

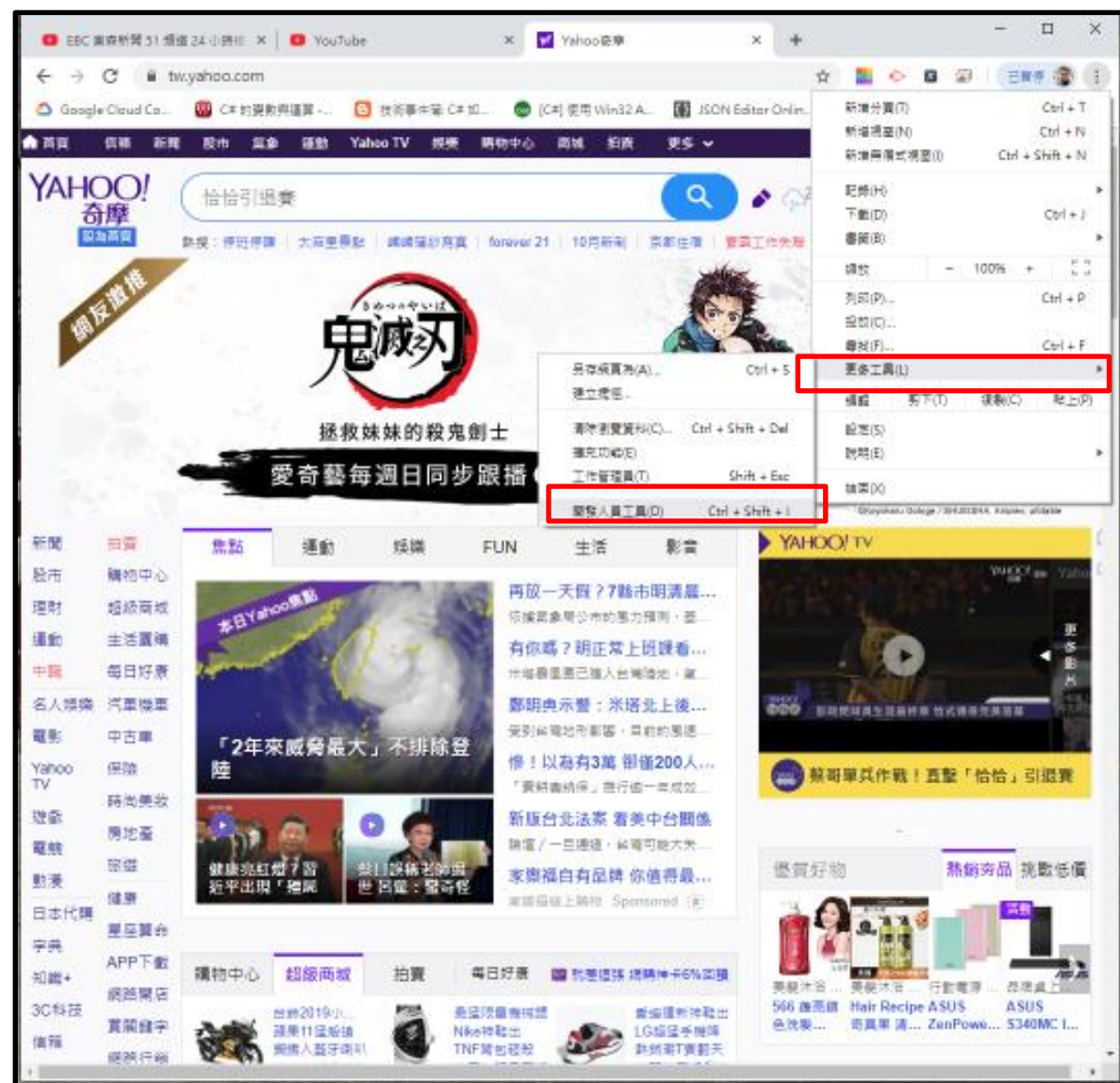


Python 程式教學 Chapter21

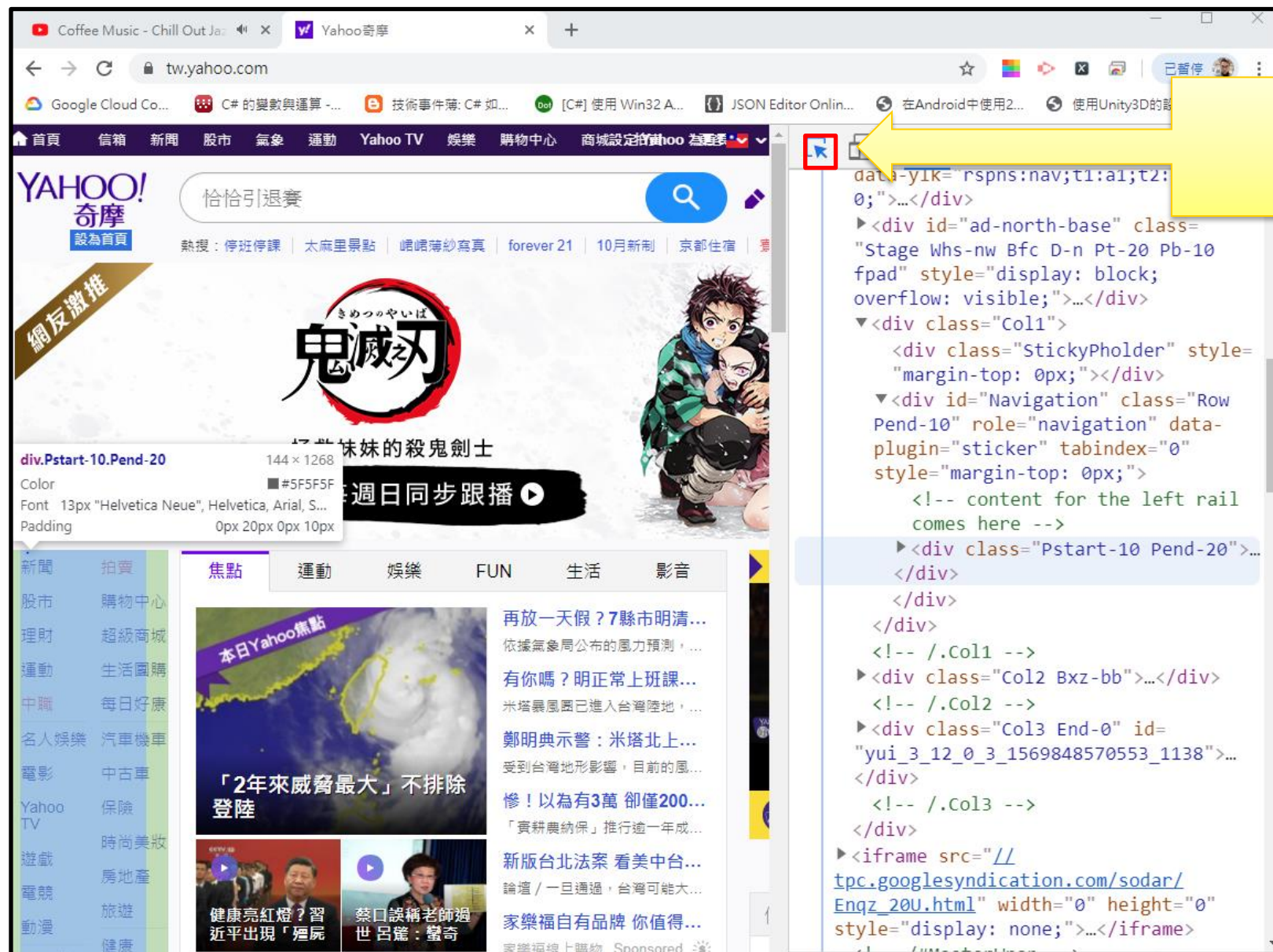
- 網頁爬蟲v1
- 使用chrome開發工具
- 爬各大網站熱門新聞
- 爬網頁表格方式(台灣銀行匯率)

■ 使用chrome → 更多工具 → 開發人員工具

(ctrl+shift+i)

