

~ 獻給我的父母 ~



國立交通大學

資訊管理研究所

碩士論文

Web 2.0 概念的圖書館個人化推薦系統

A Web 2.0-based

Personalized Recommendation System for Library

研究生：羅子文

指導教授：柯皓仁 教授

中華民國九十六年七月

# Web 2.0 概念的圖書館個人化推薦系統

A Web 2.0-based Personalized Recommendation System for Library

研究生：羅子文

Student：Tzu-Wen Lo

指導教授：柯皓仁

Advisor：Hao-Ren Ke



Information Management

July 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年七月

# 國立交通大學

## 博碩士論文全文電子檔著作權授權書

(提供授權人裝訂於紙本論文書名頁之次頁用)

本授權書所授權之學位論文，為本人於國立交通大學 資訊管理研究所，  
九十五 學年度第二學期取得碩士學位之論文。

論文題目：Web 2.0 概念的圖書館個人化推薦系統

指導教授：柯皓仁

同意  不同意

本人茲將本著作，以非專屬、無償授權國立交通大學與台灣聯合大學系統圖書館：基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學及台灣聯合大學系統圖書館得不限地域、時間與次數，以紙本、光碟或數位化等各種方法收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行線上檢索、閱覽、下載或列印。

論文全文上載網路公開之範圍及時間：

本校及台灣聯合大學系統區域網路	<input checked="" type="checkbox"/> 中華民國 98 年 7 月 1 日公開
校外網際網路	<input checked="" type="checkbox"/> 中華民國 100 年 7 月 1 日公開

授權人：羅子文

親筆簽名：羅子文

中華民國 96 年 7 月 1 日

# 國立交通大學

## 博碩士紙本論文著作權授權書

(提供授權人裝訂於全文電子檔授權書之次頁用)

本授權書所授權之學位論文，為本人於國立交通大學 資訊管理研究所，九十五 學年度第二學期取得碩士學位之論文。

論文題目：Web 2.0 概念的圖書館個人化推薦系統

指導教授：柯皓仁

### ■ 同意

本人茲將本著作，以非專屬、無償授權國立交通大學，基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學圖書館得以紙本收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行閱覽或列印。

本論文為本人向經濟部智慧局申請專利(未申請者本條款請不予理會)的附件之一，申請文號為：\_\_\_\_\_，請將論文延至\_\_\_\_年\_\_\_\_月\_\_\_\_日再公開。

授權人：羅子文

親筆簽名： 羅子文

中華民國 96 年 7 月 1 日

# 國家圖書館博碩士論文電子檔案上網授權書

ID:GT009434523

本授權書所授權之學位論文，為本人於國立交通大學 資訊管理研究所，  
九十五 學年度第二學期取得碩士學位之論文。

論文題目：Web 2.0 概念的圖書館個人化推薦系統

指導教授：柯皓仁

茲同意將授權人擁有著作權之上列論文全文（含摘要），非專屬、無償授權國家圖書館，不限地域、時間與次數，以微縮、光碟或其他各種數位化方式將上列論文重製，並得將數位化之上列論文及論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

※ 讀者基於非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

授權人：羅子文

親筆簽名：羅子文

民國 96 年 7 月 1 日

1. 本授權書請以黑筆撰寫，並列印二份，其中一份影印裝訂於附錄三之二(博碩士紙本論文著作權授權書)之次頁；另一份於辦理離校時繳交給系所助理，由圖書館彙總寄交國家圖書館。

# Web 2.0 概念的圖書館個人化推薦系統

A Web 2.0-based Personalized Recommendation System for Library

研究生：羅子文      指導教授：柯皓仁博士

國立交通大學資訊管理研究所

## 摘要

Web 2.0 的精神是想透過社群的集體力量，創造、分享並評論屬於使用者自身或他人觀點的內容。而自 2002 年以後，從 Wikipedia 可以證明此種以群眾意見為基礎的內容創造及評論模式，不論在客觀數據及社會觀感中均具有可信的份量。

而在 Web 2.0 之前，網路商店提供的推薦清單，個別使用者並無法回饋有關推薦精確度的訊息，而其他使用者也無法自他人的意見回饋中獲益。

因此本論文將 Web 2.0 的精神與個人化推薦系統相結合，並應用在圖書館推薦系統中。本論文使用資料探勘 (Data Mining) 的協力式過濾 (Collaborative Filtering) 得出個別讀者的推薦清單，再經由讀者們對書籍難易度的評價、與個別讀者設定難易度的等級，過濾出難易適中的推薦書籍；同時經由讀者們對書籍 tagging 等 Web 2.0 的活動，重新對館藏進行分類，使得圖書館的藏書以一種更貼近當代讀者的面目呈現，以消除讀者對圖書館的隔閡。

本論文希望由量變產生質變，透過社群參與難易度的評價。讓讀者自行決定書籍的適當閱讀順序與分類，讀者將透過更親切、容易的方式找書，同時也讓前人的閱讀經驗得以留存，幫助後進者的求知之路。

關鍵字：Web 2.0、群眾標記、推薦系統、關聯規則探勘、圖書館

# **A Web-2.0 based Personalized Recommendation System for Library**

Student: Tzu-Wen Lo    Advisor: Dr. Hao-Ren Ke

Institute of Information Management  
National Chiao Tung University

## **ABSTRACT**

True to the Web 2.0 spirit of creating, sharing and tagging by open-communities, the content of websites are no longer provided by site owners but users. After 2000, Wikipedia, as one of the paradigms of Web 2.0 websites, proved that this kind of running model which made of people, tagging and review by people has earned trustworthy reputation in objective data and general impression.

Before Web 2.0 era, users could not response their feedback to recommendation list of online stores, with the result that stores could not improve the system by collecting feedback.

In order to solve the problem and offer an adaptive recommendation system that automatically adjusts recommendation result to users' preference by collecting response of users, our research combined Web 2.0 features with personal recommendation system and put in use in library. First of all, our research applies Collaborative Filtering, one of solutions of Data Mining, to obtain individual recommendation list. Secondly, system filters out unsuitable results depends on personal rating records, and in proportion to overall rating by all of the users. Therefore, the final recommendation list should be more accommodate to each user. Furthermore, users' tagging would also reconstruct library catalog, which could break down the barrier between library and readers.

Our research expects of quantitative and qualitative change phenomenon by encouraging community participation. Hence, by archiving and analyzing forerunners' rating, review and tagging records, following readers would be easier to find right books quickly.

Keywords: Web 2.0; collaborative tagging; recommendation system; Association Rule Mining; library



## 誌謝

二零零五年我到了島嶼內知名的晶圓城市，學校就在這些晶圓巨人旁，中午時分會看到圍繞在學校週邊公司的員工走向學校餐廳用午餐，從他們識途老馬的行走路線，我可以知道他們不久前曾是這裡的學生。

這裡大抵是這個國家產學結合得最緊密的示範園區，範圍內兩所學校與園區公司保持著少見的親密。學校建築以公司名號起名、實驗室研發計畫與業界合作；學生畢業了，也到附近的公司上班。我一個好朋友，從大學到博班都在另一所以梅為別名的學校，未來也該是在這個穩定發展的聚落，支持著百分之三十的國家 GDP。而我深深知道，我是多麼想望著遠方的風景。

一個夏天我認識了我的指導教授，在一個極短的時間內他對我投注極大的信任，給我完成論文與計畫足夠的發揮空間與權限。我像是一個擁有珍稀畫材與無窮顏料的畫家，在一個愉快的環境內，完成了二件重要作品。姑且不論我承受著歷來最大的身心壓力，從他而來的支持與收穫是我能夠安然渡過的最大原因，而對此我深深感激，他是柯皓仁博士，我的指導教授。有些人以姓名諧音叫他為好人，我覺得名符其實。

圖書館的同仁，思羽、玉菱、媛媛與佳欣，感謝你們傳授的 know-how，希望我不是歷來最難纏的學生。圖書館數位典藏工作室林孟玲小姐，身為一位專業的館員，對於我的論文與計畫給予的建議往往啟迪了更多創意的發想，而這價值非凡；我要謝謝孟玲姐在生活上的照料與忠告，這與平常的閱讀分享，共同成為我在圖書館工作期間最美好的經驗。

以交情度認購最低消費額、幫助填寫問卷的眾家親朋好友們，謝謝你們的鼎力相助，助我完成最後一章。謝謝實驗室歷屆學長的技术支援與學弟的支持，希望這互相支援的義氣（或說風格）得以傳承下去。

我的父母，我無後顧之憂的最大支柱，我們每年見面如報稅的次數，但任誰都知道彼此心裡的牽掛。胞弟子仁，雖然你大部份作的事都令我頭痛、但有在成長；我個人希望你速度快一點，我不想習慣性頭痛。

資訊大樓前莫內的畫，凌晨四點伴我歸途的月。我要離開了，再見。

# 目錄

<b>壹、緒論</b> .....	<b>1</b>
一、研究目的與動機.....	1
二、研究方法與目標.....	3
三、論文架構.....	4
<b>貳、WEB 2.0 相關研究</b> .....	<b>5</b>
一、WEB 2.0 的定義與起源.....	5
二、WEB 2.0 的現象與意義.....	9
三、WEB 2.0 的相關技術.....	11
四、與本研究相關的 WEB 2.0 網站.....	14
(一) <i>Eurekster</i> .....	14
(二) <i>Library Thing</i> .....	15
(三) <i>Shelfari</i> .....	17
<b>參、推薦系統與資料探勘相關研究</b> .....	<b>19</b>
一、推薦系統與資訊過載.....	19
二、圖書館的圖書推薦與資料探勘.....	20
(一) 使用者導向之網路資源組織.....	20
(二) 使用者導向之網路資源檢索.....	24
三、推薦系統主要架構.....	27
(一) 輸入與輸出.....	28
(二) 推薦方法.....	30
(三) 個人化與傳遞.....	31

四、協力式過濾與關聯規則探勘 .....	33
(一) 關聯規則探勘.....	33
(二) Microsoft SQL Server 2005 的資料探勘解決方案.....	36
<b>肆、系統實作.....</b>	<b>38</b>
一、系統架構.....	38
二、資料預處理.....	40
三、建立關聯規則模型及演算法預測查詢.....	41
四、排序個人書單及難易層級.....	47
五、系統實際操作畫面.....	51
<b>伍、資料分析與討論.....</b>	<b>54</b>
一、受訪者網路使用行為及接觸圖書資訊行為分析.....	54
二、受訪者圖書館使用行為分析.....	57
三、WEB 2.0 推薦功能需求分析.....	62
四、ABU 使用滿意度分析.....	63
五、綜合討論.....	64
<b>陸、結論與未來發展方向.....</b>	<b>67</b>
一、結論.....	67
二、未來發展方向.....	68
<b>柒、參考文獻.....</b>	<b>70</b>

## 表目錄

表 1 XMLHTTPREQUEST 示例 .....	12
表 2 XMLHTTPREQUEST 於不同平台瀏覽器示例.....	12
表 3 APRIORI 演算法。資料來源：(AGRAWAL AND SRIKANT 1994).....	35
表 4 ABU 必要軟體需求。資料來源：本論文.....	40
表 5 PREDICTIONASSOCIATION 預測查詢示例。資料來源：本論文 .....	45
表 6 受試者對於 ABU 功能的滿意度分析—平均數與標準差.....	64
表 7 受試者對於 ABU 系統的綜合滿意度—平均數與標準差.....	65



## 圖目錄

圖 1	WEB 2.0 概念圖。資料來源：(TIM O'REILLY 2005/9/30)	7
圖 2	WEB 2.0 BUBBLE MAP。資料來源：(MARKUS ANGERMEIER 2005/11/11)	8
圖 3	TIME BAR OF WEB 2.0 BUZZ WORDS。資料來源：(JÜRGEN SCHILLER GARICA 2006/9/21)	11
圖 4	EUREKSTER 以 GOOGLE WEB API 為基礎的社群搜尋引擎。資料來源：EUREKSTER 網站	13
圖 5	使用 GOOGLE MAPS 混搭而成的 FON MAPS。資料來源：FON 網站	14
圖 6	EUREKSTER 的回饋意見機制。資料來源：EUREKSTER 網站	15
圖 7	LIBRARYTHING 的書籍頁面。資料來源：LIBRARYTHING 網站	16
圖 8	LIBRARYTHING 列出書籍的評價分佈。資料來源：LIBRARYTHING 網站	17
圖 9	SHELFARI 的書架畫面。資料來源：SHELFARI 網站	18
圖 10	推薦方法分類。修改自資料來源：(卜小蝶 2007)	21
圖 11	交通大學學科資源網。資料來源：交通大學圖書館	21
圖 12	賓州大學 PENNTAGS。資料來源：賓州大學圖書館 PENNTAGS 網站	22
圖 13	交通大學圖書館讀者興趣屬性。資料來源：交通大學圖書館 PIE@NCTU 網站	23
圖 14	交通大學圖書館的個人化新書通告。資料來源：交通大學圖書館 PIE@NCTU 網站	24
圖 15	圖像檢索示例。資料來源：GETTYIMAGES 網站	25
圖 16	GOOGLE 的關鍵字推薦。資料來源：GOOGLE.COM	26
圖 17	交通大學圖書館目前提供的推薦服務。資料來源：本論文	27
圖 18	推薦系統主要架構。修改自資料來源：(SCHAFER 2001)	28

圖 19	AMAZON.COM 的個人化推薦。資料來源：AMAZON.COM 網站.....	32
圖 20	MICROSOFT SQL SERVER 2005 的圖形化探勘結果示例。資料來源：本論文.....	37
圖 21	ABU 架構圖。資料來源：本論文.....	39
圖 22	ABU 功能架構圖。資料來源：本論文.....	39
圖 23	ABU 借閱資料。資料來源：本論文.....	41
圖 24	ABU 關聯規則模型。資料來源：本論文.....	42
圖 25	ABU 探勘結果之相依性網路。資料來源：本論文.....	44
圖 26	PREDICTIONASSOCIATION 預測查詢結果。資料來源：本論文.....	46
圖 27	預先批次處理完所有讀者之預測查詢。資料來源：本論文.....	46
圖 28	ABU 網站之個人推薦畫面。資料來源：本論文之 ABU 網站.....	47
圖 29	ABU 網站中書籍評價畫面。資料來源：本論文 ABU 網站.....	48
圖 30	ABU 網站中讀者之閱讀階段畫面。資料來源：本論文 ABU 網站.....	49
圖 31	ABU 過濾法則。資料來源：本論文.....	49
圖 32	阿布的個人化書單。資料來源：本論文 ABU 網站.....	50
圖 33	圖書分類目錄。資料來源：本論文 ABU 網站.....	51
圖 34	ABU 的 TAG CLOUD。資料來源：本論文 ABU 網站.....	52
圖 35	在 ABU 查詢 AMAZON.COM 書目。資料來源：本論文 ABU 網站.....	52
圖 36	ABU 名家書單。資料來源：本論文 ABU 網站.....	53
圖 37	受訪者每天上網時間.....	54

圖 38 受訪者上網資歷.....	55
圖 39 受訪者逛書店或接觸書籍的頻率.....	56
圖 40 受訪者閱讀新書的比例.....	56
圖 41 受訪者閱讀學校教科書、參考書的比例.....	57
圖 42 受訪者親自到訪交大圖書館的頻率.....	57
圖 43 受訪者使用交大圖書館電子資源的頻率.....	58
圖 44 受訪者親自到訪與使用電子資源之比較.....	58
圖 45 受訪者使用館藏目錄與新書目錄之比例.....	59
圖 46 受訪者對於圖書館電子資源查詢及排序的結果滿意程度.....	59
圖 47 受訪者對於圖書館提供新知的滿意程度.....	60
圖 48 受訪者接觸新書的管道.....	61
圖 49 受訪者選書時會參考其他熱中閱讀讀者意見之比例.....	61
圖 50 受訪者對於 WEB 2.0 推薦功能的需求分析.....	62
圖 51 受訪者對於 ABU 的滿意度分析.....	63
圖 52 受訪者對於 ABU 系統的綜合滿意度.....	64

## 方程式目錄

方程式 1	PROBABILITY 公式示例。資料來源：本論文.....	42
方程式 2	IMPORTANCE 公式示例。資料來源：本論文.....	43
方程式 3	書籍評價計算公式。資料來源：本論文.....	47
方程式 4	讀者閱讀階段計算公式。資料來源：本論文.....	48





## 壹、緒論

### 一、研究目的與動機

圖書館在歷史上，對於保存及交流人類智慧佔有極重要的角色。諸如馬克斯（Marl Marx）、狄更斯（Charles Dickens）及蕭伯納（George Bernard Shaw）都是大英圖書館（British Library）的常客(The British Library)，孫中山更是在大英圖書館寫下了三民主義(Casey 2002-06-24)，足見圖書館的實質貢獻。

然而，在 1979 年美國 Pittsburgh 大學調查報告中指出(Kent and And Others 1979)，圖書館的館藏資源只有少部份被有效利用。各圖書館無不利用各種方式，帶領讀者博覽更多尚未被發掘的藏書。

其中，為發揮導覽群書的功能，圖書館通常了圖書分類系統。以國內常用的中國圖書分類法為例，將中文圖書依哲學、宗教、自然科學、應用科學、社會科學、史地、語文及美術等略分為十類。針對英文圖書則採用美國國會分類法，以學科分成二十四類。此類的分類方式，本文稱為 Taxonomy 分類法。分類的目的是為了檢索與瀏覽，然而，這類相對屬於小眾的、專業人士才熟知了解的分類方式，卻也限制了大眾的、業餘讀者面對知識的角度。

博客來網路書店張天立總經理指出，中國圖書分類法的十大類除圖書館員外，一般讀者之反應皆不佳(張天立 民 93)。若仔細觀察圖書館的分類並與網路書店比較，可以發現兩者的分類方式完全不同。網路書店為了更接近讀者的需要、同時也基於便於行銷上的考量，使用了對讀者而言較為親近的分類方式。

若能更進一步讓讀者自行「標記」(tagging)書籍的分類，再透過網站排序、重新分類書籍，便能夠提供比中國圖書分類法及美國國會分類法更為貼近讀者的分類制度，進而幫助讀者從檢索中更容易看到相關書籍。(游子賢 2005)也提到：「組織分群後的標籤的確能讓使用者更快熟悉其搜尋的概念，以及幫助其導覽」。因此，本研究採用了讓讀者自行標記的方式，重新定義書籍的分類。

欲將分類的權力重新交回讀者手中，便需依賴眾多讀者的共同參與，才能使圖書館的分類隨著標記資料筆數的累積愈來愈貼近讀者需求。而標記以及依賴群眾力量，即是 Web 2.0 典型的活動之一。

Web 2.0 的精神，是想透過社群的集體力量，創造、分享並評論屬於使用者自身或他人觀點的內容。而自 2000 年以後，社會大眾從 Wikipedia 可以證明此種以群眾意見為基礎的內容創造及評論模式，不論在質與量上、於客觀數據及社會主觀評價中均具有可信的份量。其中著名的一項事證，即是國際知名期刊《自然》(Nature) 在 2005 十二月發表的一份研究報告指出，Wikipedia 條目的正確性，與大英百科全書 (Encyclopedia Britannica) 不相上下 (Jim 2005/12/14)。研究者相信，同樣藉由讀者共同標記的活動，一種更符合讀者需求的分類制度將可由讀者自己創造。

除了分類系統之外，推薦亦是圖書館常用的推廣方法之一。推薦系統 (Recommendation Systems) 應用最廣的領域原是電子商務，目的是欲在無人服務的電子商業環境中塑造有專人服務的效果，以補足電子商務較不人性化的缺失。在傳統書店交易，由於經常往來，老闆常會根據個人需求與喜好額外推介相關書籍，此類客戶關係管理 (Customer Relationship Management, CRM) 的活動在電子商務裡就以推薦系統取而代之。而以圖書館而言，讀者的借閱歷史即是讀者與圖書館往來的記錄，也是圖書館利用資料探勘 (Data Mining) 技術實踐推薦時進行分析的資料來源。

本論文提出的推薦方法，主要是基於協力式過濾 (Collaborative Filtering) 的概念，以關聯規則探勘 (Association Rules) 找出在借閱記錄上具有相同興趣的讀者，作為推薦的依據。而為了能夠進一步蒐集讀者對於推薦的反應，以及對讀者提供更為精準的推薦清單，本論文提出一種讓讀者自定書籍難度的功能，該功能另一方面也定義了讀者本身的閱讀階級。

讀者可以提出他對推薦結果是否符合閱讀現況的回饋，以簡單、中間偏易、中間偏難到困難以四級表示。舉例而言，若推薦一讀者一系列關於經濟學的書籍，讀者大部份皆回饋書籍屬於中間偏難或困難，即表示讀者目前屬於入門階段，其他進階甚至專業書籍並不符合他

的需求。因此在系統儲存回饋後再次顯示的結果，將只顯示被歸類為入門的書籍。而書籍屬於四級中的哪一級，則視群體讀者的回饋而定。前述讀者的分級，同樣也自入門、進階、高級至專業分為四級，相同級別的讀者只會得到相同難易程度的書籍。本論文籍由此種協力式標記（Collaborative Tagging）的方式，進一步可為讀者過濾掉將近四分之三的推薦結果，留下的結果將更貼近讀者在閱讀程度上的需求。

(吳晨帆 2004)指出，推薦的形式中，以「專家測試報告」的命中率最高，而「討論區文章」效果最差。(余明哲 2003)也提到，人工選粹（Manually Selected）式、由專家人工建立推薦清單讓讀者自行參考的方式，具有相當高的準確度。意即與其讓讀者自行「爬文」、找尋資訊，不如由圖書館提供各式專家推薦清單。因此本論文建立的示範網站中建有名家書單功能，提供讀者閱讀的導引，同時也是讀者除了借閱歷史及自行檢索外，第三種標記及評價難易程度的來源。

