案例分析

基于医疗知识图谱的问答机器人

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

基于规则匹配实现问答：

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