



# 致理科技大學

## 資訊管理系專題報告

人工智慧與機器學習

**Artificial intelligence  
and  
Machine learning**

專題生：

(10710140)游竣傑  
(10710151)謝雨倫  
(10710130)鄭宇傑  
(10710123)邱 邦  
(10710137)王盛弘  
(107101A1)吳竑慶

指導老師：曹祥雲 老師

中華民國 111 年 5 月



致理科技大學

資訊管理系

畢業專題

爬蟲與PII資料庫建置之研究



一一〇學年度



# 致理科技大學

## 授權書

本授權書所授權之專題報告在致理科技大學

110 學年度第 2 學期所撰寫。

**專題名稱：人工智慧與機器學習**

本人具有著作財產權之論文或專題提要，授予致理科技大學，得重製成電子資料檔後收錄於該單位之網路，並與台灣學術網路及科技網路連線，得不限地域時間與次數以光碟或紙本重製發行。

本人具有著作財產權之論文或專題全文資料，授予教育部指定送繳之圖書館及本人畢業學校圖書館，為學術研究之目的以各種方法重製，或為上述目的再授權他人以各種方法重製，不限時間與地域，惟每人以一份為限。並可為該圖書館館藏之一。

本論文或專題因涉及專利等智慧財產權之申請，請將本論文或專題全文延至民國 111 年 7 月 1 日後再公開。

上述授權內容均無須訂立讓與及授權契約書。依本授權之發行權為非專屬性發行權利。依本授權所為之收錄、重製、發行及學術研發利用均為無償。

(上述同意與不同意之欄位若未勾選，本人同意視同授權)

同意 不同意

**學生簽名：**

(親筆正楷簽名)

**指導老師姓名：**

(親筆正楷簽名)

中華民國 111 年 5 月

# 摘要

專題報告名稱： 人工智慧與機器學習

頁數：總頁數

校系別：致理科技大學資訊管理系

完成時間：110 學年度第 2 學期

專題生：游俊傑、謝雨倫、鄭宇傑、邱邦、王盛弘、吳竑慶

指導教授：曹祥雲

關鍵詞：論文、格式規範、中文摘要、關鍵詞

現今這個世代人人隨身都有智慧行動裝置，台灣和與台灣相鄰的香港等各區域的搜尋成為一項新興的熱點服務。然而為了讓區域搜尋這方面提供更準確和最新的資訊服務，必須讓使用者能夠方便快速的準確搜尋到附近的興趣點(Point of Interest, POI)，例如人們最基本的生活要素，食、衣、住、行、育、樂等，一些店家的地點。

所以我們組員在曹祥雲老師的協助下，根據原本的 POI 資料庫，去查詢店家的最新地址資訊，看看有沒有店家關店或搬家的，還是新增或維持不變的，並更新資料庫，提供最新的資訊。

如今很多商家或企業都會建立官方網頁，或是透過社群平台建立粉絲專頁，在網頁上詳細的介紹他們的資料，隨著這類型網頁的數量累積，現在網際網路已成為了最大的 POI 資訊來源。

# ABSTRACT

Thesis Title : Artificial Intelligence and Machine Learning

Pages :

University : Chihlee University of Technology

Graduate School : Department of Information Management

Date : May, 2022

Degree : Master

Researcher : YOU,JIUN-JIE; SHIE,YU-LUEN; JENG,YU-JIE; CHIOU,BANG;  
WANG,SHENG-HUNG; WU,HUNG-CHING

Advisor : CAO XIANG YUN

Keywords :

Nowadays, everyone has a smart mobile devices with them , and search in Taiwan and Hong Kong , which is adjacent to Taiwan, has become an emerging hot service. However , in order to provide more accurate and up-to-date information services in area search, users must be able to quickly and accurately search for nearby Point of Interest (POI) , such as people most frequently exposed to food, clothing , housing , transportation , and education. The location of some stores.

Therefore, our team members , with the assistance of Teacher Cao , according to the original POI database, check the latest address information of the store to see if there is any store closing or moving, or new or unchanged, and update the database to provide the latest information.

Nowadays, many businesses or enterprises will set up official webpages, or create fan pages through social platforms to introduce their information in detail on the webpages. With the accumulation of such types of webpages, the Internet has now become the largest POI information sources.

## 誌謝

誌謝所有對研究提供協助的個人或單位表達謝忱。

感謝學校的教學資源以及曹祥雲老師給予我們的幫助，也感謝學長姊們的資料輔助，幫助我們完成專題。

感謝組員們的幫忙，平均分配蒐集、並更新台灣及香港各地方店家資料。大學四年學習時光已經接近尾聲，感謝身邊的家人在大學四年學習的默默支持；感謝致理科技大學的老師和同學們四年來的關心和鼓勵。

特別感謝曹祥雲老師的幫助，我們才能處理這龐大的資料，專題指導老師曹祥雲老師的關心和支特尤為重要。每次遇到難題，曹祥雲老師每次不管忙或平常空閒，都會抽空時間開會或提供一些建議，然後一起商量解決的辦法。

