信息资讯通用采集插件

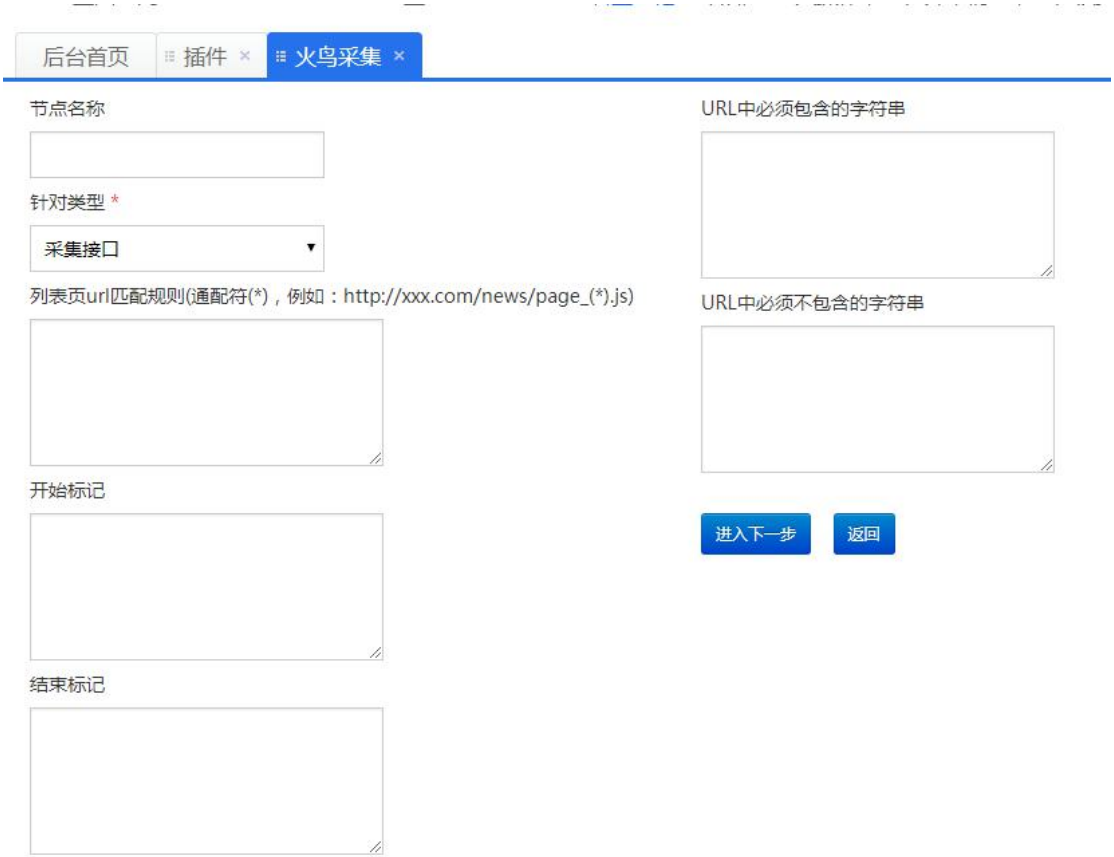
（使用手册）

[Huoniao.Plugins]

苏州酷曼软件技术有限公司
地址：苏州园区领汇商务广场 1 幢 2503 室　电话：0512-67581578　www.ikuman.cn

一、添加任务

点击添加任务之后出现以下界面：



1. 节点名称可以自己根据要采集的内容进行填写

2. 针对类型有三种
   采集单个界面：就是采集单独的一个 HTML 页面
   采集多个界面：
   例如：
   https://ihuoniao.cn/sz/article/yule/?page=1
   https://ihuoniao.cn/sz/article/yule/?page=2
   https://ihuoniao.cn/sz/article/yule/?page=(*)
   所以类似这种可以匹配页码规则的都为此类型。

