

CATAR 工具在紙本書與電子書研究趨勢分析之應用

The Application of CATAR Toolkit to the Analysis of Print Books and E-books Research Trends

謝羽涵

Hsieh, Yu-Han

國立臺灣師範大學圖書資訊學研究所研究生

摘要

隨著數位出版品的普及，圖書館開始大量引進電子資源產品，使用者可以更便利地在圖書館網頁檢索相關的電子文獻、電子書等，能夠即時地滿足資訊需求。國家圖書館於 2015 年進行大專校院圖書館相關統計，數據顯示大專校院圖書館自 101 學年度起，電子書平均可使用冊數已高於紙本圖書平均冊數，大學圖書館紙本書與電子書平均館藏冊數各佔將近 50%。由此數據可知，電子書與紙本書在圖書館的館藏發展是數位化時代中必須面對的議題。Goodwin (2014) 更是認為電子書和紙本書的使用之間的關係，比我們所認知的更為複雜，未來研究應該在這個主題上繼續進行更具創造性的研究。過去至今，學者對於紙本書與電子書研究分別關注哪些研究議題？因此，本研究使用 CATAR 工具嘗試分析歷年相關領域分別探討電子書與紙本書的議題與研究趨勢，最後根據 CATAR 工具產出的結果，以國家別、年代別、最具生產力作者、最有生產力機構，以及主題詞彙等五個面向進行初步探討。

Abstract

With the popularity of digital publications, libraries begin to introduce a large number of electronic resources to users in order to instantly meet the information needs. Nowadays, users can more easily retrieve the relevant electronic documents, e-books, etc., in the library web page. However, Goodwin (2014) argues that the relationship between the usage of e-books and print books is more complex than we know, so that future research should continue to do more creative research on this subject. Therefore,

CATAR toolkit is used to analyze the topics and research trends of e-books and print books in the past years. Finally, based on the output of CATAR toolkit, I will make a preliminary discussion on the five aspects of the results, such as country, published year, most productive author, institution of author, and subject vocabulary.

關鍵字：紙本書、電子書、CATAR、趨勢分析

壹、研究背景與動機

在 2007 年，亞馬遜推出了第一臺電子書閱讀器 Kindle；在 2010 年，電子書閱讀器銷量達到巔峰，超越精裝及平裝書籍。當時許多人對紙本書的前景抱持悲觀想法。儘管現今電子書數位閱讀的議題十分火紅，但是仍有一派說法是紙本書不會輕易地消失，也不會輕易地被電子書取代。由於電子書的使用趨勢的崛起，開始有學者擔心是否會影響讀者紙本圖書使用，因而進行相關不同文本閱讀行為比較研究。Bounie, Eang, Sirbu & Waelbroeck (2013) 研究是否電子書的銷售會對紙本圖書銷售產生不利的影響。他們在 2007-2010 年間比較了紙本版本和 Kindle 格式的亞馬遜暢銷書，最後研究結果發現在電子書市場的鼓勵下，同時也會促進傳統紙本圖書的銷售。由此可見紙本書與電子書的關係並非僅有競爭關係與替代性關係，需要學者持續進行相關議題的探討。

貳、研究目的與問題

研究者好奇過去至今，學者對於紙本書與電子書研究分別關注哪些研究議題？因此，本研究欲使用 CATAR 工具嘗試分析歷年相關領域分別探討電子書與紙本書的議題與研究趨勢，本研究以國家別、年代別、最具生產力作者、最具生產力機構，以及主題詞彙等五個面向進行初步探討。根據前述欲探討的研究目的，本研究的研究問題如下：

- 一、電子書與紙本書研究文獻產量及被引用次數名列前茅的國家為何？
- 二、電子書與紙本書研究文獻主要產出年代為何？
- 三、最具有生產力的作者為何？
- 四、最具有生產力的機構為何？
- 五、電子書與紙本書研究文獻中出現的主題詞彙為何？

參、研究工具

本研究使用的 CATAR (Content Analysis Toolkit for Academic Research) 工具為國立臺灣師範大學曾元顯教授所開發的工具，專為分析具有學術價值的文獻，提供相關科學計量分析使用。CATAR 工具可適用於任何具備文字的中英文文獻，如學術論文、新聞、訪談稿等具備主題論述的文件，可針對分析文件中的作者、機構、國家、年代、被引用情況等進行擷取及統計；此外，也可根據書目對或共現字進行文件主題分析 (曾元顯, 2011)。目前有網頁說明 CATAR 的下載、使用與注意事項，網址如下 <http://web.ntnu.edu.tw/~samtseng/CATAR/Readme.html>。

