**地理空间情报数据采集器**

**项目使用说明**

日 期：二〇二一年八月

## 后台API代码

#### 运行

在code/node-api文件夹下右键管理员方式打开，输入pm2 start index.js即成功运行（运行成功后可关闭对话框）

#### 停止

pm2 stop all

#### 重启

pm2 restart index.js

## 文本任务代码

在部署成功的前提下，可能会出现以下问题。

#### 获取cookie

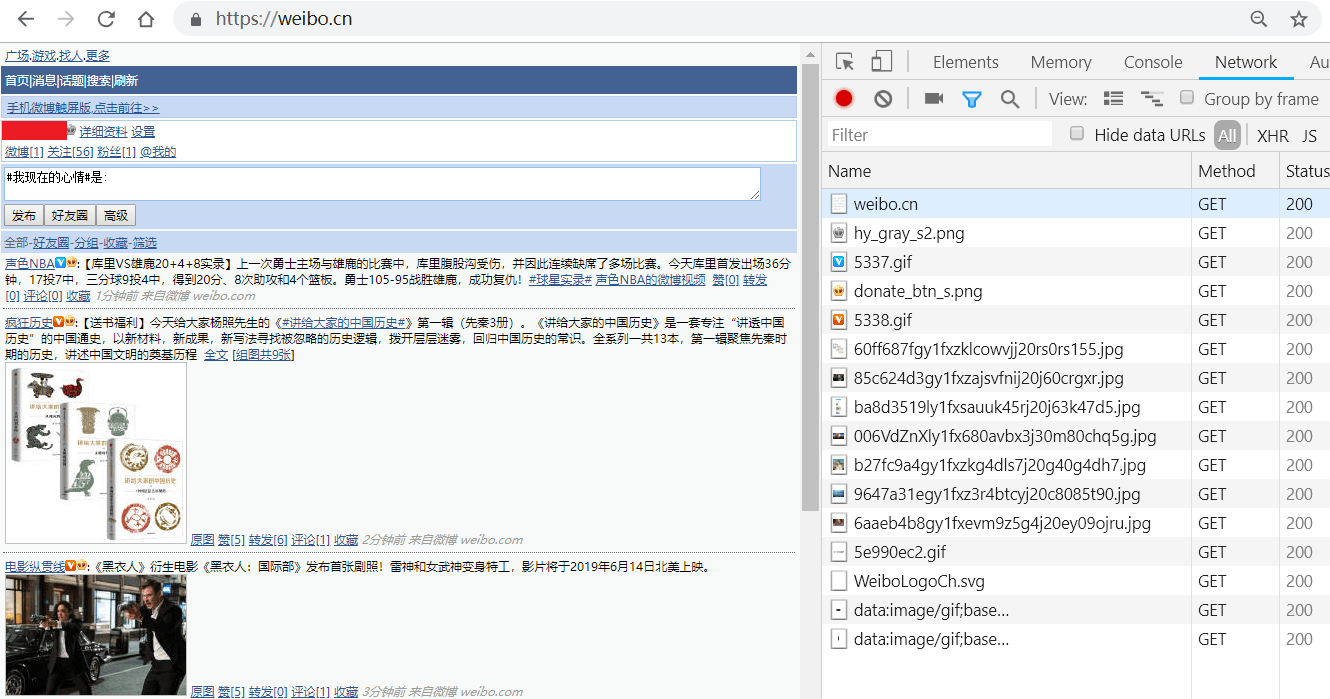
用Chrome打开<https://passport.weibo.cn/signin/login>；

