

第六章

全國檔案知識管理系統之研究

The Research of National Archives Knowledge Management System

楊新章

Hsin-Chang Yang

長榮大學資訊管理系 助理教授

Associate Professor,

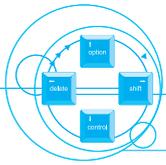
Department of Information Management,

Chang Jung University

壹、前言

一、背景

2001年全球各國政府資訊化的評比，台灣名列第二，僅次於美國；2002年台灣名列第一。這項評比結果令從事資訊工程與圖書資訊研究的學者，一則以喜，一則以憂，台灣各級政府資訊化的程度，真是如此的名列前茅嗎？台北市政府自1999年起，已舉辦五次「網路新都金像獎 — 市府各單位資訊網站評比」，行政院研考會也制訂「政府網站評鑑指標」，並進行政府各機構網站的評估，其目的不外乎，藉由評鑑各級政府機構網站，提升公務人員之資訊素養、加速政府資訊之數位化、提供及時的政府資訊、建置便民的資訊管道、發展有效的檢索機制、



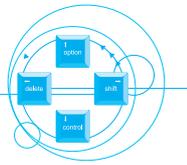
建構統一的資訊平台。

然而，若是檢視目前各級政府的資訊網站，可以發現目前政府的資訊網站仍處於提供原始資料的階段，提供有限的全文檢索功能，或是著重於各部門公文線上簽核、管理等功能。各機構將比較多的心力用於版面構成、視覺呈現，這部分雖然是極為重要的一環，對於資料的瀏覽與取用有很大的幫助，但是對於龐大的政府資訊，如何由縱向取得各時期相關的資訊，或是橫向瀏覽各機構不同卻相關的資訊，如何有效地組織整理政府資訊、如何處理老舊的政府資訊（**Legacy Government Information**）、如何有效地描述政府資訊，就目前的階段而言，並沒有滿足上述的各項需求。

「知識經濟」與「知識管理」已然成爲二十一世紀最熱門的話題，企業組織視之爲成長轉型的靈丹妙藥，國家也視其爲提振國家競爭力的必要手段。事實上，彼得·杜拉克（**Peter Drucker**）早在1965年就曾經提到，知識將取代機器設備、資金、原料或勞工，成爲企業經營最重要的生產要素，而世界經濟的主導權，也將落在那些主導知識經濟的知識工作者身上。（杜拉克，2000，頁6）彼得·杜拉克同時強調，未來是知識經濟的時代，知識生產者是組織和社會最重要的資產，知識生產者的生產力是組織利基的指標。如何將資訊轉而爲知識，從而爲大眾所用，便是知識經濟時代，重要且嚴肅的課題。在推動知識管理的大潮流之下，部門亦無法置身事外；台灣地區各級政府單位推動知識管理的行動亦方興未艾，有的是致力於知識管理觀念的導入，有的是實際落實在「知識管理系統」的建置與使用，企圖透過知識管理的落實，創造更有效率的政府施政。

但每個不同政府單位在執行知識管理的行動方向，亦應依其不同的業務屬性和單位特性而有所不同，在知識管理方案乃至於系統的選用與發展亦有所差異；以知識管理落實在政府資訊與檔案管理上，有必要先就檔案資訊的特性加以分析，方能選擇適當的知識管理解決方案。

檔案是提供不同機關間業務聯繫與經驗交流，以及民眾瞭解與認識政府的絕佳管道。國家檔案籌備處於民國89年3月1日成立後，即積極進行全國檔案資訊系統與機關檔案管理系統的建置工作，其中「機關檔案管理資訊網」（<http://online.archives.gov.tw>）收錄全國各機關上載的檔案目錄資料。爲了進一步讓一般民眾使用這些資料，檔案管理局亦建立「全國檔案目錄查詢網」（**National Electronic Archives Retrieve**，簡稱 **NEAR**，<http://near.archives.gov.tw>）。



gov.tw)，提供簡易查詢與進階查詢，進階查詢提供機關名稱、案件流水號、檔號、案由、案名、主要發/來文者、文件產生日期、媒體型式、保存狀況、附件名稱、涉及之事項主題、涉及之人名、涉及之地點地名、涉及之相關時間、收文者、發文字號、收文字號等查詢點。從上述的分析可以發現，NEAR提供的功能，仍屬一般網站的檢索功能，如何在這樣的平台上，針對政府資訊的特性，發展適合的知識管理相關技術，是未來e-government重要的課題，也是提昇國家競爭力的重要指標。

