

系統評估研究之文獻探討：
從電子圖書館角度考量
Literature Review on System Evaluation Research：
From the Perspective of Digital Library

張嘉彬 Chia-bin Chang

國立台灣大學法學院圖書分館館員

Librarian, Law Library, National Taiwan University

E-mail: r5126006@ms.cc.ntu.edu.tw

【摘要】

本文以文獻分析法分析系統評估之研究。首先介紹系統評估之標準、系統評估之方法、系統評估之大型實證研究，以做為電子圖書館評估研究之基本背景。其次介紹電子圖書館之評估研究，並將此部分之焦點集中於學者專家們所提及的電子圖書館評估指標。最後，並針對電子圖書館評估研究提出建議，包括系統評估宜從使用者角度進行評估、系統評估應考慮系統週邊人員之觀點、系統評估應廣及各層次之評估、系統評估應採用量化及質化兼具之方法及從事電子圖書館之使用者研究。

【Abstract】

This paper uses literature review method to analyze system evaluation research. It introduces researches on system evaluation criteria, system evaluation methods, and some empirical digital library evaluation researches. Finally, this article suggests that digital library evaluations should consider both system and user perspectives.

關鍵詞：電子圖書館評估；系統評估；電子圖書館；

Keywords: Digital library evaluation; System evaluation; Digital library

壹、前言

資訊檢索系統評估在資訊科學中是一門歷史悠久的研究領域，至今已有40多年歷史。較早的評估研究從50年代即已開始，但直到1960年接受美國科學基金會（National Science Foundation）贊助後，相關研究領域才漸趨成熟。（註1）早期學者進行系統評估研究時，多以回現率及精確率為評估標準，隨著評估實驗的進行，也將評估標準擴展至了包含回應時間、顯示格式、讀者找到相關文獻所花費的精力及成本因素等標準上。至於系統評估方法，也從早期的實驗法，逐漸融合問卷法、訪問法、觀察法及查詢過程記錄分析等方法。而系統評估實證研究，從大家耳熟能詳的Cranfield研究，到MEDLARS、SMART、STAIRS及TREC進行的一連串評估實驗，其研究成果已逐漸開花結果。

本文以文獻分析法分析「系統評估」之概念，共分為如下四部份進行探討：第一部份是系統評估之標準；第二部份是系統評估之方法；第三部份是系統評估之大型實證研究；第四部份是電子圖書館之評估研究。最後，並提出對電子圖書館評估研究之建議。

貳、系統評估之標準

系統評估標準一直是待突破的研究領域，有些學者主張用回現率及精確率評估系統即已足夠，有些學者則認為只利用此兩個測量值顯得相當不

足，必須輔以其他的測量標準。綜合文獻，可將這些評估標準分為兩類：第一類是針對線上資訊檢索系統（包含一般資訊檢索系統及線上公用目錄）的標準而論，第二類是就網站（Web site）檢索引擎（search engines）的標準而論。

一、線上資訊檢索系統

1969年，F. W. Lancaster進行MEDLARS系統評估，其評估標準如下：

1.資料收錄範圍 2.回現率 3.精確率 4.回應時間 5.顯示格式 6.讀者找到相關文獻所花費的精力。（註2）

1970年，B. C. Vickery整理系統評估標準，將其分為兩類，第一類是與資訊可得性（availability of information）相關的標準，包括資料收錄範圍、回現率及回應時間。第二類是與輸出結果選擇性（selectivity of output）相關的標準，包括精確率、文獻的可使用性及其顯示格式。（註3）

1971年，Allen Kent提出以使用者觀點評估資訊檢索系統，他綜合Lancaster及Cyril Cleverdon的意見，認為系統評估標準應有下列六項：1.回現率 2.精確率 3.讀者找到相關文獻所花費的精力 4.回應時間 5.提供豐富資訊的能力（informativeness）及顯示格式 6.文獻輸入資料庫內之規則。（註4）

1979年，Eva L. Kiewitt評估ERIC的PROBE程式，此程式是1970年由Ronald Tschudi所發展。他提出的評估標準共可分為三個面向，第一個面向

是使用者研究，包括使用者分析、使用目的、回應速度及讀者滿意程度等。第二個面向是系統檢索表現，包括回現率、精確率、檢索成功／失敗分析、資料收錄範圍、新穎性及資料正確性等。最後一個面向則是成本研究，包括人事、設備及電腦檢索費用等成本分析。（註5）

1981年，Carol Hansen Fenichel評估ERIC資料庫，主要目的在比較具有不同線上檢索經驗的個人於檢索上之表現，使用的評估標準為回現率、精確率及單元成本（檢索所得相關文獻除以連線時間）。（註6）

1988年，Geraldene Walker 針對BRS/After Dark系統進行評估，由兩組受試者進行實際檢索。一組是資訊需求者，一組是資訊中介者。評估依據是讀者判斷檢索結果的相關程度，而使用的評估標準包括檢索效率（回現率及精確率）及檢索效能（時間及成本效率分析）。（註7）

同年，D. Alasdair Kemp綜合多位學者意見，提出評估一般資訊檢索系統的標準。這些標準分別為：1.系統功能 2.資料收錄範圍 3.資料正確性 4.檢索效益 5.文獻的可使用性 6.顯示格式 7.減少讀者找尋資料的時間 8.回應時間 9.系統特殊協助 10.系統可信賴性 11.成本 12.安全性 13.系統彈性 14.文獻輸入資料庫花費的精力。（註8）

1996年，William Hersh及Jeffrey Pentecost針對2個不同介面的MEDLINE系統進行評估。其中一系統在索引典

上加上布林邏輯功能進行檢索；另外一系統則使用自然語言於題名、摘要及索引詞彙上進行檢索。實際的評估標準包括讀者是否成功地回答問題、讀者是否肯定地回答問題、回答的時間、找出相關文章的能力及使用者介面的滿意程度。（註9）

在國內方面，蔡明月教授認為線上檢索評估主要涉及效果（effectiveness）、成本效率（cost-effectiveness）及成本效益（cost-benefit）三大準則。效果準則乃針對檢索花費、檢索時間及檢索品質加以探究。成本效率準則主要在探究每檢索出一篇相關文獻之書目資料或原文資料的成本。成本效益則在研究整個線上系統的運作與其所產生價值之評估。她並提出評估回現率與精確率、資料庫主題範圍的完整性、評估調查及檢索者評估作為系統評估之標準。（註10）

民國81年，周曉雯綜合系統評估文獻，認為線上檢索結果之評估可從三方面探討：1.相關性之評估：包括回現率、精確率、誤引率（fallout ratio）、普遍率（generality ratio）及專指程度（specificity）等 2.檢索效率之評估：包括成本、成本效率及成本效益等 3.其他：包括資料收錄範圍、資料正確性、反應時間及新穎率等。（註11）

民國85年，黃慕萱教授綜合學者及個人意見，認為系統評估標準可從下列幾方面探討：1.回現率 2.精確率 3.檢索效益（每篇相關文章所花費之時間、金錢） 4.檢索系統（親和力、

回應時間、顯示格式及資料收錄範圍)
5. 檢索者 (個人能力及檢索技巧) 6.
新穎程度 (檢索出讀者事先不知道的
相關文章數目) 7. 讀者對檢索之主觀
滿意程度 8. 錯誤率 9. 猶豫程度 (線上
停頓時間和線上移動時間之比值)。
(註12)

