

# GIS Visualisation for the Administrative Document Catalogue of the Government-General of Taiwan

Hajime Murai and Toshio Kawashima Future University Hakodate

## 1. Goal of this research

To grasp an overview of the enormous amounts of administrative documents

Whole documents have not yet been digitized in text data format.  
However, an overview of the contents of the document could be extracted based on digitized catalog of the documents

To investigate the usefulness of digitally archived text data of official documents for historical and social studies

### Example of catalog data

1 教育ニ関スル勅語勝本各学校：台北県／三角湧公学校・老梅公学校・海樹厝公学校・楊梅壠公学校・私立東門学校・私立台北簡易商工学校・台中県／台中小学校・苗栗小学校・土庫公学校・魚池公学校・他	1900/08/03
2 二 皇孫御降誕ニ付與、序管内各学校及七国語学校、医学校ニテ本月五日儀式ヲ挙げ祝 意ヲ表スヘキ様具 事・序長並ニ右学校長へ通牒ノ件	1901/05/02
3 三 臨時台灣土地調查局事務員及測量員ニ充ツハキ見習員設置ニ関スル勅令案ヲ内務大臣へ稟申	1901/01/14
4 四 台灣總督府度量衡司檢査官制及職員官等俸給令 勅令第六二号・勅令第六三号	1901/01/17
5 五 台灣總督府官制中改正 勅令第七四号	1901/01/23
6 六 教科用図書・編修及検定ニ關スル職員ノ件勅令 勅令第八三号	1901/01/23
7 七 台湾總督府編修官及編修官補官等俸給令 勅令第八四号	1901/01/23
8 八 台灣總督府地方官制中改正案ヲ内務大臣へ稟申	1901/01/23
9 九 明治三四年度予算編制ニ關スル官制並俸給令制定若ハ改正ヲ要スルモノノ草案提出方ノ件	1900/08/10
9 九一— 法院条例改正律令按 法院職員官等俸給令改正案提出方ノ件	1900/08/10
9 九二— 二法院職員官等俸給及定員令改正案	1900/11/06
9 九三— 法院職員官等俸給及定員令改正 勅令第八五号	1901/01/18
10 台湾總督府國語學校官制中改正 勅令第八六号	1901/01/23
11 —— 台湾總督府國語傳習所官制中改正 勅令第八六号	1901/01/23
一二 台湾總督府電話交換局官制中改正並ニ台湾總督府電話交換局職員官等俸給令 勅令第七七号・勅令第12七八号	1901/01/25

