





















## Other sources

- 法鼓山 **DDBC Chinese Calendar database**
  - 中西曆日期轉換
- **CBDB**（中國歷代人物傳記資料庫）詞彙
  - 宋代人名 **35632** 筆（原 **40738** 筆，扣除明顯的雜訊後剩下 **35632** 筆；其中有 **9470** 筆出現於《宋會要》）
  - 宋代官名 **5815** 筆（其中有 **3366** 筆出現於《宋會要》）
  - 宋代地名 **4010** 筆（其中有 **2420** 筆出現於《宋會重覆與要》）
- 利用文本採礦 (**text mining**) 從全文擷取出
  - 可能的人名 **11901** 筆（粗估約 **33%** 正確率）

# 系統功能概要 Functions:

- 查找文件 **full text search, faceted search**
  - 全文檢索：查找內容有出現指定字詞的文件
  - 後分類：從查詢結果，利用「年代」、「類別」等來縮小搜尋範圍
- 人物生平資訊（取自 **CBDB**） **biography display**
- 詞頻（人名、地名、官名）分析 **term-list based analysis**
- 文件集在年代的分佈與比較 **temporal distributions**
- 儲存自訂的文件集，方便日後調閱 **saving selections**
- 使用者回饋機制 **user feedback mechanism**
- 前後綴詞工具 **context search**

2012/11/29

# 《宋會要》系統全文檢索 Search frame

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:   [清除](#) | [進階檢索](#)

全文檢索查詢「蝗」

最近幾次檢索

蝗	147
太平興國	1318
興國	1528
興國-太平興國	210

詮釋資料欄位

年代	帝號	主類別	分類	除錯
<input checked="" type="checkbox"/> 年代 (TM)				147
<input checked="" type="checkbox"/> 太宗太平興國				3
<input type="checkbox"/> 太平興國7年 (0982)				2
<input type="checkbox"/> 太平興國1年 (XXXX)				1
<input checked="" type="checkbox"/> 太宗雍熙				1
<input type="checkbox"/> 雍熙2年 (0985)				1
<input checked="" type="checkbox"/> 太宗淳化				1
<input type="checkbox"/> 淳化2年 (0991)				1

呈現模式: [檢索結果](#) | [逐篇檢視](#) | [詞頻與全文](#) | [列印全文](#)

排列方式: [檔名順序](#) | [年代](#)

找到筆數: 147 頁次: [1](#) | [2](#) | [3](#) | [下一頁](#)

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)

1. 帝系九之一三/仁宗/皇祐五年==/十月/九日

十月九日, 詔:「春夏以來, 蝗旱於災, 民間利害有未達者。其令諸路轉運、提點刑獄司徧諭[親]民官採訪以聞。」□

帝系: (皇帝·年代) ◆ ◆ \*\*\*\* ◆ 十月九日 (XXXX) ◆ SongHW\_002149\_p1\_02150

# 後分類 Faceted search (search/analysis within the results)

- 系統除了回傳查找的文件，還會利用「年代」等資訊，將回傳文件進行分類 **see results by period and section topic**
  - 可利用後分類來縮小查詢範圍、觀察檢索結果在這些特徵的分佈狀況、或者發現全文和 **metadata** 的一些明顯錯誤 **see notes on metadata issues**

詮釋資料欄位	
年代   帝號   主類別   分類   除錯	
<input type="checkbox"/> 年代 (TM)	147
<input type="checkbox"/> 太宗太平興國	3
<input type="checkbox"/> 太平興國7年 (0982)	2
<input type="checkbox"/> 太平興國1年 (XXXX)	1
<input type="checkbox"/> 太宗雍熙	1
<input type="checkbox"/> 雍熙2年 (0985)	1
<input type="checkbox"/> 太宗淳化	1
<input type="checkbox"/> 淳化2年 (0991)	1

詮釋資料欄位	
年代   帝號   主類別   分類   除錯	
<input type="checkbox"/> 主類別 (CL)	147
<input type="checkbox"/> 瑞異	56
<input type="checkbox"/> 食貨	42
<input type="checkbox"/> 禮	14
<input type="checkbox"/> 職官	12
<input type="checkbox"/> 刑法	9
<input type="checkbox"/> 兵	3
<input type="checkbox"/> 銓選	3

