

國立臺北大學華語語料庫應用軟體設計共學社群

陳淑惠、許銘芳、熊禮揚、黃潔金、莊雯樺、林婉霓 徐珮瑜、

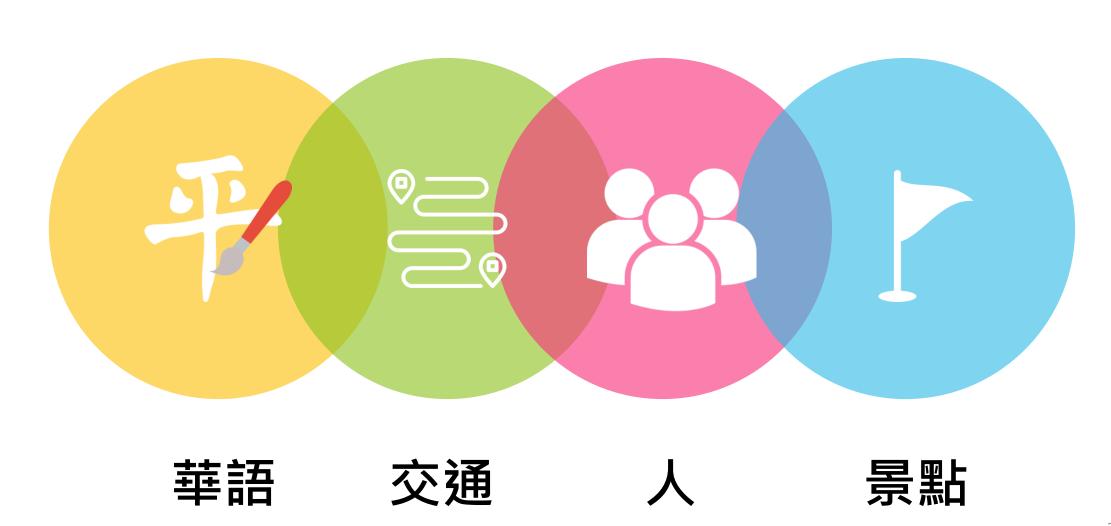


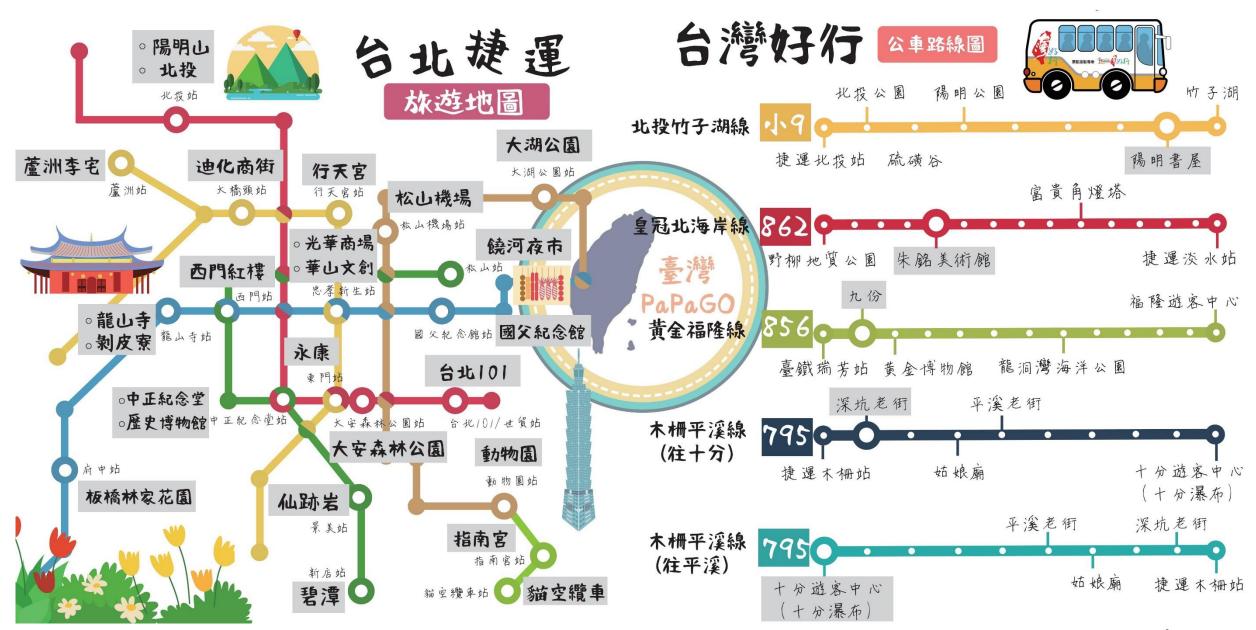
目錄



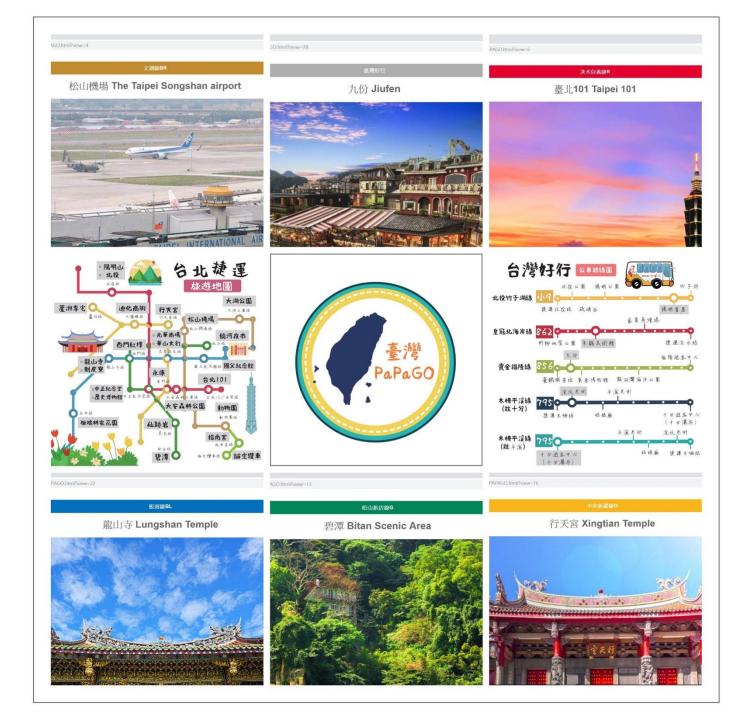
- 一.創作理念
- 二. 系統設計
- 三.未來發展

■一、創作理念-從臺灣旅遊到華語





https://reurl.cc/W4O7nD



■二、系統設計:旅遊華語,閱讀悅懂



1.首頁-捷運線5線+臺灣好行及景點。

2.景點一

- 旅遊資訊:Google照片、Google地 圖、英文官網連結
- 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:一篇短文目前提供1個生詞,漢語拼音、詞性。
- ✓ 例句:一個生詞提供2~3個例句,
- **3.主選單一**提供階層式展開選單,可以查找各路線下的景點。





景點

旅遊

華語文













1

【短文】:景點相關詞彙的短文。

- 〔國教院華英雙語索引典系統〕: 光華雜誌
- ・提供台灣光華雜誌原文網址・查看原文

2

【難易度】: 短文、生詞、例句。

• 〔國教院例句編輯輔助系統〕 自動斷詞、勾選:國教院詞表(初稿)

