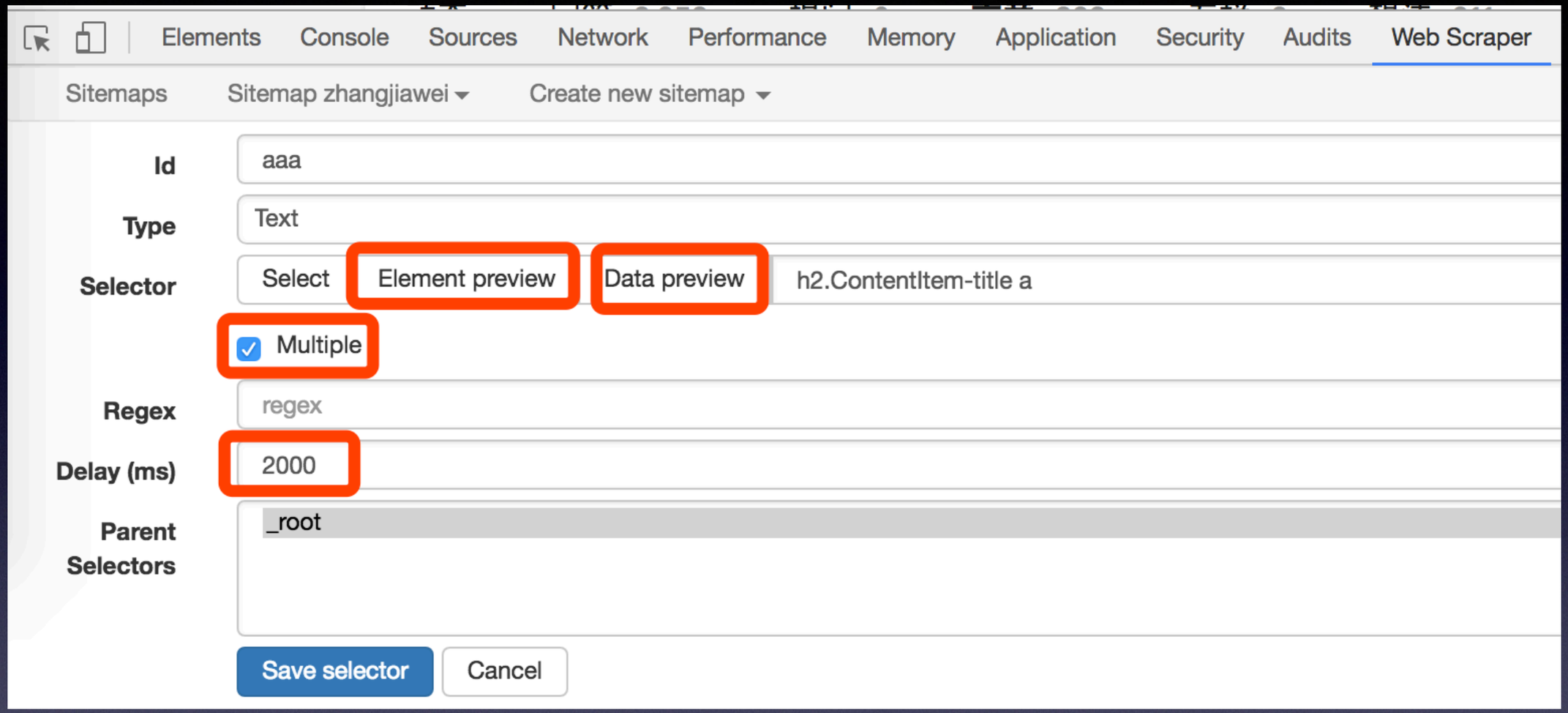
上节例题答疑

https://zhuanlan.zhihu.com/p/38345088

三、web scraper 原理浅析

1、选择器（selector）参数讲解
2、爬虫数据抓取原理（如何应用到所有网页）
3、选中元素顺序原理
4、csv 文件讲解
5、selector 操作选项讲解
6、sitemap 详情选项讲解

