

多线程技术的 Office 对象模型阅卷系统^①

吴建军

(浙江师范大学 行知学院, 金华 321004)

摘要: 一个采用多线程技术的软件可以更高效地利用系统资源。基于对象模型的 Office 文档阅卷准确实现了测试文档进行分析、比较, 从而实现自动评分。通过编程实践和应用测试, 多线程技术的 Office 对象模型阅卷是准确、稳定和高效的。

关键词: 多线程; Office 对象模型; .Net Framework 3.5

Multithreading Technology Office Object Model Scoring System

WU Jian-Jun

(Xingzhi College, Zhejiang Normal University, Jinhua 321004, China)

Abstract: A Multithreading Technology software can use the system resources more efficiently. Office document scoring which is based on the Object Model has accurately achieved to analyze and compare to the test documentation, in order to achieve automatic scoring. Multithreading Technology Office Object Model Scoring is accurate, stable and efficient through program practice and application test.

Keywords: multithreading; office object model; .Net framework 3.5

1 引言

从 2008 年秋季开始, 浙江省高校计算机等级考试增设了“二级办公软件高级应用”的等级考试。该考试项目, 以微软的 Office 2003 办公软件为主体, 要求考生掌握 Office 软件的高级应用, 实现掌握编排复杂的 Word 文档、制作功能全面的 PowerPoint 演示稿和 Excel 的高级函数应用等应用能力, 提升各类大学生的计算机应用能力, 适应学习与工作的需要。

作为一门新课程, “办公软件高级应用”(也称 Office 高级应用)引起了高校相关教学工作的高度重视, 各高校相继开设该课程, 进行对计算机应用能力教学的改革与创新。作为一门以计算机实践操作为主的课程, 最重要的实践练习和测评迫切需要合适的教学练习平台和评价工具。纵观当前的主流模式, 使用计算机辅助测评(Computer Assisted Assessment, 简称 CAA), 把计算机应用于教育测量和测评的全过程, 已经成为一种发展趋势^[1,2]。综上所述, Office 高级应用

课程教学的实施迫切需要一套为其定制的 CAA 系统。本项目正是以此为目标, 在已有的教学和科研的基础上, 结合省高校计算机等级考试的大纲要求, 设计完成了 Office 高级应用教学评测系统(软件), 已在浙大城市学院等多所高校推广使用。

面对难度较高的复杂 Office 文档阅卷, 准确、高效是设计的中心思想, 本文对项目中的多线程机制的 Office 对象模型阅卷进行研究与应用。

2 Office对象模型的自动阅卷

2.1 概况

Office 文档的阅卷方式主要有两种基本途径: 第一种途径是通过记录并分析考生的操作步骤来实现阅卷评分, 第二种途径是通过分析考生实际操作的结果文档来实现阅卷评分^[3]。构建被测软件的仿真系统是基于一第一种途径的阅卷方法, 但由于构建仿真系统工作量较大; 而且, 一旦出现新版本的软件系统, 又需

① 基金项目:2009 年度浙江省教育厅科研项目(理工科类)(Y200909441)

收稿时间:2010-08-12;收到修改稿时间:2010-09-06

要重建此仿真系统, 从而此方法适应性较差。而第二种途径是以文档的操作结果来进行评判, 鉴于 Office 文档的实际需求, 这种方法的适应性更广, 更能准确地评判文档是否达到了考试要求, 因此, 浙江省高校计算机相关能力考试中普遍采用了这种设计方法。作为第二种途径的 VBA(Visual Basic for Applications)技术, 出发点是作为宏语言使用, 但它并没有封装 Office 对象的所有属性和方法, 因而限制了 Office 操作的范围。

微软 Office 是基于 COM(Component Object Model)技术构建的软件系统^[4], 具有 COM 自动化功能。它可视为第二种途径的另一种方法。本项目即采用该方法, 并且使用微软的 Visual Studio 2008 开发平台, 基于 .Net Framework 3.5 环境, 能充分发挥微软开发平台的优势, 为软件开发提供可靠保障。采用 COM 技术的 Office 文档阅卷, 尤其对复杂性高的 Office 高级应用文档是非常合适的。它通过导入 Office 应用程序的对象库, 使用 COM 接口, 可编程操纵 Office 文档中的所有对象, 按考题需求提取、分析相关子对象及属性, 获得正确的阅卷信息。

