

資訊檢索技術之核心 The Core Technique for Information Retrieval

陳光華 Kuang-hua Chen

國立臺灣大學圖書資訊學系助理教授

Assistant Professor, Department of Library and Information Science

National Taiwan University

E-mail: khchen@ccms.ntu.edu.tw

【摘要】

資訊檢索研究的目的是在解決人類對於資訊的需求。然而，隨著不同資訊型態的出現，資訊檢索的技術亦逐漸多樣化，以適應各種型態的資訊。本文說明透過元資料進行檢索，可以適用於各種型態的資訊，故可稱之為資訊檢索核心技術。作者並探討三種不同層次的元資料：靜態權威元資料、動態權威元資料、個人化元資料，及其可能的應用方式。

【Abstract】

The purpose of researches for information retrieval is to fulfill the information need. The various techniques for information retrieval are proposed to adapt to the diversified information types. This paper manages to deliberate the importance of metadata and suggests that information retrieval via metadata be applied to various types of information. From this viewpoint, the information retrieval via metadata could be regarded as the core technique. In addition, the author also discusses three levels of metadata: statically authority-controlled metadata, dynamically authority-controlled metadata, and user-oriented metadata.

關鍵詞：資訊檢索；資訊擷取；元資料

Keywords: Information retrieval; Information extraction; Metadata

壹、前言

Borko在1968年提出" Information Science: What Is It?"的重要論述(註1)，曾經提出資訊科學的研究範疇應包括資訊的需求與使用、文件的建立與複製、語言分析、翻譯、摘要分類編碼與索引、系統設計、分析與評估、樣式辨識、適應性系統等九項，並且認為資訊科學有理論的與應用的兩種不同的面向，而圖書館學的研究可以視為應用面向的資訊科學。

雖然國內對於「圖書資訊學」到底是「圖書館學」與「資訊科學」的簡稱，還是真正有所謂的「圖書資訊學」仍然有所爭議，但是自Borko論述的數十年後的今日，即將進入公元二千年，圖書館學研究的典範也已經有所變遷，其與資訊科學之間的關係也與以往不同，同時學科間的交流日益頻繁，跨學門的研究逐漸成為主要的研究模式，電子圖書館（Electronic Library）與數位圖書館（Digital Library）的研究則具體地展現前述的研究趨勢（註2）。圖書館學的研究長期著重於實體圖書館（Physical Library），然而隨著網際網路成為訊息交流的重要管道，透過網路擴大圖書館服務的對象、提升圖書館服務的品質、加強圖書館服務的內涵，應是圖書館學界鞏固既有研究領域之外，亟待努力的方向。有鑑於部分學門的學者專家認為在電子圖書館的架構之下，僅需要全文檢索技術，吾人必須基於實體圖書館研

究的豐富經驗，提出圖書資訊學界的看法，導正前述稍嫌粗糙的觀點，投入這項跨學門、新興、重要的研究。

資訊檢索是電子圖書館研究的一項課題，如何滿足使用者或讀者的資訊需求，是資訊檢索的重要目標。本文並不討論整個電子圖書館的研究，主要著重於在電子圖書館的架構下，資訊檢索應該以何種面貌呈現。一般而言，因應不同的資訊型態，亦有各種不同的檢索技術，然而是否有核心的資訊檢索技術可適用於各種資訊型態，作者企圖對此問題提出圖書資訊學的觀點。本文第二節將討論各種資訊型態及相應的檢索技術；第三節將以資訊組織與整理的觀點，討論圖書館處理資訊的作法，並基於前述的討論，提出所謂核心的資訊檢索技術。第四節則依據第三節的討論，提出靜態的、權威控制的元資料格式（Metadata Format）與動態的、權威控制的元資料格式、個人化元資料格式及其可能的應用方式。第五節則是簡短的結論。

貳、資訊型態與檢索

自有人類歷史以來，資訊就有各種不同的型態，隨著文明不斷地演進，資訊型態亦越見繁複，不同資訊型態的結合在今日亦是司空見慣，而有「多媒體」資訊的術語。若以今日科技的觀點，吾人可以將資訊區分為以下數種型態：

