

實務專題計畫成果摘要報告 租屋資料智慧整合搜尋系統

專題編號：112-CSIE-S017

執行期限：111 年第 1 學期至 112 年第 1 學期

指導教授：陳彥霖

專題參與人員：109590011 陳彥宇

109590026 黃亮維

109590035 林子皓

一、摘要

本專題以租屋族群為主要研究對象，目的是開發一個能讓使用者快速方便地瀏覽及搜尋所有租屋資料的網站。

相較於市面上大多網站，採用屋主或仲介登入房屋資料，導致有些照片與提供的資訊不符。本專題自行搜集以及標註資料，並採用 YOLOv7，訓練了一個家具辨識模型，判斷房屋網站中的照片所提供的和家具資訊是否與實際相同。

也透過爬蟲與人工智慧整合不同網站租屋資訊（591、好房網、Facebook、批踢踢實業坊 … 等）將所有資訊整理成相同的格式存入資料庫，並開發了一個網頁應用程式讓使用者能夠更輕鬆地檢視，並有效率找到理想的租屋處。

本專題也使用了 Docker 將資料處理的相關程式容器化，以便進行部屬與測試。

關鍵詞：網路爬蟲、人工智慧、容器化技術、網頁程式開發、雲端運算

二、緣由與目的

本專題旨在協助有在外租屋需求的族群了解租屋網站中實際情況，確認房仲或屋主提供的資訊是否正確，讓租屋者可以更有效的挑選房屋。

由於現在的租屋房源四散在網路各處，像是 591 租屋網、好房網、PTT 或 Facebook 上的多個租屋社團。租屋族在找房時，常常需要在各個網站中切換，尤其 Facebook 上的貼文無法針對文字內容設定篩選條件，所以需要一篇篇貼文查看，實在不方便，也浪費許多時間。因此我們

想要整合各大租屋網、Facebook 租屋社團與 PTT 的資料，並開發一個利用 AI 幫助使用者快速方便搜尋租屋處的網站。

三、研究報告內容

（一）研究方法

1. 研究目標

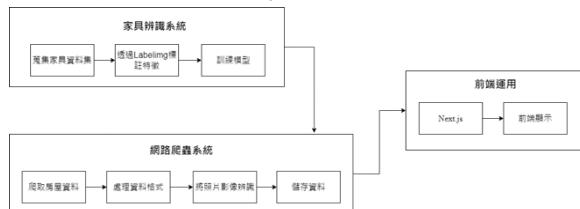
本專題的目標是完成一個租屋資料的整合及智慧分析系統，並開發一個方便使用者搜尋租屋資料的網站。藉由網路爬蟲程式將房屋網站中的資訊抓取下來，並透過基於物件偵測演算法訓練的家具辨識模型以及基於自然語言處理技術訓練的租屋貼文分析模型，對網站所提供的資訊（如：地址、在幾樓、雅房/套房、臥室/客廳/浴室數量、坪數、是否有電梯/網路等等）進行整理過後寫入資料庫。其中家具辨識模型用於判斷租屋網站的租屋圖片中家具和屋主自行提供的資訊是否相同，租屋貼文分析模型用於擷取來自 Facebook 與 PTT 租屋貼文中的房屋資訊。並透過將爬蟲以及處理資料相關的程式容器化，方便我們進行測試以及部屬。另外我們也將使用網頁應用開發的技術開發一個網頁，並部屬至雲端，讓使用者對房屋資料進行瀏覽及查詢。

2. 實驗階段

爬蟲程式會定期抓取資料，並將需要處理的資料向家具辨識系統發請求並交由該系統處理，爬蟲程式拿到處理完成的資料後會將資料寫入 MySQL 資料庫。

使用者透過在網頁應用程式，設定自己想要的租屋搜尋條件後，藉由前後端溝通，程式會查詢資料庫找出與使用者需求相同的房屋資料，並顯示在網頁上，讓使用者可以點開房屋資料的來源連結，查看房屋詳細資料。

(二) 研究步驟



圖一：開發流程圖

專案開發流程如圖一所示，包括家具辨識系統、網路爬蟲、與前端運用三個階段。

在開發家具辨識系統的步驟中，我們透過 LabelImg 自行標註約 20,000 張影像，基於 YOLOv7 進行模型訓練，並透過 Docker 封裝為容器，在啟動時自動抓取資料庫中的照片集進行辨識，並存入資料庫中。

在網路爬蟲系統的步驟中，我們使用了 Scrapy 這個爬蟲框架進行開發，並將處理完成的資料寫入 MySQL 資料庫。

在網頁應用程式開發的部份，我們使用了 TypeScript、React.js、Next.js、Prisma、tRPC、TailwindCSS … 等技術開發網頁端的使用者介面，並將網頁部屬在 Vercel 上，方便使用者能進行資料的查詢。

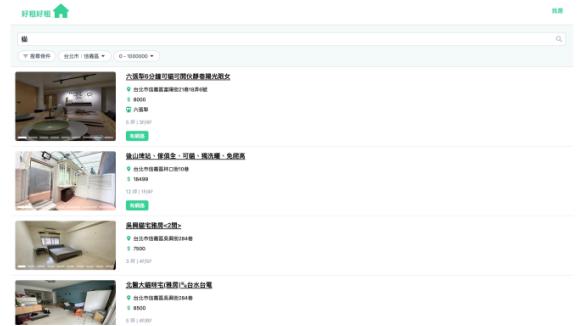
其中家具辨識系統和網路爬蟲系統及資料庫系統都有透過 Docker 容

器化以方便進行開發、測試以及部署。

四、實際成果



圖二：搜尋條件設定 Modal
使用者到首頁可設定搜尋條件 Modal。



圖三：關鍵字搜尋示意圖
使用者也可以透過關鍵字進行搜尋（如圖三所示）

五、未來方向

在目前租屋網站大多是選擇格式較為固定的（例如：特別標示租金、地址、租屋標題、樓層數等）的房屋資料，對於 PTT 以及 Facebook 上社團所提供的資訊，因格式不統一，一開始有打算透過 spaCy 訓練模型進行處理，但在最後因為效果不佳以及考量成本的狀況下沒有實作，但在大型語言模型技術崛起的現在，或許可以考慮使用 Langchain 並配合著線上開源的 LLM 進行實作。

參考文獻

- [1] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” arXiv.org, <https://arxiv.org/abs/2207.02696> (accessed May 18, 2023).