本课题的 Office 评测软件主要由 Word、Excel 和 PowerPoint 共三个模块组成, 以 Visual Studio 2008 为软件开发平台, 本文以 Word 模块为例进行阅卷分析与设计。

2.2 Word 对象模型

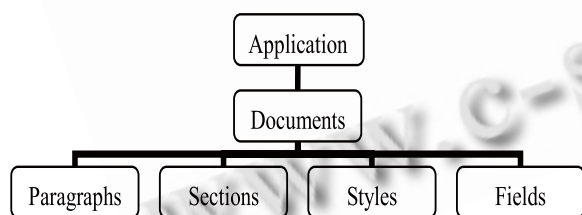


图 1 Word 文档对象

在 Visual Studio 2008 的 .NET Framework 3.5 技术框架下, 引用 Office 2003 办公软件的 COM 组件进行开发, 这是本项目实现自动阅卷的基本开发构架。在该框架下, 采用 COM 技术, 以 Office 对象模型为基础, 获取 Office 层次对象模型的相关属性, 实现软件的自动测评功能。Word 文档对象的基本组织(仅小部

分)如图 1 所示。

Application 对象代表了 Word 应用程序, Documents 是指由当前 Word 应用程序打开的所有 Document 对象所组成的集合。在 Document 对象集合下, 包含了 Paragraphs(段落)、Fields(域)、Sections(节)和 Styles(样式)等对象集合, 它们是 Word 阅卷中的主要对象, 通过对测试文档的这些对象属性值进行访问, 将获取的访问值与答案标准进行比较, 即可实现对测试文档的阅卷。

2.3 Word 阅卷实例

Word 对象模型阅卷的程序设计流程如图 2 所示。

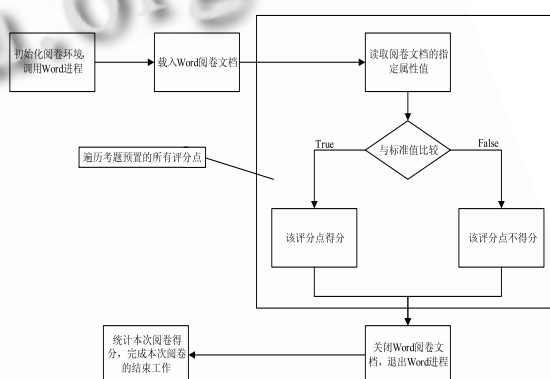


图 2 Word 对象模型阅卷流程图

2.3.1 创建 Word 阅卷对象

假设, 文档“C:\Test\Word\Word.doc”是测试文档, 对其创建阅卷对象, 基本代码如下。

```

Public myWord As Word.Application
Public myDoc As Word.Document
myWord = CreateObject("Word.Application")
myDoc = myWord.Documents.Open("
C:\Test\Word\Word.doc")
  
```

myDoc 对象即代表了测试文档, 访问它下属的对象集合的属性值, 即可准确地评判考生对文档的操作是否达到了考试要求。

2.3.2 Word 阅卷设计

Office 高级应用的考题比较复杂, 但设计阅卷代码的方式是相似的, 都是以测试文档对象的相关属性值与答案标准对照的结果作为得分依据。例如, 考题要求: 使用域的方法添加正文的页眉, 对于奇数页, 页眉中的文字为“章序号”+“章名”。其对应的程序

设计如下:

(1) 声明字符串变量 oddHeader 存放测试文档的相关属性值, oddHeaderkey 预置该测试点的答案标准。

```
Dim oddHeader As String = ""
```

```
Dim oddHeaderkey As String = "STYLEREF " +  
Chr(34) + " 标题 1" + Chr(34) + " \n \\  
MERGEFORMAT" + "STYLEREF " + Chr(34) + "标  
题 1" + Chr(34) + " \\  
MERGEFORMAT"
```

