



「XML 技術與圖書館應用研習班」研習心得

宋瓊玲

壹、前言

邁入數位時代，網路資源百花齊放，圖書館此刻面臨的最大挑戰，即是如何有效率的將各類型資源進行蒐集、整理及推廣應用。網路資源多且廣，如何將其加以整理、交換，以利資源共享，急需一標準化的技術來進行整合。由 World Wide Web Consortium(W3C)訂定的 XML 標準，是繼 HTML 之後，新一代 Web 的整合技術，為網路帶來第二波革命性的改變，促使網路從資訊處理階段跨越到知識管理階段，並在電子出版、電子商務、數位博物館/圖書館、電子資料交換等領域展現其強大的應用潛能。

由淡江大學資圖系承辦「XML 技術與圖書館應用研習班」，八月十六日至二十日為期五天，假淡江大學文學院舉辦。課程安排理論與實務並重，配合一人一機的實作課程，使學員們對 XML 的語法及其相關技術有更深刻的瞭解。研習班課程由林信成老師、余顯強老師及葉慶隆老師授課，課程內容含括以下主題：XML 及 DTD 的介紹及實作、XML 的應用層面、XML Schema 及其他相關技術、XML/Metadata 的管理系統設計、數位圖書館的標準與版權管理應用、XML 與 Semantic Web(語意網)的實現等。

貳、課程歸納整理

XML 語言是用於設計資料格式和結構的規則和方法，在不同電腦及應用程式間，利用 XML 可互相轉換。建置一數位圖書館，有大量資料需要轉換，XML 有效表達網路上各種知識，為資料的交換和處理提供新的機制。以下就上課內容，將 XML、相關技術及應用，開放式數位圖書館標準及數位版權管理及 XML 的未來發展進行整理歸納。

一、XML 的介紹

(一)XML 的特性

XML 為可擴展標記語言(Extension Markup Language)，開發的用於網路環境下網頁設計和資料交換、管理的新技術，並已成為標準，具有很好的應用和發展前景。XML 為 SGML 的精簡子集(Subset)，具有結構性、可擴展性、自我描述性、資料和樣式分離(指資料內容外觀分離)原則及整合性。

(二)XML 與 HTML

XML、HTML 均是從 SGML 演變而來而來，但是 HTML 是 SGML 的一個應用語言，而 XML 卻是 SGML 的一個精簡子集(subset)。XML 具備了 SGML 所沒有的簡易性與靈活性；又有著 HTML 所欠缺的擴展性與結構性。以現今的 Web 而言，HTML 仍是發行電子文件的標準規格，但因 HTML 著重於版面編排與外觀格式，對於文件結構的規範及內容語意的描述則乏善可陳；更因 HTML 不具備可擴展性，所以使用 HTML 著錄 Metadata 的成效不彰，XML 的誕生，正好提供了一個基礎平台。

(三)XML 與 HTML 的差異

1. XML 文件的作者可以自訂標籤 (Tags) 和屬性 (Attribute) , HTML 則否。
2. XML 是屬於一般用途 (General Purpose) 的標誌語言, 而 HTML 則是一種特殊用途 (Special Purpose) 的標誌語言。換言之, XML 是一種元語言 (Meta-Language) , 可以用以生成其他語言, HTML 則否。
3. XML 著重於文件的結構, 而 HTML 則擅長於文件的表現。
4. XML 文件的作者可以選擇性的利用 DTD 或 XML 綱要 (XML Schema) 來確認文件的有效性, HTML 則否。

二、XML 相關技術

XML 相關技術, 諸如 DTD、XML Schema、XML Namespaces、RDF、XLink、DOM、CSS、XSL 等。以下就 XML 的特性來分項說明 XML 的相關技術:

(一) 結構性而言: 如 DTD、XML Schema

「文件類型定義」(Document Type Definition, 簡稱 DTD) 和「XML 綱要」(XML Schema) 是 XML 處理器據以確定文件有效性的兩大重要機制。