输入微博的用户名、密码，登录，如图所示：

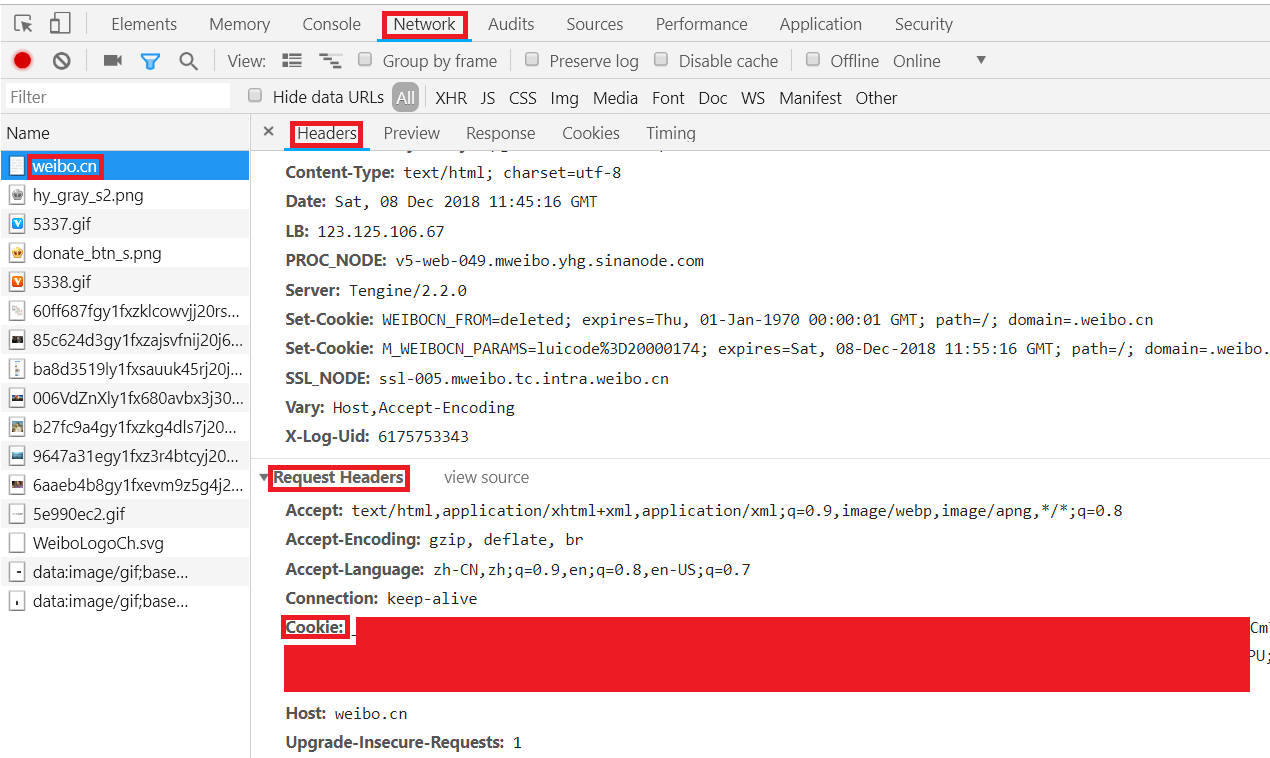


登录成功后会跳转到[https://m.weibo.cn](https://m.weibo.cn/);

按F12键打开Chrome开发者工具，在地址栏输入并跳转到[https://weibo.cn](https://weibo.cn/)，跳转后会显示如下类似界面:



依此点击Chrome开发者工具中的Network->Name中的weibo.cn->Headers->Request Headers，"Cookie:"后的值即为我们要找的cookie值，复制即可，如图所示：



