

TranXearch：專門為譯者而設的智能搜尋引擎

蕭世昌

摘要

搜尋引擎是重要的翻譯工具。然而，目前的引擎並非專為譯者而設，搜尋所得未必切合翻譯所需，使用也難免有所不便。本文遂提出一個專為譯者而設的搜尋引擎，名為 TranXearch (www.tranxlab.com/tranxearch)，令譯者更快找到翻譯各種相關資料，提升效率。更具體來說，譯者輸入字詞語句或段落，搜尋引擎自動「參考」不同翻譯工具，製作「翻譯報告」，彙整各類資料，方便譯者分析及翻譯原文。本研究強調翻譯與搜尋引擎之融合，冀有助於線上翻譯工具的發展及整合。

關鍵詞：搜尋引擎、翻譯系統、翻譯科技、電腦輔助翻譯

蕭世昌，恒生管理學院翻譯學院助理教授。

本文部分初稿內容曾發表於「2013 臺灣翻譯研討會——科技與翻譯發展」，作者在此向該場研討會主持人董大暉教授、評論人史宗玲教授及所有與會人士，以及本文的兩位匿名審查者所給予的修正意見致謝。

TranXearch: An Intelligent Search Engine for Translators

Sai-Cheong Siu

Abstract

Commercial search engines have become an important translation tool for translators. However, existing systems, few of which are specifically for translation, may not be always translator-friendly, and the information provided may not be directly relevant to their translation tasks. This paper, therefore, presents TranXearch, an intelligent search engine for translators. It is an online service providing translators with one-stop access to linguistic information on translation. The system also allows users to make multilingual queries by using not only keywords but also sentences and paragraphs. More specifically, after receiving a query, the tool automatically “consults” an array of translation tools and generates a “translation profile”, which provides the following information that helps analyze and translate the ST: (a) a word-for-word translation of the input; (b) similar expressions from translation memory databases; (c) useful sentence patterns that can be used in translation; (d) definitions as well as multilingual equivalents of important terms in the query; and (e) other relevant details that might facilitate the comprehension of the ST and the preparation of the TT. With the above features, the engine could enable translators to conduct research more effectively and efficiently.

This paper is divided into 4 sections: Section 1 is the introduction. Section 2 discusses the architecture, development and features of TranXearch. This is followed by a system demonstration in Section 3, which illustrates the use of the engine. We conclude with future work in Section 4.

TranXearch will be available on the Internet at www.tranxlab.com/

tranxearch. We hope that TranXearch, the name of which symbolizes the fusion of translation and search engines (i.e., “Tran” X “Search”), will encourage future progress on the development and integration of web-based translation tools.

Keywords: search engines, translation systems, translation technology, computer-aided translation

壹、引言

翻譯科技發展迅速，譯者工作時可運用各種翻譯工具，包括機器翻譯、翻譯記憶庫及術語庫等，提升翻譯速度及質量，而有關工具之應用及教學，不少學者已作深入探討，如 Somers (2003)、史宗玲 (2004, 2013) 及 Quah (2005) 等，在此不再贅述。除了上述專門工具外，譯者往往也利用搜尋引擎查閱線上資料，一方面方便分析原文，另一方面尋找更恰當的譯法，搜尋引擎遂成為了重要的翻譯工具。然而，目前的搜尋引擎並非專門為譯者而設，搜尋所得未必切合翻譯所需，使用也難免有所不便。為此本文提出一個專為譯者而設的搜尋引擎，名為 TranXearch，務求讓譯者更快找到翻譯相關資料，提升效率。

本論文共分四節：第壹節為引言。第貳節深入探討 TranXearch 的概念與架構，說明系統特色及開發方法。第參節系統測試，展示工具的運作情況。第肆節總結，討論未來研究方向。

貳、TranXearch 之概念及架構

一、TranXearch 之緣起：現有搜尋引擎之不足

自互聯網及搜尋引擎普及以來，不少研究討論了如何運用搜尋引擎協助實際的翻譯工作，如朱明炬、謝少華 (2003) 及王勇 (2005) 等。此外，也有研究提出多語的翻譯搜尋引擎，如劉東飛、陳智賢、鄧俊濤及余俊 (2007) 與吳丹 (2009)。此舉擴充了傳統單語搜尋引擎的性能或搜尋範圍，然而輸出資料之類型或方式，不少仍與一般的搜尋引擎或雙語索引器相似，未必能全面解決搜尋引擎應用於翻譯時的問題，以滿足譯者需求。

事實上，從譯者的角度而言，現有的搜尋引擎一般有以下問題：首先，互聯網網頁眾多，搜尋引擎常常提供大量網頁連結，譯者難以從中找到真正相關的資料。其次，搜尋引擎以搜尋網頁為主，較少兼及其他常見

的翻譯工具或數據庫（如翻譯記憶庫及術語庫等），不少有用的資料無法提供。譯者因此不能單靠搜尋引擎，而是需要來回不同工具，取得所需資訊。此外，搜尋引擎多僅支援關鍵詞搜尋，也就是說譯者若有疑問，便要以合適的關鍵詞表達，才會較容易找到答案；若所譯句子或段落有多處疑問，譯者更須自設多個關鍵詞，分別搜尋，方能釋除疑惑。

現舉一例以說明有關情況。設有原文如下，內容介紹香港的衍生產品市場：

衍生產品市場方面，香港期貨交易所（期交所）及聯交所提供一系列期貨及期權產品，包括指數期貨、股票期貨、利率期貨、債券期貨、黃金期貨、指數期權及股票期權。在兩個交易所內進行的交易透過三間互連的結算公司進行交收及結算，該三間公司分別是香港中央結算有限公司、香港聯合交易所期權結算公司及香港期貨結算公司。（香港特別行政區政府，2013年8月）

