**一．需求分析**

利用已有的数据构建一个小型的报考知识图谱(知识库)，通过调用该图谱可以实现如下功能：

1.已知自己某分数能上什么学校

2.某个特定的专业哪个学校分数最高

3.已知自己的分数判断自己能学什么样的专业

4.查询某学校的特定专业

5.我只想学XX专业，能去什么学校？

1. **初始设计**

使用基于图数据库的**Neo4j生成此次的知识图谱**

**知识图谱设计：**

**实体**：学校，专业，分数，年份，对照表

**属性**：（专业）属于某学校的文/理/不分文理专业，（属于哪个专业的）对照表

**对照表属性设计**：开设该专业的所有学校，每年的录取分数

**三．小组任务**

**任务1：数据源**

本项目需要用到的数据源：是第一组清洗的包含学校，专业，省份，分数，年份的csv文件。

任务：从之前的小组获取报考省份及相应分数线的讯息

## **任务2：知识图谱设计与优化**

利用已有的数据构建一个小型的报考知识图谱(知识库)，通过调用该图 谱可以实现如下功能：

1.已知自己某分数能上什么学校

2.某个特定的专业哪个学校分数最高

3.已知自己的分数判断自己能学什么样的专业

4.查询某学校的特定专业

5.我只想学XX专业，能去什么学校？

## **任务3：知识图谱数据准备**

此任务包括：

1.对专业名称进行消歧处理

2.为有需要的实体生成标识符

## **任务4：创建可以导⼊Neo4j的csv文件**

在第一个任务里，我们已经分获取了了包含所有信息的csv，但这些文件不能直接导入到Neo4j数据库。所以需要做⼀些处理，并生成能够直接导入Neo4j的csv格式。 我们需要生成这⼏个文件：暨设计环节决定的实体及这些实体之间相互关系对应的csv文件。对于格式的要求，参考：<https://neo4j.com/docs/operations-manual/current/tutorial/import-tool/>

## **任务5：利用上面的csv文件生成数据库**

使用neo4j命令把所有的数据导入到Neo4j中，数据默认存在 graph.db 文件夹中。

重启Neo4j服务，通过localhost:7474观察知识图谱。

使用自带的命令进行简单查询的测试，如：

# 查询node

MATCH (n:Concept) RETURN n LIMIT 25# 查询relationship

MATCH p=()-[r:industry\_of]->() RETURN p LIMIT 100

## **任务6：基于构建好的知识图谱，构建显示网页**