

Wiki 協作系統在數位典藏內容加值之應用研究

林信成 (Sinn-Cheng Lin)

淡江大學資訊與圖書館學研究所 副教授
sclin@mail.tku.edu.tw

陳瑩潔 (Ying-Chieh Chen)

淡江大學資訊與圖書館學研究所 研究生
693070020@s93.tku.edu.tw

摘要

數位時代知識的累積、傳播及管理已進階至數位內容品質的提升，本研究以參與國科會數位典藏計畫之「台灣棒球運動珍貴新聞檔案數位資料館之建置」為基礎，透過 Wiki 協作系統進行數位內容之分類、管理，提出一個適用於 Wiki 協作環境的金三角模式，並實際應用索引、控制詞彙與索引典等知識組織技術，彙整台灣棒球新聞與其相關資料，並針對台灣棒球各類主題延伸出多樣化的加值應用，以彌補原始棒球新聞資料庫提供使用者閱讀時的不足之處。

關鍵詞：數位典藏內容加值、Wiki 協作系統、數位典藏國家型科技計畫

1. 前言

數位內容品質的提升與加值應用，有賴知識組織技術的運用，如分類(Classification)、索引(Index)、索引典(Thesauri)與控制詞彙(Controlled Vocabulary)等，以創造知識的最大價值。本研究為參與國科會數位典藏計畫「台灣棒球運動珍貴新聞檔案數位資料館之建置」[17]的衍生成果，在先前的研究中，研究者致力於台灣棒球新聞內容之資訊組織與 Metadata 研究[15]，提升數位典藏內容之傳播與利用。然而，在研究過程中，基於研究者推測：僅止於提供原始數位新聞內容，若從使用者角度而言，可能無法滿足其真正的資訊需求。為突顯數位環境下資訊加值的重要性，因此延伸出本研究對台灣棒球新聞進行加值處理的動機。為了彙整多元化的相關資料，提升資訊內容的豐富性，研究小組乃研擬透過協同創作的過程，以知識組織技術來進行內容加值，彌補棒球新聞資料庫提供使用者閱讀時的不足之處。至於協同創作平台之選擇，本研究藉由系統評估，以建置時所需花費的時間、成本、人力及技術等面向作為考量，分析數種可行方案並加以評估後之結果如表 1 所示。

表 1 系統評估

方式 評估	單純網頁製作	自行開發系統	採用 OSS 方案	購置現成系統
時間	高	高	低	中
成本	低	高	低	高
人力	中	高	低	低
技術	中	高	中	低

由上表可知，OSS (Open Source Software，開放原始碼軟體)[4]是上述眾多方案中較適合本研究的選擇。而在眾多開放原始碼協作系統當中，近年來又以維基(本文以 Wiki 稱之)與部落格(本文以 Blog 稱之)較適合應用於知識組織領域。然 Blog 較著重於個人化觀點之表達[14]，類似討論版的運作模式；而 Wiki 系統除強調內容之中立觀點外[20]，更具備較完善的多元分類、組織與易於管理的功能特性，並結合超連結、重定向、可搜尋、界面親和度高等特點，最重要的是可透過協同創作的過程來完成數位內容的加值，因此本研究選用 Wiki 作為技術平台。

本論文架構如下，第一節為前言；第二節提出 Wiki 協作金三角模式；第三節為系統建置；第四節討論設群建構與管理策略；第五節則依據知識組織的概念，以 Wiki 系統進行數位內容之加值應用；第六節則描述本系統綜合應用的成果；第七節為結論與建議。

2. Wiki 協作模式

2.1 Wiki 金三角

Wiki 系統的創造者 Ward Cunningham 曾為 Wiki 下了定義[2]，他認為 Wiki 不只是一套用來儲存與修改資訊的超文件系統，且要每個頁面皆能開放任何使用者編輯，因此必須是一個可自由擴展且相互連結的資料庫系統，只要藉由具備表單功能的瀏覽器即可進行協同創作 (Collaboration)。

