

● 虞惠达 朱庆华

## 元数据在教育资源发现中的应用\*

**摘要** 元数据在资源发现中起一种间接作用。元数据应用的直接作用是网络资源的结构合理、层次分明的组织,而结构化的组织直接促进了资源的发现和利用。元数据在中文教育资源发现应用中应注意元数据的信任问题、互操作问题、描述粒度问题、元素扩展问题。图2。参考文献14。

**关键词** 教育资源 资源发现 元数据 中文元数据

**分类号** G254

**ABSTRACT** Metadata play an indirect role in resource discovery. The direct effect of the application of metadata is the rational and hierarchical organization of network resources, and the structured organization directly promotes the discovery and utilization of resources. We should pay special attention to the trustfulness, interoperability, granularity and element extendibility in the discovery and application of Chinese educational resources. 2 figs. 14 refs.

**KEY WORDS** Educational resources. Resource discovery. Metadata. Chinese metadata.

**CLASS NUMBER** G254

元数据按功能可以划分为管理型元数据、描述型元数据、保存型元数据、技术型元数据和使用型元数据5种类型<sup>[1]</sup>。在网络资源组织方面,元数据主要有5个作用:描述、定位、发现、评估和选择<sup>[2]</sup>。本文以教育资源元数据为对象,探讨元数据在资源发现中的原理,介绍元数据在教育资源发现中的作用和应用实例,并对中文元数据在教育资源发现应用中注意的问题进行分析。

### 1 元数据在资源发现中的原理和作用

#### 1.1 元数据在资源发现中的原理

资源发现是元数据最大的一个应用领域,它是解决目前网络资源缺乏组织从而导致资源发现困境的治本方案。元数据通过使用规范的描述符,对网络资源进行准确和一致的描述,形成优化结构,从而趋向整个网络资源的自组织。元数据在资源发现中起一种间接作用,元数据应用的直接作用是网络资源的结构合理、层次分明的组织,而结构化的组织直接促进了资源的发现和利用。以往主题指南和搜索引擎是从自身的角度去解决混沌的网络中用户的迷路问题,它们的中心问题是用户,而元数据问题的中

心是混沌的网络,它的目标是改变网络的无序。创建一个有序的网络是元数据的终极目标,而此时资源发现问题迎刃而解。

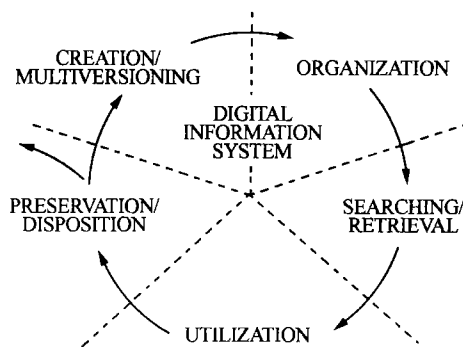


图1 信息对象的生命周期模型

元数据应用于资源发现的机制可以通过信息对象的生命周期模型来进一步说明<sup>[3]</sup>。一个信息对象的生命周期一般有5个阶段:创建或版本转换、组织、搜寻和检索、利用、保存或处置(见图1)<sup>[4]</sup>。如果将整个因特网看做一个数字信息系统,则网络上的各类资源可以看做一个一个信息对象,而元数据作用于

\* 本文为科技部基础条件平台工作重点项目“数字图书馆标准语规范建设”(2003DEA4T035)子课题“基本数字对象描述元数据规范”的研究成果之一。

整个生命周期的两个阶段,一个是信息对象的创建或版本转换阶段,一个是信息对象的组织阶段。根据元数据在信息对象中生成的不同阶段可分为两类:一类是内部参考类(internally-referenced)元数据,一类是外部参考类(externally-referenced)元数据<sup>[5]</sup>。内部参考类元数据是信息对象的创建者或维护者选择的描述对象的元数据。这类元数据一般通过嵌入方式作为内容结构的一部分或者给出参考 URI。外部参考类元数据由信息对象的创建者和维护者之外的机构创建和维护。这类元数据可能对象的创建者根本察觉不到,但却可能大量存在这些外部参考类元数据。

