# 文本批次標記工具

# 操作手册

# (含建庫檔管理工具:Tags)

# 撰稿人:洪一梅

撰稿時間:2020年8月

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

<u>https://docusky.org.tw/DocuSky/docuTools/ContentTaggingTool/index.html</u> 工具頁面如圖1所示:

Doce Skynets	· 出土內田市田橋定工用 · 输入 Decored 书 Guest 茶杯 · 输出规定输出 Decorted	8218	IC REAL
Contentingping Test Alt BELLE MILEON			
7.进入Denatore 信載:	巨额人的 ow 次作表引用。		
	5658 2495	8 20 <i>8</i>	
2 80.000 88			
2992 +2999982	日期人知為自己的		
			_
	1270 STAIN 21100	*##22259 ####=R	
	REAR STATE STATE RELEASED STATE AND STATES AND STATES AND STATES REARINGS		
1024294		**##268 	
L NZARDA • Billion Briggathe - Banican		우리 전실 위해 위치 가지 않는 것이 있었다. 같은 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전	
i NTATON • Brion Gratator Bracts • Brian	100000000000000000000000000000000	≑≢Ш2344 ∦ПБ÷С 31294€955564≟#E—1666°+¥000925584 ¥€0520659	
<ul> <li>1 日本市政会会</li> <li>第二日本市政会社主任</li> <li>第二日本市政会社主任</li> <li>第二日本市政会社主任</li> <li>第二日本市政会社主任</li> </ul>	1         1	中華國政治的 왕태는 18 19월 23일 (김유王) ( 19월 23일 (김유王) (	
1 824298 • 25 Son State State • 57 Son State State • 57 State State • 58 Soon St	Image: State     Image: State     Image: State       Image: State     Image: State     Image: State     Image: State       Image: State     Image: State     Image: State     Image: State	우료법과의 최종국·대 일이 [14:53:17] (우리 2:30 이 수 있다. 17] (유지 17) ( ight ( ig	

#### 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」 <u>https://docusky.org.tw/DocuSky/docuTools/ComposeThdlExportXml/composeSi</u> <u>mpleThdlExportXml.html</u> (二)「表格文本轉換工具(新版)」 <u>https://docusky.org.tw/DocuSky/devtools/CSVXLS2DocuXML/</u>

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		C	0	1	1	á	H:
tagName	tagVal	@Term	@ Refld	@ Aux	filter:filename		0.0
Jost Dillao	元用	漢元帝	dila_A006752	漢孝元皇帝	ZarYi_p0006ZaiYi_p00222ZaiYi_p0055		
Udef_DiHao	元帝	普元帝	dila_A002675	晉元皇帝	ZaiYi p0167ZaiYi p0236ZaiYi p0313ZaiYi p03	7/ZaiYi p0423	
Udef_DiHao	元慶	魏元帝	dila_A00646	魏光泉帝			
User Dillan	大府	宋太宗	dfla_A001640	宋文武皇帝	ZaiYi_p0768ZaiYi_p6191		
Udef DiHao	太常	唐太宗	dila_A002414	微文集帶	ZarYi p10432arYi p10442arYi p11082arYi p14	9ZaiYi_p1433Z	ai'ii 16350
Udef_DiHao	太別	魏太宗	dila_A003322	魏明元最帝	ZaiYi_p0875		
Gdef_DiHao	太洲	肾大宗	dila_A002771	普爾文皇帝	ZarYi_p0836		
Udef Dillao	大学	金太密	dila_A006629	金文亦皇帝	ZaYi p3606		
Udef_Dillan	大宗	元太宗	dila_A006663	元英文皇帝			
1 Goof InHao	太祖	魏太祖	dia_A003207	教武皇帝	ZeiYi p087422eiYi p08302asYi p0911		
2 Usef DiHao	文帝	普佩文明	dila_A002771	普種文皇帝			

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 人名别名、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:

4	*	8
1	tagName	tagVal
2	比機填該標碼名稱。預設有: Office職會 LocName地名 PersonName人名別名 Date日開 共四個標範,可自訂標籤,自 訂者加上前網Udef_、工具若 到別該標業非預設,將自動加 上Udef_的前缀。	此爾埃上該情記為A獨優毀名稱 的內容或詞彙。
1	Office	//////////////////////////////////////
4	LocName	打狗
5	LocName	進時
E	LocName	廣安
7	LocName	行天面
	LocName	10
-	PersonName	太貴
10	PersonName	唐三城
11	PersonName	装飾
12	Date	光譜十三年
13	Date	大月初二日
14	Udef_MainRole	集服王