## 2. Data resource

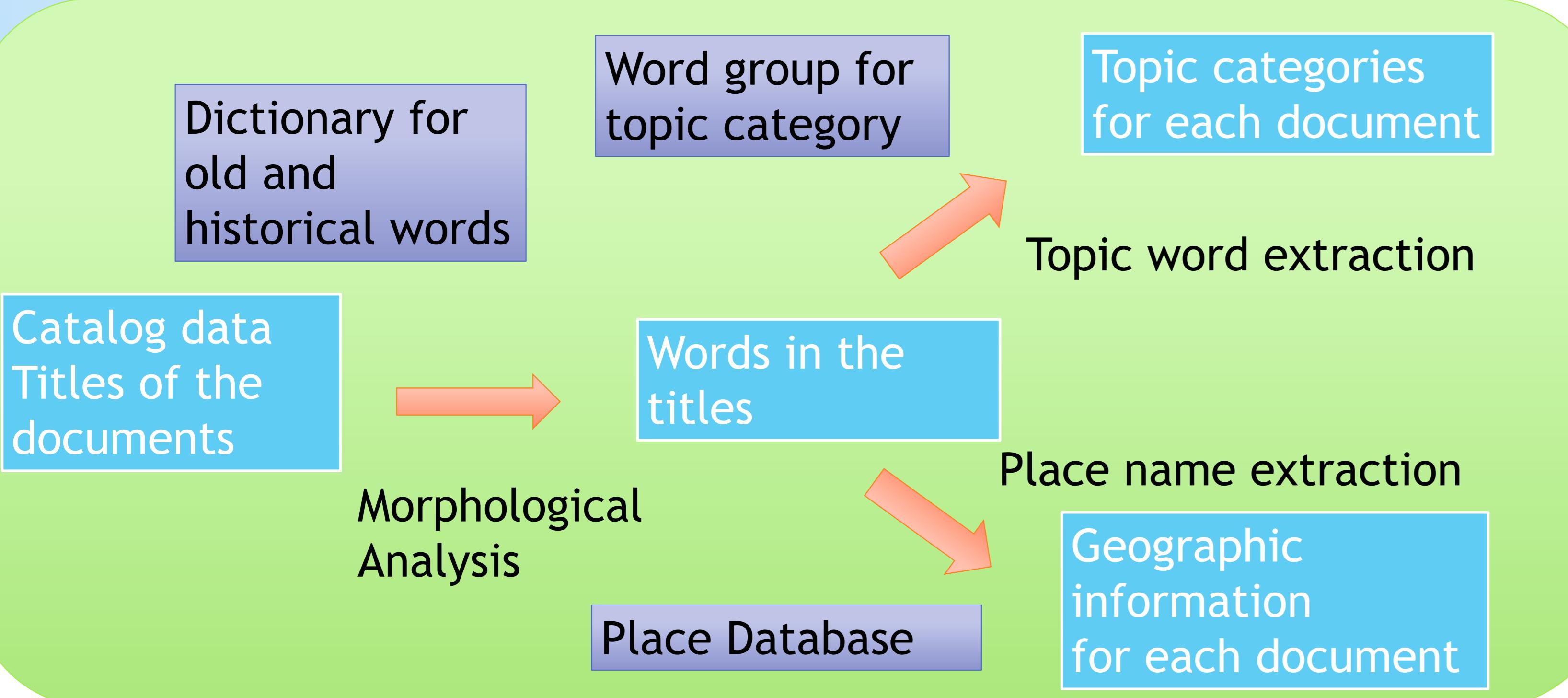
### Period

The first part (1895–1900)	24985	1~579
The second part (1901–1904)	24510	580~1050
The third part (1905–1908)	23752	1051~1451
The fourth part (1909–1912)	24782	1453~2087
The fifth part (1913–1914)	8859	2088~2340

