

國立空中大學 113 學年上學期

Python 程式設計與實務應用

《開發文件說明書：出遊小幫手》



第二組 派森萌新小組

成員：蔡文蕙、李佳怡、張嘉蘭

劉秋姣、林筱琪、曹韡

# 目錄

一、簡介與分工 .....	2
(一) 簡介 .....	2
(二) 小組分工 .....	2
三、開發環境 .....	3
四、系統開發流程 .....	3
(一) 任務分配 .....	4
(二) AWS 部署與自動化 .....	4
(三) 儀錶板創建 .....	4
五、程式碼主要技巧及重點解說 .....	5
(一) Python 寫作主要技巧 .....	5
(二) 原始碼重點解說 .....	6
六、Amazon Web Services(AWS) .....	7
(一) 雲端虛擬環境建置 .....	7
(二) 雲端自動爬蟲設定 .....	7
七、「出遊小幫手」儀表板製作 .....	8
(一) 縣市別搜尋 .....	8
(二) 單點搜尋： .....	9

# 一、簡介與分工

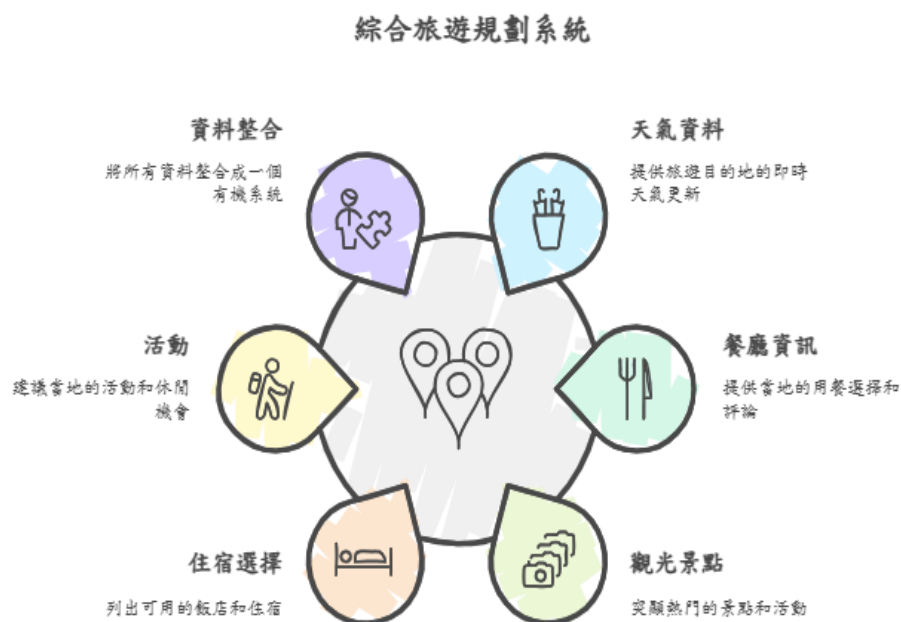
## (一) 簡介

本專題名為「出遊小幫手~今天去哪裡玩?」，旨在透過 Python 爬蟲技術，從政府公開資訊中蒐集有關台灣境內各縣市的觀光景點、餐廳、活動、天氣及民宿等資訊，協助使用者規劃出遊行程。使用者可透過一個整合網站，輕鬆搜尋想去的地點並獲取相關資訊。

這個系統的設計思路是將不同類型的旅遊資訊整合到一個平台上，讓使用者能夠方便地查詢和比較各種選擇。透過自動化的爬蟲技術，定期更新數據，確保提供給使用者的資訊是最新的。

## (二) 小組分工

- 林筱琪：負責景點爬蟲程式 ( viewpoint.py ) 。
- 蔡文蕙：負責活動爬蟲程式 ( activities.py ) 。
- 張嘉蘭：負責餐飲爬蟲程式 ( restaurants.py ) 。
- 劉秋姣：負責旅館爬蟲程式 ( hotels.py ) 。
- 曹 韓：負責天氣預報爬蟲 ( weather.py ) 。
- 李佳怡：主程式整合、AWS 環境架設、Google Looker Studio 儀表板設計。



