

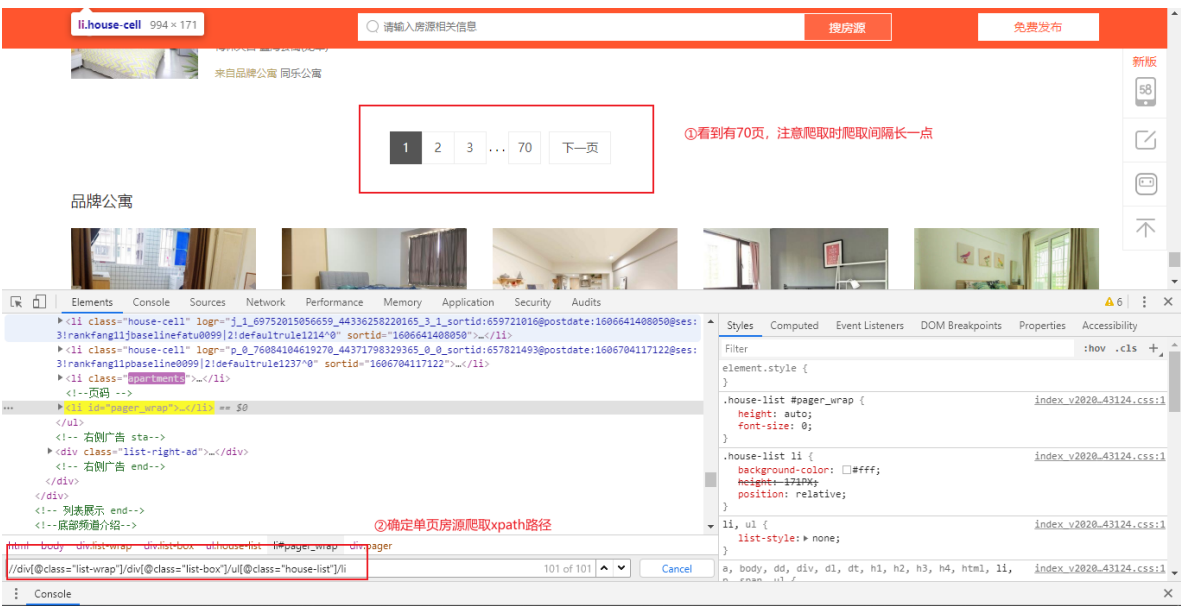
任务要求

1、爬取深圳市龙华区在58同城上的租房信息（租房价格2000-4000）

①打开58同城，确定筛选条件



②确定页数以及单页xpath提取规则



③确定每页url构成规则



完成以上分析步骤之后，就可以开始编写爬虫脚本：

①找到反爬参数headers

②解析页面，找到参数1——图片链接

③脚本编写完毕！但是ip访问次数受到限制

```
63 def save_data_to_csv(url):...
84
85
86 if __name__ == '__main__':
87     f = open("深圳龙华房源信息.csv", mode='a+', encoding='utf-8-sig', newli
88     csv_write = csv.writer(f)
89     csv_write.writerow(['名称', '房屋情况', '区/县', '小区', '价格', '图片链接'])
90     for j in range(1,71):
91         if (j == 1):
            if __name__ == '__main__' > for j in range(1,71) > if (j == 1) > try
                list index out of range
                list index out of range
                list index out of range
                list index out of range
                list index out of range
                list index out of range
                list index out of range
                list index out of range
                list index out of range
                list index out of range
                list index out of range
                list index out of range
                list index out of range
```

正常情况下只能获取到1页房源信息

④购买付费ip代理，打造ip代理池

检查ip是否有效果，可以查看这个网站：<https://www.ip.cn/>

← → ↺ ↻ ☆

https://www.kuaidaili.com/cart?p=month

☆

快代理
kuaidaili.com

免费代理 购买代理 产品列表 文档中心 开发者 帮助和支持

晚上好, 2142781703@qq.com 会员

购买开放代理

产品选择

套餐选择

时间选择

购买时长

价格

代理用途 (选填)

支付方式

私密代理

隧道代理

独享代理

开放代理

普通套餐

VIP套餐

SVIP套餐

专业版套餐

包天

包周

包月

包年 优惠

1

天

有效期至: 2020-12-01 22:15:39

10

元

爬虫

选择后方便我们为您提供技术支持

补充描述

支付宝 ALIPAY

微信支付

对公转账

产品规格

每日IP数50000+

高匿名; 支持HTTPS;

稳定性一般, 可用率20%~40% ⑦

API调用频率: 5秒1次

了解详情>>

购买后: (建议购买svip)

