



Mapping and Geocoding Historical Chinese Placenames

——CBDB 工作报告

袁钰莹

2019.2.21

提纲

➤ WorldMap

1350-1900年50年间隔的政区图

➤ 唐代地名的Geocoding

- 1) 在线数据: TGAZ
- 2) 本地数据: DDBC
- 3) 现代地图: 高德

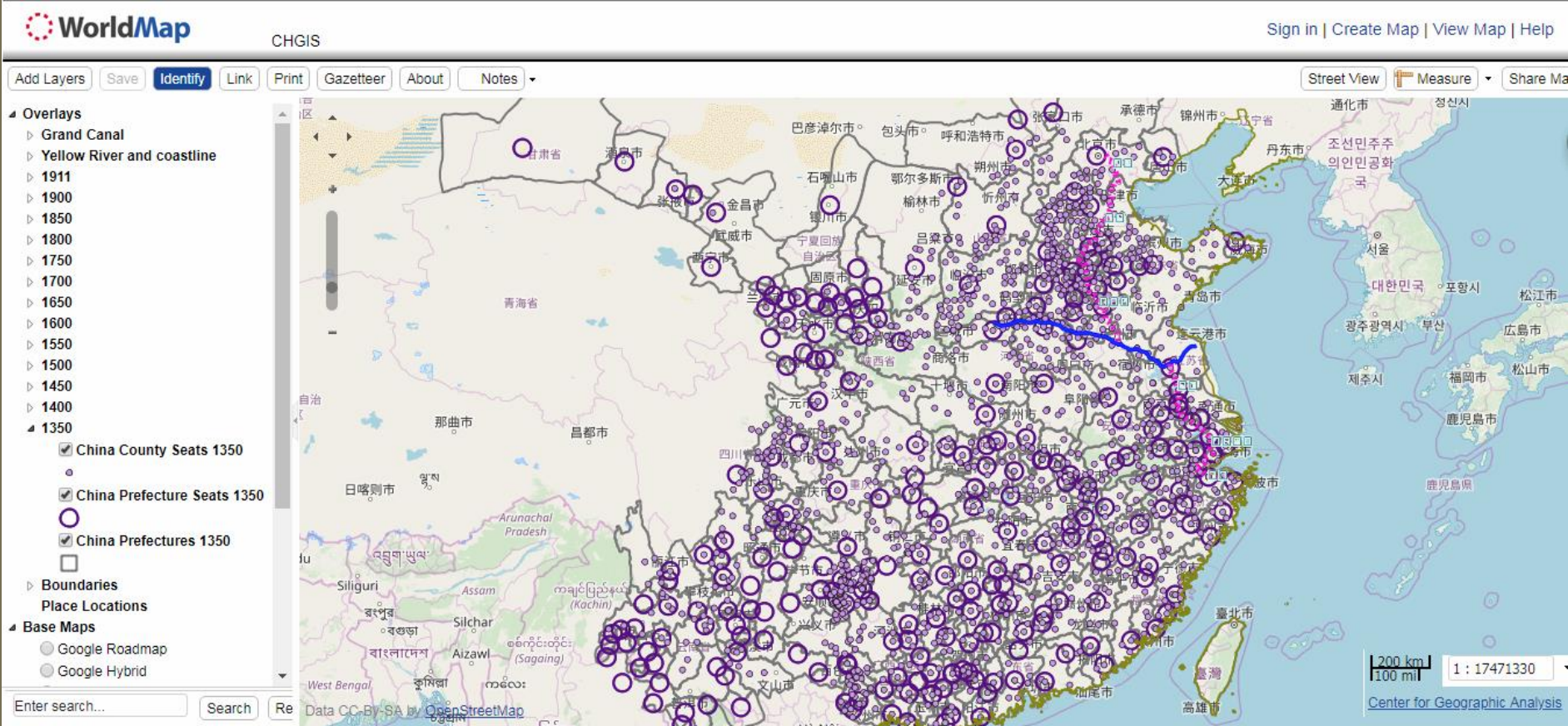
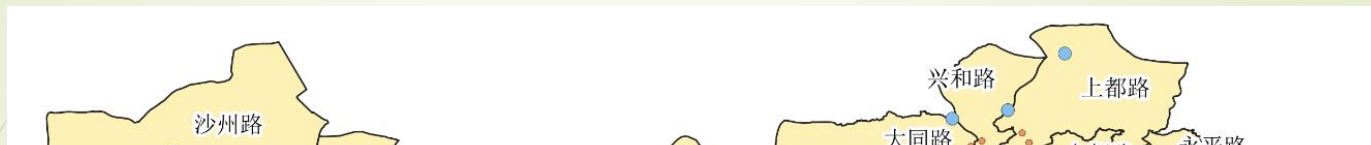
➤ 其他辅助工作

