

2025 年全國大專校院智慧創新暨跨域整合創作競賽 系統需求書

1. 系統名稱

基於開源軟體部屬的即時手語生成(Real-time Multimodal Sign Language Generation Leveraging Open-source Frameworks)

2. 系統目的與範圍

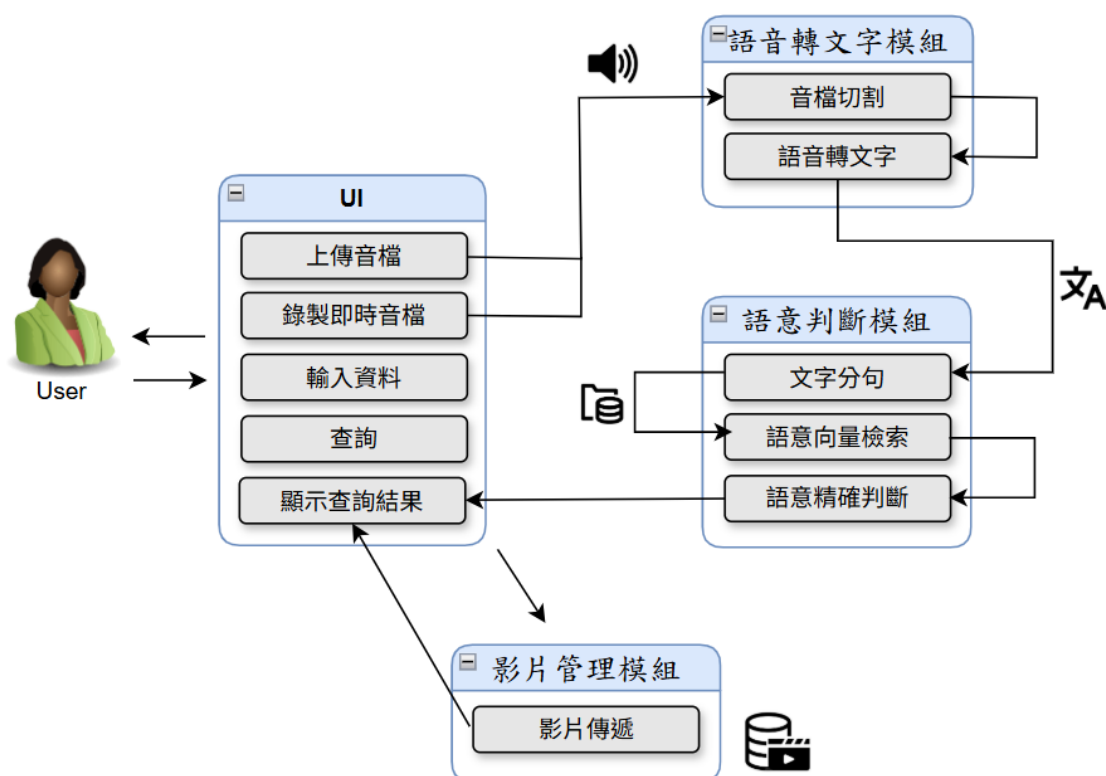
(1) 系統目的

- 解決溝通障礙：建立一套語音轉台灣手語的服務，協助聽障者理解口語語言，保障聽障族群的權利。
- 公共應用與社會福利：目前資料尚少，設計易資料擴增，為擴展至教育、醫療與新聞等需要手語服務的公共場域，建立基礎架構。

(2) 系統範圍

- 前端資料接收：接收使用者輸入的查詢資料，包含文字、音檔、錄音。
- 語音輸入與處理：將輸入音檔後轉換為文字。
- 語意比對：用向量比較與精確判斷語意篩選出資料庫中符合查詢資料的條目，並回傳解果。
- 結果展示：即時顯示手語動畫生成結果供使用者檢視。
- 資料貢獻：使用者貢獻更多手語資料的途徑。

3. 系統架構



目前系統可大致分為四個部分，分別是前端顯示與互動模組(UI)、語音轉文字模組、語意判斷模組、影片管理模組如上圖。前端適用 HTML 撰寫的網頁；後端則是剩下的三個模組，採用了 Open AI 的開源轉體 Whisper、gte-qwen2-1.5B 的句子向量化模型與 Deepmind 開發的 Gemma 模型(暫定)做語意精確判斷，打包部署至 Kserve 端原生平台部署系統。

根據個模組的需求，可在下方依重要程度分成**紅色**的最重要需求、**橘色**的次重要需求、**綠色**的普通重要需求：

- 前端顯示與互動模組(UI)
 - UI 設計
 - 文字輸入處理
 - 音檔輸入處理
 - 即時錄音處理
 - 查詢結果呈現
- 語音轉文字模組
 - 音檔切割
 - 語音轉文字
- 語意判斷模組
 - 文字分句
 - 語意向量檢索
 - 語意精確判斷
- 影片管理模組
 - 虛擬手語影片管理

