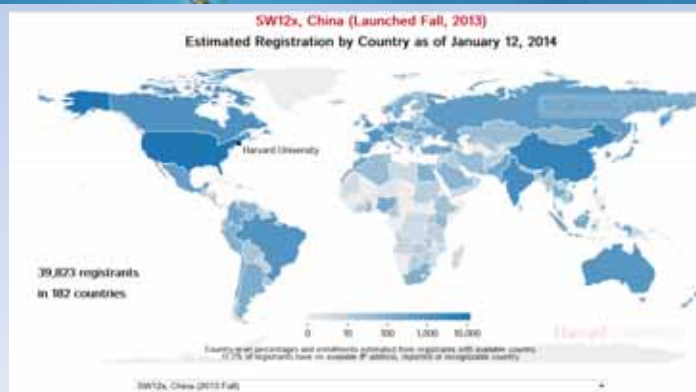


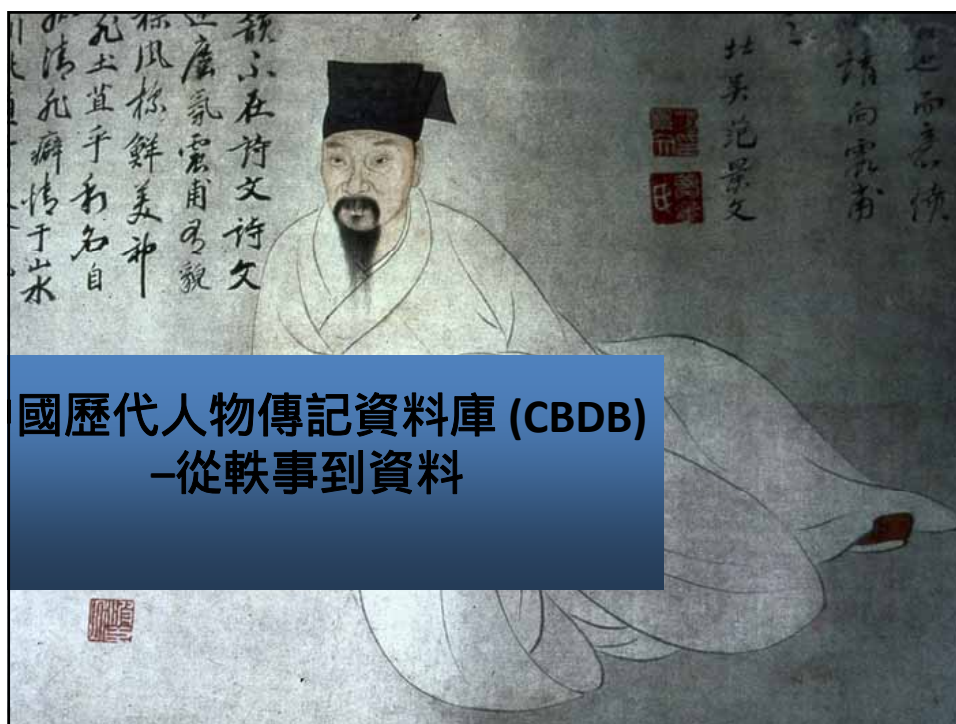
人文學、數位資源、與數位工具的
 跨界整合：
 來自中國史研究和教學的兩個範例
 - 中國歷代人物傳記資料庫(CBDB)
 和哈佛中國課(ChinaX)

哈佛大學包弼德 Prof. Peter K. Bol, Vice Provost for
 Advances in Learning

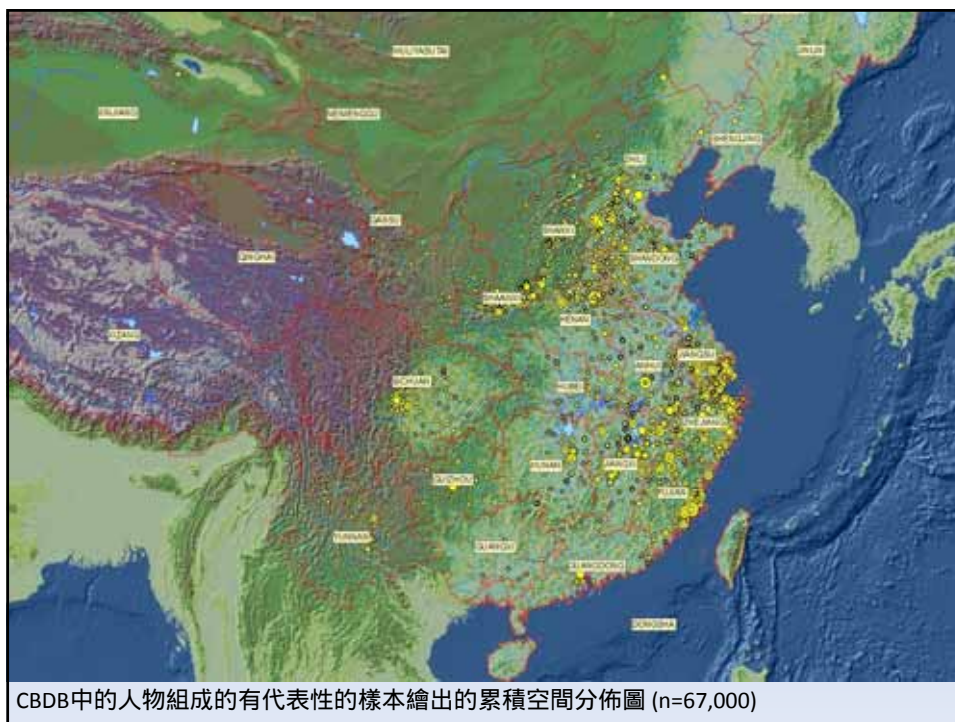
Max Planck Institute 陳詩沛 Dr. Chen Shih-pei,
 Researcher, Max Planck Institute, Berlin

哈佛大學于文 Yu Wen, Ph.D. candidate,
 Department of History, Harvard University





時代	人數
唐	45386
五代	374
宋	45072
遼	318
金	275
元	19887
明	145692
清	37915
民國	3222
其它	23879

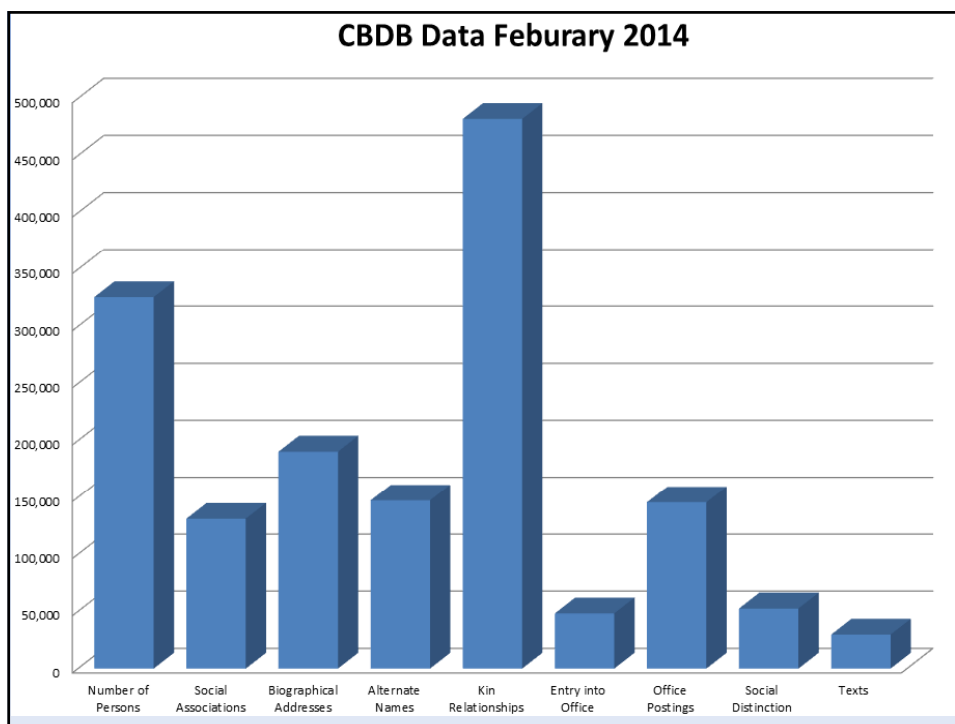


中國歷代人物傳記資料庫

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

(1) Biographical data are *coded* and stored in *tables*.
傳記資料以代碼的形式存儲於資料表中。

- | | |
|---------------------------------|----------|
| • BIOG_MAIN | • 基本資料 |
| • Biography Addresses | • 地址資料 |
| • Alternate Names | • 別名資料 |
| • Writings | • 著述資料 |
| • Postings | • 任官資料 |
| • Mode of Entry into Government | • 入仕途徑 |
| • Kinship | • 親屬資料 |
| • Associations | • 社會關係資料 |
| • Social Status | • 社會區分資料 |
| • Possessions | • 財產資料 |
| • Events | • 事件資料 |





Lü Zuqian, whose style name was Bogong, was a grandson of the Right Assistant Director to the Imperial Secretary Haowen. His family had lived in Wuzhou since his grandfather's generation. The learning of Zuqian was based on family [tradition], and embodied the textual transmission from the Central Plain. When he grew up, Zuqian studied with Lin Zhiqi, Wang Yingchen, and Hu Xian respectively. Then he also befriended Zhang Shi and Zhu Xi, and his explication and inquiry became more sophisticated.

First he obtained official rank by way of the protection privilege. But later he obtained his Jinshi degree and also passed the special decree examination for "Erudite Learning and Exceptional Literary Composition." Then he was appointed to the School for the Imperial Clan in the Southern Outer Office. During the mourning period for his mother, when he stayed in Mt. Mingzhao (in Wuyi), literati from all directions raced there. He was appointed Erudite in the National University.

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。丁內艱，居明招山，四方之士爭趨之。除太學博士。

傳記文文本中的“Factoids”

[English] Using “Regular Expressions” 正則表達式 to find data in digital texts

- In computing, a **regular expression** provides a concise and flexible means for "matching" (specifying and recognizing) [strings](#) of text, such as particular characters, words, or patterns of characters. Abbreviations for "regular expression" include "regex" and "regexp". The concept of regular expressions was first popularized by utilities provided by [Unix](#) distributions, in particular the editor [ed](#) and the filter [grep](#). A regular expression is written in a [formal language](#) that can be interpreted by a regular expression processor, which is a program that either serves as a [parser generator](#) or examines text and identifies parts that match the provided [specification](#). Historically, the concept of regular expressions is associated with Kleene's formalism of regular sets, introduced in the 1950s.
- http://en.wikipedia.org/wiki/Regular_expression

[中文]使用“正則表達式”在數位文本中發掘資料

- 正規表示式，又稱正則表達式、正規表示法、常規表示法（[英語](#)：Regular Expression - 在代碼中常簡寫為regex、regexp或RE），[電腦科學](#)的一個概念。正規表示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。
- 許多[程式語言](#)都支援利用正則運算式進行字串操作。例如，在[Perl](#)中就內建了一個功能強大的正則運算式引擎。正則運算式這個概念最初是由[Unix](#)中的工具軟體（例如[sed](#)和[grep](#)）普及開的。
- <http://zh.wikipedia.org/zh-tw/正则表达式>

正則表達式

reign title number & year character death character

(元祐|元豐|...|紹興)[0-9]+年卒

WDY

丁罕 (?~999), 潁州人。應募補衛士, 以戰功累遷指揮使。淳化中為澤州團練使, 知霸州, 河決, 以私錢募築, 民咸德之, 擢領靈環路都部署, 破李繼遷有殊功。後拜密州觀察使, 徙貝州。
咸平二年卒, 子守德能世其家。

丁明 (1127~1211), 舊名騫, 字希閱, 後改名明, 字子公, 金壇人, 權子。閉門讀書二十年, 手編事類及史通考等書百餘卷。奉祠家居, 嘉定四年卒, 年八十五。鄉里私謚博雅先生。

13

正則表達式和專名辨識(Named Entity Recognition-NER)

至元十八年為東陽縣丞

年號	年號代碼	年	year	地名	地名代碼	官名	官名代碼
至元	623	18	1281	東陽	18340	縣丞	841



- 基本資訊: 姓名, 性別, 日期(與時間的關係)
- 地址(與空間的關係)
- 別名(與不同種類稱呼的關係)
- 著作(學術關係)
- 任職(與政府的關係)
- 入仕方式
- 親屬(與別人的親屬關係)
- 社會關係(與別人的非親屬關係)
- 社會區分(與社會區分模式的關係)
- 財產(贈與和接受財產的關係)
- 事件(參加的事件)

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			
A Subset of the Data on Sima Guang: 有關司馬光的部分數據：			
Name 姓名	Dates 日期	Offices 任官	Associations 社會關係
Sima Guang 司馬光	1019-1086	(1) 1059 度支勾院 Budget Auditor; (2) 1085 門下侍郎 Executive of the Chancellery; (3) 1086 左僕射兼門下侍郎 Left Executive, Dept of Ministries [....]	(1) Yuanyou coalition member (元祐黨); (2) An Dun 安惇 Desires opposed by; (3) Chao Buzhi 晁補之 Sacrificial prayer written by; (4) Chen Jian 陳薦 Sacrificial prayer written for; (5) Chen Min 陳敏 Honored by; (6) Cheng Yi 程頤 Recommended; (7) Ding Du 丁度 Sacrificial prayer written for; (8) Fan Chunli 范純禮 Patron of; [....]

