

資訊檢索全文資料庫績效評估工作量模式之 建立與應用—以傳播學數位圖書館為例

譚家蘭、林佳慧

政治大學資訊管理學系

摘要

以資訊檢索為基礎的網路搜尋技術，已逐漸成為 Internet 上相當受到重視的技術之一。雖然全文資訊檢索的技術已存在多時，然而電腦產業一直缺乏依廣泛標準的績效評估，來評估全文資訊檢索系統的產出及價格/效能比，網際網路年代對全文資訊檢索軟體的大量需求，極需要一個標準、一致的方法，來比較各系統間效能的差異。本研究即針對傳播中文全文資訊檢索資料庫系統一個傳播數位圖書館的雛形為探討的主題，試圖以資料庫績效評估的理論及方法，來建構其工作量模式。本研究針對目前績效評估的瓶頸，即是無法正確、合理、公允表達不同使用者應用系統的真實績效表現，提出新的研究模式以解決問題根源在領域特定和領域專屬，無法代表使用者工作量的現在評估方法。透過應用 Benchmark 方法論和 Information Retrieval 方法論為理論基礎，設計以使用者需求為導向的開放型工作量模式，根據不同使用者需求產生不同工作量，執行不同測試能夠產生比較正確、合理、公允表達績效結果。本研究模式分為資料模式、交易模式和控制模式，用以標準承載資訊檢索資料庫一般化工作量，提供管理者和系統者有效率和效果評估工具。

關鍵字：全文資訊檢索、資訊檢索、自然語文資訊檢索、文件資料庫、數位圖書館、工作量模式、資料庫績效評估

感謝：本研究論文係由政治大學鄭瑞城教授主持之國科會專題研究計畫 NSC87-89-2412-H-004-021 所支持。感謝：資管學報評審委員寶貴而深入之評審意見。

A Full-Text Information Retrieval Database Workload Model for Benchmark —Communication Digital Library Case

Jia-Lang Seng, Chia-Hui Lin

Department of Management Information Systems, National Chengchi University

ABSTRACT

Information retrieval represents the vital technology in the Internet era. With rapidly growing Web sites and Internet servers, it is important to understand the performance of information retrieval and information search. However, the traditional database benchmarks lack the ability to express the unstructured data model and path expressions, not to mention document storage. In this research, we adopt the principles of benchmark methodology and information retrieval methodology to develop a requirements-driven full text information retrieval database workload model for benchmark. This development incorporates non-structured and path expressional design and domain-independent characteristics. It provides an alternative to solve the ir-representativeness and misleading performance issues. The new workload model is composed of data model, transaction model, and control model to offer reliable, accurate, real, and scalable.

Key words: Full-Text Information Retrieval, Information Retrieval, Free-Text Information Retrieval, Document Database, Digital Library, Workload Model, Database Benchmark

壹、緒論

一、研究背景與動機

(一)前言

近數百年來，人類知識的保存、傳承與相互間的溝通，主要透過印刷式的紙本資料，然而印刷科技在擅揚多時之後，已有日漸被電子數位式資料取代或作為補充的趨勢，不同形式的新媒體以超強的功能，優越的姿態陸續問世。如 Internet 網際網路和 www 全球資訊網 (World wide web, Web, WWW) 上多元的搜尋系統，是目前網路上最常被利用的服務之一，由於網路上的資源包羅萬象、種類繁多、又隨時變動，若不能好好利用這類系統，尋找資源將有如大海撈針般的困難。在 Internet 的時代，如何正確且有效率地利用網路資訊檢索技術，查詢到所需的資源，也成為人人必須學習的知識。

(二) Internet 資源搜尋技術的發展

資訊檢索 (Information Retrieval) 一直是資訊學界的重要研究領域，但長久以來並未能在其它的學門中發揮其重要性，然而藉由 Internet 的普及與網路資源的激增，以資訊檢索為基礎的網路搜尋技術，已逐漸成為 Internet 上相當受到重視的技術之一，正如 Datapro 的調查顯示，資源搜尋系統已成為少數熱門新興發展軟體之一 (Robert, 1997)。

Internet 資源搜尋技術發展至此看似完善，但隨著資料型式更多元、資料種類不再侷限於文字檔案，如一些聲音、圖片、影像等多媒體資料的急速成長。此外也有很多使用者開始覺得 Gopher 田鼠搜尋器層級式的瀏覽方式不夠親切自然，於是取而代之的是以超文件型式、瀏覽多媒體資訊的 WWW 系統。然而隨著 WWW 網站的增加，找尋適合的 WWW 網站也變得很困難，於是有人將這些網站分門別類建立起主題目錄 (Subject Directory)，方便使用者以主題類別瀏覽網站資源，不過由於資源的成長十分的驚人，這些主題目錄無法有效地應付分門別類的查詢需求，於是所謂的 Web 搜尋引擎 (Search Engine) 也就應運而生。這些系統主要提供 Web Page 的全文查詢，雖然以資訊檢索理論觀點來看，這些系統主要提供的檢索功能還相當簡單，但由於收錄資料、查詢速度、使用者層面等，都遠比以往的資訊檢索系統表現的更為突出，而深受使用者的重視。如 Alta Vista 宣稱收集了六十萬個 Web Server 站、五千多萬篇 Web Pages、及總計超過一萬四千多個網路新聞群上的三百多萬篇文件，而且每隔幾個月就成長數倍，其使用率更高達每天一千兩百多萬人次；又如 InfoSeek 每天也有七百多萬次的查詢筆數 (卜小蝶, 民 87 年)。

雖然這些搜尋引擎可以讓使用者快速地查詢到網路資源，但使用者普遍認為查詢出來的資料品質不佳，很多改進構想因此產生，諸如美國麻州大學的 Croft 教授就提出所謂的十大當其為重要的檢索技術方向，除了加強傳統索引搜尋技術 (Indexing and Searching)、分散式檢索技術 (Distributed Information Retrieval)，也包括新興技術如：強調從網路中過濾不具備檢索價值的資料，並萃取重要資訊的資訊過濾技術 (Information Filtering)，針對各種聲音、影像資料所發展的多媒體資訊檢索技術 (Multimedia Information Retrieval)，都成為研究網路資源檢索專家們所亟思解決的問題。另外目前資源搜尋技術最渴望達到個人化資訊服

務的目標，於是開發出所謂的智慧型資訊檢索代理人(Intelligent Information Agent)，如此每個人都能根據自己的需求，命令系統收集整理資訊，而不再受限於現行系統複雜的指令，以及瀏覽一些不相關的資訊(卜小蝶，民 87 年)。

(三)全文資訊檢索系統績效評估

對大部份的文字記錄歷史來說，文件資料主要是以拷貝方式儲存，而相關的文件檢索過程(可能包含跨目錄的檢索)則是以人工方式完成。1960 年代匹茲堡大學發展法律文獻的電腦輔助系統，來完成上述的檢索工作，同時間內以電腦輔助為主的文件檢索系統開始朝氣蓬勃，例如紐約時代的 InfoBand 系統，這類系統最大的特色是以布林運算和近似比對查詢為主，使用此類的資訊技術其所謂的全文資訊檢索(Full-Text Information Retrieval: FTIR)：即根據文章內容從資料庫找出文件，而不是使用之前定義好的關鍵字或主題目錄來進行檢索(Harrington, 1984-1985)。

FTIR 最大的優點就在於文件可依據其內容來進行存取，然而為了提供此種能力，常需要花費大量電腦資源。隨著文件資料庫的成長，電腦的運算能力也以直線方向增加，因此當資料庫以 N 倍成長，電腦運算能力也須相對應的增加，以保持相同的回應時間。目前 FTIR 系統應用 Surrogate 代理方式檔案結構技術來改善回應的時間，Surrogate 技術是在文件資料庫的每個符號處(Token)(如文字、數字、時間、日期等)，提供一個進入點(Entry)，因此造成 Surrogate 檔案、搜尋處理時間與資料庫的大小是以等比例的方式增加，即當搜尋數目與資料庫是以 N 成長，則搜尋所需的運算能力要以 N^2 速度成長(Faloutsos, 1985)。

(四)目前研究之困境

FTIR 資料庫因此面臨績效和資料結構系統化的取捨問題。FTIR 績效來自於快速搜尋到全文檢索的結果，進一步要求全文檢索的結果的準確度、容錯性、模糊比對性。FTIR 資料結構來自於文件資料庫在運算邏輯上的假設，純讀取和搜尋與全面讀寫和搜尋的資料結構是不相同的。斷字和斷句的資料結構技術，以及依照搜尋（也就是依照全文檢索）的形式和路徑來設計資料庫和開發演算法，都和傳統關聯式資料庫不同。FTIR 資料庫因此亟需一套系統化和需求導向的績效評估方法，作為評估和檢視 FTIR 資料結構和查詢演算法是否符合績效的要求？是否 FTIR 資料庫存在績效瓶頸？如何加以偵測？有沒有什麼方法可以改善？這是本研究企圖提供一項思考的方向和解決的模式。

FTIR 資料庫績效評估方法目前在學術界分為資料檢索領域和資料庫領域，資料檢索領域限於傳統對 Benchmark 的認知不同，將績效評估縮小為測試回應時間和測試查詢結果。資料庫領域由於不專長於全文檢索資料結構和演算法，企圖套用傳統關聯資料庫和物件資料庫績效評估方法，以表格和物件來設計測試，未能顧及 FTIR 資料庫同時包含平結構和無結構化的資料結構和演算法。在實務產業界，FTIR 資料庫績效評估方法採用快速、實用、方便的網頁流量和動態超連結的執行率為主要績效指標，對於背後 FTIR 資料庫引擎的本質和特色，並不考慮。對於是否要設計相關資料結構和演算法，就放在較後的順序中，更遑論嚴謹討論中文資訊檢索資料庫績效評估方法。

因此，本論文企圖提出一個開放型(Open)和以需求為導向(Requirements-Driven)的一般化(Generalized)資訊檢索資料庫績效評估方法，希望在研究模式上能突破目前在

學界和業界所不願和不能的評估瓶頸。這些包括非結構化資料結構的模式化、資訊檢索運算工作量的模式化、全文資料庫產生的模式化。因此，本研究模式使用需求分析為一般化的標準載具 (Carrier)，將開放型的資料和搜尋型態 (Pattern) 由使用者載入工作量需求，經由系統和機率分布產出不同工作量，以符合不同使用者使用和不同準確評估的要求。藉此，提供領域獨立性 (Domain Independence) 和領域代表性 (Domain Representativeness)。不過，由於研究本質龐大複雜，本論文僅限於應用領域，僅能以傳播學的數位資料庫圖書館為例，實作部分亦以傳播學全文檢索資料庫為主。研究模式本身是以開放式和一般化為各構成元素和各子模式的理論基礎和開發理念。理論架構根據 (TPC 2001) (Gray 1993) (Turbyfill 1993) (Dewitt 1983) 開發資料結構模式、資料運算和應用模式、和測試控制模式。

