

doi:10.3969/j.issn.1673-9833.2025.06.008

规范格式文档的自动排版技术研究

申 剑¹, 廖友媛¹, 姜利民¹, 鲍春燕²

(1. 湖南工业大学 期刊社, 湖南 株洲 412007; 2. 湖南工业大学 经济与管理学院, 湖南 株洲 412007)

摘要: 针对具有规范格式的文档自动排版问题, 提出了采用正则表达式结合样式的方案。首先, 在文档中定义符合规范的样式; 然后, 通过正则表达式查找并设置段落样式; 最后, 通过宏录制代码执行日常编辑任务。在期刊的验证案例中实现了该方案, 结果证明该方案能够高效地实现规范文档的自动排版, 大部分文档可以在3 s内完成排版。

关键词: 规范文档; 自动排版; VBA; 正则表达式

中图分类号: TP391

文献标志码: A

文章编号: 1673-9833(2025)06-0050-05

引文格式: 申 剑, 廖友媛, 姜利民, 等. 规范格式文档的自动排版技术研究[J]. 湖南工业大学学报, 2025, 39(6): 50-54.

Research on Automatic Typesetting for Standardized Format Documents

SHEN Jian¹, LIAO Youyuan¹, JIANG Limin¹, BAO Chunyan²

(1. Department of Academic Periodicals, Hunan University of Technology, Zhuzhou Hunan 412007, China;

2. School of Economics and Management, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: A solution with regular expressions and styles combined together has thus been proposed to address the issue of automatic typesetting of documents with standardized formats. Firstly, a definition is given for the standardized style in the document; next, paragraph styles are to be found and set by using regular expressions; finally, daily editing tasks are to be executed by recording code through macros. The proposed scheme can be implemented in the validation case of the journal, with the results showing that it can efficiently achieve automatic typesetting of standardized documents, and most documents can be typeset within 3 seconds.

Keywords: standardized document; automatic typesetting; VBA; regular expression

1 研究背景

随着信息技术的飞速发展, 文档处理需求日益增长, 自动排版技术应运而生并不断发展。它旨在通过计算机程序自动完成文档的格式设置、布局调整等任务, 提高排版效率与排版质量。而对于期刊论文、毕业论文这类规范文档, 一般遵从某种排版规范的要

求, 其格式要求非常严格, 而对于它们的编辑任务, 往往是一个枯燥、易错、效率低下的过程。通过自动排版技术可以将编辑、学者、学生从日常文字编排的烦恼中解放出来, 提高办公和科研的效率。

对于自动排版技术, 研究者们进行了许多卓有成效的研究。李渊本等^[1]在新信创办公环境下, 应用JSA宏技术对银河麒麟OS下的金山WPS文字软件

收稿日期: 2025-01-20

基金项目: 株洲市社会科学成果规划评审委员会课题基金资助项目(ZZSK2024178); 湖南省高等学校学报研究会编辑学基金资助项目(PY2024005)