詮釋資料欄位	
年代   帝號   主類別   分類   除錯	
<input type="checkbox"/> 除錯 (DT)	147
<input type="checkbox"/> -	147
<input type="checkbox"/> EraYMD;OK	36
<input type="checkbox"/> C3Era:EraYMD;OK	33
<input type="checkbox"/> EraYMnD;OK	33
<input type="checkbox"/> C3Era:EraYMnD;OK	15
<input type="checkbox"/> EraYnMnD;OK	15
<input type="checkbox"/> DatePatNotMatch	6

# 地理分布 geographical distribution

2012/11/29

婺州 ?vs .all 208  
 婺州 ?vs 處州 208  
 婺州 -贈婺州 197  
 婺州 208

搜索結果 | 逐篇檢視 | 詞頻與全文 | 詞頻全文 | 檔名順序 | 年代

找到筆數: 208      頁次: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | >> | 下一頁  
 相關人名: 江山 · 周綱 · 周麟之 · 秦昌時 · 陳正同 · 曾經   
 相關地名: 婺州 · 衢州 · 東陽 · 台州 · 湖州 · 處州 · 明州 · 蘭溪

依 df 降冪排列 | 依 t-q 降冪排列

人名	Term t	(tf)	df	t-q
江山	生平	(5)	5	0.294
秦昌時	生平	(4)	4	1.000
劉子翼	生平	(4)	4	0.571
陳正同	生平	(8)	4	0.400
周綱	生平	(4)	4	0.333
周麟之	生平	(4)	4	0.174
曾經	生平	(4)	4	0.007
趙師昂	生平	(3)	3	0.750
毛友	生平	(6)	3	0.600
季南壽	生平	(3)	3	0.600

更多詞彙...

依 df 降冪排列 | 依 t-q 降冪排列

地名	Term t	(tf)	df	t-q
婺州		(227)	207	1.000
衢州		(34)	31	0.171
東陽		(38)	30	0.638
台州		(30)	27	0.130
湖州		(36)	27	0.092
處州		(26)	26	0.088
蘭溪		(32)	25	0.926
明州		(27)	25	0.068
常州		(24)	23	0.094
溫州		(32)	23	0.073

更多詞彙...

依 df 降冪排列 | 依 t-q 降冪排列

官名	Term t	(tf)	df	t-q
戶部		(39)	30	0.008
通判		(47)	27	0.016
臨安府		(31)	16	0.011
西華公		(16)	12	0.005

搜索結果分佈圖 (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)

地名文件數量分佈圖 (此圖僅計入含有地名資訊的文件)

x 尋找: 宋會要      下一筆 (N)      上一筆 (P)      全部強調標示 (A)      符合大小寫 (C)

# 人物生平資訊 calling up biographical data in CBDB via CBDB API

2012/11/29

宋會要輯稿

文獻集: [宋會要輯稿](#)

測試版

最近幾次檢索

- 呂夷簡 124
- 朱熹 22
- 呂祖謙 8
- 王安石 230

呈現模式: [檢索結果](#) | [逐篇檢視](#) | [詞類與全文](#) | [列印全文](#)

排列方式: [檔名順序](#) | [年代](#)

找到筆數: 124 頁次: [1](#) | [2](#) | [3](#) | [4](#) | [5](#) | [6](#) | [7](#) | [下一頁](#)

相關人名: [呂夷簡](#)<sub>生平</sub> · [王曾](#)<sub>生平</sub> · [張士遜](#)<sub>生平</sub> · [曾公亮](#)<sub>生平</sub> · [魯宗道](#)<sub>生平</sub> · [王旦](#)<sub>生平</sub>

呂夷簡 [close](#)

CBDB: 中國歷代人物傳記資料庫 [連外網路連結](#)

呂夷簡:8100

CBDB ID: 8100

索引中文/英文名稱: 呂夷簡/呂夷簡/Lv Yijian

指數年 (index year): 1038

生年: 北宋太平興國4年 (979)

卒年: 北宋慶曆4年 (1044)

享年: 66

郡望: 東平

**出處:** 宋人傳記資料索引(電子版), 頁6615

**註:** Lu(9) Yijian [8100] When Zhao(1) Renzong [9004] (qv.) took personal control over the government, Yijian and Renzong planned to dismiss from positions of power partisans of Yang(2) Huangtaihou