●文字（Text）：包括以 ASCII、

EBCDIC、Latin-I、BIG5、GB、JIS、KS、EUC、UNICODE等不同編碼方式為各種語言文字建立的數位表示系統所呈現的資訊。

- 影像 (Image)：包括以GIF、JPEG、TIFF、BMP等方式建立的靜態影像。通常以解析度 (Resolution) 與色階 (Color Depth) 表示影像的品質。
- 音訊 (Audio)：通常區分為語音 (Speech) 與樂音 (Music)，因為特性不同而採用不同的數位表示方式。例如CD的樂音品質為44.1KHz取樣率，樣本值以16位元表示。音訊常見的格式有MID、WAV、AU、MP2、MP3等編碼方式。
- 視訊 (Video)：可以視為動態影像，亦即以連續播放不同的畫面影像即可構成視訊，例如電影每秒鐘播放24張畫面。常見的格式有MPEG、MOV、FLI等編碼方式。

既然有以上不同的資訊型態，檢索這些資訊的技術亦有所不同。

傳統的資訊檢索領域集中於文字資訊檢索的研究，發展出向量空間模型 (Vector-Space Model)、機率模型 (Probabilistic Model)、布林模型 (Boolean Model)。(註3) 由於是處理文字資訊，自然語言處理 (Natural Language Processing) 的相關技術便廣為資訊檢索研究者採納，例如索引詞彙的選擇、索引典的建置、快速剖析程序 (Parser) 的研製。至於檢索的方式多數是全文檢索，亦有提供欄位檢索的系統，以滿足使用者不同的資訊

需求。影像檢索與視訊檢索的研究通常是由影像處理與電腦視覺研究者所主導。其檢索的方式通常是所謂的內容為基礎的檢索 (Content-Based Retrieval)，例如，使用者的檢索需求為「山丘上白色教堂」的影像圖片。這種技術便與前述「文字資訊」的檢索完全不同，因而無法使用相同的技術。至於以內容為基礎的檢索技術，IBM公司研發QBIC (Query By Image Content) 技術，以顏色、材質、形狀等等方式檢索影像資訊 (註4)；張世富教授亦發展許多影像檢索的技術。(註5)

檢索音訊的使用者，其需求亦與文字資訊的檢索不同。例如，使用者可能只記得一段旋律，卻希望可以檢索原始樂曲；或是透過音高、響度等音訊的物理特性進行檢索。音訊檢索的技術多數是由數位訊號處理 (Digital Signal Processing, 簡稱 DSP) 的研究者所研發的，例如Blum發展的內容為基礎的音訊檢索即是採用 DSP 技術。(註6)

基本上，視訊檢索的本質與影像檢索相似，但是卻更為複雜，因為視訊本來就是動態的影像。使用者對於視訊檢索的需求也類似於影像，可能是某個畫面的特徵，包括色彩、背景、形狀、人數等等。當然，目前有許多單位正進行視訊檢索的研究，如卡內基美崙大學 (Carnegie Mellon University) (註7)，堪薩斯大學 (Kansas University)。(註8)

一份電子文件通常包含數種型態的資訊，吾人可說多媒體資訊是目前資訊的主要型態。例如，一段柯林頓總統與葉爾辛總統的高峰會新聞片段，其中包括美俄二位總統的對話（音訊），影像中會有美俄高峰會相關的文字，以及連續的畫面構成的視訊。這種複雜的資訊是由各種不同型態的資訊所構成，因此要檢索這一類型的資訊必須考慮各種檢索技術，以及其間的互動過程。

