# 《数据科学导论》课程作业

**课程作业要求：**

选择一个特定领域（例如，医疗、电商、金融、娱乐等），利用数据科学导论课程学习的数据采集、数据清洗、数据存储、数据统计等技术，构建该领域的知识图谱，基于neo4j实现知识图谱的存储、查询、统计及可视化。根据知识图谱构建过程撰写报告，并录制视频展示知识图谱构建过程的关键环节。

**参考题目：**

|  |  |  |
| --- | --- | --- |
| **序号** | **名称** | **参考资源** |
| 1 | 医疗常识知识图谱 | ICD编码，常用临床医学名词，百度百科 |
| 2 | 中医药知识图谱 | WHO中医药术语国际标准、中国药典、中药材名录、医案 |
| 3 | 电商知识图谱 | Amazon数据集, 天池，淘宝等各类开源电商数据集 |
| 5 | 金融知识图谱 | 股票相关公开数据集、接口 |

以上题目仅做参考，可根据自己的兴趣选择其他领域构建知识图谱。

【推荐资源：】

阿里天池竞赛平台、Kaggle

**报告要求：**

1. 课程作业可以由小组或个人完成，小组人数不超5人。
2. 课程报告要求结构合理、逻辑清晰、图文并茂，展示知识图谱构建过程中的数据收集、清晰、知识图谱构建与展示等过程。
3. 每个小组需要制作一个5分钟左右的知识图谱视频介绍，介绍项目的创作以及最终的展示效果
4. 报告模板格式：报告模板格式不做要求，但是要求结构合理、逻辑清晰。
5. Latex模板： 建议使用山东大学本科毕业设计模板或者ACM、 IEEE等会议或者期刊模板，附件给了山东大学本科毕业设计模板（github）供大家参考。
6. Word模板： 建议使用本科毕业设计模板或者实验模板，附件给了山东大学学位论文模板
7. 禁止抄袭、剽窃等不端行为，一旦发现，本次作业将会被取消成绩。

**评分标准：**

1. **必备模块：数据收集、数据清洗、知识图谱查询以及可视化**
2. **评分规则：报告质量+知识图谱展示视频+知识图谱数据质量**
3. **评分准则： 1/3的小组为优秀小组**

**提交要求：**

1. 作业完成后，需要提交知识图谱源文件（实体列表、关系列表、三元组列表、实体-概念关系等），课程报告，视频。

2. 课程结束前将作业文件打包上传到山大云盘（推荐zip,tar格式）

3. 提交链接

https://icloud.qd.sdu.edu.cn:7777/#/link/361A5B35217F58C9EC9C2C092AA2B8D9

有效期限截止：2022-12-25

【注意】：作业需要在课程结束后提交，山大云盘会设置一个截止日期，截止日期之后提交的，会有相应的扣分。

3. 命名方式【请按照规范格式命名】：班级\_(个人或者小组负责人学号)\_（个人或者小组负责人姓名）.zip。