

# 第四天

---

张向宇 18281059 安全1801

## 第四天

- 一、HTTP网页服务器编写
  - 1) 服务器源代码
  - 2) 服务器测试
- 二、脚本编写
  - 1) 脚本源代码
  - 2) ReadAD: 读取待爬取网页列表
  - 3) crawler: 爬虫网页数据抓取部分
  - 4) mail: 邮件发送部分
  - 5) main: 主函数主线程部分
- 三、脚本测试
  - 1) 运行HTTP服务器
  - 2) 运行脚本
  - 3) 查看邮箱情况
  - 4) 查看运行日志

## 一、HTTP网页服务器编写

### 1) 服务器源代码

我继续采用的是Golang语言编写我自己的HTTP服务器

```
package main

import (
    "fmt"
    "log"
    "net/http"
)

func main() {
    http.HandleFunc("/", index)
    log.Fatal(http.ListenAndServe("localhost:8080", nil))
}

func index(w http.ResponseWriter, r *http.Request) {
    fmt.Fprintf(w, "18281059@bjtu.edu.cn")
}
```

### 2) 服务器测试

访问测试成功



## 二、脚本编写

### 1) 脚本源代码

我这里的脚本编写语言为Python，主要分为以下几个模块编写：

- main：主函数主线程部分
- crawler：爬虫网页数据抓取部分
- mail：邮件发送部分
- ReadAD：读取待爬取网页列表
- log.txt：运行日志文件
- mailAD.txt：网页列表

主要特色：

- 使用了多线程并发处理邮件发送服务，可以实现尽可能地压缩电脑带宽性能，提供更疯狂的垃圾邮件递送服务
- 各模块功能划分清晰，便于代码的重复利用与再组合
- 具备系统日志功能，便于运行维护

### 2) ReadAD：读取待爬取网页列表

源代码：

```
from datetime import datetime
class retxt:
    Mlist = []
    def __init__(self):
        with open('mailAD.txt', 'r') as f1:
            list1 = f1.readlines()
            self.wrLog("读取网页列表成功！")
            for i in range(0, len(list1)):
                self.Mlist.append(list1[i].rstrip('\n'))
            #print(self.Mlist[i])

    def wrLog(self, str2):
        dt = datetime.now()
        str1 = dt.strftime('%Y-%m-%d %H:%M:%S %f')
        str1 = str1 + " " + str2 + "\n"
        with open('log.txt', 'a') as f:
            f.writelines(str1)
```

### 3) crawler：爬虫网页数据抓取部分

源代码：

```

import urllib.request
import re
from datetime import datetime

class GetAD:
    _Add = ""
    _URL = ""
    _HTM = b''
    def __init__(self,url):
        self._URL = url
        self._HTM = self.getHtml()
        self.Remod()

    def getHtml(self):
        page = urllib.request.urlopen(self._URL)
        html = page.read()
        return html

    def Remod(self):
        if re.match(r'^[0-9a-zA-Z_]{0,19}@[0-9a-zA-Z.]{1,13}\.[com,cn,net]{1,3}$',self._HTM.decode('utf-8')):
            c = re.compile(r'^[0-9a-zA-Z_]{0,19}@[0-9a-zA-Z.]{1,13}\.[com,cn,net]{1,3}$',re.I)
            self._Add = c.search(self._HTM.decode('utf-8')).group()
            str1 = "Email address is Right! " + "The Email address is " +
str(self._Add)
            print(str1)
            self.wrLog(str1)
        else:
            str1 = 'Can\'t match Email address!'
            print(str1)
            self.wrLog(str1)

    def wrLog(self,str2):
        dt = datetime.now()
        str1 = dt.strftime('%Y-%m-%d %H:%M:%S %f')
        str1 = str1 + " " + str2 + "\n"
        with open('log.txt', 'a') as f:
            f.writelines(str1)

    def ShowAdd(self):
        return self._Add

```

#### 4) mail: 邮件发送部分

源代码:

```

import smtplib
from email.mime.text import MIMEText
from email.header import Header
from datetime import datetime

class SentMail:
    _data = ""
    def __init__(self, data):
        self._data = data
        self.mail2()

