

网页中视频 url 采集

建立采集任务

1. 选择一个视频网站进行采集



2. 打开八爪鱼采集器，登录账号



3.创建采集任务，编辑采集任务及任务组名称，点击“下一步”，开始采集规则的设置

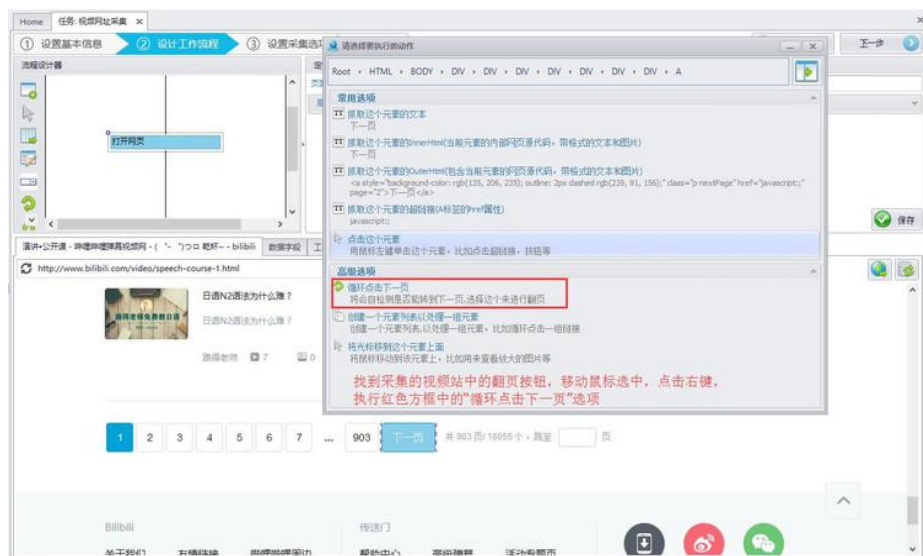


创建翻页循环

1.进入采集规则设置界面，将视频网站的 URL 输入到地址栏，如图点击右侧的按钮，在八爪鱼采集器的内置浏览器中打开该网页



2.找到采集的视频站中的翻页按钮，移动鼠标选中，点击右键，执行红色方框中的“循环点击下一页”选项。操作如图所示



3.完成后上述操作后，左侧的流程设计器中显示出一个翻页循环框，翻页循环创建完成，如图所示

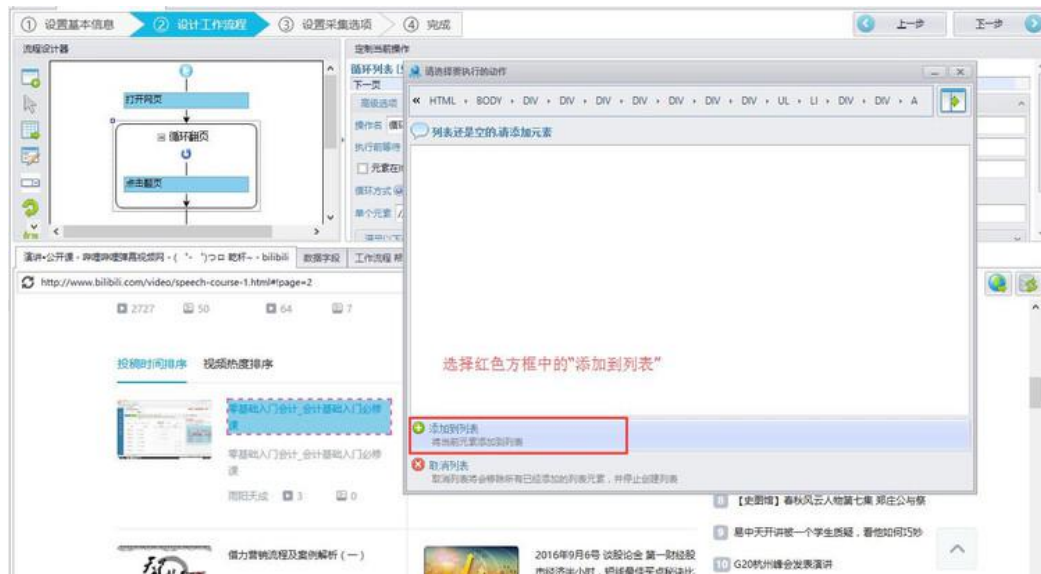


创建循环点击列表

1. 移动鼠标选中视频列表中的标题，点击鼠标右键，选择红色方框中的“创建元素列表”