感謝我們的客戶 Here 公司，願意給我們參與此次的計畫合作，讓我們學習到遇到困難、解決困難的能力，也讓我們有貢獻學習的機會。

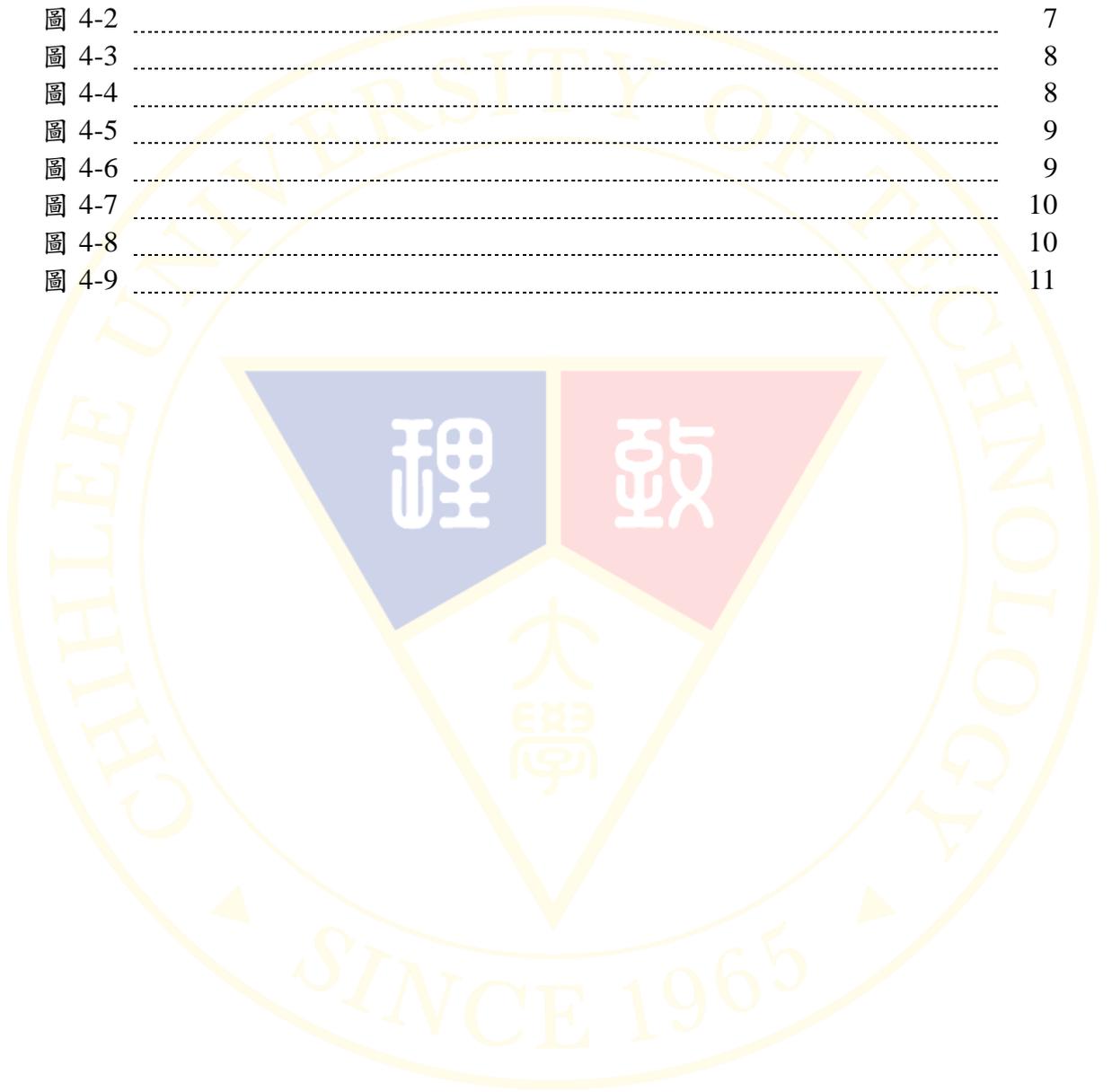
所有組員：游俊傑、謝雨倫、鄭宇傑、邱邦、王盛弘、吳竑慶 謹致  
致理科技大學 資訊管理 學士班  
中華民國 111 年 5 月

# 目錄

目錄	i
圖目錄	ii
第壹章 緒論	1
第一節 研究背景	1
第二節 研究動機	1
第三節 研究目的	1
第貳章 文獻回顧與探討	2
第一節 POI 興趣點	2
第二節 Here 公司	3
第三節 技術	4
第參章 研究內容與方法	5
第一節 研究內容	5
第二節 研究流程	6
第肆章 實驗結果	7
第一節 結案報告	7
第二節 開發工具	11
第伍章 結論與建議	13
第一節 結論	13
第二節 未來發展	13
參考資料	14

# 圖目錄

圖 3-1	.....	5
圖 3-2	.....	6
圖 3-3	.....	6
圖 4-1	.....	7
圖 4-2	.....	7
圖 4-3	.....	8
圖 4-4	.....	8
圖 4-5	.....	9
圖 4-6	.....	9
圖 4-7	.....	10
圖 4-8	.....	10
圖 4-9	.....	11



# 第壹章 緒論

## 第一節 研究背景

隨着智慧型手機與網路的普及，出門在外，一台手機就可以處理和幫助大大小小的各種事，給人們帶來極大的便利，例如在人生地不熟的地方，一台手機就能幫你找到你要走的路。

網路地圖能帶來給我們極大的便利，不過除了一些基本的地理資訊外，還必須要仰賴其背後持有大量資料的資料庫，才有資料可以搜尋使用。

在我們的日常生活中，例如遊客使用智能手機查找附近餐廳或停車場，因此，包含業務/組織的 POI 數據庫名稱、地址、位置（經度，緯度）和支持的簡單描述讓用戶查詢特別有用。

## 第二節 研究動機

我們受 Here 這家地圖資訊公司的委託，與他們合作，要取得台灣及香港相關圖資，所以請我們先利用來自政府開放平臺(OpenData)上的資料，接著再根據他們的資料庫格式轉入，其中的資料清理(data cleaning)需要大量的成本，為了解決地圖資料庫在處理龐大資料上的過程需要花大量人工及時間，試圖想辦法研究一個深度學習來縮短這些成本。

## 第三節 研究目的

在現今科技發達下，雖有政府開放平臺(OpenData)企業的登記資料，但是有些店家名稱的顯示或一些欄位還是會有不一致的情況發生，例如店家興趣點 POI (Point of Interest, POI) 會有一些漏掉，在查詢的過程中還發現部分 POI 會出現在除了開放平台的其他網頁中，例如 Facebook，官網等。

在現有的深度學習大多用於影像分析，而資料分類則存在著需要藉由人員來資料標記。在網路上深度學習分類的文獻較少，且我們試圖解決日後不需倚靠著大量人工標記成本，所以希望透過結合深度學習並加入訓練資料分類及修改甚至清除的動作，讓訓練模式越來越好。

## 第貳章 文獻回顧與探討

### 第一節 POI 興趣點(Point of Interest)

電子地圖上某個地標、景點，用以標出該地所代表的政府部門、各行各業的商業機構(加油站、百貨公司、超市、餐廳、酒店、便利商店、醫院等)、旅遊景點(公園、公共廁所等)、古蹟名勝、交通設施(各式車站、停車場、超速照相機、速限標示)等處所。由於興趣點必須包含名稱、類別、經緯度、海拔等資料才能在電子地圖上呈現。

