**SJTU公司**

**立项建议书**

项目名称：教学辅助平台

项目组组号：5

项目组负责人： 李耀

联系电话：13016060293

电子邮箱：lixyao0615@163.com

**2022年 10月**

# 项目的必要性

2020年,由于新冠病毒在全球的广泛传播，世界范围内开展了前所未有的在线教育实践。其中，中国教育改革发展的实践尤为成功，已经步入了全球最前列。中国互联网教学的新样态正在加速形成，无论是规模庞大的信息化基础设施改造工程,还是高校师生素养提升工程，都在推动着中国教育向现代化与国际化发展。

技术推动教育进步，教育推动社会进步。在“互联网+教育”的价值受到普遍认可的当下，如何借助新一代信息技术支撑、深化互联网学习体系，完善各领域与区域互联网学习的实践应用，促进优质教学资源开放共享，正成为国内高校及教育机构普遍关注的新议题。“互联网+”这一理念的提出并非意味着对传统纸质文件的摒弃，在尚且无法以电子存储设备完全替代纸质存储介质的情况下，通过互联网来实现对存储于纸质介质的知识的充分利用依旧是当前主要的发展方向。互联网与教育的结合突破了传统的课堂在教学时间和空间上的限制，但是随之而来的也有互动和沟通的问题。

根据本项目组成员以学生的身份对各校多个网络教学平台的实际使用体验以及从导师处获取的实际需求，建立一个能够通过较为方便的手段获取纸质文本信息、有助于师生之间以及学生之间交流的教学辅助平台，实现对教学管理、使用交互体验的加强，是有确实的需求的。

本项目以教学辅助系统为基础平台，通过运用数据仓库、光学字符识别（OCR）等新技术，结合教学辅助系统的设计，完成对包含公式的纸质文件的扫描、文字识别和数据的存储，从而构建一个可以适用于文本与公式识别、能够实现权限管理的“算法教学辅助平台”，最终实现对后续应用的算法课程的稳定支持。本项目软件还可以向其他不同的课程进行扩展，帮助教师完成教学发展培训，协助教师建设新课程，并最终形成集成多门学科的教学辅助系统。

# 项目外部条件落实情况

## 技术基础

本项目采用前后端分离的开发方式，基于浏览器-服务器（B/S）架构进行部署。前端基于HTML以及Javascript进行开发，采用vue框架；后端基于Python进行开发，采用Django框架。数据通过Mysql数据库进行存储，测试通过单元测试组件Jest和PyUnit等进行。

本项目采用的技术发展的比较成熟，前后端使用的vue以及Django框架是目前主流的开发框架。同时，本项目旨在实现实际纸质教材的文本识别，通过对当下准确度较高的OCR技术的集成和深化开发，能够对包含公式的文本进行扫描转换。

## 项目组织机制设计

本项目开发成员由3名成员组成，按照项目管理规范和基于UML的面向对象的开发方法，实现对项目的设计与版本管理；由项目组组长对本项目的开发情况进行管控，早期迭代版本由两名成员负责前端开发，一名成员负责后端开发以及数据库设计，后续迭代版本根据实际开发情况调整人力资源分配以保证项目能够在预定的时间内完成。

