



美國哈佛大學包弼德教授專題演講紀要 ——中國歷代人物傳記資料庫（CBDB）和哈 佛中國課（ChinaX）

曾彩娥 國家圖書館知識服務組助理編輯

【館務報導】

一、緣起

國家圖書館於2014年3月11日舉辦知識饗宴，在前館長黃寬重教授引介下邀請美國哈佛大學副教務長、東亞語言文明學包弼德（Peter K. Bol）講座教授、陳詩沛博士（德國馬克斯普郎克科學史研究所研究員）及于文小姐（哈佛大學歷史學系博士候選人）蒞臨本館發表演說，講題為「人文學、數位資源與數位工具的跨界整合：來自中國史研究和教學的兩個範例——中國歷代人物傳記資料庫（CBDB）和哈佛中國課（ChinaX）」，並邀請國立臺灣大學數位人文研究中心主任項潔教授擔任主持人。與會來賓包括長庚大學人文及社會醫學科黃寬重講座教授、中央研究院王汎森副院長、中央研究院歷史語言研究所劉錚雲研究員兼傅斯年圖書館主任等學界、圖書館界、資訊界人士。

二、中國歷代人物傳記資料庫（CBDB）

中國歷代人物傳記資料庫（The China Biographical Database 簡稱 CBDB）之始祖為賓州大學已故教授郝若貝教授（Robert M. Hartwell）

（1932 – 1996）。設計之初是研究宋代士大夫階級之構成，故原始資料庫主要包含宋代士人之傳記及親屬等資料。郝若貝教授將本資料庫初版及其他財產遺贈哈佛燕京學社。目前本資料庫的開發工作係由哈佛大學燕京學社、中央研究院歷史語言研究所及北京大學中國古代史研究中心三方合作進行。目前資料庫中約有33萬筆人物資料，自唐朝到民初，明代尤其多。其中唐朝有45386人、五代有374人、宋朝有45072人、遼有318人、金有275人、元朝有19887人、明朝有145692人、清有37915人、民國有3222人、其他23879人等。CBDB人物傳記資料包括：基本資料、地址資料、別名資料、著術資料、任官資料、入仕途徑、親屬資料、社會關係資料、社會區分資料、財產資料、事件資料等。

（一）設計概念

中國歷代人物傳記資料庫之設計具有如下飛躍概念：

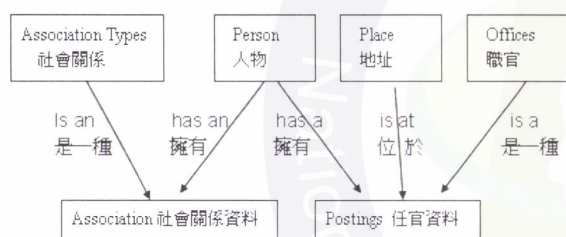
1、文本挖掘/文本建模（Text mining / text modeling）

CBDB資料庫由傳記文文本中蒐集資料，並利

用正則表達式（Regular Expression）發掘分析文本。正則表達式使用單個字串來描述、匹配一系列符合某個句法規則的字串。在很多文字編輯器裡，正則運算式通常被用來檢索、替換那些符合某個模式的文字。CBDB資料庫主要的任務是蒐集訊息，而非考證。

2、基於關聯式資料的建模（Modeling lives with relational databases）

CBDB資料庫利用人工或電腦技術將傳記文本中各種片斷資訊抽出、分門別類，運用關聯式資料庫（relational databases）將人物傳記各種資訊實體（人物、職官、地址等）結構化、模型化，如圖一。



（資料來源：陳靜沛博士2014年3月11日演講簡報資料）

圖1：CBDB 結構化圖

3、群體傳記學（Prosopography）

CBDB是研究群體傳記學的重要資料庫。群體傳記學乃係透過對一群歷史人物的生平所做的集體性的研究，探討其共同之背景特徵。其方法乃係先建立一個研究範圍，然後就此提出一組相同的問題：包括出生、死亡、婚姻、家庭、社會出身、經濟地位、居住地、教育、個人財富與來源、職業、宗教、宦歷等方面。然後，將此研究範圍中所有人物的各類信息加以羅列、組合，通過對這些信息的考察找出具有顯著意義的變量。研究者可以檢測這些信息的內在相關性，及其與其他行為形式或行動形式的相關性。（Stone,

