

2020

創新設計 研・討・會

Conference on Innovation and Design

12/01(二)



論文集 Proceedings

主辦單位 Organizer

致理科技大學 創新設計學院

Chihlee University of Technology College of Innovation and Design

承辦單位 Implementers

創新設計學院

College of Innovation and Design

資訊管理系

Department of Information Management

多媒體設計系

Department of Multimedia Design

商務科技管理系

Department of Commerce Technology and Management

目 錄

項次	論 文 篇 名	頁次
	序 言	i
1	理財機器人系統需求評估與操作界面設計	1
2	LINE 小書僮-AI 個人化學習輔助系統	3
3	美國川普政府反全球化正政策對台灣電子產業之影響	5
4	網路購物之網路行銷研究	7
5	以 SCOR 模式為基礎之供應鏈績效管理數據儀表板	9
6	特色節慶桌上遊戲設計之研究	11
7	智慧型手機周邊產品之網路行銷研究	13
8	探討 5G 開台對雲遊戲獨立開發者影響效益之探討	15
9	虛擬實境自殺防治解謎遊戲設計之研究	17
10	台灣網紅(kol)在商業互動數位影音所扮演之多媒體影像探討	19
11	台灣品牌商在社群平台行銷策略之網紅(YouTuber)影響	21
12	LINE BOT 心情抒發平台:趣工作	23
13	機器學習技術應用於台灣員工離職預測	25
14	消費者閱讀網路評論動機與行為之探索性研究	27
15	香港社會運動對香港金融與貿易影響之探究	29
16	探討相同時間下不同運動比重對熱量消耗的影響	31
17	Digital payments, advances in ICT, and gender: Cross-country evidence	33
18	家電品牌行銷之探討-以聲寶公司為例	35
19	長照 2.0 創新服務研究	37
20	次世代遠距健身教練系統之服務設計與商業模式研究	39
21	寵物渡假旅館創新服務研究	41
22	綠色物流對臺灣網購消費者行為之影響	43
23	台灣冷凍食品進入越南市場可行性研究:以即食珍珠產品為例	45
24	運用人工智慧及文本探勘解析新書主題趨勢:以博客來網站日本文學翻譯書訊為例	47
25	擴增實境用於遠距艾灸之可行性研究	49
26	台灣設計研究所畢業論文之動畫創作論述分析	51

運用人工智慧及文本探勘解析新書主題趨勢：

以博客來網站日本文學翻譯書訊為例

1 鄧敏君, 2 彭百謙

1 致理科技大學應用日語系

2 中央研究院資訊科學研究所

minchun@mail.chihlee.edu.tw

緒論

為瞭解近年台灣翻譯書籍的出版概況，本研究結合人工智慧自然語言處理、文本探勘的詞語共現網絡等分析手法，以網路書店博客來2015-2019年間日本翻譯文學書籍為研究對象，探索日譯中出版書籍的主題趨勢與語言使用特徵。

研究方法

研究方法流程如下：自博客來網路書店抓取書名、出版年份等書籍資訊，接著運用台灣中央研究院詞庫小組開發之中文分詞斷句系統CKIP Tagger[1]，針對書名進行分詞斷句及詞性賦予的語言處理，篩選具實質意義的內容詞（名詞、動詞、形容詞）且出現頻率較高的詞語之後，依據其共現情形繪製成詞語共現網絡[2]。詞語共現網絡的原理是將出現頻率較高的關鍵詞作為節點(node)，節點大小反映其出現頻率，兩詞同時出現則代表節點的連結(links)，同時出現在相同書名的次數則構成連結的權重[3]。運用上述方法製成的日本翻譯文學書名的詞語共現網絡如圖1所示，自2015-2019年間共1,134冊書籍、2,376個相異詞中，節點為20個出現頻率大於15的詞語，詞語間連結共91條。

本研究過程使用的網路爬蟲、CKIP Tagger分詞斷句系統、詞語共現網絡繪製皆是以程式語言Python撰寫而成。本分析方法可迅速地探索大量資料，解析書名的關鍵資訊與趨勢，大幅減少傳統的資訊整合所需的時間與勞力，提供肉眼觀察所不易發現的解讀觀點。

