胡雅涵、李彥賢、林正賢（2015），『結合社會性標籤及文獻內容於個人化學術文章推薦』，*中華民國資訊管理學報*，第二十二卷，第二期，頁 171-198。

結合社會性標籤及文獻內容於個人化學術文章推薦

胡雅涵
國立中正大學資訊管理學系
國立中正大學前瞻製造系統頂尖研究中心

李彥賢*
國立嘉義大學資訊管理學系

林正賢
國立中正大學資訊管理學系

摘要

學術文章推薦是近年來熱門的研究議題，過往針對學術文章推薦研究上，普遍利用學術文章內的屬性資料，如：標題、摘要、關鍵字、作者名稱以及參考文獻標題等進行推薦。然而除了上述的「內部資訊」外，學術文章中亦包含其他與該研究相關的「外部資訊」，像是參考文獻摘要及具相同社會性標籤（Social tagging）之文件等。藉由分析外部資訊，應能有助於取得與原始學術文章相關之其他研究主題的關鍵字詞，進而推薦更符合學術研究需求的文章。本研究同時考量使用者喜好文章之內、外部資訊，包括標題、摘要、關鍵字、參考文獻標題、參考文獻文章內容、社會性標籤、以及具相同社會性標籤文章內容，並以內容導向式推薦方法為基礎來建構推薦系統。此外由於不同文章屬性應具有不同程度的重要性，本研究進一步運用層級分析法（Analytic Hierarchy Process）制定出各個文章屬性之權重值，用以對文章之各個屬性相似度進行加權運算，並產生最終之推薦清單。本研究最後以實驗方式進行推薦效能評估，並以不同屬性組合之推薦方法做為評估比較基準。實驗結果顯示，在本研究採用之成對比較法（Pair Match）及命中率（Hit Rate）兩個評量指標下，本研究提出的推薦方法相較於傳統僅考慮內部資訊之推薦方法，能有較高之推薦命中率，且在文章推薦排序上能有更顯著地改善，亦即將使用者喜好程度較高之文章給予優先的推薦順位，說明本研究之學術文章推薦方法具有較佳的推薦效果。

關鍵詞：推薦系統、內容導向式推薦方法、學術文章推薦、社會性標籤、文字探勘

* 本文通訊作者。電子郵件信箱：yhlee@mail.ncyu.edu.tw
2014/06/10 投稿；2014/07/26 修訂；2014/11/12 接受

Hu, Y.H., Lee, Y.H. and Lin, J.H. (2015), 'Combining social tagging and reference content for personalized academic document recommendation', *Journal of Information Management*, Vol. 22, No. 2, pp. 171-198.

Combining Social Tagging and Reference Content for Personalized Academic Document Recommendation

Ya-Han Hu
Department of Information Management, National Chung Cheng University
Advanced Institute of Manufacturing with High-tech Innovations, National Chung Cheng University

Yen-Hsien Lee*
Department of Management Information Systems, National Chiayi University

Jheng-Hsien Lin
Department of Information Management, National Chung Cheng University

Abstract

Purpose－This study aims at developing a novel academic article recommender system. The abstract of the articles that share similar social tags with and that are referred by the preference articles of the targeted user are usedto improve the effectiveness of the recommender system.

Design/methodology/approach－This study adopts the content-based method to determine the similarity between two academic articles and make recommendations. Seven attributes relevant to academic article, including title, abstract, keyword, reference, reference article, social tag, and article with similar social tag, are used to extend the original preference document vector. The analytic hierarchy process method is applied to determine the weights among the seven attributes by their degree of importance. Aweb-based recommender system was developed for the evaluation purpose. We invited over 90 subjects and adopted Pair Match and Hit Rate as performance criteria in the evaluation experiment.

* Corresponding author. Email: yhlee@mail.ncyu.edu.tw
2014/06/10 received; 2014/07/16 revised; 2014/11/12 accepted

Findings－Theevaluation results show our recommendation method outperforms the five benchmarks. External information such as articles with similar social tags and reference articles used in the system can contribute to the better ranking list of recommending articles, in terms of the Pair Match and the Hit Rate.

Research limitations/implications－This study only considers a limited set of academic articles as the investigated corpus. Future research shall expand the array of articles and considerco-authorship and co-citation relationships as important features.

Practical implications－Though the rise of digital libraries makes the acquisition of academic resources easier, a sharp increase of academic articles leadsthe search for relevant articles ineffective. The proposed method can facilitate researchers in searching and retrieving relevant academic articles based on their individual preference profile.

Originality/value－This paper is the first that investigates the influence of both internal and external information onmaking recommendation of academic article. It advances literature in determining valuable article features for optimizing the performance of recommender system.

Keywords: recommender systems, content-based recommendation, academicarticle recommendation, social tagging, text mining.

壹、緒論

網際網路蓬勃發展使得網路上的資訊呈現爆發性成長，如何讓使用者在廣大的資料中快速獲取合適的資料已成為相當重要的議題。儘管網路搜尋引擎能讓使用者進行關鍵字搜尋，然而搜尋的結果卻往往令使用者陷入另一個資訊負載的情境之中，亦即使用者通常需要耗費相當多的時間去檢視搜尋結果並過濾想要的資訊。因此，若能學習分辨使用者喜好且能主動提供符合使用者需求的相關資訊之方法，將可減少使用者在資訊搜尋上的成本，並同時提高效率（Lin et al. 1998）。事實上，推薦系統（recommender systems; RSs）已行之有年，目前被廣泛地運用在各個不同的領域之中，例如：商品推薦、評論推薦、影音推薦及書籍推薦等（Choi & Ahn 2011; Choi et al. 2006; Wen et al. 2012; Chen et al. 2006; Weng et al. 2009），其藉由分析使用者的喜好，以提供自動化推薦服務（Cao & Li 2007; Liu et al. 2009; Senecal & Nantel 2004）。

另一方面，據 Jinha（2010）統計指出，全世界已被刊登之學術文章數量，從 1665 年到 2009 年累積已達到 5000 多萬篇。National Science Foundation（NSF）的統計數據也顯示，現今全球學術文章的輸出量正在快速的成長，過去 20 年全球發表於國際同儕審查期刊的研究文章，從 1988 年 460,200 篇成長到 2009 年約 788,300 篇。因此如何在眾多學術文章中，有效地提供符合研究人員需求的相關學術文獻資料，來提升研究人員的研究能量，已成為推薦系統研究的主要議題之一。

相較於一般文章，學術文章具有其特定格式。一篇學術文章的內容包含標題（Title）、關鍵字（Keyword）、摘要（Abstract）、本文內容（Content）與參考文獻（References）。其中關鍵字為文章作者所制定，通常關鍵字會包含本文的關鍵技術及所屬領域，旨在提供索引以利圖書書目系統編碼，供日後有興趣的人士依關鍵字查詢；摘要是整篇文章的縮影，以簡短的方式說明研究目的、方法、與結論等內容，能反應整篇文章的精髓；參考文獻則是作者文章撰寫上曾參考到的其他學術文章。上述格式主要是希望讀者在閱讀學術文章能快速地了解文章主要的內容，以達成有效的知識傳遞。

由於學術文章亦屬於文本文件（textual document），原則上可以利用傳統文件分析方式進行文章推薦；學術文章中所包含之各項內容隱含不同之文章資訊，因而需要發展其特有的推薦方法，以期能提高推薦的準確度。過去已有不少學者針對學術文章推薦進行研究（Ahlgren & Colliander 2009; Chen et al. 2006; Choochaiwattana 2010; Gemmell et al. 2012; Hwang et al. 2010; Xu et al. 2012）。大抵來說，許多研究多半利用文章中的標題、摘要、關鍵字、及文獻標題等文章資訊，做為計算文章推薦優先順序的依據。研究結果也證實，若將學術文章中的各種資

訊分別進行處理，將有效提升推薦系統的效能。儘管如此，過去研究基於分析使用者喜好文章的「內部資訊」，如標題、關鍵字、摘要、本文、以及參考文獻標題等資訊，做為推薦系統評估推薦文章之依據，可能會導致推薦文章過於特定的問題，亦即僅推薦與使用者喜好文章內容相似的文章。相似文章雖能讓研究人員瞭解目前的研究狀況；但從另一方面來看，研究人員亦同時需要與研究主題有關之其他文獻資料。舉例來說，對於進行內容基礎推薦系統研究的使用者來說，除了與內容基礎推薦系統相似的文獻之外，其他文獻如文件分類、文本分析、網絡分析等亦可能有關於使用者的研究，亦須被納入推薦的考量之中。

