暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | Django | | | | | | |
| 姓 名 | 王行健 | 专业 | 人工智能 | 班级 | 2 | 学号 | 09118220 |
| 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**  本次软件实践的数据基于上学期同学们所爬取的近三年的大学专业录取分数和各省一分一段表，利用Django制作一个关于高考志愿填报的软件。 | | | | | | | |
| **二、小组任务和个人任务**  小组任务:1.安装Django，并制作表格  2.导入数据库数据  3.清洗数据  个人任务:1. 安装Django，并制作表格  2.导入数据库数据  3.清洗数据  4.重新爬取东北林业大学数据，作为第一次作业补交  5重新爬去山东2018文科一份一段表，作为第二次作业补交 | | | | | | | |
| **三、个人任务需求分析**  1. 安装Django配置环境，并制作表格等简单操作，熟悉Django的相关操作  2. 把其他组处理好的仓库数据导入到我们制作的表格中  3. 后期我们组全体加入第一组工作中去，进行数据清洗，筛查错误数据，补全确实数据，查漏补缺。  4. 数据清洗中，爬去了东北林业大学数据和山东2018文科一份一段表，作为前期两次作业的补交。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  1.安装Django,配置环境，制作简单的登录界面。  2.制作表格，导入数据。  3.检查序号95-99的大学数据是否正确或缺失。  4.负责爬取东北林业大学的数据。    5.检查五的一份一段表文件，并负责爬取山东2018文科一份一段表。 | | | | | | | |
| **五、实验结果与分析**  1.Django的使用过程并非一成不变，各自环境不同，操作过程也各不相同。  2.导入数据过程中发现数据本身的问题对导入存在较大影响，因而后期再次对数据进行了清洗。  3.爬取东北林业大学录取分数线数据过程发现，网页的多样性对于代码的要求还是较高的，对于Python基础不好的我过程还是有难度。  4.爬取山东2018文科一份一段表发现存在的问题主要还是我们代码的不严谨导致的，稍微处理一下就可以避免诸多问题。 | | | | | | | |
| **六、实验总结与心得体会**  1.初始数据错误较多，乱序，格式错误，数据错误很多，难以写出高准确率的代码进行清洗，进行了人工处理。工作难度低，但工作量很大。  2.小组的团队合作让我感受到了团队力量，因为我自己存在诸多问题，但是我的组员给予了我很多帮助，以致于我不会太拖后腿。  3.利用excel，记事本，python处理数据的能力得到提高，了解并学会一些了xlrd和json库的函数。  4.之前的Python课程并没有学好，导致了此次课程也存在着诸多问题，这也是我最大的问题所在。 | | | | | | | |

2020年9月制