結果與討論

分析結果顯示，博客來網路書店的日本翻譯文學類書名的詞語使用多樣性逐年增加，網絡密度越趨緊密而連結更加複雜化。整體而言，近年「夏目漱石」「芥川龍之介」「太宰治」的「小說」經常再版；「日本」「文豪」「經典」「文學」成為書名的頻率高且彼此連結性強；「陰陽師」「偵探」「殺人」「事件」「深夜」，與「經典」「文豪」的連結薄弱，其中「深夜」完全獨立於其他語

詞，自成一格，顯示詞語共現一定程度反映日譯中書籍的屬性。另外，透過比較美國翻譯文學、華文創作文學等不同分類的書名詞語共現網絡，更可凸顯日本翻譯文學的主題特色。

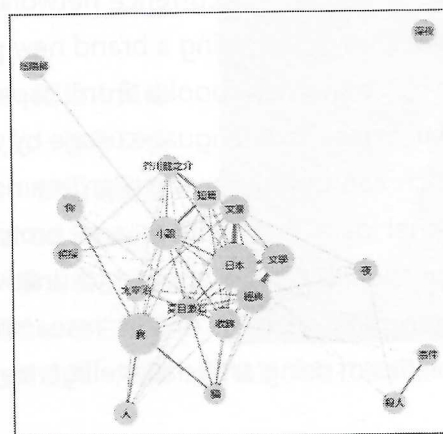


圖1：2015-2019年日本翻譯文學書名詞語共現網絡（結點為出現頻率15次以上之詞語）

結論

本研究運用人工智慧以及文本探勘的技術，嘗試設計出一套分析方法，可以快速處理大量資料，解讀新書主題趨勢。透過中文自動斷詞系統以及詞語共現網絡，解析書名詞語使用的系統性與多樣性，提供日本文學翻譯書籍的主題特徵及發展趨勢的參考依據。詞語共現網絡代表了詞語使用情形的累積結果，藉由視覺化圖示呈現詞語間的連結強度，亦可廣泛應用在挖掘其他領域的出版概況。本研究另一重要的成果是透過跨領域合作，開拓AI技術應用於人文研究的可能性。

參考資料

- [1]CKIP Tagger. Retrieved November 10, 2020. from <https://github.com/ckiplab/ckiptagger>
- [2]NetworkX: Network analysis in python. Retrieved November 10, 2020. from <https://networkx.org/documentation>
- [3]Radhakrishnan S., Erbis S., Isaacs JA, Kamarthi S.,(2017). Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. *PLoS ONE*. 12(3): e0172778.

Using AI and Text Mining to Analyze the Trends of New Book Titles: Case Study of Japanese-to-Chinese Translated Literature on Books.com.tw

¹ Minchun Teng and ² Pai-Chien Peng

¹ Department of Applied Japanese, Chihlee University of Technology

² Institute of Information Science, Academia Sinica

Corresponding author email: minchun@mail.chihlee.edu.tw

To analyze the trends and language usage characteristics of the publishing of Japanese translated literature in Taiwan in recent years, this research integrates natural language processing (CKIP, Chinese Word Segmentation System) of artificial intelligence with co-occurrence network analysis in text mining, and uses data on online bookstore Books.com.tw from 2015 to 2019. Through analyzing with CKIP Tagging and the co-occurrence network, the systematization and diversity of the words used in book titles are shown, providing a brand new perspective to the thematic features and changes of publishing trends of translated books from Japanese literature. The co-occurrence network represents the cumulative results of language usage by visualizing the strength of the links between words. This analysis approach can quickly process large amounts of data, greatly reduce the time and labor required for traditional data observation, and provide new insights that are not easy to obtain by researcher introspection. It can be applied to uncover the publication trends of other categories as well. Another important achievement of this research is that the cross-disciplinary collaboration can expand the possibilities of using artificial intelligence to humanities research.

Keywords: natural language processing, CKIP Tagger, text mining, co-occurrence network, python