✔ 您已成功购买：开放代理普通服务包天 1天

订单号: 930674583357868

付款金额: 10.00元

付款时间: 2020-11-30 22:17:50

服务有效期至: 2020-12-01 22:17:38

[查看我的订单](#) | [马上提取代理](#) | [生成API链接](#)

欢迎加入快代理技术交流群，快代理售后工程师会给您提供技术支持：



QQ群号: 719916596

[加入QQ群](#)

直接提取ip：

[←](#) [→](#) [↺](#) [↻](#) [☆](#) [https://www.kuaidaili.com/fetch/?orderid=930674583357868](#)当前可用开放代理数: **6953** (28秒前更新)

订单号* 930674583357868

[查看我的订单](#) [生成API链接](#)

提取数量* 100

[数量该怎么填?](#)

匿名度

☒ 高匿名☐ 匿名☐ 透明[了解匿名度](#)

代理协议

☐ HTTP☒ HTTPS (同时也支持HTTP)[何时用HTTPS?](#)

代理稳定性

☒ 一般☐ 稳定 **VIP**☐ 非常稳定 **SVIP**[了解稳定性和可用率](#)

支持访问 https 网站，可用率约 20% ~ 40%

[展开更多筛选项](#)[立即提取](#)

复制ip到ip.txt文件替换：

[←](#) [→](#) [↺](#) [↻](#) [☆](#) [https://www.kuaidaili.com/getproxy/?orderid=930674583357868&num=100&an_ha=1&protocol=2&quality=0&area=&ar...](#) [☆](#)

开放代理提取结果

订单号: 930674583357868

有效期: 1天 (2020-11-30 22:17:38 ~ 2020-12-01 22:17:38)

提取时间: 2020-11-30 22:25:10

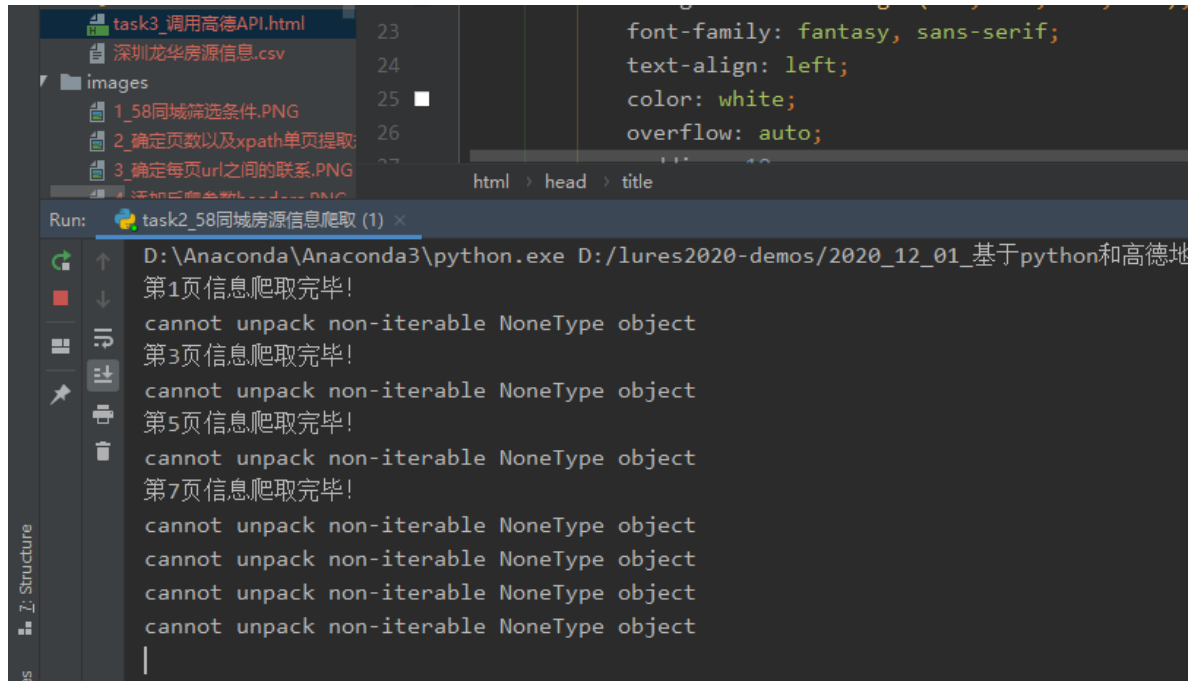
当前可用代理: 3552 本次提取: 100

今天不重复IP: 100

123.169.127.130:9999
182.46.101.84:9999
115.221.240.233:9999
110.243.16.64:9999
115.221.242.255:9999
36.249.48.48:9999
114.239.199.38:9999
175.43.35.54:9999
218.62.124.125:9999
123.163.121.90:9999
183.166.21.163:9999
115.221.240.25:9999
222.189.191.160:9999[下载为文本](#)[复制到剪贴板](#)

