

The China Biographical Database User's Guide

中國歷代人物傳記資料庫 用戶指南

[中文版]

傅君勸

(Michael A. Fuller)

北京大學 CBDB 編輯組 譯

二〇一一年四月

© 中國歷代人物傳記資料庫 (CBDB)
哈佛大學 中央研究院 北京大學

序 言

包弼德 (Peter K. Bol)

作為關係型數據庫，“中國歷代人物傳記資料庫”（以下簡稱 CBDB）能夠根據不同檢索條件反饋相應的傳記資料。這些檢索條件可以是簡單的問題如“哪些人來自同一地方”，也可以是更為複雜的問題如“在某一時段來自同一地點通過科舉入仕的人之間有著怎樣的社會和親屬關係”。用戶可以訪問在線數據庫（<http://isites.harvard.edu/icb/icb.do?keyword=k16229>）進行檢索，也可以從網站下載包括用於空間及網絡分析的查詢表格及應用工具的整個數據庫，在任何安裝有微軟 Access 的計算機上使用。《用戶指南》將向讀者介紹單機版數據庫的結構及應用。

CBDB 為關係型數據庫。該數據庫將中國歷史上衆多男性和女性的傳記資料根據不同的方面進行分類和編碼。在正式展開介紹數據庫的設計、使用和各種細節之前，希望用戶能夠瞭解以下幾點：

1. 將人置于背景中觀察的方法 CBDB 是一種將生命歷程模型化的方式，也是一種組織信息的方式。數據庫的主體是社會中的人，然而我們將其視為數據庫中的“實體”（Entity），這些實體與其親屬和社會團體、其工作和居住的地點、其生存和活躍的時段、被給予和行用的姓名、撰寫的著作、進入政府及其他機構的方式，以及社會區分相關聯。與生平敘述不同，CBDB 將“人”視為被關係網絡界定且能夠被量化和分析的實體。

2. 時間範圍 CBDB 中 90% 以上的數據屬於唐代（618-907）至 20 世紀早期。截至 2011 年 1 月，CBDB 共包含約十萬個人物，另有約七萬個人物的信息正在處理之中。已有人物的更多信息也隨工作進度及時更新。提取較早時代的人物信息將與官名表和地名表的準備工作同時進行。

3. “仿真陳述”（Factoids）與“事實”（Facts） 和世界上其他地區的群體傳記數據庫一樣，CBDB 在大部分情況下主要處理材料中的“仿真陳述”，即某一“事實”（如蘇軾是眉山人）所涉及的材料對這一事實的說法。CBDB 將這些說法聯繫在一起——包括了那些互相矛盾的表述而不是對其可靠性進行判斷。不過，CBDB 並不將所有材料視同一律。

4. 主要材料 CBDB 的開發始於已故的郝若貝教授（Robert Hartwell）對中古中國歷史的研究。此後，CBDB 開始廣泛地從已出版的索引（如王德毅先生等學者編撰的《宋人傳記資料索引》及其他同類著作）、在線數據庫（如中研院“明清人名權威檔案”及京都大學“唐代人物知識庫”）、文

本資料專著(如墓誌銘合集)以及正史中的傳記資料中提取數據。儘管哈佛大學和北京大學的 CBDB 編輯們正在嘗試從其他材料中挖掘數據，上述主要材料仍然需要一段時間才能處理完畢。

5. 文本挖掘(Text-Mining) 利用計算機“文本挖掘”技術從中研院提供的電子文本中揀選信息是目前建設 CBDB 最有效的手段。在美國“國家人文基金”(National Endowment for the Humanities)的資助下，計算機專家正在進行這項工作。哈佛大學編輯團隊與計算機專家一起設計合適的“正則表達式”(Regular Expressions)，北京大學編輯團隊檢查據表達式標記的電子文本，管理人員則負責最後給載入 CBDB 的數據編碼。這一工作流程並不保證材料中所有的信息都被發現，但載入數據庫的信息會準確反映所挖掘的材料。

6. 誤差幅度 機器與人工相比，在篩分大量數據時具有優勢，卻無法給予學術判斷和解釋。數據庫中是可能存在錯誤的：史料本身即可能不正確，編輯者在標記的過程中可能會漏過這些錯誤，編碼器可能未能適當地對兩個同名的實體做消歧處理。用戶必須確保數據庫的查詢返回了足夠的結果，在此情況下誤差範圍才不會損害根據結果所得出的結論的可信度。材料記載與 CBDB 數據間的差異是明顯的，而 CBDB 中的數據也經過了相當長時間的修訂，在應用計算機技術之後這些差異被控制在 1% 以下。我們不妨這樣考慮：根據一千個事例得出的觀點，即使其中有十個錯誤，也比根據有一例錯誤的十個事例得出的觀點好些。

7. 數據庫而非辭典 CBDB 可以用作關於某一個體傳記信息的指南，且能夠比傳記辭典提供更多的人物各方面關係的數據。然而，辭典的標準在於在人物信息所有方面的完全準確，而對數據庫的期望則在於，其提供的案例在範圍和數量上都是大規模的，因而對研究十分有用。

《用戶指南》將解釋以下三方面的內容：CBDB 作為關係型數據庫的邏輯性，CBDB 內容的結構，以及用於從 CBDB 中獲得數據的主要查詢界面。

目 录

序 言	i
導 言	1
第一章 關係型數據庫.....	3
A. 關係型數據庫及複雜數據的組織.....	3
B. 關係型數據庫的結構規則	5
C. 關係型數據庫及複雜數據的互動.....	6
第二章 CBDB 的結構.....	9
A. 數據庫“實體”概要.....	9
B. “實體”的詳細說明.....	10
小結：CBDB 中的表格	23
第三章 CBDB 的分析工具	26
A. 使用 LookAtEntry 窗體	27
B. 使用 LookAtAssociations 窗體	32
C. 使用 LookAtOffices 窗體	37
D. 使用 LookAtKinship 窗體	42
E. 使用 LookAtNetworks 窗體	49

導 言

中國歷代人物傳記資料庫（CBDB）是關於前近代中國傳記信息的關係型數據庫。通過大範圍收集數據，CBDB 提供許多檢視過去個體或群體生平的方法。在提供人物具體信息及作傳記辭典使用外，CBDB 更強大的功能在於其能夠作為“群體傳記學”（一定人群生平研究）的分析工具：

“‘群體傳記學’是指通過對一群人物生平的集體性研究，來探討他們的共同背景特徵。其採用的方法為：先建立一個研究範圍，然後詢問一組系統的問題——關於出生與死亡、婚姻與家庭、社會出身與其所繼承的經濟地位、居住地、教育、個人財富的數量與來源、職業、宗教、仕歷經驗等等。之後將這一範圍內所有個人的各種信息對比、組合，並探析其重要的變數。研究者會分析這些信息的內在相關性，以及與其他類型的行為和活動的相關性。”（L. Stone, “Prosopography”, in F. Gilbert and S. Graubard eds., *Historical Studies Today*. New York, 1972)

CBDB 還支持另一種近年來用於分析大量人物生平的研究取徑。“社會網絡分析”（SNA）過去幾十年在社會科學領域一直用於研究群體結構。最近，學者將這一技術用於分析從歷史文獻中獲取的資料。Charles Wetherell 如此形容“歷史社會網絡分析”（HSNA）：

（歷史社會網絡分析）將“共同體”（Community）界定為人際關係的集合……並為歷史學者提供了研究的藍圖，以此能夠評估過去的人們在他們的生命歷程中的哪些時間、出於什麼樣的原因、以怎樣的方式運用親屬和非親屬關係。從事社會網絡分析的學者發現，人們需要並且尋求不同類別和來自不同人群的情感及經濟支持，這一發現提示了新的分析的必要性。如今只是關注危機時期人們如何運用親屬關係已不足夠，歷史學家更應當探究歷史上的人們如何在人生歷程中，出於不同的目的，在由當時人口和文化條件帶來的機遇與限制之下運用親屬和朋友關係。其他取徑或許也可用於處理這一問題，但是“歷史社會網絡分析”不僅包含了深化這一討論的必要視角，也能夠幫助歷史學家回應 Tilly 提出的挑戰：將普通民衆的生活與長時段的變化以有意義的方式結合起來。（Charles Wetherell, “Historical Social Network Analysis,” *International Review of Social History* 43 (1998), Supplement)

在很大程度上，歷史學家已經將社會網絡分析的方法用於所有成員間關係均為已知的小樣本人群，然而 CBDB 希望能提供極大規模人群中個體社會關係及親屬關係的數據，並且這些數據的密度足以得出關於前近代中國社會模式的有統計學意義的結果。

由於 CBDB 記錄了諸如住址、求學地、仕宦地、任官、父母、配偶和相識人的信息，因此這些生活層面會與相當大的人群相關。利用 CBDB，用戶能夠查詢在特定時段或某一地區，或是特定層級的官員、職業當中是否存在典型的本地聯姻情況；查詢任何時段及（或）區域中各類職業的親屬模式；查詢與黨派對立或支持有關的區域模式；查詢在僧侶階層內部晉升中社會、親屬和地域因素的影響；查詢哪些人將自己與特定的思想家、老師相聯繫以及這些人的生活區域。總而言之，用戶幾乎可以對數據庫中的人物進行任何條件的查詢。

目前的挑戰在於如何提出 CBDB 能夠回應的問題。《用戶指南》旨在為您提供從普通的界面查詢到利用其他工具對數據集做進階查詢的各種信息。有關 CBDB 的說明分為三部分：關係型數據庫的概要介紹、CBDB 的結構（其包含的數據類型），以及 CBDB 的查詢界面。

第一章 關係型數據庫

A. 關係型數據庫及複雜數據的組織

社會史學家郝若貝（Robert Hartwell）教授很早就開始關注宋代官員的親屬和社會網絡并最早提出利用關係型數據庫研究群體傳記。CBDB 即源自他最初使用的數據庫模型。

郝若貝教授的重要進展在於認識到自己需要一個有力的組織工具來實現他所提出的研究計劃。他希望能夠觀察人與人之間關係，人們的親屬、社會網絡、曾任官職以及與他們有聯係的地方。這一長串的目標，加上各個元素之間的互動關係，增加了追蹤這些問題的複雜度和難度。郝若貝教授意識到可以將他在傳記資料中所見的互動關係分為以下幾類：人物、地點、官僚系統、親屬結構和社會網絡，后兩項均被認定為社會制度。他建立了一個專門提取這五者之間關係的關係型數據。

組織不同實體間的關係正是關係型數據的主旨：在世間各種相互關聯的複雜事物之間建立多種形式的關係。例如，“地點”是“事物”中的一類，在這一類別下，我們列舉包含信息或我們感興趣的地址。類似的，“人物”也是“事物”的一類，在此類別下我們列舉包含傳記信息的人物，接著就能列舉所有我們關心的人物和地名之間的聯係：出生地、遷居地、葬地等等。以下是實體間關係的抽象模型：

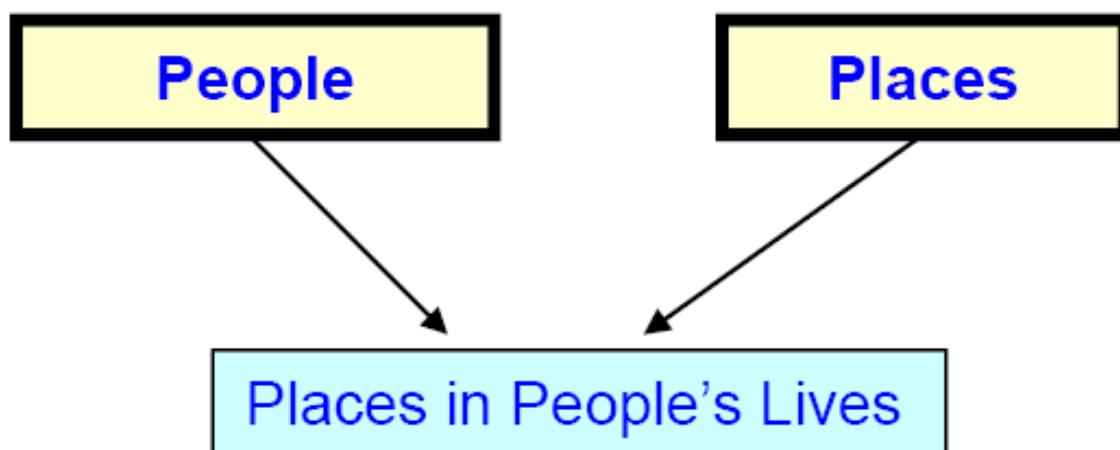


圖 1

這一抽象模型在轉化為關係型數據庫后，變成一系列填有數據的表格：

PEOPLE		
ID	Name	Dates
1	Lü Benzhong 呂本中	1084-1145
2	An Dun 安惇	1042-1101
3	Chao Buzhi 晁補之	1053-1110
4	Chen Jian(5) 陳薦	fl. 1069

PEOPLE-PLACES		
Person ID	Place ID	Relation Type ID
1	1	1
1	3	2
1	2	3

PEOPLE-PLACE TYPES	
Relation Type ID	Relation Type
1	Basic Affiliation
2	Moved to
3	Ancestral addr

PLACES	
ID	Placee Name
1	Jinhua 金華
2	Shouzhou 壽州
3	Kaifeng 開封

圖 2

請注意：經過表格的整理，人物、地名及二者關係的類別都沒有數量限制了。

通過以上關聯人物與地名的例子，我們看到關係型數據庫中有三種基本的表格：

- 1、描述基本實體的表格。（如表2中“人物”與“地名”的黃色表格）在CBDB中，這些表格包括人物、地名、親屬類別、官僚制度等等。這些表格中的字段抓取了我們想知道的“實體”的屬性。人物表包括姓名、生卒年、性別等。地名表包括了某地的行政級別、其上下級區劃及設立的起始時段。官名表包括了某官職在特定朝代的行政等級。
- 2、描述基本實體之間關係的表格。（藍色的“人物—地名”表格）在 CBDB 中，這些表格將人物與其社會、物理及文化環境的關係轉譯為結構化的格式。表格中的字段抓取了這些關係的特徵，對描述關係有重要作用。例如，當某人在官僚機構中得到了一個職位，除了此人和他官職的基本信息外，我們也想知道這一職位在哪裏？此人是否赴任？他何時任職？
- 3、描述實體之間關係類型的表格（粉色的“人物—地名類型”表格）世上兩事物有時存在多種互動方式，我們需要更具體地記錄這些相互關係的細節。在以上的例子中，人物與地名的關係有多種形式：可以是他們的籍貫、遷住地或葬地。我們可以將這些聯繫分成不同的類別並建立結構。

B. 關係型數據庫的結構規則

在數據庫中，我們力圖做到任何特定的數據只被記錄一次。如上面的例子所示，“晁補之”這個名字只出現在 CBDB 的一筆記錄中，也就是“人名”實體中的基本實體（這張表也被稱為“BIOG_MAIN”）。關於晁補之的其他所有信息通過他的 ID 與其發生關聯。因此，假如出現錯誤，也只需要在一處更正即可。比如，在記錄“洪适”的數據時，羅馬字母本應該輸入 Hong Kuo，因為以為“适”是“適 *shi*”的簡化字，卻誤輸成 Hong Shi，那麼只需更正一處即可。這個“一條數據，一筆記錄”的原則被稱之為標準化。儘管在某些情況下，為了加快工作進度，CBDB 也會違反這個原則，但是如果用戶想添加額外的表格到個人的 CBDB 版本中時，我們強烈建議考慮數據庫標準化的優點。

在以上提到的人物與地名關係的例子中，我們面臨的實際情況是，一個人往往會遷徙到很多地方。這就是“一對多”的關係。有兩種選擇可供我們通過簡單的行列表格呈現這種關係：第一種是在基本的傳記表格中創建一些“縱列”（比如“遷徙地 1”、“遷徙地 2”等），第二種是把所有的欄目添加進一個獨立的單元格。如果我們為“遷徙地”創建一些列，我們很有可能無法處理特殊情況，比如遇到一個人物他遷徙了很多次，以至於超出了我們所創建縱列的數目；此外，在傳記表中的每一筆獨立的記錄中都包括所有的“遷徙地”單元格，對於許多人物來說，這些也很有可能是空置的；如果我們僅為“遷徙地”的信息創建一個縱列，那麼在這個欄目中進行檢索，將會使重新獲得關於每個人的數據變得很困難。鑒於以上兩種保存“遷徙地”方法的缺點，CBDB 的原則是：只要我們在基本實體（這裡指“人物”和“地名”）中發現這類一對多關係，我們需要一個獨立的實體，例如“人物-地名”（以及用與呈現這一實體的表格），使得我們能夠獲取二者相互關係的信息。

當為朱熹和呂祖謙合著的《近思錄》這樣的著作編碼時，我們會遇到新問題。著述有著一種被稱之為“多對多”的關係：一部著作可能有很多作者或編者，並且每一個編著者都可能不止一種著述。CBDB 與許多數據庫一樣，將這種情況作為一對“人名”和“著述”間的“一對多”關係處理，並且引入一個新的實體“人名-著述”以獲取這些數據。

如果想給 CBDB 添加新的數據類型，那麼以上三條原則——數據標準化、為“一對多”關係創建新表格、像處理“一對多”關係那樣來處理“多對多”關係——是很重要的。

C. 關係型數據庫及複雜數據的互動

試想如下一組實體及其與基本實體“人物”之間的關係：

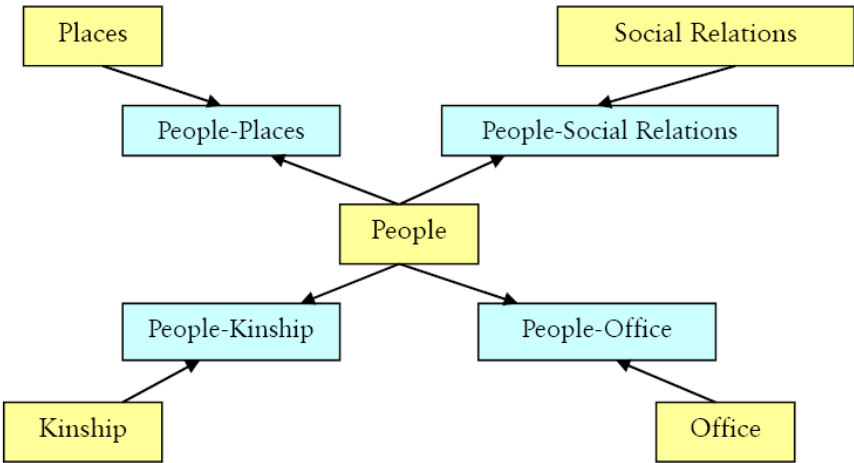


圖 3

儘管在“親屬”和“職官”之間沒有直接的聯繫，我們仍然可以通過已經採集到的人物數據發現二者的關係。由此，我們可以提出類似這樣的問題：醫官的職位是世襲的嗎？或者說，醫官的子侄仍舊是醫官嗎？還有，醫官家庭之間是否聯姻呢？那些擁有中級武階的人們又是怎樣的狀況：由武轉文的人是否更傾向于將他們的女兒嫁給擁有文職的人呢？

