以文件關聯性為基礎之企業知識客服管理模式

侯建良、楊綠淵
清華大學工業工程與工程管理學系

摘要

由於網際網路技術發達，使用者透過資訊網路取得資訊、進行交易之頻率已顯著提升。為使企業之知識文件、銷售資訊能有效而正確地提供予潛在需求對象，以實現文件保密或一對一行銷之理念，本論文乃以文件關聯性為基礎發展企業知識分群法則，並配合使用者之閱讀趨勢（包括瀏覽網頁或閱讀電子文件），作為掌握使用者偏好趨向與企業知識分群管理之依據。此模式之重點精神乃首先建構以關鍵字為基礎之文件關聯性分析模式；再以此模式為基礎，發展知識文件分群法則；最後根據客戶過去瀏覽之文件，得知其閱讀權限範圍或趨勢。若目標文件（如欲發佈之領域知識或廣告）與客戶過去閱讀文章之關聯性高或隸屬同一分群，則可推斷該客戶為有權限或有意願閱讀此目標文件之對象。藉由本研究所發展之推論模式，除可促成企業體實現一對一行銷之理念外，尚可應用於企業知識文件管理體系，協助企業組織發展智慧型知識文件管理機制，使電子化知識管理與顧客關係管理理念能相互整合支援，並帶動知識服務型產業之發展。

關鍵字：知識管理、顧客關係管理、分群、權限管理、關聯性分析

An Integrated Knowledge Management and Service Model Based on Document Correlation Analysis

Jiang-Liang Hou, Lu-Yuan Yang
Industrial Engineering and Engineering Management, National Tsing Hua University

Abstract

Owing to the popularity of the information technologies, more and more business transactions and services are conducted over the Internet. To provide the advertisements or documents to potential customers, one of the typical issues is to accurately acquire the demands of customers. Based on the document correlation and browse history of Internet users, this research develops a three-phase reasoning model for integrated knowledge management and customer relationship management. The model consists of three decision algorithms namely document correlation analysis, document clustering and document authority determination. According to the keyword distribution analysis, correlation of the target document and the historical documents of the specified user can be derived and the probability for each user to access the target document can also be obtained. A prototype system as well as a demonstration case is provided to evaluate the performance of the proposed model. The attempt of this paper is to provide an applicable and intelligent approach to improve the operation efficiency of enterprise knowledge services.

Keywords: Knowledge Management, Customer Relationship Management, Clustering, Security Management, Correlation Analysis.

壹、前言

隨著電子商務技術之蓬勃發展，使用者之交易模式已漸由傳統店面交易轉化為網路交易，網際網路市場已成為各企業必爭之地。是故，企業經營者或決策者必須了解網路消費生態，才能保有企業之競爭力。伴隨網際網路應用之快速成長，大量資訊已可不受時空限制快速被取得，且市場產品、競爭對手、使用者等相關資訊或知識已不斷更新，令企業組織應接不暇。各企業可利用資料挖礦（Data Mining）或網頁挖礦（Web Mining）等技術，從龐大客群資訊中分析線上客戶或瀏覽者之資料，掌握顧客的閱讀偏好趨勢，使企業體能更了解使用者的潛在性需求，並將之運用於行銷活動上，使企業營運更能因應使用者之需求差異性，提供個別性之資訊與服務。

另一方面，在知識經濟時代下，企業體為充分保有核心競爭力，必須有效掌握產業之領域知識與智慧財產，以資訊技術協助創研中心與知識中心成立已成為企業共同發展趨勢。知識文件為企業溝通、交易活動之核心，企業員工平均花費60%的工作時間於文件處理；故利用先進儲存技術、計算速度與系統整合技術，以改善企業文件處理效能；亦即利用電子化文件管理獲致企業利益（Meier & Sprague 1996；Tyrvainen & Paivarinta 1999）。然各項資訊、文件與知識之存取權限與使用者之業務特性、專業背景、閱讀趨勢等因素相關，企業體可利用資料探勘技術進行知識文件分群與關聯分析，進而搭配客戶之閱讀趨勢，決策知識文件之授權對象，以減輕知識工程師之作業負荷，並實現企業知識管理系統自我服務之能力。

有鑑於上述需求，本研究發展一套以使用者閱讀趨勢為基礎之知識文件管理模式。此模式為一套三階段推論模式，包括「文件關聯性分析」、「文件分群推論」及「文件授權客戶推論」三大研究課題。文件關聯性分析模組乃利用文件關鍵字搜尋之結果，分析文件間之相關性，並將分析之結果運用於後續之文件分群與文件授權對象推論等應用；除此之外，此文件關聯性分析之結果尚可應用於文件分類檢索、文件模糊搜尋等任務。整體而言，藉由本研究所發展之模式與技術，期能提供企業之行銷決策者或知識工程師一初始決策方案，作為其進行一對一行銷計劃或知識中心管理之依據，提升企業顧客關係管理（Customer Relationship Management；CRM）與知識管理（Knowledge Management；KM）之整合性效能。長程而言，此技術之發展期能提供知識提供業者（如顧問公司）或輔導服務機構（如法人機構）一套有效率服務其客戶之利器，帶動國內知識服務型產業之發展。

貳、文獻回顧

根據知識文件特徵進行分群與分類，可有效提升知識文件之再利用率；另一方面，藉由分析使用者閱讀趨勢，可應用於分析商品行銷對象或文件權限對象。大體而言，相關研究課題乃包括使用者閱讀趨勢資料之收集與探勘、文件分群/分類及文件權限推論；

故以下乃針對此些課題之研究成果加以回顧。

一、使用者閱讀趨勢資料之收集與探勘

透過網頁伺服器或者是附於 HTML 內之控制碼，可取得每位使用者瀏覽網頁時所留下之紀錄，此些紀錄可應用於分析使用者的喜好或特殊興趣，而從瀏覽網頁記錄所得之使用者行為特徵即可作為個人化服務之依據；但由於網站具有匿名瀏覽之特性，導致使用者之瀏覽紀錄分析有所困難。一般而言，瀏覽紀錄之形式大致分為三種：網頁伺服器瀏覽日誌檔（Log File）、網頁轉換與代理人系統三類（陳佳鴻 2001；卜小蝶 2002）。網頁伺服器瀏覽日誌檔為 WWW 中網站與使用者間溝通之網頁伺服器所自動產生之紀錄檔，此種記錄方法之缺點在於無法確定特定身份之使用者，而且對於動態內容之互動式網頁有分辨之困難（蔡聰洲 2001；何昶毅 2001）。網頁轉換方式乃使用者進入系統前，網頁伺服器會暫時將執行權交予紀錄伺服器，待紀錄工作完成後，再將執行權回交給網頁伺服器執行原本預定之網頁工作。這種作法的主要缺點為造成瀏覽時間延遲與畫面停頓，故應用較少。代理人系統乃是在不影響使用者瀏覽作業的情況下，由一個電腦執行程序自動紀錄使用者瀏覽歷程，並回報予伺服器。

此外，針對單一使用者之識別，則有 Cookie 與 Session 兩種方式。Cookie 是一種記錄使用者資訊的技術，當使用者的瀏覽器連上伺服器後，伺服器便於客戶端電腦中記錄蒐集網頁設計者指定之資料，並以 Cookie 形式存放於客戶端電腦（Samar 1999）。日後，當使用者的瀏覽器連上伺服器，伺服器便可藉由此些資訊辨別使用者身分（Lee & Kim 1999）或探索使用者之瀏覽序列（Monticino 1998；陳振東 & 朱志浩 2003），進行更具深度之服務與管理，並進行服務績效評估（Gschwind 等人 2002）。Session 與 Cookie 的運作方式類似，但不同的是 Session 是在網頁伺服器端所執行，而非下載至客戶端執行，所以伺服器可以明確進行控制，並取得所需之資訊（Thomas 1997；Gutzmann 2001）。