```

```

def mail2(self):
    dt = datetime.now()
    str1 = dt.strftime('%Y-%m-%d %H:%M:%S %f')
    # 发信方的信息: 发信邮箱, QQ 邮箱授权码
    from_addr = 'xxxxxxx@xxxx.com'
    password = 'xxxxxxxxxxxxx'
    # 收信方邮箱
    to_addr = self._data
    # 发信服务器
    smtp_server = 'smtp.xxxx.com'
    # 邮箱正文内容, 第一个参数为内容, 第二个参数为格式(plain 为纯文本), 第三个参数为编码
    msg = MIMEText(str1+" send by 18281059-张向宇 ", 'plain', 'utf-8')
    # 邮件头信息
    msg['From'] = Header('18281059张向宇')
    msg['To'] = Header(to_addr)
    msg['Subject'] = Header(self._data)
    try:
        # 开启发信服务, 这里使用的是加密传输
        server = smtplib.SMTP_SSL(smtp_server)
        server.connect(smtp_server, 465)
        # 登录发信邮箱
        server.login(from_addr, password)
        # 发送邮件
        server.sendmail(from_addr, to_addr, msg.as_string())
        # 关闭服务器
        server.quit()
        str1 = "邮件发送成功 发送地址邮箱地址为: " + to_addr
        self.wrLog(str1)
    except smtplib.SMTPException:
        str1 = "Error: 无法发送邮件 " + to_addr
        self.wrLog(str1)

def wrLog(self, str2):
    dt = datetime.now()
    str1 = dt.strftime('%Y-%m-%d %H:%M:%S %f')
    str1 = str1 + " " + str2 + "\n"
    with open('log.txt', 'a') as f:
        f.writelines(str1)

```

## 5) main: 主函数主线程部分

源代码:

```

import mail
import crawler
import ReadAD
import threading
import time
from datetime import datetime

class myThread (threading.Thread):
    addr = ""
    def __init__(self, addr):
        threading.Thread.__init__(self)
        self.addr = addr
    def run(self):

```

```

        print ("开始线程: " + self.name)
        senthread(self.addr)
        print ("退出线程: " + self.name)

def sentmail():
    while 1:
        time.sleep(1)#加入了定时功能
        a = ReadAD.rettxt()
        for i in a.Mlist:
            try:
                print(i)
                thread1 = myThread(i)
                thread1.start()
                #thread1.join()
            except:
                print("Error: 无法启动线程")

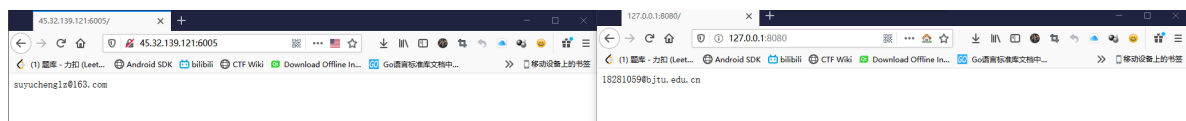
def senthread(addr):
    print(addr)
    b = crawler.GetAD(addr)
    c = mail.SentMail(b.ShowAdd())

if __name__ == '__main__':
    with open('log.txt', 'w') as f:
        dt = datetime.now()
        str1 = dt.strftime('%Y-%m-%d %H:%M:%S %f')
        str2 = str1 + " 轰炸机起飞呼~~~~~"
        f.writelines(str2)
    sentmail()
    while 1:
        pass

