

# Team 33



組員：

邱祥鴻：臺大財務金融所

黃柏森：東吳資料科學系

陳星儀：東吳資料科學系

洪鈺嫻：東吳企業管理學系

南山人壽題目一  
「文本生成--市場焦點機器人」

# Checkpoint1



# CONTENT 目錄

01 / 分工情況

02 / 會議進度與反饋

03 / 階段性實驗成果

04 / 待解決的問題

05 / 困難與阻礙

# 現行分工情況

邱祥鴻

API串接（資料取得）

黃柏森

文本生成格式、網頁介面設計

陳星儀

NLP研究、網頁架構設計

洪鈺嫻

簡易詞頻分析、網頁架構設計


PART

01

# 會議進度與反饋

PART

02



## 會議記錄 | Meeting minutes

Icons: Home, Checkmark, Question mark, Globe, Notepad, Megaphone

### 會議一覽

會議主題	日期	項目	類別	參與者	地點
第一次實體會議 <span>OPEN</span>	2021/04/01 3:00 PM-6:00 PM	flask	執行 學習 策劃	柏森 黃 (祥) 祥鴻 邱 星儀 陳 Cindy Hung Zuro Chang	臺大教研館
第二次實體會議	2021/04/08 3:00 PM-6:00 PM	flask python web crawler	討論 策劃 學習	柏森 黃 (祥) 祥鴻 邱 星儀 陳 Cindy Hung	臺大教研館
第三次實體會議	2021/04/15 4:00 PM-5:45 PM	python streamlit	學習 執行	(祥) 祥鴻 邱 (星儀) 星儀 陳 Cindy Hung	臺大教研館

## 第一次實體會議

參與者 柏森 黃 (祥) 祥鴻 邱 星儀 陳 Cindy Hung Zuro Chang

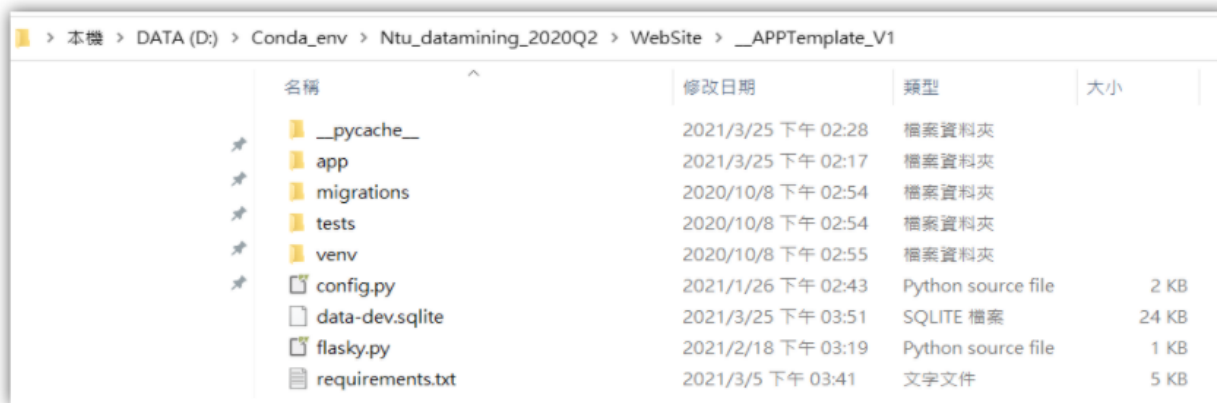
地點 臺大教研館

日期 2021/04/01 3:00 PM-6:00 PM

項目 flask

## 設定開發環境-1

- 在C碟建立 2020Q2-T1 資料夾，做為虛擬環境的目錄
- 建立虛擬環境與設定專案目錄
  - Conda create -p C:/2020Q2-T1 python=3.7
- 啟用虛擬環境
  - Activate C:/2020Q2-T1
- 在虛擬環境中，建立子目錄 WebSite
  - Mkdir WebSite
- 放入下載專案檔案
  - URL: <https://reurl.cc/OXe7bR> (目錄: \_\_APPTemplate\_V1)



