|  |
| --- |
| 班級：ZZZ001 小組名稱：衝吧!勇者無懼 指導老師：蕭翔之 |
| 專題報告-開發文件 |
| 爬蟲系統（包含天氣查詢、房屋實價查詢、臺灣地圖顯示氣溫查詢、各地新聞彙整） |

|  |
| --- |
| 小婷 鄭  2024/11/3 |

目錄

[**一、 小組成員 3**](#_Toc181550041)

[**二、 專題主題： 3**](#_Toc181550042)

[**三、 簡介： 4**](#_Toc181550043)

[**四、 開發工具與環境： 5**](#_Toc181550044)

[**開發工具與套件 5**](#_Toc181550045)

[**開發環境 6**](#_Toc181550046)

[**五、 系統開發流程： 7**](#_Toc181550047)

[**1. 需求分析：確定系統的主要功能和目標。 7**](#_Toc181550048)

[**2. 可行性確認： 7**](#_Toc181550049)

[**3. 系統設計： 8**](#_Toc181550050)

[**4. 開發環境建立： 8**](#_Toc181550051)

[**5. 功能開發： 9**](#_Toc181550052)

[**6. 程式設計主要技巧： 10**](#_Toc181550053)

# 小組成員

|  |  |  |
| --- | --- | --- |
| 學號 | 姓名 | 工作 |
| 112121931 | 林智鴻 | 架構規劃、技術支援、GitHub建置與管理、主程式、查詢指定地點實價登錄資訊地圖、查詢台灣各地天氣、查詢區域新聞、開發文件校訂 |
| 112120900 | 鄭怡婷 | 需求分析與功能確認、資料源收集、可行性評估、下載實價登錄資訊、查詢指定地點實價登錄資訊、列印功能、海報、開發文件撰寫 |
| 102120085 | 李佳慧 | 地點實價登錄資訊-泡泡圖 |
| 112122019 | 劉芳圻 |  |
| 112122858 | 黃涴淨 | 貸款計算、海報 |
| 112121260 | 李欣諭 | 開發文件：系統開發流程／程式設計主要技巧／原始碼重點解說 |
| 110128077 | 葉美麗 | 海報 |

# 專題主題：

本專題以台灣為核心，整合多種實用功能與資訊，包括房屋實價查詢、統計圖表、台灣各地即時氣溫顯示及地方即時新聞等。專題使用Python作為後端開發語言，建立網頁應用程式，並整合多個資訊來源。透過各種套件的運用，如Bootstrap、Folium、Pandas等，截取並呈現資料，並使用GitHub進行程式碼管理。此專題不僅讓小組成員深入了解Python，還能增進他們對應用程式開發的認識。

# 簡介：

1. **房屋實價登錄查詢：**動態下載台灣房屋實價登錄資料，讓使用者依房價區間篩選，並結合 Google Maps 查詢位置。系統會分析實價登錄資料，並以各式圖表的方式展示，提供使用者更直觀的數據視覺化。
2. **實價登錄資訊地圖：**以地圖方式呈現實價登錄資訊。
3. **查詢台灣各地天氣：**透過 API 與地圖系統整合，顯示各縣市即時天氣資訊。使用者可以輕鬆查詢當前天氣狀況，並獲得詳細的氣象數據。
4. **各地即時新聞：**系統會截取各地即時新聞，無廣告且提升瀏覽速度。使用者可以快速瀏覽新聞摘要，對有興趣的內容，仍可透過聯結檢視原新聞內容。
5. **貸款試算：**提供房屋貸款試算功能，使用者可以輸入貸款金額、利率及還款期限，系統會自動計算每月還款金額及總利息，幫助使用者更好地規劃財務。
6. **列印：**所有查詢結果能進行列印，或輸出成PDF格式儲存。