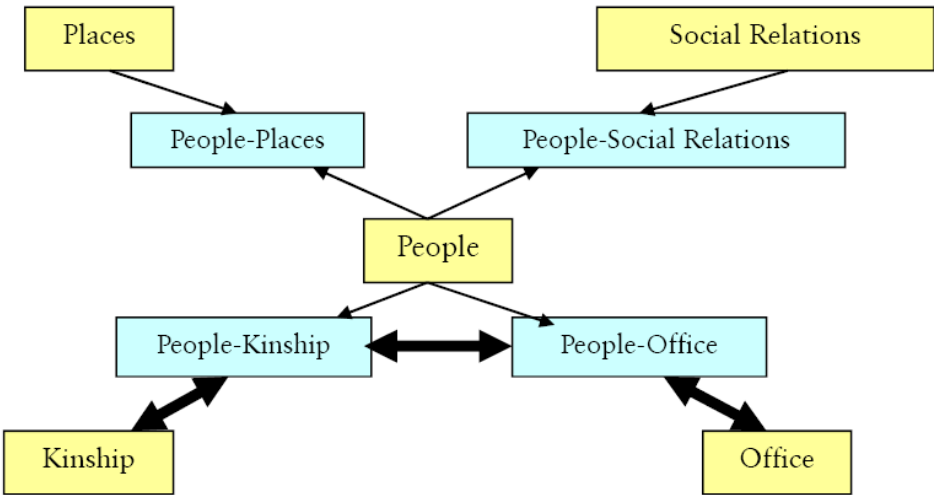


圖 4

查詢“職官”和“親屬”之間的關係

我們可以問出很多很多關於“親屬”和“職官”二者關係的問題。高層官員和低層官員內部是否存在不同的婚姻模式？這些群體組成的婚姻聯盟是否形成了新的階層？這些模式會隨時間而改變嗎？我們也可以針對“地址”和“社會關係”的關係提出類似的問題。比如，四川人更傾向與同鄉交往還是會建立全國性的關係網？這些關係類型從北宋初期到北宋後期是否發生了變化？從北宋後期到南宋後期呢？

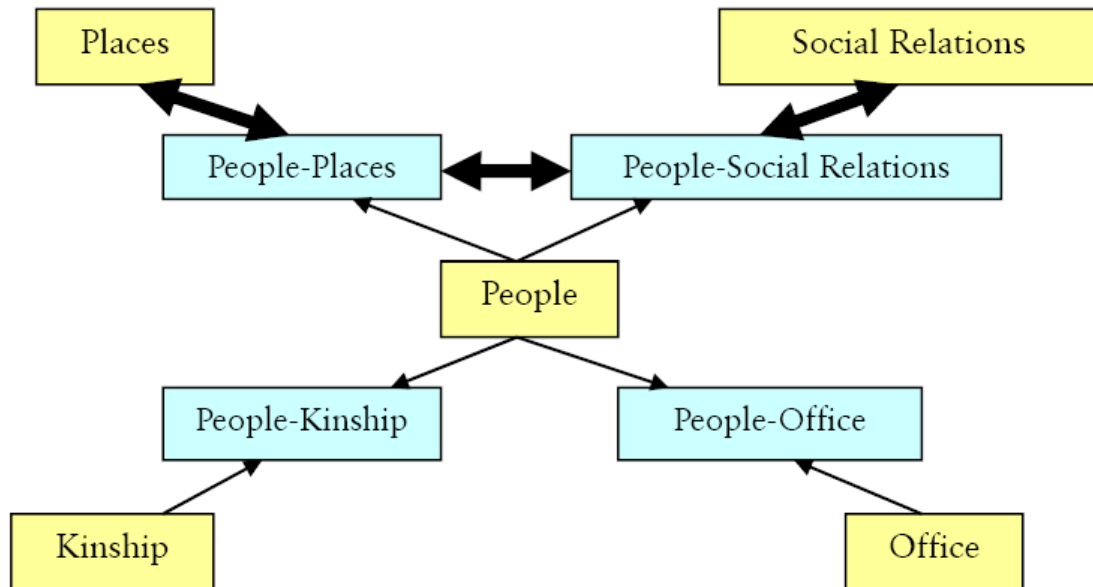


圖 5

查詢“地址”和“社會關係”之間的關係

最後，我們可以查看多種因素的相互作用，比如“地址”在“親屬”和“職官”的關係中所起的作用：

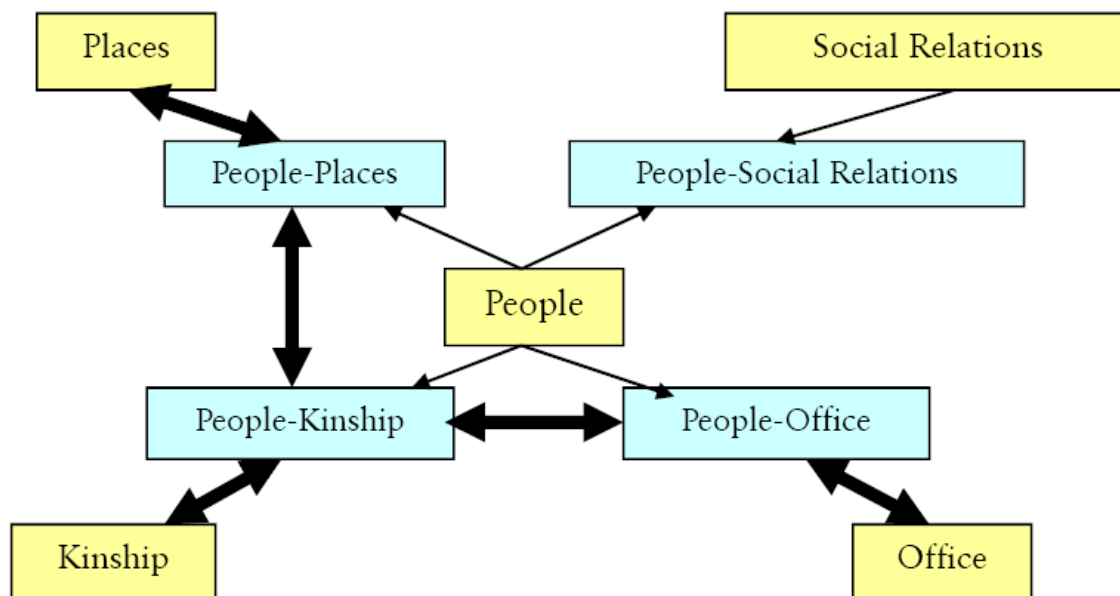


圖 6

查詢“地址”在“親屬-職官”關係中所起的作用

可以設問：來自福建的官員是否更可能與來自浙江的官員建立地方性的親屬關係網絡？其模式是否會隨著官員等級及時間的變化而變化？

在關係型數據庫中，能否對 CBDB 中各種實體的相互關係進行有效提問，僅取決於個人在多大程度上瞭解這一數據庫及其數據的組織方式。

第二章 CBDB 的結構

A. 數據庫“實體”概要

數據庫設計用表格把抽象的對象（我們簡稱為“實體”）具體化。數據庫的目標是獲取關於實體間相互關係的信息，因此將抽象的對象與呈現他們相互關係的表格相分離是有效的。這樣，人們可以更容易地發現如何改進表格才能更好地代替他們所要呈現的實體。

定義數據庫中人物傳記的核心實體如下：

1. 人物

由於關係型數據庫追蹤人們與其他人、他們所在的組織（例如社會和經濟機構）以及物理世界發生關係的途徑，我們也需要建立一些能與“人物”相互作用的實體，首先是人與人之間的關係（這些實體將在後文詳細討論）：

2. 親屬

3. 非親屬關係

其次是人與政治、社會文化組織及活動的關係：

4. 社會區分（社會區分的模式，例如以“書法家”或者“僧侶”而出名）

5. 進入行政或其他事業的方式（例如通過科舉考試）

6. 任官（例如地方官或將軍）

7. 人物參與的重大事件

8. 財產

還有人寫作的文本——借此能夠進一步了解這些人物：

9. 著述（包括撰著作品、編纂作品和古代文書資料）

還有人們賴以活動的世界的結構信息，也包括在 CBDB 中。我們所關注是政區地理和官僚制度這兩方面的信息：

10. 政區等級（以行政單位的政治術語來劃分）

11. 經緯度（為歷史比較提供空間定位）

12. 官僚組織（官僚體制的變化和相應的職責要求）

B. “實體”的詳細說明

說明：數據庫允許用戶記錄信息來源，允許用戶增添其他合適的注解。數據庫中記錄個體信息的每一個條目都有來源、頁碼和注解的屬性。因此接下來的討論中將不再解釋這些。

1. 人物

a. 基本資料：姓名、性別、生卒日期。

出生和死亡的精確時間通常無法獲得，我們所知道的是一個人活動的時間段。有時候，甚至連這也無從得知：我們僅僅知道年號或者朝代。為了獲得精確的資料，數據庫中所有的時間都允許使用年號信息。有的人物信息可能有某一年號的具體年份，但是也有可能的僅僅標識了“始”“中”“末”，或者“不詳”。為了分析方便，數據庫會以出生、死亡、活動時間的年號信息以及其他任何以傳統中國年號標誌的日期推算出公元日期，但是在年號代碼中會保留模糊信息。

b. 種族

CDBD 記錄的種族有漢、回鶻、吐蕃等。我們目前有 170 多個代碼，這些代碼都在“ETHICITY_CODES”表中。

c. 郡望

在六朝時，家族成員身份對確定一個人的社會地位至關重要。宋代之前，人們會說他們的祖先來自于某特殊地區的某特殊家族（比如崔姓來自於博陵），這種說法很難得到查證。地名和家族名的結合就構成了郡望。這些郡望名稱的代碼都在“CHORONYN_CODES”表中。

d. 指數年

為了計算的需要，CBDB 需要一個單一的年份值以在年代上定位人物。指數年（六十歲）是用於分析的一個人造數值。計算這些數值的規則較為複雜：

本人規則 1：本人卒年（如果享年 ≤ 59 ）

本人規則 2：本人出生年份+59

本人規則 3：本人的進士年+30

以父親的相關時間推算本人指數年：

父親規則 1：父親的指數年+30

父親規則 2：父親的生年+30（如果父親卒年 $>$ 本人卒年）

父親規則 3：父親的生年+89

父親規則 4：父親的卒年+ 29

以其他親屬的相關時間推算本人指數年

祖父規則 1：祖父的指數年+ 60

祖父規則 2：祖父的生年+ 119

祖父規則 3：祖父的卒年+ 59

兄規則 1：兄的生年+ 61

弟規則 1：弟的生年+ 57

子規則 1：長子的生年+ 29

婿規則 1：長女婿的生年+ 29

CBDB 記錄這些基本傳記資料的表格是“BIOG_MAIN”。“BIOG_MAIN”給每個人物都分配了唯一的 ID。

e. 在世年

CBDB 提供“在世始年”和“在世終年”。通常沒有資料顯示個人的指數年或者和生卒年時，但是文本中會提供可以確定某人時間的參考資料。CBDB 提供了在目前查證過的文本資料中可以獲知的最早和最晚日期。

2. 親屬關係

個人的親屬關係都由三個部分構成（加上資料信息）

本人

親屬關係

親屬

建立親屬關係有 9 種基本的類別：

e 本人（親屬關係正被搜尋的人）	F 父
M 母	S 子
D 女	Z 姐妹

B 兄弟	H 丈夫
W 妻 C 妾	

在這些類別的關係中有多種變化，我們也用附加性的符號來代表“核心家庭”（Nuclear Family）以外親屬關係的類型。

+	更長的（比如：B+ 兄）
-	較小的（比如：Z- 妹）
*	被收養的繼承人（比如 S*，養子）
°	被收養的
!	庶出
^	後、繼（比如 S^ 表示繼子）
½	比如 Z½，同母異父姐妹或同父異母姐妹
~	名義上的（比如 M~，合法的妻子是妾所生孩子名義上的母親）
%	有婚約的丈夫或妻子（在一定的時間內還沒有完成婚禮）
y	最小的（比如 Sy 是最小的已知的兒子）
1, 2, 3...	區別順序的數位（比如：S1, S2 分別代表第一個和第二個兒子；W1, W2 分別代表第一任和第二任妻子）
n	不能準確知道是哪代
G-#, G+#	直系的祖先 (-) 或者子孫 (+)
G-n, G+n, Gn	未知幾代前的直系親屬 (G-n)，或者未知幾代後的直系親屬 (G+n)，或者前後也未知直系親屬 (Gn)
G-#B, BG+#	在若干代 (#) 前有共同直系祖先的兄弟；兄弟的若干代 (#) 直系子孫
K, K-#, K+#,	宗親，與之前相同，前代為(-)，後代為(+)，未知的為(n)。 (CBDB

Kn	使用“宗親”一詞，因為有時有的親屬關係得到了證實，卻不能獲知準確的關係。)
K-, K+	同一代人的親屬關係，年齡較小 (-) 或者年齡較大 (+)。
P, P-#, P+#, Pn	與父親的姐妹或者母親的兄弟姐妹的親屬關係，前代(-)，後代(+)，未知(n)。用漢語的“表”來表示。(只有不能確切知道人物的關係時，CBDB 才會使用這些代碼)。
P-, P+	通過父親的姐妹或者母親的兄弟姐妹所生成的同輩人的親屬關係，年齡小的 (-)，年齡大的 (+)。
A	男方或女方的姻親，通過婚姻形成的親屬關係

這些關係類別的代碼在“KIN_CODES”表中。

親屬關係的資料在“KIN_DATA”表中。

3. 非親屬關係

a. 簡單的非親屬關係

這些由三部分構成：人物+關係名稱+關係人。在記錄個體人物在他們一生當中所形成的非親屬的關係時，最大的挑戰就是如何控制社會關係類別的增加。

因為這些社會關係總是在人與人之間成對出現，所以在交往中就一定存在著對稱關係。也就是說，如果資料庫中出現“A是B的學生”，那麼也應該出現“B是A的老師”。實際上，當前的系統中會自動生成第二個條目。因而，“關係”作為一種實體就有了一種內在結構：

1. 關係類別
2. 成對的關係類別
3. 關係的種類/子類（目前有三種）

當創建一條新的社會關係類別時，也一定要創建與之對應的社會關係。當然，對“相同關係”而言，其相對的社會關係正是本身：“A是B的朋友”也就是“B是A的朋友”。然而，在大多數關係中，兩個人扮演著截然不同的角色，CBDB 需要一種對應類別以便於兩人從不同的角度獲取各自的角色：在記錄A時用“A跟隨B”，這對B而言意味著“B被A跟隨”。有些關係並不是雙重的，因為這種關係並不是指向某一個人，而是指向一個更抽象或者普遍的事物。這種類別中最重要的關係就是黨派。因而，我們說{A “是元祐黨的成員” ∅}，這裏的∅

是一個無效元素。

b. 间接關係

在一些重要的事件中，關係是通過間接的组织或者人形成的。CBDB 通過增添附加資料到關係關係中，來獲取這些關係的類型。比如，我們可能知道 X 和 Y 之間的關係，是因為 X 要求 Y 為他父親的文集作序。

c. 未來的發展

1). 喪事

在中國，某人會參加其老師（或者通過其他非親屬關係認識的人）的喪事。數據庫呈現了與喪事有關的五方面因素：

哀悼者

被哀悼者

喪事的時間

服色

日期

或許可以證明這種關係類型在早期中國極為重要，但是我們可以預料到隨著時間的推移，中國文化的轉變，傳記會強調這些活動和關係的不同形式。數據庫需要考慮這些歷史的轉變。目前 CBDB 不能記錄“喪事”關係的具體細節，不過一旦獲得早期人物的更多信息，我們就會探尋最好的處理辦法。

2). 饋贈禮物

另外一種關係類型是通過饋贈禮物建立的。這種行為也包括五個方面：

送禮者

收禮者

禮物

禮物的價值（或數量）

日期

目前，CBDB 沒有記錄這些信息，而這些信息可能在早期的傳記材料中更為重要。正如喪事一樣，一旦獲得更多的資料，我們就會開始處理這一問題。

4. 身份

CBDB 有單獨的表格記錄個人的“社會區分”。“社會區分”被用來確定個人在社會中以何種身份被認知。儘管日期通常不確定，但是表格中有一欄可以記錄發生次序。此外，有些形式的“社會區分”會把各種角色相疊加（一位佛教徒可能因為他的書法而出名，或是一位文人創辦了一家印書坊）。目前，CBDB 在不同類別下記錄身份的不同方面。這一問題有待於進一步的研究：最好可以混合兩種不同的社會角色，但其操作性仍是問題，涉及到混合角色的出現頻率、相關人物是否一直都被認為具有一個單獨“身份”而不是兩個。個人身份資料的構成是：

人物

身份代碼

身份次序

日期

資料的來源和注解

“身份”作為一種社會經歷類別（相對於特定人在社會區分結構中的身份而言）是一個簡單的“實體”，包括：

身份代碼

身份描述

身份種類和子類 1

身份種類和子類 2

因為社會區分隨著時間不斷變化，CBDB 在處理更早或者更晚的資料時就需要增加目前社會區分的類別。

5. 入仕方式

“入仕”本身是一個簡單的實體，僅僅包括姓名、入仕類型和子類型。目前“入仕”主要描述進入政府部門的途徑，但 CBDB 也已經開始探索像僧官制度這樣的類型。不同的入仕途徑包含了不同類型的信息，使得個人入仕的情況更為複雜。比如，某人通過科舉入仕，如果可能的話，我們需要知道科舉的類型和中第時間（CBDB 也記錄落第情況）。相反，某人得益於他人而入仕，如果此人及其與入仕者的關係可知，也應被記錄下來。因此，如果 A 是通過他的伯父 B 蔭補入官，那麼這條入仕信息就是：

