



# 计算机软考操作题自动阅卷技术实现路径的研究分析

蒋华锋<sup>1</sup>, 谢国技<sup>2</sup>

(1. 工信部教育与考试中心, 北京 100040; 2. 深圳市卓帆技术有限公司 广东 深圳 518000)

**摘要:**针对上机操作题人工阅卷效率偏低, 同时大量需要专业技术人员阅卷, 致使无法大规模推广上机考试的问题, 通过对目前具有代表性的机考操作题自动评分的方法的实例分析研究, 评估在进一步在计算机软考中推广上机考试的可行性。

**关键词:**计算机软考; 操作题; 自动评分; 自动阅读

中图分类号: G642.475

文献标识码: A

文章编号: 1671-931X (2021) 05-0108-09

DOI: 10.19899/j.cnki.42-1669/Z.2021.05.020

我国从 1991 年开始第一次组织全国性计算机软件著作权专业技术资格和水平考试(以下简称软考), 到如今在全国范围内已经实施了三十年。近十年来, 报考规模持续增长。即便是 2020 年受到了疫情影响, 全国报考人数仍超过了 40 万人。由此可见, 软考在国内的含金量和关注度之高。

软考热度的持续增长, 无形中也促进了中国软件水平的进一步提升。近年来, 随着计算机技术的不断发展, 计算机在教育领域中的应用也越来越广泛和深入。软考自身也正在探索和推广以计算机上机考试模式逐步取代传统纸笔考试。“进一步完善信息系统, 建设考试综合服务系统”更是列入了软考的“十三五”规划之中。同时, 国家“十三五”规划, 以及“中国制造 2025”“互联网+”“一带一路”大数据发展等战略机遇, 极大地加大了软考市场关注度, 为软考提出了新的要求。

软考设置了三个级别 27 个岗位的资格考试,

考查内容涉及 Windows、Office、C、C++、Java、.Net、Visual Basic、Photoshop、Flash、Dreamweaver、FrontPage、Python 以及数据库等多种软件。目前, 只有信息处理技术员考试进行了 Office 上机考试。全行业计算机应用水平的日益提高, 也使得以理论笔试为主的考试方式, 已不能完全检验考生的实际技术水平和业务能力。在软考中扩大参与机考的岗位范围, 已成为必然的趋势。

目前, 计算机上机考试在初级信息处理技术员考试中已应用多年, 面向客观题(填空题、选择题、判断题)、部分主观题(简述题)和基于 Office 的应用系统操作题的命题和自动阅卷评分技术已经趋近成熟。本文将 Office 操作题自动阅卷的设计和实现过程为前导, 总结和归纳适用于软考多岗位上机考试自动阅卷的设计思想。

## 一、自动阅卷评分设计思想

收稿日期: 2021-01-08

**作者简介:**蒋华锋(1983-), 男, 北京人, 工业和信息化部教育与考试中心考务处工程师, 研究方向: 职业资格教育考试信息化技术研究与应用; 谢国技(1979-), 男, 广东深圳人, 深圳市卓帆技术有限公司计算机软件著作权和水平考试项目信息系统研发负责人, 研究方向: 计算机操作题上机考试、自动评分研究与应用。

计算机上机考试的自动评分是机考发展的核心。其核心思想就是建立一套符合知识体系和学科岗位标准,能够满足考查目的的评分体系,如图1。

在评分体系中,将整个学科或岗位标准分拆为多个知识点,在对知识点进行考查时,可以根据考查,维度(如准确性、表达方式等)设置多个评分点。评分点是整个评分体系中的最小单元。在操作题的命题中,从考生操作形式的多样性的角度考虑,可以把对操作过程和目标结果文件的检查拆分成若干个

评分点,并根据评分点编制评分细和相应的分值权重<sup>[1]</sup>。

阅卷系统应从考生的作答结果中解析并获取评分所需关键数据,以其关键数据与评分参数值进行对比,评价贴近度。按照各评分点在整个试题得分中所占的比例,评定其小分,最后累加答案中各评分点的得分,并适当考虑考生答题时的操作方式所产生的误差,最后计算得出本试题的成绩。

