

# 報紙新聞專卷的數位典藏與加值利用

林信成

淡江大學資訊與圖書館學研究所 副教授

sclin@mail.tku.edu.tw

鄭國祥

光武技術學院資訊管理系 兼任講師

孫正宜

淡江大學資訊與圖書館學研究所 研究生

## 摘要

在數位化新聞高度普及的今日，新聞的加值處理反而不若以往受到重視，原因在於使用者過度依賴全文檢索技術。本研究旨在探討數位化報紙新聞的加值與利用，爲了比較數位化新聞加值與未加值的不同效益，我們以台灣棒球運動相關之新聞專卷爲例，將 Metadata 及圖書館知識組織的概念導入數位化新聞專卷的製作，完整呈現新聞事件的發展歷程；並對於新聞事件的大事紀、新聞人物的生平事蹟與相關專卷，藉由適度的資訊組織方法加以分類、歸納，成爲更有價值之數位典藏資料，能讓使用者更深入、更有效率的探索相關新聞事件。

本文所提出之相關概念與作法，並不只適用於棒球新聞的數位典藏與加值利用，亦可應用於其他主題的新聞專卷。

關鍵詞：

新聞加值、新聞專卷、Metadata、NRMF、NITF、newsML、PAM。

## 壹、前言

報紙是訊息傳遞最快速的園地之一，也是記錄當代人文、社會、科技等活動的重要媒體。傳統上，爲了查找方便，報社會將某一專門主題的相關新聞加以群組，製作成主題性的「新聞專卷」，以方便記者、編輯利用。近年來，紙本報紙皆與電子報同步發行，使管理、保存、查找、利用更便利；舊報紙也逐年回溯建檔，以全頁影像掃描、重新打字和光學文字辨識等方式逐步數位化<sup>1</sup>。然而，在

---

<sup>1</sup> 孫正宜、林信成，「中文報業數位化技術與現況探討」，淡江大學資訊與圖書館研究所編，《2003 年資訊科技與圖書館學術研會論文集》，台北，2003，P73-93。

電子報逐漸普及後，新聞專卷的加值處理反而不若以往受到重視，原因在於使用者過度依賴全文檢索技術，造成主事者也認為有了全文檢索便不再需要花費人力進行新聞專卷製作了。其實，報紙除了將一則則新聞數位化之外，若能再進一步加值成數位化新聞專卷，必能更彰顯其利用價值。

本研究主要在探討報紙新聞專卷的數位典藏與加值利用，爲了比較加值與未加值的不同效益，我們以「台灣棒球運動相關之新聞專卷」爲例，將知識組織的概念導入數位化新聞專卷的加值與製作，完整呈現新聞事件的發展歷程；並對於新聞事件的大事紀、新聞人物的生平事蹟與相關的新聞，藉由適度的資訊組織方法加以分類、歸納，成爲更有價值之數位典藏資料，能讓使用者更深入、更有效率的探索相關新聞事件。

## 貳、數位化新聞加值探討

### 一、新聞加值與新聞專卷

「加值」的意義在於創造附加價值。因此，資訊加值的目的是在於提升資訊產品的價值<sup>2</sup>。本文將新聞加值定義爲創造更具利用價值的新聞資訊。我們結合知識組織方法與Metadata技術，將新聞加值的過程分爲「初步加值」與「進階加值」兩個等級。初步加值是利用關鍵詞加上布林邏輯運算，將數位化新聞歸檔到適當的專卷類目中，再由資訊管理者進一步檢覆、篩選、組織與整理；而進階加值則是依據初步加值的結果，對新聞事件深入解讀、分析、歸納，衍生成更有秩序更有條理的文獻。例如新聞事件的發展中，新聞大事紀、新聞人物生平事蹟等都是極重要的衍生資訊，必須經過深度加值後，才有可能一一產出，無法由電腦自動完成。

「新聞專卷」是由原始新聞所衍生出的文獻。其製作乃是專業的新聞編輯人員，依據或人、或事、或主題的重點議題，經分析、篩選、擷取、分類、組合等步驟，將同質新聞素材重組的創意加值行爲。<sup>3</sup>這種新聞資訊的加值，有助於新聞資訊的再使用與新聞提供者的競爭力，並能提供使用者專題選粹與新知通告的服務。早期製作「新聞專卷」的初衷，大都是新聞業者爲了提供編採同仁在新聞事件發生時，能有快速、完整且可靠的資料來源可供參考，於是由資料整理人員針對不同的新聞主題，把相關的資料加以收集、組織、整理，再分門別類的建檔，以供每日與時間賽跑的編輯、記者使用，不必再另花時間搜尋或查證。長期下來，竟然建立了一個又一個完整且可信度高的資料檔，累積成一座完善的專門圖書館，記錄了整個社會發展的軌跡。不同於每日出刊的報紙全文，每個新聞專卷都是分別針對單一主題長期追蹤與整理所得，比原始報紙全文更精鍊，專卷中有許

---

<sup>2</sup> 宋瓊玲，「資訊加值與圖書館服務」，國立中央大學圖書館通訊，可得自<<http://www.lib.ncu.edu.tw/book/n37/37-2-1.htm>>，2004/6/29。

<sup>3</sup> 饒大業，傳播媒體加值創新策略之應用-以聯合報系加值創新爲例，世新大學傳播研究所碩士論文，2001，頁69。

多額外加值的珍貴資料並未公開見報或出版，屬於灰色文獻，卻往往比見諸報端者詳盡、完整，更具參考與利用價值。因此，若能將新聞專卷加以數位化，必然更適合對於某個專門議題感興趣的人士使用。

