

面向多粒度层级学习对象的元数据研究与设计*

□ 张东 朱礼军 张运良 乔晓东 / 中国科学技术信息研究所 北京 100038
肖峰 / 中国民生银行培训学院 北京 100084

摘要: 文章针对当前企业e-Learning和知识管理系统中存在的可重用性和自主路径设计的需求, 设计了适用于多层粒度学习对象的元数据应用框架, 研究并解决了元数据方案设计中的关键问题, 包括知识模块的内容模型和语义结构模型。

关键词: 学习对象, 元数据, 粒度, 知识模块, 内容模型, 知识管理

DOI: 10.3772/j.issn.1673-2286.2012.04.009

1 前言

学习对象的可重用性是当前企业e-Learning和知识管理系统研究的核心问题, 可重用性与学习对象的大小(即粒度)有直接的关系, 粒度越低, 潜在的可重用性就越高, 反之亦然^[1]。综观各类e-Learning网站, 学习资源大多以课程或者一节课为最小粒度单位, 在一定程度上虽能够满足学习者的需求, 但在某些情况下, 学习系统的用户常常有更高的要求, 即希望学习系统能够提供基于知识点的学习服务, 例如企业的知识管理系统或在线培训系统, 或即时信息服务系统(如银行前台客户服务)。事实上, 当前知识管理和知识服务系统中的知识粒度问题正在引起广泛重视, 并成为

一个研究的热点。在企业的知识管理系统中, 设计失败的一个关键原因是未能在项目初始阶段, 确定一个科学合理的知识粒度层次^[2,3]。由于系统提供给用户结果的颗粒度过大, 导致可重用性较差, 降低了知识共享水平。

学习对象的粒度通过e-Learning系统的元数据方案来实现。元数据的设计直接涉及学习对象的粒度模型(内容模型)、知识模块的语义结构等核心问题, 并与知识服务系统的整体需求密切相关。本文将首先介绍相关研究背景, 然后在分析本研究用户需求的基础上, 就元数据设计中的几个关键问题进行研究, 最后设计出适用于多粒度层级的学习对象元数据框架。

2 研究背景

1994年加拿大学者Wayne Hodgins构造出“学习对象”一词^[4], 将共建共享、学习资源重用的理念引入开放教育, 许多教育机构或企业培训部门纷纷开始研究如何将所拥有的知识库内容重新开发组织, 更方便地进行检索和重用。显然, 元数据描述是知识组织最为重要的方法, 对于任何一个学习资源来说, 如何通过元数据描述使其被方便地检索和重用, 是项目初期重点考虑的问题。虽然LOM元数据标准提供了较全面的、可资借鉴的标准, 但具体到某个特定的应用时, 仍然存在着需要研究探讨的问题, 其中包括学习对象的粒度(也称聚合度)^[5]、学习对象的语义结构等问题。

* 本文系十二五国家科技支撑计划项目课题“科技知识组织体系共享服务平台建设”(编号: 2011BAH10B03)、中国科学技术信息研究所预研课题《科技文献知识服务中基于不同粒度的内容模型及其描述体系研究》(编号: YY-201130)的研究成果之一。

学习对象的粒度问题之所以得到重视,源于教育实践中来自学习者和教学方的需求。一方面,学习者有着各自不同的知识结构和学术背景,或不同的工作岗位和发展阶段,对同一知识内容的需求角度是不同的,希望能够按照自己的路线或知识结构获取知识,而不是依照一门课程或一节课所给定的学习路径或知识结构。另一方面,教学方也希望在教师备课过程中最大程度地重用知识库中已有的内容,而这些都是都需要对学习对象的大小和可拆分性进行研究。

