

## 台灣古契約文書全文資料庫的建置

項 潔\*、陳詩沛\*\*、杜協昌\*\*\*

### 摘 要

國立臺灣大學所製作的「台灣歷史數位圖書館」中的「古契書」文獻集已收錄了兩萬三千餘件的台灣古契約文書，本文將介紹這批契書資料的內容來源，並且介紹「台灣歷史數位圖書館」為這批契書資料建置的各項工具，包括全文與詮釋資料檢索、檢索後分析工具、與探索契書關係的工具。我們希望，透過結合史料與工具，「台灣歷史數位圖書館」能成為一個讓史家盡情探索史料、發現議題的「數位研究環境」。

**關鍵詞：**臺灣史、古契書、全文資料庫、研究環境、檢索後分析

---

\* 臺灣大學資訊工程系教授、臺灣大學數位典藏研究發展中心主任

\*\* 臺灣大學資訊工程系博士班學生

\*\*\* 臺灣大學資訊工程系博士後研究員

## 壹、前言

台灣古契約文書是清代台灣平民遺留的第一手史料。藉由古契約文書記錄的訊息，史家或得以深入或重構台灣歷史上經濟活動、族群關係與土地交易等議題，其重要性不言可喻。然而，當前台灣的古契書大多分散在不同典藏單位與私人收藏家手中，更有許多契書尚未公開。即便如此，台灣史學界已經能利用零散的契書，對清代台灣之土地交易與漢番關係等課題，做出相當精彩的研究。我們認為，若能將這些契書加以集中、串連，其所透露的歷史資訊將更為豐富而有意思。從中浮現的整體圖像，或許也能讓研究者發現當前研究中未能觸及之意義。

有鑑於此，臺灣大學圖書館與臺大數位典藏研究發展中心合作，將文建會委託國立臺中圖書館、臺大圖書館、及台中縣立文化中心所做之全文古契書資料集中在一起，再加上臺大人類學系的古契書資料，現已累積了兩萬三千件的古契書，並仍持續增加中，應已是台灣古契書研究者可以盡情探索的最大資料庫。

這批契約文書的打字全文與詮釋資料，均已建置於「台灣歷史數位圖書館」(Taiwan History Digital Library, 以下簡稱THDL)之中<sup>1</sup>。THDL乃一專為服務台灣史研究者而建的全文資料庫，提供了全文檢索與詮釋資料(metadata)檢索，讓研究者得以在大量的史料中查找所需的資料。但提供檢索只是第一步的工作，我們亦發展了多項分析資料、輔助觀察、進行探索的工具。這些工具或許能夠協助研究者，在大量的史料中，挖掘以人力閱讀時不易察覺的現象，甚至協助研究者深入史料的叢林進行探索。我們希望，透過資料與工具的結合，THDL不僅是台灣史研究者查找資料的必經之地，更成爲一個能讓史家盡情探索史料、發現議題的「研究環境」。

以下，本文將先介紹 THDL 中「古契書」文獻集所收錄的契書來源；其次將說明系統所具有之功能及其設計理念；最後我們將做一結語，探討未來的可能性。

## 貳、資料來源與內容

「台灣歷史數位圖書館」(THDL) 中的古契書是來自國立臺中圖書館、臺灣大學圖書館、臺大人類學系、與台中縣立文化中心近年來的數位化成果，包括：國史館台灣文獻館所藏「臺灣總督府檔案」中抄錄之契約文書、臺大圖書館藏岸裡大社文書、竹塹北門鄭利源號古契書、臺北市文獻委員會典藏臺灣北部相關古契書、臺大圖書館藏臺灣南部古契書、臺灣大學人類學系藏古文書、台南市政府文化局民族文物館暨永漢民藝館古契書、高雄市立歷史博物館古契書、與其他已

<sup>1</sup> 「台灣歷史數位圖書館」的網址位於<http://thdl.ntu.edu.tw>，已開放學術使用申請帳號。

刊之契書，目前總計有 23,250 件契約文書的打字全文與詮釋資料，新的資料建置亦在陸續進行中。

這些契書的數位化工作得以進行，乃有賴近幾年國家努力推動各項數位化計畫的成果。當中最主要的便是行政院文化建設委員會「國家文化資料庫」建置計畫，其指導補助了三十四項古文書數位化計畫<sup>2</sup>、總計完成了超過三萬件古文書的數位化<sup>3</sup>。THDL收錄了「國家文化資料庫」中，由國立臺中圖書館、臺灣大學圖書館、與台中縣立文化中心三單位所執行的計畫成果。而臺大人類學系的古契書數位化工作，則是參與「臺灣大學典藏數位化計畫」<sup>4</sup>的成果。詳細的契書數量與其數位化單位，請見表 1。以下並進一步詳細說明各數位化單位的契書資料來源。

表1 「台灣歷史數位圖書館」收錄古契書的來源與數量列表

數位化單位	件數	來源類型	來源	件數
國立臺中圖書館	17,901	原件資料 (92年)	臺灣總督府檔案抄錄契約文書－15年保存公文類纂	7,299
			台南市政府文化局民族文物館暨永漢民藝館古契書	92
			高雄市立歷史博物館古契書	79
	已刊印資料 (92年)	竹塹古文書	108	
		草屯地區古文書專輯	281	
		台灣省文獻委員會典藏北部地區古文書專輯(1)(2)	476	
		大肚社古文書	106	
		台灣平埔族文獻資料選集：竹塹社(上)(下)	703	
		泉州台灣張土箱家族文件匯編	157	
		原件資料 (93年)	臺灣總督府檔案抄錄契約文書－永久保存公文類纂	5,675
	臺灣總督府檔案抄錄契約文書－15年保存公文類纂		60	
	臺灣總督府檔案抄錄契約文書－土地調查公文類纂		2,397	
	臺灣總督府檔案抄錄契約文書－高等林野公文類纂		468	
	臺灣大學圖書館	4,927	原件資料	岸裡大社文書
竹塹北門鄭利源號古契書				383
臺北市文獻委員會典藏臺灣北部相關古契書				153
國立臺灣大學圖書館藏臺灣南部古契書				87
已刊印資料		十三行博物館館藏古文書	67	
		大崗山地區古契約文書匯編	256	
		水沙連眉社古文書研究專輯	193	
		道卡斯古契文書圖文冊	173	
		道卡斯新港社古文書	78	
		臺灣大學人類學系	362	原件資料
台中縣立文化中心	944	已刊印資料	台中縣立文化中心古文書	392
			台灣中部平埔族古文書	552
<b>總契書件數</b>			<b>23,250</b>	

<sup>2</sup> 國家文化資料庫→古文書，<http://nrch.cca.gov.tw/ccahome/oldbook/> (2009/02)。

<sup>3</sup> 此三萬件並非全是與土地、契約相關的契約文書，尚包含老照片、古信封、明信片...等非契約、非土地相關的文書，因此其中的契書資料應比三萬件要少。

