

前后端交互都是通过http1.1的request发送post/get请求实现的, 然后返回response
前端解析

网易云课堂<0基础Python实战：爬虫课程免费试听---第三节课程：真实世界中的网页解析>总结

1. 爬取多页数据：

观察每一页的地址, 找出规律, 把地址列表构造成list型, 再循环遍历每一页, 并在爬取完一页后

sleep几秒钟防止被反扒

2. 爬取登陆后才能看到的数据：

伪造http的request的heads的User-Agent和cookie信息(可通过浏览器的Network找到),

就会以登陆状态访问个人空间, 然后爬取个人空间的数据

3. 切换移动端爬取

如果有网页爬取不到数据, 可以通过浏览器的伪装成移动端浏览器爬取

(右键点击'检查'后的左上角第二个图标), 移动端浏览器会照顾各个版本的浏览器, 所以考虑不多