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			
Relational Database: 關係型數據庫		Many Entities 多個實體 People 人物 Offices 職官 Association Types 社會關係	
Name 姓名	Dates 日期		
Sima Guang 司馬光	1019-1086		
Person 人物	Posting Date 任命日期	Office Title 官名	
Sima Guang 司馬光	1059	度支勾院 Budget Auditor	
Sima Guang 司馬光	1085	門下侍郎 Executive of the Chancellery	
Sima Guang 司馬光	1086	左僕射兼門下侍郎 Left Executive, Dept of Ministries	
Person 人物	Association Type 社會關係	Associate 社會關係人	
Sima Guang 司馬光	Yuanyou member (元祐黨)	(not applicable)	
Sima Guang 司馬光	Desires opposed by	An Dun 安惇	
Sima Guang 司馬光	Sacrificial prayer written by	Chao Buzhi 晁補之	
Sima Guang 司馬光	Patron of	Fan Chunli 范純禮	
Sima Guang 司馬光	Sacrificial prayer written for	Ding Du 丁度	

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

One can now sort on the separate columns:
 現在我們可以按照不同欄位分別排序：

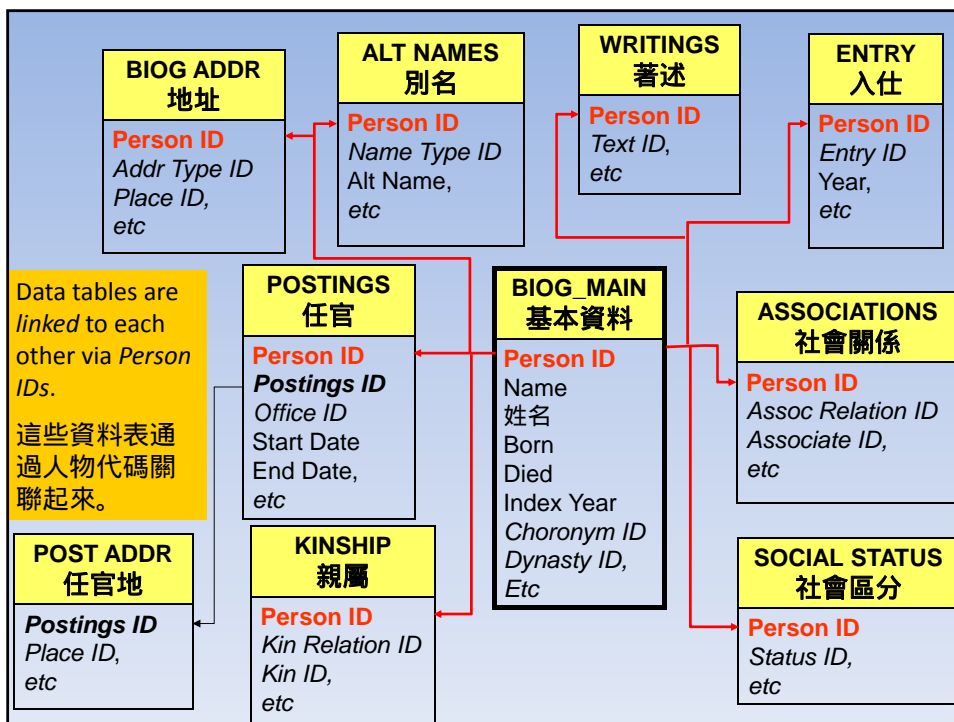
Name 姓名	Dates 日期
Sima Guang 司馬光	1019-1086

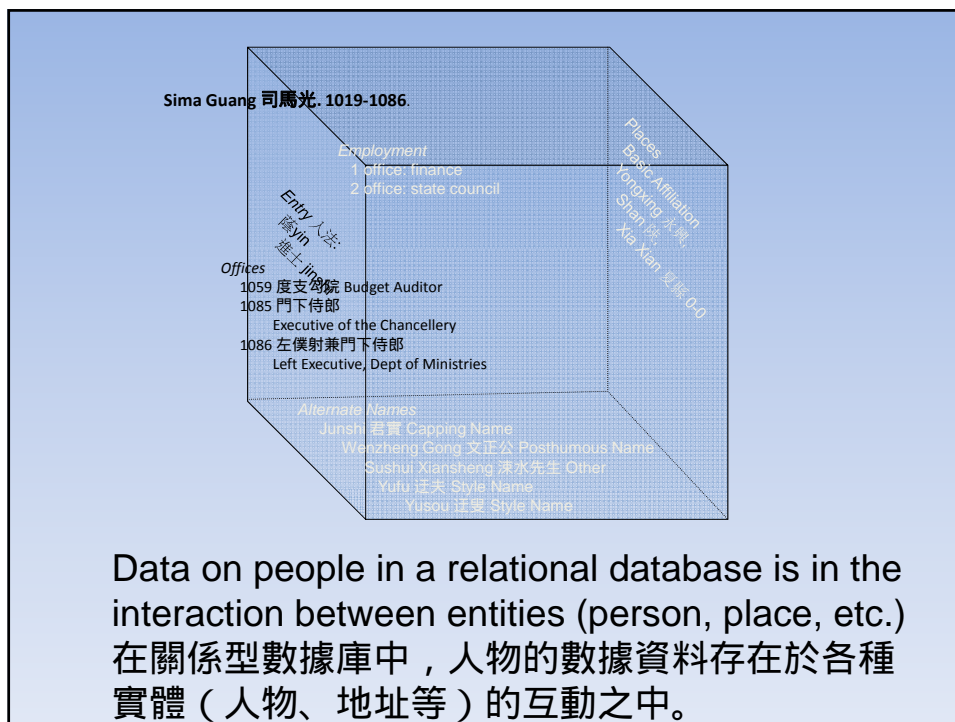
↓

Person 人物	Posting Date 任命日期	Office Title 官名
Sima Guang 司馬光	1059	度支勾院 Budget Auditor
Sima Guang 司馬光	1085	門下侍郎 Executive of the Chancellery
Sima Guang 司馬光	1086	左僕射兼門下侍郎 Left Executive, Dept of Ministries

↓

Person 人物	Association Type 社會關係	Associate 社會關係人
Sima Guang 司馬光	Yuanyou member (元祐黨)	(not applicable)
Sima Guang 司馬光	Desires opposed by	An Dun 安惇
Sima Guang 司馬光	Sacrificial prayer written by	Chao Buzhi 晁補之
Sima Guang 司馬光	Patron of	Fan Chunli 范純禮
Sima Guang 司馬光	Sacrificial prayer written for	Ding Du 丁度



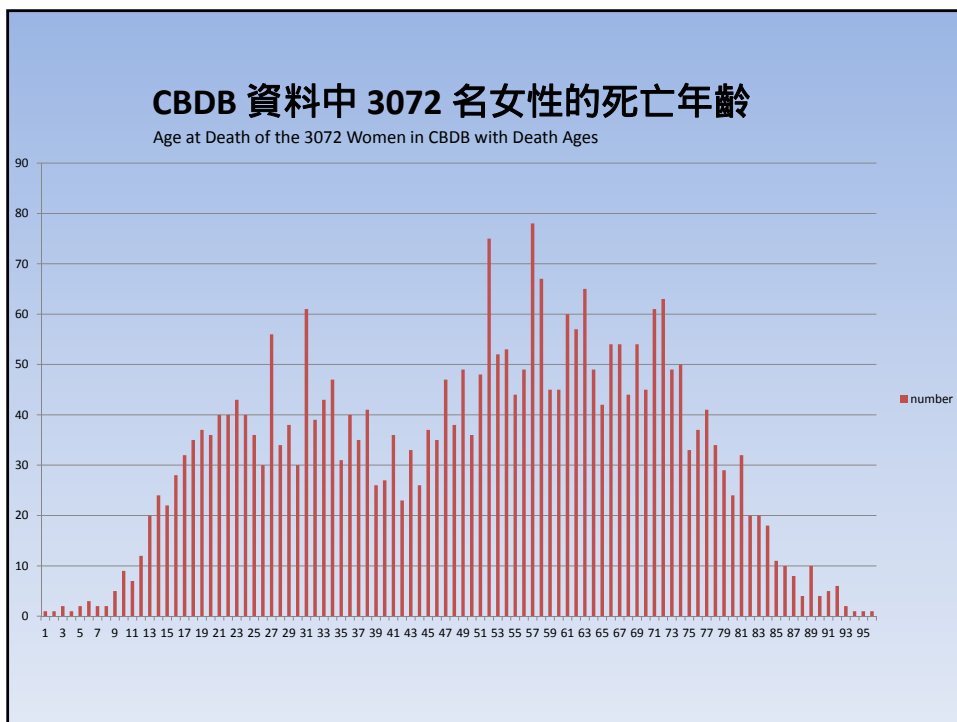
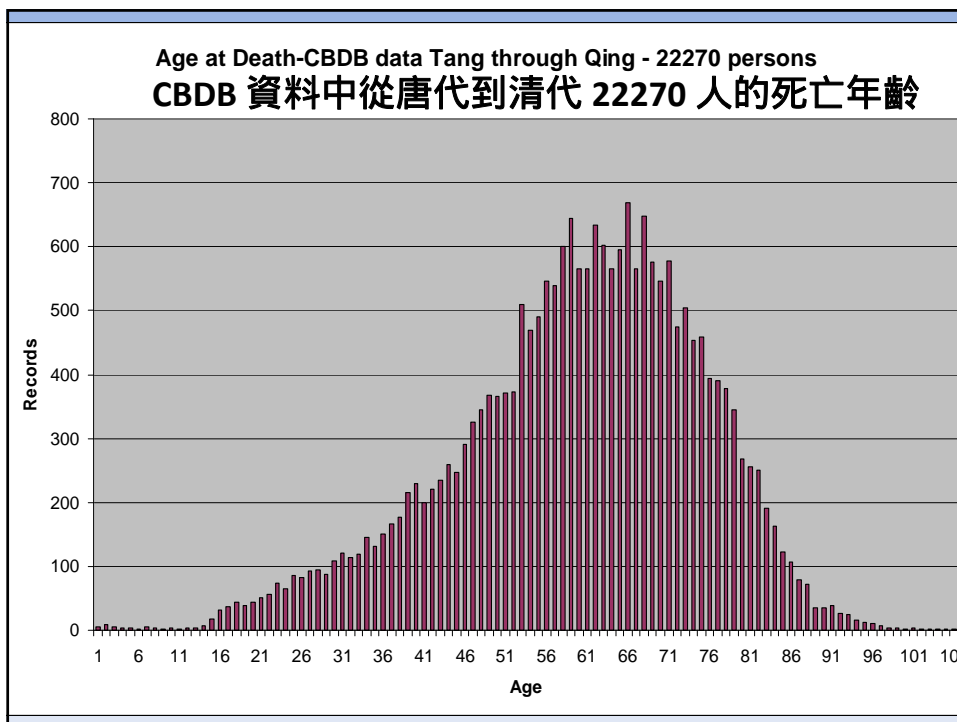


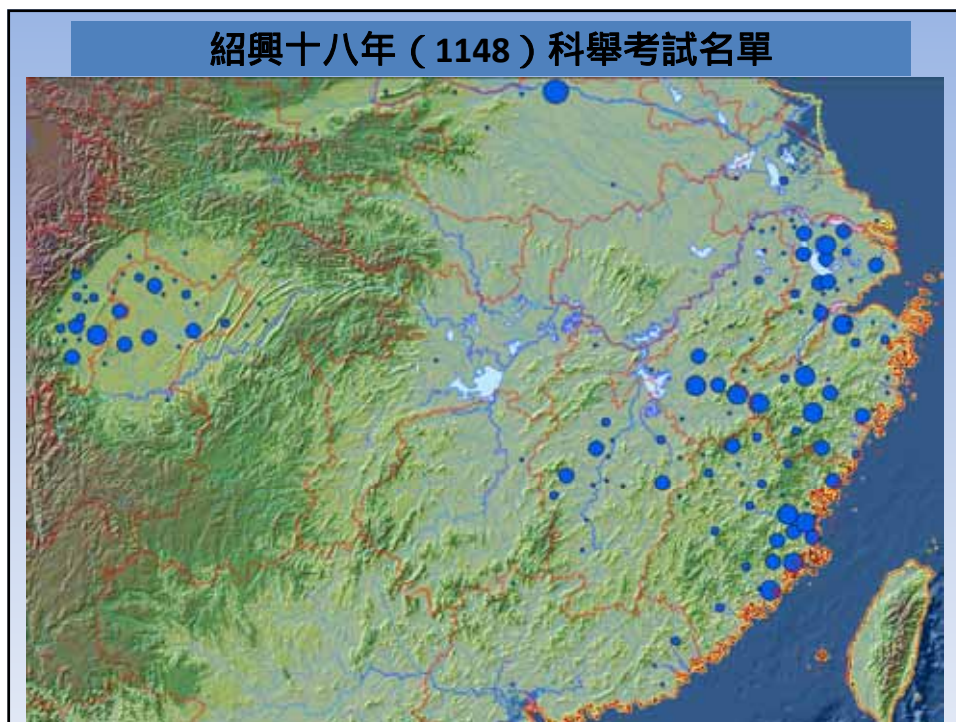
[English]Prosopography

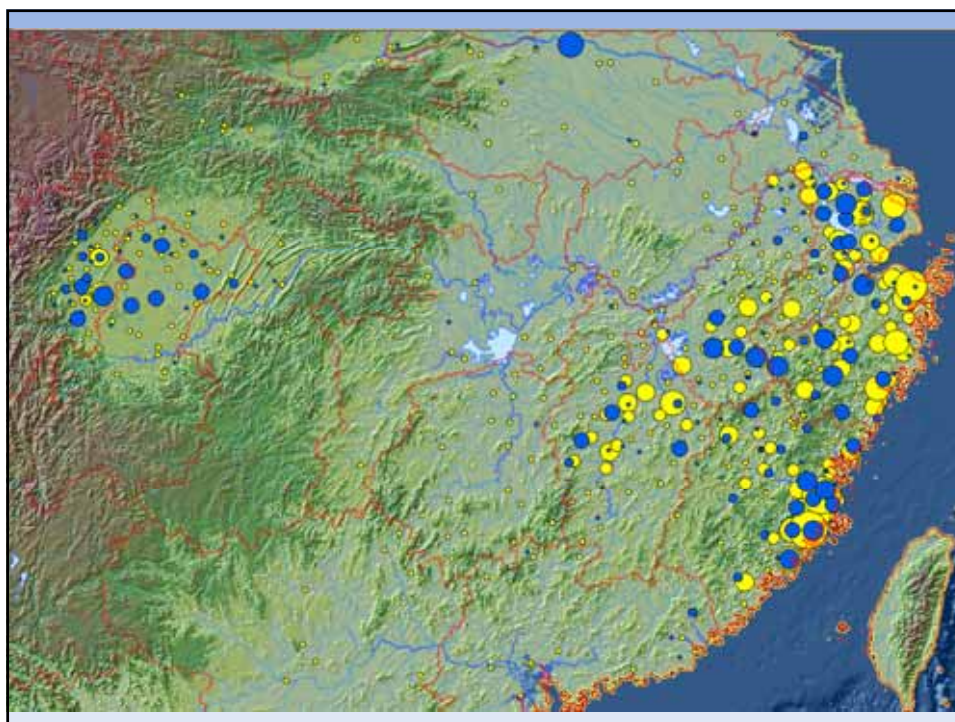
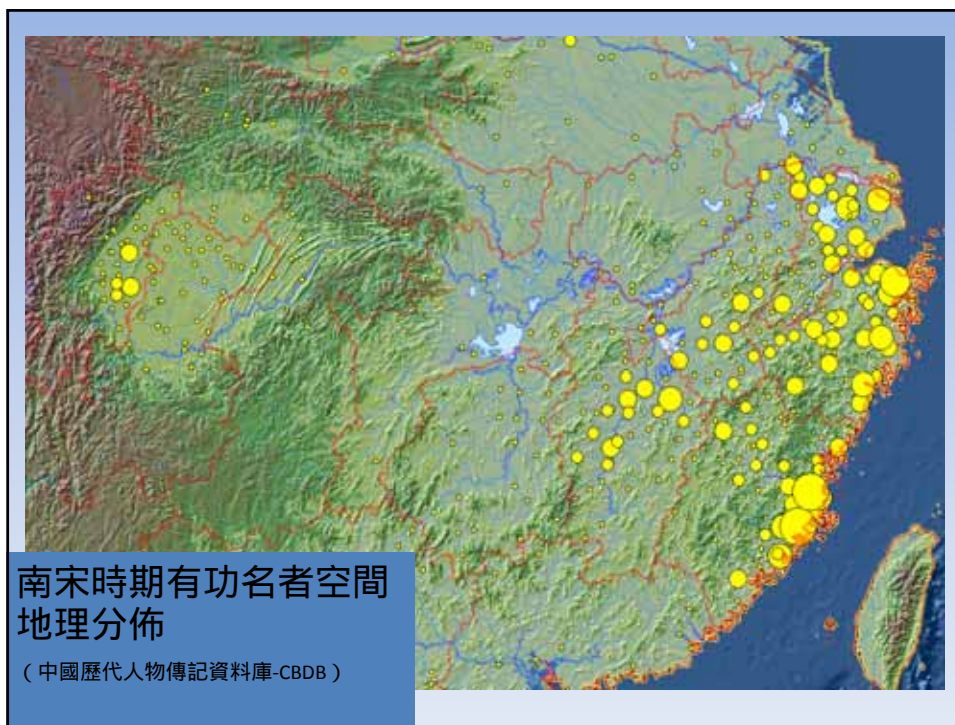
'Prosopography' is the investigation of the common background characteristics of a group of actors in history by means of a collective study of their lives. The method employed is to establish a universe to be studied, and then to ask a set of uniform questions – about birth and death, marriage and family, social origins and inherited economic position, place of residence, education, amount and source of personal wealth, occupation, religion, experience of office and so on. The various types of information about the individuals in the universe are then juxtaposed and combined, and are examined for significant variables. They are tested both for internal correlations and for correlations with other forms of behaviour or action. (L. Stone, 'Prosopography', in F. Gilbert and S. Graubard eds., *Historical Studies Today* (New York, 1972))

