知识库语义搜索实践

1. **任务说明**

在给定的三国演义数据集基础上，构建一个知识库，以该知识库为基础，实现一个简单的语义搜索系统，能够根据用户的查询返回需要查询的实体列表。

1. **任务要求**
   1. **构建三国演义知识库**

在“data/数据集/”路径下，存放有三国演义数据集，本次实践需要在观察该数据集的基础上，构建一个三国演义本体，并依据构建的本体抽取出三元组。本体构建工具可选择[Protégé](https://protege.stanford.edu/)。

可参考已给出的三国演义本体文件和三元组，具体文件见“data/三国演义本体.xlsx”和“data/triple.example.nt”。

* 1. **实现简易语义搜索系统**

实现一个支持进行关键词查询的语义搜索程序，主要包括查询，排序及结果显示功能：

* 1. 查询：该系统输入的用户查询包括多个关键词和类的筛选条件，例如需要执行的查询为：“刘备 C:人物”，其中“刘备”为关键词，而“C:人物”为类筛选条件。语义搜索系统应当根据讲用户查询转换为SPARQL查询，从而搜索出类型为人物，并且在字面量取值种包含“刘备”的实体。
  2. 排序：该系统应对筛选出的实体结果集合进行排序，将对于用户而言更“重要”的实体排在更靠前的位置。较为简单的排序方法可以依据关键词在相应的三元组中取值部分的字符数量占比进行打分（例如，关键词 “刘备”在“刘备投奔荆州刘表礼遇甚厚”中的得分为2/12）。
  3. 结果显示：类似于网页搜索引擎，该系统应当为每个检索到的实体，显示实体的IRI、标签、命中的字面量取值及其属性信息。
  4. **示例查询**

实现的问答系统需要能够显示出data/querys.txt中给定的用户查询。

* 1. **示例程序代码说明**

示例程序包含一个简易的语义查询Web系统，数据库使用Virtuoso，Web使用Tomcat，使用Java实现。在代码文件src\cn\edu\nju\ws\SemanticSearch\servlet\QueryServlet.java中实现了包含一个关键词的查询功能，可以在此程序的基础上进行修改实现目标功能。

* 1. **其他**

可自行实现其它复杂的语义查询功能，例如，查询支持针对特定属性的筛选、拓展查询（Query Expansion）等。数据集也可自行寻找，不必拘泥与以上说明，将实现方法写成文档提交，助教会根据文档、代码及实验结果进行打分。

1. **提交说明**

将以下内容以zip格式打包，并命名为**学号+姓名.zip**上传到指定服务器：

* 1. **知识库：**基于三国演义数据集构建的知识库
  2. **说明文档：**说明文档中说明知识库构建方法和问答系统实现方法，以及实验结果截图。
  3. **程序代码：**可以运行的系统代码。

FTP服务器：

1. 地址：ftp://114.212.83.207:21/
2. 账号：SemanticWeb
3. 密码：SemanticWeb