若譯者對文中金融術語有疑問（例如香港期貨交易所、指數期貨、股票期貨、利率期貨、債券期貨、黃金期貨、指數期權及股票期權等），欲加深究，則可利用搜尋引擎。但若把整段文字輸入，則系統可能把整段文字視作一個關鍵詞，造成關鍵詞過長，出現傳輸或搜尋問題。以谷歌搜尋引擎為例，輸入上述文字後，系統發出以下警告：

The requested URL /... is too large to process. That's all we know.（谷歌搜尋引擎，2013年10月16日之搜尋結果）

即使只搜尋引文第一句，系統仍出現以下信息：

“期貨”（及其後字詞）無法查詢，因為 Google 的查詢上限是 32 個字詞。（谷歌搜尋引擎，2013 年 10 月 16 日之搜尋結果）

如此一來，分開輸入關鍵詞是解決方法之一。然而，若原文關鍵詞或術語甚多，用家首先要自行辨識抽取，然後逐一在搜索引擎輸入，此舉對譯者而言實在不甚方便。

用家即使改為分開搜尋關鍵詞，搜尋引擎仍有改善空間。如用家輸入「股票期權」一語，谷歌的搜尋結果見圖 1（僅顯示結果的第一頁）。

[可供買賣的股票期權名單](#)
www.hkex.com.hk > 主頁 > 產品及服務 > 衍生產品 > 股票期權 ▾
編號. 股票編號. 正股名稱. HKATS 代號. 合約買賣單位(股數). 正股買賣單位倍數. 期權類別. 持仓限額(合約張數). 1. 1288. 中國農業銀行股份有限公司. ABC. 10,000.

[股票期權 - Hong Kong Exchanges and Clearing Limited](#)
www.hkex.com.hk/stockoptions/ ▾
2013年4月9日 - 香港交易所股票期權專頁為投資者提供一個全面的股票期權資訊平台。初學者可以先參閱期權ABC及香港股票期權投資指南以學習期權基礎知識。
[股票期權搜尋](#) - [期權工具](#) - [十大成交股票期權系列](#) - [交易資料](#)

[PDF] [理解股票期權](#)
www.hkex.com.hk/chi/prod/drprod/so/.../Stock%20Option_c.pdf ▾
本冊子旨在介紹如何使用於交易所買賣之股票期權作... 股票期權的特點與任何其他相關資產的期權相似。然... 方面與其他並非以股票作為相關資產的期權顯著不同。

[PDF] [香港股票期權投資指南 - Hong Kong Exchanges and Clearing Limi...](#)
www.hkex.com.hk/chi/prod/drprod/so/.../Stockoption_handbook_c.pdf ▾
股票期權最吸引投資者之處是其槓桿作用，因為投資者買賣合約時只需付出合約總值的... 股票期權在現今的全融市場中已被廣泛應用，在不同的市況之中，無論利用其...

[股票期權簡介 | 股票期權 | 產品及服務 | 輝立證券集團](#)
www.poems.com.hk/zh-hk/product-and-service/stock-options/ ▾
股票期權的買家與賣家。投資者可以選擇買入(LONG)或沽出(SHORT)股票期權。買家-擁有權利但無義務在到期日前(Expiration Date)以行使價(Strike price)跟賣方...

[股票期權 - MBA智庫百科](#)
wiki.mbalib.com/wiki/股票期權 ▾ 轉為繁體網頁
股票期權 (Stock Option) 股票期權交易是70年代才發展起來的一種新的股票交易方式，在美國的普遍使用是在90年代初期。股票期權一般是指經理股票...

(圖續下頁)

[股票期權簡介 - 致富證券有限公司](#)
www.chiefgroup.com.hk/options/o_intro.php ▾
 股票期權簡介。期權合約賦予合約持有人在合約到期日以指定價格買入或沽出相關股票的權利。購入期權合約者需要付出金額，(俗稱「期權金」Premium)以換取這個 ...

[股票期權令活化你也可以「做輪商」! - 經濟通ET Net](#)
www.etnet.com.hk/www/tc/news/topic_news_detail.php?category... ▾
 2013年5月2日 - 港交所(00388)早前宣布一系列改革股票期權措施，由今日(2日)起，股票期權交易費將下調，又加強莊家服務及新增到期月份，務求活化股票期權市場 ...

[股票期權服務 - 英皇證券集團有限公司Emperor Capital Group Limited](#)
www.emperorcapi.com/tch/product/stock-option.php ▾
 有關可供買賣的股票期權名單，請參閱由香港交易所不定時更新的資料。... 客戶可參考由香港交易所延續發放之股票期權最低按金詳情及市場報價(資料每日更新)。

[股票期權\(頁1\) - 衍生工具討論區- 香港討論區- Powered by Discuz ...](#)
www.discuss.com.hk > 香港討論區 > 衍生工具討論區 ▾
 2009年10月4日 - 30 篇文章 - 9 位作者
 第一次玩股票期權，預十萬蚊玩首先開倉入貨，short put 10月2628@32 1張收期權金\$760
 希望每個月會帶黎千零蚊期權全收入香港討論區- Discuz!

