

浙江农林大学

本科生毕业设计（论文）
中期检查报告

（ 2019 届）

题 目：在线商品用户评论分析系统设计与实现

学生姓名：陈怡

学 号：201505010323

专业班级：计算机科学与技术 153

学 院：信息工程学院

指导教师：黄雷君 职称：讲师

2019 年 3 月 10 日

目录

1	毕业设计的进展情况	1
1.1	课题工作完成情况	1
1.2	知识技能学习情况	1
1.3	职业素养和项目管理学习培养	1
1.4	课题相关技术（计算机技术）国际研究前沿了解情况	1
2	存在问题与解决方案	2
2.1	存在的主要问题	2
2.2	解决方案	2
3	可行性研究	3
3.1	经济可行性	3
3.2	技术可行性	3
3.3	操作可行性	3
4	前期任务完成度与后续实施计划	4

1 毕业设计的进展情况

1.1 课题工作完成情况

系统初步划分为六个主要模块，分别为管理模块、图表显示模块、爬虫模块、分析模块、交互模块、邮件发送模块。系统通过这六个模块组成，由交互模块处理用户的请求和完成不同模块间的调度。系统使用 Python 进行编写，Web 框架使用 Django 框架搭建。

目前已完成爬虫模块、交互模块、邮件发送模块的编写和测试，接下来需要完成管理模块、图表显示模块、分析模块的编写。在评论文本分析的分析模块中，使用不同的计算模型，分别将使用基于情感字典、基于机器学习以及基于百度 AI 开发平台进行编写多个分析引擎，通过配置文件进行切换。

1.2 知识技能学习情况

通过前期的摸索尝试，已完成对 Python 语言和 Django 框架使用的学习和实践。目前已经可以熟练使用 Python 语言和 Django 框架进行 Web 页面的开发。

对于自然语言处理部分开始学习主流的构建和使用方法。使用 Python 的 NLTK 库和 jieba 库进行文本的处理和分类器训练。

1.3 职业素养和项目管理学习培养

项目开始存放在本地，但有难维护、易丢失的缺陷。在 github 上创建私有库，通过 github 来维护代码，使得代码版本分明。

1.4 课题相关技术（计算机技术）国际研究前沿了解情况

中文文本的自然语言处理目前较为常用的是 Python 语言，Python 语言提供了较多且优异的第三方库以供使用，如 NLTK、jieba、sklearn 等。目前的研究方法主要有两类：一种是基于情感词典的方法，另一种是基于机器学习的方法。前者需要用到标注好的情感词典（英文/中文），通过情感词的褒贬性判断文本的情感倾向。后者是将传统的文本分类方法如朴素贝叶斯（NB）、k 近邻（kNN）、最大熵、支持向量机（SVM）等算法应用于情感分类。需要大量的人工标注的语料作为训练集，通过提取文本特征，构建分类器来实现情感的分类。

2 存在问题与解决方案

2.1 存在的主要问题

- （1）评论爬取和分析的时间周期较长，对用户的体验不够友好。
- （2）自然语言处理目前对于英文研究较为广泛，资料也较多，但对于中文的处理资料较少。

2.2 解决方案

引入数据库保存爬取的数据和分析的结果，在用户第一次输入的时候开始评论的抓取和分析，并告知用户“已加入分析稍后再试”。同时设置邮件提醒功能，在用户输入商品 URL 的同时可以输入邮箱地址，在数据分析完成后，给用户发送提醒邮件，提醒用户分析已完成，可以进行查看结果。翻阅更多的资料，查找与中文自然语言处理相关的文档，进行深入的研究分析。

3 可行性研究

3.1 经济可行性

本系统开发工具采用 pycharm、vscode、datagrip，其中数据库使用 Mysql，开发语言使用 python，Web 框架使用 django。数据库部署平台为 Linux，Web 应用部署平台为 Windows。以上产品，对于个人为免费产品。开发成本低，经济方面可行。

3.2 技术可行性

系统使用 django 作为基准开发框架。python 的用途广泛，目前分别可以应用于爬虫技术、数据挖掘、数据分析、web 交互等。技术比较成熟，可行性高，并且具有众多第三方开发库。对于爬虫，python 具有强大的 requests 第三方开发库，可以使用较少的语句实现一个简单的 http/https 爬虫请求发送和处理。对于数据分析处理，python 可以使用 jieba 分词进行中文文本的分词。NLTK 是在自然语言处理中最常使用的一个 python 库。SnowNLP 可以进行中文文本内容的分析和挖掘。更进一步，各大 IT 公司具有许多强大的 AI 平台，如百度 AI 开放平台、腾讯 AI 开放平台等开放平台可以使用，其中百度 AI 开放平台具有免费、使用方便等优点，本系统借助百度 AI 开放平台完成一部分的评论分析引擎的编写。系统使用 Mysql 作为数据库引擎，Mysql 具有简单易用，对于个人开发者免费，具有轻量化，性能较强，可移植性高、运行速度较快等优点，且部署方便。系统各方面技术均成熟可用，在技术方面可行。

3.3 操作可行性

本系统设计清晰，具有简单易操作的用户界面，有完善的异常处理机制和错误信息提醒和后台日志记录，用户通过操作界面的提示即可完成操作。通过系统的后台管理界面可以操作系统数据库，方便完成日常的数据维护。操作方面可行。

4 前期任务完成度与后续实施计划

基本完成系统框架的搭建和设计，接下来需要将管理模块、图表显示模块、分析模块编写完成。主要工作在于分析模块的编写，考虑到系统的后续扩展，将分析模块进行独立编写。使用 HTTP 接口进行调用。最后完成系统的整体调试和验收。