

文本批次標記工具
操作手冊
(含建庫檔管理工具：Tags)

撰稿人：洪一梅

撰稿時間：2020 年 8 月

目錄

壹、	工具網址.....	3
貳、	文本及標記詞彙表格式.....	3
	一、文本格式.....	3
	二、標記詞彙表規範格式.....	3
參、	操作程序說明.....	8
	一、讀入 DocuXML 檔案.....	9
	二、讀入 EXCEL 檔案.....	10
	三、設定標記參數.....	10
	四、額外輸出.....	11
	五、輸出 DocuXML 檔案.....	11
	六、標記後分類設定.....	12
肆、	匯出標籤詞彙表單.....	15
	一、匯出 TAGS 到 .CSV 檔.....	15
伍、	進階運用.....	16
	一、文本擷詞工具匯出之表單整理.....	16
	二、權威編號(REFID)的查找.....	16
陸、	練習用文本範例.....	16
	一、DocuXML.....	16
	二、UTF-8 純文字詞彙表.....	16

壹、 工具網址

<https://docusky.org.tw/DocuSky/docuTools/ContentTaggingTool/index.html>

工具頁面如圖 1 所示：



圖 1 文本批次標記首頁

貳、 文本及標記詞彙表格式

本工具目的是在一份文本資料上進行批次的詞彙標記。
使用者應備有一份文本資料及欲標記的詞彙表。

一、文本格式

文本資料應為 DocuXML 格式。

格式轉換請參考：

(一)「純文字文本轉換工具」

<https://docusky.org.tw/DocuSky/docuTools/ComposeThdlExportXml/composeSimpleThdlExportXml.html>

(二)「表格文本轉換工具(新版)」

<https://docusky.org.tw/DocuSky/devtools/CSVXLS2DocuXML/>

二、標記詞彙表規範格式

標記詞彙表應為 Excel 格式。

格式擬訂請參考：

(一)線上規範表

<https://reurl.cc/n0zm2l>



(二) 格式解說

1. Excel 表單

在 Excel 檔中，可有一個或多個詞彙表單(工作表)。

每一個詞彙表單(工作表)的第一列必為標記規範名稱；

第二列起為欲標記之資訊值。

每一個詞彙表單(工作表)的 A 欄必為「tagName」，B 欄必為「tagVal」，此二欄為必填欄。

如圖：

	A	B	C	D	E	F	G	H
1	tagName	tagVal	@Term	@RefId	@Aux	file:filename		
2	Udef_DiHao	元帝	漢元帝	dila_A006752	漢孝元皇帝	ZaYi_p0006ZaYi_p0023ZaYi_p0255		
3	Udef_DiHao	元帝	晉元帝	dila_A009575	晉元皇帝	ZaYi_p0157ZaYi_p0236ZaYi_p0313ZaYi_p0421		
4	Udef_DiHao	元帝	魏元帝	dila_A00646	魏元皇帝			
5	Udef_DiHao	太帝	宋太帝	dila_A001640	宋文武皇帝	ZaYi_p0388ZaYi_p0191		
6	Udef_DiHao	太帝	唐太帝	dila_A002414	唐文皇帝	ZaYi_p042ZaYi_p044ZaYi_p1108ZaYi_p1433ZaYi_p1433ZaYi_p0390		
7	Udef_DiHao	太帝	魏太帝	dila_A003321	魏明元皇帝	ZaYi_p0875		
8	Udef_DiHao	太帝	晉太帝	dila_A002777	晉懷文皇帝	ZaYi_p0836		
9	Udef_DiHao	太帝	金太帝	dila_A006629	金文宗皇帝	ZaYi_p3926		
10	Udef_DiHao	太帝	元太帝	dila_A006663	元英文皇帝			
11	Udef_DiHao	太祖	魏太祖	dila_A003207	魏武皇帝	ZaYi_p0874ZaYi_p0830ZaYi_p0911		
12	Udef_DiHao	文帝	晉懷文帝	dila_A002777	晉懷文皇帝			