### 三、開發環境

- (1) 作業系統：Amazon Linux 2023
- (2) Python 版本：3.11.5
- (3) 虛擬環境：myenv
- (4) 依賴套件：列於 requirements.txt
- (5) 雲端環境：AWS EC2
- (6) Google Sheets API：用於存儲與更新爬蟲結果（憑證檔案：google-sheet-api-token.json）
- (7) 專案目錄結構：

```
project/
├── main.py          # 主程式
├── credentials.json # Google Sheets API 憑證
├── taiwan_travel/   # 子資料夾，包含小程式
│   ├── __init__.py # 空檔案，將資料夾設為模組
│   ├── viewpoint.py # 台灣景點函式
│   ├── activities.py # 台灣活動函式
│   ├── hotels.py     # 台灣旅館函式
│   ├── restaurants.py # 台灣餐廳函式
│   └── weather.py    # 台灣天氣預報函式
```

### 四、系統開發流程

系統開發流程如圖 2



圖 2、系統開發流程時間軸順序

### (一) 任務分配

根據小組內部討論後，各成員根據分工開發旅遊相對應的 Python 爬蟲模組。

- 景點爬蟲：[https://media.taiwan.net.tw/XMLReleaseALL\\_public/Scenic\\_Spot\\_C\\_f.csv](https://media.taiwan.net.tw/XMLReleaseALL_public/Scenic_Spot_C_f.csv)
- 活動爬蟲：[https://media.taiwan.net.tw/XMLReleaseALL\\_public/Activity\\_C\\_f.csv](https://media.taiwan.net.tw/XMLReleaseALL_public/Activity_C_f.csv)
- 餐廳爬蟲：[https://media.taiwan.net.tw/XMLReleaseALL\\_public/Restaurant\\_C\\_f.csv](https://media.taiwan.net.tw/XMLReleaseALL_public/Restaurant_C_f.csv)
- 旅館爬蟲：[https://media.taiwan.net.tw/XMLReleaseALL\\_public/Hotel\\_C\\_f.csv](https://media.taiwan.net.tw/XMLReleaseALL_public/Hotel_C_f.csv)
- 天氣爬蟲：<https://opendata.cwa.gov.tw/api/v1/rest/datastore/>

### (二) AWS 部署與自動化

- 整合測試：主程式 main.py 整合各爬蟲模組進行測試，檢查數據流和功能是否正常。
- 雲端環境建置：創建 Amazon Web Services(AWS)帳號，新增一台 EC2 虛擬機。
- 部署上傳：將專案程式碼打包上傳至雲端 AWS 下的 EC2 虛擬機。
- 自動排程：使用 crontab 每 3 小時執行一次爬蟲與資料更新。

### (三) 儀錶板創建

- 資料連結：主程式整合爬取的資料並上傳至 Google Sheets。
- 自動排程：將爬取的數據存入 Google Sheets 並利用 Looker Studio 進行可視化展示



圖 3、系統開發流程-各階段任務項目

## 五、程式碼主要技巧及重點解說

### (一) Python 寫作主要技巧

#### 1. 網頁爬取與解析

- 使用 Python 的 requests 套件進行網頁爬取，搭配其他工具高效提取 HTML 文檔中的所需信息。
- 爬蟲模組化：依據功能分工，將景點、活動、餐廳、旅館及天氣資料分開處理，提升代碼結構的清晰度與可維護性。

