

主題分析在圖書資訊組織之角色探討

A Study of the Role of Subject Analysis in the Information Organization

陳明來

Ming-lai Chen

屏東縣立新園國中教師

Teacher, Pingtung County

Shin-Yuan Junior High School

E-mail: minglai@pchome.com.tw

【摘要 Abstract】

本文內容主要從主題分析的意義、目的及理論說明主題分析在圖書資訊組織的角色，進而探討主題分析的程序、原則以及與主題分析相關的爭議；最後說明目前圖書資訊界對網路資源主題分析概況，以及主題分析未來趨勢和筆者個人對主題分析之淺見。

The paper discusses the role of subject analysis within the contexts and theories of information organization. The paper also covers the procedure, principles, and disputes on subject analysis. A discussion of contemporary methods about subject analysis of internet resources is followed by some suggestions on this topic.

關鍵詞 Keyword

主題分析 主題編目 主題標目

Subject analysis ; Subject cataloging ; Subject heading



壹、前言

印度圖書館學大師 S. R. Ranganathan 所提出的圖書館學五律中，圖書館界多以第三條——書貴乎用 (Books are for use) 做為圖書館經營最高原則，由此可見圖書館經營最大的宗旨不是建立成千上萬的館藏，而是要儘可能把這些館藏推銷給讀者使用。因此，為了達到「書貴乎用」的目的，長久以來圖書館界學者一直在探討如何改進圖書館服務，方便讀者使用圖書館各項資源。

隨著資訊技術的推陳出新，各種文獻也與日俱增，如何在浩瀚的資料中迅速地獲取所需，成為現代人利用資訊上面臨的一大挑戰。為了解決讀者資訊利用的問題，圖書館必須針對文獻的內容有系統的加以整理、組織，以方便讀者利用各項資源，其中「主題分析」便能在這方面發揮其功效。在網路時代裡，「主題分析」在圖書資訊組織所扮演的角色更顯其重要。本文內容主要從主題分析的意義、目的及理論說明主題分析在圖書資訊組織的角色，進而探討主題分析的程序、原則以及與主題分析相關的爭議；最後說明目前圖書資訊界對網路資源主題分析概況，以及主題分析未來趨勢和筆者個人對主題分析之淺見。

貳、何謂主題分析

一、意義闡釋

所謂主題是一組具有共同性事物的總稱，用以表達文獻內容和研究具體對象和問題，即文獻的中心內容。(註 1)對於主題分析的定義，圖書資訊界學者所給予的界定大同小異，筆者認為 Vickery 所下的定義較為完整，她認為主題分析 (Subject analysis)，就是「資訊的分析」(Analysis of information)，即自文獻的內容中衍生出文字或符號，形成該文獻要義之簡要描述。這種描述可用來

辨識文獻、提供查驗文獻的檢索點、指示文獻的內容，或作為文獻的替代。它主要是根據文獻顯著的特性解析內容，以名詞、或是形容詞與名詞的組合、或是片語或句子、甚至數學等特殊符號，來標示文獻所揭示之重要主題。(註 2)簡言之，主題分析即根據文獻儲存和檢索系統的需要，依據一定的分析方式對文獻內容進行分析，從中提煉出其主題概念、確定主題類型及剖析主題結構的過程。(註 3)

根據以上的定義，一般人大多會認為主題分析即圖書館技術服務作業流程中的圖書資訊之主題編目——予以分類號及主題標目。但實際則不僅限於此，它與資訊的結構及資訊系統有關，包括自然語言關鍵字、控制敘述詞、分類和自動主題索引及主題的階層等；也涉及各種資訊系統，如紙本式及線上目錄、資料庫、全文系統及網際網路，甚至也包括了資訊系統的設計、建構及應用、使用等。(註 4)綜上所述，主題分析實際上大致包含語彙結構 (Vocabulary structuring) 及主題索引 (Subject indexing) 兩大範疇。前者是資訊組織與資訊檢索工具的建構，如分類法、標題表、索引典等，其研究重點為語彙的範圍、語彙的要素與特性、語意關係的建立等；後者是結構化或非結構化語彙的應用，以顯示文獻的主要內容，其研究重點為文獻內容的理解、索引語彙的選用、索引製作方法、索引語彙與文獻內容關係之建立及其主題關連性之剖析等方面。(註 5)

二、目的

根據上述對主題分析的闡釋，可歸納主題分析的目的有以下幾點：(註 6)

- (一)以精確、具體的語彙或符號，對文獻資料提供主題內容之適當描述，促使讀者能夠迅速查檢出所需要的資料。
- (二)揭示出特定主題範圍內研究主題、研究方



法、或觀念及知識應用上的相關性。

- (三)對於各學科領域都能提供一般性和專門性詳簡不同的分析款目。
- (四)提供任何專家、學者或一般人習慣的術語，來檢索他所想要的有關資料。
- (五)將不同國情、不同學科專家，或觀念上之變遷所造成的實際上相同的標目予以聚集。
- (六)利用參照方式將具有關連性的主題做有效的連結，以利讀者蒐集到更完整的資料。

參、主題分析的理論

主題分析所探求的範圍相當廣泛，對其理論的探討，學者亦多從不同的角度加以研究。主觀論 (Subjective idealism) 學者主張作品的主題根據讀者、編目員、作者或出版個人的認知與經驗；客觀論 (Objective idealism) 學者強調概念或一主題係獨立於人類認知；實用觀 (Pragmatic view) 學者則以讀者需求作為考量文獻主題、建立文獻與讀者需求關係的依據。現實/唯物主義 (Realist/Materialist) 學者則綜合各種觀點，主張主題應依據類別及文獻屬性的優先順序來決定。(註 7)

近來有關主題分析理論的探討相當熱烈，其中以本文理解模式 (Text comprehension model) 和多面分析 (Facet analysis) 兩者較為有名。(註 8)

一、本文理解模式

本文理解模式主要的目的是從本文的表層語句，探尋其深層的關連性 (Aboutness)。此模式主要引用認知心理學「上而下的處理」(Top-down processing) 與「下而上的處理」(Bottom-up processing) 之觀念來說明人們解讀真實語言系統的過程。上而下的處理是以概念為導向，而且為演繹式的處理，當閱讀一文獻時，閱讀者會自其語意記憶中檢索出既有的文外知識 (Extra-textual

knowledge)，插入文內閱覽以便迅速瞭解文意；下而上的處理是以資料為導向，而且為歸納式的處理，它對關連性分析做了詳細描述。這種以認知心理學的角度解讀概念形成的過程，可以幫助主題分析人員了解資訊檢索者在資訊需求形成過程中的各項概念，及其如何利用這些概念進行資料的檢索。