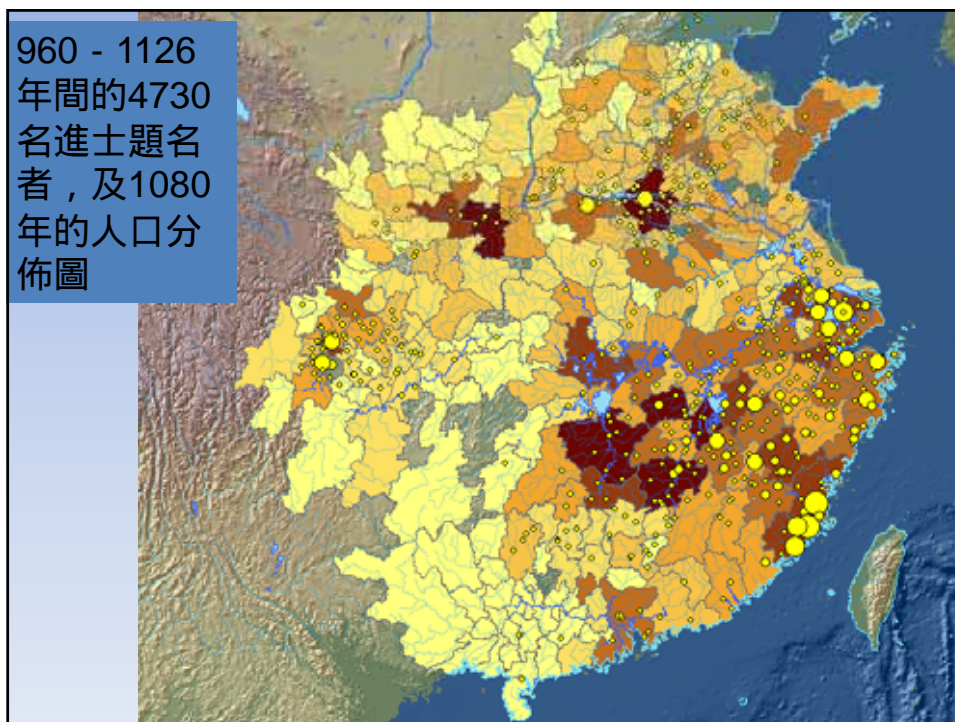
[中文]群體傳記學

- 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。(L. Stone, 'Prosopography', in F. Gilbert and S. Graubard eds., *Historical Studies Today* (New York, 1972))



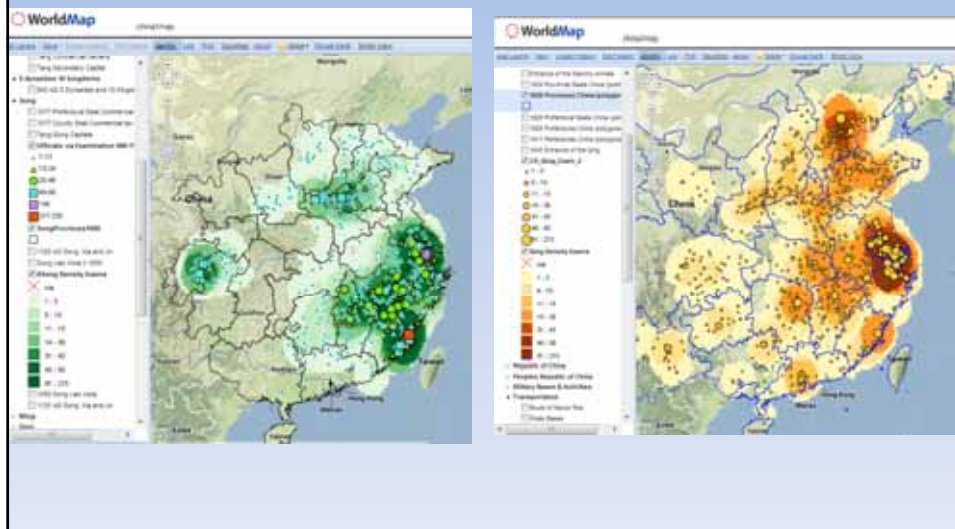






CBDB中北宋(960-1126)進士的空間分佈

CBDB 中清代(1644-1905)進士的空間分佈





Social Networks 社會網絡

“短期來講，個體創造了網絡；從長遠來講，網絡創造了個體。”

“In the short term individuals create networks; in the long term networks create individuals.”

(borrowing from John Padgett, University of Chicago)

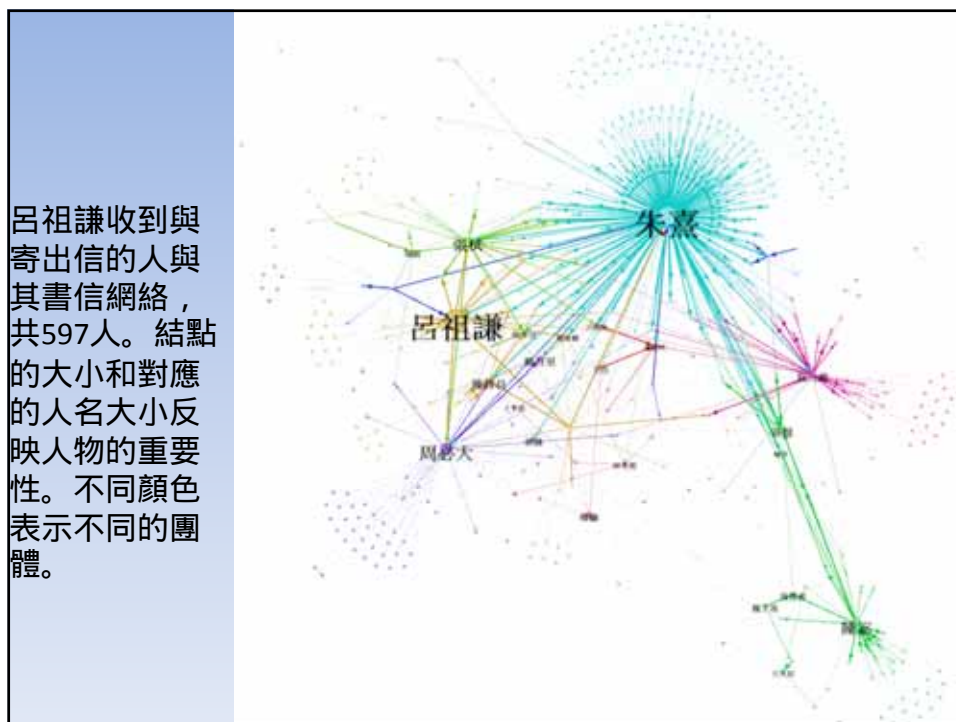
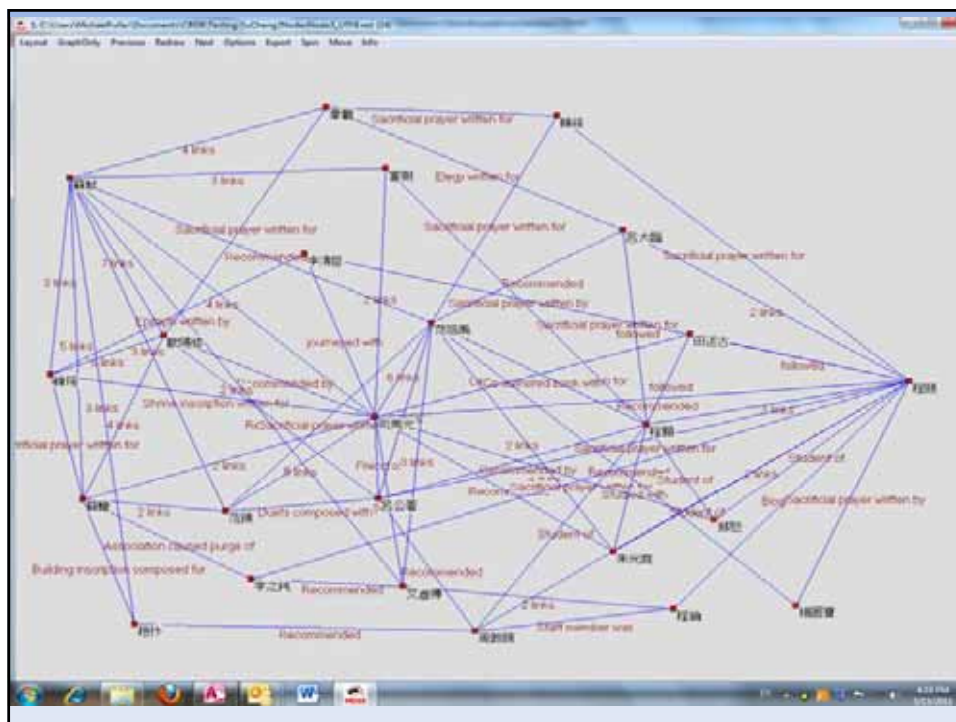
[English] Social Network Analysis

社會網絡分析

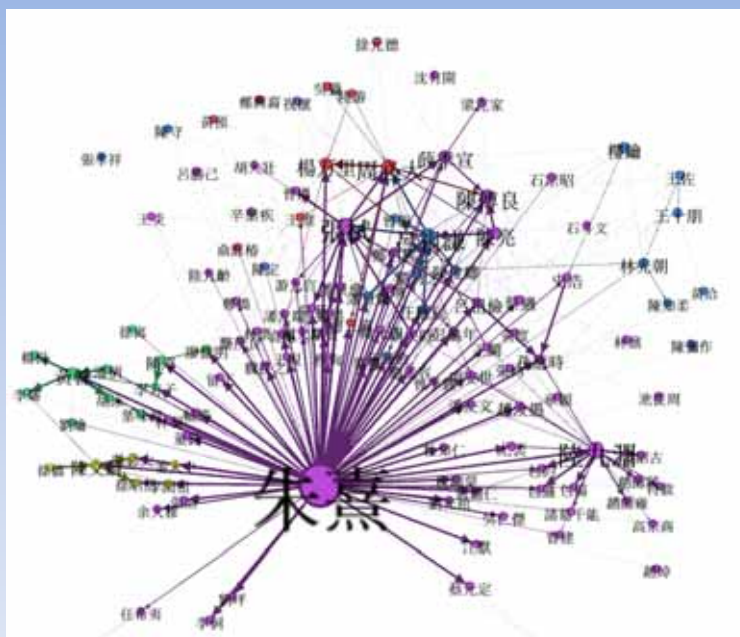
Conceptualizing community as collections of personal relationships ... provides historians with a blueprint for evaluating when, how and why people in the past used kin and non-kin in the course of their lives. The findings of social network analysts that people need and seek emotional and economic support of different kinds, from different kinds of people, suggest new analytical imperatives. It is not enough now to look solely at how people used kin in times of crisis. Rather, historians need to pursue how people in the past used the kin and friends they had, for different things, throughout the life course, and in the context of the opportunities they enjoyed and the constraints they faced courtesy of demography and culture. Other approaches might be applied to the problem, but HSNA contains the essential perspectives that cannot only advance the debate, but also help historians to meet Tilly's challenge to connect the lives of ordinary people to large-scale change in meaningful ways. (Charles Wetherell, "Historical Social Network Analysis," *International Review of Social History* 43 (1998), Supplement)

[中文] 社會網絡分析

社會網路關係(Social network analysis, SNA)係L. Stone關於群體傳記學方法論討論的例子之一。如同Charles Wetherell所述：「個人關係組成之集合體的概念化，提供歷史學家評估古人於何時、如何，及為何利用親族與非親族關係。社會網路關係分析家發現，人們須從不同的社會關係中、不同的人身上，尋求情緒上與經濟上的支持。因此，僅研究人們如何於危機時刻利用親族關係已不足夠；相反地，歷史學的研究必須涵蓋過去人們如何為不同目的而利用親族與朋友關係，以及此一利用關係的優勢與限制。事實上，社會網路關係做為一種研究方法不僅有助於此一論辯，更幫助歷史學家達到Tilly所提出的挑戰：將平民百姓的日常生活與大規模的社會變遷作有意義的連結。(Charles Wetherell, "Historical Social Network Analysis," *International Review of Social History* 43 (1998), Supplement)

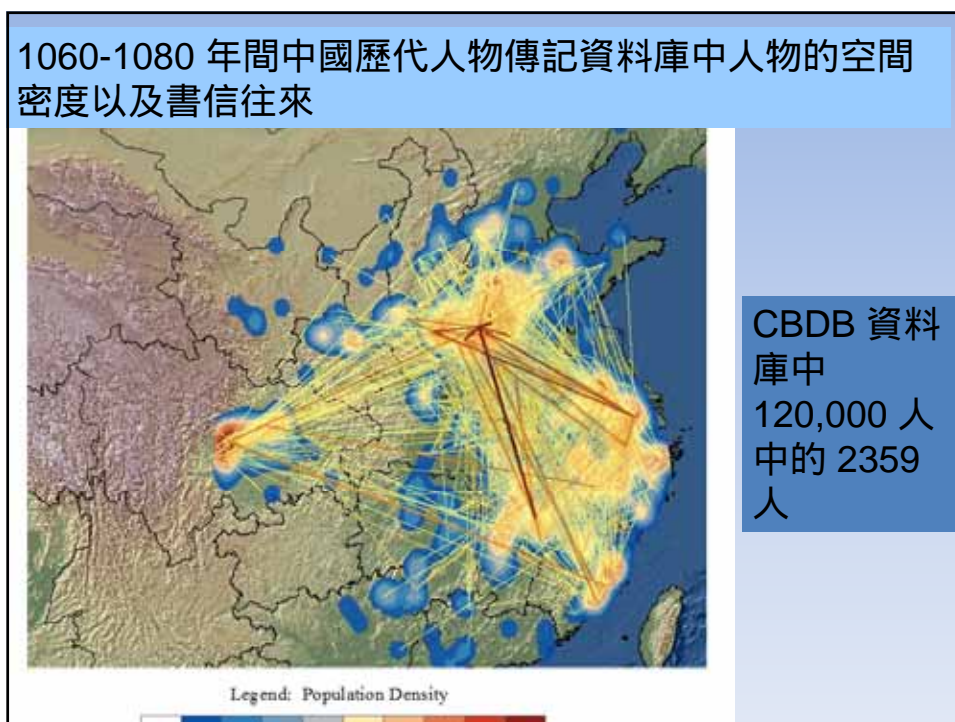
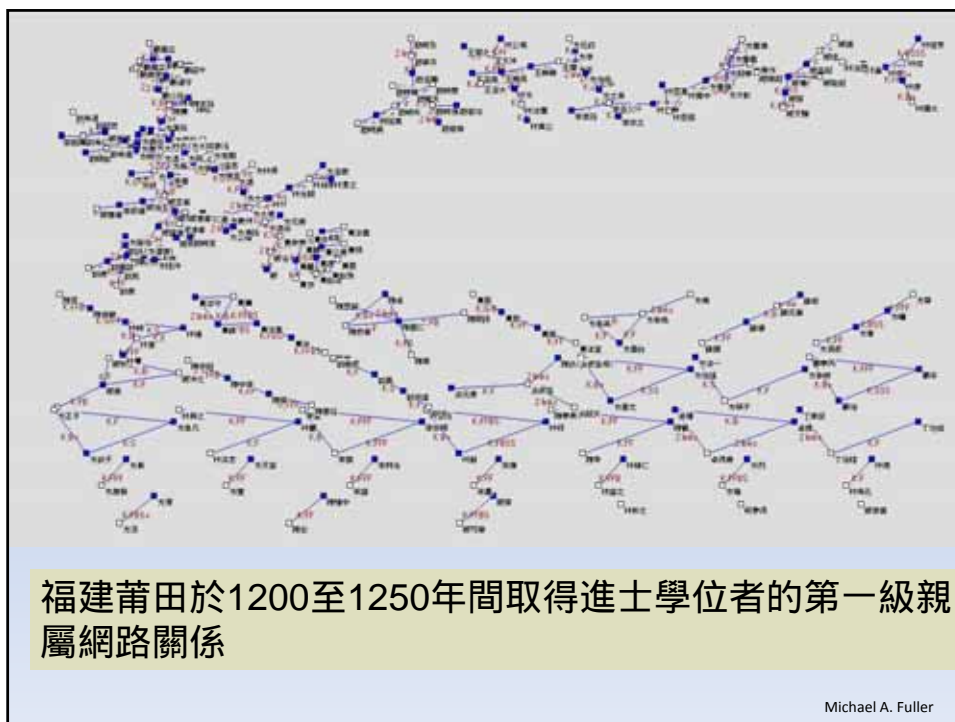


朱熹的書信網絡。這幅圖顯示了接收過朱熹兩封以上書信的人，以及他們之間的書信往來。人名大小反應了書信往來的數量；顏色用於區分子網絡



福建莆田於1050至1100年間取得進士學位者的第一級親屬網路關係

Michael A. Fuller



三種使用方式

中國歷代人物傳記資料庫 (CBDB)

首頁 | 關於我們 | 資料來源與涵蓋範圍 | 方法論 | 研討會 | 下載 | English

首頁

歡迎蒞臨中國歷代人物傳記資料庫!