依 df 降序排列 | 依 t→q 降序排列

Term t	人名	(tf)	df	t→q
<a href="#">呂夷簡</a>	<a href="#">生平</a>	(130)	122	1.000
<a href="#">王曾</a>	<a href="#">生平</a>	(20)	18	0.167
<a href="#">張士遜</a>	<a href="#">生平</a>	(17)	17	0.213
<a href="#">魯宗道</a>	<a href="#">生平</a>	(13)	11	0.324
<a href="#">曾公亮</a>	<a href="#">生平</a>	(12)	11	0.095
<a href="#">曹瑋</a>	<a href="#">生平</a>	(12)	10	0.204
<a href="#">韓琦</a>	<a href="#">生平</a>	(10)	10	0.047
<a href="#">王旦</a>	<a href="#">生平</a>	(10)	10	0.032
<a href="#">晏殊</a>	<a href="#">生平</a>	(9)	9	0.087
<a href="#">富弼</a>	<a href="#">生平</a>	(9)	9	0.083

[更多詞彙...](#)

系統若發現查詢的文件中有人物生平的資訊，會顯示一系列「相關人名」，在人名旁有個「生平」的小框，點選後可顯示該人名在 CBDB 的資訊

## 詞頻分析 term-list based frequency analysis -by person

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:   [清除](#) [進階檢索](#)

測試版

最近幾次檢索

- 呂夷簡 124
- 朱熹 22
- 呂祖謙 8
- 王安石 230

呈現模式: [檢索結果](#) | [逐篇檢視](#) | [詞頻與全文](#) | [列印全文](#) 排列方式: [檔名順序](#) | [年代](#)

找到筆數: 124 頁次: [1](#) | [2](#) | [3](#) | [4](#) | [5](#) | [6](#) | [7](#) | [下一頁](#)

相關人名: [呂夷簡](#) [王曾](#) [張士遜](#) [曾公亮](#) [魯宗道](#) [王旦](#)

相關地名: [崇德](#) [彰武軍節度](#) [太平](#) [壽州](#) [永安](#) [杭州](#) [洪州](#) [鎮江軍節度](#)

依 df 降序排列 | 依 t→q 降序排列

Term t	人名	(tf)	df	t→q
呂夷簡	生平	(130)	122	1.000
王曾	生平	(20)	18	0.167
張士遜	生平	(17)	17	0.213
魯宗道	生平	(13)	11	0.324
曾公亮	生平	(12)	11	0.095
曹瑋	生平	(12)	10	0.204
韓琦	生平	(10)	10	0.047
王旦	生平	(10)	10	0.032
晏殊	生平	(9)	9	0.087
富弼	生平	(9)	9	0.083

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)

