

第二十二章

檔案索引典探討

Thesauri for National Archives

吳美美

Mei-Mei Wu

臺灣師範大學圖書資訊學研究所教授兼所長

Professor & Chairperson, Graduate Institute of

Library & Information Studies, National Taiwan

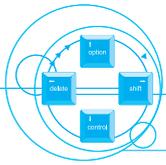
Normal University

壹、前言

檔案和圖書館藏品有許多相似之處，都是收藏各類記載內容、知識和紀錄的媒體，但是在典藏和資料組織、建立檢索線索的方法上，卻有不同的發展軌跡。主要的不同在於圖書館藏品有許多複本存放於本館和其他圖書館，此項特質和檔案的獨特（**unique**）性質有很大的不同。

圖書館由於藏品複本的特質，透過合作編目，或書目共用中心（**bibliographic utility**）可以節省人力，促進資料的交換。圖書館在20世紀中期有驅力發展編目規則標準化和自動化，即是為配合節省複本作業的人力和加速資訊交換。至於檔案的標準化和自動化遲了將近二十年，直到網際網路興起，許多分散各處的相關檔案有需要透過標準化的過程彼此集結，檔案的資料組織和標準化的需求才受到重視。

過去曾經有反對檔案標準化的聲浪，認為檔案和檔案之間的關聯很少，一個機構的編目資料，很難為其他機構再利用，抄錄編目並不需要等，但是網際網路



使得世界互相關聯，自動化、數位化或標準化不是為節省人力，卻是為使世界知識網可以拼湊、結合和推廣，是為檢索的需求，大於減輕重複工作的需求。

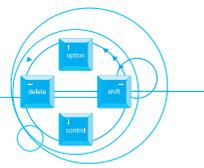
網際網路透過數位傳輸的過程，促進全球知識結網、互相交換、成長。檔案資料內容本屬獨特的資訊類型，不易和其他資料分享、交換，卻因網際網路傳輸便利的特質，全球有關聯的檔案可以有機會互相連結比對、交換。由於將相關聯的資料資訊和知識互相交換的需要，相關檔案集合在一起，便於從事比較研究，數位檔案的資料組織方法受到重視。本文主要目的在探討檔案索引典的功能、效益及可行性，尋求檔案搜尋、檢索更有效的方法。

貳、定義

「檔案」是指由機關、政府、或個人所產生的任何形式的「記錄」或「歷史資料」，也有認為「檔案」是指保存及維護這些「記錄」或「歷史資料」的「所在」，例如Archive is a central place to store and maintain records and historical materials (regardless of format) created by an organization, government or individual(s)。(註1) 討論檔案的整理、組織和檢索，必須先思考所處理的檔案是指個別檔案，或是針對某個所在地保管的所有檔案，或是人類社會某個相關活動的所有檔案，例如全球經濟、法律、全球犯罪檔案等，才能規畫檔案組織的格式。

資料處理、組織，如分類、編目等，有兩個主要功能，其一，指引使用者原文件的所在(physical location)，是「實際取得」(physical access)；其二，指引使用者是否有某相關的資料，是提供「是否有哪些相關資料存在」的訊息(intellectual access)。文件的所在地是一對一的，一個存放地，一個路徑指引；但是主題的指引可以有許多指引路徑(access points)，例如，一份檔案文獻只能有一個索書號，索書號能提供檔案所在位置的指引，但是要找尋這份檔案文獻的線索應該愈充分愈好，甚至可以有幾個主題類號、主題標目等，在檔案「文本表徵」(text representation)方面，能善盡思惟及內容方面的路徑指引。

「索引典」是「主題詞」的集合。一般而言，資料處理在主題分析方面有兩種方法，其一，集結相關的資訊，如圖書館採用「分類號」和「主題標目」集結



相關的資訊，兩者的層次相當，都是一個較大的主題範圍，用以集結相關的資訊。另一種方法是分析，「索引」和「關鍵字」通常用以分析一篇文獻、一本書的主題內容，例如一本書的書後索引。索引典收集的「主題詞」可能是指和分類號層次相當的「主題標目」，也可能是指範圍比較專精的「關鍵字」。索引典收錄詞彙的層次和細膩程度，由索引典編製者依使用需求和使用行為決定。

索引典在紙本時代，其收錄的詞彙稱為「控制詞彙」，因為資料排列或檢索時，以索引典收錄的詞彙作為索引及檢索項（**access points**）。電腦化之後，檢索者透過自然詞彙，可以利用全文檢索機制來查詢，如果系統設計者建立一套自然詞彙和控制詞彙的對照表作為系統介面，對於檢索者而言，採用任何詞彙檢索即由系統自行對照轉換，並無所謂控制詞彙和自然詞彙的區別。

