方面的一项特殊工程项目,任何人都可以通过 GEM 搜索或浏览分布于 200 个 GEM 联盟成员网站上的高质量的多学科教育资源,这些资源包括社区学院、成人继续教育、职业教育、高等教育,具体围绕课程计划、项目等进行,同时 GEM 提供部分免费的学习资源。

2.5 其它元数据标准

除了上述几个标准外,很多国家都建立了满足自己国家需要的教育资源元数据,它们大多以 DC 标准的核心元素为参考模型,在其基础之上根据具体应用的需要进行了扩充和推广,因此每种元数据标准都有自己的特色,并且在处理某些问题中占有一定的优势。

DC 标准的优势在于方便资源在网上的描述,但是它没有考虑教学方面的特性;IEEE LOM 模型庞大而复杂;GEM 在 DC 的基础上虽然增加了 8 个核心元素,但是它不是为了媒体素材的可重复使用特性而设立的,因此我们无法直接使用已有的教育资源元数据标准,必须制订一个满足现代远程教育媒体素材库需要的元数据标准。具体是仔细分析媒体素材资源的特点,参考 DC 标准,借鉴 IEEE LOM 模型和 GEM,设计我国现代远程教育系统需要的媒体素材资源元数据模型。

3 现代高等远程教育媒体素材库元数据模型

3.1 现代高等远程教育资源库组成及媒体素材特点

现代高等教育资源库^[9]包括媒体素材库 (文本、图形、图像、音频、动画及视频)、课件库、题库、案例库和网络课程库。在教学资源库中,网络课程库、课件库、案例库可以是事先开发好的比较完整或自成体系的资源,它们可以直接提取使用;而媒体素材库中的内容则可根据一定的原则由相关的各类媒体组建成满足不同层次、不同人员需要的灵活多样的教学内容,本文讨论的重点是媒体素材库。

随着学习资源的急剧增加,如何使学习资源的开发和设计成为一项增值工作是人们日渐关心的问题,这也是现代远程教育发展到一定阶段面临的新问题。目前很多学习资源的设计都是针对某个特定目的,例如为了某门课程,不是为了扩充媒体素材内容进行设计的。我们希望把媒体素材设计成能够重复使用的学习对象,然后根据这些媒体素材构造出各种可能的学习内容。即媒体素材的设计不能仅仅满足某门课程的需要,应该能够融入其它上下文应用中,根据一定的方法组建出满足各种需要的学习内容,这样既可以按照一定的方法设计出各种可能的学习内容又可以不断积累优秀素材,丰富教学资源,加快远程教育建设进程,降低资源建设成本。

总之,现代远程教育资源库中媒体素材设计的目的是为了能够构造出各种可能的学习内容,称之为可重复使用学习对象,亦即各类媒体素材具备可重复使用的特点,这种可重复使用特点在远程教育系统中具有以下五大优势:

1) 灵活性:如果资源是可以重复使用的,即可以在多种上下文中使用,那么资源建设的主要工作不是对每个新内容重新进行撰写,而是设计相对独立的媒体素材知识点,组织这些

媒体素材资源来满足各类学习者的需要。灵活性将大大减少资源设计和开发过程的时间。

2)方便修改、搜索和管理:元数据提供的过滤和选择能力,可以对相关内容进行快速搜索和修改操作,从而实现内容管理。

3)互操作性:可重用学习对象思想允许根据设计、开发、以及对象的表示需要来设置说明,同时保留与其它学习系统和上下文之间的互操作性。

4)便于能力学习:能力学习的重点是能力模型的交叉技术、知识、属性等内容,而不是课程模型,这是教育所追求的最佳境界。制约能力学习的主要因素是缺乏相应的模块,而可重用学习对象思想则能满足这种需求,通过一定的内容组建方法,可以构造出具备一定适应能力的学习内容。

5)增加内容的价值:从商业观点来看,内容的价值与重复使用的次数成正比。

媒体素材资源的设计和开发必须提供双重能力的支持,既可以直接投入使用,也可以在将来重复使用。此重复使用能力将媒体素材作为课程的一部分并在希望的粒度级上成为一个相对独立的信息或知识点,这正是现代远程教育媒体素材建设的初衷。为使我国现代远程教育系统的媒体素材库资源早日投入高效率的实际使用,为保证资源的质量和可扩充性,对媒体素材进行研究就显得非常有必要,它的研究在一定程度上对现代远程教育资源建设起着指导和促进作用。

为保证设计的媒体素材满足可重复使用特点,描述可重复使用特点的媒体素材的元数据应该包括下述五点内容:

