

國立臺灣大學電機資訊學院資訊網路與多媒體研究所

碩士論文

Graduate Institute of Networking and Multimedia
College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

日治法院檔案系統及其後分類呈現

On the Design and Implementation of a Faceted
Retrieval System for the Taiwan Judicial Court Archives

from 1895 to 1945



Hsiao I-Ling

指導教授：項潔 博士

Advisor: Jieh Hsiang , Ph.D.

中華民國 97 年 6 月

June, 2008

誌謝

研究所放榜的那個早晨，我仍依稀記得外頭下著綿綿細雨，轉眼間一千多個日子飛逝，在整個研究所的生涯中受到了老師、學長姐、同學、學弟妹許許多多的照顧，無論是在課業上、生活上，感謝你們的支持這本論文才得以順利完成。

首先感謝法律學院王泰升教授，在『日治法院檔案』建置過程中的指導，並引領著我們向前邁進。感謝法律系阿亮學長期的照顧，並為我解答資料的疑惑。特別感謝台大圖書館特藏組的嘉文學姐與邱主任，在整個日治法院檔案的製作過程中，你們是背後最大的推手，從檔案的檢查、整理、與法學院的溝通等等大小事情一手包辦，讓我能心無旁騖的專注於系統的建置，沒有你們就不會有今天的成果。

感謝項老師的教誨，在老師的帶領下我體會到作學問的真諦，學會如何認清問題的本質，並教導我們正確的觀念，再感謝老師於整個論文過程中給予我的指導與支持，讓我能順利的完成學業。感謝詩沛學姐與浩洋學長在我求學期間不斷的鼓勵與指點，無論我做的如何，學姐與學長從不責備並提點我該如何做的更好，給我許多信心與勇氣。感謝杜老大認真開導，無論對於我的論文、課業或系統都提供我許多建議，帶領我們正確的思考並改正我們走偏的方向，令我得到莫大的幫助。感謝筱盈學姐的細心與孜孜不倦的提醒，幫助我找到盲點讓我能順利的通過考驗。

感謝元哥、代昕學長與已畢業的萬慶、開昱、尙斌、宗和、榮佐學長們，你們的熱情相待，讓我很快的融入了新的環境，迎接新的挑戰。感謝同屆的國延、鈞韜、永倫、譽仁，研究所的路上有你們的陪伴，我才得以不會感到孤單。感謝家慶、承恩、思靜不嫌棄我這散漫的學長，受到學弟妹們的照顧令我十分慚愧並萬分幸運與感恩。感謝于鳴、小玉、嘉翔、雋凡對實驗室的付出，未來實驗室就拜託你們了。最後感謝父母無條件的支持讓我毫無後顧之憂，我的朋友們時常陪伴情義相挺，再多的言語都不足以形容對大家的感謝，僅將這篇論文獻給大家。

中文摘要

『日治法院檔案』為台灣高等法院與王泰升教授於西元 2000 年所發現整理並數位化之台灣日治時期司法相關文件的總集，自台北、新竹、台中、嘉義四個地方法院以及司訓所藏台中法院檔案，共 5640 冊。其內容包含民事與刑事案件判決檔案、公正證書、各類行政文書、工商名冊登記簿等法律相關文件，尤其以記錄詳細司法案件審理內容之民事、刑事判決原本與公正證書原本三類檔案所佔數量最多，共 3049 冊，因此我們除了為每冊檔案建制詮釋資料外，亦針對民事、刑事判決原本與公正證書原本三類檔案，為其書冊內各案件建制詮釋資料。

法院檔案記錄深刻影響台灣人民生活的法院各種作為，記載了一段台灣人民與日本人民的共同歷史經驗，是了解台灣在日治五十年間，經濟、社會、文化等各方面朝向 近代化演進的第一手史料；而其描述台灣殖民法制之運作，在日本近代法史的研究中亦具有非常重要的地位。因此對於不同領域的研究人員都可從『日治法院檔案中』找到所關心的材料。

為了使得典藏檢索系統能發揮檔案的價值，我們使用『屬性標籤』的資料結構整合詮釋資料，並引入了多維度的後分類導覽方式；更在此基礎下，進一步為系統加入檢 索詞組控制介面，發展出擁有檢索詞組控制與後分類架構的整合式檢索介面系統之模型，以解決一般檢索系統之缺點，達成連貫檢索流程的目的。

利用此模型，依據檢索前後兩階段之目的，1.檢索詞組建議與組合、2.檢索詞組調整與篩選，實做『日治法院檔案』之二階段檢索詞組控制檢索系統，希望藉由系統的協助下，使用者能夠迅速掌握檢索資源，並擁有可彈性調整檢索策略的檢索操作模式，以達成完善的檢索目標。

英文摘要

In 2000, the Taiwan High Court and Professor Tay-Sheng Wang of the Department of Law of the National Taiwan University re-discovered the archive of judicial court records of the judicial courts of Taipei, Taichung, Jiayi and Hsinchu during the Japanese colonial occupation of Taiwan from 1895 to 1945. Digitization of the findings was carried out under the direction of Professor Wang with help from the National Taiwan University Library. After several years of work, the digitization effort is now near its completion. This thesis describes the design and implementation of the *TCCRA, the Taiwan Colonial Court Record Archives*.

The court cases, 5640 volumes in total, gave a vivid account of the economic, cultural, and societal evolution of Taiwan during the Japanese colonization. It is not only invaluable for anyone interested in the development of Taiwan, but also for researchers of Japanese colonial laws. The digitization effort includes producing all the images (through digital camera) and metadata of each court case. Because of the sheer volume of the data, it is important to design a system that allows the users not only find what they want but also helps them discover the meaning and conduct further exploration. We start by utilizing an "Attributive tag" data format to integrate the metadata and to provide the backbone of *faceted browsing* of query results. A sophisticated yet easy to use query interface is then designed to guide the user to refine queries and to classify query results in different ways.

Our faceted retrieval system has two main additional features: query term suggestion and combination, and query adjustment and document selection. Our features enable the user to analyze query results as a collection and refine them easily.

目錄

誌謝.....	i
中文摘要	ii
英文摘要	iii
目錄.....	iv
第一章 緒論	1
1.1 研究背景.....	1
1.2 研究動機.....	1
1.2.1 『日治法院檔案』之重要性.....	1
1.2.2 一般典藏檢索系統之缺點.....	2
1.3 研究目的.....	4
1.4 研究對象.....	6
1.4.1 『日治法院檔案』資料來源.....	6
1.4.2 『日治法院檔案』內容與檔案架構.....	7
1.4.3 『日治法院檔案』之案件類型資料.....	8
1.4.4 『日治法院檔案』之數位化.....	9
1.5 論文架構.....	12
1.6 範圍限制.....	12
第二章 相關技術討論	14
2.1 資訊索引型式.....	14
2.1.1 多維度導覽之後分類資訊檢索系統.....	14
2.1.2 標籤樹資訊呈現架構.....	15
2.2 系統引用概念.....	17
2.2.1 XML 標籤架構.....	17
2.2.2 協助檢索.....	18
2.2.3 影像圖層概念.....	20
第三章 日治法院檔案檢索系統之資料架構、系統模型、與特徵介紹	21
3.1 屬性標籤的資料架構.....	21
3.1.1 Meta-Data Format & Attribute Tag Format	21
3.2 檢索介面三元素.....	23
3.2.1 檢索詞組控制之檢索系統.....	24
3.2.2 兩階段檢索詞組控制系統.....	28
第四章 日治法院檔案系統實做	33
4.1 系統概述.....	33
4.1.1 系統模型.....	33
4.1.2 系統環境.....	34
4.2 系統資料準備 – 屬性標籤樣式.....	35

4.3 系統實做.....	38
4.3.1 Agent - 檢索程序資料處理流程.....	38
4.3.2 UI Stage 1 – 檢索建議與組合介面	48
4.3.3 UI Stage 2 – 檢索檢視與篩選介面	54
第五章 系統簡介、實例分析與討論	61
5.1 『辜顯榮』研究範例.....	61
第六章 結論與未來工作	65
6.1 結論.....	65
6.2 未來工作.....	66
參考文獻	67



第一章 緒論

1.1 研究背景

西元 2000 年 7 月間，經由台灣高等法院以及其所屬各地方法院，與台灣大學法律學院王泰升教授等人員努力搜尋下，發現新竹、台中、嘉義等三個地方法院尚保存台灣在日治時期法院相關各卷宗，再加上原本已知之台灣高等法院檔案，將這四法院總合稱為『日治法院檔案』。

在以『先使學術界能利用此檔案』為努力的目標下，自 2002 年起以王泰升教授以申請國科會研究計畫之方式整理這批『日治法院檔案』，而後台灣大學圖書館亦參與整理工作；希望透過原件之翻拍、複製典藏以及目錄的編制，提供學術界得以一窺日治時期台灣司法體制運作的概貌。

1.2 研究動機

1.2.1 『日治法院檔案』之重要性

『日治法院檔案』為集結台灣日治時期司法相關文書之大成，其內容記錄台灣本地人民與近代的法律的第一次接觸，法院檔案文書內深刻的記錄著當時法院如何地影響當時台灣人的生活。這些日治法院司法文書的內容，為了解台灣在日本統治五十年間，於經濟、社會、文化等等各方面近代化過程變遷的第一手史料，甚至對日本史的研究，亦不可或缺。

在王泰升教授於『台灣總督府法院文書目錄編纂』內所述[王 02]，『日治法院檔案』在學術上能提供各方面研究領域資源，除了法律史，還包括社會、經濟史等領域。在法律的規範上，可探究法院如何透過判決，從台灣在清治時期民間習慣，轉化為日治時期的近代國家法規範。從法院使用者的身份、性別、

有無訴訟代理人、紛爭類型等，可觀察當時一般人民與法院之間的關係與當時社會概況。在訴訟案件或執行程序的卷宗中，可發現各種社會生活事實，例如一般人使用哪一些傢具，價格如何。而眾多公證書內所附的契約內容，更足以全面地了解當時人民的財產交易模式。各種登記簿內有關土地和產業事項的記載，尤其是經濟史研究不可或缺的素材。乃至行政卷內有關人事或建築等資料，也可用來做人物、建築史的參考。對於日治時期各方面的研究者，都可由『日治法院檔案』的史料中，尋得自身領域中所關心的資料。因此，將『日治法院檔案』完整並有效的呈現給所有的使用者即為我們最重要的任務。

1.2.2 一般典藏檢索系統之缺點

在『數位典藏國家型科技計畫』下，許多的珍貴資源紛紛數位化並建立典藏檢索系統，以便於保存資源並進而推廣使用，『日治法院檔案』亦屬於其中之一。實體資源在完善的數位化的過程中，往往須匯集大量的金錢與人力，也因此更顯得數位資料的可貴。然而因為歷史文件的特性，使得典藏資料在數位化流程中必須仰賴大量的人力支援，但人本身就是變動的因子，在這樣的先天條件下，常常使得最終完成的數位資料存在期望上的落差。當我們在進行數位典藏資料系統製作的時候，應仔細思考如何降低因資料的不完善所產生的使用障礙，並更進一步對資料作更深入並有效的運用。

我們觀察到目前大多數的數位典藏檢索系統中，普遍存在一些缺點：

第一、在非複合資源的典藏資料庫下，資料庫的內的文件通常具有內容特徵的性質，舉例來說，司法的資料庫通常會有人名、案件名稱等詞彙；生物相關資料庫通常會有生物學名、色彩、形狀相關描述。在資源推廣的意義下，當非專門使用者來使用資料庫時，系統應盡可能提供資料庫內相關的檢索資源，在檢索的期間予以協助，而不是僅僅告知使用者，查無您所要的資訊。

第二、大部分的典藏檢索系統通常都以直線性的檢索流程來製作，缺乏彈性的檢索調整機制，因此當使用者在檢索的歷程中，無法中途根據檢索結果改變或替換檢索詞彙，以至於當使用者希望調整檢索策略的時候，整個檢索歷程就必須重頭來過。在資料數量龐大但可運用的檢索資源卻相對短缺時，無法提供足夠的文件辨識資訊的檔案資料，一次又一次的重複檢索，將造成使用者檢索上的困擾；『日治法院檔案』就符合此特性。

第三、目前多數檢索系統對於檢索結果的處理方式，通常是利用檢索欄位排序控制結果的呈現，但是除了日期、數量等數字概念的資料外，在中文為基礎的數位資料下，文字的排序往往無法突顯檢索結果的特徵；除此之外，當符合檢索條件的文件資料高達很大的數量，欄位排序的方式通常無法有效協助使用者進行資料的篩選。

第四、雖然典藏資源數位化的工作通常都有訂定的程序與規格，然而在許多因素影響下最後所得到的數位化資料仍無法避免出現內容的差異，在本論文中所使用的資料 - 『日治法院檔案』就是這樣的例子。如下《表 1-1》所顯示，

資料類型 欄位名稱	民事判決 - 台中法院	民事判決 - 台北,新竹,嘉義法院	公證書 - ALL	刑事判決 - ALL
出處	◎	◎	◎	◎
案件編號	◎	◎	◎	◎
書類名稱		◎		◎
案件事由	◎	◎	◎	◎
關係人			◎	
原告	◎	◎		
被告	◎	◎		◎
檢察官				◎
原告代理人		◎		
被告代理人		◎		◎
法官		◎		◎

書記官		◎		◎
判決日期	◎	◎	◎	◎
是否上訴		◎		◎
所在法院		◎		◎
備註	◎	◎	◎	◎

《表 1-1》『日治法院檔案』 Metadata 欄位對照表

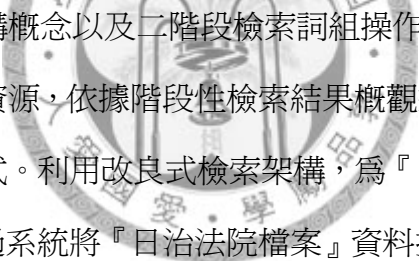
由於不同類型的司法文件擁有不同的特性與內容，因此在詮釋資料 (Metadata) 的建構上就出現了欄位名稱無法統一的狀況。除此之外，因各法院資料數位化的時間與人員單位的不同，導致檔案內 Metadata 所規定註錄的欄位也不一致，造成『台中地方法院』相較於其他法院的欄位更為精簡。這些不一致的資料狀況將造成系統實做的障礙，假使沒有仔細思考如何統整各項資料，完善解決詮釋資料差異的問題，最後所製作的系統將不利於檢索使用，如此不僅浪費了數位化的成本，亦將遮蔽典藏資料價值的光輝。

1.3 研究目的

為了建立『日治法院檔案』之檢索系統，並解決上述所觀察到的四點檢索系統製作上的問題，資料庫檢索系統除了提供使用者符合檢索條件的資料文件外，對於檢索結果文件的特徵應加以組織與呈現，提供使用者所有符合檢索條件文件之重點或概觀的描述，以減輕使用者在檢索篩選上的工作。因此在本篇論文中，我們將引入『屬性標籤』的資料架構，解決在『日治法院檔案』中不同類型案件資料詮釋差異的問題以及增加資料操作彈性。除此之外，更援引過去本實驗室在檢索資訊系統中所發展出的『檢索後分類』資訊整合模式，以及階層展示的『屬性標籤樹』架構等兩項資訊索引方案於『日治法院檔案』系統中，並在此基礎下，更進一步的將資訊索引轉化為檢索的新動力。

一般典藏數位化中資料通常以 Metadata 格式為註錄基礎，然而在檢索系

統中，以描述與保存為目的的 Metadata 欄位架構缺乏系統操作的彈性，因此我們利用屬性標籤 (Tag) 的概念，將 Metadata 中賦有檢索意義的欄位在系統內轉化為屬性標籤架構，並將整個檢索流程分為二階段：第一階段為檢索前期的詞彙探測與關聯標籤擴展，將與檢索目標存在相依關聯的詞彙提供給使用者，做檢索條件的組合、篩選等檢索策略運用，進而形成一到數個預備檢索詞組。第二階段為檢索期間對各檢索詞組內的標籤元素進行增、刪、遮蔽的操作，透過即時檢索結果後分類與階層展示的文件特徵呈現，進而調整檢索詞組的組成結構，以及篩選檢索結果的文件，並予以命名後形成最後的文件集；而後可更進一步的將文件集合作為進階服務的依據，針對各檢索詞組或個別文件集間進行詞頻統計與分析、標籤與標籤的間接關聯等協助，挖掘潛藏在深層的相關資訊，提供與使用者做檢索的分析參考。



透過屬性標籤架構概念以及二階段檢索詞組操作的設計，以期望協助使用者能夠迅速掌握檢索資源，依據階段性檢索結果概觀描述，形成可彈性調整檢索策略的檢索循環模式。利用改良式檢索架構，為『日治法院檔案』建立完善的檢索系統，希望透過系統將『日治法院檔案』資料推廣，以提供日治時期臺灣之社會經濟等各領域的相關研究資源。

1.4 研究對象

1.4.1 『日治法院檔案』資料來源

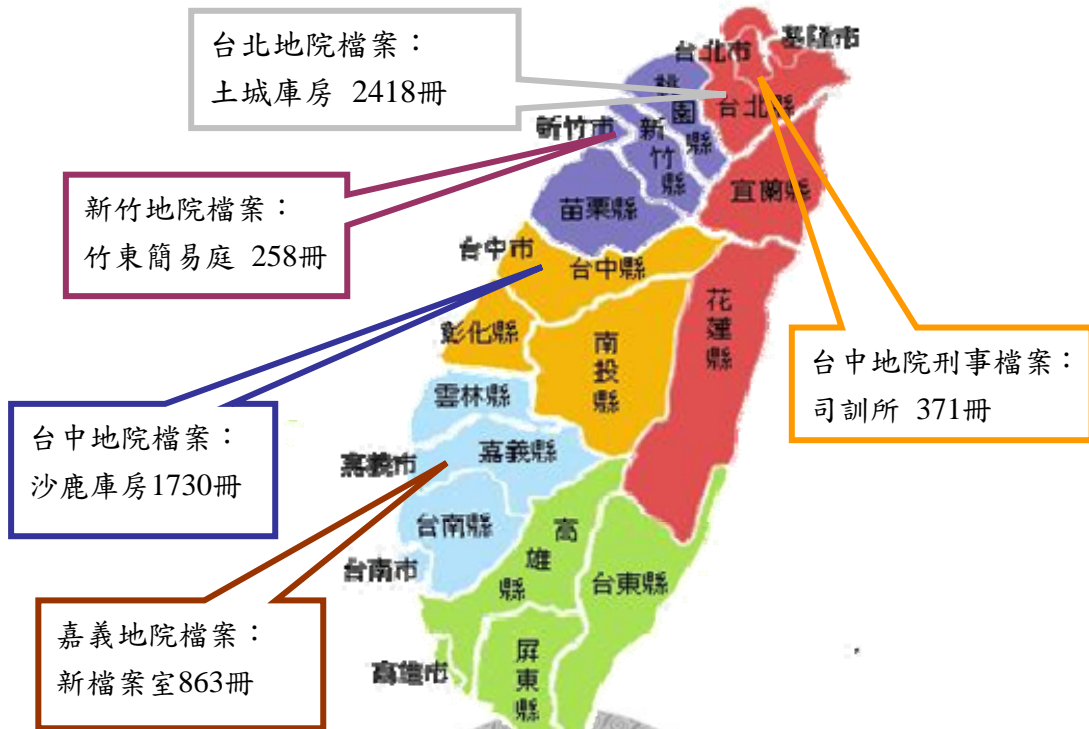
『日治法院檔案』的內容橫跨四個法院，六個資料群，為日治時期臺灣司法文件的大集合，經過統計目前一共包含書目 5916 冊，如下《表 1-2》所示，

