檔案與數位人文：數位人文視野下的檔案系統

項潔 國立臺灣大學資訊工程學系特聘教授
兼國立臺灣大學數位人文研究中心主任

本文在寫作過程中得到國立清華大學陳麗華教授和國立臺灣大學數位人文中心董家兒小姐許多的協助，在此致謝。文中所提的系統為蕭屹灵先生建置，詮釋資料為胡乃文與董家兒小姐規劃，在此一併致謝。

一、前言

數位化打破了傳統紙本檔案編排脈絡單一化的限制，大大增加了檔案在呈現和使用上的彈性與便利性，使用數位化的檔案已經成為歷史研究者日常工作中不可或缺的一部分。然而，大多數檔案檢索系統仍然建立在傳統的檔案管理原則上，這種作法雖然保留了檔案的原始脈絡，卻無法掌握眾多檔案之間內在的關聯和錯綜複雜的關係。在本文中，將透過一個多重脈絡分析的概念、一個數位檔案分析系統的方法論和一個具體的系統（地方議會議事錄總庫）中的簡單例子，探討如何使用數位檔案分析系統來達到檔案與檔案、文件與文件間多重脈絡的發掘、分析與呈現，其目的在於一方面指出檔案分析系統和傳統檢索系統的不同，另一方面亦比較數位人文和數位典藏之間在觀念和作法上的差異。

二、數位檔案：歷史研究的新契機

傳統的歷史文獻，受限於紙本實體的特性，只能以線性的方式進行編排；同時遷就於典藏管理機關或人員的編排邏輯或管理原則，只能呈現單一的脈絡，而使用者在使用整編好的檔案時，視野往往受到很大的限制，無法一一關注到文件間可能存在的各種關聯。

檔案的數位化風潮，打破了紙本檔案單一脈絡編排的限制，大大的增加了檔案在使用上的彈性與便利性。不僅可在系統上重現典藏單位的編排脈絡，更可透過系統同時呈現檔案的多重脈絡（如按照時間、地點、人物等重新排列，或顯示人名共同出現的頻率等），並可藉由系統串聯起不同文件，甚至不同卷宗內個別文件之間的關聯，大大提升檔案在研究上的便利性與可用性。（註1）

本文所談的數位化，不僅是將原始檔案進行數位掃描後，原封不動地放入系統而已，而是同時建置較為詳細的詮釋資料（metadata），好的詮釋資料才能使系統功能彰顯，而所謂好的詮釋資料，內容要翔實並且要符合檔案特性，如此才能透過系統呈現出檔案特色，並透過系統進行多元的分析（如我們在處理日治法院檔案的詮釋資料時，即設計有原告、被告、審理法官等欄位）。（註2）另外一個重點在於建置詮釋資料的層次，過去數位典藏時期，部分機關囿於經費與時程，往往以「案」為單位進行檔案數位化建置，但僅做「案」的題名，無法看出檔案內部文件間錯綜複雜的關係。因此，更細緻地在「件」的層次進行描述，才能在呈現檔案多重脈絡時發揮更大的效果。

為數位檔案建置了詳細的詮釋資料後，接下來要設計的檔案檢索系統更需要建置者的審慎思考。因為它的功能將不只停留在關鍵詞檢索上，而是一個既符合檔案自身的特性、亦契合研究者分析需求的「好用」檢索系統。

三、傳統檢索系統的問題

提到檔案檢索系統中的脈絡，人文學者可能還頗覺陌生。但是為了搜尋各種資料庫中有意義的資料，關鍵詞查詢卻幾乎是每日必做的功課。然而，這一資訊方法能夠滿足人文學者的需求嗎？

事實上，我們也的確聽到不少人文學者對於檔案檢索系統的抱怨，特別是檢索結果常常不是相關的條目太少，便是不相關的條目太多。資訊學者觀察到在檢索資料的過程中，使用者會有兩種不同的需求出現。其一是求準率（precision），即檢索結果中，有多少（尤其是排在前面的）是使用者要的；其二則是求全率（recall），即使用者要的文件，有多少被檢索到了。這兩者之間其實是有內在矛盾的；要高的求準率，就會犧牲求全率，而高求全率則會造成求準率的降低。（註3）大體來說，大家多認為一般使用者需要高的求準率，因為一般使用者下某個關鍵詞可能只是想對它的概念做大略的瞭解，而這個需求通常透過檢視前面幾個檢索成果即可滿足；然而一個學者型的使用者（尤其是人文學者）則深怕漏掉重要的資訊，所以要高的求全率，要求檢索結果多多益善。