**将获取到的cookie复制到****code/text-search/weibo/settings.py的DEFAULT\_REQUEST\_HEADERS的cookie中即可**

#### 修改MySQL数据库

对应修改code/text-search/weibo/settings.py中的MYSQL开头相关配置参数

#### 刷新text-search部署文件

修改过cookie或MySQL数据库文件后，应重新部署一下代码。

只需在code/text-search下右键管理员运行scrapyd-deploy weibo-search -p weibo即可

## 图像任务代码

#### 获取cookie

与1.1相同，使用相同的即可

#### 图像保存位置

修改code/photo-search文件中settings.py中的IMAGES\_STORE，注意修改为绝对路径即可，以防出错。

#### 图像查询位置

修改code/node-api/index.js，搜索photo\_num定位代码位置，修改dir的路径与图像保存位置相同即可在前端界面中正确查询数量。

#### 刷新photo-search部署文件

修改过cookie或图像保存位置后，应重新部署一下代码。

只需在code/photo-search下右键管理员运行scrapyd-deploy photo -p weibo\_photo即可

还需重新运行后台API代码，具体操作见（一.后台API代码）

## 视频任务代码

#### 视频保存位置

修改code/node-api/lujing.py文件即可

#### 视频查询位置

修改code/node-api/index.js，搜索video\_num定位代码位置，修改dir的路径与图像保存位置相同即可在前端界面中正确查询数量。

#### 刷新photo-search部署文件

重新运行后台API代码，具体操作见（一.后台API代码）

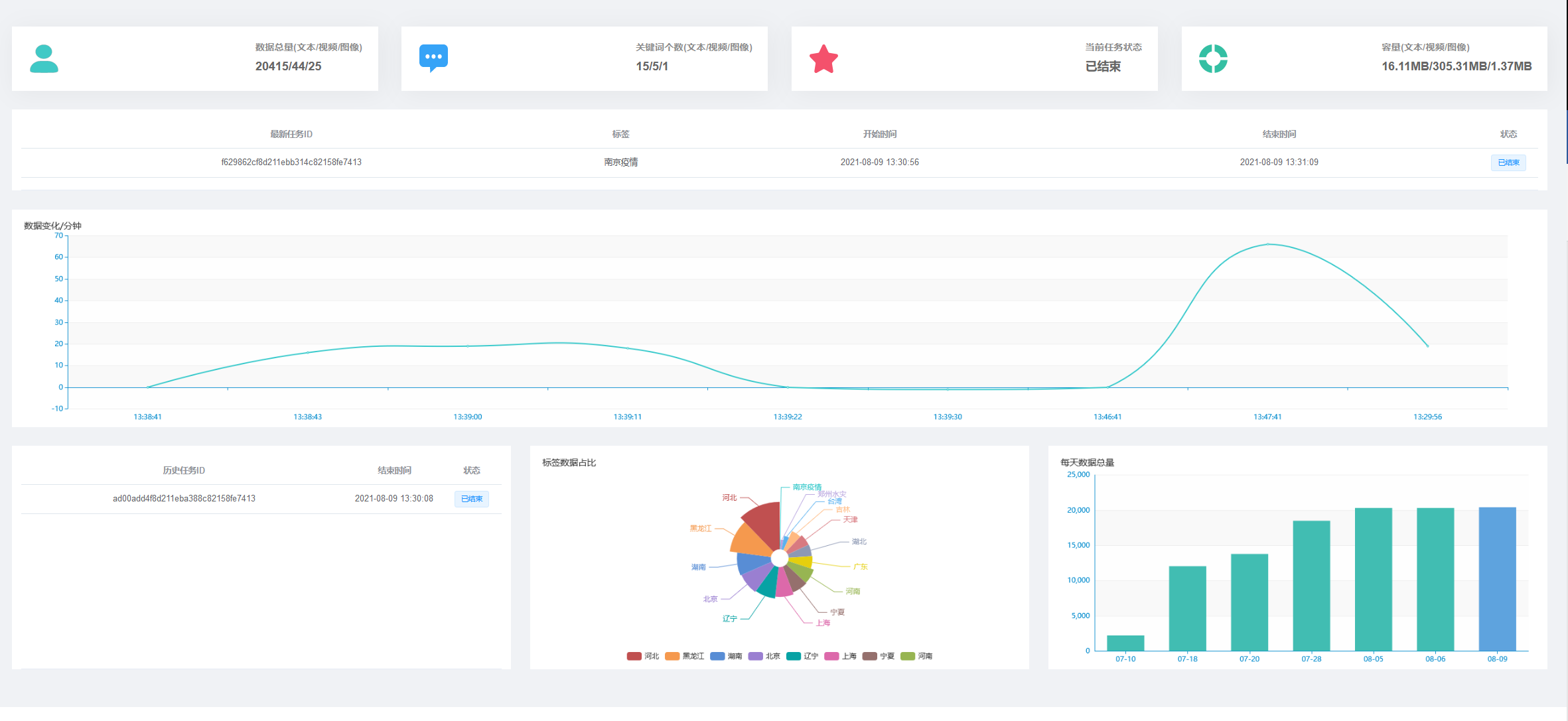
## 前端界面

#### 访问

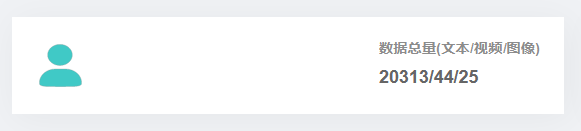
Ip:端口号即为前端地址（本地访问IP即为127.0.0.1）。

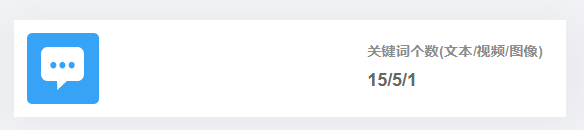
#### 首页

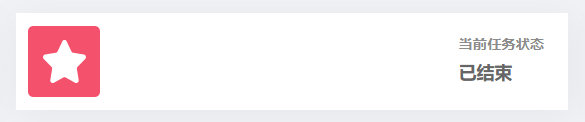
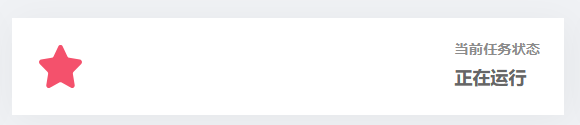
每60秒刷新一次数据，更新所有文字以及图表信息。



