汕头大学大学生创新创业训练计划项目

中期检查表

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

院　　系：工学院

专业/年级：计算机科学与技术 2020 级

学　　号：2020611073

联系电话：18218805493

指导教师：许建龙、王珊珊

填表日期：2023/10/27

汕头大学教务处制

填表说明

一、《汕头大学大学生创新创业训练计划项目中期检查表》由项目负责人根据项目实施情况实事求是填写，并用 A4 纸双面打印，于左侧装订成册。

二、本表第三项以附录方式提交《汕头大学大学生创新创业训练计划项目进展日志》。

三、本表第五项以《汕头大学大学生创新创业训练计划项目经费使用历史记录》内容为准，由学院对内容进行审核。

四、本表第六项由指导教师核查填写。

五、填写本表内容统一使用宋体，小四号字，单倍行距。

<table>
<tr><td colspan="2">项目名称</td><td colspan="5">面向计算机程序设计竞赛的在线评测系统</td></tr>
<tr><td colspan="2">起止时间</td><td colspan="2">√2023 年 7 月　□20　年　月</td><td>项目类型</td><td colspan="2">√创新训练　□创业训练　□创业实践</td></tr>
<tr><td rowspan="2">负责人</td><td>姓名</td><td>所在学院</td><td>专业</td><td>入学年份</td><td>联系电话</td><td>E-mail</td></tr>
<tr><td>张健</td><td>工学院</td><td>计算机科学与技术</td><td>2020</td><td>18218805493</td><td>20jzhang@stu.edu.cn</td></tr>
<tr><td rowspan="3">参加成员</td><td>张嘉豪</td><td>工学院</td><td>计算机科学与技术</td><td>2020</td><td>15382857905</td><td>20jhzhang@stu.edu.cn</td></tr>
<tr><td>张雷明</td><td>工学院</td><td>数据科学与大数据技术</td><td>2020</td><td>17620663547</td><td>20lmzhang@stu.edu.cn</td></tr>
<tr><td>谢璧蔚</td><td>工学院</td><td>计算机科学与技术</td><td>2022</td><td>17703074697</td><td>22bwxie@stu.edu.cn</td></tr>
<tr><td rowspan="4">导师</td><td>姓名</td><td>王珊珊</td><td>学院</td><td>工学院</td><td>职务/职称</td><td>副教授</td></tr>
<tr><td>E-mail</td><td colspan="2">sswang@stu.edu.cn</td><td>联系电话</td><td colspan="2">15142813890</td></tr>
<tr><td>姓名</td><td>许建龙</td><td>学院</td><td>工学院</td><td>职务/职称</td><td>讲师</td></tr>
<tr><td>E-mail</td><td colspan="2">xujianlong@stu.edu.cn</td><td>联系电话</td><td colspan="2">13415177616</td></tr>
<tr><td colspan="7">一、项目研究进展情况</td></tr>
</table>

项目申报书预期成果/指标	项目已取得的成果/指标
（按照项目申报书内容填写） 1. 完成 DOJ 测试版的搭建工作，进入迭代开发。 2. 采集用户数据完善推荐系统以及熔断、降级机制。 3. 实现文件服务系统并完成部署。 4. 实现判题服务系统并完成部署。 5. 实现图床服务系统并完成部署。	1. 完成 DOJ 测试版的搭建工作，进入迭代开发。 2. 实现文件、判题、图床服务系统并完成部署。 3. 完成 atcoder、cf、hackerrank、hdu、poj、spoj、uoj、ural、uva、vijos 等题库的爬取和提交转发。 4. 完成题目批量上传时自动生成标签。 5. 完成推荐系统。

二、项目成员参与情况（附《项目小组研讨与交流会议纪要》）

成员贡献：

张健：完成数据设计、后端设计与代码编写、服务器搭建、服务部署。
张嘉豪：完成前端 UI 设计与代码编写。
张雷明：完成推荐系统设计与代码编写。
谢璧蔚：完成题目爬取、功能测试。

三、项目研究存在的主要问题及应对思路与措施（附《汕头大学大学生创新创业训练计划项目进展日志》）

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。

四、项目研究下阶段主要任务及时间进程安排

阶段二：(2022/11)

1. 完成 Beta 版本产出并部署。
2. 进行项目组内的服务测试，寻找潜在的漏洞。
3. 收集用户数据，用于完善压力测试以及推荐系统。
4. 收集用户反馈。

阶段三：(2022/12~2023/04)

1. 完成 RC 版本产出并部署。
2. 修复潜在漏洞。
3. 根据用户反馈对功能进行增修。
4. 进行功能测试以及压力测试。
5. 实现文件服务系统并部署。
6. 实现推荐系统并部署。

阶段四：(2023/04~2023/05)

1. Stable 版本产出并部署。
2. 网站公开。

五、**项目经费使用情况**（说明购置实验材料、试剂、药品、加工测试、资料、复印、调研、交通等已开支经费数额，以《汕头大学大学生创新创业项目经费使用历史记录》为准）

暂无

项目负责人签字	年　月　日

六、指导教师意见（学生开展项目情况，对已存在问题和下一阶段工作的意见、建议等）

导　师（签名）：

年　　月　　日

七、学院检查意见

☐ 通 过　　☐ 不 通 过（限 20　年　月　日前整改）

考核小组（签名）：

单位（盖章）：

年　　月　　日