

植物教學園區數位典藏詮釋資料研究

劉士華 葉日進

國立臺灣師範大學圖書資訊學研究所

tbrain@tp.edu.tw ron.yeh@msa.hinet.net

摘要

隨著數位典藏計畫的推展與實施，詮釋資料(Metadata)已成為不可或缺的基礎研究與工作項目之一。開發植物教學園區數位典藏資料庫，藉由網際網路，提供民眾一個認識植物知識的管道，並做為後續增值應用的基礎。本研究進行典藏數位化時，以學科與社群為導向的詮釋資料核心集，以使同質性之資料共用同一檢索介面、共用檢索結果之顯示格式、而且容許跨資料庫檢索，以便提高讀者使用上之便利性為目標。因資料群的增加、以及未來整合之需要，分析植物的內容及特色時，比較不同的Metadata標準，建置便於檢索的資料庫系統為目的，設計屬於植物的詮釋資料，以提供國內外研究參考。

關鍵詞：詮釋資料、數位典藏

前言

中小學教育中認識植物的教學通常是零星而鎖碎，在尚未具備統整的植物分類觀念與經驗之前，教學上只能以有趣的單元切入。就學習者的觀點而言，有下列困難尚待克服：

一、科屬名稱的分類法則太過玄奧：例如為何玫瑰是薔薇科？為何月桃是薑科。反之，夾竹桃屬於那一科別？欖仁樹屬於那一科別呢？

二、必須確知正確的植物名稱才能查詢檢索：在不確知正確的植物名稱時，無法正確的以關鍵字檢索出目標植物的資料。

三、無法以功能性的查詢方式檢索出相關主題的植物，以供比較與建構多面向的相關概念。例如：原生於台灣的觀花植物有那些？每年十月份開花的植物有那些？可以防風定砂的植物有那些？

我們希望教學場域上能有一個支援認識植物的輔助教學系統，系統內有充實的數位內容，更有良好的檢索機制，是教導者隨手可得輔助教學之教材。初期學習者可以自主的(自動的、自由的)學習，後續的學習者、進度落後者可以自主性地補救學習，進度超前者可以自主性地增廣學習。傳統植物資料庫雖然已保存相當數量的典藏資料，可是面對上述尚待克服的問題，仍不易解決中小學生學習的困境。然而，最新發展的數位典藏技術將是進一步解決問題的途徑。

研究目的

為了資料的交換、典藏之需要，本研究設計中文詮釋資料的格式，並與XML技術結

合，以協助使用者更方便、準確的從數位資料庫中擷取資料，希望學生能對植物的構造、分類的依據、現行分類系統有整體系統性的概念，而且能使用檢索表查詢生物名稱，引導、啟發學生對校園植物的的欣賞與珍惜。未來如能就資料內容建立內容標記的標準，全面進行內容標記(Content Markup)的工作，將可使數位資料庫的使用邁向一個新的境界。

文獻探討

由於資訊科技的進步及網際網路的快速發展，國內近來對於數位學習的接受度提高，相關產業連帶蓬勃發展，數位學習已成為輔助傳統學習的新趨勢。

Metadata一般而言，指描述資料的資料。其真實涵義在於對資訊的內容與外觀，進行適當性描述，為典藏品建立詮釋性記錄，可因各種不同需求進行交換及儲存(Ron & Carl, 1997)；為了發揮學習資源再利用的最高效果，最好能讓電腦系統自動揀選學習元件，此時電腦就必須靠詮釋資料來作判讀與分析，以發現、過濾、篩選出合適的學習元件來組成課程。

隨著數位典藏計畫的推展與實施，詮釋資料(Metadata)已成為不可或缺的基礎研究與工作項目之一。數位學習是數位典藏非常重要的加值應用。Metadata的主要目的，是協助資料的尋找與資料檢索。一般依據下列要點進行Metadata分析：使用環境(environment of use)、功能(function)、產生(creation)、稱號(designation)、相關的查詢協定(associated search protocols)、狀態(status)(Dempsey & Heery, 1998)。

目前e-Learning的環境下，有許多LMS(Learning Management System)或資訊系統(ERP or Content management System)散佈在網路上，由於它們都是為了某個應用目的或是想解決某些問題而存在；所以對於文件存取的格式及重複使用的功能上，並不相同也未考量彼此互通性。因此若從技術的角度來實作e-Learning 環境，上述的系統在資料交換上比較不容易，因此為了使系統或異質資料庫能達到資訊交換及再使用的目的，所以採行XML的技術，是目前可行的方法，且XML 標準也被W3C推薦為現行統一的文件交換格式。