(二) 資料和樣式分離而言: 如 CSS、XSL/XSLT

XML 強調的是如何以適當的結構來組織資料, 對於外在的表現則必須透過其他顯示機制才能達成, 這就是 XML 文件的資料、樣式 (即內在、外貌) 分離原則。這使得文件作者只要專注於內容的撰寫, 而將顯示資訊的任務交由版面設計者或使用者, 依據不同的需求來展現。如此一來, 同一份文件或資料, 將可在不同的場合呈現出不同的風貌。

1. 以 CSS 層級樣式表顯示文件
2. 以 XSL 可擴展樣式語言顯示文件
3. 以程式剖析方式顯示文件

(三) 文件整合、傳遞與操縱而言: 如 XML 名稱領域、XLink

透過不同的協定轉換, 各種不同格式的資料可以轉成 XML 格式, 使得 XML 在文件整合 (Document Integration) 方面, 扮演了一個通用集成器 (Universal Hub) 的角色 (註 19), 而 XML 的名稱領域、XLink 等正是文件整合不可或缺的重要技術

1. 整合功臣: 名稱領域

資源整合時必須注意元素名稱的唯一性問題, XML 的名稱領域 (XML Namespaces) 便是為了解決元素的命名衝突所發展出來的。

根據 W3C 的定義: 一個 XML Namespace 是使用 URI (Universal Resource Identifier, 一致性資源識別碼) 識別的名稱集, 用於 XML 文件中作為元素型別和屬性名稱。由於 URI 是唯一的, 所以 Namespace 可以保證文件中所使用的名稱是獨一無二的。

2. 連結機制: XLink

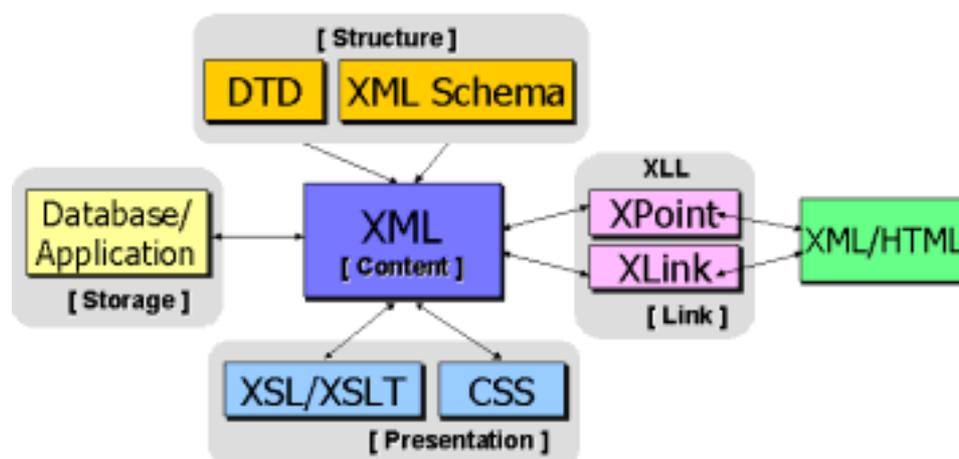
XML 便肩負了增強連結功能的重任, 不過在 XML 規格中並未定義任何的連結方法來改變 HTML 的現狀, 而是以原有 HTML 的連結概念為出發點, 延伸出更強大的方式來處理連結資訊, 此一延伸機制稱為「XML 連結語言」(XML Linking Language), 簡稱 XLink。XLink 目前仍是 W3C 的一個工作草案 (Working Draft), 尚未成為建議標準。為了與 HTML 的簡易連結相容, 並提供更強的延伸功能供人們使用, 在制訂 XLink 規格時, 簡易性、可讀

性、延伸性、實用性等，便成了首要考量的幾個重要的設計目標。

3. 文件剖析:DOM

XML 文件是一種基於文字模式的開放規格，但若要對 XML 文件的資料內容，進行其他的加值處理或操縱，需透 W3C 制訂的介面標準「文件物件模型」(Document Object Model)，簡稱 DOM 為之。W3C 所倡議的 DOM，是一種語言獨立的規格，可以在不同的作業系統中，使用任何程式語言加以實現。