同年，卜小蝶教授認為，評估資
源搜尋系統可由以下五方面進行評估：
1. 收錄範圍：包含資料量、收錄項目、
索引深度、新穎性及更新頻率、索引
建立方式、多國語言處理及提供評論。
2. 檢索功能：包含布林邏輯、複層布林
邏輯、切截、相近運算元、片語查詢、
限制欄位、大小寫及特殊符號等處理、
關鍵詞、自然語言輸入、修正檢索
(refine search)、重複偵測及相關回
饋 (relevance feedback) 等。3. 檢索結
果展示：包含相關性排列、限制顯示
筆數、顯示資料的詳細程度及直接連
結至資源 (direct links to resources) 等。
4. 檢索效率：包含查獲率、精確率、反
應時間及連線容易程度。5. 使用者介面：
包含直觀上判斷其是否容易使用、是
否提供線上輔助說明、是否提供常問
問題集 (Frequently Asked Questions，
簡稱FAQ) 及是否考慮到文字模式的
使用者。(註13)

二、網站檢索引擎

網站之檢索引擎直到1994年才開
始出現，因此關於評估網站檢索引擎
的文章從1995年後才陸續出現。現就文
獻所得分析如下：

1995年，Martin P. Courtois、
William M. Baer及Marcella Stark三位學
者，針對CUI W3 Catalog、Harvest、
Lycos、Open Text Web Index、
WebCrawler、World Wide Web Worm及
Yahoo等七個檢索引擎進行評估，評估
標準包括文件說明 (資料庫、檢索能
力之介紹及範例說明)、資料庫建構
(自動或手工)、資料庫大小、欄位
索引情況 (索引URL、題名或全文)、
檢索能力、顯示格式、更新頻率及回
應時間等。(註14)

1996年，Heting Chu及Marilyn
Rosenhal兩位學者，針對Altavista、
Excite及Lycos等三個檢索引擎進行評
估。評估標準包括三者之檢索能力
(包括布林邏輯、切截、欄位檢索及
字詞檢索等) 及檢索效益 (包括精確
率及反應時間)。作者並草擬了一評
估檢索引擎之標準，包括Web索引的
情況、檢索能力、檢索效益、顯示格
式及讀者找到相關文獻所花費的精力。
(註15)

同年，Wei Ding及Gary Marchionini
兩位學者，針對InfoSeek、Lycos及Open
Text等三檢索引擎進行評估，評估方
式是以五個檢索問題進行評估。其中
三個問題是由作者所授資訊科學課程
Dialog線上檢索作業隨機選擇而來，另
外兩個問題是依據讀者個人興趣所形
成。評估標準包括精確率、資訊顯著
程度 (salience) 及相關萃取程度 (re-
levance concentration)。(註16)

1997年，Xiaoying Dong及Louise Su

兩位學者，進行評估檢索引擎文獻回顧。他們綜合先前學者所使用的評估標準如下：使用者介面、顯示格式、精確率、回應時間、資料收錄範圍、URL連結的有效性、檢索能力、檢索特色、更新頻率、資訊顯著程度、排序輸出、欄位索引情況、資料庫建構、相關萃取程度、檢索結果的完整性、成本、檢索結果的重複性及回現率。此外，兩位學者並建議，今後研究應特別著重於從使用者角度進行評估。（註17）

在國內方面，陳秀美針對Infoseek、Lycos、Open Text、Web Crawler、Yahoo、Altavista、Inktomi、Excite、Magellan、GAIS及YAM Web Navigator-The Taiwan Index等11個檢索引擎進行評估，其評估標準如下：1.資料收錄範圍：包括網路資源包含的內容、資料的類

別、更新頻率與增加的速度及索引的深度等 2.檢索的方法：包括布林邏輯、相近運算元、字尾切截、檢索欄位及資料來源等 3.使用者介面 4.檢索效率：包括反應時間、查獲率及精確率 5.顯示格式：包括選擇每次檢索結果的展示筆數、選擇展示文件的詳細程度、提供URL連結、展示檢索結果的先後順序、提供計分或評論及檢索結果的修正。（註18）

綜合上述學者專家們所提出之線上資訊檢索系統及搜尋引擎評估標準，研究者將之歸納整理成下表一。其中較常被學者專家們提及的標準包括回應時間、精確率、檢索結果展示、回現率、資料收錄範圍、新穎性、檢索功能及成本研究等。

表一 系統評估標準

提出者 項目	Lancaster	Kiewitt	黃慕萱	卜小蝶	Courtois、 Baer & Stark	Chu & Rosenhal	Dong & Su	陳秀美
資料收錄範圍	√	√	√	√			√	√
檢索功能				√	√	√	√	√
回現率	√	√	√	√			√	√
精確率	√	√	√	√		√	√	√
回應時間	√	√	√	√	√	√	√	√
檢索結果展示	√		√	√	√	√	√	√
使用者研究		√	√					
成本研究	√	√	√			√	√	
資料新穎性		√	√	√	√		√	√
資料正確性		√						
檢索者			√					
錯誤率			√					
猶豫程度			√					
易於使用				√			√	√
輔助說明				√			√	√
容易連線程度				√				
資料庫簡介					√			
資料庫建構					√		√	
資料庫大小					√			
欄位索引情況					√	√	√	√
連結性							√	√

參、系統評估之方法

進行系統評估可使用的方法很多，不過一般研究者較常採用的有問卷法、訪問法、觀察法、查詢過程記錄分析法及實證研究法等。有些學者進行系統評估時，會同時使用多種方法，有些則僅使用一、兩種。現就文獻所得分析如下：

1983年，Ray R. Larson 及 Vicki Graham 評估 MELVYL 系統，其使用的評估方法包括查詢過程記錄分析及線上問卷。使用查詢過程記錄分析的原因在於可以得到足夠資訊以重新建構讀者和系統之間的互動，包括使用指令、錯誤分析及系統回應等。而使用線上問卷之因主要可以獲知讀者對系統的意見。（註19）

1989年，Ben Shneiderman、Dorothy Brethauer、Catherine Plaisant 及 Richard Potter 等四位學者評估超媒體系統 Hyperties 在三個博物館的使用情形，此三個博物館分別為 B'nai B'rith Klutznick、International Center of Photography 及 Smithsonian Institution's National Museum of Natural History。此評估研究可分為兩部分。第一部分是學者評估前兩個博物館超媒體系統，僅使用查詢過程記錄分析進行；第二部分是學者評估第三個博物館超媒體系統混合使用觀察法、晤談法及更深層的查詢過程記錄分析進行。評估結果顯示，後者的方法（觀察法、晤談法及更深層的查詢過程記錄分析）在分析資料

使用情形、使用者介面及資料庫結構改善上，都是較合適的方法。（註20）

1991年，Micheline Hancock-Beaulieu、Stephen Robertson 及 Colin Neilson 等三位學者評估 CLSI CL-CAT 線上公用目錄。使用的評估方法包括全螢幕查詢過程記錄分析、互動式問卷（包含檢索前和檢索後問卷、線上作業或線下作業問卷及檢索中間卷）、回顧檢索（search replay）及大聲思考（talk-aloud）。（註21）