学习对象的粒度设计视项目需求而各异,如国外较为著名CONNECTION项目和欧盟ARIADNE项目在粒度设计上就有较大的差异。CONNECTION项目^[6]探索了网上在线撰写教材的全新方法,将线性的教材转变为非线性的知识模块,提出了二层级的粒度划分方法:知识模块和知识集合。知识模块是知识集合(如一门课程、一门教材,或其他类型的知识集合)的基本模块^[7],由创作者来编撰定义,它可以被组装成结构化的知识集合。而欧盟ARIADNE项目由于其在线学习平台的更加复杂的功能,包括支持作者利用网站的编辑工具,通过新建或重用的方法在线构建文件^[8],因此粒度设计较为复杂,包括六个层级^[9]:素材(asset),教学信息(pedagogical information),教学实体(pedagogical entity),教学情境(pedagogical context),教学文档(pedagogical document)及教学范式(pedagogical schema),而且每一层都有相应的元数据描述^[10]。特别值得借鉴的是其对四种类型的语义结构属性的定义:概念(concept)、

论元(argument)、已解决的问题(solved problem)和简单文本(simple text),清晰地描述了学习对象在语义内容上的分类^[11]。

另外,由于特定项目涉及的领域不同,因此,元数据核心集和扩展集的设计也是一个较为重要的问题。对于核心集的复用,需要根据项目特点首先确定出元数据的设计原则。对于扩展集则难以用复用的方法解决,一般来说,无论是DC元数据对资源内容、知识产权、外部属性的描述,还是LOM元数据标准对教育信息、技术信息、关系信息及注释信息的精确描述^[12],都不能全面覆盖特定领域内文献描述的需求,在这方面的探索如北京大学的拓片元数据^[13],清华大学的建筑元数据^[14]以及农业元数据^[15]等。因此,扩展元数据需要结合实际需求和文献特点进行专门的设计。

3 项目需求分析及元数据设计原则

本研究的目标是结合客户(某银行培训学院)实际需求,将学习对象拆分为基于单个知识点的内容模块(本项目称为知识模块),并辅助以相应的组织和存储,经过元数据标注后的学习对象,可以得到最大程度的重用,并进而成为未来构建智能的知识服务系统的数据基础。

3.1 需求分析

在元数据设计之初,用户需求是必须考虑的问题,需求越复杂,要求元数据越丰富。但过于复杂的元数据规范会影响其易用性,因此需要在繁简间取得平衡。本项目的

用户需求可以用两个具体的情境来说明:

(1) 满足基于自主路径设计的在线学习。这种需求对传统的以一节为最小粒度的学习对象提出了挑战。所谓自主路径设计,是指学习者不需要跟随教师安排好的课程结构进行学习,而是按照自己实际的知识结构进行跳跃式的或查漏补缺式的学习。另外还包括基于知识点的学习,即学习者(如银行前台)只需要知识体系中的一个很小的片断甚至一个知识点。

(2) 基于现有资源知识重用的辅助组课。组课的过程是基于现有知识模块,面向特定教学群体的知识重用过程。研究表明,学习资源或课程文本具备一个基本的逻辑结构,这个逻辑结构与课程类型及知识模块的结构属性有关^[16,17],本研究中的案例类课程、基础理论类课程、制度规范类课程的内在逻辑结构就有较大的差异。因此知识重用要求能够对知识模块的语义结构进行准确的表达。

综上所述,设计一个合理的知识粒度层次,并对其语义结构属性进行准确表达,是本研究必须解决的关键问题。

3.2 元数据设计原则

根据上述元数据的应用情境,为保证元数据规范在功能、结构、设计方法、结构格式等各方面能够实现与其他学习资源的互操作和资源重用,本文制定了以下原则:

(1) 用户需求原则:用户的需求反映了元数据应用的具体情境,为了更好地满足用户需求,需要考虑用户的具体应用场景下对元数据检索、选择、管理、评估、交互等方

面的要求。在元数据的结构设计、元素取舍时,尽可能从用户角度出发,满足用户的各种情境下的需求。

(2) 互操作原则:元数据应支持与其他元数据系统的语义互操作,为此,应建立一个适用的开放的元数据体系框架模型,在结构、格式、内容编码体系等方面进行规范定义,元数据内容描述使用标准的内容编码体系,如主题词表、资源类型、日期等,从而保障内容描述方式的标准化和描述内容的可交换;采用SCORM标准化封装。