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	*	· · · · · · · · · · · · · · · · · · ·	1	-*	and the second s	1. AL
	tagName	tagVal	@Term	@Refid	@PlaceRefid	@Aux
	記覺清却得到三編,預給表: Office得到 LocNaem的考 PersonName人名別名 DateO時 科互提供書:司在記標畫:面 訂書加上用現込前,:工用單 升別加得量が現象:馬音動加 上Udef,的前書。	兵環境上37億紀為A間標業名称 和內容成詞集 -	<b>4.標準上京機力回或設施</b> (TagVa)設勢預防機度同。 可応定、也可不循・概位名 緒可為 <sup>7</sup> Attribute:Term。道 <sup>7</sup> ØTasm。-	式環境上tagVal設容預防提数條碼,但; 配合EDDB人名bobb,-TGAZ地名 Ind,-TWGIS地名bugis,-這是規範構 個礼,雪,有品信包互指控人這種,以 「xy=程度」等量。頃人, 可准,一 電台名稱可為"AttributeRefID」還 "保險fD」, 訂點等PersonName.Biochame有点響 以完。	御用於其它非LocName的模量 , 南雪模示地點, 首呈现在GS 上。 司:在Office的"邮册, 健上 "httl,32025, . 16GS上景示 此"通路,酒酒和我的能量。 (周於智慧工程中,若"就似子 帮你,了整路"Place/深tag毛 有;) 目:Place/Office	<b>此標項上賞設信款於該標定</b> 與樂祥·欲難示的羅性·何 間:讓太魚·欲願示私「太 筆文單雪」,供如:尤是十 二年,欲動市為「1886」
1	Offke	(E)#			host \$2025	
1	Lochame	2294	in the second	tagia_19281		
4	Lochtamo	(41)	和目	dia_PL00000025018		
6	Linchiame	<b>希室</b> : 2	長家	hwd_70628		
1	LOCHARME	的天路	自大理	ry=121.531713,25.0630698		
	Lauchiator	41	机性	hvd_3565		
	PersonName	太法	质太常	cbdb_13040		大氣文皇海
15	Personitiame	(1) [ 4]	<u>東壁</u>	dia_A000284		Substance .
- 12	Personhiame	11日 -	有意			
12	Date	水園十二日				1894
-10	Date	ARREA				1888
34	Udet_MainNoie	##T	斯相望。			

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 選填:

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: liifManifest 或@liifManifest - 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 Lesile 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:

	A.	10		100	
1	tagName	tagVal	eun	@hifManifest	extraMetaTags
	定電磁路構築式場、預販 有: Office構成 LocNameを表 PersonName人型取得 DateII 用 力量で加速 中可確定上に導 Uset、工程研究用 に 時間、資産取加止 Uset、工程研算	広餐後上記標記為 A餐燈第64條約内 容式開業。	走欄墳土tagVal 敬参考的算道機 社・環境・現不 頃,	可於此欄中版上 tagVal認識話的書 傳的IIF Manifest,	編動使用率在標記詞論的影響体與時間層性以积聚團的獲得。 在近欄中可以進入使用者自定義的醫師與不可與其外容(中定則可)、醫性類別可以而多個,要此之間以早有";」編 個,物理、有種物土類個(George Losile Maskay)的使力aMetaTags中,可以加上與醫療會(Atflation),也知( Georder )、佛愛加其傷(George Losile Maskay)的使力aMetaTags中,可以加上與醫療會(Atflation),也知( Georder )、佛愛加其傷(George Losile Maskay)的使力aMetaTags中,可以加上與醫療。(Atflation),也知 Georder )、佛愛加其他(George Losile Cheronination ) 馬爾爾(Attinution),不能,在古國中使制動可以加 A:Atflation 加拿大賞老像、Georder: 開始, Location: 童蜜酒水, Denomination: 長寒會、Nation: 加拿大 是影響力的一個意味了使用Attinution。「新聞要求」使用Attinution }。 Atflation。George Common Nation 醫療的社会認識。他們以及當不可做影响的思想是出設計使 整要文化者中,這種類的含量性生活在豐厚不同意是能的資源。 ——包訓論的的名信是中不問題的生物。etaMetaTags的資源的資源。 ——包訓論的「Atflation Oneonination Nation醫師的情況」 ——包訓論的合為是性生活在豐厚不同意思想的情況。 和此一個一個一個一個一個一個一個一個一個一個一個一個一個一個一個一個一個一個一個
1	Office	18 H			
4	Littlare .	灯中			
	tornane .	NR .			
	Lochame	RR		https://iif.lib.harva nd.edu/manifests/iid v.3098546	
	Abilitare .	对关系	1		Loistins 🗄 🗓
	Loutsete	18			
	PersonNare	2.8			
-	Ferderitate	818	and an and a second		
	Persondhame	**	https://zh.wikipe dia.org/wiki/%28 %84%x1%27%89 %81		

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:

	8-			A		. 44	
	tagName	tagval	fiterfilename	negativeLookbehind	negativeLookaheed	positiveLookbehind	positiveLookahead
	上規模的学校では、20 名号 2015年2月日 2015年2月日 2015年2月日 2015年2月 2015年2月 2015年2月 2015年2月 2015年2月 2015年2月 2015年3月 2015 2015 2015 2015 2015 2015 2015 2015 2015	出環境上記標記 現人養年勤を取出 八章石刻集。	1.5月に経営事業構築的です。日本管理構築の約34444000年間の構築の構築で、 使生産物であった時に、日本の生産業務の構成の、日本の経営が増生る あ。 当該当由の生活の含素、創造者やいため、11、満年、自然・12(0-13)に ないたまして、12(0-13)に、22(0-13)に、23(0-13)に などの支払いため、 当業業務内理量がない。 13(14)に15(14)になった場合、13(14)にのようなない。 13(14)に15(14)にのより、25(14)には、13(14)にのような 13(14)に15(14)にのより、25(14)には、13(14)にのよう 13(14)に15(14)にのよう、25(14)には、13(14)にのよう 13(14)に15(14)にのよう 13(14)に15(14)には、13(14)には、13(14)には、13(14) 13(14)に15(14)には、13(14)には、13(14)には、13(14) 13(14)には、13(14)には、13(14)には、13(14)には、13(14) 13(14)には、13(14)には、13(14)には、13(14)には、13(14) 13(14)には、13(14)には、13(14)には、13(14)には、13(14) 13(14)には、13(14)には、13(14)には、13(14)には、13(14)には、13(14) 13(14)には、13(14)には、13(14)には、13(14)には、13(14) 13(14)には、13(14)には、13(14)(14)には、13(14)(14) 13(14)には、13(14)(14)(14)(14)(14)(14)(14)(14)(14)(14)	▲人立時記試員打古將將・陸道総 起來的[如果不是小傳經一 而近 1000/01 月前。 1000/01 月前。 1000/01 月前。 1000/01 月前。 1000/01 月前。 1000/01 月前。 1000/01 月前。 1000/01 月前。 1000/01 月前 1000/01 月前 1000/01 1000/0000 1000/00000 1000/00000000	は、ご見た日点の法が高・有価 高小規則で利益不満に予想 にない、予想に予想 にない、予想に「二」 に、大力でなるまで、二」 に、大力でなられませいからは 第二、大力であるまで、 第二、大力であるまで、 第二、大力であるまで、 第二、大力であるまで、 第二、「一」 「一」 「一」 「一」 「一」 「一」 「一」 「一」 「一」 「一」	■人口得受过美加加非联,通 但在和我所们是美工业们号 4、你们工业的化学,是是,, 可以生美机会是基础的平时 「永喜」、才等。	<ul> <li>(二号之対点)TBATH</li> <li>第二号 (二号之対点)TBATH</li> <li>第二号 (二)(1)(2)</li> <li>第二号(二)(1)(2)</li> <li>第二号(二)(1)(2)(2)</li> <li>第二号(二)(1)(2)(2)</li> <li>2)(2)(2)(2)</li> <li>2)(2)(2)(2)(2)</li></ul>
	Mita	対策	060_00011060_0000		-		
+1	-pullares	178					() () () () () () () () () () () () () (
1.1	artierie	NCH.	P				
.1	nitiene	**				-	
-1	pillerie	10.00					
	Lifef Office	1 <b>1</b> 1		3.0	e can	in .	72
+ 1	do filement	HE .					
	a contine to	2.0	085_p0001008_p001	1	S	<u></u>	2
	Amerilante	増える			-		
		-			1		
184	25W	主要十二世			3		
34	Lang	<b>在時間目標</b>		2		5	
	del Weinflum	清朝王					

# **参**、 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:

Decusity at 11	22/04/19/04/12/1 BADISAN FEMALE	新 - Weiter(Ell)(160) Decisions	83180
ContentTagging Test (1)###222.8			
1 BA Seculari SE	○第八前 == ○平長方高		
(()))) - 2791( +10.2	9,5,69	22548	200
D BLA GOM BE			
2842 +2840az	EmAn and Ref.		
	電流工程 (月18月2日) 日本工程を行うためがある (月18日日) 日本1月1日日日 (日本日本日本日本日本日本日本日本日本日本日本日本日本日本日本日本日本日		A BEREN BREAN TERAPRO-LEU - FURTHERA - PARINER
1 K2862#		2017# #224.874 - 972248 - 2846-5248-	×888488 5000000 1934 - 844 - 840 - 890 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 8 1935 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880 - 880
i Gongen + Wilcon Workship - Gönurski	NEXT         REAL           NEXT         REAL         RE	antina Millaita - n7mmta - mandatska S (ve. heler) - 00 (se. slove)	8888488 596538 1934 - 1945 - 1949 - 1959 - 1959 - 1959 1934 - 1945 - 1949 - 1959 - 1959 - 1959 - 1959 - 1959 - 1959 - 1959 - 195 1959 - 1950 - 1959 - 1959 - 1959 - 1959 - 1959 - 1959 - 1959 - 1959 - 1959 - 195
1 128628 • #46con96639855 - 84502286	20223 201223 2022490.201002 20 4000 201200 201700200 20170020 20170020 20170020	enites HELLAIA - 57mpti : IAndsoyaa (vs. hater) () D (vs. sinve)	8888888 596538 1934 - 1945 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997
n Konkove + William Monaghta - Báthacove - Brien + Báthacetarida - Banach Venter	12225 2000 12280229023 8/ 400 20840 2078292: 2078292: 2078292: 2070202000000000000000000000000000000	BUILE BUILE - N7207B - IAROSUUS Buile Autori OB (n. sinvel	⊼882,428 उत्तरत्व 1728,499,4543,470,779,728 €.423,7759
5 私立地石字紙 ・ 第三 Cool 第二第三第三第二第二第二第二第二第一 ・ 第二第二章 ・ 第二章 第二章 5 第二 DooJon 第二章	20238 2000 0422890_000023 20 400 208400 242802020 242802020 24200_007Amodes 0	BUTE BERARA - N7205B - IARGESUS (vs. hele) OS (n. devel	ABBRIAN SPREAM YEAAAAA KULUU HUMBARKA €24539HB

一、讀入 DocuXML 檔案

步驟1,載入一份 DocuXML 檔案。工具將列出檔名、文獻集名稱、及內含 文件數。

操作如圖:

Doos Sky Jic In	247/RESELLA . MA Devote R Loss 24 . MURCHE Douber		AR I BOAS
Conserfagging has the GELLE MARKED			
1 B.A. Decolute WE	三年人的144 支付書刊書		
(夏季·唐葉) 200001-9世 - 第33第22時 - mp	电压力电 202000时间提出数约3.属红数 HH	Bannada Bannada	81418 8799
2 B). Exist WR			
	En.All-oos hit		
		TANIS DI MANAGANA - LET - NA	HILLS AMARINA - BURLANAW
) UZSER • BREver MILEZSER - BERLIZZBYRS	E ( S.W. & " & . S & B.W. Door BY BIGLE WERE ) . • & (ee, tester) O E (ee, sever		
4. WP-MI2			
· SIRGLERICA ASSERTATE	Uder_CTT.Aukodee		
5. Brz Doceter S.R.			
の日本に注意場合、これの日本に基定曲片文集記録 いた	(Attan)		
A second s			

DocuXML 檔如圖:



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:

locu fix y <sub>MCDA</sub>	スキカ谷田田橋巨工県 ( BA Docome R Doce R F A B REE)	BE Delation			BRI BRIA
Centern/Tagging Test 作會應任工具 通信協会					
B. Denter SE	王教人的 em 全中最利用				
	5718		745.58	-	298
	2000000年度世界的12年度19900005		ERRARAEN		4799
H. L. Basel W.R.					
THER ADDITION	EttiAtt even fott -				
		RHAR			87875
	20200013.2.進行世間的ind vite	衛務	80	78	1
		office .	007	877	-
		1.8	338	236	
		地域	1722	3210	
	六年三萬會位上方打裝片等,圖入 even 算好基本的計算經過範疇 由單句到還行還起。	**724##	- 三基単信券接著に基中的結果由中生一般素質	·UNIXVALUA	s-nutrie
STREPT .					6
· ** Los Britzer:	an i feith an " R	e off reason	41		
Miniate:					

## Excel 檔如圖:

1	A	*	C	D	E	₹ G	
1	tagName	tagVal	@Term	@Refid	@Aux	filter:filename	
2	Odef_DiHao	元産	演元帝	dia_A006752	漢字元皇帝	ZaiYi, g0006ZaiYi p0023ZaiYi g0255	
3	Uder_DiHao	発揮	實元章	dila_A002675	晋元莫等	2mYi.p01672mYi.p02362mYi.p681322mYi.p03332mYi.p0423	
4	Oder_Dillao	光律	親元帝	dila_A00646	魏元皇帝		
5	Uder_DiHao	太清	宋太帝	dda_A001640	宋文武皇甫	ZaiYi, p0708ZaiYi, p6191	
÷	Uder DiHoo	大定	慶太洋	dila_A002414	康文皇帝	ZaiYi g30432aiYi p10442aiYi g10802aiYi g543002aiYi p143302aiYi	1. p6355
7	Uder DiHao	太津	魏太宗	dila ,0003322	魏明元法律	ZmY) g8875	
8	Met_DiHu)	太常	增太宗	dilu_A002777	世间文量世	ZarVi 20036	
4	Mer DiHae	太理	重大定	dila_A006529	金文別量單	ZaiYi 93636	
10	Uder DiHae	太准	元志宗	dila A00963	元英文皇帝		
11	Uder DiHuo	太田	魏太祖	dila: A003207	魏武弘帝	ZaiYi 2087472iYi 108302aiYi 20913	
12	Mer DiHat	2.9	皆開文市	dita A002777	普爾文基堂		
19	lider_DiHae	太平	観文帝	dElla_A000361	載文皇室		
14	Uder_D(Hao	枕窪	韓武宗.	dia_A002044	宝武虽变	ZaiYi 10803ZaiYi 10833ZaiYi 10859ZaiYi 10884ZaiYi 10800ZaYi	p0908Za
15	Uder DiHao	世祖	增加过过	dila_A002961	魏文皇帝	2mYi 20002aiYi 20002aiYi 200002aiYi 200002aiYi 20002	
16	Uder_DiHac	開虹	邁壯祖	dila_A002860	满龙武帝	ZaiYi_p0083	
17	Uler Dillie	世初	<b>带世祖</b>	dila_A023151	使尤星常	ZaYi 30876	
18	Gdet_DhRao	世祖	元與祖	dila_A002085	元文武益帝	ZarYi 27470	
19	lder_DiHas	安川	漢安帝	dia_A006455	演家安息晋	ZaiYu 20029ZaiYi p0037ZaiYi p0038ZaiYi p0074ZaiYi p0089ZaiYi	i_p01107a
20	Eder_DiFlace	安治	置安華	dilu_A002828	晋安县等	ZaiYi_p021	
21	Uder_DiHat	夜夜	青成帝	dila_A002514	暂成基章	ZaiYi_p0168ZaiYi_p0203	
22	Gdef InHao	新使用	后花桃	izhdb 20609	大泉斎	A STATE AND AND A STATE AND A STAT	