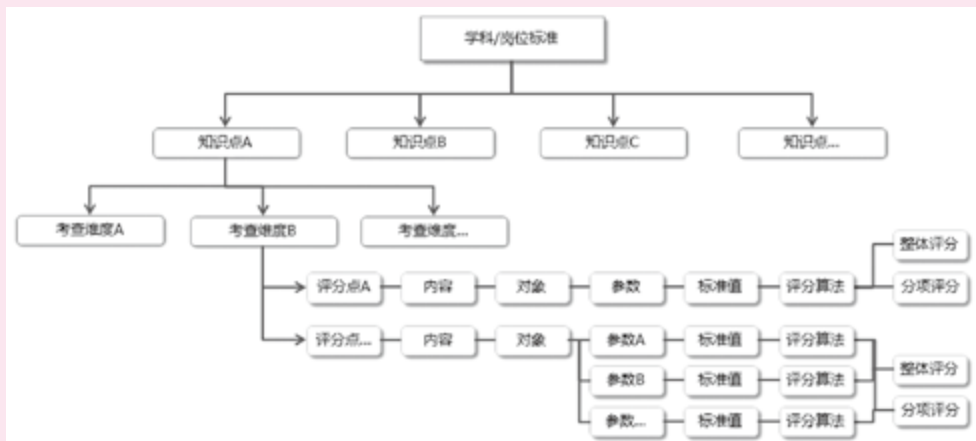


图1 评分体系模型

## 二、自动阅卷技术

随着软考的影响力在教育行业和全国的进一步扩大,以及软件技术的蓬勃发展,针对计算机上机考试的智能阅卷评分技术也得到长足发展。在客观题的自动阅卷评分方面,相关技术已经比较成熟。但操作题自动阅卷评分技术仍有所不足。

机考软件常见的自动阅卷评分原理,大致都是通过虚拟仿真技术、真实环境下的全程跟踪评测技术、基于对象的接口访问技术、基于文本文件的分析技术、基于二进制文件扫描技术等方法实现。

本文重点解析基于Office软件的系统自动阅卷评分的设计思想和实现过程,并探讨其他自动阅卷评分技术的实现方法。

### (一) 基于Office软件的阅卷技术

Microsoft Office是微软开发的一套办公软件系统,在国内各行业均有着非常广泛的应用,是各企业、教育单位必备的办公软件,也是个人软件市场占有率最高的常用软件。Office应用操作是计算机软考初级岗位的重要内容,在软考某些中级岗位的考试大纲中也有所涉及。

基于Office软件的阅卷评分技术主要包括VBA方式和Office Open XML方式,其中:

VBA(Visual Basic for Applications)方式是利用OLE技术通过Office应用程序提供的对外访问接口来访问目标结果文件的结构及相关属性<sup>[2]</sup>。使用

该技术的关键是被Office软件提供了OLE或COM访问对象,微软的Word、Excel、PowerPoint等Office类文档都可以通过这种技术来实现。VBA技术对Office软件的各个版本都具有良好的兼容性。

如:操作目标Word文档,当执行改变文档中指定文字的大小并保存后,就可以通过类似以下的VBA代码来获取目标Word文档指定文字的大小:

```
Set wd = CreateObject("Word.Application") // 创建 Word 对象
```

```
wd.Documents.Open("Target.doc") // 打开目标结果文件
```

```
wd.Application.Selection.Font.size // 获取选择字体的大小
```

Office Open XML方式是通过Open XML SDK提供的类库对Office目标结果文件内容进行解析,从中提取关键信息和属性实现与评分参数的对比。Office 2007版开始支持Office Open XML标准。Office 2007以前的版本仍需要依赖VBA技术进行文件解析。<sup>[3]</sup>

如上例中,使用Office Open XML方式获取指定文字的大小:

```
Public void GetJC()  
{  
    string filename=@" D:\Documents\Target.docx"; // 文档路径
```

```
using (WodrpocessingDocument myDocument =
```

```
WordprocessingDocument.Open(fileName,true))
{
    Paragraph p = myDocument.MainDocumentPart.
Document.Body.Elements<Paragraph>().First();// 获取文
档第一段段落对象(段落元素)
    pFont = p.font; // 取段落字体
    pFontSize = pFont.size; // 取字体大小
    Console.WriteLine(pFontSize); // 控制台输出
}
}
```

无论 VBA 技术,还是 Office Open XML 技术,都是通过 Office 开放的接口或类库建立 Office 文档模型,并获取文档属性值与评分点的评分对象的标准属性值进行比较,最终根据评分标准中设置的分值给出评分。考生可以真实环境下作答,专家在试题编制时也不需要考虑操作上的局限性,可以有很大的命题空间。

## (二)其它类型的自动阅卷技术

### 1. 虚拟仿真技术

虚拟仿真技术的原理是模拟被考查软件(或程序)的环境界面及部分功能,考生作答时在模拟软件中进行操作,机考系统通过记录操作者的操作痕迹与试题评分标准进行对比,对操作者操作步骤和结果进行评判。

虚拟仿真技术适用于可能会对电脑运行环境产生影响场景的操作题考试,如 Windows 操作、网络技术。这一类的操作题会改变操作系统及网络的运行配置,可能影响到计算机系统的正常运行(如蓝屏、网络中断等),采用虚拟仿真技术,让考生在虚拟环境中作答,作答结果不会影响到真实的电脑环境。

