自然语言处理小组工作周报

|  |
| --- |
| 工作内容 |
| 1. 对词性标注、地名实体识别的数据进行了训练和测试 2. 运用维特比算法。 3. 学习学长的项目，能够提取关键字和给出权重 4. 优化前端数据存储问题，给词性标注添加了UI云标签 5. 学习Python相关的知识 |
| 问题记录 |
| 1. 对文本的分析还不是很智能 2. 词云展示方法 3. 对gitlab、github的上传时对一些文件上传不了 |
| 下周计划 |
| 1. 训练和优化分词、词性标注、地名实体识别model 2. 训练出命名实体识别model，能够跑自己的数据 3. 训练出依存句法分析model，并完成数据的测试 4. 训练出语义决策model，并完成数据的测试 5. 对比之前的nlp方法并在前端展示 |

时间：2018年10月29日星期一~2018年11月4日星期天

时间：2018年10月22日星期一~2018年10月28日星期天

|  |
| --- |
| 工作内容 |
| 1. 本周全组成员都在一起学习crf，使用crf++完成一些实例，对数据的训练、测试，对一些文本的分词、词性标注、提取关键字。 2. 对张涛给的一些项目（命名实体识别）和一些工具都使用了起来，配置pycharm环境，运行张涛项目。 3. 学习python的基本语句，一些基础用法。 4. 对前端展示的优化。 |
| 问题记录 |
| 1. Python使用不熟练。 2. 加了一个URL提取文本的功能模块之后，对数据的提取有时会出现bug。 3. 对数据的训练集和测试集测试数据之后会出现分词不正确的现象。 |
| 下周计划 |
|  |