

基于XML的SQL Server数据库技能操作 自动阅卷^①

陈尧妃 陈焕通 (金华职业技术学院 信息工程学院 浙江 金华 321017)

倪应华 (浙江师范大学 行知学院 浙江 金华 321004)

摘要: 针对SQL Server数据库技能测评系统缺乏的现状,分析了现有逻辑形式化自动阅卷技术,提出了基于XML实现SQL Server数据库技能操作自动阅卷方案。介绍了自动阅卷的实现原理和流程,重点阐述了两个关键技术环节:XML形式化表达和XML形式化自动阅卷技术。阅卷效果表明,基于XML的SQL Server数据库技能操作自动阅卷是切实可行的。

关键词: 自动阅卷; SQL Server; 形式化; XML; 数据库

Automatic Marking in SQL Server Skills Assessment Based on XML

CHEN Yao-Fei, CHEN Huan-Tong (Jinhua College of Profession and Technology, Jinhua 321007, China)

NI Ying-Hua (Xingzhi College, Zhejiang Normal University, Jinhua 321004, China)

Abstract: In light of lacking SQL Server skills assessment system, this paper analyzes the existing formal logic automatic marking technology and proposes a model about skills assessment of SQL Server based on XML. It introduces the principle of automatic marking and its procedures. It focuses on two key techniques: XML Formal Expression and XML technology automatic marking. Practice shows that the automatic marking of SQL Server database skills based on XML is practicable.

Keywords: automatic marking; SQL server; formal; XML; database

随着计算机应用教育教学改革的进一步普及和深入,高职院校越来越注重学生实践技能的培养。目前SQL Server数据库课程考试仅局限于理论知识的纸质测评,这种考核方式无法评价学生的实践技能水平。解决此问题的根本出路在于实现数据库技能操作测评的自动化。技能操作测评是计算机辅助测评CAA(Computer Assisted Assessment)领域中一个重点研究方向^[1]。目前研究比较薄弱,尤其缺乏一般性的理论与方法。从技能测评对象来看,现有的数据库技能测评系统都是针对VFP和Access数据库^[2,3],针对SQL Server的技能测评系统相当匮乏。实际上

SQL Server数据库在高校教学、技能考证、工程开发等方面应用相当广泛。从技能测评阅卷方法来看,现有测评系统主要采用逻辑形式化方法,依赖专家知识构建解析系统实现自动阅卷,该方法存在实现技术复杂,维护不便等不足。针对SQL Server技能操作的特点,利用XML在描述上的优势实现基于XML的SQL Server数据库技能操作自动阅卷。

1 现有数据库技能测评技术简介

目前数据库技能测评自动阅卷大都采用逻辑形式化方法实现。阅卷控制信息采用逻辑形式化描述^[4,5],

① 基金项目:2008年度浙江省教育厅项目(Y200805671)

收稿时间:2009-06-23

该方法将技能操作试题描述成一个逻辑表达式，表达式中包含技能点、技能点的信息提取参数、标准答案信息和评分标准等，阅卷时分析逻辑表达式中的技能点，根据技能点从学生操作数据库中获取参数值代入表达式计算，通过返回的布尔值判断该知识点是否得分。逻辑形式化法的形式化描述结构如图 1 所示。

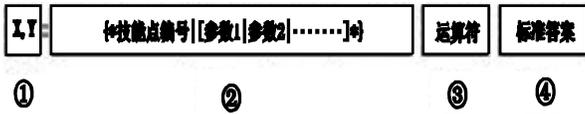


图 1 逻辑形式化描述结构图

说明：①X 试题题号，Y 试题分数。题号是为了方便后期数据分析。②、③、④组成一个逻辑表达式，返回结果为 True 时技能点操作得分。②描述了技能点编号及由该技能点决定的获取学生技能操作数据库操作信息的必须参数序列，②最终返回的是技能点学生操作答案。③运算符，主要是实现操作判断的关系运算符。④标准答案，标准操作结果。

在逻辑表达式构建完整、规范的前提下，逻辑形式化方法具有很高的测评精度。但是它也有不足：①逻辑表达式依赖于解析系统解析执行，解析系统的构建必须依赖专家支持。②解析系统只能解析规则的表达式，无法解析带有语法错误或逻辑错误的表达式。③技能操作阅卷信息必须依赖人工输入，无法自动生成，容易因为阅卷控制信息错误导致无法阅卷或者错误阅卷。

