

任务一：阅读文献，理解和掌握现有的各种关系抽取方法和模型

周一：

登录知网下载了各大中文期刊有关实体关系的论文，期刊包括电子学报、软件学报、中文电子学报、计算机科学等，共下载 34 篇论文

周二：

阅读论文：电子病历命名实体识别与实体关系抽取研究论述

法律文档与电子病历有很多相似点，故首先选择了这篇论文，篇幅较大，由于对实体关系概念没有明确的认识，该论文作为了解实体关系的切入点，花费一下午仔细阅读了文章，对实体关系有一个初步的了解

周三：

阅读论文：基于实体关系网络的微博文本摘要

采用关系网络有向图模型的方法来反应文本之间的结构关系

阅读论文：复杂中文文本的实体关系抽取研究

对基于最短依存路径核的中文实体关系抽取方法进行改进，引入了 7 条启发式规则，使之更适合于包含多个实体的句子的实体关系抽取

周四：

阅读论文：基于网络挖掘的实体关系元组自动获取

提出一种新的抽取实体关系的方法，利用两个 NE 和一个关键词构造实体关系，NE 可以是相同类型，也可以是不同类型

周五：

阅读论文：中文实体关系抽取中的特征选择研究

提出将实体关系分为包含实体关系和非包含实体关系，提出新的句法特征，构造不同的特征空间，采用了基于特征的学习算法，使用 CRF 模型对实体关系进行抽取

阅读论文：实体关系自动抽取

详细介绍了基于特征向量的机器学习算法，其中最重要的一步是实例特征向量的构造

阅读论文：基于多核融合的中文领域的是实体关系抽取（看不懂）

阅读论文：基于核方法的中文实体关系抽取研究（看不懂）

总结：

本周了解了抽取实体关系的具体方法和模型：

- （1）计算句法树之间核函数的方法（Zelenko 提出）：对于句法分析要求比较高
 - （2）计算依存关系树之间相似度的核函数（Culotta 提出）：对于句法分析要求比较高
 - （3）基于特征向量的机器学习算法：通过构造分类器进行关系分类，适合于只包含两个实体的句子的关系抽取
 - （4）最大熵学习算法：使用 SVM 模型作为分类算法
 - （5）基于卷积核函数
 - （6）基于多核融合
 - （7）基于改进的语义序列核方法：句子比较短、比较简单的情况，
 - （8）基于 Bootstrapping 的方法：关键在于种子的选择和迭代模式的生成
- 模型：SVM 模型、Maximum Entropy 模型、条件随机场模型（CRF）

