

## 二、统计知乎大V文章标题

实例：知乎大V张佳玮文章主页

<https://www.zhihu.com/people/zhang-jia-wei/posts>



# 目标效果图

1	web-scrapers-order	web-scrapers-start-url	title
2	1528021461-4	https://www.zhihu.com/people/zhang-jia-wei/posts	冲着馆子的招牌菜来，却被惊艳的小菜吓呆
3	1528021461-2	https://www.zhihu.com/people/zhang-jia-wei/posts	最燃的时刻..... 莫过于变成超级赛亚人吧
4	1528021461-1	https://www.zhihu.com/people/zhang-jia-wei/posts	世上总有追不上的人，所以，专注按自己的节奏跑吧
5	1528021461-7	https://www.zhihu.com/people/zhang-jia-wei/posts	长大了多哄几次妈，就懂小时候她哄我们的心情了
6	1528021461-19	https://www.zhihu.com/people/zhang-jia-wei/posts	两个詹姆斯的 40+，骑士与火箭的不同命运
7	1528021461-15	https://www.zhihu.com/people/zhang-jia-wei/posts	勒布朗串烤猛龙·霍福德力战双英
8	1528021461-20	https://www.zhihu.com/people/zhang-jia-wei/posts	生死之际凯尔特人少年的执行力·勇士如何锁了朗多与戴维斯
9	1528021461-3	https://www.zhihu.com/people/zhang-jia-wei/posts	心爱的球星离开时，我们能做些什么呢
10	1528021461-5	https://www.zhihu.com/people/zhang-jia-wei/posts	伤不伤身体我们不管，你揭露了，就活该被追杀！
11	1528021461-18	https://www.zhihu.com/people/zhang-jia-wei/posts	凯尔特人怎么在 48 小时内列起残阵，扣住西蒙斯与费城的？
12	1528021461-12	https://www.zhihu.com/people/zhang-jia-wei/posts	杜兰特 & 保罗：三分球时代，中投也能投死人！
13	1528021461-13	https://www.zhihu.com/people/zhang-jia-wei/posts	勒布朗举重若轻地，塔图姆华丽多变地，挽救了最后时刻
14	1528021461-9	https://www.zhihu.com/people/zhang-jia-wei/posts	沈从文先生的气性，及他与汪曾祺先生的不同
15	1528021461-10	https://www.zhihu.com/people/zhang-jia-wei/posts	33 岁零 3 天的克里斯·保罗，扼住了一向虐待他的命运
16	1528021461-6	https://www.zhihu.com/people/zhang-jia-wei/posts	连谢尔顿都改变了，结婚了..... 你呢？
17	1528021461-11	https://www.zhihu.com/people/zhang-jia-wei/posts	换一个首发，西蒙斯活了，猛龙溃了
18	1528021461-16	https://www.zhihu.com/people/zhang-jia-wei/posts	我知道那些都是很好很好的，但不是我想要的呀！
19	1528021461-14	https://www.zhihu.com/people/zhang-jia-wei/posts	比起 2016 年之前，哈登究竟哪里脱胎换骨了？
20	1528021461-8	https://www.zhihu.com/people/zhang-jia-wei/posts	还记得十年前 5 月 12 日汶川地震时，在做什么吗？
21	1528021461-17	https://www.zhihu.com/people/zhang-jia-wei/posts	勒布朗与库里的威慑力，猛龙和鹈鹕的杯弓蛇影