接著，我們藉由探討「聯合知識庫」、「時報知識贏家」、「CNN」、「BBC」、「Yahoo!奇摩新聞」等五家新聞資料庫及新聞資訊提供者對新聞專卷的處理方式，以瞭解業者對新聞專卷的看法與做法。

#### (一)聯合知識庫

聯合知識庫新聞專卷作法是由專業編輯者依新聞的人物、事件、主題等特性來設定題目，按新聞題目將相關新聞資料內容匯集編成專卷，有如新聞機構資料室原本的剪報業務，也像是圖書館的專題選粹服務(Selective Dissemination of Information, SDI)的功能，<sup>4</sup>聯合報系新聞專卷約三萬卷。

#### (二)時報知識贏家

時報知識贏家是將新聞資料以專題式的檔案卷夾方式呈現，編輯人員編排時以人物、事件為主軸，專卷架構以大類、中類及小標細分，目前共分 23 大類、132 中類。熱門新聞訂為專卷架構，整理分類，把相關新聞聚集入專卷。專卷卷夾內，相關新聞又依新聞屬性，以小標分類，新聞資料與照片搭配呈現。<sup>5</sup>

#### (三)有線新聞網(Cable News Network, CNN)

CNN的專卷稱之為special reports，CNN以人物(如Osama bin Laden)、事件(如Massacre in Madrid)、主題(如Global Influentials)相關之重點新聞，製作為特別報導(Special Reports)，整理一系列頭條報導、相關新聞資源(如地圖、專家評論、新聞背景資訊)、新聞照片、影片等匯集為新聞專卷。<sup>6</sup>

#### (四)英國廣播公司(British Broadcasting Corporation, BBC)

BBC的專卷稱之為in depth，BBC將重要的新聞報導與熱門議題，依事件(如Vote USA 2004)、主題(如Teacher shortages)整理為專題報導(In Depth reports)，以專題題目與專題發生地點作為排序依據，專題內包括新聞、背景資訊、以及專家分析資料等訊息。<sup>7</sup>

#### (五)Yahoo!奇摩新聞

在Yahoo!奇摩新聞中，新聞專卷又稱新聞專輯。為Yahoo!奇摩新聞編輯針對當日熱門的新聞議題，從各種角度收集相關的新聞素材製作而成的相關新聞專區，便利讀者在最短的時間裡利用新聞專輯了解該新聞事件的全貌。<sup>8</sup>

<sup>4</sup> 林燕飛，圖書館新聞研究-以聯合知識庫為例，淡江大學資訊與圖書館學系碩士論文，2002，頁34。

<sup>5</sup> 時報資訊，「新聞專卷」，可得自<<http://www.infotimes.com.tw/newspro.htm>>，2004/6/1。

<sup>6</sup> CNN.com, "Special Reports", available at <<http://www.cnn.com/SPECIALS/>>, 2004/6/1.

<sup>7</sup> BBC News, "In Depth", available at <[http://news.bbc.co.uk/2/hi/in\\_depth/default.stm](http://news.bbc.co.uk/2/hi/in_depth/default.stm)>, 2004/6/1.

<sup>8</sup> Yahoo!奇摩新聞，「新聞專題」，可得自<<http://help.yahoo.com/help/tw/news/news-05.html>>，2004/6/1。

## 二、新聞專卷製作要素

由上可知，新聞專卷強調的是主題導向，它是有計畫的進行資訊組織，以滿足使用者的資訊需求。從實務面而言，新聞專卷製作過程是「動態的」(Dynamic)，必須因時、因地制宜。當某一主題之新聞經過持續報導，逐漸衍生成一個重大議題之後，便有建立相關專卷來管理的必要。然而，新聞專卷的產生雖是動態的，但並非完全「不可預測的」(Unpredictable)。有經驗的新聞從業人員，對於新聞報導中的事件，應思考：過去有發生過這類事件嗎？那時是怎麼處理？結果怎樣？未來可能衍生成重大議題嗎？若是，應對此事件多加留意、提早準備，以便將來時機成熟時，可隨時將該主題之相關新聞整合，製作成專卷。

當決定將某個新聞主題製作成專卷時，應考慮的要素除 5W1H 外，可再另加 1P1R 和參考資料，說明如下。

(1) 5W 指的是何人 (Who)、為何 (Why)、何時 (When)、何地 (Where)、何物 (What)，乃構成任何新聞事件之要素，故專卷也不例外。

(2) 1H 指的是如何 (How)。新聞雖有突發事件，但有的新聞在形成前往往會有伏筆或徵兆，故在製作專卷時應盡可能把如何形成的因素列出。

(3) 1P 指的是過程 (Procedure)。新聞事件的過程是整個專卷議題的主軸，應該描述得愈詳盡愈好。如此一來，使用者在查詢時才較能掌握該議題的來龍去脈，同時也提升了專卷的價值。

(4) 1R 指的是結果 (Result)。新聞專卷來源為報紙新聞，但報紙對某事件的報導有時因外在因素或版面問題，最後不了了之。故整理新聞專卷時，對於有結果但未被報導者，亦應將結果一併紀錄。

(5) 參考資料：新聞專卷中的參考資料除了包含與該議題有關之網站資料、書目資料外，亦應包含相關專卷，使性質相近的專卷相互關聯，形成更完整的知識體系。

## 參、新聞專卷的 Metadata 探討

Metadata 具有定位、探索、文件記錄、評估、選擇的功能，可以用來管理新聞專卷資源，也可以用來描述或辨識新聞專卷內容。不僅有助於檢索資訊，亦有助於資訊管理與系統建置。綜言之，Metadata 有增進網路檢索效益，達到辨識、搜尋及處理資訊的目的。

