

104年特種考試地方政府公務人員考試試題

代號：32440

全一頁

等 別：三等考試

類 科：圖書資訊管理（選試英文）

科 目：資訊系統與資訊檢索

考試時間：2小時

座號：_____

※注意：(一)禁止使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

- 一、早期稱為圖書館自動化系統的資訊系統，近年來已逐漸改稱為整合式圖書館管理系統（Integrated Library Management System, ILMS）。最近數年，雲端圖書館管理系統（Cloud Library Management System, CLMS）亦受到大眾的矚目，部分雲端圖書館管理系統供應商允許圖書館採購個別的功能模組（例如僅購買編目功能模組），因此便可能形成某些功能模組的分散運轉情形。請討論整合式圖書館管理系統與功能模組分散運轉模式的優缺點。（25分）
- 二、一般而言，全文檢索系統會以全文內容，讓使用者進行檢索。然而，文件通常會分為數個區塊（Zone），以學術文件為例，可以區分為：題名、作者、單位、聯絡資訊、摘要、內文、參考文獻等部分。請討論一個不分區塊的檢索系統，以及一個分區塊的檢索系統，對於系統建置者與使用者而言，分別有何利弊得失，可以由成本面向與功能面向討論。（25分）
- 三、近年來，圖書館的編目實務漸漸引入FRBR（Functional Requirements for Bibliographic Records）相關的精神，企圖建構圖書館館藏資源之間的關係，例如莎士比亞的馬克白劇本（實體書）與莎士比亞的馬克白舞台劇（DVD）。請討論對於圖書館的OPAC（Online Public Access Catalogue）模組或是最近備受重視的資源探索服務（Discovery Service）而言，如何將這樣的關係引入檢索系統，讓使用者在進行資訊檢索時，可以得到提供這種關係的好處。（25分）
- 四、資訊檢索的評估可由質性的面向切入，亦可由量化的面向切入。1960年代的Cranfield計畫開創量化的資訊檢索評估模式，1992年啟動的TREC（Text Retrieval Conference）基於Cranfield計畫，建構了現今大規模量化資訊檢索評估的模式，請說明並比較Cranfield的資訊檢索評估模式與TREC的資訊檢索評估模式。（25分）

□ 申論題解答

一、【擬答】

起：直接破題，論述整合性圖書館自動化系統

承：圖書館服務平台的興起與特色

轉：比較兩套系統的分散運轉與整合的優缺點

合：

圖書館自動化系統，又稱整合型圖書館系統，為一整合資料與管理資訊系統，包括流通、編目、採訪、期刊、公用目錄查詢模組以執行圖書館各種作業功能，乃至於參考服務與館際互借與利用資料挖掘理論而有決策資源系統。總之，整合型圖書館自動化系統是圖書館核心的資訊系統，目的在於整合圖書館各種資料庫成為一個整理系統，以分享共用資料庫，並從不同模組分享資訊，減少資料重複。隨著時間的演變與資訊技術的發展，圖書資訊系統所面對的環境也有了重大變遷，近廿年來的網路化、數位化、Google 化、web 2.0 化與手機行動化改變了圖書資訊系統的演變與發展，包括加入 URL，擴展既有的館藏範圍與項目；全球資訊網服務化（Web Service），採用 OpenURL 作為服務的語法結構，以利不同系統間的串連與查詢，擴展查詢範圍；提供聯合查詢功能，除了傳統的館藏資源外，更將圖書館內其他資訊系統與資料庫整合提供一致化的查詢功能；提供 web 2.0 功能，包括社會性標籤、後分類（層面分析）。

Ken Chad 認為有五大資訊科技趨勢影響圖書館自動化系統尤深，分別是一、內容（content）。圖書館管理系統不僅要能夠管理實體與紙本作品，更能夠進一步管理數位內容與購買的流程管理。二、資訊架構與雲端運算（infrastructure/cloud computing）。新一代圖書館自動化管理系統可透過雲端資訊系統的建置，讓各館原本獨立的資源共享。三、開放性（open）：包括開放原始碼、開放資料、開放內容與開放創新的方式。四、商業智慧與分析（business intelligence/analytics）：大數據與雲端環境的影響下，藉由資料挖掘找出對組織決策分析有用的數據。五、消費（consumption）：消費行為不是僅有閱讀，而是包括社群行為與內容。在這樣的趨勢引領下，圖書館服務平台（Library Service Platform，簡稱 SP），即雲端圖書館管理系統於近幾年開始興起。Marshall Breeding 將圖書館服務平台分成三個面向來解析，一、Library：特別為圖書館而設計的軟體，可協助圖書館內部流程改善、館藏管理、滿足館員與讀者需求與傳遞服務。二、Service：以服務為導向的架構、開放的網路服務與 API 與讓圖書館可提供讀者更優化與簡化的服務。三、Platform：提供圖書館自動化資訊平臺架構，結合雲端平臺即服務的實際應用，並可讓圖書館的資訊人員利用開放性的 API 延伸功能，並可與其他圖書館資訊系統即時進行資料的互通。雲端圖書館服務系統具有高度共享的資料模型、透過軟體及服務的方式、跨資料格式與媒體格式的一致性的工作流程、彈性運用的詮釋資料管理與開放性的 API 可供跨系統之間的互通與服務。