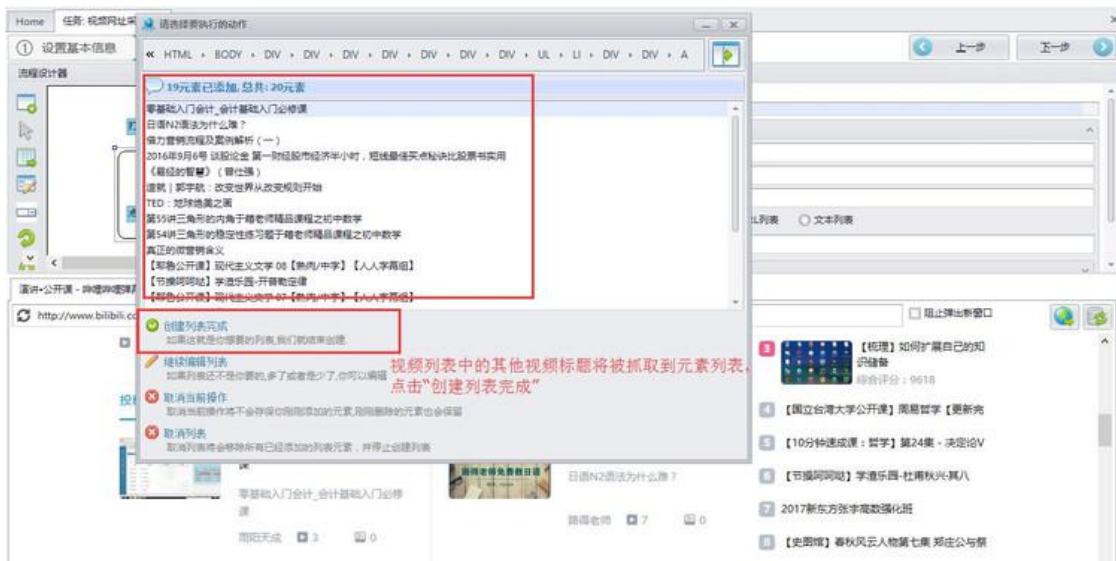
2. 选择添加到列表



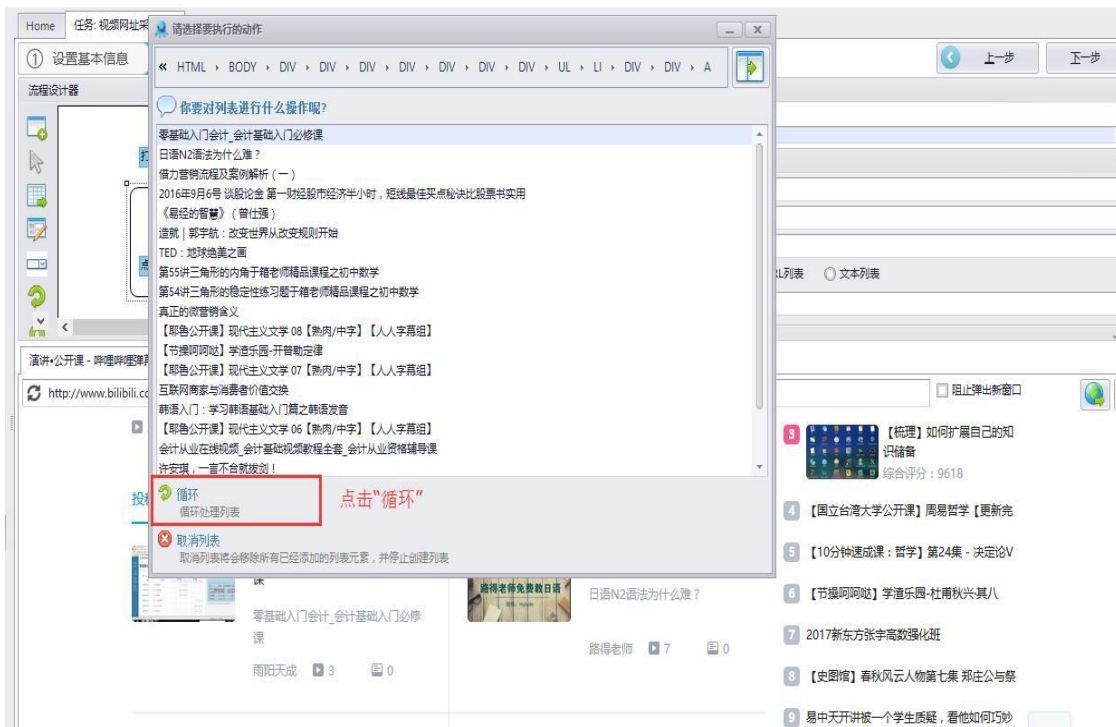
3.这个元素就会添加到上方的列表中，在执行“继续编辑列表”



