

基于 Aspose 技术的自定义模板文档生成方法^①

周千明, 朱欣娟, 胡西民

(西安工程大学 计算机科学学院, 西安 710048)

摘要: 针对当前信息系统(Information System, IS)中文档批量生成与报表技术繁琐复杂的问题, 提出了一种基于 Aspose 技术的自定义模板文档生成方法。首先建立模板文档, 划分常量域与变量域, 对其通用格式进行自定义设置, 然后重点借助 Aspose 组件调用模板文档并初始化, 进而对其变量域进行实际数据填充, 最后导出为目标格式的文档。通过多次应用实践, 证明了该方法具有较高的鲁棒性与有效性。

关键词: 文档生成; 自定义模板; Aspose 组件; 数据报表; 信息系统

Document Generation Method Based on Aspose Technology and Custom Template

ZHOU Qian-Ming, ZHU Xin-Juan, HU Xi-Min

(College of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: Aiming at the problem that document generation in batch and report technology is overly burdensome and complex in current information system(IS), this paper proposes a document generation method based on aspose and custom template. First, the template document is built, the constant and variable domains of which are divided, and the custom settings for general format of which are completed. Then, the template document is called and initialized by means of aspose component. After actual data filling for the variable domains through aspose component, the document in target format is exported. Many application results show the high robustness and effectiveness of this method.

Key words: document generation; custom template; aspose component; data report; information system

文档批量生成、数据报表及展示是信息系统(Information System, 简称IS)的一项重要功能。就目前而言, 多种开发技术被用于该项功能的研发, 其中比较常见的有引用 Microsoft Office 相关的 COM 组件生成并控制 Office 文档的操作, 基于 XML 技术的文档生成技术, JAVA 体系下的基于中间件 JACOB、POI 与 ITEXT 的文档生成技术等。然而, 这些文档生成技术主要不足在于文档内容的生成需要通过复杂的程序编码实现, 鉴于某些待生成文档的内容格式、样式、版式等的复杂性, 则实现难度异常艰巨。此外, 如果出于实际需要要求调整待生成文档的内容格式、样式、版式等, 则不得不逐一修改相应的源程序。

针对上述问题, 一种可行的解决方案是预先自定义模板文档, 然后在程序中调用模板文档以写入数据,

进而导出目标文档。生成文档的内容格式、样式、版式等可以在模板文档中预先定制, 而程序代码则完成数据填充与文档导出, 这样保证了模板文档与程序代码各自的相对独立性。该方案使得程序的逻辑结构简单, 灵活便捷, 具有很强的可操作性。目前, 关于自定义模板的文档生成方法方面的研究较少。文献[1]提出了一种文档自动生成模型, 侧重于解决企业在编辑数据汇总文档时存在的问题。文献[2]提出了一种基于 Word 自定义模板的文档管理模式, 主要对信息系统中文档的管理方法进行了探索。这两种方案均面向不同的实际应用, 尚未总结出通用的文档生成技术。基于此, 本文提出一种高效简洁并且通用的基于自定义模板的文档生成方法, 并以在高校迎新系统的应用为例, 结合.NET 技术对该方法进行了研究。

① 基金项目: 陕西省科技厅自然科学基金(2013JM8034); 陕西省教育厅自然科学专项(12JK0947)

收稿时间: 2014-10-06; 收到修改稿时间: 2014-11-14

1 Aspose技术概述

Aspose 技术是 Aspose 公司研发的一种主要用于文档操作的组件技术^[3,4],其核心产品为 Aspose 组件。该组件包括 Aspose .NET Components、Aspose Java Components 等一系列支持不同开发技术和语言的类库包,提供了丰富的文档操作与管理的应用程序接口(API),通过该组件可以在应用程序中实现对 Microsoft Office Word、Microsoft Office Excel、Microsoft Office PowerPoint、Adobe Portable Document Format(PDF)等常用格式文档的创建、编辑与格式转换功能,从而完成相应的文档处理任务。