#### 2. 資料處理與清洗

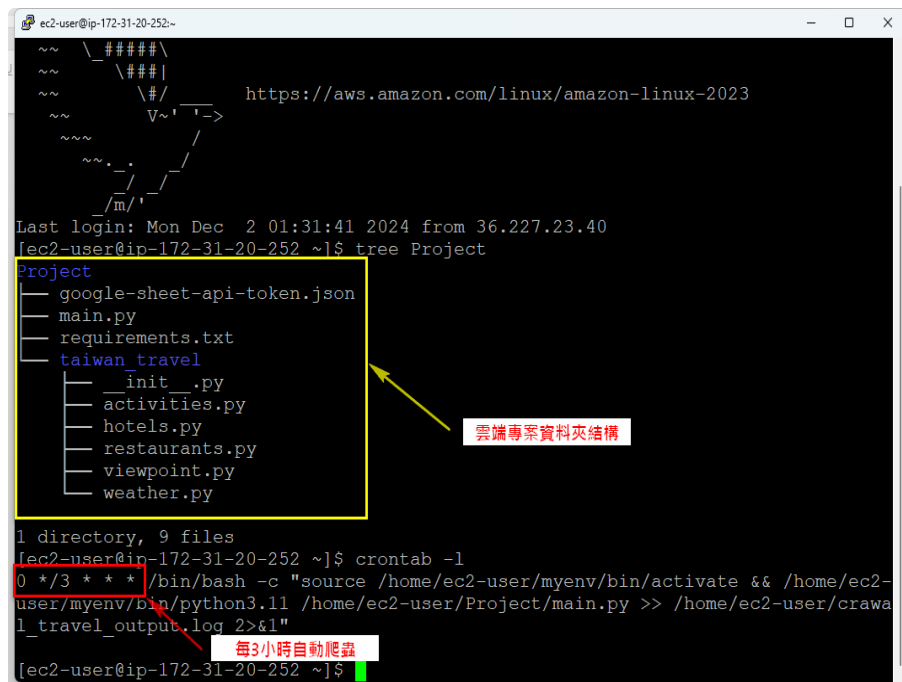
- 篩選所需欄位，處理空值、數值轉換及日期格式。
- 解析天氣描述，提取詳細參數（如降雨機率），以增強數據的實用性。

#### 3. 雲端數據存儲與共享

- 利用 Google Sheets API 將數據動態存儲至 Google Sheets，實現雲端共享與同步更新功能。

#### 4. 自動化執行

- 使用 Linux 的 crontab 指令設定定時任務，自動執行爬蟲，減少人力干預並提高工作效率。



The image shows a terminal window with the following content:

```
ec2-user@ip-172-31-20-252:~$ tree Project
Project
├── google-sheet-api-token.json
├── main.py
├── requirements.txt
├── taiwan_travel
│   ├── __init__.py
│   ├── activities.py
│   ├── hotels.py
│   ├── restaurants.py
│   ├── viewpoint.py
│   └── weather.py
└── 1 directory, 9 files
```

A yellow box highlights the `taiwan_travel` directory. A red box highlights the crontab command:

```
ec2-user@ip-172-31-20-252 ~]$ crontab -l
0 */3 * * * /bin/bash -c "source /home/ec2-user/myenv/bin/activate && /home/ec2-user/myenv/bin/python3.11 /home/ec2-user/Project/main.py >> /home/ec2-user/crawla1_travel_output.log 2>&1"
```

Annotations in the image include:

- A yellow arrow pointing to the `taiwan_travel` directory with the text "雲端專案資料夾結構".
- A red arrow pointing to the crontab command with the text "每3小時自動爬蟲".

圖 4、使用 crontab 進行定期爬蟲

## (二) 原始碼重點解說

### ◆ 各程式主要用途

#### 1. 主程式 (main.py) :

- 負責執行所有模組，包括資料抓取 ( 景點、活動、旅館、餐廳、天氣 ) 和將結果上傳至 Google Sheets 。

#### 2. Google Sheets API 憑證 (google-sheet-api-token.json) :

- 提供 API 使用的授權憑證，讓程式能夠連接和操作 Google Sheets 。

#### 3. 模組整合 (\_\_init\_\_.py) :

- 整合所有功能模組，包括 viewpoint、activities\_crawler、restaurants\_crawler、hotels\_crawler 和 weather\_crawler，方便主程式調用。

