

教育資料與圖書館學

Journal of Educational Media & Library Sciences

<http://joemls.tku.edu.tw>

Vol. 46 , no. 1 (Fall 2008) : 055-080

主題地圖與網頁資源保存之探討

Discussion on Web Archives Using Topic Maps

余顯強 Shien-Chiang Yu

Assistant Professor and Director of Library

E-mail: ysc@cc.shu.edu.tw

[English Abstract & Summary see link](#)

[at the end of this article](#)

JoEMLS

<http://joemls.tku.edu.tw/>

主題地圖與網頁資源保存之探討

余顯強

副教授兼圖書館館長
世新大學資訊傳播學系
E-mail: ysc@cc.shu.edu.tw

摘要

網路資源已成為目前重要的資訊管道，其內容豐富多元，可提供一般民眾資訊檢索及學術界研究參考所需。然而網站成長的快，消失的也快，今日所看到的網站內容，很可能隨時因伺服器關閉、網站改版等因素而無法再次被利用。網頁資源可以是記錄人類社會某一段時間的數位化資訊，包括當時的視覺設計藝術、風格，及包含在其網頁內的各類資源，十分具有文化保存的價值。由於數位資訊保存的脆弱特性，在網頁資源的保存上必須透過實驗發展策略性的典藏與管理方式，因應不同版本HTML描述的網頁，必須考慮瀏覽器的支援及長久保存的問題，包括標準化、符合未來應用的長久格式、資料萃取與重組等因素。基於這些需求，本研究主要目的在於探討網站資源長期保存的解決方案，因此採用標準化的方式，依據資料結構的角度分析網站資源的組成模式與特性。在研究設計主要是採用轉置方法，將網頁資源原始資訊依據主題地圖語法重新組織網頁結構，以提供網頁內涵描述與整體資訊內容的有效管理與應用。

關鍵詞：網路資源，長久保存，主題地圖，網頁轉置

前 言

Web興起與普及化後，諸如紙張等傳統載體經過數位化而延伸其再生與利用（再生數位）的範圍，使得各式各樣電子資源蓬勃發展，特別是網路資源的成長更是迅速與多元。網路資源已成為目前重要的資訊管道，其內容豐富多元，可提供一般民眾資訊檢索及學術界研究參考所需。然而網站成長的快，消失的也快，今日所看到的網站內容，很可能隨時因伺服器關閉、網站改版等因素而無法再次被利用。從文化保存的角度來看，這些網頁資源如同新聞報紙資

2008/08/05投稿; 2008/10/22修訂; 2008/10/24接受

料一般，可以是記錄人類社會某一段時間的數位化資訊，包括當時的視覺設計藝術、風格，及包含在其網頁內的各類資源，十分具有文化保存的價值。因此，各國國家級典藏機構多已開始進行Web的保存計畫(註1)，其目的即在於網路資源之保存，不僅要提供現時使用，更須要考量人類文明的延續，及提供未來世代知識的使用。

資訊編排與描述等組織行為是資訊利用的基礎，如何對各式各樣數位資源作有效與一致性的描述，在Web環境下更顯其重要性。如何有效地保存網頁資源的內容，並詳實地描述這些被保存之網路資源便有其迫切性與重要性。但綜觀網頁資源的內容，多以非結構化的HTML檔所組成，而這些HTML檔，有的僅是靜態網頁的形式，有的卻是透過後端程式執行的輸出結果。網頁使用的HTML語法屬於程序式(procedure)的標記方式，標示的目的在於資料呈現的效果，因此如瀏覽器這一類的剖析系統，無法透過呈現的結果推導出內容所包含的意義，而使得網頁無法提供精確查詢、自動文件處理與內容分析(註2)。由於HTML語法缺乏結構化的因素，造成網頁資源資訊再利用與資料交換的限制，使得網頁資源保存的長久性受到約束，並妨礙了其未來的應用範圍。因此研究網頁資源的保存，必須能夠提供後續應用與萃取網頁資源內部的資訊，才能達到其長久保存的目的。

因此，保存的資訊應該包含網頁資源的相關描述，也就是metadata紀錄；及網頁資源本身。詳細記錄網頁資源的內涵與產生方式，必須要有一完整規劃的metadata格式予以描述與管理；而個別網頁的結構由於涉及共用、連結、框架區塊(frame)等非結構化的特性，必須另外透過解構的技術，將內容的各種屬性分離出來，予以結構化重組。主題地圖(Topic Map)是結合語意網路的基本模型與索引的功能，能夠提供知識表達的架構(註3)，因此，本研究嘗試以主題地圖來解構網頁的結構，利用主題之間的關聯模式，不僅保存網頁原本的連結關係，也可將各網頁之間的相關性依據使用者的需求予以重組串連起來，達成原始網頁資源長久保存與永續利用的目的。

二、網頁資源保存的意義與方法

因應不同版本HTML描述的網頁，必須考慮瀏覽器的支援及長久保存的問題，包括標準化、符合未來應用的長久格式、資料萃取與重組等因素。但數位資訊的長久保存必須考量資訊科技典範的轉移，也就是說，數位資訊容易因技術的原因而無法被使用，這些技術因素包括：資訊儲存載體的失效、閱讀器/瀏覽器不再支援以致無法讀取、解譯儲存資訊的軟體喪失。資訊儲存載體的失效多由於儲存媒體的損壞造成，例如儲存於光碟的資料可能因記錄媒體的一個位元(bit)的錯誤，造成接續資料無法讀取的狀況。閱讀器/瀏覽器不再支

援常常是因載體的規格改變，例如早期使用的磁帶、磁片，由於儲存媒體的日新月異而淘汰，以致縱使仍完整保存的磁帶、磁片，卻無對應的硬體設備能夠讀取。而解譯儲存資訊的軟體喪失則是因多數檔案必須要有對應的應用軟體處理，而應用軟體或資料格式不斷因應電腦技術、作業系統、標準的改變而改版，或因市場競爭的因素被淘汰，例如試算軟體Lotus的檔案，如無適當的應用軟體能夠處理，縱使保存良好的檔案，亦無法順利解譯顯示於螢幕上。

基於數位資訊保存的脆弱特性，在網頁資源的保存上必須透過實驗發展策略性的典藏與管理方式。Wauth等人(註4)提出數位資訊長期保存包括五項關鍵點：(一)將資訊儲存於描述 metadata 記錄內的封裝(encapsulation)方式；(二)資訊轉換成能夠直接被解讀與呈現的自我文件化(self documentation)；(三)最低依賴系統、資料或文件的自我充分化(self sufficiency)；(四)能夠提供未來使用者發現或實作軟體來呈現資訊內容的內容文件化(content documentation)；(五)儲存的資訊能夠被組織機構使用的組織保存化(organization preservation)。據此，本研究在設計時，便必須考量多層次資料描述的方法，配合合適的 metadata 格式以滿足封裝的關鍵要求，且採用資料著錄的標準標示語法，以達到自我文件化與內容文件化的目標，並研擬採用標準的結構化方法，以符合內容文件化與組織保存化的應用需求。

若欲解決保存網頁內容所具備的資訊能夠在未來應用的目標，儲存網頁資源時可考慮處理的方式包括：(一)完整儲存網頁原始的內容與結構，並透過內容分析與人工標記方式，編制該網頁資源的 metadata 紀錄，提供爾後檢索利用；(二)將原始網頁資源重新解構，再依據採用現今 XML 語法、涵蓋資源描述的特性，並具備連結或資源指引的描述方式，重新建構網頁資源，使能夠在保存原有呈現內容的需求之下，還可延伸其應用範圍。網頁資源的典藏必須兼顧「長久保存」與「最小損害」(do minimal harm)的原則，比較上述兩種處理模式，很明顯地第二種方式能提供長久保存較佳的應用性，但必須解決其實務上執行的可行性。

依據資料內涵的特性與應用目的的差別，來達成長久使用數位資訊的內容，可以有系統保存(system preservation)、更新(refreshing)、轉置(migration)、模擬(emulation)、標準化(standardization)、封裝(encapsulation)、複製(redundancy)、轉成紙本或類比媒體(converting to paper or analog media)等方式(註5)。為求能符合前述數位資訊保存的五項關鍵，依據Lawrence等人的研究結論與建議，提出轉置較能滿足風險管理的品質要求(註6)，且分析HTML文件格式的特性，本研究認為轉置是最適合的方式。

