

A guide to verifying machine- extracted person names from the Song Hui Yao

March 2012

By Chen Shih-Pei (陳詩沛)

CBDB Project Manager

shihpeichen@fas.harvard.edu

Background of this task

- Taida (台大, National Taiwan University) applied text mining/natural language processing techniques to the full texts of Song Hui Yao (宋會要輯稿), and extracted 11,901 possible person names. Those names need to be verified.
- The estimated precision rate is 33%, meaning that there may be 4,000 new names that can be added to CBDB after verification.

Splitting up the task

- **25** Journal of Song Yuan Studies (JSYS) scholars registered to help. Taida also has also created a research group.
- To coordinate the work of Taida and JSYS scholars we divided the 80,396 Song Hui Yao documents into **81 parts**, each containing 1,000 documents.
- **Parts #56 – #80** are assigned to the Song Yuan scholars. We expect each of you to **pick one part** and to verify the computer-extracted person names appearing in the part.

To begin with

- The Song Hui Yao system is at:
http://thdl.ntu.edu.tw/L303_SongHuiYao/RetrieveDocs.php
- You will soon receive email containing the *password* for accessing the system. Your *email address* will be your username for login.
- To see all the 81 parts, you may type “.all” into the search box after logging in to the system. (See the following slide.)

1. Type **.all** (dot + all) into the search box to retrieve all the documents in Song Hui Yao

2. Click on 部份
(parts) to see the 81
parts

thdl.ntu.edu.tw/L303_SongHuiYao/RetrieveDocs.php

Google

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: .all

送出 清除 進階檢索

測試版

最近幾次檢索

Click on 部份 (parts) to see the 81 parts

選擇資料欄位

年代 | 帝號 | 主類別 | 部份 | 檢錯

部份 (DC) 80396

Part_001 1000

Part_002 1000

Part_003 1000

Part_004 1000

Part_005 1000

Part_006 1000

Part_007 1000

Part_008 1000

Part_009 1000

Part_010 1000

Part_011 1000

Part_012 1000

Part_013 1000

Part_014 1000

Part_015 1000

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: 80396 頁次: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | >> | 下一頁

相關人名: 曾經 王旦 蔡京 王安石 司馬光 韓琦

相關地名: 四川 河東 太平 廣南 福建路 江州 兩浙路 建康府

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. 帝系一之一/帝系 帝系一 /帝號 == //

【宋會要】僖祖立道肇基積德起功懿文憲武睿和至孝皇帝，諱眺，漢京兆尹廣漢之後。生於燕薊，仕唐，歷永清、文安、幽都三縣令。建隆元年三月，追尊曰文獻皇帝，廟號僖祖。謚議，翰林學士竇儼撰；冊文，中書舍人扈蒙撰。陵曰欽陵。在幽州。大中祥符五年閏十月，加謚曰文獻睿和。冊文，樞密使、同中書門下平章事王欽若撰。以...

帝系: 帝號 ◆ **** ◆ ◇ SongHW_000001_p1_00001

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。建隆元年三月，追尊曰惠元皇帝，廟號順祖。謚議，翰林學士竇儼撰；冊文，中書舍人趙遂擇撰。陵曰康陵。在幽州。大中祥符五年閏十月加今謚。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。熙寧五年十一月二十五日，以奉僖祖

Please pick one part to work on:

Part_056 食貨	Part_061 刑法	Part_066 兵, 方域	Part_071 蕃夷	Part_076 班序上
Part_057 食貨	Part_062 刑法, 兵	Part_067 方域	Part_072 蕃夷, 道釋	Part_077 班序上, 商稅, 銓選
Part_058 食貨	Part_063 兵	Part_068 方域	Part_073 道釋, 補編	Part_078 銓選
Part_059 食貨, 刑法	Part_064 兵	Part_069 方域	Part_074 補編, 班序上	Part_079 銓選, 食貨
Part_060 刑法	Part_065 兵	Part_070 方域, 蕃夷	Part_075 班序上	Part_080 食貨

Please **register** your pick in this **Google spreadsheet** to avoid conflict:
[https://docs.google.com/spreadsheet/cc?
key=0Avbe4kTWw59TdDJuYllUTkozNW44Qk1zYURVNHQ3RFE#gid=0](https://docs.google.com/spreadsheet/cc?key=0Avbe4kTWw59TdDJuYllUTkozNW44Qk1zYURVNHQ3RFE#gid=0)

The Verifying Process (1)

- To begin the verification task, type “{DC:Part_0XX}” into the search box (replace XX with the number of a part) and click the 送出 button.