肆、研究步驟

研究者首先至 Web of Science (WoS) 引用文獻索引資料庫系統分別檢索紙本書與電子書相關文獻，由於紙本書自印刷術的出現便開始普及，因此相關文獻量眾多，因此本研究檢索限縮在社會科學引用文獻索引 (Social science citation index, SSCI) 以及藝術與人文引用文獻索引 (Art & humanities citation index, A&HCI)；此外，在主題欄位使用的檢索詞彙為 print+books，以鎖定紙本書相關文獻研究；電子書的部分，則在主題欄位使用 digital+books OR ebooks OR e-books，以盡量使紙本書與電子書相關研究的文獻量達到一致。由於本研究希望了解兩種文本形式書籍的研究趨勢，因此不限定年代，最後紙本書相關文獻量共 838 筆，電子書相關文獻則有 859 筆。檢索結果以純文字檔案格式進行儲存，並使用 CATAR 工具進行分析，CATAR 工具分析結束後便輸出 Excel 檔案格式，如 Figure 2。

The screenshot shows the Web of Science search results page. At the top, it says '檢索歷史: Web of Science™ 核心合輯'. Below this is a table with columns for '檢索集' (Search Set) and '結果' (Results). There are two search sets listed:

檢索集	結果	編輯	刪除
# 2	859 主題: (digital+books) OR 主題: (ebooks) OR 主題: (e-books) 索引=SSCI, A&HCI 時間範圍=所有年份	<input type="checkbox"/>	<input type="checkbox"/>
# 1	838 主題: (print+books) 索引=SSCI, A&HCI 時間範圍=所有年份	<input type="checkbox"/>	<input type="checkbox"/>

On the right side of the interface, there are controls for '組合檢索集' (Combine Search Sets) with radio buttons for 'AND' and 'OR', and buttons for '組合' (Combine), '全選' (Select All), and '刪除' (Delete).

Figure 1 檢索策略

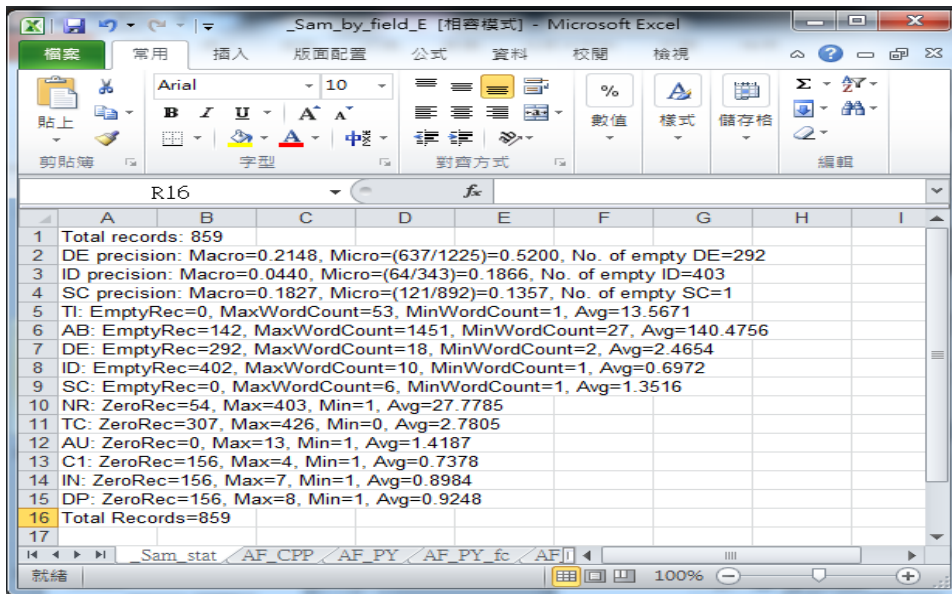


Figure 2 CATAR 工具分析結果

伍、研究結果

本研究根據 CATAR 最後產出的結果，以國家別、年代別、最具有生產力作者、最具有生產力機構，以及主題詞彙等五個面向進行初步探討。

一、以國家別來看

以國家別來看紙本書與電子書相關研究的文獻產量，紙本書的部分前五名國家依序為美國、英國、法國、西班牙及德國；電子書文獻出版量前五名的國家則是美國、英國、加拿大、澳大利亞及荷蘭。由此可推測，英國與美國對於不同閱讀文本形式的議題研究不遺餘力，其中電子書相關文獻量遙遙領先其他國家。