2 XML形式化自动阅卷技术

2.1 XML 形式化自动阅卷原理和流程

XML(eXtensible Markup Language)，即可扩展标记语言，是由万维网联盟 W3C 定义的一种语言。它以一个统一、开放、基于文本格式的模式来描述和交换数据。XML 是一种元标记(meta-markup)语言，它提供了一种描述数据的格式，能够运行在任何平台和操作系统之上。

借鉴 XML 在半结构化信息表达上的优势提出基于 XML 的数据库技能测评自动阅卷方案^[6,7]。整个 XML 形式化阅卷原理和流程如图 2 所示，整个 XML 形式化自动阅卷经过 3 个主要流程。第一步，教师根据 SQL Server 技能操作试卷对 SQL Server 初始数

据库进行做答，生成 SQL Server 标准操作数据库；学生做答生成 SQL Server 学生操作数据库。针对 1 份试卷教师只要做答一次即可，而学生则根据不同的学生会做答产生一系列学生操作数据库。第二步，使用程序设计工具从生成的操作数据库中提取数据库信息并通过 XML 形式化表达生成操作 XML 文档。教师做答的标准数据库最终生成 SQL Server 标准操作 XML 文档，学生做答的学生操作数据库最终生成 SQL Server 学生操作 XML 文档。第三步，根据 SQL Server 技能操作评分规则，将一系列 SQL Server 学生操作 XML 文档和 SQL Server 标准操作 XML 文档进行比对实现自动阅卷，将最终阅卷分数录入成绩库。从自动阅卷原理和流程的介绍中可以发现：XML 形式化表达和 XML 形式化自动阅卷是两个关键技术。

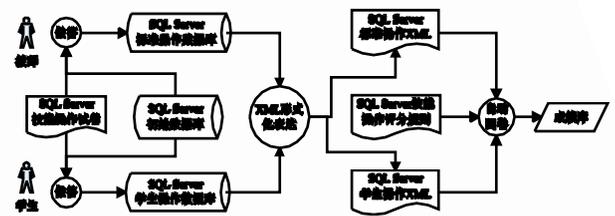


图 2 XML 形式化自动阅卷原理和流程

2.2 关键技术分析

2.2.1 XML 形式化表达

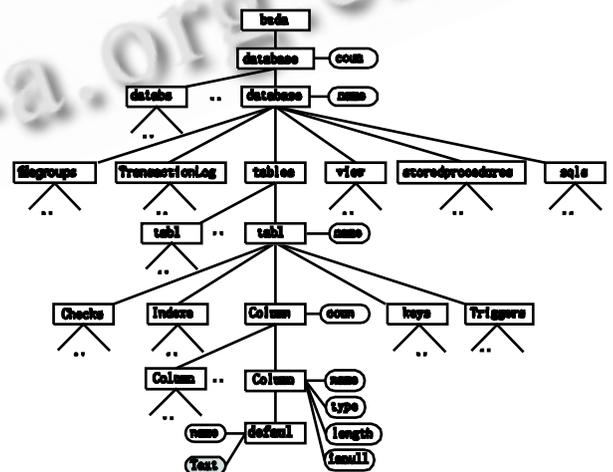


图 3 数据库 XML 树型文档描述结构

XML 形式化表达是实现 XML 自动阅卷的前提。XML 形式化表达主要分成两个技术环节：一是数据库信息获取技术，二是数据库信息 XML 表示技术^[8]。数

数据库信息获取可以通过查询 SQL Server 系统表的方法实现信息获取,具体方法请参阅参考文献[9]和[10];数据库信息 XML 表示主要是根据数据库对象结构和相互之间的关系建立数据库 XML 半结构化树型文档描述结构,该描述结构如图 3 所示。

根据图 3 的树型文档描述结构采用 SQL Server 系统表方法提取生成的 XML 文档结构如下(以 table 节点为例):

```
<?xml version="1.0" encoding="gb2312"?>
<bzda>
  <databases count="">
    <database name="">
      .....
    <tables count="">
      <table name="">
        .....
        <columns count="">
          <column name="" type="" length=""
isnull="">
            <default name=""/>
          </column>
        </columns>
        .....
      </table>
    </tables>
    .....
  </database>
</databases>
</bzda>
```