# 開發工具與環境：

## 開發工具與套件

|  |  |
| --- | --- |
| 工具與版本 | 說明 |
| Python 3.12.5 | 使用程式語言與版本  <https://www.python.org/downloads/> |
| Visual Studio Code 1.93.1 | 主要開發IDE工具，與版本簽入GitHub。  [Download Visual Studio Code - Mac, Linux, Windows](https://code.visualstudio.com/Download) |
| GitHub | 版本控制軟體，並於每次Commit時自動透過Pylint進行檢核。  [GitHub 存放路徑](https://github.com/112121931/PY/tree/main/Source/midtermwork) (未開放) |
| GitHub DeskTop | **GitHub Desktop** 是一個免費且開源的應用程式，旨在簡化使用 Git 進行版本控制的流程。  [Download GitHub Desktop | GitHub Desktop](https://desktop.github.com/download/) |
| Pylint 3.3.1 | 確保程式開發風格符合PEP8標準。  無使用其它套件。 |
| Flask 3.0.3 | Flask 是一個輕量級的 Python 網頁框架，適合快速開發網頁應用程式。 |
| Folium 0.17.0 | Folium 是一個用於地理數據可視化的 Python 套件，基於 Leaflet.js。 |
| Ipywidgets 8.1.5 | Ipywidgets 是 Jupyter Notebook 的互動小工具庫，用於創建交互式小部件。 |
| Ipython 8.27.0 | Ipython 是一個強大的互動式 Python Shell，提供了豐富的工具和功能。 |
| Pandas 2.2.2 | Pandas 是一個強大的數據處理和分析工具，提供了高效的數據結構和操作方法 |
| bs4 0.0.2 | bs4 (Beautiful Soup) 是一個用於解析 HTML 和 XML 文件的 Python 庫。 |

## 開發環境

作業系統：Windows 10 / 11

# 系統開發流程：

1. **需求分析：確定系統的主要功能和目標。**

經過小組討論，我們決定開發一個爬蟲系統，旨在收集並呈現多種資料。具體功能包括：

* 房屋實價查詢：動態下載並分析台灣房屋實價登錄資料，提供使用者依房價區間篩選和地圖查詢功能。
* 天氣查詢：整合即時天氣資料，顯示各縣市的天氣狀況，並提供詳細的氣象數據。
* 即時氣溫顯示：顯示台灣各地的即時氣溫，讓使用者能夠快速了解當前氣溫狀況。

1. **可行性確認：**

* **資料來源評估：**尋找並評估相關的資料來源，確保資料的可用性和準確性。主要資料來源包括：
* [內政部不動產成交案件實際資訊資料供應系統](https://plvr.land.moi.gov.tw/Index)：提供房屋實價登錄資料。
* [交通部中央氣象](https://www.cwa.gov.tw/V8/C/)：提供即時天氣資訊。
* [即時影像攝影機](https://tw.live/)：提供各地即時影像資料。
* [自由時報電子報](https://news.ltn.com.tw/)：提供即時新聞資料。
* [Google 地圖](https://www.google.com.tw/maps)：提供地理位置和地圖服務。
* **技術可行性驗證：**進行初步的程式撰寫和測試，驗證技術方案的可行性。確保各資料來源能夠順利整合到系統中，並能夠穩定地提供所需的資料。

1. **系統設計：**

* **框架選擇：**使用 Flask 框架開發網頁應用程式，確保系統的靈活性和可擴展性。
* **前端設計：**前端使用 HTML 和 Bootstrap 完成響應式頁面設計，主要負責用戶輸入和顯示查詢結果。確保頁面在各種設備上都能良好顯示。
* **後端開發：**後端使用 Python 進行開發，負責資料的查詢和計算。將結果以表格、圖片和地圖的方式組成頁面，並回傳給前端顯示。
* **開發架構：**遵循 MVC（Model-View-Controller）開發架構，以保持程式碼的簡潔和易於維護。這樣的架構有助於分離資料處理、業務邏輯和顯示層，提升系統的可維護性。
* **使用者體驗：**系統設計以簡單易用為前提，確保使用者能夠輕鬆上手並高效使用各項功能。

