

# Prosopographical Databases, Text-Mining, GIS and System Interoperability for Chinese History and Literature

Peter Bol

Jieh Hsiang

Grace Fong

# 中國歷代人物傳記資料庫

中央研究院歷史語言研究所、哈佛大學、北京大學中國古代史研究中心合作開發

Developed through collaboration among Academia Sinica, Harvard University, and Peking University

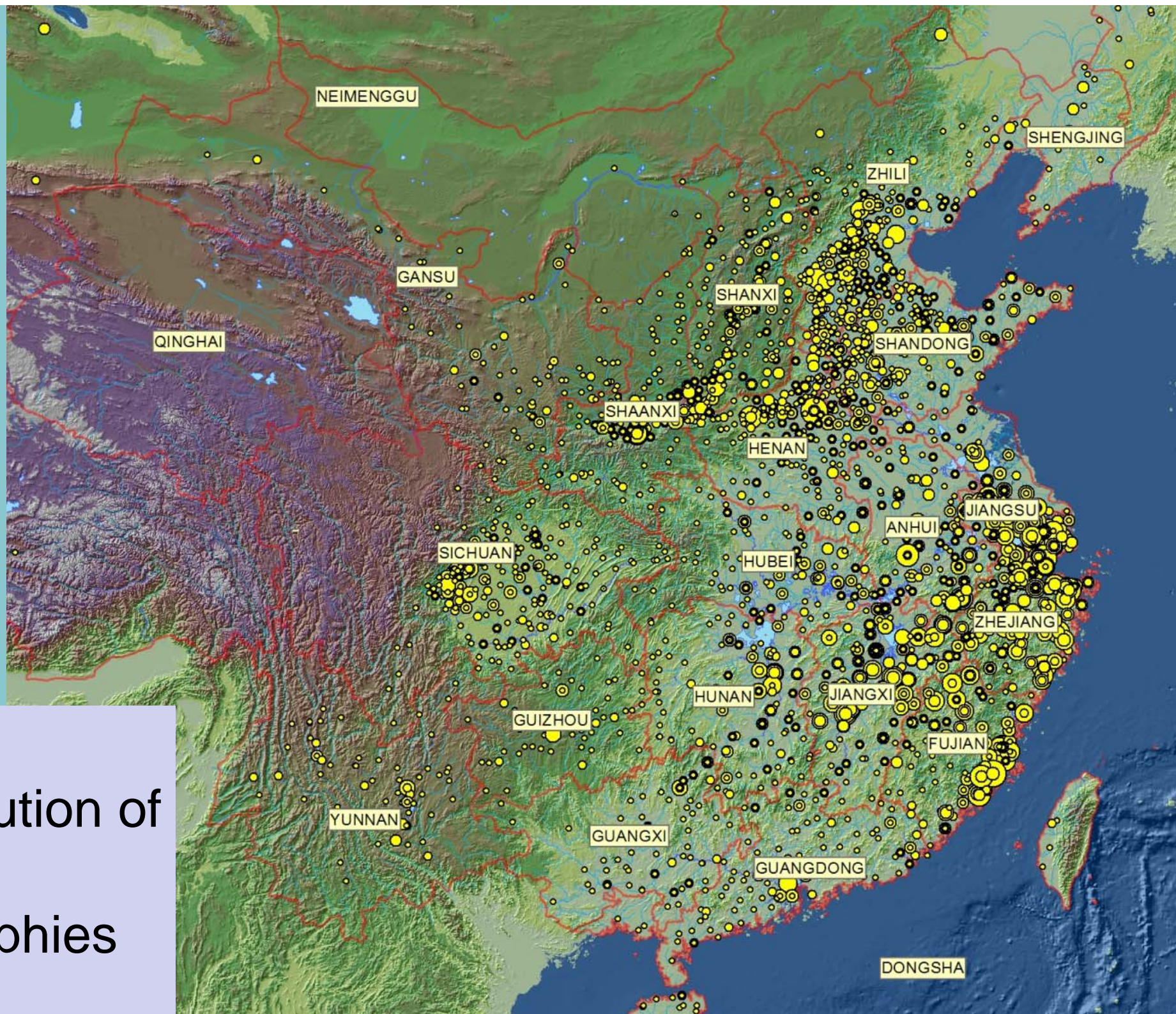


## The China Biographical Database –

*Modeling Life Histories – from anecdote to data*

**Biography**  
**Prosopography**  
**Social Network Analysis**  
**Geospatial Analysis**

Spatial  
Distribution of  
CBDB  
Biographies  
210217



# CBDB's origins and prospects

- Begun by Robert Hartwell at the University of Pennsylvania in the 1980s. Prof. Hartwell left his digital projects to the Harvard Yenching Institute on his death in 1996.
- In 2004-05 Michael Fuller at UC Irvine revised and modernized the database structure
- In 2005 CBDB became a collaboration between Harvard, The Institute of History and Philology (IHP) of Academia Sinica, and the Center for Research on Ancient Chinese History at Peking University
  - The PKU editors took on the task of emending the Hartwell database, which then had data on about 25,000 individuals.
  - IHP provided a digital edition of Wang Deyi's *Index to Song Biographical Materials* (王德毅：宋人傳記資料索引（電子版）)
  - IHP later provided Wang's Yuan dynasty index and digital texts of the dynastic histories and other texts
- 2008 The first international workshop on Chinese biographical data held at Harvard.
- 2008-2011 Harvard and thereafter the NEH funded the development of text-mining texts for Chinese biographical texts in collaboration with Stuart Shieber, Rani Nelken, and Elif Yamangil
- 2011-2014 The Canadian Social Science and Humanities Research Council funds the MQWW-CBDB editorial group at PKU
- 2012-2015 The Luce Foundation provides a sustaining grant to CBDB

# Biography in China

- China has over 2000 years of elite biography
- There are perhaps 250,000 biographies in the historical record: including biographies in dynastic histories and local gazetteers and grave biographies
- Local gazetteers (Song-Qing) mention at least 2 million people
- The tradition of compiling genealogies in China is unique in world history
- Chinese history is an important source for the history of humanity (we can know more about more people, over more territory, over a longer period)

# How CBDB data is selected

- Names——人名
- Time——時間
- Places——地址
- Offices——職官
- Modes of entry into office——入仕途徑
- Writings——著作
- Social Distinctiveness——社會區分
- Kinship——親屬關係
- Social associations——社會關係
- Possessions——財產
- Events——事件

Lü Zuqian, whose style name was Bogong, was a grandson of the Right Assistant Director to the Imperial Secretary Haowen. His family lived in Wuzhou beginning in his grandfather's generation. The learning of Zuqian was based on family tradition, and embodied the textual transmission from the Central Plain of the north. When he grew up, Zuqian studied with Lin Zhiqi, Wang Yingchen, and Hu Xian. He also was friends with Zhang Shi and Zhu Xi, and thereby his understanding gained in clarity.

At first he obtained official rank by way of the protection privilege but later he obtained his Presented Scholar degree and also passed the special decree examination for "Erudite Learning and Exceptional Literary Composition." Then he was appointed as the Instructor at School for the Imperial Clan in the Southern Outer Office of the Hostel for the Imperial Clan.

呂祖謙字伯恭，尚書右丞好問之孫也。自其祖始居婺州。祖謙之學本之家庭，有中原文獻之傳。長從林之奇、汪應辰、胡憲游，既又友張栻、朱熹，講索益精。初，蔭補入官，後舉進士，復中博學宏詞科，調南外宗正。

“Factoids” in biographical texts

c_role
c_year
c_nh_code
c_nh_year
c_yr_range

<b>EVENT_CODES</b>
c_event_code
c_event_name_chn
c_event_name
c_fy_yr
c_ly_yr
c_fy_nh_code

<b>ADDRESSES</b>
c_addr_id
c_addr_cbd
c_name

tts_sysno
c_personid
c_addr_id
c_addr_type
c_sequence

<b>KIN_DATA</b>
tts_sysno
c_personid
c_kin_id
c_kin_code
c_kin_rel

<b>KINSHIP_CODES</b>
c_kincode
c_kin_pair1
c_kin_pair2

type_id
---------

<b>EVENTS_ADDR</b>
c_event_record_id
c_personid
c_addr_id
c_year
c_nh_code
c_nh_year

<b>OFFICE TYPES</b>
type_id
type_attr
type_attr_chn
type_desc
type_desc_chn

<b>OFFICE CODE_TYPE REL</b>
sys_id
code_attr
code_attr_chn
code_id
type_id
c_office_type_type_code

<b>BIOG_MAIN</b>
tts_sysno
c_personid
c_name
c_name_chn
c_jia
c_zu
c_index_year
c_female
c_ethnicity_code
c_tribe
c_birthyear
c_by_nh_code
c_by_nh_year
c_by_range
c_deathyear
c_dy_nh_code
c_dy_nh_year
c_dy_range
c_death_age
c_death_age_approx
c_fl_earliest_year
c_fl_ey_nh_code
c_fl_ey_nh_year
c_fl_ey_notes
c_fl_latest_year
c_fl_ly_nh_code
c_fl_ly_nh_year
c_fl_ly_notes
c_surname
c_surname_chn
c_surname_code
c_mingzi
c_mingzi_chn
c_dy

<b>STATUS_DATA</b>
tts_sysno
c_personid
c_sequence
c_status_code
c_firstyear

<b>STATUS_CODES</b>
c_status_code
c_status_desc
c_status_desc_chn

<b>OFFICE_DATA</b>
c_office_id
c_office_category_id
c_reports_to

<b>OFFICE_CODES</b>
tts_sysno
c_office_id
c_office_pinyin
c_office_chn
c_office_pinyin_alt
c_office_chn_alt
c_office_trans

<b>POSTING_DATA</b>
c_personid
c_posting_id

<b>ENTRY_DATA</b>
tts_sysno
c_personid
c_entry_code

<b>ENTRY_CODES</b>
c_entry_code
c_entry_desc
c_entry_desc_chn
c_entry_weight
c_sortorder

<b>ENTRY CODE_TYPE REL</b>
code_id
type_id

type_id
type_a
type_b
type_d

<b>POSTED_TO_ADDR D...</b>
tts_sysno
c_posting_id
c_personid
c_office_id

<b>POSTED_TO OFFICE D...</b>
tts_sysno
c_personid
c_office_id
c_posting_id
c_sequence
c_firstyear
c_fy_nh_code
c_fy_nh_year
c_fy_range
c_lastyear
c_ly_nh_code
c_ly_nh_year
c_ly_range
c_appt_type_code
c_assume_office_code
c_source

<b>ADDR_CODES</b>
c_addr_id
c_addr_cbd
c_name
c_name_chn
c_firstyear
c_lastyear

<b>ADDR BELONGS DATA</b>
c_addr_id
c_belongs_to
c_addr_cbd
c_firstyear
c_lastyear
c_editor

<b>APPOINTMENT_TYPE_CODES</b>
c_appt_type_code
c_appt_type_desc_chn
c_appt_type_desc

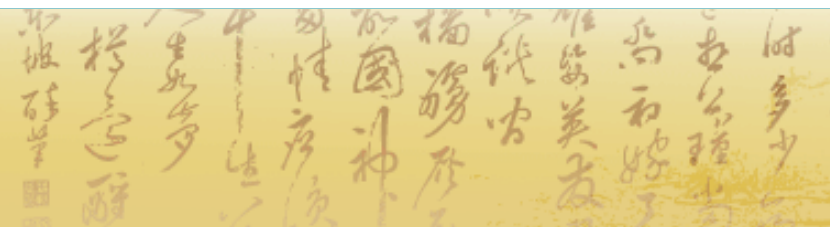
<b>DYNASTIES</b>
c_dy
c_dynasty
c_dynasty_ch

<b>TEXT_CODES</b>
tts_sysno
c_textid
c_title_chn
c_suffix_version

<b>FIX_AUTHORS</b>
c_textid
c_personid
c_name

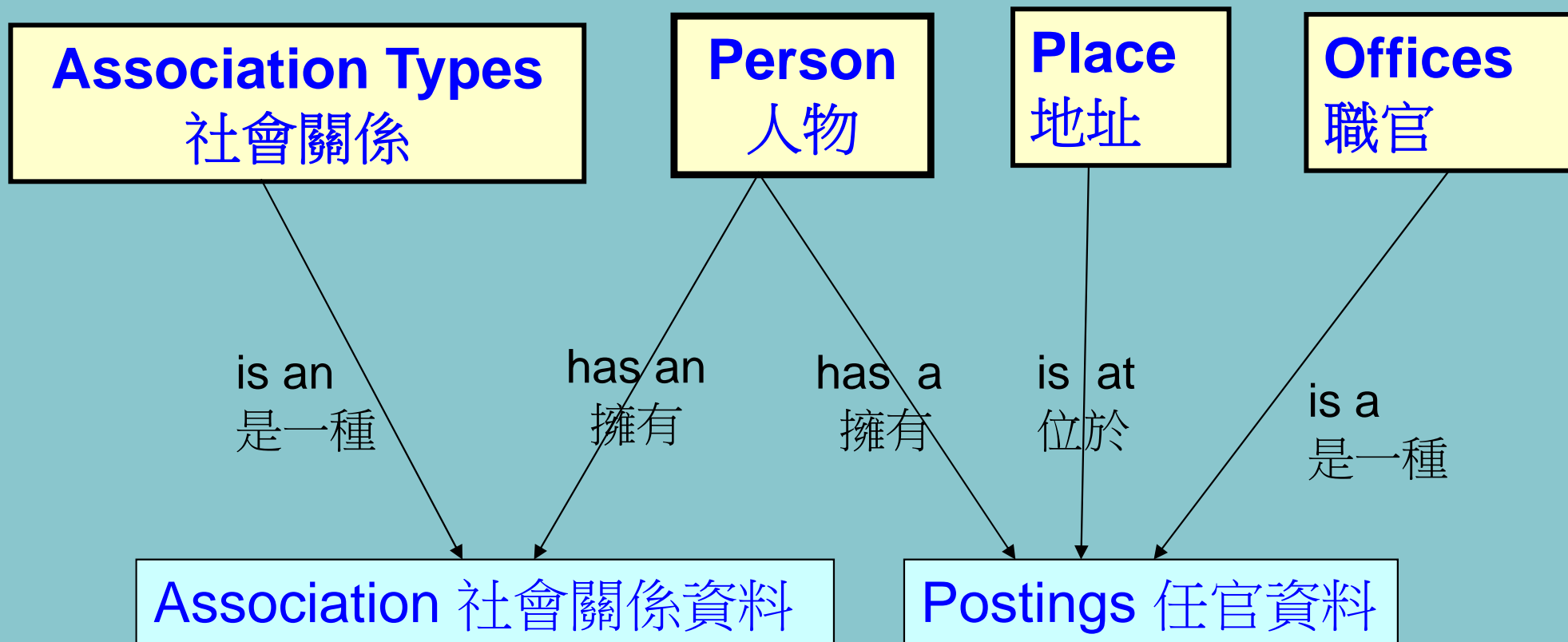
<b>OFFICE_CATEGORIES</b>
c_office_category_id

<b>CHORONYM_CODES</b>
-----------------------



## Entity Relations Modeling: Abstracting the features of the Biographical World

### 實體關係建模： 將傳記資料的特徵抽象化



Lü Zuqian, whose style name was Bogong, was a grandson of the Right Assistant Director to the Imperial Secretary Haowen. His family lived in Wuzhou beginning in his grandfather's generation. The learning of Zuqian was based on family tradition, and embodied the textual transmission from the Central Plain of the north. When he grew up, Zuqian studied with Lin Zhiqi, Wang Yingchen, and Hu Xian. He also was friends with Zhang Shi and Zhu Xi, and thereby his understanding gained in clarity.