有關各種資訊型態的檢索技術，曾元顯教授在1996年「21世紀資訊科學與技術的展望」研討會發表的「多媒體資訊檢索技術之探討」論文，有更詳盡的討論，讀者參見該論文可以獲得進一步的說明。（註9）簡言之，吾人可以得到一個初步的結論，意欲建構適用多種資訊型態的檢索系統，必須由各學門領域的學者與專家研發特用的檢索技術。然而，是否有一種檢索方式是適用所有的資訊型態，如果答案是肯定的，或許吾人可以稱之為資訊檢索的核心技術。圖書資訊學作為資訊組織與整理的核心學門，對於資訊的處理有豐富經驗，且讓吾人從圖書資訊學的觀點審視這個問題。

參、資訊組織與整理

實體圖書館存在的歷史極為久遠，圖書的組織與整理在中國也有長久的歷史，中國第一套實際使用的圖書分類法是由漢朝劉向劉歆父子所創之「七略」（註10），至於目前廣泛使

用的杜威分類法與國會圖書分類法也有很長的歷史。除了分類法之外，圖書館行之有年的機讀編目格式（Machine Readable Cataloging Format，簡稱MARC）記載著圖書的分類、主題、題名、作者、出版者、以及其他的稽核資訊，並且搭配著標題表、索引典、權威檔、與編目規則（如英美編目規則，中國編目規則），透過如此詳密的組織與整理，圖書館的使用者才能夠非常有效地檢索圖書。至於一般人常誤認圖書館僅收藏「圖書」，以致於認為目前所採用的圖書組織與整理的方法無法運用於其他型態的資訊，作者以為這是必須要釐清的觀念。

翻看中國編目規則，吾人可以清楚看到處理的資訊型態，有「圖書」、「連續性出版品」、「善本圖書」、「地圖資料」、「樂譜」、「錄音資料」、「電影片及錄影資料」、「靜畫資料」、「立體資料」、「拓片」、「微縮資料」、「電腦檔」等等。（註11）因此，若是從描述圖書或藏品的觀點而言，圖書館使用的方法足以處理目前所有的資訊型態，若是將這些描述性資訊記錄於機讀編目格式，透過電腦系統的輔助，使用者可以檢索所有經過組織與整理的資訊。以道林（Dowlin）所著的“*The Electronic Library*”一書為例，館員加註的詮釋性資料如圖一所示。前述的分類編目可以分為兩種：一為記述編目；一為主題編目。圖一的書名／作者、出版項、稽核項、叢書名、附註項、ISBN/

價格等屬於記述編目，主要是記載藏品實際的資料，不必經由編目館員進一步的分析。至於標題與索書號內的分類號則屬於主題編目的範圍，編目館員必須分析藏品的內容，經過思考然後加註適當的標題與分類號。一旦館藏皆加註上述的資料，圖書館的讀者或是使用者可以透過編目卡片、OPAC、WEBPAC線上檢索系統，有效地檢索館藏資料。

對於目前受到大眾極為重視的網路資訊，圖書資訊學的作法仍然是極為可行的，網路上也有不少網站是使用圖書館的分類法、標題表、或索引

典組織並整理資訊的。例如，Cyber-Dewey（註12）使用杜威分類法；CyberStacks(SM)（註13）使用國會圖書分類法；CliniWeb（註14）使用美國醫學標題表（MeSH）；INFORMINE（註15）則使用美國國會圖書標題表。此外，著名的網路主題指引Yahoo，其分類架構（如圖二所示）雖然並非遵循任何一種圖書館使用的分類法，卻也聘請具有圖書資訊學專業知識的員工整理並組織網頁。以上種種情形在在說明了，圖書資訊學在網路資訊組織與整理這項工作上所扮演的角色。

書名／作者	The electronic library : the promise and the process/ Kenneth E. Dowlin
主要作者	Dowlin, Kenneth E
出版項	New York, N.Y. : Neal-Schuman Publishers, c1984
稽核項	xi, 199 p. : ill. ; 23 cm
叢書名	Applications in information management and technology series
附註	Includes bibliographical references and index
ISBN／價格	0918212758 (pbk.) : \$24.95
標題	Libraries -- Automation Information technology
索書號	Z678.9 D68 1984

