哈尔滨理工大学

本科毕业设计（论文）工作

周(日)志

（2022～2023年 第一学期）

学院：计算机科学与技术学院

专业：数据科学与大数据技术

班级：大数据19-1班

姓名：马超

学号：1914050120

指导教师：李双翼

|  |
| --- |
| 2022年 9月 12日 星期一 指导教师签字：李双翼 |
| 本周的主要工作是确定选题，并撰写任务书。经过对专业知识的回顾，结合老师的意见，最终确定选题为基于机器学习的智能单词打卡系统。在确定选题后，结合软件工程所学，参照软件的开发流程，对系统的开发任务进行规划，并基于此撰写任务书。 |
| 2022年 9月 19日 星期一 指导教师签字：李双翼 |
| 本周的主要工作是完善任务书，并进行技术调研。技术调研工作如下：  1. 了解机器学习的智能单词打卡系统：了解单词打卡系统的历史发展、现状及其发展趋势，了解当前市面上的单词打卡，以及它们之间的优劣势。 主要对不背单词进行了调查，调查方式为体验系统的功能。  2. 收集智能单词打卡系的信息：收集关于单词打卡的资料，可以在网上搜索、查阅图书、参加学术会议和研讨会、调查用户反馈等。主要通过查阅文献，对机器学习算法进行学习。  3. 对相关技术的分析：分析目前单词打卡系统的技术架构，以及它们的优势和劣势，以及它们之间的区别和相似之处；对比不同的技术方案，评估其优势和劣势，确定最优的技术方案。这一步主要在各大开源网站和技术论坛对已有的开源项目进行对比分析。 |
| 2022年 9月 26日 星期一 指导教师签字：李双翼 |
| 本周的工作主要是基于上周完成的技术调研工作，对系统进行需求分析。在软件工程中，需求分析是一个系统开发过程中非常重要的环节，它可以帮助设计人员和开发人员了解用户的需求并确定单词打卡的功能和范围。本周主要结合现有的成熟的单词打卡，对系统的观感需求、易用性需求、完整性需求和可执行需求以及可维护性需求进行分析。 |
| 2022年10月 3日 星期一 指导教师签字：李双翼 |
| 本周继续完成对系统的需求分析。采用了问卷调查的方式对用户需求进行收集，以便得到更客观的需求。问卷的内容主要还是从系统的观感、易用性进行设计。另外还设计了系统的功能性需求的收集问卷，主要从用户和打卡两个角度来设计问卷。 |
| 2022年10月 10日 星期一 指导教师签字：李双翼 |
| 本周主要对上周问卷调查收集所得数据进行分析，整合两周以来的数据，采用单一职责模式，将系统的目标和功能进行划分，采用开闭原则满足系统的可维护性。根据以上需求分析对应的业务流程，并基于此对系统的架构进行分析，确定系统的架构。 |
| 2022年10月 17日 星期一 指导教师签字：李双翼 |
| 本周的主要工作是继续完善系统的需求分析，了解当下热门开发技术，根据需求分析，完成对项目的技术选型，确定项目的开发环境和运行环境。上网查阅文献资料，了解常见的软件需求说明书格式，对各种标准的软件需求说明书进行对比，选择合适的标准，完成项目需求说明书的撰写，并基于当前所完成的工作，进行开题的报告撰写。 |
| 2022年10月 24日 星期一 指导教师签字：李双翼 |
| 这周的任务主要是根据各位老师提出的开题评审意见对开题报告进行修改和完善。主要对项目的技术选型进行适当调整，使项目的技术栈更贴合专业以及实际需求，并对项目所用技术进行回顾、学习。 |
| 2022年10月 31日 星期一 指导教师签字：李双翼 |
| 这周的主要工作是基于前面的需求分析，对系统进行概要设计，主要内容是结合系统的功能性需求，绘制用例图，将之转化为系统的软件结构和数据结构。软件结构上，主要将系统的复杂功能按模块进行划分，简历系统的层次结构。数据结构上则主要从用户和打卡两个角度出发，确定系统的数据结构。 |
| 2022年11月 7日 星期一 指导教师签字：李双翼 |
