

## 信源板块拆分说明

根据目前配置平台,我们所能直接配置的信源为版块信源(即网站的“栏目”,我们称之为版块),例:<http://www.xinhuanet.com/mil/yaowen.htm>,“新华网-军事-要闻”即为我们可配置的信源形式。



第一步:查找网站。根据需求,如:地域属性(包含网站地理位置及关注地域)、网站类型(新闻、论坛等)、关注领域(如经济、军事等),利用搜索引擎(百度、Google等)通过关键词(如地区+新闻)来搜索所需网站。

**注:** 现在查找网站基本不用,由客户直接提供所需配置网站。

第二部:拆分版块。找到该网站下所有频道、子版块、版块,根据要求拆分相应类型版块,如拆分新闻类、军事类、经济类等。一般网站都会有“网站导航”、“网站地图”,这样拆分的时候比较方便,又不会漏掉版块。

第三步：属性标注。版块拆分过程中所需标注的属性主要是拆分模板中所列的项目，如图所示：

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
必填	必填	必填	必填	必填	必填				必填	必填																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
网站名称	首页URL	域名	通道类型	板块名称	板块URL	地域大小	关注地域	地理位置	舆情	商情	板块描述	导入批次																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
									<div><div>1. 时政</div><div>1.1 国家政治</div><div>1.2 军事</div><div>1.3 外交</div><div>1.4 民族</div><div>1.5 人口与计划生育</div><div>1.6 人力资源</div><div>1.7 社会保障</div></div>																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							

网站名称、首页 URL、域名、通道类型、版块名称为必填项，其他项若有需求，根据实际需求填写。

拆分原则：

1、网站名称统一使用网站 logo 上的名称，网站首页一定要打开网站首页复制 url（所提供的网站名称、首页存在不规范的情况）。如图所示



网站名称应为：新华网，网站首页为：<http://www.xinhuanet.com/>。

2、域名填写，域名一定要填写顶级域名，不要加网络名（如 www.）子频道或者子网站不算做单独网站，遇到提供的网站首页为子频道或者子网站的情况，要回溯到网站，填写网站名称、网站首页、和网站域名。如：新华网，<http://www.xinhuanet.com/>，域名为 [xinhuanet.com](http://www.xinhuanet.com/)；

新华时政, <http://www.news.cn/politics/>, 不能作为单独的网站, 所属网站应为新华网, <http://www.xinhuanet.com/>, 域名为 [xinhuanet.com](http://www.xinhuanet.com/)。

3、版块名称填写。版块名称命名规则为: 网站名称-具体版块名称, 也可以理解为到达该版块的路径。如: 新华网-军事-要闻, <http://www.xinhuanet.com/mil/yaowen.htm>。

4、子版块、子频道一定要拆分到版块, 若无特殊要求, 不拆分网站首页(包含子版块、子频道)信源。

5、若页面显示有“更多”、“加载更多”等字样, 一定要复制点击“更多”之后的版块 url。如: 新华网-新闻-即时要闻, <http://www.news.cn/jsyw.htm>, 复制到表格的版块 URL 应为: [http://www.xinhuanet.com/jsyw\\_more.htm](http://www.xinhuanet.com/jsyw_more.htm)。

注: 简而言之, 板块页就是有文章列表和翻页的页面。(如果文章数少可能会没有翻页。)

6、电子报。电子报网站信源版块和网站相同, 拆分时尽量找短地址。如人民日报, <http://paper.people.com.cn/>, 网站名称: 人民日报, 网站首页: <http://paper.people.com.cn/>, 版块名称: 人民日报-首页, 版块 url: <http://paper.people.com.cn/>。

7、舆情属性填写。舆情属性根据版块实际性质填写, 在“信源拆分模板”中有下拉框, 可直接选择相应属性。