1)和主题相容的、规范的术语。相容的术语允许离散的媒体素材进行分解和重新组装,规范的术语是媒体素材可以被重复利用的前提。

2)信息的表述以容易访问和容易理解的格式,以尽可能简洁的形式对媒体素材进行最大限度的说明。

3)媒体素材内容的建设需要满足 Web 内容设计的标准技术,例如密集的内容需要被削减成较小的单元,从而方便在 Internet 环境下的高效访问使用。

4)对象间信息的非时序性:各类媒体素材应该自由、方便地适应于多种上下文应用中。在设计媒体素材时,为避免在信息重组时发生不必要的链接,媒体素材之间不要出现前后互相参考。

5)媒体素材的编辑应该保持一致的风格。

3.2 媒体素材库元数据模型

现代远程教育系统媒体素材资源存储和管理大量的多媒体教学资源,这些资源所具备的高效管理、检索和查询、动态更新以及扩展性问题,都是在设计资源库结构时必须考虑的问题。制订统一的远程教育媒体素材资源元数据,有利于各类教育信息系统和教学共享系统按照一定的规范开发,真正发挥网络没有边界的特色。

在分析和领会 DC、LOM 标准及 GEM 的过程中,本文提出一种满足中国现代远程教育系统资源库中媒体素材库建设需要的、容易被多数用户接受和使用的多层次元数据模型,不同的用户只用到不同的层次。在具体设计和使用时,考虑到数据用户对数据要求的层次性和已有元数据内容标准的复杂性,作者认为现代远程教育系统媒体素材库的元数据应该针对不同数据用户的服务需要而设计成不同层次上的。由于元数据的可抽象性、简洁性等特点要求,这种结构是可以实现的。现代远程教育系统媒体素材库的元数据分为三个层次(见

图1)。

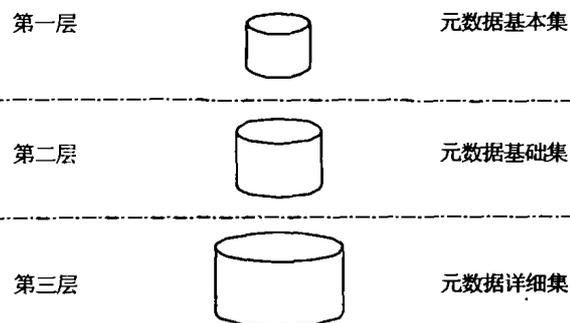


图1 现代远程教育系统媒体素材库元数据模型

1)第一层:元数据基本集,其服务对象为现代远程教育系统的一般使用者(学员)了解和检索数据的集合,它描述了数据集最基本、一般数据用户最想了解的特征。

2)第二层:元数据基础集,服务对象为有一定知识背景的非数据处理和管理目的的数据用户(教师),它描述了数据集较详细的特征,是对第一层基本集中元数据内容的进一步具体化。

3)第三层:元数据详细集,详细列出了描述数据集内容的各种要素项,服务对象是专业数据处理和管理人员,他们利用元数据详细集操作、维护和控制媒体素材库内容。

现代远程教育系统媒体素材库三个层次的元数据之间存在着密切的联系,对于元数据制作者来说需要创建包含最广泛的元数据详细集(第三层管理层级),第一层基本集和第二层基础集的元数据分别是元数据详细集的子集,它们可以从元数据详细集派生而出。根据元数据的可抽象和归并特征由系统或专业软件自动形成上层元数据。例如:根据基础集中有关教学内容的特征和属性可以得出基本集中所需要的信息;而从下级元数据中抽取有典型意义的元数据项可以作为上层元数据集中对应项的内容。

由于我们重点是在实施现代远程教育系统环境下,如何维护、发布、搜索和检索媒体资源素材,即不仅方便学习者使用的需要,而且还必须方便管理人员维护系统的需要,因此我们讨论的是现代远程教育系统资源建设中媒体素材库第三层元数据详细集。一旦拥有包罗万象的元数据详细集的设计,其元数据基本集和基础集的元素可以通过一定的操作获取。

3.3 媒体素材元数据详细集内容及其含义

现代远程教育系统媒体素材库元数据详细集的组成是一个拥有目录、数据元素和抽象数据类型的三层结构模型。其具体表现为:目录存在于顶层一级,而每一个目录包含中间级的数据元素,最后由抽象数据类型定义。我们设计的现代远程教育系统媒体素材库元数据详细集内容包括9大类目录,这些目录又分别具有多个元素,而每一类元素又拥有多个属性,其属性类型有两类:一般数据库支持的简单属性和系统根据需要自定义的复杂属性,属性的取值可以是必须(mandatory)、可选择(optional)、条件(conditional)。元数据详细集的9个目录元素的示意图如图2所示。

