暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | 知识推理和图谱可视化 | | | | | | |
| 姓 名 | 赵基藤 | 专业 | 人工智能 | 班级 | 091182 | 学号 | 09118202 |
| 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**  背景：基于Web的高考志愿推荐系统。  目的：为高考结束的学子志愿填报提供建议。 | | | | | | | |
| **二、小组任务和个人任务**  小组任务：1、 依据高校分数线进行统计分析，生成相应图表，展示变化趋势；  2、根据用户输入的动态个人信息进行相应的可视化；  3、以地图为依据生成可视化界面，展示高校信息；  4、基于知识图谱的推理和问答；  5、对应的django框架编写。  个人任务：1.统计各省份985、211以及双一流大学的数量，用于地图显示；  2.根据高校录取分数线和各省份的一分一段表得到各高校在各省的录取位次，用于推断可申报的高校；  3.计算各高校近三年在不同省份的录取位次占该省高考人数的百分比，用户输入大学和文理科，在网页上展示图表；  4.分配和统筹组员的编写工作。 | | | | | | | |
| **三、个人任务需求分析**  任务1:  目标：统计各省份985、211、双一流大学的数量供地图组使用。  实现思路：根据学校类型的csv，进行统计并保存输出。  任务2:  目标：用户选择xx省xx名次能上xx学校。  实现思路：根据各高校录取分数表首先获得某年份某校在某省的最低录取分数，然后根据各省份的一分一段表，找到最低分数对应的名次，最后可根据输入的名次和省份查找并输出大学名称。  任务3:  目标：根据用户输入的学校和文理科信息，展示该高校在各省录取位次占高考人数的百分比，一定程度上体现了高校在不同省份的性价比。  实现思路：获得高校在各省录取的最低排名，然后除以各省当年高考的总人数，获得百分比。对比各省的录取百分比，得到最高性价比。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  任务2：  清理“高校录取分数线整合.csv”的数据    根据“高校录取分数线整合.csv”获得一个标签为“年份”“高校”“省份”“最低分”的Dataframe，即得到高校所有专业录取分数线的最小值，同时过滤掉过低的不符合实际的分数线。    根据一分一段表获取最低分数线对应的录取位次，需对文理科以及一分一段表的数据格式进行分类讨论：        任务3:  确定各省高考人数，获取用户输入的专业（文理科）和大学。  6BEA331731B885B603E61A0CA3F704FB  获取数据库中该学校对应的所有省份的三年最低录取分数，继而获取最低录取分数对应的排名，除以各省份历年高考的总人数，获得百分比，并用柱状图展示。  88DB23702FA99E8A4965FF30D60E7A3D  4AF1068B0EA83D6036FBF6AC6C4E94CA  BA1E616F3F72C7B131E7A74350FD3110 | | | | | | | |
| **五、实验结果与分析**  任务1:  结果节选：  屏幕快照 2020-09-25 下午3.55.40  分析：这里截取了部分省份的985、211以及双一流大学的数量。  任务2:  结果节选：    分析：此处截取了北京大学在各省录取位次的部分数据，由于部分省份一分一段表的缺失和不完整，导致有一小部分录取位次缺失。  任务3:  前端输入文理科和大学名称，得到相应的图表。  BF721F3431BE73E8DF3B0734E81CE9EE  4A563FD5DD110EF927044EF06369C77B  D5851385DAB58A7C3A1167FEE3224A84  95AE7A32D5A2C06A09679A5DAE8F01F5  分析：可以看到南京大学三年来在各省的招生比例虽然有不同但差距并不大。福建、北京两个省级行政区百分比特别高可能为艺体生分数线，除此之外南京大学在江苏省内招生占比较高，符合实际情况。有几个省份出现0的原因在于数据缺失。 | | | | | | | | |
| **六、实验总结与心得体会**  总结：在本次项目中我主要负责基础数据的统计分析工作。代码实现的难度不大，但由于原始数据出错率较高，我们进行了一定的修正工作。例如，在获取高校最低录取分数的时候删除了330分及以下的数据，确保了一定的真实性。除此之外，原本关于高校在不同省份高地洼地的分析，由于各省份各高校的录取政策多样且无法量化，最终决定以录取位次占各省高考人数百分比来展示给用户，比较直观。  与此同时，作为本组的组长，要负责协调好各部分负责同学的工作，例如前后端编写的统筹安排，确保每个同学都有任务。 | | | | | | | | |

2020年9月制