

# 圖書館網站個人化推薦系統研究

韓瑛緋 高國峰 廖宜恩

國立中興大學資訊科學系

EMAIL: [harn002@yahoo.com.tw](mailto:harn002@yahoo.com.tw), [ieliao@nchu.edu.tw](mailto:ieliao@nchu.edu.tw)

## 摘要

圖書館的資料眾多且龐大。如何提供恰當的引導，讓讀者快速的找到需要的資料，是圖書館重要的課題。在以往，館藏查詢系統是最常被使用的工具。但傳統的系統只能提供被動的、制式的查詢服務，並不能滿足多樣化的現代讀者的需求。為了克服這些缺陷，本研究提出了一個個人化且具時間概念的主動式查詢系統。每一位讀者，一登入系統，系統將主動的提供針對其個人特性而製定的推薦書單，讓讀者能得到更好的服務。

本研究設計的個人化主題推薦系統，將分析讀者的借閱記錄，找出讀者真正有興趣的主題書目。並運用本體論(Ontology)與語意索引(Semantic Indexing)等技術，加上圖書館特有的分類編目知識，來提供使用者感興趣的主題館藏資訊。推薦主題清單將依讀者興趣值高低做個人化的排序，於圖書館網站上推薦給讀者。實驗結果顯示，這套系統確實獲得使用者的認同，將可有效的提升圖書館的服務品質。

**關鍵詞：**圖書館、個人化、推薦系統、本體論(Ontology)、語意索引(Semantic indexing)。

## 1. 前言

何光國[1]提到，圖書館不是書店，也不是出版商店。它的重心不在提供讀者片斷的資訊，而是根據知識分類的方法，將同一主題(Subject)的資訊片斷，作系統性和計劃性的徵購、組織和整理，然後再分類地、有序地排列在圖書館裡，儘量提供給讀者「主題知識」(Subject Knowledge)的全貌。其中，「主題知識」，是指根據統一的編目分類法則，將同屬性的圖書編組在一起，而成的知識。

而如何提供讀者良好的服務，協助讀者獲取資訊、運用資訊，從而產生知識，是圖書館最重要的使命。因此為了讓讀者快速、有效、完整地滿足其資訊需求，圖書館應善加利用資訊技術以有效幫助讀者使用館藏資源。現今在資訊科技發達的環境下，圖書館之發展重心也應逐漸由圖書館本位轉移至個人本位。

隨著網路及全球資訊網的發展，人們已漸漸習慣於網路上尋找所需的資訊，也促成圖書館開始思索如何運用網路及資訊科技來往「電子化服務」的

目標邁進。而雖然現今大部份的圖書館皆有線上資訊檢索系統，但並非每一位讀者皆能妥善運用。讀者做線上檢索時常遇到以下的問題：

- (1)讀者在館藏查詢系統上檢索所得之館藏數量往往非常龐大且雜亂，加上讀者若不熟悉檢索系統的功能，常會找不到需要的資源。國內圖書資訊研究學者吳美美[2]提到，臨床觀察檢索系統使用情況，發現使用者經常無法順利查檢到所需要的資訊，不是高比率的零筆檢索結果，就是因為檢索筆數太多放棄相關判斷。
- (2)傳統資訊檢索技術所提供的是被動式(Passive)、短期(Short-Term)、一般化的、零碎的、片斷的資訊服務，而非有組織、有主題的、擁有領域知識的資訊服務。
- (3)讀者不知道有哪些相關主題館藏，讀者或許一開始有設定要借哪些主題的書，但他可能不知道此主題可牽涉到很多其它相關主題，而且這些主題是對研究很有幫助的。
- (4)館藏量日漸龐大，而讀者不知道有哪些相關館藏，或是館藏中有哪些重要的主題？圖書館長期經營後必然會有館藏愈來愈多的傾向，如果不能給予讀者適當的建議，那麼可能有許多值得一讀的館藏不會被讀者借閱。

另外，根據施毓琦[3]對現今國內外幾所大學具有個人化功能的圖書館網站所進行研究發現，在個人化、資訊過濾方面，多是讓使用者自行輸入某些限制條件或是偵測使用者瀏覽路徑來建立使用者興趣檔，系統則以這些條件來作為推薦資訊的依據。而本研究者根據其論文中之論述，發現其所共同出現的問題如下：

