舆论监控系统

——数据挖掘大作业

**实验名称： 舆论监控系统**

**姓名： 于洋**

**学号： 2017326603141**

**班级： 17信科（2）班**

**Github:** <https://github.com/yu627482453/PublicOpinionSystemV1>

目录

[舆论监控系统 1](#_Toc53856138)

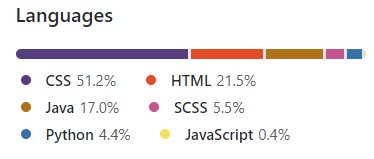
[一、 业务逻辑总构图 3](#_Toc53856139)

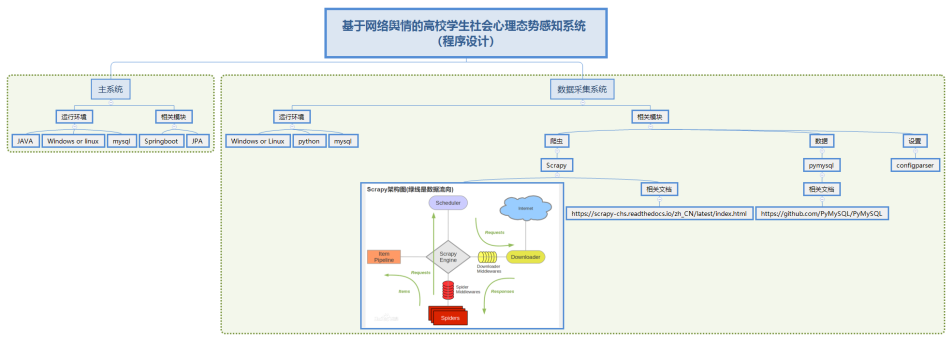
[二、 数据采集（爬虫） 4](#_Toc53856140)

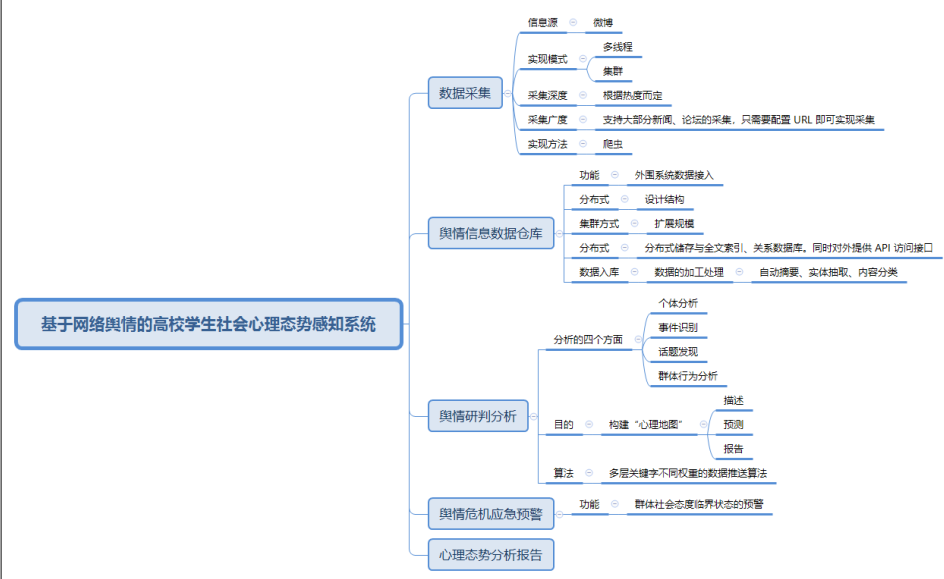
[三、 Hadoop 数据仓库 6](#_Toc53856141)

[四、 总服务器 7](#_Toc53856142)