1. **開發環境建立：**

* **版本控制：**使用 GitHub 建立版本控制系統，以便進行小組內程式碼、文件等管理作業。透過 GitHub 整合 Pylint，確保程式開發風格符合標準，提升程式碼品質。
* **開發工具：**選擇 Visual Studio Code 作為主要開發工具，提供豐富的擴展功能和便捷的開發環境。
* **系統安裝文件：**建立詳細的系統安裝文件，供小組成員參考，確保每個成員都能順利設置開發環境。這些文件也將用於後續的系統維護和升級，提供清晰的指導。

1. **功能開發：**

* **分階段開發：**根據需求分析和系統設計，分階段開發各項功能模組，包括：
* **房屋實價登錄查詢：**動態下載並分析台灣房屋實價登錄資料，提供使用者依房價區間篩選和地圖查詢功能。
* **天氣查詢：**整合即時天氣資料，顯示各縣市的天氣狀況，並提供詳細的氣象數據。
* **即時新聞顯示：**截取各地即時新聞，無廣告且提升瀏覽速度，使用者可以快速瀏覽新聞摘要，並透過聯結檢視原新聞內容。
* **貸款試算：**提供房屋貸款試算功能，使用者可以輸入貸款金額、利率及還款期限，系統會自動計算每月還款金額及總利息，幫助使用者更好地規劃財務。
* **單元測試：**每個功能模組開發完成後，進行單元測試，確保功能正常運作。
* **程式碼品質檢查：**使用 Pylint 工具進行程式碼品質檢查，確保程式碼符合標準風格和最佳實踐。Pylint 會自動檢查程式碼中的錯誤、潛在問題和不符合規範的部分，幫助開發者維持高品質的程式碼。

1. **程式設計主要技巧：**

**❶ 爬蟲技術：使用 requests 庫來爬取資料，下載實價登錄檔案並解壓縮**

**處理。**

**❷ 資料篩選與轉換：使用 pandas 篩選資料，例如篩選符合特定價格範圍的房屋交易，並將地址轉換為可點擊的 Google Maps 連結。**

**❸ 地理位置處理：系統內建台灣各縣市的經緯度數據，結合房屋地址與天氣查詢結果生成地圖顯示。**

**❹ 網頁前端顯示：利用 Flask 搭建簡易網頁應用，顯示篩選後的房屋交易與天氣查詢結果。**

**❺ API 整合：透過 HTTP 請求調用外部天氣 API，並將數據格式化後顯示在網頁中。**

**6. 原始碼重點解說：**

**❶ 房屋實價查詢系統主要邏輯：**

**🞛 下載與解壓資料 🞛**

**⦿ 此函數用於下載最新的實價登錄資料，並將其解壓縮到指定的資料夾中。**

**def download\_and\_extract\_data():**

**response = requests.get(ZIP\_URL)**

**with open(ZIP\_FILE\_PATH, "wb") as zip\_file:**

**zip\_file.write(response.content)**

**with zipfile.ZipFile(ZIP\_FILE\_PATH, "r") as zip\_ref:**

**zip\_ref.extractall(DATA\_DIR)**

**🞛 篩選資料與生成 Google Maps 連結 🞛**

**⦿ 使用者可透過城市、最低與最高房價來篩選資料，並生成房屋的 Google Maps 連結。**

|  |
| --- |
| **def query\_real\_estate(city, min\_price, max\_price):**  **df = read\_city\_data(city\_files[city])**  **filtered\_df = df[(df['總價元'] >= min\_price \* 1000000) & (df['總價元'] <= max\_price \* 1000000)]**  **df['土地位置建物門牌'] = df['土地位置建物門牌'].apply(lambda x: f'<a href="{generate\_google\_maps\_link(x)}" target="\_blank">{x}</a>')**  **return filtered\_df[['鄉鎮市區', '土地位置建物門牌', '總價元', '單價元平方公尺']].to\_html(escape=False, render\_links=True)** |

**❷ 天氣查詢邏輯：**

**⦿ 使用 requests 庫查詢天氣 API，並將結果格式化為可在地圖上顯示的格式。**

**❸ 系統運行流程：**

**⦿ 程式啟動後，Flask 會開啟網頁伺服器，使用者可透過網頁輸入查詢條件，如城市與價格區間，系統會即時回應並顯示結果。**