

三大商城比價程式



2023年6月8日

目錄

[項目名稱 2](#_Toc137083933)

[介紹 2](#_Toc137083934)

[這個主題的由來 2](#_Toc137083935)

[開源數據源。 2](#_Toc137083936)

[程式碼相關工作（使用包的介紹） 2](#_Toc137083937)

[主要功能介紹（代碼介紹） 2](#_Toc137083938)

[函式的結果 3](#_Toc137083939)

[結論。 3](#_Toc137083940)

[參考。 3](#_Toc137083941)

[心得（每個人） 3](#_Toc137083942)

[分工。 4](#_Toc137083943)

# 項目名稱

三大購物商城比價系統

# 介紹

## 這個主題的由來

現代年輕人都喜歡上網購物，但同學們時常都無法快速尋得一個最低的價格，往往為了要貨比三家而找了好多的資料，為此，我們做出這個小專題來幫助大家最快速的取得自己想要的數據。

我們這次的小專題式利用三個爬蟲去爬取三個不同的價格，再用程式去進行一個簡單的比價，就可以得到理想的答案了。

## 開源數據源。

<https://shopee.tw/>

<https://www.momoshop.com.tw/main/Main.jsp>

<https://24h.pchome.com.tw/>

# 程式碼相關工作（使用包的介紹）

# 匯入必要的函式庫

import urllib.parse # 用於解析URL

import sys # 用於訪問系統特定的參數和函數

import json # 用於處理JSON數據

import requests # 用於發HTTP請求

from bs4 import BeautifulSoup # 用於解析HTML和XML文檔

import lxml # 用作BeautifulSoup的解析器

import pandas as pd # 用於數據操作和分析

# 主要功能介紹（代碼介紹）

PChome副程式:主程式會傳入搜尋產品的關鍵字和要不要顯示PChome所有搜尋結果，藉由靜態爬取來取得所有的的商品資料，然後一筆一筆資料做比較，最後輸出最便宜的商品。

momo副程式:同PChome的格式與功能。

主程式:把PChome和momo的最低商品放入到finish變數，最後依照排序排列出最便宜商品的名次，並產生.csv檔案。

# 函式的結果

印出每個平台爬到的商品，最後印出各自平台最便宜的商品名稱與價格，並輸出成.csv方便觀察與比較，使使用者可以更方便的比較價格。

# 結論

這個三大購物商城比價系統的小專題是為了幫助現代年輕人更快速地找到最低價格而開發的。我們使用了三個爬蟲來爬取三個不同的購物商城的價格，並使用式進行簡單的比價，最後輸出最便宜的商品。我們使用了Python的許多函式庫，包括urllib、sys、json、requests、BeautifulSoup和pandas等，來實現這個小專題的功能。我們的程式碼包括PChome副程式、momo副程式和主程式，主程式會呼叫PChome和momo副程式來取得商品資料，並將最低價格的商品放入finish變數，最後依照價格排序並輸出成.csv檔案。這個小專題的結果可以印出每個平台爬到的商品，最後印出各自平台最便宜的商品名稱與價格，並輸出成.csv方便觀察與比較，使使用者可以更方便地比較價格。

# 參考資料

* <https://blog.jiatool.com/posts/pchome_spider01/>
* <https://clu.gitbook.io/python-web-crawler-note/71-momogou-wu-wang-pa-chong>
* <https://ithelp.ithome.com.tw/articles/10232131>
* <https://www.learncodewithmike.com/2020/02/python-beautifulsoup-web-scraper.html>
* <https://www.learncodewithmike.com/2020/05/python-selenium-scraper.html>
* <https://scriptwine.com/ireck/2022/05/18/python%E5%A4%9A%E8%B3%87%E6%BA%90%E7%88%AC%E8%9F%B2-%E6%AF%94%E5%83%B9%E7%88%AC%E8%9F%B2>

# 心得（每個人）

李杰儒：

經過這次實作，我了解到要爬取網站也是相當的困難，除了要好好了解網站的結構之外，也要避開網路上的許多防爬限制，甚至還需要使用官方提供的API才可以調取資料，以下詳細描述。

要理解網站的結構和架構之外，我們也需要注意到許多網站為了防止爬蟲，會設置各種防爬機制，像是CAPTCHA驗證、IP限制、User-Agent檢查等。要成功爬取這些網站，我們就需要設法避開或繞過這些防爬措施。

有時候，網站也會提供開放的API，讓我們可以透過API來取得資料，這樣就不需要直接爬取網頁了。使用API不但可以避開防爬機制，也可以獲得更加乾淨和結構化的資料。但是，並非所有的網站都提供開放的API，有些網站的API也可能需要付費使用，例如蝦皮就需要商家申請才可以使用。

在開發三大商城比價程式的過程中，我們就需要運用爬蟲技術來取得商品資料和價格資訊。但是，鑑於大型商城網站的防爬機制通常都非常強大，我們很有可能需要使用他們提供的開放API來取得部分資料。所以不可避免的就有可能失敗，但還是非常有趣，希望未來我能繼續進步。

張修齊：

這次在做比價程式時遇到蠻多的問題，像是momo的網頁格式有改，所以設計上要再研究一下，後來在輸出最便宜價格結果的時候，發現輸出的結果有點問題，找了很久才發現原來是min函數在處理Datafriend資料時，它好像把商品名稱來找最小值，所以輸出的結果是最短的文字而不是最便宜的價格，所以我改變了設計模式，我在處理一筆一筆資料時，就讓它做比較最小值的動作，最後執行的結果終於正常了。

嘗試了很久，蝦皮的API還是找不到，主要是因為需要蝦皮官方授權，所以無法順利執行，雖然有點小可惜，但也發現蝦皮網站很好的防爬蟲機制。

李翊銍：

做完這次的小專題後，我學到了基本的爬蟲應用，可以應用於一些簡單的網頁，可以豐富我們的生活的同時也可以增加未來就業的優勢，老師上課時手把手的帶我們寫程式效果也遠勝於以往的教學方式，讓我在這次的小專題收穫許多。

余孟哲：

在這次的小專題當中，對於python程式有了更進一步的認識，之前老師教過的爬蟲程式也更了解的更深入。查找出網頁的程式代碼得出網頁資訊，但實作時完全不像網路上的作法那樣容易，感謝同學的幫助，也感謝課堂上的實作，在問題處理上更好應對，在這次的程式中學習很多。

# 分工。

李杰儒：爬蟲程式設計

張修齊：爬蟲程式設計

李翊銍：程式整合

余孟哲：優化搜尋系統