

教育資料與圖書館學

Journal of Educational Media & Library Sciences

<http://joemls.tku.edu.tw>

Vol. 49 , no. 3 (Spring 2012) : 405-445

評估以MeSH做為PubMed資料庫搜尋之建議詞彙的
有效性檢索行為研究

A User Study of the Effectiveness of MAP

(Multiple-Access to PubMed):

A MeSH Based Query Suggestion Tool

唐 牧 群 Muh-Chyun Tang

Assistant Professor

E-mail: mctang@ntu.edu.tw

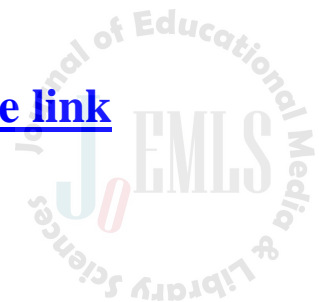
洪 承 理 Cheng-Li Hung

Librarian

E-mail: R96126013@ntu.edu.tw

[English Abstract & Summary see link](#)

[at the end of this article](#)



評估以MeSH做為PubMed資料庫 搜尋建議詞彙的有效性及 檢索行為研究

唐牧群*

助理教授
國立臺灣大學圖書資訊學系
E-mail: mctang@ntu.edu.tw

洪承理

組員
國立清華大學圖書館人社分館
E-mail: R96126013@ntu.edu.tw

摘要

本研究設計了一針對PubMed資料庫的詞彙建議系統介面，並以實驗法比較該介面與PubMed介面。以44名生物科學及醫學系所研究生為研究對象，每人以兩個自己的真實需求作為檢索任務，就使用者查詢行為及檢索結果進行分析。研究結果顯示：(一)主題熟悉度會影響使用者檢索詞的修改，不熟悉主題任務其檢索前後相似度會低於熟悉主題任務。(二)不同介面會影響使用者查詢行為，實驗介面能提供使用者大量建議相關詞彙，因此使用者改變檢索詞的機率增加，而在PubMed介面中使用者查詢詞彙的變化有限。(三)不同主題熟悉度與不同介面對使用者查詢行為產生交互作用，實驗介面在使用者較不熟悉檢索主題時可發揮較大作用。(四)實驗介面能輔助使用者取得在大量檢索結果中排序較低的相關文件。

關鍵詞：互動式資訊檢索，查詢擴展，使用者行為，詞彙建議系統，醫學標題表

前 言

在文件檢索技術演進之下，目前以自然語言進行全文索引已成為主流，使得一般使用者可不需專家協助也能夠獨力進行檢索。然而，在面對如PubMed

*本文主要作者兼通訊作者。

如此大型且內容資料異質性較高的資料庫時，一般使用者往往無法完全利用系統的強大檢索功能，且搜尋結果過於龐大也使得使用者難以選擇真正相關的文件。PubMed的許多預設功能，如擴展(explode)、題名和摘要欄位的全文索引等，其目標都在於協助一般使用者獲得足夠的檢索結果。但這些提升索引詳盡性(exhaustivity)的功能也使得使用者往往必須面對數量龐大的檢索結果，若無有效的工具協助，使用者將無法有系統地縮小檢索範圍，往往只能選擇倉促瀏覽檢索結果的前數十筆資料而已。過多的檢出文件至少造成兩個人機互動的問題。其一是使用者所需的相關文件可能堆積在檢出文件集的深處，而使用者倉促之下瀏覽檢索結果並無法取得利用之。其次，由於使用者只能窺見檢索結果的冰山一角，往往無法針對其初始的關鍵字進行有效性的判斷。

在資訊檢索過程中，檢索詞對於檢索結果有著最直接的影響。如何協助使用者表達他們的資訊需求是此一領域重要的議題(Belkin et al., 1982)。在自動索引的環境中，文件索引之詳盡性相當高，然而一般搜尋引擎使用者的檢索詞卻非常短，約2-3個詞(Jansen, Spink, & Saracevic, 2000)，換言之，使用者檢索問題的詳盡性卻沒有相對提高。簡短的檢索詞彙通常會造成龐大的搜尋結果且缺乏精確性，而此一現象可透過系統提供詞彙建議的方式來建議並引導使用者提出較多的檢索詞彙以獲得改善。

本文所測試的MAP(Multiple Access to PubMed)介面(見附錄一)，即嘗試利用由使用者初始的搜尋結果擷取詮釋資料中與原始檢索詞共同出現的MeSH控制詞彙，並將這些詞彙呈現給使用者，以作為使用者改善其檢索詞的依據，MAP介面建議詞彙主要是由資料庫紀錄中後設資料的MeSH主題詞來完成，並依層面來組織詞彙以協助使用者瀏覽，其輔助使用者探索資訊空間的方法與Hearst et al.(2002)所提出的層面式瀏覽大致相同。

由於MAP介面所提供的詞彙建議，必須由使用者瀏覽辨識選擇後才能發揮功效，因此傳統資訊檢索評估的作法並不適用於該介面的評估。傳統的系統評估是由專家對相關文件判斷並事先建立相關文件集(relevant set)，以批次處理(batch-mode)方式投入系統。本研究提出一套互動式檢索系統評估方法(Tang, 2009; Tang, 2006; 洪承理, 2010)，由使用者自身真實的研究問題來做為測試對象，希望能夠反映該介面在互動檢索環境中的檢索效益。

二、相關研究

眾多詞彙建議的技術都是建立在當使用者面對複雜的檢索問題是無法確切表達其資訊需求的前提下，在使用控制語言(control vocabulary)索引的資料庫，詞彙建議可透過索引典的方式執行。在沒有控制語言的環境下，則可利用詞彙共現頻率來建立自動化索引典。而相關回饋的技術則是透過抽取由使用者

認定相關的文件中抽取關鍵詞，以自動查詢擴展 (automatic query expansion) 或經由互動式查詢擴展 (interactive query expansion)，將抽取的關鍵詞建議給使用者，由使用者決定是否採用 (Koenemann, & Belkin, 1996)。此一技術已被證明可改善檢索結果 (Harman, 1988, 1992)，近來也有研究者嘗試利用像是閱讀時間、捲軸移動以及文件取得等隱含性相關回饋 (implicit relevance feedback) 之指標來作為改善檢索詞品質的依據 (White, Ruthven, & J. M. Jose, 2005)。

而不需要透過使用者相關回饋的詞彙建議方式則有局部內容分析 (local content analysis) (Croft, 1996; Belkin et al., 2000; John & Jose, 2006)，以及由其他使用者在相似搜尋中使用過的詞彙作為推薦的依據 (Huang, Chien, & Oyang, 2003)。其中局部內容分析是事先分析資料庫內概念之間的關係，之後當使用者輸入檢索詞後，再找出與該檢索詞共現的詞彙作為詞彙建議的來源。類似局部內容分析在控制語言環境下的運用則是以詮釋資料作為分析的對象以輔助檢索，透過動態擷取初始檢索結果集之詮釋資料來做詞彙推薦 (Hearst et al., 2002; Lin, 1999; Pollitt, 1998)。此外，在醫藥科學文獻檢索中，亦有由研究透過控制詞彙 (Doms, & Schroeder, 2005) 及文摘摘要全文 (Goetz, & von der Lieth, 2005; Perez-Iratxeta, BORK, & Andrade, 2001; Plikus, Zhang, & Chuong, 2006) 中抽取與原檢索詞的共現詞彙來提供詞彙建議。

接下來作者將過去以索引典作為詞彙建議來源的相關使用者研究做一回顧，Shiri 和 Revie (2006) 將利用索引典建構的資訊檢索系統分為：

1. 標準人工建構索引典：標準索引典中階層、等同、相關的關係已廣泛應用在詞彙選擇及查詢擴展的目的上。這些索引典也以應用於一些專家系統的雛型中以幫助使用者選擇檢索詞及擴展查詢。
2. 搜尋索引典 (search thesaurus)：搜尋索引典又被認為是終端使用者索引典，包含大量的同義詞、類同義詞、變形詞的進入詞 (entry term) 用以輔助使用者找尋替代詞彙加入查詢中。
3. 自動化建構索引典：基於詞彙共現關係所製作之索引典，又稱關聯式索引典 (association thesaurus)，詞彙之關係建立於詞彙與詞彙間同時出現的次數作為彼此關係密切與否之判斷。

Kristensen 和 Jarvelin (1990) 研究小型的搜尋索引典在回收率及精確率的效能，使用者僅參與提供搜尋及回饋的過程。其結論是搜尋索引典可提升全文資料庫的檢索效能。其亦調查了 5 種檢索模式：基本檢索、同義字檢索、狹義詞檢索、相關詞檢索、綜合檢索中回收率及精確率的效率。研究顯示搜尋索引典是加強回收率的工具，而使用者的參與可提升精確率。

Spink (1995) 檢視了不同詞彙選擇的來源檢索相關文獻的效能，包含使用者問題陳述、使用者互動、中介者、索引典、從相關文件抽取。其中索引典為

第三有效的工具，次於問題陳述及中介互動。結果顯示詞彙相關回饋的檢索詞較專業建議檢索詞及選自索引典的檢索詞更能有效檢索到相關的物件。

多數研究顯示使用索引典可幫助使用者選擇替代詞彙來擴展查詢，同時能改善檢索結果。但研究多關注於檢索結果而非檢索過程。此類研究多建立在使用者對建議的詞彙做選擇以改善在初始檢索詞的前提上，因而使用者不可避免的也成為系統評估的一環 (Dumais, & Belkin, 2005)。

Efthimiadis (2000) 認為 IR 領域迫切需要真實而非模擬的資訊需求來調查真實使用者所進行的詞彙選擇和替換。其透過 INSPEC 資料庫調查電機學院教職員生所進行 25 次真實資訊需求的開放式檢索，以問卷、log 檔及相關評估等方法來蒐集資料，其在系統評估上的主要貢獻在於使用由受測者自身的資訊需求，而非傳統的指定 (assigned) 問題的作法，並將使用者的詞彙選擇行為納入研究課題。其問卷分為四個部分：前檢索問卷蒐集受測者背景資料；查詢擴展問卷蒐集受測者所選擇之詞彙及其與原始查詢之詞彙關係；檢索後問卷則蒐集受測者之滿意度、整體檢索印象、查詢與結果之評價、預期檢得相關筆數，以及紙本之檢索結果相關評估；最終問卷包含整體評價及使用者評論。

該研究詞彙選擇的分析主要是記錄系統所產生的建議詞彙數量及受測者選擇其作為檢索詞的數量。受測者必須指出清單上所有適合用於檢索的詞彙及其選擇的原因，包括 1. 變形或同義字，2. 相關詞，3. 妥協詞 (no better term)，4. 新觀念的產生等。另一部分則是要求受測者說明所選擇最適合檢索的五個詞彙與其原始查詢詞彙的關係，如索引典中的階層架構。研究結果顯示：多數詞彙對初始查詢而言多為同義字或相關詞，而部分詞彙則是新的概念。而選擇詞彙與初始詞彙的關係為：34% 無明顯關係，66% 的選擇詞彙與初始詞彙有關聯。這些關係之比例為下位詞 46%、上位詞 3%、相關詞 17%。而檢索效益的評估則是比較初始查詢及擴展後的查詢之間的效益，主要是基於受測者對於檢索結果的相關判斷後進行比較。研究結果顯示，查詢擴展提升了檢索的精確性且提供較多的相關文獻 (Efthimiadis, 2000)。

Shiri 和 Revie (2006) 調查於利用人工索引典輔助的系統中使用使用者查詢擴展互動的行為，其研究對象為獸醫學系的研究者與學生。資料收集自 30 人操作的 90 次透過 OVID 介面檢索 CAB 摘要資料庫的檢索。資料蒐集工具為問卷、螢幕錄影、訪談。其研究結果發現，不同背景受測者所選擇之詞彙有顯著不同。同時搜尋任務的複雜性也會影響使用者與索引典系統的互動，較複雜的任務需要更多的查詢擴展及詞彙選擇。另外，使用者曾否檢索過類似問題的經驗亦會影響其選擇行為及對系統的評價。

Tang (2007) 使用縱貫性研究 (longitudinal study) 研究 19 名生醫領域研究者如何使用一個 MeSH 索引典所設計的瀏覽介面。研究者以遠端方式觀察受測者在其工作場所，以其自己的真實檢索問題使用索引典瀏覽介面的情形，並

輔以問卷方式，蒐集關於檢索問題特性的資料，觀察時間為期10週。結果發現，受測者對於該索引典瀏覽介面的選擇，以及對介面有用性的評估會受到檢索問題特性的影響，索引典介面對於使用者較為不熟悉的問題較為有效(Tang, 2007)。

