Price Scrape Quick User Guide

Bought to you by:

Chng Yan Hao

Lim Kah Ghi
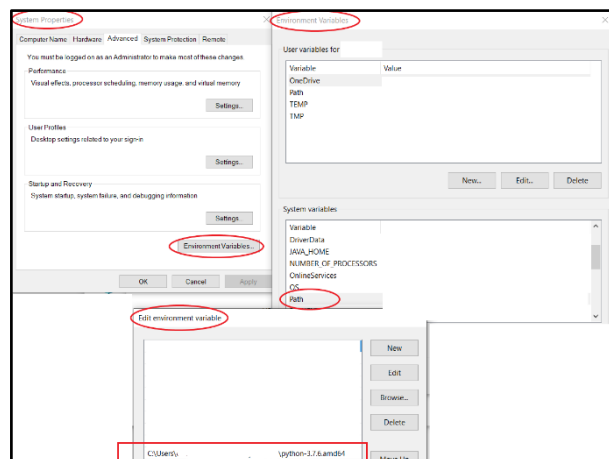
Contents
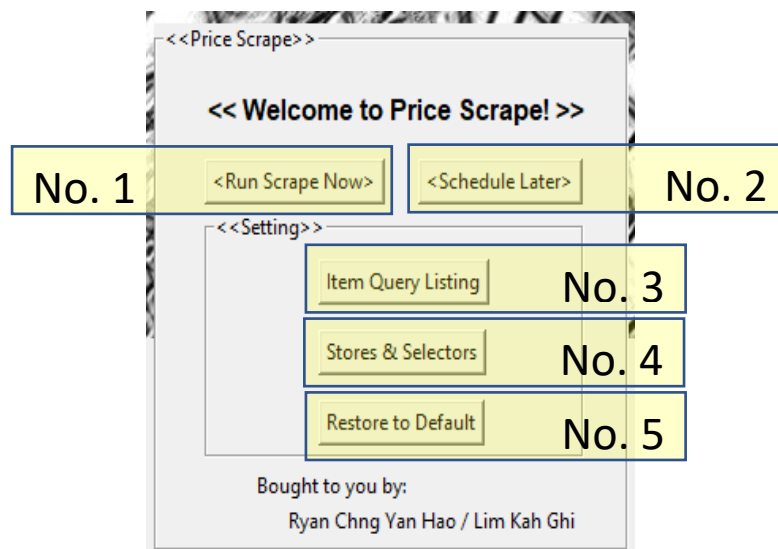
1) System Requirements & Built

    i.   Application is built under Windows 10 Pro OS

    ii.   Internet connection required

    iii.   Google Chrome Browser newly installed (recommended)

    iv.   Python 3.7.6

    v.   Packages

1)	-	Web scraping	: tagui
2)	-	GUI	: tkinter
3)	-	Data and system related	: os, datetime, re, schedule, time
4)	-	Numbers and text	: pandas, nltk, nltk (punkt & stopwords)
5)	-	Images	: PIL

    vi.   Download PriceScrape.zip

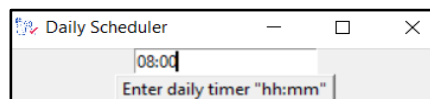   vii. Directory must not contain space in the path

2) Deployment

    i.   Allow python to run from CMD.
i.e. Windows OS: System properties -> Environment Variables -> Path -> Edit environment variable (Python path required)

    ii.   Unzip PriceScrape.zip to any preferred location.

    iii.   Click on PriceScrapeGUI.bat to start.
(Windows OS)
>> python UserInterface.py
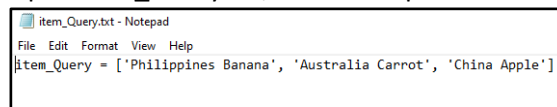(Mac OS)