3

【詞性】: 生詞、同一種詞性之例句。

• 〔國教院分詞系統〕

4

【漢語拼音】:生詞。

· 《國語辭典簡編本》

5

【例句(產生)】:以完整度高、可直接使用,不修改句型優先,以期達到未來「自動產生」的目的。

• [國教院語料庫索引典系統] COCT 書面語語料庫2017 詞系統 】可以結合成一個系統。**建議:【例句編輯輔助系統】和一



【短文】:景點相關詞彙的短文。

- (國教院華英雙語索引典系統):光華雜誌
- 提供台灣光華雜誌原文網址,查看原文



國教院華英雙語索引典系統

語料庫 光華雜誌 ▼ 關鍵詞: 九份 搜尋

總共找到8句

- 1. The fate of the temple was a microcosm of the fate of **Chiufen**; when the gold ran out, both fell into serious decline.
 - 福山宮和九份金礦命運相繫。廿年前九份金盡人散,廟裡的香客便日益稀少。
- 2. The people of Taiwan should have deeper concern than others, yet because the people of Hong Kong and **Kowloon** are closer to mainland China and have intentionally maintained a distance from Taiwan, the two groups have never had a strong **sense** of affinity; nevertheless, they feel a bond of close interdependency.
 - 和香港一水相隔,應當有著比他人更深一層的關懷的台灣,素來對於因 靠近大陸而刻意與台灣有所距離的港<mark>九</mark>,雖略覺得難以交心,卻總有著 一份「唇齒相依」的情感,對香港的前途也覺得憂心忡忡。
- 3. At the end of September Palau 's foreign minister wrote a letter to Fredrick Chien , the ROC 's minister of foreign affairs , stating that the claims made in the ad were " false , " and that the Palau government denied them and refused all responsibility for them . 帛琉 外交部長 烏貝勞 在 九 個 月 底 致函 外交部長 錢復 , 表示 這 份 廣告 內容 「 虛偽 不實 」 , 帛琉 政府 必須 全 盤 否認 , 並 拒絕 承擔 任何 責任 。
- 4. The **journal**, called Jiu Wan Shiba **Guai**, published its first issue around the time that the Taipei-Ilan Expressway opened and now has about 2,000 subscribers. 終於,大約在北宜高速公路通車的同時,這份名為「九彎十八拐」的刊物出刊了,如今已有兩千多個訂戶。

自動對應翻譯

- chiufen (2) report (1)
- journal (1)
- menu--the (1)
- sense (1)

自動抽取搭配詞

九份 [N]

[N~] 金瓜石~,

[~N]

~金瓜石,~一帶,~山城,

2

【難易度】:短文、生詞、例句。

- (國教院例句編輯輔助系統)
- 自動斷詞、勾選;國教院詞表(初稿)



國教院例句編輯輔助系統

例句編輯工具

例句編輯工具

☑ 自動斷詞 ◎ 國教院詞表(初稿) ◎ 華語八千詞

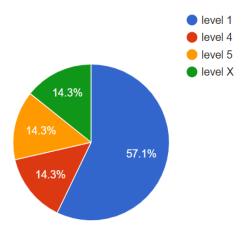
九份是台灣最富傳奇色彩的地方,因為它曾經有黃金。

輸出詞彙分級訊息

分級詞表:國教院詞彙分級表(初稿)

九份 是 台灣 最 富 傳奇 色彩 的 地方 , 因為 它 曾經 有 黃金 。 X 1 1 1 X 5 4 1 1 1 4 1 5

詞彙等級分佈



3

【詞性】: 生詞、同一種詞性之例句

(國教院分詞系統)



國教院分詞系統

請輸入需要分詞的文本

九份是台灣最富傳奇色彩的地方,因為它曾經有金。

開始分詞 清空

九份(Nd) 是(SHI) 台灣(Nc) 最(Dfa) 富(VJ) 傳奇(Na) 色彩(Na) 的(DE) 地方(Na), (PUNC) 因為(Cbb) 它(Nh) 曾經(D) 有(V_2) 金(Na)。(PUNC)

【漢語拼音】: 生詞

《國語辭典簡編本》



:::我想找	
--------	--

- ○字詞 ○音讀 ○釋義 ○全文
- ○義類關鍵詞

字詞查詢符號代表意義及檢索範例图

搜尋

注音小鍵盤 漢語拼音小鍵盤

5 为		里		Y	历	5	IL
女 士	« U	1	アー	乙	7	4	
口马	万く	P	ちメ	さ	4	大	
口为	FT		4 4	せ	ヌ	L	
			空	白	鍵		