國際上為了解決新聞資訊交換、包裝和保存的問題，已發展出數種新聞 Metadata 格式，例如，News Industry Text Format (NITF)<sup>9</sup>、News Markup Language (NewsML)<sup>10</sup>、以及 PRISM Aggregator Message (PAM)<sup>11</sup> 等；而台灣近年來在數位典藏方面的研究不遺餘力，也制訂了許多不同領域的 Metadata 標準，以適合中文

<sup>9</sup> IPTC, "News Industry Text Format", available at <<http://www.nitf.org/>>, 2004/7/18。

<sup>10</sup> IPTC, "News Markup Language", available at <<http://www.newsml.org/>>, 2004/7/18。

<sup>11</sup> IDEAlliance PRISM Working Group, "PRISM: Publishing Requirements for Industry Standard Metadata", available at <[http://www.prismstandard.org/PAM\\_1.0/](http://www.prismstandard.org/PAM_1.0/)>, 2004/7/19。

環境使用。<sup>12</sup>其中，與新聞有關的是「新聞紀錄Metadata格式」(News Records Metadata Format，本文簡稱為NRMF)<sup>13</sup>。

這些標準皆本於XML語言，資訊管理者可依需求定義新聞文件的DTD或XML Schema，既規範文件結構又保留擴充彈性；經過XML標示的新聞文件，也可利用程式轉換到其它格式或平台，有助於新聞文件的儲存、檢索以及重複使用。<sup>14</sup>

## 一、NITF

NITF為報業電訊委員會(International Press Telecommunications Council, IPTC)<sup>15</sup>於1990年代初所進行的一項計畫，主要目的是提供一個表示新聞的結構化架構，並可用以標示新聞文件的專門格式。NITF這套標示規格設計的初衷在加強新聞文件重覆使用的便利性，讓通訊社、報業、網站等新聞同業可以透過NITF將新聞稿件轉換成爲不同的資料格式，並呈現在各種新聞平台之上，而達成多重發行的目的。

## 二、NewsML

NewsML同樣是IPTC於1999年開始發展的一個在新聞產生、交換、使用的生命週期中，可用以表示與管理新聞的XML標準。NewsML的概念是伴隨Metadata來涵蓋描述多種不同的媒體(如文字、照片、圖像、錄影資料等)，主要用來描述電子出版、傳送、典藏的新聞文件，並不提供傳統紙本出版的相關描述。<sup>16</sup>其主要功用固然是新聞交換，像是通訊社發稿給報紙或電視台，或媒體之間的新聞傳遞，並可做爲新聞的儲存格式，甚至用在製作、編輯、管理、出版上。因爲其XML的邏輯結構，可與任何實體內容結合，而且是「媒體中立(media-independent)」，對新聞內容的格式不預設立場，所以NewsML可以用在文字、圖片、語音、影視，以及未來發展的媒體型態。<sup>17</sup>

以描述的內容類型而言，NewsML可以分成文字、結構特別的資料、與圖片三部份。文字採用原有的NITF標準及其XML DTD；圖片影像則採用JPEG規格；結構特別的資料則另行設計，如體育新聞的結構就有Sports ML1.0版(草稿)，若遇別的項目而現有格式不敷使用，也可制訂專用標準。基本上，NewsML可將許多標準包裹在一起使用。

---

<sup>12</sup> 行政院國家科學委員會，「數位典藏國家型科技計畫」，可得自 <<http://www.ndap.org.tw>>，2004/7/18。

<sup>13</sup> 行政院文化建設委員會國家文化資料庫知識管理系統，「News Records Metadata Format」，可得自 <<http://km.cca.gov.tw/download/rule.html>>，2004/7/17。

<sup>14</sup> Hall, Richard, "Why XML is Important for Printing and e-publishing?" Online Technology, available at <[http://www.newsandtech.com/issues/2000/05-00/ot/05-00\\_hall.htm](http://www.newsandtech.com/issues/2000/05-00/ot/05-00_hall.htm)>, 2003/9/5.

<sup>15</sup> IPTC成立於1965年，自1970年以來致力於發展國際新聞同業間交換新聞資料的業界標準。

<sup>16</sup> 同註10。

<sup>17</sup> 那福忠，「新聞的標準語言NewsML」，可得自 <[http://www.brainnew.com.tw/Article/na2002/n\\_113002.asp](http://www.brainnew.com.tw/Article/na2002/n_113002.asp)>，2004/4/20。

### 三、PAM

PAM是由IDEAlliance PRISM<sup>18</sup> Working Group於2003年10月所發佈的一個可讓出版者於線上傳送內容給聚集者(Aggregators)和報業組織的新標準。PAM融合了W3C的XHTML標準以及PRISM Metadata，是為滿足雜誌、新聞、期刊等平面媒體出版者的商業需求而設計。<sup>19</sup>更深入探討之，PAM可為出版者、聚集者、訂閱者等三方面的商業需求提供更好的服務。此外，PAM可讓新聞內容應用於多種輸出媒體與產品上，它的應用目的主要包括：(1)二次授權：雜誌出版者為二次傳播將內容授權給資料庫聚集者；(2)雜誌網站的內容描述；(3)傳送內容與Metadata的需求。