<sup>4</sup> 臺灣大學典藏數位化計畫的網站請見：臺灣大學數位典藏資源中心 (NTU Digital Archives Resource Center, DARC)，<http://www.darc.ntu.edu.tw>。

## 1. 國立臺中圖書館

國立臺中圖書館（國中圖）在行政院文化建設委員會（文建會）的委託下，於民國 92、93 年針對文建會委託臺灣大學歷史系李文良教授所進行之「臺灣關係已刊印古契約文書調查收集計畫」與「臺灣總督府檔案收錄之古契約文書調查蒐集計畫」兩計畫中收集之契書、以及台南市政府文化局博物館課民族文物館暨永漢民藝館以及高雄市立歷史博物館兩單位所典藏之古契書原件進行數位化工作，總計於民國 92 年完成 9,301 件、93 年完成 8,600 件古契約文書的 metadata 建置與全文打字數位化工作，兩年度成果合計為 17,901 件。當中最主要的契書來源是國史館臺灣文獻館所藏之「臺灣總督府檔案」，李文良教授從中蒐集了日本政府為進行地籍整理而徵集抄錄的契約文書與相關土地行政文書，共有 15,899 件<sup>5</sup>。這兩項契書調查蒐集計畫中也進行了已刊印契書的徵集，共八冊，計 1,831 件<sup>6</sup>。

以上數位化成果已分為 92、93 兩年度個別展示於「國家文化資料庫」內<sup>7</sup>，92 年度的成果並另展示於「臺灣地區古文書資訊網」<sup>8</sup>。其中臺灣總督府檔案抄錄契約文書的數位化成果也已部分出版於《臺灣總督府檔案抄錄契約文書》中<sup>9</sup>。

## 2. 臺灣大學圖書館

臺大圖書館於近十年來陸續數位化了幾批古文書資料，包括臺大圖書館館藏之「岸裡大社文書」<sup>10</sup>、鄭華生先生收藏之「竹塹北門鄭利源號古契書」、臺北市文獻委員會典藏臺灣北部古契書、與臺灣大學圖書館典藏臺灣南部古契書，後

<sup>5</sup> 關於「臺灣總督府檔案」抄錄契約文書的詳細內容，請詳見李文良教授的兩份計畫研究報告，以及〈土地行政與契約文書—臺灣總督府檔案抄存契約文書解題〉一文。李文良，《國家文化資料庫：臺灣總督府檔案收錄之古契約文書調查蒐集計畫研究報告書》（臺中：國立臺中圖書館，2003 年）。李文良，《國家文化資料庫：臺灣關係已刊印古契約文書調查收集計畫研究報告書》（臺北：行政院文化建設委員會，2003 年）。李文良，〈土地行政與契約文書—臺灣總督府檔案抄存契約文書解題〉《臺灣史研究》11（2），（臺北：中央研究院臺灣史研究所），頁 221-240。

<sup>6</sup> 八冊已刊印古文書的書名請詳見表 1。

<sup>7</sup> 國中圖 92 年成果展示於「國家文化資料庫」→「古文書」類別下之：「九十二年度臺灣地區古文書數位化」項目或「92 年度台灣地區古文書數位化作業」項目（兩項目內容相同）。93 年成果則展示於同網站之「93 年臺灣地區古文書數位化」項目下。國家文化資料庫，<http://nrch.cca.gov.tw>。

<sup>8</sup> 臺灣地區古文書資訊網，<http://od.ntl.gov.tw/tl>。

<sup>9</sup> 臺灣史料集成編輯委員會編，《臺灣總督府檔案抄錄契約文書·第壹輯》，臺北市：行政院文化建設委員會，2005 年。與：臺灣史料集成編輯委員會編，《臺灣總督府檔案抄錄契約文書·第貳輯》，臺北市：行政院文化建設委員會，2006 年。

<sup>10</sup> 臺大圖書館館藏「岸裡大社文書」的內容並不止於契書，尚有案簿、帳冊...等其他文書，但本資料庫目前仍將之全數收錄，並未排除非契書類的文書。關於臺大圖書館館藏岸裡大社文書的內容與文書類，請見：〈岸裡大社文書〉，《國立臺灣大學圖書館：特藏資源—臺灣研究資源》，[http://www.lib.ntu.edu.tw/CG/resources/Taiwan/taiwan\\_al\\_a.htm](http://www.lib.ntu.edu.tw/CG/resources/Taiwan/taiwan_al_a.htm)。

三項是於國家文化資料庫建置計畫內的「臺灣古契書數位化計畫」執行，共計 3,276 件。臺大圖書館進行的古文書數位化工作不僅包含全文打字與著錄詮釋資料建置，亦進行了原件掃瞄，其原件影像檔均已展示於「國家文化資料庫」內<sup>11</sup>。

近一年來，臺大圖書館更有計畫的調查、蒐集已刊印書籍與學術期刊中所收錄的契約文書，並進行建檔、全文打字、與著錄詮釋資料的建置，成果將陸續加入 THDL。臺大圖書館仍持續進行其他已刊印契約文書的蒐集與全文數位化工作<sup>12</sup>。臺大圖書館亦已規劃挑選其館藏「淡新檔案」內的契約文書，加入 THDL。

### 3. 臺灣大學人類學系

臺灣大學人類學系參與了「臺灣大學典藏數位化計畫」之下的「人類學系典藏文物數位化計畫」。此計畫中完成了該系典藏之 362 件古文書的全文打字、著錄詮釋資料建置、與原件影像掃瞄的數位化工作，其成果已展示於「臺灣大學數位典藏資源中心」<sup>13</sup>。

### 4. 台中縣立文化中心

台中縣立文化中心先後出版了兩批古契書資料，分別是洪麗完女士編纂之「臺灣古文書專輯」<sup>14</sup>，與洪麗完女士所撰之「台灣中部平埔族群古文書研究與導讀：道卡斯族崩山八社與拍瀑拉族四社」<sup>15</sup>。這兩套專書內所收錄的契書亦已完成數位化打字、詮釋資料、與原件影像檔的建置工作，共有 944 件，其成果已展示於「國家文化資料庫」內<sup>16</sup>。台中縣立文化中心也另外架設了「古文書藏品展示」<sup>17</sup>與「台灣中部平埔族群古文書數位典藏」<sup>18</sup>兩網站，同步展示這批數位化契書資料。

<sup>11</sup> 臺大圖書館藏「岸裡大社文書」之數位化成果展示於「國家文化資料庫」→「古文書」之「岸裡大社」項目下。「臺灣古契書數位化計畫」的成果則展示於「國家文化資料庫」→「古文書」之「國立臺灣大學圖書館、行政院文化建設委員會合作數位化古契書計畫」項目下。國家文化資料庫，<http://nrch.cca.gov.tw>。