為了確切找出關連性，本文理解模式歸納出聚集規則 (Macrorules)，以便將不重要的資訊自記憶中刪除，有效進行聚集分析。這些規則包括有：

- (一)弱刪除 (Weak deletion)：對於文意無太大作用的偶發性資訊予以刪除；
- (二)強刪除 (Strong deletion)：將普通或預期有關的局部性相關資訊予以刪除；
- (三)概括 (Generalization)：將具有相同特質或屬性者統以概括的語彙涵蓋，以減少需要記憶的資訊量；
- (四)建構 (Construction)：結合並統整重要的屬性、原因、成份、結果，而將複雜的文意建構聚集的主題。

二、多面分析原理

多面分析 (Facet analysis) 的觀念首先於 1933 年阮甘納桑的冒號分類法 (Colon classification)。根據他的解釋，多面分析法是區分一複雜主體的不同要素，並將其關連至一組抽象概念的技術。它是一種智力處理過程，能列出形成主題之分類基礎的特徵，同時決定附隨於主體之屬性的處理方式。其目的在於提供一個架構以容納不同類型的語彙，並訂定語彙組合的規則。

此種分析法主要是根據對事物的辨識，因為每一主體都是由許多元素組成，各種元素不同的組合即形成事物的特徵。在運作上將知識體系依概念、用語、註記三大層面加以剖析。在概念層中根據特殊的屬性 (即 Personality、Matter、Energy、Space



和 Time，簡稱 PREST）區分為數個次分面。然後依語意將各次分面的概念轉換為單獨的詞彙，最後在註記層將詞彙轉為單獨的代碼，並予以組合為特定類號。阮甘納桑建構面之邏輯順序的原理，可以引用到字順序的主題目錄與索引，其理論對於現代化的主題索引的製作具有深遠的影響。也因其具有科學理論的基礎，因此被認定為現代資訊儲存與檢索系統，關於主題分析處理上重要的原理。

肆、主題分析的步驟及原則

一、步驟

主題分析在圖書館作業流程中屬於智力密集的工作。一般而言，主題分析的過程包括了文獻判讀、主題概念的提煉、隱含主題概念的分析、文獻主題的取捨和主題類型的確定等五個環節。（註 9、10）

（一）文獻判讀

對文獻的研究是主題分析的第一個步驟，即研讀能呈現文獻主要概念的地方，這也是概念分析首要工作。一般而言，大多來自文獻本身的內容，但有些則是來自文獻附件資料中，如使用手冊、容器或標籤等。從文獻本身內容部分可從以下幾個地方找出文獻主要概念：（註 11）

1. 題名和副題名；
2. 目次；
3. 緒論或前言等相似的文字；
4. 不同字體（鉛字體）之索引詞彙、字句、超連接、摘要或書後索引；
5. 圖例、插圖說明、表格等。

（二）主題概念的提煉

主題分析人員一面進行文獻閱讀，一面進行判斷，並對文獻中關鍵性的語句做準確而精練的概括，此即為主題提煉的過程。但為了確保主題提煉的精確性和一致性，主題分析人員應依照各單位所訂定的主題分析政策來進行文獻主題的提煉。

（三）隱含主題概念的分析

主題分析人員應本著「由此及彼，由表及裡」的原則進行文獻分析，因為隱含主題概念往往不容易被發現，稍不注意就會一晃而過，或者把主題概念會錯意。因此，在進行主題分析時，主題分析人員一定要認真細心，反覆琢磨，切忌走馬看花，囫圇吞棗。例如「無污水排放的養豬場」中便隱含著「廢水處理」的概念。

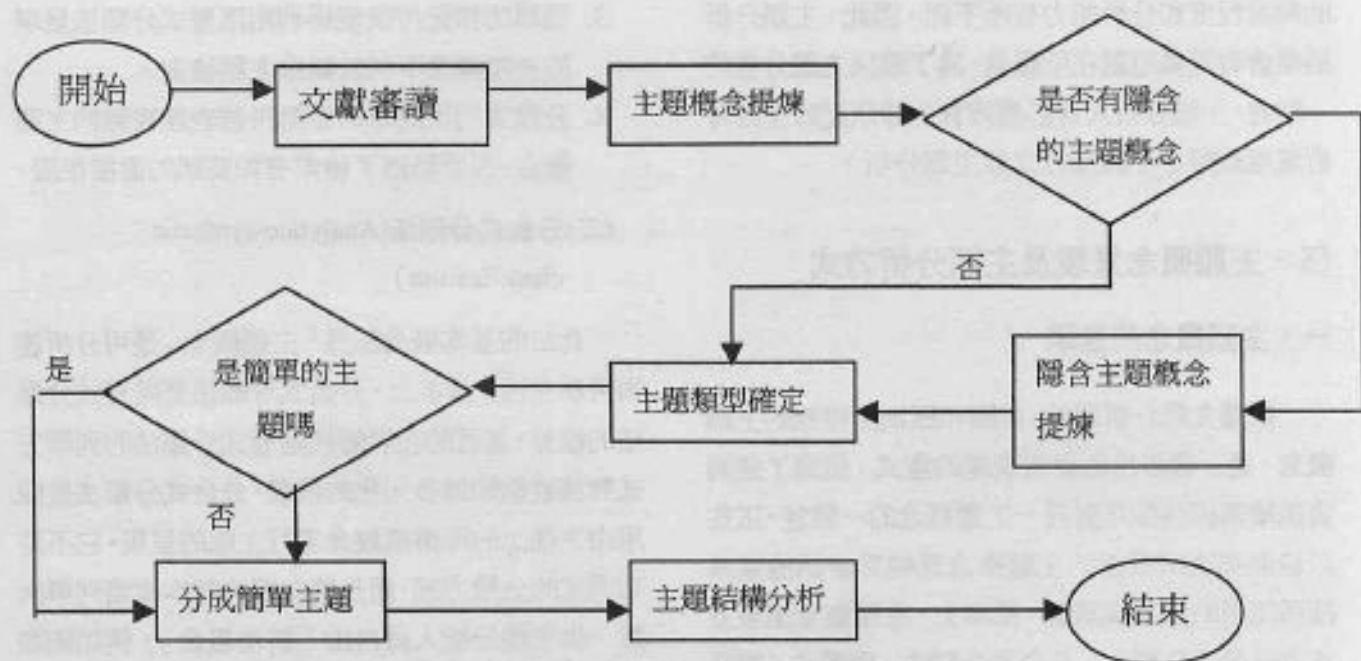
（四）文獻主題的取捨

文獻主題的選擇應按照「去蕪存菁」的原則，「去蕪」即捨棄文獻中參考價值不大、不符合檢索工具或檢索系統要求的內容特徵和不必要的外部特徵；「存菁」即把文獻中詳細論述、有參考價值、可做為檢索對象且符合檢索工具或檢索系統要求的主題內容，按照既定的政策加以確定、提煉。為了提高讀者資訊檢索結果的回現率及準確性，學者近來也多認為文獻主題的取捨，應以使用者的觀點來考量。