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： {DC.Part_001} [清除](#) [進階檢索](#)

最近幾次檢索

{DC.Part_007}.all	1000
.all	80396
{DC.Part_001}	1000
{DC.Part_055}	1000

選擇資料欄位

年代	帝號	主類別	部份	檢錯
☒ 年代 (TM)				1000
☐ 太祖建隆				2
☐ 建隆1年 (0960)				2
☐ 太祖開寶				5
☐ 開寶4年 (0971)				1
☐ 開寶6年 (0973)				1
☐ 開寶8年 (0975)				1
☐ 開寶9年 (0976)				2
☐ 太宗太平興國				2
☐ 太平興國7年 (0982)				1
☐ 太平興國8年 (0983)				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: 1000 頁次: [1](#) | [2](#) | [3](#) | [4](#) | [5](#) | [6](#) | [7](#) | [8](#) | [9](#) | [10](#) | >> | [下一頁](#)

相關人名: [王允](#)_(生平) · [王元](#)_(生平) · [武勝](#)_(生平) · [王宗](#)_(生平) · [王指](#)_(生平) · [馮翊](#)_(生平)

相關地名: [武寧](#) · [高密郡](#) · [長子](#) · [安化](#) · [保信軍節度](#) · [封國](#) · [徐州](#) · [保塞](#)

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. 帝系一之一/帝系 帝系一/帝號 == //

【宋會要】僖祖立道擊基積德起功懿文憲武睿和至孝皇帝，諱眺，漢京兆尹廣漢之後。生於燕薊，仕唐，歷永清、文安、幽都三縣令。建隆元年三月，追尊曰文獻皇帝，廟號僖祖。謚議，翰林學士竇儼撰；冊文，中書舍人扈蒙撰。陵曰欽陵。在幽州。大中祥符五年閏十月，加謚曰文獻睿和。冊文，樞密使、同中書門下平章事王欽若撰。以…

帝系:帝號 ◆◆◆◆◆ () ◆ SongHW_000001_p1_00001

The Verifying Process (2)

- The default 呈現模式 is 檢索結果. Switch to 詞頻與全文 (term-list based analysis) by clicking on it.

宋會要輯稿

文獻集: 宋會要輯稿

Click

文件檢索: {DC:Part_001}

送出 清除 進階檢索

測試版

最近幾次檢索

檢索條件	文件數量
{DC:Part_001}	1000
.all	80396
{DC:Part_055}	1000
王仲湜	16

選擇資料欄位

欄位	數量
年代 (TM)	1000
太祖建隆	2
建隆1年 (0960)	2
太祖開寶	5
開寶4年 (0971)	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: 1000 頁次: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | >> | 下一頁

相關人名: 王允平, 王元平, 武勝平, 王宗平, 王楷平, 馮翊平

相關地名: 武寧, 高密縣, 長子, 安化, 保信軍節度, 封國, 徐州, 保寧

261

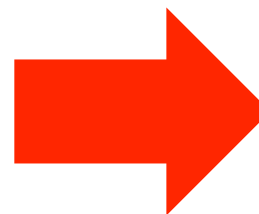
檢索結果分佈圖 (此圖僅計入含有年代資訊的文件; 橫軸: 西元年, 縱軸: 文件數量)

The terms in the former 3 lists (人名, 地名, and 官名) come from CBDB.

9

The 4th list, 其他, contains the strings which are identified by computer procedures as person names. *This is the list we need you to verify.*

依 df 降冪排列			
其他			
Term t	df	t→q	回報
祁王模	3	1.000	■■■
華王侯	3	0.750	■■■
榮易	3	0.750	■■■
康孝	3	0.500	■■■
康簡	3	0.500	■■■
魏惠憲	3	0.429	■■■
侯有儀	2	1.000	■■■
侯有鄰	2	1.000	■■■
侯孝哲	2	1.000	■■■
梁襄	2	1.000	■■■
更多詞彙...			