名稱	修改日期	類型	大小
__pycache__	2021/3/25 下午 02:28	檔案資料夾	
app	2021/3/25 下午 02:17	檔案資料夾	
migrations	2020/10/8 下午 02:54	檔案資料夾	
tests	2020/10/8 下午 02:54	檔案資料夾	
venv	2020/10/8 下午 02:55	檔案資料夾	
config.py	2021/1/26 下午 02:43	Python source file	2 KB
data-dev.sqlite	2021/3/25 下午 03:51	SQLITE 檔案	24 KB
flasky.py	2021/2/18 下午 03:19	Python source file	1 KB
requirements.txt	2021/3/5 下午 03:41	文字文件	5 KB

1 專案流程介紹

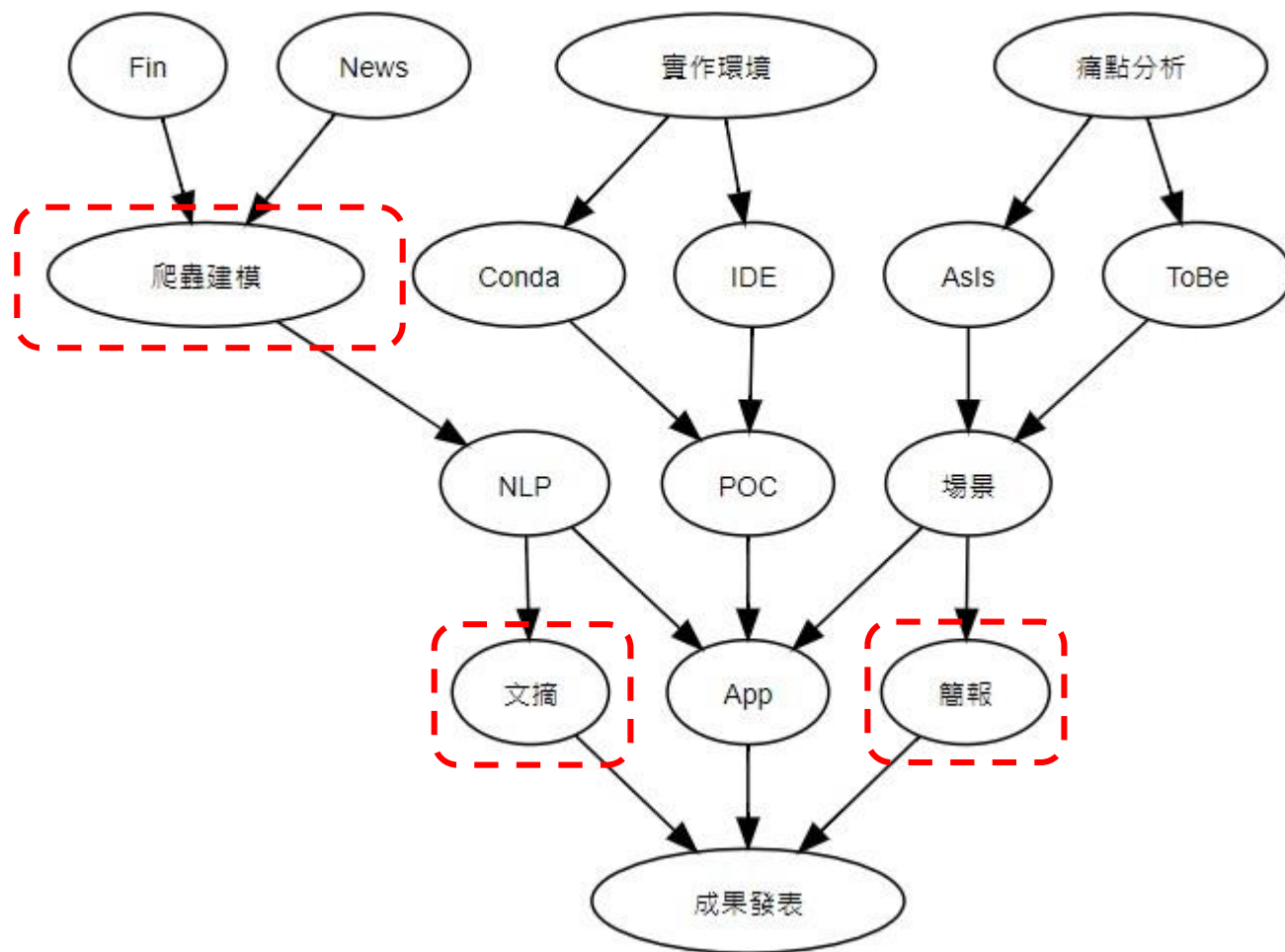
2 介紹Flask、MVC框架

3 建置開發環境

4 說明專案架構

5 確立聯絡工具

2021 金融科技專題 - 生成文本摘要 (南山人壽 - 台灣大學、東吳大學):



