

從蘋果到奇異果，詞彙夥伴在哪裡？——語料庫科技在語文教學的應用(II)

【語文教育及編譯研究中心助理研究員 吳鑑城】

因語料庫呈現語言的實際使用情形的特性，多年來已被普遍的運用於許多的教學、研究之中。近年來，電腦運算能力的提昇以及深度學習理論的成熟，更促使人工智慧更加成熟，提供各行各業更多便利的輔助。在 173 期，已先介紹了國教院語料庫索引典系統、例句編輯輔助系統，作為語料庫科技在語文教學應用的示範，本文所將介紹的語義場關聯詞系統更是結合語料庫及人工智慧的輔助教學應用。

所謂的語義場，簡單來說，就是一組詞彙在詞義上所具有的類型關係，例如「炎熱」和「寒冷」屬於意思相反的反義語義場，「牛」和「馬」屬於（動物）類型語義場。檢索本系統所獲得的詞彙，都將跟輸入的詞彙有著一定的關係，而這也是本系統為何稱為語義場關連詞的由來。接著，就讓我們來實際看看它的神奇之處。

當輸入蘋果進行查詢，系統傳回了許多的關聯詞，包括藍莓、葡萄、橘子、柳丁……等（見圖 1）。聰明的你，應該已經發現，回傳的詞彙彷彿是一份水果列表。對於還在想明天要教導學生哪些水果的老師而言，這不正是一份現成的素材嗎！

選擇語料庫	查詢詞(組)語義場關聯詞	輸出
<input checked="" type="radio"/> 遠流語料 <input type="radio"/> 中國時報 <input type="radio"/> 國語日報 <input type="radio"/> 平衡語料庫 <input type="radio"/> Chinese Giga Words	輸入 正關聯詞(組): 蘋果 負關聯詞(組): 請輸入負關聯詞，例如：賈伯斯 輸出詞數: 20	語義場關聯詞: 藍莓 0.6044 -- 檸檬 0.5820 4,進階 橘子 0.5804 3,基礎 柳橙 0.5728 -- Apple 0.5664 -- 梨子 0.5659 -- 桃子 0.5615 5,進階 草莓 0.5603 3,基礎 柳丁 0.5554 -- 葡萄 0.5551 3,基礎 Mac 0.5506 -- 櫻桃 0.5392 5,進階

圖 1. 「蘋果」一詞在「遠流語料」中獲得的查詢結果

選擇語料庫	查詢詞(組)語義場關聯詞	輸出
<input type="radio"/> 遠流語料 <input checked="" type="radio"/> 中國時報 <input type="radio"/> 國語日報 <input type="radio"/> 平衡語料庫 <input type="radio"/> Chinese Giga Words	輸入 正關聯詞(組)： 蘋果 負關聯詞(組)： 請輸入負關聯詞，例如：賈伯斯 輸出詞數: 20 <input type="button" value="送出"/>	語義場關聯詞: APPLE 0.6890 -- 橘子 0.6484 3,基礎 Apple 0.6471 -- iPhone 0.6454 -- 黑莓 0.6450 -- iPad 0.6439 -- IPHONE 0.6381 -- IWATCH 0.6265 -- IPAD 0.6240 -- IPHONE5 0.5997 -- 賈伯斯 0.5873 -- 戴爾 0.5866 -- 賈伯斯 0.5856 -- WIN8 0.5828 --
選擇分級詞表 <input type="radio"/> 華測八千詞 <input checked="" type="radio"/> 國教院分級詞表		
選擇詞彙等級 <input checked="" type="radio"/> 不限制 <input type="radio"/> 查 7 級(含)以下詞		

圖 2.「蘋果」一詞在「中國時報」中獲得的查詢結果

讓我們試著把語料從書籍類型轉換成新聞類型，再度檢索的結果，水果們不見了，取而代之的，是 iPhone、iPad、iWatch、Win8 等資訊產品（見圖 2）。同一個詞，兩次檢索為何有著天壤之別？別太訝異，當選擇不同的語料庫，就像是跟不同學習、成長背景的人討論一樣，會根據他們熟悉的領域，提供不同的意見，誰讓蘋果本身就是跨領域的呢！但，既然應用了人工智慧技術，當然我們也可以試著跟它對談一番。不是想要資訊產品，那試著在負關聯詞欄中輸入 iPhone，告訴它，我們不是想看到像 iPhone 這樣的資訊。檢索後，它很貼心的換成橄欖、杏仁、奇異果等再度比較貼近「水果食物義」的關聯詞（見圖 3）。