采集接口类型：

以下例子为网易新闻的列表页面，数据是通过

http://temp.163.com/special/00804KVA/cm_guonei_03.js?callback=data_callback

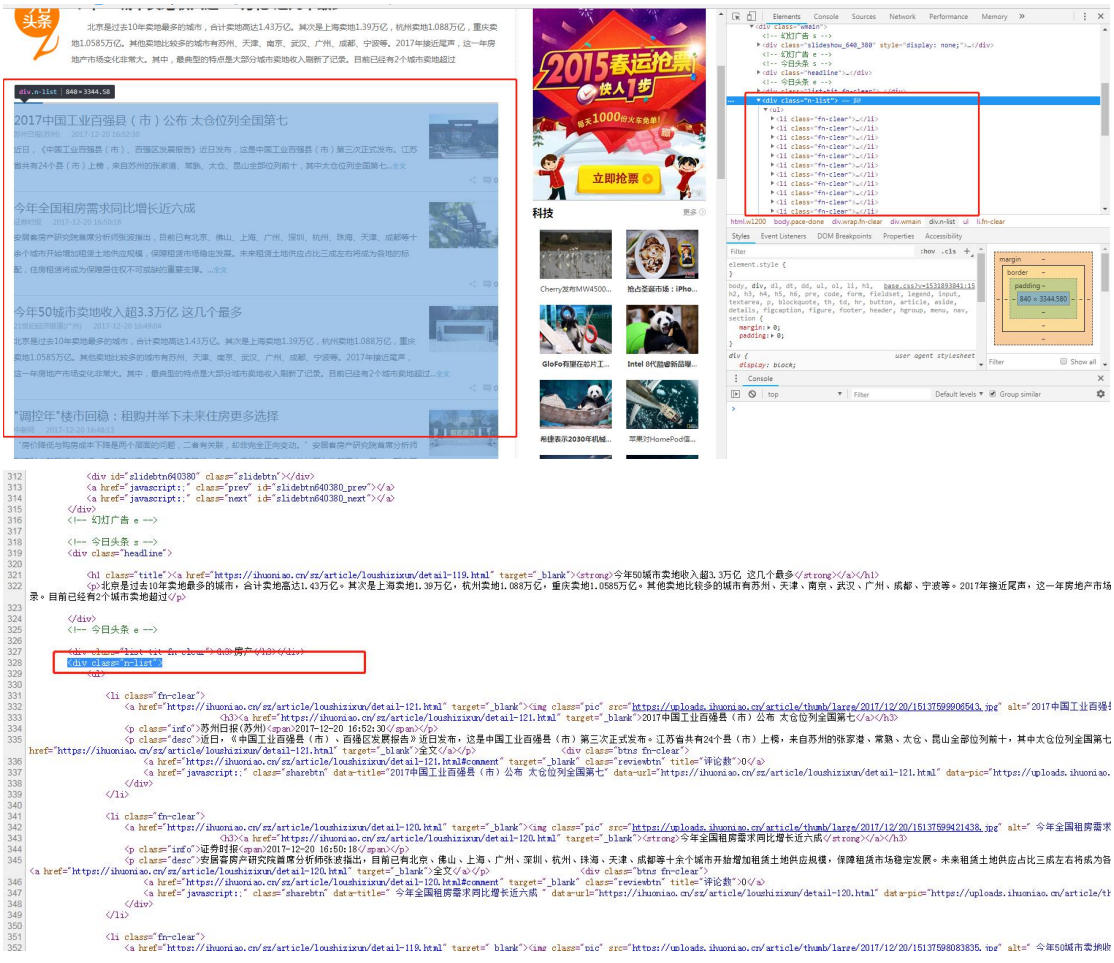
http://temp.163.com/special/00804KVA/cm_guonei_02.js?callback=data_callback

等接口获取的。若没有出现类似的接口，则说明数据与页面是同步的，则选择针对类型为网页 HTML

3. 第三个输入框根据选择的针对类型填写通配符统一填写 (*)

