



# 數位人文研究平台之觀點變遷和 年代劃分工具發展與應用

林正和\* 張 鐘\*\* 徐志帆\*\*\*

## 【摘要】

本研究旨在開發支援數位人文研究之「觀點變遷和年代劃分工具」系統，輔助使用者於短時間內快速掌握具有時間連續性相關文本群中「出現於文本內容中的特定主題」脈絡，不必花費過多時間閱讀、篩選文本群中「未出現欲探勘主題的文本」，且能找出先前未想到的關聯、未想到應該去閱讀的文本。本工具能將具有時間連續性的相關文本，切分為數個時期，並以詞向量作為探勘輸入之關鍵字在各時期中關聯程度最高的關聯詞彙依據，使用者依據結果搜尋關聯詞彙所出現的文本，從而瞭解這些文本與欲探勘主題間的潛在關聯；並以分群方式協助使用者分析時期是否需要分裂或合併，使用者依此可修正原先對文本所下的時期切分，以對整體文本可分為的時期有更全面性的理解。透過上述兩項功能，協助使用者短時間內掌握時間連續性的相關文本中「出現於文本內容中的特定主題」的脈絡。本研究以實驗研究法比較研究對象（具人文、歷史、哲學、民族背景研究者 11 人，其中碩士以上 2 人、碩士生 8 人、大學生 1 人）使用此系統後對於「出現於文本內容中的特定主題」的理解是否具有顯著提升，輔以科技接受度問卷、半結構式訪談等了解實驗對象對此系統的感受。實驗結果發現，對於「出現於文本內容中的特定

---

\* 國立政治大學圖書資訊與檔案學研究所碩士生  
通訊作者 E-mail: handsonicver28@gmail.com

\*\* 國立政治大學華人文化主體性研究中心資訊工程師  
E-mail: changc@nccu.edu.tw

\*\*\* 國立政治大學圖書資訊與檔案學研究所碩士生  
E-mail: billxu0521@gmail.com

主題」的掌握度具有顯著的提升。在系統滿意度調查中，除訪談資料對系統輔助分析文本的便利性予以肯定外，科技接受度量表（六點量表，其中 1 為非常不同意、6 為非常同意）於認知有用性平均分數為 5.17 分、認知易用性平均分數為 4.39 分。由訪談資料及科技接受度量表得知，實驗對象對於此系統持高度肯定的正面評價。

## 關鍵詞

數位人文 文本探勘 資訊視覺化 羅家倫先生文存

## 壹、緒論

傳統上人文研究者要進行研究，必須取得大量的紙本文本，並以人工閱讀方式對其內容進行歸納與整理，往往需要耗費大量的時間、心力，使得人文研究者的研究能量受到侷限。透過全文數位資料及數位工具的輔助，可以幫助人文研究者觀察不易透過解讀傳統紙本文本所能觀察到的現象 (Schreibman, Siemens, & Unsworth, 2008)。「數位人文(digital humanities)」已逐漸成為人文領域研究的新興發展方向，也促使人文研究者的研究過程與行為產生巨大的改變 (Schreibman, Siemens, & Unsworth, 2004)。

一般而言，人文研究者於研究過程中對於數位化文本最常使用兩類數位工具：其一是大規模文本的詞頻統計分析工具，其二是關鍵字詞的檢索工具 (項潔、翁稷安, 2011)。實際上，分析詞頻即為一種找出關鍵字、主題、重要概念的研究方法。在此之上若能再將「時間」變化因素納入，即能進行的具時間變遷之主題分析，此一概念與德國歷史學家柯塞雷克 (Reinhart Koselleck) 所提出之概念史 (Begriffsgeschichte) 學派理念相同。柯塞雷克認為一個「概念」的語意場域，將會隨著其所處的時空位置而變，重要的概念可以被視為一個時刻或一個時期的特徵，觀察一個概念從「過去轉移到相對未來」的語意場域變化，便可協助我們了解歷史概念的轉變 (Motzkin, 1996)。

以柯塞雷克為首的概念史學家，專注於探究影響「政治／社會」概念的誕生、變化、挪用等現象 (陳建守, 2013)。例如，Knights (2010) 使用

關鍵字的方式分析英國於 1450 年至 1800 年有關「國協( Commonwealth)」一詞的文件，探討這一詞彙在政治、社會、文化語境上的變化。除此之外，概念史研究方法也可以用在探討其他領域學科的變化上，Astobiza (2018) 就使用知識本體( ontology)的方式探究心理學兩大經典著作：康德( Kant) 的 *Critique of Pure Reason* 和詹姆斯( James) 的 *Principles of Psychology* 中，有多少的概念隨著學科的發展而被反駁、證明與改變。

於中文研究的部分，鄭文惠等人(2016)利用詞頻分析技術，從 14,566 份唐詩中取出 2,293 份含有「白」這一關鍵字為構詞的詩歌，搭配中文古典詩歌之構詞方式，分析「白色」這一概念在情感與政治應用上的變化；詹筌亦與王乃昕(2011)利用「中國近現代思想史專業數據庫(1830-1930)」，透過詞頻分析於近代中國思潮中興起或衰落的各式「主義」，從字義之緣起、應用乃至各年間的使用頻率等。

但是目前對於中文「概念史」數位人文研究的方式，仍處於以數位人文技術分析的階段，「專以概念史方向」出發、輔助人文研究者進行研究的數位人文研究工具仍非常不足。「中國近現代思想史研究專業數據庫(1830-1930)」是一以思想史、觀念史研究為出發點建立的全文數據庫。但就其功能屬性而言，仍屬於「全文資料庫」，並未發展進階的數位分析工具。若欲進一步進行資料分析，則需要依靠使用者自行將文本匯出後執行。劉吉軒等人(2012)利用人工關鍵詞與自動關鍵詞輔以社會網路模型，呈現臺灣海外史料政治思想輪廓，但其研究係以比較自動與人工關鍵詞的精確度差異方式進行，結果顯示相較於人工關鍵詞，自動關鍵詞精準度仍有疏漏。

劉青峰(2011)指出，要觀察代表某一概念詞彙的脈絡，就必須觀察文獻中使用該詞的歷史宏觀走勢，亦即概念在時間因素上的改變。以「民主」這一概念為例，就應分析其古典意義至現代意義的演變，文本中「民主」一詞所指的對象、語境等。在資訊技術上，Word2Vec 詞向量能以詞彙在文本中的語義位置將詞彙數值化 (Mikolov, Chen, Corrado, & Dean, 2013)，以相對快速而數值化的方式輔助研究者觀察詞彙意義的位置變化。本研究則嘗試以詞向量語義上的特性為出發點，結合概念史研究中概念隨時間變化的理念，以資訊探勘技術提高人文研究者的研究能量。其理念如下所示：

1. 詞彙於詞向量中的語意位置，將由訓練詞向量的文本決定。透過更

換訓練的文本產出不同的詞向量模型，同一個詞彙語意空間位置會產生變化，同時在語意空間中與該詞彙相關的關聯詞彙亦會產生改變。

2. 文本有時間紀錄，因此可透過給予時期切分點，集合不同時期的文本作為訓練語料訓練詞向量模型，將可觀察詞彙在不同時期中於語意空間上的變化、觀察文章數值化後於向量空間上的分布情形。

對於人文學者來說，史料本身有其使用限制和門檻（王泰升，2011），史料的背景、語言、數量、時間順序等，皆是人文學者研究時的障礙。與人文學者的討論中也能得知，在面對具有一定順序的龐大文本時，研究能量常被花費在排列文本、排除不相關文本、於文本中來回尋找目標資訊中。因此，本研究發展支援概念史研究的「觀點變遷和年代劃分工具」，支援對文本具有一定、但非全盤理解的初階研究人員或研究生，輔助該類研究者閱讀具有一定關聯性、年代跨度的文本內容。並以對近代中國思潮具有重大影響的人物羅家倫作為發展此一工具之主軸，透過其經由後人整理的著作《羅家倫先生文存》作為發展「觀點變遷和年代劃分工具」之目標分析文本。希望研究者能藉由原先較為侷限的先備知識，探詢文本中以特定概念為中心的概念變遷，以快速而自動的方式輔助區分文本時期，從而對目標文本中特定概念的脈絡變化有較為深刻的理解。選擇《羅家倫先生文存》為對象文本的原因如下：

1. 《羅家倫先生文存》符合本工具欲分析的文本要素，具一定時序脈絡（每篇文章有時間記載）、一定關聯（全為羅家倫先生所寫，雖有少數為與羅家倫先生來往之函札，但其以附錄形式附於正文後，且內文仍與正文有關）。
2. 《羅家倫先生文存》所涵蓋時間範圍包含民國前、初期等 50 年，時逢中國救亡圖存、中西文化交匯衝擊，思潮觀念改變迅速而劇烈，能夠觀察概念的改變。
3. 尚未有以概念史方向出發並以數位方式分析《羅家倫先生文存》的研究。

為了驗證此一工具輔助數位人文研究的效益，本研究評估使用者利用此一工具前後對於文本的理解、科技接受度，也將透過行為記錄技術，分析使用者行為模式，最後再進行訪談，以兼具量化與質性資料分析方式，

評估人文研究者使用此一工具支援數位人文研究的成效與看法，並根據研究結果提出對於此一系統發展的未來方向。

## 貳、文獻分析

### 一、數位人文工具發展

在資訊科技未發展至近代般如此方便前，人文研究總是與閱讀整理大量書卷脫離不了關係。在豐富的數位典藏資源被建置後，人文研究者即開始利用數位典藏資源結合各種創新數位工具，嘗試解決傳統人文研究所無法觀察與分析的研究問題，此即為「數位人文」的概念 (Schreibman et al., 2008)。此後越來越多人文研究者開始使用數位工具進行數位人文研究，儼然成為現代人文學研究的發展趨勢 (Kirschenbaum, 2012)。現代的數位人文研究者已不再滿足於建立單純只具典藏功能的資料庫，而開始建置各式各樣的數位人文研究工具。其中法鼓山的「CBETA 數位研究平台」除了提供漢文佛經的線上閱讀介面之外，亦提供統計、關係圖等供佛學研究者參考 (Tu, Hung, & Lin, 2012)。中央研究院的「中華文明之時空基礎架構 (Chinese Civilization in Time and Space, CCTS)」系統，則可讓使用者可以不同透明度重疊的方式比對不同時代的地圖，以快速了解古今地名、行政區的差異 (中央研究院人文社會科學研究中心地理資訊科學研究專題中心, 2018)。吳承翰等人 (2018) 即將古籍中災害紀錄文本以 Word2Vec 詞向量數值化後，以分群技術進行整理，利用 CCTS 系統中地名時空對位應用服務進行座標定位，建立涵蓋 1647 至 1795 年中國災害紀錄的時空資訊應用系統。

「中國歷代人物傳記資料庫 (China Biographical Database, CBDB)」是由哈佛大學、臺灣中央研究院歷史語言研究所及北京大學三個機構合作建置，該資料庫收錄超過 360,000 中國歷代人物傳記和譜系資料，研究者可以透過此資料庫檢索歷史人物之籍貫、官職等資料，並提供 API 介接的方式供數位人文研究者使用數位工具取得其中的資料 (中國歷代人物傳記資料庫, 2018)。「史料人物關係圖工具」便是使用 CBDB 所提供的本名、別名、官名及諡號等人名資訊建立人名權威檔，用以判斷明代文集中的人物

資訊（陳志銘、張鐘，2018）。

於「依時間產生語意變化」的研究中，有研究使用以十年為一期的方式訓練語意向量，探勘美國於兩次世界大戰間外交文件的語彙變化（Jo & Algee-Hewitt, 2018）。Hamilton, Leskovec, 與 Jurafsky（2016）利用四種語言、橫跨兩個世紀的語料訓練詞向量，觀察同一詞彙於不同時期中於語一空問中附近詞彙的變化。本研究因而希望能透過詞向量方式將詞彙、文件量化，提供資訊技術計算，以輔助人文學者解讀文本。

數位人文的發展並非否定「人文研究者必須讀萬卷書」，而是探究是否能發展數位工具或資訊探勘技術，來幫助人文研究者能以更有效率的方式來「讀萬卷書」。因而本研究發展輔助人文研究者解讀具有一定時序脈絡、一定關聯文本的「觀點變遷和年代劃分工具」來支援數位人文研究，希望藉由發展友善、介面容易操作的數位人文研究工具，提升人文研究者進行數位人文研究的效益。

## 二、概念史

概念史(Begriffsgeschichte)學派起源於德國史學家柯塞雷克(Reinhart Koselleck)之提倡。一般而言，在政治與社會的語彙中經常充斥著許多值得探討的「概念」，而這些概念通常以關鍵字(Keywords/Schlüsselwörter)或標語(Slogans/Schlagwörter)的方式呈現(Koselleck, 2011)。因此，研究這些概念的變遷就相當於研究社會與政治的變化。柯塞雷克認為，概念(Grundbegriffe)包含了多種不同的觀念，概念的意義取決於它被使用的語境，此類研究應專注於探索「新詞」如何與「舊詞」開始產生重疊，甚至改變含意，因此概念史的研究中，「時間變化」是一重要的元素。

金觀濤(2011)認為柯塞雷克是將「概念」、「觀念」與數位人文結合的重要推手，其提出之「歷史沉澱於特定概念」的想法與文本分析中「詞彙於歷史中的語意學考察」，在概念上不謀而合。

從「概念」、「觀念」被引入歷史研究後，便離不開「尋找關鍵詞」這一研究方法，Williams(1958/彭淮棟譯, 1985)便藉由探究工業、民主、階級、藝術、文化五項關鍵詞，於18世紀下半葉至20世紀中葉中，英國作家與思想家所使用的「意義變遷」，來推論英國社會於工業革命至近現代

文化的變化。如同西方世界中劃時代的劇烈社會變化，近代中國於清末民初也經歷了一場外來思想的洗禮，知識份子為救亡圖存，引進了大量的西方文明思想與概念，除了啟迪民智外，也造成了新舊思想的抗拮與結合。

姚育松 (2011) 則使用「改造」作為關鍵詞，觀察該關鍵詞於 1895 年至 1921 年間的使用，以詞頻分析統計的方式嘗試解釋共產主義於中國的發生背景。鄭文惠、邱偉雲、劉昭麟、林書佑 (2016) 則使用與「世界」為中心之一定視窗單位 (collocation window size) 的共現關係作為相關依據，觀察 69 個關鍵詞於《新青年》雜誌中與「世界」相關的概念變化。此外，金觀濤 (2016) 等人也利用齊夫定律計算高偏離值的關鍵詞，配合關鍵詞間彼此連結的概念節點網絡，分析《新青年》雜誌中立場的變化。

本研究擬發展之「觀點變遷和年代劃分工具」係以詞向量中字與字之間的距離作為相似／相關度計算的依據，可讓使用者以自身對「概念」的關鍵詞先備知識，探勘出文本中對使用者來說未知的觀點脈絡變化。此外，也以詞向量技術將文本數值化，以分群方式輔助使用者適當地劃分出文本大致上的觀點變遷時期。

### 三、羅家倫先生與五四運動關係

羅家倫，字志希，籍貫浙江紹興，生於江西進賢。為民國教育家、政治家，被認為是最早提出五四運動一詞的人，以下簡要說明實驗題目與文本選擇的關係。

羅家倫於民國 3 年 (1914 年) 進入上海復旦公學就讀，此時羅家倫便對政治、思想領域感興趣，他曾擔任復旦雜誌編輯，發表〈二十世紀中國之新學生〉文章，也與當時的一些重要政治人物 (如戴季陶) 進行對話。

