**需求介绍**

**在知识库里面的疾病的描述中抽取三元组 然后在进行推理**

**做这个辅助诊断 根据症状 去 推出疾病是啥 最终目标是有症状之后出来具体的病**

**老板说：这个解决的方式有两种 一个是基于三元组 在去推理找到这个疾病，这个时候会存在多对多的关系**

**第二个就是训练三元组模型去看疾病的里面一些描述要么存到知识库里面去 要么就直接通过模型去输出这个病 我把描述病的一段话直接交给模型 模型 根据 原来训练的方法能找到 出来什么病（关系抽取模型泛化能力 看到这段话能知道 疾病和症状的关系 以及症状和疾病的关系）**

**只要能做到这个事情就行 不管用啥方式**

**我目前做的事情：**

这一步是做的是 抽取这句话的三元组 “肺泡蛋白质沉积症(简称PAP)，又称Rosen-Castle-man-Liebow综合征，是一种罕见疾病。该病以肺泡和细支气管腔内充满PAS染色阳性，来自肺的富磷脂蛋白质物质为其特征，好发于青中年，男性发病约3倍于女性”

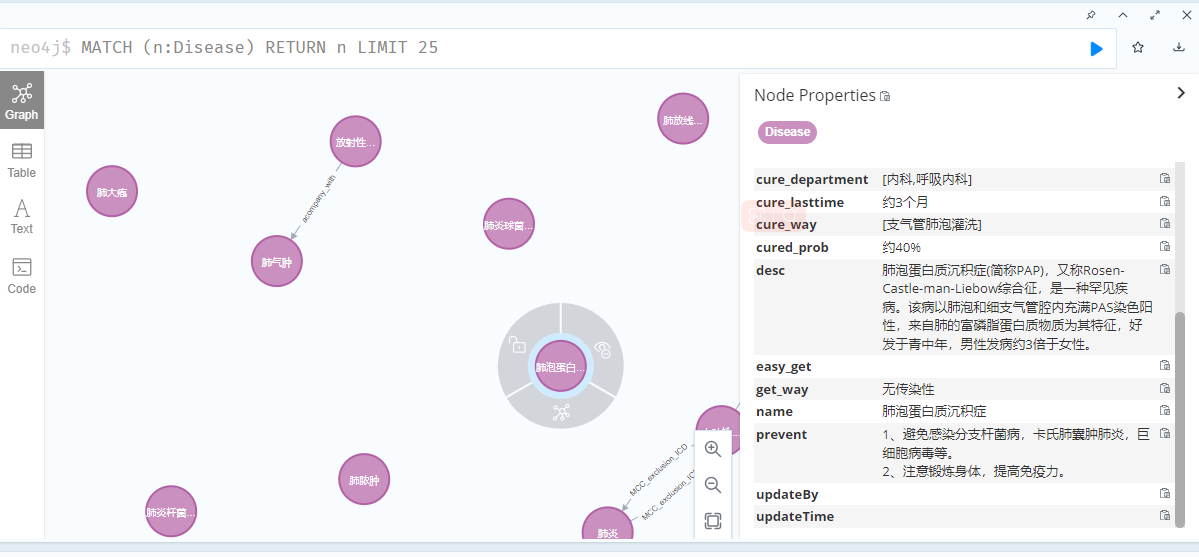
疾病 具有 症状

具有是关系

第二步是推理三元组 疾病 推理症状，症状反过来推理疾病

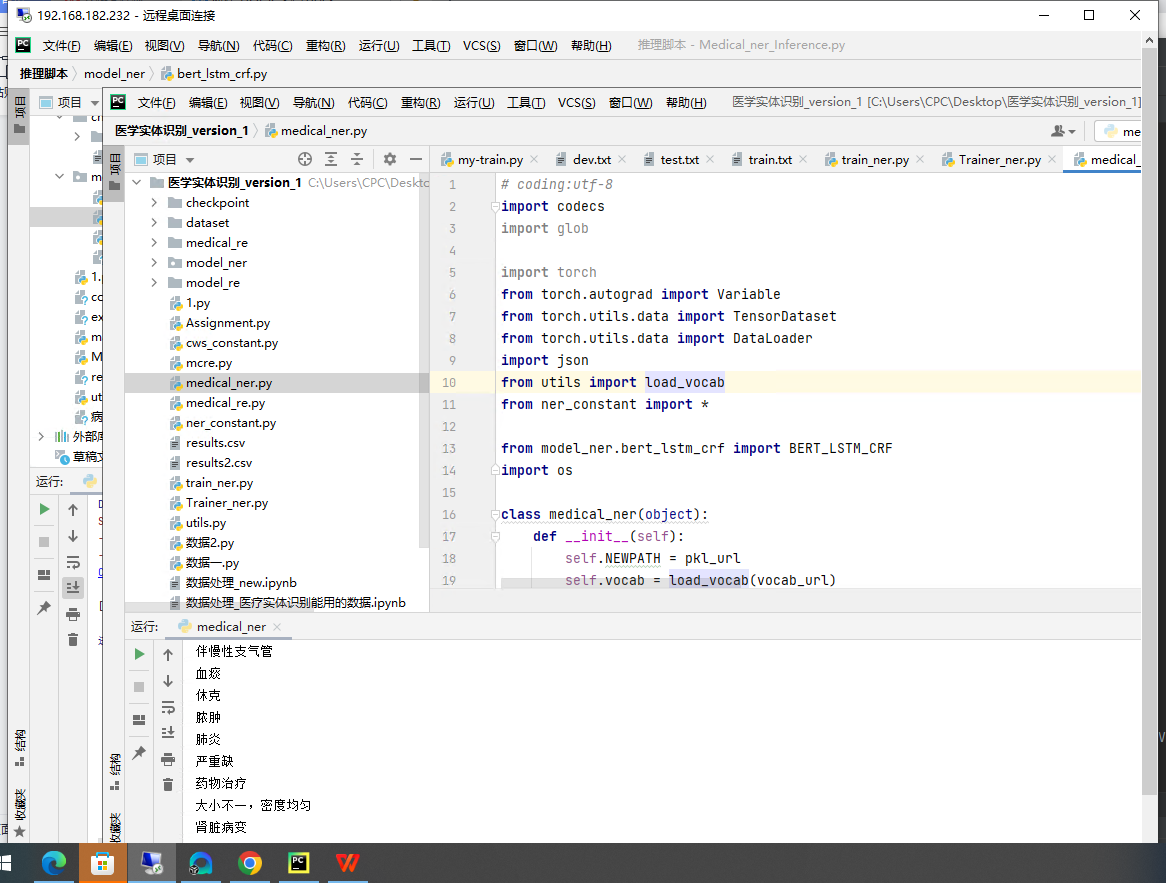
输入疾病描述（我扁桃体发炎了） 返回症状（嗓子疼，发烧）

输入症状描述（我嗓子疼，发烧） 返回疾病（扁桃体发炎）



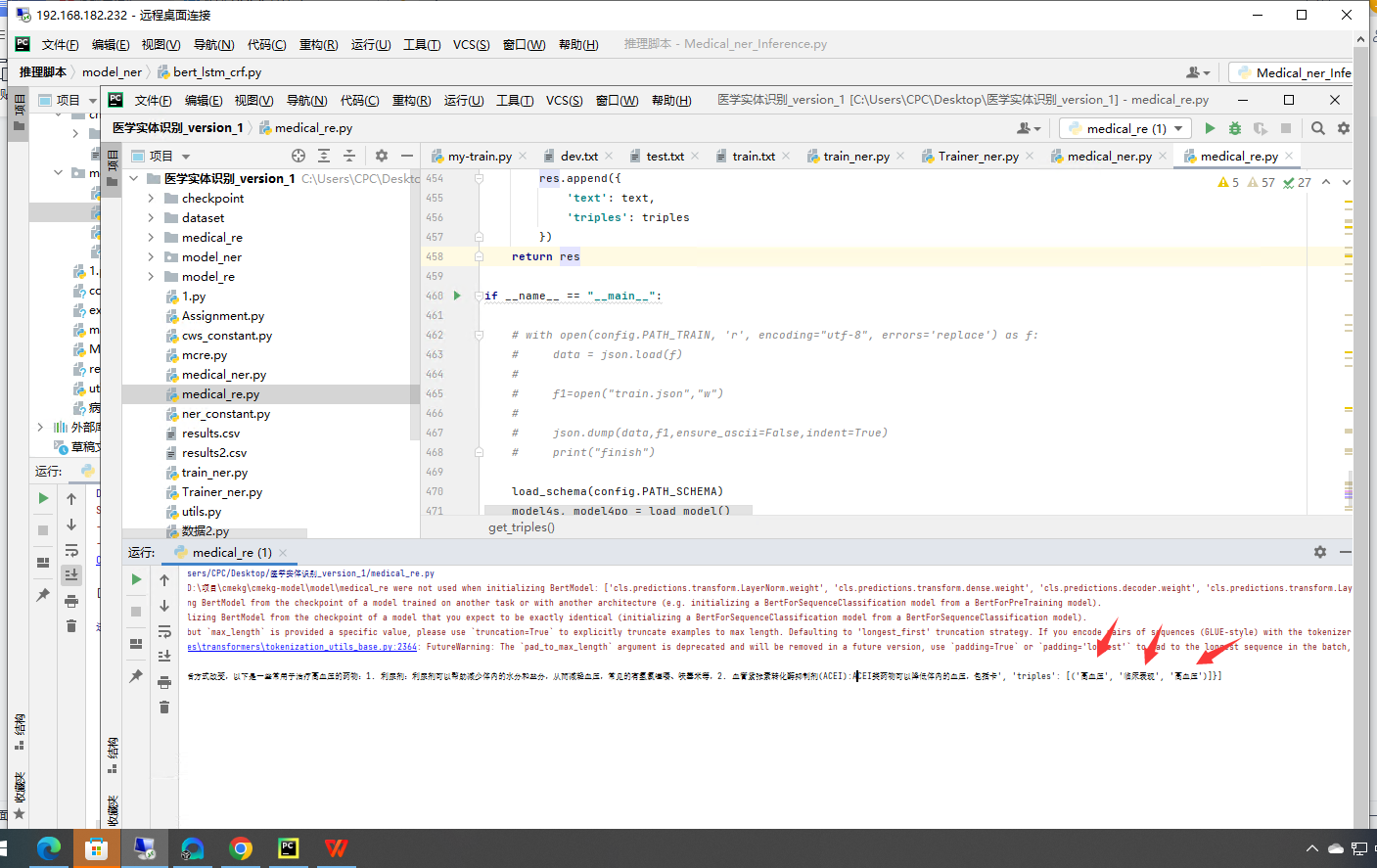
目前实体识别 只认识.症状的这个实体