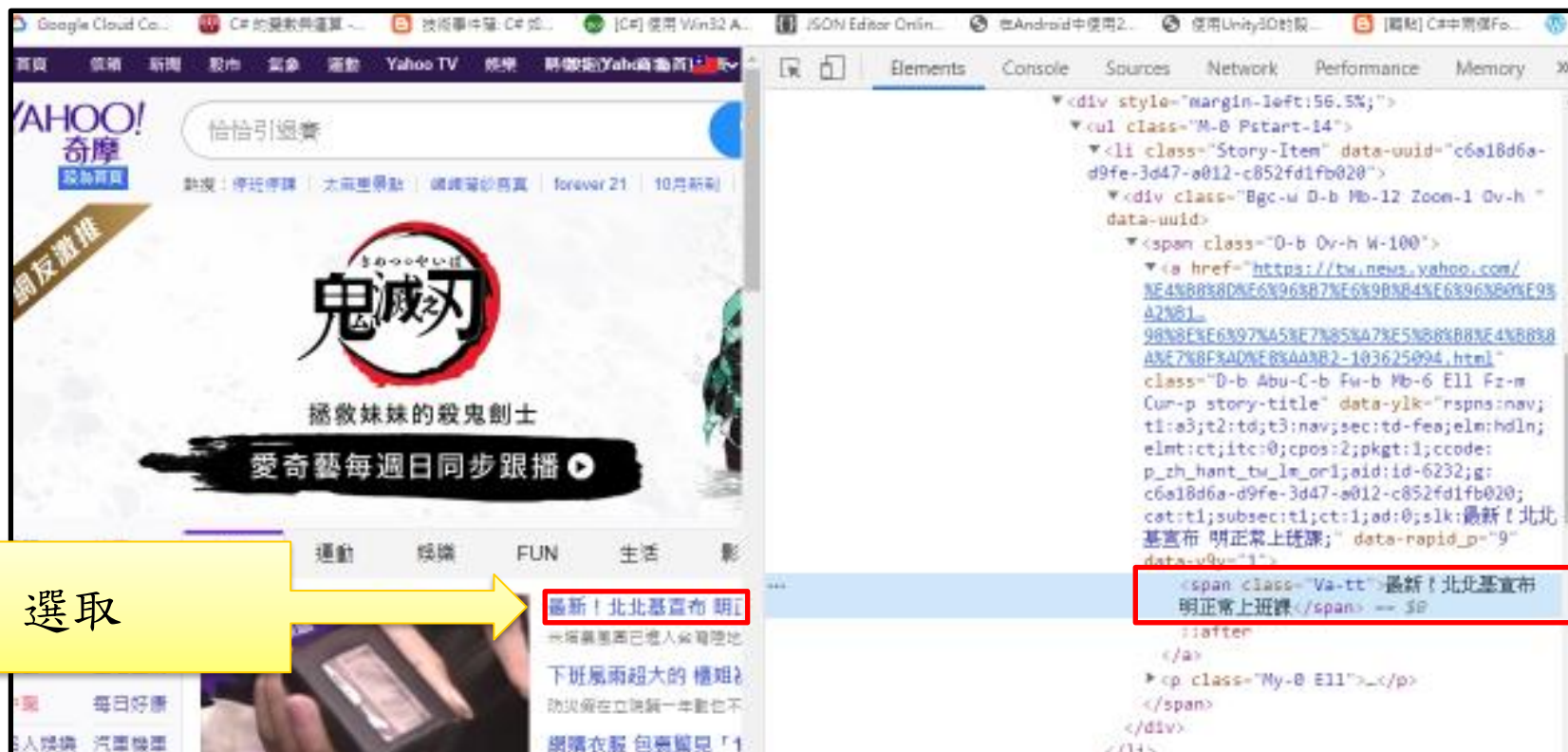


■ 使用選取元件模式



選取箭頭

檢視元件位置



The screenshot shows the Yahoo! homepage with the '鬼滅之刃' (Demon Slayer) banner. The browser's developer tools are open, showing the 'Elements' panel. A red box highlights the HTML element containing the text '最新！北北基宣布 明正堂上課課'.

選取

定位

■ 爬找yahoo奇摩熱門新聞

■ <https://tw.news.yahoo.com/most-popular>



yahoo! 新聞

搜尋

搜尋新聞 搜尋網頁

熱搜：梅穎梅花季 美國國會暴動 信貸利率最低 菲律賓空姐 谷關便宜住宿 中途島海域失聯 Vpn 美元別放定存

首頁 娛樂 國際 政治 論壇 社會地方 財經 運動 生活 科技 健康 天氣 影音 萊豬來了 byet 回顧20

Yahoo民調 台灣的防疫作為有需要再精進嗎？ 4 小時前

熱門新聞

EBC東森新聞 • 3 天前

第2波寒流要來了！濕冷5天急凍7度時間點曝 這天才回暖

本周將迎來寒流級冷空氣！中央氣象局指出，明天（5日）開始受到東北季風影響，北台灣降雨明顯，各地早晚較涼，低溫約15到19度，接著周四（7日）寒...

廣告 樂易購商城

打呼嚕不是睡的香！很可能是病，用“它”當晚止鼾！14天告別打呼嚕，再也不被踹下床了！

歡笑，那麼整個世界將和你一起歡笑；打鼾，那麼將沒有人和你一起睡覺。睡覺打鼾，好多人都有這情況打鼾者自己不算苦惱苦的是枕邊人他在旁邊早就做起...