4. 开始标记为页面列表部分 div 的开始。

采集网页 HTML 类型：

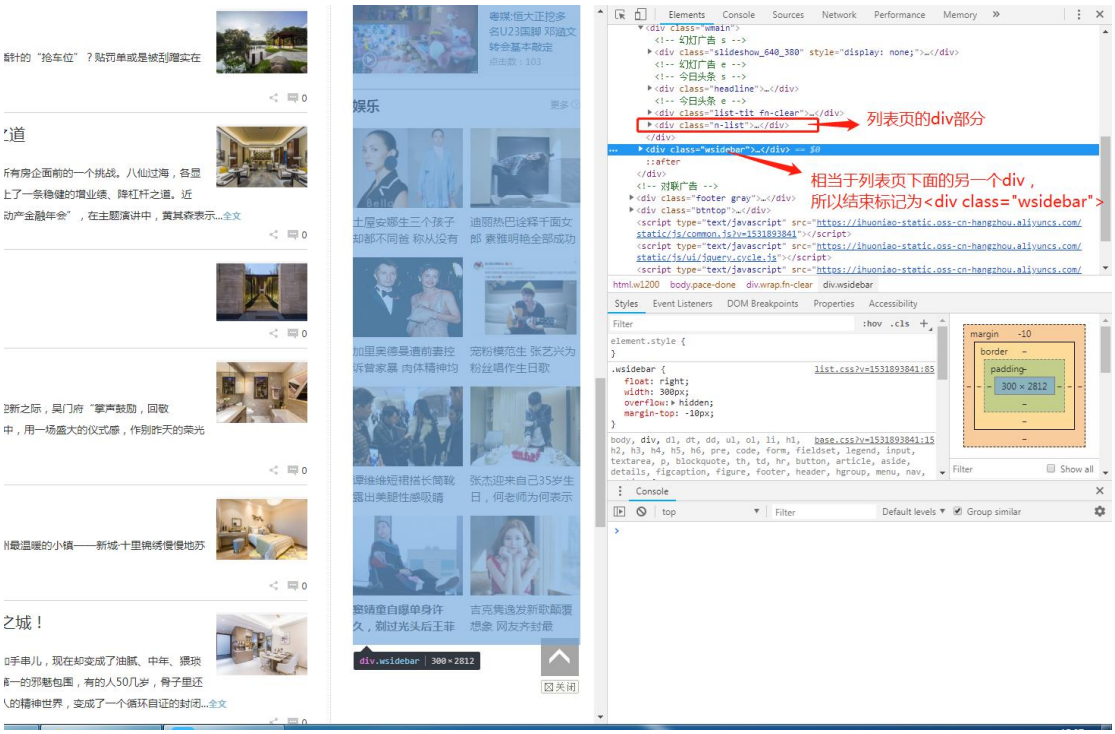



```
<div id="slidebtn640380" class="slidebtn"></div>
<a href="javascript:;" class="prev" id="slidebtn640380_prev"></a>
<a href="javascript:;" class="next" id="slidebtn640380_next"></a>
</div>
<!-- 幻灯广告 e -->

<!-- 今日头条 s -->
<div class="headline">

<h1 class="title"><a href="https://ihuoniao.cn/sz/article/loushixinxun/detail-119.html" target="_blank"><strong>今年50城市卖地收入超3.3万亿 这几个最多</strong></a></h1>
<p>北京是过去10年卖地最多的城市，合计卖地高达1.43万亿。其次是上海卖地1.39万亿，杭州卖地1.088万亿，重庆卖地1.0585万亿。其他卖地比较多的城市有苏州、天津、南京、武汉、广州、成都、宁波等。2017年接近尾声，这一年房地产市场录。目前已经有2个城市卖地超过</p>

</div>
<!-- 今日头条 e -->

<div class="n-list">

<li class="fn-clear">
<a href="https://ihuoniao.cn/sz/article/loushixinxun/detail-121.html" target="_blank"><img class="pic" src="https://uploads.ihuoniao.cn/article/thumb/large/2017/12/20/151375999065
<h3><a href="https://ihuoniao.cn/sz/article/loushixinxun/detail-121.html" target="_blank">2017中国工业百强县（市）公布 太仓位列全国第七</a></h3>
<p class="info">苏州日报(苏州)<span>2017-12-20 16:52:30</span></p>
<p class="desc">近日，《中国工业百强县（市）、百强区发展报告》近日发布，这是中国工业百强县（市）第三次正式发布。江苏省共有24个县（市）上榜，来自苏州的张家港、常熟、太仓、昆山全部位列前十，其中太仓位列全国第七
href="https://ihuoniao.cn/sz/article/loushixinxun/detail-121.html" target="_blank">全文</a></p>
<a href="https://ihuoniao.cn/sz/article/loushixinxun/detail-121.html#comment" target="_blank" class="reviewbtn" title="评论数">0</a>
