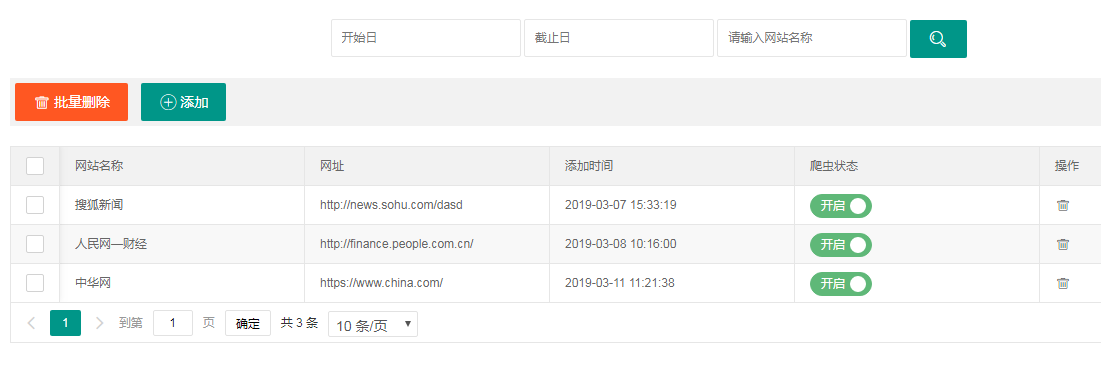
# 爬虫后台管理系统功能文档

## 一 网址管理

### 1.1在采集网址管理



#### 1.11网址列表

显示网站名称，网站域名，添加时间，采集时间，采集状态，该网站今日被采集信息量，在采集网址数量，网址更新频率，所属爬虫引擎等信息。

#### 1.12日期查询

根据开始采集的起止时间查询在此时间范围内开始采集的网站，显示列表信息，或输入网站名称查询网站信息。

可以对正在采集的网站进行关闭处理，关闭之后的网址会被剔除到被采集网站列表之外，可以根据网站被开启采集的时间段或网站名称查询某个或某些网站信息，对网站状态进行处理。可以批量关闭一些网站。

### **1.2待采集网址管理**

查询未被开启采集的网站信息，显示网站名称、网站域名、停用时间、采集状态以及一些操作功能。可以对网站进行启动处理，启动后的网站被添加到启用网址列表，开始采集，可以批量启动网址采集功能，可以添加新网站（针对通用爬虫），批量导入网站（针对通用爬虫），可以删除网站信息，可以根据网站被停止采集的时间段或网站名称查询某个或某些网站信息，对网站状态进行处理。

### **1.3问题网址管理**

查询在被采集过程中出现异常的网站，所在爬虫服务器，异常信息，发生异常的时间，异常是否解决等等信息，有提交反馈信息功能，以便运维人员后期维护，查询某时间段发生异常的信息，批量提交等等。

## **二 爬虫管理**

### **2.1在启动爬虫列表**

查询服务器中正在运行的爬虫程序，显示该爬虫的类型（通用爬虫，定向爬虫，高效率实时爬虫等），爬虫所采集的网站数量，采集的数据量，运行时间，采集效率等信息，可暂停/批量暂停爬虫（暂停的爬虫保持先前的配置信息）。

### **2.2待启动爬虫列表**

查询服务器中未启动的爬虫程序，科添加新的爬虫任务，对未启动的爬虫程序做配置处理，如定时启动爬虫，为爬虫程序批量导入需要采集的网站，采集深度，采集量等信息，可一键格式化爬虫配置信息。

## **三 数据统计**

### **3.1服务器监控**

### **3.2爬虫监控**

## **四 权限管理**

### **4.1成员管理列表**

管理员创建账号登录密码可供其他成员登录，没有注册功能，成员对爬虫，网站的操作信息都会展示在权限管理模块中（这一块还在构思中）

## **五 开发设计**

### **5.1 mysql数据库**

#### **4.11 mysql5**

### **5.1 mysql数据库**

#### **5.11mongodb**