At first he obtained official rank by way of the protection privilege but later he obtained his Presented Scholar degree and also passed the special decree examination for "Erudite Learning and Exceptional Literary Composition." Then he was appointed as the Instructor at School for the Imperial Clan in the Southern Outer Office of the Hostel for the Imperial Clan.

呂祖謙字伯恭，尚書右丞好問之孫也。自其祖始居婺州。祖謙之學本之家庭，有中原文獻之傳。長從林之奇、汪應辰、胡憲游，既又友張栻、朱熹，講索益精。初，蔭補入官，後舉進士，復中博學宏詞科，調南外宗教。

Query by person, place, kin, association, office, and various combinations

# CHINA BIOGRAPHICAL DATABASE PROJECT (CBDB)

## 中國歷代人物傳記資料庫

Enter Biographical Data 輸入傳記資料

### Retrieve Data 數據查詢

Look Up Data on an Individual 按人查詢

Query by Methods of Entry into Government 按入仕途徑查詢

Query Office Holding 官職查詢

Query Kinship 查詢親屬關係

Query Associations 查詢社會關係

Query Social Networks 查詢社會關係網絡

Query Pair-wise Associations 查詢兩人之間社會關係

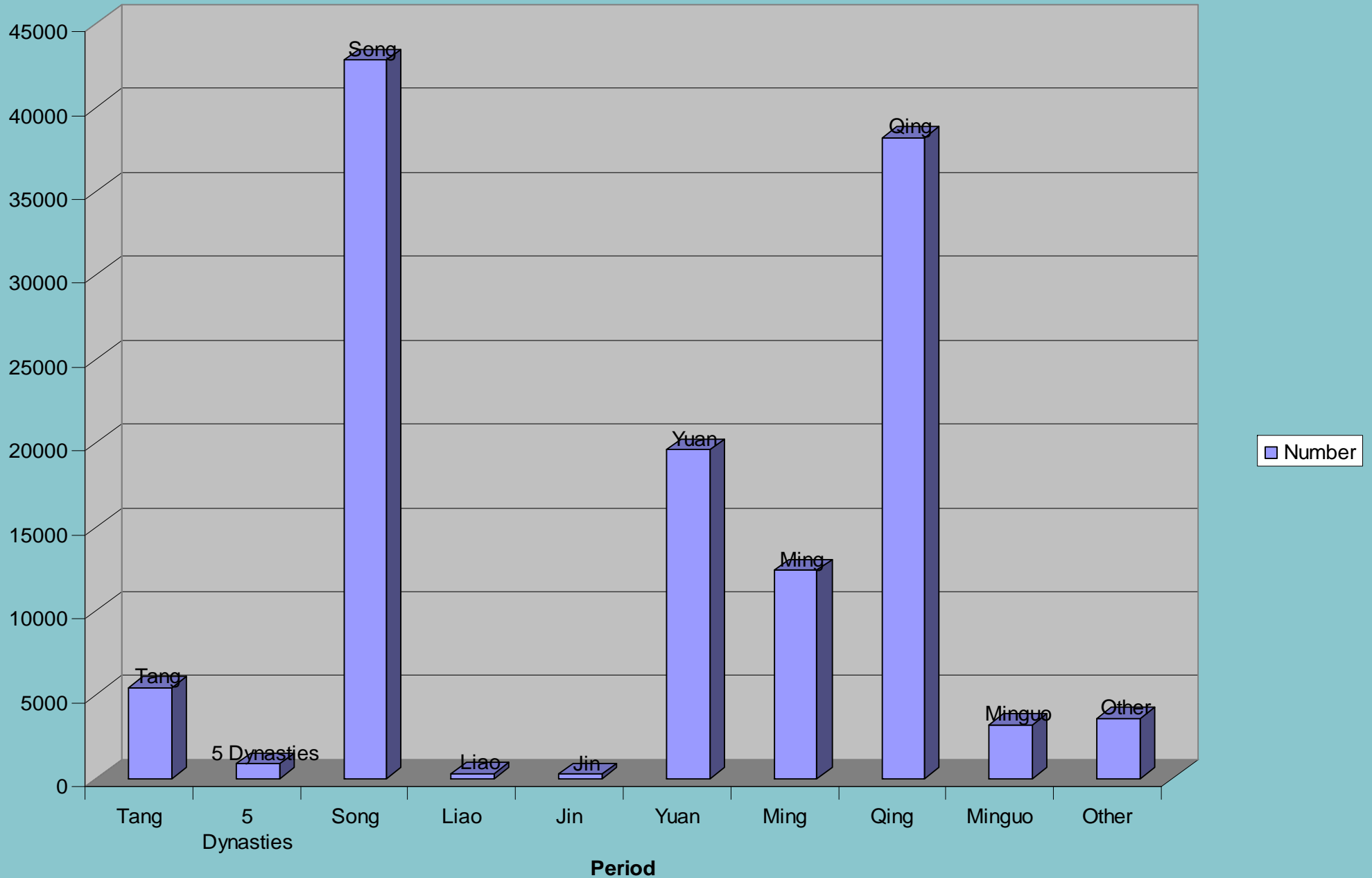
Exit 退出

Desktop Version

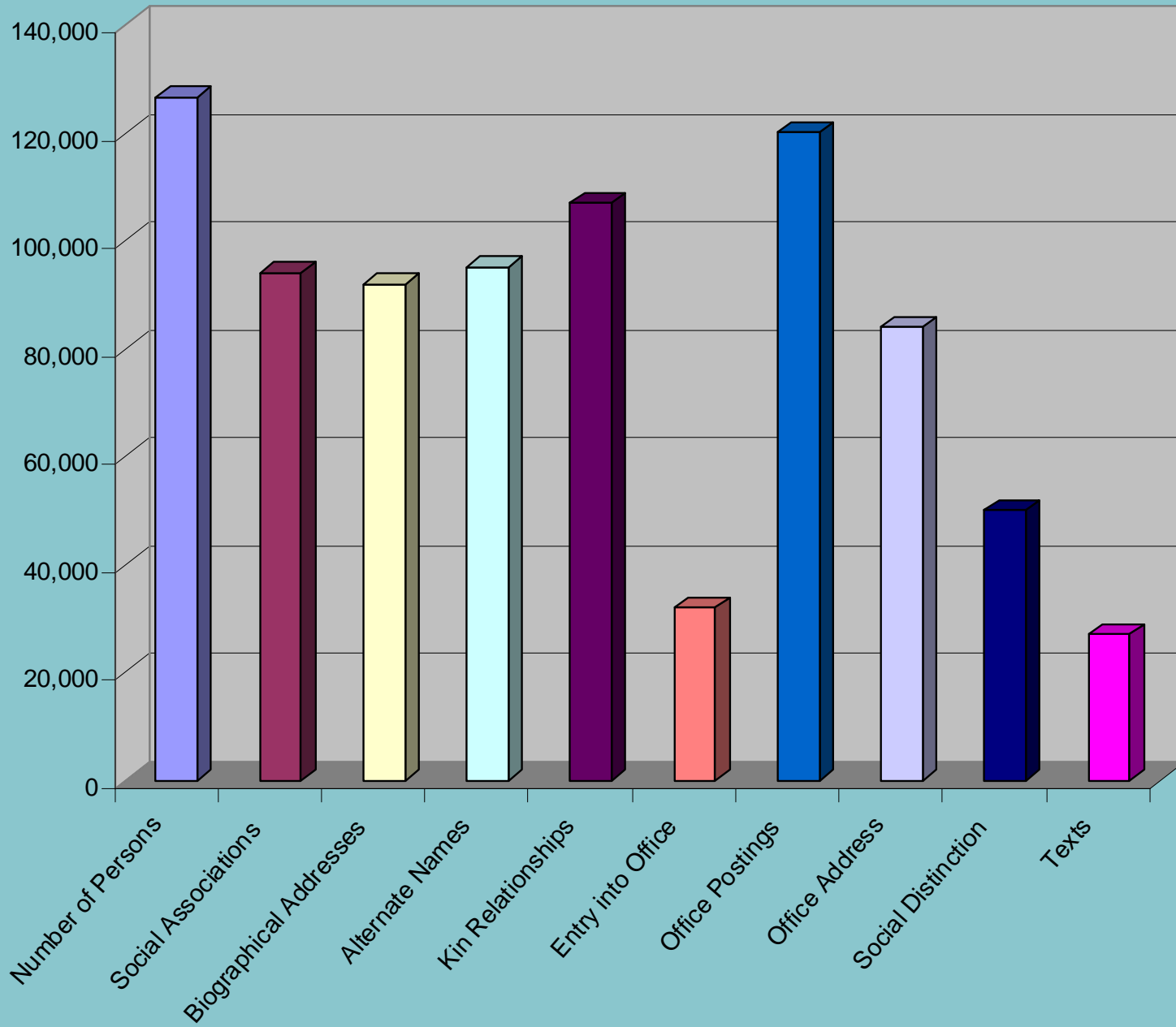
Last update: June 28 2011

單機版 2011年6月28日最後更新

# Current Contents



信息



- Number of Persons
- Social Associations
- Biographical Addresses
- Alternate Names
- Kin Relationships
- Entry into Office
- Office Postings
- Office Address
- Social Distinction
- Texts

CBDB

# **OUTPUTS FOR STATISTICAL ANALYSIS, SNA, GIS**

# CHINA BIOGRAPHICAL DATABASE PROJECT (CBDB)

## 中國歷代人物傳記資料庫

Enter Biographical Data 輸入傳記資料

### Retrieve Data 數據查詢

Look Up Data on an Individual 按人查詢

Query by Methods of Entry into Government 按入仕途徑查詢

Query Office Holding 官職查詢

Query Kinship 查詢親屬關係

Query Associations 查詢社會關係

Query Social Networks 查詢社會關係網絡

Query Pair-wise Associations 查詢兩人之間社會關係

Exit 退出

Desktop Version

Last update: June 28 2011

單機版 2011年6月28日最後更新

Select Person

Import People

All People

Select Place

Import Places

All Places

 Restrict to PlaceFrom: To:  Use Index Year Include Male Include FemaleMax Node Dist. Max Loop # 

Run Query

Save to UCINET

Save to Pajek

Save to Pajek (漢字)

Save to GIS

Save to GUESS

Exit

简体

繁體

To save the data to the Clipboard, click on the square in the upper left corner to select all the records, then right-click on it and copy.

Select Types of Relations

Social Network Relationship

People Participating in the Social Network

 Include Non-Kin  Select All Friendship Family Religion Finance Medicine Military Supportive Oppositional Scholarship Teacher-Student Scholarly Affiliation Scholarly Topic Common Membership Academic Patronage Literary / Artistic Affiliations Intellectual Attacks Politics Equal Relation Subordinate Relation Superior Relation Supportive Association Recommendation / Sponsorship Opposition Writings Commemorative Texts Epitaphs Prefaces / Postfaces Ritual Texts Biographical Texts Explanatory Texts Admonitory Texts Correspondences Texts for Social Occasions Include KinMax Generation UP Max Generation Down Max Collateral Links Max Marriage Links  Use Kinship Parameters UTF-8  Big-5  GB UTF-8  GB

# China Biographical Database Project (CBDB)

哈佛大学、北京大學中國古代史研究中心合作開發  
in collaboration with Academia Sinica, Harvard University, and Peking University

Search by Name | Search by Address | Search by Office | **Keyword Search** | Advanced Search | Kinship and Social Associations | Export Search Results | Social Network Analysis | User Guide | Explanation of Terms

**Data Export**

Export Configurations

- Individual Data
- Individual Data+Office Postings
- Individual Data+Kin+Kin Social Associations
- Individual Data+Kin+Kin Office Posting
- Individual Data+Kin+Kin Mode of Entry
- Kin with Five Degree of Mourning