檔案典藏單位	冊數	已匯入資料庫	資料數位化進度	
			影像拍攝完成	Meta 建制完成
台北地方法院	2418 冊	1516 冊	2418 冊	1516 冊
新竹地方法院	258 冊	258 冊	258 冊	258 冊
台中地方法院	1635 冊	290 冊	1635 冊	290 冊
台中地方法院刑事檔案	95 冊	0	95 冊	0
司訓所藏台中地方法院 刑事檔案	371 冊	0	371 冊	0
嘉義地方法院	863 冊	863 冊	863 冊	863 冊
總計	5640 冊	2927 冊	5640 冊	2927 冊

《表 1-2》『日治法院檔案』內容統計

新竹、嘉義兩個地方法院的資料已全數建置完成，台北地方法院尚有 900 餘冊之 Metadata 仍在建制當中；而台中地方法院由於是整個計畫中第一個製作的法院，在不熟悉的狀況下因此資料的品質較差，詮釋資料的建置也比後來所製作的法院檔案少了部份欄位，例外處理也出現了不一致的狀況，因此台中地方法院的資料雖已建置完畢，仍需要經過詳細的檢查才能納入資料庫當中。

『日治法院檔案』的原始資料分別存放在五個地方法院之倉庫，如下頁《圖 1-1》所展示，在檔案資料的數位化處理上存在一些限制，1.資料不可搬離倉庫、2.書冊不可拆裝、3.檔案年代久遠紙質脆弱、4.檔案數量眾多。因此最後決定以數位相機拍攝的方式作為『日治法院檔案』的數位化方案。將拍攝完成的影像最後再帶回台北作進一步的查驗與詮釋資料編目。



《圖 1-1》『日治法院檔案』原始資料分佈圖

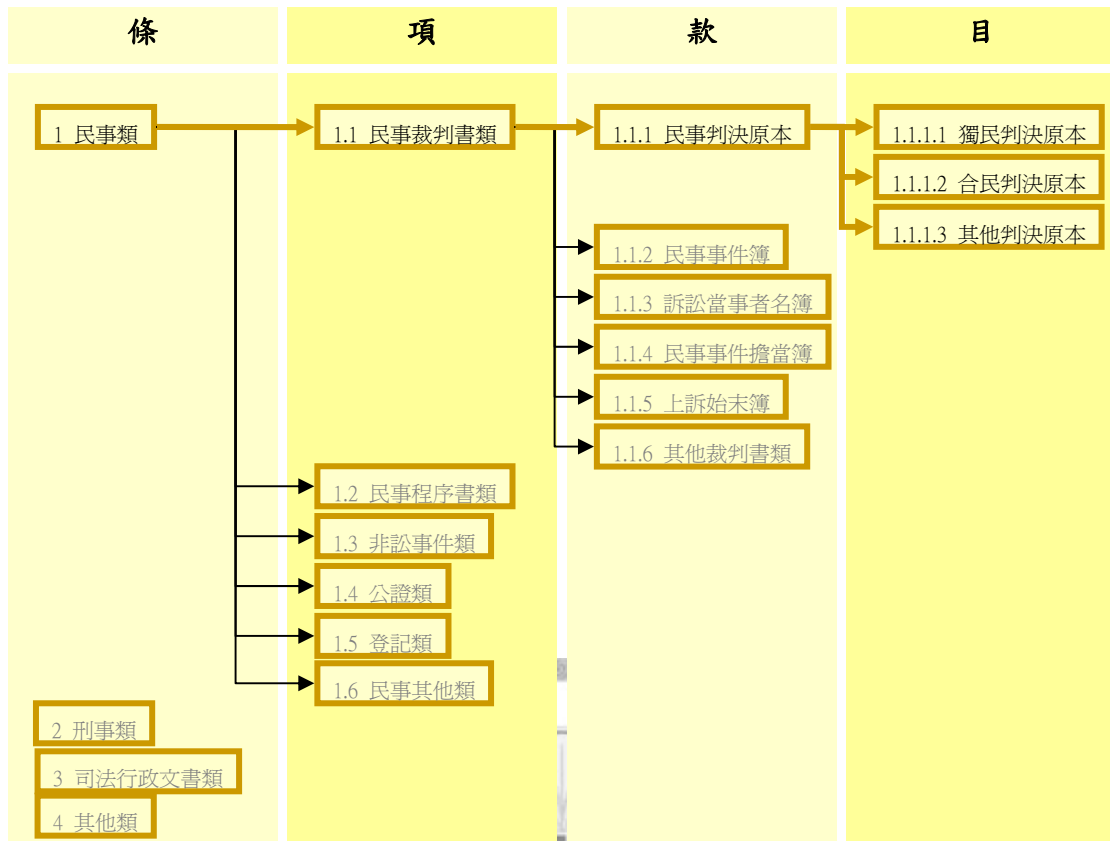
1.4.2 『日治法院檔案』內容與檔案架構

『日治法院檔案』的內容包括了所有當時經由法院處理的相關文件，除了民事與刑事的案件資料外，也包含了許多行政文書、表格名冊、登記簿等等法律相關文件，如右《圖 1-2》所示，『以嘉義地方法院』為例，除了『民事裁判』的案件資料外，還包含了『民事程序』、『登記類』、『公證書』等民事資料，在司法行政文書下亦包括『文書記錄管理』、『法院人事』、『法院庶務』等等文書資料，內容包羅萬象十分豐富。因此在檔案建置的過程中，我們將『日治法院檔案』所有資料依據資料的性質概分為『條』、『項』、『款』、『目』四個階

嘉義地院	
1	民事類
+ 1	民事裁判書類
+ 2	民事程序書類
+ 3	非訟事件類
+ 4	公證書
+ 5	登記類
+ 6	其他民事類
3	司法行政文書
* 2	文書記錄管理
* 4	法院藏書
* 5	法院人事
* 6	法院令達
* 7	法院往復文書
* 8	法院庶務其他文書
4	其他類
* 3	中華民國文書

《圖 1-2》嘉義地方法院檔案類型

層，如下頁《表 1-3》所示，『此為台中地方法院』

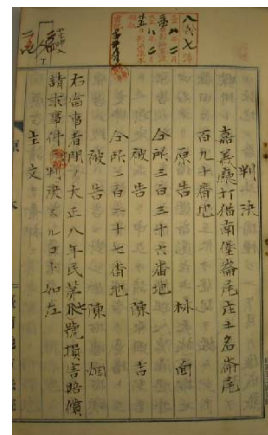


《表 1-3》『日治法院檔案』台中地方法院檔案架構圖（暫）

之檔案架構，系統也依據此架構建立瀏覽機制，讓使用者能夠快速的了解整個『日治法院檔案』的概貌。然而並非所有的檔案都擁有四項階層，其內容架構在各地方法院也不盡相同，因此《表 1-3》為目前暫定的架構，未來將根據所有法院建置完成後的資料，進行整體的修正。

1.4.3 『日治法院檔案』之案件類型資料

『案件』類型資料為各地方法院紀錄其法院所經手的司法案件，其內容記載案件的相關人員、案件原委、審理過程以及判決結果等資料，每年依據案件編號排序，集結成冊以資保存。案件類型資料共有三種：『民



『刑事判決原本』、『公正證書原本』與『刑事判決原本』，案件類型資料在『日治法院檔案』各類型的法院資料中所佔的比例最多，如下《表 1-4》所示，除了『嘉義地方法院』外，其餘地方法院所收藏案件資料

法院	台北地方法院	新竹地方法院	臺中地方法院	嘉義地方法院	司訓所藏台中地方法院刑事檔案
民事判決原本	377 冊	99 冊	419 冊	181 冊	-
刑事判決原本	-	-	95 冊	-	371 冊
公正證書原本	1139 冊	18 冊	272 冊	78 冊	-
小計	1516 冊	117 冊	786 冊	259 冊	371 冊
法院檔案總數	2418 冊	258 冊	1730 冊	863 冊	371 冊
比例 (案件冊數/總冊數)	62.7%	45.3%	45.4%	30%	100%

《表 1-4》『日治法院檔案』案件類型資料書冊統計

都佔有相當高的比例，因此在 Metadata 的建制上，除了為每冊『日治法院檔案』書目建制書目 Metadata 外，我們將更進一步為案件類型資料建制案件 Metadata。

1.4.4 『日治法院檔案』之數位化

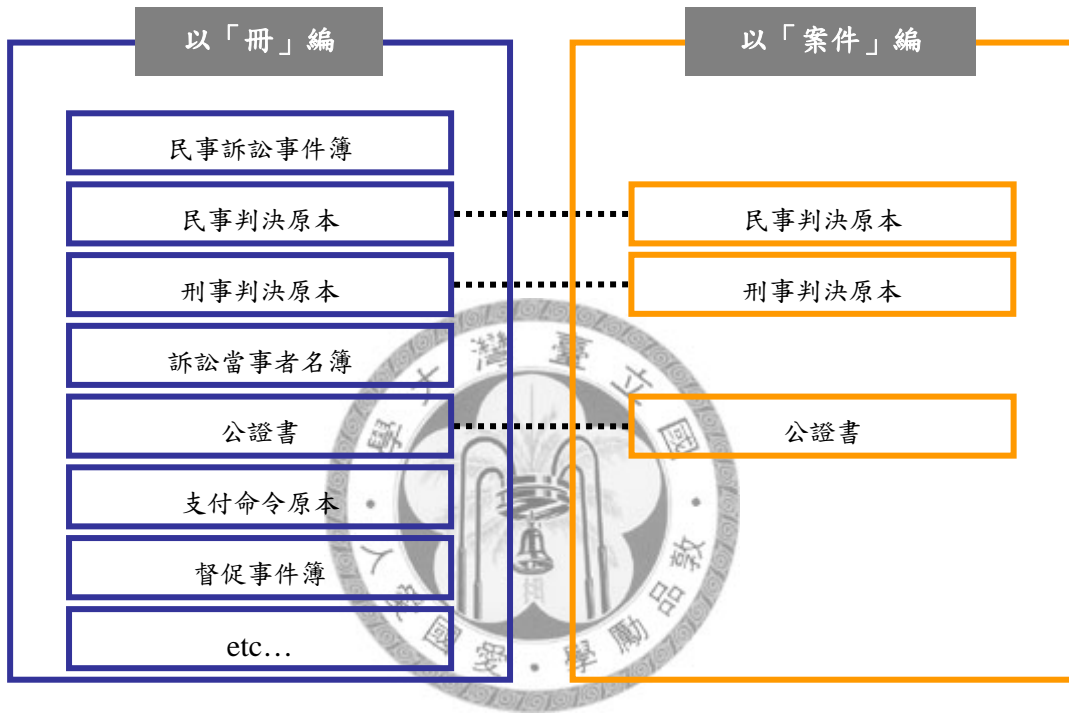
『日治法院檔案』數位化的重點在於所有法院檔案文件影像翻拍，並輔以 Metadata 建置，目前在資料庫內已收納的 2927 冊的檔案資料，一共擁有 1,457,447 個影像檔案，如下《表 1-5》所示，每張影像大小為 300Kbytes 不等，

檔案典藏單位	已匯入資料庫	影像檔案	容量
台北地方法院	1516 冊	951,943	227.0GB
新竹地方法院	258 冊	84,058	23.0GB
台中地方法院	290 冊	115,025	43.2GB
台中地方法院刑事檔案	0	0	0
司訓所藏台中地方法院刑事檔案	0	0	0
嘉義地方法院	863 冊	306,421	85.2GB
總計	2927 冊	1,457,447	380GB

《表 1-5》『日治法院檔案』影像統計

目前系統內共 380GBytes，未來將持續的增加。

在 metadata 的建置方面，『日治法院檔案』將 metadata 分為兩個種類，一、是以『冊』為單位建置 metadata，二、再由冊中，將『民事判決原本』、『刑事判決原本』與『公證書』三種案件類型的檔案更進一步為其書冊的內容，為每一個案件製作『案件』的 Metadata。



《表 1-6》『日治法院檔案』 Metadata 建置對照表

『冊』與『案件』的 Metadata 所註錄的內容如下《表 1-7》所示。『冊』的 Metadata 一共有四個欄位，『案件』 Metadata 比較詳細，為 15 個欄位，

『冊』 Metadata	1.分類層次(條項款目) 2.冊名 3.內容與年代 4.書籍冊次
『案件』 Metadata	1.冊名 2.事件號 3.判決標題 4.原告

	5.原告訴訟代理人 6.被告 7.被告訴訟代理人 8.事件名稱 9.判決日期 10.法院 11.法官姓名 12.書記官姓名 13.上訴及上訴人姓名 14.影像位置 15.備註 (以台北地方法院民事判決原本為例)
--	--

《表 1-7》『日治法院檔案』Metadata 欄位對照表

因此目前系統內的 Metadata 資料統計如下《表 1-8》所示，書冊 Metadata 共 2927 筆，案件 Metadata 共 157,300 筆，未來依據其他法院的進度資料將會隨之增加。

	書冊 Metadata	案件 Metadata		總計
		民事判決原本	公正證書原本	
台北地方法院	1,516 筆	48,355 筆	55,886 筆	104,241 筆
新竹地方法院	258 筆	12,022 筆	1,004 筆	13,026 筆
台中地方法院	290 筆	9,590 筆	8,950 筆	18,540 筆
嘉義地方法院	863 筆	18,520 筆	2,973 筆	21,493 筆
總計	2927 筆	88,487 筆	68,813 筆	157,300 筆

《表 1-8》『日治法院檔案』Metadata 資料統計

1.5 論文架構

本篇論文分為六個章節，第一章主要闡述本論文系統緣起的背景，以及經由當前數位典藏資訊檢索系統，觀察到多數系統對於數位資料運用過於單純，並在檢索系統的製作上存有可改進的地方；以及因應『日治法院檔案』的資料缺陷，進而提出利用屬性標籤架構的資料運作方式，配合資訊後分類處理與呈現，發展出檢索詞組控制的整合檢索介面。

第二章將說明論文中所引入的技術與概念，首先介紹本系統在建置上所利用的資訊索引基礎 - 『後分類』與『標籤樹』兩項資訊索引與呈現模式。再來介紹系統中引用的相關技術，由介紹標籤系統的與應用方式，進而提及系統在發想時引入的概念。

第三章為介紹系統建置上所應用概念、方法與模型。首先介紹為了系統運作而準備的屬性標籤資料架構；而後經由描述一般檢索流程的缺點，為了讓系統加入後分類架構與增進檢索流程彈性的目的，我們為『日治法院檔案』系統的設計了新的系統模型。最後我們運用新的模型提出了兩階段的檢索流程，並針對各階段進行介紹。

第四章為系統實做，針對檢索系統的模型，根據各檢索階段的架構元件進行仔細的描述，並依據在系統實做上所遇到的問題給予解決的方案。

第五章為系統應用實例分析與討論，以實例來展示檢索系統建置的成果以及分析討論是否達到本論文目的。

第六章為結論與未來工作，作為本篇論文的研究總結，並提出一些應改進以及可增進、深入發展之處，引領未來努力方向。

1.6 範圍限制

本論文索引用的資料為具有詮釋資料的典藏資源，根據詮釋資料的架構進行整合分析，在系統運作階段將具有檢索參考價值的詮釋資料轉換成屬性標籤的資料架構。因此對於沒有詮釋資料的數位資源，或是詮釋資料內建構的內容

不具有檢索價值，將不囊括在本論文所論述的範圍內。

本系統所援用的資料為單一來源，針對於複合資源資料庫的系統設計與考量將更形複雜，依據資料庫的規模以及內容來源的數量，在整合資源儲存架構規範的訂定與系統運算分散處理的設置都必須納入系統建置的考量，此部份亦不再論文的討論範圍之內；然而無論是單一資源或是複合資源的系統，本章第三節內所論述的檢索的問題依然會存在，本論文亦可作為系統建置的參考方向。



第二章 相關技術討論

2.1 資訊索引型式

爲了增進檢索的效率，並有效的彰顯資料的價值與意義，我們爲『日治法院檔案』引入了兩項互動式資訊檢索模式，1.多維度導覽之後分類資訊檢索系統 2.標籤樹資訊呈現架構，以下兩小節將分別概述之。

2.1.1 多維度導覽之後分類資訊檢索系統

(Post-Query Multi-Dimensional Navigation System) (Faceted Browsing)

多維度導覽之後分類資訊檢索系統爲杜協昌博士在 1999 年所提出的『互動式網際網路檢索』模型[Tu99]，在資訊檢索系統中，將 Metadata 中特定之欄位視爲描述資料的獨立維度 (dimensions)，如『年代』、『作者』等，使用者可利用維度進行檢索，例如『年代：昭和 15 年』，當系統將符合檢索條件之文件，依據各個維度進行分類與排序的功能，例如將搜尋的結果依『年代』、『出處』、『作者』等分別列出維度下的分類，提供給使用者進行下一步的檢索動作。

The screenshot shows the THDL (Taiwan Historical Digital Library) search results page. The search term is '林爽文'. The results are filtered by 'Year (TM)' to '1795' (Qianlong 47th year). The page displays a list of search results, including titles like '1. 哆囉囉大武壠派二社番租碑' and '2. 哆囉囉大武壠派二社番租碑'. The interface includes a search bar, a sidebar for faceted browsing, and a main content area with search results. A red arrow points to the sidebar, highlighting the 'Faceted Browsing' feature.