整合型圖書館自動化系統與雲端圖書館自動化系統在有許多不同，包括系統硬體是自己準備主機或是連線到資料中心，模組的內容差異，系統購買方式是買斷還是租賃，傳統的圖書館自動化系統大多是以 MARC 與 ISO2709 儲存書目資料而雲端圖書館自動化系統則是能容納多種詮釋資料格式，整合性圖書館自動化系統以本館書目資料為核心，而雲端圖書館自動化系統則以聯合目錄的精神運作等不同。而部分的雲端圖書館自動化系統，因含括的模組包括探索系統、電子資源管理系統、智慧型連結系統與知識庫、數位物件保存、決策支援系統等並非圖書館必備功能與模組，因此，圖書館可能在架構與應用面的實際考量下僅購買部分模組，形成部分模組分散運轉的情形。整合式圖書館管理系統與功能模組分散運轉模式的優缺點表列如下：

	整合式圖書館管理系統	功能模組分散運轉模式
優點	<ol style="list-style-type: none"> 1.書目集中，各模組資料可跨模組使用。 2.資料在自己的主機中，沒有資料安全的議題。 3.利用整合式圖書館管理系統將館內資料彙集成大書目庫，再與其他模組與外部網路資訊做資訊加值與應用較為方便。 	<ol style="list-style-type: none"> 1.擁有雲端的優點，館藏資料為參與雲端圖書館自動化系統共同維護，並包括出版商之知識庫，方便抄錄標目。 2.利用軟體及服務的方式取用資料，不用負擔主機硬體的購置與維護費用。
缺點	<ol style="list-style-type: none"> 1.因傳統的書目資料受限於以機讀編目格式，若需整合到圖書館管理系統則需要進行資料映射與轉檔，有時效性的即時性的考量。 2.需額外付當主機硬體的費用與維運成本。 3.抄錄編目需透過 z39.50 或轉檔，無法從現有的知識庫或聯合目錄中抄錄。 	<ol style="list-style-type: none"> 1.若需要進行數據資料分析與資料整合時需透過另一套資訊系統或 API 方式串結。 2.館方有資料安全議題的考量。 3.有資料重複建置與如何即時獲取的缺點，例如，館方已有在整合型圖書館自動化系統轉入電子書，若有購買智慧型連結系統與其知識庫，則需要重複建置。 4.廠商為統一的介面，無法針對單一圖書館進行客製化修改。

圖書館購置資訊系統都有其考量因素，在選擇資訊系統時，往往因為經費的問題與圖書館的功能與任務，並非都要購買完整模組。到底整合性圖書館還是雲端圖書館自動化系統好，完全視圖書館的需求而定。但不可否認的是，如何使用與善用圖書資訊系統，滿足不同類型的資訊資源及圖書館管理與讀者使用資訊的需求，由圖書館經由圖書資訊系統適時地提供適當的資訊服務，讓讀者與圖書館共同創造、分析、組織、選擇、取得、儲存與保存這些資訊資源。

參考書目

黃鴻珠，陳亞寧（2010）。圖書資訊系統的演變與發展，教育資訊與圖書館，48（4），頁 403-428。

二、【擬答】

這題較難，難以從既有的圖書資訊參考文獻的基礎上論述。

全文檢索系統是利用關鍵字進行索引的建立、組織、存取與呈現的一個系統。通常是利用自動化索引的方式進行索引，其技術與資料庫的索引技術不同，資料庫是利用多元樹的方式建立索引結構，而全文檢索系統是利用赫序(Hash)函數的方式，建立硬碟中的索引結構。

資料庫的索引技術主要建立在 BTree 這種資料結構上，是利用多元樹的結構排序方式，讓系統可以在數次的存取後，取得所要找的記錄，當多元樹的分支度夠高時，就可以在 2-5 次的存取當中，取得所要的記錄。而全文檢索的索引方式，則是直接將每個關鍵字出現的地點，全數編入到索引結構中。採用赫序 (Hash) 函數的方式可以減輕作業系統與資料結構的負擔，使得整體的效率提升，也降低了索引程式的複雜性。