今日新聞NOWnews • 20 小時前

test.html

滑鼠右鍵檢查



下載網頁內容

```
import requests
from bs4 import BeautifulSoup

url= 'https://tw.news.yahoo.com/most-popular'
resp = requests.get(url)
soup=BeautifulSoup(resp.text, 'lxml')
```

進行定位

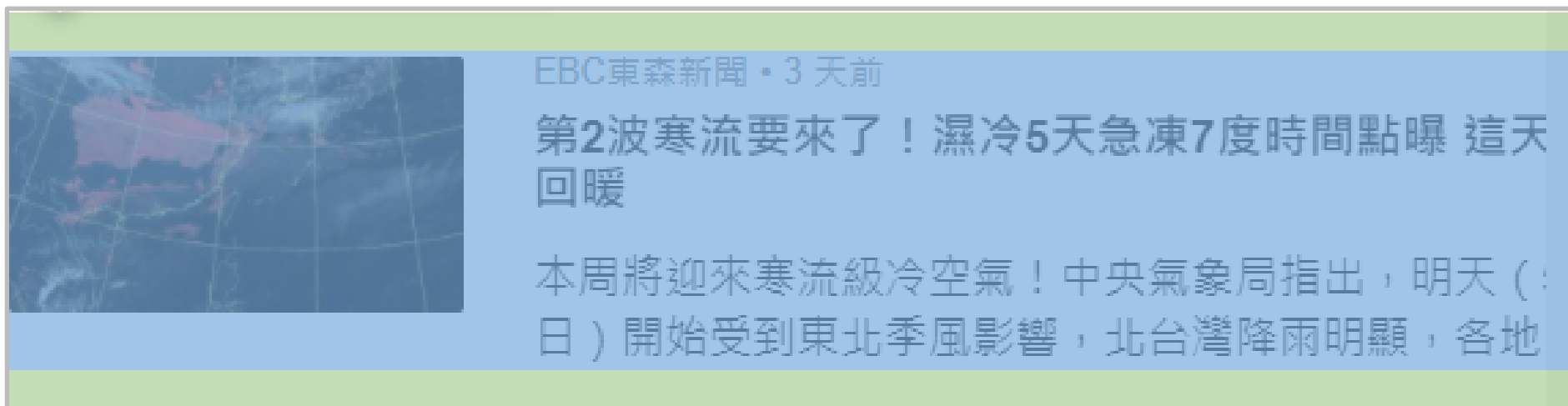
```
In [7]: 1 soup.find('u')
```

```
Out[7]: <u class="StretchedBox" data-reactid="14"></u>
```

資料沒找到?

The screenshot shows a web browser displaying the Yahoo! News website. The developer tools are open, showing the DOM tree on the right. The selected element is a link with the class "StretchedBox" and data-reactid="14". The console on the right shows the output of the soup.find('u') command, which is the same HTML element. The browser address bar shows the URL tw.news.yahoo.com/most-popular.

往上搜尋更大區間(包含href)



```
▼<div class="Cf" data-reactid="6"> == $0
  ::before
  ▶<div class="Fl(start) Pos(r) Mt(2px) W(2
    6.5%) Maw(220px)" data-reactid="7">...</div>
  ▼<div class="Ov(h) Pend(44px) Pstart(25p
```


In [12]: `1 soup.find('div',class_='Cf').text`

Out[12]: 'EBC東森新聞第2波寒流要來了！濕冷5天急凍7度時間點曝 這天才回暖本周將迎來寒流級冷空氣！中央氣象局指出，明天（5日）開始受到東北季風影響，北台灣降雨明顯，各地早晚較涼，低溫約15到19度，接著周四（7日）寒流南下，北台灣明顯轉冷，氣溫一路下降13'

In [14]: `1 soup.find('div',class_='Cf').find('a')`

Out[14]: `<u class="StretchedBox" data-reactid="14"></u><!-- react-text: 15 -->第2波寒流要來了！濕冷5天急凍7度時間點曝 這天才回暖<!-- /react-text -->`

找出內文跟連結

In [15]: `1 soup.find('div',class_='Cf').find('a').text`

Out[15]: '第2波寒流要來了！濕冷5天急凍7度時間點曝 這天才回暖'

In [16]: `1 soup.find('div',class_='Cf').find('a').get('href')`

Out[16]: `'/%E7%AC%AC2%E6%B3%A2%E5%AF%92%E6%B5%81%E8%A6%81%E4%BE%86%E4%BA%86-%E6%BF%95%E5%86%B75%E5%A4%A9%E6%80%A5%E5%87%8D7%E5%BA%A6%E6%99%82%E9%96%93%E9%BB%9E%E6%9B%9D-%E9%80%99%E5%A4%A9%E6%89%8D%E5%9B%9E%E6%9A%96-125900588.html'`

導入連結

In [17]: `1 url+soup.find('div',class_='Cf').find('a').get('href')`

Out[17]: `'https://tw.news.yahoo.com/most-popular/%E7%AC%AC2%E6%B3%A2%E5%AF%92%E6%B5%81%E8%A6%81%E4%BE%86%E4%BA%86-%E6%BF%95%E5%86%B75%E5%A4%A9%E6%80%A5%E5%87%8D7%E5%BA%A6%E6%99%82%E9%96%93%E9%BB%9E%E6%9B%9D-%E9%80%99%E5%A4%A9%E6%89%8D%E5%9B%9E%E6%9A%96-125900588.html'`