(2) 获取测试文档的指定值。

```
oddHeader=myDoc.Sections(i).Headers(Word.WdH  
eaderFooterIndex.wdHeaderFooterPrimary).Range.Fields  
(1).Code.Text+myDoc.Sections(i).Headers(Word.WdH  
eaderFooterIndex.wdHeaderFooterPrimary).Range.Fields(2  
) .Code.Text
```

(3) 评分

将变量 oddHeader 与 oddHeaderkey 比较, 产生相应得分。当完成所有测试点的评判后, 统计本次阅卷得分。

(4) 关闭测试文档

在阅卷完成后, 关闭测试文档、Word 进程是必要的, 代码如下:

```
myDoc.Save()
```

```
myDoc.Close()
```

```
Dim processList() As Process
```

```
processList=Process.GetProcessesByName("winwor  
d")
```

```
For Each proc As Process In processList
```

```
proc.Kill()
```

```
Next
```

除了关闭阅卷后的文档, 还采用了 For Each 结构对系统中可能未退出的 word 进程(winword)进行强制退出, 可避免在后续多次的操作或阅卷中因多个 word 进程存在于操作系统中可能带来难预测的影响。

3 多线程机制的阅卷设计

由于 Office 高级应用的考题复杂, 用本课题设计的方法完成一套试题的阅卷大约需要一分多钟, 对学生练习和考试阅卷都带来了明显的不利影响; 同时, 单线程程序在阅卷中对用户与应用程序交互的不响应也影响了使用效果, 学生练习时的阅卷出现的长时间

不响应用户操作, 甚至会让某些学生误以为程序死机。由于考题的阅卷主要由 Word、Excel 和 PowerPoint 三个独立模块组成, 如果在阅卷中能使用多线程技术, 则能对阅卷速度产生积极的影响。

3.1 Net 3.5 中的多线程技术

在 .Net Framework 3.5 环境中, 进程用于描述当前运行的应用程序使用的一组资源(如外部代码库和主线程)以及必要的内存分配^[5]。每个 Windows 进程都确切地有一个“主线程”, 该线程的作用是应用程序的入口点, 进程的入口点创建的第一个线程成为主线程。包含一个执行主线程的进程在本质上是安全的, 但可能对用户的交互不灵敏。对于单核 CPU 而言, 只有一个 CPU 的机器不能够同时确切地处理多个线程, 多线程的进程提供了大量活动差不多同时发生的假象, 额外的开销使可能会使单 CPU 系统上的应用程序总体执行变慢。

不过, 多 CPU 系统可以在不同的处理器上执行不同的线程, 多线程的程序就会运行更快, 因为每个线程可以在单独的处理器上运行^[6]。当前的个人计算机普遍采用了多(双)核 CPU, 这为本项目的设计提供了广泛的硬件支持。

多线程编程比单线程复杂和不直观, 它要求线程之间进行额外的协调和安全通信。为此, Visual Studio 2008 提供了 BackgroundWorker 组件, 组件利用了与异步委托基本相同的线程化语法, 在后台执行并调用 RunWorkerAsync 事件^[5]。调用线程(主线程)继续正常运行, 可响应用户操作, 而工作者线程则异步运行。当工作者线程完成操作时, BackgroundWorker 组件就通过激活 RunWorkerCompleted 事件通知主线程, 它使多线程应用更容易和更安全。

3.2 多线程机制的阅卷流程设计

对于单线程的阅卷系统, 其阅卷流程如图 3(a)所示, 在阅卷中, 主线程依次进行各个模块的阅卷, 并且不响应用户与软件的交互, 因而造成阅卷过程较慢、界面不够友好等问题。

而在图 3(b)的多线程阅卷设计中, 由主线程同时开启了各个阅卷模块, 这些阅卷模块在多核 CPU 系统中被有效执行; 同时, 主线程监视各阅卷模块是否全部完成, 还可以同时响应用户对软件的交互。当各模块的阅卷完成后, 其对应的 BackgroundWorker 组件抛

