以本體論為基礎之新聞事件檢索與瀏覽

許孟淵 [1]，黃純敏 [2]
g9323703@yuntech.edu.tw[1]，huangcm@yuntech.edu.tw[2]
[1] 資訊管理系，雲林科技大學，斗六，台灣
[2] 資訊管理系，雲林科技大學，斗六，台灣

摘要

當前電子新聞的瀏覽，存有以下缺點：(1)新聞文件的瀏覽缺少以事件觀點來加以呈現 (2)電子新聞專輯的內容不包括新聞多文件摘要(Multi-Document Summarization) (3)欠缺社會大眾與網友對於該新聞的評論與看法。當讀者欲全盤掌握新聞內容事實時，須額外找尋數個新聞網站來比較整理，以得到特定新聞事件全貌；此外新聞報導具備前因後果的特性(如白米炸彈客事件到後續處理)，以現今新聞入口網站所提供的瀏覽功能而言，並無法滿足讀者的需求。

本研究主要藉由本體論(Ontology) 理論，提出適性模型來處理電子新聞，提供給讀者更易了解新聞事件發展始末的新聞呈現方式。研究中首先利用事件偵測(Event Detection)與追踪(Event Tracking)之群聚技術，產生新聞事件群集；之後運用自動建構出的新聞本體論及應用到主題地圖(Topic Map)理論的主題地圖索引萃取模型，針對單一事件找出其中蘊含的人、事、時、地、物等主要概念，形成事件中的主要議題及其關聯，並以圖解方式的主題索引地圖來呈現事件中所涵蓋之議題和關聯。本研究另一重點，以建構出的新聞本體論為基礎，找出概念間連結，針對新聞內文中重要字詞加權，擷取出新聞多文件摘要；另外利用新聞本體論結合事件合併演算法，可將相似新聞事件群集做合併處理，便於讀者瀏覽相關新聞事件發展；最後，新聞本體論的重要概念，會被擷取並做為本體論分類概念檢索之用，可讓讀者瀏覽感興趣的新聞人、事、時、地、物概念，節省讀者寶貴閱讀時間，快速找到新聞事件的重點！本研究將上述技術應用到新聞檢索與瀏覽(News Retrieval and View)的目的，是希冀讓讀者在閱讀電子新聞時，能夠了解到新聞事件發展的始末，以及得到更加精確的資訊檢索結果。

本研究之系統評估採公開發佈方式進行系統測試，評估時程為期五天，共回收 72 份問卷。評估結果顯示，本研究確實能增進新聞事件的呈現內容、改善主題地圖呈現之品質，每項系統評估指標都有七成左右受測者能滿意地接受。研究中所提及的新聞事件檢索與瀏覽機制，得到了多數受測者認可。

關鍵字:本體論、主題地圖、事件偵測、事件追踪、新聞多文件摘要、事件合併、改良式新聞檢索

1. 緒論

1.1. 研究背景與動機

現今線上新聞入口網站，除 Google 提供較為完整的新聞事件分群閱讀方式、及 Yahoo！奇摩新聞網提供類似事件閱讀方式的新聞專輯之外，其他新聞網大多只提供基本新聞分類瀏覽。此外，新聞事件有多人同時撰述特性，缺少客觀論點來描述特定事件，使讀者欲客觀了解一新聞事件，須多方閱讀比較，花費在新聞瀏覽搜尋的時間十分可觀。

上述問題，在先前研究(許登傑, 2005; 黃純敏 *et al.*, 2003)以事件分群分類、多文件自動摘要以及主題地圖視覺化新聞呈現等技術，提出對應的解決方案。然而在多文件摘要的可讀性及正確度、以特定事件為基之主題地圖主題及關聯擷取有意義與否，以及資訊檢索的成效上，受限於並非是基於新聞文件語意為主的處理方式，使得呈現效果上還有改善空間。本研究運用新聞本體論(Ontology)及主題地圖(Topic Maps)理論，利用本體論概念間關聯及定義，針對個別新聞事件產生所屬主題索引地圖，提供讀者創新且符合新聞原文語意的新聞關聯知識瀏覽介面。讀者藉由基於新聞本體論之主題地圖幫助，可快速正確地理解事件內包含到的主題及關聯。在文件字詞處理技術及新聞群聚分類的方法論，本研究沿用過去研究成果(許登傑, 2005)，將線上新聞文件依事件相似度分群分類，找出新聞事件群集，並透過新聞本體論之助，剖析出更正確且具內文代表性之多文件摘要內容。隨後透過新聞本體論與主題地圖索引萃取模型，剖析出特定事件之主題地圖，並藉由主題合併機制之輔助，產生事件之完整知識索引介面；利用新聞本體論，將彼此語意相近的新聞事件合併在一起，類似 Yahoo！奇摩新聞網的「新聞專輯」的方式呈現。最後將本體論重要概念擷取出，並顯示成個別新聞類別中重要的新聞人、事、時、地、物概念，可讓讀者針對有興趣的概念加以檢索。有感於讀者在瀏覽新聞事件上的需要，本研究引入社會大眾對於該特定新聞事件的觀感與評論，收集網友對於該事件相關內容電子佈告欄討論及各大新聞網的相關新聞圖片連結，希望讓讀者對於新聞事件，有更加深入且多元化的看法。

1.2. 研究目的與貢獻

本研究目的，乃是利用新聞本體論，達到：

1. 改善新聞多文件摘要的內文代表性及正確性，提升閱讀的流暢度。
2. 結合事件合併演算法，將概念相近新聞事件合併，供讀者在檢索相關事件的方便性。
3. 利用本體論概念語意剖析與連結，有效解決先前研究(許登傑, 2005)中事件主題地圖之主題與關聯不夠相關的缺點，提高讀者閱讀滿意度。

在資訊檢索上，利用新聞本體論包含之人、事、時、地、物，提供更符合使用者語意的資訊檢索、事件專輯瀏覽功能，並納入社會大眾意見，提供特定事件的多元化觀點，強化閱讀深廣度。