三、研究問題

本研究的目的是在於了解像MAP這樣一個提供控制詞彙作為詞彙建議的介面，在互動環境中是否對於使用者的搜尋有所幫助。在沒有傳統資訊檢索評估的相關文件集的協助下，本研究必須嘗試使用新的評估指標和方法來比較使用者在搜尋MAP與PubMed原本介面的搜尋成效。此外，本研究也希望了解系統的效能是否會受到使用者檢索問題不同而異。換言之，我們想要了解不同種類的問題，是否於系統功能有交互作用，這是根據於比較複雜或不熟悉的問題可能更需要詞彙推薦幫助的假設。就搜尋成效而言，又可分為兩個面向：

1. 檢索詞彙的品質：比較受測在與介面互動前後檢索詞的品質是否有所不同，關於兩者(與系統互動前後)的比較，可以表現在檢索詞彙的改變，以及受測者主觀對於介面互動後修正檢索詞之適用性(goodness)的判斷。
2. 檢索結果的品質：結果品質又可反映在兩方面，一為使用者對於找到相關文件的評價，二為實驗介面能否幫助使用者更有效率的找到使用傳統介面難以發現的相關文獻。

檢索行為則在觀察受測者在於不同介面互動時，是否產生不同的搜尋行為。搜尋行為的不同可表現在檢索詞的數量、檢索次數以及瀏覽的模式等。以下筆者將詳述本研究為此一互動檢索情境下所設計的研究方法及評估指標。

四、研究方法

(一) 搜尋任務

由於本研究關注於實驗介面能支援與解決何種類型需求情境，如何描述並分類如使用者需求情境這樣一個抽象概念就成為一相當重要的課題。目前以使用者為導向的互動式檢索評估中，研究者都事先精心設計一套模擬的搜尋情境，然後以實驗要求受測者針對這些情境進行檢索。本研究嘗試在傳統模擬情境模式外，探索其他的可能性，亦即讓受測者檢索自己真正(genuine)的問題，而非如一般檢索實驗由實驗者指定(assigned)檢索任務(task)的方式。

本研究將探索以使用者的真實資訊需求而非研究者所指定的檢索任務作為基本分析單位(unit of analysis)。如此設計的原因有二：一為若採用指定檢索任務，在描述任務時將難免給與受測者檢索詞彙的提示，如此一來，受測者或許

會過於仰賴任務描述中已出現的詞彙，而失去對瀏覽式界面的需求。即使已有研究顯示語意開放的任務描述能讓參與者融入假想的需求情境，並足以獲得與真實情境相當的評估結果 (Borlund, & Ingwersen, 1997, 2000)，然在對於醫學如此具備高度專業知識的領域而言，設計具有開放語意的任務描述是有困難的。

第二個採取真實檢索問題的原因，在於本研究欲找出檢索問題與工具間的對應性。在本研究研究者將有機會觀察、紀錄不同檢索工具在不同搜尋情況下的效用。若採取使用研究者所假想的檢索任務模式，則必須將不同的搜尋情境用敘事方式操作化，否則將喪失研究兩者相容性的機會。因此本研究將以受測者的真實資訊需求做為檢索任務，並利用問卷方式來描述受試者檢索問題的屬性。

(二) 實驗設計

本研究以重複量測 (repeated measure) 方式設計實驗，受測者被要求以兩個介面 (PubMed 及 MAP) 檢索問題，即每一檢索問題都要求受測者交互使用兩個介面進行檢索。重複量測可以控制因檢索問題 (request) 不同所產生的變異量 (variance) 並加強統計檢定的力量。但此一設計最明顯的風險在於次序效應會影響實驗結果，即前一介面之檢索結果會影響後一介面之檢索結果。為控制可能發生的次序效應，受測者的檢索需求會被隨機指派到替代的處理順序，而使得每一個檢索需求都有均等機會以任何介面進行第一次檢索。受測者將被隨機指派到四者其中之一進行檢索，然後再轉移到另一介面。

實驗之分組如表 1 所示，共有四組，以介面之差異與主題熟悉度作為控制，受測者將隨機指派至四個類別，先以單一介面檢索不熟悉主題及熟悉主題，而後以另一介面進行不熟悉主題及熟悉主題之檢索，其目的在於比較不同主題熟悉度及不同介面對於使用者的檢索行為及查詢擴展行為所產生的差異。

表 1 實驗分組表

類型	PubMed		MAP	
1	熟悉主題	不熟悉主題	熟悉主題	不熟悉主題
2	不熟悉主題	熟悉主題	不熟悉主題	熟悉主題
	MAP		PubMed	
3	熟悉主題	不熟悉主題	熟悉主題	不熟悉主題
4	不熟悉主題	熟悉主題	不熟悉主題	熟悉主題

(三) 研究工具

測試介面

為避免動態的計算並呈現詞彙共現資訊的困難，我們預先下載 2006 到 2007 年 PubMed 所有的書目紀錄，總計包含約 120 萬筆書目紀錄，以及 1 萬 9 千餘期刊，下載作者人次達 153 萬，並由此事先建立 1 萬 6 千多個 MeSH 詞彙以及詞彙與其他書目資料欄位的共現資料詞。介面處理使用者檢索詞的具體作法如下¹：

¹ 實際介面網址為：<http://morris.lis.ntu.edu.tw/map/new/medlineFre2.php>

1. 當使用者使用MAP輸入檢索詞，系統會將檢索詞投入PubMed搜尋列。
2. 此時，系統會利用PubMed本身的MeSH Translation Table功能，找出與使用者原檢索詞對應的MeSH詞。例如若使用者輸入「lung cancer」，MeSH Translation Table會將它轉譯為「lung neoplasm」。
3. 在取得適當MeSH詞彙之後，系統將會從預先建立的詞彙共現資料庫找出前200個與該詞彙經常出現的其他MeSH詞彙作為參考建議詞。在沒有合適MeSH敘述詞的情況下，將會啟動從MEDLINE檢索結果中擷取前200個敘述詞的機制作為建議詞彙的來源。
4. 這200個建議詞的排序挑選則是由其與原檢索詞共現(co-occurrence)的頻率以及其IDF(Inverse Document Frequency)來決定。
5. 接著介面會根據MeSH最上層分類的呈現方式，將它們呈現給使用者。使用者可決定增加或捨去這些詞彙。
6. 當使用者改變其檢索詞，新的推薦詞彙也將動態的更新並呈現給使用者。

問卷包含背景資料問卷、搜尋前問卷及搜尋後問卷。背景資料問卷詢問受測者的背景資料，如職業、性別、專業領域、使用PubMed資料庫及Google Scholar頻率、對於MeSH熟悉度、使用進階檢索指令頻率等。搜尋前問卷則是詢問受測者欲檢索之問題及對該議題之熟悉度等。搜尋後問卷則詢問受測者對於檢索結果、介面滿意度，及介面對於受測者的影響，包含最佳檢索詞彙、熟悉程度等。問卷內容詳如附錄二。

(四) 研究對象

由於PubMed資料庫屬於醫學資料庫，使用者多具有醫學相關專業知識且基於特定專業資訊需求而檢索該資料庫，故本研究之研究對象為生命科學及醫學相關系所專業人士而非一般普通使用者。而研究對象之取得則是於臺灣大學電子佈告欄系統Ptt及校內訊息公佈欄網頁張貼告示，公開徵求生命科學及醫學相關系所之研究生，並以新臺幣500元做為參與實驗之報酬。受試者在參與實驗之前都被告知須事先準備一個較熟悉，一個較不熟悉的檢索問題。

(五) 實驗步驟

當受測者抵達時，會先徵得其同意實驗，同時要求其填寫一問卷收集其專業、教育背景及PubMed/Medline使用經驗。此外，實驗者也以影片教學方式來解釋如何操作介面以讓使用者熟悉介面。

在每個檢索開始前，要求使用者寫下搜尋前問卷說明其欲檢索的任務。同時寫下其所認為最適合的查詢者。每一受測者會用兩個介面交互檢索兩個查詢，因此會有四個階段。當每一檢索階段開始時，受測者輸入其在搜尋前問卷

所寫下的查詢詞，並要求其用指派的介面在該階段檢索10筆有用的紀錄。在輸入初始的查詢詞之後，受測者可透過介面的建議詞彙（在MAP搜尋時）及檢索結果來修改其查詢。受測者被告知每一任務均有15分鐘去完成，但可提前結束。

每一階段結束後，受測者會寫下此時其所認為最適合的查詢詞。並以六點尺度要求其評估檢索前及檢索後的理想查詢詞的適用性。另外，受測者也會評估對於檢索結果的滿意度及介面的好用程度。所有受測者與介面的互動均以螢幕錄製軟體所記錄。本研究特別關注於查詢次數、選擇關鍵字數量、檢視紀錄數，此有助於理解使用者在不同系統互動時所呈現的檢索行為。

在同一查詢的兩個介面都完成後，受測者被要求用六點尺度在先前所找到20筆紀錄的相關程度評分（每一介面10筆），使用者對於這些搜尋系統的評分將作為比較兩介面搜尋結果之依據。

具體實驗流程如下：

1. 於檢索開始先請受測者填寫檢索前問卷，記錄其背景資料及所欲檢索的問題及初始之關鍵字，並為初始關鍵字的適用性評分。
2. 檢索開始後使用螢幕記錄軟體錄製受測者的檢索過程，每一檢索均耗時15分鐘。
3. 每一受測者都要搜尋兩個自己感興趣的檢索問題，且每一個檢索問題均透過MAP及PubMed介面進行兩次檢索。
4. 檢索完畢後請受試者填寫檢索後問卷，記錄此刻其認為最適合的關鍵字，評估其適用性，以及其對介面的觀感。
5. 對檢索過程所挑出的相關文件評估。

五、研究結果

(一) 使用者基本資料

本實驗共募集44名受測者，其就讀學院及性別等資料如表2所示，受測者除1人為研究助理外其餘均為研究生。

表2 受測者基本資料表

學 院	性 別		總和
	男	女	
醫學院	3	6	9
公衛學院	4	2	6
生命科學院	3	4	7
理學院	2	2	4
工學院	4	3	7
農學院	6	3	9
其他	1	1	2
總 和	23	21	44



為符合實驗之目的，即測試PubMed資料庫之一般使用者，因此將受測者背景限於相關生醫理工系所，其中之例外分別為電機所及特殊教育研究所，故分於其他類之中。

除受測者就讀系所及性別資料外，實驗前問卷並蒐集受測者利用PubMed及Google Scholar的頻率與對於MeSH醫學索引典的熟悉程度和檢索時利用布林邏輯進行進階檢索的頻率，要求受測者以李克特6點尺度進行選擇，0表示程度最低，5表示程度最高，其結果如表3及圖1。結果顯示受測者在利用PubMed及Google Scholar的頻率大致相近，而對於MeSH醫學索引典及布林邏輯檢索則較不熟悉。由是觀之，可發現一般使用者並不熟悉檢索系統所提供之進階檢索功能，對於相關之資訊組織工具MeSH也缺乏理解，在進行檢索時仍只會以關鍵字組合進行檢索。

表3 受測者檢索背景資料

	平均數	標準差	個數
使用PubMed資料庫之頻率	3.68	0.86	44
使用Google Scholar之頻率	3.39	1.59	44
MeSH醫學索引典之熟悉程度	1.07	1.50	44
使用布林邏輯等進階檢索功能之頻率	2.52	1.58	44

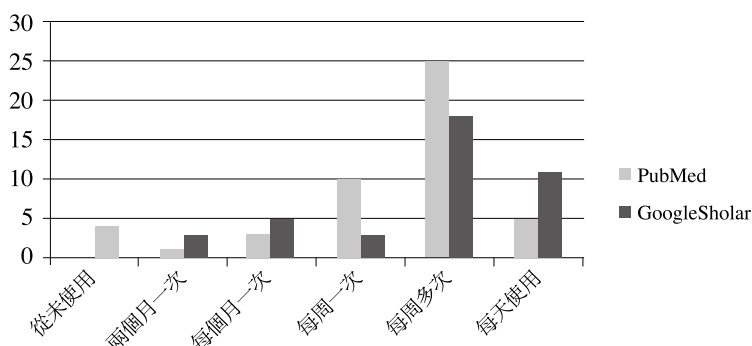


圖1 受試者利用PubMed及Google Scholar頻率圖

本實驗採用使用者之真實資訊需求做為檢索任務，每一使用者提供熟悉與不熟悉各1個檢索任務，共有88個檢索任務，並在PubMed及MAP各進行一次檢索，總計176次檢索。在88個檢索任務中受測者未曾檢索之主題為28個，實驗前已檢索過之主題為60個。受試者之檢索任務詳如附錄三。

在搜尋前問卷部分則要求受測者填寫欲檢索之問題及查詢關鍵字，而後詢問其受測前是否檢索過該主題，並要求受測者以李克特六點尺度，0表示程度最低，5表示程度最高，說明其對檢索主題之熟悉程度、對於檢索結果完整度的要求以及由受測者評估其問題可由他人代為查詢的取代程度。由於缺乏一能夠客觀評估查詢(query)之指標，因此設計查詢「適用性」指標，由受測者於搜