《圖 2-1》台灣歷史數位圖書館資訊檢索系統

如《圖 2-1》所展示，此系統為『台灣歷史數位圖書館資訊檢索系統』，在圖中我們以『明清檔案』資料庫作為系統範例。『明清檔案』為明清時期與台灣相關的奏摺、行政文書等文件蒐集而成的資料庫。當我們用『林爽文』作為查詢的關鍵字進行系統查詢後，在圖上紅色箭頭所指示的地方，系統將所有符合的文件，依據『年代』、『出處』、『作者』三項維度對檢索結果進行各別的分類，藉由這樣的機制，我們將可以很容易的利用後分類資訊，掌握檢索結果文件的概況，並更容易進行下一步的檢索。在例子中我們可以藉由文件的統計，發現與『林爽文』相關的文件在『乾隆』51-53 年間大幅度的增加，這樣的統計資料極具意義，配合資料的特性，使用者從後分類的機制中立刻就可以知道『林爽文』在這段時間於台灣發生緊密的關聯，這樣的結果是很難由文件排序中顯現出來。



因此運用多維度的後分類資訊索引系統，將更容易彰顯資料本身的意義，並利用簡單的屬性運算增進檢索效率，並提供使用者有效的檢索資源，達成檢索目標。

2.1.2 標籤樹資訊呈現架構

(Presentation of Query Result Using Tag Trees)

在後分類的資訊索引下，標籤樹透過樹狀的結構與屬性組織檢索結果的特徵，為一種屬性標籤的階層架構分類及整合呈現的方式[莊 06]。標籤樹組織文件中屬性明確的關鍵詞，如『人名』、『地名』。使用者透過標籤樹的資訊呈現架構，能夠簡單直覺的判別出檢索結果中的重點，並依據不同主軸的觀察，如以『人名』為主軸、以『時間』為主軸，了解重要特徵的關聯性。並透過可直接點選的方式，能夠方便的縮小文件範圍，修正檢索結果。

下圖《圖 2-2》為 THDL Prototype System V0.83，使用『古契書』文件集，

以『金廣成』為關鍵字檢索後，系統所顯示的標籤樹呈現介面。在本例子當中，由時間的主軸下，我們可以觀察到系統將詞彙標籤，依據在文件中出現的次數，利用字型的大小以及顏色排序與呈現。



《圖 2-2》標籤樹 - 階層式呈現架構

『金廣成』為墾號之名，是由一群人以此名稱向政府租地，再分租給其他人的組織。由標籤樹的階層式呈現介面我們可以觀察到，第一階層的時間主軸由『清光緒十三年』開始，表示與『金廣成』相關的契書文件從此年才開始再資料庫內出現，而由第二階層的地名資訊中我們也可以發現，『竹北二堡』與『十寮庄』兩個地名在『清光緒十三年』的 13 件文件中，就各都出現了 11 次，『清光緒十五年』的 15 件文件中，就各出現了 13 次與 14 次，顯然的可以推論出『金廣成』與這兩個地方有著高度的相關性，而從文件的全文中，也證實了這個推論。

在後分類的檢索系統下，運用階層式資訊整合的標籤樹呈現架構，將檢索結果作多主軸的方式呈現，以提供使用者在一次檢索下能夠得到更完整的文件的資訊，這將提供無全文檢索的系統另一種資訊索引的幫助。

2.2 系統引用概念

『日治法院檔案』系統的製作上，為了解決系統的問題，以及配合系統的設計方向，我們參考了當前資訊系統的特色，並引用資訊架構處理系統所面臨到的問題，以下將分別為『日治法院檔案』內所運用到的資訊系統概念作簡要的介紹。

2.2.1 XML 標籤架構

XML 原文為 Extensible Markup Language，又稱『可擴展標記語言』，簡稱 XML，意即電腦所能解讀的標記符號並可自由擴充[W3C]。XML 是從標準通用標記語言（SGML）中簡化修改而成，自 IBM 從 1960 年代就開始發展的 GML（Generalized Markup Language）標準化後的名稱。其中 GML 有兩點重要的概念，1.文件中能夠明確的將標示與內容區隔、2.所有文件的標注方式必須統一，這兩項特色後來也成為 XML 的特性。XML 被廣泛用來作為跨平台間資訊交換的型式，而根據 XML 的內容，透過不同的格式化描述(XSLT，CSS 等)將可以成為最後的資料呈現樣式(HTML，PDF 或者其他的文件格式)。

XML對資料的描述結構為樹狀的屬性標籤樣式，其撰寫的樣式必須遵循一定的規則，在此不對XML的文件格式加以詳述，僅以例子描述紀錄內容的標籤樣式。以下為一組XML的文件標記樣式，Tag『name』隸屬於Tag『root』，而『name』中含有屬性標示『attribute="value"』作為這筆資料的特性註記。運用XML可自訂架構的特徵，可以用來補足因資料整合而產生的資訊誤差，使得系統能夠同時保有建置的彈性以及原資料的完整性。

```
<root >
```

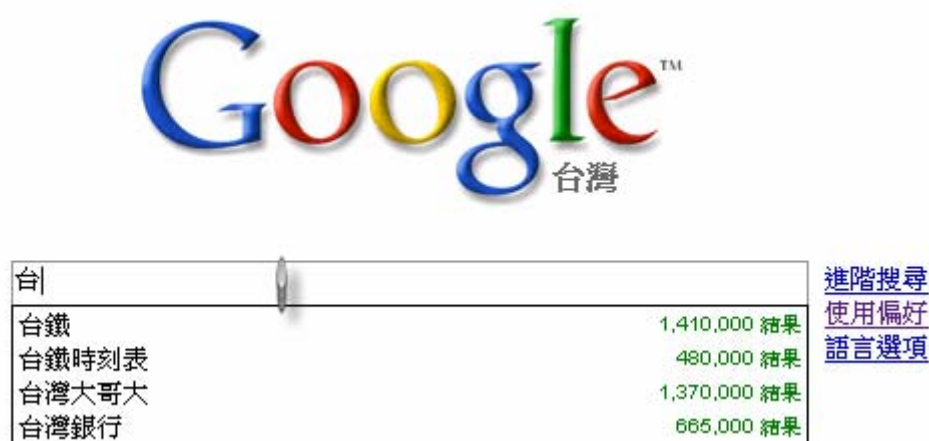
```
<name attribute="value">Content</name>
```

```
</root>
```

2.2.2 協助檢索

許多資訊檢索系統中，依據不同的目的，系統都會利用所擁有的資源製作協助檢索的工具，在操作檢索系統的過程中提供給使用者作為檢索的建議。以下提出兩種目前資訊檢索系統中，協助檢索的實做範例以供參考。

1.在網路搜尋引擎方面，除了在搜尋介面上列舉出近來常被檢索的熱門關鍵字，近年來 Google 與 Yahoo 分別都利用 ajax 的技術製作搜尋關鍵詞建議的功能，如下圖《圖 2-3》所示，當使用者在填寫檢索詞彙的過程中，利用目前



《圖 2-3》Google Suggest System

輸入的片段的資訊，運用演算法猜測使用者可能的檢索詞彙，並將可能性最高的資料提供給使用者作參考。除了提供相關的字彙以供使用者以選擇代替輸入，這樣的方式還可以幫助使用者，從建議的詞彙中檢驗自己所下的關鍵字是否正確。

2.在商業網站中，常利用一件商品關聯到其他可能相關的商品上，以增加商品的曝光度，下《圖 2-4》是在 amazon.com 的網站上經由搜尋系統所找到的一本書，當使用者點選之後系統除了提供此書的相關內容，並利用商品特價資訊或其他人相關的購買行為，如圖中紅色箭頭所指的地方，將其他關聯商品一併提供給使用者作為選擇的建議。

SEARCH INSIDE!™
REVISED PRINTING
COMPUTER ORGANIZATION AND DESIGN
THE HARDWARE/SOFTWARE INTERFACE
DAVID A. PATTERSON
JOHN L. HENNESSY

Computer Organization and Design: The Hardware/Software Interface. Third Edition, Revised (Paperback)
by [David A. Patterson](#) (Author), [John L. Hennessy](#) (Author)
★★★★☆ (29 customer reviews)

List Price: ~~\$64.95~~
Price: **\$58.46** & this item ships for **FREE with Super Saver Shipping**. [Details](#)
You Save: **\$6.49 (10%)**

In Stock.
Ships from and sold by **Amazon.com**. Gift-wrap available.

Want it delivered Thursday, June 19? Order it in the next 6 hours and 17 minutes, and choose **One-Day Shipping** at checkout. [See details](#)

[Share your own customer images](#)
[Search inside this book](#)

103 used & new available from **\$14.95**

Are You an Author or Publisher?
[Find out how to publish your own Kindle Book](#)

Also Available in: List Price: Our Price: Other Offers:
Paperback (3) \$64.95 \$64.95 84 used & new from \$11.00

Most Frequently Bought Together
4.8% of customers who bought this item bought it with:

Computer Architecture, Fourth Edition: A Quantitative Approach by John L. Hennessy
★★★★☆ (9) \$76.46
In Stock. Ships from and sold by **Amazon.com**.

Price For Both: \$134.92 [Add both to Cart](#)

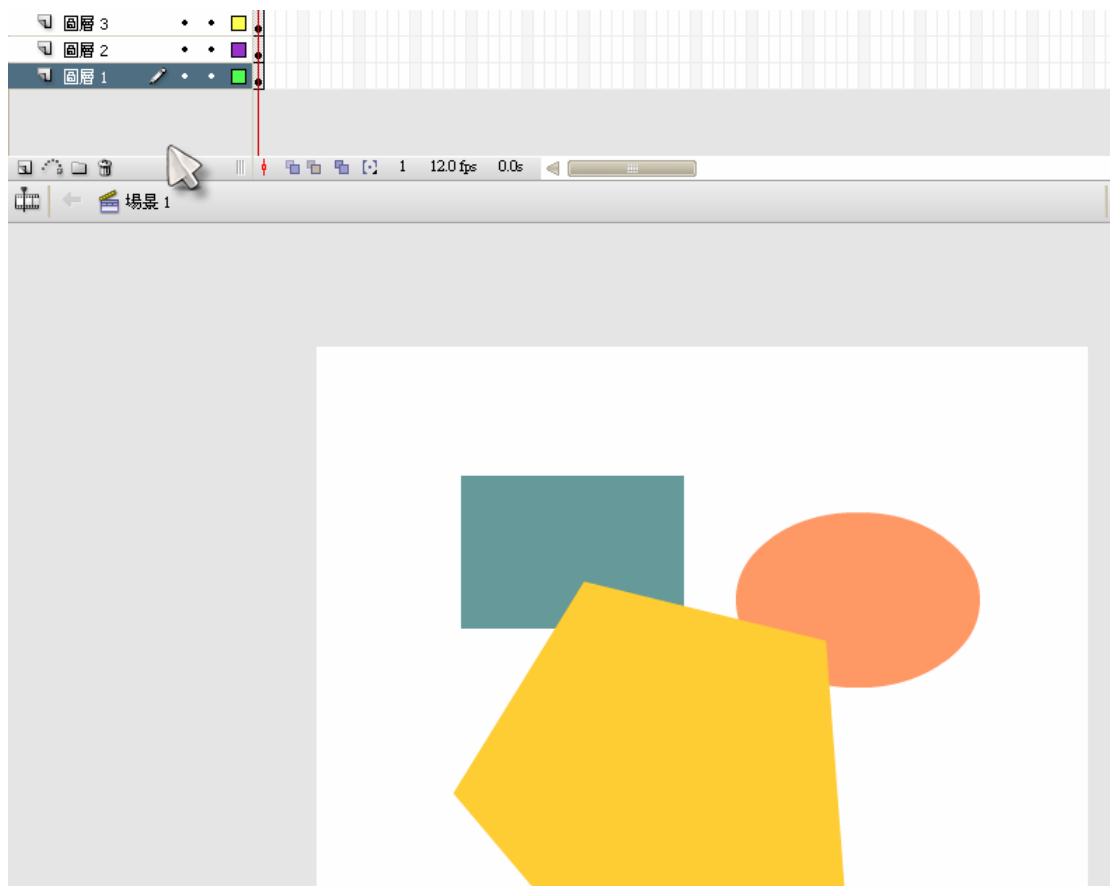
Customers Who Bought This Item Also Bought

《圖 2-4》amazon.com 購買系統

商業網站提供檢索建議是爲了增加自身的經營績效，而數位典藏系統雖然不以營利爲目的，但對於提供完善的資訊服務應是所有資訊檢索系統一致的目標。目前大多數的數位典藏系統中卻鮮少含有檢索建議的功能，典藏資源通常不屬於一般人的知識領域中，因此在系統的使用上除了利用瀏覽的機制，檢索的方式常常會導致使用者毫無頭緒的窘境，以至於檢索品質下降。然而在數位典藏系統中製作檢索建議也必須做好考量，以專家的方式建立輔助資料會耗費大量的資源並不符合效益，而利用資訊計算的方式自動產生的輔助資訊雖然比較經濟與快速，但仍存在準確率的問題，假使系統提供了錯誤的意見將會造成使用者對系統的信心下降，因此如何拿捏系統輔助尺度將是系統製作者的考驗。

2.2.3 影像圖層概念

在影像處理的技術中存在一種元素，稱之為『圖層』(Layer)，圖層的作用在於可以層疊於另外一張影像上，使得原來的影像經由層疊後含有更多的元素。許多的影像製作軟體都具有圖層操作的功能，其操作圖層系統具有三種特性，1.圖層具有可重疊性，可以在一張影像上層疊一到數個圖層。2.圖層有次序性，上方的圖層將會遮蔽下方的圖層。3.圖層具有可隱藏性，可以將某一層圖層暫時的隱藏，除去其圖層對於影像的影響。如下圖《圖 2-5》箭頭所指示的地方，即為影像處理軟體的圖層操作介面。



《圖 2-5》影像處理軟體 – 操作圖層

圖層層疊操作的特性，啟發我們對於檢索詞組的處理的作法，一張影像可由多個圖層所堆疊構成，而檢索結果也視為由數項檢索條件所交集而成，由此將引發我們之後對於檢索介面設計的概念。

第三章 日治法院檔案檢索系統之資料架構、系統模型、與特徵介紹

3.1 屬性標籤的資料架構

3.1.1 Meta-Data Format & Attribute Tag Format

在數位典藏各計畫中，通常採用 Metadata 的欄位架構來建置典藏資源的描述，以都柏林核心集(Dublin Code)為基礎[國家文化資料庫]，依據各種不同的主題或形式再加以延伸。然而欄位形式的詮釋資料屬於制式化的架構，其格式、內容的建置都必須遵循規定，這項特性卻在資訊檢索系統中侷限了資料操作的彈性，因此為了在檢索系統中，發揮資料的關聯特性，我們將欄位架構的詮釋資料，在檢索期間轉換成屬性標籤架構，兩種資料描述架構範例如下《圖 3-1》所示：



Meta-Data Format :

Data1

Identifier	案號	事由	原告	被告	判決日期
D1	單民第 2 號	土地引渡並不當利得請求事件	謝阿雨	謝大双	T08/11/28

Attribute Tag Format :

Data1 {

```
<Identifier>D1</Identifier>  
<案號>單民第 2 號</案號>  
<事由>土地引渡並不當利得請求事件</事由>  
<相關人 value='原告'>謝阿雨</相關人>  
<相關人 value='被告'>謝大双</相關人>  
<判決日期>T08/11/28</判決日期>  
}
```

《圖 3-1》 Metadata & Attribute Tag Format

『**Meta-Data Format**』與『**Attribute Tag Format**』之差別在於以下三點：

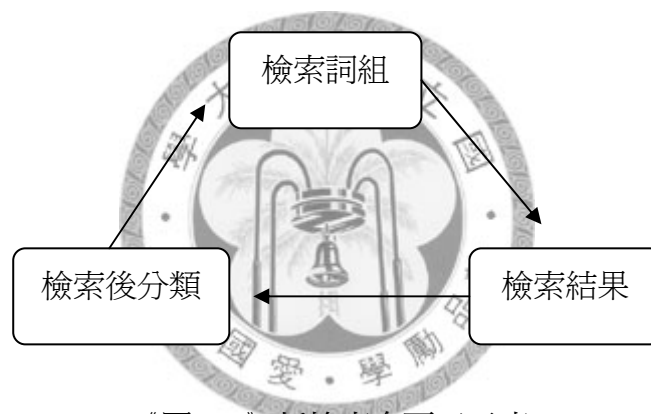
一、在觀察角度上，『**Meta-Data Format**』的建構方式為依據範例中”Data1”此資料的特徵，根據所規定的六個欄位填入對應的內容，建構方向為『單向』的由資料到欄位，描述方向則反之。而『**Attribute Tag Format**』的方式則為符合 Data1 資料的形容標籤，方向為『雙向』的互相索引，並且因為雙向的特性，標籤與標籤藉由 D1 的中介而互相產生關聯。

二、資料擴充性上，『**Meta-Data Format**』在規定上無法隨意新增欄位或填入資料，但『**Attribute Tag Format**』卻沒有這限制，舉例來說，我們可以在《圖 3-1》的『**Attribute Tag Format**』中增加<事由>土地引渡</事由>、<事由>不當利得</事由>兩個形容標籤，既不會破壞對原始資料的描述與格式，反而增加資料關聯的描述性；而這兩項標籤可經由自動化的方式產生，例如：斷詞系統，亦不會增加數位資料建置的成本。

三、在資料的操作性上，由於雙向關聯的特性，『**Attribute Tag Format**』在檢索建議系統的操作上能有更好的表現，可藉由使用者的所輸入的檢索詞彙，進而由此詞彙擴展出與其有關聯的標籤，予以使用者作為檢索策略建議。在檢索期間的詞彙組合操作、檢索結果後分類與標籤樹的呈現上，以及更進一步的關聯文件詞頻分析等協助工具，『**Attribute Tag Format**』都能提供更大的資料操作彈性給予系統製作者，藉由標籤運算彰顯資料本身資料價值。

3.2 檢索介面三元素

在資訊檢索系統後分類架構中，我們將符合檢索結果的文件，依據限定屬性的標籤或是欄位，進行資訊的統合整理，以提供使用者目前檢索結果的概觀資訊。而標籤樹則是立基於檢索後分類的架構上，更進一步將結果資訊，依據不同的主軸面向，以樹狀的階層式架構，提供文件中重要特徵關聯性；然而本論文所提出的檢索詞組操作，即在這兩項研究成果中，更進一步所衍生出的檢索介面。我們將一般檢索系統中瀏覽檢視檢索結果的地方，加入了後分類系統以及檢索詞組的操作，再加上原本系統中所必須的檢索結果呈現，藉由三方面的互相交流影響，我們建立的一個交互檢索的機制，如下《圖 3-2》所示，



