

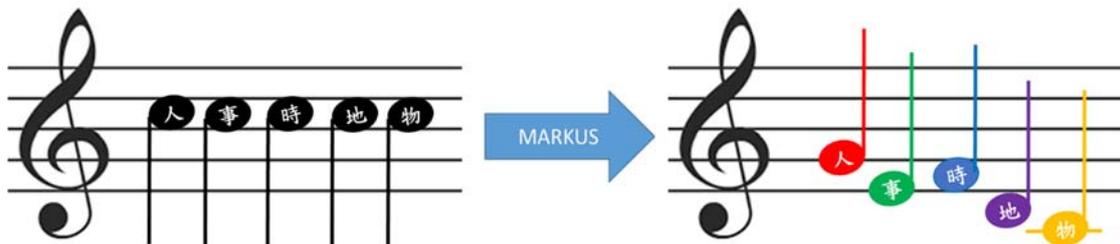
MARKUS 標記工具

方珮雯* 胡其瑞**

什麼是 MARKUS 標記工具

MARKUS 的正式名稱為「古籍半自動標記平台」，由荷蘭萊頓大學 Hilde De Weerd (魏希德) 教授與何浩洋博士設計開發，是一個線上文本標記工具。透過自動或手動標記，使用者可以將文本檔案，標出人名、地名、年號及官名等重要關鍵字。再以 MARKUS.html 格式輸出後，以其他數位人文工具進行進一步的分析。

簡單地說，一個純文字的文本檔案 (.txt) 對於電腦而言，是一個個平面的數值。但是對使用者來說，文本中的每個字詞，可以代表著不同的意義。在一篇文本中，可能包含了人、事、時、地、物，甚至更多不同性質的詞彙；但未經標記處理的文本，對電腦來說，就像是一個沒有聲調的音符一樣。而經過 MARKUS 標記過的文本，則可以讓電腦知道每個字詞所屬的不同音階，使文本對電腦來說，成為一個有意義的曲譜，並在其他的平台與工具上，做進一步的分析。



網址：<http://dh.chinese-empires.eu/markus/beta/index.html>

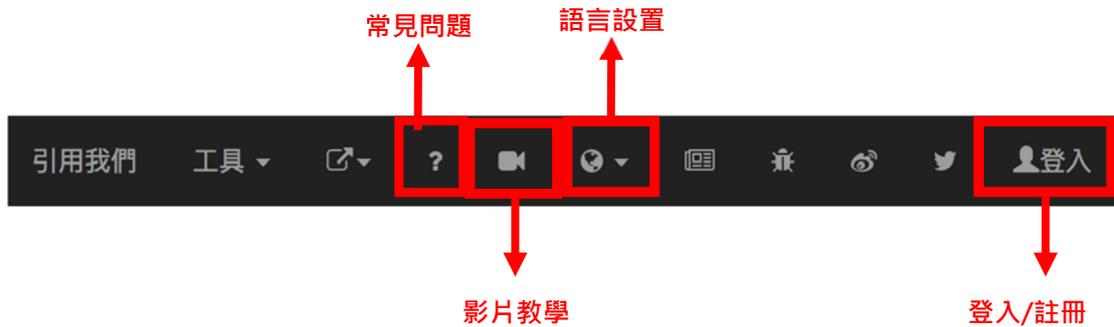
(目前只能運行於 google chrome 瀏覽器)

* 台灣大學資訊工程學系碩士生

** 台灣大學數位人文研究中心博士後研究人員

首頁工具列介紹

首頁上方有一個工具列，分別為：



1. 語言設置

· MARCUS 現有三種語文可供使用者選擇，分別是：英文、正體中文與簡體中文。

- 點擊工具列右上方的地球圖示
- 選擇「正體中文」
- 完成設定
- 可參考操作影片「如何更改語言設置」：

http://dh.chinese-empires.eu/MARKUS/beta/video_zhtw.html#languageSettings



2. 登入/註冊

MARKUS 大部分功能在未登入的狀況下仍可使用。若使用者註冊 MARKUS 帳戶，則可以在網站中保存工作進度、新增標記、使用段落標記功能等。登入/註冊步驟如下：

- 點擊工具列右上角「登入」，進入登入頁面



- 若使用者已經擁有 MARKUS 帳號，輸入註冊使用之 email 及密碼，按下登入，便能以您的個人身份回到網站首頁。

➔ Login with MARKUS account

The image shows a login form with two input fields: 'Email' and 'Password'. Below the fields is an orange 'Login' button. A red box highlights the entire form area, with a red '1' to its right. A second red box highlights the 'Login' button, with a red '2' below it.

· 若使用者尚未註冊，且已經擁有社群網站帳戶，可選擇以 Facebook/Twitter/Google 帳戶登入，點選對應按鈕即可。

🔒 Login with your social account



· 使用者亦可在 MARKUS 網站直接註冊。點選「Signup here」，依序輸入電子信箱、密碼、名、姓與服務單位，按下「Sign up」按鈕，即完成註冊

➔ Login with MARKUS account

The image shows a login form with two input fields: 'Email' and 'Password'. Below the fields is an orange 'Login' button.

Need an account? [? Signup Here](#) 1

➔ Signup

The image shows a signup form with five input fields: 'Email', 'Password', 'First Name', 'Last Name', and 'Affiliation'. To the right of each field is a red label: '電子信箱', '密碼', '名', '姓', and '服務單位'. Below the fields is an orange 'Signup' button. A red box highlights the entire form area, with a red '2' to its right. A red arrow points from the 'Signup Here' link in the previous block to the 'Email' field. A red '3' is below the 'Signup' button.

· 可參考操作影片「如何登入 MARKUS 或建立新帳號」：

http://dh.chinese-empires.eu/markus/beta/video_zhtw.html#loginAndCreateAccount

3. 影片教學

MARKUS 為使用者拍攝了一系列操作教學影片，除了閱讀本說明手冊，使用者也可以直接透過影片學習網站操作。

- 點擊工具列右上方的攝影機圖示，進入影片教學頁面
- 在左方列表選擇您需要的教學影片，或直接在右方列表上下滑動至所需的影片，按下播放鍵即可。



除了第一個「什麼是 MARKUS? 如何使用?」演講影片外，其他皆為英文發音教學影片。若預設字幕為英文，可改為中文字幕。

- 點擊影片視窗內的設定圖示
- 選擇「設定」
- 選擇「中文(繁體)」
- 影片即能顯示中文字幕





4. 常見問題

其他關於文檔上傳、詞彙標記的疑問，可參閱此常見問題頁面。

- 點擊工具列右上方的問號圖示，進入常見問題頁面
- 在列表中選擇需要的疑問解答



上傳文本

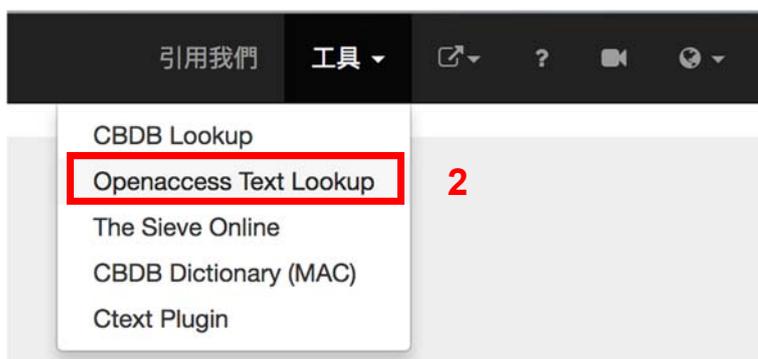
1. 尋找文本

首先，使用者準備欲標記的文本檔案。MARKUS 目前僅能針對純文字，或是轉換成 UTF-8 編碼的 txt 純文字檔進行標記。若使用者沒有文本，也可以透過 MARKUS 為使用者連結的開放文本平台（Openaccess Text Lookup）尋找需要的資料。