## 二、研究方法與目標

本論文建立的「交大阿布 A.bu」系統（以下簡稱 Abu），以國立交通大學浩然圖書館讀者借閱記錄為基礎，利用資料探勘的技術，探索讀者借閱書籍的相關程度，並藉由讀者自行評價書籍的活動，擬出一份屬於個別讀者的推薦清單，提供讀者作為借閱的參考。

此外，本論文亦希望藉由讀者參與書籍 tagging 及評價的活動，累積讀者對於圖書館藏書的回饋，並期望藉由評價資料的量變造成質變，使得書籍的標記與評價記錄可以成為除了既有的圖書館詮釋資料（Metadata）之外，讀者借閱書籍前的另一重要參考資源。一方面除了使後人能夠減少在書目之間重新摸索的時間，另一方面也藉由此一電子化服務強化圖書館在閱讀導引上的服務。

本論文提出的推薦系統，結合了自動化推薦、讀者自行分群及社群的特色，提供了推薦清單、清單過濾及群體標記的功能，最大的目的是希望經由讀者對於書籍的回饋，重新組織圖書館藏書，以另一種別於 Web 1.0 的、小眾的、專家的方式，重新以一種更能夠讓新世代

大學圖書館讀者能夠接受的方式，引介圖書館的藏書，讓讀者能夠更有效利用圖書館的資源，並提昇圖書館的價值。

### 三、論文架構

本論文第二章，將會介紹 Web 2.0 發展的相關研究，以及將之與推薦系統整合後所顯現的價值。第三章說明推薦系統與資料探勘的相關研究，以及 Microsoft SQL Server 2005 提出的解決方案。第四章介紹推薦系統與讀者評價的整合方法，及應用於國立交通大學浩然圖書館 A.bu 系統的狀況。第五章以問卷調查分析讀者對於 A.bu 系統的使用滿意度。第六章為結論及針對 Web 2.0 之推薦系統未來的發展方向。



## 貳、Web 2.0 相關研究

本論文所提出的個人化推薦系統與現今大部份推薦系統的差異在於：使用了 Web 2.0 的技術手段，讓讀者能夠回饋推薦系統，並且能夠即時反應每次回饋的結果；同時，讀者有能力重塑圖書館藏書的分類。而為了能夠達成這些目的，必須使用一些 Web 2.0 的技術，達成 Web 2.0 網站所能夠期待的結果。

在本章中，第一節介紹 Web 2.0 的定義與起源；第二節說明 Web 2.0 在文化上的現象與意義；第三節則為 Web 2.0 網站經常使用（但非充要條件）的技術；第四節介紹與本研究或推薦系統相關 Web 2.0 網站。

### 一、Web 2.0 的定義與起源

Web 2.0 一詞由提姆·歐萊禮（Tim O'reilly）在 2004 與 MediaLive 公司的一場會議中第一次出現。這位創立 O'reilly Media 同名公司的網路先驅和他的共同創辦人戴爾·多爾帝（Dale Dougherty）認為，2001 年後的 dot-com 泡沫化狂潮及股市的衰退不盡然像是媒體所宣稱的，是一種過份炒作的結果，而是所有技術革命的共同特徵。歐萊禮認為，dot-com 的泡沫化及股市的衰退，不過是這些嶄新的技術開始佔領產業的中心舞台、而實力不夠堅強的冒牌貨（pretender）被淘汰的過程（Tim O'Reilly 2005/9/30）。

歐萊禮觀察，能夠在 dot-com 泡沫化後倖存的網站都有一些共同特徵，而這些共同特徵正是網站為了生存下來而進行的改變。歐萊禮認為，Web 2.0 即是一種下個世代軟體的設計形態與商業模式（a design patterns and business models for the next generation of software）（Tim O'Reilly 2005/9/30）。由於這種觀察在當時獲得眾人的支持，Web 2.0 一詞便開始廣為流傳。

由於 Web 2.0 並不完全偏重技術性或文化性，而是一個融合技術與文化的新詞；Web 2.0 的內涵也並未明確定義，遂引起諸多爭論，使得各家評論者紛紛以自身觀察定義了 Web 2.0。其中，又以活躍於各式媒體的科技評論家及專業部落客（blogger）的見解最為精采。例如，英國巴斯大學（University of Bath）UKOLN 研究中心的 Davis, Ian 認為：「Web 2.0 是一種態

度而非技術」(an attitude not a technology) (Ian 2005/7/4)。Web 2.0 Journal 及 AjaxWorld Magazine 的主編—Dion Hinchcliffe 認為，對他而言最好的詮釋(Dion Hinchcliffe)，是「Web 2.0 是由群眾所構成 (Web 2.0 is made of people)」。而台灣的網路家庭董事長詹宏志認為：「Web 2.0 是指網路上的服務、內容和行為」(吳靖雯 2006)。

有鑑於各家言論不一，歐萊禮於 2005 年的 Web 2.0 研討會提出了一個更完整的說明，認為 Web 2.0 應該包含以下原則(Tim O'Reilly 2005/9/30)：

將 Web 視為一種平台；能夠駕馭群體智慧；資料將變成未來的「Intel Inside」(即資料在未來即是最核心的元素)；軟體不斷發行與升級的循環將會終結(即 Web 2.0 應該是「永遠的 Beta 版」)；輕量型程序設計模型；透過內容與服務的整合使得輕量型的商業模式變得可行；軟體執行將跨越單一設備；豐富的使用者體驗；分享和參與的架構所驅動的網路效應；透過帶動分散的、獨立的開發者把各個系統和網站組合形成大整合的改革；具備拉動長尾的能力；快速的反應與具備雙向的互動。

歐萊禮將以上原則，以一張 Web 2.0 概念圖表現(見圖 1)。這張圖目前仍在演變，由於 Web 2.0 並沒有一個明確的界限，因此歐萊禮建議在解釋 Web 2.0 時以原則視之。同樣的，圖 1 的各項概念，以概念距離核心遠近的方法組成類似太陽系的圖形，方便表示 Web 2.0 的各項原則。

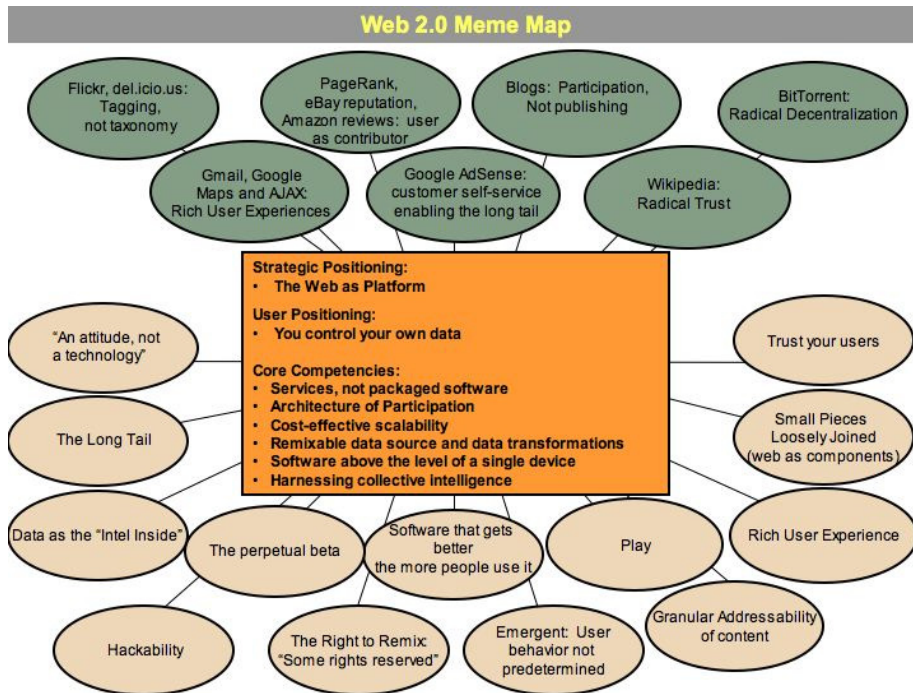


圖 1 Web 2.0 概念圖。資料來源：(Tim O'Reilly 2005/9/30)

之於歐萊禮的 Web 2.0 概念圖，Markus Angermeier 提供了一個變形(Markus Angermeier 2005/11/11)，以一種更 Web 2.0 的表示法、即 tag cloud 來視覺化歐萊禮的定義（見圖 2）。不同的是，Angermeier 加上了一些詮釋，諸如在「可用性」層面上（Usability）考量的「使用的愉悅」（Joy of Use）、「專注把一件簡單的事作好」（Focus on Simplicity）等。相較於原始的歐萊禮定義，Angermeier 的圖示更具體而微、視覺化了 Web 2.0 的內涵。





當然，依據 Web 2.0 的精神，Web 2.0 的精簡定義不會只有一家之言。《數位時代》認為(盧諭緯 2006)，Web 2.0 三大特點為：

1. 網站所有做的事，都要以使用者為中心出發。
2. 透過開放的討論，使用者的經驗才能愈來愈豐富。
3. 強調使用者網路的外部延展性。

而台灣師範大學圖書資訊學研究所卜小蝶認為(卜小蝶 2006)，Web 2.0 是：

1. 由大量使用者個別提供少數資訊，集成龐大的資料庫。
2. 全民參與：任何人可以輕易加入的網路環境，而不只是電腦怪才 (Geek) 的玩具。
3. 混搭系統：Google、eBay、Amazon 與 Yahoo 提供 API，新興系統創意無限。

## 二、Web 2.0 的現象與意義

網路家庭董事長詹宏志認為：「進入 Web 2.0 時代，對整個社會來說，是一個很大的權力轉移…這個權力轉移的過程，從中心化 (centralized) 到去中心化 (decentralized)，去中心化並不見得沒有力量，去中心化是透過每一個人獨力表達，最後可以找到中心化的方法」(吳靖雯 2006)。

以圖書館現存的分類系統而言，不論是中國圖書分類法或是美國國會分類法，皆是以專家意見為依歸，即是一種 centralized 的成果。此類分類法，即是 taxonomy。

Taxonomy 中譯為「分類學」，原義是表示生物體間的進化發展關係，然而演變至近代，在不同的領域下即有不同的意義，但主要功能在於對特定領域進行分類組織，形式上不脫樹狀圖或各式筆劃、字母等排序清單。然而，taxonomy 訂定後由於缺乏定期更新，反而使得當代讀者難以藉由分類找尋資訊。(Godfray 2002)也提到，若欲定期更新解決此一問題，又要找回同一批或相同領域的專家重新訂定分類，所費不貲。

與 taxonomy 互補的 folksonomy，是一個由 folks 與 taxonomy 組成的複合字，由 Thomas Vander Wal 在網路論壇 (mailing list) 中的討論所創(Gene Smith)，意即非由特定專家、而是由讀者自發性的經由大量的標記定義出分類的活動，意即詹宏志提到去中心化的具體案例。Folksonomy 最適合使用的場合，在於有很多使用者共同描述特定領域或範圍的資料。例如在 del.icio.us 中每個人都對 Web 2.0 相簿網站 Flickr 下了不同的 tag，在這些不同的 tag 中，del.icio.us 會顯示最常被使用的語彙，例如 photo、sharing、photography 等。相同的 tag 在其他網站，諸如 Gmail 也常使用，然而在 Gmail 下 tag 不會是 folksonomy 的原因在於，Gmail 的使用者並不能互相分享所下的 tag，進而形成某種分類。

有趣的是，若以圖書館為例，在 taxonomy 的分類制度下，一本書僅能屬於一個分類。但在 folksonomy 的多標籤定義下，書籍可以同時屬於好幾種分類。圖文書、跨領域的書種，在 folksonomy 的分類制度下將能夠更有效的被分類，而不只是就既有分類擇一。使用者也能就分類不當的書重新下 tag，當出現共識時（相同標籤數量多於門檻值）就可以取代舊有的、較不適合的分類，圖書館藏書分類遂逐漸趨向讀者的認知。Jon Udell 認為(Jon Udell 2004/8/20)：這類系統其基本上的差異，在於使用者的回饋 (the fundamental difference in these systems is feedback)。由於 folksonomy 貼近讀者的優點，Amazon.com 書店也開放讀者能夠自行標記書籍，同時保留 Amazon.com 現有的分類。由此可知，taxonomy 與 folksonomy 並不是取代而是互補的關係。

也由於使用者的共同參與，在分類制度、知識創造及分享上的成就，時代雜誌 (Time Magazine) 2006 首次以抽象的「你」作為年度風雲人物。時代雜誌認為：「你」為數位社會奠定了新的框架，無償的提供內容並在專業領域中擊敗專家 (And for seizing the reins of the global media, for founding and framing the new digital democracy, for working for nothing and beating the pros at their own game, TIME's Person of the Year for 2006 is you) (Lev 2006)。前者指的是 folksonomy，而後者指的便是 Wikipedia 條目的正確性與大英百科全書不相上下的成就(Jim 2005/12/14)。

### 三、Web 2.0 的相關技術

*Enterprise 2.0. Web 2.0 im Unternehmen* 一書作者，德裔的 Jürgen Schiller Garica 針對 Web 2.0 相關技術提供了一張加上時序尺度的圖表（見圖 3），從其中可以看到數項 Web 2.0 關鍵技術的演進(Jürgen Schiller Garica 2006/9/21)。

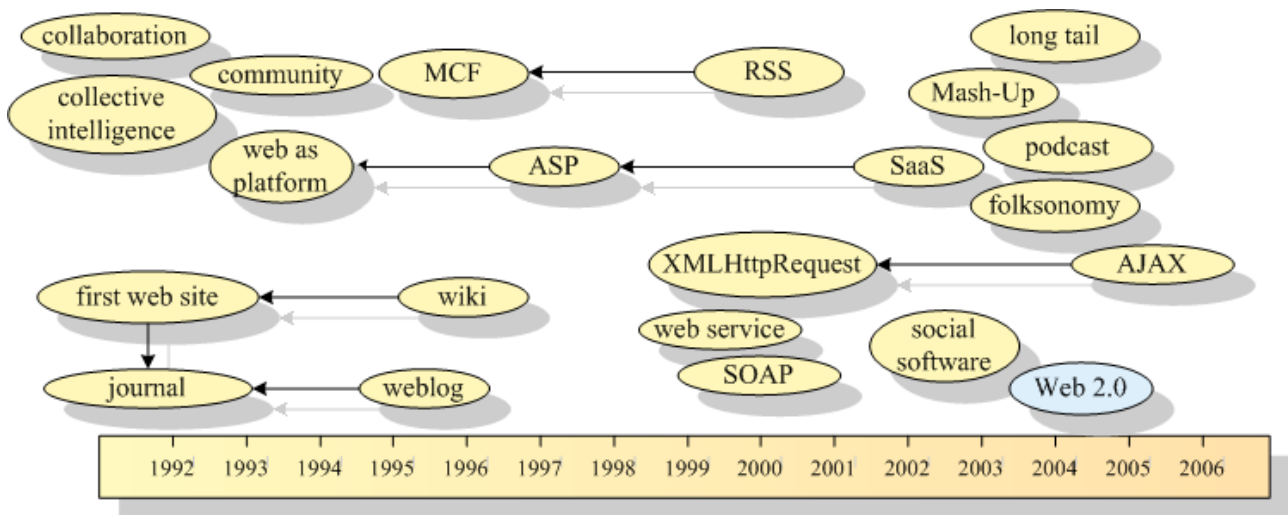


圖 3 Time bar of Web 2.0 buzz words。資料來源：(Jürgen Schiller Garica 2006/9/21)

其中最值得注意的，是名為 AJAX（Asynchronous JavaScript and XML）的技術。傳統的 Web 1.0 應用程式，在瀏覽器與伺服器之間傳遞資訊必須依靠 get、post 等動作完成，在使用經驗（User Experience）上是一連串的点選動作，網頁並無法在不重新載入網頁的情況下顯示最新的資訊。以往在視窗應用程式（Windows Applications）可以輕易達成從下拉式選單選擇一個項目、就可以直接帶出詳細訊息的畫面，反而在網頁上難以實現了。

AJAX 綜合了多種呈現（Presentation）、傳遞資訊、交換格式上的既有技術，包括 XHTML+CSS、JavaScript、DOM（Document Object Model）、SOAP、XML 等等，重新包裝成一種新的應用方式，因此它並不是指一種單一的技術。在前述的 Web 1.0 應用困境上，AJAX 採取了以 SOAP、XML 等技術傳遞必要資訊的方式，並在瀏覽器端以 JavaScript 和 CSS 處理來自伺服器端的回應，因此不需要重新載入整個畫面；因為傳遞的訊息變少，使用者可以發現網頁的反應速度變快了。此外，很多本來要送回伺服器端完成的工作，可以在瀏覽器端完

成，所以伺服器的工作負擔也減輕了。

AJAX 的核心是 JavaScript 的 XMLHttpRequest，用來處理伺服器與瀏覽器間的非同步訊息交換而不需重新載入網頁。一個最簡單的 XMLHttpRequest 範例如表 1 所示。

表 1 XMLHttpRequest 示例

```
<script type="text/javascript">
    var xmlhttp = new XMLHttpRequest();
</script>
```

AJAX 在本質上是一個瀏覽器端的技術，無可避免的會面臨瀏覽器的相容性問題。由於各家瀏覽器在 JavaScript、DOM、CSS 的支援不完全相同，甚至同一瀏覽器的不同版本之間也有支援度上的差異，使得 AJAX 的大部份開發時間都必須調整成相容成各家瀏覽器。以表 1 的範例而言，XMLHttpRequest 在不同瀏覽器的宣告方法則要改成表 2 之示例。

表 2 XMLHttpRequest 於不同平台瀏覽器示例

```
xmlhttp_request = new ActiveXObject("Msxml2.XMLHTTP.3.0");
xmlhttp_request = new ActiveXObject("Msxml2.XMLHTTP");
xmlhttp_request = new ActiveXObject("Microsoft.XMLHTTP");
xmlhttp_request = new XMLHttpRequest();
```

此外，由於 AJAX 依靠非同步的訊息傳遞，如果有多個 AJAX 同時送出要求，會造成伺服器應用程式的多程序（process）或多線程（thread）的競爭，因此程式設計師必須手動處理或避免此種情況發生，無形之中加重了工作負擔、也恐怕模糊了工作重心。然而也有許多有志之士致力研究 AJAX，想要開發一套解決方案根本解決上述問題，例如開放原始碼 ZK 及 Microsoft ASP.NET AJAX Framework。

其他純粹的技術性名稱，諸如 Social Software、ASP、SOAP 等，是達成 Web 2.0 功能的背後技術手段，本論文中不詳述。值得注意的是透過 RSS、SOAP、Web Services 等基礎建設，一種新的服務應用方式在 Web 2.0 時代應運而生，那便是「混搭」（Mash-Up）。

《數位時代》認為(盧諭緯 2006)：「Mash-up 的觀念，就是將各個網站不同的內容，不同的工具組合在一起」。Wikipedia 的定義提到(Wikipedia contributors)：「Mashup 是一種結合多種內容來源成為一個完整體驗的網站或軟體」(A mashup is a website or application that combines content from more than one source into an integrated experience.)。BusinessWeek 的解釋言簡意賅(Robert D. Hof 2005)：「主流服務的尋常(意指技術皆是現有的)組合」(homespun combinations of mainstream services)。

對於傳統的網站來說，能夠吸引人潮、並且盡量使人潮停留在自己的網站，是在 Web 1.0 時代重要的獲利關鍵(以吸引廣告量)。然而既有的網站無法提供滿足使用者的服務，有些專業使用者便開始抓取網站的內容、並剖析成自己需要的格式，重新呈現在自己的網站上。例如 Findbook.tw 即是一例 (<http://findbook.tw>)。Web 2.0 的網站開始開放了 API，使得使用者可以免費的、無償使用來自各種不同網站的內容，建立或拼湊出自己需要的服務。

例如 Eurekaster (<http://www.eurekaster.com>) 是一個以 Google 為基礎的社群搜尋網站(見圖 4)，本身並不擁有一個搜尋引擎，而是藉由 Google Web API 送回搜尋結果再加值成以社群偏好為主的特定需求搜尋引擎，目的是希望能夠更準確命中社群成員的內在需求。

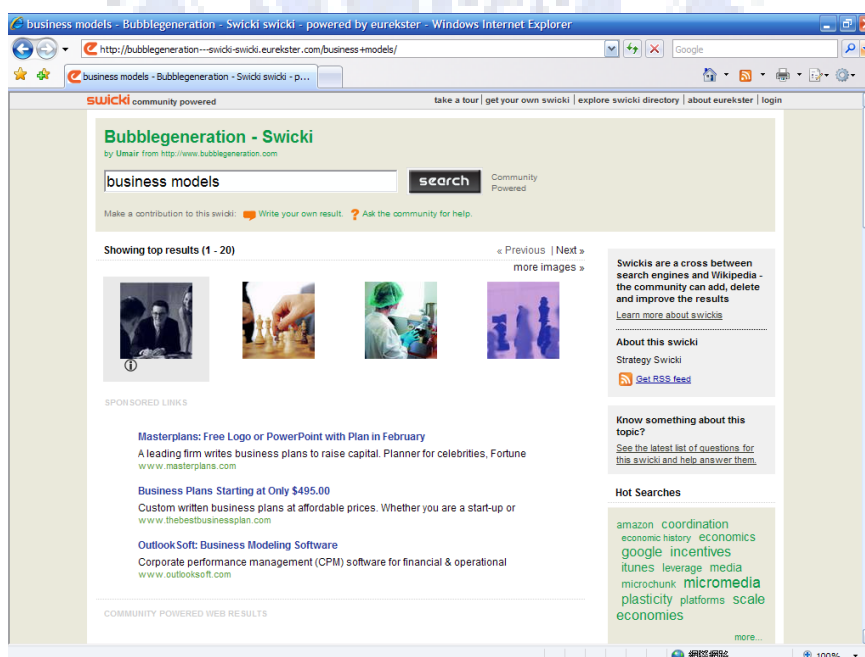


圖 4 Eurekaster 以 Google Web API 為基礎的社群搜尋引擎。資料來源：Eurekaster 網站

另一個有趣的混搭案例是 FON Maps (<http://www.fon.com>)。FON 是一個提倡使用者與其他需使用無線網路的使用者分享其無線頻寬，藉由 FON 用戶的串連，FON 的使用者走遍全球便可以免費使用其他同是 FON 使用者的無線網路（假設自己也是免費分享）。為了方便使用者快速尋找可以使用 FON 上網的地點，FON Maps 使用了 Google Maps API，在 FON Maps 上呈現世界地圖，以便讓使用者在地圖上標明自己開放 AP 的位置；同樣的，其他使用者旅經異地時，也可透過 FON Maps 找到開放的 AP（見圖 5）。