## 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:

DOCUSRYACTA	2*ABHERELER . BADWARE REAW EV	- MIDIELECT Decement			R.K. Mr.
ContentTagging Son THE RECT MILLION					
B.A. Denation B.B.	世歌人们 and 文:本義的音				
	8228			6 <b>8</b>	248
	202000118年生会社工具	Eiter	観察者が注	構築が	8768
建人主worl 编篇:					
AND ADDEND	世間人的 exten 打得				
	AREA		AT IN A CONTRACT OF	4.8522,88	BRIATIL
	2828年6月2日萬行河增肥加4186	<b>E</b> H	23	78	,
	Property Construction and the	infea	887	877	2
		208	1588	1568	
		324	3.58	3.08	
	以下工具を注土力引用元章・塗人 annu 算り用中的公式 由者の力値で使用。	IERARA N711148 IA	HESERNMONTAINE	-EES WSW74515A	*
	11年三員會注土有計算改革,建人 annu 建石质中的公司 業務約增予確認。	REMARKA NYUHNA IKI	NES-EXHECTERANC		-

## 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:

DOGRAMMENTS.	2478686ULA BA Down R tool BE-	MILES Dealer			REAL PROPERTY.
ContentTagging The (108E1A 111EE					
) 購入 Demulters 運業	普戴人紀 ame 艾卡莫拉哥				
· · · · · · · · · · · · · · · · · · ·	80048 2000-98 # 3 P.1	A2.9.ml	0.415 #1:520	anii Ace	150 CW
1 BA Ever WB					
	包裹人的 ever 形態				
	62.5 <b>8</b>	SHELK	a iline	Fatton	2868.74
	2009001333進出寸確Efmal.sht	<b>9</b> 11	10	10	1
		office	887	417	2
		(A)#	1588	1505	3
			3.58	2,90	
	に本工具要作上で引用で第一級人 avoir 営ん用やけいで を営い引きて用け、	BEMANA N7DAWN - IAB	SSEANNELERANCE-	-DEG - BURHACH	A BUREROW
1024094					
all - avander Statesaland	CARE OF THE CARE AND POST OF THE POST OF T	the ball of the states			
· Buttle strategies - sectors and the	C(AV)W R. CURRY DOD BUILDING REF.	Chart-Dattach is an International			
A BORD		1			
+ REPORTERIES AND REPORT +8	ung strander 👘 📲				
	Record				

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:

BACKTRATER TA	TENRORELLE B. Instruction St.	- MARINE AN INCOME			SRI STO
Content Tagging Test 7 # #1212 A 1992					
E.1. Deculted WE	三角八川 200 三十長万角				
■用葉葉 20200111度 M32用記録 1000	202000 - 20 C	AC8 ===	310		3118 8766
1 種人 Record 種類					
	ER. <sup>1</sup> , C Annal 7(R)				
	67.44		00.00	Automatical (	
	2020081333高行中第EPeet Alex	带线	. 90	18	,
		1908	907	877	2
		100	1958	1988	
	and the based on the state of the state of the state	2.6	1722	3208	
	※二三集會任上市民務計算、第人 mon 第只要未找 会員以前會行業長。	VELIGATIAN N711498-IAB	CPURLEY, DANY	★-出意味: 副製用目前代的	A BLACKER
<b>私世界已来来</b>					
	HARREN THE REAL PROPERTY AND REPORTED BEARING	a toni tatleri -E ma itaneti			
and an international states in the second se	- Mark Marken and				
STRUCTURE RESILT OF A	· · · · · · · · · · · · · · · · · · ·				
R T CHIER BR					
the second					