- 點擊工具列右上方的「工具」
- 選擇「Openaccess Text Lookup」，瀏覽器會開新視窗。
- 點擊搜尋窗格，輸入需要的中國古代原典文獻，例如輸入：太平天國



1



MARKUS open access text (Ctext)



MARKUS open access text (Ctext)



- 系統會依據關鍵字，列出含有此關鍵字標題、不同版本或相關連的文獻



- 在這裡 MARKUS 提供了兩種下載文獻的方式，如果使用者所在的 IP 位置是有訂閱 CText 的機構 IP 範圍內，則可以使用 MARKUS+ 的直接下載功能。若非享有此項附加功能的使用者，則可以點選 CTEXT 進入哲學電子書資料庫 (<http://ctext.org/zh>) 進行逐篇的下載步驟。

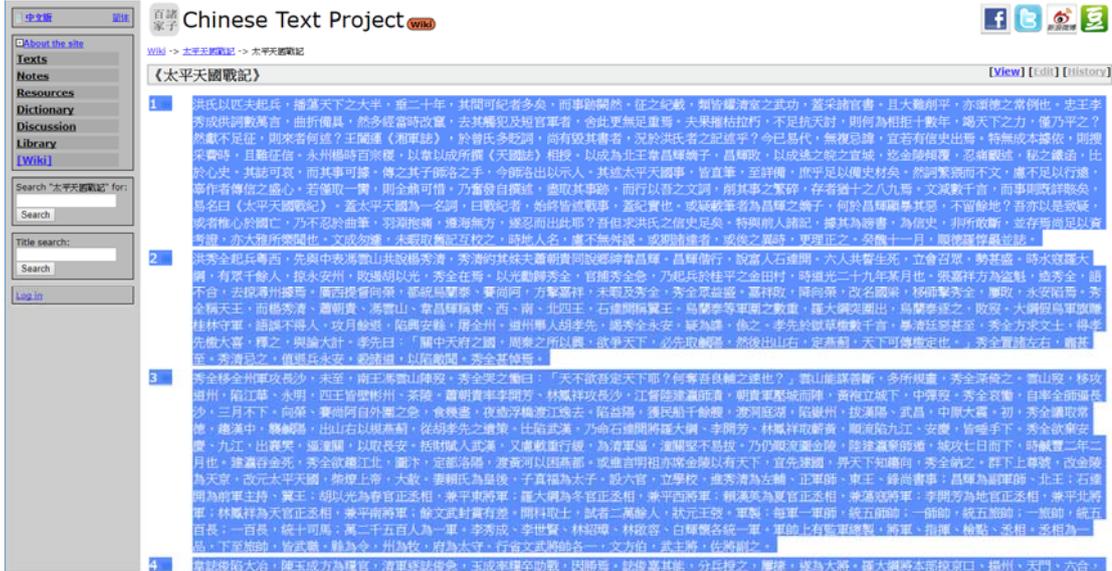
2. 上傳文本

我們可以將所需文字複製貼上，或儲存成 UTF-8 編碼的 txt 純文字檔。

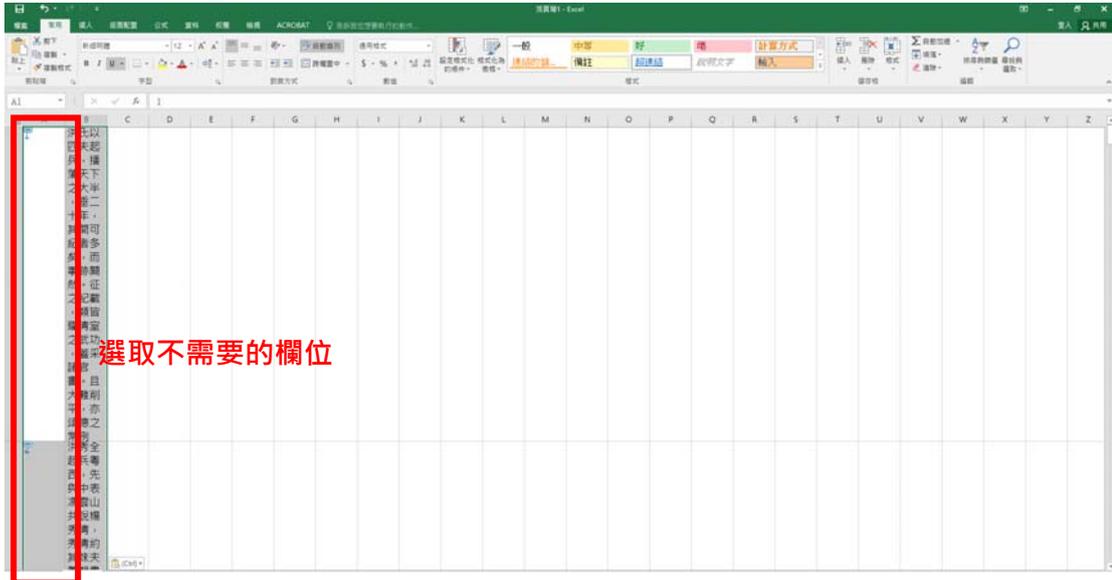
(1) 將 CText 文字複製貼上

由於 CText 文字在複製時會連同分段資訊一同複製起來，在貼入 MARKUS 之後需要再另外刪除不必要的資訊。因此，我們建議可以先將 CText 的文字複製之後，先貼入 EXCEL 的表單內，刪除不必要的資訊後，再貼回 MARKUS 裡面。以下為範例試做：

