

This the note for 电子病历命名实体识别和实体关系抽取研究综述(2014)

电子病历的主要部分：出院小结、病程记录。自由文本等无结构数据为搜索、统计分析等研究制造了障碍。

电子病历中包含的实体：疾病、症状、检查、治疗

电子病历中设计的实体关系：疾病与治疗（这种病应该怎么治）、症状与治疗（这种症状应该怎么治）、疾病与检查（这种病应该做什么检查）、症状与检查（这种症状应该做什么检查）、疾病与症状（疾病导致症状）、疾病与疾病（一种疾病导致一种疾病）、症状与症状（伴随症状）

修饰语很重要。比如无发热，与发热就是完全相反的意思。

根据检查中的实体关系，生成描述病人情况的摘要。

专业术语有不同变体。（有人维护中文医学专业术语的变体吗，是否有资源？）

不同的实体间关系分别作为**不同的任务**研究。

概念的识别、医疗问题的修饰、概念间的关系是电子病历信息抽取的三个核心任务。

命名实体识别

是监督或半监督学习，先找到一些词的特征

可以看作词的分类问题：SVM、ME

可以看作序列标注问题：HMM、CRF

医疗实体词典是重要的资源。有？

实体关系抽取

需要得到关系类型、关系参数（关系的主体、客体）

有监督：SVM

无监督：Affinity propagation

高精度句法分析对实体关系识别有重要影响