以下將 XML 相關延伸技術用圖來展現（資料來源：余顯強老師上課講義）



三、XML 的應用

XML 因其具有結構性、可擴展性、自我描述性及整合性等特性，應用層面廣泛，可用來進行資料庫的交換、使用者端的文件處理、使用者可用不同樣式觀看資料、資訊過濾等，就數位圖書館相關計畫或電子商務等網際網路應用而言，XML 已逐漸成為網路界及資料庫界積極推廣的語言。以下根據課程的內容及相關文章的論述，將 XML 應用的方向加以敘述：

(一) XML 與電子出版

從電子出版的角度來看，目前 Web 上為數眾多的網站及互相糾結的 HTML 文件，在寫作之初由於缺乏有效的資料著錄，造成資料雜亂、不易檢索的問題，因此亟需資訊檢索技術與方法來解決；此外，除了致力於技術的發展，在寫作及出版的過程中加強 Metadata 的著錄，亦不失為另一個解決網路上電子文件亂象並提昇檢索精確率的有效方案。XML 與生俱來的結構化及自我描述特性，能有效的組織資料、描述資料，以進行語意層次的有意義檢索；而其無限擴展的能力，更使其跳脫既定的框架，而在版面編排、多媒體出版、文件共享等方面發展出獨特的應用，在在使 XML 成為網路出版內、外兼顧的有效方案。

(二) XML 與電子商務

從電子商務的角度觀之，不同商業系統採用相同的 XML 詞彙集，便能隨時自動更新來自各供應商的商品目錄和價格資訊，並根據商品類別為客戶提供最直觀的購物環境。圖書館更可利用 XML 達到自動化的館際合作、館際互借、線上採購等工作。

(三) XML 與圖書資訊組織與資料交換

隨著網路的發展，許多新的標準紛紛出現，XML 以其具有可擴展性、高度結構化和良好的資料組織能力，能夠有效的表達網路上各種知識，為資料的交換和處理提供新的機制。XML 並能配合 DTD 驗證其資料內容與結構是否合乎規範，確保其正確性。

圖書館較長時間以來採用 MARC 格式交換書目資料。MARC 格式僅僅用於圖書館系統間數

據交換，MARC 格式不能取代系統的內部格式。這些異質系統要想在網路環境下實現資源共享，還必須採用 Z39.50 協定。圖書館界一向使用 ISO 2709 做為資料交換格式的標準，但其並無檢查記錄正確性的功能，隨著網路的興起，多數的圖書館自動化系統皆提供 Web 式的查詢介面，因此直接使用 XML 做為資料交換格式，可以統一所有的輸出格式，避免額外製作許多的不同格式相互轉換工具。

(四)XML(Topic Maps)在數位圖書館的應用

主題地圖為建構知識管理系統所發展的一套方法論，主要用於架構、擷取和引導資訊資源工作上所需的模型建立。即是電子化的主題索引，架設在 WWW 環境，利用 XML 語義標準、及 DTD 定義來提供運用，將資源內容所代表的主題及領域，呈現在同一層級上。功能類似於書本後面所附的關鍵詞索引

主題地圖在數位圖書館的應用，可利用 XTM(XML Topic Maps)來標示數位化物件，依其主題的串聯，來儲存、取得與 metadata、數位物件的群組知識與推衍關係，開發出以 XML Topic Maps 為主的知識管理平台，讓數位圖書館只是 metadata 與數位物件之管理系統。

四、開放式數位圖書館標準及數位版權管理

XML 是網路環境下對電子資源進行處理的一種行之有效的新技術。它適用於多種領域各種電子資源(如全文、多媒體)的處理，同樣也適用於簡潔明瞭的目錄資訊處理，如 Dublin Core 的 15 個欄位或 MARC 中的各欄位的處理。Dublin Core 與 XML 相結合，可將 XML 在各種領域需要處理的資料語義辭彙予以簡化和規範，使 XML 的更能廣泛應用。