《圖 3-2》新檢索介面三元素

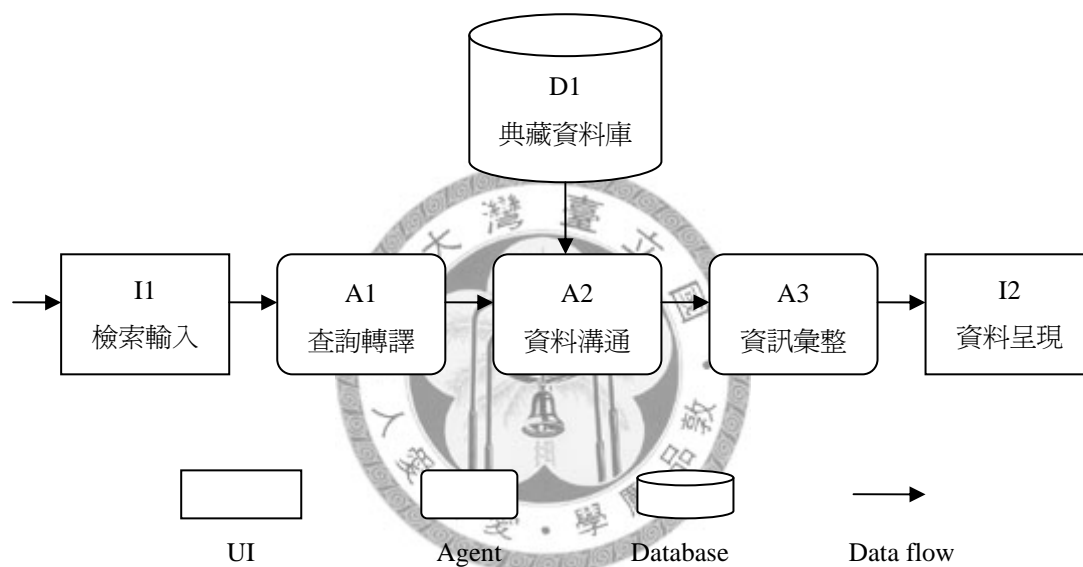
當使用者能在檢索中改變檢索詞組的組成，將立即影響檢索結果的內容，而利用檢索結果文件內的標籤資訊，經由統計整理產生檢索結果後分類，使用者藉此觀察檢索結果的變化，並了解符合檢索條件的文件內容特徵，進而由後分類中選擇標籤詞彙加入檢索詞組，亦或調整檢索詞組的內容而改變檢索結果。如此，三方面交互影響，形成循環不息的檢索流程，以解決研究動機中所提及的檢索歷程中斷之缺失。將資料作簡單的運算即可提供使用者更方便的參考資訊，提供簡易的操作功能讓使用者能方便運用資訊，發揮系統檢索力量。

3.2.1 檢索詞組控制之檢索系統

首先我們會先概述一般檢索系統的檢索流程，根據『日治法院檔案』系統中我們所必須面臨的改變而新增檢索流程的元件，整合成一新的檢索系統模型，並詳細敘述新的模型中增加元件的來由以及因應改變之處，以作為未來系統製作參考的資源。

3.2.1.1 一般檢索系統模型

目前大多數的典藏資訊檢索系統所使用的概略檢索流程模型如《圖 3-3》，



《圖 3-3》一般檢索系統模型

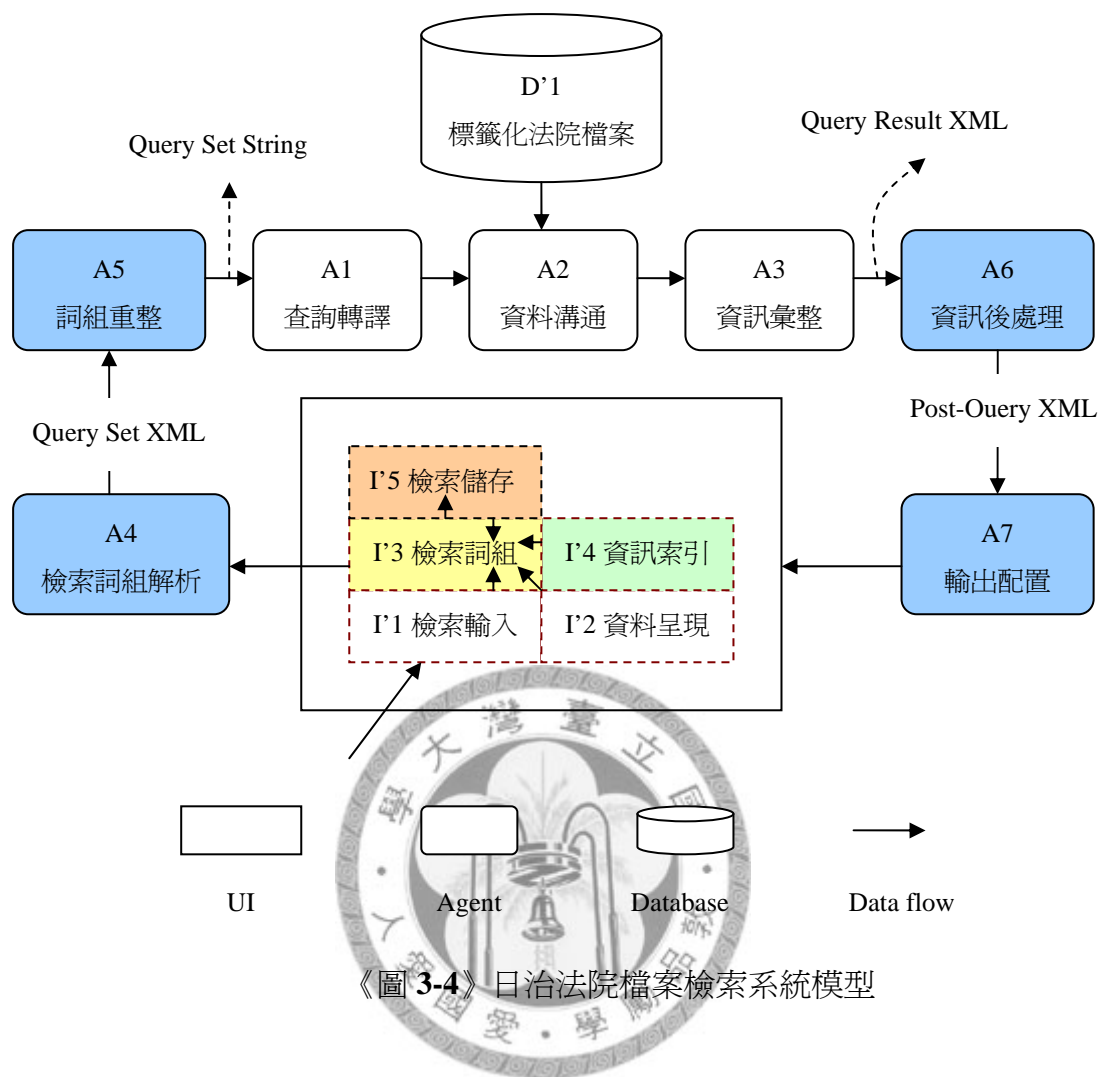
由《圖 3-2》所示，使用者經由檢索系統輸入介面 - **I1** 所制定的檢索輸入架構填寫檢索詞彙或詞組，檢索詞組由檢索轉譯程式 - **A1** 轉譯成資料庫查詢語彙 (Database SQL)，轉譯完成後傳遞給資料庫仲介程式 - **A2** 與資料庫 - **D1** 溝通，取得在資料庫中，符合使用者檢索條件的相關資料；最後由資訊彙整程式 - **A3** 統整所取得的資料，進行分頁、標注與排序等後續工作後組成查詢結果 XML，讓資料呈現介面 - **I2** 將最後的檢索結果輸出呈現。使用者可利用系統所提供的資料輸出方式，例如：欄位排序，重新調整檢索結果的輸出（在此各階段模組為一概略的分切，某些系統可能因不同實做方法，資料流程或工作模

組會在此模型中產生誤差，但並不影響此檢索流程的結構，在此於以註明)。

由《圖 3-3》中我們觀察到兩個問題，第一、一般典藏資訊系統的檢索模組為直線性的檢索流程，也就是說，當使用者在檢索過程中，依據系統回傳的檢索結果希望調整或者更換檢索策略時，通常必須回到 **I1** 重新進行整個檢索流程。第二、在檢索的過程中，系統與資料的交流上僅僅只限於使用者檢索條件的搜尋比對，或是檢索結果的呈現的排序，然而在某些情況下，檢索結果排序並無法提昇檢索效率或解決使用者的檢索需求。系統對於資料的運用過於單純，以至於資料使用的效率不彰，將導致增加使用者在資料篩選的工作負擔。

3.2.1.2 檢索詞組控制之整合介面檢索系統模型

承上 3.2.1.1 節，依據一般資訊檢索系統的檢索流程所提出的兩個問題，首先我們先在系統內加入過去本實驗室所發展的『後分類』資訊整合索引與『標籤樹』的階層顯示的架構，以提高系統對於資源運用的效能。更在此條件下，重新提出一個能即時調控的互動式檢索詞組操作介面，整合整個檢索流程，修正後的新式檢索流程如下圖《圖 3-4》所示。



《圖 3-4》日治法院檔案檢索系統模型

在使用者介面上，我們將一般檢索流程上的兩個使用者介面 **I1**、**I2** 變更為模組化元件 **I'1**、**I'2**，並加入屬性階層顯示的後分類資訊索引架構的『資訊索引』介面模組 - **I'4**。以及即時顯示目前檢索結果的成因，匯集檢索詞組的內容來源，以及提供檢索操作的目標的『檢索詞組控制』介面模組 - **I'3**。用以暫存使用者建構的檢索詞組，以及作為文件篩選後的文件集儲存索引的『檢索儲存』面介面模組 - **I'5**。以此五項介面模組整合成新的使用者綜合檢索介面，而檢索詞彙的建構來源則由 **I'1** 的使用者輸入、**I'2** 的文件篩選、**I'4** 屬性標籤選擇，以及 **I'5** 的過去檢索儲存，四種方式，並由 **I'3** 彙整檢索條件並傳遞給系統進行檢索流程。

由於新的使用者檢索介面產生，在檢索流程上發生了兩項改變，第一、檢索詞彙的來源不在只限於使用者輸入。第二、由於檢索詞組操作介面的產生，系統所必須解譯的查詢指令也不再僅止於不同檢索詞彙的組合，更增加了檢索詞組的串連結構變化，因此原 **A1** 查詢轉譯的工作必然無法因應需求；而新增的後分類資訊整合索引以及詞組儲存與操作介面，也將導致原流程 **A3** 所輸出的資訊彙整結果，不足以供應整合介面的資訊顯示需求。因此在處理資料流的 **Agent** 架構上，我們在一般檢索流程 **Agent A1、A2、A3** 的前後各新增兩個分析處理應用程式，串接原檢索架構。

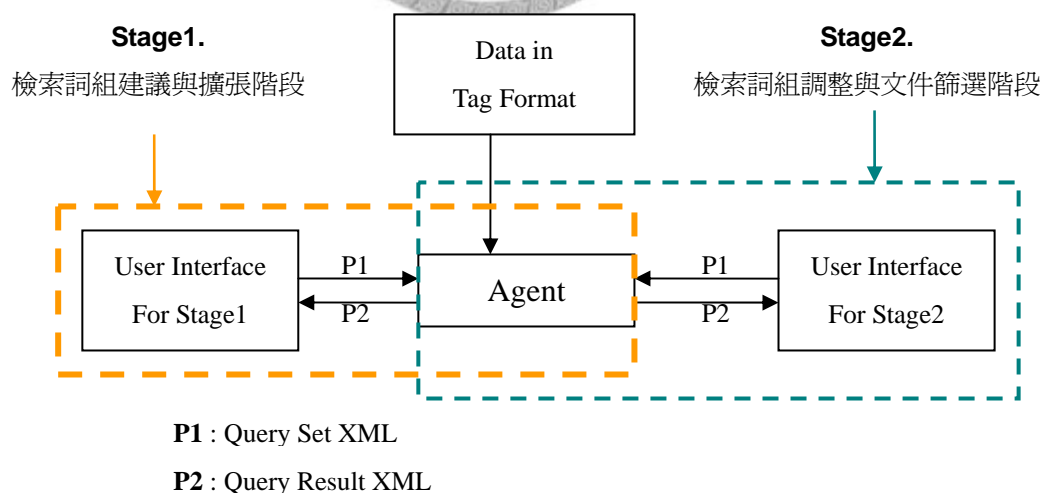
我們在檢索流程中，檢索詞組輸出的地方增加一個可偵測檢索詞組變化的『檢索詞組解析』程式 – **A4**，針對檢索詞組內的檢索詞彙標籤組合的任何改變進行相對應的處理動作，其工作有二：1.取得檢索詞組改變訊息，並啟動檢索程序。2.取得檢索詞組的內容『Query Set XML』，並傳遞給下一個處理程序。而後『檢索詞組重整』程式 – **A5** 進行檢索語句的轉換，將標籤型式的檢索詞組組合轉變成符合原系統查詢轉譯的詞組規則『Query Set String』。而後經由原一般檢索系統的 **A1、A2、A3** 流程，得到所有符合檢索條件查詢結果『Query Result XML』後，令『資訊後處理』程式 – **A6** 針對文件的各種維度進行後分類資訊統計整理，並將所得到的結果『Post-Query XML』。最後由『輸出配置』程式 – **A7** 根據介面組合形式，將所有系統產出的 XML 內容分別配置到因應的位置，顯示檢索結果。

經由上段所概述的第一輪檢索流程結束後，使用者可以依據系統所提供的資訊，藉由點選的方式改變 **I3** 介面內檢索詞組的內容，而 **A4** 將會收到 **I3** 內容改變訊息，重新啟動下一輪的檢索流程。經由介面的統整，使得系統中檢索與檢視兩種需求介面合而為一，介面與介面間交互影響並即時顯示，而達成檢索歷程操作的一貫性。在檢索詞組操作介面的輔助下，可以依據各種不同的

資料與系統特性，利用複合式的介面模組自由調整使用者介面，並依循相同的檢索流程模式，產生不同的檢索系統功能，以達成檢索各階段不同的目的。

3.2.2 兩階段檢索詞組控制系統

在一般的檢索系統中，我們可以將使用者的檢索行為概分為兩個階段，第一階段為檢索詞彙建立，對應到一般檢索系統即為一般或進階檢索；第二階段則為檢索結果的檢視，而其中兩階段的分界點以使用者點選『進行檢索』的按鈕為依據。在資訊檢索系統中加入檢索後分類的架構後，大幅度的增進第二階段檢索結果檢視的功能性；然而在檢索建議功能的啟發下，後分類的資訊索引架構亦可對使用者於檢索第一階段產生協助功能，藉由關聯標籤的牽引，可以檢視檢索目標的正確性以及相關欄位的關聯詞彙擴張，加上檢索詞組介面的統整與輔助，更能彈性的訂定檢索策略，以複數檢索組合進行檢索操作，以達到更完善的檢索目標。因此我們可以經由《圖 3-4》的檢索模型中，配合兩階段的檢索，衍生出新的系統模型架構，如圖《圖 3-5》，以下兩小節分別介紹各階段檢索之特色。



《圖 3-5》兩階段檢索詞組控制系統模型

3.2.2.1 檢索詞組建議與擴張階段

檢索擴張意即檢索策略中 - 引用文獻滾雪球法 (Citation Pearl Growing)[Citation Pearl Growing]與後分類資訊索引架構的綜合應用。當使用者在建立檢索條件的階段，對使用者當前所提供的檢索目標 (root) 進行資料庫系統的檢索前探測，在標籤架構下將關聯詞彙依屬性統計整理後，提供給使用者作為檢索詞彙修正建議，或是檢索條件組合運用。在檢索擴張階段有兩個主要概念：a.檢索建議、b.檢索組合。

a.檢索建議：

檢索建議系統與後分類的演算法相同，將原本由檢索結果的詞彙統計與屬性分類的特性，在檢索策略形成期間即提供給使用者，除了給予使用者欲檢索目標在系統內的概略描述外，也鼓勵使用者根據系統整理的資料庫內容，修正檢索目標的文字詞彙，亦或進而讓使用者以選擇替代輸入，減低因文意用詞的誤差而發生檢索結果不完全、甚至失敗的情況。以『日治法院檔案』系統為例：當使用者希望查詢日治時期台灣在土地抵押借貸方面的司法程序，可以藉由『土地』的關鍵字，從系統提供『相關事由』排序，如下圖《圖 3-6》中紅色虛線內所示，發現『借貸』這同意詞彙，將之加入檢索詞彙組合的策略之中，抑或是直接挑選系統建議作為檢索目標，以達到最小的檢索失誤。



關鍵字 : 土地

關鍵字 土地

存在 4569 筆 相關案件資料

2008/06/24,10:02 儲存 送出查詢

相關詞條	土地借貸借契約	[818]
- 相關人員	土地引渡其他請求事件	[213]
- 案件事由	土地賣買豫約	[183]
- 結案年代	土地引渡請求事件	[165]
- 所屬類別	土地賣買	[138]

《圖 3-6》系統檢索建議介面以『土地』為例

b. 檢索組合：

檢索組合的概念系由標籤樹的階層式資訊整合架構衍生而來，標籤樹的階層架構中，將文件由屬性標籤自樹根 (root) 一層層解析到樹葉節點，整個解析路徑所通過的節點，可視為一連串的標籤詞彙組合。在後分類的架構上，標籤樹的目的為聚焦特定屬性特徵的文件群；然而

輸入檢索條件:

關鍵字 土地

and

關係人 臺灣土地建物株式會社

存在 94 筆 相關案件資料

2008/05/01,05:54 儲存 送出查詢

相關詞條

- 相關人員
- 案件事由
- 結案年代
- 所屬類別

臺灣土地建物株式會社 [92]

株式會社日本勸業銀行 [6]

三十四銀行 [5]

日本勸業銀行 [3]

周氏阿貴 [2]

黃奕謙 [1]

《圖 3-7》系統檢索建議介面詞組組合

此處檢索組合的目的則為使用者保存多面向彈性的檢索策略，我們提供簡易的邏輯中介『AND』、『OR』、『NOT』，由使用者自行輸入、從檢索建議中選擇欲使用的詞彙增加檢索條件，利用邏輯中介將標籤組合成檢索詞組，如上《圖 3-7》所示，以供使用者為之後第二階段檢索結果篩選與調整作準備，也可成為使用者未來檢視檢索結果的入口，並在檢索結果篩選調整後成為檢索儲存的標示與未來個人文件的索引。在檢索組合形成階段，檢索建議會依據目前已被挑選的詞彙組成，即時調整檢索建議的內容。

3.2.2.2 檢索詞組調整與文件篩選階段

當資料庫使用者決定檢索策略後，無論是利用簡單的單一關鍵字，或是一到數個檢索詞組查詢，開始進行檢索後，系統將會自資料庫內取出符合檢索條件的所有文件，並依據資料的各屬性標籤作後分類資訊整合處理。經由後分類的統整，使用者將不必瀏覽所有系統列出的資料，即能概略了解檢索策略是否正確，使用者可選擇所關注的資訊，利用後分類作為進一步的查詢索引。