作者简介: 申 剑, 男, 湖南工业大学教师, 主要研究方向为计算机应用技术和编辑理论, E-mail: 319779@qq.com

通信作者: 鲍春燕, 女, 湖南工业大学教师, 主要研究方向为可持续发展管理, E-mail: 319731@qq.com

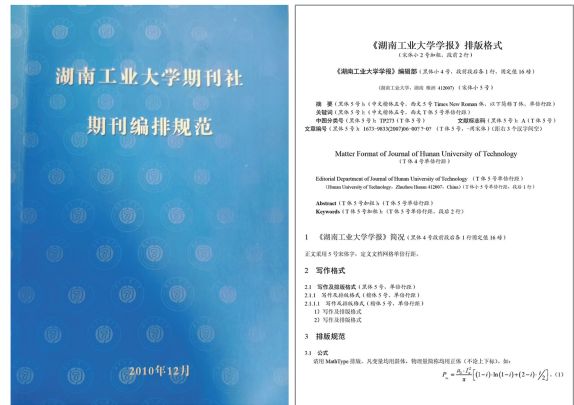
进行二次开发, 研发了自动编排统计信息报送格式程序。杨桂华等^[2]以《环境科学学报》为例, 构建了基于 XML 一体化的科技期刊数字化生产与新媒体融合出版模式, 采用 XML 一体化生产平台后, 实现了内容一次制作、多元发布、支持单篇和整期稿件快速发布、生产即发布, 提前 4 个月网络首发, 大大缩短了论文发表周期。张官福^[3]探讨了使用 Word 中宏录制汇编 VBA (Visual Basic for Applications) 代码, 实现公文自动排版的技巧。雷群泌^[4]研究了 Word VBA 自动排版技术, 并以最新的公文格式规范要求为标准, 通过 VBA 编程实现了公文的自动排版。吴智^[5]利用 Word VBA 快速生成新闻稿, 其操作简单, 生成的新闻稿结构完整、格式规范, 大大提高了新闻写作速度。李秀贤^[6]基于 Python 对 Word 格式论文进行了自动排版, 提高了编辑工作效率。殷克涛等^[7]在 XML 基础上构建了电子书 E-book, 并利用专业设计的参数文档结合 XSL 转化实现电子书自排版, 基本达到了专业排版的效果。张晓多^[8]对在线科技论文编辑与自动排版平台的相关内容进行了讨论。王超^[9]利用 VBA 技术构建并实现了一键公文自动排版, 其满足了政务公文排版的需求, 同时可减轻编排人员的工作量, 并且提供标准化、规范化的公文排版格式。李响^[10]使用 VC++ 环境开发了一种实现 Word 文档自动生成和文档格式自动排版系统。何伟平等^[11]开发了学位论文自动规范化排版系统, 该系统能满足论文完整性检查、各结构部分的识别、格式纠正等要求。余姝华^[12]利用 VBA 在 Word 中的应用, 开发一个自带题库, 教师可以对题库进行添加、修改和删除试题等操作, 并能够对所选择的题目进行自动排版。郑根创^[13]采用 Delphi 语言与 Word 无缝连接的方法, 通过智能识别技术, 识别相应内容并应用相应样式, 最终实现对本科论文及相关文件的自动化排版。余梓唐^[14]对自动组卷系统中的选择题类型的排版方法进行了研究, 提出了两种基于 VBA 编程实现的方法, 给出了具体实现的代码, 并对实际运行情况进行了比较。

综上所述, 对于期刊论文、毕业论文这类规范文档, 通常有明确的排版规范要求, 急需找到一种安装简单、节省费用的自动排版技术。现有技术有的需要安装 Python 的 Office 接口, 通过编程控制接口, 实现对文档的自动排版; 有的需要通过付费方式, 通过软件定制自动排版方式来实现; 有些技术则对文档的格式有要求, 不具备通用性。本研究拟通过 Office VBA 语言, 通过正则表达式和样式的配合, 实现一种简单高效地对规范文档的自动排版技术。

2 研究对象及技术路线选择

2.1 研究对象

本文要实现自动排版的对象为《湖南工业大学学报》, 该期刊是由湖南省教育厅主管、湖南工业大学主办的一本综合性学术期刊。期刊论文本身具有一定的规范要求的, 《湖南工业大学学报》遵守期刊社制定的内部编排规范, 该规范对编排的各个方面进行了定义和要求, 并参考许多编辑专业的国标或者行业标准, 对论文的内容和形式进行了约束。该内部规范要求如图 1 所示。