資料庫簡介

中國歷代人物傳記資料(或稱數據)庫係線上的關係型資料庫。其遠程目標在於系統性地收入中國歷史上所有重要的傳記資料，並將其內容毫無限制地、免費地公諸學術之用。截至 **2013年10月**為止，本資料庫共收錄約**328,000**人的傳記資料。這些人物主要出自七世紀至十九世紀，本資料庫現正致力於增錄更多的明清代人物傳記資料。本資料庫除可作為人物傳記的一種參考資料外，亦冀可數統計分析與空間分析之用。下圖是 CBDB 中已知籍貫的 67,000 歷史人物的**地理分佈圖**。(點擊可放大)

訂閱 CBDB 更新消息

若您希望未來持續收到 CBDB 更新的消息，請點選以下連結，輸入您的 email 帳號及相關資訊，我們未來會自動將 CBDB 更新的消息送至該信箱。

[請點選此連結以前往訂閱表填寫資料](#)
(註：若您無法連結該訂閱表，請 email 至 shihpelchen [at] fas.harvard.edu 通知計畫經理陳詩珍。)

[請點選此連結回報錯誤](#)

如果您在CBDB中發現了任何錯誤或問題，請告訴我們!

鏈接

- 中文**查詢系統**。
- 英文 **查詢系統**。
- **線上輸入系統**

—您可以使用“Guest”帳號(無需密碼)，請任

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

人名查詢
地名查詢
官名查詢
關鍵詞查詢
畫廊查詢
異屬/人際關係查詢
檢索結果輸出
社會關係分析
操作說明
名詞說明

人名查詢

沒有查詢結果, 請再來開始查詢

傳主資料

姓名/別名/外文名:

年代(指數年): 至

朝代:

地名: [尚未定義]

人名查詢說明

人名查詢可以根據姓名、指數年、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。預設狀態下，查詢範圍包含 CBDB 中的所有人物，但您也可以通過調整左邊任一查詢欄位中的參數來進一步篩選查詢結果。

請進一步閱讀【操作說明】中的“分類查詢”

範例

利用人物姓名中的關鍵字來查找人物。比如說，在【姓名】欄輸入 Wang Anshi (拼音) 或王安石 (漢字)，都將產生一筆查詢結果 (截止至 2011 年二月份為止)：十一世紀的王安石 (Wang Anshi)。

如果只知道人物姓名中的一部份的話，您可以通過調整時間範圍來篩選查詢結果。比如說，維字在人名中很常見 (截止至 2011 年二月份共有 463 例記錄)，但當我們將查詢的時間範圍限定在唐朝時，查詢結果就會被縮減到 5 例；在【朝代】一欄的選單中選擇唐，則王維 (Wang Wei) 將出現在查詢結果中的第五位。

另一種常用的情況是，透過輸入人物姓氏來篩選出一批相同姓氏的人物。例如，在【姓名】欄中輸入 狐狐 兩字，將會產生 54 例結果。我們可以對這 54 例結果進行進一步篩選，使其只顯示活躍在北魏 (5 例) 或者活躍在唐朝 (32 例) 的人物。

在查詢結果中點選任一人物的姓名，便可顯示此人在 CBDB 中的完整記錄。

注：請用 "v" 來代替拼音中的 "u"。

cbdb.fas.harvard.edu/cbdb/cbdbedit3271261483211-20140671125219

Getting Started Imported From... Imported From IE Add to Del...

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

管理員: PBOL

最新新增人名:

書名新增:

人名新增:

人名查詢: (0筆)

查詢說明

- 於空格處輸入人名，點選 **查詢** 可列出包含所鍵入字的人名來：
例如：輸入「文」一字，系統列出的是姓名中包含「文」一字的人名

于仲文 (Yu Zhongwen)
干文傳 (Gan Wenzhuan)
文天祥 (Wen Tianxiang)

- 於空格處輸入人名，點選 **瀏覽** 可列出以鍵入字帶頭的人名來：
例如：輸入「文」一字，系統列出的是「文」字帶頭的人名

文士慶 (Wen Shiqing)
文元發 (Wen Yuanfa)
文天祥 (Wen Tianxiang)

- 也可以輸入人名的拼音進行查詢：
例如：輸入「Ouyang」一字

歐陽公瑾 (Ouyang Gongjin)
歐陽文記 (Ouyang Wenji)

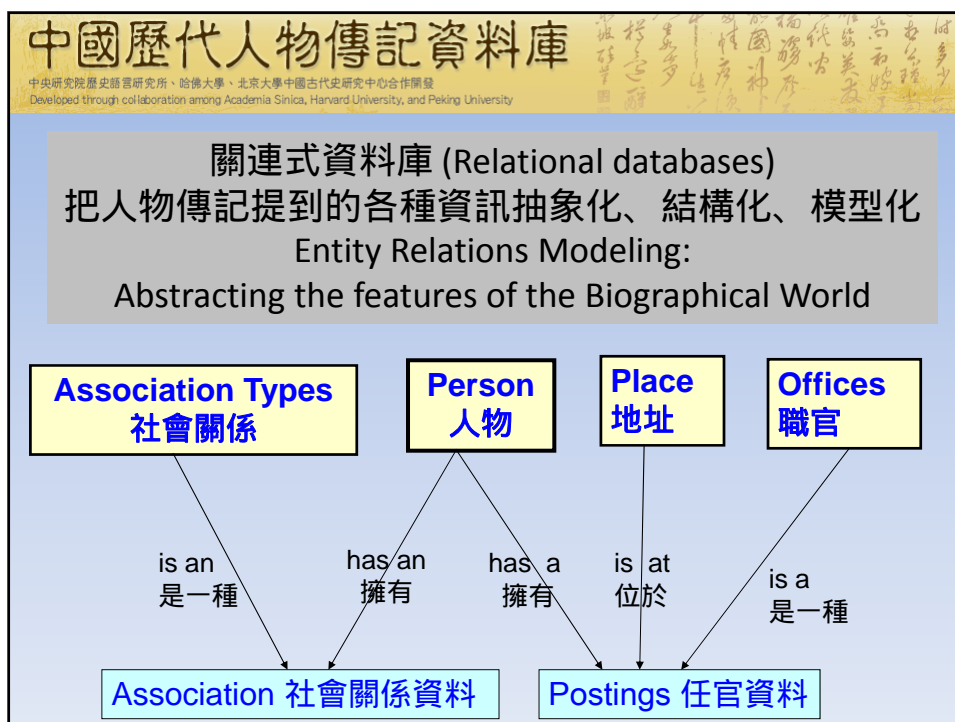


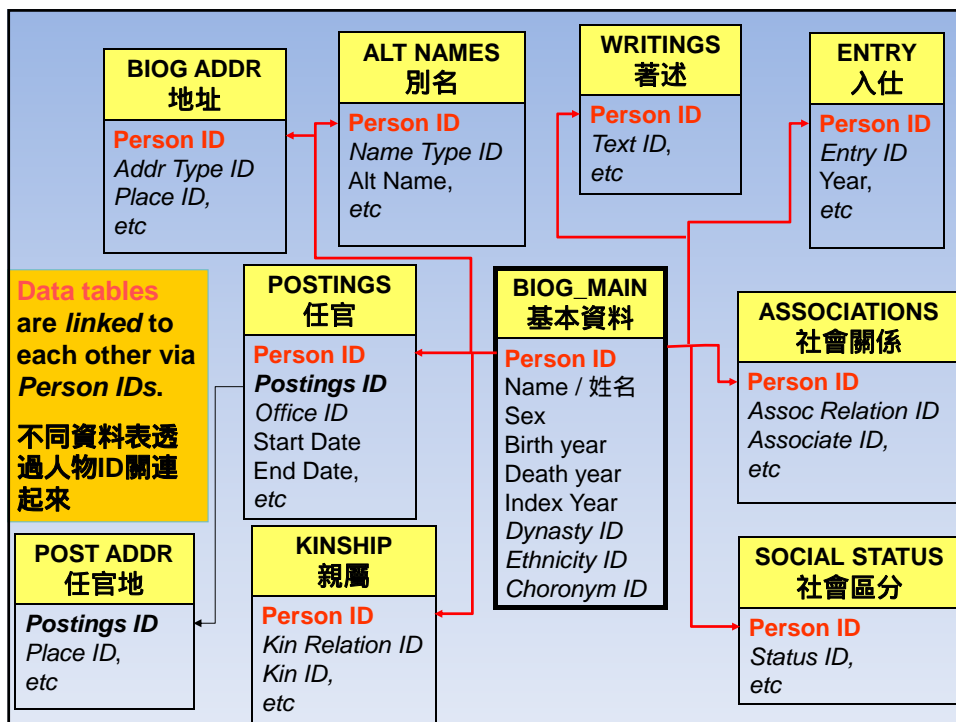
Chen Shih-pei 陳詩沛

DATA ACQUISITION FOR CBDB

CBDB用到的數位技術/工具

- Relational database 關連式資料庫
- 沒有數位全文: 明代進士登科錄
- 有數位打字全文
 - XML Tagging: 全文標記
 - 人工標記: 唐代墓誌銘
 - 自動標記: 宋/元/明人傳記資料索引
 - Regular expressions 正則表達式
- 工具
 - CBDB Regex Machine
 - 地方志 interface
- 跨資料庫合作: CBDB API





明進士登科錄

第一甲三名

賜進士及第

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	姓名	籍貫	戶籍	科目	字	年齡	生日	曾祖	曾祖任官	祖	祖任官	父	父任官
2	第一甲三名 賜進士及第												
3	吳伯宗	江西撫州府金谿縣	儒籍	書經	伯宗	三十八	八月二十九日	可	宋登仕郎書院進士	泰運		儀	元鄉貢
4	郭紳	山西潞州府襄垣縣	民籍	春秋	子翔	三十三		唐		成		希賢	
5	吳公達	浙江處州府麗水縣	儒籍	詩經	致中	二十四	二月初十日	良之	宋進士	繼祖		世德	元鄉貢
6	第二甲一十七名 賜進士出身												
7	楊自立	江西吉安府泰和縣	儒籍	春秋	子晉	三十八	八月十一日	復圭	元贈朝列大夫富川尹	會可		觀	
8	趙友庭	浙江紹興府會稽縣	儒籍	春秋	宗典	四十二	十二月初五日	崇構	宋禮興府通判	必奎		良通	
9	仇敬	山西平陽府曲沃縣	民籍	書經	仲立	三十二	六月二十日	守信		維		惟邦	

傳記內容

No大和005【蓋】失。
【志文】唐故朝散大夫臨晉縣令上柱國李府君墓誌銘並敘從弟河南府洛陽縣尉**行方**撰公諱鼎，字**鼎**，其先隴西成紀人也。昭行茂績，煥乎前史。五代祖孝恭，封河間王，行台尚書左僕射、三十州大總管謚曰**元**，配饗神堯皇帝廟；高祖崇義，土襲舊封，官列右揆，河東道大總管，圖形陵煙閣。自陵煙而上至於景皇；凡五世，皆以人倫傑秀，河岳挺生，德被幽明，道該寰宇。曾祖尚道，太子通事舍人；祖汪，絳州翼城令；父饗，彭州唐昌令；並公方利器，休稱藹和，位不展才，名屈於命，

唐代墓誌銘： 全文+人工標記

基本資料 郡望 種族

別名資料 字 諡號 其他別名 (如小字、行第)

地名資料 籍貫 祖籍 祖先遷徙 遷居地 實際居住地 卒地 卒時 享年 本誌葬地 本誌葬時

親屬關係 曾祖 曾祖母 祖 祖母 父 母 嗣子 子女 孫 孫女 女婿 孫女婿 夫/妻

入仕資料 進士 明經 進士或明經登科年(年紀) 蔭補 蔭補時間 相關親屬

職業生涯 官名 任官地 朝代

社會關係 社會關係 關係人 關係類別

著述資料 書名

社會區分 社會區分 所讀書/經 聲譽

1	VP recordID	s_personid	c_name	s_person_cite	c_links_view	c_family
2		146386	Kuili Qian	庫狄干		FALSE
3		146387	Kuili Luo	庫狄洛		FALSE
4		146388	Cui Mo(Cuilihezhengfa)	崔某(庫狄真相夫)		FALSE
5		146389	Cui Xingcan	崔行鑑		FALSE
6		146390	Cui Xingbo	崔行範		FALSE
7		146391	Li Mo(Lihezhengfa)	李某(庫狄真相妻)		FALSE
8		146392	Mou Shao(Mouyuanzhengfa)	某(某月相曾祖)		FALSE
9		146393	Mou Tongzhi(Mouyuanzhengfa)	某(某月相祖父)		FALSE
10		146394	Mou Chuanzhi(Mouyuanzhengfa)	某(某月相父)		FALSE
11		146395	Mou Junzhi(Mouyuanzhengfa)	某(某月相子)		FALSE
12		146396	Mou Gu(Mouyuanzhengfa)	某(某月相曾祖)		FALSE
13		146397	Mou Wei(Mouyuanzhengfa)	某(某月相祖父)		FALSE
14		146398	Mou Qi(Mouyuanzhengfa)	某(某月相父)		FALSE
15		146399	Mou Xingbo(Mouyuanzhengfa)	某(某月相子)		FALSE
16		146400	Yuan Xiang	元匡		FALSE
17		146401	Nu Lu	那盧		FALSE
18		146402	Nu Luhe	那盧和		FALSE
19		146403	Yuan Yi	元乂		FALSE
20		146404	Qo Jin	麴進		FALSE

