

# 七、统计58同城房屋信息

实例：58同城房屋信息

[http://bj.58.com/chuzu/?](http://bj.58.com/chuzu/?key=%E7%A7%9F%E6%88%BF&cmcskey=%E7%A7%9F%E6%88%BF&final=1&jump=1&specialtype=gl)

[key=%E7%A7%9F%E6%88%BF&cmcskey=%E7%A7%9F%E6%88%BF&final=1&jump=1&specialtype=gl](http://bj.58.com/chuzu/?key=%E7%A7%9F%E6%88%BF&cmcskey=%E7%A7%9F%E6%88%BF&final=1&jump=1&specialtype=gl)



# 目标效果图

intro	price	style		address	source
整租   石景山区八角中里南向 价格便宜 中间楼层 屋里干净	3200元/月	1室1厅1卫	45m²	八角 八角中里出租 距1号线八角游乐园站938米	来自经纪人： 我爱我家 - 八角南里店买卖B组 李勤 李勤 6年 我爱我家 58同城认证商家
整租   望京SOHO旁 望京西园一区 南向大一居出租 520	5200元/月	1室1厅1卫	65m²	望京 望京西园一区出租 距15号线望京站807米	来自经纪人： 中天置地 - 中天置地望京店 郭正军 郭正军 1年 中天置地 58同城认证商家
整租   亦庄 青年公寓 1室1厅	2700元/月	1室1厅1卫	28m²	亦庄 青年公寓	null
整租   泰和宜园 精装修 南北通透 家电家具齐全 出租300	3000元/月	2室1厅1卫	92m²	马坡 泰和宜园出租	来自经纪人： 中联置家 - 中联置家龙湖店D组 郑兴飞 郑兴飞 1年 中联置家 58同城认证商家
单间   龙跃苑一区 4室1厅1卫 转租房，房东住主卧	1650元/月	次卧(4室)	18m²	回龙观 龙跃苑一区出租	null
单间   亦庄线次渠南站次渠南里小区紧邻京东总部酒仙网鸿坤云时	1350元/月	次卧(3室)	16m²	次渠 南里出租	来自经纪人： 美丽屋资产 田森 田森 1年 美丽屋资 58同城认证商家
整租   六里桥南里 1室1厅 20平米 精装修 押一付一	2600元/月	1室1厅1卫	20m²	广安门内 六里桥南里出租 距10号线六里桥站232米	来自经纪人： 美丽家园 张自然 张自然 1年 美丽家园 58同城认证商家
单间   当天订房送3天房租地铁六里桥站免所有杂费押一付一免费	1980元/月	次卧(3室)	16m²	七里庄 六里桥家园出租	来自经纪人： 韩静 韩静 1年 58同城认证商家
单间   次渠锦园次卧出租 紧邻地铁 全新入住	1300元/月	次卧(3室)	15m²	亦庄东区 次渠锦园出租	来自经纪人： 发源地产 包梅珍 包梅珍 1年 发源地产 58同城认证商家
整租   望京南地铁站 大山子 高家园 精装复试一居出租 随时	3800元/月	1室1厅1卫	40m²	酒仙桥 高家园五区出租 距14号线高家园站171米	来自经纪人： 广发地产 - 高家园店 凡梦豪 凡梦豪 1年 广发地产 58同城认证商家
整租   定福家园南里3区 精装大两居腿儿这到地铁口 好住不贵	4500元/月	2室1厅1卫	77m²	定福庄 定福家园南里3号院出租 距6号线褡裢坡站167米	来自经纪人： 蛋壳公寓 - 蛋壳公寓朝阳门店 高永华 高永华 1年 蛋壳公寓 58同城认证商家
整租   出租香江北岸 精装大开间 紧邻地铁14号线与15号线	3500元/月	1室1厅1卫	56m²	孙河 康营家园12期出租 距15号线孙河站919米	来自经纪人： 蕴实地产 - 康营家园总店 赵贺宁 赵贺宁 3年 蕴实地产 58同城认证商家
单间   个人出租大主卧1850元 独立卫生间 招稳定年轻（中	1850元/月	主卧(3室)	35m²	天通苑 天通苑西三区出租	null
整租   京旺家园百平米大两居出租 地铁14号将台站 15号崔	5300元/月	2室1厅1卫	109m²	望京 京旺家园出租 距15号线崔各庄站3160米	来自经纪人： 昌胜源地产 - 昌胜源地产京旺家园店 田晶 田晶 1年 昌胜源地 58同城认证商家