1.3. 研究範圍與限制

Yahoo！奇摩新聞網為新聞資料來源，若新聞內文中夾雜英文詞句，將其標記為外來語而不處理。

2. 文獻探討

2.1. 人名辨識

人名是整篇文件的關鍵，已有許多研究(Chang *et al.*, 1994; Chen *et al.*, 1998; Miller *et al.*, 1999; Radev & McKeown, 1998; 李振昌, 1994; 李振昌 *et al.*, 1994; 黃燕萍, 1999; 楊昌樺 & 陳信希, 2004) 投身於人名辨識中。本研究提出一套辨識系統架構，在擷取人名的成效上相當不錯。

2.2. 向量空間模型

以 VSM Model(Salton & McGill, 1983)和相似度計算 Cosine 公式，執行文件分群分類處理。

2.3. 文件分群方法

採用「Single-pass clustering」(Salton & McGill, 1983) ，執行文件分群分類處理。

2.4. 新聞事件偵測與追踪

事件分群處理大致沿用先前研究(許登傑, 2005)事件偵測與追踪系統架構，但因應研究需要做修改。當完成新聞下載或達到設定之單次處理新聞量時，對新聞文件進行字詞處理與文件群聚。

2.4.1 斷句斷詞子系統

包含二個步驟：(1)執行中研院之 CKIP 斷詞和斷句系統 (2)中文詞字過濾器。之後將取回的已逐句逐詞標註詞性之標記文件，剖析器依標記進行判斷，取出斷詞、詞性及其未知詞類別。本研究保留重要名詞和動詞，做為事件主題地圖主題及關聯候選詞。在建置新聞本體論概念時，由於新聞事件特性，會將概念分為人、事、時、地、物，可藉由觀察 CKIP 詞性標註，配合辨識演算法，可取出上述斷詞。本研究採用先前研究(翁頌舜 & 許正欣, 2004; 許登傑, 2005)詞性合併法則，以降低字詞擷取後所產生語意不符問題。

2.4.2 字詞權重計算子系統

本研究使用 TFIDF 公式。考量本研究之斷詞詞性，如主題候選詞包含的專有名詞與一般斷詞，以及詞性類別為人、事、時、地、物的詞類、出現在新聞標題的字詞，會進行特別的字詞加權處理。

2.4.3 事件偵測子系統

將已經由字詞權重處理之新進新聞文件，與既有群集比對相似度後，觀察其是否顯示為新群集文件，其相似度若低於所設定之門檻值則表示此文件表示為新群集，反之則暫時表示為舊群集，再交由時間區間的時間衰退公式，計算出其分數。最後直接將通過相似度與時間區間門檻之文件交由事件追蹤子系統進行處理。新進文件與事件群集間相似度計算採用 Cosine 相似度計算。時間區間之處理主要計算新進文件與候選群集間之新事件信心度，若高於門檻值即視新進文件為新群集，反之則交由事件追蹤子系統，以安排該文件歸屬至適合的候選群集內。時間區間新事件信心度公式如下：

$$score(x) = 1 - \max_{c_i \in window} \{(1 - \frac{k}{m}) \times sim(\vec{x}, \vec{c_i})\} \quad \text{(公式 1)}$$

2.4.4 事件追踪子系統

在找出新進文件之對應候選群集後，此系統會分別計算新進文件與其之間的相關分數，完成計算所有相關分數之後，挑選最大分數之對應候選群集為此新進文件之歸屬群集。本研究採 two-way kNN 法，衡量新進文件所對應目標群集與其他群集間之相關分數。相關分數之計算公式，列示如下：

$$relevance\_score(\vec{x}, kp, kn, D) = \frac{1}{|U_{kp}|} \sum_{y \in U_{kp}} \cos(\vec{x}, \vec{y}) - \frac{1}{|V_{kn}|} \sum_{z \in V_{kn}} \cos(\vec{x}, \vec{z}) \quad \text{(公式 2)}$$

2.5. 語意網

伯納斯李(Berners-Lee & Fischetti, 1999)提出網路資訊架構「語意網(Semantic Web)」，主張將網路上文件有意義的結構化，建立資訊可充分分享與知識重複利用的網路。它能讓本體論除了表達特定領域詞彙定義與詞彙間關係，還包含詞彙和資訊間、以及資源和資源間關係等，能讓電腦交換、搜尋和認同文字意義，提供使用者以語意來搜尋資料而言。

2.6. 本體論(本體知識庫, Ontology)概論

本體論用以定義說明某一特定領域的知識或主題(陳雅絹, 2003)。其內容包含物件(Object)、物件特徵(Property)及物件間關係(Relation)；物件也被稱為概念(Concept)或類別(Class)，用來描述領域中概念；物件特徵則用來描述概念特性；關係則闡述概念間關係。目前語意網上有名的本體知識庫參考資料，如SchemaWeb(Lindesay, 2003)和Swoogle(Finin *et al.*, 2004)，是符合語義網文件格式(如RDF、OWL及DAML+OIL)的資料收集處，收集個人定義本體知識庫，提供使用者或語意網軟體開發者存取。由於本研究新聞資料為中文格式，故採用自訂新聞本體論，做為本新聞事件系統的領域知識。

2.7. 自動建構本體論(Ontology)

在(翁頌舜 & 許正欣, 2004)研究提到，本體論建構方法可分四大類型，分別是以字典為基、以文字分群為基、以關連式法則為基及以知識庫為基之建構法。本研究大致採用(龔俊杰, 2000)關連式法則方式自動建構新聞本體論，但做部分修改，以符合後續階段運作。

2.8. 新聞多文件摘要

新聞多文件摘要可讓節省讀者瀏覽時間、掌握事件重點。在先前研究(戴尚孝, 2003)已取得不錯的多文件摘要研究成果，其步驟包含：(1)斷句與斷詞 (2)群聚語句 (3)形成多文件摘要。但過程並沒有依據文章語意生成摘要，使得摘要內文代表性及語意不足。本研究透過新聞本體論，了解文章描述內容，增加摘要品質。在(吳家威 & 劉昭麟, 2002)研究中，利用 Ontology 幫助摘要系統分析文章語意，其 Precision 和 Recall 的成效較傳統摘要方法為好。

2.9. 主題地圖(Topic Map)

