

第二周周记

周一	
完成内容	为更好阅读论文了解了中文分词与命名实体识别的概念
内容描述	通过百度了解到中文分词指的是将一个汉字序列切分成一个一个单独的词。命名实体识别，又称作“专名识别”，是指识别文本中具有特定意义的实体，主要包括人名、地名、机构名、专有名词等。
未解决问题	尚未了解它们之间的关系与互相影响。

周二	
完成内容	比较了法律文本和论文里的专利文献的区别
内容描述	法律文本要使用正式的词语，意义准确。专利文献涉及的领域多，术语嵌套现象较多。
未解决问题	未了解这些区别会造成哪些影响。

周三	
完成内容	学习了论文里中文分词的研究方法
内容描述	用到了统计和规则相结合的多策略分词方法
未解决问题	第一次接触这种方法，还没完全理解

周四	
完成内容	了解了中文分词的实验及实验结果分析
内容描述	了解了评测方法，如准确率的求法和各分词算法的比较结果
未解决问题	术语较多，公式和定义还未完全理解

周五	
完成内容	了解结合微博评论和转发信息的微博命名实体识别算法
内容描述	步骤是微博文本、分词、去常用词、统计词频、实体规则、实体抽取
未解决问题	无

周末	
完成内容	了解论文里具体的方法和实验结果
内容描述	比如微博文本预处理的方法和实验的评价标准
未解决问题	无

工程汇总	
完成任务	阅读了两篇论文
任务描述	了解了中文分词和命名实体识别的概念和实现的基本步骤
代码量	无
未解决问题	无

论文汇总	
论文列表	[1] 《面向专利文献的中文分词技术的研究》

	[2] 《面向微博文本的命名实体识别》
论文摘要	<p>[1]第一篇论文先介绍了专利的概念，专利文献的特点，然后提出统计和规则相结合的多策略分词方法和这个方法的具体流程。最后用求准确率的评测方法对实验结果作出分析</p> <p>[2]第二篇论文先介绍了微博的含义和特点，提出了一个结合微博评论和转发信息的微博命名实体识别算法，然后举了一个算法的例子，对算法作出具体描述，接着收集新浪微博的 30000 条微博作实验。评价并分析实验的结果。</p>
未解决问题	几个公式看不太懂

下周任务	
工作	继续查找相关的论文和继续阅读相关文献。
论文	继续寻找与中文分词和命名实体识别相关的论文
其他	无
汇总	了解更多与我的课题相关的知识

日期:2018/01/08 -

2018/01/13