

# 《数据结构》在线考试及评分系统的研究与开发

骆 健,王海艳,罗卫兰

(南京邮电大学 计算机学院,江苏 南京 210003)

**摘 要:**文中简要概括了考试及评分系统的系统结构、开发模式及核心功能。以经典测量理论(Classical Test Theory, CTT)为基础,全面分析了题库试题应包含的主要属性,设计了配合自动组卷算法执行的题库结构。在比较、分析常用自动组卷算法优缺点的基础上,给出了随机搜索逐步优化的自动组卷算法。在比较、分析常用程序题评分算法的基础上,提出了基于程序内部结构特征分析的编程题自动评分算法。开发实践证明,组卷成功率较高,评分算法具有较高的准确率,系统性能稳定。

**关键词:**在线考试系统;题库;自动组卷算法;自动评分算法

**中图分类号:**TP319

**文献标识码:**A

**文章编号:**1673-629X(2012)07-0181-04

## Research and Development of Online Examining and Grading System for Data Structure

LUO Jian, WANG Hai-yan, LUO Wei-lan

(College of Computer, Nanjing University of Post and Telecommunications, Nanjing 210003, China)

**Abstract:** In this paper, the system architecture, development pattern and the core functions of the examining and grading system is briefly introduced. Based on the classical test theory, the characters which the examination database must have is analyzed, a structure of the examination database for the automatic composing paper system is proposed. Based on comparing and analyzing the common algorithms of composing test paper, an algorithm of random search with optimization is developed. After comparing and analyzing the often used automatic grading algorithm, a new algorithm based on analyzing the internal character of the program is proposed. It is proved by implementation that the success rate of generating paper is improved and the accuracy of grading is also good and the performance of the system is stable.

**Key words:** online examining system; examination database; automatic composing paper algorithm; automatic grading algorithm

## 0 引言

目前,针对《数据结构》课程建立的试题库系统已有不少,主要集中在对题库结构和组卷算法的研究,对评分系统的研究相对较少。文献[1,2]对组卷算法进行研究,提出了随机抽取法,算法对试题约束条件少,导致试题重复率高、试卷的总体质量难以控制。《数据结构》考试中存在着编程题多的特点,而对编程题的评阅一直是阅卷技术中的一个难点问题。对编程题评阅技术的研究中,文献[3]提出使用动态测试法对程序题评分,但其仅通过结果进行评分,忽视考生程序的内部结构特征,0分或满分的极端现象较多,存在很大的不合理性。文献[4]提出使用编译原理技术对程

序的内部结构进行分析评分,克服了文献[3]中存在的主要缺点,但算法对程序运行失败的情况没有考虑。

针对以上出现的问题,文中对组卷算法与程序题评分算法进行了改进,主要创新性工作表现如下:

(1)以经典测量理论为指导,设计了配合自动组卷算法执行的题库结构。

(2)针对随机抽取法、回溯试探法等组卷算法中存在的试题重复率高、组卷效率低等问题,提出基于随机搜索逐步优化的适合在线考试的自动组卷算法。

(3)针对《数据结构》考试中编程题多的特点,着重研究编程题评阅技术,所设计的算法从程序的语法、语义角度衡量学生程序与正确答案的接近程度进行评分,并对程序运行失败的情况进行分析处理。

## 1 在线考试及评分系统的系统设计

### 1.1 系统的结构

文中所设计的在线考试及评分系统的基本结构可

收稿日期:2011-11-21;修回日期:2012-02-25

基金项目:江苏省基础研究计划(自然科学基金)滚动资助项目(BK2011072)

作者简介:骆 健(1976-),女,江西赣州人,硕士,讲师,主要研究方向为计算机网络、信息安全。

以抽象为以下三个层次<sup>[5]</sup>:应用层、业务逻辑层和数据层,如图 1 所示。

应用层即用户直接操作界面层,它是系统与用户交互的最外层。它的主要功能是接受用户的操作指令,并且传送给业务逻辑层和数据层,同时将操作处理所得输出给用户。例如,通过这一层,用户可进行输入注册信息、查询试题类型等操作。