主題地圖可追溯於標準通用標記語言(Standard Generalized Markup Language, SGML)即開始發展，主要用以實現索引與辭典之建構過程，並由國際標準組織制定為 ISO/IEC 13250 標準。此模型包含有三大組合元素，分別為 Topic、Association、Occurrence，這些元素代表了主題、關聯與參照。在先前研究(許登傑, 2005)中，提出植基於主題地圖的新聞事件網頁檢索與瀏覽介面，將新聞事件的主題和關聯，以圖解方式呈現，使得讀者能快速地瀏覽新聞事件中的主題、主題間的關聯關係，以及此新聞主題內容出處(亦即類似書後索引的查找)，可快速呈現新聞事件的重要資訊。主題地圖以圖像與中心發散式加以呈現，將權重越高且越具代表性之關聯式置於主題地圖中央，逐一向外擴展，將事件中最重要的主題突顯出來。主題與關聯具有參照的性質，事件下的每個關聯式皆能透過參照索引功能 — 見(See)，對應出其所屬內文中不同句子位址。主題合併下的關聯式則同樣以參見(See also)功能，對應顯示出此關聯式所屬之不同事件位址。其架構如圖 1 所示：

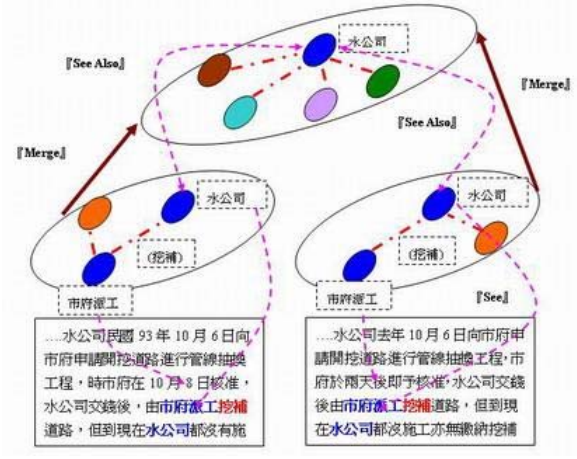

圖 1. 主題地圖中主題與關聯參照示意圖

如圖 1，橢圓形部分是單一新聞事件主題地圖，包含找到的主題，如「水公司」、「市府派工」，主題間關聯「挖補」則串起主題間關係；之後將不同新聞事件間相同主題合併，可查看相同主題下有那些不同的事件內容。經由主題地圖的視覺化呈現技術，吾人可快速掌握住新聞事件重點所在。

2.10. 主題地圖索引合併系統(TMs-merge System)

在「應用 Topic Maps 理論建置知識索引於線上新聞事件檢索研究」(許登傑, 2005)中，透過主題地圖之合併(Merge，亦即「索引的合併」)機制，擴大主題索引的涵蓋範圍，藉以關聯與接續可能相關之事件。而主題地圖索引的合併，是利用 Cosine 相似度公式結合 TDT 分群分類法，如圖 2 所示：

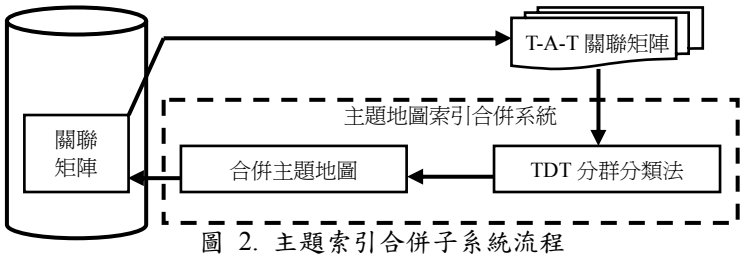

圖 2. 主題索引合併子系統流程

2.11. 中文句結構樹

本研究主題地圖的建構過程中，主要是以名詞篩選出主題候選詞，而以動詞做為主題間關聯。本研究採用先前研究(許登傑, 2005)發展出的中文詞性結構句法則，將重要的主題及關聯候選詞擷取出來，以做為主題地圖中主題和關聯的選取依據。

3. 研究架構

3.1. 系統架構

如圖 3 所示，先從 Yahoo!奇摩新聞網下載新聞文件。接著將原先散落於網路上之各新聞來源的新聞文件，透過相似度比對形成事件群集。新聞本體論自動建構系統產生出新聞本體論後，利用改良式多文件摘要以及主題地圖索引萃取系統，配合新聞事件群集，從事件之新聞文件中產生多文件摘要、萃取其主題地圖知識關聯索引，並將相關主題合併，建構主題地圖間合併關係；這部分是改進先前研究(許登傑, 2005)中無法考量到詞彙語意的缺憾。新聞事件合併處理系統，將語意內容相近事件合併，讓讀者查詢符合某特定主題的新聞事件；新聞事件關鍵字暨重要分類概念檢索詞擷取系統，將產生事件重要關鍵詞、相關新聞圖片連結，加上新聞類別重要人事時地物概念檢索功能，供讀者新聞檢索之用。檢索與瀏覽介面則加入文獻探討中提及的技術，希望提升讀者新聞檢索效能。

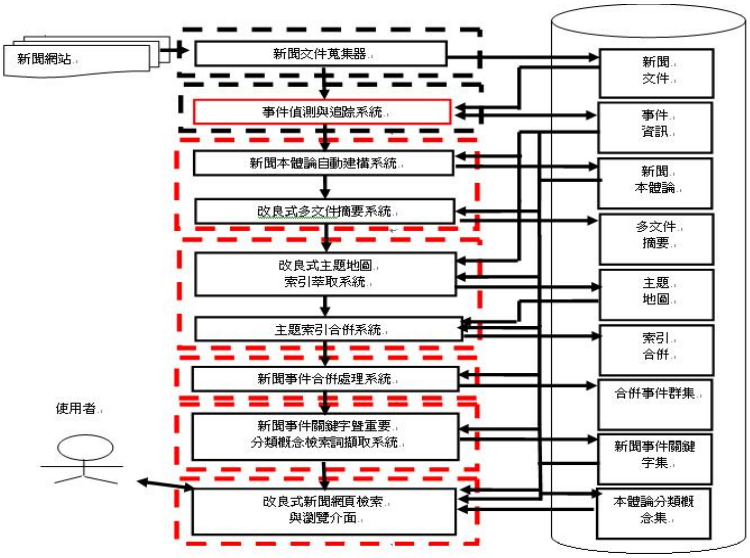

圖 3. 本研究系統架構

3.2. 事件偵測與追踪系統

在完成線上新聞下載、新聞文件的前置處理並儲存至資料庫後，接下來的事件偵測與追蹤系統沿用先前研究(許登傑, 2005)處理的流程，但做部分修改。此系統針對新聞文件進行字詞處理與文件群聚，流程如圖 4，最終產出事件群聚資訊。流程內相關步驟則分述如下。