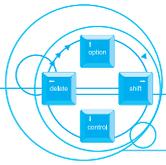
索引典是某個領域詞彙的集合，並界定詞彙的關係，如廣義詞、狹義詞、相關詞等。嚴謹的索引典，可以視如該領域的知識樹。索引典的功能作為知識樹、知識結構的功能，大於所謂控制詞彙的功能。利用電腦設計檢索機制之後，索引典更進一步作為檢索的工具，檔案的編目項中，主題分析就更有必要了。

簡言之，索引典是促進檢索效益的有效工具，索引典的具體目的有五：（註2）

- （1）將使用者的一般檢索詞彙轉為系統使用的控制詞彙；
- （2）使索引詞彙一致；
- （3）同義字控制；
- （4）表達語意關係；
- （5）作為檢索輔助工具。

參、檔案主題分析的重要趨勢

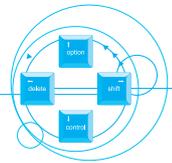
一般而言，資料整理及檢索工具有下列幾種：書目（**bibliographies**）、目錄（**catalogs**）、索引（**indexes**）、指引（**finding aids**）、註記（**registers**），其中每一種工具都包括資料形式描述和主題描述，以便於資料查找。檔案過去一般都是利用指引（**finding aids**）作為查詢工具，指引通常就是檔案清單，列舉檔案的來源和特質等，檔案的主題一向不受特別重視，檔案數位化之後，主題分析就顯得重要了。



Foster (1997) 接受英國檔案局委託於1996年進行一項檔案主題分析研究，(註3) 總共調查七所國家級的檔案典藏中心、三所地區檔案局、四所專業典藏中心以及四所大學檔案典藏中心，期望了解英國檔案工作者對於檔案主題分析的看法。節錄幾個重要的研究發現：(一) 18所檔案機構大多有主題分析，並都希望建置詞彙標準；(二) 美國國會主題標目(LCSH)不適用於檔案主題分析；(三) 英國國家檔案局的基本主題詞彙表雖然有用，但是不夠詳盡。Foster依據研究所得，提出建置檔案索引典應注意的事項，包括成立檔案索引典的管理單位、建立索引典編製說明、該索引典必須有層級、不同地區名詞對照、地理名詞、文件類型說明，如個人日記或自傳等，並應有六人小組，每年定期集會兩次，以檢討監督該檔案索引典的發展。

肆、檔案相關索引典發展近況

亞利桑那州立大學圖書館有檔案管理網頁，(註4) 介紹檔案組織的方法，並提供線上索引典，便於從事檔案整理人員查詢，其中介紹SiteSearch檔案資料庫的內容，重要欄位主要取自Dublin Core中的幾個要素：檔案名稱、創造者、主題、摘要、出版者、外型描述、類別、媒體形式、箱子編號、卷宗編號、提供者、日期、館藏名稱、檔案型態、電子檔案型態、標示號、原始來源、語言別、大小、版權、電子聯絡信箱、網路檔案位址等，並有各要素的描述、標準化作業、處理形式及範例等(如表一)。在「主題」部分，連結到11個不同主題的詞彙表，例如「藝術、建築索引典」、「國會圖書館索引典」、「醫學主題標目樹狀結構」、「地理名稱索引典」、「教育索引典」、「U C Davis照片索引典」等(如表二)，顯然是在建議檔案編目人員，自其中選擇合適的詞彙，作為該檔案的主題詞。亞歷桑那大學圖書館並未建立自己的索引典，但是鼓勵採用控制詞彙，作為檔案的主題描述。以該系統目前處理主題的做法，顯然尚無索引典可以作為檢索工具。不過參考其他索引典來編目檔案主題，應是檔案主題分析的第一步，也是未來發展索引典的基礎。



為組織檔案所發展的索引典愈來愈多，Kim Schroeder自Wayne 州立大學圖書資訊學研究所畢業，也獲有檔案管理的學分證明證照，在1994年成立Archive Impact 公司，提供檔案相關產品服務，由於創辦人圖書資訊學及檔案學背景，其網頁整體風格十分善盡訊息性（informative），提供相當完整充分的檔案組織資訊，其中介紹幾個檔案索引典：（註5）

表一 亞歷桑那州立大學檔案資料組織及檢索網頁

Collection Types for Sitesearch

Archive

The following table describes the descriptive information that is used to create a record for an Archiverecord in a SiteSearch database. There are separate collection types for images, journal articles and monographic material.