學術文章中除了「內部資訊」外，還包含其他與該研究相關的「外部資訊」，像是論文參考文獻及具相同社會性標籤（Social tagging）之文件。基本上，學術研究建構於過去相關研究的基礎之上，在文章篇幅的考量之下，多半會以引文型式呈現於文章中；換句話說，參考文獻與引用它的學術文章必有一定程度的相關性，而藉由分析參考文獻內容或許可取得與該研究相關的其他資訊。其次，隨著網路互動技術的發展，許多網路社群允許成員對網站內容或項目依自己的喜好給予關鍵字或標籤，讓成員參與分類的過程，而非單純接受網站的資料分類方式，稱之為社會性標籤（Social tagging）制定或協同式標籤制定（Collaborative tagging）（Chi & Mytkowicz 2007; Passant & Laublet 2008）。在學術社群中，不同的專家們可依自己的角度來對同一篇學術文章給予標籤。此方式可視為是傳統學術文章中關鍵字的延伸，其透過多位專家，而非作者本身來定義該篇文章的重要領域與概念。因此，擁有相同社會性標籤之文章應具備一定程度的相關性。

要取得與一篇學術文章的相關外部資訊，在現在已不是難事。近年來學術文章電子化及線上數位圖書館的風行，要取得特定學術文章的相關外部資訊已非難事；網際網路的超連結特性，使得學術文章與引文之間的連結更加緊密，進而讓人容易地蒐尋及取得與論文相關的外部資訊。舉例來說，美國 Institute for Scientific Information（ISI）公司建置的引用文獻索引資料庫系統（Web of Science Core Collection; WOS[1]）即提供各學科領域之學術文章文獻書目及摘要資料，並以超連結方式讓讀者跳躍於文章內容與引文內容之間。而近年來興起的學術文章社群網站，例如 CiteULike[2]則吸引許多學者前來搜尋其感興趣的文章，並給予文章個人化標籤，來進行個人化的文章典藏管理（Kim et al. 2010）。

整體來說，學術文章的外部資訊隱含與該文章相關研究的資訊。因此，如能藉由分析參考文獻內容與具相同社會性標籤之文件內容，應能取得與原始學術文章相關之其他研究主題的關鍵字詞，而這些關鍵字詞將有助於推薦系統在評估推

1 http://apps.webofknowledge.com/UA_GeneralSearch_input.do?product=UA&search_mode=GeneralSearch&SID=Q1peChFLU5dp4icBKdh&preferencesSaved=

2 http://www.citeulike.org

薦文章時，能進一步考量與使用者喜好主題相關之其他研究，進而推薦更符合學術研究需求的文章。基於上述，本研究提出一個以內容為基礎、考量的文章外部資訊之學術文章推薦系統（content-based recommender system using external information; CRSEI），並進行文章推薦實驗評估。相較於先前以文章內部資訊為基礎的推薦研究，本研究試圖瞭解文章相關外部資訊是否有助於提升學術文章推薦系統之推薦效能，亦即在推薦準確度及推薦排名上的提昇。本研究指出一篇學術文章中所具備的七個文章屬性包括：標題、摘要、關鍵字、參考文獻標題、參考文獻文章內容、社會性標籤、以及具相同社會性標籤文章內容；其中前四項屬於文章內部資訊，餘三項則屬文章外部資訊。此外，在評估文章的推薦程度時，出現在不同文章屬性中的字詞應具有不同程度的重要性。儘管如此，過去研究大多以經驗法則方式權衡文章屬性的重要性。對此，本研究利用層級分析法（analytic hierarchy process; AHP），藉由學者專家的意見來制定學術文章屬性間之相對權重值，並在評估文章推薦順序時用以調整各屬性推薦分數之加權值。

本文後續內容編排如下，第貳節將回顧與本研究相關之文件推薦研究；第參節詳細說明本研究提出之學術文章推薦系統；第肆節說明實驗設計；第伍節討論相關實驗結果；最後則於第陸節綜合探討研究結果，並提出未來研究方向。

貳、文獻探討

個人化文件推薦（Personalized Document Recommendation）係藉由分析使用者個別特性或相關資料來萃取出使用者喜好特徵，然後依據其個人喜好特徵進行文章推薦（Chiu et al. 2010; Kim et al. 2010; Lai & Liu 2009; Lavie et al. 2010; Liu et al. 2008; Liu et al. 2011; Zheng & Li 2011）。過去研究中常見的推薦技術大致分為協同過濾推薦（collaborative filtering recommendation; CFR）、內容基礎推薦（content-based filtering recommendation; CBR）、以及上述兩種技術的綜合體（Basilico & Hofmann 2004; Fouss et al. 2007; Kim et al. 2010; Liu et al. 2011; Martín-Vicente et al. 2012）。協同過濾推薦方法主要尋找與目標使用者具有相同品味或特性的使用者（鄰居），並評估這些鄰居們的喜好項目（items），從中找出適合推薦給目標使用者的項目。內容基礎推薦方法則透過分析目標使用者喜好的項目內容或屬性，從中找出目標使用者偏好項目的重要特徵，並找尋具有相似重要特徵之項目來推薦給目標使用者。以文件推薦來說，協同過濾推薦方法會先找出與目標使用者閱讀或喜好相似文章的鄰居使用者，並找出鄰居使用者所擁有但卻不在目標使用者清單中的文件，做為推薦候選文件；最後則綜合所有相似使用者的喜好程度，來決定是否推薦某候選文件給目標使用者。相較於協同過濾推薦在找尋相似品味的使用者，內容基礎推薦則著重在目標使用者喜好文件內容的分

析。內容基礎推薦方法通常先將目標使用者閱讀或喜好的文章進行相關的文件前處理，像是斷字或同義字處理，以形成字詞集合；藉由分析工具找出可能代表使用者偏好的關鍵字詞集合，例如透過字詞權重公式（如 term frequency-inverse document frequency; TF-IDF），計算出字詞權重以剔除掉不具重要性的字詞；接著利用相似度比對方式，找出與關鍵字詞集合相似度高之其他文件，做為要推薦給使用者之目標文件。過去研究的實證指出，相較於協同過濾推薦方法，內容基礎推薦方法基本上有較佳的效能與效率，特別是在面對資料稀疏的問題上（Pazzani & Billsus 1997; Kazienko 2007; Zhang & Koren 2007; Semeraroet. al. 2009）。

目前常見文件推薦應用有新聞時事文件推薦（Wen et al. 2012）、評論文件推薦（Martín-Vicente et al. 2012）、部落格文件推薦（Li & Chen 2009）及一般文件推薦（Chen et al. 2006）。過往在文件推薦的研究之中，許多學者考慮各種不同的特徵屬性進行推薦系統研究。Liu 等（2011）考量時間及信任（Trust）兩個關鍵因子於推薦方法中，針對目標使用者，首先藉由評估信任度來找出其相似使用者，再利用時間的先後順序調整不同文章的重要程度，最後進行推薦。Li 與 Chen（2009）研究部落格文章推薦，除考慮文章之內容外，也考慮信任與作者間的社群關係（Social Relation）等屬性，以進行推薦預測數值的計算。Liu 等（2011）提出一種在行動裝置上進行之部落格推薦系統，其做法是將使用者之點擊次數以及文章發表的時間做為考量的因子，利用公式計算出文章熱門程度與時間遞移的關係，最後找出兼具高熱門程度以及最新的部落格文章進行推薦。Wen 等（2009）將圍繞在圖片或影音附近的文字敘述內容視為分析資料，並進行字詞的分析，最後則以內容字詞分析結果來代表圖片或影音，達成後續推薦評估的目的。

在學術文章的推薦研究當中，Hwang 等（2010）利用共同作者的關係來做為文件推薦的考量因素。其利用作者之間的研究合作關係所建構出的社群網路（Social Network）中計算每位作者對彼此的影響程度，最後再針對目標學者進行個人化學術文章推薦。Xu 等（2012）判別學術文章關鍵字之語意，並建立學者間社群網路來進行學術文章的推薦。在關鍵字處理部分，其擷取文章作者所制定的關鍵字，利用 WordNet 來進行字詞語意的判別，再透過關鍵字萃取（Terminology extraction）、語意解釋（Semantic interpretation）、以及知識本體建構（Ontology construction）等步驟，以定義出字詞之間在語意上的相似程度，用以決定兩文章在關鍵字上的語意相似程度。在建立學者間的社群網路方面，則透過計算共同發表文章的次數來定義出作者之間的強弱關係。最後則是將兩學術文章語意相似度與作者間強弱關係度進行加總，並根據加總後數值的高低來決定文章推薦順序。

Chen 等（2006）將學術文章切割成標題、摘要、關鍵字以及參考文獻標題等四個屬性來進行相似度的運算，其中標題與關鍵字的字詞向量處理使用布林模式；摘要關鍵字的處理則是利用 TF-IDF 來計算；參考文獻的處理則是計算兩文章