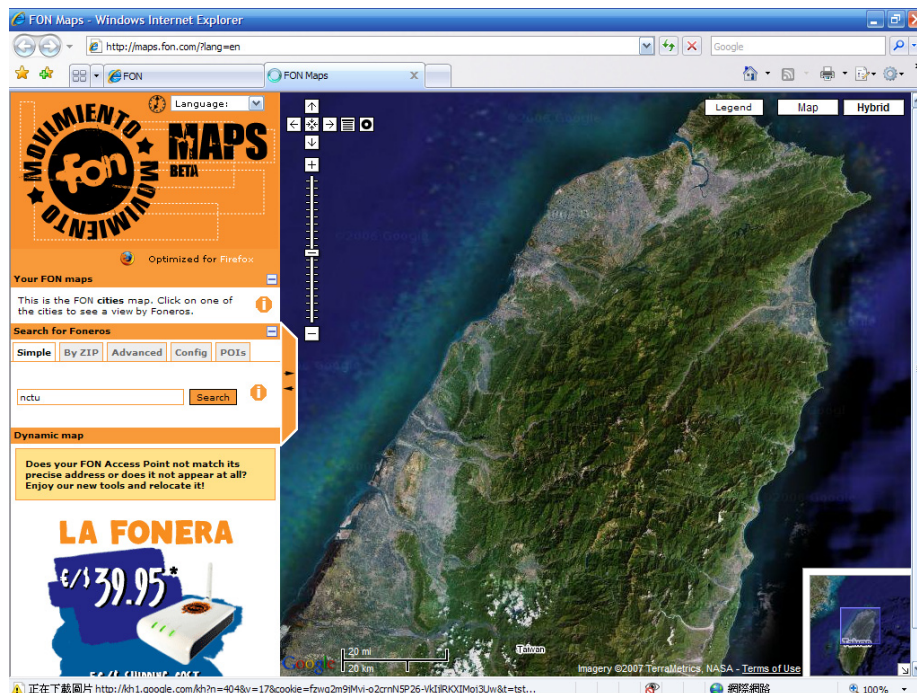


圖 5 使用 Google Maps 混搭而成的 FON Maps。資料來源：FON 網站

#### 四、與本研究相關的 Web 2.0 網站

與本研究相關的 Web 2.0 網站皆是利用某一特定技術，例如 community-based search 或 social web 完成一件應用。本節列舉三個與本研究相關的新應用，並說明其應用的功能與特色。

##### （一）Eurekster

Eurekster (<http://www.eurekster.com>) 是一個以 Google 為基礎的社群搜尋網站。網站使

用者可以成立一個自己的搜尋引擎並預先定義好該社群的偏好事物關鍵字，並邀請朋友加入社群。每當一位使用者使用該社群搜尋引擎查詢某一關鍵字並回饋意見，下一位同樣搜尋相同議題的使用者就可以分享到之前的搜尋結果。

以圖 6 為例，當使用者查詢 Mash up 一詞，並對結果清單的項目表示符合所需(點選 vote for this result 連結)，下次相同的搜尋時，該項目就會提高順位。反之，便會愈來愈往下排序。針對完全不滿意的結果，可以表示「很不滿意」(點選 vote against for this result 連結)，那麼該結果會直接刪除，下次相同搜尋便不再出現。

由於社群成員的稀少性，每個人的意見相較於 Google 有全世界的使用者而言，權重顯然大得太多，也因此每次意見的回饋都可以更快速地反應在下一次的搜尋。在社群成員具有相同愛好的偏好下，Eurekster 藉此種回饋機制增進搜尋的精確度。

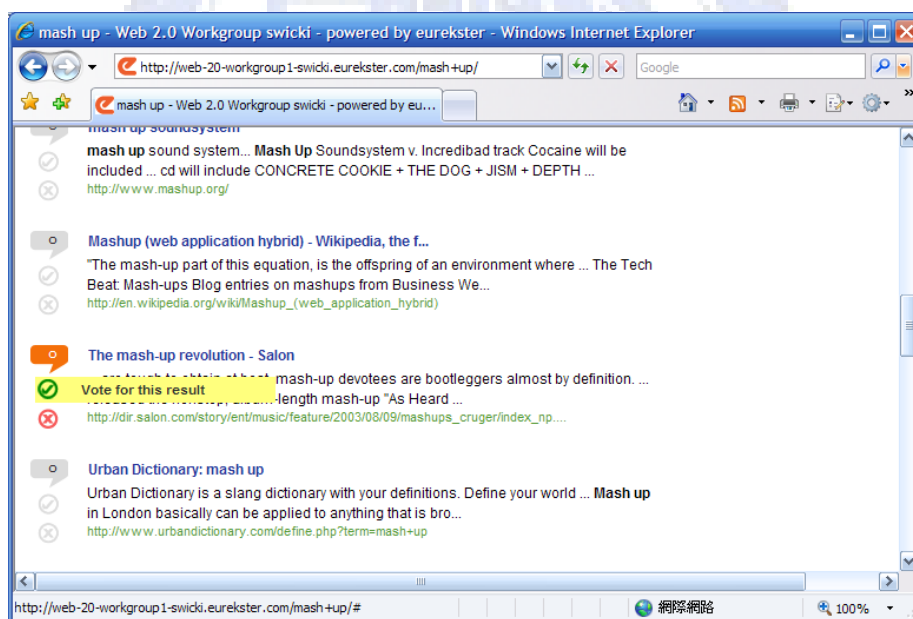


圖 6 Eurekster 的回饋意見機制。資料來源：Eurekster 網站

## (二) Library Thing

LibraryThing 對有大量閱讀習慣的讀者而言相當實用，讀者可以記錄已經讀過、準備閱讀以及想要推薦的書目，是目前使用者數量最多的個人藏書管理網站。LibraryThing 的一大

優勢在於可查詢全球近七十個圖書館的藏書，因此年代久遠或較學術類的書都可納入；

LibraryThing 不想讓使用者自行建入書籍資料，實際上也沒有這個功能。使用者僅需輸入書名或 ISBN，LibraryThing 即會自 Amazon.com 或其他資料來源擷取書籍的詮釋資料(Metadata)及封面。讀者可自行評價書籍、加上 tag。如圖 7，研究者加進了 *Tuesday with Morrie* 一書，便可以看到其他讀者為這本書下了什麼 tag，進而透過該 tag 找到類似書籍，發揮分類導引的功能。

此外，LibraryThing 還會列出其他同樣讀過這本書的讀書，推薦你其他類似的書籍。因此，Mitch Albom 繼 *Tuesday with Morrie* 後再次推出的同類型動人小說，*The Five People You Meet in Heaven* 就會在推薦清單之列。同樣的，LibraryThing 也利用了直觀的評價機制，書籍的評價從一到五顆星分為五級，並會列出評價分佈（見圖 8）。



圖 7 LibraryThing 的書籍頁面。資料來源：LibraryThing 網站



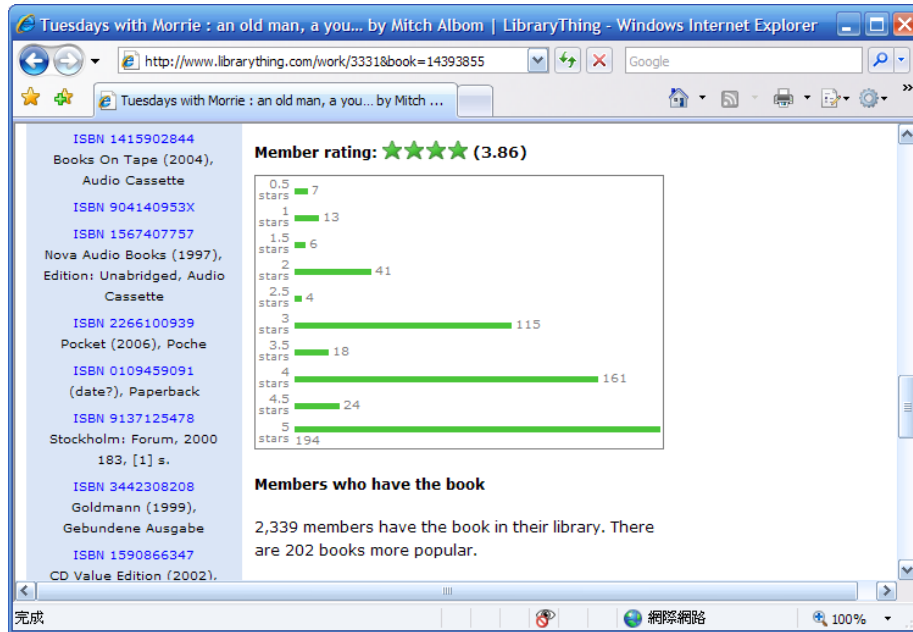


圖 8 LibraryThing 列出書籍的評價分佈。資料來源：LibraryThing 網站

### (三) Shelfari

Shelfari 是另一個藏書管理網站，它的圖書資料來源完全來自 Amazon.com，營運模式也是透過連結導引使用者到 Amazon.com 購書獲取佣金。而它的營運模式也被證實可行，是極少數 Web 2.0 網站、尤其是此類藏書管理網站能有正向現金流量的典範，在 2007 年獲得 Amazon.com 投資( Josh )。

與 LibraryThing 不同的是，Shelfari 強化了社群的機制。讀者可自行成立 My Groups，一個類似讀書會的討論版並邀請朋友加入。另可將書籍歸類到 Top 10 List (十本必讀好書)、Wish List (最想要的贈書清單) 等內建分類，於是每位讀者將會有自己的 Top 10 List，無疑是另種去中心化後的「名家書單」。



圖 9 Shelfari 的書架畫面。資料來源：Shelfari 網站



## 參、推薦系統與資料探勘相關研究

自 Amazon.com 以它優異的推薦系統精準地分析客戶的偏好，適時推薦客戶購買商品，極大程度滿足了客戶的需求，並藉此成為市場上最成功的網路書店後，人們對推薦系統的熱情又再度燃起。而要達成推薦，背後需依靠資料探勘技術的支援，才能在大量的書籍及客戶資料中找出其關聯，主動提供推薦清單，免除客戶資訊過載的困擾。

在本章中將說明推薦系統的目前發展。第一節說明資訊過載與推薦系統的應用；第二節為現今推薦系統的主要架構及設計上需考量的議題。第三節說明推薦系統常用的協力式過濾，以及本研究使用的 Microsoft SQL Server 2005 在關聯規則探勘上提供的解決方案。

### 一、推薦系統與資訊過載

以國立交通大學浩然圖書館為例，館藏數量至 2007 年二月底止有近二百萬筆書籍、期刊、微縮資料片等物件，每月另有新書一批批上架，在如此大量的資料內找尋所需資料，造成了讀者的資訊過載 (Information Overloading)。資料過多，變成一個問題。如何過濾這些「大量」資料找到自己需要的「少量」資料，成為一個更關鍵的課題；對於在圖書館內遍尋不著的讀者而言，更是一種不愉快的尋書經驗。高震宇的研究提到，在網路購物情境中，消費者在高資訊量以及沒有資訊過濾機制的情況下，所認知到的資訊過載的確比在低資訊量以及有過濾機制的情況下來得高…此外，認知到較高的資訊過載的消費者，對於他們的購物決策結果傾向於有較差的感受(高震宇 2004)。

為了能更有效率的檢索資料，各種加強搜尋結果的技術與研究遂百花齊放，諸如資訊過濾、自然語言查詢、多媒體資訊檢索技術、個人化、網路代理人、資料檢索等等工具便應運而生。

其中，資訊檢索 (Information Retrieval) 是使用者最為熟知的查詢方式之一，普遍應用在各種搜尋引擎上，包括輸入搜尋條件的介面、搜尋演算法及資料庫(Oard and Marchionini 1996)。此類的搜尋引擎都是使用搜尋關鍵字與資料的內容進行比對並找出符合條件的資料，

但前提條件是，使用者必須先知道自己要查詢的確切字彙。而在進階搜尋中提供較多的功能選項，諸如布林邏輯、同音字元比對等，目的是幫助使用者透過更仔細的描述，找出為數較少的候選資料。然而曾元顯提到(曾元顯 民 86)，一般使用者對布林邏輯的運用較不熟悉、檢索結果沒有依照符合程度排序、檢索字串要求精確無誤等原因，導致較高的檢索失敗率 (search failure) 與資訊過載 (information overloading)。

資訊過濾 (information filtering) 則是藉由去除資訊流 (information stream) 中重覆或不符條件的資料以達成過濾的功能，希望藉此降低資訊過載及語意上的雜訊 (semantic signal-to-noise ratio)。為了達成過濾，技術上是比對使用者身份設定 (profile) 與文件的特性，並篩選出相符的資料。

然而，不論是資訊檢索或過濾，皆是被動式地反應使用者的輸入而回應，對於資訊超載並不能起較大的效用，因此需要一個能夠因應不同使用者偏好而主動出擊的解決方案，減少使用者需手動檢索，再逐一從結果清單中挑選所耗費的心力 (effort)。而推薦正是一個適當的選擇，也普遍應用於各電子商務網站。Resnick 認為(Resnick 1997)，推薦除了過濾資訊之外，還能夠給使用者有興趣的資訊。Schafer 的觀點較偏向電子商務，他認為推薦系統即是電子商務網站對客戶提供足夠的訊息，以幫助他們購買時的決策(Schafer 2001)。

## 二、圖書館的圖書推薦與資料探勘

資料探勘大多使用於具有大量交易或歷史資料的應用情境，諸如零售業、網站書店等等，由於資料都不具匿名性、可追蹤到單一客戶的偏好，因而能夠推薦個別使用者有興趣的物件。對網路書店而言，推薦能夠提高顧客的回店誘因，進而促進消費；對圖書館而言則屬於經營層次，能夠提高圖書資源的能見度，吸引讀者來館借閱，提高資源的利用率。

### (一) 使用者導向之網路資源組織

能夠達到推薦目的有很多種方法，大致可以區分為「被動式目錄」與「主動式檢索」。

卜小蝶以更 Web 2.0 的方式，將應用在圖書館的推薦區分為「使用者導向之網路資源組織」與「使用者導向之網路資源檢索」(卜小蝶 2007)。該架構以圖 10 表示如下。

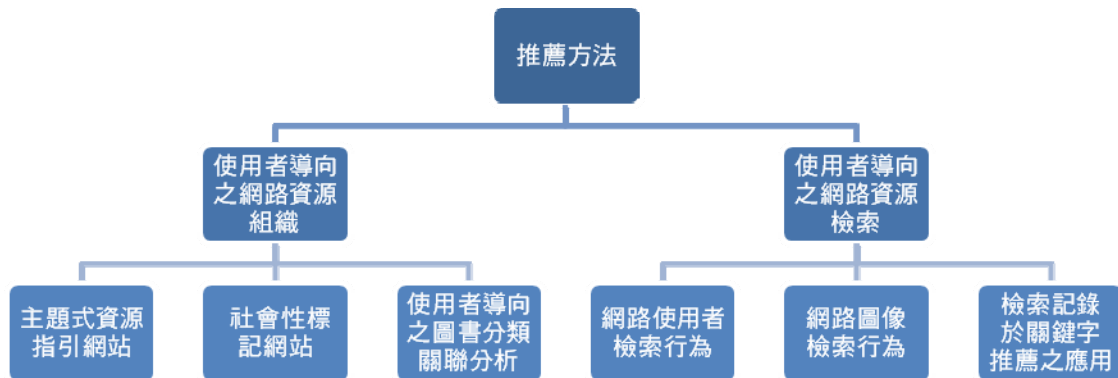


圖 10 推薦方法分類。修改自資料來源：(卜小蝶 2007)

主題式資源指引網站 (SG, Subject Gateway) 是一種經過高品質篩選、並以各特定學科或研究領域將網路資源分類的目錄索引。SG 通常由圖書館建置維護，由專業館員及學科專家貢獻內容，故多以學術性見長，也符合圖書館要服務的目標讀者。國內目前建置有 SG 的圖書館包括了交通大學、台灣大學等校。



圖 11 交通大學學科資源網。資料來源：交通大學圖書館

交通大學圖書館採取自行建置的方案，提供分類目錄及搜尋的功能，現階段以人工處理的方式建置，並以經營社群的方式鼓勵各學科專家無償貢獻內容。圖 11 即為交大學科資源網的分類索引畫面。以交通大學圖書館而言，設有常態性的推廣活動鼓勵學生主動推薦學科資源，以增加學生參與的誘因快速累積資源，是國內圖書館較少見的「虛實整合」的行銷案例。台灣大學圖書館則與 Yahoo! 奇摩合作，以搜尋方式為應用主軸，同樣以社群方式鼓勵學生主動推薦學術資源。

為了能解決人工建置 SG 的成本及效率問題，國際上另在進行 SG 自動索引及分類的研究計畫，包括 INFOMINE、GERHARD 及 Scorpion。目前執行成果較佳的 INFOMINE 是由美國加州大學河濱分校 (UCR, University of California, Riverside) 圖書館建置，已具有超過十萬筆網路資源索引，其中約四分之三是機器自動產生、四分之一是人工建置。UCR 圖書館用來建置 INFOMINE 的開放原始碼軟體相關套件名為 iVia，主要功能是網頁的抓取 (Crawling) 及詮釋資料自動編製 (Metadata Assignment)，UCR 圖書館也無償提供下載<sup>1</sup>。



圖 12 賓州大學 PennTags。資料來源：賓州大學圖書館 PennTags 網站

<sup>1</sup> iVia Project 網址 <http://ivia.ucr.edu/>

「社會性標記網站」(Social Tagging Web) 則是圖書館較少參與的領域。一方面原因是 Web 2.0 的風氣與應用在 2000 後才陸續發酵，二方面圖書館對於標記的精確度與品質也持保留態度，反倒是圖書館之外的網站對於 tagging 的應用顯得更為熱中。

學術性的社會性標記網站有賓州大學圖書館的 PennTags<sup>2</sup> (見圖 12)，其標記目標較多元，包括網頁、文件、書籍等，也與賓州大學圖書館相結合，可反查標記書籍是否可供借閱及館藏地點等。LibraryThing 則是專注在書籍的標記與評價 (見「與本研究相關的 Web 2.0 網站」一節，頁 14)，不能標記其他電子資源。

總類	哲學類	宗教類	自然科學類	應用科學類	社會科學類	史地類	世界史地類	語文類	美術類
0.1235	0.2672	0	0.224	0.218	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2672	0	0.224	0.218	0.2963	0	0.1287	0.1481	0.134
0.1235	0.2672	0	0.224	0.218	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2672	0	0.224	0.218	0.2201	0	0.1287	0.5926	0.134
0.1235	0.2672	0	0.224	0.218	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2672	0	0.224	0.218	0.2201	0	0.1287	0.6942	0.134
0.1235	0.2672	0	0.224	0.218	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2672	0	0.224	0.218	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2672	0	0.224	0.218	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2672	0	0.5326	0.218	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2672	0	0.224	0.218	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2138	0	0.5326	0.2725	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2138	0	0.4092	0.2725	0.2963	0	0.1287	0.1481	0.134
0.1235	0.2138	0	0.7196	0.2725	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2138	0	0.5326	0.2725	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2138	0	0.4092	0.2725	0.2201	0	0.1287	0.1481	0.134
0.1235	0.2138	0	0.4092	0.2725	0.2201	0	0.1287	0.1481	0.134

圖 13 交通大學圖書館讀者興趣屬性。資料來源：交通大學圖書館 PIE@NCTU 網站

「使用者導向之圖書分類關聯分析」為圖書館著力最多的部份。透過大量讀者的借閱歷史，圖書館研究的目的是在於分析讀者的借閱類別及數量，以便將有限的預算花費在最值得投資的館藏中；或分析讀者的閱讀模式，以了解讀者中是否形成數個具有明顯特徵的族群，預測其未來的借閱需求並提供推薦服務。技術上常採用關聯規則探勘或分群分析，找出書籍與

<sup>2</sup> PennTags 網址 <http://tags.library.upenn.edu/>

書籍的關係、讀者與讀者的關係及圖書與讀者的關係等三種預期的結果。

戴玉旻就曾以交通大學圖書館借閱記錄作資料探勘(戴玉旻 2001)、找出圖書與圖書之間的關係，並針對不同系所讀者找出相關規則。為了讓規則更容易使用，探勘結果導入交通大學個人化數位圖書資訊環境 PIE@NCTU 中(楊雅雯 民 90)，讀者可以在網站上了解自己的興趣屬性，如圖 13。

余明哲更將交通大學圖書館的推薦系統推進了個人化的層次(余明哲 2003)，以乏析式的協力式過濾找出讀者間的關聯性並得出推薦清單，再以內容導向過濾排序，找出符合讀者興趣的館藏。

## (二) 使用者導向之網路資源檢索

「網路使用者之檢索行為」則指使用者自行輸入關鍵字查詢，從分類目錄瀏覽、全文網頁搜尋到使用者導向排序，網路使用者歷經了 Yahoo! 早期、AltaVista 到 Google 等三個搜尋時期的應用典範。關鍵字查詢的方式無疑是現今最容易使用且無需學習的檢索方式。