例如,模拟出 Windows IP 地址的设置程序。考生在作答时,由考试软件自动调出模拟的 IP 地址设置框,然后考生按照题目的要求在模拟的 IP 设置框中进行作答。作答过程中,模拟软件自动记录考生的操作步骤及操作结果,最后通过这些操作步骤与操作结果实行自动阅卷。

虚拟仿真技术相比真实环境的评测,需要去模拟被考查软件的真实运行环境及功能。在虚拟环境中进行操作不会影响到真实环境,“安全可控”是虚拟仿真技术最突出的优点。虚拟仿真技术也有比较大的缺点:只能对做了模拟开发的相关知识点操作进行考查,命题开放性、评分灵活性及知识点覆盖度都比较有限;升级的代价太高,本身去模拟被考查软件的真实运行环境及功能,软件开发、测试的工作量就比较大,如果被考查的软件版本发生较大升级时,比如 Office 2003、2007、2010 这三个版本,操作方式、操作界面有较大的不同。那么就需要开发新版本的模拟环境,代价就非常巨大。

### 2. 全程跟踪评测技术

真实环境下的全程跟踪评测技术的原理是通过

监视程序(即采用钩子技术)全程监视考生对被测软件的操作,并将操作步骤记录下来,然后通过分析步骤记录实行自动阅卷。

比如,通过监视程序监视考生对 Windows 画图软件的操作,可以记录到以下类似的信息:14:32 分,考生使用矩形工具在画布中画出了 69mm\*96mm 的图形,图形的左上角坐标为 67,158……,有了这些信息,实行自动阅卷就比较容易了。

钩子的本质是一段用以处理系统消息的程序。通过调用,将其与指定的系统挂接,截取系统的内部消息。钩子的种类有很多,每种钩子可以截获并处理相应的消息。每当特定的消息发出,在到达目的窗口之前,钩子程序先行截获该消息、得到对此消息的控制权。

由此可见,真实环境下的全程跟踪评测技术是一种基于操作过程的考查方式,通过监视程序记录考生的操作过程用来作为阅卷的依据。比较适用于考查考生软件操作步骤、操作熟练度和反应能力的考查场景。

全程跟踪评测技术的优点是,考生在真实环境下作答,操作没有局限性,阅卷评分不需要特定的算法,处理起来相对比较简单。但全程跟踪评测技术也存在其特有的缺点:阅卷程序设计起来比较复杂。因为考生在真实环境中作答,往往可以通过多种操作方式来达到操作要求,如要求给 Word 文档中的文字加粗,考生可以通过菜单设置文字格式为加粗,也可以直接按快捷键 Ctrl+B 进行加粗,还可以通过格式刷将目标字体刷成加粗样式等等。所以,在评分点的评分标准时,需考虑到考生所有可能的操作方式。否则,就有可能导致考生作答了,但不能得分的情况;另外,监视程序可能因电脑环境的影响停止运行(某个 DLL 版本冲突)或被阻止运行(被杀毒软件误认为木马)。监视程序所使用的技术与传统“木马”及“病毒”非常相似,绝大多数的杀毒软件都有可能将监视程序阻止。

### 3. 基于文本文件的分析技术

基于文本文件的分析技术的原理是将被考查软件生成的目标结果文件直接以文本方式打开进行解读分析,实行自动阅卷的关键就在于如何分析文本文件。

基于文本文件的分析技术适用于像 FrontPage、Dreamweaver、VB、Java 等这类程序填空题等类型软件的面向结果的考查场景。早期 Office 版本也通常通过 RTF 技术将 Office 文档转换成 RTF 文本文件进行内容解析。文本文件方式不能考查考生的具体操作过程,只能对最终目标结果文件进行自定义评判。

基于文本文件的分析技术可以有效的解决真实环境下的全程跟踪评测技术及基于对象的接口访问技术的缺点,但同时也有其弊端:首先是同一操作下生成的目标结果文件不唯一,阅卷程序实现逻辑



比较复杂,比如在 FrontPage 软件中,要设置某些字体的颜色为红色,生成的目标结果文件有可能是在 <Font> 标签中进行体现,也有可能在 CSS 中样式中体现等等,为了兼容这两种截然不同的目标结果文件,在阅卷程序在设计及实现上就必须考虑各种可能的情况,否则就会导致考生作答正确了但不得分;另外就是大多数被考查软件产生的目标结果文件都是特有的格式,基于商业保密目的,这种格式一般不对外公开,所以也就难以使用这种技术实现自动阅卷。

#### 4. 基于二进制文件的扫描技术

基于二进制的扫描技术的原理是通过对被考查软件生成的操作记录文件和目标结果文件的研究分析,建立文档模型,并通过与评分标准的贴进度分析给出评分。与文本文件分析技术一样,实现自动阅卷的关键就在于如何分析被考查软件产生的操作记录文件和结果文件。