- 1) 郭书地名和CBDB的对比
- 2) 全宋诗目录整理

一、WorldMap

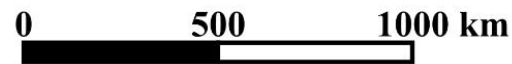
- 开始时间：1350年
- 截止时间：1900年+1911年
- 时间间隔：50年
- 数据来源：CHGIS V6
- 地图内容：府州级&县级行政单位分布
县级 (county)：点
府州级 (prefecture)：点&面
- 网页链接：
<http://worldmap.harvard.edu/maps/chgis>

样例



- county seats in 1350
- prefecture seats in 1350
- prefecture boundaries in 1350

吉安军 万安军



制图工具：QGIS

■ 制图方法：

- 1) 下载CHGIS Version 6 的数据
- 2) 在QGIS中利用“属性表” (attribute table) 中的利用公式筛选功能，选出对应年份存在的行政区数据
举例： "BEG_YR" <= 1350 and "END_YR" >= 1350
- 3) 将选出的数据另存为shp文件
- 4) 在“属性” (Properties) 中调整样式

■ QGIS的用途：

制图

空间分析

打开编码问题的csv、txt、xls文件

二、唐代地名Geocoding

邢云唐代地名的地理坐标查补

三、地理坐标的确认

3137个郭ID有地理坐标

- 郭书与CBDB数据存在对应关系。且CBDB ID有坐标

共有3250个郭ID没有地理坐标

- 郭书中有而CBDB中没有的地名
- 郭书与CBDB数据存在对应关系的地名，但CBDB没有相应的地理坐标

运用Python从TGAZ上抓取地理坐标 (by Yuying)

运用Regex从文本中抓取地理信息

TGAZ (在线数据)

➡ 网址: <http://maps.cga.harvard.edu/tgaz/>

TGAZ API

:: home

[intro](#)

[usage](#)

[examples](#)

[credits](#)

[cga](#)

The Temporal Gazetteer API is a webservice for querying historical placenames.

Introduction

TGAZ API - is a read-only interface designed to search the contents of the [China Historical GIS](#) placename database. The TGAZ system architecture -- based loosely on the CHGIS XML API (2006) -- has been normalized and made more generic in order to integrate multiple data sources, and to allow for vernacular spellings or transcription methods in many languages.

This interface accepts queries for placenames in UTF-8 encoded strings or glyphs (for example, Chinese Simplified or Complex characters). Each placename recorded in the database is referred to as a *spelling*. A brief look at the project in diagrams ([PDF](#)). Demonstration on [Youtube](#)

see also: [A Research Bibliography on Temporal Gazetteers](#)

Usage

Try this simple query form, or continue reading below for the API Usage Instructions:

Search for Placename 地名查询:

TGAZ API

place name 地名 ▶

朔州

year 年份 ▶

feature type 分类 ▶

data source 数据来源 ▶

any ▼

TGAZ (在线数据)

举例：
朔州

TGAZ API

TGAZ API

:: search results : name=朔州

Results: 1-13

Results: 1-13 of 13 total hits



TGAZ API URI :: http://maps.cga.harvard.edu/tgaz/placename/hvd_9

JSON XML RDF

placename:

朔州 (traditional Chinese)

朔州 (simplified Chinese)

Shuo Zhou (Pinyin)

type: 州 州 zhou sub-prefecture

temporal span: from 1369 to 1911

spatial info: POINT point N 39.30807 E 112.42152 (geo data source: FRC)

part of:

大同府 Datongfu from 1369 to 1726

朔平府 Shuopingfu from 1727 to 1911

sub units:

data source: CHGIS

source note: 朔州 (1369 ~ 1911年) 治所。编写者：王尚义、王杰瑜 明洪武区。民国元年更名朔县。注：⑤《明史》卷四十一《志》第十七《地理二》山平府：清初为右玉、左云、平鲁三卫。雍正三年，于右玉卫置府，并改三卫为隶。嘉庆元年，所领马邑省入为乡”。（中华书局P2037，1977年。）光绪

Copyright: 2019 © CHGIS, License: [CC BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

Published by: China Historical GIS [Harvard University and Fudan University]

[hvd_112886](#) 朔州

(Shuo Zhou) begin 319 CE end 350 CE [106.27810, 36.01409]

[hvd_113129](#) 朔州

(Shuo Zhou) begin 325 CE end 351 CE [109.21136, 40.64219]

[hvd_112886](#)

(Shuo Z

[hvd_113130](#) 朔州

(Shuo Zhou) begin 422 CE end 428 CE [109.46991, 36.59181]

[hvd_113129](#)

(Shuo Z

[hvd_113131](#) 朔州

(Shuo Zhou) begin 444 CE end 524 CE [111.82157, 40.37236]

