

抓取网站地址（i 问财）：<https://www.iwencai.com/data-robot/extract-new?>  
页面如下



右侧上部输入表单中键入关键词，表单位置处 html:

```
<div class="left_panel">
  <div class="robot-header">
    <div class="robot-header-top" id="robotHeaderTop">...</div>
    <div class="robot-header-search">
      <div class="questionInputWrap2">
        <input class="robot-search-input" id="questionInput" type="text" maxlength="250" value="" />
        <i class="robot-search-icon" id="inputSubmit"></i>
      </div>
    </div>
    <div class="robot-header-tab">...</div>
  </div>
  <div id="cardList">...</div>
</div>
```

搜索输入框

搜索提交键

输入关键词通过以下方式获得：

(1) 单个关键词：输入“指标.xlsx”或“形态.xlsx”中的数据。

“指标.xlsx”部分数据如下，

财务指标	每股收益	小于0		技术指标	1日主力资金	小于0	
		大于0				大于0	
每股净资产		大于0.3		5日主力资金		大于5000万	
		大于0.5				大于1亿	
		小于0				大于5亿	
		大于0				大于0	
每股未分配利润		大于1元		均线		大于5000万	
		大于3元				大于1亿	
		大于5元				大于5亿	
		大于0元				均线多头排列	
每股资本公积		0-1元		MACD		均线粘合	
		1元-3元				5日线上穿10日线	
		大于3元				20日线支撑	
		小于1元				股价超过250日线	
每股股利		1元-2元				股价上穿250日线	
		2元-5元				MACD金叉	
		大于5元				MACD上升	
		小于0.2元				MACD顶背离	
		0.2元-0.5元				MACD底背离	
		大于0.5元					

输入格式可以为：

6月每股收益大于0，  
2017年一季度每股净资产大于0，  
4月18日MACD金叉等等

即输入格式为：

**[时间][指标][数量关系]**  
其中时间为可选项，数量关系可以选用推荐的关系或自行设置设置。

“形态.xlsx”部分数据如下，

1	红三兵	长阳重炮
2	双重底	尖三兵
3	蜻蜓点水	旭日初升
4	缺口回补	向下跳空缺口
5	三金叉	倒锤线
6	升势三鸦	向上跳空缺口
7	倾盆大雨	长阳不破
8	旭日东升	叠叠多方炮
9	曙光初现	长阳包阴
10	淡友反攻	双阳包阴
11	好友反攻	巨量长阴
12	绝地反击	阴孕十字
13	射击之星	阳孕十字
14	倒拔杨柳	十字孕线
15	出水芙蓉	假阳十字星
16	回踩五日线	底部t形

输入格式与之前类似：

[K 线类型][时间][形态]

如：

周线周期 6 月 20 日假十字星，

2016 年 5 月多方炮

其中时间是可选项，需自行设定；最好针对 X 年，X 年 X 月，X 年 X 月 X 日都有涉及。K 线类型也是可选项，最好日线周期，周线周期，月线周期都有涉及。

(2) 多个关键词：将若干个关键词组合起来，如，

6 月每股收益大于 0 且 2016 年 5 月多方炮

尽可能多的组合不同的关键字（关键字可属于任意类型关键字）。

(3) 网页表单中输入过程中，表单一般会对输入自动联想，可以将联想得到的关键字输入，联想关键字的获取在文档末尾介绍。

输入关键字后通过 search 标签提交表单。

表单提交之后页面显示示例（部分）如下：

收集对应搜索问题的前 15 只股票的信息（股票代码和股票简称），股票信息处 html 代码如下：

需要保存内容 [问题（输入的关键字），[股票 1，股票 2，...，股票 15]，搜索时的时间（YYYY-MM-DD）]。

```
<div class="static_con">
  <div class="thead_table_wrap static_thead_table_wrap" style="left: auto; top: auto; width: 243px;">...</div>
  <table class="static_table tbody_table static_tbody_table" border="0" cellspacing="0" cellpadding="0">
    <tbody>
      <tr class="even_row">
        <td width="56">...</td>
        <td width="28" class="tdCheckbox">...</td>
        <td width="66" class="item" colnum="0">
          <div class="em">603579</div>
        </td>
        <td width="90" class="item" colnum="1">
          <div class="em graph alignCenter graph">
            <a href="/stockpick/search?tid=stockpick&q=stockpick_diag&ts=1&w=603579" target="_blank">荣泰健康</a>
          </div>
        </td>
        <td class="addList"></td>
      </tr>
    </tbody>
  </table>
</div>
```

股票代码

股票名称

注意!!! 联想关键词的获取。

在输入过程中，i 问财提供自动联想功能，如下：



可以任意选取联想的词条作为新的关键字进行下一轮的搜索。为了覆盖更多的概念，进行联想时键入的自串的长度尽可能短，使网站的联想功能可以有更大的联想空间。例如在键入关键字“每股收益增长率”的时候，可以选取输入“每股”后的联想词条，而不必选择输入“每股收益”后的联想词条。

联想词条的 html 位置：



数据爬取中输入的关键词要尽量覆盖所有“指标.xlsx”和“形态.xlsx”中所有的指标和形态，每一种指标或形态至少在爬取数据中出现 10 次，尝试不同的关键字组合，并且为了提高数据的质量，尽可能多的使用联想关键词。