二、研究問題陳述

目前一般的資料庫績效評估其測試的工作量模式通常是使用合成工作量 (Synthetic Workload) 作為測試的依據，並且也都依據特定的應用領域來建構某些典型的應用。例如一系列的 TPC (Transaction Processing Performance Council) 績效評估是根據企業的 OLTP (On line Transaction Processing) 及 DSS (Decision Support System) 系統而制定；Bucky 績效評估方法針對學校的運作狀況來建構其資料庫綱目；而 001 和 007 是針對工程上的應用，這些由學術界和資料庫廠商所自行針對特定領域所訂定的測試工作量，在許多的研究文獻中是屢見不鮮的 (Gray, 1993) (Carey. et al., 1997) (Cattell, 1992) (Carey. et al., 1993)。另外由於近年來電腦科技及網路技術快速發展，使得資料庫的應用環境與領域比較過去而言更加複雜與多元，因此績效評估規格為了能夠更具全面與深度地測試資料庫管理系統的各項功能，所訂定的工作量模式之資料庫綱目與測試運算要能更符合特定領域的需求規格。

本研究將探討下列的問題：

1. 前人在資料庫績效評估的研究，如何應用到網際網路的中文自由語文資訊檢索資料庫系統上。
2. 應用資料庫特性建構一般性的傳播數位圖書館資料庫的工作量模式。
3. 如何實作評估傳播中文自由語文資訊檢索資料庫績效。

貳、文獻探討

一、資訊檢索系統

(一) 資訊檢索技術簡介

資訊檢索的目的在利用電腦的儲存和計算能力，來幫助使用者由大量缺乏結構化資料中，快速地擷取到所需的資訊。利用電腦科技協助分類管理、儲存、搜尋、過濾大量的電子資訊，就是所稱的資訊檢索技術 (Information Retrieval: IR)。在這裡所要檢索的資訊包括各種形式的資料，如表格、文件、圖片等，這些不同形式的資料在資訊內涵上有結構化 (Structural) 與非結構化 (Non-Structural) 之區分。

1. 傳統資訊檢索技術

傳統上資訊檢索技術研究包括關鍵字自動萃取(Keyword Extraction)、全文資訊檢索(Full-Text Information Searching)、文件自動分類(Document Classification)、文件自動摘要(Document Abstraction)、關鍵字索引(Keyword Indexing)、超文本技術(Hypertext)等(卜小蝶, 民 85 年)。

2. 資訊檢索的區分

文件資料庫系統依檢索方式的不同可再區分為指標式檢索 (Indicative-text Retrieval) 全文資訊檢索 (Full-text Information Retrieval)。指標性檢索只能讓使用者針對某些特定關鍵語加以檢索；而全文資訊檢索允許檢索文件任一文句、任一字串。顧名思義，既然稱"全文"，即是將使用者所下的查詢直接和文件內容比對。

(二) 全文資訊檢索技術

全文資訊檢索技術的需求起源來自於關鍵詞索引技術的限制，早期資訊檢索多半仰賴關鍵詞索引，如查詢新聞事件資料庫時，只能利用系統預先提供的關鍵詞為檢索點，使用者在檢索時，必須先知道哪些檢索詞屬於控制詞彙，其檢索彈性因而大為受限。而全文資訊檢索就是為了提高資訊檢索便利而產生出來的技術。

1. 全文資料庫

討論全文資訊檢索技術必須先從瞭解全文資料庫開始，而在瞭解全文資料庫之前必須先對文件加以定義。

(1) 文件(Document)

根據國際標準組織(International Standard Organization)中第 8879 號標準對文件的定義：「所謂文件或全文是由自然語言或人工語言中的符號、片語、句子，以及圖形、公式和表格等所組成之二度空間形式，藉以表達人類所能理解所有資訊」。所以書名、摘要、法規、論文、報告等皆可稱為文件。

(2) 全文資料庫(Full-Text Database)

所謂全文資料庫(Full-Text Database)是指將文獻的全部原文以忠於原來形式的電子方式，逐字存入電腦之資料庫中，並建立各種檢索或查詢之方法，以便使用者能透過電腦網路或相關的電子通訊設備，做即時線上之查閱資料或進一步做統計、分析、整理等應用之工作的系統(徐惠文, 民 77 年)。

2. 全文資料庫與傳統資料庫

傳統的資料庫多是針對數值性、可格式化的資料加以設計，特別是關連式資料庫(RDBMS, Relational Database Management System)，其所處理的對象是具有格式化性質的資料(Formatted Data)，而其目的便是定義關連式的記錄(Record)，並記載其屬性(Attribute)以便檢索；而全文資料庫所處理的對象是屬於全文性的資料，較缺乏明顯和具體的結構屬性，且檢索的內容亦無法事先確定，這是因為其所面對的檢索需求可能包括文件中的任一字串，或是文件中所隱含的任一概念。

3. 全文資訊檢索技術之發展

早期的全文資料庫經常藉由已有的資料庫，配以全文檔案所組成，他們把全文中與檢索相關的屬性(Attribute)資料或關鍵字(Keywords)利用人工或電腦程式整理出來，並利用資料庫管理系統建立為欄位化的資料庫，在此資料庫中以指標(Pointers)指向相關原文的檔案中的地址，故其設計主要是以書目資料庫的模式為基礎發展而來，因此全文資料庫的檢索

方式和傳統的資料庫一樣，只是多提供了參閱原文的機會而已。由於從原文中摘出檢索所需的屬性或關鍵字，存在許多與語文相關的技術上的困難，例如：同義字、同形異義字、相關字等，因此必須由各種專家建立索引典等權威工具對詞彙加以控制，以做為處理資料及檢索資料的依據。

4. 全文資訊檢索特性

全文資訊檢索系統通常具有下列特點(蔡明月, 民 80 年)：

- 可任意選擇某特定詞句查詢所需文獻的全部內容。
- 可利用作者、主題等不同需求條件，檢索相關文獻的部份內容。
- 無須預定設定關鍵語，即可做單字或片語的檢索。
- 使用者可依不同需求，以各種指令建立查詢程序。
- 操作過程採畫面指引方式，使使用者容易學習。
- 共通的檢索技術有：布林邏輯運算 AND、OR、NOT。
- 依檢索用語出現次數的多少判斷其與主題相關性。
- 為了使全文資料庫檢索更加便利，全文資料庫與索引資料庫的結合乃一必然的走向。

5. 全文資訊檢索技術

目前已發表的全文資訊檢索技術，最常用的分別為全文掃描檢索法(Full Text Scanning)、轉置法(Inversion)、叢集法(Clustering)和特徵檔(Signature File)等四類，分別說明如下(Faloutsos, 1985)：

(1) 全文掃描法(Full-Text Scanning)

此搜尋方式是逐一的比對(Compare)文章與搜尋字串(Search String)中的每一個字。而每一次的比對都從文章的起始位置開始，如果兩者的比對結果不一致(Mismatch)，則右移文章中一個字的位置且繼續比對，直到比對的條件完全符合或是到達文章的終了為止。此種搜尋方式是簡單、容易實作、不需要額外的空間(Space Overhead)，但其最大的缺點是搜尋的速度太慢，平均需要花相當多的時間才能找到所需的資料。

(2) 轉置法(Inversion)

此搜尋方式的主要精神是在於對關鍵字作索引(Keyword Indexing)，在搜尋資料之前，首先需對文章中的關鍵字作索引，並將索引的結果儲存在字元轉置檔(Word Inversion File)，而在查詢時，只需要去比對字元轉置檔中的資料即可，此種搜尋方式的優點在於其檢索速度快，但缺點是需要很大的額外空間。

(3) 叢集(Clustering)

此搜尋的主要概念是將相同性質的文章歸類在一起形成一個叢集，因為在叢集法中提出一個前提假設(Clustering Hypothesis)，其認為相同類型的文章常伴隨著同一類型的查詢條件，因此若是將這些相似的文章整合在一起必能加快搜尋的速度，但這個方法與轉置法一樣，有著不易新增及更新資料的缺點。

(4) 特徵檔(Signature file)

特徵法的主要概念在於將文章中的字透過某一種轉換方式，例如雜湊函數(Hashing Function)，轉換成位元字串(Bit String)的表示，並且透過適當的資料結構整合這些轉換後的位元字串。這個搜尋方式的優點在於其有很快的搜尋速度，且資料容易加入、更新、適於動態的資料環境中。但是這個搜尋方法最大的缺點在於會有錯誤查詢(False Drop)的產生，而所謂的錯誤查詢是指在實際資料庫中並無符合搜尋條件的文章，但是在特徵檔的比對中

確有符合搜尋條件的結果發生稱之。

6. 數位圖書館的趨勢

自從 1993 年 World Wide Web 的興起，Internet 上的文件數量一直是以幾何級數在成長。現在大部份的學生、專業人士、學者在尋找資料或參考文件時，最常用也最方便的方式就是上網找資料，網際網路儼然如一虛擬的數位圖書館。數位圖書館(Digital Library)泛指各式各樣的數位文件(Digital Document)之集合，它是數位的資訊庫、文件庫、同時也是知識庫；數位文件泛指各式各樣數位化的文件，不論是文字或圖片、單一媒體或多媒體、簡單文件或複合文件(張品臻，民 87 年)。

從傳統的圖書館自動化系統衍生到數位圖書館系統，最基本的差別當然就是數位化文件與非數位化文件。但這數位化後的環境卻完全的改觀，文件本體從紙張為主的變成多媒體複合數位文件；文件展示閱讀媒介變成是以電腦終端機為主；文件儲存空間改為數位儲存媒體；文件儲存擷取方式改為數位儲存媒體上的存取方式；文件查詢檢索除了對編目索引之外，更可對文件的本文內容做查詢檢索，其它演變功能像智慧型代理人(Intelligent Agent)及其他的服務更是不勝枚舉。

數位圖書館系統的功能架構大致上可分成幾個層次(張品臻，民 87 年)：

- (1)網路功能(Network)：提供數位圖書取用之通道，此功能大量運用網際網路的特性與結構。
- (2)儲存功能(Storage)：提供數位圖書及相關資料的儲存功能。數位文件通常為多媒體複合文件，而此儲存功能即是提供多媒體複合文件和相關資料之儲存功能的文件庫。
- (3)存取功能(Access)：基於多媒體、複合文件及相關資料的存取方式迥異，文件存取功能大致上可分為格式化資料存取、多媒體資料存取及複合文件的存取功能。
- (4)展現功能(Presentation)：此功能是将數位文件展現給使用者。