基于二进制的扫描技术的自动阅卷处理流程如下:

##### 步骤 1:生成扫描文件

阅卷子系统从考试系统中接收到考生作答的目标结果文件后,从题库数据库中读取该试题对应的评分标准数据。逐条根据评分标准中的评分点编码从扫描代码文件中读取的对应的扫描代码,写入考生作答结果文件所在的目录下的扫描代码文件中。

##### 步骤 2:执行扫描代码文件

阅卷子系统执行扫描代码文件自动解析并执行扫描代码,以生成字符串形式的执行结果,并将执行结果保存在目标结果文件所在的目录下关键数据文件中,该关键数据文件的内容为字符串形式。

##### 步骤 3:产生阅卷结果

阅卷子系统在执行自动阅卷时撕开执行扫描后生成的关键数据文件,以试题设定的评分点逐一和评分标准进行对比分析。判断作答目标结果文件所涉及的各评分点与标准答案的贴进度和准确性,最后根据评分标准设置的分值比例计算分值,得出阅卷结果。

基于二进制文件的扫描技术适用于像 Flash、Photoshop 等这类没有提供对外的 OLE 或 COM 访问对象,没有办法直接应用基于对象的接口访问技术<sup>[4]</sup>;生成的目标结果文件是二进制格式,且其文件格式未对外公开的,既要面向结果,又要面向操作过程的考查场景。由于一般的软件都支持对结果文件格式的兼容,因此,软件版本变更对结果文件分析的正确性影响的风险极小。其缺点就是文件解析方法复杂,阅卷系统的开发、测试、发布所需要的周期较长,开发成本也较高。

#### (三)几种自动阅卷技术的对比

如今常用的自动阅卷技术一般有六种:VBA、Open XML、虚拟仿真技术、全程跟踪评测技术、基于文本文件的分析技术、基于二进制的扫描技术。其特点及优缺点对比如表 1。

表 1 几种自动阅卷技术的对比

序号	阅卷技术	特点	适用场景	考试环境	考查对象	优点	缺点
1	VBA	使用 OLE 技术和 Office 提供的接口,访问目标结果文件的结构及属性。	Office 软件	真实环境	操作结果	对 Office 有良好的兼容性	过度依赖 Office
2	Open XML	通过 Open XML SDK 提供的类库对目标结果文件进行解析。	Office 等开放文档结构标准的软件	真实环境	操作结果	对 Office 等开放文档标准的软件有良好的兼容性和稳定性	不支持 Office 2007 以前的版本;不支持没有开放文档标准的软件
3	虚拟仿真技术	模拟被考查软件的界面及部分功能,考生在模拟环境中操作和作答。	对考试机运行环境会产生影响的操作考试	虚拟环境	操作过程	操作过程不会影响到真实环境	只能对做了模拟开发的知识点做考查;升级的代价太高
4	全程跟踪评测技术	采用钩子技术全程监视考生对被测软件的操作,通过分析步骤记录实现自动阅卷	考查考生软件操作步骤、操作熟练度和反应能力的考试	真实环境	操作过程	阅卷评分不需要特定的算法,实施简单方便	阅卷设计需要穷举所有可能的操作;容易被防病毒软件拦截
5	基于文本文件的分析技术	目标结果文件直接以文本方式进行解读分析	程序填空题类型软件	真实环境	操作结果	对文本直接解析,不需要接口和文档格式	同一操作下生成的目标结果文件不唯一
6	基于二进制的扫描技术	对操作记录文件和目标结果文件进行研究分析,建立文档模型,并扫描关键值	未提供接口和文档标准,结果文件是二进制格式的软件	真实环境	操作过程 操作结果	不易受软件版本变更影响,兼容性强	文件解析方法复杂;开发成本高

### 三、自动阅卷评分流程

计算机上机考试的自动评分过程包括形成试题制作、考生作答和自动阅卷三个主要步骤。

#### (一) 形成电子试题

由命题专家命制考题并形成电子试题,包括试题描述、操作素材及试题评分标准。

#### (二) 考生作答

考生根据试题的要求进行作答,生成目标结果文件。

#### (三) 自动阅卷

计算机根据评分标准及目标结果文件进行自动阅卷并报告评分结果。

据此,上机考试自动阅卷涉及了三个子系统的

功能设计:题库子系统、考试子系统和阅卷子系统。其中:

题库子系统主要提供试题命制功能,供命题专家填写试题描述、试题属性、收集试题素材、设置评分标准。

考试子系统主要提供考生考试作答功能,并最后生成目标结果文件。

阅卷子系统一般由系统的管理人员监控系统的自动阅卷过程。系统根据考生作答生成的目标结果文件及命题制定的评分标准,动态生成阅卷指令以执行自动阅卷,最后产生阅卷结果。

如图 2,整个自动阅卷评分过程大致分为 11 个业务过程,涉及“题库子系统”“考试子系统”和“阅卷子系统”三个子系统。

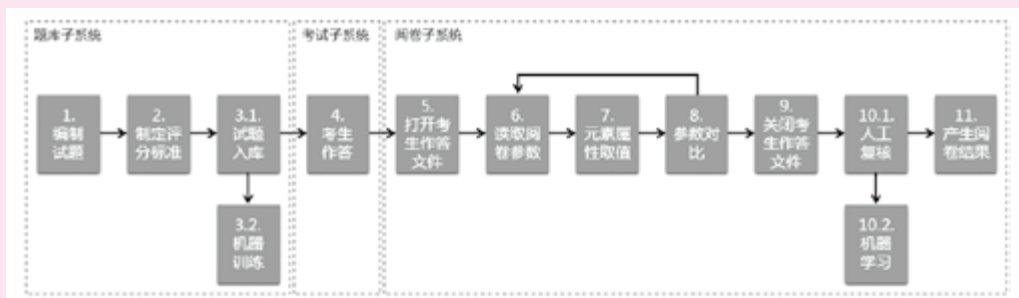


图 2 阅卷评分流程

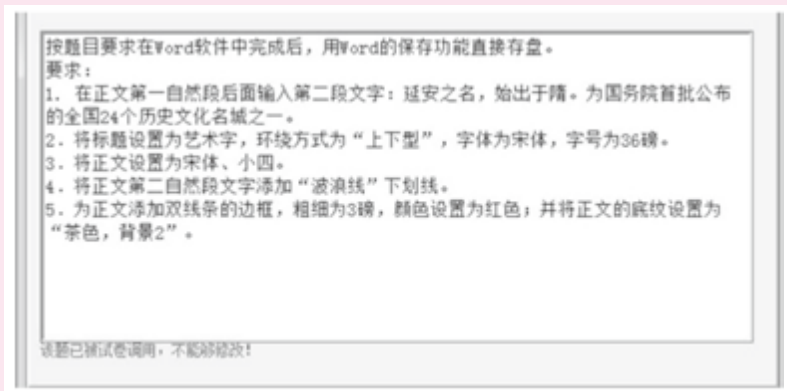


图 3 试题

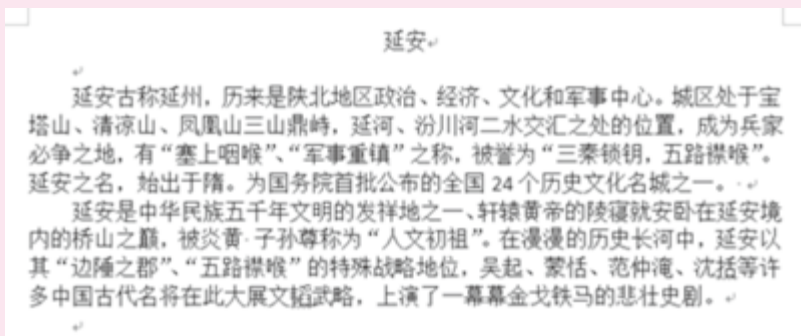


图 4 试题素材

#### (四) 编制试题

命题专家根据知识点和考查维度, 编制试题内容及制定与试题相关的素材文件。

图 3、图 4 是考查 Word 文档编辑能力的试题示例。

#### (五) 制定评分标准

试题和素材编制完成后, 命题专家需要编制该

试题的评分标准。评分标准可以由多条评分点的评分细则记录组成, 每条评分细则除了必须的标准信息(基本内容、评分参数、分值等)以外, 还可以包括其它属性信息, 如文件名称、场景、操作类型等, 每条评分点的评分细则可以具有不同的数据结构。

以 Word 编辑试题为例, 根据阅卷要求生成如表 2 的评分标准。

表 2 表的评分标准

序号	评分点	内容	评分对象	评分参数和值	分值
1	输入文字	在正文第一自然段后输入第二段文字:延安之名,始于隋朝。为国务院首批公布的全国 24 个历史文化名城之一。	第二自然段	<内容> 延安之名,始于隋朝。为国务院首批公布的全国 24 个历史文化名城之一。	1
2	标题设置	将标题设置为艺术字,环绕方式为“上下型”,字体为宋体,字号为 36 磅。	标题	<字型:艺术字><环绕方式:上下型> <字体:宋体><字号:36 磅>	1
3	字体设置	将正文设置为宋体、小四。	正文	<字体:宋体><字号:小四>	1
4	文字设置	将正文第二自然段文字添加“波浪线”下划线。	第二自然段	<下划线:波浪线>	1
5	文字设置	为正文添加双线条的边框,粗细为 3 磅,颜色为红色;并将正文底纹设为“茶色,背景 2”。	正文	<边框类型:双线条><边框粗细:3 磅> <边框颜色:红色><底纹:茶色,背景 2>	1

