# Chinese Word Sketch Engine操作手册

#### 1. 網址

http://wordsketch.ling.sinica.edu.tw/



# 2. 登入

連線到 wordsketch.)	ling.sinica.edu.tw 🛛 🛛 🔀
	G P
Word Sketch Engine	
使用者名稱(U):	2
密碼(P):	
	記憶我的密碼(R)
	確定 取消

### 3. 檢索首頁



### 4. 檢索關鍵詞或詞組(Concordance)的語料

點選 New Query 後進入以下頁面:

🗿 Word Sketch Engine - Mic	rosoft Internet Explorer		
檔案(E) 編輯(E) 檢視(V)	我的最愛(A) 工具(D) 說明(H)		
🕒 l-ā - 🕥 - 💌	📓 🏠 🔎 搜尋 🤺 我的話	ee 🚱 🔗 🎍 🖬 🛛 📙 🧕 🚳	
商址 🛈 🙋 http://wordsketch.lin;	g sinica.edu.tw/cws/		🖌 🎦 移至 🛛 連結 🛛 Norton AntiVirus 😡 🗸
Home Concordance W	Vord Sketch Thesaurus Sketch-	Diff	
Corpus: gigaword2	*		
Kevword(s)			Make Concordance
Phrase:			
Word Form:	PoS: unspec	ified 🗸 Match case:	
CQL:			
Default	t attribute: word 🎽 Tagset summ	lary	
Context			
Oner Time:			
Query rype.	I of these items.	Right context	
Window Size: 5	Y tokens	5 × tokens	
Word Form:			
Do.S.	ioun 🔨	noun	
(use Ctrl+click for	erb 🗧	verb	
multiple selection)	dyerb 🖌	adjective adverb	
Make Concordance			
「完成			新院網路

#### 4.1 上方選項

點選上方 Home、Concordance、Word Sketch、Thesaurus、Sketch-Diff 等五選 項,可連結至主頁(Home)與其他檢索功能頁面。

#### 4.2 語料庫 (Corpus) 選單

此下拉選單提供繁體字版與簡體字版語料,以及次語料庫檢索設定選項:

(1) 設定 gigaword2\_cna,可查詢繁體字版的 CNA 中央通訊社語料。

(2) 設定 gigaword2\_xin,可查詢繁體字版的 XIN 新華社語料。

(3) 設定 gigaword2\_zbn,可查詢繁體字版的 ZBN 新加坡聯合早報語料。

(4) 設定 sinica,可查詢繁體字版的中央研究院平衡語料庫 5.0 版語料。

如果已經建有其他次級語料庫,次語料庫的名稱亦會出現在下拉選單中。

#### 4.3 關鍵詞(Keyword(s))設定

可輸入單一詞彙(word),含二個或二個緊鄰詞彙的詞組(phrase),或者以語料 庫檢索語言(Corpus Query Language (CQL))來設定檢索的關鍵項目。 (1)詞組(Phrase)設定 每個詞彙之間應空一個半形空白間隔,例如:「覺得說」、「我想你」、 「一般認為,這是」等等。 (2)詞彙(word) 可輸入所要查詢的關鍵詞彙。例如:「覺得」、「想」、「認為」。 (3)語料庫檢索語言(Corpus Query Language (CQL)) 設定細則請參看語料庫檢索語言說明: http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPSyntax.html

# 4.4 語境 (Context) 設定

本項目提供使用者進一步設定關鍵詞前後搭配的詞語環境。使用者可以檢索到非緊鄰的詞彙搭配語料。

# 4.4.1 檢索類型(Query Type)

下拉選單中提供三種進階設定方式:

- (3) None (of these items):過濾掉同時含以下欄位中關鍵詞左、右二詞彙的語料。

### 4.4.2 設定與關鍵詞的距離(Window Size)

下拉選單中提供 1 到 10 個詞(token)的設定方式。預設值為 5,表示在和關鍵詞 距離 5 個詞彙之內的範圍。

範例一:欲檢索含「逮捕」的語料。可如下設定:

🗿 Word Sketch Engine - Microsoft Internet Explorer			- 2 🛛
檔案 (E) 編輯 (E) 檢視 (Y) 我的最愛 (A) 工具 (D) 說明 (H)			A.
🌀 上—頁 • 💿 · 💌 💈 🏠 🔎 搜尋 🌟 叛的最爱 🤣 🐼 • 🌽 🔟 • 📜 💦 • 🌋			
網址① (劉 http://wordsketch.ling.sinica.edu.tw/cws/	💌 🄁 🕴	移至 連結	i Norton AntiVirus 😡 🗸
Home Concordance Word Sketch Thesaurus Sketch-Diff Corpus: gigaword2 v Keyword(s) Phrase:		Make	Concordance
Word Form: 逮捕      PoS: unspecified      Match case			
CQL:			
Detault attribute: word Y lagset summary			
Context			
Query Type: All v of these items.			
Window Size: 5 V talcos			
Word Form:			
PoS: (use Ctrl+click for multiple selection) Make Concordance			
<ol> <li>網頁發生錯誤。</li> </ol>			2 網際網路

