

個人化及群體化圖書館資訊服務初探

柯皓仁

國立交通大學圖書館副教授兼數位圖書資訊組組長

楊雅雯 吳安琪 戴玉旻

國立交通大學資訊科學研究所碩士

楊維邦

國立交通大學資訊科學系教授兼圖書館館長

【摘要】

近年來，電子圖書館已成為圖書館界努力追尋的目標。電子圖書館的主要元素有三：電子化館藏、電子化作業和電子化服務。由於個人化和社群化乃是圖書館電子化服務的發展趨勢，為提供圖書館界發展個人化與群體化服務的參考，本論文探討如何以圖書館自動化系統為基礎，在圖書館建置個人化與社群化的服務。本文主旨如下：(1) 提出以圖書館自動化系統為基礎之個人化與群體化服務的系統架構；(2) 研究動態追蹤使用者興趣的相關理論，並將之應用在個人化服務；(3) 將資料探勘應用於圖書館館藏借閱共同性及順序性的探索，以提供群體化服務；(4) 根據以上的理論實作一套具備個人化與群體化功能的交通大學個人化圖書資訊環境系統。

關鍵詞：多電子圖書館、個人化資訊環境、社群、動態偵測使用者興趣、個人興趣關聯圖、資料探勘、相關規則探勘、循序規則探勘

一、緒論

近年來，電子圖書館 (Electronic Library, or E-Library) 的觀念已經深深衝擊圖書館的角色、任務與運作。^[1] 電子圖書館之主要目的在於運用電腦與網路科技，輔助圖書館業務之執行與讀者需求之滿足，並加速讀者知識的產出。我們認為電子圖書館的構成要素有三，即：電子化館藏 (E-Collection)、電子化作業 (E-Operation)，以及電子化服務 (E-Service)，而此三要素又以圖書館自動化系統為核心。^[2]

另一方面，電腦與網路科技的蓬勃發展，帶動了電子商務 (Electronic Commerce) 的熱潮。電子商務的經營者皆期望經由網際網路接觸更廣大的客戶，更無遠弗屆地深入所有消費族群。然而，若想在眾多電子商務經營者中脫穎而出，其關鍵就在於是否能了解客戶的需求，提供每位客戶量身訂作之個人化服務，亦即採用客戶關係管理 (Customer Relationship Management, CRM) 的理念，並利用電腦系統紀錄並分析客戶的需求，從而建立一對一之個人化服務，以期增加客戶的滿意度與忠誠度，並確保在激烈商業競爭的環境裡獲利。^{[3][4]}

儘管圖書館並非以營利為目的之商業機構，然而圖書館的最終目的，乃在於使讀者有效地利用圖書館的資料，協助讀者獲取資訊、運用資訊，從而產生知識。因為每位讀者都有其特別的資訊需求，圖書館的服務應該把每位讀者視為不同的個體，儘量去滿足每一讀者個別的資訊需求。從這個觀點來看，若能採用 CRM 的理念並推動個人化服務，相信必能提升圖書館對讀者的服務。

與個人化相對的概念是群體化。有時知識的產生不能光靠單一個體，而是得藉由具有相同興趣、專長的個體組成社群 (Community)，彼此激發靈感與分享心得，方能促成知識的產生。表面上看來，個人化和群體化似乎是互斥的，但實際上對圖書館的服務而言，卻是一體兩面。

[1] 陳亞寧，另類圖書館：電子圖書館綜觀，《資訊傳播與圖書館學》，5：3 (1999.3)，頁 59-73。

[2] 蔡淑琴、柯皓仁，圖書館自動化系統轉換經驗談與電子圖書館時代之自動化系統 - 以交通大學圖書館為例，《圖書館自動化系統的新發展及系統轉換研討會》(臺北縣淡水鎮：淡江大學覺生紀念圖書館，2000.10)，頁 65-75。

[3] 蔣以仁，一對一個人化服務機制，《電腦與通訊》，95 (2001.3)，頁 88-93。

[4] 張進群、陳建良，客戶關係管理，《機械工業雜誌》，89：12 (2000.12)，頁 165-176。

由於電子圖書館三要素的核心為圖書館自動化系統（以下簡稱自動化系統），在自動化系統中儲存了圖書館館員專業知識的成果（編目資料）與讀者背景資料，且記錄了館藏的流通狀況及流通歷史，因此當圖書館界欲實施CRM 理念並推動個人化和群體化服務之時，若能以自動化系統為基礎，將可收事半功倍之效。觀察現有圖書館自動化系統所提供的功能，大多侷限於圖書館本身作業的自動化，極少部分的功能是著眼於讀者服務，遑論個人化與群體化服務。以讀者最常接觸的線上公用目錄（Online Public Access Catalog，OPAC，或稱館藏查詢系統）而言，除館藏查詢之外，僅具備基本的個人化服務，如館藏續借、館藏預約、借閱狀況查詢、讀者基本資料查詢與修改等。至於群體化服務更是付之闕如。

本文的主旨在闡述如何以圖書館自動化系統為基礎，設計個人化與群體化服務。在個人化服務方面所採用的技術包括動態偵測個人興趣（Dynamically Tracking User Interests）和資訊過濾（Information Filtering）；在群體化服務方面則是採用資料探勘（Data Mining）的技術。文中亦介紹我們所實作的個人化圖書資訊環境雛形——PIE@NCTU。本文組織如下：第二節探討個人化以及群體化服務的相關文獻；第三節提出以圖書館自動化為基礎之個人化與群體化資訊服務的系統架構；第四節提出動態偵測讀者興趣的演算法；第五節敘述如何將資料探勘技術應用於群體化服務；第六節則是介紹我們實際開發的個人化與群體化圖書資訊系統——PIE@NCTU；第七節則為結論與未來發展方向。

二、文獻探討

本節分別針對個人化以及群體化服務的相關文獻加以探討。

（一）個人化服務

目前許多網際網路服務都提供了個人化服務，例如：「我的 Yahoo! 奇摩」（<http://tw.mykimo.yahoo.com>）、「MyNetscape」（<http://my.netscape.com>）等；與圖書館相關的個人化服務則有「MyLibrary」^[5]、「MyLibrary@NCState」^[6]、

[5] S. Cohen, J. Ferreira, A. Horne, B. Kibbee, H. Mistlebauer, and A. Smith, "MyLibrary: Personalized Electronic Services in the Cornell University Library," *D-Lib Magazine*, April 2000.

[6] K. Morgan and T. Reade, "Pioneering Portals: MyLibrary@NCState," *Information Technology and Libraries* 19:4 (2000), pp. 191-198.

「MyGateway」^[7]等。

French 與 Viles 二位學者在 1999 年提出個人化服務環境的架構。^[8] 概括來說，個人化服務環境應該要具備以下條件：（1）個人化的使用者界面（Customizable User Interface），讓使用者依自己的喜好組織使用環境；（2）有效的檢索（Effective Search），能提高檢索結果的正確性，引導使用者尋找資料，提高查全率（Recall）；（3）確保使用者的隱私權（Privacy）。我們認為現階段的個人化服務系統應該具備以下條件：

1. 個人化使用界面

個人化使用界面讓使用者依據個人的喜好來規劃其使用界面，前述幾種個人化服務均具備讓使用者設定個人化使用界面的功能。

2. 個人資料紀錄

個人化服務的每位使用者都有其個人的資料紀錄（User Profile），儲存背景、興趣、學科專長、檢索歷史等資料，做為個人化服務的主要依據，其中尤以個人興趣紀錄最為重要。個人興趣紀錄的主要來源有二：（1）由讀者人工填寫；（2）運用電腦科技推導自動產生。

若個人興趣紀錄是由人工填寫產生的，則此類系統均提供一個興趣關鍵字詞的輸入界面，讓使用者能自行設定感興趣的關鍵字。像 MyLibrary@NCState 以及日本奈良先端科學技術大學院大學附屬電子圖書館 (<http://dlw3.aist-nara.ac.jp>) 即是根據讀者自行設定的關鍵字詞提供資訊選粹服務。這種作法雖然很直覺（因為是由使用者輸入興趣資料），但是使用者往往會選擇過於普遍的字詞來描述自己的興趣^[9]，導致對於興趣的描述不夠精確。再者我們亦不能期望使用者都是勤勞、有足夠耐心，且總是能正確輸入關鍵字詞的。

至於自動產生個人興趣紀錄的系統，通常是根據使用者的使用歷程或特定行為來推導興趣，常見的方法有：（1）將使用者瀏覽過的網頁中所含的關鍵字詞

[7] My Gateway at University of Washington, from <http://www.lib.washington.edu/resource/help/MyGateway.html>.

[8] J. C. French and C. L. Viles, "Personalized Information Environments: An Architecture for Customizable Access to Distributed Digital Libraries", *D-Lib Magazine*, June 1999.

[9] I. B. Crabtree and S. J. Soltysiak, "Identifying and tracking changing interests", *International Journal on Digital Libraries* 2 (1998), pp. 38-53.

