

第__9__周周记

周一	
完成内容	凌雅娟.新闻文本中的中文机构名识别和机构关系抽取研究
内容描述	该论文创新点为本文中的关系与常规意义上的关系不同，不是由一个关键词组成，而是由能表示两个机构实体之间关系的短语或者词序列构成，关系表示形式为（实体 1，实体 2，关系描述信息[时间属性][来源属性][.....]），其中[]表示可选的属性值。
未解决问题	

周二	
完成内容	刘一正.基于信息增益的互联网二元关系抽取
内容描述	人工给出了一些种子元祖，采用机器学习的方法从待抽取语料中得到更多符合要求的种子元祖，首先由种子元祖得到种子元祖共线句；处理由种子元祖得到的搜索语料，得到关系抽取模板，根据关系抽取模板得到模板共线句；从模板共线句中提取得到关系新元祖；判断是否满足自举循环的结束条件，若满足，则结束，若不满足，则将新元祖加入种子元祖集合。
未解决问题	

周三	
完成内容	郝博.基于句法模式识别的中文关系抽取方法研究与实现
内容描述	首先对句子进行分词、词性标注和依存句法分析，得到句子的依存句法分析图，然后抽取句子中的名词块和候选关系词，将句子中的名词块作为关系元祖的论元，将句子中的谓语动词作为关系元祖的关系词，对于抽取到的名词块要按照规则合并一些必要的修饰成分，如果符合名词块优化的条件，还要进行名词块的优化，对于抽取到的动词同样进行轻动词结构处理和特殊句式结构的处理，最后将关系元祖进行特征提取。
未解决问题	

周四	
完成内容	整理法律案例实体关系抽取的大体思路
内容描述	文本预处理（分词、标注、命名实体识别，筛选句子）->识别四大实体对类型(实体 1，实体 2)->依存句法树识别关系特征短语（关键词不能完整表达实体关系） 实体对类型参考语料标注{原告-原告，被告-被告，原告-被告，被告-原告} 采用用户满意度对抽取结果进行评价
未解决问题	

周五	
完成内容	
内容描述	
未解决问题	

周末	
完成内容	
内容描述	
未解决问题	

工程汇总	
完成任务	
任务描述	
代码量	
未解决问题	

论文汇总	
论文列表	[1]凌雅娟.新闻文本中的中文机构名识别和机构关系抽取研究 [2]刘一正.基于信息增益的互联网二元关系抽取 [3]郝博.基于句法模式识别的中文关系抽取方法研究与实现
论文摘要	
未解决问题	

下周任务	
工作	完成毕业设计开题报告
论文	
其他	
汇总	

日期:xx/xx/xx - xx/xx/xx