1995年，學者 Egbert De Smet 針對丹麥一公共圖書館之電腦化社區資訊系統進行評估。評估方法是使用查詢過程記錄分析及讀者調查，其中讀者調查包含線上問卷及抽樣讀者進行問卷調查。（註22）

同年，Micheline Beaulieu 及 Victoria Mellor 之研究針對倫敦國家畫廊之西歐繪畫超媒體系統 Micro Gallery 進行評估，目的是希望檢視使用者介面對檢索行為的影響。所使用的評估方法包括系統使用前、後之問卷調查及直接觀察伴隨大聲思考（註23）

1997年，Michael Dolenko 針對加拿大國家統計局（Statistics Canada）的網站進行評估，目的是希望能夠幫助網站重新設計時，節省系統設計者所花費的心力。他所使用的評估方法包括查詢過程記錄分析、線上問卷、焦點團體法及觀察法等。（註24）

同年，William E. Moen 等多位學者針對政府資訊定位服務系統（Government Information Locator Services，

簡稱GILS)進行評估。其評估的時間為1996年9月至1997年5月，所採用的評估方法包括參觀網站之經驗(site visit)、焦點團體法、調查法、內容分析法、查詢過程記錄分析法、使用者填答評估(scripted user assessment)及政策檢視(policy review)等。(註25)

在國內方面，關於系統評估方法的文章並不多，目前就文獻所得僅得三篇。民國82年，王喜沙綜合線上公用目錄評估方法，將OPAC之評估方法分為下列三種：1.查詢過程記錄分析 2.調查法、問卷法及訪問法 3.觀察與實驗控制法。(註26)

邱韻鈴在其碩士論文以查詢過程記錄分析國立清華大學讀者使用線上公用目錄之情況。調查時間從八十二年十一月一日至十一月七日，共蒐集3,663筆查詢過程資料。調查結果顯示，

清華大學線上公用目錄使用者最常使用的檢索點是關鍵字，得到零筆結果的比例為31.04%，主要原因係清華大學的系統設計不當，而使用者也不了解系統的輸入規則。(註27)

民國85年，林麗娟教授提出以問卷調查、查詢過程記錄或線上調查、定點式的實地觀察及口語分析(protocol analysis)進行系統評估，並且特別強調以使用者為本之資訊系統評鑑理念。她並建議以交叉評估的方法進行系統評估，以提升評估的品質。(註28)

綜合系統評估之方法主要可分為三類，分別為查詢過程記錄分析、一般調查研究(調查、問卷及訪問法)及觀察與實驗控制法。茲將此三類評估方法之優缺點及特色以表二表示。(註29)

表二 系統評估方法之優缺點

優缺點 評估方法	優點	缺點
查詢過程記錄分析	<ol style="list-style-type: none"> 1.使用Transaction logs法，研究人員可以類似無感式(unobtrusively)的觀察方式來觀察讀者的行為，使其在查詢過程中不受打擾。 2.可提供一些實際的查詢記錄，了解讀者真正進行了那些查詢活動。 3.記錄下來的查詢過程資料可以重複觀察讀者與系統之間的互動關係。 4.可被用於實驗性的研究，也可用於實地的研究（field study） 5.由於查詢過程記錄是由系統軟體記錄，故此評估方法成本不高，具經濟效益。 	<ol style="list-style-type: none"> 1.大部分Transaction logs處理無法掌握個別使用者於查詢進行時產生的不同變數。如：讀者以匿名的方式進行查詢時，系統軟體若無法辨識時，則其記錄將會影響研究者了解不同讀者特性的機會。 2.無法真正代表讀者的查詢意向、查詢結果是否滿意及查詢過程中發生的困難或問題。 3.分析過程須花許多時間。 4.須有一功能很強的軟體來處理大量的資料。 5.由於各種檢索系統的檢索結構、提供的指令、協助等不盡相同，各查詢過程記錄軟體所包含的項目也不同，因此查詢過程記錄分析大多只用於單一系統的分析，較無法同時比較兩個以上的系統優劣，也難以將某系統所得的研究結果套用在其他系統之上。 6.查詢過程記錄分析是一種相當量化的研究方法，因此有學者批評其不夠深入。

(接下頁)

(承上頁)

表二 系統評估方法之優缺點

一般調查研究 (調查、問卷 及訪問法)	1.較適合於蒐集大量資料—可大規模抽樣個人或團體作為調查的對象，以便蒐集大量資料。 2.調查研究在成本的考量上較經濟—若在樣本抽樣上具代表性，則可以節省許多的時間及花費。	1.調查研究僅及表面不夠深入，其無法深入了解內層的問題，特別是有關讀者的問題及意見的探討上。 2.調查研究是樣本的調查，易有樣本誤差及是否具代表性的問題，此會嚴重影響調查結果的正確性。
觀察與實驗控制法	觀察法： 真正應用觀察法於OPAC研究者並不多，通常實際觀察讀者與系統的互動，大部分屬於無感式的觀察。讀者沒有感受到自己被觀察，較能表現實際的行為。 實驗控制法： 實驗控制法最主要的特徵是對某些特定的研究項目，控制其於特定情況下而觀察受試者的反應與變化。	觀察法： 觀察法涉及的主觀因素較多，其信度與效度的控制不易。 實驗控制法： 1.需設計測試的情境與問題的內容，而測試的情況與實際情形往往會有誤差。 2.重複使用同一樣本，使得樣本備受折磨，如：有些實驗經常以一組受試者來測試不同系統間的功能或不同的檢索問題。

肆、系統評估之大型實證研究

系統評估實證研究雖從50年代即已開始，但直到1957~1966年Cleverdon完成Cranfield研究後，才算是第一個真正大型的系統評估實證研究。爾後，Lancaster進行MEDLARS系統評估、Gerard Salton進行SMART系統評估、

David C. Blair及M. E. Maron進行全文檢索系統STAIRS評估及TREC所進行的評估研究，其研究成果可說已相當豐富。回顧這幾個大型的系統評估實證研究，可以了解在系統評估史上的一些重要里程碑及其重要成果。

Cranfield研究在系統評估的歷史上是一件家喻戶曉的事，其研究共分為前後兩期，而研究時間分別為1957年

及1966年，皆由Cleverdon所完成。(註30)一般將前期的研究稱為Cranfield I，而將後期的研究稱為Cranfield II，此前後兩期的目的都是為了找出較佳之索引語言。Cranfield I 僅測試四種索引語言，分別為國際十進分類法、層面分類法、按字母順序排列之主題索引及單詞組合索引法 (uniterm system of coordinate indexing) 四種。而用來評估系統的館藏是Western Reserve University 圖書館中的1,100篇冶金學文獻，評估系統的標準是回現率及精確率。評估結果顯示，此四種索引語言的檢索表現難分軒輊，Cleverdon認為因為此四種索引語言都同時是提高回現率及精確率所設計出之混合產品，因此很難測試出其對檢索表現所產生之影響。(註31)

Cranfield II主要在測試下列三類語言所產生之33種索引語言，包括單一詞彙語言、簡單概念索引語言 (single concept language) 及控制詞彙索引語言。評估系統的館藏是Cleverdon所蒐集之有關太空動力學 (aerodynamics) 的279個問題及1,400篇文獻 (註32)，而使用的測量值是常態化回收率。鑑於Cranfield I未能找出最佳之索引語言，Cleverdon希望能將各種提高回現率及精確率的方法 (如切截、同義詞、類同義詞等) 抽離出，並做實際的結合，以測量每一種方法對檢索表現的影響。評估結果顯示，這三類檢索語言以單一詞彙語言的總體表現最好，其次是控制詞彙索引語言，表現最差的是簡單概念索引語言。Cleverdon認為造成