範例二:欲檢索含「連...都」的語料。若如下設定:

🗿 Word Sketch Engine - 1	Microsoft Internet Explorer		
檔案(E) 編輯(E) 檢視(⊻	) 我的最愛(A) 工具(T) 説明(H)		At 1997
G +-頁 • ② • [	🗴 💈 🏠 🔎 搜尋 🥎 我的最	愛 🔗 🗟 • 🌺 🗹 • 📙 🔝 • 🚳	
網址① 顲 http://wordsketch	ling.sinica.edu.tw/cws/		💙 芛 移至 🛛 連結 🛛 Norton AntiVirus 🚱 🗸
Home Concordance	Word Sketch Thesaurus Sketch-	Diff	
Corpus: gigaword2	~		Mala Oreconderes
Keyword(s)			Make Concordance
Phrase:			
Word Form: 連	PoS: unspec	ified 👻 Match case: 🗔	
CQL:			
Def	ault attribute: word 🚩 <u>Tagset summ</u>	lary	
Context			
Query Type:	All 🗸 of these items.		
	Left context	Right context	
Window Size:	5 👻 tokens.	5 💌 tokens.	
Word Form:		者問	
PoS:	noun	noun	
(use Ctrl+click for	adjective	adjective	
manple selection)	adverb 🔛	adverb	
Make Concordance	e		
) ▲) 網頁發生錯誤。			

在關鍵詞(keyword(s))的 word form 中輸入「連」, 在語境(Context)的 Query Type 中設定 All, 在右側語境(Right context)中輸入「都」, 在右側語境,距離範圍(Window Size)中設定 5 個詞長。 則表示:「擷取關鍵詞「連」右側五個詞的範圍內,出現「都」的語料。」 範例三:檢索「以」,且欲過濾掉「不...以...為」的語料,若如下設定:

🗿 Word Sketch Engine - Microsoft Internet Explorer	
檔案(P) 编辑(E) 檢視(Y) 我的最愛(A) 工具(I) 說明(H)	
🔇 上—頁 • 🚫 · 💌 🛃 🏠 🔎 搜尋 🌟 我的最爱 🤣 🔗 è 🎍 🗃 • 🗾 👧 • 🦓	
網址 D 🎒 http://wordsketch.ling.sinica.edu.tw/cws/	🖌 🔁 移至 連結 Norton AntiVirus 😵
Home Concordance Word Sketch Thesaurus Sketch-Diff	
Corpus: gigaword2 🗸	
Keyword(s)	Make Concordance
Phrase:	
Word Form: 🔍 PoS: unspecified 💌 Match case:	
CQL:	
Default attribute: word 👻 Tagset summary	
Context	
Query Type: All 👻 of these items.	
Left context Right context	
Window Size: 3 v tokens. 10 v tokens.	
Word Form: 本	
PoS: noun noun	
(use Ctrl+click for adjective adjective	
adverb	
Make Concordance	
▲) 網頁發生錯誤。	網際網路

在關鍵詞(keyword(s))的 word form 輸入「以」

在 Query Type 中設定 None,

在語境(Context)的左側語境(Left context)中輸入「不」,

在右側語境(Right context)中輸入「為」,

- 在左側語境,距離範圍(Window Size)中設定3個詞長,
- 在右側語境,距離範圍(Window Size)中設定10個詞長。
- 則表示:「擷取「以」的語料,但刪除關鍵詞「以」左側三個詞的範圍內,出現「不」,或右側十個詞的範圍內,出現「為」的語料。」

### 5 關鍵詞(組)(Concordance)語料檢索結果頁面操作

以檢索「逮捕」爲例,關鍵詞或詞組(Concordance)語料檢索結果頁面可能如下:

🚰 Word Sketch Engine - Microsoft Internet Explorer	
檔案(F) 編輯(E) 檢視(Y) 我的最愛(A) 工具(T) 説明(H)	A.
😋 上—頁 • 📀 · 💌 🙆 🏠 🔎 搜尋 🌟 我的最爱 🌮 🔗 🔗 - 🌺 🔟 • 📴 🥵 • 🌋	
網址 🛛 🕘 http://wordsketch.ling.sinica.edu.tw/cws/	💌 🔁 移至 連結 Norton AntiVirus 😡 🗸
Home       Concordance       Word Sketch       Thesaurus       Sketch-Diff       Frequency       Collocation         KWIC/Sentence       View options       Sample       Filter       Sort       IIII       Page 1       of 3661       Mext       Last	Corpus: <b>gigaword2</b> Hits: 7 <b>3213</b> <u>conc description</u>
CNA CMN 199101010146 CNA CMN 199101020107下午 被高雄市 警察局 三民 分局 逮捕 , 晩間 移送法辦。 ( $p>+1$ 九CNA CMN 199101020106 CNA CMN 199101030022 CNA CMN 199101030022 CNA CMN 199101030022 CNA CMN 199101030022 	
<u>e</u>	<ul> <li>網際網路</li> </ul>

