2003 資訊技術應用與發展研討會，屏東商業技術學院，頁 42-51，2003.6.28

模糊理論與標示語言在電子新聞管理系統之應用

陳勇任*　　林信成**　　蕭勝文**
*淡江大學教育科技學系
**淡江大學資訊與圖書館學系
sclin@mail.tku.edu.tw

摘要

本研究結合模糊理論與全文標示兩種不同的作法，達到提升檢索系統回現率與精確率之目的。我們以電子新聞作為實驗對象，主要研究成果有二：(1) 針對新聞資料內容進行軟性的模糊分類，依據模糊理論的概念賦予每筆新聞在各新聞類別上的模糊歸屬函數值，以便檢索系統能依據新聞資料之歸屬度進行模糊關聯搜尋，使得檢索方式更具彈性，亦可提供較佳的回現率；(2) 針對新聞資料內容以 XML 格式進行全文標示，利用 XML 進行語意定義與描述，使資料內容具備自我描述性，以便檢索模組進行較精確的語意檢索，使檢索結果能夠達到較高的精確率，以更符合使用者的資訊需求。

關鍵字：模糊理論、模糊分類、全文標示、XML、Fuzzy、電子新聞管理系統

壹、前言

網路的發達改變了人們取得資訊的方式及習慣，以往得在浩瀚的書籍與報章之間奔波查找，現今透過網路只要在彈指之間即可取得數量龐大的資訊。然而，當大量電子文件透過網路出版、傳播之後，如何組織這些龐大的資訊以及提供有效的檢索機制，便成為數位典藏與電子出版等相關研究領域所考慮的重要課題之一。一般而言，回現率（Recall Rate）和精確率（Precision Rate）是評估檢索效能的兩大重要指標[1]。回現率指的是與查詢條件相關的資料被檢索出來的比例；而精確率則是指被檢索出的資料與查詢條件相關的比例。這兩者經常成反比關係，如同魚與熊掌無法兼得。

本研究以電子新聞資料庫管理系統為例，提出一個結合模糊理論（Fuzzy Theory）與標示語言（Markup Language）的作法，達到提升回現率與精確率之目的。透過實驗研究的方法，完成二項主要成果：首先，為了提升資訊檢索之回現率，針對資料進行軟性的資料分類，依據模糊理論的概念賦予每筆新聞資料在各新聞分類上的模糊歸屬函數值，我們並發展一個「模糊檢索引擎」(Fuzzy Search Engine)，可針對所有新聞資料錄所歸屬的模糊集合，依據使用者的檢索條件進行資料的搜尋，使得其檢索的方式更具彈性，可提供較佳的回現率；再者，為了提升資訊檢索之精確率，本研究採用國際公認標準的可擴展標示語言（eXtensible Markup Language，簡稱 XML)，針對新聞資料內容進行全文的標示，利用 XML 賦予資料自我描述性的特質，進行資料內容定義與語意描述，並發展一個「語意檢索引擎」(Semantic Search Engine)，使用者進行資料檢索時可針對每筆新聞資料進行內文語意檢索，使得檢索的結果能夠達到更高的精確率。

本文的架構如下：第一節為前言；第二節針對模糊理論與 XML 標示語言加以回顧；其後第三節探討分類模式與提出本研究之新聞資料模糊分類法，探討 Fuzzy 在電子新聞管理系統之應用；第四節提出以 XML 進行新聞資料全文標示之作法；第五節為系統實驗成果與展示，印證模糊檢索引擎與語意檢索引擎分別在回現率和精確率二方面

所努力的成果；第六節為結論及建議。

貳、模糊理論與全文標示

一、模糊理論

模糊理論實際上是模糊集合（Fuzzy Set）、模糊關係（Fuzzy Relation）、模糊邏輯（Fuzzy Logic）、模糊控制（Fuzzy Control）、模糊量測（Fuzzy Measure）… 等理論之泛稱[2]，是一門用以將模糊概念量化的學問，起源於 1965 年扎德（L.A. Zadeh）教授所發表的著名論文－「模糊集合」[3]。模糊理論以模糊集合為基礎，以研究不確定事物為目標，接受模糊現象存在的事實，根據不清晰訊息，透過近似推理（Approximation Reasoning）[4]過程而得到正確結果，這與人腦「過程模糊，結論清晰」的思維方式極其類似，因此已被廣泛的應用於各種不同領域的智慧型系統中[5]。

