



C 組 ·

Common Voice 族語聖經錄音

Irvin Chen

台灣維基媒體協會 專案執行

 HackMD

Common Voice 族語 聖經錄音

@CCNDA 2025/3/16

hackmd.io/@moztw/common-voice-ccnda 

tags: Common Voice, Mozilla

Irvin Chen

台灣維基媒體協會 專案執行

Mozilla Firefox 瀏覽器（火狐）志工、MozTW 台灣
社群聯絡人、網頁工程師



Common Voice 族語錄音計劃

- <https://moztw.org/common-voice>

Mozilla Common Voice



Common Voice 願景與目標

希望促成的願景

- 讓電腦 / 手機的語音助理可以說各種語言
- 教 AI 用族語說話
- 讓手機可以翻譯族語與世界上的其他語言

目標

- 準備一組沒有版權問題的包含所有語言的免費語音資料，讓任何開發者能實現上述的願景

Common Voice 簡歷

- 2017 在台北「如何設計一個網站來『收集語音』workshop」；英語開放錄音
- 2018 國語（華語）
- 2022 台語（台灣閩南語）
- 2024/12 開放九種族語：排灣, 賽德克, 布農, 撒奇萊雅, 泰雅, 魯凱（含萬山, 多納, 茂林）

Common Voice 運作步驟

準備階段

- [專案志工] 收集錄音用的句子，並上傳到錄音後台：
 - 無版權限制的句子
 - 生活對話為主
 - 單句為主、單詞為輔

開放錄音階段

1. [錄音者] 自行上網站錄音
2. [錄音者] 自行在網站上聽現有的錄音，並進行驗證（確認錄音）
3. [Mozilla] 每三個月以 **CC0 授權**（公眾領域）在網站上免費公布語音資料庫
4. [研究者與商業公司] 自由下載語音資料，研究與開發相關科技。例如：
 - 語音辨識 speech-to-text
 - 文字轉語音/語音合成 text-to-speech, voice synthesis
 - 口譯, AI, 字典...任何用途

語音資料庫現況

@2025-3-14

<https://commonvoice.mozilla.org/languages>

- 華語 129 小時 / 2260 人錄音
- 台語 24 小時 / 271 人錄音
- 族語：各 11~15 小時，預定七月開放下載

華語（台灣）

🕒 小時

129

👤 使用者數

2260

台語

🕒 小時

24

👤 使用者數

271

族語語料庫

- 因為想開放族語錄音，需要找各語言的句子
 - 在去年12月獲得教育部國教署授權，將政大原民中心編的 1~9 年級族語課本的句子整理後上傳，各族約各 600 句
- 在一月底獲得台灣聖經公會授權，可取用聖經族語版 500 節以內的經文

族語聖經

<https://cb.fhl.net/>

- 舊約、新約：布農、賽德克、魯凱、泰雅
- 馬可福音：萬山、茂林、多納



聖經網站

上帝永存的話語

GOD'S WORD, LIGHT FOR THE WORLD.

[台灣聖經公會](#) 聖經網站

目標

從各族語聖經，整理出各一百句的句子，以便上傳
Common Voice 錄音網站

問題

- 如何選擇經文？
- 如何把經文整理成一句一句？
- 相同的方法可應用在國語、台語、客語聖經嗎？

目前進度：[github.com/irvin/
formosan_bible_corpus_processing](https://github.com/irvin/formosan_bible_corpus_processing)