八爪鱼自定义(非模板）数据采集

采集非常方便，我也给本科生试讲过了，可以非常快速的上手！可以应对多种网站，多种网页结构。。。

什么是自定义采集呢？进入八爪鱼界面之后，红圈圈出的就是啦。



下面逐一介绍常见的抓取类型。

1. 单网页列表信息采集

保存网址-> 选中有规律的数据区域块（列表）单击->选中子元素->选中全部->采集以下数据->保存并开始采集

1. 单网页列表详情页信息采集

保存网址-> 单击详情页链接->选中全部->循环单击每个元素->选中要采集的数据单击->采集该元素文本->采集以下数据->保存并开始采集

3.多网页列表信息采集

保存网址-> 点击下一页->循环点击下一页->选中有规律的数据区域块（列表）单击->选中子元素->选中全部->采集以下数据->保存并开始采集

注意：设置Ajax点击翻页 时间间隔

4. Ajax加载技术 网页地址不发生变化，网页内容出现局部变化

5.多网页列表详情页信息采集

保存网址-> 点击下一页->循环点击下一页->单击详情页链接->选中全部->循环单击每个元素->选中要采集的数据单击->采集该元素文本->采集以下数据->保存并开始采集

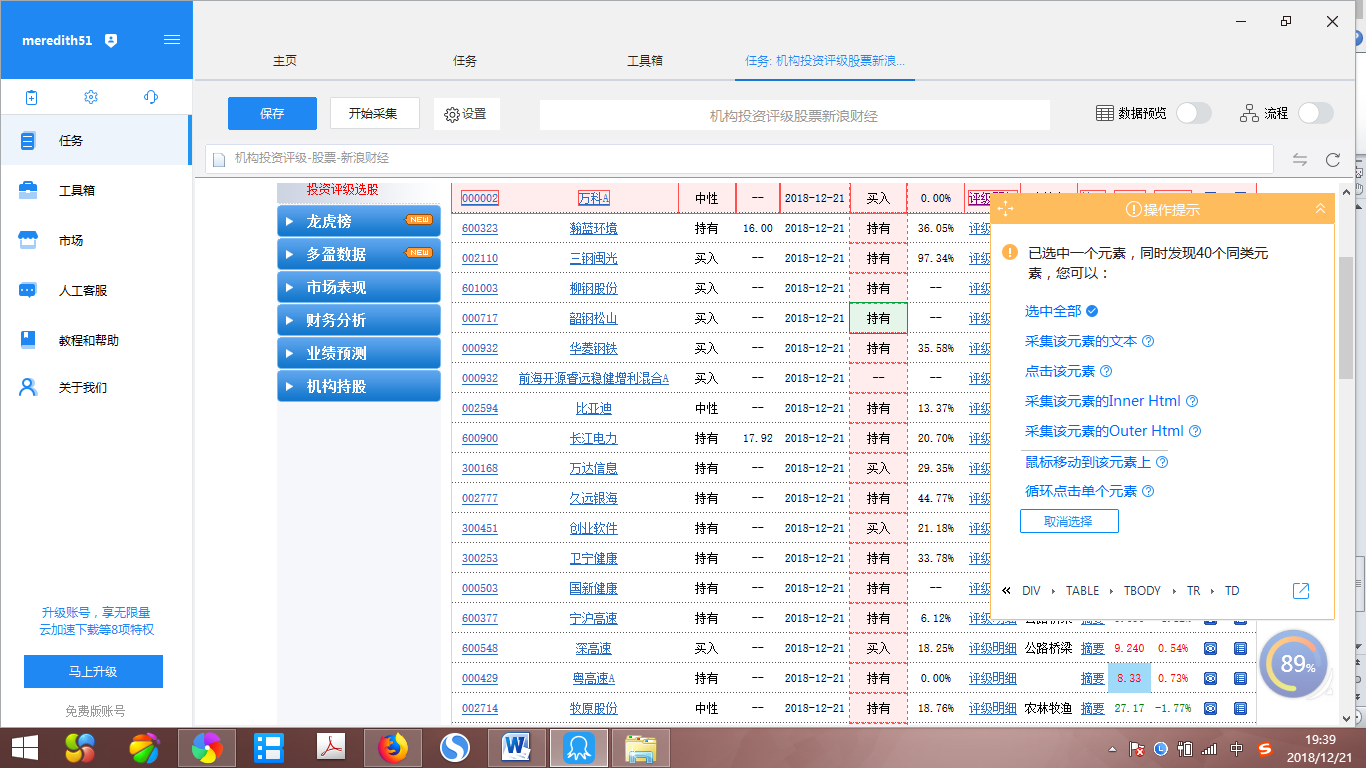
6.正则表达式使用

保存网址-> 点击下一页->循环点击下一页->单击详情页链接->选中全部->循环单击每个元素->选中要采集的数据的整个区域->采集该元素的Inner Html->点选“流程”进入流程操作界面->点选 自定义数据字段 小图标 ->格式化数据->添加步骤->正则表达式匹配->不懂正则？试试正则工具->确定要采集数据的起始位置->生成->匹配->应用->确定->确定。-> 采集数据->保存并开始采集

7.表格数据抓取 例如：同花顺表格数据抓取

<http://vip.stock.finance.sina.com.cn/q/go.php/vIR_CustomSearch/index.phtml>

<http://vip.stock.finance.sina.com.cn/mkt/#cyb_root>

保存网址-> 点击下一页->循环点击下一页->选中要抓取的某个数据，在点选界面选择 扩展区域按钮 ->选中子元素->选中全部->采集以下数据->保存并开始采集。 此图右下角图标为扩展区域按钮