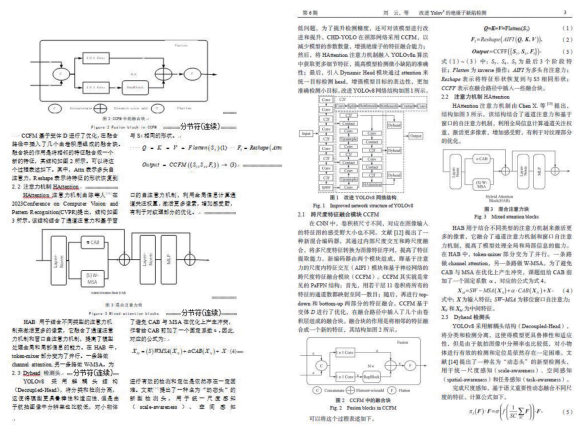


a) 内部编排规范 b) 排版规范要求

图 1 《湖南工业大学学报》内部排版规范

Fig. 1 Internal Layout Standards of Journal of Hunan University of Technology

编辑人员通过对作者文档的编辑加工, 并通过数次修改和校对过程, 最终形成了美观大方、排版紧凑的印刷版本, 如图 2 所示。



a) 编辑前 b) 编辑后
图 2 编辑前后的论文版式对比

Fig. 2 Comparison of papers layout before and after editing

2.2 技术路线选择

由于编辑所面对的文档大多是 Office Word 文档, 许多学生和研究者最熟悉的编辑软件都是 Office Word, 最终选择采用 Office VBA 语言作为实现自动排版的技术平台。VBA 是微软开发的一种编程语

言,它嵌入在 Microsoft Office 应用程序(如 Word、Excel、PowerPoint 等)中,允许用户通过编写代码来自动化和扩展这些应用程序的功能。Word VBA 则是专门用于 Microsoft Word 的 VBA 编程,用户可以利用它来创建宏、自定义功能区、开发文档处理工具等,以实现文档处理的自动化和个性化。

采用 VBA 的方式,可以不安其他编程工具,仅使用 Office 平台就能实现文档的自动排版技术。这种方式对于移动办公环境,尤其方便。在实践中,利用 Word VBA 的功能,在 Word 文档中,先定义一些符合期刊规范要求的样式,通过正则表达式(regular expression, 简称 regex 或 regexp)在论文中查找和定位内容,将不同类别的内容赋予要求的样式,从而实现文档的自动排版;对于某些特定的任务,比如说页面设置、对于参考文献某些内容的替换,则采用录制宏以后,将宏代码嵌入处理流程中。

3 自动排版的实现

3.1 自动排版流程

自动排版的流程如图 3 所示。

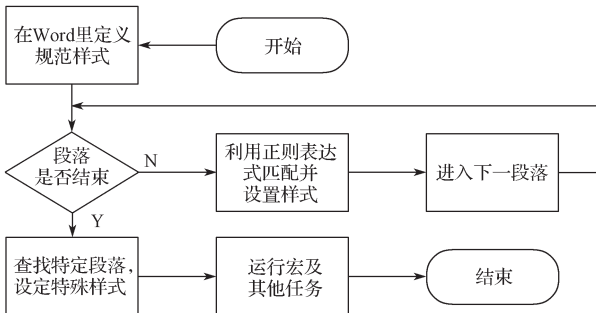


图 3 自动排版流程图

Fig. 3 Automatic typesetting flowchart

如图 3 所示,首先需要在文档中定义符合规范要求的样式,由于期刊本身对于论文的表达和格式有规范的要求,所以论文这种规范文档是有严格的要求的。图 4 为在 Word 内定义刊物要求样式的过程。

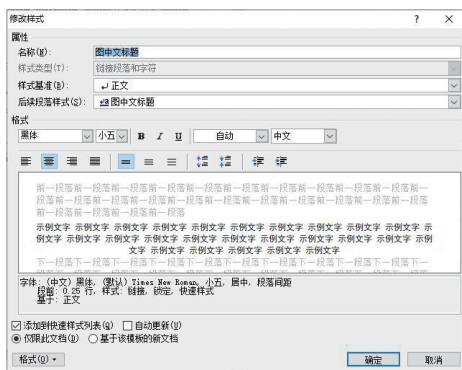


图 4 建立刊物样式示例

Fig. 4 Example of establishing publication style

在 Word 模板中定义了三级标题、文档总标题、中英文图表标题、中英文单位、分标题、参考文献等各种需要的样式,它们也匹配内部规范对论文格式的各种要求。

然后,通过正则表达式识别和定位文档内容,将段落的样式设置为定义好的样式。

最后,运行宏任务和用 VBA 写好的一些常规编校任务。

3.2 Word VBA 文档结构

Word VBA 通过层次化的对象,方便开发者使用,其层次模型如图 5 所示。

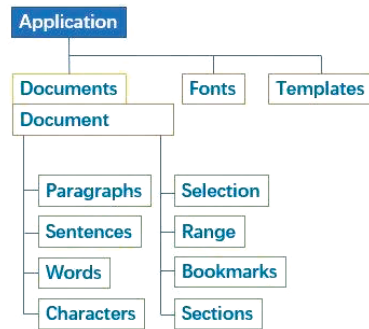


图 5 Word VBA 的主要层次对象模型

Fig. 5 Word VBA main hierarchical object model

在图 5 中,最常见的对象有:

- 1) Application 对象,代表 Word 应用程序本身;
- 2) Document 对象,代表 Word 文档,类似于 Excel 工作簿对象;
- 3) Paragraphs 对象,代表文档的段落;
- 4) Sentences 对象,代表句子;
- 5) Words 对象,代表单词;
- 6) Characters 对象,代表字符;
- 7) Selection 对象,代表文档中所选择的内容;
- 8) Bookmarks 对象,代表文档中的书签;
- 9) Range 对象,代表文档中一个由起始点和终止点所确定的连续区域;
- 10) Sections 对象,代表节;
- 11) Fonts 对象,代表字体;
- 12) Templates 对象,代表 Word 文档模板。

由于规范要求中,大多数对于格式的要求是基于段落的,对于一个段落的格式是唯一的,所以主要采用 Paragraph 对象来进行操作。通过对 Paragraph 对象的样式属性进行设置,可以完成对一个段落样式的设置,并最终完成对全文样式的设置。Paragraph 由许多段落组成,可以通过定位特定段落来实现某些使用正则表达式难以发现的特殊文本。

3.3 正则表达式

正则表达式是一种用于描述文本模式的强大工

具。它通过使用特定的字符组合来定义匹配规则, 从而可以在文本中搜索、匹配、替换或提取符合这些规则的字串。正则表达式广泛应用于文本处理、数据验证、日志分析、字符串操作等众多领域。

正则表达式利用一个模式串来匹配字符串, 从而实现字符串的查找、替换、数据验证等功能。由于规范文档对于文档的格式有严格的要求, 因此文档的段落可以通过正则表达式来进行识别或者定位, 从而实现相应的功能要求。刊物部分正则表达式识别模式如表 1 所示。

表 1 刊物部分正则表达式模式串

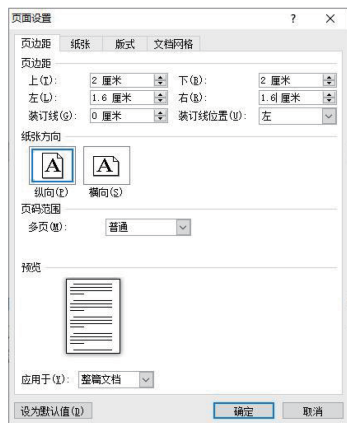
Table 1 Regular expression pattern strings for publication section

样 式	正则表达式模式串
一级标题	$\backslash d\{1,2\}\backslash s+\backslash S+$
二级标题	$\backslash d\{1,2\}\backslash s*\backslash d\{1,2\}\backslash s*\backslash S+$
图中文标题	$\backslash s*\text{图}\backslash d\{1,2\}\backslash s+\backslash S+$
图英文标题	$\backslash s*(\text{Figfig})\backslash s*\backslash s*\backslash d\{1,2\}\backslash s*\backslash S+$
分图标题	$\backslash s*[a-z](\backslash))\backslash s*$