1	kin_id	kin_name	VP recordID	s_personid	person name	c_kin_code	kin_rtl	c_source	c_pages
2	146386	庫狄干	3	138708	庫狄真相	62	祖父	32038	Wade 3
3	146387	庫狄洛	3	138708	庫狄真相	75	父		
4	146388	崔某(庫狄真相夫)	3	138708	庫狄真相	134	丈夫	32038	Wade 3
5	146389	崔行鑑	3	138708	庫狄真相	180	子	32038	Wade 3
6	146390	崔行範	3	138708	庫狄真相	180	子	32038	Wade 3
7	146391	李某(庫狄真相妻)	3	138708	庫狄真相	181	女婿	32038	Wade 3
8	146392	某(某月相曾祖)	4	138709	□月相	48	曾祖	32038	Wade 4
9	146393	某(某月相祖父)	4	138709	□月相	62	祖父	32038	Wade 4
10	146394	某(某月相父)	4	138709	□月相	75	父	32038	Wade 4
11	146395	某(某月相子)	4	138709	□月相	180	子	32038	Wade 4
12	146396	某(某月相曾祖)	5	138710	□某先	48	曾祖	32038	Wade 5
13	146397	某(某月相祖父)	5	138710	□某先	62	祖父	32038	Wade 5
14	146398	某(某月相父)	5	138710	□某先	75	父	32038	Wade 5
15	146399	某(某月相子)	5	138710	□某先	180	子	32038	Wade 5
16	146400	元匡	7	138712	元贊孫	62	祖父	32038	Kaiming 1
17	146401	那盧	7	138712	元贊孫	134	丈夫	32038	Kaiming 1
18	146402	那盧和	7	138712	元贊孫	560	嗣子	32038	Kaiming 1
19	146403	元乂	7	138712	元贊孫	75	父	32038	Kaiming 1
20	146404	麴進	8	138713	麴舉	62	祖父	32038	Kaiming 2

宋/元/明人傳記資料索引: 電腦自動標記 + 人工驗證

CHGB ID	Final	姓名	姓名	生年	卒年	傳記文
126290		一如	一如	1352	1425	字一範，號道翁。上虞孫氏子。年十三為僧。居杭州上天竺寺。永樂中奉命輯釋宗語錄。後又輯法華科注、大明法數等書。屢授僧錄司右翼教。年七十四卒。
126291		一清天	一清天			號蕭祖。處州歙縣氏子。宣德初落髮，主雲山弘慶寺。道風浩蕩，朝野知名。正統中入住雙徑禪院。道紅欽仰，而不喜其勸助。
126292		一覺	一覺			僧。祝髮後嘗刺血書華嚴經八十一卷。韻吟詠，有寒泓稿。與太子正字性彥良彥，一日太祖問彥良在鄉里與誰游，對曰僧覺性原，有詩贈之。因誦其詩。太祖稱善，即賜和。命徐瑛書之寄贈一覺。
126293		乙邦才	乙邦才			字奇山，貴州人。崇禎中以陳長擊就於河南江北間，累著戰功。史可法鎮揚州，以邦才為副將。弘光元年城破，自斃死。諡烈憲。
126294		丁一統	丁一統			鄞州諸生。崇禎十四年李自成陷登州，抗節死。
34287		丁士美	丁士美			字邦彥，號後溪，淮安清河人。嘉靖三十八年進士第一。萬曆初歷吏部左侍郎，居官絕偽造，與人處迥然若無能。意所不可，輒拂然見辭色。遭父復，以毀卒。諡文恪。
126295		丁川	丁川	1431	1478	字大容，號東陽，浙江新昌人。天順八年進士，授監察御史。成化中上言會昌侯孫繼宗父子權重，宜加裁損。萬妃專寵，干預外政，復上疏論諫。官至會都御史。歷歷延綏，進安邊十策，悉允行。成化十四年母喪歸，以毀卒。年四十八。有東陽文集。
126296		丁文暹	丁文暹			號竹坡，瑞金人。精於翎毛，兼善山水。時作枯木，蕭疎有致。
126297		丁天鏡	丁天鏡			字秀夫，宜興人。性耿介，登萬曆二十九年進士。令金谿六載，有惠政。績毫不諂。擬擢御史，候命卒於京。
126298		丁元復	丁元復	1525	1609	字見心，號玉陽，長洲人。隆慶五年進士，除陽信知縣，擢御史。官終浙江參議。卒年八十五。有片玉齋存稿。

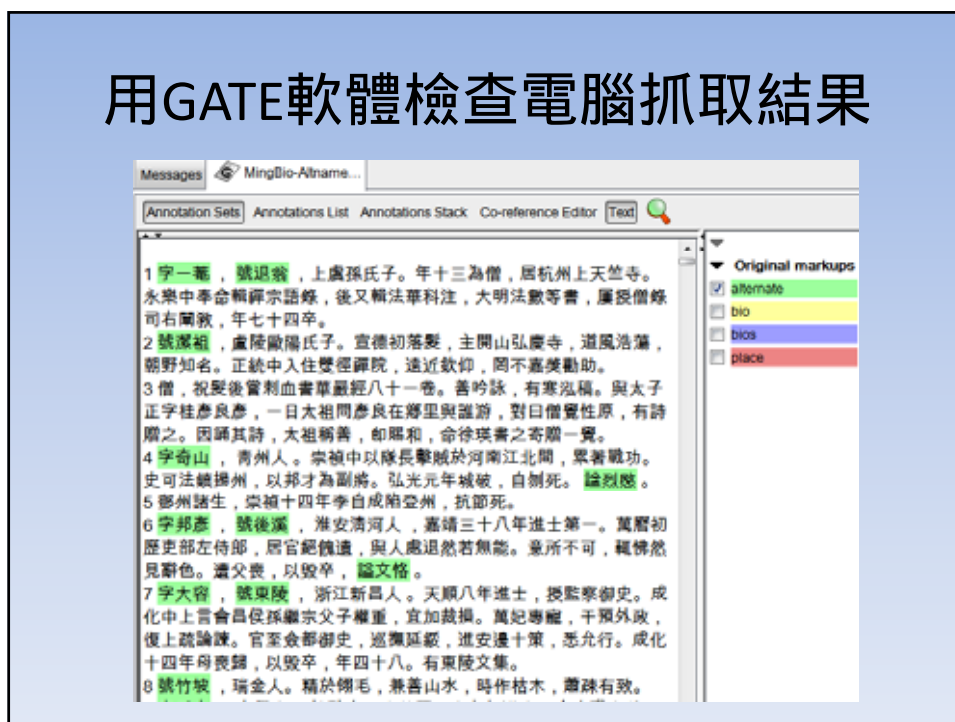
Regular expressions: 正則表達式

告訴電腦要抓「號」、「名」後面的文字 - 抓到標點符號為止

號(?P<name>[^ ,.。]{2,4})(?=[,.。:;])

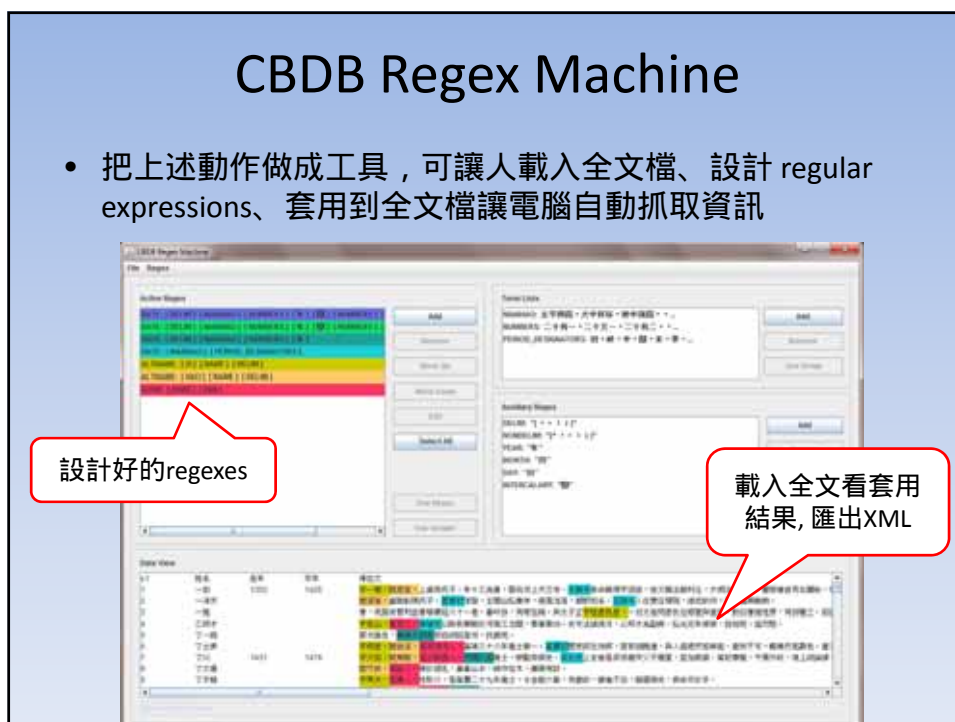
名(?P<name>[^ ,.。]{1,2})(?=[,.。:;])以字行

用GATE軟體檢查電腦抓取結果



CBDB Regex Machine

- 把上述動作做成工具，可讓人載入全文檔、設計 regular expressions、套用到全文檔讓電腦自動抓取資訊

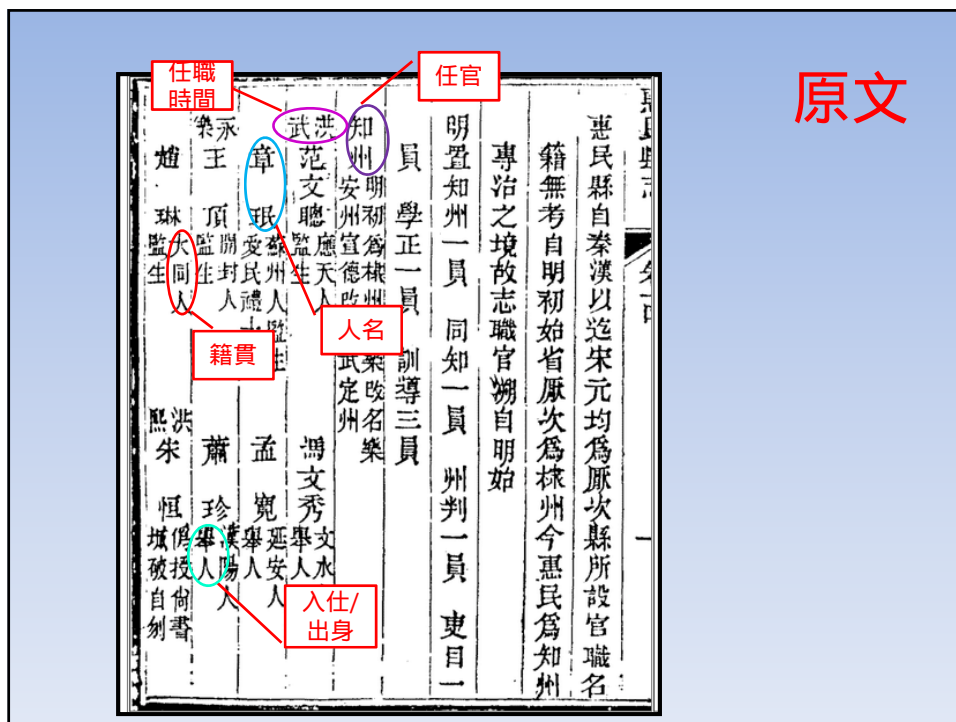


用 Excel 開 XML 轉成表格

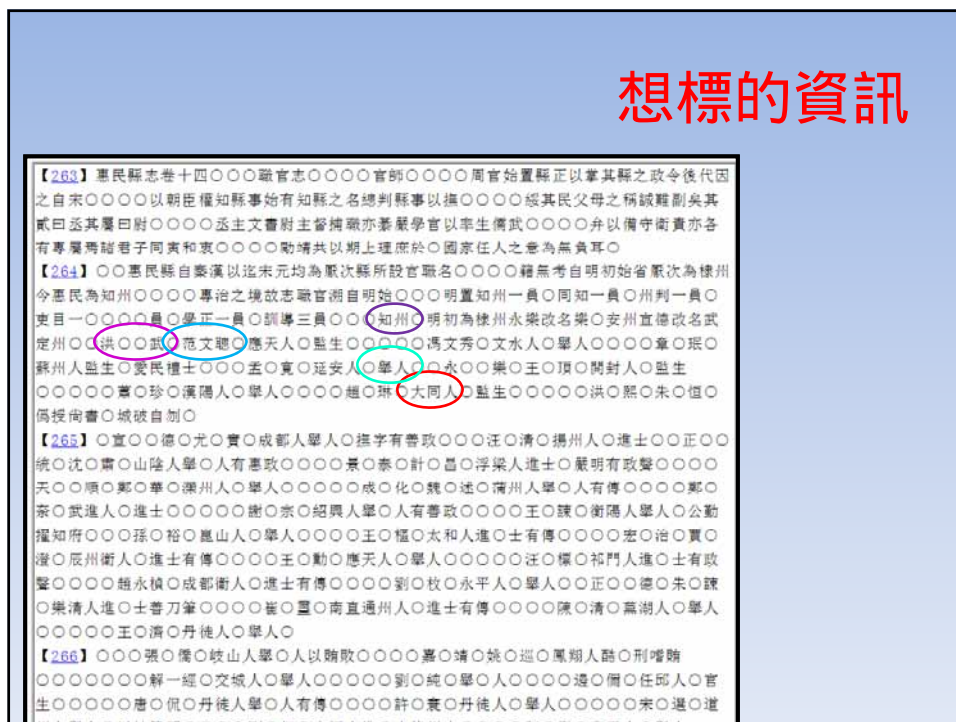
line	ALTNAME	name_type	name	ADDR	addr2
1					
2	字一	號	一		
3	2	號	遜翁		
4	3	號	潔祖		
5	4	字	桂彦		
6	5	字	奇山		
7				青州人	青州
8					
9					
10	7	字	邦彦		
11	7	號	後洪		
12				淮安清河人	淮安清河
13				與人	與
14	8	字	大容		
15	8	號	東陵		
16				浙江新昌人	浙江新昌
17	9	號	竹坡		
18				瑞金人	瑞金
19	10	字	秀夫		
20				宜興人	宜興

Computer Assisted Tagging and Data Extraction

原文



想標的資訊



- 作法: A computer Assisted Tagging and Data Extraction System

運用 regular expressions 來標記有特殊寫法的文字區段

流程:

Import
-> Tag
-> Export

#	Book ID	Section ID	方志	部	朝代	入仕地點	入仕時間	入仕方法	籍貫	職銜	後銜	年表	無蓋	物產	別名	人名	任官地點	任職時間	任職種類	前官	出處	其他
1	35213	98798	臨海縣志	縣志	明				肥鄉						縣令	蔣胤	成化年間					
2	35213	98798	臨海縣志	縣志	明										縣令	高顯	成化年間					
8	35213	98798	臨海縣志	縣志	明	怡鄉		舉人							縣令	趙璧	弘治年間					

【539】○○○教諭○毛希賢○見大○成碑○翟瑋○○○○主簿○高塔不歸○見舊○志○史進○至元時任見○峯道南縣志○明○○○○縣令○歐陽銘○慶陵○人○張惟德○葛文明○孫隆○以上皆洪○武年間任○姚著○□□□人○程朋○○○○王謙○江都○歲貢○彭自翳○真溪舉人○以○上皆永樂間任○蔣鳳○肥鄉○□□□高顯○余本政○江○夏○舉人○李仲良○漢陽舉人○以上○皆成化年間任○侯敬○靈寶○舉人○王世臣○開州○舉人○周致○鄞縣○進士○趙○璧○柏鄉舉人○以上○皆弘治年間任○許讚○靈寶○進士○馬暹○威縣○舉人○馬冕○□寧○進士○楊權○揚州舉人○以○

系統開發: 彭維謙 (台大資工系)

抓取「誰什麼時候任什麼官」

- Tagging Names
Separate the names into different rows

【264】○○惠民縣自秦漢以迄宋元均為厥次縣所設官職名○○○○籍無考自明初始省厥次為棗州

【262】惠民縣志卷十四○○○職官志○○○○官師○○○○周官始置縣正以掌其縣之政令後代因之自宋○○○○以明臣權知縣事始有知縣之名總判縣事以攝○○○○統其民父母之稱誠難副矣其或曰恐其覆曰對○○○亦主之責對主督攝職亦甚重舉官以事生備或○○○○亦以備中則書亦各