### 5.1 左上方第一排五選項

點選上方 Home、Concordance、Word Sketch、Thesaurus、Sketch-Diff 等五選 項,可連結至主頁(Home)與其他檢索功能頁面。

#### 5.2 右上方訊息

右上方藍色方框內,顯示檢索的語料庫(例如,為 chinese\_giga\_trad),以及檢索所得關鍵詞(組)的總筆數,(如,共有 73213筆「逮捕」語料)。

#### 5.3 左上方第二排選項:KWIC/sentence

提供檢索結果顯示的不同選擇,預設值為「以關鍵詞為中心」(key word in context (kwic))的排列方式,呈現檢索結果。如上5中之圖所示。 點選「KWIC/sentence」,可切換至以整句的方式呈現檢索結果。

#### 5.4 左上方第二排選項:顯示選項(View Options)

提供其他檢索結果顯示的選擇。點選此項,將會出現另一頁面,提供使用者更 改結果顯示的設定,包括:

🕽 Word Sketch Engine - Microsoft Internet Explorer
檔案的 編輯 (2) 檢視 (2) 裁約最愛 (4) 工具 (1) 說明 (1) [2] [2] [2] [2] [2] [2] [2] [2] [2] [2]
網址 🕖 🥘 http://wordsketch.ling.sinica.edu.tw/cws/ 🔽 🔁 連結 Norton AntiVirus 💡 🗸
Home       Concordance       Word Sketch       Thesaurus       Sketch-Diff       View       Frequency       Collocation       Corpus: gigaword2 conc description         View options       View       Frequency       Collocation       Conc description       Conc description
Attributes Structures References
Word <headline>         tag       <headline>         Cliplay attributes          For each token       <doc>         KWIC tokens only</doc></headline></headline>
Page size (number of lines): 20 KWIC Context size (number of characters): 40 Change View Options
② 完成

### 5.4.1 標記 (Attributes)

提供「詞」(word)與「標記」(tag)二種可複選的顯示選項。 在此二選項下可進一步設定所有詞彙(For each token)或者只有關鍵詞(KWIC tokens only)才顯示前述選項。

預設値未點選「標記」(tag),則語料呈現時不顯示詞類標記。 例如:

今天下午被高雄市警察局三民分局 逮捕,晚間移送法辨。

可更改設定選項,如:

(1)若點選「詞」(word)、「標記」(tag),以及「只顯示關鍵詞標記」(KWIC tokens only),則會顯示關鍵詞的詞類標記。 例如:

今天下午被高雄市警察局三民分局 逮捕/VC31,晚間移送法辨。

(2)若點選「詞」(word)、「標記」(tag),以及「顯示每個詞的標記」(For each token),則會顯示句中所有詞彙的詞類標記。 例如:

高雄市/Nca 警察局/Ncb 三民/Nc 分局/Ncb <u>逮捕</u>/VC31 , /COMMACATEGORY 晚間/Ndc 移送法辦/VB12

### 5.4.2 結構 (Structures)

結構 (Structures) 欄提供標記開頭(<)與結尾(/>)的結構顯示,如: (1) 標題(headline),例如: <*headline*>大陸主要城市天氣預報<*/headline*>

- (2) 時間(dateline) , 例如: <dateline>( 中央社 紐約 二日 專電 )</dateline>
- (3) 文本(text) , 例如:

<text>針對中共政權迫害知識份子與逮捕劉曉波,

翁山蘇姬目前已被緬甸軍事執政當局逮捕。 </text>

(4) 段落(paragraph),例如:

警方表示, 逮捕 過程中並未如外傳發生槍戰事件。

, 今天下午被高雄市警察局三民分局逮捕
 , 晚間移送法辦。
 (5) 檔案(doc) 例如:

, 並報以熱烈掌聲。</doc><doc>美警方逮捕 十三名涉及 綁架 大陸 人民 (中央社 紐約

#### 5.4.3 參考資料 (References)

參考資料 (References)顯示在結果頁面的左方,以藍色字體呈現,預設值為該 語料出現的檔案代號(doc.id),如<u>CNA19910102.0196</u>。

除檔案代號(doc.id)外,參考資料 (References)欄提供語料來源的相關訊息顯示設定,如檔案類型(doc.type),以及檔案來源(doc.src)等等。

#### 5.4.4 頁面長度 (Page Size)

頁面長度(Page size)欄位,提供自由的(行數 number of lines)設定,預設值為 每頁顯示二十行,但亦可設定成每頁顯示十行或者五百行。(每頁顯示行數越 多,檢索速度可能越慢。)