通常興趣點的來源可分為三種：

#### (一)公司

是專門建構此類資料的商業公司，如 Google 地圖、Sensis、WhereiS、CarteBlanche、Papago、高得、百度地圖、Here 等業者。這類興趣點的優勢是資料眾多，不過相對而言內容稍嫌制式化、且內容豐富度各家不一。

#### (二)自製

是網友自發性的製作，這一類資料最符合「興趣點」的本意，因為網友自行編輯上傳該地點的文字、照片、影片、甚至個人評論心得等，使之成為真正符合旅遊指南的興趣點。還有一種另類的興趣點資料便是「測速照相」的興趣點，喜歡開快車的網友搜羅各路段的速度限制，以及各地的固定式、流動式超速照相機的設置地點，上傳網路供駕駛們使用參考。

#### (三)客製化

來源是第一種資料的延伸運用，也就是客製化興趣點。譬如某地方政府打算振興當地的觀光及商業活動，於是可以和電子地圖資料公司專案合作，除了介紹當地的名勝遊覽行程，更可結合當地商家的資料推出優惠活動，以吸引旅客前往消費。

## 第二節 Here 公司

Here 是一個總部設於荷蘭埃因霍溫的跨國企業集團，提供電子地圖、地理位置服務及導航服務。曾是諾基亞旗下的子公司，現為德國汽車廠商奧迪、BMW、戴姆勒合資持有。

HERE 地圖包括建築、公園、交通等資訊，同時提供到商業公司的導航，例如：Garmin、BMW、亞馬遜、甲骨文公司。

此外，HERE 地圖不僅在諾基亞、微軟 Lumia 手機上提供服務，還能執行在其他作業系統上，包括 iOS、Android 和 Firefox OS，和 BMW、福特等廠牌的汽車車載系統。Here 地圖提供定位服務、地理資訊系統和政府的解決方案。通過 Bing，MapQuest 和雅虎地圖提供電子地圖。Here 地圖有 200 多個國家地圖，在 94 個國家提供語音導航，在 33 個國家提供交通資訊，在 45 個國家為 49000 特色建築提供室內地圖。

跨平台：

### (一)Here 地圖

目前，HERE 地圖在 196 個國家可用，有內建路線規劃導航功能、離線地圖、3D 路標和 45 個國家 29000 獨特建築室內地圖。收藏列表顯示最喜歡的 25 個地方，並且顯眼。搜尋問題和其他使用者資料。這個應用整合了基於擴增實境技術的 LiveSight 功能。之後手機就會掃描使用者周圍的環境，並給出相關資訊，如位址、電話號碼和消費者評級等，並在手機相簿中顯示預覽。

### (二)Here Drive

提供 94 個國家駕車導航。提供視覺和聲音的限速提醒。在超過 60 個國家和 50 種語言的街名的離線語音。使用者介面為駕駛員設計，而且地圖資料包括轉彎限制、邊界和單行道等 260 多個方面。

### (三)Here Transit

擁有 50 多個國家 700 多個城市公共運輸資訊，一個應用整合了火車、輪船、公交、電車和步行資訊。

## 第三節 技術

### 一、Minecraft office:

由微軟公司開發的辦公軟體，其中我們都是用 Excel。

### 二、Opendata(政府開放平臺):

指的是一種經過挑選與許可的資料。這種資料不受著作權、專利權，以及其他管理機制所限制，可以開放給社會公眾，任何人都可以自由出版使用。

### 三、.Tgos (地理資訊圖資雲服務平台):

台灣內政部資訊中心維運的系統，定位為全國地理空間資料及網路服務搜尋取用、瀏覽查詢與加值媒合之入口，擁有最完整及正確的全國地理資料與網路服務之查詢目錄與詮釋資料庫，可以很方便地在平台上找到自己想要的地理資料與網路服務。

### 四、Google 地圖:

由 Google 公司所開發，是 Google 公司向全球提供的電子地圖服務，地圖包含地標、線條、形狀等資訊，提供向量地圖、衛星相片、地形圖等三種視圖。街景功能提供水平方向 360°及垂直方向 180°的街道全景，讓使用者能檢視所選城市地面上街道不同位置及其兩旁的景物。

### 五、香港地址解析器:

由於香港的地址和台灣不一樣，所以需要這個用來拆解複雜的地址、地址轉座標、中英翻譯地址、統一地址格式。

### 六、Python 軟體:

是一種廣泛使用的直譯式、通用型程式語言。相比於 C++或 Java，Python 讓開發者能夠用更少的代碼表達想法。

#### (一) spider :

是一個使用 Python 語言的開放原始碼跨平台科學運算整合開發環境 (IDE)。

#### (二) jupyter notebook:

是一個介於 IDE 以及 Editor 之間的一個讓你可以寫 code 的工具。並利用直譯語言的特性，可以很容易的逐行執行並且做到資料視覺化，而且可輸出成 Html5 嵌入到任何網頁或是 Blog 上，或是用 notebook 的形式分享。

# 第參章 研究內容與方法

## 第一節 研究內容

### 一、台灣資料

由於資訊會不斷的更新變動，台灣政府開放資料平台上的資料與 Here 公司的內部資料庫資料會有一些不同，我們使用了內政部的 Tgos (地理資訊圖資雲服務平台)，這個平台裡面可以進行各門牌批次比對的動作，也就是說可以先比對店家是否還在，以及門牌地址的正確性，經過輸出後的檔案再整理完後，提供給我們的合作對象 Here 公司查詢及檢查的動作，除了使用 Tgos，也有使用 Google 地圖，以及直接去店家的官網去查詢。以下有流程與解說。

### 二、香港資料

由於香港的資訊有些比較不容易取得，地址格式也與台灣不同，所以我們先利用 Python 寫出一支網頁爬蟲程式，爬取各店家官網上分店的各種資訊，因為抓取下來的資料還沒有很詳細，接著再利用香港地址解析器來拆解複雜的地址、地址轉經緯度、中英翻譯地址、統一地址格式。彙整到 Excel 及資料庫上，最後統整的結果結合到 Here 公司的資料庫進行篩選與檢查。以下有流程與解說。