GU  GD  CO  MA

Kinship Analysis table 1  Kinship Analysis table 2

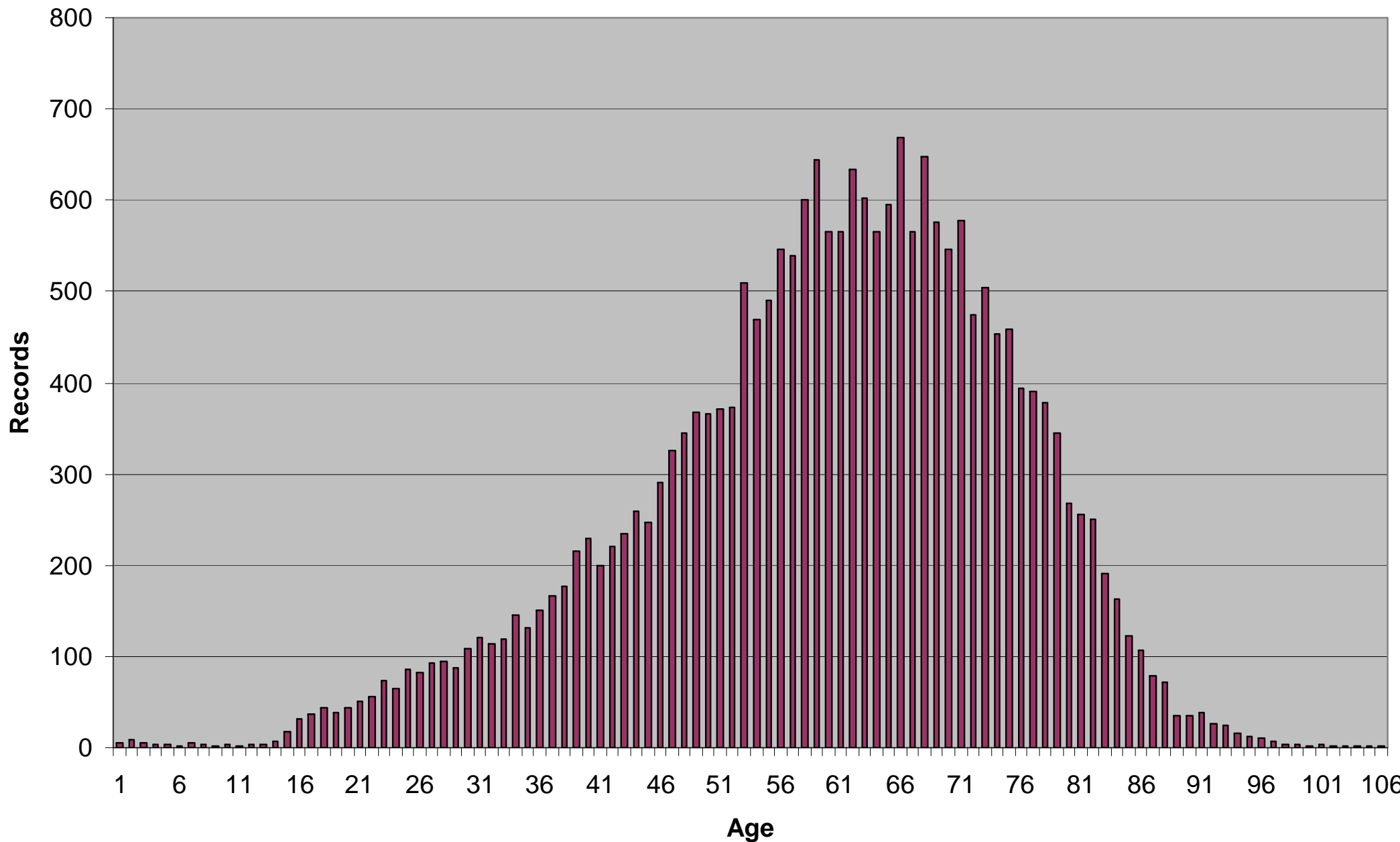
**Kinship Analysis**

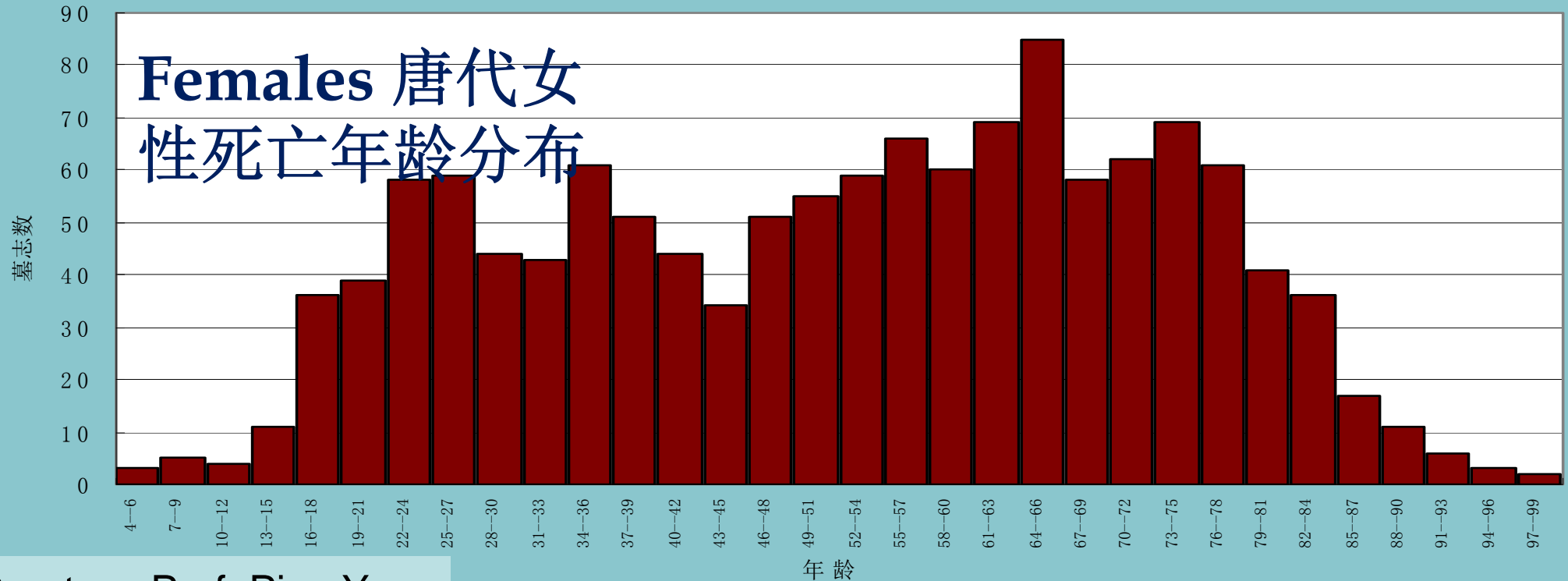
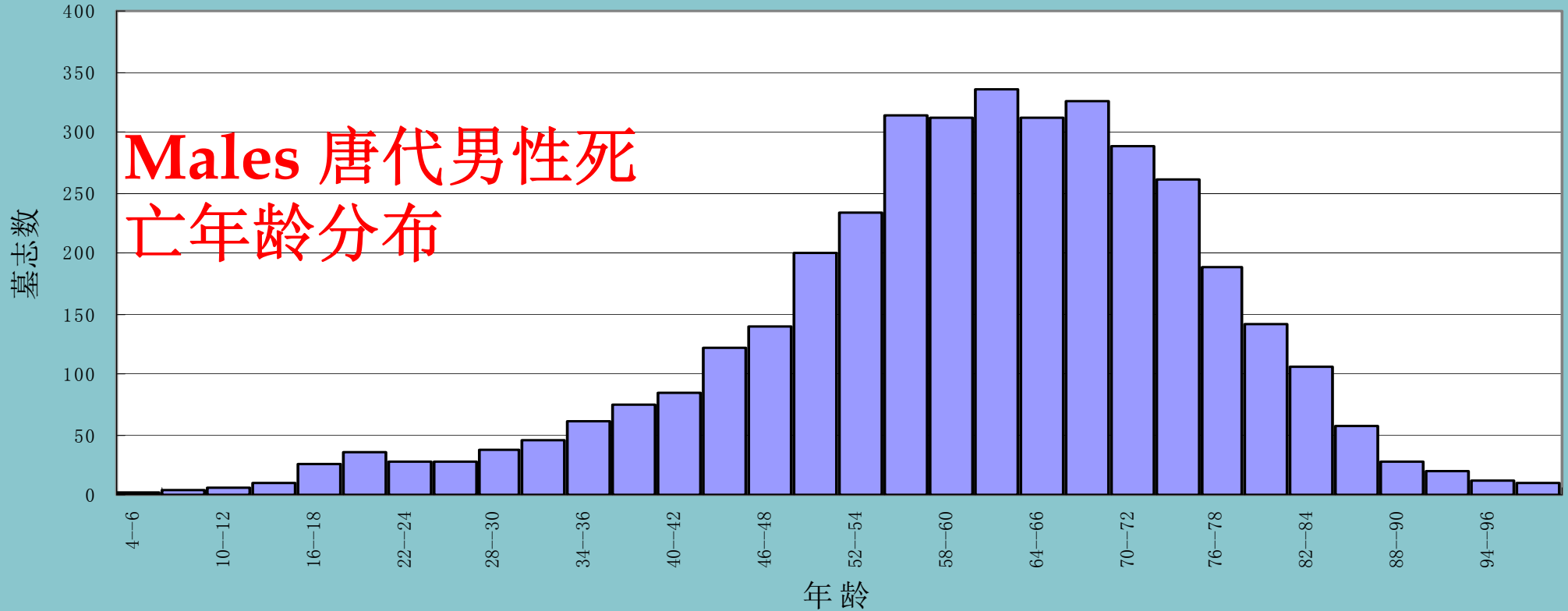
#	System Id	Name-Pinyin	Name-Chinese	Index Year	Place of Basic Affiliation	層級	Mourning Grade	Kinship Type	Kinship Type Id	Kin's Name-Pinyin	Kin's Name-Chinese	Kin's Index Year	Kin's Place
1	0007055	Lv Zuqian	呂祖謙	1181	金華	2	WF	妻父	0000669	Han Yuanji	韓元吉	1177	上饒
2	0007055	Lv Zuqian	呂祖謙	1181	金華	0	FF	祖父	0001320	Lv Bengzhong	呂朋中	1139	金華
3	0007055	Lv Zuqian	呂祖謙	1181	金華	2	FFB	伯叔祖父	0001322	Lv Chenzhong	呂忱中	1157	開封
4	0007055	Lv Zuqian	呂祖謙	1181	金華	2	FFB	伯叔祖父	0001333	Lv Yongzhong	呂用中	1140	金華
5	0007055	Lv Zuqian	呂祖謙	1181	金華	0	MF	外祖父	0001692	Zeng Ji	曾幾	1143	山陰
6	0007055	Lv Zuqian	呂祖謙	1181	金華	0	F	父	0003652	Lv Daqi	呂大器	1169	金華
7	0007055	Lv Zuqian	呂祖謙	1181	金華	1	FB	伯叔父	0003653	Lv Datong	呂大同	1178	金華
8	0007055	Lv Zuqian	呂祖謙	1181	金華	1	B	兄弟	0003655	Lv Zujian	呂祖儉	1196	金華
9	0007055	Lv Zuqian	呂祖謙	1181	金華	1	B	兄弟	0003656	Lv Zutai	呂祖泰	1208	宣興
10	0007055	Lv Zuqian	呂祖謙	1181	金華	1	FFF	母祖父	0003901	Zeng Zhun	曾準	1093	洛陽
11	0007055	Lv Zuqian	呂祖謙	1181	金華	1	MB	母之兄弟	0003910	Zeng Dai	曾達	1183	山陰
12	0007055	Lv Zuqian	呂祖謙	1181	金華	2	FBW	伯叔母	0005086	Fang Shi	方氏 (呂大同妻)	1176	桐廬
13	0007055	Lv Zuqian	呂祖謙	1181	金華	1	W	妻	0005087	Han Shi	韓氏 (呂祖謙妻)	1162	上饒
14	0007055	Lv Zuqian	呂祖謙	1181	金華	2	FFFF	高祖	0007052	Lv Xizhe	呂希哲	1119	開封
15	0007055	Lv Zuqian	呂祖謙	1181	金華	3	FFF	母祖父	0007220	Fang Meng	方蒙	1097	桐廬
16	0007055	Lv Zuqian	呂祖謙	1181	金華	3	FFF	曾伯叔祖父	0007302	Lv Qiewen	呂切問	1125	開封
17	0007055	Lv Zuqian	呂祖謙	1181	金華	1	FFF	曾祖	0008099	Lv Haowen	呂好問	1123	開封
18	0007055	Lv Zuqian	呂祖謙	1181	金華	3	WB	妻兄弟	0010187	Han Biao	韓澆	1218	上饒
19	0007055	Lv Zuqian	呂祖謙	1181	金華	2	BS	姪	0010321	Lv Qiaonian	呂喬年	1191	金華
20	0007055	Lv Zuqian	呂祖謙	1181	金華	2	FFB	伯叔祖父	0010703	Lv Benzhong	呂本中	1143	金華
21	0007055	Lv Zuqian	呂祖謙	1181	金華	0	BS	姪	0011651	Lv Kangnian	呂康年	1238	金華
22	0007055	Lv Zuqian	呂祖謙	1181	金華	0	S	子	0011652	Lv Yannian	呂延年	1248	金華
23	0007055	Lv Zuqian	呂祖謙	1181	金華	2	MF	外祖父	0017033	Fang Yuanju	方元矩	1130	桐廬
24	0007055	Lv Zuqian	呂祖謙	1181	金華	1	MB	母之兄弟	0017599	Zeng Feng(2)	曾逢	1180	山陰
25	0007055	Lv Zuqian	呂祖謙	1181	金華	1	FB	伯叔父	0017903	Lv Dalun	呂大倫	1173	金華
26	0007055	Lv Zuqian	呂祖謙	1181	金華	1	FB	伯叔父	0017904	Lv Dayou	呂大猷	1175	金華
27	0007055	Lv Zuqian	呂祖謙	1181	金華	2	FFFBS	族伯叔祖父	0017943	Lv Wenzhong(2)	呂文中	1173	金華
28	0007055	Lv Zuqian	呂祖謙	1181	金華	1	MFBS	堂舅之子	0021701	Zeng Song	曾松	1193	山陰
29	0007055	Lv Zuqian	呂祖謙	1181	金華	2	FFM	曾祖母	0041922	Wang Shi	王氏 (呂好問妻)	1123	金華
30	0007055	Lv Zuqian	呂祖謙	1181	金華	3	FFFM	高祖母	0041954	Zhang Shi	張氏 (呂希哲妻)	1111	金華
31	0007055	Lv Zuqian	呂祖謙	1181	金華	1	M	母	0042142	Fang Shi	方氏 (呂大器妻)	1175	金華
32	0007055	Lv Zuqian	呂祖謙	1181	金華	1	M	母	0044808	Zeng Shi	曾氏 (呂大器妻)	1175	金華
33	0007055	Lv Zuqian	呂祖謙	1181	金華	0	D	女	0045258	Lv Huanian	呂華年	1218	金華
34	0007055	Lv Zuqian	呂祖謙	1181	金華	0	DH	女婿	0045259	Pan Jingliang	潘景良	1200	金華

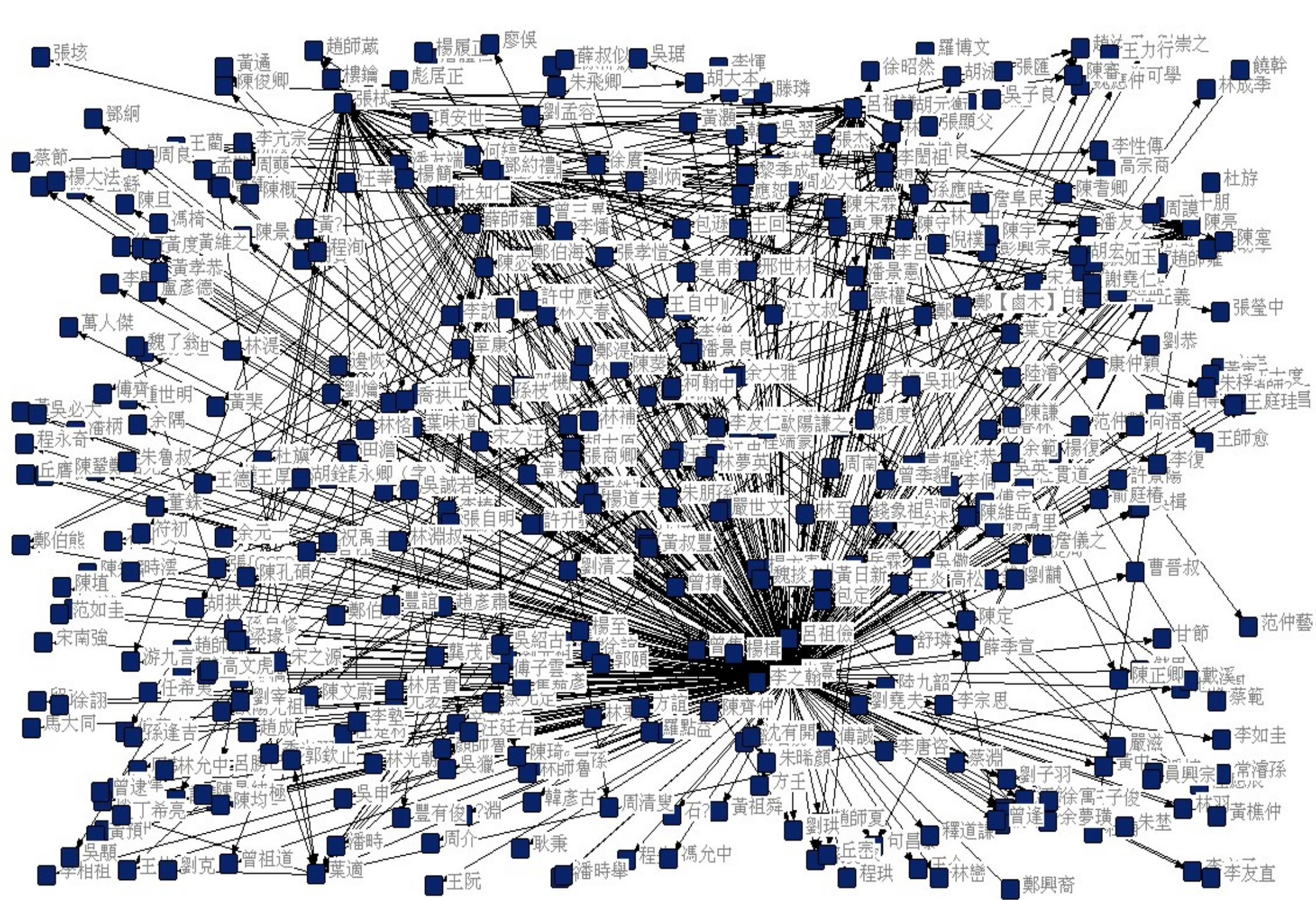
Bookmarks 81 | By Alpha A-Z

- Apache UIMA - Apache ...
- Biographical Dictionary ...
- BSPSPProjectHome - berke...
- Buij database De Weerd
- Captured Data
- Captured Text
- CBDB Access Data
- CBDB API Calls
- CBDB archive
- CBDB Chinese page
- CBDB code tables
- CBDB inputting
- CBDB Query Page
- CBDB-Tang epitaphs
- CCT-database
- CCT-Database | Informatie
- China Biographical Datab...
- China Biographical Datab...
- China Biographical Datab...
- Chinese Family relations...
- Chinese kinship - Wikipe...
- Cousin - Wikipedia, the fr...
- Data Analysis - The Catal...
- Data Minina Work Sched...