#### 5.5 左上方第二排選項:樣本(Sample)

樣本(Sample)功能,提供從多量結果語料中,隨機擷取少量語料的功能。 例如,「逮捕」共有 73213 筆,可點選樣本(Sample)鍵,在「樣本行數」 (Number of lines in the sample)中設定 250 或者 500 等,隨機抽取 250 筆或 500 筆 語料,進行觀察與分析。

#### 5.6 左上方第二排選項:過濾(Filter)

過濾(Filter)提供多重過濾功能。可設定「擷取」(positive)或者「過濾」 (negative),搭配詞語(Selected token)出現在關鍵詞之前(first)或之後(last)。 搜尋範距(Search Span)亦可自由設定。可輸入與關鍵詞搭配的詞組(phrase)或 者詞(word)等目標檢索項目。亦可以語料庫檢索語言(Corpus Query Language (CQL))檢索。例如,若如下設定:

🗿 Word Sketch Engine - Microsoft Internet Explorer	
~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	
③ 上-頁 · ⑤ · ☑ ⑥ / 2 搜尋 ☆ 我的最爱 Ø ▷ · 曼 図 · □ ◎ · 恣	
網址 🕖 🗃 http://wordsketch.ling.sinica.edu.tw/cws/	🖌 🄁 移至 🧵 連結 Norton AntiVirus 😡 🗸
Home Concordance Word Sketch Thesaurus Sketch-Diff View Frequency Collocation	Corpus: gigaword2 conc description
Concordance ritter	
Filter:	
	🥑 網際網路

則表示從「逮捕」73213 筆中抽取在關鍵詞「逮捕」左側 5 個詞的範圍內,出現「警方」的語料。則會出現含「警方.....逮捕」的語料 2595 筆。若再如下設定:

🗿 Word Sketch Engine - Microsoft Internet Explorer	
檔案·E 編輯·E 檢視·E 我的最爱(A) 工具(I) 說明·E	an 1997
③ 上─頁 · ● · ▲ ②  ◇ 搜尋 ☆ 我的最爱  ⊗ · ▲ 図 · ▲ ◎ · ▲	
網址 D @ http://wordsketch.ling.sinica.edu.tw/cws/	💽 🔁 移至  連結 Norton AntiVirus 😣 🗸
Home Concordance Word Sketch Thesaurus Sketch-Diff View Frequency Collocation	Corpus: <b>gigaword2</b> conc description
Concordance Filter	
Filter: <ul> <li>positive [] negative</li> <li>selected token:</li> <li>first <ul> <li>last</li> <li>search Span:</li> <li>from 0</li> <li>to 5</li> </ul>            Word Form:         #32         PoS:         unspecified ♥         Match case         CQL:        </li></ul>	
	#E:

則表示從含「警方.....逮捕」的語料 2595 筆中,再抽取在關鍵詞「逮捕」右側 5 個詞的範圍內,出現「嫌犯」的語料。點選「Filter Concordance」後則會出現 含「警方.....逮捕....嫌犯」的資料 245 筆。

# 5.7 左上方第二排選項:排序(Sort)

可利用排序(Sort)左方三鍵, ┋┋┋┋, ӭ┋, ѕб 便設定「依關鍵詞左邊」、「依關 鍵詞」、「依關鍵詞右邊」的詞首排序。亦可點選排序(Sort)鍵進階設定排序條 件。可依照詞類排序,可設定排序詞數,並可做多重排序選項的設定。

### 5.8 頻率 (Frequency)

提供二種相關頻率訊息的設定:

(1) 多層頻率分佈(Multilevel frequency distribution)設定 提供關鍵詞在語料庫中的詞類分佈,與前後詞語搭配的頻率等。例如檢索關 鍵詞「花」,進入「頻率」(Frequency)設定頁面,點選「關鍵詞詞類」 (Node tags)鍵,或者在「多層頻率分佈」(Multilevel frequency distribution)第 一層,屬性(Attribute)下拉選單中選「詞類」(tag),位置(Position)關鍵節點 (node),設定如下:

🚰 Word Sketch Engine - Microsoft Internet Explorer		I I I X
檔案 (P) 編輯 (E) 檢視 (Y) 我的最愛 (A) 工具 (I) 說明 (B)		A.
③ 上─頁 ▼ ② ▼ 図 ◎ ↓ 20 10 10 10 10 10 10 10 10 10 10 10 10 10		
網址 (D) (酱) http://wordsketch.ling.sinica.edu.tw/cws/index.html	💙 🄁 移至	連結 Norton AntiVirus 😡 🗸
Home         Concordance         Word Sketch         Thesaurus         Sketch-Diff         View         Frequency         Collocation           Node tags         Node forms         Doc IDs         Text Types         Text Types         Text Types		Corpus: gigaword2 conc description
Multilevel frequency distribution		
Frequency limit.		
Image: Second level Attribute:       Image: Second level Attribute:		
Text Type frequency distribution		
Frequency limit: 0 doc.id doc.type		~
		網際網路