圖 1：「股票期權」的搜尋結果（谷歌搜尋引擎，2013 年 10 月 16 日）

如圖 1 所示，譯者要知道「股票期權」的含義，不能從搜尋結果頁面直接找到答案，而是需要先進行至少兩重的篩選工序。其一，用家須先檢視搜尋結果中各網頁的內容摘要，判斷哪些網頁較有可能提供所需資料。若第一頁沒有合適網站，則還要瀏覽其他結果頁面。其二，找到可能有用的網站後，還要通過超連結訪問外部網站，再檢查網站內容能否確實提供所需資料。簡言之，用家一方面要排除不合適的搜尋結果，另一方面要檢視外部網頁的內容，進行重重「瀏覽—篩選」的工作，才能解答疑難。如此譯者難以集中心力解決文字問題，甚至可能花了大量時間瀏覽搜尋結果及訪問外部網站，卻仍空手而回。

此外，譯者若不只是要了解個別用語的意思，還希望知道其譯法，則可能需要作另一次搜尋，並修飾關鍵詞。例如譯者若要知道「股票期權」的英譯，從圖一所見，單靠搜尋「股票期權」一語並不能知道譯文。此時若對關鍵詞稍加修改，如改作「股票期權英文」或「股票期權

English」，再作搜尋，則有較大機會找到答案，但此舉也增加了譯者搜尋資料的時間。

從上可見，一般引擎搜尋過程較為繁瑣，往往需要多重搜尋以取得翻譯所需的資料，加上篩選搜尋結果及網站資料需時，未必能完全方便譯者。因此，本文作者提出 TranXearch 系統，冀補上述不足。接下來討論有關系統的概念、特色及架構。

二、TranXearch 的概念及重要特色

TranXearch 屬於專為譯者而設的搜尋引擎，旨在方便譯者取得各種翻譯相關資訊。引擎將於網路開放（www.tranxlab.com/tranxearch）。操作流程見圖 2：

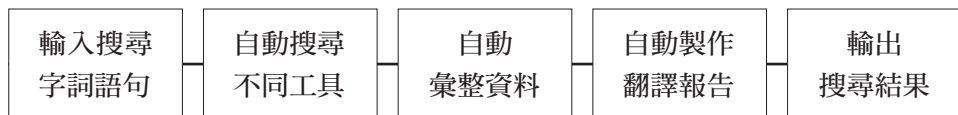


圖 2：TranXearch 的基本操作流程

如圖 2 所示，譯者首先在指定位置輸入要搜尋的字詞語句，這與傳統搜尋引擎相似。但系統取得用家的輸入後，並非如傳統引擎般尋找含有該語句的網頁，而是自動「參考」不同翻譯工具（如辭書及翻譯記憶庫），然後製作「翻譯報告」，報告彙整各類資料，方便譯者分析及翻譯原文。報告內容包括：（一）用家所輸入語句之逐詞翻譯；（二）翻譯記憶庫中相似的語句；（三）有用句型；（四）語句中關鍵術語之解釋及翻譯；以及（五）其他背景資料。報告備妥後向用家輸出，即為搜尋結果。系統內部的設計及運行情況於下一節「系統架構及開發」詳論。

TranXearch 系統旨在令譯者更快找到翻譯相關資料，提升效率。更進一步來說，本系統具以下重要特色：第一，系統允許同時輸入不同語言（如香港常見的中英夾雜句子），也可同時輸出不同語言的結果，用

家可進行跨語言搜尋，更具彈性。再者，此系統除了傳統的關鍵詞外，還支援整句甚至整段文字的輸入，用家只需搜尋一次，即可取得整句或整段的資料，省卻逐詞搜尋之勞。此外，本系統以「翻譯報告」彙整資料，而並非僅提供大量網頁或文件連結，用家參考報告，可專心翻譯工作，不必費神親自篩選有用連結。最後，系統參考不同翻譯資源，包括內部數據庫及外部網頁，然後一次顯示結果，譯者不用穿梭往返各種翻譯工具。

三、系統架構及開發

TranXearch 系統的架構如圖 3 所示：

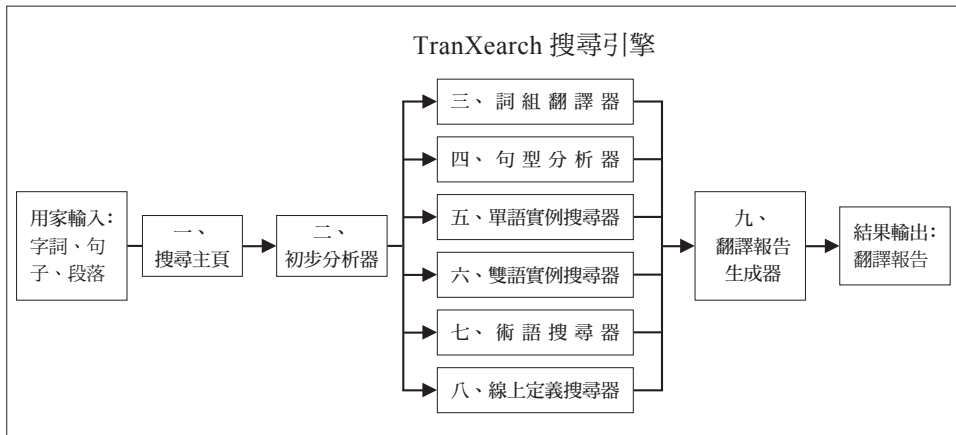


圖 3：TranXearch 的內部架構及主要部件

從圖 3 可見，系統共設九個部件。用家首先在搜尋主頁（部件一）輸入文字，然後由初步分析器（部件二）作基本處理，再由另外六個部件（部件三至八）分別搜尋不同資訊，如單語及雙語實例，以及句型等。各部件的搜尋結果由翻譯報告生成器（部件九）統一處理，製成翻譯報告，送回搜尋主頁，供用家閱覽。

