**社交网络事件事态挖掘**

吴松泽 2120151048

徐程程 2120151049

刘琳 2120151012

1. 数据采集

社交网络的事件获取主要来源是新浪微博，方式为爬虫程序。

我们的爬虫程序是模仿Scrapy框架自行开发实现的单机爬虫程序，使用python的模块BeautifulSoup和urllib2进行开发，数据库采用MongoDB。

事件微博的获取基于微博搜索接口，获取主要字段包括：微博正文，发布时间，评论数，转发数，点赞数等。

1. 数据处理

根据事件的发展关系，对事件进行事态分析，并最终训练出相应的预测模型，能对特定的数据事件进行事态分析。事件事态主要分为：开始、突发、发展、拐点、衰退、结束等事态标签，实现将事件情况进行数据分析并最终获得事件的各个事态阶段的标签。

事态分析过程

1. 结果展示