95

一次選取(或取消)全部勾選

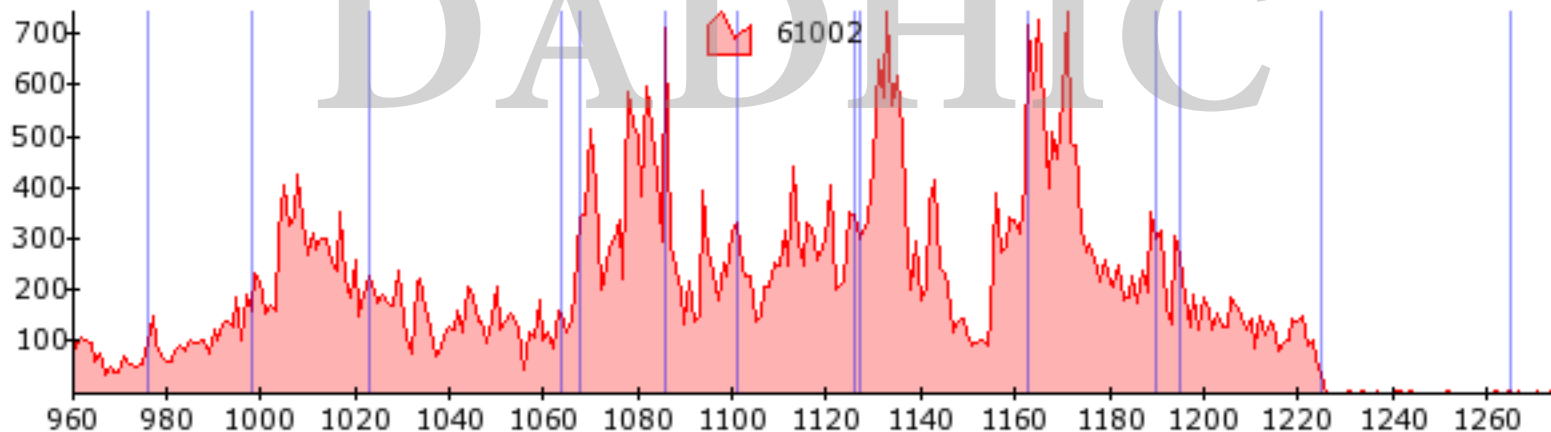
系統在顯示全文時, 將以不同的顏色標示人名、地名、官名、與其他詞彙

查詢「呂夷簡」



## 文件集在年代上的分佈 temporal distribution of all records

- 若取樣具代表性，文件在時間上的分佈通常是有意義的
- 在《宋會要》的 **80396** 筆條目中，目前有 **61002** 筆可轉譯出西元日期



# 兩個查詢結果在年代分佈上的比較

## comparing the temporal distribution of different terms ?vs

宋會要輯稿

文件檢索: 溫州 ?vs 饒州

找到筆數: 313

頁次: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 下一頁

136 213

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)

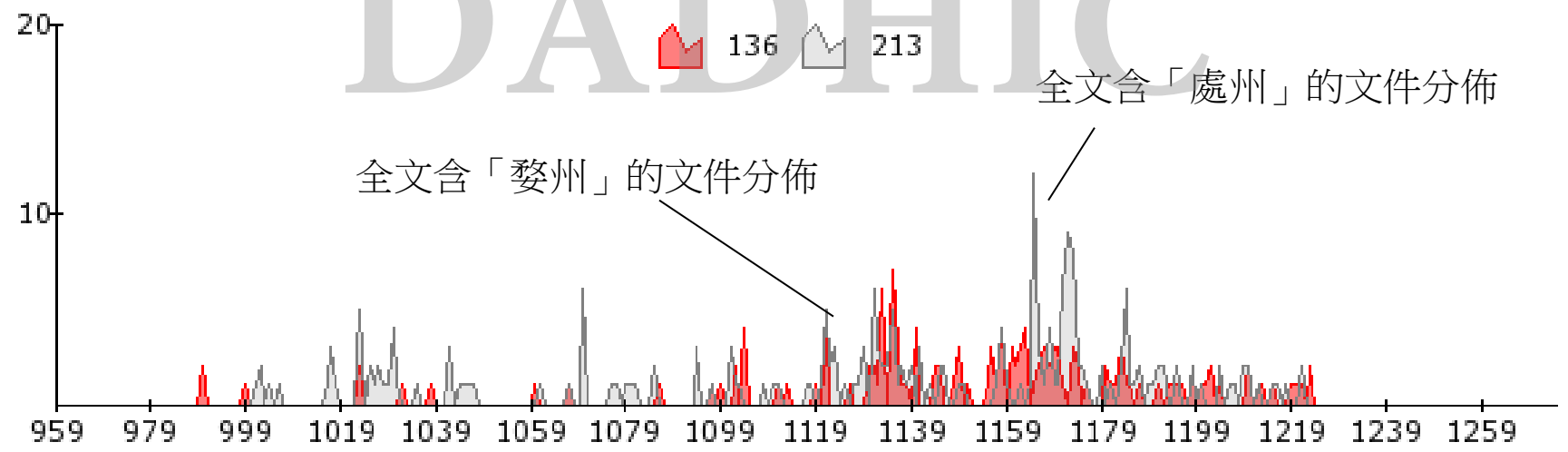
1. 帝系二之五/雜錄 孝宗/淳熙十一年 == 十一年十一月二十四日

十一月二十四日, 大宗正司言: 「奉詔於瀛王下懷王位選擇「不」字宗室一員比換南班官, 日赴朝參及諸般行事, 填不諱身故關。本司今選擇到修武郎、新添差東南第五副將、饒州駐劄不預, 持身廉謹, 無疾病過犯, 見年五十歲, 欲填不諱關。其請給、人從並乞依不諱等已得指覆。、詔依, 與撫寧府率。 □

• 語法：q1 ?vs q2

- 其中 q1, q2 為一般的查詢字串
- 在「年代-文件數量」分佈圖中，將以紅色線顯示q1 的查詢結果，而以灰色線顯示q2 的查詢結果

• 例如：婺州 ?vs 處州



## 進階工具：前後綴詞的分析 advanced tools: discovering terms proximate to a search term

- 有時我們會想知道：給定一個文件集合，某個特定詞彙的前後會出現那些綴詞
  - 「右丞相**XXX**」的出現情形
  - 「**X**國公」的出現頻率
  - 「**X**稅」的出現統計
- 「進階檢索」下方，有提供一個簡化工具
  - 使用者僅需輸入一個關鍵詞，並選擇該關鍵詞左方（右方）欲顯示的綴詞長度