尋後問卷以李克特六點尺度，0表示最低，5表示最高，主觀評估其初始查詢以及與介面互動後最終查詢能夠表達其資訊需求的程度。

(二) 使用者檢索特性分析

對於受測者檢索行為的分析包含初始與最終查詢之間的相似度、最終查詢的詞彙、新增詞彙、移除詞彙之字數以及送出查詢之次數和瀏覽文件數，經由統計分析以了解在初次檢索及再訪檢索和不同介面上其檢索特性有否明顯差異。

查詢相似度指使用者最終查詢與初始查詢之相似程度，係以Jaccard係數 (Jaccard's Coefficient) 進行運算，查詢之相似度為 $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$ ， $A \cap B$ 為初始查詢與最終查詢字彙之間的交集， $A \cup B$ 為初始與最終檢索詞彙的聯集，以初始查詢與最終查詢字彙之聯集比較兩種介面之查詢相似度能說明MAP介面是否較傳統PubMed介面能改變受測者之查詢。最終查詢、新增詞彙、移除詞彙之字數則是受測者在與檢索系統互動後所產生之明顯變化。根據受測者新增及刪除詞彙的數量，可看出受測者受到介面而改變其原始檢索詞彙的影響程度。送出查詢之次數則表示受測者改變其檢索之次數，瀏覽文件數則是指受測者在檢索結果中所點閱的頁面數量。

(三) 不同介面檢索特性比較

不同介面之檢索特性如圖2到圖7，經過獨立樣本平均數檢定後，結果顯示不同介面之查詢相似度平均數有明顯差異 ($t=3.31, p=.001$)，PubMed介面 ($m=.54$) 之相似度高於MAP介面 ($m=.40$)。此結果說明了使用者利用PubMed介面時，由於僅能透過與檢索結果互動來產生新的檢索詞彙，因此其最終查詢與初始查詢的相似度會較高，而在MAP介面時，由於介面提供了大量的相關建議詞彙，故而使用者能從中獲得較多的檢索詞彙並透過檢索驗證其相關性而置入最終查詢詞彙中，因此其相似度會較低。

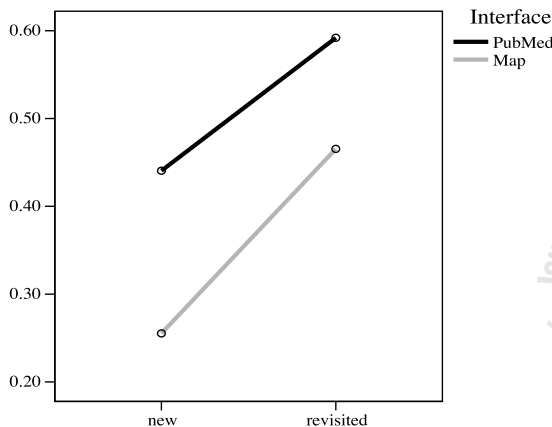


圖2 不同介面及不同熟悉度之相似度平均數圖

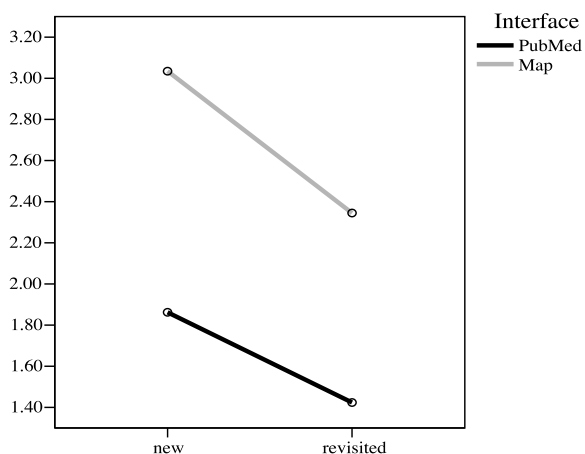


圖3 不同介面及不同熟悉度之新增詞彙數平均數圖

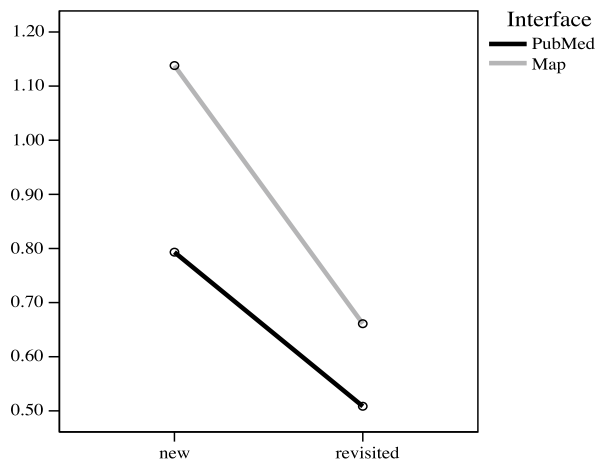


圖4 不同介面及不同熟悉度之移除詞彙數平均數圖

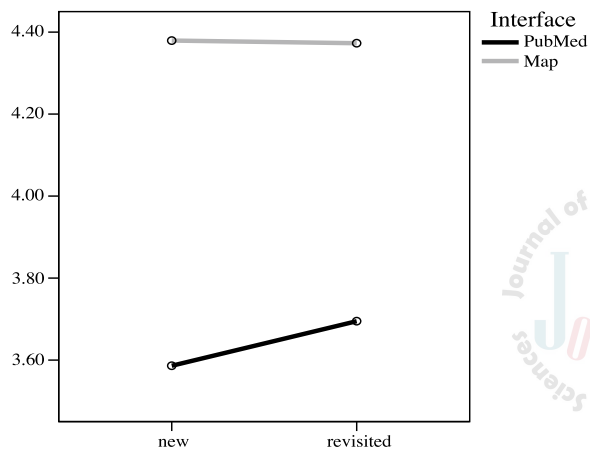


圖5 不同介面及不同熟悉度之最終查詢字數平均數圖



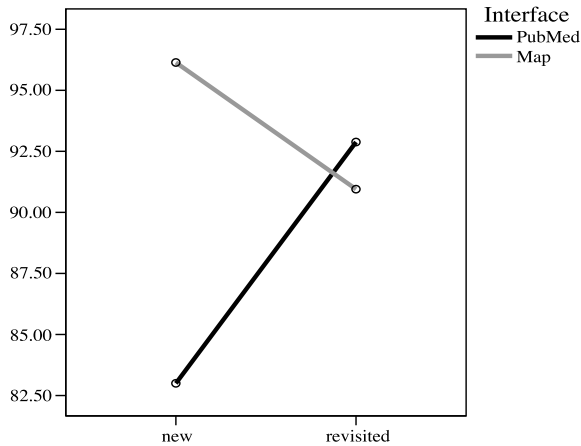


圖6 不同介面及不同熟悉度之瀏覽文件數平均數圖

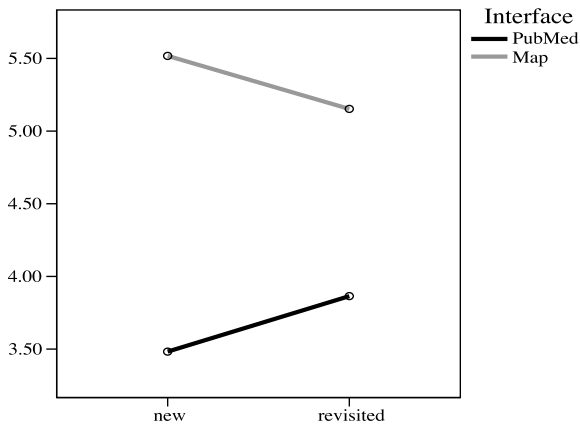


圖7 不同介面及不同熟悉度之查詢次數平均數圖

此外，不同介面之新增詞彙數亦有明顯差異 ($t=-3.71, p=.000$)，MAP介面 ($m=2.58$) 明顯高於PubMed介面 ($m=1.57$)。由於使用者在利用PubMed介面檢索時，其查詢詞彙僅能從與檢索結果互動中所產生；反之，MAP介面能提供大量的相關MeSH詞彙供使用者點選利用，因此兩者之新增查詢詞彙數會有明顯不同。

不同介面在最終查詢字數方面亦有明顯差異 ($t=-2.37, p=.02$)，MAP介面之最終查詢字數 ($m=4.38$) 明顯高於PubMed介面 ($m=3.67$)。這亦可說明透過MAP，使用者經由瀏覽詞彙介面辨識，並選擇之前未想到的詞彙中取用合適的查詢詞，因此其最終查詢詞彙字數會較PubMed介面來的高。

而不同介面在查詢次數具有明顯差異 ($t=-3.68, p=.000$)，MAP介面之查詢次數 ($m=5.27$) 明顯高於PubMed介面 ($m=3.74$)。但於瀏覽文件數 ($t=-.33, p=.75$) 方面，不同介面間並無明顯差異。但MAP介面的查詢次數較多。此結果顯示使用者在使用MAP時，平均每次瀏覽的文件比較少 (見表4)。

表4 不同介面之檢索特性比較

	介面	平均數	標準差	個數
相似度	PubMed	.54	.31	88
	Map	.40	.27	88
新增詞彙數	PubMed	1.57	1.54	88
	Map	2.58	2.01	88
刪除詞彙數	PubMed	.60	.90	88
	Map	.82	1.09	88
最終查詢詞彙數	PubMed	3.66	1.91	88
	Map	4.38	2.09	88
查詢次數	PubMed	3.74	2.74	88
	Map	5.27	2.78	88
瀏覽文件數	PubMed	89.63	60.10	88
	Map	92.66	63.32	88

***p<.001, *p<.05

(四) 檢索詞適用性分析

為了比較檢索詞的品質，受測者根據其主觀判斷來為檢索的適用性評分。ANOVA的檢定顯示MAP介面的檢索適用性要高於PubMed介面($F_{(1,86)}=7.88$, $p=.006$)。表5顯示不同檢索情境在兩個介面下詞彙適切性的平均分數及標準差。兩者之間的差異尤其在新的檢索問題，換言之，在受測者對問題較不熟悉的情況下更為明顯(見圖8)。

表5 受測者互動後檢索詞適切性(n=88)

	介面	Mean	SD	N
全新檢索	PubMed	4.21	1.10	28
	MAP	4.79	.69	28
再訪檢索	PubMed	4.55	.89	60
	MAP	4.70	.87	60
整體	PubMed	4.44	.97	88
	MAP	4.72	.81	88

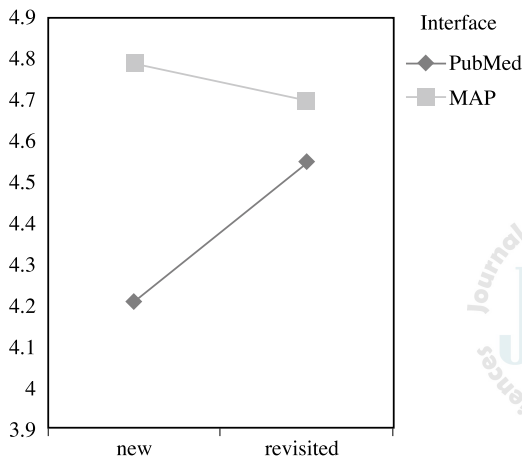


圖8 不同介面最終檢索詞彙適切性之比較

(五) 檢索結果過濾分析

使用者初始查詢多使用過於廣泛的關鍵字進行檢索，而後透過與檢索結果互動後更進一步重構 (reformulation) 其查詢。本研究的研究目的之一即比較使用者透過不同介面重構其查詢而過濾初始查詢中不相關的資訊，同時挖掘位於檢索結果底部相關資訊的能力。

一般而言，使用者缺乏領域內的相關知識或僅有模糊的資訊需求時，從而以少量關鍵字進行檢索，其檢索結果則是以詞彙共現之程度進行排序的紀錄。在PubMed之檢索結果是以年代排序，造成在大量檢索結果之中，排序較低的相關文獻則被忽略。而透過MAP提供的大量相關詞彙進行查詢重構，能進一步強化其資訊需求表述，其目的在於使能搜尋到在大量檢索結果中排序較低但相關程度亦高的文獻。

為比較PubMed與MAP介面何者較能輔助使用者取得在大量檢索結果中排序較低的相關文獻，研究者也嘗試分析使用者所挑選的文件其在初始搜尋結果的排序位置 (rating position)。如此比較的前提為在實驗過程中，使用者必須用相同的初始檢索來檢索PubMed與MAP介面，以得到相同的初始結果。