基於這些需求，本研究主要目的在於探討網站資源長期保存的解決方案，因此採用標準化的方式，依據資料結構的角度分析網站資源的組成模式與特

性，並將資源保存分成兩個層面，分別是 metadata 格式與網頁內容重新建構成結構化文件。

關於 metadata 格式的分析，可參考學者王麗蕉(註7)所規劃之 Web 資源多元層次描述 (multi-level description) 模式，以尊重資源產生者及其原有資源結構所具體實施的控制層次。此模式有別於主題內容分類的主觀性，故具有組織工具的客觀性與目的合理性，並能呈現資源產生的背景情境，提供資源歷史價值的運用；而在網頁資訊內容的層面上，內容將包含文字、圖形、聲音、影像、視覺等多種不同元件所組成。

在網頁內容保存方面的研究設計，主要便是採用轉置方法，及分析處理過程所需的程序來達成研究目的。因此，本研究將應用系統分析方法解構網頁，若內容所代表的資源過於專屬(例如 JavaScript、Applet、Flash 等程式或物件)，可依據所屬類型以資源描述框架 (Resource Description Framework, RDF) 標準加以描述，再透過主題地圖 (Topic Maps) 將這些資源之間的關聯性連起來，並結合多元層次描述的 metadata 紀錄，以提供網頁內涵描述與整體資訊內容的有效管理與應用。

三、主題地圖

(一)主題地圖之緣起與發展

主題地圖是用來提供一套建立資訊資源之間的關係，又可提供檢索及連結的功能，以達成資訊探索的目標。最初，為提供 ISO/IEC 8879-1986 標準通用標示語言 (Standard Generalized Markup Language, SGML) 多媒體與超連結的功能，Goldfarb 與 Newcomb 嘗試設計一種在任何時間與空間均能對多媒體與文件檔案間達成超連結的邏輯結構，因此發展出多媒體/時間為基礎的結構化語言 (Hypermedia/Time-based Structuring Language, HyTime)，並於 1992 年通過成為 ISO/IEC 10744 標準(註8)。

HyTime 的語法承襲於 SGML，由於過於複雜，因此 1991 年在 San Antonio 舉辦的 ACM Hypertext 成立之 Davenport 群組在 1993 年後分成兩個群組，一個群組專注於發展 DocBook 格式，另一個是歸屬於 GCA Research Institute (GCARI) 的 Conventions for the Application of HyTime (CApH) 群組，以期能發展簡化 HyTime 的應用子集，用來提供一般性主題在分享時的識別資訊物件(註9)。該計畫所發展出的解決方案稱為「主題導航地圖」(Topic Navigation Maps)，於 1996 年移轉至 ISO 組織，並於 1999 年 12 月通過成為 ISO/IEC 13250 標準。主題導航地圖採用 HyTime 作為主題地圖的定義語法，屬於 SGML 的文件型別定義 (Document Type Definition, DTD)，不僅可定義各欄位類型的子集合，亦可將欄位、屬性等名稱重新命名，因此能夠將各種資訊概念描述成為主題，各主題能

夠具備自己的名稱、屬性、資源指引等內容，並可定義主題與主題之間的連結關係。

由於SGML的複雜、軟體工具成本過高與使用不易等因素，限制了其應用的範圍(註10)。而HTML則受限於其在Web上使用的專屬且缺乏結構化的功能，因而W3C發展了一套新一代資料標示語言XML，其不僅能適用於Web作業環境，並能夠定義結構化檔案交換的資料格式。XML是由SGML的專家群和W3C合作制定簡化SGML的子集合規範，並於1998年2月正式公佈XML1.0版本(註11)。不像HTML由固定的標籤(tag)集合所構成，XML支援語言中立(language neutral)的定義和平台中立(platform neutral)的特性，允許使用者自行定義所需的標示語言，可將資料內容以清楚的標籤表現其意義，並可廣泛地應用在各種領域(註12)。因此，2000年成立的TopicMaps.Org組織即著手研擬採用XML作為主題地圖的語法規則，於2001年3月推出XML Topic Maps(XTM)1.0規格(註13)，並於2002年通過包含了HyTime與XTM兩種語法架構的ISO/IEC 13250: Topic Maps版本(註14)。

(二)主題地圖語法之主要元素

XTM標準描述了主題地圖的主要關鍵的元素(elements)，而這些關鍵元素包括主題(topic)、參照(occurrence)(註15)、關聯(association)、標題描述(subject descriptor)和範圍(scope)(註16)。這些元素的功能與意義分述如下：

1. 主題

主題表示文件中的一個標題，依據XTM標準，標題同義於資源，因此可將主題視為一個資源。但嚴格來講，標題(subject)表示一個**what**，而主題就是一個**what**的資訊表述(information representation)，因此，一個主題能夠用來呈現出被涉及的標題。一個主題是建構在主題地圖中必要的區塊，透過其所指的資源來代表其標題。資源可以是標題的組成物(constitutes)或指引(indicate)，主題的標題也可以是藉由資源的參照關係表示。一個主題可歸屬一個以上的主題型態。此外，主題可有名稱，也可沒有名稱，或有多個名稱，甚至還可包括顯示名稱(display name)和排序名稱(sort name)，提供顯示與排列使用。

2. 參照

參照是指一個包含某些特定資訊資源的主題，該資源可以僅是一個存在於外部的URI(使用<resourceRef>元素標示)，或存在於該筆主題地圖內部的實際資料，或另一個主題(使用<resourceData>元素標示)。因此，參照在主題地圖扮演與主題相關資訊的連結角色，將與該主題相關的資訊群聚起來，表達與該主題有關的其他資訊。

3. 關聯

一個關聯，其連結了兩個以上主題之間的關係。在 XTM 語法結構中是使用 <association> 元素來表示，透過此元素不僅能夠描述各主題之間的關係，亦可應用 <roleSpec> 元素表達主題之間的「角色」與使用 <instanceOf> 元素表達彼此的「關係」狀況。在主題地圖中，比較參照與關聯兩者的差別，參照在於表達與主題有關的其他資訊，而關聯則是將主題之間建立連結的關係予以明確描述。因此，主題之間的參照關係明確，但關聯則會依據各種情況，而允許主題之間有不同的關聯情況出現。

4. 主題描述

原在 ISO/IEC 13250 最初稱為標題描述 (subject descriptor)，在 XTM 則稱為標題指引 (subject indicator)，表示資料原始提供者在主題地圖中針對標題所定義的絕對且明確的識別名稱。主題指引是採用聯合國制定電子商務標準的組織 OASIS (Organization for the Advancement of Structured Information Standards) 制訂的公開標題指引 (Published Subject Indicators, PSI) 作為主題的統一識別名稱，在應用上只要是標題指引相同的主題均可合併。

5. 範圍

範圍是一個特殊的主題，它定義了一組或一定界限內的相關主題，範圍的作用與名稱空間 (namespace) 類似：基本名稱在某個範圍內應是唯一的，如果在同一個範圍內有兩個相同名稱的主題就應合併。

(三) 主題地圖與 RDF 之比較

RDF 是應用在 Web 上呈現資訊的語言，提供在不同應用程式或代理程式 (agent) 之間交換之展現 meta 資訊的共通性框架 (common framework) (註 17)，而主題地圖則是與 W3C 的語意網 (Semantic Web) 技術同屬於 meta 等級之描述資訊資源關係的語意地圖 (註 18)。如表 1 所示，比較主題地圖與 RDF 兩個標準，無論應用目的或技術實作上都非常類似：兩者均可採用 XML 作為描述的標準，也都具備約束語言 (constraint language) 與查詢語言 (query language)。所有相關於主題地圖的概念都可相同地以 RDF 表示，但 RDF 較專注於資源的描述，而主題地圖則著重於資源之間關聯的連結，以 RDF 描述的資源轉換成主題地圖的元素時，某些語意會因此而喪失 (註 19)。