<a href="javascript:;" class="sharebtn" data-title="2017中国工业百强县（市）公布 太仓位列全国第七" data-url="https://ihuoniao.cn/sz/article/loushixinxun/detail-121.html" data-pic="https://uploads.ihuoniao.
</div>
</li>

<li class="fn-clear">
<a href="https://ihuoniao.cn/sz/article/loushixinxun/detail-120.html" target="_blank"><img class="pic" src="https://uploads.ihuoniao.cn/article/thumb/large/2017/12/20/151375894214
<h3><a href="https://ihuoniao.cn/sz/article/loushixinxun/detail-120.html" target="_blank"><strong>今年全国租房需求同比增长近六成</strong></a></h3>
<p class="info">证券时报<span>2017-12-20 16:50:18</span></p>
<p class="desc">安居客房产研究院首席分析师张波指出，目前已有北京、佛山、上海、广州、深圳、杭州、珠海、天津、成都等十余个城市开始增加租赁土地供应规模，保障租赁市场稳定发展。未来租赁土地供应占比三成左右将成为各
<a href="https://ihuoniao.cn/sz/article/loushixinxun/detail-120.html" target="_blank">全文</a></p>
<a href="https://ihuoniao.cn/sz/article/loushixinxun/detail-120.html#comment" target="_blank" class="reviewbtn" title="评论数">0</a>
<a href="javascript:;" class="sharebtn" data-title="今年全国租房需求同比增长近六成 " data-url="https://ihuoniao.cn/sz/article/loushixinxun/detail-120.html" data-pic="https://uploads.ihuoniao.cn/article/th
</div>
</li>

<li class="fn-clear">
<a href="https://ihuoniao.cn/sz/article/loushixinxun/detail-119.html" target="_blank"><img class="pic" src="https://uploads.ihuoniao.cn/article/thumb/large/2017/12/20/151375980838
```