記錄下來當成使用者的興趣^[10]；(2) 利用電子郵件通信紀錄來抽取關鍵字詞當成使用者興趣^[11]；(3) 以交易行為推導興趣（例如：Amazon 會以顧客買過的書之關鍵字為興趣）

在本文中，我們將提出一個能考慮時間因素，顧及現在與過往興趣，並能偵測使用者改變興趣的演算法，用半自動的方式協助使用者建立個人興趣紀錄。

3. 資訊選粹服務

有了個人興趣紀錄之後，便可提供資訊選粹（Selective Dissemination of Information, SDI）服務。資訊選粹利用資訊過濾技術分析出個別使用者感興趣的新進資訊。資訊過濾技術可分為內容式資訊過濾（Content-based Information Filtering）以及協力式資訊過濾（Collaborative Information Filtering）二種。^[12]內容式資訊過濾主要是以資訊的內容為過濾的依據並加以分析比較，使用者在興趣檔中只要記錄感興趣的關鍵字詞，系統便會比對新進資訊和興趣檔，以達成資訊過濾。諸如：「LA Times Custom News Services」（<http://www.latimes.com/>）「MyLibrary @NCState」均採用內容式資訊過濾來提供資訊選粹服務。

協力式資訊過濾不直接分析資訊內容，而是找出與使用者背景、知識、興趣接近的同好或社群，再針對使用者的查詢主題，從這些同好或社群成員感興趣的資訊中，分析並選取最可能相關的資訊提供參考。^{[13] [14] [15]}

內容式資訊過濾與協力式資訊過濾並不是完全互斥的，由於二種方法各有其優缺點，若能將二者結合，可以突顯二者的優點、抵銷二者的缺點。^{[16] [17]}

[10] 同註 9。

[11] 同註 9。

[12] 卜小蝶，網路資訊過濾技術與個人化資訊服務，《21世紀資訊科學與技術國際研討會》（臺北：世界新聞傳播學院圖書資訊學系，1996.11.7-9），頁339-350。

[13] 卜小蝶，提供個人化服務的線上公用目錄檢索系統初探，《中國圖書館學會會報》，59（1997.12），頁127-133。

[14] R. Mooney and L. Roy, "Content-Based Book Recommending Using Learning for Text Categorization," *Proceedings of the Fifth ACM Conference on Digital Libraries*, (2000), pp.195-204.

[15] M. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligence Review*, (1999), pp.393-408.

[16] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as Classification: Using Social and Content-Based Information in Recommendation," *Proceedings of the AAAI-98*, (Madison, WI: AAAI Press, 1998), pp. 714-720.

4. 個人化檢索

所謂個人化檢索是指能依個別使用者的背景、興趣或需求，幫助使用者尋找資訊。個人化檢索的可能應用有：(1) 提供適合使用者的背景、年齡等因素的資訊，例如同樣是檢索「網路多媒體」這個主題，提供給小學生和提供給電腦相關系所博士生的資訊就應該要加以區別^[18]；(2) 將檢索結果根據使用者的個人興趣加以排序(Ranking)，把使用者感興趣的檢索結果排列在較明顯的位置。

根據個人興趣紀錄提供資訊選粹服務的系統頗多，然而將其應用在個人化檢索上卻很少見，這對建置個人化資訊環境是一個很大的遺憾。有鑑於此，我們在本文中提出一個利用個人興趣紀錄及資訊過濾技術達成檢索個人化的方法。

(二) 資料探勘與群體化服務

資料探勘(Data Mining)也叫做資料庫探勘(Database Mining)或資料庫知識發掘(Knowledge Discovery in Database)。簡單地說，資料探勘是從儲存於資料庫(Database)、資料倉儲(Data Warehouse)或其他資訊儲存器(Information Repository)的大量資料中，發掘出具有價值的知識之過程。資料探勘在近年來廣泛地運用在各種領域或行業，例如：行銷、財務、銀行、製造、通訊、保險等，用以發掘潛在客戶、管理異常狀況、管理客戶關係、或作為企業決策的參考。例如：超級市場能將資料探勘運用於發掘顧客的消費模式，並利用所發掘的消費模式研擬促銷或貨物排架策略，以提升超級市場的業績。

Cabena 等人提到^[19]，資料探勘包含下列幾個步驟：

1. 確定目標：雖然定出目標是很直覺的一件事，但在資料探勘中卻是不可缺少的一部分，所以必須明確地定義出問題所在及質疑的地方。
2. 預備資料：此步驟是資料探勘處理過程中最花時間的一個步驟，在整個資料探勘的過程中，大約佔了60%的時間。此步驟又分為三部分：
 - (1) 資料選取：從所有的來源資料中選擇需要的資料。
 - (2) 資料的前置處理：確保選取資料的品質，使分析更容易。

[17] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining Content-Based and Collaborative Filters in an Online Newspaper," *Proceedings of ACM SIGIR Workshop on Recommender Systems*, 1999.

[18] 同註13。

[19] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, and A. Zanasi. "Discovering Data Mining: From Concept to Implementation," Prentice Hall, Upper Saddle River, NJ, 1998.

- (3) 資料轉換：根據下一步驟擬採用的資料探勘演算法，將資料轉換成適合的格式。
3. 資料探勘：選擇適當的資料探勘演算法來探勘上一步驟處理過後的資料。
 4. 分析結果：評估步驟3所探勘出來的資訊和知識。
 5. 產生知識：將步驟4所分析出來的資訊和知識融入企業化的組織或資訊系統中。

資料探勘主要功能有分類規則歸納(Classification) 推估分析(Estimation) 預測分析(Prediction) 相關規則探勘(Association Rules Mining) 同質分群(Clustering) 等五種，關於此五種功能的意義及可能使用的技巧，請參見陳淑燕撰寫之「資訊勘測初探」一文。^[20]

在本文中，我們將資料探勘中的相關規則探勘和循序規則探勘(Sequential Pattern Mining) 應用在圖書館的群體化服務，以自動化系統中的書目、館藏、借閱與預約歷史紀錄為來源資料，探索讀者的社群特性。本文擬探索的讀者社群關係包含館藏借閱的共同性及順序性。以下簡單介紹相關規則探勘的基本概念。

相關規則探勘最常應用在商店交易紀錄資料庫，用以分析並發掘顧客的交易模式，並根據發掘出來的交易模式採取適當的行銷策略，以提升商品的銷售率。例如：20% 買牙刷的顧客也會買牙膏、毛巾、和香皂就是一個典型的相關規則。^[21] 相關規則探勘的正規敘述^[22] 如下：

令 $I = \{i_1, i_2, \dots, i_m\}$ 是由交易項目(Item) 組成的集合，由一個或一個以上的項目所組成的集合稱為項目集(Itemset)。令資料庫 D 是由一群交易(Transaction) T 所組成的集合，每個 T 為一項目集，代表交易紀錄， $T \subseteq I$ ，每個交易紀錄有其唯一的識別碼，稱為 TID。如果 $X \subseteq I$ 且 $X \subseteq T$ ，則稱 T 包含(Contain) X 。以商店的應用來看，每一種商品就是一個交易項目，一個顧客在某次來訪時中所購

[20] 陳淑燕， 資訊勘測(Data Mining) 初探， 《國立臺灣師範大學圖書館通訊》，41 (2000.4)，頁21-23。

[21] 當商店發掘出「20% 買牙刷的顧客也會買牙膏、毛巾、和香皂」這個相關規則時，便能夠擬定行銷方案，提升這些商品的銷售率。可能的行銷方案包括：(1) 把上述四種商品排在一起—讓顧客快速且方便地取得此四種商品；(2) 把上述四種商品散置於商店各處—讓顧客必須逛遍整個商店才能夠取得這四種商品，如此有可能在逛的過程中多買一些其他的商品；(3) 把這四種商品結合在一起銷售，例如同時買這四種商品打85折。

[22] R. Agrawal, T. Imieliski, and A. Swami. "Mining Association Rules between Sets of Items in Large Databases," *Proceedings of the 1993 ACM SIGMOD Conference*, pp.207-126.

買的商品所成的集合即為一交易。

一個相關規則 (Association Rule) 表示成 $X \rightarrow Y$, 其中 $X \subseteq I, Y \subseteq I, X \cap Y = \emptyset$ 。若 D 中包含 X 的交易裡有 $c\%$ 也同時包含了 Y , 我們就說規則 $X \rightarrow Y$ 的確信值 (Confidence) 為 $c\%$; 如果 D 中包含 $X \rightarrow Y$ 的交易紀錄有 $s\%$, 我們就說規則 $X \rightarrow Y$ 的支持度 (Support) 為 $s\%$ 。相關規則探勘定義為: 給定交易紀錄資料庫 D , 在當中找出所有確信值和支持度大於最小確信值跟最小支持度的規則。

Agrawal 等人^[23] 將相關規則探勘分為二個子問題:

1. 子問題 1: 找到所有支持度大於最小支持度的項目集。支持度大於最小支持度的項目集稱為大項目集 (Large Itemset)。
2. 子問題 2: 用子問題 1 中所找到的大項目集來產生所期望的規則。此步驟的演算法非常直接, 即: 對於任一大項目集 L , 找出其所有非空子集合。對於每個非空子集合 a , 如果規則 $a \rightarrow (L-a)$ 的確信值 (也就是 $\text{support}(L) / \text{support}(a)$) 大於最小確信值, 則此規則即符合所求。