各部件的內容及開發情況說明如下：

（一）搜尋主頁：如圖 4 所示，此為用家與系統互動的主要地方，

設有文字框，供用家輸入欲搜尋之字詞語句。用家點擊開始按鈕後，系統即處理所輸入的文字。主頁同時預留地方顯示搜尋結果。

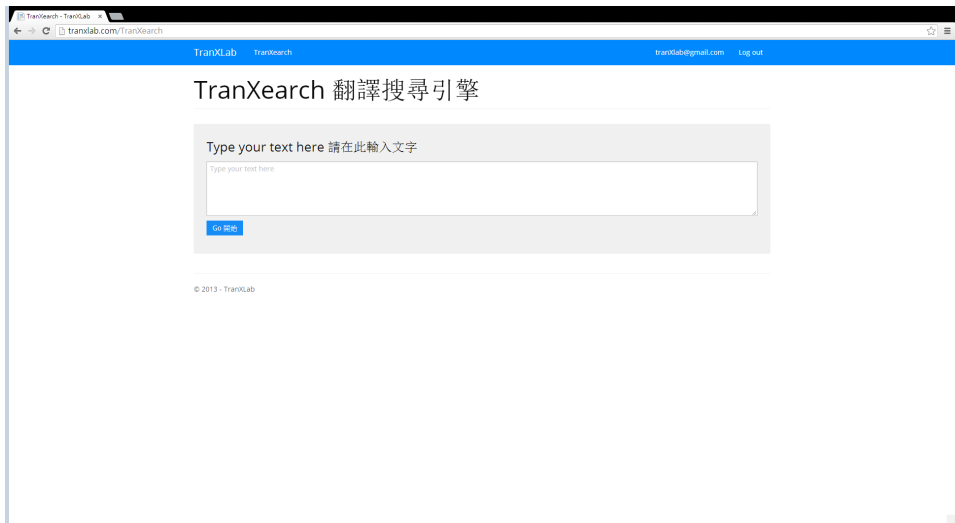


圖 4：TranXearch 搜尋主頁

(二) 初步分析器：此分析器將所輸入的文字分割為句子，以便系統內其他部件處理。比如說，用家輸入一段文字，內含數個句子，初步分析器則先把文字劃分成句子，然後每句再依次交由其他部件分析，搜尋各句的相關資料。搜尋結束，翻譯報告也以句子為單位，顯示結果，如此譯者即使輸入一大段的文字，也可按句檢視搜尋所得的資料，而不必一次過消化整段文字的搜尋內容。

(三) 詞組翻譯器：此翻譯器內設雙語辭書及翻譯引擎，能把輸入的句子逐詞對譯，尋找解釋，方便譯者了解原文各詞的含義，省卻逐詞查閱辭書的工夫。處理方法是先作分詞，如採用最大匹配算法 (Maximum Matching) (劉源、譚強、沈旭昆，1994)，再逐詞翻譯及尋找解釋。雙語辭書內容主要根據 MDBG 之英漢詞典，共收錄 108,112 項條目 (詳情可參閱網站：<http://www.mdbg.net/chindict/chindict.php>)；此外也有筆者日常教研收集所得的詞彙。

(四) 句型分析器：此分析器內含句型庫及句型匹配引擎，匹配引擎能根據輸入的文字，從句型庫尋找合適的翻譯句型，供譯者使用。目前句型庫內之句型，主要由筆者自行從雙語的香港政府文件及財經文獻分析所得。

(五) 單語實例搜尋器：內含中英單語語料庫，可針對輸入文字中各詞尋找原語及譯語的實例。比如說「股票期權」，經上述「詞組翻譯器」可知英譯為「Stock Options」，則實例搜尋器分別在中文及英文的單語語料庫尋找二詞的實例，列出含有「股票期權」或「Stock Options」的句子。搜尋結果的用途有二：其一，在原文分析階段，用家分析各原語實例，更深入了解個別詞彙在不同語境下之意思，以便更準確把握原文的含義；其二，在譯文製作階段，用家參考各譯語的實例，確保譯文用詞得宜。中文實例的部份內容選錄自香港政府出版的《香港便覽》及中文維基百科，英文實例則部份參考美國政府網頁。

(六) 雙語實例搜尋器：內設翻譯記憶庫及翻譯匹配引擎。原理跟一般翻譯記憶系統相似，翻譯匹配引擎首先比較輸入文字及記憶庫中句子，計算相似度，然後找出若干最相似者，供譯者修改重用。相似度的計算方法眾多，如編輯距離 (Levenshtein distance) 及 n 元語法方法等 (王正，2011)，而以下方法則專門為本搜尋器而設：

設 $\text{Similarity}(A,B)$ 為兩個句子 A 及 B 的相似度，則有

$$\text{Similarity}(A,B) = \frac{1}{2} \left(\frac{\sum_{m=1}^{|A|} \text{CharSimB}(a_m)}{|A|} + \frac{\sum_{n=1}^{|B|} \text{CharSimA}(b_n)}{|B|} \right)$$

其中

(續下頁)

1. A 為記憶庫中某句子，由 $|A|$ 個字符 a_m 組成，即 $A = a_1 a_2 \dots a_{|A|}$ ，且定義字符 a_m 在句中之絕對位置為 m ，相對位置為 $\text{rpos}(a_m) = \frac{m}{|A|}$ 。

2. B 為新輸入的句子，由 $|B|$ 個字符 b_n 組成，即 $B = b_1 b_2 \dots b_{|B|}$ ，且定義字符 b_n 在句中之絕對位置為 n ，相對位置為 $\text{rpos}(b_n) = \frac{n}{|B|}$ 。

3. $\text{CharSimA}(b_n)$ 為字符 b_n 在 A 中之最高相似度，通過比較 b_n 與 A 各字符而取得：若 b_n 不曾在 A 中出現，該值為 0；若出現至少一次， A 中相符者統記為 a^* ，則計算各 a^* 與 b_n 在各自句中相對位置之差的絕對值，再取其最少者，然後從 1 減去之，以得最高相似度之值。具體如下所示：

$$\text{CharSimA}(b_n) = \begin{cases} 1 - \min_a (|\text{rpos}(b_n) - \text{rpos}(a^*)|) & \text{若 } A \text{ 中存在 } a^* = b_n \\ 0 & \text{若否} \end{cases}$$

