

可视化传染病知识图谱及其应用项目介绍

Author: 陈屹峰, 修曾琪, 和隆

2022年秋北航人工智能研究院《知识图谱》课程大作业 202211222

有问题请联系组长 20373860@buaa.edu.cn

可视化传染病知识图谱及其应用项目介绍

项目介绍

环境配置

项目运行

构建知识图谱

启动问答系统

可视化系统部署

小组分工

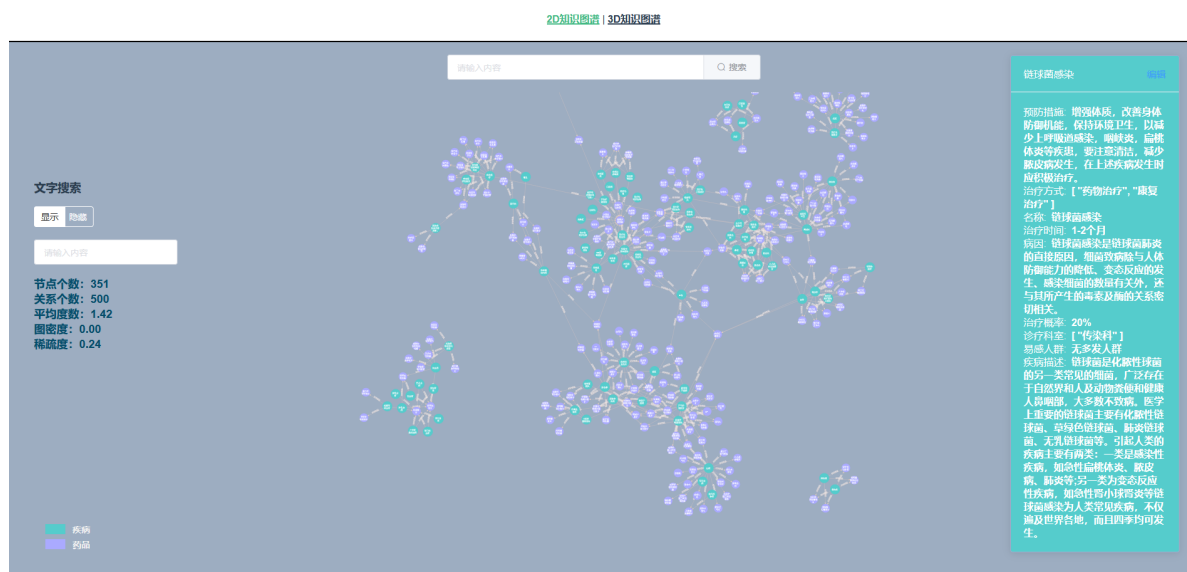
文件列表

项目介绍

通过爬取互联网中传染病的知识, 存储为格式化数据, 将其导入到neo4j图数据库中, 利用jieba与fasttext工具搭建知识问答系统, 生成字典树, 问题标签类型和查询语句样本并进行训练, 能够支持15类问题的智能问答。

利用d3, vue等框架搭建可视化页面, 使用户能够与知识图谱进行交互。

点击此进行体验: [传染病知识图谱](#)



环境配置

- python3.6
- jdk11(java11)
- neo4j-community-4.4.15 下载链接: <https://neo4j.com/download-center/#community>
- jieba

- fasttext

项目运行

原始爬取数据存储在 `./data/newmedical.json`

构建知识图谱

- 终端运行命令，启动neo4j数据库

```
neo4j.bat console
```

- 运行 `build_medicalgraph.py`，将数据导入neo4j图数据库，并生成实体特征分词，存储在 `./data_for_trans_diseases`

启动问答系统

- 运行 `data_producer_2.py`，加载实体特征词和问句指示词，生成文本分类训练数据，并将每类问题打上标签，生成训练文件 `training_data_for_trans_diseases.txt`
- 若非首次运行，直接运行 `chatbot_graph_test.py`，训练模型，并启动问答系统。

问答系统共支持15类问题：

```
# 由症状找疾病 # 由疾病找症状 # 由疾病找原因 # 由疾病找并发症 # 由疾病找不宜食食物 # 由疾病找宜食食物 # 由疾病找药物 # 由药物找疾病 # 由疾病找检查 # 由检查找疾病 # 由疾病找预防 # 持续时间 # 治疗方式 # 治愈可能性 # 易感人群
```

问答系统运行实例：

```
Building prefix dict from the default dictionary ...
Loading model from cache C:\Users\10140\AppData\Local\Temp\jieba.cache
模型已训练完成！
Loading model cost 0.555 seconds.
Prefix dict has been built successfully.
医疗字典已录入！
用户：得了登革出血热有什么症状
disease_symptom
小勇：登革出血热的症状包括：肝大；昏迷；昏睡；发烧；休克；咯血；淋巴结肿大；医师；出血倾向
用户：患上痢疾应该吃什么药
disease_drug
小勇：痢疾通常的使用的药品包括：盐酸左氧氟沙星片；盐酸左氧氟沙星胶囊；穿心莲内酯片；诺氟沙星片；乳酸左氧氟沙星片
用户：应该去哪个科室检查痢疾
disease_check
小勇：痢疾通常可以通过以下方式检查出来：纤维结肠镜检查；胸部B超；粪便脓液；钼靶X线检查；小肠镜检查；痢疾杆菌检测；粪便显微镜检查
用户：患了血吸虫病推荐什么食谱
disease_do_food
小勇：血吸虫病宜食的食物包括有：南瓜子仁；腰果；芝麻；松子仁
推荐食谱包括有：豆浆南瓜汤；扁豆糕；白扁豆参米粥；薏米扁豆老黄瓜汤；苋菜豆腐汤；豆腐苋菜羹；绿豆杂面条；玉米粉燕麦粥
```

可视化系统部署

具体流程见 `./visualize-graph/README.md`

小组分工

陈屹峰：知识问答、项目报告、PPT制作

和隆：知识可视化、系统部署、PPT制作

修曾琪：知识抽取、图谱构建、PPT制作

文件列表

```
data --- 原始爬取数据

data_for_trans_diseases --- 处理爬取数据生成的实体名单

spider --- 爬虫

visualize-graph --- 可视化

answer_search.py --- 查询程序

build_medicalgraph.py 建立知识图谱程序

chatbot_graph_test.py --- 问答主程序

data_producer_2.py --- 问答训练数据生成程序

dict_for_trans_diseases.txt 实体总表

entity_detection_test.py --- 实体检测程序

question_classifier_test.py --- 问题分类程序

README.md --- readme

training_data_for_trans_diseases.txt --- 问答训练数据
```