Click on the arrow to expand the list

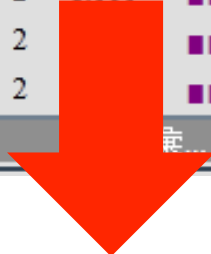
依 df 降冪排列			
其他			
Term t	df	t→q	回報
祁王模	3	1.000	■■■
華王侯	3	0.750	■■■
榮易	3	0.750	■■■
康孝	3	0.500	■■■
康簡	3	0.500	■■■
魏惠憲	3	0.429	■■■
侯有儀	2	1.000	■■■
侯有鄰	2	1.000	■■■
侯孝哲	2	1.000	■■■
梁襄	2	1.000	■■■
更多詞彙...			
唐王俊	2	0.667	■■■
侯允熙	2	0.667	■■■
侯德潤	2	0.667	■■■
華王植	2	0.667	■■■
楊王昉	2	0.667	■■■
張卓	2	0.500	■■■
王材	2	0.400	■■■
武容和	2	0.400	■■■
張掖	2	0.286	■■■
宰輔表	2	0.059	■■■
20			
周顯德	2	0.057	■■■
彭信	1	1.000	■■■
王臨川	1	1.000	■■■
王舊無	1	1.000	■■■
仲江	1	1.000	■■■
仲沃	1	1.000	■■■
仲汲	1	1.000	■■■
仲防	1	1.000	■■■
仲杵	1	1.000	■■■
仲旆	1	1.000	■■■
仲決	1	1.000	■■■
仲啟	1	1.000	■■■
仲處	1	1.000	■■■
仲偃	1	1.000	■■■
仲登	1	1.000	■■■
仲銑	1	1.000	■■■
仲墳	1	1.000	■■■
仲懣	1	1.000	■■■
仲盤	1	1.000	■■■
仲翮	1	1.000	■■■
40			
仲嘩嘩	1	1.000	■■■
仲禦	1	1.000	■■■

The Verifying Process (3)

For each string, click on the **2nd column** (the one with digital numbers). A yellow window will pop-up. You will see **the text in which the string appears**. (The term will be highlighted.)

依 df 降冪排列

Term t	df	t→q	回報
祁王模	3	1.000	■
華王侯	3	0.750	■ ■ ■
榮易	3	0.750	■ ■ ■
康孝	3	0.500	■ ■ ■
康簡	3	0.500	■ ■ ■
魏惠憲	3	0.429	■ ■ ■
侯有儀	2	1.000	■ ■ ■
侯有鄰	2	1.000	■ ■ ■
侯孝哲	2		■ ■ ■
梁襄	2		■ ■ ■



Click on the number to show the context of the string

依 df 降冪排列

Term t	df	t→q
祁王模	3	1.000
華王侯	3	0.750
榮易	3	0.750
康孝	3	0.500
康簡	3	0.500
魏惠憲	3	0.429
侯有儀	2	1.000
侯有鄰	2	1.000
侯孝哲	2	1.000
梁襄	2	1.000

更多詞彙

唐王俊	2	0.667
侯允熙	2	0.667
侯德潤	2	0.667

+!specific:祁王模 +{DC:Part_001}

顯示: [後分類](#) | [詞頻分析](#) | [文件部分內容](#)

