taobao控价工具使用说明

■ Date	@December 5, 2023
⇔ Status	Not started
∷ Tags	

安装依赖

pip install -r requirements.txt

运行

cookie过期处理

如果cookie过期了,执行python文件的时候,执行

python taobaoCrawler.py needlogin

后续当天再登录就不需要新cookie了,执行

python taobaoCrawler.py

安装好依赖之后,找到taobaoCrawler.py文件,运行即可

输入文件

文件可以用Execl编辑,.csv格式存储就好

文件当中的关键字应当尽量精炼,因为是模糊匹配的

传入格式为如下图所示:

Α	В	С	D	E	F	G
关键词	版本	适用地区	科目	最低价格	起始页	结束页
有道名师	全有道	高一	数学,物理	29, 29, 29	1	5
写作能力	暴涨秘籍	高一	写作,阅读	29, 29, 58	1	5
文言文能	力暴涨秘籍	高一		35	1	5

其中:

• 版本号对应

版本: 纸上的作文直播课 重构阅读思维+文语方程式+破解古典诗词

包君成用中华典籍打造高分作文

包君成写作/阅读能力暴涨秘籍

- 适用地区对应
 - 。高一、高二、等等

适用学龄 段: 初中通用

• 科目包括

科目之间用中文,隔开!

- 。 语文、数学、英语。。。
- 。 科目组合:格式可以为"语文+数学+。。。",如果害怕匹配不上多写一些格式就 行比如:语文+数学,语数,语+数。。。。

传文件demo

crawl.csv

验证码跳过

简单验证码

遇到普通的滑块验证码,需要进行不断尝试(一般尝试6~7次就会进入),如果尝试的次数超过了10次,大概率是监测到机器人,解决方案就是在电脑上手动辅助进行滑动,一般手动两三次就可以

复杂验证码



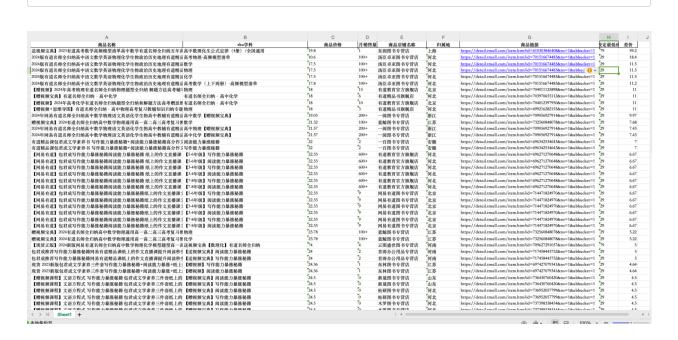
若遇到类似的验证码,脚本会尝试删除浏览器cookie,然后重新尝试登录,需要人手动配合(最好能有Cookie池)

若重新登录之后仍然带有该验证码检验,则需要人手动进行验证

输出文件

<u>淘宝爬取商品结果2023-12-05 14-34-38.xlsx</u>

<u>淘宝爬取商品结果2023-12-05 15-27-12.xlsx</u>



分为以下几列,商品名称、sku学科、价格、销量、商品店铺名称、归属地、商品链接、设定的最低价、差价(指与设定的价格的差距)