Chukyo University has been made catalog data since 1982

Text format catalog has developed for documents from 1895 to 1914

## 3. Method



### 3.1 Constructing Dictionary for Historical Words

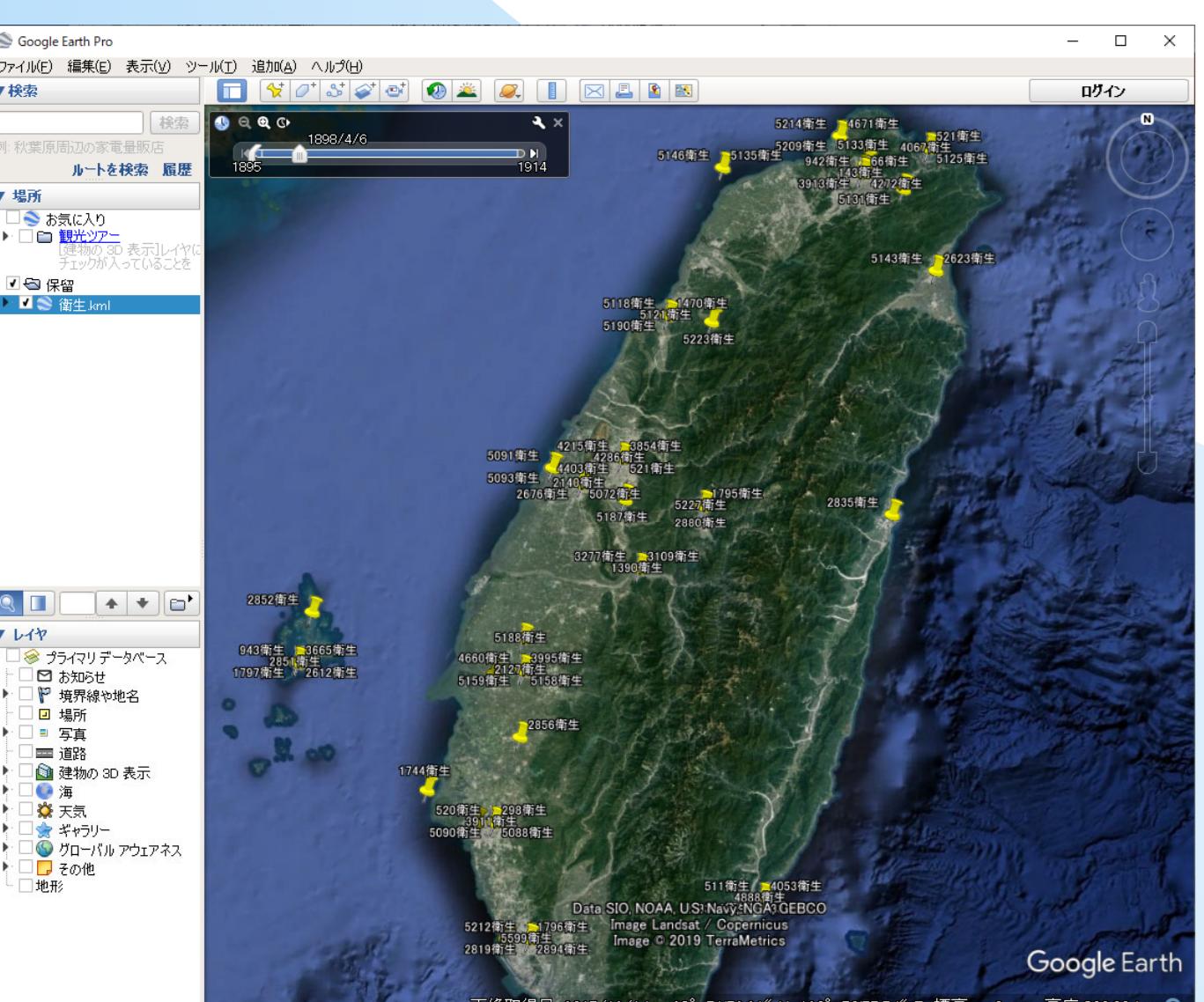
Name of the historical organization	Historical Japanese words	Historical words related to Taiwan
撫墾署 (Indigenous office)	勅詔 (Command paper)	隘勇 (Defense line)
弁務署 (Police)	聖詔 (Command paper)	土匪 (Indigenous)
警保 (Security, Police)	命免 (Appointments and dismissal)	蕃務 (Indigenous operations)
郵便電信局 (Post office)	紀元節 (Old holiday)	樟脑 (Camphor)

Total words: 197 words  
 • Old place names: 136 words  
 • Other words: 61 words

