基于自然语言处理的智能编辑辅助软件设计与实现

1. 项目背景调研

1、问题提出

编辑部的稿件审阅工作包括稿件内容审查和稿件格式审查两大方面。其中格式审查的范围主要是对章节编号、页眉页脚、引用文献格式的检查。相较于对内容的审查，该项工作较为枯燥繁琐。特别是对于字数较多的文稿来说，其中的格式错误不易被人工所排查，且检查字数较多的文稿将会耗费审稿员很大时间和精力，影响工作效率。利用远程稿件处理系统（采编系统）可极大的方便稿件处理，自动校对稿件格式、要素是否齐全，提醒作者重新修改再投，编辑将省去因格式修改带来的大量工作。

基于此,研发软件测试文档自动审查系统。

为学报编辑设计软件对投稿的文章进行格式审查，自动检测不符合要求的格式并生成针对论文格式的修改意见并生成反馈。即可极大的方便编辑人员的工作。

2、国内市场发展现状

国内编辑部采用的采编系统包括勤云8.0采编系统，中国知网“腾云"采编系统,三才采编系统,玛格泰克采编系统 Journalx 2.0等等。目前市场占有率较高的采编系统是玛格泰克采编系统和勤云采编系统。

玛格泰克长期从事新闻出版行业的应用系统开发和推广工作，其系统安全性和稳定性、系统功能强大和自定义的灵活性、用户界面、系统响应速度等方面均处于国内领先水平。但是目前采编只是实现部分电子化, 并未实现编校工作的电子化，采编系统无法实现参考文献的在线校对。

黑马校对等其他文字性校对工具可进行文字内容的校对进行错别字,政治性敏感词等的校对。黑马校对系统是国内市场上占有率最高、实用性最强的专业校对软件，为各类文稿的校对提供最佳解决方案。可是并不能做到对于论文的格式要求这方面的处理

虽然很多采编系统，文字校对软件已经很大程度的提高了编辑们的审稿效率，但是文章中一些微小的格式错误,却并没有在他们的考虑范围之内，且我们希望能够让用户根据需求设置审查范围

3、项目创新点

不同编辑社对于文稿审查要求不同。本项目可针对不同的需求设计文档自动审查系统的功能模型，提供不同的文档审查自动化系统应实现的功能。即可实现系统功能自定义化。

中文自然语言处理与英文自然语言处理有着很大的不同，因为英文中每个句子里的每个词是分隔开的，而在中文中不是。所以本项目将会面向中文稿件的识别、编辑与处理，基于中文文本进行自然语言处理，实现采编系统中文化。