XML(eXtension markup language)是由World Wide Web Consortium(W3C)在1996年底所制定的標準，一種延伸式的標記語言，從SGML(Standard Generalized Markup Language)衍生出來的簡化格式，是SGML的子集，具有擴展性(Extensibility)、結構性(Structure)、描述性(Description)、確認性(Validation)等特性。同時XML具有跨平台的功能，以及HTML可在網路上瀏覽、傳送的優點，並去除了SGML 複雜不常用的規則，對於不同的作業系統、硬體設備、應用軟體、多元的輸入模式，開發者可以自行制定符合己身需求的標記(tag)，做結構性的描述，促使相同的一份文件呈現不同的規格，適用於不同的軟體，符合不同的設備、滿足多重的輸入方式。XSL (Extension Stylesheet Language) 及CSS (Cascading Style Sheets) 也是W3C所提出，都是排版的語言，而XSL更是專為XML 而設計的排版語言，CSS 原本是針對HTML所發展出的排版語言，XSL與CSS 在XML文件的排版上，是相輔相成的(Hakon & Janne, 1999; Serge, 1999; W3C, 2006)。

對於XML文件的資料處理方法，大致可分下列四種：

- 1.檔案系統：例如DOS模式的目錄與檔案管理。
- 2.關聯式資料庫：傳統RDBMS。
- 3.物件導向管理模式：OODBMS。
- 4.特定資料庫：專為XML而發展的DBMS。

除了特定資料庫仍未形成標準，國內外已有許多適當的資料處理方法，余顯強(2004)以程式驗證實驗性方法評估效能的結果，無論使用何種資料結構，關聯式資料庫的效率明顯優於其他各種檔案架構或資料庫形式。至少在XML資料庫並未標準化之前，關聯式資料庫仍是儲存管理與運作最適當的儲存體。

資料庫的優點在資料管理、維護及查詢上，而 XML 文件的優點在資料的遷移、呈現、架構欄位的修改上，其比較如下表(陳百薰&洪政欣，2001)。

表 1：資料庫技術與XML技術在資料建置、管理及維護上的不同

	資料庫技術	XML文件
資料新增、刪除、修改	須藉由特定資料庫軟體來進行	一般的文字編輯工具皆可
資料的查詢	資料庫技術以SQL的語法或全文檢索的方式來進行查詢，方便、功能強。	在查詢語法上尚未有標準，正在發展中，更何況查詢的工具，除非各別以程式語言來寫出符合自己專屬 XML 架構的查詢工具。
資料遷移到另一個平台或機械	在資料遷移前，必須確定有其專屬的資料庫軟體，再行設定作業。	非常容易，只須將文件複製或移動過去即可。
欄位或標籤的修改	為保持資料的一致性、完整性，欄位最好不要修改，也不允許修改，在輸入資料時最好已完成欄位的架構。	非常簡單，加在 DTD 中即可，再修改 XML 文件內容。
資料管理與維護	資料庫軟體本身提供很好的機制。	一般只藉由最基本的檔案管理。
資料的呈現	與網頁程式結合	只須有樣本表的配合。

研究內容

在擬規劃時應以「學習」為考量重點，「數位之技術」只是使學習更有效率。「檢索表的使用及製作」：主要是學習分類最重要的工具---檢索表的類別、製作及使用，並讓學生實際收集植物，進行檢索表的製作。

學生以數位相機或手機拍攝得到的植物影像，可以自由地在系統中加以比對、分析、歸納、統整，將學習者自行獲取的初步資料在線上交給專家確認並回覆給學習者，一方面可以正確的確認所看到的植物資料，另一方面更可以充實資料庫的內容；這種互動式學習模式是配合數位學習的最佳典範。

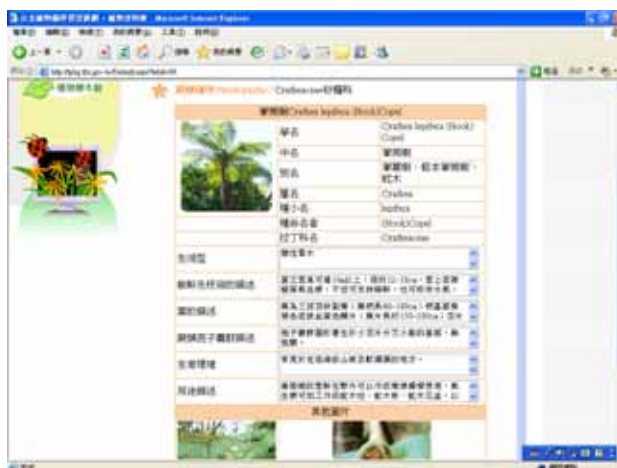
植物的數位典藏應具有科學性、教育性及實用性，故而產生分類的主觀性。依教育的觀點，植物可分為觀賞植物、藥用植物、食用植物、水生植物、經濟作物、有毒植物……等。不同分類的植物，其描述屬性的結構必有所差異，例如：觀賞植物需注重其生長期及培養方法等屬性，經濟作物會有產量、產值等屬性，藥用植物會有主治、療效、藥用