业务逻辑层是系统的中间层,它的主要功能是完成客户的应用功能,即 Web 服务接受客户请求,并与后台数据库连接,进行业务处理,最后将处理结果返回 Web 服务器,再传递至客户端。该层主要分为试题管理、组卷管理等功能模块。

数据层主要用来存储系统运行中所需的各类持久数据,是系统的最底层。该层由用户表、试题表等组成。

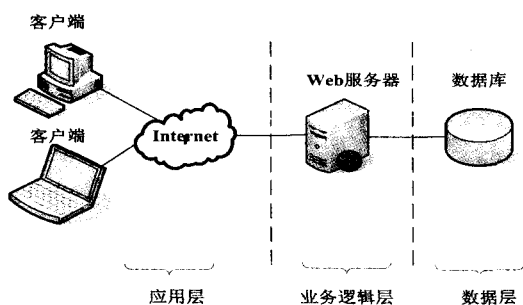


图 1 系统的结构图

## 1.2 系统的开发模式

本系统采用浏览器/服务器(B/S)模式,由浏览器、Web 服务器和数据库组成,结合 C#语言、JavaScript 脚本语言、ASP、net、ADO、net、ODBC 等技术,Web 服务器采用 IIS,后台数据库采用 SQL Server 2000。

## 1.3 核心功能描述

(1)题库管理:组卷管理员和任课教师可以设置试题类型、管理章节知识点、录入试题以及对已有试题进行更新与维护,还可以查询和浏览试题。

(2)试卷管理:试卷管理功能又分为组卷管理和成卷管理两大子模块。组卷管理中主要的功能是自动组卷、手动组卷和模板管理。对于自动组卷模式,系统将根据模板中的难度和知识点分布等组卷参数,调用组卷算法自动组卷,组卷完成后以模板要求的格式将试卷和答案导出到 Word 文档中。对于手动组卷模式则是由教师从题库中手动抽出符合自己的试题。成卷管理即试卷档案管理,可对已组成的卷子进行查询、分析等管理。

(3)角色及权限管理:本系统中分四类角色,分别是:系统管理员、组卷管理员、任课教师及学生。系统管理员由系统生成,拥有对各类用户管理的权限及对本系统的数据备份和恢复等管理权限;任课教师对题

库、已考试卷进行管理;组卷管理员对组卷进行管理;学生则可进行在线章节自测和在线考试。

(4)测评管理:包括评分和成绩管理两个子模块。评分管理是对试卷进行评分并写入数据库。成绩管理实现成绩的查询、统计分析等功能。

(5)学生测试:包括自测和在线考试。学生可以利用自测功能进行章节自测,并可将做错过的题放入错题集,以便进行复习和总结;在线考试功能则是根据后台已设定好的模板,调用自动组卷算法,生成在线试卷以供考生使用。

## 2 系统关键问题及解决方案

### 2.1 题库系统理论基础

在考试系统建设过程中,优良的组卷算法固然重要,但一份好的试卷的生成,必须以良好的试题质量控制与合理的试题结构为前提。

所谓题库<sup>[6]</sup>是由许多适用于不同目的、技能和任务考核需求,且由大量优质题目组成的某个学科试题集合。传统的题库往往只是收集大量试题以备选择使用,对试题本身并无太多要求,试卷质量无法得到有效控制。因此,为了保证试卷质量,题库系统中试题属性的计算与评估要以合理的理论基础作为指导。

目前对题库进行评测的教育测量理论有:经典测量理论<sup>[7]</sup>和项目反应理论。其中,经典测量理论体系成熟,理论假设较弱,对实施条件要求低,适用性广,以其为基础的测验在各行各业得到了成功的应用。

经典测量理论对题库系统的评测主要分成两方面:项目分析和整体分析。

(1)项目分析是针对试题进行分析。利用试卷的测验结果对每道试题属性(如:难度、区分度等)进行统计与分析。

(2)整体分析是对整个试卷进行分析和评价。为了考查试卷质量,利用试卷的测验结果对整个试卷从难度、区分度、信度和效度方面进行分析,构成了整个试卷的信度和效度。

### 2.2 题库结构设计

本系统以经典测量理论为基础建立题库。经典测量理论有一套完整的题目分析指标体系和评价标准,要求进入题库的每一试题,除了试题本身的内容外,还要包括题目的编号、知识点、难度、效度、区分度、使用情况等多种定性、定量的属性。这样既可以利用这些参数对试题进行分类管理,同时也能有效地约束入库的试题质量。