人物：A 的 ID

入仕途徑：“蔭補”的代碼

入仕關係類型：“伯父”的代碼

入仕關係：B 的 ID

由於個人還有可能通過非親屬關係的蔭補入仕，“入仕事件”就需要一種方式來記錄這種非親屬關係。最後，入仕有很多種屬性，但是任何個例都只涉及到其中的一部分：

人物 ID

入仕類型代碼

入仕關係類型代碼（適用於親屬）

入仕關係類型代碼（適用於非親屬）

入仕關係 ID（適用於親屬和非親屬）

科考日期（公元年及年號+年份）

科第名次

6. 職官和除授

CBDB 目前囊括了 6000 多個官名，（至少在宋代部分）還包含這些官名在政府官僚結構中的位置。“除授”這一實體包括了人、官僚機構和地點（因為大多數都遠離首都）。即一個人在一定的時間、地點和官僚層級中供職。然而，有時候一次“除授”授予的管轄權超出一個行政單位，或是一次“除授”承擔一個以上的行政職務。遵循“一對多”的關係（如：“一次除授<>多個地址”和“一次除授<>多個官名”），我們需要建立單獨的表格，除授信息需要三個實體：基本的“除授”（**postings**）表，“任官”（**posted-to-office**）表和“任官地點”（**posted-to-office-address**）表。由於行政單位通常總是明確地與職官相聯繫（而不是除授），因此地址表格記錄“除授-職官”信息和位址之間的聯繫。

除授數據

除授 ID（這是獨有的數字）

人物 ID

出處和注解

除授-職官

除授 ID

職官 ID

職官類型

序列（因為通常我們只知道職官的順序，而對於除授年份沒有更具體的數據）

年份（公元年和年號+年份：因為個人擔任某官時可能會有其他的職務）

出處和注解

任職-地址

除授 ID

官名 ID

地址 ID

未來發展的思考

1) 佛教和道教

佛教和道教管理機構的職位最終會被添加到“官名/官職/除授”實體中。不過，這也需要有價值研究成果來闡明佛教和道教管理機構的歷史變化。

2) 追蹤官僚組織的歷史變化

我們需要再次考慮，數據庫應該在多大程度上反應中華帝國官僚體制的複雜性。從漢朝到清朝的官僚體制中，雖然一個職官的名稱不變，但其職責可能會發生變化。或者職責保持不變而官名發生變化。學者曾經抱怨 Charles Hucker 在其《中國古代官名辭典》中嘗試把職能的延續性強加於某一官名之上，而簡單地承認這種轉變可能會更有用。因此，我們計劃在 CBDB 中提供能夠反映官名功能歷史變化的表格。我們認為“官名”是個實體，而“官職”是另一個實體。大多數特定時期實際的官職和 CBDB 無關，因為這些細節對數據庫的分析能力並沒有多大幫助；重要“官”的屬性有（1）官是否被當作酬勞和等級的象徵？是否具有職能？（2）向什麼樣的官負責；（3）官的類型（中央禁軍，州級文官，等等）。目前，CBDB 獲取這種信息很少，因為澄清官名的變化本身就是一個重要的研究項目。

7. 地點

CBDB 使用“中國歷史地理信息系統”（CHGIS）的方法來給地點編碼，這種策略依靠兩類空間實體：“地址”和“地點”（經緯度）

“地址”這一實體，是有地名的歷史“場合”——空間中有特定名稱的行政轄區。其位置由 x、y 座標的交匯點來確定。如果其邊界或者名稱中任何一個改變，就要建立一個新的地址。然而，這些歷史場合是行政層級的一部分：這些信息都保存在“belongs-to”表格中，這張表格和 CHGIS 中的“part-of”表格功能相同。某一區域只有面積或名稱發生改變時，地址 ID 才會變化，當其成為別的上級行政區的一部分時，地址 ID 並不發生變化。因此，CBDB 中有兩張表格：

地址表

地址代碼

地址名稱

行政類別

X 座標

Y 座標

地址始年

地址終年

從屬表

地址代碼

從屬地址代碼

從屬始年

從屬終年

CBDB 的一項主要目標是允許考察跨越朝代的變化趨勢，因此數據庫需要一種方法檢索在較長的一段時期內某個特定區域發生的事情。CBDB 依靠“地點”實體實現這一要求：地點指向一個物理位置，即地圖上的 x-y 座標。Hartwell 教授已經記錄了相應朝代的行政代碼及 1990 個縣的邊界，CBDB 目前只用這 1990 個縣作為地點單位。目前數據庫的“地點表”包括 1990 個縣和其他 Robert Hartwell 代碼表之外的縣，此表涵蓋了歷史上所有的縣（或是處於行政區劃底端的地點）。地點單位的最重要屬性是其 x-y 座標。這樣我們就有了地點表格：

地點代碼

地點名稱

X 座標

Y 座標

CBDB 有一張最終的表格把地址和地點信息合到一起，這僅需要兩個實體：

地址代碼

地點代碼

總之，CBDB 有四張表格構成了部分對地點的基本描述：

行政單位：ADDR_CODES

行政層級：ADDR_BELONGS_DATA

地球上的物理地點：PLACE_CODES

政區在物理地點上的映射：ADDR_PLACE

未來發展的思考

應該強調的是當 CHGIS 項目完成時，我們希望用它的表格來代替 CBDB 內部的地理代碼系統。然而，可能需要面臨這樣一個重要的問題：CBDB 的方法不是基於 CHGIS 的。也就是說，CHGIS 並不關注長時間的趨勢，也沒有對應的 Place 和 Address_Place 表格。CHGIS 只有一張表格，記錄從一個位址代碼到另一個位址代碼的變化，但是這種方法不能滿足 CBDB 的需要。比如，設想如果中央機構為節省管理費用，決定把三個人口減少的縣 A、B 和 C 合併成一個大的縣 D。假如三十年後政府意識到這樣做是錯誤的，把縣 D 再次按照原來的界限重現劃分為 A'、B' 和 C'。A 的變化表格會記錄 A→D 和 D→A'，但是沒有辦法重建 A→A' 的事實。為了提供經度資料，CBDB 需要跟蹤隨著時間的推移相應行政區所對應的物理位置。

8. 傳記地點信息

人物和地點之間有很多關係：他們在哪里出生、生活、死亡、安葬，他們在哪里供職，他們在哪里擁有財產、經營買賣，他們到哪里遊歷過。這些與地點的關係產生于 CBDB 中用單獨的表格記錄的各項活動（如：擁有的職官和財富），所以，這些信息會出現在那些不同的表格中，而不是只出現在一個地方。記錄人物和地點信息的表格有：

關於地點的基本傳記信息(BIOG_ADDR_DATA)

任職地 (POSTED_TO_ADDR_DATA)

在建立非親屬關係時的地點(ASSOC_DATA)

人們參與事件的相關地點(EVENT_ADDR)

人們擁有財產的地點(POSSESSION_ADDR)

注意目前的 CBDB 並沒有保存有關人們到哪里進行過短暫遊覽，在哪里接受教育，或者在哪里寫了作品等信息。CBDB 也無法為用戶把所有地點信息的資源整合到一起。

9. 文本

與數據庫有關的文本類型主要有三種：碑刻及其他古文書資料，印刷的原始文本和二手著作（印刷和數碼形式都可以）。如黃宗羲《宋元學案》這樣的著作既是一部梳理早期作品的學術綱要，又帶有個人創作的性質；而古代文書資料也由數據庫感興趣的作者撰寫。對前現代文本類別的區分，既無法十分明確，也沒有多大用處，因此 CBDB 把以上三種類型都當做“文本”。我們認為“文本”有以下屬性：

標題

作品類型（碑刻、手稿或印刷品）

流派

當前的出版日期

當前的出版者

當前的出版地

人們可能通過各種方式與文本產生聯繫

作者

出版者

編纂者

校對者

翻譯者

註釋者

文本的表格有：

文本代碼

文本 ID

文本名稱

寫作日期

當前的情況：存世還是已佚

當前的出版信息（如果存世）

文本資料

文本 ID

人物 ID

角色 ID

文本角色代碼

角色 ID

角色描述

未來發展的思考

碑刻材料有少許附加屬性被記錄在獨立的表格中，而這些表格還沒有得到開發，因為我們還不能處理這種類型的資料：

別名

發現地

發現日期

目前的位置

資料來源

這些材料也可能涉及其他的人物：

捐贈人

接收人

因為文本可以作為傳記信息的來源，所以 CBDB 記錄了所用現代版本的出版信息，也包括頁碼。然而 CBDB 並不希望把這作為傳記信息的標準參考書。CBDB（至少在現在和不久的將來）不會提供某一作者所有文本的現存版本，也不會判定在現存的版本中哪些最可信。未來計劃的一部分是在 CBDB 在線版與其他網絡資源建立鏈接。（書目網站當然會在這些鏈接當中）

10. 事件

CBDB 也是最近才開始記錄重要事件中的群體信息。目前列出的事件都來自于《宋史紀事本末》，這些事件由不同的類別構成，包括比如和議成功的技巧、謀反、皇后廢立，以及各種黨爭。我們在這些事件中發現有共同目的人形成了團體，因此這些信息值得保存。CBDB 通過三張表格記錄這些事件：

事件-代碼

事件 ID

事件名稱

事件日期

事件-數據

事件 ID

事件-記錄 ID（這是追蹤一件事例的唯一數字）

人物 ID

日期（有時需要一點時間把事件和人物匹配）

事件-地點

事件-記錄 ID

地址 ID（以防參與到一個事件中的人物會和多個地點相關聯）

11. 財產

CBDB 才剛開始搜集有關財產的信息。“財產”信息的一個層面是記錄在歷史文獻中的交易類型（購買、捐贈等）。我們可能會改變 CBDB 表格的結構以容納收集資料時得到的這類信息。

小結：CBDB 中的表格

1. 基本實體

這些實體代表了前近代中國社會的基本要素。每一要素都有著複雜的歷史和結構（這些都在另外的附屬表格中）。CBDB 在次級表中記錄了人們與其生活諸層面的相互作用。

表格名稱	註釋
ADDR_CODES	中國行政地理單位
ADDRESSES	顯示等級便於參考的表格
ASSOC_CODES	（非親屬）社會關係
BIOG_MAIN	前近代中國的人物
ENTRY_CODES	入仕的方式
EVENT_CODES	重大事件
KINSHIP_CODES	前近代中國的親屬類別
OFFICE_CODES	政府的官僚機構單位
PLACE_ID	帶有經緯度的標準單位的替代物
STATUS_CODES	人們取得社會區分的方式
TEXT_CODES	前近代中國著述的彙編+重要的研究著作

2. 基本實體間的相互關係

表格名稱	註釋
ADDR_BELONGS_DATA	行政單位等級結構的數據
ALTNAME_DATA	為人所知的其他稱呼
ASSOC_DATA	人們之間的非親屬關係
BIOG_ADDR_DATA	人與行政地理之間的關係
ENTRY_DATA	人與機構之間最初的關係
EVENTS_ADDR	人、事件與地址的關係

EVENTS_DATA	人和事件的關係
KIN_DATA	連結人們的親屬關係
OFFICE_DATA	官僚機構的等級結構
POSSESSION_ADDR	人名、財產和地址的關係
POSSESSION_DATA	目前可表述的“財產”狀況
POSTED_TO_ADDR	人、職官和地址的關係
POSTING_DATA	除授的集合表：人與職官的連結
POSTED_TO_OFFICE_DATA	人與職官關係的細節
STATUS_DATA	在社會聲望的體系中人所處的位置的數據
TEXT_DATA	人與著述的關係

3. 關係類型信息

表格名稱	註釋
BIOG_ADDR_CODES	人和地址關係的類別。
ALTNAME_CODES	為人所知的各種別名類別。
APPOINTMENT_TYPE_CODES	除授類別：正授、權，等等。
ASSOC_TYPES	寬泛的社會關係類別，將多種親屬關係代碼編組。
ASSUME_OFFICE_CODES	表示個人是否赴任。
ENTRY TYPE	寬泛的入仕類別，將入仕代碼編組。
EXTANT_CODES	表示資料的存佚情況
GENRE TYPES	文本的目錄學歸類
GENRE_CODES	寬泛的目錄學分類
LITERARYGENRE_CODES	文學作品的形式
OCCASION_CODES	人們參與的事件
OFFICE TYPES	職官類別

POSSESSION_ACT_CODES	人與物品之間的關係類別
SCHOLARLYTOPICS_CODES	學術主題類別
TEXT_ROLE_CODES	人與文本關係類別
YEAR_RANGE_CODES	日期的相對準確度

4. 歷史信息輔助表

表格名稱	註釋
CHORONYM_CODES	用於確認中古家族的“地名+姓氏”代碼
COUNTRY_CODES	國家代碼
DYNASTIES	朝代與日期的代碼
ETHNICITY_CODES	民族代碼
GANZHI_CODES	六十個干支代碼
KIN_MOURNING	五服等級中的親屬和服喪代碼
MEASURE_CODES	物品、貨幣、書籍和空間的數量單位
NIAN_HAO	年號
SCHOOL_CODES	學校與書院（將被加入）代碼

5. 分析輔助表

表格名稱	註釋
ASSOC CODE_TYPE REL	特定“社會關係”與更廣範疇社會關係的關係
ENTRY CODE_TYPE REL	特定“入仕法”與更廣範疇入仕法的關係
GENRE CODE_TYPE REL	特定“流派”與更廣範疇流派的關係
OFFICE CODE_TYPE REL	特定“職官”與職官類型的關係
OFFICE_CATEGORIES	職官的類別：階官、勳爵，等等

第三章 CBDB 的分析工具

“中國歷代人物傳記資料庫”包含了大量的信息。如果沒有分析這些信息的方法，這些信息將幾無價值。目前，CBDB 的 Access 版本特別設計了五種查詢窗體，以方便用戶查詢數據庫中重要的信息類別。這些形式的名字就反映了它們的功能。

1. **LookAtEntry** “查詢入仕”可以找到在特定時段通過特定途徑取得做官資格的群體。
2. **LookAtAssociations** “查詢社會關係”可以找到通過特定類別的社會關係而聯繫起來的群體。
3. **LookAtOffice** “查詢職官”不僅可以找到擔任特定官職的人，還可以找到官僚系統中擔任較低（或較高級別）官職的群體。
4. **LookAtKinship** “查詢親屬關係”可以查詢個體之間的親屬關係網絡。既包括五服以內的親屬關係，也包括更加擴展的關係。
5. **LookAtNetworks** “查詢關係網絡”可以為個人、群體，甚至一個特定區域找到所有的關係網絡（包括親屬關係與社會關係）。

除這五種形式之外，利用 Access 軟件，熟悉數據庫結構的用戶還可以設計查詢程序來檢索 CBDB 的所有信息。編寫“結構化查詢語言”（SQL）的方法需另辟專文介紹。

A. 使用 LookAtEntry 窗體

LookAtEntry（查詢入仕）是最簡單的形式。打開後點擊“Select Entry”按鈕來挑選一個入仕類別。

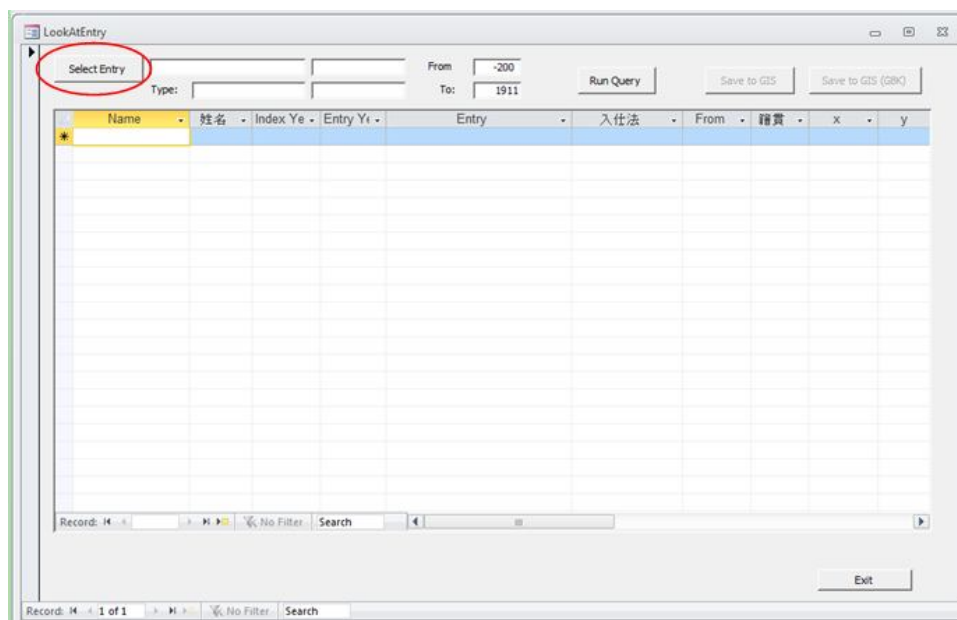


圖 1

可以看到，有多種不同的入仕法：

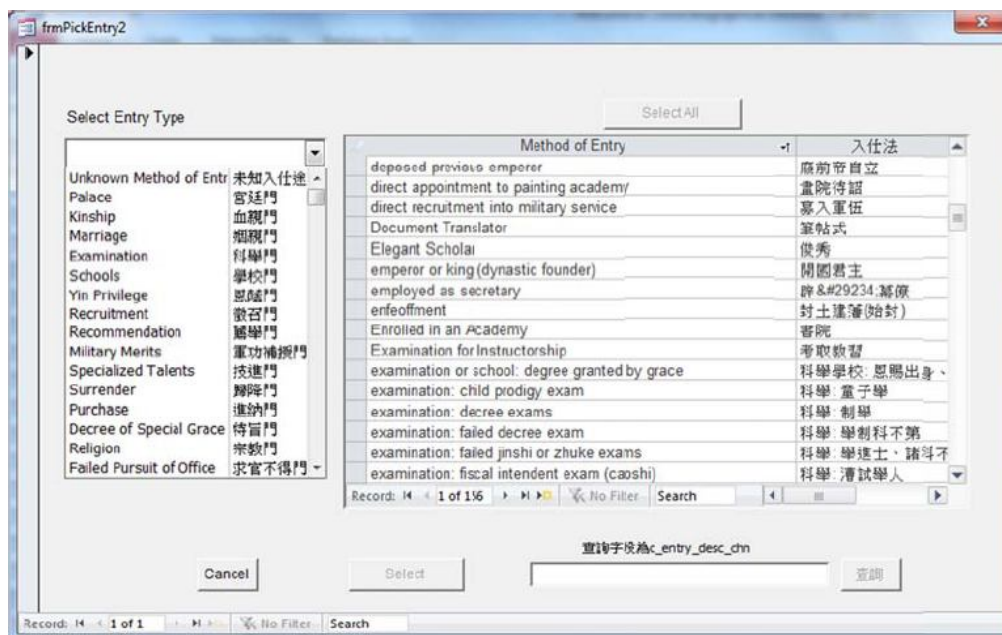


圖 7

在（1）中選擇一種特定的入仕類型可以簡化選擇：

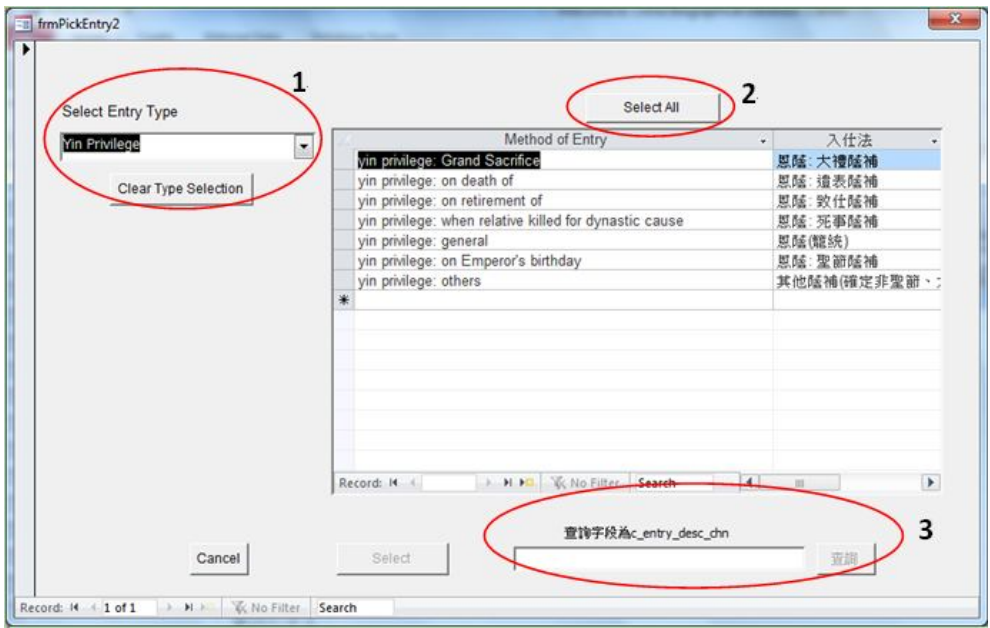


圖 8

用戶可以在普通類型中選擇一種入仕法，或者點擊（2）以選擇所有列出的入仕法。也可以在（3）中檢索某一特定的入仕法。在選定之后，回到 LookAtEntry 界面，在（4）中選擇年代，然後點擊（5）運行。