(3) 通用与专用兼顾原则:由于客户所提供的学习对象涉及多种类型和粒度,因此,不可能只采用一种元数据标准就覆盖所有类型和粒度的学习对象资源。这也是本研究的复杂性所在。一方面,应尽可能复用或嵌套国际标准的元数据元素(LOM),避免创建新的元素,同时兼顾适用于具体的资源类型和粒度的元数据元素。

(4) 继承性原则:原始文献经拆分整理后,所形成的知识模块脱离了原有的语境信息,并且独立存储与检索使用,在新的应用情境下产生了新的语境。因此知识模块的内容及描述需求均不同于其原始文献。但在设计和标注过程中,还要尽可能直接继承原始文献的元数据元素。

(5) 最小集原则:过多或过复杂的元数据元素会不可避免地增加标注的工作量,降低工作效率,因此在满足用户需求的前提下,尽可能减少元数据元素的数量,以最小的元数据集来最大程度地涵盖所有类型和粒度学习对象的需求,在词汇的选择上,选取那些概括性较强的词汇,如发布者可以代替出版

社(书籍)、发文机构(文件)或发文部门(报文)等,专用标识符代替书籍的ISBN号、报文的发文号、文件的文件编号等。

4 元数据设计中的关键问题及其解决方案

在综合分析用户需求及现有相关研究基础上,本项目根据客户的实际需求,将现有散在于各种文献资料中的知识或信息抽取出来,加

工转变成粒度较小的知识模块(拆分示意图见图1)。通过粒度设计和元数据描述,赋予知识模块语义信息和结构属性。基于知识模块的学习系统在知识的表达和传递上更加灵活,具有更强的适应性。原始文献经过拆分后,其中蕴含的知识点转化为一个个独立的知识模块,脱离了原有的上下文语境,不再具有任何情境信息,极大地方便了知识的重用。

具有一定语义结构属性和粒度

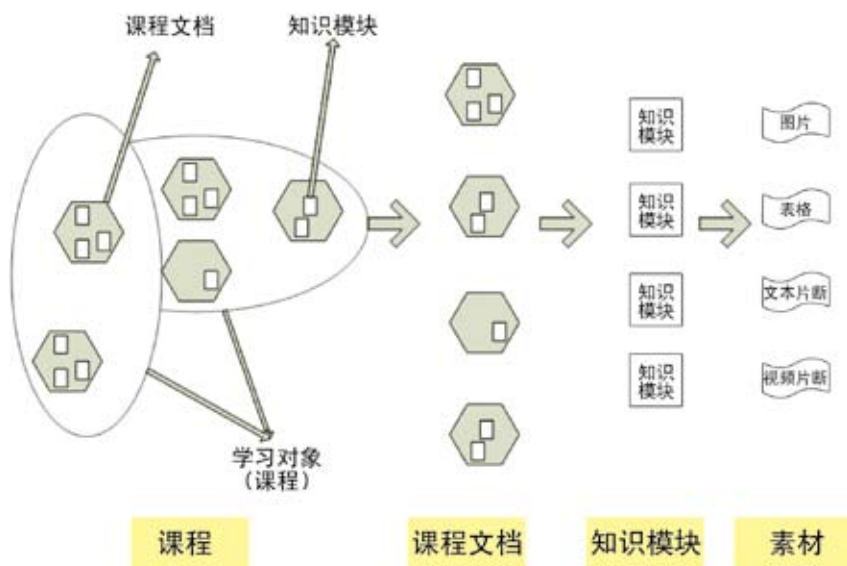


图1 学习对象拆分示意图

层级的学习对象是e-Learning系统知识库的核心,在一些应用中,粒度是用户检索使用得最多的一个元数据字段之一^[18],因此本研究设计中的关键问题在于学习对象的粒度层级设计及其语义结构属性设计。