但事實上是不是如此？每個常用資料庫或網路搜尋引擎尋找資料的研究者應該都

有被資料（檢索結果）淹沒的經驗，而面對數以萬計的檢索結果有望洋興嘆的感覺。這個想像其實有更深層的理由，主要原因是因為搜尋引擎通常假設文件之間沒有關聯，而用一個排序函數來決定檢索成果呈現的順序。因而作為檢索結果的文件之間，其實存在彼此競爭的關係：贏家排在前面，受到使用者第一眼的青睞，輸家則排在後面，有可能被忽視或遺漏。而決定文件輸贏的因素，則取決於不同系統的設定條件，它有可能是根據關聯度、引用率，也有可能是根據商業或其它方面的考量。

但是對歷史研究者而言，檔案是不是一堆沒有關聯的烏合之眾呢？當然不是。檔案之間不但有關聯，而且一個歷史學家訓練很重要的一環就是看出別人無法看到的存在於檔案／文件之間的脈絡。既然如此，一個號稱為學者使用而設計的檔案檢索系統，就不能不把「檔案間的關聯」這個因素放在系統設計的考慮裡。那麼，從這個「檔案間有關聯」的假設出發而設計的檔案檢索系統，又應該提供哪些功能，以便滿足研究者的需求呢？

四、數位檔案分析系統與脈絡分析

我們想像中的數位檔案檢索系統不但要可以幫助研究者找到需要的文件，並且應該能夠找到文件間相關聯的脈絡，以便於探索未知的問題。也就是說，它應該不只是設置一個檢索的窄門，等待研究者投擲敲門磚後依序回應，而應該是一個開放式的寬敞大廳，研究者可以透過不同的路徑進入其中，而且從自己關心的角度仔細觀察大廳中各個部分、家具、器物和裝飾的功能等，揣測其結構、建材來源，甚至推想曾經穿梭其中的人們的物質和精神世界。換言之，我們希望建構的不僅僅是一個數位檔案檢索系統，而是一個數位檔案分析系統。

本文所指出的脈絡（context）的觀念，闡述的便是這樣一個檔案分析系統之功能要捕捉的主要觀念：除了讓使用者可以檢索之外，更應當提供檢索結果之間的關聯。因此，系統的建置便不應該只是簡單的檔案紙本的數位化及檢索，它更應該從研究者的角度，思考研究者所需要的資訊是什麼，系統可以透過怎樣的辦法呈現，甚至將具有研究價值的基本資訊附加於檔案文本之上，為研究者搭建一個更高的平臺。這一問題意識之下，原始檔案文本、系統建置者和研究者的關係，便呈現出不同的樣貌。（註4）

這樣一個有意識的為研究者提供便利數位環境的系統，目標在協助研究者很快地掌握文本間隱含的脈絡，我們把這些脈絡總稱為「文本脈絡」（textual context）。加上「文本」二字，是因為對一個人文研究者而言，文本僅是資料來源的一部分，人文學者所看到的線索，往往有形無形加入了自己累積的知識、文化環境、當下的社會氛圍、

乃至個人偏見等因素。我們稱這種結合無形的、無法量化的訊息形成的脈絡為「無形脈絡」（intangible context），這些無法具形的脈絡不是電腦僅僅透過文本可以掌握的。但是即使資訊科技無法顧到無形脈絡，單單文本脈絡就已經包含了豐富的訊息。