- (1)個人興趣檔只由使用者自行建立的方式，當使用者所挑選的詞彙或興趣主題太過專指或太過廣泛時，會造成資訊選粹服務所提供的資訊過於狹隘或資訊太多不符合自己所需的情況。
- (2)使用者的興趣會隨著時間而有所不同，使用者必須時常上網更改設定才能得到自己想要的資訊。
- (3)若是系統自行偵測瀏覽路徑來推導，則可能使用者非正式或無目的瀏覽及檢索網站的各項行為，亦會被歸納為使用者興趣或需求，而造成系統所提供的資訊不符合自己所需的情況。

針對上述問題本研究設計一主動式的推薦系統，以站在讀者個人本位及提供讀者主題知識的角

度，結合圖書館特有的館藏資源分類架構來設計。再者，根據讀者的實際借閱行為來動態追蹤其興趣，將讀者想要的圖書資訊自動送上門。

## 2. 相關研究

本系統係以本體論及語意索引的技術應用於圖書館網站個人化推薦系統，因此，在相關文獻方面，將以下列幾個方向為探究目標。

### 2.1 領域本體論(Domain ontology)

本體論是對真實世界中存在的實體進行有系統的說明。而領域本體論是將特定知識領域作概念化，並以此特定領域(如醫學或電腦)來描述其相關詞彙及關係。具體而言，即是許多知識術語(Knowledge Term)所組成的集合，包含詞彙(Vocabulary)，語意上的相互連結(Semantic Interconnection)，以及推論及邏輯上的簡單規則。領域本體論如同專家的專門知識，每一份專業知識都運用於記載該領域中的事物。

而一個特定領域的本體論是如何建構出來的？對於定義該領域的專家來說，必須作以下三個的分析[8]：

- (1) 檢查作為描述節點內實體的詞彙。
- (2) 對於詞彙裡面的名詞，建立起正規的敘述，亦即將他們轉換為概念及關係。
- (3) 描繪出名詞間在概念上的關係為何？

依上述定義，本研究所使用的圖書分類架構-中國圖書分類法即為一領域本體論，以分類號的方式將書籍做概念化的分類，且描述了概念間的遠近及階層關係。

### 2.2 參考本體論與個人本體論

參考本體論(reference ontology)指的是現存的一種分類方式，而個人本體論(personal ontology)則是使用者個人偏好所產生的一種分類方式。Chaffee[5]於其做個人化的網頁瀏覽與搜尋的研究中提出一種將參考本體論與個人本體論加以對應的方式，以解決統一的本體論與個人偏好的衝突。利用個人本體論主要著重於個人良好的瀏覽經驗，既然系統已經知道使用者個人的偏好，就不需要給所有的使用者同樣的分類方式。

中國圖書分類法為絕大多數人所認同的一種中文書籍之分類方式，可視為參考本體論，且可以樹狀結構來呈現。本研究透過分析使用者的借閱各類別的行為及借閱時間等因素，自動產生使用者的個人偏好分類(personal preference categorization)。這樣的偏好分類以個人本體論的方式呈現。並且可以與由圖書分類結構所產生的參考本體論加以對應，做較深入的或廣泛的知識探索。

### 2.3 語意索引(Semantic Indexing)

由於傳統的技術只提供文件內的單一關鍵詞供檢索，無法讓使用者找到與關鍵字詞語意上相類似的文件，因此有了語意索引的方法產生。語意字詞除了能找出文件內的關鍵詞外，並能找出所有語

意概念上相近的關鍵詞，使索引技術更為強大[10]。

建立語意索引的方法就是將兩兩關鍵詞用語意相似性串連起來，進而建構成一兩兩相連的網路脈絡，實際的呈現方法稱之為「語意矩陣(Semantic Matrix)或語意網路(Semantic Network)」。

