采集说明

首先需要有一定的python 基础，环境搭建这里不做说明比较复杂

1. 安装依赖

Pip install ast

Pip install cProfile

Pip install pydoc

Pip install attr

Pip install requests

Pip install urllib

Pip install selenium

Pip install bs4

Pip install openpyxl

Pip install cfscrape

Pip install json

Pip install string

Pip install re

Pip install time

Pip install math

Pip install \_thread

Pip install dateutil

Pip install ssl

Pip install sys

1. 安装google浏览器
2. 安装selenium python浏览器插件

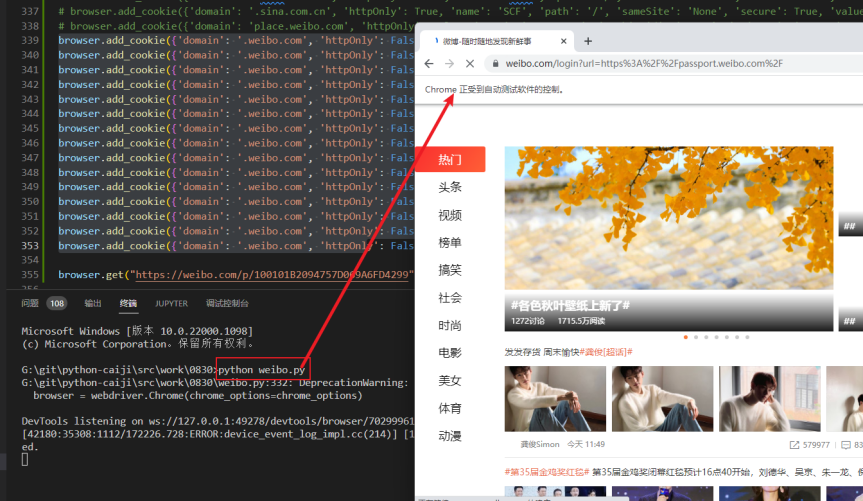
地址：<http://chromedriver.storage.googleapis.com/index.html>

按照google 浏览器版本安装相应的插件

1. 运行程序

运行命令行工具CD进入程序目录(最好是有开发工具)

运行 python weibo.py



如图跑起来会启动浏览器插件 给程序所控制，程序会抓取浏览器中相应区域的信息，以及自动跳转链接加载信息

1. 中断继续采集

程序做了异常后继续以断点位置进行采集，如果发现程序异常退出，请重新执行采集命令启动程序

1. 采集结果

采集结果为excel， 文件名称以当前时间yy-MM-ddHHmmss 的格式，多次中断会有多个文件。最总需要将所有采集结果汇总