需要右键查看源代码，复制开始标记：

5. 结束标记

火鸟门户



采集接口类型：

所以开始标记为 data_callback(

结束标记为

6. 必须包含为此列表页包括的所有新闻内页的 url 中共有的部分（只要是公共部分就行，

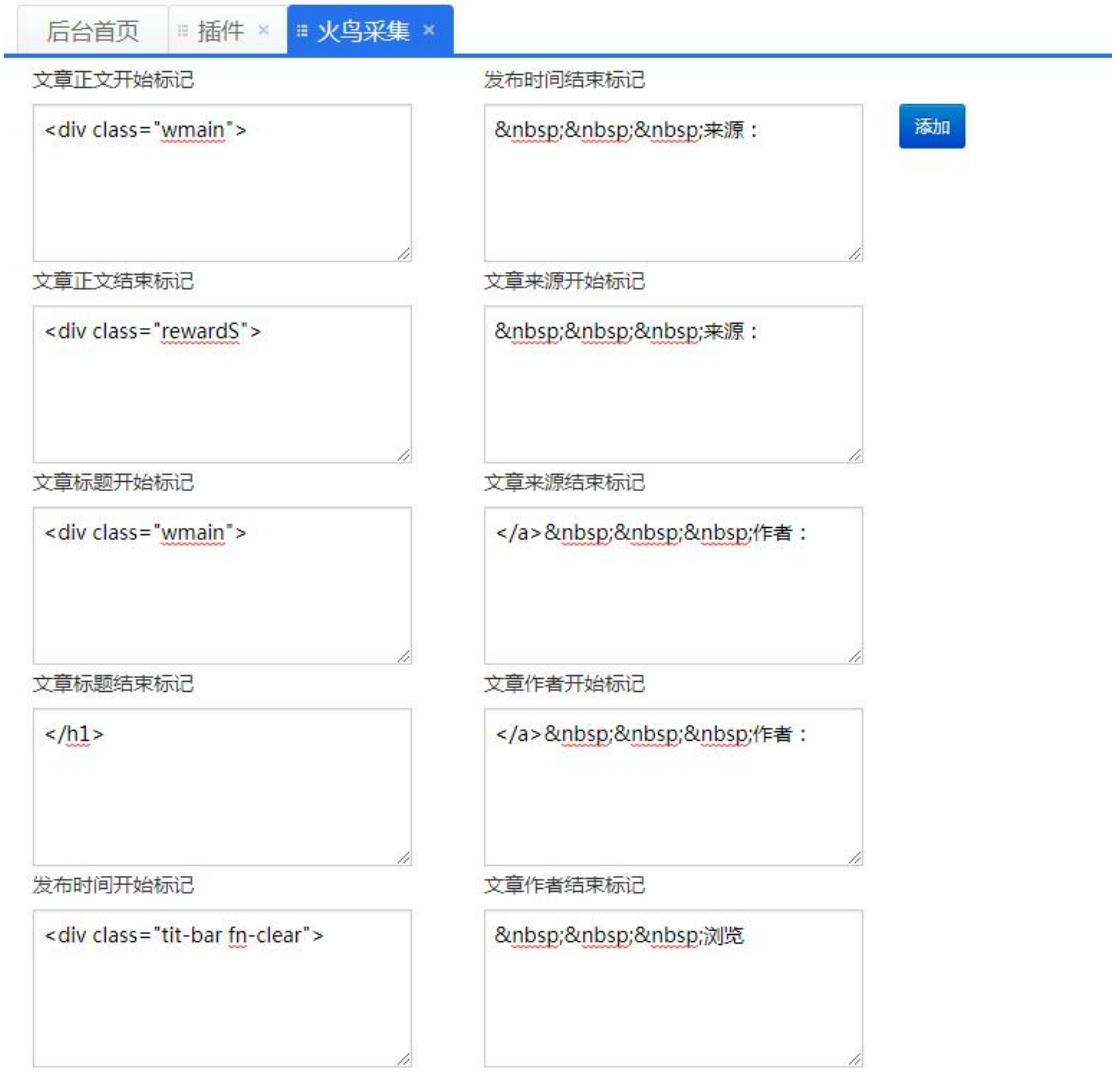
也可直接填写网页后缀）

7. 必须不包含一般为列表页面中的一些图片的链接或者关键词评论的链接，填写之后采集时会过滤这些内容。

