

「2019 華語文教學應用軟體競賽」企劃書

甲、創作主題

一、創作緣起

國家教育研究院為促進對華語文語料庫的認識，於 108 年建置了各類華語文語料庫及標準體系整合應用系統，讓使用者不必在網路中大海撈針，即可快速地做相應的檢索。但是，語料庫的專業性及真實語料不適宜直接作為華語為第二語言學習者的使用。因此，本團隊以師生共學方式組成，從推廣華語文教學、開發相關應用軟體出發，為縮短學習者與華語文的距離，就「2019 華語文教學應用軟體競賽」擇定以「(七)其他華語文教學相關應用軟體等」做綜合開發。

二、創作範圍與目標

1.本團隊以【認識臺灣、學習華語文】為目標，開發「臺灣 PaPaGO」，提供旅遊資訊及學習華語的主題平台，讓「華語為第二語言學習者」來臺灣學習華語、在臺灣旅遊，除了遠端連網規劃臺灣旅遊，也可在旅途中使用。



2.以【臺灣在地城市小旅行】的方式接近學習者，學習者應用本系統查詢景點相關訊息：景點照片、景點地圖、景點英文官網等，運用對於景點的好奇激發華語學習的興趣，運用本系統做華語學習。

3.利用多元的華語文語料庫，藉由規劃設計「臺灣 PaPaGO」以【雙北區域】為開發先導，提供臺灣旅遊景點資訊平台進而探索華語。

乙、創意構想

一、企劃概念：從臺灣旅遊到華語

□關係連結：交通、人、景點與華語

□區域：雙北-新北市、臺北市

□路線：臺北捷運線(5 線)+臺灣好行(4 線)

選擇北部地區作為開發的先導，外國人士來台，抵達桃園機場、松山機場，經由交通的連結臺北市、新北市將會是旅遊的首站。而串聯雙北的主要交通工具為臺北捷運(MRT)及接駁特色景點的臺灣好行(BUS)，以兩大重要大眾交通工具路線，把人的關係連接起來，透過觀光：深入旅遊、了解在地文化、置入華語學習。



二、系統設計：旅遊華語，閱讀悅懂

以路線作延伸，提供一站式的景點資訊平台，除了基本的旅遊資訊還有提供「光華雜誌」的短文，旅途中閱讀及學習華語文。

□現階段完成簡易版：

- 1.首頁(地圖)—路線地圖(捷運線 5 線+臺灣好行)及景點。
- 2.景點—

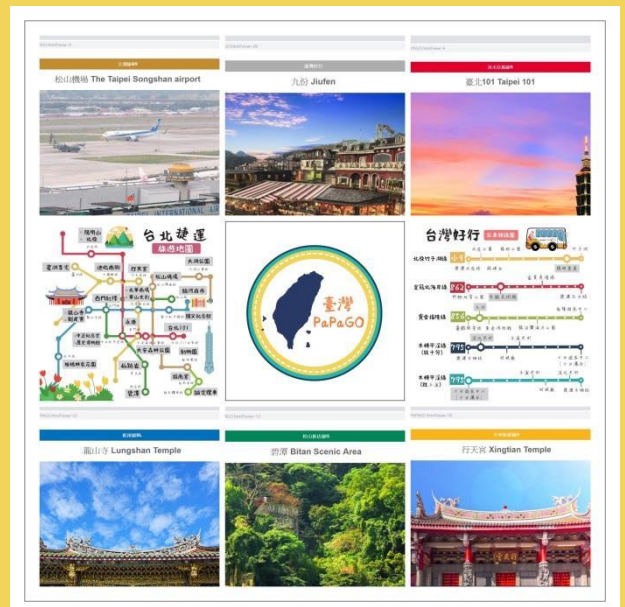
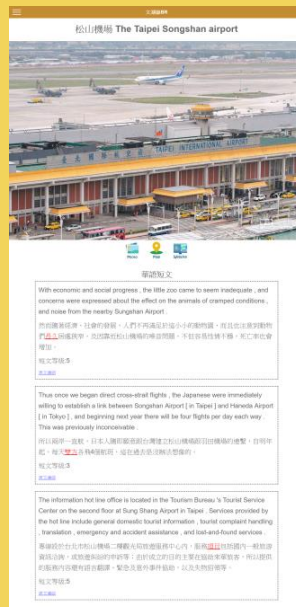
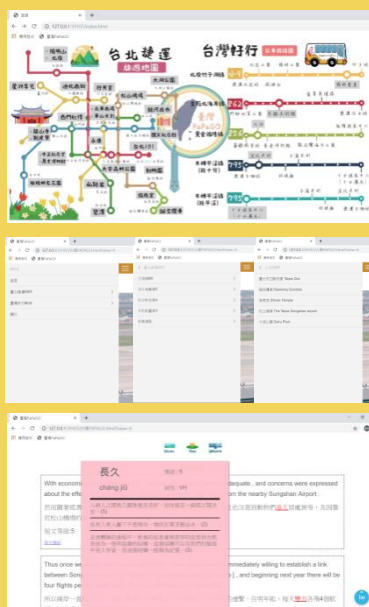
■旅遊資訊：從交通路線地圖查找景點，點選景點，提供 Google 照片、Google 地圖、英文官網連結，以查找景點照片、位置及路線規劃、及連結英文官網。

■華語文學習：提供景點相關詞彙的中英「短文」、認識「生詞」、學習「例句」。

- #短文：一個景點提供 1~3 篇短文，及短文等級。
- #生詞：一篇短文提供 3~5 個生詞，生詞為 \geq 短文等級，漢語拼音、詞性。
- #例句：一個生詞提供 2~3 個例句，生詞的例句需與生詞同詞性、且 \leq 生詞等級，做為華語文學習。

- 3.主選單—提供階層式展開選單，可以查找各路線下的景點。

□實際畫面：景點使用滿版的圖片，吸引目光！(景點照片取自於 Google)





三、未來發展：具 AI 的華語教材

目前「臺灣 PaPaGO」現階段只有北部地區的雙北。依照整體架構的未來可能性進行設計，可以開發其他區域，如北部、中部、南部、東部、離島地區等，用交通把臺灣 368 個鄉鎮景點串聯起來，豐富臺灣旅遊華語教材。

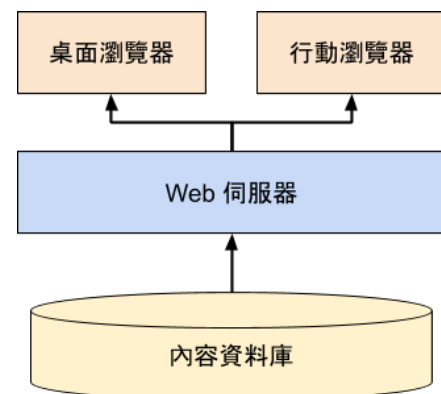
更期望未來可以使這套架構具有 AI 自動生成，甚直能深度學習產生所需主題的華語教材的功能。



丙、系統架構

「臺灣 PaPaGO」是一種 Web 系統，同時可在行動裝置網頁瀏覽使用，也可以無礙地在電腦桌面上執行。使用 MS-SQL 資料庫儲存相關材料，經由 Web 伺服器應用程式的處理，將旅遊資訊提供給使用者瀏覽，並且與之雙向互動，客戶端的執行環境為各平台的 App 與瀏覽器。

目前資料庫儲存試題材料來源為人工整理，以累積將語料庫轉為題庫材料的處理經驗，明白其差異程度，記錄注意事項，以作為日後開發自動化轉換之重要參考。



丁、系統應用

一、華語文標準體系

□光華雜誌：短文分級

□華語文學習：生詞、例句分級

以國教院「華語文標準體系」作為整體華語文學習的指標。「臺灣 PaPaGO」為查詢與閱讀的系統開發，使用者等級建議為 3 級以上，能在公共場域中閱讀簡短段落並理解基本訊息；而系統內容也以此體系做為分級鑑別。

