每天的词，hashtag哪个最多

@的人，re-tweet的人每天的时间序列

舆论讨论模型：参与者、议题和行动，议题和行动从文本中找出，参与者从文本和rt@和@中来

计算某一个人发tweets的时间序列

计算某一个词随着时间的变化它的时间序列，但是现在不知道哪一个词更重要，计算TF、RTF值

怎么样使用Python计算TF/RTF,

词频 (term frequency, TF)=一词语出现的次数除以该文件的总词语数

逆文本频率指数” (Inverse documentfrequency 缩写为IDF)=一个文本出现的数量/总文本的数量 的倒数，取log，乘以词出现的总次数

1 从原始数据中先取一部分测试数据

1. 通过line的方法将原始数据分为一个tweet一个tweet的格式
2. 通过for loop 可以计算出tweets总数，通过分词再for loop可以计算出每个tweet中总的词数量，通过+=可以计算出一个tweet中一个词的数量，通过+=可以计算出出现过这个词的tweet文本数量