4. 同理， $\text{CharSimB}(a_m)$ 為字符 a_m 與 B 中各字符比較而得出之最高相似度。若 a_m 在 B 中不曾出現，該值為 0；若出現至少一次， B 中相符者統記為 b^* ，則計算各 b^* 與 a_m 在各自句中相對位置之差的絕對值，再取其最少者，然後從 1 減去之，以得其值。情況如下：

$$\text{CharSimB}(a_m) = \begin{cases} 1 - \min_b (|\text{rpos}(a_m) - \text{rpos}(b^*)|) & \text{若 } B \text{ 中存在 } b^* = a_m \\ 0 & \text{若否} \end{cases}$$

5. 欲計算兩個句子的相似度，分別求 A 及 B 中各字符在對方句子之最高相似度，兩者各求其總和，再分別除以 A 及 B 的長度，再取兩個數值之平均數，即為其值。

翻譯記憶庫內容主要參考香港政府編印的《香港便覽》雙語文本（詳見：<http://www.gov.hk/tc/about/abouthk/factsheets/>）。

（七）術語搜尋器：此搜尋器為專業文本而設，內含專業術語庫。用家所輸入的文本若含專業術語，本搜尋器會自動找出，加以標示，提醒譯者注意。TranXearch 暫支援財經及佛教領域的詞彙，財經詞彙選錄自恒生管理學院商務術語翻譯系統之數據庫（詳見：www.hsmcbti.com），佛教詞彙則取自 William Edward Soothill 及 Lewis Hodous 編撰的 *A Dictionary of Chinese Buddhist Terms*（詳情請參閱法鼓佛教學院圖書資訊館數位典藏組的電子文檔，網址為：<http://buddhistinformatics.ddbc.edu.tw/glossaries/files/soothill-hodous.ddbc.pdf>）。

（八）線上定義搜尋器：此搜尋器能根據「詞組翻譯器」的結果，逐詞在網上尋找其定義。再以「股票期權」為例，系統自行於網上尋找「股票期權」的中文及英文定義。現階段搜尋器以中英文維基百科為主要的搜尋對象。

（九）翻譯報告生成器：內含 HTML 編寫的翻譯報告模板，把部件三至八所輸出的結果加以整理，嵌入 HTML 編碼，製成互動式翻譯報告。報告輸出後於搜尋主頁顯示，配合主頁所包含的 JavaScript 程序，方便用家以不同方式閱覽報告，例如選擇展開或隱藏報告關於不同句子的內容（即僅顯示某一語句之報告）。

參、系統測試及演示

接下來上文關於香港衍生產品市場的文字作測試，說明系統的運作與應用。如圖 5 所示，用家首先在搜尋頁面輸入文字；注意用家不必先抽取關鍵字，只要把整段文字輸入即可。

TranXearch 翻譯搜尋引擎

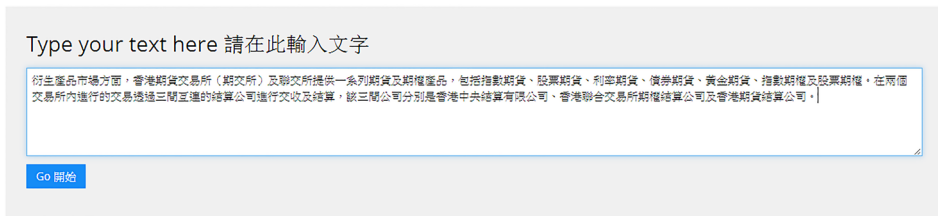


圖 5：用家輸入中文內容

按下「開始」按鈕，系統進行搜尋，製作翻譯報告，結果如圖 6 所示。翻譯報告按句子排列，即圖中之「句子一」及「句子二」：

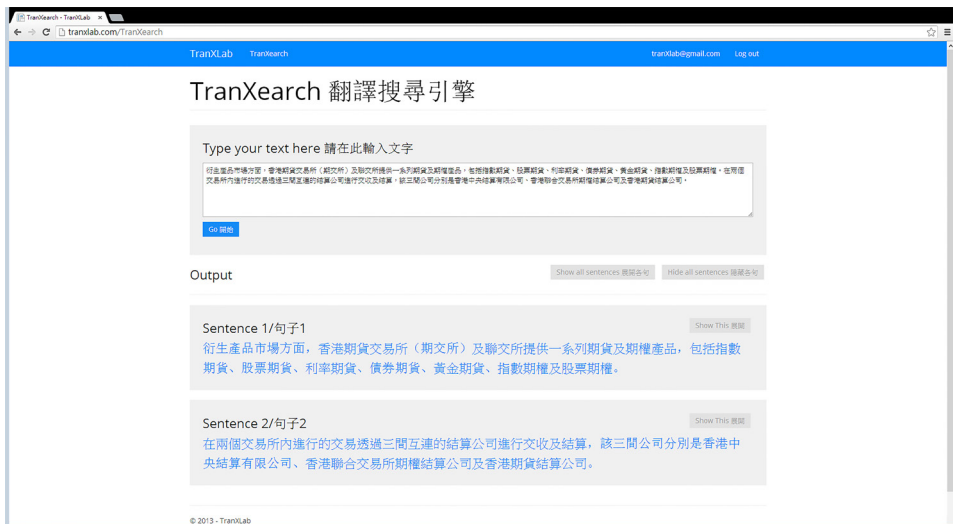


圖 6：搜尋結果

按下「展開各句」或「展開」按鈕，可閱覽個別句子的搜尋結果，也就是翻譯報告的具體內容。如圖 7 所示，每句再劃分為詞組翻譯等不同內容，用家可通過「展開 / 隱藏」按鈕展開或隱藏報告各部份：