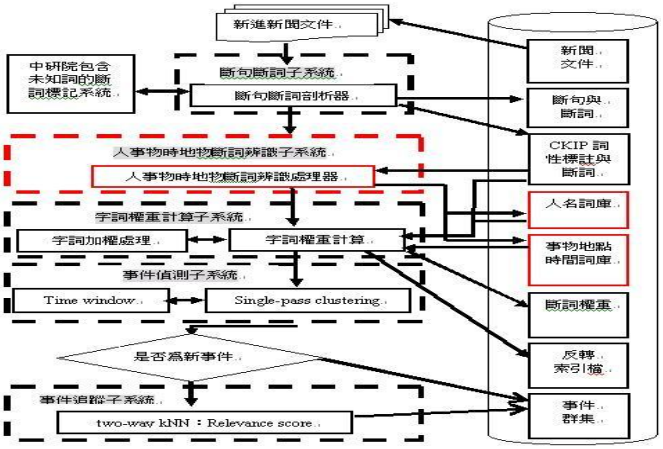

圖 4. 事件偵測與追踪系統流程

3.2.1 人名辨識子系統

執行完文獻探討提及的斷句斷詞子系統，擷取出新聞斷句及關鍵字後，執行此系統以找到新聞中重要人名。首先會將已先經由中研院 CKIP 斷詞處理後所斷出的詞彙做一判斷：和人名無關的詞類(如時間詞、地方詞..等)略過不考慮。接下來，由於人名在 CKIP 的詞性標註中，只會被標註為專有名詞，但並無法單獨判斷為是"人名"的詞性，所以再經由下列人名辨識四流程來加以判斷：

1. 挑出CKIP詞性為"Nb"且未知詞詞性為"CN、EN、MA和xx"的斷詞：根據觀察和實測，人名常出現在詞性標註為專有名詞(Nb)的斷詞裡，其中又以未知詞判斷詞為"CN"(中國人名)、"EN"(歐美譯名)、"MA"(由下而上合併詞)和"xx"(一般詞性)裡最常見。
2. 新聞文件斷詞和當篇新聞未知詞比較：由於CKIP中對於人名並沒有特別的判斷，只會將之歸類為專有名詞(Nb)(CKIP會將之當做未知詞)，故將每一篇新聞文件的前述斷詞和當篇新聞未知詞列表做比對，依據未知詞出現次數和詞性，做特別加權。
3. 運用百家姓詞庫比對：利用以下幾條規則，給予特別加權。當詞性:
   (1) 為"Nb"且為"CN"的斷詞，可能是"中國人名"；再利用百家姓詞庫比對，將含有姓氏的斷詞且長度(介於二至四字詞)給予特別加權，可斷出如"林曉培"。
   (2) 為"Nb"且為"EN"的斷詞，可能是"歐美譯名"；之後將斷詞長度(大於四字詞)的斷詞給予特別加權。可斷出如"凱薩琳麗塔瓊斯"等人名。
   (3) 為"Nb"且為"MA"的斷詞，可能是"日本人名"； 再利用百家姓詞庫比對，將含有姓氏的斷詞且長度(介於三至五字詞)給予特別加權，可斷出如"角川歷彥"。
   (4) 為"Nb"且為"xx"的斷詞，可能是"藝人人名"； 再利用百家姓詞庫比對(，將含有姓氏的斷詞且長度(介於二至四字詞)給予特別加權，可斷出如"侯孝賢"。
4. 挑出大於姓名門檻值的斷詞：整個流程如圖5所示。

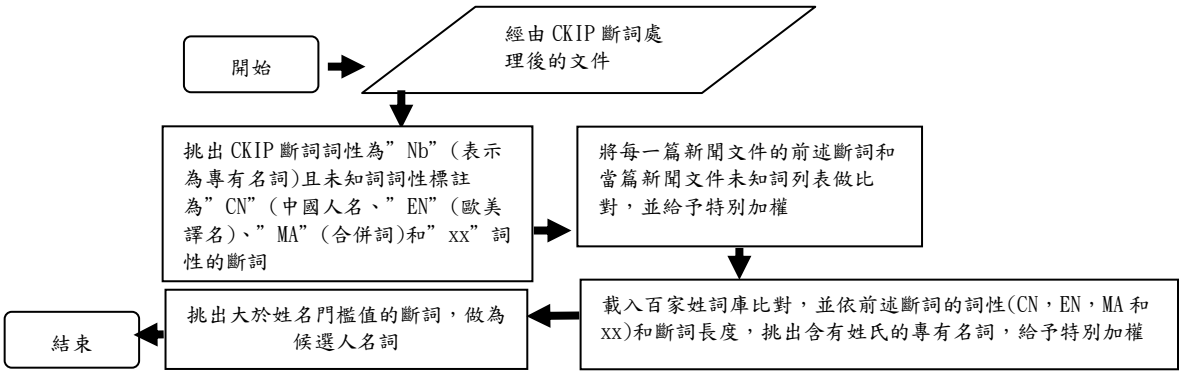

圖 5. 新聞文件人名擷取程序

系統在處理上有所限制，像原住民人名命名規則複雜、沒有一定規則，故不加以處理；外國人英文譯名，在各大新聞網站的新聞報導並沒有統一，也不在處理範圍。在此系統運作完成後，接下來依照文獻探討中的「字詞權重計算」、「事件偵測」、「事件追蹤」等步驟執行，最終產出事件群聚。

3.3. 新聞本體論自動建構系統

3.3.1 物件導向本體論(Object Oriented Ontology)架構

圖 6 為(Lee *et al.*, 2002)所定義的物件導向本體論架構，包含了領域、分類、概念和關係。領域是本體論所要描述的特定標的；分類是領域下的分類項目，會繼承領域的一些特性；概念類似物件，由物件名稱、屬性和操作所構成；關係則類似物件之間具有的三種關係：關聯、概化及聚合關係等。