按照属性对试题质量的约束力的不同,可将题库中试题的属性分为两类:性能属性和内容属性。试题的内容属性主要是指试题的内容及答案等信息,这些

信息不作为组卷算法中的约束参数。性能属性则要做为组卷算法中的约束参数参与到组卷过程中,这里主要对影响组卷算法的性能属性进行讨论研究。通过对不同考核需求、组卷算法及成卷速度要求的分析,以下给出了对试卷质量起关键作用的主要性能属性:题型、难度、知识点和出题频度<sup>[8]</sup>。

(1) 题型。题型是指试题类型。可分为填空题、选择题、编程题等。并可根据每次组卷的需要,动态增加新的题型。

(2) 难度。难度<sup>[9]</sup>是衡量试题难易程度的指标。在使用计算机进行自动组卷过程中,试题难度分布设计得是否合理至关重要,若一份试卷使学生成绩普遍都很高,或是普遍都很低,都是失败的。因此,必须合理地设计试卷中试题的难度分布,使其达到或是基本达到组卷的预期要求。

在经典教育测量理论中,对于每道试题的难度是使用通过率来表示。例如:

① 对于客观性试题,其计算公式为:

$$D_i = \frac{C_i}{Z_i}$$

其中,  $D_i$  是第  $i$  题的难度;  $C_i$  表示答对第  $i$  题的人数;  $Z_i$  表示参加第  $i$  题考试的总人数。

② 对于主观性试题,其计算公式为:

$$D_i = \frac{P_i}{M_i}$$

其中,  $D_i$  是第  $i$  题的难度;  $P_i$  表示第  $i$  题的平均得分;  $M_i$  表示第  $i$  题的总分。

由此可看出,  $0 \leq D_i \leq 1$ , 而且  $D_i$  越小, 试题越难,  $D_i$  越大, 试题越简单。本系统难度采用五级分档, 即  $D_i = \{ \text{难, 较难, 中等, 较容易, 容易} \}$ , 并将此五级分档通过五个连续的难度等级分段来表示, 依次为  $[0, 0.2]$ 、 $[0.2, 0.40]$ 、 $[0.40, 0.60]$ 、 $[0.60, 0.80]$ 、 $[0.8, 1]$ 。

(3) 知识点。可用十进制的整型类型字段表示知识点信息, 例如: 用 6 位十进制数表示时, 可以用前两位表示章, 中间两位表示小节, 最后两位表示每小节中存在的不同知识点。在自动组卷时, 可以根据这一参数精确对知识点进行选取, 可避免同一知识点在同一试卷中出现多次的现象, 以保证知识点在试卷中尽量分布均衡。

(4) 出题频度。用于统计试题库中, 某试题在实际试卷中出现过的次数。通过该参数可以控制题目在试卷中出现的概率。每次组卷前, 将所有试题的频度字段累加并求平均。抽题时, 按频度由小到大的顺序, 优先选取频度小于等于平均数的试题。

### 2.3 组卷算法

目前组卷算法主要有随机抽取法、回溯试探法等,

各种算法特点不同, 同时也存在不足之处。

(1) 随机法。即随机从试题库中提取试题。其优点是: 算法简单易行、运行速度快; 缺点是: 由于约束条件少, 组卷中试题重复率、相似度较高, 试卷的总体质量难以控制。

(2) 回溯试探法。将随机产生的每一道题状态记录下来, 当搜索失败时回溯到上次记录的状态, 然后再根据已定的规律变换出一种新的状态进行新的试探, 通过不停地回溯、试探直到试卷生成完毕。该算法在理论上可以遍历所有可能的组合状态, 但实际上当试卷总题量较大时, 状态的组合数将成为天文数字, 使得算法难以实施下去, 导致最终组卷失败。该算法存在着内存占用量大、组卷效率低、组卷失败率高等缺点, 只能适用于约束参数和试卷总题量不多的题库, 显然不适合在线组卷。

本系统组卷算法采用的是随机搜索逐步优化算法。随机算法的特点是运行速度快、占用内存少、试题重复率低、知识点分布均匀、难度均衡、组卷成功率高, 尤其适合组卷时需要在线等待的情况。