[hvd_113130](#)

(Shuo Zhou)

[hvd_113134](#) 朔州

(Shuo Zhou) begin 422 CE end 428 CE [109.46991, 36.59181]

[hvd_113131](#) 朔州

(Shuo Zhou) begin 444 CE end 524 CE [111.82157, 40.37236]

[hvd_113134](#) 朔州

TGAZ (在线)

- 特点：可以——便于程
- 举例：朔州

网址形式：

TGAZ API URI :: http://maps.cga.harvard.edu/tgaz/placename/hvd_112886

[JSON](#) [XML](#) [RDF](#)

placename:

朔州 (traditional Chinese)

朔州 (simplified Chinese)

Shuo Zhou (Pinyin)

type: 州州 zhou region

temporal span: from 319 to 350

spatial info: POINT point N 36.01409 E 106.27810 (geo data source:)

part of:

[后赵 Houzhao](#) from 319 to 350

no
subunits

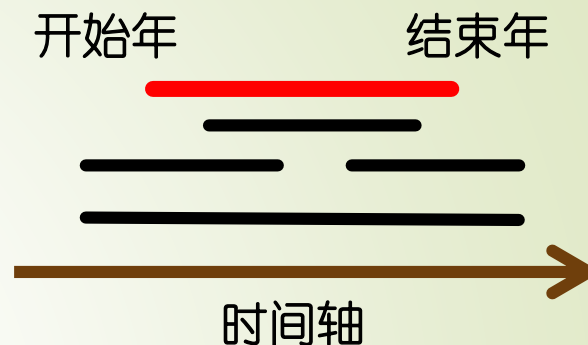
data source: CHGIS

```
{
  "system" : "CHGIS - Harvard University & Fudan University",
  "memo" : "Results for query matching key '朔州'",
  "count of displayed results" : "13",
  "count of total results" : "13",
  "placenames" : [
    {
      "sys_id" : "hvd_112886",
      "uri" : "http://maps.cga.harvard.edu/tgaz/placename/hvd_112886",
      "name" : "朔州",
      "transcription" : "Shuo Zhou",
      "years" : "319 ~ 350",
      "parent sys_id" : "hvd_113665",
      "parent name" : "后赵 (Houzhao)",
      "feature type" : "州 (zhou)",
      "object type" : "POINT",
      "xy coordinates" : "106.27810, 36.01409",
      "data source" : "CHGIS"
    },
    {
      "sys_id" : "hvd_113129",
      "uri" : "http://maps.cga.harvard.edu/tgaz/placename/hvd_113129",
      "name" : "朔州",
      "transcription" : "Shuo Zhou",
      "years" : "325 ~ 351",
      "parent sys_id" : "hvd_113665",
      "parent name" : "后赵 (Houzhao)",
      "feature type" : "州 (zhou)",
      "object type" : "POINT",
      "xy coordinates" : "109.21136, 40.64219",
      "data source" : "CHGIS"
    },
    {
      "sys_id" : "hvd_113130",
      "uri" : "http://maps.cga.harvard.edu/tgaz/placename/hvd_113130",
      "name" : "朔州",
      "transcription" : "Shuo Zhou",
      "years" : "422 ~ 428",
      "parent sys_id" : "hvd_113677",
      "parent name" : "夏 (Xia)",
      "feature type" : "州 (zhou)",
      "object type" : "POINT",
      "xy coordinates" : "109.21136, 40.64219",
      "data source" : "CHGIS"
    }
  ]
}
```

TGAZ (在线数据)

检索逻辑：

- 1) 地名匹配
- 2) 年份上有交叉或者是包含关系
- 3) 只保留point类型的结果



储存信息：

- 1) 原始数据的信息
- 2) TGAZ中获得的数据

多个检索结果的保存

——使用id_x的方式，显示是检索结果的第几条

	ID	NAME	FIRST	LAST	count	source	name_found	latitude	longitude
205_0	205	終南	619	634	0				
210_1	210	始平	618	709	2	tgaz	始平县	34.2987	108.4826
210_2	210	始平	618	709	2	tgaz	始平县	27.9047	115.28114
225_1	225	扶風	620	627	1	tgaz	扶风郡	34.2916	107.83877

DDBC (本地文件/数据)

- DDBC (Dharma Drum Buddhist College)
- 网址: <http://authority.dila.edu.tw/place/>
- 法鼓文理學院

地名規範檢索

Buddhist Studies Place Authority
Databases
(Beta Version)

歡迎來到地名規範資料庫

地名共 58902 筆, 地名群組共 93 筆



地名檢索字串(Name)

朔州

☒ 部份符合 ☐ 完全符合
☐ 進階選項

編號搜尋(Authority id)