此現象的原因主要是控制詞彙索引的專指程度及詳細程度 (exhaustivity) 不夠。(註33)

Lancaster於1966年8月至1967年7月進行了相當著名的MEDLARS系統評估研究，此資料庫主要是蒐集有關生物醫學方面的期刊文章。共有二十一個使用者組群 (user group) 參與研究，分析了302個檢索問題。而使用的評估標準為資料收錄範圍、回現率、精確率、回應時間、顯示格式及讀者得到相關文獻花費的精力。系統操作者在接到檢索問題時，先將檢索敘述轉成MeSH (Medical Subject Heading) 所使用的詞彙，並將檢索所得結果 (即參考書目) 交由資訊需求者評估。由於參考書目相當多，因此利用隨機抽樣僅選擇25筆至30筆進行相關判斷。在302個檢索問題中，得到的平均回現率是57.7%，精確率是50.4%。(註34)至於造成回現率及精確率不佳的原因，主要可以歸因於索引語言上 (如錯誤的詞彙組合)、索引上 (如索引過於詳細)、檢索上 (如檢索詞彙太過專指) 及人機互動不良等原因 (註35) 且其中23%的回現率及37%的精確率不佳是導因於「人機互動不良」 (inadequate user-system interaction) 的結果。(註36)

Salton 於 1960 年代初開始進行SMART系統評估研究 (註37)，所使用的資料庫是以Cranfield II計畫中所蒐集的實驗性館藏為本 (註38)，而其主要目的在證明圖書館界長期使用的內容分析方法，應比一般關鍵字檢索

能達到更好的檢索效益。(註39) 1965年，初期的研究成果推翻了其原先的預測和假設，加權關鍵字加上切截後的表現相當突出，而一些複雜的內容分析方法，其表現卻遠較預期為差。(註40) 1960年底，資料庫內的文獻增加到了1,000篇，Salton再次進行SMART系統之評估。此次評估研究主要在比較題名和摘要檢索、加權詞彙和未加權詞彙及單一詞彙加切截和索引典關係之檢索表現。評估結果顯示，摘要在檢索上比題名好，而加權詞彙的表現也較未加權詞彙為佳，而單一詞彙和切截結合之表現並不比複雜的索引典關係遜色。(註41) 1969年，Salton企圖比較SMART關鍵字自動索引方式和MEDLARS人工索引方式在檢索效益上的差異。評估結果發現，SMART自動索引的表現並不會比MEDLARS人工索引的表現來的遜色。(註42) 1983年起，SMART系統的實驗性館藏由1,000篇文獻增加到50,000篇文獻，而評估的方法也擴增到包括群聚檔案組織(cluster file organization)、相關回饋(relevance feedback)及向量空間(vector space)等，即Salton已逐漸將其研究轉向檢索方法對於檢索效益之影響。(註43)

1985年，Blair及Maron針對全文檢索系統STAIRS(STorage And Information Retrieval System)進行評估。他們在STAIRS上儲存了40,000篇文章，大約是350,000萬頁的全文資料，主要作為大型律師事務所法律辯護之用，而其使用的評估標準為回現率及精確率。

評估結果顯示，平均回收率是0.20，而精確率則為0.79。作者對於檢索結果並不是相當滿意，尤其是回收率只有0.20，因此試圖找出為何回收率只有0.20的原因。作者認為律師並沒有意識到低回收率的問題，他們甚至以為其回收率高達0.80，因此他們對檢索結果相當滿意。其主要原因在於律師只關心他們所檢索到的相關文章，而不在乎整個資料庫中其他的相關文章。(註44)

1989以來，Su進行一連串有系統評估的實證研究，她的目的是想找出評估資訊檢索系統的合適指標。在本研究中，作者選擇了20個關於系統評估的指標，包括下列四類評估標準：第一類是相關的標準，只包括精確率。第二類是效用(utility)的標準，包括以金錢、花費時間、花費體力及花費心力來衡量，其檢索結果的價值及檢索整體結果之價值。第三類是讀者滿意程度的標準，包括檢索者的貢獻、檢索晤談的過程及檢索結果。最後一類是效能(efficiency)的標準，包括檢索時間、相關評估時間及實際檢索花費。實驗結果顯示，檢索整體結果之價值(value of search result as a whole)是20個指標中最適合評估資訊檢索系統的單一指標。而傳統系統評估中最重要之評估標準—精確率，與檢索成功之間的關係並不高，使用者反而比較關心完全的回現率(absolute recall)。而時間相關的標準，如檢索時間及相關評估時間等，與檢索成功並無太大的相關性。(註45)

由於評估檢索表現所使用的指標—回現率及精確率只是資訊尋求中之一環，再加上早期所使用的測試集（test collection）規模均不大，且大多數是使用同質性較高之文件集，因此起源於1992年的文件檢索會議（Text Retrieval Conference，簡稱TREC）便是在此種情況下產生，試圖解決如上之缺失。（註46）TREC是由美國國家標準暨技術局（National Institute of Standards and Technology，簡稱NIST）及美國國防部高等研究計畫局（Defense Advance Research Projects Agency，簡稱DARPA）所共同舉辦。（註47）從1992年舉辦了第一屆，其後持續在每年年底舉辦會議，至1998年底已進行了七屆。除了與會者依據大會提供的測試集送回各測試項目的資料進行評估之外，尚有一個為期三天的研討會，與會者可以在會中發表系統的架構，評估結果，並相互討論切磋。（註48）

TREC的主要目標有如下六項：（註49）

- 1.以大型測試集為基礎，鼓勵文件檢索的研究。
- 2.創造一個意見交流的開放性論壇，以增加產業界、學界及政府之間的溝通。
- 3.經由對真實檢索環境的模擬與實質的論證，加速將實驗室研究技術轉移為可運作的系統。
- 4.發展適當且具應用性的評估技術，以供產業界及學界採用。
- 5.發展各種不同的測試項目，希望能

在一致的模式中對各種檢索技術進行評估。

6.提供一個一般項目的評估，以允許跨系統（cross-system）之比較。

每一次TREC會議總是環繞著兩項主要工作項目，即所謂的routing項目及ad hoc項目。並且自TREC-4開始，正式產生一些正式的測試項目，稱為tracks。Routing項目，其概念類似圖書館學中的資訊選粹服務（SDI），主要目的在探討利用穩定的測試主題檢索新文件的能力；而ad hoc項目，主要目的在探討系統使用新主題檢索固定文件集的效益。至於TREC所使用的系統評估標準，主要還是傳統的回現率及精確率。但是有些測試項目，仍有其特殊的評估標準。（註50）

TREC所產生的影響相當深遠，這可由如下三方面來分述：（註51）

- 1.在測試集方面：TREC測試集至今已有5GB的文件集以及350個具有相關判斷的主題，並已被整個文件檢索的研究社群所廣泛採用。這樣一個大規模測試集的可得性，使得研究者得以發展出更符合真實運作環境的檢索系統，也由於此測試集的發展目標為模擬真實的檢索環境，因此檢索系統亦可測試新技術在真實環境中可能的成效，增進其原有技術，提昇檢索效益。
- 2.在測試項目方面：TREC持續地致力於研究發展許多新的測試項目，使得不同的檢索技術均能在一致的測試環境中進行評估，也將傳統的文