在此，我們參酌了國內外相關針對政府資訊與檔案利用的相關研究，提出以資料探勘（Data Mining）、文本探勘（Text Mining）或知識探索（Knowledge Discovery）為本的知識管理系統雛形，以供檔案管理局進行知識管理落實於檔案利用的第一步。

二、研究之目的

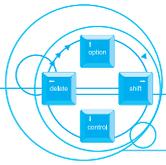
本研究之目的乃基於下列的原因：

- (一) 全國檔案目錄資料庫資料龐大，如能經由知識探索技術，針對不同類型機關使用者需求提供專業知識，形成檔案知識庫雛形，進而提供決策支援，形成檔案相關業務推動策略。
- (二) 由檔案管理邁向知識管理為檔案管理發展的必然趨勢，現有系統已有許多相當珍貴的檔案資料，缺乏更有效的系統機制，發揮檔案內含之價值。

由於檔案目錄資料相當龐大，本研究的研究方向將定為擷取全國檔案目錄資料適用的資訊、擷取檔案目錄之相關資訊形成檔案知識、規劃檔案知識管理系統可能應用模式。

三、研究之範圍與限制

一個完整的檔案知識管理系統，必須結合組織內部的資訊系統，並至少包含三大模組：使用者管理模組、知識分享模組、知識探索模組。然而，由於市場上知識管理系統產品，多數都有使用者管理與知識分享的模組，這方面的討論與實作亦較為常見，因此，本研究受限於時間的因素，系統實作部分將著重於知識探索模組，發展有效探索檔案知識的演算法。



此外，受限於部分檔案資料的敏感性，本研究並不能處理檔案目錄資料所有的欄位，也無法接觸所有的檔案目錄資料，因此，在資料的完整性上有實際的缺憾。

四、研究方法與步驟

本研究擬使用之研究方法如下所示。運用系統分析法分析實際的政府資訊系統與NEAR網站，再將以系統實作法，建構配合全國檔案資訊系統的知識管理雛形系統，並撰寫能夠輕易整合於NEAR網站的前端介面（Front End），以提供使用者有效使用檔案資訊與知識。

(一)系統分析法

調查國內外相關的資訊系統與NEAR網站，分析其瀏覽以及檢索的模式，並探討是否有實際運用自然語言、資料探勘、資訊檢索、資訊擷取等技術。分析的結果將作為本研究實作的參考。

(二)系統實作法

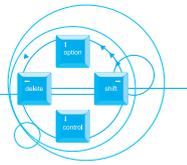
截至民國92年2月全國檔案資訊系統約有1千8百萬筆檔案目錄資料，也就是描述檔案的詮釋資料（Metadata），這些是全國各機關陸續上載的資料，資料的完整性不一。另外經檔案管理局同仁整合的「二二八檔案」，具有完整的目錄資料，但是檔案本身較具敏感性，且多為影像檔，手稿居多，無檔案全文資料。因此，系統實作將建構配合全國檔案資訊系統的知識管理雛形系統。

貳、全國檔案知識管理系統架構

基本上，我們將一個知識管理系統分為三大模組，第一是使用者管理模組；第二是知識分享模組；第三是知識探索模組。以下針對各功能模組做一概略描述。

一、使用者管理模組（User Management Module）

使用者為知識管理系統之主要標的，使用者為部份知識之貢獻者，亦為知識



之享用者。然知識管理並非無價，並非所有的使用者皆可享有同樣的知識，同時在知識建置之過程中，使用者也會貢獻不同形式與不同種類的知識。因此對於使用者身份、權限與職掌之管理，亦為知識管理系統重要之一環。此外，透過使用者的管理，分析使用者資訊需求與行為，亦可獲得相當的知識，可作為爾後開發新系統的參考。因此，一個知識管理系統必須記載詳系的使用者資料，在本研究中，使用者管理模組包含了下列功能：