傳統明確集合（Crisp Set）的特徵函數（Characteristic Function）採用非 0 即 1 的二分法，而模糊集合的基本精神則是將其擴展成由 0 至 1 的任何值，稱為歸屬函數（Membership Function），當一個元素屬於某集合的程度越大時，其歸屬程度就越接近於 1，否則越接近於 0。若以 X 代表論域（Universe of Discourse），A 代表 X 中的任一離散型模糊集合，則可將 A 表示成如下型式：

$$A=\sum_{i=1}^{n} m_i / x_i = m_1 / x_1 + m_2 / x_2 + \ldots + m_n / x_n \quad (1)$$

其中，$x_i \in X$ 為 X 中任一點。藉由歸屬函數對模糊概念進行量化之後，便可以利用精確的數學方法進行模糊資訊的分析和處理了。

二、全文標示

Web 是網路時代最重要的資訊傳播平台，而 HTML 則是發行 Web 電子文件的標準規範。然而，HTML 擅長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相對不足；XML 的誕生正好提供了一個可行的解決方案，彌補 HTML 之短。XML 自 1998 年日 2 月 10 正式標準[6]發佈至今，歷經五年多的發展[7]，已經成為一個陣容龐大的技術家族：DTD 和 XML Schema[8]用以定義文件的結構；CSS[9]和 XSL/XSLT[10]分別作為呈現文件版面和轉換文件格式之用；DOM[11]是剖析文件時的標準物件模型；RDF[12]則作為 Metadata 之整合框架；Nmaespace[13]是一致性名稱識別機制；以及各種衍生的應用語言，例如數學專用的 MathML[14]、向量繪圖的 SVG[15]、可攜式網路圖形的 PNG[16]、同步多媒體的 SMIL[17] … 等，再加上由各個不同組織基於 XML 所發展出適用於各行各業的應用語言將近千種[18]，真可謂族繁不及備載。至於近期 XML 的研究則逐漸朝向智慧型的 Web 語意網（Semantic Web）[19]和 Web 知識體（Web Ontology）[20]的方向發展，使得 Web 系統與文件皆愈來愈智慧化。

以 XML 進行全文標示（Full-Text Markup）的電子文件不但內容和外觀可分離處理，而且具備了結構性、整合性和自我描述性等重要特色，不但能使網路上的電子文件更有組織，同時亦有助於開創智慧型電子出版的新契機。

？、電子新聞模糊分類

分類（Classification）是知識管理（Knowledge Management）的基礎工程之一，各行各業在自己的專業領域中隨時隨地都在進行著分類工作。分類規則有時必須遵循共通的規範，有時則以自訂的方式為之；而分類模式又可分為單一主題明確分類模式與多重主題模糊分類模式。

一、單一主題明確分類模式

單一主題分類模式是將被分類對象依主題範圍歸屬於單一類別中，圖書館的圖書分類法可為代表。「圖書分類」是依主題或性質將圖書分門別類，使讀者可以根據索書

號找到所需要的圖書。圖書館的圖書分類規則通常遵循既定的共通規範，國外以「美國國會分類法」（Library of Congress Classification：LCC）[21]或「杜威十進分類法」（Dewey Decimal Classification：DDC）[22]使用者居多，國內則以「中國圖書分類法」（New classification Science for Chinese Libraries：CCL）[23]使用者最多。

「新聞分類」簡單的說，就是將各類不同主題性質之新聞分門別類，然而，新聞分類方式目前並無共通的規範，大都是由新聞業者自訂，因此，該如何分類可能會涉及到主觀性的判斷而有所差別。以幾個較知名的電子新聞網站為例，年代 TVBS 分為：台灣政治、兩岸三地、台灣社會、財經新聞、娛樂新聞 … 等[24]；聯合新聞網分為：國內要聞、兩岸國際、地方新聞、財經產業、體育賽事…等[25]；東森新聞報分為：政治要聞、紫禁城、大城小調、股市理財、運動場等[26]… ；中時電子報分為：政治、大陸、地方、股市財經、運動…等[27]。