Element is the field title that will be used to index objects in the SiteSearch Database. The Dublin Core Element is interpreted here for our specific use.

Description outlines what the content of the field should be.

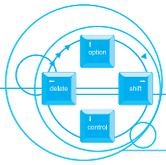
Standard states the name of a national or internationally recognized convention that is used for the specific field, if one applies.

Form illustrates how the content of the field should appear.

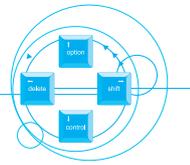
Example illustrates how the data that is entered into a specific field should appear.

(Link to Document Format Type and Document Type Definition)

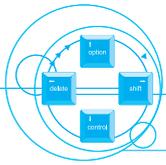
Element	Description	Standard	Form	Example
Title	The name given to the resource by the CREATOR or PUBLISHER.	Taken from original		Arizona:Jewel in the Sun
Creator	The author or person(s) or organization(s) primarily responsible for the intellectual content of the resource.	Taken from original	Lastname, Firstname MI	Strachman, Gordon



	intellectual content of the resource.			
Subject	The topic of the resource, or keywords or phrases that describe the subject or content of the Resource.	Various	See <u>Thesauri</u> page.	Marketing Chamber of Commerce Arizona Agriculture Tourism
Abstract	A textual description of the content of the resource, including abstracts in the case of document-like. May be copied directly from a document, journal article or other monograph.	Taken from original, or produced by indexer	Descriptive sentences	This pamphlet is a black and white reprint from the August 1948 edition of the Santa Fe Magazine.
Publisher	The entity responsible for making the resource available in its present Form, such as a publisher, a University department, or a corporate entity.	Taken from original		Santa Fe Magazine
Physical Description				
Genre				Marketing Brochure
Medium				Paper
Box				11
Folder				26
Contributor				Santa Fe Magazine



Date	The date the resource was made available in its present form.	ANSI X3.30-1985	YYYYMMDD	1948?
Collection				University of Arizona, Special Collections
Type	The category of the resource, such as home page, novel, poem, working paper or technical report.	<u>List</u>		text.pamphlet
Format	The data representation of the resource, such as text/html, ASCII, Postscript file, executable application, or JPEG image.	<u>List</u>		.pdf
Identifier	String or number used to uniquely identify the resource.		Alphanumeric	C9791Pam1
Original	The work, either print or electronic, from which this resource is derived, if applicable.			Reprint of Santa Fe Magazine article
Language	Language(s) of the intellectual content of the resource.	<u>NISO</u> <u>Z39.53</u>	Three letter abbreviation	eng
Size	Physical description of the object.	Taken from original		1MB
Copyright	Copyright notice or a rights-management			University of Arizona Library,



	statement, name of the organization or persons who hold copyright on this resource.			Special Collections
Email	The email address of the Creator or other email address if appropriate.			andy@any.com
URL	The Web address associated with this resource.			http://xxx.yyy.edu

Send comments and inquiries to Garry Forger, Assistant Librarian
 University of Arizona Library, 520-621-8753
 資料來源: <http://dizzy.library.arizona.edu/sitesearch/archive.html>

表二亞歷桑那州立大學圖書館檔案管理線上索引典

Online Thesauri

The following resources can be used to select terms that will be used when indexing items in SiteSearch.

The Art and Architecture Thesaurus Browser - from the Getty Information Institute
Library of Congress Thesauri Home Page - Connect to: Thesaurus for Graphic Materials I & II, Legislative Indexing Vocabulary (LIV), and the Global Legal Information Network (GLIN) Thesaurus.

NLM - MeSH Tree Structures - Medical Thesaurus

The Union List of Artist Names Browser - from the Getty Information Institute

Thesaurus of Geographic Names - from the Getty Information Institute

UCDavis photographic thesaurus - Headings from the Library of Congress Graphic Materials Thesaurus used in the Eastman Collection metadata.

Acronym and abbreviation list - Locate definitions for TLA's (Three Letter Acronyms) and others.

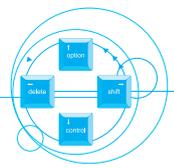
NASA Thesaurus - This is an electronic version of the NASA Thesaurus, an experimental project.

Welcome to the Search ERIC Wizard - Information on Education.