## 1 專案流程介紹

### 專案目標概要：

1. 爬取新聞資料
2. 自動生成摘要
3. 自動輸出成簡報



## 第二次實體會議

參與者 柏森 黃 (祥) 祥鴻 邱 (星儀) 陳 Cindy Hung

地點 臺大教研館

日期 2021/04/08 3:00 PM-6:00 PM

項目 flask python web crawler

類別 討論 策劃 學習

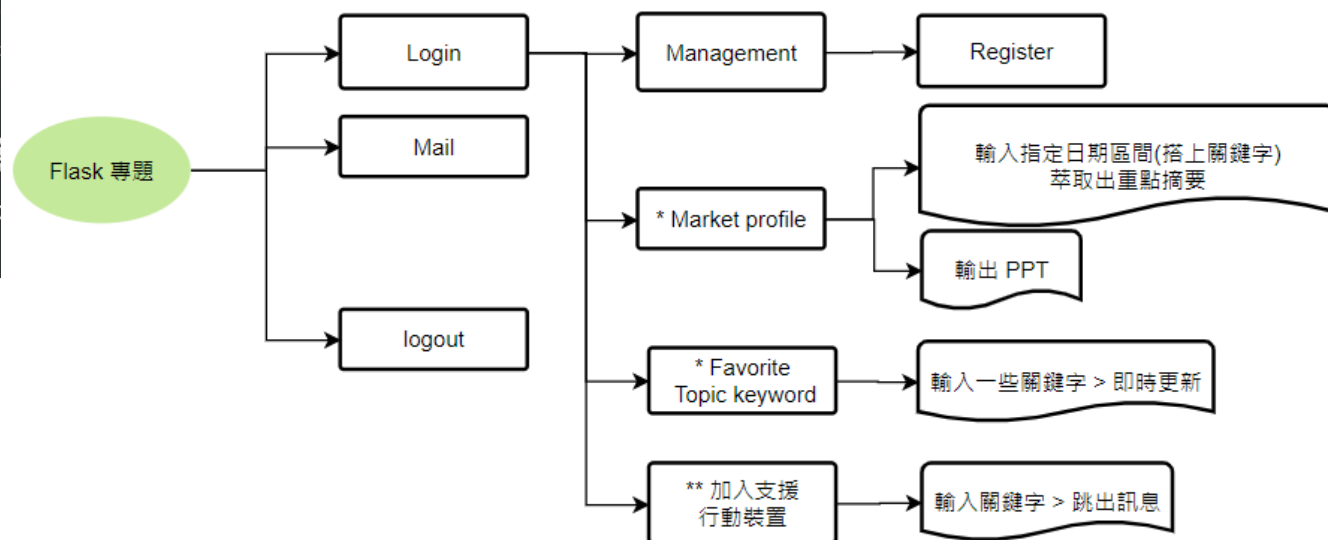
備註 大致分工完成

+ Add a property

Add a c

會議

1. 設定 ID  
• 專



1 專案目標及要求

2 設定 IDE 開發工具 ( VSCode )

3 說明Flask專案目標及流程

4 News API資料蒐集初步實做

5 分工：輸入、  
英文文本分析、  
輸出 ( 自動生成簡報 )

### News API 資料收集

- 首先知道該網站的請求方式為 `get`，因此參數可以直接加在網址後面也能透過 `JSON` 的形式再透過 `requests` 的 `get` 方法中的 `params` 參數做傳遞。

```
parameters_esg = {  
    'q': 'ESG',           # 查詢關鍵詞  
    'pageSize': 100,      # 查找頁數  
    'apiKey': api_key,    # 輸入個人api key (必須輸入否則會報錯)  
    'language': 'en',     # 查詢語言  
    'from' : '2021-04-01' # 查詢時間段  
}
```

