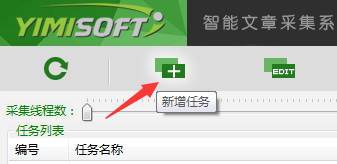
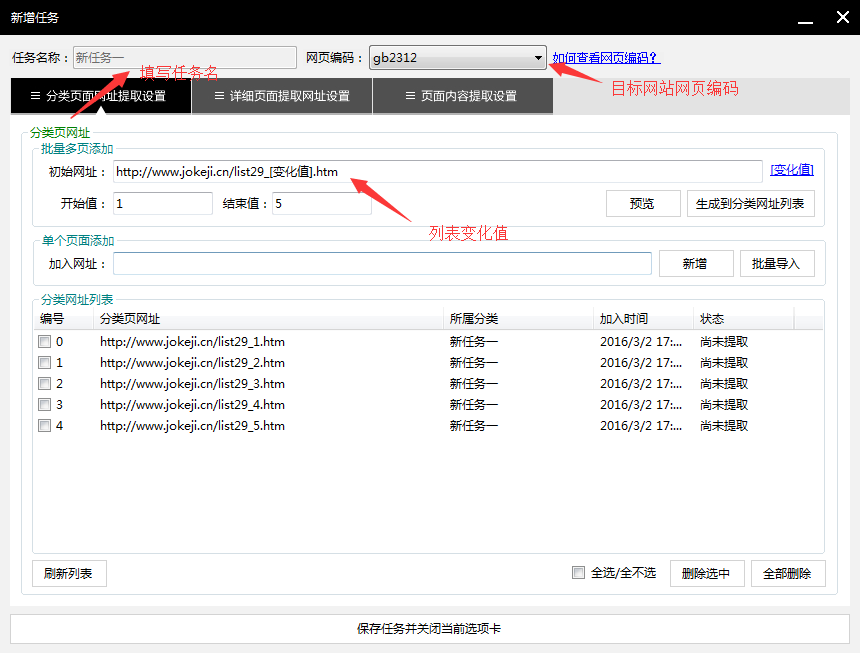
**智能文章采集系统教程**