透過語意索引的檢索，使用者所下達的索引關鍵詞不僅僅針對文字上的相似性做檢索，而是針對檢索關鍵詞整個語意上的概念來作檢索。

## 3. 系統架構與設計

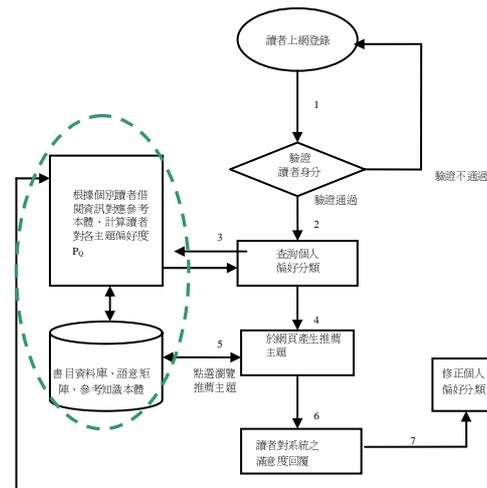


圖 1 推薦系統基本架構及流程圖

圖 1 為本研究設計的圖書館網站個人化推薦系統之基本架構及推薦流程圖，圖中箭號旁的數字為讀者造訪網站與推薦系統互動的流程。使用者透過圖書館網站的登錄來進入「個人化圖書館」網頁，登錄後系統將個人借閱記錄取出，做為推薦的基礎，其中以橢圓形虛線圍繞起來的部分為推薦系統需事先做好的工作。之後，再針對讀者的反應，以進行接下來的流程。以下章節將會針對流程中的主要部分加以說明。

### 3.1 使用者借閱記錄處理

由於本研究是以讀者借閱各類別(分類號)的情形來推導讀者興趣，因而讀者的借閱記錄需轉換為以統計借閱各類別的次數的形式存在，以作系統後續處理。

由於大多數人的興趣會隨時間的流逝而產生改變，因而推導讀者興趣，不止考量借閱各類別次數，亦需考量時間上距今的遠近，如<方程式 1>。

$$W_Q = \sum_{x=1}^n (f_{Qx} * T_x) \quad \dots \langle \text{方程式 1} \rangle$$

$W_Q$  : 對個別使用者而言，標題 Q 的借閱權重。  
 $f_{Qx}$  : 第 x 個時間區段中，標題 Q 出現的次數。

$T_x$  :  $1/2^{(x-1)}$  where  $1 \leq x \leq n$ , 第 x 時區之時間係數。採半衰期方式遞減權

重。其中 n 為時區數目。

### 3.2 參考本體論及個人本體論對應

中國圖書分類法可視為一領域本體論，用此本體論當作參考本體由所有讀者所共用。藉由分析讀者借閱各類別的多寡及分佈上的聚散情形，作為讀者興趣之推論。

#### 3.2.1 參考本體論及個人本體論對應之構想

圖 2 為圖書分類參考本體架構圖之某一部分可從中觀察到其分類號的左右排列及上下階層有一定的規則，階層愈往下分類號數字愈增加。

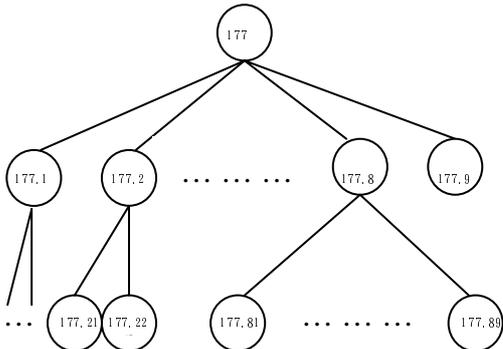


圖 2 圖書分類參考本體架構圖(部份)