表一為一個交易資料庫範例。在此交易資料庫中共包含四筆交易, 其交易紀錄識別碼分別為 100、200、300、400。此交易資料庫中共有五種交易的項目 (或商品): 牙刷、牙膏、毛巾、手錶、香皂, 為方便電腦運算起見, 我們將這五種商品依序賦予 1、2、3、4、5 的代碼。假設最小支持度為 50%, 則只要出現在交易資料庫二次或以上的項目集便為大項目集, 因此如表二所示, 本範例中的大項目集包含 {1}、{2}、{3}、{5}、{1, 3}、{2, 3}、{2, 5}、{3, 5}、{2, 3, 5}; 本範例中可能出現的相關規則列於表三, 以 {2, 5} \rightarrow {3} 此一相關規則來看, 其確信值為 67%, 這代表有 67% 買 2 和 5 的客人也會同時買 3 (在 200, 300, 400 這三筆交易中都有 2 和 5, 但是只有 200 和 300 這兩筆交易有 3), 而同時購買 2, 3, 5 的交易有二次 (即 200 和 300 這兩筆交易)。

若設定之最小確信值為 100% (在實際狀況下不可能如此設定), 則探勘得到的相關規則有: {1} \rightarrow {3}、{2} \rightarrow {5}、{5} \rightarrow {2}、{2, 3} \rightarrow {5}、{3, 5} \rightarrow {2} 等五個規則, 分別代表 {牙刷} \rightarrow {毛巾}、{牙膏} \rightarrow {香皂}、{香皂} \rightarrow {牙膏}、{牙膏, 毛巾} \rightarrow {香皂}、{毛巾, 香皂} \rightarrow {牙膏}。

[23] 同註 22。

表一：交易資料庫範例

資料庫 <i>D</i>	
交易紀錄識別碼 (TID)	項目集 (Itemset)
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

表二：大項目集 (最小支持度為2)

大項目集	支持度
{1}	2
{2}	3
{3}	3
{5}	3
{1, 3}	2
{2, 3}	2
{2, 5}	3
{3, 5}	2
{2, 3, 5}	2

表三：相關規則

相關規則	確信值
{1} {3}	100%
{3} {1}	67%
{2} {3}	67%
{3} {2}	67%
{2} {5}	100%
{5} {2}	100%
{3} {5}	67%
{5} {3}	67%
{2} {3, 5}	67%
{3} {2, 5}	67%
{5} {2, 3}	67%
{2, 3} {5}	100%
{2, 5} {3}	67%
{3, 5} {2}	100%

儘管從前面的例子來看，相關規則探勘似乎只是簡單的統計，然而當資料庫中具有大量的項目與交易資料時，潛在的（Potential）大項目集會是呈指數（Exponential）成長，而如何在潛在的大項目集中找出真正的大項目集以及相關規則，便需要複雜的演算法（Algorithm）來達成，而非僅是簡單的統計。

從相關規則探勘延伸出各式各樣的問題，其解決方法和基本的問題息息相關，且可應用在更多領域。以下簡述目前的相關研究：^[24]

1. 有些研究考慮到顧客多次交易的狀況，同一位顧客會有先後時間關係的交易紀錄。此類研究的目的是想要在這些交易紀錄中找出顧客最常買的商品序列（Sequence），此序列是由項目集所組成，而每一個項目集內的項目是不考慮個數且沒有順序關係的。此種相關規則探勘又稱為循序規則探勘（Sequential Pattern Mining）。
2. 支持度和確信值是用來控制產生規則多寡的兩個限制，除了這二個限制外，有些研究還定義更多的限制，繼而利用這些限制的特性來加速探勘。
3. 在實際的應用環境中，資料庫內的交易紀錄筆數會隨著時間而增加或減少，而資料探勘要耗費龐大的計算時間，若是每次資料庫一更新就重新探勘，將會浪費很多的時間；相對地若只探勘資料庫更新的部份便可節省很多時間，因此有些研究希望只探勘資料庫增加或減少的部分就能得到所有的規則，此即漸進更新（Incremental Update）問題。
4. 有些研究除了考慮交易紀錄之間的順序關係外，也探討交易紀錄內的項目順序關係。
5. 有些研究將交易紀錄中每一種交易項目的個數考慮進去，或者將交易項目分類，以分類來當相關規則的項目。

三、系統架構

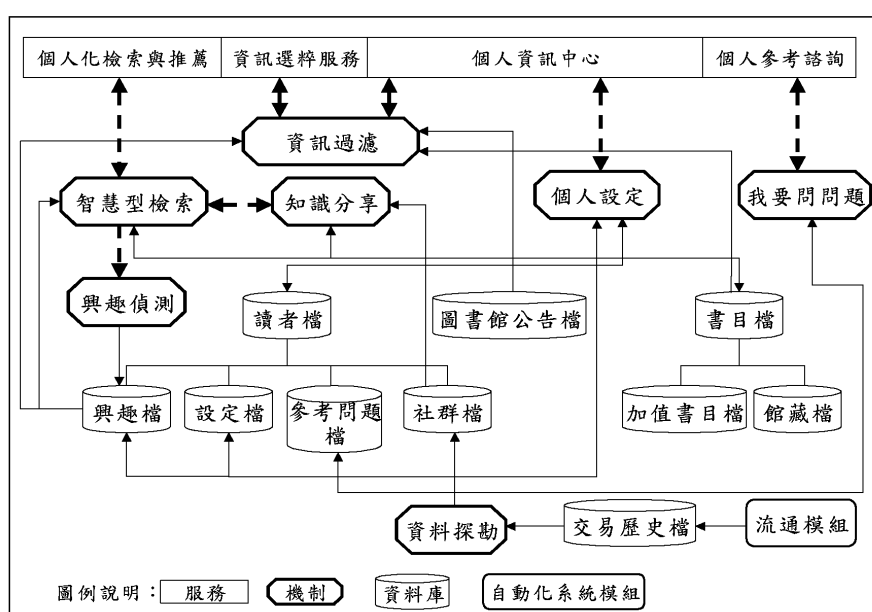
本節提出在圖書館自動化系統的基礎之上，建構個人化與群體化服務的系統架構。系統架構如圖一所示，包含服務、機制以及資料庫等三個層面。個人化與群體化服務和自動化系統的其他模組^[25]有著密切的關係，其中較重要的關係

[24] 林高煌，「一個有效率的大參考序列探勘方法及其在全球資訊網上的應用」，（新竹：國立交通大學資訊科學系碩士論文，2000.6）。

[25] 各層級圖書館資訊系統規範工作小組，各層級圖書館自動化網路系統軟硬體規範，教育部

有：

1. 為線上公用目錄查詢模組的延伸；
2. 為參考服務暨文獻傳遞模組的延伸；
3. 以流通模組的交易歷史檔為資料來源，運用資料探勘技術探索社群關係；
4. 整合流通模組的讀者檔；
5. 整合編目與期刊模組的書目與館藏檔。



圖一：以圖書館自動化為基礎之個人化與群體化服務系統架構

以下我們說明在此個人化與群體化服務中提供的相關服務——個人資訊中心、個人化檢索與推薦、資訊選粹服務以及個人參考諮詢，並指出與每一項服務相關的機制和資料庫。

(一) 個人資訊中心

顧名思義，個人資訊中心讓使用者組織其認為有價值的資訊，以及圖書館認為需要讓使用者知道的資訊。個人資訊中心包含現今圖書館自動化系統中既有的讀者基本資料與權限資料、借閱狀況、預約狀況、歷史借閱狀況等資訊；除此之

外，還包含個人化桌面、個人書籤、個人興趣、個人電子報 等個人化服務的使用及設定。

與個人資訊中心相關的機制包括：(1) 資訊過濾機制；以及(2) 個人桌面、書籤、興趣的設定機制。相關的資料庫則有讀者檔、讀者興趣檔、讀者設定檔、圖書館公告檔、流通模組的交易檔及交易歷史檔。

(二)個人化檢索與推薦

個人化檢索依個別使用者的背景、興趣或需求，提供使用者量身訂作的資訊檢索服務。實現個人化檢索的可行方法有：(1) 考量使用者的背景、年齡等因素，提供不同的檢索結果；或(2) 將檢索結果依照使用者的興趣加以排序。

若將個人化檢索與社群知識分享機制結合，可進一步達成個人化推薦。^[26]可行的方法之一是將相關規則探勘和循序規則探勘套用在圖書館的借閱與預約交易歷史紀錄，藉以發掘出一些相關規則和循序規則，並將之儲存於社群檔中。例如：「有 5% 借閱過大漠英雄傳的讀者也會借閱成吉思汗傳」(相關規則)、「有 5% 借閱過大漠英雄傳、神鵬俠侶、倚天屠龍記的讀者是依照大漠英雄傳 神鵬俠侶 倚天屠龍記的順序來借閱」(循序規則)，有了這些規則之後，當有一位讀者檢索(或借閱)大漠英雄傳時，系統便可以推薦讀者也檢索(或借閱)成吉思汗傳，同時，也可以建議讀者依照大漠英雄傳 神鵬俠侶 倚天屠龍記的順序來閱讀這三本書。

此外，為促進知識分享，我們認為應該多鼓勵讀者撰寫書評、參與討論、成為積極活躍的資訊提供者。為達到此一目的，系統必須具備一個加值書目檔，提供每一館藏之加值資訊，而這些加值資訊的來源不是編目館員，而是讀者。加值書目檔的主要內容包含：