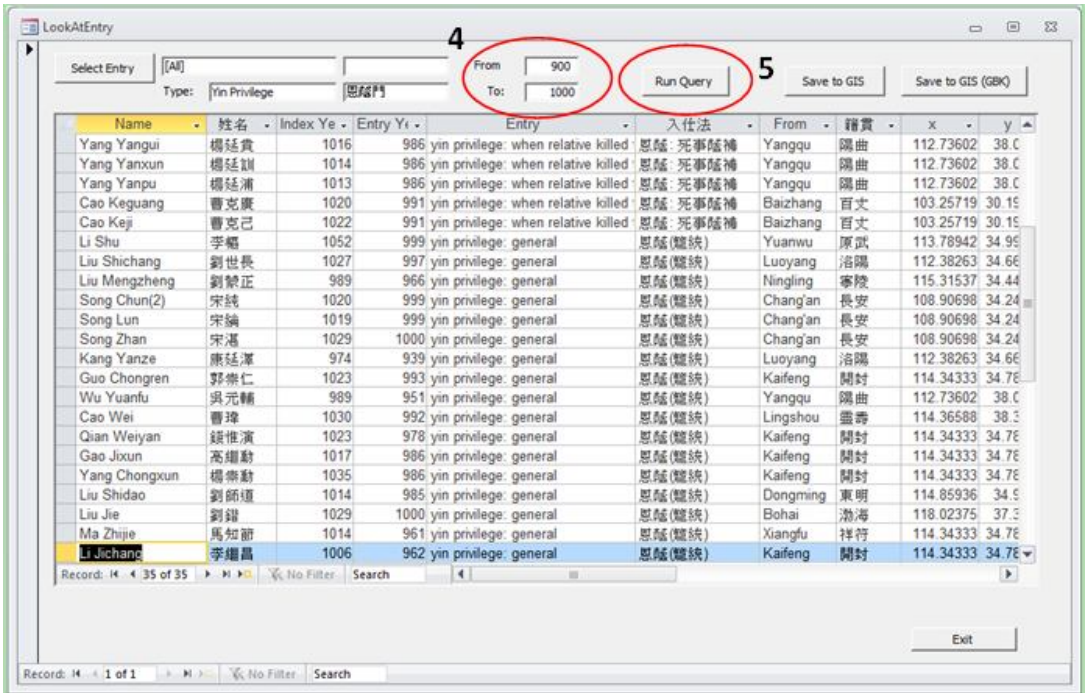


圖 9

檢索結果的表格包括以下 15 列：

人名（拼音）	
人名（中文）	
指數年	
入仕年	
入仕描述（英文）	
入仕描述（中文）	
籍貫（拼音）	CBDB 希望得到個人的索引地址（index-place），但是如果沒有相關記載，就使用 BIOG_ADDR_DATA 表中的第一個地址。
籍貫（中文）	
籍貫的橫座標	目前 CBDB 用橫縱座標的交匯點來確定地點，但是使用的是政區治所。
籍貫的縱座標	
座標總數	CBDB 計算表格中同一籍貫地的人數。
考試排名	如果知道的話就給出。這顯然和通過蔭補入仕的人無關。
親屬關係	蔭補特權，使個人可以通過某親屬的優勢獲得入仕資格。如果 CBDB 知道誰是親屬以及二者的關係，表格將給出這個信息。
親屬人名（拼音）	
親屬人名（中文）	

用戶可以利用任一欄目對表格進行整理。按“入仕年”排序可能便於使用。

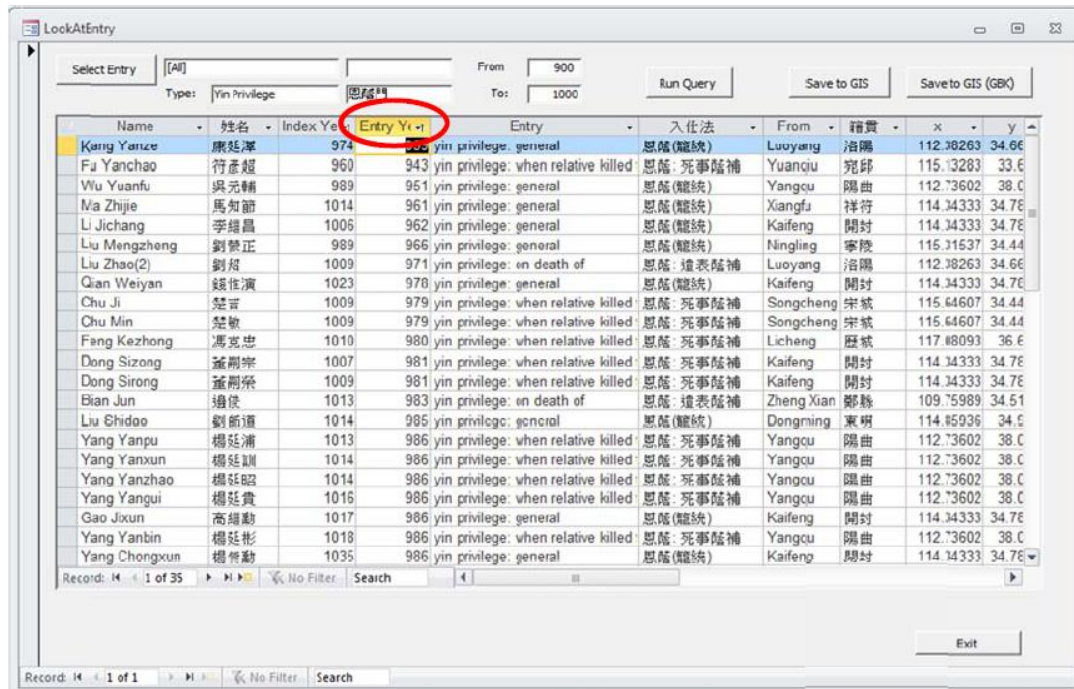


圖 10

如果想保存這張表格，最簡單的辦法是點擊表格左上角以選定整張表格，然後按 **Ctrl+C** 保存到剪貼板上。

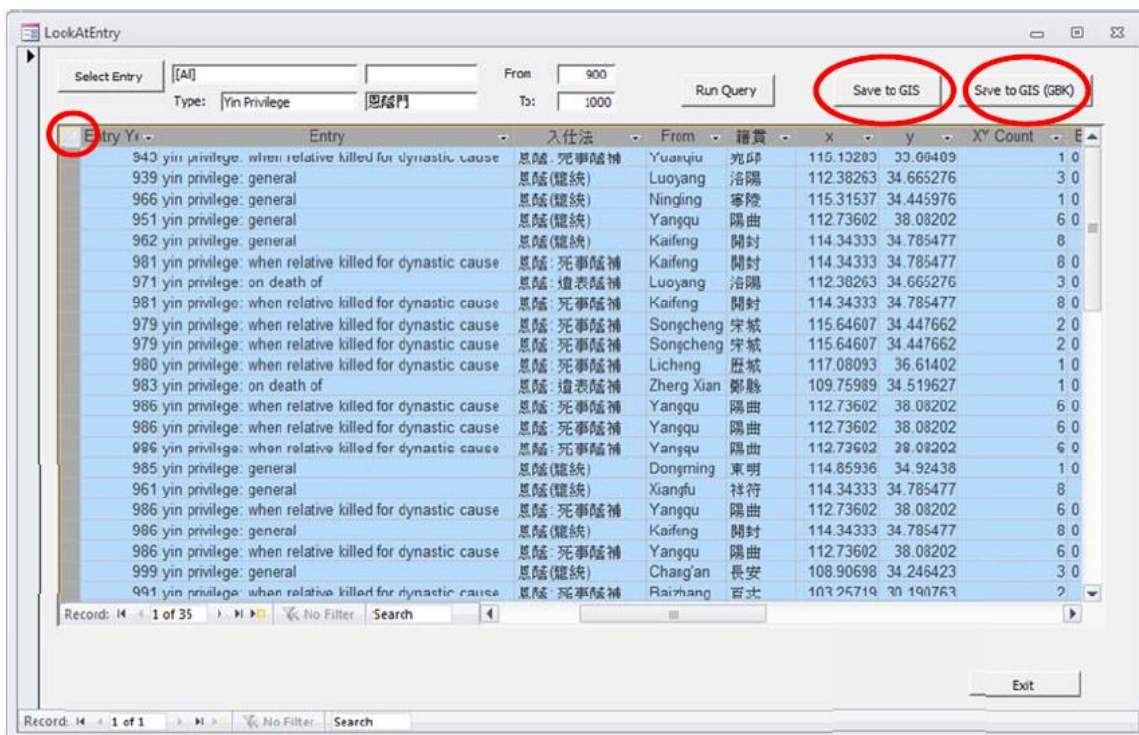


圖 11

這張表格可粘貼到任一識讀此格式的程序中。如果想將結果保存為能被 ArcGIS 識讀的文件，依據用戶是否想要使用漢字，選擇合適的命令鍵，這一表格就被存為文本文件，默認文件名為“entry_gis.txt.”

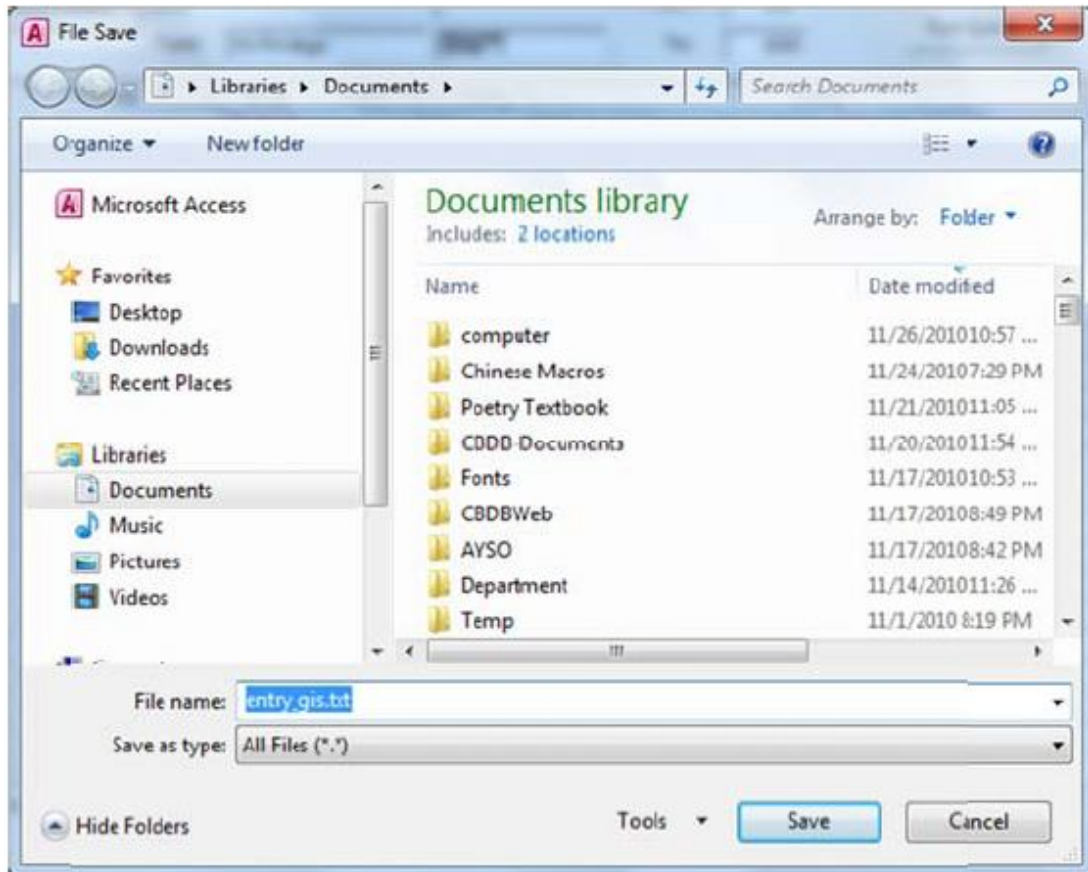


圖 12

B. 使用 LookAtAssociations 窗體

LookAtAssociations（查詢社會關係）窗體可使用戶查詢從屬於某一類特殊關係（或者幾種類型關係）的人物信息。打開LookAtAssociations窗體之後，點擊“Select Association”選擇想要查詢的關係類型。

