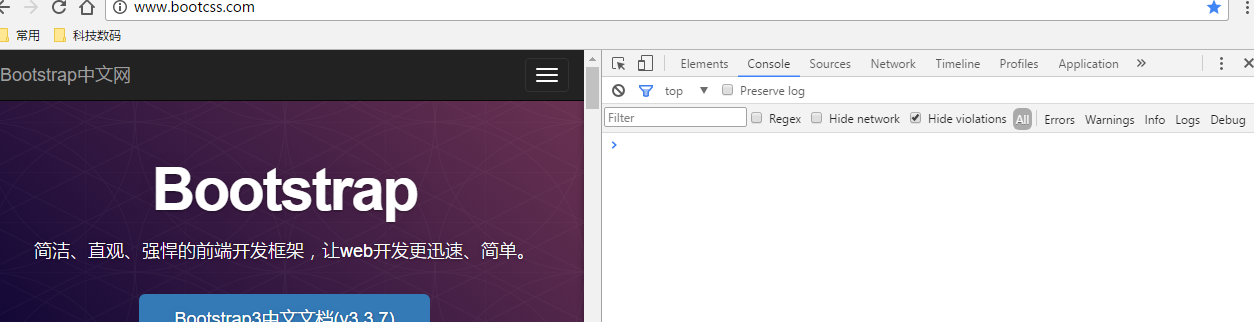
1. 打开chrome浏览器

输入<http://www.bootcss.com/> 截图为打开效果图



1. 按下键盘F12键

打开**Chrome DevTools** 截图为打开效果图



1. Console下输入代码并回车

便可以获取网站的目标数据信息

以下为输入的代码

$.ajax({

type : 'POST',

url : 'http://www.neeq.com.cn/disclosureInfoController/infoResult.do',

dataType: 'JSONP',

data :{disclosureType:"5",page:"1",companyCd:"",isNewThree:"1",startTime:"2017-03-01",endTime:"2017-03-31",keyword:"",xxfcbj:"1"},

success : function(data) {

console.info(JSON.stringify(data));

},

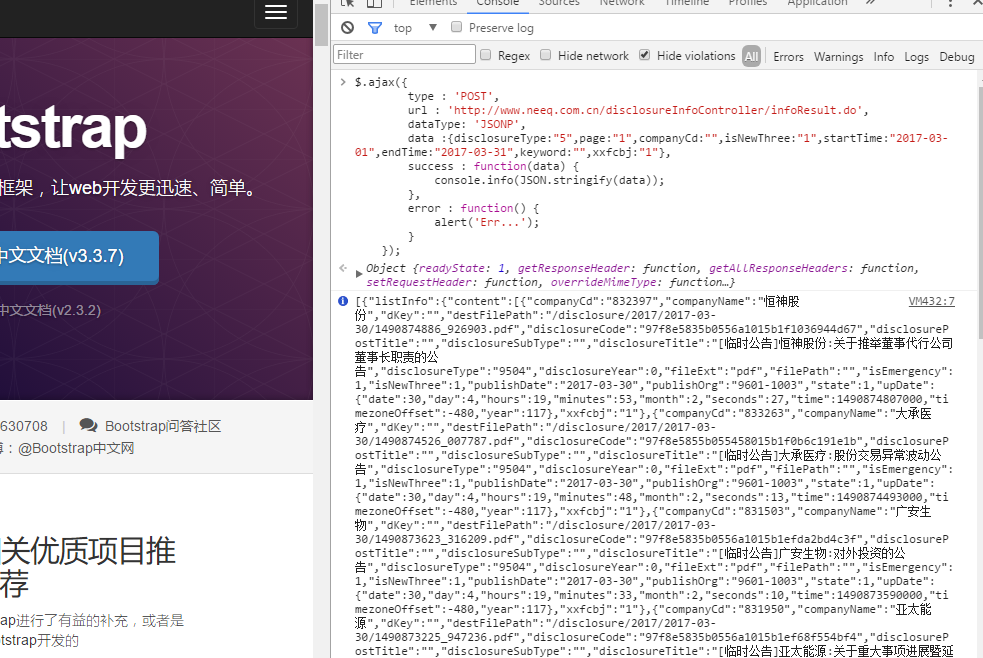
error : function() {

alert('Err...');

}

});