1. 屬性欄位，由讀者根據其個人主觀價值給予評等，如某讀者可能給予某本書以下評等：家庭倫理度(0.9)、悲劇度(0.5)、人性關懷度(0.1)、科幻驚悚度(0.0)、五星評等()
2. 不定長欄，由讀者分享相關經驗，如讀後感想或評語、留給後人的話、錯誤校正、推薦其他參考書籍或論文、相關網站連結等。

當加值書目檔擁有豐富資訊之後，將能夠進一步促進智慧型檢索的發展。現

[26] 同註 13。

今館藏查詢系統的使用者必須要知道所查詢館藏的基本資料，如書名、作者等，因此當使用者若在館藏查詢系統中以「中國家庭倫理悲劇性的小說」為檢索策略尋找資料時，將無法找到任何符合的館藏。但有了加值書目檔，配合自然語言處理（Natural Language Processing），智慧型檢索系統能夠在前述檢索策略中切出屬性值「中國」、「家庭倫理」、「悲劇」、「小說」，再根據其他讀者給予的屬性評等執行模糊查詢（Fuzzy Search），如此一來可能會找出紅樓夢、梁山伯與祝英台等相關館藏。同理，當使用者輸入紅樓夢進行檢索時，除了紅樓夢之外，智慧型檢索系統可能還會找出西遊記、曹雪芹傳、清史、甚至是雍正王朝VCD。

與個人化檢索與推薦相關的機制包括：智慧型檢索、資訊過濾、興趣偵測、知識分享、資料探勘；相關的資料庫則有：讀者檔、讀者興趣檔、社群檔、書目檔、館藏檔、加值書目檔、流通模組的交易歷史檔。

（三）資訊選粹服務

資訊選粹服務能依據個別使用者的需求與興趣過濾出新進資訊，並將過濾結果傳送給個別使用者。在圖書館中新進資訊的來源包含新進館藏、圖書館公告等。

與資訊選粹服務相關的機制主要為資訊過濾；相關的資料庫則有讀者檔、讀者興趣檔、讀者設定檔、書目檔、圖書館公告檔等。

（四）個人參考諮詢

儘管有部分學者認為電子圖書館和個人化服務的發展可能會導致圖書館參考服務的困難、使圖書館參考服務的角色降低，然而我們認為參考諮詢館員和讀者間的互動即使在電子圖書館時代仍然是非常重要的。在電子圖書館時代的參考諮詢服務亦必須走向電子化、網路化。所謂的個人參考諮詢就是讓讀者線上輸入問題內容、用途、服務類型、資料類型、回覆時限等，由參考諮詢館員負責答覆。讀者的問題與館員的解答並儲存於資料庫中，讓館員得以後續整理為常問問題（FAQ）供讀者瀏覽及查詢。

與個人參考諮詢相關的機制主要為我要問問題；相關的資料庫則是參考問題檔。

四、動態學習使用者興趣

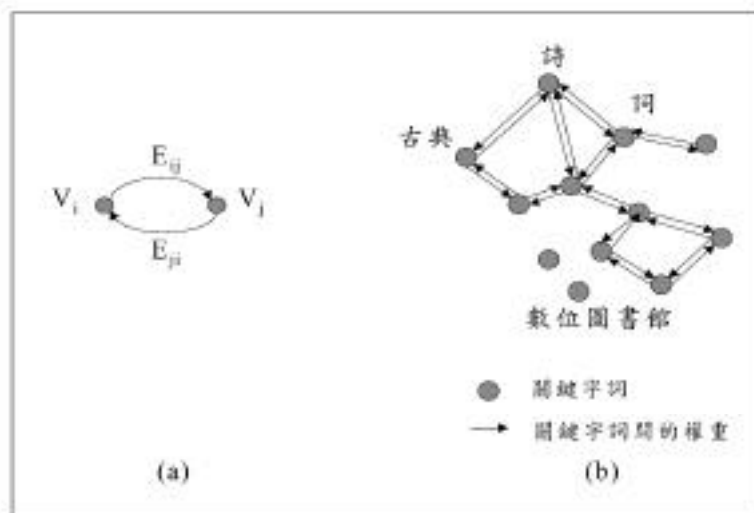
個人化服務的關鍵技術在於如何準確得知使用者的興趣。在圖書館自動化系

統（如線上公用目錄）中，使用者的資訊需求經常是透過檢索來滿足的，因此，我們認為使用者的興趣或資訊需求能夠從其曾經用過的檢索策略來解讀。基於此一概念，在本論文中，我們考慮個別使用者曾經在檢索策略中用過的關鍵字詞的頻率、各關鍵字詞間的相關性，以及時間對興趣的影響，來動態學習其興趣。

本節先提出一個動態學習使用者興趣的演算法，並透過一個範例來說明演算法；接著說明如何根據此演算法實作資訊選粹服務和個人化檢索。

（一）個人興趣關聯圖

若要提出一個能有效偵測使用者興趣的方法，首先必須對「興趣」的特性有所了解。我們認為每個人的興趣通常會維持一段時間，且隨著歲月流逝而有所改變，因此，偵測使用者興趣的演算法必須要能隨著時間修正使用者的興趣（興趣加強或減弱）。如前所述，本文假設使用者的興趣可由其曾使用過的檢索策略來解讀，在綜合考量時間因素和檢索行為的情形下，我們針對每位使用者曾用過的檢索策略，建立個人興趣關聯圖（Personal Keyphrase-Relationship Graph，簡稱 PKRG）。個人興趣關連圖的主要用途在於計算單一使用者曾用過的各關鍵字詞的權重，以及該使用者認為各關鍵字詞間的關聯性，從而推演使用者的興趣。



圖二：個人興趣關聯圖（Personal Keyphrase-Relationship Graph, PKRG）

1. 個人興趣關連圖的建立

對於每一位使用者，我們會為其建立專用的個人興趣關聯圖（PKRG）。

PKRG 為一有向圖 (Directed Graph), 圖二為 PKRG 的範例。如圖二(a)所示, PKRG 中的每一個端點 (Vertex) V_i 代表使用者曾在檢索策略中用過的關鍵字詞; 每一條邊線 (Edge) E_{ij} 表示使用者曾用過以該邊線兩端點 (V_i 及 V_j) 代表之關鍵字詞做「且 (AND)」運算所產生的檢索策略 (例如: 「詩AND 古典」)。

我們為 PKRG 中的每個端點和邊線都賦予權重 (Weight): 端點 V_i 的權重代表使用者對關鍵字詞 V_i 感興趣的程度, 邊線 E_{ij} 的權重則代表使用者對「 V_i AND V_j 」此檢索策略感興趣的程度, 同時也代表使用者認為 V_i 和 V_j 這兩個關鍵字詞的關聯性。系統在使用者執行一檢索策略時, 便會自動修正該使用者的 PKRG。

2. 權重計算

PKRG 中的每個端點和邊線都具有一個權重: 端點的權重代表使用者對相對應的關鍵字詞感興趣的程度; 邊線的權重則代表使用者對相對應的檢索策略感興趣的程度, 同時也代表相對應的二關鍵字詞的關聯性。計算權重時需考量三個因素: (1) 單一關鍵字詞出現在檢索策略中的次數; (2) 單一檢索策略曾使用的次數; 以及 (3) 時間對使用者興趣的影響。我們假設時間對使用者興趣的影響可依固定時間劃分成若干區段, 在此假設下, 我們採用下列步驟來計算 PKRG 中每一個端點 (代表關鍵字詞) 和邊線 (代表檢索策略) 的權重。

(1) 計算個人關鍵字詞的權重

計算單一時間區段內曾使用過的關鍵字詞之權重;

計算各個時間區段內時間對興趣的影響係數;

結合各個時間區段內的個人關鍵字詞權重, 以及時間對興趣的影響係數, 求得整體的關鍵字詞權重。

(2) 計算個人檢索策略的權重

計算單一時間區段內曾使用過的檢索策略之權重;

計算各個時間區段內時間對興趣的影響係數;

結合各個時間區段內的個人檢索策略權重, 以及時間對興趣的影響係數, 求得整體的檢索策略權重。

接下來我們將詳細介紹計算個人關鍵字詞和檢索策略權重的方法。

(1) 單一時間區段內關鍵字詞的權重計算

我們將單一時間區段內某關鍵字詞的使用頻率加以正規化之後所得的值, 作

為單一時間區段內該關鍵字詞的權重。計算公式如下：

$$W_{ik} = \frac{TF_{ik}}{\sum_{j=1}^{n_k} TF_{jk}}$$

W_{ik} ：第 k 個時間區段中，關鍵字詞 i 的權重。

TF_{ik} ：第 k 個時間區段中，關鍵字詞 i 的使用頻率。

n_k ：第 k 個時間區段中，用過的關鍵字詞總數。

(2) 單一時間區段內的檢索策略權重計算

我們依據關鍵字詞和檢索策略的頻率來計算單一時間區段內檢索策略的權重。公式如下：

$$Sim_{ijk} = \frac{EF_{ijk}}{\sqrt{TF_{ik} \times TF_{jk}}}$$

Sim_{ijk} ：在第 k 個時間區段中，「 i AND j 」這個檢索策略的權重。

TF_{ik} ：在第 k 個時間區段中，檢索策略中含有關鍵字詞 i 的正規化頻率。

TF_{jk} ：在第 k 個時間區段中，檢索策略中含有關鍵字詞 j 的正規化頻率。

EF_{ijk} ：在第 k 個時間區段中，檢索策略中含有「 i AND j 」這個檢索策略的正規化頻率。

m_k ：在第 k 個時間區段中，用過的檢索策略總數。

(3) 時間對興趣的影響係數

儘管使用者通常會隨著時間的流逝改變其興趣，但我們假設使用者的興趣在某一段固定時間內都不會有所改變，且越久之前的興趣紀錄對使用者而言越不重要。為了表現時間流逝對興趣的影響，我們將使用者用過的檢索策略依固定時間間隔劃分成若干區段，每一時間區段對使用者整體興趣的影響呈半衰期遞減。假設將時間劃分成 n 個區段，則在第 k 區段中 (k 值越大時間越遠)，時間對使用者整體興趣的影響率為：

$$HI_k = \frac{2^{n-k}}{2^n - 1} \quad \text{where } 1 \leq k \leq n$$

HI_k ：第 k 時間區段的半衰期時間係數。

例如當 $n=5$ 時，五個時間區段的半衰期時間係數為 $(HI_1, HI_2, HI_3, HI_4, HI_5)$
 $= (16/31, 8/31, 4/31, 2/31, 1/31)$