- 會員管理（Member Management）：包括會員登錄，身份確認，權限設定，密碼設定，描述檔設定等
- 群組管理（Group Management）：包括使用者群組設定，群組關連設定，群組權限設定等
- 職掌管理（Duty Management）：包括工作職掌設定，職掌關連設定等
- 描述檔管理（Profile Management）：包括使用者描述檔管理，分析，統計與檢視等
- 日誌管理（Log Management）：分析使用者日誌。

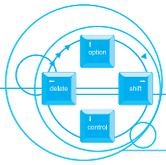
二、知識分享模組（Knowledge Sharing Module）

知識分享模組主要提供知識檢索、分享、傳播的機制，使得系統內部儲存的知識，或是使用者的腦中的經驗，可以讓其他的使用者分享。通常可以透過下列的方式達成這些目標。

（一）資訊檢索

提供使用者檢索系統內部儲存的文件與資料，可以分為兩種方式：

- 內容為本之檢索（Content-based Retrieval），涵蓋文字檢索、影像檢索、音訊檢索（含語音與樂音）、視訊檢索或多模檢索（Multi-Modal Retrieval）等。
- 詮釋資料之檢索（Metadata Retrieval），檢索描述原始物件的資料，這些資料可能人工賦予的，如編目人員進行圖書編目；亦可能是機器生成的，如應用資訊擷取（Information Extraction）的技術，自動



生成詮釋資料。

目前市場上的知識管理系統多數提供搜尋引擎的功能，可以檢索文字資料，亦有極為少數的系統提供影像檢索的功能。

(二) 知識代理人

使用推播技術 (Push Technology) 主動將檔案知識傳遞給使用者。

- 依據日期 (Current Date) 將相關檔案資料主動傳遞給使用者
- 依據職掌 (Duty) 將所藏相關檔案資料主動傳遞給使用者
- 依據使用者自訂需求將所藏相關檔案資料主動傳遞給使用者
- 依據使用者之使用紀錄描述檔 (usage profile) 將所藏相關檔案資料主動傳遞給使用者

(三) 使用者推薦

使用者可傳遞相關檔案資料給其他使用者

- 依據群組 (Group)
- 依據職掌 (Duty)

(四) 佈告欄與討論區

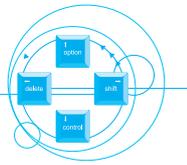
使用者可以依據不同群組，不同職掌，不同議題設定討論議題，並進行知識與資料分享。或是簡單地將個人意見或經驗發佈於佈告欄。

三、知識探索模組 (Knowledge Discovery Module)

目前市場上常見的知識管理系統，多數是知識分享資訊系統，提供討論群、FTP、搜尋引擎、訊息傳遞等功能。應用資訊技術創造有效的分享機制是知識管理系統非常重要的，但並不是唯一重要的；另一項重要的工作是，從大量的資訊中探索資訊間的關係，這卻是多數知識管理系統欠缺的功能。

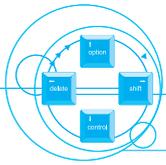
知識的探索與資料或資訊的類型密切相關，本研究探討的是檔案知識管理系統，因此我們將在這個範疇內，探討可以探索的知識類型以及所使用的技術。此模組應包含下列功能：