1. 业务逻辑总构图

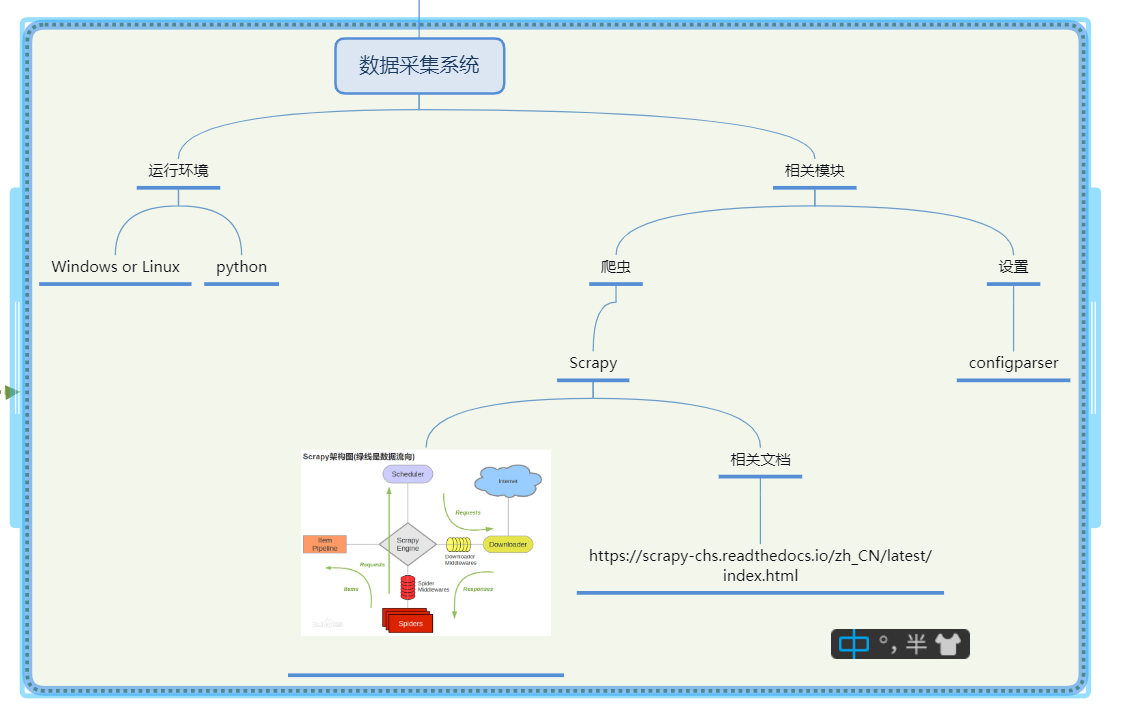


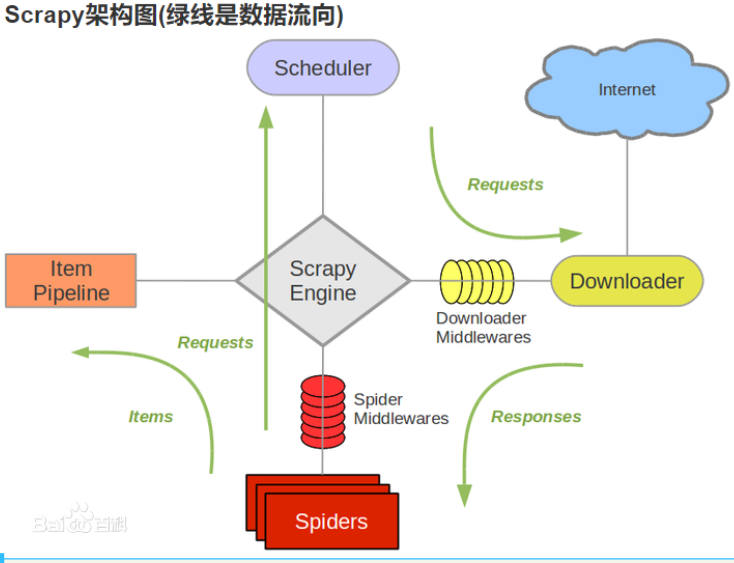




1. 数据采集（爬虫）

Python3 + Scrapy + Jieba





