

基于 XML 模式的作业描述语言的设计与处理

朱新华, 黄立和, 罗辉, 张显全

(广西师范大学计算机科学与信息工程学院, 桂林 541004)

摘要: 作业描述语言为远程教育中的 Web 作业管理系统, 提供了一套基于 XML 的分割、表示与处理半结构化作业数据的解决方案, 并可作为异构型作业管理系统之间交换作业文档提供一种技术规范。介绍了基于 XML 模式的作业描述语言的设计过程与使用方法, 并对 XML 作业文档的验证、显示与存储等处理方式进行了讨论。

关键词: XML 模式; 半结构化数据; 远程教育; Web 作业管理系统; 作业描述语言

Design and Processing of Assignment Description Language Based on XML Schema

ZHU Xin-hua, HUANG Li-he, LUO Hui, ZHANG Xian-quan

(College of Computer Science and Information Engineering, Guangxi Normal University, Guilin 541004)

【Abstract】 The assignment description language provides an XML-based approach to partitioning, representing and processing the semi-structured assignment data in the Web-based assignment management systems of distance education, and it can provide a technology specification for heterogeneous assignment management systems to exchange assignment documents with each other. This paper introduces the designing process and using method of the assignment description language based on XML schema, and it also discusses the processing methods of the validation, display and storage of assignment's XML document.

【Key words】 XML schema; semi-structured data; distance education; Web-based assignment management system; assignment description language

1 概述

Web 作业管理系统是网络教学支持系统中一个非常重要的子系统^[1], 它提供在线作业管理环境, 真实模拟课堂教学中的作业布置、答题、提交与批改过程, 实现作业管理的自动化和无纸化, 可有效克服传统课堂教学中批改作业工作量大、效率低、反馈周期长等缺点, 并因无纸化而有利于环境保护。

Web 作业管理系统与以标准化试题为主的在线测试系统相比有着很大的不同, 主要表现在: (1) 它不仅要实现单选题、多选题等标准化试题内容的描述、存储与管理, 而且还要实现问答题、分析题等非标准化试题的描述、存储与管理。(2) 除需存储作业的试题文档外, 它还需记录学生对作业的答卷与教师对作业的批改内容。(3) 一门课程不同章节的作业, 试题重复出现的概率很低, 因此作业库可以直接由粒度较粗的、由若干试题组成的作业文档构成。

作业文档是一种典型的同类不同构的半结构化数据, 它所包含的每一道试题都有着相对固定的结构, 但由于不同章节的作业文档所包含的试题数量与试题类型各不相同, 所有作业文档却没有一个统一的结构。因此, 很难在传统的关系数据库中, 使用一个关系模式统一刻画不同章节的作业结构, 而必须使用多个子关系分别描述每类试题的结构, 然后通过子关系的连接运算重组作业文档, 其实现方法非常复杂。但是, 采用可扩展标记语言 XML, 却可以非常简单、直观地描述作业的内容与结构。

XML 是面向数据而非面向显示的, 它集中于描述数据本身的性质与结构。XML 文档的形式是自由的文本格式, 使用

带有语义的自描述性标记区分不同的数据成份, 并允许所描述的同类数据具有不同的结构, 即 XML 所描述的数据可以是半结构化的^[2]。XML 已在电子商务、数据交换、Web 服务等各个领域有了相当多的应用, 在各领域内开发了基于 XML 的工具语言, 并定义了面向应用的词汇表、语法及语义^[3]: 在数学领域, 采用 MathML 描述数学表达式的结构和内容; 在化学领域, 基于 XML 的 CML 主要用于描述分子结构和序列、光谱结构与化学数据库等; 在教育领域, 美国 IMS 全球学习委员会发布了系列基于 XML 的学习技术规范。

XML 不是一个独立的预定义的标记语言, 它是一种元语言, 是用来描述其他语言的语言。在实际应用中, 首先要使用 XML 针对具体的领域设计出相应的标记语言。为实现 XML 在 Web 作业管理系统中的应用, 本文设计了一套作业描述语言, 该描述语言由 2 大类标记组成: 作业正文标记与答卷标记。

2 作业描述语言的设计

2.1 数据模型分析^[2,3]