中共同參考之文章數量。利用上述方法研究計算出個別屬性的相似度數值，再以倒傳遞類神經函數（back propagationnetwork; BPN）進行四個屬性相似度數值的合併，以產生推薦清單。Gemmell 等（2012）利用使用者所制定的標籤進行學術文章的推薦，其透過具有社會性標籤（social tagging; ST）之網站（如：CiteUlike、Bibsonomy 等）所提供之資料集，並結合內容基礎與協同過濾方法所組成的混合式推薦方法來進行學術文章的推薦。

綜觀本文所回顧的研究，文章推薦系統多半仍以文章的內部資訊，亦即標題、摘要、關鍵字、參考文獻標題等，做為評估推薦文章的基礎，而未考量學術文章特有的參考文獻的文章內容與具相同標籤文章內容等相關外部資訊。因此，本研究將試圖運用外部資訊於推薦系統之中，期望除了能推薦與研究人員喜好相似的文章外，更能夠廣泛推薦與其研究相關的文章，以滿足研究人員在研究相關資訊上的需求。此外過去研究在評估推薦文章時多未考量不同資料屬性的相對重要性，或是以經驗法則方式決定其相關屬性權重。因此，本研究試圖搜集學者專家的意見，並利用層級分析法來制定學術文章屬性間之相對權重值，以做為進一步調整推薦文章排序的依據。

參、研究方法

本研究建置 CRSEI 文章推薦系統之流程如圖 1 所示。第一階段為學術文章蒐集，主要從電子期刊代理商網站抓取特定領域的學術文章建立資料集（Complete Set of Documents），並同時建立外部資料集（Expansion Documents），做為本研究提出之方法所需的外部資料。緊接著將資料經由資料前處理（Data Preprocessing）的步驟後，便可在下一階段形成資料集文章向量（Document Profile）。此外目標使用者可由本研究資料集中挑選其所喜好的文章，來建立個人向量（User profile）。最後階段則透過目標使用者向量與資料集所有文章向量進行相似度計算，並在考量不同屬性加權值之下，產出最終的文章推薦順序清單。以下將分別說明本研究在學術文章蒐集、文件前處理、以及文章屬性權重制定、以及文章推薦評估等階段的細節。

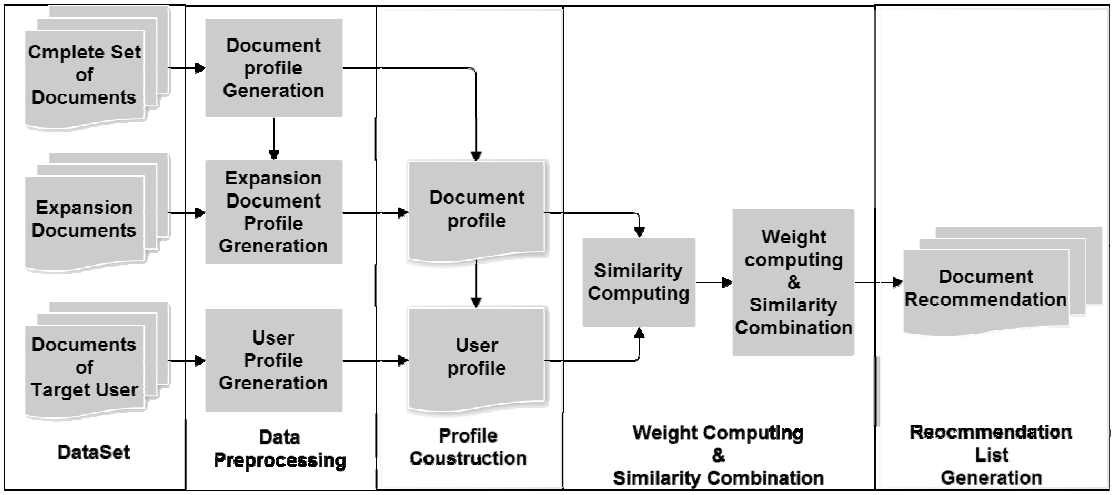

圖 1：CRSEI 文章推薦系統流程圖

一、學術文章蒐集

在文章資料蒐集上，本研究需分別為每個文章屬性建構特徵向量，因此需先蒐集文章屬性的相關資料，包括：

1. 文章標題（title; TI）：目標文章之標題。
2. 文章摘要（abstract; AB）：目標文章之摘要。
3. 文章關鍵字（keyword; KY）：目標文章之關鍵字。
4. 參考文獻標題（reference; R）：目標文章中所有參考文獻的標題。
5. 參考文獻文章（reference article; RA）：目標文章中參考文獻之文章內容。
6. 文章標籤（tag; T）：外部使用者對於目標文章所訂定的關鍵字。
7. 標籤文章（tag article; TA）：資料集中與目標文章具有相同標籤之其他文章摘要集合。

除學術文章本身的內容外，本研究仍需取得各文章所引用的文獻摘要及相同標籤文章摘要等資料。在考量資料取得的便利性與完整性之下，本研究將實驗資料來源限定在 ScienceDirect 及 CiteULike 網站上。其中，學術文章是從期刊代理商 ScienceDirect 網站（http://www.sciencedirect.com/）進行收集，一篇由 ScienceDirect 所收錄的文章，可直接由網站上取得其 TI、AB、KY、及 R 等資訊。而 RA 的部份，本研究僅考量目標文章中被 ScienceDirect 中所收錄之 RA 摘要文字做為 RA 的內容，此部分之資料亦可經由 ScienceDirect 網站上所附之連結取得。每一篇學術文章之參考文獻資料有所不同，可能包含書籍資料、會議資料、期刊資料或是網站資料等，雖然近年來各出版商已努力建構相關文獻資料庫，但以目前來說，完整資料在實際取得上仍有難度。因此本研究僅能將參考文獻資料之擷取限縮在

ScienceDirect 提供的期刊範圍內，此為本研究在學術文章資料取得上的限制。

在標籤及具相同標籤文章摘要之資料取得上，本研究主要來源為 CiteULike 網站（http://www.citeulike.org/）。CiteULike 網站提供使用者自訂文件標籤的功能，因此透過 CiteULike 網站資料庫可以取得 T 屬性資料，而藉由標籤連結，可同時收集其他具有相同標籤的文章摘要（即 TA）。此外，雖然 CiteULike 中的每篇文章大多被讀者給予一個以上的標籤，但是與目標文章完全具有相同標籤之文章卻相對不多，對此本研究利用 CiteULike 內建的標籤搜尋引擎，找尋與目標文章標籤集相似的文章，做為資料蒐集對象。

圖 2 為 CiteULike 網站上利用標籤進行搜索之後所顯示的畫面，這些文章都共同地被“recommendation”所標記，而同樣被標記的文章總數，則有 200 多篇。在資料量龐大且考慮系統效能的情況下，本研究僅針對前 10 篇具高度相關之標籤文章進行蒐集，以此做為具相同標籤文章摘要之屬性資料，並進行後續研究。

圖 2：CiteULike 標記文章

本研究蒐集之學術文章分為兩大類別：推薦系統類與電子商務類文章。文章資料是由五個知名期刊取得，包括：Decision SupportSystems、Information Processing & Management、Data & Knowledge Engineering、Knowledge-Based Systems 以及 Information Systems，以 2000 年至 2012 年文章資料為主要收集範圍，而所收集期刊文章之 T 與 TA 資料則由學術社群網站 CiteULike 所取得。

在資料收集階段，原始未經處理的期刊本文資料總篇數為 1,256 篇；參考文獻標題共 21,501 筆；參考文獻內容總文章數 19,023 篇；標籤文章內容總篇數 147,320

篇。上述資料皆經由下列資料過濾條件進行篩選：

1. 刪除原始本文期刊文章缺少標題、摘要以及關鍵字之文章。
2. 刪除未引用期刊的本文文章。
3. 刪除未被標籤的期刊本文文章。
4. 刪除參考文獻內容缺乏摘要的文章。
5. 刪除標籤文章內容中，未同時具備文章標題以及摘要的文章。

經上述條件過濾後，本研究實驗最終考量的期刊論文共 1,241 篇，其中包含 642 篇推薦系統類與 599 篇電子商務類論文，各期刊所蒐集的論文篇數整理如表 1；參考文獻標題供 19,830 筆；參考文獻摘要共 17,410 筆；文章標籤共 3,815 個；具相同標籤文章摘要 124,076 篇。

表 1：期刊資料分布表

期刊名稱	篇數
Decision Support Systems	250
Information Processing & Management	246
Data & Knowledge Engineering	252
Knowledge-Based Systems	245
Information Systems	248