件檢索研究領域擴展至新的領域。

- 3.在會議及論壇分面：TREC的舉行使得系統能透過系統的測試以及研究者相互間的觀摩切磋，使其檢索技術得到改良，並獲致更高的檢索效益。會議的影響力亦使許多新的優良技術迅速傳播而成爲基本的檢索技術之一。另外，許多TREC的參與者開始習慣利用此會議的進行作爲其研究發展的趨力。

伍、電子圖書館之評估研究

從1975年電子圖書館此名詞被提出後，此名詞便甚囂塵上充斥於整個圖書資訊界，而關於電子圖書館的研究也越來越多。簡單來說，電子圖書館應該是從實體圖書館的延伸或是從資訊檢索系統的延伸，在多媒體、分散式以及協力的網路環境中，提供以使用者爲導向的資源之儲存、搜尋、處理與檢索的功能。（註52）由於電子圖書館與傳統圖書館在館藏發展、採訪、編目及索引、借閱及讀者服務等方面的不同（註53），並且電子圖書館與一般資訊檢索系統性質不同，因此用於評估傳統圖書館與一般資訊檢索系統的指標是否也適用於電子圖書館，許多學者展開了熱烈討論，並且積極從事相關的研究。

電子圖書館的評估研究如果從 American Memory 的評估研究算起（註54），至今已將近有十年的時間，但由於科技進展的關係，其研究也是從數位圖書館先導研究計畫（The

Digital Library Initiative）（註55）後才慢慢趨多。

早期的評估研究很多都是根據電子圖書館的雛形系統進行評估。並且，大部份的評估研究都是針對電子圖書館之使用者進行評估，而非針對其系統進行評估，因此關於評估電子圖書館的指標，目前尚無定論。由於評估電子圖書館的相關文獻不多，因此研究者也將網站評估的文獻納入其中，並且將此節的焦點集中於學者專家們所提及的電子圖書館或網站之評估指標。現就文獻所得分析如下：

American Memory 隸屬於美國國會圖書館（Library of Congress, LC），其基本概念在推廣文獻電子化，並針對美國歷史、文化或罕有文獻進行整理。早在1989至1990年間，American Memory 已經先行從事過二次小型的評估研究，而本研究是1991至1993年所從事的二年大規模評估研究。本研究以44個各類型的圖書館爲對象，目的是決定核心使用者，並了解其需求及期望。評估目標主要有如下五項：1.誰是核心使用者 2.何種資料被使用，並應增添什麼資料 3.資料如何被使用 4.如何支援資料的利用 5.American Memory 的使用性如何。

本研究共包含量化及質化的研究方法，前者包括使用記錄檔（21/44）和使用者問卷（1801份），而後者包括電話訪談、個別訪談（包含55位職員、121位使用者）和觀察法（資料庫瀏覽）等。具體之結論包括如下八項：

1. 中等學校是核心使用者，但隨著內容成長，希望將使用對象擴及高等教育群，以支援他們從事學術研究。
2. 資料庫內容在各類型圖書館中是評估的最重要因素。
3. 對於資料庫收錄範圍及內容的描述，是每個系統的基本要求，因為如此可讓使用者對文獻的完整性及出處有所概念。
4. 使用者希望自己上系統查閱資料，因此建議American Memory只需說明基本概念及展示方法即可。
5. 一般使用者傾向線上求助或詢問館員，因此系統需要完整的索引及方便指導的系統說明書。
6. 考慮CD-ROM系統的使用性。
7. 介面簡化和搜尋引擎的強化。
8. 擁有對資料進行整合的能力，以因應個別使用者的需求。（註56）

1995年，Anne Peterson Bishop以身為數位圖書館先導研究計畫使用者研究小組的成員，提出其工作報告。此小組為一非正式組織，並且一年舉行兩次會議。在1995年春季的會議，他們討論了電子圖書館使用者評估的基本方法，並且對於達到不同目標使用不

同層次的評估有了共識。總結他們所提出的使用者研究之評估層次總共有四：

1. 館藏的合適性、系統功能、介面及使用性。
2. 檢索表現及行為。
3. 對於使用者工作改變的影響。
4. 對於公共政策的影響。（註57）

同年，June P. Mead及Geri Gay利用概念比對法（concept mapping）進行電子圖書館之系統設計及評估。其概念比對的程序分為三階段：一、發展互動式多媒體小組（Interactive Multimedia Group, IMG）的概念圖（主要是發展電子圖書館檢索工具的概念圖）；二、建構數位圖書館工作小組（Digital Library Working Group, DLWG）的概念圖（主要是發展評估數位圖書館使用者介面的概念圖）；三、建構MoA（The Making of America）電子圖書館之概念圖（進行中）。

目前本研究只進行至互動式多媒體小組的概念圖及數位圖書館工作小組的概念圖類型比對(pattern matching)的階段。以下為此兩者類型比對之結果：（註58）

數位圖書館工作小組概念圖	互動式多媒體小組概念圖
一般特色	個人化的參考的工具
	排序功能
	修正工具
	更高階檢索工具
檢 索	檢索及瀏覽工具
	其他檢索工具
檢索結果展示	介面展示功能
文件展示	
航 行	個人化的航化
可得性／適用性	易於使用
	使用者的因素
	資源分享
列印及轉錄	個人化的輸出方式
登入及監測	視覺／使用者方面
版 權	沒有比對到

1997年，Linda L. Hill 等多位學者針對加州大學聖塔芭拉拉分校（University of California at Santa Barbara）之亞歷山卓電子圖書館（Alexandria Digital Library）進行使用者評估。亞歷山卓電子圖書館是數位圖書館先導研究計畫的六個子計畫之一，主要目的在進行地理資訊數位化；而在本研究中，Hill等學者針對亞歷山卓電子圖書館Web版的雛形系統進行使用者評估，所採取的資料蒐集方法包括線上問卷、線上註冊（registrations）、民族誌研究（ethnographic studies）及焦點團體法。具體之結論包括下列幾項：

1. 大部份的早期使用者都是具備高學

歷、理解能力強，並且分布世界各地的網路使用者。此外，經由焦點團體法得知，地球科學家工作於高科技的環境，因此他們希望系統可以與其工作環境相結合，以方便他們使用資料；資訊專家會多花時間於學習系統的廣泛功能，因為他們可能較會處理檢索與查詢的工作項目，而非使用資料本身。教育學者需要將系統納入他們的課堂活動中，因此他們會著眼於系統支援團體活動、內容及功能性上。

2. 本研究所採用的使用者評估方法（線上問卷、線上註冊、民族誌研究及焦點團體法）雖然蒐集到許多寶貴

的資料，但仍有其限制，因此未來進行後續評估時，應將其評估方法進行修正。

- 3.大部份使用者對於系統介面都相當讚許，但仍有少數使用者對於系統介面感到困擾。例如使用back及stop鍵容易混淆、網頁上包含太多資訊、重新載入(reload)時間太長等。此外，使用者並建議將來輔助說明功能應再加強，並在系統中應描述系統之目的及使用本系統所需要的基本技能等。
- 4.系統應將擴增資料內容列為優先順序，並重新設計Web介面以提供這些資料及提醒使用者資料的侷限性。此外，在系統功能上，使用者需求檢索結果排序、連結之前的檢索、直接查看及瀏覽其結果、提供地理資訊系統等功能。(註59)