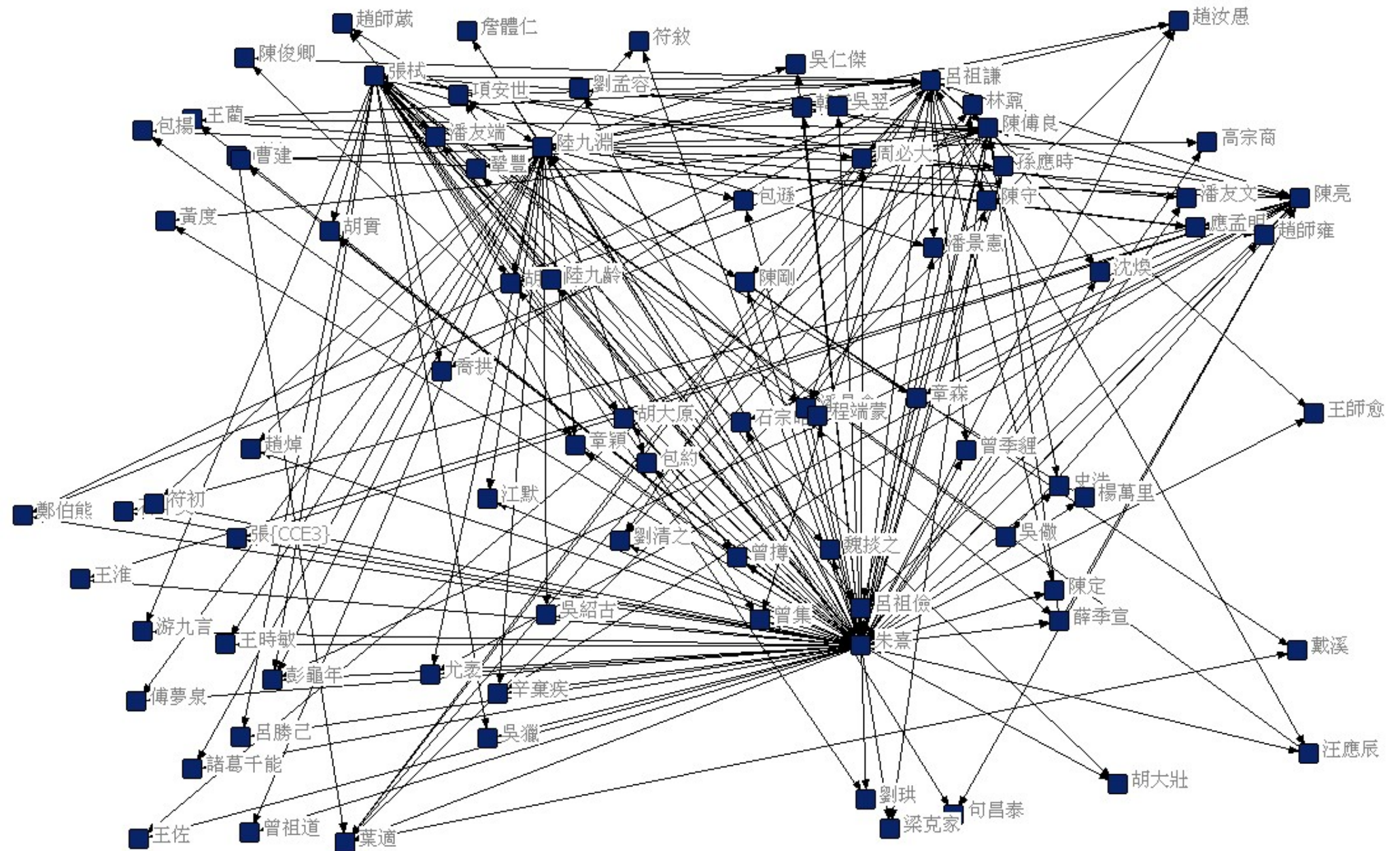
# Age at Death-CBDB data Tang through Qing - 22270 persons





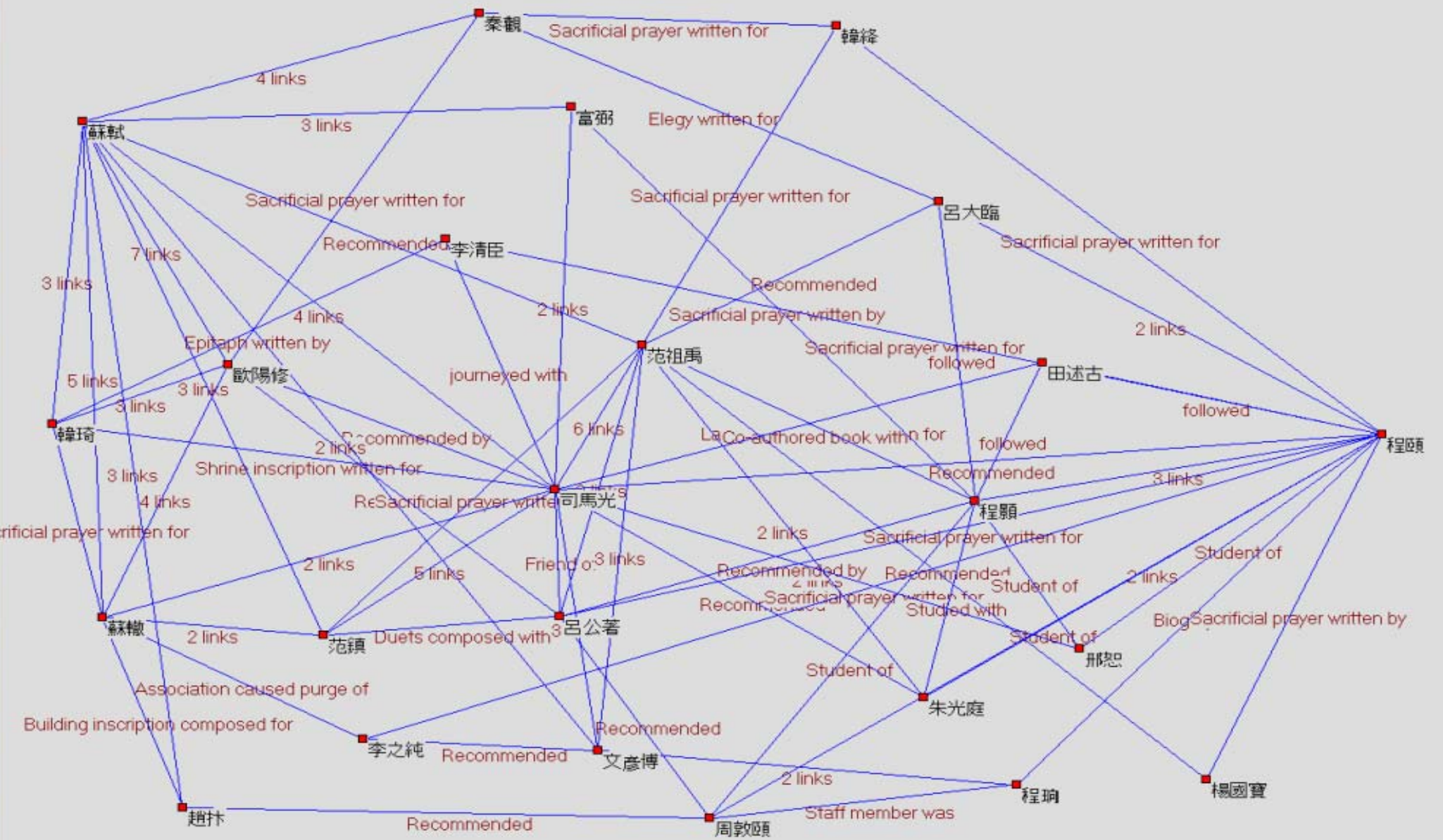


2717 letters to or from leading Neo-Confucians of the mid-12<sup>th</sup> century; 453 persons

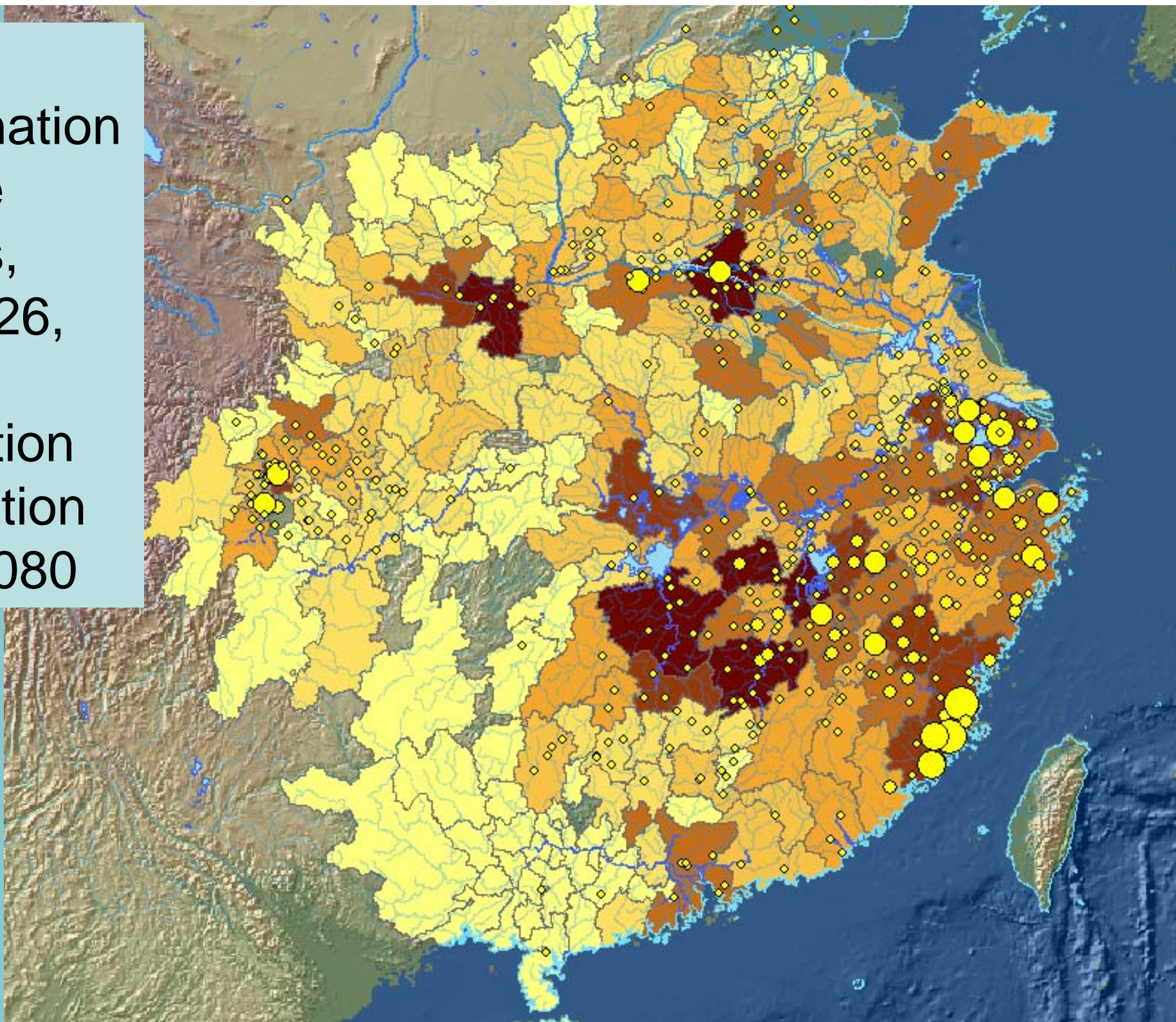


Showing only those who corresponded with at least two people in the dataset





4730  
Examination  
Degree  
holders,  
960-1126,  
with  
population  
distribution  
as of 1080



CBDB

# **PROVIDING CODES TABLES FOR OTHER PROJECTS**

# Address codes

c_addr_id	c_name	c_name_cn	c_admin_type	c_firstyear	c_lastyear	x_coord	y_coord	belongs1_ID	belongs1_Name	belongs2_ID	belongs2_Name	belongs3_ID	belongs3_Name
13071	Lanshan	藍山	Xian	999	1003	112.1778	25.37241	13058	郴州	13015	荆湖南路	10989	宋朝
13072	Shao Zhou	邵州	Zhou	999	1224	111.4749	27.24932	13015	荆湖南路	10989	宋朝		
13073	Baoqing Fu	寶慶府	Fu	1225	1279	111.4749	27.24932	13015	荆湖南路	10989	宋朝		
13074	Shaoyang	邵陽	Xian	999	1224	111.4749	27.24932	13072	邵州	13015	荆湖南路	10989	宋朝
13075	Xinhua	新化	Xian	1072	1224	111.2748	28.00967	13072	邵州	13015	荆湖南路	10989	宋朝
13076	Wugang	武岡	Xian	999	1105	110.6223	26.73071	13072	邵州	13015	荆湖南路	10989	宋朝
13077	Shizhu	蒔竹	Xian	1081	1105	110.0268	26.49105	13072	邵州	13015	荆湖南路	10989	宋朝
13078	Quan Zhou	全州	Zhou	999	1279	111.065	25.93181	13015	荆湖南路	10989	宋朝		
13079	Qingxiang	清湘	Xian	999	1279	111.065	25.93181	13078	全州	13015	荆湖南路	10989	宋朝
13080	Guanyang	灌陽	Xian	999	1279	111.1447	25.48865	13078	全州	13015	荆湖南路	10989	宋朝