數位化時代之知識管理與「系統」(System)、「社群」(Community)、「內容」(Content)等三者相關；又受「平台技術」(Platform Technology)、「知識組織」(Knowledge Organization)和「管理策略」(Management Strategy)等三大面向之影響，研究者因此提出一個適用於 Wiki 協作環境的金三角模式 (Triangle Model)，如**錯誤! 找不到參照來源。**所示。在此 Wiki 金三角模式中，知識的創造乃是由「Wiki 三要素」，藉由「Wiki 三面向」共同形塑而成，詳述如後。

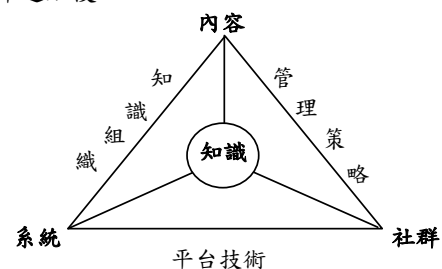


圖 1 Wiki 協作環境金三角模式

2.2 Wiki 三要素

在 Wiki 金三角模式中的「Wiki 三要素」是指「系統」、「社群」與「內容」三者：

(1) 系統

Wiki 系統是 Wiki 協作環境的基礎建設，通常由 Web server (網頁伺服器)、server-side script program (伺服器端程式) 和 DBMS (資料庫管理系統) 所構成，統稱為 Wiki 軟體(Wiki Software) 或 Wiki 引擎(Wiki Engine)。

(2) 社群

在 Wiki 協作環境中，社群主要是由讀者、作者與管理者所組成。讀者是指僅閱讀資料而無進行任何編輯動作的使用者；作者則不僅止於閱讀，而是實際參與協作編輯的核心使用者，更是 Wiki 數位內容之重要來源，任何讀者只要願意編寫，隨時皆可成為作者，不受任何條件限制；管理者則是具有文章刪除/恢復、帳號封鎖/開放、系統設定/維修等能力的超級使用者，具有較高的權限，以維護系統正常運作。

(3) 內容

Wiki 協作環境中的數位內容，是由社群中的成員共同完成，任何成員都可編輯與修改內容，更可透過各種知識組織技術進行分類、索引等加值處理，以提升數位內容的品質。至於內容主題可以是百科全書、技術報告、研究論文、新聞動態、旅遊指南 ... 等，並無一定限制。

2.3 Wiki 三面向

「社群」、「內容」與「系統」三要素之間的運作，則須由「平台技術」、「知識組織」和「管理策略」三大面向加以串接，形成一個完整體系：

(1) 平台技術

「社群」與「系統」間乃是透過平台技術為媒介，將知識完整呈現並提供利用，因此在 Wiki 環境中有賴系統管理者建構協作平台。目前已有眾多 Wiki 系統平台被發展出來，如 MediaWiki、MoinMoin、TWiki、PhpWiki 等[8]。

(2) 知識組織

「內容」經由知識組織的過程存放於「系統」之中，以達成內容加值的目標；同時系統亦應提供強有力的知識組織工具，以方便內容之編輯。

(3) 管理策略

「社群」之間需制訂一套管理策略(如編輯守則、品質守則等)，以作為「內容」協作時遵循的依據，並可進行內容品質的控管；此外，亦需有行銷策略，用以提升網站之可見度和內容之使用率；而版權策略則是數位內容版權管理之指導原則。

3. 系統建置與平台技術

本研究旨在將執行國科會數位典藏計畫「台灣棒球運動珍貴新聞檔案數位資料館之建置」過程中所收集之台灣棒球相關資料，延伸建構一個線上開放的台灣棒球歷史教育館，旨在記錄台灣棒球發展史中值得記錄的人物、事件、時代、地點、文物 ...