（五）主題類型的確定

根據上述主題取捨的方法，將文獻主題解析出來後，再依照主題結構分析的概念，剖析主題結構，如此便完成主題分析工作。從附圖一可看出整個主題分析的過程。





圖一：主題分析流程圖

資料來源：美國資訊科學會台北分會編，索引典理論與實務（臺北市：編者，民國 83 年），頁 3。

二、原則

主題分析人員在進行文獻主題分析時，常因主題概念選取錯誤，而導致主題分析結果的誤差。因此，主題分析人員在進行主題分析時應儘量秉持以下幾點原則：（註 12、13）

（一）需求性

主題分析人員必須針對使用者的需求來分析文獻的主題，既要滿足使用者一般檢索需求，也要能滿足使用者特別檢索需求，以達到既要查準、查全，又要有較快的查詢速度之要求。

（二）客觀性

就是按照文獻原來的面貌及作者的撰寫目的來分析內容，不可摻雜個人的主觀猜測或褒貶之意，以避免或減少主題分析誤差的產生。

（三）全面性

即對文獻論述的各個主題概念提煉達到全面程度，主題分析的全面度越高，提煉出來的主題就越多，用以表達這些主題的主題標目也就越多，使用者檢索結果的回現率也就越高。但要注意的是，主題分析的全面性並不是越高越好，而是要恰如其份。

（四）專指性

從文獻中提煉出來的每個主題概念的外延和內涵，必須與文獻中論述的主題之外延和內涵相符。

（五）一致性

即要求不同主題分析人員或同一人員在不同時間對同一文獻的主題分析達到基本一致，但由於主題分析人員的觀點、專業認知、對主題檢索語言



的熟悉程度和分析能力有所不同，因此，主題分析結果會有差異也就在所難免。為了確保主題分析的一致性，主題分析人員必須確實依據既定的主題分析處理政策之步驟進行文獻主題分析。

伍、主題概念呈現及主題分析方式

一、主題概念的呈現

根據主題分析理論、步驟和原則所得到的主題概念，是一種以自然語言表達的型式，但為了達到資訊檢索與儲存時對同一主題概念的一致性，這些以自然語言所表達的主題概念應轉為資訊檢索系統所使用的代碼或詞彙。基本上，主題概念呈現方式有階層式分類法、分合式分類法、標題法、單元詞描述法、敘述詞描述法及關鍵詞法六種，前二種概稱為分類法；後四種概稱為主題法。以下分別說明之。(註 14、15)

(一)階層式分類法 (Hierarchical classification)

階層式分類法是一種直接表現知識分類層級概念的標識系統。它是依照文獻內容及部份外表特徵而進行邏輯分類和系統排列的一種分類法，即利用傳統的分類表如杜威十進分類法(Dewey Decimal Classification, DDC)或中國圖書分類法，給予文獻分類號。此法運用概念邏輯的原理，根據文獻內容的特徵，進行系統化的組織，因此，以此法所呈現的知識架構，就如同一張「知識地圖」，可以表達主題概念的層次關係。

雖然此種主題概念呈現方式具有邏輯性且層次分明，但亦有幾項問題，包括有：

1. 若檢索者的思考邏輯與階層式思考邏輯不同時，將會影響主題檢索結果。
2. 對於包含多種主題概念的文獻，會因分類標準的差異，而被歸入不同的類目中，而形成同類文獻分散的現象。

3. 簡單的標記符號使得利用階層式分類法呈現的主題概念不利於做為主題檢索。
4. 分類法利用簡單的標記符號表達複雜的主題概念，因而阻絕了檢索者和資訊的直接接觸。

(二)分合式分類法(Analytico-synthetic classification)

此法的基本概念認為「主題概念」是可分析性與可綜合性。基本上，分合式分類法是階層式分類法的延伸，其目的在於解決階層式分類法的列舉方式無法容納知識多元化的發展。分合式分類法是採用由下往上的反傳統觀念進行主題的呈現。它不是列舉式的分類方式，而是將主要的基本術語列舉出來，供主題分析人員自由「拼湊組合」，例如國際十進分類法(Universal Decimal Classification, UDC)和冒號分類法(Colon Classification)均屬此法。因為分合式分類法對主題概念呈現具有相當的彈性，十分適合線上檢索系統。

(三)標題法

標題法是一種以標題詞作為文獻主題概念呈現的主題法。所謂標題詞，是一種經過規範化處理，用來代表文獻主題的詞或詞組。其構成的基本原理可歸納成以下四點：

1. 依照主題集中文獻；
2. 以控制的詞彙直接標示文獻主題概念；
3. 以參照方法間接顯示主題用詞間相互關係；
4. 以字順排列方式直接提供主題檢索。

和之前兩種分類法相較，標題法較特別的特性是對於同一文獻可同時給多個標題，表達主題概念時較分類法完整；將文獻主題概念文字化，較能直接表達主題。不過，標題法和分類法一樣仍會產生文獻集中和分散的問題—即主題標目以字順排列，從學科角度來看便是分散了文獻資料。另外還有標題表詞彙新穎度、詞量、專指度等問題



(四)單元詞描述法 (Unit term)

單元詞描述法是以單元詞為文獻主題標識，通過字面組合方式表達文獻主題的主題法。所謂單元詞是指最小、最基本的詞彙單位，且能夠描述文獻所涉及的主題事物，同時這些單元詞在概念上不能再進一步分解。因此，單元詞是構成標題的要件，但其本身並非具體的標題詞，須由若干個單元詞相互組合，才能構成一個專指款目，以精確表達文獻主題的內容。此外，單元詞法缺乏完善的參照系統，不利於讀者進行主題檢索，因此也才有敘述詞法的出現。

(五)敘述詞描述法 (Descriptors)

敘述詞描述法是以敘述詞作為文獻主題標識，藉由概念組合方式表達文獻主題的主題法。所謂敘述詞是經過規範化處理，以基本概念為基礎的表達文獻主題的詞或詞組。和標題法相較，敘述詞描述法有幾點不同之處：

1. 此種主題法多用來編制索引典，多半用於期刊文獻的主題概念分析；
2. 不像標題表中的有多種呈現主題的詞彙結構，敘述詞少有複分或倒裝等詞彙形式；
3. 敘述詞大都有範圍註，其架構較標題表嚴謹；
4. 敘述詞可用各種方式排列，較標題表容易呈

