

六、统计列车时刻表

实例：北京列车时刻表

<http://qq.ip138.com/train/beijing/BeiJing.htm>

目标效果图

1	web-scraper-order	web-scraper-start-url	车次	列车类型	始发站	始发时间	经过站	经过站	经过站	终点站	到达时间
2	1529897207-778	http://qq.ip138.com/train/beijing/BeiJing.htm	6452	普客	杨村	07:28	北京	null	null	北京	10:59
3	1529897207-789	http://qq.ip138.com/train/beijing/BeiJing.htm	D19	动车	北京	14:18	北京	null	null	长春	20:44
4	1529897207-793	http://qq.ip138.com/train/beijing/BeiJing.htm	D22	动车	珥春	12:43	北京	null	null	北京	22:30
5	1529897207-888	http://qq.ip138.com/train/beijing/BeiJing.htm	K4729	快速	北京	22:36	北京	null	null	海拉尔	07:17
6	1529897207-834	http://qq.ip138.com/train/beijing/BeiJing.htm	D7	动车	北京	17:45	北京	null	null	沈阳北	22:53
7	1529897207-883	http://qq.ip138.com/train/beijing/BeiJing.htm	K43	快速	敦煌	07:30	北京	null	null	北京	19:42
8	1529897207-790	http://qq.ip138.com/train/beijing/BeiJing.htm	D2	动车	沈阳	09:13	北京	null	null	北京	14:08
9	1529897207-879	http://qq.ip138.com/train/beijing/BeiJing.htm	K411	快速	北京	23:47	北京	null	null	威海	14:42
10	1529897207-988	http://qq.ip138.com/train/beijing/BeiJing.htm	T63	特快	北京	19:57	北京	null	null	合肥	07:40
11	1529897207-806	http://qq.ip138.com/train/beijing/BeiJing.htm	D34	动车	大连	09:08	北京	null	null	北京	15:12
12	1529897207-889	http://qq.ip138.com/train/beijing/BeiJing.htm	K473	快速	北京	16:16	北京	null	null	昆明	12:14
13	1529897207-792	http://qq.ip138.com/train/beijing/BeiJing.htm	D21	动车	北京	06:39	北京	null	null	珥春	17:19
14	1529897207-890	http://qq.ip138.com/train/beijing/BeiJing.htm	K4730	快速	海拉尔	23:45	北京	null	null	北京	08:53
15	1529897207-843	http://qq.ip138.com/train/beijing/BeiJing.htm	K1177	快速	北京	13:35	北京	null	null	银川	08:13
16	1529897207-945	http://qq.ip138.com/train/beijing/BeiJing.htm	K885	快速	天津	21:38	北京	null	null	兰州西	05:42
17	1529897207-1031	http://qq.ip138.com/train/beijing/BeiJing.htm	Z62	直特	长春	21:58	北京	null	null	北京	06:08

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

车次↓	列车类型	始发站	始发时间	经过站	经过站 到达时间	经过站 发车时间	终点站	到达时间
Z9	直达特快	北京	19:08	北京	当天19:08	19:08	杭州	09:29
Z80/Z81	直达特快	大连	19:18	北京	第2日06:03	06:03	北京	06:03
Z8	直达特快	青岛北	21:00	北京	第2日06:10	06:10	北京	06:10
Z79/Z82	直达特快	北京	20:25	北京	当天20:25	20:25	大连	07:50
Z72	直达特快	三明北	14:45	北京	第2日06:28	06:28	北京	06:28
Z71	直达特快	北京	18:56	北京	当天18:56	18:56	三明北	10:55
Z7	直达特快	北京	22:10	北京	当天22:10	22:10	青岛北	06:35
Z64	直达特快	长春	22:33	北京	第2日07:50	07:50	北京	07:50
Z63	直达特快	北京	20:50	北京	当天20:50	20:50	长春	06:12
Z62	直达特快	长春	22:00	北京	第2日06:08	06:08	北京	06:08
Z61	直达特快	北京	22:10	北京	当天22:10	22:10	长春	06:37
Z601	直达特快	北京	08:30	北京	当天08:30	08:30	菏泽	13:55