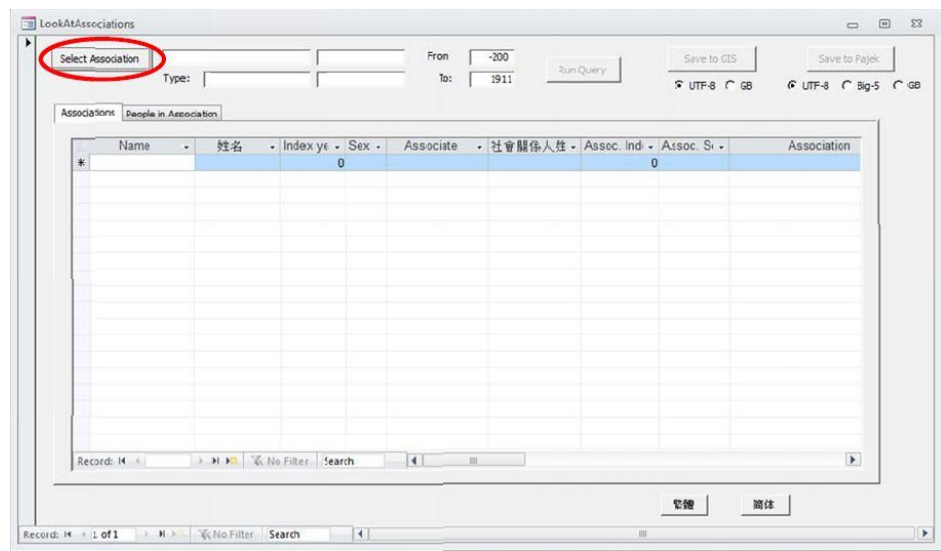


圖 13

CBDB包含了超過四百種社會關係類別，用戶可以按類別和子類別進行查詢。

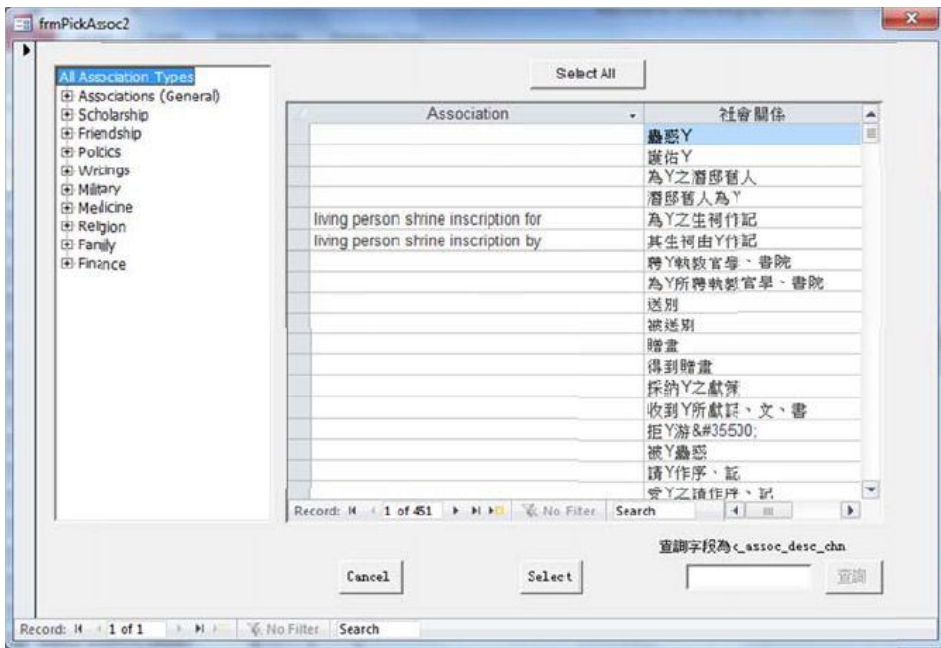


圖 14

以“軍事關係”為例，其中有三種子類型。“軍事對抗”子類則包含了八種關係。

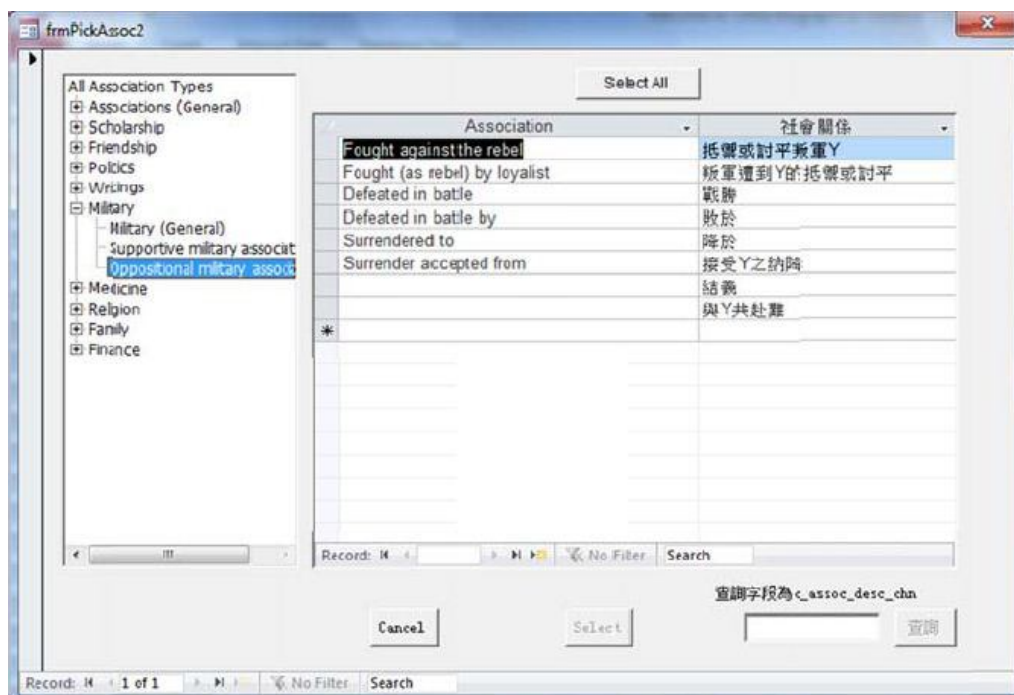


圖 15

在下面的視窗中，我們在區域(1)選擇所有可能的軍事關係，然後在區域(2)中把年代限定在900-1200年。最後點擊(3)執行查詢。

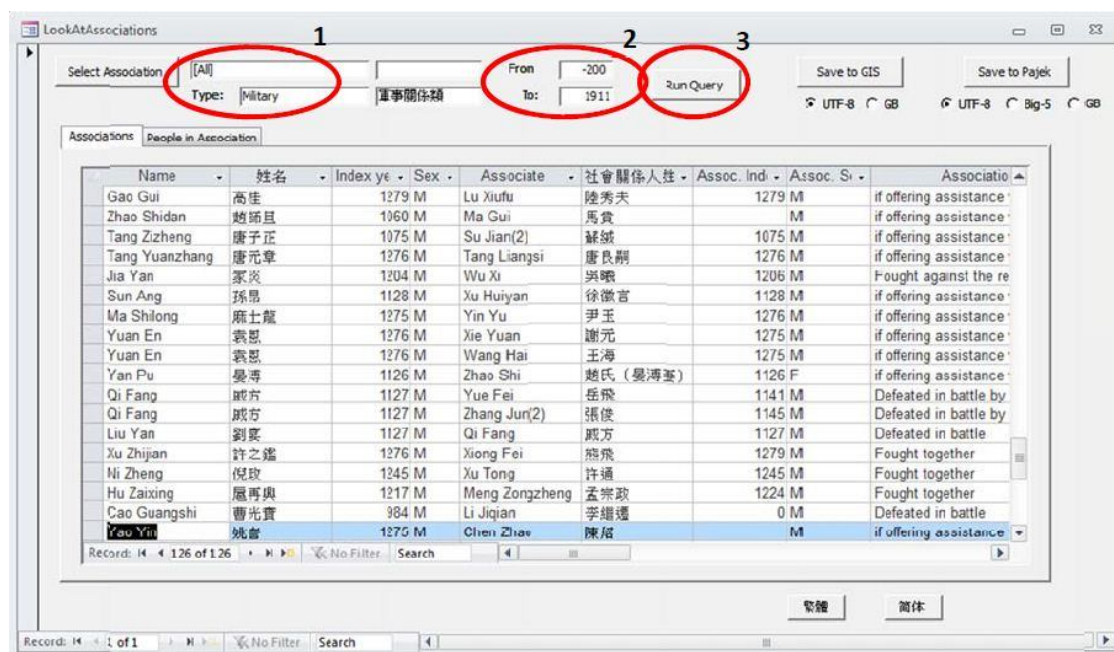


圖 16

雖然日期是ASSOC DATA表的一部分，對於大多數社會關係而言，我們無法得知具體的日期信息。LookAtAssociations使用個人的“指數年”來確定他們的關係是否在我們給定的年代範圍之內。LookAtAssociations中的關係表有18縱列，展示了ASSOC DATA中的關係類型。

Name (Pinyin)	姓名（拼音）
Name (Chinese)	姓名（中文）
Index Year	指數年
Sex	性別
Associated Person's Name (Pinyin)	社會關係人姓名（拼音）
Associated Person's Name (Chinese)	社會關係人姓名（中文）
Associated Person's Index Year	社會關係人的指數年
Associated Person's Sex	社會關係人性別
Association Category (English)	社會關係類別（英文）
Association Category (Chinese)	社會關係類別（中文）
Address (English)	地址（英文）。如果可能，CBDB 會選擇索引地址，否則會選擇個人第一條地址信息。
Address (Chinese)	地址（中文）
X-coordinate	X-Y 坐標。這是以上地址的地理中心座標
Y-coordinate	
Associate's Address (English)	社會關係地址（英文）
Associate's Address (Chinese)	社會關係地址（中文）
Associate's X-coordinate	關係地址 X 座標
Associate's Y-coordinate	關係地址 Y 座標
Kinship Relation (English)	親屬關係（英文）。以下四欄是為了親屬關係而設計的。
Kinship Relation (Chinese)	親屬關係（中文）

Kin Name (pinyin)	親屬姓名（英文）
Kin Name (Chinese)	親屬姓名（中文）
Associate's Kinship Relation (English)	社會關係人的親屬關係（英文）。以下四欄是為存儲關係人的親屬關係而設計的
Associate's Kinship Relation (Chinese)	社會關係人的親屬關係（中文）
Associate's Kin Name (pinyin)	社會關係人的親屬姓名（拼音）
Associate's Kin Name (Chinese)	社會關係人的親屬姓名（中文）

除此關係表之外，LookAtAssociations還將在此關係中涉及到的人物列成表。只要點擊“People in Association”選項即可。這張“人物表”同時也提供了附有地點的關係信息。

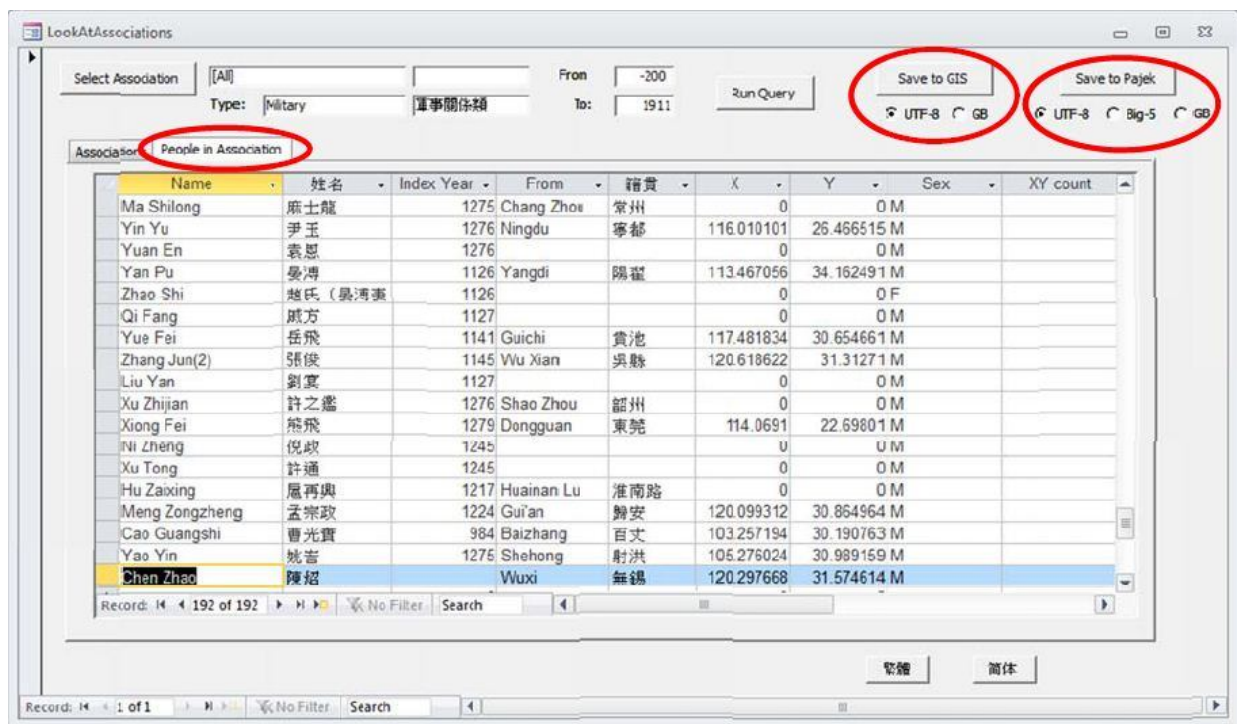


圖 17

人物表有以下9列信息：

Name (pinyin)	姓名（拼音）
Name (Chinese)	姓名（中文）

Index Year	指數年
Sex	性別
Place Association(pinyin)	社會關係地址（拼音）
Place Association(Chinese)	社會關係地址（中文）
X-coordinate	X坐標
Y-coordinate	Y坐標
XY-count	XY地址計數

點擊“Save to GIS”，我們可以將地址信息存儲為能用 GIS 程序展示的文件。社會關係數據把相關人群聯繫起來，展示出直觀上并不明顯的社會網絡。我們可以點擊“Save to Pajek”，將此網絡存儲為 Pajek 格式。Pajek 格式為在社會網絡分析(SNA)中實現可視化而設計，很多程序可以識別及將其轉換為其他格式。CBDB 允許 GIS 文件與 Pajek 文件用不同的字符編碼存儲，以便於使用漢字。

C. 使用 LookAtOffices 窗體

中華帝國的官僚制度非常繁雜，並且隨時間而變化。目前，CBDB記錄了超過六千個官職的編碼，並且數目肯定會隨著數據庫逐步擴展到整個前近代中國而繼續增加。由之而來的挑戰是：如何提供一種有效的手段去檢查官員在官職上的角色，使得繁蕪的職官可以被組織成較大的單元以供分析。LookAtOffices窗體提供了官僚等級和官職功能的分組。LookAtOffices窗體與其他簡單的分析窗體並無二致。用戶可以點擊“SelectOffices”來開始檢索。

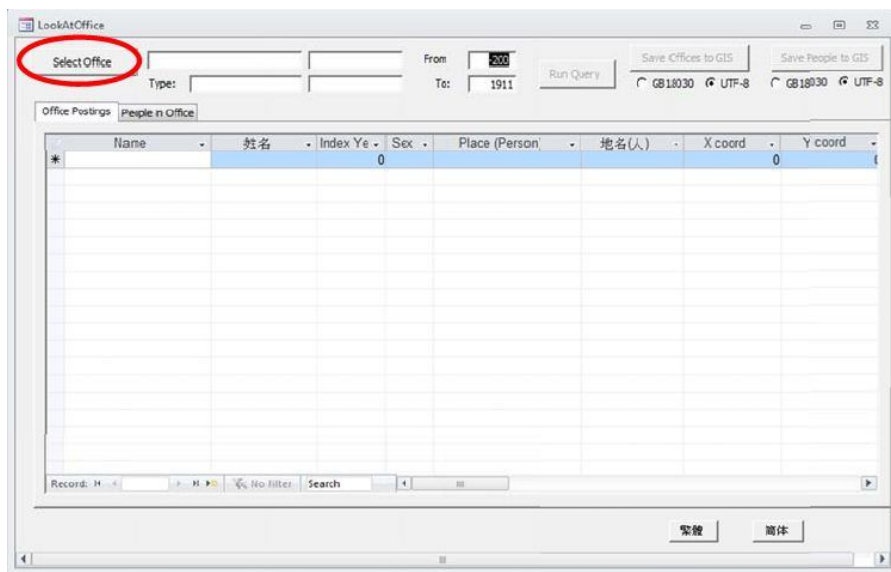


圖 18

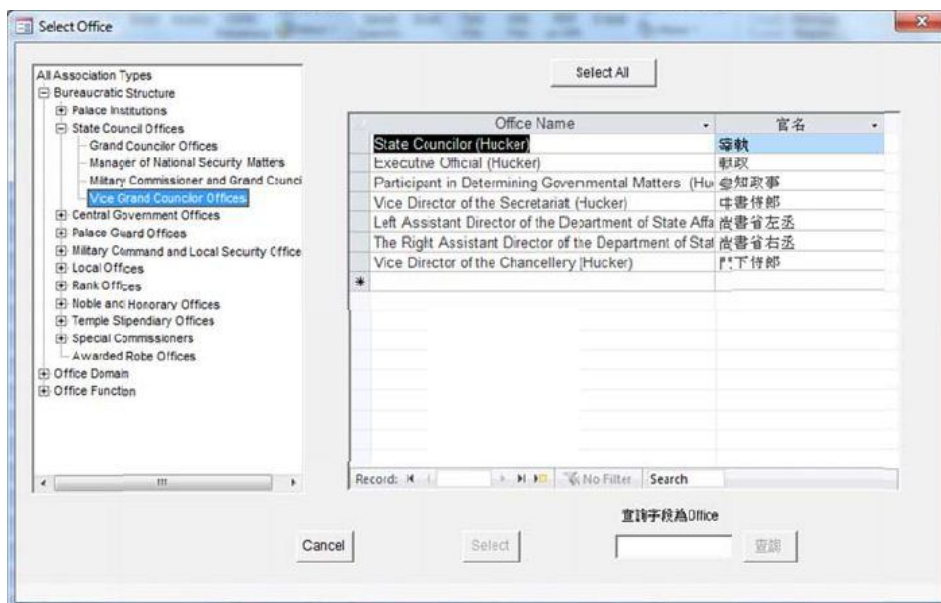


圖 19

“SelectOffices”窗體展示了職官類別的樹形結構。第一個節點，“Bureaucratic Structure”展示了官僚體制。我們可以查看並且選擇制度中任何一級的職位。上面的表格顯示了所有與“執政”(Vice GrandCuncilior)相關的職官。我們同樣可以通過職官所在地（中央、路、州或縣）以及類別（軍事、民政、宮廷等）來查詢。

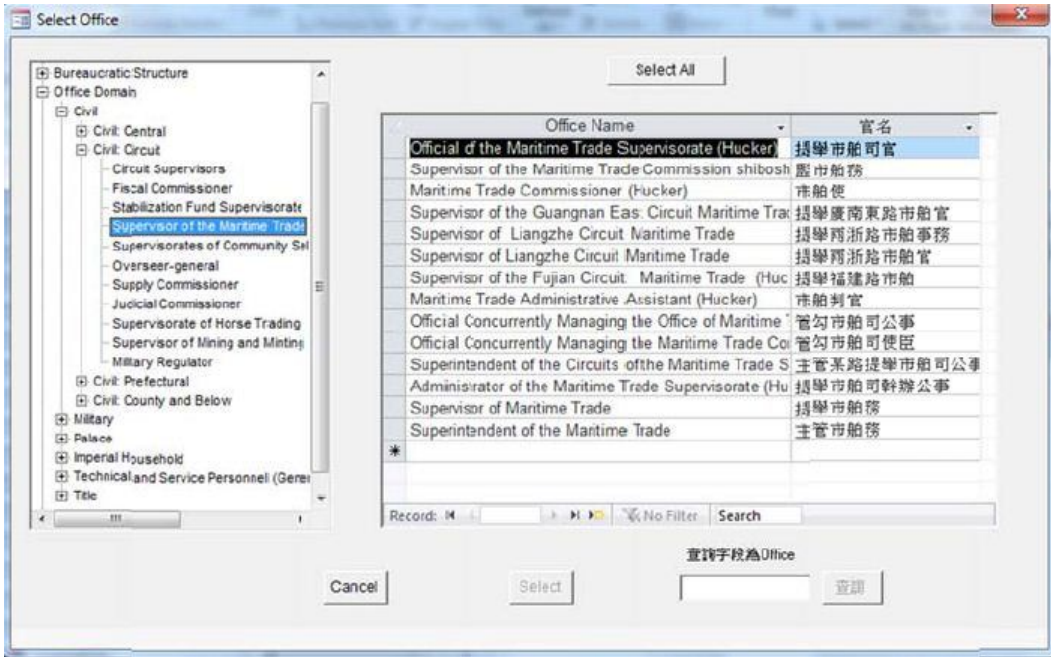


圖 20

下面的窗體是檢索所有與執政相關的職官之後得到的結果，起于公元前200年止于1911年。

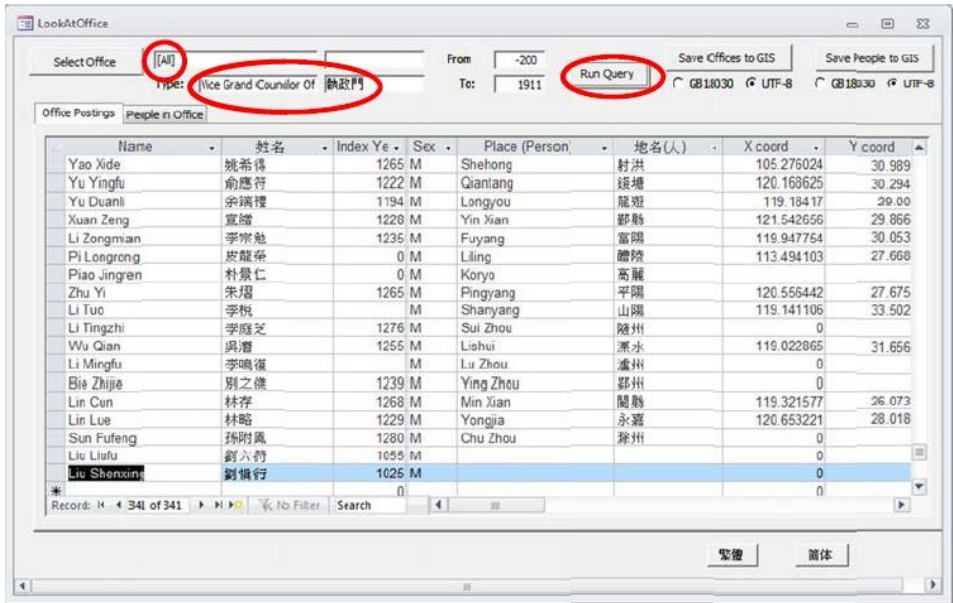


圖 21

這次查詢生成了兩張結果頁面，第一張頁面顯示了所有所查職官的除授情況，第二張頁面列出了獲得授官的人物信息。當我們想把相關人物導入LookAtNetworks窗體中以嘗試去研究某一特定官職的官員之間的聯繫時，這張人物表尤為有用。