2.2.2 XML 形式化自动阅卷

(1) XML 形式化自动阅卷流程

XML 形式化自动阅卷根据执行流程可以分成 3 个步骤:第一步,读入两个数据库操作 XML 文档进行 DOM 解析^[11,12],在内存中形成两棵 DOM 树,将标准操作数据库 XML 记为 KeyTree,将学生操作数据库 XML 记为 KsTree;第二步,获取技能操作评分规则,规则描述结构如<试题编号,数据库名称,操作对象,对象名称,操作类型,题分>。第三步,通过技能操作评分规则描述中分析操作对象确定 DOM 树的对象节点,根据操作类型通过比对 KeyTree 和 KsTree 的相应节点实现评判。

(2) 数据库不同操作阅卷规则

技能测评中所有的操作类型分成三类: Create(创建)、Modify(修改)和 Delete(删除)。创建和修改的操作结果都会保存在数据库操作 XML 文档中。删除操作由于对象不存在因此也就不可能存在于 XML 文档内。因此,删除操作只需判断 KsTree 中当前操作数据库内是否仍然存在该对象的节点信息,如果不存在则得分,否则不得分。创建和修改操作由于对象节点众多,评判要复杂得多。具体阅卷规则如下:

① 在 KeyTree 中提取当前操作对象的节点(包括子节点信息),判断并去除该节点信息内部的冗余子节点,最终得到 XML 片段为 KeyXML。

② 在 KsTree 中提取当前操作对象的节点(包括子节点信息),如果不存在则结束判断,否则判断并去除该节点信息内部的冗余子节点,最终得到 XML 片段为 KsXML。

③ 提取 KeyXML 中的所有节点。这些节点包括元素节点和文本节点。依次从 KsXML 中比对相应节点是否存在,若存在则给分。

(3) 操作赋分规则

在进行 XML 形式化自动阅卷时,将操作分数平均分配到各个节点,每个节点的分数再根据属性的个数平均分配到各个属性,这样节点各属性的评分累计就是该节点的分数,各个节点的评分累计就是该操作的得分。在 XML 文档中可以发现除了根节点以外都带有属性。这些节点分成两个大类:带 Name 属性的节点和不带 Name 属性的节点。对于带 Name 属性的节点其他属性只有在 Name 属性值正确的前提下才有实际意义。因此对于带 Name 属性的节点首先判断 Name 属性值的节点在 KsXML 是否存在,若不存在则该节点其他属性不用再判,该节点不得分。若存在则继续判断其他属性。对于非独立节点,将该节点分配的分值按照属性个数再一次平均分配给各个属性。对于不带 Name 属性节点,阅卷时依次判断属性值是否一致,若一致则得分。

3 XML形式化自动阅卷实例和效果分析

3.1 XML 形式化自动阅卷实例介绍

下面以创建数据表为例说明 XML 形式化自动阅卷,技能操作要求“在已有数据库[GXC]中创建名为[货物损耗表]的数据表,……”。技能操作评分规则描述

为<3, GXC, Table, Create, 货物损耗表, 4>, 表示试题编号为 3 号的试题操作数据库名称是 GXC, 操作对象是 Table, 操作类型是 Create(创建), 数据表名称为货物损耗表, 题分是 4 分。从标准操作数据库中提取标准操作 XML 文档, 去除冗余节点得到 table 节点信息 KeyXML 片段内容如下:

```
<table name="货物损耗表">
  <columns count="6">
    <column name="条形码" type="varchar" length="30" isnull="0"/>
    <column name="货单号" type="char" length="20" isnull="0"/>
    <column name="损耗数量" type="numeric" length="9" isnull="0"/>
    <column name="原因" type="text" length="16" isnull="-1"/>
    <column name="处理结果" type="text" length="16" isnull="-1"/>
    <column name="经手人" type="char" length="20" isnull="-1"/>
  </columns>
</table>
```

由于 KeyXML 中的 table 节点个数共有 8 个, Table、columns 以及 6 个 column。按照操作赋分规则, 每个节点分配分值为 0.5 分。table 节点是带 name 属性节点且只有 1 个属性, 因此该 Name 属性具有 0.5 分的节点分数。Columns 节点不带 name 属性节点, 且只有 count 属性, 因此该属性也分配到 0.5 分。Column 是带 name 属性节点且共具有 4 个属性, 因此该节点四个属性都分配到 0.125 分。该节点分数评判首先判断 Name 属性值是否一致, 若一致则继续判断其他属性值; 若 Name 属性值不一致则整个节点不得分。

