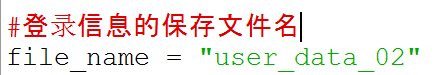
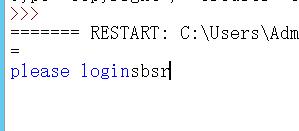
## 步骤1： 获取登录信息（selenium\_cookie.py）

编辑文件

输入 登录信息的保存文件名(此处为user\_data\_02)



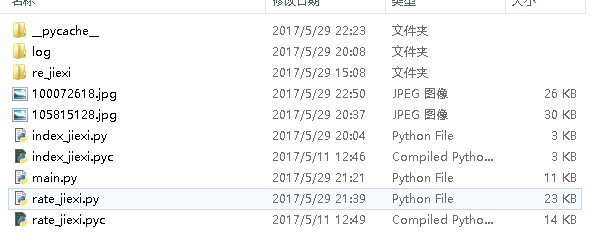
输入保存登录信息的文件名后，运行程序，会启动chrome浏览器，在60秒内完成手动点击登录输入账户密码,后再shell处随意输入字母按回车



等待 shell 打印cookie信息即可，此时在目录下会生成相应的登录文件夹（user\_data\_02,见上图），里面就保存了登录信息。 若需要在多个线程下使每个线程都能登录。需要多次操作 输入不同的保存文件名及淘宝帐号密码

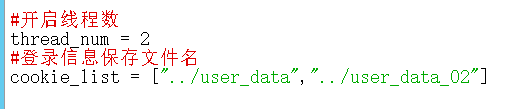
登录信息**有效期为**6天左右 ，之后需要再更新

1.获取店铺主页文件列表：



**任务简介**：读取all\_url.csv文件，获取店铺主页商品列表（shop\_homepage）和店铺评价页面（seller\_info），生成has\_craw\_shop.txt 文件

**全局变量**

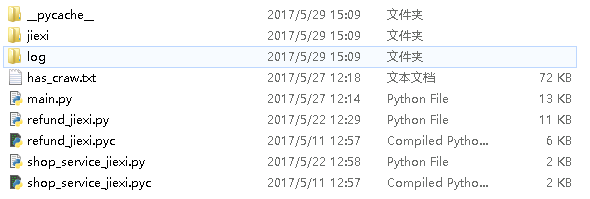


设置线程数以及保存文件名

**断点续传**：程度会读取all\_url.csv 和has\_craw\_shop.txt 做对比，筛选为爬取的店铺

**新的一轮**：需要清空has\_craw\_shop.txt

2.获取商品中店铺的信息文件列表



**任务简介：**读取任务一生成的has\_craw\_shop.txt 文件，更新seller\_info 以获取商品中的店铺信息，生成has\_crawl.txt文件

**全局变量**



**断点续传**：程序会对比has\_craw\_shop.txt 和has\_crawl.txt 筛选未更新的店铺

**新的一轮**：需要清空该文件夹下的has\_crawl.txt 文件