二、文件分群與分類

Tyrvainen 與 Paivarinta（1999）曾於研究中整理各類產業的十一種文件類型，該研究發現若產業特徵不同，其應用之文件類型差異亦甚大，故資訊系統專家、組織規劃者及特定產業領域專家合作時，企業文件類型必須列為文件管理系統發展的重要考量。分群技術乃將一群體分隔為數個性質相近之子群，而與分類技術所不同者在於分群技術無須預先定義分類所需之類別，其乃根據群體特徵之相近性而劃分群集。因此，群集化可視為分類之前置作業，也是進行應用區隔的首要任務。將共同主題或相關性高之文件集合為同一族群，可協助進行文件分類或文件管理等工作。楊傑勝（2000）乃提出適應性群集演算法（Clustering Algorithm），該方法可在每個類別文件中找出一個具代表性之特徵文件，再根據群集之結果可找到與此代表性文件相關之文件。詹智凱（2000）則以詞彙關聯性為基礎進行文件自動分類，亦即利用詞彙與詞彙間的關聯性（Hou & Chan 2003），將關聯性高之詞彙聚成一群集，形成代表該群集（類別）之關鍵字，再利用此些成形之群集將文件分類。

針對網路文件自動分類之研究，顧皓光（1997）提出一適合網路文件自動分類之模

型，並充分利用 Web 文件提供超文件連結特性及 HTML 標籤加註等功能，提昇系統分類能力。Ng 等人（2001）則利用機率模式將網路文件分類為有意義文件及無意義文件，此機率性模式乃以多變量統計分析為基礎，利用典型網路文件進行測試，得知此機率性模式較適用於網路文件之二元分類。此外，過去關於文件分類之研究尚有許多學者提出關鍵字分類法（侯永昌&楊雪花 1998）、經驗分類法（Lin 等人 2002）及其他分類法（Haruechaivasak et al. 2002）等，可做為文件自動化分類之基礎。

類神經網路方法普遍應用於學習及資訊解析等課題，其可針對文件多項特徵擷取相關屬性（Dasigi 1998；Salvador et al. 2002）；此些文件特徵屬性可應用為文件分群的依據。Lam &Low（1997）應用特徵擷取及機率推論等觀念，發展一套 Bayesian 推論網路，作為文件分群推論方法；而 Lam &Chao（1999）則結合實例為基（Instance-based）與規則為基（Rule-based）之概念進行文件分群。此外，Farkas（1993、1995）運用語意概念建立量化之向量，發展學習演算法和自我組織圖示（Self-Organizing Maps），建構名為 NeuroClass 之自動文件分群功能。

三、文件授權對象推論

在消費者導向之市場環境下，顧客關係管理已成為產學界高度重視之課題。過去關於顧客關係管理之議題以探索消費者行為趨勢者居多（Yuan &Chang 2001；Kim &Han 2001），其目的乃藉由此行為趨勢協助企業體進行相關決策。由於電子商務環境興起，近年來網頁歷程探勘（Web Usage Mining）已被廣泛應用於探索網頁瀏覽者行為，作為改進企業體網站架構設計或資訊內容提供之參考（Jenamani， 2003；Cooley et al. 1997）。然在電子環境下，如何針對顧客之行為趨勢與特徵，給予閱讀資訊、文件之授權，為企業進行市場行銷或確保資訊安全之重要課題。網頁使用歷程探勘主要乃透過使用者於全球資訊網上之日誌檔進行解析，藉由分析使用者在該網站伺服器上所留下之網頁瀏覽紀錄，發掘網路使用者之瀏覽行為模式。一般可藉此得知使用者興趣偏好、網站瀏覽狀況、網頁受歡迎程度，藉以調整網頁結構，提供個人化服務予使用者（Feng &Murtagh 2000；Kim et al. 2002）。

過去關於電子化文件權限與安全管理的相關研究著重於應用認證技術（Feldella & Prandini 2000）、加密技術（Wewers &Wargitsch 1998）於資訊/文件權限控管。由於安全性協定屬文件管理結構之最上層，且與網頁資訊管理高度相關，Dridi &Neumann（1998）乃提出一套根據文件內容進行文件分類之模式，以作為網路資訊管理之依據。基於文件內容與文件權限對象高度相關，部分研究學者乃提出以文件分類方式，作為權限控管之參考（孫銘聰&侯建良 2003）。為解決資料庫中因目錄或種類繁多所引起之管理問題，Navathe &Yong（1998）提出多指標之文件分類法解決繁雜文件分類之問題，並依此進行權限控管。知識文件提供者在分享文件時，常指定訊息的接受對象；然基於網際網路帶動之知識爆漲，知識接受者亦當有權利指定其所期望取得之知識類型。因此，Dengel（1997）即提出一套 Office MAID 系統，根據企業之制式採購文件，自動決策作業文件之接受對象與工作流程。另外，若一公司或組織同時有數專案針對同一份文件進行處理時，則可根據文件內容進行相關性遞減排列，再以文件分類及權限控制觀念解決多角色

同步處理一份文件之問題（Carrere et al. 1998）。

綜上所述，考量過去關於使用者瀏覽行為、知識分類、知識管理之相關研究趨勢，本研究所發展之三階段式 CRM 與 KM 整合管理模式，乃整合文件關聯性分析、文件分群及文件權限對象推論等課題，作為自動化知識管理與顧客關係管理之基礎。當中，文件關聯性乃根據文件關鍵字進行推論，並由文件相關性高低與使用者閱讀趨勢決定文件群集與權限對象。研究中並發展一雛形技術展示其可行性，並與既有相似之方法論進行比較。期能藉由此些模式與技術之研發，作為產業發展智慧型 CRM 系統或 KM 系統之基礎，減輕企業組織之專家決策負荷。

參、三階段式 KM 與 CRM 整合推論模式

本研究乃發展以文件關聯性、使用者閱讀趨勢為基礎之知識文件分群與客戶關係管理模式，整合 KM 與 CRM 機能建構一知識服務中心，使組織之知識管理與客戶關係管理效獲得顯著提升。本研究與一般 Web Usage Mining 技術不同；以資料探勘範圍而言，本研究所探勘之範圍為使用者瀏覽系統知識文件之歷程，故需要資料庫技術記錄各使用者點選檔案超連結。然而一般 Web Usage Mining 技術乃應用 Log 檔案或 Cookie 取得使用者瀏覽不同網站之資訊。另一方面，就資料探勘內容而言，一般 Web Usage Mining 技術乃探勘已知、變動性較小之網頁（指僅探勘網站內部網頁者），且分析之內容較非詳細逐字針對網頁內容進行解析。本研究所探勘之內容為使用者使用系統之知識文件，當中知識管理系統之文件變異較一般網站之網頁變動大，且文件內容為自由形式，而非如網頁之結構式 HTML 內容。

此知識服務中心之架構如圖 1 所示，乃包含三大核心決策階段—即「文件關聯性分析」、「文件分群推論」及「文件授權客戶推論」。於「文件關聯性分析」中，乃以關鍵字擷取為基礎，根據文件關鍵字之分佈狀況，判斷文件間之相關性。「文件分群推論」則以文件相關性分佈為依據，運用 K 平均法之精神，自動完成文件分群。而「文件授權客戶推論」則根據文件相關性分佈與使用者閱讀趨勢，自動決策文件之接受對象。於說明此些推論模式之前，將模式中所採用之符號定義如下：

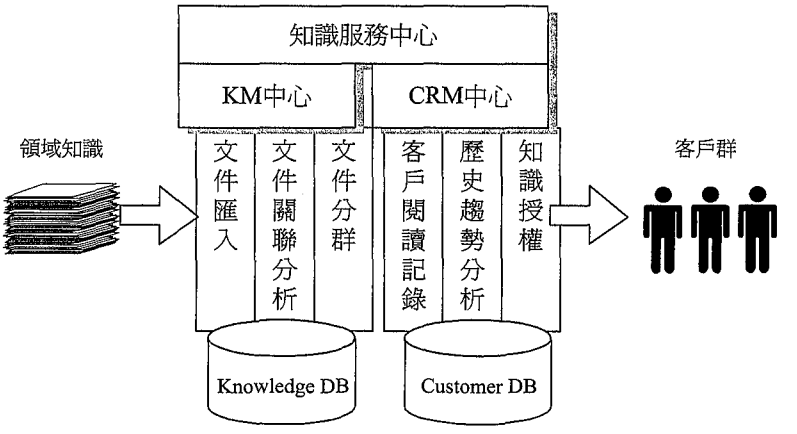


圖 1：整合 KM 與 CRM 之知識服務中心

A	分群維度
$B(M_i,DU)$	M_i 是否擁有 DU 文件權限之指標函數（$B(M_i,DU)=1$ 代表具有權限、$B(M_i,DU)=0$ 則否）
δ	文件權限開放之門檻值
D_i	文件庫中第 i 份文件
DG_i	第 i 份文件所屬之文件群集
$Dist_{i,k}$	D_i 與 A 份種子文件之相關向量與第 k 個種子向量之距離
DU	權限對象未知之目標文件
G_k	第 k 個文件群集，k =1~H
H	分群群數
$K_{i\bullet}$	D_i 所有關鍵字所成之集合
K_{ij}	D_i 之第 j 個關鍵字
$K(DU)$	DU 文件之權限開放對象所成之集合
M_i	第 i 位文件需求者
M_iD_j	M_i 已閱讀之第 j 份文件
M_iR_j	M_iD_j 與 DU 文件間之相關性
n	文件庫之文件數
N_i	D_i 除去無意義字後所剩餘之總字元數
$N(G_j)$	第 j 個文件群集下之文件數
$N(M)$	文件需求者個數
$N(M_i,D)$	M_i 已閱讀之文件份數
$N(K_{i\bullet})$	D_i 之關鍵字個數
$N(K_{i\bullet}\cap K_{j\bullet})$	D_i 與 D_j 相同之關鍵字個數
P_i	M_i 被認定為目標文件權限對象之機率