评分规则可以采用“整体评分规则”也可以采用“分项评分规则”。当采用“整体评分规则”时,系统根据评分参数给出总体的成绩评定。当一个评分点中有多项评分参数时,系统按平均分给每项打分并计算评分点得分;当采用“分项评分规则”时,系统根据每项设定的分值和答案贴近度给出单项评分,并计算评分点得分。

在评分标准的设计时,要考虑到考生误操作的情况。如上表中的第 4 个评分点:将正文第二自然段

文字添加“波浪线”下划线。考生误将“波浪线”加到了第一自然段,这明显不是考生能力问题,而是定位错误。又如上表中第 1 个评分点,考生可能在录入时少打了几个字,或录入错了几个字。这类情况在评分标准设计时,可以考虑给部分得分,以提高阅卷评分的容错能力。

评分标准制定后,命题系统自动根据用户设定的评分参数、标准值和评分规则,生成如表 3 的评分参数表。

表 3 表的评分参数

序号	评分点	定位	评分参数	标准参数值	分值
1	输入文字	第二自然段	InnerText	延安之名,始于隋朝。为国务院首批公布的全国 24 个历史文化名城之一。	1
2	标题设置	标题	extent	延安	0.25
3			wrap type	topAndBottom	0.25
4			rFonts	宋体	0.25
5			fontSize	36pt	0.25
6	字体设置	正文	rFonts	宋体	0.5
7			fontSize	12pt	0.5
8	文字设置	第二自然段	underline	wave	1
9	文字设置	正文	w:val	double	0.2
10			w:sz	24	0.2
11			w:color	FF0000	0.2
12			w:themeFill	background2	0.2
13			w:themeFillShade	E6	0.2

### (六) 试题入库

题库子系统接收命题专家编制的试题和与之相对应的评分标准,并存储到题库子系统的题库数据库中。在组卷时,题库子系统会将基础数据、试题、评分标准、组卷规则等重要数据打包在一起,形成加密的压缩数据包,这里称为试题包。在考试和阅卷时,由考试子系统和阅卷子系统分别导入试题包,作为考试作答和阅卷评分的基础数据来源。

### (七) 机器训练

在试题通过审核入库的同时,系统会根据试题评分点和评分参数启动机器训练。机器训练的数据主要来源于命题管理单位组织的试考正确样本,以及来自互联网的相近样本资源。系统通过分析样本中的特征因子,提取数据类型、模型参数和数据关联,构建出数据网络模型。通过机器训练使题库和阅卷系统掌握更多的掌握更多的阅卷知识,可以更全面地覆盖和识别考生的操作习惯和方式。

例如,我们要求考生对一段文字设置字体、字号并设为粗体字。比较常见的处理方式有以下五种:通过菜单的工具栏分别设字体、字号和黑体;通过快捷工具分别设置字体、字号和黑体;通过右键菜单在字体属性设置工具中分别设置字体、字号和黑体;通过“Ctrl+B”的快捷键设置黑体,再用上述三种方式分别设置字体、字号;通过选择样式设置字体、字号和黑体。

如果对考查操作结果,通过对结果目标文件的分析,前四种的处理结果是相似的,而第五种则完全不同。如果考查操作过程,前四种方式可以形成多种操作组合。

无论是面向操作结果,还是面向操作过程的考查。由于考生的操作习惯和方式的多样性,如果采用常规的正则匹配的方式进行评判,数据分析过程会变得非常复杂,评分效率会大幅降低,而且不能确保完全覆盖各种操作组合。而通过机器训练对各种操作方式的特征因子分析,构建完善而内容丰富数据网络模型,就可以通过特定的算法函数,快速、准确的完成作答结果评判。

### (八) 考生作答

在上机考试开考前,考试监考系统将封装的试题包解密还原,并将试卷推送到考生端的考试子系统。考生在考试子系统中打开试题,并在素材文件中作答。作答完成后交卷,系统自动将考生作答内容和目标结果文件打包生成考生的作答包上报到阅卷子系统。