(4) 個人關鍵字詞權重計算

在將使用者檢索歷程劃分成若干時間區段、個別計算單一時間區段內關鍵字詞的權重後，我們將之合併計算，求得整體的個人關鍵字詞權重。計算公式如下：

$$W_i = \sum_{k=1}^n (W_{ik} \times HI_k)$$

W_i ：關鍵字詞 i 的權重。

W_{ik} ：第 k 個時間區段中，關鍵字詞 i 的權重。

HI_k ：第 k 時段中時間對興趣的影響係數。

n ：劃分的時間區段數目。

(5) 個人檢索策略權重計算

在將使用者檢索歷程劃分成若干時間區段、個別計算單一時間區段內檢索策略的權重後，我們將之合併計算，求得整體的個人檢索策略權重。計算公式如下：

$$W_{ij} = \sum_{k=1}^n (Sim_{ijk} \times HI_k)$$

W_{ij} ：「 i AND j 」這個檢索策略的權重。

Sim_{ijk} ：在第 k 個時間區段中，「 i AND j 」這個檢索策略的權重。

HI_k ：第 k 時區中時間對興趣的影響係數。

n ：劃分的時間區段數目。

我們用一個範例來說明前述的演算法，在這個範例中，我們以二週為一個時間週期。表四為該使用者六週內用過的所有檢索策略與次數。由表四可算出使用

者用過的關鍵字詞之權重（結果如表五）以 Digital Library 為例，在每一週結束時 Digital Library 的權重如下：

$$\begin{aligned}
 \text{第一週} & \quad \frac{12}{20} = 0.6000 \\
 \text{第二週} & \quad \frac{18}{30} = 0.6000 \\
 \text{第三週} & \quad \frac{12}{20} \times \frac{2}{3} + \frac{12}{20} \times \frac{1}{3} = 0.6000 \\
 \text{第四週} & \quad \frac{6}{15} \times \frac{2}{3} + \frac{18}{30} \times \frac{1}{3} = 0.4666 \\
 \text{第五週} & \quad \frac{2}{23} \times \frac{4}{7} + \frac{12}{20} \times \frac{2}{7} + \frac{12}{20} \times \frac{1}{7} = 0.3068 \\
 \text{第六週} & \quad \frac{2}{25} \times \frac{4}{7} + \frac{6}{15} \times \frac{2}{7} + \frac{18}{30} \times \frac{1}{7} = 0.2457
 \end{aligned}$$

以下，我們說明 Digital Library 在第六週結束時的權重計算方式。在第六週時，檢索歷程可分為三個週期，第五、六週為第一週期，第三、四週為第二週期，第一、二週為第三週期。在第一週期中，半衰期時間係數為 $4/7$ ，Digital library、Interface、Literature、Dumas、Shakespeare 被使用過的次數分別為 2、2、8、4、9 次，因此所有關鍵字詞出現的總次數為 25 次，Digital Library 在第一週期的權重為 $2/25$ 。

在第二週期中，半衰期時間係數為 $2/7$ ，Digital library、Interface、Literature 被使用過的次數分別為 6、8、1 次，因此所有關鍵字詞出現的總次數為 15 次，Digital Library 在第一週期的權重為 $6/15$ 。

在第三週期中，半衰期時間係數為 $1/7$ ，Digital library、Interface、IEEE 被使用過的次數分別為 18、6、6 次，因此所有關鍵字詞出現的總次數為 30 次，Digital Library 在第一週期的權重為 $18/30$ 。

結合三個週期內 Digital Library 的權重以及時間對興趣的影響係數，Digital Library 在第六週結束時的權重為 $2/25 \times 4/7 + 6/15 \times 2/7 + 18/30 \times 1/7 = 0.2457$ 。

由表四也可算出使用者用過的檢索策略的權重（結果如表六）。以 Digital Library AND Interface 這個檢索策略為例，在每一週結束時的權重如下：

$$\begin{aligned} \text{第一週} \quad & \frac{6}{12} / \sqrt{\frac{12}{12} \times \frac{6}{12}} = 0.7071 \\ \text{第二週} \quad & \frac{6}{18} / \sqrt{\frac{18}{18} \times \frac{6}{18}} = 0.5773 \\ \text{第三週} \quad & \frac{4}{12} / \sqrt{\frac{12}{12} \times \frac{4}{12} \times \frac{2}{3}} + \frac{6}{12} / \sqrt{\frac{12}{12} \times \frac{6}{12} \times \frac{1}{3}} = 0.6206 \\ \text{第四週} \quad & \frac{4}{11} / \sqrt{\frac{6}{11} \times \frac{8}{11} \times \frac{2}{3}} + \frac{6}{18} / \sqrt{\frac{18}{18} \times \frac{6}{18} \times \frac{1}{3}} = 0.5773 \\ \text{第五週} \quad & \frac{4}{12} / \sqrt{\frac{12}{12} \times \frac{4}{12} \times \frac{2}{7}} + \frac{6}{12} / \sqrt{\frac{12}{12} \times \frac{6}{12} \times \frac{1}{7}} = 0.2660 \\ \text{第六週} \quad & \frac{4}{11} / \sqrt{\frac{6}{11} \times \frac{8}{11} \times \frac{2}{7}} + \frac{6}{18} / \sqrt{\frac{18}{18} \times \frac{6}{18} \times \frac{1}{7}} = 0.2474 \end{aligned}$$

以下我們說明 Digital Library AND Interface 在第六週結束時的權重計算方式。在第六週時，檢索歷程可分為三個週期，第五、六週為第一週期，第三、四週為第二週期，第一、二週為第三週期。在第一週期中，半衰期時間係數為 $4/7$ ，使用者共執行 17 次檢索，含有 Digital Library 的檢索策略之正規化頻率為 $2/17$ ，含有 Interface 的檢索策略之正規化頻率為 $2/17$ ，含有 Digital Library AND Interface 的檢索策略之正規化頻率為 $0/17$ ，因此 Digital Library AND Interface 在第一週期的權重為 $\frac{0/17}{\sqrt{(2/17 \times 2/17)}} = 0$ 。

在第二週期中，半衰期時間係數為 $2/7$ ，使用者共執行 11 次檢索，其中含有 Digital Library 的檢索策略之正規化頻率為 $6/11$ ，含有 Interface 的檢索策略之正規化頻率為 $8/11$ ，含有 Digital Library AND Interface 的檢索策略之正規化頻率為 $4/11$ ，因此 Digital Library AND Interface 在第二週期的權重為

$$\frac{4/11}{\sqrt{(6/11 \times 8/11)}}。$$

在第三週期中，半衰期時間係數為 $1/7$ ，使用者共執行 18 次檢索，其中含有 Digital Library 的檢索策略之正規化頻率為 $18/18$ ，含有 Interface 的檢索策略之正規化頻率為 $6/18$ ，含有 Digital Library AND Interface 的檢索策略之正規化頻率為 $6/18$ ，因此 Digital Library AND Interface 在第三週期的權重為

$$\frac{6/18}{\sqrt{(18/18 \times 6/18)}}。$$

結合三個週期內 Digital Library AND Interface 的權重以及時間對興趣的影響

係數，Digital Library AND Interface 在第六週結束時的權重為

$$0 + \frac{2}{7} \times \frac{4}{11} \frac{1}{\sqrt{(6/11 \times 8/11)}} + \frac{1}{7} \times \frac{6}{18} \frac{1}{\sqrt{(18/18 \times 6/18)}} = 0.2474。$$

表四：某一讀者在六週內曾經用過的檢索策略

週	關鍵詞A	運算符號	關鍵詞B	頻率
1	Digital Library			4
1	Digital Library	AND	IEEE	2
1	Digital Library	AND	Interface	6
2	Digital Library			2
2	Digital Library	AND	IEEE	4
3	Digital Library			2
3	Digital Library	AND	Interface	4
4	Literature			1
4	Interface			4
5	Digital Library			2
5	Literature	AND	Dumas	4
5	Literature	AND	Shakespeare	4
6	Interface			2
6	Shakespeare			5