- 查找所有新聞
 - 使用find_all

查找所有新聞

```
In [29]: 1 len(soup.find_all('div',class_='Cf'))|
```

```
Out[29]: 20
```

```
In [30]: 1 datas=soup.find_all('div',class_='Cf')
2 datas
```

...

```
In [31]: 1 for data in datas:
2     print(data.find('a').text)
3     print(url+data.find('a').get('href'))
```

...

```
In [32]: 1 [[data.find('a').text,url+data.find('a').get('href')] for data in datas]
```

...

```
In [33]: 1 import pandas as pd
```

```
In [34]: 1 pd.DataFrame([[data.find('a').text,url+data.find('a').get('href')] for data in datas])
```

■ 程式練習

■ NBA 新聞查找

■ <https://nba.udn.com/nba/news/>

補強後場火力 籃網將簽下強森
根據《ESPN》知名記者沃納羅斯基 (Adrian Wojnarowski) 的消息，布魯克林籃網將簽...
<https://nba.udn.com/nba/story/6754/4657332>

湖人填補布萊德利缺口 傳史密斯成頭號人選
湖人主力後衛布萊德利 (Avery Bradley) 因家庭因素確定不參與復賽，這也讓湖人後場戰力出現...
<https://nba.udn.com/nba/story/6780/4657009>

讚詹皇堪稱史上最佳 格林盼其未來出馬選總統
場上有著過去騎士與勇士連續4年碰頭總冠軍賽宿敵情懷，格林 (Draymond Green) 表示，自己...
<https://nba.udn.com/nba/story/6780/4656942>

確定婉拒參與復賽 布萊德利歸因家庭因素
日前厄文 (Kyrie Irving) 表示挺黑人平權而罷賽的立場，當時有2位湖人球員加以響應，哈沃德...
<https://nba.udn.com/nba/story/6780/4656839>

昔嘆歧視數十年未變 小男孩作為今感動羅素
傳奇球星羅素 (Bill Russell) 日前才感慨，自己活了86歲如今竟又見到黑人平權運動再起，證...
<https://nba.udn.com/nba/story/6780/4656759>

相信哈沃德將參與復賽 格林：我們都支持他
就在厄文 (Kyrie Irving) 拋出不支持賽季重啟的看法後，隨即引發了復賽與否的巨大討論，其中...
<https://nba.udn.com/nba/story/6780/4656707>

車手挺平權疑收套索警告 詹皇柯瑞齊發推持相挺
日前以實際行動力挺黑人平權運動，NASCAR車手華勒斯 (Bubba Wallace) 驚傳所屬停車位...
<https://nba.udn.com/nba/story/6780/4656687>

透露離開溜馬原因 喬治：球團拒絕追逐明星大前鋒
曾經被視為溜馬救世主的喬治 (Paul George)，在2017年夏天透過交易轉戰雷霆，兩年之後，...
<https://nba.udn.com/nba/story/6780/4656661>

傳火箭將簽下努瓦巴 預計本季不復出繼續養傷
日前才傳出對於2月份遭釋出的後衛強森 (Tyler Johnson) 有興趣，《The Athleti...
<https://nba.udn.com/nba/story/6780/4656643>

離開活塞接受聯盟延攬 羅斯出任籃球事務營運副總
根據《ESPN》記者史比爾斯 (Marc J. Spears) 報導，NBA官方將聘請球員出身的羅斯 (...
<https://nba.udn.com/nba/story/6780/4656601>

輸出成csv格式

自動儲存 ☒ 關閉 📄 ↶ ↷ 🔍 nba_news.csv 🔍 搜尋

檔案 常用 插入 繪圖 頁面配置 公式 資料 校閱 檢視 說明

C16 ✕ ✓ fx

	A	B	C
1		title	link
2	0	美國國會山莊爆衝突 瑞佛斯：令人	https://nba.udn.com//nba/story/6780/5155730
3	1	開季連8場至少20分 塞克斯頓打破詹	https://nba.udn.com//nba/story/6780/5155740
4	2	希爾德關鍵四分打 華頓：他會找回	https://nba.udn.com//nba/story/6780/5155690
5	3	近3年首度打背靠背 雷納德：身體已	https://nba.udn.com//nba/story/6780/5155487
6	4	太陽三分砲火擊退暴龍 席亞康：我	https://nba.udn.com//nba/story/6780/5155409
7	5	雷納德第四節接管戰局 快艇客場力	https://nba.udn.com//nba/story/6780/5155099
8	6	爵士慘遭逆轉 小瑞佛斯：這支尼克	https://nba.udn.com//nba/story/6780/5154988
9	7	亞當斯首次大三元做白工 希爾關	https://nba.udn.com//nba/story/6780/5154879
10	8	首輪菜鳥致命補籃 綠衫軍驚險滅火	https://nba.udn.com//nba/story/6780/5154811
11	9	左膝前十字韌帶撕裂 佛爾茲賽季提	https://nba.udn.com//nba/story/6779/5154747
12			

爬表格資料

爬取人事行政局資料

<https://www.dgpa.gov.tw/typh/daily/nds.html>

表格資料

dgpa.gov.tw/typh/daily/nds.html

Google Cloud Co... C# 的變數與運算... 技術事件簿: C# 如... [C#] 使用 Win32 A... JSON Editor Onlin... 在Android中使用2... 使用Unity3D的設...