表1 主題地圖與RDF之比較

		主題地圖	RDF
相同點	約束語言	Topic Maps Constraint Language (TMCL)	RDF Schema
	查詢語言	Topic Maps Query Language (TMQL)	RQL (RDF Query Language)
	描述語言	具備XML Topic Maps (XTM) 語法規範	W3C 建議以XML作為RDF的描述語言
	具備唯一識別	使用屬性id標示主題唯一性識別，亦使用URI作為資源唯一性識別	使用URI作為資源唯一性的識別
不同點	標準制定機構	ISO	W3C
	應用層面 (註20)	建立資訊資源較高階的索引方式，以提供資源的尋找	透過結構化描述資源的metadata與邏輯推論的基礎支援語意網的展現
	範圍	提供scope元素，支援範圍的功能	不具備範圍的功能
	定址方式	提供直接與xlink:href屬性的間接定址方式	只提供直接定址的方式

四、應用主題地圖語法之設計

歸納前述的說明可知，主題地圖就其本質而言是很單純，可透過各項元素的應用與串連，保存完整的網頁資源，並可將整個網站內各個網頁之間的關係建立，透過這些主題之間關聯的建立，便可應用主題地圖將非結構的網頁內容重新組織成有座標概念的結構化文件。

1. 保存完整的網頁資源

以「主題」作為網頁內容最基本的單位，個別的「主題」透過主題型態 (type) 的描述，將主題歸類成群，再藉由「參照」將事件相關的「主題」聚合一起，而相關事件之間的語意則可以使用「關聯」予以串連起來，並利用「範圍」限制名稱、資源指引和關聯的有效範疇，如此便可在具備結構化的語法之下，保存原始的整個網頁資源。

2. 建立整個網站內各個網頁之間的關聯

配合「關聯」結合 <roleSpec>「角色」與 <instanceOf>「關係」的標示，便能夠明白表示主題之間上下層級的關聯性，如此可明確的呈現出如圖1所示之索引典 (thesaurus) 與概念類別 (concept class) 所表現的上下位詞與相關詞的架構 (註21)，即能將整個網站內各個網頁之間的關係建立起來。

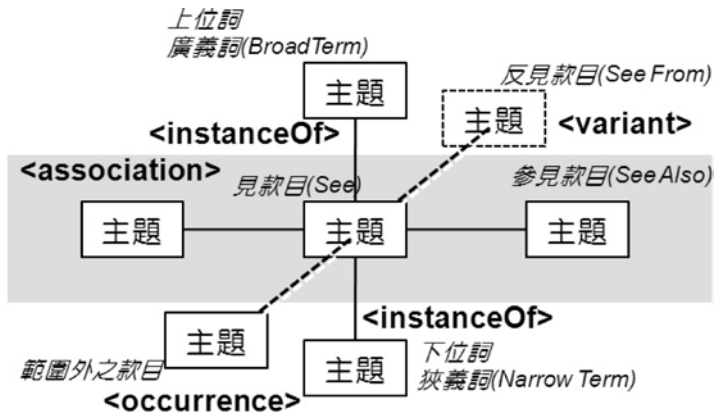


圖1 依據主題地圖表現文獻內容主題之關係架構圖
資料來源：本研究整理

3. 本研究設計之作業流程與資源本體保存的剖析程序

分析上述程序，本研究設計之作業流程如圖2所示，在網頁資源的保存包含了metadata紀錄的描述與資源本體的保存兩個主要作業項目。而資源本體的保存方式是經由下列執行的剖析 (parsing) 程序：將網頁HTML先行解構及內容分群 (zoning)、轉換成XTM語法的主題地圖文件、結合 metadata 紀錄連結相關的外部多媒體資源……等處理項目，最後再將剖析過程中所獲得的主題地圖文件與 metadata 紀錄儲存於資料庫。

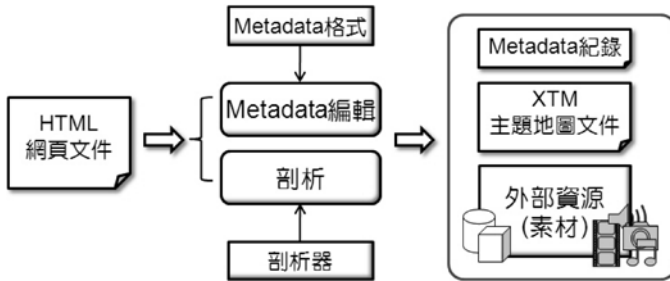


圖2 本研究設計之作業流程圖

4. 本研究設計處理模式的優點

主題地圖所使用的XTM語法即為XML，自然具備了XML的優點。因此採用XTM將HTML主要元素予以解構，並描述與標示成為個別的主題，再應用XTM語法之<association>與<occurrence>元素保存原有網頁資源之間超連結與參照的關聯性，將網頁資源轉換成具備結構化的文件結構，即可解決現有網頁保存所產生後續利用的問題。因此，本研究提出的處理模式能夠包含如下所述的優點：



- (1)具備文件長久保存的可靠性。
- (2)提供資料再製與重組的方便性。
- (3)實現文件內容採擷 (mining) 的可用性。
- (4)支援系統之間的通用性。
- (5)解決HTML文件缺乏的結構性。

五、研究實施

本研究在網頁解構的設計分析，將分三大部分說明：HTML 語法分析與解構、主題地圖建構，與流程說明與實證檢驗。

(一)HTML 語法分析與解構

呈現於使用者端瀏覽器上的單一網頁，文件格式多以HTML為主要語法。一個完整的HTML網頁由標題、段落、清單、表格，及內嵌的各種物件所組成，邏輯上這些物件稱為元素 (element)，HTML使用標籤 (tag) 來分割並描述這些元素。如圖3所示，實際上整個HTML網頁就是由元素與標籤組成 (註22)，其內容以<HTML>標籤宣告整個文件的範圍，並表達文件使用之語法為HTML。文件的結構包含兩個部分：提供應用軟體，如瀏覽器、搜尋引擎辨識資訊的<HEAD>標籤；與提供網頁呈現之實際內容的<BODY>標籤(註23)。

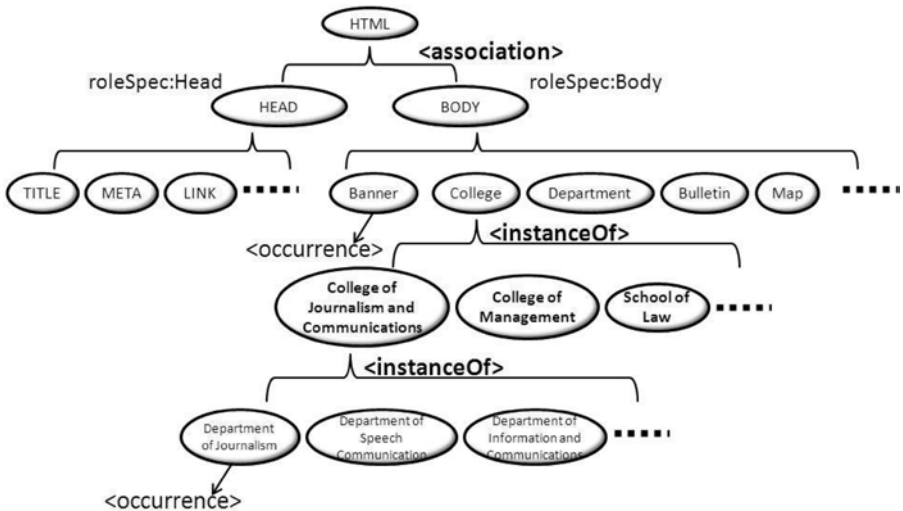


圖3 使用主題地圖描述語法解構網頁範例

資料來源：本研究整理

因此，初步可先將<HEAD>與<BODY>兩個元素所標示的文件區域，視為個別topic。若將每個網頁資源視為一個主題地圖，則各個主題地圖便可以通過XTM的<mergeMap>標籤合併為一個主題地圖。因此，若網頁內容的設計是