未來的資訊資源將會集中在網際網路上，網際網路也將取代現有的傳播方式成為主要的資訊傳遞、檢索與擷取的媒介。許多的網站本身功能可看成數位圖書實儲存所在地，譬如網路上的搜尋引擎 Yahoo、Infoseek、AOL 等網站，對整個網際網路的網站提供了目錄、索引、搜尋等服務。因此我們可以說網際網路可視為一功能具足的虛擬數位圖書館。本論文所研究的對象傳播學中文自由語文資訊檢索資料庫系統其提供檢索查詢的機制和文件存取、展現的基本功能，就概念上來說，已具備一小型數位圖書館的規模，但要邁入成為一全方位的數位圖書館，仍需不斷地擴充、發展。而期望藉由本研究所整理、定義出的中文全文資訊檢索資料庫系統的評估準則，能繼續成為未來傳播中文自由語文資訊檢索資料庫系統擴充、發展後績效評估的基石和依據。

二、資料庫績效評估

(一)績效評估的定義

績效評估是一套評估效能的方法或技術。通常績效評估定義一組規則用來比較相同系統或不同系統下的效能，因此我們可定義績效評估是：一組標準的指令用來比較不同系統的相對績效(Gray, 1993)。

績效評估提供了系統建置及使用的參考標準，其主要目的是提供不同系統間績效評估的標準，或是廠商用來彰顯其系統的特定，亦或是用以設計某個可預測系統績效表現的模式。

(二)績效評估的特色

由於沒有一個簡單的指標能夠衡量電腦系統在所有的應用程式上的績效。系統的表現

可從一個應用領域至另一個，而每個系統是用來設計解決一些特定的(Domain Specific)問題領域(Problem Domain)，因此極有可能無法解決其他的任務。領域專屬的標竿(Domain-Specific Benchmark)也就應運而生，每個此類的標竿皆規範了一個模擬的工作量(Synthetic Workload)，可用來特徵化(Characterize)該問題領域的標準問題。

(三)資料庫績效評估的組成要素

績效評估在電腦界所測試的指標，可以指的是硬體的運作績效，也可是某應用程式的效能。就本研究而言，績效評估所測試的是全文資訊檢索資料庫的效能。一個完整的績效評估除了需有明確的目標之外，必需具備有下列的組成要素：

1. 實驗自變數(Experimented Factors)：

在測試過程中可予以改變的參數或系統組態。舉凡資料庫大小(Database Size)、查詢(Query)的複雜度、系統的設定情形(Configuration)等。

2. 績效指標應變數(Performance Metrics)：

測試的產出結果以指標形式呈現以供比較。一般而言績效指標可以是產出(Throughput)、回應時間(Response Time)、價格/表現比(Price/Performance)。

3. 實驗設計與衡量方式的選擇(Experimental Design & Measurement Techniques)：

在這個階段中，績效評估必須產生工作量(Workload)模式。所謂的工作量的定義即在特定時段內指定要完成的工作，每套績效評估必定擁有並詳細載明其專屬的工作量模式。

(四)全文資訊檢索之績效評估

基本上全文資訊檢索績效評估主要著重在某些硬體設定下文件檢索引擎的效能。其所引用的績效評估衡量指標為每分鐘所完成的搜尋(Search Completed Per Minute：SPM)，並計算每 SPM 的價格/效能比 (Price/Performance Ratio)。直至今日，全文資訊檢索績效評估已廣為使用在 Sequent 電腦系統，其測試的結果指出其為可行之績效評估方法。(Samuel 1993)設計績效評估方法包括組織模式 (Business Model)、資料庫模式 (Database Model) 和系統模式 (System Model)。

1. 組織模式

全文資訊檢索績效評估的運作環境是以多人的全文資訊檢索系統進行文件的檢索為主，我們將此種應用程式稱為文件檢索服務(Document Retrieval Service：DRS)。組織模式是以文件檢索為主、多人應用環境。

2. 資料庫模式

概念上來說資料庫是一群文件區(Partitions)的集合。每個文件區包含一組文件，以變動長度記錄資料型態儲存，每一筆記錄包含一份文件的全部內容，可看成是文件檔案的進入點(Entry)。此外每個文件區段都會有一索引值對應，使 FTIR 能藉由索引增進回應時間。

3. 系統模式

在系統模式中可將每個使用者視為 Client 端對 DRS 所發出的檢索處理要求，使用者輸入感興趣的主題(此為執行搜尋：Search Transactions)；執行檢索(Retrieval Transactions)就在文件中取出對應的內容，而兩種動作的權限只在讀取的層次上。

(五)目前全文資訊檢索績效評估之缺失

(Samuel 1993) 績效評估方法提供之後相關研究以及本研究的參考架構。但是，(Samuel 1993) 的三模式仍有改良的空間。組織模式著重主從式系統架構，並未對資訊檢

索的搜尋、運算、比對、排列組合、重組等重要演算法加以設計測試。資料庫模式已提供不同於表格資料庫的文件區段概念，但是重要的 Surrogate 檔案結構未能設計在內，無法避免 FTIR 軟體以循序性方式搜尋檢索的文件，而浪費時間。系統模式過於簡單，比較類似固定式關鍵字式搜尋，並非 FTIR 檢索。再則 (Samuel 1993) 及相關研究和業界常用方法，均 Domain Specific 會發生不同問題領域產生錯誤評估結果。

參、研究方法與工作量模式之建立

一、研究方法

發展一套績效評估方法必須訂定所測試之工作量。工作量模式由三個次模式所組成，亦即在產生工作量模式的過程，我們必須依序產生如下三個模式，此三個模式之建構為傳播學網際網路圖書館之一般性模式，圖 1 為本研究的研究架構圖。

本研究如下所述，先提出一般化研究模式，透過實作或工程方法，將研究模式建製，不過，本研究僅以傳播學數位圖書館為例。本研究著重於提出目前資訊檢索績效評估在無系統性、無通用性、無代表性，對於有關 Domain Independence 及 User Workload Requirements 重要評估屬性的缺失上，開發開放型、系統化、一般化，以使用者工作量需求為導向的資訊檢索資料庫績效評估工作量模式。本研究以研究底層和基本的工作量模式為宗旨，不以實作不同檢索系統，執行測試程式，產生比較數據。而是希望提供不能描述和表達以及不能產生真實具代表使用者應用環境績效的目前研究困境一項解決方案，核心議題在於工作量模式的產生和設計。因此，在研究模式部分，它成為本論文重心和主要貢獻來源。本研究模式將以 Benchmark Methodology 基本理論將資料模式、交易模式、和控制模式成為一般化載具 (TPC 2001) (Gray 1993) (Turbyfill 1993) (Dewitt 1983)。接著，依據 (Faloutsos, 1985) (Samuel 1983) 對於全文資訊檢索的文件區段定義、進入點定義、代理檔案建立、索引檢索建立、比對、字串搜尋、字元轉置、叢集概念、和雜湊演算法等，產生資料模式、交易模式和控制模式。

圖 1 中我們蒐集資訊檢索資料庫工作量需求，導入工作量模式當中包括資料需求模式、交易需求模式和控制需求模式。使用者輸入工作量需求，系統自動導入產生三項子模式，再經由系統中轉換/翻譯/產生器產生工作量規格書。資料需求規格書以表單方式將欄位式和結構化的資料儲存，並以將來全文資料庫產生的語法儲存，形成未來可程式化的資料庫需求規格書。交易需求規格書以 SQL 及擬程式語言 (Pseudocode) 撰寫而成，撰寫基礎來自於原來使用者輸入和填選的交易資料。控制需求規格書以擬程式語言語法完成，訂定實驗設計和執行的內容和細節，包括測試時間、Startup、Steady State，在冷/暖/熱執行狀態、交易執行順序、交易執行次數、和收集測試中時間統計值，以計算回應時間 (Response Time) 和產出效率 (Throughput)。回應時間為每一單位工作 (Unit of Work) 所花費的時間，產出效率為每一單位時間 (Unit of Time) 所完成的工作單位數，彼此恰巧互為相反運算。在設計和產生資料模式和交易模式時要注意到。本研究為領域獨立、開放型、一般化之合成工作量模式。因此，機率分佈極為重要。機率分佈用來產生測試資料庫、測試交易、和實驗控制變數。理論上，分為 Uniform 機率分佈及 Non-Uniform 機率分佈，都可以設計在三項模式中。本研究實作部分目前僅建置 Normal, Zipf, Weighted, Exponential, Bayesian, 和 Uniform 機率分佈。系統自動按照控制模式要求，資料模式建立資料庫，採用電腦模擬方法模擬交易模式進行測試，蒐集時間統計值。