### 四、NRMF

NRMF是由文建會所制訂的新聞Metadata格式，主要以都柏林核心集(Dublin Core Metadata Element Set，本文簡稱DC)<sup>20</sup>的15個核心元素為基礎架構，再以子元素(Subelements)延伸出的欄位來著錄更詳盡的新聞資料。由於DC具有簡易、明確、彈性等特色，因此，NRMF也具備了相類似的性質。採用NRMF的優勢是極容易便可與大量採用DC作為Metadata格式的系統進行資料互通與共享。NRMF目前的最新版本是1.2版，於2004/5/24修訂頒佈(前一版1.1-1版頒佈日期是2003/12/18)。<sup>21</sup>由於NRMF的設計是以DC為基準所擴展，因此也具備了DC的簡易、明確、彈性等特色。

## 肆、案例探討 -- 棒球新聞資料庫加值系統實作

為了更明確的說明與印證前述所提的種種概念，我們選擇以聯合報的「台灣棒球運動」專卷為案例，進行系統實作。此批專卷是聯合報資料中心長期整理與珍藏的近三萬個新聞專卷的一部份，詳細的將五十年來台灣棒運的發展，依不同主題、事件、人物...等歸納成「棒球人事」、「棒球比賽」、「棒球場地」、「職業棒球」、「職棒賭博」...等專卷，是研究台灣棒運發展的珍貴文獻，如能加以數位化並進行加值處理，必能建立一個極具價值的棒球主題數位資料館，以供對台灣棒運有興趣的人士檢索利用。

### 一、資料庫結構與系統模組規劃

本研究之實驗系統採用Microsoft Windows 2000 Server 中文版作業系統；Web伺服器則選用Apache Tomcat 5.0；資料庫使用MySQL 4.0.17；程式語言和

---

<sup>18</sup> PRISM是Publishing Requirements for Industry Standard Metadata的縮寫。

<sup>19</sup> IDEAlliance PRISM Working Group, "Guide to PRISM Aggregator DTD V 1.0", available at <[http://www.prismstandard.org/PAM\\_1.0/PAMGuide1\\_0.pdf](http://www.prismstandard.org/PAM_1.0/PAMGuide1_0.pdf)>, 2004/4/20.

<sup>20</sup> Dublin Core Metadata Initiative (DCMI), "Dublin Core Metadata Element Set, Version 1.1: Reference Description", 2003-06-02, available at <<http://dublincore.org/documents/dces/>>, 2004/7/18.

<sup>21</sup> 同註13。

軟體工具有 SUN Java 2 Platform, Standard Edition, v 1.4.2\_03 和 GPL Type-4 (all Java) JDBC driver for MySQL MM.MySQL v 3.0.1。

首先，我們對新聞專卷之資料內容進行分析，規劃出適用的實體關聯圖 (Entity Relation Diagram, ERD)，以作為進一步設計資料庫結構之依據。我們所規劃的棒球資料庫 ERD 如圖 1 所示。

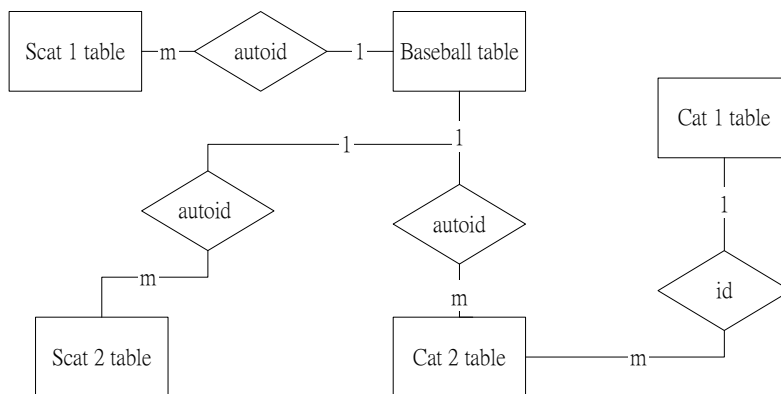


圖 1 棒球資料庫 ERD 圖示

接著，我們進一步設計完整的資料表欄位及彼此的關聯圖，以完成實際的資料庫結構設計，如圖 2 所示。baseball 為單篇新聞表格；cat1 為組織表表格；cat2 為新聞歸屬表格；SCAT1 為主要專卷表格；SCAT2 則為相關專卷表格。

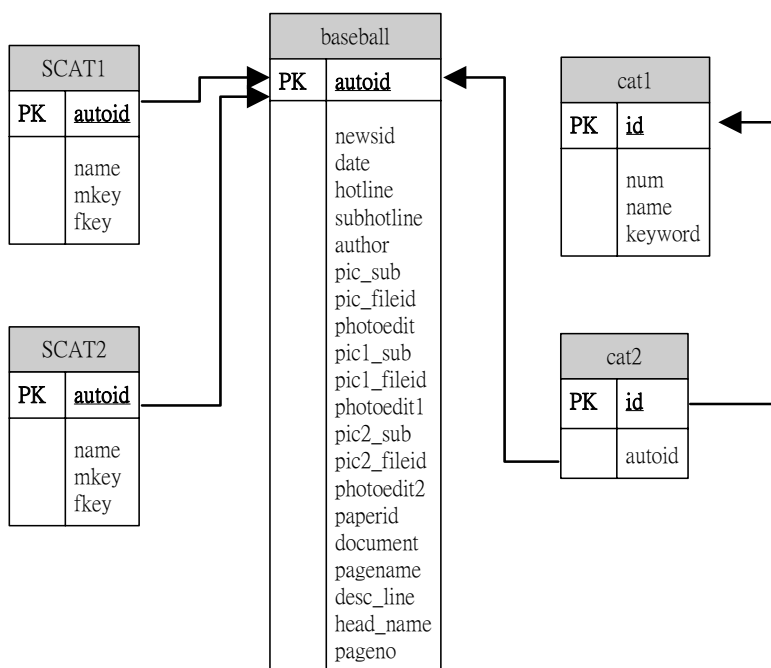


圖 2 棒球資料表關聯圖

為了在全文檢索之外，提供使用者更方便的檢索與管理工具，本研究擬藉由電腦與人腦整合的加值過程，將特定主題的新聞專卷建構彼此關聯的網狀知識體系，以更符合使用者的資訊需求。本系統所規畫的系統模組如下表所示。