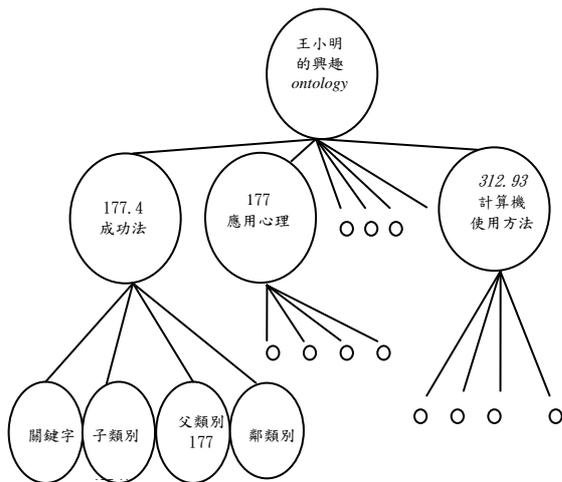


圖 3 圖書分類個人本體架構概念圖

圖 3 為本研究所提出之圖書分類個人本體架構概念圖，此種個人本體論結構是以讀者的興趣為主要考量，中國圖書分類法為龐大的分類架構，但大多數的人皆不了解其階層關係，亦不清楚各分類號所代表的涵意。因而本研究以圖 3 結構，讓讀者以另一種方式來針對感興趣的主題，做相關書籍的搜尋。

在圖 3 根節點下的第二層節點，為系統所推薦給讀者的主題書目，皆為系統依讀者的借閱各分類主題的情形所推算出來的  $P_Q$  值高低，而左到右依序列出。節點內含有所屬所有書目的資訊，由於考慮到在中國圖書分類下，部份分類節點(分類號)有數百本的書籍，為減少讀者搜尋特定主題書目的時間，採用語意索引的方式來協助讀者(詳細方法

見 3.3 節)。

若要針對系統所推薦的特定主題來做較深入的研究，可以就名稱為「子類別」的子節點中裡，作相關書籍的搜尋；若要針對系統所推薦的特定主題來做較廣泛的研究，可以就名稱為「父類別」或「鄰類別」的子節點裡，作相關書籍的搜尋。而這些特定主題的子類別、父類別及鄰類別等須與參考本體—中國圖書分類法對應。

#### 3.2.2 個人本體論之建構

而在本體論推論的設計上，特定 Q 類別週遭相關類別(父類別、子類別及鄰類別)的興趣值  $W_Q$ (由<方程式 1>算出)和個數，對於 Q 都有影響。 $W_Q$  因為代表使用者對某個類別感興趣的程度，所以很直觀的會有影響。而個數的影響在於，如果某個類別(例：Q)的子類別(例： $C_1 \sim C_9$ )中，9 個裡有 6 個感興趣，那麼可以斷定你對於 Q 類別的興趣也會不小。另外，由於  $W_Q$  是相對的，所以我們再計算本體論的影響程度時會先找到一個  $W_Q$  最大的類別  $W_{QMAX}$ ，把這個類別當作基準，做一個正規化的動作。

基於以上的設想，我們參考曾信誠[4]以本體論推論造訪網站之使用者興趣的方法，做部分修改以推論讀者對圖書館中某個特定的圖書類別 Q 的支持度 (Support) 為  $S_Q$ 。如<方程式 2>：

$$S_Q = \beta_1 \left( \frac{\sum_{k=1}^{C_{total}} W_{Qck}}{W_{QMAX}} \right) \frac{1}{C_{total}} + \beta_2 \frac{W_{QP}}{W_{QMAX}} + \beta_3 \left( \frac{\sum_{k=1}^{S_{total}} W_{Qsk}}{W_{QMAX}} \right) \frac{1}{S_{total}} \quad \langle \text{方程式 2} \rangle$$

$S_Q$ ：讀者對 Q 類別的本體論推論支持度。  
 $\beta_1, \beta_2, \beta_3$ ：分別表示子、父及鄰類別之偏好的加權係數，其中  $\beta_1 > \beta_2 > \beta_3$ 。代表我們認為子類別的影響程度大於父類別，而父類別又大於鄰類別。

$W_{QMAX}$ ：Maximum Subject Weight, 在圖書分類法的所有類別中，讀者最有興趣的類別(主題)。亦即由<方程式 1>所算出其值最大者。