<sup>12</sup> 臺大圖書館目前已蒐集了近百冊其內收錄契約文書的書籍、期刊，例如楊雲萍藏台灣古文書、大臺北古契字二集／三集、台灣古書契、外埔鄉藏古文書專輯、平埔百社古文書專輯、宜蘭古文書第壹～陸輯、笨港古文書選輯。

<sup>13</sup> 臺灣大學數位典藏資源中心（NTU Digital Archives Resource Center, DARC）：  
<http://www.darc.ntu.edu.tw>。

<sup>14</sup> 洪麗完編撰，《臺灣古文書專輯》，豐原市：臺中縣立文化中心，1996 年。

<sup>15</sup> 洪麗完，《台灣中部平埔族群古文書研究與導讀：道卡斯族崩山八社與拍瀑拉族四社》，豐原市：臺中縣立文化中心，2002 年。

<sup>16</sup> 展示於「國家文化資料庫」→「古文書」之「台灣古文書資料數位化計畫」與「台灣中部平埔族古文書數位計畫」兩項目下。國家文化資料庫：<http://nrch.cca.gov.tw>。

<sup>17</sup> 「台中縣立文化中心」首頁→「數位資料庫」→「古文書藏品展示」：<http://ca.tchcc.gov.tw/histbook/>。

<sup>18</sup> 台灣中部平埔族群古文書數位典藏：<http://ca.tchcc.gov.tw/pingpu/>。

這些契書的內容含括明清以降，台灣各類契約文書，其中又以土地契約為最多，包含買賣、借貸、典當、鬮分等契字；但亦包括非關於土地的契約，如婚姻契；此外還包括了其他相關之公私文書，如契尾、丈單、執照、呈、稟、理由書、證據書類文書……等<sup>19</sup>。

根據王世慶先生的估計，台灣各公私機關擁有之古文書約有兩萬件。<sup>20</sup>就目前學界蒐集的情形看來，應該還不只於此。但無論如何，目前THDL中集結了兩萬三千餘件的契書，已經是相當可觀的數量。我們認為，契約文書的集結對於史家進行研究有絕對正面的幫助，因此，為了讓契書資料更為完整，臺灣大學圖書館正著手蒐集更多的古契約文書，目前已經規劃的有：清代臺灣大租調查書，臺灣私法中收錄之契書、戰後已刊印之古契約文書；而臺大館藏之淡新檔案（已數位化）中亦有不少的契約文書。在推動台灣契約文書研究的前提下，我們亦期待能與更多單位合作，讓研究者能夠更方便地獲得所需之史料。

### 參、系統設計理念與功能介紹

THDL 是一個專為服務台灣史研究者而建的全文資料庫，但它不只是一個提供大量史料、便於研究者查找的資料庫；它還是一個具備了各項「工具」，以幫助研究者更能有效運用這批大量史料的「研究環境」。

THDL提供了查找資料過程中所需的輔助工具，盡可能讓研究者能找到所需之史料。另一方面，我們開發了多項分析檢索結果的工具，幫助研究者觀察以人力閱讀時不易察覺的現象。最後，我們還試著以自動的方式重構契書之間的關連性，或許可作為研究者探索這批契書的另一種途徑。在此我們要略為說明三者之功用，及其設計理念。<sup>21</sup>

<sup>19</sup>這兩萬餘件的契書乃來自不同的數位化單位，對於契書的分類並無一致的分類準則。為了讓研究者易於使用這批大量的契約相關文書，我們對其進行了初步的分類。我們發展了一套基於「契書標題」的電腦自動分類方法，將這批契書重新以一致的分類準則進行分類，目前分為：杜賣契、鬮分契、公文書類、合約字、典契、丈單/執照、開墾契、證明/協議/決議、贖耕字、胎借字、添典/找洗、目錄、合股、其他等十四類。但此分類成果目前仍不夠完善，請研究者斟酌使用。關於此自動分類的詳細方法，請見：盧家慶，〈台灣古契書自動分類與依分類定義契書角色〉，臺北：國立臺灣大學資訊工程學研究所碩士論文，2008年。

<sup>20</sup>王世慶先生曾於〈介紹臺灣史料：檔案、古文書、族譜〉一文中估計：『目前古文書的保存數量，除日治時代排印的不算……初步估計，目前臺灣古文書至少有二萬多件。』。王世慶，〈介紹臺灣史料：檔案、古文書、族譜〉，《臺灣史與臺灣史料：臺灣史料評析講座紀錄（一）》（臺北市：自立晚報，1993年），頁71。

<sup>21</sup>更深入的討論，請見以下兩篇論文：Chen, S.P., Hsiang, J., Tu, H.C., and Wu, M.C. "On Building a Full-Text Digital Library of Historical Documents". In *Proceeding of the 10th International Conference on Asian Digital Libraries Held in Hanoi, Vietnam, December 10-13, 2007*, edited by Dion Hoe Lian Goh, Tru Hoang Cao, Ingeborg SÅlvberg, Edie Rasmussen, 49-60. Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers. New York: Springer Berlin, 2007。

## 1. 找到使用者需要的：查找資料過程中的輔助工具

THDL的古契書文獻集已經集結了兩萬餘件古契書，這麼大量的契書該如何發揮其輔助研究的效用呢？我們認為最首要、根本的就是「讓研究者找得到所需資料」，而「全文檢索」便是資料庫是否具備此能力的關鍵。全文檢索，提供了史料的完整全文作為檢索時可利用的線索，讓研究者逐字檢視的檢索需求變成可能。換言之，即使在詮釋資料（metadata）欄位內不具相關描述，只要史料的全文當中出現了該線索，便可利用全文檢索找到該件史料。然而，全文檢索可能導致找到的雜訊太多，這時候，「詮釋資料檢索」便可提供更為精確的檢索方式，將查詢詞彙限定於詮釋資料欄位內進行檢索。THDL中提供了全文檢索與詮釋資料檢索兩項功能，可視研究者的需要，互相搭配使用。下圖便是以「金廣成」<sup>22</sup>進行檢索後的畫面。

目前位於何文獻集 (全文暨所有欄位檢索)

文件檢索

THDL 台灣歷史數位圖書館

文獻集: [全部] | 明成德 | 古契書

文件檢索: 金廣成

最近幾次查詢

金廣成	61
[TM]日明治三十四年 金廣成	15
[OldDeeds_SimpleDi...	1
+章程 +金廣成	7

個人文件集

個人文件集	2
[金廣成查詢]	2

檢索後分類

年代 (TM)	61
光緒	36
清光緒十三年 (1887)	13
清光緒十五年 (1889)	15
清光緒十六年 (1890)	2
清光緒十七年 (1891)	2
清光緒十八年 (1892)	1
清光緒十九年 (1893)	1
清光緒二十年 (1894)	1
明治	26
日明治二十九年 (1896)	2
日明治三十年 (1897)	1