使用<FRAME>組合多個HTML文件成為一個網頁，便可使用<mergeMap>將個別網頁所轉換的主題地圖合併成一個完整代表該網頁資源的主題地圖。

1. <HEAD>元素內容定義網頁文件的表頭(head)資訊，用於提供網頁相關的宣告，包括文件標題的<TITLE>元素、解決相對路徑的<BASE>元素及其他如<META>等用於提供搜尋引擎處理的輔助資訊(註24)。

2. <BODY>元素包含了一個網頁資源的實際內容，依據用途區分，其內部的資訊可分為描述、文字資料、超連結、內嵌物件四個主要元素類型，以下就此四部分，探討各元素轉換成為XTM的標示方式：

(1)描述：HTML描述的元素多用於處理資料的呈現，如段落、換行、字型等。一般性元素可以轉換成獨立的主題，具備屬性的元素可以使用XTM<variant>之<parameters>標籤逐一記錄。而連續重複出現之元素，其唯一性識別值(id)則可依據原標籤名稱再另加序號識別。

(2)文字資料：呈現於網頁內容的文字資料。本研究將內容所代表的資源，依據所屬的類型以RDF標準加以描述，再透過主題地圖將這些資源之間的關聯性串連起來。

(3)超連結：<a>、、<embed>等具備href或src屬性的標籤，均是透過超連結方式提供連結至其他網頁資源，或將其他多媒體物件內嵌置網頁內的功能，XTM均使用xlink語法提供網頁內、外超連結的方式。

(4)內嵌物件：如Java Applet、Plug in軟體(如Flash)等程式或物件。於儲存管理上，可將此一類視為多媒體物件予以各別儲存。此外，如樣式宣告(style sheets)、手稿語言(scripting)等混合於網頁內容的元素，則如同HTML標籤一般轉換成主題。

(二)主題地圖建構

本研究以世新大學英文網站(<http://english.shu.edu.tw/>)作為測試網頁，嘗試以自動化方式解構網頁內容之區塊，如圖4所示，解構的方式與瀏覽器剖析HTML的程序相同，將HTML內所標示的各元素或標籤辨識並解析完成，然後再逐一將各區塊內容之HTML語法，轉換成為XTM語法。

網頁中左邊的選項，於網頁文件中的HTML，以學院資料(Academics)與行政單位(Administrator)為例，其HTML語法如表2所示，連結至Java Server Pages(JSP)網頁互動程式np.asp，並依據傳遞參數ctNode的值決定產生的動態網頁內容。

將網頁內容依據設計之標籤對映(mapping)與流程(processing)，轉換之結果包括網頁內容的<head>元素，以及學院資訊與行政單位網頁轉換成以XTM語法描述的主題地圖檔案之部分內容，請參見文末之附件。



圖4 網頁解構成各別區塊

表2 選項內容依據傳遞參數產生動態網頁內容

```

<td class="leftbg">
<h3> <a href="np.asp?ctNode=6248&amp;mp=1015" title="Introduction"> Introduction </a>
</h3>
<h3> <a href="np.asp?ctNode=6252&amp;mp=1015" title="Academics"> Academics </a>
</h3>
.....
</td>

```

(三)流程說明與實證檢驗

一個網頁資源可以視為相關資訊集合的單位，網頁資源透過超連結 (hyper-link) 連結其他相關的網頁資源。當使用主題地圖描述網頁內涵時，不僅可作為網頁資源所包含之資訊的指引，同時還可以反映網頁資源的結構。如圖5所示，使用者不僅可藉由 metadata 的描述提供檢索找尋所需的網頁資源，亦可藉由主題地圖的結構性，提供資源之間關係的連結與重組。提供使用者需要查找某一項資源內容時，不必在大量的資料庫內毫無方向的尋找，只要在主題地圖中通過其提供的連結，就可很快找到相關的主題及與之相關的事件。

經由分析 HTML 文件內容的結構與其語法特性，如圖5所示的解構示意圖，提供典藏應用程式將解構之各組成元件，依據其 HTML 語法與內涵，對應轉換成為 XTM 的語法，逐一將 web page 內容轉換成主題地圖文件。

本研究採用 TM4J (Topic Maps For Java) 計畫 (註25) 所發展的開放程式碼工具驗證，將圖6所示的網頁範例轉換的主題地圖文件視覺化的呈現來檢視轉置結果與資訊內涵。TM4J 具備視覺化環境顯示主題地圖的介面，除了提供互動功能之外，主題的概念也更能夠清楚且完整地呈現，以輔助使用者能夠經由文件中任何一個主題了解其內容的概念，縱觀單一主題相關的所有其他主題，

並能追溯主題之間先後發生的事件。本研究即是藉由此視覺化的設計介面，達到檢視主題之間的關係與事件之間關聯正確性的目的。

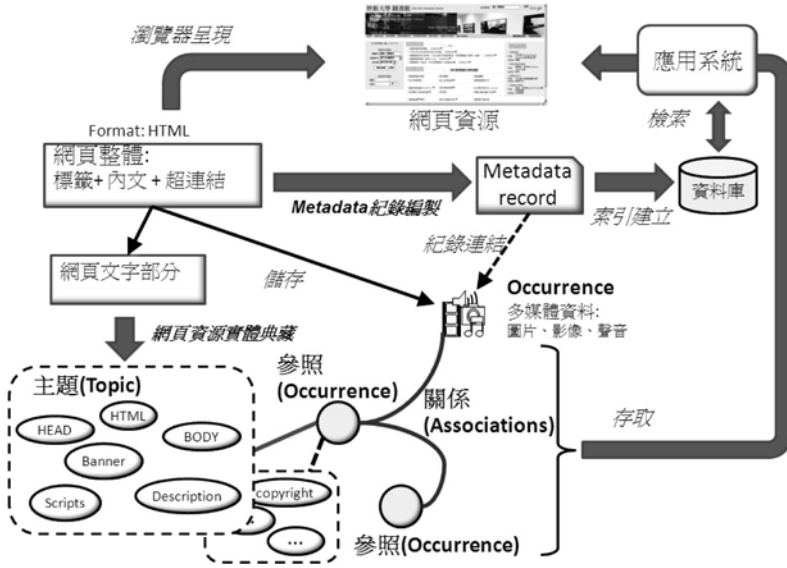


圖5 本研究設計之系統整體功能流程圖

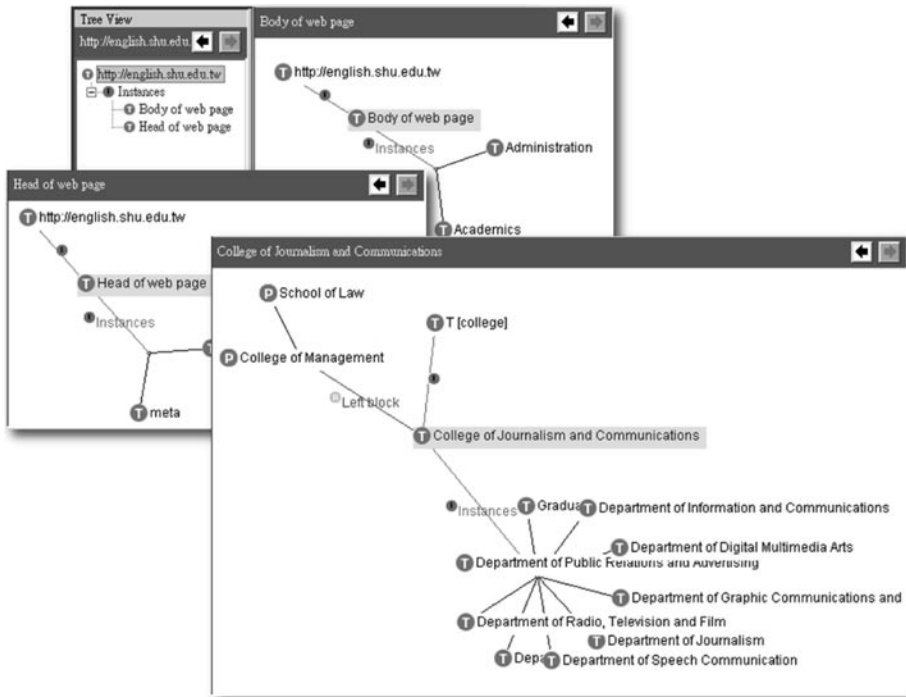


圖6 使用TM4J工具軟體檢視網頁轉換結果的主題地圖文件內容

基本上，典藏應用程式除了應能提供將網頁資源轉換成以 XTM 語言描述的主題地圖文件，亦必須具備爾後將主題地圖文件還原回原始網頁資源的功能。由於將網頁資源由非結構化的 HTML 轉成結構化的 XTM 語法，如此還可兼顧爾後還原時能依據網頁資源描述語法的改變，轉換成符合未來網頁標示語言的可能性，提供長久保存的實質目的。