在元数据作用于信息对象的生命周期后,生命周期 5 阶段模型中第 1、第 2 阶段开始模糊,若元数据在信息对象生命周期的第 1 阶段生成,即内部参考类元数据,信息对象在创建或版本转换过程中本身具有某种程度的组织,或者说信息对象生命周期直接从第 1 阶段跳跃至第 3 阶段,我们将原有模型的第 1、第 2 阶段整合成一个阶段,创建或版本转换与组织阶段,生成信息对象的生命周期改进模型(见图 2)。

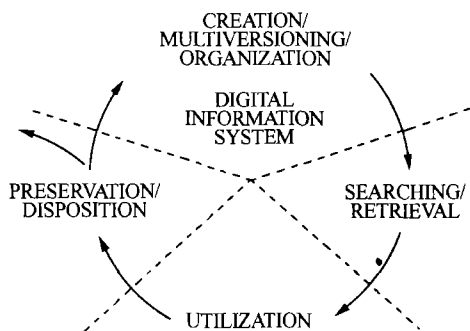


图 2 信息对象的生命周期改进模型

### 1.2 元数据在教育资源发现中的作用

教育资源是在学习过程中被使用或参考的一个实体<sup>[6]</sup>。教育资源包括用于教学过程及其评价、管理的教材、课件、辅导资料、参考资料、习题集试题集、工具型资料、教学标准与过程规范等<sup>[7]</sup>。因特网是当今虚拟和分布式学习系统的普遍的支持环境,各种教育系统和资源纷纷出现,然而目前的用关键词进行资源发现的搜索引擎技术或者基于分类法的主题指南方式在帮助人们准确定位、发现资源方面仍然难以令人满意。

在教育资源中创建元数据将促进教育资源的发

现,教育元数据将通过准确描述教育资源,定位和区分不同类型资源,依据相关性准则发现并给出定位信息。教育资源的特殊性使元数据的应用大大提高了其资源发现的能力。

教育资源一般都面向一定的使用对象,比如包括教师、学生、作者和管理人员等,而学生又包括小学生、中学生、大学生等。如果用户要查找适合于小学六年级学生使用的教育资源,搜索引擎将无法构造简单的关键词,而元数据却通过提供对教育资源的受众的描述使这类资源的发现能力大大提高。

教育资源中存在着大量的非文本资源。传统的搜索引擎和主题指南无法对视频、音频等多媒体教育资源进行标引和发现,而元数据通过对资源类型的描述,只要用户在资源检索时对资源类型进行限定即可发现它们。

许多资源受知识产权保护,不会全文公开,而搜索引擎的标引软件只有在对资源有完全权限时才会对其标引,这会造成许多存在访问限制的教育资源不被发现,从而影响资源的查全率。访问元数据不需要对资源内容本身访问,而是资源的替代内容,这个替代内容又能准确描述资源本身的内容,如果用户觉得资源重要,元数据会提供获取资源本身的信息。

教育资源的交互性特别强。教学过程中存在着教师与学生的交互,教师与教师的交互,在全球化的背景下,这种交互已经突破了地域的限制。因特网上存在大量教育资源,这些资源的质量参差不齐,传统的搜索引擎和主题指南没法提供一个质量评估机制。而这正是教育元数据的一个特征,它通过提供教育资源使用者的评估来反映不同资源之间的质量差异,有助于用户发现高质量的教育资源。

## 2 元数据在教育资源发现中的应用实例

### 2.1 学习对象元数据(LOM)标准

LOM 是 IEEE 学习技术标准委员会在 1998 年推出的一个描述学习对象的元数据标准,是在 IMS 和 ARIADNE 提交给 IEEE 的提议基础上生成的标准。LOM 是一个详细说明学习对象元数据的多部分标准。在 LOM 中,学习对象被定义成了可用于学习、教育或训练的任何实体——数字的或非数字的。一个学习对象的元数据实例描述它所应用的学习对象的相关特性。

