

作業說明：
我在作業三爬取了Yahoo股市的"單日熱門股"的各項數據，但因為這次作業是以文字進行關聯圖分析，作業三都是數字為主，幾乎沒有文字，所以後來對作業三進行一些更動，讓其可以根據輸入的內容爬取指定的文字

```
[ ] stock_search=input()
    link=("https://tw.stock.yahoo.com/quote/" + stock_search)
    print(link)
```

1314.TW
<https://tw.stock.yahoo.com/quote/1314.TW>

輸入股票代號後就可得到網址並進行爬蟲，而爬蟲的內容為該股票的當日相關新聞，並存入google試算表

```
spreadsheet_id = '1Ci8bo2Fw1IyyuYvKP9icxP1olukYwAvAmPLYHfVObes'
sheet_name = 'stock_news title/stock_news content'

def truncate_to_limit(cell_value, limit=10000):
    if len(cell_value) > limit:
        return cell_value[:limit] + '... [內容過長，已停止寫入]'
    else:
        return cell_value

[ ] df_truncated = df.applymap(lambda x: truncate_to_limit(str(x)))
df_truncated
```

web_crawler ☆ 📁 ☁
檔案 編輯 查看 插入 格式 資料 工具 擴充功能 說明

🔍 ⏪ ⏩ 🔄 📄 100% NT\$ % 0.00 123 預設 (...) - 10 + B I ⚡ A 🗑

H2B ▾ | ㄆ

	A	B	C	D
1	Title	Content		
2	《熱門族群》內資領頭衝 資產股發燙	【時報-台北電】台股29日反彈行情熱絡，類股維持多頭		
3	中石化啟動轉型 營運添力	中石化 (1314) 啟動精細 / 特用化學品、電子級化工材		
4	中石化利基轉型就位發酵 營運添力	中石化 (1314) 啟動精細 / 特用化學品、電子級化工材		
5	《熱門族群》中石化、炎洲 資產開發報喜	【時報-台北電】中石化 (1314)、炎洲 (4306)、大洋		
6	中石化、炎洲 資產開發報喜	中石化 (1314)、炎洲 (4306)、大洋 (1321) 資產開		
7	資產開發報喜 中石化、炎洲、大洋多元收益添利	中石化 (1314)、炎洲 (4306)、大洋 (1321) 資產開		
8	《塑膠股》中石化：最嚴峻狀況已過	【時報-台北電】中石化 (1314) 12日法說會，總經理陳		
9	中石化：最嚴峻狀況已過	中石化 (1314) 12日法說會，總經理陳穎俊表示，已內		
10	中石化：終端需求回溫 上半年營運添力	中石化 (1314) 12日法說會，總經理陳穎俊表示，已內		
11	【公告】召開中石化113年股東常會事宜	日期：2024年04月11日公司名稱：中石化(1314)主 旨：		
12	【公告】中石化 2024年3月合併營收27.3億元 年增17.04%	日期： 2024 年 04 月 10日上市公司：中石化(1314)單位		
13	PVC、CPL漲價 台塑、華夏、聯成、中石化營運加分	PVC亞洲4月外銷報價月漲10美元 / 噸，隨著印度節慶結		
14	華友聯 標得中石化亞灣建地	高雄房產因台積電進駐更形火熱，中石化位於高雄亞灣		
15	【公告】中石化處分高雄市前鎮區經貿段五小段1-1地號土地情形	日期：2024年03月26日公司名稱：中石化(1314)主 旨：		
16	《塑膠股》華友聯斥資20億元取得高雄前鎮區土地	【時報-台北電】華友聯(4436)公告斥資約20億元，取得		

作業四即是根據此試算表資料進行LLM與文字視覺化關聯圖的分析

```

) from google.colab import auth
  auth.authenticate_user()

import gspread
from google.auth import default
creds, _ = default()

gc = gspread.authorize(creds)

import pandas as pd
# read data and put it in a dataframe
# 在 google 工作表載入 gsheets
gsheets = gc.open_by_url('https://docs.google.com/spreadsheets/d/1Ci8bo2Fw1IyyuYvKP9icxP1olukYwAvAmPlYHfV0bcs/edit?usp=sharing')
# 從 gsheets 的工作表
sheets1 = gsheets.worksheet('stock_news title/stock_news content').get_all_values()

df = pd.DataFrame(sheets1[1:], columns=sheets1[0])

] df.head()

```

	Title	Content
0	《熱門族群》內資領頭衝 資產股發燙	【時報-台北電】台股29日反彈行情熱絡，類股維持多頭輪動格局，在本土資金領軍衝鋒下，資產概念...
1	中石化啟動轉型 營運添力	中石化（1314）啟動精細／特用化學品、電子級化工材料、能源、氢能低碳產品等四大轉型計畫，力...
2	中石化利基轉型就位發酵 營運添力	中石化（1314）啟動精細／特用化學品、電子級化工材料、能源及氢能低碳產品四大轉型計畫，力拼...
3	《熱門族群》中石化、災洲 資產開發報喜	【時報-台北電】中石化（1314）、災洲（4306）、大洋（1321）資產開發部署就位，多元...
4	中石化、災洲 資產開發報喜	中石化（1314）、災洲（4306）、大洋（1321）資產開發部署就位，多元收益添利。中石化...

▲將試算表資料讀入並輸出

```

[ ] def remove_spaces(text):
    # 使用 replace 方法來刪除空格
    return text.replace(" ", "")

output_text = remove_spaces(summary)
output_text

'時報台北電台股日反彈行情熱絡類股維持多頭輪動格局在本土資金領軍衝鋒下資產概念股強勢
出，有「」的股票是指有「」。」'

```

```

[ ] import jieba
import jieba.analyse

jieba.analyse.extract_tags(output_text, topK=10)

Building prefix dict from the default dictionary ...
DEBUG:jieba:Building prefix dict from the default dictionary ...
Dumping model to file cache /tmp/jieba.cache
DEBUG:jieba:Dumping model to file cache /tmp/jieba.cache
Loading model cost 0.889 seconds.
DEBUG:jieba:Loading model cost 0.889 seconds.
Prefix dict has been built successfully.
DEBUG:jieba:Prefix dict has been built successfully.
['炎裕', '時報', '電台', '股日', '反彈', '熱絡類', '股維持', '多頭', '輪動', '資金

```

▲套用jieba提取出關鍵字

▲最後即可得到關聯圖