找到筆數: 3 頁次: 1

1. 帝系二之二二/徽宗/宣和四年 == 四年/五月/六日

五月六日, 詔:「皇子**祁王模**已經裹頭出閣, 本閣官吏、諸色祇應人並依昨**文安郡王**裹頭例各與轉一官。」

帝系:(皇帝·年代) ← + 08.徽宗 + 宣和四年五月六日 (宣和4年) + SongHW_000359_p1_00359

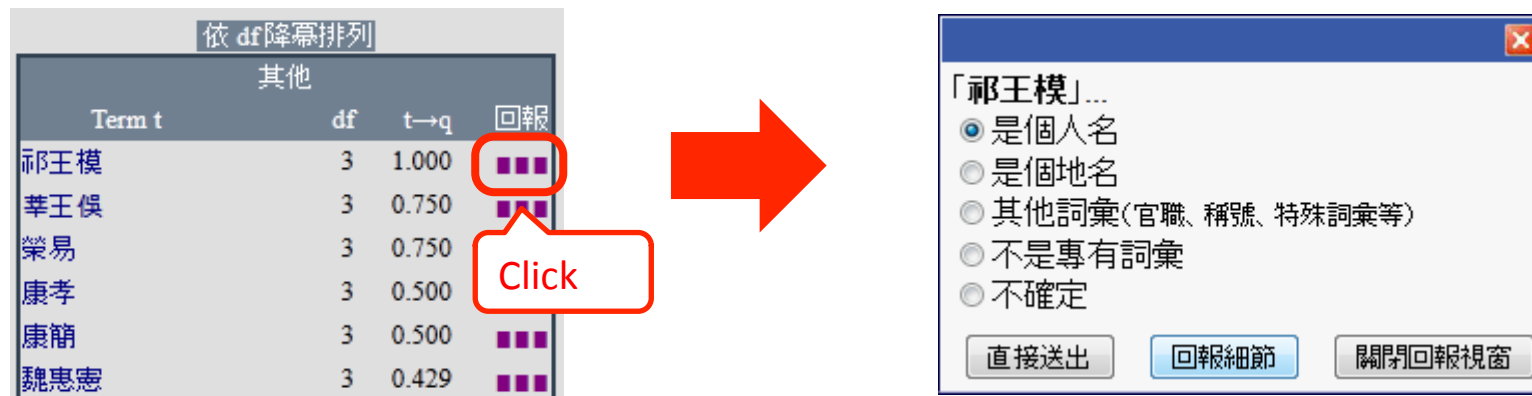
原文參照:(尚未建立相關資料)

2. 帝系一之四二/皇子諸王/**祁王模** == //

祁王模, 徽宗子。大觀元年十二月生。二年三月賜名, 授**武勝軍節度使**、檢校太尉, 封**鎮國公**。政和三年正月, 正官名, 改授檢校太保。宣和二年十一月, 改淮南節度使, 加開府儀同三司, 進封樂

The Verifying Process (4)

- To give your feedback,
 1. Close the yellow window to return to the original view
 2. On the term list, click on the *3rd column* (the one with 3 solid blocks ■■■). The *feedback window* will pop-up to gather your feedback.



依 df 降序排列

Term t	df	t→q	回報
祁王模	3	1.000	■■■
莘王侯	3	0.750	■■■
榮易	3	0.750	■■■
康孝	3	0.500	■■■
康簡	3	0.500	■■■
魏惠憲	3	0.429	■■■

Click

「祁王模」...

☒ 是個人名

☐ 是個地名

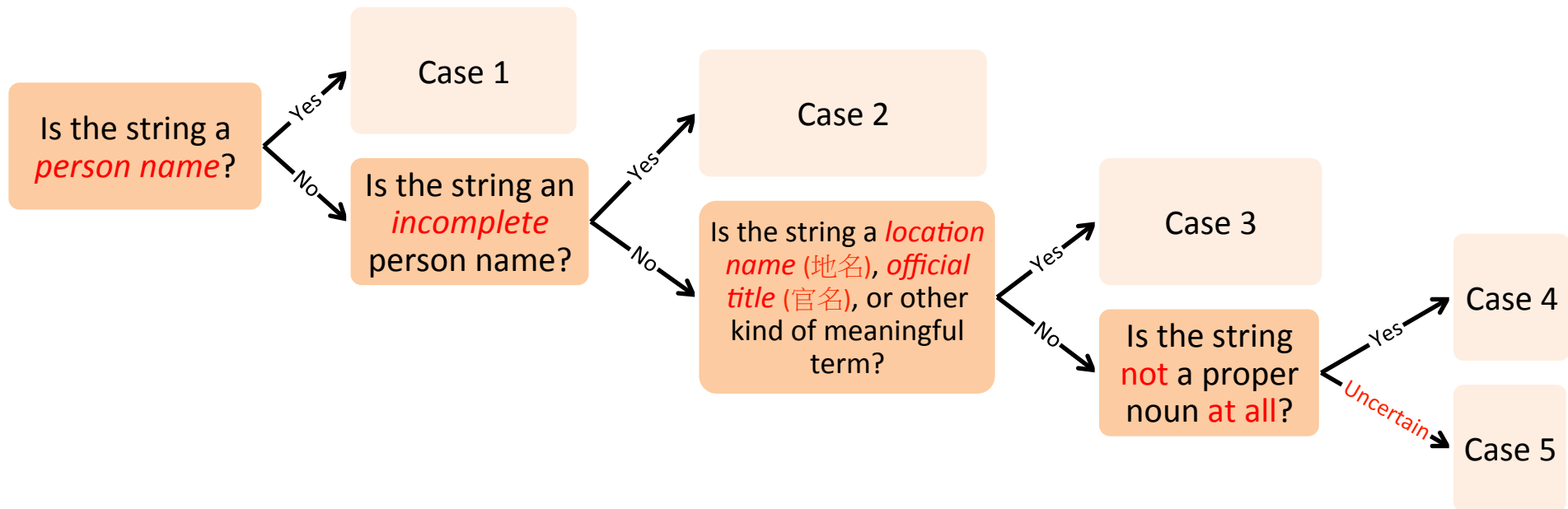
☐ 其他詞彙(官職、稱號、特殊詞彙等)