圖一、館藏的加值處理

Main Class	Class	Main Class	Class
Arts & Humanities	Literature, Photography...	Business & Economy	Companies, Finance, Jobs...
Computers & Internet	Internet, WWW, Software, Games...	Education	Universities, K-12, College Entrance...
Entertainment	Cool Links, Movies, Humor, Music ...	Government	Military, Politics, Law, Taxes ...
Health	Medicine, Diseases, Drugs, Fitness...	News & Media	Full Coverage, Newspapers, TV...
Recreation & Sports	Sports, Travel, Autos, Outdoors...	Reference	Libraries, Dictionaries ...
Regional	Countries, Regions...	Science	Biology, Astronomy...
Social Science	Archaeology, Economics...	Society & Culture	People, Environment, Religion...

圖二、Yahoo之分類構 (註16)

如果仔細地審視目前資訊檢索的研究，應該會瞭解一件事實，也就是有各種不同學科背景的學者與專家投入，因此可以視之為跨領域的研究。在電子圖書館／博物館的架構之下，資訊型態是多元化的，資訊檢索的角色亦隨之呈現多元化的發展，在各特定領域皆有其展現與組織的方式。然而，在這些相異的地方是否有共同之

處，或是有所謂的「核心」？經過這幾年學者專家的研究，吾人認為應該有「核心」，否則已建構的異質性電子圖書館／博物館，如何從事最低程度的資料交換，如何進行分散處理、整合檢索，如何建構所謂的"Interoperability"。(註17) 在前述的基本看法之下，無論是任何形式的媒體，其檢索方式也應有共同的核心。因之，若是

從宏觀的角度審視電子圖書館／博物館的研究、建構，必不能忽視各個學術領域對於不同型態資訊的研究與貢獻。長期以來，圖書資訊學的研究就著重於資訊的組織與整理，無論是紙本資訊、非書資料、或是多媒體，因之，自然比其他學術領域更容易掌握資訊的組織與整理。舉一例說明，加州大學聖塔芭芭拉分校（University of California at Santa Barbara）進行的 Alexandria計畫，對於地圖等特殊資料也採用文字形式的描述（註18），Dublin Core雖然不如FGDC專門為地理資訊設計的元資料詳細，僅有的15個欄位，但是也能夠用於描述地圖等形式的資訊。（註19，20）同理，對於檢索技術而言，全文檢索行之有年，有些領域採用這種作法即可，例如WWW上比較粗糙的檢索，有些領域可能需要圖形、色彩的檢索，更有一些領域需要音律、音符的檢索，這些就是所謂的相異之處，需要各領域的努力、協助。然而，是否有共同之處，答案是肯定的。經過這幾年電子圖書館的研究，各國的研究者已逐漸達到一個共識，答案就是透過元資料進行細緻、高價值的檢索，這也是為何ACM、IEEE等各個重要單位目前積極舉辦元資料相關學術會議與研討會的原因。（註21）