出其阅卷完成标记,由主线程获得后,完成阅卷、进入成绩处理,最后结束本次阅卷。

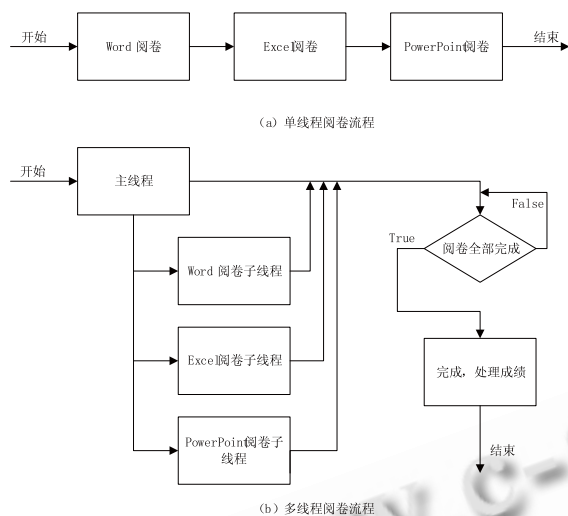


图 3 单线程和多线程阅卷流程图

3.3 多线程机制的阅卷程序设计

在程序设计中,使用 Visual Studio 2008 控件工具箱的 BackgroundWorker 组件分别为 Word、Excel 和 PowerPoint 阅卷模块设计线程对象 Yuejuan_word、Yuejuan_Excel 和 Yuejuan_Ppt。BackgroundWorker 组件的应用主要是 DoWork 和 RunWorkerCompleted 事件,前者是运行该线程对象的程序内容,也就是进行阅卷;后者则是在该线程对象完成时响应的事件,在程序中可设计一个逻辑值用以表示该线程已经完成。主要代码如下(包含了注释和省略的代码说明):

启动 word、Excel、PowerPoint 阅卷线程

```
Yuejuan_word.RunWorkerAsync()
```

```
Yuejuan_Excel.RunWorkerAsync()
```

```
Yuejuan_Ppt.RunWorkerAsync()
```

下面的 Yuejuan_word_DoWork 为 word 阅卷线程。Excel、PowerPoint 阅卷线程与此相似,省略。

```
Private Sub Yuejuan_word_DoWork(ByVal sender As System.Object, ByVal e As System.ComponentModel.DoWorkEventArgs) Handles Yuejuan_word.DoWork
```

```
word 阅卷, 代码省略。
```

```
End Sub
```

上面的 word 阅卷线程完成后,响应下面的 RunWorkerCompleted 事件。

```
Private Sub Yuejuan_word_RunWorkerCompleted (ByVal sender As Object, ByVal e As System.ComponentModel.RunWorkerCompletedEventArgs) Handles Yuejuan_word.RunWorkerCompleted
```

word 阅卷线程完成后,其阅卷完成标记置为 True。

```
YuejuanWordFlag = True
```

```
End Sub
```

下面是主线程的定时器监控 Timer_Yuejuan_Tick,定时扫描阅卷线程是否完成。

```
Private Sub Timer_Yuejuan_Tick(ByVal sender As System.Object, ByVal e As System.EventArgs) Handles Timer.Tick
```

当所有阅卷模块的线程都结束,则进入成绩统计、结束工作;否则,继续监视。

```
If YuejuanWordFlag And YuejuanPptFlag And YuejuanExcelFlag Then
```

```
'成绩统计、结束工作,代码省略。
```

```
Else
```

```
Exit Sub
```

```
End If
```

```
Timer_Yuejuan.Enabled = False
```

```
End Sub
```

上述代码的设计和运行,实现了多线程阅卷的过程。通过使用 Visual Studio 2008 的 BackgroundWorker 组件,创建了阅卷所需的多线程对象,以较简单、可靠的方式解决了多线程设计中的复杂问题。

4 性能测试和结果分析

4.1 测试模型

运行效果的测试是决定本项目是否适合采用多线程技术的关键。本项目以当前高校机房和学生计算机配置性能相似的平台为测试对象,选用了 CPU 为单核 P4(2.4GHz)、双核 T5600(1.86GHz)和 P8400(2.26GHz)等三个模型,内存均为 2GB,操作系统为 windows XP。其中,单核 P4 是配置中最低的,大约为 2005 年左右的机房计算机配置,后两个模型代表了学生用笔记本的类似性能,而当前学习机房主流配置的台式机 CPU