民國 6 年 (1917 年) 進入北京大學就讀，於此結識傅斯年、顧頡剛等人。民國 7 年秋 (1918 年) 開始與同道成立「新潮社」，藉由當時李大釗等師長的協助，創辦《新潮》雜誌，並於民國 8 年元旦出版第一號。民國 8 年春 (1919 年)，五四愛國運動爆發，羅家倫為其寫「五四運動宣言」，被認為是最早提出五四運動一詞的人。

思想研究多專注於羅家倫於新文化運動中的影響，運動期間誕生了不少刊物，其中著名的刊物有《新潮》、《新青年》等，而研究則分為探討理

念、文學體裁等。新文化運動期間，亦爆發了五四愛國運動，羅家倫身為該運動的重要領導者、宣言書寫者，其重要性可想而知。例如馮夏根與胡旭華（2015）分析羅家倫於文學革命中白話小說與新詩的創作、田小華與陳磊（2008）整理羅家倫於「新潮」社的活動等。

本研究將「羅家倫對五四運動觀點變化」作為實驗問題，並選擇《羅家倫先生文存》作為對象文本以建置「觀點變遷和年代劃分工具」。

## 參、系統設計

### 一、資料前處理與後設資料設計

本研究欲發展「觀點變遷和年代劃分工具」，以協助人文研究者探究具一定時間跨度的文本中，某一特定概念的時間變遷過程，因此本系統所採用之文本必須包含以下幾項特質：

1. 文本需具有一定的時間跨度，其時間跨度應至少不少於 3 年。
2. 文本之間需具有一定程度關聯，如出自同一著者手筆，或來自同一份雜誌。

本研究將以《羅家倫先生文存》作為發展此一系統之目標文本。《羅家倫先生文存》原為紙本書籍，需進行資料前處理。本研究為避免先行區分文本會造成實驗對象閱讀文本時有先入為主的限制，並未於實驗前依照主題對文本進行區分作業，僅有給予以原先於紙本中就有存在的體裁標記。本研究將取出文章中卷次、體裁形式、標題、日期、內文等資訊，為了後續系統功能使用及效能考量上，於本階段將內文進行斷詞，為內文斷詞後欄位，處理後的各篇文章各項欄位如表 1 所示。卷次、體裁形式、標題、作者、作品日期、作品來源等資訊，邀請專家以都柏林核心集架構為基礎，並依《羅家倫先生文存》之特性設計後設資料，呈現系統介面中，如表 2 所示。

機器自動判斷斷詞正確率與人工判讀難免有出入，本研究採用的 Jieba 斷詞器允許使用者加入自定義的「詞典」（結巴中文分詞，2020），在本部分係透過觀察斷詞器所得結果，逐步將可能的詞彙加入詞典中重新斷詞，以提高斷詞的準確度，詞彙必須要完全符合才算找到。

表 1

前處理後資料範例

卷次	7
體裁形式	函札
標題	復易君左函 釋「今日之世界思潮」文義
西元日期	1919-3-15
民國年	8
西元年	1919
月	3
日	15
內文	——民國八年三月十五日—— 君左先生：來書與拙作「今日之世界新潮」一文相辯難，佩甚！感甚！今以時間匆促，謹簡單致答如下： 一、思想革命是各種革命的總因。因為人民的思想變了，所以會起革命。(略)
內文斷詞後	民國 八年 三月 十五日 君 左 先生 來書 拙作 今日 世界 新潮 一文 相 辯難 佩 甚 感 甚 時間 匆促 簡單 答 如下 思想革命 革命 總因 因為 思想 變 革命 (略)

表 2

系統呈現後設資料欄位

標題 (Title)	復易君左函 釋「今日之世界思潮」文義
卷次 (Volume)	7
體裁 (Type)	函札
作者 (Creator)	羅家倫
作品日期 (Date)	1919-3-15
作品來源 (Source)	羅家倫先生文存

機器自動判斷斷詞正確率與人工判讀難免有出入，本研究採用的 Jieba 斷詞器允許使用者加入自定義的「詞典」(結巴中文分詞，2020)，在本部分係透過觀察斷詞器所得結果，逐步將可能的詞彙加入詞典中重新斷詞，以提高斷詞的準確度，詞彙必須要完全符合才算找到。

## 二、系統架構

本研究發展之「觀點變遷和年代劃分工具」系統架構如圖 1 所示，整個系統可以區分為兩個部分。其中「觀點變遷探勘」可讓使用者輸入某一關鍵詞組，由系統找出各劃分時期中文本內與該關鍵詞組相關聯的詞彙，並以視覺化方式顯示。「年代劃分工具」則會以分群方式分析各時期、兩兩分析相鄰兩時期的文本，並提供視覺化顯示與說明，輔助使用者判斷是否該將此兩時期文本合併為同一時期，或者重新切分年代，再進行「觀點變遷探勘」。以下針對圖 1 架構內的流程進行說明：

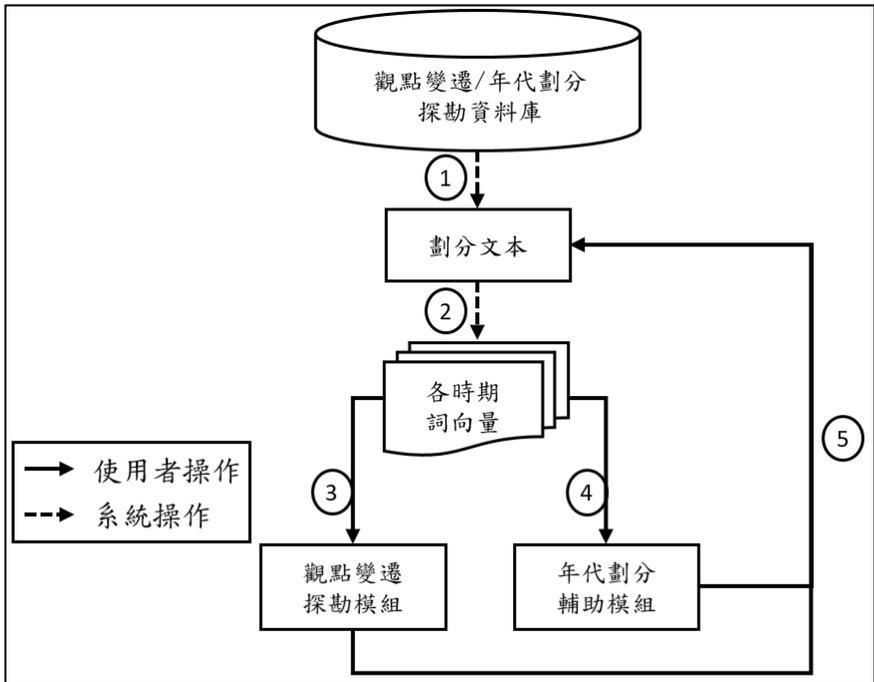


圖 1 系統架構

1. 結構化的文本已經以章節為單位，將時間資訊、體裁資訊、標題、閱讀用文本、訓練用文本、數值化結果等資訊依照結構化的方式儲存。使用者可利用本系統的各項功能，對上述資料進行閱讀分析。

使用者在使用「觀點變遷探勘」與「年代劃分工具」前，需先給定一組時間切分點，並輸入要探勘的關鍵詞組。使用者可以依照自身的認知，如對民國初中期大事記、特定的歷史事件年份甚至是羅家倫生平中所擔任的職位進行切分(後續可於參考內容後進行修正)。

2. 系統會依選定的時間切分點，將文本劃分為  $n$  個時期。在文本被劃分後，可以使用「觀點變遷探勘」與「年代劃分工具」兩項功能。
3. 「觀點變遷探勘」功能中，系統讀取使用者輸入的關鍵詞組，系統將以餘弦相似性計算各個時期中與該關鍵詞組關係最靠近的關聯詞彙。
4. 「年代劃分工具」可輔助人文研究者將每一文本視為一個資料節點，以分群方式計算相鄰兩時期的所有資料節點(文本)，是否應該被分在同一群；或其中一個時期是否有能夠分裂的可能性，以提供使用者作為決定年代劃分的參考。
5. 人文研究者參考結果後，可以重修正時間切分點，再重新切分文本，進行下一輪的分析。

### 三、系統使用元件

本研究發展「觀點變遷和年代劃分工具」所使用外部軟體元件，依序說明如下：

#### (一) Jieba中文斷詞器

本系統透過 Word2Vec 技術產生之詞向量，將文本進行數值化，為此需要先將文本進行中文斷詞之前置處理，本研究使用 Jieba 斷詞器進行文本的斷詞。Jieba 斷詞器係以開放源碼方式釋出的中文斷詞套件，同時支援繁體中文斷詞，並允許使用者加入自定義的「詞典」，來提高斷詞的準確度(結巴中文分詞, 2020)。本研究使用 Python 版本的 Jieba 斷詞套件進行中文斷詞。

#### (二) Word2Vec詞向量

Word2Vec 詞向量技術是由 Google 所開發，將詞彙向量化的技術

(Mikolov et al., 2013)。詞向量透過神經網路模型將訓練語料中的詞彙向量化，經由 Word2Vec 技術而數值化的資料可供資訊分析技術計算。詞向量的數值會受到訓練語料中詞彙語意資訊的影響，若將詞彙投影至向量空間中，意義相近的詞彙會較為靠近，反義詞則相反。部分的詞彙於詞向量中可能會呈現具邏輯的線性關係，如： $\text{vector}('King') - \text{vector}('Man') + \text{vector}('Woman') \sim \text{vector}('Queen')$ 。本研究使用 Word2Vec 詞向量將文本數值化以供分群演算法計算，並使用餘弦距離 (cosine distance) 計算功能，作為判斷詞彙間相似度的標準。

本研究中，餘弦距離的計算對象為「詞彙」；分群計算的對象為「文章」，本研究採用的計算方式為詞向量 (Word2Vec) 而非文件向量 (Doc2Vec)，無法直接以餘弦距離直接對文章進行計算，因此先利用詞向量將文章轉換為數值後，以下述密度基礎分群演算法對文章進行群集分析。

### (三) 密度基礎分群演算法 (density-based clustering algorithm)

分群又稱聚類分析 (cluster analysis)，是可以應用於數位人文資料分析的資料探勘技術之一。分群演算法主要是將輸入的資料分為不同的子集，使得每個子集中所包含的資料具有某些相近特質，再由研究者解讀命名各項子集，從而發現各項子集中所包含資料節點的共同特質。

本研究使用 Rodriguez 與 Laio (2014) 所提出密度基礎之分群演算法，可以快速輔助使用者判斷資料有幾個群集中心，該演算法基於「密度  $\rho$ 」與「距離  $\delta$ 」概念來計算群集中心，其核心思想如下：

1. 群集中心被其他局部密度較低的相鄰點包圍
2. 群集中心與其他局部密度高的節點間具有相對較高的距離

「密度  $\rho$ 」是指某資料節點於固定的範圍下，其周圍有多少資料節點。「距離  $\delta$ 」是指一資料節點距離另一「密度  $\rho$  相對較高」的資料節點有多遠，在通常的情況下會取最小值。每個資料節點取得上述兩項數值後，以密度  $\rho$  為 x 軸與距離  $\delta$  為 y 軸作圖，稱為「決策圖」。假設有範例資料分布為圖 2，其決策圖如圖 3 所示。

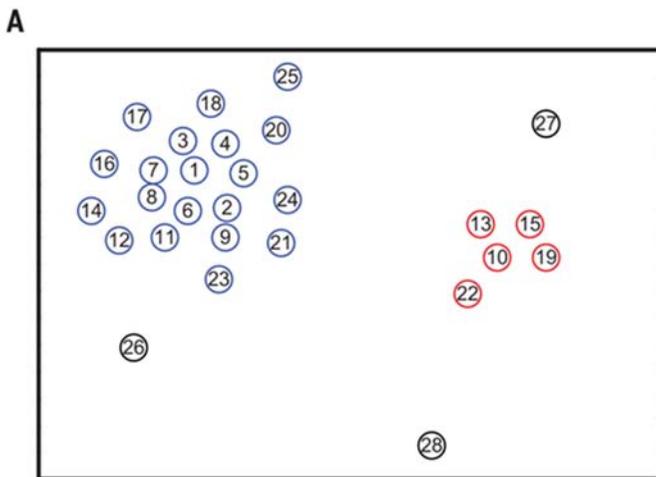


圖 2 範例資料分布

資料來源：“Clustering by fast search and find of density peaks,” by A. Rodriguez, & A. Laio, 2014, *Science*, 344(6191), p.1493.

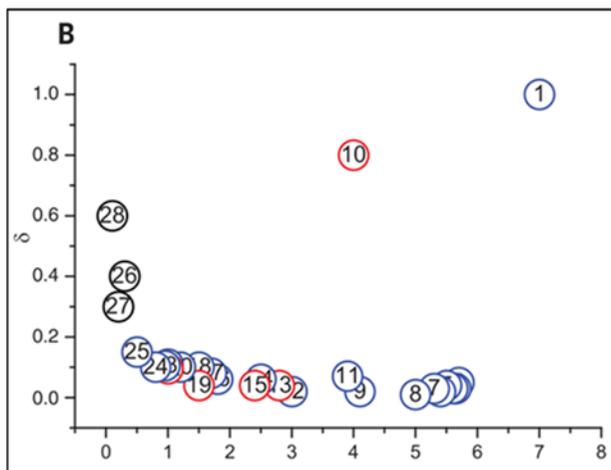


圖 3 密度基礎分群演算法決策圖

資料來源：“Clustering by fast search and find of density peaks,” by A. Rodriguez, & A. Laio, 2014, *Science*, 344(6191), p.1493.