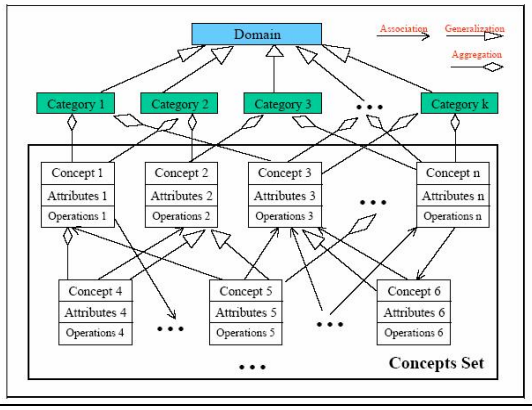

圖 6. 物件導向本體論架構

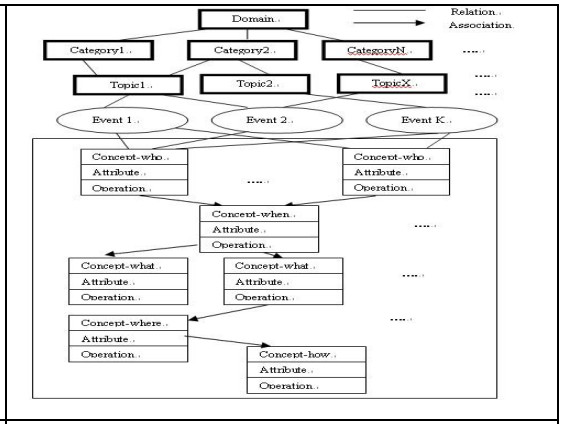

圖 7. 延伸新聞領域物件導向本體論架構

然而此種架構若要對映到新聞本體論的建置，會有所不足。由於新聞領域可用事件觀點來看待相關新聞報導，但圖6無法反應出事件概念；故本研究延伸先前研究(陳雅絹, 2003)提出的Domain Ontology結構，發展出新聞領域物件導向本體論架構。除了原先領域、類別及原先概念層級外，加上主題與事件概念。將經由斷句斷詞、字詞權重與關聯法則所選出之重要名詞建構成新聞本體論中的概念與關係，並把新聞文件人、事、時、地、物等元素納入考量，以反應出新聞文章中重要的主題。圖7為延伸新聞領域物件導向本體論架構。

如圖7所示，將新聞本體論中的概念區分成人、事、時、地、物等分類概念，不但可幫助新聞事件主題地圖主題的瀏覽檢視，更可實作出符合語意的新聞事件分類概念的檢索(如要查找人的話，可以找出和"張錫銘"此分類概念相關的各項主題資訊)。圖8是新聞本體論自動建構的過程。

如圖8所示，此新聞本體論建構暨主題索引萃取系統流程，除沿用先前研究(許登傑, 2005)主題索引萃取流程外，另外加上新聞本體論建構部分，如圖紅色區塊所示。此等改良不但可以產生出事件主題地圖中主題和關聯之外，可產生出新聞本體論，以供用於後續多文件摘要、主題地圖索引合併系統、語意相近新聞事件合併處理擷取所需資料之用。下一節解釋新聞本體論建構理念步驟。

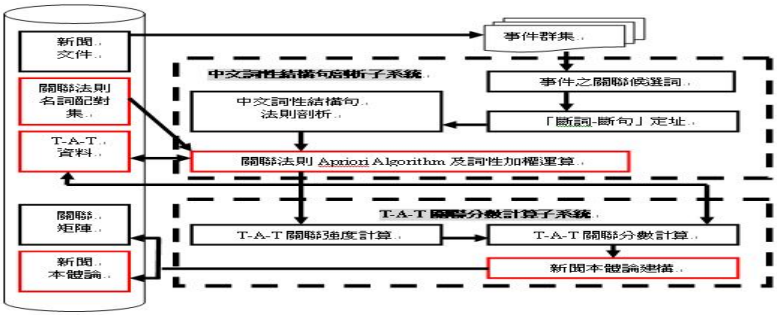

圖 8. 新聞本體論暨主題地圖索引萃取系統流程

3.3.2 新聞本體論建構理念與步驟

理念是使用自然語言處理，亦即中研院的詞性標註和資料探勘的關聯式法則，過濾出重要的名詞與動詞，當作是重要的事件主題地圖主題與關聯候選詞。新聞本體論建構步驟：

1.步驟 1：以事件群集為單位，利用 CKIP 先挑選出有意義的名詞和動詞。所擷取的重要名詞，包含了 CKIP 詞性辨識為普通名詞(Na)、專有名詞(Nb)、人名詞、地方詞(Nc)、位置詞(Ncd)及時間詞(Nd)。動詞則包含 VA、VAC、VB、VC、VCL 及 VD 等；將找出的重要名詞與動詞，依權重高低降冪排序，先挑出動詞以確認概念間關係。

2.步驟 2：使用資料探勘 Apriori Algorithm，挑出兩兩間具有強烈相關名詞，形成關聯法則。再以先前挑選出的動詞為中心，往外找二個名詞串成候選 T-A-T(應用中文句結構樹和詞性合併法則，可參閱文獻探討與本研究 3.5.1 章節的作法)，並配合找出的關聯名詞集合及名詞、動詞對應到的詞性，做適當 T-A-T 加權，再設立門檻值，去除掉主題間較無關聯的 T-A-T 集合，使其每個新聞事件的 T-A-T 集合更具代表性(可參閱本研究 3.5.2 章節作法)。

3.步驟 3：將上述的 T-A-T 做計算與篩選，(應用 T-A-T 關聯強度與分數公式，挑出合適的 T-A-T，可參閱本研究 3.5.3 章節的作法)，挑出要形成新聞事件本體論的 T-A-T 集合。

4.步驟 4：利用新聞類別新聞事件(不小於 2 篇以上新聞的新聞事件群集)的代表性 T-A-T 集合，形成初步新聞本體論(過程可參閱本研究 3.5.4 章節)，並調整本體論的概念。其調整是將主題間較沒有關聯的 T-A-T 集合，左邊 Topic 視為父概念，右邊 Topic 視為子概念，再將 Topic 間關聯，併入到父概念操作(operation)，視為父概念方法之一；把子概念併入到父概念屬性。經上述步驟可形成單一類別中單一特定事件本體論。然而建成的本體論，需經由新聞領域專家檢閱和修改，並調整產生本體論的演算法(如關聯式法則的門檻值設定、T-A-T 關聯強度與分數公式)，反覆測試本體論正確性，直到符合新聞領域專家的期望。

