MKG项目

需求文档

部门：二组

撰写：二组全体

文档名称：MKG(医疗知识图谱)项目\_

版本号：1。0。0\_需求文档

文档修订记录

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本号 | 日期 | 修订页/修订描述 | 作者 | 审批人 |
|  |  |  |  |  |
|  |  |  |  |  |

1. 产品概述

1.1产品定位

在线医疗是当前互联网与医疗行业结合所产生的新业态，其研究也逐渐得到管理学界的广泛关注。利用知识图谱技术将各种医疗信息系统中琐碎、零散的知识相互连接，以支持综合性知识检索以及问答、辅助决策支持等智能医疗应用。

方法：在构建医疗领域本体和对医疗知识库进行语义标注的基础上，以39健康网（http://www.39.net/）为基础，找到最常见的疾病及与疾病相关的其他实体，构建医疗知识图谱并将其应用于智能医疗中。过程：提出了医疗知识图谱的构建方法，利用命名实体识别技术，对客户输入的关键信息进行查询，给与回应。结论:基于知识图谱的智能医疗应用将给用户提供最准确和最需要的医疗知识，在解决国内优质医疗资源供给不足和医疗服务需求持续增加的矛盾中产生重要的作用。

1.2产品范围及阶段规划

本产品主要面对常见病，对用户提出的与常见疾病相关的问题给予比较准确的回答。

第一阶段将会建立整体的框架，初步实现疾病查询、症状查询、疾病所需药品、疾病对应相关检查、疾病对应的相关科室的查询，用Django框架实现前台的显示

第二阶段将会与智能机器人对接，实现医疗的智能问答

1.3词汇表

**知识图谱实体类型**

| **实体类型** | **中文含义** | **举例** |
| --- | --- | --- |
| Department | 科室 | 消化内科;呼吸内科 |
| Disease | 疾病 | 血栓闭塞性脉管炎;胸降主动脉动脉瘤 |
| Symptom | 疾病症状 | 乳腺组织肥厚;脑实质深部出血 |
| Drug | 药品 | 京万红痔疮膏;布林佐胺滴眼液 |
| Check | 诊断检查项目 | 支气管造影;关节镜检查 |
| Operation | 手术 | [激光心肌血管重建](http://jbk.39.net/shoushu/jgxjxgzjs/);[血液透析用血管插](http://jbk.39.net/shoushu/xytxyzxgcg/) |

**知识图谱属性类型**

**疾病（Disease）**

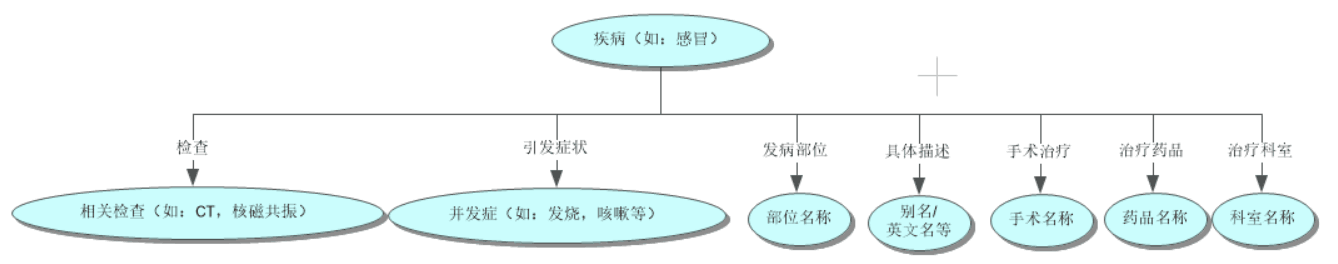
| **属性类型** | **中文含义** | **举例** |
| --- | --- | --- |
| name | 疾病名称 | 喘息样支气管炎 |
| desc | 疾病简介 | 又称哮喘性支气管炎... |
| cause | 疾病病因 | 常见的有合胞病毒等... |
| prevent | 预防措施 | 注意家族与患儿自身过敏史... |
| cure\_way | 治疗方式 | "药物治疗","支持性治疗" |
| easy\_get | 疾病易感人群 | 无特定的人群 |