1972）

4、空間分析（Spatial analysis）

CBDB根據人物籍貫結合地理資訊系統（Geographic Information System, 簡稱GIS）可以分析人物傳記資料空間分佈，如CBDB中已知籍貫的67,000歷史人物的地理分佈主要在華北邊緣、四川西部、福建、江蘇、浙江、江西等地。

5、社會網絡分析（Social network analysis）

社會網絡分析（Social Network Analysis, SNA）係指社會關係所構成的結構，透過研究網絡關係，把個體間關係、社會角色層次的結構（微觀結構）與大規模的社會系統的“巨集觀”結構結合起來，讓研究者探究個人關係間之連結意義。

（二）使用方式

中國歷代人物傳記資料庫之使用方式有三種：

1、線上查詢

中國歷代人物傳記資料庫提供之查詢服務包括：

- （1）人名查詢：可以根據姓名、指數年（年代）、朝代、地址等人物的基礎生平資訊，從CBDB中篩選出相應的人物。指數年是指該人物六十歲時的年份，若該人物在六十歲前去世，則指數年為其去世的年份。查詢條件欄位包括：姓名/別名/外文名、年代、朝代（下拉式選單）、地名（選單）。
- （2）地名查詢：可以根據人物的籍貫、本貫、祖籍、遷住地、出生地、埋葬地等地址來查詢人物。查詢條件欄位包括：地名、地址類別（下拉式選單）、年代、朝代（下拉式選單）。
- （3）官名查詢：可以根據人物的任官及官職相關



資訊、篩選出相關的人物或人物列表。查詢的參數包括有入仕的途徑（科舉考試、薦舉等）、不同等級和類別的官名、（地方官的）為官地址，以及為官年代。查詢條件欄位包括：朝代、官名、入仕類別、為官年代、為官地址。

- (4) 關鍵詞查詢：可以根據CBDB的：基本資料、地址、著述、官名、入仕、社會區分、親屬、社會關係等八個代碼表中所使用的中英文關鍵詞，來篩選出所要尋找的人物。查詢條件欄位包括：關鍵詞、類別、指數年、為官年代
- (5) 進階查詢：可以透過對一系列額外參數的設定如傳主基本資料、地址、官名、入仕等來查詢所要尋找的人物。查詢條件欄位包括：姓名/別名/外文名、生年、卒年、指數年、郡望、朝代、地址類別、地址、官名、為官年代、除授類別、職官類別、官名地址、入仕類別、入仕年、社會區分、著作類別。
- (6) 親屬／人際關係查詢：可根據人物之親屬關係、親屬關係人、社會關係、社會關係人等查詢人物相關親屬或相關人物資料。

2、下載單機版離線使用

CBDB 提供下載 .mdb 檔案在自己 local 主機上查詢運用。

3、利用 CBDB API

API（application programming interfaces）是一種允許資料庫之間互相溝通的介面。透過API，每個資料庫都可以取用其他資料庫的資訊，來補充自己未能著重的資料面相，而不必在自己資料庫重新輸入這些資料。CBDB為其他資料庫提供API服務，讓任何資料庫都可以取用CBDB 人物傳記，並於自己的資料庫呈現出來。CBDB API 目

前有兩種查詢方式：（1）用人物CBDB ID查詢；（2）用漢字或拼音人名查詢。

（三）資料來源

中國歷代人物傳記資料庫的資料來源包括：

- 1、近代傳記資料索引：宋人傳記資料索引、元人傳記資料索引、明人傳記資料索引、清代人物生卒年表；
- 2、史料：正史列傳、墓誌銘、墓表、地方志列傳等；
- 3、文學作品：文集之祭文、序、記、書信等；
- 4、官方文書：郡守年表、會要、實錄等；
- 5、其他資料庫：由哈佛燕京圖書館建置之明清婦女著作（Ming Qing Women's Writings）、由國立故宮博物院圖書文獻處與中央研究院歷史語言研究所合作建置之明清檔案人名權威資料庫（Ming Qing Name Authority files）。

