標記與詞彙統計分析工具 (含 Palladio 視覺化工具) 操作手冊

目錄	. 1
一、 工具簡介、網址與範例檔下載	. 2
二、 首頁功能鍵介紹	. 2
三、 操作流程	. 3
1. START 開始	. 3
2. 載入文本或文獻集	. 3
(1) 方法一:上傳純文字	. 4
(2) 方法二:從 DOCUSKY 中上載文獻集	. 4
3. TERMLIST 上傳詞彙表	. 5
4. ANALYSIS 開始分析	. 6
5. 統計結果與匯出	. 8
(1) BASIC TERM FREQUENCIES 詞彙詞頻統計	. 8
(2) CATEGORIZED FILE RESULT 依段落統計詞頻	. 9
(3) CATEGORIZED TERM RESULT 依詞彙類別統計詞頻	10
四、 匯出資料與視覺化呈現	11

一、工具簡介、網址與範例檔下載

標記與詞彙統計分析工具由謝博宇先生設計開發,是一款對文字資料進行詞 彙統計的工具。當我們在進行文本研究的時候,常會好奇某些特定的字詞出現的 頻率,藉由對這些語詞出現頻率的統計,來觀察作者對於某些概念的重視程度, 或作者所要表達的意涵。

工具頁面如圖1所示:

Versione to Locksky termsdart tool The too big hys to clocksky termsdart tool The too big hys to clocksky termsdart tool The tool is also appoint by boxedy, if we want to bion how to do place dol (tow to add how to a	Webserste Des Clark metho						
Total visit airs supported by OccUMA, if you want there is do, place did [New Form 1x1 Form Liver/SU]. Take water is airs supported by OccUMA, if you want there is do, place did [New Form 1x1 Form Liver/SU]. Take water is airs supported by OccUMA, if you want there is do, place did [New Form 1x1 Form Liver/SU]. Take water is airs supported by OccUMA, if you want there is an interview of the prome, Term Liver is airs supported by OccUMA, if you want there is an interview of the prome, Term Liver is an interview of the prome of	This tool helps you to calculate term status	ica from docume	rits and categorated term lists.				
CORPUS CONFUS TERMILIST AMALYSIS AREAL ADDING Md Rew Corpus 1. free load 2. free load 2. free load 3. free lo	This tool is also supported by DocuSky, if y To know more about Docusky, click 'Docus	ou want to know ky" link on the r	r how to do, please click <mark>. How to add from Doct</mark> ght top corner, Thank yout	riky .			
Teol Usage Flow Chart CORPUS TERMLIST AMULYSIS RESULT Add Rew Corpus							
CORPUS TERMLIST AMULYSIS Result Add Rev Cepus Add Rev Term List Confine your Capue, Term Lists Result and Disput 1. fram Lind 1. fram Lind 1. fram Lind 2. Introm Capue, Term Lists Result and Disput 2. fram Dandary 1. fram Lind 2. Jast Ingle 1. fram Lind Jast Ingle 2. fram Dandary 1. fram Lind 2. Jast Ingle 1. fram Lind	Tool Usage Flow Chart						
Add New Corpus Add New Terms List Confirm your Corpus, Term Lists Result and Dulput 1. Franciscal 1. Franciscal 1. Franciscal 1. Franciscal 1. Franciscal 2. Franciscal 2. Franciscal 2. Stress Dansfary 2. Stress Term Sea 2. Oxycols by desagered memory and provided memory 1. Franciscal 1. Franciscal 1. Franciscal 2. Oxycols by desagered memory 1. Franciscal 1. Franciscal 2. Oxycols by desagered memory	CORPUS		TERMLIST		ANALYSIS		RESULT
1. Area Load min Inform Load min L. Franc Load min L. Franc Load 2. Franc Dendlay 2. Franc Dendlay 2. Start Dendlay 2. 2019 tags 2. Darpa by comparison terms 3. If Terms are togged in SN4 tags, and no come area Microscoped, on to the nome 3. All tags 3. All tags 3. All tags	Add New Corpus		Add New Term List		Confirm your Corpus, Term Lists	-10	Result and Output
2. From Dendlay 2. From Dendlay 2. Then Dendlay 2. Then Dendlay 3. If them are topped in 94% top, ont in occus and the next-top to the next	1. From Local	-	1. From Local	0	1. Extra term lists	-	1. From Local
J. If trans are tagged in XM, tags, and no J. Output by documents data turn kits roquind, go to thy not	2. From DecuBility		2. From DecuSily		2.3390 mps		2. Output by categorized terms
dap. 4. Word Caus			 If terms are tagged in XHL tags, and no outra term lists required, go to the next step. 				3. Output by documents 4. Word Cloud