LOM 说明了一个用于定义学习对象元数据实例

结构的、概念性的数据范式。概念性数据范式允许学习对象和用于描述这些学习对象的元数据实例存在语言差异,它详细说明了组成一个学习对象的元数据实例的数据元素,并能够被其他定义数据范式实现算法的标准所参考。一个学习对象的元数据实例可以被学习技术系统用于管理、定位、评价或交换学习对象。

LOM详细说明了学习对象元数据的语法和语义,定义了学习对象的各种属性,包括元素名、定义、数据类型、类属、词汇和字段长度。LOM的数据模型是一个数据元素的层次结构,包括了聚合数据元素和简单数据元素(层次结构的叶结点)。在LOM1.0的基本范式中,只有叶结点拥有通过相关值域和数据类型定义的值,而聚合数据元素没有值。因此,聚合数据元素没有值域或数据类型。LOM数据模型总共定义了60个元素,数据模型的最上层定义了9个基本类目。

## 2.2 教育管理系统(IMS)元数据标准

IMS项目是由EDUCAUSE发起的,它是一个北美教育机构和他们的工业伙伴成立的一个协会,旨在为分布式的学习程序和服务定义一个互操作的技术标准。他们的主要工作是定义分布式学习系统的体系结构和参考模型。IMS意识到在学习资源标准化的进程中对教育资源的元数据达成一致是目前的首要任务。从1998年开始,IMS和ARIADNE一起制定了LOM,而IMS也在定期地发展自己的标准。IMS认为LOM定义的元素太多了,许多IMS内的组织建议使用一个精简的基本元素来描述教育资源,所以IMS通过提供两套不同的元素集来使LOM元数据实施更灵活。一套是IMS核心(IMS Core),是描述基本元数据的简化集,共19个LOM元素;另一套是IMS标准扩展库(IMS-SEL),它包括LOM元数据的其他元素。

IMS是在调查了协会内大量的教育机构、培训组织和软件提供商后确定IMS核心元素集的,这些团体致力于将学习资源元数据嵌入他们的软件产品。IMS调查这些机构和组织在采用LOM元素时的相对重要性,并根据调查结果提出了19个元素的IMS核心元素集。

## 2.3 澳大利亚教育网络元数据标准

澳大利亚教育网络(EdNA)是用于在因特网上教育和培训方面进行协作的国家网络框架<sup>[8]</sup>。EdNA元数据标准基于都柏林核心元数据元素集(DCMES),与澳大利亚政府定位服务(AGLS)相一致。制定

EdNA元数据标准的目的是,为了在在线资源发现和管理领域,支持澳大利亚教育培训各部门的协同工作。采用这个标准将帮助在教育培训领域的人们生成和使用充分描述的数字式的内容。它还将满足这些具有良好结构编码内容的技术需要,从而来改变和服务于需求的数据。这个标准目前主要应用是为EdNA在线的各州各地区的的所有教育培训部门加强教育资源的聚合。

EdNA元数据标准承诺,澳大利亚教育培训系统内部的特殊团体将更进一步扩展和限定元素集,来满足当地的需要。EdNA元素集共有23个元素,分成两类,其中15个元素完全取自DCMES,而其他8个元素是专门为了EdNA在线资源管理而设立的。

## 3 教育资源发现领域中文元数据的探讨

### 3.1 国内教育资源元数据开发的现状

我国从1999年开始着手远程教育技术的标准化工作。2001年,我国启动了中国教育信息化技术标准(CELTS: China E-Learning Technology Standards, 原称DLTS)研究项目。目前,主要是引进国外标准并进行本地化工作。中国教育信息化技术标准委员会先后制定了学习对象元数据规范、教育资源建设规范和基础教育教学资源元数据应用规范。

