任务二中存在

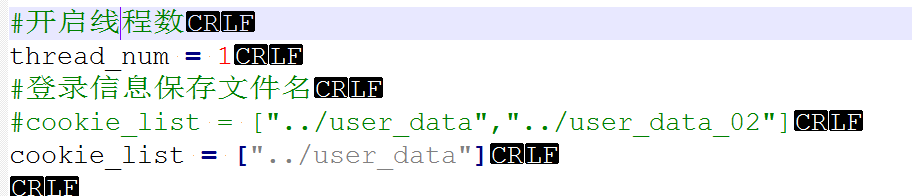
任务一中也存在

任务一与任务二的区别在于任务一使用模拟浏览器的方法进行爬取

任务二中使用动态IP的方法爬取

由于任务二的方法无法应对 访问限制的情况 可使用任务一中的代码爬取

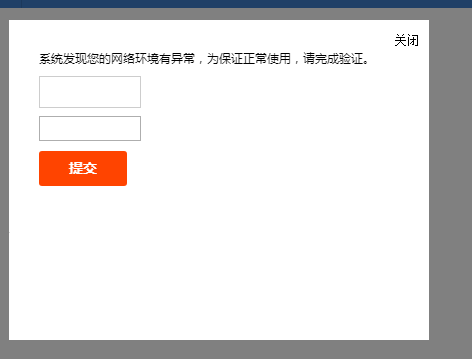
同抓取店铺评价信息



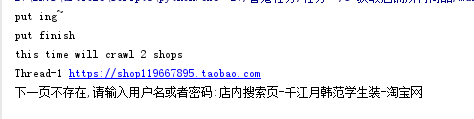
设置 登录信息保存 (一个淘宝帐号一个登录信息) thread\_num与 登录信息数量保持一致

该代码读取all\_url.csv 爬取店铺商品列表存入数据表product\_list, 及has\_craw\_shop.txt

当发生访问限制时需要人工输入密码及验证码

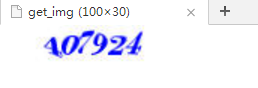


出现以上信息时 先等待console窗口打印相关信息



然后输入密码 验证码

为加快访问 模拟浏览器默认屏蔽图片 在验证码图片右键 复制图片地址 ，打开另一个浏览器输入及可看到 验证码



输入完成密码 验证码后 在console 窗口 按下回车 则爬虫继续

这种方法属于半人工的爬取方法，

速度取决于能找到的淘宝帐号数量 及一些人工的效率