即可得到「花」以下的詞類分佈訊息:

🗿 Word Sketch Engine – Microsoft Internet Explorer	
檔案·E 編輯·E 檢視·(Y) 我的最爱(A) 工具(I) 說明·E	an 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 19
🌀 上—頁 • 🛞 · 🖹 🐔 🔑 搜尋 🌟 挑的最爱 🤣 🙆 • 🌺 🔟 • 📜 👧 • 🦓	
網址 🛛 🝓 http://wordsketch.ling.sinica.edu.tw/cws/index.html	▼ 秒 移至 連結 Norton AntiVirus 😵 -
Home         Concordance         Word Sketch         Thesaurus         Sketch-Diff         View         Frequency         Collocation	Corpus: <b>gigaword2</b> conc description
Frequency list	
Frequency limit: 0 Set limit	
tag Free Rell%	
VC31 38559 597.1	
Nab 9805 36.1	
VH11 60 0.4	
	(mar) (mar)
(2) 元以	1 網際網路

(2) 文類頻率分佈(Text Type frequency distribution)設定

提供檢索的關鍵詞(組)在文本檔案中分佈的頻率訊息,例如檢索關鍵詞「花」,點選「檔案類型」(doc.type),則可得出如下「花」在不同文類中出現的分佈訊息:

🗿 Word Sketch Engine - Microsoft Internet Explorer	- 7 🛛
檔案(F) 編輯(E) 檢視(V) 我的最愛(△) 工具(I) 説明(田)	At
③ 上一頁 · ③ · ▲ Ø 🏠 🔎 搜尋 🌟 我的最要 Ø 🔗 · ॆ 図 · □	
網址 🕖 🙋 http://wordsketch.ling.sinica.edu.tw/cws/index.html	🔽 🔁 移至 連結 Norton AntiVirus 😡 🗸
Home         Concordance         Word Sketch         Thesaurus         Sketch-Diff         View         Frequency         Collocation	Corpus: <b>gigaword2</b> conc description
Frequency list	
Frequency limit. 0 Set limit	
doc.type Freq Rel [%]	
story 47381 104.9	
multi 579 79.7	
other 447 19.5	
advis 17 6.7 i	
, 2011年1月1日1日1日1日1日1日1日1日1日1日1日1日1日1日1日1日1日1	安全地移住面朝

### 5.9 共現訊息(Collocation)

提供與關鍵詞(組)所搭配詞彙的相關共現訊息,包含 T-score、MI、MI3、log likelihood、min.sensitivity,以及 sailence 等值的設定。

### 5.10 關鍵詞(組)(Concordance)結果頁面的設定

欲檢視其他關鍵詞(組)(Concordance)結果頁面,可在「頁數」(page)欄位中 指定欲檢視的頁數,按「前往」(go)鍵,或者點選「下一頁」(next)、「最後一 頁」(last)、「第一頁」(first),以及「前一頁」(previous)等鍵移動頁面。

### 5.11 察看某行關鍵詞(組)(Concordance)結果

點選結果頁面中紅色關鍵詞(組),例如點選第一行「逮捕」,會在頁面下方顯示更多的語境,如下:

🗿 Word Sketch Engine - Microsoft Internet Explorer						
檔案 [2] 编辑 [2] 依視 (2] 我的最爱 (3) 工具 (1) 說明 (3) [3] [3] [3] [3] [3] [3] [3] [3] [3] [3]						
③ 上-頁 · ③ · Ⅰ 2 4 /2 按尋 ☆ 我的最爱 ④ ② · ↓ 図 · □ 3 · ↓						
網址 🛛 🕘 http://wordsketch.ling.sinica.edu.tw/cws/index.html	💙 🄁 移至	連結 Norton AntiVirus 😵 🗸				
Home       Concordance       Word Sketch       Thesaurus       Sketch-Diff       Frequency       Collocation         KWIC/Sentence       View options       Sample       Filter       Sort       III       IIII       IIIII       IIIII       IIIII       IIIII       IIIII       IIIII       IIIII       IIIIII       IIIII       IIIIII       IIIIII       IIIIIIIII       IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII		Corpus: gigaword2 Hits: 73213 conc description				
CNA CMN 199101010146       下午 被 高雄市 警察局 三民 分局       一方       「方       「方       「方       「方       「方       「方       「方       「う       「う						
<u>CNA CMN 19910104.0226</u> (CNA CMN 19910104.0226 雙十路 二段 四十八巻 民宅 内速捕 通線犯 賴育成(中央社 豐原 雙十路 二段 四十八巻 民宅 内速捕 違反 槍砲 彈藥 管制 條例 通緝犯						
CNA CMN 19910104.0226 甚虚脫、幾乎休克的賴育成逮捕。警方在完成偵訊後,將 CNA CMN 19910104.0226 台中地檢署。答方表示,逮捕過程中並未如外傳發生槍戰		×				
expand left 兒嫌 就達(中央社 高雄一日 電)高雄縣 鳥松 鄉民 蔡仁貴 、 簡 焜 保 兩 人 昨天 深夜 在 縣市 交界 的 醉翁國 人 , 今天 下午 被高雄市 警察局 三民 分局 <mark>逮捕</mark> , 晩間 移送法辦 。 十九 歲 的 許俊銘 身 中 三 刀 , 深及 , 胛骨 遭 砍傷 , 經 長庚 醫院 高雄 分院 緊急 治療 後 , 都 無 生命 危險 。 警方 expand right	δ 餐廳 殺傷 肺部。 二十	許俊銘與 楊世傑 兩 - 歲的 楊世傑 左 肩 ✓				
		🥥 網際網路				