# Office codes

c_office_id	c_office_pinyin	c_office_chn	c_office_pinyin_alt	c_office_chn_alt	c_office_trans
53	lu lu fa yun shi	陸路發運使			Supply Commissioner of the Land Forces
54	men xia shi lang	門下侍郎	men shi; shi lang; er dong sheng; sheng zuo; suo gui can yi; dong tai guan xia	門侍;侍郎;貳東省;省佐;瑣闈參議;東臺管轄	Vice Director of the Chancellery
59	san si du ping you si	三司都憑由司	ping you	憑由	Certificate Validation Office of the State Finance Commission
69	san si gou dang gong shi	三司勾當公事	san si gan dang gong shi; gou dang gong shi	三司幹當公事;勾當公事	Office Manager of the State Finance Commission
70	san si shi	三司使	ji xiang; zong ji; sheng zhu	計相;總計;省主	State Finance Commissioner
71	ti ju cha chang mai ma	提舉茶場買馬			Supervisor of Horse
77	shao zai	少宰	ci fu; kui zai	次輔;揆宰	Junior Steward
79	shui lu fa yun shi	水路發運使			Supply Commissioner
81	shu mi shi	樞密使	shu shi; shu mi; mi shi; yuan shu; shu zhou; shu zhang	樞使;樞密;密使;元樞;樞軸;樞長	Military Affairs Commissioner
82	si kong ping zhang jun guo shi	司空平章軍國事			Minister of Works -Manager of National Security Matters

# Using CBDB code tables in text analysis

宋會要輯稿

文獻集: [宋會要輯稿](#)

[回首頁](#) [功能說明](#) [修改密碼](#) [登出](#)

文件檢索: 方臘

[送出](#) [清除](#) [進階檢索](#)

測試版

最近幾次檢索

方臘	39
{DC:Part_0XX}	0
{DC:Part_081}	396
all	80396

呈現模式

排列方式

檢索結果 | [逐篇檢視](#) | [詞類與全文](#) | [列印全文](#) | [檔名順序](#) | [年代](#)

找到筆數: 39 頁次: 2 | [下一頁](#)

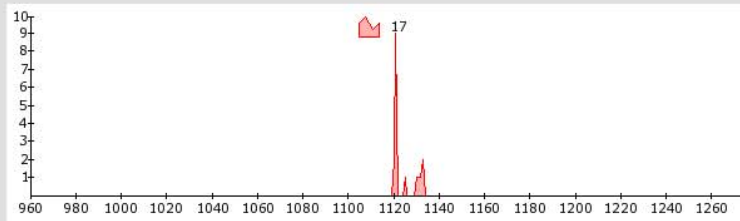
相關人名: [方臘](#) [王粲](#) [劉鎮](#) [童貫](#) [辛興宗](#) [曾經](#)

相關地名: [秀州](#) [杭州](#) [兩浙路](#) [湖州](#) [婺州](#) [平江府](#) [平江](#) [江陰軍](#)

依 df 降序排列 | 依 t→q 降序排列

Term t	(tf)	df	t→q
<a href="#">方臘</a> 生平	(19)	18	1.000
<a href="#">王粲</a> 生平	(7)	7	0.636
<a href="#">劉鎮</a> 生平	(8)	7	0.583
<a href="#">童貫</a> 生平	(5)	5	0.058
<a href="#">辛興宗</a> 生平	(4)	4	0.500
<a href="#">楊惟忠</a> 生平	(4)	4	0.308
<a href="#">曾經</a> 生平	(5)	4	0.007
<a href="#">毛友</a> 生平	(6)	3	0.600
<a href="#">范汝為</a> 生平	(3)	3	0.158
<a href="#">劉光世</a> 生平	(4)	3	0.024

[更多詞彙...](#)

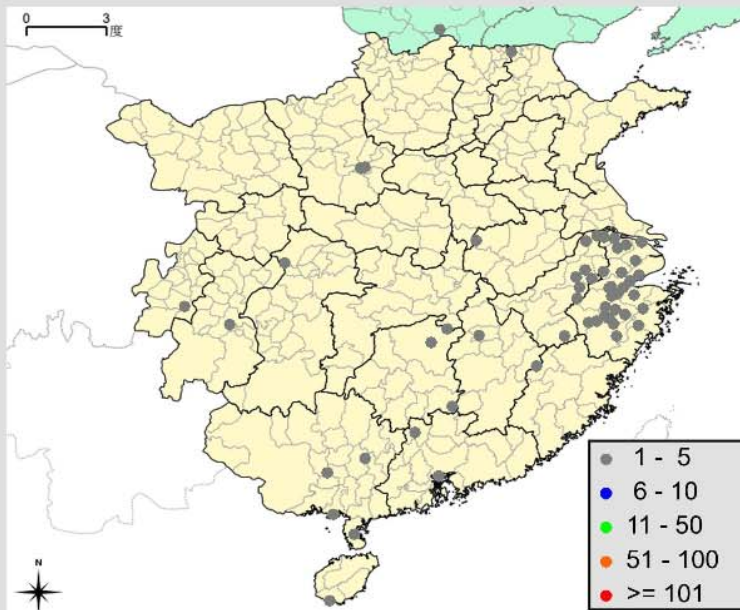


檢索結果分佈圖(此圖僅計入含有年代資訊的文件; 橫軸:西元年, 縱軸:文件數量)

依 df 降序排列 | 依 t→q 降序排列

Term t	(tf)	df	t→q
<a href="#">秀州</a>	(20)	5	0.017
<a href="#">婺州</a>	(8)	4	0.019
<a href="#">杭州</a>	(4)	4	0.018
<a href="#">湖州</a>	(4)	4	0.014
<a href="#">兩浙路</a>	(4)	4	0.008
<a href="#">睦州</a>	(3)	3	0.035
<a href="#">平江</a>	(4)	3	0.028
<a href="#">江陰軍</a>	(3)	3	0.027
<a href="#">臨安</a>	(6)	3	0.017
<a href="#">常州</a>	(3)	3	0.012

[更多詞彙...](#)



地名文件數量分佈圖(此圖僅計入含有地名資訊的文件)

依 df 降序排列 | 依 t→q 降序排列

Term t	(tf)	df	t→q
<a href="#">尚書省</a>	(29)	6	0.002
<a href="#">戶部</a>	(8)	6	0.002
<a href="#">吏部</a>	(15)	4	0.002
<a href="#">宣撫使</a>	(3)	3	0.009
<a href="#">大理寺</a>	(5)	3	0.004
<a href="#">刑部</a>	(11)	3	0.002
<a href="#">三省</a>	(3)	3	0.002
<a href="#">臨安府</a>	(15)	3	0.002

一次選取(或取消)全部勾選

■ 系統在顯示全文時, 將以不同的顏色標示人名、地名、官名、與其他詞彙

Elif Yamangil [yamangil@fas.harvard.edu](mailto:yamangil@fas.harvard.edu)

# **CBDB REGEX MACHINE**

**THE CBDB REGEX MACHINE IS AVAILABLE FOR  
DOWNLOAD AT THE CBDB WEBSITE**

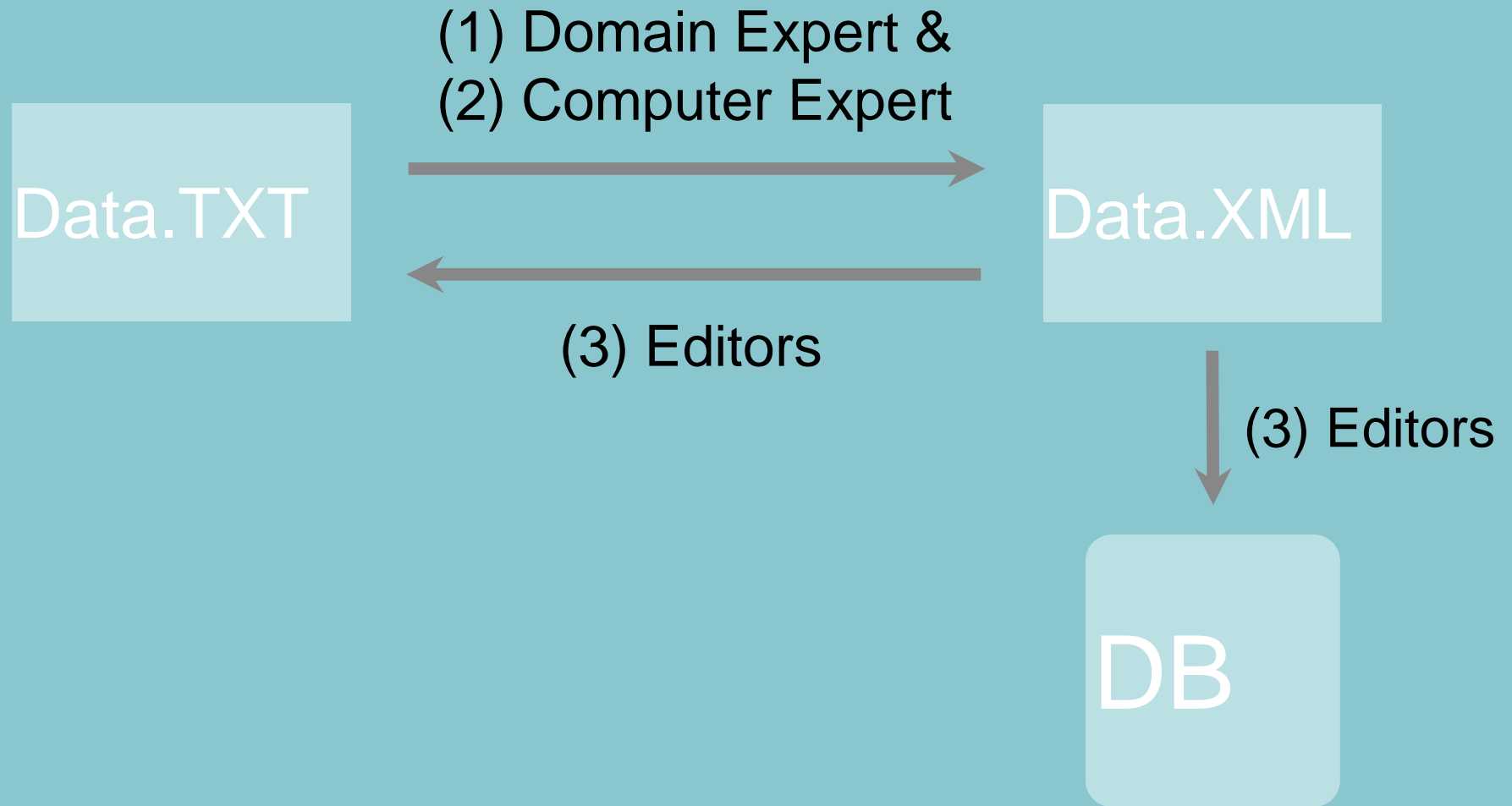
# Data-Mining Biographies

- Regular Expressions (Regex)
- **Simple** yet **sufficient** model

# Task Flow

1. “**Domain Expert**” comes up with “patterns”
2. “**Computer Expert**” programs the patterns
3. “**Editors**” check results / make revisions
  - a)if performance **acceptable**, commit to database (DB)
  - b)otherwise return to **step 1**

# Task Flow



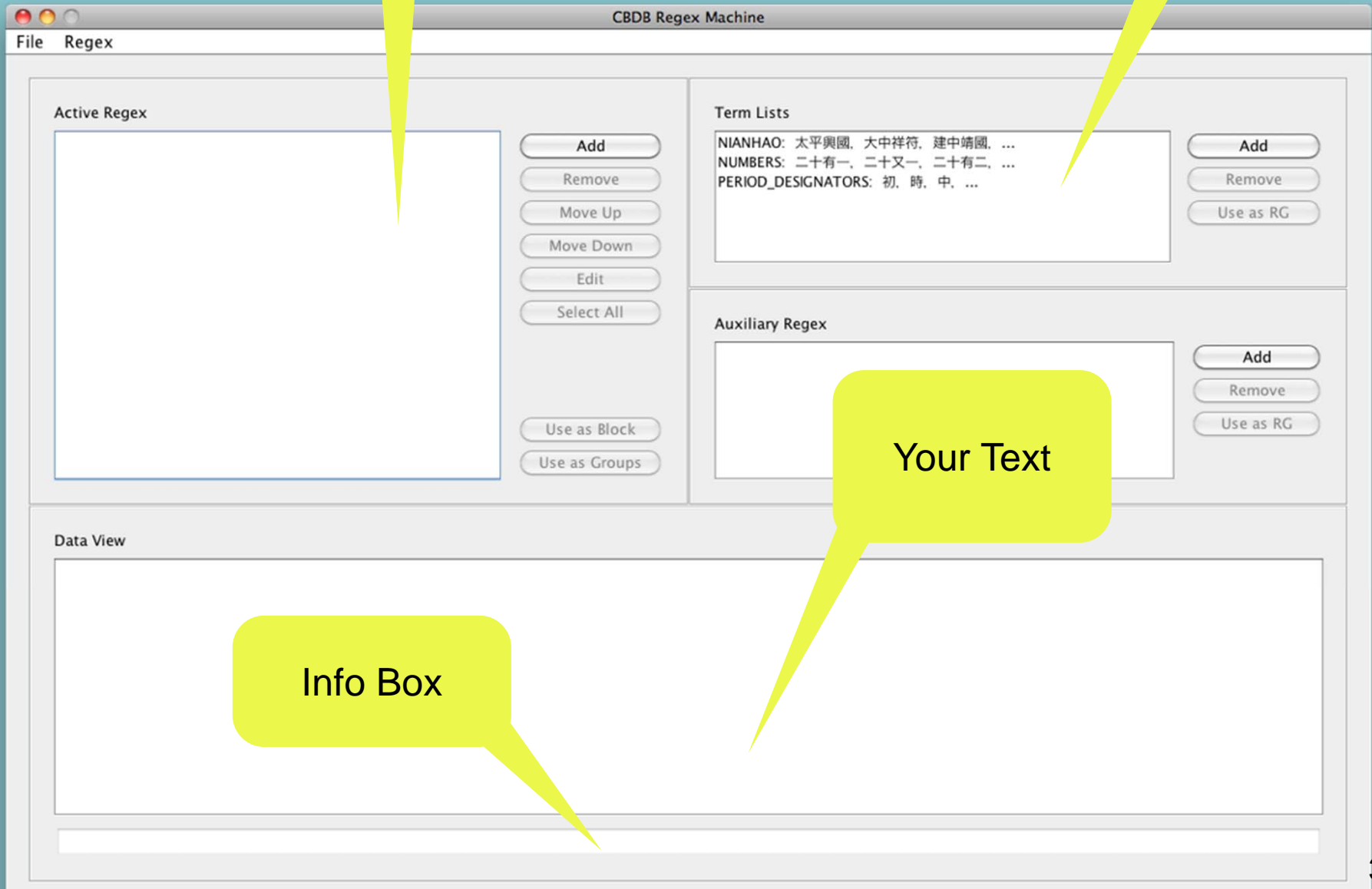
# CBDB Regex Machine

- **Step 2** requires programming skills!
- Our graphical user interface (GUI) allows end users to achieve both **step 1** (design regex) and **step 2** (implement regex) without computer skills

# GUI

List of active regex

List of "terms"



# Using little bricks of regex to compile the big ones: clarity + reusability

Little bricks 2: term lists

- Nianhao
- Chinese numbers
- Period designators
- ...