4.1 学习对象内容模型

本项目提出了一种基于知识模块的学习对象内容模型,即以知识模块为核心描述对象,由素材、知

识模块、知识图谱、课程文档及课程构成的粒度图谱(见图2)。

素材:指不具有独立使用意义的图片、表格、一段视频或音频、文本片段等。

知识模块:知识模块是内容模型的核心,是指在本应用研究框架下,具有独立使用意义、粒度最小、不可再拆分的知识点,如一个概念的定义或其某个具体属性。知识模块好比是一个个内容完整独立的学习“积木”,可以从其原有上下文中

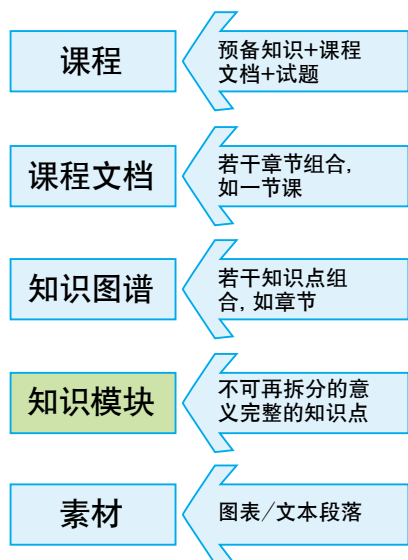


图2 学习对象内容模型

抽出来单独使用，或可以与相关模块组合成更高粒度的内容模块。图片或表格不能脱离文字描述单独存在，因此，带有说明性文字的图片/表格也作为知识模块。

知识图谱：是若干知识模块的集合，如课程的章节，一般围绕某一个主题展开论述，是知识模块之上更高的粒度层级。

课程文档：若干章节组合，如一节课。

课程：是包含预备知识和试题的课程文档。

4.2 知识模块的语义结构

研究表明，当一个知识点的粒度处于不可再拆分状态时，其内容往往是围绕着一个特定概念的某种属性而进行的论述^[19]，换句话说，一个概念加一个属性就可以完整表达一个基本的知识内涵，因此本研究将只涉及一个主要概念及其属性的知识模块称为简单知识模块。但有些知识模块可能涉及两个或多个概念的联系与对比，或一个概念的

多个属性间的联系，这类知识模块称为复合知识模块。

知识模块的元数据语义描述中，概念及其属性的描述是十分重要的，对于用户的检索和精确定位、教师组课都至关重要。据初步估算，简单知识模块一般占知识库知识总量的70%以上，复合知识模块占30%以下。知识模块内容模型的这种划分方法，为原始课程文档的拆分确定了依据和原则。本研究围绕着客户业务内容和特点，设计了基于主题词的分类表和属性词表，用以规范表示知识模块的内容，其他不能够用属性词穷尽的知识模块，则由标注者标注后补充到属性词表。对于复合知识模块，主题词和属性词涉及两到三个词。属性词元素的设计是为了从语义上更加准确地表达学习对象的语义内容，有助于帮助教师组课，或辅助确定知识模块在教学文档中可能的位置信息。

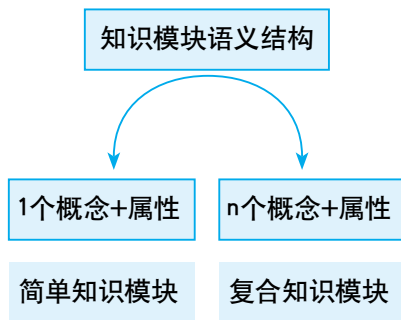


图3 知识模块的语义结构

5 学习对象元数据应用框架

本研究经过研究与分析，最终参照LOM标准，依据银行实际需求，设计了一套学习对象元数据框

架，整体框架见图4。“核心+扩展”模式的优势在于既可以达到与其他元数据标准的互操作性要求（核心集），也可以兼顾应用领域资源的特征和专业性需求（扩展集）。该方案较好地描述了多层级粒度学习对象的属性，为下一步的技术工具开发和应用奠定了良好的基础。