現出詞彙概念間的關係。

(六)關鍵詞描述法

關鍵詞描述法是直接引用文獻所使用的主題詞彙，用以代表文獻主題概念，並作為檢索用詞，包括題名、作者、乃至全文等出現的原始詞彙。此種主題法出現在美國 1950 年代中期，是隨著電腦檢索的應用，因應主題索引自動編制的需要產生的主題分析法。這種方法不需要詞表，一般只將不適合作為檢索的非關鍵詞編為禁用詞，以便電腦編制主題索引時，將其排出。其最大的優點就是主題索引編製成本低、速度快而且詞彙專指度高。但和上述之控制詞彙之主題法相較，關鍵詞描述法也會產生一些問題，包括：

1. 關鍵詞描述法使用的關鍵詞屬於自然語言，因此並沒有對所用詞彙建立參照及層次附屬關係，因此，同一主題款目的文獻，常會因作者用詞的不同而分散；及增加主題檢索的困難度。
2. 電腦無法判別詞彙的語意及語法結構。
3. 有些電腦自動索引產生的關鍵詞是不具檢索意義，浪費電腦空間。

從下表可看出上述四種主題法的異同：(註 16)

主題法	主題詞	組配方式	組配原理	詞彙關係
標題法	標題詞	前組合	字面組配	等同、層級、互見
單元詞描述法	單元詞	後組合	字面組配	等同關係
敘述詞描述法	敘述詞	後組合	概念組配	等同、層級、聯想
關鍵詞描述法	關鍵詞			

二、主題分析方式

主題分析方式可從不同的角度加以區分：(註 17)

(一)按分析單元不同可分為宏觀分析和微觀分析

宏觀分析是根據需要，以某一主題內容相同或基本相同的一份或者一組文獻作為索引單元，進行



組合級的主題分析；微觀分析是以單份文獻作為索引單元，進行文件級的主題分析。

(二)按分析單元所概括範圍可分為整體分析、描述分析、針對性分析

整體分析是只指出文獻材料中的整體主題；描述式分析則不只指出文獻材料中的整體主題，還指出若干相關或局部主題；針對性分析是從某類型圖書館、專業需求出發，從不同角度或偏重某一方面地對文獻材料進行細緻深刻的主題分析，以提煉概括出適合某類型圖書館、專業需求的主題。

(三)按主題結構所概括的多少可分為概要分析 (Summarization) 和深入分析 (Depth of indexing)

概要分析是指在分析時以文獻整體為分析之單元，因此概要分析對主題較寬的百科全書或主題較窄的期刊論文而言並無差別，亦即無論主題寬窄，都是在對整份資料有全盤了解之後再給予適當的主題描述，資料範圍的寬窄與描述語並無直接關

係；而深入分析是指所分析的是一資料中之各部分內容，而不是以整體為分析單位，也因分析之深淺程度不一，所以在程度上深入分析並無明確的定義，完全視目的與實際狀況而定。

陸、主題分析相關議題的爭論

一、控制詞彙與非控制詞彙之爭

主題分析結果的良窳關乎主題檢索結果的品質，而主題分析時用以代表文獻主題概念的主題詞到底應該採用控制詞彙或非控制詞彙的方式，哪一種模式對讀者進行主題檢索時產生較好效能，至今仍是沒有定論，可謂是「公說公有理，婆說婆有理」。所謂「控制詞彙」即主題分析所使用的主題詞來自事先規範處理之索引工具，如標題表或索引典等；而「非控制詞彙」即所謂的「自然語言」，主題分析所引用的主題詞來自文獻本身內容，如題名、摘要、目次等。從下表中可瞭解主題索引採用控制詞彙及自然語言的優缺點。(註 18、19)

	控 制 詞 彙	自 然 語 言
優 點	<ol style="list-style-type: none"> 1.主題用語明確、清楚固定，使得索引用詞與檢索用詞一致，避免資料分散與不易查尋的問題 2.解決很多語意上的問題（如同義詞、同形異義字） 3.表達主題之階層附屬關係（如上層主題、下層主題、相關主題） 4.主題詞選用得當，可提高主題檢索之回現率及精確率 5.為一種知識地圖的呈現 	<ol style="list-style-type: none"> 1.主題索引作業快，成本較低 2.讀者進行主題檢索時較為簡單 3.容易產生新的及複雜概念 4.減少主題分析人員產生人為誤差 5.主題分析人員的學識背景及經驗較不重要，因為主題詞來自文獻本身 6.不必應付控制詞彙工具易於老化帶來的問題 7.字義表達較為自由



缺點	<ol style="list-style-type: none"> 1. 成本高 2. 使用工具無法隨時增添新的或複雜概念，更新較困難； 3. 用詞受限於控制詞彙工具，不夠彈性 4. 控制詞彙工具包含的用詞未必詳盡 5. 主題分析人員必受過相當的訓練，製作較慢 6. 所使用的標準詞彙易於老化過時，必須隨時更新 7. 會有主題分析人為錯誤，產生不一致的問題 8. 控制詞彙用詞常無法與讀者的用語、觀念相容，影響主題檢索的結果。 	<ol style="list-style-type: none"> 1. 無法呈現主題概念層次附屬關係，影響讀者主題檢索的結果 2. 讀者需對主題概念要有相當的認知，對讀者檢索資料時負擔較重 3. 因主題來自文獻本身，所以可能會遺漏未呈現在文字中但隱含在文內的主題 4. 主題檢索結果回現率較低
----	--	--

雖然圖書館界對於主題用詞是否採用控制詞彙或非控制詞彙的方式沒有定論，但一般而言，兩種方式並非是完全互斥的，學者之間有一個共識，即控制詞彙與自然語言這兩種模式，應該是可以互相截長補短。

二、前組合方式與後組合方式之爭

前組合方式是指主題分析人員選擇主題用詞時，如主題包含兩個以上的概念，須先將其作好適當之組合。此種方式使得讀者可對一個主題各方面主題進行瀏覽，但只能按已建立的標題的次序和預先組配好的標題來檢索，無法根據檢索者要求調整，也不能根據檢索需要靈活縮小或改變檢索範圍，傳統的分類法和主題標目法均屬此方式。而後組合方式是指讀者檢索前，主題索引中的主題詞是獨立的，讀者進行主題檢索後，才會根據檢索需要進行組配。這種方式可根據檢索需要，用檢索主題中的任一個詞進行檢索，也就是可以靈活擴大或縮小組配的範圍，對檢索內容進行調整。(註 20)

從上述得知，後組合方式較前組合方式有較大

的彈性與自由，因為前者可利用各種檢索技術如布林邏輯運算 (Boolean logic operators)，調配檢索字彙間的關係，從任何角度或觀點查尋資料。此種方式是「綜合」的，而非「列舉」的；是「多元」的，而非「直線」的。(註 21)但對於未自動化作業的圖書館而言，採用前組合方式則是主題分析最佳模式，因其在卡片主題目錄便於管理。而因應電腦科技的應用，後組合方式則有成為主流之趨勢。