有專署焉諸君子焉

【264】○○惠民縣為知州○○○民目一○○○○縣州○○洪○○○蘇州人監生○愛世

○○○○○嘉○珍○漢陽人○舉人○○○○籍○齊○天同人○監生○○○○○洪○熙○朱○德○高授尚書○城破自刎○

洪○熙○○
朱○恒○○偽授尚書○城破自刎○

Tag: person
Value: 洪○熙
Remove this Remove this(with newline) Remove all Remove all(with newline)

• Adding shared information

Select certain words and add them as titles

Add shared information to a number of records at one time

知州 明 惠 縣 志

(知州) (知州) ○明初為棣州永樂改名樂○齊州官德改名武定州○

#	Book ID	Section ID	方志	部	頁數	朝代	入仕地點	入仕時間	入仕方法	籍貫	親屬關係	後官	年號	無意	物產	官職	別名	人名	任官地點	任職時間	任職種類	前官	出處	字	其他
1	35006	92077	惠民縣志	職官志												知州	范文聰	洪武							應天人監生
2	35006	92077	惠民縣志	職官志												知州	馮文秀	洪武							文水人舉人
3	35006	92077	惠民縣志	職官志												知州	童珉	洪武							蘇州人監生 愛民禮士
4	35006	92077	惠民縣志	職官志												知州	孟寬	洪武							延安人舉人
5	35006	92077	惠民縣志	職官志												知州	王頂	永樂							開封人監生
6	35006	92077	惠民縣志	職官志												知州	蕭珍	永樂							漢陽人舉人
7	35006	92077	惠民縣志	職官志												知州	趙琳	永樂							大同人監生
8	35006	92077	惠民縣志	職官志												知州	朱恒	洪熙							偽授尚書城 破自刎

(洪○熙) (知州) 朱○恒○偽授尚書○城破自刎○

• Creating your own Regexes

Finding patterns in the text

Load and run predefined regular expressions

(知州) (知州) (洪○熙) (洪○熙) (洪○熙) (知州) (永○樂) (永○樂) (永○樂) (知州) (洪○熙)

#	Book ID	Section ID	方志	部	頁數	朝代	入仕地點	入仕時間	入仕方法	籍貫	親屬關係	後官	年號	無意	物產	官職	別名	人名	任官地點	任職時間	任職種類	前官	出處	字	其他
1	35006	92077	惠民縣志	職官志					監生	應天						知州	范文聰	洪武							
2	35006	92077	惠民縣志	職官志					舉人	文水						知州	馮文秀	洪武							
3	35006	92077	惠民縣志	職官志					監生	蘇州						知州	童珉	洪武							愛民禮士
4	35006	92077	惠民縣志	職官志					舉人	延安						知州	孟寬	洪武							
5	35006	92077	惠民縣志	職官志					監生	開封						知州	王頂	永樂							
6	35006	92077	惠民縣志	職官志					舉人	漢陽						知州	蕭珍	永樂							
7	35006	92077	惠民縣志	職官志					監生	大同						知州	趙琳	永樂							
8	35006	92077	惠民縣志	職官志												知州	朱恒	洪熙							偽授尚書城 破自刎

regex: [input field]
[button]
[button] (en Group)

• Customizing the Regexes

Can modify the regular expressions to fit the variation of patterns in each gazetteer

The number.

The character "年"

can be modified by user

Delete Regex group

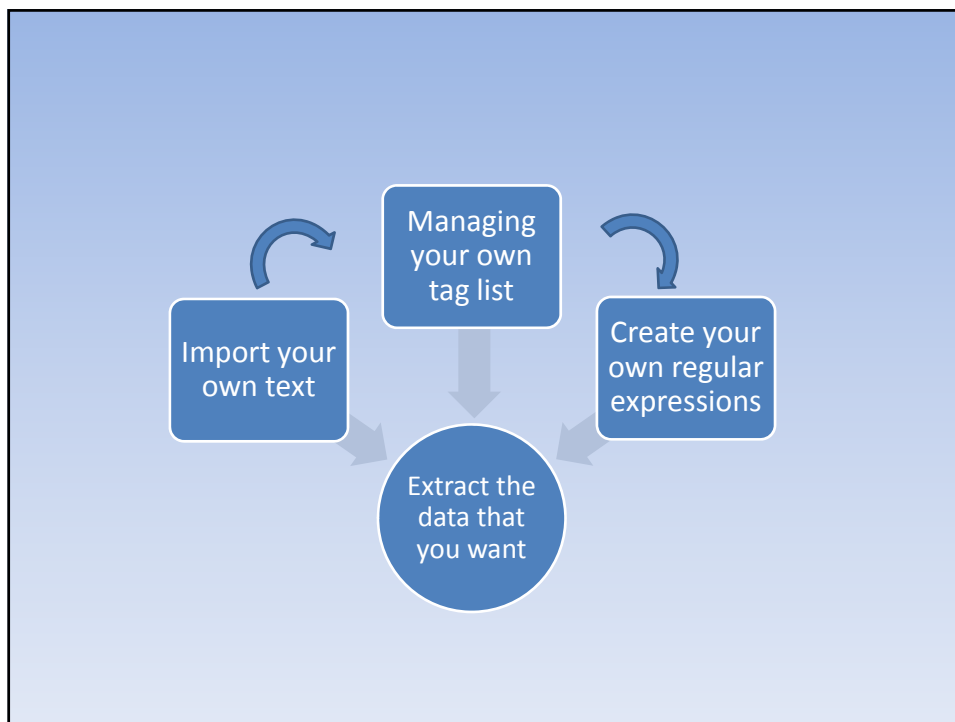
• Export

Export extraction result as a table

Exported text examples:

- (縣令) (明) (以上皆永樂間任) ○
- (成化年間) (縣令) (明) 蔣鳳○肥鄉○□□○
- (成化年間) (縣令) (明) 高顯○
- (明) (縣令) (成化年間) 余本致○江○夏○舉○人○

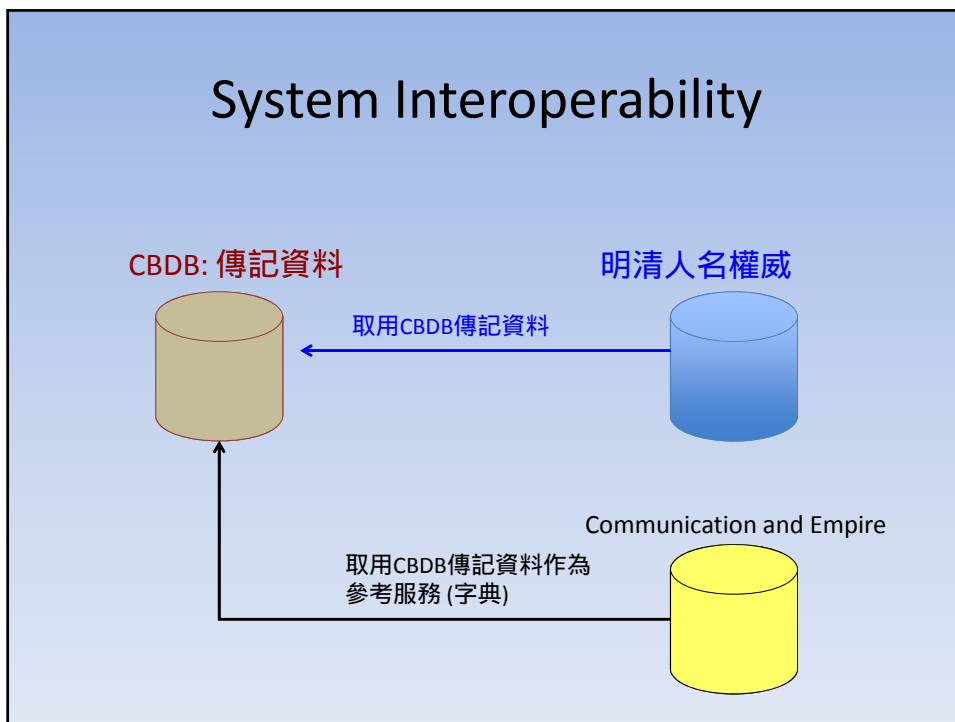
Book ID	Section ID	方志	部	有數	朝代	入仕地點	入仕時間	入仕方法	籍貫	親屬關係	後官	年號	無官職	物產	官職	別名	任官地點	任官時間	任職種類	前官	出處	字	其他
1	35213	98798	臨海縣志	職官	明				肥鄉						縣令	蔣鳳	成化年間						□□
2	35213	98798	臨海縣志	職官	明										縣令	高顯	成化年間						
3	35213	98798	臨海縣志	職官	明	江夏		舉人							縣令	余本致	成化年間						
4	35213	98798	臨海縣志	職官	明	漢陽		舉人							縣令	李仲英	成化年間						
5	35213	98798	臨海縣志	職官	明	靈壽		舉人							縣令	侯敏	弘治年間						
6	35213	98798	臨海縣志	職官	明	開州		舉人							縣令	王世臣	弘治年間						
7	35213	98798	臨海縣志	職官	明	鄞縣		進士							縣令	周致	弘治年間						
8	35213	98798	臨海縣志	職官	明	柏鄉		舉人							縣令	趙靈	弘治年間						



CBDB API: 資料庫之間的互通

- 不同團隊建立的資料庫, 各有不同著重的面相
 - CBDB
 - 明清人名權威資料 (中研院史語所, 故宮)
 - Communication and Empire (Leiden University)
- 若要求資料庫廣泛涵蓋各種面相的資料, 花很多力氣, 模糊焦點, 不容易
- =>互通, 互相串連資料

System Interoperability



Make an API call to get external data

The screenshot shows a search result for 'Zhang Qizao' in the CBDB database. The search results are highlighted with a red box:

CBDB: 183
 索引/中文/英文名稱 / 侯爵暨 Zhang Qizao
 指數年 (index year): 1002
 生卒:
 卒年: 北宋大中祥符7年 (1014)
 享年: 72
 鄉望: [未詳]
 出處: 宋人傳記資料索引(電子版), 頁13588
 2: Zhang(1) Qizao (183) His family moved to Luoyang when he was 8 because of disaster during the Five Dynasties. Zongjia's (107) and Zongjia's (1167) father, Ziqao's (1712) grandfather. Father-in-law of Wang(2) Jiahong's (1143) brother. Father (1945) Jiahong's son-in-law was Zhang(1) Qizao's (1099) older cousin. Jing's (1730) son, Zuo's (1745), Zhang(1) Zongjiao was Wang(2) Shao(1)'s (1877) son-in-law. XCD, 24.4a.8a. SHY SH, 49.3a.55, 268.15a.18a. CB, 3, 2420-21.
 別名: 諸姓文定, 字顯可。
 地理資訊, 籍貫(原本地址): 河南府 / 洛陽

Peter Bol 包弼德

HARVARDX AND CHINAX MOOCS 大型線上免費開放課程



“Knowledge is the most important currency of the 21st century.”

Drew Faust, Harvard University President

Goal 1: Improve Teaching

Enhanced online & on-campus learning

- 1st humanities course on edX
- integrated tools for campus students

Brought faculty together

- 59 faculty in HX pipeline
- 5 faculty town halls
- Joint Harvard-MIT summit



Goal 2: Advance Research

Built a team of research fellows

- creating an online learning community at Harvard and beyond
- received a grant from the Gates Foundation to study personalizing MOOCs
- HX & HILT fellows are collaborating



Published findings

- "Studying Learning in the Worldwide Classroom: Research into edX's first MOOC" (Breslow, Ho)

HarvardX

Goal 3: Expand Access

Engaged global learners

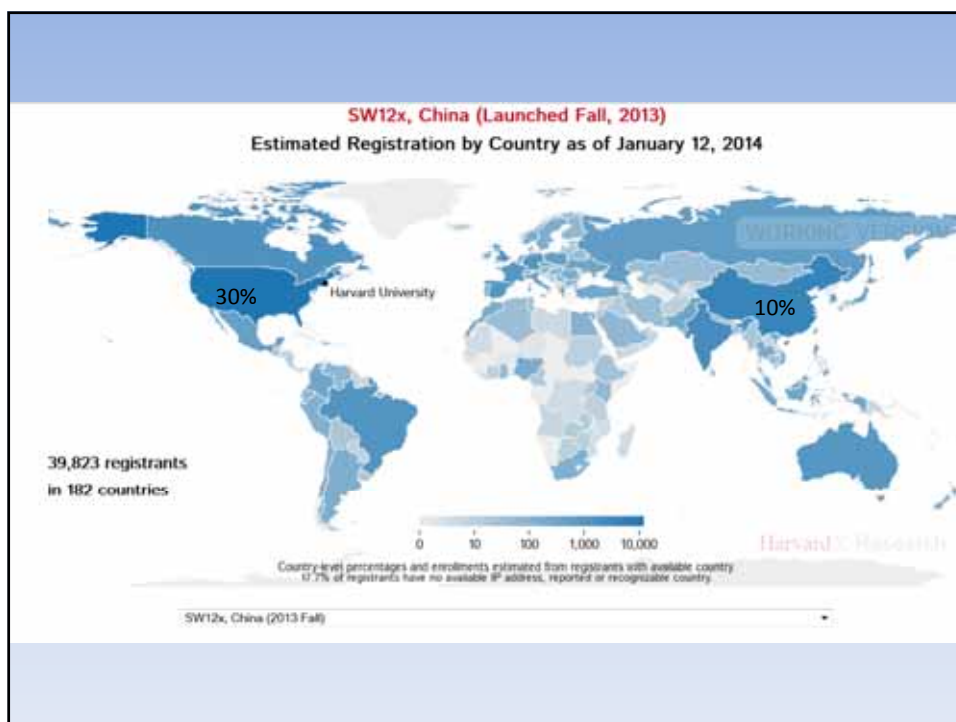
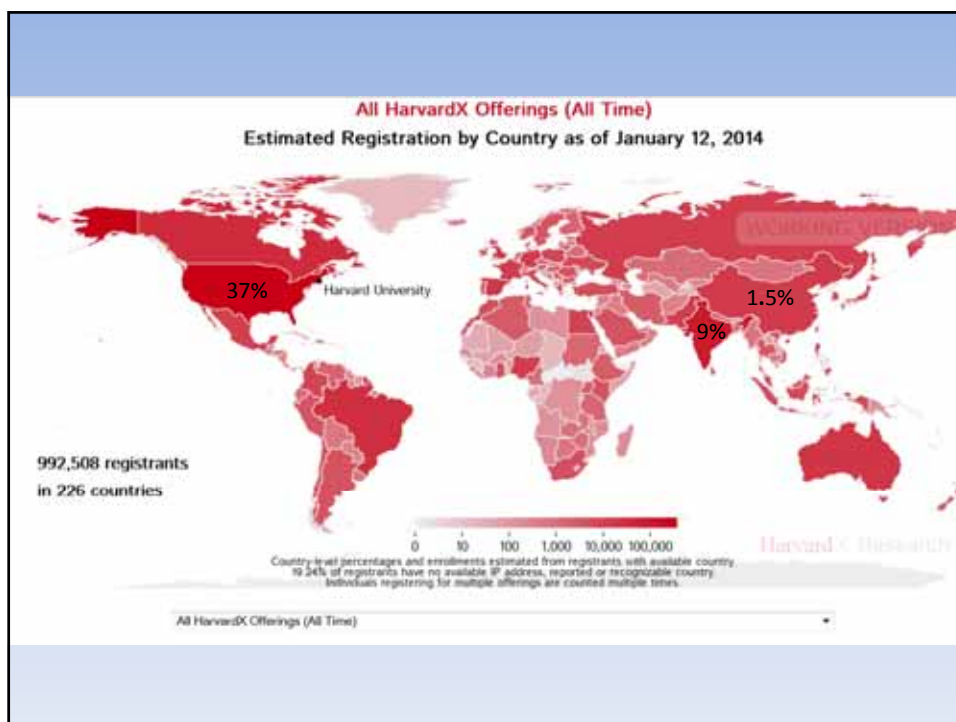
- 800,000+ students enrolled
- 10 completed courses and 30 future