圖 14 交通大學圖書館的個人化新書通告。資料來源：交通大學圖書館 PIE@NCTU 網站



然而隨著使用者的多元發展，逐漸無法照顧到個別使用者的需要，搜尋引擎開始利用檢索記錄分析（Log Analysis）的方法來記錄使用者的操作，並依此產生一些系統關鍵字來代表使用者的興趣。楊雅雯應用在交通大學圖書館的 PIE@NCTU 便是一套以智慧型個人化查詢系統來處理查詢館藏的應用範例(楊雅雯 民 90)。當使用者查詢館藏資料時，系統會利用系統判斷關鍵字、使用者自訂關聯字和使用者自選類別等三類資訊，過濾、排序出最需要的館藏並得出清單。另外，此份個人興趣設定也同步套用到圖書館的新書通報功能上。當圖書館有新書到館時，系統會過濾使用者的興趣，並發送符合條件的新書通告給讀者，如圖 14 的操作畫面。

「網路圖像檢索行為」指的是搜尋目標為「圖像」而非文字，例如輸入關鍵字「林懷民」，應該出現的結果是「林懷民」的圖像而不是文字資料。概念式的圖像檢索（Concept-based）通常以圍繞在圖像旁邊的文字（Surrounding Texts）建立文字索引，而使用者輸入關鍵字後，系統再以關鍵字查詢索引，這點與文字搜尋並無相異之處。圖像內容式的檢索（Content-based）則與圖學有關，以顏色、形狀、樣式（Texture）、空間（Spatial）等特徵為主，利用圖像擷取技術依此建立索引，使用者再利用這些特徵查詢圖像。

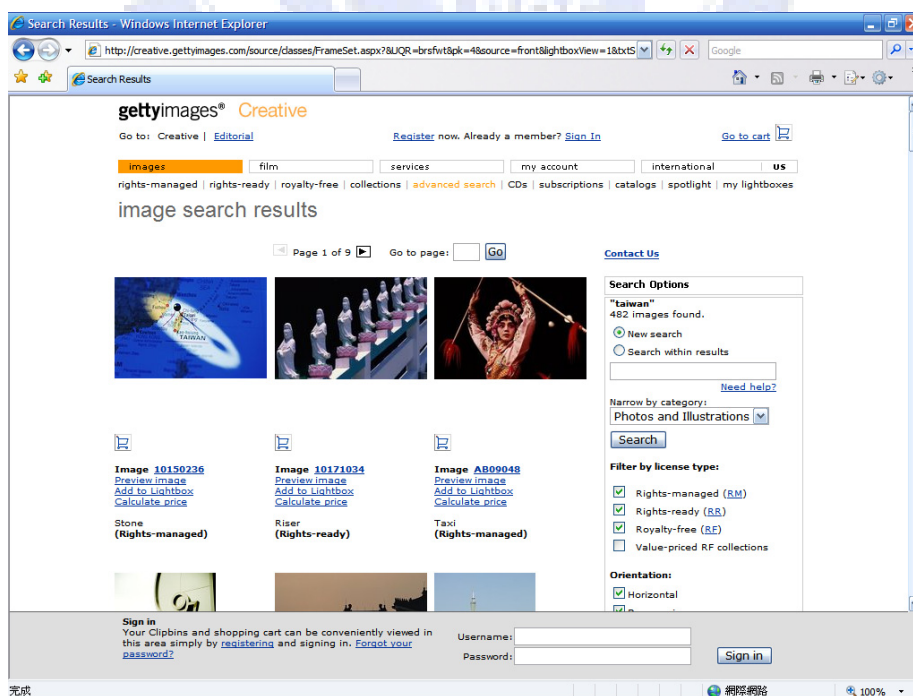


圖 15 圖像檢索示例。資料來源：GettyImages 網站

著名圖庫網站 GettyImages 便使用概念式的索引方法為使用者建立品質較高（因為人工處理）的圖像搜尋引擎。從圖 15 輸入 taiwan 回傳結果包含有台灣地圖、大中至正門、京劇演員、台北 101 等內容的多樣性看來，的確是圖像內容式檢索較難達成的效果。

以圖書館較偏重文字閱讀的發展方向而言，類似的圖像檢索較少見到實際應用。

「搜尋記錄於關鍵字推薦之應用」是一種利用關鍵字自動抽取（Keyword Extraction）及互動式資訊檢索技術（Interactive Information Retrieval）發展的推薦方法，其推薦的方法是在輸入關鍵字時、自動建議相關的關鍵字，而兩關鍵字經常能組成一關鍵詞。利用此種方法，能夠有助聯想並擴大檢索範圍。



圖 16 Google 的關鍵字推薦。資料來源：Google.com

Google 應是目前最好的關鍵字應用示範，圖 16 是輸入交通大學的畫面。Google 依據搜尋的結果數量及 PageRank 分數過濾並排序建議關鍵字清單。

圖 17 整理了目前交通大學圖書館提供的推薦服務，楊雅雯、戴玉旻與余明哲合力接續研究發展的 PIE@NCTU 為交通大學圖書館使用者導向之圖書分類關聯分析與智慧型檢索功能奠下基礎；2007 新開放的交大學科資源網則為大學圖書館讀者建立了主題式資源索引的交

換分享平台；本論文實作的 Abu 系統兼具社會性標記網站及使用者導向之圖書分類關聯分析功能，補足了交大圖書館在「使用者導向之網路資源組織」上的功能。

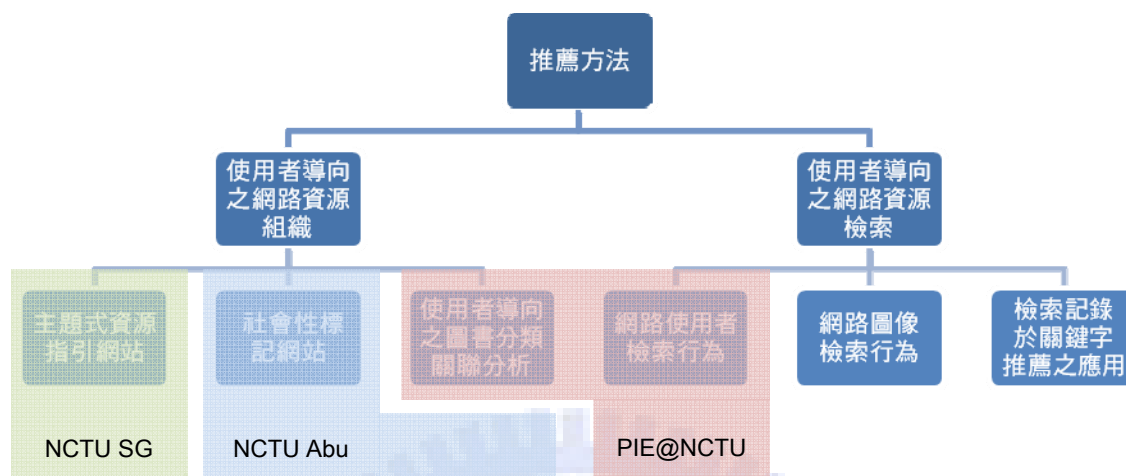


圖 17 交通大學圖書館目前提供的推薦服務。資料來源：本論文

### 三、推薦系統主要架構

推薦系統的主要架構，一般認為以 Schafer 提出的架構較為完整(Schafer 2001)。在此架構下，Schafer 將系統分為社群輸入變數 (Community inputs)、個人化程度 (Degree of personalization)、遞送 (Delivery)、個人輸入變數 (Target customer inputs) 及推薦方法 (Recommendation method)，其架構如圖 18。

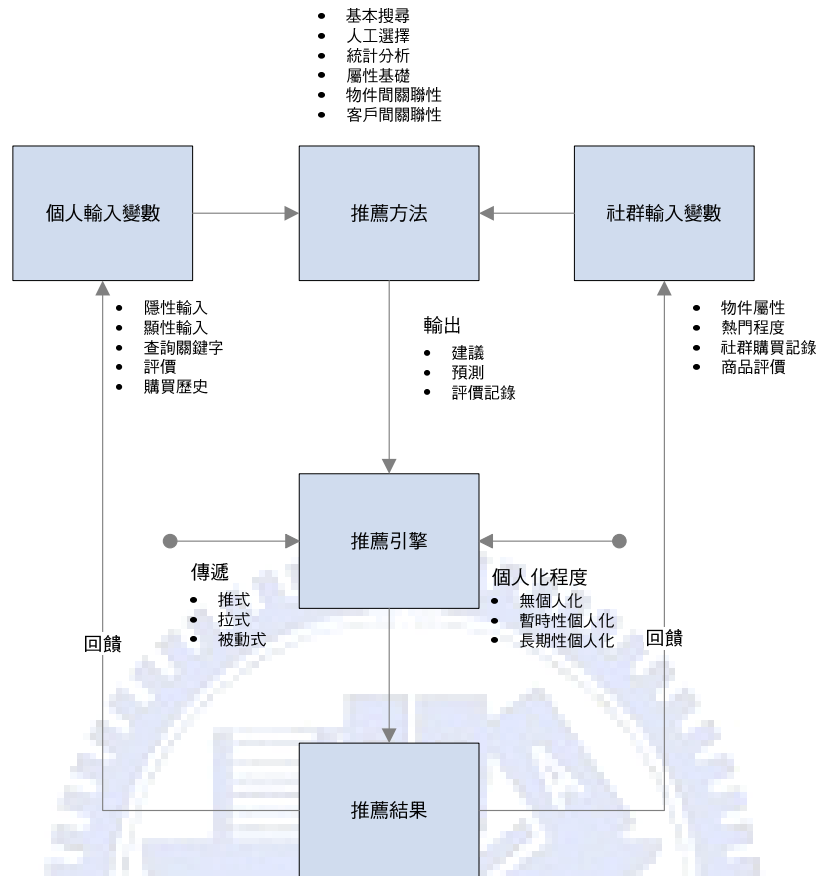


圖 18 推薦系統主要架構。修改自資料來源：(Schafer 2001)

## (一) 輸入與輸出

「輸入」指進行推薦分析之前需要輸入的資料，Schafer 將之區分為個人輸入變數及社群輸入變數。個人輸入變數包括了以下五個項目(Schafer 2001)，若能完整記錄則能較完整的描繪使用者的「樣貌」(profile)。

1. 隱性輸入 (Implicit Navigation)：由使用者的瀏覽記錄可以得知使用者有興趣的物件，這些平常的瀏覽活動，包含停留時間、順序、點選流向等都在表示了使用者獨特的興趣與偏好。例如 Amazon.com 就是使用隱性輸入的典型網站。
2. 顯性輸入 (Explicit Navigation)：以問卷方式請使用者自行填入偏好，通常在加入會員時輸入，平時也可修改。Amazon.com 在新會員註冊時也會要求 (選填) 新用戶

填寫。

3. 查詢關鍵字 (Keywords/Item Attribute)：將使用者輸入的關鍵字也視為喜好。
4. 評價 (Rating)：使用分級制 (如一至五顆星) 或二元式 (喜歡不喜歡) 對購買過的物件評分。
5. 購買歷史 (Purchase History)：使用者真正去購買的物件表示了最強的偏好，此類記錄也應視為是一種隱性輸入。

社群輸入變數可以描繪群體對物件的偏好。由於是社群總體對物件的大量記錄，因此可以取得物件的屬性資料，這是與「個人輸入變數」欲取得個人資料最大的不同。社群輸入變數包括了：

1. 物件屬性 (Item Attribute)：包括物件目標、客戶標籤與物件類別。例如「影片」類型相較於「二次世界大戰後期紀錄片」的等冷門類型，前者較可能擴大目標客戶群。由此可以判定商品是屬較熱門或冷門商品。
2. 熱門程度 (External Item Popularity)：特定節日如耶誕節會有大量需求，或是本身已是熱門商品，通常銷量較高。由此可以判定商品是平常即暢銷或節日熱銷。
3. 社群購買記錄 (Community Purchase History)：社群的購買記錄是由個人記錄構成，因此可以找出商品的銷售趨勢，或是相似商品之間的關係 (例如替代性、互補性)。應視為另一種隱性輸入。
4. 商品評價 (Rating and Text Comments)：有些網站特別鼓勵使用者留下文字評價。文字比星等或喜不喜歡等更容易吸引顧客。

一個優良的推薦系統應該綜合個人以及社群輸入變數，以便作出最符合客戶需求的推薦。

在「輸出」部份，Schafer 的架構將推薦的輸出分為三級(Schafer 2001)，包括：

1. 建議 (Suggestion)：指未排序過的推薦。通常直接放在網頁上以「試試這個 (try this)」

的形式呈現、通常是單獨項目（只會一次出現一個推薦項目），例如 Amazon.com 的個人化首頁。

2. 預測（Prediction）：有結合或不結合個人化偏好二種預測推薦。Schafer 指出 MovieFinder 對電影即提供了「大家的評價」與「我們的評價」兩種預測推薦。Amazon.com 對書籍也有專家書評以及讀者書評兩種，其評分也是獨立計算。
3. 評價記錄（Rating）：讓使用者除了看到總分之外，還可以看到個別評分記錄，以方便了解大家的評價是否一致。

## （二）推薦方法

Schafer 的架構(Schafer 2001)將推薦方法整理為五類，而此五類各有優缺點、並非是替代關係，通常可整合多種方法達成所需的推薦效果。

1. 基本搜尋（Raw Retrieval）：即網站上的搜尋功能，以關鍵字比對。以 Amazon.com 為例，其提供的是全文檢索，然而大部份網站僅提供特定欄位之搜尋，在此狀況下使用者不一定可以找到未登錄在主要欄位關鍵字之資料。
2. 人工選擇（Manually Selected）：人工建立推薦清單，由於通常是專家建立，具有較高的參考價值，例如《中國時報》人間副刊的年度選書、誠品選書等名家書單。
3. 統計分析（Statistical Summaries）：提供總體使用者的統計數據，這也是網站常使用來推薦熱門商品的方法，例如熱門一百大、每月暢銷書榜等。此類排行榜通常能發揮導引功能並成為閱讀指標，然而也招致可能造成閱讀普及化障礙的疑慮(陳慧敏)。
4. 屬性基礎（Attribute-based）：以產品的詮釋資料屬性作為推薦的判斷，通常會與使用者的偏好或動作連結。例如推薦新商品給曾買過同類產品的使用者，即內容導向過濾（Content-based filtering）。

5. 商品間關聯性 (Item-to-Item Correlation): 商品之間可能存在某些關聯, 例如印表機、印表耗材的互補關聯, 或單冊書籍、散文精選的替代性關聯等。例如 Amazon.com 在使用者將一商品放入購物車時, 會看到一頁「買這件商品的人, 也會同時買下列商品」的資訊。
6. 顧客間關聯性 (User-to-User Correlation): 根據使用者與使用者之間的關聯性作為推薦基礎, 即協力式過濾 (見協力式過濾與關聯規則探勘, 頁 33)。使用者之間可能因類似的操作習慣或偏好, 可歸為同一類別中, 而相同類別的群體意見會影響個人的推薦結果。

### (三) 個人化與傳遞

Schafer 提到仍有其他在設計上需要考慮到的議題(Schafer 2001), 包括了個人化與傳遞。

個人化包括了三種, 分別是無個人化 (Non-Personalized)、暫時性的個人化 (Ephemeral Personalization) 以及長期性的個人化 (Persistent Personalization)。無個人化指的是大家在同樣的操作下會得到相同的推薦, 例如查詢某本書, 網站會推薦同類型的書籍。暫時性的個人化是以一個時期 (Session) 作為推薦基礎, 例如瀏覽時的點選流向作為推薦基礎 (見圖 19 紅框處, Recommended Based on Your Browsing History)。通常是以「項目間的關聯性」與「屬性基礎」兩種方法作出推薦。長期性的個人化需要追蹤使用者長期的使用記錄, 例如交易歷史或圖書館的借閱歷史檔, 並利用資料探勘的方法找出使用者之間的關聯。



圖 19 Amazon.com 的個人化推薦。資料來源：Amazon.com 網站

在經過輸入、分析後，最後即是傳遞推薦的方法。好的網站通常會巧妙的融合以下三種傳遞方式，在最適當的時機送出推薦結果，得到事半功倍的效果；反之，則容易干擾使用者，甚至降低推薦效果。傳遞包括了推式（push）、拉式（pull）及被動式（passive）。推式即指使用者無需任何動作網站即會自動送出推薦結果，常見的例子如廣告信；此種效果通常最低，因為大部份廣告推薦信皆未作到個人化，且發送太頻繁造成困擾；至今除了 Amazon.com 少量使用外，其他網站少見廣告推薦信。拉式即是使用者主動要求，例如點選連結後才看到推薦清單。被動式是現在最常見的方式，也是效果最好的一種；原因在於它藉由使用者的操作作為推薦基礎，在行為上產生互動，使用者的接受程度較高，例如輸入關鍵字後再送出相應的推薦清單（見圖 19 黃框處，Customer with similar Searches Purchased）。

在 Schafer 的架構中可以發現，推薦的輸入是後續分析以及輸出的關鍵，具有連動的影響性。若需要作到長期性的個人化，便需要在顧客資料之間找出關聯、定義所要解決的商業問題並因應資料結構使用適合的推薦演算法（Recommendation Algorithm），輸入變數部份就必須包含歷史交易資料。也就因為從輸入到輸出整個過程之中，包含了商業面及技術面的考量，而使得推薦系統各不相同。



推薦系統的終極願景，在於每個使用者都能夠看到一份屬於自己偏好的網站內容。使用者不再需要自行一一挑選、過濾想看的內容。資訊沒有超載的問題，因為出現的都是根據偏好及互動而過濾出來。Amazon.com 的執行長 Jeff Bezos 說(Schafer 2001)：「如果我有三百萬個客戶，那麼我應該就有三百萬個網路商店」。

#### 四、協力式過濾與關聯規則探勘

協力式過濾 (Collaborative Filtering) 由 Goldberg 在 1992 提出(Goldberg et al. 1992)。它的原始想法很直觀，來自使用者決策之前常會徵詢前人的經驗，作為自己的參考。技術上是建立一個顧客—商品的矩陣 (Customer-Product Matrix)，矩陣中的值為使用者對商品的評分。透過觀察商品評分矩陣，可以得知使用者間的好惡相似程度，再以此結果和其他使用者對商品的評分推估使用者對未評分商品的評價。關聯規則探勘 (Association Rule Mining) 即是協力式過濾的一種應用。

##### (一) 關聯規則探勘

關聯規則常用來找出物品間的關係 (Item Association) 與使用者間的關係 (User Association) (Lin 2002)。物品間的關係係指商品之間可能存在某種關係，造成特定商品常常被一起購買。這類規則可以應用在若使用者購買了某種商品，系統即推薦其他也同樣常被一起購買的商品，此例在「個人化與傳遞」一節中已提到 (頁 31)，應用圖例請見圖 19。使用者間的關係係指找出具有相似購買習慣的使用者，而當要提供某使用者推薦時，由於歸屬相同類的群組具有相似購買習慣，便可一起送出推薦。例如某使用者群組常常購買某類書，若歸屬同類的使用者欲購買新書，便可推薦所屬群組中其他人購買過的書籍作為推薦。

關聯規則可以找出來的隱含資訊分為三大類(尹相志 2006)：

1. 有用的規則 (Useful)：指高品質的有效情報。通常不難驗證，而且可以用來設計成有意義的產品組合。例如買了 A 產品的了也會購買 B，那麼促銷時就可以合併兩種

產品為促銷組合，或是在購物動線中將兩種商店放在一起（讓顧客不會忘記購買）；或擺遠一點（既然兩種商品有高度關聯，那麼就算擺放的很遠，顧客也會購買，讓客戶多繞一點路還能增加其他商品的銷售機會）。

2. 常識 (Trivial)：指該領域專家之既知事實。此類規則由於早有因果關係可循，雖然它是正確的、而且也有充份資料佐證，但仍是無用的。例如顧客會同時購買電鑽與螺絲、油漆與油漆刷等等。然而，有時候一些規則可能是反映了過去某時刻行銷活動的成效或是佣金制度的成果。例如電腦和線上遊戲可能配套販售。因此在找出關聯規則後，必須去除過往受到行銷活動或佣金制度操控的部份。
3. 無法解釋的結果 (Inexplicable)：結果看起來沒有合理解釋，而且難以運用在決策上。例如精油蠟燭與烤雞總是一起賣出，但生鮮食品似乎較難與家用品放在一起陳列。

那麼，要如何判斷哪些規則是可用的？尹相志提到可以從兩個指標著手(尹相志 2006)：信心水準 (Confidence) 及支援度 (Support)。信心水準係指該規則的準確度，信心水準愈高其參考價值也愈高。支援度係指符合規則的交易次數，舉例而言若某一規則具有高信心水準，但一年發生的次數極低，那麼規則就不一定可用。在實務上，會先訂定規則的最小信心水準及最小支援度門檻先去除不合格的規則。

由於關聯規則探勘的來源是龐大的交易資料庫，鉅量的資料量造成演算效率低落，後續的演算法多著重在效率的改進上。其中，Apriori 演算法(Agrawal and Srikant 1994)是目前最普遍運用的關聯規則演算法（見表 3），也是 Microsoft SQL Server 2005 內建的演算法之一，後續的演算法也多以 Apriori 為基礎改良。

表 3 Apriori 演算法。資料來源：(Agrawal and Srikant 1994)

```
L1 = {large 1-itemsets};
for (k=2;Lk-1≠0;k++) do
begin
  Ck=apriori-gen(Lk-1);
  for all transaction tID do
  begin
    Ct = subset(Ck,t);
    for all candidates cICt do
      c.count++;
    end
    Lk={cICk|c.count≥minsup}
  end
end
Answer = ∪kLk;
```

首先 Apriori 會掃描過全部交易記錄，產生一個物件數的所有物件組合 (Item Sets)。若有一萬種商品，每個商品的物件組合便會有一萬組，數量極為龐大，因此需要設定最小支援 (Support) 參數，意即一個規則出現在總交易中的頻率或是百分比；若物件組本身都達不到最小支援的門檻，那麼包含這個物件組的規則就一定達不到門檻。除去未達最小支援的物件組之後，剩下的物件組合即為候選大型物件組 (Candidate Large Item Set)，即是一般所稱的「候選項目集」(Candidate Item Set)。