3) Using Price Scrape – User Interface

- ➢ No. 1, **\<Run Scrape Now>**:   Execute PriceScrape.py, main algorithm.
- ➢ No. 2, **\<Schedule Later>**:   Trigger a popup, user can set "hh:mm", execute PriceScrape.py daily at "hh:mm" (24-hour format, invalid input will cause scheduler to fail).
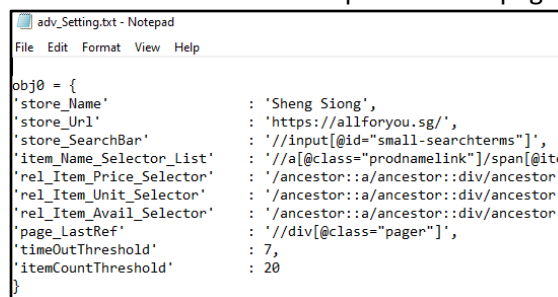

- ➢ No. 3, **Item Query Listing**:   Open item_Query.txt, user can input their item list here.

  To add an item, replace last "]" with ", 'Your Item']"
  The list should look "['item1', 'item2', … , 'itemN']"
  Changing "item_Query = " will cause malfunction

- ➢ No. 4, **Stores & Selectors**:   Designed for **advanced users** to change scrape targets. Able to add more scraping objects too.
  timeOutThreshold: wait time before going to next step
  itemCountThreshold: max number of items user wish to scrape from first page

  Users with TagUI / selectors knowledge may wish to change other settings here to better suit their needs.

- ➢ No. 5, **Restore to Default**:   Trigger default item_Query.txt & adv_Setting.txt. Previous settings are stored into output folder.

4) Known Errors

i. Failed to create data directory

If your Windows 10 had been updated recently, an error window may occur due to the new privacy policy update of Windows 10 onto one of the Chrome folders.
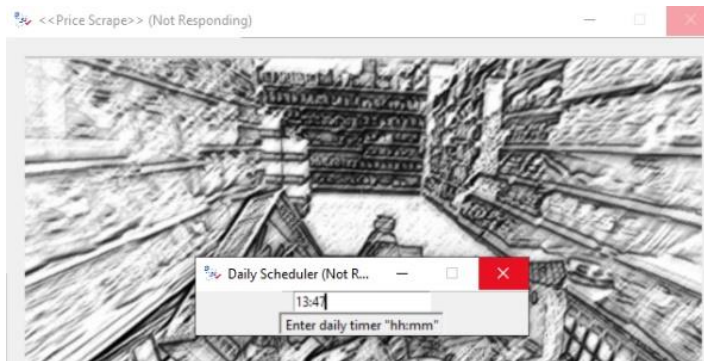
This will cause new Chrome session to fail to open properly and Price Scrape will be disrupted both in its <Run Scrape Now> and <Scheduler Later> modes.

It is recommended to
- i.   Reinstall Google Chrome Browser OR
- ii. Reset Google Chrome Browser settings to default OR
- iii. Remove HKEY_LOCAL_MACHINE\SOFTWARE\Policies\Google\Chrome

You are reminded to reinstall tagui or set its Chrome path if chrome.exe has changed its location after reinstallation.

ii. Application not responding

The main UI may appear to be "Not Responding" after the first run from scheduler while the scheduler is actually waiting for its next job.

Closing either of the windows that Price Scrape opened up will disrupt the scraping:
1) 1 main window (the main UI), 1 small hh:mm window (scheduler mode)
2) 2 CMD windows (behind the main UI)
3) 1 Google Chrome Browser (while scraping)

iii. Browser loaded but items not scraped
As different query item will return different list, different internet connections and websites will affect the speed of the webpage being loaded, the tool may start to scrape too early if **Stores & Selectors: timeOutThreshold (the number of seconds before tool gives up)** for the stall is set too small. Kindly set a larger **timeOutThreshold** that suit your needs if you encounter a problem with any of the webpage on this.

--- End of Guide ---