二、文件前處理

本研究以 Vector Space Model（VSM）來表示文件。在進行文件相似度比對前，需先定義出能代表文章內容的所有關鍵字詞，以便後續將文章進行前處理以形成文章向量。關鍵字詞（key phrases）是用來描繪某個概念的單字或單詞；換句話說，關鍵字詞是具有能夠表達文章意涵的詞語，也就是該字詞具備詞彙辨別力（Term Discrimination Value），可用來有效區分不同的文章。因此，若能準確地擷取到文章關鍵字詞，便能掌握該文章之焦點。由於本研究分別考量七個文章屬性，因此針對所蒐集的文章內容，本研究分別建立不同屬性各自的關鍵字詞集合，並形成七個屬性向量。

文章前處理的過程包含斷字（Word Segmentation）、停字（Stop-word List）處理、以及一字多形（Morphological Processing）處理等。首先，英文文章的單字之間是以空白做區隔，故在處理斷字上，是以空白作為字詞間切割條件，以取得文章中個別的字詞。停字（stop-word）通常不帶有任何資訊，因此需被刪除而不會當作關鍵字詞用以檢索之用（Ur-Rahman & Harding 2012; Yang & Lee 2004）。在停字處理的部份，本研究直接依據過去專家的所建立的停字表（Stop-word list）來去

除文章中不具意義的詞彙。在詞性處理方面，則使用史丹福大學自然語言研究團隊所開發出來的字詞詞性判斷工具 Part Of Speech Tagger（http://nlp.stanford.edu/software/tagger.shtml），利用 Tagger 所回傳的字詞詞性，擷取其中的動詞以及名詞作為關鍵字詞集。一字多形（Morphological Processing）的問題是指詞彙在不同的文章內容裡可能會有複數、動名詞、或過去式的變形，這些詞彙必須經過抽取詞幹（stemming）的處理後，方能予以一致化（Gajendran, Lin & Fyhrie 2007）。本研究使用 Porter Stemming 演算法進行一字多形的處理，此演算法的原理是運用一個已建立好的字尾列表及一連串的判斷規則，用以切截詞彙，只留下詞彙字根。例如，詞彙 *addresses* 可由規則＜*sses→ss*＞切截詞彙，留下字根 *address*。

如表 2 所示，經由上述的文件前處理步驟後，所找出的字詞仍然非常多，除造成向量維度過高的問題外，並同時產生相當稀疏的文件字詞矩陣，亦即資料稀疏性的問題，因而影響文件間相似度比對的效率。因此，為降低向量維度、提高系統效能並降低文章中字詞雜訊的影響，本研究進一步使用 Term frequency-Inverse Document Frequency（TF-IDF）進行文字權重計算，針對每個文章屬性，保留前 1000 個重要字詞（若不足 1000 則全部保留）做為文件表達的關鍵字集。

表 2：不分類資料集屬性數量表

	標題	摘要	關鍵字	參考文獻標題	參考文獻內容	標籤	標籤文章內容
電子商務	1474	3280	1421	6910	13277	863	8053
推薦系統	1482	3438	1453	4016	12338	912	7925
總計	2198	4777	2216	8343	16287	1406	9697

三、層級分析法制定文章屬性權重

本研究使用層級分析法來制定文章屬性的權重，專家依其對於學術文章的經驗，來評估不同文章屬性對於自身判斷文章喜好的重要程度。表 3 為本研究使用之層級分析法評估尺度說明。表 4 為某專家之受測結果，以比較矩陣的形式呈現，其中對角線為屬性自身的比較，故其值均為 1，而下三角部分的數值則為上三角數值的倒數，藉此建立出比較矩陣。隨後利用層級分析法進行幾何平均數的計算，其做法係將矩陣之橫向數值相乘後，再開 n 次方得出數值，最後將各屬性所計算出來的幾何平均數除以總幾何平均數，即可獲得屬性的權重值。

表 3：層級分析法評估尺度意義及說明

尺度	定義	說明
1	Equal importance	兩項屬性的要重程度具同等重要性
3	Weak importance	稍微偏向某一屬性較為重要
5	Essential importance	傾向某一屬性較為重要
7	Demonstrated importance	非常強烈傾向某一屬性為重要
9	Absolute importance	肯定絕對傾向某一屬性為重要

表 4：比較矩陣範例

	TI	**AB**	**KY**	**R**	**RA**	**T**	**TA**
TI	1	3	3	7	5	9	9
AB	1/3	1	3	7	5	7	5
KY	1/3	1/3	1	9	5	7	5
R	1/7	1/7	1/9	1	1/5	1/7	1/5
RA	1/5	1/5	1/5	5	1	5	3
T	1/9	1/7	1/7	7	1/5	1	1/3
TA	1/9	1/5	1/5	5	1/3	1/3	1

本研究使用上述步驟求得各個文章屬性之間權重數值，來表示專家對於文章各個屬性間的重視程度。其數值越高，表示專家認為此屬性在構成影響文章喜好的程度較高；反之則程度較低。本研究收集多位專家的意見，並利用群體專家權重制定方法，亦即將個別專家計算出的權重結果進行平均，來求得整體屬性權重值。

本研究利用問卷調查方式來取得專家對於屬性相對權重意見，主要對象為任職於大專院校的教授及博士生。問卷共發放 20 份，回收問卷 18 份，回收率為 90%。將回收後之問卷以問卷分析工具 Expert Choice 11.5 決策分析軟體進行權重運算以及一致性檢定，剔除未達一致性指標（consistency index; C.I.）的受測問卷，以確保權重值計算的正確性。最後所剩之有效問卷為 16 份，其中包含博士生 2 位、講師 2 位、助理教授 8 位、副教授 6 位，年齡分布介於 30 至 40 歲之間。問卷的職位分佈以助理教授居多，其次為副教授，顯示主要受測對象皆有多年學術研究經驗。

表 5 為 AHP 問卷權重計算之結果，其中顯示每一個族群所重視的文章屬性有著些微的差異，但綜觀而言，不同族群的專家都認為文章摘要為判斷是否喜好某

學術文章的主要依據，其次則為文章標題與文章關鍵字等屬性。本研究最後採用整體平均權重，做為評估推薦文章時，各屬性相似度調整之加權值。

表 5：AHP 運算結果

屬性	博士生	講師	助理教授	副教授
TI	0.2173	0.2605	0.2617	0.2660
AB	0.3979	0.3601	0.2979	0.3206
KY	0.1556	0.1639	0.1247	0.1676
R	0.0500	0.0555	0.1034	0.1146
RA	0.0393	0.0376	0.0848	0.0390
T	0.0950	0.0979	0.0461	0.0458
TA	0.0448	0.0245	0.0815	0.0464

四、文章相似度計算

本研究將一文件之七個屬性均表示為關鍵字詞的二元向量，並利用 Jaccardcoefficient 來計算兩文件屬性之間的相似度。以文章標題為例，令 d 為文章資料庫 D 中之一篇文章，$K^{TI}=\{k_1^{TI}, k_2^{TI},..., k_n^{TI}\}$ 為依據所有文章標題所萃取出之關鍵字詞集合，則針對 d 之標題內容可依 K^{TI} 建立文章標題向量 $\overline{d^{TI}}=\{w_1^{TI}, w_2^{TI},..., \underline{w_n^{TI}}\}$，其中若字詞$k_i^{TI}$（$1 \le i \le n$）出現在文章 d 之標題中，則w_i^{TI}= 1；否則w_i^{TI}= 0。令 $\overline{d_x^{TI}}$ 與 $\overline{d_y^{TI}}$ 為兩篇文章 x、y 之標題二元向量，其標題相似度$S_{u,x}^{TI}$可以 Jaccard coefficient 求得如下：

$$Jaccard\left(\overline{d_x^{TI}}, \overline{d_y^{TI}}\right)=\frac{f_{11}}{f_{10}+f_{01}+f_{11}} \quad (1)$$

其中 f_{10} 為出現在文章 x，但不出現在文章 y 的關鍵字詞數量；f_{01} 為出現在文章 y，但不出現在文章 x 的關鍵字詞數量；f_{11} 為共同出現在文章 x、y 的關鍵字詞數量。

依據上述定義，兩篇學術文章可依七個屬性分別計算其相似度（Jaccard coefficient）。此外，藉由 AHP 所得到的各個文章屬性權重，本研究利用加權方式調整各個屬性之相似度，並以加總方式計算兩文章之間的整體相似度。舉例來說，令$S_{u,x}^{TI}$，$S_{u,x}^{AB}$，$S_{u,x}^{KY}$，$S_{u,x}^{T}$，$S_{u,x}^{TA}$，$S_{u,x}^{R}$，$S_{u,x}^{RA}$依序分別為目標使用者 u 整體喜好與文章 x 之標題相似度、摘要相似度、關鍵字相似度、參考文獻標題相似度、參考文獻內容相似度、文章標籤相似度與文章標籤內容相似度，a_1 至 a_7 為 AHP 所計算出之文

