

資工系實務專題研究計畫成果報告

Tourist Blog Information Integration System

專題編號：098-CSIE-S023

執行期限：97 年 2 學期至 98 年 1 學期

指導教授：王正豪

專題計劃參與人員：李韋翰、劉耕銘、范閔捷

一、 中文摘要

本專題設計了一個介面提供使用者可以查詢有關於旅遊資訊當中的「食、衣、住、行」等日常生活相關的資訊，讓使用者可以更容易搜尋到旅遊景點相關的資訊。在搜尋方面，我們使用「Google AJAX Search API」來查詢各網友在 Blog 中所發表的文章並針對回傳的資訊做分類。中文斷詞方面，使用了中研院研發的「中央研究院資訊科學所詞庫小組中文斷詞線上服務」來對中文文章做處理，然後再針對回傳的資訊做分類^{註(1)}。在地圖方面，我們使用了「Google Maps API」來標示位置。以及關於各個景點，可以看到 Wikipedia、Youtube 影片、Panoramio 照片^{註(2)}來讓使用者更加清楚該地方的特色與文化。

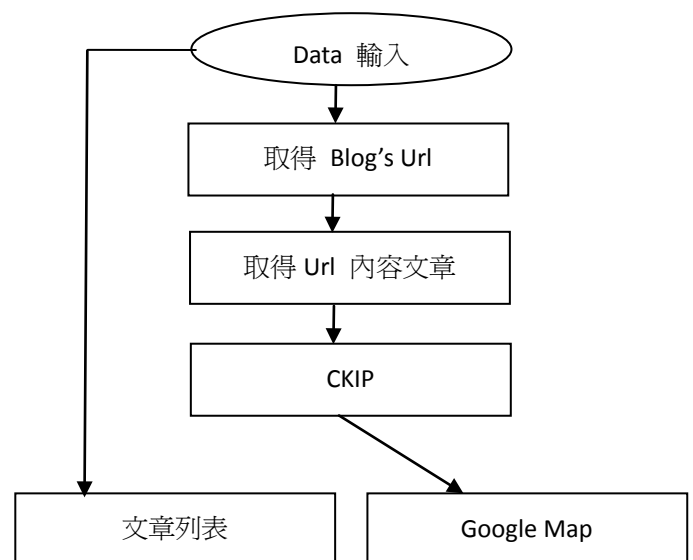
關鍵詞：

Google AJAX Search、Google Maps、PHP、CKIP、Named Entity Recognition ^{註(3)}。

二、 緣由與目的

網路世界發展至今，資訊成長速度從爆炸式演變成倍增式的增加，各種資訊充斥於網際網路之中，從純文字的 BBS 站，到各大論壇的出現，及近期所出現的個人 Blog。每個人都可以擁有屬於自己的 Blog。可以用來分享屬於自己的經驗、遊記、美食心得，也就是關於各類「食、衣、住、行」的文章。在眾多雜亂無章的文章中，搜尋並掌握適當的資訊，已成為現今科學研究的基石。

三、 系統簡介與架構



(流程圖)

本系統分為四大架構：

(1) Data 輸入：

資料來源：來自無名小站 Blog 之文章。

資料格式：中文。

由使用者輸入關鍵字，藉由 Google AJAX Search API 搜尋，將結果限定在無名小站並輸出文章列表在畫面上。

(2) 取得內容

收到 Google AJAX Search API 傳回的結果，傳回搜尋結果後擷取 Blog 的網址，再藉由取得網址之內容後，進行斷詞。

(3) CKIP

中研院中文斷詞線上服務不支援簡體中文。欲斷詞的資料在送出前會去除符號、英文並且限制字數，以增加斷詞服務的處理效率。

將處理過的文字送到中研院中文斷詞線上服務做斷詞服務，因中研院伺服器需要處理的資料量龐大，所以資料回傳的時間在一分鐘內不等的時間，此系統花費最多的時間即為此處。

(4) 網頁輸出

搜尋後的結果是以清單列表方式以及地點標記在 Google Map 上來呈現。除了有地標 marker、Panoramio(照片)、Youtube(影片)以及 Wikipedia。也會分別以「食、衣、住、行」等四種標籤來分類。

四、 相關技術與工具

程式語言： HTML, PHP5, Javascript

Web Server： Apache + MySQL

斷詞分析： 中研院線上中文斷詞服務

地圖顯示： Google Maps API

網誌搜尋： Google AJAX Search API

五、 研究成果



(圖 1)搜尋出來的結果



(圖 2)於 Google Maps 上的標示

六、 結論

本系統使用 Google AJAX Search API 提供使用者能快速的搜尋他所想查詢文章與資訊，並結合 Google Maps API 讓使用者知道他所查詢的地點，更連結到 Wikipedia、Panoramio、Youtube，可以讓使用者更加了解其所查詢之地點及其資料。計畫未來還可以記錄每個使用者的習慣，個人化的設定，進而對使用者進行搜尋喜愛來幫使用者設計出專屬的搜尋器。

七、 參考資料

- 1.PHP Example and Source Code
<http://phpbb-tw.net/phpbb/>
- 2.CKIP Client Example
<http://ckipclient.sourceforge.net/>
- 3.中文斷詞系統
<http://ckipsvr.iis.sinica.edu.tw/>
- 4.Google AJAX Search API Developer's Guide
<http://code.google.com/intl/zh-TW/apis/ajaxsearch/documentation/>
- 5.Google Maps API Developer's Guide
<http://code.google.com/intl/zh-TW/apis/maps/documentation/index.html>
- 6.PHP Manual
<http://www.php.net/manual/en/>

八、 註解

註(1)：根據回傳網址連結之內容，取回網站內容進行斷詞功能，再藉由標籤進行分類，傳送至網頁表單。

註(2)：從 Google API 取得參考資料，並且新增在 Google Maps API 圖層中。圖層本身與其所有元件被視為是 Google Maps API 的一個疊加層，可透過顯示或隱藏圖層的方法，決定地圖中所得到的資訊。

註(3)：本專案應用在中文斷詞系統服務中，主要是從大量文字資料中自動抽取特定訊息，對明確的資料進行分類、判斷上下文以及語義化的分析。