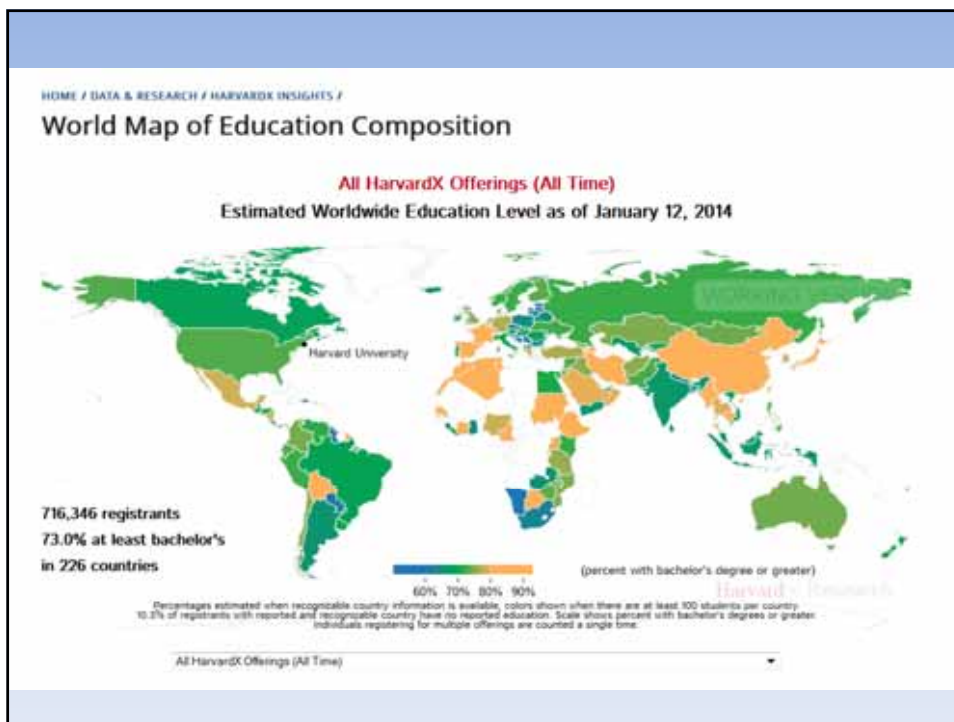
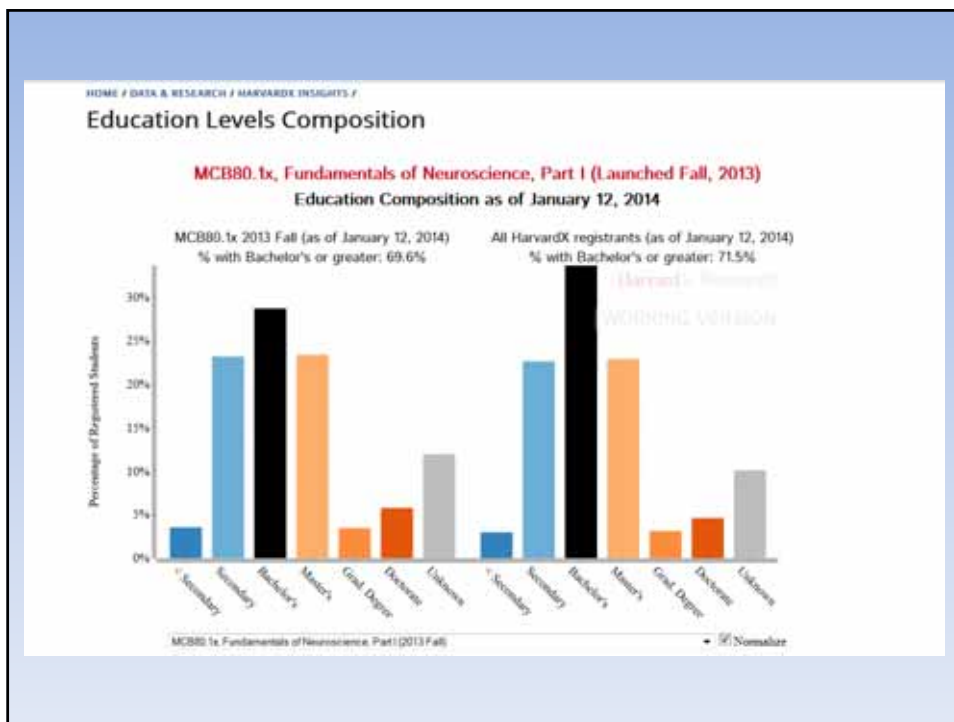
Offered breadth and depth

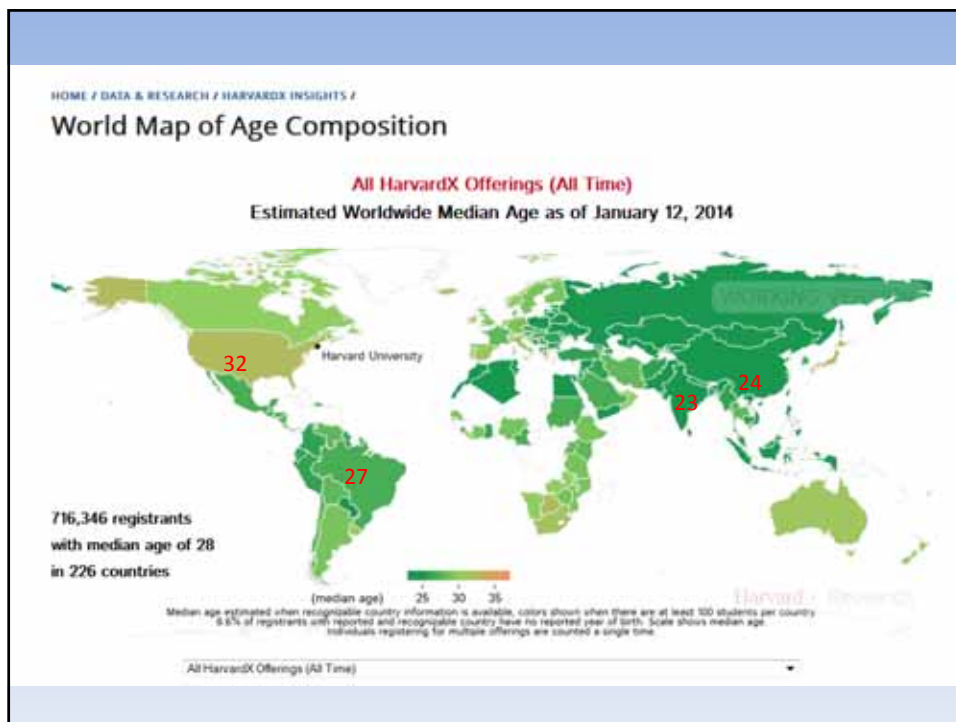
- 10 Harvard schools
- topics in the humanities, government, health, medicine

70%
of students
are international

HarvardX







Yu Wen 于文

CHINAX

The screenshot shows the edX course page for 'China'. At the top, there is a navigation bar with 'edX' logo, 'HOW IT WORKS', 'COURSES', 'SCHOOLS & PARTNERS', and a 'dashboard' button. Below the navigation bar is a large banner image showing a collage of Chinese landscapes and modern cityscapes. To the right of the banner is a video player with the title 'Watch the Course Intro Video'. Below the banner, the course title 'China' is displayed, followed by a description: 'China's past, present, and future: through history and geography, economy and ecology, philosophies and politics, literature and art.' There is also a section 'About this Course' with a note: '*Note - This is an Archived course*'. A sidebar on the right contains course details: 'School: HarvardX', 'Course Code: SW2x', 'Classes Start: 31 Oct 2013', and a 'Prerequisites' section. At the bottom of the sidebar is an 'Access Courseware' button. Below the main content area, there are social media icons for Twitter, Facebook, and YouTube.

Course trailer :
<https://www.edx.org/course/harvardx/harvardx-sw12x-china-920>

The screenshot shows a video player interface. On the left side, there is a navigation menu with the following items: 'What's this all about?', 'Meet ChinaX', 'The Big Picture', 'Introduction: Time, Space and Identity', 'Section 1: Time Assessment', 'Section 2: Space Assessment', 'Section 3: Identity Assessment', 'Section 4: China's History Mapped', 'Glossary', 'Feedback', and 'Reference'. The main video area displays a map of China with the title '2. USING CHINAX MAP'. The map shows a satellite view of China with various geographical features. To the right of the map, there is a text overlay that reads: 'So here you have a s about the same area roads show showing Play around and die maps are best for yo what you're interested in. Now you've already zoom in and zoom o But we can also tur off. Let's lock now at the Let's change first bac base map, which wo place names or road'. A 'YouTube' logo is visible in the bottom right corner of the video player.

The screenshot shows a discussion forum interface. On the left, there is a sidebar with a search bar and a list of discussion threads. The threads include:

- If your Chinese Zodiac is Ox, Snake or Rooster, join this thread for Week 5 Discussion.
- If your Chinese Zodiac is Tiger, Horse or Dog, join this thread for Week 5 Discussion.
- If your Chinese Zodiac is Rabbit, Ram or Pig, join this thread for Week 5 Discussion.
- If you were born from January to March, please join this thread for Week 4 Discussion.
- If you were born from April to June, please join this thread for Week 4 Discussion.
- If you were born from July to September, please join this thread for Week 4 Discussion.
- If you were born from October to December, please join this thread for Week 4 Discussion.
- (ChinaX Staff) Dynasty Song Thread

The main content area shows a post by user 'iosesleep' from 4 months ago. The post text is:

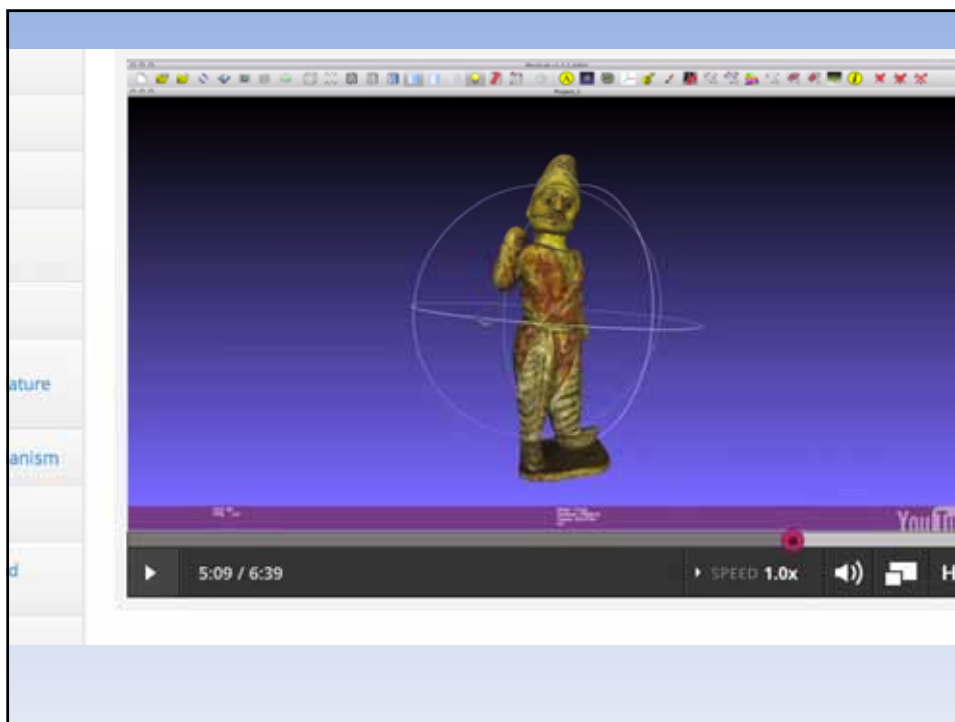
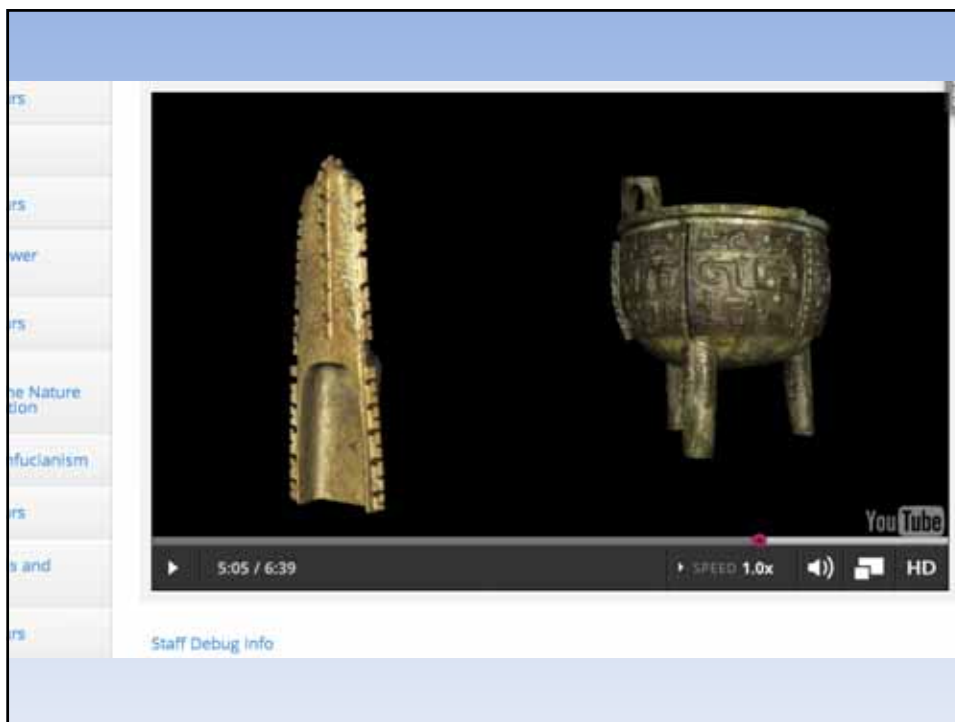
I agree with Kelghtley. From ancient times to nowadays, China is always a highly hierarchical country. 普天之下莫非王土，土之氣莫非王臣。 There is always an emphasis on the group rather than the individual, for example, the long lasting imperial examination required an eight-part essay (八股文), you should articulate for the ancient Chinese sages (代圣人立言), obviously they didn't respect individual viewpoints.