其前提為在實驗過程中，使用者並未刪除其初始查詢中的任一關鍵字，以確保最終查詢是透過新增詞彙過濾初始查詢結果內的不相關文獻。並非所有的檢索均能符合此一條件。在88組檢索中，共有10組檢索符合此一條件 (11.36%)。相關文獻之中數、最大值、最小值、全距如表6。

表6 不同介面相關文獻於初始檢索結果之位置 (ranking)

使用者	介面	中位數	平均數	標準差	最小值	最大值	全距	個數
2	PubMed	134.5	131.17	54.25	39	197	158	6
	MAP	503	516.17	259.39	219	829	610	6
24	PubMed	52	54.80	38.17	11	139	128	10
	MAP	78.5	99.50	50.36	38	167	129	10
25	PubMed	440	487.83	182.87	345	820	475	6
	MAP	1393.5	5078.83	6513.36	366	13812	13446	6
27	PubMed	38.5	44.90	34.60	2	103	101	10
	MAP	42	54.10	44.73	7	128	121	10
29	PubMed	133	115.10	42.80	52	171	119	10
	MAP	90	114.60	58.50	58	254	196	10
30	PubMed	60	69.50	30.05	39	121	82	10
	MAP	476	483.00	207.92	194	867	673	10
31	PubMed	695.5	687.63	38.74	629	745	116	8
	MAP	1371.5	1459.13	793.21	616	2549	1933	8
32	PubMed	107	103.29	38.90	33	160	127	7
	MAP	151	162.71	97.45	33	297	264	7
44	PubMed	88	84.90	32.79	37	125	88	10
	MAP	71	141.00	198.15	26	676	650	10
46	PubMed	583.5	873.70	803.87	90	2420	2330	10
	MAP	2519	2404.30	1049.45	1063	3586	2523	10

在10組符合檢索初始查詢結果案例中可發現，PubMed及MAP介面其相關文獻位置之最小值(highest ranking)未有明顯差距，但在最大值(lowest ranking)部分則MAP介面多遠高於PubMed介面，顯示MAP介面確能較PubMed介面輔助使用者取得在原始查詢中排序較低的相關文獻。

以使用者2號為例，其不同介面之文獻排序如圖9，在PubMed介面中使用者選取之相關文獻集中於檢索結果的前197筆；而在MAP介面下，在相同檢索結果中，使用者反而能取得原始檢索結果排序末端的第829筆紀錄。

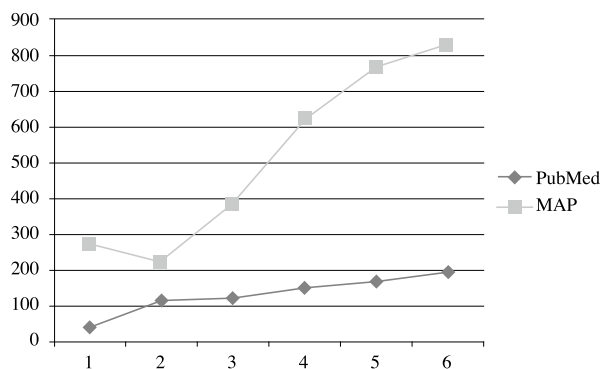


圖9 使用者2號之不同介面文獻排序

而使用者30號在不同介面之文獻排序如圖10，在PubMed介面中使用者選取之相關文獻則集中於檢索結果的前121筆，而在MAP介面下使用者反而能取得第867筆紀錄。

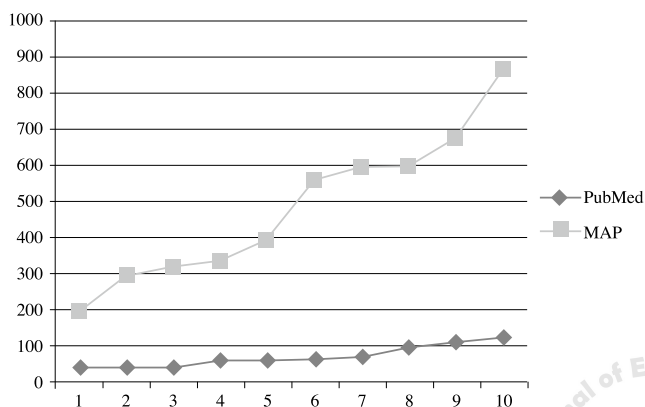


圖10 使用者30號之不同介面文獻排序

這些案例反映了MAP介面較PubMed介面有助使用者進一步取得大量檢索結果底層之相關文獻。

(六) 使用者評論

本研究在實驗過後要求使用者填寫問卷，說明MAP介面在何時較有幫助，並給予相關的改進建議。經過分析後，使用者認為MAP介面的主要的功能在於資料過濾、關鍵詞聯想等兩個面向。舉例而言，使用者6、9、18、23、26、31、34均提出了資料過濾的概念。

如果研究主題屬於大範圍，詞彙建議會有幫助。(6)

當搜尋範圍過大以致資料難以消化(數百、數千筆資料)則可以此減少資料彙整。(9)

對牽涉到較多領域的主題能夠幫助集中焦點。對於比較細部的研究能提供較精確的資料，幫助節省過濾資料的時間。(18)

縮小 search 的 range。(23)

在分類下，有系統的更多相關TERM更容易符合較細的檢索目標。(26)

在做研究時，往往需要擬定方向，才能夠縮小範圍，在詞彙建議下，可協助使用者找到相對應的資源。(34)

而在幫助聯想部分，使用者1、3、11、23、27則認為MAP介面有助聯想相關的檢索詞彙。

提供更精確的字眼或同義字幫助聯想。(1)

適時的提供keyword、新的想法。(3)

可提供研究者當初無想到之關鍵詞彙。(11)

找到更有興趣的主題。(23)

提供新的關鍵字或新的訊息，也許原本不知道(不熟)的字。也許原本知道，但仍可快速的分類資料或提醒提供更精確的字眼或同義字幫助聯想。(27)

使用者認為在熟悉主題與不熟悉主題的情況下，MAP的功用互有不同。在不熟悉主題的情境下，MAP介面可提供MeSH中的語意架構，輔助使用者了解領域內的知識架構並取得相關的關鍵詞進行檢索，37名使用者均提出相似的概念，以使用者2、4、16、20、30、33為例：

題目不熟時；不知下關鍵字時；有助help與知道下何種keywords。(2)

在對該主題不熟悉時，可觸類旁通，知道某些keyword後，MeSH可提供我其他可能的word是我之前所不知道的，剛進入此領域時有幫助。(4)

我認為在一開始完全進到新的、完全不懂的領域下，MAP的MeSH功能有很多分類可以了解這疾病如何在人類產生，及其機制，這可幫助剛進來的生手。(16)

當對查找主題不熟悉時，MeSH提供了許多對該主題的建議，幫助使用者更快取得對主題的概觀。(20)

當使用者想要收集某一較general的主題時，MAP是很好的幫手，使用者可得到很多新的靈感也不受關鍵字限制，可得到更多資料。(30)

在僅知大範圍的關鍵字時，詞彙建議可幫助提供原先沒有想到的term或narrow down搜尋範圍，更容易找到目標文獻。(33)

而在熟悉主題的情境下，使用者認為MAP介面的輔助效益不大，但仍有幫助相關概念的聯想及指引，以使用者14、20、21、22、30、36、37、43為例：

對於熟悉的主題，差異則不大，但會使搜尋更精確。(14)

而在針對熟悉領域時，MeSH也扮演了一個提醒者的角色，可確保任一關鍵字不被遺忘。(20)

在平常已有確定目標的搜尋中，PubMed/map的效果差不多。(21)

較熟悉的研究主題，能提供更多不同層面的知識。(22)

當使用者已有一非常SPECIFIC的題目時，MAP幫助不大，甚至會使資料搜尋速度下降。(30)

再者，若對所蒐集詞彙了解十分透徹，可藉詞彙建議了解其他研究面向。(35)

在已熟悉的前提下，MAP仍有助於縮小尋找範圍，並有助於條列式搜尋。(36)

當對主題很了解時，詞彙建議能限制出想要的主題並開啟新的思維。(37)

但在對於要檢索的文獻已相當熟悉情況下，有時甚至毋需使用到MeSH即可完成檢索工作，但MeSH仍能提供新的方向，發現自己熟悉題目的新走向，有助於開展視野。(41)

在熟悉的題目上，感覺詞彙建議的功能就較貧弱，因此其提供的關鍵字，對於熟悉的題目而言，實在不夠具體。並且知識上較清楚，相關的重要程度。因此在策略上寧可運用既有知識去搜尋。(44)

在實驗過後，要求使用者以五點尺度填答對於建議詞彙之滿意度及以七點尺度填答，題目為若MAP介面有所幫助則是在何種面向較有幫助，其面向分別為1.其輔助產生新的概念進行下一步搜尋，2.幫助釐清檢索問題，3.能夠顯示資料庫文獻的架構，4.管理大量的檢索結果。

研究結果如表7，雖然新舊檢索對於建議詞彙之滿意度無顯著差異，但新的檢索對於建議詞彙的滿意度 ($m=4.28$) 略高於重複檢索 ($m=4.08$)。而在MAP的輔助面向部分，則只有在「產生新的概念」此一面向新舊檢索有明顯差異 ($p=.036$)。分析結果與使用者意見相符，顯示MAP介面在使用者面臨不熟悉主題時能夠提供相關主題詞彙供使用者利用檢索之。這一點也呼應之前對於互動式詞彙擴展的評估研究 (Efthimiadis, 2000)。

表7 使用者對於MAP輔助功能評估

		平均數	標準差	個數
詞彙滿意程度	全新檢索	4.39	1.10	28
	重複檢索	4.03	1.68	60
產生新的概念*	全新檢索	4.57	1.10	28
	重複檢索	3.95	1.35	60
釐清檢索問題	全新檢索	4.18	1.09	28
	重複檢索	4.03	1.57	60
顯示文獻結構	全新檢索	4.18	1.28	28
	重複檢索	4.50	1.10	60
管理檢索結果	全新檢索	4.29	0.90	28
	重複檢索	4.42	1.16	59

*p<.05

六、結 論

本研究旨在了解不同介面下的使用者查詢擴展行為，以及MAP介面輔助使用之成效，期望能提供資訊檢索系統設計者及使用者行為研究者參考。以下綜合本研究分析使用者詞彙選擇之結果以及相關特殊案例之說明，針對使用者查詢擴展行為提出以下結論。

(一) 主題熟悉度會影響使用者查詢行為

研究結果顯示，使用者在熟悉主題以及不熟悉主題的查詢行為上有明顯不同。在不熟悉情境下，使用者會大量嘗試並修正檢索詞彙並進行多次檢索，此一現象在兩個檢索介面均有出現，在MAP介面中尤為明顯。從查詢相似度來看，熟悉主題之查詢相似度會較不熟悉主題之相似度來得高，意即在熟悉主題情況下，使用者之最終查詢與初始查詢間並不會產生明顯變化。這是在熟悉主題情境下，使用者對於其所需的資訊有一定程度的了解，對於資訊需求之表述亦能有效表達，因此，其初始及最終查詢彼此之間並不會有太大差距。

而在不熟悉主題情境下，使用者缺乏對於該主題的知識以及對於資訊需求的理解，因此只能粗略的進行資訊需求表述，與資訊檢索系統互動後而進一步改善其資訊需求，以檢索其所需的資訊。透過與系統互動的過程，重構使用者初始查詢，因此其查詢相似度會明顯低於熟悉主題之查詢。

而在瀏覽文件數及檢索次數，不同主題熟悉度並無明顯影響，此說明無論在何種情境下，使用者仍須瀏覽一定數量的檢索結果來確認資訊需求以得到滿足，並經過一定次數的檢索過程才能驗證檢索詞彙之適用性。

(二) 不同介面會影響使用者查詢行為

查詢相似度是以使用者初始查詢及最終查詢互相比較，用以說明不同介面影響使用者修改其檢索詞彙的能力。研究結果顯示，不同介面之查詢行為

的確有所不同。MAP介面之查詢相似度低於PubMed介面，其原因在於MAP系統能提供使用者大量的建議相關詞彙，因此使用者可從中辨認出之前未想到的檢索詞，而在PubMed介面中，使用者僅能就檢索結果來獲得詞彙的反饋，其查詢詞彙之刺激有限，所以其最終查詢高度與初始查詢相似。在同樣原因之下，PubMed之新增詞彙數及移除詞彙數亦較MAP介面來得少。在瀏覽文件數不同介面並無明顯差異，但在與MAP互動時，查詢次數有明顯增加，由此可看出使用者在用MAP介面搜尋，平均瀏覽結果的件數比較少。

(三) 不同主題熟悉度與不同介面對使用者查詢行為產生交互作用

研究結果中主題熟悉度與不同介面對於查詢行為之交互作用雖然不顯著，然而根據使用者之評論以及相關變數之測量，結果顯示MAP介面在使用者較不熟悉檢索主題時可發揮較大作用，其能提供與檢索主題相關之建議詞彙，輔助使用者重構其資訊需求，釐清模糊的初始查詢。而在熟悉的情境下，使用者因對於該主題有所認知，並掌握一定數量的關鍵詞彙，因此認為MAP系統的輔助作用不大，無助於改善資訊檢索結果。