## 第二節 研究流程

至 Tgos、店家的官網、Google 地圖等等進行搜尋，經由已建好的資料庫執行新增、修改、刪除的動作，最後供客戶查詢，如下圖流程由此可見。

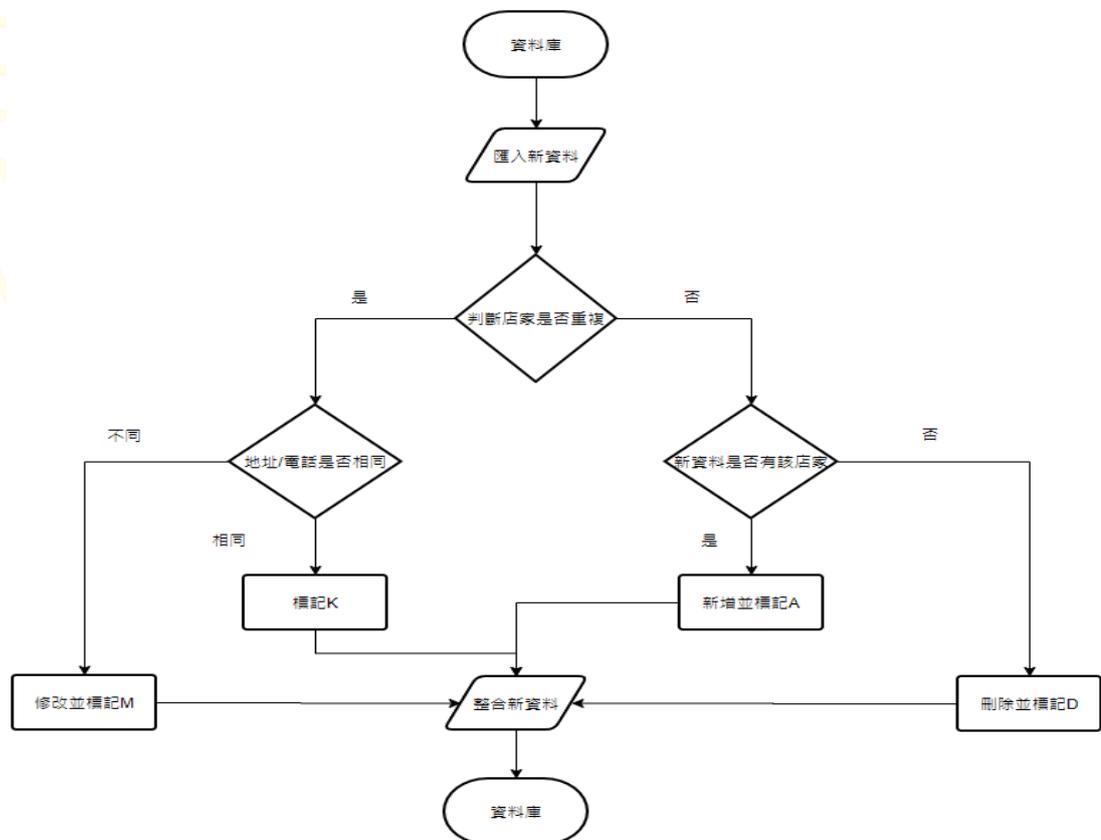


圖 3-1

## 研究流程其一

將網站上的店家資料利用 Opendata、Tgos 或 Python 抓出來並匯成 Excel。

```
In [1]: import requests
import csv

from bs4 import BeautifulSoup
req = requests.get("https://www.welcome.com.hk/wd2shop/zh/html/customer-services/store-locator.html?currLoc=R_HK")

#print(req.text)
f = open("hk_welcome.csv", "w", encoding = 'utf8')
w = csv.writer(f)
sp = BeautifulSoup(req.content,"html.parser")

st = sp.find("div",class_="storeList").findAll("dl")
for dl in st:
    dd = dl.findAll("dd")
    data = [dd[0].text,dd[1].text,dd[2].text,dd[3].text,dd[4].text]
    print(data)
    w.writerow(data)
f.close()
```

```
[ '安定', '新界屯門鄉事會路2A號安定邨H.A.N.D.S N區3樓N321號舖', '2690 4668', '08:00-23:00', '' ]
[ '安基', '九龍牛頭角安基商業中心地下及1樓', '2799 9218', '07:00-22:30', '-' ]
[ '安蔭', '葵涌安蔭村安蔭購物中心地下', '2429 0850', '07:30-23:00', '星期一至日及公眾假期,於上午10時至晚上10時,購物滿$150,即享1小時免費泊車。' ]
[ '柏宜中心', '九龍旺角彌敦道771-775柏宜中心2號地舖及地庫', '2154 1367', '08:00-22:00', '' ]
[ '百匯軒', '九龍荔枝角道33號百匯軒地下3-4號舖', '2521 6328', '08:00-23:00', '-' ]
[ '保安道', '九龍李鄭屋村李鄭屋購物中心地下102號', '2725 0426', '08:00-23:00', '星期一至日:上午10時至晚上9時30分 購物滿$150可享一小時免費泊車' ]
[ '卑路乍街', '香港卑路乍街46A隆基大廈 G/F', '2818 5277', '08:00-23:00', '-' ]
[ '北角', '北角英皇道 338 號華懋交易廣場地下', '2508 6197', '07:30-24:00', '-' ]
[ '碧湖花園', '新界粉嶺碧湖花園上層31號', '2677 9324', '08:00-22:00', '星期一至五上午7時至晚上10時:於商場兩間商戶購物滿$200,首2小時可以半小時收費泊車;星期六,日及公眾假期上午7時至晚上10時:於商場兩間商戶購物滿$400,首2小時可以半小時收費泊車' ]
[ '博康', '沙田博康村第一期商業中心126室', '2649 5936', '08:00-22:00', '-' ]
[ '彩虹', '九龍彩虹村全幢樓地下5-9號', '2350 4526', '07:30-22:00', '-' ]
[ '彩明3', '新界將軍澳將軍澳彩明街1號彩明商場3樓3078舖', '3104 4991', '08:00-23:00', '' ]
[ '彩園超級廣場', '上水彩園村彩園購物中心地下G1號舖', '3124-0140', '07:00-24:00', '星期一至五:任何一間商戶購物滿$200可享兩小時免費泊車 星期六,日:任何一間商戶購物滿$200可享一小時免費泊車' ]
[ '彩雲', '九龍牛池灣清水灣道45號彩雲一村三期商場三樓B301號舖', '2701 1244', '07:30-23:00', '' ]
```