# 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



# 3、设置 selector

北京出租

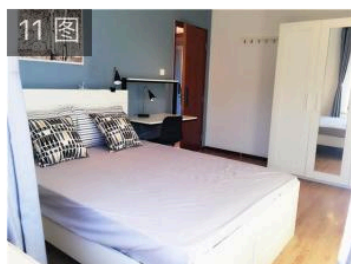
个人房源

经纪人

热租房源

品牌公寓

房屋求租



**整租 | 吉祥里小区 蛋壳公寓直租 物业齐全 南北向采光好** 

**6500** 元 / 月

1室1厅1卫 68m<sup>2</sup>

朝阳门 吉祥里小区出租  距离 2 号线朝阳门地铁站 483 米

来自经纪人: 蛋壳公寓 - 蛋壳公寓朝阳门店 程中霖



**整租 | 为你而选 华彩国际公寓 2室1厅117平** 

**13500** 元 / 月

2室1厅2卫 117.0m<sup>2</sup>

望京 华彩国际公寓出租  距离 14 号线 (东) 东湖渠地铁站 165 米

来自经纪人: 链家地产 - 链家地产望京花园店 C 组 汪洋



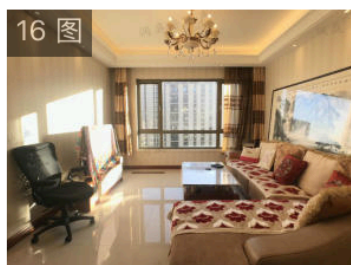
**整租 | 为你而选 荣丰 2008 1室1厅28平** 

**4800** 元 / 月

1室1厅1卫 28.49m<sup>2</sup>

广安门外 荣丰 2008 出租  距离 7 号线达官营地铁站 471 米

来自经纪人: 链家地产 - 链家地产荣丰店 A 组 范月琴



**整租 | 亦庄金茂悦 急租 133 平米家具家电全齐 五星物业 + 新** 

**9000** 元 / 月

3室2厅2卫 133m<sup>2</sup>

瀛海镇 金茂悦 (亦庄金茂悦南区) 出租

来自经纪人: 润庭房地产 - 润庭房地产总店 王炳成





# 3、设置 selector

html网页中的父子元素



web scraper中的父子选择器

整租 | 精装全齐二居室，南北通透三空调，房间明亮全地瓷，随时  热  3200元/月

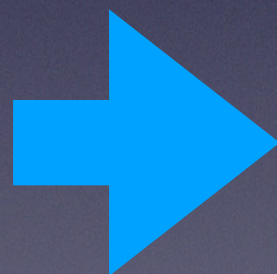
2室1厅1卫 94m<sup>2</sup> B

宋庄 李庄新村出租 C

来自经纪人：我爱我家 - 我爱我家华龙店1组 于冬冬 D

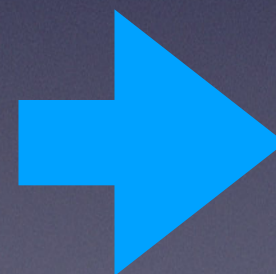
F

```
1. <F>
2.   <A></A>
3.   <B></B>
4.   <C></C>
5.   <D></D>
6.   <E></E>
7. </F>
```



F: 父亲

ABCDE: 儿子



F: Element selector: 整体  
A: Text selector: 简介  
B: Text selector: 房屋样式  
C: Text selector: 地址  
D: Text selector: 来源  
E: Text selector: 租金



# 3、设置 selector

如何选中 F ? ? ? (创建父选择器)