本系统组卷算法的基本思想是设计一随机函数  $\text{random}()$ , 对题库随机抽取试题, 然后匹配用户录入的组卷要求, 若匹配, 则记录试题编号; 否则, 放弃该试题, 反复此过程直到达到试卷要求的题量, 即组卷完毕。

算法的伪代码如下:

```

1. input(题型类型、大题型数量、小题型数量、整卷难度)
2. for n=1 to 大题型数量
3. for m=1 to 小题型数量
4. call random() //调用函数抽取此种题型的最近未使用或使用频度少的试题
5. if 知识点 ∈ {已抽到的题目所属的知识点}
6. then
7. 此题被抽取
8. m=m+1
9. endif
10. endfor
11. DX=sum(各小题难度系数)//调用函数求各小题难度系数之和
12. if DX>整卷难度+0.05 or DX<整卷难度-0.05
13. then
14. if DX>整卷难度
15. then
16. 淘汰已取题中难度最大的(设此题难度为 DMAX), 从题库选出与此
17. 题知识点相同, 难度最小的题或是难度在 DMAX-DX-TD 范围的题替换
18. endif

```

18. if  $DX < \text{整卷难度}$

19. then

20. 淘汰已取出题中  $D_i$  最小的题(设此题难度为  $DMIN$ ),从题库选出与此知识点

21. 相同,难度最大的题或是难度系数在  $DMIN+TD-DX$  范围的题替换

22. endif

23. endif

24.  $n = n + 1$

25. endfor

为了验证算法性能,采用 10~30、30~50、50~70 的 3 组不同的抽题数进行 30 次的测试,仅有一次不成功,成功率为 96.67%,重复题出现 2 题。抽取试题的难度与出题者的试题难度要求相差在 3.5% 内,满足允许的误差范围。由以上测试可知,组卷算法设计科学,组卷合理,系统稳定可靠,完全能满足教学要求。

## 2.4 程序题评分算法

在评分系统中,常用程序题评分算法主要有两类<sup>[10]</sup>:动态测试法和程序特征分析法。这两种算法特点不同,同时也存在不足之处。

(1) 动态测试法的基本思想<sup>[11]</sup>是:先将源程序编译为可执行文件,然后输入测试数据集,将运行结果与答案进行对比评分。动态测试法的优点是简单易行。其缺点是仅通过结果进行评分,忽视考生程序的内部结构特征,0 分或满分的极端现象较多,存在很大的不合理性。

(2) 程序特征分析法的基本思想<sup>[12]</sup>是:通过对程序特征(如:代码行、变量个数以及复杂度等)的分析评价学生程序的质量,给出评分。其优点是度量参数容易计算;其缺点是忽视考生程序的内部结构特征,评分结果不够准确、公平。

为了克服以上两类程序题评分算法的缺陷,文中提出的程序题评分算法将从程序的语法、语义角度衡量学生程序与正确答案的接近程度,从而给出评分。算法的基本思想是:对学生提交的程序首先进行编译,如果编译成功,则运行程序,如果运行成功则输入测试数据集进行结果对比评分。如果运行不成功,如存在死循环等问题,则采用正则表达式对程序代码进行分析,按得分点给出相应的分数。如果编译程序不成功,则借助算法进行错误修正,如果错误修改后可正常编译运行,则根据存在的错误情况适当扣分;如果错误修正后程序仍不能编译运行通过,则采用正则表达式对程序代码进行分析,按得分点给出相应的分数。程序题评分算法流程图如图 2 所示。