觀察以上資料可發現，資料節點可分為兩個群集。其中「節點 1」和

「節點 10」處於「密度  $\rho$  和距離  $\delta$  兩項數值均高」的情形，此一情形也可輕易的從決策圖中辨別，對照資料分布圖後可確認「節點 1」和「節點 10」的確為群集中心。此外，假設參數  $\gamma = \text{距離 } \delta \times \text{密度 } \rho$ ，將各資料節點依照參數  $\gamma$  排序後，以參數  $\gamma$  為 x 軸、次序為 y 軸作圖，可發現「群集中心與下一次序資料節點」的傾斜程度明顯大於「非群集中心與下一次序資料節點」的傾斜程度。意即「群集中心與下一次序資料節點」斜率，極度小於「非群集中心與下一次序資料節點」之斜率。

本研究並非要將欲進行分析文本中的每一篇文章（資料節點）完整歸類在不同的子集中，而是「透過計算相鄰兩時期文章（資料節點），觀察其應該被分成幾群，從而推論出年代劃分脈絡」。此一部分理念將於「年代劃分輔助模組」中詳述。

#### 四、系統開發模組

本研究發展「觀點變遷和年代劃分工具」，用以輔助使用者短時間內掌握具有時間跨度的大量關聯文本中，特定主題隨著時間經過在文章中的變化。為此需要讓使用者能夠以時間資訊（年代）、詞彙為出發點分析文本，並以視覺化的方式呈現輔助使用者解讀文章中的資訊，或判斷是否能再進行進一步的年代分析。所開發使用之內部模組功能，依序說明如下：

##### （一）劃分文本模組

本模組為進行文本探勘的前置作業，輔助使用者依照時間資訊分析文本，使用者輸入時期切分點與欲探勘之關鍵詞，工具於本部分利用時期切分點資訊劃分文本，並訓練對應的詞向量，作為工具進行語意分析的基礎，詞向量依照訓練語料可分為「單時期詞向量」、「相鄰時期詞向量」。假設使用者將文本依照切分為四個年代，則工具首先會集合四個年代內的文本作為訓練語料訓練詞向量，分別為一至四時期之單時期詞向量；其後工具集合一、二時期內文本作為訓練語料訓練詞向量，為「一、二時期相鄰詞向量」，並依序訓練「二、三」、「三、四」時期的相鄰詞向量。「單時期詞向量」用於「觀點變遷探勘模組」中計算使用者輸入關鍵詞彙的對應關聯詞彙，以及「年代劃分輔助模組」中計算單一時期是否有能分裂的可能性；

「相鄰時期詞向量」用於「年代劃分輔助模組」中計算相鄰時期是否有合併的可能性。

由於訓練語料的不同，個別的「單時期詞向量」、「相鄰時期詞向量」中同一詞彙投影至向量空間的位置也有所不同，工具以此做為「能夠代表該詞彙於該文本集中語意位置」的數值。

## (二) 觀點變遷探勘模組

本模組主要目的在於找出各時期之間的關聯詞彙，並以視覺化、依時間序的方式呈現，從而輔助人文研究者進行文本中觀點變遷的解讀。本模組使用「劃分文本模組」中使用者所輸入的關鍵字以及產出的「單時期詞向量」，以餘弦相似性計算各時期詞向量中與關鍵字最為相關的前十名關聯詞彙，依時期序方式以視覺化的表格與圖形顯示。使用者參考結果後，可以選擇是否要重新修正關鍵詞或是重新劃分時期。

## (三) 年代劃分輔助模組

本模組將針對各時期間的文章進行分群計算，把文章用「劃分文本模組」中所得的詞向量數值化視為資料節點，計算資料節點的「分群結果」，藉此輔助人文研究者決定兩個時期文章是否應該被合在一起，或是其中的時期應該被分裂。於計算「兩個時期文章是否應該被合在一起」時，將使用「相鄰時期詞向量」將文章數值化；於計算「其中的時期應該被分裂」時，將使用「單時期詞向量」將文章數值化。本模組功能並非著重在「使用分群技術精確將文章分為數個群集」上，而是採用以下的想法：

1. 相鄰兩時期的文章群，在分群結果上若只有一個群集中心，則這兩個時期內的文章應被視為同一時期，此時應該進行這兩個時期的合併。
2. 相鄰兩時期的文章群，在分群結果上若有兩個群集中心，則顯示這項劃分結果是正確的。
3. 同一時期的文章群，在分群結果上若有多個群集中心，則說明這項劃分過於粗略，應該切為更細的時期，此時應該進行這個時期的分裂。

由於時間切分並非一個具有「絕對正解」且「適用於任何主題」的答

案，本研究並未特意為其設定一個由系統計算的正解，在年代區分輔助模組中，也僅是告知使用者能參考區分結果，並非規定使用者一定要按照系統的年代區分。

## 五、介面與功能

本節說明本研究所發展之「觀點變遷和年代劃分工具」的使用者介面和功能，讓使用者能以「劃分文本」、「觀點變遷探勘」、「年代劃分輔助」、「記事本」等功能輔助閱讀文本，進行基於文本脈絡之概念變遷探勘。各功能依序說明如下：

### （一）劃分文本

劃分文本功能為「觀點變遷和年代劃分工具」計算並解讀文本的前置作業，於初次執行時，順序將先於其他功能。使用者在執行其他功能之前必須先以本功能輸入欲進行探勘的關鍵字，並輸入時期切分點來切分文本。分項說明如下：

1. 使用者輸入欲探勘的關鍵詞，可輸入一個或多個。
2. 輸入欲探勘的起始時間點，系統將以此時間點開始探勘文本。若未輸入則預設以系統中時間點最早的文本為起始點。
3. 輸入欲探勘的結束時間點，系統將以此時間點為結束時間開始探勘文本。若未輸入則以系統中時間點最晚的文本為結束點。
4. 若使用者輸入一個或多個欲切分的時間切分點，系統會將文本切分為數個時期。若未輸入，則預設以「起始點~結束點」為一時期。

上述資料均取得後，便可開始劃分文本、計算各時期的詞向量，以進行之後的處理流程。此一部分的使用者介面如圖 4 所示。

請輸入欲查詢詞彙、時間

關鍵詞(可輸入多個，請使用空格隔開)

1 中國 五四

欲查詢起始時間

2 1918/01/14

欲查詢結束時間

3 1968/09/01

進行年代分割

切割年代

4 1930/01/01 刪除

年/月/日 刪除

新增分割年代

開始搜尋 取消

圖 4 劃分文本功能使用者介面

## (二) 觀點變遷探勘

觀點變遷探勘功能為「觀點變遷和年代劃分工具」兩大主要解讀文本之輔助功能之一，使用者在進行上述文本選擇與劃分後，可執行本項功能。

系統將使用使用者輸入關鍵詞組對「劃分文本」階段所得之 Word2Vec 詞向量，依序對各時期進行各關鍵詞的相似度計算，並取出各關鍵詞前十名最近似的關聯詞彙，並以視覺化的圖形介面顯示，稱為「觀點網路圖」，如圖 5 所示。其中中心深藍色節點為使用者輸入之關鍵詞彙節點，外側節點為工具依使用者輸入之關鍵詞彙所找出的關聯詞彙，各節點中詞彙旁括弧內數字表示該詞彙於該時期中出現在多少篇文章中。外側關聯詞彙只會與中心關鍵詞彙相連，節點間連線越靠近代表關聯度越大，關聯詞彙可不只與一關鍵詞彙節點相連，如圖 5 中，「西洋」與「五四」及「中國」兩關鍵詞彙皆有關聯。

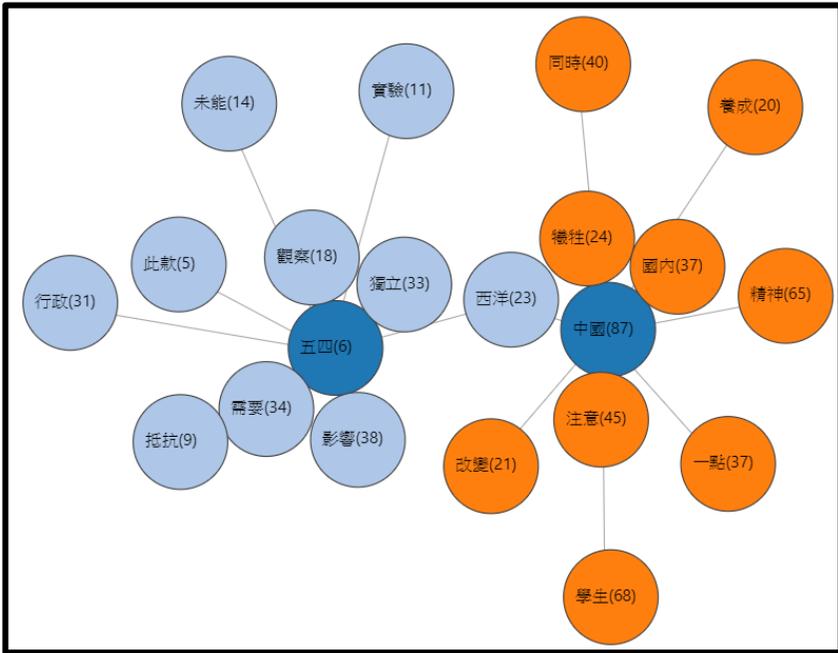


圖 5 探勘結果顯示之使用者介面

使用者可以搭配「時期分頁」功能，查看各時期的相似詞彙結果，而「時期分頁」能選擇的分頁為「劃分文本」步驟所切分的時期，如圖 6 所示。

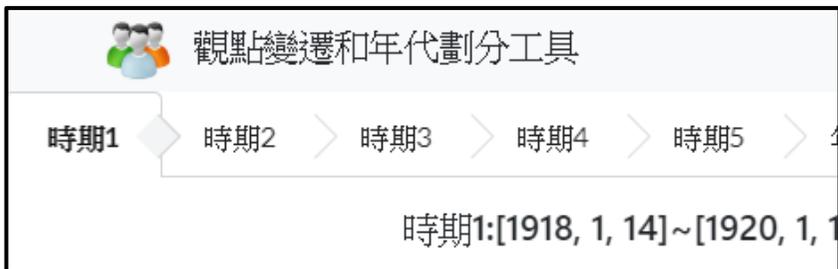


圖 6 時期分頁之使用者介面

當使用者點擊關聯詞彙，可以顯示出此關聯詞彙的相關資訊，包括該辭彙於文本中的出處，以及與各關鍵字的相似程度，如圖 7 所示。點擊其中文本的題名、體裁、內文等資訊皆可查詢相對應的文本，如圖 8 所示。

時期2:[1920, 1, 1]~[1933, 1, 1]

五四  中國

詞彙	出現文章數	與關鍵詞向量
+	五四	6
		該點為主要節點
+	影響	38
		0.9996081590652466
-	觀察	18
		0.9995607733726501

	體裁	年份	題名
	論著	1921	美國大選與新政局
出處	...美國以後，就從各方面 <b>觀察</b> 這個變遷，早思有所陳...		
	論著	1929	贊成「知難行易」說
出處	...耳。故吾人就常識上的 <b>觀察</b> ，頗覺「知難行易」說...		
	論著	1920	一年來我們學生運動底成功失敗和將來應取的方針
出處	...前的失敗，而且照我的 <b>觀察</b> ，我們現在的舉動，實...		
	論著	1921	維持北京國立高等專門以上學校建議
出處	...(二)以各省財政現狀 <b>觀察</b> ，無論何處，欲創辦此...		

圖 7 關聯詞彙詳細資料顯示之使用者介面

觀察範圍  文本資訊  記事本

搜尋條件:內文含有「甚麼叫做「五四運動」呢?民國八年五月四」

題名	卷	體裁	民國年	西元年	月	日期	內文
「五四運動」的精神	1	論著	8	1919	5	26	甚麼叫做「五四運動」呢?民國八年五月四日北京學生幾千人因山東曹錕失敗在政府高壓的壓下，屈於強迫示威，作正當民意的表示。這是中國學生的創舉，是中國教育界的創舉，也是中國國民的創舉。大家不可忘了一句話示威，在外國是常有的事，何以我們竟把也看得大驚小怪呢?不知這次運動底有三種真精神，可以驅使中國民族的存亡。第一、這次運動，是學生犧牲的精神，從前我們中國的學生，口裏法學味天，筆下天花亂墜，到了實行的時候，一個個縮頭縮腦，比起德國、朝鮮的學生來，真是慚愧死人哩!惟有這次一級責任學生，奮空拳，擁白手，和那強勢力相鬥，強的也有，

圖 8 文本查詢之使用者介面

觀察上述結果後，使用者可以再次回到輸入關鍵詞的階段，重新修正關鍵詞進行「簡易搜尋」，或回到「劃分文本」階段重新修正時期切分點，以得到新一輪的分析結果。

### （三）年代劃分輔助

「年代劃分輔助」功能係利用自動分群技術，輔助使用者判斷相鄰的兩年代中的文章「是否應該被分為同一群」。此一功能使用的分群演算法會產出「決策圖」，顯示欲分析之文本中應該有幾個群集中心，系統判斷之依據為經驗法則，該節點的密度 $\rho$ 與距離 $\delta$ 皆高於平均值，即過系統判斷之門檻值，將其判斷為群集中心，以紅點表示，再由使用者判斷是否應該將這兩個年代合併或者分裂，如圖 9、圖 10 所示。

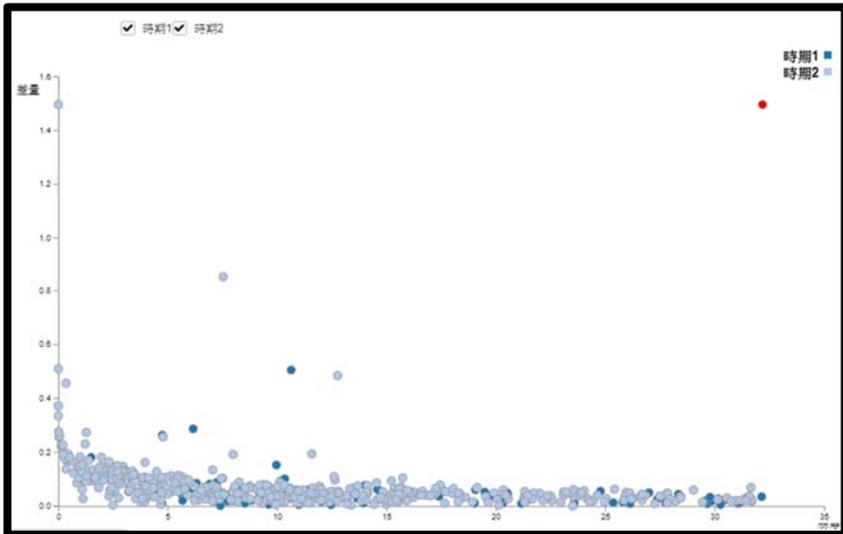


圖 9 兩時期合併判斷（決策圖）之使用者介面

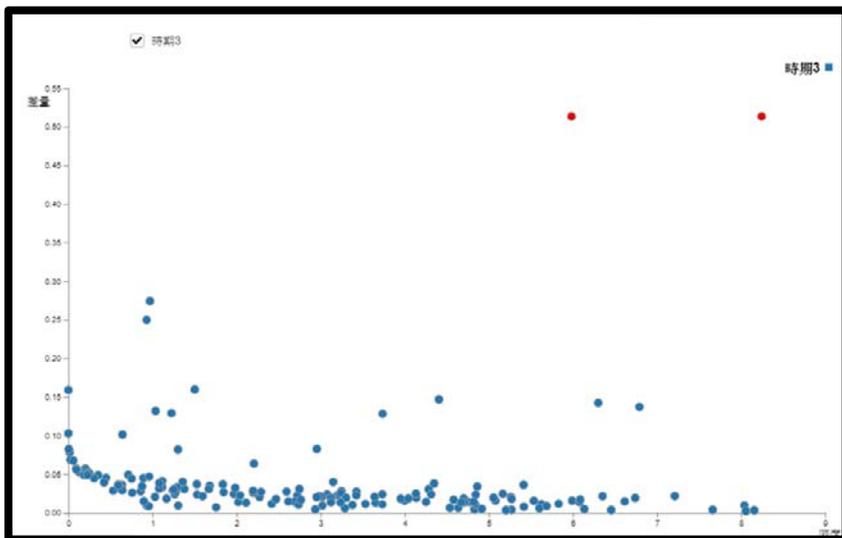


圖 10 單一時期分割判斷 (決策圖) 之使用者介面

於圖 9 中，右側「分群的結果」X 軸為密度  $\rho$ 、Y 軸為距離  $\delta$ 。可以觀察到座標圖中，有一個紅色資料節點，相較於其他資料節點處於「密度  $\rho$ 、距離  $\delta$ 」都較大的情形，決策圖便是以此對使用者建議「可將這一資料節點視為群集中心」。換句話說，就是「此兩個時期具有可以合併為一個時期的可能性」，而圖 10 為單一時期的決策圖，圖上有出現兩個紅點，代表演算法判斷出此時期內有兩個群集中心，換句話說，就是「此一時期具有可以分割為二個時期的可能性」。系統亦會顯示各時期的資訊，如起訖時間、文章數量、時期內各文章資訊供使用者參考，如圖 11 所示。