章屬性權重值，則兩者之總相似度 $S_{u,x}$ 為：

$$S_{u,x} = a_1 \cdot S_{u,x}^{TI} + a_2 \cdot S_{u,x}^{AB} + a_3 \cdot S_{u,x}^{KY} + a_4 \cdot S_{u,x}^{T} + a_5 \cdot S_{u,x}^{TA} + a_6 \cdot S_{u,x}^{R} + a_7 \cdot S_{u,x}^{RA} \quad (2)$$

本研究最後則依據各文章與目標使用者喜好之總相似度值高低來進行排序，並推薦前 k 篇最相似文章給予目標使用者。

肆、實驗設計與評估指標

一、實驗設計

本研究建構一個能讓受測者進行推薦系統測驗的網路平台，其目的在於收集受測者對於文章的喜好排序，並利用此喜好排序驗證各個推薦方法的推薦效果優劣。實驗系統架設於 Microsoft Windows 7 作業系統上，並採用 Apache 作為網站伺服器，以 PHP 作為網頁開發語言，後端資料庫軟體使用 MySQL 5.0。實驗受測者只需透過任意網頁瀏覽器進行操作後即可完成實驗。

在實驗中，本研究將會建構六種不同屬性組合的學術文章推薦方法，並與本研究所提出之 CRSEI 方法進行比較。此外六種考量不同屬性組合的推薦方法皆考慮 AHP 專家問卷所產生的權重數值，各推薦方法所使用之文章屬性與其權重如表 6 所示。

表 6：各推薦方法中所使用之文章屬性及其屬性權重

方法	文章屬性						
	TI	AB	KY	R	RA	T	TA
F1	0.3641	0.6359					
F2	0.3418	0.4882	0.1700				
F3	0.3230	0.4329	0.1675			0.0766	
F4	0.2783	0.3917	0.1481	0.1166		0.0653	
F5	0.2757	0.3558	0.1386	0.1043	0.0677	0.0579	
CRSEI	0.2573	0.3256	0.1469	0.0942	0.0589	0.0586	0.0588

本研究之實驗區分為三個階段進行。第一階段為文章選取階段，首先系統隨機選取 20 篇文章讓受測者瀏覽，受測者依據系統所提供之文章標題、摘要與關鍵字來從中選取至少一篇的喜好文章，受測者也可以針對所選擇的文章，自行給予喜好的標籤。在第一階段的受測過程中，為確保受測者有充足時間且能夠確實觀看文章，系統設定每篇文章至少需停留 90 秒，方能繼續閱讀下一篇文章。被受測

者選取的文章及訂定的相關標籤，經過資料前處理後，將會轉換為受測者個人特徵屬性集合（即圖 3 的 user profile）。爾後系統會將受測者個人特徵屬性集合與排除受測者已瀏覽文章後的所有文章一一進行相似度比對，並使用六種推薦方法分別找出前 k 篇最高相似度文章。在第二階段中，系統利用第一階段六種方法所推薦的文章進行隨機混合排序以建立文章推薦清單，並將文章標題、摘要與關鍵字呈現在網頁上，並設定每篇文章至少需停留 60 秒，而受測者則需於網頁中勾選符合其所需求或喜好之推薦文章。由於在評估階段已經將受測者第一階段閱讀過的文章排除，因此相關文章不會出現在第二階段的文章推薦清單中；此外，不同推薦方法產生相同推薦文章時，系統會自動排除重複之文章。在實驗的第三階段中，受測者需針對前階段選取之推薦文章進行個人的喜好排序，而系統亦設定受測者至少需停留 30 秒方能送出排序結果。

總和來說，系統在第二階段結束時，將可記錄個別受測者對於六種推薦方法所推薦之文章的喜好程度；而在第三階段結束時，則可進一步瞭解個別受測者對於選取之推薦文章的喜好順序。

二、評估指標

在本研究中，我們使用了成對比較法（Pair Match）以及命中率（Hit Rate）兩種評估方法來驗證各個推薦方法的推薦效果，以下為兩種評估方法之介紹：

（一）成對比較法（Pair Match）

受測者對於系統所推薦之學術文章進行喜好程度的排序，之後將利用受測者的文章喜好排序與推薦方法所評估之喜好排序分別產生各自的項目對（item pair），之後再進行兩個項目對集合的比對計算。有別於傳統的相似度計算方法比較受測者對於喜好整體排序的差異，成對比較法是比較項目對的相對排序差異。如果受測者與推薦方法擁有越多相同的項目對，則兩者的相似度就越高，反之則越低，成對比較法計算如下：

$$PM(u_i, F_a) = \frac{Pair(u_i) \cap Pair(F_a)}{Pair(u_i)} \tag{3}$$

其中 $Pair(u_i)$代表受測者 u_i 對於文章喜好排序產生的項目對；$Pair(F_a)$代表推薦方法 F_a 對於文章推薦排序產生的項目對。

舉例來說，假設受測者 u_i 對於文章的喜好排序為$(d_1, d_2, d_3, d_4, d_5)$，推薦方法 F_a 的文章推薦排序為$(d_5, d_2, d_1, d_4, d_3)$。則受測者 u_i 喜好排序產生的項目對為{(d_1, d_2), (d_1, d_3), (d_1, d_4), (d_1, d_5), (d_2, d_3), (d_2, d_4), (d_2, d_5), (d_3, d_4), (d_3, d_5), (d_4, d_5)}；推薦方法 F_a 的推薦排序產生的項目對為{(d_5, d_2), (d_5, d_1), (d_5, d_4), (d_5, d_3), (d_2, d_1), (d_2, d_4),

(d_2, d_3), (d_1, d_4), (d_1, d_3), (d_4, d_3)}；兩者共同項目對為{(d_1, d_3), (d_1, d_4) , (d_2, d_3), (d_2, d_4)}；$PM(u_i, F_a) = \frac{4}{10} = 0.4$。

在實驗中，本研究收集了多位受測者對於學術文章的喜好排序，故以平均成對比較方法（Mean Pair Match, MPM）做為比較各推薦方法之優劣指標。平均成對比較方法計算公式如下：

$$MPM(F_a) = \frac{\sum_{i=1}^{N} PM(u_i, F_a)}{N} \tag{4}$$

其中 $PM(u_i, F_a)$為受測者 u_i 與推薦方法 F_a 之成對比較值；N 為受測者總數。

（二）命中率（Hit Rate）

Hit Rate 計算方式為以受測者所喜好的文章為基準，推薦方法所推薦文章命中之數量百分比。如果受測者於第二階段實驗中所選取之喜好的文章與推薦方法所推薦的文章重複性越高，則 Hit Rate 就越高，Hit Rate 公式如下：

$$HR(u_i, F_a) = \frac{Prefer(u_i) \cap Recommend(F_a)}{Prefer(u_i)} \tag{5}$$

其中 $Prefer(u_i)$代表受測者 u_i 於第二階段實驗中所選取之喜好文章；$Recommend(F_a)$代表推薦方法 F_a 所推薦之所有文章。

伍、實驗評估

本研究依三個不同的學術文章資料集分別邀請不同受測者進行實驗分析，其中電子商務類別受測者人數為 33 人、推薦系統類別受測者人數為 31 人以及混合兩種分類之受測者人數為 30 人。此外，本研究以各個推薦方法所產生之推薦文章排序中的前三、六、九篇文章（以下分別以 TOP3、TOP6、以及 TOP9 表示），來評估其推薦效能。

一、電子商務分類資料集

圖 3 為利用受測者所喜好之 TOP9、TOP6 以及 TOP3 等電子商務類文章進行 Pair Match 評估結果。由圖 3 可知，推薦的文章數目越多，Pair Match 值越高，在只考量 TI、AB 與 KY 的情況下，Pair Match 值普遍偏低（即方法 F1 與 F2）；比較 F2 與 F3 的結果，我們可以發現若單純只加入考量標籤屬性（T），對於 Pair Match 的提升相當有限；除 F5 方法在 TOP3 實驗中表現不佳外，基本上可以看出考量 RA