在後分類的基礎上，檢索結果除了顯示符合的文件，我們也將上一階段『檢

索擴張』所得到的檢索詞組一併加入資訊呈現的介面中，並賦予即時操作的特性，讓使用者在檢索過程中能夠方便的調整檢索詞組內元素的組合，觀察檢索結果的變化，分析比較並得到所需要的資訊。在此階段內包含兩特性：a.圖層式檢索操作、b.檢索需求整合。

a. 圖層式檢索操作：

在一般的檢索系統中，檢索歷程(歷史)為許多檢索系統的基本功能之一，其主要的目的在於紀錄使用者所使用過的檢索詞彙，提供使用者作為檢索結果的索引或調整檢索的參考。然而在後分類系統的幫助下，使用者增加檢索詞彙的方式不再限於自行輸入關鍵詞，自後分類的項目中選取檢索詞彙則更加簡單、直覺並快速，因此我們調整過去的作法，將以紀錄的目的檢索歷程，轉換成可操作的檢索詞組。

圖層的概念來自於影像層疊的技術，在本論文 2.2.3 中有簡單的介紹，在一般系統的檢索過程中，文件篩選的作法通常是利用增加條件詞彙來限縮檢索結果，當所使用的條件越多，在交集的運作下，所得到的結果將越精確，資料數量也隨之降低。當我們將一個個的檢索條件詞彙以條列式的方式呈現時，這樣的格式剛好與影像處理軟體中，對圖層的操作樣式相仿，也由此提供我們對於檢索詞組操作設計的靈感。因此我們在檢索條件詞彙組合的地方，加入『新增』、『刪除』、『遮蔽』三種操作方式予以使用者，使用者可以利用這三種方式來調整檢索詞彙組合，並觀察檢索結果。



《圖 3-8》系統檢索詞組操作介面

b. 檢索需求整合：

使用者增加檢索條件的方式除了藉由後分類系統，迅速的選擇所需要的詞彙標籤加入目前檢索詞組外，還可利用手動增加檢索詞彙以及篩選符合條件之文件。在此階段，檢索詞組控制介面將接收所有的檢索需求，成為介面中所有使用者動作的匯集地，並成為啟動系統檢索機制之鑰。藉由此即時的調整機制，當使用者在操作檢索詞組的時候，能夠立刻的根據後分類的整合資訊，了解檢索結果的變化，而不再如同過去訂定檢索組合後，一次又一次的從頭開始瀏覽檢索結果，並判斷是否存在符合檢索目的的資訊。而系統中新式的檢索詞彙操作，將為後分類帶來新的延伸應用，讓後分類系統除了反應檢索結果的文件，亦幫助詮釋檢索詞組的意義。



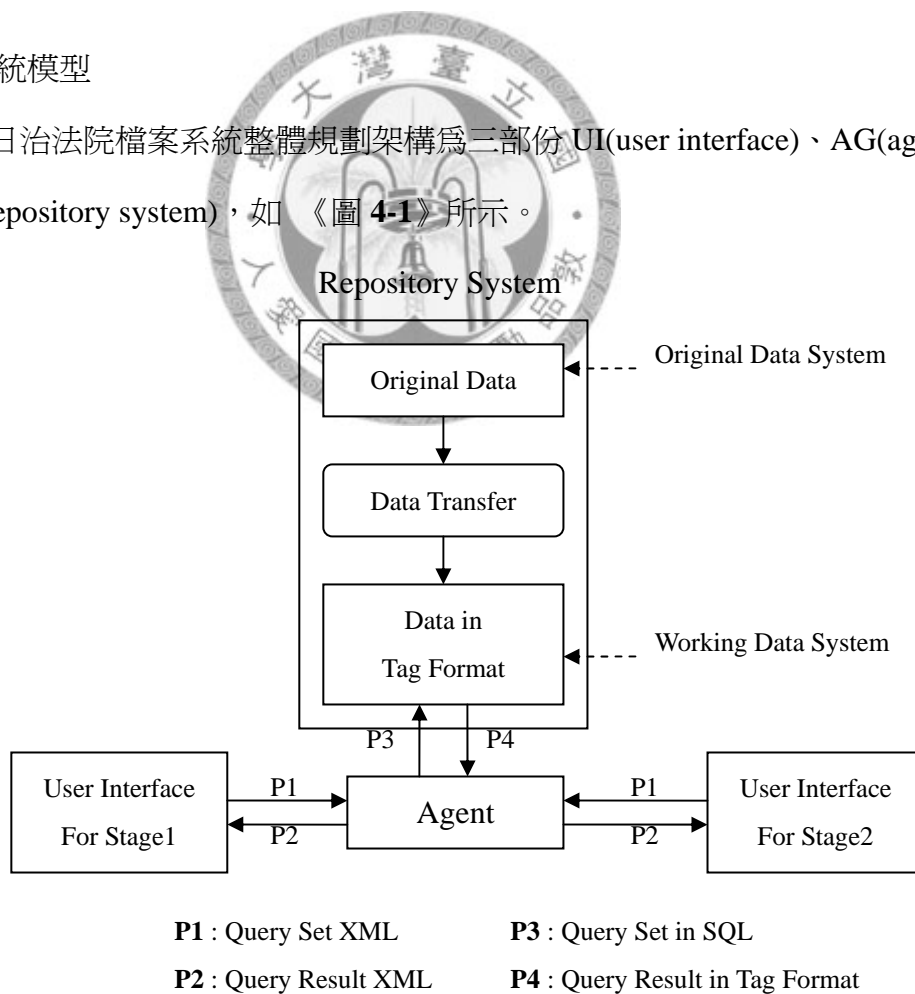
第四章 日治法院檔案系統實做

4.1 系統概述

本章我們將利用第三章所論述的『二階段檢索詞組控制』系統模型，再加上標籤資料轉換程序，我們得到了一個以檢索三元素為基礎的完整系統模型架構，我們將依據此模型的各元素進行系統實做，並考量實際資料的問題，建構網頁模式的『日治法院檔案』資料庫檢索系統；時間範圍在台灣日治時期，自明治 28 年（西元 1895 年）到昭和 20 年（西元 1945 年）間，台灣四個地方法院的相關司法文件，詳細資料將在下一節介紹。

4.1.1 系統模型

在日治法院檔案系統整體規劃架構為三部份 UI(user interface)、AG(agent) 與 RS(repository system)，如《圖 4-1》所示。

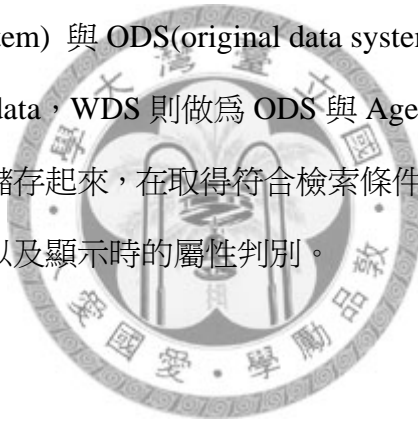


《圖 4-1》日治法院檔案系統模型

1.UI 的部份負責系統與使用者之間的溝通，接受使用者所輸入的檢索詞彙，並呈現系統查詢結果以及檢索結果後處理的相關資料給與使用者。依據檢索階段的不同分為兩部份，檢索詞組操作與輔助工具的呈現也在此處呈現。

2.Agent 負責取得使用者的檢索條件組合，並轉譯使用者的檢索詞彙成為資料庫檢索語言（SQL）的樣式；與資料庫溝通並組織查詢結果、進行屬性標籤後分類整理，將最後整合資料以 XML 資料格式回傳，並依據 UI 的顯示模組予以適當的資料進行呈現。

3.Repository system 儲存系統將用到的各種資料，其中又包含了 WDS(working data system) 與 ODS(original data system)兩項系統。ODS 存放各地方法院的原始 metadata，WDS 則做為 ODS 與 Agent 之間的橋樑，以標籤的形式轉換詮釋資料並儲存起來，在取得符合檢索條件的資料後，提供給 AG 作進一步的系統運算，以及顯示時的屬性判別。



4.1.2 系統環境

以下是我們系統建置與運作的環境

- Apache Web Server Version 2.2.4
- PHP Script Language Version 5.2.3
- MySQL Database Version 5.0.45
- phpMyAdmin Database Manager Version 2.10.2

測試瀏覽器：

- Internet Explorer Version 6.0.2900.2180
- Mozilla Firefox Version 2.0.0.14

4.2 系統資料準備 – 屬性標籤樣式

在『日治法院檔案』的詮釋資料分為兩種：

一、為以書冊為單位的 metadata，主要描述書籍的資料，內容一共有五個欄位，如下表《表 4-1》所示。

《表 4-1》法院檔案 - 書籍 Metadata 格式

欄位代號	代表意義	範例
ID	書籍序號	Cu0001
Class	書籍層次類別	新竹地院 民事類 民事判決書類 民事判決原本
Title	書籍封面提名	大正 8 年判決原本第 1 冊
Content	書籍封面內容	第 1-200 號
Volume	上冊或下冊	A

二、以案件為單位的 metadata。在日治法院檔案的檔案架構中，隸屬於民事判決、刑事判決、公證書三類的司法文件，有另外為書籍中的每件案件建立各別的 Metadata。也就是說，在此三類的書籍中，除了建置了書籍的 Metadata，其書冊內的案件亦一併額外建置案件 Metadata。

然而整個法院檔案進行數位化的時間橫跨數年之久，因各階段數位化的策略不同，導致各法院的案件資料在建構 Metadata 時出現部份差異；除此之外，由於不同類型的司法案件，其所紀錄的內容亦有其個別特徵；舉例來說，民事判決的法律文件內的相關人員，有『原告』與『被告』兩種不同的身份，但是在公證書中的相關人員則無上述的兩個身份，並且紀錄的關係人也不一定為兩灶雙方。以下《表 4-2》就各種類型案件 Metadata 作欄位比較，並在最後作統一欄位的歸納設計。

《表 4-2》各類型案件 Metadata 欄位差異比較表

資料類型 欄位名稱	民事判決 - 台中法院	民事判決 - 台北,新竹,嘉義法院	公證書 - ALL	刑事判決 - ALL
出處	◎	◎	◎	◎
案件編號	◎	◎	◎	◎
書類名稱		◎		◎
案件事由	◎	◎	◎	◎
關係人			◎	
原告	◎	◎		
被告	◎	◎		◎
檢察官				◎
原告代理人		◎		
被告代理人		◎		◎
法官		◎		◎
書記官		◎		◎
判決日期	◎	◎	◎	◎
是否上訴		◎		◎
所在法院		◎		◎
備註	◎	◎	◎	◎

由上表可得知各案件內容差異最大的地方，在於案件相關人員紀錄的內容，也就是表《4-2》以桃紅色所標示的部份，而其中以台北、新竹、嘉義三個地方法院的『民事判決』類型的案件 Metadata 欄位最為完整，因此我們以此類案件的欄位型式作為『日治法院檔案』系統中的案件資料 WDS 的欄位架構基礎，並利用屬性標籤的架構紀錄各案件中差異的部份。我們將案件人員分成三種，『主動關係人』、『被動關係人』與『其他相關人員』三種，並分別將各類型案件內容的關係人以標籤型式個別對應到這三種欄位中，並在標籤的名稱屬性上註明其原始的欄位名稱。以下的對應表《表 4-3》則為我們最後的案件資料在 WDS 內 Metadata 格式與內容，並以實例註明其紀錄型式。

《表 4-3》法院檔案 – 案件 Metadata 格式

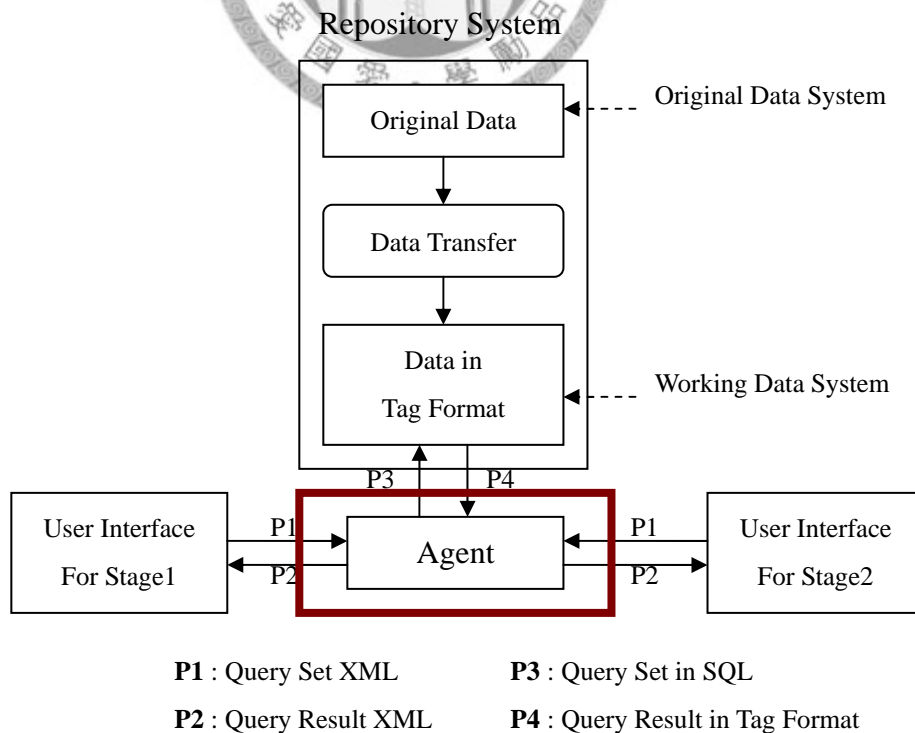
欄位代號	代表意義	各類案件相關欄位對應	ODS	WDS
Identifier	案件序號		Ntu_cu10101000000 1	
Class	所屬類別		cu10101	<Class value='cu10101'> 新竹地方法院 </Class>
BookId	所屬書籍編號		Cu0001	<Book id=' Cu0001'> 大正 8 年判決原本 1 冊 </Book>
TccNum	案件編號		大正 8 年單民第 2 號	<Tccnum> 大正 8 年單民第 2 號 </Tccnum>
Title	書類名稱		判決	<Title>判決</Title>
Content	案件事由		土地引渡並不當利得請求事件	<Content> 土地引渡並不當利得請求事件 </Content>
Date	判決日期		大正 08/11/28	<Date num='19191128'> 大正 08/11/28 </Date>
PersonM	主動關係人	關係人 原告 檢察官	謝阿 X	<PersonM value='原告'> 謝阿 X </PersonM>
PersonB	被動關係人	關係人 被告	謝 X 雙	<PersonB value='被告'> 謝 X 雙 </PersonB>
PersonO	其他關係人	原告代理人 被告代理人 法官 書記官	謝 X 泉 篠崎 X 太郎	<PersonO value='原告代理人'> 謝 X 泉 </ PersonO > < PersonO value='法官'> 篠崎 X 太郎 </ PersonO >
Appeal	是否上訴			
Location	所在法院		臺北地方法院新竹	<Location>

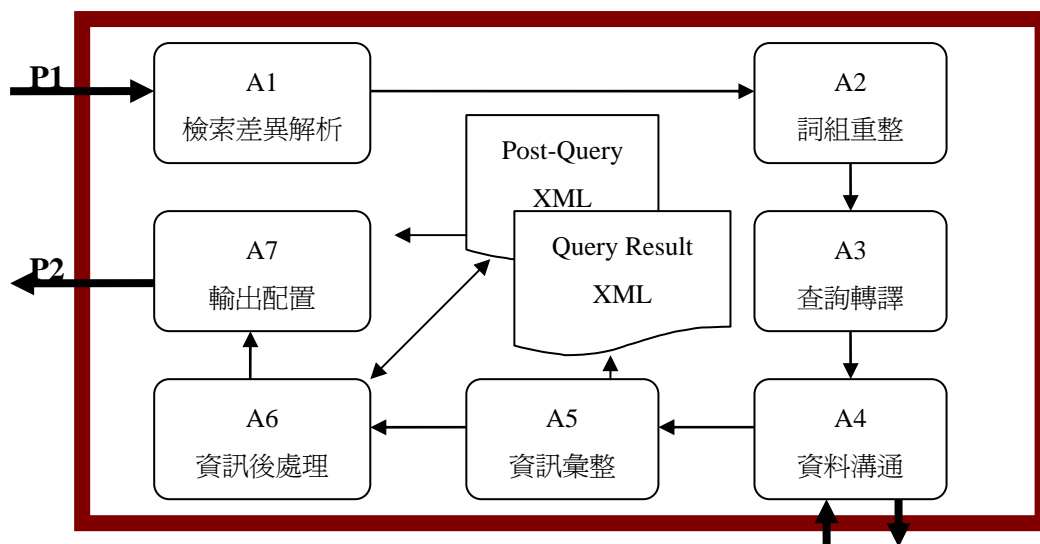
			出張所	臺北地方法院新竹出張所</Location>
ImgSite	影像位置		9-14	
Temp	備註			

4.3 系統實做

系統的實做上，主要分為三部份依序討論。第一部份為檢索程序資料處理 Agent 建構，針對資料在各階段的 Agent 的 I/O 格式與運作方式作詳細的描述。第二部份為二階段檢索操作之檢索建議與組合 UI 的建置，以及根據其所必須的檢索建議與詞組組合所作的介面設計與調整。第三部份為二階段檢索操作的檢索檢視與篩選 UI 建置，並依據其所需要的特性，檢索詞組操作與後分類階層資訊索引，以及差異性資料架構顯示作設計與調整。

4.3.1 Agent - 檢索程序資料處理流程





《圖 4-2》檢索程序資料處理流程

在上《圖 4-2》所示，檢索程序資料處理的 Agent 依共有 7 個元件，以下依據每個 Agent 的起動因素 (Start up)、資料輸入格式 (Data Input)、資料輸出格式 (Data Output)、運作方式 (Function) 四項分別描述之：

A1：檢索差異解析

Start up：接收到 Query Set XML file name

Data Input：Query Set XML

Input 的 Query Set XML 格式範例如下，

Ex：

```

<Query name=' QSxml' count='3'>
  <Term Tid='t1' Field='F1' Not='0' Hidden='0'>V1</ Term>
  <Term Tid='t2' Field='F2' Not='0' Hidden='0'>V2</ Term>
  <Term Tid='t3' Field='F2' Not='1' Hidden='0'>V3</ Term>
  <Combine> t1+t2|t3</Combine>
</Query>
  