等，並以 Wiki 開放協作方式供大眾對新聞報導之外的相關資料加以彙整，以充實我國棒球發展的歷史紀錄。

在上述的 Wiki 金三角模式中，系統是首要的基礎建設，本研究經由「平台技術」面向進行評估後，選用 MediaWiki 作為系統平台，它因被用來撰寫號稱全球最大的線上百科工程「維基百科」(Wikipedia) [10]而頗具盛名，該系統功能特性亦符合本研究之應用。目前本研究採用的規格如下：

- 軟體版本：MediaWiki 1.44；
- 程式語言：PHP；
- 資料庫管理系統：MySQL + MySQLAdmin；
- Web 伺服器：Apache。

4. 社群建構與管理策略

完成初步之系統建置後，研究者決定將此實驗網站命名為「台灣棒球維基館」[18]，並開始進行「社群建構」與「管理策略」等面向之構思。起初，以研究團隊五人小組為核心社群，藉由小組討論及協同創作過程，先行以輔助說明(Help)方式完成本站編輯原則與手冊之撰寫，以供協作者遵循；再者，在數位版權協議部分，則採用 GNU 自由文檔許可證(GNU Free Documentation License)之方式，以避免版權糾紛；且因 Wiki 開放協作過程具有「先出版後編輯」之特性，其頁面可能出現過渡、不完整或甚至錯誤之內容，因此，研究者以「免責聲明」方式進行告知，提醒使用者使用本實驗網站內之資源時必須有此方面之認知。至於更深入的「管理策略」面向，如頁面品質管理、社群規模擴增、網站推廣 ... 等，則留作下一階段之研究重點。

5. 知識組織與內容加值

本研究主要重點乃在於 Wiki 金三角模式中的「知識組織」面向之研究，以實踐知識組織技術應用於數位典藏內容加值之概念。

知識組織為數位內容加值的方法之一，其理論是建立在知識單元——概念的基礎上，是將無序或分散的特定知識，根據一定的原則與方法，使之有序、集中、定址，以方便知識的提供、利用和傳播[13]。在網際網路與科技的進步下，最新型態的知識組織為知識處理(Knowledge Processing)的一個過程，主要指知識的分類及構造，其包括：編目(cataloging)、索引(indexing)、分類(clustering and classification)、過濾(filtering)及聯結(linking) [19]。因此，「台灣棒球維基館」乃是以這樣的理念為基礎來建構的。

5.1 分類架構的設計

分類是依據物件間的關係將其排序分組的一種活動[5]，John Dewey 更曾提出分類就是知識(all knowledge is classification)的理念[3]。因此，也說明了分類架構在知識體系的範疇中是極具關鍵性影響力的。

本研究首先針對棒球運動相關主題進行分類架構的設計，如表 2 所示，並利用 MediaWiki 系統

稱，兩者被視為同義，結果如圖 5 所示。

表 3 同義概念的製作舉例

類別	同義詞
球隊	嘉義農林棒球隊 = 嘉農隊
人物	郭泰源 = 東方特快車
賽事	1984 年第二十三屆洛杉磯奧運會表演賽 = 1984 年第廿三屆洛杉磯奧運會表演賽 = 洛杉磯奧運 = 1984 年奧運 = 第二十三屆奧運 = 第廿三屆奧運.....
組織	少年棒壘聯盟 = 少棒聯盟 = LLBS
地點	威廉波特 = 美國威廉波特 = 美國賓州威廉波特 = Williamsport
其他	全壘打 = Home runs = 紅不讓



圖 5 同義概念的製作—檢索結果

5.3.2 階層與關聯的概念

除同義關係外，索引典尚具有階層與關聯概念。本研究採用的方式是以主題詞為標題，在其 Wiki 頁面中，將相關詞彙列出，再描述其廣義詞、狹義詞或相關詞，以呈現其上下層主題間的關係，並透過連結的功能將彼此串聯起來。如此可提醒使用者有另一種關係的存在，並引導使用者加以利用。以圖 6 所示為例，運用似類索引典的詞彙關係，在紅葉少棒的頁面中，將該詞彙的相關主題列出，並揭示其相互關係。



圖 6 階層與關聯概念的製作

5.3.3 同形異義的概念

在知識組織過程中，也時常遇到同形異義的情形，本研究的作法是採用另加一個分歧頁面的方式處理。以圖 7 為例，台灣棒壇人物中曾有兩位「林建良」，同名同姓卻不同人，因此先以分歧頁面說

明以便判讀，再採間接連結方式將其區隔取用，搜尋時也能有所區別。



圖 7 同形異義概念的製作

6. 綜合應用

除以知識組織技術的實踐外，本研究更進一步將內容加值成果予以綜合應用。以下即針對目前「台灣棒球維基館」的綜合應用情形加以說明。

6.1 台灣棒球發展史

台灣棒球發展史是以電子書方式訴說台灣棒球運動發展的軌跡，讓讀者可像閱讀書籍般，並在資訊架構中搭配上下單元的情境式導覽設計，導引讀者依章節閱讀相關資訊，如圖 8 及圖 9 所示。



圖 8 台灣棒球發展史之目錄

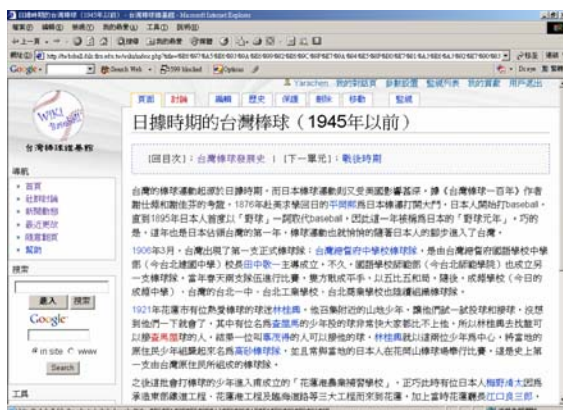


圖 9 台灣棒球發展史之內容

6.2 台灣棒球大事記

以年度為單位，依據年份排列整理台灣棒壇大事記，並於文中運用變數引入功能輔以超連結將相關資訊作更詳盡的說明及建構，如圖 10所示。

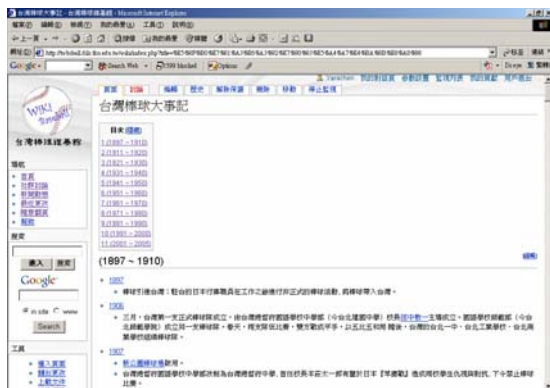


圖 10 台灣棒球大事記

6.3 台灣棒球名人錄

將台灣棒球史上各個不同時期的重要人物加以彙整，按照年代加以排列，並附上簡要說明，以增加頁面的可讀性，如圖 11所示。

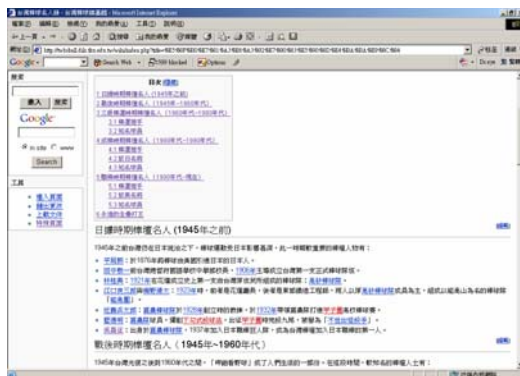


圖 11 台灣棒球名人錄

6.4 台灣棒球影像館

此區域展示具歷史價值的照片，包括人物、球隊、場地、賽事、球具、球裝、球卡...等，將其以相簿的形式呈現，如圖 12所示。



圖 12 台灣棒球影像館一個人照片集

6.5 台灣棒球重要新聞報導

收錄由聯合報授權提供的歷史性棒球新聞，並呈現其新聞標題、記者、日期、來源、版數、內文等，運用超連結功能可將新聞中所提及相關的人、事、時、地、物的概念串連起來，讓知識的呈現形成網絡的形態，如圖 13所示。



