

美食探索 Big Data

專題編號：106-CSIE-S022

執行期限：105年第1學期至106年第1學期

指導教授：王正豪

專題參與人員：103820010 商資穎

103820019 王怡萱

一、摘要

本專題為 — 美食探索系統。我們將收集網路上現有的部落客撰寫的美食文章。分析不同作者對各個店家的評論，彙整出每家店家的綜合評價和熱門度。之後我們會將上述的分析結果及店家資訊呈現在 App，並且利用 Google 提供的 API 將餐廳位置標記在 Google 地圖上面。有了這樣的資訊平台，使用者可以很方便的透過店家類型來對應查詢其感興趣的熱門店家之參考評價。

關鍵詞：

網路爬蟲 (web crawler)、關鍵字比對 (keyword matching)、情緒分析 (Sentiment Analysis)、情緒傾向程度 (Sentiment Orientation)

二、緣由與目的

每當與朋友、家人聚餐時，上網搜尋「中式餐廳」，卻是百萬筆以上的資料，各式餐廳推薦文良莠不齊，店家資訊、價位參考、特色料理等…無法輕易的得到所需要的資訊及評價。以致於勞神費力的進行分析，卻事倍功半而錯過真正有特色的餐廳。

所以我們想藉由此次專題為美食愛好者製作出美食探索平台。這樣以後享受美食時不會再踩地雷，再也不會盲目的選擇哪家餐廳，方便又省時。

三、研究報告內容

(一)研究範圍

痞客邦美食部落格文章

(二)使用技術方法及工具說明

1. API：透過痞客邦 API 取得指定用戶底下所有文章。透過 Google Maps API 進行定位。

2. 使用 jsoup library 去下載網頁內容，透過解析、切割 html 裡面各種 tag 去獲得餐廳的資訊、地址及評論。

3. 斷詞系統 (word segmentation service)：以中文描述的文章架構與英文不同，字詞之間沒有空格區隔。在分析文意上，必須先將文章切成句，句切成具有意義的詞才能進行比對。我們利用中央科學研究院研發的 CKIP 中文斷詞系統以及 Jieba，鑑別出最小且具有意義的語言文詞單位，並擷取其傳達的訊息。

4. 字典法 (lexicon-based method)：利用 NTUSD 字典預先定義的情緒字詞，對評論內容進行搜尋比對，將識別出的字詞代入公式計算，以估計該美食專文對於餐廳的評價。

5. 機械訓練學習法 (machine learning approach)：由正負面文章訓練統計分析情緒字詞出現頻率，計算該詞彙之正負面情緒傾向程度 (Sentiment Orientation)，並加入自建立之情緒字典以增加後續分析文章之準確度，透過大量的資料分析訓練後便能提高辨別文章語意的能力。

6. 結果呈現方式：使用 Android

Studio 來開發 App，XAMPP 框架開發後端資料庫系統，將店家資訊地址以及評價分數存進資料庫，使用者再透過 App 存取資料庫查詢資料。

(三)研究架構流程



圖1. 架構流程圖

四、結果呈現

使用者可透過店家地圖搜索附近店家(如圖2)，並可以收藏自己喜歡的店家(如圖3)。

五、參考文獻

- [1] 游和正、黃挺豪、陳信希，「領域相關詞彙極性分析及文件情緒分類之研究」，中文計算語言學期刊，Vol. 17, No. 4, December 2012, pp. 33-48

[2] 李孟潔，利用機器學習作法之中文意見分析，碩士，國立清華大學，新竹，2009。

[3] 李啟菁，中文部落格文章之意見分析，碩士，國立臺北科技大學，臺北，2010。



圖2. 附近店家地圖



圖3. 使用者收藏列表