選擇語料庫	查詢詞(組)語義場關聯詞	輸出
<input type="radio"/> 遠流語料 <input checked="" type="radio"/> 中國時報 <input type="radio"/> 國語日報 <input type="radio"/> 平衡語料庫 <input type="radio"/> Chinese Giga Words	輸入 正關聯詞(組)： 蘋果 負關聯詞(組)： iPhone 輸出詞數: 20 <input type="button" value="送出"/>	語義場關聯詞: 橄欖 0.4808 -- 杏仁 0.4384 7,精熟 菠菜 0.4207 -- 葡萄乾 0.4134 -- 胡蘿蔔 0.4122 -- 核桃 0.4111 7,精熟 大王 0.4108 5,進階 紅蘿蔔 0.4004 -- 奇異果 0.3961 -- 蕃茄 0.3910 -- 堅果 0.3885 -- 果仁 0.3848 -- 無花果 0.3832 -- 檸檬 0.3790 4,進階
選擇分級詞表 <input type="radio"/> 華測八千詞 <input checked="" type="radio"/> 國教院分級詞表		
選擇詞彙等級 <input checked="" type="radio"/> 不限制 <input type="radio"/> 查 7 級(含)以下詞		

圖 3.加入「iPhone」作為負關聯詞獲得的查詢結果

讓我們再換個情境：當學生好奇並想要多學習在平常的生活情境中會運用的詞彙時，他查詢了「廚房」，系統就提供了流理台、客廳、浴室等關聯詞。在看到流理台後，他更加好奇還有哪些詞彙會跟廚房有關，此時，只要同時輸入流

理台一詞，系統就提供了瓦斯爐、洗碗槽、抽油煙機等等更多廚房相關的設備（見圖4），進一步豐富個人詞彙。而老師們更可利用此一檢索方式，輔助設計不同的情境詞彙組，如地點類型（客廳、浴室等）、設備類型等，再加上本院所研訂的對外華語教學適用的詞語分級資訊（基礎、進階、精熟），更可做為詞彙難易度參考，冀能縮短教材設計時間，也提昇教材之品質。

選擇語料庫		
<input type="radio"/>	遠流語料	
<input checked="" type="radio"/>	中國時報	
<input type="radio"/>	國語日報	
<input type="radio"/>	平衡語料庫	
<input type="radio"/>	Chinese Giga Words	

選擇分級詞表		
<input type="radio"/>	華測八千詞	
<input checked="" type="radio"/>	國教院分級詞表	

選擇詞彙等級		
<input checked="" type="radio"/>	不限制	
<input type="radio"/>	查 7 級(含)以下詞	
<input type="radio"/>	查 8 級(含)以下詞	

查詢詞(組)語義場關聯詞		
輸入		
正關聯詞(組):		
廚房 流理台		
負關聯詞(組):		
請輸入負關聯詞，例如：賈伯斯		
輸出詞數: 20		
<input type="button" value="送出"/>		
輸出		
語義場關聯詞:		
瓦斯爐	0.7299	--
洗碗槽	0.7282	--
抽油煙機	0.7227	--
水槽	0.6910	--
浴廁	0.6864	--
客廳	0.6794	2,基礎
浴室	0.6791	3,基礎
料理台	0.6595	--
臥室	0.6584	3,基礎
飯廳	0.6540	2,基礎
爐台	0.6500	--
洗手台	0.6407	--
餐桌	0.6404	3,基礎
鍋子	0.6392	4,進階
廚櫃	0.6372	--

圖 4.同時查詢「廚房」以及「流理台」獲得的查詢結果

資料來源：

林慶隆、柯華葳、吳鑑城、白明弘、陳茹玲（2019）。《建置應用語料庫及標準體系》期末研究報告。國家教育研究院研究計畫成果報（編號：NAER-107-12-F-1-01-00-1-11）。新北市：國家教育研究院。

*全文下載網址：<https://www.grb.gov.tw/search/planDetail?id=12562546>