表五：各關鍵字詞的權重

週	關鍵詞	權重	週	關鍵詞	權重
1	Digital Library	0.6	5	Digital Library	0.3068
1	Interface	0.3	5	Literature	0.2236
1	IEEE	0.1	5	Interface	0.1994
2	Digital Library	0.6	5	Dumas	0.0994
2	Interface	0.2	5	Shakespeare	0.0994
2	IEEE	0.2	5	IEEE	0.0714
3	Digital Library	0.6	6	Digital Library	0.2457
3	Interface	0.2333	6	Interface	0.2267
3	IEEE	0.1667	6	Shakespeare	0.2057
4	Digital Library	0.4667	6	Literature	0.2019
4	Interface	0.2333	6	Dumas	0.0914
4	Literature	0.1333	6	IEEE	0.0286
4	IEEE	0.0667			

表六：各檢索策略的權重

週	關鍵詞A	運算符號	關鍵詞B	權重
1	Digital Library	AND	Interface	0.7071
1	Digital Library	AND	IEEE	0.4083
2	Digital Library	AND	Interface	0.5773
2	Digital Library	AND	IEEE	0.5773
3	Digital Library	AND	Interface	0.6206
3	Digital Library	AND	IEEE	0.5209
4	Digital Library	AND	Interface	0.5773
4	Digital Library	AND	IEEE	0.1924
5	Digital Library	AND	Interface	0.2660
5	Digital Library	AND	IEEE	0.2233
5	Literature	AND	Dumas	0.3810
5	Literature	AND	Shakespeare	0.3809
6	Digital Library	AND	Interface	0.2474
6	Digital Library	AND	IEEE	0.0825
6	Literature	AND	Dumas	0.4041
6	Literature	AND	Shakespeare	0.2694

(二)PKRG 與資訊選粹服務、個人化檢索的關係

運用 PKRG 計算出每個關鍵字詞以及檢索策略的權重之後，可用以推演出使用者的興趣並應用在資訊選粹服務與個人化檢索。

運用 PKRG 進行資訊選粹服務的方法如下：

1. 透過 PKRG 計算出使用者用過的關鍵字詞的權重；
2. 將關鍵字詞依照權重排列，挑出權重最大的前幾個關鍵字詞，當作該使用者的興趣，並存入個人興趣紀錄中；
3. 比對新進資訊所含關鍵字詞是否和使用者的興趣紀錄檔中的興趣關鍵字詞相符。若相符，則視該資訊為使用者有興趣的新進資訊。

本文中所實現的個人化檢索，乃指系統能考慮各關鍵字詞間的相關性，將檢索結果依使用者興趣重新排列。在概念上，我們認為經常共同出現的關鍵字詞對

使用者而言代表這些關鍵字詞的關聯性很強，因此當使用者使用某一檢索策略搜尋資料時，若檢索結果中含有和該檢索策略經常共同出現的關鍵字詞，便會是使用者比較有興趣的。例如：若使用者常使用「古典AND 詩詞」這個檢索策略，代表使用者認為「古典」和「詩詞」間的關聯性很強，因此當使用者用「詩詞」來尋找資料時，我們便認為包含「古典」這個關鍵字詞的檢索結果會是使用者較有興趣的。運用PKRG 進行個人化檢索的方法如下：

1. 透過 PKRG 計算出使用者用過的檢索策略的權重，此權重亦代表關鍵字詞間的關聯性，並將檢索策略權重資訊存入個人興趣紀錄中；
2. 當使用者輸入一檢索策略時，根據個人興趣紀錄中的檢索策略權重資訊篩選出和當次檢索策略最有關的前幾個關鍵字詞；
3. 比對檢索結果和篩選出的關鍵字詞，含有愈多關鍵字詞的檢索結果代表使用者愈有興趣。

五、資料探勘與群體化服務

我們將資料探勘技術中的相關規則探勘與循序規則探勘應用在圖書館，期能發掘出讀者借閱館藏的社群性，以作為實施群體化服務的依據，我們欲探索的讀者社群關係包含：

1. 館藏借閱的共同性：興趣相同的讀者們往往會借閱類似的館藏，若我們能發掘出館藏借閱的共同性，當有某位讀者借閱某館藏時，我們便可推薦給他借過此館藏的讀者亦曾借閱的其他館藏。
2. 館藏借閱的順序性：對於某些館藏，讀者可能會依據一定的順序來閱讀（例如先借入門，再借進階），若我們發現許多讀者都按照一定的順序來閱讀某些館藏，那麼當有某位讀者借閱這些館藏中的某一本時，我們便可建議他按照順序來閱讀相關書籍。

(一)前置作業

在第二節中曾提及完整的資料探勘處理過程中，預備資料的步驟是最耗時的，也對資料探勘的正確性影響很大。我們在此步驟採取的動作如下：

1. 確定資料來源
資料來源為圖書館自動化系統中的書目檔、館藏檔、讀者檔及交易歷史檔。
2. 資料選取

- (1) 確定所要分析紀錄的時間範圍：交易歷史檔中通常包含許多年份的歷史紀錄，因此首先必須確定要分析的時間範圍。本研究的時間範圍是從1998年1月1日至2000年8月31日。
- (2) 確定所要分析紀錄的類別：交易歷史檔中通常包含許多類型的交易紀錄，例如：借閱、歸還、預約、聲明歸還等，由於本文所要探討的是館藏借閱的共同性和順序性，因此只需要用到交易歷史檔中的借閱和預約歷史紀錄。

經過本步驟的處理，我們在交易歷史檔中選取了1998年1月1日至2000年8月31日間有關借閱及預約的交易歷史資料，共有487,786筆。

3. 資料的前置處理及轉換

針對發掘館藏借閱共同性和順序性的資料處理需求，進行必要的資料前置處理及轉換。

- (1) 館藏借閱的共同性：採用相關規則探勘發掘館藏借閱的共同性。我們將每一筆書目資料視為一個項目，以書目號作為項目代碼，而每位讀者在一段時間內（如一學期或一學年）所借閱或預約^[27]的書目所成之集合即為一筆交易，資料庫 *D* 便是由所有交易組成的集合。舉例來說，若圖書館中有二位讀者 *A* 及 *B*，*A* 在一段時間內借了「1343」及「253」這二本書，*B* 在一段時間內借了「3423」、「34636」及「9689」三本書，資料庫 *D* 中就會有 {1343, 253} 和 {3423, 34636, 9689} 二筆交易。
- (2) 館藏借閱的順序性：採用循序規則探勘發掘館藏借閱的順序性。我們先將讀者借閱或預約的交易歷史紀錄依時間排序，並將每一筆書目資料視為一個項目，以書目號作為項目代碼，讀者同時間借閱或預約的書目所成之集合即為項目集，同一位讀者所借閱或預約的項目集依時間排序而成的序列即為讀者的交易序列。例如：圖書館中有一位讀者 *A* 在10月1日借了「3425」，在10月15日又借了「9823」及「4875」，則其交易序列即為 { (3425) (9823, 4875) }。

[27] 在探討相關規則探勘與循序規則探勘時，我們發現圖書館和商店最大的不同點為：在商店中同樣的物品會有許多個，但在圖書館中同樣的館藏往往只有一本或少數幾本複本，若有讀者將館藏借走，則其他的讀者便不能同時借閱，因此我們假設讀者預約館藏即表示有興趣借閱，並將預約書資料加入交易紀錄中。

(二)探勘成果及討論

在經過前置處理之後，我們進行相關規則以及循序規則的探勘。我們將1998年1月1日至2000年8月31日間的借閱及預約歷史資料納入探勘，每位讀者在這一段時間內借閱和預約過的館藏都視為同一筆交易；然後，我們將只含一本書的交易刪除（因為不可能探勘出共同性和順序性），以減少交易紀錄的筆數。經過這些處理後，探勘的資料包括11,398筆交易，且最長的交易有542個項目。由於讀者借閱館藏的期限大多為一個月，因此，若假設館藏只有一本，且讀者一借就是一個月，則二年半內最多只有30位讀者可借閱，因此我們將最小支持度設為0.21%（即23人，約是二年內可借閱的讀者人數），並針對書籍館藏運用相關規則探勘^[28]^[29]以及循序規則探勘^[30]的技術，分析出讀者借閱館藏的共同性，部分成果如表七。書籍借閱順序性的部分成果則如表八。

表七：館藏借閱的共同性（部分成果）

書名	借閱人數
· 精通Borland C++ Builder:視覺化C/C++程式設計.基礎篇 · Borland C++ Builder視窗程式設計經典	52
· MPEG video:compression standard · Digital video:an introduction to MPEG-2	51
· CMOS circuit design, layout, and simulation · Low-power cmos wireless communications:a wideband CDMA system design	42
07.MATLAB入門引導 · PC MATLAB入門與實例應用	38
· 親蜜心事 · 是誰拿走了那一雙雪靴	37
· CDMA systems engineering handbook · CDMA techniques for third generation mobile systems	36
· FreeBSD 抓得住 INTERNET:伺服器架設與管理 · FreeBSD 網路應用	34

[28] 楊雅雯，「個人化數位圖書資訊環境——以 PIE@NCTU 為例」(新竹：國立交通大學資訊科學系碩士論文，2001.6)

[29] 吳安琪，「利用資料探勘的技術及統計的方法增強圖書館的經營與服務」(新竹：國立交通大學資訊科學系碩士論文，2001.6)