不同的新聞業者都有屬於自己的分類方式，這是因為新聞內容廣泛，主題多變，天生具備極大的「模糊性」。然而，對於牽涉較廣以致涵蓋二種以上主題範圍的新聞，分類者在進行分類時，通常必須依賴個人主觀判斷，硬將該篇新聞歸類至某個單一類別中，這是採用「明確集合」的二分法將資料強制歸類，抹煞了新聞內容因多重主題而與生俱來的模糊性質，也增加使用者在查找資料時的困難。茲以「陳水扁總統參加中華職棒十四年開幕儀式致詞」為例，此則新聞至少可歸屬於「政治」、「體育」兩類別，若將其歸屬於「政治」類，則在「體育」類中便無法找到此篇報導；反之，若將其歸屬在「體育」類，則在「政治」類又找不到，導致分類查找時的回現率降低。

由此可知，若要使讀者在分類查找新聞時能找到更多相關資料以增加回現率，可捨單一主題分類法而採用多重主題分類法。而模糊分類不但是有效的多重分類法，更是一種軟性分類法，可增強不同類別的關聯性，使分類結果更平順自然，也更符合一般人的思維模式。

二、多重主題模糊分類模式

若被分類對象涵蓋不只一個主題，單一主題分類模式便不適用。此時，採用多重主題分類模式的模糊分類法便可派上用場。

假設某新聞資料庫中共有 n 篇新聞，其中第 i 篇記為 d_i，並以 D 表示所有新聞的集合，即 $d_i \in D, i = 1,2,..., n$；現有 m 個新聞類別，第 j 類記為 c_j，並以 C 表示所有類別的集合，亦即 $c_j \in C, j = 1,2,..., m$。若將 d_i 視為一個定義在論域 C 上的離散型模糊集合，且將 d_i 歸屬於某個類別 c_j 的程度以 m_{ij} 表示，則此一模糊集合可記為：

$$d_i = \sum_{j=1}^{m} m_{ij} / c_j \qquad (2)$$

亦即：

$$d_1 = \mu_{11}/c_1 + \mu_{12}/c_2 + \ldots + \mu_{1m}/c_m$$
$$d_2 = \mu_{21}/c_1 + \mu_{22}/c_2 + \ldots + \mu_{2m}/c_m$$
$$\vdots$$
$$d_n = \mu_{n1}/c_1 + \mu_{n2}/c_2 + \ldots + \mu_{nm}/c_m$$

舉例而言，$d_1 = \mu_{11}/c_1 + \mu_{12}/c_2 + \ldots\ldots\mu_{1m}/c_m$ 代表 d_1 此則新聞在 c_1 這一類別裡的歸屬值是 μ_{11}，在 c_2 這一類別裡的歸屬值是 μ_{12}，在 c_m 這一類別裡的歸屬值是 μ_{1m}。若將其歸屬函數畫成離散式模糊集合，則如圖 1 所示。圖中顯示出 d_1 新聞分別在 c_1、$c_2 \ldots c_m$ 類別中所佔的歸屬函數值高低，愈接近於 1，表示屬於某類別的程度愈高，反之，愈接近於 0，表示屬於某類別的程度愈低。

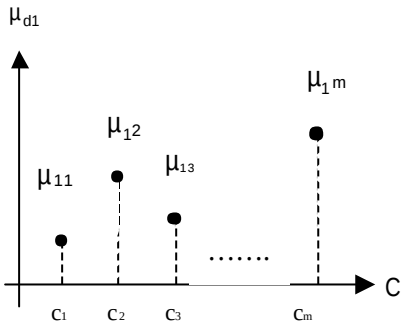

圖 1 將新聞以離散型模糊集合表示

再者，我們可將各則新聞(d_1~ d_n)在各個類別(c_1~ c_m)中分屬的歸屬函數值(μ_{11}~μ_{nm})，以一個 $n \times m$ 的模糊關係矩陣表達如下：

$$\begin{array}{c} \\ d_1 \\ \vdots \\ \vdots \\ \vdots \\ d_n \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_m \end{array} \\ \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1m} \\ \vdots & \vdots & & \\ \vdots & \vdots & & \\ \vdots & \vdots & & \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nm} \end{pmatrix} \end{array}$$