$\mathbf{R}_i$	D_i 與各種子文件之相關性所形成之向量（其中 i=1~n）
R_{ij}	D_i 與 D_j 之相關性係數
$R_{i,a}$	D_i 與 SD_a 之相關性（其中 i=1~n、a=1~A）
SD_a	第 a 份種子文件（其中 a=1~A）
S_{ka}	第 k 個群集的第 a 維度種子值（其中 k=1~H、a=1~A）
S'_{ka}	第 k 個群集的第 a 維度之新種子值（其中 k=1~H、a=1~A）
$S(K_{i\bullet})$	D_i 所有關鍵字之出現頻率總和（即 $\sum_j S(K_{ij})$）
$S(K_{i\bullet} \cap K_{j\bullet})$	D_i 與 D_j 相同關鍵字之出現頻率總和
$S(K_{ij})$	K_{ij} 之出現頻率

一、文件關聯分析模式

此模式乃以文件關鍵字為基礎判斷文件間之相關性，再以此相關性，達成文件分群之目標，進而作為後續授權決策之依據。本節即詳述文件關聯性分析之細節，而文件接受對象推論之說明則於下一節闡述。文件關聯性分析乃首先針對各單一文件之詞彙頻率（孫銘聰&侯建良，2003）與語意關聯（Hou&Chan，2003）擷取其關鍵字；待所有文件完成對應之關鍵字擷取後，再以兩兩文件之關鍵字進行比較，計算任兩份文件之相關性。此分析含兩種作法：其一乃僅考慮關鍵字內容（以下簡稱 Index A），另一作法則是除關鍵字內容外，並考量關鍵字於文件之出現頻率（以下簡稱 Index B）。本推論模式共有四大步驟，其完整規劃架構與流程如圖 2 所示。

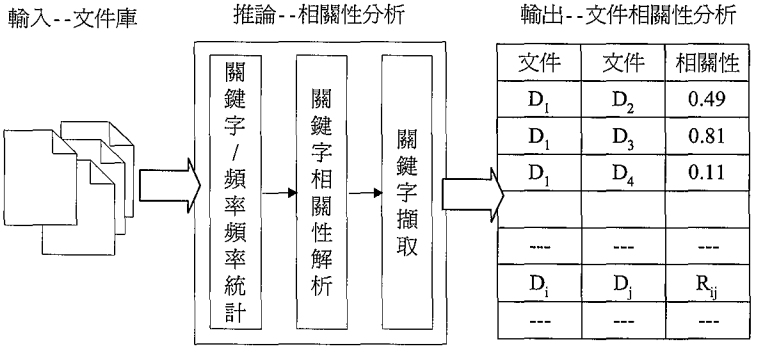


圖 2：文件相關性分析模式之輸入/輸出

步驟(A1)—文件前處理

本階段乃搭配非關鍵字集去除文件內容中無意義之文字（如「我們」、「是否」等詞彙），再根據詞彙於文件之出現頻率與語意關聯整併之結果擷取關鍵字。此部份所整合乃參照孫銘聰&侯建良（2003）所使用的關鍵字擷取方法論，透過字節解析、字詞解析、

字詞關聯比對、字詞頻率整併、候選詞庫關鍵字擷取與待確認詞庫關鍵字擷取等步驟，擷取文件庫中各文件（D_i）之關鍵字（$K_{i\bullet}$）。擷取各文件之關鍵字後，並統計之各關鍵字出現頻率，其結果可整理如表 1。

步驟(A2)—關鍵字集相關性分析

取得表 1 之資料後，即可以關鍵字針對表中任兩份文件解析其相關性。解析方式有以下兩種作法：

(a) Index A：即找出兩文件間相同之關鍵字個數$N(K_{i\bullet}\cap K_{j\bullet})$，則其相關性即可由下式推導：

$$R_{ij}=\frac{\dfrac{N(K_{i\bullet}\cap K_{j\bullet})}{N_i}+\dfrac{N(K_{i\bullet}\cap K_{j\bullet})}{N_j}}{\dfrac{N(K_{i\bullet})+N(K_{j\bullet})}{N_i+N_j}\times 2} \tag{1}$$

表 1：文件關鍵字擷取列表

文件	D_1		D_2		…		D_i		…	
	Index A	Index B	Index A	Index B	Index A	Index B	Index A	Index B	Index A	Index B
關鍵字	K_{11} K_{12} ⋮ K_{1j} ⋮	$S(K_{11})$ $S(K_{12})$ ⋮ $S(K_{1j})$ ⋮	K_{21} K_{22} ⋮ K_{2j} ⋮	$S(K_{21})$ $S(K_{22})$ ⋮ $S(K_{2j})$ ⋮	…	…	K_{i1} K_{i2} ⋮ K_{ij} ⋮	$S(K_{i1})$ $S(K_{i2})$ ⋮ $S(K_{ij})$ ⋮	…	…
次數	$N(K_{1\bullet})$	$S(K_{1\bullet})$	$N(K_{2\bullet})$	$S(K_{2\bullet})$	…	…	$N(K_{i\bullet})$	$S(K_{i\bullet})$	…	…

(b) Index B：即找出兩文件共同關鍵字之總出現頻率$S(K_{i\bullet}\cap K_{j\bullet})$，其相關性則可由共同關鍵字頻率佔總頻率之比例加以計算：

$$R_{ij}=\frac{\dfrac{S(K_{i\bullet}\cap K_{j\bullet})}{N_i}+\dfrac{S(K_{i\bullet}\cap K_{j\bullet})}{N_j}}{\dfrac{S(K_{i\bullet})+S(K_{j\bullet})}{N_i+N_j}\times 2} \tag{2}$$

步驟(A3)—文件間相關性分析

依據前一步驟之概念，針對所有文件進行兩兩文件間之相關性分析，可求得R_{ij}（當中$R_{ij}=R_{ji}$），並建立文件間相關性對照表（參見表 2）。此表可應用於產業文件管理系統之文件分群、文件授權決策或文件庫資料模糊搜尋之基礎。

二、文件分群模式

根據前述步驟所得之文件相關性，可將文件庫內各文件予以分群；亦即針對文件相關性分佈狀況，將相關係數相近者歸為同一群集，以利後續管理與授權推論。過去不同領域之分群研究普遍採用之分群法則為 K 平均法，此方法乃先由使用者指定欲分群之群數（在此以 H 表示之），再隨機產生對應相同數目之種子值（Seed Value）作為群集質心；將各分群目標歸列予最接近之質心，形成 H 個分群。此 H 個分群形成後，再重新計算質心，並以上述觀念重新分群。如此反覆執行，直到各群集包含之分群目標不再變動為止。本研究利用此方法進行文件分群，方法論中乃自文件庫隨機擇取 A 份種子文件（A 視為分群維度），並於初始階段隨機產生 H 個種子向量作為群集質心；取得各文件與 A 份種子文件之相關性後，再以此些相關性尋找其最接近之群集質心，給予一初步之群集分配。之後反覆計算新群集質心並重新分群，直到群集所包含之文件不再變異為止。此模式主要有以下五大運作步驟（其運作架構如圖 4 所示）。