```

檢索詞組以 XML 中 <Query> 標籤所群集，其屬性有兩項：

Name : 註明檢索詞組的代號

Count : 檢索詞彙數量

各檢索詞彙以 Term 標籤所包覆，V1 代表檢索內容，其屬性分為四項：

Tid : 詞彙編號

Field : 檢索的目標欄位

Not : 是否為反檢索

Hidden : 是否暫時遮蔽

檢索詞彙標籤的組合以<Combine>標籤註明，以各檢索詞彙標籤的代號為元素，連結代號有兩種組合樣式：

+ : 代表詞組間以 and 連接，為系統預設值。

| : 代表詞組間以 or 連接，為檢索建議中，複數選則的預設值。

Data Output : Query Set XML

Content :

- 1.Load Query Set Xml by Query Set XML file name from UI
- 2.Pass Query Set Xml to A2

A2 : 詞組重整

Start up : 接到 Query Set XML

Data Input : Query Set XML

Data Output : Query term set string

Content :

藉由 Query Set XML 中 Combine Tag 內所註記的檢索詞彙組合樣板，在樣板分解的過程中利用標籤 id 自 Term Tag 取得檢索詞彙的相關內容，並依據解析規則分別賦予樣板中介『and』、『or』、『not』的組合，最後輸文字型的 Query term string。以下為轉換演算法：

Function1 :Count(Array) = the number of element in Array

Get the 'Pattern1' from the value of Combine Tag

Split 'Pattern1' by ' + ' to Plv1 ,Plv1 is an array

For i=1 ~ Count(Plv1)

Split Plv1[i] by ' | ' to Plv2[i] , Plv2[i] is an array

End for

SET TermSet is an Array

For i=1 ~ Count(Plv1)

If Count(Plv2[i] >1)

Foreach Plv2[i] as Tag_id

SET temp is an empty Array

Get the value Field = f1 , value Not = n1 and Tag Value=v1 from

Query/Term tag where Tid= Tag_id and Hidden != 1 ;

If n1==1

v1=v1+'

temp[]= v1

Join temp by ',' to String Term

TermSet[i]=f1+' : '+ Term

Else

Get the value Field = f2 , value Not = n2 and Tag Value=v2 from

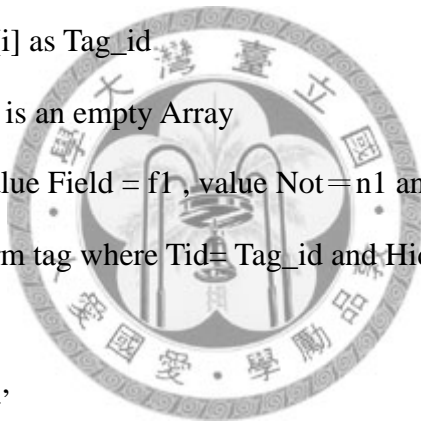
Query/Term tag where Tid= Tag_id and Hidden != 1 ;

If n2==1

v2=v2+'

end if

TermSet[i]=f2+' : '+v2



```

End else

End for

Join TermSet by '@' to String TermString

Return TremString;

```

以下為四種樣式的 Tag form 到 Term string structure 的轉換範例，分別為 Single、And、Or、Not。

1. Single term

```

<Query name='QSx' count='1'>
  <Term Tid='t1' Field='F1' Not='0' Hidden='0'>V1</Query>
</Query>
=> F1 : V1

```

2. And

```

<Query name='QSx' count='1'>
  <Term Tid='t1' Field='F1' Not='0' Hidden='0'>V1</Query>
  <Term Tid='t2' Field='F2' Not='0' Hidden='0'>V2</Query>
  <Combine> t1 +t2</Combine>
</Query>
=>F1 : V1@F2 : V2。

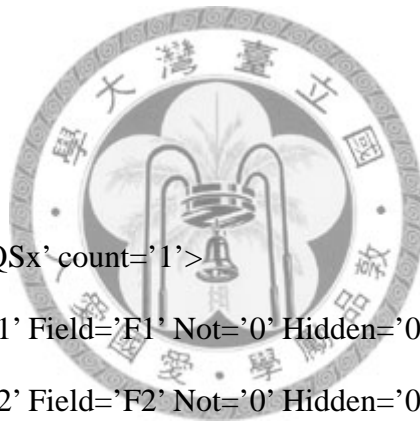
```

3. Or

```

<Query name='QSx' count='1'>
  <Term Tid='t1' Field='F1' Not='0' Hidden='0'>V1</Query>
  <Term Tid='t2' Field='F1' Not='0' Hidden='0'>V2</Query>
  <Combine> t1 |t2</Combine>
</Query>
=> F1 : V1,V2。

```



4. Not

```
<Query name=' QsX' count='1'>  
  <Term Tid='t1' Field='F1' Not='1' Hidden='0'>V1</Query>  
  <Combine> t1 </Combine>  
</Query>  
=> F1 : V1' .
```

A3 : 查詢轉譯

Start up : 接收到 Query term set string

Data Input : Query term set string

Data Output : SQL

Content :

Function1 :Count(Array) = the number of element in Array

Set Query term string = QTS , QTS is a String

Split QTS by '@' to QT , QT is an Array

Set Pattern is an empty Array

For i=1 ~ Count(QT)

Switch(QT[i])

Case " F : V " :

Pattern[i]="(F like V)";

Break;

Case " F : V1,V2 " :

Pattern[i]="(F like V1 or F like V2)";

Break;

Case " F : V' " :

```
Pattern[i]="( F not like V )";
```

```
End Switch
```

```
End for
```

```
Set Sql ="Select * from Table where ";
```

```
Sql = Sql + join Pattern by " and " + " order by id ";
```

```
Return Sql;
```

Ex:綜合範例 F1 : V1,V2 @ F2 : V3'

=>select * from table where (F1 like V1 or F1 like V2) and (F2 not like V3)

A4 : 資料庫溝通介面

Start up : 接收到 SQL

Data Input : SQL

Data Output : data result for sql

Content :

與資料庫建立連結，並使用資料庫讀取函數與資料庫溝通，在本系統是以 php 的三個 mysql 函數所構成：

1. mysql_pconnect (string server, string username, string password)
2. mysql_select_db (string database_name, resource link_identifier)
3. result = mysql_query (string SQL)

Return result

A5 : 資訊彙整

Start up : 接收到 result for sql

Data Input : result for sql

Data Output : Query Result XML

Content :

讀取 Result 取得資料庫內所有符合 SQL 的 Data，利用 mysql_fetch_array function，依序將 data 內的 working data 取出並組合成 Result XML。其中 ClassLevel、BookTitle、BookContent 標籤的內容必須參照另外的資料表才能取得。Result XML 內容完成後，將 xml file name 傳遞給 A6。其中一筆 Result XML 樣式如下：

```
<Doc>

  <No>1</No>

  <Id>ntu_cu101010000001</Id>

  <ClassId>cu10101</ClassId>

  <ClassLevel>新竹地院,民事類,民事判決書類,民事判決原本</ ClassLevel >

  <BookId> cu0001</BookId>

  <BookTitle>大正 8 年判決原本第 1 冊</BookTitle>

  <BookContent>第 1-200 號</BookContent>

  <Tcnum>大正 8 年單民第 2 號</Tcnum>

  <Title>判決</Title>

  <Content>土地引渡並不當利得請求事件</Content>

  <PersonM value='原告'>謝 X 雨</PsesonM>

  <PersonB value='被告'>謝 X 双</PsesonB>

  <PersonO value='原告代理人'>謝 X 泉</PsesonO>

  <PersonO value='法官'>篠崎 XX 郎</PsesonO>

  <Date value='大正 08 年' dateCount=' 19191128' >T08/11/28</Date>

  <Location>臺北地方法院新竹出張所</Location>

  <ImgSite>9-14</ImgSite>

</Doc>
```

A6：資訊後處理

Start up：接收到 Query Result XML file name

Data Input：Query Result XML(Rxml)

Data Output：Post-Query XML(PQxml)

Content：

Agent 6 啟動後將會依據收到的 Query Result XML file name 讀取 Result XML，依循每一筆文件的 xml，將系統需求的後分類標籤自 xml 中取出，並暫存與計算與排序後，重新製成新的 XML，稱作 Post-Query XML。在『日治法院檔案』系統中，我們選用五個欄位標籤作為後分類索引的目標分別為『所屬類別』 - <ClassLevel>、『相關事由』 - <Content>、『關係人』 - <PersonM>and<PersonB >、『結案年代』 - <Date>、『承辦法官』 - <PersonO value='法官'>，為改善資料在索引上使用的便利性，每項標籤在後分類資訊整理過程中都有經過小幅度的資料修正。Xml 完成後將 Post-Query XML file name 與 Query Result XML file name 傳遞給 A7。最後 Post-Query XML 內容樣式，以『相關事由』為例如下所述：

<PostClassification>

<Content value='相關事由'>

<cue count='16' >貸金請求事件</cue>

<cue count='8' >賣掛代金請求事件</cue>

<cue count='4' >損害賠償請求事件</cue>

<cue count='4' >離婚請求事件</cue>

<cue count='2' >業主權移轉登記手續請求</cue>

</Content>

</ PostClassification >

A7：輸出配置

Start up：收到 Query Result XML & Post-Query XML file name

Data Input：Query Result XML & Post-Query XML

Data Output：Query Result XML

Post-Query XML

Content：

根據不同階段的檢索流程，呼叫對應介面執行跟新動作，其運作的演算法內容如下：

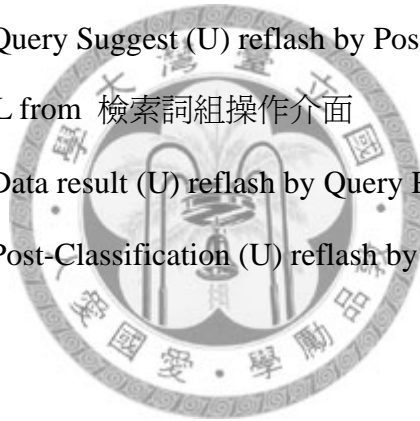
If Query Set XML from 檢索詞組建構介面

Call interface Query Suggest (U) refresh by Post-Query XML

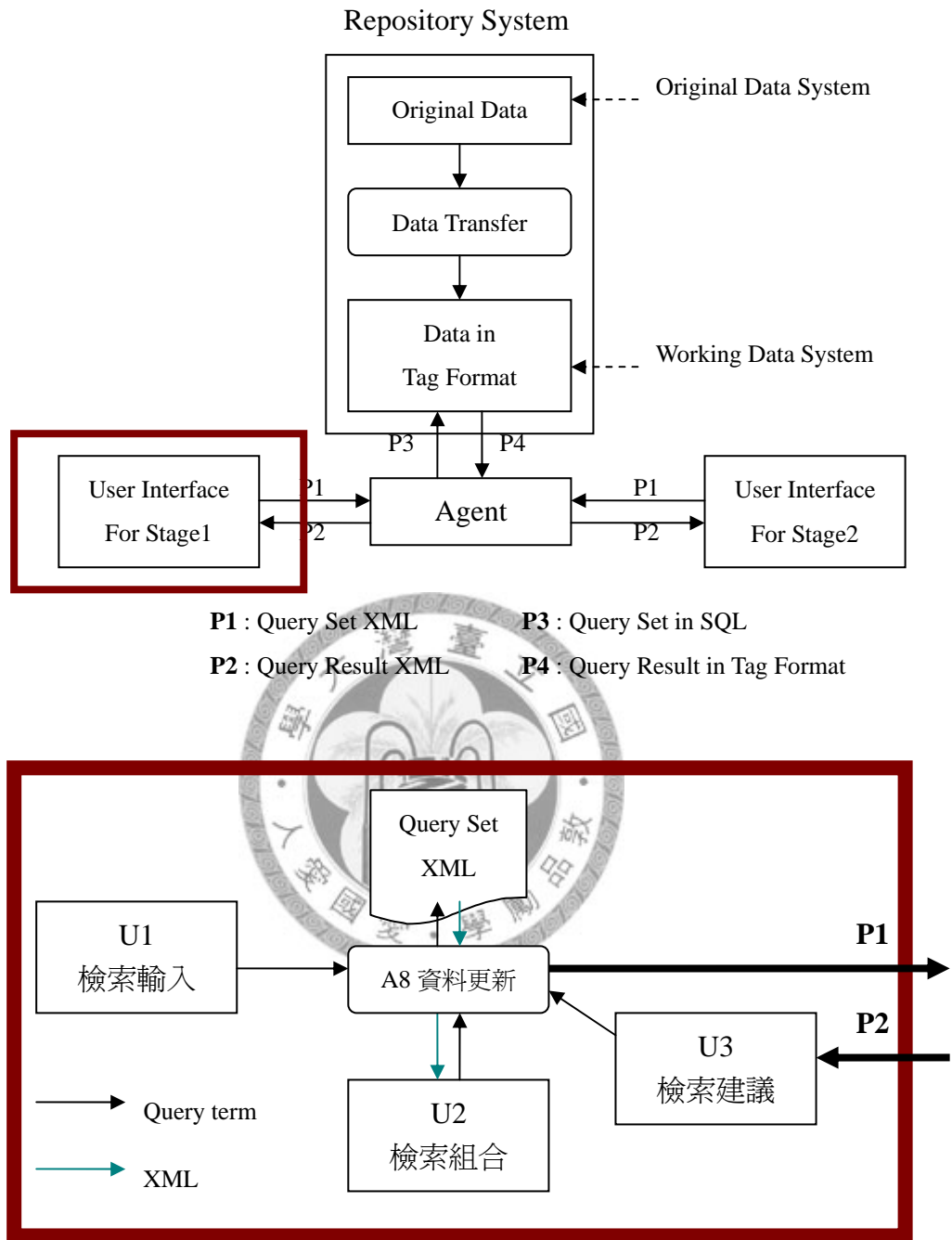
If Query Set XML from 檢索詞組操作介面

Call interface Data result (U) refresh by Query Result XML

Call interface Post-Classification (U) refresh by Post-Query XML



4.3.2 UI Stage 1 – 檢索建議與組合介面



《圖 4-3》檢索建議與組合介面

檢索建議與組合介面主要由一個 Query Set XML，一個中介資料溝通 Agent A8 作 XML 的資料更新，與三個顯示介面：檢索輸入介面 - U1，檢索組合介面 - U2，以及檢索建議介面 - U3，五項元件所共同組成，以下分別介紹之。

1. Query Set XML

Query Set XML 的資料樣式在 4.3.1 的 A1 agent 有介紹，在此便不再贅述，而 XML 的內容則由 U1、U2、U3 三個介面提供，並由 A8 進行 XML 的資料更新。

2. A8 - 資料更新程式

資料更新程式主要的工作在接收各使用者介面的資料輸出，並依據介面的要求更新 Query Set XML 的內容，並在資料更新後啟動 A1 執行檢索循環程序，以及更新檢索組合介面 U2。

Start up : AJAX function in UI

Data Input : Query term & function name(AddTag,DelTag,SetModify,TagModify)

Data Output : Query Set XML

Content :

資料更新分為以下四種方式：

1.AddTag (FieldName , Query)

給予新增的 Query term 標籤序號 Tid，並依據<Term>標籤格式將 Query term 加入 Query Set XML 中。將 String "+Tid" 串連到<Combine>標籤的內容後面。

2.DelTag (TagId)

根據收到的標籤序號 TagId 將對應的<Term>標籤自 Query Set XML 中刪除。將<Combine>標籤的內容以依序以'+',';'分解成二維陣列，移除 Tid 後在以','、'+ '的順序組合回來。如果 TagId 為'All'，則刪除 XML 所有資料。

3.SetModify (TagId1, TagId2, x{and|or})

根據收到的兩個標籤序號 TagId1, TagId2 以及應變更的串連樣式 x，將找出<Combine>標籤值內容中，TagId1 與 TagId2 之間符合'+ '或',';'的符

號，並置換成 x 後回存<Combine>標籤。

4.TagModify (TagId, An , Av)

根據收到的標籤序號 TagId, 將找出對應 id 的<Term>標籤並根據標籤內屬性名稱爲 An 的屬性，將更新其值爲 Av。

資料更新完成後，Agent A8 將傳遞 A1Query Set XML 的 file name，啓動檢索循環程序，並更新檢索組合介面 U2 的內容。

3. U1 - 檢索輸入介面

一般檢索系統都具備的檢索輸入介面，通常由 HTML 的文字輸入框，配合<select>或<checkbox>等選項標籤，以及按鈕等元件共同組合而成的輸入介面，通常依據系統需求而

查詢台灣地區重要圖書館館藏中西文報紙目錄

報刊名：

出版單位：

出版國家：

作品語文：

館藏單位：

類別：

查詢結果依： 排序

《圖 4-4》

決定輸入的欄位的樣式，如右《圖 4-4》所示，使用者需將所有的檢索條件資料都決定並輸入後，才進行查詢動作。而在本系統中，我們針對檢索輸入的方式與介面進行修改，運用 U2 檢索詞組介面可操作、堆疊、並即時更新的特點，我們將一次到位的檢索條件輸入方式，改成分段式的條件堆疊。系統檢索輸入介面如下《圖 4-5》，當我們將 Query term 送出後，利用 AJAX 函數 - AddQuery 啓動 A8 將 Query term 寫入 Query Set XML 中，當 Query Set XML 的內容改變，

The screenshot shows a search interface with two tabs: '一般查詢' (General Search) and '案件查詢' (Case Search). Under '案件查詢', there is a section titled '輸入檢索條件:' (Input Search Conditions). A dropdown menu is set to '-原告(民事判決)' (Plaintiff (Civil Judgment)). To its right is a text input field containing '徐大' (Xu Da) and a '+' button. Below this is a yellow box containing a search result: '關係人 徐大頭' (Related Person: Xu Dadao). At the bottom of this box, it says '存在 24 筆 相關案件資料' (24 related case records exist). At the very bottom of the interface, there is a timestamp '2008/06/16,12:31', a '儲存' (Save) button, and a '送出查詢' (Send Query) button.