### (九) 阅卷过程

阅卷过程就是利用自动阅卷技术对于考生作答产生的目标结果文档和操作记录进行解析,然后依据评分标准进行评分参数标准值和作答实际值对

比,计算考生试卷得分的过程。该过程如图5所示。

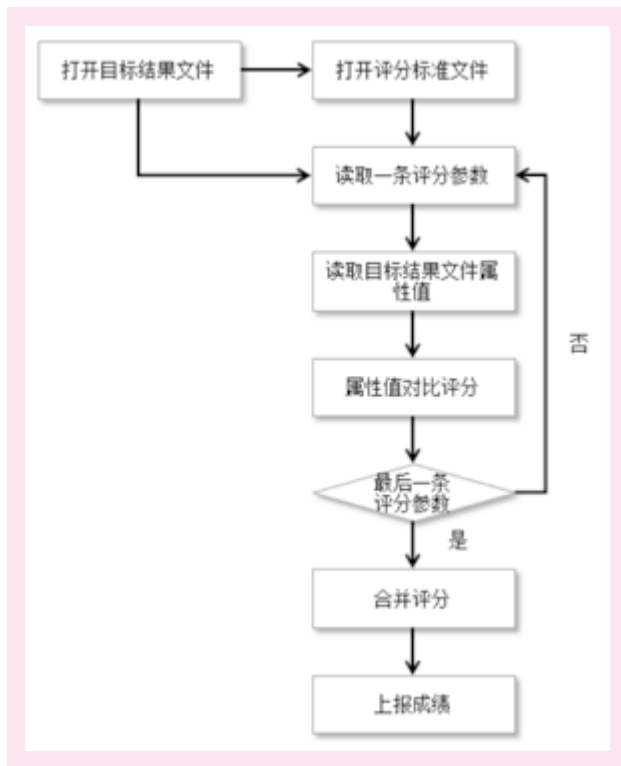


图5 阅卷流程

#### 1. 打开考生作答文件

阅卷子系统接收到考生作答包后,根据被考查软件的类型声明目标结果文件的文档对象模型,打开考生作答待评分试题的目标结果文件,并根据文档对象模型对目标结果文件做结构化处理。

#### 2. 读取阅卷参数

阅卷子系统调取试题评分标准,根据评分点排序逐条获取评分参数,并以此为循环逐条与考生目标结果文件中的对应属性进行对比,并计算出各评分点的评分。

以 Word 编辑试题为例,获取评分参数顺序如表4。

#### 3. 元素属性取值

根据评分参数,系统在结构化的文档对象模型中查找试题对应的特征因子,并以此通过算法函数进行文档属性定位获取对应的属性值。

如,在目标结果文件中获取第1个评分参数的属性值,阅卷系统所使用的读取属性值代码为:

```
string strDocText = Paragraph.InnerText();
```

又如,在目标结果文件中获取第4、5个评分参数的属性值,通常可以采用下面的方式:

```
Fonts fsz = Paragraph.Properties.Elements<Fonts>();
```

```
String fn = fsz.GetAttribute("rFonts").Value;
```

```
String fs = fsz.GetAttribute("fontSize").Value;
```

当考生采用选择样式的方式设置字体时,系统



找到特征因子“w:pStyle”,并以此查找“w:pStyle”所关联的字体属性:

```
Fonts fsy = ParagraphProperties.Elements<Styles>();
```

```
Fonts fsz = fsy.GetAttribute("val", "a3").Value;
```

```
String fn = fsz.GetAttribute("rFonts").Value;
```

```
String fs = fsz.GetAttribute("fontSize").Value;
```

表4 评分参数

序号	定位	评分参数	标准值	评分算法	分值
1	Paragraph2	InnerText	延安之名,始于隋朝。为国务院首批公布的全国 24 个历史文化名城之一。	模糊	1
2	Title	extent	延安	相等	0.25
3	Title	wrap type	topAndBottom	相等	0.25
4	Title	rFonts	宋体	相等	0.25
5	Title	fontSize	36pt	相等	0.25
6	body	rFonts	宋体	相等	0.5
7	body	fontSize	12pt	相等	0.5
8	Paragraph2	underline	wave	相等	1
9	body	w:val	double	相等	0.2
10	body	w:sz	24	相等	0.2
11	body	w:color	FF0000	相等	0.2
12	body	w:themeFill	background2	相等	0.2
13	body	w:themeFillShade	E6	相等	0.2

#### 4. 参数对比

获取目标文件对应评分参数的文档属性值后,系统将获取的数值与评分参数值进行对比。数值对比时要充分考虑数值属性,不能单纯以两值相等来评分。例如,对输入文字内容进行对比时,考生录入错了几个字,或是少录入了几个字,如果以两值不相等

判不得分,则稍显不够公允,系统应该以正确内容给予适当分数。

以 Word 编辑试题为例,根据评分点逐条从目标结果文件中获取元素属性值后,与评分参数标准值进行对比,形成如表 5 的评分结果。

表5 评分结果

序号	评分参数	标准值	作答值	得分
1	InnerText	延安之名,始于隋朝。为国务院首批公布的全国 24 个历史文化名城之一。	延安之名,始于随[错]朝,[错]为国务院首批公布的全国 24 个[漏]历史文化名城之一。	0.83
2	extent	延安	延安	0.25
3	wrap type	topAndBottom	topAndBottom	0.25
4	rFonts	宋体	宋体	0.25
5	fontSize	36pt	34pt[错]	0
6	rFonts	宋体	宋体	0.5
7	fontSize	12pt	12pt	0.5
8	underline	wave	wave	1
9	w:val	double	Single[错]	0
10	w:sz	24	24	0.2
11	w:color	FF0000	FF0000	0.2
12	w:themeFill	background2	background2	0.2
13	w:themeFillShade	E6	E6	0.2
本题得分:				4.38