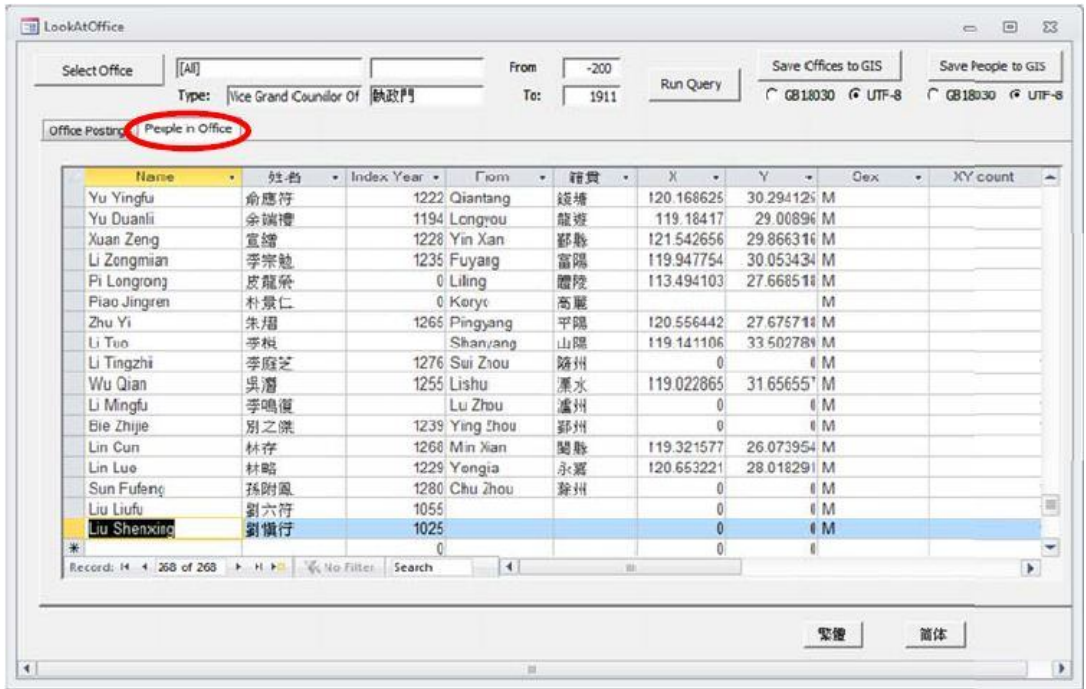


圖 22

有關“除授-官職”的表格包含了28條信息：

Person ID	人物ID
Person Name (pinyin)	人物姓名（拼音）
Person Name (Chinese)	人物姓名（中文）
Index Year	指數年
Sex	性別
Address ID	地址ID
Address (pinyin)	地址（拼音）
Address (Chinese)	地址（中文）
X coordinate of Address	地址X坐標
Y coordinate of Address Posting ID	地址Y坐標

Office ID	除授ID
Office (translation)	職官（英文翻譯）
Office (Chinese)	職官（中文）
First year of appointment	任職始年
Last year of appointment	任職終年
Dynasty (useful in cases where the years are very uncertain)	朝代（在年代不確定時有用）
Appointment type (regular, provisional, etc .)	除授類型（正授、權攝等）
Appointment type (Chinese)	除授類型（中文）
Information on assumption of office (accepted, declined, etc.)	赴任情況（接受、推辭等）
Information on assumption of office (Chinese)	赴任情況（中文）
Office Address ID	職官地址ID
Office Address (pinyin)	職官地址（拼音）
Office Address (Chinese)	職官地址（中文）
X coordinate of Office Address	職官地址X坐標
Y coordinate of Office Address	職官地址Y坐標
XY count (number of postings) for the Office Address	職官地址計數（除授數）
Notes	備註

任官人物表包括的信息有：

Person ID	人物ID
Person Name (pinyin)	人物姓名（拼音）
Person Name (Chinese)	人物姓名（中文）
Index Year	指數年

Sex	性別
Address ID	地址ID
Address (pinyin)	地址（拼音）
Address (Chinese)	地址（中文）
X coordinate of Address	地址X坐標
Y coordinate of Address	地址Y坐標
XY count (number of people) for the Address	地址計數（人數）

因為我們可能會查詢除授或在任官員的空間分佈，LookAtOffice窗體還可以將這些信息保存為GIS軟件可識別的文件。我們可以選擇UTF-8或GB18030的編碼。

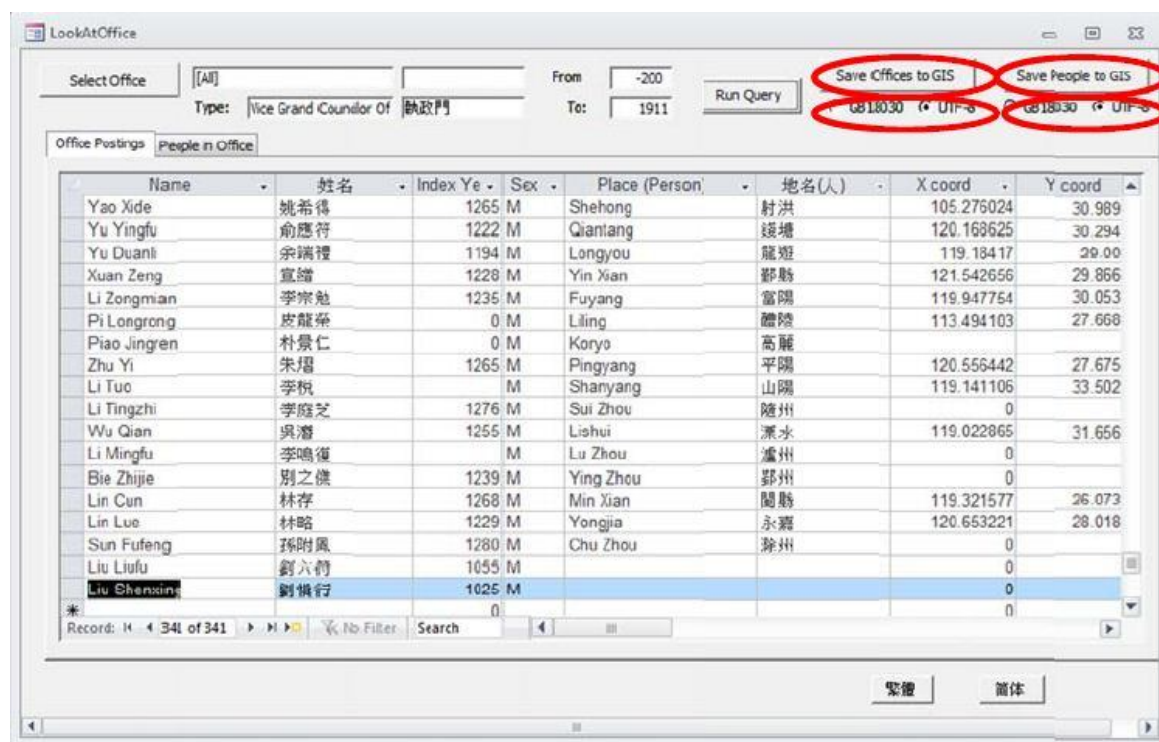


圖 23

D. 使用 LookAtKinship 窗體

涉及親屬關係的查詢要比查社會關係種類或者任官資格之類更為繁雜。個人的親屬信息通常只含有較少的記錄，因此CBDB從這些記錄開始檢索與此人有親戚關係的人，再檢索親戚的親戚.....這樣一直持續檢索，直到特別的條件被滿足。首先是對檢索循環次數的限制，通常5000次循環即可。其次是對於親族關係的親疏度（不妨稱之為親屬距離）進行檢查：

Max. Ancestor Gen. 限定循環中包括幾代祖先的信息。一個人的父親輩是1，祖父輩是2，曾祖父輩是3，依此類推即可。

Max. Descend. Gen. 限定循環中包括幾代後人的信息。一個人的兒子輩是1，孫子輩是2，曾孫輩是3，依此類推即可。

Max. Collateral Kin 限定能檢索多少層同輩間的關係。例如，一個人妻子的姐妹，與此人有1個單位的婚姻距離，1個單位的旁系距離。

Max. Marriage Dist. 限定檢索時能有多少層通過婚姻結成的關係。一個人的妻子的姐妹的丈夫有2個單位的婚姻距離。

要直觀地認識這些區別，可參考下面的親屬網絡

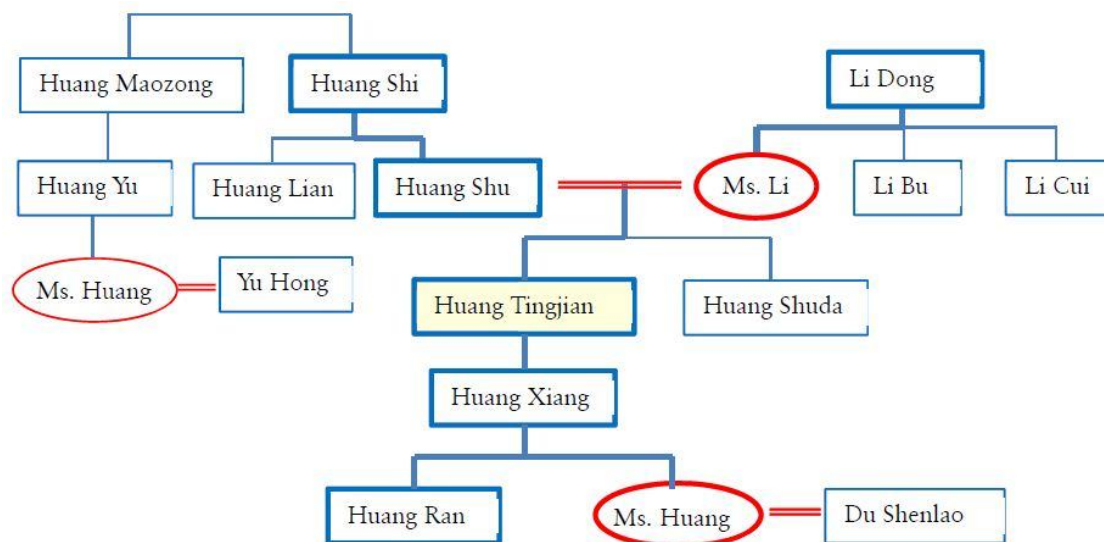


圖 24

對黃庭堅而言，有粗線標記的矩形和橢圓是直系血親（父母，子女）。由兩道線連接的是婚姻關係。其他單線標記的為旁系親屬。在LookAtKinship的系統中，我們得出：

Huang Yu 黃育	is FFBS	(Up = 2, Down = 1, Collateral = 1)
Yu Hong 余宏	is FFBSDH	(Up = 2, Down = 2, Collateral = 1, Marr. = 1)
Li Cui 李萃	is MB	(Up=1, Collateral=1, Marr.=1)
Du Shenlao 杜莘老	is SDH	(Down = 2, Marr. = 1)

因為LookAtNetwork窗體在檢索時會持續地進行循環檢查，直至達到設定的距離限制，因此由一條檢索請求而生成的親屬關係表可能非常龐大。

請注意：檢索親屬距離很大的旁系血親和婚姻關係可能產生龐大的數據集。

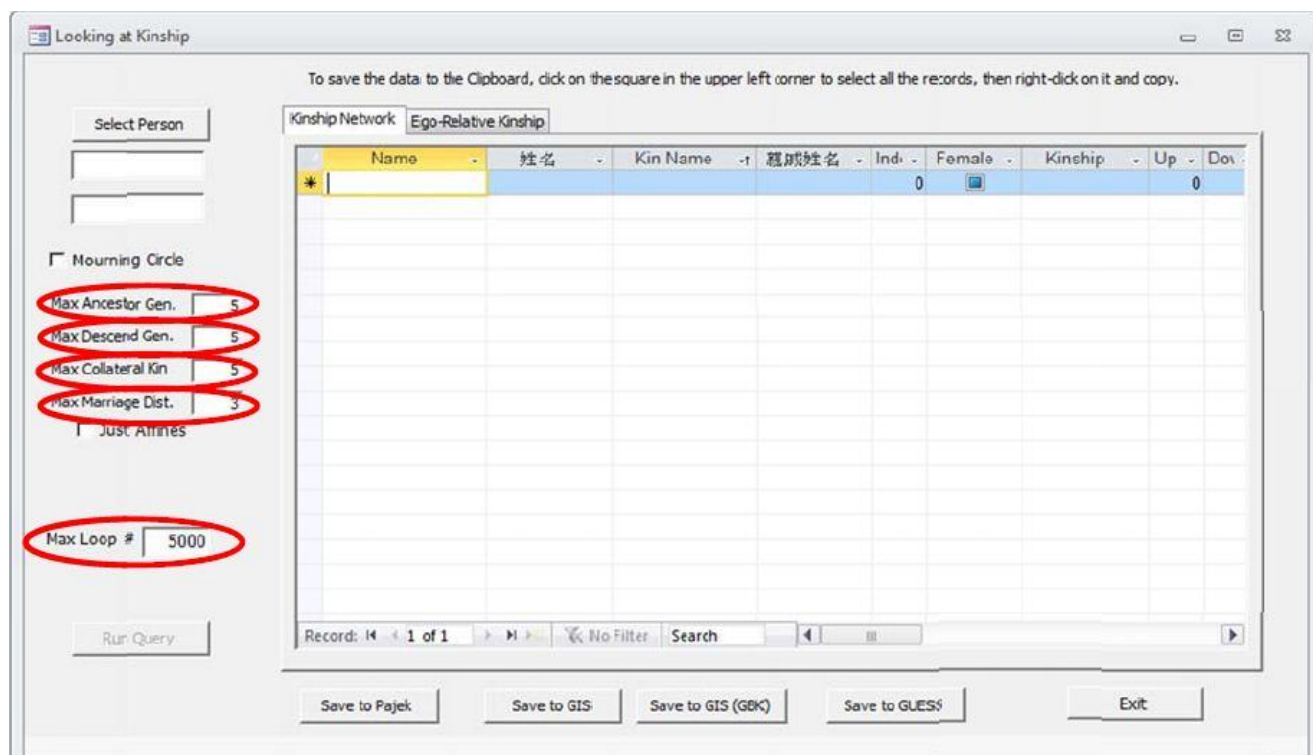


圖 25

另一中國家族研究中受到廣泛關注的議題即“五服”。LookAtKinship窗體允許我們通過點擊“五服”選項，讓數據庫重建特定人物的“五服親屬”。不過，當我們選擇“五服”選項時，其他四個限制條件就無效了。