除了看國家文獻產量，研究者也來看文獻被引用的次數最多的前五名國家，以此來了解其文獻的重要程度，是否為其他學者所參考。被引用次數前五名的國家，紙本書依序為英國、美國、荷蘭、法國及比利時；電子書則是美國、英國、愛爾蘭、法國及荷蘭。數據中最顯目地尤其是美國，美國在電子書的文獻產量較其他國家多 2 倍以上，被引用次數較其他國家更高達 8 倍以上，可見美國在電子書的議題上領先他國，並且是其他學者所共同承認的。

		No.1	No.2	No.3	No.4	No.5
紙 本 書	文獻量	美國 (86)	英國 (65)	法國 (17)	西班牙 (10)	德國 (9)
	被引用 次數	英國 (46)	美國 (33)	荷蘭 (26)	法國 (20)	比利時 (16)
電 子 書	文獻量	美國 (122)	英國 (58)	加拿大 (15)	澳大利亞 (14)	荷蘭 (13)
	被引用 次數	美國 (831)	英國 (102)	愛爾蘭 (29)	法國 (27)	荷蘭 (27)

Figure 3 文獻量及被引用次數前五名國家

二、以年代別來看

若是來看英國與美國兩大主要文獻產量國家其文獻產出年代，數據顯示最早美國於 1975 至 1979 年間便有學者研究紙本書相關議題，之後 2000 年開始紙本書與電子書議題開始有討論趨勢，文獻量增加的速度十分迅速，電子書 (USA_E 及 UK_E) 斜率更是較紙本書 (USA_P 及 UK_P) 大，可以看出電子書議題尤其熱門。

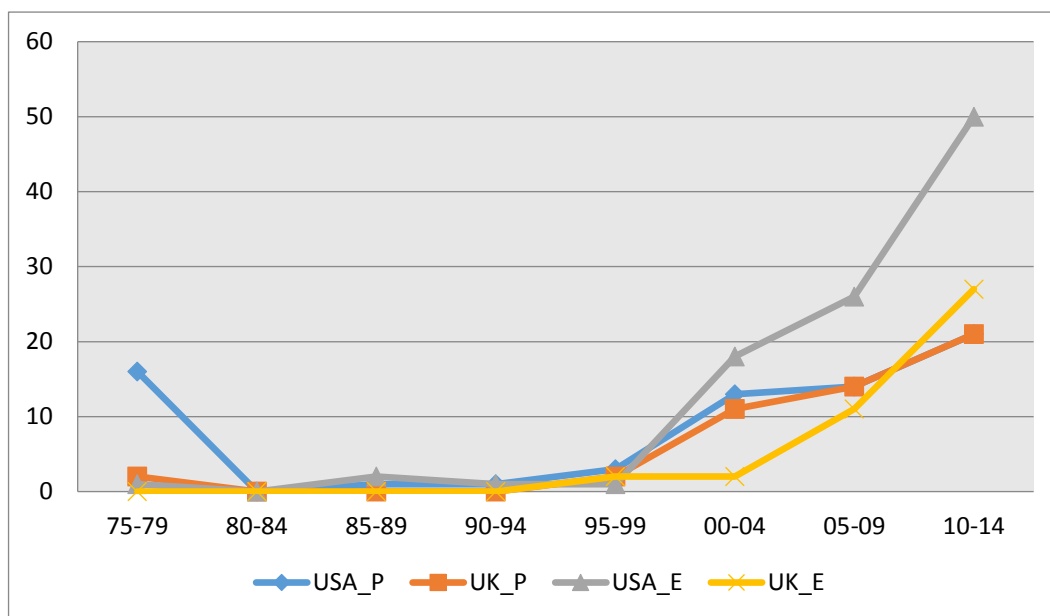


Figure 4 美國與英國其電子書與紙本書文獻產出年代

三、以最具生產力的作者來看

研究者將最具有生產力之紙本書與電子書文獻研究議題的作者前十名整理出來。其中 NC(normal count)是一般累計文獻篇數的計算方式;FC(fractional count)