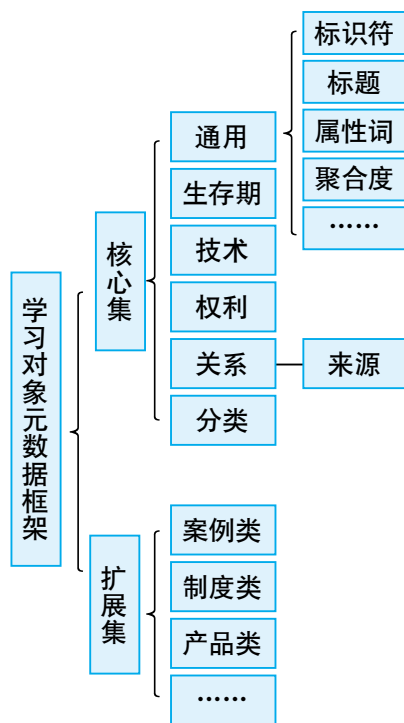


图4 学习对象元数据框架

5.1 核心集

核心集分为6大类元数据，共14个元素。为了确保互操作性，尽量复用现有标准如LOM元数据或DC元数据中的元素。对于较通用的元素，直接进行复用，如标题、关键词、分类等。对于较有领域特色的元素，采用部分复用的方法，对于值空间进行自定义。

完全复用的元素：指命名空间、定义、值空间完全相同。包括标题和关键词等元素。

部分复用的元素：指元素名称和定义相同，但命名空间和值空间不同。包括聚合度、关系、格式等元素。

自定义元素：指元素名称、定义、值空间完全自定义，包括来源等元素。

粒度元素即聚合度，值空间分为5个值，分别是素材、知识模块、知识图谱、课程文档和课程。虽然在前期拆分实践中只涉及知识模块和课程文档两个粒度层级，其他粒度在后期组课时会半自动生成，最终知识库中会包含全部五个粒度层级的学习对象。

5.2 扩展集

扩展集共12个元素，按照不同的文献体裁，涉及4大类扩展集，如案例类、产品类、制度类等，全部采用自定义方式。

5.3 元数据框架应用

图5表明了元数据框架在工具中的实现。显然在应用中，对于不同粒度层级的学习对象，需要从标准框架中部分选取使用，其中知识模块采用核心元数据集，而课程文档则采用核心集+扩展集，其他粒

表1 学习对象粒度层级及其元数据方案

学习对象粒度层级	元数据方案
课程	核心集+扩展集
课程文档	核心集+扩展集
知识图谱	核心集+扩展集
知识模块	核心元数据集
素材	主题，来源，发布时间，作者

度层级的元数据会由系统自动标引或工具自动生成。

6 结论

本项目设计了一种面向客户需求的多粒度层级学习对象元数据框架，满足了用户对e-Learning和教学活动中自主路径设计和计算机辅助组课的需求。在设计元数据应用框架过程中，对学习对象的内容模型、知识模块的语义结构属性等关键问题进行了研究，这些问题的解决使元数据能够更好地满足学习者、教师、教学管理方对于学习对象的多方面需求。下一步将就不同粒度学习对象的形式化表示、数据库存储与用户检索等问题展开研究。