1、选择器选项



Element preview：查看信息是否选中

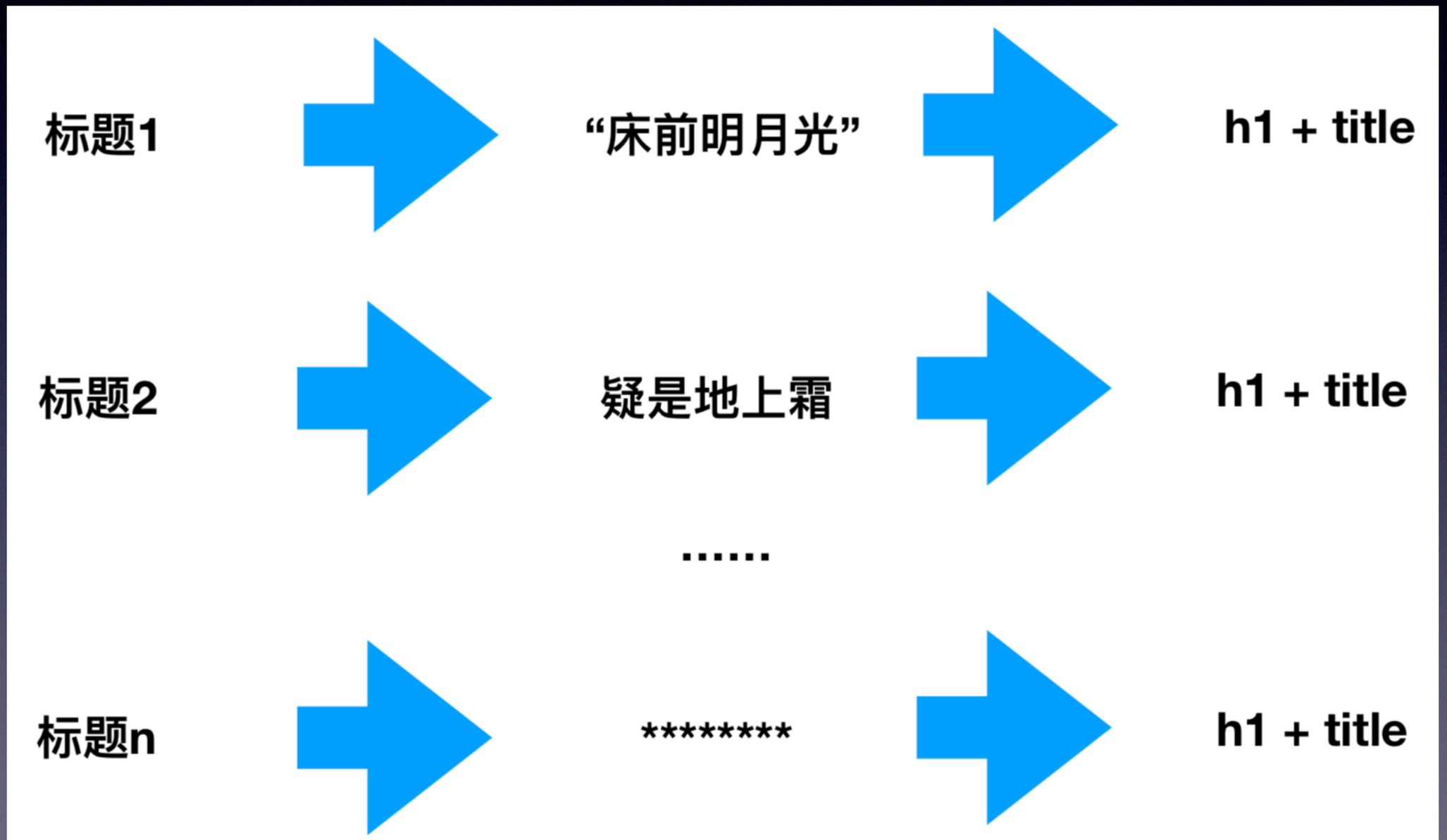
Data preview：预览抓取数据

Multiple：抓取多个

Delay：延时、确保数据加载成功（2000 - 5000）

2、数据抓取原理

为何选中 2 个标题后，所有的标题都被选中?