4.选中视频列表中的第二个视频，同样的方法，将其添加到元素列表中，视频列表中的其他视频标题将被抓取到元素列表，这是因为我们添加的是具有两个相似特征的元素，八爪鱼采集器会智能的将页面中其他具有相似特征的元素都添加进来。点击“创建列表完成”



5.如图，点击“循环”

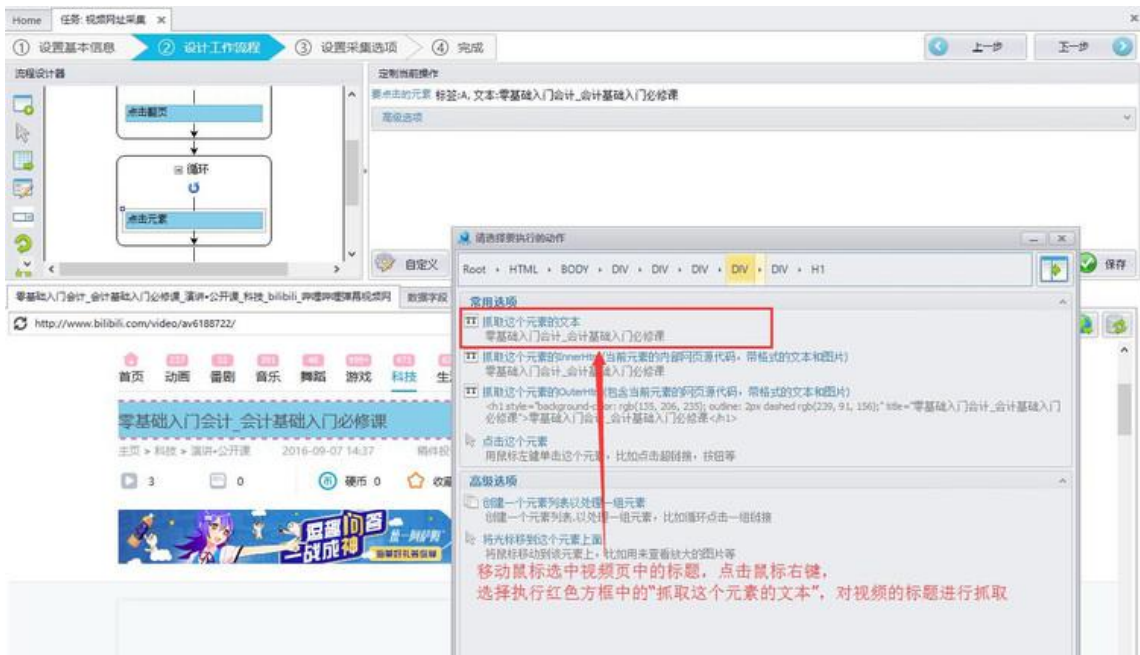


6.左上方的流程设计器会显示一个点击元素的循环框。完成循环点击列表的创建



抓取视频 URL

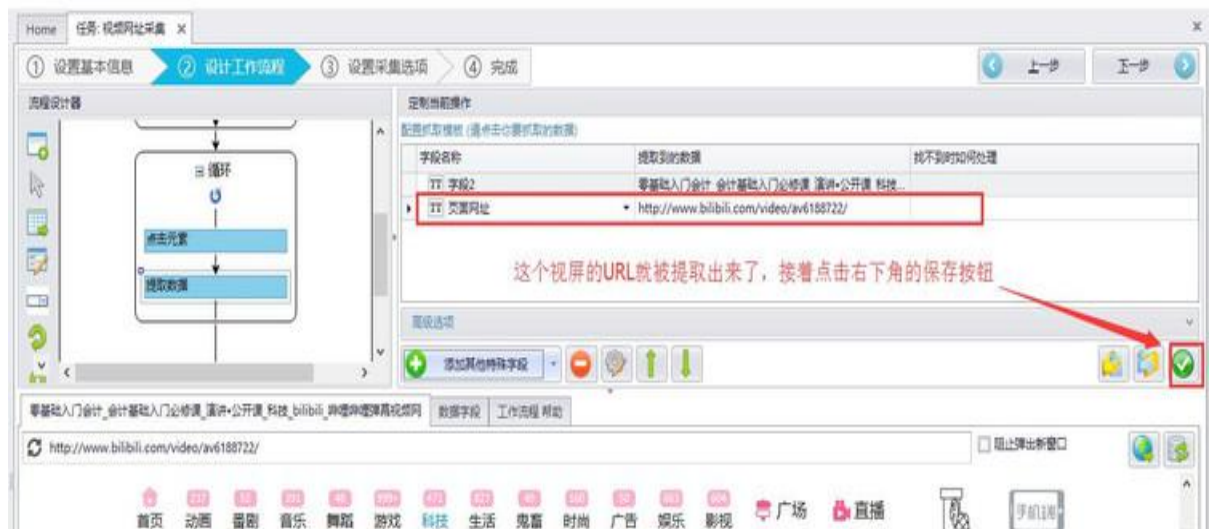
1. 移动鼠标到视频标题处，点击右键，选择执行红色方框中的“抓取这个元素的文本”，对视频的标题进行抓取



2. 抓取到的标题会显示在右上方的操作框中，点击“添加其他特殊字段”，选中“添加当前页面的网址”