# 步骤（通用！！！！）

1、打开 web scraper

2、创建 sitemap

3、设置 selector（抓取规则）

4、启动抓取程序

5、导出 excel 到本地电脑



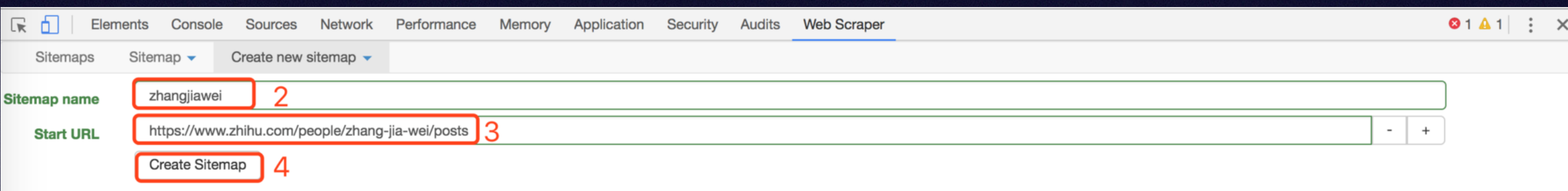
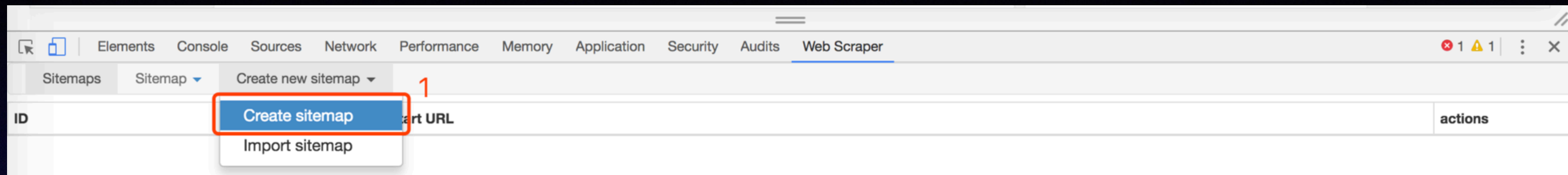
# 1、打开 web scraper

1、鼠标右键 ——> 审查元素

2、鼠标右键 ——> 检查



## 2、创建 sitemap

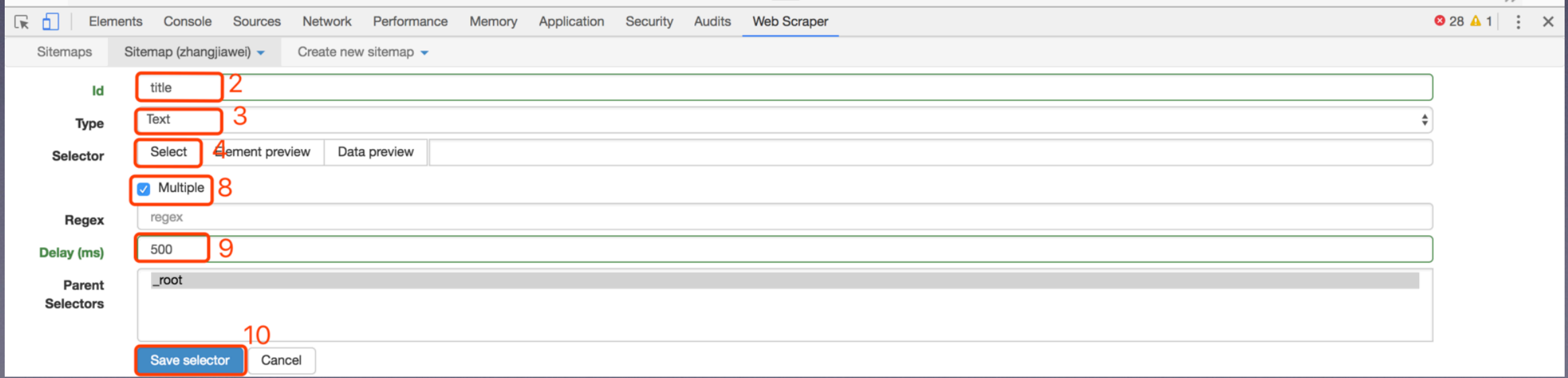
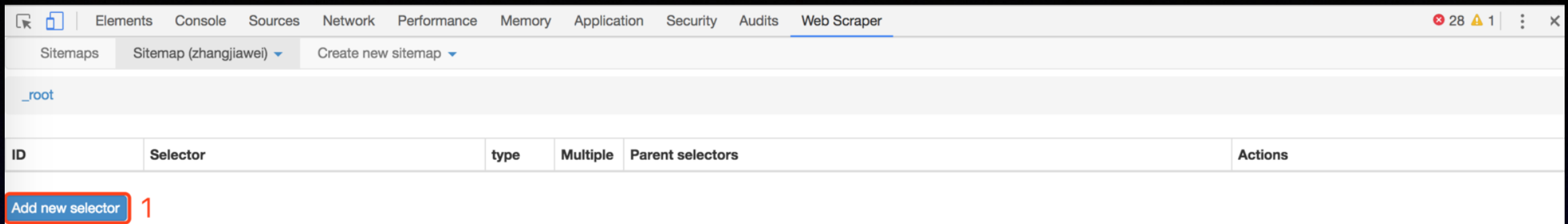