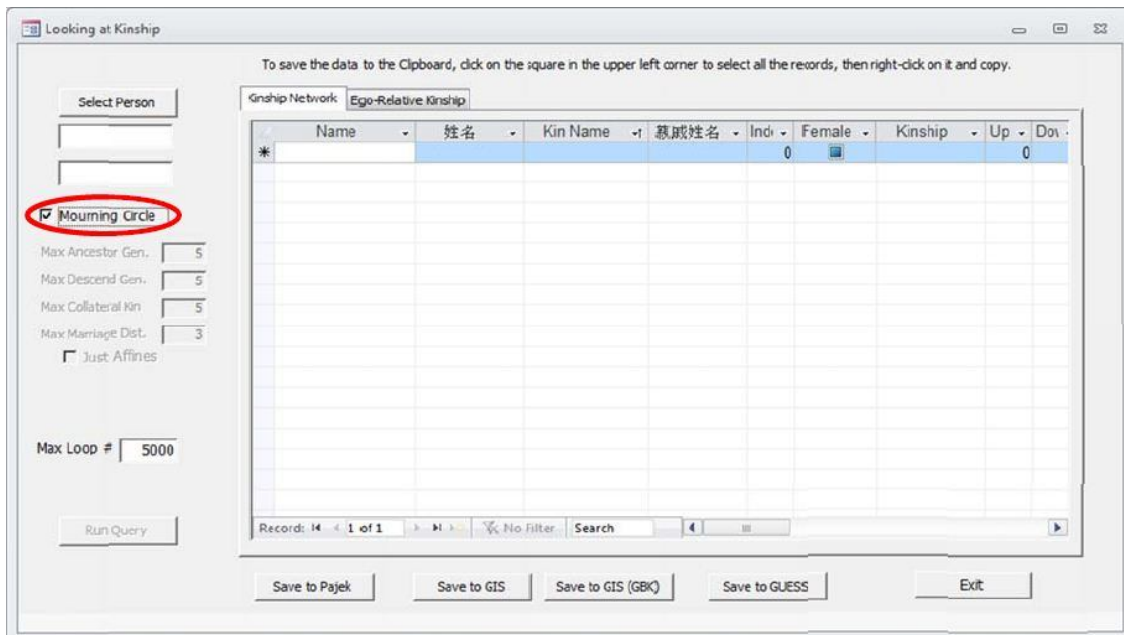


圖 26

為了檢查親屬關係，我們首先點擊“Select Person”選項，彈出窗體中會出現數據庫中所有人的列表。如果知道待搜索之人的中文名，就可直接查找：

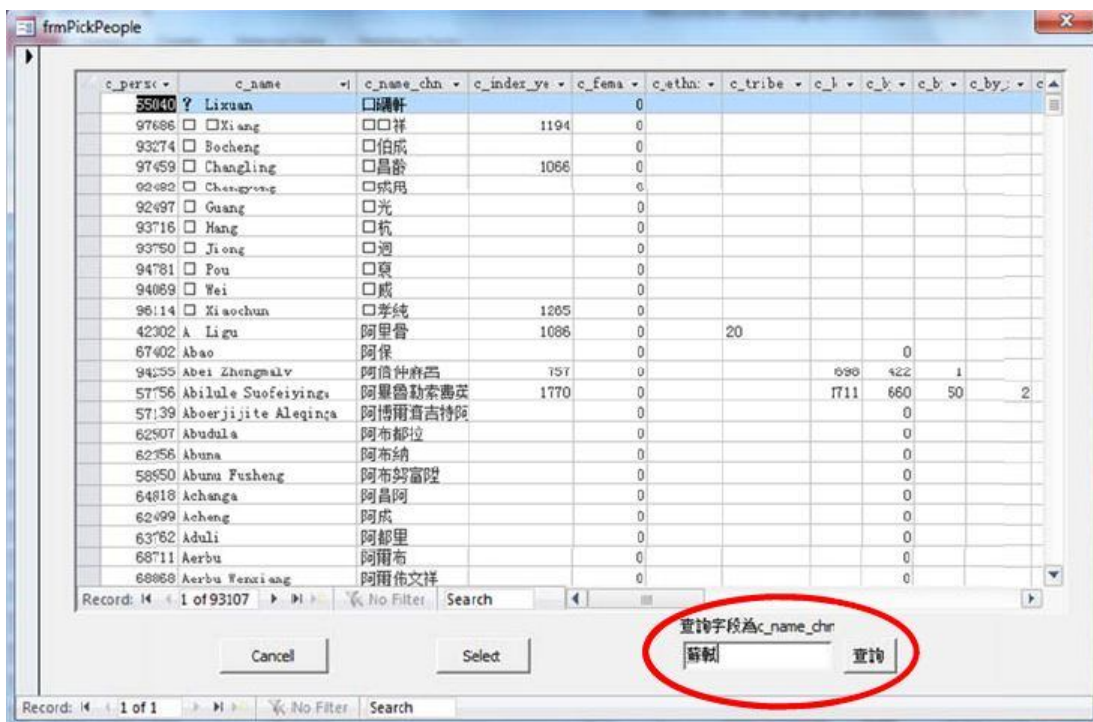


圖 27

當已經選定待檢索之人物后，設定搜索的條件限制，點擊“Run Query”開始檢索。鑒於檢

索可能耗費很長時間，CBDB會彈出窗口，提示當前檢索的最新動態和檢索到的人數：

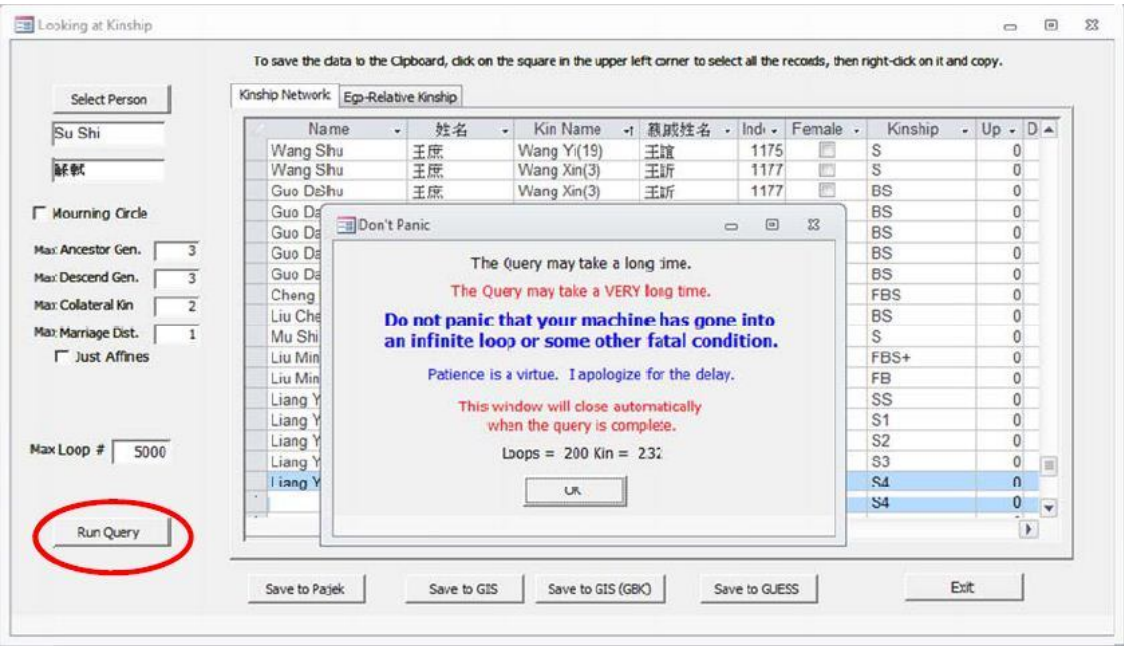


圖 28

當檢索結束，我們能看到兩張表。首先是“親屬網絡表”（Kinship Network），列出了所有找到的親屬關係：

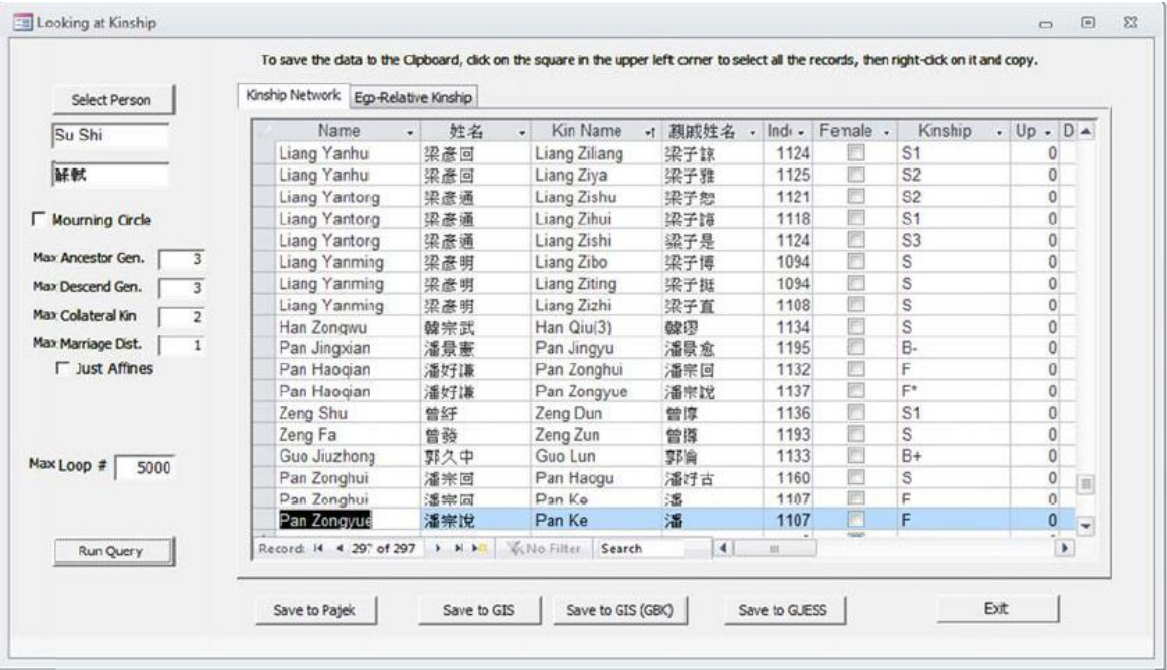


圖 29

親屬網絡表有18列:

Name (pinyin)	姓名（拼音）
Name (Chinese)	姓名（中文）
Kin Name (pinyin)	親屬姓名（拼音）
Kin Name (Chinese)	親屬姓名（中文）
Index Year of Kin	親屬指數年
Sex of Kin	親屬性別
Kinship Relation	親屬關係
Address of Person (pinyin)	人物地址（拼音）
Address of Person (Chinese)	人物地址（中文）
Address Type	地址類別
X- Coordinate	X坐標
Y- Coordinate	Y坐標
Address of Kin (pinyin)	親屬地址（拼音）
Address of Kin (Chinese)	親屬地址（中文）
X- Coordinate of Kin Address	親屬地址X坐標
Y- Coordinate of Kin Address	親屬地址Y坐標
Address Type of Kin Address	親屬地址類別
Notes	備註

第二張表名為“自我相關親屬表”（Ego-Relative Kinship），展示了第一張表中的人物與最初檢索之人的親屬關係。



圖 30

例如，梁子美是蘇軾的弟弟的兒子的妻子的父親。從“蘇軾”搜索到“梁子美”的路徑依次是，首先通過旁系血親(B2)，然後向下搜索一代(S)，通過婚姻關係(W)，最後向上一代(F)。這樣親緣距離就可以表示為{1,1,1,1}。不過，“自我相關親屬表”表有一個附加列，給出了一條形如(FS>B)2SWF的搜索路徑，說明CBDB是如何得出B2SWF的搜索路徑的。這裡有一個簡單的變換：CBDB把“FS”（即“父親的兒子”）用“B”（兄弟）代替了。

與其他窗體中生成的表一樣，如果點擊左上角，我們能夠全選表中的所有記錄，然後可以剪切或者粘貼到其他程序中。同樣，鼠標右鍵點擊任意一列的頂部可以根據此列進行排序。

最後，我們可以將親屬關係數據導出為三種格式的文件，前兩種針對不同的社會網分析工具：Pajek（1）和GUESS（2）。第三種為在GIS系統中可視化呈現，我們可以選擇（3）中文GBK或（4）非中文編碼保存文件，以用於ArcGIS軟件。

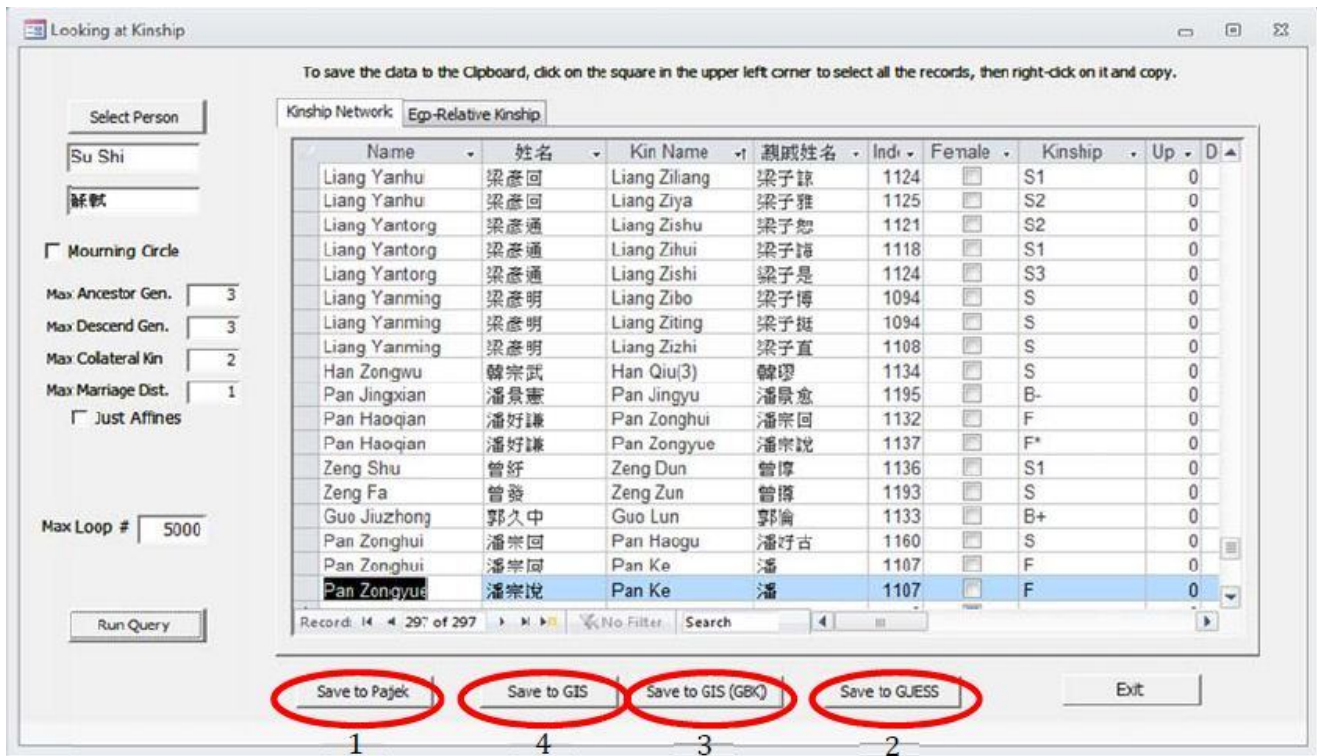


圖 31

E. 使用 LookAtNetworks 窗體

LookAtNetworks（查詢關係網絡）窗體是CBDB-ACCESS版功能最強大的查詢界面，能夠讓用戶檢索由親屬關係或者其他社會關係而生成的社會網絡。用戶可以選擇一個人或者導入一批由其他檢索得到的人物，也可以從地點或者地點列表開始檢索。LookAtNetworks允許用戶選擇特定形式的社會關係來檢索，還可限定檢索的年代範圍。與LookAtKinship窗體中一樣，LookAtNetworks中的檢索也是迭代進行的。從基本人群出發，找出與他們相關的人，然後不斷循環，找出更多相關的人。LookAtKinship有五個限定條件，但是LookAtNetworks只有兩個：一個最大的循環計數量（檢索進行了多少次循環），以及一個最大的距離數，即從網絡中的個人與檢索初始所認定的團體成員之間的連接層數。如果最初用戶選擇了一個特定的人，那麼距離從此人算起；如果最初選取一個團體，那麼整個團體中的人都可以作為起點。如果是從地點或者地點列表開始，那麼最初與這個地點有關係的人會被視為起始點。

請注意：較高的距離數級別可能花費相當長的時間，並產生龐大的數據集。

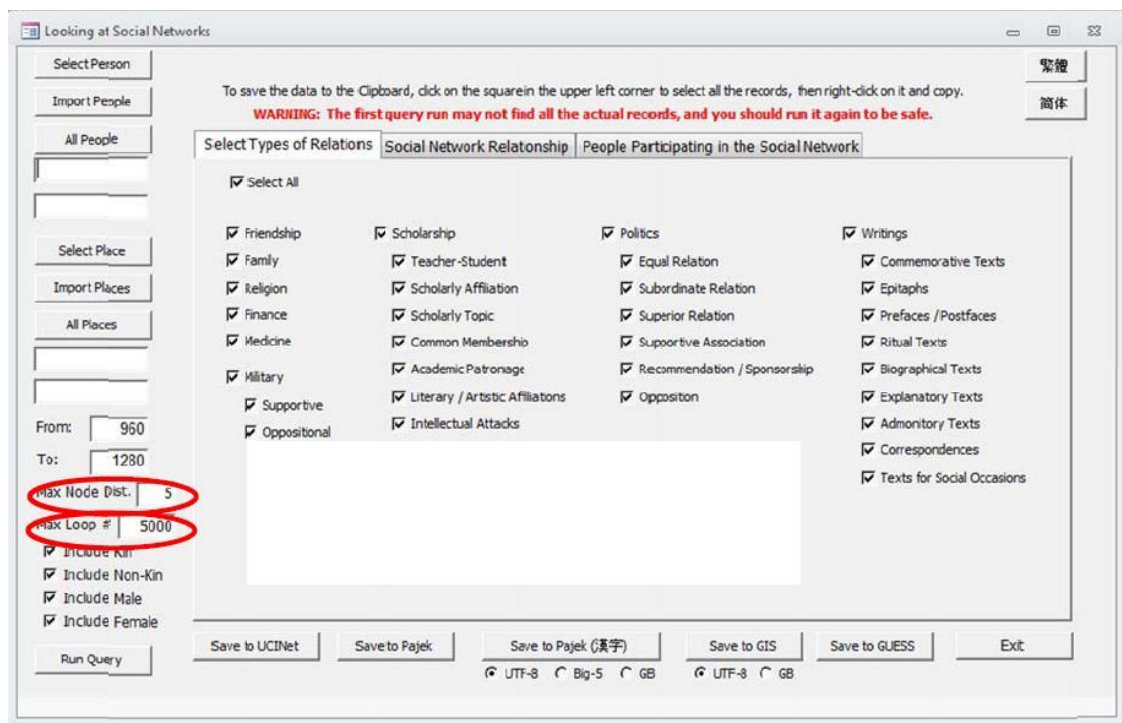


圖 32

說明：距離數設置為2的檢索會生成：

1. 所有與選中人物存在社會關係的人；
2. 所有在選中人物的社會網絡中出現的社會關係。

我們可以對檢索得到的表中關係進行排序，用戶也可以刪除後續分析不需要的記錄。

執行檢索從選擇檢索的元素開始：

1. 選擇人物

a. 選擇單個人物

如果我們想查看特定個人的社會關係網，我們可以點擊“Select Person”打開人物列表。

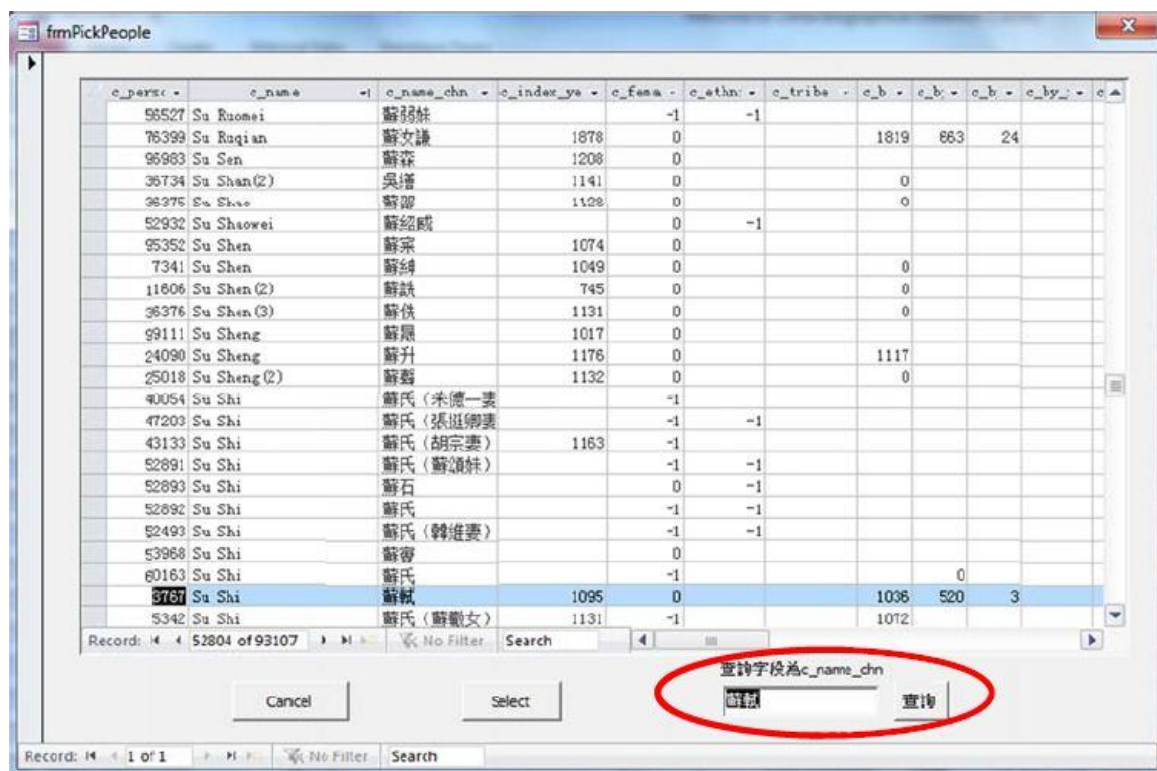


圖 33

由於列表包含了數據庫中的所有人物，用戶可以通過中文名直接搜索。

b. 導入人物群體

第二種對研究社會網絡很有用的方法是導入一組通過其他查詢獲得並擁有相同特徵的人群。例如，用戶可以選擇宋代通過科舉考試取得任官資格的士人群體。我們把通過LookAtEntry檢索得到的結果存入一個文本文件中（.txt文件，可以通過Word或Excel創建）並且進行編輯：



