

財團法人演譯基金會

美兆健康資源中心



美兆健康數據資料庫
地理資訊編碼流程概要

MJHRF

技術報告編號：MJHRF-TR-05

2016/05/18

引用本文參考文獻格式：

莊淵傑(2016)。美兆健康數據資料庫地理資訊編碼流程概要。財團法人演譯基金會技術報告(編號：MJHRF-TR-05)。

美兆健康數據資料庫地理資訊編碼流程概要

一、前言

地理資訊系統（GIS）的發展，讓數據資料的研究跳脫平面的思考向度，躍入了空間維度的概念分析。為能讓美兆健康數據資料庫的應用，結合空間維度的研究發展，美兆健康資源中心進一步建置了地理編碼資料庫，期望透過地理資訊與健康數據資料的結合，發展出更多健康數據與空間維度的分析。

二、資料來源

美兆會員於每次健檢時會提供地址資訊，以利健檢報告等文件的寄送，因此資料庫中保留了會員歷次來健檢時的地址，這些資訊從1994年開始至2014年底總共累積了1,536,964筆資料，而其中包含了早期極少數的郵政信箱以及有缺漏的地址（0.32%）。

三、地理編碼工具

目前地理資訊系統運用最廣範的平台是Google Map，因此我們選擇利用Google Map API將會員地址轉換為經緯度，因此得到的經緯度座標系統，就是Google Map所使用的WGS84經緯度座標系統，可以直接或間接透過轉換，串連到各種地理資訊應用平台上。

四、地理編碼程序

由於google提供的地址轉換經緯度介面，是在特定網址輸入地址資訊，然後經過編碼回傳一整頁HTML格式的經緯度相關資料，所以不容易進行批次處理，因此我們開發了一個excel的VB模組，來達成批次處理的目標。過程中會出現無法順利取得經緯度的狀況，我們會嘗試刪除可能混淆地理編碼的內容，例如：樓層、幾號”之幾”、鄰、里等資訊，然後再重新進行地理編碼。

五、地理編碼結果與驗證

經過Google Map API成功完成編碼的地址有1,529,476筆，大約是佔總筆數的99.5%，為了驗證地址轉換的準確率，我們隨機抽取1,000筆轉換後的經緯度，進行地理反編碼（Reverse Geocoding）的程序，為了批次作業，我們採用網路上現有的資源來進行。

在獲得反編碼的地址後，我們將分三個層次與原始會員地址做比對，第一層是比對縣市層級是否一致？在1,000筆資料中，共有6筆資料無法成功反編碼，而且全部都是金門縣的地址，但如果實際用經緯度看Google Map，發現座標點還是很接近金門，但這部份在驗證時，仍然計算為不一致。最後1,000筆驗證資料中，有970筆的縣市是一致的，也就是在縣市層級的正确率是97%。

接下來是要比對地址中的鄉鎮市區層級是否一致，在驗證資料中，共有958筆地址的鄉鎮市區是一樣的，正确率是95.8%，不過再驗證過程中，會發現有些鄉鎮市區不一致，是因為該地址剛好位於兩個鄉鎮市區的交界處，轉換座標上一點的偏移，就造成了鄉鎮市區的改變，但反編碼所得到的位置，與原始地址是很接近的，因此實際上地理編碼的正确率，應該是會比95.8%更好。

最後是路名或街名的比對，在1,000筆地址中，有880筆的路名或街名是一致的，也就是路名或街名的正确率為88%，在這個層級，會發現有些地址剛好位於交叉路口（例如：忠孝東路與復興南路），原本地址是在忠孝東路上，但因為轉換的誤差，而讓地址變成復興南路，但實際的座標位置，是相去不遠的。

六、個人隱私保護

為了保護會員的個人隱私資訊，轉換後的經緯度並不會直接提供原始數據做為研究分析使用，而是會將經緯度數據進行模糊化，處理過後的經緯度無法精確的對應到實際的地址，但會在一定的誤差範圍內，並不影響一般的地理資訊相關分析。

在個人隱私資料日益受重視的情況下，先進國家已著手建置統計區做為資料收集、分析與發佈的基本單元，而近年我國內政部國土資訊系統也完成「最小統計區」的建置（內政部統計處，2016），主要是考量社經指標、地形與行政疆界等，劃訂出穩定不變的單元界線，以利時間轉換後的時間序列分析。未來美兆健康資料庫也會將地理資訊轉換為最小統計區，提供會員更好的隱私保護機制，但也給研究者更理想的地理資訊分析資源。

七、結語

美兆會員的地址資料，累計了超過20年的內容，資料可能因為鍵入的不便、經歷轉換格式以及地理行政區域變革等因素，比較容易產生不完整或錯誤的地址，未來會再加強會員地址紀錄的完整性與正確性，以利後續地理編碼的順利運作，將來也會每年根據新收集的會員地址進行經緯度的轉換，成為例行

的工作，讓美兆健康資料庫擁有更完整與更新的地理資訊。

八、參考文獻：

1. 內政部統計處，2016，國土資訊統計區比對服務系統概述。取自
http://moisagis.moi.gov.tw/moiap/match/system_common.cfm
(2016/5/18)。