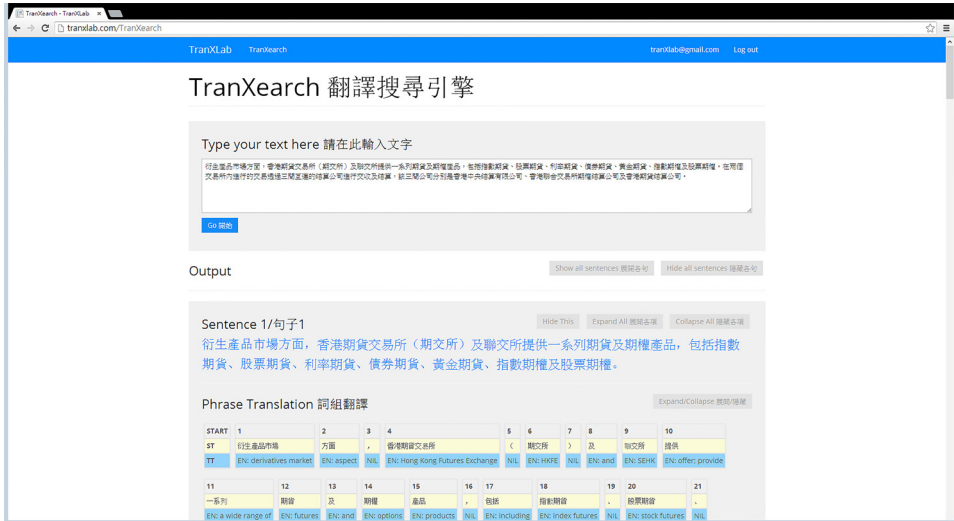


圖 7：翻譯報告（句子一）

接下來以第一句的搜尋結果為例。圖 8 為該句的詞組翻譯結果，提供逐詞中英對譯，原文若有多處疑難，不必多次查閱辭書。每個詞組均獲編號，黃色部份為用家輸入的內容，藍色部份為系統從雙語辭書找到的譯法。由於頁面空間有限，此處只提供主要或常見譯法，字詞如有詳細解釋，則見於翻譯報告的「詳細分析」一節，具體情況下文討論。

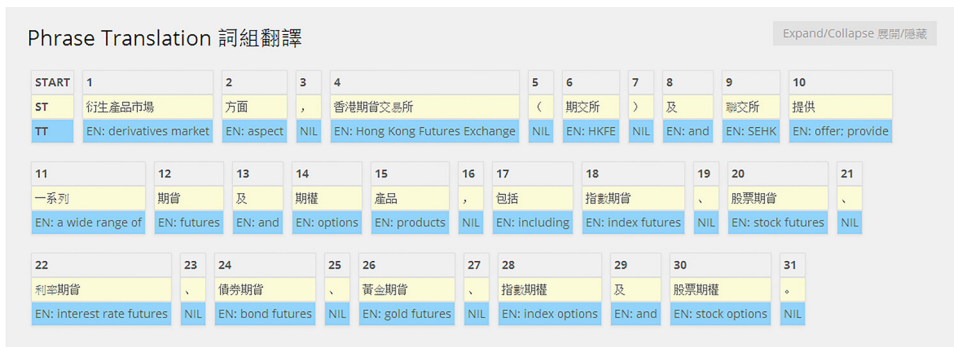


圖 8：詞組翻譯的結果（句子一）

除了詞組翻譯外，系統也提供翻譯實例，從翻譯記憶庫尋找近似譯文，供譯者參考或修改；系統還提供相似度，標明異同，方便用家取捨。圖 9 為第二句之翻譯實例，共有三條結果，相似度分別為 100%、60.51% 及 32.43%，黃色標示的字詞為所輸入語句與例句相同的地方：

Expand/Collapse 展開/隱藏

Translation Examples 翻譯實例

Source Text 在兩個交易所內進行的交易透過三間互連的結算公司進行交收及結算，該三間公司分別是香港中央結算有限公司、香港聯合交易所期權結算公司及香港期貨結算公司。

Candidate: 1	Similarity: 100%
<p>ST: 在兩個交易所內進行的交易透過三間互連的結算公司進行交收及結算，該三間公司分別是香港中央結算有限公司、香港聯合交易所期權結算公司及香港期貨結算公司。</p> <p>TT (en): The transactions on the two exchanges are cleared and settled through their three associated clearing houses, namely, the Hong Kong Securities Clearing Company (HKSCC), the Stock Exchange of Hong Kong Options Clearing House Company (SECH) and the Hong Kong Futures Exchange Clearing Corporation (HKCC).</p>	<p>ST: 在兩個交易所內進行的交易透過三間互連的結算公司進行交收及結算，該三間公司分別是香港中央結算有限公司、香港聯合交易所期權結算公司及香港期貨結算公司。</p>
Candidate: 2	Similarity: 60.51%
<p>ST: 目前，香港的證券及期貨市場包括兩間交易所及三間結算所，分別為香港聯合交易所（「聯交所」）、香港期貨交易所（「期貨交易所」）、香港中央結算有限公司（「香港結算公司」）、聯交所期權結算所有限公司（「期權結算公司」）和香港期貨交易結算有限公司（「期貨結算公司」）。</p> <p>TT (en): Currently, the securities and futures market in Hong Kong comprises two exchanges and three clearing houses, namely the Stock Exchange of Hong Kong, the Hong Kong Futures Exchange, Hong Kong Securities Clearing Company Limited, the SEHK Options Clearing House Limited and HKFE Clearing Corporation Limited.</p>	<p>ST: 目前，香港的證券及期貨市場包括兩間交易所及三間結算所，分別為香港聯合交易所（「聯交所」）、香港期貨交易所（「期貨交易所」）、香港中央結算有限公司（「香港結算公司」）、聯交所期權結算所有限公司（「期權結算公司」）和香港期貨交易結算有限公司（「期貨結算公司」）。</p>
Candidate: 3	Similarity: 32.43%
<p>ST: 你可透過配售銀行，香港中央結算有限公司或能夠並願意為你辦理申請的證券經紀申請債券。</p> <p>TT (en): You can apply through a Placing Bank, HKSCC or a securities broker who is willing and able to make application for you.</p>	<p>ST: 你可透過配售銀行，香港中央結算有限公司或能夠並願意為你辦理申請的證券經紀申請債券。</p>