表 1 系統模組一覽表

棒球新聞資料庫 Web 瀏覽平台（前端）	棒球新聞資料庫加值管理平台（後端）
台灣棒球報紙新聞前端檢索系統 <ul style="list-style-type: none"> <li>○ 棒球資訊全文檢索模組</li> <li>○ 棒球資訊加值檢索模組               <ul style="list-style-type: none"> <li>* 棒球專卷瀏覽與檢索</li> <li>* 棒球大事記瀏覽與檢索</li> <li>* 棒球名人錄瀏覽與檢索                   <ul style="list-style-type: none"> <li>- 棒球名人大事年表</li> <li>- 棒球名人詳細資料</li> <li>- 棒球名人相關專卷</li> </ul> </li> </ul> </li> </ul> 單篇新聞呈現瀏覽介面（HTML） 單篇新聞呈現瀏覽介面（XML）	台灣棒球報紙新聞後端管理系統 <ul style="list-style-type: none"> <li>○ 內文 Metadata 格式制定模組</li> <li>○ 新聞專卷分類模組</li> <li>○ 新聞專卷檢覈模組</li> <li>○ 新聞專卷管理模組</li> <li>○ 單篇新聞管理模組</li> <li>○ 主要關鍵詞加值模組</li> <li>○ 相關關鍵詞加值模組</li> <li>○ 類別加值模組（棒球名人錄、棒球大事記、棒球大事年表判定）</li> </ul>

## 二、數位化棒球新聞加值步驟

前曾述及，本研究將新聞加值的過程分為「初步加值」與「進階加值」兩個等級。實際作法說明如下：

### （一）初步加值

本文所謂的初步加值主要是將單篇新聞歸入適當專卷的過程。原始單篇新聞各自獨立，彼此並無關聯，傳統作法是以人工方式一則一則歸入適當專卷，使相同議題之新聞合成群組，以方便查找利用。此種加值過程雖然最準確但也最耗時。如今，新產出的新聞皆已是數位資料，在此前提下，本研究提出的初步加值做法是以「新聞專卷分類模組」利用關鍵詞加上布林邏輯運算，以便自動化的將大筆數位化新聞資料歸檔到適當的專卷之中。以我們所要建立的棒球新聞專卷資料庫為例，如「國內青少棒賽」專卷，可以下達如下關鍵詞：

[青少棒賽 or (青少棒 and 比賽)]not 國際

由於這個主題明確，關鍵詞加上布林邏輯運算便可將適當的新聞歸入此專卷中。然而，並非所有專卷都能以此方式準確運作，例如像「棒球球員出路」之類的專卷，由於這個主題比較抽象，在新聞內文中不易出現如主題的關鍵詞，故可能下達的關鍵詞是：

棒球 or 出路 or 前途 or 生涯

如此會查找出新聞內容有棒球或出路或前途或生涯字眼的新聞，其中可能有許多與「棒球球員出路」的主題不切合者，如「XX將和尤敏合作拍片」<sup>22</sup>等不相關的新聞。

一般而言，利用關鍵詞加上布林邏輯運算將新聞自動歸類，其準確度並不高。因此，還需再進一步由管理者檢覈自動分類結果是否正確，並加以篩選、組織與整理，並加註專卷名稱、主要關鍵詞、固定關鍵詞、相關關鍵詞等。其中，

<sup>22</sup> 此處所舉的例子是一則 1961/02/12 與棒球無關的新聞，但內文出現「他是高中時代的**棒球**選手」、「我想這是我**生涯**第一次約會」，故這則新聞在初步加值時會被歸入「棒球球員出路」這個專卷。

相關關鍵詞內含人、事、時、地、物五個要素，方便日後系統運作帶出相關專卷，達到網狀知識組織效果。

## (二) 進階加值

本文所謂進階加值是由專業人士對新聞事件深入解讀、分析、歸納，衍生成更有秩序更有條理的文獻。例如新聞發展的重要大事紀、特殊人物的生平事蹟、相關新聞與一個新聞事件的完整性等。在本實驗系統中，我們實作了「棒球大事紀」、「棒球名人錄」和「相關專卷」的進階加值。做法如下：

第一、對於重大新聞事件，管理界面允許管理者將其加入棒球大事紀，如果該則新聞有許多後續報導，亦可加入相關專卷中。屆時，使用者除可在瀏覽介面的棒球大事紀中看到該則新聞，同時也可連結至與該則新聞相關的專卷進行檢索。

第二、對於重要人物，以棒球名人為例，管理界面允許管理者將其加入「棒球名人錄」中，同時也會登錄在此人物的大事年表中。只要是與此人物有關的報導，可在主要關鍵詞上註明，屆時點選該人物，只會列出有關此人的新聞報導，與全文檢索的意義不同。