The diagram shows a rental listing with several elements highlighted by red boxes and numbered green circles:

- Circle 1:** Points to the price "3200元/月" (3200元/月), labeled with a red **F**.
- Circle 2:** Points to the empty space between the address and the price, labeled with a red **F**.
- Circle 3:** Points to the empty space between the room details and the address, labeled with a red **F**.
- Circle 4:** Points to the main title "整租 | 精装全齐二居室, 南北通透三空调, 房间明亮全地瓷, 随时", labeled with a red **A**.
- Circle 5:** Points to the agent information "来自经纪人: 我爱我家华龙店1组 于冬冬", labeled with a red **D**.

Other elements and their labels:

- B:** "2室1厅1卫 94m²"
- C:** "宋庄 李庄新村出租"
- E:** "3200元/月"

要选中父元素 F, 就必须用鼠标点击 **在 F 内, 但是又不在 ABCDE 内** 的区域。

否则要么选中子元素, 要么选中其他 F 之外的元素, 都不是我们想要的结果。

口诀: 多次尝试, 直到成功

# 3、设置 selector

## 创建父选择器

58同城·房产

租房

搜房源

免费发布

北京出租

个人房源

经纪人

热租房源

品牌公寓

房屋求租

房源推广请拨打: 4006401189

7图

整租 | 大一居室、万寿寺北里、紧邻法华寺小区、魏公村南区、韦

5500 元 / 月

1室1厅1卫 33m²

魏公村 万寿寺北里出租

距离 4 号线魏公村地铁站 1182 米

来自经纪人: 我爱我家 - 我爱我家 人济山庄店租赁 A 组 谢尚东

8图

单间 | 远洋一方林语苑 蛋壳公寓直租 贴心管家服务随时看房

2500 元 / 月

主卧 (2 室) 28m²

管庄 远洋一方林语苑出租

来自经纪人: 爱程平

整租 | 丰台地铁 9 号线 四合欣园 大兴家园 二居室 全朝南

5500 元 / 月

房源推广 套餐选购

帮您快速成交房源

点击购买 ▶

热推房源 hot

宋浩男 17600722119

朝阳 / 朝阳公园

3 室 130 平

13000 元 / 月

ul.listUI li

S

P

C

Done selecting!

Elements

Console

Sources

Network

Performance

Memory

Application

Security

Audits

Web Scraper

Sitemaps

Sitemap (tongcheng58)

Create new sitemap

Id

house

Type

Element

Selector

Select

Element preview

Data preview

ul.listUI li

Multiple

Delay (ms)