(四) MAP介面能助使用者取得大量檢索結果中排序較低相關文件

PubMed在處理使用者查詢時，是將使用者查詢之關鍵字以布林邏輯AND連接，然後傳回含有該關鍵詞之文獻，並以出版年份排序。然而在使用者輸入語意較為廣泛之詞彙時，則檢索結果也趨於龐大，而使用者又僅會瀏覽點選前數頁之檢索結果，使得使用者所需的相關文獻埋於大量檢索結果之底部，且無從取得。

在符合初始查詢及最終查詢均檢索同一回收結果(return set)的十個案例之中，研究結果發現透過MAP提供之建議詞彙，使用者能在相同檢索結果中，利用增加關鍵字的方式來縮小檢索結果，逐步逼近位於大量檢索結果底部之相關文獻，並從而取得其所需資訊。

七、建 議

綜合文獻分析、使用者查詢查詢擴展行為分析及使用者意見調查，研究者提出下列建議供資訊檢索研究者參考。

(一) 發展能夠支援不熟悉情境之檢索工具

資訊檢索學界傳統上認為資訊檢索系統只要存有使用者所需的相關資訊，而使用者有能力檢索之，則該系統必對使用者有所幫助。然而不同的使用者即使擁有相同的檢索主題，其對於檢索結果的要求以及對檢索主題的理解也不盡相同，因此認為檢索系統能夠一體適用(one size fits all)的觀念，忽略了使用

者在不同搜尋情境下對主題熟悉程度所產生的資訊需求。

White et al. (2006) 提出探索式檢索 (Exploratory search)，認為在使用者缺乏領域內的專業知識時，亦無適當的關鍵字表述其資訊需求，於此檢索系統必須要提供瀏覽及探索的功能，以輔助使用者能進一步理解此一不熟悉領域內之知識架構，從而形塑其查詢以進行資訊檢索。

本研究結果亦發現，詞彙建議系統在使用者於不熟悉情境下能夠發揮較大的效用，使用者可藉由相關的建議詞彙理解其不熟悉的領域，並從中取得相關之關鍵字，而能檢得其所需的資訊。

由此可見，使用者對於檢索主題之熟悉度以及其對於檢索結果之要求均會影響其利用資訊檢索系統，因此，未來之研究可進一步區分使用者之檢索情境，而設計與其需求相符之資訊檢索系統進行研究，以求提供使用者最大的支援 (Kules, & Capra, 2008)。

(二) 採用使用者之真實資訊需求進行研究

傳統資訊檢索評估，主要以研究者先建的檢索主題進行，並以事先決定的相關判斷做為比較的基準，而後以使用者的檢索結果進行比較。此種設計的優點在於結果較為客觀且容易計算，且不同系統間能互相比較，而缺點在於研究之生態效度較低，無法真實呈現使用者在真實情境下的資訊檢索，同時忽略了使用者在操作過程中與資訊互動之後，認知狀態及資訊尋求動態的改變。而採用使用者真實資訊需求作為資訊檢索主題進行檢索，其目的即在於提高研究之生態效度，期能進一步貼近使用者的真實檢索情境。

然而，採用使用者真實資訊需求所產生之問題在於不同主題間缺乏比較之基準點。在研究者先建的主題下，可比較在不同處理下所產生的檢索結果，從而比較資訊檢索系統之優劣。但採用使用者真實資訊需求做為檢索主題時，因使用者行為會受到檢索主題的強烈影響，是故檢索結果亦會劇烈變化，使得系統間難以進行比較。如何在控制與真實之間取得平衡，仍是未來系統評估研究的一大課題。

參考文獻

- 黃慕萱(1996)。檢索系統評估。在資訊檢索(第十章,頁269-275)。台北市:台灣學生。
- 洪承理(2010)。使用者查詢擴展行為研究:以MeSH為詞彙來源的建議詞彙工具為例。未出版之碩士論文,台灣大學圖書資訊學系,台北市。
- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. New York: ACM Press.
- Bates, M. J. (1979). Information search tactics. *Journal of the American Society for information Science*, 30(4), 205-214.
- Belkin, N. J. (1993). Interaction with texts: Information retrieval as information-seeking

- behavior. *Information retrieval*, 93, 55-66.
- Belkin, N. J., Chang, S. J., Downs, T., Saracevic, T., & Zhao, S. (1990). Taking account of user tasks, goals and behavior for the design of online public access catalogs. In *Proceedings of the 53rd ASIS annual meeting* (pp. 69-79). Medford, NJ: Learned Information.
- Belkin, N. J., Cool, C., Head, J., Jeng, J., Kelly, D., Lin, S., Lobash, L., Park, S. Y., Savage-knepshild, P., & Sikora, C. (2000). Relevance feedback versus local context analysis as term suggestion devices: Rutgers' TREC-8 interactive track experience. In *TREC-8, Proceedings of the Eighth Text Retrieval Conference* (pp. 565-574), Maryland: Gaithersburg.
- Belkin, N. J., Kelly, D., Kim, G., Kim, J. Y., Lee, H. J., Muresan, G., Tang, M. C., et al. (2003). Query length in interactive information retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 205-212). New York: ACM.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38(2), 61-71.
- Borlund, P., & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3), 225-250.
- Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1), 71-90.
- Huang, C. K., Chien, L. F., & Oyang, Y. J. (2003). Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology*, 54(7), 638-649.
- Cleverdon, C. W. (1967). The Cranfield tests on index language devices. *Aslib Proceedings*, 19(6), 173-194.
- Doms, A., & Schroeder, M. (2005). GoPubMed: Exploring PubMed with the gene ontology. *Nucleic Acids Research*, 33(Web Server Issue), 783-786.
- Dumais, S. T., & Belkin, N. J. (2005). The TREC interactive tracks: Putting the user into search. In E. M. Voorhees, & D. Harman (Eds.), *TREC: Experiment and evaluation in information retrieval* (pp. 123-153). Cambridge Mass: MIT Press.
- Efthimiadis, E. N. (1996). Query expansion. In M. E. Williams (Ed.), *Annual Review of Information Systems and Technology*, 31, 121-187.
- Efthimiadis, E. N. (2000). Interactive query expansion: A user-based evaluation in a relevance feedback environment. *Journal of the American Society for Information Science*, 51(11), 989-1003.
- Fidel, R. (1984). Online searching styles: A case-study-based model of searching behavior. *Journal of the American Society for Information Science*, 35(4), 211-221.
- Goetz, T., & Lieth, C. W. V. D. (2005). PubFinder: A tool for improving retrieval rate of relevant PubMed abstracts. *Nucleic Acids Research*, 33(Web Server Issue), 774-778.
- Harman, D. (1988). Towards interactive query expansion. In *Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 321-331). New York: ACM.

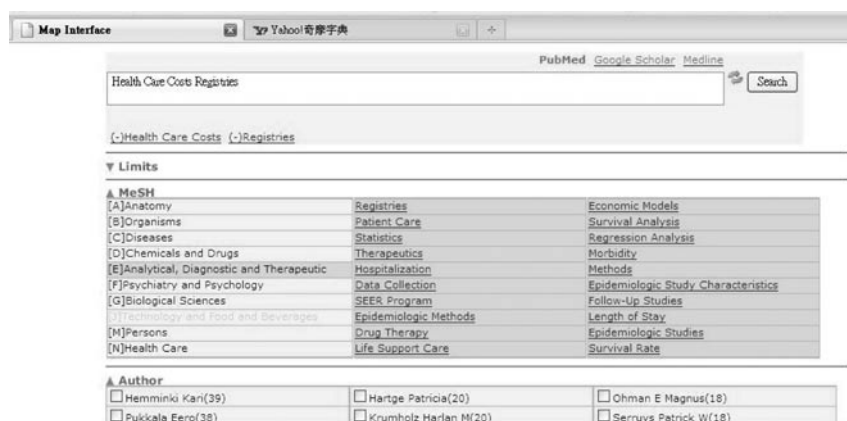
- Harman, D. (1992). Relevance feedback revisited. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 1-10). New York: ACM.
- Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., & Yee, K. P. (2002). Finding the flow in web site search. *Communications of the ACM*, 45(9), 42-49.
- Hersh, W., Pentecost, J., & Hickam, D. (1996). A task-oriented approach to information retrieval evaluation. *Journal of the American Society for Information Science*, 47(1), 50-56.
- Hersh, W. R., & Hickam, D. H. (1998). How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. *JAMA*, 280(15), 1347-1352.
- Huang, C. K., Chien, L.F., & Oyang, Y.J.(2003). Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology*, 54(7), 638-349.
- Ingwersen, P., & Järvelin, K. (2005). *The turn: Integration of information seeking and retrieval in context*. Dordrecht, Netherlands: Springer.
- Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. (1998). Real life information retrieval: A study of user queries on the web. *SIGIR Forum*, 32(1), 5-17.
- Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users and real needs: A study and analysis of users queries on the web. *Information Processing and Management*, 36(2), 207-227.
- Joho, H., & Jose, J. M. (2006). Slicing and dicing the information space using local contexts. In *Proceedings of the 1st international conference on Information interaction in context* (pp. 66-74). New York: ACM.
- Koenemann, J., & Belkin, N. J. (1996). A case for interaction: A study of interactive information retrieval behavior and effectiveness. In *Proceedings of the ACM SIGCHI conference on human factors in computer systems* (pp. 205-212).
- Kristensen, J., & Järvelin, K. (1990). The effectiveness of a searching thesaurus in free text searching of a full-text database. *International Classification*, 17(2), 77-84.
- Kules, B., & Capra, R. (2008). Creating exploratory tasks for a faceted search interface. In *Second Workshop on Human-Computer Interaction (HCIR 2008)*.
- Lin, X. (1999). Visual MeSH. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 317-318). New York: ACM.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York: Cambridge University Press.
- Pennanen, M., & Vakkari, P. (2003). Students' conceptual structure, search process, and outcome while preparing a research proposal: A longitudinal case study. *Journal of the American Society for Information Science and Technology*, 54(8), 759-770.
- Perez-Iratxeta, C., Bork, P., & Andrade, M. A. (2001). XplorMed: A tool for exploring MEDLINE abstracts. *Trends in biochemical sciences*, 26(9), 573-575.
- Plikus, M. V., Zhang, Z., & Chuong, C. M. (2006). PubFocus: Semantic MEDLINE/PubMed citations analytics through integration of controlled biomedical dictionaries and ranking

- algorithm. *BMC Bioinformatics*, 7(1), 424.
- Pollitt, A. S. (1998). The key role of classification and indexing in view-based searching. *International cataloguing and bibliographic control*, 27(2), 37-40.
- Robertson, S. E., & Hancock-Beaulieu, M. M. (1992). On the evaluation of IR systems. *Information Processing and Management*, 28(4), 457-466.
- Ruthven, I. (2008). Interactive information retrieval. In B. Cronin (Ed.), *Annual Review of Information Science & Technology* (Vol. 42, pp. 43-91). Medford, NJ: Information Today.
- Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension and applications. In *Proceedings of the American Society for Information Science* (Vol. 34, pp. 313-327).
- Saracevic, T., & Kantor, P. (1988a). A study of information seeking and retrieving. II. Users, questions, and effectiveness. *Journal of the American Society for Information Science*, 39(3), 177-196.
- Saracevic, T., & Kantor, P. (1988b). A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science*, 39(3), 197-216.
- Saracevic, T., Kantor, P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science*, 39(3), 161-176.
- Schatz, B. R., Johnson, E. H., Cochrane, P. A., & Chen, H. (1996). Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval. In *Proceedings of the first ACM international conference on Digital libraries* (pp. 126-133). New York: ACM.
- Shiri, A., & Revie, C. (2006). Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation. *Journal of the American Society for Information Science and Technology*, 57(4), 462-478.
- Shiri, A., Revie, C., & Chowdhury, G. (2002). Thesaurus-assisted search term selection and query expansion: A review of user-centred studies. *Knowledge Organization*, 29(1), 1-19.
- Spink, A. (1995). Term relevance feedback and mediated database searching: Implications for information retrieval practice and systems design. *Information Processing and Management*, 31(2), 161-171.
- Spink, A., & Saracevic, T. (1997). Interaction in information retrieval: Selection and effectiveness of search terms. *Journal of the American Society for Information Science*, 48(8), 741-761.
- Su, L. T. (1992). Evaluation measures for interactive information retrieval. *Information Processing and Management*, 28(4), 503-516.
- Su, L. T. (1994). The relevance of recall and precision in user evaluation. *Journal of the American Society for Information Science*, 45(3), 207-217.
- Su, L. T. (2003). A comprehensive and systematic model of user evaluation of web search engines: II. An evaluation by undergraduates. *Journal of the American Society for Information Science and Technology*, 54(13), 1193-1223.
- Tang, M. C. (2006, November). Multiple accesses to PubMed: A proposed user study of term suggestion tools for PubMed. In *ASIS&T (American Society for Information Science and*