为了测试算法的性能,采用 5 道编程题,以 80 份程序进行实验。对于正确的程序,能 100% 的准确评分;对于存在错误的程序,评分结果与人工评分标准相

比较,平均有 91.4 % 的程序的评分准确率大于 95 %,具有较高的准确性,避免了不合理的极端分数的出现。

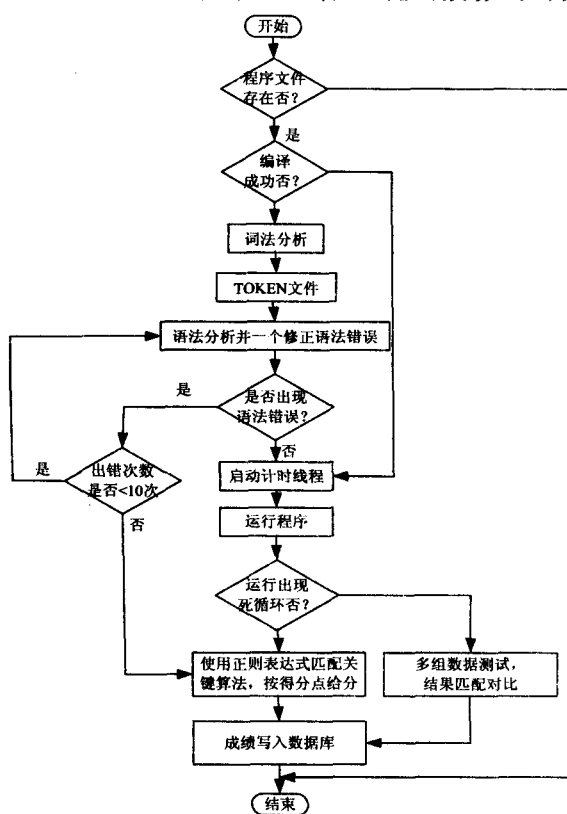


图 2 程序题评分算法流程图

## 3 结束语

文中在分析和设计了《数据结构》试题库结构的基础上,给出了随机搜索逐步优化的组卷算法和基于编译技术的程序题评分算法。经过对系统的测试比较可知,试题库设计合理,试卷成卷、试卷评阅质量较高,对于实现无纸考试具有重要的实际应用意义和推广价值。

### 参考文献:

- [1] 邓霖,戴敏. 数据结构试题库管理与组卷系统的开发[J]. 天津理工学院学报, 2003, 19(2): 83-86.
- [2] 赵耀红. 数据结构组卷系统的设计与实现[J]. 长春大学学报, 2010, 20(12): 26-29.
- [3] Morris D S. Automatic grading of student's programming assignments: an interactive process and suit of programs[C]// Proc of the 33rd ASEE/IEEE Frontiers in Education Conf. Piscataway, NJ: IEEE, 2003: 1-6.
- [4] 赵晓,刘学锋,王晓霞. 编译原理理论在 C 程序题自动评分系统中的应用[J]. 陕西科技大学学报, 2010(6): 135-137.
- [5] 陈荟慧,张新颜. 在线考试系统的分析与设计[J]. 计算机技术与发展, 2009, 19(10): 214-221.

序来实现系统新增硬件设备的控制。

驱动程序位于操作系统的内核和硬件设备之间,是实现两者通信的中介,可为应用程序提供设备接口。这样,操作系统对设备的操作即可转换为对映射为设备文件的文件打开(open)、读(read)、写(write)和关闭(release)等操作。

温度传感器 DS18B20 对温度的采集是通过温度对振荡器频率的影响来实现测量操作的。因此,可通过 ARM2440 控制核心对 DS18B20 进行设备文件的操作而完成:

1) 初始化设备总线;

2) 跳过总线上的 ROM,启动 DS18B20 并进行温度转换;

3) 复位;

4) 跳过 ROM 并发出读取温度指令;

5) 读取温度数据,进行温度数据处理和转换;

6) 返回。

该过程的伪代码实现如下:

```
void ds18b20Read(void) {
    while(ds18b20Reset()); //复位温度传感器
    udelay(120); //时间延迟
    ds18b20WriteByte(0xcc); //利用指令码跳过 ROM
    ds18b20WriteByte(0x44); //发出温度转换命令
    .....
    ds18b20WriteByte(0xbe); //发出读取温度的指令
    data[0] = ds18b20ReadByte(); //读取低位处温度数据
    data[1] = ds18b20ReadByte(); //读取高位处温度数据
    .....
}
```

湿度传感器、摄像头的初始化和数据采集控制步骤与温度传感器的控制过程类似,文中不再赘述。

值得说明的是,由于在 ARM 嵌入式开发板上已经具有了 Qt 基本环境,故而可直接利用 Qt 提供的 C++ 类和函数,如 QWSKeyboardHandler 的 processKeyEvent()、processKey-code() 等,快速实现触摸屏的键盘、鼠标输入事件。这也正是引入 Qt 应用程序框架的重要意义所在。