圖 13 台灣棒球重要新聞報導

7. 結論與建議

無論人類知識的載體如何變化，知識組織的最終的目的仍是為了讓人們能便於使用及管理為出發點。本研究透過分類、索引、控制詞彙與索引典等知識組織技術的實際應用，已具初步雛型，目前完成「台灣棒球發展史」、「台灣棒球大事記」、「台灣棒球名人錄」、「台灣棒球影像館」、「台灣棒球重要新聞報導」等主題的加值利用，並已開放讓熱愛棒球的人士共同參與編輯，發揮協同創作的特點。透過此種加值應用方式，不僅能提升數位內容品質，也創造知識的最大價值。

本研究依此次之實作經驗，歸納出 Wiki 協作系統具備下述特點：

(1) 開放性協同創作：開放所有使用者對數位內容進行編輯與修改，以合作接力的方式完成某一主題，使該領域知識越來越完善。

(2) 結構化組織知識：將各個概念或詞彙以類似語意網路之網狀形式連結，並藉由資料庫管理，使知識更趨結構化。

(3) 技術面門檻較低：在系統架設方面，採開放原始碼作法，並有多種套件組合可供選擇，使用者不需花費太多心力於系統開發上，可專心於內容的建構；此外，其編輯語法簡易，非常容易學習。

(4) 以版本控制維持品質與安全：由於採開放式編輯，因此系統通常會記錄下每一筆異動，以便發生未預期之變動時，能迅速回復到先前的版本。

(5) 最近更新有助於編輯與管理：依據內容的更改時間排序，提供近期內更新之差異、修改日期、修改幅度、修改者等資訊，並配合特殊標記，使編輯與管理工作更容易。

(6) 持續成長的有機體：因開放編輯的特性，理想上，Wiki 的內容與結構將隨著時間成長，越來越多元且豐富。

然而,Wiki 協作系統並不是實現知識組織的唯一選擇,因其仍存在如開放編輯不易掌控其內容品質、可能遭駭客入侵破壞、需大量人力與管理不易等問題。在數位典藏內容加值的過程中,有更多環節是不容忽視的,因此提出下述問題與建議,可供未來研究與改善之參考:

(1) 內容專家的重要性:創造知識的同時,技術固然重要,但如何確保內容之正確性及可信度也是不容忽視的,而這則是需要仰賴該領域之內容專家,輔以知識的過濾與分析,才能維護其品質。

(2) 界面呈現應妥善規劃:在數位資源中,時常運用大量的超連結功能將知識串連,因此特別容易造成使用者迷失於其中,因此在知識呈現的界面規劃應明確指引其所在位置。

(3) 權威檔製作的必要性:數位典藏搭配權威檔的製作,不但可達到其標引的一致性,對於往後資訊儲存、檢索、流通及系統設計上更有其效益。本研究藉由系統協同創作與組織的特性,截至目前為止共集結台灣棒球相關之球隊(148筆)、人名(1212筆)、重要賽事(69筆)、組織(16筆)、地名(21筆)、術語(129筆)、特殊名詞(6筆),目前正持續增加當中,期望可供日後作為棒球相關之權威檔的依據。

(4) 與知識組織新概念的結合:近年來在各領域的努力研究下,發展出許多知識組織的新概念,如:Ontology[13]、ISO 13250 Topic Map[7]、Semantic Web[8]等,也有較具體且頗具成效的成果,如:國外的有IEEE提議的SUMO(Suggested Upper Merged Ontology,建議上層共用知識本體)[6]、普林斯頓(Princeton University)的詞彙網路(WordNet)[11]等,國內也有中央研究院平衡語料庫(Sinica Corpus)[12]、中文詞知識庫資料等。而這些概念都是將知識內容結構化,且將其資訊架構加以描述與定義,因此未來各領域都應發展屬於該領域的知識組織架構,才能在知識體系中與世界接軌。