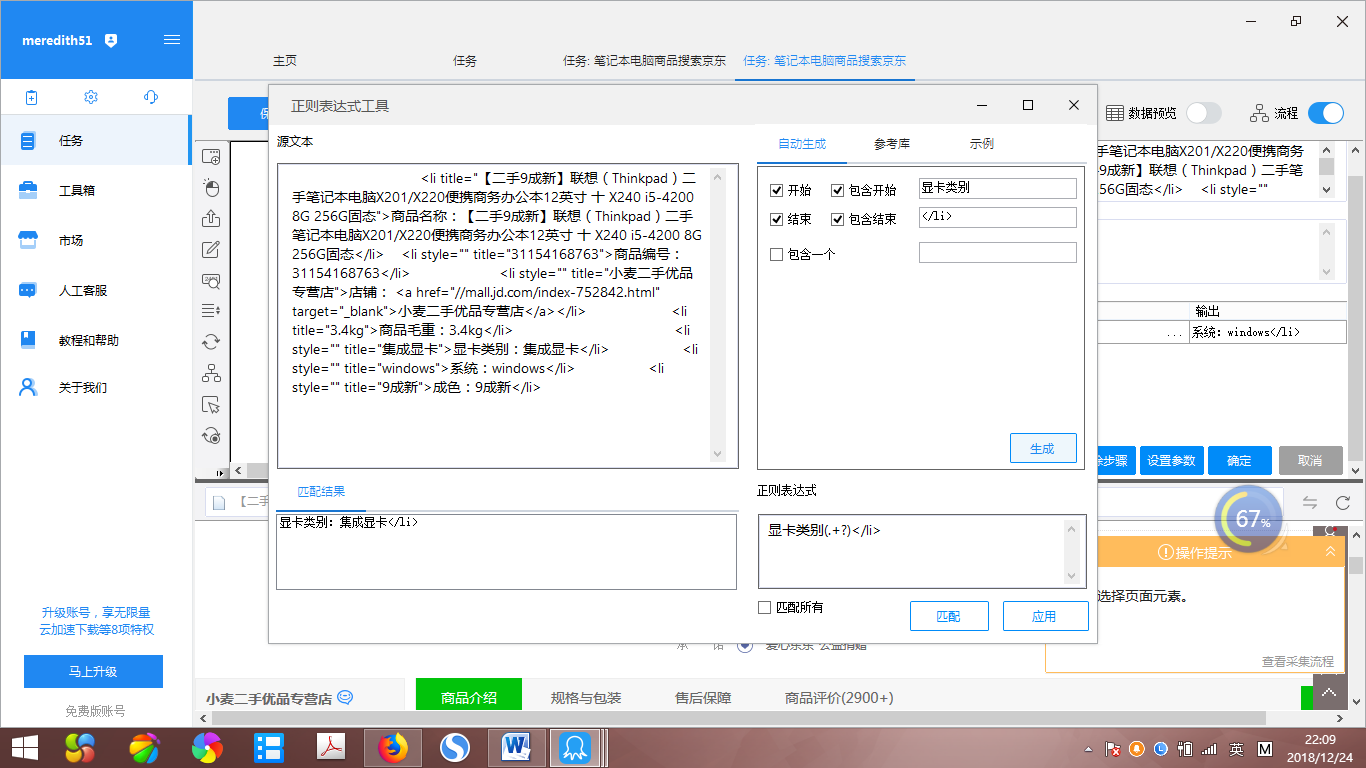
在流程页面 修改变量名称。

自定义数据字段见下面红色圈





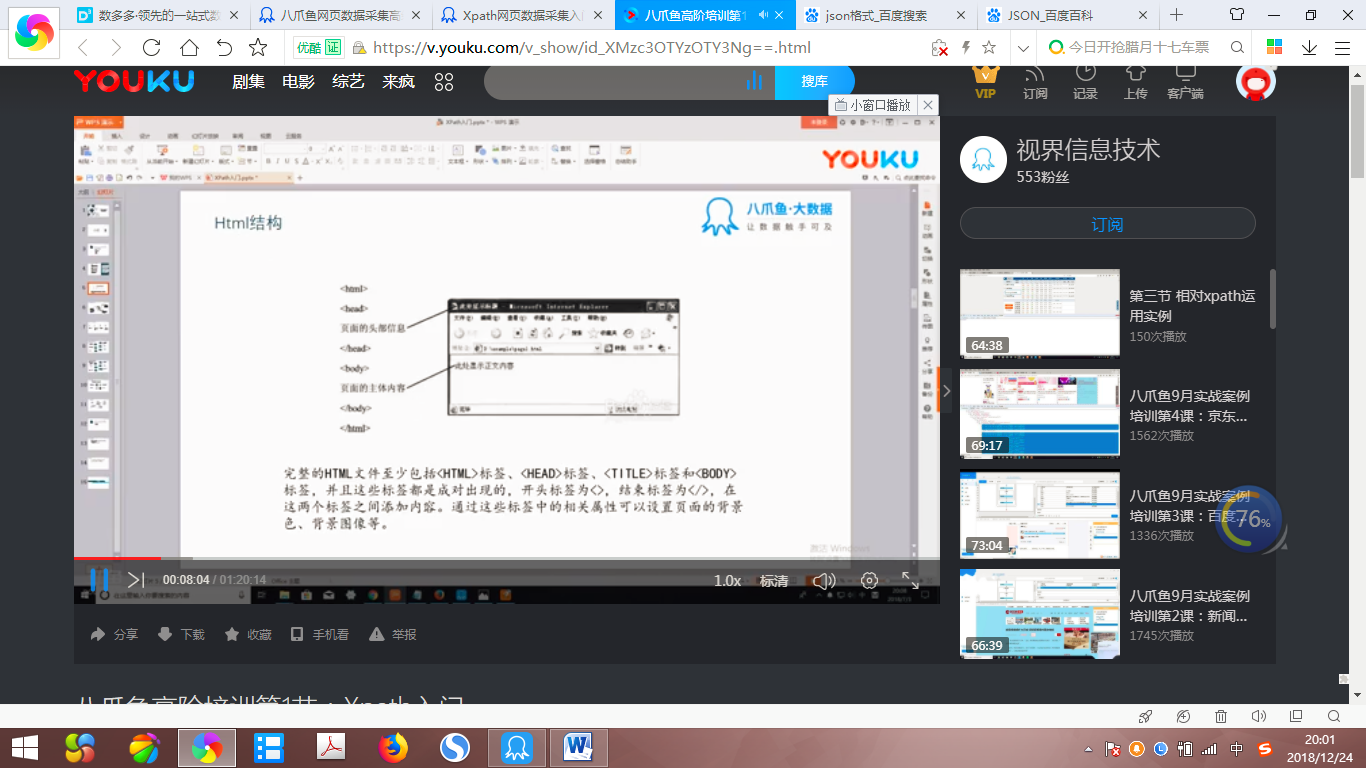