- (一) 聚集相同檔案之目錄資料：在檔案中有許多重複性資料 (redundant data)，同一文由同一機關發文至不同單位。此類資料佔據了一大部份的



- 檔案內容，不但造成空間浪費，也會降低資料分析之準確度，因此有必要將這類重覆性資料偵測出來。
- (二) 建構檔案資料之時間序列：當原始發文機關發文後，後續機關之處理應有其固定之模式。時間序列之建立目的在統計以便觀察公文傳遞流程。此類資訊除可用以改善公文流程外，也可以據以發掘組織間因公文傳遞所產生之關聯。
 - (三) 建立檔案資料之相互參照：檔案資料眾多繁雜，其中的關聯可以許多不同面向存在、相互參照是其中之一。藉由不同欄位間之同質性，我們可以建立檔案間之相互參照，以提供簡易之檔案檢索。
 - (四) 建立相關檔案資料之群集：檔案資料間有許多類似的資料，我們可由其間之相似性加以分群。分群之依據又可分以精確資訊分群與以模糊資訊分群兩者。精確資訊乃必須為某一欄位或某些欄位相同者為一群。模糊資訊分群則只需檔案資料間有某種相似性即可，甚至可以不用任何欄位相同。資料之群集可以提供我們相當多的知識，因此在任何知識建置與管理系統中為必備之功能。
 - (五) 抽取重要事件與人物：在眾多資料中抽取出所謂的重要資料—包括事件與人物—是非常重要的知識。除了提供事件的摘要外，它並可以讓使用者瞭解重要事件的起源，人物之間的關聯，透過人物瞭解事件的關聯，透過事件瞭解人物的關聯等眾多應用。
 - (六) 分析機關檔案分類表：統計與分析各機關所處理之檔案型態、數量、內容、其他參與機關等資訊，並建立分類。此類知識有助於瞭解機關檔案之特性，輔助各機關針對業務職掌進行調整。
 - (七) 建構動態統計共現索引典：檔案資料間經常共同出現之檔案必定有相當之關聯性，此關聯性不容易以欄位分析之方式取得。此類知識亦有助於瞭解欄位間之關聯，以提供公文內容設計與調整之用。

上述之功能，只能涵蓋部分的檔案知識管理系統的範疇，然本研究之目的為建構全國檔案知識管理系統之雛形，並探究其功能上應具備之基本組成元件，實



無法一一列舉所有可能的知識管理系統功能。知識管理並非僅限於文件管理，如何自文件中產生知識並應用該等知識於管理中，方為知識管理系統之真諦。本研究將以此為導向探討如何運用全國檔案目錄資料以進行知識探索與應用。

如前文所言，本研究將知識探索模組規劃為下列數項：

1. 聚集相同檔案之目錄資料
2. 建構檔案資料之時間序列
3. 建立檔案資料之相互參照
4. 建立相關檔案資料之群集
5. 抽取重要事件與人物
6. 分析機關檔案分類表
7. 建構動態統計共現索引典

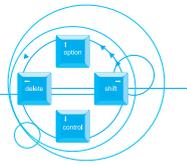
這七項功能我們再以其所需運用之核心技術加以區分並介紹如下：

- 資料分群（data clustering）－功能 1、3、4
- 主題偵測（topic detection）－功能 5
- 關聯分析（association analysis）－功能 2、5、7
- 屬性分析（attribute analysis）－功能 1、6
- 資料歸類（data classification）－功能 6

以下則對這些技術予以簡單的介紹。

一、資料分群

資料分群的目的在將原始資料依其相似度加以分成群組，在事先並無法獲知資料的類別為何，甚至無法獲知類別之數量的情況下，對資料來進行分群有其困難。但若能夠對資料進行分群，則此群組關係將可提供許多有用的資訊。本研究中許多知識探索的功能，都必需倚靠分群的結果進行。資料分群的核心在於相似度的定義，如何定義相似度將對分群結果之好壞產生影響。



二、主題偵測

主題偵測的目的在於自文件資料中偵測出重要的主題，可作為摘要或類別使用。這個功能大多應用在全文文件中，因其提供了大量的語句以供分析。本研究之資料僅為目錄資料，所能獲得之文字資料有限，因此必須輔以屬性分析的結果來完成主題偵測。

三、關聯分析

關聯分析的目的在於發現紀錄（文件）間之關聯性。在此關聯性的定義甚廣，可為某兩事件間之因果關聯，兩機關依文件順序之關聯，兩文件共現關係之關聯等等。由於關聯之呈現對使用者而言意義頗大又頗易瞭解，關聯分析可以說是知識探索必備的技巧之一。本研究將針對不同功能分析資料間之不同關聯。

四、屬性分析

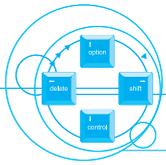
屬性分析的目的在發現屬性（欄位）間之關聯，分析無效屬性，挑選重要屬性等等。由於全國檔案目錄資料皆存在於不同欄位中，屬性分析可以對眾多的欄位加以解析，除了直接提供屬性相關統計知識外，亦有助於其他知識探索功能之執行。

