**知识图谱问答系统设计**

疫情知识图谱是专有领域知识图谱，所以问答系统是基于模板的。

用户在网页上输入一个问题，需要先对问题进行分类，然后根据四类模板完成问题解析，提取出关键词，自动生成查询语句，完成查询。

从用户角度考虑，用户可能对下面几类问题感兴趣：

1. 某地、某国的疫情情况；或者某国、某机构的防控政策、疫情事件等；
2. 某地、某国、某机构、某人的介绍；
3. 新冠肺炎相关的知识，包括来源、如何预防、如何治疗、药物、疫苗等；

模板以外的问题先关闭。

|  |  |  |  |
| --- | --- | --- | --- |
| **问题分类** | **模板** | **例子** | **查询方法** |
| 实体疫情 | [国家、地区、省市、机构]的[疫情、防疫、确诊、死亡、治愈] | 广元的疫情情况如何？  美国疫情情况如何？  四川有哪些防控措施？  四川信息职业技术学院有哪些防控措施？  四川信息职业技术学院的防疫政策？  注：暂时不考虑历史情况，只查询最新疫情数据。 | 数据来源：微博、官网、openkg；  实体：地名、机构、  关系：疫情、防疫、确诊人数、死亡人数、治愈人数  查询语句：  1、疫情、防疫措施从neo4j查询：match (n:term{name: ’xxx’} )-[r:rel{relation:’防疫措施’}]-()  2、人数从表格直接查询，不通过neo4j |
| 实体情况 | [国家、地区、省市、机构、人]的[简介、情况、介绍] | 钟南山是谁？  美国是什么？  尼日利亚是哪里？ | 数据来源：百科、openkg  实体：地名、人物、机构  关系：是、哪里  查询语句：  Match (n:term{name:’xxx’}) return n  在查询结果中可能再搜索属性名 |
| 肺炎知识 | [新冠病毒、新冠肺炎、COVID-19肺炎、新型冠状病毒、冠装病毒肺炎]的[来源、症状、预防、药物、治疗、防护、疫苗] | 新冠肺炎来源？  新冠肺炎是什么？  怎么预防新型冠状肺炎？  哪些药物可以治新冠肺炎？  新冠肺炎有特效药吗？  新冠肺炎有什么药？  新冠病毒有疫苗吗？ | 数据来源：百科、openkg  实体：疾病、病毒、药物  关系：来源、属性、预防、治疗、  查询语句：  Match (n:term{name:’新冠肺炎’})-[r:rel] where r.relation=’药物’ … |
| 未知 | 其他问题 | 今天星期几？ | 对不起，我不懂这个问题。 |
|  |  |  |  |

1、问题文本分类

分词、词性、句法

2、查询获得答案列表

3、网页显示答案列表

近义词查询

实体、关系名单中找到对应关键词

根据实体、关系、问句中的动词，完成问题分类

文本查询各国/地区确诊人数

实体、关系查询neo4j

pyltp词性、句法标注

seg\_user\_dict.txt

pos\_user\_dict.txt

nodes.txt

relations.txt

synonyms.txt

cilin.py 词林

parse\_question ()

1、先根据句法和规则分类

2、后续用深度学习分类（Attention+LSTM）

子线程，每隔半小时从网页查询最新确诊人数数据，更新到covid\_patient.txt

自动生成cypher语句，完成查询，并解析查询结果，形成答案列表。

现有数据，后续都保存到nCov\_KG\data目录下：

|  |  |  |
| --- | --- | --- |
| **数据** | **文件名** | **来源** |
| 百科信息 | csv\baike.csv | 爬虫：互动百科关键词查询  www.baike.com |
| 实体数据 | csv\nodes.csv | 和neo4j数据库保持一致 |
| 关系数据 | csv\relations.csv | 和neo4j数据库保持一致 |
| 确诊人数 | csv\covid\_patient.csv | 爬虫每小时自动更新 |
| 实体名 | txt\nodes.txt | 列出实体名 |
| 关系名 | txt\relations.txt | 列出关系名 |
| 同义词典 | txt\synonyms.txt | 同义词字典 |
|  |  |  |

从网站爬取实时疫情数据，样例如下：

<https://blog.csdn.net/weixin_43130164/article/details/104113559?depth_1-utm_source=distribute.pc_relevant.none-task&utm_source=distribute.pc_relevant.none-task>

kgqa\cilin.py 词林类CiLin，实现查询同义词get\_synonym()，查询两个词相似度get\_sim2018()

query\_graph.py KGQA\_Parser类，实现对问题文本的解析，分类，获取关键词

|  |  |  |  |
| --- | --- | --- | --- |
| 分工 | 说明 | 实现方法 | 责任人 |
| 问题分类 | 按照模板和正则表达式分类  Source\kgqa目录 | 1. pyltp分词，词性标注，句法标注； 2. 在 | 江勇、杨源斌 |
|  | 深度学习方法完成问题分类 |  | 叶茂 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |