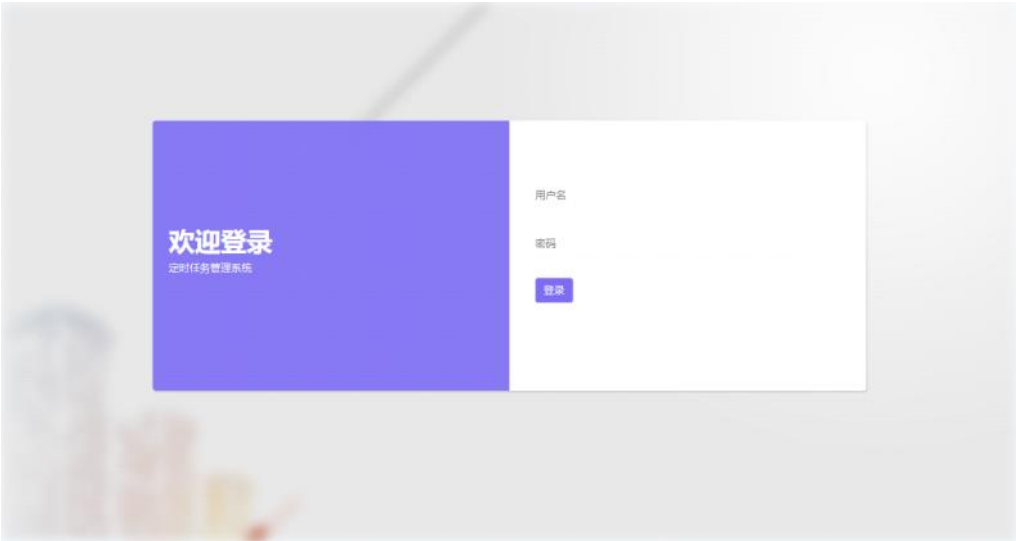


# 定时任务管理系统-操作说明

2020年12月16日 14:39

## 一、前端介绍

1、登录界面，用户创建及修改等操作请参见后台管理相关内容。



## 2、主界面



在主界面上可以看到已经配置的任务。

## 3、双击任务，可以查看任务详情



4、启动、关闭任务

默认新创建的任务及后台重启后，任务状态均为停止状态，需要在主界面中，对相应的任务点击“启用”按钮



二、后台管理

1、后台管理员账号及密码

默认后台管理员为：admin 密码为：1!qQ1111

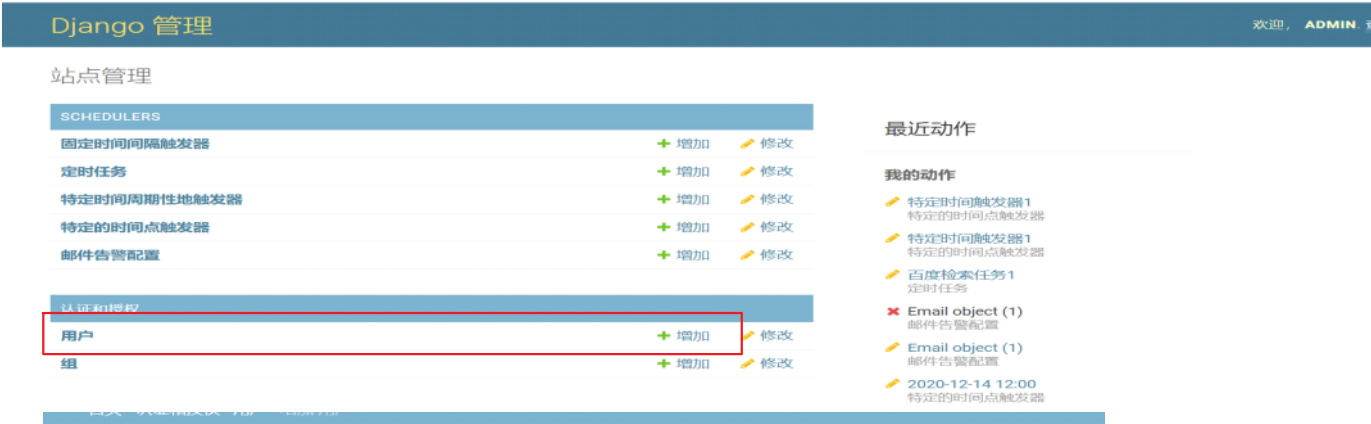
登录系统后台后，界面显示如下：



修改密码可点击右上角的“修改密码”完成

2、创建新用户

点击左侧“用户”行的“增加”按钮，并按提示完成用户信息及权限的设置。



SCHEDULERS	
固定时间间隔触发器	+ 增加
定时任务	+ 增加
特定时间周期性地触发器	+ 增加
特定的时间点触发器	+ 增加
邮件告警配置	+ 增加
认证和授权	
用户	+ 增加
组	+ 增加

### 增加 用户

输入用户名和密码后，你将能够编辑更多的用户选项。

用户名:

必填。150个字符或者更少。包含字母，数字和仅有的@/./+/-。

密码:

你的密码不能与你的其他个人信息太相似。  
你的密码必须包含至少 8 个字符。  
你的密码不能是一个常见密码。  
你的密码不能全都是数字。

密码确认:

为了校验，请输入与上面相同的密码。

### 3、配置“邮件告警配置”（只能有一条记录）

如果没有添加过，则点击“添加”，如果已有添加过，仅能做修改。

### 增加 邮件告警配置

邮箱服务器地址:

发送方邮件地址:

发送方邮件账户:

发送方邮件密码:

接受方邮件地址(多个用|分开):

邮箱服务器地址：为发送服务器地址，以163邮箱为例：smtp.163.com

发送方邮件地址：xxx@163.com

发送方账户：为邮箱的用户名

发送方邮箱密码：邮箱登录密码

接受方邮件地址（多个用|分开。该符号为shift+\\）：123@163.com|231@163.com.....

### 4、设置触发器

触发器类型共有三种，分别为：特定的时间点触发器、固定时间间隔触发器和特定时间周期性地触发器。区别如下：

- 特定的时间点触发器**：指定为了某个时间点执行，只能执行一次
- 固定时间间隔触发器**：指定与当前时间或与指定的开始时间 间隔一段时间执行。只执行一次  
例：如果与当前时间间隔半小时执行，则只需在“间隔分钟数”处输入30即可（**不支持小数点**）  
如果为16日下午3点起计算，半小时后执行，则在“间隔分钟数”处输入30；在“开始时间”输入16日和15：00
- 特定时间周期性地触发器**：依据特定时间间隔重复执行  
例：在每天22和23点的25分，运行一次 job 方法；则：小时字段输入：22-23；分钟数字段输入：25即可  
工作日周一、周三，23点执行；则：工作日字段输入0，2；小时字段输入23 即可

5、绑定任务

注：一个触发器只能和一条任务绑定，不能重复绑定。绑定任务后，默认不启动，请至前端点击启动

1) 通用单页面定时爬虫任务：是指需爬取的内容在遍历的网页内已有，无需获取链接后再次深度爬取。

增加 1、通用单页面定时爬虫任务

任务名称:	<input type="text"/>	用来区别各个任务
网址:	<input type="text"/>	需要爬取数据的网址
关键字(多个用   分开):	<input type="text"/>	需关心的关键字
<input type="checkbox"/> 是否为懒加载网站		默认即可，无需关心
存放招标信息的关键词(多个用   分开):	<input type="text"/>	
存放详细网页网址的关键词:	<input type="text"/>	
触发器类型:	<div><div></div></div>	依据任务执行周期，选择触发器类型
特定的时间点触发器	<div><div></div></div>	
固定时间间隔触发器	<div><div></div></div>	根据所选触发器类型，选择触发器。
特定时间周期性地触发器	<div><div></div></div>	三种触发器只能选用一种
当前状态:	<div><div>停用</div></div>	当前任务的状态，默认即可，调整可在前端完成
接受方邮件地址(多个用   分开):	<input type="text"/>	配置任务获取数据后，接受数据的邮箱。
<input type="checkbox"/> 是否需要登录		
登录地址:	<input type="text"/>	特殊网站需要登陆后才能查询数据，
用户名输入框的XPath:	<input type="text"/>	可通过配置登录信息，来模拟实现自动登录
登录用户名:	<input type="text"/>	XPath可以通过浏览器的开发人员工具进行获取。
密码输入框的XPath:	<input type="text"/>	
登录密码:	<input type="text"/>	
提交按钮的XPath:	<input type="text"/>	
单页爬取/多页爬取:	<div><div>单页爬取</div></div>	
下一页按钮的名称:	<input type="text"/>	如有要爬取的数据有多页，
爬取页数:	<input type="text"/>	可以选择多页爬取，默认为单页爬取
备注:	<div><div></div></div>	

2) 二级页面定时爬虫任务：部分网站需通过获取详细页地址后才能爬取数据，此时就需要适用二级页面定时爬虫任务。

增加 2、二级页面定时爬虫任务

任务名称:	<input type="text"/>	用来区别各个任务
网址:	<input type="text"/>	列表界面网址
关键字(多个用   分开):	<input type="text"/>	需关心的关键字

网址:	<input type="text"/>	需关心的关键字
关键字(多个用   分开):	<input type="text"/>	获取详细页面网址的正则表达式
正则表达式模式:	<input type="text"/>	拼接详细页面网址时需要用到的网站根网址
根网址:	<input type="text"/>	
触发器类型:	<input type="text"/>	依据任务执行周期, 选择触发器类型
特定的时间点触发器:	<input type="text"/>	根据所选触发器类型, 选择触发器。 三种触发器只能选用一种
固定时间间隔触发器:	<input type="text"/>	
特定时间周期性地触发器:	<input type="text"/>	
任务id:	<input type="text"/>	默认即可, 无需关心
当前状态:	<input type="text"/>	当前任务的状态, 默认即可, 调整可在前端完成
接受方邮件地址(多个用   分开):	<input type="text"/>	配置任务获取数据后, 接受数据的邮箱。
<input type="checkbox"/> 是否需要登录		特殊网站需要登录后才能查询数据, 可通过配置登录信息, 来模拟实现自动登录 XPath可以通过浏览器的开发人员工具进行获取。
登录地址:	<input type="text"/>	
用户名输入框的XPath:	<input type="text"/>	
登录用户名:	<input type="text"/>	
密码输入框的XPath:	<input type="text"/>	
登录密码:	<input type="text"/>	
提交按钮的XPath:	<input type="text"/>	
单页爬取/多页爬取:	<input type="text"/>	如有要爬取的数据有多页, 可以选择多页爬取, 默认为单页爬取
下一页按钮的名称:	<input type="text"/>	
爬取页数:	<input type="text"/>	
备注:	<input type="text"/>	