☐ 不是專有詞彙

☐ 不確定

直接送出 回報細節 關閉回報視窗

Cases you might encounter in the verification process



Case 1: the string **IS** a person name

- Ex.: 趙充國 is a person name.

The screenshot shows a search interface. On the left, a table lists terms and their frequencies. The main area displays search results for the query '+!specific:趙充國 +{DC:Part_055}'. A red arrow points to the first result, which is a historical document snippet mentioning 趙充國.

Term t	df	t→q
趙充國	7	0.304
鄭民憲	4	0.444
強科	3	0.600
張選	3	0.600
吳浚	3	0.500
陳未	3	0.333
李若愚	3	0.250
羅長康	2	1.000
暴斂	2	0.667
韓通	2	0.667

依 df 降序排列

其他

找到筆數: 7 頁次: 1

1. 食貨六三之七六/神宗/熙寧五年 == 五年/四月/十日

五年四月十日，權發遣延州趙高乞差通判范子儀及機宜官魏璋、左文通等根括閑地，乃提舉招置弓箭手。從之。先是，高管勾本路機宜文字，上《營田議》曰：「昔趙充國興屯田以破先零，唐宰相裴師德嘗為檢校營田使，而河西、隴右三百六十屯歲入六十餘萬石。今陝西雖有曠土，而未嘗耕種。朝廷屯戍不可撤，而遠方有輸納之勤。願以閑田募民耕種，以紓西顧之憂。」上以其事下經略安撫使郭達。達言：「今懷寧寨新得地百里，已募漢、蕃戶使為弓箭手，實無閑田以募耕者。」故至是尚復乞根括焉。

食貨：(皇帝·年代) ◀ ▶ 06.神宗 ▶ 熙寧五年四月十日 (熙寧5年) ▶ SongHW_054876_p4_03656

- How to send your feedback?
=> Choose the 是個人名 option.
=> Then click the 直接送出 button.

The screenshot shows a feedback dialog box titled 「趙充國」... with radio button options for classification. A red arrow points to the '是個人名' option, and another red arrow points to the '直接送出' button.

「趙充國」...

☒ 是個人名

☐ 是個地名

☐ 其他詞彙(官職、稱號、特殊詞彙等)