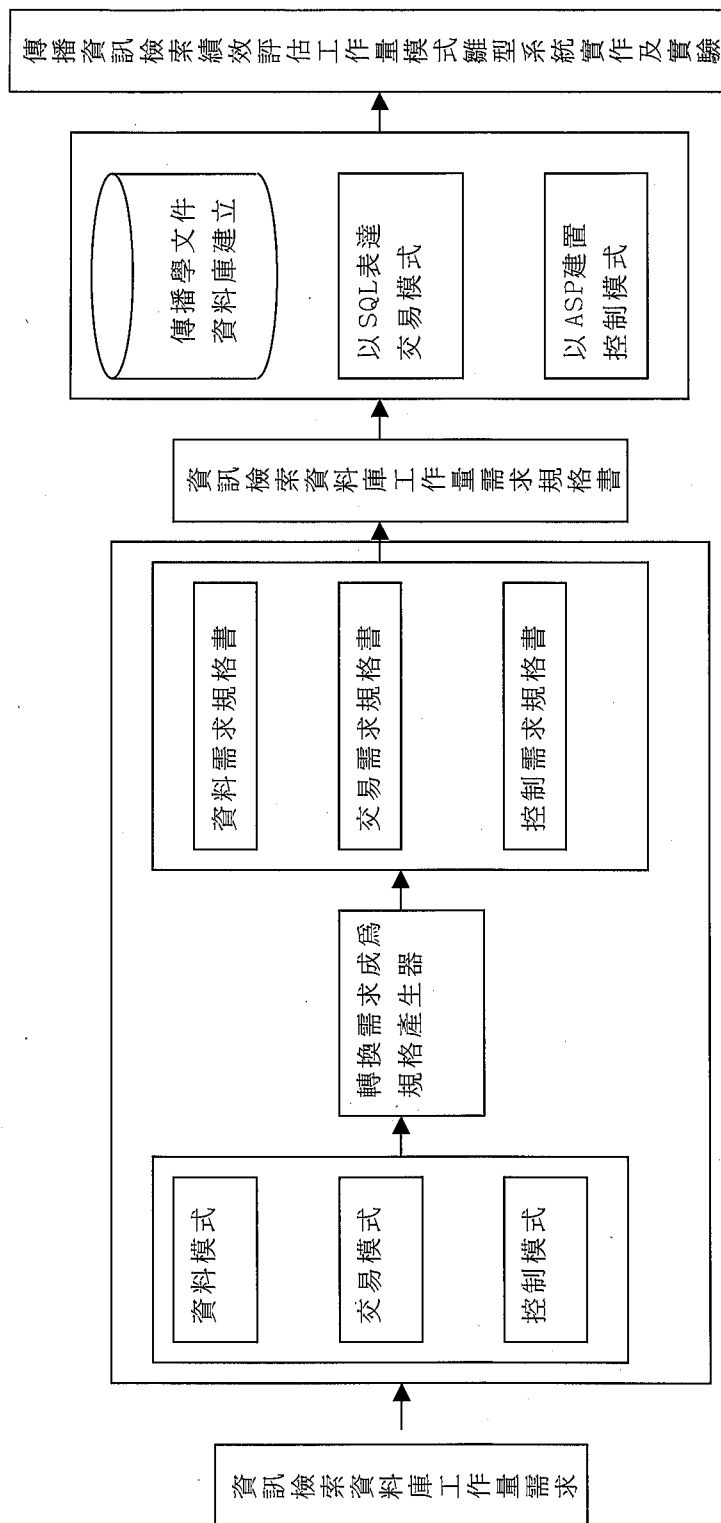


圖 1：本研究的研究架構圖

二、工作量模式之建立

(一)資料模式之建立

傳播學數位圖書館資料模式的細目初步歸納有以下的分類：文獻資料、統計資料、人才與機構資料、以及系統管理時所需之資料。茲以上述各分類為主，簡介各分類為下：

1. 文獻資料

文獻表格裏所收集之資料是以政府、學術機構及其人員自行或委託其他機構或個人所作之傳播相關學術性研究，其範圍可包括期刊論文、政府出版品、會議論文等，每篇文章登錄在資料庫裏的深度內容有作者、書名(刊名)、出版者、出版日期、關鍵字、摘要、全文等。

2. 統計資料

統計表格中所收錄之資料以產業相關單位所發表之統計資料為主，此外也收錄了傳播相關之年鑑資料，如電影年鑑、無線電視年鑑、出版年鑑、網際網路年鑑、電信年鑑等，每筆資料內容包括資料名稱、刊名、篇名、作者、關鍵字、摘要、全文等。

3. 人才與機構資料

人才資料表格中收錄了各相關學校系所、學術單位之學者、研究者、教師之基本資料，其每筆資料內容包含作者姓名、地址、目前所服務之機關、電話、著作等。機構資料擬收錄傳播學術、產業單位及政府相關部門之組織資料，其內容可包括組織名稱、聯絡方式、領域等。

4. 指南資料

指南資料表格中所收錄之資料是以傳播類各期刊書目與年鑑的詳細資訊為主，如登錄各期刊書目的創刊日期、狀況、總期數、發刊期間、出版地、出版者等資料。

5. 系統管理所需之資料

基於便利性之考量，系統管理者與資料收集者應可以自遠端存取資料庫，並作必要之變更，故特別設計一個虛擬工作室，其功能意在提供系統管理者必需由遠端登錄，進行線上資料的新增等動作。這部份資料庫所需之資料有管理者編號、姓名、帳號及密碼、和管理者權限、最近一次存取時間等。

(二)資料模式之架構圖

在此以最常運用的實體關係圖 (Entity-Relationship Diagram)，來表示其各資料表格間的關係。如圖 2 所示。

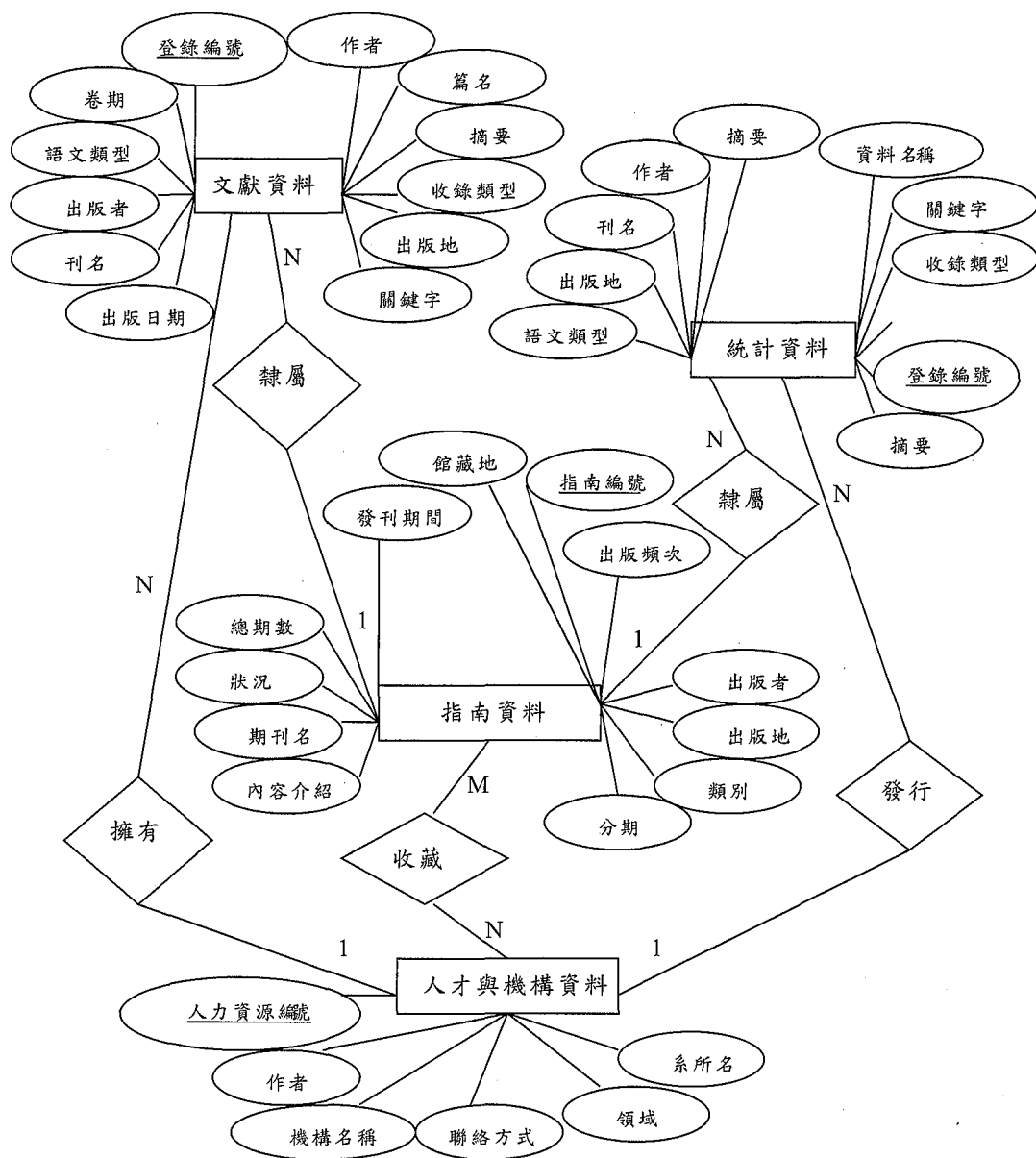


圖 2：資料模式的實體關係圖

(三)交易模式之建立

基本上傳播中文全文資訊檢索資料庫可分單一資訊檢索和多資料庫查詢。而單一資訊檢索和多資料庫查詢的交易模式所具備之功能，初步分類為查詢功能：其又可細分為基本查詢、縮小範圍查詢、進階查詢如基本查詢可依作者、書名、關鍵字、分類時間、及不分欄位的查詢；縮小範圍的查詢功能可依語文類型、出版日期等查詢。此外還包含檢索結果顯示功能、個人資料輸入及遠端動態管理資料庫等，如表 1 及表 2 所列示：

表 1：交易模式之分類及其細目(本研究整理)

功能別	範例
基本查詢	可依作者、書名、關鍵字、及不分欄位來查詢。
縮小範圍查詢	如依語文類型(中文或英文)、出版日期、關鍵字來查詢。
進階查詢	使用者可啟動進階查詢的同音、容錯、中英文近似概念查詢功能，來提高檢索的合理性。
檢索結果顯示	其可包括大綱及細目的搜尋結果顯示。
個人資料輸入	其資料輸入活動包括註冊成為資料庫的成員、使用資料庫的檢索、工作區等功能。
遠端資料庫管理	系統管理者由遠端登錄，進行線上文獻、統計、人才、機構資料的新增、修改、儲存等

表 2：查詢檢索畫面與連結點(本研究整理)

		1.文獻與統計資料	2.指南資料		3.傳播人力資源	
檢 索 點	主要	篇名(書名)、資料名稱、作者、刊名、出版者、關鍵字	--		人名、機構名(校名)、系所名	
	輔助	語文類型、日期、出版地、收錄類別、卷期	--		領域	
檢 索 結 果 大 綱		篇名(書名)、資料名稱、作者、刊名	期刊名(按照:學術傳播、學術傳播相關、非學術傳播、非學術傳播相關)與年鑑		人名、機構名、系所名	
檢 索 結 果 細 目	登錄	連結	登錄	連結	連結	連結
	篇名、資料名稱、作者、刊名、出版者、關鍵字、語文類型、日期、出版地、收錄類型、卷期、摘要	篇名、作者、刊名、出版者、關鍵字、卷期	創刊日期、狀況(現刊\已停刊)、總期數、發刊期間、出版頻次、出版者、出版地、類別、館藏地、內容介紹、目錄(分期)	每期目錄、單篇論文	人名 機構名 聯絡方式 領域	著作(作者)

(四)控制模式之建立

定義資料模式與交易模式之後，績效評估的測試以及所進行的測試交易將有根據。而關於如何進行資料庫的績效評估，則屬控制模式的範疇。控制模式可包含測試的持續時間、查詢順序、穩定的測試狀態、績效衡量指標、測試模式、查詢次數，本研究實作這六項控制變數，茲分述如下：

1.測試的持續時間(Test Duration)

測試的持續時間意謂著績效評估的時間區段。評估的時間區段需要在系統達到穩定的狀態後才能進行評估，此外時間區段的選取須滿足能重複地產生查詢結果及不半途中斷的要求。

2.查詢順序(Test Sequence)

查詢的順序指的是查詢執行的順序，例如有些查詢可複合一起執行，制定其應該的查詢順序，以防止不適當的執行順序。

3.穩定的測試狀態(Steady State)

在此所進行之測試是假設在一個穩定的狀態下進行系統的測試，才能真實反應出系統的效能。

4.績效衡量指標(Performance Metric)

所謂的績效衡量指標指的是回應時間(Response Time)及產出(Throughput)。回應的時間可定義為發出查詢請求至收到查詢回覆所經歷的時間，也就是所謂的回應時間。產出為回應時間的為分母的反值，代表每一單位時間所完成的查詢。

5.測試模式(Test Mode)