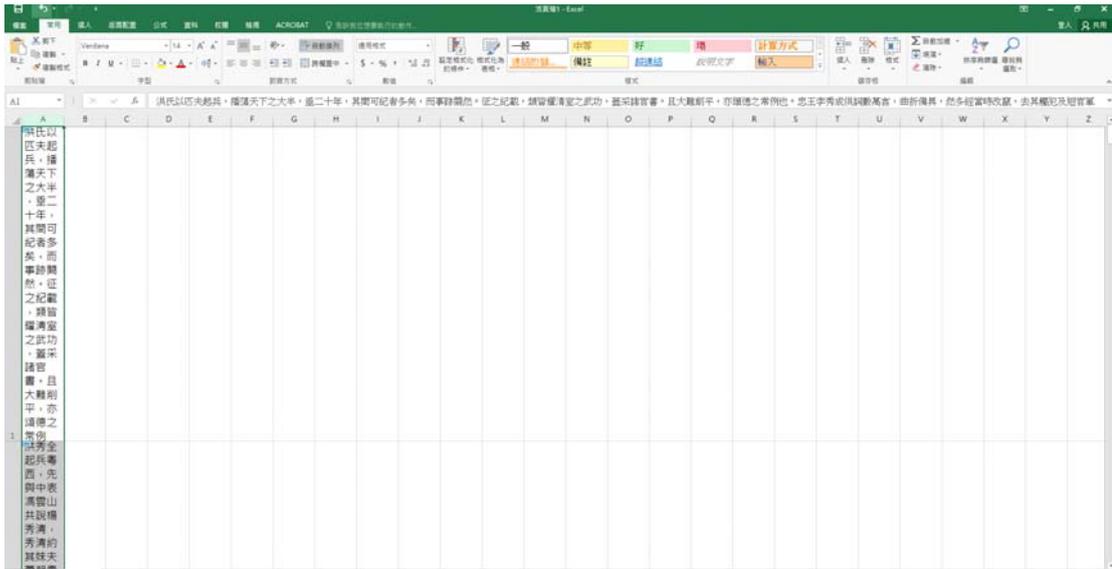
- 複製您需要的文本文字（選取文字後按 Ctrl+c）



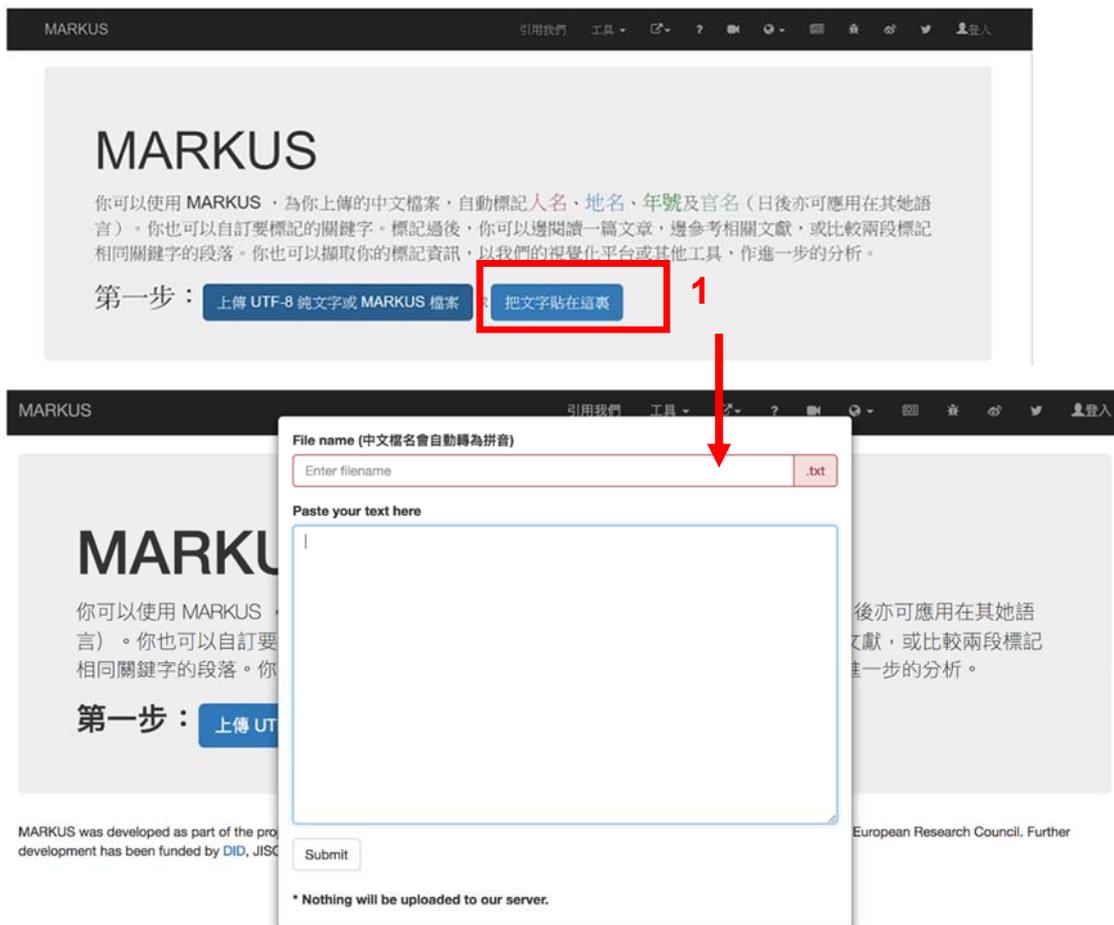
- 開啟 EXCEL 後貼上 (Ctrl+v) 文字



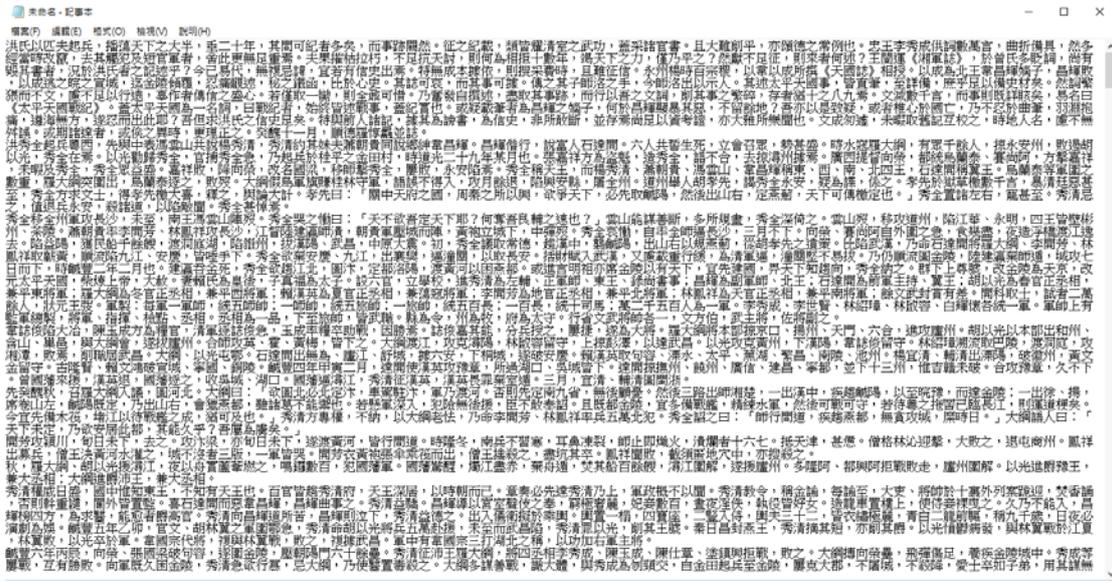
- 將不需要的欄位內容選取刪除



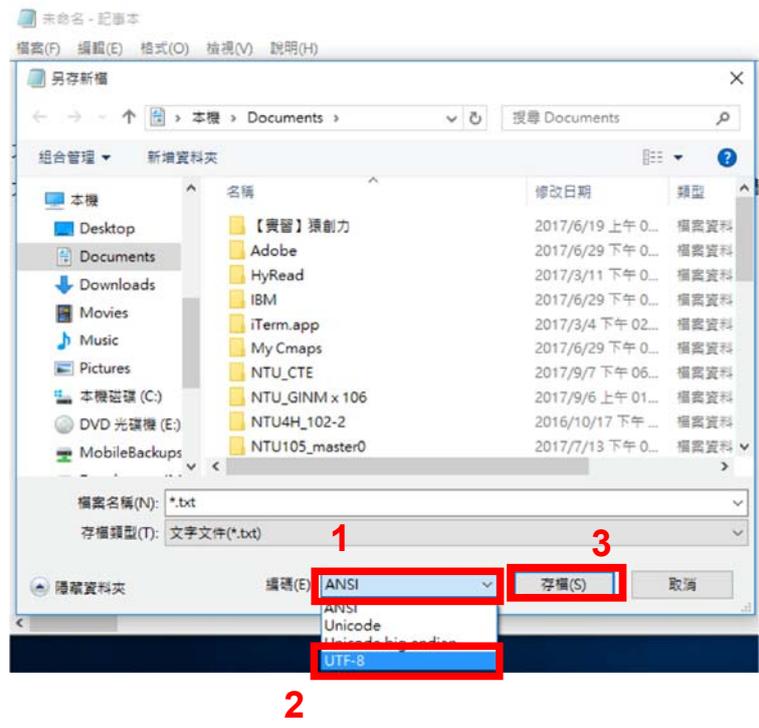
- 選取剩下的文字部分，複製文字 (Ctrl+c)，然後回到 MARKUS 首頁。
- 點選「把文字貼在這裡」，會跳出編輯視窗



- 將檔案命名，並貼上複製的文字文字，完成後按下「submit」

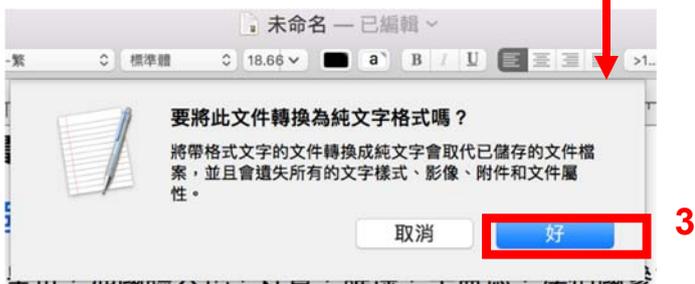