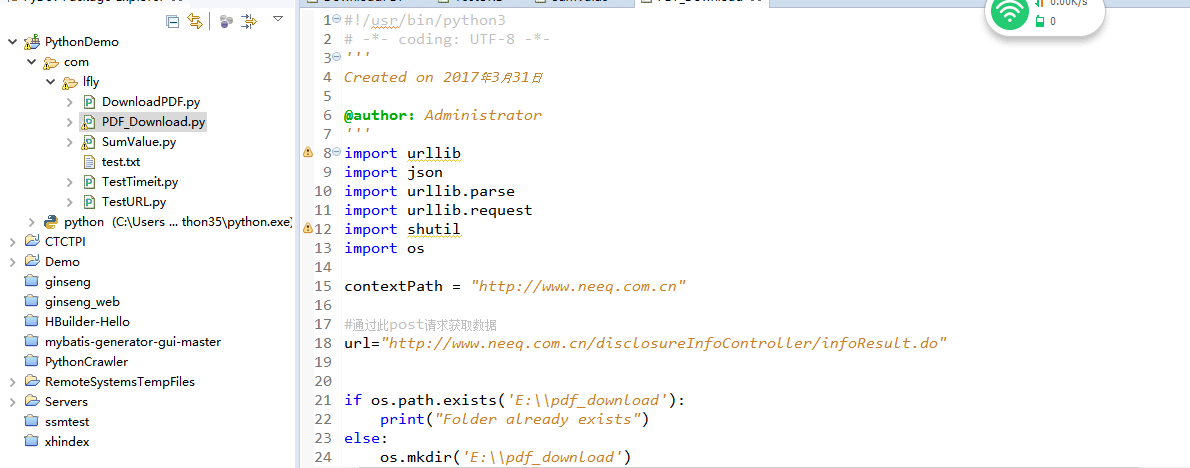
以下为回车后，控制台打印输出的结果



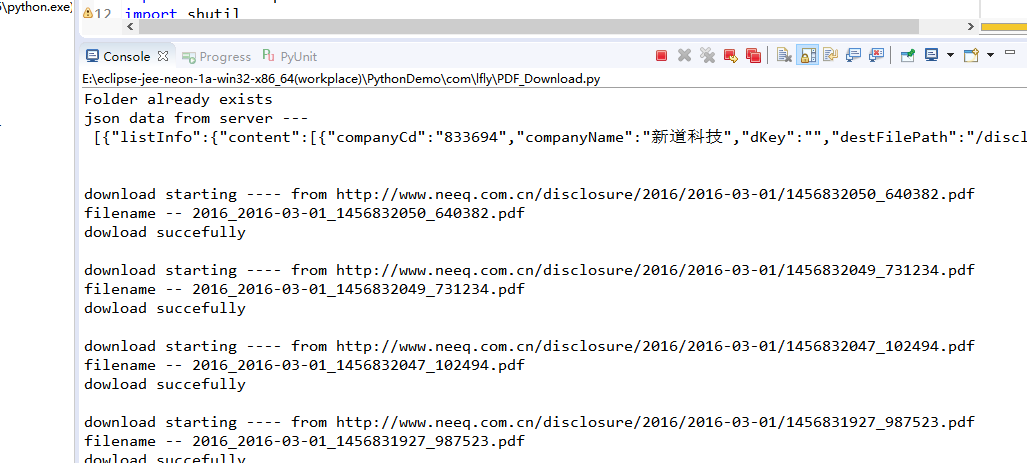
说明：返回的json数据包含了真实pdf的下载路径，则需要解析json数据，拼接为实际的下载地址并下载