五、資料歸類

資料歸類乃根據資料之內容，將其歸類於某特定類別下。根據不同類別的定義，我們可以依歸類結果，獲知有用的知識。因此在執行資料歸類時，類別的定義，以及如何進行歸類為此技術之核心。

參、系統實作與評估

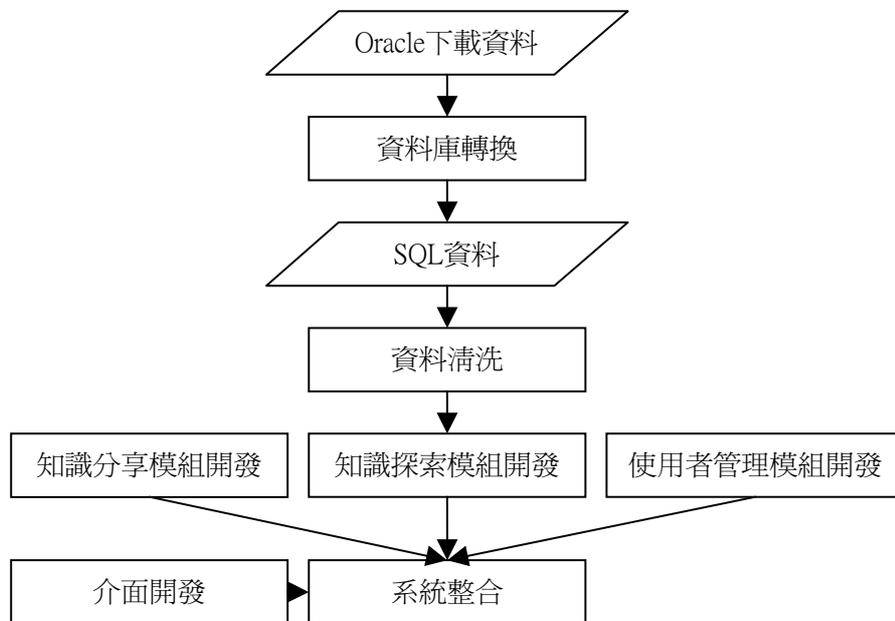
為驗證本研究之規劃與制定之系統功能之正確性與可行性，本研究開發了一雛形系統以供驗證參考。由於本研究執行期間僅十個月，限於人力與時間因素實



難以將全國檔案知識管理系統所規劃之全部功能加以完整呈現。另一方面，本研究屬研究性質，其目的在於功能之規劃與規格之制定，因此實作時僅以雛形系統方式呈現以供未來完整系統製作時參考。緣此本研究在實作時並不以功能之完整性為優先考量，而以資料之淺、深層分析為主，期能提供規劃各功能時之參考外，亦能成為未來實作各功能時之基礎。因此本章將詳述本研究實作時應涉之流程與技術，並探討其間之問題與解決方式。

一、實作流程概觀

圖一為本雛形系統之開發流程圖。



圖一 計畫雛形系統開發流程圖

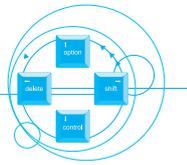
(資料來源：研究者繪製)