#### 4. 各爬蟲模組 :

- 景點 (viewpoint.py) : 抓取台灣景點的相關資訊，如名稱、電話、地址等。
- 活動 (activities.py) : 抓取台灣活動的資訊，並篩選尚未結束的活動。
- 餐廳 (restaurants.py) : 抓取餐廳的詳細資料，包括地址、營業時間、停車資訊等。
- 旅館 (hotels.py) : 抓取旅館資訊，如電話、地址、停車資訊等。
- 天氣 (weather.py) : 抓取未來 2 天的逐 3 小時天氣預報，並解析天氣描述提取詳細參數。

### ◆ 各程式間關聯性

#### 1. 資料流程 :

- 主程式透過模組呼叫各爬蟲函式，分別抓取不同類型的資料 ( 景點、活動、旅館、餐廳、天氣 ) 。
- 抓取完成後，統一格式化為列表格式，準備上傳至 Google Sheets 。

#### 2. Google Sheets 整合 :

- 所有抓取的資料最終由 upload\_to\_google\_sheet 函式上傳至 Google Sheets，實現資料的雲端共享與管理。

#### 3. 模組化設計 :

- 每個爬蟲模組專注於特定的資料來源和處理邏輯，確保程式結構清晰、易於維護。
- \_\_init\_\_.py 提供統一的模組匯入接口，讓主程式能簡單調用。

## 六、Amazon Web Services(AWS)

### (一) 雲端虛擬環境建置

- 在 AWS EC2 上啟動 Amazon Linux 2023 實例。
- 建立虛擬環境：

```
conda create -n myenv python=3.11.5  
source myenv/bin/activate
```

- 安裝 Python 3.11.5 及相關依賴：

```
pip install -r requirements.txt  
requests  
beautifulsoup4  
pandas  
google-auth  
google-auth-oauthlib  
google-auth-httpplib2  
gsread
```

### (二) 雲端自動爬蟲設定

- ◆ 確保所有爬蟲模組已正確編寫並存放於 taiwan\_travel 資料夾中。
- ◆ 設定 AWS EC2 上的 crontab 以每三小時執行一次主程式：

```
crontab -e
```

使用 Vim 編輯器開啟 crontab，在內容內添加以下行：

```
0 */3 * * * /bin/bash -c "source /home/ec2-user/myenv/bin/activate && /home/ec2-user/myenv/bin/python3 /home/ec2-user/project/main.py"
```

(PS: crontab 屬於全域變數，設定時須使用絕對路徑)



# 七、「出遊小幫手」儀表板製作

本專案最終成果是利用 Google Looker Studio 製作出的「出遊小幫手」儀表板，該儀表板包含兩個主要分頁，使用者可透過 Google Looker Studio 進入儀表板，選擇不同的搜尋方式

## (一) 縣市別搜尋

允許使用者根據不同縣市快速查找相關旅遊資訊，提供直觀且易於操作的界面。

- (1) 景點查詢：提供各縣市的觀光景點資訊，包括名稱、門票和位置等詳細內容。
- (2) 活動查詢：列出當地活動及其詳情，例如活動截止時間、地點。
- (3) 餐飲查詢：推薦當地餐廳、提供餐廳地址、營業時間以便使用者做出選擇。
- (4) 旅館查詢：提供旅館、民宿的地址、電話及官網連結網址，方便使用者訂房。
- (5) 天氣預報：顯示未來三天的天氣狀況，包括溫度、降雨機率等重要信息。



圖 5、出遊小幫手-縣市別搜尋

(二) 單點搜尋：

使用者可根據篩選條件設定距離範圍（例如 5 公里），查詢某特定景點附近的所有旅遊資訊。  
此功能允許使用者根據自己的需求靈活調整搜尋條件。



圖 6、出遊小幫手-單點搜尋