表 2：文件相關性對照表

	D_1	D_2	D_3	D_4	…	D_i	…
D_1	1	R_{21}	R_{31}	R_{41}	…	R_{i1}	…
D_2	R_{12}	1	R_{32}	R_{42}	…	R_{i2}	…
D_3	R_{13}	R_{23}	1	R_{43}	…	R_{i3}	…
D_4	R_{14}	R_{24}	R_{33}	1	…	R_{i4}	…
⋮	⋮	⋮	⋮	⋮	…	⋮	…
D_j	R_{1j}	R_{2j}	R_{3j}	R_{4j}	…	R_{ij}	…
⋮	⋮	⋮	⋮	⋮	…	⋮	…

步驟(B1)：文件相關性計算

由決策者設定進行文件分群時所使用之維度數（即 A）後，即隨機選定文件庫中之 A 份文件作為種子文件。以此些種子文件為基礎，透過前一小節所述之文件相關性推論方法，進行各文件 D_i 與 SD_a 之相關性分析（相關係數為 $R_{i,a}$），並建立文件相關性對照表。如表 3 所示，可得到 n 個一維陣列 $\mathbf{R}_i$，其元素為 D_i 與各種子文件之相關性 R_{ia}。

步驟(B2)—取得初階種子值

由決策者設定所需之文件分群數（以 H 代表之），並以隨機產生 H*A 個介於[0,1]間之亂數值（即 $S_{ka} \sim U(0,1)$，k=1~H、a=1~A，且 $U(0,1)$ 代表介於[0,1]間之 Uniform 分配）。所產生之隨機亂數值即為第一階段進行分群之種子值，後續步驟即以此為群集質心，作為其他文件歸列至各群集之基礎。

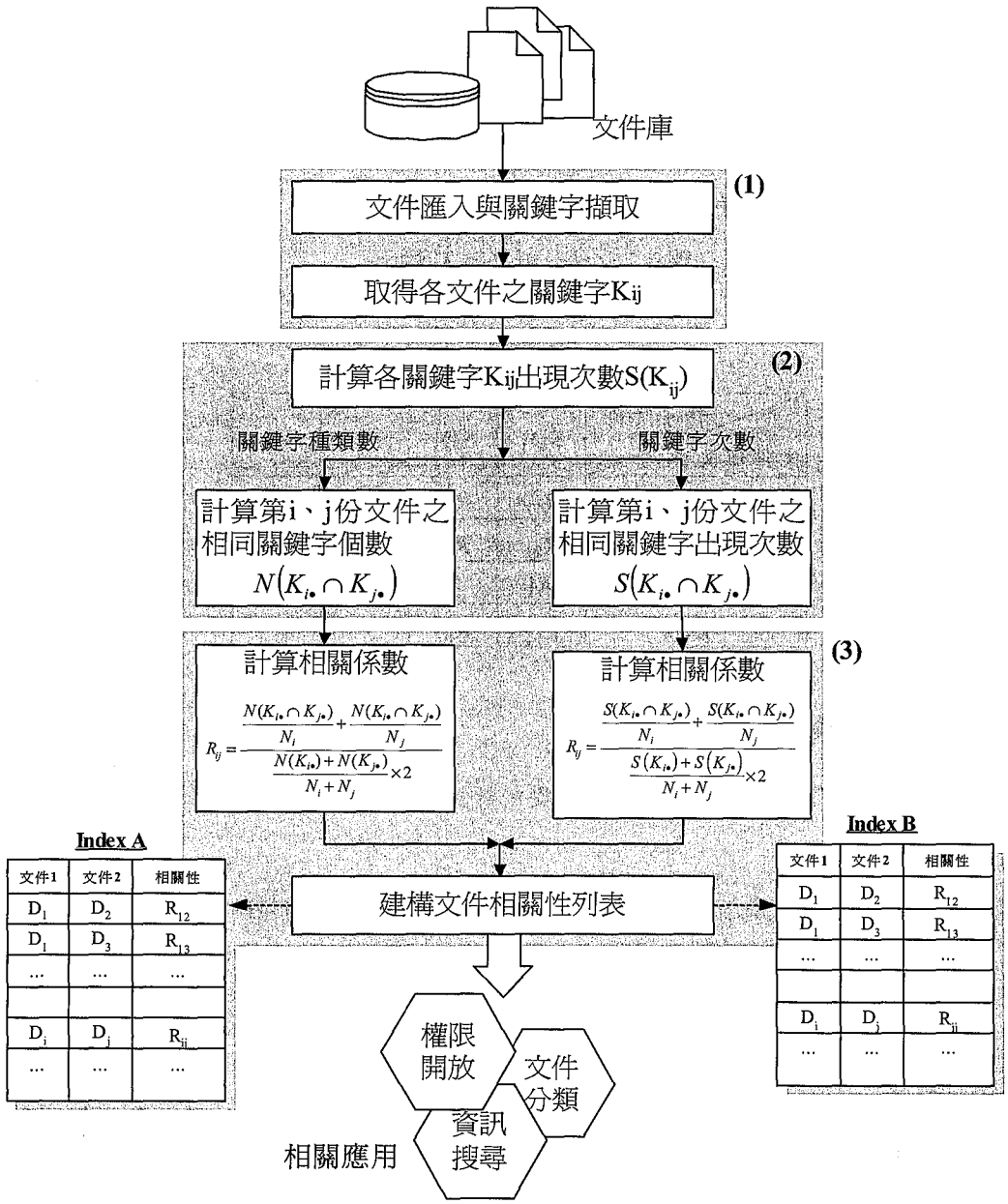


圖 3：文件相關性分析模式流程

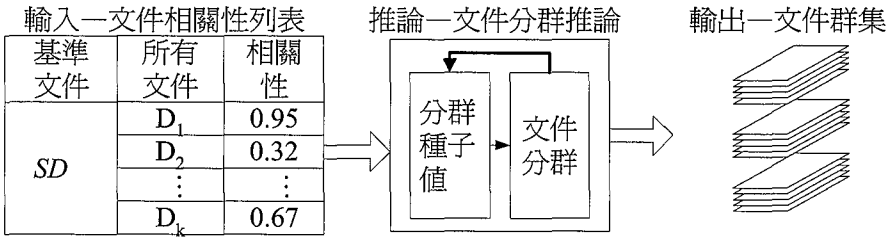


圖 4：文件分群之輸入/輸出

表 3：文件相關性分析列表

種子文件 / 文件庫文件	SD_1	SD_2	…	SD_A
D_1	R_{11}	R_{12}	…	R_{1A}
D_2	R_{21}	R_{22}	…	R_{2A}
M	M	M	…	M
D_n	R_{n1}	R_{n2}	…	R_{nA}

步驟(B3)—文件群集歸列

以各文件對應之相關係數向量 $\mathbf{R}_i$ 為基礎，計算各文件 D_i 與各種子向量 S_j 之距離：

$$Dist_{i,k} = \sqrt{\sum_{a=1}^{A}\left(R_{ia} - S_{ka}\right)^2} \quad \text{，i=1~n、a=1~A、k =1~H} \tag{3}$$

各文件乃選擇與其最接近之群集質心（即相關性較高者）歸列為所屬群集，即：

$$DG_i = j \quad \text{if} \quad \min(Dist_{i,\bullet}) = Dist_{i,j} \quad \text{，for i=1~n} \tag{4}$$

步驟(B4)—新群集質心計算

將文件庫中各文件依步驟(B3)歸列至各群集後，將各群集中每一文件所對應之相關係數予以平均，可得到各群集之新質心。即：

$$S'_{ka} = \frac{\sum_{i=1}^{n}\left(R_{i,a} \mid DG_i = k\right)}{N(G_k)} \quad \text{for a=1~A} \tag{5}$$

步驟(B5)—反覆分群

以前一步驟所得之新質心 S'_g 為基礎（即 $S_g = S'_g$），重複上述步驟(B3)、(B4)，取得新

群集文件，直至各群集內含之文件不再變動為止。最後即可得到一系列之文件群集G_k與對應文件$\{D_i \mid DG_i = k\}$，k=1~H，如此即完成文件分群。

