《基于OOML的文档格式检查程序》文献综述

齐家宝 20201103035 2020级少数民族预科班 翟晔

一、研究现状

目前主流的办公软件是微软公司的Microsoft 365 （Office），其中文字处理软件是Word，可以说微软的office主导着办公软件的领域。微软公司开发的Word软件也是占据着文字处理软件中的核心位置。但是微软公司对于Word软件的开发文档或者API的共开较少。利用程序代码操作Word文档的技术公布也相对较少。微软公开了Office Open XML标准，又给提供了Open XML SDK。但是此SDK提供微软自家语言C#的方法和接口。C/C++.Java等语言调用此SDK并不容易，也SDK本身没有提供相关方法和接口。关于C++操作Word可值得一提的是微软提供了COM API可以借助COM API用C++编程实现Word的自动化。但是COM API需要依赖office组件，在一台没有office的计算机上借助COM API开发的程序是无法正常运行的，这也是致命的问题。

目前，国内外许多研究和开发机构都在致力开发文档格式解析程序的研究和开发。例如，国内的一些高校和科研机构正在进行相关项目的研发，如北京理工大学的“毕业设计论文格式检查系统”等，但是此类相似研究目前还是较为匮乏。

Word文档格式检查校验需求市场份额并不是很乐观。再加上企业和各个科技公司对于Excel（电子表格）的各类操作需求要远大于Word文档格式校验的需求。所以在Excel表格的处理技术上更用心导致Excel表格的操作处理手段和方法技术要比Word文档丰富的多，Word在这一方面属于在匮乏的状态，再一方面，开源社区对于Word格式处理的重视度并不高。再加上Word格式处理的开发文档与技术文档和各类研究文献匮乏导致开发关于Word的格式解析和处理程序较为困难所以Word格式处理问题上研究并现状不容乐观。[4][8]

现有的一些Word格式解析检查程序的程序大部分基于微软的.NET平台的VBA技术或者VSTO（全称Visual Studio Toolfor Office,是.net平台下的Office开发技术）技术。VBA技术缺陷也是很明显，无法达到全自动化，程序操作期间还需要大量人工辅助完成工作。利用VBA技术完成的Word格式检查程序要依靠大量的人工操作，这显然不是我们想要的结果。[8]替代VBA技术的方法之一就是用微软新提供的office开发技术VSTO进行office处理的程序开发。但是VSTO技术开发对于平台的依赖性特别大，还需要特定的开发工具，VSTO技术需要依靠微软提供的IDE（集成开发环境）：Visual Studio。Visual Studio是windows平台的开发工具，在其他平台也是不方便使用。[9]

这一现状维持到Microsoft Office 2007问世以后，微软提出了Open XML格式随后ECMA-376 标准问世。[7]并且微软公开了Open XML标准相关的文档和说明。其中包含针对Word文档处理的相关参考手册。现如今对于word的操作我们可以依靠xml处理技术完成相关操作。[10][11][12]

二、参考文献

本人论文主要参考了以下文献。

1. 王杏龙 & 毛永江.(2008).本科毕业论文质量下降的原因及对策. 高等农业教育(09),69-71.
2. 喻健 & 朱三元.(2016).基于Entity Framework和easyUI的毕业论文管理系统. 软件工程(12),40-42.
3. 阚运奇.(2012).毕业论文格式检测系统的设计与研究. 无线互联科技(11),17.
4. 徐东风,彭红星 & 廖俊杰.(2010).基于Java的文档格式检查技术的研究及其应用. *计算机工程与设计*(19),4309-4311+4315. doi:10.16208/j.issn1000-7024.2010.19.018.
5. 褚佳强.(2009).计算机辅助教学管理——本科毕业论文规范化排版. 科技资讯(28),232. doi:10.16661/j.cnki.1672-3791.2009.28.128.
6. 赵振洲 & 乔明秋.(2013).基于OXML的Office文档结构分析与取证调查. 计算机安全(08),44-46.
7. (2008).微软文档格式获批为ISO/IEC标准. 每周电脑报(11),41.
8. 孙晓伟,毕晓清,吴洪越 & 赵卫东.(2013).基于DOM和VBA的Word文档数据交互式抽取. 计算机应用与软件(02),113-115.
9. 刘艳平.(2016).基于VSTO的word文档格式检查插件的开发与研究. 电脑知识与技术(26),106-108. doi:10.14004/j.cnki.ckt.2016.3435.
10. 李琳.(2003).浅谈XML及其在Microsoft Office中的应用. 衡阳师范学院学报(自然科学)(03),69-72
11. 文档格式标准攻防战尘埃落定 微软过关各方积极应对[J]. 中国计算机报,2008, (12).
12. 李高超. 微软 OOXML 文档格式获批成为国际标准[N/OL]. 国际商报,2008-04-07(A04).