中國歷史行政地名(行政疆域)
之時空基礎架構系統, 第一版,
台灣, 2002.」

參考書目
回首頁

地名規範檢索

Buddhist Studies Place Authority Databases
(Beta Version)

地名檢索字串(Name)

朔州

查詢

☒ 部份符合 ☐ 完全符合
☐ 進階選項

編號搜尋(Authority id)

查詢

中國歷史行政地名(行政疆域)資料來源:「中華文明之時空
基礎架構系統, 第一版, 中央研究院, 台北, 台灣,
2002.」

參考書目
回首頁

*Z碼系統使用GRS87經緯度座標系統, 與GPS或Google Earth使用WGS84經緯度座標系統略有不同, 有需要可利用「Web版坐標轉換程式」進行座標轉換

檢索【朔州】(地名: 9 筆, 群組: 0 筆):

- 。朝代過濾器(不含群組): 顯示全部
- 。檢視所有結果於: [Google Map](#) 或 [下載KML](#)

朔州(shuò zhōu)



規範碼: PL000000006308

舊規範碼: CN0150123Z13AA(僅供參考請勿使用)

地名分類: 中研院歷史地名

經緯度: 緯度: 40.4506 經度: 111.8203(精確值) [download KML](#) [Map](#)

行政區: 中國-內蒙古自治區-呼和浩特市-和林格爾縣

屬: 朔州

備註: (420 ~ 589) 州級行政中心所在地。

朝代: 北魏

註解: 朔州(本漢五原郡, 延和二年置為鎮, 後改為懷朔, 李昌中改為州。後陷, 今寄治并州界。)領郡五, 縣十三(魏書, 卷一百六上。地形二上: 28)

[問題回報\(Feedback\)](#)

太原市



別名:

[正體中文]并州(bìng zhōu),朔州(shuò zhōu),并部

DDBC (本地文件/数据)

- ➡ 数据提取：从htm文件提取数据到csv
- ➡ htm文件：

琅邪

別名：

[正體中文]琅琊(láng yá)

規範碼：PL000000020554

舊規範碼：CN0370284Z08AA(僅供參考請勿使用)

地名分類：中研院歷史地名

經緯度：緯度：35.7035經度：119.8554(精確值) [download KML](#) [Map](#)

行政區：中國-山東省-青島市-黃島區

屬：琅邪郡

備註：(-220 ~ -206) 郡級行政中心所在地。

朝代：秦

註解：秦始皇帝置，治琅邪縣（今山東膠南市琅邪台西北）。（<http://zh.wikipedia.org/琅邪郡>，2011.12.09）

Occurs in：

[g009p0523\(普陀洛迦新志\)](#) { 琅琊 }

DDBC (本地文件/数据)

恆州(héng zhōu)

 引用網址

別名：

[正體中文]鎮州(zhèn zhōu),常山郡(cháng shān jùn),成德軍節度使(chéng dé jūn jié dù shǐ)

規範碼：PL000000000871

舊規範碼：CN0130123Z21AA(僅供參考請勿使用)

地名分類：中研院歷史地名

經緯度：緯度：38.1394經度：114.5733(精確值) [download KML](#) [Map](#)

行政區：中國-河北省-石家莊市-正定縣

屬：河北道

備註：(618 ~ 907) 府級、州級、郡級行政中心所在地。

朝代：唐

註解：歷代為常山郡。周、隋改為恆州，後廢。義旗初，復置恆州，領真定、石邑、行唐、九門、滋陽五縣，州治石邑。武德元年，陷竇建德。四年，賊平，徙治所於真定，省滋陽縣，又割廉州之冀城來屬。天寶元年，改為常山郡。乾元元年，復為恆州。興元元年，昇為都督府。元和十五年，改為鎮州。(<http://zh.wikisource.org/zh-hant/舊唐書/卷39> ; <http://baike.baidu.com/view/804670.htm> , 2012.02.07)

Occurs in :

