## 广东东软学院本科毕业设计（论文）任务书

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **学 院** | 计算机学院 | **专 业** | 软件工程 | **班 级** | 22软工专升本1班 |
| **学生学号** | 22215150135 | **学生姓名** | 何超华 | **指导教师** | 宫翠峰 |
| **选 题** | 基于机器学习的中文情感识别系统的设计与实现 | | | | |
| **选题类别** | **毕业设计 （√） 毕业论文 （）** | | | | |
| **选题类型** | **理论性课题（） 实践性课题（√）**  注：企业项目都属于实践性课题 | | | | |
| **选题来源** | **企业项目 （） 自拟题目 （√）**  注：实践性课题也可是自拟题目 | | | | |
| **一、主要内容**  基于机器学习的中文情感识别系统是对大量文本数据进行分析，将文本数据与情绪类别进行关联（正性、中性、负性），在关联情感类型后再与地区关联，最后对各种数据分析的结果进行可视化，供使用人员针对具体问题分析并采取必要措施。该系统具有前端和后端，前端给用户提交数据集以及展示数据可视化结果，后端的情感分析主要分为如下流程：   1. 文本获取：通过爬虫爬取数据或使用用户提交的csv等格式的语料数据。 2. 文本预处理：对收集的数据进行预处理，包括去除无关字符、分词、去除停用词等。 3. 数据标注：对预处理后的数据进行标注，标注的内容一般是情感极性，如正性、中性、负性。 4. 特征提取：从预处理后的数据中提取出有用的特征。比如词频、词性、情感词汇、N-gram等。 5. 模型训练：选择SVM、决策树、朴素贝叶斯等机器学习算法对上面提取的特征和标注的数据进行模型的训练。 6. 模型评估：使用测试集来测试模型的性能，如准确率、召回率、F1值等。 7. 运用模型：使用训练好的模型对数据进行识别分析。 8. 数据分析可视化：生成词云图、雷达图、区域分布图等。 | | | | | |
| **二、基本要求**  学生在系统的分析与设计阶段，应通过问卷调查方式了解用户需求，查阅和体验国内外优秀的情感识别分析系统，分析和总结用户痛点。着重对爬虫、SVM、朴素贝叶斯等技术收集资料和参考文献。根据软件工程思想和软件开发过程，经过需求调研、需求分析、系统总体设计、系统详细设计、系统编码实现和系统测试与试运行等关键性阶段，设计并实现一个基于机器学习的情感识别系统。  学生所呈交的毕业设计（论文），应是本人在指导老师的指导下，独立进行的设计（研究）工作及取得的成果，论文中引用他人的文献、数据、图件、资料均已明确标注出。对本文的研究做出贡献的个人和集体，应在论文中作明确的说明。  学生应加强与老师的沟通，在规定时间内按要求完成各个阶段材料的撰写、提交和归档工作。  毕业论文应根据广东东软学院本科生毕业设计（论文）撰写规范（2024年版），并根据学校提供的广东东软学院本科毕业设计（论文）参考格式进行撰写。论文正文不得少于10000字，查重低于等于30%。 | | | | | |
| **三、工作进度安排**   |  |  |  |  | | --- | --- | --- | --- | | 序号 | 工作内容 | 工作程序及要求 | 完成时间 | | 1 | 开题报告 | 1.查阅文献，熟悉中文情感识别系统的业务流程，和机器学习的技术要求，加深对本选题的理解。  2.学生在老师指导下，完成开题报告；  3.指导教师对开题报告的质量进行审查 | 2023.11.6-12.8 | | 2 | 毕业设计作品实现 | 根据审查后的开题报告，学生基于机器学习实现一个完整的中文情感识别系统，通过测试并正确运行。 | 2023.11.20-2024.2.29 | | 3 | 中期检查 | 1.学院组织专业进行毕业设技（论文）中期自查工作；  2.质保部组织校级中期检查,包括前期检查的整改落实情况，开题报告质量检查等 | 2023.12.11-12.22 | | 4 | 论文定稿及查重工作 | 1.学生在老师指导下完成论文定稿，并在规定时间内上传系统；  2.指导教师在系统中完成学生论文定稿查重检测 | 2023.12.25-2024.3.8 | | 5 | 评分工作 | 1.指导教师进行论文评阅并给出成绩和评语；  2.指导教师成绩合格后再由评阅教师评阅并给定成绩和评语 | 2024.3.11-3.22 | | 6 | 答辩资格审查、答辩准备 | 学院审查学生答辩资格、制定答辩工作安排表并做好学生通知工作 | 2024.3.11-3.29 | | 7 | 答辩及成绩评定 | 学院答辩委员会根据答辩流程组织答辩工作，各答辩小组做好答辩记录，并填写答辩记录表和答辩成绩评定表，并根据情况确定是否组织第2次答辩 | 2024.4.1-5.10 | | 8 | 成绩登记 | 学院指定老师填写成绩评定总表，并汇总学生成绩，通过教务系统登记成绩。 | 2024.5.6-5.24 | | 9 | 优秀毕业设计（论文）、优秀指导教师评选 | 教务部组织开展优秀论文、优秀指导教师评选工作：推荐、审核、公示、发文 | 2024.5.6-6.7 | | 10 | 材料存档、工作总结 | 教务部、各学院做好毕业设计（论文）材料存档和工作总结。 | 2024.6.3-6.21 | | 11 | 论文质量检查 | 质保部组织校、学院两级督导和校外专家抽查学生论文 | 2024.8.26-9.6 |   以上毕业设计（论文）安排的各阶段完成时间将根据学校及二级学院实际工作安排适当调整 | | | | | |
| 1. **应收集的资料及主要参考文献**   [1] 梁锡豪. 面向社交文本的情感识别研究[D].清华大学,2019.  [2] 刘鑫. 在线医疗社区中文本热点主题识别与情感分析方法研究[D]. 重庆:重庆邮电大学,2019.  [3] 贺小雨.基于情感识别的案件舆情统计分析[D].黑龙江大学,2020.  [4] 李彩玲.网络健康社区的主题识别与情感分析研究[D]. 湖北:华中师范大学,2020  [5] 吴飞.面向新闻实体的内容识别与情感分析算法应用[D]. 湖北:中南财经政法大学,2020.  [6] 左洋.基于网络文本数据挖掘的大学生关注点识别及情感分析研究[D].山东:山东科技大学,2020.  [7] 王韩林.面向社交媒体的短文本反讽识别与情感分析[D].重庆:重庆邮电大学,2022.  [8] 郑少伟.基于统计机器学习的微博用户角色识别及情感分析[D].河北:燕山大学,2022. | | | | | |

**指导教师签名：**

**任务下达日期：** **年** **月** **日**