| 本周的主要工作内容是根据上周的概要设计内容对当前项目的总体设计的可行性进行评估，并检查现有的模块是否完整，每个功能是否都有对应的模块来实现，并结合技术选型中的前端技术，对用户界面进行设计。 |
| 2022年11月 14日 星期一 指导教师签字：李双翼 |
| 本周的主要工作内容是根据前两周的工作内容完成对项目概要设计说明书的撰写。先对系统进行总数，阐明系统的需求和目标，并指明系统的环境。利用功能结构图对系统的总体设计进行阐述，并说明系统的组织结构，处理流程以及模块间的关系。 |
| 2022年11月 21日 星期一 指导教师签字：李双翼 |
| 本周的主要工作内容是基于前面的工作对系统进行详细设计。主要是在开发环境中完成对数据库的构建。数据库首先设计了用户表和机器学习表用于保存用户信息和电影信息，并在这两个表的基础上引出其他必要的表，如模型训练表、用户数据表以及单词数据表等。使用Navicat软件对MySQL数据库进行操作，简化数据库的构建工作。 |
| 2022年11月 28日 星期一 指导教师签字：李双翼 |
| 本周的主要工作内容是继续系统的详细设计。主要是对用户界面进行详细设计，包括用户的登录注册界面及其前后端的交互，数据库中为用户添加一个cookie表来保存用户的登录信息，为解决冷启动问题，添加用户偏好表保存用户的初始标签；打卡任务和详情页的界面以及业务流程的设计；打卡任务评判功能的设计，包括评分业务流程的设计，并对数据库进行改进，添加基于用户表的机器学习训练模型表用来保存用户对单词的学习情况的记录。 |
| 2022年12月 5日 星期一 指导教师签字：李双翼 |
| 本周的主要工作内容是继续完成系统的详细设计，主要是对系统的核心算法进行设计。朴素贝叶斯算法可以用于识别每个单词属于哪一类别，例如识别一个单词是名词、动词、形容词等。具体来说，朴素贝叶斯算法可以通过统计每个单词在不同类别中出现的频率来计算概率，然后根据概率大小进行分类。在设计算法时，考虑到的因素主要有：数据预处理。对于从网上获取的单词数据，其维度复杂，并且脏数据较多，需要对数据进行去重、降维、缺失值处理以及归一化处理。相似度计算。对用户之间与电影之间的相似度进行计算时，可选择的算法较多，如余弦相似度、皮尔逊相似度等。 |
| 2022年12月 12日 星期一 指导教师签字：李双翼 |
| 本周的主要工作内容是继续完成系统的详细设计，包括vue框架的搭建，数据库的连接和API的开发。vue选用比较稳定的正式版，查看vue的官方文档，使用 vue 的 sequelize 框架来连接数据库，并实现数据的增删改查操作。学习vue的API框架，设计登录、推荐、缓存等API。基于以上工作，完成对项目详细设计说明书的撰写，内容包括背景、程序设计说明、功能、算法、 流程逻辑接口等内容。 |
| 2022年12月 19日 星期一 指导教师签字：李双翼 |
| 本周的工作内容主要是中期报告的撰写。主要完成中期报告的工作进展模块。内容主要包括需求分析、概要设计、详细设计的呈现。再次对软件系统开发生命周期进行复习，检查和完善系统设计目标以及技术方案的可行性分析，修改概要设计和详细设计中存在的问题包括系统架构图和程序流程图的修改。分析系统对社会、对健康、对安全、对文化等方面的影响。查阅资料，分析实际开发中，项目的管理、维护成本来源。 |
| 2022年12月 26日 星期一 指导教师签字：李双翼 |
| 本周的工作内容主要是继续完成中期报告的撰写。主要完成中期报告中毕业设计工作存在的问题及解决方案以及下一步工作预测及可能存在的问题两个模块。在前一个模块中主要回顾前面需求分析、概要设计以及详细设计所完成的工作，对其中比较深刻的问题进行重点论述，并说明解决思路和方案。在后一个模块中，基于当前已完成的工作，对后续编码阶段的工作进行预测，分析可能存在的问题，重点关注朴素贝叶斯算法以及随机森林算法的学习应用。 |