The screenshot shows the 'CBDB Regex Machine' application window. It has a menu bar with 'File' and 'Regex'. The main area is divided into several sections:

- Active Regex:** A list of four active regex patterns:
  - DATE: [ DELIM ] [ NIANHAO ] [ NUMBERS ] [ 年 ] [ DELIM ]
  - DATE: [ DELIM ] [ NIANHAO ] [ NUMBERS ] [ 年 ] [ 閏 ] [ NUMBERS ] [ ]
  - DATE: [ DELIM ] [ NIANHAO ] [ NUMBERS ] [ 年 ] [ 閏 ] [ NUMBERS ] [ ]
  - DATE: [ NIANHAO ] [ PERIOD\_DESIGNATORS ]
- Term Lists:** A list of three term lists:
  - NIANHAO: 太平興國, 大中祥符, 建中靖國, ...
  - NUMBERS: 二十有一, 二十又一, 二十有二, ...
  - PERIOD\_DESIGNATORS: 初, 時, 中, 間, 末, 季, ...
- Auxiliary Regex:** A list of auxiliary regex patterns:
  - DELIM: "[ , . : ; ]"
  - NONDELIM: "[ ^ , . : ; ]"
  - YEAR: "年"
  - MONTH: "月"
  - DAY: "日"
  - INTERCALARY: "閏"

Buttons for 'Add', 'Remove', 'Move Up', 'Move Down', 'Edit', and 'Select All' are visible between the sections.

Regex in use (active)

- Date patterns
- Office appointment patterns
- ...

Little bricks 1: auxiliary regex

- the characters for year/month/day/intercalary...
- Delimiters: Chinese comma, period...
- ...

# GUI example

- Four **Date** patterns that we use:
- [reign period] [number] [*“year”*]
- [reign period] [number] [*“year”*] [number] [*“month”*]
- [reign period] [number] [*“year”*] [number] [*“month”*] [number] [*“day”*]
- [reign period] [period designator]

# GUI

- **Office** appointment patterns use **Date** patterns as building blocks:
- [*“appointed to”*] [office] [*“on”*] [**Date**]

# GUI

- What's in a **Date** pattern?
- [reign period] [number] [*“year”*]

太平興國,大中  
祥符,建中靖國  
,開平,乾化,貞  
明,  
...

元,一,  
二,三,  
四,五,  
...

“年”

consists of three  
**“Regex Groups”**

# GUI

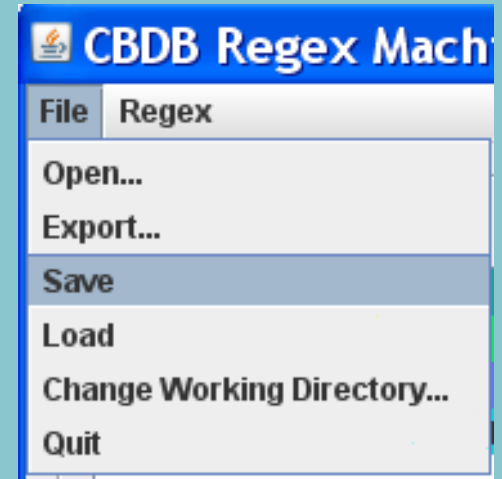
- We make use of “**Term Lists**”
  - Nianhao (reign period)
  - Numbers
  - Period designators (beginning of, during)
  - Verbs of appointment (assigned, appointed, promoted...)

# Operations

- Load Auxiliary Regex
- Load Date Patterns



- Save: save your environment
- Load: load your environment



- Open: open a text file to extract info from
- Export: export the extraction result as XML

# CBDB Regex Machine in Action

- Load Date Patterns
- Open the text file: **unicode.txt**
- Click “Select All” or select the date patterns one by one

Data View

2	帝系一之一	帝系	帝系一	帝號	順祖惠元睿明皇帝，諱珽，僖祖子。歷備鎮從事，兼御[史]中丞。建隆元年三月，追尊曰惠元皇帝，廟號順祖。諡議，翰林學士曾
3	帝系一之二	帝系	帝系一	帝號	翼祖簡恭睿德皇帝，諱敬，順祖子。歷營、薊、涿三州刺史。周顯德中，贈左衛上將軍。建隆元年三月，追尊曰簡恭皇帝，廟號
4	帝系一之三	帝系	帝系一	帝號	宣祖昭武睿聖皇帝，諱弘殷，翼祖子，母曰簡穆皇后劉氏。仕周為龍捷左廂都指揮使，岳州防禦使。顯德三年七月二十六日崩
5	帝系一之三	帝系	帝系一	帝號	太祖啓運立極英武睿文神德聖功至明大孝皇帝，諱匡胤，宣祖第二子，母曰昭憲皇后杜氏。後唐天成二年丁亥歲二月十六日，生
7	帝系一之四	帝系	帝系一	帝號雜錄	《十朝綱要》太宗至仁應道神功聖德文武睿烈大明廣孝皇帝，諱炅，宣祖第三子，母曰昭憲皇后杜氏。晉天福四年己亥歲十月七
8	帝系一之五	帝系	帝系一	帝號雜錄	李直口土《十朝綱要》仁宗體天法道極功全德神文聖武睿明哲孝皇帝睿：原作「濬」，據《宋史》卷九《仁宗本紀》一改。諡
9	帝系一之七	帝系	帝系一	帝號雜錄	【宋會要】孝宗高宗皇帝之子也。初，藝祖皇帝六世孫秀王娶夫人張氏，以建炎元年丁未歲十月二十二日生上於秀州嘉興縣。紹
12	帝系一之八	帝系	帝系一	太子諡	【宋會要】昭成太子元僖，初諡恭孝。不懈為德曰恭，慈惠愛親曰孝。乾興元年改諡昭成。明德有功曰昭，安民立政曰成。
94	帝系一之二〇	孝宗	淳熙二年	次詣宮中奉上壽聖齊明廣慈備德太上皇[后]冊寶。冊文丞相梁克家撰，書冊文參政施師點，篆寶參政黃洽。冊文曰：「皇帝臣慎謹稽首再拜言：臣	
118	帝系一之二五	皇子諸王	高密郡王德恭	高密郡王德恭此上原有眉批云：「高密郡王以下係諸王之子，已見後」。	，魏王廷美子。太平興國四年二月，以皇子授貴州防禦使。七年五月，奉
130	帝系一之二八	皇子諸王	同安郡王惟正	同安郡王惟正此上原有眉批云：「同安郡王以下係諸王之子，已見後」。	，燕王德昭子。太平興國八年十月，授左千牛衛將軍。四遷左龍武軍。與
131	帝系一之二八	皇子諸王	翼王惟吉	翼王惟吉，燕王德昭子。幼養宮中，太祖視之如子，與諸叔聯名德雍。	太平興國八年十月，授右監門衛將軍。端拱元年二月，改今名，進左驍衛
132	帝系一之二八	皇子諸王	舒國公惟忠	舒國公惟忠，燕王德昭子。初名文都，太平興國九年十月賜今名，授右千牛衛將軍。四遷左龍武軍。端拱元年二月，改今名，進左驍衛	將軍。端拱元年二月，改今名，進左驍衛

97 match(es) found.

# GUI

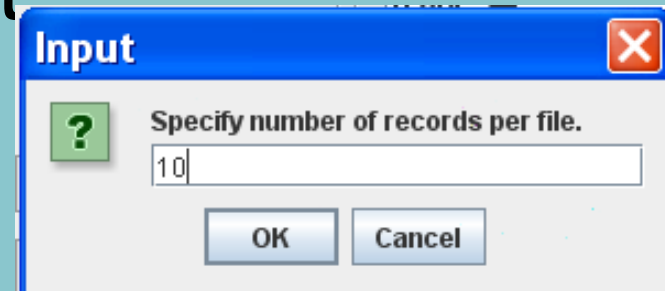
The screenshot displays the CBDB Regex Machine interface. At the top, the title bar reads "CBDB Regex Machine". Below it is a menu bar with "File" and "Regex".

The main interface is divided into several sections:

- Active Regex:** A list of four regex patterns: `DATE: [ DELIM ] [ NIANHAO ] [ NUMBERS ] [ 年 ] [ DELIM ]`, `DATE: [ DELIM ] [ NIANHAO ] [ NUMBERS ] [ 年 ] [ 同 ] [ NUMBE`, `DATE: [ DELIM ] [ NIANHAO ] [ NUMBERS ] [ 年 ] [ 同 ] [ NUMBE`, and `DATE: [ NIANHAO ] [ PERIOD_DESIGNATORS ]`. To the right are buttons for "Add", "Remove", "Move Up", "Move Down", "Edit", "Select All", "Use as Block", and "Use as Groups".
- Term Lists:** A list of terms: `NIANHAO: 太平興國, 大中祥符, 建中靖國, ...`, `NUMBERS: 二十有一, 二十又一, 二十有二, ...`, and `PERIOD_DESIGNATORS: 初, 時, 中, ...`. To the right are buttons for "Add", "Remove", and "Use as RG".
- Auxiliary Regex:** An empty text area with buttons for "Add", "Remove", and "Use as RG".
- Data View:** A large text area containing a block of Chinese text with several date-related phrases highlighted in green and blue. Below the text, it says "97 match(es) found.".

# Export the extraction result automatically

- Top menu: File -> Export



- Go to your working directory:
  - unicode.txt\_0.xml
  - unicode.txt\_1.xml
  - unicode.txt\_2.xml
  - unicode.txt\_3.xml

# The text is encoded in XML using the CBDB codes

- `<DATE nianhao="顯德" nianhao_code="441" period_designator="中" western_year="957">顯德中</DATE>`
- `<DATE nianhao="建隆" nianhao_code="505" year="元" month="三" western_year="960">建隆元年三月</DATE>`

```
<?xml version="1.0" encoding="UTF-8" ?>
```

```
<bios>
```

```
- <bio line="1">
```

2 帝系一之一 帝系 帝系一 帝號 順祖惠元睿明皇帝，諱珽，僖祖子。歷藩鎮從事，兼御[史]中丞。

`<DATE nianhao="建隆" nianhao_code="505" year="元" month="三" western_year="960">建隆元年三月</DATE>`

，追尊曰惠元皇帝，廟號順祖。諡議，翰林學士竇儼撰；冊文，中書舍人趙逢撰。陵曰康陵。在幽州。大中祥符五年閏十月加今諡。冊文，參知政事陳彭年撰。《要》云：冊文，樞密使陳堯叟撰。

`<DATE nianhao="熙寧" nianhao_code="528" year="五" month="十一" day="二十五" western_year="1072">熙寧五年十一月二十五日</DATE>`

，以奉僖祖為太廟始祖，告遷神主，奉藏於夾室，依禮不諱不忌。以上《續國朝會要》。

```
</bio>
```

```
- <bio line="2">
```

3 帝系一之二 帝系 帝系一 帝號 翼祖簡恭睿德皇帝，諱敬，順祖子。歷營、蜀、涿三州刺史。周

`<DATE nianhao="顯德" nianhao_code="441" period_designator="中" western_year="957">顯德中</DATE>`

，贈左衛上將軍。

`<DATE nianhao="建隆" nianhao_code="505" year="元" month="三" western_year="960">建隆元年三月</DATE>`

，追尊曰簡恭皇帝，廟號翼祖。諡議，翰林學士竇儼撰；冊文，中書舍人趙逢撰。陵曰靖陵。在幽州。本曰定陵，乾興元年七月改。大中祥符五年閏十月加今上《國朝會要》。《續會要》云：冊文，參知政事丁謂撰。元祐元年正月，以神宗祔廟，告遷神主，奉藏於夾室，依禮不諱不忌。以上《續國朝會要》。

```
</bio>
```

```
- <bio line="3">
```

4 帝系一之二 帝系 帝系一 帝號 宣祖昭武睿聖皇帝，諱弘殷，翼祖子，母曰簡穆皇后劉氏。仕周為龍捷左廂都指揮使，岳州防禦使。顯德三年七月二十六

`<DATE nianhao="建隆" nianhao_code="505" year="元" month="三" western_year="960">建隆元年三月</DATE>`

，追尊曰昭武皇帝，廟號宣祖。諡議，翰林學士竇儼撰；冊文，中書舍人扈蒙撰。陵曰安陵。在開封府開封縣，今奉先資福禪院即其地。

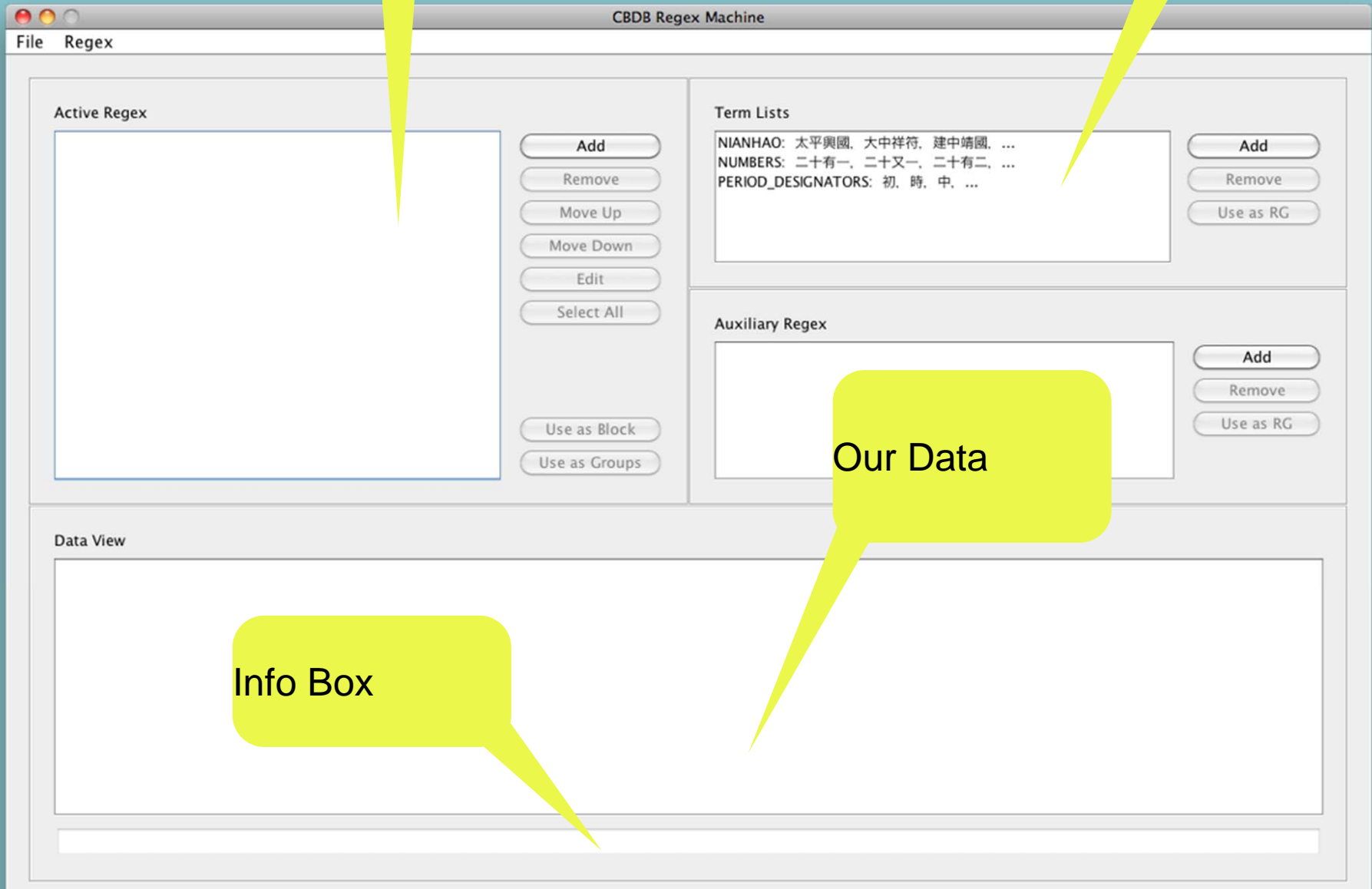
`<DATE nianhao="乾德" nianhao_code="506" year="二" western_year="964">乾德二年</DATE>`

，改卜於河南府鞏縣。

# GUI

List of active regex

List of "terms"



- File
- Regex
- Open...
- Export...
- Save
- Load
- Change Working Directory...
- Quit

] [NUMBERS] [年] [閏] [NUMBERS] [  
 ] [NUMBERS] [年] [閏] [NUMBERS] [  
 ] [NUMBERS] [年]  
 D\_DESIGNATORS ]

- Add
- Remove
- Move Up
- Move Down
- Edit
- Select All
- Use Regex
- Use Groups

Term Lists

NIANHAO: 太平興國，大中祥符，建中靖國，，...  
 NUMBERS: 二十有一，二十一，二十有二，，...  
 PERIOD\_DESIGNATORS: 初，時，中，間，末，季，...