第三、對於一個重要人物（棒球名人錄）或是一個重要事件（棒球大事記）的相關新聞報導，管理者可在相關關鍵詞中註明人、事、時、地、物，以關聯至相關專卷。

## 伍、系統實作成果

本節說明系統實作成果，並以未加值的全文檢索與加值後的加值檢索兩者相互比較，以驗證加值成效。系統首頁如圖 3 所示，有「全文檢索」與「加值檢索」兩個選項。



圖 3 台灣棒球報紙新聞數位典藏庫首頁

### 一、棒球新聞全文檢索

我們以查找郭泰源的新聞作為範例。首先，在首頁中選擇全文檢索功能；其

次，下達檢索條件。搜尋字串輸入：郭泰源；日期範圍輸入：1970/1/1～1989/12/30止；查詢範圍全選：聯合報、經濟日報、民生報、聯合晚報、星報。如圖 4 所示。

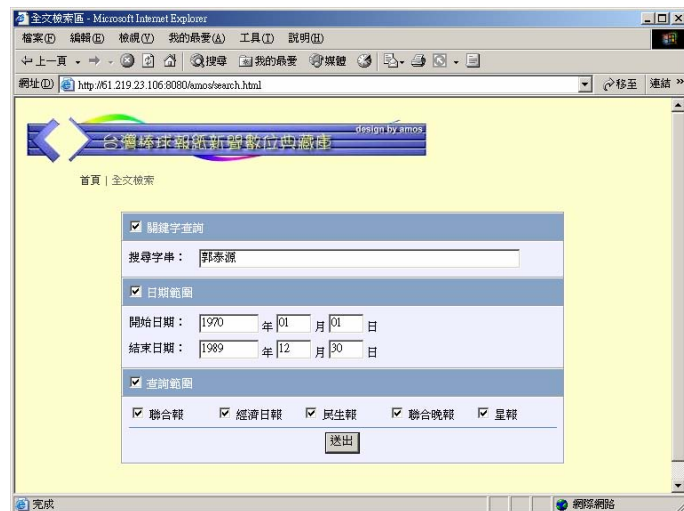


圖 4 全文檢索檢索條件值頁面

全文檢索結果如圖 5 所示，顯示郭泰源資料共有 579 筆，只要文中有郭泰源三字者皆被檢索出來，不論其與郭泰源的關聯性為何，此乃全文檢索以字詞比對為導向的特性使然。