行政院人事行政總處 全球資訊網

108年 9月 30日 天然災害停止上班及上課情形

資料來源：各縣市政府

縣市名稱	是否停止上班上課情形
基隆市	今天停止上班、停止上課。 明天照常上班、照常上課。
臺北市	今天停止上班、停止上課。 明天照常上班、照常上課。
新北市	今天停止上班、停止上課。 明天照常上班、照常上課。
桃園市	今天停止上班、停止上課。 明天照常上班、照常上課。 復興區:今天停止上班、停止上課。 復興區:明天停止上班、停止上課。
新竹市	今天晚上06:00起停止上班、停止上課。 明天照常上班、照常上課。
新竹縣	今天停止上班、停止上課。 明天照常上班、照常上課。
苗栗縣	今天晚上06:00起停止上班、停止上課。 明天照常上班、照常上課。
臺中市	今天照常上班、照常上課。 明天照常上班、照常上課。
彰化縣	今天照常上班、照常上課。 明天照常上班、照常上課。
雲林縣	今天照常上班、照常上課。 明天照常上班、照常上課。
南投縣	今天照常上班、照常上課。 明天照常上班、照常上課。
嘉義市	今天照常上班、照常上課。 明天照常上班、照常上課。
嘉義縣	今天照常上班、照常上課。 明天照常上班、照常上課。
臺南市	今天照常上班、照常上課。 明天照常上班、照常上課。

■ 定位表格



dgpa.gov.tw/typh/daily/nds.html

English 中文

行政院人事行政總處 全球資訊網

108年 9月 30日 天然災害停止上班及上課情形

更新時間：2019/09/30 21:49:04
歷次天然災害停止上班上課訊息
民眾洽詢天然災害停止上班及上課各地方政府連絡一覽表
天然災害停止上班及上課作業辦法
天然災害停止上班及上課作業Q&A

資料來源：各縣市政府

tbody.table_body 786 x 1353

是否停止上班上課情形	
基隆市	今天停止上班、停止上課。 明天照常上班、照常上課。
臺北市	今天停止上班、停止上課。 明天照常上班、照常上課。
新北市	今天停止上班、停止上課。 明天照常上班、照常上課。
桃園市	今天停止上班、停止上課。 明天照常上班、照常上課。 復興區:今天停止上班、停止上課。 復興區:明天停止上班、停止上課。
新竹市	今天晚上06:00起停止上班、停止上課。 明天照常上班、照常上課。
新竹縣	今天停止上班、停止上課。 明天照常上班、照常上課。
苗栗縣	今天晚上06:00起停止上班、停止上課。 明天照常上班、照常上課。
臺中市	今天照常上班、照常上課。 明天照常上班、照常上課。
彰化縣	今天照常上班、照常上課。 明天照常上班、照常上課。

```
<html lang="zh-TW">
<head></head>
<body>
<title>行政院人事行政總處全球資訊網-天然災害停止上班及上課情形查詢</title>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<meta content="MSHTML 6.00.2800.1400" name="GENERATOR">
<meta http-equiv="X-UA-Compatible" content="IE=edge">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<style>...</style>
<script type="text/javascript" async src="https://ssl.google-analytics.com/ga.js"></script>
<script type="text/javascript">...</script>
<div class="Version_btn_Div">...</div>
<div class="Header_Div">...</div>
<div id="Content">
  <div class="f_right Content_Update">...</div>
  <br>
  <br>
  <div style="clear:both;">...</div>
  <div style="clear:both;">...</div>
  <div style="clear:both;">...</div>
  <table id="Table" rules="All">
    <tbody>...</tbody>
    <tbody class="Table_Body">
      <tr>
        <td headers="city_Name" valign="center" align="middle" width="13%">
          <font>基隆市</font> == $0
        </td>
        <td headers="StopWorkSchool_Info" valign="center" align="left" width="70%">...</td>
      </tr>
      <tr>...</tr>
      <tr>...</tr>
      <tr>...</tr>
      <tr>...</tr>
      <tr>...</tr>
    </tbody>
  </table>
</div>
</body>
```

