

# 信息内容安全实验技术报告

项目名称：微博的内容识别与控制

韦昆杰

1181000420

348971397@qq.com

2021 年 6 月 23 日

## 目录

<b>1 概述</b>	<b>3</b>
1.1 测试目的	3
1.2 测试内容	3
<b>2 测试环境</b>	<b>3</b>
2.1 硬件环境	3
2.2 软件环境	3
<b>3 测试方案</b>	<b>4</b>
3.1 功能测试	4
3.1.1 微博热搜子系统	4
3.1.2 微博搜索子系统	4
3.1.3 微博过滤子系统	4
3.1.4 微博管控子系统	4
3.1.5 数据库子系统	4
3.2 性能测试	4
<b>4 测试方法与结果分析</b>	<b>5</b>
4.1 功能测试	5
4.1.1 微博热搜子系统	5
4.1.2 微博搜索子系统	6
4.1.3 微博过滤子系统	6
4.1.4 微博管控子系统	7
4.1.5 数据库子系统	7
4.2 性能测试	7
<b>5 测试总结</b>	<b>7</b>
5.1 基本功能	7
5.2 性能指标	7
5.3 改进建议	8

## 1 概述

### 1.1 测试目的

对微博内容识别和控制进行测试，检查是否能完成预期功能，此外还有对性能的测试

### 1.2 测试内容

1. 微博热搜子系统
2. 微博搜索子系统
3. 微博过滤子系统
4. 微博管控子系统
5. 数据库子系统

## 2 测试环境

### 2.1 硬件环境

处理器: Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz 1.80 GHz

内存: 8.00 GB (7.85 GB usable)

System type 64-bit operating system, x64-based processor

### 2.2 软件环境

操作系统: Windows 10

IDE: Visual Studio Code

开发语言: TypeScript

界面实现: HTML CSS

数据库系统: MongoDB 数据库

## 3 测试方案

### 3.1 功能测试

用户进入相应的功能页面，完成相应操作，实现对应功能

#### 3.1.1 微博热搜子系统

用户在 localhost:3000/trend 网站点击热搜按钮，系统获取实时热搜并展示到网站页面上

#### 3.1.2 微博搜索子系统

用户在 localhost:3000/search 网站输入要搜索的微博关键字并提交，系统获取对应的微博存储到数据库，并展示到网站页面上

#### 3.1.3 微博过滤子系统

用户在 localhost:3000/search 网站输入要过滤的微博关键词，系统根据 KMP 或 AC 算法匹配微博，将匹配到的微博展示到网页页面上

#### 3.1.4 微博管控子系统

用户在 localhost:3000/control 网站输入要管控的微博 URL 或在 localhost:3000/search 选择管控过滤的微博，系统将举报需要管控的微博

#### 3.1.5 数据库子系统

用户可以在 MongoDB 数据库看到存储的热搜和微博

### 3.2 性能测试

测试 BF、KMP、AC 算法匹配相同字符串需要的时间

测试网页交互各项功能是否正常且流畅

## 4 测试方法与结果分析

### 4.1 功能测试

#### 4.1.1 微博热搜子系统

进入 localhost:3000/trend 页面，点击热搜榜，获取热搜

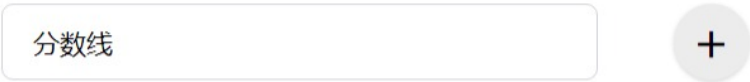
热搜榜

序号	关键词	热搜次数
1	高考分数线	5253547
2	高考成绩	2201025
3	教育部要求严禁炒作高考状元	1694231
4	班主任把全班45位同学名字写成诗	1491483
5	匪我思存 妈我有个剧要播了	1214416
6	航天员太空吃播	973856

4.1.2 微博搜索子系统



4.1.3 微博过滤子系统



1 分数线 —

可以查看到过滤的微博结果

微博URL	发送微博用户	微博创建时间	具体内容	赞数	评论数	转发数
https://weibo.com/2656274875/Klv9yidqz	央视新闻	2021-06-23 11:11	#高考分数线# 【转存！#多地公布高考分数线#！持续更新】今天，安徽、宁夏、广西等地公布高考分数线！祝考生取得好成绩，考入理想大学！	62947	6228	2993
https://weibo.com/2202896360/KlvDCCZL3	西安身边事	2021-06-23 14:56	【转存！！3005所高校全名单（名单未包含港澳台地区高校）】全国各地#高考分数线#陆续公布，#陕西2021高考分数线#也即将公布，这份3005所高校全名单你指的拥有	12	3	9
https://weibo.com/5557742758/KlvOdb4vS	九月遇山城	2021-06-23 15:24	#高考分数线# 笑的，文科生就不配上大学了是吗，我也是文科生毕业的，那些看不起文科生的大可不必，如果觉得文科生有脑子会背就行那我是不是也可以理解为理科生会背公式会用就行	48	12	1
https://weibo.com/1916780343/KlvQF476p	微成都	2021-06-23 15:30	【转存！教育部解读高考志愿填报8大热点问题】俗话说“七分考，三分报”，志愿填报也是一门学问。平行志愿105%投档，那剩余的5%就被退档了吗？专业级差是每个学校都有吗？高考志愿填报8个热点问题 转发收藏！#高考成绩##南开录取通知书送两粒嘉兴莲花种子##高考成绩##14个省份新高考录取变化#	3	1	1

4.1.4 微博管控子系统

对于过滤到的微博或者特定的其他微博，可以在 localhost:3000/control 网站输入要管控的微博 URL 对该微博进行举报

 请输入管控的URL

4.1.5 数据库子系统

在数据库子系统中，可以查看存储的热搜和微博信息

```

    _id: ObjectId("60cc2f437decb80fc0aaee92")
  user: Object
    id: 2609400635
    username: "TFBOYS-王俊凯"
    created_at: "2014-09-21 10:21"
    content: "今天我十五岁了，#我的十五岁#有那么多的你们陪伴我，谢谢这几年来你们的一直陪伴，《给十五岁的自己》不仅是给自己的生日歌，也是送给所有支持我的..."
    likes_count: 866723
    comments_count: 0
    reposts_count: 0
    __v: 0

    _id: ObjectId("60cc2f437decb80fc0aaee94")
  user: Object
    created_at: "2016-11-01 12:57"
    content: "#王俊凯触电张艺谋长城#@TFBOYS-王俊凯 MEQQME//@马雅瑶酒K://@王俊凯的特战一组:#王俊凯触电张艺谋长城# #从开始到未..."
    likes_count: 0
    comments_count: 0
    reposts_count: 0
    __v: 0
```

4.2 性能测试

测试 BF、KMP、BM 算法匹配相同字符串需要的时间，结果如下：

```
PS C:\Users\wkj\Desktop\my-app\utils> deno run test.ts
bf: 4ms
bm: 6ms
kmp: 3ms
```

5 测试总结

5.1 基本功能

基本功能得到实现

## 5.2 性能指标

根据测试，实际系统中采用 KMP 算法作为单模式匹配，AC 算法作为多模式匹配

## 5.3 改进建议

还可以采用更加先进的匹配算法，如多模式匹配的 Wu-Manber 算法

同时，对微博的管控需要一个登录用户，会受到单个用户举报或者发言频率的限制，也难以实现并发 (需要多个账号或单个账户共享)