- Technology*) 2006 annual meeting, Symposium held at the Nation Taiwan University, Taiwan .
- Tang, M. C. (2007). Browsing and searching in a faceted information space: A naturalistic study of PubMed users' interaction with a display tool. *Journal of the American Society for Information Science and Technology*, 58(13), 1998-2006.
- Tang, M. C., Wu, W. C., & Hung B. W. (2009). Evaluating a metadata-based term suggestion interface for PubMed with real users with real requests. In *Proceedings of the Annual Meeting of the American Society for Information Science & Technology 2009* (pp. 35-42).
- Vakkari, P. (1999). Task complexity, problem structure and information actions integrating studies on information seeking and retrieval. *Information Processing and Management*, 35(6), 819-837.
- Vakkari, P. (2000). Cognition and changes of search terms and tactics during task performance: A longitudinal case study. In *Proceedings of the RIAO 2000 Conference* (pp. 894-907).
- Vakkari, P. (2001). A theory of the task-based information retrieval process: A summary and generalisation of a longitudinal study. *Journal of Documentation*, 57(1), 44-60.
- Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, 56(5), 540-562.
- Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing and Management*, 39(3), 445-463.
- White, R. W., Kules, B., Drucker, S. M., & Schraefel, M. C. (2006). Supporting exploratory search, introduction, special issue, communications of the ACM. *Communications of the ACM*, 49(4), 36-39.
- White, R. W., Ruthven, I., & Jose, J. M. (2005). A study of factors affecting the utility of implicit relevance feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 35-42). New York: ACM.
- Xie, I. (2008). *Interactive information retrieval in digital environments*. Hershey: IGI Publishing.

附 錄

一、介面截圖



二、正式實驗問卷

感謝您同意參與本研究。本研究試圖了解使用者參與「醫學資料庫詞彙建議研究」的相關資訊行為，藉以評估該介面系統的使用效能。本研究立意於了解使用者的資訊尋求行為及該資料庫介面的使用性，測驗的過程以及結果並非測研您的搜尋能力，請您放鬆心情作答。您的個人資料在研究報告結果中將會以匿名方式處理，所有資料將只用在本研究之中。

在介面測試時，您將參與以下程序，請您仔細閱讀，若有問題也可以提出：

介面測試前：

1. 您將會閱讀及簽署本同意書，您可以在這個階段提出任何有關的問題。
2. 您將會填寫一份簡短的問卷，藉以了解您先前的背景資料和搜尋經驗。

介面測試開始時：

1. 本測試的輔助程式將會紀錄您在使用介面時的螢幕畫面，在測驗期間您的使用過程將會被記錄下來。

2. 在檢索過程中請嘗試解釋選取檢索詞的理由，您的對話將被錄音。

測試的任務：

您將與我們所提供的介面互動並搜尋您兩個自己的檢索問題。請盡可能至少選擇一個您比較不熟悉的主題。

1. 您有30分鐘時間來完成每個問題，在檢索每一個問題時，您各有15分鐘時間分別使用兩種檢索模式，一為一般PubMed介面（無詞彙建議功能），另一有詞彙建議功能，若提早完成可舉手告知研究者。

2. 在搜尋過程中，您的主要任務有二：一為找出並存取20筆對您最有用的文獻（每一檢索模式各10筆； $10 \times 2 = 20$ ）；二為在與介面互動之後，寫下您認為最能夠表達您的資訊需求的檢索詞彙。

介面測試後：

1. 每一問題搜尋結束後，您將針對該檢索經驗填答搜尋後問卷，並為找到的每筆資料評分。

本研究乃自願性質，為時約一個半小時，結束後您將獲得新台幣三百八十元報酬。若您中途放棄測試將不會得到任何酬勞。

我，_____已確實閱讀以及了解這份同意書，並同意參本研究的測驗。

_____（使用者簽名）日期： 年 月 日

_____（施測者簽名）日期： 年 月 日

背景資料

1. 請問您的主要身分是？

- 大學生
研究生
研究人員
教師
醫技／醫療護理人員
其他

2. 性別

- 男 女

3. 請問您的研究／工作專業領域為何？如物理治療，生物化學，皮膚科等。

4. 請問您使用PubMed或Medline的頻率為何？

從未使用	兩月一次	每月一次	兩周一次	每周一次	每天使用

5. 請問您使用Google學術搜尋(Google Scholar)的頻率為何？

從未使用	兩月一次	每月一次	兩周一次	每周一次	每天使用

6. 請問您對醫學標題表(MeSH)的熟悉程度為何？直接圈選下排數字作答。

完全不熟悉					非常熟悉	
0	1	2	3	4	5	6

7. 請問當您使用PubMed或Medline檢索問題時，您運用進階檢索技巧(如布林邏輯AND/OR/NOT，查詢擴展，限制條件等)的頻率為何？

從未使用					每次都	
0	1	2	3	4	5	6

搜尋前問卷

1. 請敘述您想要檢索的問題。

2. 假設下面的空格代表檢索欄位，您會輸入哪些檢索詞來檢索這個問題呢？

3. 是否曾經檢索過這個問題？

- 是 否

4. 請問您對此檢索問題涉及的主題熟悉程度為何？

完全不熟悉					非常熟悉	
0	1	2	3	4	5	6

5. 請問您對檢索結果完整度的要求有多高？

有一篇代表性文獻即可					完整的文獻回顧	
0	1	2	3	4	5	6

6. 如果今天您無法親自檢索這個問題，必須請他人代為檢索，您認為他人的檢索結果能夠滿足您資訊需求的程度是？

完全不能						完全能	
0	1	2	3	4	5	6	

搜尋後問卷

請根據您剛剛利用介面的經驗回答下列問題，直接圈選下排數字作答：

經過方才與介面互動之後，您或許將修改一開始的檢索詞。請在下方空格寫下現在您認為最能理想地表達您原本檢索問題的檢索詞彙（可加入布林邏輯，如：AND, OR, NOT）。

1. 請問您認為該介面容易使用的程度為何？

完全不易			部分容易			非常容易
0	1	2	3	4	5	6

2. 請問您認為此介面對檢索的有用程度為何？

完全沒用			部分有用			非常有用
0	1	2	3	4	5	6

3. 整體而言，您對使用關鍵詞推薦介面所搜尋到的結果滿意程度為？（請針對剛剛的檢索經驗作答）

完全不滿意				部分滿意		非常滿意
0	1	2	3	4	5	6

4. 請問您對推薦詞彙的滿意程度為何？

完全不滿意				部分滿意		非常滿意
0	1	2	3	4	5	6

5. 相較於傳統介面，若關鍵詞推薦介面對您有所幫助，請問是在哪些方面有助於檢索呢？請指出以下敘述和您經驗的符合程度。

a. 介面提供的關鍵字給予我新的想法，使我可能因此發展出新的研究方向。

完全不符合			部分符合			非常符合
0	1	2	3	4	5	6

Example

b. 介面提供的關鍵字使我的檢索詞更加明確地表達我原本的問題。

完全不符合			部分符合			非常符合
0	1	2	3	4	5	6

Example

c. 此介面呈現了資料庫相關文獻的範圍和結構

完全不符合			部分符合			非常符合
0	1	2	3	4	5	6

Example

d. 此介面幫助我處理龐大的檢索結果

完全不符合			部分符合			非常符合
0	1	2	3	4	5	6

請判斷您所檢索出的每一篇文件與您的檢索主題相符程度 (relevance) 與提供您之前不知道 (novelty) 的程度。由0到6，0代表完全不，6代表非常。

Topic /Interface

#	Seen Before? (Y/N)	Relevance	Novelty
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

End of the search questionnaire

1. 請評估您在檢索前所填理想檢索詞彙能夠完整表達資訊需求的程度

完全不能						完全能
0	1	2	3	4	5	6

2. 請您評估在與 PubMed 互動後所填理想檢索詞彙能夠完整表達您資訊需求的程度。

完全不能						完全能
0	1	2	3	4	5	6

3. 請您評估在與 MAP 互動後所填檢索詞彙能夠完整表達您資訊需求的程度。

完全不能						完全能
0	1	2	3	4	5	6

4. 請談談您認為在什麼情況下詞彙建議的功能會對搜尋有幫助。

三、受試者檢索問題

USER	REVISIT	REQUEST
2	全新檢索	Legionella sp. 與氣象因子的關係
2	重複檢索	呼吸系統與空氣中微生物的關係
6	重複檢索	ER的CPG island分布
6	重複檢索	HBV影響的 signaling pathway
7	全新檢索	不同亞型流感病毒血球凝集素的鍵結親和力
7	重複檢索	近年流感病毒神經胺酸酶的變異性及其在病毒複製過程所扮演角色
8	全新檢索	the effect of TMV-encoded p126 protein in suppression of VIGS
8	重複檢索	the role of Grb7 regulation of cellular functions and tumor formation
9	全新檢索	果膠酯酶抑制劑
9	重複檢索	大豆異黃酮及其相關水解酵素
10	全新檢索	以導電高分子製作奈米及尺寸表面，觀察對心肌細胞行為之影響
10	重複檢索	利用表面地形的改變觀察肝細胞型態及功能之影響關聯性
11	全新檢索	心理衛生相關醫療資源
11	重複檢索	肺癌病患接受外科手術後之醫療品質
12	重複檢索	traumatic brain injury (TBI) & dementia & Alzheimer's disease (AD)
12	重複檢索	traumatic brain injury (TBI) & social behavior & theory of mind (ToM)
13	重複檢索	總額預算制度對於抑制醫療費用成長之成效
13	重複檢索	paruresis 之危險因子
14	全新檢索	如何分離犬隻的 T-reg Cells
14	重複檢索	TGF-B 如何影響 canine dendritice cell 的 mafuration
15	重複檢索	光動力療法治療與奈米粒子包覆 PPIX
15	重複檢索	癌症幹細胞的抗藥性
16	重複檢索	asthma 的體內免疫反應形況
16	重複檢索	類風濕性關節炎的動物模式
17	全新檢索	biochip 用於檢測癌症方面
17	重複檢索	preparing nanoparticle, drug delivery 如何製備 nanoparticle, 包附化療藥物 dox
18	全新檢索	奈米粒子在 EPR 效應下，是否有效果上的提升
18	重複檢索	了解 hydrogel 在藥物釋放方面所具有的特性及效果
19	全新檢索	嗎啡計量的耐受性和相關的研究數值及作用途徑
19	重複檢索	rumination depression
20	重複檢索	protein docking prediction
20	重複檢索	protein function prediction using network property
21	全新檢索	drug resistance in mTOR signaling
21	重複檢索	Rheb
22	重複檢索	探討 GSK3-B 在肺炎細胞中的調控
22	重複檢索	應用蛋白質體學探討丹參酚酸 B 調控血管內皮細胞
23	全新檢索	吞嚥障礙患者之電刺激療法
23	重複檢索	研究使用於失語症患者的健康生活品質
24	全新檢索	PBDE 的毒性資料
24	重複檢索	針對 fFuran 分析方法
25	全新檢索	bioreactor 生物反應器
25	重複檢索	幾丁聚醣相關文章
26	全新檢索	low back pain, LBC, prevalence
26	重複檢索	芳香煙受體功能
27	重複檢索	斑馬魚的肌肉發育過程為何
27	重複檢索	lactoferricin 的功能為何