呈現模式與排列方式

檢索結果年代分佈圖

檢索結果分佈圖 (此圖僅計入含有年代資訊的文件: 橫軸: 西元年, 縱軸: 文件數量)

1. 奉諭開墾合約字 | 臺灣總督府檔案

2. 奉諭開墾合約字 | 臺灣總督府檔案

圖1 THDL 檢索後的系統呈現

研究者有時或需以多個詞彙聯合檢索、或需排除某詞彙出現的文件，THDL也提供了這樣的檢索功能，讓研究者得以彈性地搭配、組合查詢詞彙，以查找所需的資料。例如以「+金廣成 +開墾」進行檢索，系統會回傳「同時含有“金廣成”、又含有“開墾”」的契約文書，而「+金廣成 -開墾」則會在「含金廣成」的契書中、排除「含開墾」的契書。

杜協昌，〈互動式網際網路檢索：模型、度量及實驗〉，臺北：國立臺灣大學資訊工程學研究所博士論文，1998年。

<sup>22</sup>「金廣成」為一墾號，開墾今日的關西地區。

THDL 也對於研究者查找資料的過程，以及找到的資料，提供了簡單的管理工具：「最近幾次檢索」自動紀錄了最近四十次的檢索歷史，讓研究者便於回顧查找資料的過程；「個人文件集」則讓研究者可將有用的資料儲存在自訂名稱的文件集中，下次想調閱資料時便不需重新檢索，尤其當研究者正為特定主題蒐集資料時，更可陸續將不同時候、不同狀況下找到的資料加入文件集中。

此外，THDL 可依照研究者的需求，以不同的「呈現模式」、「排列方式」顯示檢索回傳的文件，「詞彙標示」的功能會在全文中醒目標示查詢詞彙，幫助研究者快速檢閱回傳結果。

## 2. 觀察使用者找到的：檢索後分析工具

THDL 想為研究者提供的，不僅是「檢索」罷了。對於如此大量的契書資料，僅僅提供檢索並不足以發揮其輔助研究的功用。我們認為，系統在檢索之後，應為研究者分析檢索結果中的重要特徵，讓研究者得以快速的掌握文獻，系統更應藉由檢索後分析的工具，提供研究者用以觀察檢索結果的「新眼光」。或許這樣的工具，能夠協助研究者在大量的史料中，挖掘以人力閱讀時不易察覺的現象。

我們認為，經由同一檢索條件而回傳的多件契書之間，應該是彼此相關的，若將檢索結果視為整體來看，所能呈現的意義絕非單件契書能夠呈現。例如以「地區 A」進行檢索得到的契約文書，若將之串連起來，有可能勾勒出地區 A 的土地開發史；這便是這群契書所共同訴說的「整體意義」。基於這樣的理念，我們開發了以下對檢索結果進行分析的工具，希望能透過檢索後分析的工具，協助研究者觀察檢索結果所呈現的整體意義。

### (1) 「檢索結果年代分佈圖」：

是將檢索回傳的契書中，統計每一年份的契書數量共有幾件<sup>23</sup>，再以橫軸為西元年、縱軸為契書數量，繪製成折線圖。「立契時間」原本就是契書的重要特徵，藉由畫出視覺化的年代分佈圖，研究者更能感受到檢索結果在時間上的分佈狀況。以圖 2 為例，可明顯看出回傳文件的年代集中於特定區段，且有兩個顯著的高峰。

「年代分佈圖」僅提供回傳文件於時間軸上分佈狀況的概貌，當研究者需要詳細了解每一年份的契書數量，則需進一步使用「檢索後分類」工具中的「年代」後分類。

<sup>23</sup> 檢索結果中，立契時間不詳的契書則不計入。

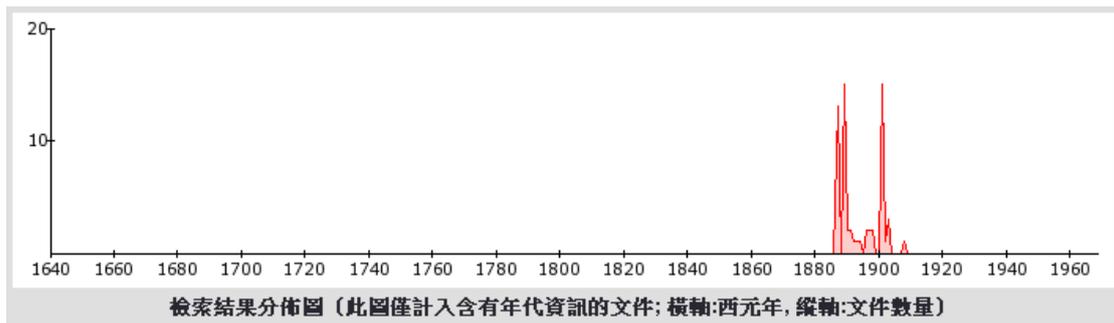


圖2 「檢索結果年代分佈圖」

(2) 「檢索後分類」：

是系統對於檢索回傳文件進行的「檢索後分析」之一，此工具乃根據年代、出處、作者、契書分類等四種詮釋資料欄位，將檢索回傳的文件分類至各個類別之中，並計算每個類別擁有的文件總數。藉由觀察檢索後分類的結果，使用者可以得知檢索回傳文件在年代、出處、作者、契書分類上的分佈情況。尤其是當檢索結果的量太大時，使用者並不需將回傳文件一一檢視完畢，即可快速掌握檢索結果的概貌。

圖 3便是對檢索「金廣成」所得的 61 件契書，再進行「檢索後分類」的結果。在「年代後分類」表中詳細列出了 61 件契書如何分佈於各個年份。可以觀察到：其中最早的契書年代是光緒十三年，有 13 件，接下來則到光緒十五年，有 15 件契書，而到了光緒十六年則僅有 2 件；總計，清光緒年間共有 35 件「含金廣成」的契書。而往後到了日據明治年間，又有 26 件契書。從「出處後分類」中可以看到，THDL內「含金廣成」的契書都是出自於「臺灣總督府檔案抄錄契約文書」之中，有 60 件是出自「永久保存公文類纂」，1 件出自「高等林野公文類纂」。而從「作者後分類」則可看到這 61 件契書中的人物、與其在契書中扮演的角色各為何，例如「張秀欽」的角色曾是「墾首」、「墾戶」、「股夥」，甚至曾是「賣主」。最後，「契書分類」中則可初步解析這 61 件契書的性質，鬮分契有 27 件、杜賣契有 13 件……等。以上的檢索後分類工具，或許能幫助研究者，在尚未開始閱讀回傳文件以前，便對「含金廣成」的檢索結果有了大致的了解。

「檢索後分類」也可作為「縮小範圍檢索」之用：以滑鼠點擊任一類別便是對該類別的進一步檢索，幫助研究者聚焦於具有特定特徵的契書上。例如點選「作者後分類」中的「張秀欽（賣主）」，便可進一步閱讀檢索結果中，張秀欽做為賣主的那 2 件契書。