二、語料庫應用系統

□「國教院華英雙語索引典系統」之光華雜誌：作為景點短文的内容來源。

□「國教院例句編輯輔助系統」：查看各等級詞彙所佔的比例，藉以判斷短文的難易度，再從短文中選擇該等級或等級以上生詞作為學習。及在生詞之例句產生後，



判斷例句的難易度。

□「國教院分詞系統」：查詢在短文中生詞的詞性。

□「國教院語料庫索引典系統(含國教院華語中介語索引典系統)」之 COCT 書面語料庫 2017：以生詞做查詢，找到與生詞相同詞性的例句。

三、語料及公開授權資料

□《教育部國語辭典簡編本》：「臺灣 PaPaGO」設定為 3 級以上，具有閱讀學習的基本能力，因此提供生詞漢語拼音，作為華語文學習。



綜上系統的應用，在製作「臺灣 PaPaGO」，遇到許多華語文語料庫的錯誤：

□語料庫的錯誤：牽一髮而動全身，源頭的錯誤，將會使不正確資料導出。

1.「國教院華英雙語索引典系統」之光華雜誌內容：中英翻譯錯誤、中文錯字、中英文對照不出來等，造成後續資料無法正確的對應。如：文湖線 BR-景點：木柵動物園的第一段短文，中文有「因此...規劃」，但是英文句子裡卻沒有表示這段意思；且另有中文日期「十月卅一日」，然英文卻是「October 30th」這根本性錯誤將會造成華語文學習的誤解，及未來自動產生的連環錯誤。

2.「國教院語料庫索引典系統」：有亂碼或文字錯誤，也造成查詢的困難。



611	5b334ab9027f3cbd21faca0	ceng2 M iuN7 樣種 zhong3 M iuN7 樣樣 yang4 M jin5-chhu3 人次 人次 ren2ci4 M jin5-chhu3 人次 人次 ren2ci4 M kai2 改回 hui2 M
612	5b334ab9027f3cbd21faca0	M iuN7 樣種 zhong3 M iuN7 樣樣 yang4 M jin5-chhu3 人次 人次 ren2ci4 M jin5-chhu3 人次 人次 ren2ci4 M kai2 改回 hui2 M khoan2
613	5b334ab9027f3cbd21faca0	zhong3 M iuN7 樣樣 yang4 M jin5-chhu3 人次 人次 ren2ci4 M jin5-chhu3 人次 人次 ren2ci4 M kai2 改回 hui2 M khoan2 款樣 yang4 M khou
614	5b334ab9027f3cbd21faca0	M iuN7 樣樣 yang4 M jin5-chhu3 人次 人次 ren2ci4 M jin5-chhu3 人次 人次 ren2ci4 M kai2 改回 hui2 M khoan2 款樣 yang4 M khou 雜
615	5b334ab9027f3cbd21faca0	gu3 M kun5 群群 qun2 M liap8 粒頭 tou2 M lin5 人次 人次 ren2ci4 M mai2 羅回 hui2 M mia5 名名 ming2 M

□內容年代、不合時宜

隨著時代變遷，「國教院華英雙語索引典系統」之光華雜誌、「國教院語料庫索引典系統」有些內容對於現在觀念、知識的更新，而不合時宜。

若是可以依照華語文取材時間、來源、類型作標籤上的區分，加上篩選的功能作為快速擇取出需要的內容或排除不適當的華語文語料。

□詞性與等級的結合

「國教院例句編輯輔助系統」可以查詢斷詞及等級，而「國教院分詞系統」可以查詢斷詞及其詞性，這兩者具有共同的斷詞功能，建議可以結合此兩大系統。

□語料庫應用系統與語料及公開授權資料並沒有勾稽

同上，我們建置系統的生詞，斷詞系統是很重要的依據。但以此在《教育部國語辭典簡編本》查詢生詞漢語拼音時，發現部分查找時無法在辭典中找到該詞彙，也就表示斷詞出的詞彙在辭典中不被收錄。

如：生詞「商機」，在辭典中並沒有直接被查找到。

戊、軟體清單

軟體名稱	用途	授權
Brackets	程式編輯環境	免費軟體，MIT 授權
HTML	程式語言	HTML5 網路標準
CSS	程式語言	HTML5 網路標準
JavaScript	程式語言	HTML5 網路標準
IIS 7.5	Web 伺服器	Microsoft 合法授權軟體
Windows 2008 R2	伺服器作業系統	Microsoft 合法授權軟體
網頁瀏覽器	使用者界面	免費軟體

附錄

本師生共學團隊成員及分工如下：

召集人	陳淑惠
總籌、系統功能架構規劃	徐珮瑜
原始資料製作	徐珮瑜、陳淑惠、許銘芳、熊禮揚 黃潔金、莊雯樺、林婉霓
系統程式製作	許銘芳、黃潔金
美工設計	林婉霓



影片製作

莊雯樺

影片協助(畫面拍攝、聲音錄製、演員等)

徐珮瑜、林婉霓、黃潔金、熊禮揚