4. 系統非功能需求

非功能需求編號	非功能性需求描述
RSLG-NF-001 實時性	語音輸入至手語動畫輸出應在 10 秒內完成
RSLG-NF-002 系統負載處理能力	每個 pod 最多可以同時處理 4 個推論請求 系統應針對流量自動擴展及釋放閒置資源
RSLG-NF-003 系統穩定性	連續運行至少 30 分鐘不中斷
RSLG-NF-004 部屬可攜性	支援本地端與 k8s 容器部屬
RSLG-NF-005 開源模組整合	核心模組皆由開源框架組成，使用者可在無授權費用下部屬和修改

5. 系統功能需求

功能需求編號	功能需求描述
前端顯示與互動模組	
RSLG-FR-001 輸入接收	接收使用者的輸入，包含 文字、音檔、錄音
RSLG-FR-002 查詢結果展示	從後端取得向量語料庫與精確判斷後的結果，並依結果決定展示手語影片或「沒有相似資料」
語音轉文字模組	
RSLG-FR-003 語音轉文字	將使用者輸入的音檔或即時錄音轉換為文字，並傳送給語意判斷模組
RSLG-FR-004 檔案切割	當音訊過長時做適當的切割，避免轉換檔案過大造成判斷輸出失真
語意判斷模組	
RSLG-FR-005 分句處理	將語音轉文字模組傳遞的資料(純文字)進行需要的分句處理
RSLG-FR-006 語意向量檢索	將經處理過後的句子與以 gte-qwen2-1.5B 製作的向量語料庫進行向量檢索，從所有資料中擷取出最相近的 5 筆資料
RSLG-FR-007 語意精確比對	將向量檢索出的 5 筆資料與查詢資料交給 chat LLM 進行精確語義比對，比對成功輸出手語路徑給前端，反之告知無資料
影片管理模組	
RSLG-FR-008 影片管理	儲存所有虛擬人物的手語動畫，當前端送出請求時，依路徑傳遞需要的影片

6. 一般性的系統功能操作使用案例(Use case, UC)之劇本(Scenario)描述如下。

使用案例編號：RSLG-UC001	使用案例名稱：文字輸入查詢
系統反應動作	使用者操作動作
	輸入欲查詢的語言資料(詞、句、段落、文章、對話等)
	按「送出查詢」鍵
系統執行分句處理	
用迴圈將分出的句子依序作向量檢索，輸出 5 筆資料	
比對 5 筆資料與查詢句，作出精確判斷，決定提供影片路徑或查詢失敗	

	得到查詢結果 各句本身與對應的影片 或 「沒有相似句子」
--	---------------------------------

使用案例編號：RSLG-UC002	使用案例名稱：音檔輸入查詢
系統反應動作	使用者操作動作
	上傳欲查詢的音檔資料
	按「送出查詢」鍵
判斷為音檔傳遞給 Whisper API	
Whisper 將音檔轉為文字傳輸給語意判斷模組	
系統執行分句處理	
用迴圈將分出的句子依序作向量檢索，輸出 5 筆資料	
比對 5 筆資料與查詢句，作出精確判斷，決定提供影片路徑或查詢失敗	
	得到查詢結果 各句本身與對應的影片 或 「沒有相似句子」

使用案例編號：RSLG-UC003	使用案例名稱：即時錄音查詢
系統反應動作	使用者操作動作
	按下開始錄音鍵
系統詢問權限問題	
	同意系統取得權限
取得權限開始紀錄錄音	
	網頁標籤紅點出現開始講話，結束後按下停止錄音鍵(網頁標籤紅點消失)
結束錄音	
	按「送出查詢」鍵
判斷為錄音將系統紀錄的錄音檔案傳遞給 Whisper API	
Whisper 將音檔轉為文字傳輸給語意判斷模組	
系統執行分句處理	
用迴圈將分出的句子依序作向量檢索	

，輸出 5 筆資料	
比對 5 筆資料與查詢句，作出精確判斷，決定提供影片路徑或查詢失敗	
	得到查詢結果 各句本身與對應的影片 或 「沒有相似句子」

7. 追溯矩陣

	RSLG-UC001	RSLG-UC002	RSLG-UC003
RSLG-FR-001	✓	✓	✓
RSLG-FR-002	✓	✓	✓
RSLG-FR-003		✓	✓
RSLG-FR-004		✓	✓
RSLG-FR-005	✓	✓	✓
RSLG-FR-006	✓	✓	✓
RSLG-FR-007	✓	✓	✓
RSLG-FR-008	✓	✓	✓