現於 11 篇契書中，與其他三位股首的出現件數具有明顯落差。而作為第一股首的「張秀欽」也僅出現於 38 篇契書，並未在全部 61 篇當中佔有高比例。再反觀地名的詞頻分析結果，出現於最多件契書的地名是：「十寮庄」（57 篇）、「竹北二堡」（56 篇），佔了 61 篇契書中的相當高的比例，顯示「竹北二堡十寮庄」與金廣成具有密切的關係，或許便是金廣成墾號當年主要的開墾地區。

在詞頻列表中點選任一人名或地名，也具有「縮小範圍檢索」之功用，系統將回傳：此檢索結果中，含有該人名或該地名的契書，以供研究者查閱。例如在圖 4 中點選「徐連昌」，便可閱讀「含金廣成」的檢索結果中、「又含有徐連昌」的契書，以進一步了解「徐連昌」在契書中出現的情況，或許有助於了解他在金廣成各股東之中扮演的角色。

我們在「人名」的詞頻列表中，另安插了一個小工具：若人名的右方有「生平」的連結，點選後便可閱讀國家圖書館「臺灣歷史人物小傳」<sup>26</sup>與故宮「清代檔案人名權威資料查詢」<sup>27</sup>對於該人物的生平描述。

THDL 使用的人名、地名部分來自詮釋資料中的「作者」、「關係人」、「內容地點」等欄位，然而為了能完整分析曾出現於契書內文的人物、地點，我們又以電腦輔助，從契書全文中擷取了更細緻的人名與地名<sup>28</sup>。

#### （4）比較工具：「?vs」與「?in\_cnt」

是用來「比較兩種檢索條件所得結果」的工具。其中的「檢索一 ?vs 檢索二」，是將兩檢索所得的結果繪製於同一張年代分佈圖上，紅色線是「檢索一」的年代分佈曲線，而灰色線則為「檢索二」的年代分佈曲線。此工具可便於研究者比較兩檢索條件所得的契書數量差異、以及分佈的年代區段。

例如，若想比較「金廣成」與「金廣福」的檢索結果（兩者同為新竹地區的墾號），可在文件檢索框中輸入「金廣成 ?vs 金廣福」，便會得到如圖 5 的畫面。此時的「年代分佈圖」同時呈現了含「金廣成」與含「金廣福」的年代分佈曲線（圖 6），可從圖中觀察到，在 THDL 中含「金廣福」的契書數量，比含「金廣成」的契書數量多上許多，並且，含「金廣福」契書的出現年代也比含「金廣成」契書的出現早了很多。

需請使用者注意的是，在「?vs」工具的結果畫面中，僅有「年代分佈圖」是同時使用了「檢索一」與「檢索二」的結果；畫面中的其他資訊，皆與「檢索一」的結果相同。再以「金廣成 ?vs 金廣福」為例說明，圖 5 中「找到筆數」指

<sup>26</sup> 國家圖書館，臺灣記憶→人與事→臺灣歷史人物小傳：  
[http://memory.ncl.edu.tw/tm\\_cgi/hypage.cgi?HYPAGE=toolbox\\_home.hpg](http://memory.ncl.edu.tw/tm_cgi/hypage.cgi?HYPAGE=toolbox_home.hpg)。

<sup>27</sup> 故宮博物院，清代檔案人名權威資料查詢：  
<http://www.npm.gov.tw/zh-tw/learning/library/archives.htm>。