同类型自动识别

3、selector 选中元素顺序

自上而下，从“第一个”开始

编程 ——————— 遍历从当前开始

产品 ——————— 选择的权利交给用户

4、csv 文件讲解

	web-scraper-order	web-scraper-start-url	title
2	1528021461-4	https://www.zhihu.com/people/zhang-jia-wei/posts	冲着馆子的招牌菜来，却被
3	1528021461-2	https://www.zhihu.com/people/zhang-jia-wei/posts	最燃的时刻…… 莫过于变成
4	1528021461-1	https://www.zhihu.com/people/zhang-jia-wei/posts	世上总有追不上的人，所以
5	1528021461-7	https://www.zhihu.com/people/zhang-jia-wei/posts	长大了多哄几次妈，就懂小
6	1528021461-19	https://www.zhihu.com/people/zhang-jia-wei/posts	两个詹姆斯的 40+，骑士与
7	1528021461-15	https://www.zhihu.com/people/zhang-jia-wei/posts	勒布朗串烤猛龙 · 霍福德力
8	1528021461-20	https://www.zhihu.com/people/zhang-jia-wei/posts	生死之际凯尔特人少年的执
9	1528021461-3	https://www.zhihu.com/people/zhang-jia-wei/posts	心爱的球星离开时，我们能
10	1528021461-5	https://www.zhihu.com/people/zhang-jia-wei/posts	伤不伤身体我们不管，你揭

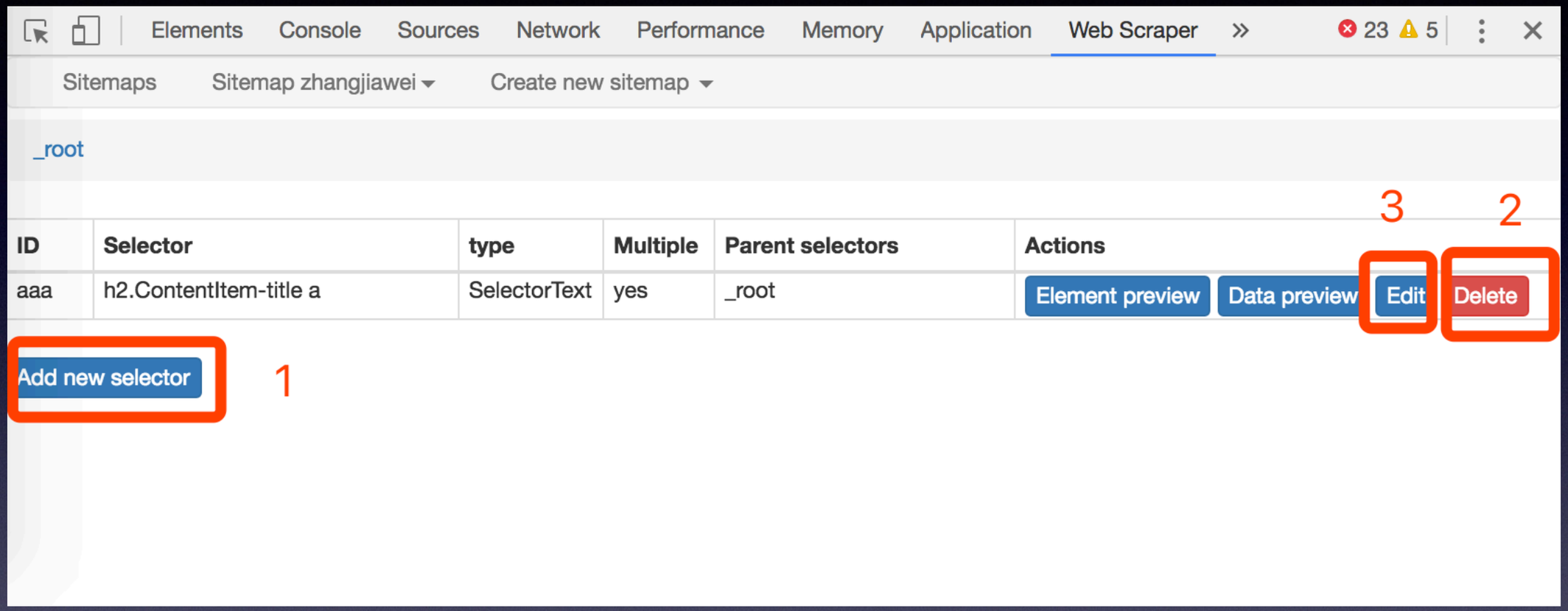
1、zhangjiawei.csv：sitemap name(ID)

