案例分析

基于医疗知识图谱的问答机器人

**一、引言**

大家都知道，人类历史发展到现在，已经经历了三次工业革命，每次工业革命都会带来翻天覆地的变化，生产力的进步带来了整个社会的变革，生活方式随之发生深刻变化。第一次工业革命使人类走向机械化时代，第二次是带来了电力，第三次是信息化革命。我们很幸运，今天处在第四次工业革命的开端，第四次工业革命的核心驱动力就是人工智能。当然，人工智能是为了模拟人的能力，需要包括语音、视频、图像、AR/VR 等感知方面的技术，也需要知识图谱、语言理解等认知方面的技术。当然，还有机器学习，以及最近这些年很重要的深度学习等等。

过去这些年，AI 技术通过深度学习取得了非常好的效果，尤其是在语音、视觉等感知技术上取得了非常大的突破，甚至在很多方面，深度学习达到的效果已经超过了人类。深度学习也给自然语言处理以及知识相关的技术带来了非常大的帮助，但是我们继续深入研究、应用实践的时候会发现，我们还需要更好地结合知识、推进知识图谱相关的工作，才能取得更好的效果。

基本的知识图谱，会涉及到实体的属性关系，每一个实体可能有若干个属性，实体和实体之间有很多关系，每一个关系基本上可以理解为是一个事实。因此知识图谱就是对客观世界的描述。本次案例的目的在于理解知识图谱，学习知识图谱构建。用途在于推进、结合知识图谱工作来加强人工智能的技术。

**二、案例内容**

**三、案例解决方案与分析**

知识抽取：

通过爬虫（lxml-etree、xpath）来提取页面中相关的信息，进行持久化

将持久化的数据进行整理与格式化，得到有用的数据

图谱构建：

如何去构建图谱，各实体间的关系怎么去确定？都是影响图谱执行效率的因素。在这里，实体设计与数目如下图所示：



图1 医疗知识图谱实体

实体关系类型的设计与数目如下：



图2 医疗知识图谱实体关系

属性类型的设计与数目如下：

图3 医疗知识图谱属性

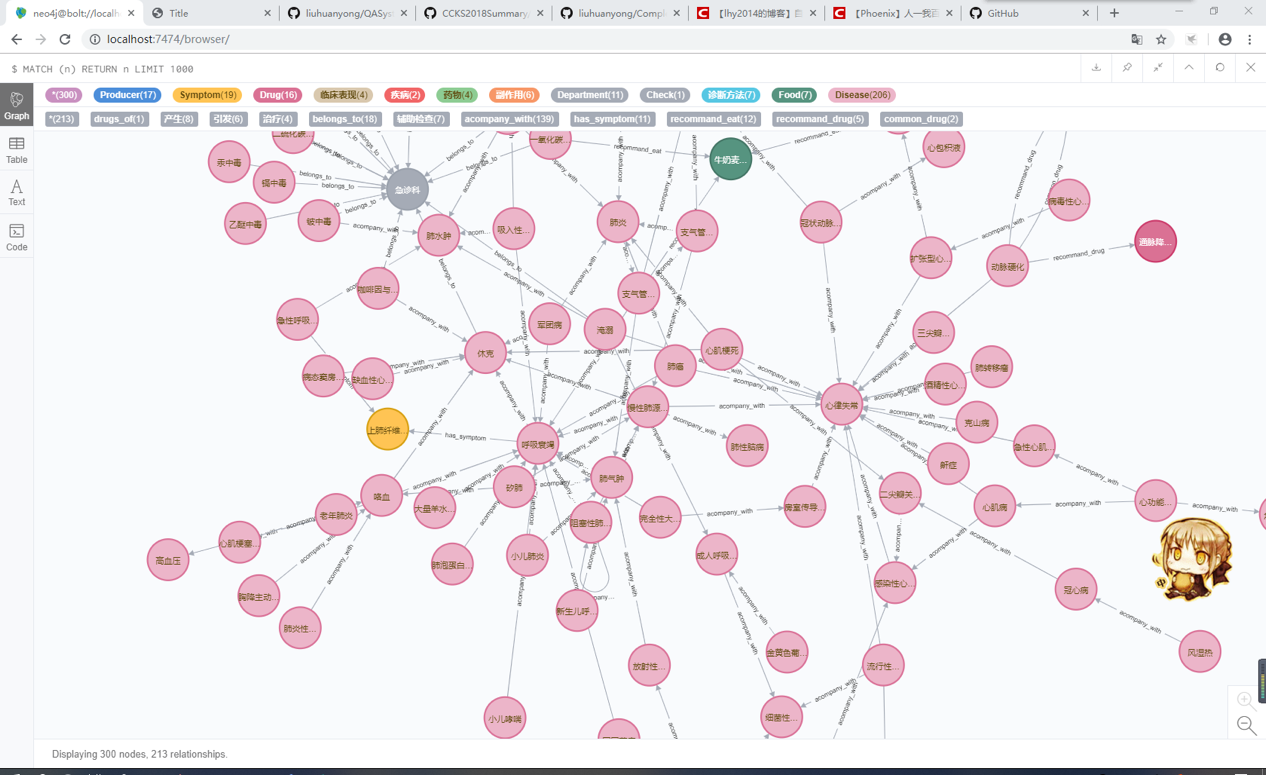


图4 本地搭建的依赖图谱

基于规则匹配实现问答：

**四、启发**

* 通过复现一个基于知识图谱的医疗智能问答程序，我们可以大致的清楚所提及的知识图谱是什么东西。按照我个人的理解，他实际上是一个以图的结构进行存储的数据库，节点与边权各自有不同的含义，但是他比起其他最具有优势的地方，在于他的推理能力
* 以图作为存储结构，比起常规的SQL数据库，没有行列的限制；比起No-SQL数据库，图数据库突出了存储的数据间的关系。我们后续的一些操作，都是基于这样的图数据库来进行实现
* 我们可以依据实体、关系、属性，实现基于水务设备的物联网图谱服务，目前打算的是基于oneM2M推荐的服务方法，实现中间件的服务，后期会计划在此基础之上实现基于推理的设备智能控制