## 市场前景与市场基础

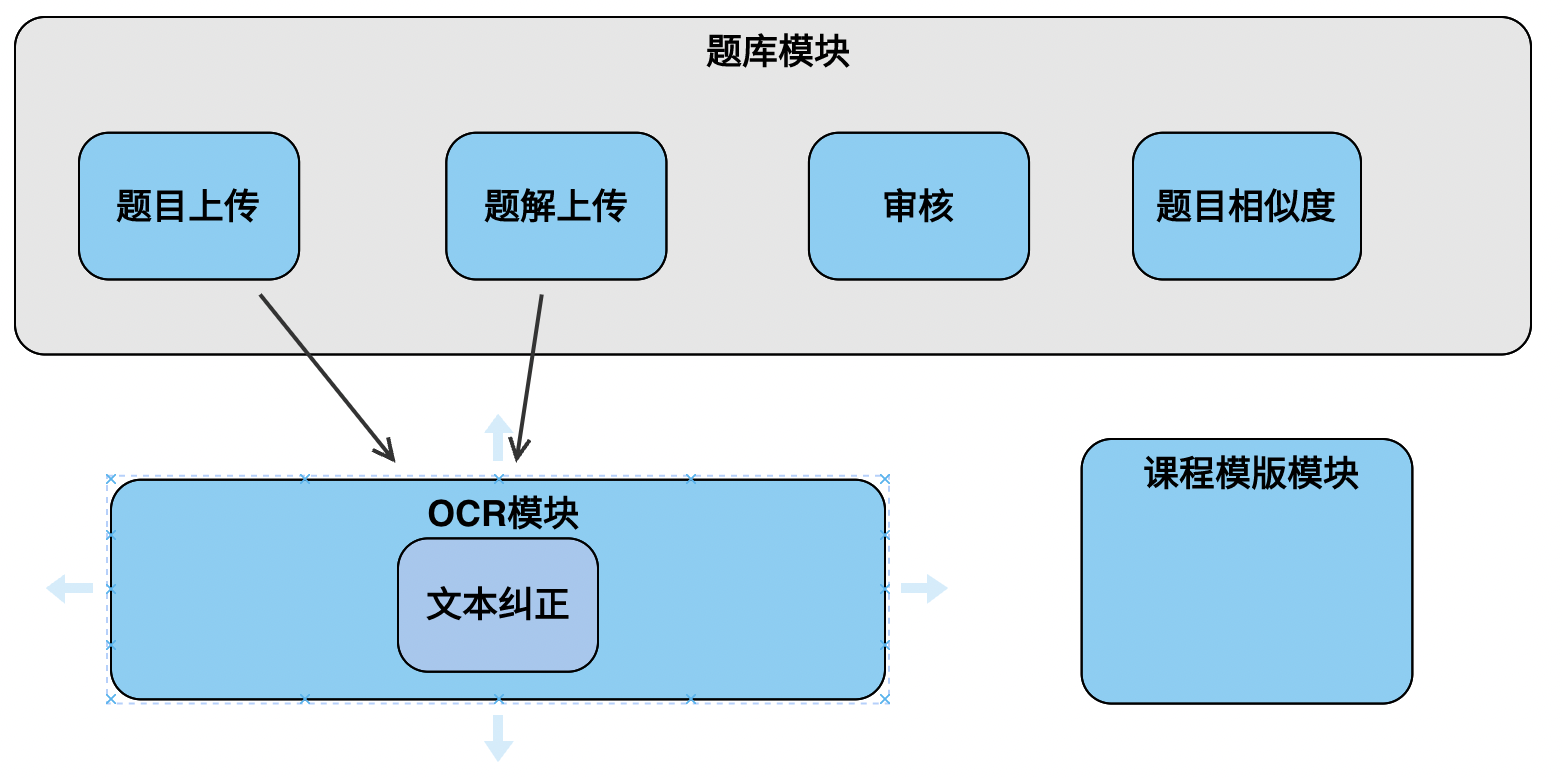
中国网络教育市场总体仍处于起步阶段。随着中国的信息化程度、网民对网络教育认知程度、学历教育社会的认可度的提高，网络教育市场规模增长速度很快。2022年全国教育经费总投入为53014亿元，相较去年同比增长5.65%。在线教育行业是教育技术变革的产物，在互联网和移动智能设备的发展和普及下，在线教育使得传统线下教育的时空限制被打破。