2. 基本欄位

用以定義詞彙標籤及欲標記的文本中詞彙，為必填欄：

A 欄 - tagName

此欄填錄標籤名稱，

預設有：Office 職官、LocName 地名、PersonName 人名別名、Date 日期，共四個標籤，

可自訂標籤，自訂者加上前綴 Udef_。工具若判別該標籤非預設，將自動加上 Udef_ 的前綴。

B 欄 - tagVal

此欄填上欲標記為前欄標籤名稱的內容或詞彙。

該內容或詞彙用以比對文本中的內容或詞彙，完全相符即可標記成功。

範例如圖：

	A	B
1	tagName	tagVal
2	此欄填標籤名稱，預設有： Office 職官 LocName 地名 PersonName 人名別名 Date 日期 共四個標籤，可自訂標籤，自訂者加上前綴 Udef_。工具若判別該標籤非預設，將自動加上 Udef_ 的前綴。	此欄填上欲標記為 A 欄標籤名稱的內容或詞彙。
3	Office	巡撫
4	LocName	打狗
5	LocName	臺南
6	LocName	鹿寮
7	LocName	行天宮
8	LocName	橋
9	PersonName	太室
10	PersonName	唐三藏
11	PersonName	蘇奧
12	Date	光緒十二年
13	Date	六月初二日
14	Udef_Mainhole	蘇羅王

3. 聚合、消歧、參照欄位

用以設定標籤詞彙的權威及顯示功能，非必填，依使用者希望達成的目的選填：

Attribute:Term 或@Term - 此欄填上前欄內容或詞彙(tagVal)欲指定的權威詞。可指定，也可不填。如有指定權威詞，則系統會將 tagVal 的值都聚合為權威詞。

例如：tagVal 「唐三藏」「三藏」全部指定權威詞為「玄奘」，系統即會將文本中所有的唐三藏、三藏、玄奘等詞皆聚合為「玄奘」。

Attribute:RefID 或@RefID - 此欄填上 tagVal 欲參照的權威編碼，目前有：來自 CBDB 人名 cbdb_、TGAZ 地名 hvd_、TWGIS 地名 twgis_、法鼓規範庫 dila_ 等。也可直接放入座標，以「xy=經度,緯度」填入。可填，可不填。目前僅 PersonName 及 LocName 有此權威功能。如有填入權威編碼，系統會將 tagVal 的值聚合為該權威編碼，且可參照該權威編碼的內容資訊。

例如：tagVal 「打狗」，指定權威編號為「twgis_19291」，系統會將文本中的打狗聚合為「twgis_19291」，並可參照 twgis_19291 的地名權威資訊，也可在 GIS 上顯示打狗的坐標。

Attribute:PlaceRefId 或@PlaceRefId - 適用於其它非地名(tagName)的標籤，但兼含有地點的多重屬性。於此欄填上 tagVal 兼含的地點屬性的地名權威編碼。需於建庫檔管理工具中，將「類別/子類別」更動為「Place/原 tag 名稱」

例如：在 Office 的「巡撫」標上「hvd_32025」，於 GIS 上表示此「巡撫」為臺灣府的巡撫。於建庫檔管理工具中，將「類別/子類別」更動為：Place/Office。

Attribute:Aux 或@Aux - 用來進行游標顯示。於此欄填上游標至文本中

tagVal 的值時，欲指定顯示的資訊。例如：tagVal「唐太宗」，欲顯示為「太宗文皇帝」；例如：tagVal「光緒十二年」，欲顯示為「1886」範例如圖：

1	tagName	tagVal	@Term	@Refid	@PlaceRefid	@Aux
2	北屬填標前名稱，開始有： Office標首 LocName地名 PersonName人名標首 Date日期 再訂定標首，可自訂標首，自訂標首上標首Udef_，工具欄則用於標首管理，將自訂標首上Udef_的標首。	此欄填上欲標記為A屬標首名稱的內容或詞彙。	此欄填上標首內容或詞彙 (tagVal) 欲參照的標首詞，可指定，也可不填，標首名稱可為： "Attribute:Term" 或 "@Term"。	此欄填上tagVal欲參照的標首編號，如： 來自CDBB人名cddb_，TGAZ地名 hvd_，TWGIS地名twgis_，法經類編 dila_等，目前也可直接放入座標，以 "xy=經度,緯度" 填入。 可填，可不填。 標首名稱可為 "Attribute:RefID" 或 "@RefID"。 目前僅PersonName及LocName有此標首編號。	適用於其它非LocName的標首，需要標首地點，當座標在GIS上。 如：在Office的「座標」欄上 "hvd_32025"，於GIS上表示此「座標」為臺灣的座標，(用於管理工具中，將「類別/子類別」變動為 "Place/座標名稱") 如：Place/Office	此欄填上標首標記的標首詞彙時，欲顯示的屬性。例如：唐太宗，欲顯示為「太宗文皇帝」；例如：光緒十二年，欲顯示為「1886」。
3	Office	府署				
4	LocName	打狗		tagid_19281		
5	LocName	鹿港		ala_P100000025016		
6	LocName	鹿寮		hvd_78628		
7	LocName	打米港		xyz121.531713,25.0630098		
8	LocName	板橋		hvd_8565		
9	PersonName	太宗	唐太宗	cbdb_19360		太宗文皇帝
10	PersonName	蔡三發	蔡三發	ala_A000294		
11	PersonName	蔡三發	蔡三發			
12	Date	光緒十二年				1886
13	Date	光緒十二年				1886
14	Udef_MainRole	蔡三發	蔡三發			