接下來進一步產生兩個物件數的物件組合，而且其中兩個物件都必須符合最小支援門檻，故實際的最大組合數會從  $\binom{10000}{2}$ ，也就是  $\frac{10000 \times 9999}{2}$  降為約五十萬種組合，即  $\binom{10000}{2}$ 。

繼續將此五十萬種組合根據最小支援度刪去，反覆以上步驟直到產生出包含「最大規則物件數」+1 個物件的物件組為止。例如若設定規則物件數最多為 3，那麼就要一直迴圈產生物件數為 4 的物件組為止。

最後，根據產生的物件組頻率相除產生關聯規則，再去除未達到最小信心水準的規則(尹相志 2006)。

## (二) Microsoft SQL Server 2005 的資料探勘解決方案

在資料探勘市場，Microsoft SQL Server 2005（以下簡稱 SQL SERVER 2005）算是起步較晚，一直到 1998 年推出的 SQL Server 7.0 版才納入 OLAP 模組，但後續請來資料探勘之父 Usama M. Fayyad 欲急起直追。

對微軟來說，資料探勘是一種特殊的資料彙總形式（Aggregate），將原始資料彙總成為特殊的規則與知識，其規格內容也以特殊的表格形式儲存。尹相志認為(尹相志 2006)，SQL SERVER 2005 的資料探勘即是：

1. 建立資料探勘模型結構：產生彙總表格定義。
2. 訓練資料探勘模型：過程即是將歷史資料插入（Insert）至彙總資料表中。
3. 資料探勘演算法：特殊的彙總函數。
4. 產生預測結果：將輸入資料表與彙總資料表連結（Join）起來。
5. 後續處理：允許使用者查詢、新增及刪除，並可利用 XML 交換資料。

以關聯規則探勘為例，SQL SERVER 2005 提供了互動式的圖形化介面來呈現探勘結果，比起數據清單更能提供給決策人員更直觀的參考。如圖 20 所示，綠色節點表示目前選取的節點，藍色節點表示目前點選節點（綠色）預測出的結果，當目前節點往外連結愈多預測節點（藍色）表示此商品愈能夠促銷其他商品。

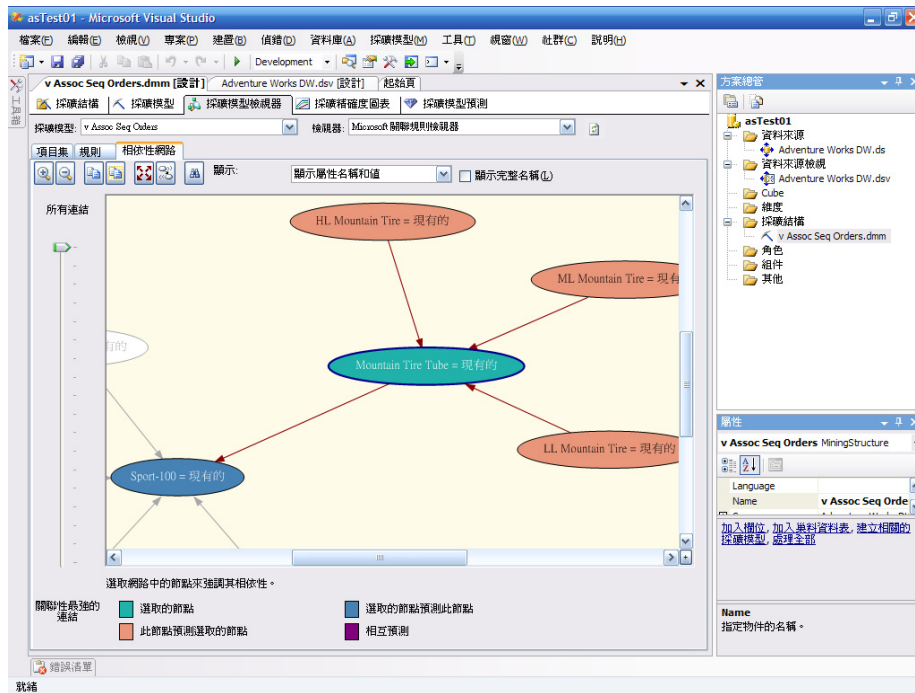


圖 20 Microsoft SQL Server 2005 的圖形化探勘結果示例。資料來源：本論文



## 肆、系統實作

在本章中會介紹本論文所提出的「A.bu, 阿布個人化推薦系統」(以下簡稱 Abu) 於交通大學圖書館中建置的技術細節，並說明讀者群的回饋及個別讀者對於書籍的難易回饋如何運用在過濾書單中，達成 Web 2.0 應用所能期待的典型效果。

第一節為系統架構簡介，並說明必要的建置環境與需注意之處；第二節介紹資料的預處理與異質資料的轉換；第三節說明如何應用 SQL SERVER 2005 的關聯規則探勘技術於系統中，以及使用 Data Mining Extension (DMX) 預測查詢的過程；第四節說明本系統排序個人書單的方法；最後說明本系統希望如何運用 Web 2.0 的群體評價方式，界定書籍的難易程度並反應到個人的推薦書單中。

### 一、系統架構

Abu 的名稱起源自西方諺語——女人衣櫃裡總少一件衣服——研究者認為讀者的書架上也永遠少一本好書。「一本」好書可以是引領入門的啟蒙、或是解決關鍵問題的經典好書；「一本」的概念是開始，因為閱讀與學習永無止境；因此研究者為此系統定名為 A.bu，取其 A book 的意象，希望能夠讀者找到適合的「一本」書。

Abu 的設計是希望除了個人化推薦功能之外，能夠另外加入對於推薦的回饋機制，並將使用者回饋反應到個人書單中，讓使用者感受到其回饋機制是具有效果的。個人化推薦因涉及大量的資料運算，故設計為離線處理；個人書單則為線上即時處理。整個系統的架構以圖 21 表示。

Abu 的使用者可以分為認證過的交大讀者與一般的使用者，前者為下圖的 Individual reader (個別讀者)、後者為 Readers (讀者群)。如此區分的目的在於，雖然 Abu 系統是設計給交大讀者使用的，但由於 Web 2.0 活動需要大量的群眾基礎，研究者希望有更多讀者可以貢獻書籍評價，因此仍保留了非交大讀者可以註冊並使用 Abu 的功能；唯一的差別在於，非交大讀者因於系統中無歷史借閱資料，故無法提供推薦書單。

個別讀者註冊登入後，若成功使用交通大學圖書館之讀者證號及密碼認證後，就可以使用個人化書單及查閱借閱歷史，由此兩種書單作為評價的來源。讀者群們可經由查詢交大館藏、Amazon.com 資料庫、交大分類書目及首頁的標籤雲 (tag cloud) 進行評價。

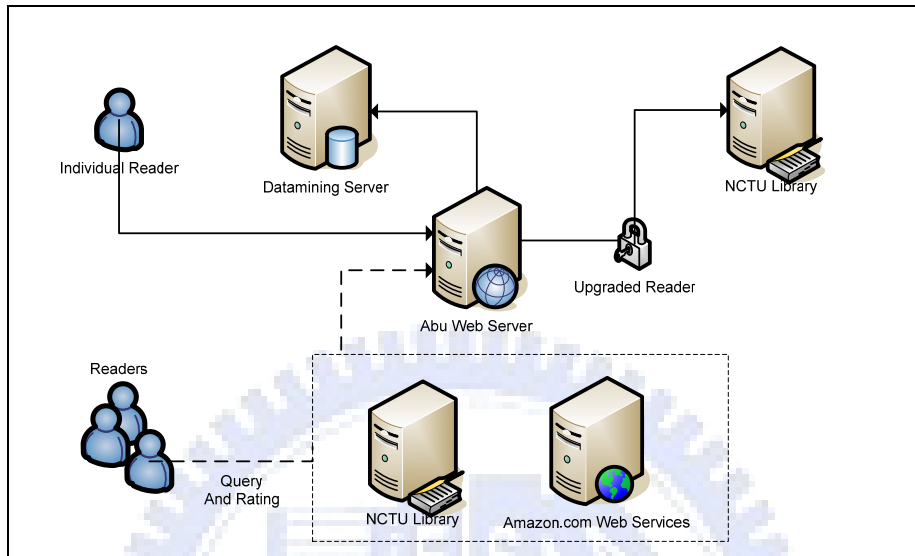


圖 21 Abu 架構圖。資料來源：本論文

Abu 的功能架構如圖 22 所示，主要的功能集中在「2 你的阿布」項目下。

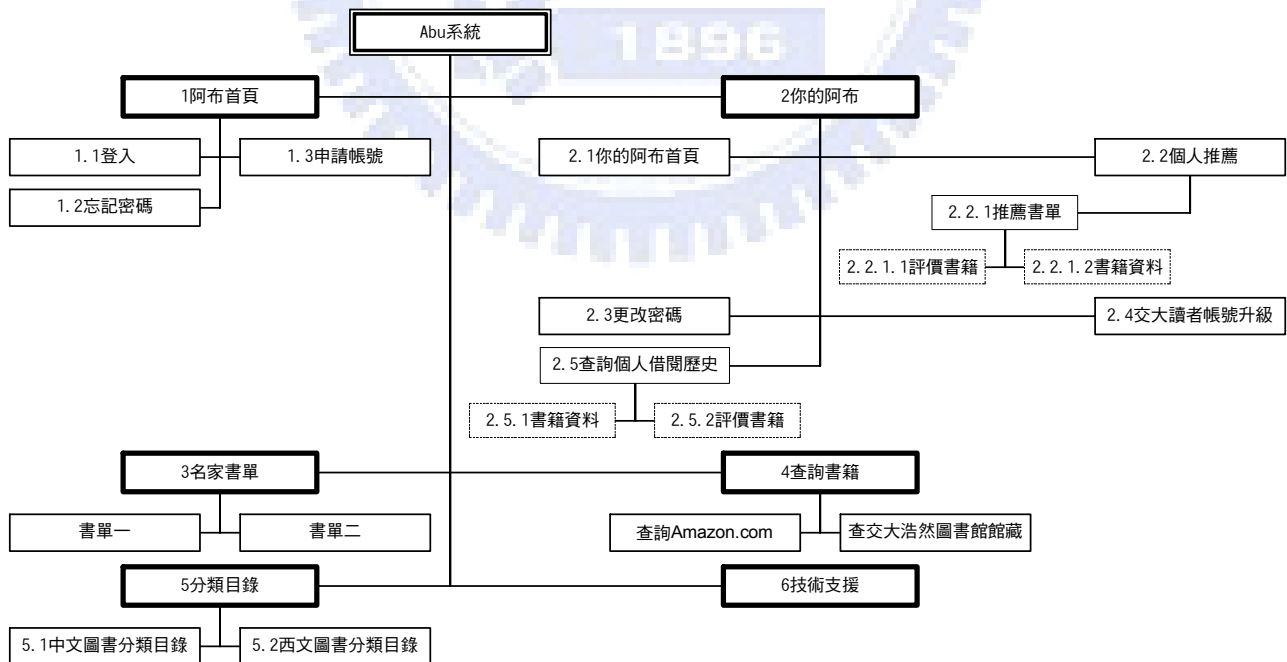


圖 22 Abu 功能架構圖。資料來源：本論文

Abu 系統使用的必要軟體需求如表 4。其中，由 Abu 連結至圖書館 Sybase 資料庫可透過 ODBC 或 OLEDB 兩種方式，實務上發現 OLEDB 在透過 Linked Server 以 openquery、openrowset 查詢 Sybase 資料庫時，無法處理 Sybase 內中文字碼的問題；而透過 ODBC 連結之效能也比 OLEDB 高出許多。因此研究者建議在表 4 的相似環境中採用 ODBC 連結。

表 4 Abu 必要軟體需求。資料來源：本論文

功能	軟體名稱
Web Server	Microsoft IIS 6
	Microsoft Windows 2003 Standard SP1
Development Tools	ASP.NET 2.0
	Visual Studio 2005 Professional
	.NET Framework 2.0.50727
Database Server	Microsoft SQL Server 2005 Enterprise SP1
Data Access Component	MDAC 2.8 SP2
Web Services	Amazon.com Web Services
ODBC Driver	Sybase PC Client for ASE 11
	Microsoft ODBC.NET

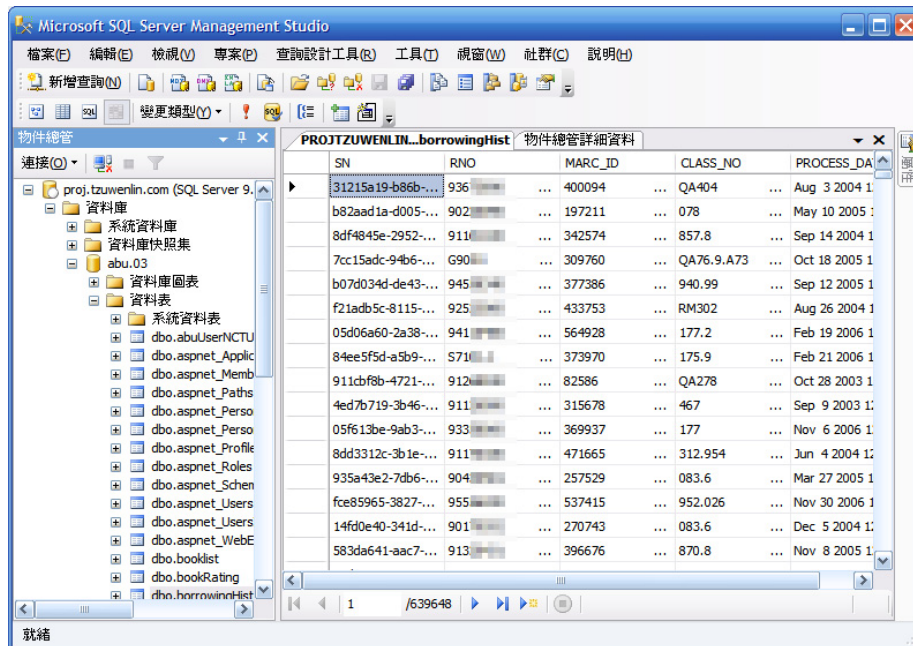
## 二、資料預處理

本論文採用 SQL SERVER 2005 的關聯規則解決方案，在關聯規則設計模型中需將資料來源設計為「案例資料表」(Case Table) 與「巢狀資料表」(Nested Table) 兩類。以本論文而言，分析目標為讀者所借閱過的書籍，案例資料表為讀者主檔 (以讀者編號為主鍵)，而借閱明細為巢狀資料表。

研究者將此兩類的資料自交大圖書館 T2 系統中匯出借閱明細及讀者主檔至 SQL Server 2005 中。借閱歷史以有效 (active) 讀者之活動年份往前推四年，因此借閱歷史自 2003 年至 2007 年共 63 萬餘筆。建立資料探勘模型後，可發現以書籍登錄號為鍵值之關聯探勘未能產生關聯規則，原因在於圖書館相同一本館藏的數量通常不多、有限的數量借出後就不能再給其他讀者外借，多位讀者借閱相同書籍的記錄不多。面對這個問題，(余明哲 2003)提出了以



圖書類別替代相同書作為推薦單位的作法，本研究續用其方法。



SN	RNO	MARC_ID	CLASS_NO	PROCESS_DA
31215a19-b86b-...	936	400094	QA404	Aug 3 2004 1
b82aad1a-d005-...	902	197211	078	May 10 2005 :
8df4845e-2952-...	911	342574	857.8	Sep 14 2004 1
7cc15adc-94b6-...	G90	309760	QA76.9.A73	Oct 18 2005 1
b07d034d-de43-...	945	377386	940.99	Sep 12 2005 1
f21adb5c-8115-...	925	433753	RM302	Aug 26 2004 1
05d06a60-2a38-...	941	564928	177.2	Feb 19 2006 1
84ee5f5d-a5b9-...	S71	373970	175.9	Feb 21 2006 1
911cbf8b-4721-...	912	82586	QA278	Oct 28 2003 1
4ed7b719-3b46-...	911	315678	467	Sep 9 2003 1:
05f613be-9ab3-...	933	369937	177	Nov 6 2006 1
8dd312c-3b1e-...	911	471665	312.954	Jun 4 2004 1:
935a43e2-7db6-...	904	257529	083.6	Mar 27 2005 1
fce85965-3827-...	955	537415	952.026	Nov 30 2006 1
14fd0e40-341d-...	901	270743	083.6	Dec 5 2004 1:
583da641-aac7-...	913	396676	870.8	Nov 8 2005 1

圖 23 Abu 借閱資料。資料來源：本論文

以圖 23 為例，CLASS\_NO 即為該本書的所屬類號。交大圖書館的中文圖書採用中國圖書分類法，西文圖書則採用美國國會圖書館分類法。以圖書類別為推薦單位之下，若讀者借閱相同類別的書籍，將之視為借閱同一本書。

### 三、建立關聯規則模型及演算法預測查詢

在關聯規則模型成功建立並部署後，可以在「採礦模型檢視器」瀏覽資料探勘結果。如果分析的是大型資料庫之交易資料，因為商品數量極為巨大，所產生之規則支援會小至找不出符合最小支援門檻之規則。在建立資料探勘模型時可以先將最小支援度設小以便找到規則，再逐步調高最小支援門檻或最小規則信心水準來排除較不重要的規則，如圖 24 所示。或是直接修改探勘模型的構成，如本論文採用圖書類別作為巢狀資料表之主鍵。

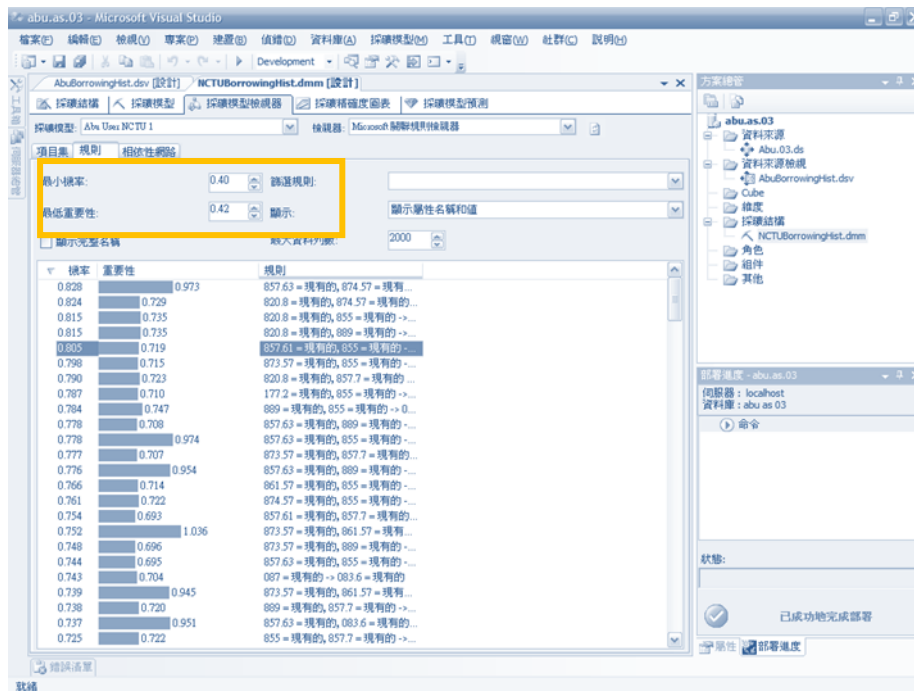


圖 24 Abu 關聯規則模型。資料來源：本論文

SQL SERVER 2005 提供數項參數供設計者調整資料探勘模型，設計者必須仔細調整這些參數以獲得較好的探勘結果。

參數 MINIMUM\_SUPPORT 表示最小支援門檻，是表示物件集出現頻率的指標，若物件組低於此參數將會被刪除。值得注意的是，若系統記憶體不足，演算法可能會動態調整此參數以降低系統負擔。若資料探勘結果不如預期可試著檢視此參數是否已被更動。

參數 MINIMUM\_PROBABILITY 表示最小信心水準門檻（即圖 24 中之「最小機率」），輸入值範圍在 0~1 之間，信心水準低於此參數則該規則會被刪除。指定最小機率值便能限制形成的規則數量。借了 A 後會借 B 的機率公式表示如下：

$$\text{Probability}(A \Rightarrow B) = p(B | A) = \frac{\text{Support}(A, B)}{\text{Support}(A)}$$

方程式 1 Probability 公式示例。資料來源：本論文

參數 MINIMUM\_IMPORTANCE 表示規則最小重要性，當規則重要性低於此參數將會被刪除。

機率高不一定表示是有意義的規則。例如「借了 A 則有 80% 的機率借 B」這條規則，如果進圖書館的任一讀者借 B 的機會不大、只有 25%，那麼這條規則就有意義。如果隨意一位讀者本來借 B 的機率就高達 90%，那麼這條規則就屬於必然的組合，意義不大了。因此，必須要比較在有 A 以及沒有 A 的條件下，發生 B 事件的機率比例，公式如下：

$$\text{Importance}(A \Rightarrow B) = \log\left(\frac{p(B|A)}{p(B|notA)}\right)$$

方程式 2 Importance 公式示例。資料來源：本論文

公式中分子、分母的比例可以相差很大，故開對數取得 Importance 指數。當「在 B 的條件下發生 A 的機率」高於「在無 B 的條件下發生 A 的機率」則對數後會大於零。Importance 愈大，表示規則愈顯著。

SQL SERVER 2005 中提供一項圖形化檢視的工具，稱為「相依性網路」。以本論文為例，點選分類號 177（綠色圓環）可以發現另外指向兩個藍色圓環，分別為分類號 083.6 及 855。當點選的圓環能對外指向愈多藍色圓環，表示此項分類號愈能帶動其他分類號的借閱行為。對照中圖書館分類法可以發現，「應用心理」（177）可以促進「現代叢書」（分類號 083.6）與「中國散文·雜著」（分類號 855）的借閱。