## 5 结束语

图书馆特种文献典藏库监测系统以三星 S3C2440 微处理器 ARM9 嵌入式开发板为控制核心,使用 NOKIA 公司最新的 Qt Embedded for Linux 嵌入式开发环境,集成温度、湿度传感器和摄像采集设备,既有效满足了文献典藏库半封闭环境要求的实际需要,又充分降低了系统的整体功耗,使其能够满足文献典藏库的不间断监测需求。系统在软件选用上使用了开源的非商业软件,因而具有较好的灵活性和可移植性,更为文献典藏库的管理工作提供了有效的科学平台。

### 参考文献:

- [1] 单广荣. 藏文古籍数字图书馆设计研究[J]. 计算机工程与设计, 2009, 30(1): 255-258.
- [2] 杨彦辉. 客家文献数据库管理系统建设探讨[J]. 图书馆论坛, 2007, 27(3): 67-69.
- [3] 闫月香, 谢峰. 基于汇文系统的书标自动生成功能的改进[J]. 科技情报开发与经济, 2005, 15(12): 221-222.
- [4] 钱承军. 论高校图书馆纸质文献保护工作[J]. 河南图书馆学刊, 2007, 27(2): 65-67.
- [5] 中华人民共和国文化部. 图书馆古籍特藏书库基本要求(WH/T24-2006)[S]. 2007.
- [6] Kobayashi N. [http://www.nlc.gov.cn/newtsgj/ifylaygt/gjtlb-cbhxx\\_zgxx/hydt/hy/zrhhhyzt/hylw/201011/t20101130\\_21900.htm](http://www.nlc.gov.cn/newtsgj/ifylaygt/gjtlb-cbhxx_zgxx/hydt/hy/zrhhhyzt/hylw/201011/t20101130_21900.htm). 2012-01-02.
- [7] 蔡锦达, 李玉路. 基于 ARM 的图书管理系统的研究与实现[J]. 微计算机信息, 2011, 27(9): 108-109.
- [8] 李自强, 张小波, 周光辉, 等. 基于 ARM 的移动图书导航系统[J]. 软件, 2011, 32(9): 26-30.
- [9] NOKIA. [http://developer.qt.nokia.com/wiki/Support\\_for\\_Embedded\\_Linux\[EB/OL\]](http://developer.qt.nokia.com/wiki/Support_for_Embedded_Linux[EB/OL]). 2012-01-06.
- [10] Blanchette J, Summerfield M. C++ GUI Qt 4 编程[M]. 闫锋欣, 曾泉人, 张志强译. 第 2 版. 北京: 电子工业出版社, 2008.
- [11] 周育红, 闫锋欣. 数字资源跨平台整合系统的设计与实现[J]. 计算机技术与发展, 2010, 20(1): 243-246.
- [12] 张敏, 张井岗. Linux 平台下支持触摸屏的 Qt/Embedded 移植[J]. 计算机技术与发展, 2011, 21(10): 154-156.

(上接第 184 页)

- [6] 刘博. 智能教学系统中个性化题库的设计与实现[J]. 中国电化教育, 2010(9): 110-114.
- [7] 洪飞. 经典测量理论与项目反应理论的比较研究[J]. 社会心理科学, 2006(6): 15-17.
- [8] 殷晓玲, 夏启寿, 范训礼. 网络考试系统中成卷模式分析研究[J]. 计算机技术与发展, 2009, 19(2): 205-212.
- [9] 郭敏华, 兰雨晴, 金茂忠. 一种海量题库难度系数和区分度动态优化模型[J]. 计算机工程与应用, 2007, 43(21): 246-248.
- [10] Ala-Mutka K. A survey of automated assessment approaches for programming assignments[J]. Computer Science Education, 2005, 16(2): 83-102.
- [11] 李炜, 王喜乐. 网络教学中程序题机器批改的研究与实现[J]. 计算机技术与发展, 2011, 21(3): 229-235.
- [12] Ala-Mutka K, Uimonen T, Hannu-Matti J. Supporting students in C++ programming courses with automatic program style assessment[J]. Journal of Information Technology Education, 2004(3): 245-262.