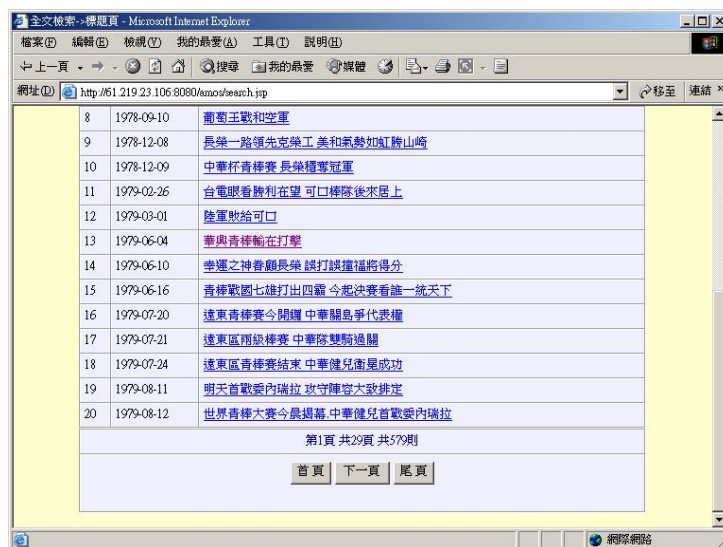


圖 5 全文檢索結果顯示頁面

## 二、棒球新聞加值檢索

加值檢索針對的是專卷資料庫，因已事先經過了系統與人工等加值過程，因此其所得結果準確性理應高於全文檢索。

### (一) 棒球新聞專卷組織表

初步加值以程式將單篇新聞歸檔到適當的樹狀專卷類目中後，再由管理者檢覈，完成篩選、整理工作。圖 6 所示為加值結果，視窗左側呈現棒球新聞專卷組

織表，點選特定專卷可列出該專卷所屬新聞於視窗右側中。

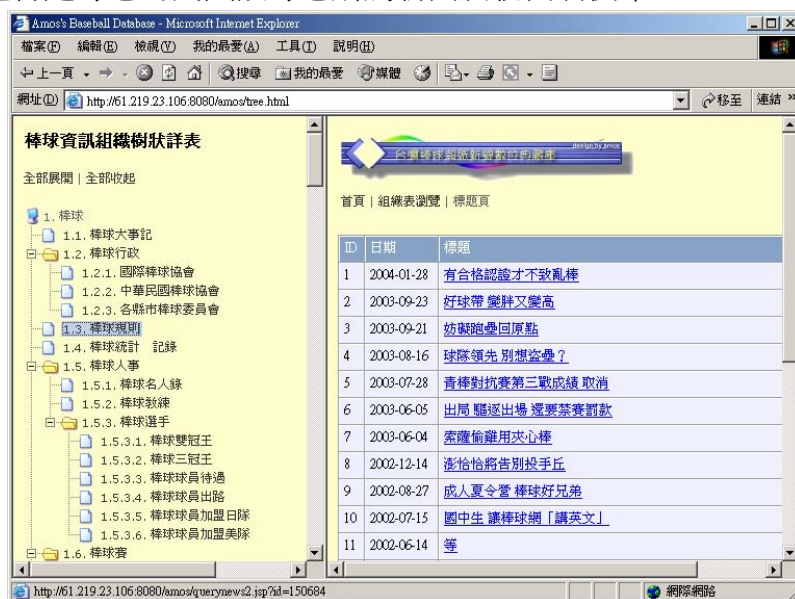


圖 6 棒球新聞專卷組織表呈現頁面

## (二) 棒球大事紀

「棒球大事記」主要是將歷年來棒球重要事件列表呈現，點選大事紀中某筆新聞後可瀏覽該新聞內容，同時也可進一步連結至所屬專卷，如下圖所示。

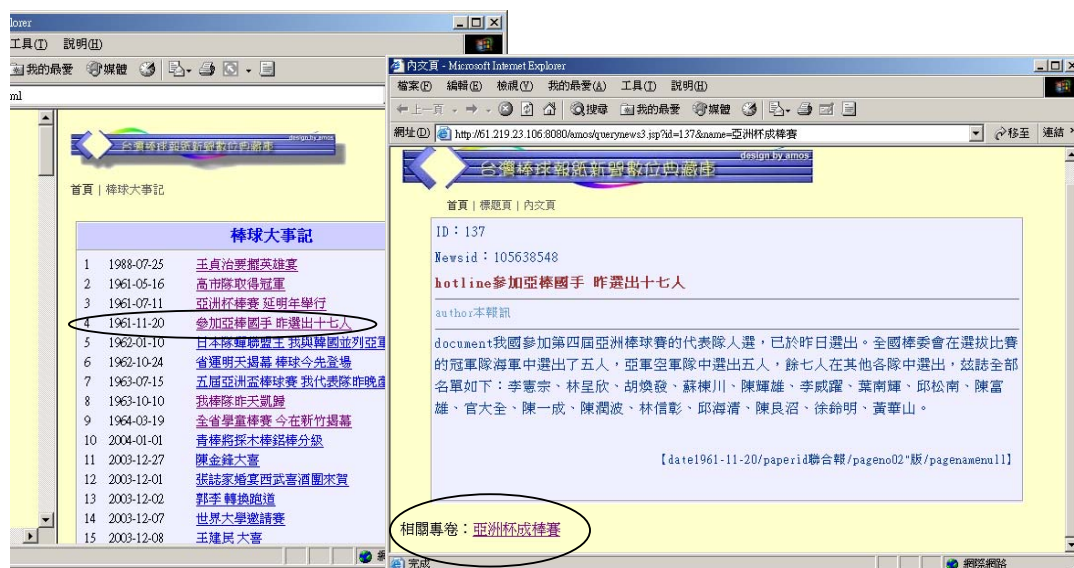


圖 7 棒球大事紀、新聞內容與相關專卷

## (三) 棒球名人錄

「棒球名人錄」主要是將歷年來棒球重要人物列表呈現，又分為「大事年表」、「詳細新聞」與「相關專卷」三個部分。

「大事年表」主要是該棒球名人歷年所發生的重大新聞事件，如下圖所示。以郭泰源為例，其大事年表僅有經篩選過的 26 筆重要新聞，遠比全文檢索的 579

筆精鍊許多。點選大事年表中某筆新聞後可瀏覽該新聞內容，同時也可進一步連結至所屬專卷。

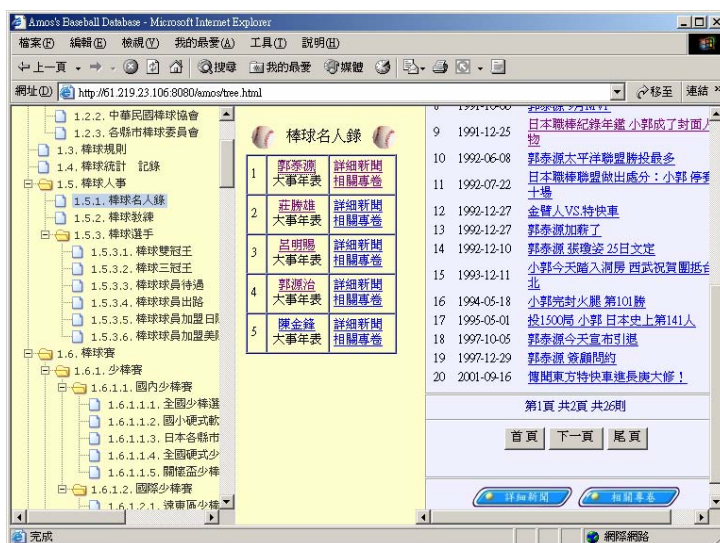


圖 8 棒球名人錄之大事年表呈現頁面

「詳細新聞」則是與某棒球名人真正相關之重要新聞。同樣以郭泰源為例，其詳細新聞畫面中僅列出與郭泰源真正有關的新聞 168 筆，如下圖所示，也遠比全文檢索的 579 筆精鍊許多。

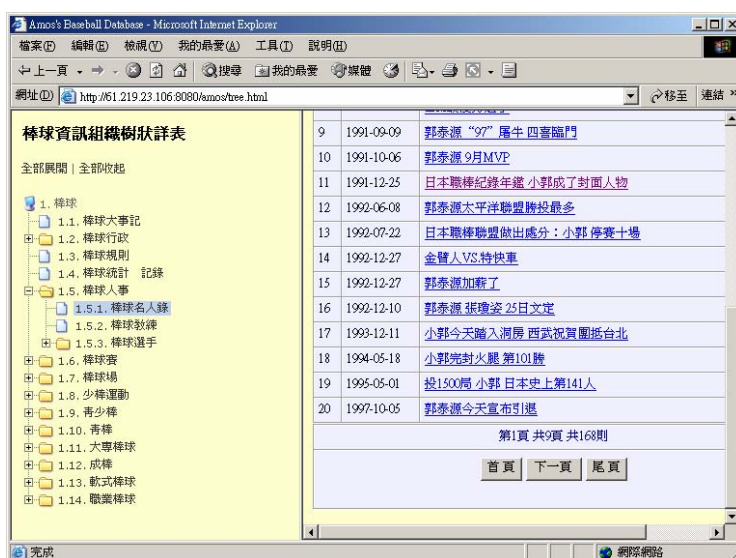


圖 9 郭泰源詳細新聞頁面

「相關專卷」主要是列出與某棒球名人有關的相關專卷，並由這些相關專卷再連結出去，如此可形成網狀的資訊組織結構。同樣以郭泰源為例，其相關專卷有「張瓊姿」、「張純芳」、「二郭一莊」... 等，如下圖所示。