<sup>28</sup> 張尚斌，〈詞夾子演算法在專有名詞辨識上的應用 - 以歷史文件為例〉，臺北：國立臺灣大學資訊工程學研究所碩士論文，2006 年。

的是含「金廣成」的契書件數，畫面下方列出的契書也是含「金廣成」的契書，而「檢索後分類」亦是針對含「金廣成」契書的分析結果。

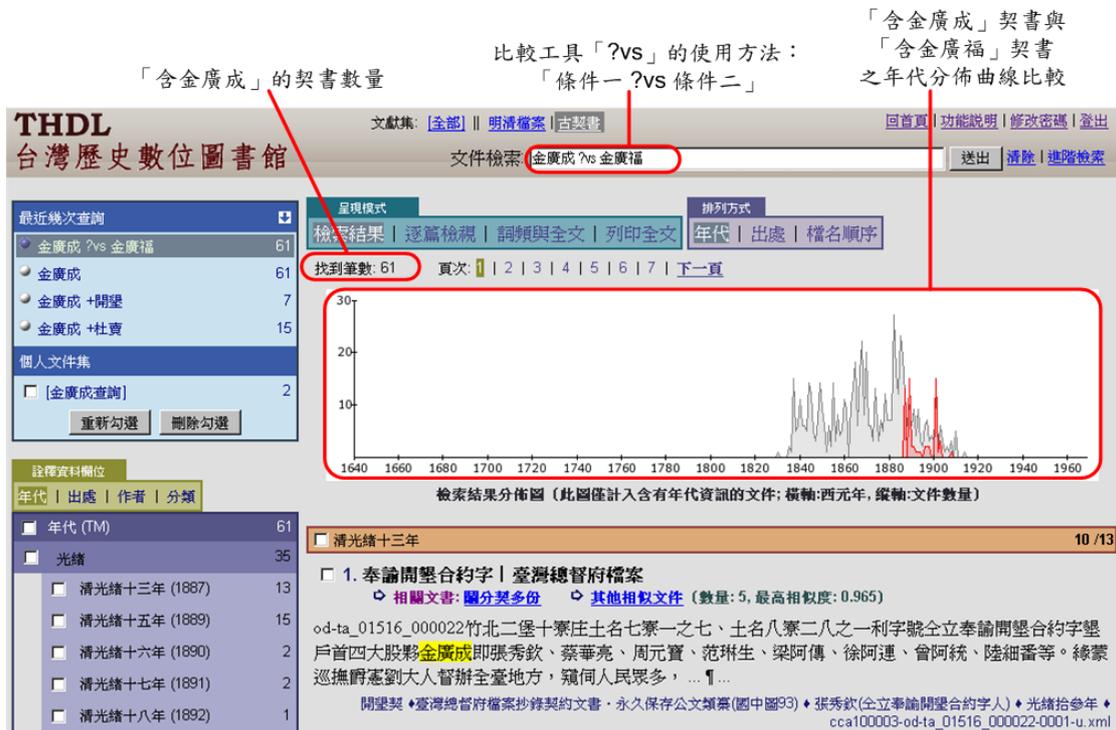


圖5 「金廣成 ?vs 金廣福」檢索後的系統畫面

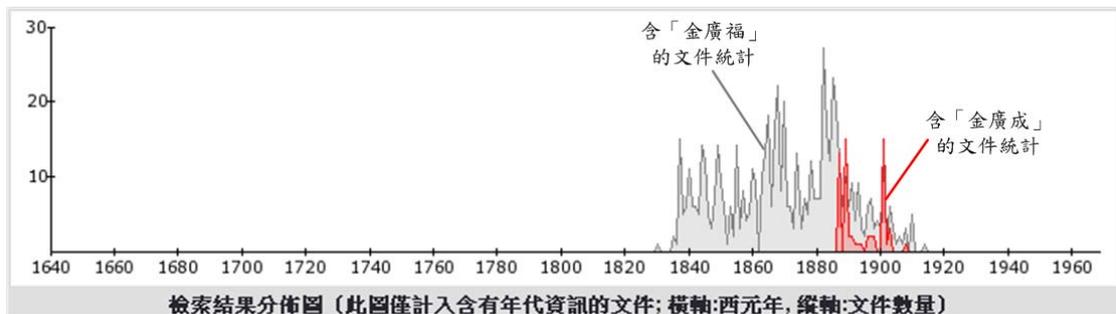


圖6 「金廣成 ?vs 金廣福」的年代分佈比較圖

而另一個比較工具「檢索一 ?in\_cnt 檢索二」則可幫助研究者找出兩檢索條件的交集文件。在檢索後，於「找到筆數: X筆」的右側會多出一項「in: Y」的訊息，其中X代表符合「檢索一」的契書件數，而Y代表同時符合了「檢索一」與「檢索二」的契書件數。例如以「找洗 ?in\_cnt 杜賣」檢索，會得到如圖 7 的畫面。可見到「找到筆數」為 1,946 筆，代表THDL中含「找洗」的契書有 1,946 件，其右側的「in: 1,473」則代表既含「找洗」、又含「杜賣」的契書有 1,473 件。而在畫面下方列出的含「找洗」契書中，若於標題處出現了「in」的標記，代表

該件契書除了含「找洗」、也同時含了「杜賣」。

「含找洗」者有 1,946 件  
其中同時「含杜賣」者有 1,473 件

比較工具「?in\_cnt」的使用方法：  
條件一 ?in\_cnt 條件二

THDL 台灣歷史數位圖書館

文獻集: [全部] | 明清檔案 | 古契書

文件檢索: 找洗 ?in\_cnt 杜賣

最近幾次查詢: 找洗 ?in\_cnt 杜賣 1946

檢索結果 | 逐篇檢視 | 詞類與全文 | 列印全文

找到筆數: 1946 in: 1473 頁次: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | >> | 下一頁

檢索結果分佈圖 (此圖僅計入含有年代資訊的文件; 橫軸: 西元年, 縱軸: 文件數量)

1. 杜賣盡根茅屋及厝地契字  
此件既「含找洗」又「含杜賣」

5. 找洗契字  
此件僅「含找洗」, 不「含杜賣」

圖7 「找洗 ?in\_cnt 杜賣」的檢索結果

以上對於檢索結果進行各式分析的工具，皆是 THDL 提供研究者以「不同於閱讀」的新眼光、觀察檢索結果的嘗試。藉由這些工具，或能協助研究者快速掌握檢索結果的特徵，或能輔助研究者發現以人力閱讀時不易察覺的現象。而當檢索結果的量太大，難以人力進行閱讀時，這些「檢索後分析」工具便能以電腦分析代替人為的閱讀，先一步將檢索結果文件中的重要特徵呈現給研究者，讓研究者先行掌握檢索結果的整體圖像。或許研究者便能從中得到足夠的線索，知道下一步該如何篩選檢索結果、聚焦在自己關心的文件上，而不至於被大量的檢索結果所淹沒。

### 3. 探索使用者可能不知道的：相似文件與相關文件

契約文書彼此之間可能存在著關連性，例如兩件契書是關於同一塊土地，或兩者的內容極為相似。這些彼此有關的契書，其實體物件或許擺放在一起，但數位化之後，關係便不復存在；又或者，相關的契書早在流傳的過程中便分散了。

當 THDL 匯集了來自各處的契書之後，或許能意外將相關的契書重新聚集在一處。然而，如何在 THDL 兩萬餘件的契書中，找到彼此有關的契書，也因為契書的數量太大，而難以人力挖掘。

為此，THDL的建置團隊藉助了資訊科技的幫忙，試圖重新建構契書之間的關連性，目前重構的關係包括內容相似的契書，以及關於同一土地的相關契書。我們將重構後的契書關係以連結的方式置於THDL的畫面上（如圖 8），希望能提示研究者「在THDL中還有其他與此件相關的契書存在」，研究者若循線探索，或許可以串連出有意思的契書脈絡，這樣的線索也或許作為研究者探索這批大量契書的另一種途徑。

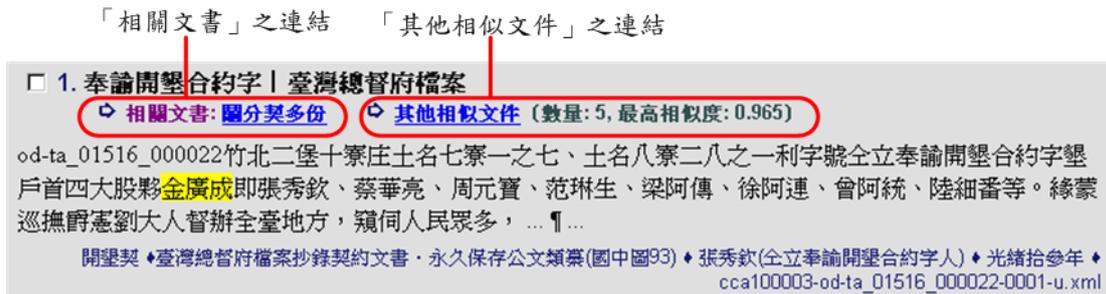


圖8 「相似文件」與「相關文書」之連結

(1) 「相似文件」：

我們已對THDL收錄的兩萬三千餘件契書，計算其兩兩之間的内容相似度。研究者可利用THDL上的「其他相似文件」連結，進一步探索與此篇契書相似的其他文件。例如點選圖 8的「其他相似文件」後，系統會呈現與此件（基準文件）相似的其他 5 件契書（圖 9）。此例中，相似的各篇契書乃是「金廣成墾號」於光緒十三年合股開墾時，抽得不同圖的各股所持的契約。

2. 奉諭開墾合約字 | 臺灣總督府檔案

- 相關文書
- 相似文件 (5, 0.967) 0.965
  - 臺灣總督府檔案抄錄契約文書，永久保存公文類彙(國中圖93) · 光緒拾參年
  - · cca100003-
  - od-ta\_01516\_000172-0001-u.xml

3. 奉諭開墾合約字 | 臺灣總督府檔案

- 相關文書
- 相似文件 (5, 0.971) 0.946
  - 臺灣總督府檔案抄錄契約文書，永久保存公文類彙(國中圖93) · 光緒十三年
  - · cca100003-
  - od-ta\_01516\_000054-0001-u.xml