Sitemap name: 自定义

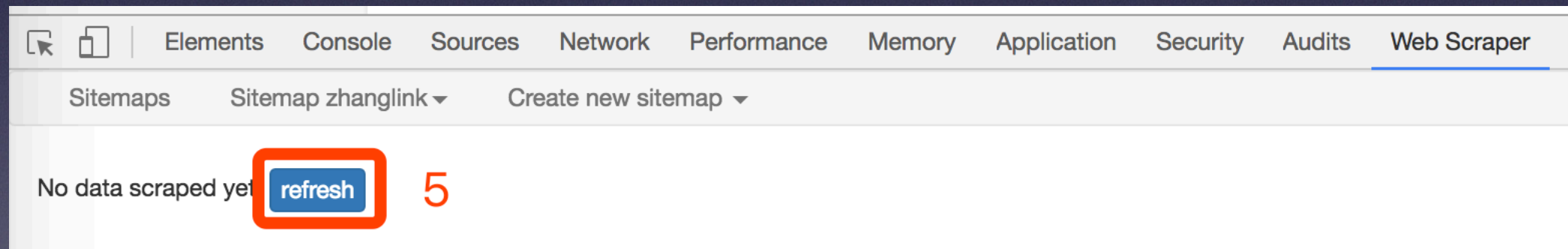
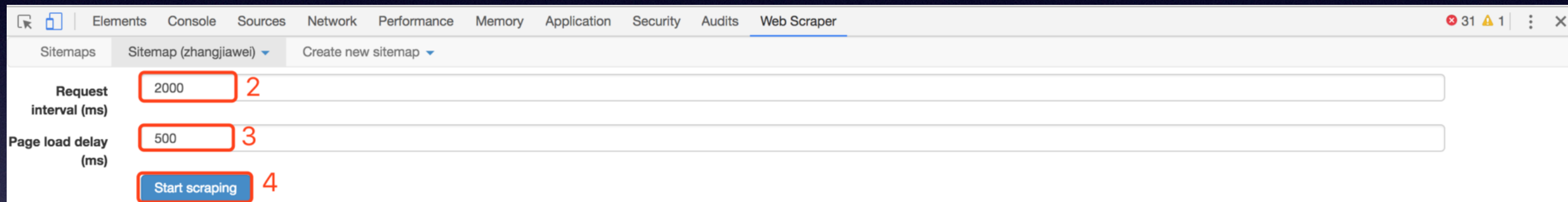
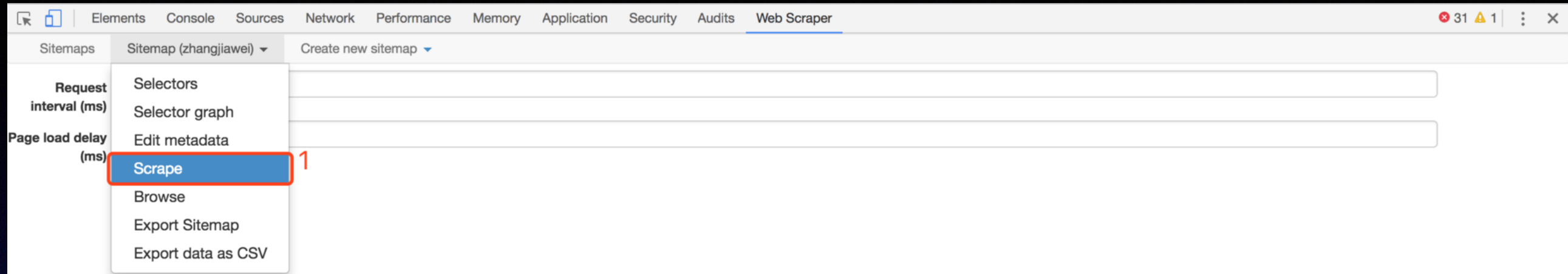
Start URL: 当前网址 (可直接从浏览器复制)