1998 PACS - Physics and Astronomy Classification Scheme.

Access the PLANTS National Database - United States Department of Agriculture Natural Resources Conservation Service.

Send comments and inquiries to Garry Forger, Metadata Librarian



University of Arizona Library, 520-621-2916

資料來源：<http://dizzy.library.arizona.edu/sitesearch/thesauri.html>

一、澳洲照片索引典 Australian Picture Thesaurus (APT)

澳洲影像主題詞彙的集合，以澳洲當代術語描述澳洲圖書館、博物館以及檔案館所收藏的影像中包括的物件、人物、地方、建築、活動等。

二、加拿大素養索引典 Canadian Literacy Thesaurus

加拿大素養索引典由加拿大素養索引典聯盟發展維護，為成人素養領域的雙語標準詞彙。

三、國會圖書館索引典 Library of Congress Thesauri

包括下列四個索引典：

(一) 全球法律資訊網絡索引典

Thesaurus for the Global Legal Information Network (GLIN)

1950年代以來，幾個國家共同協力發展以英文為主的全球法律資訊網絡索引典。

(二) 法律索引詞彙

Legislative Indexing Vocabulary (LIV)

美國國會圖書館發展的法律索引典用於法律及公共政策相關出版品。

(三) 圖像資料索引典I：主題詞

Thesaurus for Graphic Materials I: Subject Terms (TGM I)

(四) 圖像資料索引典II：類別及外型特質詞彙

Thesaurus for Graphic Materials II: Genre and Physical Characteristic

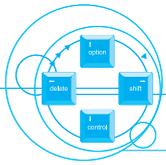
Terms

(TGM II)

圖像資料索引典I,II由美國國會圖書館印刷及照片部門開發，其中II就包括有600詞彙。詞彙來源參考其他檔案資料的特質。

四、聯合國教科文組織索引典 UNESCO Thesaurus

英國倫敦大學電算中心協助聯合國教科文組織（UNESCO）製作一個索引典（註6）放在網頁上，提供非商業的自由使用。該索引典目前尚未發現在教科文組



織的資料庫中應用。一個索引典如果只是一個詞彙的集合，沒有檢索的功能，只能說是一個知識結構、知識樹。國內有圖書資訊學索引典，也未被利用於相關主題的資料庫檢索，都十分可惜。

UNESCO索引典還是有為其他機構採用為檢索機制，如AIM25 (Archives in London and the M25 area) 即採用UNESCO的索引典，並用以檢索文獻。(註7) 經過作者試用，可能由於AIM25資料庫資料量不大，該系統雖然提供檢索和索引典主題的功能，試驗檢索兩個案例，檢索結果都不怎麼理想。

Haynes教授在奧克拉荷馬州立大學任教，也整理幾個動態影像索引典，(註8) 如：Film and video glossaries and dictionaries、Library of Congress Subject Headings (LCSH)、PAIS、Political Science Thesaurus、Moving Image Genre-Form Guide (MIG)、Moving Image Materials: Genre Terms (MIM) 等。Haynes教授為了開發、建置動態影像資料索引典，也整理並建議參考有關機構建置的相關索引典如：

Public Affairs Video Archives at Purdue University <http://pava.purdue.edu>

Special Media Archives at NARA <http://www.nara.gov/naramedia.htm>

Vanderbilt University Television News Archive <http://tvnews.vanderbilt.edu/>

Library of Congress Motion Picture/Broadcasting/Recorded Sound Division

<http://lcweb.loc.gov/rr/mopic/migintro.html>

Baylor Collections of Political Materials

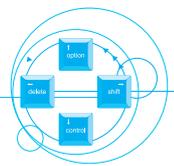
<http://diogenes.baylor.edu/WWWproviders/Library/LibDepts/BayCollPolMat/>

[BCPM.html](http://diogenes.baylor.edu/WWWproviders/Library/LibDepts/BayCollPolMat/BCPM.html)

Carl Albert Center for Congressional Study

<http://www.ou.edu/special/albertctr/main.html>

檔案索引典的發展正方興未艾。主要的原因當然是檔案管理從剛開始發展時關心儲存和管理的議題，已經逐漸發展成熟，足以典範轉移到重視檢索的議題。為促進檢索效益，檔案主題分析就很重要，檔案索引典的編製，也就受到各國國家檔案局的重視。



伍、索引典編制方法

利用google查詢，可以發現英、美、澳等國家檔案局都十分重視檔案索引典的編製。澳洲國家檔案局並在其網頁提供檔案索引典的編製方法及發展步驟（如圖一、二），內容詳盡，很值得借鏡。