肆、元資料

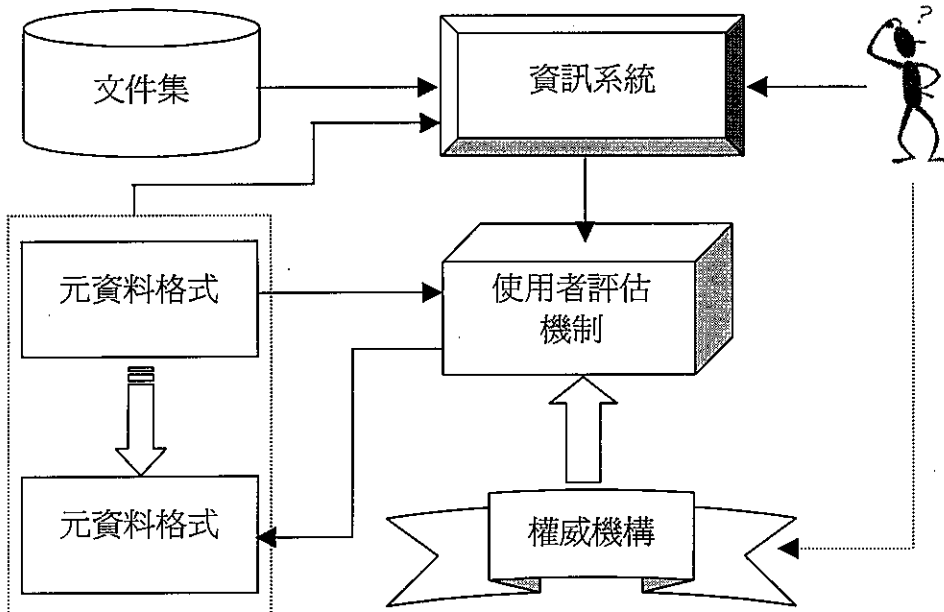
所謂的元資料，指的是用以說明、描述其他資料的資料，目前有眾多的

翻譯方式，作者以元資料稱之，其原因已於另文說明。（註22）元資料的格式一般是由權威機構（Authority）經過使用者研究與藏品性質研究，並遵循典藏政策，從而決定應該使用何種欄位以描述藏品。元資料提供的資料欄位可視為檢索點，使用者透過這些著述資料，便能夠檢索藏品或是資訊。因而，吾人可以說凡是經過組織與整理的資訊，都可以透過元資料提供的檢索點進行檢索，所以資訊檢索的核心技術應由元資料的研究入手。

正由於元資料是由權威機構制訂，元資料的格式通常是靜態的、固定的、而甚少更動的，這種作法有其優點亦有其缺點，茲分別說明如下。對於設定使用群的資源提供者，必定對使用群的資訊尋求行為有一定的瞭解，因而元資料格式具備的欄位應該是這群使用者經常使用的檢索點，所以使用者的資訊需求應經常獲得滿足。但是，即使是穩定的使用群，其資訊的需求會隨時間的遞移而變化，終究會出現使用者所需的資訊無法直接透過元資料欄位檢索的情形。一個可能的解決辦法是，採用動態的元資料格式，欄位會隨著多數使用者需求的變化而改變，圖三可用以說明前述的作法。然而，即使採用動態元資料格式，多數的時候元資料格式仍然是靜態的、權威控制的。對於使用群較廣泛而不確定的系統而言，使用最少的元資料欄位（例如Dublin Core）是比較可行的作法，另外採用資訊擷取（Information

Extraction) 的技術輔助，滿足不同使用者特定的需求。資訊擷取是由文件中擷取使用者事先設定所需的資訊。資訊擷取可視為比資訊檢索更深一層的資訊服務，資訊檢索僅僅送回系統認為相關的文件，至於有關的資訊還需使用者進一步閱讀文件。對於資訊擷取的研究，訊息理解會議 (Message Understanding Conference, 簡稱MUC) 認為不僅僅需要辨識重要的個體 (Entity)，還必須決定個體之間的關係 (Relationship)。(註23) 然而，因為資訊擷取工作的特殊性，所以到底擷

取何種資訊是依資訊服務系統服務的範疇而定。以MUC會議歷年的主題為例，MUC-5會議處理的文件為聯合貿易行為以及微電子產品相關的文件；MUC-6則是有關管理層級變化的新聞報導。(註24) MUC-6會議訂定的工作項目為：辨識專有名詞 (Name Identification)、照應詞解析 (Coreference Resolution)、腳本樣版 (Scenario Template) 等三項。專有名詞的辨識正如字面上的意思，企圖擷取文件中的專有名詞；而照應詞的解析是串連專有名詞及其對應的代名詞；腳本樣版



圖三、動態元資料格式模型

則是依照預先訂定的樣版，由文件中擷取相關的資訊填入樣版的欄位。吾人可以將這三項工作視為是有層級的關係，唯有專有名詞辨識完成，才能夠進行照應詞解析，而後進行腳本樣版的記錄。事實上，前述工作中有兩項（辨識專有名詞、腳本樣版）正如圖書館編目館員進行的分編工作一般，館員首先進行記述編目然後是主題編目，將所得的資料填入元資料格式的欄位（此處指的是MARC），前述的腳本樣版即是吾人所稱的元資料格式，而所謂的資訊擷取就是此處的第三項工作。腳本樣版是欄位名（屬性名）與欄位值（屬性值）的集合，而資訊擷取系統則是針對不同的屬性由文件擷取適當的值填入腳本樣版，也就是填入元資料。