圖 2 新聞與類別之模糊關係矩陣

由此可知，d_i 與 c_j 之間具備一種「模糊歸屬關係」，從語意的觀點而言可描述成「d_i 歸屬於 c_j 的程度為 m_{ij}」。因此，上述的模糊關係矩陣實際上等同於下圖所示的模糊語意網路。我們可以很清楚的看出，模糊歸屬函數值在此網路中，扮演了一個將相關概念以不同的強度加以連結的重要角色。

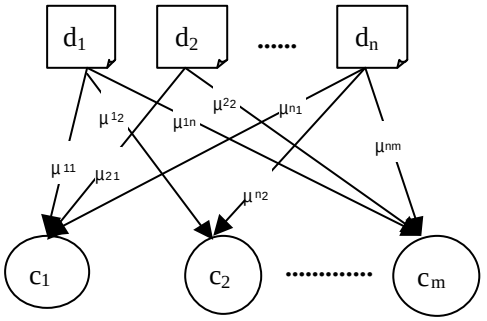

圖 3 新聞與類別之模糊語意網路

三、實例探討

為了比較上述單一主題明確分類法與多重主題模糊分類法的差異，我們實際從某個著名新聞網站中下載了兩則新聞作為範例，並將其標題、內容及類別整理如表 1 所示。從表中可知，在此新聞網站中，分類者將「布希要求 750 億美元的戰費」歸類於「財經產業」類，而將「一瑞士銀行將遭凍結伊拉克資金交美國」歸類於「兩岸國際」類。

編號	d_1	d_2
標題	布希要求 750 億美元的戰費	一瑞士銀行將遭凍結伊拉克資金交美國
內容	美國白宮 24 日表示，布希總統今年將向國會提出 747 億美元的額外支出要求，以支應伊拉克戰爭、人道援助與國土安全的開支。	一家瑞士銀行把它凍結的一筆伊拉克資金交給美國。UBS 銀行只說這筆錢從 1990 年起就遭到凍結，但沒說出數額。美國政府已經宣佈，它要將伊拉克存在美國的將近二十億美元沒收。
類別	財經產業 (c_1)	兩岸國際 (c_2)

表 1 兩則範例新聞

如果我們分別以 d_1、d_2 代表這兩則新聞，而以 c_1、c_2 代表這兩個類別（參見上表），則此新聞網站的分類方式顯然是採用傳統的單一主題明確分類法，我們可用下表的明確集合特徵函數值表示之。

新聞\特徵函數\類別	c_1 (財經產業)	c_2 (兩岸國際)
d_1 (布希要求 750 億美元的戰費)	1	0
d_2 (一瑞士銀行將遭凍結伊拉克資金交美國)	0	1

表 2 明確分類特徵函數值

此表可寫成如下的明確關係矩陣（Crisp Relation Matrix）：

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

然而，若仔細研讀以上新聞的標題或內容，

可發現將「布希要求 750 億美元的戰費」分類至「兩岸國際」，或許比「財經產業」來得讓使用者更容易取得此則報導；而另一則新聞「一瑞士銀行將遭凍結伊拉克資金交美國」，假使分類至「財經產業」也許會比原本的「兩岸國際」更適合！適當的給予每則新聞分類是每位新聞分類者的責任，當然分類過程中多少牽涉到個人對於新聞的主觀性判斷。對於具備模糊主題性質的新聞通常很難以單一主題方式明確分類，若採用模糊分類法，在詮釋內容意涵後，給予適當歸屬函數值便可解決這類型問題。例如，我們可以如下表所示，透過模糊歸屬函數值的指定，使得 d_1、d_2 這兩則新聞，在不同的輕重程度上分屬 c_1、c_2 兩個類別，如此一來，透過數值的高低判別其歸屬程度，更接近一般人的思考模式。

新聞\歸屬函數\類別	c_1 (財經產業)	c_2 (兩岸國際)
d_1 (布希要求 750 億美元的戰費)	0.8	0.5
d_2 (一瑞士銀行將遭凍結伊拉克資金交美國)	0.6	0.9