[30] 戴玉旻，「圖書館借閱記錄探勘系統」(新竹：國立交通大學資訊科學系碩士論文，2001.6)

· 精通Borland C++ Builder:視覺化C/C++程式設計.基礎篇 · Borland C++ Builder完全征服手冊	33
· JPEG still image data compression standard · Win 32多緒程式設計:執行緒完全手冊=Multithreading Applications in Win 32	30
· 仙河飲馬 · 淨土之春	30
· RF power amplifiers for wireless communications · Microwave circuit design using linear and nonlinear techniques	28
· Win 32多緒程式設計:執行緒完全手冊=Multithreading Applications in Win 32 · PC影像處理技術. (一), 圖檔壓縮篇	28
· FreeBSD抓得住INTERNET:伺服器架設與管理 · 抓住你的PhotoImpact 4.2中文版	28
· 線性代數 · 通訊系統	27
· Numerical Recipes in C:The Art of Scientific Computing · An introduction to wavelets	24
· Visual C++ 6.0程式開發手冊 · Visual C++ 6.0程式設計指南, 應用程式架構篇	24
· 麵包樹上的女人 · 賣海豚的女孩	24
· Delphi 4.0 徹底研究 · 煞死你的網頁設計絕招	24

表八：館藏借閱的順序性（部分成果）

書名	借閱人數
· MATLAB入門引導 · PC MATLAB入門與實例應用	37
· Wideband CDMA for third generation mobile communications · CDMA systems engineering handbook	32
· CDMA systems engineering handbook · CDMA techniques for third generation mobile systems	29
· 仙河飲馬 · 淨土之春	29

在分析出館藏借閱的共同性和順序性之後，我們可以將分析的結果存入第三節提到的社群檔中，供個人化檢索與推薦時使用。舉例而言，當使用者經由個人化檢索找到《麵包樹上的女人》這本書時，在使用者檢視這本書的詳細資料時，系統可以將《賣海豚的女孩》推薦給這位使用者（在二年半的時間裡有 24 位讀者看過《麵包樹上的女人》與《賣海豚的女孩》這二本書）；類似的，若使用者檢索到《PC MATLAB 入門與實例應用》這本書時，我們可以建議他先閱讀《MATLAB 入門引導》（在二年半的時間裡有 37 位讀者是依照《MATLAB 入門引導》《PC MATLAB 入門與實例應用》的順序來閱讀這二本書）。

在此必須特別指出，並非資料探勘的所有結果都是有價值的；例如我們在探索館藏借閱的順序性時就發掘出以下的順序性：《名流劍客沒羽箭．第一部》《名流劍客沒羽箭．第二部》《名流劍客沒羽箭．第三部》《名流劍客沒羽箭．第四部》《名流劍客沒羽箭．第五部》。很明顯地，這是一個無用的循序規則。

六、PIE@NCTU 系統之實作

根據第三節提出的系統架構，並輔以第四節動態偵測讀者興趣的演算法，以及第五節以資料探勘實施社群化服務的技術，我們實作了一套適用於圖書館之個人化圖書資訊系統，稱之為交大個人化圖書資訊環境 (PIE@NCTU, Personalized Information Environment at National Chaio-Tung University)。^[31] PIE@NCTU 的發展目的為：

1. 提供一個 User-friendly、Web-based 的界面以便於讀者使用交大圖書館的圖書資源；
2. 能針對使用者需求給予個別的使用環境；
3. 具有社群化功能，促進讀者間的知識分享；
4. 動態偵測讀者興趣，協助讀者尋找交大圖書館館藏；
5. 提供資訊選粹服務，根據讀者興趣，定期通知讀者有興趣的新進資訊；
6. 提供跨平臺書籤功能，方便讀者利用館藏資源；
7. 提供讀者與圖書館互動的園地，促進讀者與圖書館的交流，並協助讀者解決利用館藏資源時遭遇到的困難。

[31] 同註 29。

從使用者的觀點來看，PIE@NCTU 是一個提供個人化和群體化服務的 WWW 網站。使用者可透過任何 WWW 瀏覽器進入。經由連接交大圖書館自動化系統的讀者資料檔，交大的讀者都能夠使用 PIE@NCTU 所提供的各項服務。讀者以其在交大圖書館自動化系統中既有的帳號、密碼成功登入系統後，便能享用 PIE@NCTU 的所有服務。圖三為 PIE@NCTU 首頁。

PIE@NCTU 所提供的服務有：(1) 個人化環境設定；(2) 個人化搜尋引擎；(3) 個人書籤；(4) 群體化服務；(5) 資訊選粹服務；(6) 我要問問題；(7) 個人通告。由於資訊選粹服務、我要問問題和個人通告與第三節系統架構中的資訊選粹服務、個人參考諮詢、個人資訊中心功能相近，不再贅述，以下僅針對前四項服務詳加說明。

(一) 個人化環境設定

PIE@NCTU 提供三項個人化環境設定的功能：個人化桌面、個人服務設定、個人興趣設定。

個人化桌面設定讓使用者根據其需要組織桌面環境。設計上是將整個桌面環境依不同功能劃分成數大類，各類別中含有子集合，然後提供選單讓使用者點選希望出現在自己桌面上的項目與功能。個人化桌面提供的服務類別包括：圖書館服務、圖書館導覽、館藏查詢、資料庫查詢、新書通告、借閱狀況、檢索界面、系統公告、圖書館連結等。圖三右方畫面為系統預設的桌面環境，但在某使用者根據其需求設定之後的個人化桌面如圖四。



圖三：PIE@NCTU 首頁 (<http://pie.e-lib.nctu.edu.tw/pie>)



圖四：個人化桌面

個人服務設定讓讀者選擇希望收到的資訊選粹服務，包含：個人新書目錄、藝文活動通告、圖書館公告。

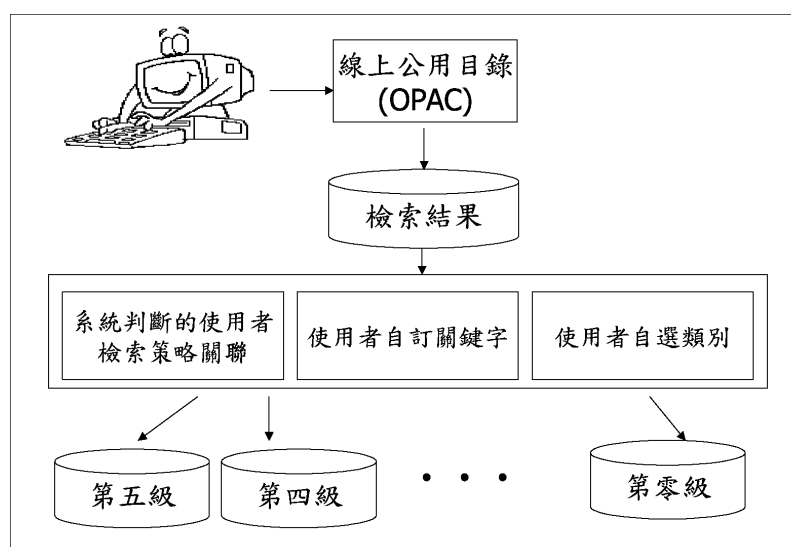
至於個人興趣設定，顧名思義就是讓使用者設定其個人的興趣。儘管在 PIE@NCTU 中我們已經應用了在第四節中提出的動態偵測興趣演算法來判斷讀者興趣，但是系統自動判斷出來的興趣可能沒有辦法完全符合使用者的興趣，為了輔助系統的不足，PIE@NCTU 提供讀者手動設定環境的功能：

1. 選擇系統關鍵字詞：由於系統判斷的關鍵字詞可能並非讀者真正的興趣，因此 PIE@NCTU 提供此項功能讓讀者能取消系統判斷的關鍵字詞。藉由此動作，我們也可知道對讀者而言，系統根據使用者檢索行為所判斷出來的個人興趣關鍵字詞是否符合使用者的需求。
2. 設定個人興趣關鍵字詞：為了彌補系統關鍵字詞的不足，PIE@NCTU 提供此功能讓讀者加入有興趣的關鍵字詞。藉由此動作，我們可知道有多少使用者有興趣的關鍵字詞是 PIE@NCTU 尚未找出來的。
3. 設定個人興趣類別：PIE@NCTU 除了讓讀者設定個人興趣關鍵字詞之外，也可讓讀者設定感興趣的書目類別。書目類別選項的依據是以「中國圖書分類法」為基礎的二階層「中國圖書分類法與美國國會圖書館分類法對照簡表」。^[32]

[32] 同註 28。

(二)個人化搜尋引擎

個人化搜尋引擎為第三節中個人化檢索的實現。PIE@NCTU 提供二種個人化搜尋引擎服務：個人館藏查詢、個人新進館藏查詢。圖五為PIE@NCTU 如何達成個人化搜尋引擎的示意圖。



圖五：個人化搜尋引擎示意圖

當使用者輸入一檢索策略時，系統首先會根據此檢索策略修正該使用者的 PKRG，然後將此檢索策略傳遞給線上公用目錄搜尋館藏，當線上公用目錄將檢索結果回傳給 PIE@NCTU 後，PIE@NCTU 根據以下三條件將檢索結果重新排列，以達到最符合讀者需求的呈現方式：