```

### 三、脚本测试

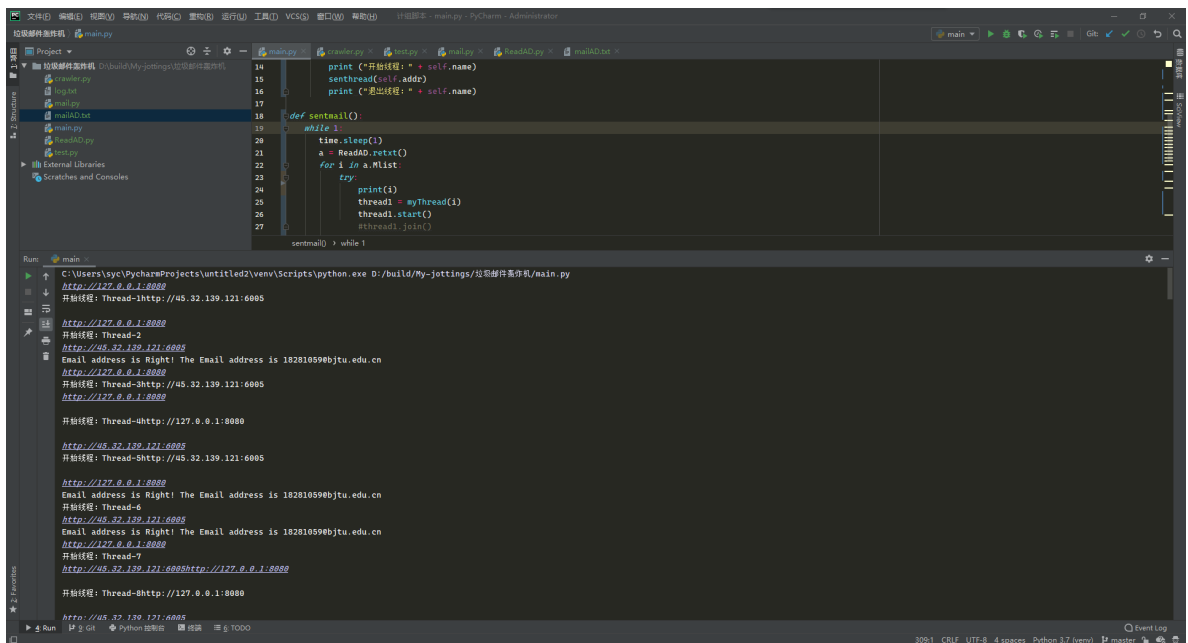
#### 1) 运行HTTP服务器



如图所示，共运行着两台自行编写的简易HTTP服务器，网页可以获取到邮箱信息，其中共有两个邮箱：

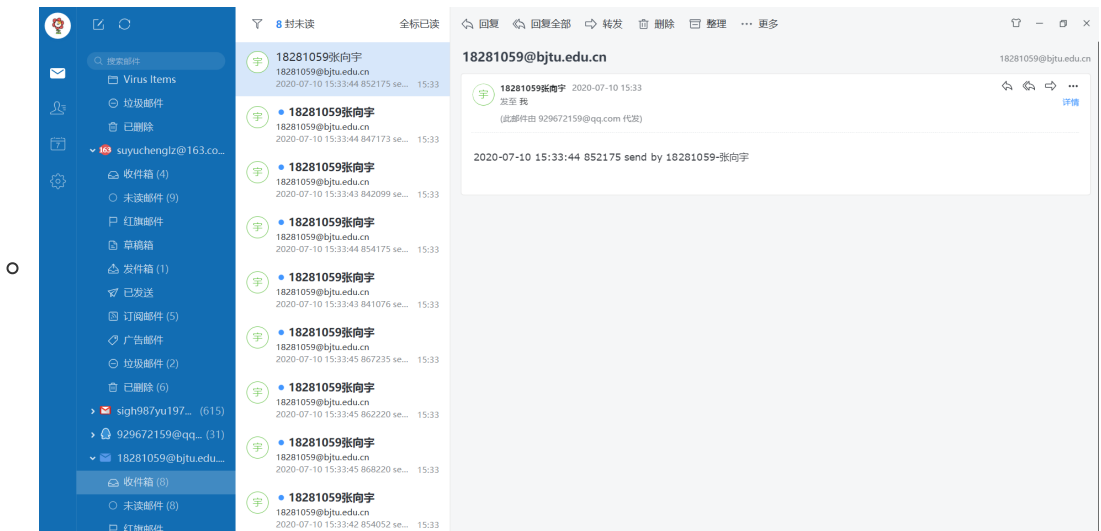
- [18281059@bjtu.edu.cn](mailto:18281059@bjtu.edu.cn)
- [suyuchenglz@163.com](mailto:suyuchenglz@163.com)