XML 是采用树形逻辑结构描述数据, 设计基于 XML 的描述规范的首要任务是分析有关数据对象的数据模型, 并将其转换为树的形式。在作业管理系统中有 2 类数据对象存在:

基金项目: 广西自然科学基金资助项目(0447034)

作者简介: 朱新华(1965 -), 男, 高级工程师, 主研方向: 计算机辅助学习技术, 网络教学系统, XML 应用; 黄立和, 硕士; 罗辉, 工程师; 张显全, 教授

收稿日期: 2007-02-05

E-mail: zxb429@263.net

作业正文文档与答卷文档。作业正文文档用于完整保存构成一个作业的所有试题内容；答卷文档用于记录学生对作业各试题的回答及老师的批改内容，为减少数据冗余，答卷文档中不再重复出现试题正文。作业文档与答卷文档之间存在着 1:N 联系，即教师布置的每一个作业文档都对应着不同学生的多个答卷文档。经过对作业数据的全面分析与抽象，得出如图 1 与图 2 所示的作业正文文档与答卷文档的数据模型树。

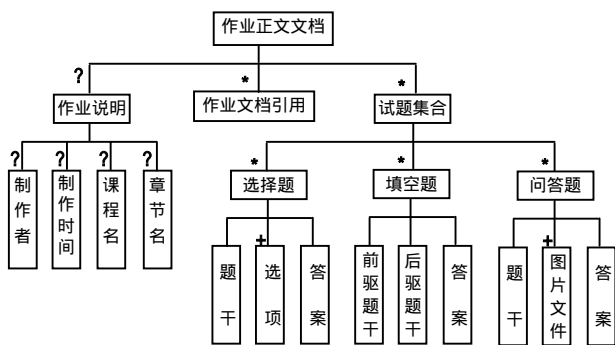


图 1 作业正文文档数据模型树

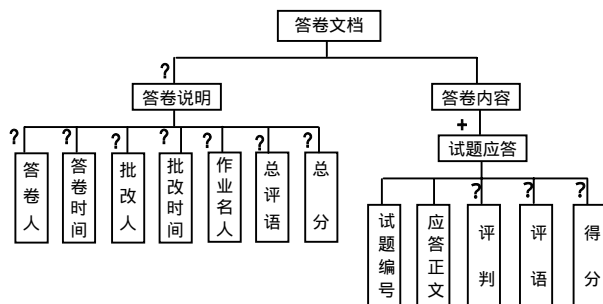


图 2 作业答卷文档数据模型树

图 1 与图 2 给出的是一种半结构化的数据模型，它们允许一些节点在实例数据中不出现或重复出现多次。图 1 中所标注的符号借用 XML 的文档类型定义 (document type definition, DTD) 中的使用方法：加号 (+) 表示可出现一次或多次，问号 (?) 表示可出现零次或一次，星号 (*) 表示可出现零次或多次，没有做标注的元素表示必须出现且仅出现一次。在作业正文文档中设置“作业文档引用”元素，可以实现作业内容的共享，并可通过多层次的作业文档引用，形成作业间的层次结构。“问答题”元素是所有综合题、计算题、编程题、分析题与简答题等主观试题的抽象与代表。在“填空题”元素中，将位于填空位置之前的题干定义成“前驱题干”子元素，而位于填空位置之后的题干定义成“后驱题干”子元素。在实际应用时，必须为各元素定义一些必要的属性，如“作业正文文档”元素应具有“作业编号”属性，“选择题”元素应具有“试题编号”属性等。

2.2 Schema 模式设计^[4]

模式 (schema) 是 W3C 制定出来专门用于定义与验证 XML 文件内容和结构的一种最佳机制，它除了提供早期的文档类型定义 DTD 机制所具有的绝大部分功能外，还引入了许多新的功能，如丰富的数据类型、对域名的支持等，目前它有取代 DTD 机制的趋势，因此笔者采用 XML Schema 对作业描述语言进行定义与约束。

为确保不同系统或教师能够以统一的词汇与符合要求的

结构描述作业的内容，必须为所设计的作业描述语言定义一套 XML 模式。这个模式还可作为一种作业技术规范，各种异构型作业系统，可按照该规范自由交换与共享作业文档资源。