- Add
- Remove
- Use Group

Auxiliary Regex

DELIM: "[, . : ; ]"  
 NONDELIM: "[^, . : ; ]"  
 YEAR: "年"  
 MONTH: "月"  
 DAY: "日"  
 INTERCALARY: "閏"

- Add
- Remove
- Use Group

Data View

2	帝系一之一	帝系	帝系一	帝號	順祖惠元睿明皇帝，諱珽，僖祖子。歷蒲鎮從事，兼御[史]中丞。建隆元年三月，追尊曰惠元皇帝，廟號順祖。諡議，翰林學士賈
3	帝系一之二	帝系	帝系一	帝號	翼祖簡恭睿德皇帝，諱敬，順祖子。歷管、薊、涿三州刺史。周顯德中，贈左衛上將軍。建隆元年三月，追尊曰簡恭皇帝，廟號
4	帝系一之二	帝系	帝系一	帝號	宣祖昭武睿聖皇帝，諱弘殷，翼祖子，母曰簡穆皇后劉氏。仕周為龍捷左廂都指揮使，岳州防禦使。顯德三年七月二十六日崩
5	帝系一之三	帝系	帝系一	帝號	太祖啟運立極英武睿文神德聖功至明大孝皇帝，諱匡胤，宣祖第二子，母曰昭憲皇后杜氏。後唐天成二年丁亥歲二月十六日，生
7	帝系一之四	帝系	帝系一	帝號雜錄	《十朝綱要》太宗至仁應道神功聖德文武睿烈大明廣孝皇帝，諱昚，宣祖第三子，母曰昭憲皇后杜氏。晉天福四年己亥歲十月七
8	帝系一之五	帝系	帝系一	帝號雜錄	李直二土《十朝綱要》仁宗體天法道極功全德神文聖武睿哲明孝皇帝睿：原作「濬」，據《宋史》卷九《仁宗本紀》一改。諡
9	帝系一之七	帝系	帝系一	帝號雜錄	【宋會要】孝宗高宗皇帝之子也。初，藝祖皇帝六世孫秀王娶夫人張氏，以建炎元年丁未歲十月二十二日生上於秀州嘉興縣。紹
12	帝系一之八	帝系	帝系一	太子諡	【宋會要】昭成太子元禧，初諡恭孝。不懈為德曰恭，慈惠愛親曰孝。乾興元年改諡昭成。明德有功曰昭，安民立政曰成。
94	帝系一之二〇	孝宗	淳熙二年	次詣宮中奉上册聖齊明廣慈備德太上皇[后]冊寶。冊文丞相梁克家撰，書冊文參政施師點，簽寶參政黃洽。冊文曰：「皇帝臣慎謹稽首再拜言：臣	
118	帝系一之二五	皇子諸王	高密郡王德恭	高密郡王德恭此上原有眉批云：「高密郡王以下係諸王之子，已見後」。魏王廷美子。太平興國四年二月，以皇子授貴州防禦使。七年五月，	
130	帝系一之二八	皇子諸王	同安郡王惟正	同安郡王惟正此上原有眉批云：「同安郡王以下係諸王之子，已見後」。燕王德昭子。太平興國八年十月，授左千牛衛將軍。四遷左龍武軍。真	
131	帝系一之二八	皇子諸王	翼王惟吉	翼王惟吉，燕王德昭子。幼養宮中，太祖視之如子，與諸叔聯名德雍。太平興國八年十月，授右監門衛將軍。端拱元年二月，改今名，進左驍衛	
132	帝系一之二八	皇子諸王	舒國公惟忠	舒國公惟忠，燕王德昭子。初名文起，太平興國八年十月賜今名，授左千牛衛將軍。四遷右龍武軍。真宗即位，改右千牛衛大將軍原無「大」字	

257 match(es) found.

## Active Regex

DATE: [ DELIM ] [ NIANHAO ] [ NUMBERS ] [ 年 ] [ 閏 ] [ NUMBERS ] [ ]  
 DATE: [ DELIM ] [ NIANHAO ] [ NUMBERS ] [ 年 ] [ 閏 ] [ NUMBERS ] [ ]  
 DATE: [ DELIM ] [ NIANHAO ] [ NUMBERS ] [ 年 ]  
 DATE: [ NIANHAO ] [ PERIOD\_DESIGNATORS ]

Add

Remove

Move Up

Move Down

Edit

Select All

## Term Lists

NIANHAO: 太平興國, 大中祥符, 建中靖國, , ...  
 NUMBERS: 二十有一, 二十又一, 二十有二, , ...  
 PERIOD\_DESIGNATORS: 初, 時, 中, 間, 末, 季, ...

Add

Remove

Use Group

## Auxiliary Regex

DELIM: "[ , . : ; ]"  
 NONDELIM: "[ ^ , . : ; ]"  
 YEAR: "年"

Add

Remove

Use Group

輸入



Specify number of records per file.

1000

確定

取消

## Data View

至仁應道神功聖德文武睿烈大明廣孝皇帝，諱昉，宣祖第三子，母曰昭憲皇后杜氏。晉天福四年己亥歲十月七日，生于開封府浚儀縣崇德坊護聖營官舍。《宋朝會要》：開寶九年詔以其日為乾明節，淳化元年改謂之聖祖節。仁宗體天法道極功全德神文聖武睿明孝皇帝睿：原作「濬」，據《宋史》卷九《仁宗本紀》一改。諱禎，真宗第六子，母曰章懿皇后李氏。大中祥符三年庚戌四月十四日生。《宋朝會要》：乾興元年詔以真宗皇帝之子也。初，娶祖皇帝六世孫秀王娶夫人張氏，以建炎元年丁未歲十月二十二日生上於秀州嘉興縣。紹興三十二年八月二十六日，詔以其日為會慶節。是夜，赤光滿室，如日正中，高宗皇帝知其生有聖瑞，故以子元億，初諡恭孝。不懈為德曰恭，慈惠愛親曰孝。乾興元年改諡昭成。明德有功曰昭，安民立政曰成。

上皇[后]冊寶。冊文丞相梁克家撰，書冊文參政施師點，篆寶參政黃洽。冊文曰：「皇帝臣慎謹稽首再拜言：臣聞慶都，是開放勳；塗山，寔肇與子。越周文臣，亦有任，徵音相繼，《雅》什載歌。永惟三五之隆，高密郡王以下係諸王之子，已見後」。魏王廷美子。太平興國四年二月，以皇子授貴州防禦使。七年五月，秦王得罪，削籍。九年四月，復以皇姪授峰州刺史。雍熙元年四月，封安定侯。端拱元年二月，進公，同安郡王以下係諸王之子，已見後」。燕王德昭子。太平興國八年十月，授左千牛衛將軍。四遷左龍武軍。真宗即位，進左千牛衛大將軍。大中祥符八年十一月，進簡州團練使。天禧二年八月，進齊州防禦使。太平興國八年十月，授右監門衛將軍。端拱元年二月，改今名，進左驍衛大將軍。至道二年二月，進開州觀察使。真宗即位，進武信軍節度使。景德二年十一月，加同中書門下平章事。太平興國八年十月賜今名，授左千牛衛將軍。四遷右龍武軍。真宗即位，改右千牛衛大將軍原無「大」字，據《宋史》卷二四四本傳補。又「軍」下原衍「軍」字，徑刪。大中祥符二年正月，進叙州刺史。端拱元年二月，授右武衛將軍。三遷右神武軍。真宗即位，進右龍武軍。大中祥符四年二月，加左千牛衛大將軍。八月卒，贈懷州刺史，追封河內侯。明道二年九月，贈保靜軍節度觀察留後，改封郡公。

<?xml version="1.0" encoding="UTF-8" ?>

- <bios>

- <bio line="1">

2 帝系一之一 帝系 帝系一 帝號 順祖惠元睿明皇帝，諱珽，僖祖子。歷藩鎮從事，兼御[史]中丞。

<DATE nianhao="建隆" nianhao\_code="505" year="元" month="三" western\_year="960">建隆元年三月</DATE>

，追尊曰惠元皇帝，廟號順祖。諡議，翰林學士竇儼撰；冊文，中書舍人趙逢撰。陵曰康陵。在幽州。

<DATE nianhao="大中祥符" nianhao\_code="515" year="五" intercalary="閏" month="十" western\_year="1012">大中祥符五年閏十月</DATE>

加今諡。冊文，參知政事陳彭年撰。以上《國朝會要》。《續會要》云：冊文，樞密使陳堯叟撰。

<DATE nianhao="熙寧" nianhao\_code="528" year="五" month="十一" day="二十五" western\_year="1072">熙寧五年十一月二十五日</DATE>

，以奉僖祖為太廟始祖，告遷神主，奉藏於夾室，依禮不諱不忌。以上《續國朝會要》。

</bio>

- <bio line="2">

3 帝系一之二 帝系 帝系一 帝號 翼祖簡恭睿德皇帝，諱敬，順祖子。歷營、薊、涿三州刺史。周

<DATE nianhao="顯德" nianhao\_code="441" period\_designator="中" western\_year="957">顯德中</DATE>

，贈左衛上將軍。

<DATE nianhao="建隆" nianhao\_code="505" year="元" month="三" western\_year="960">建隆元年三月</DATE>

，追尊曰簡恭皇帝，廟號翼祖。諡議，翰林學士竇儼撰；冊文，中書舍人趙逢撰。陵曰靖陵。在幽州。本曰定陵，

<DATE nianhao="乾興" nianhao\_code="517" year="元" month="七" western\_year="1022">乾興元年七月</DATE>

改。

<DATE nianhao="大中祥符" nianhao\_code="515" year="五" intercalary="閏" month="十" western\_year="1012">大中祥符五年閏十月</DATE>

加今諡，冊文，參知政事王曾撰。以上《國朝會要》。《續會要》云：冊文，參知政事丁謂撰。

<DATE nianhao="元祐" nianhao\_code="530" year="元" western\_year="1086">元祐元年</DATE>

正月，以神宗祔廟，告遷神主，奉藏於夾室，依禮不諱不忌。以上《續國朝會要》。

</bio>

- <bio line="3">

4 帝系一之二 帝系 帝系一 帝號 宣祖昭武睿聖皇帝，諱弘殷，翼祖子，母曰簡穆皇后劉氏。仕周為龍捷左廂都指揮使，岳州防禦使。

<DATE nianhao="顯德" nianhao\_code="441" year="三" western\_year="956">顯德三年</DATE>

七月二十六日崩，贈武清軍節度使。

<DATE nianhao="建隆" nianhao\_code="505" year="元" month="三" western\_year="960">建隆元年三月</DATE>

，追尊曰昭武皇帝，廟號宣祖。諡議，翰林學士竇儼撰；冊文，中書舍人扈蒙撰。陵曰安陵。在開封府開封縣，今奉先資福禪院即其地。

<DATE nianhao="乾德" nianhao\_code="506" year="二" western\_year="964">乾德二年</DATE>

，改卜於河南府鞏縣。

<DATE nianhao="景德" nianhao\_code="514" year="四" western\_year="1007">景德四年</DATE>

，析鞏、偃師、緱氏、登封縣地，置縣曰永安，以奉陵寢。

<DATE nianhao="大中祥符" nianhao\_code="515" year="五" intercalary="閏" month="十" western\_year="1012">大中祥符五年閏十月</DATE>

加今諡。冊文，宰臣王旦撰。以上《國朝會要》崇寧二年十二月，以哲宗祔廟，告遷神主，奉藏於夾室。三年十月二十九日，立九廟，復祀。以上《續國朝會要》

</bio>



Shih-pei Chen shihpeichen@fas.harvard.edu

**CBDB API**

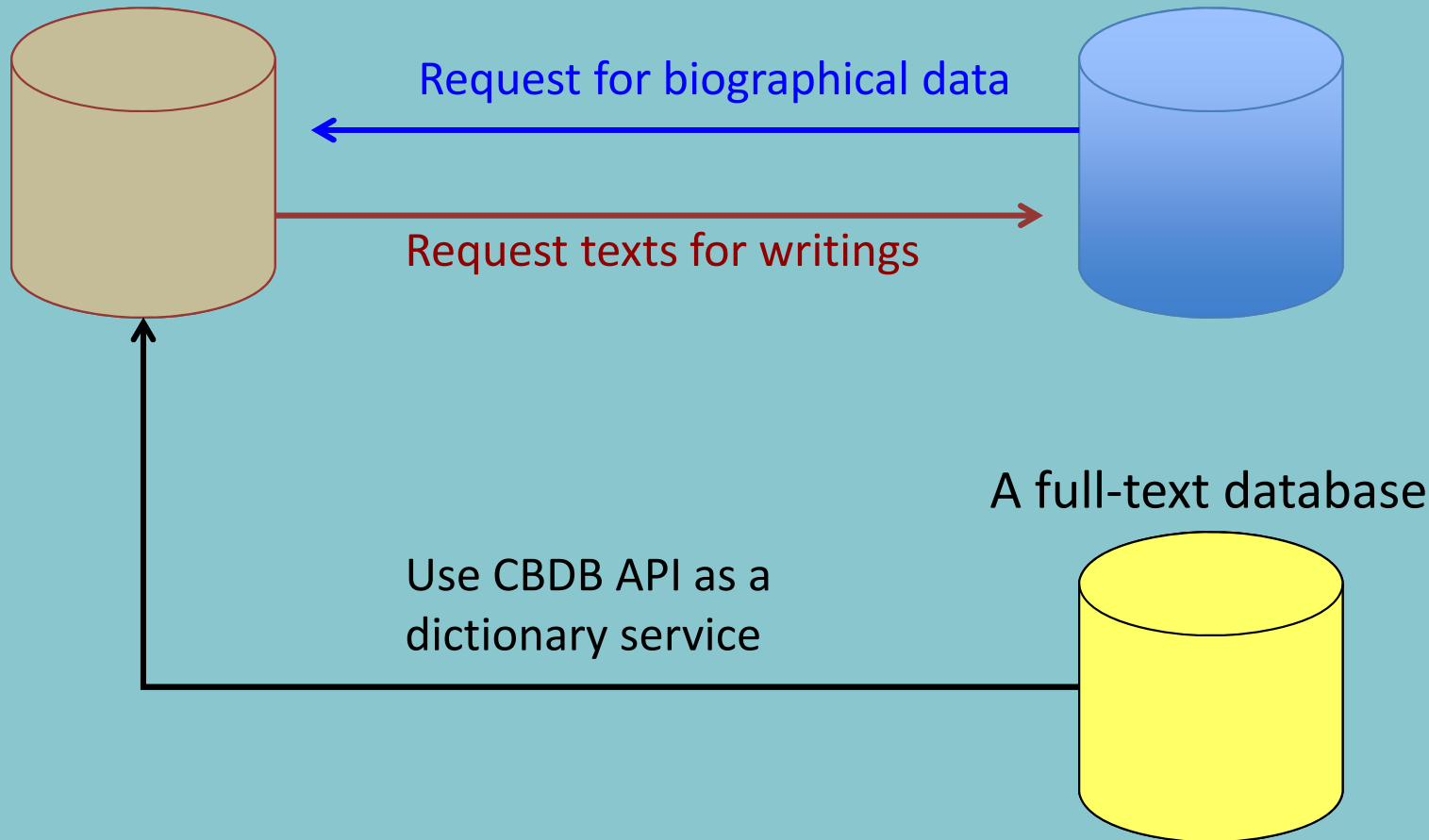
# Interoperability between databases

- Scholars are building databases to pursue particular research goals or to support specific fields. **Each database has its own focus and strength.**
  - E.g., **CBDB**'s focus is on collecting the complete biographical data of a person, while **MQWW**'s focus is on the writings of women writers.
- It ought to be unnecessary for any one database to replicate what others specialize in.