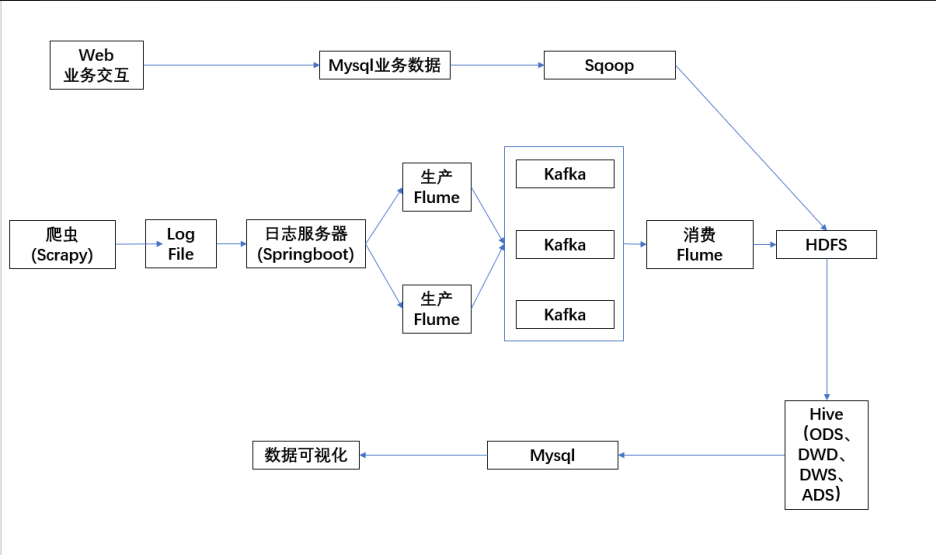
以Scrapy框架下的爬虫，爬取新浪、知乎等相关论坛的舆论信息。

Scrapy框架可以规范爬虫的编写，便于扩展其他论坛的爬取。