XML Schema 模式是在一个独立的命名空间，使用 XML 的基本语法规则进行描述的。为便于对作业正文文档与答卷文档分别进行验证，作业正文标记与答卷标记必须在两个不同的 Schema 模式中进行格式定义。以作业正文标记的模式文件 Aml.xsd 定义为例，Schema 模式命名空间的框架结构如下：

```
<xs:schema xmlns:xs = "http://www.w3.org/2001/XMLSchema"
  xmlns = "http://www.AssignmentSchema.org/xsd/Aml"
  targetNamespace = "http://www.sequencingObject.org/xsd/Aml">
  <元素定义语句块>
</xs:schema>
```

元素的定义包括类型、属性、子元素与子元素间的顺序等多方面的描述，以下给出作业正文标记中“试题集合”元素的模式定义片段：

```
<xs:element name = "试题集合">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref = 选择题 minOccurs = "0"
maxOccurs = "unbounded"/>
      <xs:element ref = 填空题 minOccurs = "0"
maxOccurs = "unbounded"/>
      <xs:element ref = 问答题 minOccurs = "0" maxOccurs =
"unbounded"/>
      <xs:any namespace = "##other" processContents = "strict"
minOccurs = "0" maxOccurs = "unbounded"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name = "问答题">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref = 题干 type = "xs:string"/>
      <xs:element ref = 图片文件 minOccurs = "0"
maxOccurs = "unbounded" type = "xs:anyURI"/>
      <xs:element ref = 答案 type = "xs:string"/>
    </xs:sequence>
    <xs:attribute name = "试题编号" use = "required" type =
"xs:string"/>
  </xs:complexType>
</xs:element>
...
```

在“试题集合”元素中所定义的“any”子元素，可允许该元素包含任何在其他模式命名空间中定义的合法元素，从而为兼容其他描述语言，以及未来新题型的扩展留下了契机。

2.3 XML 文档的编写

使用上述模式所定义的作业标记，就可以编写作业正文与答卷的 XML 文档。根据 XML 语言规范的要求，这两类文档应符合以下原则^[2]：

- (1) 包含所需要的 XML 声明，即给出文档必需的版本号、字符集、名字空间、模式与样式引用等信息；
- (2) 作业正文文档必须以<作业正文文档>标记作为顶级

元素(即根元素), 答卷文档必须以<答卷文档>标记作为顶级元素;

(3)元素、属性和实体必须符合模式中要求的层次和包含关系。下面给出一个只包含一道问答题的作业正文 XML 文档:

```
<?xml version = "1.0" encoding = "UTF-8"?>
<作业正文文档 xmlns = " http://www.AssignmentSchema.org
/xsd/Aml" 作业编号= " 001 " >
  <作业说明>
    <课程名>Visual Basic 程序设计</课程名>
    <章节名>第七章 常用控件</章节名>
  </作业说明>
  <试题集合>
    <问答题 试题编号= " 001 " >
      <题干>在如图所示的时间显示程序中, 如果要求每隔
      0.5 秒刷新一次时间, 则时钟控件的 Interval 属性应设置为多少? </
      题干>
      <图片文件>Interface.Gif</图片文件>
      <答案>500</答案>
    </试题集合>
  </作业正文文档>
```

3 XML 文档的处理

3.1 XML 文档的验证

可通过 XML 解析器验证用户所编写的 XML 作业文档, 是否遵守了作业描述语言的 Schema 模式。XML 解释器可采用 W3C 的基于对象的 DOM(document object model)模型或基于事件的 SAP(simple API for XML)模型来实现。鉴于完整地实现一个 XML 解析器是一项庞大的工程, 在实际应用中, 可直接调用各大商业公司所提供的 XML 解析器来验证 XML 文档。目前, 比较通用的 XML 解析器主要有 3 种: (1)SUN 公司的 JAXP(Java API for XML processing) 解析器; (2)Oracle 公司的 XML Parser for Java v2 解析器; (3)IBM 公司的 XML Parser for Java 4J 解析器。它们都同时提供 DOM 与 SAP 接口。