#### 2) 运行脚本

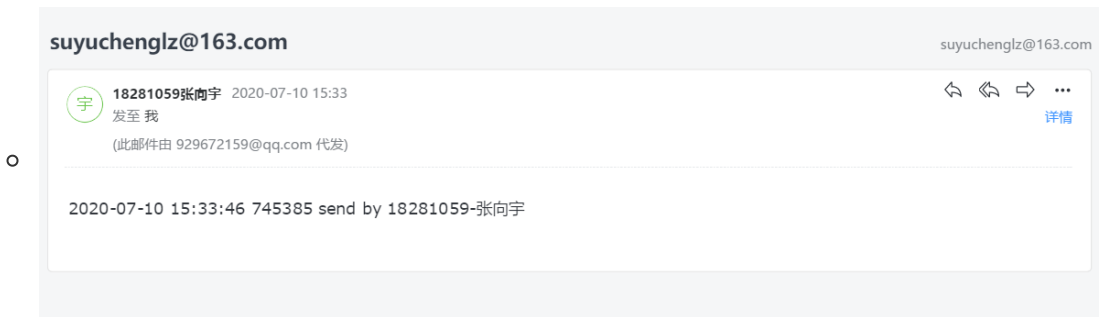


### 3) 查看邮箱情况

- [18281059@bjtu.edu.cn](mailto:18281059@bjtu.edu.cn)



- [suyuchenglz@163.com](mailto:suyuchenglz@163.com)



### 4) 查看运行日志

log.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

2020-07-10 15:33:41 825341 轰炸机起飞~~~~~2020-07-10 15:33:42 827044 读取网页列表成功!  
2020-07-10 15:33:42 853053 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:43 833094 读取网页列表成功!  
2020-07-10 15:33:43 838079 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:43 841076 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:44 839191 读取网页列表成功!  
2020-07-10 15:33:44 845191 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:44 850173 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:44 853176 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:45 850233 读取网页列表成功!  
2020-07-10 15:33:45 855219 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:45 858219 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:45 865363 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:45 867235 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:46 281249 邮件发送成功 发送地址邮箱地址为: 18281059@bjtu.edu.cn  
2020-07-10 15:33:46 476263 Email address is Right! The Email address is suyuchenglz@163.com  
2020-07-10 15:33:46 743367 Email address is Right! The Email address is suyuchenglz@163.com  
2020-07-10 15:33:46 743367 Email address is Right! The Email address is suyuchenglz@163.com  
2020-07-10 15:33:46 864299 读取网页列表成功!  
2020-07-10 15:33:46 872290 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:46 875485 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:46 880291 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:46 886295 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:46 891296 Email address is Right! The Email address is 18281059@bjtu.edu.cn  
2020-07-10 15:33:46 945296 Email address is Right! The Email address is suyuchenglz@163.com  
2020-07-10 15:33:47 396347 邮件发送成功 发送地址邮箱地址为: 18281059@bjtu.edu.cn  
2020-07-10 15:33:47 458327 邮件发送成功 发送地址邮箱地址为: 18281059@bjtu.edu.cn  
2020-07-10 15:33:47 496331 Email address is Right! The Email address is suyuchenglz@163.com  
2020-07-10 15:33:47 496331 Email address is Right! The Email address is suyuchenglz@163.com  
2020-07-10 15:33:47 496331 Email address is Right! The Email address is suyuchenglz@163.com  
2020-07-10 15:33:47 496331 Email address is Right! The Email address is suyuchenglz@163.com  
2020-07-10 15:33:47 504330 邮件发送成功 发送地址邮箱地址为: 18281059@bjtu.edu.cn  
2020-07-10 15:33:47 652355 邮件发送成功 发送地址邮箱地址为: 18281059@bjtu.edu.cn  
2020-07-10 15:33:47 891344 读取网页列表成功!  
~~~~~

第 1 行, 第 1 列 100% Windows (CRLF) ANSI