文本脈絡至少有下列三種：首先便是「詮釋資料脈絡」（metadata context）。在數位化過程中，賦予數位物件詮釋資料是非常重要的工作。以歷史文件而言，時間、出處（來源）、作者、主題都可能是重要的詮釋資料，而依據不同文件的特性，詮釋資料也會十分不同。系統可以利用不同的詮釋資料欄位（metadata attributes）將檢索成果進行俯瞰式的編排、分析和呈現，對於研究者而言可能便有意想不到的發現。其次便是「統計脈絡」（statistical context），系統可以迅速對檢索結果、詮釋資料進行量化統計和分析，協助研究者判斷文本的性質、傾向性和代表性等問題。為了讓統計脈絡更有意義，我們往往需要對文本做前置性的處理，如標注人、時、地、物等，這些標注可以用來呈現一些重要的統計資訊，如某人或某地在檢索成果中和關鍵詞的共現性（co-occurrence）。第三種脈絡則比較複雜，即針對文本的特性，設計特殊的方法找出原始文本間特有的關聯，並展示其彼此之間的關係。我們稱這種脈絡為「語義脈絡」（semantic context）。前面兩種脈絡對於任何經過前處理（即已萃取出時間、空間、人物等因素）的文本均適用，但語義脈絡則需要針對文本的特性做設計，用前置作業（pre-processing）的方式，同時找出一個大型資料集所含所有符合所定義的語意脈絡。在一篇較早的文章裡，我們透過文本挖掘（text mining）找出清代地契間的上下手契的關係，並串聯起一塊土地移轉的關係，就是一個語義脈絡的例子。（註5）

當然，上述系統所呈現出來的脈絡，都可以歸納為文本脈絡（textual context），也就是資料庫所包含文件之間的脈絡。但系統不應只是靜態的提供這些脈絡，而且也要提供一個可以讓研究者與脈絡互動的環境。換句話 ，這些脈絡應是可以被人文學者觀察和解釋的（interpretable context）。透過這些觀察、解釋和互動以產生新的文本脈絡，便已經進一步接近人文學者的工作路徑了。雖然系統並不能告訴我們非文本的脈絡，例如使用者的知識、背景、偏好等等，但是作為有「記憶」的系統，最大限度地「接近」這一目標，卻不是不可能的事。

以下，本文試圖以「地方議會議事錄總庫」為例，說明數位檔案分析系統如何利用簡單的系統工具，協助研究者進行檔案分析。

五、「脈絡」：以地方議會議事錄總庫為例

臺灣省議會議政史料的數位化工作從余

玲雅諮議長時代開始，初始合作單位為中研院臺史所，為數位典藏國家型計畫的一部分，數位化內容以議會檔案為主。自 2008 年起與國立臺灣大學數位人文中心持續進行議事錄和公報的數位化，在李源泉諮議長及李雪津秘書長接續努力推動下，完成全部議事資料的數位化與系統開發。這些數位的議事資料，涵蓋自 1946 年臺灣省參議會時代至 1998 年底精省為止的完整內容，記錄議員提案、質詢、省政府報告與提案、人民請願案等發生、審查與執行進度，為透過代議政治，實現臺灣民主政治的過程，在使用上提供了論述臺灣民主政治發展的重要議題，為瞭解臺灣民主政治、相關法案與地方制度的推動的重要資源。從 2013 年開始，再將議政史料的數位化向下延伸，將各個縣市的議事錄和公報也納入數位化的範圍，這項工作在現任議長鄭永金的指導下持續進行。現在建置可查詢的系統如下：

「臺灣省議會史料總庫」（網址：http://ndap.tpa.gov.tw/）：收錄臺灣省參議會、臨時省議會與省議會時期之資料，計有 464,037 筆詮釋資料，影像 183 萬 1 千頁，包含檔案、議事錄、公報三大內容，共 62,852 卷、263 冊，時間自 1946 年迄 1998 年 12 月 20 日，涵蓋臺灣戰後初期臺灣省參議會、臨時省議會、省議會時期。並整合議員小傳與前傳，共計 369 位議員生平介紹。（註6）

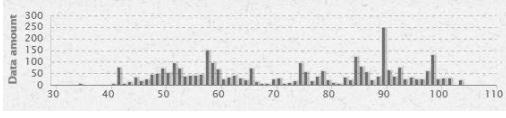
「地方議會議事錄總庫」（網址：http://journal.tpa.gov.tw/）：收錄臺灣各地方議會議事錄與公報，目前系統共有 135 萬餘筆詮釋資料、近 228 萬張影像，並持續擴充中。除臺灣省議會外，並包含臺北市議會、新北市議會（含原臺北縣議會）、臺中市議會（含原臺中縣議會）、臺南市議會（含原臺南縣議會）、高雄市議會（含原高雄縣議會）、基隆市議會、桃園市議會（含原桃園縣議會）、新竹縣議會、新竹市議會、苗栗縣議會、彰化縣議會、雲林縣議會、嘉義縣議會、嘉義市議會、宜蘭縣議會、花蓮縣議會、臺東縣議會、連江縣議會、金門縣議會、澎湖縣議會等地方議會的議事數位資料。