索引典的編製應注意過程及標準化的問題，首先要界定該索引典的主要功能、提供哪些服務，有哪些資料轉換程序等，寫成文件之後，即開始根據文件需求，選擇詞彙、自現成的索引典或分類表中轉化而來，並決定詞彙和詞彙之間的關係，以編製完成索引典，以澳洲國家檔案局為例，有幾個檔案管理相關標準可參考（如圖三）。

編製索引典最耗神、費時者，是選擇詞彙及建構基本詞彙表，澳洲國家檔案局有一套名為AAA的基本詞彙表，作為政府內部各部門編製索引典的基本詞彙參考來源，利用該基本詞彙表，接著決定可用的詞彙及不採用的詞彙，並說明詞彙的應用範圍等，一個屬於該機構合用的索引典就可以產生（如圖四）。

Haynes教授就其親自設計並建置動態影像資料索引典的經過，說明索引典的編製過程，包括選擇詞彙和決定詞彙之間的關係，步驟簡述如后：（註9）

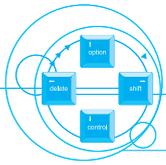
步驟一：選擇詞彙的方法

- 1、界定範圍--首先決定索引典主題的深度和廣度，分為核心範圍、次要範圍和外圍範圍。
- 2、決定詞彙來源--決定核心範圍的文獻、檔案。
- 3、擇原始詞彙--從詞彙來源中，依詞彙出現頻率選擇。

步驟二：詞彙建構的方法

Haynes教授說明傳統方法是將原始詞彙寫在卡片中，加以排列組合。現在大多是擇一索引典軟體來經營詞彙。處理詞彙的過程包括：

- 1、分析互相衝突的詞彙關係。
- 2、建立互有關係的詞彙。
- 3、慮多語言互通需求。

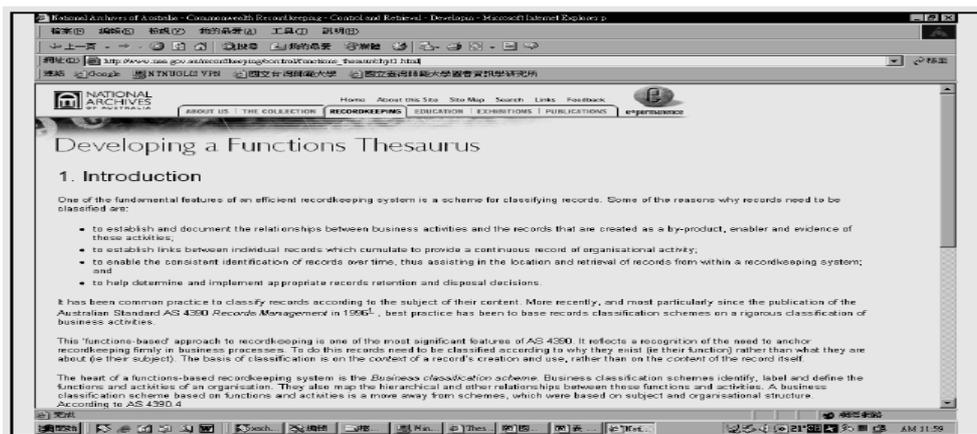


- 4、慮詞彙群聚和多面分析的需求等。同時註記詞彙來源註、詞彙選入索引典日期註、範圍註、詞彙史註、類別號等。

陸、結語

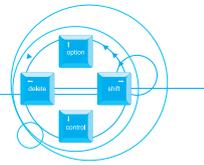
編製索引典需注意兩項議題，詞彙層次的選擇和詞彙關係的決定。至於詞彙的來源可以從現有文獻產生，包括中英文都有發展自然詞彙的關鍵字擷取程式，中文方面清華大學張俊盛教授開發一套「文後索引詞彙擷取」程式，程式依據詞頻等規律，自動選出詞彙。

中英文自然語言研究領域都有wordnet計畫，對於詞彙的知識大有幫助。雖然自然語言的研究發展大有進展，但是目前編製索引典所利用的方法，仍沿用傳統索引典編製方法，如圖一。相信不久會有自動產生詞彙，並將詞彙和詞彙之間的關係呈現的系統。