4. 附加資訊欄位

用以設定標籤詞彙的附加資訊，非必填。依使用者希望附加的資訊性質選擇：

Attribute:Url 或 @Url - 此欄填上 tagVal 的值欲參考的資源網址。可填，可不填。例如：在 tagVal「蔡騫」的此欄填上蔡騫的維基百科網址 <https://zh.wikipedia.org/wiki/%E8%94%A1%E7%89%B5>。在 DocuSky 雲端資料庫中可以小視窗呼叫維基百科的資訊。

Attribute: IiifManifest 或 @IiifManifest - 於此欄中加上 tagVal 的值欲連結的圖像的 Iiif Manifest。例如：在 tagVal「長安」的此欄填上 <https://iiif.lib.harvard.edu/manifests/ids:3098346>。在 DocuSky 雲端資料庫中可以小視窗嵌入該圖象。

extraMetaTags - 在標記詞彙時，可自訂詞彙的附加屬性以利後續的運用。在該欄中可以填入使用者自定義的屬性類別（英文）與其內容（中英皆可），屬性類別可以有多個，彼此之間以半形「;」隔開。例如：在傳教士馬偕（George Leslie Mackay）的 extraMetaTags 中，可以加上所屬差會（Affiliation）、性別（Gender）、傳教區域（Location）、所屬宗派（Denomination）或國籍（Nation），所以，在該欄當中我們就可以填入：Affiliation: 加拿大長老會; Gender: 男性; Location: 臺灣淡水; Denomination: 長老會; Nation: 加拿大。填錄後，於 DocuSky 資料庫中即可出現 Affiliation、Gender、Location、Denomination、Nation 等附加屬性的後分類。

範例如圖：

tagName	tagVal	@Uri	@Manifest	extraMetaTags
此欄位標記名稱，預設為：Office職銜 LocName地址 PersonName人名 Date日期 此欄位標記，可由訂標者 訂定，工業級的訂標標 記符號，從自動加上 Udel訂標。	此欄位標記的 A欄位標記的 訂定。	此欄位標記的 A欄位標記的 訂定。	可於此欄中加上 tagVal的標記的 Manifest。	協助使用者在標記詞彙的同事增加詞彙的屬性以方便檢索。 在這種情況下，使用者自定義的屬性類別（英文）與內容（中文）可。 屬性類別可以有許多，彼此之間以半形「 」隔開。 例如：在標記詞彙（George Leslie Mackay）的extraMetaTags中，可以加上 Affiliation（性別）、Gender（性別）、Location（地點）、Denomination（宗 教）、Nation（國家）、Affiliation（性別）、Gender（性別）、Location（地 點）、Denomination（宗教）、Nation（國家）等。所以，在這種情況下我們 可以在多個人名上完成了extraMetaTags的標記，完成Content Tagging的 DocuSky標準標記會在TAG後分類區主或Affiliation、Gender、Location、 Denomination、Nation等新的TAG類別。我們可以在這些TAG類別中的 每個詞彙中，這個類別內會顯示出在標記詞彙中的標記。 同一個詞彙的TAG標記不同點在於，extraMetaTags的設計與，這一定 會出現在文本中（例如：加拿大、長老會等）。通過extraMetaTags的 設計，讓使用者可以從標記詞彙中，直接獲取DocuSky的標記後分 類功能掌握某些特定類別在文獻集中出現的情況。
Corpus	通志			
LocName	行旅			
LocName	通志			
LocName	通志	https://ref.lib.harvard.edu/manifests/id/0308246		Location: 通志
LocName	行旅			
LocName	通志			
PersonName	通志			
PersonName	通志	https://zh.wikipedia.org/wiki/%E5%85%B4%E5%85%B6%E5%85%B7%E5%85%B8		