1. Python代码实现

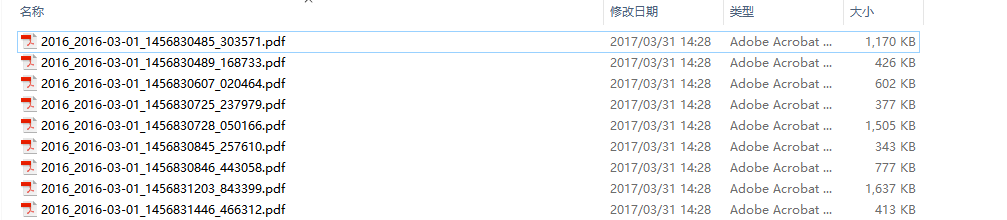
Eclipse下代码文件图 PDF\_Download.py



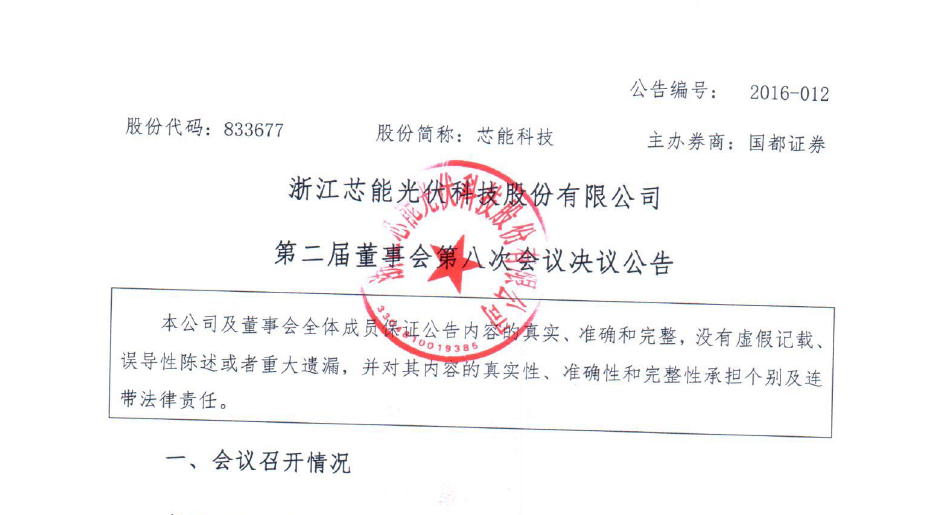
Eclipse下运行效果图



设置E盘符下pdf\_download文件夹生成的文件效果图



文件打开效果图



参考

Python源码：

#!/usr/bin/python3

# -\*- coding: UTF-8 -\*-

*'''*

*Created on 2017年3月31日*

**@author:** *LFly*

*'''*

import json

import urllib.parse

import urllib.request

import os

#网站源：全国中小企业股份转让系统

contextPath = *"http://www.neeq.com.cn"*

#通过此post请求获取数据

url=*"http://www.neeq.com.cn/disclosureInfoController/infoResult.do"*

#在E盘符下建立文件夹

if os.path.exists(*'E:\\pdf\_download'*):

print(*"Folder already exists"*)

else:

os.mkdir(*'E:\\pdf\_download'*)

#切换到该目录下

os.chdir(*'E:\\pdf\_download'*)

def **downloadPDF**(url):

print(*'download starting ---- from %s '*%(url))

#拼接文件名

filename = url[url.find(*'disclosure/'*)+11:].replace(*'/'*,*'\_'*)

response = urllib.request.urlopen(url)

print(*"filename -- %s"*%(filename))

fp = open(filename,*'wb'*)

block\_sz = 8192

while True:

buffer = response.read(block\_sz)

if not buffer:

break

fp.write(buffer)

fp.close()

print(*'dowload succefully \n'*)

def **postHttp**(disclosureType=None,page=None,companyCd=None,

isNewThree=None,startTime=None,endTime=None,

keyword=None,xxfcbj=None):

#提交的参数

postdata=dict(disclosureType=disclosureType,page=page,companyCd=companyCd,isNewThree=isNewThree,

startTime=startTime,endTime=endTime,keyword=keyword,xxfcbj=xxfcbj)

#url编码 post请求参数

postdata=urllib.parse.urlencode(postdata).encode(encoding=*'utf\_8'*)

request = urllib.request.Request(url,postdata)

request.add\_header(*'Content-Type'*,*"application/x-www-form-urlencoded"*);

request.add\_header(*'Accept'*,*"text/plain"*);

#返回json数据

response = urllib.request.urlopen(request)

#根据返回的数据作出相应的处理 截取json字符串

rawData = response.read().decode(*"utf\_8"*)

#根据返回数据格式做相应操作

if(rawData[0:4]==*"null"*):

jsonArray = rawData[5:-1]

else:

jsonArray = rawData[:-1]

#打印截取后的json字符串

print(*"json data from server --- \n %s"*%(jsonArray))

json2dit = json.loads(jsonArray)

contents = json2dit[0][*'listInfo'*][*'content'*]

print(*"\n"*)

for companyInfo in contents:

downloadPDF(contextPath+companyInfo[*'destFilePath'*])

#对应的参数 需要自己设置

postHttp(*"5"*,*"1"*,*""*,*"1"*,*"2016-02-01"*,*"2017-03-31"*,*""*,*"1"*)