现代远程教育系统媒体素材库元数据详细集标准的9个目录对象可以很方便地实现与 DC 标准的互换。由于篇幅关系,此处略去媒体素材库详细集元素的具体内容。

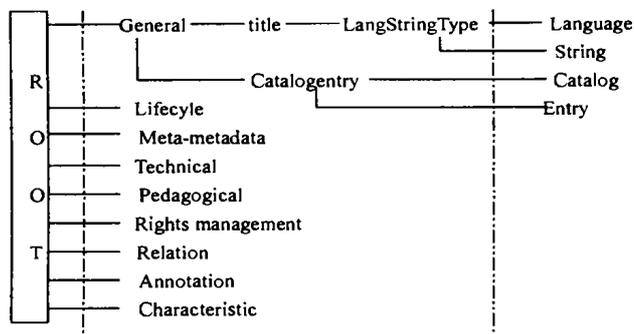


图2 媒体素材库元数据详细集元素示意图

小结 本文在分析了已有的教育资源的元数据基础上,仔细分析了现代远程教育系统中媒体素材的特点,针对远程教育系统中用户的不同,提出了媒体素材库的层次型元数据模型。现代远程教育媒体素材库元数据标准规范和统一了远程教育系统中媒体素材的描述,保证了使用媒体素材的一致性和简洁性,为远程教育环境下由媒体素材组建个性化的

(上接第152页)

体上表达某项问题,就使人相信问题已经得到了解决,但实际上将问题传给了下一个阶段。

尽管 UML 太概括了以至它不能成为表达设计的自然选择,但在 SDL 设计中仍然可以从 UML 中获得益处。利用 UML 紧凑的符号提供 SDL 架构的概括综述。UML 中的图非常适合描述 SDL 包之间的关系,而 SDL 却没有这种图示,因此可以借助 UML 来表示。SDL 类型之间的继承层次关系也可以图示化,在继承层次中 SDL 类型表示成 UML 中的类。

结束语 UML 和 SDL 结合起来使用非常有效,比单独使用能够提供更好的开发支持。现在的 UML/SDL 工具通常分解成两个独立的部分,一个是基于 UML 用于面向对象的分析,另外一个基于 SDL 用于设计与实施,然后提供两模型间的链接与转换。这是当今 SDL 工具最通常的结构,能够给用户提供一个良好的解决方案,充分利用 UML 和 SDL 两种语言的优势。

然而总是用两种以上的工具或语言,必然增加开销,因为在两种不同的工具或语言之间需要切换。因此将 UML 和 SDL 合并在一起这是当今的趋势。合并 SDL 和 UML 工具是可能的,能够使得 UML 和 SDL 结合使用更好方便平滑。从 UML 开发者来看,他能够得到由 SDL 带来的便利,如在 UML 工具中他也可以仿真、形式化验证,完全的应用生成和大规模分布系统的直接定义;从 SDL 开发者来说他可以在他的 SDL 工具中直接使用强有力的面向对象的分析工具,他还

学习内容奠定了基础。本文提出的媒体素材库元数据模型仅是一种新的尝试,希望对远程教育资源建设的规范化有一定的指导和促进作用。

参考文献

- 1 The Dublin Core element set. <http://WWW.purl.org/DC/about/element-set.htm>
- 2 The IEEE Learning Object Meta-data (LOM) base document. <http://ltsc.ieee.org/doc/wg12/scheme.html>
- 3 The Instructional Management System Project (IMS) website. <http://www.imsproject.org>
- 4 <http://www.fgdc.gov/metadata/metadata.html>
- 5 Gateway of Education Material (GEM) website. <http://www.gem.org>
- 6 英国 UKOLN 网站. <http://www.ukoln.ac.uk/>
- 7 澳大利亚 EDNA 元数据网站. <http://standards.edna.edu.au>
- 8 教育部高等教育教育资源规划可行性报告,1999
- 9 教育部现代远程教育工程教育资源建设规范,1999

能从其它可以利用的 UML 模型中获得益处;从研发单位来看 UML 和 SDL 合并也是非常重要的,因为这可以减少用户使用 SDL 和 UML 的培训费用,可以相互重用,工程师可以方便地从基于 SDL 的开发转到基于 UML 的开发,可以大大地减少因此而引起的投资。而且当不同的符号统一规范了之后,由分析团队建立的分析模型就能很方便地由设计团队来使用。

我们可以得出这样的结论,将非形式化的 UML 世界和 SDL 的形式化的语义相结合在各类系统的开发中带来很多便利。我们认为有必要将它们集成在一起,建立性能更为优越的建模设计环境,以便更加适应复杂分布式实时系统的大规模的工业化分析与设计。

参考文献

- 1 OMG,UML1.3 June1999
- 2 ITU-T Recommendation Z.100.Specification and Description Language(SDL)Nov.1999
- 3 bourhfir C,Aboulhamid E. Test cases selection from SDL specifications,Computer Network. 2001. 693~708
- 4 ITU-T Recommendation Z.120:Message Sequence Chart(MSC), Nov.1999
- 5 魏定国,吴时霖. 基于 UML 和 Petri 网的用户界面原型的研究. 计算机科学,2001,12
- 6 Engels G,Heckel R,Saucer S. UML -A universal Modeling Language? ICATPN2000,LNCS,2000. 24~38