5. 條件設定

用以設定在什麼狀況下，詞彙該如何標記。

就同一標記詞彙(tagVal)來說，條件之間是 and，也就是說當設有多個條件時，必須多個條件都符合才會進行標記或不標記。

filter: 詮釋資料欄位名稱或? 詮釋資料欄位名稱 -

利用詮釋資料進行篩選，在特定的詮釋資料值中進行標記。

篩選條件若為多值，則在欄中以半形「|」隔開。例如：f1|f2|f3|f4 可以比對到 f1 或 f2 或 f3 或 f4，符合此條件詮釋資料的文本，就會被標記起來；空值則表示不設條件。

例如：filter:corpus 內容「通志 88-200」，即是指定 corpus 名稱為「通志 88-200」的所有文件皆進行標記；filter:metatags/myTopic 內容「客家研究」，即是指定 metatags/myTopic 中有「客家研究」的所有文件才進行標記。

如需利用以詮釋資料反向篩選的方式，設定「!」在特定詮釋資料的值，即不進行標記(排除標記)。欄位若多值，以半形「|」隔開。因此，若「!f1|f2」就是指詮釋資料若為 f1 或 f2 即不標。

negativeLookbehind -

填入標記詞彙的前綴詞，在文本中有這些前綴詞的詞彙不進行標記。

例如：tagVal「永曆」，negativeLookbehind「吳|洪」，表示如果是吳永曆、洪永曆的這些詞中的「永曆」不標。

negativeLookahead -

填入標記詞彙的後綴詞，在文本中有這些後綴詞的詞彙不進行標記。

例如：tagVal「永曆」，negativeLookahead「元|二|三|四|」，表示如果是永曆元年、永曆二年、永曆三年、永曆四年...的這些詞中的「永曆」不標。

positiveLookbehind -

填入標記詞彙的前綴詞，在文本中有這些前綴詞的詞彙才進行標記。

例如：tagVal「永曆」，positiveLookbehind「明」，表示如果是明永曆這個詞中的「永曆」才標。

positiveLookahead -

填入標記詞彙的後綴詞，在文本中有這些後綴詞的詞彙才進行標記。

例如：tagVal「永曆」，**positiveLookahead**「年間」，表示如果是「永曆年間」這些詞中的「永曆」才標。

範例如圖：

TagName	tagVal	filterName	negativeLookbehind	negativeLookahead	positiveLookbehind	positiveLookahead
此欄填標記名稱，種類繁多		1. 專用此種標記詞彙於指定，並指定該標記詞的start與end進行標記。 2. 標記名稱可為「Altair」或「Altair」後綴詞標記名稱，而「」後綴詞標記名稱。				
Office類名		3. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
LastName類名		4. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
PersonName.姓名類名		5. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
Date類名		6. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		7. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		8. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		9. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		10. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		11. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		12. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		13. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		14. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		15. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		16. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		17. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		18. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		19. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		20. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		21. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		22. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		23. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		24. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		25. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		26. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		27. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		28. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		29. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		30. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		31. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		32. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		33. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		34. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		35. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		36. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		37. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		38. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		39. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		40. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		41. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		42. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		43. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		44. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		45. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		46. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		47. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		48. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		49. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				
此欄填標記詞彙		50. 標記名稱後綴詞，則在標記中指定「」。標記，則此「Altair」或「Altair」後綴詞標記名稱。				

參、 操作程序說明

「文本批次標記工具」是讓使用者可以一次性為文本進行多標籤詞彙的標記工作。

使用者應先備有一份 DocuXML 的檔案，以及一份符合規範格式的 Excel 標記詞彙表。

補充：DocuXML 的轉換工具，有

1. 純文字文本轉換工具

<https://docusky.org.tw/DocuSky/docuTools/ComposeThdlExportXml/composeSimpleThdlExportXml.html>

2. 表格文本轉換工具(新版)