表 3 模糊分類歸屬函數值

此表可寫成如下的模糊關係矩陣（Fuzzy Relation Matrix）：

$$\begin{bmatrix} 0.8 & 0.5 \\ 0.6 & 0.9 \end{bmatrix}$$

肆、電子新聞全文標示

利用模糊分類可以擴展新聞之間的關聯性，進而提升檢索時的回現率。而在提升檢索精確率方面，我們則是藉由對電子新聞進行全文標示的方式進行。由於 XML 非常適合 Web 環境，且具備優越的結構化與自我描述性，能使電子文件更「智慧化」，因此我們選擇以 XML 作為標示語法。實際的作法說明如下：

首先，必須制訂一組標示用的 Metadata。在參考國內外的新聞 Metadata，包括政大謝瀛春教授發表的有關科學新聞的內容標示[28]，及國際上的兩大主流 NITF（News Industry Text Format）[29] 與 NewsML[30]之後，基於本研究實際需求之考量，我們認為國際標準過於複雜，因此化繁為簡自訂了一個適用於本系統的簡易版新聞 Metadata DTD。其中，每則新聞除了包含諸如新聞編號、索引、日期、標題、作者 … 等基本資料，另將新聞內容分以人、事、時、地、物等加以標示，以便進行內文語意搜尋之用，最後，尚含有一個排版用的標籤作為不同樣式之套版。接著，便以 XML 語法進行新聞資料之全文標示，且在新聞檢索模組中加入依標示語意進行檢索之機制，測試加註了這些描述性資料之後的檢索結果，探討與印證此法對於提升資訊檢索之精確率是否有實質之助益。

伍、系統實作

一、系統架構與功能說明

為了印證上述論點，我們採用實驗法進行系統實作。實驗的重點有二：(1) 針對資料進行軟性的模糊分類，依據模糊理論的概念賦予每筆新聞資料分屬各新聞類別的模糊歸屬函數值，擴展新聞資料的關聯性，使得「模糊檢索引擎」可對資料進行更廣範泛的關聯搜尋，藉以達到較佳的回現率；(2) 針對新聞資料內容進行全文標示，利用 XML 具備自我描述性的特質，對資料內容進行更精確的語意定義與詮釋，使得「語意檢索引擎」可針對每筆新聞資料進行語意檢索，以達到更高的精確率，滿足使用者的資訊需求。

本研究所設計之電子新聞管理系統，可分為前端使用者子系統與後端管理者子系統兩部分，如圖 4 所示。

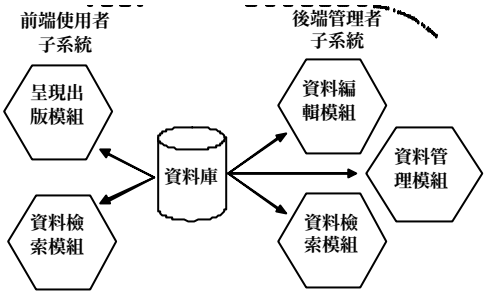

圖 4 系統功能模組

若依功能劃分則可區分為五大部分，分別是：(1) 呈現模組（Presentation Module）；(2) 檢索模組（Searching Module）；(3)編輯模組（Editing Module）；(4) 管理模組（Management Module）；(5) 資料庫（Database）。主要功能說明如下：

(1) 呈現模組

本模組負責將資料內容呈現給使用者，主要負責呈現使用者經過檢索後所需閱讀的資料。由於本系統內之文件資料儲存方式以 XML 為主，因此本模組具備解讀 XML 文件的功能。此模組透過 DSO 以及 DOM 來解讀 XML 文件，並結合該資訊所需之相關功能及超連結，加以包裝、排版，只要相容於本系統之 DTD 規範的 XML 文件，皆可透過此模組呈現內容。

(2) 檢索模組

本模組提供新聞資料檢索功能，主要由「模糊檢索引擎」與「語意檢索引擎」所構成：

(a) 模糊檢索引擎：提供新聞分類的模糊檢索功能，使用者可決定欲查詢之新聞資料其所屬分類的歸屬度(0 至 1)，模糊檢索引擎針對所有新聞資料錄之歸屬函數值，在其對應的模糊集合當中依據使用者的檢索條件進行搜尋，使得其檢索的方式更具彈性，以提供較佳的回現率。