**知识图谱关系类型**

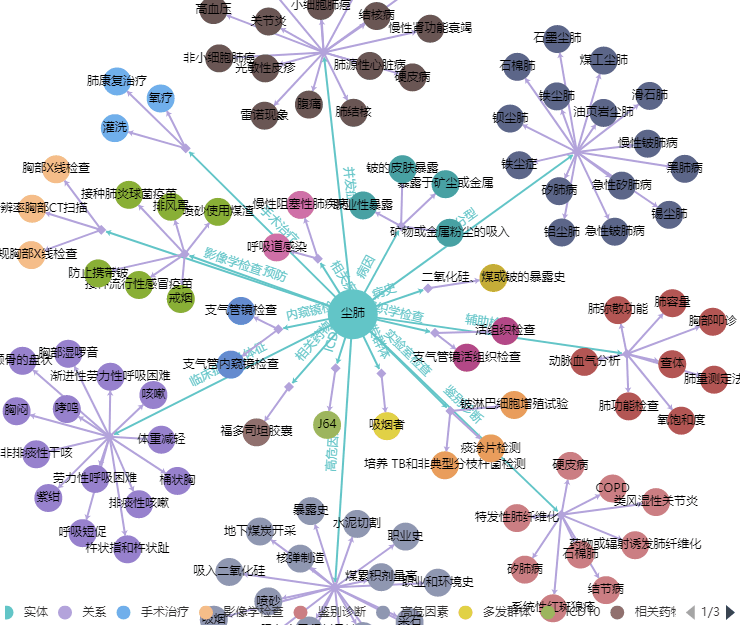
| **属性类型** | **中文含义** | **举例** |
| --- | --- | --- |
| belongs\_to | 属于 | <妇科,属于,妇产科> |
| common\_drug | 疾病常用药品 | <阳强,常用,甲磺酸酚妥拉明分散片> |
| need\_check | 疾病所需检查 | <单侧肺气肿,所需检查,支气管造影> |
| recommend\_surgery | 疾病推荐手术 |  |
| has\_symptom | 疾病症状 | <早期乳腺癌,疾病症状,乳腺组织肥厚> |
| acompany\_ symptom | 疾病并发症 | <下肢交通静脉瓣膜关闭不全,并发疾病,血栓闭塞性脉管炎> |

1. 整体思路

第一步，基于scrapy爬虫框架获取的json 格式文件，从中提取图谱的实体，进过数据预处理，目前爬取了7类实体，其中以疾病Disease为中心，可以将获取得到实体构建的关系，分别组成三元组；每一类疾病，都有自己的属性，方便我们查询。利用



第二步，将提取出的实体和三元组关系导入Neo4j 图形数据库中，方便前台调用



第三步，通过Django框架建立前台显示页面，通过前后台交互实现医疗图谱的功能

3．技术选型及项目的主要步骤：

1.1利用scrapy爬虫框架进行数据的的爬取

1.2 基于scrapy爬虫框架获取的json 格式文件，从中提取图谱的实体和关系

1.3 将数据导入到Neo4j图形数据库中

1.4 用idcnn+crf模型进行命名实体识别

1.5 用Django框架实现前台与后台的交互

4. 项目计划

|  |  |  |
| --- | --- | --- |
| 任务 | 时间 | 负责人 |
| 数据的爬取 | 21日数据爬取完毕 | 朱尧枫 |
| 实体提取和标注 | 23日提取完毕 | 焦芝琳 |
| 关系的提取 | 23日提取完毕 | 董皓 |
| Neo4j数据库的导入 | 26日导入完毕 | 左天赋 |
| 命名实体识别模型 | 24日模型基本训练完毕 | 赵冠智 |
| Django框架 | 27日框架搭建完毕 | 高菲 |
| 文档的填写，机动 | 项目期间文档的填写，各分项的代码跟进 | 张振锋 |