<https://docusky.org.tw/DocuSky/devtools/CSVXLS2DocuXML/>

3. 碼庫思格式轉換工具

<https://docusky.org.tw/DocuSky/docuTools/MarkusConverter/Markus2DocuXml.html>

工具首頁：

二、讀入 Excel 檔案

步驟 2，載入一份依規範格式整理好的 Excel 檔案(標記詞彙表)。工具將列出檔案名稱、資料表名稱、資料筆數、不重覆詞彙數及資料表次序。檔案如不符合規範格式，將會提示錯誤。

操作如圖：



Excel 檔如圖：

	A	B	C	D	E	F	G	H
1	tagName	tagVal	@Term	@Refid	@Aux	filter:filename		
2	User_Duhao	元帝	漢元帝	dha_A000752	漢孝元皇帝	ZaYi_p0009ZaYi_p0023ZaYi_p0255		
3	User_Duhao	元帝	晉元帝	dha_A002675	晉元皇帝	ZaYi_p0167ZaYi_p0236ZaYi_p0333ZaYi_p0423		
4	User_Duhao	元帝	魏元帝	dha_A00646	魏元皇帝			
5	User_Duhao	太帝	宋太宗	dha_A001640	宋文皇帝	ZaYi_p078ZaYi_p0891		
6	User_Duhao	太帝	唐太宗	dha_A002434	唐文皇帝	ZaYi_p1043ZaYi_p1044ZaYi_p1106ZaYi_p1433ZaYi_p0380		
7	User_Duhao	太帝	魏太宗	dha_A003322	魏明元皇帝	ZaYi_p0875		
8	User_Duhao	太帝	晉太宗	dha_A002771	晉簡文皇帝	ZaYi_p0836		
9	User_Duhao	太帝	金太宗	dha_A009639	金文宗皇帝	ZaYi_p936		
10	User_Duhao	太帝	元太宗	dha_A009663	元文皇帝			
11	User_Duhao	太祖	魏太祖	dha_A003337	魏武皇帝	ZaYi_p0874ZaYi_p0830ZaYi_p0911		
12	User_Duhao	文帝	晉周文帝	dha_A002777	晉周文皇帝			
13	User_Duhao	文帝	魏文帝	dha_A002861	魏文皇帝			
14	User_Duhao	世宗	魏世宗	dha_A002044	宣武皇帝	ZaYi_p0913ZaYi_p0832ZaYi_p0839ZaYi_p0844ZaYi_p0900ZaYi_p0908ZaYi_p0912		
15	User_Duhao	世宗	魏世宗	dha_A002861	魏文皇帝	ZaYi_p0815ZaYi_p0760ZaYi_p0811ZaYi_p0843ZaYi_p0912		
16	User_Duhao	世宗	漢世宗	dha_A002860	漢光武帝	ZaYi_p0083		
17	User_Duhao	世宗	唐世宗	dha_A0023151	唐文皇帝	ZaYi_p0876		
18	User_Duhao	世宗	元世宗	dha_A002085	元文宗皇帝	ZaYi_p7470		
19	User_Duhao	安帝	漢安帝	dha_A006456	漢孝安皇帝	ZaYi_p0029ZaYi_p0037ZaYi_p0038ZaYi_p0070ZaYi_p0890ZaYi_p0110ZaYi_p0110ZaYi_p0221		
20	User_Duhao	安帝	晉安帝	dha_A002838	晉安皇帝	ZaYi_p0021		
21	User_Duhao	成帝	晉成帝	dha_A002514	晉成皇帝	ZaYi_p0168ZaYi_p0303		
22	User_Duhao	編後補	編後補	cbdb_20609	宋皇帝			

三、設定標記參數

步驟 3，設定參數。使用者可決定在標記前是否清除文本中的舊有標記，會清除的部份僅限詞彙表中欲標記的標籤種類，例如：在 Excel 中有 LocName 的標籤詞彙，即清除文本中舊有的 LocName 標記。