5.步驟 5：產生每一新聞類別的新聞事件本體論資料後，利用事件合併演算法(two-way KNN)，將語意相近新聞事件做事件合併處理，找出新聞事件本體論中主題和相關事件的關聯，以建成完整的新聞領域物件導向本體論。詳細事件合併處理過程，會於本研究 3.6 章節提到。

3.4. 改良式多文件摘要系統

若事件群集內包含兩篇以上新聞文件，即針對該事件產生一篇多文件摘要，協助使用者在極短的時間內判讀事件內容，並快速尋找其所感興趣之新聞事件。本研究之多文件摘要技術大體沿用先前(戴尚孝, 2003)研究，利用 Single-pass clustering 技術群聚語句，但加上了時間性考量、專有名詞加權、新穎性偵測、本體論加權及句子詞彙的關聯字詞加權等，如圖 9：

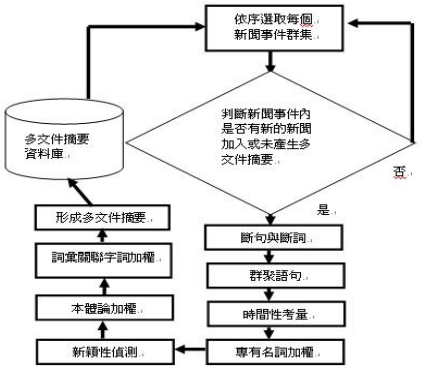

圖 9. 改良式多文件摘要流程

其步驟說明如下：

1. 斷句與斷詞：利用中研院 CKIP 斷詞、本研究自行開發的斷句處理及 CKIP 未知詞偵測，分析出新聞文件包含的關鍵詞彙及新聞語句。
2. 群聚語句：將各文件中描述同一事實的句子群聚起來，而群聚完成後輸出摘要時，是將各句子群集中分別取一句出來即可(不過其句子需經過第 2 步驟後的其它考量及運算)。

句子群聚的方法採用 Single-pass clustering，句子分類的方式則採用 two-way KNN。

3. 時間性考量:加上時間考量後的某事件某句子群集之中的句子，會加重在前面時間及後面時間的句子;而中間句子則依其天數,逐漸加強其時間所佔權重,會影響到句子的 TFIDF。
4. 專有名詞加權：依句子中的專有名詞出現次數，給予特別加權。當句子出現的專有名詞越多，則句子整體的 TFIDF 值也將越高，將提升該句子被選入摘要的機會。本研究使用一個句子中 CKIP 詞性辨識為 Nb 的斷詞做為計算依據。
5. 新穎性偵測：當要形成某事件的多文件摘要時，是將各句子群集中分別取一句出來即可，以避免輸出描述重複事實之語句。其使用的作法是：判斷句子中 name entities 的數量是否通過某個門檻值，也就是判定詞性標註(POS)為 Nb(專有名詞)、Na(普通名詞)、Vc(動詞)的數量是否通過門檻。倘若通過時，則取出此句當成摘要的一部分。
6. 本體論加權：先從單一新聞事件本體論中取出重要的本體論概念(包含人、事、時、地、物等分類概念)後，再判斷語句群集中的句子，其詞彙有出現在新聞事件本體論的任一重要概念上時，則特別予以加權，以突顯出句子裡重要的主題。
7. 句子詞彙的關聯字詞加權：透過事件本體論概念之間相互連結的資訊，以句子中的某特定字詞角度來看，若句子裡的其它詞彙，都和某特定字詞相關，代表它們彼此之間可能在探討某一個特定的事件或訊息(如「張錫銘」此特定字詞來看，「槍擊案」和「制式手槍」可能和它相關)，則依其相關字詞的個數，給予特別加權。
8. 形成多文件摘要：在完成句子群聚後，針對語句的輸出順序是依據該句子在原始文件的相對位址來決定，如公式(3)所示：

   P=句子在原始文件的位置／原始文件總句數　　(公式 3)

   計算所有輸出句子的 P 值後，P 值小的句子會先輸出，而若 P 值相同，則依文件編號順序，最後形成一篇多文件摘要。

3.5. 改良式主題地圖索引萃取系統

大體沿用先前研究(許登傑, 2005)做法，以事件群集為單位，產出代表該事件群集主題地圖索引。此系統運用中文詞性結構句法則與關聯強度計算公式，並依此建立完整索引萃取評估模型，以產出能顯示為主題地圖之關聯式資訊結構矩陣，此矩陣結構由主題 x、關聯 y、主題 z 相關資訊所組成，研究中以 T-A-T 代表關聯式。本研究改良先前研究不足，將新聞本體論建構與 T-A-T 關聯式萃取結合，並利用 Apriori Algorithm，將語意上較不相關 T-A-T 去除，可找到彼此間更具關聯主題。之後利用 T-A-T 關聯式與新聞本體論,即可建置以事件為單位的主題索引地圖建置。流程可參閱圖 8。

3.5.1 中文詞性結構句剖析子系統

利用完成詞性標注之句子與具備詞性之主題、關聯候選詞，透過中文詞性結構句法則的分析，判斷出哪些主題候選詞能經由關聯候選詞而結合成候選 T-A-T。先將該事件群集之所有元素候選詞依權重降冪排序後逐一定址，找出其位於原文件內文之所在短句。以關聯元素為中心，定址之短句上下各另取一句短句，使元素候選詞之定址對象成為三句短句的句組。定址完成後以關聯候選詞為中心，進行中文詞性結構句法則的剖析，找出所有以此候選詞為關聯之 T-A-T，再交由 T-A-T 關聯分數計算子系統加以評估。表 1 將舉例說明以接近優先原則之中文詞性結構句法則剖析後的產出。