三、文件授權客戶推論模式

文件授權客戶模式乃以前述之文件關聯性分析結果為基礎，首先分析目標文件與客戶過去閱讀文件之關聯性，即可針對每一客戶取得一系列相關係數值。此些相關係數值經整併後（包括平均值法、最大值法及中位數法），即可產生一代表該使用者被認定為此目標文件接收對象之推斷值。最後，以系統管理者所指定之門檻值或隨機亂數值，篩選實際接收對象。此模式概念如圖 5 所示，以下詳述各步驟之作法。

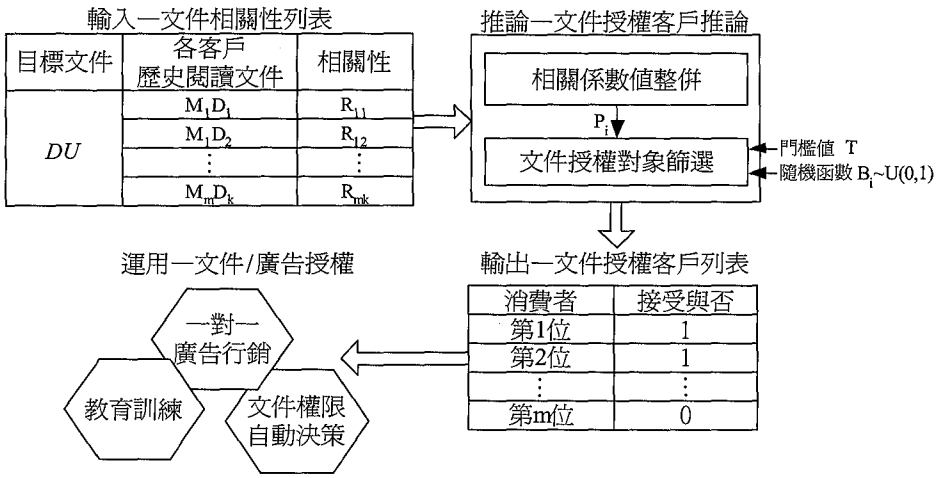


圖 5：文件授權客戶推論流程示意圖

步驟(C1)—關聯性分析

擷取權限未知之目標文件*DU*關鍵字，並與各客戶已閱讀文件進行相關性分析，取得文件相關性對照表（如表 4）；此部分之觀念及手法與前述「文件關聯性分析」之觀念相同。

表 4：文件相關性對照表

權限未知文件	文件需求者已閱讀文件	相關性
DU	M_1D_1	M_1R_1
	M_1D_2	M_1R_2
	⋮	⋮
	M_iD_j	M_iR_j
	⋮	⋮
	M_mD_n	M_mR_n

步驟(C2)—權限開放機率計算

由上述步驟所得之列表，可以下述多種方法計算第 i 位文件需求者擁有目標文件 DU 權限之機率 P_i。

(a) 平均值法：

此方法乃將所有文件之相關係數全部納入考慮，即認定所有使用者瀏覽之文件皆具有權限推論之代表性，以整體之平均值作為判斷分享機率之標準。其計算方式如下：

$$P_i = \frac{\sum_{j=1}^{n} M_i R_j}{N(M_i D)} \tag{6}$$

(b) 最大值法：

即取第 i 位文件需求者所有閱讀之文件與權限未知文件 DU 相關性之最大值，作為判斷之標準（認定該文件具有高度代表性）。其計算方式如下：

$$P_i = Max(M_i R_j) \tag{7}$$

(c) 中位數法：

考量文件需求者其閱讀趨勢可能差異甚大，此時相關性之中位數便可以用來作為折衷判斷之標準。其計算方式乃將 $M_i R_1$、$M_i R_2$、...、$M_i R_n$，由小到大依序排列，則：

- 當 $N(M_i D)$ 為奇數時，$P_i = M_i R_L$，當中 $L=(N(M_i D)+1)/2$。
- 當 $N(M_i D)$ 為偶數時，$P_i = (M_i R_K + M_i R_{K+1})/2$，當中 $K=N(M_i D)/2$。

(D) 區間估計法：

在平均值法中，考量其機率推算結果可能受到某些相關係數特低或特高（outlier）之文件影響，因此給定一判斷區間，將未落在此判斷區間內之相關係數剔除後，再計算整理後之整體之平均值，作為判斷之標準。其計算方式如下：

$$P_i = \frac{\sum_{j=1}^{n}\left(M_i R_j \mid M_i R_j \in \overline{X} \pm 3S\right)}{N\left(M_i R_j \mid M_i R_j \in \overline{X} \pm 3S\right)}，其中 \overline{X} = \frac{\sum_{j=1}^{n} M_i R_j}{N(M_i D)}，S = \sqrt{\frac{\sum_{j=1}^{n}\left(M_i R_j - \overline{X}\right)^2}{n-1}} \tag{8}$$

步驟(C3)—判斷目標文件之需求者

根據系統管理者所指定之門檻值 δ 及上述計算所得之機率值，可以下式判斷文件需求者是否具有目標文件之權限：

$$B(M_i DU) = \begin{cases} 1, if\ P_i \geq \delta \\ 0, if\ P_i < \delta \end{cases} \tag{9}$$

當 $B(M_i DU)=1$，則代表文件需求者具有目標文件 DU 之權限，故 DU 文件之權限對象集合為 $K(DU)=\{M_i \mid B(M_i DU)=1\}$。

肆、矩陣式推論模式

前一小節所述之方法論可以矩陣模式予以表現，於說明此模式前，將相關變數定義如下：

K　文件庫中所有文件關鍵字所成之關鍵字集合（K_j 即關鍵字集合的第 j 個關鍵字）

M　文件關鍵字隸屬關係矩陣（M_i 即第 i 份文件之關鍵字隸屬矩陣）

M'　整理文件關鍵字擷取列表後，文件庫中所有文件關鍵字出現次數與關鍵字集合之隸屬矩陣

M_i'　文件庫中第 i 份文件中關鍵字出現次數對應關鍵字集合之隸屬矩陣

M_u　文件庫各文件之權限對象矩陣（以文件庫各文件為 x 軸、權限對象為 y 軸）

M_u'　目標文件之權限開放對象集合

P　權限開放予各對象之機率矩陣

R'　文件庫內兩兩文件之相關性對照矩陣

R_{ij}　第 i 份文件與第 j 份文件間之相關性係數

$R_{i,u}$　第 i 份文件與目標文件之相關係數

R_u'　目標文件與文件庫內各文件間之相關性係數之集合

若僅考慮關鍵字內容，首先以文件庫各文件為行（Column）、關鍵字集合為列（Row），將文件關鍵字擷取列表轉換成矩陣形式，得到一文件關鍵字隸屬關係矩陣 M：

$$M=\begin{bmatrix} B_{1,1} & B_{1,2} & \Lambda & B_{1,i} & \Lambda & B_{1,n} \\ B_{2,1} & B_{2,2} & \Lambda & B_{2,i} & \Lambda & B_{2,n} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ B_{w,1} & B_{w,2} & \mathrm{K} & B_{w,i} & \Lambda & B_{w,n} \end{bmatrix} \tag{10}$$

當中 $B_{i,j}$ 代表第 j 份文件與第 i 個關鍵字之隸屬關係。以 Index A 而言，則 $B_{i,j}=1$ 代表第 i 個關鍵字為第 j 份文件之關鍵字；否則，$B_{i,j}=0$。如欲計算兩文件 D_i 與 D_j 之相關性，可先取得各文件之關鍵字集合矩陣：

$$M_i+M_j=\begin{bmatrix} B_{1,i} \\ B_{2,i} \\ \mathrm{M} \\ B_{m,i} \end{bmatrix}+\begin{bmatrix} B_{1,j} \\ B_{2,j} \\ \mathrm{M} \\ B_{m,j} \end{bmatrix}=\begin{bmatrix} B_{1,i}+B_{1,j} \\ B_{2,i}+B_{1,j} \\ \mathrm{M} \\ B_{m,i}+B_{1,j} \end{bmatrix} \tag{11}$$

若 $B_{h,i}+B_{h,j}=2$，則令 $V_h=1$；否則 $V_h=0$。兩文件之共同關鍵字集合矩陣可推導為：

$$M_i \cap M_j = \begin{bmatrix} V_1 \\ V_2 \\ \mathrm{M} \\ V_w \end{bmatrix} \tag{12}$$

