**程序目的：**

获取黑猫投诉平台的八家快递公司（中国邮政、京东物流、顺丰、中通、圆通、申通、百世快递、韵达）的投诉单信息和投诉流程信息

**环境准备：**python3以上版本，以下库的安装

import os

import codecs

import requests

import urllib3

from bs4 import BeautifulSoup

import pymysql

import time

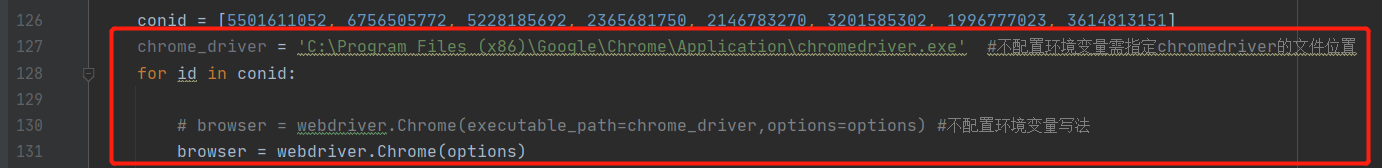
import re

from selenium import webdriver

import json

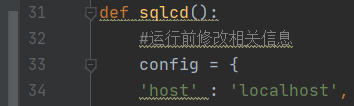
**Chrome浏览器安装**

chromedriver下载并配置，如果不配置环境变量则需修改utils中如下



**数据库修改**

修改utils中有关数据库的设置，def sqlcd 和def sqlct方法内数据库参数(用户名密码等)，不用建库建表，程序运行可自行建库建表,修改def selpath中chromedriver驱动位置路径，如果配置环境变量则不用修改





**Fiddler安装并配置**

下载后按照网上的基础配置配置好

打开后按ctrl+r快捷键打开CustomRules.js这个文件找到OnBeforeResponse将内容改为如下（注意修改保存路径，并将程序getinfo.py中路径与该路径保持一致）：

static function OnBeforeResponse(oSession: Session) {

if (m\_Hide304s && oSession.responseCode == 304) {

oSession["ui-hide"] = "true";

}

if (oSession.fullUrl.Contains("https://tousu.sina.com.cn/api/company/received\_complaints")){

oSession.utilDecodeResponse();

//消除保存的请求可能存在乱码的情况

var fso;

var file;

fso = new ActiveXObject("Scripting.FileSystemObject");

//文件保存路径，可自定义

file = fso.OpenTextFile("C:/Users/lht/Desktop/heimao\_Response.txt",8 ,true, true);

//file.writeLine("Response code: " + oSession.responseCode);

file.writeLine("Response body: " + oSession.GetResponseBodyAsString());

//file.writeLine("\n");

file.close();

}

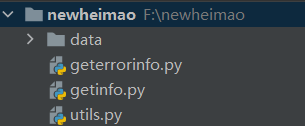
}



getinfo.py中路径：

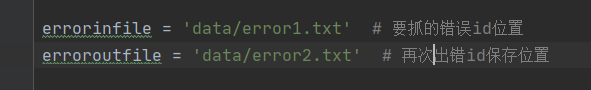


**程序内容：**



**操作流程：**

1. 数据库建库名为heimaodata ，打开fiddler并开启抓包
2. 运行getinfo.py程序
3. getinfo.py运行结束后，观察data目录下error文件看看有没有id信息，有的话运行geterrorinfo.py程序即可处理error中的id（也可直接运行，没有内容不会有影响），如果出现新error文件，更改geterrorinfo.py文件中两个error文件目录位置，重新运行



**详细说明：**

目前设置的是每个快递公司页面滑动60页，也就是每个快递公司获取600个投诉单，每个公司投诉数量不同，现情况60页都可以覆盖到上星期的信息，如有需要可加大获取页数，经测试程序翻页最大页数在100页左右，自己手动翻页可在300页左右

详细流程：

1. 开启fiddler软件，selenium模拟滑动页面，抓包相应内容，将抓到的信息保存到文件
2. 解析抓包信息文件，将所需投诉id保存到data目录下
3. 依次访问投诉id相应页面，保存html信息到data目录下
4. 解析html信息，得到投诉单信息和投诉流程信息，插入数据库

快递公司id：中国邮政 5501611052

京东物流 6756505772

顺丰 5228185692

中通 2365681750

圆通 2146783270

申通 3201585302

百世快递 1996777023

韵达 3614813151

**Utils工具类：**包含连接数据库方法，头部信息，sql语句提交，替换方法，修改数据库信息在这里，其中sqlcd是建库用的，sqlct是建表用的,以及getsnid

是解析抓包信息得到id的方法，selpath模拟滑动方法，getinfo解析html获得投诉单信息

getstep解析html获得投诉流程信息

**备用：**建库建表在程序中都有，如出意外以下语句备用

建库语句：

CREATE DATABASE if not exists heimaodata DEFAULT CHARACTER SET utf8

建表语句：

投诉信息表：

create TABLE if not exists tousu\_sina(

`title投诉标题` varchar(255) DEFAULT NULL,

`sn投诉编号` varchar(255) DEFAULT NULL,

`cotitle投诉对象` varchar(255) DEFAULT NULL,

`reason投诉问题` varchar(255) DEFAULT NULL,

`appeal投诉要求` varchar(255) DEFAULT NULL,

`request\_repay\_fee涉诉金额（元）` varchar(255) DEFAULT NULL,

`status投诉进度` varchar(255) DEFAULT NULL,

`attitude服务态度评分` varchar(255) DEFAULT NULL,

`process处理速度评分` varchar(255) DEFAULT NULL,

`satisfaction满意度评分` varchar(255) DEFAULT NULL,

`evalContent处理后用户的评论` LONGTEXT DEFAULT NULL,

`appeal\_date投诉产生时间` varchar(255) DEFAULT NULL,

`crawler\_time爬取时间` varchar(255) DEFAULT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

投诉流程表：

create TABLE if not exists tousu\_steplist (

`sn投诉编号` varchar(255) DEFAULT NULL,

`operator该流程操作人` varchar(255) DEFAULT NULL,

`status该流程的当前状态` varchar(255) DEFAULT NULL,

`operate\_content该流程的操作内容` LONGTEXT DEFAULT NULL,

`operate\_time该流程的时间节点` varchar(255) DEFAULT NULL,

`提取时间` varchar(255) DEFAULT NULL,

`小时` varchar(255) DEFAULT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

**导出语句：**注意修改路径

投诉信息表：

select \* from

(SELECT

'title投诉标题' ,

'sn投诉编号' ,

'cotitle投诉对象' ,

'reason投诉问题' ,

'appeal投诉要求' ,

'request\_repay\_fee涉诉金额（元）' ,

'status投诉进度' ,

'attitude服务态度评分' ,

'process处理速度评分' ,

'satisfaction满意度评分' ,

'evalContent处理后用户的评论' ,

'appeal\_date投诉产生时间' ,

'crawler\_time爬取时间'

union all

select \* from `tousu\_sina`) as t1 INTO OUTFILE 'D:/heimaodata/tousu\_sina.csv'

CHARACTER SET gbk FIELDS TERMINATED BY ',';

投诉流程表：

select \* from (

select

'sn投诉编号' ,

'operator该流程操作人' ,

'status该流程的当前状态' ,

'operate\_content该流程的操作内容' ,

'operate\_time该流程的时间节点' ,

'提取时间',

'小时'

union all

select \* from `tousu\_steplist`) as t1 INTO OUTFILE 'D:/ztodata/tousu\_steplist.csv'

CHARACTER SET gbk FIELDS TERMINATED BY ',';