[T50n2060_p0507c20](#) 唐高僧傳: 釋僧朗傳

[T50n2060_p0615c11](#) 唐高僧傳: 釋法礪傳

[T50n2060_p0669c04](#) 唐高僧傳: 釋道璩傳

[T50n2060_p0670a01](#) 唐高僧傳: 釋僧蓋傳

[T50n2060_p0672c29](#) 唐高僧傳: 釋靈達傳

[T50n2061_p0722b01](#) 宋高僧傳: 般若傳

[T50n2061_p0775c13](#) 宋高僧傳: 從諗傳

[T50n2061_p0873c25](#) 宋高僧傳: 自覺傳{ 鎮州 }

[g009p0362\(普陀洛迦新志\)](#) { 鎮州 }

[g009p0362\(普陀洛迦新志\)](#) { 鎮州 }

[g032p0309\(徑山志\)](#) { 鎮州 }

总ID	link▶	地名	別名	規範碼	舊規範碼	地名分類	緯度	經度	行政區	行政級別	屬	朝▶年代	註解	引用位置	URL
44	0	蜀郡		PL0000000034783	CN0510108Z09AA	中研院▶	30.6501	104.0864	中國-四川省-成都市-成華區	郡級行政	蜀郡	秦 -220 ~ ->	周赧王元年 (前314) , 秦惠王封子通		
44	1	蜀郡		PL0000000034783	CN0510108Z09AA									g084p0202(雞足山志) http://	
44	2	蜀郡		PL0000000034783	CN0510108Z09AA									g049p0290(峨眉山志) http://	
45	0	郿道		PL0000000037061	CN0511822Z08AA	中研院▶	29.7951	102.8456	中國-四川省-雅安市-榮經縣	郡級行政	徙郿	秦 -220 ~ -206			
46	0	漢中郡		PL0000000043872	CN0610702Z31AA	中研院▶	33.0761	107.0439	中國-陝西省-漢中市-漢台區	郡級行政	漢中郡	秦 -220 ~ ->	漢中郡, 秦惠文王后十三年 (前312)		

DDBC（本地文件/数据）

经纬度获取

检索逻辑（类TGAZ）

① 年代筛选：只在DDBC唐代的数据中检索

② 名称对应：使用纯字符包含关系

——“地名” VS “地名” & “别名”