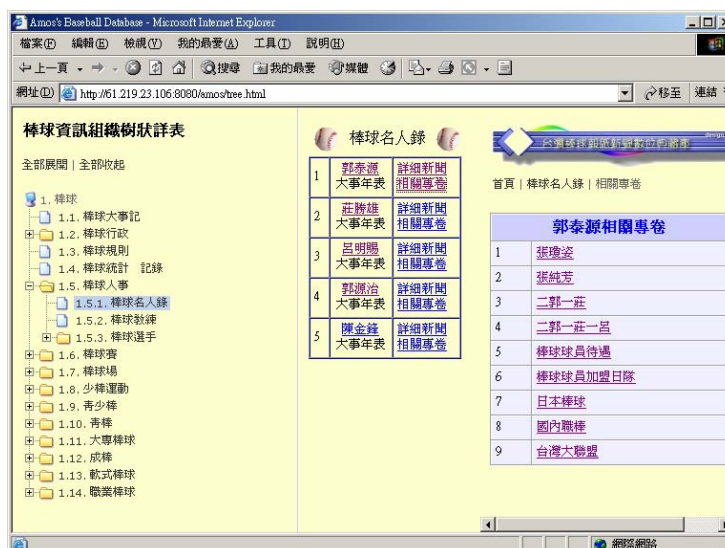


圖 10 郭泰源的相關專卷

### 三、Metadata 轉出功能

本系統亦提供新聞 Metadata 轉出功能，目前之轉出格式初步依據 NRMF 加以增刪修訂，未來將再陸續探討與 NITF、NewsML、PAM 等規格之相容性議題，故本系統目前轉出為測試版之 Well-Formed XML 文件。

```

- <description>
- <newsProfileDescription>
  <newsProfileTitle>棒球球員待遇</newsProfileTitle>
  <newsProfileClassification>1.5.3.3</newsProfileClassification>
  <newsProfileAbstract>棒球球員薪水、福利、工作環境等
  </newsProfileAbstract>
</newsProfileDescription>
- <newsContent>
  <personGroupName>郭泰源</personGroupName>
  <personGroupName>王貞治</personGroupName>
  <personGroupName>西武隊</personGroupName>
  <personGroupName>日本巨人職業棒球隊</personGroupName>
  <abstract>金臂投手郭泰源昨天在接受本報記者訪問時說，他在前天晚間已和西武隊簽下了「協議書」。郭泰源說，他和西武隊簽下的「協議書」，並非正式合約，其中也未註明簽約金的數目，他認為這並不影響到他的業餘資格。</abstract>
  <Introduction />
  <content>金臂投手郭泰源昨天在接受本報記者訪問時說，他在前天晚間已和西武隊簽下了「協議書」。郭泰源說，他和西武

```

### 陸、結論

眾所周知，全文檢索功能強大，但卻常對使用者造成資訊溢載的困擾；而經過人工篩選、粹煉、加值的資訊，往往能更精準的滿足使用者需求。新聞專卷原為報業極重要之加值處理過程，然而，在數位化新聞與全文檢索技術興起之後，反而不受重視。本研究旨在喚起新聞重業人員重新重視新聞專卷加值的重要性與方便性。

不過，我們要強調的是本研究非但未否定全文檢索技術之功用，反而相信其將來必然朝向更智慧化的方向發展，以滿足更廣泛的檢索需求。只是電腦能力畢竟有其侷限性，因此除了發展智慧型檢索系統之外，亦不應忽略人工加值的重要性。將數位化資料分類歸檔或以Metadata加註額外的語意描述，可視為提升電子文件智能的手段之一，智慧型系統與智慧型文件若能有效整合，必能提供使用者更便利的資訊服務。<sup>23</sup>本文藉由系統實作，也證實了有效結合電腦技術與人工加值，不但可突顯出真正重要的新聞，亦能更精確的查找出滿足使用者需求的新聞，同時也將主題相近的新聞集結一起，方便使用者瀏覽、檢索，更可旁徵博引，連結更多的相關新聞資訊。

科技無垠的前進，人類對電腦的需求與日俱增，已達無可取代的地步，有時反而忽略了「人」的重要性。然而，電腦僅是人類完成偉大夢想的輔助工具，文明要更臻完美，「人」應扮演比電腦更重要的角色才是。當然，數位化新聞專卷的加值是必須付出人力、物力、金錢與時間等代價的，未來研究者應思考的是如何更有效率的結合人腦與電腦的優點，以創造更具利用價值的數位化資訊。

## 誌謝

承蒙國家科學委員會經費補助（計畫編號：NSC 93-2422-H-032-001），使本研究得以順利進行，特此致謝。

---

<sup>23</sup> 林信成，「智慧型文件與智慧型系統整合之研究」，教育資料與圖書館學，第四十卷，第四期，頁 481-495，2003 年 6 月。