目前国内的远程教育市场主要分为三个发展方向：一是以优质课程内容取胜的富媒体教学网站，如各大院校及培训机构的线上课程等；二是以社交网络为核心理念的平台型网站，如谷歌等企业的在线教育平台；三是针对特定群体用户，主打线下教育与线上平台相结合的社交网站教育平台，如大街网教育等。本项目用户群体清晰、功能创新点明确，且对用户需求有持续、深入的调研，有一定的市场基础。

# 项目目标和创新点

## 项目目标

1. 本项目“算法教学辅助平台”主要是为教学活动服务，用户受众群体主要为学生和老师。系统通过整合教学教材资源，提供一个平台给学生巩固和交流所学知识，加深对知识的理解和运用。系统主要包括三个模块：题库模块，OCR模块和课程模版模块。



项目整体架构图

1. 题库模块主要包括题目的上传、题解上传、审核和题目相似度四部分。题目和题解的上传可以使用OCR技术进行扫描，同时题解也可以使用.md格式的文件。题目相似度部分可以对用户上传的题目与题库中存在的题目进行相似度匹配，然后自动归到正确的类别下面，实现自动分类功能。审核部分主要是对用户上传的题解进行审核，解答正确则打上正确标签并置顶展示。
2. OCR模块主要使用OCR（光学字符识别）技术对资源进行扫描识别上传，增加了系统使用的便利性。系统还会对OCR识别出来的文本进行语法语义纠正，提高识别的准确性。
3. 课程模版模块通过制定一个课程模版，可以方便快速的进行其他课程的扩展。本系统主要面对算法课程，但使用课程模版可以简便的应用于其他的课程。

## 创新点

1. 项目集成了OCR技术对课程资源的上传，减少了用户手动输入题目的繁琐，大大提升了系统的便捷性。OCR模块能够识别复杂的公式，且同时支持多道题目同时识别，具备自动分题功能，极大的减少了用户的工作量。并且系统集成文本语法纠正功能，可以对OCR识别出来的文本进行检查纠正，提高识别的正确率。
2. 系统具备题目相似度计算功能，能够对两道题目之间进行相似度的计算和匹配，从而能够将题目放到最相似的集合之中，实现自动分类题目的功能，减少人工在类别中添加题目的工作量。
3. 项目具备课程模版模块。通过使用模版可以快速生成另一门课程的教学辅助平台，使系统不仅仅局限于一门课程，大大增加了其的可扩展性。