4. 奉諭開墾合約字 | 臺灣總督府檔案

- 相關文書
- 相似文件 (5, 0.980) 0.939
  - 臺灣總督府檔案抄錄契約文書，永久保存公文類彙(國中圖93) · 光緒拾參年
  - · cca100003-
  - od-ta\_01516\_000154-0001-u.xml

2. 奉諭開墾合約字 | 臺灣總督府檔案

- 相關文書: 鬮分契多份 (Groups: 585,590,596,602) 其他相似文件 (數量: 5, 最高相似度: 0.967) 與基準文件相似度: 0.965

檔名: cca100003-od-ta\_01516\_000172-0001-u.xml  
 標題: 奉諭開墾合約字 | 臺灣總督府檔案  
 作者: 張秀欽(股夥)  
 出處: 臺灣總督府檔案抄錄契約文書，永久保存公文類彙(國中圖93)  
 分類: -  
 成文日期: 光緒拾參年

od-ta\_01516\_000172竹北二堡十寮庄土名箕箕窩一三八利字號全立奉諭開墾合約字墾戶首金廣成四大股夥即張秀欽、周元寶、梁阿傳、曾阿統、蔡華亮、范琳生、徐阿連、陸細番等。緣蒙巡撫爵憲劉大人督辦全臺地方，窺伺人民眾多，生齒日繁，目擊青山田野曠廢，土番戕害生靈。爰欲辟土地，正經界，發政施仁，撫綏番民，因于大嵙崁設撫墾總局，示諭撫番招墾，以廣國家王土，以安兆姓耕耘。但一人難以兼攝，獨力莫能維持。欽等崇奉爵憲至意，爰邀集同心合夥四大股等，各自傾囊，共津需費，築成墾戶首，公賦金廣成。於本年三月間，各股夥親造總局，具稟呈請局憲委員陸、陳尊前，稟明開墾情形等。因經總局憲施恩，通詳爵憲大人鴻裁，即蒙爵憲俯憐垂恩，不棄庸劣，准將所墾大竹坑界後崇墾山名曰福興庄一帶青山前後左右山場。東至彩和山龍崗直透為界；西至金山嶼橫透

各相似文件之條列

以深灰色背景標示「與基準文件相同的文句」

圖9 「相似文件」

## (2) 「相關文書」：

土地契約之間可能存在有土地的移轉、典賣、鬮分等關係，這些關係是研究土地開發及地權轉移過程與模式的重要依據。然而彼此相關的契書，有可能分散在各來源之中，若要靠人力找出這樣的關係，需要花費相當大的力氣。並且，研究者不太可能同時關注多條主軸。借助資訊科技的幫忙，我們可以節省下很多時間，並且同時找到許多條脈絡，串連起原本分散的契書。

我們利用契書的詮釋資料以及數位化全文的內容，目前已自動重構了部分的契書關係：上下手契關係、原契與契尾、鬮分契一式多份、與契書內容相同<sup>29</sup>。研究者可利用THDL中的「相關文書」連結，檢視具有這四種關連性的契書。例如點選圖 8 的「相關文書：鬮分契多份」之後，系統便會呈現鬮分後的各鬮契字。

除了上述工具以外，THDL也提供了「原文參照」的工具，讓研究者便於參照契書的原始出處與參考原件影像。另，由於大量的資料在建置過程中難以避免錯誤，我們亦提供了「錯誤回報」的工具，歡迎研究者協助我們發現、改正錯誤。這兩項工具於THDL中的顯示位置請見圖 10。<sup>30</sup>

<sup>29</sup> 契書可能因為由鬮分之後不同的人提供給總督府登記，造成上手買契出現重複抄寫（例如可能只有登記的地號不同）的情況。

<sup>30</sup> 「原文參照」工具與metadata及全文的「錯誤回報」工具，乃置於「逐篇檢視」與「詞頻與全文」兩種「呈現模式」之下。而人名、地名等詞彙的錯誤回報，請至「詞頻與全文」模式下使用。

3.1. 杜賣股份山業字 | 臺灣總督府檔案

相關文書: [契書內容相照](#) 其他相似文件 (數量: 1, 相似度: 0.928)

檔名: cca100003-od-ta\_01516\_000021-0001-u.xml  
 標題: 杜賣股份山業字 | 臺灣總督府檔案  
 作者: 張秀欽(立抽出祖墳山批字人)  
 出處: 臺灣總督府檔案抄錄契約文書·永久保存公文類彙(國中圖93)  
 分類: 杜賣契  
 成文日期: 光緒十三年

od-ta\_01516\_000021

**竹北二堡十寮庄土名七寮一六之四**  
 立抽出祖墳山批字人公號**金廣成**四股首張秀欽、周元寶、徐連昌、蔡華亮等。緣因光緒十三年間奉爵撫憲諭，自備工本招佃墾闢崇墘山後面，至彩和山前面，至**糞箕窩**，至水(石祭)仔，以及**大粗坑**前后兩面山場田園等處。茲有股夥蔡華亮出首前來，請得眾股允意，到地面踏分明，抽出祖墳四穴，其壹穴址在**糞箕窩**一穴，坐東向西，前后各四丈，左右各三丈。又祖墳壹穴址在**七寮庄**背西片小廍透上大嵵下壹穴，坐北向南，前后各四丈，左右各三丈。又祖墳址在**八寮庄**背由龍透上嵵下壹穴，坐北向南，前后各四丈，左右各三丈。其祖墳址在**八寮庄**面前**橫路**面祖墳壹穴，坐北向南，前后各四丈，左右各三丈。計共祖墳四穴。即日股夥面踏，交付蔡華亮殯葬祖墳，任從扞穴，修整築墳，永作佳城。眾股人等日後不能騎龍斬穴，亦不敢佔穴生端滋事。此係二比歡悅，各無反悔。恐口無憑，筆乃有據，立抽出祖墳山批字壹紙，付為永照。

代筆人 陳五雲 (花押)  
 在場人 陳阿生 (印)  
 股夥知見人 梁阿傳 (花押)  
 徐連昌 (印)  
 張秀欽 (印)  
 光緒十三年三月 日立抽出山批字公號**金廣成**四股首 周元寶 (花押)

原文參照: 國立台中圖書館整理, 《臺灣總督府檔案抄錄契約文書·永久保存公文類彙》 (國家文化資料庫, <http://nrc.ch.cca.gov.tw/ccahome/index.jsp>), 編號 ta\_01516\_000021-0001

錯誤回報 (metadata) ↑

錯誤回報 (全文) ↓

原文參照, 便於查找原始資料來源

圖10 「原文參照」與「錯誤回報」工具

## 肆、結語

本文介紹了臺大數位典藏研究發展中心與臺大資工數位典藏與自動推論實驗室所開發的「台灣歷史數位圖書館」(THDL), 尤其是其中古契書的部分。目前內容方面已集結了兩萬三千多件的全文, 內容亦會陸續增加。我們希望, 透過大量契書的集結, 能反映出更為完整的台灣基層社會圖像, 引領史家們進行更為精彩的研究。