部位等屬性，有毒植物則有毒性描述、解毒方法等屬性。

如何能將不同分類植物之必要屬性全部納入此一植物資料庫之metadata，使得XML結構兼容並蓄，完整的適用於各類植物，將是本研究的設計重點。

本研究以台北植物園學習資源網(<http://tpbg.tfri.gov.tw/Default.aspx?tabid=56>)的植物資料庫為樣本，內有蕨類植物(Pteridophyta)、裸子植物(Gymnospermae)、被子植物(Angiospermae)。

學生若要查筆筒樹，必要知道要依序點選，才能得出右圖的畫面。



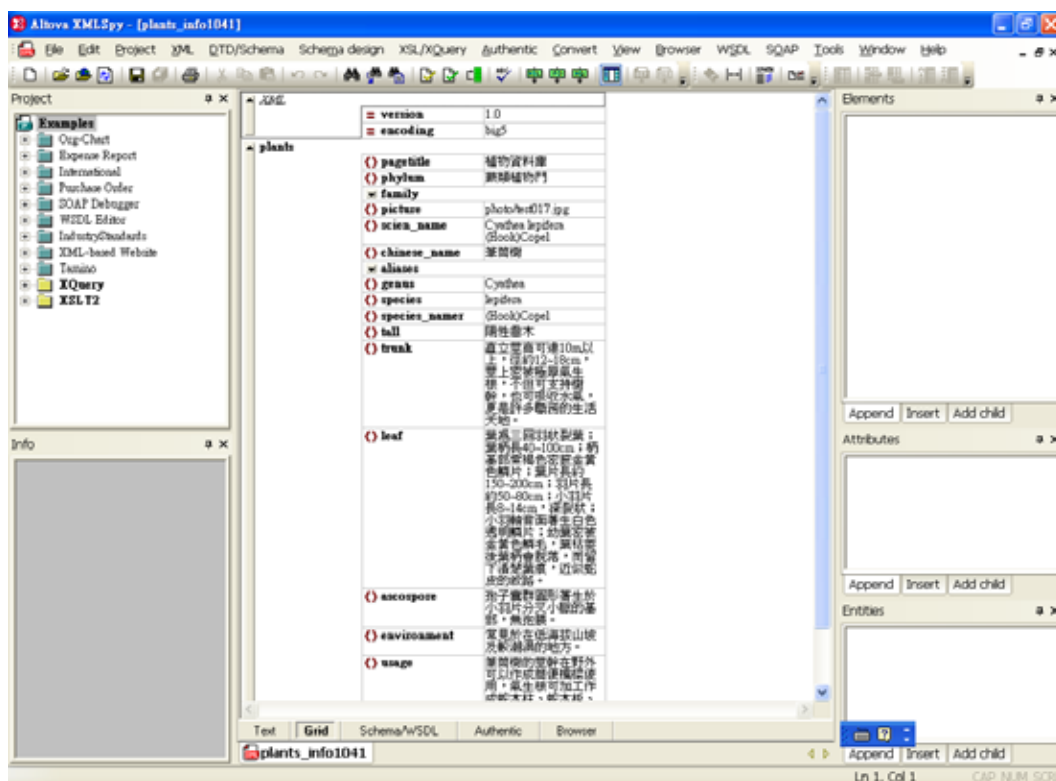
同樣的資料庫，我們給予新的結構。以筆筒樹為例，其 xml 定義內容如下：

項目名稱	中文說明		資料型態	必填	多值	
plants	pagetitle		頁面標題	文字	*	
	phylum		門分類名	文字		
	family	family_en	科名	中文科名	文字	
		family_ch		英文科名	文字	
	picture		主要圖片檔名	文字		
	scien_name		學名	文字		
	chinese_nam		中名	文字		
	english_name		英名	文字		
	aliases	alias	別名群	別名	文字	*
	origins	origin	原產地群	原產地	文字	*
	genus		屬名	文字		
	species		種名	文字		
	species_namer		種命名者	文字		
	ordinal		目分類名	文字		
	tall		生活型	文字		
	trunk		枝幹描述	文字		
	leaf		葉描述	文字		
	flower		花描述	文字		
	seed		種子描述	文字		
	climate		物種氣候	文字		
	environment		生活環境	文字		
	usage		用途描述	文字		
	otherspec		其他敘述	文字		
pics	pic	其他圖片群	其他圖片檔	文字	*	

```

<?xml version="1.0" encoding="big5" ?>
<!ELEMENT plants (pagetitle | phylum | family | picture | scien_name | chinese_name | english_name | aliases |
origins | genus | species | species_namer | ordinal | tall | trunk | leaf | flower | seed | climate | environment |
usage | otherspec | pics)+>
<!ELEMENT family (family_en | family_ch)+>
<!ELEMENT aliases (alias)*>
<!ELEMENT origins (origin)*>
<!ELEMENT pics (pic)*>
<!ELEMENT pagetitle (#PCDATA)>
<!ELEMENT phylum (#PCDATA)>
<!ELEMENT family_en (#PCDATA)>
<!ELEMENT family_ch (#PCDATA)>
<!ELEMENT picture (#PCDATA)>
<!ELEMENT scien_name (#PCDATA)>
<!ELEMENT chinese_name (#PCDATA)>
<!ELEMENT english_name (#PCDATA)>
<!ELEMENT alias (#PCDATA)>
<!ELEMENT origin (#PCDATA)>
<!ELEMENT genus (#PCDATA)>
<!ELEMENT species (#PCDATA)>
<!ELEMENT species_namer (#PCDATA)>
