基于知识库的问答实践

**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，实现一个简单的问答系统，能够回答一些基本问题。

**2、任务要求**

**2.1 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.example.xlsx”和“data/triple.example.nt”。(注意：以上本体文件和三元组文件仅作为示例，最后的提交结果需要以规范格式提交，如N-Triples、RDF、OWL等格式)。

**2.2 简易问答系统实现**

**2.2.1、基于模板的问答**

本次基于知识库的问答实践给出了基于模板问答的示例程序，可参照该示例程序，自行完成一个基于模板的问答系统。先将示例程序主要思路介绍如下：

对于给定的一个问句，第一步，槽识别：识别出问句中出现的实体槽及属性槽；第二步，模板选择：基于识别出来的槽+问句中的关键词，找到相匹配的模板；第三部，知识库查询，根据匹配的模板，生成相应的SPARQL查询语句，在知识库中进行查询（可使用Virtuoso储存知识库，安装教程可查看[地址](http://ws.nju.edu.cn/blog/2017/03/virtuoso%E5%AE%89%E8%A3%85%E5%92%8C%E5%AF%BC%E5%85%A5%E6%95%B0%E6%8D%AE/)）。

**1.1示例程序模板文件说明**

在“KBQA\_TC/src/main/resource/data/QA/template\_example.json”中给出了两个问答模板，每个模板由九个字段构成，“name”表示模板的名字，“slotList”表示该模板相关的槽列表，“Triggers”表示该模板相关的关键词/触发词，用于模板打分，“regex”表示该模板匹配的正则表达式，“sparql”表示该模板相关的SPARQL语句，“param”表示SPARQL语句返回的参数，“reply”表示该模板生成的回复信息。

**1.2 示例程序代码文件说明**

“edu.nju.ws.kbqa.Talk”是KBQA程序入口。

“edu.nju.ws.kbqa.SentenceRec”是用于问句实体/属性槽识别。

“edu.nju.ws.kbqa.TemplateManage”用于根据问句中识别的关键词和识别的槽选择合适的模板。

“edu.nju.ws.kbqa.ServiceCall”用于往选择的模板中填槽值。

“edu.nju.ws.kbqa.Loadsource”初始化加载资源。

**1.3 基本任务要求说明**

实现的问答系统需要能够回答出data/questions.txt中给定的问题。

**1.4示例程序下载地址**

链接：https://pan.baidu.com/s/1FfK8\_yU4zgh5jmgB\_EUcsQ

提取码：6yqh

**2.2.2、其他**

CMD可自行实现其它类型的问答系统，例如基于深度学习的端到端问答系统，数据集也可自行寻找，不必拘泥与以上说明，将实现思路写成文档提交，助教会根据文档及代码进行打分。

**3、提交说明**

需要提交构建的**知识库**+**说明文档**+**问答代码**，说明文档中说明知识库构建思路和问答系统实现思路，打包发送到以下地址，压缩包命名为**学号+姓名**。

提交地址如下：ftp://114.212.83.207:21/

用户名：SemanticWeb

密码：SemanticWeb