圖 3-2

## 研究流程其二

將資料進行分類:

新增:A

修改:M

刪除:D

不變:K

P22	D	E	F	G	H	I	J	K	L	M	N	O
1	EnglishName	ChainName	Category	Subcategory	FoodType	Address	Lat	Lon	Phone	E-Mail	WebURL	ModifyK
2	EatonHotel	EatonHotel	Hotel			九龍油蔴地佐敦道30號	22.307863	114.1718915	27821818			A
3	Rimowa	Rimowa	SpecialtyStore			九龍油蔴地尖沙咀柯士頓道西1號(匯方)	22.3052815	114.1604951	2789300	store.HongKong@rimowa.com		A
4	Rimowa	Rimowa	SpecialtyStore			新界離島區赤鱗角翔天路1號(機場)	22.3162916	113.93064	29189968	store.HongKong@rimowa.com		A
5	Rimowa	Rimowa	SpecialtyStore			香港中西區中環皇后大道中15號(置地廣場)	22.28134805	114.1581147	25232780	store.HongKong@rimowa.com		A
6	Rimowa	Rimowa	SpecialtyStore			九龍油蔴地尖沙咀廣東道322號(海港城)	22.2971829	114.1690001	61463000	store.HongKong@rimowa.com		A
7	Rimowa	Rimowa	SpecialtyStore			香港灣仔區軒尼詩道139號(利園2)	22.276567	114.1654426	26326983	store.HongKong@rimowa.com		A
8	Rimowa	Rimowa	SpecialtyStore			香港中西區金鐘金鐘道88號(太古廣場)	22.2778174	114.1646916	23115777	store.HongKong@rimowa.com		A
9	Rimowa	Rimowa	SpecialtyStore			九龍油蔴地尖沙咀廣東道322號	22.2971829	114.1690001	23026888	store.HongKong@rimowa.com		A
10	Rimowa	Rimowa	SpecialtyStore			九龍油蔴地尖沙咀廣東道322號	22.2971829	114.1690001	23026888	store.HongKong@rimowa.com		A
11	Rimowa	Rimowa	SpecialtyStore			香港灣仔區銅鑼灣勿地臣街1號(時代廣場)	22.2781892	114.1827268	24132211	store.HongKong@rimowa.com		A
12	ChromeHearts	ChromeHearts	ClothingStore			香港中西區中環皇后大道中10號	22.28149884	114.1593171	28106932	hongkong@chromehearts.com		A
13	ChromeHearts	ChromeHearts	ClothingStore			九龍油蔴地尖沙咀梳士巴利道22號	22.2948076	114.171797	27396113	hongkong@chromehearts.com		A
14	Mulberry	Mulberry	ClothingStore			九龍油蔴地尖沙咀廣東道322號	22.2971829	114.1690001	34662960			A
15	Hublot	Hublot	SpecialtyStore			九龍油蔴地尖沙咀廣東道322號	22.2971829	114.1690001	27031808	hongkong@hublot.com		A
16	Hublot	Hublot	SpecialtyStore			九龍油蔴地尖沙咀廣東道322號(力保木舖)	22.2962363	114.1696905	23026888			A
17	Hublot	Hublot	SpecialtyStore			九龍油蔴地尖沙咀廣東道2A號(188時尚廣場)	22.29551924	114.1702984	27812728	Hongkong@hublot.com		A
18	Hublot	Hublot	SpecialtyStore			香港中西區中環金鐘道88號(國際金融中心)	22.28411043	114.1590205	21167008	hongkong@hublot.com		A
19	Hublot	Hublot	SpecialtyStore			香港中西區中環德輔道中111號	22.2816217	114.1562376	25283968	hongkong@hublot.com		A
20	AnteService&Maintenance	AnteService&Maintenance	AutoService&Maintenance			香港九龍灣觀塘祥泰街大高行集團大廈少樓	22.3244257	114.2083536	21721655	enquiry@hauheymotors.com.hk		A
21	DCH	DCH	AutoService&Maintenance			香港九龍灣觀塘祥泰街大高行集團大廈少樓	22.3244257	114.2083536	21721204	asia@hauheymotors.com.hk		A
22	DCH	DCH	AutoService&Maintenance			香港九龍灣觀塘祥泰街大高行集團大廈少樓	22.3244257	114.2083536	21721204	asia@hauheymotors.com.hk		A
23	DCH	DCH	AutoService&Maintenance			新界元朗新田粉嶺地段7449	22.5047884	114.0770942	27783006			A
24	DCH	DCH	AutoService&Maintenance			香港九龍灣觀塘祥泰街大高行集團大廈少樓	22.3244257	114.2083536	25231300	service_gmail@hauheymotors.com.hk		A
25	DCH	DCH	AutoService&Maintenance			香港九龍灣觀塘祥泰街大高行集團大廈少樓	22.3244257	114.2083536	23975111	hauheymotors@hauheymotors.com.hk		A
26	DCH	DCH	AutoService&Maintenance			九龍旺角彌敦道403-406號(東區大廈地下)	22.3247911	114.1675771	2399-7679	traco@trangle.com.hk		A
27	DCH	DCH	AutoService&Maintenance			香港九龍灣觀塘祥泰街大高行集團大廈少樓	22.3244257	114.2083536	2768-2438	info@trimenotors.com.hk		A
28	Maxim'sPalace	Maxim'sPalace	Restaurant	CHINESEFOOD		香港中環德輔道中廣東銀行對面大都會大廈少樓	22.2824488	114.1619257	25211303			A
29	Maxim'sPalace	Maxim'sPalace	Restaurant	CHINESEFOOD		香港上環干諾道中188-200號怡和中心地庫B13-B18號舖	22.2876555	114.1518223	25210098			A
30	Maxim'sPalace	Maxim'sPalace	Restaurant	CHINESEFOOD		九龍灣觀塘祥泰街一期54號	22.3245376	114.2114027	27508028			A
31	Maxim'sPalace	Maxim'sPalace	Restaurant	CHINESEFOOD		新界荃灣區荃灣街10-16號	22.3725419	114.1556554	24968983			A
32	Maxim'sPalace	Maxim'sPalace	Restaurant	CHINESEFOOD		新界屯門屯門街三號屯門大會堂	22.3931677	113.9772655	24300688			A
33	Maxim'sPalace	Maxim'sPalace	Restaurant	CHINESEFOOD		新界屯門屯門街三號屯門大會堂	22.3931677	113.9772655	24300688			A