(b) 語意檢索引擎：傳統的搜尋引擎僅提供資料欄位的檢索功能，並未提供針對內容語意上之特定目標加以檢索，如人名、地名等，本模組除提供傳統的欄位與關鍵字詞檢索功能之外，藉由 XML 具備資料自我描述之特性，可針對新聞內容的人、事、時、地、物等內文語意加以檢索，提高檢索結果的精確率。

(3) 編輯模組

本模組提供管理者編輯新聞文件內容。透過本模組可將新聞內容編輯成符合系統之 DTD 規範的 XML 文件，並針對新聞內容給予人、事、時、地、物不同的語意標示，以及可針對新聞系統內各新聞分類給予歸屬函數值，以利提供檢索模組進行內文語意檢索以及模糊分類檢索等功能。另外經由 XML 文件內容與呈現資料分離的特點，同一份文件可選擇不同樣式來排版。

(4) 管理模組

本模組提供管理者異動／修改資料。透過系統之資料管理模組，可針對新聞資料做新增、刪除、修改等各項異動，並且內建資料檢索功能，以利於管理者於後端進行資料維護時提供便捷的資料查詢功能。

(5) 資料庫

關聯式資料庫（Relational Database）是目前最通用的資料庫系統，因此，在本研究中我們採用關聯式資料庫系統。我們依據上一單元所分析的模糊分類模式，可以很容易的將模糊關係矩陣轉換為關聯式資料表的模糊分類欄位，如表 4 所示，並給予歸屬函數值，使得檢索模組可藉以判別新聞所屬之模糊類別。

	c_1	c_2	…	c_m
d_1	μ_{11}	μ_{12}	…	μ_{1m}
d_2	μ_{21}	μ_{22}	…	μ_{2m}
⋮	⋮	⋮	…	
d_n	μ_{n1}	μ_{n2}	…	μ_{nm}

表 4 模糊關聯式資料表

二、系統建置與實驗結果

本系統建構於 Microsoft Windows 2000 Advance Server（NT 技術平台）之上：Web 伺服器為 Microsoft IIS，中介軟體及各模組使用 ASP（Active Server Pages）程式語言開發，至於後台資料庫系統則採用 Microsoft SQL Server。使用者端則可以使用支援 XML（如 Microsoft Internet Explorer 5.0 以上版本）之瀏覽器，進行連線；新聞分類則分為：『政治類』、『社會類』、『國際類』、『財經類』、『生活類』等五類。

此系統的運作流程如圖 5 所示，分為使用者與管理者兩種流程。

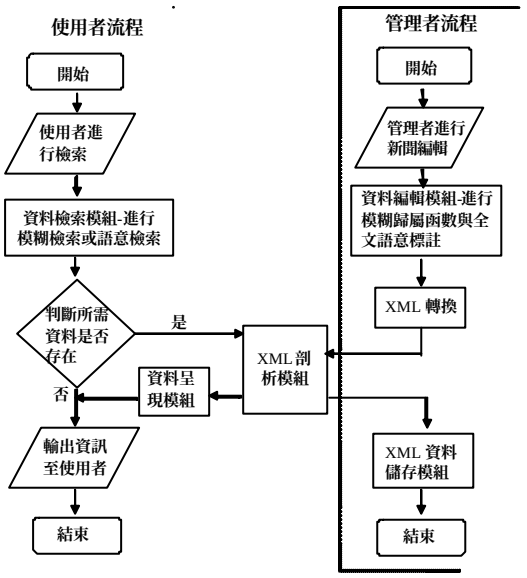

圖 5 系統運作流程圖

使用者可經由資料檢索模組，輸入欲查詢之條件，而且可選擇搜尋模式為模糊檢索或語意檢索，如圖 6 所示，隨後資料檢索模組便依據使用者輸入至資料庫中搜尋所需資料。檢索模組的檢索工作完成後，由資料庫取出之資料便交由資料呈現模組，依照系統所訂定之 DTD 來確認 XML 資料的合法性，隨後將呈現給使用者。呈現的 XML 資料透過 DSO 以及 DOM 的解讀，可在不更改原始新聞內容之下，套上各種不同的排版樣式，如圖 7 所示。