則是加權累計文獻篇數算法，意即一篇文章若有 n 個作者，則每位作者僅累計 1/n 次，可以反映作者平均發表多少篇；TC (times cited) 為被引用次數；FTC (fractional times cited) 為每位作者平均被引用的次數，算法為該文獻有 n 位作者，每位作者被引用次數都累計 1/n 次；CPP (citation per paper) 為 TC 除以 NC 的值；FCPP (fractional citations per paper) 則為 FTC 除以 FC 的值，可以顯示每篇文獻的平均影響力。

AU	NC	FC	TC	FTC	CPP	FCPP
Bodenmann, R	5	5	0	0	0	0
Woudhuysen, HR	5	5	1	1	0.2	0.2
Edwards, ASG	4	3.5	13	6.5	3.25	1.86
Mckitterick, D	4	4	0	0	0	0
Flood, JL	4	4	0	0	0	0
Barbieri, E	3	3	1	1	0.33	0.33
Rumelin, C	3	3	0	0	0	0
Barker, N	3	3	0	0	0	0
Stenuit, B	3	3	1	1	0.33	0.33
Pettegree, A	2	1.5	21	14.5	10.5	9.67

Figure 5 紙本書研究文獻最有生產力作者（前 10 名）

AU	NC	FC	TC	FTC	CPP	FCPP
Greco, AN	8	4	27	13.7	3.38	3.42
Wharton, RM	5	1.5	18	6.2	3.6	4.13
Hatami, N	3	0.9	1	0.3	0.33	0.33
Hatami, H	3	0.9	1	0.3	0.33	0.33
Hatami, M	3	0.9	1	0.3	0.33	0.33
Gooding, P	2	1.3	10	6	5	4.62
Estelami, H	2	0.6	9	2.6	4.5	4.33
Price, L	2	2	15	15	7.5	7.5
Chou, IC	2	2	7	7	3.5	3.5
Unsworth, L	2	1.3	2	2	1	1.54

Figure 6 電子書研究文獻最有生產力作者（前 10 名）

四、以最具生產力的機構來看

紙本書相關議題之文獻量產出排名前十名的機構為英美國家大學及圖書館，如

下圖所示，其中英國聖安德魯斯大學（University of St. Andrews）被引用次數結果十分突出。後續若是做紙本書相關研究，可以留意此大學學者所撰寫的文獻進行觀察，繼續追蹤與瞭解此大學撰寫文獻的優勢為何，為何引用次數明顯高於其他單位。

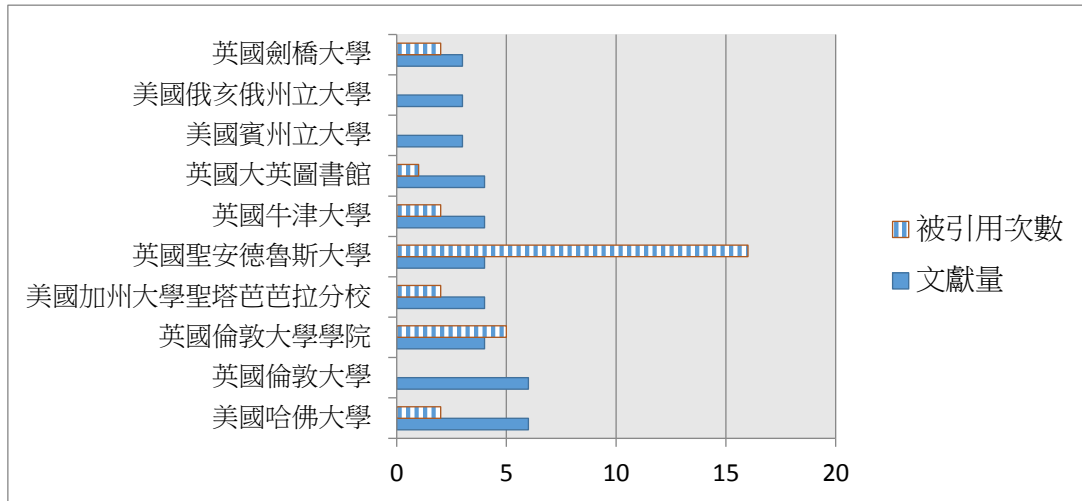


Figure 7 紙本書相關議題之文獻量產出排名前十名機構

電子書的部分依舊主要出產的機構是位於英國與美國兩個國家，比較特別的是，荷蘭一所烏特勒支大學（Universiteit Utrecht）也上榜，這所大學為荷蘭最古老大學之一，也是歐洲規模最大的大學之一。被引用次數的部分，相較於紙本書，每個機構單位產出的文獻其引用次數都不低，或許是電子書議題是現今全球討論的趨勢，因此各界皆透過相關文獻進行討論與研究。

