暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | Django开发 | | | | | | |
| 姓 名 | 黄华程 | 专业 | 人工智能 | 班级 | AI 2班 | 学号 | 09118211 |
| 实验时间 | 2020.8.31-2020.9.23 | | 指导教师 | 孔祥龙 | | 成绩 |  |
| **一、实验背景和目的**  2020年高考之后，我们力求做一个高考志愿填报系统，力求根据高考生所填入的分数，选出符合保一保，稳一稳，冲一冲的战略方针的高考志愿填报方案 | | | | | | | |
| **二、小组任务和个人任务**  小组任务：·完善后台管理系统，根据需求修改setting.py和url.py等配置文件  ·设计sqlite3数据库存入和读取的接口，从第一组获取数据，向其他组提供数据  个人任务：  ·创建登录界面  ·创建数据表格  ·制作majors类  ·进行数据修改  ·进行数据库第一，二，三，四次更新 | | | | | | | |
| **三、个人任务需求分析**  ·需求确定:首先我需要将数据表格可视化，并将第一组交给我的数据录入进我负责的majors类中  ·需求抽取：数据表格可视化相对简单，majors类的创建过程中需要不少外键的参与，录入的几十万条数据是否保持格式统一  ·需求协商和验证：跟第一组以及我负责对接的小组进行协商，了解第一组能给我提供什么格式的数据，以及我需要向其他小组提供什么数据，哪些冗余的数据可以省略，哪些数据不合要求需要重新清洗。遇到问题及时与小组负责人协商，讨论出解决方案 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**    def get(request):  df=pd.read\_csv('./extFiles/高校录取分数线整合（省份名字统一）.csv')  num=df.shape[0]  a=[]  for i in range(10000)：  c1=models.Majors(majorName=df.at[i,'Major'],  year=df.at[i,'Year'],  id=i,  categoryID=models.Category.objects.get(categoryname=df.at[i,'category']),  minScore=df.at[i,'score'],  collegeID=models.Colleges.objects.get(collegeName=df.at[i,'College']),  provinceID=models.Provinces.objects.get(provinceName= df.at[i,'Province']))  c1.save()  return HttpResponse("Finished!%s"%app\_name)  由于我负责的majors类中在数据库中包含其他实体的主键，因此在填入categoryID，collegeID，provinceID时需要使用models.Colleges.objects.get的格式，当然在导入数据时我们需要先将我们在majors类中使用到的数据表格全部导入到文件中，上述代码仅为所有文件中的一部分，majors类中所使用的部分实体由小组内其他同学提供 | | | | | | | |
| **五、实验结果与分析**  由于我负责的是majors这块的数据录入，因此实验是否成功就很显而易见了，在代码运行后，  所有录入的数据均准确无误，其中的省份学校id均正确对应，没有什么问题 | | | | | | | | |
| **六、实验总结与心得体会**  这是我整个人生中第一次参与较为大规模的团体合作，整个团队的人数达到了一百人，  在参与实验的过程中，我发现了很多问题也有很多的收获：  问题：  ·对于我而言，最显而易见的问题就是关于小组之间的接洽问题，我们组的工作需要第一组的数据来支持，但是由于第一组并不知道我们小组的具体需求，提供给我们的数据并不是统一格式的，比如有xlxs，csv，excel。当然这并不是很难解决的问题，但这种小的问题应该是可以克服的而不是交给我们这组头疼。  ·其次，是其他小组的工作质量问题，就比如第一小组提供的数据并没有清洗完毕，我们在制作数据库和录入数据时，发现经常报错，去第一小组提供的数据表格中寻找相关数据，我们发现就西藏这个地区的叫法就有n种，比如内地西藏，西藏内地，西藏，西藏自治区等，这导致我们在录入数据的时候会麻烦，更有甚者，地名之前还多一个空格键，这都是很难发现的错误，需要花不少时间来debug。我在最后运行的时候甚至发现college表格中的大学有不与分数整合表中相对应的  收获：  ·很显然在这种分组式的工程作业中，沟通和合作是很重要的，在面对一些我自己无法短时间解决的问题时，我都会积极的询问同组成员，问一下他们有没有类似的经验以及相对应的处理方法。在发现第一组提供的数据有问题时，我积极地与小组组长以及小组对接员沟通，让第一组进行了相对应的数据清洗和删改 | | | | | | | | |

2020年9月制