③ 从属关系（道）对比

结果保存

——形式类似TGAZ

高德地图（现代地图）

- ➡ 数据特点：现代地址的检索

- ➡ 前提：知道对应的现代地址

- ➡ 检索方式

- 1) 结构化数据：国家-省-市-区县-街道号码

- 2) 利用高德地图API，批量输入地址，获取地址的经纬度信息

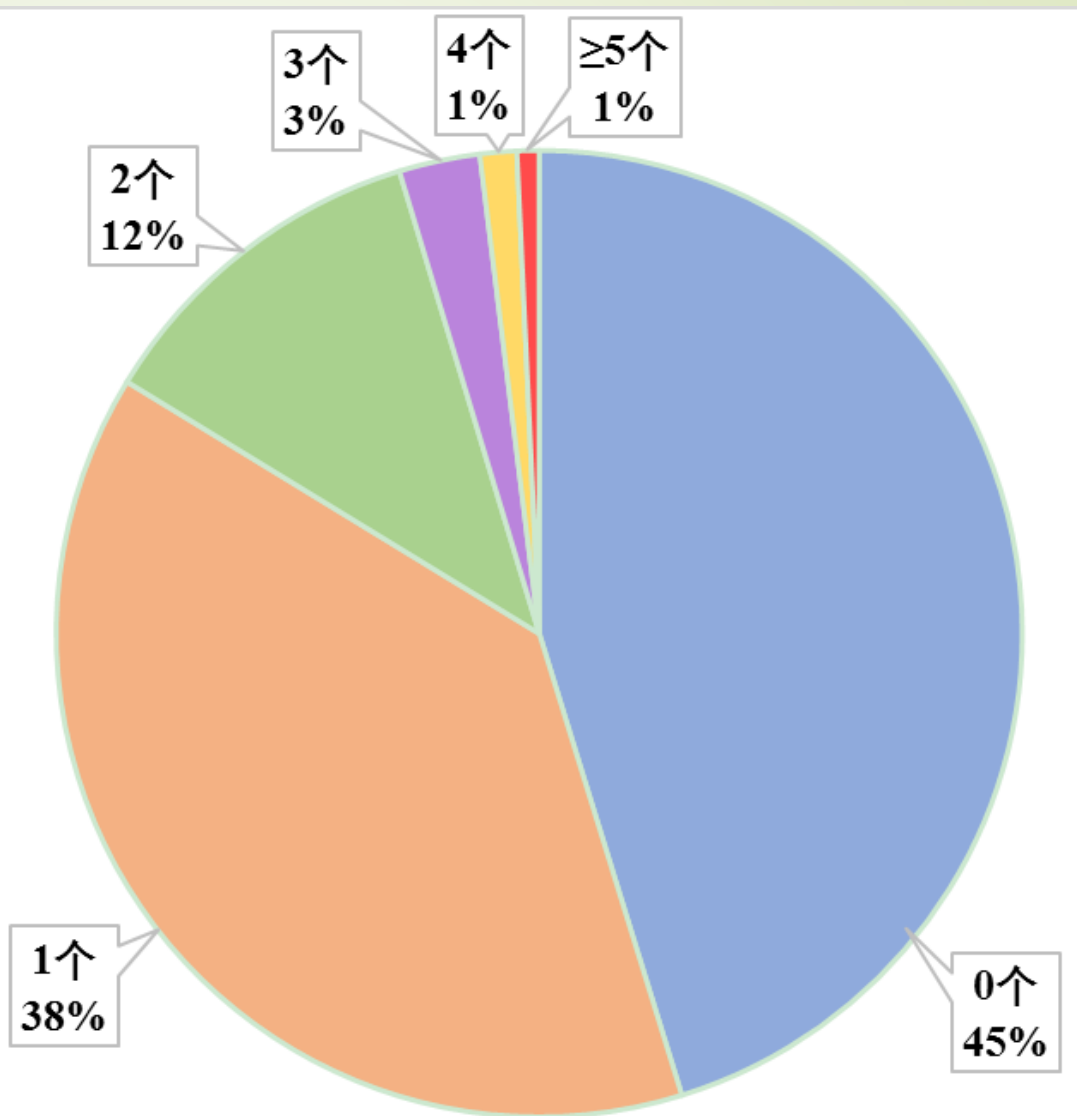
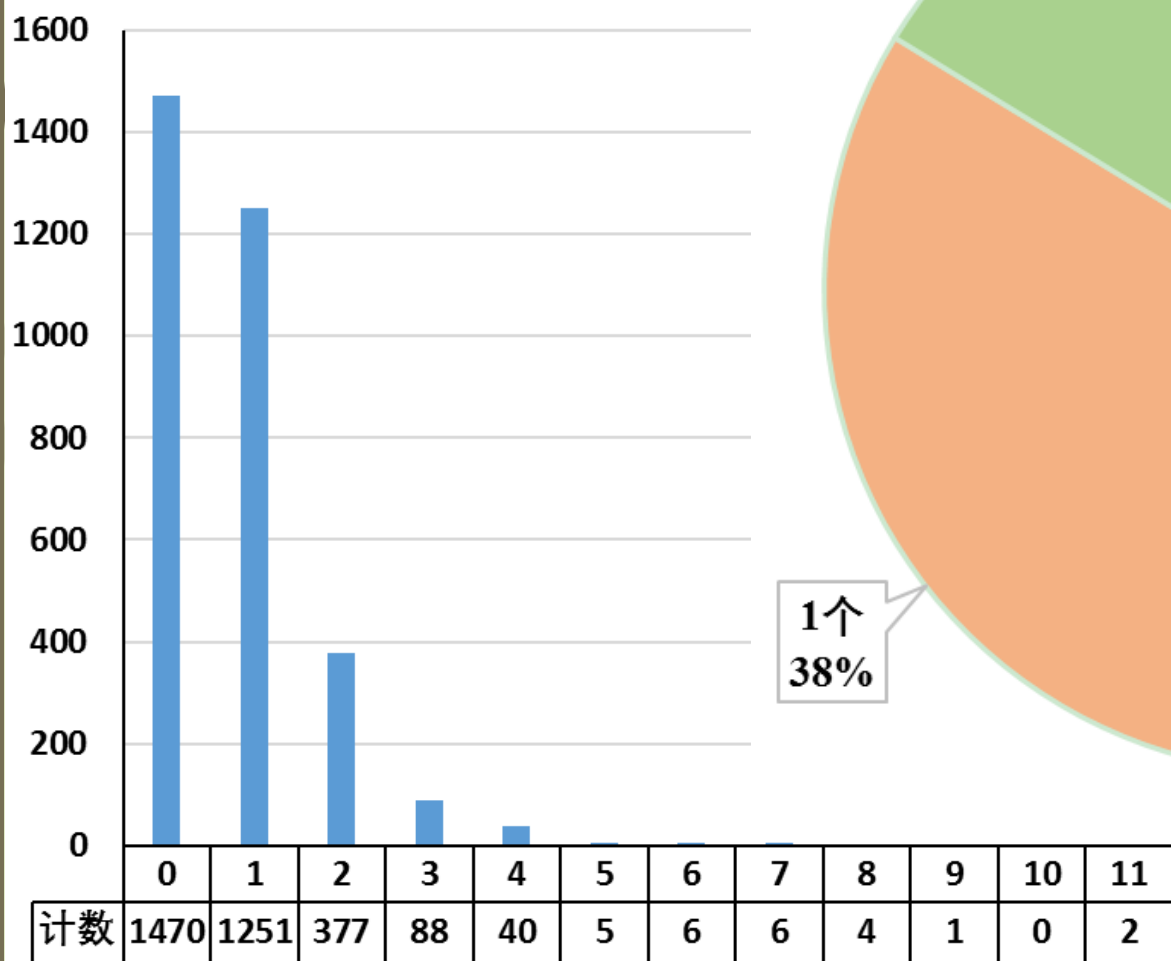
- ➡ 其他同类数据源