最终效果：

二、点击进入下一步添加新闻详细页面的匹配规则

正文开始：
正文结束：
来源开始：
来源结束标记：
以上为举例。所有的结束标记都是匹配到**开始标记之后遇到的第一个结束标记**为止。
最终效果：



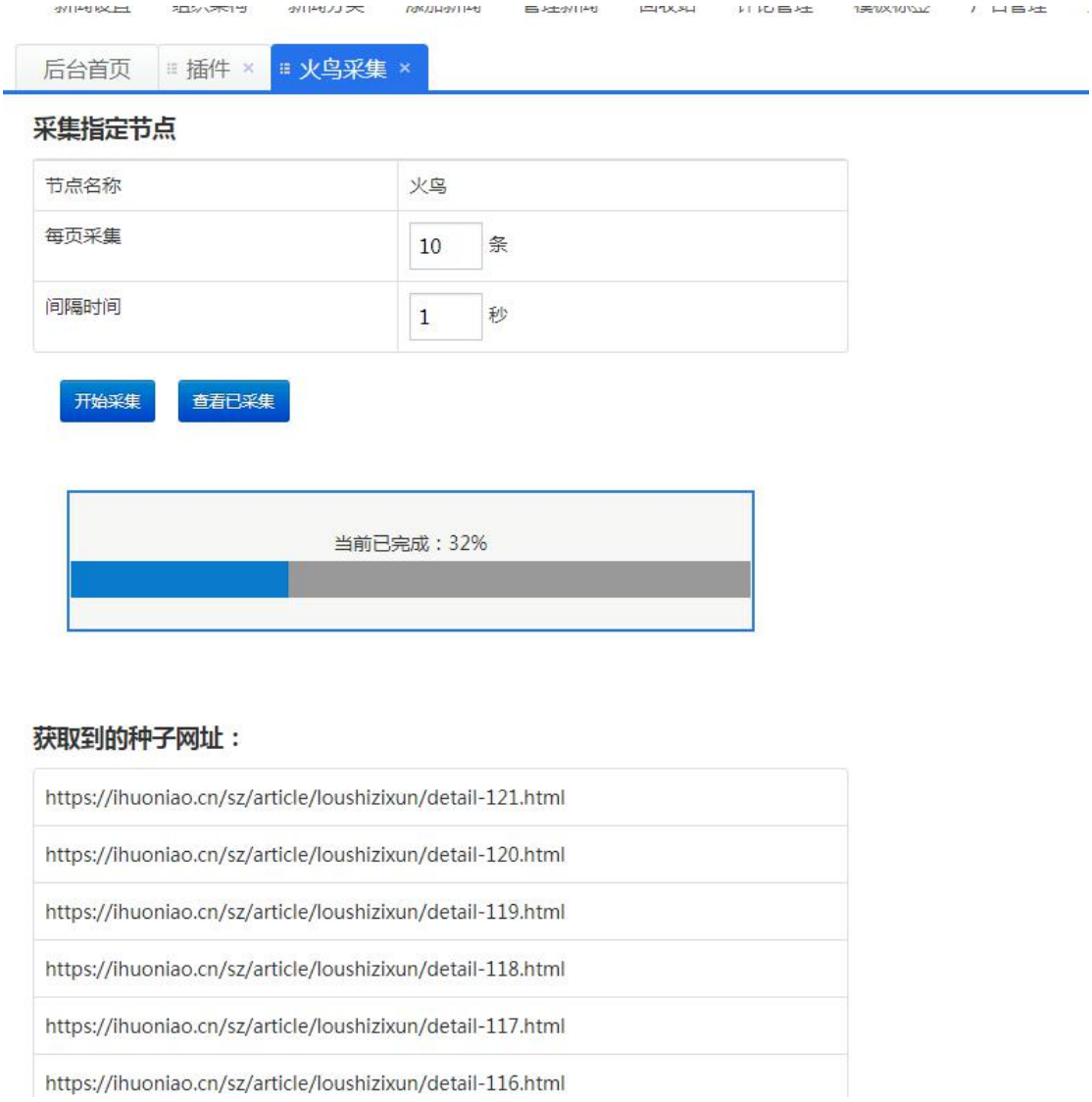
添加添加之后返回主页，则可以看到刚才添加的节点。

三、开始采集

点击采集按钮

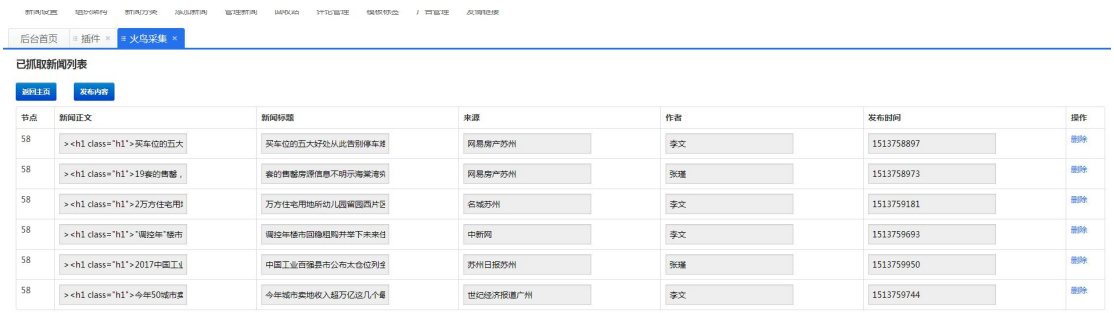