六、結 論

HTML 自 1991 年問世以來已有多次的版本修訂，加上 1995 年之後微軟與網景 (Netscape) 瀏覽器的競爭，使得不同瀏覽器廠商使用各自的網頁標準，此外，W3C 也因應網際網路與資訊傳播應用的需求，不斷推出新版本的 HTML 修訂。因此，研究網頁資源的保存必須能夠提供後續轉置、應用與萃取網頁資源內部的資訊，才能達到其長久保存的目的。而 HTML 不具備結構化的特性，使得單純儲存既有的網頁資源，不僅無法提供自動化萃取資訊的功能，也可能會因 HTML 未來的改變而導致無法使用現有保存之網頁資源，雖然透過現有軟體系統的保存、模擬、更新等方式，可解決資訊改版所造成無法使用的問題，但成本與時效卻是另外衍生的問題。

現有網頁版面的資訊除以 XML 搭配 CSS 之外，均是以 HTML 語法的標籤透過瀏覽器剖析而呈現，所以使用 XML 來封裝或保存 HTML 網頁，只要將 HTML 標籤與內文能保存於 XTM 文件內，即可確保將 XTM 文件還原原本之 HTML 文件，如此就可再依循原瀏覽器剖析之方式，還原原有之呈現樣式。因此，本研究依據標準化的方式，採用依據 XML 所制訂的主題地圖語法，將 HTML 文件內的標籤與內文予以結構化描述，提供了網頁保存的另一種解決方案。

研究也發現，由於 HTML 非結構化與缺乏文件組織規範的特性，造成許多網頁編排的散亂與結合多種規範的混雜，所需花費最大的負荷並非實作的系統，而是建立特定網站內容各 HTML 標籤的特徵參數 (profile)，以提供轉置程式能夠正確解讀該標籤的使用內涵。因此，除了依靠資訊技術的開發，網頁性質的分析亦是需要搭配的作業要項，兩者處理得宜便能夠確實提供保存現有網頁資源完整的內容與呈現樣式，亦可提供自動化的批次資訊處理要求，解決長久保存所必須考量的資訊重組問題。

註 釋

註 1 Michele Kimpton and Jeff Ubois, "Year-by-Year: From an Archive of the Internet to an Archive on the Internet," in *Web Archiving*. Edited by Julien Masanès (Berlin: Springer, 2006), 201-212.

註2 James H. Coombs, Allen H. Renear, and Steven J. DeRose, "Markup Systems and the Future of Scholarly Text Processing," *Communications of the ACM* 30, no. 11 (November 1987): 933-947.

註3 Kal Ahmed et al., *Professional XML Meta Data*. Birmingham (UK: Wrox, 2001), 251-253.

註4 Andrew Waugh et al., "Preserving Digital Information Forever," in *Proceedings of the Fifth ACM Conference on Digital Libraries* (USA: Texas, 2000), 175-184.

註5 歐陽崇榮,「數位資訊保存之探討」, 檔案季刊 1 卷, 2 期(2002): 36-47。

註6 Gregory W. Lawrence et al., *Risk Management of Digital Information: A File Format Investigation* (Washington, DC: Council on Library and Information Resources, 2000), 12.

註7 王麗蕉,「網路資源長期保存:以多元層次描述模式建構之探討」, 教育資料與圖書館學 44 卷, 4 期(2007): 455-471。

註8 SGML SIGhyper, "A Brief History of the Development of SMDL and HyTime," <http://www.sgmlsource.com/history/hthist.htm> (accessed February 25, 2008).

註9 Steve Pepper, "Euler, Topic Maps, and Revolution," in *Proceedings of XML Europe 99 Conference* (Alexandria, VA: GCA, 1999), 2.

註10 余顯強, XML 標準與技術簡介(台北市:數位典藏訓練推廣分項計畫, 2004), 22。

註11 Tim Bray, Jean Paoli, and C. M. Sperberg-McQueen, "Extensible Markup Language (XML) 1.0," (February 1998), <http://www.w3.org/TR/1998/REC-xml-19980210> (accessed September 28, 2001).

註12 Shien-chiang Yu and Ruey-shun Chen, "An XML Framework for an Electronic Document Delivery System," *The Electronic Library* 19, no.2 (2001): 102-110.

註13 Steve Pepper and Graham Moore, "XML Topic Maps (XTM) 1.0," (2001), <http://www.topicmaps.org/xtm/1.0/> (accessed February 25, 2008).

註14 Steve Pepper, "The TAO of Topic Maps: Finding the Way in the Age of Infoglut," <http://www.gca.org/papers/xml europe2000/pdf/s11-01.pdf> (accessed November 11, 2005).

註15 Occurrence 一詞, 國內文獻有翻譯為「資源指引」、「呈現」、「事件」等用語, 本研究認為「資源指引」較能表現 Occurrence 在設計上的意義, 但依據 XTM 對於 Occurrence 的解釋在於使用其表現主題之間的相關性。以圖書資訊學的角度來探討其功能性, 本研究認為「參照」較能符合其本意, 因此將 Occurrence 翻譯為「參照」。

註16 Jack Park and Sam Hunting, *XML Topic Maps: Creating and Using Topic Maps for the Web* (Boston, MA: Addison-Wesley, 2002), 37-40; Steven R. Newcomb, Michel Biezunski, and Martin Bryan, "ISO/IEC 13250 Topic Maps: Information Technology- Document Description and Processing Languages," <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf> (accessed February 25, 2008); Michael C. Daconta, Leo J. Obrst, and Kevin T. Smith, *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management* (Indiana: Wiley, 2003), 170-176.

註17 Ramesh R. Sarukkai, *Foundations of Web Technology* (Boston, MA: Kluwer, 2002), 249.

註18 Anne Cregan, "Building Topic Maps in OWL-DL," <http://www.idealliance.org/pa>

pers/extreme/proceedings/xslfo-pdf/2005/Cregan01/EML2005Cregan01.pdf (accessed September 6, 2008).

註19 Uwe M. Borghoff et al., *Long-Term Preservation of Digital Documents* (Berlin; New York: Springer, 2005), 114-119.

註20 Lars Marius Garshol, "Living With Topic Maps and RDF," <http://www.ontopia.net/topicmaps/materials/tmrdf.html> (accessed September 6, 2008).

註21 Gerald J. Kowalski and Mark T. Maybury, *Information Storage and Retrieval Systems: Theory and Implementation* (Boston: Kluwer Academic Publishers, 2000), 34-35.

註22 在標示語言的定義中，元素是由成對的起始標籤、結束標籤再加上內容資料所組成。若起始標籤與結束標籤是同一個時，則稱為空元素。由於HTML的容錯性，當文件中有起始標籤並無對應的結束標籤（反之亦然），系統仍能正確的處理，因此，在HTML文件中常有資料內容僅是由單一標籤而非元素所組成。

註23 Richard P. Smiraglia, *Metadata: A Cataloger's Primer* (NY, Binghamton: Haworth, 2005), 8-9.

註24 Ramesh R. Sarukkai, *Foundations of Web Technology* (Boston: Kluwer, 2002), 43.

註25 Kal Ahmed, "TM4J Developer's Guide," <http://tm4j.org/tm4j/docs/devguide/> (accessed February 25, 2008).