性能与 P8400 相似。

4.2 结果分析与应用策略

(1) 单套考题的阅卷比较

首先,以 T5600 模型对单套考题的单线程和多线程阅卷进行了比较,如图 4 所示。选用了五套不同的考题进行阅卷,从图中可见,多线程阅卷速度明显高于单线程阅卷。

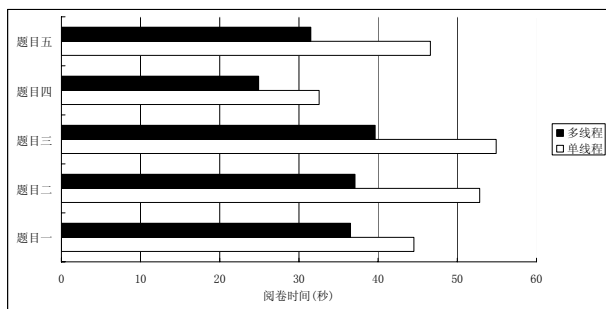


图 4 不同考题的阅卷速度比较

(2) 多套考题的阅卷比较

为了更显著地比较两种方式的阅卷速度,采用不同的硬件配置,对五套试题进行连续阅卷,其比较结果如图 5 所示。

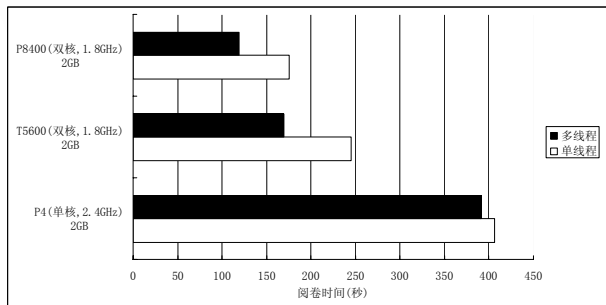


图 5 不同计算机配置的阅卷速度比较

(3) 应用策略

从上述的测试数据比较可以发现,CPU 对阅卷速度有显著的影响,在连续五套试题阅卷中,最快的 P8400 所需时间仅为 P4(2.4GHz)的三分之一;同时,

多线程方式有比单线程方式快出了约三分之一。可见,如果需要对大批考生进行批量阅卷,采用更先进的多核 CPU、多线程阅卷的优势明显。

同时,对较早的单核 CPU 来说,单、多线程阅卷的速度差别不大,这也是符合多线程理论的;但毕竟多线程可以在阅卷的同时,程序还可响应用户对程序的交互,采用多线程也是能更好地适应用户需要的。

5 结束语

综上所述,基于 .NET Framework 3.5 的多线程阅卷方式可以普遍地适应不同的计算机应用,它提高了阅卷速度,可构建更友好的人机交互界面。同时,基于 Office 对象模型的阅卷方式采用了与 Office 文档的最佳接口,可确保精确地获取文档属性,可靠地实现阅卷评分。本项目以此为基础,实现了办公软件高级应用阅卷软件的准确、稳定和快速运行。

参考文献

- 1 Brydges R, Sidhu R, Park J, Dubrowski A. Construct validity of computer-assisted assessment: quantification of movement processes during a vascular anastomosis on a live porcine model. The American Journal of Surgery, 2008,193(4):523—529.
- 2 Sim G, Read J, Holifield P, Brown. Heuristic Evaluations of Computer Assisted Assessment Environments. World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA). 2008: 896—905.
- 3 汤克明,陈峻.Word 自动阅卷系统的设计与实现.计算机工程与应用,2008,44(35),69—72.
- 4 Templeman J, Mueller JP. COM Programming with Microsoft .NET. Microsoft Press, 2003, Chapter 3.
- 5 Troelsen A. Pro VB 2008 and the .NET 3.5 Platform. Apress, 2008,(3):537—569.
- 6 Stephens R.侯普秀,曹俊译.Visual Basic 2005 设计与开发专家教程.北京:清华大学出版社,2008.