Google Map、百度地图等

检索结果

TGAZ: 1780

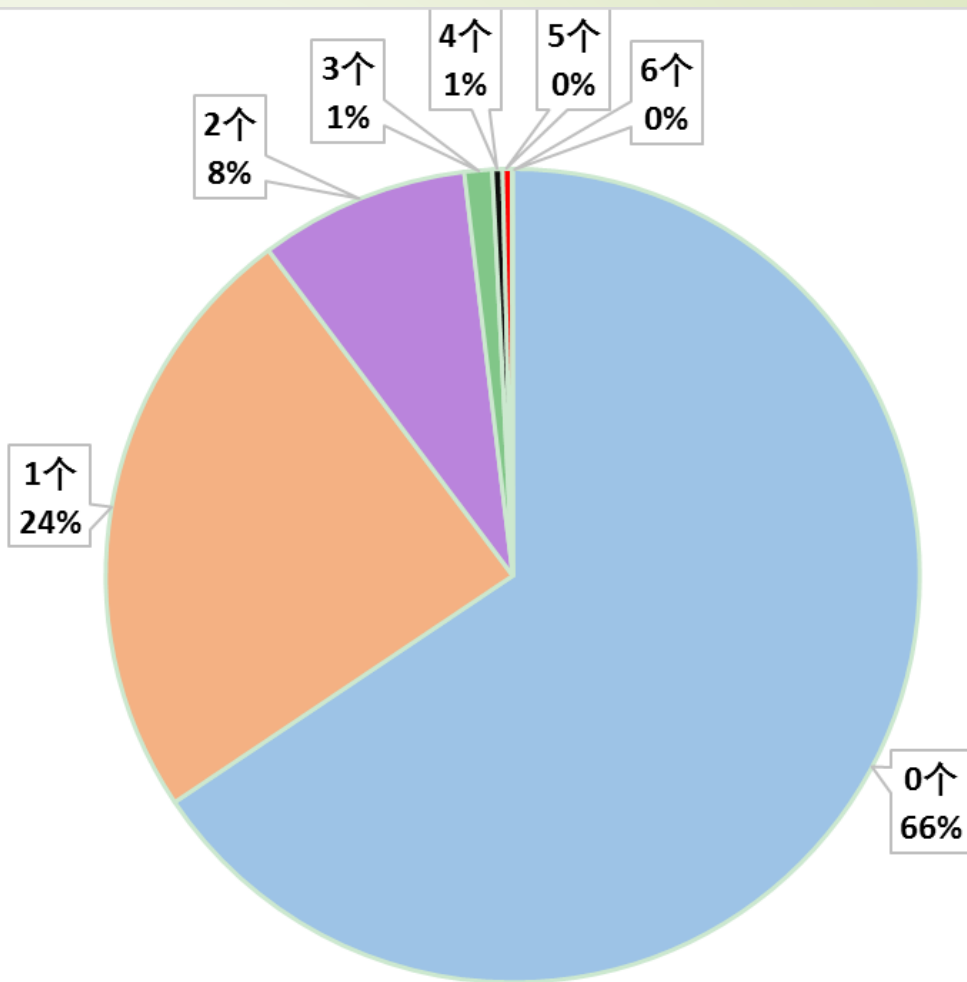
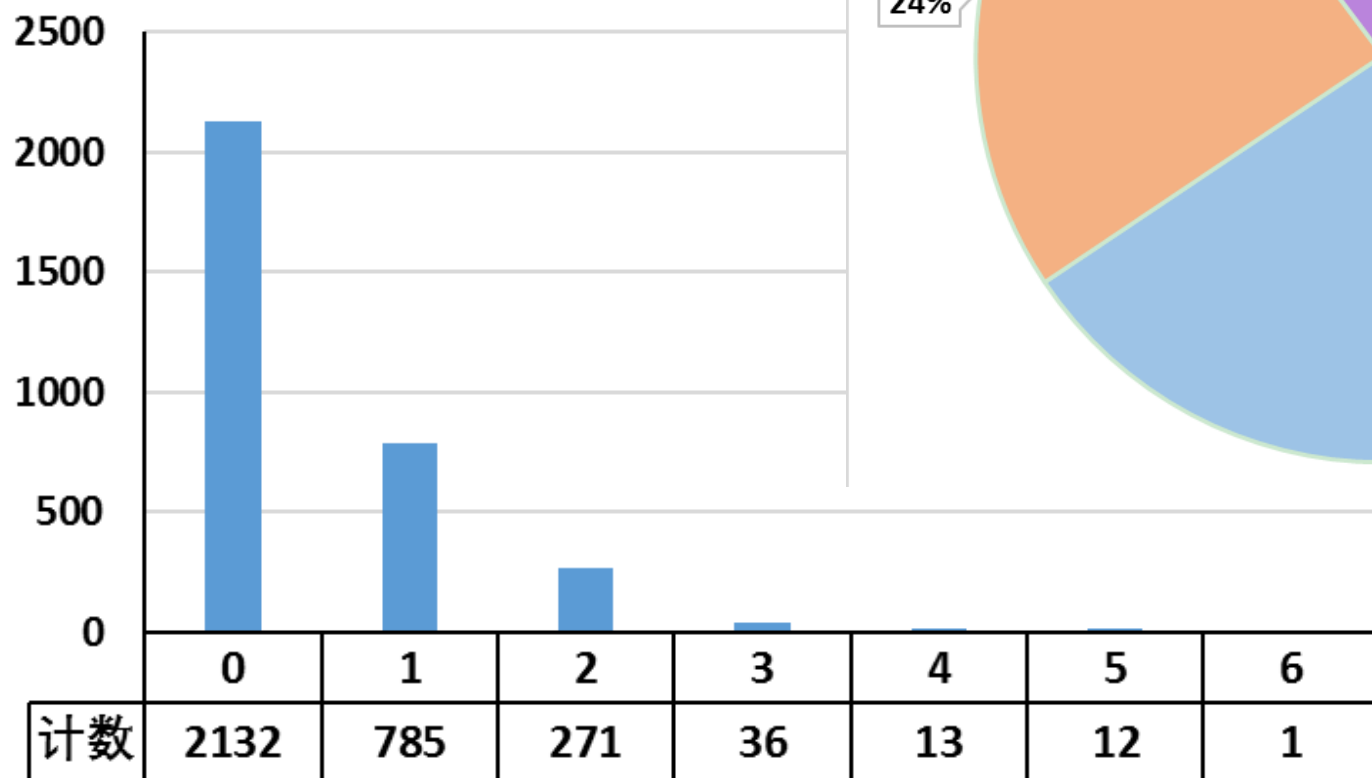
Tgaz地名匹配结果



