Python爬虫常用模块（一）：requests模块

2017-10-12 [每日一Python](https://mp.weixin.qq.com/s?__biz=MzI2Mjg5NDA4Ng==&mid=2247483695&idx=1&sn=010185730a81d4851b6c49125a746b12&chksm=ea456646dd32ef50e82ffe5eeb7ff61e70a5b7747887af577c3c4b9d82ab868cdeb3f74a18d5&scene=27&key=c5e6aaaa5d1e2425c7d19b75b0cca19634af270203b447eaf656c8b516c3e49364e81d6c59835934c0e93197e4d2301973db607c2b1cdaa623c097fa91d58fb848261147d26db00f42f1ae76a9619ed9&ascene=7&uin=Nzk3NjIyNQ%3D%3D&devicetype=Windows+7&version=6205051a&pass_ticket=pVlB793F3uK4YJsVaPHBeSBPYMcGIImT6eadIav5PlM%3D&winzoom=1##)

爬虫

爬虫的基本框架是获取HTML页面信息，解析页面信息，保存结果，requests模块是用于第一步获取HTML页面信息； requests库用于爬取HTML页面，提交网络请求，基于urllib,但比urllib更方便；

(一）、发送请求

对应http的不同请求类型，requests库有不同的方法：

1.requests.get():

 获取HTML网页的主要方法，对应于HTTP的GET

2.requests.post():

向HTML网页提交POST请求的方法，对应于HTTP的POST

3.requests.head():

获取HTML网页头信息的方法，对应于HTTP的HEAD

4.requests.put():

向HTML网页提交PUT请求，对应于HTTP的PUT

5.requests.patch():

向HTML网页提交局部修改请求，对应于HTTP的PATCH

6.requests.delete():

向HTML页面提交删除请求，对应于HTTP的DELETE

(二）、响应内容

当我们用上述方法向HTML页面发送请求后，会获得一个Response对象；

例如 r =  requests.get('https://www.python.org/')

 r 就是一个Response对象，我们可以从这个对象中获取我们想要的信息；

Response对象的一些属性有：

1. r.status\_code:  HTTP请求的返回状态，200表示连接成功，404表示失败
2. r.text:  URL所对应的页面内容（字符串形式）
3. r.content:  页面内容的二进制形式
4. r.encoding:  从HTTP header 中推测的响应内容的编码形式
5. r.apparent\_encoding:  从内容分析出的响应内容的编码形式
6. r.headers:  以字典形式显示的服务器头信息

（三）、实例