<!ELEMENT ordinal (#PCDATA)>
<!ELEMENT tall (#PCDATA)>
<!ELEMENT trunk (#PCDATA)>
<!ELEMENT leaf (#PCDATA)>
<!ELEMENT flower (#PCDATA)>
<!ELEMENT seed (#PCDATA)>
<!ELEMENT climate (#PCDATA)>
<!ELEMENT environment (#PCDATA)>
<!ELEMENT usage (#PCDATA)>
<!ELEMENT otherspec (#PCDATA)>
<!ELEMENT pic (#PCDATA)>

```



完成的數位典藏資料庫網址如下，
http://140.122.104.15:8080/ug-102.jsp?xsd_name=tree&start_rec=1&page_size=10&sort_order=create-date&display_mode=brief：



結論

一、本研究朝多元化發展，數位檔案與產生方式以人工為主，功能訴求以結構化描述與組織為主。描述深度集中在資料結構，同時涵蓋一些資料值的標準，並未涉及資料內容方面的標準。描述層次以典藏品(item)為主，少數以全文為主，並以數位方式掃描為數位檔案，尚未進行內容標記。整體以深化描述與通用式發掘並重，將詮釋資料標準的主要功能定位在數位資訊生命週期中的組織、查詢與檢索，以及使用等三個層面。

二、本植物資料庫數位典藏有互動式的輔助學習機制，其可分為兩種循環過程，描述如下：

(一)、學生學習循環：

1. 學生經過觀察及拍攝取得所見植物的影像及認知的描述。
2. 學生將影像及描述登載進入植物資料庫數位典藏學習網頁。
3. 典藏學習網進入資料庫搜尋特性特徵符合的植物列表以供學生辨認。
4. 學生可進一步觀察、比較、篩選之後確認目標植物，或者不能確認者，將發出通知，留待專家鑑定及回覆。
5. 回饋給學生目標植物的完整資料，或提供進一步學習的建議。

(二)、資料庫學習循環：

1. 任何一筆學生登載的影像及其描述均記錄在前期資料庫內。

- 2.將一-3搜尋特性特徵符合的植物列表附記在前期資料庫內。
- 3.將一-4確認或者不能確認目標植物附記在前期資料庫內。
- 4.由系統發出通知，由植物專家確認學生所做的篩選是否正確，及正確的目標植物資料，回覆給學生。
- 5.系統將已經專家確認的正確影像與描述資料加入植物資料庫內。
- 6.刪除前期資料庫相關內容。

參考文獻

- 陳百薰，洪政欣(2001)，數位博物館虛擬展示之模型、概念與系統。新世紀數位圖書館與數位博物館趨勢研討會，國立交通大學。
- Dempsey, L., Heery, R.(1998). "Metadata: a current view of practice and issues". Journal of Documentation, 54(2), 145-172.
- Hakon Wium Lie, Janne Saarela(1999)," Multipurpose Web publishing using HTML, XML, and CSS", Commun. ACM 42, 10, Pages 95-101.
- Ron Daniel, Carl Loagoze(1997),"From Metadata containers to active digital objects," D-Lib Magazine November.
- Serge Abiteboul(1999), "On views and XML", In Proceedings of the eighteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, Pages 1-9. W3C, <http://www.w3.org>.