圖 9：翻譯實例（句子二）

圖 10 為句子二的「實用句型」，用家翻譯時可加利用，紅色方框所示為可作替換的地方。系統同時通過比較所輸入句子及數據庫中的句型，為各個句型估算實用的程度，以百分比顯示（見圖中「Score」位置）：

Expand/Collapse 展開/隱藏

Useful Patterns 實用句型

<p>Candidate: 1 Score: 16.67%</p> <p>SL [*0]，分別是[*1:某物]及[*2:某物]。</p> <p>Pattern: [*0:Statement]，namely [*1:Something] and [*2:Something].</p> <p>TL [*0] are [*1] and [*2] respectively.</p> <p>Pattern [*0] are [*1] and [*2] respectively.</p> <p>(en):</p>	<p>SL [*0]，分別是[*1:某物]及[*2:某物]。</p>
<p>Candidate: 2 Score: 16.67%</p> <p>SL [*0]分別是[*1]及[*2]。</p> <p>Pattern: [*0] are [*1] and [*2] respectively.</p> <p>TL [*0] are [*1] and [*2] respectively.</p> <p>Pattern [*0] are [*1] and [*2] respectively.</p> <p>(en):</p>	<p>SL [*0]分別是[*1]及[*2]。</p>

圖 10：實用句型（句子二）

圖 11 顯示第二句的「重要術語」，系統偵測文本中的專門術語，加以翻譯，同時特別標示出來。

Key Terms 重要術語 Expand/Collapse 展開/隱藏

在兩個交易所內進行的交易透過三間互連的**結算公司/clearing houses**進行交收及結算，該三間公司分別是**香港中央結算有限公司/Hong Kong Securities Clearing Company**、**香港聯合交易所期權結算公司/Stock Exchange of Hong Kong Options Clearing House Company**及**香港期首結算公司/Hong Kong Futures Exchange Clearing Corporation**。

圖 11：重要術語（句子二）

TranXearch 強調搜尋一次即可取得資訊，然而文本若較複雜或較長，則可能導致輸出資料過多，用家未必能夠在短時間內全部消化。有見及此，翻譯報告特別加設「詳細分析」一項，把基本資料（即上述詞組翻譯、翻譯實例、有用句型及重要術語）跟一些詳細資料區分開來，以使用家因應實際需要閱讀搜尋結果。若用家需要對句中各詞有更深入了解，則可參閱此部份。在「詳細分析」中，句中每個詞語均有一「小檔案」，提供詳細解釋、線上的多語定義及用法實例。圖 12 為「期貨」一語的示例：

No.	12
Source Text	衍生產品市場方面，香港 期貨 交易所（期交所）及聯交所提供一系列 期貨 及 期權 產品，包括 指數期貨 、 股票期貨 、 利率期貨 、 債券期貨 、 黃金期貨 、 指數期權 及 股票期權 。
Other Languages	EN: futures
Multilingual Definitions	<p>期貨合約（英語：Futures contract），簡稱期貨，是一種跨越時間的交易方式。買賣雙方透過簽訂合約，同意按指定的時間、價格與其他交易條件，交收指定數量的現貨。通常期貨集中在期貨交易所，以標準化合約進行買賣，但亦有部分期貨合約可透過櫃台交易進行買賣，稱為場外交易合約。</p> <p>期貨是一種衍生工具，按現貨標的物之種類，期貨可分為商品期貨與金融期貨兩大類。參與期貨交易者之中，套期者（或稱對沖者）透過買賣期貨，鎖定利潤與成本，減低時間帶來的價格波動風險。投機者則透過期貨交易承擔更多風險，向標在價格波動中牟取利潤。</p> <p>不少期貨市場發展自遠期合約，指一對一個別簽訂的跨時間買賣合同，交易細節由買賣雙方自行約定。期貨合約則由交易所統一標準化，讓四方八面的交易者可在同一個平台上方便地撮合交易。期權（選擇權）是從期貨合約上再衍生出來的另一種衍生工具。</p>
SL Examples	<p>股票期貨合約是指在一個指定未來日子，按預定價格買賣預定數額相關股票的承諾。</p> <p>證監會最近就一個新股票期貨合約類別及一個新股票指數期貨合約類別訂立類似的規限。</p> <p>股票期權及股票期貨合約是根據莊家制度進行買賣的。</p> <p>為了便利市場莊家進行對沖活動，證監會現訂立修訂規則，使股票期權和股票期貨市場莊家就對沖風險而進行的賣空，毋須受到《證券條例》第 80(1)條所規限。</p> <p>賣空的股份必須是被對沖的股票期權或股票期貨持有的指定股份。</p>
TL Examples	<p>Hong Kong Exchanges and Clearing Limited ("HKEX") proposed to launch four new stock futures contracts and four new stock options contracts in respect of BOC Hong Kong (Holdings) Limited ("BOCHK"), Denway Motors Limited ("Denway"), Esprit Holdings Limited ("Esprit") and the Kowloon Motor Bus Holdings Limited ("KMB"). The contract specifications for these new stock futures and stock options contracts will be the same as those for other existing stock futures and stock options contracts.</p> <p>Introduction of new stock futures and stock options contracts is encouraged because investors will be provided with more choices and hedging activities can be facilitated.</p> <p>The Commission has to make rules to put statutory contract limits and reportable positions for these new stock futures and options contracts in place, which will mirror those contained in the HKFE Rules and SEHK Trading Procedures.</p> <p>HKEX has consulted the market makers of stock futures and options about the introduction of the new contracts.</p>