如表中,考生一共答错了三地方。其中第1个评分参数的评分算法是采用“模糊算法”,因此将考生作答值与标准值做了贴近度计算,给了部分分数。而第5、9个评分参数的评分算法采用了“相等算法”,因此都不能得分。

#### 5. 关闭考生作答文件

阅卷子系统逐条完成各试题的阅卷评分后,关闭考生作答的目标结果文件,并将所有试题的评分相加生成考生本次考试的最终成绩。

#### (十)人工复核和机器学习

在系统自动阅卷结束并阅卷结果后,由专家对机器阅卷结果的高分区、低分区、中分区抽取不同比例的试卷样本进行抽查复核。根据以往的数据统计,专家复核的阅卷评分和系统自动阅卷评分的分值平均差距在正负0.5分之间,专家和系统评分一致比例达到96%以上,系统自动阅卷评分已经具备了较高的准确性。

专家复核与系统阅卷差异的试卷,经专家组复审确定后,由系统对评判错误的试卷进行解析,进行深度学习,修正特征因子及所关联的模型参数,进一步完善数据网络模型。

### 四、结束语

人工阅卷的准确率通常受诸多方面因素的影响,其中评卷专家个人的水平、状态、情绪和评卷结果数据处理的手段都会较大地影响阅卷的准确性。为避免人为因素对评分准确性的影响,业务上大多采取“每人一题,一题二评,二评不一致进三评,三评以后做仲裁”务求零误差。一般情况下,第一轮人工阅卷产生的误差会在10%左右,既使进行了二轮、三轮之后也很难做到零误差。不仅如此,随着软考报考量的逐年上升,人工评卷压力也巨增。以信息处理技

术员考试为例,近年总报考量达到5万以上。如果全部采用人工阅卷,则需要百余名阅卷专家,每人每天阅卷200份以上,耗费数天才能完成。

采用计算机阅卷则可以有效回避人工水平、状态、情绪和评卷结果数据处理的手段所产生的影响,单轮的阅卷准确率即可达到99%以上。而且完成单份阅卷只需零点几秒的时间,五万份试卷,单轮大致需要用时4小时左右。如果采用多线程处理技术,用时可以压缩到更低。虽然计算机自动阅卷仍然不能完全避免发生考生成绩争议的情况,但是机考系统在记录考生作答结果的同时,也记录了考生的作答过程。通过作答日志可以有效还原考生作答的操作全过程,为解决争议提供了数据证据。大大缩短了争议的处理周期,也降低了争议升级的可能性。

目前,计算机软件水平考试中只有信息处理技术员考试采用了上机考试的方式。经过这几年的实施验证,采用上机考试和计算机自动阅卷技术,可以大幅度提高阅卷的准确性,缩短阅卷周期,降低阅卷成本。而且,上机考试可以使软考命题内容更丰富、方式更灵活,更具创新性,其成绩也能较客观地反映考生的技能水平和操作熟练程度。

#### 参考文献:

- [1] 邵朝友.评分规则的理论和技术[M].杭州:浙江大学出版社,2018:38-69.
- [2] 林关成.基于VBA的Office操作题自动阅卷技术研究[J].信息技术,2013,(7):83-86.
- [3] 吴国辉.面向Office操作题的自动阅卷系统设计与实现[D].南京:南京师范大学,2017:24-25.
- [4] 倪应华.基于XML自动阅卷算法的设计与实现[J].仪器仪表学报,2005,(z1):704-705.

[责任编辑:胡大威]

## Design and Implementation of Automatic Marking Technology for Operation Questions of Computer and Software Technology Proficiency Examination

Jiang Huafeng<sup>1</sup>, Xie Guoji<sup>2</sup>

(1.Education and Examination Center of the Ministry of Industry and Information Technology, Beijing100040, China; 2.Shenzhen zhuofan Technology Co., Ltd. Shenzhen518000, China)

**Abstract:** In view of the low efficiency of manual marking of computer operation questions and the need for a large number of professional and technical personnel to mark papers, it is impossible to popularize computer examination on a large scale. Through the analysis and research of the representative method of automatic scoring of computer operation questions, this paper evaluates the feasibility of further promoting computer examination in computer software examinations.

**Key words:** computer software examination; operational questions; automatic scoring; automatic marking