## 搜狗微信爬虫

获取搜索关键字分别为微星，机械革命，炫龙，机械师，神州五个笔记本品牌的前10页微信文章信息，字段包含（标题，文章链接，简介内容，来源，发布时间【时间格式为yyyy-mm-dd，年份月份分期用-连接】），信息保存在MongoDB中。

主页地址：<https://weixin.sogou.com/>



## 电视猫爬虫

1. 剧情爬虫

爬取电视剧《冰糖炖雪梨》剧集的标题，剧情

<https://www.tvmao.com/drama/aG0gXGRh/episode>

数据存储mysql数据库

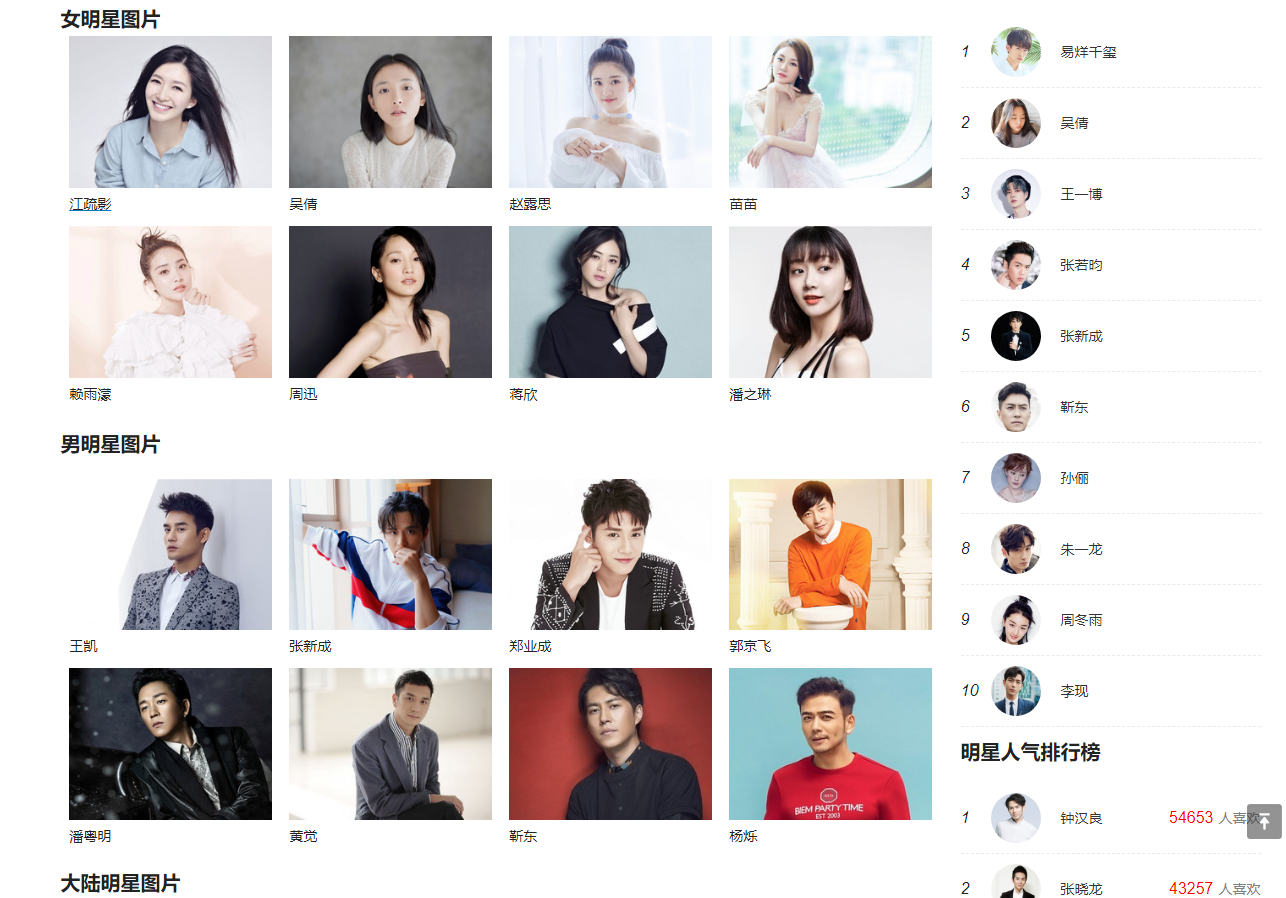


1. 明星库爬虫

<https://www.tvmao.com/star>

爬取所有明星姓名、英文名、血型、身高、国家地区、身高、体重、生日、毕业院校、出生地、星座、简介、电视剧作品、电影作品

数据存储mysql数据库





## 全国城市天气爬虫

**入口地址：**http://www.weather.com.cn/weather1d/101180101.shtml

自动爬取，提取**省份、城市、天气情况、最高气温、最低气温、日期**

数据存储excel 格式：weather.xls







## 拉勾网

<https://www.lagou.com/zhaopin/Python/?labelWords=label>

爬取python职位相关信息：

标题，链接，发布时间，薪资待遇，学历，公司名称，福利待遇，公司说明

爬取10页

数据存储csv格式



## 全国企业破产重组信息网

<http://pccz.court.gov.cn/pcajxxw/gkaj/gkaj?lx=999>

提取标题，链接，案件类型，日期，经办法院，被申请人，申请人



爬取10页，存储为excel文件

## 思诺爱普软件著作权爬虫（备用）

1. 思诺爱普软件著作权爬虫，获取第一层第一页前10条，对应的第二层前10页的计算机软著信息，第二层每一页20条，相当于10\*10\*20=2000条数据。字段包含（第一层的日期，第二层的公司名，软件名，登记号【即用红框括起来的数据】），信息保存在MySQL中。

主页地址：<http://www.ipyear.cn/copyright/index.html>

第一层：



第二层：