時期6
 時期7
年代合併

	時期序	時期起	時期迄	總文章數
-	時期6	[1945, 1, 1]	[1955, 1, 1]	218
<ul style="list-style-type: none"> <li>• <a href="#">論著 人類良心備忘錄</a></li> <li>• <a href="#">雜著 中華速記班二週年紀念題詞</a></li> <li>• <a href="#">函札 復張國興函 建議以秋瑾為題材拍攝電影</a></li> <li>• <a href="#">函札 復狄燦芝函 奉還「碧血英魂」稿</a></li> <li>• <a href="#">函札 致姚漁湘函 告中央日報增設專欄其編務係由編輯小組負責</a></li> <li>• <a href="#">論著 北京大學的精神</a></li> <li>• <a href="#">函札 復王雲五函 告鄭逸先因事改期趨謁學理兄</a></li> <li>• <a href="#">函札 致葉公超電 英、印、緬均將承認偽政權請速電示</a></li> <li>• <a href="#">函札 上總裁書 報告伊犁軍事進展情形</a></li> <li>• <a href="#">函札 上總裁書 呈國民大會宣言草稿</a></li> <li>• <a href="#">函札 上總裁書 檢呈「陸軍軍官學校三十年」稿本及「黃埔建軍三十年概述」</a></li> <li>• <a href="#">函札 致沈昌煥、許蘭淵函 保舉歐陽無畏在革命實踐研究院受訓</a></li> <li>• <a href="#">函札 致張羣函 謝收回山田所藏史料並請訪查他處藏件</a></li> <li>• <a href="#">序跋 「黃花崗革命烈士畫史」編輯後記</a></li> <li>• <a href="#">雜著 吉隆坡中國報四十二年元旦特刊題詞</a></li> <li>• <a href="#">函札 致王世杰函 告知起程赴印附奉上總裁書請轉呈</a></li> <li>• <a href="#">函札 上總裁書 呈總理民國十一年照片</a></li> <li>• <a href="#">函札 致尼赫魯函 對西藏貿易考察團任何商談結果概不予承認</a></li> </ul>				

圖 11 各時期資訊顯示之使用者介面

若選擇合併，將會合併「時期 1」與「時期 2」，即是刪除「時期 1」與「時期 2」之間的時期切分點後，再次進行運算。觀察上述結果後，使用者可選擇是否重新進行時間區分，或回到「劃分文本」階段重新修正時期切分點，以得到新一輪的分析結果。

#### (四) 記事本

「記事本」功能係提供使用者記錄上述功能解讀文本時所得到的新資訊之用。本系統將記事本功能整合於工具中，是為了方便使用者免於頻繁切換成其他工具來記錄思想軌跡，也方便使用者記錄輸入的資訊，如圖 12 所示。



圖 12 記事本使用者介面

### 肆、研究設計與實施

#### 一、研究架構

本研究旨在探討使用「觀點變遷和年代劃分工具」輔助解讀文本脈絡，在解讀文本脈絡上高低成效的人文學者，其操作「觀點變遷和年代劃分工具」的行為歷程上是否具有顯著的差異。將成效分為「掌握主題」、「對主題認知產生變化」的兩個面向，分別探討其高低成效人文學者作「觀點變遷和年代劃分工具」的行為歷程。

## 二、研究方法

### (一) 實驗研究法

由於本研究之關注目標特性，難以建立「未使用觀點變遷和年代劃分工具」之環境，因此本研究採用單組前後測實驗法，評估本研究發展之「觀點變遷和年代劃分工具」是否有利於輔助人文研究者進行文本解讀後之概念變遷觀察。本研究雖採用單組之前實驗研究法，但利用人文學者所得之結果將其分為高低成效兩組，比較兩組人文學者在使用「觀點變遷和年代劃分工具」行為歷程上的差異。

### (二) 訪談法

本研究於實驗結束後，將對實驗對象進行半結構式訪談，以獲得更為全面且有助於分析之質性資料。訪談的內容題目設計將注重實驗對象於實驗中所進行之研究探索歷程、結果，以及系統使用後感受。例如：「為何選擇加入某些關鍵字？」、「為何一開始選擇某一跨度之初始跨度年份？」、「使用系統後是否能簡述觀點或脈絡變遷？」、「使用系統前後是否對於主題的變化有更深程度認識？」、「使用系統各項功能時有無遭遇困難或疑問？」、「系統介面是否功能明確且流暢？」、「對於本系統是否有未來發展建議或看法？」等。

### (三) 問卷調查法

本研究欲了解實驗對象對於本研究所發展之「觀點變遷和年代劃分工具」輔以數位人文研究之看法與感受，因此於實驗結束後，邀請實驗對象填寫科技接受度量表，藉此了解實驗對象對於使用此一工具協助進行數位人文研究之文本解讀時，在認知有用性及認知易用性上的感受，並據此作為系統改善之依據。

## 三、研究工具

### (一) 觀點變遷和年代劃分工具

本研究所發展之「觀點變遷和年代劃分工具」，可以輔助人文研究者處

理、解讀具有年代時序記錄之文本，其功能又可以區分為「觀點變遷探勘」與「年代劃分輔助」兩個部份。其中「觀點變遷探勘」功能可輔助使用者擴展在文本中所出現的觀點相對應之關鍵詞或關鍵詞集，並依照年代分別顯示。「年代劃分輔助」則可以比較相鄰兩年代文本彼此之間的相似度，讓使用者決定是否該將兩年代進行合併或分裂之分析，從而讓人文研究者能以更精確方式劃分文本年代。

## (二)「摘要短文」填寫表

本研究所採用之實驗問題—「羅家倫對五四運動觀點變化」，設計為一自由、未有正確答案和定論的議題，因此本研究將集中探討人文研究者藉由「觀點變遷和年代劃分工具」輔助後，能否對於欲解讀文本之中的「觀點變遷」有更進一步、全面的理解。

為記錄實驗對象於實驗前後對《羅家倫文存》中「五四運動」這一主題概念變遷的理解程度變化，本研究設計一「摘要短文」填寫表，讓實驗對象於使用工具前，就目前的背景知識填寫「羅家倫對五四運動觀點變化」摘要短文，並於使用工具後，再以實驗流程所得之新知識修改該摘要短文，並對摘要短文進行評分，以作為成效評估依據。另外以摘要短文分數平均數將實驗對象分為高分組與低分組，觀察兩組實驗對象於實驗中的行為。

本研究邀請歷史學者依照羅家倫先生與五四運動相關生平資料為基礎，訂立摘要短文評分機制，針對研究對象於使用工具前後所撰寫、修改之摘要短文分別進行評分，評分機制標準如表 3 所示。摘要短文每符合表 3 中一項計分準則，則獲得 1 分，分數最大為 7 分，其內容必須正確符合歷史事實才予以計分。

表 3

### 摘要文章評分機制

序號	內容	設計緣由
1	提及「五四運動」發生始末	五四運動，原是 1919 年，由反對一次世界大戰後，巴黎和會決議讓日本拿走山東半島主權所引發。

(續下表)

(接上表)

序號	內容	設計緣由
2	提及「五四運動」發生後一年內反思	羅家倫注意到五四在當時社會所引起的混亂，包括盲從各種說法，追求近效而非遠慮，以及因各種因素而失焦等等。
3	提及羅家倫於「教職」時，對「五四運動」的認知	作為教育家的羅家倫完成了身份的轉換，從走上街頭抗議的學生轉變為了諄諄教誨的師長，因此針對青年的培養，也少了過去的激進，多了些實際的看法。
4	提及羅家倫於「抗日」時，對「五四運動」的認知	約莫在 1938 年左右，或許隨著中日戰爭的展開，五四相關討論中，國家及侵略出現的頻率也開始增高。
5	提及羅家倫於臺灣時，對「中共利用五四運動」的看法	羅家倫發現五四開始被有意識地塑造，譬如共產黨想要把五四說成是由自己所領導。
6	提及「新文化運動」對傳統中國文化相關影響	新文化運動為五四運動的前身，探討新文化運動，或可了解五四運動的精神。
7	提及年代資訊	本工具「觀點變遷和年代劃分工具」著重於觀點的變化，於文章中則以年代的記述顯現。

### (三)「五四運動對近代中國的影響」評估表

本研究工具主要期望可輔助使用者於短時間內能觸及更多對其專注議題有幫助的文本，於此之上協助使用者可「獲取更多原先未知的知識」或「對於議題的理解產生變化」。基於此一目的，本研究採用「關鍵詞彙變動量」作為實驗上觀察使用者「認知變化」的依據，以實驗對象第一次於工具中輸入之關鍵詞組作為其初始認知基準，並請實驗對象於流程結束後依自身對於此議題的認知變化進行整理與歸納，撰寫一關鍵詞評估表，以此評估表與初始關鍵詞組進行比較，變動越多代表使用者經由閱讀文本，對於主題的認知產生越大的變化。本研究將關鍵詞彙變動量作為成效評估依據。另外以變動量之平均數將實驗對象分為高分組與低分組，觀察兩組實驗對象於實驗中的行為。

因此，為瞭解實驗對象於實驗中是否有針對議題獲得新資訊，本研究設計「五四運動對近代中國的影響」評估表。讓實驗對象於工具使用流程結束後填寫此一表格，以歸納自身對於此總體文本中「五四運動」的認知的關鍵詞組、切分的時期與各時期對應的關聯詞彙，並探討實驗前後關鍵詞變動是否有顯著變化，其示例如表 4 所示。

表 4

關鍵詞組示例

詞組名	詞組內容	詞彙數目	萊文斯坦距離
探索前關鍵詞組	政治、世界大戰、外交	3	6
探索後關鍵詞組	新文化、創新、革命	3	

上述例子中前後關鍵詞彙的數目皆為 3，直接計算前後關鍵詞數目差為 0，但其內容構成實際上具有顯著的變動。本研究將詞視為一個單位，計算前後關鍵詞組的萊文斯坦距離 (Levenshtein distance)，記為關鍵詞變動數量。在資訊理論和電腦科學中，萊文斯坦距離是一種兩個字串序列的距離度量 (Haldar & Mukhopadhyay, 2011)。

#### (四) Google分析

Google 分析 (Google Analytics) 為 Google 公司所開發的網站流量統計服務，網站開發者只需在網站內嵌入特定程式碼，便可以使用 Google 分析記錄使用者在網站的使用歷程，諸如點擊、輸入、滑鼠移動等使用行為，並可產生視覺化的圖表供網站開發者進行分析與觀察 (Google Analytics, 2018)。本研究中採用 Google 分析技術，記錄「觀點變遷和年代劃分工具」之使用歷程，將其整理為可分析的資料，並於資料分析階段進行使用者行為模式探討。

#### (五) 科技接受度問卷

本研究參考 Hwang、Yang 與 Wang (2013) 所編製的科技接受量表，並修改該量表之語句以符合本研究之需求。本量表採用李克特六點量表，

包含兩大構面，分別為「系統認知有用性」共 6 題；「系統認知易用性」共 3 題，總計 9 題。其中「系統認知有用性」構面之 Cronbach's  $\alpha$  值為 0.95；「系統認知易用性」構面之 Cronbach's  $\alpha$  值為 0.94，皆具有良好的信度。本研究將於實驗結束後邀請使用者填寫此一問卷，用以了解實驗對象對於本研究發展之「觀點變遷和年代劃分工具」的科技接受度。

#### 四、研究對象

本研究之實驗對象需對於實驗文本具有一定認識和了解的人文研究者。考量到時間、空間、成本等因素，本研究以某國立大學文學院學生、畢業生為主，邀請對近代中國發展具有一定認識（對於五四運動、羅家倫、民國初期有一定知識而非完全無概念）或接觸過相關史料（閱讀過民國初期相關史料或羅家倫相關著作）之人文、歷史、哲學（即文史哲）以及民族研究者 11 人為實驗對象。如表 5 所示，其中包含碩士（含）以上 2 人、碩士生 8 人、大學生 1 人。

表 5

實驗對象學術程度

學術程度	人數
碩士（含）以上	2
碩士生	8
大學生	1

#### 五、實驗流程

本研究以《羅家倫先生文存》為目標文本進行處理，目標文本共有 12 冊，形制為白話文，文體包含信札、日記、文章、演講等，以本研究所發展之「觀點變遷和年代探勘工具」輔助實驗對象進行文本之解讀，進行欲觀察概念之時代變遷與理解。整體實驗流程如圖 13 所示，實驗可區分為三個階段，合計實施時間為 120 分鐘，依序說明如下：

### (一) 利用工具前對於文本觀念的理解

第一階段於實驗前向實驗對象說明整個實驗流程，使用 5 分鐘。其後，請實驗對象以先備知識就欲解讀目標文本《羅家倫先生文存》中，羅家倫對「五四運動」這一主題的觀點變化，作一摘要之撰寫，並控制實驗對象撰寫摘要的時間為 20 分鐘。

### (二) 實驗說明與利用輔助工具進行實驗

於第二階段一開始進行輔助工具的用途說明及操作介紹、說明本實驗欲達成的目標，預計使用 15 分鐘。

其後，要求實驗對象利用「觀點變遷和年代劃分工具」輔助閱讀目標文本，解讀目標文本中能代表「五四運動」這一主題觀點之關鍵詞彙變化，以及嘗試將目標文本依照實驗對象認知的要素，例如：主題、風格、觀點變化等進行年代劃分，並記錄劃分的年代與其對應的關鍵詞(群)。此一部分使用歷程記錄器、螢幕錄影記錄使用行為，時間為 40 分鐘。

此一階段完成後，請實驗對象修改其使用「觀點變遷和年代劃分工具」輔助解讀文章前所撰寫之摘要，並控制實驗對象修改摘要的時間為 20 分鐘。

### (三) 回饋與訪談

本階段為最終步驟，請實驗對象進行半結構訪談，使用時間為 20 分鐘。

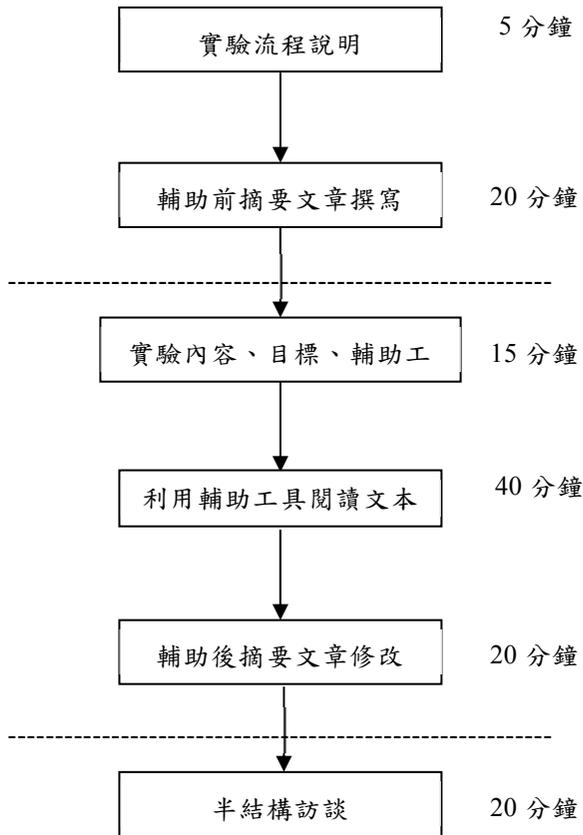


圖 13 實驗流程

## 伍、實驗結果分析

### 一、前後摘要短文分數之成效分析

因實驗對象數量為小樣本，本研究採用無母數檢定方法中的 Wilcoxon 符號等級檢定，檢定實驗對象使用工具前後所撰寫的摘要短文所得分數是否具有顯著提升 ( $\alpha=0.05$ )，敘述統計及 Wilcoxon 符號等級檢定結果如表 6 所示。結果顯示實驗對象於使用工具後，其使用工具前與後所撰寫的短文分數具有顯著的提升 ( $Z=-2.70, p=0.007<.05$ )。

表 6

使用工具前後摘要短文分數之 Wilcoxon 符號等級檢定

項目	人數	平均數	標準差	Z 值	顯著性 (雙尾)
文章分數 (使用工具前)	11	1.91	0.70	-2.70**	.007
文章分數 (使用工具後)	11	3.91	1.70		