# 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:

	出口四可可用用 TETE KAA. M M M M P I Particulate 用			- KR   MC#1
事業的叫:主工具可提上一台 Depuikes 谜:伊里马希诺拉家	WRAI metadata 電気計算法制刷用可用口、形体解出系一	- Data Ref 1		
1 昭二 編人 Google (田田)				
ATTA MARS ADDINES				
<ul> <li>● 80 Copus(10)</li> <li>* 2008 - 60000</li> <li>* 2008 - 60000</li> <li>* 200000</li> <li>* 200000</li> <li>* 200000</li> <li>* 200000</li> </ul>				
▶ TABLE #1. 空観人的 moladana (国動名): (文書編):				
STR. 2.2 Treasure BE	(). (). (). (). (). (). (). (). (). ().		素な影響にお	
<ul> <li>DRZPANBIEC (RUTHADDRAGE)</li> <li>THELE RUTHER (REALING AUGUST AUGUST</li></ul>	Faring) kaland i Kalingtoka kesaing birasi-la	(8+a)		
2012-7488/00-0 (017-852)/7482 (0)     1022-01-2171 (0424-01-4)/2008/7212 (01     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     1022-01-2171 (0424-01-4)/20     102	Andres Harabart I Franklight Mar Armalia Archadi - 2 History Andres Harabart Constantion Andres Harabart Constantion	(k+a) States Constant BARL House B2	autoriant.	97479+8
construction of the second secon	Танбан нарадары і падарацы аналуа Арлаал-2 Паланда Аларианана анд Барадарац состоральная Ард улард — арасала ала алулаа	Land Land Land Land Land Land Land Land	a-inatiat	
A TAR A STATE OF THE ADDRESS AND ADDR	анскига с нарадни с на акциански в нарадин народ на 10—3 Окончул — Зама с нарадин на нарад на на нарадина с простор на прила на нарадина с прила на нарадина с прила на прила с прила на прила на прила с прила на п	(k+a) 1 ottos+kastela 1788: Helende SII - Stoartela	10000000000000000000000000000000000000	RZARTENE RATE – EREINARI REFE GAMME BURK FARTAN
THE 2 PARTIES CONTINUES OF THE PARTIES OF THE	анскиги с марались I на а средстви — анализи и Арстании — а Попоница на се и полосий Кам со со со со со со на се и полосий Кам со со со со со со со на со	(k+a) - 	10000000000000000000000000000000000000	SZBZYBÓB 1975 – SZBZINBE 1975 – Landon I. H.; Kin & AM.Y. († 1
THE PARTIES (IN TRADITIONS) 10     THE REPORT OF TRADITION OF TRADITION OF TRADITION     THE PARTY OF TRADITION OF TRADITICO OF TR	анский какаласті на абратива жана (какаласта) Папачар Запачала какала анда какаларана собарана арда какаларана собарана арда жана арда аластар аласта	(k+a) 1 01105+HARTHE 81881 Helende SIZ -020047488	enserat Recent	SZBATENE MAR – SZELINE MAR – SZELINE MAR formati filmene Mar formati filmene
2012-2018 18-4 (11-7-825/17-851-5)     10212-21 (112-96/18-94/18-26/18-18-18-     10212-21 (112-96/18-18-18-18-     10212-21 (112-96/18-18-18-18-     10212-21 (112-96/18-18-18-      10212-21 (112-96/18-18-18-     10212-21 (112-96/18-18-18-     10212-21 (112-96/18-18-18-     10212-21 (112-96/18-18-18-     10212-21 (112-96/18-18-18-     10212-21 (112-96/18-18-18-     10212-21 (112-96/18-18-18-     10212-21 (112-96/18-18-18-     10212-21 (112-96/18-18-18-18-     10212-21 (112-96/18-18-18-18-18-18-18-18-18-18-18-18-18-1	ански, и каралис; ; на акциански к-налин и кастенит-а Папана, 2004 акциона и иза на и какцатаки состата - алики Какции - история и и акции и или-или и и и или или-или	(k+a) = ==================================	AN AND AN	SZARTANA MARA – ERELINAR ERE forozoni filmén filmente
2227222222222222222222222222222222	аланы наланы: практарына анылуу бургалт-а ала изана аны аныланан ана изана ана старыта-анаан ана изана ана старыта-анаан ана алананан ана алананан ана алананан ана алананан ана алананан ана алананан ана алананан ана алананан алананан алананан алананан алананан алананан алананан алананан алананан алананан алананан алананан алананан алананан алананан алананан алан ала	(£+2) 		LESSON STATE

(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 DoucXML 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:

TABLE OF TAG ANALYSIS THE HERBERT	1811 / Log in the Internation	14 2742118-8040181281 mile	將tag编碼 更改為自訂 的顯示標籤	
AND THE OWNER	LT.	MIN-FIELD	and the second se	TACK - URBANES
1.Unit.dem	<ul> <li>Unit_Zx</li> </ul>	m- (3.8	1.	
1 Unit Dition	· Unit Di	Had-	Tanks	
3 Lochane	Page-	116	74/81	~
4.0808	Official	81	Tatkit	*
S.Udet_stamage	<ul> <li>UNL</li> </ul>	trappi (24)	Taribit	*
A Low Comprohi	<ul> <li>Ues_D</li> </ul>	Augory#14 DLBA	Tatka	
7 unet_Canegori/0	<ul> <li>Usef_Cu</li> </ul>	Jagory 22-	E Farmer	*
	10000	And and a second s		

DocuSky 雲端資料庫顯示如圖:

DocuSky <sub>DCTA</sub>		RAINDANA (BA
**** *********************************	NANA ANA ANA ANA ANA ANA ANA ANA ANA AN	
	Guary: all           Manches: 6760           1. 第第五/2回(不知)14年	
	→二三「秋・末 (第二), ●第三部以為許會、市比市為末日、新之配 は、前日前減益市業構成、第八日前、日可業業、解約公之等。	5555-888-51890-8518479288-827
日本市の 日、市営へ株工業と成果と基準に行用、1942、 工業株式工業の本社に可の使いたのが 私に大量に内になりためたまた。 工業株式社会社に支援の業法、 工業株式社会社、工業業業業業業業	2. ##5/76(7#3# #85#80 · (20) · #6 (20) (20=5+## ·	
<ul> <li>BQ:</li> <li>□ T.Y. (± n.m. ski, ski, *E, *ki ×ki 1333</li> <li>2.6.0* sli ti ski ski ski ski ski 1341</li> <li>□ S.Y. (Ξ k. * ski ski ski ski 1341</li> </ul>	2. 東慶五行町(不詳)4年	-+ 0, # <b>2</b> + + Det, #00

