FB采集流程：Facebook登入——主页搜索框自动输入——搜索列表每一条信息进入详情页——详情页中获得邮箱 200-1000个邮箱

[https://www.facebook.com/pg/bfsteelstructure/about/?ref=page\_internal](https://www.facebook.com/pg/bfsteelstructure/about/?ref=page_internal" \t "https://www.zhipin.com/geek/new/index/_blank)

试试在公共主页点点





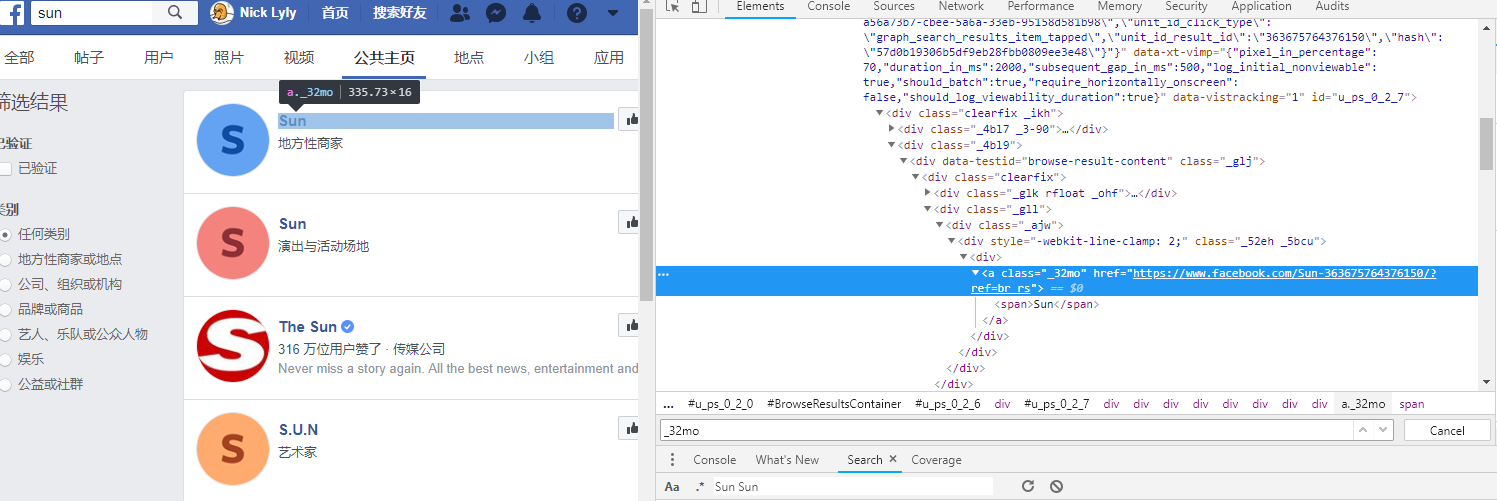
IMG_256

输入关键词的路径：

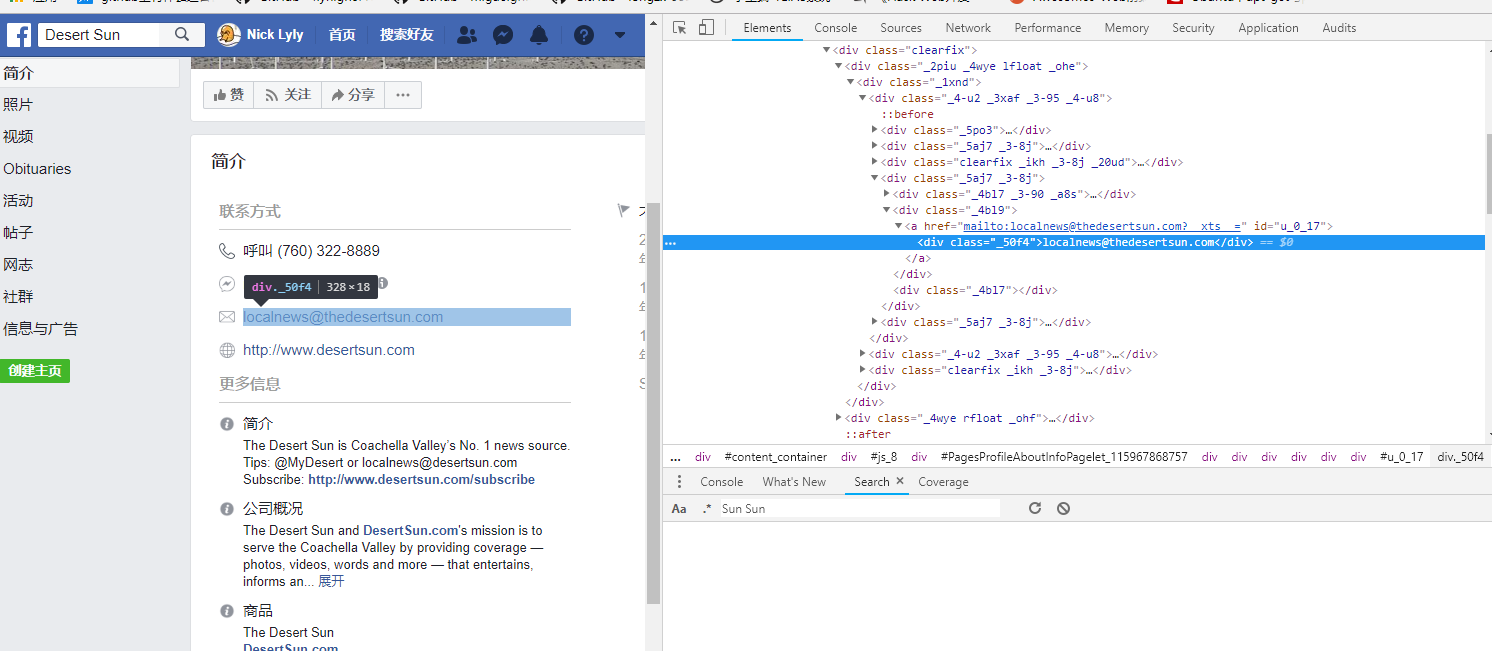


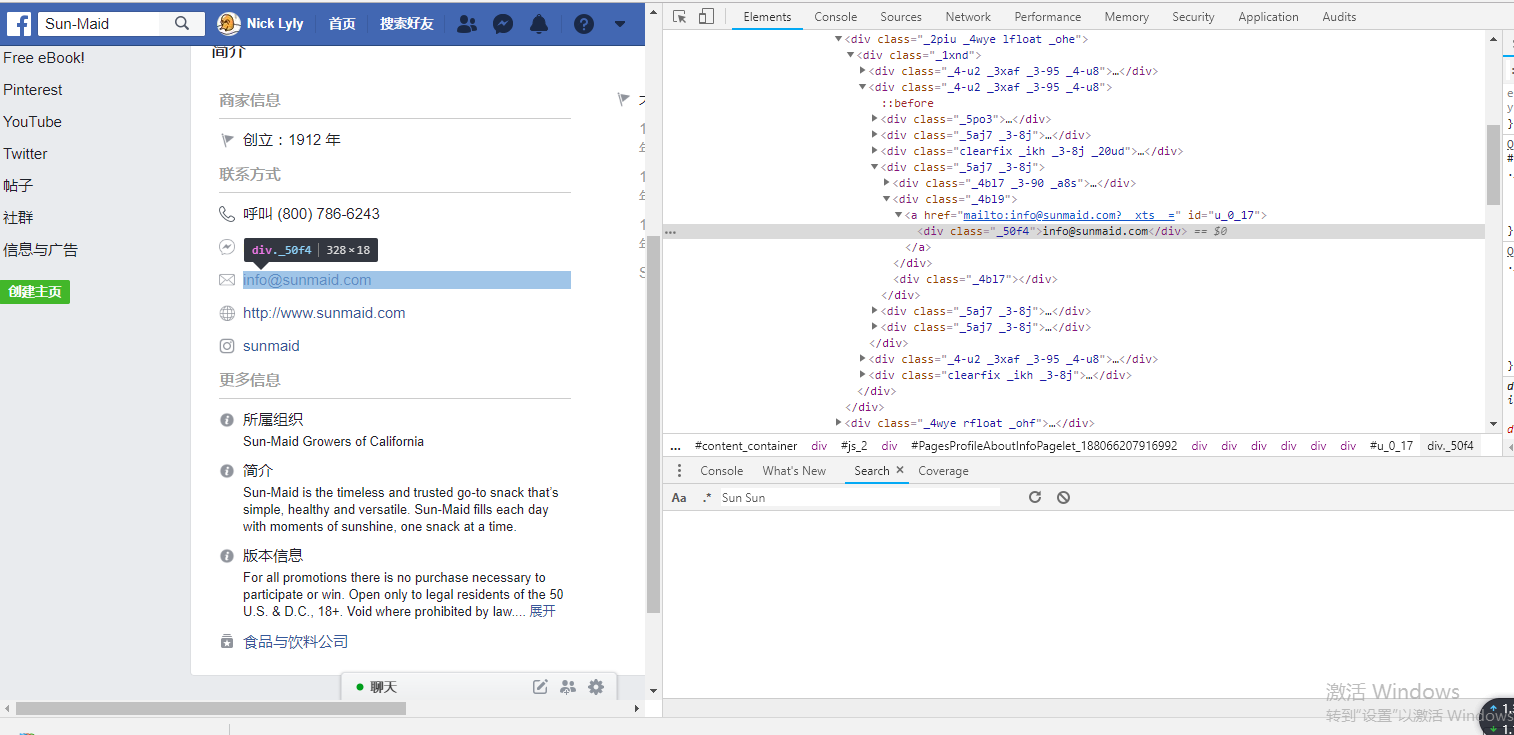
找到符合a=browser.find\_elements\_by\_class\_name("\_32mo")所有的标签块

[<selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-1")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-2")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-3")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-4")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-5")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-6")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-7")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-8")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-9")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-10")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-11")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-12")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-13")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-14")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-15")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-16")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-17")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-18")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-19")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-20")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-21")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-22")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-23")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-24")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-25")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-26")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-27")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-28")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-29")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-30")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-31")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-32")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-33")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-34")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-35")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-36")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-37")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-38")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-39")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-40")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-41")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-42")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-43")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-44")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-45")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-46")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-47")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-48")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-49")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-50")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-51")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-52")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-53")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-54")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-55")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-56")>, <selenium.webdriver.remote.webelement.WebElement (session="002e5e74dbd694cb22852ce049cbebf7", element="0.01762768532376824-57")>]



