即使第一次和服务器连接后并且登录成功后,第二 由来:http无状态 次请求服务器依然不能知道当前请求是哪个用户 第一次登录后服务器返回一些数据(cookie)给浏 什么是cookie 览器,然后浏览器保存在本地,当该用户发送第二 目的:解决http无状态问题 次请求的时候,就会自动的把上次请求存储的 cookie数据自动的携带给服务器,服务器通过浏览 器携带的数据就能判断当前用户是哪个了 存储大小有限,一般不超过4KB NAME: cookie的名字 VALUE: cookie的值 Expires: cookie的过期时间 cookie格式 Set-Cookie: NAME=VALUE; Expires/Max-age=DATE; Path=PATH; Domain=DOMAIN_NAME; SECURE Path: cookie作用的路径 Domain: cookie作用的域名 SECURE: 是否只在https协议下起作用 from urllib import request, parse from http.cookiejar import CookieJar headers = { 'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_13_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/ 68.0.3440.106 Safari/537.36' # 1. 登陆 def get_opener(): # 1.1 创建一个cookiejar对象 cookiejar = CookieJar() # 1.2 使用cookiejar创建一个HTTPCookieProcess对象 handler = request.HTTPCookieProcessor(cookiejar) # 1.3 使用上一步创建的handler创建一个opener opener = request.build_opener(handler) CookieJar: 管理HTTP cookie值、存储HTTP请求生成的cookie、向传出的HTTP请求添加cookie return opener cookie 的对象。整个cookie都存储在内存中,对CookieJar实例进行垃圾回收后cookie也将丢失。 def login_renren(opener): # 1.4 使用opener发送登陆的请求(人人网的邮箱和密码) data = {'email': '970138074@qq.com', 'password':'pythonspider'} data = parse.urlencode(data).encode('utf-8') login_url = 'http://www.renren.com/PLogin.do' req = request.Request(login_url, headers=headers, data=data) opener.open(req) # 2. 访问个人主页 def visit_profile(opener): url = 'http://www.renren.com/880151247/profile' req = request.Request(url, headers=headers) resp = opener.open(req) with open('profile.html', 'w', encoding='utf-8') as fp: fp.write(resp.read().decode('utf-8')) http.cookiejar模块 FileCookieJar (filename,delayload=None,policy=None):从CookieJar派生而来,用来创建FileCookieJar实例, 检索cookie信息并将cookie存储到文件中。filename是存储cookie的文件名。delayload为True时支持延迟访问访 问文件,即只有在需要时才读取文件或在文件中存储数据。 from urllib import request from http.cookiejar import MozillaCookieJar cookiejar = MozillaCookieJar('cookie.txt') handler = request.HTTPCookieProcessor(cookiejar) 保存cookie到本地 opener = request.build_opener(handler) resp = opener.open('http://httpbin.org/cookies/set? course=abc') MozillaCookieJar (filename,delayload=None,policy=None):从FileCookieJar派生而来,创建与Mozilla浏览器 cookiejar.save(ignore_discard=True) cookies.txt兼容的FileCookieJar实例。 from http.cookiejar import MozillaCookieJar cookiejar = MozillaCookieJar('cookie.txt')

LWPCookieJar (filename,delayload=None,policy=None):从FileCookieJar派生而来,创建与libwww-perl标准的

Set-Cookie3文件格式兼容的FileCookieJar实例。

从本地加载cookie

cookiejar.load(ignore_discard=True)

for cookie in cookiejar: print(cookie)