這些資料年代從 1946 年迄今，不僅完整的記載了每次會議的情形，反映過去社會政治、經濟等縮影，呈現民意及政府互動過程，更提供了臺灣民主發展之重要脈絡，也對於當前政府政策與制度改革、臺灣民主政治、議政研究都深具參考價值。將不同議會資料整合在一個系統中讓研究者可以對一個事件觀察不同議會反應與處理的方式，這是分散式的系統無法達到的。

本文用「地方議會議事錄總庫」，以關鍵詞「風災」做為檢索案例做簡單的觀察。

如果將檢索所得的 3,114 筆資料用系統的「議會名稱」後分類進行簡單觀察，可以發現有關風災的議事資料散落在 32 個議會時期（各議會因行政區劃或升格而有不同時期，如桃園縣議會、桃園市議會）。也可以透過詮釋資料中建置的「時間」資訊，觀

察檢索成果的年代分布（這也是一個「詮釋資料脈絡」），可見風災資料自 1946 年起至 2016 年止均有相關資料，而資料分布在各年代的情況與數量同時可透過系統的柱狀圖呈現及數量統計工具，進行觀察。



然而 3,114 筆資料如果要全面翻閱仍然需要相當人力，如果有一個方法能夠先檢視與這些風災關聯的颱風是哪些，就可以對這 3,114 筆資料做比較好的掌握。為了這個目的，我們設計了一個「前後綴詞檢索查詢」的統計分析工具（也就是「統計脈絡」的一種）。這個簡單的工具讓使用者可以對文件進行關鍵詞的前後詞分析，可以設定前後不確定詞的數量，以操作系統的檢索結果並進行觀察與分析，如在檢索欄位填入關鍵詞以及前後綴詞的字數，如「XX 風災」，「XX 公園 XX」等。

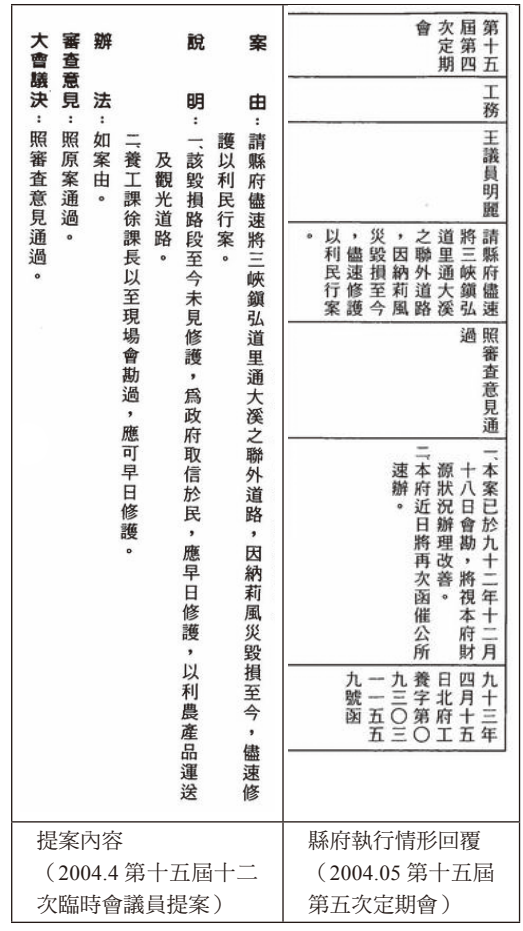
1946 - 民國35年	[7]
1947 - 民國36年	[3]
1948 - 民國37年	[1]
1949 - 民國38年	[2]
1950 - 民國39年	[2]
1951 - 民國40年	[1]
1952 - 民國41年	[7]
1953 - 民國42年	[77]
1954 - 民國43年	[7]
1955 - 民國44年	[15]
1956 - 民國45年	[34]
1957 - 民國46年	[16]
1958 - 民國47年	[25]
1959 - 民國48年	[44]
1960 - 民國49年	[48]
1961 - 民國50年	[70]
1962 - 民國51年	[52]
1963 - 民國52年	[96]
1964 - 民國53年	[72]
1965 - 民國54年	[37]
1966 - 民國55年	[39]
1967 - 民國56年	[41]
1968 - 民國57年	[45]
1969 - 民國58年	[148]
1970 - 民國59年	[97]
1971 - 民國60年	[69]
1972 - 民國61年	[26]
1973 - 民國62年	[32]
1974 - 民國63年	[39]
1975 - 民國64年	[30]
1976 - 民國65年	[22]
1977 - 民國66年	[70]
1978 - 民國67年	[12]
1979 - 民國68年	[4]
1980 - 民國69年	[6]
1981 - 民國70年	[25]
1982 - 民國71年	[28]
1983 - 民國72年	[4]
1984 - 民國73年	[8]
1985 - 民國74年	[18]
1986 - 民國75年	[97]
1987 - 民國76年	[55]
1988 - 民國77年	[19]
1989 - 民國78年	[35]
1990 - 民國79年	[60]
1991 - 民國80年	[20]
1992 - 民國81年	[11]
1993 - 民國82年	[4]
1994 - 民國83年	[34]
1995 - 民國84年	[23]
1996 - 民國85年	[123]
1997 - 民國86年	[78]
1998 - 民國87年	[55]
1999 - 民國88年	[23]
2000 - 民國89年	[37]
2001 - 民國90年	[246]
2002 - 民國91年	[66]
2003 - 民國92年	[38]
2004 - 民國93年	[74]
2005 - 民國94年	[26]
2006 - 民國95年	[34]
2007 - 民國96年	[24]
2008 - 民國97年	[26]
2009 - 民國98年	[62]
2010 - 民國99年	[132]
2011 - 民國100年	[24]
2012 - 民國101年	[29]
2013 - 民國102年	[28]
2014 - 民國103年	[3]
2015 - 民國104年	[21]
2016 - 民國105年	[1]