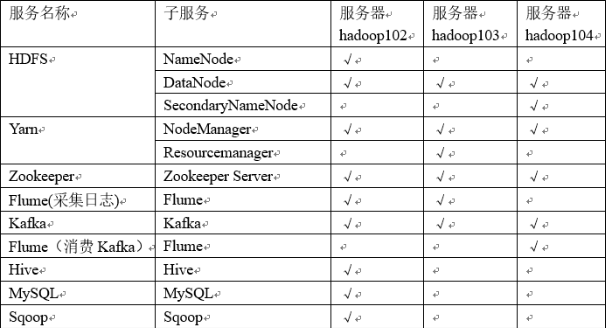
1. Hadoop 数据仓库

Linux（CentOS 6）+ Hadoop 2.7.2

结构流程图：

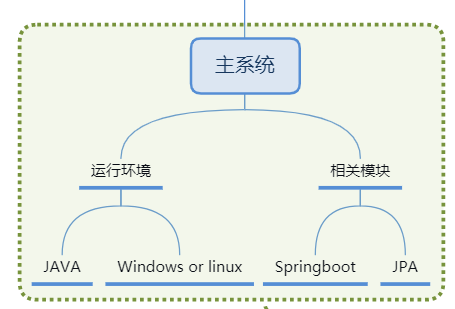


服务器装配服务：