如此任一文件之關鍵字個數為 $N(M_i) = B_{1i} + B_{2i} + \Lambda + B_{wi}$，而兩文件之共同關鍵字個數為 $N(M_i \cap M_j) = V_1 + V_2 + \Lambda + V_w$。故兩文件之相關性可以下式推導：

$$R_{ij} = \frac{\dfrac{N(M_i \cap M_j)}{N_i} + \dfrac{N(M_i \cap M_j)}{N_j}}{\dfrac{N(M_i) + N(M_j)}{N_i + N_j}} \tag{13}$$

若以 Index B 而言，則建立文件關鍵字出現頻率矩陣 M'：

$$M' = \begin{bmatrix} N(K_{1,1}) & N(K_{1,2}) & \Lambda & N(K_{1,i}) & \Lambda & N(K_{1,n}) \\ N(K_{2,1}) & N(K_{2,2}) & \Lambda & N(K_{2,i}) & \Lambda & N(K_{2,n}) \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ N(K_{w,1}) & N(K_{w,2}) & \mathrm{K} & N(K_{w,i}) & \Lambda & N(K_{w,n}) \end{bmatrix} \tag{14}$$

當中元素 $N(K_{i,j})$ 代表第 j 份文件之第 i 個關鍵字出現頻率。如欲計算兩文件 D_i 與 D_j 之相關性，可以關鍵字交集關係重新建立文件之關鍵字頻率矩陣；若 $V_i = 1$，則 $N'(K_{h,i}) = N(K_{h,i})$ 且 $N'(K_{h,j}) = N(K_{h,j})$，否則 $N'(K_{h,i}) = N'(K_{h,j}) = 0$。故文件之新關鍵字頻率矩陣為：

$$M_i' = \begin{bmatrix} N'(K_{1,i}) \\ N'(K_{2,i}) \\ \mathrm{M} \\ N'(K_{w,i}) \end{bmatrix} \tag{15}$$

故兩文件之相關性可以下式推導：

$$R_{ij} = \frac{\dfrac{N'(K_{\bullet,i})}{N_i} + \dfrac{N'(K_{\bullet,j})}{N_j}}{\dfrac{N'(K_{\bullet,i}) + N'(K_{\bullet,j})}{N_i + N_j}} \tag{16}$$

依據前述步驟，可對文件庫內兩兩文件進行相關性分析，可求得 R_{ij}（當中 $R_{ij} = R_{ji}$），並建立文件間相關性對照矩陣：

$$R' = \begin{bmatrix} R_{1,1} & R_{1,2} & \Lambda & R_{1,i} & \Lambda & R_{1,n} \\ R_{2,1} & R_{2,2} & \Lambda & R_{2,i} & \Lambda & R_{2,n} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ R_{n,1} & R_{n,2} & \mathrm{K} & R_{n,i} & \Lambda & R_{n,n} \end{bmatrix} \tag{17}$$

透過文件關聯性分析模式，可求得目標文件與文件庫各文件間之相關性係數矩陣 R_u'，再結合文件權限對象矩陣，可得知目標文件權限開放予各對象之機率：

$$P = M_u \times R_u' = \begin{bmatrix} B_{1,1} & B_{1,2} & \mathrm{L} & B_{1,i} & \mathrm{L} & B_{1,k} \\ B_{2,1} & B_{2,2} & \mathrm{L} & B_{2,i} & \mathrm{L} & B_{2,k} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ B_{m,1} & B_{m,2} & \mathrm{K} & B_{m,i} & \mathrm{L} & B_{m,k} \end{bmatrix} \times \begin{bmatrix} R_{1,u} \\ R_{2,u} \\ \mathrm{M} \\ R_{k,u} \end{bmatrix} = \begin{bmatrix} P_{1,u} \\ P_{2,u} \\ \mathrm{M} \\ P_{m,u} \end{bmatrix} \tag{18}$$

其中元素 $P_{i,u}$ 即代表目標文件權限開放予第 i 位對象之機率。由 $U(0,1)$ 隨機產生 m 個亂數值 $V_1, V_2, 、、、, V_m \sim U(0,1)$；若 $V_i \le P_{iu}$，令 $B(D_{i,u}) = 1$，即目標文件權限開放予第 i 位使用者；否則 $B(D_{i,u}) = 0$。故目標文件之授權客戶集合為：

$$M_u' = \begin{bmatrix} B(D_{1,u}) \\ B(D_{2,u}) \\ \mathrm{M} \\ B(D_{m,u}) \end{bmatrix} \tag{19}$$

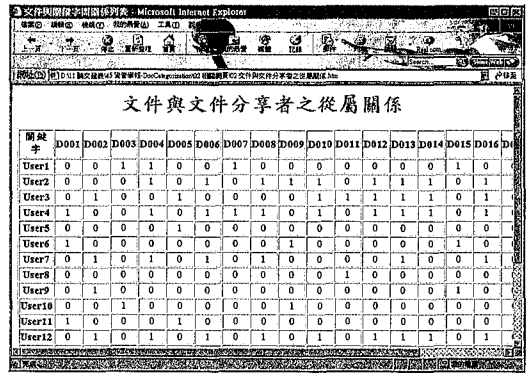
伍、應用案例

本研究乃以 JSP 搭配 SQL Server 2000 為工具，發展一套於 Web 環境下具有文件關鍵字擷取、文件相關性分析、文件分群與文件授權解析功能之系統平台。此系統在各文件匯入後，即依其內容擷取文件關鍵字及關鍵字出現頻率等資料，並整理文件與關鍵字間之關係矩陣（如圖 6 所示）。之後，即自資料庫取得各文件與客戶之授權資料，整理成從屬關係列表（如圖 7 所示），以作為分析新目標文件與客戶間從屬關係之依據。藉由關鍵字之種類與出現頻率兩向度，取得文件庫內各文件間與目標文件間之相關性，並且產生相關性列表。最後利用各文件與客戶之從屬關係、及目標文件與各既有文件之相關性，求得目標文件授權予各客戶之機率值。以此機率值為篩選標準，即可決定目標文件之授權對象。

文件與關鍵字間關係列表

關鍵字	D001		D002		D003		D004		D005		D006		D007		D008		D009		D01	
	個數	頻率	個數	頻率	個數	頻率	個數	頻率	個數	頻率	個數	頻率	個數	頻率	個數	頻率	個數	頻率	個數	頻
KW1	0	0	1	0.011	0	0	1	0.009	0	0	0	0	0	0	1	0.007	0	0	0	0
KW2	1	0.025	1	0.021	0	0	0	0	1	0.027	0	0	1	0.006	0	0	1	0.031	0	0
KW3	0	0	0	0	1	0.029	0	0	0	0	1	0.013	1	0.024	0	0	0	0	1	0.0
KW4	1	0.013	0	0	0	0	1	0.006	0	0	0	0	1	0	0	0	0	0	0	0
KW5	0	0	0	0	1	0.014	1	0.031	0	0	0	0	0	0	0	0	0	0	0	0
KW6	1	0.009	0	0	0	0	0	0	1	0.006	0	0	0	0	1	0.019	0	0	1	0.0
KW7	1	0.010	1	0.009	0	0	0	0	0	0	0	0	1	0.021	0	0	0	0	0	0
KW8	0	0	0	0	0	0	0	0	0	0	1	0.023	0	0	0	0	0	0	0	0
KW9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.022	0	0	0	0

圖 6：文件與關鍵字群組之關係



文件與文件分享者之從屬關係

關鍵字	D001	D002	D003	D004	D005	D006	D007	D008	D009	D010	D011	D012	D013	D014	D015	D016	D0
User1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	
User2	0	0	0	1	0	1	0	1	1	1	0	1	1	1	0	1	
User3	0	1	0	0	1	0	0	0	0	1	1	1	1	1	0	1	
User4	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0	1	
User5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
User6	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	
User7	0	1	0	1	0	1	0	1	0	0	0	0	1	0	0	1	
User8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
User9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
User10	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	
User11	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
User12	0	1	0	1	0	1	0	1	0	1	0	1	1	1	0	1	