圖1 工具首頁

本工具的網址為:

<u>https://docusky.org.tw/DocuSky/docuTools/TagStatsTool/index.html</u> 為了便於引導使用者使用本工具,使用者可以至

https://tinyurl.com/y6nqbjyq

下載本使用說明所使用的建庫檔「西遊記_DocuSky 範本.xml」,將此建庫檔 在個人的 DocuSky 上建庫後,即可按本使用說明的操作方式依序操作各項功能。

二、 首頁功能鍵介紹

在工具頁面的左側是如何操作本工具的流程,分別為:「開始」(START), 「上傳文本或文獻集」(CORPUS),「上傳詞彙表」(TERMLIST),「開始分析」 (ANALYSIS),以及「結果呈現」(RESULT)。如果上傳的文本是經過「碼庫思 (MARKUS):古籍半自動標記平台」或「批次標記工具」(ContentTagging Tool) 標記的文本,則可省略「上傳詞彙表」的步驟,直接以標記過的詞彙進行分析。

START	開始
CORPUS	上傳文本或文獻集
TERMLIST	上傳詞彙表
ANALYSIS	開始分析
RESULT	結果呈現

圖 2 操作面板

三、 操作流程

1. START 開始

nis tool helps you to calculate term stati	istics from docume	nts and categorized term lists.				
you are new to the tool, please click	low to use -					
his tool is also supported by DocuSKy, it	f you want to know	how to do, please click How to add from Doc.	iSky -			
know more about Docusky, click "Doc	uSky" link on the ri	ght top corner, Thank you!				
ool Usage Flow Chart						
•						
CORPUS		TERMLIST		ANALYSIS		RESULT
Add New Corpus		Add New Term List		Confirm your Corpus, Term Lists		Result and Output
Add New Corpus 1. From Local		Add New Term List	1	Confirm your Corpus, Term Lists		Result and Output
Add New Corpus 1. From Local 2. From DocuSky	Û	Add New Term List. 1. From Local 2. From DocuSity	Ŷ	Confirm your Corpus, Term Lists 1. Extra term lists 2. XML tags	Ŷ	Result and Output 1. From Local 2. Output by categorized terms
Add New Corpus 1. From Local 2. From DocuSky	Ŷ	Add New Term List 1. From Local 2. From DocuSky 3. If terms are tagged in XML tags, and no	Ŷ	Confirm your Corpus, Term Lists 1. Extra term lists 2. XML tags	Ŷ	Result and Output 1. From Local 2. Output by categorized terms 3. Output by documents
Add New Corpus 1. From Local 2. From DocuSky	Ð	Add New Term List 1. From Local 2. From DocuSky 3. If terms are tagged in XML tags, and no extra term bis required, go to the next	Ð	Confirm your Corpus, Term Lists 1. Extra term lists 2. XML tags	ß	Result and Output 1. From Local 2. Output by categorized terms 3. Output by documents 4. Wood Cloud

圖 3 進入工具後的頁面

在開始步驟中,有簡明的使用流程,使用者也可以直接點選 How to use 來 連結本使用說明的 PDF 檔。關於如何由 DocuSky 載入文獻集,也會在本使用說 明中說明。

2. 載入文本或文獻集

DocuSky TermStats Tool		Decisity
	Add New Corpus	Corpus List
00045	You have two methods to add corpus you want to analyzen Nethor 1: Add from locat	No Xinot Carpui Anst.
	Corpus Name	#~~~~
ANALYSES (Choose Rise as To choose Select 万法一・上得※	۳XF
	Method 2: Add from OccuBly	
	teg in Dec./Sky 188, then select the corpus you w方法二:從 Do Doc./Sy Corput Widget	ucSky 上載文獻
	Next	

圖4 載入文本或文獻集

點選左側的「CORPUS」按鈕,進入上傳文本或文獻集。本工具提供兩種文本上傳方式,一種是上傳純文字(txt 檔,UTF-8 格式);另一種是從 DocuSky 中上載文獻集。