$W_{Qck}$ ：對 Q 類別的第 k 個子類別的興趣值，根據<方程式 1>算出。

$W_{QP}$ ：對 Q 類別之父類別的興趣值，根據<方程式 1>算出。

$W_{Qsk}$ ：對 Q 類別的第 k 個鄰類別的興趣值，根據<方程式 1>算出。

$C_{total}$ ：Q 類別之子類別(Child)數目。

$S_{total}$ ：Q 類別之鄰類別(Sibling)數目。

讀者借閱各類別的  $W_Q$  值與個人本體推論之  $S_Q$  相乘，即為個人對各類別之偏好度，如<方程式 3>。

$$P_Q = W_Q \times S_Q \quad \langle \text{方程式 3} \rangle$$

### 3.3 建置各類別語意索引

本節所用的<方程式 4>~<方程式 8>乃參考建置語意索引(semantic indexing)[6][10]及資訊擷取(information retrieval)[7][9]技術，這些技術

原用於文件上做重要訊息的擷取，現在針對本研究的需要予以修改，以適應研究擷取圖書館書目關鍵詞及關鍵詞間之關聯性的需要。分為下列四步驟：

<方程式 6>

(1) 書名斷詞：用中研院所開發出之中文自動斷詞系統(CKIP)做書名斷詞。

(2) 斷詞合併：由於斷詞系統所斷出的詞多為二字詞，語意較不完整，因此，以用來計算兩個事件連結強度的方法「Mutual Information」，進行前後相鄰字詞合併，合併成較長且有代表性的詞，如：計算機組織、數據結構、數值分析…等，以產生較語意較完整的詞。

$$MI(t_i, t_j) = \log \left( \frac{\frac{tf_{ij}}{sf_{ij}}}{\frac{tf_i}{sf_i} \times \frac{tf_j}{sf_j}} \right) \quad \text{<方程式 4>}$$

其中， $t_i$  及  $t_j$  代表兩個單一詞，考量其前後相鄰的情形來作合併，合併成有較完整語意的長詞， $t_i$  是第一個詞， $t_j$  是第 2 個詞。 $tf_i$  是指  $t_i$  這個中文詞在同一分類號下的所有書名裡出現的頻率， $sf_i$  代表的就是「 $t_i$ 」這個中文詞出現的在相鄰類別之類別數目。同理可得  $tf_{ij}$  就是代表著此合併長詞在小類中所有書名中出現的頻率，而  $sf_{ij}$  就是該合併長詞出現的在相鄰類別之類別數目。

(3) 計算關鍵字重要性

計算詞彙  $t_i$  在類別  $k$  中的權重，即在特定類別中，關鍵字的重要性排序，其精神是依據其獨特性程度、館藏數量的多寡及關鍵字字數的多寡決定其重要性。

$$W_{ik} = tf_{ik} \times \log \left( \frac{N}{sf_i} \times (2^2)^{w_i-1} \right) \quad \text{<方程式 5>}$$

$W_{ik}$ ：詞彙  $t_i$  在類別  $k$  中的重要性。

$tf_{ik}$ ：詞彙  $t_i$  在類別  $k$  中的頻率。

$sf_i$ ：代表的就是「 $t_i$ 」這個中文詞出現的在同一父類別下之相鄰類別的類別數目。值愈小，表示此詞彙具有獨特性，且在此類別愈有代表性。

$N$ ：表示同一父類別下之相鄰類別總數。

$W_i$ ：詞彙  $t_i$  若由一個斷詞所組成其值為 1，若由二個斷詞合併而成其值為 2，以此類推。

此公式中，詞彙所含的名詞個數愈多，所計算出來的權重也愈高，這是因為名詞片語由愈多的文字組成的會有愈強的語意。

(4) 語意關連性計算

語意索引的方法就是將兩兩關鍵詞用語意相似性串連起來，進而構成一兩兩相連的網路脈絡，以<方程式 6>~<方程式 8>計算兩兩關鍵詞之間的關聯性。

$$Weight(t_i, t_j) = \frac{\sum_{k=1}^n W_{ijk}}{\sum_{k=1}^n W_{ik}} \times WeightingFactor(t_j)$$

$$W_{ijk} = tf_{ijk} \times \log \left( \frac{N}{sf_{ij}} \times (2^2)^{w_i-1} \right) \quad \text{<方程式 7>}$$

$$WeightingFactor(t_j) = \left( \frac{\log \frac{N}{sf_j}}{\log N} \right) \times (2^2)^{w_j-1} \quad \text{<方程式 8>}$$