找到邮箱的地址

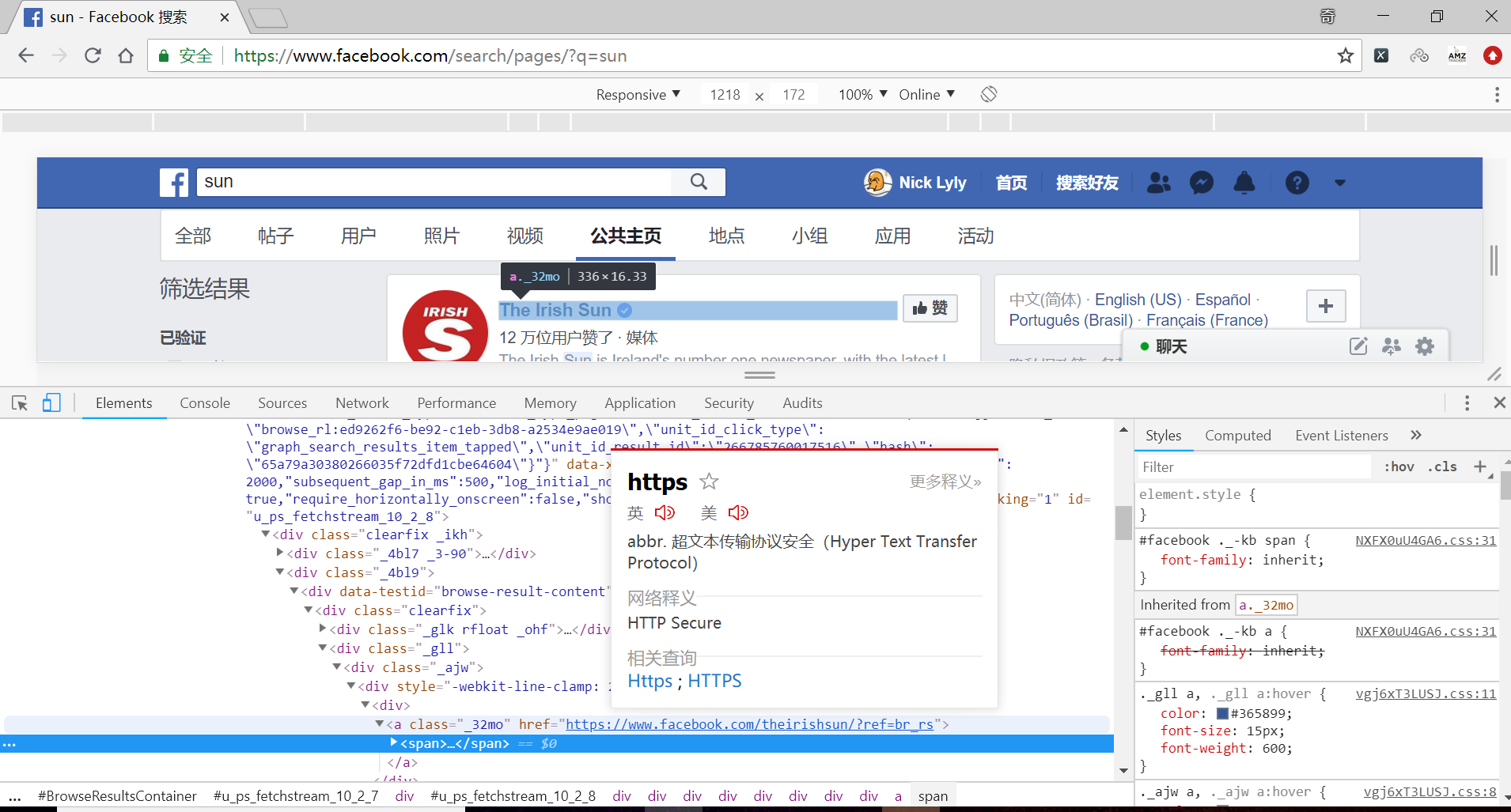




加入多线程

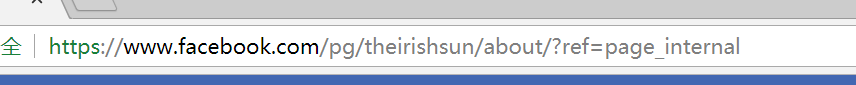
两个进程要两个账号

说明

1：

找到所有关键词用户，这些用户的共同点就是class为上面那个。

然后直接得到他们的超链接就是该用户的主页url



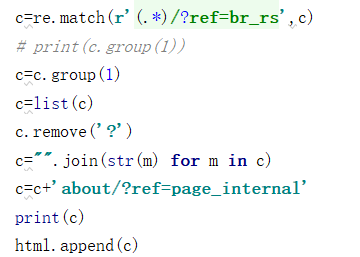
根据这个找到简介页面于个人主页的页面url的规律，构造url。

