6月24日-6月25日

进行微博网页源代码分析，根据源代码特点对微博正文评论进行初步爬取。

6月26日

今天的主要任务就是根据爬取的微博内容，做出微博正文高频词的词云分析。

首先研究分词库，包括jieba库、SnowNLP库和PkuSeg库，根据分析各类库的特点，最终选择jieba库。安装jieba库的过程中，我先是遇到了报错信息，经过检查，是pycharm的配置环境不是3.7，经过修改后，jieba安装成功。

成功运用jieba库实现分词后，再对分词算法进行优化，我们对结果分析可知，初步的分词算法无法识别人名，因此我们手动添加需要识别的人名到分词辞典里。

接下来在绘制词云过程中，首先学习了wordcloud库，了解各个函数的基本作用，以及每个参数的意义，最后成功生成了词云图。