所有的正则表达式模式串都经过了测试和验证, 保证能够无误地识别文档的内容, 可以起到查找、定位、替换等功能。

3.4 宏录制及应用

对于一些常规编辑任务, 可以通过录制 VBA 宏来完成。在 Microsoft Excel 中, 录制 VBA 宏是一种非常方便的方式, 可以让用户快速生成用于自动化任务的 VBA 代码, 而无需手动编写代码。通过录制宏, Excel 会自动记录用户在软件中的操作, 并将其转换为 VBA 代码。VBA 宏实质上是一段可以执行的 VBA 代码, 录制宏的过程实际上是将日常需要执行的过程转换成 VBA 代码的过程。日常的页面设置、查找替换等功能都可以通过录制宏的过程转换成 VBA 代码。图 6 显示的是页面设置过程转换成 VBA 代码的过程。



a) 录制“页面设置”宏

```
With Selection.PageSetup
    .LineNumberingActive = False
    .Orientation = wdOrientPortrait
    .TopMargin = CentimetersToPoints(2)
    .BottomMargin = CentimetersToPoints(2)
    .LeftMargin = CentimetersToPoints(1.6)
    .RightMargin = CentimetersToPoints(1.6)
    .Gutter = CentimetersToPoints(0)
    .HeaderDistance = CentimetersToPoints(1.5)
    .FooterDistance = CentimetersToPoints(1.75)
    .PageWidth = CentimetersToPoints(21)
    .PageHeight = CentimetersToPoints(29.7)
    .FirstPageTray = wdPrinterDefaultBin
    .OtherPagesTray = wdPrinterDefaultBin
    .SectionStart = wdSectionContinuous
    .OddAndEvenPagesHeaderFooter = False
    .DifferentFirstPageHeaderFooter = True
    .VerticalAlignment = wdAlignVerticalTop
    .SuppressEndnotes = False
    .MirrorMargins = False
    .TwoPagesOnOne = False
    .BookFoldPrinting = False
    .BookFoldRevPrinting = False
    .BookFoldPrintingSheets = 1
    .GutterPos = wdGutterPosLeft
    .LayoutMode = wdLayoutModeGrid
End With
```

b) “页面设置”宏转换后的 VBA 代码

图 6 “页面设置”宏转换示例

3.5 自编制任务

由于期刊论文对于形式和格式都有许多硬性要求, 主要包括易错的专业名词、错误的单位形式等。因此在项目中利用 VBA 编制了一些任务过程来辅助编辑工作:

- 1) 对于易错的专业名词, 建立词库, 在全文进行排查, 进行突出显示;
- 2) 对于文中出现的错误单位形式, 利用正则表达式进行定位, 纠正为正确的格式;
- 3) 设置英文标题和参考文献的大小写, 将除停用词以外的单词设置成首字母大写。

4 自动排版测试

将编制好的代码融合到 Word 程序中, 对未经加工的 Word 文档进行自动排版, 测试环境为: CPU 为 Inter(R) Core(TM) i5-8500@3.00 GHz; 运行内存为 8 GB; 测试软件环境为 Word 2010。

采用 100 篇各类投稿到本刊物且大体符合刊物格式的文档进行测试, 测试结果如表 2 所示。

表 2 自动排版测试结果

Table 2 Automatic typesetting test results

最快时间/s	最慢时间/s	平均时间/s	成功排版	未成功排版
			篇数	篇数
1.5	7.6	2.9	97	3

从表 2 可以得出, 大部分文档都能在 3 s 之内完成任务; 少量文档由于本身图形较多, 机器载入困难, 导致运行较慢; 还有极少数文档出现了问题未能进行自动排版, 经核实, 是由于来稿作者经验不够, 设置了错误的格式, 需手动解决问题后再进行自动排版。

而在期刊的实际使用过程中, 所设计的程序能快速地进行自动排版, 大大提升了编辑的工作效率。

5 结语

针对规范文本格式自动排版问题,首先采用VBA编程,采用正则表达式匹配不同样式的文本,并设置为预置的样式;然后运行宏任务和用VBA编制的任务。自动排版测试结果表明,绝大多数文档可以在3s内完成排版,结果无误,可以为期刊、报纸、论文等规范文档排版机构提供一种免费、高效的排版解决方案。下一步的研究目标将与AI语料大模型结合,对文本的错别字、语义逻辑进行建议和修改,减少研究者和相关从业者的压力。

参考文献:

