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 | 薛翔天 | 专业 | 人工智能 | 班级 | 091182 | 学号 | 09118215 |
| 实验时间 | 2020.8.31-2020.9.23 | | 指导教师 | 孔祥龙 | | 成绩 |  |
| 1. **实验背景和目的**   **背景：**高考完查到自己的成绩后，会有很多考生疑惑“我这成绩能考上XX大学吗？”“如果考不上，我能考哪儿？”。考生和家长对于学校录取的了解较少，而高考志愿填报往往决定了考生的未来，需要一个网站能够提供专业的评估和预测。  **目的：**本次课的课程设计是全体同学协作实现一个基于Python的高考志愿推荐网站。通过爬虫自动获取不同梯队的学校信息，构建院校、专业、各省分数的知识图谱，训练模型以实现分数同目标学校专业的适配度的精准评估。 | | | | | | | |
| **二、小组任务和个人任务**  **小组任务：**小组负责后端的Django开发工作。通过Django框架下的ORM操作来搭建一个包含大学、专业、排名等实体的数据库。将经过融合清洗的数据导入数据库，提供算法组使用数据库的接口，并永久负责数据库的更新和维护。  **个人任务：**负责合理分配组内任务，严格要求高效完成老师布置的小组任务。积极与第一组对接，负责数据实时的更新；积极与其他组对接，及时解决反馈的问题。 | | | | | | | |
| 1. **个人任务需求分析**   **功能性需求：**   1. 搭建包含大学、专业、排名等六个实体和多个属性的数据库。严格规定数据库的输入和输出，确保数据库内在逻辑关系正确，确保数据库的完整性和一致性。 2. 针对其余各组，安排1-2人对接，永久对数据库负责，确保数据的正确完整和算法组的正常使用。   **非功能性需求：**  1、组员任务难度和时长分配平均，实时记录督促完成情况。  2、这个系统在任何时候都必须是可用的，其他小组只能查询数据库，不能增、删和修改数据表内容。  3、确保数据库的可靠性、易使用性、可维护性、和可扩充性。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  **1、任务分配：**创建在线表格实时记录组员任务完成情况，绿色表示已完成，红色表示未完成。其中，包含2次集体任务，针对老师两次授课内容的掌握；其余任务均为搭建完善数据库的分任务。  截屏2020-10-02 下午3.57.29  **2、数据库ER图：**数据库包含六个实体对象，分别是“大学”、“专业”、“一级学科”、“省份”、“分数排名”、“科类”。大学类中包含六个属性，其中“大学ID”为主键，“省份ID”为省份类的外键。专业类中包含九个属性，其中“大学ID”为大学类的外键，“省份ID”为省份类的外键......以此类推。    **3、搭建数据库：**在/ormDesign/models.py中编写搭建数据库的代码并定义各个实体对应属性的数据类型。  **截屏2020-10-02 下午4.53.57**  **4、插入数据：**在/ormDesign/views.py中定义插入数据的函数，并在/ormDesign/urls.py中调用。  截屏2020-10-02 下午4.56.58**截屏2020-10-02 下午4.56.28**   1. **保存数据：**为方便其余各组查询数据库，在/db.sqlite3中存放插入的数据   **截屏2020-10-02 下午4.59.41**  **6、与第一组合作，负责数据的清洗和融合：**从大学专业和一分一段表两个方面全面检查数据，同时对其他组反映的问题积极尽快解决，不读更新完善数据库。  **截屏2020-10-02 下午5.04.40** | | | | | | | |
| 1. **实验结果与分析** 2. 数据库搭建完成，“专业”实体中共有170,000余条数据，“分数排名”实体中共有29,000余条数据。平均导入数据时间为200条/分钟。 3. 数据库第一次更新，针对majors实体，更新内容为：修改东南大学为西南交通大学；增填东南大学，南京农业大学，吉林大学，苏州大学数据。 4. 数据库第二次更新，针对majors实体，更新内容为：将所有非本一批次的大学数据删除，约10,000条数据。删除了下面三所大学的数据：贵州大学；东北农业大学；东北林业大学。 5. 数据库第三次更新，针对majors类，更新内容为：修正清华大学数据，新增南京理工大学、浙江大学等十四所大学数据。 6. 数据库第四次更新，针对majors类，更新内容为：修改复旦大学、华东师范大学、南京财经大学、西南大学数据。 7. 数据库第五次更新，针对rankings类，更新内容为：存放正确的2017-2019一分一段表，一分一段数据并不全，但能保证已有数据完全正确，具体情况参考检查表格。 | | | | | | | | |
| **六、实验总结与心得体会**  小组任务基于PyCharm开发平台，感受到Django框架下ORM操作的魅力：只需要面向对象编程, 不需要面向数据库编写代码；对数据库的操作都转化成对类属性和方法的操作；不用编写各种数据库的sql语句；实现了数据模型与数据库的解耦, 屏蔽了不同数据库操作上的差异；通过简单的配置就可以轻松更换数据库, 而不需要修改代码。  在实验过程中，遇到了各种各样意想不到的问题，通过查阅资料和询问老师，掌握了很多知识。我既然作为组长，那就要尽到组长的责任。我不能要求每一个组员都尽力完成，但我能严格要求自己必须尽心尽力。刚搭建完数据库的时候，觉得任务比较轻松，并没有想象中的重。但当别的组陆陆续续反馈出数据的问题的时候，我担心了。我和第一组组长沟通过，第一组组长给我的意见是，都快结束了现在改也来不及了。我知道问题在当初部分同学小作业的爬虫问题，但第一组的任务就是负责数据的清洗和融合，如果当时大家做的都没有问题，那第一组也不会有那么多人。对于我而言，我们组是数据的最后一道关，我一定要把住。我很感谢三十多位同学能够在短短几天的时间内完成全部的检查工作，数据库的五次更新离不开第一组组员和我们组的努力。我们组保证，会永远负责数据库的更新和维护。  感谢组员，有时候任务很急，我会要求组员在很短的时间内完成。没有一个组员有异议，大家聚在一起互相帮助互相吐槽，完成任务之后，虽然很累但笑得很开心。其实我们组在之后做了很多额外的工作，甚至是大多数数据清洗的工作，我向大家表示抱歉。我看到大家聚在一起为了共同的目标努力的时候真的很感动，感谢大家！  最后感谢老师，老师真的在整个任务过程中帮了我们很多，有时候弄到很晚发老师消息老师也一一回复。感谢孔老师，学到很多，后会有期！ | | | | | | | | |

2020年9月制