(一)開放式數位圖書館標準

數位圖書館的問題之一即在各系統間各有獨特的作業模式，大部分的數位圖書館之間，並不能互相透過地溝通以及分享彼此的資訊。數位圖書館採用標準的原因，期許能達成系統互通、資源分享、平台獨立、系統移轉、資料長久保存的功能。而資料處理相關標準，包括資訊組織標準(Metadata)、檔案格式(RDF)、命名原則(DOI、URN)、分散式資訊查尋與檢索標準(Z39.50、OAI)、使用者介面等層面。

(二)數位版權管理

數位版權管理可將一些先進的技術整合在一起，使得內容擁有者和版權所有者能保護他們手上已經數位化的智慧財產權。數位版權管理能將數位內容加密，而且限制內容的流通，只讓有權限或執照的人使用、播放或觀看特定的數位內容。數位版權管理應用目的在於內容保護與追蹤、防止版權內容遭到非法使用。企業界可用以確保機密文件等智慧財產不被內部員工外洩，重點是文件的管理機制。出版業使用數位版權管理技術，重點是內容的使用範圍加以界定。

五、未來展望：XML 與 Semantic Web(從資訊網到語意網)

將全球資訊網上的資料，變成電腦能理解的資料型態，和「類神經網路」相似，希望網路能了解人們所輸入的字的真正意思，進而提供更好、更方便的全球資訊網。語意網的實現需要依靠 Ontology(知識本體)，Semantic Web 才能成為下一代的網際網路。Ontology 是語意網成功的因素之一，其用以描述及表達特定領域知識的一組概念或術語，用來研究特定領域知識的類別、類別的屬性以及類別與類別間的關係，進一步達成概念語意的資訊檢索。隨著電腦科技的發展，知識本體論已漸漸從人工智慧應用到知識管理的領域，用以分享知識和

再使用知識。

以圖書資訊為對象的資訊組織，和以知識資源為對象的知識組織，在層次上畢竟有所不同。語意網即是以較深層、形而上的方式，讓網際網路上流通的知識資源夾帶更多意義，不僅供人類；也讓電腦能理解，當它以知識本體邏輯等方式來表達，越模糊也就不侷限於文字的束縛。

參、結論與心得

經由授課老師對 XML 的理論與應用的傳授，將 XML 的效益歸納如下：(一)資訊檢索面：由於 XML 具有自我描述性質，因此可以提供語意層次的搜尋，避免盲目搜索，進而提昇檢索結果的精確率。(二)系統的開發面：由於具備可擴展性、資料與樣式分離等特色，各個系統可以根據自身的需求，對 XML 資料進行其他加值處理，這使得 Web 應用程式的發展，更具彈性。(三)資料庫管理系統面：由於 XML 文件的高度結構化，使得 XML 可以很容易和目前已發展成熟的各種資料庫管理系統(DBMS)進行資料交換，這意味著來自不同管道的資料，可以輕易的藉由 XML 加以整合。(四)知識管理面：XML 能夠有效的表達網路上各種知識，為資料的交換和處理，提供新的機制，因此 XML 將促使網際網路，從資訊處理階段跨越到知識管理階段。

此次課程的三位授課老師在 XML 領域各有所長，但課程安排豐富且深入，雖然以往在文章上或研討會中曾聽過 XML 的論述，但藉由此次研習課程中 XML 實務課程的安排，使學員們對 XML 有進一步的認知，有助於瞭解 XML 相關技術發展及與圖書館業務的關聯性。數位時代的來臨，圖書館員若對 XML 加以瞭解，將有助於將來建置電子資料庫時，對系統的開發與設計可提出具體需求與規範，但對於技術的部分，還需仰賴程式設計師的專業所長。

參考書目

- 1、林信成，〈XML 相關技術與下一代 Web 出版趨勢之研究〉，教育資料與圖書館學 37 卷 2 期 (民 88 年 12 月)，頁 184-210。
- 2、林信成，〈智慧型文件與智慧型系統整合之研究〉，在 XML 技術與圖書館應用研習班研習資料(台北市：中國圖書館學會，民 93 年 8 月)，頁 125-142。
- 3、陳昭珍，〈以 XML 技術設計詮釋資料格式(Metadata)及數位圖書館管理系統之理論與實作〉，大學圖書館 5 卷 2 期(民 90 年 9 月)，頁 12-24。