2000

Parent Selectors

\_root

house

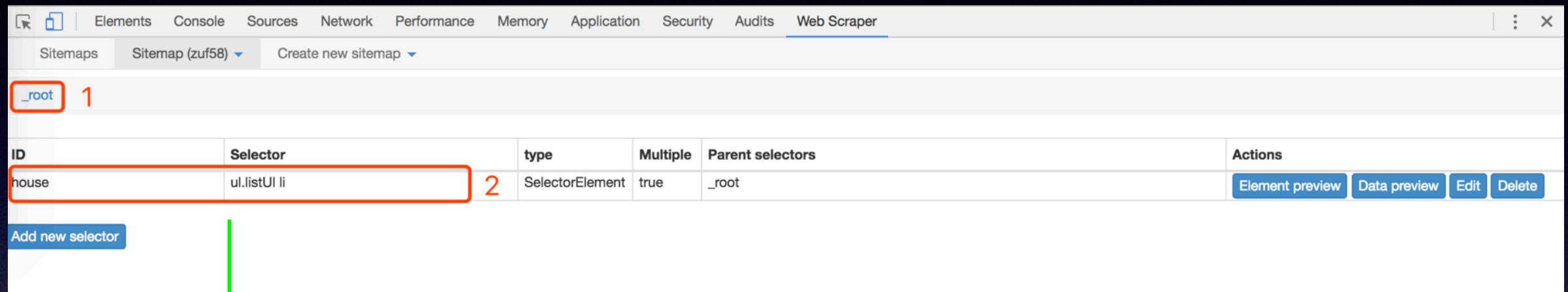
Save selector

Cancel

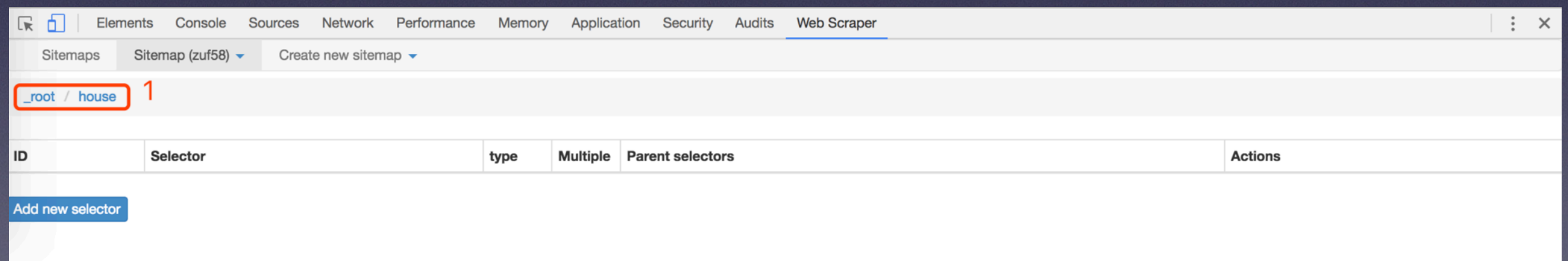


# 3、设置 selector

## 创建子选择器



- 1. **\_root** : 根选择器
- 2. **house**: 根选择器的子选择器。



- 1. **\_root / house** : 当前是在根选择器下的 **house** 选择器下操作



# 3、设置 selector

## 创建第一个子选择器

The screenshot shows a web browser displaying a real estate website (58同城·房产) and a web scraper interface (Web Scraper). The website has a red header with the logo and a search bar. Below the header are navigation tabs: 北京出租, 个人房源, 经纪人, 热租房源, 品牌公寓, 房屋求租. A yellow box highlights a specific rental listing. The web scraper interface at the bottom shows a configuration form with fields for Id, Type, Selector, Regex, Delay, and Parent Selectors. Red numbers 1 through 8 are placed over the interface to indicate steps:

1. Id field (intro)
2. Type dropdown (Text)
3. Selector dropdown (Select)
4. Yellow box on the website (highlighting the rental listing)
5. Price field (3200元/月)
6. "Done selecting!" button
7. "house" in Parent Selectors
8. "Save selector" button

4. 注意！！（点击“Select”）之后，你会发现 4 号区域（黄色）自动被选中。这是因为我们是在 house 下建立子选择器，而 house 在第一步已经被确定为黄色区域。



## 3、设置 selector

## 创建其余四个子选择器

## 步骤同第一个

整租 | 精装全齐二居室，南北通透三空调，房间明亮全地瓷，随时

精

热

3200元/月

2室1厅1卫 94m²

B

宋庄 李庄新村出租

C

来自经纪人： 我爱我家 - 我爱我家华龙店1组 于冬冬

D

F

## 五个子选择器都创建完毕后如下：

Sitemaps Sitemap (zuf58) Create new sitemap

[\\_root](#) / [house](#)

ID	Selector	type	Multiple	Parent selectors	Actions
intro	h2 a	SelectorText	false	house	<a href="#">Element preview</a> <a href="#">Data preview</a> <a href="#">Edit</a> <a href="#">Delete</a>
price	div.money	SelectorText	false	house	<a href="#">Element preview</a> <a href="#">Data preview</a> <a href="#">Edit</a> <a href="#">Delete</a>
style	p.room	SelectorText	false	house	<a href="#">Element preview</a> <a href="#">Data preview</a> <a href="#">Edit</a> <a href="#">Delete</a>
address	p.add	SelectorText	false	house	<a href="#">Element preview</a> <a href="#">Data preview</a> <a href="#">Edit</a> <a href="#">Delete</a>
source	div.jjr	SelectorText	false	house	<a href="#">Element preview</a> <a href="#">Data preview</a> <a href="#">Edit</a> <a href="#">Delete</a>