图5 知识模块加工平台

参考文献

- [1] HODGINS W. Searching for the Future of Metadata [OL]. [2010-10-19]. <http://dc2004.library.sh.cn/english/prog/ppt/hodgins.ppt>.
- [2] WANG KESHENG. Introduction to knowledge management: principles and practice [M]. Trondheim: Tapir Academic Press, 2001:109.
- [3] Architect zone. Knowledge Management: It's All About Granularity [OL]. [2010-10-19]. <http://architects.dzone.com/news/knowledge-management-it%E2%80%99s-all>.
- [4] The Current State of Open Educational Resources [OL]. [2010-10-19]. <http://opencontent.org/blog/archives/247>.
- [5] Reusable Learning -Glossary [OL]. [2010-10-08]. <http://www.innovativelearning.ca/archive/resources/examples/glossary.htm>.
- [6] Connexions [OL]. [2011-05-11]. <http://cnx.org/>.
- [7] Ariadne foundation [OL]. [2011-05-11]. <http://www.ariadne-eu.org/>.
- [8] ARIADNE project. ARIADNE Educational Metadata Recommendation [OL]. [2011-05-11]. <http://vs.fernuni-hagen.de/methoden/ils/Organisation/ariadne.html>.
- [9] FERNANDES E, MADHOUR H, MINIAOUI S, et al. Phoenix Tool: A Support to Semantic Learning Model [OL]. [2011-10-02]. <http://www.win.tue.nl/SW-EL/2005/swel05-icalt05/final/W3-3.pdf>.
- [10] FERNANDES E, MADHOUR H, MINIAOUI S, et al. Phoenix Tool: A Support to Semantic Learning Model [OL]. [2011-10-03]. <http://ali2.unil.ch/articles/icalt.pdf>.
- [11] FERNANDES E, MADHOUR H, MINIAOUI S, et al. Phoenix: An XML-Based Tool for E-Learning Environment [OL]. [2011-10-03]. <http://ali2.unil.ch/articles/>

ArticleEdMedia.pdf.

- [12] IEEE 1484.12.1-2002. Draft Standard for Learning Object Metadata [OL]. [2010-06-01]. http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.
- [13] 肖珑,陈凌,冯项云,等.中文元数据标准框架及其应用[J].数字图书馆论坛,2001(5).
- [14] 罗德胤,吴开华,邢春晓.元数据与建筑学知识管理[J].数字图书馆论坛,2002,20(3).
- [15] 崔运鹏,钱平,苏晓莺.农业科技信息核心元数据标准框架研究及其著录信息管理系统[J].中国农业科学,2007,40(4).
- [16] HORN R E. Two Approaches to Modularity: Comparing the STOP Approach with Structured Writing [OL]. [2011-08-14]. http://www.stanford.edu/~rhorn/a/topic/stwrng_infomap/artclModularity.html.
- [17] 维基百科. Information Mapping [OL]. [2011-05-11]. http://en.wikipedia.org/wiki/Information_mapping.
- [18] NAJJAR J, TERNIER S, DUVAL E. The Actual Use of Metadata In Ariadne: AN Empirical Analysis [OL]. [2011-05-12]. http://www.cs.kuleuven.ac.be/~najjar/papers/EmpiricalAnalysis_ARIADNE2003.pdf.
- [19] NAMAHN. Information Mapping™. a research note by Namahn [OL]. [2011-06-19]. <http://www.namahn.com/resources/documents/note-IM.pdf>.

作者简介

张东 (1968-), 女, 博士, 中国科学技术信息研究所, 副研究员, 主要研究方向为知识工程与知识管理。E-mail: zhangdong@istic.ac.cn

Research and Design of a Metadata Framework toward Multi-layer Granular Learning Object

Zhang Dong, Zhu Lijun, Zhang Yunliang, Qiao Xiaodong / Institute of Scientific and Technical Information of China, Beijing, 100038
Xiao Feng / China Minsheng Banking Corporate, Beijing, 100084

Abstract: Towards the requirements of reusability and the self learning path design that exist in current enterprise e-Learning and knowledge management system, a metadata application framework that is suitable for multi-layer granular learning objects has been designed, and the key issues have been resolved, including content model and semantics structure model of the knowledge chunks.

Keywords: Learning object, Metadata, Granularity, Knowledge chunks, Content model, Knowledge management

(收稿日期: 2012-01-06)