首页第一栏为数据总量，关键词个数，当前任务状态（已结束/正在运行）以及数据容量。

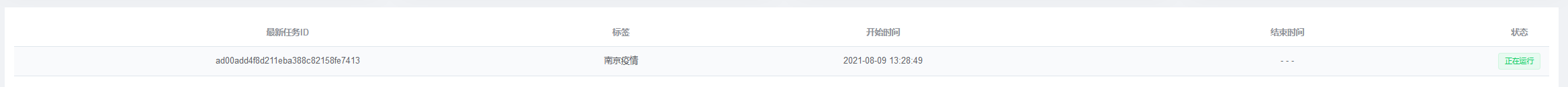


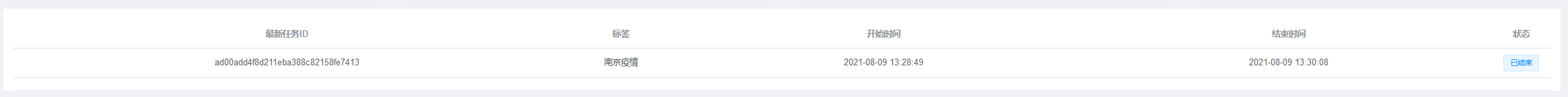


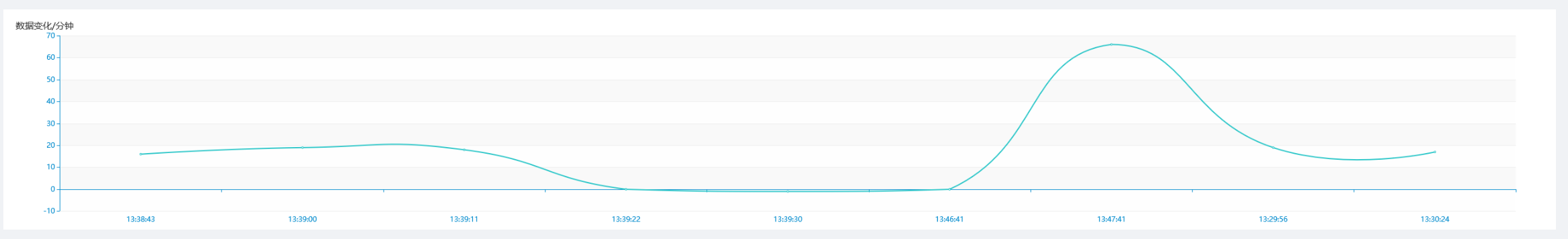


首页第二栏为最新一条任务的具体情况，包括任务ID，标签，开始时间，结束时间和运行状态（正在运行/已结束）。

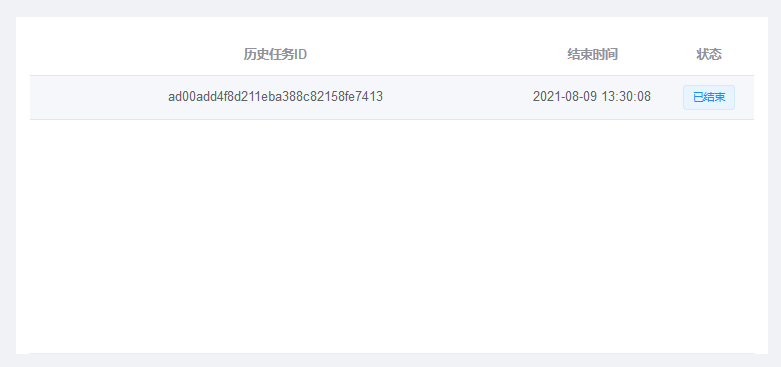




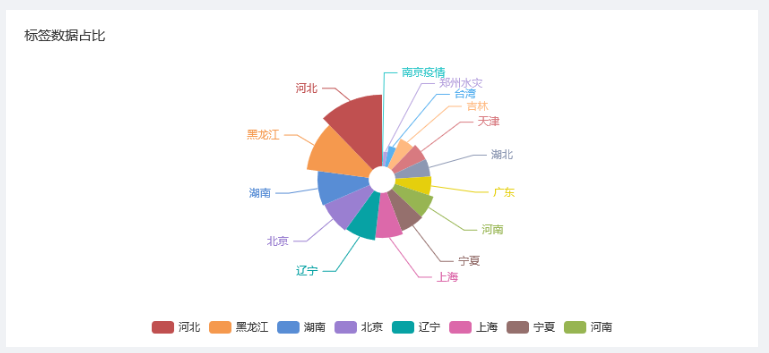
首页第三栏为文本数据变化折线图/分钟，展示每分钟数据的增量。



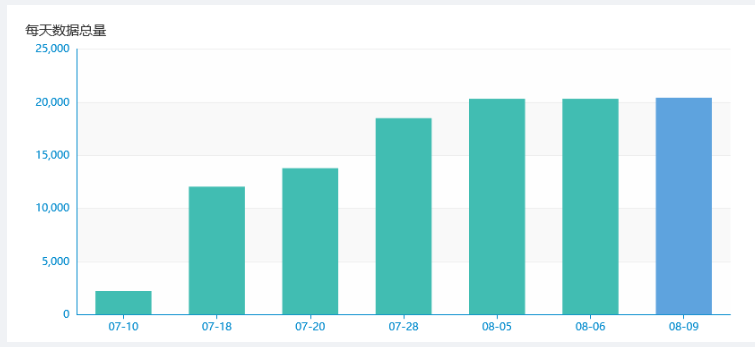
首页第四栏的第一部分为历史任务，显示当前创建过的所有任务ID、结束时间以及状态。



首页第四栏的第二部分为标签数据占比（饼状图），可以清晰的查看每个标签在数据中的占比。