$Weight(t_i, t_j)$ ：值愈大，則表示  $t_i$  與  $t_j$  有愈重要的關聯。

$W_{ijk}$ ：表示詞彙  $t_i$  及詞彙  $t_j$  同時出現在類別  $k$  中的權重。

$W_{ik}$ ：表示詞彙  $t_i$  出現在類別  $k$  中的權重，由<方程式 5>而來。

$tf_{ijk}$ ：表示取詞彙  $t_i$  及詞彙  $t_j$  在類別  $k$  中，各別的出現頻率值，較少的那一個。

$sf_{ij}$ ：表示詞彙  $t_i$  及詞彙  $t_j$  同時出現在同一父類別下之相鄰類別之類別數量。

$N$ ：表示的是同一父類別下之相鄰類別總數。

$w_i$ ：表示詞彙  $t_i$  所包含的單一詞個數，例如，程式語言，是由程式及語言這兩個單一斷詞合併起來的，所以其值為 2。

### 3.4 收集讀者對推薦類別的回應及個人偏好分類修正

#### 3.4.1 收集使用者對推薦系統之回應

讀者在網頁上依類別是否符合其興趣來回覆滿意度(如圖 4 右方)。這裡的滿意度指的是讀者對此主題下所屬書目的興趣程度，分成五個等級讓讀者選擇，分別是“非常滿意”、“滿意”、“普通”、“不滿意”及“非常不滿意”五種，實際上系統內部代表的分數為 0.9, 0.7, 0.5, 0.3, 0.1。

#### 3.4.2 修正個人偏好分類

將讀者對各類別主題所回覆的滿意度與系統針對個別讀者所推論出的各類別主題的  $P_Q$  值「相乘」。所得的結果即為讀者對各類別最新的  $P_Q$  值，當讀者再次上網時，系統的推薦排序將以此值為準。

## 4. 系統實作及實驗結果

本論文在實作及實驗過程中所使用的書目資料及讀者借閱資料係來自中興大學圖書館，並徵求其中 37 位讀者配合進行使用滿意度實驗。

### 4.1 網站建置

本研究以 MS SQL-2000 為資料庫，採用 ASP 為讀者及系統之間溝通的網頁程式，取得讀者個人的借閱資料後，以<方程式 1>~<方程式 3>推論出讀者對各類別的興趣值之後，依興趣值高低排列，推薦給讀者點選其相關圖書資訊，畫面呈現如圖 4。



圖 4 個人化推薦主題呈現畫面

當讀者點選其中某一類別時，則系統將呈現此類別下所屬全部書目如圖 5，畫面左方為此類別的語意索引，當點選其中某一關鍵詞(如：IBM PC)時，系統則呈現書名中有此關鍵詞的所有書目，如圖 6。另外，圖 6 左方為系統提供與此關鍵詞有語意相關之語意索引。

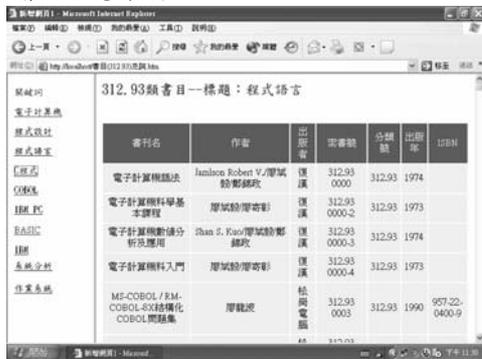


圖 5 特定「類別」之書目及其「相關詞」呈現畫面



圖 6 「關鍵詞」書目及其語意相關之「相關詞」呈現畫面

若讀者對系統所推薦的特定主題下所有的書目瀏覽之後，想搜尋更較深入、更細分的知識，可以點選系統提供「子類別」超連結裡，作相關書籍的搜尋，如圖 7 為 312.93 之子類別，共有四個讓讀者選擇、點選；若要搜尋較廣泛的、上層的知識，可以點選名稱為「父類別」或「鄰類別」的超連結裡，作相關書籍的搜尋，其中「鄰類別」顯示的方式與「子類別」類似，例如讀者點選 312.931 的鄰類別時，系統會顯示圖 7 的畫面讓讀者選擇、點選；而若讀者點選 312.931 的「父類別」時，則系