使用者若欲更新整份文本的標記，則建議設定為「是」，清除舊有標記；如只是想增加標籤詞彙，則建議設定為「否」，保留舊有標記。注意：選擇保留有時會造成重複標記的狀態。

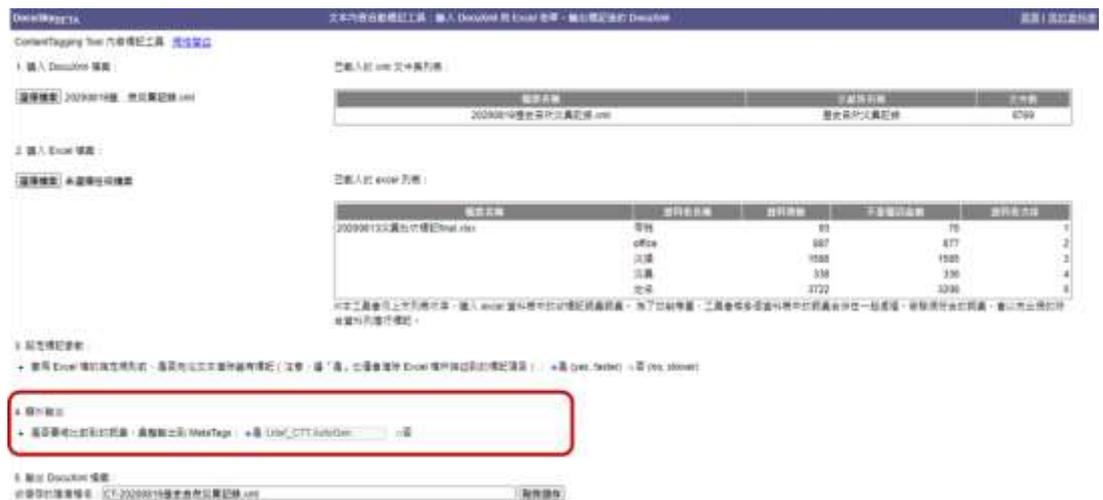
操作如圖：



四、額外輸出

步驟 4，將文本中比對到的詞彙，依文件彙整成為該文件的 metatags(多值的 metadata)。如選擇「是」，必須自訂輸入欄位名稱。

操作如圖：



五、輸出 DocuXML 檔案

步驟 5，完成批次標記，輸出完成檔。使用者可自訂欲儲存的建庫檔案。操作如圖：



六、標記後分類設定

步驟 6，將步驟 5 的完成檔，載入「建庫檔管理工具」中，進行建庫到 DocuSky 雲端資料庫後，顯示與功能的相關設定。
建庫檔管理工具，如圖：



(一) 自訂 tag 的顯示名稱

在建庫檔管理工具中的步驟 1► TABLE #3 中，可將原 DocuXML 檔中 tag 的英文編碼名稱，更改為自訂的顯示名稱。如有啟動 tag 的後分類，此自訂的顯示名稱即顯示為後分類的標題。