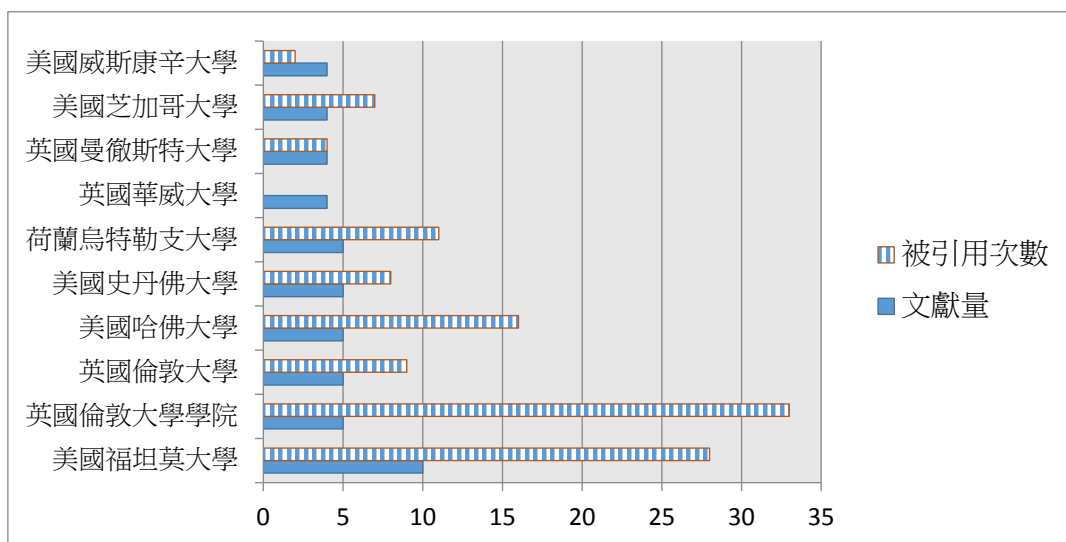


Figure 8 電子書相關議題之文獻量產出排名前十名機構

五、以主題詞彙來看

為顯現待分析文件隱含的知識結構，有三種常見的文件關聯計算方式：共被引 (co-citation)、書目對 (bibliographic coupling) 以及共現字 (co-word) (Moed, 2005)。共被引是指 2 篇文獻如果共同被後來出版的另 1 篇文獻所引用，則此 2 篇文獻之間有共被引關係，當同時被愈多其他文獻引用，表示二者之間的主題關聯性愈強；書目對是指 2 篇文獻若引用到共同的參考文獻，則 2 篇文獻產生耦合關係 (coupling)，共同參考的文獻資料量愈多，表示此 2 篇文獻在主題上愈相似、關係愈強；共現字則是指 2 篇文獻的標題或摘要裡，使用到相同的詞彙，因此當共現字出現愈多，則此 2 篇文件的主題相似度也愈高 (曾元顯、林瑜一，2011)。

本研究針對紙本書與電子書關鍵字檢索出的文獻，研究者以書目對與共現字進行三階段歸類，並將兩種歸類結果進行比較，希望可以由不同的角度來瞭解文獻的討論主題趨勢。首先紙本書相關文獻的部分，從 Figure 9 可以看出書目對與共現字分析可以看出其共通性，其中相同的詞彙有：「手稿」、「複本」、「圖文」、「圖書館」、「編目」、「歷史」等。

1	手稿、複本、現代、早期	紙本書、手稿、複本
2	歷史、檔案、文化	館藏、圖書館、地圖、編目、卷期
3	科學、植物學	歷史、傳播、跨文化
4	圖文、文化、小說、教學	大學、數位的
5	鄂圖曼帝國、十七世紀	藝術、圖文書、作品
6	紙本書、印刷	觀點、素養、傳統的
7	神話、補救、藝術家、性別	英國人、澳大利亞人、政治的
8	編目、國家圖書館	英國、威尼斯、義大利
9	聲望、荷蘭、期刊	現代、早期
10	蘇格蘭人、英國人、商人、高貴	翻譯、讀者、閱讀