身處於這個世代中,共享及協同創作的理念日益盛行,不再侷限於個人知識的創造,因此,在數位典藏運作的同時,僅提供使用者端數位內容已無法真正滿足其需求,其內容的加值應用層面更是不容我們忽視的課題。

致謝

本文為 NSC 94-2422-H-032-002 研究計畫部分成果,國家科學委員會經費補助,聯合報授權使用所需新聞資料,使本研究得以順利進行,特此致謝。

參考文獻

- [1] ANSI, "ANSI Z39.19-1993(R1998) Guidelines for the Construction, Format, and Management of Monolingual Thesauri", p.1, <<http://www.niso.org/standards/resources/Z39-19.pdf>> (2002/12/27)
- [2] Bo Leuf, Ward Cunningham, The Wiki Way: Quick Collaboration on the Web, Addison

Wesley, Addison Wesley, 2001.

- [3] M. P. Satija. "Classification : some fundamentals, some myths, some realities." Knowledge Organization 25: 1-2(1998) : 32.
- [4] Open Source Initiative, " Open Source Initiative OSI", <<http://www.opensource.org/>>(2005/07/15)
- [5] Sherry Koshman. "Categorization and classification revisited: a review of concept in library science and cognitive psychology." Current Studies in Librarianship Spring/Fall 1993: 26.
- [6] Ontology portal, "Suggested Upper Merged Ontology (SUMO)", <<http://www.ontologyportal.org/>>(2005/07/15)
- [7] TopicMap.Org, "XML Topic Maps (XTM) 1.0 - TopicMaps.Org Specification," 8 Aug. 2001, <<http://www.topicmaps.org/xtm/1.0/index.html>> (2004/1/5).
- [8] W3C, "Semantic Web", <<http://www.w3.org/2001/sw/>>, (2005/7/30).
- [9] WikiEngines, <<http://wikifeatures.wiki.taoriver.net/moin.cgi/WikiEngine>> (2005/07/15)
- [10] Wikipedia, <<http://wikipedia.org/>>(2005/07/15)
- [11] WordNet <<http://www.cogsci.princeton.edu/~wn/>>(2005/07/15)
- [12] 中央研究院平衡語料庫 <<http://www.sinica.edu.tw/SinicaCorpus>>(2005/07/15)
- [13] 阮明淑、溫達茂, ontology 應用於知識組織之初探,佛教圖書館館訊,32期(民91年12月),頁6
- [14] 官容鳳、張裕幸,從出版寫作與傳播媒介看「部落格」的發展與應用,國家圖書館館刊,93:2(民93年12月),頁143-161。
- [15] 林信成、康珮熏, "報紙新聞數位典藏 Metadata 轉換系統之設計與應用", 中文媒體數位典藏與新聞標示語言研討會,台北·國家圖書館,頁B2-1~B2-23(2005/5/11~5/12)。
- [16] 林信成、歐陽慧、歐陽崇榮, "以主題地圖建構索引典之語意網路模型", 圖書與資訊學刊,48期(民93年2月),頁35-56。
- [17] 淡江大學資圖所數位典藏計畫小組,台灣棒球運動珍貴新聞檔案數位資料館之建置, <<http://ndap.dils.tku.edu.tw>>。
- [18] 淡江大學資圖所數位典藏計畫小組,台灣棒球維基館, <<http://twbsball.dils.tku.edu.tw>>。
- [19] 陳和琴,中國圖書館學會九十一年資訊組織基礎班, <<http://datas.ncl.edu.tw/catweb/datas/3-1-007-1.pdf>> (2005/07/15)
- [20] 維基百科, Wikipedia: 中性的觀點, <<http://zh.wikipedia.org/wiki/Wikipedia:%E4%B8%AD%E6%80%A7%E7%9A%84%E8%A7%82%E7%82%B9>>。
- [21] 數位典藏國家型科技計畫後設資料工作組,建構數位典藏「主題索引典」:簡介與實務,民國91年3月27日,頁1。