圖 12：「期貨」的詳細分析

TranXearch 各種搜尋結果的示例如上所述。不同結果均在同一頁面呈現，譯者可自由展開及隱藏報告不同地方，一方面省卻來回切換頁面之煩，另一方面確保版面內容不會過於繁多。

還有一點特別值得注意，用家所輸入的語言若同時涉及中英文，系統也能處理。香港常見中英夾雜的文章，為方便譯者處理這類文字，TranXearch 不要求用家指定所輸入文字的語言，系統可自動偵測原文中不同語言的字詞語句。若該語句某部份為中文，則提供英文資料，反之亦然。理論上，只要系統的數據庫能提供資料，用家所輸入的文字可同時包含三種甚至更多的語言。譯者若翻譯與跨語言、跨文化相關的著作，或遇到原文同時涉及多種語言，此時 TranXearch 支援多語混合分析，譯者不必按不同語言分別搜尋，節省時間。

肆、總結

TranXearch 強調翻譯與搜尋引擎之融合，研究成果冀促進線上翻譯工具的發展及整合。展望未來，以下工作值得注意：

其一，擴充數據庫。TranXearch 目前仍在初步研發階段，亟需增加數據庫詞條、翻譯實例及語言資訊，以便為譯者提供更多有用信息。除此以外也會注意發展專業領域的數據庫（如宗教、醫藥、化工及法律等），令系統不但能處理一般用語，也能支援更多專業概念及字詞語句的查詢。

其二，目前 TranXearch 的搜尋結果主要基於兩大資料來源：（一）系統內部的數據庫；（二）指定的網上辭典及百科。系統日後或可允許企業及專業譯者，自行上傳各種數據庫（如常用的翻譯記憶庫、術語庫及語料庫等）。如此一來，系統可結合用家自訂的數據庫，提供客製化搜尋體驗，令搜尋結果更切合實際個別翻譯環境所需。

其三，進行用家滿意度調查。本文寫作之時，搜尋引擎仍在內部調試階段，待測試妥當，一切就緒，即可開放公測，屆時可於線上設問卷

調查，收集用家關於操作界面及搜尋結果的意見，以便持續改良系統，令工具更切合譯者所需。

其四，系統可提供更多種類的功能，並且加強與其他翻譯科技的融合。例如用家輸入文字後，引擎可一併提供該段文字的機器翻譯結果，甚至檢查該段文字的拼寫及語法，令搜尋引擎得以更靈活運用，不但可協助用家尋找原文資訊，也可幫助確保譯文正確。

其五，加強對其他語言的支援（如日語及法語等），特別關注對方言的處理。港式中文動輒以粵語入文，部份影視節目字幕也採用粵語，對於自動分析、翻譯及搜尋粵語，自然有其需求。然縱觀目前的自動翻譯軟體及其他電子翻譯工具（如詞典），支援粵語乃至其他方言者較少，TranXearch 未來會對此多加注意，務求除了支援一般書面語，也能分析方言及口語。此舉對於將搜尋工具應用至字幕翻譯，甚至電腦輔助方言口譯，有推動作用。

關於未來的研究工作，還有一點應該強調：搜尋引擎（乃至其他翻譯工具）能否全面發揮功用，除了取決於系統設計及用家態度外，語言數據之多寡也是關鍵。儘管目前網絡發達，單語及雙語文本隨處可見，辭書及術語詞典也輕易取得，且不少均屬免費，但有關資源多受著作權保護，軟體開發人員不能自由複製採用。因此，希望日後有更多個人或單位，願意將其創作或成果置於公有領域，或允許非商業性的自由重用，此舉大大方便開發人員建立數據庫及設計系統，對於翻譯科技的整體發展及融合當有莫大裨益。

參考文獻

中文文獻

- 王正 (2011)。翻譯記憶系統的發展歷程與未來趨勢。編譯論叢，4 (1)，133-160。
- 王勇 (2005)。搜索引擎與翻譯。中國科技翻譯，2005 (1)，28-30、41。
- 史宗玲 (2004)。電腦輔助翻譯：MT & TM。臺北市：書林出版公司。
- 史宗玲 (2013)。電腦輔助商務翻譯教學——MT 及 TM 之案例研究。編譯論叢，6 (1)，113-145。
- 朱明炬、謝少華 (2003)。充分利用搜索引擎，準確地道英譯詞語。上海科技翻譯，2003 (1)，59-62。
- 吳丹 (2009)。英漢交互式跨語言檢索系統設計與實現。現代圖書情報技術，2009 (2)，89-95。
- 香港特別行政區政府 (2013 年 8 月)。香港的金融制度。香港便覽。2013 年 10 月 16 日登錄。取自 http://www.gov.hk/tc/about/abouthk/factsheets/docs/financial_services.pdf。
- 劉東飛、陳智賢、鄧俊濤、余俊 (2007)。一種輔助翻譯搜索引擎的研究與實現。計算機工程，2007 (1)，216-218。
- 劉源、譚強、沈旭昆 (1994)。信息處理用現代漢語分詞規範及自動分詞方法。北京市：清華大學出版社。

英文文獻

- Quah, C. K. (2005). *Translation and Technology*. New York: Palgrave Macmillan.
- Somers, H. (Ed.) (2003). *Computers and Translation: A Translator's Guide*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

網路資源

- 谷歌搜尋引擎 (2013)。2013 年 10 月 16 日登錄。取自 <http://www.google.com>。