因此全文檢索系統若有分區塊 (Zone) 與不分區塊，首先面臨到的問題是如何利用自動化的技術識別區塊，進行區塊內全文索引。於此，我們透過成本面與功能面進行討論。

	分區	不分區
功能	1.對系統建置者來說，需要針對不同資料類型判別區塊的規則，加深系統建置之困難度。 2.對於檢索者可透過檢索欄位進行查詢，ex：只要找尋摘要內的關鍵字。如此一來可提升資訊檢索的求全與求準率。 3.檢索速度較快。但對使用者而言，若想要找全部資料時不一定會比不分區的檢索速度快。	1.因不分區，因此每個區塊同等重要，不用費心判別區塊的規則。 2.對於檢索者，無法透過檢索欄位以及加權方式進行檢索與排序。 3.檢索速度較慢。
成本	對硬體空間要求比不分區還要高，因在進行索引架構時需要紀錄分區（Zone）資訊。	硬體空間的要求較分區少，但因處理不分區資訊文本內容量極大，因此須強化資訊系統的規格（ex：CPU 效能與記憶體）

不管是分區，還是不分區的全文檢索系統，在進行全文自動化索引時，往往會配合系統內部建置詞庫，關鍵詞抽詞系統與知識庫系統可將自動與人工索引整併等。其中的核心技術則是關鍵字自動擷取的機制，透過詞庫比對法、文法剖析法與統計分析法進行關鍵字擷取。新一代資訊檢索系統，尤其是允許全文式查詢的系統，必須能夠運用更具效率的自動化技術，以提供簡易有效的檢索服務。

三、【擬答】

起：新一代的圖書館目錄系統

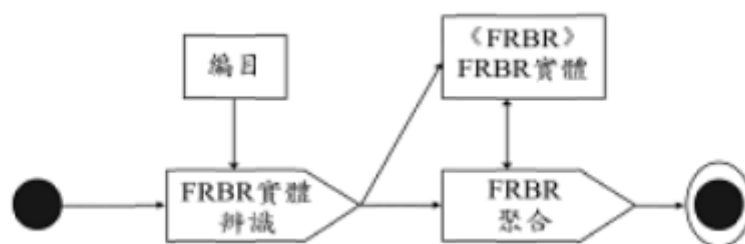
承：何謂 FRBRization

轉：如何引入到檢索系統

合：FRBR 對線上公用目錄的四方面好處

新一代的圖書館公用目錄功能包括整合 web2.0 技術，供多語言主題檢索，行動版的 OPAC，無所不在的 OPAC，優質的操作介面及符合 FRBR 概念架構。其中，自 1991 年 IFLA 著手研究，書目紀錄功能基本需求，目的在利用結構化架構，以連結書目紀錄中的編目資料，滿足使用者需求，達到目錄的尋找、識別、選擇與獲得。其中，在公用目錄或探索系統上呈現 FRBRization 指的就是將書目紀錄分成 FRBR 提出之作品群組，以便符合 FRBR 的需求，這樣的過程被稱做 FRBR 化，即 FRBRization。

如何進行 FRBRization，首先要先經過 FRBR 實體辨識，提取書目紀錄中的實體進入 FRBR 模式中，再將重複的實體聚合起來如下圖所示。



說來簡單，但是實際上在每一個環境都有其眉角。

1.從自動化系統的機讀資料格式抽取資料，進行實體辨識。抽取資料的過程中就面臨到即使是同一個圖書館自動化系

統，內部就可能具有 CMARC 或 MARC21，系統單是在判別機讀編目格式就有些困難，判斷屬於何種編目格式後，尚必須抽取機讀編目格式中的欄號與分欄內容。在這一步驟中碰到的就是書目資料品質、AACR2 與中國編目規則對標點符號的處理方式的差異而必須再進行資料正規化處理等議題。

2.重複的資料進行聚合。進行聚合的過程中就是在建立書目紀錄之 W-E-M-I 的關聯，但往往實務上所碰到的困難點在於書目紀錄在概念上是被放置在載體版本的層次，但套在 FRBRization 中缺乏作品對內容版本層次提出的具體關係。這樣的具體關係，很難以自動化的技術進行自動聚合。例如，哈利波特有多國語言的版本，中文版，英文版，法文版這樣的關係是無法透過系統自動判斷而產生聚合。

3.機讀目錄中常有與書目紀錄無關的資料連結，會產生雜訊。

有些圖書館自動化系統利用網路上的其他工具進行類似 FRBRization，而非透過抽取書目紀錄進行，例如透過 xISBN 進行電子書經由不同出版社發行而有不同 ISBN 的應用加值。