Add new selector



# 要点解析

## 子选择器设置

**Multiple:** 空

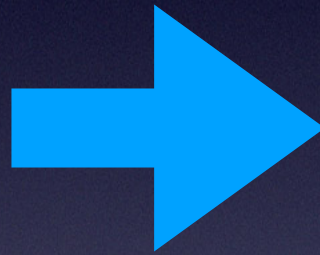
一个house只有一个price

**Regex:** 空

未用到

**Delay:** 空

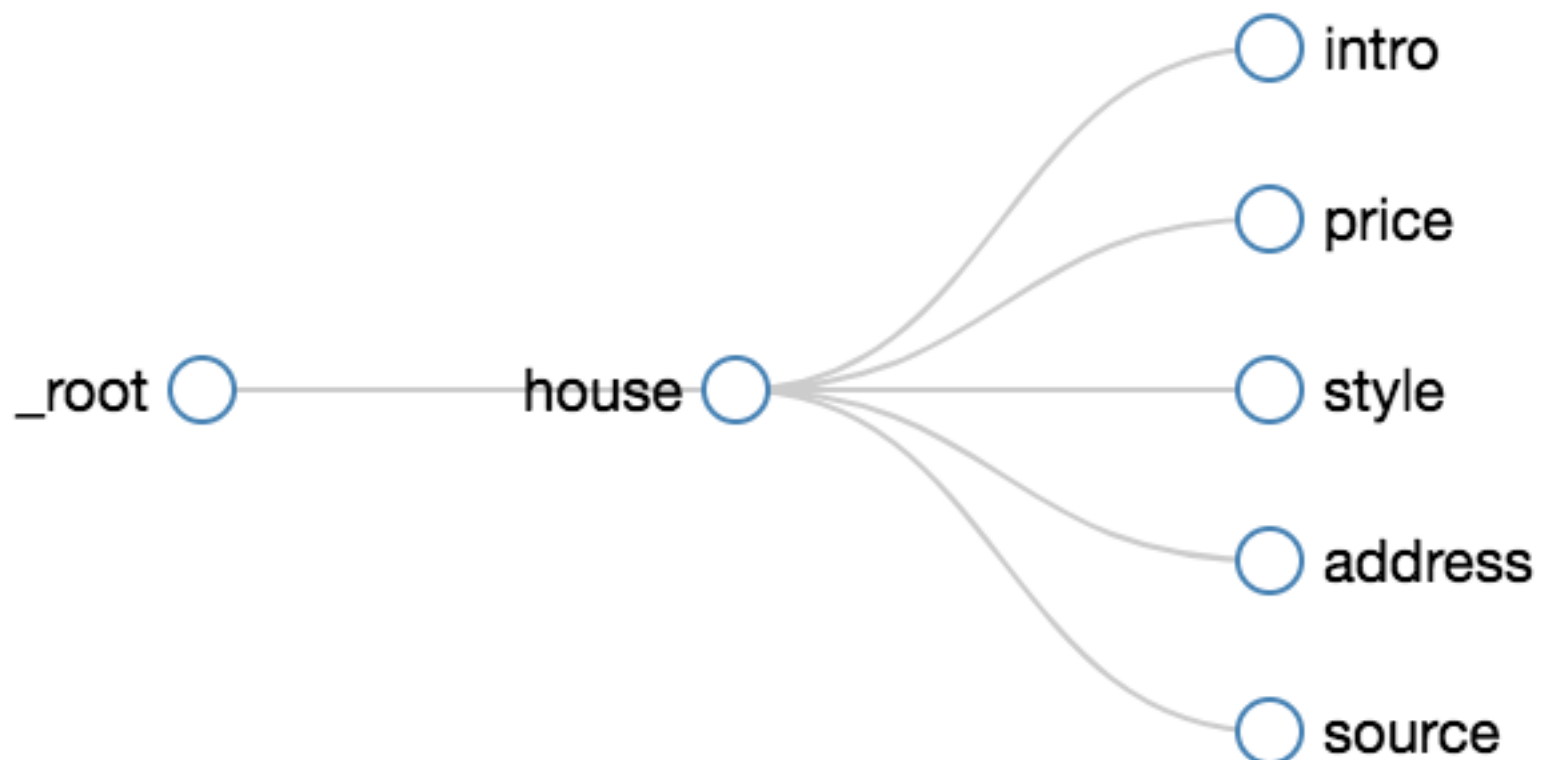
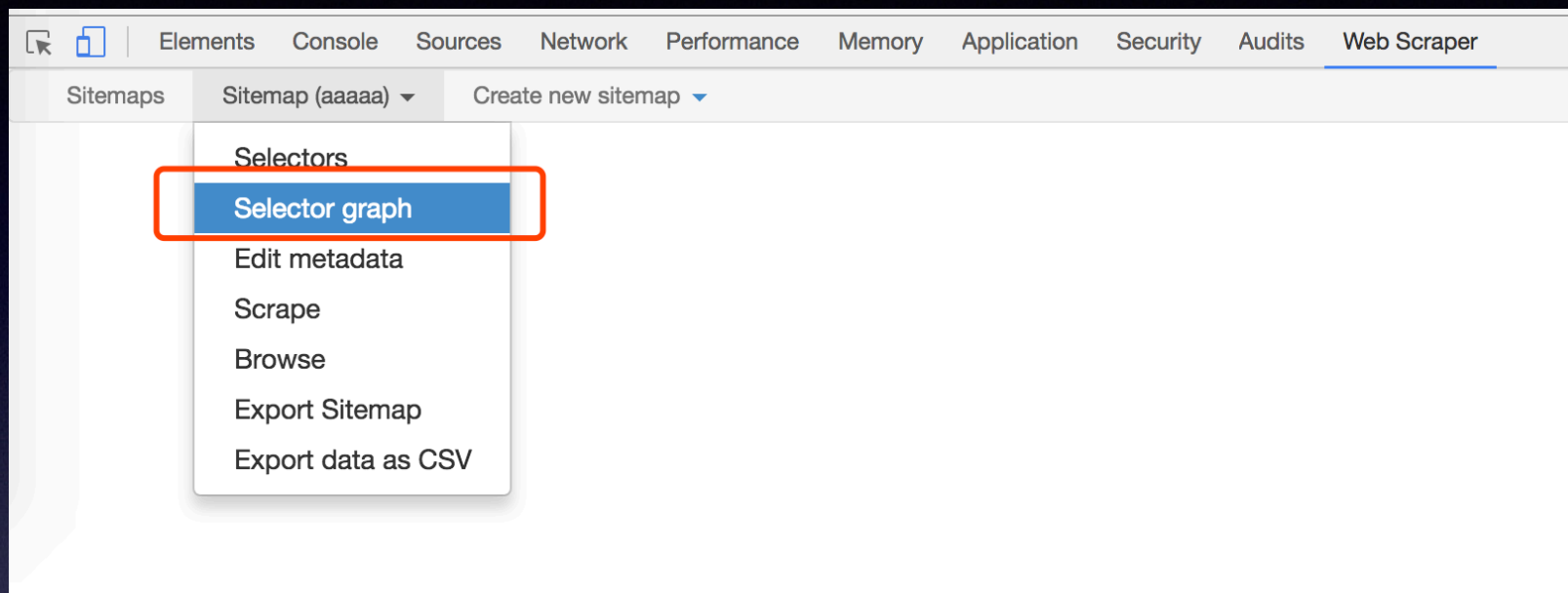
父选择器已经Delay