測試模式可包含 Cold Mode、Warm Mode 及 Hot Mode。Cold 模式表示在 cache 或 buffer 內沒有之前的資料存在，因此當我們在 Cold 模式時，反應時間是高於 Warm 模式的；而 Warm 模式表示的是 cache 內已有之前的資料，查詢時間會大幅縮短；Hot 模式基本上與 Warm 模式相似，其結果包含第一次以 Cold 模式執行，之後再多次執行同一個查詢，來求其平均回應時間。

6.查詢次數(Number of Repetitions)

查詢次數規定了某交易在同一實驗執行中，所執行的次數。

將以上六個控制變數應用到建構傳播中文全文檢索資料庫工作量的控制模式中，如下所述：

- 1.傳播中文全文資訊檢索資料庫的測試的持續時間：分別找出一天中使用傳播中文全文檢索資料庫的尖峰、離峰與一般狀態的時間區段，是分別位於哪一時候，並針對尖峰、離峰及一般時段測量其時間區段的長度、頻率等。
- 2.傳播中文全文資訊檢索資料庫中的查詢順序：針對資料庫中的各種查詢可能，初步可分為沒有查詢的先後順序之分和有查詢的先後順序，以分別進行評估。如在文獻資料庫中，基本查詢功能可依作者、或依關鍵字等分別進行查詢，此時為無先後順序關係存在的查詢；而有前後順序的查詢，如在文獻資料庫中依作者查詢完後，可再縮小範圍以出版日期進行查詢，將查詢的範圍縮小，此時使用者需先在前一網頁中依作者查詢後，才能在下一網頁裏進行縮小範圍的查詢。
- 3.傳播中文全文資訊檢索資料庫的穩定的測試狀態：所謂的穩定狀態在此可規劃出離峰至

一般時段、一般至尖峰的測試狀態。

4. 傳播中文全文資訊檢索資料庫的績效衡量指標：本研究選取回應時間與產出為評估傳播中文全文資訊檢索資料庫中各查詢的主要衡量依據。
5. 傳播中文全文資訊檢索資料庫的測試模式：對於每一個或每一組查詢分別就 Cold Mode、Warm Mode 及 Hot Mode 模式進行評估，以找出在各個模式下查詢的績效。
6. 傳播中文全文資訊檢索資料庫的查詢次數：衡量傳播中文全文檢索資料庫中單一或一組查詢，在特定的一段時間內重複執行的次數的平均績效。

肆、系統實作與結果分析

一、測試資料庫之建立

(一) 傳播資料庫表格設計

1. 文獻資料表格

文獻資料表格主要收錄傳播相關學術期刊及會議論文(如新聞學研究)、政府出版品與政府委託之傳播研究報告、國科會委託研究計劃裏的各筆詳細資料，如收錄作者、篇名、刊名、著者、出版地、出版者、出版日期、卷期、版次、引用來源、關鍵字、資料描述等。

表 3：文獻資料表格

	欄位名稱	型別	長度(Bytes)	欄位描述(key, null..)
1	登錄編號	int	4	Primary
2	資料名稱	nvarchar	50	Not null
3	篇名	nvarchar	255	Not null
4	刊名	nvarchar	255	Not null
5	著者	nvarchar	255	Not null
6	出版地	nvarchar	20	Null
7	出版者	nvarchar	100	Null
8	出版日期	nvarchar	25	Null
9	卷期	nvarchar	20	Null
10	版次	float	8	Null
11	頁數	nvarchar	20	Null
12	引用來源	nvarchar	255	Null
13	關鍵詞	nvarchar	255	Null
14	資料描述	ntext	16	Null
15	分類	nvarchar	255	Null
16	資料起始週期	nvarchar	255	Null
17	語文類型	nvarchar	10	Null
18	收錄類型	nvarchar	20	Null
19	備註欄	nvarchar	255	Null
20	本筆資料時間	nvarchar	255	Null
21	著者英文名	nvarchar	255	Null
22	篇名英文名	nvarchar	255	Null

23	刊名英文名	nvarchar	255	Null
24	關鍵詞英文	ntext	16	Null
25	資料描述英文	ntext	16	Null

2. 產業資料表格

產業資料表格主要記錄傳播相關年鑑、雜誌，如中華民國電視年鑑、中華民國出版年鑑、廣告雜誌等來源之資料，其中各筆詳細資料收錄資料名稱、篇名、刊名、著者、出版地、出版者、出版日期、卷期、版次、表格名稱、本筆資料時間、引用來源、關鍵字、資料描述等。

表 4：產業資料表格

	欄位名稱	型別	長度(Bytes)	欄位描述(key, null..)
1	登錄編號	int	4	Primary
2	資料名稱	nvarchar	255	Not null
3	篇名	nvarchar	255	Not null
4	刊名	nvarchar	255	Not null
5	著者	nvarchar	255	Not null
6	出版地	nvarchar	255	Null
7	出版者	nvarchar	255	Null
8	出版日期	char	10	Null
9	卷期	char	10	Null
10	版次	char	10	Null
11	頁數	char	20	Null
12	引用來源	nvarchar	255	Null
13	關鍵詞	nvarchar	255	Null
14	資料描述	nvarchar	255	Null
15	分類	nvarchar	255	Null
16	表格名稱	nvarchar	255	Null
17	資料起始週期	nvarchar	255	Null
18	語文類型	char	10	Null
19	收錄類型	char	20	Null
20	本筆資料時間	ntext	16	Null
21	英文期刊名稱	char	40	Null

3. 人才資料表格

人才資料表格主要記錄傳播相關學者專家的個人資料以及其聯絡方式，其中各筆人才詳細資料收錄著各學者的姓名、服務單位、職稱、學歷、專長和電子郵件。

表 5：人才資料表格

	欄位名稱	型別	長度(Bytes)	欄位描述(key, null..)
1	登錄編號	int	4	Primary
2	姓名	char	255	Not null
3	服務單位	Int	255	Null
4	職稱	char	255	Null
5	最高學歷	char	255	Null
6	專長	ntext	255	Null
7	電子郵件	char	255	Null

4. 機構資料表格

機構資料表格主要記錄傳播相關機構如各大學傳播學院等的機構基本資料以及其聯絡方式，其中各機構單位詳細資料收錄著機構名稱、住址、郵遞區號、電話、傳真、網址。

表 6：機構資料表格

	欄位名稱	型別	長度(Bytes)	欄位描述(key, null..)
1	登錄編號	int	4	Primary
2	機構名稱	char	40	Not null
3	住址	char	50	Null
4	郵遞區號	char	5	Null
5	電話	char	30	Null
6	傳真	char	20	Null
7	網址	char	100	Null

(二) 測試資料表格之產生與輸入

1. 決定每資料表格的記錄數

整體的運作環境是在作業系統 Windows NT，其主記憶體為 128MB，根據 (Gray 1993) (Turbyfill 1993) 對 Cold Mode、Warm Mode、Hot Mode 的考量，測試資料庫至少應該與記憶體大小相似或成為倍數，以利績效結果分析，並且避免產生主記憶資料庫的效應。資料庫為 Microsoft SQL Server 主從架構資料庫系統，開發工具為 ASP 及 Java。故定其測試資料庫大小約為 128MB，並以文獻整理中實務上傳播學文獻來源分配各資料表的大小。

表 7：各資料表格之 Tuple 計算表

	資料表格名稱	分配大小 (MB)	每列 Byte 數 (Bytes/Tuple)	Tuples 數
1	文獻表格	99.84	135.64	7,687,506
2	產業表格	25.6	259.13	113,304
3	人才表格	2	6.4	419,430
4	機構表格	1	266.24	3150
總計		128.44		

二、測試查詢和檢索交易之建立

(一)文獻資料查詢

1. 依篇名(Title)

- (1)查詢名稱：以篇名，繼以著者和文獻資料進行查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第二筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，T11 是以查詢在文獻資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

2. 依篇名及啟動 AND 功能查詢

- (1)查詢名稱：啟動 AND 功能，以篇名和作者名稱同時查詢，繼以著者進行深度查詢。
- (2)查詢詳細：以特定的篇名和作者名稱進行查詢，可瀏覽此筆資料的篇名、著者、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，TA11 是在啟動 AND 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

3. 依篇名及啟動 OR 功能查詢

- (1)查詢名稱：啟動 OR 功能，以篇名和作者名稱同時查詢，繼以著者進行深度查詢。
- (2)查詢詳細：以特定的篇名“或”作者名稱進行查詢，則可瀏覽每筆資料的篇名、著者、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，TO11 是在啟動 OR 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

4. 依作者(Author)

- (1)查詢名稱：以作者，繼以著者和文獻資料進行查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，A11 是以查詢在文獻組的資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

5. 依作者及啟動 AND 功能查詢

- (1)查詢名稱：啟動 AND 功能，以作者和刊名名稱同時查詢，繼以著者進行深度查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，AA11 是在啟動 AND 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

6. 依作者及啟動 OR 功能查詢

- (1)查詢名稱：啟動 OR 功能，以作者和刊名同時查詢，繼以著者進行深度查詢。
- (2)查詢詳細：以特定的作者“或”刊名名稱進行查詢，則可瀏覽每筆資料的篇名、著者、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，AO11 是在啟動 OR 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

7. 依刊名(Book)

- (1)查詢名稱：以刊名，繼以著者和文獻資料進行查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，B11 是以查詢在文獻組的資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

8. 依刊名及啟動 AND 功能查詢

- (1)查詢名稱：啟動 AND 功能，以刊名和關鍵字同時查詢，繼以著者進行深度查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，BA11 是在啟動 AND 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

9. 依刊名及啟動 OR 功能查詢

- (1)查詢名稱：啟動 OR 功能，以刊名和關鍵字同時查詢，繼以著者進行深度查詢。
- (2)查詢詳細：以特定的刊名名稱“或”關鍵字進行查詢，則可瀏覽每筆資料的篇名、著者、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，BO11 是在啟動 OR 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

10. 依關鍵字(Keyword)

- (1)查詢名稱：以關鍵字，繼以著者和文獻資料進行查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，K11 是以查詢在文獻資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

11. 依不分欄位檢索(Non-Field)

- (1)查詢名稱：以不分欄位方式檢索，繼以著者和文獻資料進行深度查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，N11 是以查詢在文獻資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

(二)統計資料查詢

1. 依資料名稱(Title)

- (1)查詢名稱：以資料名稱，繼以著者和文獻資料進行查詢。
- (2)查詢詳細：可瀏覽每筆資料的資料名稱、刊名；此外可選擇想閱覽的特定筆數，如第

一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，ST11 是以查詢在統計表格的資料筆數為主，最終可查詢出此作者在統計資料庫中所擁有的文章總數及內容。