(1) 方法一:上傳純文字

先在文獻集名稱(Corpus Name)中輸入自訂的文獻集名稱,然後點選 Select 按鈕開啟上載檔案視窗。然後點選 Add 按鈕。

have two me	thods to add corpus you milocal	want to analyze:
Corpus Name	西遊記	1. 輸入文獻集名稱
Choose File	Xiyuji.txt Select	2. 上載純文字檔
\dd		3. 點選 Add 按鈕

圖 5 上載純文字檔

此時,右方欄中將會出現文獻集的列表,若要刪除該份文獻集,則點選右側 紅色的 ⊠ 圖示。使用者也可以一次上傳多份 txt 檔,並將多份檔案放在同一個 文獻集當中,以利後續的詞頻統計分析。

Corpus List							
ID	Title	#Doc	Del				
1	西遊記	100	×				

圖 6 完成文本載入

(2) 方法二:從 DocuSky 中上載文獻集

點選 DoucSky Corpus Widget 按鈕連結 DocuSky 資料庫(需登入),然後點 選載入即可將文本帶入工具中。

ou have two m	1. 太平廣記 2. 西遊記1-5回 3. 禁辭類一	太平廣記樣本 上課用的西遊記	載入 載入
ou have two m	2. 西遊記1-5回 3. 禁辭類一	上課用的西遊記	載入
ou have two m	3 林窩空街		
ou have two m	· AERIT #R	<u></u>	載入
	4. 楚辭類—	楚辭補注	載入
1ethod 1: Add fr	5. 楚辭類—	离住影子	載入
	6. 唐詩宋詞資料庫	唐詩	載入
Corpus Name	7. 唐詩宋詞資料庫	宋詞	載入
Choose File	8. 紅樓夢	元智版	載入
	9. 翰林版國小課本	00.2014_國語首冊課文	載入
Add	10. 翰林版國小課本	01.2014_國語一上課文	載入
1ethod 2: Add fror Log in DocuSky DocuSky Corpu	n DocuSky DB, then select the corpus you us Widget	want to add:	

圖7從DocuSky 載入文本

上述兩種方法中擇一完成文獻集上載後,按 Next 按鈕進入下一步。

3. TERMLIST 上傳詞彙表

上傳詞彙表的方式在系統的規劃中也有兩種,分別為上傳詞彙表(txt 檔, UTF-8 格式),以及從 DocuSky 載入詞彙表,惟目前後者的功能因故暫停使用。 但若使用者從 DocuSky 中載入的文本已經經過標記,則可以標記過的詞彙直接 進行分析。故此時便不用上傳任何詞彙表,直接按 Next 按鈕進入下一步。上傳 的詞彙表一樣會顯示在右側的欄位中。

You have two met Method 1: Add fr	hods to add term l om local	st you wan	t to calculate stati	stical result in previous loaded cor	pora:
You can create a save as Unicode	a new term list file UTF-8 text file (.t	at local. Ple kt).	ease separate each	n term with a newline (or comma)	and
Choose Files	No file chosen	Select	方法一:	上傳詞彙	
Add					
Method 2: Down	oad from DocuSky	<u> </u>			
Upload Termlist	to DocuSky			方法二:從 DocuSk	xy 上載詞如
1. Log in DocuS	ky and update file	list: Docus	Sky File Widget	(目前暫停使用)	
2. View the term	n list groups saved	on DocuSky	y View Term List	Groups	
	No sa	ved term list	t group in your Doc	uSky account.	
Terms in XML tag	JS				
If you use XML	format documents	with tagged	d terms, please go	to the next step.	
				I	Next

圖 8 上傳詞彙列表

由使用者直接上傳詞彙表 txt 檔的方式同前,這裡不再贅述。詞彙表 txt 檔的格式,必須每個詞彙以換行分開。

🥘 西遊	記妖怪列表	長 - 記事本					-	×
檔案(F)	編輯(E)	格式(O)	檢視(V)	說明(H)				
混寅熊特黑白凌黃虎鎮鎮白白屍黃世將山處風衣虛風先元元骨骨魔袍魔軍軍士怪秀子怪鋒大子精夫 怪	王 士 仏 人							^

