

1.get传参

(1)汉字报错:解释器ascii没有汉字 url汉字转码

`urllib.parse.quote safe="string.printtable"`

(2)字典传参

`urllib.parse.urlencode()`

post:

`urllib.request.urlopen(url,data = "服务器接受的数据")`

handler:处理器的自定义:

User-Agent:

(1)模拟真实的浏览器发送请求:(1)百度批量搜索(2)检查元素(百度搜索useragent大全)

(2)`request.add_header`(动态添加head数据)

(3)响应头 `response.header`

(4)创建request:`urllib.request.Request(url)`

2.IP代理:

(1)免费的IP:时效性差,错误率高

(2)付费的IP:贵花钱,也有失效不能用的

IP分类:

透明:对方知道我们真实的ip

匿名:对方不知道我们真实的ip,知道了你使用了代理

高匿:对方不知道我们真是的IP.也不知道我们使用了代理

handler:

(1)系统的`urlopen()`不支持代理的添加

创建对应的处理器(handler)

1.代理处理器:`ProxyHandler`

2.拿着`ProxyHandler`创建opener:`build_opener()`

3.opener.open(url)就可以请求数据

auth认证handler

`Cookieshandler`

URLError

requests(第三方模块):简单易用

数据解析:

数据存储:json csv MongoDB resdis mysql