三、主題分析的品質之爭

談到主題分析的品質就要先提到主題分析之詳盡性 (Exhaustivity) 和明確性 (Specificity)。詳盡性是指以足夠的詞彙來描述一文獻之主題，描述詞越多越詳盡，那麼文獻被檢索到的機率就愈大。理論上主題分析越詳盡，檢索者主題檢索結果的回現率 (Recall) 也就相對提高，但實際上並不是沒有限制數量，至於主題分析人員應給予多少主題詞，則應視各館之政策。而明確性則是主題分析過程中最高指導原則，早在 1876 年時，Cutter 即主張要以明確的詞彙描述文獻主題，因此，主題用詞



明確性愈高，讀者檢索所得的文獻相關性也就愈高。

一般人會認為，詳盡性與明確度呈反比的，也就是說主題分析愈詳盡，雖可增加檢索結果的回現率，但可能會犧牲精確度；而主題敘述詞愈明確，雖可增加檢索所得之文獻相關性，但會犧牲回現率。但實際上，檢索之回現率與精確度除受主題分析程度影響外，尚須其他因素的配合，如檢索系統設計是否具有彈性。所以，主題分析之詳盡性與明確度並非完全互斥的，若在檢索系統上加上若干的設計如加權處理 (Weighting)，便能使主題分析兼具詳盡性與明確度。

一致性是主題分析品質最常討論的議題，因為主題分析是否一致性關乎檢索者之效率與檢索結果。從本文理解模式可瞭解到主題分析過程涉及相當複雜的認知歷程，因此影響主題分析的因素非常多，諸如分析者的背景、採用的索引語彙系統、語彙量的大小及精確度、賦予的語彙量、文獻內容的多寡、索引者使用的工具等。而 Lancaster 歸納影響主題分析結果的因素主要可分為五大類：主題分析人員因素 (Indexer factor)、詞彙因素 (Vocabulary factor)、文獻因素 (Document factor)、過程因素 (Process factor) 和環境因素 (Environment factors)。(註 22)以上諸多的因素，均會影響到主題分析的一致性。而根據 C. M. Chu 的研究結果，影響主題分析人員表現的因素亦有五種：文獻的學科領域、文獻內容的事實性 (Factual) 或主觀性

(Subjective)、文獻主題的複雜度、文獻是否有參考書目、文獻內容的明確度 (Clarity)，其中也發現文獻最顯著主題之辨識能力較強，而對於次要主題之辨識力通常較弱；在學科領域方面，科學及社會科學的主題較易辨識，人文方面的主題則較困難。(註 23)因此，在探討主題分析一致性議題時，應先就相關因素加以考量。

柒、網路資源之主題分析

圖書館界學者很早便會大膽地推論未來的社會將是無紙的社會 (Paperless society)，雖然這個大膽的預測未必會真正實現，但可以確定的是，網路電子資源將是未來圖書館的主流館藏。在這浩瀚的網路資源裡，讀者要如何迅速地找到所需的資訊，則要靠圖書館對網路資源的資訊組織及主題分析，因此主題分析工作在網路時代更顯現出其重要性。

目前圖書資訊界對網路資源的整理多以主題概念方式加分類整理，其主要方式可分為下列三種：

一、主題指引 (Subject guides)

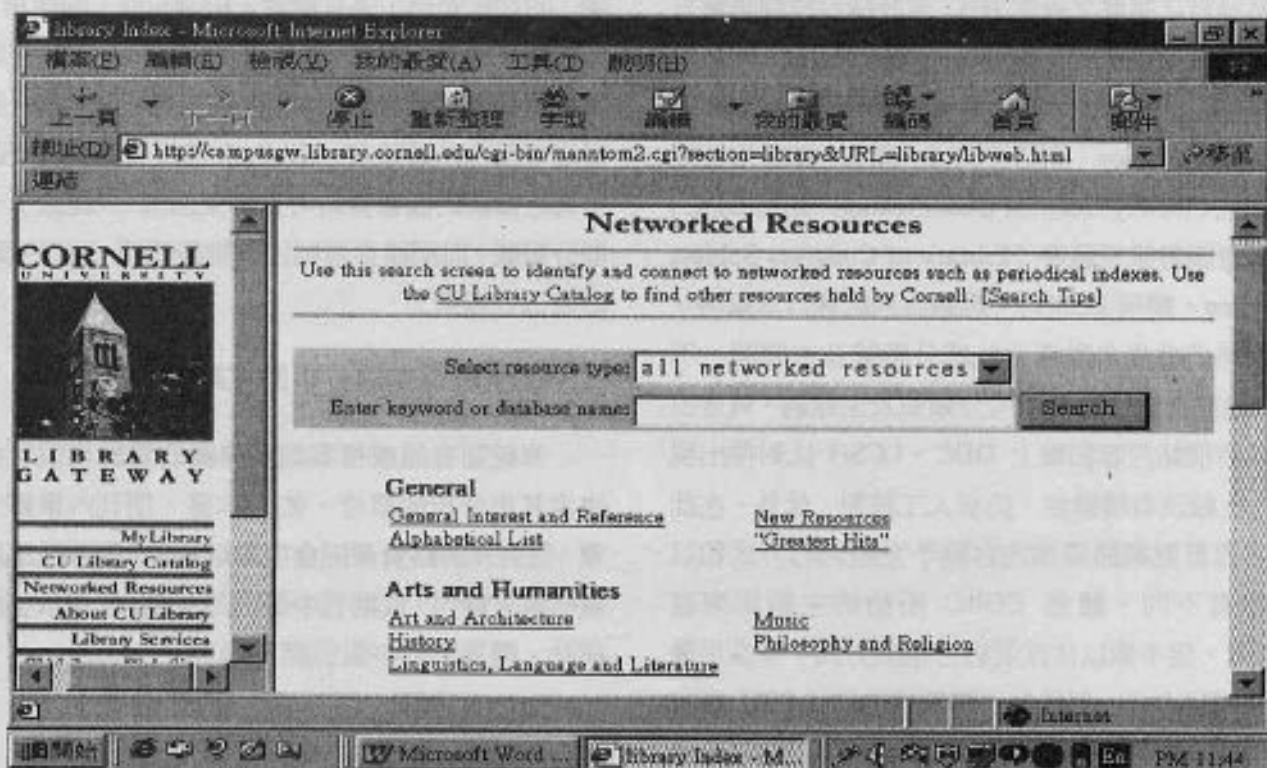
即網路上常使用的搜尋引擎的分類體系，如 Yahoo (見圖二)。在這種模式下，同一網路資源在不同版本裡，根據使用者需求，而分在不同的主題類目之下。