2、web-scraper-order ： 由于排序

3、web-scraper-start-url：start url

4、title：selector Id

5、selector 操作选项讲解

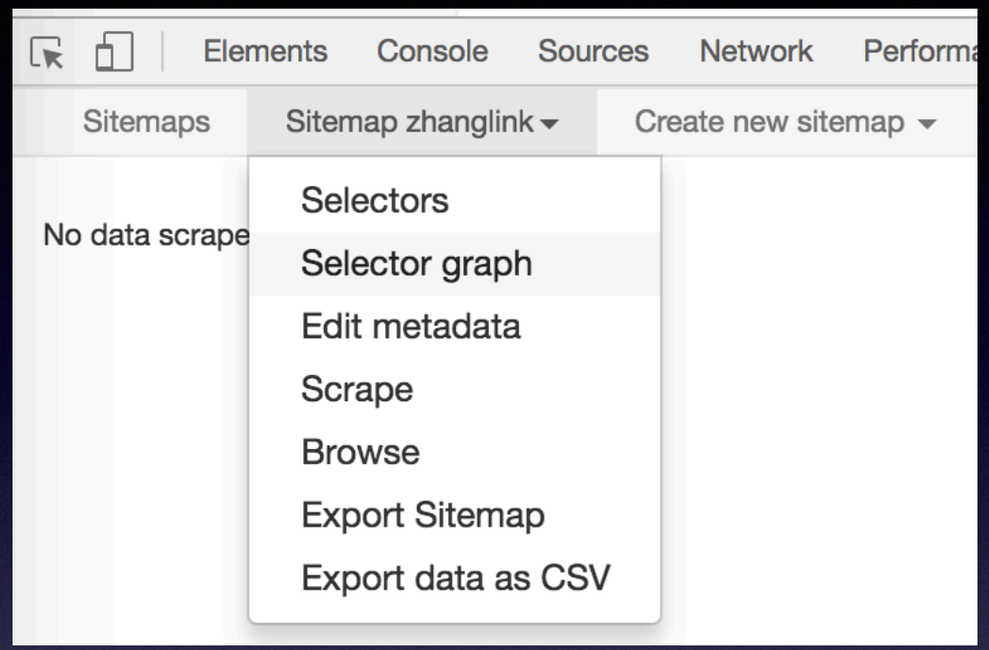


1、增加

2、删除

3、编辑（查看）

6、sitemap 详情选项讲解



1、Selectors：显示 Selector 列表
2、Selector graph：Selector 树状结构图
3、Edit metadata：修改 sitemap name 和 start url
4、Scrape：开始抓取程序
5、Browse：浏览抓取结果
6、Export Sitemap：导出 sitemap 设置
7、Export data as csv：导出抓取结果 csv 文件到本地电脑