**项目研究进展情况**

**一、小组成员分工**

**二、研究内容与方法**

我们想要对一个纸质文本进行文字识别转化成电子版以便于存储和识别，我们首先会将文本文献拍摄成为一份图片，然后通过机械视觉的方法将图片中的文字识别出来，由图片文件转化文本文件。

1、UI

我们目标是制成一个app，UI提供程序与外部的接口

2、DAO

利用python对外部传来的图像进行一定处理再传给底层识别

3、OCR底层

我们利用了Python开源代码Tesseract，对DAO层处理后的图片进行文字识别，

但是Tesseract主要针对英文识别设计，所以我们会针对中文进行机械学习，重新自定义中文语言库。

三、进度及阶段性成果展示

1、项目需求分析

2、GitHub存放语言库

3、

4、app的服务窗口UI设计完成

1）主界面

2）打开本地文件

3）识别界面

4）识别结果

5）

**项目研究存在的主要问题及应对措施（如项目的研究内容及项目团队成员有变动，也请在此栏说明）**

**应对措施：**

**项目研究下阶段主要任务及时间进程安排**