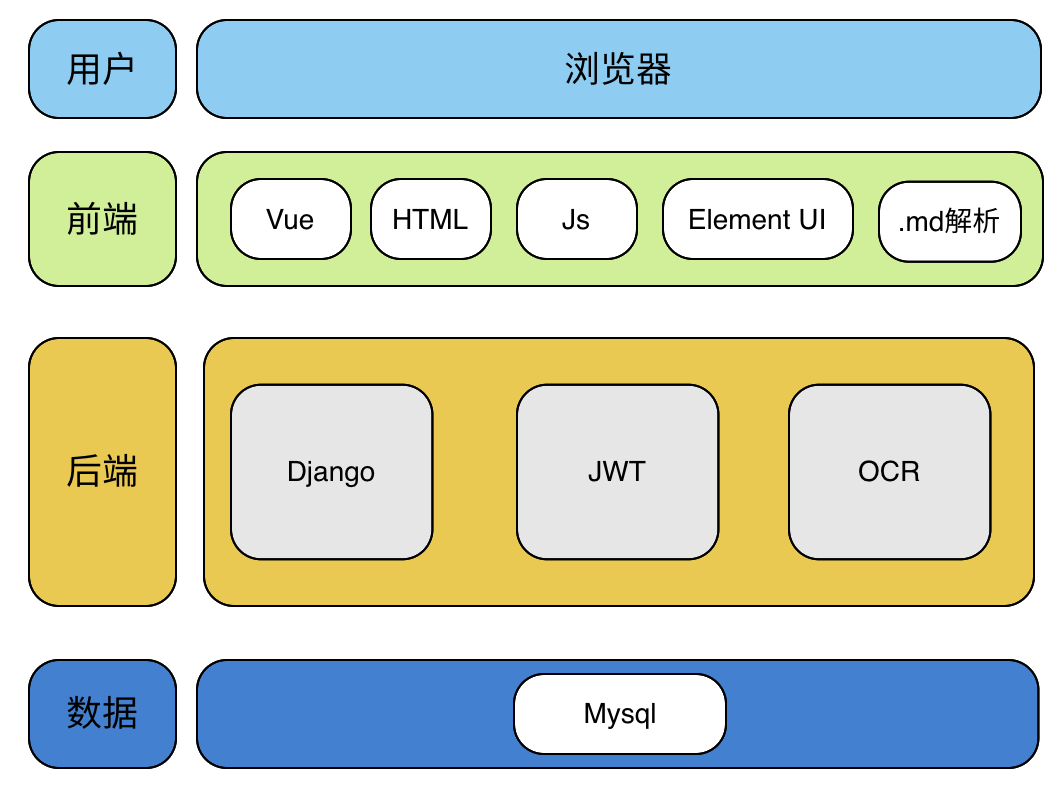
# 项目方案和可行性分析

## 项目前景

1. 项目包含算法课程的相关题库，按照类别/章节进行分类。用户可以上传.md格式的题解文件作为解答。上传的题解会经过审核，正确通过则会打上正确标签并置顶展示。用户可以在系统平台上对相关知识点进行巩固和交流。
2. 项目使用OCR技术进行相关资源的提取上传，可以很好的减少题目上传的工作量，同时具有自动分类题目的功能，大大增加了系统的便利性。
3. 项目同时具备课程模版的功能，能够快速的生成另一个课程的辅助平台，通过使用模块快速导入课程的资源，扩大了系统的使用范围，增加了系统的可扩展性。

## 技术方案

1. 项目使用基于UML的面向对象的开发方法，通过用例图、类图等明确系统的整体情况过。采用前后端分离的开发模式，使用BS(Browser-Server)架构进行开发。前端主要采用vue框架,使用WebStorm软件进行开发，后端主要采用python语言和Django框架，使用PyCharm软件进行开发。前后端通过使用ajax和REST进行通讯。
2. 项目数据库使用的是Mysql关系型数据库进行数据的存储和处理；建模工具主要使用VP Online；测试工具主要使用单元测试Jest和PyUnit，性能测试使用Locust。



技术架构图

## 可行性分析

1. 项目所使用的技术都比较成熟可靠。前后端开发分别使用主流的开发框架，具有比较完整的应用链路。而OCR技术目前也比较成熟，市面上有相应的插件可以调用，我们只需要对其进行适当的改进。总的来说，整个项目并不具备无法攻克的技术难点，能够顺利完成开发任务。
2. 项目的定位是“教学辅助平台”，旨在通过集成相关教学资源帮助学生更好的学习和巩固知识点。项目的用户群体主要为学生和老师，受众群体存在且数量充足，可以进行开发。
3. 项目组成员为3人，所用的技术框架都为开源的，开发周期约为3个月，开发的人力成本和时间成本都不高。系统完成后能很好的服务于教学活动，提高教学的质量和效率。

# 计划进度

本课题计划采用迭代开发过程，自2022年10月8日开始，于2023年1月13日结束，用时三个月。

## 风险分析

5分制度

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **优先级** | **风险名称** | **说明** |
| 1 | 5 | 需求不明确和可能变更 | 可能性：80%｜影响度：70｜风险暴露量：50 |
| 2 | 5 | 开发人员不够熟悉Vue和Django框架 | 可能性：80%｜影响度：60｜风险暴露量：48 |
| 3 | 5 | 项目进度不能被维持 | 可能性：80%｜影响度：60｜风险暴露量：48 |
| 4 | 3 | 开发人员同时充当测试人员，影响测试质量 | 可能性：50%｜影响度：50｜风险暴露量：25 |
| 5 | 2 | OCR识别精度不够 | 可能性：60%｜影响度：20｜风险暴露量：16 |
| 6 | 1 | 题目相似度计算的准确率不够高 | 可能性：30%｜影响度：50｜风险暴露量：5 |
| 7 | 1 | OCR识别矫正功能无法实现 | 可能性：20%｜影响度：25｜风险暴露量：2.5 |