二、資料概觀

本研究所使用資料為全國檔案目錄資料中淬取部份欄位內容組合而成，並未包含全部欄位。所使用的欄位整理於表一。

(一)欄位簡介

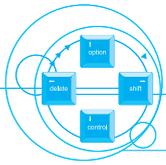
以下就各欄位於知識探索時所可能提供之資訊與扮演之角色作一介紹。



1. 機關代碼：機關代碼主要可提供為文件歸類使用，藉以統計各機關的檔案數量與型態。
2. 案件流水號：流水號為每一檔案之唯一性編號，大多僅供識別所用。
3. 年度號：提供年度資訊，可供歸類與時序分析使用。
4. 分類號：直接提供檔案內容分類資訊，可提供歸類與群集結果驗證。
5. 案次號、卷次號、目次號：提供檔案歸類參考資訊。
6. 案名、案由：以全文方式提供本記錄之摘要性資訊，可據以歸類與提供分群標籤。亦可提供文本探勘之資料來源。
7. 主要發文者、受文者：提供歸類與文件流程追蹤資訊。亦可依此進行關聯分析。
8. 發文字號、收文字號、來文字號：提供資料清洗之依據，亦可據以進行關聯分析。
9. 西文文件產生日期：提供詳細時間資訊，可供時序分析使用。
10. 本文媒體型式：可提供公文傳遞方式歸類資訊。
11. 主題名稱：提供本案主題分類資訊，可供淺層歸類參考使用。
12. 主題內容：本案主題本資訊，提供主題偵測及資料歸類使用。
13. 附件名稱：可併入主題內容等其他欄位以提供全文資訊。
14. 附註名稱、附註內容：大部份資料並無此二欄位資料，可併入其他欄位作為文本探勘資料使用。

(二)依Oracle資料型態名稱

1. 資料清洗：主要發文者、受文者、發文字號、收文字號、來文字號、附註名稱、附註內容
2. 淺層資料歸類：機關代碼、年度號、分類號、主要發文者、受文者、本文媒體型式、主題名稱
3. 深層資料歸類：案名、案由、主要發文者、受文者、主題內容
4. 資料分群：案名、案由、主要發文者、受文者、主題內容、附件名稱、附註名稱、附註內容
5. 關聯分析：年度號、分類號、主要發文者、受文者、發文字號、收文字



號、來文字號

6. 時序分析：年度號、西文文件產生日期

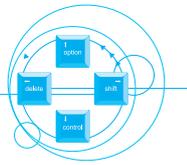
7. 主題偵測：案名、案由、主題內容、附件名稱、附註名稱、附註內容

8. 文本探勘：案名、案由、主要發文者、受文者、發文字號、收文字號、來文字號、主題內容、附件名稱、附註名稱、附註內容

依據上述之欄位簡介，以下將執行各任務時可用／需用之欄位整理如下：

表一 實作檔案目錄資料庫欄位

欄位代碼	欄位名稱	欄位儲存值格式
A01	機關代碼	CHAR(10)
A02	案件流水號	NUMBER(10)
A03	年度號	CHAR(4)
A04	分類號	CHAR(20)
A05	案次號	CHAR(12)
A06	卷次號	CHAR(4)
A07	目次號	CHAR(3)
A08	案名	VARCHAR2(100)
A09	案由	VARCHAR2(300)
A10	主要發文者	VARCHAR2(60)
A11	受文者	VARCHAR2(60)
A12	發文字號	CHAR(40)
A13	收文字號	VARCHAR2(40)
A14	來文字號	CHAR(40)
A15	西元文件產生日期	CHAR(8)
A16	本文媒體型式	CHAR(1)



A17	主題名稱	CHAR(1)
A18	主題內容	VARCHAR2(50)
A19	附件名稱	VARCHAR2(50)
A20	附註名稱	CHAR(1)
A21	附註內容	VARCHAR2(100)

資料來源：檔案管理局

(三) 資料清洗

在實作時我們針對所得資料觀察而得到下列結果：

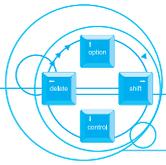
- 1.許多欄位皆以數字表示，所能提供資訊不夠清楚直接。
- 2.部份欄位填列不夠詳細或不盡合理，例如主題內容有填列發文日期者，有填列部份案由者，致喪失了應具功能。
- 3.重覆性資料過多，同一公文因收發文機關不同會出現在不同記錄中，產生重覆性資料。

依據初步資料分析結果，資料漏失現象並不嚴重，部份未填列欄位明顯為原先未有資料者。惟空白欄位究竟代表著原來就不具該欄位資料者或漏填，應作一區分，例如可填列一固定標記以標示此欄位原來就未具資料。如此可以很輕易的偵測出漏失值。

在雜訊處理上，由於欄位資料大都為文字，因此並未發生太大問題。可能會發生雜訊問題的欄位有案次號、卷次號、目次號、本文媒體型式、主題名稱、附註名稱等，因其皆為數字或字元型態之編號。這類欄位因資料簡單且固定，我們可以結合電腦與人工檢視方式很容易的偵測出雜訊所在並以合理值修正之。