《圖 4-5》檢索簡易與組合介面 – 檢索詞彙輸入

將會觸發 4.3.1 的 A1 agent 執行檢索流程，並即時的更新 U2 檢索組合的介面，也就是《圖 4-5》膚色色塊的部份。配合 U2 介面的彈性，使用者將不再受限於系統提供的檢索輸入介面的數量，而每一個檢索詞彙的建立都將經過檢索過程的檢驗，因此將解決詞組檢索失敗卻不清楚為何項條件所造成的檢索困擾。

Function :

```
QueryTermAdd ( FieldName , Query){ AddTag( FieldName, Query ) }
```

透過《圖 4-5》中『+』按鈕，將啟動此函數並傳遞紅色箭頭所指的檢索目標欄位內容 (FieldName)，以及綠色箭頭使用者輸入的檢索詞彙 (Query)，並執行 A8 中的 AddTag 程序。

3.U2 – 檢索組合介面

檢索組合介面提供檢索詞彙標籤的暫存與組合，介面的內容依據 Query Set XML 所生成，並提供兩種檢索詞彙標籤的操作，1.刪除檢索詞彙標籤，2.變更詞組串連中介，也就是『And』、『Or』、『Not』的變換。Query Set XML 中含有兩種標籤，<Term>標籤紀錄檢索詞彙、<Combine>標籤紀錄檢索詞彙組合方式，在介面中以 HTML 的超連結標籤 (Hyperlink) <a>格式，以及 onClick 動作連結兩個 AJAX 函數，並分別根據對應的功能對 Query Set XML 內容進

行修改。兩個 AJAX 函數分別為 QueryTagDelete & QuerySetCombination。

一般查詢 | 案件查詢

輸入檢索條件:

-原告(民事判決) : 徐大

事由 遺言

and

x 關係人 徐大頭

2008/06/17,08:11 儲存 送出查詢

《圖 4-6》檢索簡易與組合介面 – 檢索詞彙組合

Function :

1. QueryTagDelete (TagId) { DelTag (TagId) }

透過《圖 4-6》中綠色箭頭所指的『x』符號，將啓動此函數並傳遞目標的標籤代號 (TagId) 並執行 A8 中的 DelTag 程序。

2. QuerySetCombination (TagId 1 , TagId 2 , x { and/or }) {

SetModify (TagId1, TagId2, x) }

透過《圖 4-6》中紅箭頭所指的『and』符號，將啓動此函數並傳遞『and』前後兩個標籤代號 (TagId 1 , TagId 2) 以及變更內容 (x) 並執行 A8 中的 Set Modify 程序。

4.U3 – 檢索建議

檢索建議介面提供目前檢索組合中，與檢索詞組相關連的 Query term，依據不同的欄位區分類型，並利用統計方式將相關詞彙由多到少排序；目前檢索建議所提供的欄位一共有四種 1.相關人員 2.案件事由 3.結案年代 4.所屬類別，其內容來自於 Agent (A1-A7)執行完後所產生的 Post-Classification XML。系統檢索建議介面如下圖《圖 4-7》所示，依據檢索組合介面的內容，

× 事由 遺言

存在 1003 筆 相關案件資料

2008/06/17,08:11 儲存 送出查詢

相關詞條	<input type="checkbox"/>	遺言	[897]
- 相關人員	<input type="checkbox"/>	遺言公正證書	[32]
- 案件事由	<input type="checkbox"/>	遺言取消	[15]
- 結案年份	<input type="checkbox"/>	遺言變更	[11]
- 所屬類別	<input type="checkbox"/>	遺言公正	[9]
	<input type="checkbox"/>	遺言證書	[3]
檢索歷史	<input type="checkbox"/>	遺言贈與	[3]
	<input type="checkbox"/>	遺言全部取消	[3]
	<input type="checkbox"/>	遺言全部取消公正	[2]
	<input type="checkbox"/>	遺言一部取消	[2]
	<input type="checkbox"/>	遺言變更	[2]

《圖 4-7》檢索簡易與組合介面 – 檢索詞彙建議

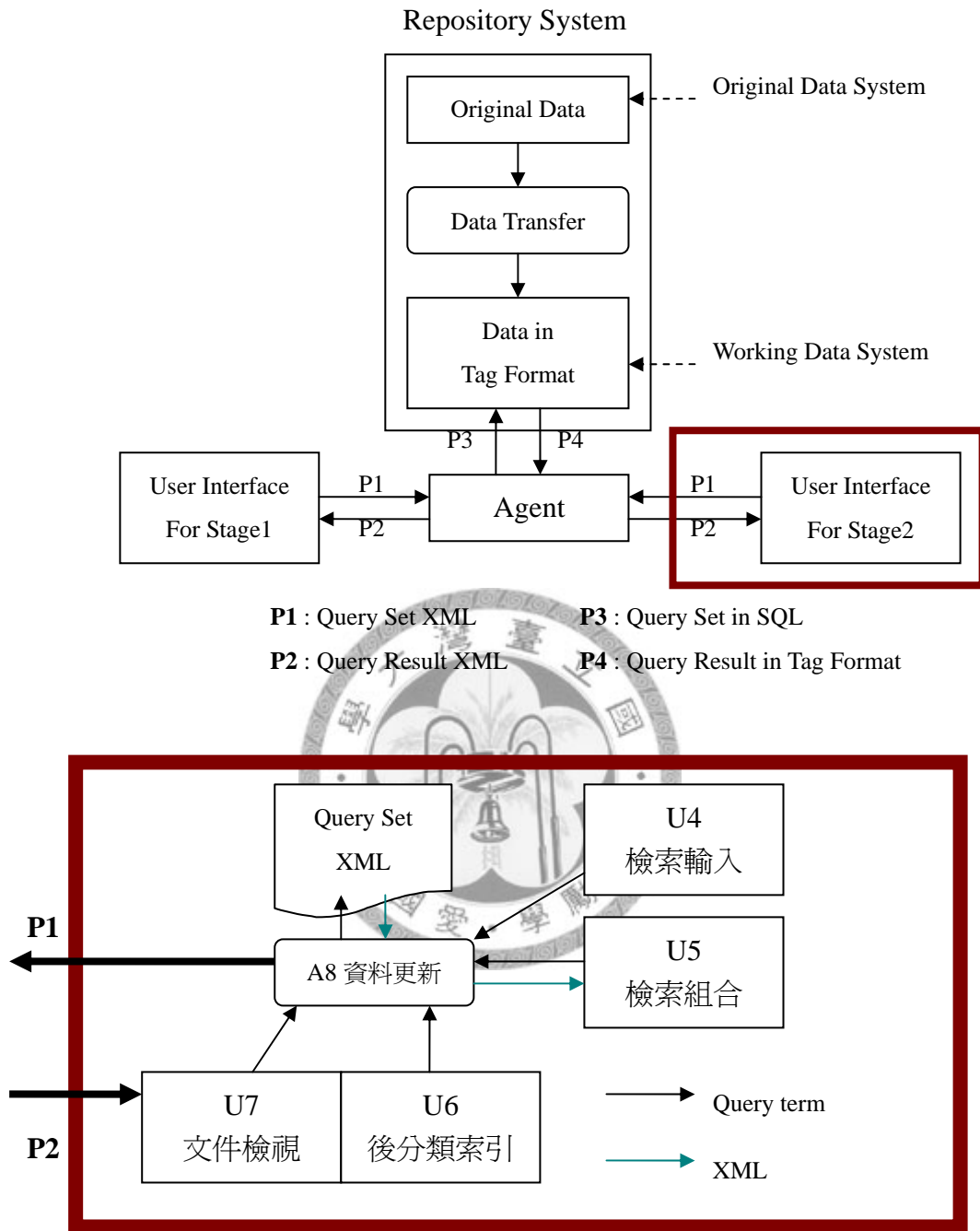
在《圖 4-8》中以檢索『事由 - 遺言』為例，在檢視方面，經過系統先行處理而整理出來的建議資料，在介面中分為四項，使用者可利用左邊黃色區塊以點選的方式變所關心換類別。在資料新增方面，系統提供兩個方式，1.直接點選右邊的詞彙標籤，加入檢索組合中，2.利用勾選的方式，將所關心的檢索詞彙勾選起來，送出後將整組詞彙加入檢索組合；利用勾選的方式，系統將會將勾選的詞彙以' Or'方式串連。介面製作方式與 U2 相同，都使以 HTML 的超連結標籤（Hyperlink）<a>格式，以及 onClick 動作連結 AJAX 函數，函數與資料新增介面 U1 所用的 QueryTermAdd 函數相同。

Function :

```
QueryTermAdd ( FieldName , Query){ AddTag( FieldName, Query ) }
```

啓動此函數將傳遞使用者所選擇的欄位類別作為 FieldName 的值，以及所選擇的標籤詞彙作為 Query，並執行 A8 中的 AddTag 程序。

4.3.3 UI Stage 2 – 檢索檢視與篩選介面



《圖 4-8》檢索檢視與篩選介面架構

檢索檢視與篩選介面的組成元件有以下六項：1.XML 文件 - Query Set XML 儲存檢索詞組的資料。2.資料中介溝通的資料更新 Agent – A8，作為 XML 的資料更新。3.檢索輸入介面 – U4，作為使用者輸入檢索詞會的介面。4.檢索組合介面 – U2，用以顯示與操作檢索詞組的介面。5. 後分類索引介面 – U6，

提供檢索結果的整合資訊與索引。6.文件檢視介面 – U7，提供符合檢索條件的相關文件瀏覽檢視的介面。一共六項元件組合而成檢索檢視與篩選的使用者介面，以下分別介紹之。

1. Query Set XML

Query Set XML 的資料樣式在 4.3.1 的 A1 agent 有介紹，在此便不再贅述，而 XML 的內容則由 U4、U5、U6、U7 三個介面提供，並由 A8 進行 XML 的資料更新。

2. A8 - 資料更新程式

資料更新程式主要的工作在接收各使用者介面的資料輸出，並依據介面的要求更新 Query Set XML 的內容，並在資料更新後啟動 A1 執行檢索循環程序，以及更新檢索組合介面 U5。與 4.3.2 UI Stage 1 – 檢索擴張介面中的 2.A8 資料更新程式相同。



3. U4 - 檢索輸入介面

提供文字輸入框讓使用者輸入 Query term，其介面如下《圖 4-9》所展示，



《圖 4-9》檢索檢視與篩選介面架構 – 檢索輸入

介面提供兩種選擇給使用者，如圖箭頭所標示的地方，1.縮小範圍，將使用者的 Query term 加入到原本的檢索詞組中，對應的 AJAX 函數為 QueryTermAdd。2.重新查詢，以此 Query term 作為新的查詢，當查詢結果為找不到任何相符的文件時，此選項將被系統設為預設值；對應的 AJAX 函數為 QueryTermNew。

Function :

1. QueryTermAdd (FieldName , Query){ AddTag(FieldName, Query) }


選擇『縮小範圍』並按下『查詢』按鈕後將啓動此函數，預設 FieldName 爲'ALL'查詢所有欄位，並執行 A8 中的 AddTag 程序。

2. QueryTermNew (FieldName , Query) {

DelTag('ALL') , AddTag(FieldName, Query) }

選擇『重新查詢』並按下『查詢』按鈕後將啓動此函數，預設 FieldName 爲'ALL'查詢所有欄位，並依序執行 A8 中的 DelTag，刪除所有 XML 內容，而後進行新增檢索標籤 AddTag。

4. U5 - 檢索組合操作介面



檢索組合操作介面提供檢索詞彙標籤的組合與操作，介面的內容依據 Query Set XML 所生成，並以 HTML 的超連結標籤 (Hyperlink) <a>格式，以及 onClick 動作連結啓動 AJAX 函數，並根據對應的功能進行 Query Set XML 內容的修改。介面提供兩種檢索詞彙標籤的操作，1.刪除檢索詞彙標籤，對應的 AJAX 函數爲 QueryTagDelete。2.檢索標籤遮蔽，暫時屏除檢索詞彙不進行查詢，但仍保留在檢索詞組當中。與標籤刪除的不同，在於提供使用者檢索策略操作的彈性 (舉例)，對應的 AJAX 函數爲 QueryTagHidden。

Function :

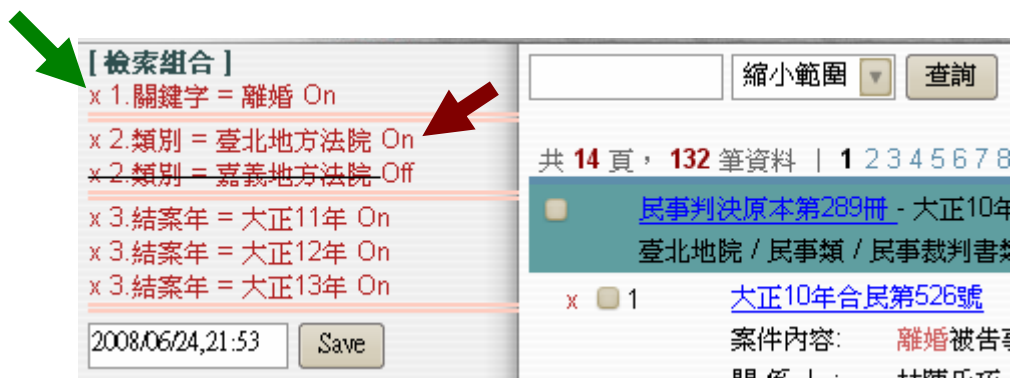
1. QueryTagDelete (TagId) { DelTag (TagId) }

透過《圖 4-10》中綠色箭頭所指的『x』符號，將啓動此函數並傳遞目標的標籤代號 (TagId) 並執行 A8 中的 DelTag 程序。

2. QueryTagHidden (TagId) { TagModify (TagId, 'hidden', x{0|1}) }

透過《圖 4-10》中紅色箭頭所指的『On/Off』符號，將啓動此函數並傳遞目標的標籤代號 (TagId) 並執行 A8 中的 TagModify 程序，其中傳遞

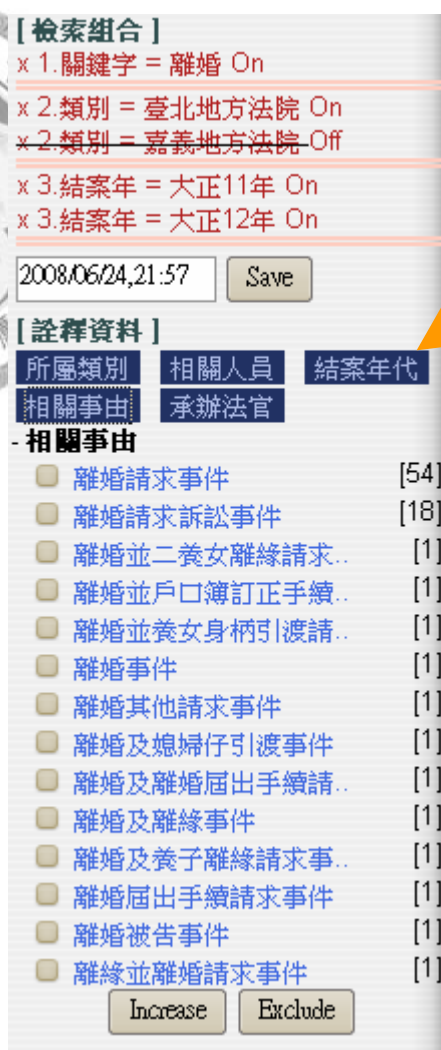
TagModify 的參數有兩種，如果標籤為正常狀態，則 $x = 1$ ，如果標籤為遮蔽狀態，則 $x = 0$ 。



《圖 4-10》檢索檢視與篩選介面架構 – 檢索操作

5. U6 – 後分類索引介面

後分類索引介面的內容來自於 Post-Classification XML，其為根據目前的檢索詞組，自所有符合檢索條件的文件中的五種標籤作統計與排序，分別為 1. 相關人員 2. 案件事由 3. 結案年代 4. 所屬類別 5. 承辦法官等五種，當 Post-Classification XML 完成後，由 Agent A7 啟動與更新此介面。系統的後分類索引介面如右《圖 4-11》所示，介面分為兩部份，1. 為資料顯示控制，在右圖橘色箭頭所指示的地方，使用者可以利用點選的方式選擇介面下半部資料的類別，用意在於當資料量過多的時候，『相關事由』與『承辦法官』兩類資料將被排到頁面的下方，使用者可利用顯示控制選擇後分類資料的類型，不需使用捲



《圖 4-11》後分類索引

軸的方式而使得頁面瀏覽產生混亂。2.為後分類資訊索引的部份，使用者直接點選標籤詞彙，將所關注的資料加入目前檢索詞組當中，其對應的 AJAX 函數為 QueryTermAdd。

Function :

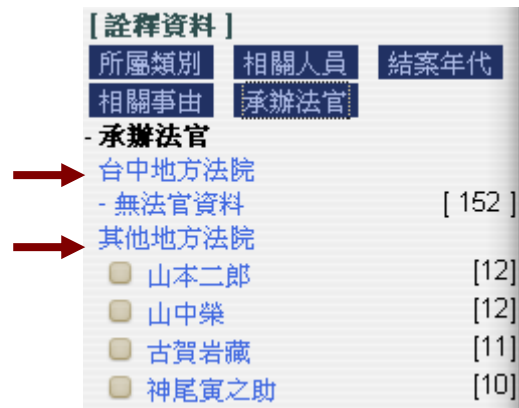
```
1. QueryTermAdd ( FieldName , Query){ AddTag( FieldName, Query ) }
```

點選後分類的標籤詞彙後將啟動此函數，預設 FieldName 為階層的標題，Query 為點選的詞彙，並執行 A8 中的 AddTag 程序。

而在此介面的實做上出現兩個問題，第一，當後分類資料量過大時，網頁型式的介面會呈現輸出上的瓶頸，因此我們的應對辦法為階段性的顯示方式，第一階段顯示排序最高的 20 項資料，其他則暫時不輸出，以『其他相關資料』按鈕替代，當使用者點選『其他相關資料』，再由 Post-Query XML 中讀取剩下的資料，如此將解決資料傳輸的效能問題，不會因為網頁型的限制，形成系統的使用的障礙。

第二個問題為案件資料內容不對等的問題，台中地方法院的『民事案件』資料相較於其他法院簡單，因此在後分類的顯示上，其案件資料中並無『承辦法官』此項內容，因此當後分類資訊統整得時候，『承辦法官』的統計資料將與檢索結果的文件造成資訊不對等的狀況發生，也就是含有法官資料的文件並不等於檢索結果的所有文件。因此我們利用『階層顯示』的方式來解決案件資料內容差異的問題，首先在『承辦法官』後分類資料的統計上，我們將台中地方法院的案件資料納入『未紀錄資料』的選項中，而後在後分類資料呈現上，假使檢索結果中含有台中地方法院的文件，則在『承辦法官』的後分類資料上將會以雙層的資訊索引顯示，第一層為『台中地方法院』與『其他法院』，在『台中地方法院』下顯示『未紀錄資料』的內容並且設定為無法點選，而『其他法院』下則依據原承辦法官資料的結果進行顯示。如此將可以解除使用者在

操作資料庫時產生的誤解。



《圖 4-12》後分類階層顯示架構

6. U7 – 文件檢視介面

文件檢視介面內容來自於 Result XML，為經由檢索詞組進行檢索循環後所產生的所有符合檢索條件的文件。其介面如下《圖 4-13》所示，其主要顯示



《圖 4-13》檢索檢視與篩選介面架構 – 文件檢視

文件的 Metadata，以及標示出符合檢索條件的地方，使用者可以點選文件的標

題檢視案件的影像資料，而介面也提供 **CheckBox** 與『文件刪除』選項，如圖
圖紅色箭頭所示，讓使用者進行資料的篩選。使用者進行的文件篩選動作，也
將成為檢索詞組的內容，其所對應的 **AJAX** 函數為 **QueryTermAdd**。

Function :

1. **QueryTermAdd (FieldName , Query){ AddTag(FieldName, Query) }**

使用者勾選文件後將啟動此函數，預設 **FieldName** 為 **ID**，**Query** 為文件
id，並執行 A8 中的 **AddTag** 程序。



第五章 系統簡介、實例分析與討論

本章我們將以實例，實際的操作『日治法院檔案』系統的循環檢索架構，並檢驗本系統是否將本論文第一章所提及一般檢索系統的四項不足。

5.1 『辜顯榮』研究範例

The screenshot shows the '日治時期法院檔案資料庫' (Taiwan Court Archives, 1895-1945) search results for the keyword '辜顯榮'. The interface includes a search bar, filters, and a list of search results. The results are organized into sections based on court type and case number.