Figure 9 紙本書的書目對與共現字歸類比較的結果

接下來電子書相關文獻的部分，研究者同樣進行書目對與共現字的比較，試圖將其中的關連性找出來，與紙本書操作的方式一樣，研究結果顯示從 Figure 10 可以看出書目對與共現字分析可以看出其共通性，其中相同的詞彙有：「數位化」、「大學」、「檔案」、「書」、「新穎性」、「互動」、「兒童」、「文化」等。

1	數位化、數位、線上、公司	研究、語言、新穎性、文章
2	書、虛構、發展、新穎性、語言	字典、資料庫、政策、設計
3	作品、媒體、小說、閱讀	互動、兒童、繪本、文化、問題
4	文學、文化、數位作品、文件	檔案、數位化、電子的、文學
5	歷史、全球化、數位人文	出版者、學術、大學、電子書
6	市場、大學、預訂書籍、圖書館	歷史、書、世界、當代
7	電腦、生產、電腦可讀的	技術、音樂、書、紙本、作品
8	兒童、親子、時間、課堂互動	文學、讀者、改變、小說

Figure 10 電子書的書目對與共現字歸類比較的結果

陸、結論

本研究根據 CATAR 最後產出的結果，以國家別、年代別、最具生產力作者、最具生產力機構，以及主題詞彙等五個面向進行初步探討。首先國家別的部分，英國與美國對於不同閱讀文本形式的議題研究不遺餘力，為主要紙本書與電子書議題研究文獻出產的國家，其中電子書研究文獻量遙遙領先其他國家；另外，被引用次數的部分，英美仍舊是主要國家，不過其中值得注意的是，美國在電子書研究文獻產量較其他國家多 2 倍以上，被引用次數較其他國家更高達 8 倍以上，可見美國在電子書議題研究上領先他國，並且是其他學者所共同承認的。

年代別的部分，儘管紙本書與電子書皆是在 2000 年後開始有探討的趨勢，然而紙本書相關文獻增加的速度相對趨緩，電子書文獻量產出的速度十分驚人，可以看出數位化時代的出現，電子書議題尤其熱門；以作者來看可以一窺文獻生產力高的作者為何人，並在未來研究時進行關注，但是由於紙本書文獻最高產量的作者也僅產出 5 篇文獻，電子書最高產量作者產出 8 篇文獻，可能不容易看出顯著差異；以最具生產力機構來看，與國家別來看產出的結果具有一致性，主要都是英國與美國的大學機構；但是電子書議題，荷蘭烏特勒支大學亦有不錯的文獻產出，其文獻引用次數也不低，值得未來繼續關注。

最後在主題詞彙部分，以書目對與共現字歸類並進行比較的結果，可以看出從過去到現在，紙本書與電子書議題的主題詞彙。從主題詞彙來看，紙本書為「手稿」、「複本」、「圖文」、「圖書館」、「編目」、「歷史」等，可以觀察出紙本書的研究主要仍舊著重於傳統圖書館事業與角色研究，如透過編目以館藏珍貴手稿、複本及其他具有歷史代表性之書籍；電子書主題詞彙則是「數位化」、「大學」、「檔

案」、「書」、「新穎性」、「互動」、「兒童」、「文化」等，可以呼應目前大學圖書館致力於採購電子書資源供讀者使用的趨勢，大學讀者相較於其他類型讀者，更有機會使用電子書；此外，電子書應用於兒童繪本，其多元互動性具有一定優勢，也吸引研究者進行相關研究。整體而言，以 CATAR 為工具的確可以觀察出電子書與紙本書不同文本的研究趨勢與發展現況。

參考文獻

- Bounie, D., Eang, B., Sirbu, M., &Waelbroeck, P. (2013). Superstars and outsiders in online markets: An empirical analysis of electronic books. *Electronic Commerce Research and Applications*, 12(1), 52–59.
- Goodwin, C. (2014). The e-Duke Scholarly Collection: e-Book v. print use. *Collection Building*, 33(4), 101–105.
- 國家圖書館（主編）（民 104）。中華民國一〇四年圖書館年鑑。臺北市：國家圖書館。
- 曾元顯（2011）。文獻內容探勘工具-CATAR-之發展和應用。《圖書館學與資訊科學》，37(1)，31-49。
- 曾元顯、林瑜一（2011）內容探勘技術在教育評鑑研究發展趨勢分析之應用。《教育科學研究期刊》，56（1），129-166。
- 鄭懿君（2016）。紙本與電子孰強孰弱。檢自：
<http://castnet.nctu.edu.tw/castnet/article/10048?issueID=632>