圖 34

請注意，我們可以在錄入文本信息之前，添加一行標注（如圖34中的Person ID、Name等）或者其他文本信息，程序會自動省略這一行。用戶可以用制表符或者逗號來分隔字段。CBDB從中抽取的信息是Person ID，而ID必須是每一行的第一項。點擊“Import People”命令鍵，選擇文件，當LookAtNetworks成功讀取文件之後，窗體會變成以下狀態：

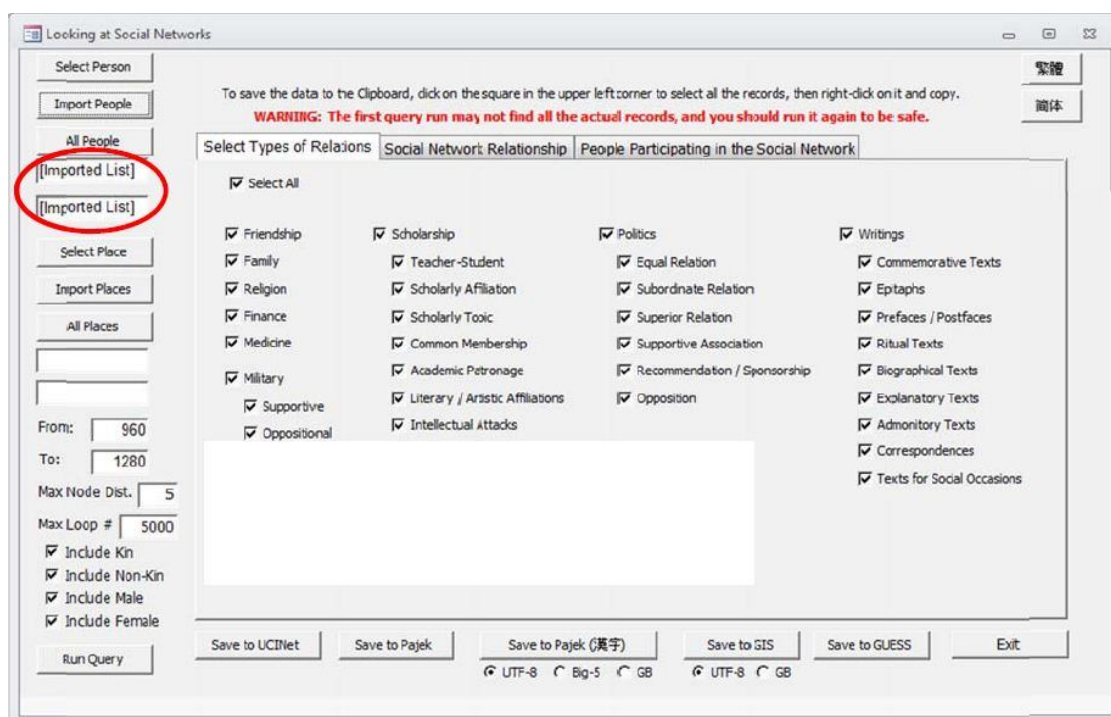


圖 35

兩個分別給出人名拼音和漢字的文本框會顯示 “[Imported List]”。

2. 選擇地點

a. 選擇單個地點

點擊“Select Place”，會打開一個窗體，讓我們選擇一個特定的地點。現在資料庫中已經有超過21000個地名。如果知道中文名，可以進行地名搜索。由於同樣的行政區劃可能隨著時間而改變其邊界，地名表向用戶提供了附加信息，讓用戶決定在查詢的時候使用哪一個地名ID。如果我們選擇了一個比縣大的行政區劃，那麼CBDB會自動識別從屬於這個行政區的低級政區（以及ID）。如果我們選擇了特定的年代範圍，但是這一政區並不在整個年代範圍內都存在，那麼CBDB可能會把佔據相同地域的政區列出（用ADDR PLACE表）。

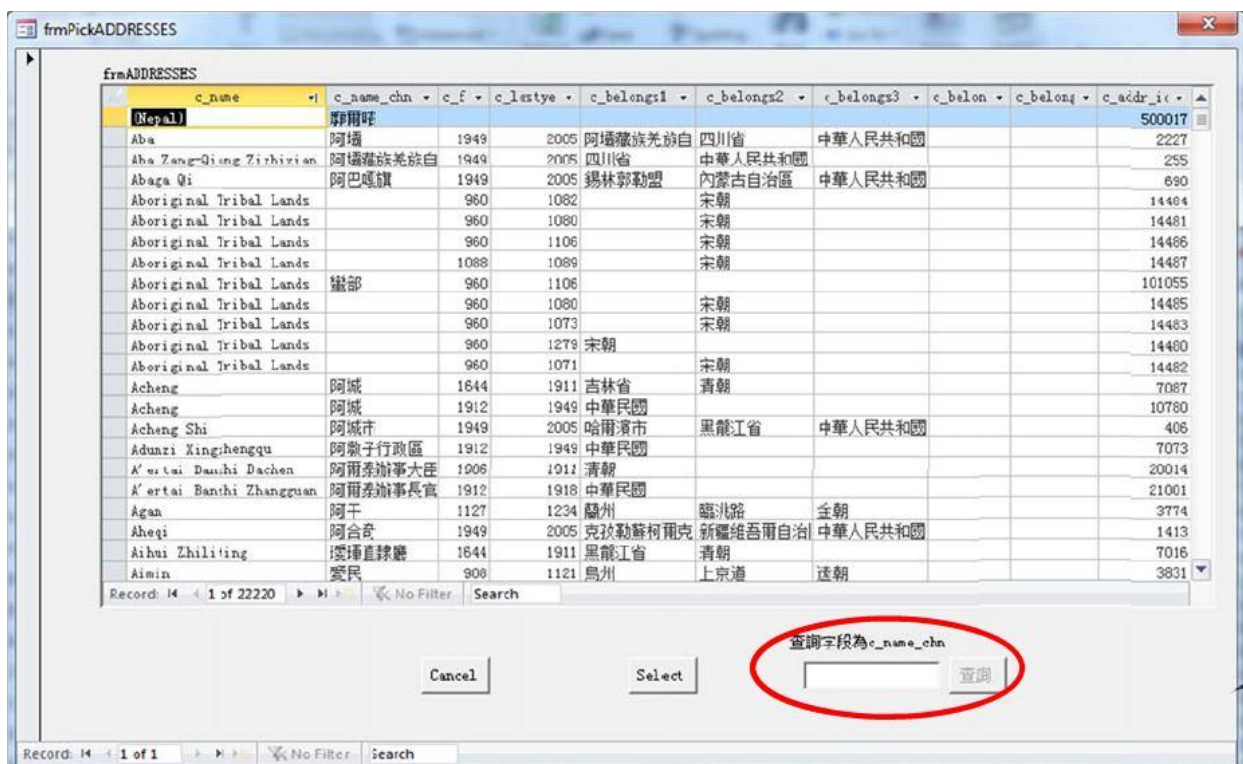


圖 36

b. 導入地點列表

有時，研究者更希望導入一組地點，來精確研究一個區域的社會網絡。導入一組地點的ID與導入人群(Person IDS)方法相同。CBDB會自動忽略以文本開始的行以及Address ID後面的文本——這要求Address ID是每一行的第一項。選擇單個地點或者地點列表只會影響開始時如何找出一群人來圍繞其構建社會網絡。找到與選擇的地點有聯繫的人之後，CBDB會對所選的社會關係進行下一步檢索。如果我們同時使用人物和地點設定檢索的起點，CBDB會查詢來自某個特定地點的人，這些人與選中的個人或人群有選定的的社會關係。

c. 決定時間段

顯然，用戶只需填寫待檢索人物的起始和終結的指數年即可。

d. 選擇節點距離

我們需要注意：隨著距離數的增加，檢索得到的人物數量可能呈指數級的上升，所以建議用戶從較小的數字開始。在示例的檢索中，用九個通過科舉取得任官資格的士人，節點距離為3，得到了超過5000條關係。

e. 設置循環數限制

由於檢索過程速度較慢，因此用戶可以設定較小的循環數限制。每一個新添加到列表中的人都需要一次循環，因此檢索無法獲得超過最大循環數下的生成人數。

f. 選擇親屬關係，社會關係，男性或者女性

LookAtKinship窗體不允許用戶搜索一個群體的親屬關係，因此LookAtNetworks窗體向我們提供了替代的方法。為了檢索親族關係，我們可以選擇“包括親屬關係”，并勾除“包括社會關係”。我們也可以排除基於女性的關係，或者檢查限定在女性之間的網絡。LookAtNetworks 窗體中能提供這樣的選項。

g. 選擇非親屬關係的類型

因為有很多非親屬關係的類別，而絕大多數在特定的檢索中並無用處，因此我們可以將搜索限定在一個較大的關係類別中，例如：

Friendship	朋友
Family	家庭
Religion	宗教
Finance	財政
Medicine	醫藥
Military	軍事
Scholarship	學術
Politics	政治
Writing	著述

最後的四種非親屬關係還有可供選擇的子類。軍事關係有兩種，學術關係有七種，政治關係

有六種，著述關係有九種。我們可以將這些子類進行混合，關係的類別一旦選定，這些限制在整個搜索過程中都有效。

所有選擇完成之後，即可開始查詢，這可能會花費很長時間。下面的例子使用了滿足下列條件的人物列表

- A. 通過明法科入仕（第一版選擇960-1280年）
- B. 距離數限制為3
- C. 循環數限制為5000
- D. 對“是否為親屬關係”沒有限制，也沒有限制各種社會聯繫。

如同 LookAtKinship 窗體，檢索時會出現一個提示檢索動態的窗體。在社會網絡分析的術語中，關係中的個人被稱之為“節點”（Node），關係被成為“邊緣”（Edge）。窗體顯示經過兩百次循環，程序已經找到了在 3100 筆關係中的 1483 個人物（E）。

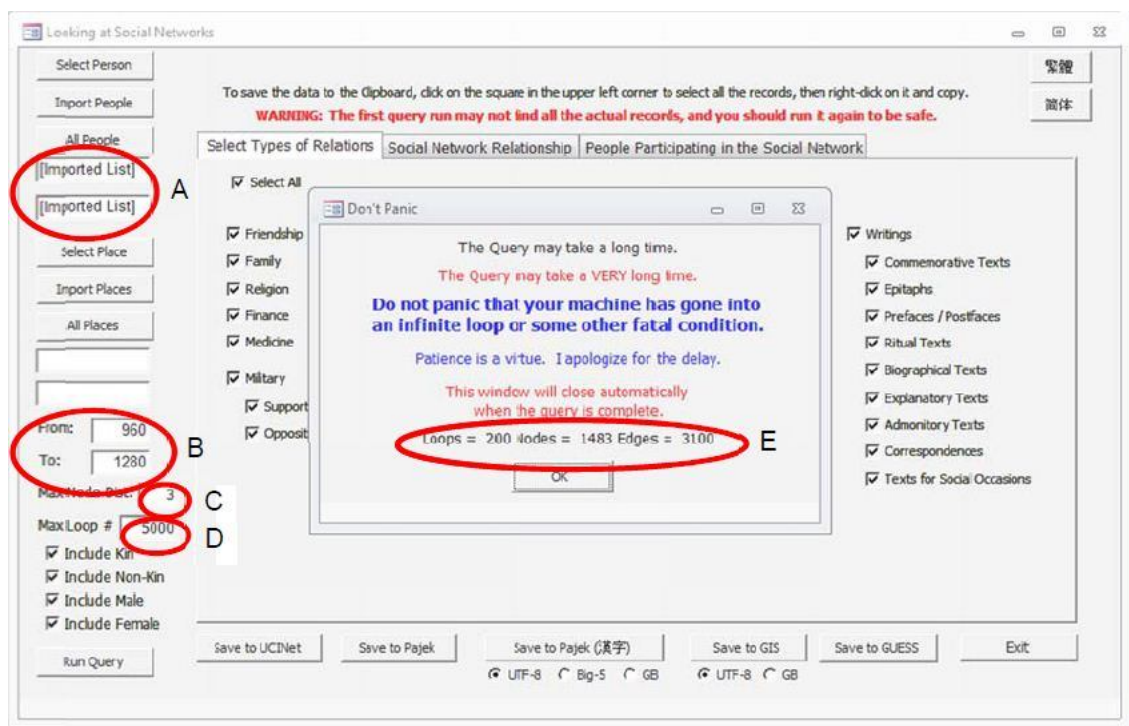


圖 37

最後結果顯示了5482筆關係中的1483個人物（圖38）。

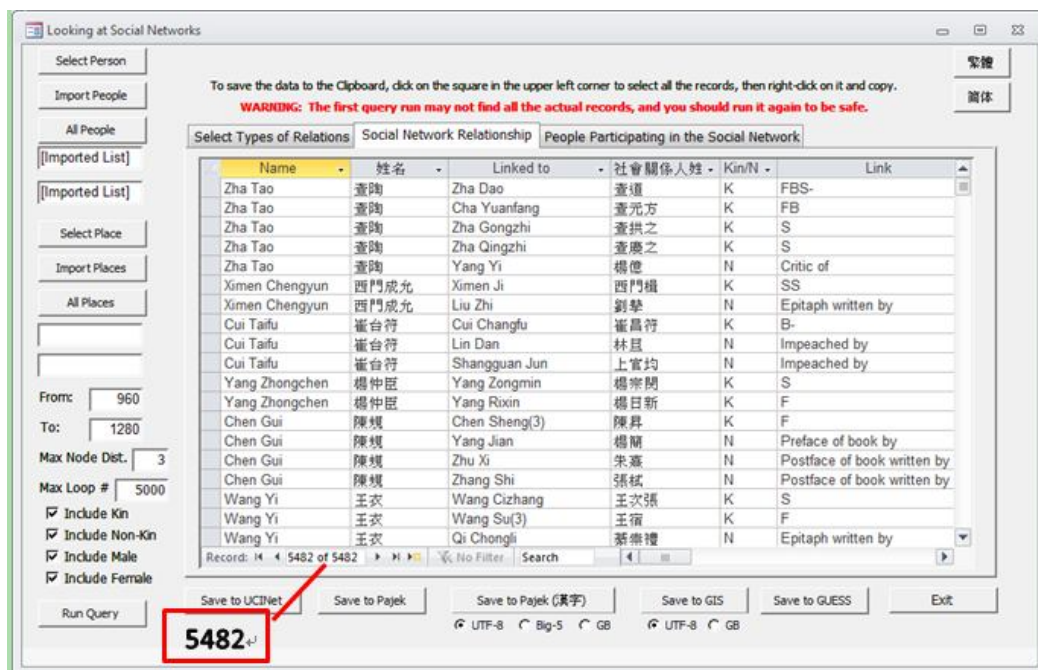


圖 38

如果我們只檢索目標群體的親屬關係，CBDB找到了通過98條關係相關聯的89個人（距離數為5，時間範圍相同）。如果我們想要查詢通過著述關係相連的人物，當節點距離為3時會找到2903條關係中的823個人物。

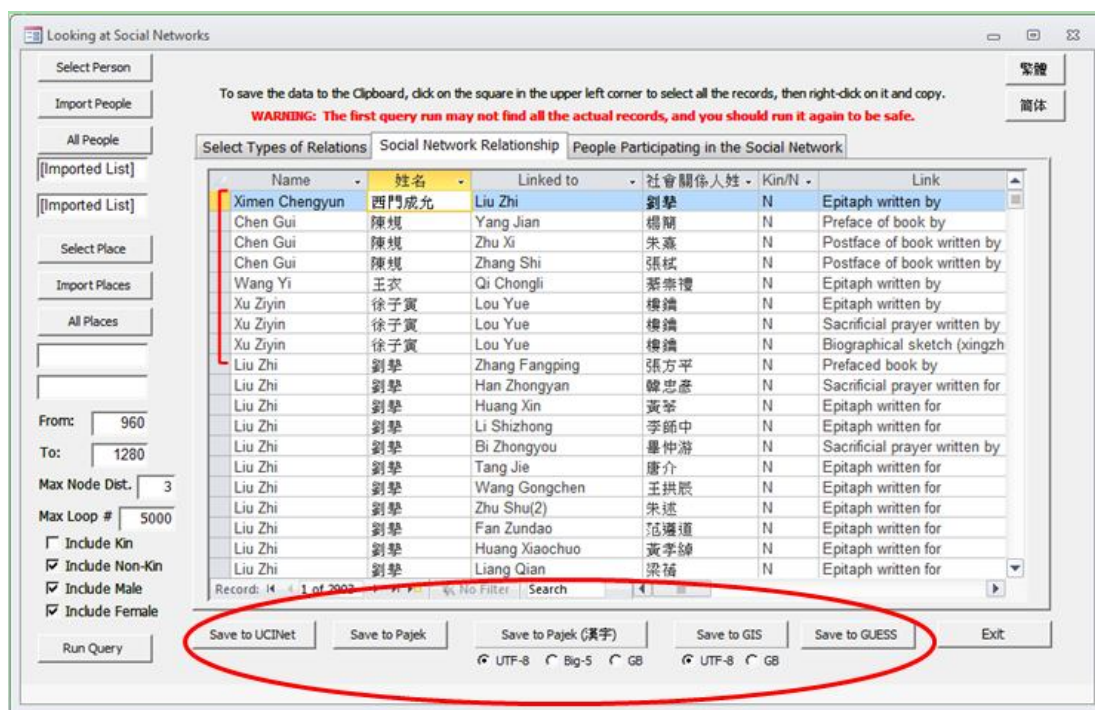


圖 39

以上的查詢結果看上去似乎還不錯：關係條目不多也不少。不過，從圖39可以看到，只有最開始的8條記錄是與最初的那些由明法科入仕的群體直接相關，而最初導入的9個人中只有4個人有社會關係，並且只與6個人相關聯。在這6條關係中，5條與那些有龐大社會網絡且貢獻了社會網絡中絕大部分關係的人相連（劉摯、楊簡、朱熹、張栻、樓鑰）。因此，最好返回到較大的5482條關係的集合，然後應用社會網絡分析的工具來對數據進行分類。

LookAtNetworks 窗體可以將查詢結果輸出到3種社會網絡分析（SNA）程序中：**UCInet**, **Pajek** 和 **GUESS**。**Pajek**已經能支持中文，因此**CBDB**能以3種不同的漢字編碼保存查詢結果：**UTF-8**、**Big-5** 和 **GB**。**GIS**軟件也支持中文，但是對漢字編碼的處理基於個人電腦的區域設置。