圖 3-3

# 第肆章 實驗結果

## 一、結案報告

(一) 主要做台灣和香港兩個部分

此圖表為台灣和香港的類別與資料的筆數

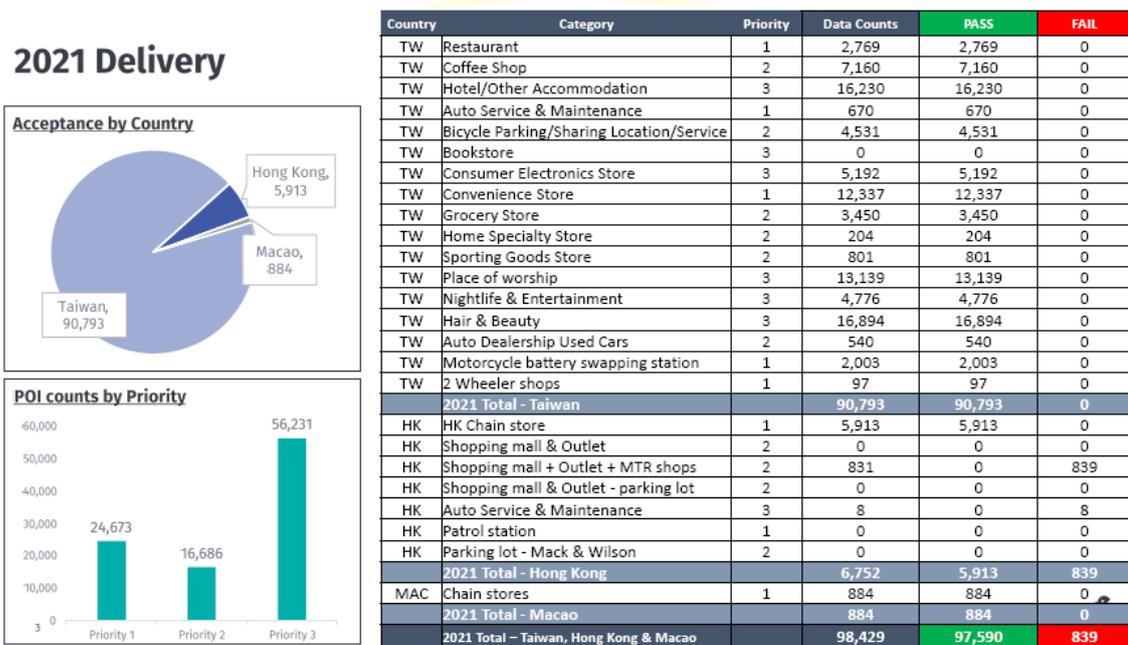


圖 4-1

(二) 台灣的 POI 類別筆數

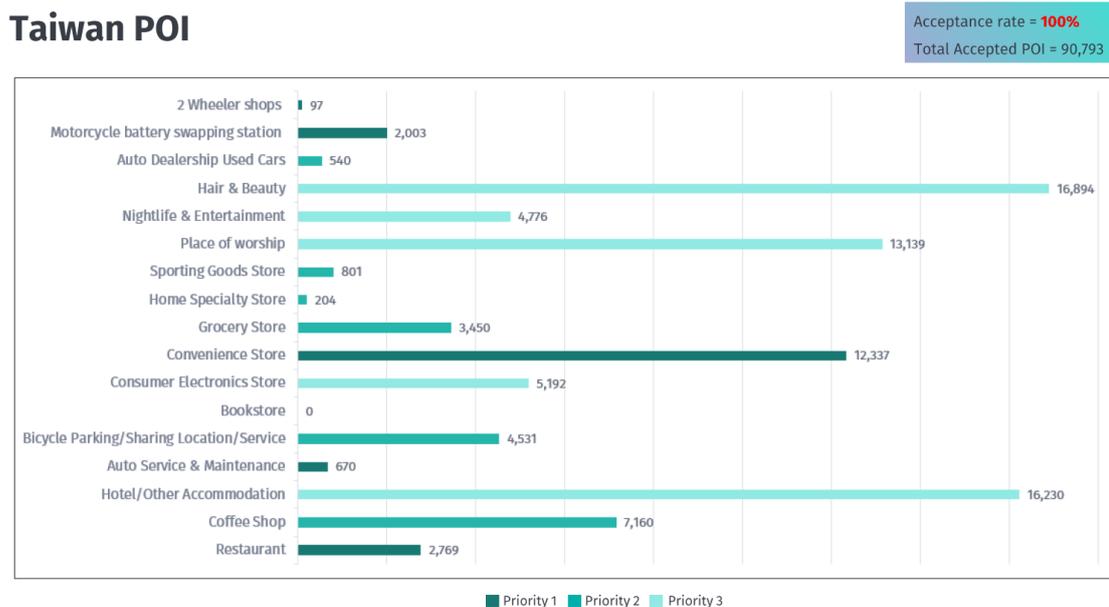
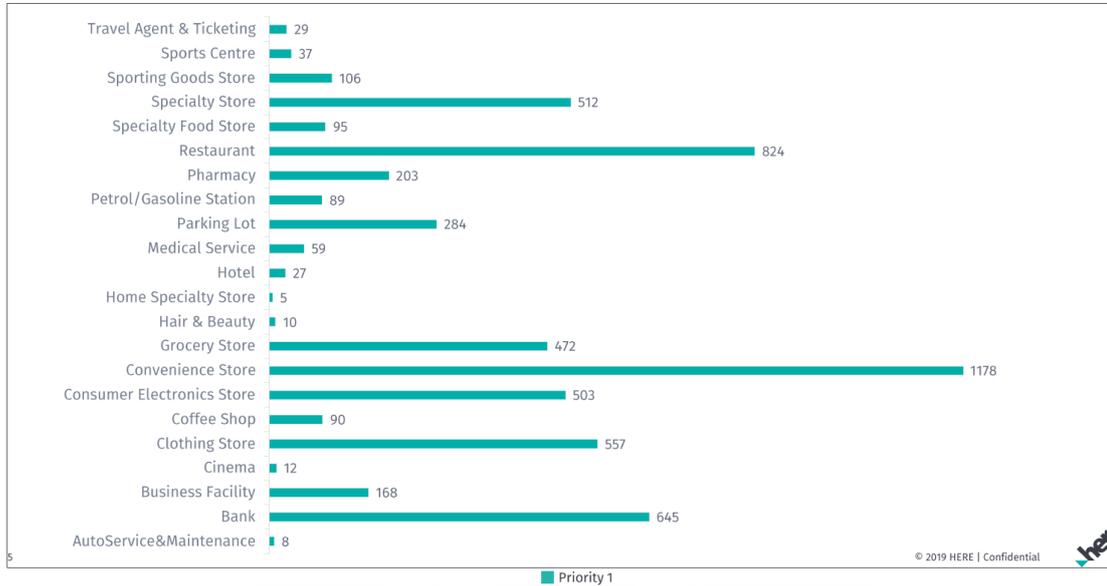