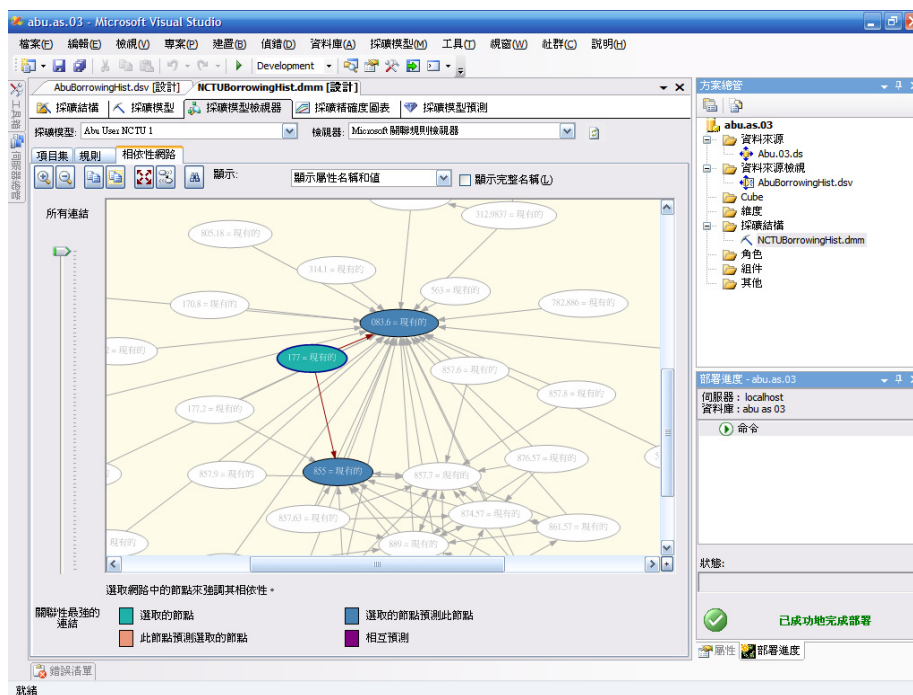


圖 25 Abu 探勘結果之相依性網路。資料來源：本論文

然而相依性網路並無法應用在 Web 化的環境中，研究者需要知道的是在探勘模型中，輸入使用者的變數、再得出推薦的結果。因此，在 SQL Server 2005 的解決方案中，便需要使用到資料探勘延伸語言 (Data Mining Extension, MDX)。

在「Microsoft SQL Server 2005 的資料探勘解決方案」一節中提到 (頁 36)，Microsoft 的資料探勘模式是一種特殊型態的資料表，整個模型從建置、訓練到預測都可以使用 DMX 來完成。而具有產生預測結果功能性的則稱為 Prediction Query。使用 Prediction Query 的優點為，只要以近似 T-SQL 的 JOIN 語法連結欲預測資料與資料探勘模型便較容易得出預測結果，並順利應用到於各式應用系統中。

以本系統而言，若想要作到如 Amazon.com 的自動商品推薦，則需要參考已有的資料探勘模型，並輸入個別讀者的借閱項目以取得推薦結果，需要用到的函數即為 PredictionAssociation。

PredictionAssociation 是一個「巢狀表格函數」(Nested Table Function)，它會讀取「案例資料表」(Case Table) 中該讀者已借閱的圖書類別去比對所有關聯規則之推薦項目，同時去

除該讀者已借閱過的類別而產生推薦清單。若無符合的規則，PredictionAssociation 則會以訓練組資料的總體平均值帶入。在圖 25 看到可能有多個綠色圓環指向同一藍色圓環表示有多條規則推薦同一類別，若有此一情況，則以機率較高者為準。

例如，隨意輸入五項某讀者曾借閱過之圖書類別，分別為 019.9、581.31、582.8、584.3 及 584.94 欲預測其推薦結果，其 MDX 內容如表 5 所示：

表 5 PredictionAssociation 預測查詢示例。資料來源：本論文

```
SELECT
  [Abu User NCTU 1].[Borrowing Hist],
  PredictAssociation([Abu User NCTU 1].[Borrowing Hist], INCLUDE_STATISTICS, 5)
From
  [Abu User NCTU 1]
NATURAL PREDICTION JOIN
(SELECT (SELECT '019.9' AS [CLASS NO]
  UNION SELECT '581.31' AS [CLASS NO]
  UNION SELECT '582.8' AS [CLASS NO]
  UNION SELECT '584.3' AS [CLASS NO]
  UNION SELECT '584.94' AS [CLASS NO]) AS [Borrowing Hist]) AS t
```

其中，引數 INCLUDE\_STATISTICS 表示預測結果會在傳回的巢狀表格中加入 \$Support、\$Probability 以及 \$AdjustedProbability 三樣統計資訊。數字 5 表示過濾機率最高的預測物件數量為 5，所以最多只會傳回五筆結果。

傳回的預測查詢結果如圖 26 所示。根據輸入的引數所預測的類別依機率大小排序依序為 083.6、857.7、855、889 及 312.932C。

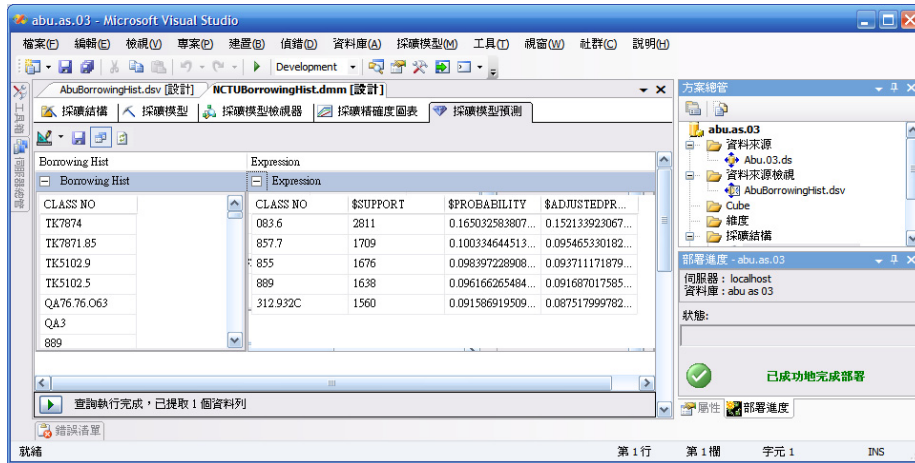


圖 26 PredictionAssociation 預測查詢結果。資料來源：本論文

然而，此類預測查詢若等待到讀者線上要求時才處理效率較差，在實務上 Abu 以窮舉法將所有讀者的預測結果預先批次處理完成並存入資料庫中（如圖 27），便可以在 Web 上以較快速度呈現推薦結果（如圖 28）。

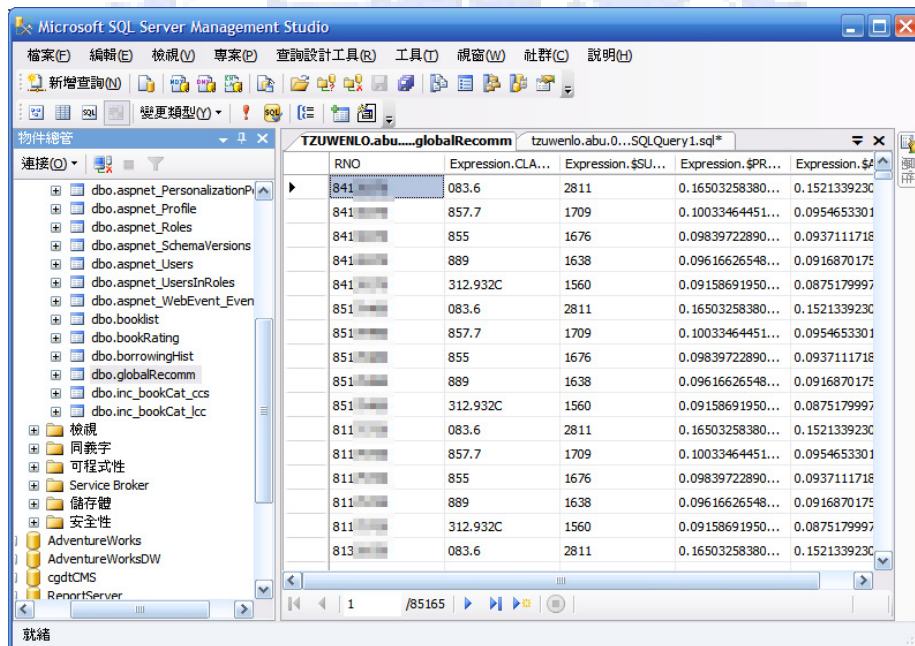


圖 27 預先批次處理完所有讀者之預測查詢。資料來源：本論文

圖 28 為實際之 Abu 個人推薦畫面，呈現時以 \$AdjustedProbability（調整後機率）以大到小排序。



圖 28 Abu 網站之個人推薦畫面。資料來源：本論文之 Abu 網站

#### 四、排序個人書單及難易層級

傳統的推薦系統於此一階段通常採用群集演算法，以 K-means 或 E-M 計算圖書（或圖書類別）與讀者的向量式，並以此建立書籍與讀者之間的關係。

本論文在此一階段，使用的是讀者自行評價的方法找出圖書與讀者之間的關係。本論文將書籍分為四種難度，分別為專業程度、玩家程度、進階程度及入門程度，其分數的計算公式如下：

$$AvgBookRating = \left( \frac{\text{所得分數之總和}}{\text{所得之評價次數}} \right), \text{ where 單次評價分數} = \begin{cases} 1, if = \text{入門程度} \\ 2, if = \text{進階程度} \\ 3, if = \text{玩家程度} \\ 4, if = \text{專家程度} \end{cases}$$

方程式 3 書籍評價計算公式。資料來源：本論文

若某本書籍所受評價次數愈多，其得分會愈接近讀者心中的真實難度。圖 29 是 Abu 系統中顯示書籍評價的畫面。



圖 29 Abu 網站中書籍評價畫面。資料來源：本論文 Abu 網站

讀者也依照自身對書籍的評價活動，計算其自己的閱讀階段，亦分為四級，依序為專業程度、玩家程度、進階程度及入門程度。算式如下：

$$IndividualReaderRating = \left( \frac{\text{所評分數之總合}}{\text{所評之評價次數}} \right), \text{ where 單次評價分數} = \begin{cases} 1, if = \text{入門程度} \\ 2, if = \text{進階程度} \\ 3, if = \text{玩家程度} \\ 4, if = \text{專家程度} \end{cases}$$

方程式 4 讀者閱讀階段計算公式。資料來源：本論文

讀者登入後在「你的阿布」頁面上可獲知目前的閱讀階段，如圖 30。





圖 30 Abu 網站中讀者之閱讀階段畫面。資料來源：本論文 Abu 網站

在讀者點選「個人推薦」頁面任一推薦書籍類別時，便會依照所有讀者對該類書籍的總評分推薦數量，計算出每本書的累積推薦數（星級）及平均推薦數（星級），再從中過濾掉與讀者閱讀階段不合的書籍，才呈現出最終之推薦書單，如圖 32。

其過濾法則如下所示：

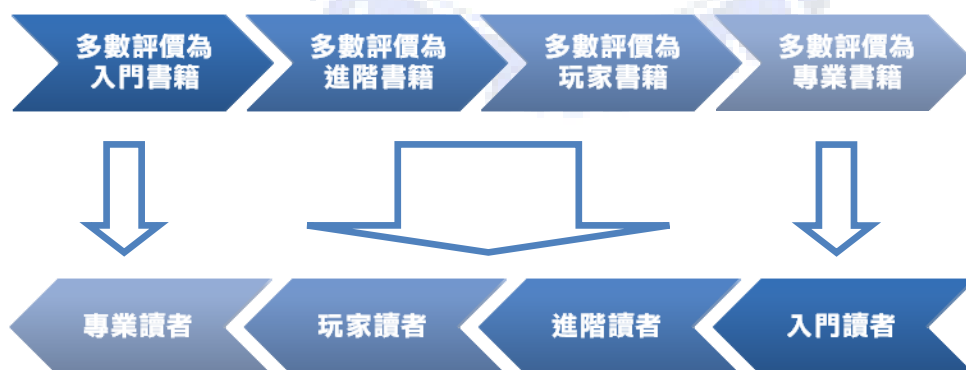


圖 31 Abu 過濾法則。資料來源：本論文

若讀者將大部份書籍都評為入門書籍，表示自己的閱讀階段已屬專業階段，在書單部份將只出現被讀者評為專業程度的書籍；反之，若多數都評價為專業書籍，則將讀者歸為入門

讀者，在書單部份將只出現被讀者評為入門程度的書籍。玩家及進階讀者因屬性較相近、也為增加閱讀的可能性，因此目前將之視為同一類別，並只推薦同一類的書籍。



圖 32 阿布的個人化書單。資料來源：本論文 Abu 網站

在初期評價數量不多時，個人書單呈現的結果可以預期將較單一。隨著評價資料的累積及讀者自身以評價調整閱讀階段，個人書單的結果將漸趨多元且一致。

以 Web 2.0 精神觀察此份書單，可以發現書籍的排序及可通過過濾的書籍，完全是透過社群讀者所決定的。若讀者發現有某本書根本不屬於某難易程度、尤其是該書評價資料尚不多時，也可以透過多多地重新評價，讓系統重估其難易層級。

## 五、系統實際操作畫面

Abu 中建有中、英文圖書分類目錄（見圖 33），提供 taxonomy 式的瀏覽方式。此種分類方式不易突顯出社群目前熱中的議題，有些書目更是可能同時存在於多種類別中，這是傳統式分類法不能處理的問題。



圖 33 圖書分類目錄。資料來源：本論文 Abu 網站

Abu 累積的標記資料會以標籤雲（tag cloud）的方式呈現在首頁，讀者可一目了然獲知社群中目前最熱門的閱讀趨勢，藉以作為閱讀的參考。圖 34 是 Abu 在首頁中的標籤雲。



圖 34 Abu 的 tag cloud。資料來源：本論文 A

為了使評價書籍的來源更多元，讀者可在「查詢書籍」中透過 Amazon.com Web Services 查詢 Amazon.com 的書目資料，並連結到圖書館 T2 系統查詢是否有館藏，若有館藏則可繼續評價活動。圖 35 即是自 Abu 中查詢 Amazon.com 書目的畫面。



圖 35 在 Abu 查詢 Amazon.com 書目。資料來源：本論文 Abu 網站

另外，為發揮館員推薦及專家推薦的角色，Abu 另外建有「名家書單」的功能。



圖 36 Abu 名家書單。資料來源：本論文 Abu 網站

名家書單功能可補足 Abu 中專家推薦的角色（見圖 36），為了日後維護方便，此部份名家書單以 XML 格式處理，可交由館員日後繼續新增書單，不需直接接觸資料庫。

## 伍、資料分析與討論

本章就交通大學圖書館 Abu 系統之使用者滿意度調查結果，進行統計分析。第一節為受訪者使用網路之行為分析；第二節為受訪者平時使用交通大學浩然圖書館之行為分析；第三節為受訪者對於 Web 2.0 相關功能的需求分析；第四節為使用者使用過 Abu 系統後、對於 Abu 達成 Web 2.0 功能的使用滿意度分析。第五節檢視受訪者的網路使用資歷、平時使用圖書館的行為、對 Web 2.0 的功能需求與最後的滿意度之間，是否隨著涉入程度而造成差異，並綜合討論以上的分析內容。

本問卷調查範圍包括交大、台大、清大及成大學生，其中以交通大學學生（包括大學部、碩、博班）為大宗，共蒐集 268 份問卷。

### 一、受訪者網路使用行為及接觸圖書資訊行為分析

#### 1. 每天上網時間

過半（53.90%）受訪者每天使用網路的時間為三至五小時，二至三小時占四成多。每天使用網路二小時以上的受訪者占九成以上，如圖 37 所示。

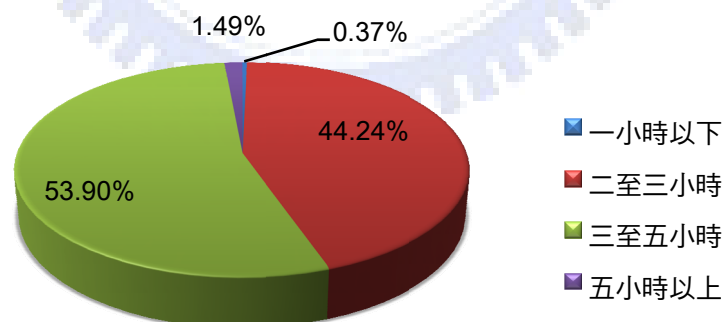


圖 37 受訪者每天上網時間

## 2. 上網資歷

由圖 38 可以發現受訪者之上網資歷有四分之三強具三年至五年的上網經歷，五至七年的使用者更占有二成左右，一至三年上網資歷的讀者非常少數(1.87%)，一年以下完全沒有，詳情如圖 38 所示。

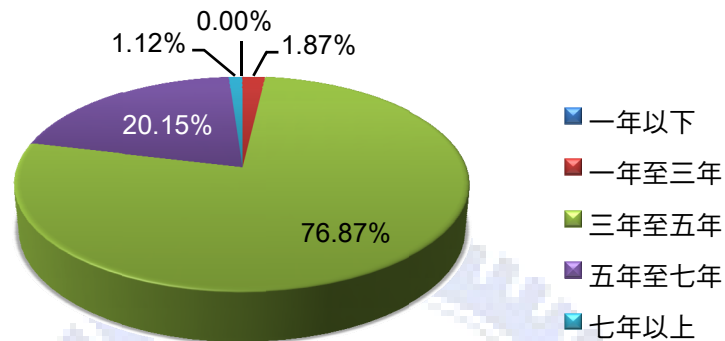


圖 38 受訪者上網資歷

## 3. 逛書店或接觸書籍的頻率

本題主要想了解圖書館的讀者是否會主動接觸圖書資訊。圖 39 表示有八成左右(82.22%)每週皆會逛書店或接觸書籍，每月逛書店的讀者仍有一成。幾乎每天逛書店或接觸書籍的不到一個百分點。可見受訪讀者仍是屬於會主動閱讀的族群。

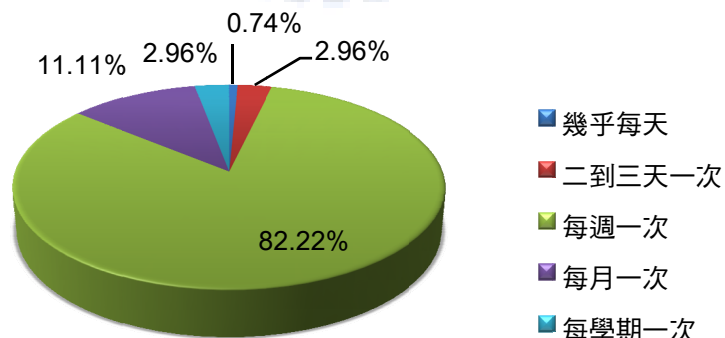


圖 39 受訪者逛書店或接觸書籍的頻率

#### 4. 閱讀行為中讀新書的比例

圖 40 顯示約五成（51.49%）的讀者在平時的閱讀行為中是閱讀新書，顯見新書對於讀者的吸引力仍大。接近三成（29.48%）的讀者有一至三成的比例是在閱讀新書，約一成五（14.55%）的讀者嗜讀新書，平時的閱讀中有五至七成都在讀新書；更有 1.87% 左右的讀者讀的書九成以上皆是新書。若圖書館的讀者喜好新書更甚於其他館藏，圖書館如何更好的推薦新書、推廣更貼近閱讀趨勢的書籍，應更能提高圖書的利用率。

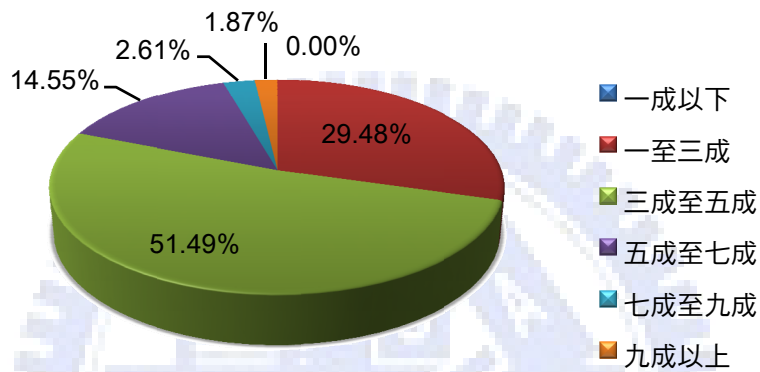


圖 40 受訪者閱讀新書的比例

#### 5. 閱讀行為中讀學校教科書、參考書的比例

相對於閱讀新書之外，受訪讀者又是如何處理課內的閱讀呢？圖 41 顯示有近八成（79.10%）的讀者在平時閱讀學校教科書、參考書的比例在一成以下。有約二成（20.52%）的讀者平常讀的書中有一至三成是學校教科書。受訪者中平時閱讀課內書籍占整體閱讀三成以上，調查結果為零。



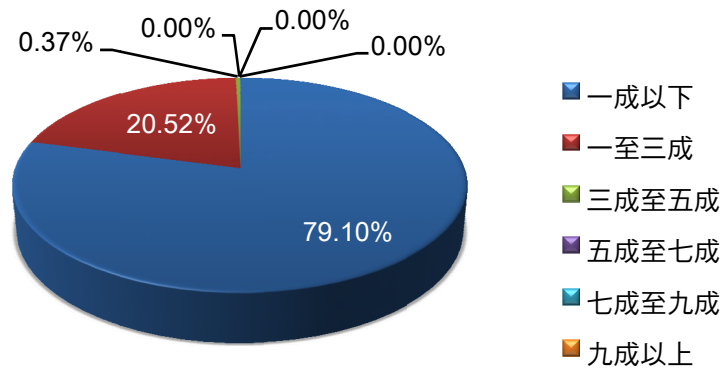


圖 41 受訪者閱讀學校教科書、參考書的比例

## 二、受訪者圖書館使用行為分析

### 1. 親自到訪交大圖書館的頻率

圖 42 顯示在受訪者中約有四分之三強（75.66%）的比例親自到訪交大圖書館的比例為每學期數次；約一成五（15.93%）的讀者每月都有數次親自到訪。每週到訪數次的讀者僅占 7.78%。