或 TA 之方法，即 F5 與 CRSEI 皆能有效提升 Pair Match 值。

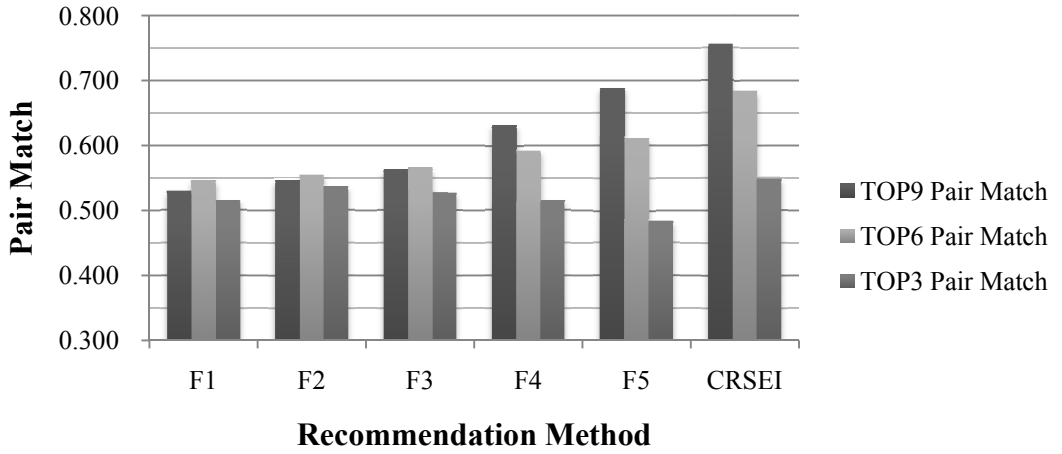

圖 3：電子商務類 Pair Match 實驗結果

圖 4 為利用受測者所喜好之 TOP9、TOP6 以及 TOP3 等電子商務類文章進行 Hit Rate 評估結果。由圖 4 可知，推薦的文章數目越多，Hit Rate 值越高，且差異非常顯著；在 TOP3 實驗中，所有方法的表現均不佳，Hit Rate 值大約只有 0.1；然而在 TOP6 與 TOP9 實驗中發現考量 RA 或 TA 之方法可提升 Hit Rate 約 5%。

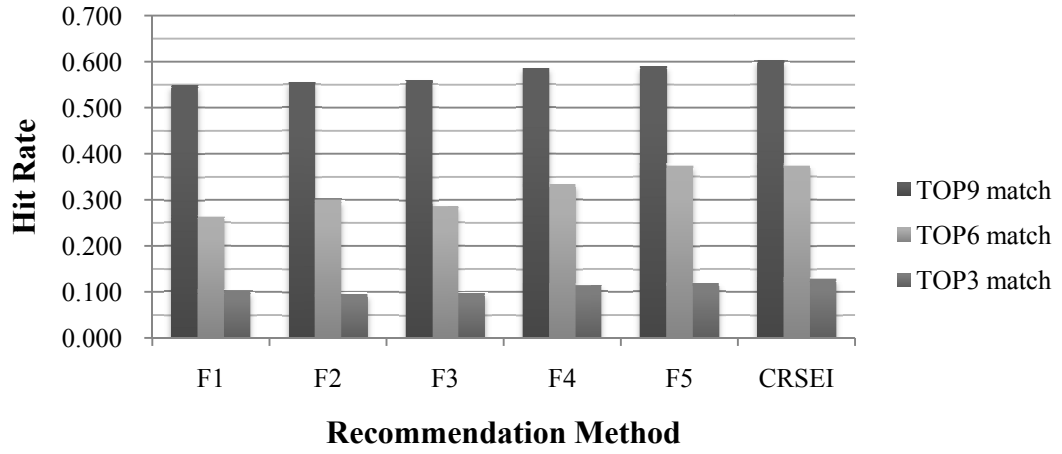

圖 4：電子商務類 Hit Rate 實驗結果

二、推薦系統分類資料集

圖 5 為利用受測者所喜好之 TOP9、TOP6 以及 TOP3 等推薦系統類文章進行 Pair Match 評估結果。在 TOP6 與 TOP9 的實驗中，結果與電子商務類結果近似，F5 與 CRSEI 表現最佳，其中 CRSEI 的 Pair Match 值超過 0.8，再次說明考量 RA 或 TA 屬性能有效提升推薦效能；在 TOP3 的實驗中，CRSEI 仍保持最佳的推薦效能，但 F5 的 Pair Match 值卻下降到接近其他四種方法（F1 至 F4）。

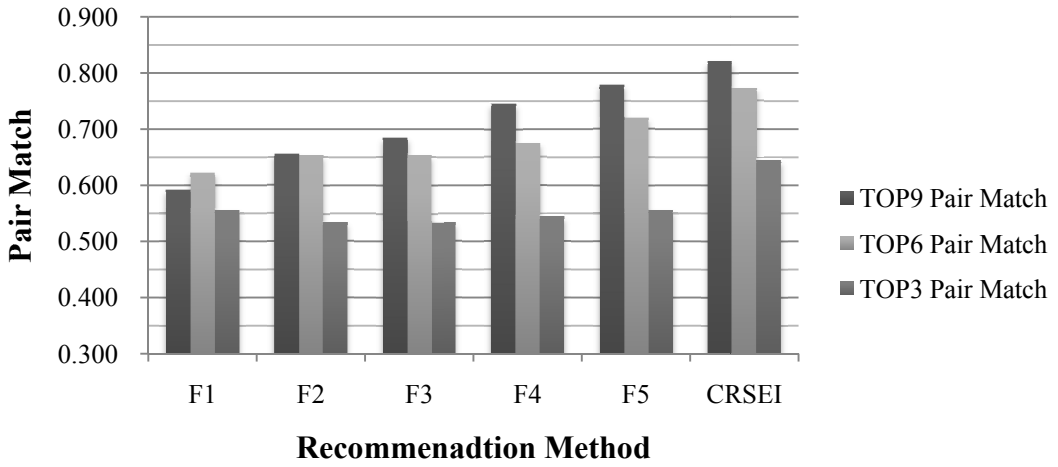

圖 5：推薦系統類 Pair Match 實驗結果

圖 6 為利用受測者所喜好之 TOP9、TOP6 以及 TOP3 等推薦系統類文章進行 Hit Rate 評估結果。大致來說 CRSEI 仍然是整體最佳的方法，但是在 TOP6 的實驗中，F4 獲得與 F5 方法與 CRSEI 不相上下之 Hit Rate 值。僅管如此，由於 Hit Rate 只考慮是否命中，而未考慮文章推薦的先後順序，在兩者同時考量的情況下，仍是以本研究提出的 CRSEI 方法具有較佳之推薦效能。

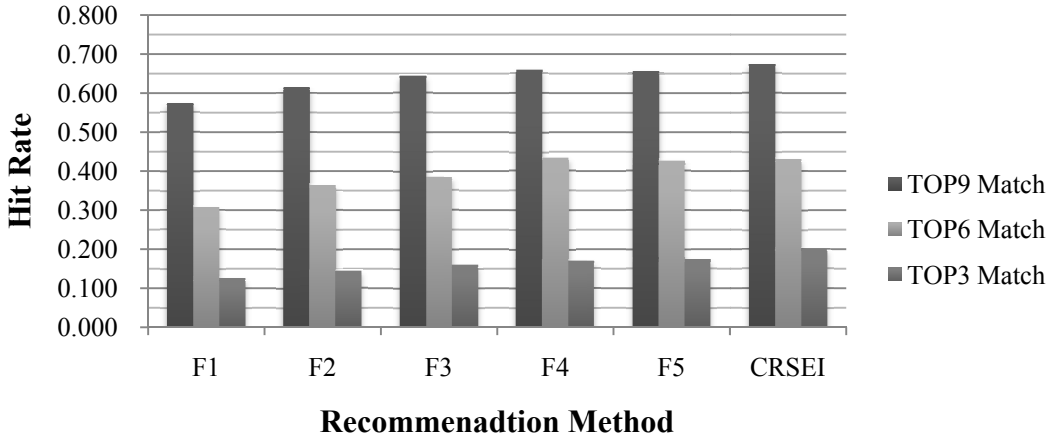

圖 6：推薦系統類 Hit Rate 實驗結果

三、混合式資料集

圖 7 為利用受測者所喜好之 TOP9、TOP6 以及 TOP3 等混合式資料集文章進行 Pair Match 評估結果。在學術文章異質性較高的情況下，實驗結果亦與之前的兩個資料集結果相符：方法 F1、F2、F3、與 F4 相較於 F5 與 CRSEI 獲得較低的 Pair Match 值。圖 8 為 Hit Rate 的實驗結果。同樣地，F4 方法在 TOP9 的實驗中獲得與 F5 與 CRSEI 不相上下的結果，但在同時評估推薦順序的情況下，仍是以本研究提出的 CRSEI 方法具有相對較佳之推薦效能。