本文以綴詞查詢：「XX 風災」（臺灣風災名稱多為兩個字），觀察在系統中相關詞彙的狀況，可看出風災名稱顯示在系統中的狀況。

其中，以 2001 年 9 月橫掃臺灣的納莉風災數量最多。因此透過縮小範圍工具，將原有「XX 風災」的相關資料現縮到「納莉風災」的 174 筆進行觀察，可見相關討論分布在 10 個議會之中，其中又以臺北市最多。

另外使用檢索結果後分類工具中的階層顯示，將各議會相關資料年代進行顯示與統計，可見大部分議事資料集中 2001-2002 年，但新北市議會卻集中在 2003-2004 年間（共 33 筆，有 22 筆集中在 2003-2004 年）。

透過內容檢視，新北市至 2004 年為止仍有部分道路尚未修復。議員於第十五屆第十二次臨時會（2004 年 4 月）提案「請縣府儘速將三峽鎮弘道里通大溪之聯外道路，因納莉風災毀損至今，儘速修護以利民行案」。縣府於第十五屆第五次定期會（2004 年 5 月）回覆將再次函催公所速辦。（如下圖）



提案內容（2004.4 第十五屆十二次臨時會議員提案）	縣府執行情形回覆（2004.05 第十五屆第五次定期會）
案　由：請縣府儘速將三峽鎮弘道里通大溪之聯外道路，因納莉風災毀損至今，儘速修護以利民行案。 說　明：一、該毀損路段至今未見修護，為政府取信於民，應早日修護，以利農產品運送及觀光道路。 二、養工課徐課長以至現場會勘過，應可早日修護。 辦　法：如案由。 審查意見：照原案通過。 大會議決：照審查意見通過。	第十五屆第四次定期會 工務 王議員明麗 請縣府儘速將三峽鎮弘道里通大溪之聯外道路，因納莉風災毀損至今，儘速修護以利民行案。 照審查意見通過 一、本案已於九十二年十二月十八日會勘，將視本府財源狀況辦理改善。 二、本府近日將再次函催公所速辦。 九十三年四月十五日北府工養字第〇九三〇一一五五九號函

臺北縣議會第十五屆第十二次臨時會 - [內容]
005b-15-07-000000-0154 @ P.244
» 會議階層 ：臺北縣議會/第十五屆/第十二次臨時會/議決案/本次臨時會議案
» 類別階層 ：議決案/本次臨時會議案
» 檔案日期 ：2004-04-05 ~ 2004-04-08