統提供其父類別所有書目及關鍵字提示，如圖 5。而這些特定主題的子類別、父類別及鄰類別等須與統一的參考本體論—中國圖書分類法對應。



圖 7 特定類別之「子類別」、「鄰類別」呈現畫面

## 4.2 實驗結果

在徵求讀者配合進行使用推薦系統後，請其依各類別做滿意度回覆，收集所有讀者所回覆的滿意度之後則進行分析。

### 4.2.1 實驗結果分析方法

以三種方式來計算讀者滿意度，以評估推薦系統是否可行：

- (1) 平均滿意度：計算所有讀者個人對各類別主題滿意度之平均值的平均。
- (2) 平均誤差值：計算讀者實際滿意度及系統預測之間的差距。

$$S_x = \frac{1}{m_x} \sum_{j \in p_x} |p_{x,j} - v_{x,j}| \quad \langle \text{方程式 9} \rangle$$

$S_x$ ：讀者  $x$  對推薦主題的整體滿意度

$m_x$ ：讀者  $x$  得到的推薦數量

$p_x$ ：為系統給讀者  $x$  的所有主題推薦

$p_{x,j}$ ：為系統預測讀者  $x$  對主題  $j$  的滿意度

$v_{x,j}$ ：讀者  $x$  對主題  $j$  的實際滿意度

- (3) 排序分數度量值：計算使用者對系統推薦的排序是否滿意。

$$R_x = \sum_j \frac{\max(v_{x,j} - d, 0)}{2^{(j-1)/(\alpha-1)}} \quad \langle \text{方程式 10} \rangle$$

$R_x$ ：讀者  $x$  對主題排序的滿意度

$j$ ：為在推薦清單中排序的位置， $j$  愈小表示排的愈前面，推薦程度愈高。反之， $j$  愈大表示排的愈後面，推薦程度愈低。

$v_{x,j}$ ：是讀者  $x$  對推薦主題  $j$  的實際滿意度

$d$ ：為讀者滿意度的平均值。

$\alpha$ ：推薦數目，其作用為此公式中的半衰期。

$R_x$  的意義在於如果讀者比較喜歡的館藏都在推薦清單比較前面的位置，而比較不喜歡的都在排序比較後面的位置，則  $R_x$  可以得到最大的值。反之，那  $R_x$  的值就會很低。所以  $R_x$  可以看出讀者對推薦清單排序的滿意度。通常排序分數度量值還需要正規化，即將  $R_x$  除  $R_x^{max}$ ，使得到的值能介於某個範圍內。 $R_x^{max}$  代表當推薦排序完全依照讀者喜

好排序時，亦即讀者喜歡的都排在前面，較不喜歡的都排在後面，所能得到的最大滿意值。這樣得到的分數可以比較明確地看出系統排序的結果到底有多接近讀者理想的排序。正規化的排序分數度量值計算方式如〈方程式 11〉。

$$R = 100 \times \frac{\sum_x R_x}{\sum_x R_x^{\max}} \quad \langle \text{方程式 11} \rangle$$

#### 4.2.2 實驗結果分析

表 1 讀者滿意度回覆統計分析結果

評估方式 項目	平均滿意度	平均誤差值	排序分數 度量值
第一項	0.736364	0.163636	49.71154
前三項	0.725758	0.174243	52.74962
前六項	0.711053	0.147718	67.86519
前十項	0.717165	0.157273	81.88048

平均滿意度，由表 1 可觀察出，讀者平均滿意度皆在 0.7 以上。也顯示讀者對於此推薦系統的認同度頗高，本研究所設計的圖書館網站推薦系統有一定的可行性。

平均誤差值，大約落在 0.14-0.18 之間。因為系統中給讀者選擇的滿意度間隔區間為 0.2，顯示系統預設的推薦分數與讀者選擇的滿意度差距在一個誤差內，代表誤差在可接受的範圍內