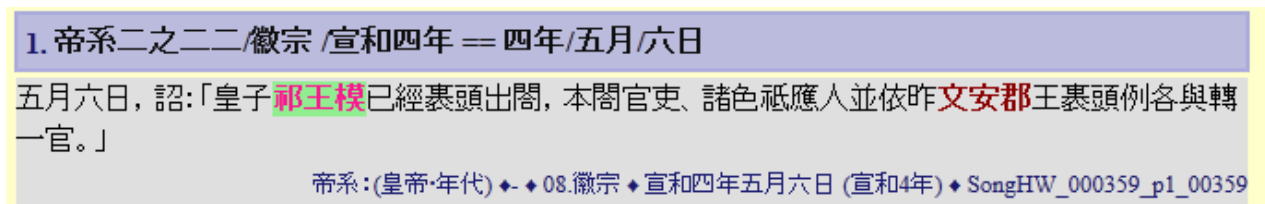
☐ 不是專有詞彙

☐ 不確定

直接送出 回報細節 關閉回報視窗

Case 2: the string is an **incomplete** person name

- Ex: 祁王模 does refer to a person. But the string “祁王模” is not a complete name: 祁王 is the title and 模 is only the first name. The complete name should be 趙模.



- In this case you need to report two things:
 - 祁王模 does refer to a person
 - The complete name of 祁王模 is 趙模
- To do so, you need to:
 - choose the 是個人名 option.
 - click the 回報細節 button to add more details.



Case 2 continued

3. Another form -- 人地特殊詞彙編輯 -- will pop up. Fill in the following information:

the term

the complete name

type of the term

related document ID

comments

The form is titled '人地特殊詞彙編輯' (Person/Place Special Vocabulary Editing) and includes a 'close' link in the top right corner. Below the title is a note: '(您的指正將交管理者處理並儘快...)' (Your correction will be handled by the administrator and processed as soon as possible...). The form contains several input fields with green labels on the left:

- 詞名** (Term Name): The input field contains '祁王模'.
- 完整人地名** (Complete Person/Place Name): The input field contains '趙模'.
- 類別** (Category): The input field contains '人名' (Person Name).
- 相關檔名** (Related Document Name): The input field contains 'SongHW_000001_p1_00001'.
- 參照出處** (Reference Source): An empty input field.
- 同義詞** (Synonyms): An empty input field.
- 任何相關建議** (Any related suggestions): An empty input field.

At the bottom of the form are two buttons: '送出' (Submit) and '重設' (Reset).

Annotations in red speech bubbles provide instructions:

1. Type in “趙模” as the complete person name of “祁王模”
2. Type in “人名” to specify this term is a person name.
3. Click the 送出 button to submit your feedback.

Case 3: the string is *not a person name* but
it *is a proper noun*

- Ex., 黃西 is a location name

1. 食貨六三之七八/神宗 /元豐二年 == 二年/六月/十五日

六月十五日，都大提舉淤田司請以雍[丘]縣黃西等十棚牧地為官莊田，從之此條與下文元豐三年六月十五日一條全同，按《長編》繫此條於元豐三年，則此處當衍。。

食貨：(皇帝·年代) ◆ ◆ 06.神宗 ◆ 元豐二年六月十五日 (元豐2年) ◆ SongHW_054888_p4_03668

=> choose 是個地名, then click 直接送出

「黃西」...

☐ 是個人名

☒ 是個地名 **Click**

☐ 其他詞彙(官職、稱號、特殊詞彙等)

☐ 不是專有詞彙

☐ 不確定

Click 直接送出 回報細節 關閉回報視窗

Case 3: another example

- Suppose a term T is not a person name, nor a place name. But it is *a proper noun* such as an official title.
 - Choose 其他詞彙, then click 回報細節 to give more details.
 - Specify the *type* for this term. You can use any of the *predefined types* (人名, 地名, 官名, 其他) or *make up a new type*.

「周馬帥」...

☐ 是個人名

☐ 是個地名

☒ 其他詞彙(官職、稱號、特殊詞彙等)

☐ 不是專有詞彙

☐ 不確定

Click

直接送出 回報細節 關閉回報視窗

Click

close

人地特殊詞彙編輯

(您的指正將交管理者處理並儘快更正)

詞名	周馬帥
完整人地名	
類別	官名
相關檔名	SongHW_054371_p4_03151 (optional)

Specify the type.