(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」

W. 116			· · · · · · · · · · · · · · · · · · ·	- 伊斯内西 其原来)	1.00
1 User_ZaW	5	URM_ZWVI-	1.8	(monitority zer	
1. Udet DiFlat	0.0	Uper Cirkan-	710	THICKIT.	
3 Lochlame	5.	Place!-	e4	Fanan	
4.000		Cfficer	81	Ferfan	
6. Liter darage		Liner damage-	18	Tarikr	
6. Liber Committe		limit Calegraphic	1944	Tarter	
1 littlet ( discount)		light Calmer City	10000	TAUNU	
T. Ober, Camporna		Contraction of the second seco		- THURLE	
a user_uara		neet neet-	1	A TRUNCT	
AL BERTSARE IN	72.815	##C71-#25#	A1144.000	##())(=##	
		##UCTIFESTS #	ALGORITH .	46118-841	
A rape	72335		ALTR	a a 111 - 4 4 1	
AL BARY ARE (199) A Depen	72835 1248	##10-434	8178	4611(=63)	
AL BLANDARD ACTION	72815 1524 1249048	4803-404	AILTR +1 (#178 -1 (#178 -1	400(#41	
ELECTRACE (111)      ELECTRACE (111)      ELECTRACE (111)      CONSTRUCT (111)      CONS	TERES ETHE ETHERES ETHERES	(101612-117)	8178	44)))(=34))	
Capet C. RCH (111)     Capet C. RCH (111	72815 1228 1288 1286248 228822549 229822549 229822549	(स्वार्टास्ट न) ((स्वार्टास्ट) सं)	8178	##(I)(i):=####	
(d) - Saletta Acte (170) 	728418 21248 21248248 21248248 2124828248 21248282 21248282	1 ( 100 k 44) 1 ( 100 k 44) 1 (	8118	a 6,111) - =8108(1	
24.         Status Actin (mm)           1000         Status Actin (mm)           1         Status Actin (mm)           1         Status Actin (mm)           2         Status Actin (mm)           4         Status Actin (mm)           4         Status Actin (mm)           4         Status Actin (mm)           5         Status Actin (mm)           6         Status Actin (mm)           7         Status Actin (mm)           7         Status Actin (mm)           7         Status Actin (mm)	72815 1258 124624 224624 224624 2258 2258 8428 1258 828 1258 828 1258 828 1258 828 1258 828 1258 828 1258 125	- (#1012-0)# (#1012-0) #) #)	101000000 10170	a 40/11(i = 40 <b>4</b> [1	
24.         Calentin, Archin (1991)           Calentin, Archin (1991)           Calentin, Service	72/815 122-8 120-8224 220-8224 220-8224 122-8182 122-8182 122-8182 122-81 122-8		8178	a 6,111) - aug);	
24.         Status Activity           1         Conject           2         Conject           3         Conject           4         Conject           4         Conject           4         Conject           5         Conject           6         Conject           6         Conject           7         Conject           8         Conject           9         Conject           10         Conject           11         Conject           12         Conject           13         Conject	788×5 12+8 12+8 12+82-48 22+82-48 22+82-84 12+81-82 14+8-82 2348 2348 2348	(100000) (1000000) (1) (1)	8178	a 6.711) i = 8100 ()	
24.         Calentina Alexan (1991)	T2855 124524 2745248 2756248 2756248 255624 25624 245 245 245 245 245	480124034 ((KOTRIA)) 4) 4)		444110-3041	
24.         Status Activity           Constant         Constant           1         Constant           2         Constant           3         Constant           4         Constant           4         Constant           5         Constant           6         Constant           7         Constant           8         Constant           10         Constant           10         Constant           10         Constant           10         Constant	728355 2248523 2248523 2248523 2250653 2250853 2240 2443 2443 2443 2443 2443 2443 244		8118 0100 8118 8118 8118 8118 8118 8118 8118 8118 8118 8118	a 6,1111 - =0.08(1)	
24.         Status Activity           102.         Status Activity           1         Status Activity           2         Status Activity           3         Status Activity           3         Status Activity           3         Status Activity           4         Status Activity           5         Status Activity           6         Status Activity           7         Status Activity           7         Status Activity           8         Status Activity           8         Status Acti	72855 2298 2248248 2440248 24402549 24402549 244028 2440 2440 2440 2440 2440 2440 24		001000000 001700	a 46/11)( = =0.00[0	
21.         Call VI. A City (VIII)           Call VIII)           Call VIIII)           Call VIIII)           Call VIIII)           Call VIIII)           Call VIIIII)           Call VIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	72815 2245 22482254 2442254 2442254 2442254 2442254 2443 2445 2445	**********	8178 -	a 6,1111819()	
24.         Call of the Architecture           1         Call of the Architecture           2         Call of the Architecture           3         Call of the Architecture           4         Call of the Architecture           5         Call of the Architecture	72895 2240 2240 2240 2240 2240 2240 2240 224	ARCORDA	Alteriore           Alteriore <td< td=""><td>a 6.711) = =0.98(0)</td><td></td></td<>	a 6.711) = =0.98(0)	
Call Control Arg in (1991)           Call Processing Arg in (1991)	72815 72855 7248 7248 7248 7248 7248 7248 748 748 748 748 748 748 748 748 748 7	48078404 ((REFERSE)) (#) (#)	MITE	a 46 (11) i - atanti (	
21.         Call 20, ACM (1990)           Call 20, ACM (2000)         Call 2000           1         Call 2000           2         Call 2000           3         Call 2000           4         Call 2000           4         Call 2000           4         Call 2000           5         Call 2000           6         Call 2000           7         Call 2000           8         Call 2000           9         Call 2000           10         Call 2000           11         Call 2000           12         Call 2000           13         Call 2000           14         Call 2000	7 2 8 9 % 2 2 4 % 2 4 % 2 2 4 % 2 4 % 2 2 4 % 2 4 % 2 2 4 % 2 % 2 2 4 % 2 2 4 % 2 4 %	ARCORDA	NTE         NTE           NTE         -           NTE	a 6,1111/ -= 8198 ()	