## 6、设置举例

1) 如果需要定期查询全国中高风险地区中是否有指定地区信息时, 可如下操作

网站: [https://m.thepaper.cn/baijiahao\\_16036962](https://m.thepaper.cn/baijiahao_16036962)

关键字: 上虞区

周期: 周一、三每日10点

由于每周的一、三都要执行一次, 所以选择“特定时间周期性地触发器”

触发器配置如下:

触发器名称:	周一、三每日10点
年份(2020):	
月份(1-12):	
日期(1-31):	
周数(1-53):	
工作日(0-6); 0为周一:	0,2
小时(0-23):	10
分钟(0-59):	
秒数(0-59):	

由于在[https://m.thepaper.cn/baijiahao\\_16036962](https://m.thepaper.cn/baijiahao_16036962)网页中已近有我们需要查询的数据了，无需深度爬虫，所以使用“通用单页面定时爬虫任务”。只需配置如下内容即可。

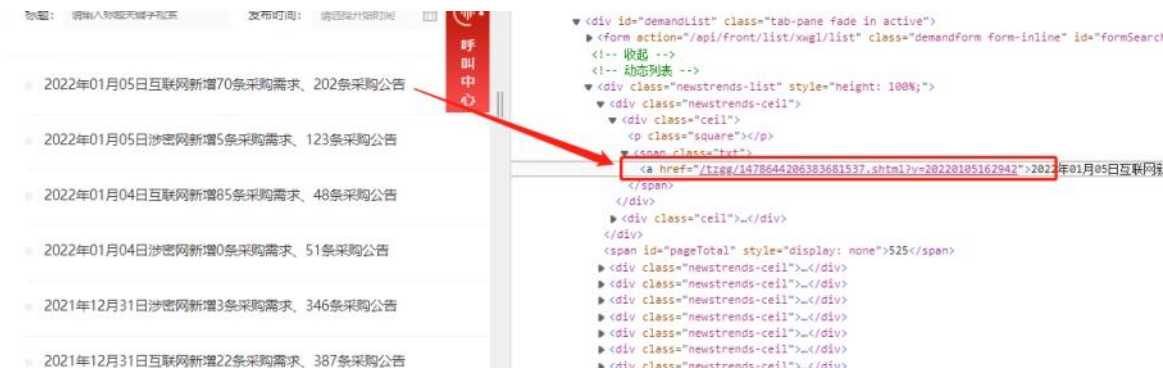
任务名称:	全国中国风险区查询
网址:	<a href="https://m.thepaper.cn/baijiahao_16036962">https://m.thepaper.cn/baijiahao_16036962</a>
关键字(多个用   分开):	上虞区
<input type="checkbox"/> 是否为懒加载网站	
存放招标信息的关键字(多个用   分开):	
存放详细页网址的关键字:	
触发器类型:	特定时间周期性地触发器 ▼
特定的时间点触发器:	----- ▼ ✎ +
固定时间间隔触发器:	----- ▼ ✎ +
特定时间周期性地触发器:	周一、三每日10点 ▼ ✎ +
当前状态:	启用 ▼
接受方邮件地址(多个用   分开):	zhangsan@163.com

2) 爬取 全军武器装备采购信息网上的招标信息。由于每个招标信息都有单独的详细页面，所以适用“二级页面定时爬虫任务”

**网址:** <https://weain.mil.cn/tzgg/list.shtml> 招标信息的列表界面

**关键字:** 核|应急|机房

通过分析网页，可以发现，每条列表内容对应一个详细页面的地址



适用正则表达式，获取这个地址

**正则表达式为：**href="/tzgg/[0-9].\*")>

点击某条信息后，实际的详细页面网址为：

https://weain.mil.cn/tzgg/1478644206383681537.shtml?v=20220105162942

所以需要对上步获取的地址进行一个拼接过程。

**根网址为：** https://weain.mil.cn

修改 2、二级页面定时爬虫任务

任务名称:	采购信息定时任务
网址:	https://weain.mil.cn/tzgg/list.shtml
关键字(多个用   分开):	核 应急 机房
正则表达式模式:	href="/tzgg/[0-9].*")>
根网址:	https://weain.mil.cn
触发器类型:	特定时间周期性触发器
特定的时间点触发器:	-----
固定时间间隔触发器:	-----
特定时间周期性触发器:	周一、三每日10点
任务id:	job_id_3_two
当前状态:	启用
接受方邮件地址(多个用   分开):	12345679@qq.com
<input type="checkbox"/> 是否需要登录	