3.2 XML 文档的显示

XML 文档所关注的是对内容的描述, 它不能直接在浏览器中显示。要想在浏览器中显示 XML 文档, 必须以某种方式告诉浏览器如何显示。在一个作业管理系统中可以通过以下 2 种方式显示作业的 XML 文档:

(1)样式单法。可使用层叠样式单 CSS(Cascading Style Sheets)与可扩展样式单语言 XSL(eXtensible Stylesheet Language)两种样式单显示 XML 的作业文档。一个样式单可作用于多个 XML 作业文档; 而一个 XML 作业文档也可根据不同的使用者分别使用不同样式单, 呈现出不同的内容。该方法简单、易于实现, 一般用于对 XML 作业文档的查询显示。

(2)编程法。在网页的脚本程序中, 通过 XML 文档的 DOM 树模型提取作业文档各元素中的数据进行显示。该方法可将作业正文文档与答卷文档混合显示, 并可在显示界面中嵌入各种输入表单, 可广泛应用于作业的查询、录入、修改、应答与批改等各类界面。

3.3 XML 文档的存储

在 Web 作业管理系统中, 根据目前业界 XML 文档的存储技术, 并结合作业文档对象自身的特点, 作业的 XML 文档有以下 4 种存储方式^[5]:

(1)作为普通文件存储于服务器的文件系统中。该方法将作业的 XML 文档按照不同的类别, 分别保存在不同的目录中, 通过 Web 链接或操作系统的文件系统搜索作业的文档文件, 使用 XML 的 XPath 与 XQuery 技术查询文档内容。该方法需存在查询速度慢、并发控制性能差等缺点, 只适合小型的作业管理系统。

(2)利用关系数据库索引作业文档。在该方法中, 作业的 XML 文档仍保存为独立的文件形式, 但通过数据库对其进行引用, 即将文档文件的存储路径保存在数据表中的一个字段中。为提高查询速度, 还可以在数据库中镜像存储作业文档的说明信息。

(3)存储于专门为 XML 设计的原生 XML 数据库(Native XML Database)中。目前比较著名的商用 XML 数据库有美国 Software AG 公司的 Tamino 数据库, 以及美国 dbXML 公司的 dbXML 数据库等。该方法实现简单, 支持 XML 的节点查询, 但查询速度慢, 查询方法通用性差。

(4)存储于关系数据库中。该方法将作业的 XML 文档按照某种方式进行转换, 然后存储到关系数据表中。为降低转换的实现难度, 在作业系统中可采取一种简单转换方法: 将作业文档说明节点中的各元素分别对应数据表中的一个字段, 而将“试题集合”与“答卷内容”的整个节点打包, 分别对应数据表的一个可变长字符串或二进制大对象字段。

4 结束语

对于上述所介绍的作业描述语言, 笔者在所开发的 Visual Basic 教学网站中的作业管理子系统对其进行了试用, 结果表明该作业描述语言在作业内容的分割、表示、验证、显示与存储等各方面都提供了非常强的功能, 获得了比纯数据库方式更好的性能与效果。

本文所介绍的作业描述语言, 既可以直接在作业管理系统中用于表示与存储作业文档, 也可作为一种技术规范, 为异构型的作业管理系统间的作业交换, 提供一种标准的作业表示方式。

参考文献

- 1 王宇, 张五红. 高校网络辅助教学平台选择、实施及分析[J]. 电化教育研究, 2006, 27(2): 43-46.
- 2 许斗, 陈恩红. XML 的半结构化数据表示方法及其在医学文档处理中的应用[J]. 计算机工程, 2002, 28(1): 57-58.
- 3 刘晓平, 毛峥强, 石慧, 等. 面向网络化协同模板技术的模板设计标记语言[J]. 计算机辅助设计与图形学学报, 2006, 18(8): 1224-1230.
- 4 Walmsley P. XML 模式权威教程[M]. 陈维军, 乔安平, 莫宇, 译. 北京: 清华大学出版社, 2003.
- 5 杨小花, 杨宗凯, 吴砥. 基于 XML 数据库的标准化教育资源内容服务系统的设计与实现[J]. 计算机应用研究, 2006, 25(12): 461-462.