進行綴詞分析： 關鍵詞  左綴字數  右綴字數

# 「右丞相XXX」的出現情形 example: office title.(3) gives all occurrences of that title with the 3 characters following it

(進階檢索的綴詞工具)

進行綴詞分析: 關鍵詞  左綴字數  右綴字數

## 宋會要輯稿

文獻集:

[回首頁](#) | [功能說明](#) | [修改密碼](#) | [登出](#)

文件檢索:

[清除](#) | [進階檢索](#)

### 最近幾次檢索

- ?termpat:右丞相.{3} .all +右丞相 150
- ?termpat:賜.{1,8}使 .all +賜 +使 2433
- ?termpat:.[1]稅 .all +稅 4329
- 稅 4329

### 詮釋資料欄位

年代 | 帝號 | 主類別 | 分類 | 類目 | 除錯

- 年代 (TM) 150
- 孝宗隆興 1
  - 隆興2年 (1164) 1
- 孝宗乾道 6
  - 乾道7年 (1171) 1
  - 乾道8年 (1172) 1
  - 乾道9年 (1173) 4
- 孝宗淳熙 29
  - 淳熙1年 (1174) 5
  - 淳熙2年 (1175) 4

呈現模式

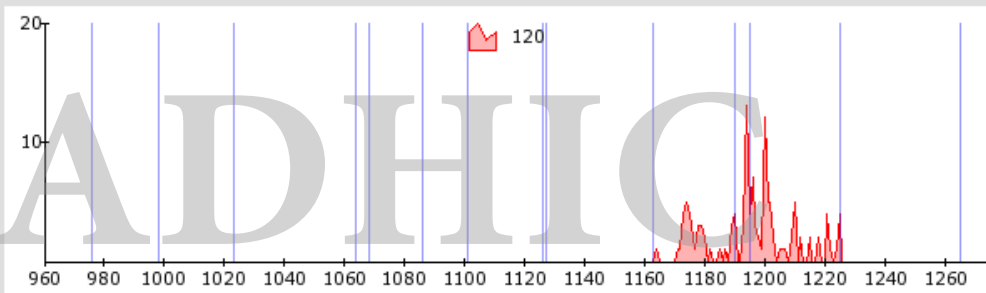
[檢索結果](#) | [逐篇檢視](#) | [詞頻與全文](#) | [列印全文](#)

排列方式

[檔名順序](#) | [年代](#)

找到筆數: 150

頁次: [1](#) | [2](#) | [3](#) | [下一頁](#)



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)

term	df	tf
右丞相謝深甫	14	15
右丞相趙汝愚	11	14
右丞相史彌遠	8	9
右丞相陳自強	7	7
右丞相兼樞密	5	5
右丞相京鏜等	3	3
右丞相京鏜攝	3	3
右丞相梁克家	3	3

term	df	tf
右丞相趙雄等	3	3
右丞相趙雄為	2	2
右丞相葛邲提	2	2
右丞相一員掌	2	2
右丞相葉衡等	2	2
右丞相周必大	2	2
右丞相京鏜為	1	1
右丞相葉衡率	1	1

term	df	tf
右丞相葛邲特	1	1
右丞相陳自強	1	2
右丞相張九齡	1	1
右丞相曾懷朝	1	1
右丞相臣京鏜	1	1
右丞相余端禮	1	1
右丞相臣留正	1	1
右丞相史浩後	1	1

term	df	tf
右丞相留正率	1	1
右丞相留正提	1	1
右丞相曾懷提	1	1
右丞相史浩以	1	1
右丞相曾懷除	1	1

# 「X國公」的出現頻率

2012/11/29

(進階檢索的綴詞工具) 進行綴詞分析: 關鍵詞  左綴字數  右綴字數

term	df	tf
封國公	109	110
魯國公	81	85
魏國公	75	84
開國公	46	52
楚國公	34	36
榮國公	30	32
秦國公	24	25
衛國公	23	23
韓國公	22	23
祁國公	20	21
建國公	20	20
英國公	19	22
永國公	18	20
崇國公	17	17
燕國公	17	17
和國公	17	17
許國公	16	16
成國公	16	17
申國公	16	17
徐國公	15	15
惠國公	15	15
嘉國公	15	15
濮國公	15	16
澗國公	14	15