6

7

Elements

Console

Sources

Network

Performance

Memory

Application

Security

Audits

Web Scraper

Sitemaps

Sitemap (table-test)

Create new sitemap

Id

table-bj

1

Type

Table

2

Selector

Select

Element preview

Data preview

div#checilist table

3

Header row selector

Select

Element preview

thead tr

Data rows selector

Select

Element preview

tbody tr

Delay (ms)

Multiple

4

2000

5

Parent Selectors

_root

3、设置 selector

☒ Multiple

Delay (ms)

2000

Parent Selectors

_root

Table columns

Column	Result key	Include into result
车次↓	车次↓	<input checked="" type="checkbox"/>
列车类型	列车类型	<input checked="" type="checkbox"/>
始发站	始发站	<input checked="" type="checkbox"/>
始发时间	始发时间	<input checked="" type="checkbox"/>
经过站	经过站	<input checked="" type="checkbox"/>
经过站到达时间	经过站到达时间	<input checked="" type="checkbox"/>
经过站发车时间	经过站发车时间	<input checked="" type="checkbox"/>
终点站	终点站	<input checked="" type="checkbox"/>
到达时间	到达时间	<input checked="" type="checkbox"/>

Save selector

Cancel

8

9

CSV 文件讲解

1	web-scraper-order	web-scraper-start-url	车次	列车类型	始发站	始发时间	经过站	经过站	经过站	终点站	到达时间
2	1529897207-778	http://qq.ip138.com/train/beijing/BeiJing.htm	6452	普客	杨村	07:28	北京	null	null	北京	10:59
3	1529897207-789	http://qq.ip138.com/train/beijing/BeiJing.htm	D19	动车	北京	14:18	北京	null	null	长春	20:44
4	1529897207-793	http://qq.ip138.com/train/beijing/BeiJing.htm	D22	动车	珲春	12:43	北京	null	null	北京	22:30
5	1529897207-888	http://qq.ip138.com/train/beijing/BeiJing.htm	K4729	快速	北京	22:36	北京	null	null	海拉尔	07:17
6	1529897207-834	http://qq.ip138.com/train/beijing/BeiJing.htm	D7	动车	北京	17:45	北京	null	null	沈阳北	22:53
7	1529897207-883	http://qq.ip138.com/train/beijing/BeiJing.htm	K43	快速	敦煌	07:30	北京	null	null	北京	19:42
8	1529897207-790	http://qq.ip138.com/train/beijing/BeiJing.htm	D2	动车	沈阳	09:13	北京	null	null	北京	14:08
9	1529897207-879	http://qq.ip138.com/train/beijing/BeiJing.htm	K411	快速	北京	23:47	北京	null	null	威海	14:42
10	1529897207-988	http://qq.ip138.com/train/beijing/BeiJing.htm	T63	特快	北京	19:57	北京	null	null	合肥	07:40
11	1529897207-806	http://qq.ip138.com/train/beijing/BeiJing.htm	D34	动车	大连	09:08	北京	null	null	北京	15:12
12	1529897207-889	http://qq.ip138.com/train/beijing/BeiJing.htm	K473	快速	北京	16:16	北京	null	null	昆明	12:14
13	1529897207-792	http://qq.ip138.com/train/beijing/BeiJing.htm	D21	动车	北京	06:39	北京	null	null	珲春	17:19
14	1529897207-890	http://qq.ip138.com/train/beijing/BeiJing.htm	K4730	快速	海拉尔	23:45	北京	null	null	北京	08:53
15	1529897207-843	http://qq.ip138.com/train/beijing/BeiJing.htm	K1177	快速	北京	13:35	北京	null	null	银川	08:13
16	1529897207-945	http://qq.ip138.com/train/beijing/BeiJing.htm	K885	快速	天津	21:38	北京	null	null	兰州西	05:42
17	1529897207-1031	http://qq.ip138.com/train/beijing/BeiJing.htm	Z62	直特	长春	21:58	北京	null	null	北京	06:08

1、web-scraper-order：由于排序

2、web-scraper-start-url：start url

3、表头：表格真实数据

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！！！！**