- [1] 李渊本,周俊.信创办公环境下一种实现统计信息文档格式自动编排的方法[J].统计科学与实践,2023(7):61-64.
LI Yuanben, ZHOU Jun. A Method to Realize Automatic Layout of Statistical Information Document Format in Xinchuang Office Environment[J]. Statistical Theory and Practice, 2023(7): 61-64.
- [2] 杨桂华,卜庆杰,王虎,等.基于XML一体化的科技期刊数字化生产与新媒体融合出版实践:以《环境科学学报》为例[J].中国科技期刊研究,2024,35(6):781-787.
YANG Guihua, BU Qingjie, WANG Hu, et al. Practice of Digital Production and New Media Convergence of Scientific Journals Based on XML Integration: A Case Study of Acta Scientiae Circumstantiae[J]. Chinese Journal of Scientific and Technical Periodicals, 2024, 35(6): 781-787.
- [3] 张官福.巧用“宏”录制的方式实现公文处理自动化[J].办公自动化,2022(12):17-21.
ZHANG Guanfu. Use Skillfully the Way of “Macro” Recording to Realize the Automation of Documents Processing[J]. Office Informatization, 2022(12): 17-21.
- [4] 雷群泌.Word VBA自动排版方法在公文处理中的应用[J].中国新技术新产品,2021(4):32-34.
LEI Qunbi. Application of Word VBA Automatic Typesetting Method in Document Processing[J]. New Technology & New Products of China, 2021(4): 32-34.
- [5] 吴智.基于WordVBA新闻稿快速生成与排版系统的研究[J].电脑知识与技术,2020,16(33):214-216,226.
WU Zhi. Research on the Rapid Generation and Typesetting System of WordVBA Press Release[J]. Computer Knowledge and Technology, 2020, 16(33): 214-216, 226.
- [6] 李秀贤.基于Python的学位论文自动排版研究[J].现代职业教育,2020(9):128-129.
LI Xiuxian. Research on Automatic Typesetting of Dissertation Based on Python[J]. Modern Vocational Education, 2020(9): 128-129.
- [7] 殷克涛,张焯.基于XML的电子书自排版系统的实现[J].出版科学,2017,25(4):40-44.
YIN Ketao, ZHANG Xuan. Realization of the E-Book Self-Publishing System Based on XML[J]. Publishing Journal, 2017, 25(4): 40-44.
- [8] 张晓多.在线科技论文编辑和自动排版平台研究[J].新闻研究导刊,2016,7(8):298,301.
ZHANG Xiaoduo. Research on Online Editing and Automatic Typesetting Platform for Scientific Papers[J]. Journal of News Research, 2016, 7(8): 298, 301.
- [9] 王超.基于VBA技术的一键公文自动排版的实现[J].淮海工学院学报(自然科学版),2015,24(3):42-46.
WANG Chao. On Instant Automatic Typesetting of Official Document Based on VBA[J]. Journal of Huaihai Institute of Technology (Natural Science Edition), 2015, 24(3): 42-46.
- [10] 李响.VC中用Word实现文档的自动生成和排版的研究[D].北京:华北电力大学,2015.
LI Xiang. Research on Automatic Generation and Typesetting of Documents by Word in VC[D]. Beijing: North China Electric Power University, 2015.
- [11] 何伟平,张晓东.学位论文自动规范化排版系统的设计[J].广东化工,2015,42(1):139-140.
HE Weiping, ZHANG Xiaodong. The Automatic Software System Design for the Format of Dissertation[J]. Guangdong Chemical Industry, 2015, 42(1): 139-140.
- [12] 余姝华.VBA在Word中的应用:自动生成试卷[J].新课程学习,2014(2):124-125.
YU Shuhua. Application of VBA in Word: Automatic Generation of Test Paper[J]. New Courses' Study, 2014(2): 124-125.
- [13] 郑根创.论文自动排版系统的设计与实现[D].广州:华南理工大学,2012.
ZHENG Genchuang. Design and Implementation of Paper Automatic Typesetting System[D]. Guangzhou: South China University of Technology, 2012.
- [14] 余梓唐.VBA组卷中自动排版系统[J].电脑编程技巧与维护,2009(1):15-17.
YU Zitang. Automatic Typesetting System in VBA Test Paper Composition[J]. Computer Programming Skills & Maintenance, 2009(1): 15-17.

(责任编辑:申剑)