吾人可以從使用者的觀點審視上述的流程。系統可以將使用者的資訊需求以表格化的方式呈現，如同早期IBM發展的Query By Example (QBE)（註25），系統依據表格的要求，從文件集擷取適當的資訊填入適當的欄位。因為使用者的資訊需求若以表格化的方式呈現，也就是欄位名／欄位值配對的集合，所以吾人可將其視為個人化的元資料格式（User-Oriented Metadata Format），此種元資料格式是複雜而多變的，並非權威機構所制訂，卻更能滿足使用者特定的資訊需求，但是，從系統的觀點而言，困難度當然更高。

綜合言之，吾人可以從元資料與

使用者的角度說明資訊檢索核心技術的三個層次。

●靜態權威控制元資料格式

使用者從權威機構制訂的元資料檢索資訊，此種元資料能夠滿足設定使用群的資訊系統之多數使用者的需求。

●動態權威控制元資料格式

權威機構依據系統的使用者查詢記錄，並配合自動化學習機制，修訂元資料格式。動態元資料格式不會面臨老舊、過時的問題，使用群資訊需求變動時，仍然可以得到滿意的服務。

●個人化元資料格式

個別使用者的特定資訊需求，無法由權威元資料滿足時，可以提出個人化的元資料格式，並以表格化的方式呈現，系統依據此表格各欄位名即時由文件集擷取適當資訊填入表格。

伍、結論

網際網路的發展與電子圖書館概念的興起，對於吾人取用資訊的方式有重大的影響。本文思索資訊檢索技術在電子圖書館的架構之下，如何因應不同的資訊型態。作者認為各種型態的資訊有特定的資訊檢索技術，能夠發揮該資訊的特性；但是同時有一種核心的檢索技術適用所有的資訊型態，亦即透過元資料進行資訊檢索。

從目前學術界紛紛從事元資料的相關研究，可以瞭解以嚴謹的態度從

事資訊的組織與整理已經受到大家的重視。檢索經過組織與整理的資訊，其求全率 (Recall) 與求準率 (Precision) (註26) 均能夠大幅提昇，更能滿足使用者的需求。至於所謂的經過組織與整理的資訊，也就是吾人爲其建立適當的元資料，因此透過元資料檢索不僅是核心的檢索技術，也是重要的、高效益的檢索技術。

此外，作者從三個層次討論如何進行元資料的檢索，並且結合資訊擷取技術，說明靜態權威元資料、動態權威元資料、個人化元資料在資訊檢索過程中扮演的角色。如果資訊系統能夠提供上述不同層次的元資料，使用者也能夠有效運用，再搭配各種資訊型態特定的檢索技術，必定可以滿足不同使用者的資訊需求。

註釋

註 1 : H. Borko, "Information Science: What Is It?" American Documentation 19:1 (Jan 1968):3-5.

註 2 : 本文不擬討論電子圖書館、數位圖書館、虛擬圖書館之異同，國內學界對此項議題並沒有一致的看法，不同的論述散見於相關期刊、會議論文集，讀者可以自行參考。作者只能說圖書館學界傾向使用「電子圖書館」，電腦學界傾向使用「數位圖書館」。

註 3 : G. Salton, Automatic Text Pro-

cessing (New York: Addison Wesley, 1989), 313-373.

註 4 : Christos Faloutsos et al., "Efficient and Effective Querying by Image Content," Journal of Intelligent Information Systems 3 (July 1994):231-262. 讀者請參見 IBM 數位圖書館首頁, <http://www.software.ibm.com/is/diglib/>，可以獲得更詳細的說明。

註 5 : S.F. Chang, "Content-Based Indexing and Retrieval of Visual Information," IEEE Signal Processing Magazine 14:4 (July 1997):45-48.