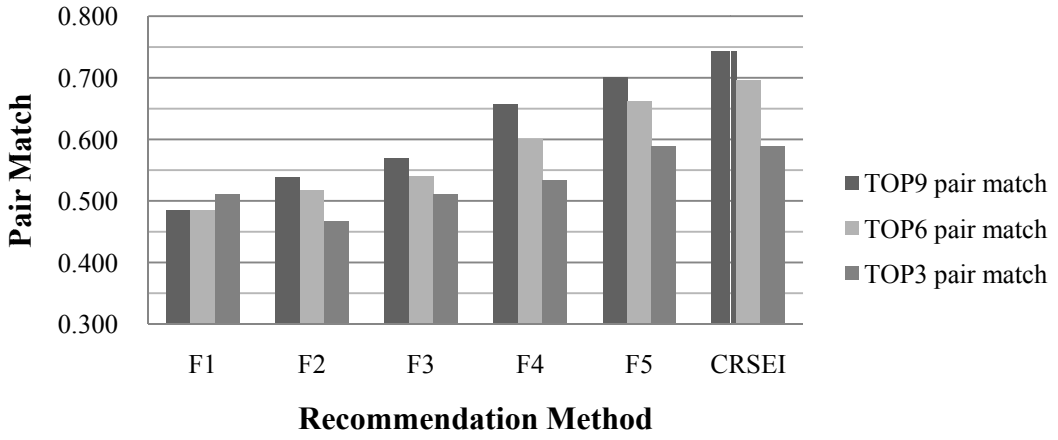

圖 7：混合式資料集 Pair Match 實驗結果

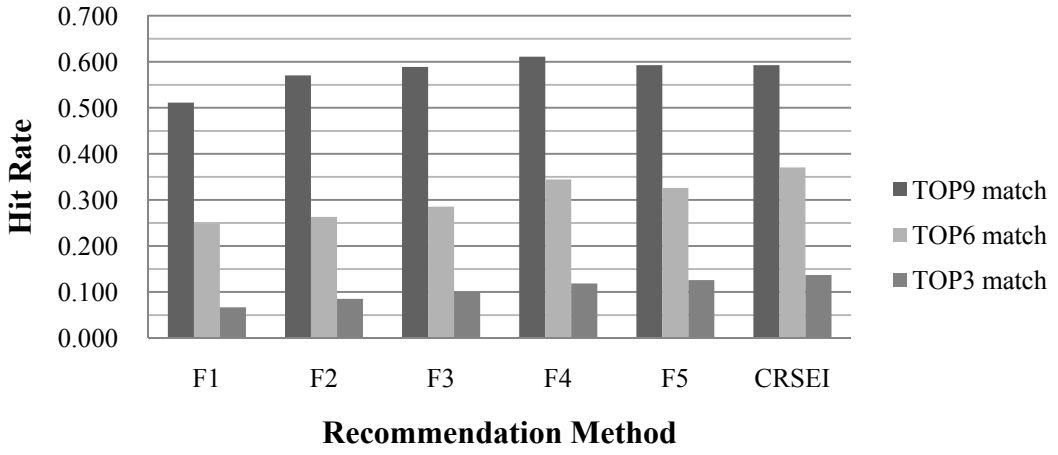

圖 8：混合式資料集 Hit Rate 實驗結果

四、討論

本研究進一步以成對樣本 *t* 檢定來檢驗兩兩方法間之推薦成效差異。本研究方別檢測在 TOP9、TOP6、以及 TOP3 實驗情境下，各方法所獲得之 Pair Match 與 Hit Rate 是否有顯著差異。表 7 為針對 Pair Match 與 Hit Rate 檢定之結果；其中數值格的左邊為 Pair Match，右邊則為 Hit Rate 之檢定結果。檢定結果顯示本研究所提出的 CRSEI 推薦方法所得之 Pair Match 值，在所有實驗情境下皆顯著優於其他推薦方法；而 Hit Rate 部分，除 TOP9 的實驗中 CRSEI 僅顯著優於 F1 與 F2 推薦方法外，其他實驗情境中，CRSEI 大致上顯著優於其他推薦方法。基本上，受測者對於推薦方法所推薦之文章選取的數量越多，則其 Hit Rate 越高；反之則越低。另一方面，被選取的推薦文章若受測者的喜好程度越高，亦即具有較高的喜好排序，則會獲得較高之 Pair Match 值；反之則越低。因此從檢定的結果可以知道，當推薦的文章較少時（如 TOP3），CRSEI 推薦方法所獲得的 Hit Rate 顯著優於其他方法，表示 CRSEI 即使在推薦數量較少的情況下，其推薦之文章依然能夠滿足使用者的喜好。隨著推薦文章數量的增加（如 TOP9），所有推薦方法所獲得之 Hit Rate 亦較先前增長，然而彼此之間卻不存在顯著差異。此現象可能表示各推薦方法在後續增加的推薦文章中，皆能先後滿足使用者的文章喜好。然而從推薦順序的角度來看，本研究提出之 CRSEI 方法無論推薦數量在多或少的情況下，其獲得之 Pair Match 值皆顯著優於其他方法，表示其推薦排序較前之文章能同時獲得使用者給予較高的喜好排序。

表 7：不同推薦方法間成對樣本 *t* 檢定結果（Pair Match/Hit Rate）

	F2	F3	F4	F5	CRSEI
TOP9					
F1	0.00***/0.015*	0.00***/0.00***	0.00***/0.00***	0.00***/0.00***	0.00***/0.00***
F2		0.00***/0.07	0.00***/0.00***	0.00***/0.037*	0.00***/0.00***
F3			0.00***/0.046**	0.00***/0.296	0.00***/0.105
F4				0.00***/0.621	0.00***/0.703
F5					0.00***/0.235
TOP6					
F1	0.095/0.00***	0.031*/0.00***	0.00***/0.00***	0.00***/0.00***	0.00***/0.00***
F2		0.170/0.208	0.00***/0.00***	0.00***/0.00***	0.00***/0.00***
F3			0.013*/0.00***	0.00***/0.00***	0.00***/0.00***
F4				0.00***/0.511	0.00***/0.044*
F5					0.00***/0.068
TOP3					
F1	0.589/0.184	0.903/0.00***	0.921/0.00***	0.712/0.00***	0.137/0.00***
F2		0.516/0.012*	0.495/0.00***	0.417/0.00***	0.034*/0.00***
F3			0.734/0.019*	0.586/0.024*	0.046*/0.00***
F4				0.683/0.453	0.034*/0.00***
F5					0.026*/0.012*

*** p< 0.001; ** p < 0.01; * p < 0.05

總結上述的實驗結果，本研究所提出之 CRSEI 推薦方法，亦即利用參考文獻文章（RA）以及具相同標籤文章（TA）等內容來擴充評估推薦文章時之關鍵字集，其無論在推薦命中率或是推薦喜好排序上，皆優於其他推薦方法。不論是在不同的資料集，或是不同推薦數量的實驗之中，本研究之推薦方法基本上皆顯著優於其他利用不同屬性資料進行推薦評估之推薦方法，這說明本研究所探討之外部資訊，在某種程度上能夠有助於提升推薦系統之推薦效能。

陸、結論

在學術領域中，使用期刊文獻資料庫進行學術文章搜尋一向是學者專家所仰賴的工具，然而卻往往耗費更多時間針對搜尋結果進行資料過濾。傳統以內文為基礎之學術文章推薦方法主要考量使用者喜好文章之內部資訊，以做為評估文章

是否值得推薦之基礎。此方式雖能找出與使用者喜好相似之文章，但對從事學術研究的學者來說，應該同時存在與研究偏好相關之其他學術文章的需求。因此本研究利用文章外部資訊，亦即將參考文獻內容以及相同標籤的文章內容納入評估文章是否值得推薦之資訊，試圖改善學術文章推薦效能。實驗結果顯示，本研究所提出的推薦方法，優於傳統只考慮文章內部資訊的推薦方法，說明本研究提出的方法具有提升傳統以內文為基礎之學術文章推薦效能。在實驗所採用的三個資料集合中，皆反應出本研究提出的推薦方法不僅在推薦命中率上有所提升，更顯著地改善文章推薦排序，將使用者喜好程度較高之文章給予優先的推薦順位。

儘管實驗結果顯示本研究提出之方法能改善學術文章推薦效能，本研究仍存在一些研究限制有待未來研究改善。首先，受限於受測資料來源，亦即電子期刊代理商對於每日下載文章數量之管控，導致本研究在短時間內能夠蒐集之學術文章數量有限。雖然目前的資料量已能滿足實驗之基本需求，但未來若能持續增加文章數量，提高涵蓋率，將更能提高實驗結果之可信度。其次，參考文獻文章內容之收集上，礙於目前資料整合以及取得的問題，本研究僅蒐集並採用參考文獻中之期刊資料，其餘參考文獻，如會議論文、書籍、報導、或是網站等資料，均未納入本研究中考量。再者，本研究為學術文章推薦，因而受測者必須具備相關領域知識及研究需求，致使增加尋找受測者之難度。最後，使用者喜好可能會隨時間經過而改變，然而本研究僅考量特定時間點上使用者的喜好來進行推薦評估，而未將使用者喜好之變動納入推薦評估。

本研究期待未來能朝向以下方向進行進一步探討。首先，本研究僅蒐集資管領域中之推薦系統與電子商務主題進行實驗評估。雖然實驗結果顯示本研究的方法在上述資料中皆能提升推薦效能，但是否能外推至其他研究領域，仍有待進一步的研究。其次，本研究雖同時考慮使用者喜好文章之內部資訊與外部資訊，但仍是在關鍵字詞的基礎上進行推薦文章評估。事實上，除了關鍵字詞之外，仍有許多資訊可用於提升推薦系統效能，例如作者關係、引用關係、發表期刊名稱、及發表時間等相關資訊，本研究希望未來能進一步將上述資訊納入推薦文章評估考量。最後，使用者喜好可能會隨時間經過而改變，而導致用來評估推薦文章之屬性內容改變。因此，未來研究應考慮如何因應使用者喜好變動，或是如何將喜好變動資訊納入推薦系統評估與考量。

誌謝

感謝審查委員提供許多的寶貴建議，使本論文之內容更臻完美；本研究承蒙科技部專題研究計畫經費補助，計畫編號為 NSC 102-2410-H-194-104-MY2，謹致謝忱。

參考文獻

Ahlgren, P. and Colliander, C. (2009), 'Document document similarity approaches and science mapping: Experimental comparison of five approaches', *Journal of Informetrics*', Vol. 3, No. 1, pp. 49-63.