1. 系統判斷與當次檢索策略相關的關鍵字詞：當系統收到使用者的檢索策略後，PIE@NCTU 首先會根據 PKRG 的檢索策略權重（參見第四節）找出對使用者而言與當次檢索策略最有關聯的關鍵字詞。例如：當系統收到讀者甲要求尋找「貝多芬」時，PIE@NCTU 檢查讀者甲的 PKRG，發現甲的 PKRG 中存有二個包含「貝多芬」的檢索策略：「貝多芬 AND 小提琴」（權重 0.8）、「貝多芬 AND 交響曲」（權重 0.5），這表示對甲而言，「小提琴」和「交響曲」此二關鍵字詞與「貝多芬」具有關聯性，因此在當次檢索中，包含有「小提琴」或「交響曲」此二關鍵字詞的檢索結果對使用者愈相關。
2. 使用者自訂興趣關鍵字詞：PIE@NCTU 比對使用者自訂的興趣關鍵字詞是

否出現在檢索結果中。愈多自訂的關鍵字出現在某一筆檢索結果中，PIE@NCTU 即認為該資料對使用者而言越重要。

3. 使用者自選興趣類別：PIE@NCTU 比對檢索結果的書目類別是否和使用者自選興趣類別相同，若是，PIE@NCTU 則認為該資料對使用者而言是重要的。

比對完所有的檢索結果是否符合上述三個條件後，PIE@NCTU 會將檢索結果分成六個等級加以排序。愈符合上述三個條件的檢索結果等級越高，亦即系統認為該檢索結果對使用者較有用，在結果呈現時，會將其排列在較明顯的位置。系統將檢索結果分成零至五共六個等級，其中皆不符合為第零等級，剩餘五個等級依檢索結果符合上述三個條件的程度排列，每個等級差1分。其中，與當次檢索策略相關的關鍵字詞至多為3分（依檢索策略的權重累加），使用者自訂關鍵字至多為2分（每個符合的關鍵字為1分），使用者自選興趣類別為1分，當總分大於5分以上時皆歸在第五等級。圖六為某使用者檢索「事件」這個關鍵字詞的結果，列出的六筆資料為等級三的結果，其中的《77 班機事件幕後案：夏樹靜子推理系列之二》之所以被歸為等級三的原因如下：

1. 在 PKRG 中「推理 AND 事件」此檢索策略的權重為0.6；
2. 「推理」為使用者自訂的興趣關鍵字（1分）；
3. 分類號 861.37 為使用者自選興趣類別（1分）

加總後得分為2.6分，四捨五入後為3分，故歸類為等級三。



圖六：以「事件」為檢索策略進行個人化檢索的結果（等級三）

(三)個人書籤

PIE@NCTU 提供跨平臺的書籤功能：「個人書籤」，可讓使用者記錄有興趣書目的超連結及使用者的註解。圖七為個人書籤的範例。



圖七：個人書籤

(四)群體化服務

在資訊爆炸與網路科技的時代，圖書館在校園中所扮演的角色必須重新定位。圖書館除了提供使用者快速、精確、完整的查詢之外，還要能累積前人的智慧、綜合讀者的知識、提升校園閱讀文化。未來，圖書館除了是大量資料儲存的地方，也要是知識的所在之地。為了（1）增加圖書館館藏資源的附加價值；（2）增進讀者學習興趣，彼此激勵學習興趣；（3）提升圖書館與讀者互動讓讀者能充分利用圖書館的資源。在PIE@NCTU 中我們融入了群體化服務的功能。

1. 投票系統：藉由讀者投票讓圖書館更加了解讀者的意見。
2. 討論區：討論區讓讀者針對各主題發表意見與分享別人的看法；圖書館亦能經由討論區了解讀者的意見及看法，作為業務參考的依據。
3. 知識分享與網路讀書會：網路讀書會讓讀者針對特定書目／主題發表意見、獲得其他讀者的看法；也可以透過資料探勘的成果將讀者組成社群，讓彼此分享社群成員的學習心得、搜尋策略等，藉此促進知識分享，激發更多靈感。
4. 整合 PIE@NCTU 與第五節中資料探勘的成果，當使用者在查閱某一館藏的詳細資料時，PIE@NCTU 會顯示與該館藏經常被同一讀者借閱的其他館藏。範例如圖八。



圖八：整合資料探勘與PIE@NCTU的範例。當某位讀者在點選「溫馨廚房咖啡座」這本書的詳細資料時，系統會提供與這本書經常共同被借閱的其他三本書籍

七、結論與未來研究方向

本文主旨在探討電子圖書館環境下，圖書館如何運用客戶關係管理的理念，記錄、分析、並了解讀者的需求，從而建立個人化及群體化服務，以期增加讀者之滿意度與忠誠度。本文主要探討以下議題：

1. 以圖書館自動化系統為基礎，提出個人化與群體化服務的系統架構，分為系統服務、機制和資料庫三個層次。系統包含的服務有個人資訊中心、個人化檢索與推薦、資訊選粹服務、個人參考諮詢；系統機制有資訊過濾、智慧型檢索、知識分享、個人設定、我要問問題、興趣偵測和資料探勘；相關的資料庫包括：讀者檔、讀者興趣檔、讀者設定檔、參考問題檔、社群檔、圖書館公告檔、書目檔、館藏檔、加值書目檔、交易歷史檔。
2. 提出一個動態偵測使用者興趣的演算法。本演算法考慮個人使用檢索關鍵字詞的頻率、檢索策略的頻率以及時間對興趣的影響，並藉由個人興趣關聯圖清楚了解使用者的興趣分佈狀況，不僅可得知使用者感興趣的關鍵字詞，尚能了解使用者常用關鍵字詞的相互關係。我們並運用本演算法實作資訊選粹服務以及個人化檢索。
3. 運用資料探勘技術於探索圖書館館藏借閱的共同性和順序性，並將之應用於

群體化服務。

4. 實作一套具備個人化與群體化功能的交通大學個人化圖書資訊環境系統。

個人化與群體化的研究正方興未艾，即便我們已經實作出PIE@NCTU，但仍然有許多議題有待持續的探討。以下略述我們未來的研究方向：

1. 在 PIE@NCTU 中落實知識分享機制：實踐第三節中提及之知識分享概念，在 PIE@NCTU 中建構加值書目檔，設計讓讀者分享加值資訊的機制。
2. 進一步探討時間對使用者興趣的影響：在本文中我們提出一假設，認為時間對使用者興趣的影響是以半衰期遞減的。未來可進一步探討時間對使用者興趣可能展生的變化，以求得更好的結果。
3. 協力式資訊過濾：在本文中我們係以個人興趣為依據來實現內容式資訊過濾，未來可加入群組興趣實現協力式資訊過濾。
4. 在資料探勘中考量讀者分類：目前我們是將圖書館的所有讀者視為一個大社群，從中了解成員在館藏借閱的共同性與順序性。然而，讀者背景與學科領域可能會影響到其借閱行為，因此若能先將讀者分群（如根據系所、學院、性別、年級等分群），再針對每一群讀者探索其借閱的共同性和循序型，相信資料探勘的結果更能切合讀者的需求。
5. 進行評鑑與使用者調查：本文所提的技術皆已融入 PIE@NCTU，然而 PIE@NCTU 是否合乎使用者的需求、PKRG 是否能夠推導出使用者的興趣、資料探勘的技術是否足以應用於群體化服務，皆有待深入進行評鑑與使用者調查。

個人化與群體化的趨勢不只出現在電子圖書館，舉凡代理人或者全球資訊網的領域，皆可以看見越來越多的努力投入到個人化與群體化系統的研究。個人化與群體化的趨勢拉近服務以及使用者的距離，期待能有更多的研究投入這個領域，創造更理想的個人化與群體化資訊環境。

致 謝

感謝交通大學圖書館蔡淑琴小姐及交通大學資訊科學研究所陳莉君同學在建置 PIE@NCTU 的技術支援。另，本論文研究係國科會專題研究計畫之部分成果，計畫編號 NSC90-2213-E-009-083。

Preliminary Study of Personalized and Group Services in Electronic Libraries

Hao-Ren Ke

Associate Professor & Head of Digital Library Division

National Chiao-Tung University Library

Ya-Wen Yang An-Chi Wu Yu-Min Tai

Master of Institute of Computer and Information Science

National Chiao-Tung University

Wei-Pang Yang

Professor & Library Division, National Chiao-Tung University Library

【Abstract】

In recent years, libraries have decided to move toward electronic libraries. Electronic collection, operation, and service are the three aspects of an electronic library. The current trend in electronic service is to provide users with personalized services and group them into communities for sharing common interests and knowledge. In order to foster the development of a personalized and group information system in an electronic library format, this article studies how to establish such a system using library-automata-

tion systems. This article aims at the following:

1. Proposing a framework for a personalized and group information environment based on library-automation systems;
2. Studying how to dynamically track user interests and apply the research result to a personalized electronic service;
3. Studying the data mining technique and applying it to on electronic service that enables the formation of user communities; and
4. Implementing a prototype system, called Personalized Information Environment at NCTU (PIE@NCTU).

Keywords: Electronic Library, Personalized Information Environment, Community, Tracking User Interests, Personal Keyphrase-Relationship Graph, Data Mining, Association Rule Mining, Sequential Pattern Mining