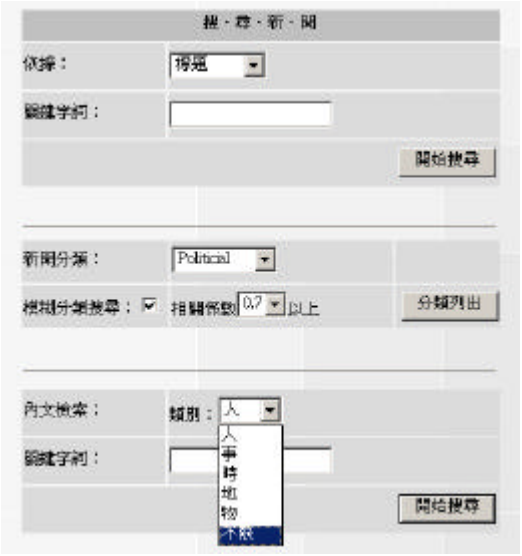

圖 6 資料檢索模組

圖 7 各種排版樣式

(1) 模糊檢索結果

模糊檢索引擎的目的在於提供較佳的回現率。以本系統資料庫內之新聞『今起台灣民眾赴港可辦電子簽證節省時間金錢』為例，由於內容描述香港政府之旅遊發展局等相關單位有鑑於台港旅遊人次繁多，簽證手續繁雜，因此推出新的簽證方式以節省時效。在實質上可歸屬於『生活類』的新聞；但此篇新聞內容亦與香港政府相關，因此與『國際類』亦有若干關聯，若以一般傳統的單一類別的分類法，直接將此篇新聞歸屬於生活類，不但有欠周詳，也會造成使用者在『國際類』中找不到此篇報導的問題。因此，如果我們適度的給予此篇新聞模糊歸屬函數值，使其歸屬於『生活類』與『國際類』

的程度分別為 0.9 與 0.7，則藉由模糊分類檢索功能，使用者不但可以在『生活類』中檢索出這篇新聞（如圖 8 所示），亦可在『國際類』中檢索到（如圖 9 所示），不但增加了檢索的彈性，也提升了回現率。

圖 8 檢索『生活類』新聞之結果

圖 9 檢索『國際類』新聞之結果

此外，我們也在模糊分類檢索的功能中提供使用者自由選擇相關係數（即模糊歸屬函數值）之選項，以利使用者篩選新聞分類的相關度，相關係數愈高表示歸屬程度必須愈高者才會被檢索出來，如圖 10 所示。

圖 10 模糊分類檢索功能

(2) 語意檢索結果

本系統之新聞資料在經過 XML 標示加值處理之後，使用者除可依據標題、作者、時間 … 等資料欄位進行檢索外，也可針對新聞全文內容，依據人、事、時、地、物等語意條件進行更精確的內文語意檢索。例如，圖 11 乃是以「大學」作為關鍵字詞，選擇以不限標示的方式進行檢索之結果，共找到八篇文章；對於同樣的檢索詞，如果將內文檢索條件限制於「地」，表示使用者欲檢索之條件僅為內文語意上與「大學」相關的地方、地名或地點，而非所有與「大學」概念相關的文章，結果如圖 12 所示，找到七篇符合內文語意檢索條件之文章；再者，若檢索詞仍為「大學」，但將內文檢索條件限制於「事」，即表示使用者欲查詢之條件僅為有關大學的「事件」而非地點或其他，則如圖 13 所示，更精確的只找出兩篇在內文語意上含有大學相關事件的文章。

圖 11 不限檢索標示之結果

圖 12 限定檢索標示為「地」之結果

圖 13 限定檢索標示為「事」之結果

陸、結論與建議

本研究提出了一個結合模糊理論與全文標示之實際作法，並發展一套具有模糊分類搜尋與精確語意檢索的電子新聞管理系統。主要研究成果之一是藉由 Fuzzy 理論的技術，以歸屬函數的高低來判別各則新聞所屬類別輕重，並發展出具有模糊搜尋的檢索功能；另一個成果是藉由 XML 進行新聞內容的語意描述，各個進行資料處理及交換的模組皆以 XML 為基礎，系統內之所有資料亦採用 XML 格式，並發展出具有精確語意的檢索功能。由實驗結果可以清楚的看出結合此兩種不同作法，的確可以有效的提升回現率與精確率。

誌謝

本研究承蒙國科會計畫編號 NSC 91-2413-H-032-007 所支持，使研究得以順利進行，特此致謝。

參考文獻

[1] Robert R. Korfhage, *Information Storage and Retrieval*, Wiley Computer Publishing, New York, 1997, pp. 196-199.