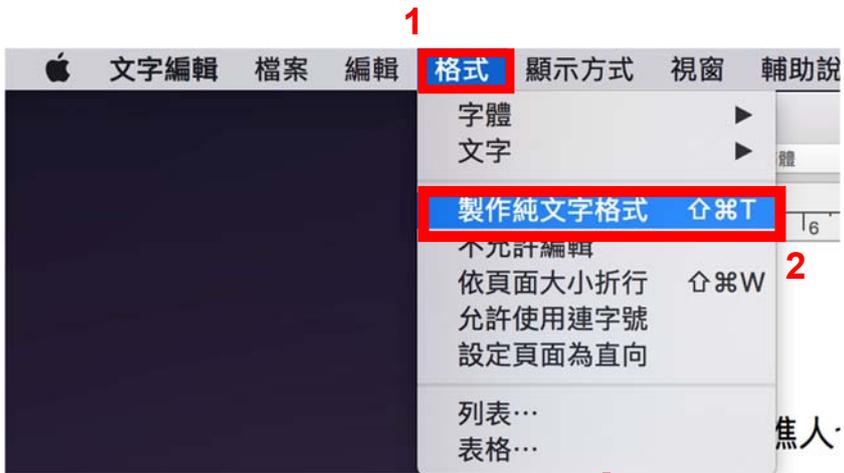


· Windows 記事本的預設編碼是 ANSI。因此，在另存新檔時，命名完後請點選視窗下方「ANSI」旁邊的箭頭，將編碼改成「UTF-8」，再按下存檔。

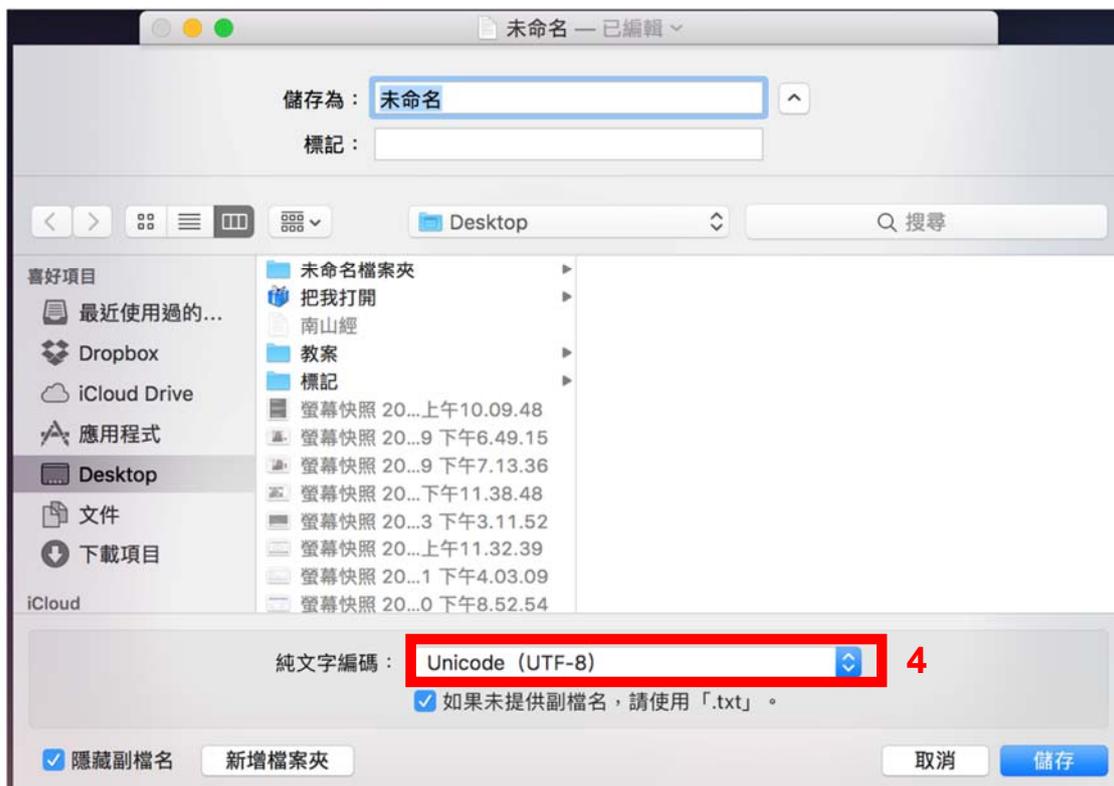


[OS 系統]：以 macOS Sierra 系統為例：

- 開啟文字編輯工具
- 貼上所需文字
- 點選「格式」->「製作純文字格式」，按下「好」，將文本轉成純文字



- 存檔時，點選「純文字編碼」旁邊的箭頭，將編碼形式改成 Unicode (UTF-8)，再儲存檔案





· 在您的作業系統 (windows/OS) 完成儲存後，請回到 MARKUS 首頁，點選「上傳 UTF 純文字或 MARKUS 檔案」，選擇您存好的 txt 純文字檔