\*\* p<.01

## 二、前後關鍵詞變動數量之成效分析

本研究以各單詞為一單位，計算前後關鍵詞組的萊文斯坦距離 (Levenshtein distance)，記為關鍵詞變動數量。以 Wilcoxon 符號等級檢定關鍵詞變動數量是否顯著的大於 0 ( $\alpha=0.05$ )，Wilcoxon 符號等級檢定結果如表 7 所示。結果顯示實驗對象於使用工具後，其關鍵詞變動數量顯著的大於 0 ( $Z=-2.81, p=0.005<.05$ )。

表 7

使用工具後關鍵詞彙變動量之 Wilcoxon 符號等級檢定

項目	人數	平均數	標準差	Z 值	顯著性 (雙尾)
起始值 (0)	11	0	0	-2.81**	.005
關鍵詞彙變動量 (使用工具後)	11	5.18	3.19		

\*\* p<.01

## 三、觀點變遷和年代劃分工具之科技接度分析

藉此了解實驗對象感受本工具有助於解讀文本中特定主題概念變遷之有用性，以及對工具操作難易度主觀感受的易用性，問卷中 1 為非常不同意、6 為非常同意，問卷附於後方附錄中，敘述統計結果如表 8 所示。結果顯示在實驗對象在認知有用性 (平均數=5.17)、認知易用性 (平均數=4.39)、整體科技接受度 (平均數=4.91) 等面向上的科技接受度均持相當高之正面態度。

表 8

實驗對象科技接受度之敘述統計

項目	人數	平均數	標準差
認知有用性	11	5.17	0.47
認知易用性	11	4.39	0.89
科技接受度整體	11	4.91	0.42

#### 四、觀點變遷和年代劃分工具之使用行為分析

為了探討實驗對象使用觀點變遷和年代劃分工具解讀文本的使用行為，本研究基於 Google Analytics 所發展之微歷程行為記錄器，記錄實驗對象於實驗過程中操作系統之行為歷程，並以「摘要短文前後差距」、「前後關鍵詞變動量」兩面向作為成效高低之依據，分別探詢「高成效」與「低成效」的實驗對象在操作工具的行為上是否具有顯著差異。

##### (一) 觀點變遷和年代劃分工具使用歷程之行為事件編碼與說明

參與實驗對象使用觀點變遷和年代劃分工具之行為為編碼，列舉如圖 14 所示，並以表 9 進行說明。



圖 14 行為事件對照觀點變遷和年代劃分工具介面圖

表 9

## 實驗對象使用觀點變遷和年代劃分工具之行為編碼及說明

編號	事件名稱	事件內容
1	概念搜尋	設定搜尋條件進行概念搜尋，或進行僅修改關鍵詞彙的簡易搜尋。本功能將會繪製出觀點網路圖。
2	外部搜尋	使用外部資源檢索，包括 Google、Wiki 等外部資源。
3	時期切換	切換檢視各時期分頁。
4	合併分析	包括切換合併檢視合併決策圖、使用時期合併按鈕等功能。
5	觀點網路圖	切換至觀點網路圖顯示的頁面。
6	回到文本資訊	切換分頁回到檢視內文搜尋結果。
7	筆記	包括檢視、記錄資訊、修改、複製文字至記事本等功能。
8	過濾	以遮罩功能篩選顯示的搜尋結果。
9	操作網路圖	操作網路圖以檢視詞彙出處文章的標題、年代、體裁、前後文等資訊。
10	內文搜尋	透過搜尋功能輸入關鍵字搜尋內文，或經由標題、年代、體裁等資訊快速搜尋內文，進行「內文搜尋」後會自動跳自內文搜尋結果頁面，使用者無需使用「檢視搜尋結果」功能。

本次參與實驗的 11 位實驗對象的行為統計資料如表 10 所示，結果顯示最常使用的系統功能前三名分別為筆記（420 次）、內文搜尋（272 次）、操作網路圖（226 次）三項。轉移圖繪製基準為「轉移顯著性大於 1.96」才會繪製行為轉移，轉移圖箭頭上的數字顯示該行為轉移的轉移顯著性，其中轉移顯著性越顯著者，其轉移行為的箭頭越粗。本節將對轉移圖中具有特定解釋性的部分轉移行為及潛在行為進行比較與解讀。

表 10

## 觀點變遷和年代劃分工具之使用歷程行為統計

事件編碼	出現頻率	出現百分比
筆記	420	28.4%
內文搜尋	272	18.4%
操作網路圖	266	18.0%
時期切換	161	10.9%

(續下表)

(接上表)

事件編碼	出現頻率	出現百分比
過濾	145	9.8%
概念搜尋	130	8.8%
回到文本資訊	118	8.0%
觀點網路圖	97	6.6%
外部搜尋	17	1.2%
合併分析	16	1.1%
事件總數	1477	100.0%

## (二)前後文章摘要差距面向上高低成效使用者之行為歷程差異分析

本研究將實驗對象操作工具前後所得摘要短文分數差距作為成效高低分組依據，觀察成效高低不同組別實驗對象於操作工具上的歷程行為是否具有顯著的不同。其中高成效組別為 4 人，其行為統計資料如表 11 所示；低成效組別共 7 人，其行為統計資料如表 12 所示。將上述兩組實驗對象之行為歷程資料分別利用滯後序列分析 (lag sequence analysis) 進行行為轉移圖繪製，高成效組別行為轉移圖如圖 15 所示、低成效組別行為轉移圖如圖 16 所示。

表 11

高成效 (文章分數) 實驗對象之行為歷程統計

事件編碼	出現頻率	出現百分比
筆記	226	36.8%
內文搜尋	130	21.2%
時期切換	80	13.0%
操作網路圖	75	12.2%
觀點網路圖	62	10.1%
回到文本資訊	58	9.4%
概念搜尋	56	9.1%
外部搜尋	4	0.7%
合併分析	3	0.5%
事件總數	614	100.0%

表 12

低成效 (文章分數) 實驗對象之行為歷程統計

事件編碼	出現頻率	出現百分比
筆記	194	22.5%
操作網路圖	191	22.1%
過濾	145	16.8%
內文搜尋	142	16.5%
時期切換	81	9.4%
概念搜尋	74	8.6%
回到文本資訊	60	7.0%
觀點網路圖	35	4.1%
合併分析	13	1.5%
外部搜尋	13	1.5%
事件總數	863	100.0%

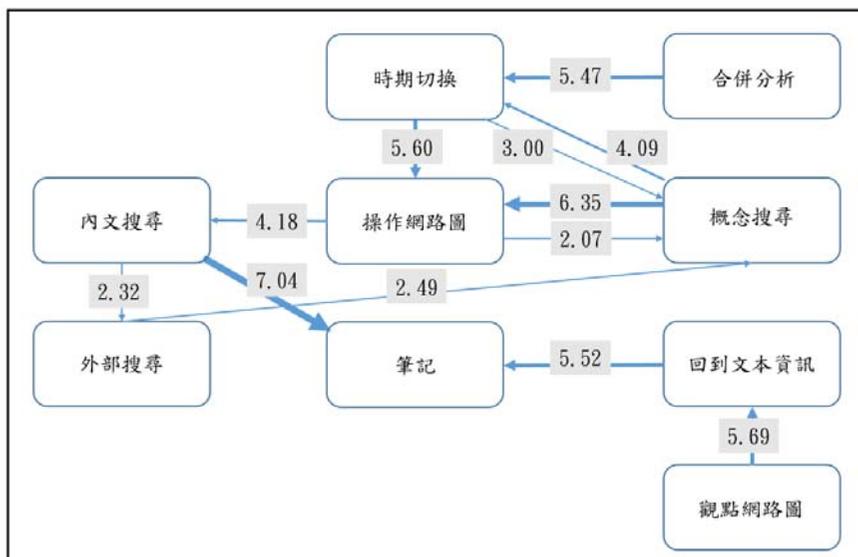


圖 15 高成效 (文章分數) 組別實驗對象之行為轉移圖

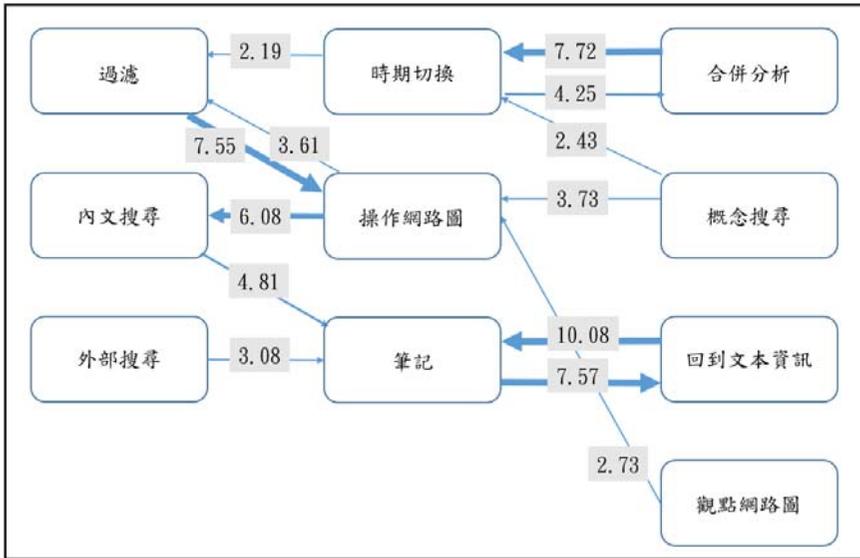


圖 16 低成效（文章分數）組別實驗對象之行為轉移圖

高成效實驗對象在「外部搜尋」到「概念搜尋」中出現顯著轉移 ( $Z = 2.49 > 1.96$ )，低成效實驗對象則在「外部搜尋」到「筆記」出現顯著轉移 ( $Z = 3.08 > 1.96$ )。推測高成效實驗對象在檢視完外部檢索資訊後，傾向於利用該資訊再度進行概念檢索後，再次繪製觀點網路圖；低成效使用者則在檢視完外部檢索資訊後，傾向於直接記錄於筆記中，而非使用「概念搜尋」繪製觀點網路圖，並以資料庫中的《羅家倫先生文存》文本驗證該資訊是否符合探索之主題。

高成效實驗對象的行為中，「概念搜尋」到「操作網路圖」出現顯著轉移 ( $Z = 6.35 > 1.96$ )、「操作網路圖」到「內文搜尋」出現顯著轉移 ( $Z = 6.08 > 1.96$ )、「內文搜尋」到「外部搜尋」出現顯著轉移 ( $Z = 2.33 > 1.96$ )，顯示高成效使用者頻繁使用概念搜尋功能繪製觀點網路圖搜尋內文，並於內文得到新資訊後，會利用外部檢索該資訊，以獲取更全面的整體資訊。結合上述「外部搜尋」到「概念搜尋」出現的顯著轉移，可以歸納出高成效實驗對象具有「使用概念搜尋功能繪製觀點網路圖，從圖中得到詞彙的相關資訊（如題名、出處等），並以該資訊搜尋內文之顯著行為轉移。在從

內文處得到新知識後，會利用外部檢索得到更全面的知識，再將得到的新知識輸入工具中進行概念搜尋，繪製新的觀點網路圖」的潛在行為。

### (三) 前後關鍵詞差距面向上高低成效使用者之行為歷程差異分析

本研究將實驗對象操作工具前後所得關鍵詞變動量作為高低成效分組依據，觀察高低成效組別實驗對象於操作工具上的行為歷程是否具有顯著的不同。其中高成效組別為 6 人，其行為統計資料如表 13 所示；低成效組別為 5 人，其行為統計資料如表 14 所示，高、低成效組別行為轉移圖分別如圖 17 及圖 18 所示。

表 13

高成效 (關鍵詞變動量) 實驗對象歷程行為統計

事件編碼	出現頻率	出現百分比
筆記	244	28.1%
內文搜尋	194	22.3%
操作網路圖	145	16.7%
過濾	95	10.9%
回到文本資訊	77	8.9%
觀點網路圖	73	8.4%
時期切換	70	8.1%
概念搜尋	51	5.9%
外部搜尋	7	0.8%
合併分析	4	0.5%
事件總數	869	100.0%

表 14

低成效 (關鍵詞變動量) 實驗對象歷程行為統計

事件編碼	出現頻率	出現百分比
筆記	176	28.9%
操作網路圖	121	19.9%

(續下表)

(接上表)

事件編碼	出現頻率	出現百分比
時期切換	91	15.0%
概念搜尋	79	13.0%
內文搜尋	78	12.8%
過濾	50	8.2%
操作網路圖	41	6.7%
概念網路圖	24	3.9%
合併分析	12	2.0%
外部搜尋	10	1.6%
事件總數	608	100.0%

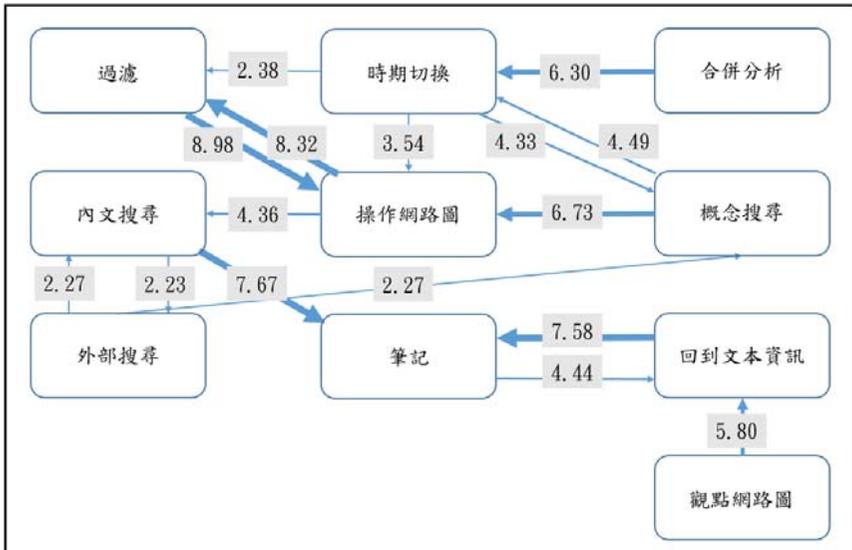


圖 17 高成效（關鍵詞變動量）組別行為轉移圖

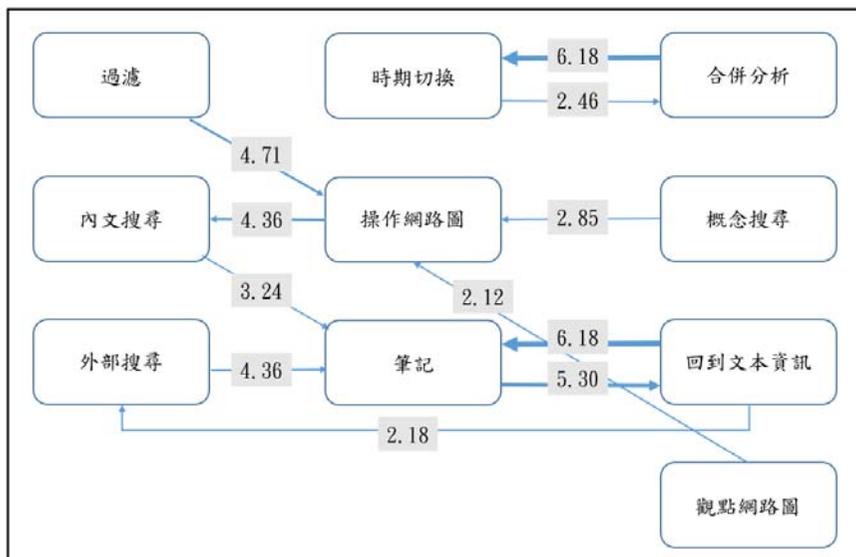


圖 18 低成效 (關鍵詞變動量) 組別行為轉移圖

高關鍵詞變動成效實驗對象在「內文搜尋」到「外部搜尋」出現顯著轉移 ( $Z = 2.229 > 1.96$ )、低關鍵詞變動成效實驗對象則在「回到文本資訊」到「外部搜尋」出現顯著轉移 ( $Z = 2.18 > 1.96$ )。雖然「內文搜尋」和「回到文本資訊」的情形皆顯示兩組使用者具有「從文本中得到資訊，再經由外部搜尋取得資訊」的潛在行為，但低成效組別的「回到文本資訊」則顯示實驗對象曾執行過其他行為動作以切換至「並非顯示查詢內文結果」的頁面，於進行「外部搜尋」前才再度切回內文搜尋結果頁面，與高成效組別相比動作較為零碎，並且沒有從「特定行為」轉移至「回到文本資訊」的顯著轉移。

