

## 8 爬虫 使用urllib2获取数据

### Python中的Urllib2

<https://docs.python.org/2/library/urllib2.html>

我的python版本：2.7

#### 发起GET请求

<http://kaoshi.edu.sina.com.cn/college/scorelist?tab=batch&wl=1&local=2&batch=8&year=2013>

---

```
request = urllib2.Request(url=url, headers=headers)
response = urllib2.urlopen(request, timeout=20)
result = response.read()
```

---

#### 发起POST请求

<http://shuju.wdzy.com/plat-info-59.html>

---

```
data = urllib.urlencode({'type1': x, 'type2': 0, 'status': 0, 'wdzyPlatId': int(platid)})
request = urllib2.Request('http://shuju.wdzy.com/depth-data.html', headers)
opener = urllib2.build_opener(urllib2.HTTPCookieProcessor())
response = opener.open(request, data)
result = response.read()
```

---

#### 处理返回结果

Html：BeautifulSoup，需要有一些CSS基础

API：JSON

<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

我的个人网站：<http://zhanghonglun.cn/>

雪晴数据网：<http://www.xueqing.tv/course/73>

毕业僧：<http://www.biyeseng.cn/front/couinfo/34>

研趣网：<http://www.rinterest.cn/course/72>

优酷：[http://v.youku.com/v\\_show/id\\_XMTc1NDY1NjM2QA==.html?f=28356198&from=y1.7-3](http://v.youku.com/v_show/id_XMTc1NDY1NjM2QA==.html?f=28356198&from=y1.7-3)

腾讯：<http://v.qq.com/vplus/334d685e3cea789f67d5bd9617482a49/foldervideos/1ad000101bqho2b>

土豆：<http://www.tudou.com/listplay/IstkSpQCB1o.html>

爱奇艺：[http://www.iqiyi.com/v\\_19ruejtj7x.html?list=19rkraftq](http://www.iqiyi.com/v_19ruejtj7x.html?list=19rkraftq)