附件 網頁轉換成以XTM語法描述 的主題地圖檔案之內容範例

1. head 元素

```

<topic id="default.html">
  <baseName>
    <baseNameString>http://english.shu.edu.tw</baseNameString>
  </baseName>
</topic>

<topic id="head">
  <baseName>
    <baseNameString>Head of web page</baseNameString>
  </baseName>
  <instanceOf><topicRef xlink:href="#default.html"/></instanceOf>
</topic>

<topic id="title">
  <instanceOf><topicRef xlink:href="#head"/></instanceOf>
  <baseName>
    <baseNameString>Welcome To Shih Hsin University</baseNameString>
  </baseName>
</topic>

<topic id="meta">
  <instanceOf><topicRef xlink:href="#head"/></instanceOf>
  <baseName>
    <baseNameString>meta</baseNameString>
    <variant>
      <parameters id="http-equiv">Content-Type</parameters>
      <parameters id="content">text/html; charset=utf-8</parameters>
    </variant>
    <variant>
      <parameters id="name">Education</parameters>
      <parameters id="content">Welcome To Shih Hsin University</parameters>
    </variant>
  </baseName>
</topic>

```

2. 學院資訊 (部分內容)

```

<association id="unit">
  <instanceOf><topicRef xlink:href="#Academics"/></instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#session1"/>
    </roleSpec>
    <topicRef xlink:href="#smgmt"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#session1"/>
    </roleSpec>
    <topicRef xlink:href="#jc"/>
  </member>
</member>

```

```

    <roleSpec>
      <topicRef xlink:href="#session1"/>
    </roleSpec>
    <topicRef xlink:href="#law"/>
  </member>
</association>

```

3. 行政單位 (部分內容)

```

<association id="dept">
  <instanceOf><topicRef xlink:href="#department"/></instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#session2"/>
    </roleSpec>
    <topicRef xlink:href="#oga"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#session2"/>
    </roleSpec>
    <topicRef xlink:href="#osa"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#session2"/>
    </roleSpec>
    <topicRef xlink:href="#ord"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#session2"/>
    </roleSpec>
    <topicRef xlink:href="#oga"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#session2"/>
    </roleSpec>
    <topicRef xlink:href="#library"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#session2"/>
    </roleSpec>
    <topicRef xlink:href="#po"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#session2"/>
    </roleSpec>
    <topicRef xlink:href="#oa"/>
  </member>
</association>

```

Discussion on Web Archives Using Topic Maps

Shien-Chiang Yu

Associate Professor and Director of Library
Department of Information & Communications
Shih-Hsin University, Taipei, Taiwan
E-mail: ysc@cc.shu.edu.tw

Abstract

Web resources have already become the present important information channel, its pluralism abundant content to offer people for information retrieval and academia to study the needs of consulting. But web resources not only grow up quickly, but also disappear fast. Web resources will probably be unable to be utilized again because of factors such as server shutdown, revision, etc. at any time. Web resources are among the most important human societal information, including visual design art, style, and included various kinds of resources in these web pages. Due to the fragile characteristics of digital information, this study aims to address strategic, methodological and practical issues in archiving and managing digital preservation. Because the webpage may be described with the HTML edition and version, the support of the browser and long-term preservation must be considered, including standardization, format, data utilization, mining and restructured employment in the future. Based on these requirements, the main purpose of this study is analysis in probing the solution for long-term preservation of web resources; adopting the standardized method; and analyzing composition components and characteristics of web resources according to the perspective of the data structure. This study adopted the migration method to reorganize the structure but keep primitive information of web resources according to the Topic Maps metadata, in order to offer web resources integrity, management and application effectively with the whole information content.

Keywords: *Web resources; Long term preservation; Topic maps; Webpage migration*

SUMMARY

Introduction

The amounts of web resources have increased rapidly in recent years. Libraries and the information science professionals turn to cooperation in long-term preservation of digital resources to pass down human civilization, and many national-class libraries and archives around the world have launched web archives projects to ensure long-term preservation and provide continued access to digital resources¹.

¹ Michele Kimpton and Jeff Ubois, "Year-by-Year: From an Archive of the Internet to an Archive on the Internet", in *Web Archiving*. Edited by Julien Masanès. (Berlin: Springer, 2006), 201-212.

Observing the contents of the web resources, it could be noticed that most of them are written in unstructured html texts. Some of those html texts are static, and others are driven by computer programs. Therefore, to record the content of the web resource, is absolutely vital. The holistic plan of the description and management of the metadata format is necessary in order to preserve and to use the web resources in the future. Individual structures of web have to be analyzed, to deconstruct the attributes of the web content. The analysis of the web resource not only helps preserve the original relationship of the content, but also reconnects different web pages by the demand of the users.

HTML is a kind of procedural markup language and it is impossible to infer the meaning of the content. Because of that, it can't be used for exact searching, automatic documents processing and content mining.² This paper focuses on the execution project of the long-term preservation of the web resource. Topic Maps are base on basic model of the Semantic Web and function of index, can offer structure of the knowledge expression.³ So, this study is attempting to solve the restructuring of web pages with the Topic Maps, not only keep the original relation of hyperlinks, can also recombine the relevance between web pages according to the archiving institution's demand.

Topic Maps

1. The origin and development of Topic Maps

To enrich Standard Generalized Markup Language (SGML, ISO/IEC 8879-1986) with functions of multimedia and hyperlinks, Goldfarb and Stev Newcomb tried to design an architectural form which could make hyperlinks available to multimedia and documents at any time. That caused Hypermedia/ Time-based Structuring Language (HyTime) to be introduced, and HyTime became an ISO and International Electrotechnical Commission (IEC) joint standard upon publication as ISO/IEC 10744 in 1992.⁴ Being inherited from SGML, the syntax rule of Hytime is very complicated, and that intrigued the Graphic Communication Association Research Institute (now known as IDEAlliance) in an activity called Conventions for the Application of HyTime (CAPH) that proposed a revised clause for identifying information objects that share a common topic.⁵

² James H. Coombs, Allen H. Renear, and Steven J. DeRose, "Markup Systems and the Future of Scholarly Text Processing," *Communications of the ACM* 30, no. 11 (November 1987): 933-947.

³ Kal Ahmed et al., eds., *Professional XML Meta Data*. (UK: Wrox, 2001), 251-253.

⁴ SGML SIGhyper, "A Brief History of the Development of SMDL and HyTime," <http://www.sgml-source.com/history/hthist.htm> (accessed February 25, 2008).

⁵ Steve Pepper, "Euler, Topic Maps, and Revolution," (Alexandria, VA: GCA, 1999), 2.

The solutions developed were called “Topic Navigation Maps” which became an ISO/IEC standard as ISO/IEC 13250 in December of 1999. Topic Navigation Maps adopts HyTime as the definition syntax, and it is a kind of Document Type Definition (DTD) of SGML. It can define subsets of various types of fields and rename their names and attributes. According to that, it could describe various kinds of information concepts as topics which possess their own name, attribute, resource guide, etc. Besides, it could define the relations that topics bear to one another.

Moreover, to break through the restriction for application of SGML and HTML, World Wide Web Consortium (W3C) developed a new generation markup language-eXtensible Markup Language (XML)⁶ which could be used in web environments and can define interchange formats of structured data files. Not composed by specific tags like HTML and supporting language neutral as well as platform neutral, XML allows users to define markup languages needed by themselves and could be used in various areas widely. So TopicMaps.Org established in 2000 adopted XML syntax to develop XML Topic Maps (XTM) 1.0 Specification in 2001.⁷ In 2002, ISO/IEC 13250: Topic Maps containing 2 syntax structures which are HyTime and XTM were approved.⁸

2. Elements of Topic Maps

The XTM standard identifies the key of Topic Maps. The key concepts summed up as the “TAO” of Topic Maps, comes from the initials of the constructs for representing find aids: topics, associations, occurrences, subject descriptor, and scope:⁹

(1) Topics

A topic is a representation of the subject; according the XTM standard, it acts as a resource that is a proxy for the subject. Each topic is implicitly an instance of a topic type-that is, the class of the topic.

⁶ Tim Bray, Jean Paoli, and C. M. Sperberg-McQueen, “Extensible Markup Language (XML) 1.0,” <http://www.w3.org/TR/1998/REC-xml-19980210> (accessed September 28, 2001).

⁷ Steve Pepper and Graham Moore, “XML Topic Maps (XTM) 1.0,” <http://www.topicmaps.org/xtm/1.0/> (accessed February 25, 2008).

