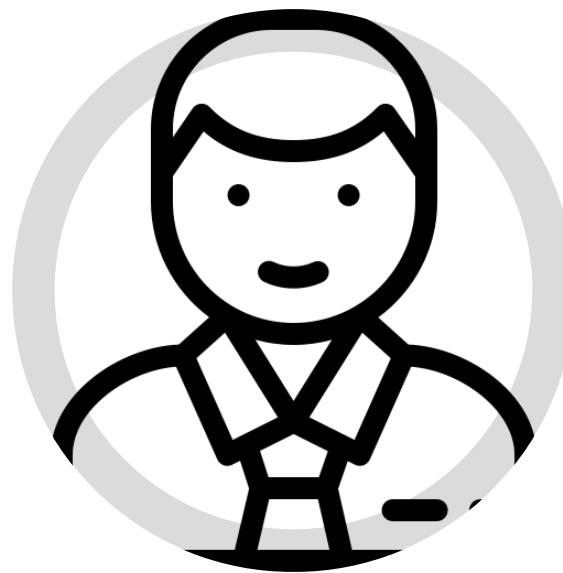


Day 14

# 靜態網頁資料爬蟲

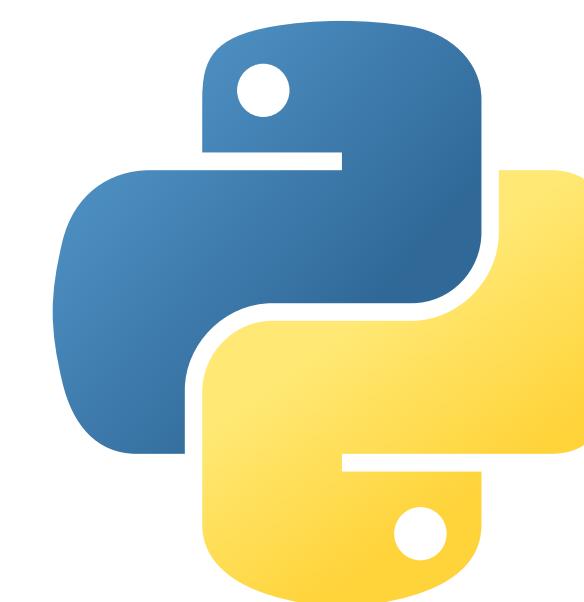


Coding 練習日



出題教練：張齊文

Yahoo電影網頁爬蟲練習



python

# 重要知識點

## Yahoo電影首頁

yahoo! 電影

搜尋

電影 · 明星 搜尋

網頁 搜尋

首頁 | 本週新片 | 上映中 | 即將上映 | 預告片 | 排行榜 | 情報站 ▾ | 電視時刻 | 追蹤我們 | [f](#)

Show ad | +

戲院總覽 | 訂票客服中心 | [冰雪奇緣2](#) | [孔劉](#) | [金馬56](#)

電影時刻 | 選擇 片名 | 請選擇片名 | 選擇 地區 | 檢索

最新電影 | 編輯推薦

  
**抓狂博物館**  
 上映日期：12月6日  
 期待度：42%  
[預告片](#) | [時刻表](#)

  
**人間失格：太宰治與**  
 上映日期：12月6日  
 期待度：88%  
[預告片](#) | [時刻表](#)

  
**絕世情歌**  
 上映日期：12月6日  
 期待度：33%  
[預告片](#) | [時刻表](#)

  
**我的靈魂是愛做的**  
 上映日期：12月6日  
 期待度：71%  
[預告片](#) | [時刻表](#)



**電影排行榜**

台北票房榜 | 全美票房榜 | 預告片榜

  
**冰雪奇緣2**  
 席捲全球票房的《冰雪奇緣》又回來了！電影故事設定在前作的3年後，在艾莎聽到來自北方神秘的...

服務中心 | 隱私權(更新) | 建議 | 關於我們的廣告 | 服務條款(更新)