讀取網頁內容

```
url='https://www.dgpa.gov.tw/typh/daily/nds.html'  
resp=requests.get(url)  
resp.encoding = 'utf-8'  
soup=BeautifulSoup(resp.text, 'lxml')  
print(soup)
```

發現該網頁編碼格式為utf-8

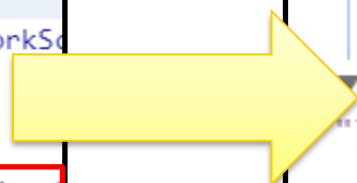
```
<html lang="zh-TW">  
<body>  
<title>行政院人事行政總處全球資訊網-天然災害停止上班及上課情形查詢</title>  
<meta content="text/html; charset=utf-8" http-equiv="Content-Type"/>  
<meta content="MSHTML 6.00.2800.1400" name="GENERATOR"/>  
<meta content="IE=edge" http-equiv="X-UA-Compatible"/>  
<meta content="width=device-width, initial-scale=1.0" name="viewport"/>
```

utf-8

```
<html lang="zh-TW">  
<body>  
<title>行政院人事行政總處全球資訊網-天然災害停止上班及上課情形查詢</title>  
<meta content="text/html; charset=utf-8" http-equiv="Content-Type"/>  
<meta content="MSHTML 6.00.2800.1400" name="GENERATOR"/>  
<meta content="IE=edge" http-equiv="X-UA-Compatible"/>  
<meta content="width=device-width, initial-scale=1.0" name="viewport"/>  
<style>
```


從tr標籤找出td

```
<tbody class="Table Body">
  <tr>
    <td headers="city_Name"
    </td>
    <td headers="StopWorkSchool_Info"
    "70%">...</td>
  </tr>
  <tr>
    <td headers="city_Name"
    </td>
    <td headers="StopWorkSchool_Info"
    "70%">...</td>
  </tr>
  <tr>
    <td headers="city_Name"
    </td>
    <td headers="StopWorkSchool_Info"
    "70%">...</td>
  </tr>
  <tr>
    <td headers="city_Name"
    </td>
    <td headers="StopWorkSchool_Info"
    "70%">...</td>
  </tr>
```



```
<tr>
  <td headers="city_Name" valign="center" align="middle" width="13%"> == $
    <font>基隆市</font>
  </td>
  <td headers="StopWorkSchool_Info" valign="center" align="left" width=
  "70%">
    <font color="#FF0000">今天停止上班、停止上課。 </font>
    <br>
    <font color="#000080">明天照常上班、照常上課。 </font>
  </td>
</tr>
<tr>
  <td headers="city_Name" valign="center" align="middle" width="13%">
    <font>臺北市</font>
  </td>
  <td headers="StopWorkSchool_Info" valign="center" align="left" width=
  "70%">
    <font color="#FF0000">今天停止上班、停止上課。 </font>
    <br>
    <font color="#000080">明天照常上班、照常上課。 </font>
  </td>
</tr>
```

```
data=[]  
for tr in trs:  
    tds=tr.find_all('td')  
    data.append([td.text.strip() for td in tds])
```

```
data=data[0:-1]  
print(data)
```

```
[['基隆市', '今天停止上班、停止上課。 明天照常上班、照常上課。'],  
 ['臺北市', '今天停止上班、停止上課。 明天照常上班、照常上課。'],  
 ['新北市', '今天停止上班、停止上課。 明天照常上班、照常上課。'],  
 ['桃園市', '今天停止上班、停止上課。 明天照常上班、照常上課。 復興區:今天停止上班、  
 停止上課。 復興區:明天停止上班、停止上課。'],  
 ['新竹市', '今天晚上06:00起停止上班、停止上課。 明天照常上班、照常上課。'],  
 ['新竹縣', '今天停止上班、停止上課。 明天照常上班、照常上課。'],  
 ['苗栗縣', '今天晚上06:00起停止上班、停止上課。 明天照常上班、照常上課。'],  
 ['臺中市', '今天照常上班、照常上課。 明天照常上班、照常上課。'],  
 ['彰化縣', '今天照常上班、照常上課。 明天照常上班、照常上課。'],  
 ['雲林縣', '今天照常上班、照常上課。 明天照常上班、照常上課。'],  
 ['南投縣', '今天照常上班、照常上課。 明天照常上班、照常上課。'],  
 ['嘉義市', '今天照常上班、照常上課。 明天照常上班、照常上課。'],  
 ['嘉義縣', '今天照常上班、照常上課。 明天照常上班、照常上課。'],  
 ['臺南市', '今天照常上班、照常上課。 明天照常上班、照常上課。'],  
 ['高雄市', '今天照常上班、照常上課。 明天照常上班、照常上課。'],  
 ['屏東縣', '今天照常上班、照常上課。 明天照常上班、照常上課。'],  
 ['宜蘭縣', '今天停止上班、停止上課。 明天照常上班、照常上課。'],  
 ['花蓮縣', '今天停止上班、停止上課。 明天照常上班、照常上課。'],  
 ['臺東縣',  
  '明天照常上班、照常上課。 成功鎮:今天上午10:00起停止上班、停止上課。 東河鄉:今天上  
  午10:00起停止上班、停止上課。 長濱鄉:今天上午10:00起停止上班、停止上課。 綠島鄉:今  
  天停止上班、停止上課。 蘭嶼鄉:今天停止上班、停止上課。'],
```