3.2 XML 形式化自动阅卷效果分析

为了进一步测试操作赋分规则的有效性, 选择了多个学生对该操作题进行测试, 从中选择了 10 个不同出错因素的学生答案和自动阅卷成绩进行分析和对比。分析数据如表 1 所示。从表 1 的阅卷分数和出错因素进行对比分析不难发现: XML 形式化自动阅卷能够根据不同的出错因素和出错程度给出合理的评分。

表 1 同一操作题不同答案自动阅卷比较分析

答案	阅卷分数	出错因素分析
答案 1	1.7	缺少 2 个列字段定义 1 个字段中 2 个属性错误
答案 2	2.0	缺少 1 个列字段定义且 2 个字段中 4 个属性错误
答案 3	2.0	缺少 2 个列字段定义
答案 4	2.0	缺少 1 个列字段定义且 1 个字段中 name 属性错误
答案 5	3.2	1 个列字段 name 属性错误且 1 个字段中 1 个属性错误
答案 6	3.3	1 个列字段定义中 name 属性错误
答案 7	3.2	增加了 1 个列字段定义且 1 个字段中 1 个属性错误
答案 8	3.3	增加了 1 个列字段定义且同列定义中有 1 个属性错误
答案 9	3.8	1 个列字段中有 1 个属性错误
答案 10	4.0	没有错误

3.3 两种形式化阅卷方法对比分析

XML 形式化方法只需关心操作试题的技能操作评分规则, 即只要关心操作对象、操作类型和操作分值等基本要素。对于分值分配和判分规则根据 XML 形式化自动阅卷规则自动完成。较逻辑形式化方法而言, XML 形式化法彻底克服了逻辑形式化法的不足, 两者性能对比见表 2 所示。

表 2 两种形式化阅卷方法比较

对比项目	逻辑形式化法	XML 形式化法
专家知识	依赖	不依赖
解析系统	需要	不需要
描述结构	固定	固定
操作语意	不清晰	较清晰
分值分配	人工分配	自动分配
规则维护	较难	容易

4 结语

本文分析了现有逻辑形式化自动阅卷方法的不足, 针对 SQL Server 数据库技能测评系统缺乏的现状提出了基于 XML 实现 SQL Server 数据库技能测评自动阅卷的方案。介绍了自动阅卷的原理和实现流程, 重点阐述了两个关键技术环节: XML 形式化表达和 XML 形式化自动阅卷技术。基于 XML 自动阅卷方案能够不依赖于专家知识, 无需解析系统支持, 只需要根据技能操作评分规则就能根据教师数据库标准答案对学生数据库操作结果进行有效测评。实践证明: 基于 XML 的 SQL Server 数据库技能操作自动阅卷是切实可行的。

(下转第 186 页)

参考文献

- 1 何克抗,许骏.计算机辅助测评CAA研究新进展.开放教育研究, 2005,11(2):78-83.
- 2 李桂英. II 技能测评系统的设计与实现.计算机系统应用, 2007,16(6):9-12.
- 3 李桂成,崔军.数据库操作题自动阅卷的设计与实现.计算机工程与设计, 2004,25(6):1005-1006.
- 4 张捷,封俊红.Windows 操作题自动阅卷的研究与实现.计算机系统应用, 2007,16(10):102-105.
- 5 金炳尧,马永进,骆红波,等.阅卷信息的形式化描述及其应用.计算机科学, 2005,32(1):106-107.
- 6 Bertino E, Carminati B, Ferrari E. Access control for XML documents and data. Information Security Technical Report, July-September 2004.134-137.
- 7 Dalamagas T, Cheng T, Winkel KJ, Sellis T. A methodology for clustering XML documents by structure. Information Systems, January 2005.51-55.
- 8 王会霞,王行仁.基于XML的知识获取工具研究与实现.航天控制, 2005,23(6):50-53.
- 9 陈尧妃,陈焕通.系统表在SQL Server 技能测评中的应用.金华职业技术学院学报, 2008,8(2):36-38.
- 10 朱更明,李方敏,李和军.基于数据库系统表的通用查询方法.计算机工程与应用, 2001,24:112-114.
- 11 李效东,顾毓清.基于DOM的Web信息提取.计算机学报, 2002,(25):526-533.
- 12 朱艳红,何东彬.基于XML文档检索的存储研究与实现.计算机系统应用, 2007,16(5):49-52.