暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | 数据采集和清洗小组 | | | | | | |
| 姓 名 | 钟倩如 | 专业 | 人工智能 | 班级 | 091182 | 学号 | 09118204 |
| 实验时间 | 2020.8.31-2020.9.27 | | 指导教师 | 孔祥龙 | | 成绩 |  |
| 1. **实验背景和目的**   2020年高考落幕，设计一个基于web的高考志愿推荐系统，让同学们能对自己的分数线可以达标的学校和专业有更多了解和选择。（原定于2020年7月开展实验项目，因为学校时间安排原因，9月才开始。） | | | | | | | |
| **二、小组任务和个人任务**  小组任务：至少收集近三年（2019、2018、2017）各所高校的录取分数线，并进行数据清洗，转化为统一的格式，剔除无用数据，向其他各组提供数据来源。  个人任务：听从组长调度 | | | | | | | |
| **三、个人任务需求分析**  第一周：读取图片形式的2019年安徽、北京、河北、湖北、江苏、山东、上海、四川、天津、浙江的一分一段表，并整理输出成json和excel文件  第二周：读取近三年（2019年、2018年、2017年）各省市（宁夏、新疆、青海、陕西、甘肃、云南、贵州、西藏、四川、辽宁、吉林、黑龙江）本科录取批次分数线，并转化成json文件  第三周：检查2019-2020-3学期收集的高校录取分数是否有缺失，并重新爬取了中国海洋大学、中国药科大学近三年（2019、2018、2017）的数据，并按格式输出  第四周：检查安徽、山东近三年（2019、2018、207）文理科数据是否有问题 | | | | | | | |
| 1. **实验过程（需附上关键代码及相关说明）**   1、生成一个关于随机省份及其id的excel     1. 将2019年安徽、北京、河北、湖北、江苏、山东、四川、天津的一分一段表输出成json文件（图以天津为例）      1. 将2019年上海、浙江的一分一段表输出成json文件（图以上海为例）     4、读取近三年（2019年、2018年、2017年）各省市（宁夏、新疆、青海、陕西、甘肃、云南、贵州、西藏、四川、辽宁、吉林、黑龙江）本科录取批次分数线，并转化成json文件（只展示部分代码）    5、重新爬取了中国海洋大学、中国药科大学近三年（2019、2018、2017）的录取分数线 | | | | | | | |
| **五、实验结果与分析**  1、生成一个关于随机省份及其id的excel（只展示部分结果）    2、将2019年安徽、北京、河北、湖北、江苏、山东、四川、天津的一分一段表输出成json文件（图以安徽为例且只展示部分结果）    3、将2019年上海、浙江的一分一段表输出成json文件（图以上海为例，且只展示部分）     1. 读取近三年（2019年、2018年、2017年）各省市（宁夏、新疆、青海、陕西、甘肃、云南、贵州、西藏、四川、辽宁、吉林、黑龙江）本科录取批次分数线，并转化成json文件（只展示部分结果）      1. 重新爬取了中国海洋大学、中国药科大学近三年（2019、2018、2017）的录取分数线（只展示部分结果） | | | | | | | |
| **六、实验总结与心得体会**  第一次全班一起做一个项目，且我们小组一开始没有进行具体的分析需求，以及我自己太过依赖组长分配任务而不是自己进行个人需求分析。如果以后再遇到多人合作的项目，希望能吸取经验，做得更好！ | | | | | | | |

2020年9月制