2 处理流程

本文提出的基于 Aspose 技术的自定义模板文档生成方法及文档展示的处理流程如图 1 所示。

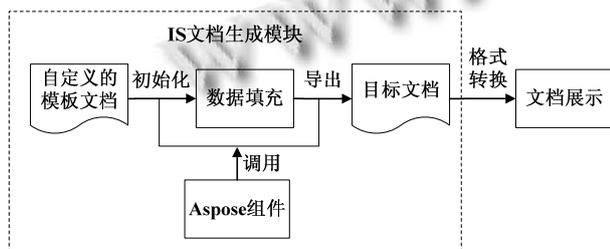


图 1 基于自定义模板的文档生成及展示流程

其中,首先建立模板文档,划分常量域与变量域,根据实际需要对其通用格式进行自定义设置,包括内容格式、样式与版式等。模板文档的格式可以为 Word、Excel、PowerPoint、PDF 等,以 Word 与 Excel 为典型代表的模板文档是本文研究的重点。在实际系统研发过程中,调用相应的 Aspose 组件,初始化自定义的模板文档。在此基础上,对其变量域进行具体数据填充,进而导出为目标格式的文档。最后,根据实际需求对目标文档进行格式转换,得到易于在系统中显示的 HTML 或 SWF 格式文档,以进行文档展示。

3 关键技术及实现

3.1 模板文档自定义

在具有文档批量生成或数据报表功能需求的信息系统中,将要批量生成的文档或报表往往包含内容相同,格式、样式与版式一致,并且位置固定的区域,定义该区域为常量域,定义常量域之外的区域为变量域,常量域与变量域中的数据既可以是文字,也可以是图

片。在系统研发过程中,首先根据实际需要建立相应格式的模板文档,进而划分模板文档中的常量域与变量域。不同格式模板文档常量域与变量域的划分方法为:

(1)对于格式为 Word 的模板文档,采用标签技术^[5]确定变量的位置。对于每个变量的实际数据显示位置,分别插入相应的“书签(Bookmark)”。这里需要指出的是,经过试验,如果要求生成的文档版面格式较为自由,则宜采用插入“文本框”的方式进行模板文档的版面布局。在此种情况下,可以通过在“文本框”内插入“书签”以确定变量域。

(2)对于格式为 Excel 的模板文档,由于电子表格一般都是由 M 行和 N 列的单元格组成,都具有复杂的主栏和宾栏,这就使得单元格之间存在大量而复杂的计算和引用关系^[6],因而可以根据单元格的行列唯一标识确定常量域与变量域。

3.2 基于自定义模板生成文档

对模板文档自定义之后,便可在信息系统中调用 Aspose 组件实现基于自定义模板生成文档。

(1)对于格式为 Word 的模板文档,首先在系统解决方案中引用 Aspose.Words.dll 动态链接库,引入命名空间“Aspose.Words”,生成文档的主要步骤如下:

Step 1: 初始化。利用自定义模板新建并初始化新的 Document 对象:

```
Aspose.Words.Document doc = new Aspose.Words.Document(templateFile); //templateFile 为自定义模板的路径。
```

Step 2: 数据填充。当实际数据为文字时,先定义 Bookmark 变量,并赋值为 Document 对象中相应的“书签”值,然后将实际数据赋给 Bookmark 变量的 Text 属性:

```
Aspose.Words.Bookmark bookmark = doc.Range.Bookmarks["ID"]; //ID 为模板文档中插入的书签名
if (bookmark != null)
{
    bookmark.Text = value; //value 表示实际数据
}
```

当实际数据为图片时,需要先新建并初始化 DocumentBuilder 对象,然后分别调用该对象的 MoveToBookmark 方法与 InsertImage 方法实现图片数据填充。

```
DocumentBuilder dBuilder = new DocumentBuilder
(doc);
```

```
dBuilder.MoveToBookmark("Photo"); //Photo 为模
板文档中插入的书签名
```

```
dBuilder.InsertImage(photoFile, width, height);
//photoFile 为图片文件路径, width×height 表示插入图
片的大小
```

