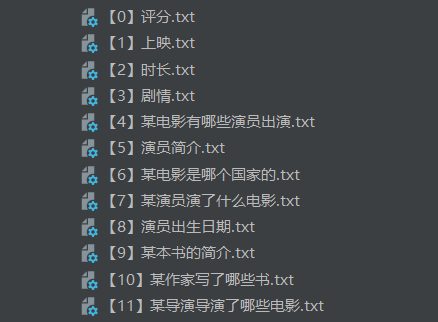
参考：https://github.com/chizhu/KGQA\_HLM

1、导入rdf数据

2、设计问题模板



每个模板下有若干相关问题，其中需要被讲识别出来的词性替换为nm（电影），nnt（人名），nb（书籍）等

3、首先对于问题分词建立词汇库



分类模型采用朴素贝叶斯生成分类器：



更新问题模板时需要删除model下的全部文件，再次运行train.py

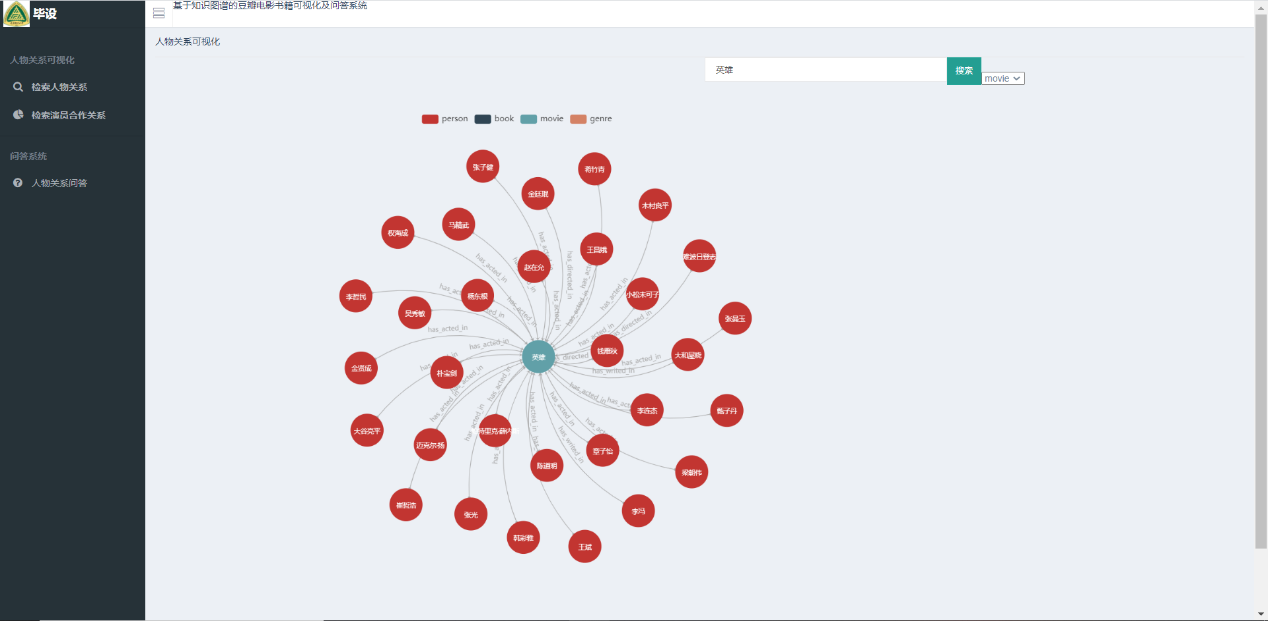
采用Jieba分词来做分词和词性识别

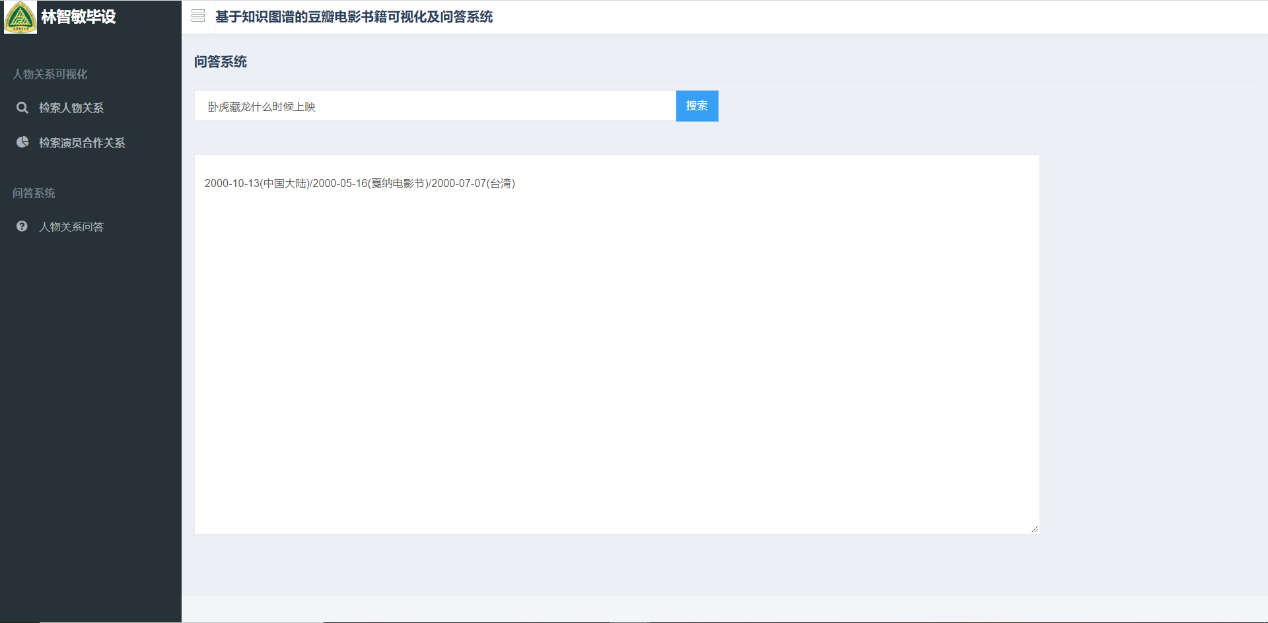
不用ltp的原因：ltp没法自定义词性，词性识别的效果不好

Ltp 底层是c++写的效率上快于jieba，后面可以研究一下ltp的源码

4、搜索模式

分为基于关系网络的搜索和基于问答的搜索





5、设计cypher语句并查询neo4j

需要修改一下neodb.config.py里的配置