操作如圖：



DocuSky 雲端資料庫顯示如圖：



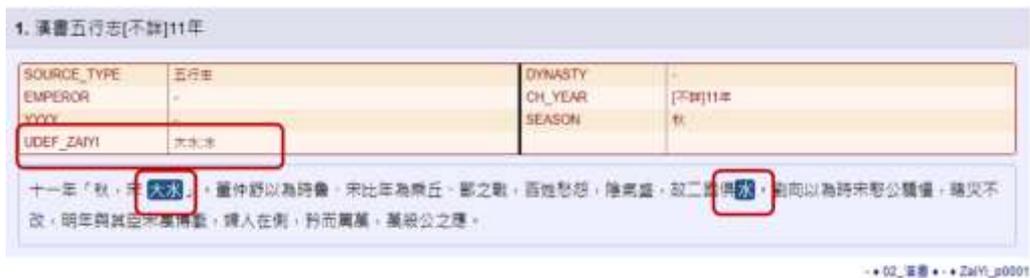
(二) 複製為 metadata 欄位資料

在建庫檔管理工具中的步驟 1► TABLE #3 中，可將原 DoucXML 檔中的某類標籤詞彙複製為 metadata 的自訂欄位。

例如：將 Udef_ZaiYi 複製為 metadata，則在 metadata 欄位中會多一欄「xml_metadata/Udef_ZaiYi」



DocuSky 雲端資料庫顯示如圖：



(三) 類別/子類別修改

於非地名標籤中使用了附加地名屬性(Attribute:PlaceRefId 或@PlaceRefId)，欲啟動 GIS 顯示，必須執行此設定。

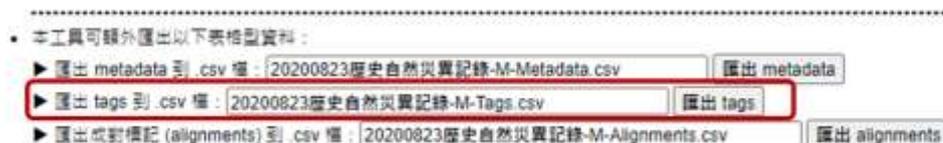
肆、 匯出標籤詞彙表單

已完成標記的 DocuXML 檔，可透過「建庫檔管理工具」，匯出檔案中的標籤詞彙，以備後續的研究使用。

一、匯出 tags 到 .csv 檔

在建庫檔管理工具中，載入已有標記的 DocuXML 檔，工具將匯出檔案中被標記的 filename 檔名、tagName 標籤名稱、tagVal 詞彙、Term 權威詞、RefId 權威編號。

操作如圖：



匯出檔案如圖：

	A	B	C	D	E
1	filename	tagName	tagVal	@RefId	@Term
2	ZaiYi_p0001	Udef_Date	十一年		
3	ZaiYi_p0001	Udef_ZaiYi	大水		雨災
4	ZaiYi_p0001	Udef_ZaiYi	水		水災
5	ZaiYi_p0002	Udef_Date	三年夏		
6	ZaiYi_p0002	LocName	夏	PL000000023563	夏朝
7	ZaiYi_p0002	LocName	漢中	hvd_115312	漢中郡
8	ZaiYi_p0002	Udef_ZaiYi	大水		雨災
9	ZaiYi_p0002	Udef_ZaiYi	水出		水災
10	ZaiYi_p0003	Udef_Date	四年秋		
11	ZaiYi_p0003	LocName	河南	PL000000023642	河南
12	ZaiYi_p0003	Udef_ZaiYi	大水		雨災
13	ZaiYi_p0003	Udef_ZaiYi	水		水災
14	ZaiYi_p0004	Udef_Date	八年夏		
15	ZaiYi_p0004	LocName	夏	PL000000023563	夏朝
16	ZaiYi_p0004	LocName	漢中	hvd_115312	漢中郡
17	ZaiYi_p0004	Udef_ZaiYi	水		水災
18	ZaiYi_p0004	LocName	沔水	PL000000026396	沔水
19	ZaiYi_p0005	Udef_DiHao	文帝	dila_A002777	晉簡文帝
20	ZaiYi_p0005	Udef_Date	三年秋		
21	ZaiYi_p0005	Udef_ZaiYi	大雨		雨災
22	ZaiYi_p0005	Udef_Date	三十五日		
23	ZaiYi_p0005	LocName	藍田	hvd_940	藍田
24	ZaiYi_p0005	Udef_ZaiYi	水出		水災

伍、 進階運用

一、文本擷詞工具匯出之表單整理

使用者可利用「文本擷詞工具(2020版)」找出研究需要的詞彙，匯出擷取到的詞彙，再將該詞彙表整理為符合規範格式的 Excel 標記詞彙表單。

文本擷詞工具(2020版)：

<https://docusky.org.tw/DocuSky/docuTools/TermClipper/TermClipper2020.html>

二、權威編號(RefId)的查找

使用者可利用「權威資訊查找工具」，批次找出特定人名、特定地名的權威編號，或是轉換中西曆日期。

權威資訊查找工具：<https://docusky.org.tw/DocuSky/DocuTools/RefLookupTools/>

(一) 中西曆轉換工具：藉由開源的 CeJS 來對「應存在的中西曆日期」進行轉換。

(二) 歷史人名查詢工具：可輸入歷史人名，輸出 CBDB/DILA 的人名 RefId。可單筆人名輸入，也可輸入一個.txt 檔(包含一批人名，一個人名一行)。

(三) 歷史地名查詢工具：可輸入歷史地名，輸出 TGAZ/DILA/TWGIS 的地名 RefId。可單筆地名輸入，也可輸入一個.txt 檔(包含一批地名，一個地名一行)。

陸、 練習用文本範例

下載網址：<https://reurl.cc/WLdZK9>

QRCode：



一、DocuXML

文本：歷史自然災異記錄.xml

二、UTF-8 純文字詞彙表

種子詞彙表：災異詞彙表.txt