同年，Betsy Anagnostelis及Alison Cooke兩位學者針對Medline進行個案研究。他們比較了許多廠商所提供之Web版Medline服務，希冀發展出其評估標準，以提供圖書館或其他資訊專家購買不同廠商同樣Web版資料庫時之參考。這些廠商分別是COS、Healthgate、Healthworks、Helix、IGM、Medscape、OBGYN.net、PubMed及Ovid等。他們所發展的評估標準分為如下七大項：

- 1.一般資訊：包括權威性、資料庫所提供之服務及容易連線的程度(accessibility)等。
- 2.資料庫內容：包括資料庫包含範圍、

時效性。

- 3.檢索機制(retrieval mechanism)：包括一般檢索功能、提供索引典功能、提供自由語言檢索及檢索結果展示格式暨輸出控制等。
- 4.易於使用：包括從資料庫中是否容易檢索、展示及輸出資訊及是否容易航行等。
- 5.特殊功能：包括系統是否有提供特殊或獨一無二的功能等。
- 6.容易連線的程度：包括連線到這個資料庫是否有限制、是否有收費？若有收費，與其他廠商所提供的相同服務相比其收費如何等。
- 7.輔助說明及使用者協助功能：包括是否有簡介資料庫的服務、是否有線上輔助說明等。(註60)

此外，Canisius College館藏發展委員會發展了評估網站之查核表，其目的是要協助圖書館員、教師及學生評估網路上的資源。他們所發展出來的查核表共分為六個項目，分別為內容方面、權威性、整體架構及組織、檢索機制、容易連線的程度及與其他相同資源的比較等查核項目。(註61)

1998年，Ron Goral發展網站評估查核表。其查核表共分為如下八部分：

- 1.權威性：包括是否有提供網站負責人之資訊及其聯絡方式、是否有更新日期等。
- 2.網站贊助者：即網站贊助者之資訊。
- 3.內容：包括網站是否有陳述清楚的目的、是否有預設的使用族群、網頁中並沒有拼字或文法上的錯誤、

是否有文獻來源、是否為版權所有、從首頁所連結出去的是相關及有用的網站、內容是否正確及無偏見等。

4. 影像：包括影像是否與網頁有所關連、影像是否能夠傳輸快速、在一網頁中之影像數量是否合適、是否大部分的影像都為GIF的影像等。
5. 展示格式：包括網頁是否容易閱讀、網頁風格是否一致、在一網頁上是否有足夠的空白空間、字體大小和影像是否合適、網頁的傳輸時間是否合適等。
6. 多重網頁使用（multiple pages）：包括母頁（parent page）是否簡短以及包含其目的、是否有內部連結等。
7. 加強整體網頁：即是否有圖片、影音加入其中。
8. 技術上之考量：包括網頁上之連結是否能夠正常運作、在每頁最後之E-mail連結是否能夠正常運作、如網頁中有計時器，其計時器是否能夠正常運作、如網頁需要影音輔助，系統是否有提供觀賞影音之設備或軟體等。（註62）

1999年，Judy Druse提出評估網站的標準。他認為評估網站可由下面五方面進行評估：

1. 目的（purpose）：包含此網站是否有陳述清楚的目的、是否有預設的使用族群及此網站是否實質上有內容，而非僅是一大堆連結所組成的網站等。
2. 權威性（authority）：包括誰是網站的管理機構、它是否有提供聯絡方

式及網站上的所有資訊是否具有版權等。

3. 內容（content）：包括資訊的正確性、時效性及與其他資訊來源比較，是否存有偏見、資訊是一手資料還是二手資料、資訊是否已被好好研究及其可靠性如何、網站所包含的主題是否夠詳細、是否有提供超連結之敘述性說明、是否有圖形或影像以加強其內容及此網站上是否有提供廣告等。
4. 技術上的品質（technical quality）：包括網站是否便利使用者及是否很容易在此網站中航行、首頁是否簡明，且快速瀏覽、整個網站的呈現架構，是否能夠幫助使用者找到其所需要的資訊、網站是否為互動性、此網站是否夠穩定、是否提供搜尋引擎以搜尋網站的內容等。
5. 評論（review）：包括此網站是否有被專業機構評論或是此網站是否有被專門從事評估網站的商業機構評比過等。（註63）

在國內方面，翁慧娟針對國內五所大學圖書館之網站進行品質評估研究。首先，她蒐集相關文獻整理出網站品質評估準則，並針對其研究所選定的國內外大學圖書館網站進行內容分析，以了解目前大學圖書館網站建置現況及服務內容。接續針對台大、政大、中央、中原、銘傳五所大學圖書館館員與讀者進行網站品質評估準則及服務項目之意見調查。以疊慧法（Delphi study）彙整館員意見，問卷調查法了解

讀者意見，進一步比較兩者之差異。最後研擬大學圖書館網站品質評估準則，供以建置網站的圖書館或尚未建置者之參考。此大學圖書館網站評估準則共分為十二項：一、連結性，二、畫面設計，三、資源分類架構，四、內容，五、權威性，六、選擇性，七、新穎性八、多元性，九、管理資訊，十、服務溝通，

十一、溝通性，十二、使用者調查。（註64）

綜合上述學者專家們所提出之線上資訊檢索系統及搜尋引擎評估標準，研究者將之歸納整理成如下表三。其中較常被學者專家們提及的評估標準包括資訊內容、權威性、檢索功能、系統管理資訊及易於使用等。

表三 電子圖書館之評估標準

提出者 項目	Bishop	Mead & Gay	Anagnostelis & Cooke	Canisius College	Goral	Druse	翁慧娟
檢索功能	√	√	√	√			
檢索表現	√						
檢索結果展示		√	√		√		
輸出控制		√	√				
輔助說明			√				√
目的						√	
資訊內容	√		√	√	√	√	√
權威性			√	√	√	√	√
系統管理資訊			√	√		√	√
航行功能	√	√	√			√	
整體架體及組織				√		√	√
連結性					√	√	√
畫面設計					√		√
互動性					√	√	√
技術上的考量					√		
社會層次的影響	√						

陸、結論

系統評估研究雖然至今已有四十年時間，但針對電子圖書館進行評估也是近幾年的事，因此不論中外，其研究仍相當少，尤其國內的研究更是貧乏。由於目前對電子圖書館之評估研究尚屬起步階段，因此學者們對於建立電子圖書館評估之準則尚無共識。本文針對系統評估研究進行文獻探討，並針對電子圖書館評估研究提出下列之建議：

一、系統評估宜從使用者角度進行評估

系統評估的最大目的是希望能夠改進系統、提昇資訊服務品質，以供使用者做最大的利用，因此使用者在系統評估過程中佔了相當重要的地位。目前中外多項電子圖書館的評估研究中越來越強調使用者角色，因此紛紛從使用者角度進行評估，畢竟使用者是影響一個系統成敗的重要因素。有鑑於此，今後之系統評估研究宜從使用者角度進行評估。