**新增任务，打开新增任务设置窗口**

一：分类页面网址提取设置



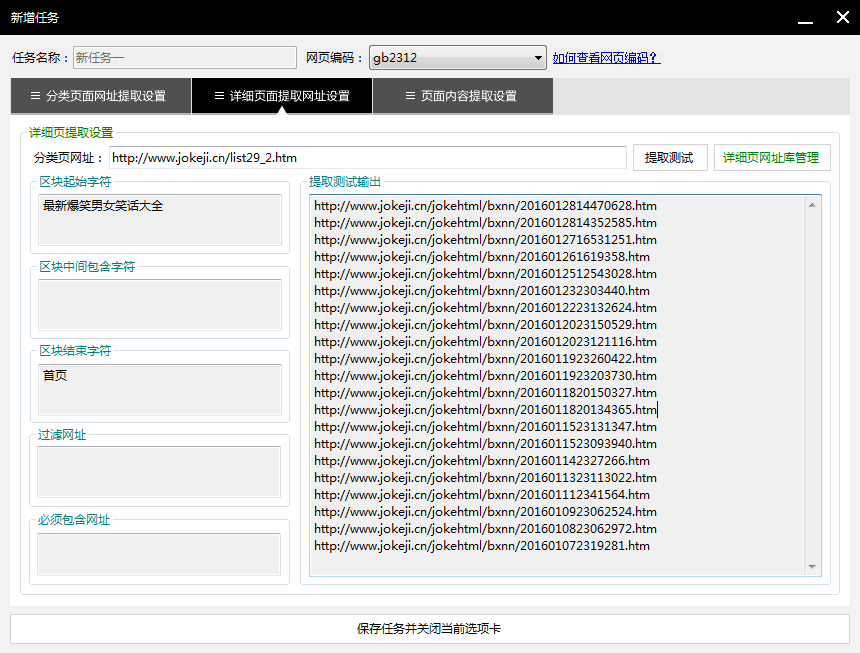
1.先填写唯一任务名（一般是按网站栏目或分类名，也可以自己取，主要是便于标识）

2.设置网页编码，网页编码在目标网页的源码中查看，选择对应的编码即可（只要编码正确，可识别任意语言）



3.生成到分类网址列表，也可单个网址添加，或多个网址整理在TXT中一行一个批量导入

二：详细页面提取网址设置



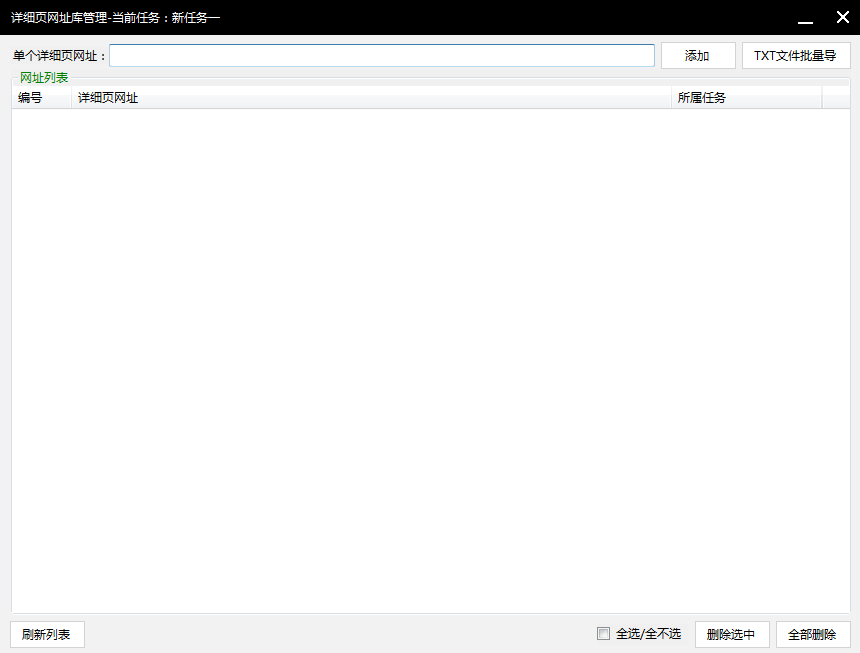
区块起始字符对应网页中列表网址头部字符，如

区块结束字符对应网页中列表网址底部字符，如



按“提取测试”按钮尝试获取详细页网址，如果有非详细页网址，则可在过滤网址和必须包含网址中设置过滤，只保留需要的详细页网址

如果想要自己添加详细页网址，可在

按钮点开的

中新增，批量导入或删除

三：网页内容提取设置

浏览器中点开任意一个设置的分类或栏目中的详细面

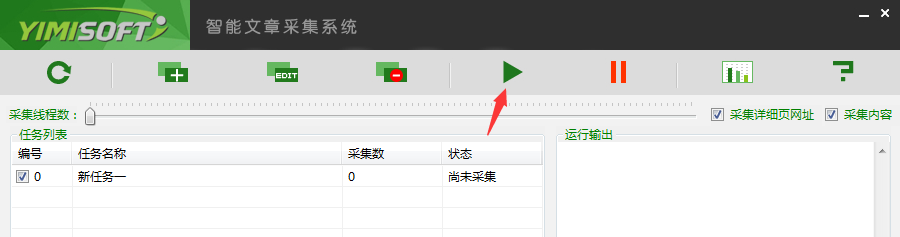
比如：<http://www.jokeji.cn/jokehtml/bxnn/2016012814470628.htm>



软件会自动识别提取正文及标题内容，如果含有一些不需要的内容，可以设置过滤

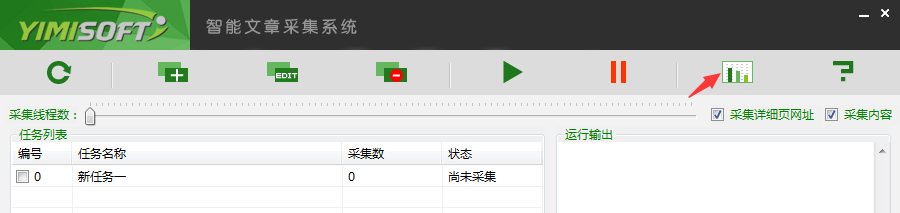
设置完点击保存任务，一个任务设置完成。

**开始采集**



选中刚才新增的任务,选择线程数和采集方式，点“开始采集”即执行采集，点停止为终止采集。可多个任务同时勾选

**采集数据管理**



点击如上图中“采集数据管理”打开“采集数据管理”窗口

**“已采集数据管理”选项卡**



单击选择一行，编辑可查看文章内容，勾选行可执行批量删除操作

**“数据批量修改及导出”选项卡**



**启用伪原创处理**

1. 插入关键词/锚链接，可在文章中文中随机插入关键词或锚链接
2. 替换同义词，可替换内容或标题中的同义词，可选择对应的同义词库，支持英文的TBS同义词库，自带有中文同义词库。

**发布设置**

一：接口生成，选择对应cms类型，填写密码，生成接口文件（在本地程序目录下的”jiekou”文件夹下），把生成的接口文件上传到网站根目录下即可。

二：数据发布设置，填写步骤一种上传的接口文件的网址和密码，填写要发布数据到cms里的栏目/分类ID，即可发布