# 要点解析

## Selector graph 选择器树





# 要点解析

Element 选择器深究

如果要用 Element 选择器抓取数据

**必须创建对应的子选择器！**



# csv 文件讲解

intro	price	style	address	source
整租   石景山区八角中里南向 价格便宜 中间楼层 屋里干净	3200元/月	1室1厅1卫 45m²	八角 八角中里出租 距1号线八角游乐园站938米	来自经纪人： 我爱我家 - 八角南里店买卖B组 李勤 李勤 6年 我爱我家 58同城认证商家
整租   望京SOHO旁 望京西园一区 南向大一居出租 520	5200元/月	1室1厅1卫 65m²	望京 望京西园一区出租 距15号线望京站807米	来自经纪人： 中天置地 - 中天置地望京店 郭正军 郭正军 1年 中天置地 58同城认证商家
整租   亦庄 青年公寓 1室1厅	2700元/月	1室1厅1卫 28m²	亦庄 青年公寓	null
整租   泰和宜园 精装修 南北通透 家电家具齐全 出租300	3000元/月	2室1厅1卫 92m²	马坡 泰和宜园出租	来自经纪人： 中联置家 - 中联置家龙湖店D组 郑兴飞 郑兴飞 1年 中联置家 58同城认证商家
单间   龙跃苑一区 4室1厅1卫 转租房，房东住主卧	1650元/月	次卧(4室) 18m²	回龙观 龙跃苑一区出租	null
单间   亦庄线次渠南站次渠南里小区紧邻京东总部酒仙网鸿坤云时	1350元/月	次卧(3室) 16m²	次渠 南里出租	来自经纪人： 美丽屋资产 田森 田森 1年 美丽屋资 58同城认证商家
整租   六里桥南里 1室1厅 20平米 精装修 押一付一	2600元/月	1室1厅1卫 20m²	广安门内 六里桥南里出租 距10号线六里桥站232米	来自经纪人： 美丽家园 张自然 张自然 1年 美丽家园 58同城认证商家
单间   当天订房送3天房租地铁六里桥站免所有杂费押一付一免费	1980元/月	次卧(3室) 16m²	七里庄 六里桥家园出租	来自经纪人： 韩静 韩静 1年 58同城认证商家
单间   次渠锦园次卧出租 紧邻地铁 全新入住	1300元/月	次卧(3室) 15m²	亦庄东区 次渠锦园出租	来自经纪人： 发源地产 包梅珍 包梅珍 1年 发源地产 58同城认证商家
整租   望京南地铁站 大山子 高家园 精装复试一居出租 随时	3800元/月	1室1厅1卫 40m²	酒仙桥 高家园五区出租 距14号线高家园站171米	来自经纪人： 广发地产 - 高家园店 凡梦豪 凡梦豪 1年 广发地产 58同城认证商家
整租   定福家园南里3区 精装大两居腿儿这到地铁口 好住不贵	4500元/月	2室1厅1卫 77m²	定福庄 定福家园南里3号院出租 距6号线褡裢坡站167米	来自经纪人： 蛋壳公寓 - 蛋壳公寓朝阳门店 高永华 高永华 1年 蛋壳公寓 58同城认证商家
整租   出租香江北岸 精装大开间 紧邻地铁14号线与15号线	3500元/月	1室1厅1卫 56m²	孙河 康营家园12期出租 距15号线孙河站919米	来自经纪人： 蕴实地产 - 康营家园总店 赵贺宁 赵贺宁 3年 蕴实地产 58同城认证商家
单间   个人出租大主卧1850元 独立卫生间 招稳定年轻（中	1850元/月	主卧(3室) 35m²	天通苑 天通苑西三区出租	null
整租   京旺家园百平米大两居出租 地铁14号将台站 15号崔	5300元/月	2室1厅1卫 109m²	望京 京旺家园出租 距15号线崔各庄站3160米	来自经纪人： 昌胜源地产 - 昌胜源地产京旺家园店 田晶 田晶 1年 昌胜源地 58同城认证商家

1、web-scraper-order： 由于排序

2、web-scraper-start-url： start url

3、其他列： 对应的抓取信息



数据不一致？

正常！



# 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！！！！**