⁸ Steve Pepper, “The TAO of Topic Maps: Finding the Way in the Age of Infoglut,” <http://www.gca.org/papers/xml europe2000/pdf/s11-01.pdf> (accessed November 11, 2005).

⁹ Jack Park and Sam Hunting, *XML Topic Maps: Creating and Using Topic Maps for the Web* (Boston, MA: Addison-Wesley, 2002), 37-40; Steven R. Newcomb, Michel Biezunski, and Martin Bryan, “ISO/IEC 13250 Topic Maps: Information Technology: Document Description and Processing Languages,” <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf> (accessed February 25, 2008); Michael C. Daconta, Leo J. Obrst, and Kevin T. Smith, *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management* (Indiana: Wiley, 2003), 170-176.

(2) Occurrence

An occurrence is a resource specifying some information about a topic. The resource is either addressable (using a URI) or has a data value specified inline.

(3) Association

An association is the relationship between (one or more) topics and it is represented as <association> in XTM Structure.

(4) Subject Indicator

Originally named as subject descriptor in ISO/IEC 13250, a subject indicator is a resource that is intended by the topic map Page author to provide a positive, unambiguous indication of the identity of a subject.

(5) Scope

Scope is a special topic and it defines a group or a specific range of related topics. The function of Scope is similar to the name space: The base name of a topic should be the only one in a certain scope.

3. Design the Web Archive Model

To sum up, the nature of Topic Maps is very simple. The basic components, topics, are clustered via the description of topic type. Then event-related topics are gathered together via the element Occurrence, and the semantics between related events are linked by the element Association. Semantics of related events are linked through association. Besides, users could take advantage of the element Scope to limit the range of names, resources guides and relations to form the original entirety web resource.

Being used for describing the webpage contents, Topic Maps could be the

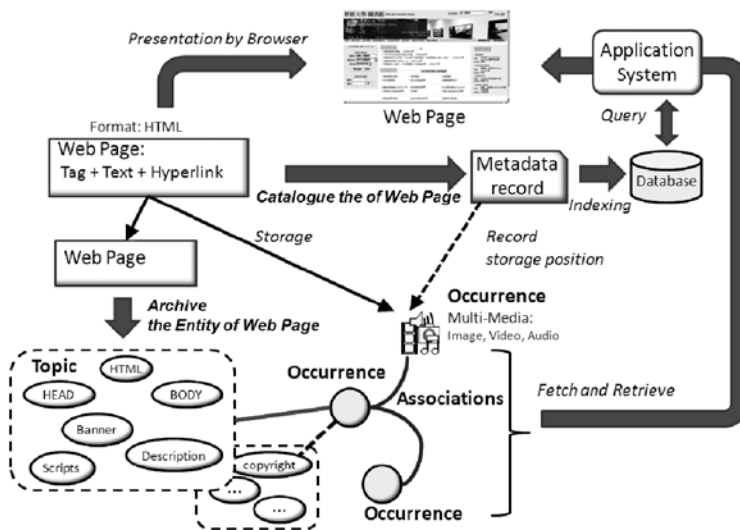


Figure 1 The Major Process Flow of This Study

guide to them, and it could also reflect the structure of web resources. As Fig. 1 shows, users can search for web resources needed through description of meta-data. They could find out links between web resources and recombine them additionally via the structure displayed by Topic Maps. When searching for a specific web resource, users don't have to search in a large database. With only the links provided by Topic Maps, users could find out related topics and events.

Considering the application would be restricted by archived web resources themselves in the future, the authors try to make use of XTM to deconstruct HTML elements of web pages, describe and markup them, then apply XTM association and occurrence to keep original links and related references of web resources. Several advantages of that are listed below:

- (1) Possess the dependability that the file is kept for a long time (stability);
- (2) Offer the convenience for rebuilds and recombines of the materials;
- (3) Realize the usability of file content mining;
- (4) Support the interoperation between the systems;
- (5) Solve the problems due to structure-less HTML document.

4. Design practice

(1) Deconstruction web page

Single web pages shown on a user's browser usually are HTML documents. As Fig. 2 shows, its content is contained within the range which the <HTML> Tag declares. The HTML document could be separated in two parts: the <HEAD> Tag providing identification information for application software, and the <BODY> Tag providing content represented by the webpage.¹⁰

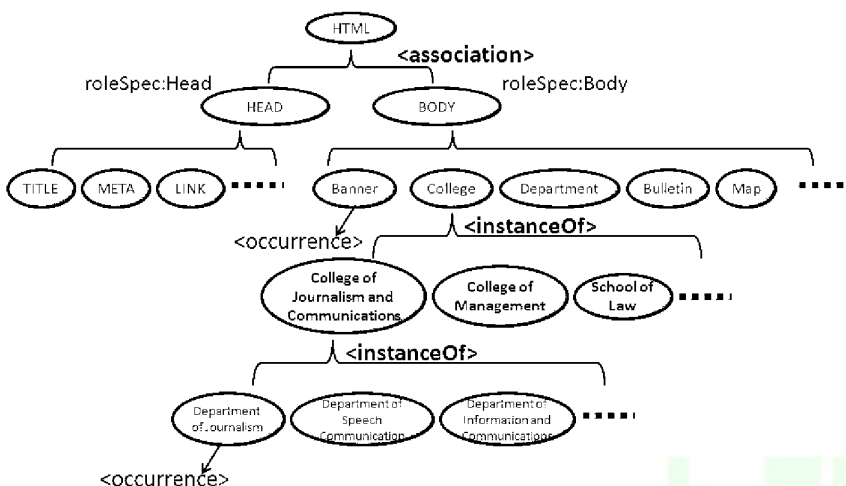


Figure 2 Deconstruction Web Page Using Topic Maps Example

¹⁰ Richard P. Smiraglia, *Metadata: A Cataloger's Primer* (New York: Haworth, 2005), 8-9.

Firstly, the two document ranges declared by the <HEAD> and <BODY> are regarded as two different topics. If a single web resource is regarded as a Topic Map, each Topic Map could be combined via the <mergeMap> Tag of XTM. If a webpage is a combination of several web pages that associate with each other through the <FRAME> Tag, the Topic Maps of single webpages could be combined via the <mergeMap> Tag too.

The <BODY> Tag declares the actual content of a webpage resource. Distinguishing webpage elements by their function, there are four kinds of information contained in the <BODY> Tag: tag, text, hyperlink, and embedded object. The way in which HTML elements converted into XTM is described below:

a. Tag: HTML Tags are usually used for dealing with the representation of data. Tags could be converted to independent Topics generally, and Tags with attributes could be recorded one by one using the <parameters> Tag of XTM <variant> Tag. Besides, the id value of repeated Tags could be identified by adding serial numbers to the original Tag name.

b. Text: Like CSS, Scripts (such as JavaScript, VBScript) and text data shown on webpage context, etc., this study describes the resources by their type using RDF standards, then links the relationship between resources through Topic Maps.

c. Hyperlink: Tags with the attribute *href* or *src* like <a>, , <embed>, etc. provide links to other web resources by hyperlinks, or make other multimedia objects embedded in web pages. XTM make hyperlinks available by using the syntax XLINK.

d. Embedded object: they are programs or objects like Java Applet, Plug-in software (such as Flash), etc., and can be regarded as multimedia objects to store separately for the purpose of preservation management.

Fig. 3 illustrates the archive application program to map the HTML syntax of each component to XTM syntax, and convert the content of web pages to a Topic Maps document. Additionally, this study used open-source tools developed by the Topic Maps for the Java program (TM4J)¹¹ to parse the webpage shown in Fig. 3 and get the converted Topic Maps document shown in Fig. 4. By representing the visualized environment of Topic Maps, TM4J can provide interactive functions of interface and display topic concepts clearly and completely to help users understand the topic concept between web resources. This study is trying to achieve the purpose of reviewing the correctness of document conversion through visualized interface.

¹¹ Kal Ahmed, "TM4J Developer's Guide," <http://tm4j.org/tm4j/docs/devguide/> (accessed February 25, 2008).