圖 7：各文件與文件分享者之從屬關係

為驗證前述推論模式之運作效能，並以一案例分析此推論模式之推論精準度。本案例參考孫銘聰&侯建良（2003）之汽車專利文件解析個案，該研究乃針對汽車專利文件之管理發展自動化文件分類與權限推論之技術。而本研究所發展之文件分群與文件授權客戶決策結果可與之比較，以進行模式績效分析。由於汽車研發、設計與生產過程中，參與之專業人員乃研發供企業內部運用之重要知識（如發明、智慧財產與生產技術改良）。若要於高度競爭之汽車產業中具領先地位及不侵犯同業之專利、智慧財產技術時，需做好企業重要知識與專利之管控。在此個案中，專利知識之授權對象乃以部門進行區分，分為「研發部門（RD）」、「引擎部門（EN）」、「車裝部門（CA）」、「烤漆部門（RP）」、「機械部門（ME）」與「壓造部門（DC）」等 6 個與汽車製造相關之部門。而文件類型則區分為：「電力、動力、散熱系統（EPRS）」、「烤漆加工作業（ROAP）」、「鈑金加工作業（STBO）」、「踏板、煞車、警示系統（TRWA）」、「車身構造、裝配（CSBA）」、「車身、車內配件（APCA）」、「排氣、滅火系統（EXFI）」、「車輪、轉向、傳動、懸吊（WTDS）」、「車燈照明系統（LALI）」與「其他（OTHE）」等十大類型。

本研究將分析過程分為「訓練」、「測試」與「評估」等三階段。其中，訓練階段乃於十種汽車專利類別下，先以 400 篇汽車生產之專利相關文件進行訓練資料建置。而系統測試則於此十大文件類別中，隨機挑選 15 篇未進行訓練之新文件進行系統推論結果測試。於評估階段則針對系統測試之結果，確認推論模式之準確性與精確性。考量文件分群之績效，分析結果如表 5 所示。本研究所之文件分群方式可將 15 份文件中的 13 份文件正確分群（正確率約 86.67%）；與孫銘聰&侯建良（2003）所提之文件分類方法論比較（正確率約 77.78%），本研究之推論方法可提升正確率約 11%。

就文件授權對象推論而言，利用文件相關性計算目標文件授權予各客戶之機率值後，分別以最大機率值與隨機亂數值決定 15 份測試文件之授權對象，並將結果整理如表 6 所示，各方法論累積準確度分配如圖 8 所示。比較採用最大機率值與隨機亂數值進行文件授權對象推論，最大機率值之推論結果較具準確性（76.5%比 70%）。若考量 Index A 與 Index B 之影響，則以考量關鍵字頻率（Index B）之推論效果較佳（準確度為 70%比 76.5%、精確度為 0.435 比 0.473）；亦即關鍵字出現頻率較高者，其對於權限推論應更具代表性。再與孫銘聰與侯建良（2003）所提之分類方法論比較，本研究所根據關鍵字頻

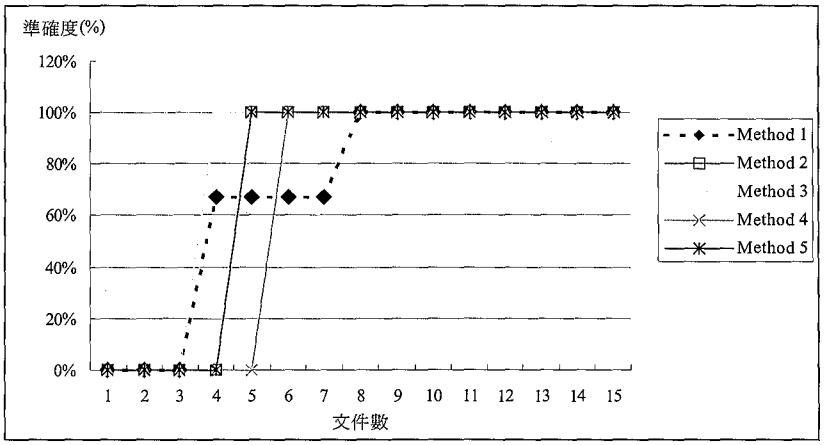
率（Index B）判斷目標文件授權予各客戶之機率值，並以最大機率值決定授權對象之方法較隨機亂數值推論者佳（準確度為 80%比 71%）。表 6 中部分文件（如編號 2 之文件）無法推論其授權對象，其原因在於測試資料與訓練資料間無共同關鍵字所致。若摒除關鍵字無共同性之文件（即無訓練基礎之測試資料），以最大機率值推論而言，其準確度為 85%與 92%，精確度為 0.376 與 0.277，故以關鍵字關聯性為基礎之文件授權推論實具有高度之準確性。

表 5：文件分類與分群之績效分析

文件編號	孫銘聰&侯建良(2003)			本研究之分群法	
	理想值	推論結果	精確度	群集	正確性
1	EPRS	EPRS	100.00%	G1	Y
2	EPRS	EPRS	100.00%	G1	Y
3	ROPA	ROPA	100.00%	G2	Y
4	ROPA	ROPA	100.00%	G2	Y
5	STBO	STBO	100.00%	G3	Y
6	STBO	STBO	100.00%	G2	N
7	TRWA	TRWA	100.00%	G4	Y
8	TRWA	TRWA	100.00%	G4	Y
9	BOAS	ASSE	0.00%	G4	N
10	BOAS	BOAS	100.00%	G5	Y
11	ASSE	ASSE、EPRS	66.67%	G6	Y
12	ASSE	ASSE	100.00%	G6	Y
13	EXFI	EXFI	100.00%	G7	Y
14	OTHE	STBO	0.00%	G8	Y
15	OTHE	BOAS	0.00%	G8	Y
		Average Precision	77.78% 0.398		

表 6：授權對象推論結果整理表

文件編號	理想值	孫銘聰&侯建良(2003)		本方法論（以最大機率值推論）				本方法論（以隨機亂數值推論）			
		推論結果	精確度	推論結果		精確度		推論結果		精確度	
				Index A	Index B	Index A	Index B	Index A	Index B	Index A	Index B
1	EN	EN	100%	EN	EN	100%	100%	EN	EN	100%	100%
2	RP	EN	0%	---	---	0%	0%	---	---	0%	0%
3	RP	RP EN	67%	RP	RP	100%	100%	RP	RP	100%	100%
4	DC	DC	100%	DC	DC	100%	100%	DC	DC	100%	100%
5	DC	DC	100%	DC	DC	100%	100%	DC	DC	100%	100%
6	ME	ME EN	67%	RD	RD	0%	0%	---	---	0%	0%
7	CA	CA	100%	CA	CA	100%	100%	CA	CA	100%	100%
8	CA	CA EN	67%	CA	CA	100%	100%	---	RD	0%	0%
9	ME	ME	100%	ME	ME	100%	100%	ME	ME	100%	100%
10	ME	ME	100%	ME	ME	100%	100%	ME	ME	100%	100%
11	ME	ME	100%	ME	ME	0%	100%	---	ME	0%	100.0 %
12	ME	ME	100%	ME	ME	100%	100%	ME	ME	100%	100%
13	CA	ME CA	67%	CA	CA	100%	100%	CA	CA	100%	100%
14	RD	CA EN	0%	---	---	0%	0%	---	---	0%	0%
15	RD	ME、CA RP、EN	0%	RD	RD	100%	100%	RD	RD	100%	100%
		平均值	71%		平均值	73%	80%		平均值	67%	73%
		標準差	0.396		標準差	0.458	0.414		標準差	0.488	0.458



[註]： Method 1: 孫銘聰&侯建良（2003）、Method 2: Index A+最大機率值、Method 3: Index B+最大機率值、Method 4: Index A+隨機亂數值、Method 5: Index B+隨機亂數值

圖 8：各方法論累積準確度分配圖

本方法論之另一特點為應用之彈性，即方法論之建構非以特定應用領域為考量。此方法論亦應用於另一個案，作為該個案進行生產設計、規劃與製造文件之管理與分享，並經實驗確認其推論精準度（整理如表 7）。在此生產相關文件之管理個案中，本方論之績效雖不若汽車專利管理個案佳，但仍優於孫銘聰&侯建良（2003）所提之方法論；故本方法論應用於其他應用案例仍具有精準度。

表 7：本方法論之推論精準度整理表

			文件分類與分群	授權對象推論
汽車專利個案	過去方法[1]	準確度	77.78%	71%
		精確度	0.398	0.396
	本方法論[2]	準確度	86.67%	80%
		精確度	----	0.414
生產文件個案	本方法論[2]	準確度	82.14%	75%
		精確度	----	0.411