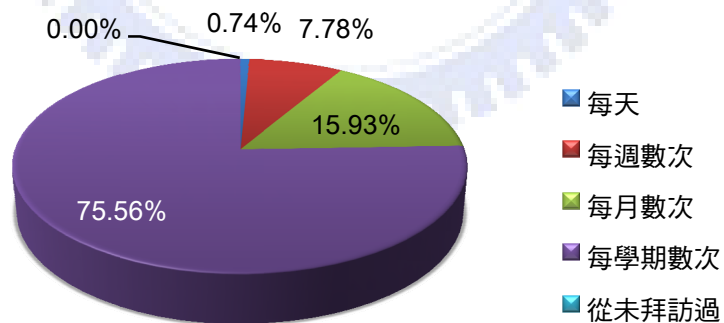


圖 42 受訪者親自到訪交大圖書館的頻率

### 2. 受訪者使用交大圖書館電子資源，包括網站、blog、電子資料庫的頻率

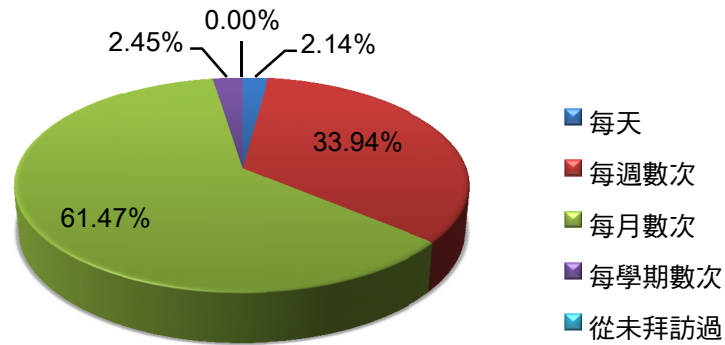


圖 43 受訪者使用交大圖書館電子資源的頻率

圖 43 顯示受訪者中有約六成（61.47%）的讀者每月有數次會使用交大圖書館所提供的電子資源，有三成多（33.94%）的讀者每週有數次會使用圖書館的電子資源。若再以圖 44 比較親自到訪與使用電子資源的比例，可以看出在使用電子資源項目中，每月使用數次的讀者為大宗；但親自到訪的讀者，每學期數次為大宗。

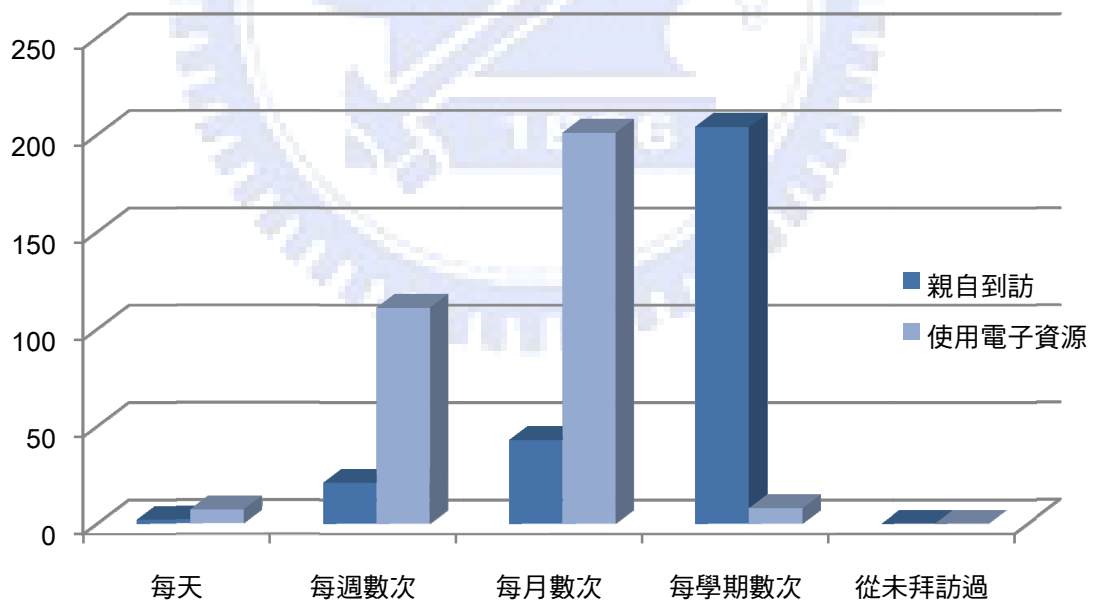


圖 44 受訪者親自到訪與使用電子資源之比較

### 3. 受訪者使用館藏目錄與新書目錄的頻率

圖 45 顯示受訪者使用館藏目錄與新書目錄的比例大致相同，以每月使用數次為大宗；

僅有少數讀者每週會查詢館藏及新書目錄。

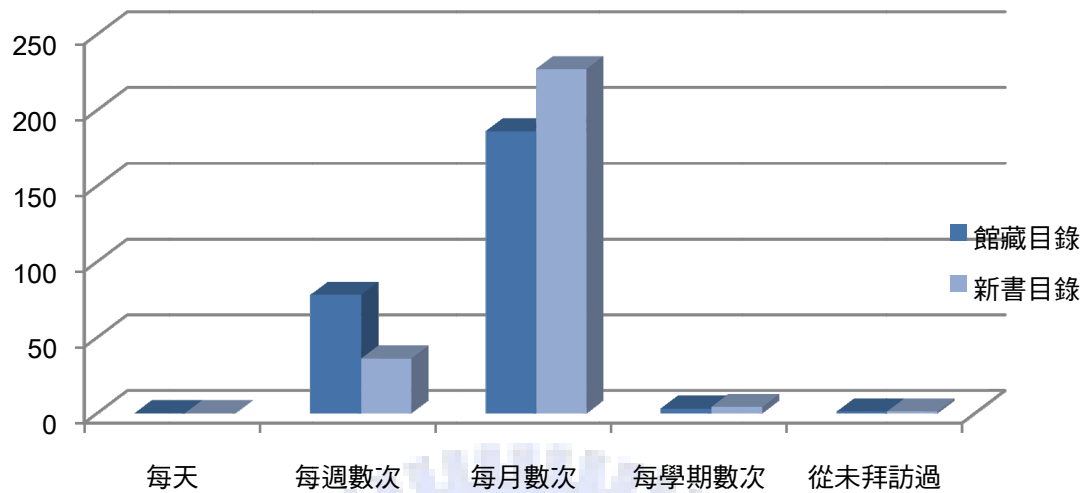
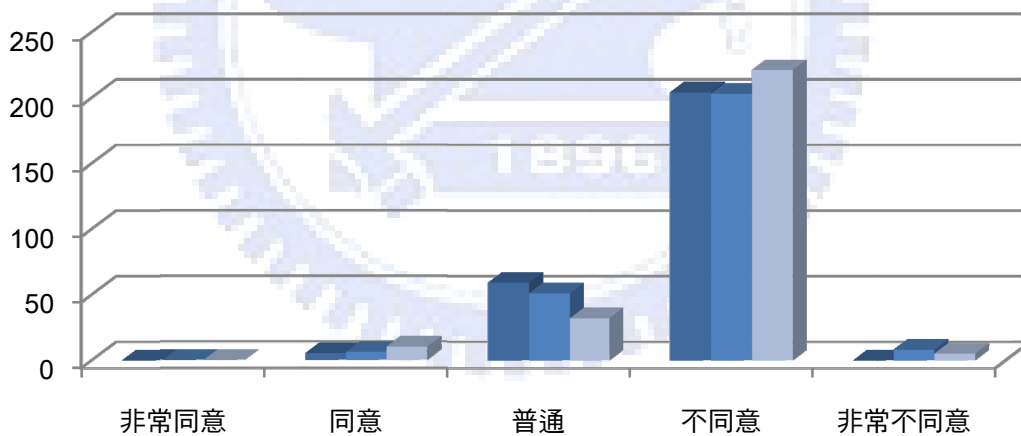


圖 45 受訪者使用館藏目錄與新書目錄之比例

#### 4. 受訪者對於圖書館電子資源查詢及排序結果滿意程度



- 以關鍵字查詢館藏資料庫時，覺得查詢的結果數量適中、沒有不相關的資訊出現
- 以關鍵字查詢館藏資料庫，可以解決找書及閱讀選書上的問題
- 以關鍵字查詢館藏資料庫後查詢結果的排序方式，可以解決找書及閱讀選書的問題

圖 46 受訪者對於圖書館電子資源查詢及排序的結果滿意程度

受訪者對於圖書館電子資源查詢及排序結果滿意程度為一題組，內有三道獨立問題，依序為「以關鍵字查詢館藏資料庫時，覺得查詢的結果數量適中、沒有不相關的資訊出現」、「以

關鍵字查詢館藏資料庫，可以解決找書及閱讀選書上的問題」以及「以關鍵字查詢館藏資料庫後查詢結果的排序方式，可以解決找書及閱讀選書上的問題」。圖 46 顯示大部份的受訪者皆對以上三道問題表示「不同意」。顯見在圖書館館藏資料愈見豐富的同时，如何提供一個有效、精確的查詢及排序方法，以幫助解決找書及選書的問題，成為愈來愈重要的課題。

### 5. 受訪者對於圖書館提供新知的滿意程度

受訪者對於圖書館提供新知的滿意程度為一題組，包含二道問題，依序為「新書目錄以圖書分類（總類、自然科學、哲學…等）的方式，可以幫助找到符合自己興趣和需求的新書資訊」及「目前圖書館電子資源或網站的設計，可以幫助你很快的了解目前學校或社會上的閱讀趨勢或熱門議題」。圖 47 顯示大部份的讀者對於圖書館電子資源及網站是否能提供閱讀上的新知及趨勢，表示了強烈的不同意。

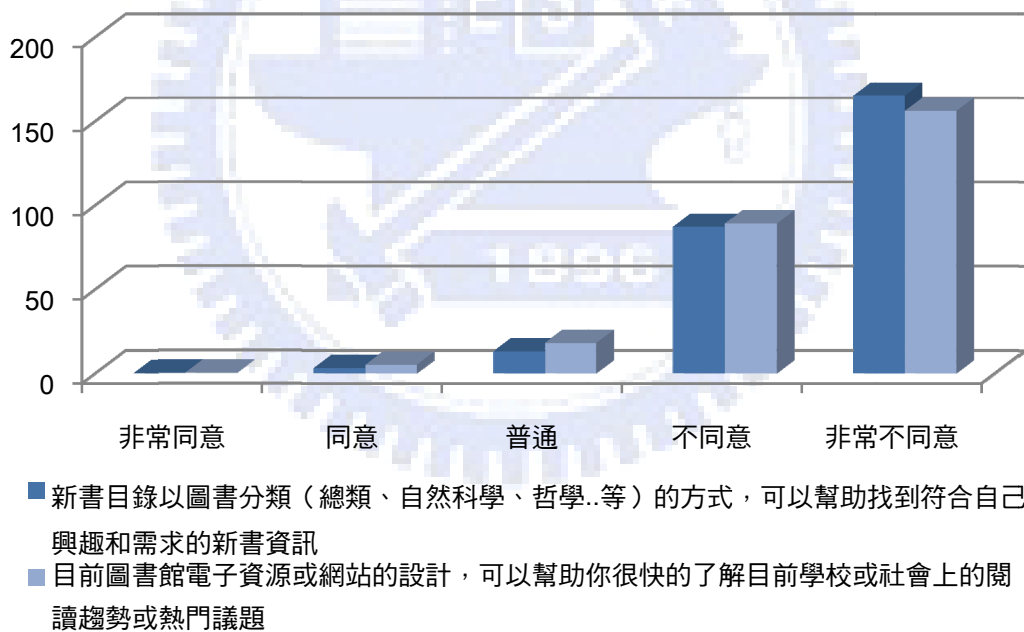


圖 47 受訪者對於圖書館提供新知的滿意程度

### 6. 接觸新書的管道

圖 48 顯示為了接觸新書資訊，受訪者近四成比例（38.29%）選擇瀏覽網路書店；有三成多（33.09%）的受訪者選擇實際到實體書店翻閱；近三成（27.46%）的受訪者會直接瀏覽

出版社的網站；以圖書館、報紙雜誌或親友介紹得知新書資訊的受訪者，合計不到二個百分比。

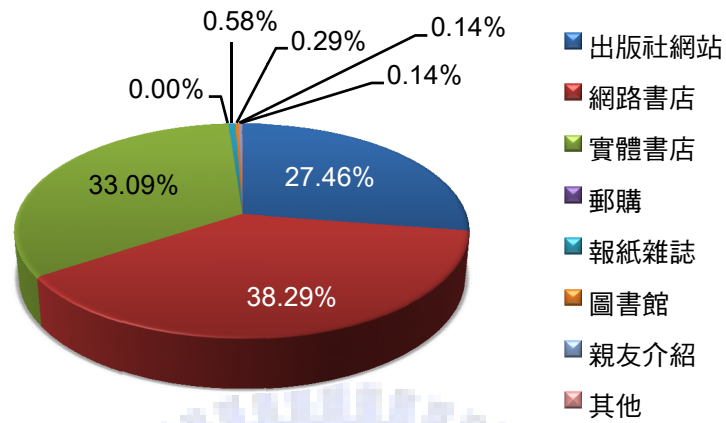


圖 48 受訪者接觸新書的管道

#### 7. 受訪者選書時會參考其他熱中閱讀讀者意見之比例

圖 49 顯示熱中閱讀讀者的意見，對於受訪者選書具有相當的影響力，有近四分之三受訪者表示選書時會參考其意見。書籍的口碑效果仍左右受訪者選書的決定。

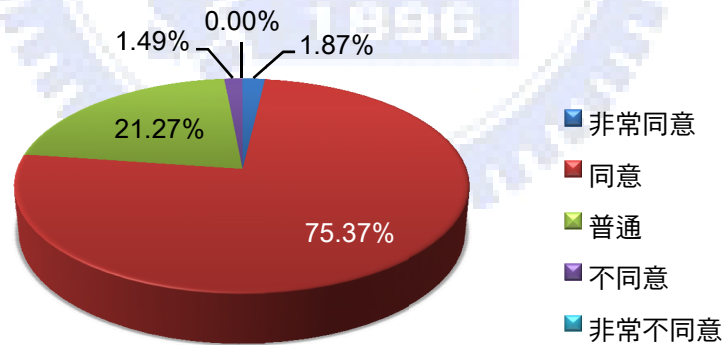
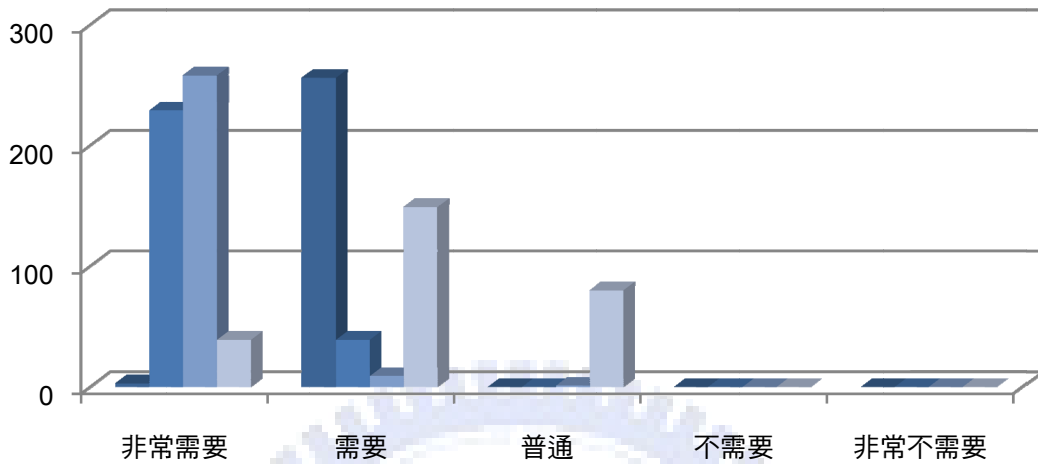


圖 49 受訪者選書時會參考其他熱中閱讀讀者意見之比例

### 三、Web 2.0 推薦功能需求分析



- 圖書館若提供Web 2.0的網站服務，你對個人化推薦的需求程度為何？
- 圖書館若提供Web 2.0的網站服務，你對Web 2.0式的圖書索引方式（例如標籤雲，tag cloud）的需求程度為何？
- 圖書館若提供Web 2.0的網站服務，你對依個人閱讀難易程度過濾個人推薦書單的需求程度為何？
- 圖書館若提供Web 2.0的網站服務，你對依大家的評價決定書籍的分群、以達到讓書籍的分類更貼近讀者需要，此項功能的需求程度為何？

圖 50 受訪者對於 Web 2.0 推薦功能的需求分析

圖 50 顯示受訪者對於圖書館欲提供 Web 2.0 之推薦服務都表達了高度的興趣，並對於即將提供的功能都表示了非常需要及需要。Web 2.0 常見的標籤雲，受試者對於採用在圖書館系上都表示了非常需要的意見；依個人閱讀難易程度過濾個人推薦書單的功能次之；個人化推薦再次之。

#### 四、Abu 使用滿意度分析

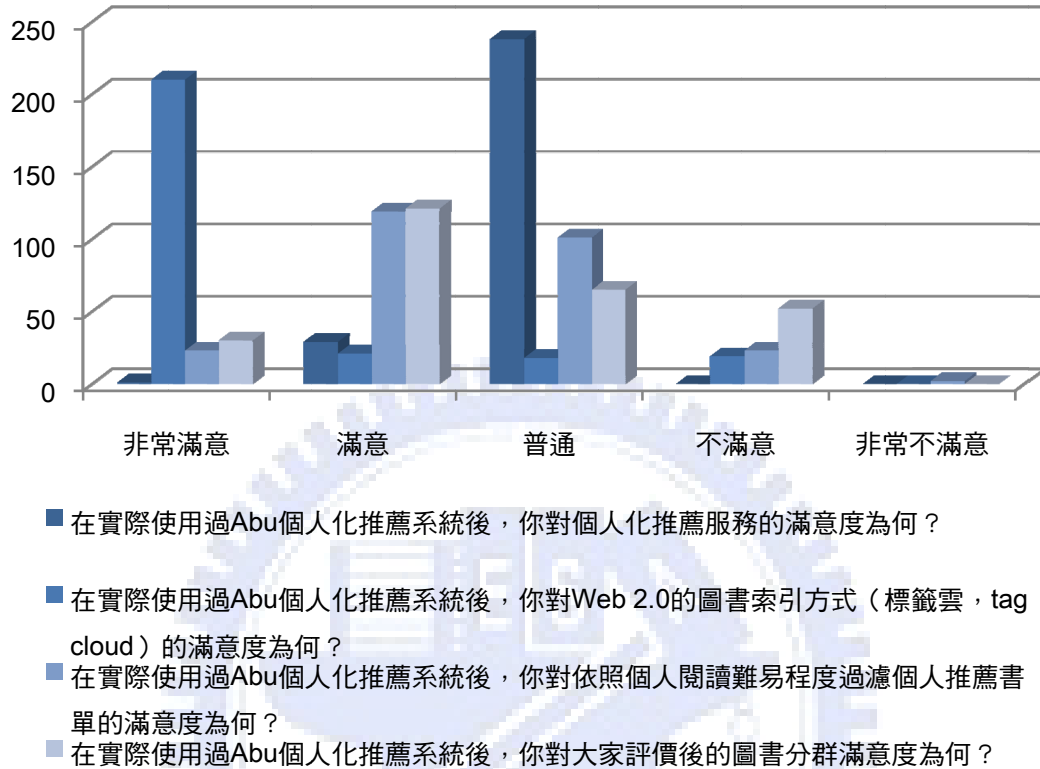


圖 51 受訪者對於 Abu 的滿意度分析

圖 51 顯示受訪者對於 Abu 系統的使用滿意度。在個人化推薦及依照個人閱讀難易程度過濾書單方面，大部份受訪者表示普通或滿意。由於推薦的功能僅推薦到類別，第二層的書單是依照社群使用者的評價而定，使用者可能對於不能直接顯示書單感到不滿意。另外，Abu 由於自上線測試後到開放問卷調查不到一個月時間，累積的書籍評價有限，依難易度過濾書單的功能未能充份發揮，隨著評價的累積這部份的問題應能逐漸舒解。

在 Web 2.0 的圖書索引方面，大部份的受訪者表示非常滿意。使用者對於標籤雲應用在圖書館網站應該是感到新奇，也從標籤雲點選中發現一些驚奇，故給予較高的肯定。

在以社群評價決定圖書分群方面，多數受訪者表示滿意、普通等意見。相信在逐漸累積圖書評價後此部份也能慢慢修正到符合使用者的需要。

表 6 顯示受試者對於 Abu 系統的滿意度排序，滿意度由高至低依序為「圖書索引方式」、「依難易度過濾個人書單」、「圖書分群」及「個人化推薦系統」。

表 6 受試者對於 Abu 功能的滿意度分析—平均數與標準差

滿意度	圖書索引方式	依難易度過濾個人書單	圖書分群	個人化推薦系統
排序	1	2	3	4
平均數	4.574627	3.514925	3.481343	3.115671642
標準差	0.898412	0.800141	0.929957	0.331911833

## 五、綜合討論

最後本研究請受試者綜合以上心得，回答綜合滿意度題組，包含「總合來說，Abu 系統的確能夠幫助你解決選書的問題」及「Abu 系統的確能夠幫助你了解現在（指圖書館）最新的閱讀趨勢及熱門議題」二道題目。