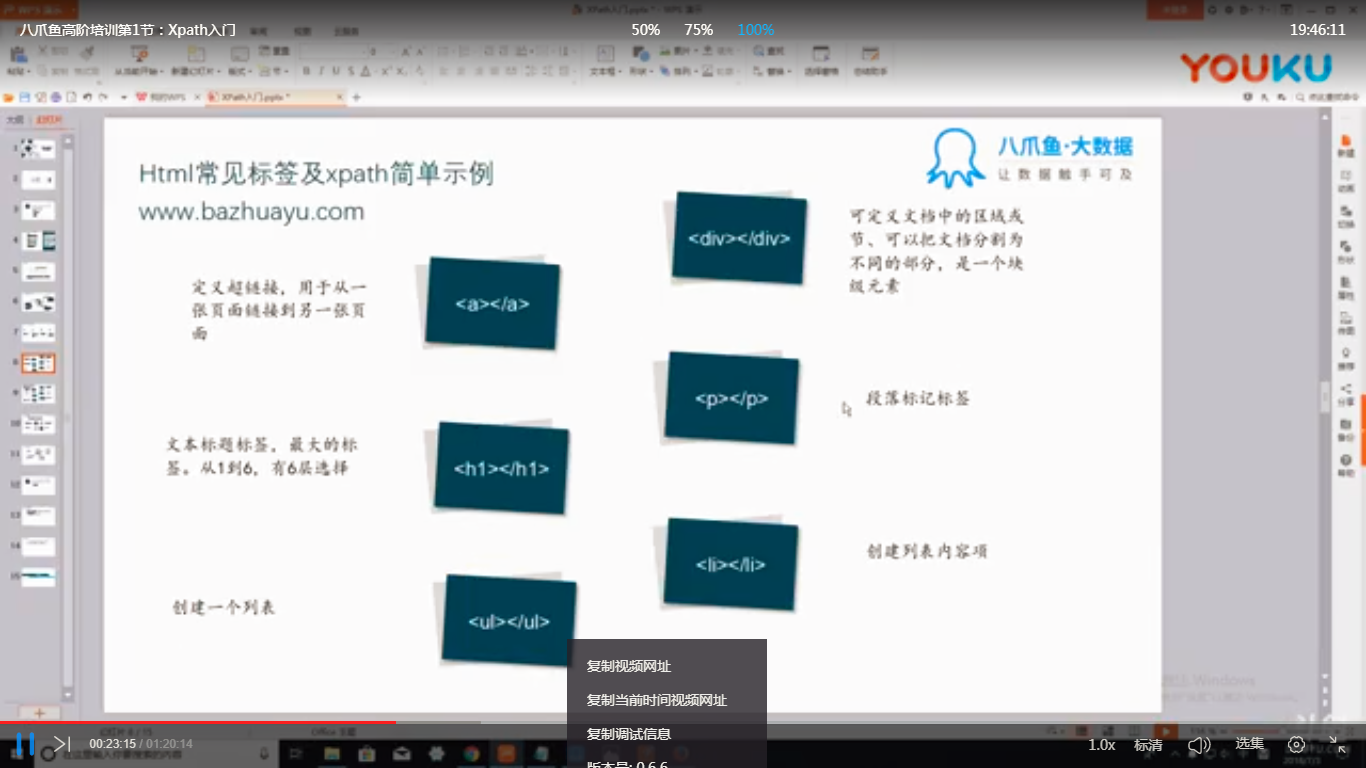


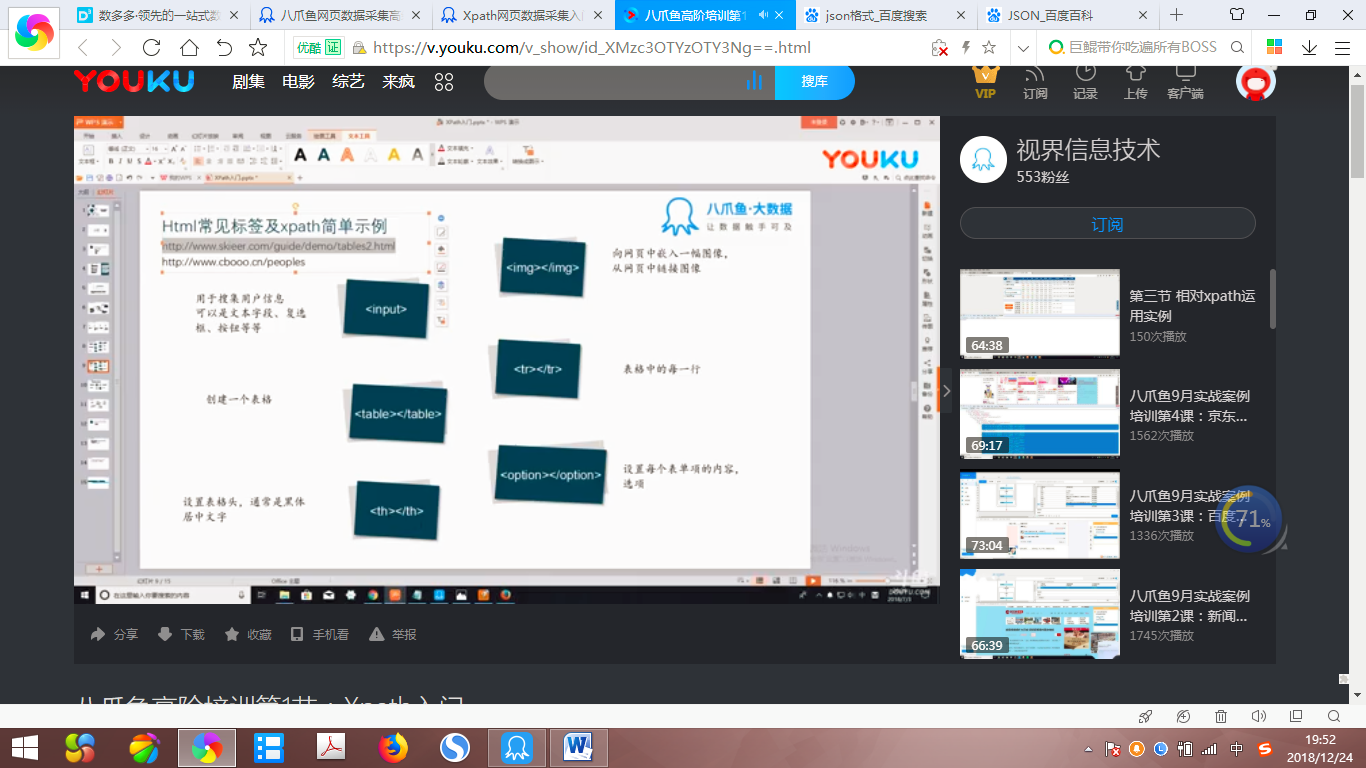
8.地图数据信息采集

保存地址-> 点击搜索文本框->输入文字（如火锅）-> 确定->点击搜索按钮->点击该按钮->点击下一页->循环点击下一页->选中有规律的数据区域块（如列表）单击->选中子元素->选中全部->采集以下数据->保存并开始采集。

9．关于网页结构，我截了以下几个屏幕，是八爪鱼官网的学习视频中介绍的，对于没有学过网页设计的人来说，是很好的入门资料。有兴趣的话可以点击听听。







常见标签：