表 1.中文詞性結構句法則剖析範例

<table>
<tr><td colspan="3">關聯候選詞「高達(VJ)」所定址之句組</td></tr>
<tr><td colspan="3">目前(Nd) 受難記(Na) 的(DE) 北美(Nc) 票房(Na) 總收入(Na)，</td></tr>
<tr><td colspan="3">高達(VJ) 3億5千480萬美元(DM)，</td></tr>
<tr><td colspan="3">在(P) 北美(Nc) 票房(Na) 排行榜(Na) 上(Ncd) 名列(VG) 第8位(DM)。</td></tr>
<tr><td colspan="3">中文詞性結構句剖析產出(T-A-T 候選關聯式)</td></tr>
<tr><td colspan="3">後面短句含有動詞(V)，判斷爲完整句子結構，不進行主題候選詞組合</td></tr>
<tr><td>總收入</td><td rowspan="4">高達(VJ)</td><td>3億5千480萬美元</td></tr>
<tr><td>票房總收入</td><td>北美</td></tr>
<tr><td>北美票房總收入</td><td>北美票房</td></tr>
<tr><td>受難記北美票房總收入</td><td>北美票房排行榜</td></tr>
</table>

資料來源：(許登傑, 2005) 應用主題地圖理論建置知識索引研究

3.5.2 關聯法則及詞性加權運算子系統

先前研究(許登傑, 2005)的 T-A-T 關聯式，主題間並沒有具備較高的關聯性；故本研究加入關聯法則和詞性加權的判斷機制，進行 T-A-T 關聯式的主題關聯程度計算，步驟如下：

1. 關聯法則的名詞配對判斷：利用資料探勘法則，針對每個新聞事件中已辨識出的人、事、時、地、物等斷詞，執行關聯法則運算，以找出具有高度關聯「人與人」、「人與事」、「人與時」、「人與地」、「人與物」…等依此類推的關聯法詞名詞配對集合。再針對先前的事件 T-A-T 關聯式，依關聯式中主題與關聯法則名詞配對間符合程度，給予不同程度加權。
2. T-A-T 關聯式的主題與關聯詞性加權判斷：經由本研究實驗觀察，發現由名詞詞性為「Na」、「Nb」、「Nc」及「Nd」所構成的主題，以及動詞為「VA」、「VB」、「VC」、「VD」所形成的主題關聯式，會讓讀者感覺到 T-A-T 關聯式的組成較能表達出新聞某一重要的片段資訊；符合以上詞性描述的 T-A-T 關聯式，依其符合程度不同，給予不同程度加權。
3. T-A-T 關聯式的 Support 值及 Confidence 值加權運算：倘若 T-A-T 關聯式中的任一主題符合關聯法則名詞配對集合，則針對原先關聯法則中名詞的 Support 值及 Confidence 值做特別的加權；以資料深勘法則來說，當關聯法則中名詞配對間的支持度(Support)與信賴度(Confidence)越高時，則此關聯式對於讀者而言，應該能得到更具有相關意義的主題閱讀機會；故本研究針對 T-A-T 關聯式中的主題，其原先在關聯法則中所對應到的支持度與信賴度，依其大小給予不同程度的加權，找出更符合事件主題 T-A-T 關聯式。

3.5.3 T-A-T 關聯分數計算子系統

沿用先前研究(許登傑, 2005)提出的 T-A-T 關聯強度公式與 T-A-T 關聯分數公式，以關聯分數評估 T-A-T 之代表性。先以關聯強度公式計算於同一事件群集中擷取之數條 T-A-T，若關聯式存在於標題中，則應適當加權其強度值。關聯強度公式列式如下，藉由類似詞頻的衡量方法，若能求出相對高於其他關聯式的數值，則表示此關聯式於該事件群集中出現之頻率比率越高，即關聯強度越強。

$$association\_strength(T_x, A_y, T_z) = Freq\frac{\{T_x, A_y, T_z\}}{T_x} \times Freq\{T_x, A_y, T_z\} \qquad \text{(公式 4)}$$

公式(4)中，Tx(Tz)為主題候選詞，Ay 為關聯候選詞 y，示為 Tx-Ay-Tz 關聯式，Freq()則表示計算該元素或關聯式之出現頻率。T-A-T 關聯分數之計算，係將該關聯式之關聯強度值乘上其主題與關聯候選詞之權重。公式中有將關聯候選詞納入權重加乘部分，主要考量動詞詞性仍有可能屬於文件之重要關鍵詞，例如關聯詞「謀殺」、「當選」。透過 T-A-T 關聯強度分數的計算，單一事件群集將產生代表其事件之 T-A-T 關聯矩陣，利用矩陣的關聯式資訊可直接呈現出此事件群集之主題地圖分布。T-A-T 關聯分數公式如下。公式(5)中，Tx(Tz)為主題候選詞 x(z)，Ay 為關聯候選詞 y，WTx(WAy、WTz)則為 Tx(Tz、Ay)之權重。

$$association\_score(T_x, A_y, T_z) = association\_strength(T_x, A_y, T_z) \times W_{Tx} \times W_{Ay} \times W_{Tz}$$

(公式 5)

3.5.4 T-A-T 新聞本體論建構

利用此 T-A-T 群集建構成事件本體論，步驟如下：

1. 新聞事件本體論的資料建構來源：一次產生某一特定類別(如政治)的事件本體論。將每個事件裡的代表性 T-A-T 群集取出後，透過本體論門檻值的設定，將低於門檻值的 T-A-T 從本體論概念的候選建構 T-A-T 群集中去除；這些去除掉的 T-A-T 群集將被合併為本體論概念中的屬性(attribute)或操作(operation)。
2. 本體論概念的特徵確認：事件本體論架構包含的人、事、時、地、物等概念特徵，可在之後的本體論加權處理中對於重要的人事時地物做特別加權之用、找出事件中重要的人事時地物等重要關鍵詞等，在此步驟會事先找到每個特定事件所辨識出的人事時地物詞彙，再和被選為本體論概念的 T-A-T 做比對，依此可找出本體論概念的特徵歸屬為何。