## 迭代

|  |  |  |  |
| --- | --- | --- | --- |
| **迭代周期** | **迭代名称与交付** | **任务** | **应对的风险** |
| 2周  10.08-10.21 | 分析需求，建立原型界面。交付界面原型图和需求文档。 | 1. 与客户沟通，获取需求。 2. 分析需求。 3. 根据需求，建模，并使用Azure设计原型图。 4. 拿原型图与客户进一步沟通，优化原型界面。 | 风险1:需求不明确和可能变更。 |
| 5周  10.22-11.25 | 软件架构设计，对题库分类和上传，OCR识别和课程模版等基本功能进行详细的需求分析、设计和实现，构建初步的原型系统。交付具备基本功能的可用的初步原型体系统。 | 1. 前端完成基本功能的页面，将mathpix集成到前端界面。 2. 后端完成基本功能的业务逻辑。 3. 分别进行单元测试。 4. 前后端完成对接和部署。 5. 进行集成测试和系统测试。 | 风险2:开发人员不够熟悉所使用框架。  风险3:项目进度不能被维持  风险4:开发人员同时充测试人员，测试质量无法保证  风险5:OCR识别精度不够 |
| 4 周  11.26-12.30 | 根据客户反馈，优化系统，同时开发OCR矫正和题目相似度匹配功能，将其集成到系统中。 | 1. 优化系统。 2. 开发OCR矫正功能和题目相似度计算功能。 3. 对纠正功能和题目相似度功能进行单元测试。 4. 将纠正功能和题目相似度计算功能集成到系统中。 5. 进行系统测试。 | 风险3:项目进度无法维持  风险4: 开发人员同时充测试人员，测试质量无法保证  风险6:题目相似度计算准确率不够高  风险7:OCR识别矫正功能无法实现 |
| 2周  1.1-1.13 | 持续优化系统，并进行项目验收。 | 1. 根据客户反馈，优化系统。 2. 整理项目成果，准备验收事宜。 3. 项目验收。 |  |

# 项目预期成果

1. 立项建议书
2. SRS软件需求规约
3. 用例图、活动图、类图、时序图、通信图、包图
4. 架构设计文档
5. “算法教学辅助平台”的源代码、文档（使用手册）以及安装包
6. 软件测试计划和软件测试报告
7. 4个迭代的迭代计划和迭代评估报告
8. 项目总结报告
9. 申请软件著作权和软件产品登记

# 项目社会经济效益

1. 社会效益

如今市面上有较多的算法平台，其中面向求职的有LeetCode、牛客网、FreeCodeCampus，面向算法竞赛的有codeforces、acwing以及各大高校自己的OJ系统。但是市面上很少有纯粹的算法教学平台，重心不在于刷题形成的“肌肉记忆”，而是像《算法导论》那样更加专注于算法本身的思想及其形式化的证明。

本平台就是面向这一需求开发。减少了普通刷题网站共有的编译和提交评分模块，更多的是通过类博客的方式来分享算法思想、发布作业习题和回复解决方法。为了方便大家，本平台会与博客平台一样，支持通过MarkDown文件来分享和发布问题。

本项目以交大软院的算法课教学为依托，以博客平台为模版，将其改造成教学、算法交流与解答于一体的算法教学辅助平台。

1. 经济效益和前景

本项目面向的是校内算法教学，本质是为算法服务的博客平台，因此本平台可以在校内进一步扩展，为校内各专业的同学提供交流和学习算法，分享算法资源的平台。具有一定的扩展潜力。