# 实验 网络疫情信息的主题分析

一、实验描述

对微博热搜数据和谣言数据利用结巴分词和词云工具进行分析和可视化展示，进一步了解自然语言的处理工具，学会对语料库进行分析。

二、数据描述：

2组数据：

第一组为微博数据： 微博热搜标注.xsl,

第二组为谣言数据：下载地址：<https://github.com/BlankerL/DXY-COVID-19-Data/blob/master/csv/DXYRumors.csv>

## 三、实验语言

Python

## 四、基本任务要求：

1. 针对微博数据中的主题为“疫情”的微博，对其“热搜内容”进行分词，计算词出现的次数，将出现次数前20名的动词输出，将出现次数前20名的名词输出，并采用“词云”可视化输出（可以定义多个时间段，看舆情的变化）
2. 针对微博数据中的主题为“政府”的微博，对其“热搜内容”进行分词，计算词出现的次数，将出现次数前20名的动词输出，将出现次数前20名的名词输出，并采用“词云”可视化输出。（可以定义多个时间段，看随着时间的舆情变化）
3. 针对谣言数据，对“title”内容进行分词，计算词出现的次数，将出现次数前10名的为动词输出，将出现次数前10名的名词输出，出现次数前10的词（不考虑词性）输出，并采用“词云”可视化输出.（不定义时间段）

（四）可以自定义其他的分析主题

（五）提示：用结巴分词工具（jieba）分词和用“pycloud”或“wordcloud”做词云；

分词中考虑到在词典中要自己添加下新词：如：“火神山”

## 五、参考代码：