高關鍵詞變動成效實驗對象在「外部搜尋」到「內文搜尋」出現顯著行為轉移 ( $Z = 2.27 > 1.96$ )、在「外部搜尋」到「概念搜尋」出現顯著行為轉移 ( $Z = 2.27 > 1.96$ )；低關鍵詞變動成效實驗對象則只在「外部搜尋」到「筆記」出現顯著行為轉移 ( $Z = 4.355 > 1.96$ )。推測高關鍵詞變動成效實驗對象在取得「外部搜尋」資訊後，傾向於利用工具的「內文搜尋」、「概念搜尋」等功能與系統內文本進行比對，以驗證是否符合主題，而低關鍵

詞變動成效實驗對象則傾向於直接記錄該資訊。

## 五、質性訪談資料分析

本研究除了以分析實驗對象操作觀點變遷和年代劃分工具所得之成效、科技接受度與使用者行為歷程外，亦希望了解實驗對象使用觀點變遷和年代劃分工具解讀文本時的感受。因此，在實驗結束後對 11 名實驗對象進行半結構式訪談，以得到使用者操作系統進行觀點變遷探索時的感受，以及希望改善系統的建議。

### (一)「觀點網路圖」功能輔助使用者解讀文本之效益

實驗對象認為「觀點網路圖」功能能夠輔助使用者解讀文本，幫助使用者以多元相關的觀點方式，檢索到許多與主題相關但並未想到的關聯詞彙，再專注於該關聯詞彙進一步尋找文本中出現該詞彙的位置。但同時也認為「觀點網路圖」中與探索主題相關之關聯詞彙的判斷依據，需要作進一步修正，例如只取實質意義較大的「名詞」、去除虛詞等；亦有使用者建議除了使用語意關聯作為相關度分析之外，亦可以嘗試其他的關聯方式取出關聯詞彙，例如主題分析、詞頻統計、共現關係等。

### (二)「年代合併」功能輔助使用者解讀文本

本研究發現實驗對象使用「年代合併」功能的次數偏少，然而觀察對象於實驗中的行為，其使用年代合併時多為「觀察」年代合併成果，並非頻繁進行操作，而是將其視為「驗證」功能以確認所下的劃分是否正確。且大多數實驗對象在訪談中認為在時間允許、量化研究、研究時代加長等情況之下，此一功能有助於觀點變遷之年代劃分。

### (三)「文本搜尋」功能有助於輔助使用者解讀文本

實驗對象皆認為「文本搜尋」功能有助於輔助使用者解讀文本，特別是可讓使用者以快速、方便的方式搜尋文本，並以結構化的方式顯示文本的標題、體裁、內文資訊，但也認為本功能在文本的可讀性上應該改善，例如固定文章位置等，也可以加上額外的資訊匯出功能。

#### (四)「外部搜尋」功能有助於輔助使用者解讀文本

大部分的實驗對象都認同「外部搜尋」功能可以快速地幫助背景知識較不完備的使用者，建立初步解讀主題的背景知識、或修正一些錯誤資訊。

#### (五)「觀點變遷和年代劃分工具」作為學習工具的適切性

實驗對象對本題的回答有分歧，有部分實驗對象認為「觀點變遷和年代劃分工具」適合作為研究工具、有部分則認為適合作為學習工具。其中，認為較適合作為「研究工具」者認為此一工具的使用門檻過高、主題不具吸引力可能導致學習者卻步，需要以專家對文本分門別類或提供學習者一個入門的前導介紹、課程或文章等。

### 陸、結論與未來研究方向

本研究分析實驗對象在使用觀點變遷和年代劃分工具前後，在「羅家倫對五四運動的觀點變化」主題上的掌握程度是否具有顯著的提升，並分析其對「五四運動的主題認知是否具有顯著變化。結果顯示使用者在使用工具後，對《羅家倫先生文存》中「羅家倫對五四運動的觀點變化」主題的掌握度具有顯著的提升，並且對「五四運動」這一主題的認知產生顯著的變化。此外，實驗對象對於使用觀點變遷和年代劃分工具輔以掌握特定主題的觀點變遷科技接受度持正面態度，認為本工具能夠輔助使用者於短時間內迅速掌握大量具關連性、時間連續性文本中的特定主題變化。

本研究藉由使用滯後序列分析實驗對象使用觀點變遷和年代劃分工具的行為歷程，將使用者依照「掌握主題」(摘要短文分數)和「對主題認知產生變化」(前後關鍵詞差距)兩個成效面向進行高低分組，觀察其行為歷程的不同。在兩個組別中，高成效實驗對象皆有「查看概念搜尋所繪製的網路圖以搜尋文本，將文本中得到資訊以外部檢索查詢，再將外部檢索得到的資料輸入系統進行概念搜尋，進而再次繪製觀點網路圖」的潛在行為序列；低成效實驗對象則傾向直接記錄透過外部檢索得到的資訊。很明顯的，高成效實驗對象循環的行為序列，正是觀點變遷和年代劃分工具的核心功能。

在訪談中，有許多實驗對象表示觀點網路圖可以方便、快速的讓使用者找到一開始未考量的關聯詞彙，幫助其以滾雪球的方式瀏覽具有關聯性的文章，使得使用者可以在短時間內找到更多可用於掌握主題的文章。然而實驗對象也建議修正觀點網路圖所萃取詞彙的精確性，以避免使用者對於其結果失去信心、或花更多時間排除機器判斷出的錯誤資訊。此外，許多實驗對象建議只需要顯示具有實質意義的「名詞」，讓使用者直覺地從圖中判斷出更多資訊。再則，研究發現實驗對象使用年代合併分析工具的次數偏少，訪談中除了有實驗對象認為此一工具不符合其使用習慣、對於單純量化圖表不信任外，也有實驗對象指出本研究所涵蓋的文本其涵蓋時間範圍過短，使得使用者可以直接單純經由解讀文本來進行年代劃分、或以歷史事件為界來劃分後再以文本內容佐證。但也認同此一功能在某些情況下能有其用處，例如文本跨度為兩三百年的時期，或者可用以佐證已有的人文時代劃分等。

本研究雖已一定程度證明《羅家倫先生文存》所發展之觀點變遷和年代劃分工具，可基於詞向量及自動分群技術協助人文學者利用分割時期、探詢各時期中與關鍵字相關的關聯詞彙，分析具有一定時間序及一定關聯性的文本，有效支援數位人文研究，但仍有幾個未來研究方向值得繼續探究。首先，本實驗所使用《羅家倫先生文存》，是一時間跨度約 50 年的白話文文本，未來可以考慮更換文本，以驗證本工具是否適合應用於其他不同體裁（如文言文）、不同時間長度（如 200 年）的文本。其次，基於詞向量可以從語意層次上探詢關聯詞彙，然而對人文學者來說這樣的關係有些抽象，有可能造成解讀上的困惑。未來可以嘗試使用其他計算關聯詞彙的技術，例如主題相關性或傳統的共現字關係，比較不同關聯詞彙計算方法在使用本工具顯示其詞彙關聯上的成效。再者，本研究的實驗對象為 11 名人文學者，整體樣本偏少，未來有機會進行後續研究應增加實驗對象數目，以大量受訪者佐證工具之有效性。最後，本研究的訪談中，有許多實驗對象建議可以將本工具發展為輔助數位人文教學的輔助工具。例如，未來可以依特定的教學主題，使用本工具輔助學生閱讀、探勘文本，以探究本工具在支援數位人文教育上的成效。

## 誌謝

此論文撰寫臺灣教育部高教深耕計畫特色領域中心之「華人文化主體性研究中心」經費補助（經費代碼 109H21）。

（接受日期：2020 年 4 月 7 日）

## 參考文獻

- Sun, Junyi(2020)。GitHub - fxsjy/jieba: 結巴中文分詞。檢自：  
<https://github.com/fxsjy/jieba>
- 中央研究院人文社會科學研究中心地理資訊科學研究專題中心。中華文明之時空基礎架構(簡易版) CCTS Lite。檢自：<http://gissrv4.sinica.edu.tw/gis/cctslite.aspx>
- 中國歷代人物傳記資料庫。CBDB querying and reporting system -online。檢自：  
<http://db1.ihp.sinica.edu.tw/cbdb/help/systemintro.html>
- 王泰升（2011）。數位化歷史資料庫與歷史研究——以明清檔案、淡新檔案、日治法院檔案等資料庫為例。在項潔主編，*從保存到創造：開啟數位人文研究*（頁 31-50）。臺北市：國立臺灣大學出版中心。  
doi:10.6327/NTUPRS-9789860300208
- 田小華、陳磊（2008）。羅家倫與新潮社。*沙洋師範高等專科學校學報*，3，37-39。
- 吳承翰、吳尚芸、白璧玲、蔡融易、黃詩芸、蔡宗翰、范毅軍（2018）。基於史籍所載氣象紀錄之事件分類與時空資訊整合研究。在第九屆數位典藏與數位人文國際研討會。臺灣數位人文學會、法鼓文理學院主辦，法鼓文理學院，新竹市。
- 金觀濤（2011）。數位人文研究的理論基礎。在項潔編，*數位人文研究的新視野：基礎與想像*（頁 45-61）。臺北市：國立臺灣大學出版中心。  
doi:10.6327/NTUPRS-9789860301656
- 金觀濤、邱偉雲、梁穎誼、陳柏聿、沈錕坤、劉青峰（2016）。觀念群變化的數位人文研究——以《新青年》為例。在項潔編，*數位人文：在過去、現在和未來之間*（頁 427-463）。臺北市：國立臺灣大學出版中心。
- 威廉斯（Williams, R.）（1958）。*文化與社會：1780 至 1950 年英國文化觀念之發展*（*Culture and society, 1780-1950*）（彭淮棟譯）。臺北市：聯經。

(原作 1958 年出版)

- 姚育松 (2011)。從「改造」一詞的使用看共產主義在中國的發生背景。在項潔編，*數位人文在歷史學研究的應用* (頁 189-217)。臺北市：國立臺灣大學出版中心。doi:10.6327/NTUPRS-9789860301632
- 陳志銘、張鐘 (2018)。古籍數位人文研究平台之史料人物關係圖工具發展與應用。在*第九屆數位典藏與數位人文國際研討會*。臺灣數位人文學會、法鼓文理學院主辦，法鼓文理學院，新竹市。
- 陳建守 (2013)。語言轉向與社會史：科塞雷克及其概念史研究。*東亞觀念史集刊*, 4, 171-221。doi:10.29425/JHIEA.201306\_(4).0004
- 項潔、翁稷安 (2011)。數位人文和歷史研究。在項潔編，*數位人文在歷史學研究的應用* (頁 11-20)。臺北市：國立臺灣大學出版中心。doi:10.6327/NTUPRS-9789860301632
- 馮夏根、胡旭華 (2015)。羅家倫與文學革命。*長江師範學院學報*, 31(1), 59-63。
- 詹筌亦、王乃昕 (2011)。「主義」的數位人文研究。在項潔編，*數位人文在歷史學研究的應用* (頁 219-245)。臺北市：國立臺灣大學出版中心。doi:10.6327/NTUPRS-9789860301632
- 劉吉軒、柯雲娥、張惠真、譚修雯、黃瑞期、甯格致 (2012)。以文本分析呈現臺灣海外史料政治思想輪廓。在項潔編，*數位人文要義：尋找類型與軌跡* (頁 83-114)。臺北市：國立臺灣大學出版中心。doi:10.978.98603/42369
- 劉青峰 (2011)。觀念史研究與數據庫的建立和應用。在項潔編，*數位人文研究的新視野：基礎與想像* (頁 63-81)。臺北市：國立臺灣大學出版中心。doi:10.6327/NTUPRS-9789860301656
- 鄭文惠、邱偉雲、劉昭麟、林書佑 (2016)。概念關係的數位人文研究——以《新青年》中的「世界」觀念為考察核心。在項潔編，*數位人文：在過去、現在和未來之間* (頁 57-102)。臺北市：國立臺灣大學出版中心。
- 鄭文惠、劉昭麟、邱偉雲、許筑婷 (2016)。情感現象學與色彩政治學——中唐詩歌白色抒情系譜的數位人文研究。在項潔編，*數位人文：在過去、現在和未來之間* (頁 207-257)。臺北市：國立臺灣大學出版中心。
- Analytics tools & solutions for your business - Google Analytics*. Retrieved from <https://marketingplatform.google.com/about/analytics/>
- Astobiza, A. M. (2018). Repensando la estructura conceptual del continuum filosofía-psicología: Uso de big data y minería de datos para analizar la transformación

- conceptual de la filosofía y la psicología. *Caracteres: Estudios Culturales Y Críticos De La Esfera Digital*, 7(1), 126-155.
- Haldar, R., & Mukhopadhyay, D. (2011). *Levenshtein distance technique in dictionary lookup methods: An improved approach*. Retrieved from <http://arxiv.org/abs/1101.1232>
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). *Diachronic word embeddings reveal statistical laws of semantic change*. Retrieved from <https://arxiv.org/abs/1605.09096>
- Huang, A. T., Hung, J. J., & Lin, Y. H. (2012). Building a text analysis platform for Chinese buddhist text –An example based on CBETA and tripitaka catalog projects. In *PNC 2012 Annual Conference and Joint Meetings*. Berkeley: UC Berkeley School of Information.
- Hwang, G.J., Yang, L.H., & Wang, S.Y. (2013). A concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education*, 69, 121-130.  
doi:10.1016/j.compedu.2013.07.008
- Jo, E. S., & Algee-Hewitt, M. (2018). The long arc of history: Neural network approaches to diachronic linguistic change. *Journal of the Japanese Association for Digital Humanities*, 3(1), 1-32. doi:10.17928/jjadh.3.1\_1
- Kirschenbaum, M. (2012). What is digital humanities and what's it doing in English departments? In M. K. Gold (Ed.), *Debates in the Digital Humanities* (pp.3-11). Minneapolis, MN: University of Minnesota Press. doi:10.5749/9781452963754
- Knights, M. (2010). Towards a social and cultural history of keywords and concepts by the early modern research group. *History of Political Thought*, 31(3), 427-448.  
doi:10.5840/thought19338340
- Koselleck, R. (2011). Introduction and prefaces to the geschichtliche grundbegriffe. *Contributions to the History of Concepts*, 6(1), 1. doi:10.3167/choc.2011.060102
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. Retrieved from <https://arxiv.org/abs/1301.3781>
- Motzkin, G. (1996). On Koselleck's intuition of time in history. In H. Lehmann, & M. Richter (Eds.), *The Meaning of Historical Terms and Concepts: New Studies on Begriffsgeschichte* (pp.41-45). Washington, D.C.: German Historical Institute.

Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks.