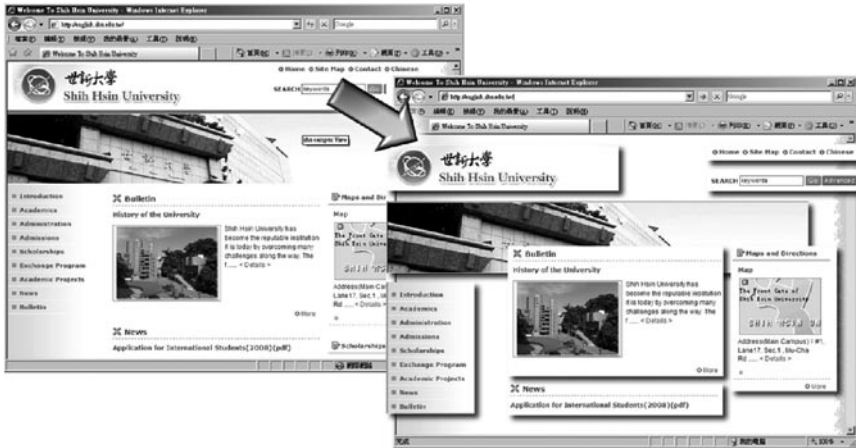


Figure 3 Deconstruction Web Page Example

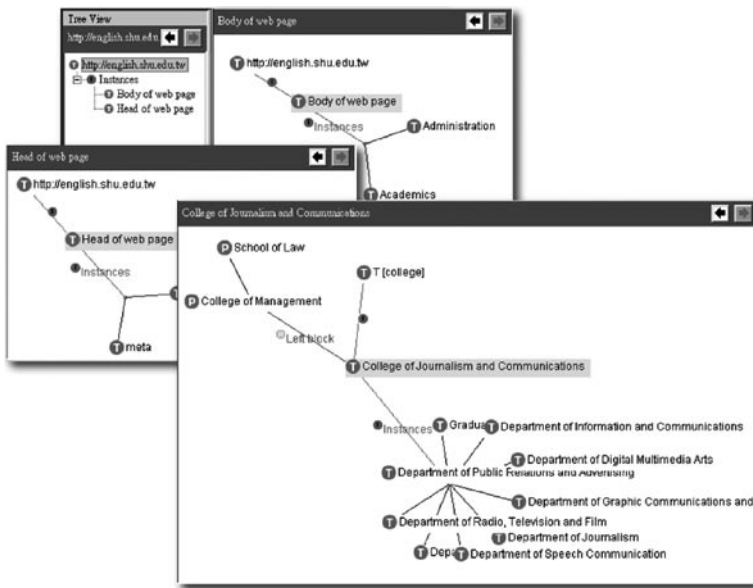
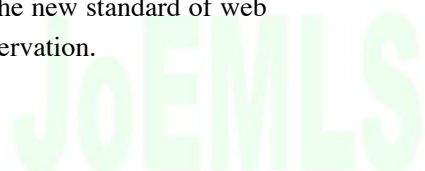


Figure 4 Apply TM4J Tools to Review the Converted Topic Maps Document

5. Conclusion

Archive application programs offer the function of converting web pages to Topic Maps documents, it also permits Topic Maps documents being restored to web pages. Moreover, considering, the structured XTM document converted from the unstructured HTML document, this could meet the new standard of web markup language to achieve the purpose of long-term preservation.



ROMANIZED & TRANSLATED NOTES FOR ORIGINAL TEXT

註1 Michele Kimpton and Jeff Ubois, “Year-by-Year: From an Archive of the Internet to an Archive on the Internet,” in *Web Archiving*. Edited by Julien Masanès (Berlin: Springer, 2006), 201-212.

註2 James H. Coombs, Allen H. Renear, and Steven J. DeRose, “Markup Systems and the Future of Scholarly Text Processing,” *Communications of the ACM* 30, no. 11 (November 1987): 933-947.

註3 Kal Ahmed et al., *Professional XML Meta Data*. Birmingham (UK: Wrox, 2001), 251-253.

註4 Andrew Waugh et al., “Preserving Digital Information Forever,” in *Proceedings of the Fifth ACM Conference on Digital Libraries* (USA: Texas, 2000), 175-184.

註5 歐陽崇榮[James C. Ouyang], 「數位資訊保存之探討」[“Shuwei Zixun Baocun zhi Tantaotao”], 檔案季刊[*Dangan Jikan*] 1卷, 2期[1, no.2] (2002): 36-47。

註6 Gregory W. Lawrence et al., *Risk Management of Digital Information: A File Format Investigation* (Washington, DC: Council on Library and Information Resources, 2000), 12.

註7 王麗蕉[Li-Chiao Wang], 「網路資源長期保存：以多元層次描述模式建構之探討」[“Wanglu Ziyuan Zhangqi Baocun: Yi Duoyuan Cengci Miaoshu Moshi Jiangou zhi Tantaotao”], 教育資料與圖書館學[*Journal of Educational Media and Library Sciences*] 44卷, 4期[44, no.4] (2007): 455-471。

註8 SGML SIGhyper, “A Brief History of the Development of SMDL and HyTime,” <http://www.sgmlsource.com/history/hthist.htm> (accessed February 25, 2008).

註9 Steve Pepper, “Euler, Topic Maps, and Revolution,” in *Proceedings of XML Europe 99 Conference* (Alexandria, VA: GCA, 1999), 2.

註10 余顯強[Shien-Chiang Yu], *XML標準與技術簡介*[*XML Biao Zhun Yu Jishu Jianjie*](台北市：數位典藏訓練推廣分項計畫, 2004) [(Taipei: Digital Archives Training and Promotion Division Program, 2004)], 22。

註11 Tim Bray, Jean Paoli, and C. M. Sperberg-McQueen, “Extensible Markup Language (XML) 1.0,” (February 1998), <http://www.w3.org/TR/1998/REC-xml-19980210> (accessed September 28, 2001).

註12 Shien-chiang Yu and Ruey-shun Chen, “An XML Framework for an Electronic Document Delivery System,” *The Electronic Library* 19, no.2 (2001): 102-110.

註13 Steve Pepper and Graham Moore, “XML Topic Maps (XTM) 1.0,” (2001), <http://www.topicmaps.org/xtm/1.0/> (accessed February 25, 2008).

註14 Steve Pepper, “The TAO of Topic Maps: Finding the Way in the Age of Infoglut,” <http://www.gca.org/papers/xml europe2000/pdf/s11-01.pdf> (accessed November 11, 2005).

註16 Jack Park and Sam Hunting, *XML Topic Maps: Creating and Using Topic Maps for the Web* (Boston, MA: Addison-Wesley, 2002), 37-40; Steven R. Newcomb, Michel Biezunski, and Martin Bryan, “ISO/IEC 13250 Topic Maps: Information Technology- Document Description and Processing Languages,” <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf> (accessed February 25, 2008); Michael C. Daconta, Leo J. Obrst, and Kevin T. Smith, *The Semantic Web: a Guide to the Future of XML, Web Services, and Knowledge*

Management (Indiana: Wiley, 2003), 170-176.

註17 Ramesh R. Sarukkai, *Foundations of Web Technology* (Boston, MA: Kluwer, 2002), 249.

註18 Anne Cregan, "Building Topic Maps in OWL-DL," <http://www.idealliance.org/papers/extreme/proceedings/xslfo-pdf/2005/Cregan01/EML2005Cregan01.pdf> (accessed September 6, 2008).

註19 Uwe M. Borghoff et al., *Long-Term Preservation of Digital Documents* (Berlin; New York: Springer, 2005), 114-119.

註20 Lars Marius Garshol, "Living With Topic Maps and RDF," <http://www.ontopia.net/topicmaps/materials/tmrd.html> (accessed September 6, 2008).

註21 Gerald J. Kowalski and Mark T. Maybury, *Information Storage and Retrieval Systems: Theory and Implementation* (Boston: Kluwer Academic Publishers, 2000), 34-35.

註23 Richard P. Smiraglia, *Metadata: A Cataloger's Primer* (NY, Binghamton: Haworth, 2005), 8-9.

註24 Ramesh R. Sarukkai, *Foundations of Web Technology* (Boston: Kluwer, 2002), 43.

註25 Kal Ahmed, "TM4J Developer's Guide," <http://tm4j.org/tm4j/docs/devguide/> (accessed February 25, 2008).