点击下一步进入采集页面。

页面下方为匹配到的新闻链接。
输入每页采集条数和间隔时间
点击开始采集。
以下为最终效果。

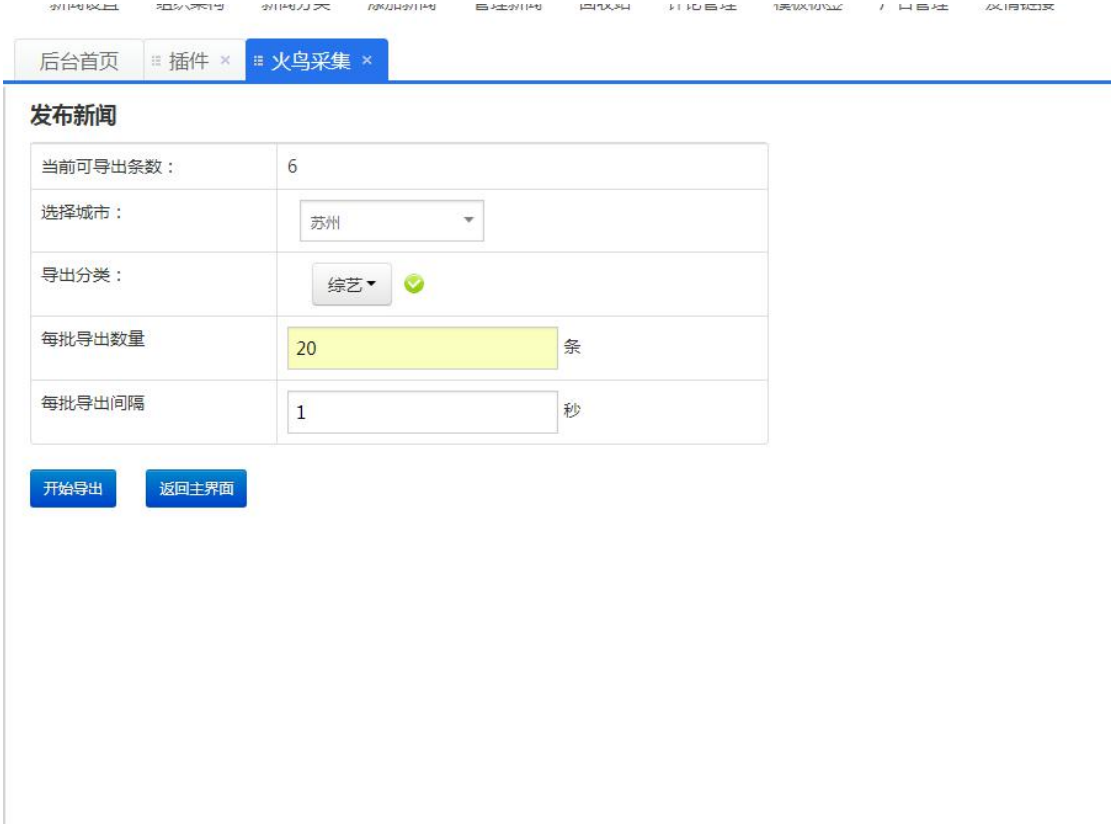


采集完成之后可以点击查看已采集内容：

火鸟门户



发布内容按钮可以把采集到的内容发布到网站的新闻中心。



同样输入每批导出数量和时间间隔。

导出完成之后，则可以在新闻管理界面看到采集的内容。