2. 依資料名稱及啟動 AND 功能查詢

(1) 查詢名稱：啟動 AND 功能，以資料名稱和刊名名稱同時查詢，繼以著者進行深度查詢。

(2) 查詢詳細：以特定的資料名稱和刊名名稱進行查詢，可瀏覽此筆資料的資料名稱、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，STA11 是在啟動 AND 功能後，最終可查詢出此作者在統計資料庫中所擁有的文章總數及內容。

3. 依資料名稱及啟動 OR 功能查詢

(1) 查詢名稱：啟動 OR 功能，以資料名稱和刊名名稱同時查詢，繼以著者進行深度查詢。

(2) 查詢詳細：以特定的資料名稱“或”刊名名稱進行查詢，則可瀏覽每筆資料的資料名稱、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，SNO11 是在啟動 OR 功能後，最終可查詢出此作者在統計資料庫中所擁有的文章總數及內容。

4. 依刊名 (Book)

(1) 查詢名稱：以刊名，繼以著者和文獻資料進行查詢。

(2) 查詢詳細：可瀏覽每筆資料的資料名稱、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，SB11 是以查詢在統計表格的資料筆數為主，最終可查詢出此作者在統計資料庫中所擁有的文章總數及內容。

5. 依不分欄位檢索(Non-Field)

(1) 查詢名稱：以不分欄位方式檢索，繼以著者和文獻資料進行深度查詢。

(2) 查詢詳細：可瀏覽每筆資料的資料名稱、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，SN11 是以查詢在統計表格的資料筆數為主，最終可查詢出此作者在統計資料庫中所擁有的文章總數及內容。

(三) 人力與機構資料查詢

1. 依姓名 (People Name)

(1) 查詢名稱：以姓名，繼以文獻資料進行查詢。

(2) 查詢詳細：可瀏覽每個人的姓名、所屬單位；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此筆資料的詳細，例如姓名、職稱、所屬機構名稱、最高學歷、專長、在文獻表格資料等；在此以選取文獻資料進行查詢，可列出此作者在本資

料庫中的文獻表格擁有的文章總數及內容。

2. 依姓名及啟動 AND 功能查詢

- (1) 查詢名稱：啟動 AND 功能，以姓名和機構名稱同時查詢，繼以文獻資料進行深度查詢。
- (2) 查詢詳細：以特定的人名和機構名稱進行查詢，可瀏覽每個人的姓名、所屬單位，此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此筆資料的詳細，例如姓名、職稱、所屬機構名稱、最高學歷、專長、在文獻表格資料等；在此以選取文獻資料進行查詢，可列出此作者在本資料庫中的文獻表格擁有的文章總數及內容。

3. 依姓名及啟動 OR 功能查詢

- (1) 查詢名稱：啟動 OR 功能，以姓名和機構名稱同時查詢，繼以文獻資料進行深度查詢。
- (2) 查詢詳細：以特定的人名“或”機構名稱進行查詢，可瀏覽每個人的姓名、所屬單位，此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此筆資料的詳細，例如姓名、職稱、所屬機構名稱、最高學歷、專長、在文獻表格資料等；在此以選取文獻資料進行查詢，可列出此作者在本資料庫中的文獻表格擁有的文章總數及內容。。

4. 依機構名稱 (Organization)

- (1) 查詢名稱：以機構名稱，繼以文獻資料進行查詢。
- (2) 查詢詳細：可瀏覽每個人的姓名、所屬單位；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此筆資料的詳細，例如姓名、職稱、所屬機構名稱、最高學歷、專長、在文獻表格資料等；在此以選取文獻資料進行查詢，可列出此作者在本資料庫中的文獻表格擁有的文章總數及內容。

5. 依不分欄位方式檢索 (Non-Field)

- (1) 查詢名稱：以不分欄位方式檢索，繼以文獻資料進行深度查詢。
- (2) 查詢詳細：可瀏覽每個人的姓名、所屬單位；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此筆資料的詳細，例如姓名、職稱、所屬機構名稱、最高學歷、專長、在文獻表格資料等；在此以選取文獻資料進行查詢，可列出此作者在本資料庫中的文獻表格擁有的文章總數及內容。

(四) 縮小範圍查詢和檢索

1. 依語文類型 (Language Type)

- (1) 查詢名稱：以語文類型，繼以著者和文獻資料進行查詢。
- (2) 查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，L11 是以查詢在文獻資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

2. 依語文類型及啟動 AND 功能查詢

- (1) 查詢名稱：啟動 AND 功能，以語文類型和出版日期同時查詢，繼以著者進行深度查詢。
- (2) 查詢詳細：以語文類型和出版日期進行查詢，可瀏覽此筆資料的篇名、著者、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵

詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，LA11 是在啟動 AND 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

3. 依語文類型及啟動 OR 功能查詢

(1)查詢名稱：啟動 OR 功能，以語文類型和出版日期同時查詢，繼以著者進行深度查詢。

(2)查詢詳細：以語文類型“或”出版日期進行查詢，則可瀏覽每筆資料的篇名、著者、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，LO11 是在啟動 OR 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

4. 依出版日期(Publish Date)

(1)查詢名稱：以出版日期，繼以著者和文獻資料進行查詢。

(2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，PD11 是以查詢在文獻資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

5. 依出版日期及啟動 AND 功能查詢

(1)查詢名稱：啟動 AND 功能，以出版日期和出版地同時查詢，繼以著者進行深度查詢。

(2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，PDA11 是在啟動 AND 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

6. 依出版日期及啟動 OR 功能查詢

(1)查詢名稱：啟動 OR 功能，以出版日期和出版地同時查詢，繼以著者進行深度查詢。

(2)查詢詳細：以特定的出版日期“或”出版地進行查詢，則可瀏覽每筆資料的篇名、著者、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，PDO11 是在啟動 OR 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

7. 依出版地查詢

(1)查詢名稱：以出版地，繼以著者和文獻資料進行查詢。

(2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，PS11 是以查詢在文獻資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

8. 依出版地及啟動 AND 功能查詢

(1)查詢名稱：啟動 AND 功能，以出版地和收錄類型同時查詢，繼以著者進行深度查詢。

- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，PSA11 是在啟動 AND 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

9. 依出版地及啟動 OR 功能查詢

- (1)查詢名稱：啟動 OR 功能，以出版地和收錄類型同時查詢，繼以著者進行深度查詢。
- (2)查詢詳細：以特定的出版地“或”收錄類型進行查詢，則可瀏覽每筆資料的篇名、著者、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，PSO11 是在啟動 OR 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

10. 依收錄類型(Collect Type)

- (1)查詢名稱：以收錄類型，繼以著者和文獻資料進行查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，C11 是以查詢在文獻資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

11. 依收錄類型及啟動 AND 功能查詢

- (1)查詢名稱：啟動 AND 功能，以收錄類型和關鍵字同時查詢，繼以著者進行深度查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，CA11 是在啟動 AND 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

12. 依收錄類型及啟動 OR 功能查詢

- (1)查詢名稱：啟動 OR 功能，以收錄類型和關鍵字同時查詢，繼以著者進行深度查詢。
- (2)查詢詳細：以特定的收錄類型“或”關鍵字進行查詢，則可瀏覽每筆資料的篇名、著者、刊名；點選符合的資料筆數之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，CO21 是在啟動 OR 功能後，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

13. 依關鍵字(Keyword)

- (1)查詢名稱：以關鍵字，繼以著者和文獻資料進行查詢。
- (2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，LK11 是以查詢在文獻資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

14. 依不分欄位檢索(Non-Field)

(1)查詢名稱：以不分欄位方式檢索，繼以著者和文獻資料進行深度查詢。

(2)查詢詳細：可瀏覽每筆資料的篇名、著者、刊名；此外可選擇想閱覽的特定筆數，如第一筆資料，點選之後會列出此篇文章的詳細資料，例如出版地、出版日期、關鍵詞、版次、著者、出版者、卷期等；在此以選取著者進行查詢，可列出此作者在本資料庫中擁有的文章總數，LN11 是以查詢在文獻資料筆數為主，最終可查詢出此作者在文獻資料庫中所擁有的文章總數及內容。

表 8：總查詢代號彙整表

查詢代號	交易名稱	查詢和 檢索代號	交易名稱	查詢和 檢索代號	交易名稱
T	以篇名資料進行查詢	ST	以資料名稱資料進行查詢	L	以語文類型資料進行查詢和檢索
TA	啟動 AND 功能，同時以篇名和其它查詢條件進行查詢	STA	啟動 AND 功能，同時以資料名稱和其它查詢條件進行查詢	LA	啟動 AND 功能，同時以語文類型和其它查詢條件進行查詢和檢索
TO	啟動 OR 功能，同時以篇名或其它查詢條件進行查詢	STO	啟動 OR 功能，同時以資料名稱或其它查詢條件進行查詢	LO	啟動 OR 功能，同時以語文類型或其它查詢條件進行查詢和檢索
A	以作者資料進行查詢	SB	以刊名資料進行查詢	PD	以出版日期資料進行查詢
AA	啟動 AND 功能，同時以作者和其它查詢條件進行查詢	SN	以不分欄位方式檢索	PDA	啟動 AND 功能，同時以出版日期和其它查詢條件進行查詢
AO	啟動 OR 功能，同時以作者或其它查詢條件進行查詢	PN	以人名資料進行查詢	PDO	啟動 OR 功能，同時以出版日期或其它查詢條件進行查詢
B	以刊名資料進行查詢	PNA	啟動 AND 功能，同時以人名和其它查詢條件進行查詢	PS	以出版地資料進行查詢
BA	啟動 AND 功	PNO	啟動 OR 功能，	PSA	啟動 AND 功

	能，同時以刊名和其它查詢條件進行查詢		同時以人名或其它查詢條件進行查詢		能，同時以出版地和其它查詢條件進行查詢
BO	啟動 OR 功能，同時以刊名或其它查詢條件進行查詢	NF	以不分欄位方式進行查詢	PSO	啟動 OR 功能，同時以出版地或其它查詢條件進行查詢
K	以關鍵字資料進行查詢	G	以機構名稱資料進行查詢	C	以收錄類型資料進行查詢和檢索
N	以不分欄位方式進行查詢	CA	啟動 AND 功能，同時以收錄類型和其它查詢條件進行查詢和檢索	CO	啟動 OR 功能，同時以收錄類型或其它查詢條件進行查詢和檢索
				LK	以關鍵字資料進行查詢和檢索
				LN	以不分欄位方式進行查詢和檢索

三、績效評估之執行

傳播中文全文資訊檢索系統包含文獻、產業、人力與機構三個資料庫，其中又以文獻資料庫資料數量最多，且查詢和檢索類別最為複雜、多樣；而產業及人才與機構資料庫的查詢設計與文獻相同，因此本實驗先針對文獻資料庫進行測試。文獻查詢可分依書名、篇名、作者、關鍵字和不分欄位方式查詢；此外也可使用 AND 和 OR 邏輯運算元輔以查詢，據此本研究將分別以書名、篇名、作者、關鍵字方式，不分欄位，AND 及 OR 等三種方式進行實驗。即第一次各別以書名、篇名、作者、關鍵字查詢類別測試；第二次以不分欄位方式測試；第三次加入 AND 和 OR 邏輯運算元進行連集和交集查詢測試。以求取最佳測試結果。如圖 3 所示：