所屬類別	相關人員	結案年代
新竹地方法院	辜顯榮	大正8年
民事判決原本	有限責任龍江信用組合	大正8年
臺北地方法院	株式會社臺灣銀行	大正4年
民事判決原本	臺灣銀行	大正4年
公正證書原本	李祿星	大正4年
嘉義地方法院	高砂鑄造株式會社	大正4年
合民判決原本	株式會社臺灣銀行	大正4年
公正證書原本	王慶忠	大正4年
	辜斌甫	大正4年
	三十四銀行	大正4年
	張沙塗	大正4年
	林和尚	大正4年
	林英	大正4年
	株式會社三十四銀行	大正4年
	株式會社日本勸業銀行	大正4年
	葉松白	大正4年
	蔭均堂	大正4年

Search Results:

- 大正8年判決原本第2冊 - 第201-456號
新竹地院 / 民事類 / 民事判決書類 / 民事判決原本
- 大正8年民第453號
判決日期: 大正08/10/08
案件內容: 不當利得及其他請求事件
關係人: 辜顯榮; 劉阿進
承辦法官: 篠崎岩太郎
- 民事判決原本第170冊 - 大正4年
臺北地院 / 民事類 / 民事裁判書類 / 民事判決原本
- 大正4年單民144號
判決日期: 大正04/09/25
案件內容: 家賃請求事件
關係人: 辜顯榮; 趙滿朝
承辦法官: 長尾景德
- 民事判決原本第178冊 - 大正4年
臺北地院 / 民事類 / 民事裁判書類 / 民事判決原本
- 大正4年單民780號
判決日期: 大正04/06/29
案件內容: 貸金請求事件
關係人: 王慶忠; 李祿星, 辜顯榮
承辦法官: 長尾景德
- 民事判決原本 - 昭和12年單民雜[一]
臺北地院 / 民事類 / 民事裁判書類 / 民事判決原本
- 昭和12年單民第773號
判決日期: 昭和12/06/21
案件內容: 貸金請求事件
關係人: 有限責任龍江信用組合(法定代理人理事辜顯榮); 周財, 周祖定, 周天賜, 周錫圭(周錫圭法定代理人後見人戶主周天賜), 周丙丁

《圖 5-1》系統操作範例 – 以關鍵字『辜顯榮』為例

如《圖 5-1》我們以『辜顯榮』作為關鍵字，欲了解台灣在日治時期與『辜家』相關的案件研究，進行查尋後我們可以藉由後分類的幫助，觀察到與『辜

顯榮』有關的案件資料在每個地方法院都存在，然而大部分都集中在『台北地方法院』，可以直覺的判斷北台灣為其勢力範圍所在。而從相關人員中我們也可以觀察到，『有限責任龍江信用組合』商社與『辜顯榮』有很強的關聯。藉由後分類的幫助，我們可以很快的掌握到檢索結果文件的概要，並瞭解到資料庫中對於使用者的檢索目標擁有多少資源。

從結案年代的後分類我們可以觀察到，每一年幾乎都有與其相關的案子結案，可以想見其在當時的社經地位存在相當的份量。在資料庫所收納資料的未幾年案件似乎有增加的趨勢，因此我們選擇『昭和 7-11 年』的資料作為後續的觀察。

The screenshot shows the '日治時期法院檔案資料庫' (Taiwan Court Archives, 1895-1945) website. The search results are filtered by the name '辜顯榮' (Gu Xianrong). The results list several cases, including:

- Case 1: 昭和6年單民第2484號 (昭0702/01), 判決日期: 昭和07/02/01. 案件內容: 為替手形金請求事件. 關係人: 有限責任龍江信用組合(理事 辜顯榮); 吳金土. 承辦法官: 高柳愿.
- Case 2: 昭和7年合民第319號 (昭0805/01), 判決日期: 昭和08/05/01. 案件內容: 貸金請求事件. 關係人: 有限責任龍江信用組合(理事 辜顯榮); 劉溪. 承辦法官: 讚井健, 小山隼太, 戶塚敬造.
- Case 3: 昭和7年單民第475號 (昭0707/05), 判決日期: 昭和07/07/05. 案件內容: 約束手形金請求事件. 關係人: 有限責任龍江信用組合(理事 辜顯榮); 王文松, 王連等. 承辦法官: 石井精一.

《圖 5-2》系統操作範例 – 年代遮蔽

我們選擇了後四年，透過後分類的案件事由，我們觀察到這四個年份結案的案件絕大多數都是『貸金請求事件』，而當我們將剛剛選擇的四個年代暫

時遮蔽起來時我們發現，全部與『辜顯榮』有關的『貸金請求事件』集中在我們剛剛所選的這四個年代當中，佔了 25/29，這是一種蠻有趣的現象。透過檢索詞組的操作，我們將不再需要爲了變更檢索詞彙而重新進行查詢的動作，想要比對查詢時，利用檢索詞彙遮蔽的功能，將可以保留原本的檢索詞組進行差異的分析。

【檢索組合】
x 1. 關鍵字 = 辜顯榮 On
x 2. 結案年 = 昭和7年-Off
x 2. 結案年 = 昭和8年-Off
x 2. 結案年 = 昭和9年-Off
x 2. 結案年 = 昭和11年-Off
x 3. 事由 = 貸金請求事件 On
2008/06/24, 2009 Save

【詮釋資料】
所屬類別 相關人員 結案年代
相關事由 承辦法官

- 相關事由

- 為替手形金請求事件 [5]
- 契約解除 [4]
- 賣掛代金請求事件 [4]
- 家賃請求事件 [3]
- 約束手形金請求事件 [3]
- 胎權設定金錢貸借 [3]
- 動產賣買及貸借借契約 [2]
- 官有原野拂下許可權條... [2]
- 家產贈與契約並分戶承... [2]
- 年賦償還金貸借契約 [2]
- 根抵當 / 胎權設定契約 [2]
- 根抵當 / 胎權設定契約 [2]
- 根抵當權設定契約 [2]
- 胎權一部拋棄 [2]
- 辨濟金受領 [2]
- 開墾土地賣買預約契約 [2]
- 債務承認并二辨濟二關... [1]
- 家屋賃貸借契約 [1]
- 抵當權設定債務更改契... [1]
- 造作物品賣買 [1]

其他相關案由(45)
Increase Exclude

縮小範圍 查詢 回到檢索頁面 / 我的書目清單 / 我的檢索歷史

共 10 頁, 91 筆資料 | 1 2 3 4 5 6 7 8 9 10

[大正8年判決原本第2冊 - 第201-456號](#)
新竹地院 / 民事類 / 民事判決書類 / 民事判決原本

x 1 [大正8年民第453號](#) 判決日期: 大正08/10/08
案件內容: 不當利得及其他請求事件
關係人: 辜顯榮; 劉阿進
承辦法官: 篠崎岩太郎

[民事判決原本第170冊 - 大正4年](#)
臺北地院 / 民事類 / 民事裁判書類 / 民事判決原本

x 2 [大正4年單民144號](#) 判決日期: 大正04/09/25
案件內容: 家賃請求事件
關係人: 辜顯榮; 趙滿朝
承辦法官: 長尾景德

[民事判決原本 - 大正6年第106-256號](#)
臺北地院 / 民事類 / 民事裁判書類 / 民事判決原本

x 3 [大正6年民第128號](#) 判決日期: 大正06/02/09
案件內容: 家賃請求事件
關係人: 辜顯榮; 李錫揚
承辦法官: 中上友三郎

[民事判決原本第224冊 - 大正7年第701-875號](#)
臺北地院 / 民事類 / 民事裁判書類 / 民事判決原本

x 4 [大正7年民第820號](#) 判決日期: 大正07/06/10
案件內容: 養子離縁請求事件
關係人: 辜顯榮; 辜皆慶
承辦法官: 安保忠毅

《圖 5-3》系統操作範例 – 條件排除

除了『貸金請求事件』外，我們還想要了解與『辜顯榮』相關的其他案件時，我們可以將欲排除的條件以『Exclude』的方式加入檢索詞組當中，如同《圖 5-3》箭頭所示，排除的條件會以『！=』的型式出現在檢索詞組當中，由此我們可以看出除了金錢借貸的案件外，其在其他方面的相關案件屬於哪些類型，並由此將可以進行更深入的研究與探討。

當我們將檢索條件設定好了之後，我們可以利用文件檢視瀏覽的介面，將不符合需求的文件排除在外，可以利用『刪去』的方式將不需要的文件加入檢索組合的作為條件的排除，如下圖《圖 5-4》箭頭所示之處，點選後將會在檢索組合

The screenshot shows the '日治時期法院檔案資料庫' (Taiwan Court Archives, 1895-1945) interface. On the left, the search criteria are listed under '【檢索組合】' (Search Combination):

- x 1. 關鍵字 = 辜顯榮 On
- x 2. 類別 = 臺北地方法院 On
- x 3. 類別 != 公正證書原本 On
- x 4. 關係人 = 有限責任龍江信用組合 On
- x 4. 關係人 = 辜顯榮 On
- x 4. 關係人 = 有限責任龍江信用組合.. On

Below these, it indicates '> 排除案件：4件 +' (Exclude cases: 4 cases +) and a timestamp '2008/06/24, 22:46' with a 'Save' button. Under '【詮釋資料】' (Metadata), there are tabs for '所屬類別' (Category), '相關人員' (Related Personnel), and '結案年代' (Case Year). The '所屬類別' section shows '臺北地方法院' [57] and '民事判決原本' [57].

The main search results area shows '共 6 頁, 57 筆資料 | 1 2 3 4 5 6' (Total 6 pages, 57 records). Three results are visible:

- 民事判決原本第176冊 - 大正4年**
臺北地院 / 民事類 / 民事裁判書類 / 民事判決原本
x 1 **大正4年單民780號**
案件內容: 貸金請求事件
關係人: 王慶超; 李祿星, 辜顯榮
承辦法官: 長尾景德
- 民事判決原本第228冊 - 大正7年第1452-1681號**
臺北地院 / 民事類 / 民事裁判書類 / 民事判決原本
x 2 **大正7年民第1467號**
案件內容: 家屋明渡家賃請求事件
關係人: 辜顯榮; 荒木武藏
承辦法官: 高田富藏
- 民事判決原本第258冊 - 大正9年單民第1391-1517號**
臺北地院 / 民事類 / 民事裁判書類 / 民事判決原本
x 3 **大正9年單民第1494號**
案件內容: 土地境界確認等請求事件
關係人: 和田富先; 辜顯榮
承辦法官: 高橋服膺

《圖 5-4》系統操作範例 – 文件篩選

介面中出現『排除案件：n 件 +』的條件，並將排除的案件以隱藏的方式列在其下方，點選開來後即如右圖《圖 5-5》的顯示樣式，會將案件以其標題的方式加入排除條件當中，使用者可依據刪除條件的方式將被排除的文件回復到檢索結果當中。當使用者篩選完畢後，可以利用儲存的功能將檢索條給予名稱並儲存起來，預設值為檢索的日期與時間，作為未來再次使用系統時可以作為文件集的索引。

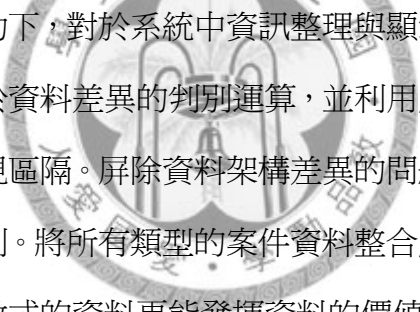
The screenshot shows the '排除案件：4件 -' (Exclude cases: 4 cases -) list:

- x 5. 案件 != 大正4年單民144號 On
- x 5. 案件 != 昭和12年單民第77.. On
- x 5. 案件 != 大正6年民第128號 On
- x 5. 案件 != 大正7年民第820號 On

第六章 結論與未來工作

6.1 結論

在整個『日治法院檔案』系統的製作中，我們發現到各法院的檔案中，案件資料內容格式不一致的問題，以及當我使用一般檢索系統的模式製作出的資訊檢索系統，因資料量龐大而發生使用上的不便。因此我們引入了屬性標籤的資訊架構，以統一『日治法院檔案』案件資料儲存格式，並將後分類的資訊索引方式帶入系統的架構中，以及在最後加入檢索詞組操作介面，而使得檢索的流程形成一種循環的模式。我們運用這個檢索循環的模式為『日治法院檔案』製作出擁有新檢索介面的系統，以解決一般檢索系統中的檢索的缺點。



在標籤架構的幫助下，對於系統中資訊整理與顯示有明顯的幫助，在統整之架構下省去系統對於資料差異的判別運算，並利用屬性的特性同中取異，將資料差異之處由此顯現區隔。屏除資料架構差異的問題後，對於後分類資訊統整運算也相形更為便利。將所有類型的案件資料整合成一個完整的資料庫，整合共享的資料將比分散式的資料更能發揮資料的價值。

當資訊檢索系統中加入了檢索詞組操作與後分類的資訊整合架構後，我們將檢所系統的檢索與瀏覽兩個部份合而為一，提供了一個更直覺並有效率檢索介面，從後分類的資訊索引幫助下，我們可以很快的掌握檢索的結果，當下利用檢索詞組的操作進行檢索組合的修正，透過循環檢索的模式，整個檢索流程將不再中斷，循環不息。而在使用者的檢索的歷程中，藉由系統的幫助下，檢索將以選擇為主，取代傳統的檢索輸入，使用者不必再困擾於如何使用檢索詞彙，而將更容易依據自己的檢索策略一步步的完成檢索目標。

6.2 未來工作

在『日治法院檔案』的資料內容當中，紀錄最多的資訊為案件的相關人物，如原告與被告、法官、書記官、原被告的代理人等資訊，而目前的系統中對於這些資訊的處理僅止於搜尋與統計，當我們將這些案件關聯人的關係一併納入系統的運算考量當中，勢必將可以提供更深層資訊意義給予使用者，而這樣的運算勢必將花費大量的系統效能，我們可以利用外部運算的方式與搜尋系統切割作為實做時的考量方案。

目前『日治法院檔案』的系統內主要的功能為搜尋與瀏覽，然而典藏檢索系統中除了檢索外，仍需要許多的輔助工具配合，例如歷史文件需要相關的背景知識與工具，如『年代轉換』、『官職架構』等等，應以工具的形式加入系統當中。本系統在介面的設計上使用模組化的架構，應此經過版面配置的設計與調整並配合 AJAX 外部運算的方式，將可以很容易的加入輔助的工具於系統當中。其他如『詞頻分析』、『詞會關聯交集集運算』，以及『中日文詞彙用語參照表』等輔助工具的建置，對於進階的文獻研究將提供許多的幫助，也視為未來系統應努力目標。

『日治法院檔案』系統中除了文字型態的 Metadata 資料外，還包含了大量的影像檔案，這些檔案在系統的製作當中僅用於顯示的用途，假使利用 web2.0 的概念，讓使用者在瀏覽影像的時候可以順便為系統建置影像相關的資料的內容，或是在影像上製作標注或連結，對於系統資料修正與內容擴張將有一定的幫助，也是未來系統可以嘗試的方向。

參考文獻

英文文獻

[Chen07] Szu-Pei Chen, Jieh Hsiang, Hsieh-Chang Tu, Micha Wu: On Building a Full-Text Digital Library of Historical Documents. Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers: 10th International Conference on Asian Digital Libraries, ICADL 2007, Hanoi, Vietnam, December 10-13, 2007, Proceedings. (Lecture Notes in Computer Science). Springer 2007. 49-60.

[Wang00] Tay-sheng Wang, Legal Reform in Taiwan under Japanese Colonial Rule, 1895-1945: The Reception of Western Law, Seattle: University of Washington Press, 2000. pp. 63-64.

[Jonathan06] Jonathan Stuart-Moore, Monte Evans, Patricia Jacobs , Interface design for browsing faceted metadata , Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries , 2006.

[Tu99] Hsieh-Chang Tu , Interactive Web IR : Focalization Model, Effectiveness Measure, and Experiment. , PH.D Dissertation , National Taiwan University,1999.

[Elmasri00] Elmasri,Navethe, “ Fundamental of Database System 3rd edition”, 2000.

中文文獻

[王 02] 王泰升, 關於台灣總督府法院文書目錄編纂, 「台灣的近代與日本」國際學術研討會, 名古屋: 中京大學社會科學研究所主辦, 2002年10月26-27日。

[莊 06] 莊萬慶, 標籤樹於文件檢索後分類與呈現之運用 – 以古文書為例, 碩士論文, 國立臺灣大學, 2006

[陳 01] 陳光華, 莊雅棻, 資訊檢索之中文詞彙擴展, 資訊傳播與圖書館學 8卷1期 : 59-75, 2001

英文網站

[Amazon] Amazon.com , <<http://www.amazon.com/>>

[Google Suggest] Google Suggest Labs ,

< <http://www.google.com/webhp?complete=1>>

[W3C] W3C , <<http://www.w3.org/>>

[Citation Pearl Growing] Citation Pearl Growing ,

< <http://newadonis.creighton.edu/HSL/searching/PearlGrowing.html> >

[Faceted Classification] Use of Faceted Classification ,

< <http://www.webdesignpractices.com/navigation/facets.html>>

中文網站

[王 07] 王泰升,日治法院檔案的保存與利用 ,

<http://www.law.ntu.edu.tw/chinese/03/professor/Tay_sheng_Wang.html>

[國家文化資料庫] 國家文化資料庫 ,

< <http://nrch.cca.gov.tw/ccahome/about.jsp>>

