



# 聯合新聞網爬蟲程式執行

## 動機

想利用爬蟲的方式爬取聯合新聞網最新的新聞，並生成文字雲以獲取現在最新的消息。

1. **數據可訪問性**：提供新聞連結的便捷訪問，否則手動收集將非常耗時
2. **自動化收集**：消除了多個頁面之間手動搜索的工作
3. **結構化數據存儲**：將收集的信息格式化為易於分析的 CSV 格式
4. **文字雲分析**：能夠透過文字雲，找到現在熱門的新聞。

## 執行結果

執行爬蟲程式後，我們成功從聯合新聞網站獲取了即時的新聞連結並生成文字雲。以下是執行結果的實際展示：

### 1. 最新新聞擷取成果

從下圖可以看到，我們成功最新新聞標題，包含了：

- 頁數
- 該頁數連結
- 抓取篇數

```
...  抓取第 1 頁: https://udn.com/api/more?page=1&channelId=1&cate\_id=0&type=breaknews
      抓取第 2 頁: https://udn.com/api/more?page=2&channelId=1&cate\_id=0&type=breaknews
      抓取第 3 頁: https://udn.com/api/more?page=3&channelId=1&cate\_id=0&type=breaknews
      抓取第 4 頁: https://udn.com/api/more?page=4&channelId=1&cate\_id=0&type=breaknews
      抓取第 5 頁: https://udn.com/api/more?page=5&channelId=1&cate\_id=0&type=breaknews
Building prefix dict from the default dictionary ...
✓ 爬取完成，共抓到 100 篇新聞，已存入 udn_news.csv
Dumping model to file cache C:\Users\User\AppData\Local\Temp\jieba.cache
Loading model cost 0.809 seconds.
Prefix dict has been built successfully.
```

## 2. CSV檔

系統運行日誌顯示了爬蟲的完整執行過程：

```
1 標題,連結,時間
2 新建室內打擊場!銅鑼國小棒球隊5戰全勝 捧車全國賽,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656306.jpg&s=Y&x=311&y=0&sw=3378&sh=2252&exp=3600,{ 'date': '2025-03-14 19:58' }
3 Fed降息動向?經濟學者估時間落在這2個月 全年共降2碼,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656304.jpg&s=Y&x=0&y=0&sw=4324&sh=2883&exp=3600,{ 'date': '2025-03-14 19:58' }
4 泊連逾200萬巴人?美以官員擬在非洲重新安置 蘇丹官員稱已拒絕,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656294.jpg&s=Y&x=0&y=0&sw=2400&sh=1600&exp=3600,{ 'date': '2025-03-14 19:52' }
5 撿刀還是撿便宜?美股賣壓壓頂 散戶照買不誤,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656272.jpg&s=Y&x=0&y=0&sw=7981&sh=5324&exp=3600,{ 'date': '2025-03-14 19:42' }
6 取得國產鮮乳標章才能稱「鮮乳」 預計明年7月上路,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656271.jpg&s=Y&x=0&y=0&sw=5568&sh=3712&exp=3600,{ 'date': '2025-03-14 19:40' }
7 駐教廷大使李世明退職 亞非司長賀忠義接任,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656269.jpeg&s=Y&x=0&y=0&sw=1600&sh=1067&exp=3600,{ 'date': '2025-03-14 19:39' }
8 川普將棄守貿易戰?策略師：美股跌到「這點」時就有可能,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656261.jpg&s=Y&x=0&y=0&sw=5500&sh=3668&exp=3600,{ 'date': '2025-03-14 19:38' }
9 「你毀了這裡！」范斯偕妻聽古典樂 大廳突爆噓聲尷尬影片瘋傳,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656260.jpg&s=Y&x=0&y=0&sw=1280&sh=853&exp=3600,{ 'date': '2025-03-14 19:37' }
10 電價漲不停 經濟部：過去電價調漲因國際燃料價格大漲,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/04/realtime/31597652.jpg&s=Y&x=0&y=0&sw=1374&sh=916&exp=3600,{ 'date': '2025-03-14 19:36' }
11 機票1位難求?金門縣府調查旅客需求 爭取清明加班機,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656253.jpg&s=Y&x=0&y=30&sw=4032&sh=2688&exp=3600,{ 'date': '2025-03-14 19:33' }
12 影!「阿哩嘎豆!」8旬日籍僑迷途捷運站 土城警聯手移民署助返家,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656251.png&s=Y&x=184&y=72&sw=839&sh=560&exp=3600,{ 'date': '2025-03-14 19:33' }
13 鮑魚大亨鍾文智判刑30年5月失聯 北檢通緝今發文註銷護照,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/13/realtime/31650567.jpg&s=Y&x=0&y=328&sw=3648&sh=2432&exp=3600,{ 'date': '2025-03-14 19:31' }
14 性侵犯死後被毀屍 德州殯葬業者割下生殖器塞嘴羞辱,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/0/31656153.jpg&s=Y&x=0&y=0&sw=1279&sh=853&exp=3600,{ 'date': '2025-03-14 19:28' }
15 標示再轉彎!「鮮乳」限國產乳使用 進口僅能稱「牛乳」,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656205.jpg&s=Y&x=0&y=39&sw=5500&sh=3666&exp=3600,{ 'date': '2025-03-14 19:19' }
16 顧問自我認同是隻「烏龜」 美國俄勒岡州心理健康委員會招募惹議,https://udn.comhttps://pgw.udn.com.tw/gw/photo.php?u=https://uc.udn.com.tw/photo/2025/03/14/realtime/31656204.jpg&s=Y&x=0&y=0&sw=1279&sh=853&exp=3600,{ 'date': '2025-03-14 19:19' }
```

此csv檔保存了標題、連結以及抓取時間。

### 3. 文字雲



## 獲取近100篇新聞最提及的關鍵字!