圖二：雅虎中文站

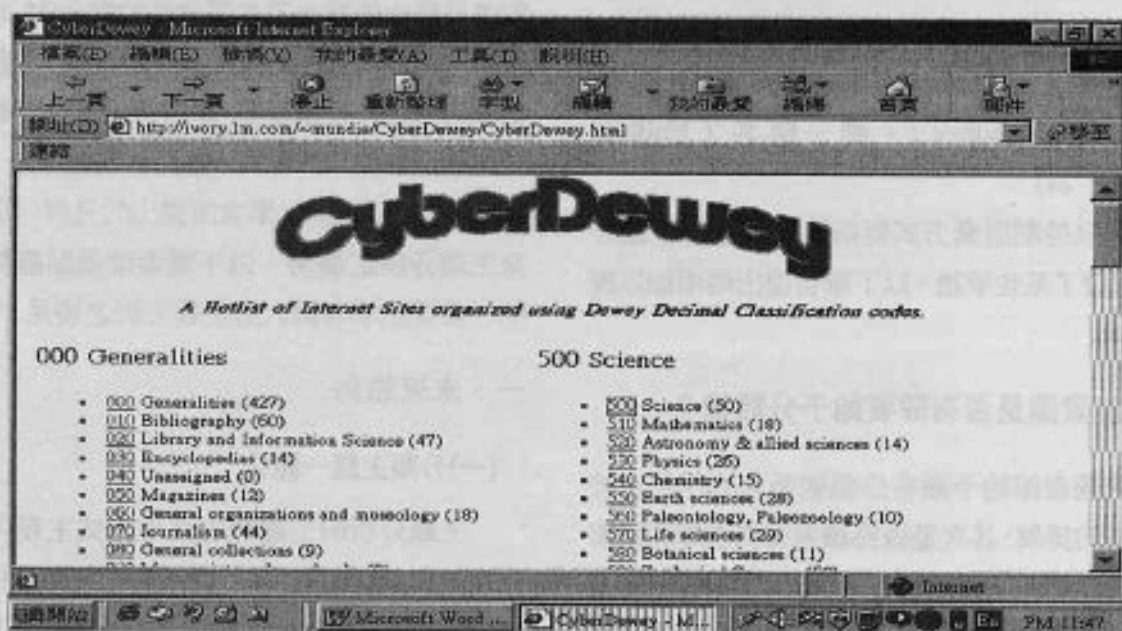


二、根據圖書館所蒐集之網路資源，自訂類別整理。(見圖三)



圖三：康乃爾大學網路資源之整理

三、根據圖書館既有的分類體系，如 OCLC's Netfist 便是根據 DDC 來整理。(見圖四)



圖四：根據 DDC 整理網路資源



以上三種主題概念所整理之網路資源均提供自然語言之關鍵字檢索方式，至於採用控制語彙方式對網路資源進行主題分析，則有 OCLC 大力推行的線上資源合作編目 (Cooperation Online Resource Catalog, 簡稱 CORC) 專案，即採用杜威分類法 (Dewey Decimal Classification, 簡稱 DDC) 及國會圖書館標題表 (Library of Congress Subject Heading, 簡稱 LCSH)。透過該系統進行草編時，編目系統會先自動產生杜威分類號及主題詞，不過，這些系統自動產生的分類號及主題詞，只是由系統將網站內容和線上 DDC、LCSH 比對後出現的，比較沒有精確性，仍須人工核對。此外，在此專案裡針對網路資源內容給予主題詞的方式和以往略有不同，雖然 CORC 所給的主題詞來自 LCSH，但不像以往採前組合組配方式，改採用後組合組配方式；另外也不再採用傳統 LCSH 原有的年代、地理、形式等複分標題，並將這些複分標題獨立成爲主標題詞，即所謂的「多面應用主題術語」(Faceted Application of Subject Terminology, 簡稱 FAST)。而這裡的 Facets 則可分爲主體 (Topical)、地理 (Geographic)、型式 (Form)、時間 (Period)、人名 (Personal names)、團體名稱 (Corporate names)、會議名稱 (Conference/Meetings)、劃一題名 (Uniform titles)。(註 24)

不過以控制語彙方式對網路資源進行主題分析，也引發了某些爭論，以下筆者提出兩項加以說明。

一、網路資源是否有需要給予分類號？

傳統圖書館給予圖書分類號最大的作用在於圖書資料的排架，其次是做爲讀者資料查尋的檢索點。但網路資源並沒有排架的問題，讀者只要知道其 URL 便可取得該資訊內容。因此，學界便常在爭論網路資源是否也要給分類號？筆者個人的看

法認爲應該要給分類號，因爲分類號並不只有「排架」的功能而已，還有瀏覽 (Browsing) 的作用，過去讀者找不到自己所預定要找的書時，可到書庫鄰近書架瀏覽找相關的資料；在自動化時代裡，讀者亦可藉由分類號在 OPAC 上進行瀏覽。只不過差異在傳統的圖書資料考量排架因素，只能給予一個分類號；而網路資源則沒有排架問題，可以同時給幾個分類號。

二、網路資源主題分析的「單位」定義何在？

傳統圖書館或檔案館所典藏的資料均很容易決定其可分析的單位，如一本書、期刊內單篇文章，但對於網路資源則會很難以決定，例如是應以整個電子期刊、或期刊中個別的主題；是要以整個網站、還是網站中個別網頁呢？

捌、主題分析之未來趨勢及建議

圖書館運用主題分析的觀念於圖書資訊組織作業的歷史可說是淵源甚遠，早在西方的亞歷山大圖書館目錄及中國漢朝劉氏父子所編撰《七略》之圖書目錄中便隱含了主題分析的概念(註 25)，而隨著資訊的爆增，爲便利讀者檢索資料，主題分析所扮演的角色亦日趨重要，學者對主題分析的理論及相關議題亦熱烈的探討。雖然對某些議題，學者的看法不盡相同，但從學者所提出的見解可窺探出未來主題分析之趨勢。以下筆者就幾點趨勢加以說明，並提出筆者個人對主題分析之淺見。

一、未來趨勢

(一)分類主題一體化

主題分析的主題概念呈現方式主要可區分爲分類法和主題法兩大類，但這兩種主題分析方式之角度不同，功能有所差異，具有相互補充的作用。分類法的優點正是主題法的缺點，主題法的長處正



是分類法的不足。若能將兩者結合，則可以截長補短，相輔相成，既方便多元檢索，又有利於簡化索引作業。大陸檢索專家張琪玉認為，理想的檢索語言應是「學科事物」型檢索語言，由學科分類系統和事物分類系統面構成，兩系統介面可以互相組配——分；當按事物聚類時，則藉助於學科及其問題面進行複分。(註 26)因此分類主題一體化將是主題分析未來的一個新趨勢，也基於分類主題一體化的優點，近年來分類主題一體化成大陸地區圖書資訊界熱門的研究主題，根據統計，目前大陸地區編輯出版的一體化詞表多達二十餘部，最著名則為 1994 年出版的《中國分類主題詞表》；而台灣地區則有何世文曾在其碩士論文中嚐試以教育類為例進行分類主題一體化之研究。