[1]指孫銘聰&侯建良（2003） [2]指 Index B、並以最大機率值推論所得之結果

陸、結論

本研究提出一套以文件相關性分析為基礎之知識文件與授權之 KM/CRM 整合模式。方法論中乃以企業內之文件庫為基礎，擷取文件之關鍵字詞；再利用各文件關鍵字個數與出現頻率，進行其間之相關性分析。利用文件間之相關性分析，可進一步作為文件分類、文件授權對象決策、資訊搜尋等相關領域之研究基礎。藉由此自動推論模式，可針對一份尚未確認權限或發佈對象之目標文件，透過與已知閱覽對象之文件進行相關性分析及納入所有文件需求者之文件閱讀趨勢後，推斷其可以或有意願閱讀此目標文件之機率；進而決策其權限對象，或提出初步之決策方案供系統管理者參考，以增加文件

權限或分享決策之彈性與效率性。在現今網路行銷與知識管理之趨勢下，此方法論將可有效應用於一對一網路行銷或智慧型文件權限開放，將文件資料有效地提供予需求對象。長遠而言，此模式之導入可減輕企業體管理龐大知識文件之人力與時間成本，進而促使企業無人化知識客服中心之成形。

參考文獻

1. 卜小蝶，民 91，以使用記錄分析探索網路使用者檢索興趣之研究，交通大學資訊管理研究所碩士論文。
2. 何昶毅，民 90，以網頁探勘技術提供一對一個人化服務」，東海大學企業管理研究所碩士論文。
3. 侯永昌，楊雪花，1998，『以模糊理論和遺傳演算法為基礎的中文文件自動分類之研究』，模糊系統學刊，第四卷：45～57 頁。
4. 孫銘聰、侯建良，2003，『建構電子化知識文件之權限指派架構與模式』，*工業工程學刊*，第二十卷・第四期：305～316 頁。
5. 陳佳鴻，民 90，發展基於使用者行為導向之智慧型財經資訊系統，交通大學資訊管理研究所碩士論文。
6. 陳振東、朱志浩，2003，『應用資料採礦技術於網際網路使用者資訊偏好分析之研究』，產業論壇，第五卷・第二期：43～64 頁。
7. 楊傑勝，民 89，適應性聚類演算法及其應用，成功大學資訊工程研究所碩士論文。
8. 詹智凱，民 89，以詞的關聯性為基礎的文件自動分類，國立台灣科技大學資訊管理研究所碩士論文。
9. 蔡聰洲，民 90，整合資料倉儲與資料探勘於網站瀏覽分析，交通大學資訊管理研究所碩士論文。
10. 顧皓光，民 86，網路文件自動分類，台灣大學資訊管理研究所碩士論文。
11. Carrere, J., Cholvy, L., Cuppens, F. and Saurel, C., "Merging Security Policies: Analysis of Practical Example," *Proceedings. 11th IEEE on Computer Security Foundations Workshop*, 1998, pp. 123-136.
12. Cooley, R., Mobasher, B. and Srivastava, J., "Web Mining: Information and Pattern Discovery on the World Wide Web," *IEEE International Conference on Tools with Artificial Intelligence*, 1997, pp. 558-567.
13. Dasigi, V., "Information Fusion Experiments for Text Classification," *IEEE Information Technology Conference*, 1998, pp. 23-26.
14. Dengel, A., "Bridging the Media Gap from the Guthenberg's World to Electronic Document Management Systems," *IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation* (4) 1997, pp. 3540-3545.
15. Dridi, F. and Neumann, G., "Towards Access Control for Logical Document Structure,"

Proceedings. The Ninth International Workshop on Database and Expert Systems Applications, 1998, pp. 322-327.

16. Farkas, J., "Neural Networks and Document Classification," *Canadian Conference on Electrical and Computer Engineering* (1) 1993, pp. 1-4.
17. Farkas, J., "Towards Classifying Full-Text Using Recurrent Neural Networks," *Canadian Conference on Electrical and Computer Engineering* (1) 1995, pp. 511-514.
18. Feldella, E. and Prandini, M., "A Novel Approach to On-line Status Authentication of Public-Key Certificates," *16th Annual Conference on Computer Security Applications*, 2000, pp. 270-277.
19. Feng, T. and Murtagh, K., "Towards Knowledge Discovery from WWW Log Data," *Proceedings of International Conference on Information Technology: Coding and Computing*, 2000, pp. 302-307.
20. Gschwind, T., Eshghi, K., Garg, P. K. and Wurster, K., "WebMon: A Performance Profiler for Web Transactions," *Proceedings, Fourth IEEE International Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems*, 2002, pp. 171-176.
21. Gutzmann, K., "Access Control and Session Management in the HTTP Environment," *IEEE Internet Computing* (5:1), 2001, pp. 26-35.
22. Haruechaivasak, C., Shyu, M.-L. and Chen S.-C.,"Web Document Classification Based on Fuzzy Association," *Proceedings, The 26th Annual International On Computer Software and Applications Conference*, 2002, pp487-492.
23. Hou, J.-L. and Chan, C.-A., "A Knowledge Content Extraction Model Using Keyword Correlation analysis," *International Journal of Electronic Business Management* (1:1) 2003, pp. 54-62.
24. Jenamani, M., Mohapatra, P. K. J. and Ghose, S.,"A Stochastic Model of e-Customer Behavior," *Electronic Commerce Research and Applications* (2:1) 2003, pp.81-94.
25. Kim, K.-S. and Han, I., "The Cluster-Indexing Method for Case-Based Reasoning Using Self-Organizing Maps and Learning Vector Quantization for Bond Rating Cases," *Expert Systems with Applications* (21:3) 2001, pp.147-156.
26. Lam, W. and Chao, Y.-H., "Modeling Textual Document Classification," *Proceedings, IEEE International Conference on Systems, Man, and Cybernetics* (3) 1999, pp. 946-949.
27. Lam, W. and Low, K.-F., "Automatic Document Classification Based on Probabilistic Reasoning: Model and Performance Analysis," *IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation* (3) 1997, pp. 2719-2723.
28. Lee, K.-W. and Kim, H.-J., "Consistency Preserving in Transaction Processing on the Web," *Proceedings of the First International Conference on Web Information Systems Engineering* (1) 2000, pp. 190-195.
29. Lin, S.-H., Chen, M. C., Ho, J. M. and Huang Y.-M., "ACIRD: Intelligent Internet

Document Organization and Retrieval," *IEEE Transactions on Knowledge and Data Engineering* (14) 2002, pp. 599-614.

30. Meier, J. and Sprague, R., "Towards a Better Understanding of Electronic Document Management," *Proceeding of the 29th Annual Hawaii International Conference on System Sciences*, 1996, pp. 53-61.
31. Monticino, M., "Web-Analysis: Stripping away the Hype," Computer (31:12) 1998, pp. 130-132.
32. Navathe, S. B. and Yong, C. O., "Avoiding Inference Problem Using Page Level Security Classification," *Proceedings. Ninth International Workshop on Database and Expert Systems Applications*, 1998, pp. 294-299.
33. Ng, Y.-K., Tang, J. and Goodrich, M., "A Binary-Categorization Approach for Classifying Multiple-Record Web Documents Using Application Ontologies and a Probabilistic Model," *Proceedings, Seventh International Conference on Database Systems for Advanced Applications*, 2001, pp. 58-65.
34. Salvador, N. S., Evangelos, T. and Donald, K., "A Feature Mining Based Approach for the Classification of Text Documents into Disjoint Classes," *Information Processing and Management* (38:4) 2002, pp. 583-604.
35. Samar, V., "Single Sign-on Using Cookies for Web Applications," *Proceedings, IEEE 8th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 1999, pp. 158-163.
36. Thomas, B., "Recipe for e-Commerce," *IEEE Internet Computing* (1:6) 1997, pp. 72-74.
37. Tyrvainen, P. and Paivarinta, T., "On Rethinking Organizational Document Genres for Electronic Document Management," *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*, 1999, pp. 1-10.
38. Wewers, T. and Wargitsch, C., "Four Dimensions of Interorganizational, Document-Oriented Workflow: A Case Study of the Approval of Hazardous-Waste Disposal," *Proceedings of the Thirty-First Hawaii International Conference on System Sciences* (4) 1998, pp. 332-341.
39. Yuan, S.-T. and Chang, W.-L., "Mixed-Initiative Synthesized Learning Approach for Web-Based CRM," *Expert Systems with Applications* (20:2) 2001, pp. 187-200.