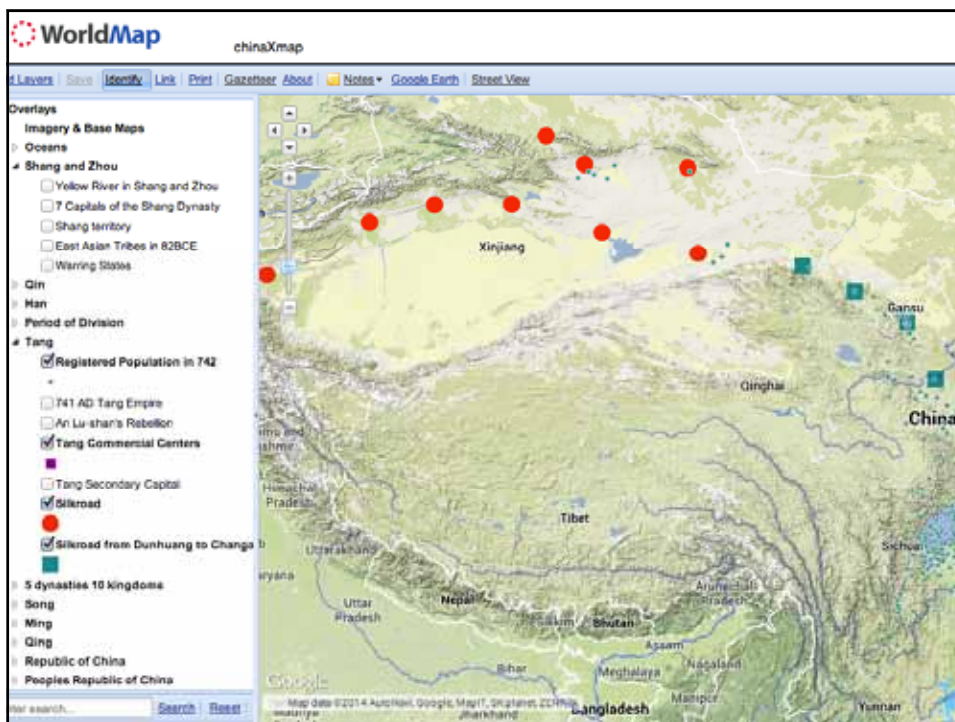
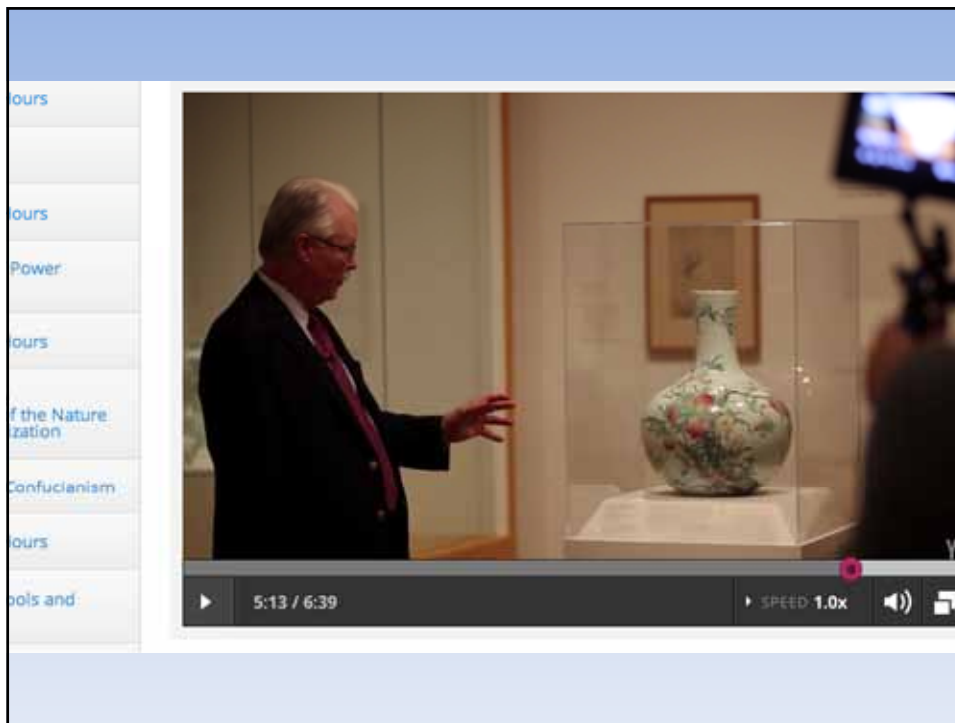
Below the main post, there are several replies:

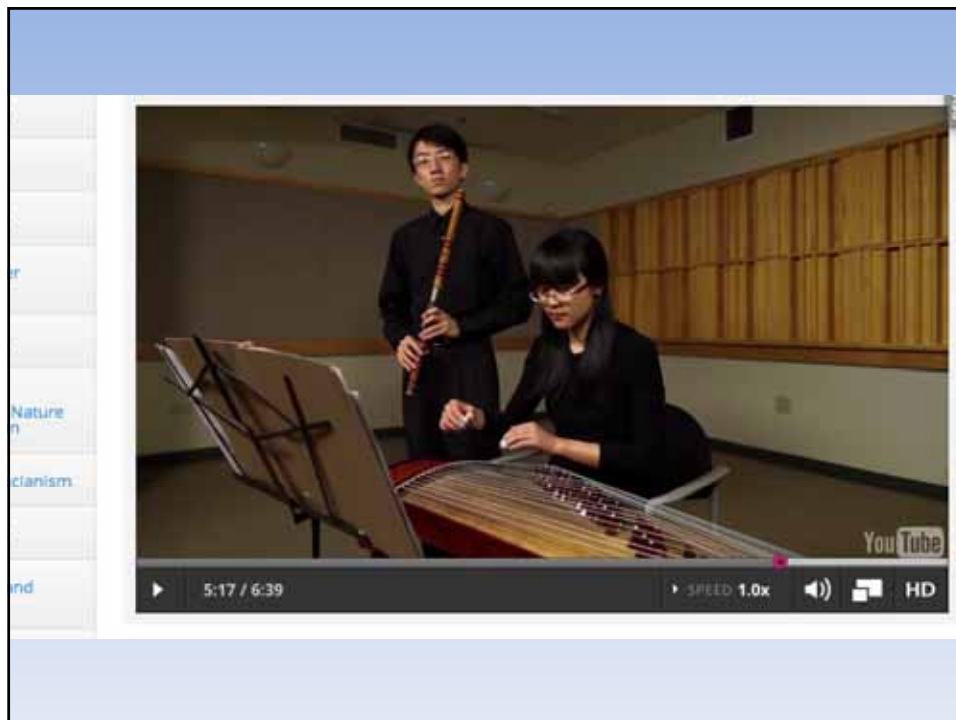
- I agree... it's impossible to decouple the development of China's society from how modern day society in China behaves. This is true even for short periods of time (i.e. how we were raised as children impacts our beliefs and actions as adults). Kelghtley says it in the introduction, "The values of the present, generated by the past, reflect back on that past; fact is seen as value, and value in turn affects what facts are seen." Living in China, I see many of the Shang values reflected today.
- Agree, I am really looking forward to the rest of this course to see how Kelghtley's arguments seem after I have learned more Chinese history.
- There were moments when and locations where this was not about the race, Chinese mindset and love.

**SOME DIFFERENT MODULE
EXAMPLES**

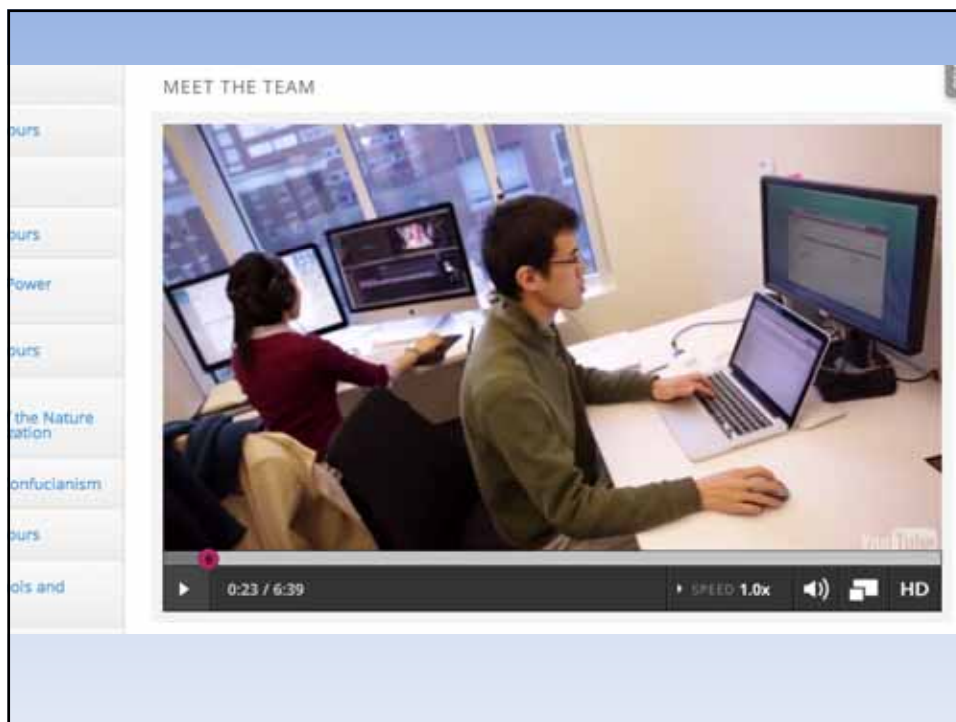
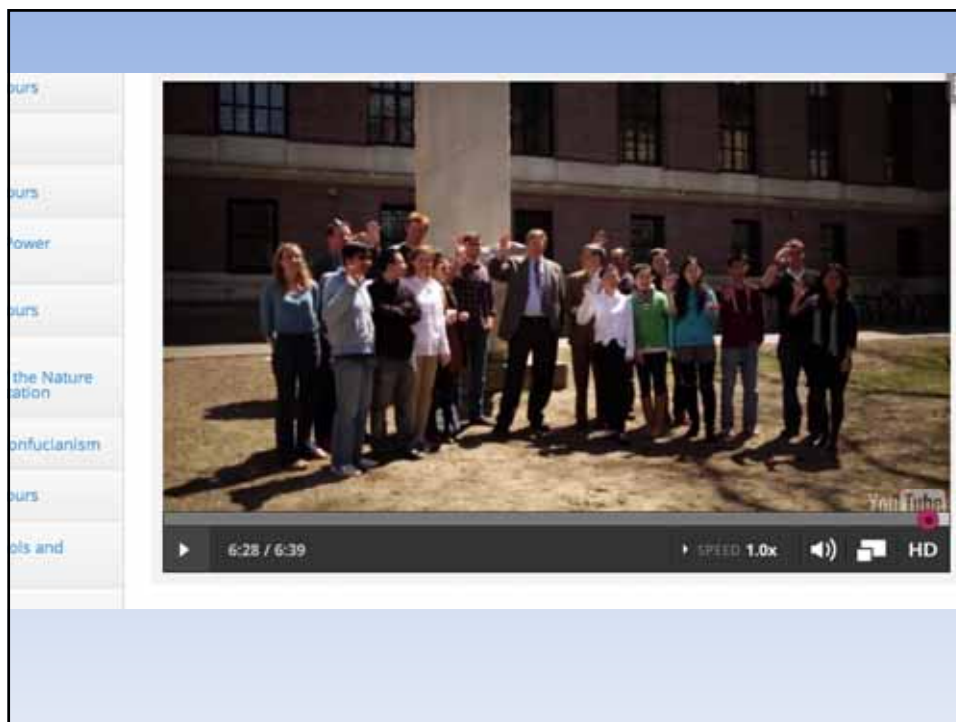


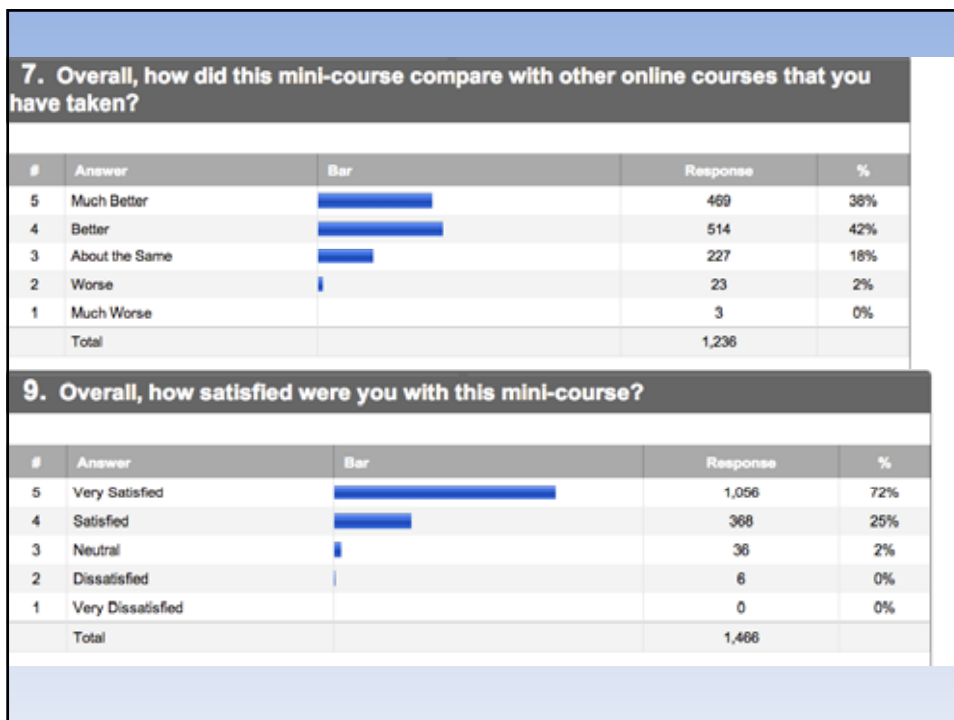












@戴玮_CASIA

@哈佛中国课ChinaX 其实assessment 12那题蛮可惜的。。因为没看到多选是选2个，所以我选了4个- -> 不过这门课真的很棒，谢谢你们。

[收起](#) | [查看大图](#) | [向左转](#) | [向右转](#)

Course Progress for Student 'loveisp'

Assessment	Progress (%)
Assessment 01	100%
Assessment 02	100%
Assessment 03	100%
Assessment 04	100%
Assessment 05	100%
Assessment 06	100%
Assessment 07	100%
Assessment 08	100%
Assessment 09	100%
Assessment 10	100%
Assessment 11	100%
Assessment 12	~85%
Assessment 13	100%
Assessment 14	100%
Assessment 15	100%
Assessment 16	100%
Assessment 17	100%
Assessment 18	100%
Assessment 19	100%
Assessment Avg	100%
Discussion 01	100%
Discussion 02	100%
Discussion 03	100%
Discussion 04	100%
Discussion 05	100%
Discussion Avg	100%
Total	99%

今天 00:34 来自新浪微博

[转发\(1\)](#) | [评论\(1\)](#)

@鹿翠腿

#ChinaX#终于搞定Part 1~~俺的2G流量啊。。😓不过若是没学这门课，可能我对三代历史还是一无所知，而外邦友人却能大谈青铜器对维系商朝统治的重要作用，那就愧对祖宗了。。。而且，以前还真不知道，如果某人没有恻隐之心，就会被孟子打落到“蛮夷or兽”的行列中去👉

🔼 收起 | 🖼 查看大图 | ↶ 向左转 | ↷ 向右转

Course Progress for Student 'palitui'

Item	Progress (%)
Assessment 01	100%
Assessment 02	100%
Assessment 03	100%
Assessment 04	100%
Assessment 05	100%
Assessment 06	100%
Assessment 07	100%
Assessment 08	100%
Assessment 09	100%
Assessment 10	100%
Assessment 11	100%
Assessment 12	100%
Assessment 13	100%
Assessment 14	100%
Assessment 15	100%
Assessment 16	100%
Assessment 17	100%
Assessment 18	100%
Assessment Avg.	100%
Discussion 01	100%
Discussion 02	100%
Discussion 03	100%
Discussion 04	100%
Discussion 05	100%
Discussion 06	100%
Discussion Avg.	100%
Total	99%

weibo.com/palitui

rk!

I really like & find these aspects of the MOOC effective:

- 15-minute length videos
- 15-degree video view of speaker improves my listening
- Progress section is the best...this will keep me in the game until the end.
- This is the piece that has been missing from other MOOCs
- Perhaps this is a tool that will improve completion rates.

Quality of the content is superb. It is wonderful when occasionally a student is asked for their opinion. This helps me change my mind and helps me engage and become more emotionally invested in this class.

Thank you so much.