检索结果

DDBC: 1118

DDBC 地名匹配结果



检索结果

➡ **≥1054条无匹配**



三、辅助工作

1) 唐代地名匹配 (郭书 vs CBDB)

➤ 筛选逻辑:

① 层级一致

② 名称匹配

③ 所属道相同

—— “關內道” & “關内道”

④ 最佳年代匹配

—— 起止年相差的绝对值之和最小

➤ 人工判断、数据导入

2) 全宋诗目录整理

➤ 使用regex对“册”、“卷”、“作者”、“诗题”进行分别标注

➤ 人工检查, 修正错误

总结

➤ 文本录入规范化

- ①统一繁简体、编码（文件名注明GBK, UTF-8）
- ②英文大小写一致
- ③统一体例

举例：①多个别名用\$区分；②无内容单元格，都空着或都填blank——不影响文本含义&编程识别的符号

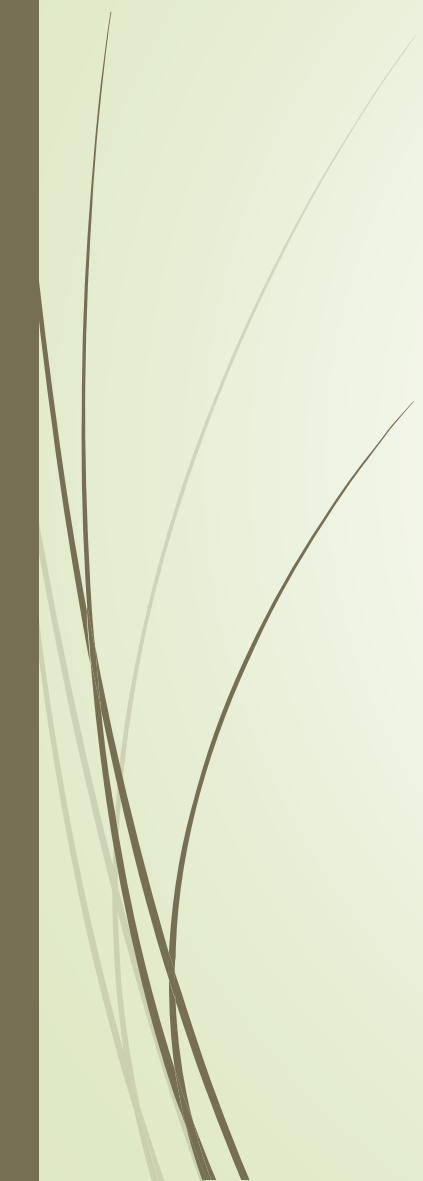
➤ 资料录入时标明data来源，便于反查

➤ 说明文件

- 录入者对数据特殊之处/数据特点/数据格式的说明
- 数据修正的记录

➤ 编程的使用

- 1) 有逻辑性、重复性工作（regex）
- 2) 大批量数据的分析：整体性分析、具体问题的辅助分析



谢谢