Step 3: 文档导出. 调用 Document 对象的 Save 方法实现文档的导出:

```
doc.Save(saveFile); //saveFile 为将要保存的文件
路径.
```

(2) 对于格式为 Excel 的模板文档, 首先在系统解决方案中引用 Aspose.Cells.dll 动态链接库, 引入命名空间“Aspose.Cells”, 主要步骤如下:

Step 1: 初始化. 首先利用自定义模板新建并初始化新的 Workbook 对象, 然后定义 Worksheet 变量, 赋值为 Workbook 对象的属性 Worksheets[0]:

```
Aspose.Cells.Workbook workBook = new Aspose.
Cells.Workbook(templateFile); //templateFile 为自定义
模板的路径
```

```
Aspose.Cells.Worksheet workSheet = workBook.
Worksheets[0];
```

Step 2: 数据填充. 利用 Worksheet 变量的 Cells 属性直接定位模板文档中单元格的位置, 进而调用 Cells.PutValue 方法写入实际数据:

```
workSheet.Cells[column+row].PutValue(value);
//column 与 row 均为 string 类型变量, column+row 表征
了模板文档中唯一单元格, value 表示实际数据.
```

Step 3: 文档导出. 调用 Workbook 对象的 Save 方法实现文档的导出:

```
workBook.Save(saveFile); //saveFile 为将要保存的
文件路径
```

3.3 文档展示

文档的展示可通过以下两种方案实现:

方案 1: 在文档导出步骤中, 调用 Document 对象或 Workbook 对象 Save 方法的一个重载, 将生成的文档导出为 HTML 格式文件. 在 C/S 架构与 B/S 架构的系统中均容易嵌入 HTML 文件, 从而实现文档展示:

```
doc.Save(saveFile, Aspose.Words.SaveFormat.
Html); //保存为 HTML 格式文件
```

```
workBook.Save(saveFile, Aspose.Cells.SaveFormat.
```

```
Html); //保存为 HTML 格式文件
```

方案 2: 在系统解决方案中调用第三方文档转换工具, 将生成的文档经过虚拟打印方式转换成 SWF 格式文件, 只要系统能够支持 Flash 播放, 就可以实现文档展示. 这类文档转换工具的典型代表是 Macromedia FlashPaper、Print2Flash 等.

4 应用实例与分析

4.1 录取通知书的批量生成

以国内某高校的新生录取通知书为例, 该校要求录取通知书上显示的项目包括条形码图片、通知书编号、考生号、考生头像图片、姓名、省份、学院、录取专业与报到日期 9 个项目. 根据实际需要, 选择格式为 Word 的模板文档生成录取通知书. 首先, 建立录取通知书模板文档, 采用插入“文本框”的方式进行版面布局, 通过在“文本框”内插入“书签”以确定变量域, 共 9 个书签标识的区域, 如图 2 所示. 常量域包括校名、校长签名图片、落款以及相应的衔接文字. 然后, 以全国普通高校招生管理系统中该校的录取数据为基础数据, 采用本文提出的基于 Aspose 技术的自定义模板文档生成方法批量生成录取通知书, 生成文档效果如图 3 所示:

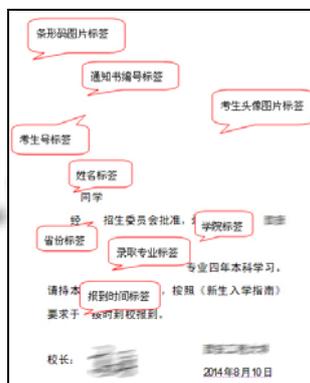


图 2 自定义模板(Word) 图 3 生成文档效果(Word)

对比图 2 与图 3 可以发现, 自定义模板中采用书签标识的变量域在生成的文档中均填充了有效数据, 并且生成文档的字体、段落及版面格式与自定义模板保持一致. 该方法具有很强的灵活性, 在变量域确定的情况下, 使得程序代码与模板文档相互独立.