DocuSky 雲端資料庫顯示如圖:

LOURCE_TYPE	五行車	DYNASTY	
MPEROR		CH_YEAR	[mailut
0000		SEASON	41.
IDEF_ZAIYI	7.8.8	a second second	
十一年 (秋,年	23. · 董仲舒以為時費。宋比年為乘丘、鄙之!	11、百姓秋日・除来盛	· 故二¥俱 <mark>题</mark> ·自向以為時未整公績優, 建灾

(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:PlaceRefld 或@PlaceRefld), 欲 啟動 GIS 顯示,必須執行此設定。

1.在載入 DocuXML 檔之前,需先於▶ TABLE #3 中點選「類別/子類別」欄,啟動\*號,如圖:



### 2.載入 DocuXML 檔

3.「類別/子類別欄」成為可修改狀態,將有附加地名屬性的標籤,修改為

「Place/標籤名稱」。如圖:

		A CONTRACTOR OF A CONTRACTOR A CONT	101101-012	Cana - URBNAL	
1. Udot, Hankardontto		Lider_HankarperBlain	88.84	37.8/28/T	
2. Stat Installation And Laboration		Liter, NemeRelgonAndCulture	E BBHRLE	(Farket)	*
3. Udal_Halteducia/Organization	1.0	Law Harradocal Organization		(Facks)	
4. Gdet_reports		Liter_keyenem-	THE REAL	Faden	
5 yak senatem		UMAC_HarmaTerro-	KPR0.H	()Earthern	
6 liter_measured		LMAP, Sala any entit	E390	( (Facility)	
P. Vall_CountyName		Wet_Courts/Name	1 Usef_CourtyName-	( Faller)	
R Date		Deter	12. 内架	(不進行能力)	
3 Lankane		Pater	1 96	(Factor)	
10. Parcentianes			11 1.4	(74/38/2)	
11. UM/L2016		Pacelishel, Zone	SDEM.	( Factor)	
12. UMCONIE		Pacelidel Church	100	Tackt	
O. Udel_Moneters	1 N	User United Sty-	100.0	- Faithr	

DocuSky 雲端資料庫顯示如圖:

DocuSky <sub>ec1a</sub>				ALTE A MARCHESIAN & STA
				Seato   Manual   [20   0 0 0
RH.	0.040	Read undo Y-W		Cocheta (Decent) (Date (Dece
ing/calling.		Query: all		
		Matches: 120 2 3 4 5 6 + RMR K		
Sort Manuferra	_	DocuGiti Lian Pant Classification (1) 22 11	=	
-		• • · · · · · · · · · · · · · · · · · ·	_	
C REENPLICED	10.0			UIDIA MUNICIPALITY DE CARENTAL
0 //#21909_1874(##2)	193	D ANT	CODE	11505
D 104 8 tvga_28880.004 81	18.1	· · · 11.0	ANTEORIARY_L.S.	44B
#3#3wgs_17451(#dd)	- 2		EVEL	4 <b>8</b>
C) #Fit tryps_20808 (#its)	2+		ASS_ALLE	17.8
O REALIZINGS_1950(Filler)	21		in a	14
C 224 twps 3651 (58)	2.4	and the second	INTRASTY	81
O #871www 1875/#871	2.1	A 19 20 20 15 12 2	WY .	*****
C =### 1608 21619/E##	1.		THOR	著作單人地理計算人/著作計書項計算人/ 武士基地大学主要
C Seeting Anticipal		and the arrest and	THERE	IN A SECOND FOR THE SECOND
C. THIR INDEPARTURE INC.			FLENAMED	which do #111112 000 and
① 九型線 tyge_45757(九甲線)	11	Cannol May new Witness Street States, 55, 57-58, mapping Street	INTERGORY LE	1818
口: 九阳州 hvgs_2029(九阳州)	1.8	which not bettern contract	-	
○ 私工業 tvg8_41943 (RX株)	1.5		10.00 W 40.00 (C)	* http://fid.nk.adu/to/THEL.Habiousmage.ph
0 DN#higk_43855(DH#)	15	作、学通用建筑事儿、自动就能完成有应注意者。而再供题在干,扎然包括		d'imge group-Ganlid les polinepe hen

# 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、 Refld 權威編號。

操作如圖:

#工具可額外匯出以下表	格型資料:			
► 🔳 🗄 metadata 🗐 .csv	場:20200823歷史自然災異記錄-M-Metadata.csv		<b>國出 meta</b>	data
▶ 匯出 tags 到 .csv 橿 :	20200823歷史自然災異記錄-M-Tags.csv	臣出	tags	<u></u>
<ul> <li>         ・         ・         ・</li></ul>	ents) 到.csv 幅;20200823歷史自然災異記錄-M-A	lignments.cs	IV V	露出 alignments

匯出檔案如圖:

	A	В	С	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	PL00000023563	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	PL00000023642	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	PL00000023563	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	PL00000026396	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(Refld)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:<u>https://docusky.org.tw/DocuSky/DocuTools/RefLookupTools/</u>

- (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 Refld。 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 Refld。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:<u>https://reurl.cc/WLdZK9</u>

QRCode :



- 
 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