**urllib模块中的方法**

**1.urllib.request.urlopen(url[,data[,proxies]])**

打开一个url的方法，返回一个文件对象，然后可以进行类似文件对象的操作。本例试着打开google

[复制代码](javascript:void(0);)

>>> import urllib

>>> f = urllib. request.urlopen('http://www.google.com.hk/')

>>> firstLine = f.readline()   #读取html页面的第一行

>>> firstLine

'<!doctype html><html itemscope="" itemtype="http://schema.org/WebPage"><head><meta content="/images/google\_favicon\_128.png" itemprop="image"><title>Google</title><script>(function(){\n'

[复制代码](javascript:void(0);)

urlopen返回对象提供方法：

-         read() , readline() ,readlines() , fileno() , close() ：这些方法的使用方式与文件对象完全一样

-         info()：返回一个httplib.HTTPMessage对象，表示远程服务器返回的头信息

-         getcode()：返回Http状态码。如果是http请求，200请求成功完成;404网址未找到

-         geturl()：返回请求的url

**2.urllib. request.urlretrieve(url[,filename[,reporthook[,data]]])**

urlretrieve方法将url定位到的html文件下载到你本地的硬盘中。如果不指定filename，则会存为临时文件。

urlretrieve()返回一个二元组(filename,mine\_hdrs)

**临时存放：**

[复制代码](javascript:void(0);)

>>> filename = urllib. request.urlretrieve('http://www.google.com.hk/')

>>> type(filename)

<type 'tuple'>

>>> filename[0]

'/tmp/tmp8eVLjq'

>>> filename[1]

<httplib.HTTPMessage instance at 0xb6a363ec>

[复制代码](javascript:void(0);)

**存为本地文件:**

[复制代码](javascript:void(0);)

>>> filename = urllib. request.urlretrieve('http://www.google.com.hk/',filename='/home/dzhwen/python文件/Homework/urllib/google.html')

>>> type(filename)

<type 'tuple'>

>>> filename[0]

'/home/dzhwen/python\xe6\x96\x87\xe4\xbb\xb6/Homework/urllib/google.html'

>>> filename[1]

<httplib.HTTPMessage instance at 0xb6e2c38c>

[复制代码](javascript:void(0);)

**3.urllib. request.urlcleanup()**

清除由于urllib. request.urlretrieve()所产生的缓存

**4.urllib.** **request.quote(url)和urllib.quote\_plus(url)**

将url数据获取之后，并将其编码，从而适用与URL字符串中，使其能被打印和被web服务器接受。

>>> urllib.quote('http://www.baidu.com')

'http%3A//www.baidu.com'

>>> urllib.quote\_plus('http://www.baidu.com')

'http%3A%2F%2Fwww.baidu.com'

**5.urllib.unquote(url)和urllib.unquote\_plus(url)**

与4的函数相反。

**6.urllib.urlencode(query)**

将URL中的键值对以连接符&划分

这里可以与urlopen结合以实现post方法和get方法：

**GET方法：**

[复制代码](javascript:void(0);)

>>> import urllib

>>> params=urllib.urlencode({'spam':1,'eggs':2,'bacon':0})

>>> params

'eggs=2&bacon=0&spam=1'

>>> f=urllib.urlopen("http://python.org/query?%s" % params)

>>> print f.read()

[复制代码](javascript:void(0);)

**POST方法：**

>>> import urllib

>>> parmas = urllib.urlencode({'spam':1,'eggs':2,'bacon':0})

>>> f=urllib.urlopen("http://python.org/query",parmas)

>>> f.read()

Urllib.parse

urllib.parse.**urlparse**(*urlstring*, *scheme=''*, *allow\_fragments=True*)[¶](https://docs.python.org/3/library/urllib.parse.html#urllib.parse.urlparse)