· 首頁一樣會跳轉到第二步驟，準備選擇標記方式



進行標記

目前能使用的標記方式有兩種：「自動標記」、「手動標記」與「關鍵字標記 (測試中)」三種。一般而言，我們建議使用者先以 MARKUS 內建的標記功能進行「自動標記」，刪除錯誤或不需要的詞彙後，再進行手動標記。

1. 自動標記

*操作影片：http://dh.chinese-empires.eu/MARKUS/beta/video_zhtw.html#automaticMarkup

- 點選二 A：自動標記下方的「Ready」按鈕

二A：自動標記

以中國歷代人物傳記資料庫 (CBDB) 標記人名 (全名、字號)、官名及年號。以中國歷史地理信息系統 (CHGIS) 標記地名。

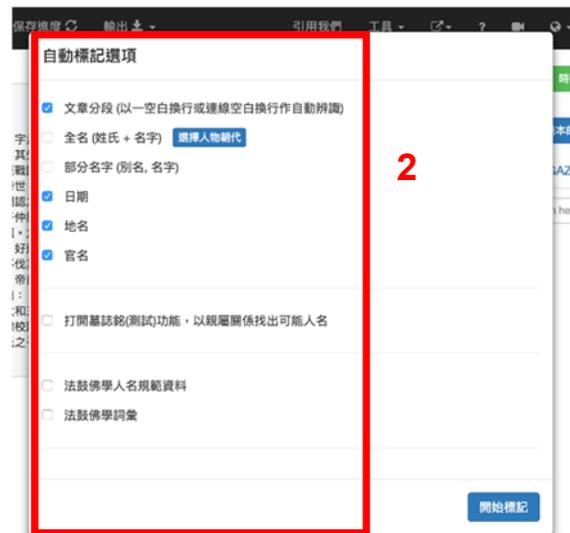
你可以勾選要標記的類別。

(暫時只支援正體中文)



1

- 網頁會跳轉到自動標記選項頁面。根據您的需求，勾選對應的方框。



文章分段：在您的 txt 檔案，文字若有空白換行，MARKUS 會自動將其分為不同段落，在資料庫存成一筆一筆資料，以利後續分析

*欲進一步了解段落標記功能，可參考操作說明影片：http://dh.chinese-empires.eu/MARKUS/beta/video_zhtw.html#passageMarkupFunction

全名 (姓氏+名字)：自動標記全名。您可依文本內容大概所屬的時間，選按「選擇人物朝代」，以限縮標記內容的範圍並節省系統比對的時間。

- 文章分段 (以一空白換行或連線空白換行作自動辨識)
- 全名 (姓氏 + 名字) 選擇人物朝代 **3**
- 部分名字 (別名, 名字)
- 日期
- 地名
- 官名

因為《太平天國戰記》屬於清代發生的事情，因此我們勾選出清代以縮小範圍。

公元前1100 ~ 西元221

西元221 ~ 589

西元589 ~ 907

西元907 ~ 1279

西元1279 ~ 1911 **4**

^

- 元 (1279 ~ 1367)
- 明 (1368 ~ 1644)
- 北元 (1371 ~ 1387)
- 南明 (1644 ~ 1661)
- 清 (1644 ~ 1911)

5 **點選箭頭可伸縮選項框**

· 按下「選取」按鈕，回到自動標記選項選項視窗

西元1911 ~

選取

6

* 墓誌銘選項為進階功能，請參考操作說明影片：http://dh.chinese-empires.eu/MARKUS/beta/video_zhtw.html#muzhimingFunction

· 點選這些顏色標籤，能改變顯示風格，以作為閱讀上的明顯區分。風格有三種，以姓名為例，分為：



· 若標記下方出現紅色虛線，表示此詞彙有不只一種搜尋結果，例如在太平天國後期的將領陳玉成，在此次的檢索中就出現了紅色虛線。



· 點擊此詞彙會看到有三個由 CBDB 所提供的標記：



這表示在使用者指定的範圍中，同一個詞彙有兩個以上的選擇，而系統判斷不出是哪一位，這個時候就必須由使用者來自行判斷。使用者可以點選右方的小書標誌以確認 CBDB 中這三個 ID 編碼所代表的人中，哪一位才是使用者需要的陳玉成。

CBDB TGAZ ZDict Wikipedia

Search 65436|124581|124582

65436 124581 124582

65436

CBDB ID: 65436
索引/中文/英文名稱: 陳玉成 Chen Yucheng
指數年 (index year): 1862
生年: 清道光17年 (1837)
卒年: 清同治1年 (1862)
享年: 26
郡望: 【未詳】
註: 《清代人物生卒年表》定其生卒年為道光17年~同治1年。【參考《清代人物生卒年表》#13124。】

出處: 人名權威資料 (中央研究院歷史語言研究所), 頁 8862

別名: 室名、別號三十檢點回馬槍, 室名、別號四眼狗, 別名、曾用名陳丕成。

地理資訊: 另一籍貫(基本地址): 清朝 / 廣西省 / 梧州府 / 藤縣
 出處: 頁13124
 註: 廣西 藤縣(清代人物生卒年表)。
 籍貫(基本地址): 清朝 / 廣西省 / 潯州府 / 貴縣
 出處: 頁8862
 註: 廣西省-潯州府-貴縣(中央研究院人名權威資料)。

· 在這裡可以看到共有「65436」、「124581」與「124582」三個 ID 是陳玉成，而其中，「65436」是正確的陳玉成。我們就可以在該關鍵詞上將其餘兩個不正確的 ID 點按消除。點按完後請記得要點選儲存。



· 但在本書中，陳玉成不只提到一次，因此我們可以利用搜尋的工具將所有的陳玉成搜尋出來，並且一次完成標記。請點選搜尋用的放大鏡圖示。跳出的視窗告訴我們，本書中陳玉成不只出現一次，而且 MARKUS 都將其標示了三個不同的 ID。而第一個陳玉成則是我們已經確認過 ID 為 65436 的，所以請點選右方小勾。再將選單拉至最底部，點選「全部套用」，即可將所有的陳玉成標記為 65436 的陳玉成了。