圖 4-2

(三) 香港的 POI 類別

Hong Kong POI

Acceptance rate = **88%**  
Total Accepted POI = 5,913



(四) 圖 4-3

(五) 獎勵  
做資料的酬勞

Incentive

	Accept NT\$/POI	Acepected POI	Reject NT\$/POI	Reject POI	Total
Taiwan	1.8	90,793	0	0	163,427
Hong Kong	4	5,913	2	839	25,330

圖 4-4

(六) 來源評估第一次繳交-台灣

資料準確度的統計

Source Evaluation

1st delivery - TW POI

Source evaluation – **PASS**  
 • Score = 92%  
 • EES ticket = [34944](#)

High level Summary - Accuracy

Chileeuni TQS Breakdown

Measurement Accuracy

Good Data : 22647, 92%  
 Bad Data : 0  
 Unusable/Reject : 1985, 8%  
 Quality level \*: Acceptable

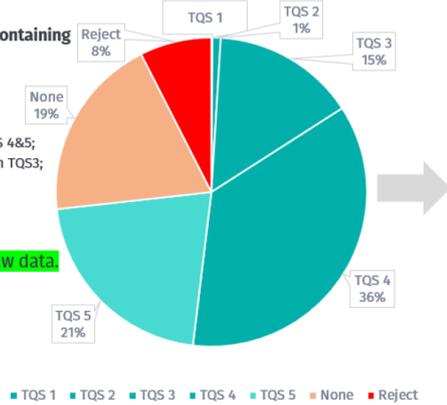
April 2021 sample file was delivered containing 197,435 records

WHAT WAS IDENTIFIED

61.97% of Raw Chileeuni is matching with TQS 4&5;  
 17.17% of Chileeuni is matching with less then TQS3;  
 20.85% is None  
 8 % is reject

WHAT DOES THIS MEANS?

High confidence on 61.97% of the raw data  
 8% rejected (useless)  
 38.03% need to be evaluated



9

© 2019 HERE | Confidential

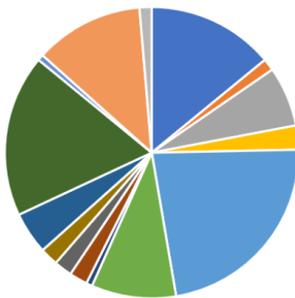
圖 4-5

(七) 來源評估第二次繳交-香港

Source Evaluation

2nd delivery - HK POI

Source evaluation – **PASS**  
 • Sample Count = 150  
 • Error Count = 6  
 • Score = 97%  
 • EES ticket = [34943](#)



- Bank
- Business Facility
- Clothing Store
- Consumer Electronics Store
- Convenience Store
- Grocery Store
- Hotel
- Medical Service
- Parking Lot
- Petrol/Gasoline Station
- Pharmacy

	Sample Count	Error Count	Address	website	Closed
Bank	21				
Business Facility	2				
Clothing Store	10				
Consumer Electronics Store	4				
Convenience Store	34				
Grocery Store	14				
Hotel	1				
Medical Service	3				
Parking Lot	3				
Petrol/ Gasoline Station	3				
Pharmacy	7				
Restaurant	27	1			1
Specialty Food Store	1	5	1	4	
Specialty Store	18				
Sporting Goods Store	2				
Grand Total	150	6			

10

© 2019 HERE | Confidential

圖 4-6

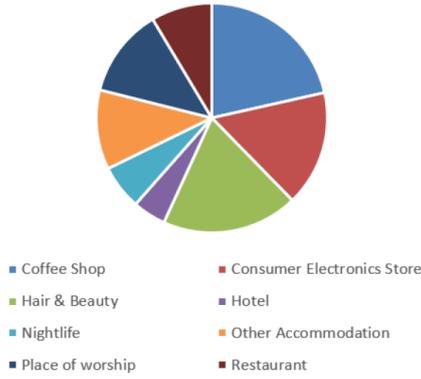
(八)來源評估第三次繳交-台灣

**Source Evaluation**

3rd delivery - TW POI

Source evaluation – **PASS**

- Sample Count = 750
- Error Count = 93
- Score = 88%
- EES ticket = [37040](#)



	Sample Count	Error Count	Closed	Address	Naming
Coffee Shop	161	10	10		
Consumer Electronics Store	122	17	14	3	
Hair & Beauty	143	22	14	1	7
Hotel	35	4	2		2
Nightlife	47	21	20	1	
Other Accommodation	84	3	1	1	1
Place of worship	94	2		2	
Restaurant	64	14	13	1	
<b>Total</b>	<b>750</b>	<b>93</b>	<b>74</b>	<b>9</b>	<b>10</b>

11

© 2019 HERE | Confidential



圖 4-7

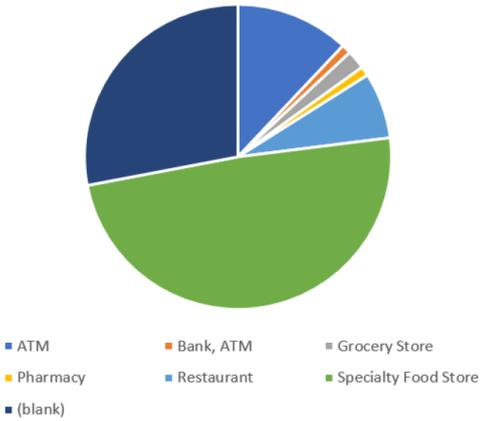
(九)來源評估第四次繳交-香港

**Source Evaluation**

4th delivery - HK POI

Source evaluation – **FAIL**

- Sample Count = 100
- Error Count = 97
- Score = 3%
- EES ticket = [37041](#)