圖 9 編寫詞彙列表

上載詞彙表到 DocuSky 存檔功能目前暫停使用,故不在此處贅述。

4. ANALYSIS 開始分析

開始分析的頁面分為兩種分析形式, Corpus and term lists 分頁表示,使用者可以利用剛剛上傳的詞彙表進行分析; Corpus with XML tags 則表示利用標記過

的詞彙進行分析。若是使用前者,確認上傳檔案無誤後,即可按 Run Analysis 進行分析。

	Α	alysis Settings
Please confirm the corpora and term in If you want to analyze XML tag conten	sts you selected. Once confirmed, click "Run" to start the analysis. Is in your documents, click on "Corpus with XML Tags".	
	Corpus and term lists	Corpus with XHL tags
	Corpus	Term List
	Corpora (1)	No term list exist.
2822		
	(F	un Analysis

圖 10 以詞彙列表進行分析 (Corpus and term lists)

使用已經標記過的文本進行統計(Corpus with XML tags)時,下方會出現 不同的標記類別供使用者選擇,依使用者的需求進行勾選即可。

	Analysis Se	ttings
	Please confirm the corpora and term lists you selected. Once confirmed, click "Run" to start the analysis. If you want to analyze XHL tag contents in your documents, click on "Corpus with XHL Tagu".	
RMUST	Corpus and term lists	Corpus with XML tags
4,7515	Corpu	15
	Corputa Corputa	(1)
	XHL tag set	tings
	XHL tag set Teg Knor	tings Fitano Ten Lat
	XHL tag set	tings fitmso Tom Lit No Fitmog 🔍
	XHL tag set	tings Filming Tess Lit Na Filming ¥ Ng Filming ¥
	XHL tag set	tings Fitmsojiem (st No Fiteino V No Fiteino V
	XHL tag set Tag Nature Udd_BuddhistDuracter Udd_DeadDuracter Udd_Foods	tings Filmring Ters Lit No Filtering V No Filtering V No Filtering V

圖 11 以標記文本進行分析 (Corpus with XML tags)

以上圖為例,在載入的「西遊記_DocuSky 範本」這個文件集中,已有 LocName (地名)、Udef_BuddhistCharacter(佛教人物)、Udef_DaoistCharacter(道教人物)、 Udef_MainCharacter(主角)......等許多的標記。使用者可以按照目前需要進行 統計的詞彙類別勾選一到多個類別進行統計。現以《西遊記》當中的佛、道教人 物、妖怪(Udef_Monster)的詞頻進行分析。便可得到如下圖 11 的畫面:

The calculation of term statistics about terms in XML tags - (Total Time:0:00:00): Udef_BuddhistCharacter Udef_Monster Udef_Monster in corpora - 西遊記
has completed.
Output CSV File Basic Term Frequencies ^O Categorized File Result ^O Categorized Term Result
Each line of data shows term frequency and document frequency of terms found.
Save as CSV Save as JSON Upload to DocuSky Copy CSV to Clipboard Open Palladio
Word Clouds
西遊記 Save as PNG
鹿力大仙
一百百姓。 四海龍王 天薙元帥 太上老君 凌空子
鐘元子 鐘教伽藍 木吒 觀世音 mmn工
翻音菩薩 ^{九曜星 100000} 五方揭諦 伽葉
虎力大仙 大丁

圖 12 統計結果

5. 統計結果與匯出

統計的結果,除可以文字雲(Word Clouds)呈現並點選「Save as PNG」將 此圖檔下載。使用者可按三種不同方式進行詞彙統計,並將統計數據數據以 CSV (Save as CSV)、JSON (Save as JSON)檔案格式下載或上載到 DocuSky 上(Upload to DocuSky),或是直接將數據複製後匯出至史丹佛大學(Stanford University)所開發的視覺化平台 Palladio 進行後續的處理與呈現。

(1) Basic Term Frequencies 詞彙詞頻統計

最基本的統計數據是「Basic Term Frequencies」,輸出的資料將包括標記詞 彙類別(Category)、詞彙(Term)、詞彙出現次數(TF)以及該詞彙出現在多 少個段落(DF)。輸出的 CSV 檔案可以 Microsoft Excel 開啟,如下圖所示:

Category	Term	TF	DF
Udef_BuddhistCharacter	如來	207	38
Udef_BuddhistCharacter	觀音	65	25
Udef_BuddhistCharacter	觀音菩薩	50	31
Udef_BuddhistCharacter	觀世音	28	23
Udef_BuddhistCharacter	伽葉	10	2
Udef_BuddhistCharacter	靈吉菩薩	9	2
Udef_BuddhistCharacter	鳥巢禪師	6	4
Udef_BuddhistCharacter	如來佛祖	3	3
Udef_BuddhistCharacter	如來佛	3	2
Udef_BuddhistCharacter	文殊菩薩	3	3
Udef_BuddhistCharacter	菩提祖師	2	1
Udef_BuddhistCharacter	釋迦牟尼	2	2
Udef_BuddhistCharacter	毗藍婆菩薩	1	1
Udef_BuddhistCharacter	大勢至著	1	1
Udef_BuddhistCharacter	普賢菩薩	1	1
Udef_DaoistCharacter	玉帝	189	33
Udef_DaoistCharacter	哪吒	73	15

圖 13 Basic Term Frequencies 統計結果

使用者可就此表進行詞彙的統計與後續的利用。

(2) Categorized File Result 依段落統計詞頻

若選擇了「Categorized File Result」,則是以每個段落(filename)為基礎, 統計各類詞彙在該段落中出現的次數。使用者也可以在匯出前選定要一併匯出 的 metadata 欄位,以作為匯出之後資料的參照或後續利用。

如以載入的「西遊記_DocuSky 範本全標記」為例,在匯出時我們先選擇「Title」欄位一併匯出。

The calculation of term statistics about terms in XML tags - (Total Time Udef_BuddhistCharacter Udef_DoalstCharacter Udef_Monster In corpora - 西姆語	::0:00:00):
has completed.	
Output CSV File OBasic Term Frequencies Categorized File Result Categorized Ti	ierm Result
Each line of data shows categorized terms found in a document.	
Select metadata output to CSV file, then click "Save" to save the result	
Select metadata columns to output:	
✓ Title	
Source	
Passageid	
Save as CSV Save as JSON Upload to DocuSky Copy	y CSV to Clipboard Open Palladio

圖 14 以「Categorized File Result」方式進行匯出

匯出的 CSV 如下圖 15:

Category	Corpus	Filename	Title	TermsCount	TotalOccur	TermList	Detail			
Udef_BuddhistCharacter	西遊記	西遊記p0001	靈根育孕源流出	1	2	菩提祖師	菩提祖師(2)			
Udef_DaoistCharacter	西遊記	西遊記p0001	靈根育孕源流出	4	4	千里眼;順風耳;玉帝;搖光	千里眼(1);順團	以耳(1);玉 帘	;(1);摇光(1))
Udef_Monster	西遊記	西遊記p0002	悟徹菩提真妙理	1	1	混世魔王	混世魔王(1)			
Udef_DaoistCharacter	西遊記	西遊記p0003	四海千山皆拱伏	18	27	西海龍王敖閏;秦廣王;玉帝	西海龍王敖閏	(3);秦廣王	(3);玉帝(3);	東海龍王敖
Udef_Monster	西遊記	西遊記p0003	四海千山皆拱伏	3	3	混世魔王;狻猊;牛魔王	混世魔王(1);翁	§猊(1);牛质	【王(1)	
Udef_DaoistCharacter	西遊記	西遊記p0004	官封弼馬心何足	4	59	玉帝;哪吒;巨靈神;托塔天王	玉帝(28);哪吒	(16);巨靈神	•(12);托塔-)	天王(3)
Udef_Monster	西遊記	西遊記p0004	官封弼馬心何足	1	1	牛魔王	牛魔王(1)			
Udef_BuddhistCharacter	西遊記	西遊記p0005	亂蟠桃大聖偷丹	1	1	觀音	觀音(1)			

圖 15 Categorized File Result 統計結果

在上圖 15 中,「Category」欄為詞彙類別,「Corpus」欄為文獻集名稱,「Filename」 則為段落檔名。「TermCount」是本段落(Filename)中,本類詞彙出現幾個? 「TotalOccur」則是本類詞彙總共出現幾次?如在 Udef_BuddhistCharacter 這類 中,出現了一個角色,即菩提祖師(TermsCount=1),而菩提祖師在第一章(西 遊記 p0001)中被提到兩次(TotalOccur=2)。而出現的詞彙(TermList)為「菩 提祖師」,在「Detail」欄中則可看出菩提祖師出現兩次(菩提祖師(2))。