(四) 資料歸類

在資料歸類的應用上，首先我們先匯集相同檔案目錄資料。因為檔案目錄資料的特性，由一機關所發文之資料，也會出現在收文機關的記錄中，因而這些記錄將會擁有大部份相同的欄位值，例如案名與案由等。因此我們可以將這些記錄視為同一筆而降低了資料筆數。其作法為檢視發文字號與收文字號或來文字號間是否有相同者，若一記錄之收文字號或來文字號與發文字號相同，則可視為同一



筆記錄予以歸類。另一方面，我們也可將這些重覆性資料加以清除，降低後續相關處理的成本，例如資料分群。但是我們要注意是某些任務（例如時序分析）需要使用完整的原始資料，不可採取刪除重覆性資料步驟。

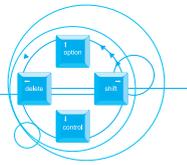
由於檔案目錄資料中已有部份欄位具有類別資訊，因此我們可以利用其作淺層歸類。例如機關代碼、年度號、分類號、案名、主題名稱等，可據以進行單欄位或多欄位歸類。要使用這些欄位進行歸類時，首先必須整理各欄位之可能值以作為類別標籤使用，而後再據以進行分類。

除了依上述欄位內容進行淺層歸類外，我們也可依部份欄位，例如案由、主題內容、附件名稱、附註內容等全文欄位，進行深層歸類。在此深層歸類的定義為使用不精確的比對方式來進行歸類，亦即我們不要求欄位值要完全一樣或部份一樣才能被歸於一類。要達成此種歸類任務，必須具備兩個要素：其一是必須定義類別標籤；其二是必須定義歸類的標準，即欄位值間相似度的定義。以下說明這兩個要素的定義方法。

- 類別標籤定義：由於深層歸類並不直接使用欄位內容作為標籤，因此我們要先自欄位內容間定義出可用之標籤。在此標籤又分為兩類，其一為單一字詞標籤；其二為全欄位值標籤。單一字詞標籤為出現在欄位內容中之某一特定字詞，其偵測方法可以統計方式為之，統計該欄位中每一字詞出現的次數，再據以定義標籤。在此或許必須輔以人工檢視以得到更合理的結果。全欄位標籤則是統計該欄位中出現最多的欄位值，再據以設定標籤。單一字詞標籤比全欄位值標籤更具深層歸類效果，但後者較易設定，也較易理解。
- 相似度定義：欄位間之相似度，我們可以採用向量空間模式（vector space model）中兩文件間之相似度計算方法，亦即將兩欄位值以向量方式表達，再計算向量間之相似度即可。如此相似度超過門檻值者方為同一類別。上述方法適用於單一字詞標籤及全欄位值標籤。

(五) 資料分群

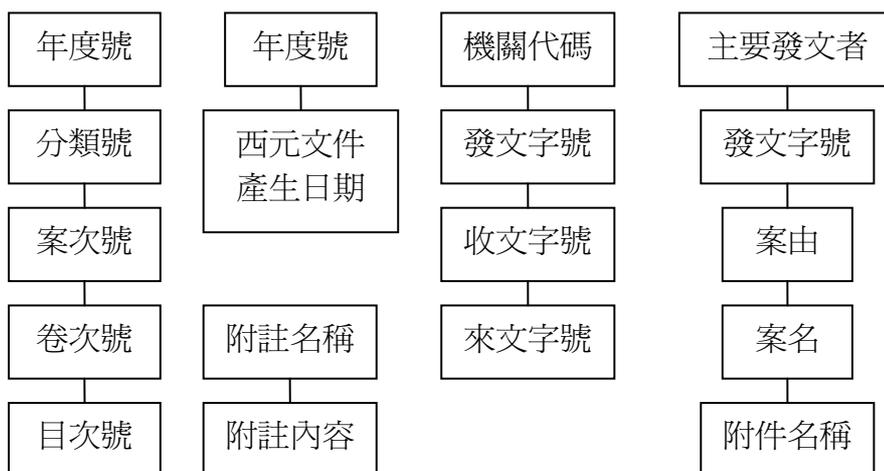
本研究在資料分群實作上可分成兩種方法進行，其一是利用數值型欄位進行分群；其二是利用文本型欄位來進行分群。簡述如下：