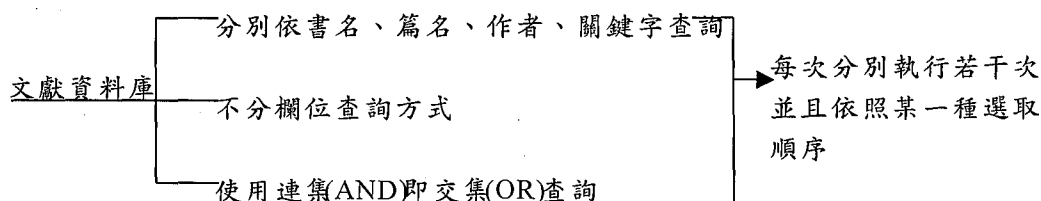
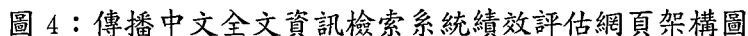


圖 3：查詢和檢索類別執行流程示意圖

此外各次查詢類別之子類別是屬於資料庫中的 Selection 查詢，其中選取因子

圖 4 為實驗系統網頁的架構分佈圖。接下來將詳細介紹每個執行之步驟。



如圖 5 所示，此網頁為本系統的主網頁，其主要目的是超連結到三個系統部分：分別是資料表格總覽、執行預設測試流程、績效評估結果查詢。

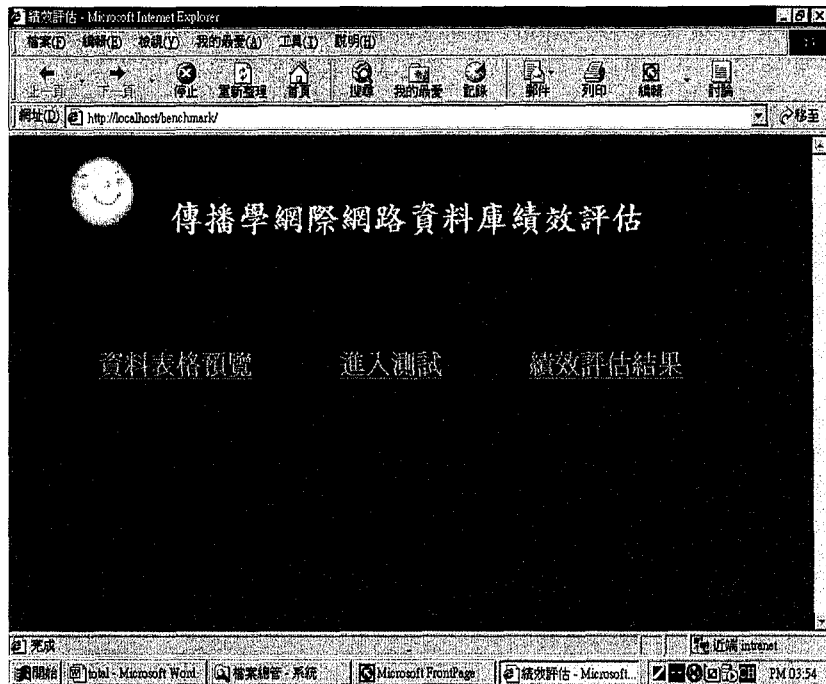


圖 5：傳播中文全文資訊檢索系統績效評估主網頁

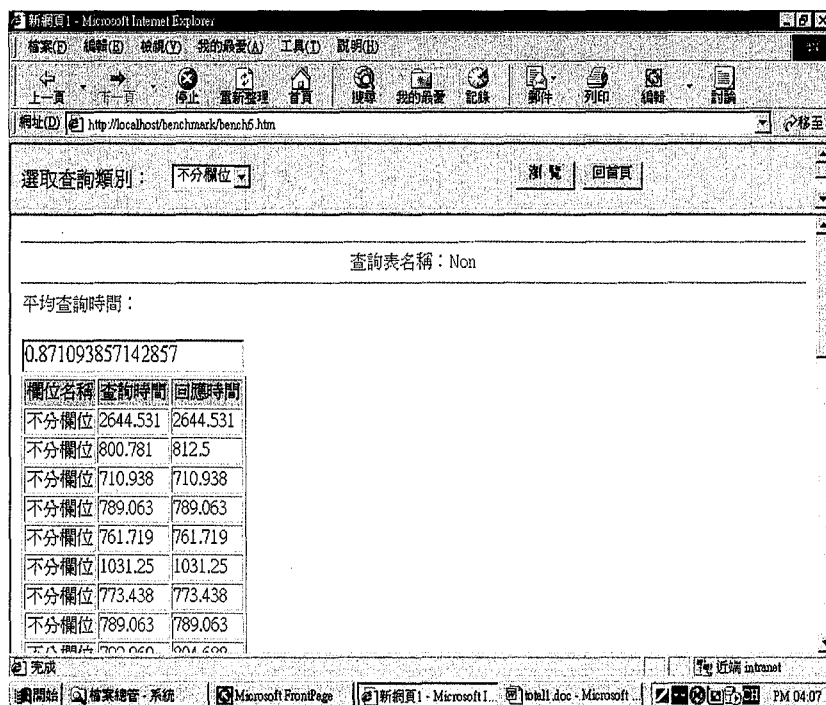


圖 6：績效評估之結果(依不分欄位查詢)

由圖 6 中可按一下「績效評估結果」，會出現回應時間、查詢處理時間的資料表格(時間單位為毫秒)，同時計算其平均時間，單位為秒，至於產出效率，將於整體測試完成時，自系統首頁中選取查詢績效評估結果。

由於本論文實作為配合政治大學傳播學院建立全國傳播學數位資料庫體系著重資訊檢索和資料分類所產生的查詢回應時間，因此系統本身顯示查詢處理時間(Response Time)，但是沒有顯示產能(Throughput)。根據學者(DeWitt 1983) (Gray 1993) (Turbyfill 1993) (Cattell 1993) (TPC 1999)定義績效評估指標，查詢處理時間和查詢處理產能，因為互為相反變數(Inverse Performance Metrics)，可以前後倒算，不能將其視為兩個獨立(Independent)的績效評估指標。但是在 AS3AP 當中所定義的績效評估指標(Equivalent Database Size and Ratio)，因為同時考慮資料庫的大小和資料庫處理查詢時間則需要顯示，因為它是一個獨立的績效評估指標。本論文實作未能同時列出查詢處理時間和查詢處理數量是本論文實作一項缺點，未來應該進行更多績效評估指標的運算和呈現。

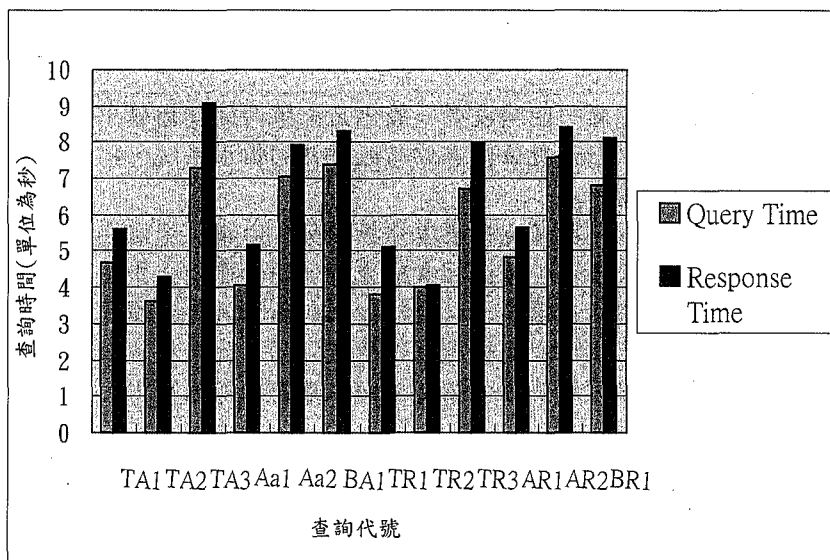


圖 7：使用 AND 和 OR 方式查詢

圖 7 針對個別依篇名(T1)、刊名(B1)、作者(A1)、關鍵詞(K1)查詢測試結果來看，分析得知依篇名、刊名、著者、關鍵詞進行查詢時，查詢時間差異不大，其執行四次後的平均時間大多在 1-3 秒之間，代表資訊檢索資料庫對於建立索引的查詢執行較快並且不分是哪一項主要欄位被查詢。惟獨依關鍵字進行查詢時，查詢時間加長，平均達 5 秒左右，所造成的原因是因為關鍵字查詢設計為不只要比對「關鍵詞」欄位，此外還需逐一比對「資料描述欄位」(記錄每篇文章的摘要性資料)，因此查詢時間明顯地比其它個別的查詢來的長。當使用者隨意輸入任何字元，此查詢會將其與資料庫中所有欄位裏的資料一一比對，因此其查詢時間是加倍地增加，不過與依篇名、刊名、作者、關鍵字查詢相比仍高出許多，平均竟達 17.5 秒左右。可以看出 1-3 秒、5 秒、17.5 秒在有主要鍵、無主要鍵、有索引、無索引，系統執行查詢時演算法的選擇和存取路徑的選擇受到很大的影響。只要使用者啟動 AND 或 OR 功能進行查詢時，兩兩欄位的比對時間比單一欄位的比對時間明顯地加長。

就如上述所說，若 AND 或 OR，即連集或交集到「關鍵詞」欄位，我們可發覺查詢時間較長，TA3(篇名和關鍵詞)、AA2(作者和關鍵詞)、BA1(刊名和關鍵詞)、TO3(篇名或關鍵詞)、AO2(作者或關鍵詞)、BO1(刊名或關鍵詞)等查詢，其查詢時間較其它查詢較長，這是因為都有搭配關鍵詞欄位進行查詢。

伍、結語與未來研究方向

一、結語和討論

網路的出現提供了一個人類知識累積的平台，使我們逐步邁向知識導向的社會。未來的資訊資源將會集中在網際網路上，網際網路也將取代現有的傳播方式成為主要的資訊傳遞、檢索與擷取的媒介，這其中全文式的資訊檢索系統便成了未來此一社會人人需學習建構的基本功能。目前國內相關的自由語文資訊檢索系統研究仍著重在使用者使用自由語文資訊檢索系統的認知行為，卻鮮少有研究者將重點放在評估各類應用領域中的全文或自由語文資訊檢索系統。