請縣府儘速將三峽鎮弘道里通大溪之聯外道路，因納莉風災毀損至今，儘速修護以利民行案。
議決案-本次臨時會議案

來源:新北市議會提供

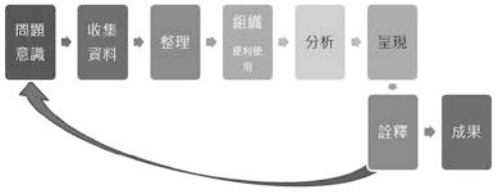
我們希望透過這個非常簡單的例子見微知著，可以窺見如何透過這種脈絡分析和呈現的設計，可以在檢索成果中發現一些並不那麼直觀的現象。研究者可以由此出發，進一步發掘和探討自己感興趣的議題。

六、從數位典藏到數位人文

以上我們描述了一個具有分析性的數位檔案檢索系統，在建置和使用上與傳統檢索系統的不同。簡單而言，它將以研究者為中心的關懷放了進去，不再將檔案視為原始的、不需任何干預的「物」，而是以豐富立體的方式呈現其中隱藏的各種「脈絡」，讓研究者不致深入寶山卻空手而歸，也不致被淹沒在無邊無際的文件裡。臺灣數位化 20

餘年來的歷史，經歷了從早期數位典藏到其後數位人文的轉變，其實便可以說是這一理念和作法轉變的歷程。

我們不妨先來看一下人文研究的歷程。一位人文研究者在運用文本進行研究時，處理步驟約略可分為提出問題，收集資料與整理，對資料進行組織分析及呈現，然後在此基礎上進行詮釋及轉化為研究成果等。

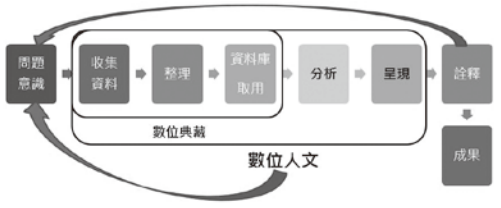


首先就是「問題意識」的提出，即研究者對某一特定議題感到興趣，進而想去理解或提出自己的假設。影響一個問題提出的要件很多，是一個十分發散而自由的過程。之後為了解決這個研究議題而開始著手「收集資料」，尋找相關的書籍與文獻。在取得自己所需的相關資料，研究者會依據自己的需要進行初步的「整理」，如在資料上標記出相關的重要名詞、畫下重點，並寫下簡短的標注和分類等。接下來回應問題意識針對資料開始「組織」，如把相關主題的放在一起，讓資料可以便於使用。對資料進行閱讀後，對前階段所整理出的重點，進行「分析」，找出和所提的問題相關的種種脈絡，並繪出統計表、地圖、樹狀圖等圖表「呈現」整理與分析的結果。之後研究者基於問題意識以及資料分析與呈現的結果進行詮釋與研究成果的發表，但這樣的流程並不是單線的，而是近乎一個研究的循環，透過分析觀察可能需要修改原來的問題意識。（註7）事實上一份好的人文研究應該是要不斷往還反覆，在詮釋的過程中逐漸深化問題意識而產出好的成果。

上述的人文研究流程隨著數位化時代的來臨，也產生了變化。在臺灣透過數位典藏國家型計畫及其他研究與典藏機構的數位化工作，在收集資料方面往往是獨立於個別研究者，針對某一方面的資料做全面性的大量收集，然後透過數位化的手段打造數位典藏資料庫，供給研究者使用。這些工作對於人文研究中「收集資料」、「整理資料」與「組織」使資料方便取用的部分，當然發揮了極大的作用，使得研究者在資料取用上，獲得前所未有的便利。但也就是因為這些數位典藏所收集資料的全面性，反而對研究者造成新的難題，也就是從資料的匱乏轉為對巨量資料的駕馭。因為資料庫並不是為個別研究議題而設計的，所以檢索出來的東西往往比研究者期待的還多很多，而傳統檔案檢索系統所強調的高求全率反而在檢索成果摻雜太多的雜訊，干擾研究者對真正需要的資料的掌握。