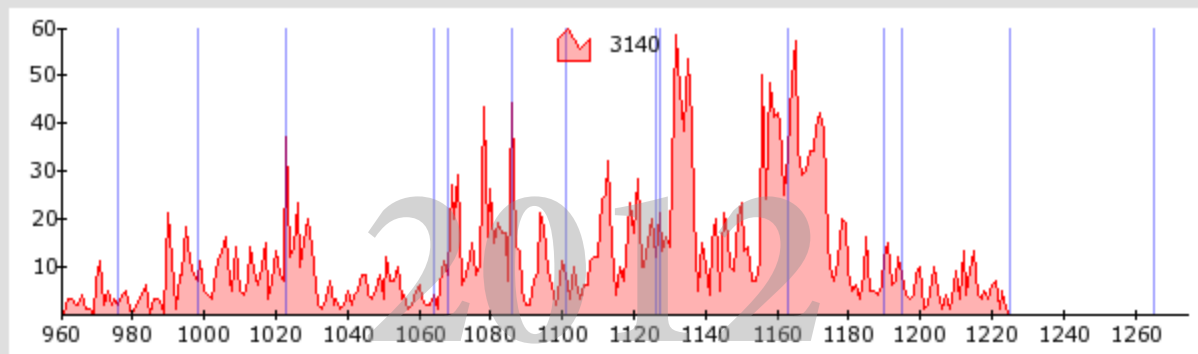
term	df	tf
遂國公	9	10
冀國公	9	10
邢國公	8	8
安國公	8	10
涇國公	8	9
鄧國公	8	10
康國公	8	8
華國公	7	8
岐國公	7	7
溫國公	6	6
房國公	6	6
蔡國公	6	7
郇國公	6	8
晉國公	6	6
鄭國公	6	6
荊國公	6	6
雍國公	6	6
鄆國公	5	5
恭國公	5	5
虢國公	5	5
齊國公	5	5
儀國公	5	5
壽國公	5	5
定國公	5	5

term	df	tf
鄒國公	4	4
景國公	4	4
紀國公	4	6
吉國公	4	4
卽國公	4	5
濟國公	4	4
趙國公	4	4
宿國公	3	3
洋國公	3	3
密國公	3	5
梁國公	3	3
宣國公	3	3
二國公	3	3
江國公	3	5
唐國公	3	3
尹國公	3	3
商國公	3	4
漢國公	3	3
子國公	2	2
隨國公	2	2
邠國公	2	3
莒國公	2	2
并國公	2	2
廣國公	2	2

term	df	tf
三國公	1	1
揚國公	1	1
中國公	1	1
令國公	1	1
號國公	1	1
姪國公	1	1
助國公	1	2
萊國公	1	1
郟國公	1	1
夏國公	1	2
豐國公	1	1
祈國公	1	1
珍國公	1	1
襲國公	1	1
隋國公	1	1
瀛國公	1	1
褒國公	1	1
昭國公	1	1
循國公	1	1
譙國公	1	1
葵國公	1	1
原國公	1	1
歧國公	1	1
忻國公	1	1

# 「X稅」的出現統計

2012/11/29



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)

term	df	tf
商稅	1049	1571
二稅	338	446
納稅	287	326
收稅	286	401
夏稅	249	303
租稅	235	287
酒稅	201	230
賦稅	182	190
秋稅	152	170
免稅	122	126
放稅	120	141
苗稅	109	169
戶稅	83	88
催稅	77	96

term	df	tf
零稅	25	25
立稅	24	31
省稅	23	27
置稅	23	24
割稅	23	27
漏稅	23	23
於稅	23	23
屋稅	23	29
舊稅	23	23
年稅	22	22
軍稅	22	23
等稅	22	24
征稅	22	24
匿稅	21	21

term	df	tf
色稅	10	11
米稅	10	12
務稅	10	10
失稅	10	10
出稅	10	12
取稅	10	10
雜稅	9	16
欄稅	9	12
木稅	9	9
住稅	9	9
元稅	9	9
合稅	9	11
依稅	9	10
由稅	9	11

term	df	tf
業稅	5	5
罷稅	5	5
得稅	5	5
府稅	5	5
故稅	5	5
委稅	5	5
上稅	5	7
京稅	5	5
受稅	5	5
回稅	5	5
鄉稅	5	5
非稅	5	5
人稅	5	5
江稅	5	5