· 若看到標記錯誤的地方，請點選該詞彙，按下垃圾桶圖示刪除標記。例如，系統將「忠王李秀成供詞數萬言」中的「萬言」當成是人名，故要將其刪除。



請注意，刪除後的標籤無法復原，如欲重新標示，需進入手動標記程序。

2. 手動標記

自動標記偶有標記錯誤或闕漏，因此，通常在自動標記完後，我們會再進行手動標記。當然，使用者一開始就可以在首頁選擇「手動標記」，避免標記增減所消耗的額外心力。而進入手動標記之後，也可以點選螢幕左側的圖示返回「自動標記」或進入「關鍵字標記」



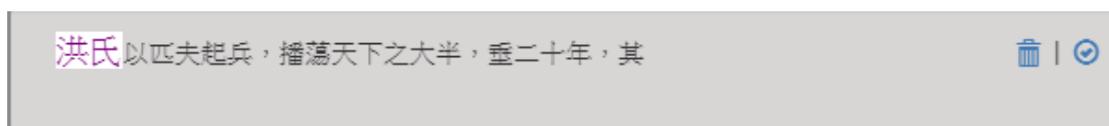
進入手動標記頁面後，您就可以添加自己的標記了。

舉例來說，本書一開始提到「洪氏以匹夫起兵」，當中的洪氏顯然指的是天王洪秀全。洪秀全在本書他處已經由自動標記標記出為人名，並且添上 CBDB 的 ID 為 65447。所以使用者也可以為「洪氏」標記為人名並加上 ID。



手動標記的方法是先將欲標記的詞選取，螢幕上就會直接出現現有預設的標記分類，使用者點選「姓名」後，就可以將「洪氏」標記姓名。但同樣地，在本書中，洪氏出現不只一次，所以我們依然可以使用搜尋（放大鏡按鈕）來檢索出所有的「洪氏」並加以標注。

我們可以把洪秀全的正確 ID 輸入後，點選儲存，勾選右方的小勾，再將選單拉至最底部，點選「全部套用」，即可將所有的洪氏標記為 65447 的洪秀全了。



需要留意的是，不見得每個洪氏在本書中都指的是洪秀全，因此，為了避免錯誤標記，建議使用者還是逐筆檢視，以確認是否標記有誤，如有錯誤者，點選右方垃圾桶按鈕即可刪除於全部標記的選單當中。

除了標記姓名，預設的標記類別中還可以選擇「別名」、「時間」、「地名」與「官名」。

在地名方面，MARKUS 提供了與 TGAZ 的連結可以標記地名。但由於中國相同的地名太多，為了簡化標記程序，系統僅會先行標記該詞彙為地名，而未帶入 TGAZ 的 ID。我們可以在右方的地圖中尋找合適的 ID，再將其填入或點選帶入標記當中。以本書中的「長沙」為例，使用者在 TGAZ 的列表中確認長沙的 ID 為 hvd-32825，如此帶入，再以同樣的方法搜尋並標記全部的長沙即可。關於 TGAZ 更詳細的說明，請參見「4. 線上參考資料」。

爭天下，必先取關，然後出山右，定燕劑，天下可傳檄定也。」秀全置諸左右，麗甚至。秀清忌之，值退兵
 與諸道，以陪敵聞。秀全甚悼焉。
 全 王馮雲山陣歿。秀全哭之慟曰：「天不欲吾定天下耶？何奪吾良輔之速也？」雲
 斷 雲山脫，移也。滇州，陷江華、永明，四王皆墜。彬州、茶陵。肅朝貴率
 林鳳祥攻長沙。地名 軍壓城而陣，黃袍立城下，中彈歿。秀全哀慟，自率全師逼長
 司不下。向榮、袁尚阿百外圍之急，夜幾盡，夜進浮橋渡江逃去。陷益陽，獲民船千餘艘，渡洞庭湖，陷嶽
 陽、武昌，中原大震。初，秀全讓取常德，趨漢中，襲歸陽，出山右以規燕劑，從胡季先之道策。比陷武
 石達開將羅大綱、李開芳、林鳳祥取贛黃，順流陷九江、安慶，皆唾手下。秀全欲棄安慶、九江，出襄
 陽關，以取長安。括財賦入武漢，又慮載重行艱，為清軍逼，潼關堅不易拔。乃仍順流圖金陵，陸建瀛棄師
 七日而下，時咸豐二年二月也。建瀛吞金死，秀全欲趨江北，圖汴，定都洛陽，渡黃河以困燕都。或
 亦席金陵以有天下，宜先建國，與天下知趨向，秀全納之。群下上尊號，改金陵為天京，改元太平天國，柴煥
 敏。妻賴氏為皇后，子真福為太子。設六官，立學校，進秀清為左輔、正軍師、東王、錄尚書事；昌輝
 、北王；石達開為前軍主持、翼王；胡以光為春官正丞相、將軍；羅大綱為冬官正丞相，兼
 軍；賴漢英為夏官正丞相，兼編寇將軍；李開芳為地官正丞相，兼平北將軍；林鳳祥為天官正丞
 南將軍；餘文武封賞有差。開科取士，試者二萬餘人，狀元王張。軍製：每軍一軍師，統五師帥；一師帥，
 帥；一旅帥，統五百長；一百長，統十司馬；萬二千五百人為一軍。李秀成、李世賢、林紹璋、林啟容、白
 一軍。軍帥上有監軍總製，將軍、指揮、丞相為一品，下至旅帥，皆武職。縣為令，州為
 太守。行省文武將帥各一，文方伯，武主將，佐將副之。
 大冶，陳玉成方為糧官，清軍逐諸侯急，玉成率糧卒助戰，因勝焉。諸侯嘉其能，分兵授之，屢捷，遂為大
 綱將本部掠京口、揚州、天門、六合，進攻廬州。胡以光以本部出和州、含山、無色，與大綱會，遂拔
 帥攻英、霍、黃梅，皆之下。大綱渡江，攻克灣陽，林啟容留守，上掠彭澤，以建武昌。以光攻克黃
 陂陽，聿諸侯留守。林紹璋溯流取巴陵，渡洞庭，攻湘潭，敗焉，別踞居武昌。大綱、以光屯鄂。石達開
 廬江、舒城，兼六安，下桐城，遂破安慶。賴漢英取句容、溧水、太平、蕪湖、繁昌、南陵、



· 不過，預設的標記分類，可能並不足夠讓使用者使用，因此，MARKUS 也提供了一個標記管理的功能，讓使用者新增自定義標記。但這個功能必須擁有並登入 MARKUS 帳號方能使用。

在登入後的標記頁面右方，您可以看到一個管理標記圖示。

- 點選標記圖示，進入管理頁面



- 預設標記有五個，您可以在左方「快捷顯示」欄位，勾選欲顯示的預設標記；亦可在右方「刪除所有標記」欄位，刪除不需要的預設標記

快捷顯示	標記	標記名稱	按鈕名稱	刪除所有標記
<input checked="" type="checkbox"/>	姓名	fullName	姓名	<input type="checkbox"/>
<input checked="" type="checkbox"/>	別名	partialName	別名	<input type="checkbox"/>
<input checked="" type="checkbox"/>	時間	timePeriod	時間	<input type="checkbox"/>
<input checked="" type="checkbox"/>	地名	placeName	地名	<input type="checkbox"/>
<input checked="" type="checkbox"/>	官名	officialTitle	官名	<input type="checkbox"/>