並且點選「擴展左側」(expand left)鍵,或「擴展右側」(expand left)鍵,可以繼續擴展左側或右側語境。

欲知某一筆語料的文本來源訊息,可點選結果頁面左側藍色字體有關文件檔案 訊息的部分,例如在以上檢索「逮捕」的頁面中,點選第一行語料左側的 <u>CNA19910101.0146</u>,則會在頁面下方,顯示檔案代號(doc.id)、檔案類型 (doc.type),以及檔案來源(doc.src)等訊息如下:

🗿 Word Sketch Engine - Microsoft Internet Explorer							
檔案 E 線輯 E 檢視 Y 我的最爱 L 工具 D III 明 E L L L L L L L L L L L L L L L L L L							
③ 上-頁 · ② · ▲ 2 6 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2							
網址 🕖 🙋 http://wordsketch.ling.sinica.edu.tw/cws/index.html	🖌 🄁 移至 🧵 連結 Norton AntiVirus 🚱 🗸						
Home       Concordance       Word Sketch       Thesaurus       Sketch-Diff       Frequency       Collocation         KWIC/Sentence       View options       Sample       Filter       Sort       III       III       III       III       III       III       III       III       IIII       IIII       IIII       IIII       IIII       IIIIIIII       IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	Corpus: gigaword2 Hits: 73213 conc description						
CNA CMN 19910101.0146       下午 被 高雄市 警察局 三民 分局 逮捕 , 晩間 移送法辦。 / CNA CMN 19910102.0107       同年 九月 随即 為 大陸 公安 人員 逮捕 , 而於 去年 十二月 九日         CNA CMN 19910102.0107       日前日 九月 随即 為 大陸 公安 人員 逮捕 , 而於 去年 十二月 九日         CNA CMN 19910102.0107       与本 位 (大幸 大幸 三要 城市 天         CNA CMN 19910103.0022       日前日 法 補匈 軍事 執政 當局 逮捕 , 命>大陸 主要 城市 天         CNA CMN 19910103.0022       二日 專電 ( $\sim$ )> 総約 市警局 昨天 逮捕 十三 名 涉及 納架 大陸 人         CNA CMN 19910103.0026       二日 專電 ( $\sim$ )> 総約 市警局 昨天 逮捕 ( $\rightarrow$ ) 大陸 主要 城市 天         CNA CMN 19910103.0026       二日 專電 ( $\sim$ )> 総約 市警局 昨天 逮捕 ( $\rightarrow$ ) 大陸 主要 城市 天         CNA CMN 19910103.0026       二日 專電 ( $\sim$ )> 総約 市警局 昨天 逮捕 ( $\rightarrow$ ) 大陸 主要 城市 天         CNA CMN 19910103.0026       二日 專電 ( $\sim$ )> 総約 市警局 昨天 逮捕 ( $\rightarrow$ ) 八声 ( $\rightarrow$ )         CNA CMN 19910104.0021       当前 中共 政權 追害 知識 份子 與逮捕 劉助義波 , 義大利 國會 諸         CNA CMN 19910104.0108       注到 ( $\sim$ )> 印尼 以 賣淫 罪名 逮捕 三 名 台灣 女子 (中央社):         CNA CMN 19910104.0108       注動 第名 , 在 稚加達逮捕 ( $\sim$ 大声 素) 國本 ( $\sim$ CNA CMN 19910104.0168       二日 專電 ( $\sim$ )> 印尼 以 賣注 罪名 逮捕 ( $\sim$ 金)> 四方 二         CNA CMN 19910104.0168       二日 專「 ( $\sim$ )         CNA CMN 19910104.0168       二日 專「 ( $\sim$ )         CNA CMN 19910104.0169       二日 專「 ( $\sim$ ) <td>&lt; + 九  &lt; + 九   - → +九   - → → → →   - → →  &lt;</td>	< + 九  < + 九   - → +九   - → → → →   - → →  <						
CNA CMN 19910104.0226         陸委會的運作。         公本         公本 </td <td>」 原 ] 通緝犯 , 將 主 槍戰</td>	」 原 ] 通緝犯 , 將 主 槍戰						
doc.id CNA_CMN_19910101.0146 doc.type story doc.src cna							
bittp://wordsketch.ling.sinica.edu.tw/cws/run.cg/fullief/pos=31/86.corpname=gigaword2	初時網路						