已复制

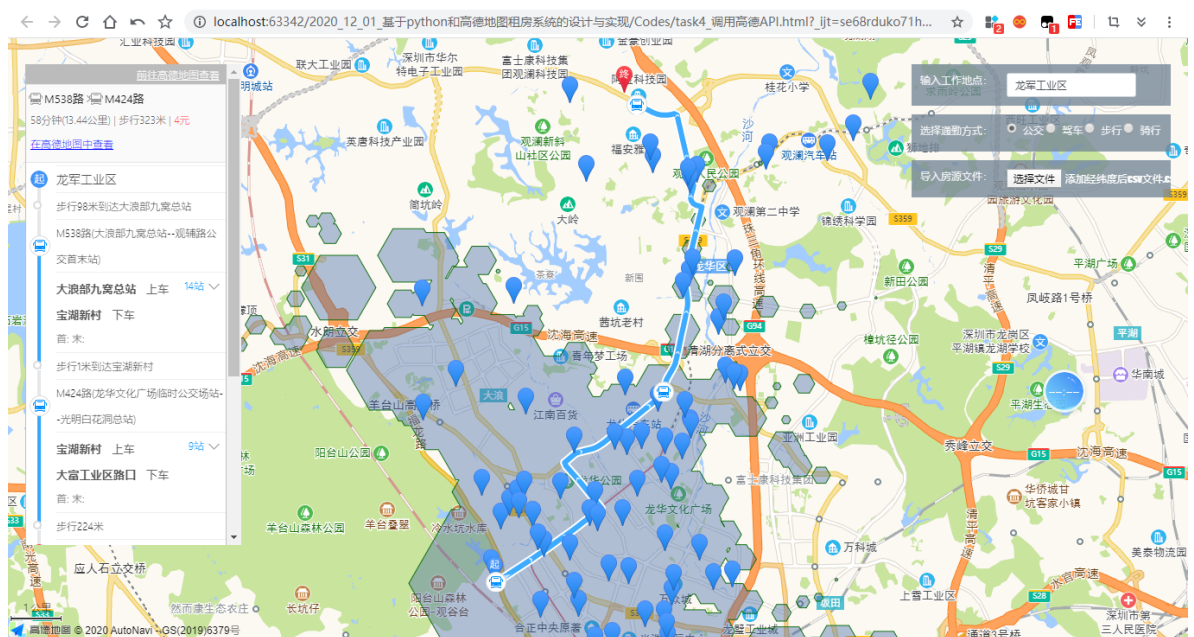
爬取过程很神奇，不知道什么问题：



估计是ip的问题，每次爬都会有这个，只是位置不一样

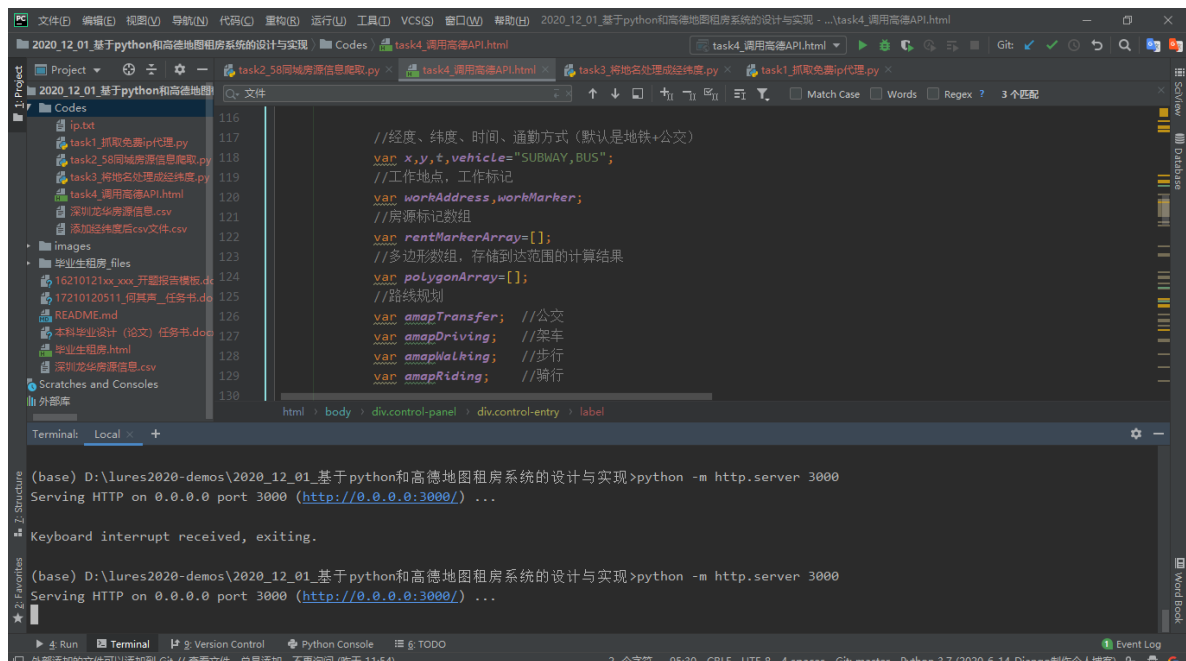
2、调动高德地图api，展示房源

地址：<https://lbs.amap.com/api/javascript-api/example/bus-info/arrival-range/>



高德展示信息

下面是Pycharm运行截图：

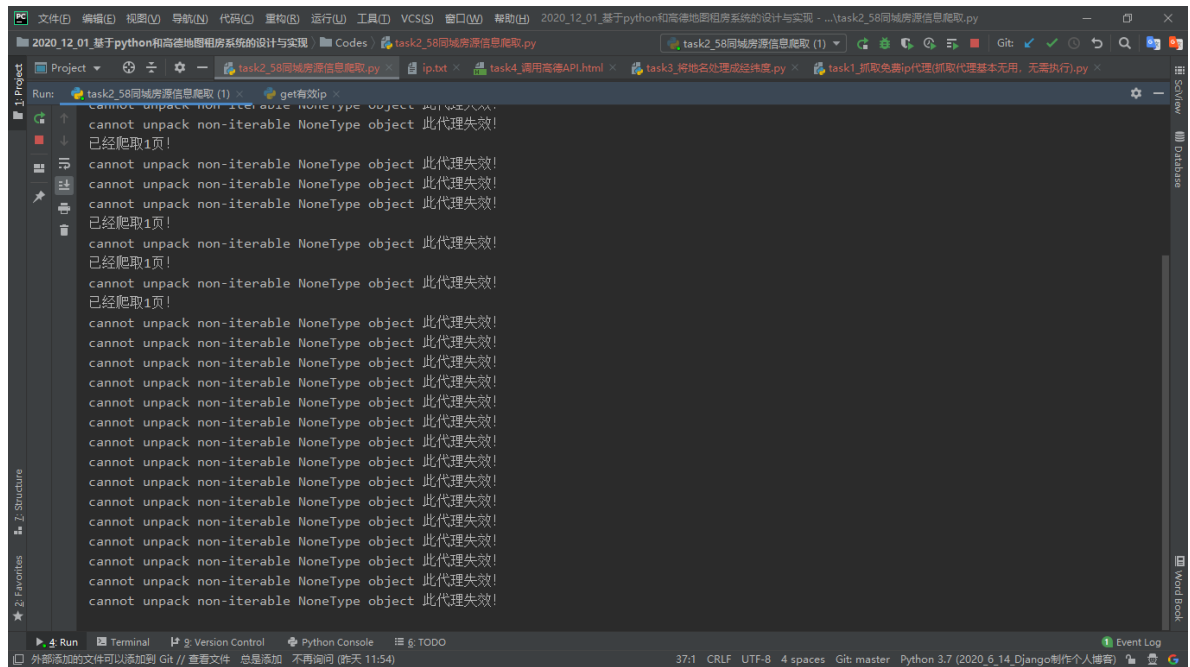


注意事项

- ①平时不用的时候，不要爬，要不之前爬的csv文件会被覆盖，演示的时候自己购买ip代理池，按照文档要求覆盖ip.txt文件内容即可
- ②申请的高德API默认免费次数是6000次，注意使用次数，超过次数需要重新申请，否则报错
- ③目前测试58爬取没问题，在网页结构不变时，代码都有效
- ④缺少的python第三方模块包，自己下载按照（通过pip命令）
- ⑤文件夹中的task1_爬取ip代理.py别执行，因为爬下来的ip都是失效的



添加多进程，加快爬取速度：



终端运行: `python -m http.server 3000`

浏览器打开: `http://0.0.0.0:3000/`

鼠标右键运行: `task4_调用高德地图API.html` 导入"添加经纬度后csv文件.csv"可以看到效果