[2] 林信成、彭啟峰，*Oh！Fuzzy 模糊理論剖析*，台北：第三波，民 83。

[3] L. A. Zadeh, “Fuzzy sets,” *Information and Control* **8** (1965), pp. 338-353.

[4] L. A. Zadeh, “Outline of a new approach to the analysis of complex systems and decision processes,” *IEEE Trans. on Syst., Man and Cybern*. SMC-**3** (1973), pp. 28-44.

[5] Jyh-Shing Roger Jang, Chuen-Tsai Sun, Eiji Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall (1996).

[6] W3C, "Extensible Markup Language (XML)", available at <http://www.w3.org/XML/> (20 Feb. 2003).

[7] W3C, "Happy Fifth Birthday to XML-10 February 2003", available at <http://www.w3.org/> (20 Feb. 2003).

[8] W3C, "XML Schema", <http://www.w3.org/XML/Schema> (20 Feb. 2003).

[9] W3C, "Cascading Style Sheets", <http://www.w3.org/Style/CSS/> (20 Feb. 2003).

[10] W3C, "The Extensible Stylesheet Language (XSL)", <http://www.w3.org/Style/XSL/> (20 Feb. 2003).

[11] W3C, "Document Object Model (DOM) ", <http://www.w3.org/DOM/> (20 Feb. 2003).

[12] W3C, "Resource Description Framework (RDF) ", <http://www.w3.org/RDF/> (20 Feb. 2003).

[13] W3C, "Namespaces in XML", <http://www.w3.org/TR/REC-xml-names/> (20 Feb. 2003).

[14] W3C, "W3C Math Home", <http://www.w3.org/Math/> (20 Feb. 2003).

[15] W3C, "Scalable Vector Graphics (SVG)", <http://www.w3.org/Graphics/SVG/> (20 Feb. 2003).

[16] W3C, "PNG (Portable Network Graphics) ", <http://www.w3.org/Graphics/PNG/> (20 Feb. 2003).

[17] W3C, "Synchronized Multimedia ", <http://www.w3.org/AudioVideo/> (20 Feb. 2003).

[18] XML.ORG, "Applying XML and Web Services Standards in Industry", <http://www.xml.org/ > (20 Feb. 2003).

[19] W3C, "Semantic Web", <http://www.w3.org/2001/sw/> (20 Feb. 2003).

[20] W3C, "Web-Ontology (WebOnt) Working Group", <http://www.w3.org/2001/sw/WebOnt/> (20 Feb. 2003).

[21] Library of Congress. Cataloging Policy and Support Office, *Library of Congress classification. H. Social sciences* , Washington, D.C. : Library of Congress, Cataloging Distribution Service, 1994

[22] Bauer, Mary C, *Dewey decimal classification : 200 schedules expanded for use*, Catholic Library Association, 1988

[23] 賴永祥，中國圖書分類法，增訂七版(台北市：商務，民 78 年)。

[24] TVBS，<http://www.tvbs.com.tw/index/> (16 Apr. 2003).

[25] 聯合新聞網，<http://udn.com> (16 Apr.

2003).
[26] ETtoday 東森新聞網，<http://www.ettody.com> (16 Apr. 2003).
[27] 中時電子報，<http://news.chinatimes.com> (16 Apr. 2003).
[28] 謝瀛春、黃學碩、維習安、雷約翰、謝清俊，「新聞內容的標示-XML 之應用」，海峽兩岸資料庫/數據庫與資訊/信息服務交流與合作論文集，民 90.1，頁 205-212。
[29] IPTC, "News Industry Text Format", <http://www.nitf.org/> (20 Feb. 2003).
[30] XMLNews.org, "XML and News Industry", <http://www.xmlnews.org/> (20 Feb. 2003).