可直接檢視綴詞附近的文件內容 the text in which the string appears can be called up

term	df	tf	term	df	tf	term	df	tf	term	df	tf
商稅	1049	1571	零稅	25	25	色稅	10	11	業稅	5	5
二稅	338	446	立稅	24	31	米稅	10	12	罷稅	5	5
納稅	287	326	省稅	23	27	務稅	10	10	得稅	5	5
收稅	286	401	置稅	23	24	失稅	10	10	府稅	5	5
夏稅	249	303	割稅	23	27	出稅	10	12	故稅	5	5
租稅	235	287	漏稅	23	23	取稅	10	10	委稅	5	5
酒稅	201	230	於稅	23	23	雜稅	9	16	上稅	5	7

.all +稅 +酒稅 close

找到筆數: 201      頁次: 1 | 2 | 3 | 4 | 5 | [下一頁](#)

**1. 食貨三六之四/太宗/至道三年 == 三年/九月/**

三年九月，詔西川峽路州軍：「自今應收酒稅、鹽諸般課利，並據合納課額，只令送納見錢，不得更折金、銀、匹帛。如官中關用，即轉運司於合收買州軍，依本處見賣時價置場收買，仍取情願，不得抑勒及虧價錢。」時川陝寇... ¶...

食貨：(皇帝·年代) ← ♦ 02.太宗 ♦ 至道三年九月(至道3年) ♦ SongHW\_049186\_p3\_13531

**2. 食貨一七之一四/真宗/咸平五年 == 五年/二月/二十三日**

二十三日，除合州赤水、巴川二縣、長利、謝市、樓灘三鎮酒稅錢，以經寇殘弊也。... ¶...

食貨：(皇帝·年代) ← ♦ 03.真宗 ♦ 咸平五年二月二十三日(咸平5年) ♦ SongHW\_046649\_p3\_10994

**3. 食貨二〇之五/真宗/景德四年 == 四年/四月/**

四月，詔曰：「權酷之法，素有定規，宜令計司，立為永式。自今中外不得更議增課，以圖恩獎。」先是，遣使檢舉天下酒稅，頗為煩擾，且競以義利邀寵，遂罷其事而條禁約之。... ¶...

食貨：(皇帝·年代) ← ♦ 03.真宗 ♦ 景德四年四月(景德4年) ♦ SongHW\_047250\_p3\_11595

# Conclusion

- In this digital age, the role of historical databases might change from providing “dictionary lookup service” to providing “discovery services”

DADHIC

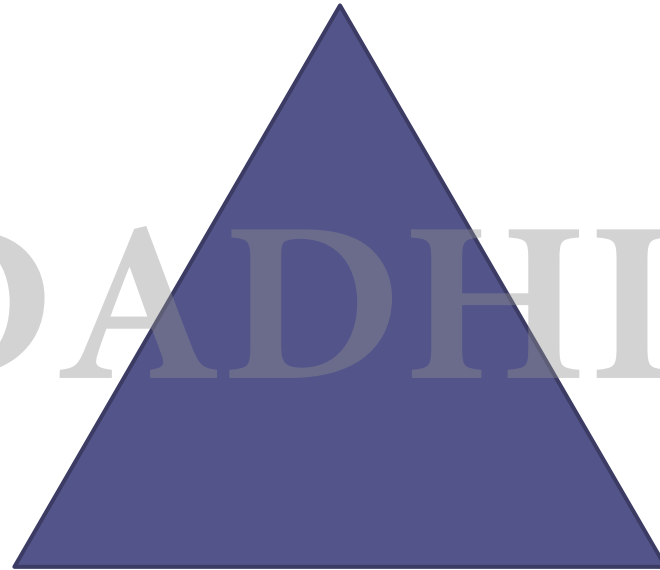


How?

2012 **Humanist**

DADHIC

**Digital  
Content**



**Computer  
Science**

## How to get access

- [http://thdl.ntu.edu.tw/L303\\_SongHuiYao/](http://thdl.ntu.edu.tw/L303_SongHuiYao/)
- You need to be a subscriber for the Journal of Song Yuan Studies
- Then apply for an account from CBDB. CBDB manages the access to the database.