圖片分類請點選下方圖片分類進入檢索

動天童概古工交建學植書 日人綜綜音武運 器具通築術物法 體合合樂器動 俗飾 比 物 文 玩 念 用 圖物 形 氣 品部 主 工 容 象 具 題 位 詞 圖 圖

5

【例句(產生)】: 以完整度高、可直接使用,不修改句型優先,以期達到「自動產生」的目的

臺灣 PaPaGO

• (國教院語料庫索引典系統)coct書面語語料庫2017

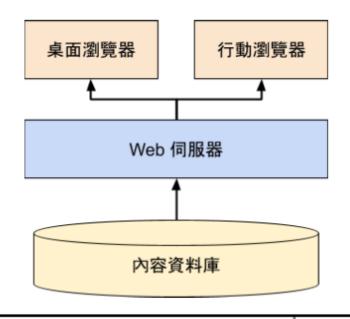
Your query "九份" returned 647 matches in 144 different texts (in 287,568,639 words [158,247 texts]; frequency: 2.25 inst per million words) [2.875 seconds]

<	<<	>>	>	Show Page: 1	Line View	Show in random order	New query	▼ Go!
---	----	----	---	--------------	-----------	----------------------	-----------	-------

No	Filename	Solution 1 to 50 Page 1 / 13
1	576a560a326d1b3fb77fd380	簡介 吳念真 本名 吳文欽 。 一九五二年 生 於 台灣 北部 著名 金礦區 <mark>九份</mark> 。 一九七六年 考入 輔仁 大學 夜間部 , 主修 會計
2	576a560a326d1b3fb77fd384	責任。 民國 三十多 年 的 瑞芳 , 正 因為 金瓜石 和 <mark>九份</mark> 的 淘金 採礦 事業 發達 之 故 , 而 自然 形成 了
3	576a560a326d1b3fb77fd458	十二 公里 處 , 有 一 塊 路 牌 寫 著 「 <mark>九份</mark> 」 , 看來 從 此 處 往 右轉 就 是 九份 了
4	576a560a326d1b3fb77fd458	「九份」,看來從此處往右轉就是 <mark>九份</mark> 了。二○○四年我曾和母親來台灣觀光
5	576a560a326d1b3fb77fd458	和母親來台灣觀光,腦中不禁浮現當時拜訪 九份的景色,令人非常懷念。到達水湳洞時,
6	576a560a326d1b3fb77fd8e1	是同一個人——義秉恆。義秉恆原為北縣 <mark>九份</mark> 地區茶藝館「八番坑」的股東之一。他說
7	576a560a326d1b3fb77fdb5c	,多半 也 會 宣稱 自己 的 芋頭 來自 甲仙 。 比如 <mark>九份</mark> 芋圓 , 即為 一 鮮明 之 例 。 久而久之 , 大甲芋
8	576a560d326d1b3fb7800808	知 多少 次 了 , 說 是 老街 新生 , 其實 與 <mark>九份</mark> 一般 完全 變質 。 剛 畢業 時 , 淡水 真 是 好
9	576a560d326d1b3fb7800b5a	不斷 切換 , 馬路 換成 山 景 又 換成 海景 , 離 <mark>九份</mark> 越來越 近 , 「 前面 的 小路 左轉 。 」 瑪立哥 指
10	576a560d326d1b3fb7800cff	還是流動的。15. 通往快樂的祕徑最近,在 <mark>九份</mark> 過了一個消逝的和風之夜。當我



軟體清單資料形成



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 13

■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。
- 未來可以開發其他區域,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。
- 更期望未來可使這套架構具有 AI自動生成,甚至能深度學習 產生所需主題的華語教材的功 能。



謝謝大家