

爬虫：定向/非定向(一个全都爬一个爬取指定页面)

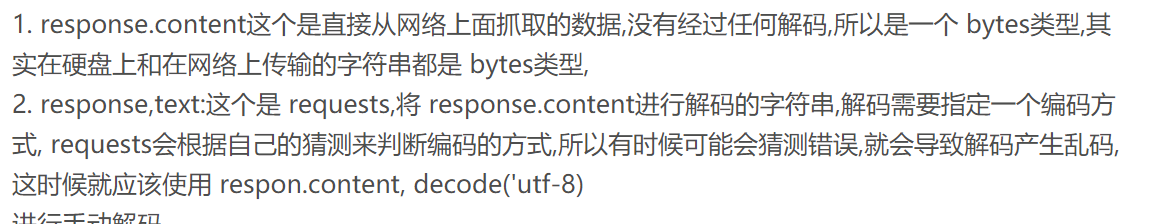
1. 下载页面：

<https://www.autohome.com.cn/all/>（网站返回字符串即页面的源代码）

1. 筛选：

正则表达式

--------------------开源模块------------------



1. requests:

pip intall requests

response = request.get(“<https://www.autohome.com.cn/all/>”)

response.text

1. beautisoup模块

pip install Beautifulsoup4

from bs4 import Beautifulsoup

soup = BeautiSoup(response.txt, parser=’html.parser’)

将Html转化为对象

target = soup.find(id=’XXXXXXXXX’)

print(target)

soup.find(‘div’) 找第一个孩子为div的