学习对象元数据规范是在参考IEEE LOM的基础上,考虑在教育资源中生成元数据的成本和教育资源自身的特点后,选取了LOM中的部分元素。它是LOM的一个子集<sup>[9]</sup>。它将描述教育资源的元素分为必须数据元素和可选数据元素。必须数据元素是在数据结构中定义,并且必须在数据结构的实例中出现的数据元素;而可选数据元素是在数据结构中定义,但不一定要求在数据结构的实例中出现的数据元素。

教育资源建设规范是具体的资源属性标准<sup>[10]</sup>,它是有很强的实践指导意义的行业专用标准。它将描述教育资源的元素分为必须数据元素、可选数据元素和分类数据元素。必须数据元素与学习对象元数据规范中的必须数据元素是一致的,是任何类型的资源都必须具备的属性标注。可选数据元素是从学习对象元数据规范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。分类数据元素的分类扩展是根据9类资源各自的特点,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。

基础教育教学资源元数据应用规范是在学习对象元数据规范的基础上,结合我国基础教育的实际,定义的一组面向基础教育的教学资源元数据元素<sup>[11]</sup>。它规定的描述基础教育资源的数据元素集包括23个元素,其中必须元素11个,可选元素12个。它的必须数据元素与学习对象元数据规范的全部必须元素(不含子元素)对应;而可选数据元素包含了学习对象元数据规范的11个可选元素。

### 3.2 元数据在中文教育资源发现应用中注意的问题

我国中文教育资源元数据的开发,在参考国际教育资源领域的基础上,已经形成了一套自己的元数据标准规范集,给中文教育资源的组织、发现和利用提供了一个指导框架。然而从全球范围来看,教育资源元数据还停留在研究和开发阶段,没有完全进入应用领域,只有个别项目在应用进行探索。下面提出一些元数据在中文教育资源发现应用中注意的问题,以求进一步探讨。

#### (1) 教育资源元数据的信任问题。

元数据应用领域的现状不是元数据标准的缺乏或者元数据生成工具的缺乏,而是内容创建者在创建资源时创建元数据的太少。这个问题的直接原因是目前的主要搜索引擎不采用元数据作为在资源发现时的标引依据<sup>[12]</sup>,其根源是信任问题,是搜索引擎对元数据创建者缺乏信任。只有元数据被主要搜索引擎充分采用,资源发现能力才会得到根本性提高。这种信任不能通过传统的司法机构强制实现,目前主要是通过数字签名来保证内容描述的真实性,只有数字签名的广泛采用才会最终让元数据应用到资源发现领域。

在教育资源发现领域,美国的GEM在元数据信任问题上给出了自己的解决方案<sup>[13]</sup>。GEM是美国教育部下面的一个项目,它在自己协会成员中利用元数据组织教育资源和改善资源发现能力,从而保证了描述教育资源的元数据的真实性。国内在实施教育资源元数据规范时信任问题当优先考虑。

#### (2) 教育资源元数据的互操作问题。

我国教育资源元数据的互操作有两层涵义。一层是中文教育资源元数据与国际上教育资源元数据的互操作。基于因特网的教育资源发现是在全球背景下进行的,在本土化与国际化之间要有完善的桥接机制,实现全球范围内的教育资源发现。另一层是教育资源元数据与其他应用领域元数据的互操

作。中文元数据的开发过程与国外有某种相似,都是先在应用领域开发。国内目前中文基本元数据规范还在开发之中,而其他某些领域的元数据规范已经出现。教育资源元数据与其他元数据规范之间的互操作需要国内教育资源发现界给予足够重视。

#### (3) 教育资源元数据描述粒度问题。

粒度问题是元数据研究领域一个热点问题。它是指对资源描述详细程度的一种衡量。在资源发现领域,用户一般在本领域内查找资源,而有时需要跨领域查找。用户在本领域与其他领域的知识范围不同,所以至少需要两套资源描述粒度的元数据集,这是一种面向用户的资源发现元数据方案,在资源发现中体现了一种人性化关怀,如美国的全球定位服务(GILS)正是基于这种关怀而设计的<sup>[14]</sup>。在教育资源领域,一个学科的专家在另一个学科可能是外行,而当今学科与学科之间的交融,交叉学科、边缘学科的出现,使跨学科资源发现成为一个很大的需求,教育资源发现的面向用户的关怀更加迫切,在教育资源元数据框架内提供不同粒度描述的资源发现将成为一个新的研究热点。