# 3、设置 selector



# 4、启动抓取程序

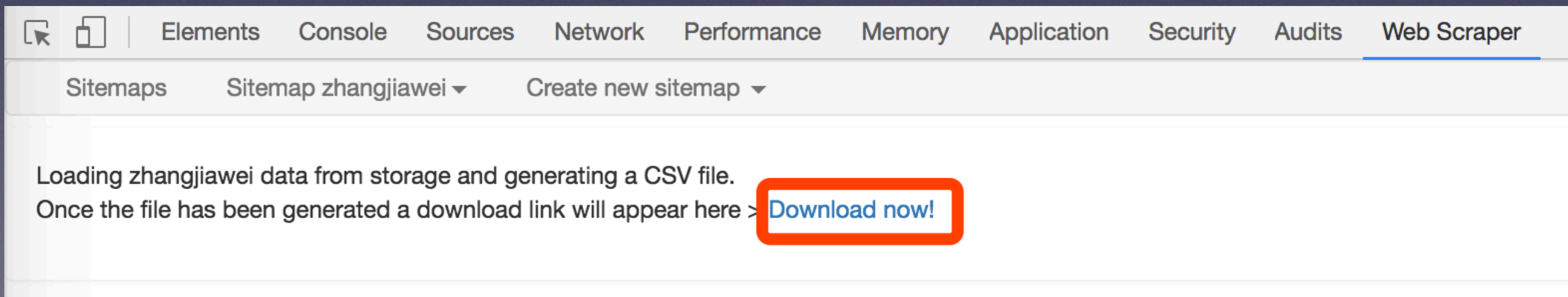
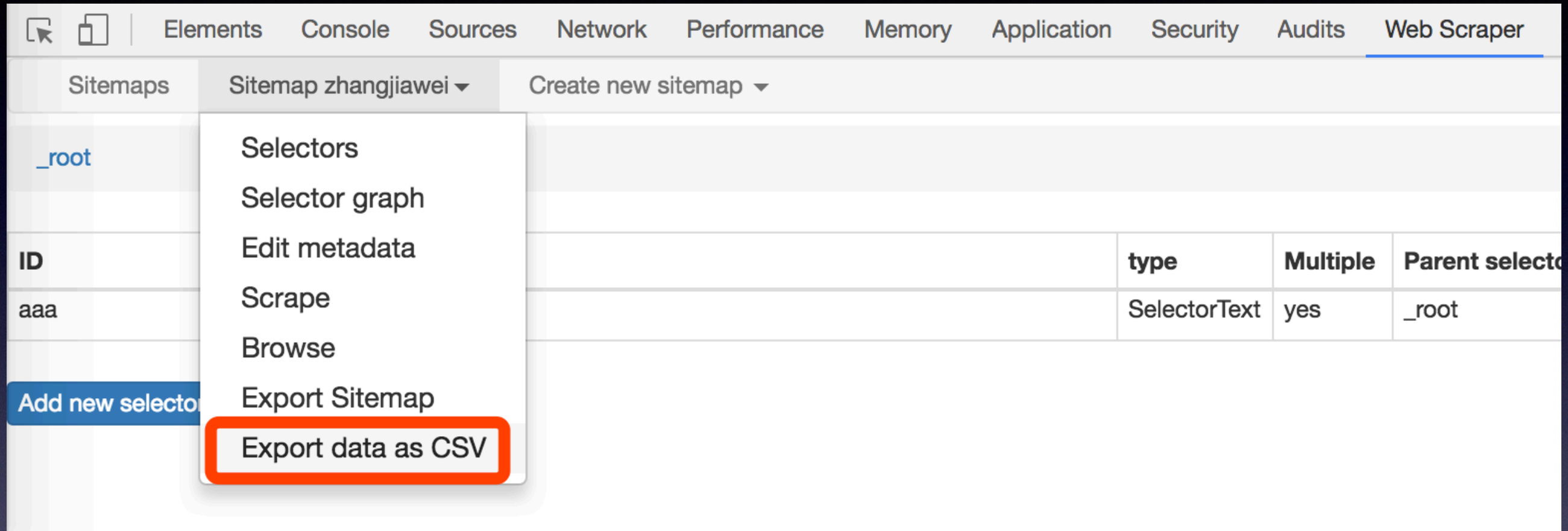


1、抓取窗口可以最小化，不能关闭！

2、可以同时启动多个抓取程序



# 5、导出 excel 到本地电脑





# CSV 文件讲解

1	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
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扶
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、web-scraper-order：由于排序

2、web-scraper-start-url：start url

3、title：selector Id



# 遗留问题

1、单页？

2、无序？

后续会详解



# 必须练习

1 **熟能生巧** 是精通一项技能型知识 永不变的秘诀

2 最低要求：练到**不用看教程步骤**，就能将**全部课程里的例子**抓取成功。

3 课程内的例子**没有达到最低要求**，不要抓**课程之外的网站**。

4 练习的时候，一定按照教程步骤，**不要想当然！！！！**



# 牛刀小试

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

214 条评论

⇌ 切换为时间排序

写下你的评论...

 有谁共鸣

8 天前

c 罗是天生的斗士，这就是他最出色的天赋

👍 119

 黎铁锤

8 天前

软广，猝不及防

👍 328

以上为精选评论 ?

 浪子天妖

8 天前

c 罗应该可以进入葡萄牙历史前十伟人了

👍 65

 智扈

8 天前

C 罗不慌，哈哈

👍 26

 郝先森

8 天前

约稿，C 罗，还以为是清扬

👍 93