暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | 知识图谱构建B组 | | | | | | |
| 姓 名 | 赵琦 | 专业 | 人工智能 | 班级 | 091182 | 学号 | 09118239 |
| 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**  高考对我们来说是一场非常重要的考试，但重要的不仅仅是考试这一环。常言道，七分考，三分报，志愿填报同样是一项非常重要的工作。对于刚刚从高中毕业的同学来说，多数人对志愿填报的各种参考信息，填报策略都不是很了解。基于此，为了更好的帮助各位考生，故开发一个能够提供志愿填报参考的智能推荐系统。 | | | | | | | |
| **二、小组任务和个人任务**  1.小组任务  构建与省份信息相关的知识图谱，包括各省拥有的学校，各省近三年的分数信息。并实现对构建的知识图谱的可视化。  2.个人任务  部分后端开发任务。使用其他小组处理好的数据，根据小组成员设计的数据库模式，搭建Neo4j图数据库，管理与省份相关的信息。写一份配置文档。 | | | | | | | |
| **三、个人任务需求分析**  后端搭建Neo4j图数据库，首先需要学会在电脑上配置Neo4j图数据库，会打开关闭等基本操作，学会用python语言连接到数据库，并能对数据库进行修改。然后才能将整理的数据中与省份相关的部分按照设定的数据库模式添加进来。  设计数据库模式。与省份相关的信息有很多，省名、ID，拥有的院校、不同年份，不同  科类的一分一档表、不同年份的一本线、二本线等等。设计一种合理的数据库模式来管理  这些数据是非常有必要的。  需要掌握对json，csv等文件的读取方法，提取相关数据，用于创建图数据库的节点和  关系。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  1.在自己的电脑上配置Neo4j数据库  ①安装JDK（11版）  ②安装Neo4j数据库（community-4.1.1版）  2.学会数据库的基础操作，开启关闭数据库等，学会使用Neo4j的网页端  3.导入py2neo包，使用python语言对数据库进行修改  *from py2neo import Graph, Node, Relationship*  4.连接到数据库（保证数据库处于开启状态）  *graph = Graph("http://localhost:7474",*  *username="neo4j",*  *password="123456")*  5.创建节点、关系  *#以创建省份节点为例，逻辑基本通用*  *for i in range(provinceinfo.shape[0]):*  *provinceNode.append(Node(‘省份’,*  *ID=int(provinceinfo.iloc[i][0]), Name=str(provinceinfo.iloc[i][1])))*  *graph.create(provinceNode[i])*  6.打开Neo4j网页端查看更新后的数据库里管理的数据。  7.写配置文档。 | | | | | | | |
| **五、实验结果与分析**  更新后的数据库内总共有三类节点，省份，学校，分数信息。四种关系，省份拥有学校，学校位于某个省份，省份指向分数信息，以及分数信息指向学校。  省份节点包含属性：省份ID、省份名称。  学校节点包含属性：所在省份的ID、学校ID、学校名称、是否是985、是否是211、是否是双一流  分数信息节点包含属性：省份ID、分数（list）、累计人数（list）、年分、分数线类型（list）、分数线值（list）、科目（文理综） | | | | | | | | |
| **六、实验总结与心得体会**  1.实验总结  实现在自己电脑上配置一个Neo4j图数据库，并用python语言对数据库进行各种操作。也学  习了如何写配置文档。  2.心得体会  ①前期学习参考的博客文档要有所保留，方便后续回顾和写配置文档。  ②将整理好的数据存入数据库的整个流程中，理清楚各种数据间的关系，设计一个合适的数据库模式是非常关键的。  ③数据清洗工作对更新数据库有很大影响。不合格的数据格式会导致程序的各种报错。 | | | | | | | | |

2020年9月制