- 28 重複檢索 the repair of hypoxanthine in DNA
- 28 重複檢索 the repair of uracil in double stranded DNA
- 29 重複檢索 database
- 29 重複檢索 text-mining
- 30 重複檢索 腦內神經傳導物質對睡眠的影響
- 30 重複檢索 兒茶素的抗癌機制
- 31 重複檢索 cell cycle的機制
- 31 重複檢索 VP28的影響及預防
- 32 重複檢索 醫院臨床癌症病患的營養狀況
- 32 重複檢索 放射線治療癌症患者營養介入成效
- 33 重複檢索 關於神經細胞於胚胎發育及 larva stage 發育的關聯機制
- 33 重複檢索 apoptosis 過程中 engulfment 的機制
- 34 全新檢索 scan zebrafish image by using ultrasound
- 34 重複檢索 the design of high-temperature superconducting coil for magnetic resonance imaging
- 35 重複檢索 the asymmetrical development of human embryo
- 35 重複檢索 the relation between the genetic variants of the COMT gene and schizotypal personality
- 36 重複檢索 了解 probiotics 定義及內涵，及在水產養殖上用於抵抗病毒病之功用或實力
- 36 重複檢索 NNN (nervous necrosis virus): 了解時空背景，近年新發現等
- 37 全新檢索 the effect of Brevetoxin on development, heart rate & c-fos expression
- 37 重複檢索 effect of Brevetoxin on Bovine Adrenal Chromaffin cell
- 38 全新檢索 the study of arabidopsis promote of ubiquitin genes
- 38 重複檢索 the functional study of Toc33 in Arabidopsis
- 39 全新檢索 micelle
- 39 重複檢索 study of cancer stem cell
- 40 全新檢索 骨質疏鬆症與運動習慣的相關性
- 40 全新檢索 營養與基因交互作用對運動表現之影響
- 41 全新檢索 gsps
- 41 重複檢索 glutaredoxin
- 42 全新檢索 flower color examination o Oncidium
- 42 重複檢索 Function of MADs-box genes in sweet potato
- 43 全新檢索 using MRI techniques to evaluate the leakage pattern of blood brain barrier in ischemic stroke model
- 43 重複檢索 source memory recollection ERP
- 44 全新檢索 motion perception
- 44 重複檢索 consciousness implicit-implicit processing
- 45 全新檢索 narcolepsy sleepdisorder orexin
- 45 重複檢索 app flatillin, lipiraft
- 46 全新檢索 aflatoxin
- 46 重複檢索 hypo cholesterolemic drug
- 47 全新檢索 關於 HIV 之治療
- 47 重複檢索 SVD 之單源抗體應用
- 48 全新檢索 目前組織工程在心肌梗塞治療上的幫助與應用
- 48 重複檢索 cartilage repair and tissue engineering

A User Study of the Effectiveness of MAP (Multiple-Access to PubMed): A MeSH Based Query Suggestion Tool

Muh-Chyun Tang*

Assistant Professor
Department of Library and Information Science
National Taiwan University
Taipei, Taiwan
E-mail: mctang@ntu.edu.tw

Cheng-Li Hung

Librarian
Humanities and Social Sciences Branch
National Tsing Hua University Library
Hsinchu, Taiwan
E-mail: R96126013@ntu.edu.tw

Abstract

An interface, MAP (Multiple-Access to PubMed) was proposed here to provide MeSH term suggestion for PubMed bibliographic search. This paper reported the results of an experiment comparing users' search behaviors and performance between MAP and regular PubMed interfaces. A total of 44 graduate students in Bio-medical domains in Taiwan participated in the study. They were asked to ask for two information problems of their own alternately with the two interfaces. Comparison were made between the interfaces in terms query "goodness", result quality, and impact on users' querying behaviors. Information regarding the attributes of their information problems was also collected through questionnaire so that the effectiveness of the interfaces with different types of information problems can be explored. The results show that 1. Topic familiarity would influence users' querying behaviors; MAP had a high impact on users' queries when searching for unfamiliar problems. 2. MAP was shown to be helpful, especially when searching for search requests that the users had not attempted before. 3. MAP was most helpful in generating new research ideas. 4. MAP was shown to be able retrieve documents buried deep down in the initial returned results.

Keywords: *Interactive information retrieval; Query expansion; User's behavior; Term suggestion system; MeSH*

SUMMARY

Introduction

In today's automatic indexing environment, users who input short queries are often faced with an unmanageable volume of results. Arguably one of the most

* Principal author for all correspondence.

difficult tasks in information retrieval (IR) is the representation of users' requests (Belkin, 1980). Search engine users typically submit very short and ambiguous queries (Jansen, Sprink, & Saracevic, 2000). The ambiguity of user requests stands in direct contrast to the comprehensiveness of document retrieval. This disparity often results in unmanageable search results for users of a heterogeneous and massive bibliographic database such as PubMed. Several system features of PubMed (e.g. default "explode" function and free-text indexing in title and abstract fields) that aim at facilitating end-user searching tend to increase indexing exhaustivity and therefore favor search recall at the expense of precision. Faced with an unmanageable volume of returned results, users are often left with few options but to hastily browse the first few pages.

The enormous size of the returned set causes a significant barrier between user-system communication, which is especially severe in search situations where users are searching for unfamiliar topics or have only vague information needs. Without timely feedback from the system, users are unlikely to be able to interact effectively with the system and refine their searches.

One approach to bridge the breakdown is to provide a term suggestion function. To address the communication breakdown, an interface was created for the PubMed search that utilizes MeSH co-occurrence information to provide support for users' query construction. A MeSH term co-occurrence database was created, using a local context analysis approach where MeSH terms and other bibliographic information, such as authors and journal titles, were extracted from metadata. After users submit their queries, the interface will input the queries to PubMed and identify proper MeSH terms so that other MeSH terms that occur with the target query terms can be retrieved and presented to users for term suggestion.

Research Questions

Our research question is two-fold; firstly, we were interested in knowing whether the system is effective in helping users in constructing their queries; and secondly, if the system is useful, under what kinds of search situations is it most useful. The effectiveness of the system was evaluated by the "goodness" of the queries, as judged by the users, as well as whether the system was able to help find relevant documents that might otherwise not be accessible. Several performance criteria were used to demonstrate the potential benefits of the term suggestion function, including comparison between users' initial and revised query, and user assessment of the "goodness" of the queries. Using genuine requests also allowed us to investigate whether the effectiveness of MAP varies in different search situations.

This paper reports on part of the results of an experiment comparing MAP and regular PubMed interfaces. A total of 44 graduate students in Bio-medical disciplines in Taiwan participated in the study. They were instructed to query two information problems of their own alternately with the two interfaces. Comparison were made between the interfaces in terms query “goodness”, result quality, and impact on users’ querying behaviors. Information regarding the attributes of their information problems was also collected through questionnaire so that the effectiveness of the interfaces with respect to different types of information problems could be explored.

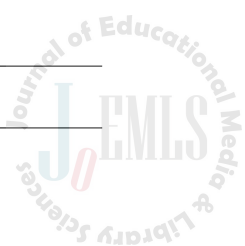
Research Method

As the function of the interface assumes active participation (i.e. selection of terms from the suggestion list) in the IR process, a traditional, batch-mode evaluation is ill-equipped to evaluate its effectiveness. Thus a novel evaluation methodology was employed to address the challenges involved in evaluating an interactive IR system such as the term suggestion interface proposed here. The most significant aspect of this methodology is that it uses real users doing real searches on real systems, instead of using assigned searches common in traditional IR evaluation.

Table 1 summarizes the research procedures and data collection instruments. In the pre-search questionnaire, participants were asked to characterize their search requests. They would then conduct the search task, which involved finding what they believed to be the 10 records most relevant to their information needs. They were also asked to write down their ideal queries at multiple points during the experiment: first before the search began, then after their session with either interface. After both search sessions for the same request had ended, participants were asked to assess the “goodness” of the queries (before and after searching) they had input and the degree of relevance of the 10 records they retrieved from each search session (20 records in total) on a 0-6 scale.

Table 1 Evaluation Procedures

Fill out pre-search questionnaire <ul style="list-style-type: none"> • Search request • Ideal query (pre-search) • Answer questions about the current search
Perform search task <ul style="list-style-type: none"> • Retrieve 10 relevant records with either interface
Fill out post-search questionnaire <ul style="list-style-type: none"> • Ideal query (post-search) • Self-assess goodness of queries (pre-and post-) • Relevance judgment of 10 records retrieved • Self-assess satisfaction with results • Perceived usefulness of interface



Experimental design

Genuine requests

Participants were asked to characterize their search requests with chosen attributes such as domain familiarity, and whether it is a new or revisited search. This provided an opportunity to investigate the interactions between these attributes and the interfaces on various performance criteria.

A repeated measure experimental design

Table 2 shows the design used in this study with the interface as the within-request factor and topic characteristics as the between-request factor. Participants were randomly assigned to one of the four groups in which they would search their own requests, starting with one interface, then moving on to the next.

Table 2 Graphic Representation of the Experiment Design

	PubMed		MAP	
1	More familiar	Less familiar	More familiar	Less familiar
2	Less familiar	More familiar	Less familiar	More familiar
	MAP		PubMed	
3	More familiar	Less familiar	More familiar	Less familiar
4	Less familiar	More familiar	Less familiar	More familiar

Research instruments

See Figure 1 for the screenshot of the interface¹:

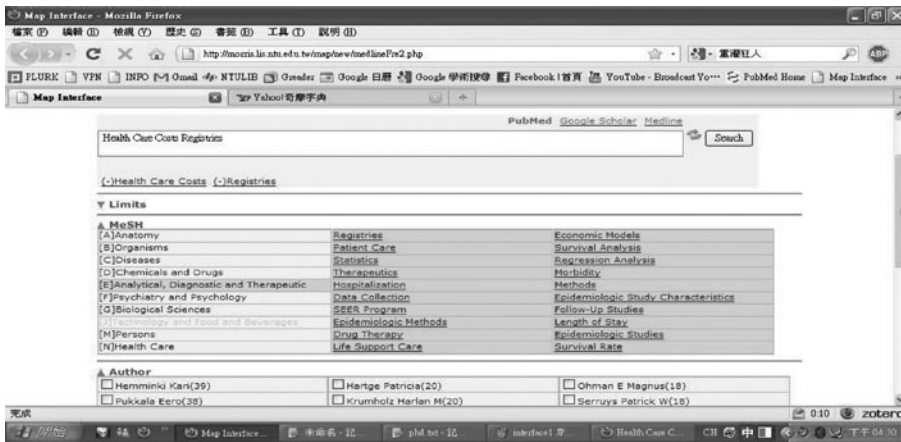


Figure 1 The MAP Interface

Results

Table 3 presents information on participants' subject discipline. All participants were graduate students from various health science related disciplines.

¹ The interface can be accessed at: <http://morris.lis.ntu.edu.tw/map/new/medlineFre2.php>

Table 3 Participants' Subject Discipline

College	Gender		Total
	Male	Female	
Medicine	3	6	9
Public Health	4	2	6
Life Science	3	4	7
Science	2	2	4
Engineering	4	3	7
Bio-Resources and Agriculture	6	3	9
Other	1	1	2
Sum	23	21	44

Table 4 compares users' querying behaviors, such as initial-final query similarity, the number of added and deleted terms, as well as the number of submissions and records browsed by users in the search process. The results show that when using MAP, users' initial and final queries (as defined by Jaccard's coefficient) were significantly less similar than when using PubMed, which indicated that the MAP interface had a large impact on user's query (re) formulation.

Table 4 Comparison of Interfaces: Users' Querying Behaviors

	Interface	Mean	SD	n
Similarity	PubMed	.54	.31	88
	MAP	.40	.27	88
Added terms	PubMed	1.57	1.54	88
	MAP	2.58	2.01	88
Deleted terms	PubMed	.60	.90	88
	MAP	.82	1.09	88
Final query terms	PubMed	3.66	1.91	88
	MAP	4.38	2.09	88
Number of submissions	PubMed	3.74	2.74	88
	MAP	5.27	2.78	88
Document browsed	PubMed	89.63	60.10	88
	MAP	92.66	63.32	88

*p<.05, ***p<.001

Table 5 compares users' judgment of the "goodness" of their final query for each interface. The results show that MAP generated queries with a higher "goodness" rating, especially when the search had not been attempted before.

Table 5 Comparison of Interfaces: "goodness" of Query Term (n=88)

	Interface	Mean	SD	n
New search	PubMed	4.21	1.10	28
	MAP	4.79	.69	28
Revisited search	PubMed	4.55	.89	60
	MAP	4.70	.87	60
Overall	PubMed	4.44	.97	88
	MAP	4.72	.81	88

We compared the mean and median rank positions of records judged to be relevant by the users when using PubMed and Map in cases where users input the same initial queries in both interfaces for the same search requests. The results showed that MAP was able to help users retrieve documents buried deep down in the returned sets (i.e. lower rank positions).

Table 6 Comparison of Interfaces: Rank Position of Initial Returned Set

User	Interface	Median	Mean	SD	Min	Max	Range	n
2	PubMed	134.5	131.17	54.25	39	197	158	6
	MAP	503	516.17	259.39	219	829	610	6
24	PubMed	52	54.80	38.17	11	139	128	10
	MAP	78.5	99.50	50.36	38	167	129	10
25	PubMed	440	487.83	182.87	345	820	475	6
	MAP	1393.5	5078.83	6513.36	366	13812	13446	6
27	PubMed	38.5	44.90	34.60	2	103	101	10
	MAP	42	54.10	44.73	7	128	121	10
29	PubMed	133	115.10	42.80	52	171	119	10
	MAP	90	114.60	58.50	58	254	196	10
30	PubMed	60	69.50	30.05	39	121	82	10
	MAP	476	483.00	207.92	194	867	673	10
31	PubMed	695.5	687.63	38.74	629	745	116	8
	MAP	1371.5	1459.13	793.21	616	2549	1933	8
32	PubMed	107	103.29	38.90	33	160	127	7
	MAP	151	162.71	97.45	33	297	264	7
44	PubMed	88	84.90	32.79	37	125	88	10
	MAP	71	141.00	198.15	26	676	650	10
46	PubMed	583.5	873.70	803.87	90	2420	2330	10
	MAP	2519	2404.30	1049.45	1063	3586	2523	10

Figure 2 and Figure 3 show the rankings of the records saved by user #2 and user #30, respectively, using PubMed and MAP. The figured demonstrated visually the capability of the MAP interface to “unearth” records buried deeper in the initial returned sets.