排序分數度量值，由第一項的排序分數度量值觀察，顯示對所有的讀者大約只有一半會對第一項推薦主題打上最高的分數。而前三項的排序分數度量值亦不是很理想，但之後的前六項及前十項的排序分數度量值則快速提升，並於在前十項得到最佳值 81.88048，顯示系統前十項推薦主題的排序非常接近讀者理想的排序

#### 5. 結論與未來展望

本研究採用資訊過濾的方法來為讀者做到長期的、主動式與個人化的資訊服務。將高達萬類以上的主題書目過濾，只呈現給讀者最感興趣的幾個主題，供讀者瀏覽搜尋。真正達到為個別讀者做資訊過濾的功能。並利用讀者個人的借閱記錄及結合圖書館既有的、大家普遍接受的知識分類架構，來推論出讀者真正有興趣的主題。是一結合圖書館學理論、現行圖書分類(知識分類)架構及個人興趣提出一個適合圖書館環境的個人化推薦系統。

在推薦方法的應用上採內容導向過濾的方法，將個人的興趣分類(個人本體論)和統一的分類架構(參考本體論)作對應。使系統既能有效的依照讀者興趣來推薦書籍，又可依既有的、統一的分類架構作相關主題書籍搜尋。由於許多分類標題下有數百本的書籍，因而在各類別下，以建立語意索引的方式，引導讀者搜尋相關主題。

在系統評估的結果上，所得到的平均滿意度、平均誤差值及排序分類度量值都達到不錯的結果，證明這個圖書館個人化系統確實可行，能帶給讀者一個新的管道來接觸圖書館的館藏。

因本論文重點在於推薦策略的運用，而在實作時，我們發覺未來本系統有以下幾點可再加強：

- (1) cold-start 的問題。未來可再考慮採用一適當的協力式資訊過濾方法以解決此項問題。
- (2) 而讀者借閱一本書後二天就還與借一個月才還的意義有很大的不同，未來可考慮以借閱期長短來當成推論讀者興趣的依據。
- (3) 西文圖書推薦，本研究只針對中文書籍來做推薦，以後可增加以杜威十進位分類法為參考本體及西文語意索引的方式，則系統就可以增加西文圖書的推薦服務。

#### 六、參考文獻

- [1]何光國(1999)。《圖書館學理論基礎》(頁 349-364)。台北市：三民書局。
- [2]吳美美(2001)。《中文資訊系統使用研究》(頁 119~142)。台北市：台灣學生。
- [3]施毓琦(2003)。《大學圖書館網站個人化服務之使用者需求研究》。(碩士論文，臺灣大學，2003)。全國博碩士論文資訊網，91NTU00448007。
- [4]曾信誠(2004)。《以本體論為基礎之使用者喜好萃取、隱私權控管與側寫建構》。(碩士論文，東華大學，2004)。全國博碩士論文資訊網，92NDHU5392012。
- [5]Chaffee, J., (2000). Personal Ontology for Web Navigation. *Proc. 9th Intl. Conf. on Information and Knowledge Management*, McLean VA, Nov 2000. pp.227-234
- [6]Chung, Y.M., He, Q., Powell, K. and Schatz, B. (1999). Semantic Indexing for a Complete Subject Discipline. *Fourth ACM Conference on Digital Libraries*, pp.39-48.
- [7]Guarino, N.(1997). Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction and Integration. In M. T. Pazienza(ed.) *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*. Springer Verlag: pp.139-170.
- [8]Navigli, R. & Velardi, P. (2003). Ontology Learning and its Application to Automated Terminology Translation. *IEEE Intelligent Systems*, Vol. 18, Jan/Feb 2003, pp.23-31
- [9]Popov, B., Kiryakov, A., Ognyanoff, D., Manov, D. and Kirilov.(2004). A. KIM – a semantic annotation platform for information extraction and retrieval. *Natural Language Engineering*, Vol.10, Issues 3-4, September 2004, pp.375-392.
- [10]Saggion, H., Cunningham, H., Bontcheva, K. and Wilks, Y. (2003) Multimedia indexing through multisource and multilingual information extraction; the MUMIS project. *Data and Knowledge Engineering*, Vol.48, 2003, 247-264.