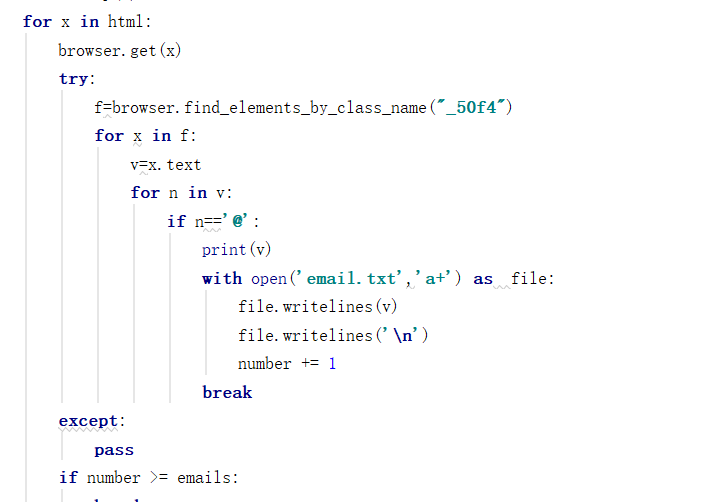
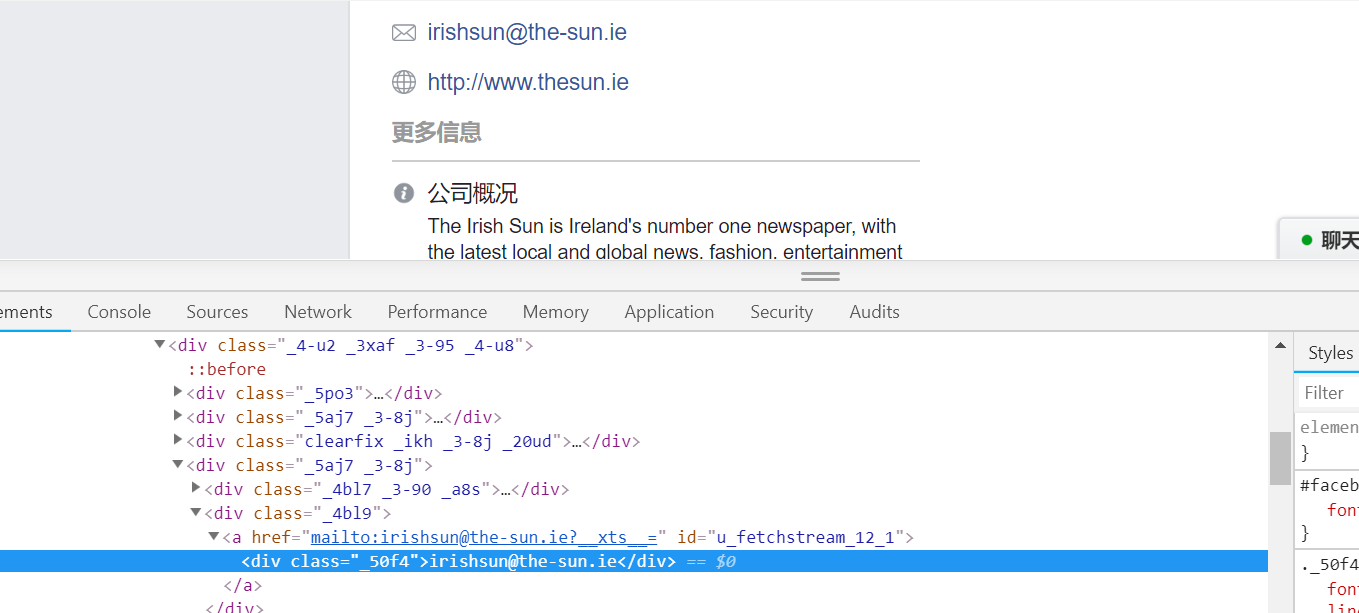


A就是每个用户的节点

X就是去遍历他

c是每个用户主页的超链接

以上就是通过正则提取然后根据找到的规律构造url,然后把他添加到所有的html列表中。

这里就是遍历刚才手机的界面主页，找到邮箱的class名为这个，可是页面中有这个名的有还几个，所以就遍历所有的，然后做个判断，如果有@这个的话就是邮箱地址了。