### 3.2 Constructing Topic Categories

Categories	Words	Examples of categorized words
Personnel shift	119	任官 (Commissioner), 採用 (Recruitment), 辞令 (Resignation), 人員 (Personnel), 雇員 (Employer), 職員 (Employee), 技師 (Engineer), 官吏 (Official)
Publicity	29	周知 (Publicity), 揭載 (Publication), 公布 (Promulgation), 告示 (Announcement), 訓諭 (Reminder)
Rules	29	官制 (Government System), 制定 (Establishment of law), 法律 (Law), 法令 (Law), 律令 (Statute), 条例 (Ordinance), 勅令 (Decree)
Office work	20	事務 (Office work), 書式 (Document form), 署名 (Signature), 印章 (Seal), 届出 (Notification), 統計 (Statistics)
Permission	8	認可 (Authorization), 許可 (Permission), 免狀 (Diploma), 免許 (License), 特許 (Patent)
Hygiene	91	衛生 (Hygiene), 病院 (Hospital), 医師 (Doctor), 製薬所 (Pharmacy), 藥劑師 (Pharmacist), 病人 (Disease), 患者 (Patient)
Livestock	18	家畜 (Livestock), 牧畜 (Pasture), 牛 (Cattle), 猪 (Pigs), 獣疫 (Veterinary Diseases), 瘟牛 (Plague Cattle), 瘟豚 (Plague Swine)
Economy and tax	36	会計 (Accounting), 予算 (Budget), 税 (Tax), 税法 (Tax Law), 地租 (Land Tax), 納稅 (Tax Payment), 關稅 (Customs)
Judgment	53	犯罪 (Crime), 非行 (Delinquency), 強盜 (Robbery), 密輸 (Smuggling), 懲罰 (Punishment), 死刑 (Death Penalty)
Infrastructure	21	水道 (Water supply), 電氣 (Electricity), 瓦斯 (Gas), 土木 (Civil engineering), 官有地 (Government-owned land), 地籍 (Land ownership)
Police	16	警察 (Police), 警察署 (Police station), 派出所 (Dispatch station), 警部 (Police officer), 警保 (Security)
Correspondence	14	郵便局 (Post office), 電信 (Telegraph), 郵便及電信局 (Postal and telegraph office), 送金 (Remittance), 為替 (Exchange)
Education	26	日本語 (Japanese language), 中国語 (Chinese language), 生徒 (Students), 学校 (School), 大学校 (University), 教育 (Education)
Monopoly	16	專賣 (Monopoly), 填田 (Salt field), 煙草 (Tobacco), 樟腦 (Camphor), 阿片 (Opium)
Establishment and closure	13	開設 (Opening), 設置 (Installation), 移転 (Relocation), 開庁 (Opening agency), 閉庁 (Closed agency)
Traffic	35	交通 (Traffic), 運搬 (Transportation), 輸出入 (Import/Export), 海事 (Maritime), 港湾 (Harbor), 灯台 (Lighthouse), 浮標 (Buoy)
Disaster and welfare	22	救恤 (Salvation), 助扶 (Assistance), 自殺 (Suicide), 年金 (Pension), 罷災 (Disaster), 台風 (Typhoon), 水害 (Flood)
Mining and engineering	12	鉱山 (Mine), 鉱業 (Mining), 石炭 (Coal), 硫黃 (Sulfur), 金銅 (Gold Copper), 石油 (Petroleum), 工場 (Factory), 工業 (Industry)
Rebellion and indigenous	22	土匪 (Indigenous), 賊匪 (Bandit), 凶徒 (Villain), 生蕃 (Rebellious indigenous), 熟蕃 (Cooperative indigenous), 蕃務 (Indigenous operations)
Military	56	大本營 (Headquarters), 海軍 (Navy), 陸軍 (Army), 軍隊 (Military), 軍務 (Military affairs), 軍艦 (Warship), 陸軍大臣 (Army minister)
Agriculture	33	糖業 (Sugar industry), 甘藷 (Sweet potato), 果樹 (Fruit tree), 茶樹 (Tea tree), 綿花 (Cotton), 護謾 (Rubber), 椰子 (Coconut)
International	25	國際 (International), 公法 (Public law), 条約 (Treaties), 外國 (Foreign countries), 欧米 (Europe and the U.S.), 領事 (Consulate)
Residents	7	住人 (Residents), 戸口調查 (Doorway survey), 戸口簿 (Family Register)
Imperial family and rituals	28	天皇 (Emperor), 明治天皇 (Emperor Meiji), 皇太子 (Crown Prince), 皇后 (Empress), 殿下 (His Highness), 祭典 (Celebrations)
Religion	29	宗教 (Religion), 布教 (Mission), 神社 (Shrine), 仏教 (Buddhism), 本願寺 (Honganji), 延暦寺 (Enryakuji), 護國寺 (Gokokuji)
Fisheries	7	水産業 (Fisheries), 水產物 (Fishery Products), 漁船 (Fishing boats), 漁夫 (Fishermen)

## 4. GIS



Selecting documents by included topic and date of promulgation

Visualizing based on longitude latitude of included city names

## 5. Conclusion and Future Works

- Visualization of the chronological and spatial shifts in historical topics
- The utility of the Catalogue of the Government-General of Taiwan demonstrated

### Utilization of named entity recognition

- It is necessary to wait for future development of NLP
- Applicable for other historical documents
- Appropriate datasets of old place names, geographical information, and words are necessary.