程式練習(爬台灣銀行匯率)

- <https://rate.bot.com.tw/xrt?Lang=zh-TW>

抓取表格資料

rate.bot.com.tw/xrt?Lang=zh-TW

Google Cloud Co... C# 的變數與運算... 技術事件簿: C# 如... [C#] 使用 Win32 A... JSON Editor Onlin...

請注意：1. 本表資料僅供參考，不代表實際交易匯率。
2. 「網路銀行」及「Easy購線上申購現鈔或旅支」之實際交易匯率，以交易時顯示之匯率為準。
3. 臨櫃實際交易匯率以交易時本行匯率為準。
4. 本網頁廣告匯率資訊為靜態顯示，顯示之廣告匯率資訊不會隨後續變動而自動更新資訊，欲得知本行最新廣告匯率資訊請按「取得最新報價」鈕。

取得最新報價 線上申購外幣現鈔或旅支

牌價最新掛牌時間：2019/09/30 16:00

幣別	現金匯率		即期匯率		遠期匯率	歷史匯率
	本行買入	本行賣出	本行買入	本行賣出		
美金 (USD)	30.64	31.31	30.99	31.09	查詢	查詢
港幣 (HKD)	3.802	4.006	3.928	3.988	查詢	查詢
英鎊 (GBP)	36.99	39.11	37.99	38.41	查詢	查詢
澳幣 (AUD)	20.58	21.36	20.85	21.08	查詢	查詢
加拿大幣 (CAD)	22.96	23.87	23.35	23.57	查詢	查詢
新加坡幣 (SGD)	21.89	22.8	22.38	22.56	查詢	查詢
瑞士法郎 (CHF)	30.46	31.66	31.12	31.41	查詢	查詢
日圓 (JPY)	0.2785	0.2913	0.2858	0.2898	查詢	查詢
南非幣 (ZAR)	-	-	2	2.08	查詢	查詢
瑞典幣 (SEK)	2.78	3.3	3.12	3.22	查詢	查詢
紐元 (NZD)	18.97	19.82	19.35	19.55	查詢	查詢
泰幣 (THB)	0.8919	1.0799	1.0005	1.0405	查詢	查詢
菲律賓比索 (PHP)	0.5252	0.6582	-	-	查詢	查詢
印尼幣 (IDR)	0.00188	0.00258	-	-	查詢	查詢
歐元 (EUR)	33.13	34.47	33.75	34.15	查詢	查詢
韓元 (KRW)	0.02418	0.02808	-	-	查詢	查詢
越南盾 (VND)	0.00097	0.00147	-	-	查詢	查詢
馬來幣 (MYR)	6.3	7.944	-	-	查詢	查詢
人民幣 (CNY)	4.253	4.415	4.325	4.375	查詢	查詢

輸出成CSV

A	B	C	D	E
幣別	現金匯率	即期匯率	遠期匯率	歷史匯率
美金 (USD)	30.64	31.31	30.99	31.09
港幣 (HKD)	3.802	4.006	3.928	3.988
英鎊 (GBP)	36.99	39.11	37.99	38.41
澳幣 (AUD)	20.58	21.36	20.85	21.08
加拿大幣 (CAD)	22.96	23.87	23.35	23.57
新加坡幣 (SGD)	21.89	22.8	22.38	22.56
瑞士法郎 (CHF)	30.46	31.66	31.12	31.41
日圓 (JPY)	0.2785	0.2913	0.2858	0.2898
南非幣 (ZAR)	-	-	2	2.08
瑞典幣 (SEK)	2.78	3.3	3.12	3.22
紐元 (NZD)	18.97	19.82	19.35	19.55
泰幣 (THB)	0.8919	1.0799	1.0005	1.0405
菲國比索 (PHP)	0.5252	0.6582	-	-
印尼幣 (IDR)	0.00188	0.00258	-	-
歐元 (EUR)	33.13	34.47	33.75	34.15
韓元 (KRW)	0.02418	0.02808	-	-
越南盾 (VND)	0.00097	0.00147	-	-
馬來幣 (MYR)	6.3	7.944	-	-
人民幣 (CNY)	4.253	4.415	4.325	4.375