註 6 : Thom Blum et al., "AudioDatabases with Content-Based Retrieval," in Intelligent Multimedia Information Retrieval, ed. Mark Maybury (Menlo Park: AAI Press, 1997), 119.

註 7 : Informedia, 1996, <<http://www.informedia.cs.cmu.edu/>> (13 Nov. 1998).

註 8 : Susan Gauch, Wei Li and John Gauch, "The VISION Digital Video Library," Information Processing & Management 33: 4 (April 1997) : 413-426.

註 9 : 曾元顯，「多媒體資訊檢索技術之探討」，21世紀資訊科學與技術的展望國際學術研討會論文集，世界新聞傳播

- 學院圖書資訊學系及國家圖書館主辦，民國85年11月7-9日，頁281-298。
- 註 10：何光國，圖書資訊組織原理（台北市：三民書局，民國79年），頁131。
- 註 11：中國圖書館學會，中國編目規則（台北市：圖書館學會，民國84年），頁xiv-xix。
- 註 12：CyberDewey, 1995, <http://www.lm.com/~mundie/Cyber_Dewey/CyberDewey.html>(13 Nov. 1998).
- 註 13：CyberStacks(sm), 1996, <<http://www.public.iastate.edu/~CYBERSTACKS/OCLC.htm>> (13 Nov. 1998).
- 註 14：CliniWeb, 1995, <<http://www.ohsu.edu/clinweb/>> (13 Nov. 1998).
- 註 15：INFORMINE, 1994, <<http://lib-www.ucr.edu/>> (13 Nov. 1998).
- 註 16：Yahoo, 1994, <<http://www.yahoo.com>>(13 Nov.1998).
- 註 17：Interoperability主要討論的是異質系統之間相容性，Interoperability有不同的層次，不同的系統若能夠達到Interoperability，則可以共享並交換各種不同層次的資訊，如資訊、程式、功能、服務等等。
- 註 18：Alexandria Digital Library, 1996, <<http://www.alexandria.ucsb.edu/>> (13 Nov. 1998).
- 註 19：Dublin Core Metadata Initiative, 1998, <<http://purl.oclc.org/dc/>> (13 Nov. 1998).
- 註 20：FGDC. "Content Standards for Digital Geospatial Metadata -- FGDC." 1994, <<http://fgdc.er.usgs.gov/>> (13 Nov. 1998).
- 註 21：ACM 與 IEEE 紛紛舉辦元資料相關會議，請參見下列網站。Conference on Digital Libraries: Post-Conference Workshops, 1997, <<http://www.sis.pitt.edu/~diglib97/Workshops.htm>> (13 Nov.1998).
- First IEEE Metadata Conference, 1996, <http://www.llnl.gov/liv_comp/metadata/events/ieee-md.4-96.html> (13Nov. 1998).
- 註 22：陳光華，「資訊的組織與擷取」，臺灣大學圖書館學刊第12期（民國86年12月），頁129。
- 註 23：Message Understanding Conference, 1994, <<http://www.tipster.org/muc.htm>>(13 Nov. 1998).
- 註 24：D. Appelt and D.Israel, Tutorial on Building Information Extraction Systems (Washington, DC, 1997), 4.
- 註 25：R. Elmasri and S. Navathe, Fundamentals of Database Systems (CA: The Benjamin/Cummings Publishing company, Inc., 1994), 249.

註 26：資訊檢索研究經常使用"Recall"與"Precision"兩個重要的術語，臺灣地區使用的中文翻譯並不統一，如Recall有譯為回現率、回收率、召回率等等，而Precision有譯為精確率、準確率等等。然而，無論回現率、回收率等等詞彙都無法表達Recall的意涵。大陸地區使用的

中文翻譯分別為「查全率」與「查準率」，確實反應Recall與Precision的意涵。但是Recall與Precision已經廣泛使用於其他研究領域，如中文斷詞，「查」這個字在這類應用卻無實際的意義，所以本文使用「求全率」與「求準率」表示之。