\* 更多參數的用法可以參考 News API 的官方網站介紹

- 得到 `response` 後，將 `JSON` 放入 `pandas` 的 `dataframe` 中做處理，先篩選出 `articles` 的欄位，因為是 `pandas` 的 `series` 格式 (基本上就是表格整行column的資料)，再來是選取一個 `raw` 的資料所以再透過中括號和數字做選取。

```
df['articles'][0]  
# {'source': {'id': 'reuters', 'name': 'Reuters'},  
#  'author': 'Reuters Staff',  
#  'title': 'Citi adds ESG scores to data platform for climate-conscious investors - Re
```

## 4 News API資料蒐集初步實做



## 第三次實體會議

參與者 祥鴻 邱 星儀 陳 Cindy Hung

地點 臺大教研館

日期 2021/04/15 4:00 PM-5:45 PM

項目 python streamlit

類別 學習 執行

備註 分工架構可能需要調整

+ Add

Add

會議

1. 重新  
環境並

## 建立必要環境-(Windows版)

\* 因重新定義專案需求，需使用Python 3.8.5執行，所以必須清除上上周Python3.7 的環境

## ● 請依序輸入以下指令：

1. 進入 C:\2020Q2-T1 資料夾內，刪除所有內容
2. 使用 Anaconda Prompt 指令，建立虛擬環境  
conda create -p C:\2020Q2-T1 python=3.8.5
3. 確認安裝環境：[Y]、安裝相關套件：[Y]
4. 啟動虛擬環境  
conda activate c:\2020Q2-T1
5. 透過檔案總管進入 C:\2020Q2-T1，新增 WebSite 資料夾
6. 將共用雲端硬碟中，找NTU\_SCU-T1，Jack5 內的專案  
全數下載後放入WebSite內
7. 回到 Anaconda Prompt 指令，安裝全部專案的必要套件  
pip install -r C:\2020Q2-T1\WebSite\all\_projects\_requirements.txt

```
Anaconda Prompt (Anaconda3) - conda create -p c:\2020Q2-T1 python=3.8.5
(base) C:\Users\style>conda create -p c:\2020Q2-T1 python=3.8.5
WARNING: A directory already exists at the target location 'c:\2020Q2-T1'
but it is not a conda environment.
Continue creating environment (y/[n])? y
```

1 針對呈現結果重新定義

2 重新建置開發環境

3 討論專案未來擴展方向  
EX：使用情境、多重功能等



針對使用者情境建立一個具有流暢性的故事。



針對自動生成摘要，要有以下前提：

1. 由於摘要是同篇，故輸入的新聞需有核心主題。
2. 摘要的標題可反映整個新聞的合集。
3. 文字雲需要有集中的樣貌。



需尋找合適套用的NLP模型，進行務實導向的POC概念驗證及整合。



可設計文本建置後的複合應用功能，以利實用性加分。

例如：利用新聞集合之股票標的進行後續簡易的股價走勢分析、投資組合等。

# 階段性實驗成果

## PART 03

executed in 31ms, finished 04:31:05 2021-04-17

---

\_\_\_\_\_



# 待解決的問題

輸出成果的  
網頁設計架構

後續使用情境下  
可擴展的功能討論  
(結合相關投資策略、  
股價趨勢等)

要再以更細的功能項  
分工並需思考及聚焦  
主題與使用情境

NLP相關應用發展  
及套件使用討論

## PART 04

# 困難與阻礙

## PART 05

# 當前的困難（人、事、物）

---

## 模型不熟稔

---

由於未曾接觸過NLP的處理，故需要花費較多的時間研讀，並對其有初步認知後，才能針對後續的實做有些許的貢獻。

## 重新定義專案呈現

---

一開始對於本專案的定義與第三周重新定義後稍嫌不同，故實做層面部分尚需些微調整，加上臨近期中考週，以致可能會推遲原先訂定的專案進度。



## 組員臨時退出

---

有位組員於第二週開會當天才告知無法參與，導致工作需要重新分配，每個組員的負擔亦相對加重。

## 使用情境未定義

---

由於尚未針對何種情境下使用，故尚且無法針對專案的網頁功能有其他的發展，已排入下次的討論動議內，期盡早訂定，以利後續專案的開發。



Team  
33



**謝謝聆聽**