二、系統評估應考慮系統週邊人員之觀點

雖然使用者在系統評估中之角色越來越受重視，但系統評估不應僅包含使用者之意見，也應該包括系統週邊人員之意見。Julian Kilker及Geri Gay兩位學者針對MOA (The Making of American) 電子圖書館之雛形系統進

行評估時，即從三群不同使用者族群、發展者、圖書館員及贊助者之觀點進行評估。(註65) Moen等四位學者針對政府資訊定位服務系統進行評估時，亦從使用者、機構本身及整個政府機關等觀點進行評估。(註66) 由上所述，目前學者們從事電子圖書館之評估研究，已逐漸將使用者觀點此範疇慢慢擴展至包含系統週邊人員之觀點，而如此所做出之系統評估才是較完整之評估研究。

三、系統評估應廣及各層次之評估

Tefko Saracevic在述及系統評估研究時，將系統評估依目標層次劃分為工程層次 (engineering level) 的系統評估、輸入層次 (input level) 的系統評估、處理層次 (processing level) 的系統評估、輸出層次 (output level) 的系統評估、使用與使用者層次 (use and user level) 的系統評估及社會層次 (social level) 的系統評估等六個系統評估種類 (註67)，其中層次一至三趨近於系統中心；層次四至六則趨近於使用者中心。系統評估除了層次一至三的技術面思考外，更應致力於真正與人類密切相關的社會層次評估，因為這是最有價值且最複雜的系統評估層次。(註68) Borgman等多位學者針對電子圖書館進行評估研究時，即特別著重於社會層次之評估。(註69) 有鑑於此，未來的系統評估研究有必要針對電子圖書館如何影響人類的工作、休閒、社會及文化等方面進行研究。雖

然此方面的評估相當困難，但卻是系統評估研究真正落實於人類生活的表現。

四、系統評估研究應採用量化及質化兼具之方法

以往從事系統評估研究不外採取兩種評估方式，即量化研究及質化研究。自從使用者角色在系統評估過程中被確立後，學者們逐漸將量化之研究方向轉向質化之研究方向。計量之研究諸如問卷法、查詢過程記錄分析法等，而計質之研究諸如深度訪談法、焦點團體法及關鍵事件回述法等。以上兩種研究導向皆有其優缺點，但重要的是如何將兩方面資料進行結合，而得到一個較完整的評估結果，進而改善系統。Moen等四位學者針對政府資訊定位服務系統進行評估時，有採量化之調查法、內容分析法、查詢過

程記錄分析法，亦有採質化之參觀網站之經驗、焦點團體法、使用者填答評估及政策檢視等。（註70）

五、從事電子圖書館之使用者研究

從1989年American Memory從事使用者研究起，一直至1995年數位圖書館先導計畫所從事的一連串電子圖書館使用者研究為止，國外在電子圖書館使用者研究方面已累積相當多的經驗，而國內則甚少針對此方面進行研究。使用者研究的效益可用於介面設計、活動設計、資源的組織與選擇及文獻記錄上（註71），因此了解電子圖書館之使用者也就深具重要性，畢竟系統設計的重點是要系統適應個人，而非讓人們學習如何去使用系統。由於電子圖書館的使用者研究比一般資訊檢索系統使用者研究更難進行，因此將來有必要再針對此方面繼續努力。

註釋

- 註 1 : Gerard Salton, "The State of Retrieval System Evaluation," Information Processing & Management 28:4 (1992): 441.
- 註 2 : F. W. Lancaster, "MEDLARS: Report on the Evaluation of Its Operating Efficiency," American Documentation 20:2 (Apr. 1969): 119-142.
- 註 3 : B. C. Vickery, Techniques of Information Retrieval (London: Butterworth, 1970), 213.
- 註 4 : Allen Kent, Information Analysis and Retrieval (N. Y.: Becker and Hayes, 1970), 314.
- 註 5 : Eva L. Kiewitt, Evaluating Information Retrieval Systems: The PROBE Program (Westport, Conn.: Greenwood Press, 1979), 125-132.
- 註 6 : Carol Hansen Fenichel, "Online Searching: Measures that Discriminate among Users with Different Types of Experiences," JASIS 32:1 (Jan. 1981): 23-32.
- 註 7 : Geraldene Walker, "The Search Performance of End-Users," In Proceedings of the 9th National Online Meeting (Medford, NJ.: Learned Information, 1988): 403-410.
- 註 8 : D. Alasdair Kemp, Computer-based Knowledge Retrieval (London: Aslib, 1988), 212-215.
- 註 9 : William Hersh and Jeffery Pen-tecost, "A Task-Oriented Approach to Information Retrieval Evaluation," JASIS 47:1 (Jan. 1996): 50-56.
- 註 10 : 蔡明月, 線上資訊檢索—理論與應用 (台北市: 學生, 民國80年), 頁 295-305。
- 註 11 : 周曉雯, 「線上檢索結果之評估」, 書府 12期 (民國80年6月), 頁117-120。
- 註 12 : a. 黃慕萱, 資訊檢索 (台北市: 學生, 民國85年), 頁280-287。
b. 黃慕萱, 資訊檢索中「相關」概念之研究 (台北市: 學生, 民國85年), 頁107-116。
- 註 13 : 卜小蝶, 圖書資訊檢索技術 (台北市: 文華, 民國85年), 頁172-181。
- 註 14 : Martin P. Courtois, William M. Baer, and Marcella Stark, "Cools Tools for Searching the Web: A Performance Evaluation," Online 19:6 (Nov./Dec. 1995): 15-32.
- 註 15 : Heting Chu and Marilyn Rosenhal, "Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology," In Proceedings of the 59th Annual Meeting of the ASIS (Medford, N. J.: Information Today, 1996): 127-135.

- 註 16 : Wei Ding and Gary Marchionini, "A Comparative Study of Web Search Service Performance," In Proceedings of the 59th Annual Meeting of the ASIS (Medford, N. J.: Information Today, 1996) : 136-142.
- 註 17 : Xiaoying Dong and Louise T. Su, "Search Engines on the World Wide Web and Information Retrieval from the Internet: A Review and Evaluation," Online & CD-ROM Review 21:2 (Apr. 1997) : 67-81.
- 註 18 : 陳秀美, 「WWW檢索工具的選擇與比較」, 資訊傳播與圖書館學 2卷4期 (民國85年9月), 頁41-63。
- 註 19 : Ray R. Larson and Vicki Graham, "Monitoring and Evaluating MELVYL," Information Technology and Libraries 2:1 (March 1983) : 93-104.
- 註 20 : Ben Shneiderman, Dorothy Brethauer, Catherine Plaisant, and Richard Potter, "Evaluating Three Museum Installations of a Hypertext System," JASIS 40:3 (May 1989) : 172-182.
- 註 21 : Micheline Hancock-Beaulieu, Stephen Robertson, and Colin Neilson, "Evaluation of Online Catalogues: Eliciting Information from the User," Information Processing & Management 27:5 (1991) : 523-532.
- 註 22 : Egbert De Smet, "Evaluation of a Computerised Community Information System through Transaction Analysis and User Survey," Libri 45:1 (1995) : 36-44.
- 註 23 : Micheline Beaulieu and Victoria Mellor, "The Micro Gallery: an Evaluation of the Hypertext System in The National Gallery, London," The New Review of Hypermedia and Multimedia 1 (1995) : 233-260.
- 註 24 : Michael Dolenko, "Making Userbased Assessment Work: Evaluating Statistics Canada's Web Site," In Proceedings of the 19th National Online Meeting (Medford, NJ.: Learned Information, 1998) : 105-113.
- 註 25 : a. William E. Moen, Charles R. McClure, June Koelker, and Erin, "Assessing the Government Information Locator Service (GLIS): A Multi-Method Approach for Evaluating Networked Services," In Proceedings of the 61th Annual Meeting of the ASIS (Medford, N. J.: Information Today, 1998): 70.
b. William E. Moen, and Charles R. McClure, "An Evaluation of

