

# 明代历史知识图谱

## 「知识图谱导论」课程作业

### 1 项目简介

本项目实现了基于明代历史人物及事件的知识图谱，让使用者沉浸在纵横开阔的历史人文知识海洋中。我们的「明代历史知识图谱」项目的关系图谱展示了关于条目主体（如[朱元璋](#)）的所有关联性信息，使用者可以通过某个主体的关系图谱，一步一步探索未知的关系可能性。

### 2 技术实现

#### 2.1 图谱结构设计

#### 2.2 数据采集

本组经过调研，选择了[全历史](#)及[百度百科](#)作为本项目的数据来源。

在数据采集方面，我们使用了[Python](#)语言中的[selenium](#)库及[urllib3](#)库爬取网页中的信息，并使用[BeautifulSoup](#)库及正则表达式进行网页信息解析，以期提取出我们需要的数据。一个可资参考的示例代码如下：

```
pass
```

在获取信息后……

我们利用[MySQL](#)数据库存储爬虫爬取得到并清洗后的信息，以供之后可视化及QA的使用。

#### 2.3 数据可视化

本项目基于[Neo4j](#)实现了项目数据的可视化。[Neo4j](#)是一个高性能的[NOSQL](#)图形数据库，它将结构化数据存储在网络上而不是表中。它将结构化数据存储在网络上(从数学角度叫做图)上而不是表中。[Neo4j](#)也提供了图数据库的可视化功能，可利用此引擎直接可视化的展示本数据库的图谱。

#### 2.4 问答系统

### 3 效果展示

#### 3.1 知识图谱可视化

## **3.2 问答系统**

# **4 项目结构**

本项目中的文件结构如下所示。

## **5 附录**

### **5.1 项目中实体间关系列表**