又如,道教人物(Udef_DaoistCharacter)在第一章出現了千里眼、順風 耳、玉帝與搖光等四位道教神祇,且各被提到一次,因此TermsCount與 TotalOccur 分別都為4。

(3) Categorized Term Result 依詞彙類別統計詞頻

「Categorized Term Result」是依照詞彙的類別進行詞頻的統計。使用者也可以在匯出前選定要一併匯出的 metadata 欄位,以作為匯出之後資料的參照或後續利用。

如以載入的「西遊記_DocuSky 範本全標記」為例,在匯出時我們先選擇「Title」欄位一併匯出。

The calculation of term statistics about terms in XML tags - (To Udef_BuddhistCharacter Udef_Monster Udef_Monster In corpora - 世語語	tal Time:0:00:00):
has completed.	
Output CSV File O Basic Term Frequencies O Categorized File Result Categorized File Result Each line of data shows. Please select metadata and HTML tag Select metadata columns to output:	orized Term Result attribute, then click "Save" to output CSV file.
Title	
Passageid	
Select XML tag attributes to output (WARNING: Tagged terms v	vith the same selected attributes will be combined)
Save as CSV Save as JSON Upload to DocuSky	Copy CSV to Clipboard Open Palladio

圖 16 以「Categorized Term Result」方式進行匯出

匯出的 CSV 如下圖:

1					
Category	Corpus	Filename	Title	TagVal	Frequency
Udef_BuddhistCharacter	西遊記	西遊記p0001	靈根育孕源流出	菩提祖師	2
Udef_DaoistCharacter	西遊記	西遊記p0001	靈根育孕源流出	千里眼	1
Udef_DaoistCharacter	西遊記	西遊記p0001	靈根育孕源流出	順風耳	1
Udef_DaoistCharacter	西遊記	西遊記p0001	靈根育孕源流出	玉帝	1
Udef_DaoistCharacter	西遊記	西遊記p0001	靈根育孕源流出	搖光	1
Udef_Monster	西遊記	西遊記p0002	悟徹菩提真妙理	混世魔王	1
Udef_DaoistCharacter	西遊記	西遊記p0003	四海千山皆拱伏	東海龍王敖廣	2
Udef_DaoistCharacter	西遊記	西遊記p0003	四海千山皆拱伏	南海龍王敖欽	2
Udef_DaoistCharacter	西遊記	西遊記p0003	四海千山皆拱伏	北海龍王敖順	2
Udef_DaoistCharacter	西遊記	西遊記p0003	四海千山皆拱伏	西海龍王敖閏	3

圖 17 Categorized Term Result 統計結果

在上圖 17 中,「Category」欄為詞彙類別,「Corpus」欄為文獻集名稱, 「Filename」則為段落檔名。「Title」是隨同匯出的 metadata 欄位。TagVal 為匯 出的詞彙,「Frequency」則是該詞彙在本段落中出現的頻率。如在第三章中 (西遊記 p0003)屬於道教人物的東海龍王、南海龍王、北海龍王各出現了兩 次,而西海龍王敖閏則出現了三次。

四、 匯出資料與視覺化呈現

本工具所搭配的視覺化呈現工具為史丹佛大學所建置的 Palladio 視覺化平 台,該平台可提供多樣化的視覺化呈現方式。因本工具與該平台的連結為 CSV 統計資料的匯出與關連分析圖的呈現,故僅就此部分進行說明,關於 Palladio 的其他功能請見: http://hdlab.stanford.edu/palladio/help/。

我們選用 Categorized Term Result 的統計結果後,直接選取「Copy CSV to Clipbord」即可將本統計結果以 CSV 格式暫存在剪貼板(Clipboard)中,此時 再點選「Open Palladio」即可以另開分頁的方式在瀏覽器中開啟 Palladio。

圖 18 複製並開啟 Palladio

進入 Palladio 之後,即可將已經暫存於剪貼板中的 CSV 資訊直接在 Palldio 的 Load .csv or spreadsheet 中,如下圖 19 所示:

Create a new project	Oracte e new Delladia eraiset by velociding your data from a percedebrat or flat file or load data from a CDADOL codesist. Not
Load an existing project	sure how Palladio works?
Try with sample data	Load .csv or spreadsheet Copy and paste out of your spreadsheets, drag-and-drop to upload tabular data (e.gcsv, .tab, .tsv), or link to a file in a public Dropbox folder to create a new Palladio project. Not sure how to structure your data?
	612 "Udef_DaoistCharacter", "西班記", "田班記", "田班上", "张成真", "班記", "I" 63 "Udef_LoudhistCharacter", "西班記", "西班記", "西班記", "田班上", "王定成真", "班記", "I" 64 "Udef_LoudhistCharacter", "西班記", "西班記", "西班記", "田班記", ""田述記", "田班記", ""
	More than one table? No problem! If you have more than one table, start by uploading your primary table. The primary table should contain the main entities you want to visualize. It could be a collection of objects, like persons or letters, or more abstract concepts, like relationships or flows. Once you have uploaded your primary table, you will be able to extend it with additional information from other tables.
	Load data from a SPARQL endpoint (beta)

圖 19 在 Palladio 貼入複製的 CSV 資訊

點選「Load」按鈕即可將貼上的 CSV 資訊載入到 Palladio 當中。 進入到圖 20 的畫面後,使用者可點選「Graph」進入繪圖功能。

2	Data	Мар	Graph	Table	Gallery
Pro	ovide a	title to	this pro	oject	
ow	details				
U	ntitled				
Pri	mary table	е			640 rows
۲	Category				Text 😑
۲	Corpus				Text
۲	Filename	•			Text
۲	Title				Text
۲	TagVal				Text
۲	Frequenc	у			Number
۲	generate	d			Text
*					Đ

圖 20 點選 Graph 進入繪圖功能區

接下來使用者要決定哪兩項資料要當作是繪製關連圖的 Source (來源)與 Target (目標)。如在本案例中, Source 可以是各段落的 Title, 也可以是各段落 的檔名 (Filename),而 Target 則為被標記的詞彙 (TagVal)。當 Source 與 Target 選定之後,就會在繪圖區域生成詞彙關連圖。

	Settings			\equiv	
		Source Filename 👻			
	ні	ghlight			
		Target TagVal 👻			
	ні	ghlight 🔀			
	Sho	w links 🔀			
	Siz	anodes 🗙			
	Accor	ding to Number of Untitleo	1 -		
			*	Download	
Choose far	cet dimensions				×
Category	Corpus File	name Frequency	TagVal Ti	tle generated	
					Close
			z 于姚元én	西遊記の	Close

圖 21 選擇 Source 與 Target

使用者可按照個人的研究需求,在 Target 或 Source 中選擇一個點選 Highlight 使之醒目提示,也可以點選 Size nodes 使點的大小依照其頻率而有所 不同。



圖 22 點選參數後生成關連圖

不過,這樣的關連圖過於繁雜,我們可以利用篩選的機制把暫不需要的點 隱藏起來。我們可以點選左下角的「Facet」開啟篩選機制。



圖 23 點選 Facet 進行資料篩選

在「Dimensions」中我們可以選擇要進行篩選的 CSV 欄位,

	_						
Category	Corpus	Filename	Frequency	TagVal	Title	generated	
							Close

圖 24 選擇要進行篩選的欄位類別

例如,我們只需要觀察《西遊記》中妖怪在各章節中出現的關連分析,我 們就在這裡選擇「Category」(詞彙類別),然後在篩選的選單中,點選 Udef_Monster,如此就可以先把道教人物與佛教人物先行隱藏,畫面也會變得 比較清爽。



之後,我們再靠著拖曳圖中的點,就可以呈現比較清楚的妖怪詞彙關連分析圖了。



圖 26 《西遊記》妖怪詞彙關連分析圖



圖 27 以牛魔王為核心的妖怪詞彙關連分析圖

那麼,要如何看待圖 27 這張詞彙關連分析圖呢?我們發現,這是一張以牛魔王 為核心的關連圖,由章節的節點所串起來的關係顯示,牛魔王與紅孩兒在西遊 記 p0053、西遊記 p0042、西遊記 p0040、西遊記 p0041、西遊記 p0059 等章節 中共同出現過,而又如鐵扇公主在西遊記 p0059 這一章中與紅孩兒、牛魔王共 同出現過。這種「共現關係」即能透過這樣的關連分析圖迅速呈現。並且可以 利用更多不同類別的詞彙、CSV 欄位的關連性,提供使用者做更多的視覺化觀 察。