- the Federal Implementation of the Government Information Locator Service (GLIS): Final Report," <<http://www.unt.edu/slis/research/gilseval>>
- 註 26 : 王喜沙, 「線上公用目錄評估及研究方法之探討」, 政大圖資通訊 5 期 (民國 82 年 5 月), 頁 32-46。
- 註 27 : 邱韻鈴, 國立清華大學線上公用目錄使用調查 (台北市: 漢美, 民國 83 年), 頁 92-95。
- 註 28 : 林麗娟, 「以使用者為本之資訊系統評鑑」, 圖書與資訊學刊 15 期 (民國 84 年 11 月), 頁 25-29。
- 註 29 : 此表係綜合王喜沙的意見及個人閱讀文獻所得。
- 註 30 : a. David B. McCarn and Craig M. Lewis, "A Mathematical Model of Retrieval System Performance," JASIS 41:7 (Oct. 1990): 495.
b. Masse Bloomfield, "Evaluating of Indexing 4. : A Review of the Cranfield Experiments," Special Libraries 62:1 (Jan. 1971): 24.
- 註 31 : a. Cyril Cleverdon, "The Cranfield Tests on Index Language Devices," Aslib Proceedings 19:6 (1967): 173.
b. 同註 12b, 頁 118。
- 註 32 : Karen Sparck Jones, Information Retrieval Experiment (London: Butterworth, 1981), 276.
- 註 33 : 同註 30a, 頁 173-192。
- 註 34 : a. 同註 2。
b. 同註 3。
- 註 35 : Gerard Salton, "A Comparison Between Manual and Automatic Indexing Methods," American Documentation 20:1 (Jan. 1969): 64.
- 註 36 : 同註 3, 頁 225。
- 註 37 : a. 同註 1, 頁 446。
b. 同註 12b, 頁 125-126。
- 註 38 : 同註 35, 頁 62。
- 註 39 : 同註 1。
- 註 40 : 同註 32, 頁 320-321。
- 註 41 : 同註 1。
- 註 42 : 同註 35, 頁 70。
- 註 43 : 同註 1。
- 註 44 : David C. Blair and M. E. Maron, "An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System," Communications of the ACM 28:3 (March 1985): 289-299.
- 註 45 : Louise T. Su, "An Investigation to Find Appropriate Measures for Evaluating Interactive Information Retrieval," In Proceedings of the 52th Annual Meeting of the ASIS (Medford, N. J.: Information Today, 1989): 13-23.
-----, "An Investigation to Find Appropriate Measures for Evaluating Interactive Informa-

- tion Retrieval" (Ph. D. diss., University of Rutgers, 1991)
- 註46：Alan F. Smeaton, and Donna Harman, "The TREC Experiments and Their Impact on Europe," Journal of Information Science 23:2 (1997): 169-174.
- 註47：Stephen P. Harter, and Carol A. Hert, "Evaluation of Information Retrieval Systems: Approaches, Issues, and Methods," In ARIST (Medford, N. Y.: Information Today, 1997): 24-27.
- 註48：江玉婷、陳光華，「TREC現況及其對資訊檢索研究之影響」（尚未發表）
- 註49：a.同上註。
b. Ellen M. Voorhees, and Donna Harman, "Overview of the Sixth Text Retrieval Conference (TREC 6)," <<http://trec.nist.org/pubs/trec6/papers/overview.ps>>
- 註50：同上註。
- 註51：同註48。
- 註52：陳雪華，「史料數位化與Metadata」，在台灣古文書數位化研討會暨成果發表會（台北市：國立台灣大學，民國87年6月25日）
- 註53：Monica Landoni, Nadia Catenazzi, and Forbes Gibb, "Hyper-Books and Visual-Books in an Electronic Library," The Electronic Library 11:3 (June 1993): 177.
- 註54：Library of Congress, "Final Report of the American Memory User Evaluation, 1991-1993," <<http://lcweb2.loc.gov/ammem/usereval.html>>
- 註55：相關的研究諸如：
a. Nancy A. Van House, "User Needs Assessment and Evaluation for the UC Berkeley Electronic Environmental Library Project: a Preliminary Report," <<http://www.csd1.tamu.edu/DL95/papers/vanhouse/vanhouse.html>>
b. Linda L. Hill et al., "User Evaluation: Summary of the Methodologies and Results for the Alexandria Digital Library, University of California at Santa Barbara," In Proceedings of the 60th Annual Meeting of the ASIS (Medford, N. J.: Information Today, 1997): 225-243.
- 註56：同註54。
- 註57：Ann Peterson Bishop, "Working Towards an Understanding of Digital Library Use: A Report on the User Research Efforts on the NSF / ARPA / NASA DLI Projects," <<http://www.dlib.org/dlib/october95/10bishop.html>>
- 註58：June P. Mead, and Geri Gay. "Concept Mapping: An Innovative Approach to Digital Library Design and Evaluation," <<http://>

- edfu.lis.uiuc.edu/allerton/95/s2/mead/mead.html>
- 註 59 : 同註 55b。
- 註 60 : Betsy Anagnostelis, and Alison Cooke, "Evaluation Criteria for Different Versions of the Same Databases a Comparison of Med-line Services Available Via the World Wide Web," In Proceedings of the 21st International Online-Information Meeting (Oxford: Learned Information, 1997) : 165-179.
- 註 61 : Canisius College Library. Collection Development Committee. "Checklist for Evaluating Web Sites," <<http://www.canisius.edu/canhp/canlib/webcrit.htm>>
- 註 62 : Ron Goral, "Web Site Evaluation Checklist," <<http://www.madison.k12.wi.us/webpub/evaluate.htm>>
- 註 63 : Judy Druse, "Web Site Evaluation Criteria," <<http://www.washburn.edu/mabee/reviews.html>>
- 註 64 : 翁慧娟, 「我國大學圖書館網站品質評估之研究」(碩士論文, 國立台灣大學圖書資訊學研究所, 民國 88 年)。
- 註 65 : Julian Kilker, and Geri Gay, "The Social Construction of a Digital Library: A Case Study Examining Implications for Evaluation," Information Technology and Libraries 17:2 (June 1998) : 60-70.
- 註 66 : 同註 25a。
- 註 67 : Tefko Saracevic, "Evaluation of Evaluation in Information Retrieval," SIGIR Forum 1995, 140.
- 註 68 : 同上註, 頁 146。
- 註 69 : Christine L. Borgman et al. "Social Aspects of Digital Libraries," <http://www-lis.gslis.ucla.edu/DL/UCLA_DL_Report.html>
- 註 70 : 同註 2, 頁 67。
- 註 71 : 林珊如, 「數位圖書館使用者研究」在 建構使用者導向的數位化圖書館 (台北市: 國科會數位博物館推廣教育計畫小組, 民國 87 年 10 月 8 日~9 日)。