

requests进阶概述

我们在之前的爬虫中其实已经使用过headers了. header为HTTP协议中的请求头. 一般存放一些和请求内容无关的数据. 有时也会存放一些安全验证信息. 比如常见的User-Agent, token, cookie等.

通过requests发送的请求, 我们可以把请求头信息放在headers中. 也可以单独进行存放, 最终由requests自动帮我们拼接成完整的http请求头.

本章内容:

1. 模拟浏览器登录->处理cookie
2. 防盗链处理-> 抓取梨视频数据
3. 代理 -> 防止被封IP

综合训练:

抓取网易云音乐评论信息