

附件 1:

汕头大学大学生创新创业训练计划项目
进展日志

项目编号: 202310560203

项目名称: 面向计算机程序设计竞赛的在线评测系统

项目负责人: 张健

项目组成员: 张嘉豪、张雷明、谢璧蔚

指导教师: 王珊珊、许建龙

资助金额: 2000

项目预计开展时间: 2023 年 5 月 至 2024 年 5 月

表 1: 研究（创新项目）记录表

项目名称		面向计算机程序竞赛的在线评测系统		
研究时间		2023 年 5 月	研究地点	汕头大学东海岸校区
参与人员	姓名	学号	年级专业	任务分工
	张健	2020611073	2020 级计算机科学与技术	相关应用后端开发
	张嘉豪	2020611072	2020 级计算机科学与技术	相关应用前端开发
	谢璧蔚	2022611047	2020 级计算机科学与技术	相关应用后端开发
	张雷明	2020612037	2020 级数据科学与大数据技术	相关模型生成实现
一、 研究（实验）内容 1. 利用无人机队汕头大学桑浦山校区进行实景采样,并使用深度学习等方法进行三维重建。 2. 将生成的模型进行简单的应用和展示。				
二、 研究（实验）目的 先前校园在线评测 CodeZ 的实现过于粗糙,缺少用户社区,判题逻辑有限。为了解决这个问题我们将利用新的技术和方法重新开发功能更加全面的在线评测平台。				

三、 研究（实验）手段

1. 通过迭代增量式开发，每次设计和实现该在线评测平台的一部分，逐步逐步完成开发。在整个迭代开发方法中，整个开发工作组织为一系列短期的小项目，即一系列的迭代。每一次迭代都包含了需求分析、设计、实现与测试。采用这种方法，开发工作可以在需求被完整确定之前启动，并在以此迭代中完成系统的一部分功能与业务逻辑的开发工作。再通过用户的反馈来细化需求，并开始新一轮的迭代。
2. 我们将传统的单体架构的在线评测系统拆分成了更细粒度的服务单位，如主要服务、图床服务、推荐系统服务、判题服务以及文件服务等。每个服务单位都可以进行独立的开发、测试、部署和交付，每个服务单位都可以做到独立的自治。

四、 研究（实验）步骤

1. 展示迭代后的网站。
2. 通过共享开发文档填写修改意见/潜在 bug。
3. 通过服务迭代完成 bug 修复、功能实现、服务升级。

五、 研究（实验）数据/结论

目前已经完成 DOJ 测试版的搭建工作，进入迭代开发。

六、 存在的问题和解决方案

主要问题：

1. 本项目开发中，部分成员已经外出实习，暂时无法保证在校服务器的网络与供电。远程使用校内服务器开发时会出现断电、网络欠佳等情况。
2. 各个 OJ 平台的题面格式各有不同，爬取不同的平台时需要设计不同的正则捕获。
3. 在追踪其它 OJ 平台的提交时不同平台的提交返回值有着不同的字段，但不同的字段均有意义，不能略去。
4. 总题目数量过多，无法人工的为所有题目增设标签。
5. 在各大现有 oj 平台。存在因为数据极其稀疏，和数据清洗问题。对模型的性能造成严重限制。比如：由于算力资源开销巨大，爬取时间窗口长度受限，能爬取到的数据集资源通常是有限的，用户的上线时间总是捉摸不定，我们无法完整的获取到一个用户在完整的活动周期下全部的题目点击量。其次，由于冷启动问题无法避免的，我们在尝试考虑上下文信息的辅助利用。

应对思路与措施：

1. 使用云服务器保证服务器的网络与供电稳定。校外开发时也可以快速的完成服务的部署。
2. 对每个 OJ 平台的题面设计专用的正则表达式捕获关键字段。
3. 跟踪提交时，除了捕获关键字段外，也会对 html 页面进行存储，避免错过其它有价值的字段。
4. 使用 bing 的翻译、搜索接口，尝试搜索每个题目的描述和题解，并捕获题解中的关键字作为该题的标签。
5. 对于数据集进一步进行更多信息的爬取，对机器人帐号等一些违法用户

进行更深层次的排查。确保模型训练的数据集是一个干净可利用的数据集。对于冷启动问题，我们在尝试考虑上下文信息的辅助利用，然后构建上下文图领域信息的传播来加强图结构信息的利用来丰富协同信息，我们还会采用 Xgcn 的做法，来缓解模型梯度下降的压力，并根据其机理尝试提升模型的泛化性。

七、 心得体会

本项目进行到中期，不断遇到新的困难，大规模数据处理的环境更加复杂，需要不断的尝试和实践才能更加深入理解其原理和问题，总的来说仍有许多需要学习的地方。

项目全体成员（签名）：

附件 2:

项目小组研讨与交流会议纪要

时间	6 月 1 日	地点	A 栋	记录人	张雷明
参加人员	张健、张雷明、谢壁蔚、张嘉豪				
会议主题	项目概况				
会议记录					
1.讲解项目基本情况					
2.分配基本任务					

注：项目负责人每 2 周应组织一次项目组成员研讨会，请作好记录。

项目小组研讨与交流会议纪要

时间	9 月 15 日	地点	A 栋	记录人	张雷明
----	----------	----	-----	-----	-----

参加人员	张雷明、谢璧蔚、张健、张嘉豪
会议主题	项目进度
会议记录 项目进度安排	

注：项目负责人每 2 周应组织一次项目组成员研讨会，请作好记录。

项目小组研讨与交流会议纪要

时间	9 月 15 日	地点	A 栋	记录人	张雷明
参加人员	张雷明、谢璧蔚、张健、张嘉豪				
会议主题	项目进度				

会议记录

项目进度反馈

注：项目负责人每 2 周应组织一次项目组成员研讨会，请作好记录。

项目小组研讨与交流会议纪要

时间	9 月 30 日	地点	A 栋	记录人	张雷明
参加人员	张雷明、谢璧蔚、张健、张嘉豪				
会议主题	项目进度				

会议记录

项目进度反馈

注：项目负责人每 2 周应组织一次项目组成员研讨会，请作好记录。

项目小组研讨与交流会议纪要

时间	10 月 14 日	地点	A 栋	记录人	张雷明
参加人员	张雷明、谢璧蔚、张健、张嘉豪				
会议主题	项目进度				

会议记录

项目进度反馈

注：项目负责人每 2 周应组织一次项目组成员研讨会，请作好记录。

项目小组研讨与交流会议纪要

时间	10 月 31 日	地点	A 栋	记录人	张雷明
参加人员	张雷明、谢璧蔚、张健、张嘉豪				
会议主题	项目进度				

会议记录

中期答辩前后相关任务安排

注：项目负责人每 2 周应组织一次项目组成员研讨会，请作好记录。