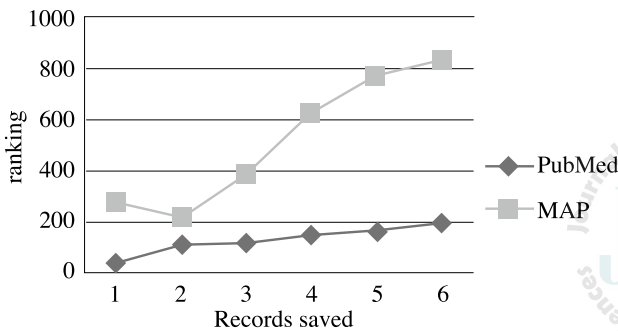


Figure 2 The Original Ranking of the Records Saved by User#2



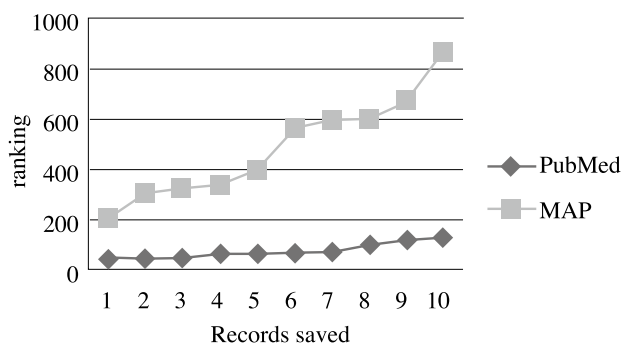


Figure 3 The Original Ranking of the Records Saved by User#30

Discussion

The results show that 1. Topic familiarity influences users' querying behaviors; MAP had a high impact on users' queries when searching for unfamiliar problems. 2. MAP was helpful, especially for search requests that the users had not attempted before. 3. MAP was most helpful in generating new research ideas. 4. MAP was able to retrieve documents buried deep in the initial returned results. Our study demonstrated the feasibility of using real participants with real information needs in the context of interactive IR evaluation. It also showed that the characteristics of users' information needs might be an important confounding factor when assessing the effectiveness of an IR system.

ROMANIZED & TRANSLATED REFERENCES FOR ORIGINAL TEXT

黃慕萱 [Huang, Mu-Hsuan](1996)。檢索系統評估 [Jiansuoxitong pinggu]。在 [In] 資訊檢索 [Zixun jiansuo](第十章, 頁269-275)[(Chap. 10, pp.269-275)]。台北市: 台灣學生 [Taipei: Studentbook]。

洪承理 [Hung, Cheng-Li](2010)。使用者查詢擴展行為研究: 以 MeSH 為詞彙來源的建議詞彙工具為例 [A user query expansion behavior study: Using MeSH as term suggestion source]。未出版之碩士論文 [Unpublished master's thesis], 台灣大學圖書資訊學系 [Department of Library and Information Science, National Taiwan University], 台北市 [Taipei]。

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. New York: ACM Press.

Bates, M. J. (1979). Information search tactics. *Journal of the American Society for information Science*, 30(4), 205-214.

Belkin, N. J. (1993). Interaction with texts: Information retrieval as information-seeking behavior. *Information retrieval*, 93, 55-66.

Belkin, N. J., Chang, S. J., Downs, T., Saracevic, T., & Zhao, S. (1990). Taking account of user tasks, goals and behavior for the design of online public access catalogs. In *Proceedings of the 53rd ASIS annual meeting* (pp. 69-79). Medford, NJ: Learned Information.

- Belkin, N. J., Cool, C., Head, J., Jeng, J., Kelly, D., Lin, S., Lobash, L., Park, S. Y., Savage-knepshield, P., & Sikora, C. (2000). Relevance feedback versus local context analysis as term suggestion devices: Rutgers' TREC-8 interactive track experience. In *TREC-8, Proceedings of the Eighth Text Retrieval Conference* (pp. 565-574), Maryland: Gaithersburg.
- Belkin, N. J., Kelly, D., Kim, G., Kim, J. Y., Lee, H. J., Muresan, G., Tang, M. C., et al. (2003). Query length in interactive information retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 205-212). New York: ACM.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38(2), 61-71.
- Borlund, P., & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3), 225-250.
- Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1), 71-90.
- Huang, C. K., Chien, L. F., & Oyang, Y. J. (2003). Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology*, 54(7), 638-649.
- Cleverdon, C. W. (1967). The Cranfield tests on index language devices. *Aslib Proceedings*, 19(6), 173-194.
- Doms, A., & Schroeder, M. (2005). GoPubMed: Exploring PubMed with the gene ontology. *Nucleic Acids Research*, 33(Web Server Issue), 783-786.
- Dumais, S. T., & Belkin, N. J. (2005). The TREC interactive tracks: Putting the user into search. In E. M. Voorhees & D. Harman (Eds.), *TREC: Experiment and evaluation in information retrieval* (pp. 123-153). Cambridge Mass: MIT Press.
- Efthimiadis, E. N. (1996). Query expansion. In M. E. Williams (Ed.), *Annual Review of Information Systems and Technology*, 31, 121-187.
- Efthimiadis, E. N. (2000). Interactive query expansion: A user-based evaluation in a relevance feedback environment. *Journal of the American Society for Information Science*, 51(11), 989-1003.
- Fidel, R. (1984). Online searching styles: A case-study-based model of searching behavior. *Journal of the American Society for Information Science*, 35(4), 211-221.
- Goetz, T., & Lieth, C. W. V. D. (2005). PubFinder: A tool for improving retrieval rate of relevant PubMed abstracts. *Nucleic Acids Research*, 33(Web Server Issue), 774-778.
- Harman, D. (1988). Towards interactive query expansion. In *Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 321-331). New York: ACM.
- Harman, D. (1992). Relevance feedback revisited. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 1-10). New York: ACM.
- Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., & Yee, K. P. (2002). Finding the flow in web site search. *Communications of the ACM*, 45(9), 42-49.

- Hersh, W., Pentecost, J., & Hickam, D. (1996). A task-oriented approach to information retrieval evaluation. *Journal of the American Society for Information Science*, 47(1), 50-56.
- Hersh, W. R., & Hickam, D. H. (1998). How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. *JAMA*, 280(15), 1347-1352.
- Huang, C. K., Chien, L.F., & Oyang, Y.J.(2003). Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology*, 54(7), 638-349.
- Ingwersen, P., & Järvelin, K. (2005). *The turn: Integration of information seeking and retrieval in context*. Dordrecht, Netherlands: Springer.
- Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. (1998). Real life information retrieval: A study of user queries on the web. *SIGIR Forum*, 32(1), 5-17.
- Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users and real needs: A study and analysis of users queries on the web. *Information Processing and Management*, 36(2), 207-227.
- Joho, H., & Jose, J. M. (2006). Slicing and dicing the information space using local contexts. In *Proceedings of the 1st international conference on Information interaction in context* (pp. 66-74). New York: ACM.
- Koenemann, J., & Belkin, N. J. (1996). A case for interaction: A study of interactive information retrieval behavior and effectiveness. In *Proceedings of the ACM SIGCHI conference on human factors in computer systems* (pp. 205-212).
- Kristensen, J., & Järvelin, K. (1990). The effectiveness of a searching thesaurus in free text searching of a full-text database. *International Classification*, 17(2), 77-84.
- Kules, B., & Capra, R. (2008). Creating exploratory tasks for a faceted search interface. In *Second Workshop on Human-Computer Interaction (HCIR 2008)*.
- Lin, X. (1999). Visual MeSH. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 317-318). New York: ACM.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York: Cambridge University Press.
- Pennanen, M., & Vakkari, P. (2003). Students' conceptual structure, search process, and outcome while preparing a research proposal: A longitudinal case study. *Journal of the American Society for Information Science and Technology*, 54(8), 759-770.
- Perez-Iratxeta, C., Bork, P., & Andrade, M. A. (2001). XplorMed: A tool for exploring MEDLINE abstracts. *Trends in biochemical sciences*, 26(9), 573-575.
- Plikus, M. V., Zhang, Z., & Chuong, C. M. (2006). PubFocus: Semantic MEDLINE/PubMed citations analytics through integration of controlled biomedical dictionaries and ranking algorithm. *BMC Bioinformatics*, 7(1), 424.
- Pollitt, A. S. (1998). The key role of classification and indexing in view-based searching. *International cataloguing and bibliographic control*, 27(2), 37-40.
- Robertson, S. E., & Hancock-Beaulieu, M. M. (1992). On the evaluation of IR systems. *Information Processing and Management*, 28(4), 457-466.

- Ruthven, I. (2008). Interactive information retrieval. In B. Cronin (Ed.), *Annual Review of Information Science & Technology* (Vol. 42, pp. 43-91). Medford, NJ: Information Today.
- Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension and applications. In *Proceedings of the American Society for Information Science* (Vol. 34, pp. 313-327).
- Saracevic, T., & Kantor, P. (1988a). A study of information seeking and retrieving. II. Users, questions, and effectiveness. *Journal of the American Society for Information Science*, 39(3), 177-196.
- Saracevic, T., & Kantor, P. (1988b). A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for information Science*, 39(3), 197-216.
- Saracevic, T., Kantor, P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for information Science*, 39(3), 161-176.
- Schatz, B. R., Johnson, E. H., Cochrane, P. A., & Chen, H. (1996). Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval. In *Proceedings of the first ACM international conference on Digital libraries* (pp. 126-133). New York: ACM.
- Shiri, A., & Revie, C. (2006). Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation. *Journal of the American Society for Information Science and Technology*, 57(4), 462-478.
- Shiri, A., Revie, C., & Chowdhury, G. (2002). Thesaurus-assisted search term selection and query expansion: A review of user-centred studies. *Knowledge Organization*, 29(1), 1-19.
- Spink, A. (1995). Term relevance feedback and mediated database searching: Implications for information retrieval practice and systems design. *Information Processing and Management*, 31(2), 161-171.
- Spink, A., & Saracevic, T. (1997). Interaction in information retrieval: Selection and effectiveness of search terms. *Journal of the American Society for Information Science*, 48(8), 741-761.
- Su, L. T. (1992). Evaluation measures for interactive information retrieval. *Information Processing and Management*, 28(4), 503-516.
- Su, L. T. (1994). The relevance of recall and precision in user evaluation. *Journal of the American Society for Information Science*, 45(3), 207-217.
- Su, L. T. (2003). A comprehensive and systematic model of user evaluation of web search engines: II. An evaluation by undergraduates. *Journal of the American Society for Information Science and Technology*, 54(13), 1193-1223.
- Tang, M. C. (2006, November). Multiple accesses to PubMed: A proposed user study of term suggestion tools for PubMed. In *ASIS&T (American Society for Information Science and Technology) 2006 annual meeting*, Symposium held at the Nation Taiwan University, Taiwan .
- Tang, M. C. (2007). Browsing and searching in a faceted information space: A naturalistic study of PubMed users' interaction with a display tool. *Journal of the American Society for*

Information Science and Technology, 58(13), 1998-2006.

- Tang, M. C., Wu, W. C., & Hung B. W. (2009). Evaluating a metadata-based term suggestion interface for PubMed with real users with real requests. In *Proceedings of the Annual Meeting of the American Society for Information Science & Technology 2009* (pp. 35-42).
- Vakkari, P. (1999). Task complexity, problem structure and information actions integrating studies on information seeking and retrieval. *Information Processing and Management*, 35(6), 819-837.
- Vakkari, P. (2000). Cognition and changes of search terms and tactics during task performance: A longitudinal case study. In *Proceedings of the RIAO 2000 Conference* (pp. 894-907).
- Vakkari, P. (2001). A theory of the task-based information retrieval process: A summary and generalisation of a longitudinal study. *Journal of Documentation*, 57(1), 44-60.
- Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, 56(5), 540-562.
- Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing and Management*, 39(3), 445-463.
- White, R. W., Kules, B., Drucker, S. M., & Schraefel, M. C. (2006). Supporting exploratory search, introduction, special issue, communications of the ACM. *Communications of the ACM*, 49(4), 36-39.
- White, R. W., Ruthven, I., & Jose, J. M. (2005). A study of factors affecting the utility of implicit relevance feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 35-42). New York: ACM.
- Xie, I. (2008). *Interactive information retrieval in digital environments*. Hershey: IGI Publishing.