FRBR 對線上公用目錄的影響從書目呈現方面來看，以作品為中心，把作品的內容版本、載體版本和單鍵等資訊統一到作品樹結構的目錄之下，在透過相關資源的聚合反映實體之間的相互關係。在檢索方面，以作品為基礎加以聚合，讓使用者執行一次檢索即可獲取所有相關資源，協助使用者辨識即選擇。一方面提高回收率，也提高檢索效率。在館員方面，減少重複編目，降低編目成本。使用者方面，因 W-E-M-I 檢索結果完整且易於瀏覽，便於使用者選擇資訊。

參考書目

張慧珠（2011），圖書館電子資源組織-從書架到網路。新北市，Airiti Press。

四、【擬答】

起：資訊檢索評估歷史

承：簡介 Cranfield Test I, II 期

轉：簡介 TREC 與比較

合：結論

資訊檢索評估的歷史可追溯至 1950 年左右，早期的評估是在正規化環境下進行，使用量化或質化的方式，企圖衡量不同檢索技術、檢索模式、索引語言之相對績效。1957 年英國 C.W. Cleverdon 主持進行 Cranfield 研究，第一期的 Cranfield 研究根據 Case Western Reserve University 圖書館中 1,100 多篇有關冶金學的文章，以檢全率和檢準率來評估四種索引語言（國際十進分類法（UDC）、主題索引、多面分類法與單一詞）之優劣，結果發現這四種索引語言的表現難分軒輊；也就是說，利用這四種索引語言來組織資訊，所得的檢索效益大體上是一致的。此次研究檢索問題是從文獻產生，而非從資訊需求者。評鑑量標則是文獻和檢索問題之間的邏輯相關。

之後，在 1960 年代中期著手進行 Cranfield Test II 研究，根據第一次的失敗經驗，Cleverdon 將索引語言大致分為三種：單一詞彙語言（single term language）、簡單概念索引語言（simple concept index language）、控制詞彙索引（controlled term index），然後再將各種能提高檢全率和檢準率的方法附加於各索引語言之下（例如切截、同義詞、或類同義詞等），總共產生了三十三種不同的索引方式。第二期 Cranfield 計畫即在比較這三十三種索引方式之檢索效益。研究結果發現『單一詞彙語言』（使用的是後組合系統，post-coordinate system）的總體表現最好，其次是『控制詞彙

索引』，表現最差的則為『簡單概念索引語言』(使用的是前組合系統，pre-coordinate system)。Cleverdon 的研究推翻控制詞彙較自然語言為佳的說法，意謂圖書資訊人員在整理組織資訊上雖投入大量時間和人力，但其檢索效益仍無法超越自動索引，實為人力資源上的極大浪費。

為了促進資訊檢索的研究與應用的發展，美國國防部高等研究計劃局 (Defense Advanced Research Projects Agency, 簡稱 DARPA) 與美國國家標準暨技術局 (National Institute of Standards and Technology, 簡稱 NIST) 於 1992 年共同舉辦了『文件檢索會議』(Text REtrieval Conference, 簡稱 TREC)。TREC 以文獻檢索評估為主，包含二部分：測試資料，有語料 (corpus) 和檢索問題集 (topics) 以及評鑑準則。透過所發展出的大型測試集，制定各種測試項目、測試程序及測量準則、組合成一評估檢索系統的機制。TREC 首創大型測試集，使測試環境得以更接近實際情況，對檢索技術發展與系統績效提升具有相當重要的貢獻。

Cranfield 資訊檢索評鑑與 TREC 的資訊檢索評估模式之比較在於評估方式不同與相關判斷的判斷機制不同。Cranfield 是以文獻與檢索問題的邏輯相關進行資訊檢索評估，前後二期計畫的目的在於找出較佳的索引語言；TREC 的資訊檢索評估模式是第一次各系統可以在相同文獻集合上，使用相同的評估方法，比較各系統的檢索成效。在相關的觀念上，TREC 主要是根據主題的 Nar-rative 欄位進行。對相關與否的判斷原則，是只要文件部分與主題相關即可 (即使只是數句)，並不要求文件的每個部分均與主題相關。

另外有一個很大的差別在於 TREC 可加速研究成果與產業應用的技術轉移與交流速度，提供一個可以呈現資訊檢索最新發展之環境並改善評估技巧。藉此，可以激盪未來資訊檢索的趨勢

以往的資訊檢索評估模式，幾乎都是外文語料，因此成立中文資訊檢索系統評比會議 (CTREC) 或是在 TREC 中持續推廣中文語料，有利於更多研究者進行中文檢索技術研究與跨語檢索。

參考書目

- 1.江玉婷、陳光華 (1999)。TREC 現況及其對資訊檢索研究之影響。《圖書與資訊學刊》，29 期，頁 36-59。
- 2.吳美美 (2001)，中文資訊檢索系統使用研究，臺北市，學生書局。