

## 第 5 周周记

周一	
完成内容	学习 Stanford CoreNLP
内容描述	完成对一篇文章的分句，分词
未解决问题	

周二	
完成内容	继续学习 Stanford CoreNLP
内容描述	完成对一篇文章的词法分析和命名实体识别
未解决问题	

周三	
完成内容	整理已完成的工作
内容描述	
未解决问题	

周四	
完成内容	学习如何标注语料
内容描述	学长发来语料标注的规范ace2005。 读ACE Chinese Annotation Guidelines for Events 2005
未解决问题	

周五	
完成内容	学习如何标注语料
内容描述	继续读 ACE Chinese Annotation Guidelines for Events 2005
未解决问题	实际操作可能有问题，等语料到位后开始尝试

周末	
完成内容	阅读论文《句子级中文事件抽取关键技术研究》丁校
内容描述	查询《音乐领域典型事件抽取方法研究》的相关论文，未查到类似同组的论文，但查到这篇深化了的硕士论文。 基于触发词聚类的事件类型的自动发现，基于同义词词林的事件触发词自动扩展。
未解决问题	

工程汇总	
------	--

完成任务	学习使用 Stanford CoreNLP，学习语料标注规范
任务描述	阅读基于触发词聚类的事件类型的自动发现的相关论文部分
代码量	
未解决问题	

论文汇总	
论文列表	[1]丁效. 句子级中文事件抽取关键技术研究[D].哈尔滨工业大学,2011.
论文摘要	[1]1.提出了基于触发词聚类的事件类型体系自动构建方法。不仅成功地发现了 ACE 2005 预先定义好的 33 类事件类型,还在金融领域和音乐领域有着不错的表现,这说明该算法具有较强的鲁棒性和领域自适应性。2.融入外部词典资源解决 ACE 语料数据稀疏问题。本文提出了基于同义词词林自动扩展触发词的算法,该方法成功的融入了外部丰富的语义资源和背景知识,并在 ACE 2005 语料上取得很好的实验结果。3.提出模板泛化方法解决基于模式匹配的事件元素抽取召回率不高的问题。本文提出了基于 BestMatch 的模板泛化算法。对任意两个事件实例模板进行匹配,计算其匹配代价并泛化。经 ACE 2005 语料实验证明该方法一定程度上弥补了模式匹配召回率不高的问题。4.在事件元素抽取任务中,本文提出,用依存句法分析结果定位事件元素,用名词短语句法分析结果识别事件元素边界。这种方法很好的结合了两种句法分析器的优点:依存句法更加擅长于分析句法成分,给出事件元素核心词,而名词短语句法更擅长识别名词短语,给出完整的事件元素。
未解决问题	

下周任务	
工作	尝试对语料进行标注，查找标注工具，继续研究基于触发词聚类的事件类型的自动发现
论文	
其他	
汇总	

日期:2018/01/29 - 2018/02/03