暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | 第10组：Django开发组 | | | | | | |
| 姓 名 | 李璟宸 | 专业 | 人工智能 | 班级 | AI 2班 | 学号 | 09118225 |
| 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**  放假在家，邻居孩子跑来问你，我这成绩能考上东南大学么？如果考不上东大，我能考上哪？本实验要求完成基于web的高考志愿推荐系统。  本项目分为模型、控制器和视图3个部分。模型构建从数据采集、数据清洗到数据融合，然后再构建知识图谱以及AI算法应用。在控制器方面，要求用构建的知识图谱实现知识推理；AI算法有聚类算法、预测模型和应用可视化等，设计好后要有内容推荐，根据高考分数推荐学习，根据个人兴趣推荐专业等。在视图方面在网站进行基础的UI设计，实现知识图谱可视化和AI应用可视化。 | | | | | | | |
| **二、小组任务和个人任务**  本组的任务是进行Django开发，要实现课程作业中的ORM操作，在Pycharm中创建一个名为ormDesign的APP，然后针对第一组的各校近三年的录取分数线和各省近三年高考的一分一段表设计出数据库，并提供各组查询的接口。各组只能查询数据库，不能增删和修改数据表内容。本组还帮助第一组整理了部分数据，对于格式不对或者数据不完善的做了清洗和整合。  个人任务包括初期的创建登录界面和数据表格，创建数据库Category查询方式，完善数据库并负责组内与第一组对接。在后期帮第一组整理数据时做了4次数据的清洗和数据库的更新。 | | | | | | | |
| **三、个人任务需求分析**  放假在家，邻居孩子跑来问你，我这成绩能考上东南大学么？如果考不上东大，我能考上哪？本实验要求完成基于web的高考志愿推荐系统，小组任务是根据高考分数线和一分一段表数据设计出供各组查询的数据库。个人任务是实现数据库的构建和完善以及帮助第一组整理数据。  这个问题分为方案级需求和问题级需求两方面。方案级需求是设计各组能查询但不能修改的数据库。项目总体采用的是Pycharm的ORM架构。数据库包含6个实体对象，分别是：专业、学科门类、大学、省份、分数排名和科类。  对应的数据库ER图如下：  图片3  图片4  在大学类中，大学ID为大学类的主键，省份ID为省份类的主键。在专业类中，大学ID为大学类的外键，省份ID为省份类的外键。其中，专业类的属性“一级学科IDs”表示该专业对应的一级学科，可以没有也可以有多个，所以这里的一级学科是一个文本数据，多个ID由特定字符分开。因此，这里的一级学科IDs与一级学科类中的一级学科ID不同，也不是外键。  其他组作为数据库的用户，能够看到的是各个数据库类的表格，并能够在表格里面根据索引和筛选到的进行查询，但是因为数据库后台和数据本身的权限问题，其他组在查看数据库的时候是不能进行改动，添加，删除的。  我个人主要负责的是科类的函数，包含科类ID和科类名称。要在db.sqlite3中建立Category表，包含categoryID和categoryname，要提供给各组能够查询科类的接口。  此外我还帮助第一组整理了4次数据，在一些表格格式不规范或者一分一段表是否包含近三年从一本线到最高分，分段是否正确，每个同学当时负责的省份和学校是否都有录入等。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  个人负责的任务主要包含：  0  在最开始创建的登录界面和数据表格如下图所示：  2  3  在对接方面，我和第一组对接，负责将各组使用数据库过程中遇到的问题收集后反馈给第一组，并与第一组协商解决问题。  个人负责的category类的创建如下：  **1**  在数据库遇到问题时，也对数据库内代码进行了修改和完善  **3**  在后续的几次数据处理方面，我主要是核查一分一段表和省份的分数线问题是否正确。  4  **5**  **6**  **7**  在数据遇到问题时候我会记录下来并联系负责该板块的同学改正  **8**  例如上图是某些大学在爬虫爬取时没有考虑要剔除艺术类的或者国家专项类，我都进行了手动记录和调整。  4  还有一些数据不太对应的，例如2017年云南的原来是5分一段表，我也将它修改为了一分一段表。 | | | | | | | |
| **五、实验结果与分析**  开始创建的登陆页面和数据库初始化页面如下图所示：  1  2  实现的category类如图所示  **2**  2017云南的一分一段表是5分一段的，修改为一分一段后如图所示：  **1** | | | | | | | | |
| **六、实验总结与心得体会**  本次实验我是参与了百人开发的一个项目：基于web的志愿填报推荐。并主要是在第十组：django开发组来参与项目，并负责后期的第一组的一些数据的清洗和处理。  本次实验我加深了对html的认识，学会了如何设计登录界面与html基本框架。在数据库构建方面，我学会了在pycharm中利用orm来构建数据库后台并为其他组提供支持。在数据清洗方面，学会了利用python的pandas和numpy进行一些表格内的数据查找和清洗处理操作。 | | | | | | | | |

2020年9月制