(二)自動索引編製之研究

文獻主題分析是智力導向的工作，因此相當花費人力及時間。但由於近十年來電腦科技的進步，圖書館界便想利用電腦運算能力及人工智慧的思想設計，協助圖書館進行文獻的主題分析工作。因此，近年來學者研究重點便在於如何改善衍生索引 (Derived indexing，即自原文中載取索引詞) 及自動索引的品質上，以減少控制詞彙索引之必要。(註 27)

利用電腦進行主題分析早期主要應用於文內關鍵字索引 (Key-Word-In-Context，簡稱 KWIC index) 與文外關鍵字索引 (Key-Word-out-of-Context，簡稱 KWOC index) 的製作。此種交替索引 (Permuted index)，即以題名為基礎，用電腦與外在的字典做比對，把不重要的字予以剔除，然後將其餘字交替做成索引款目。(註 28)基本上這屬於摘錄式的索引，分析的內容僅限於題名中的文字，未必能明確地表達文獻的主題。現階段利用電腦進行自動主題分析的技術主要可分兩種——元素分析 (Elemental analysis) 和結構分析 (Structural

analysis)。前者包括有使用停用字表、字根法、詞彙控制、用語分析、計算出現次數及相對頻率、詞彙分析、語法分析等；後者即語意分析，將文獻轉成結構性表達 (Structural representation)，以推論文獻之內涵。(註 29)總而言之，目前電腦進行之自動主題分析及索引之技術有的尚處於實驗階段，亦無法完全模擬人工所進行的主題分析方式，因此有關自動分析技術仍有待深入的研究。

(三)主題分析之工具朝向後組合式組配檢索詞方向發展

文獻主題概念最早呈現工具即是標題表，當初考慮到卡片目錄型式檢索的需求及管理的便利性，標題表原始設計採用前組合式組配檢索詞，先把單純概念詞彙間的複雜關係予以組配好，隨著電腦運算及檢索技術的進步，這些原先的考慮均可交由電腦執行；再者從各種檢索語言的發展歷程，也可看出後組合之組配方式已成為趨勢，如曾在英國大力推行的 PRECIS 系統便是以單元詞為基本單位，再組合出語句關係。近來，LCSH 修訂方向也朝著縮短標題詞的長度，一方面取消原先的年代、地理、形式等複分標題，並將這些複分標題詞獨立成為主標題詞；另一方面將複雜主題字串，分離成單純概念的獨立標題，再藉由後組合運算，並以定長欄限制檢索範圍。(註 30)目前國家圖書館亦正著手進行《中文圖書標題表》之修訂，筆者建議可參考 LCSH 模式來進行。

(四)如何結合自然語言與控制語彙的特性為主題分析研究的重點

雖然利用電腦進行自動主題分析可節省圖書館不少的人力，也有不少學者信誓旦旦宣稱已不再需要控制詞彙的主題檢索系統，但若真的完全仰賴自然語言的自動索引系統，現階段仍有技術上無法克服的問題。因此，對於要求高品質的檢索系統而言，詞彙控制仍然重要，自然語言的關鍵詞檢索可



以作為輔助，結合兩者之優點才能有高品質的檢索結果。換言之，主題分析研究的重點應朝向如何結合自然語言與控制語彙的特性。

(五)利用既有主題分析之工具或方式，著手網路資源之主題分析

在利用電腦自動主題分析技術尚未完全成熟之前，大部份學者仍認為未來圖書館所使用的主題分析方法不會有革命性的改變，換言之，圖書館仍會利用控制詞彙的工具如標題表進行文獻主題分析。基於方便讀者檢索及檢索結果全面性，學者亦認為利用既有的標準工具進行網路資源之主題分析，目前國外已有些研究機構利用圖書館傳統主題分析工具之分類表與標題表進行網路資源之主題分析，如 OCLC 之線上資源合作編目 (CORC) 專案即利用杜威分類法 (DDC) 及國會圖書館標題表 (LCSH)。如此，讀者便可透過檢索系統一次查尋到各類的資源。

二、建議

(一)在檢索系統上增加主題詞線上瀏覽功能，並配合檢索技術的改進

根據調查研究結果顯示，讀者利用主題檢索途徑檢索資料的比例有日愈增加之趨勢(註 31)，但主題檢索結果則是往往令讀者洩氣的，因為其結果不是太多就是太少。根據亨特爾對美國北卡羅萊納州大學 OPAC 使用研究，發現有 62% 的主題檢索結果為「零符合」，其原因即是讀者的檢索用語大多無法與主題索引詞彙配合(註 32)，換言之，讀者根本不清楚主題索引詞彙間的關係及內容。而國內曾繁娟對中文標題檢索效益的研究，也發現有 61.9% 檢索詞彙不符合中文標題款目，分析結果亦指出讀者對檢索系統所使用的主題詞彙不清楚為中文標題檢索不滿意及困難度的原因之一。(註 33)因此，為了使主題分析所使用的詞彙能和檢索用語一

致，圖書館應在檢索系統增加主題詞彙的線上瀏覽功能，但標題詞彙排列方式應採用索引典的設計，以顯示詞彙之間的關係。除了提供架構完整、使用說明詳細的線上索引瀏覽工具外，檢索系統應提供相關回饋等技術，及其他檢索技術如布林邏輯運算、切截法或近似運算等功能，協助檢索者進一步擴大或縮小主題檢索結果。

(二)加強主題權威控制 (Subject authority control)

主題分析所使用的詞彙到底要採用控制詞彙還是非控制詞彙一直是學者最大的爭議點，不過筆者認為不管採取哪一種方式，圖書館在進行主題分析的同時，亦要著手主題權威控制的工作，因為這是確保主題索引詞彙與檢索詞彙一致的方法之一，且其效率比修訂主題分析之標準工具來得快。不過主題權威檔的建立與維護需花圖書館相當多的時間與人力，因此，筆者建議圖書館應合作建立主題權威檔，並設置於全國書目中心 (NBInet)，以達資源共建共享。

(三)編訂主題分析的標準工具

目前台灣地區圖書館界已於民國 82 年編訂了主題分析的工具《中文圖書標題表》，並於民國 84 年修訂，但其為圖書館界最大的垢病即其詞量少於分類表之類目，影響標題表的實際應用。因此，對於國內的中文圖書標題表的修訂，應就標題詞的詞量、詞形、詞類、詞義、詞間關係與標題詞的專指度加以重新檢視。在擴大詞彙收錄的方式可經由直接從圖書期刊等文獻資料中選詞，亦可參考國內外相關的主題詞表、標題表與分類表、學科專業字詞典，以擴增詞彙收錄之數量。另外配合詞彙組配方式的趨勢，可仿 LCSH 的方式將複合概念標題詞詞組分離成單純概念的獨立標題；至於標題詞間的關係亦可仿索引典方式，建立嚴密的階層參照架構。另外，圖書館亦可透過讀者主題檢索的研究分析，藉由電腦對檢索所使用之自然語言詞彙的統