	Included place names	Appeared frequency
台北 (Taipei)	台北,大加蚋,水返脚,士林,芝蘭,北投,內湖,新莊,五股,海山,鶯歌石,土城,擺接,文山,新店,深坑,石碇,坪林,淡水,滬尾,八里,石門,瑞芳,万里,加投,大湖,山仔腳,大稻,大稻花,興直,直興,金包里,三貂	34142
台中 (Taichung)	台中,大里杙,霧峰,藍興,烏日,葫蘆,內埔,東勢,大甲,五里牌,牛罵頭,牛罵頭,梧棲,外埔,茄投,大肚,沙鹿,犁頭店,石崙,神岡,社口,三角仔,四張犁,塗葛堀,後里,潭子	17894
台南 (Tainan)	台南,新豐里,仁德,帰仁,永康,永寧,大目降,善化,安定里,左鎮,曾文,下營,茅港,六甲,北門,佳里,西港仔,學甲,新營庄,塭水港,後壁,番社,店仔口,拔仔,安平,閔帝廟,果毅,果毅後,查畝營	24689
桃園 (Taoyuan)	桃園,桃仔園,楊梅,新屋,興隆里內里,興隆外里	6717
高雄 (Kaohsiung)	鳳山,高雄,打狗,楠梓,楠仔坑,左營,三塊,苓雅,前鎮,旗後,阿公店,橋頭,橋頭仔,田寮,半路竹,大湖,湖內,弥陀港,竹仔港,梓官,港仔,大樹脚,仁武,烏松,六龜,杉林,甲仙,羅漢內門,雁爾,瀰濃,竹仔門	4774
基隆 (Keelung)	基隆,七堵,暖暖	9947
新竹 (Hsinchu)	新竹,新豐,竹東,寶山,竹北,新埔,湖口,北埔	2563
苗栗 (Miaoli)	苗栗,竹南,頭分,三灣,南庄,造橋,頭屋,公館,銅鑼,通霄,四湖,大湖,獅潭	6453
彰化 (Changhua)	彰化,鹿港,鹿仔港,馬芝,和美,樹仔腳,線西,線東,福興,秀水,芬園,員林,溪湖,田中央,埔鹽,永靖,社頭,北斗,二林,田尾,貓羅,溪州,深耕,二八水,燕霧	7812
南投 (Nantou)	南投,草鞋,埔里,魚池,集集,林圮埔,林圮埔,沙連下,東螺,霧社,皮仔柑	6431
雲林 (Yunlin)	雲林,斗六,古坑,虎尾,西螺,土庫,二崙,崙背,海口,北港,元長庄,四湖,口湖,埔姜,他里霧	7110
嘉義 (Chiayi)	嘉義,太保,布袋,双溪口,東石,中埔,大埔,新港,阿里山,打貓,六脚,鹿仔草,蒜頭,樸仔腳,朴仔腳,水堀頭,梅仔坑,大目根	12096
屏東 (Pingtung)	屏東,阿猴,阿麻,潮州,東港,恒春,万丹,麟洛,阿里港,塩埔,高樹,內埔,新園,茄苳,車城,枋山	9332
宜蘭 (Yilan)	宜蘭,礁溪,民壯團,員山,羅東,五結,蘇澳,蘇澳,叭哩沙	8095
花蓮 (Hualien)	花蓮,花蓮港,鳳林,奉鄉,林田村,七腳川,璞石閣,針柯,加礼宛,蕃薯	5759
台東 (Taitung)	台東,卑南,太麻里,大武,火燒島,新開園,池上村,新港,加走灣,馬武窟,呂家,知本,大埔尾,加路蘭,成廣灣	5920
澎湖 (Wuhu)	澎湖,馬公,湖西,白沙,西嶼,漁翁島,望安	5966

	Included place names	Appeared frequency
台北 (Taipei)	台北,大加蚋,水返脚,士林,芝蘭,北投,內湖,新莊,五股,海山,鶯歌石,土城,擺接,文山,新店,深坑,石碇,坪林,淡水,滬尾,八里,石門,瑞芳,万里,加投,大湖,山仔腳,大稻,大稻花,興直,直興,金包里,三貂	34142
台中 (Taichung)	台中,大里杙,霧峰,藍興,烏日,葫蘆,內埔,東勢,大甲,五里牌,牛罵	