4.2 报到进度情况实时报表

在新生报到过程中, 某高校为了全面地掌握新生报到进度情况, 要求报到系统能够实时生成分学院分

专业报到人数与报到率的统计报表。因此,可以选择格式为 Excel 的模板文档生成该统计报表。首先建立报到进度报表模板文档,进行相应的自定义设置。图 4 所示为该报表模板文档的部分截图。

2014级新生报到进度统计							
分学院分专业:				总人数 4701 (男 2284 女 2417)		按报到人数/人数自动填充	
序号	学院代码	学院名称	专业代码	专业名称	人数	报到人数	报到率
1	01	纺织与材料学院	01	纺织工程	180	0	0.00%
2			02	高分子材料与工程	31	0	0.00%
3			03	轻化工程	30	0	0.00%
4			04	非织造材料与工程	30	0	0.00%
5			31	纺织工程(卓越班)	30	0	0.00%
6			41	轻化工程(卓越班)	30	0	0.00%
7			合计		331	0	0.00%
8			02	工业设计	30	0	0.00%

图 4 自定义模板(Excel)

在图 4 中所示的自定义模板中,常量域的设置包括单元格大小、单元格的拆分与合并、基本数据填写、字体属性等。变量域为“报到人数”列,初始值设置为 0,“报到率”列的数据按“报到人数/人数”自动填充,设置其为百分比格式。采用本文提出的基于 Aspose 技术的自定义模板文档生成方法,在新生报到过程中,系统提取存储在数据库中的实时报到人数,生成相应的报表文档,效果如图 5 所示。

2014级新生报到进度统计							
分学院分专业:				总人数 4701 (男 2284 女 2417)			
序号	学院代码	学院名称	专业代码	专业名称	人数	报到人数	报到率
1	01	纺织与材料学院	01	纺织工程	180	176	97.78%
2			02	高分子材料与工程	31	30	96.77%
3			03	轻化工程	30	30	100.00%
4			04	非织造材料与工程	30	29	96.67%
5			31	纺织工程(卓越班)	30	29	96.67%
6			41	轻化工程(卓越班)	30	29	96.67%
7			合计		331	323	97.58%
8			02	工业设计	30	29	96.67%

图 5 生成文档效果(Excel)

通过上述高校迎新系统中两个典型的应用实例可

知,根据实际需求的不同特征选用最合适格式的模板文档,才能使得本文提出的文档生成方法执行效率最大化。如果需要在系统对生成的文档进行展示,则可以任意采用本文阐述的两种文档展示解决方案。

5 结语

本文针对当前信息系统中文档批量生成与报表技术繁琐复杂的问题,提出了一种高效简洁且通用的基于自定义模板的文档生成方法,并以在高校迎新系统的应用为例进行了研究。该方法通过建立模板文档并对其常量域与变量域进行划分,自定义设置其通用格式,然后借助 Aspose 组件调用模板文档并初始化,进而对其变量域进行实际数据填充,从而导出为目标格式的文档。应用实践证明,该方法具有很强的实用性。

参考文献

- 1 曲明成,廖明宏,吴翔虎,等.一种文档自动生成模型的构建及其应用.计算机集成制造系统,2008,14(7):1297-1305.
- 2 张琼妮,肖刚,张元鸣,等.一种基于 Word 自定义模板的文档管理模式.浙江工业大学学报,2006,34(1):101-104.
- 3 Aspose corporation. Aspose Product Doc's Home. <http://www.aspose.com/docs/dashboard.action>.2014-10-03.
- 4 李培国,叶嫣.多类型文档检索和转换技术研究.微型机与应用,2014,33(10):4-6.
- 5 车晓波,王知学,王磊.在实现 Word 自动化操作中书签的应用.信息技术,2013,(8):148-150.
- 6 赵立永,赵冲冲,时鹏,等.半监督二次划分聚类算法及其报表系统应用.小型微型计算机系统,2011,32(3):499-505.
- 7 王小科,王军,等编著.C#开发实战 1200 例(第 I 卷).北京:清华大学出版社,2011.