計，在超過某些次數後，可將這些詞彙列為控制詞彙詞庫。

(四)各館應制訂主題分析之處理政策

影響主題分析一致性的因素有人員因素、處理程序因素、文獻相關因素、環境因素，其中最主要的因素則是人為因素，因為其牽涉到主題分析人員的學科知識、專業能力、使用者需求知識、認知的不同及其經歷等，但實際上要百分之百確保主題分析的一致性是不可能的，為了能在異質環境中取得最大公因數，圖書館應根據文獻收藏範圍、使用者需求、檢索工具或系統設計等因素制訂主題分析處理政策，做為主題分析人員進行主題分析的規範，提高文獻主題分析的全面性、專指性和一致性。

(五)加強漢語研究

國內有關主題分析的理論、方法之研究大多引自國外學者的研究，但西文與中文的特性有諸多的不同之處如西文有單複數而中文則無；西文字之同義字甚多，但中文則多為同義詞；西文字往往一字代表一特定概念，但中文字則必須以複詞才能表達完整的概念。因此，適用於西文環境的主題分析方式未必能套用在中文環境。尤其近幾年檢索系統隨著資訊科技的進步與應用，減少主題分析之人力負擔，圖書館界近年來開始研究如何在中文環境下由電腦自動編制索引，其首先要加強研究及克服的課

題便是中文語意 (Semantics) 與語法 (Syntax) 的問題研究。

(六)加強讀者檢索能力

根據曾繁娟的研究指出，檢索者對中文標題檢索感到困難的原因主要是不懂得主題檢索的相關技巧，如不知如何進行主題分析、不知如何進一步檢索等。(註 34)因此，筆者建議圖書館可在圖書館利用教育相關課程中加強各種資訊檢索系統之檢索技巧及資料主題分析的能力。

玖、結語

主題檢索在網路電子時代已成為主要的資訊搜尋方式，然而唯有正確的主題分析，才有高品質的主題檢索結果。因此，主題分析與主題檢索有密切的關係，提高主題分析的精確度與詳盡度，使用加權等技術，加強主題分析的一致性等均是有效的改善途徑。不過因主題分析涉及的相關因素太多，相關問題仍須解決，始能獲得更大的效益。圖書資訊人常以能提供最便捷的服務給讀者自我期許，而圖書資訊的主題分析便是方法之一，也唯有從各種技術服務理論實務的研究，做為圖書館各種讀者服務堅實的後盾，始能期許我們達到更臻善的服務願景。

(收稿日期：2000年10月30日)

註釋：

註 1：姚杰，「文獻主題標引中的主題分析」，圖書館學刊 1996：2 (1996年3月)，頁36。

註 2：Clara M. Chu and Ann O'Brien, "Subject Analysis: the Critical First Stage in Indexing," Journal of Information Science 19(1993), p.439.

註 3：馬緒超、李傳健，「文獻主題分析初探」，文獻工作研究 1997：5 (1997年9月)，頁2。

註 4：Nancy J. Williamson, "The Importance of Subject Analysis in Library and Information Science Education," Technical Service Quarterly 15:1-2 (1997), p.73.

註 5：陳敏珍，「主題分析理論與方法之探討」，在圖書館與資訊研究論集：慶祝胡述兆教授七秩榮慶論文集，胡述兆教授七秩榮慶祝壽論文集編輯小組 (臺北市：漢美，民國85年)，頁692。



- 註 6：陳敏珍，「主題分析與主題檢索初探」，國立臺灣師範大學圖書館館訊 14（民國 83 年 12 月），頁 4。
- 註 7：同上註，頁 4-5。
- 註 8：同註 5，頁 694-697。
- 註 9：同註 3，頁 3-4。
- 註 10：陳佳君，「從知識結構探討主題分析」，書府 16（民國 84 年 6 月），頁 39。
- 註 11：Arene G. Taylor, The Organization of Information (Englewood: Libraries Unlimited, 1999), pp.133-135.
- 註 12：同註 3，頁 4-5。
- 註 13：Louis Mai Chan, Cataloging and Classification: an Introduction (New York: McGraw-Hill, 1994), pp.160-164.
- 註 14：同註 10，頁 42-46。
- 註 15：馬張華、鞏雪芹編著，中小學圖書館文獻分類與主題標引（北京市：北京圖書館出版社，1998 年），頁 160-162。
- 註 16：美國資訊科學會台北分會編，索引典理論與實務（臺北市：編者，民國 83 年），頁 3。
- 註 17：同註 3，頁 3。
- 註 18：蔡明月，線上資訊檢索：理論與應用（臺北市：學生，民國 80 年），頁 161-169。
- 註 19：陳昭珍，「控制詞彙與非控制詞彙主題索引問題之探討」，在 2000 年讀者服務研討會手冊（臺北市：實踐大學，民國 88 年），頁 137-138。
- 註 20：同註 15，頁 163。
- 註 21：張淳淳，「組合索引法」，中國圖書館學會會報 44（民國 78 年 6 月），頁 143。
- 註 22：陳昭珍，「主題索引問題初探」，ASIS 台北學生分會會訊 5（民國 81 年 6 月），頁 6。
- 註 23：同註 2，頁 452-453。
- 註 24：引用麥麟屏教授於民國 89 年 5 月 20 日於台大圖書館演講 powerpoint 資料。
- 註 25：何光國，「主題標目結構之評析」，圖書館學與資訊科學 19:2（民國 82 年 10 月），頁 33-34。
- 註 26：盧秀菊，「中文主題標目與標題表」，中國圖書館學會會報 59（民國 86 年 12 月），頁 24-42。
〈http://lac.ncl.edu.tw/會報/lac59/art_5.htm〉 (22 May 2000)。
- 註 27：同註 19，頁 140。
- 註 28：同註 5，頁 699。
- 註 29：陳昭珍，「主題索引理論之探討（上）：主題分析」，書農 9（民國 81 年 12 月），頁 24。
- 註 30：同註 25。
- 註 31：同註 19，頁 141。
- 註 32：同註 28，頁 6。
- 註 33：曾繁娟，「中文標題檢索效益之研究—以國立臺灣大學 TULIPS 系統為例」，大學圖書館 2:1（民國 87 年 1 月），頁 100-123。〈http://www.lib.ntu.edu.tw/pub/univj/uj2-1/uj2-1_9.html〉 (8 Jun. 2000)。
- 註 34：同前註。