# 6 建立次級語料庫(Creating a subcorpus)

如果您想檢索某個語料庫的一部份, (例如繁體字版的 gigaword2, 只想檢索 story 類型的文本),可在檢索主網頁中, (或者點選「首頁」(Home)鍵, 回到 主頁面),再點選「建立次級語料庫」(Create subcorpus), 在以下視窗中 點選, 並輸入該次級語料庫的名稱即可。

🗿 Word Sketch Engine - Microsoft Internet Explorer	
檔案· E 編輯· E 檢視· E 我的最爱(A) 工具· E 說明· E	
③ 上-頁 · ② · ▲ ②  ◇ 搜尋 ☆ 我的最爱  ※ ◎ · ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	<b>k</b>
網址 D) 🕘 http://wordsketch.ling.sinica.edu.tw/cws/index.html	🔽 🔁 移至 連結 Norton AntiVirus 😵 🗸
Home Concordance Word Sketch Thesaurus Sketch-Diff	
Commer nineword2	
New subcorpus name:	
Document counts Tokens	
doc.type # doc.src #	
I advis 907390 I cna 808060584	
Dother 34200589	
▼ story 754500611	
Create Subcorpus	
	御際網路
如欲檢索該次級語料庫,可在檢索頁面的「語料庫	(Corpus) 下拉選單中,選

如欲檢索該次級語料庫,可在檢索頁面的「語料庫」(Corpus)下拉選單中,選 取該語料庫,進行檢索。

如欲刪除該次級語料庫,可在檢索首頁中(點選「首頁」(Home)鍵,回到主頁面),點選「刪除次級語料庫」(Delete subcorpora),再點選欲刪除的語料庫名稱,進行刪除即可。

### 7 詞彙特性速描(Word Sketch)功能

「詞彙特性速描」(Word Sketch)是以語料庫為本對某個詞彙的語法以及和其他詞 語搭配表現的一覽表。點選首頁中的「詞彙特性速描」(Word Sketch),會出現 以下頁面:

🗿 Word Sketch Engine – Microsoft Internet Explorer		
檔案·E 编辑·E 檢視·E 微說·E 武明·E 説明·E		<b></b>
🔇 上—頁 • 🚫 · 🗷 🛃 🏠 🔎 搜尋 🌟 挑的最爱 🤣 🙆 • 🌺 🔟 • 🔜 🥵 • 🦓		
網址 (D) 🕘 http://wordsketch.ling.sinica.edu.tw/cws/index.html	✓ → 移至 連結	Norton AntiVirus 😵 🗸
Home     Concordance     Word Sketch     Thesaurus       Sketch     Entry     Form		
Corpus: gigaword2 🗸		
Word Form:		
Sort grammatical relations: 🔽		
Minimum frequency: 5		
Minimum salience: 0.0		
Maximum number of items 25		
Show Word Sketch		
<b>劉</b> 完成		網際網路

可在「語料庫」(Corpus)的下拉選單中,點選所欲察看的語料庫,在「詞彙形式」(Word Form)中輸入欲察看的詞彙,可選擇是否按照語法關係排序(Sort grammatical relations)、設定搭配詞語出現最低的頻率限制(Minimum frequency)、最低的顯著性(Minimum Salience),以及某個語法關係中最大的詞項數量(Maximum number of items)限制。

此功能適用於大型語料庫,若語料庫太小,則不適用此功能。

若在預設値設定下,輸入「逮捕」查詢,則會出現以下結果頁面:

🔮 Word Sketch Engine - Microsoft Internet Explorer										
檔案 (E) 编辑 (E) 檢視 (Y) 我的最爱 (A) 工具 (D) 說明 (B) 🧗										
G	③ 上-頁 ▼ ② · 区 ③ 《 20 按章 ☆ 我的最爱 ④ ② · 曼 圖 · □ ◎ · ▲									
網址①	網址 🕐 🎒 http://wordsketch.ling.sinica.edu.tw/cws/index.html 🔍 🎔 野 移至 🏾 连结 🛛 Norton Anth Virus 😵 •									
Hor	Home Concordance Word Sketch Thesaurus Sketch-Diff									
读捕 gigaword2 freq = 73213										
PP	將 4349 74.6	PP 被	63 13.6	Subject	11755 4.7	Object	41550 3.9	PP 從 24 3.9		
他	1722 53.83	指控	13 29.61	警方	3659 75.45	嫌犯	<u>3015</u> 75.23	父子檔 5 34.15		
嫌须	2 276 51.72	法庭	8 22.89	現行犯	<u>82</u> 52.95	名	<u>4698</u> 48.0	_		
陳頻	≹ <u>53</u> 46.78	·	_	罪名	<u>188</u> 45.44	嫌疑人	<u>182</u> 45.8			
黃姆	<mark>≹ <u>36</u> 43.6</mark>	PP_b	<u>133</u> 7.4	日警	<u>29</u> 43.18	嫌疑犯	<u>179</u> 45.63			
三娛	<mark>≹ <u>30</u> 41.8</mark>	嫌犯	<u>24</u> 34.46	當局	<u>521</u> 42.06	三嫌	<u>77</u> 44.77			
李娟	<mark>≹ <u>28</u> 40.87</mark>	罪嫌	<u>7</u> 20.71	警察	<u>305</u> 36.98	現行犯	<u>79</u> 44.29			
張頻	≹ <u>25</u> 38.29			幹員	<u>80</u> 36.96	嫌	<u>260</u> 43.81			
周頻	≹ <u>15</u> 35.11			川督趙爾豐	. <u>10</u> 34.08	毒販	<u>134</u> 42.98			
林婉	<mark>≹ <u>21</u> 34.38</mark>			公安	<u>168</u> 32.46	陳嫌	<u>91</u> 41.83			
嫌	<u>49</u> 34.12			警網	<u>38</u> 32.03	兒嫌	<u>135</u> 39.92			
劉娛	≹ <u>15</u> 32.6			員警	<u>93</u> 30.28	科学研究	<u>101</u> 37.9			
調頻	₱ <u>13</u> 32.31			間課罪	<u>16</u> 29.5	李啸	<u>49</u> 37.9			
徐娟	<u>≹ 11</u> 31.42			殊犯 ## [ #######	<u>102</u> 29.32	東嫌	<u>52</u> 37.12			
土州	≹ <u>15</u> 31.3 ¥ 00 00 00			/////////////////////////////////////	<u>19</u> 29.06	1丁動 3日開催怒!	<u>1231</u> 36.92			
見が	≹ <u>28</u> 30.68 ≇ 10.00.50			中六	428 28.90 20 00 2	クビラド外市先正/ アビカ単	<u>120</u> 30.79			
	* IU 29.59			=	<u>28</u> 20.5	****	47 30.42			×
 (合) 完成									網際網路	
🥙 完成									🔮 網際網路	

以上頁面中,顯示「逮捕」的語法搭配關係,如賓語(object)、主語「subject」、修飾語(modifier),以及搭配的「以」字介詞組等。每個欄位下,顯示搭配詞,與搭配詞出現的總頻率,與搭配詞出現的顯著性。例如「嫌犯」出現當「逮捕」的賓語共有 276筆,其顯著性為 51.72。以上結果頁面,按照搭配詞的顯著性,由高至低排序。

「詞彙概述」(Word Sketch),由統計方式自動抽取搭配詞關係,如其他檢索系統一樣,允許某些程度的噪音(noise),即可能擷取到不正確的結果。

使用者可以隨時點選頁面上排的功能鍵,轉換到其他功能檢索頁面。

#### 8 同義辭典(Thesaurus)功能

提供在搭配詞語上,與關鍵詞表現相近的詞彙訊息。

### 9 詞彙素描異同(Word Sketch Difference)功能

提供二個詞彙的語法功能、搭配詞語異同的訊息。並且以顏色顯示此二比較詞彙,在搭配某個詞彙上的顯著程度。可點選此功能鍵,進入以下頁面:

🗿 Word Sketch Engine - Microsoft Internet Explorer	
檔案(P) 編輯(E) 檢視(Y) 我的最愛(A) 工具(I) 說明(H)	At
🔇 上—頁 • 🚫 · 🖹 🖉 🏠 🔎 搜尋 🌟 我的最爱 🤣 😥 · 🌺 🗃 • 🔲 🥵 • 🦓	
網址 D 🗿 http://wordsketch.ling.sinica.edu.tw/cws/index.html	💌 🄁 移至 連結 Norton AntiVirus 😡 🗸
Home         Concordance         Word Sketch         Thesaurus         Sketch-Diff           Word         Sketch         Differences         Entry         Form	
Corpus: gigaword2 🗸	
First lemma:	
Second lemma:	
Sort grammatical relations:	
Separate blocks:	
Minimum frequency: 5	
Maximum number of items in a grammatical relation 25 of the common block:	
Maximum number of items in a grammatical relation 12 of the exclusive block:	
Show Diff	
ê	網際網路

在「第一個詞彙」(the first lemma),與「第二個詞彙」(the second lemma)欄位中輸入所欲比較詞彙,可細部設定所欲查詢的條件,以及顯示的方式。

# ● 網路資源:

The Word Sketch Engine: http://www.sketchengine.co.uk/ Chinese Gigaword 訊息:

http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2003T09