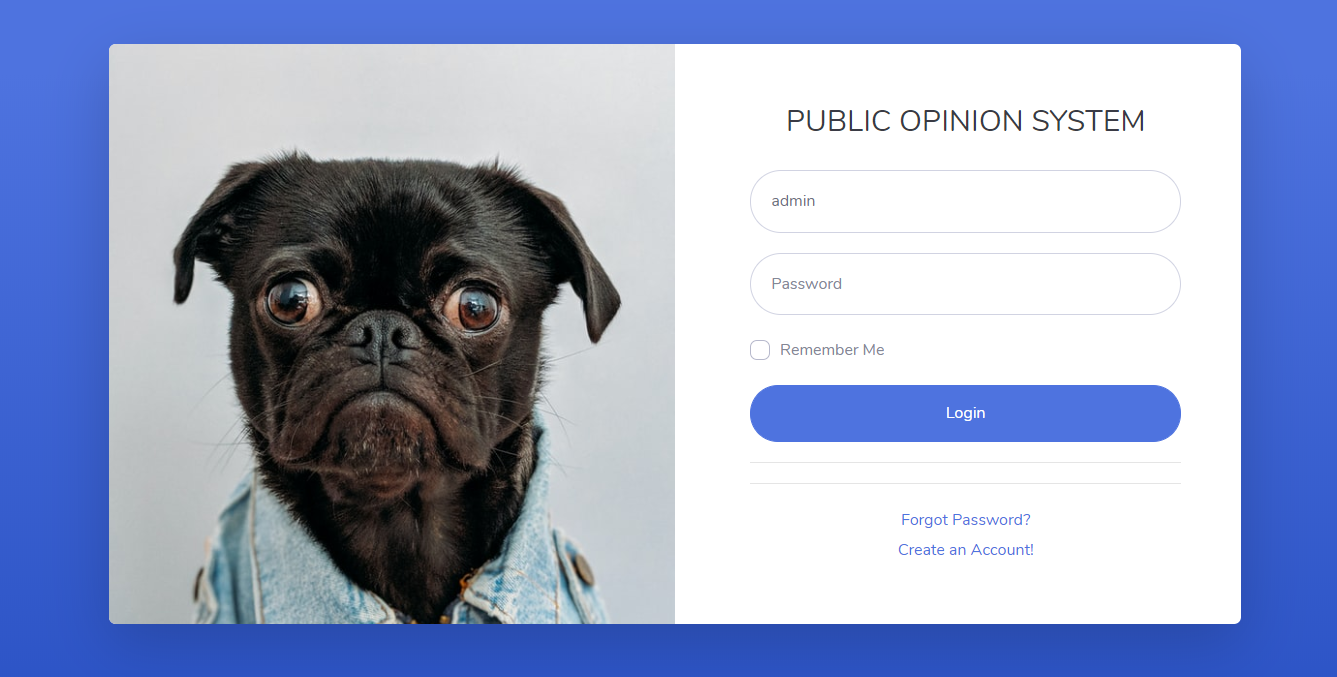


1. 总服务器

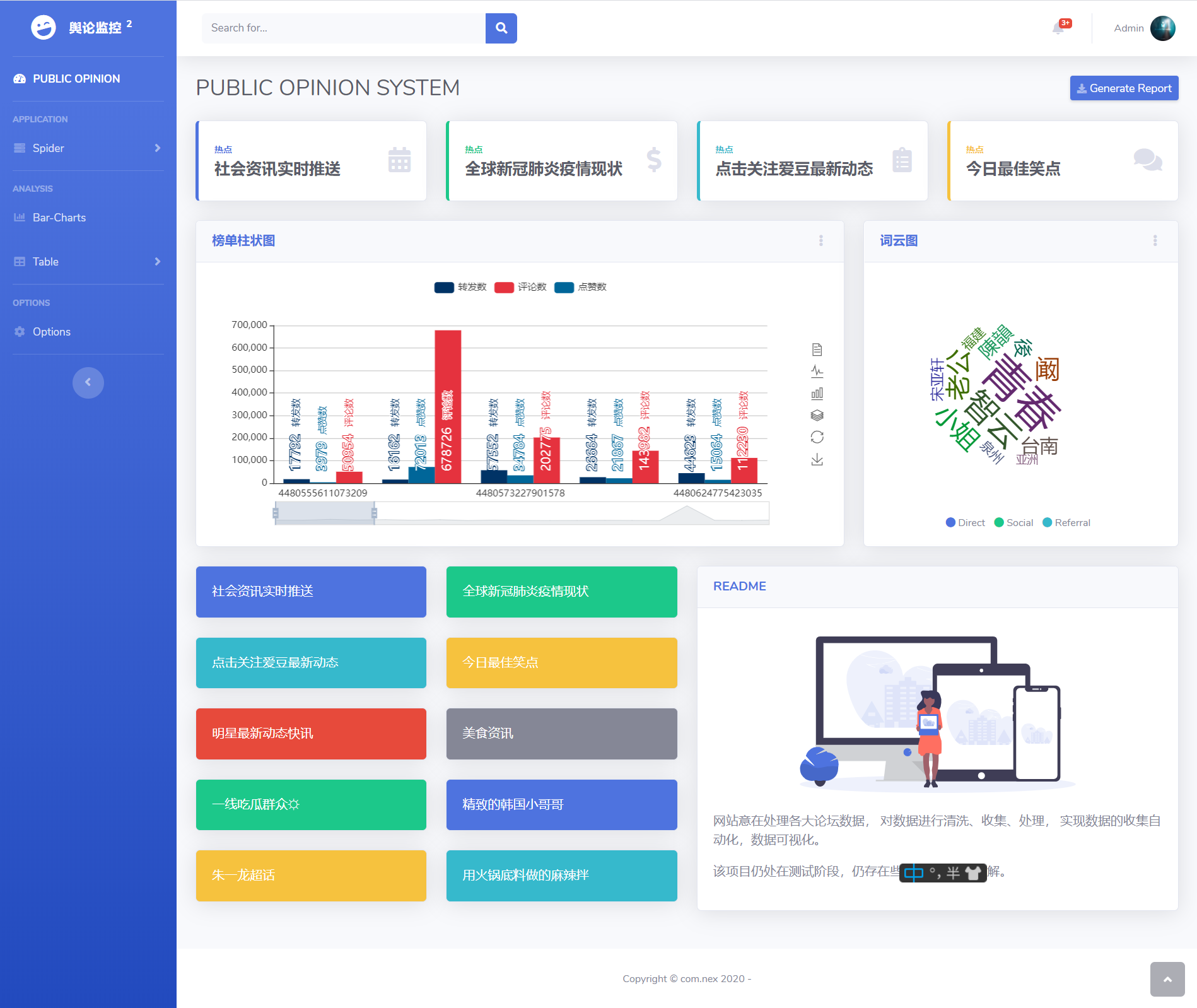
Java + Springboot + Html5 + JS + CSS