- 利用數值型欄位進行分群：此類欄位包括案件流水號、年度號、分類號、卷次號、目次號、西元文件產生日期、本文媒體型式、主題名稱、附註名稱等。使用這幾個欄位值利用上述分割法或階層法便可進行分群。
- 利用文本型欄位進行分群：此類欄位包括案名、案由、主要發文者、受文者、發文字號、收文字號、來文字號、主題內容、附件名稱、附註內容等。將這些欄位的文本內容聚集，並分離其中之字詞，再依向量空間模式方式對應至一向量，便可採取類神經網路方式進行分群。

(六) 重要欄位偵測

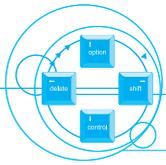
在實作上重要欄位偵測便是採取上述之屬性導向推論法進行。我們必需先行統計各欄位之不同值數量，再據以進行屬性移除或屬性一般化。但在此之前，我們必需先定義各欄位間之概念階層，其中部份定義如圖。



圖二 各欄位間之概念階層

(資料來源：研究者繪製)

圖二所定義之位概念階層，可用於屬性一般化中。例如當我們評估目次號可以進行一般化後，我們便可依此階層將其一般化至卷次號。但如果我們評估附註名稱應一般化時，因其不具更高之概念階層，因此該欄位便無法進行一般化。

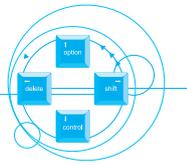


肆、研究發現與結論

本研究之目的在於針對國家檔案目錄資料研究一個可行之知識管理架構。在此架構下我們希望可以能夠提供一個完整介面以幫助使用者發掘知識、累積知識、探討知識與分享知識。因此本研究之研究認為一個適切之知識管理系統應包含下列功能元件：

- 知識探索功能：知識管理之基本應為知識之創造與發掘。在國家檔案目錄資料中儲存了來自各機關針對公務所發出之公文目錄，其中涵蓋了國家各類政策之發起、執行與結果資訊，亦隱含了各種寶貴的知識。如各機關政策走向、主要業務內容之轉變等等。然而要自其中整理或發掘出這類隱含知識，其實並非易事。因此，必須運用各項資料探勘與知識探索技術方能自資料間發掘出各項知識。然而知識之型態甚多，本研究所建議之功能為其中較適用於國家檔案目錄資料者。此外，因為目前的知識管理系統多未具完善的知識探索功能，因此希望本研究可成為未來知識管理系統開發的參考案例。
- 知識分享功能：上述之知識探索功能以資料為主產生的，所能產生的知識限於資料內容。然而，知識之運用者為人類本身，使用者獲得知識之管道不應受限於此單一管道，因此一知識管理系統也應具備其他創造知識的管道。本研究中規劃了知識分享功能，便是希冀透過使用者間之互動與分享創造知識，使知識之運用更加靈活，更能遂行知識管理之任務。
- 適切的介面：使用者介面為使用者與系統之直接接觸的第一線，介面的好壞會直接影響使用者認定系統之良窳。另一方面，雖然本研究認為知識創造與發掘為知識管理最重要的部份，但知識之呈現方式亦可影響知識之品質。因此一良好的介面亦為發展知識管理系統時必須慎重考量的一環。

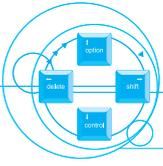
事實上，上述各項功能，主要的核心技術還是在知識探索這部份，因其牽涉



到較具難度之資料探勘、圖型識別、資料分析、統計、資料庫管理等理論與技術，而非一般網頁設計師或軟體工程師所能勝任。因此本系統實際在開發時應考量開發人員之技術背景，最好有堅強的資料探勘或知識探索技術人員，否則僅能提供淺層知識，不能完全發揮本系統應具之功能並達成應具的任務。

參考文獻

杜拉克等著（2000），知識管理（Knowledge Management）（張玉文譯）。台北市：天下遠見。



檔案資訊資源管理
