# Python爬虫

# 1.Requests库的使用

## 1-1 京东网站爬虫



## 1-2 Amazon网站爬虫

进行了浏览器访问模拟，即人为设置requests的headers字段的值



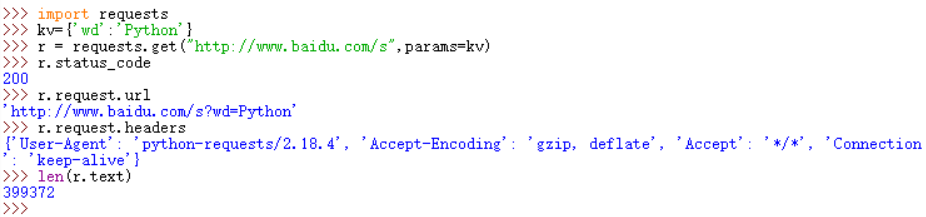
调用re.quest.get(url)方法后网站会返回一个response对象r，查看请求头部的方法：



## 1-3 使用requests库向百度和360提交关键词并得到查询结果

百度——

控制台：

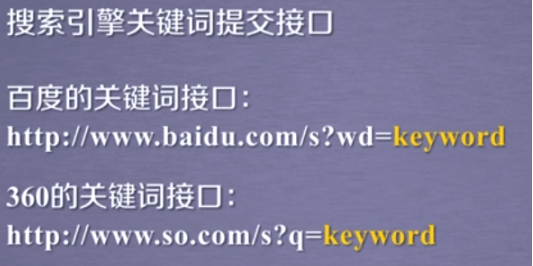


代码:



为啥加/s ？

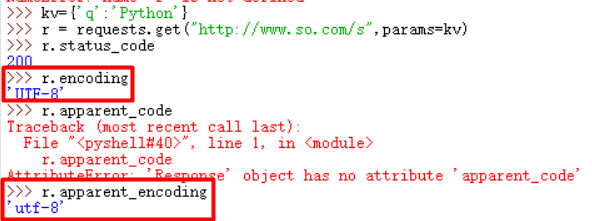
因为：



360——



控制台——



所以代码中即使没有r.encoding = r.apparent\_encoding也是可以的，不过加上什么时候也没毛病。

### 1-4 网络图片的爬取



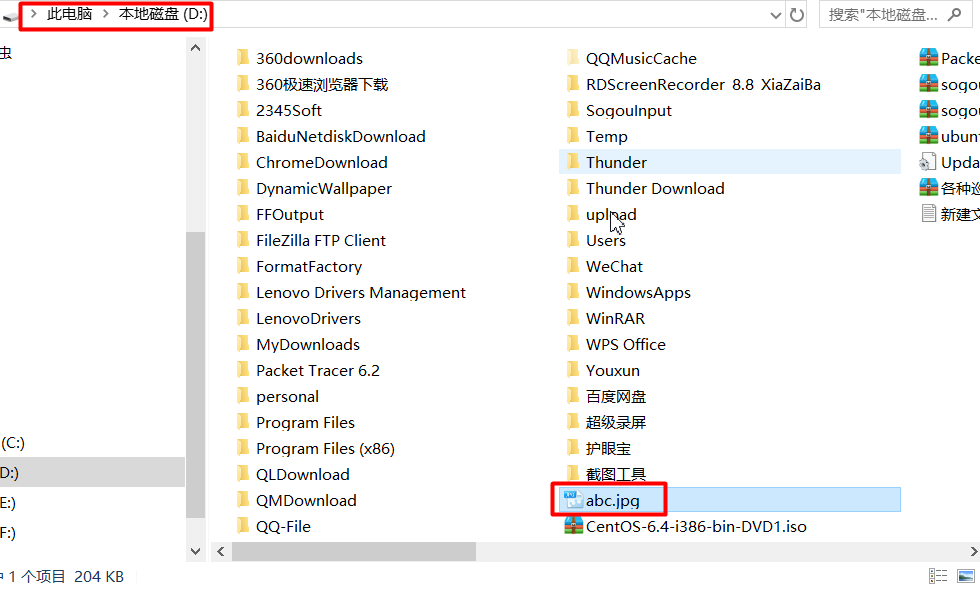
从网络上找一张图片，查看其属性，会看到图片的url地址示例如下：

<https://timgsa.baidu.com/timg?image&quality=80&size=b9999_10000&sec=1542993991904&di=739fd48f89fe5dff35fe2e72e58051dc&imgtype=0&src=http%3A%2F%2Fimglf1.ph.126.net%2FEL41V1hr6QzSsKaaccug4Q%3D%3D%2F6630935824305511510.jpg>

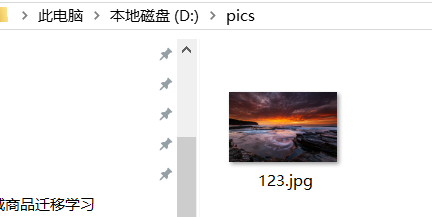
控制台——

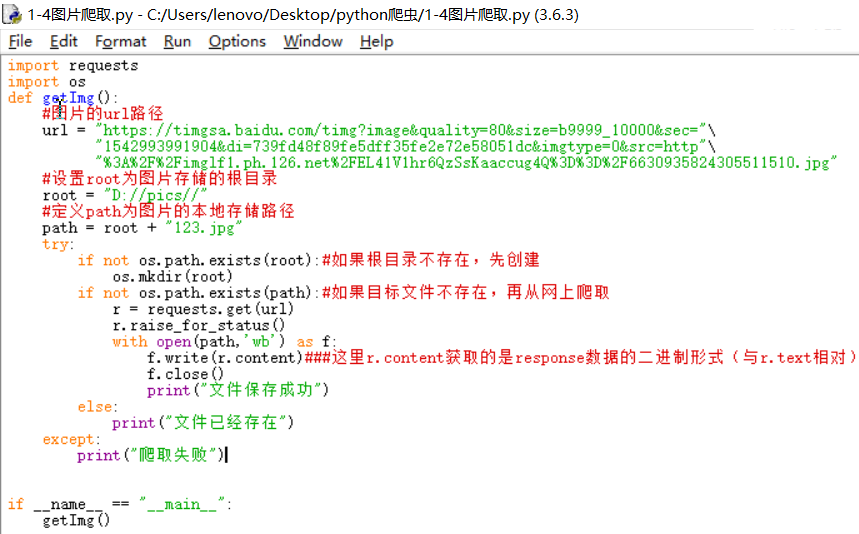


如下：在D：下生成了abc.jpg



全代码——





### 1-5 IP地址归属地查询

一个叫ip138的网站可以提供这个功能

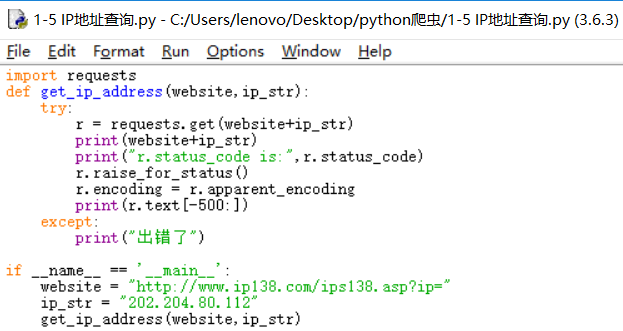




控制台——



全代码——



url的写法是根据下面来的：



输出:





就是url+6种操作