# 重要知識點

範例：爬取「上映中」的電影資料：電影名稱、期待度、滿意度、電影簡介等等。

**yahoo! 電影**

搜尋 電影 · 明星 搜尋 網頁 搜尋

首頁 | 本週新片 | **上映中** | 即將上映 | 預告片 | 排行榜 | 情報站 | 電視時刻 | 追蹤我們 | f

戲院總覽 訂票客服中心 冰雪奇緣2 孔劉 金馬56

電影時刻 | 選擇 片名 | 請選擇片名 | 選擇 地區 | 檢索

**上映中**

上映日期: 2019-11-22 | 共84筆, 目前顯示1~10筆

	<b>小小夜曲</b> Little Nights, Little Love	期待度 61% 網友想看 滿意度 4.1 綜合評分
上映日期 : 2019-11-22		
這是一段由平凡的你我，創造的小小奇蹟故事！仙台車站前人來人往的天橋上，人們聚精會神地看著戶外電視牆上播送的世界重量級拳王爭霸戰，同一時間，27歲的單身上班族佐藤（三浦春馬 飾），則在一旁焦頭爛額地作著問卷調查，正當他屢屢遭到路人拒絕，準備放棄時，卻受到旁邊街頭歌手的歌聲吸引，在他專注聆聽的同時，也注意到另一個和他一樣 ... <a href="#">詳全文</a>		
<a href="#">電影介紹</a>   <a href="#">預告片</a>   <a href="#">劇照</a>   <a href="#">時刻表</a>		
	BanG Dream! FILM LIVE	期待度 64% 網友想看 滿意度 3.6 綜合評分

**內建 INTEL® CORE™ 處理器 強大效能唾手可得**

intel msi 指定 INTEL

**電影排行榜**

台北票房榜 | 全美票房榜 | 預告片榜

**冰雪奇緣2**  
席捲全球票房的《冰雪奇緣》又回來了！電影故事設定在前作的3年後，在

# 重要知識點

- 再來練習爬取有興趣的電影，其放映資訊。
- 在首頁中，選擇片名以及地區，送出查詢後，觀察其URL的變化。

The screenshot illustrates a search process on the movies.yahoo.com.tw website. It shows two browser windows side-by-side.

**Left Window (Initial Search):**

- Address bar: movies.yahoo.com.tw
- Header menu: 首頁, 本週新片, 上映中, 即將上映, 預告片, 排行榜, 情報站 ▾, 電視時刻
- Search bar: 電影時刻 (dropdown), 選擇片名 (dropdown), 新聞記者 (dropdown), 選擇地區 (dropdown), 台北市, 按钮: 查詢
- Buttons: 最新電影, 編輯推薦
- Content: Movie thumbnails for "聖母在上" (The Virgin Suicides), "女神失序人生" (The Handmaiden), "夕霧花園" (In the Land of the Living), and "布紐爾超現實人生" (Buñuel in the Land of the Living).

**Right Window (Result Page):**

- Address bar: movies.yahoo.com.tw/movietime\_result.html?movie\_id=10217&area\_id=28
- Header menu: 首頁, 信箱, 新聞, 股市, 氣象, 運動, 名人娛樂, App下載, 購物中心, 商城
- Logo: yahoo! 電影
- Search bar: 搜尋
- Header buttons: 首頁, 本週新片, 上映中, 即將上映, 預告片, 排行榜
- Search bar: 電影時刻 (dropdown), 選擇片名 (dropdown), 請選擇片名 (dropdown), 選擇地區 (dropdown)
- Buttons: 電影介紹, 預告片, 劇照, 原聲帶, 時刻表, 網友短評
- Content: Section titled "新聞記者" (The Journalist) showing the movie poster and release date (2019-11-15).

A large blue arrow points from the search bar in the left window to the URL in the right window, indicating the change in the URL structure after performing a search.

# 重要知識點複習

- 觀察URL後發現，查詢放映時刻時，網頁會送出「movie\_id」、「area\_id」資訊給伺服器。
- 接下來要練習逐步取得「movie\_id」、「area\_id」等資訊，並指定放映日期。然後組成完整的query URL，送出請求給伺服器，再爬取全部放映時間。



# 解題時間

# LET'S CRACK IT

請跳出 PDF 至官網 Sample Code & 作業

開始解題

