

基於大型語言模型之智慧瀏覽器擴充套件：整合影音內容摘要與即時語意翻譯系統

專題編號：114-2-CSIE-S024

執行期限：114 年第 1 學期至 114 年第 2 學期

指導教授：劉建宏

專題參與人員： 112590003 蔡宗佑
112590012 陳竑愷
112590056 陳泰葦

一、摘要

本專題設計並實作一套瀏覽器擴充功能系統，旨在提升使用者在跨語言資訊理解、影片內容掌握與語音轉文字處理上的效率。系統整合網頁翻譯、選取文字翻譯、PDF 翻譯、YouTube 影片摘要、字幕整理與本地 Whisper 語音轉文字等功能，讓使用者可直接在瀏覽器中完成多種語言輔助任務。系統後端串接 Gemini、ChatGPT、Groq 與 OpenRouter 等生成式 AI 服務，並結合本地 Whisper 伺服器進行影片音訊轉錄。經測試，系統具備字幕分流機制：處理無字幕影音最高可提升 10.38 倍資訊獲取效率(節省 90.4% 時間)；處理有字幕影音時因略過轉錄僅需 28 秒，效率提升達 128.5 倍(節省 99.2% 時間)，成功以量化指標證實解決手動處理耗時之痛點。

關鍵詞：Chrome 擴充功能、網頁翻譯、YouTube 摘要、語音轉文字、生成式 AI

二、緣由與目的

隨著外語網站、技術文件與影音教學內容大量增加，使用者在日常學習與資訊查找過程中，常面臨語言障礙、影片內容冗長以及重點不易快速掌握等問題。雖然現有翻譯工具與摘要工具已相當普遍，但多數功能彼此分散，需切換不同平台才能完成完整的理解流程，使用上不夠便利。本專題源自於實際使用情境中的需求，目標是將翻譯、摘要與語音轉文字等功能整合至單一瀏覽器擴充功能中，讓使用者在不離開目前瀏覽環境的情況下，即可快速

理解外文內容、整理影片重點並取得可閱讀的文字紀錄。透過 TranslateX，希望降低跨語言資訊處理門檻，提升學習、研究與日常瀏覽的效率。

三、研究範圍與目標

本專題聚焦於建立一套可直接嵌入瀏覽器使用的 AI 輔助內容處理系統，系統的核心特色與目標包括：

1. 即時翻譯與文件輔助：提供網頁與選取文字的即時翻譯功能，並支援 PDF 文件翻譯與閱讀輔助，降低使用者閱讀外文內容的門檻。
2. 影音摘要與時間軸整理：提供 YouTube 影片摘要與時間軸整理功能，協助使用者快速掌握影片重點，並支援多觀點內容比較以全面理解主題。
3. 本地化語音轉錄能力：透過本地 Whisper 伺服器實作語音轉文字功能，強化並提供穩定可控的影片內容轉錄能力。
4. 多模型整合與擴充性架構：整合多個生成式 AI 模型服務，並建立模組化且易於擴充的 Chrome 擴充功能架構，方便後續維護與功能新增。

四、使用技術方法

本系統以 Chrome Manifest V3 擴充功能為基礎，採模組化方式設計前端與後端流程，並整合多種 AI 服務與本地轉錄機制，以完成內容翻譯與摘要處理。整體架構分為前端互動與渲染架構，以及後端處理與 AI 分析架構兩部分。

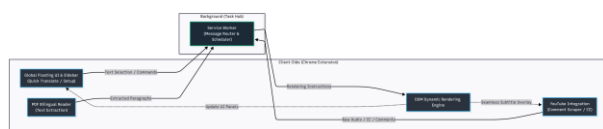


圖 1 前端互動與渲染架構

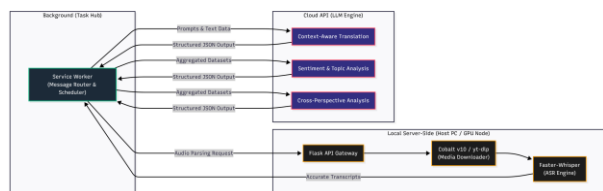


圖 2 後端處理與 AI 分析架構

(一) 前端互動與渲染架構：前端主要由 popup、content script 與 options 頁面組成，負責使用者操作、網頁內容注入與設定管理。

(二) 後端處理與 AI 分析架構：後端由 background service worker、本地 Whisper 伺服器與多個 AI 模型服務構成。系統透過 Gemini、ChatGPT、Groq 與 OpenRouter 等模型服務進行內容分析，並結合本地 Faster-Whisper 完成影片音訊轉錄。

(三) 多模型翻譯與摘要整合：根據不同使用情境提供翻譯、摘要與內容整理功能，維持較好的彈性與可用性。

(四) YouTube 影片摘要與字幕處理：系統可擷取 YouTube 影片相關資訊與字幕資料，進一步整理為摘要、重點段落與時間軸資訊。

(五) 本地 Whisper 語音轉文字：利用 Faster-Whisper 與 ffmpeg 進行影片音訊轉錄，提升影片轉錄的可控性。

五、實驗結果

本研究針對 3~60 分鐘不同長度之 YouTube 影片進行模組效能測試，並以效率提升倍數與時間節省率作為評估指標。在「無內建字幕情境」下，系統需執行完整下載、轉錄與 AI 摘要流程，各階段耗時數據如表 1 所示：

影片長度	下載時間	本地轉錄時間	AI 摘要時間	系統總耗時	效率提升倍數	時間節省率
180s	7.8s	16.9s	28.0s	52.7s	3.42 倍	70.7%
300s	10.5s	23.9s	28.0s	62.2s	4.82 倍	79.3%
600s	13.9s	47.3s	28.0s	89.2s	6.73 倍	85.1%
1020s	17.8s	99.7s	28.0s	145.5s	7.01 倍	85.7%
1500s	24.1s	130.5s	28.0s	182.6s	8.21 倍	87.8%
1800s	23.0s	128.6s	28.0s	179.6s	10.02 倍	90.0%
3600s	40.6s	278.3s	28.0s	346.9s	10.38 倍	90.4%

表 1：各模組執行效能數據表

其中，本地轉錄模組為主要時間瓶頸，然而面對 60 分鐘長片，系統仍能提升 10.38 倍效率（省下 90.4% 時間）。

此外，若為「有內建字幕情境」，系統將自動跳過下載與語音轉錄，直接送交 AI 處理。在此情境下，無論影片長短，總耗時皆約等於 AI 回饋時間（約 28 秒）。以處理 60 分鐘影片為例，效率倍數升至 128.5 倍（省下 99.2% 時間）。

從客觀量化數據證實，TranslateX 確實解決了手動耗時之痛點，大幅提升資訊獲取效率。

六、結論

本專題成功實作整合式瀏覽器輔助系統 TranslateX，將翻譯、摘要與語音轉文字功能整合於單一介面中，提升跨語言資訊處理的效率與便利性。系統以模組化方式設計，具備良好的擴充性。針對未來的後續發展，計畫將系統架構朝向「雲端原生 (Cloud-Native)」與 SaaS 服務化轉型。

預計將目前高度依賴終端算力的本地 Whisper 語音轉錄模組進行容器化並部署至雲端 GPU 叢集。在產品發布上，計畫將系統正式上架至瀏覽器擴充功能商店。透過前端上架商店與後端雲端化的結合，將能徹底消除使用者的硬體設備門檻限制，實現真正的「隨插即用」，將 TranslateX 從單機版的輔助套件，升級為具備高度可擴展性的商業級雲端服務。