#### (4) 教育资源元数据元素扩展问题。

任何一个标准都不可能一成不变,许多元数据标准在公布之初都会建立一套扩展机制,如DC元数据项目在评估新增元素时有一个决策树机制,新增元素需通过决策树的一套规则审核后决定是否扩展原有元素集。教育资源类型多样,目前国内将教育资源分为九大类型,而将来可能会出现新的表现形式和新的描述需求,所以需要一套完整的处理教育资源扩展的程序,成立专门的机构,负责接受扩展申请、审核、公布和实施,及时对新出现的资源进行描述和发现。

### 参考文献

- 1 Gilliland-Swetland, Anne J. Defining metadata. In: Baca, Murtha, ed. *Introduction to metadata: pathways to digital information*. U.S.A.: Getty Information Institute, 1998
- 2 刘嘉. 元数据导论. 北京: 华艺出版社, 2002
- 3,4 Anne J. Gilliland-Swetland. *Setting the Stage*. Accessed 13/03/2004. <http://www.getty.edu/research/conducting-research/standards/intrometadata/2-articles/index.html>
- 5 Stuart L. Weibel, Carl Lagoze. An element set to support resource discovery: The state of the Dublin Core: January 1997. *International Journal on Digital Libraries*. 1997 Volume 1, Number 2. (下转第100页)

# 四川大学图书馆

四川大学图书馆始建于1896年,是中国西南地区藏书规模最大的大学图书馆。图书馆由文理馆、工学馆、医学馆和江安馆四个分馆组成,馆舍总面积62800平方米;馆藏纸质文献519万册,并拥有丰富的电子信息资源;专用阅览室和多媒体阅览室现共有阅览座位4000多个。目前正在建设功能强大、资源丰富、全面开放的数字图书馆门户网站。

四川大学图书馆是教育部“CALIS西南地区中心”、“CASHL西南区域中心”、“文科文献信息中心”,联合国教科文组织“英文出版物收藏中心”,卫生部“西南地区医学文献中心”,教育部“科技项目查新工作站”和卫生部“医药卫生科技项目查新站”。此外,四川省高校图书情报工作委员会秘书处、四川省图书馆学会文献检索委员会、医学专业委员会等也设在本馆。



(上接第74页)

- 6 Luis E. Anido, etc. Educational metadata and brokerage for learning resources. *Computers & Education* 38 (2002)
- 7 张晓林.元数据研究与应用.北京:北京图书馆出版社,2002
- 8 The EdNA Metadata Standard. Accessed 19/03/2004. <http://www.edna.edu.au/edna/go/pid/385>
- 9 Xin Xiang, etc. Introduction of the Core Elements Set in Localized LOM Model. Accessed 20/03/2004. <http://www.celtsc.edu.cn/jiedu/3.htm>
- 10 教育资源建设规范 CELTS-41.1(CD1.0). Accessed 20/03/2004. <http://www.celtsc.edu.cn/document.htm>
- 11 基础教育教学资源元数据应用规范 CELTS-42.1

(CD1.6). Accessed 20/03/2004. <http://www.celtsc.edu.cn/document.htm>

- 12 Tony Gill. Metadata and the World Wide Web. Accessed 13/03/2004. <http://www.getty.edu/research/conducting-research/standards/intrometadata/2-articles/gill/index.html>
- 13 GEM Project. About GEM. Accessed 22/03/2004. <http://www.geminfo.org/networker.html>
- 14 Eliot Christian. A metadata initiative for global information discovery. *Government Information Quarterly*, 2001(18)

虞惠达 中国科学院文献情报中心04级硕士研究生。  
通信地址:北京。邮编100080。

朱庆华 南京大学信息管理系教授,博士生导师。  
通信地址:南京。邮编210093。(来稿时间:2004-07-07)