*Science*, 344(6191), 1492-1496. doi:10.1126/science.1242072

Schreibman, S., Siemens, R., & Unsworth, J. (2004). The digital humanities and humanities computing: An introduction. In R. Siemens & S. Schreibman (Eds.), *A companion to digital humanities*. Retrieved from

<http://www.digitalhumanities.org/companion/>

Schreibman, S., Siemens, R., & Unsworth, J. (2008). *A companion to digital humanities*.

Retrieved from <http://www.digitalhumanities.org/companion/>



# Development and Application of Digital Humanities Research Platform with Observation Tool of Viewpoint Migration and Age Division

Cheng-Ho Lin\*   Chung Chang\*\*   Chih-Fan Hsu\*\*\*

## 【 Abstract 】

This study aims to develop “Observation Tool of Viewpoint Migration and Age Division” for supporting research on digital humanities, allowing humanists dividing texts with time continuity into several phases to train word vectors. Word vectors are regarded as the reference for relational vocabulary with the highest relevance in various phases to mine input keywords. Clustering is also applied to assist users in analyzing whether two neighboring phases require split or merge to determine better perspective change observation vocabulary and phase division. Meanwhile, real-time interaction interface is provided for assisting users, through human-computer interaction, in grasping a wide range of contexts with specific

---

\* MA Student, Graduate Institute of Library, Information and Archival Studies, National Chengchi University

Principal author for all correspondence E-mail: handsonicver28@gmail.com

\*\* IT engineer, Research Center for Chinese Cultural Subjectivity, National Chengchi University

E-mail: foxx1216@gmail.com

\*\*\*MA Student, Graduate Institute of Library, Information and Archival Studies, National Chengchi University

E-mail: billxu0521@gmail.com

subjects in texts with time continuity. Experimental research is applied in this study to compare the significant promotion of comprehension of specific subjects in texts and the remarkable change in cognition of research subjects before and after using “Observation Tool of Viewpoint Migration and Age Division”. Technology acceptance questionnaire, and semi-structured interview are utilized for understanding the experiment objects’ opinions and perception of “Observation Tool of Viewpoint Migration and Age Division”. The experimental results reveal that the experiment objects, after using “Observation Tool of Viewpoint Migration and Age Division”, significantly promote the grasp of specific subjects in texts. Furthermore, the interview data and technology acceptance scale reveal highly positive evaluation of the experiment objects on “Observation Tool of Viewpoint Migration and Age Division”. It presents that such a tool could rapidly and conveniently allow them to understanding the context of specific subjects in large amount of texts with time continuity.

### Keywords

Digital humanities, Text mining, Information visualization, Works of Mr. Lo Chia-lun

## 【 Summary 】

It is considered in conceptual history school that important concepts could be regarded as the features of a moment or a period; observing the semantic field change in a concept from “the past transferring to the relative future” could help people understand the change in history concept (Motzkin, 1996). The “observation tool of viewpoint migration and age division” for supporting the research on conceptual history is developed in this study. With the support of word embedding (Mikolov, Chen, Corrado, & Dean, 2013), it would assist junior researchers or postgraduates with certain but incomprehensive understanding of texts in reading text contents with certain association and time

span. Lo Chia-Lun, presenting major influence on the trend of modern Chinese thoughts, is applied to the core to develop the tool, and “Mr. Lo Chia-Lun’s writings” are used as the target analysis text. It is expected that researchers could inquire the concept migration in specific concept centered texts, through the originally limited prior knowledge, to rapidly and automatically assist in dividing text periods in order to more deeply comprehend the context changes of specific concepts in the target text. The system architecture is shown in Figure 1.

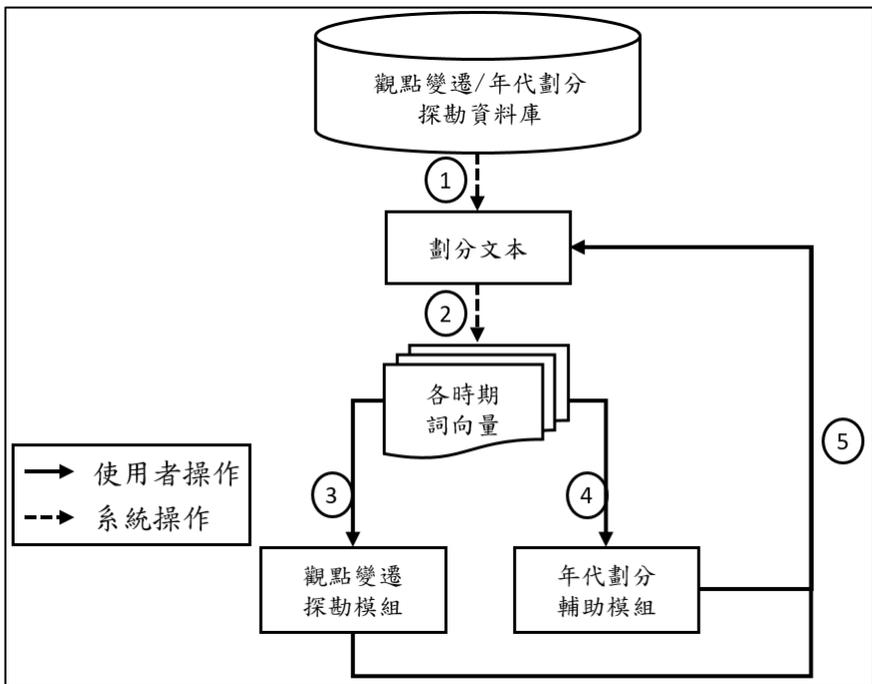


Figure 1 System architecture

• Text division:

For dividing a text with specific time span, the text is divided into several

periods for training word embedding.

• *Viewpoint migration mining module:*

It assists users in mining keyword associated words in different periods and offers users with opportunities for feedback, correction, and repeated analysis and observation. The system would sequentially calculate the similarity of keywords embedded in different periods according to the keyword set keyed in by users, extract top ten associated words with the most similarity to the keyword, and display with visual graphics interface, Figure 2.

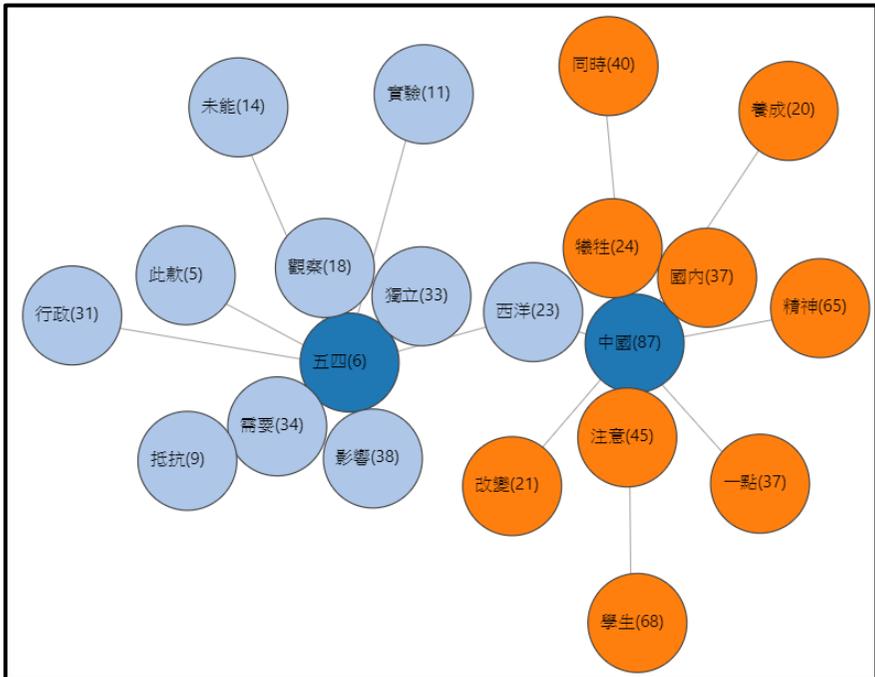


Figure 2 User interface of mining result display

Navy blue nodes in the center are keywords keyed in by the user, the outside nodes are associated words searched according to user's keywords, and

the number in the parenthesis next to the word reveals the number of articles in which the word appears in the period. The outside associated words would merely link with the keyword in the center, the closer link between nodes reveals the larger association, and associated words would not simply link with a keyword node. In Figure 2, the word “western” is associated with two associated words of “May Fourth” and “China”.

With the “period paging” function, users could search the result of similar words in various periods, and the optional pages in “period paging” are the periods divided with “text division”, Figure 3.

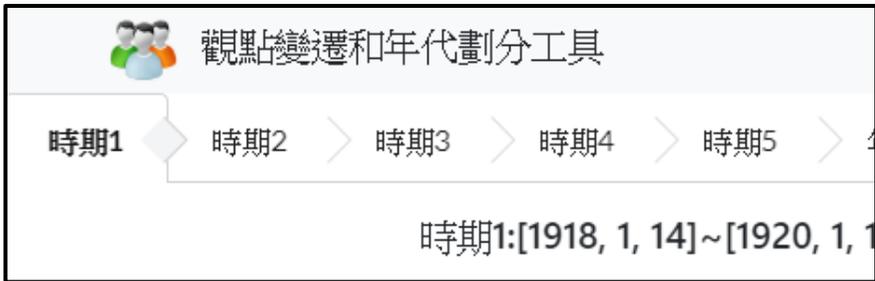


Figure 3 User interface of period paging

When users click on an associated word, the related information to the associated word would be displayed, Figure 4. Text information of title, genre, and text could be searched through the click, Figure 5.

Development and Application of Digital Humanities Research Platform  
with Observation Tool of Viewpoint Migration and Age Division

時期2:[1920, 1, 1]~[1933, 1, 1]

五四  中國

詞彙	出現文章數	與關鍵詞向量	
+	五四	6	該點為主要節點
+	影響	38	0.9996081590652466
-	觀察	18	0.9995607733726501

	體裁	年份	題名
	論著	1921	美國大選舉與新政局
出處	...美國以後，就從各方面 <b>觀察</b> 這個變遷，早思有所陳...		
	論著	1929	贊成「知難行易」說
出處	...耳。故吾人就常識上的 <b>觀察</b> ，頗覺「知難行易」說...		
	論著	1920	一年來我們學生運動底成功失敗和將來應取的方針
出處	...前的失敗，而且照我的 <b>觀察</b> ，我們現在的舉動，實...		
	論著	1921	維持北京國立高等專門以上學校建議
出處	...(二)以各省財政現狀 <b>觀察</b> ，無論何處，欲創辦此...		

Figure 4 User interface of associated words detailed data display

觀點地圖 文本資訊 記事本

搜尋條件:內文中含有「甚麼叫做「五四運動」呢?民國八年五月四」

題名	卷	體裁	民國年	西元年	月	日期	內文
「五四運動」的精神	1	論著	8	1919	5	26	甚麼叫做「五四運動」呢?民國八年五月四日北京學生幾千人因山東問題失敗在政府高壓的壓下，屈於強迫示威，作正當民意的表示。這是中國學生的創舉，是中國教育界的創舉，也是中國國民的創舉。大家不可忘了一字號示威，在外國是常有的事，何以我們竟把他看得大謬小怪呢?不知這次運動有三種真精神，可以顯發中國民族的存亡。第一、這次運動，是學生犧牲的精神。從前我們中國的學生，口裏法理喊天，筆下天花亂墜，到了實行的時候，一個個縮頭縮尾，比起俄國、朝鮮的學生來，真是慚愧死人哩!惟有這次一班青年學生，奮空拳，攜白手，和軍閥勢力相鬥，真的也有。

Figure 5 User interface of text search

• *Age division module*

With the analysis function of clustering to judge the combination or split of neighboring periods, a suitable period observation interval is then divided. Density-based clustering algorithm (Rodriguez 08270327

& Laio, 2014) is utilized for calculating the data of text numeralized with word embedding to provide users with the reference for combining or splitting the age. Clustering algorithm would output a “decision chart”, showing that there are several cluster centers in the analyzed text, and the density  $\rho$  and the distance  $\delta$  of nodes higher than the average, as the threshold of system judgment, is judged as the cluster center, denoted with a red dot. Users then judge to combine or split such two ages, Figure 6 & Figure 7.

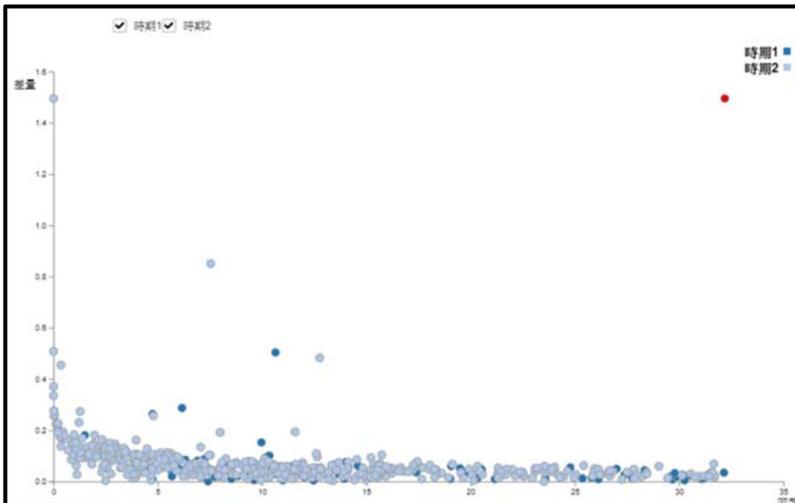


Figure 6 User interface for two-period combination judgment (decision chart)

Development and Application of Digital Humanities Research Platform  
with Observation Tool of Viewpoint Migration and Age Division

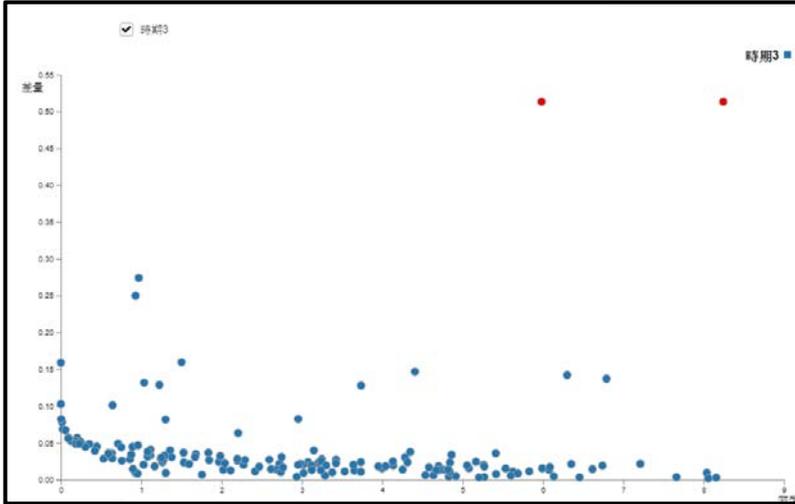


Figure 7 User interface for single-period split judgment (decision chart)

In Figure 6, X-axis and Y-axis on the right are density  $\rho$  and distance  $\delta$ , respectively. A red data node appears higher “density  $\rho$  and distance  $\delta$ ” than other data nodes; accordingly, the decision chart suggests users to “regard the data node as the cluster center”, i.e. “the possibility of combining such two periods into a period”. Figure 7 shows the decision chart of a single period, on which two red dots appear. It reveals that the algorithm judges two cluster centers in the period, i.e. “the possibility to split such a period into two periods”. The system would display the information in different periods for the reference of users, Figure 8.