Case 4: the string is **not** a proper noun **at all**

- Ex: 陳未 is not a term

1. 食貨六二之三 / 高宗 / 紹興二十四年 == 二十四年九月四日

二十四年九月四日，監登聞鼓院曹紱言：「常平之法，肇自前古，迨及聖朝及：原脫，據本書食貨五三之二五補。加惠元元，其出納之方，尤切注意。米賤則斂之於官，貴則散之於民，使農、末皆利而國用益裕，是堯、湯先具之備也。今者時和政協，歲已告豐：原作「豐」，據本書食貨五三之二五改。視向日新陳未接之際，其價益平，然能因天之所與以利于下，寔甚盛之舉也。欲望明詔有司，俾令州縣及時廣糴，使倉廩充寔，異時用以賑貸，則下民永無乏食之患。」詔令戶部措通，申尚書省。其後戶部言：「乞下諸路常平司，嚴切行下所屬遵依見行條法及已降指揮施行，毋致稍有違戾。如本司不切檢察，即仰漕、憲司按劾施行。」從之。

食貨：(皇帝-年代) ◆ 10 高宗 ◆ 紹興二十四年九月四日 (紹興24年) ◆ SongHW_054228_p4_03008

=> Choose 不是專有詞彙, then click 直接送出

「陳未」...

☐ 是個人名

☐ 是個地名

☐ 其他詞彙(官職、稱號、特殊詞彙等)

☒ 不是專有詞彙

☐ 不確定

直接送出 回報細節 關閉回報視窗

Case 5: uncertain whether the string is a proper noun or not

- Ex: not sure if 邢中和 is a proper noun or not

1. 禮二九之二四/禮二九 歷代大行喪禮上 /真宗 == 真宗乾興元年/六月/五日

六月五日，命龍圖閣直學士呂夷簡、魯宗道、入內押班岑保正、入內供奉官任守忠，即時乘傳詣永安縣相度皇堂地，仍遣司天監主簿侯道寧、周訥隨往。又令夷簡召京城習陰陽地理者三五人偕行。先是，藍繼宗與王承勳按行山陵，封域已定，又命雷允恭覆按。允恭乃與邢中和輩擅移皇堂就東南，地頗峻側，衆知非便，以允恭交結丁謂，莫敢言者。開築之際，土石相半，興作踰月，皇堂

=> Choose 不確定 (as a type), then click 直接送出

The screenshot shows a web form titled 「邢中和」... with the following options:

- ☐ 是個人名
- ☐ 是個地名
- ☐ 其他詞彙(官職、稱號、特殊詞彙等)
- ☐ 不是專有詞彙
- ☒ 不確定

At the bottom, there are three buttons: 直接送出, 回報細節, and 關閉回報視窗. A red arrow points to the 直接送出 button, and another red arrow points to the 不確定 radio button.

You can check to see which terms you have already reported on

- Go to: http://140.112.30.227/song_feedback/search.php
- Type in your *email address* (your username) as 回報者 and click 查詢, you will see all the terms you ever reported and *how* you reported them.

宋會要人名回報系統

共 9 筆 ◀ 第 1 頁 / 共 1 頁 ▶

詞名	類別	完整人地名	檔名	回報者
王仲	非專有詞彙			shihpeichen@fas.harvard.edu
王仲	非專有詞彙		SongHW_000001_p1_00001	shihpeichen@fas.harvard.edu
劉氏	非專有詞彙		SongHW_001001_p1_01001	shihpeichen@fas.harvard.edu
錢氏	非專有詞彙		SongHW_001001_p1_01001	shihpeichen@fas.harvard.edu
滕氏	非專有詞彙		SongHW_001001_p1_01001	shihpeichen@fas.harvard.edu
榮易	人名	趙允良	SongHW_001001_p1_01001	shihpeichen@fas.harvard.edu
宰輔表	非專有詞彙		SongHW_000001_p1_00001	shihpeichen@fas.harvard.edu
祁王模	人名	趙模	SongHW_000001_p1_00001	shihpeichen@fas.harvard.edu

詞名

回報者 shihpeichen@fas.harvard.edu

查詢

The reported terms

Go to the online CBDB database to check whether a person name is already in the database

- Go to:

[http://db1.ihp.sinica.edu.tw/cbdbc/ttsweb?
@0:0:1:cbdbkmeng@@0.0542755133964546
44](http://db1.ihp.sinica.edu.tw/cbdbc/ttsweb?@0:0:1:cbdbkmeng@@0.054275513396454644) and search by the name.