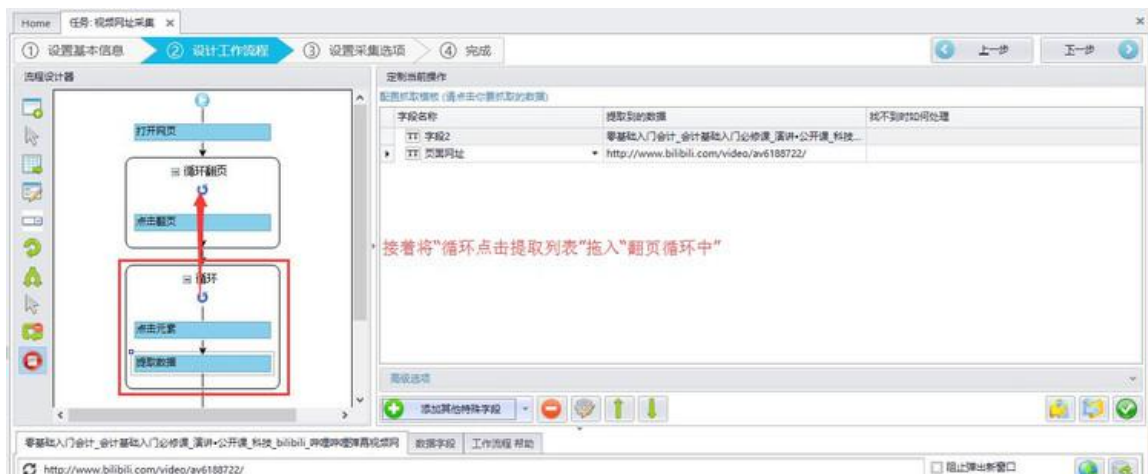
3.这个视屏的 URL 就被提取出来了，接着点击右下角的保存按钮



4.完成视频 URL 提取步骤

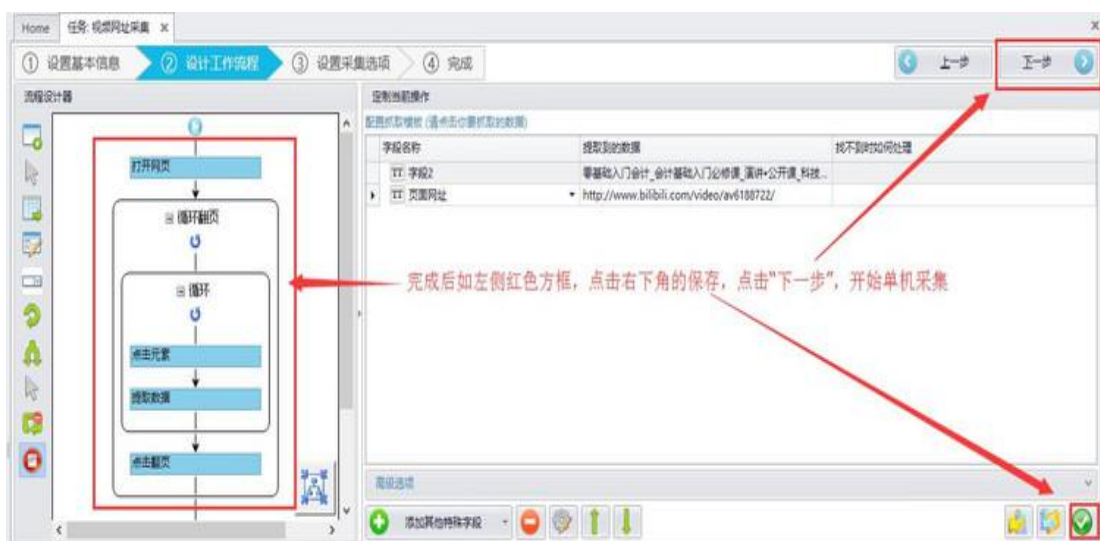


5.由于每一页都需要循环采集数据，所以我们需要将这个循环列表拖入到翻页循环里。



注意：流程是从上网页执行的，所以这个循环列表需要放到点击翻页的前面，否则会漏掉第一页的数据。

6.完成后的采集规则就如图中红色方框，点击右下方的保存后，就可以点击“下一步”，开始视频 URL 的采集



开始采集

1.选择单机采集，开始视频 URL 的提取



2.采集完成，将视频的 URL 导出

The screenshot displays the 'Video Website Collection' (视频网址采集) software interface. The main window shows a Bilibili video page for 'Bitcoin's 3rd Anniversary'. The right sidebar contains task information and a flowchart of the collection process. Below the main window, a table lists the extracted data, including video titles and their URLs. A red arrow points to the 'Export to Excel (2007)' option in the bottom right corner.

任务信息 预览

任务名: 视频网址采集
有效期: 从2016/9/7 到2017/3/7
描述:

缩略图:

采集选项 采集完成将视频URL导出

- 采集异常时提示错误信息
- 禁止浏览器加载图片,勾选后可以加快采集速度
- 使用Http代理 点击设置代理IP

字段2	页面网址
683 会计实操论坛_物业会计实操演练_会计实操题_演讲+公开课_科技...	http://www.bilibili.com/video/av6098998/