- 若欲新增使用者的自定義標記，輸入新增的標記名稱（須為英文）、按鈕名稱，再選擇喜好的標記顏色，按下「新增」按鈕。也可以直接輸入按鈕名稱（中文），系統會自動將音譯名稱帶入標記名稱當中。

· 加完所有標記，點選「確定」，使用者的自定義標記即會出現在標記列表。

範例： 標記名稱：bookName；按鈕名稱：書名；標記顏色：紫色

新增標記 (標記名稱的中文會自動轉為拼音)

標記顏色	標記名稱	按鈕名稱	1	2
	<input type="text" value="bookName"/>	<input type="text" value="書名"/>		<input type="button" value="新增"/>

3

3. 關鍵字標記

MARKUS 另有一種標記功能稱為關鍵字標記。可以透過關鍵字的檢索或詞夾子（測試功能）的方式來將一些詞迅速地標記出來。

· 使用者可以透過左側的選單進入關鍵字標記功能。



· 關鍵字標記分為「載入關鍵字/正規表達式」、「關鍵字小幫手」與「詞夾子」等三項功能。點選「載入關鍵字/正規表達式」進入頁面，使用者可以在下方輸入欲作為檢索關鍵字的詞彙。

· 以《太平天國戰記》為例，在這場戰爭中有許多的交戰團體，例如：太平軍、湘軍、淮軍、楚軍、清軍等等。使用者將這些詞彙列在當中，並且在上方增列標記名稱（如：交戰團體），選擇適合的顏色後，點選「載入關鍵字」，即可將上述詞彙以「交戰團體」分類一次標記起來。

載入關鍵字/正規表達式

標記名稱 (中文名會自動轉為拼音)

詞彙請一行一行分開

太平天國
太平軍
湘軍
淮軍
楚軍
清軍

或

- 在關鍵字標記功能中，若是要將不同的關鍵字一次標記成現有的標記類別中，此時需以該類別的英文標記名稱填入，選擇該標記名稱的顏色，便可以將這次的字詞標記呈現有的關鍵字類別中。例如，如果要將使用者所知的人名以關鍵字的方式標記成「姓名」類別時，在標記名稱中則填入「fullName」（注意大小寫需相同），選擇紅色，如此便可將關鍵字納入姓名類別中。
- 此外，MARKUS 也提供了關鍵字小幫手，讓不熟悉文本的使用者，也可以透過這個功能找到需要的資料。舉例來說，太平天國立了不少的王，但是除了一般人所熟知的東王楊秀清、翼王石達開之外，很多王一般人並不熟悉。使用者可以在關鍵字小幫手中「欲分析的關鍵字」裡填入「王」，而左方綴詞字數輸入 2，右方綴詞字數輸入 3，構成「OO 王 OOO」的格式，然後點選「開始分析」，即可檢索出許多的王的名稱。刪除不需要的部分後，點選「標記已分析的詞彙」，其後功能就像「載入關鍵字」一樣，可以將這些王一次標記在相同的分類（如「王位」）當中。

關鍵字小幫手

左方綴詞字數	欲分析的關鍵字	右方綴詞字數
2	三	3
開始分析		
左方綴詞	欲分析的長度	右方綴詞
關鍵字	min max	關鍵字
<input type="checkbox"/> 只選取分析的詞彙		
開始分析		
<ul style="list-style-type: none">、堵黃黃黃金、輔黃黃輔清、侍黃黃李世賢、聽黃黃陳炳文、西黃黃蕭朝貴、南黃黃馮雲山、北黃黃李昌輝、翼黃黃石達開、英黃黃、天官、英黃黃死後、		
標記已分析的詞彙		
重設關鍵字小幫手		

· 關鍵字小幫手的第二個功能，則是以固定綴詞來檢索字詞。例如在章回體的小說中，每個章節都會有「第某回」，但是當中所填的數字可能是一個字，也可能是兩個字或三個字，因此，使用者可以在「左方綴詞」中填入「第」，在「右方綴詞」中填入「回」，中間的文字長度最短(min)一字，最常(max)三字，如此，便能將需要的字詞一次抓取出來進行標記。

· 「詞夾子」則是以一種透過半自動的方式，在固定的行文模式中，找到同類的詞彙。特別是在小說類的文本中，因為當中許多的人物、地名不見得是MARKUS 能夠識別與標注的，透過詞夾子的功能，經過多次的反覆運算，就可以把所需標記的詞彙挑選出來。詞夾子的使用有兩種不同的方式，分別為「挖掘詞彙」與「詞夾子檢索」。以下將以《三國演義》為例進行說明。

· 挖掘詞彙功能：首先，在載入文本以後，先於「種子詞彙」欄中填入一些常見的人物，如：劉備、關羽、張飛、趙雲、孔明。

詞夾子 (測試中)

基本模式:

種子詞彙	發現的詞彙	排除詞彙	保存詞彙
劉備 關羽 張飛 趙雲 孔明			
挖掘更多詞彙		標記已保存詞彙	
重設詞夾子			

點選「挖掘更多詞彙」後，會發現這些名字被帶入「保存詞彙」欄中，而「發現的詞彙」欄中，則出現了一些可能的人名候選詞彙，如：姜維、魏延、黃忠等。

詞夾子 (測試中)

基本模式:

種子詞彙	發現的詞彙	排除詞彙	保存詞彙
請輸入保存詞彙 (一行一個詞彙)	姜維 魏延 雲長 黃忠		劉備 關羽 張飛 趙雲 孔明
挖掘更多詞彙		標記已保存詞彙	
重設詞夾子			

使用者再將「姜維、魏延、黃忠」放入「種子詞彙」欄中，再次點選「挖掘更多詞彙」，則「發現的詞彙」欄中，又會多出一些人名候選詞彙。

詞夾子 (測試中)

基本模式:

種子詞彙	發現的詞彙	排除詞彙	保存詞彙
姜維 魏延 黃忠	取將 明年 姜維 魏延 雲長 黃忠 勝負 鄧艾 蜀兵 馬岱 關興		劉備 關羽 張飛 趙雲 孔明
挖掘更多詞彙		標記已保存詞彙	
重設詞夾子			

如此不斷反覆操作，可以將許多可能的人名候選詞彙不斷挖掘出來。等到使用者挖掘到足夠的詞彙後，點選「標記已保存詞彙」就可以新增至姓名標記當中（記得標記名稱要填寫英文 `fullName` 否則會另外出現一個類別）。

載入關鍵字/正規表達式

標記名稱 (中文名會自動轉為拼音)

詞彙請一行一行分開

劉備
關羽
張飛
趙雲
孔明
姜維
魏延
黃忠
孫乾
周瑜

載入關鍵字 或 載入正規表達式

· 詞夾子檢索功能則與詞彙檢索相反。這個功能是利用作者行文的固定模式，以詞彙的前後綴詞來推測可能的候選詞彙。再藉由不斷的反覆操作，達到與詞彙檢索相同的目的。

首先，也是在「種子詞彙」欄中填入一些常見的人物，如：劉備、關羽、張飛、趙雲、孔明。

詞夾子 (測試中)