本系統仍有不足的地方, 譬如我們尚未建立人名與地名權威檔, 亦尚未將現有資料與 GIS (地理資訊系統) 結合。這些工作若能實現, 會使 THDL 的使用更加方便。

透過 THDL 的開放使用, 我們希望邀請對於古契書有興趣的朋友共襄盛舉: 來使用、來給意見、來建議需進行的方向、來提供資料.....等, 任何方式的參與都可以促進 THDL 的成熟、演進。

建置 THDL 的目的是要打造一個研究臺灣歷史的「數位研究環境」。對一個研究環境而言, 內容固然重要, 但同樣重要的是提供使用者有用且好用的研究工具。換言之, 一個好的數位圖書館並不是一堆資料的集合而已, 而要提供工具協助研究者探討資料間的關聯, 以便在大量的資料中發現議題並能深入分析。

THDL 是第一步的實驗，我們希望透過研究者的使用與回饋，讓這個臺灣史的數位研究環境更加的成熟。

THDL 的目標是成爲集合臺灣史第一手史料的全文資料庫，因此除了古契書外，尚收錄了另一文獻集：明清臺灣行政檔案。古契約文書記錄的是基層人民的角度，明清臺灣行政檔案則是從中央政府的視角來看臺灣這個邊陲島嶼。我們希望藉由集結這兩種不同面向的文獻，反映出更爲完整的臺灣歷史。

## 致謝

在此，我們要特別感謝文建會、國立臺中圖書館、臺灣大學圖書館、與臺灣大學人類學系等數位化契書資料的製作單位，讓我們得以建置此一具有豐富收藏的古契書全文資料庫。在建置資料的過程中，非常感謝臺大圖書館特藏組邱婉容組長、胡乃文小姐、與陳靖玟小姐，一直以來持續的蒐集、建置台灣史相關資料。

我們也非常感謝成功大學吳密察教授、臺灣大學李文良教授、以及兩位教授的學生們，在我們建置全文資料庫的過程中，屢屢提供協助與指導，並讓我們能長期蒐集歷史學家的使用需求。

在建置「台灣歷史數位圖書館」的過程中，很感謝臺大數位典藏研發中心的領導與幕後協助。多年來，辛苦了臺大資工數位典藏與自動推論實驗室的研究生們，他們做了很多清理資料的工作，更進行了各種實驗，找出能有效利用這批數位化史料的方式；尤其是張尙斌、盧家慶、黃于鳴，他們的研究成果大大提升了「台灣歷史數位圖書館」的功能。

感謝曾在我們建置資料與系統的過程當中，曾給予我們幫助、提供意見的所有朋友。

最後，非常感謝涂豐恩先生，在本篇論文的寫作過程中給予的諸多協助。

## 參考文獻

### 一、專書

洪麗完，〈台灣中部平埔族群古文書研究與導讀：道卡斯族崩山八社與拍瀑拉族四社〉，豐原市：臺中縣立文化中心，2002年。

洪麗完編撰，〈臺灣古文書專輯〉，豐原市：臺中縣立文化中心，1996年。

臺灣史料集成編輯委員會編，〈臺灣總督府檔案抄錄契約文書·第壹輯〉，臺北市：行政院文化建設委員會，2005年。

臺灣史料集成編輯委員會編，〈臺灣總督府檔案抄錄契約文書·第貳輯〉，臺北市：行政院文化建設委員會，2006年。

### 二、會議論文

Chen, S.P., Hsiang, J., Tu, H.C., and Wu, M.C. "On Building a Full-Text Digital Library of Historical Documents". In Proceeding of the 10th International Conference on Asian Digital Libraries Held in Hanoi, Vietnam, December 10-13, 2007, edited by Dion Hoe Lian Goh, Tru Hoang Cao, Ingeborg SÃ,lvberg, Edie Rasmussen, 49-60. Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers. New York: Springer Berlin, 2007.

王世慶，〈介紹臺灣史料：檔案、古文書、族譜〉，《臺灣史與臺灣史料：臺灣史料評析講座紀錄（一）》（臺北市：自立晚報，1993年），頁71。

### 三、期刊論文

李文良，〈土地行政與契約文書—臺灣總督府檔案抄存契約文書解題〉《臺灣史研究》11(2)，（臺北：中央研究院臺灣史研究所），頁221-240。

邱婉容，〈明清臺灣行政檔案全文建檔及標注計畫〉，《中華民國圖書館年鑑》（臺北市：國家圖書館，2006年），頁128-129。

### 四、碩博士論文

杜協昌，〈互動式網際網路檢索：模型、度量及實驗〉，臺北：國立臺灣大學資訊工程學研究所博士論文，1998年。

張尙斌，〈詞夾子演算法在專有名詞辨識上的應用—以歷史文件為例〉，臺北：國立臺灣大學資訊工程學研究所碩士論文，2006年。

盧家慶，〈台灣古契書自動分類與依分類定義契書角色〉，臺北：國立臺灣大學資訊工程學研究所碩士論文，2008年。

## 五、計畫研究報告

李文良，《國家文化資料庫：臺灣總督府檔案收錄之古契約文書調查蒐集計畫研究報告書》（臺中：國立臺中圖書館，2003年）。

李文良，《國家文化資料庫：臺灣關係已刊印古契約文書調查收集計畫研究報告書》（臺北：行政院文化建設委員會，2003年）。

## 六、網站

台中縣立文化中心，古文書藏品展示，

<http://ca.tchcc.gov.tw/histbook/>。

台灣中部平埔族群古文書數位典藏，

<http://ca.tchcc.gov.tw/pingpu/>。

台灣歷史數位圖書館（Taiwan History Digital Library，THDL），

<http://thdl.ntu.edu.tw>。

故宮博物院，清代檔案人名權威資料查詢：

<http://www.npm.gov.tw/zh-tw/learning/library/archives.htm>。

國立臺灣大學圖書館，岸裡大社文書，

[http://www.lib.ntu.edu.tw/CG/resources/Taiwan/taiwan\\_al\\_a.htm](http://www.lib.ntu.edu.tw/CG/resources/Taiwan/taiwan_al_a.htm)。

國家文化資料庫之古文書

<http://nrch.cca.gov.tw/ccahome/oldbook>。

國家圖書館，臺灣歷史人物小傳：

[http://memory.ncl.edu.tw/tm\\_cgi/hypage.cgi?HYPAGE=toolbox\\_home.hpg](http://memory.ncl.edu.tw/tm_cgi/hypage.cgi?HYPAGE=toolbox_home.hpg)。

臺灣大學數位典藏資源中心（NTU Digital Archives Resource Center，DARC），

<http://www.darc.ntu.edu.tw>。

臺灣地區古文書資訊網，

<http://od.ntl.gov.tw/tl>。