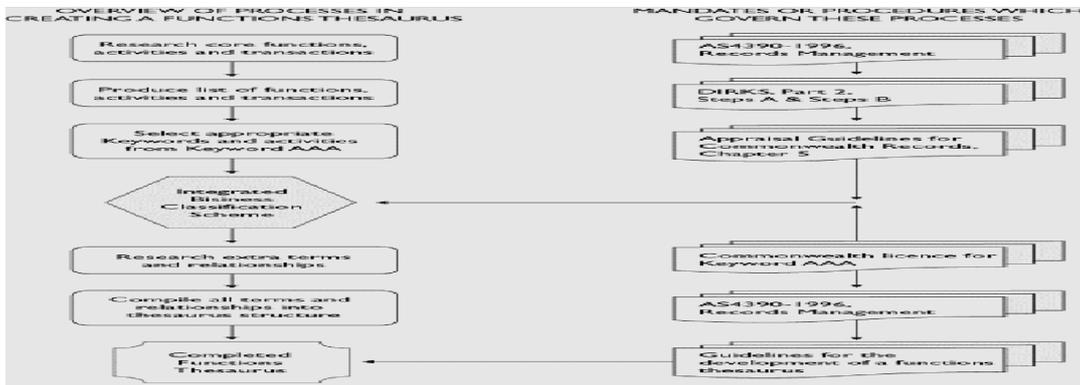
圖一 澳洲國家檔案局檔案索引典發展指引網頁之一

資料來源：http://www.naa.gov.au/recordkeeping/control/functions_thesaur/chpt1.html



圖二 澳洲國家檔案局檔案索引典發展指引網頁之二

資料來源：http://www.naa.gov.au/recordkeeping/control/functions_thesaur/contents.html



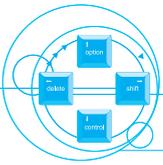
圖三 索引典建構的程序

資料來源：http://www.naa.gov.au/recordkeeping/control/functions_thesaur/chpt5.html



圖四 澳洲政府機關內部索引典建置主要步驟

資料來源：http://www.naa.gov.au/recordkeeping/control/functions_thesaur/chpt5.html



【原刊載於檔案管理局出版之檔案季刊（九十二年三月）第二卷第一期】

註釋：

註1：<http://www.archiveimpact.com/glossary1.htm>

[cited 2002/12/20]

註2：Haynes, Kathleen J. M. (1999). What's In a Name: Traditional and Emerging Methods for Subject Analysis. Design and Creation of a Thesaurus for Archival Moving Image Materials in the Political Communication Center of the University of Oklahoma. In AMIA Annual Conference, November 1-6, 1999. Montreal, Canada.

http://www.amianet.org/04_Annual/04f03_Montreal/04f03e_SubjAnlys.html

[cited 2002/12/20]

註3：Foster, Janet (1997). Subject indexing survey and analysis-- a report for JISC: NFF Archives Sub-Committee.

<http://www.kcl.ac.uk/projects/srch/reports/subindex.htm> [cited 2003/01/15]

註4：<http://dizzy.library.arizona.edu/sitesrch>

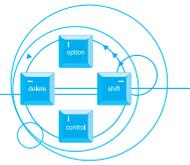
[/archive.html](#) [cited 2003/01/15]

註5：<http://www.archiveimpact.com/aboutus.htm>

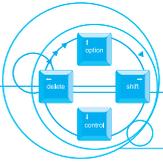
[cited 2002/12/20]

註6：<http://www.ulcc.ac.uk/unesco/terms>

[/list74.htm#Information%20processing](#) [cited 2003/01/15]



- 註7: <http://www.rslp.ac.uk/projects/research/30.htm> [cited 2003/01/15]
- 註8 : Haynes, Kathleen J. M. (1999). What's In a Name: Traditional and Emerging Methods for Subject Analysis. Design and Creation of a Thesaurus for Archival Moving Image Materials in the Political Communication Center of the University of Oklahoma. In AMIA Annual Conference, November 1-6, 1999. Montreal, Canada.
http://www.amianet.org/04_Annual/04f03_Montreal/04f03e_SubjAnlys.html [cited 2002/12/20]
- 註9 : Haynes, Kathleen J. M. (1999). What's In a Name: Traditional and Emerging Methods for Subject Analysis. Design and Creation of a Thesaurus for Archival Moving Image Materials in the Political Communication Center of the University of Oklahoma. In AMIA Annual Conference, November 1-6, 1999. Montreal, Canada. [http://www._____ amianet.org/04_Annual/04f03_Montreal/04f03e_SubjAnlys.html](http://www.amianet.org/04_Annual/04f03_Montreal/04f03e_SubjAnlys.html) [cited 2002/12/20]



檔案資訊資源管理
