HW1 Feebee比價網站 針對樹莓派硬體, 顯示出每個網站的商品名稱,售價及網址

進入https://feebee.com.tw/ 網頁, 搜尋方塊內輸入raspberry pi 4 可得到完整網址: https://feebee.com.tw/s/?q=raspberry+pi+4

程式必須加上偽裝瀏覽器的header

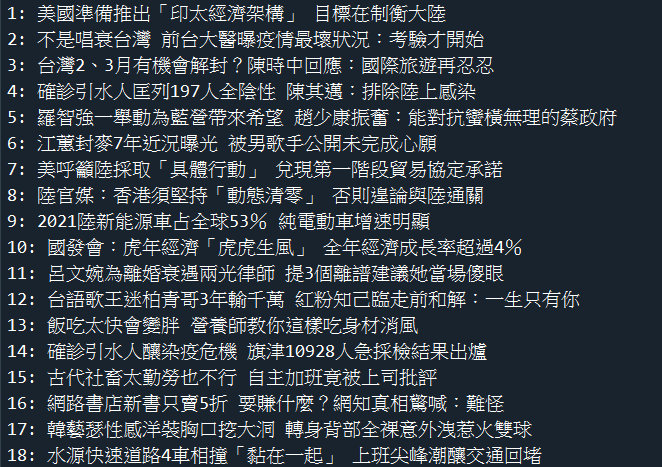
使用 物件變數. get("href") 或 物件變數["href"] 取得<a>之連結網址

部分輸出:



HW2爬取中國時報焦點新聞標題

網址:https://www.chinatimes.com/?chdtv

部份輸出:

HW3爬取百度科技新聞標題

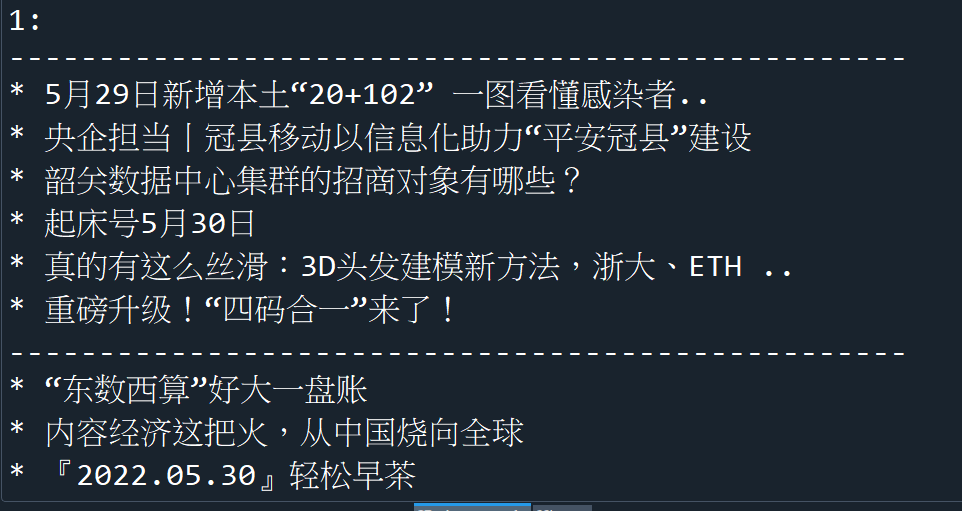
網址: <http://news.baidu.com/tech>

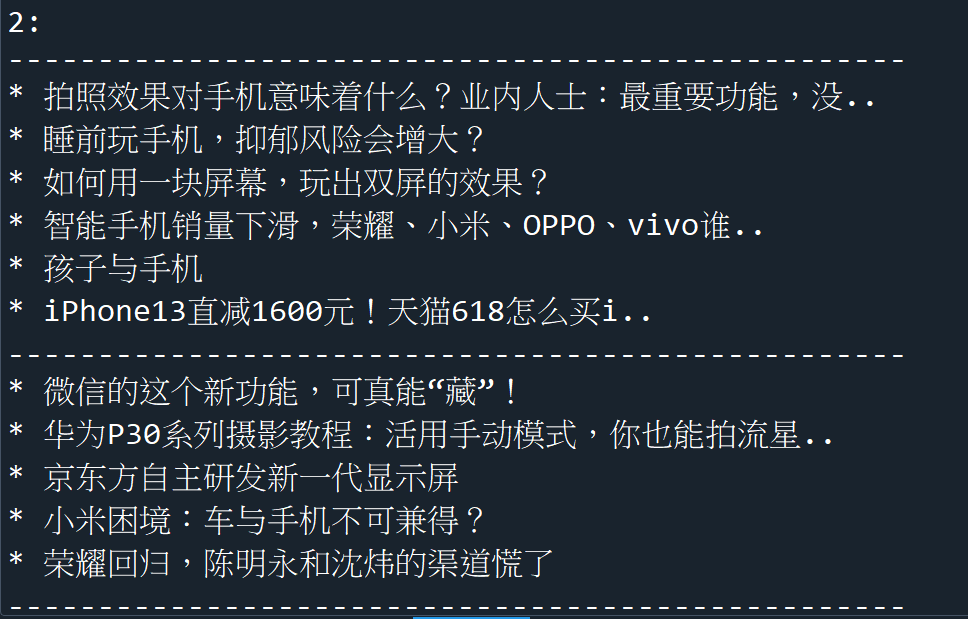
只爬中間欄位的新聞標題

<div class='middle-focus-news'…….

提示: 有多個middle-focus-news, 每個middle-focus-news內含多個<ul>,每個<ul>內含多個<ul>內含多個<li>, 每個<li>內部的<a>的text即為新聞標題, 用多層迴圈處理 (程式無法抓到全部的middle-focus-news, 只能抓到奇數項, 這不是錯誤)

部份輸出:





HW4碁峯爬蟲Ch13-3.py爬取HouseFun的台北市租房資料

(先進入<https://rent.housefun.com.tw>/網頁, 用區域找 選台北市(全區),

再按 找租屋 可得出程式要用的網址)

程式只爬取第一頁的資料, 擴充該程式使程式可爬取多頁的資料

scpt="PM({0})".format(page)

driver.execute\_script(scpt)

npage=eval(input("要爬取多少頁? 請輸入: "))

driver = webdriver.Chrome()

driver.get(url)

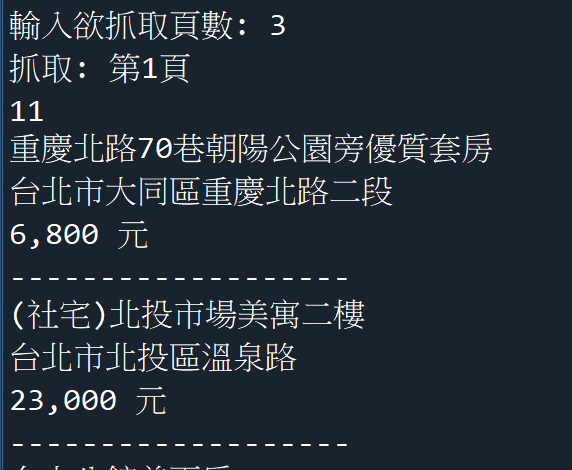
page=1

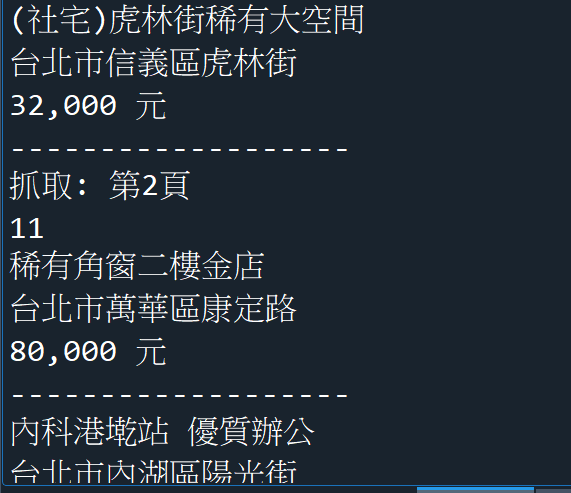
while page<=npage:

Page+=1

scpt=f"PM({page})"

driver.execute\_script(scpt)



………………………..

繳交4個程式檔,hw1.py, hw2.py, hw3.py, hw4.py