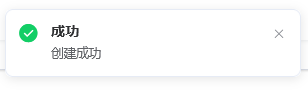
首页第四栏的第三部分为数据总量/天（柱状图），展示每天的数据总量。



#### 创建文本任务

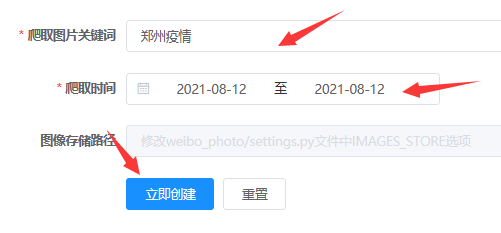
最少输入一个关键词作为标签1，程序以标签1为搜索关键词，爬取时间也必须输入。根据提示输入好信息后，点击立即创建，稍等一分钟右上角会提示创建成功，即可在数据整编和首页中看到新增数据和状态改变。





#### 创建图像任务

输入一个关键词和时间，即可开始图像程序，成功后会在右上角提示创建成功。保存的图片在自行设置的路径下。



#### 创建视频任务

支持多个关键词爬取视频，最少输入一个关键词，可以点击增加关键词按钮增加关键词，成功后会在右上角提示创建成功。开始程序后视频保存在自行设置的路径下。



#### 当前爬虫任务

展示正在爬取数据的任务和已完成的爬虫任务，如果想自行停止爬虫任务，可以点击正在运行任务后的取消按钮，取消成功后会在右上角提示取消成功，相关信息也会在一分钟后在首页刷新。



