1.爬取网页

from urllib import request

def download(url,fname):

html = request.urlopen(url)

with open(fname,'wb') as fobj:

while True:

data = html.read(4096)

if not data:

break

fobj.write(data)

if \_\_name\_\_ == '\_\_main\_\_':

download('http://f12.baidu.com/it/u=1702154046,3138678857&fm=72','/tmp/ha.jpg')

download('http://img3.duitang.com/uploads/item/201504/19/20150419H2315\_SLRyX.jpeg','/tmp/a.jpg')

2.模拟客户端

#为防止由于服务器限制,不能通过程序爬取页面,模拟使用firefox浏览

from urllib import request

url = 'http://127.0.0.1/'

header = {

'User-Agent':'Mozilla/5.0 (X11; Linux x86\_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/60.0.3112.113 Safari/537.36'

}

r = request.Request(url,headers=header)

html = request.urlopen(r)

#html = request.urlopen(r)

data = html.read()

print(data.decode('utf8'))

3.数据编码

#标准的URL只会允许一部分ASCII字符,比如数字、字母、部分符号等

#

# from urllib import request

#

# r = request.quote('hello 达内') #编码成浏览器可读字符

# print(r)

#

# ur = request.unquote(r)

# print(ur)

3.异常处理

from urllib import request,error

try:

html = request.urlopen('http://127.0.0.1/xyz/') #xyz不存在

except error.HTTPError as e:

print(e)

try:

html = request.urlopen('http://127.0.0.1/ban') #ban无访问权限

except error.HTTPError as e:

print(e)

1. 爬取网络图片

from urllib import request

import os

import re

from urllib import request

def download(url,fname):

html = request.urlopen(url)

with open(fname,'wb') as fobj:

while True:

data = html.read(1024)

if not data:

break

fobj.write(data)

def search\_url(fname,patt):

patt\_list =[]

cpatt = re.compile(patt)

with open(fname) as fobj:

for line in fobj:

m = cpatt.search(line)

if m:

item = m.group()

patt\_list.append(item)

return patt\_list

if \_\_name\_\_ == '\_\_main\_\_':

img\_dirs = '/tmp/imgs'

if not os.path.exists(img\_dirs):

os.mkdir(img\_dirs)

download('http://www.tmooc.cn/','/tmp/tedu.html') #获得首页

img\_patt = 'http://[\w./-]+\.(jpg|jpeg|gif|png)' #匹配图片url

#\w匹配数字.字母.下划线.\.表示任意字符

# print(search\_url('/tmp/tedu.html',img\_patt))

img\_list = search\_url('/tmp/tedu.html',img\_patt)

for url in img\_list:

fname = url.split('/')[-1] #获得url中的图片名

#url = 'http://www.tedu.cn/logo.jpg'

#url.split('/')

#print(fname)

fname = os.path.join(img\_dirs,fname) #os.path.join 拼接

print(fname)

try:

download(url,fname)

except:

pass

5.paramiko

1)安装paramiko

pip install paramiko

1. 基础使用

SSHClient 创建用于连接ssh服务器的实例

>>>>>>ssh = paramiko.SSHClient()

paramiko.AutoAddPolicy 设置自动添加主机密钥

#ssh.connect 连接ssh服务器

#ssh.exec\_comand 在ssh服务器上执行指定命令

3)

import paramiko

import sys

import getpass

def rcmd(host,password,cmd,port=22,username='root'):

ssh = paramiko.SSHClient()

ssh.set\_missing\_host\_key\_policy(paramiko.AutoAddPolicy())

ssh.connect(

hostname = host,

username = username,

password = password,

port = port

)

stdin, stdout,stderr = ssh.exec\_command(cmd)

data = stdout.read().decode('utf8')

error = stderr.read().decode('utf8')

if data:

print('[%s :OUT]:\n%s' % (host,data))

if error:

print('[%s :ERROR]:\n%s' % (host,error))

ssh.close()

if \_\_name\_\_ == '\_\_main\_\_':

#rcmd('192.168.4.2','123456','id zhangsan')

pwd = getpass.getpass()

ipfile = sys.argv[1]

with open(ipfile) as fobj:

for line in fobj:

ip = line.strip()

rcmd(ip,pwd,sys.argv[2])