• Did not require delivery as POI open data is acquired

	Sample Count	Error Count	Address	Naming/ category
ATM	12	12		12
Bank, ATM	1	1		1
Grocery Store	2	2		2
Pharmacy	1	1		1
Restaurant	7	7		7
Specialty Food Store	49	49	4	47
(blank)	28	25	3	20
<b>Total</b>	<b>100</b>	<b>97</b>	<b>7</b>	<b>90</b>

12

© 2019 HERE | Confidential



圖 4-8

## 二、開發工具

台灣資料的部分我們大多都是運用 Tgos 這個平台去批次查找店家門牌地址，但是還是會有漏網之魚的情況發生，所以這部分解決方案我們經過討論選擇使用 Google 街景服務去尋找，輔助我們這方面的資料的完整度，不過也不用太大擔心，因為 Tgos 很少找不到或有漏掉的資料，正確率相當的高！另外使用這個平台也有一定的限制，要特別注意的是每一帳戶的使用者似乎有每月一定的查詢量，像我們遇到的狀況是每個月限制一萬筆，所以我們要好好規劃並利用這些資源去分配，達成我們的需求。

香港地區資料的部份我們是如上面有提到就是除了利用網路上以及 Here 公司提供的現有資料去查詢外，查不到的資料就寫一支 Python 網頁爬蟲程式去抓取某店家官網上的分店資訊，但此時抓取下來的資料是相當雜亂的，而且資料部分有點遺漏加上不是我們要的資料庫格式，接著我們會利用 Excel 軟體先進行資料人工處理，例如說篩選功能，很快就能找到殘缺的欄位，再利用香港地址解析器把需要的資料填補上去，少部分甚至會使用到 FB、開飯喇等網頁，為了確保資料庫能規模化且完整度。

## 三、TGOS 平台的定位：

地理資訊圖資雲服務平台的定位是對本專題來說是網路服務搜尋取用，該網路服務之查詢目錄有很完整的資料庫當後盾，我們只需在 TGOS 平台，經過比對查詢資料就能發現想要的內容是什麼，不僅讓我們專題文獻上有所幫助，還能讓一般民眾只要申請一個帳號，就可以很方便地在平台上查詢自己想要查的資料，而且還是免費的。

## 四、TGOS 的功能：

最主要的功能是提供一般民眾完整的地理資料與網路服務查詢，只要透過 TGOS，就可輕鬆查詢全國有什麼地理資料與網路服務，如果要更清楚的內容及資訊，可向該單位承辦人申請或進一步詢問，完全免費。另外比較深層一點是還可以讓網路地圖服務的需求者先做套操作及瀏覽，判斷使否要使用。



圖 4-9

## 五、不同層次使用者之使用效益：

本平台建置成功以後，無論是政府機關一般民眾都可以有資料的收穫，以下簡單說明比較常見的使用者來進行分析。

### (一)教育單位

對於環境教育相關議題上，因為要能快速取得教學的內容和研究素材，所以如果說能透過免費的 TGOS 平台，這樣就可以不用使用成本較高的空間視覺分析，成為教育單位喜歡用的網站之一。

### (二)一般民眾

透過原始資料的供應及政府、民間 GIS 系統快速建立，都能讓一般民眾可以直接獲取資料流通的便利以及該效益。透過網路使用 TGOS 取得資料，不僅有一點社會平等的概念，且能讓民眾對社區環境衛生意識更為了解，等同有綠化環境的功效。

## 六、TGOS 平台的使用限制：

(一)雖然功能強大的 TGOS 可以快速查詢大量資料，但現在還無法百分之百保證所有的網路地圖服務永遠正常上線，因為 TGOS 平台像是匯集後的大海，網路地圖服務就像是山上分支下來的河流，有幾條路堵住或是本身沒有水(資料)就無法匯到海洋(TGOS)，所以如果有發現有這種情況發生的一般民眾可以主動通知該單位，讓他們儘速復原，以避免造成使用者的困擾。

### (二)仍未容納全國所有的地理資料

由於有些資料牽涉機密性、敏感性，所以權責單位沒有向 TGOS 平台上申請，這點還請使用者多多包涵。

## 第五章 結論與建議

### 第一節 結論

我們的工作內容主要是先參照 Here 列出的名單後，分配每個組員負責的範圍，再利用 TGOS(經緯度)、Opendata(立案店家查詢)等資源去抓名單內各店家的相關資訊後，將這些資料編輯成 Here 要求的格式，再把大家處理完的資料結合，建立一個大型資料庫以供 Here 查詢。但從一開始參照名單到最後資料編輯完成，之中需要耗費很多的時間及精力，因此作業的速度及資料的準確度也會不太穩定。

因此我們也抽空討論交流如何讓作業更有效率，後來指導老師建議我們可以使用深度學習，讓 AI 學會自動分類、處理大量資料，不論在速度、準確度都可以有飛躍式的提升。雖然還是會有少部分資料需要人工，不過這方法的確讓我們可以更省時省力的去作業，這也是我們未來的期許。

### 第二節 未來發展

本專題已將資料庫建立完畢，剩下資料清理以及後續維護的部份，然而日後可能還是會有其他新名單需要抓資料，所以我們期待能早日實現上述所提到的深度學習技術。

我們持續會對深度學習去做研究，期待這技術能早日開發出來，並被廣泛應用，讓作業流程更加方便迅速。同時也希望未來 AI 能主動去自我思考模糊的資料叢，去預測該店家應分類在哪個業種(餐廳 or 大賣場 or 專賣店...)，並讓電腦也能懂人類的語言，使這項技術能夠更加平易近人。

## 參考資料

1.TGOS 網站(2021)，擷取至 [https://www.tgos.tw/tgos/Web/TGOS\\_About.aspx](https://www.tgos.tw/tgos/Web/TGOS_About.aspx)

2.POI 維基百科

<https://zh.wikipedia.org/wiki/%E8%88%88%E8%B6%A3%E9%BB%9E>

3. Here 公司，擷取至

[https://zh.wikipedia.org/wiki/Here\\_\(%E5%85%AC%E5%8F%B8\)](https://zh.wikipedia.org/wiki/Here_(%E5%85%AC%E5%8F%B8))