時期6
 時期7
年代合併

時期序	時期起	時期迄	總文章數	
-	時期6	[1945, 1, 1]	[1955, 1, 1]	218

- [論著 人類良心備忘錄](#)
- [雜著 中華速記班二週年紀念題詞](#)
- [函札 復張國興函 建議以秋瑾為題材拍攝電影](#)
- [函札 復狄燦芝函 奉還「碧血英魂」稿](#)
- [函札 致姚漁湘函 告中央日報增設專欄其編務係由編輯小組負責](#)
- [論著 北京大學的精神](#)
- [函札 復王雲五函 告鄭逸先因事改期趨謁學理兄](#)
- [函札 致葉公超電 英、印、緬均將承認偽政權請速電示](#)
- [函札 上總裁書 報告伊犁軍事進展情形](#)
- [函札 上總裁書 呈國民大會宣言草稿](#)
- [函札 上總裁書 檢呈「陸軍軍官學校三十年」稿本及「黃埔連軍三十年概述」](#)
- [函札 致沈昌煥、許蘭淵函 保舉歐陽無畏在革命實踐研究院受訓](#)
- [函札 致張羣函 謝收回山田所藏史料並請訪查他處藏件](#)
- [序跋 「黃花崗革命烈士畫史」編輯後記](#)
- [雜著 吉隆坡中國報四十二年元旦特刊題詞](#)
- [函札 致王世杰函 告知起程赴印附奉上總裁書請轉呈](#)
- [函札 上總裁書 呈總理民國十一年照片](#)
- [函札 致尼赫魯函 對西藏貿易考察團任何商談結果概不予承認](#)

Figure 8 User interface of information display in various periods

The experimental results reveal that users, after using the tool, significantly enhance the grasp of the topic “Lo Chia-Lun’s viewpoint change in May Fourth Movement” in “Mr. Lo Chia-Lun’s writings” and show remarkable changes on the cognition of the topic “May Fourth Movement”. The experimental objects present positive attitudes toward the technology acceptance with the tool and consider that the tool could help them rapidly grasp large amount of topic changes with association and time continuity in the text. By observing the

experimental objects' behaviors in using the tool, with lag sequential analysis, it is discovered that users with higher effectiveness present the behavior sequence to recycle the core function of the tool.

It is proved in this study that texts with certain chronological order and association could be analyzed with “period split and search of words associated with keywords in various periods”. The future research could analyze different texts, change the calculation methods of associated words in the system, and develop digital humanities teaching tools.

## Romanized & Translated Reference for Original Text

Sun, Junyi (2020). *GitHub - fxsjy/jieba: 結巴中文分詞*。檢自：

<https://github.com/fxsjy/jieba> 【Sun, Jun-Yi (2020). *GitHub - fxsjy/jieba: Jieba Chinese text segmentation*. Retrieved from <https://github.com/fxsjy/jieba> (in Chinese)】

中央研究院人文社會科學研究中心地理資訊科學研究專題中心。 *中華文明之時空基礎架構(簡易版) CCTS Lite*。檢自：<http://gissrv4.sinica.edu.tw/gis/cctslite.aspx>

【Center for GIS, RCHSS, Academia Sinica. *Chinese civilization in time and space(Lite)*. Retrieved from <http://gissrv4.sinica.edu.tw/gis/cctslite.aspx> (in Chinese)】

中國歷代人物傳記資料庫。 *CBDB querying and reporting system -online*。檢自：

<http://db1.ihp.sinica.edu.tw/cbdb/help/systemintro.html> 【China Biographical Database Project (CBDB). *CBDB querying and reporting system -online*. Retrieved from <http://db1.ihp.sinica.edu.tw/cbdb/help/systemintro.html> (in Chinese)】

王泰升 (2011)。數位化歷史資料庫與歷史研究——以明清檔案、淡新檔案、日治法院檔案等資料庫為例。在項潔主編，*從保存到創造：開啟數位人文研究* (頁 31-50)。臺北市：國立臺灣大學出版中心。【Wang, Tay-Sheng (2011). Digital archives and historical studies. In Hsiang, Jieh (Ed.), *From Preservation to Knowledge Creation: The Way to Digital Humanities* (pp. 31-50). Taipei: National Taiwan University Press. (in Chinese)】

doi:10.6327/NTUPRS-9789860300208

田小華、陳磊（2008）。羅家倫與新潮社。《沙洋師範高等專科學校學報》，3，37-39。

【Tian, Xiao-Hua, & Chen, Lei (2008). Luo Jialun and the Xinchao Community. *Journal of Shayang Teachers College*, 3, 37-39. (in Chinese)】

吳承翰、吳尚芸、白璧玲、蔡融易、黃詩芸、蔡宗翰、范毅軍（2018）。基於史籍所載氣象紀錄之事件分類與時空資訊整合研究。在第九屆數位典藏與數位人文國際研討會。臺灣數位人文學會、法鼓文理學院主辦，法鼓文理學院，新竹市。【Wu, Cheng-Han, Wu, Shang-Yun, Bai, Bi-Ling, Cai, Rong-Yi, Huang, Shi-Yun, Cai, Zong-Han, & Fan, Yi-Jun (2018). Ji yu shi ji suo zai qi xiang ji lu zhi shi jian fen lei yu shi kong zi xun zheng he yan jiu. In *9th International Conference of Digital Archives and Digital Humanities*. Symposium conducted at the meeting of the Taiwanese Association for Digital Humanities & Dharma Drum Institute of Liberal Arts. Dharma Drum Institute of Liberal Arts, Hsinchu City. (in Chinese)】

金觀濤（2011）。數位人文研究的理論基礎。在項潔編，*數位人文研究的新視野：基礎與想像*（頁 45-61）。臺北市：國立臺灣大學出版中心。【Jin, Guan-Tao (2011). The theoretical foundation of digital humanistic study. In Hsiang, Jieh (Ed.), *New Eyes for Discovery: Foundations and Imaginations of Digital Humanities* (pp. 45-61). Taipei: National Taiwan University Press. (in Chinese)】  
doi:10.6327/NTUPRS-9789860301656

金觀濤、邱偉雲、梁穎誼、陳柏聿、沈錕坤、劉青峰（2016）。觀念群變化的數位人文研究——以《新青年》為例。在項潔編，*數位人文：在過去、現在和未來之間*（頁 427-463）。臺北市：國立臺灣大學出版中心。【Jin, Guan-Tao, Chiu, Wei-Yun, Leong, Yin-Yee, Chen, Po-Yu, Shan, Man-Kwan, & Liu, Qin-Feng (2016). Alternation of idea groups on *New Youth* (Xin Qingnian): A new digital humanities approach. In Hsiang, Jieh (Ed.), *Digital Humanities: Between Past, Present, and Future* (pp. 427-463). Taipei: National Taiwan University Press. (in Chinese)】

威廉斯（Williams, R.）（1958）。*文化與社會：1780 至 1950 年英國文化觀念之發展*（*Culture and society, 1780-1950*）（彭淮棟譯）。臺北市：聯經。（原作 1958 年出版）【Williams, R. (1958). *Culture and Society: 1780-1950*. (Peng, Huai-Dong trans.). Taipei: Linking Publishing. (in Chinese)】

Development and Application of Digital Humanities Research Platform  
with Observation Tool of Viewpoint Migration and Age Division

- 姚育松（2011）。從「改造」一詞的使用看共產主義在中國的發生背景。在項潔編，*數位人文在歷史學研究的應用*（頁 189-217）。臺北市：國立臺灣大學出版中心。【Yu, Yih-Soong (2011). The background of China's Communism: The study of the keyword "Gaizao". In Hsiang, Jieh (Ed.), *Digital Humanities: New Approaches to Historical Studies* (pp. 189-217). Taipei: National Taiwan University Press. (in Chinese)】doi:10.6327/NTUPRS-9789860301632
- 陳志銘、張鐘（2018）。古籍數位人文研究平台之史料人物關係圖工具發展與應用。在第九屆數位典藏與數位人文國際研討會。臺灣數位人文學會、法鼓文理學院主辦，法鼓文理學院，新竹市。【Chen, Chin-Ming, & Chang, Chung (2018). Development and application of an ancient books digital humanities research platform with characters' relationship map tool. In *9th International Conference of Digital Archives and Digital Humanities*. Symposium conducted at the meeting of the Taiwanese Association for Digital Humanities & Dharma Drum Institute of Liberal Arts. Dharma Drum Institute of Liberal Arts, Hsinchu City. (in Chinese)】
- 陳建守（2013）。語言轉向與社會史：科塞雷克及其概念史研究。*東亞觀念史集刊*，4，171-221。【Chen, Chien-Shou (2013). Linguistic turn and social history: An introduction to Reinhart Koselleck and "Begriffsgeschichte". *Journal of the History of Ideas in East Asia*, 4, 171-221. (in Chinese)】doi:10.29425/JHIEA.201306\_(4).0004
- 項潔、翁稷安（2011）。數位人文和歷史研究。在項潔編，*數位人文在歷史學研究的應用*（頁 11-20）。臺北市：國立臺灣大學出版中心。【Hsiang, Jieh, & Weng, Chi-An (2011). Digital humanities and historical research. In Hsiang, Jieh (Ed.), *Digital Humanities: New Approaches to Historical Studies* (pp. 11-20). Taipei: National Taiwan University Press. (in Chinese)】doi:10.6327/NTUPRS-9789860301632
- 馮夏根、胡旭華（2015）。羅家倫與文學革命。*長江師範學院學報*，31(1)，59-63。【Feng, Xia-Gen, & Hu, Xu-Hua (2015). Luo Jialun and literary revolution. *Journal of Yangtze Normal University*, 31(1), 59-63. (in Chinese)】
- 詹筌亦、王乃昕（2011）。「主義」的數位人文研究。在項潔編，*數位人文在歷史學研究的應用*（頁 219-245）。臺北市：國立臺灣大學出版中心。【Chan, Chuan-Yi, & Wang, Nae-Xin (2011). The "Isms" of digital humanities. In

Hsiang, Jieh (Ed.), *Digital Humanities: New Approaches to Historical Studies* (pp. 219-245). Taipei: National Taiwan University Press. (in Chinese)】  
doi:10.6327/NTUPRS-9789860301632

劉吉軒、柯雲娥、張惠真、譚修雯、黃瑞期、甯格致（2012）。以文本分析呈現臺灣海外史料政治思想輪廓。在項潔編，*數位人文要義：尋找類型與軌跡*（頁 83-114）。臺北市：國立臺灣大學出版中心。【Liu, Jyi-Shane, Ke, Yun-Er, Chang, Hui-Chen, Tan, Hsiu-Wen, Huang, Ruei-Chi, & Ning, Ke-Chih (2012). Text analysis on overseas Taiwanese journals for political thought profiling. In Hsiang, Jieh (Ed.), *Essential Digital Humanities: Defining Patterns and Paths* (pp. 83-114). Taipei: National Taiwan University Press. (in Chinese)】  
doi:10.978.98603/42369

劉青峰（2011）。觀念史研究與數據庫的建立和應用。在項潔編，*數位人文研究的新視野：基礎與想像*（頁 63-81）。臺北市：國立臺灣大學出版中心。  
【Liu, Qing-Feng (2011). Research on the history of ideas and application of the database. In Hsiang, Jieh (Ed.), *New Eyes for Discovery: Foundations and Imaginations of Digital Humanities* (pp. 63-81). Taipei: National Taiwan University Press. (in Chinese)】 doi:10.6327/NTUPRS-9789860301656

鄭文惠、邱偉雲、劉昭麟、林書佑（2016）。概念關係的數位人文研究——以《新青年》中的「世界」觀念為考察核心。在項潔編，*數位人文：在過去、現在和未來之間*（頁 57-102）。臺北市：國立臺灣大學出版中心。【Cheng, Wen-Huei, Chiu, Wei-Yun, Liu, Chao-Lin, & Lin, Shu-Yu (2016). The digital humanities research on concept relations: Focusing on the “World” concept in *New Youth*. In Hsiang, Jieh (Ed.), *Digital Humanities: Between Past, Present, and Future* (pp. 57-102). Taipei: National Taiwan University Press. (in Chinese)】

鄭文惠、劉昭麟、邱偉雲、許筑婷（2016）。情感現象學與色彩政治學——中唐詩歌白色抒情系譜的數位人文研究。在項潔編，*數位人文：在過去、現在和未來之間*（頁 207-257）。臺北市：國立臺灣大學出版中心。【Cheng, Wen-Huei, Liu, Chao-Lin, Chiu, Wei-Yun, & Hsu, Chu-Ting (2016). Phenomenology of emotion and politics of color: Digital humanities research on the lyrical genealogy of “White” in the poetry of the Mid-Tang dynasty. In Hsiang, Jieh (Ed.), *Digital Humanities: Between Past, Present, and Future* (pp. 207-257).

Development and Application of Digital Humanities Research Platform  
with Observation Tool of Viewpoint Migration and Age Division

Taipei: National Taiwan University Press. (in Chinese)】

- Analytics tools & solutions for your business - Google Analytics*. Retrieved from <https://marketingplatform.google.com/about/analytics/>
- Astobiza, A. M. (2018). Repensando la estructura conceptual del continuo filosofía-psicología: Uso de Big Data y minería de datos para analizar la transformación conceptual de la filosofía y la psicología. *Caracteres: Estudios Culturales Y Críticos De La Esfera Digital*, 7(1), 126-155.
- Haldar, R., & Mukhopadhyay, D. (2011). *Levenshtein distance technique in dictionary lookup methods: An improved approach*. Retrieved from <http://arxiv.org/abs/1101.1232>
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). *Diachronic word embeddings reveal statistical laws of semantic change*. Retrieved from <https://arxiv.org/abs/1605.09096>
- Huang, A. T., Hung, J. J., & Lin, Y. H. (2012). Building a text analysis platform for Chinese buddhist text –An example based on CBETA and tripitaka catalog projects. In *PNC 2012 Annual Conference and Joint Meetings*. Berkeley: UC Berkeley School of Information.
- Hwang, G. J., Yang, L. H., & Wang, S. Y. (2013). A concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education*, 69, 121-130.  
doi:10.1016/j.compedu.2013.07.008
- Jo, E. S., & Algee-Hewitt, M. (2018). The long arc of history: Neural network approaches to diachronic linguistic change. *Journal of the Japanese Association for Digital Humanities*, 3(1), 1-32. doi:10.17928/jjadh.3.1\_1
- Kirschenbaum, M. (2012). What is digital humanities and what's it doing in English departments? In M. K. Gold (Ed.), *Debates in the Digital Humanities* (pp.3-11). Minneapolis, MN: University of Minnesota Press. doi:10.5749/9781452963754
- Knights, M. (2010). Towards a social and cultural history of keywords and concepts by the early modern research group. *History of Political Thought*, 31(3), 427-448.  
doi:10.5840/thought19338340
- Koselleck, R. (2011). Introduction and prefaces to the geschichtliche grundbegriffe. *Contributions to the History of Concepts*, 6(1), 1. doi:10.3167/choc.2011.060102
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word*

- representations in vector space*. Retrieved from <https://arxiv.org/abs/1301.3781>
- Motzkin, G. (1996). On Koselleck's intuition of time in history. In H. Lehmann, & M. Richter (Eds.), *The Meaning of Historical Terms and Concepts: New Studies on Begriffsgeschichte* (pp.41-45). Washington, D.C.: German Historical Institute.
- Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, 344(6191), 1492-1496. doi:10.1126/science.1242072
- Schreibman, S., Siemens, R., & Unsworth, J. (2004). The digital humanities and humanities computing: An introduction. In R. Siemens & S. Schreibman (Eds.), *A companion to digital humanities*. Retrieved from <http://www.digitalhumanities.org/companion/>
- Schreibman, S., Siemens, R., & Unsworth, J. (2008). *A companion to digital humanities*. Retrieved from <http://www.digitalhumanities.org/companion/>