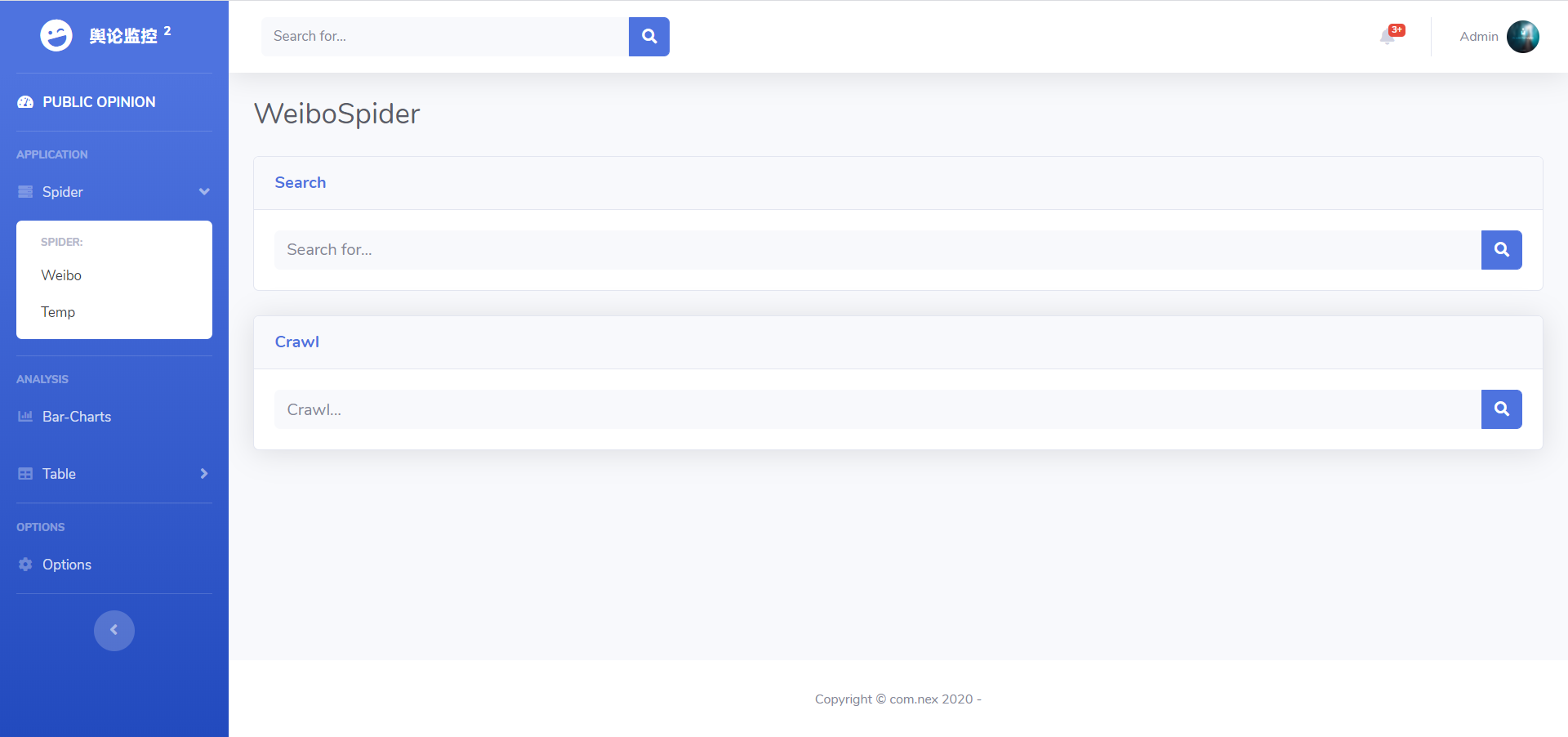
前端登录界面：



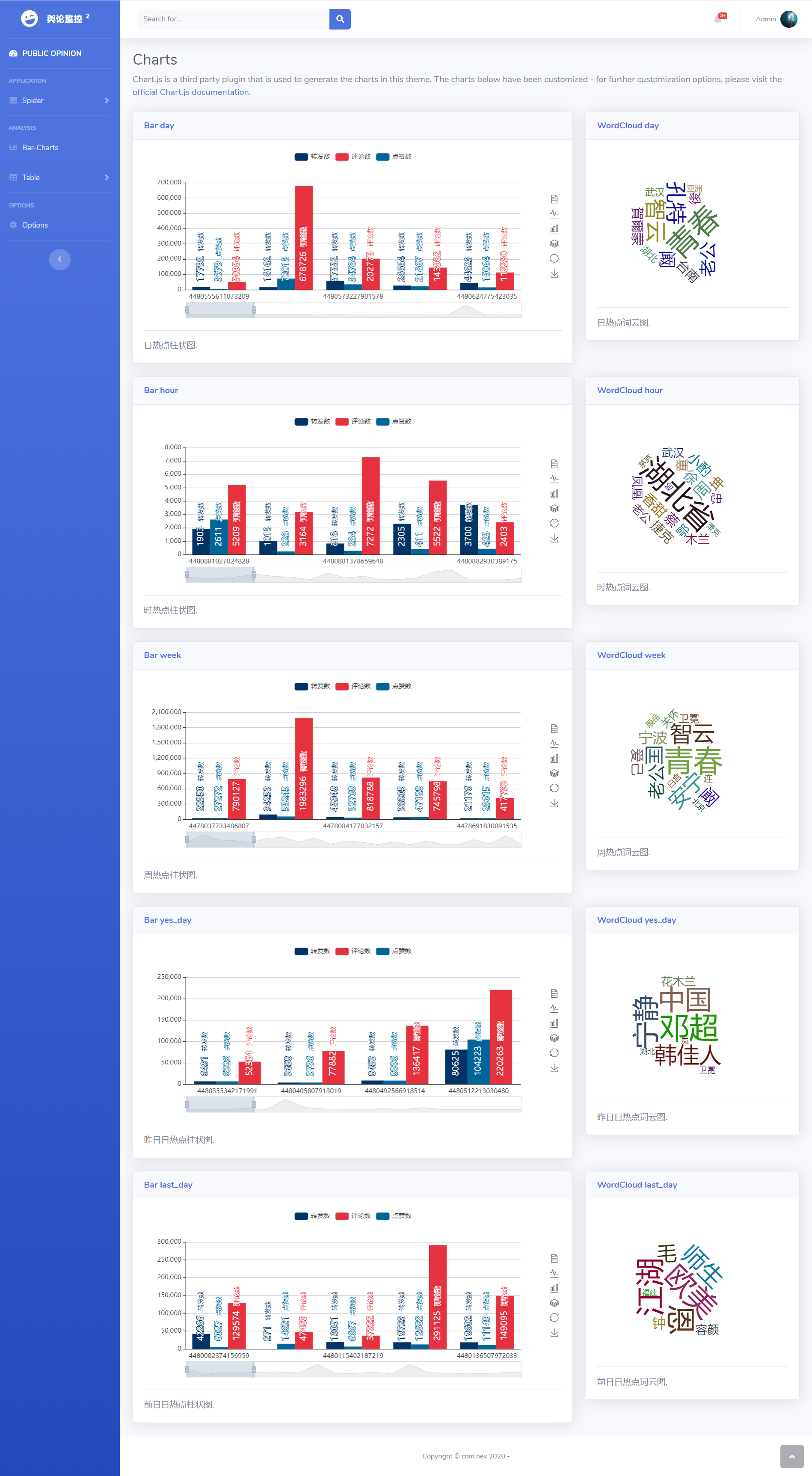
主界面：



爬虫界面：



图表界面：



表格界面：