在本研究中，我們期望能建置一個開放型一般化需求導向的傳播學中文自由語文資訊檢索系統的工作量模式，並建立一個數位圖書館績效評估的雛形系統，以滿足自由語文資訊檢索資料庫系統效能評估的需求。

在研究的過程中，我們仔細、嚴謹地收集與分析中文自由語文資訊檢索系統的基本概念，與實際運作所涵蓋的功能面為何，藉由詳盡地分析，逐步建構出中文自由語文資訊檢索系統的雛型架構，繼而以此架構為基準，以關連式資料庫績效評估模式的理論與實作基礎，進行傳播學中文自由語文資訊檢索資料庫系統績效評估工作量模式的建構，並在最後開發了一個可供運作的數位圖書館績效評估雛型系統，來衡量傳播學中文自由語文資訊檢索資料庫系統的效能。

本研究主要貢獻在於首先提出以標準載具方式來模式化中文傳播學資訊檢索資料庫績效評估工作量，希望以取得使用者工作量需求，產生績效評估的測試內容，能夠得到符合和代表使用者應用環境中的績效評估結果，是公允合理可靠的。相較於產業標準 TPC 及 (Samuel 1993) 和國內線上搜尋引擎實務測試方法，過於單一和專屬。由於事前設定造成與應用環境不相容工作量。本研究提供開放型使用者導向，可以因隨應用環境改變定義和產生不同工作量。標準載具的方法很重要，它需要根據 Benchmark Methodology 的理論基礎以及 Information Retrieval 的理論基礎，設計和開發一般化工作量模式。標準載具即以資料結構和資料運算為基礎，使用三項子模式形成框架，承載變化中資料需求、交易需求、和控制需求，實際建置-中文傳播學數位圖書館為例，進行實驗測試驗證本研究模式的可行性。

本研究雛型系統可以使 (Samuel 1993) 以文件區段及國內線上搜尋引擎以流量和超連結的工作量模式，用我們的工作量模式產生證明他們是本研究之工作量模式之子集合。再則資訊檢索文獻中，測試每秒完成多少搜尋，每次搜尋花多少時間、資料庫大小如何影響績效評估，都可以使用本研究的自變數和因變數設計在工作量中，使用基本時間統計值和資料統計值來產生不同要求的績效表達和數據。因為，基本時間統計值、資料結構統計值、交易分佈和產生的統計值，都記錄和儲存在本工作量模式中。

二、研究限制與未來研究之方向

本研究之研究重心著重並限制於建構一中文自由語文資訊檢索資料庫系統的工作量模式，其後續的研究方向可朝以下所述各點進行：

- 1.由 Free-Text Information Search 擴充至 Full-Text Information Search 來探討全文資料庫中實際 Full-Text 檢索機制的效能評估。
- 2.本研究所探討之中文自由語文資訊檢索系統，著重在文字資料格式的存取效能評估上，未來可進行多媒體之中文全文資訊檢索系統之績效評估研究。
- 3.未來中文資訊檢索資料庫系統的存取技術是朝向所謂的智慧型資訊檢索技術，如智慧型代理人(Intelligent Agent)等，可以此研究所奠定的基礎探討此領域之全文資訊檢索機制的效能。

參考文獻

- 1.卜小蝶，「圖書資訊檢索技術」，文華圖書館資訊股份有限公司，民國 85 年 11 月。
- 2.卜小蝶，「Internet 資源搜尋系統的發展與運用」，大學圖書館期刊，2 卷，1 期，民國 87 年 1 月，pp. 36-54。
- 3.簡立峰、卜小蝶，「中文全文檢索」，倚天雜誌，Jun 1992，pp.196-201。
- 4.簡立峰、卜小蝶，「中文全文檢索」，倚天雜誌，July 1992，pp.198-202。
- 5.張品臻，「數位圖書館」，圖書館學與資訊科學，卷 24，民國 87 年，pp. 13-28。
- 6.蔡明月，「線上資訊檢索-理論與應用」，台灣學生書局，民國 80 年。
- 7.謝清俊，「全文檢索的方法」，科學月刊，第 19 卷，第四期，April 1988，pp.263-267。
- 8.徐惠文，「全文資料庫的發展與現況」，科學月刊，卷 19，期 4，民國 77 年 4 月，pp. 248-260。
- 9.陳光華，「資訊檢索查詢之自然語言處理」，圖書館學會會報，第 57 期，1996，pp. 141-153。
- 10.Bai, B., Chien, L. and Lee, L., "Very-Large- Vocabulary Mandarin Voice Message File Retrieval Using Speech Queries", Proceedings of the 1996 International Conference on Spoken Language Processing, October 1996.
- 11.Bitton, C. and Turbyfill, C., "A Retrospective on the Wisconsin Benchmark", Readings in Database System, Ed. by M.Stonebraker, Morgan Kaufmann, Inc., 1988.
- 12.Cattell, R. G. G. and Skeen, J., "An Engineering Database Benchmark", The Benchmark Handbook for Database and Transaction Processing Systems, Ed. by J. Gray., Morgan Kaufman Inc., 1993, pp.397-434.
- 13.Carey, J., Dabid J., Jeffery F., Asgarian, M., Brown, P., Johannes E., and Grhrke, N., "The BUCKY Object-Relational Benchmark", ACM SIGMOD, 1997, pp.135-146.
- 14.Carey, J., Dabid J., and Jeffery F., "The 007 Benchmark", Computer Science

- Department, University of Wisconsin-Madison, Version of April 12 1993.
15. DeFazio, S., and Hull, J., "Toward servicing textual database transactions on symmetric shared memory multiprocessors", Proceedings of the Fourth International Workshop on High Performance Transaction Processing Systems, Asilomar Conference Center, Sep 1991.
 16. Dewitt, "A Systematic Approach to Evaluate Database Systems," ACM SIGMOD, 1983.
 17. Faloutsos, C., "Access Methods for text", ACM Computing Surveys, March 1985, pp. 49-74.
 18. Gauch, S., "Intelligent Information Retrieval : An introduction", Journal of the American Society for Information Science, Vol. 43, 1992, pp. 119-136.
 19. Gray, J., The Benchmark Handbook for Database and Transaction Processing Systems, Morgan Kaufmann, Inc., 1993.
 20. Harrington, W. J., "A brief history of computer-assisted legal research", Law Library Journal, 1984-1985, pp. 543-556.
 21. Workgroup application systems, full-text retrieval systems market review and forecast, 1989/90-1995, Framingham, MA : International Data Corp., February 1991.
 22. James K., Jeetu P., Joe F., Paul B., Dan C., Dhaval J., and Yue W., "Internet Text Retrieval Benchmark Study", INFORM, April 1996, pp. 24-44.
 23. Losee, R.M., "An Analytic Measure Predicting Information Retrieval System Performance", Information Processing & Management, Vol. 27, No. 1, 1991, pp. 1-13.
 24. Malone, L., Driscoll, J., and Pepe J., "Modeling The Performance Of An Automated Keywording System", Information Processing & Management, Vol. 27, Nos. 2/3, 1991, pp. 145-151.
 25. Medina, R., "Doculabs' Records Management Systems Benchmark", INFORM, Sep 1998, pp. 22-32.
 26. Muddamalle, M.R., "Natural Language versus Controlled Vocabulary in Information Retrieval : A Case Study in Soil Mechanics", Journal Of The American Society For Information Science, 49, August 1998, pp.880-887.
 27. Patel, J., and Fenner, J., "Document Management In Cyber Space—Doculabs evaluates Web-enabled DM Products", INFORM, Feb 1998, pp. 10-14.
 28. Patel, J., Fenner, J., and Watson, J., "Doculabs Looks at Document Management", INFORM, Jan 1997, pp. 20-31.
 29. Pollitt A. S., "CANSEARCH : An Expert System Approach to Document

- Retrieval”, Information processing & Management, Vol. 23, 1987, pp. 119-136.
30. Robert R. Mueller, “ Information Retrieval Tools for the Internet ”, In Datapro Management of Application Software, Delran, N.J : Datapro Research Corp., 1997.
31. Robertson, S.E., and Beaulieu, M.M., “ On The Evaluation Of IR Systems ” , Information processing & Management, Vol. 28, No. 4, 1992, pp. 457-466.
32. Samuel D., “ Overview of the Full-Text Document Retrieval Benchmark ”, The Benchmark Handbook for Database and Transaction Processing System, Ed. by Jim Gray, Morgan Kaufmann, Inc., 1993, pp.435-443.
33. Sawyer, T., “ Doing your Own Benchmark ”, The Benchmark Handbook for Database and Transaction Processing Systems, Ed. by J. Gray, Morgan Kaufmann, Inc., 1993, pp.21-40.
34. Smith, R., and Chang, S., “ VisualSEEK : A Fully Automated Content-based Image Query System”, June 1996.
35. (<http://www.ctr.columbia.edu/~jrsmith/html/pubs/acmm96/acm.html>)
36. Stonebraker, M., J. Frew, K. Gardels, and J. Meredith, "The Sequoia 2000 Storage Benchmark," Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, May 1993, pp. 2-11.
37. Su, L., “ Evaluation Measures For Interactive Information Retrieval ”, Information Processing & Management, Vol. 28, No. 4, 1992, pp. 503-516.
38. “ TPC Benchmark A Standard Specification Revision 1.1 ”, Transaction Processing Performance Council, 1 March 1992.
39. “ TPC Benchmark B Standard Specification Revision 1.1 ”, Transaction Processing Performance Council, 1 March 1992.
40. “ TPC Benchmark C Standard Specification ”, Transaction Processing Performance Council, 13 August 1992.
41. “ TPC Benchmark D Standard Specification Revision 2.0 ”, Transaction Processing Performance Council, 26 June 1998.
42. “ TPC Benchmark H Standard Specification Revision 1.0 ”, Transaction Processing Performance Council, Apr 1998.
43. “ TPC Benchmark R Standard Specification ”, Transaction Processing Performance Council, Apr 1998.
44. “ TPC Benchmark W Standard Specification ”, Transaction Processing Performance Council, Apr 2001.
45. Turbyfill, C., Orji, C. and Bitton, D., “ AS3AP-An ANSI SQL standard Scaleable and Portable Benchmark for Relational Database Systems ”, The Benchmark Handbook

for Database and Transaction Processing System, Ed by Jim Gray, Morgan Kaufmann, Inc., 1993.

- 46.<http://www.benchmarkresources.com> Benchmark 線上書籍
- 47.<http://www.elandnetwork.com/eLand/main.html> 意藍科技股份有限公司
- 48.http://www.openfind.com.tw/About_us/index.html 網擎資訊股份有限公司
- 49.<http://www.topology.com.tw/prog/default.asp> 拓璞科技股份有限公司
- 50.<http://www.tornado.com.tw/> 龍捲風科技股份有限公司
- 51.<http://www.tpc.org> TPC 協會的網站