[{'症状': '无粘液脓'}, {'症状': '无血块'}, {}, {'症状': '呕吐咖啡色胃内容物'}, {}, {}, {'症状': '大便'}, {}, {}, {'症状': '拟“上消化道'}, {}, {'症状': '黑便'},



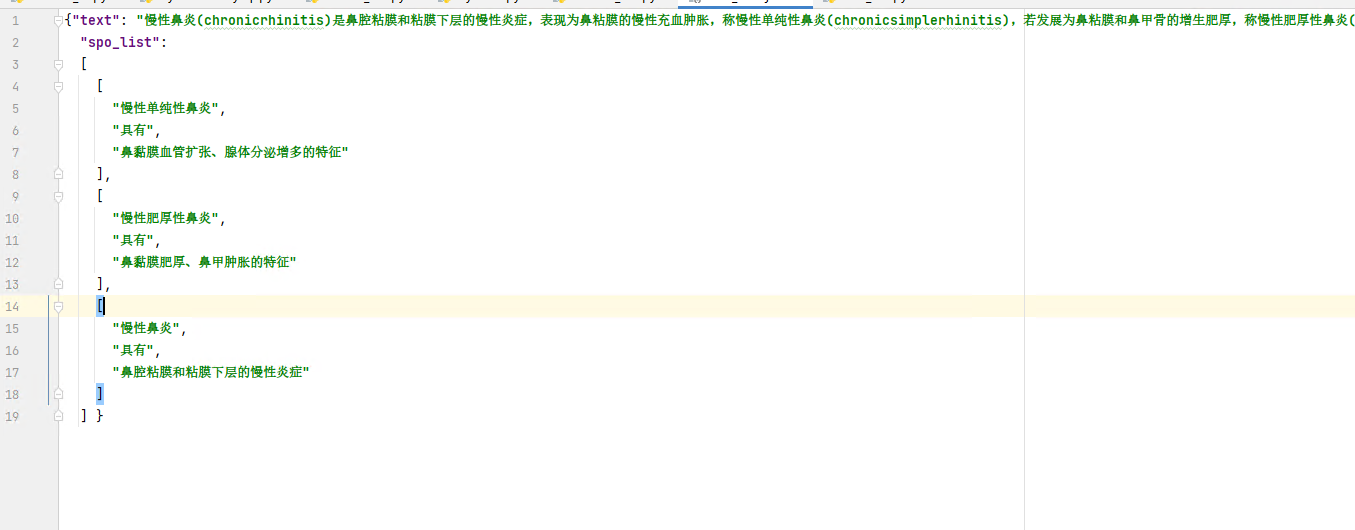
这个三元组是没训练的 所以需要把实体识别跟关系抽取 串起来 **（个人理解）**

用他model\_re默认的my\_train.py报错信息太多 没找到原因，需要轩轩帮忙**串起来 把关系抽取跑通，我现在跑不通（个人理解）**



去抽取 症状 和疾病的关系 为 “具有”

数据长这样



会有这样的数据标注

**个人理解：实体识别 和关系抽取 cmekg 都是用bert 按道理说是可以串起来的**

**目前 标注的数据 是疾病 具有 症状**