數位人文與數位典藏的不同在於除了提供資料外，也希望透過資訊工具來協助研究者進行人文研究中的「分析」與「呈現」，

不僅建置檔案檢索系統囊括數位典藏在收集、整理、取用的功能，並跨領域的結合資訊技術與人文研究，建置符合檔案特性以及研究者需求的數位人文工具。這樣的數位檔案分析系統讓研究者鳥瞰式的分析與觀察文本，提供一個可以觀察、操作與解釋文件脈絡的環境，讓文本的多重脈絡得以基於研究者的問題意識以及判斷來呈現類別、時間、頻率、地理位置等，讓研究者能透過這樣的方式，去釐清、創造自己論述所需的脈絡，並在這樣分析與觀察的過程中產生新的問題意識，逐漸深化問題意識與累積佐證資料，進而產出一份好的研究成果。



七、結語：新的可能性

經過數位典藏國家型計畫的執行，臺灣檔案數位化的經驗與數量在全世界均首屈一指。然而，若能在嫻熟的數位典藏經驗及豐富的數位資源基礎上，進一步建構能夠深入發掘與呈現文本脈絡的數位檔案分析系統，將數位檔案多重脈絡的特性充分發揮，將讓人文學者擁有更強大的鳥瞰史料能力，以及更寬廣的詮釋空間。

因此，數位人文研究的發展，能在人文研究的前段流程：史料收集、整理、取用、分析、呈現上，發揮極大的功效，不但極大地節省了研究者搜尋比對的時間，更對資料有整體鳥瞰式的分析與觀察，有助於研究者進一步專注於非文本脈絡的挖掘和詮釋。系統中脈絡分析與呈現的功能另外一個出人意表的作用，在於提供一個發現研究問題的環境，豐沛的史料加上發掘與觀察脈絡的機制，讓研究者在遨遊於史料之海的同時，亦可能激發出新的問題意識。

當然，數位人文的發展無法取代研究者的思考與詮釋，更無法取代研究者更深層的關懷。本文所提出檔案分析系統多重脈絡的觀念，正是把研究者所需要的、帶有智識性的「客觀」要素突出，而把「主觀」的工作留給研究者自己決定，一方面將數位的效能發揮到最大，另一方面也盡量減少對於研究者的干擾。在一個更直觀、立體和友善的數位環境中，研究者與巨量資料的關係不再是單一的，而是充滿了各種動態的機制。發展新知識的主動權，依然掌握在有志於拓展人文思維和解決人文問題的研究者手中。在臺灣，這樣跨領域的學科發展已經跨出了一步，前面的路還需要更多領域間的對話和激盪。

【註釋】

1. 項潔、翁稷安，〈多重脈絡——數位典藏的問題與挑戰〉，收入項潔主編，《數位人文要義：尋找類型與軌跡》（臺北：臺大出版中

心，2012 年），頁 25-59。

2. 項潔、蕭屹灵、董家兒，〈跨界的日治法院檔案研究〉，收入王泰升主編，《日治法院檔案數位典藏系統之研發與建置》（臺北：元照出版社，2009 年），頁 83-224。
3. Gerard Salton, Michael J. McGill, *Introduction to Modern Information Retrieval* (New York: McGraw-Hill, 1983), pp. 155-156.
4. 有關作者對於幾者關係的思考，亦可參見項潔、翁稷安，〈多重脈絡——數位典藏的問題與挑戰〉，收入項潔主編，《數位人文要義：尋找類型與軌跡》（臺北：臺大出版中心，2012 年），頁 25-59；項潔、翁稷安，〈數位人文和歷史研究〉，收入項潔主編，《數位人文在歷史學研究的應用》（臺北：臺大出版中心，2011 年），頁 18-19。
5. Shih-Pei Chen, Yu-Ming Huang, Jieh Hsiang, Hsieh-Chang Tu, Hou Ieong Ho, Ping-Yen Chen, *Discovering Land Transaction Relations from Land Deeds of Taiwan, Literary & Linguistic Computing*, Vol. 28, No. 2 (June 2013), pp. 257-270.
6. 項潔、彭惠萱、董家兒、蕭屹灵，〈臺灣發展的歷史脈絡——建置臺灣省諮議會史料總庫之總觀〉，《臺灣省諮議會會訊》，第 18 期（2010 年 9 月），頁 34-41。
7. 在此感謝國立清華大學楊叔卿教授指出回圈到「問題意識」的建議。