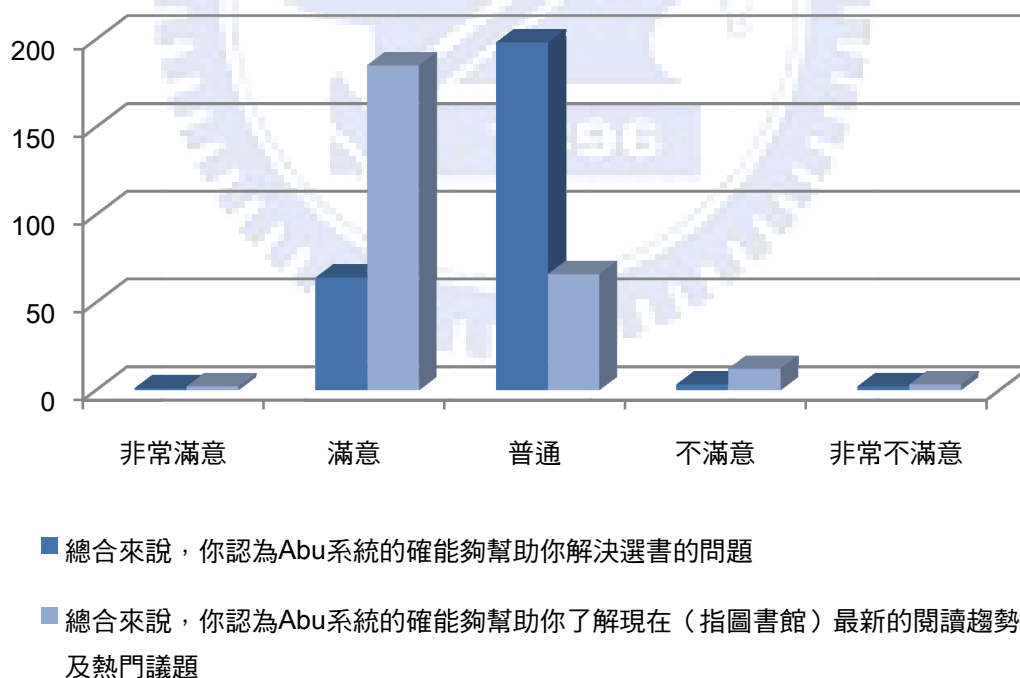


圖 52 受訪者對於 Abu 系統的綜合滿意度

圖 52 顯示受試者對於能夠解決選書的問題上，大多表示普通。在能夠了解最新的閱讀



趨勢上，多數表示滿意。表 7 顯示受試者對於解決選書及了解最新閱讀趨勢的平均滿意度與標準差。

表 7 受試者對於 Abu 系統的綜合滿意度—平均數與標準差

滿意度	解決選書的問題	了解最新閱讀趨勢
平均數	3.220149	3.63806
標準差	0.497225	0.63568

從圖 40 受訪者閱讀新書的比例中可以了解，交大的讀者對於新書的需求是較強烈的，有一半的讀者平時讀的書中是新書，顯示讀者對於最新的閱讀趨勢與議題，其興趣遠大於較經典的、較不新穎的書籍。因此當 Abu 系統以標籤雲呈現社群讀者的標記，並即時的呈現目前的熱門的閱讀趨勢，受訪者對於將此種 Web 2.0 特徵的功能應用在圖書館上，表示了極高的評價。在圖 51 受訪者對於 Abu 的滿意度分析中，多數讀者對於 Web 2.0 的索引功能是非常滿意的。若從表 6 受試者對於 Abu 功能的滿意度分析—平均數與標準差的排序來看，顯示讀者對於如何較快找到正確書籍的需要，是比較強烈的。因此未來可陸續加強索引的多樣性及即時性以滿足讀者的需求。

圖 43 中顯示多數讀者每月都會使用交大圖書館所提供之電子資源，但對於目前圖書館呈現及排序的結果多數感到不滿意（見圖 46）。由於既有的圖書館資料庫系統以關鍵字比對的方式顯示搜尋結果、也未作到過濾，因此使用者多半要在結果清單中一一查閱，顯得較為耗時耗神。在圖 51 受訪者對於 Abu 的滿意度分析中「過濾個人書單」一項，受訪者多數表示滿意及普通的評價，可見使用者對於能夠有效過濾、減少搜尋結果，只顯示有用連結的功能是期待的，同時此功能也能減輕使用者的資訊過載負擔。在「個人推薦書單」方面，系統根據讀者的借閱歷史提供推薦書目，意在減少使用者主動搜尋的次數，也獲得受訪者普通及滿意的評價。目前可能因書籍評價數量仍少的關係，受訪者對於「過濾個人書單」的滿意程度普遍在普通及滿意，但未來隨著評價的累積，應會慢慢呈現對過濾個人書單及個人推薦的效益。

圖 49 受訪者選書時會參考其他熱中閱讀讀者意見之比例中，顯示有近四分之三的受訪

者會偏好從身旁熱中閱讀的朋友中打聽書籍的口碑，以作為選書的參考。但觀察圖 51 中分群的滿意度，並未呈現多數滿意的結果。推測也是因為書籍累積之評價尚未發生「量變產生質變」的階段。由於 Abu 系統中主動功能都要依賴社群讀者的參與，多多貢獻評價，因此未來如何鼓勵使用者多多留下書籍的推薦星級、標記、難易度評價等記錄，成為日後經營 Abu 系統最重要的課題。



## 陸、結論與未來發展方向

本章總結本論文以及實作的 Abu 系統所達成的效益。第一節簡要說明本論文提出的模式對於推廣閱讀產生的應用結果，第二節則說明 Abu 系統及本論文未來可以發展的研究。

### 一、結論

本篇論文結合了協力式過濾的關聯規則探勘，透過圖書館借閱歷史資料為讀者建立了一套個人化的推薦系統，並加大讀者回饋對於推薦清單的影響力，以期吸引更多讀者的互動、豐富網站的評價資料，以期達到推廣閱讀的目的。

在個人化的推薦系統方面，以協力式過濾找出讀者之間的關聯性，使得借閱歷史相似的讀者能夠獲得具有同樣閱讀行為讀者推薦的書籍類別。而在推薦書單上，採用以群眾評價為主的排序方法，讓獲得最多正面評價的書籍優先排序，讀者能夠得到最感興趣的推薦書類、並利用群眾的力量為自己選書。

在讀者回饋機制方面，設計有一至四級直觀的評價方法，讓讀者自由評等書籍的難度。一方面透過評價建立起讀者自身同樣為一至四級的閱讀階段，另一方面豐富圖書館藏書的評價資料。在排序個人書單後，以難易度評級過濾掉不適宜的書籍，進一步減少資訊過載的問題。讀者每一筆的評價資料，都在重新形塑自己的閱讀階段與書籍評價，每一筆的回饋資料都能夠更快速的反應到推薦書單上。

Abu 圖書館個人化推薦系統同時還提供了以下功能：

1. 以「名家書單」機制補足專家推薦的不足，讓 Abu 除了能夠反映出即時、當代、folksonomy 的閱讀活動外，仍保有專業的、經典的、權威的閱讀來源。同時「名家書單」以 XML 形式儲存，更方便專業館員離線編輯而無需要求技術背景。
2. 以 Web Services 串連 Amazon.com 書目資料，並能夠連結反查圖書館的館藏資料庫，使讀者能夠接觸更多元的書目資料。

3. 提供撰寫書評功能，讓讀過的讀者留下質性的文字建議，提供給讀書選書時的另一參考資料。
4. 為了能夠擴大群眾基礎讓更多讀者留下評價資料，Abu 的帳號區隔成二類。一般的使用者註冊後即能夠評價書籍，以交大圖書館證號認證成功的讀者則能夠另外享有個人化推薦書單及查詢歷史借閱之功能。

最後在使用者滿意度調查上，使用過的讀者對於 Abu 系統皆採正面肯定的看法，證明 Abu 融合推薦系統與 Web 2.0 的群眾評價方式，的確能夠解決讀者選書及過濾資訊的問題。相信在經過更長時間的運作後，Abu 累積的評價資料會使推薦功能更貼近 Abu 讀者的需要。

## 二、未來發展方向

本研究所實作的 Abu 系統是個人化推薦與 Web 2.0 結合的一次新嘗試，雖然在初步完成後獲致使用者肯定的評價，但對於個人推薦的精確度以及與圖書館、網路書店等資源的整合性仍有很大的改進空間，同時也需根據系統上線後使用者的意見作為改進參考、實踐 Web 2.0 的精神—使系統的功能與實際需求相符—不致淪於閉門造車。

對於 Abu 系統，本研究認為仍有以下方向可供後續研究：

1. 現今圖書的難易度評價與讀者閱讀階段計算是一種全域性評價 (global rating)，亦即讀者對不同分類號甚至不同領域的書籍評價，都一致等價的累積到自身的評價分數上。未來應建立一種比分類號更為細膩的圖書分類認定方法，才能更準確的將難度評價附加至書籍上。舉例而言，同樣是分類號 312 以下的電腦相關書籍，就有 Office、程式語言、3DMax、Database 等不同專業領域的書籍，而對 Office 的難易評價就不適合套用在程式語言、Database 等不同領域上。
2. 要使 Abu 能更全面的蒐集讀者的偏好以提供更精確的推薦書單，必須加入更多 Schafer 提到的變數輸入(Schafer 2001)。例如當讀者在系統上瀏覽各種不同書籍時，

即計算 click rate、停留時間以及網頁之間的流向等數據並存入使用者描述中 (Profile)，並作為個人推薦的依據之一。

3. 就 Web 2.0 的觀點而言，Abu 的設計在先天上就限定了使用者人數的發展，因為目前的功能設計大部份為交大讀者量身設計，而若要更快速累積書籍評價需更擴大使用者基礎 (user base)，而現今並沒有吸引非交大使用者的誘因。未來可以發展更多非交大讀者可以使用的功能，並降低以借閱歷史作為推薦輸入變數的比重，以擴大可能使用者基礎。
4. Web 2.0 上最引人入勝的特色之一，即是社群內具有高度寫作產量或強烈個人風格的使用者，創造出吸引人的內容並促進閱聽人採取相關行動。Abu 現在並沒有提供讓使用者創立一個自己的社群或 blog，讓活躍的讀者得以透過一個能夠發聲的媒體留下對於書籍的評論，並直接或間接促進讀者對於所談議題、書籍的興趣，進而到訪圖書館借書，達到推廣閱讀的功效。交通大學圖書館的網路服務中，已有閱讀浩然 blog 的平台，將 Abu 和 blog 整合達到相同的效果，也是可以研究的方向。
5. 圖書館的書目資料僅有書名、作者、出版者等詮釋資料，然而對於讀者而言仍需其他選書的參考資料。目前 Abu 僅有來自 Amazon.com 的西文書籍資源，未來可增加來自網路書店的中文圖書資料，以補足圖書館詮釋資料的不足。
6. 就圖書館而言，各個學門從入門、進階、參考到專業程度的閱讀書單，應該是相近的。未來可以利用資料探勘的方法，找出同類學門中 (例如電子系) 建議的閱讀書單。這份書單是有閱讀順序的、而順序是依照難度排序。例如入門的讀者可以先從書籍 A 開始，然後可以讀 B、C、D 中擇一作為進階；若選 B 則讀完後可再自 E、F、G、H 中擇一等等，以減少讀者初入某一領域時選書上的困擾。閱讀地圖也可以利用專家推薦的方法、或是兩者兼備。

## 柒、參考文獻

- Agrawal, R., and R. Srikant. "Fast Algorithms for Mining Association Rules in Large Databases," *Proceedings of the 20th International Conference on very Large Data Bases* 1994, pp. 487-499.
- Andi Gutmans. "What is Web 2.0?" <http://news.zdnet.com/html/z/wb/6103931.html> (2007:2007/4/8), .
- Casey. "Evening Drum and Morning Bell: Sun Yat-Sen and the Revolutionary Party," <http://www.cctv.com/english/TouchChina/China20th/20020626/100013.html> (2007:2007-04-02 03:12:23), 2002-06-24, .
- Dion Hinchcliffe. "The State of Web 2.0," [http://web2.wsj2.com/the\\_state\\_of\\_web\\_20.htm](http://web2.wsj2.com/the_state_of_web_20.htm) (2007:2007/4/8), .
- Gene Smith. "Folksonomy: Social Classification (Atomiq)," [http://atomiq.org/archives/2004/08/folksonomy\\_social\\_classification.html](http://atomiq.org/archives/2004/08/folksonomy_social_classification.html) (2007:2007/4/8), .
- Godfray, . "Challenges for Taxonomy," *Nature* (417:6884), 2002, pp. 17.
- Goldberg, D., D. Nichols, B. M. Oki, and D. Terry. "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM* (35:12), 1992, pp. 61-70.
- Ian, D. "Talis, Web 2.0 and all that," <http://iandavis.com/blog/2005/07/talis-web-20-and-all-that?year=2005&monthnum=07&name=talis-web-20-and-all-that> (2007:2007/4/8), 2005/7/4, .
- Jim, G. "Internet Encyclopaedias Go Head to Head," <http://www.nature.com/news/2005/051212/full/438900a.html> 2005/12/14, .
- Jon Udell. "Collaborative Knowledge Gardening," [http://www.infoworld.com/article/04/08/20/34OPstrategic\\_1.html](http://www.infoworld.com/article/04/08/20/34OPstrategic_1.html) (2007:2007/4/8), 2004/8/20, .
- Josh. "Turning the Lights on," <http://shelfari.typepad.com/> (2007:2007/4/9), .
- Jürgen Schiller Garica. "Web 2.0 Buzz Time Bar," <http://www.scill.de/content/2006/09/21/web-20-buzz-zeitstrahl/> (2007:2007/4/8), 2006/9/21, .
- Kent, A., and And Others. "Use of Library Materials: The University of Pittsburgh Study," 1979, pp. 282.
- Lev, G. "Time's Person of the Year: You," <http://www.time.com/time/magazine/article/0,9171,1569514,00.html> (168:26), 2006/12/25, 2006, .

- Lin, . "Efficient Adaptive-Support Association Rule Mining for Recommender Systems," *Data Mining and Knowledge Discovery* (6:1), 2002, pp. 83.
- Markus Angermeier. "The Huge Cloud Lens Bubble Map web2.0,"  
<http://nonfiction.ig-gestaltung.de/archives/2005/11/11/the-huge-cloud-lens-bubble-map-web20/>  
(2007:2007/4/8), 2005/11/11, .
- Oard, D. W., and G. Marchionini. "A Conceptual Framework for Text Filtering," *College Park, MD, University of Maryland*1996, .
- Resnick, . "Recommender Systems," *Communications of the ACM* (40:3), 1997, pp. 56.
- Robert D. Hof. "Mix, Match, and Mutate,"  
[http://www.businessweek.com/@@76IH\\*ocQ34AvyQMA/magazine/content/05\\_30/b3944108\\_mz063.htm](http://www.businessweek.com/@@76IH*ocQ34AvyQMA/magazine/content/05_30/b3944108_mz063.htm) (2007:2007/4/8), 2005/7/25, 2005, .
- Schafer, . "E-Commerce Recommendation Applications," *Data Mining and Knowledge Discovery* (5:1), 2001, pp. 115.
- The British Library. "History of the British Library,"  
<http://www.bl.uk/about/history.html>2007-03-21 11:12:39), .
- Tim O'Reilly. "What is Web 2.0,"  
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>  
(2007:2007/4/8), 2005/9/30, .
- Wikipedia contributors. "Mashup (Web Application Hybrid),"  
[http://en.wikipedia.org/w/index.php?title=special:cite&page=mashup\\_\(web\\_application\\_hybrid\)&id=121188987&oldid=121188987](http://en.wikipedia.org/w/index.php?title=special:cite&page=mashup_(web_application_hybrid)&id=121188987&oldid=121188987) (2007:2007/4/8), .
- 卜小蝶. "Web 2.0, Library 2.0, Folksonomy," 2006/9/8, 2006, .
- . *使用者導向之網路資源組織與檢索*, 文華圖書館管理資訊股份有限公司, 台北, 2007.
- 尹相志. *SQL Server 2005 資料採礦聖經*, 學貫, 台北, 2006.
- 余明哲. "圖書館個人化館藏推薦系統," 2003, pp. 69.
- 吳晨帆. "推薦策略與推薦資訊對線上推薦績效影響之研究," 2004, pp. 103.
- 吳靖雯. "Web 2.0, 是社會權力轉移的巨大力量!," *數位時代:15分鐘搞懂Web2.0 特刊*  
2006/10/12, 2006, pp. 8-10.
- 高震宇. "認知資訊超載對消費者線上購物決策之影響," 2004, pp. 125.
- 張天立. "網路時代的契機與挑戰:網路書店的經驗分享," *國立成功大學圖書館館刊*第十三期),

中華民國 93 年 6 月, 民 93, pp. 1-6 頁.

陳慧敏."誠品近八成書籍 賣不到 100 本," *經濟日報* (D3 版:2004/11/26, .

曾元顯."新一代資訊檢索技術在圖書館 OPAC 系統的應用," *大學圖書館* (一:三), 民 86, pp. 82-93.

游子賢."通俗分類標記之自動組織 Automatic Organization of User-Generated Tags from the Web," 2005, pp. 34.

楊雅雯."個人化數位圖書資訊環境—以 PIE@NCTU 為例," 民 90, .

盧諭緯."完全解讀 Web 2.0," *數位時代* 134), 2006, pp. 56-63.

戴玉旻."圖書館借閱記錄探勘系統," 2001, pp. 89.







## Abu個人化圖書推薦系統使用狀況調查

作者: Tzu-Wen Lo

親愛的同學 您好：

這是一份學術性問卷，目的在於瞭解交通大學讀者使用Abu個人化圖書推薦系統的使用行為及滿意程度，以提供研究者瞭解您的使用情況及滿意度。您的回答將是本研究完成的最大關鍵。請您依據使用Abu各項功能之感受撥冗作答。您所提供的資料僅作學術性上之研究分析，內容絕對保密，敬請安心填寫。謹致十二萬分的謝意！

敬祝您

身體健康 萬事如意

1\* 請問你每天花多少時間上網？

- 一小時以下    二至三小時    三至五小時  
 五小時以上

2\* 請問你上網的資歷有多久？

- 一年以下    一年至三年    三年至五年  
 五年至七年    七年以上

3\* 請問你逛書店或接觸書籍的頻繁是？

- 幾乎每天    二到三天一次    每週一次  
 每月一次    每學期一次

4\* 您的閱讀行為中，有多大比例是讀新書？

- 一成以下    一至三成    三成至五成  
 五成至七成    七成至九成    九成以上

5\* 您的閱讀行為中，有多大比例是讀學校的教科書、參考書？

- 一成以下    一至三成    三成至五成  
 五成至七成    七成至九成    九成以上

		每 天	每週 幾次	每月 幾次	每學期 幾次	從未拜 訪過	
6*	請問您親自到訪交大圖書館的頻率是：	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
7*	請問您使用交大圖書館電子資源，包括網站、blog、電子資料庫等的頻率是：	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
8*	請問您利用圖書館館藏目錄的頻率是：	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
9*	請問您利用圖書館新書目錄的頻率是：	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
			非常	同	普	不同	非常不

		同意	意	通	意	同意
10*	就您以關鍵字查詢館藏資料庫時，您覺得查詢的結果數量適中，沒有相關的資訊出現。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11*	你覺得以關鍵字查詢館藏資料庫查到的書目，可以解決你找書及閱讀選書上的問題。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12*	你覺得以關鍵字查詢館藏資料庫後查詢結果的排序方式，可以解決你找書及閱讀選書上的問題。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13*	你覺得新書目錄以圖書分類（總類、自然科學、哲學...等）的方式，可以幫助你找到符合自己興趣和需求的新書資訊。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14*	你覺得目前圖書館電子資料庫或網站的設計，可以幫助你很快的了解目前學校或社會上的閱讀趨勢或熱門議題。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

15\* 你都以什麼方式接觸新書資訊？

- 出版社網站    網路書店網站    實體書店  
 郵購    報紙雜誌    圖書館  
 親友介紹  
 其他請註明:

16\* 在選擇書籍閱讀時，你會參考其他熱中閱讀讀者的意見。

- 非常同意    同意    普通    不同意  
 非常不同意

		非常需要	需要	普通	不需要	非常不需要
17*	圖書館若提供Web 2.0的網站服務，你對的個人化推薦服務需求程度為何？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18*	圖書館若提供Web 2.0的網站服務，你對Web 2.0的圖書索引方式（例如標籤雲，tag cloud）的需求程度是？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19*	圖書館若提供Web 2.0的網站服務，你對依個人閱讀難易程度過濾個人推薦書單的需求程度是？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20*	圖書館若提供Web 2.0的網站服務，你對依大家的評價決定書籍的分群、以達到讓書籍的分群更貼近讀者需要，此項功能的需求程度是？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

		非常滿意	滿意	普通	不滿意	非常不滿意
21*	在實際使用過Abu個人化推薦系統後，你對個人化推薦服務的滿意度為何？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22*	在實際使用過Abu個人化推薦系統後，你對Web 2.0的圖書索引方式（標籤雲，tag cloud）滿意度為何？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

23*	在實際使用過Abu個人化推薦系統後，你對依個人閱讀難易程度過濾個人推薦書單的滿意度為何？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24*	系統後，你對大家評價後圖書的分群結果滿意度為何？	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

25 (選填) 如果你對Abu個人化推薦系統不甚滿意，你覺得可能的原因為何？

- 推薦的原始資料 (借閱歷史) 不完全
- 我的借閱冊數、時間不夠久
- 系統設計的機制不夠週全
- 其他 請註明:

26\* 總合來說，你認為Abu個人化推薦系的確能夠幫助你解決選書的問題。

- 非常同意     同意     普通     不同意
- 非常不同意

27\* 總合來說，你認為Abu個人化推薦系的確能夠幫助你了解現在 (指圖書館) 最新的閱讀趨勢及熱門議題。

- 非常同意     同意     普通     不同意
- 非常不同意

28 (選填) 除Abu現有的功能外，你還希望Abu下一次更新能提供什麼功能？

---



---



---



---



---

必須回答有 '\*' 記號的問題

謝謝您完成此份問卷。

全卷完