Basilico, J. and Hofmann, T. (2004), 'Unifying collaborative and content-based filtering, in: C. Brodley (Eds.), *Proceedings of the 21st International Conference on Machine Learning*, ACM Press, New York, pp. 65-72.

Cao, Y. and Li, Y. (2007), 'An intelligent fuzzy-based recommendation system for consumer electronic products', *Expert Systems with Applications*, Vol. 33, No. 1, pp. 230-240.

Chen, Y.L., Wei, J.J., Wu, S.Y. and Hu, Y.H. (2006), 'A similarity-based method for retrieving documents from the SCI/SSCI database', *Journal of Information Science*, Vol. 32, No. 5, pp. 449-464.

Chi, E.H. and Mytkowicz, T. (2007), 'Understanding navigability of social tagging systems', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, California, USA, April 28-May 3.

Chiu, P.H., Kao, G.Y.M. and Lo, C.C. (2010), 'Personalized blog content recommender system for mobile phone users', *International Journal of Human-Computer Studies*, Vol. 68, No. 8, pp. 496-507.

Choi, S.H., and Ahn, B.S. (2011), 'Rank order-based recommendation approach for multiple featured products', *Expert Systems with Applications*, Vol. 38, No. 6, pp. 7081-7087.

Choi, S.H., Kang, S. and Jeon, Y.J. (2006), 'Personalized recommendation system based on product specification values', *Expert Systems with Applications*, Vol. 31, No. 3, pp. 607-616.

Choochaiwattana, W. (2010), 'Usage of tagging for research paper recommendation', *The 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE 2010)*, Chengdu, China, Auguest 20-22, pp. 439-442.

Fouss, F., Pirotte, A., Renders, J.M. and Saerens, M. (2007), 'Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19, No. 3, pp. 355-369.

Gajendran, V.K., Lin, J.R. and Fyhrie, D.P. (2007), 'An application of bioinformatics and text mining to the discovery of novel genes related to bone biology', *Bone*,

Vol. 40, No. 5, pp. 1378-1388.

Gemmell, J., Schimoler, T., Mobasher, B. and Burke, R. (2012), 'Resource recommendation in social annotation systems: A linear-weighted hybrid approach', *Journal of Computer and System Sciences*, Vol. 78, No. 4, pp. 1160-1174.

Hwang, S.Y., Wei, C.P. and Liao, Y.F. (2010), 'Coauthorship networks and academic literature recommendation', *Electronic Commerce Research and Applications*, Vol. 9, No. 4, pp. 323-334.

Jinha, A.E. (2010), 'Article 50 million: an estimate of the number of scholarly articles in existence', *Learned Publishing*, Vol. 23, No. 3, pp. 258-263.

Kazienko, P. (2007), 'Filtering of Web recommendation lists using positive and negative usage patterns', *Lecture Notes in Computer Science*, Vol. 4694, pp. 1016-1023.

Kim, H.N., Ji, A.T., Ha, I. and Jo, G.S. (2010), 'Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation', *Electronic Commerce Research and Applications*, Vol. 9, No. 1, pp. 73-83.

Lai, C.H. and Liu, D.R. (2009), 'Integrating knowledge flow mining and collaborative filtering to support document recommendation', *Journal of Systems and Software*, Vol. 82, No. 12, pp. 2023-2037.

Lavie, T., Sela, M., Oppenheim, I., Inbar, O. and Meyer, J. (2010), 'User attitudes towards news content personalization', *International Journal of Human-Computer Studies*, Vol. 68, No. 8, pp. 483-495.

Li, Y.M. and Chen, C.W. (2009), 'A synthetical approach for blog recommendation: Combining trust, social relation, and semantic analysis', *Expert Systems with Applications*, Vol. 36, No. 3, pp. 6536-6547.

Lin, S.H., Shih, C.S., Chen, M.C., Ho, J.M., Ko, M.T. and Huang, Y.M. (1998), 'Extracting classification knowledge of internet documents with mining term associations: A semantic approach', *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, Melbourne, Australia, August 24-28, pp. 241-249.

Liu, D.R., Lai, C.H. and Chiu, H. (2011), 'Sequence-based trust in collaborative filtering for document recommendation', *International Journal of Human-Computer Studies*, Vol. 69, No. 9, pp. 587-601.

Liu, D.R., Lai, C.H. and Huang, C.W. (2008), 'Document recommendation for knowledge sharing in personal folder environments', *Journal of Systems and Software*, Vol. 81, No. 8, pp. 1377-1388.

Liu, D.R., Lai, C.H. and Lee, W.J. (2009), 'A hybrid of sequential rules and

collaborative filtering for product recommendation', *Information Sciences*, Vol. 179, No. 20, pp. 3505-3519.

Liu, D.R., Tsai, P.Y. and Chiu, P.H. (2011), 'Personalized recommendation of popular blog articles for mobile applications', *Information Sciences*, Vol. 181, No. 9, pp. 1552-1572.

Martín-Vicente, M.I., Gil-Solla, A., Ramos-Cabrer, M., Blanco-Fernández, Y. and López-Nores, M. (2012), 'Semantic inference of usr's reputation and expertise to improve collaborative recommendations', *Expert Systems with Applications*, Vol. 39, No. 9, pp. 8248-8258.

Passant, A. and Laublet, P. (2008), 'Meaning of atag: A collaborative approach to bridge the gap between tagging and linked data', *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008)*, Beijing, China, 22 April.

Pazzani, M. and Billsus, D. (1997), 'Learning and revising user profiles: The identification of interesting web sites', *Machine Learning*, Vol. 27, No. 3, pp. 313-331.

Semeraro, G., Lops, P., Basile, P. and de Gemmis, M. (2009), 'Knowledge infusion into content-based recommender systems', *Proceedings of the Third ACM Conference on Recommender Systems (RecSys 2009)*, New York, NY, USA, 23-25 October, pp. 301-304.

Senecal, S. and Nantel, J. (2004), 'The influence of online product recommendations on consumers' online choices', *Journal of Retailing*, Vol. 80, No. 2, pp. 159-169.

Ur-Rahman, N. and Harding, J.A. (2012), 'Textual data mining for industrial knowledge management and text classification: A business oriented approach', *Expert Systems with Applications*, Vol. 39, No. 5, pp. 4729-4739.

Wen, H., Fang, L. and Guan, L. (2012), 'A hybrid approach for personalized recommendation of news on the Web', *Expert Systems with Applications*, Vol. 39, No. 5, pp. 5806-5814.

Weng, S.S., Lin, B. and Chen, W.T. (2009), 'Using contextual information and multidimensional approach for recommendation', *Expert Systems with Applications*, Vol. 36, No. 2, pp. 1268-1279.

Xu, Y., Guo, X., Hao, J., Ma, J., Lau, R.Y. and Xu, W. (2012), 'Combining social network and semantic concept analysis for personalized academic researcher recommendation', *Decision Support Systems*, Vol. 54, No. 1, pp. 564-573.

Yang, H.C. and Lee, C.H. (2004), 'A text mining approach on automatic generation of web directories and hierarchies', *Expert Systems with Applications*, Vol. 27, No. 4,

pp. 645-663.

Zhang, Y. and Koren, J. (2007), 'Efficient bayesian hierarchical user modeling for recommendation systems', *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007)*, Amsterdam, Netherlands, 23-27 July, pp. 47-54.

Zheng, N. and Li, Q. (2011), 'A recommender system based on tag and time information for social tagging systems', *Expert Systems with Applications*, Vol. 38, No. 4, pp. 4575-4587.