三、哈佛中國課（ChinaX）

哈佛中國課（ChinaX）為由麻省理工學院和哈佛大學創建的edX大型開放線上課程平台上HarvardX的一門課程，免費開放大眾線上學習。哈佛大學之HarvardX目前有24個課程，約有來自193個國家100萬個學生，其中ChinaX約有43,000人修此課程。ChinaX由哈佛大學包弼德教授（Peter K. Bol）和柯比教授（William C. Kirby）共同開課講授課程內容，哈佛中國課（ChinaX）主要介紹中國的歷史、地理、哲學、文學、政治、經濟、藝術和生態學的過去、現在與未來。此課程利用新科技、數位工具跨界整合人文學、數位資源提供大眾不受限時間、空間及地點線上學習環境。透過此課程可看到哈佛珍藏有關中國之中外古今資料，探討瞭解中國的生活方式、文化、歷史和

現在。

四、結語

中國歷代人物傳記資料庫（CBDB）和哈佛中國課（ChinaX）跨界整合了人文學、數位資源與數位工具，包教授指出CBDB和 ChinaX資料庫有五個共同特色：（1）資料內容皆是有關中國的；（2）兩者皆是國際性的合作計畫；（3）跨學科：結合文學、藝術、歷史學、社會學、經濟及資訊科學等學科領域學者合作；（4）教授與學生合作；（5）免費公開給學界查詢利用。

CBDB資料庫由文本蒐集資料，並利用正則表達式分析文本。資料庫主要的任務是蒐集訊息，而非考證。本資料庫除可作為人物傳記的一種參考資料外，亦可對中國歷史進行統計、社會網絡以及空間層面的分析之用。ChinaX 課程並非一個人專業知識就足夠，就可以完成的。數位世界學生的期待是什麼、教學者難以預知。該課程透過老師與老師、老師與學生、學生與學生間對話及學生反饋，讓學生學到更多。

透過包弼德教授、陳詩沛博士及于文小姐對中國歷代人物傳記資料庫（CBDB）、哈佛中國課（ChinaX）演講內容，讓參與人員對 CBDB、ChinaX有更深一層認識。此次演講內容豐富、精彩，獲得與會各界人士熱烈回響。



曾淑賢館長（中）、哈佛包弼德教授（右）、臺大項潔教授（左）



包弼德教授（左2）、于文小姐（左1）、項潔教授（右2）、陳詩沛小姐（右1）

參考文獻

1. 包弼德（Peter K. Bol）、陳詩沛、于文，〈人文學、數位資源與數位工具的跨界整合：來自中國史研究和教學的兩個範例——中國歷代人物傳記資料庫（CBDB）和哈佛中國課（ChinaX）〉2014年3月11日專題演講 簡報資料。
2. Stone, L. , 〈Prosopography〉, in Felix Gilbert and Stephen R. Graubard eds., 《Historical Studies Today》, New York : W.W. Norton, 1972.
2. 中國歷代人物傳記資料庫（CBDB）網站，網址：<http://isites.harvard.edu/icb/icb.do?keyword=k35201>（擷取日期 2014年4月13日）
3. 維基百科〈正規表示式〉，網址：<http://zh.wikipedia.org/wiki/%E6%AD%A3%E5%88%99%E8%A1%A8%E8%BE%BE%E5%BC%8F>（擷取日期 2014年4月13日）
4. MBA智庫百科〈社會網路分析〉，網址：<http://wiki.mbalib.com/zh-tw/%E7%A4%BE%E4%BC%9A%E7%BD%91%E7%BB%9C%E5%88%86%E6%9E%90>（擷取日期 2014年4月13日）
5. edX 網站 HarvardX，網址：<https://www.edx.org/course/harvardx/harvardx-sw12x-china-920>（擷取日期 2014年4月13日）
6. 維基百科〈edX〉，網址：<http://zh.wikipedia.org/wiki/EdX>（擷取日期 2014年4月13日）