# System Interoperability

CBDB: biographical data

MQWW: women's writings



# System Interoperability

- APIs (application programming interfaces) enable interoperability among databases.
- Through providing an API, each database can **retrieve data from another database** to supplement its own data without need for re-inputting and/or hosting.

# The CBDB API

- Allows other databases to retrieve and present CBDB data *on the fly*.
- CBDB API currently accepts two types of queries:
  - Query by person ID (CBDB ID)
  - Query by name (Chinese or pinyin)
- It then returns the complete biographical data in CBDB for the person requested:
  - Date of birth/death, alt. names, addresses, entry, postings, writings, kinship, and social association

# To make an API call to CBDB

- Simply send a http request (a URL) to CBDB:
  - Query by person ID:
    - <http://cbdb.fas.harvard.edu/cbdbapi/person.php?id=1762> (this is the ID for Wang Anshi, no disambiguation required)
  - Query by name:
    - <http://cbdb.fas.harvard.edu/cbdbapi/person.php?name=王安石> (this may require some disambiguation by the user)
- <http://cbdb.fas.harvard.edu/cbdbapi/person.php?name>

# HTML Output

1762

CBDB ID: 1762

索引/中文/英文名稱: /王安石/Wang Anshi

指數年 (index year): 1080

生年: 北宋元祐5年 (1021)

卒年: 北宋元祐1年 (1086)

享年: 66

郡望: 太原

出處: 宋人傳記資料索引(電子版), 頁1536

註: Wang(2) Anshi [1762] Yi(3)'s [7082] son, Guan(1)'s [1841] grandnephew, Anguo's [7076], Anli's [1760], and Anshang's [1761] brother, Fang's [1803] uncle, Jue(1)'s [1796] great grandfather, Zhu(1) Mingzhi's [526] and Shen(2) Jichang's [1445] brother-in-law, Cai(1) Bian's [8131] father-in-law and uncle-in-law, and Wu(3) Anchi's [1957] father-in-law. Anshi was Zhang(1) Jifu's [195], Xu(3) Xi's [753] and Yang(2) Wei's [2032] patron, Li(2) Ding(1)'s [1097] patron and teacher, Gong(2) Yuan's [941] teacher, and the teacher of Wang(2) Pin's [7381] uncle, Boqi [3970]. He was the grandnephew of Wang(2) Guanzhi [3965], the father-in-law of Zhou(1) Jiazheng's [480] son, Yanxian [3250]. His mother was Wu(3) Shi and his wife was Wu(3) Shi. He once arranged for a marriage between his wife's younger sister and Wang(2) Ling [3967] whose scholarship he admired. The sister's, and therefore Wang Anshi's wife's, grandfather, Wu(3) Min [4017], had the same generational name and address as Anshi's maternal grandfather, Wu(3) Tian [7396], so that it can be assumed that they were brothers or cousins. When Wang Anshi composed the funerary inscription for the two women's father, Wu(3) Ben [4020], in 1054, neither daughter was married. Wang(2) Ling died in 1059 at the age of 28. Wang Anshi was 32 in 1054. Wang(2) Ling was Wu(3) Yue's [1992] maternal grandfather. The mother of Yan(5) Shu's [2073] grandnephew, Fang [4118], was the younger sister of Wang(2) Anshi's wife, Wu(2) Shi. Songshi yi, 5.5b discusses the coalitional significance of the fact that Lu(9) Jiawen's [1294] son married the daughter of Wang Anshi's son, Pang(a) [3968], and that Cai(1) Bian [8131] married Pang's elder sister. Anshi was a friend of Wang(2) Ping's [1856] son, Hui [3958] and his older and younger brothers were tongnianyou with the sons of Chen(1) Jiansu [7101]. XCB, 177.2a, 179.3a, 184.15a, 188.8a, 189.3a, 18a, 191.9a-10a, 192.4a-4b, 7a; XCBSB, 4.4b, 15.17a, 16.14b; Shen Gou, WJ, 2.48b; SHY:ZG, 5.1a; Zeng Gong, WJ, 45.4b; Du Dagui, 'xia,' 14.1a; Wang Anshi, WJ, 97.998, 98.1012; Wang Ling, WJ, 'fu,' 13b. CBD, 1, 277-281. From Hartwell's ACTIVITY table: 1075: Working on the railroad all the time all the time all the time

## Basic Information

別名: 字介甫, 室名、別號半山老人, 諡號文, 小字雅郎。

## Alternative names

地理資訊: 籍貫(基本地址): 宋朝 / 江南西路 / 撫州 / 臨川

出處: , 頁

註:

## Addresses

任官:

- 參知政事: 地點: 宋朝。起始年: 1069。赴任。

出處: 宋人傳記資料索引(電子版)(頁1536)。

註: Hartwell defined the office as Canzhi zhengshi (參知政事)

- 三司度支判官: 地點: 宋朝。起始年: 1058。赴任。

出處: 宋人傳記資料索引(電子版)(頁1536)。

註: Hartwell defined the office as Duzhi panguan (度支判官)

- 正授 群牧判官: 地點: 宋朝。起始年: 1054。終止年: 1055。赴任。

出處: 未知。

註: Hartwell defined the office as Qunmu panguan (群牧判官)

- 同中書門下平章事: 地點: 宋朝。起始年: 1071。赴任。

出處: 宋人傳記資料索引(電子版)(頁1536)。

註: Hartwell defined the office as Tong Zhongshu Menxia Pingzhang shi (同中書門下平章事)

## Postings

# HTML Output (continued)

親屬關係:	兄(B+):	王安仁	<a href="#">出處:宋人傳記資料索引(電子版)(頁1535)。</a>
	弟(B-):	王安禮	<a href="#">出處:宋人傳記資料索引(電子版)(頁1543)。</a>
	弟(B-):	王安國	<a href="#">出處:宋人傳記資料索引(電子版)(頁1538)。</a>
	姪女之夫(BDH):	蔡京	
	姪孫;從孫(BSS):	王景	<a href="#">出處:宋人傳記資料索引(電子版)(頁993)。</a>
	曾姪孫女婿(BSSDH):	石斗文	
	女兒(D):	王氏 (吳安持妻)	
	女婿(DH):	吳安持	
	女婿(DH):	蔡卞	
	父(F):	王益	<a href="#">出處:宋人傳記資料索引(電子版)(頁812)。</a>
	從祖;伯叔祖(FFB):	王貴之	<a href="#">出處:宋人傳記資料索引(電子版)。</a>
	十世孫(G+10):	王熙	<a href="#">出處:元人傳記資料索引(頁732)。</a>
	母(M):	吳氏 (王益妻)	<a href="#">出處:宋人傳記資料索引(電子版)。</a>
	子(S):	王雱	<a href="#">出處:宋人傳記資料索引(電子版)(頁954)。</a>
	子(S):	王滂	
	孫女婿(SDH):	呂嘉問	
	曾孫;重孫(SSS):	王珪	<a href="#">出處:宋人傳記資料索引(電子版)(頁765)。</a>
	妻子(W):	吳氏 (王安石妻)	<a href="#">出處:宋人傳記資料索引(電子版)(頁1536;1537)。</a>
	妻兄(WB+):	吳顯道	
	岳父(WF):	吳黃	
	連襟;妻姊妹之夫(WZH):	王令	<a href="#">出處:宋人傳記資料索引(電子版)(頁600)。</a>
	妹夫(Z-H):	朱明之	<a href="#">出處:宋人傳記資料索引(電子版)。</a>

## Kinship

社會關係:	未詳:	李常	
	未詳:	林之奇	
	未詳:	邵材 (1103)	
	同年友:	王越石 (1080)	
	拒絕在Y的主政期出仕:	孫立節	
	拒絕在Y的主政期出仕:	俞汝尚	
	拒絕在Y的主政期出仕:	劉恕	
	相唱和:	吳顯道	
	同學、同門:	胡舜元	
	同學、同門:	譚揆 (1080)	
	同學、同門:	王介 (1091)	
	推薦:	呂升卿	
	推薦:	成倬 【熙寧間,王安石用事,以其通經術,置諸門下。懇歸,安石借其志未遂,特薦得右選。】	
	推薦:	段充	
	推薦:	錢公輔 (1080)	
	推薦:	錢公輔 【舉錢公輔自代狀 / 臨川集】	
		<a href="#">出處:宋人傳記資料索引(電子版)(頁22346)。</a>	
	推薦:	韓宗師 (1080)	
	推薦:	劉彝 (1080)	
	推薦:	劉彝 【舉屯田員外郎劉彝狀 / 臨川集】	
		<a href="#">出處:宋人傳記資料索引(電子版)(頁21454)。</a>	

## Social associations

# Other output formats: XML and JSON

- XML and JSON are two **machine-readable data formats** for carrying **structured** data.
- By outputting XML and JSON, CBDB API provides the flexibility of **refining the presentation** according to the needs of each each database.

# Make an API call to get external data

## Ming Qing Women's Writings

Home Introduction Project Search

54563

### AUTHOR [ BAO ZHIHUI ]

鮑之蕙，清乾隆嘉慶人，女，漢族，字仲姁，號蕙香，丹徒人(今江蘇丹徒)  
三女史詩鈔，國朝閩閣詩鈔，國朝閩秀正始集，隨園女弟子詩選，小黛軒論

#### [ Individual collections in this database authored by 鮑之蕙 Bao Zhihui ]

清娛閣吟彙：六卷(清鮑之蕙撰)。嘉慶16年[1811]刻本  
Qing yu ge yin gao : 6 juan (by Bao Zhihui, Qing dynasty) - (All texts in this collection)

#### [ anthologies in this database containing works by 鮑之蕙 Bao Zhihui ]

國朝閩秀正始集：二十卷，附錄：一卷，補遺：一卷，又附題詞：一卷(清  
Guo chao gui xiu zheng shi ji: 20 juan, fu lu: 1 juan, bu yi: 1 juan, you fu ti  
鮑之蕙 Bao Zhihui )

國朝閩閣詩鈔：一百種一百卷(清蔡殿齊輯)。清道光二十四年(1844)刻本  
Guo chao gui ge shi chao: 100 zhong 100 juan (edited by Cai Dianqi, Qing  
清娛閣吟稿：一卷 Qing yu ge yin gao: 1 juan - (All texts in this collection)

隨園女弟子詩選：六卷(清袁枚輯)。清嘉慶道光年間(1796-1850)坊刻巾箱  
Suiyuan nü di zi shi xuan: 6 juan (edited by Yuan Mei, Qing dynasty) - (All texts in this collection)

京江鮑氏三女史詩鈔合刻(清戴燮元輯)。清光緒八年(1882)刻本。  
Jingjiang Bao shi san nü shi shi chao he ke (edited by Dai Xieyuan, Qing  
清娛閣詩鈔 Qing yu ge shi chao - (All texts in this collection authored by Bao Zhihui)

#### [ remarks and biographies on 鮑之蕙 Bao Zhihui ]

《歷代婦女著作考》P. 762 著錄 作者: 胡文楷

鮑之蕙[略傳] Bao Zhihui [lue zhuan] 作者/Author: 甘晉 Gan Jin

論鮑之蕙詩 Lun Bao Zhihui shi 作者/Author: 陳芸 Chen Yun

鮑之蕙[略傳] Bao Zhihui [lue zhuan] 作者/Author: 惲珠 Yun Zhu

CBDB ID: 54563

索引/中文/英文名稱: /鮑之蕙/Bao Zhihui

指數年 (index year): 1810

生年: 清嘉慶22年 (1757)

卒年: 清嘉慶15年 (1810)

享年: 54

出處: 明清婦女著作數據庫, 頁MQWW PoetID #99, HuWenKai #762

註: 《清代人物生卒年表》定其生卒年為乾隆22年~嘉慶15年。【參考《清代人物生卒年表》#23845。】

### CBDB人物傳記

別名: 字仲姁, 未詳姬香, 室名、別號蕙香。

地理資訊: 籍貫(基本地址): 清朝 / 江蘇省 / 鎮江府 / 丹徒

出處: , 頁23845

註: 江蘇丹徒(清代人物生卒年表)。

親屬關係: 兄(B+): 鮑之鍾  
從女: 姪女(BD): 0少韞 [出處]: 明清婦女著作數據庫(頁MQWW PoetID #4810)。  
父(F): 鮑皋  
丈夫(H): 張鉉  
翁: 公公; 丈夫之父(HF): 張堂  
族子(K+1(male)): 鮑桂星 [出處]: 明清婦女著作數據庫(頁MQWW PoetID #4807)。  
族子(K+1(male)): 鮑文達  
母(M): 陳蕊珠  
姊(Z+): 鮑之蘭  
姊(Z+): 鮑之芬

社會關係: 書序由Y所作: 鮑之鍾  
書序由Y所作: 吳錫麒 [出處]: 明清婦女著作數據庫(頁MQWW PoetID #99, HuWenKai #762)。  
書序由Y所作: 吳烜 [出處]: 明清婦女著作數據庫(頁MQWW PoetID #99, HuWenKai #762)。

**Data Legend**

Map Layers

- Society & Demographics
  - CH\_NSong\_Exams
  - NSong Density Exams
  - CH\_Ming\_Exams
  - Ming Density Exams
  - CH\_Qing\_Exam\_2
  - Qing Density Exams
- Rivers, Streams, Lakes
  - Rivers\_1820
- Boundaries
  - Provincial boundaries in 1911
  - County boundaries in 1911
  - Prefectural boundaries in 1911
  - Skinner\_Macroregions
  - v3\_1820\_prov\_pgn\_utf
  - Prefectures\_1820
- Transportation
  - Ming-Qing PPostal Routes Main lines
  - Ming Qing Postal Routes - Minor
  - Ming Qing Postal Stations
  - China Railroads 1990
  - Courier\_Routes
- Place Locations
  - Three Chan Schools in Jin Dynasty
  - Prefectural seats in 1911
  - County seats in 1911
  - China 1820 "Busy" Xian
- Utilities & Infrastructure
  - Global\_NightLights
- Imagery & Base Maps
  - General
    - Jinhua shi City Map
- Background
  - OpenStreetMap
  - Google Roadmap
  - Google Terrain
  - Google Hybrid
  - ESRI World Imagery
  - Google Satellite

Enter search... Search Reset