基本模式:

種子詞彙	發現的詞彙	排除詞彙	保存詞彙
劉備 關羽 張飛 趙雲 孔明			
挖掘更多詞彙		標記已保存詞彙	
重設詞夾子			

然後點選左下欄的「計算夾子」。就會出現這些名字前後出現過的綴詞，以及可能是同類詞的前後綴詞。

進階模式 (可修正挖掘出的夾子):

種子夾子	排除夾子
┆乃...也┆ 德問...曰┆ ┆被...一槍 乃問...曰┆ 卻說...在祁	
計算夾子	夾取詞彙

使用者可以先看看這些「種子夾子」中是否有不合適的前後綴詞，若有，則可以貼入右側「排除夾子」中。然後點選「夾取詞彙」後，便可以獲得系統告知，夾取了多少詞彙出來。

dh.chinese-empires.eu 顯示：

1738 Keywords found

確定

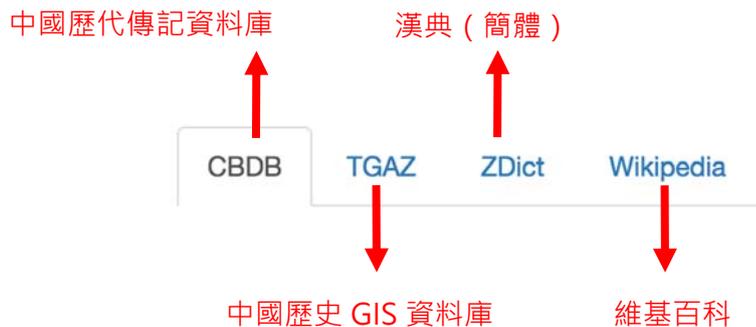
夾取的詞彙會放在「發現的詞彙」欄中，然後再將該欄中的人名放入「種子詞彙」，再重複執行「計算夾子」→「排除夾子」→「夾取詞彙」，直到使

用者挖掘到足夠的詞彙後，點選「標記已保存詞彙」就可以新增至姓名標記當中。

線上參考資料

MARKUS 提供了四種線上小工具，讓使用者可以透過網路搜尋文本詞彙。

- 若該詞彙已有標記，以滑鼠輕點一下，顯示編輯視窗。
- 點擊書本圖示，進行網路解說搜尋，數秒後右方視窗將會顯示相關資訊。



- CBDB (中國歷代傳記資料庫) 可查詢人名相關資訊
- TGAZ (中國歷史 GIS 資料庫) 可查詢地名在不同朝代的地理位置
- ZDict (漢典) 即漢語字詞典，可查詢對特定關鍵字的解釋
- Wikipedia (維基百科) 可查詢維基對此關鍵字的解釋

範例一：人物類

以「漢高祖」為例，點選「漢高祖」，按下搜尋圖示，CBDB 即顯示出漢高祖的詳細資料。(自動標記的姓名數字 ID，就是來自 CBDB 資料庫)

1 漢高祖 之起

2 姓名

3 CBDB

16622 ID

Search 16622

16622

CBDB ID: 16622
索引/中文/英文名稱: /劉邦/Liu Bang
指數年 (index year): -195
生年: 未詳
卒年: 未詳
享年: 53
郡望: 彭城
註: Liu(2) Bang [16622] Han Gaozu.

出處: 歷代帝王廟謚年諱譜:一卷, 頁0

別名: 廟號太祖, 未詳季, 謚號高皇帝, 未詳高祖。

· 點選 Wikipedia 則可以看到維基百科的說明。



· 除了在文本中點選標記，亦可直接在搜尋窗格打上關鍵字，enter 查詢



範例二：地名類

以益州為例，點選「益州」標記，按下搜尋圖示，發現 CBDB 並未顯示任何資訊，因為這並不是一個人名。點選 TGAZ，您會發現一個中國地圖，顯示出「益州」這個地名，不同年代在中國不同的地理位置。

1 **益州** 刺史种 [壽];

2

3 CBDB TGAZ ZDict Wikipedia

Search 益州

hvd_113523
益州 (Yi Zhou)
(184 ~ 263) [104.07800, 30.65039]

hvd_113524
益州 (Yi Zhou)
(264 ~ 606) [104.07800, 30.65039]

- 藍色地標代表「益州」在特定年代範圍的特定地理位置。
- 黃色圈圈中數字，代表在那一塊區域，有 15 處不同的地方被記載為益州。
- 點選藍色地標，即可看見地標 id 與年代範圍。

4

地標 id

hvd_113344

益州

433 ~ 441

年代

· 經過連續放大、點選黃色圈圈，如連續點選黃色 7，最後會出現 7 個不同地標。點選每個地標，即可分別顯示地名出現的對應年代。

1

6



綠色圈圈中數字，同樣代表在那一塊區域，記載為益州的地標數目。

每個地標顯示的 ID，未來將能在其他工具（如：DocuGIS 工具的地圖），做進一步的標記。

*操作影片：http://dh.chinese-empires.eu/MARKUS/beta/video_zhtw.html#onlineReferences

*欲了解如何管理個人線上參考資料，請參考操作說明影片：

http://dh.chinese-empires.eu/MARKUS/beta/video_zhtw.html#webReferences

輸出與保存

- 文本標示完成後，請至畫面左上角選擇「MARKUS 存檔」
- (文本名稱)_markus.html 檔將會下載至您的電腦。為了保障使用者辛苦的標記成果，建議使用者在標記的過程中能夠時常輸出存檔，以免因為意外而喪失標記的成果。



- 若您已經登入，點選「保存」按鈕，可保存您的工作進度。
- 看到存檔圖示顯示一個勾勾，表示保存完成



- 至網頁右上角，點選自己的帳號，按下「我的檔案」，進入您的線上個人檔案，即可看到您保存的所有工作進度
- 在 MARKUS 欄位下，點選「Auto」可進入自動標記頁面進行編輯；點選「Manu」則可進入手動標記頁面進行編輯；點選 Key 進入關鍵字標記。毋需下載再上傳 markus.html 檔案

Peiwen 1

我的檔案 2

登出

Upload a txt (UTF-8) or saved MARKUS file

Delete Learn Download

	File Name	dynasty	genre	MARKUS	VISUS	LEARN
<input type="checkbox"/>	KR2a0013_000.txt.html	undefined	undefined	Auto Manu Key		
<input type="checkbox"/>	KR3j0002_000.txt.html	undefined	undefined	Auto Manu Key		
<input type="checkbox"/>	KR3j0002_001.txt.html	undefined	undefined	Auto Manu Key		
<input type="checkbox"/>	WeiShuYi_DuanLaYi.txt.html	undefined	undefined	Auto Manu Key		

Showing 1 to 4 of 4 rows

MARKUS 標記完的輸出檔案，檔名為 (文本名稱)_markus.html，這個檔案將能在 Docusky 平台轉換為 DocuXml 格式，在您專屬的資料庫建立文獻集。