684 【史蒂夫·乔布斯在NeXT的内部运营策略】(1991)_演讲+公...	http://www.bilibili.com/video/av6098862/
685 【史蒂夫·乔布斯在NeXT介绍新闻发布会】(1988年)_演讲+公...	http://www.bilibili.com/video/av6098783/
686 零基础日语-表白篇_演讲+公开课_科技_bilibili_哔哩哔哩视...	http://www.bilibili.com/video/av6098641/
687 3.2 突围! 高考计算化繁为简! 第二弹: 如何用“点差法”简化...	http://www.bilibili.com/video/av6098459/
688 股票早知道 今日股票走势及龙头个股推荐(0831)_演讲+公开课...	http://www.bilibili.com/video/av6098267/
689 【存档成功】暖男阿祖再现江湖!!【氧化还原方程式归来②】...	http://www.bilibili.com/video/av6098229/
690 【书报阿阿阿】学道乐器-国周运动3_演讲+公开课_科技_bilibili...	http://www.bilibili.com/video/av6098133/
691 2016年+出的视频1_演讲+公开课_科技_bilibili_哔哩哔哩视...	http://www.bilibili.com/video/av6098131/
692 蔚蓝网络学院标准日语初级 免费学习(3) 第十一讲_标清...	http://www.bilibili.com/video/av6097995/
en2 日语零基础,零基础日语_演讲+公开课_科技_bilibili_哔哩哔哩视...	http://www.bilibili.com/video/av6097867/

导出到数据库
发布到网站
导出到Excel(2007)
导出到Excel(2003)
导出到Csv
导出到Txt
导出到Html

3.将视频 URL 导出后,使用视频 URL 批量下载工具将视频下载出来就完成了。