3. 計算新聞事件中的本體論概念出現次數與語意連結：一個新聞事件的 T-A-T 群集建成時，有可能會重複相同的主題(如張錫銘－涉入－搶案，搶案－主導－張錫銘)，在此會計算某特定本體論概念在事件中出現的次數、和它相關的其它本體論概念資訊，再透過 T-A-T 中的關聯，可找到兩兩本體論概念間的語意連結；之後依照本體論概念出現的次數，以及是否被歸類為人事時地物等其中之一的概念特徵，設立一個本體論建構門檻值，再將符合門檻值的本體論概念建構成一個初步的事件本體論。
4. 事件本體論的合併處理：由於一個新聞事件大多會針對特定的幾個主題，會有詳細且反覆的描述，倘若能將事件本體論的重要概念抽出，並透過事件間重要概念的比對及合併處理，將可找到彼此間相互有關聯的新聞事件報導，提供給讀者查閱相關新聞事件報導的方便性。本研究提出的事件合併處理機制，結合了已建構好的事件本體論和事件合併演算法(two-way KNN)，可解決先前研究(許登傑, 2005)所提到的事件群聚斷層的問題、事件偵測的時間區間設計問題(可參閱文獻探討)，將彼此間原本獨立的事件群集串連在一起，供讀者瀏覽與檢索之用。新聞事件的合併處理機制，會於論文 3.5 章節詳細說明。
5. 新聞本體論建構完畢：經前四步驟處理，可建構如圖 7 的本體論架構。

3.5.5 主題地圖索引合併系統(TMs-merge System)

在經由改良式主題地圖索引萃取系統的運作後，此時已將個別新聞事件重要的主題及關聯擷取出來，讀者可任意瀏覽它們的內容，快速了解新聞事件的重點所在；不過在不同的新聞事件中，也許會有相同的主題在討論著，在此沿用先前研究(許登傑, 2005)的主題地圖索引合併架構，將原先分散於不同事件的相同主題合併，便於讀者閱讀不同的新聞事件內容。

3.6. 新聞事件合併處理系統

合併處理對象是已經建好的初步新聞事件本體論(參閱 3.3.2 章節)，需要將一個個的新聞事件本體論做分群及分類處理，以呈現出主題和相關事件群集間的關係；所用到的分群技術，採用「Single-pass clustering」，而分類技術則採用修改過後的 two-way KNN 分類法，以衡量候選的新聞事件本體論該歸屬到的新聞事件主題目標群集。詳細過程如下所示：

1. 新聞事件本體論間的相似度矩陣計算：在執行新聞事件本體論的分群處理前，需先計算出新聞事件本體論間的相似度矩陣，以便做分群比較相似度之用。本研究採用 Consine 相似度計算公式，但做了部分修改，如公式(6)所示：

$$Sim(oi,oj) = \Sigma^{M}_{g=1} W_{goi} * W_{goj} / \sqrt{(\Sigma^{M}_{j=1} W_{goi}^{2})} * (\Sigma^{M}_{j=1} W_{goj}^{2}) \qquad \text{(公式 6)}$$

公式(6)中，sim(oi,oj)代表新聞事件本體論 oi 比對新聞事件本體論 oj 的相似度，W_{goi} 為本體論概念 g 在本體論 oi 中的權重(以本體論概念在事件中出現的次數，取代原始 Consine 公式的字詞權重)，W_{goj} 則為本體論概念 g 在本體論 oj 中的權重，M 則代表新聞本體論中本體論概念總數。利用此公式，可計算出新聞事件本體論間的相似度矩陣。

2. 新聞事件本體論分群處理：類似 two-way KNN 演算法執行過程，只是對象換成是某特定新聞類別的事件本體論集合，故不再贅述其流程。
3. 合併後新聞事件本體論的主題命名：從中觀察合併後的事件本體論所描述的內容，實際觀察事件本體論悶重要且重複的概念，以人工方式為合併事件本體論命名。

3.7. 新聞事件合併處理系統

利用已事先建好的新聞事件本體論，從中取得新聞事件的重要概念，再經過一連串運作，形成新聞事件關鍵字與分類概念檢索資料集，流程如下所述：

1. 取得個別新聞事件本體論的關鍵概念：新聞本體論於建構之初，會計算某一概念在整個新聞事件中出現的次數，故本研究將概念出現次數當作擷取關鍵概念的重要考量因素；將概念出現次數設定門檻值，並取出限制個數的新聞本體論關鍵概念。
2. 產生關鍵概念的 BBS 超連結及相關照片資訊連結：透過 Google、Yahoo！奇摩新聞網、網擎資訊等資料來源，利用 Crawler 抓取網路上關於新聞關鍵字相關討論及照片網址。

3. 依關鍵概念屬性形成本體論分類概念群集：將步驟 1 取得的所有新聞事件本體論重要概念，依其特徵(人、事、時、地及物)，做個概念上的分類群集；設立分類群集門檻值，以擷取出分類概念群集(人、事、時、地及物)中可被擷取為本體論重要分類概念的資料。

3.8. 改良式新聞網頁檢索與瀏覽介面

如圖 10 所示。除了新聞事件主題地圖，更涵蓋社會大眾對於某一特定新聞事件的相關討論、新聞圖片連結、相關新聞、新聞多文件摘要等；本系統另外提供一般新聞關鍵字檢索，以及將新聞事件的重要人、事、時、地、物詞彙擷取出、供讀者檢索之用的新聞本體論分類概念檢索功能等，希望能提供讀者一個內容豐富且具有創新性的新聞事件瀏覽平台。

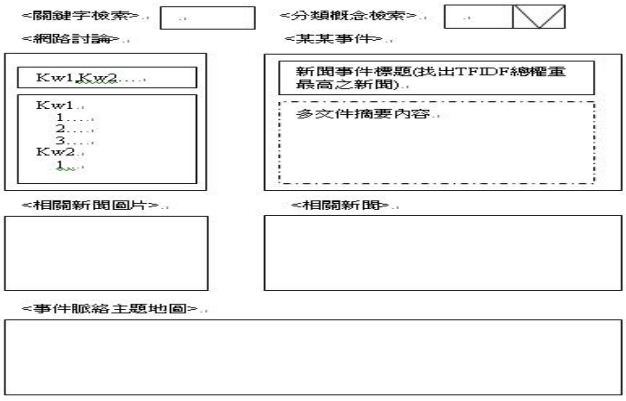

圖 10. 本研究新聞網頁檢索與瀏覽介面架構

4. 系統實作與評估

4.1. 系統開發

本研究之系統介面與功能分為後端與前端兩大部分，第一部分是後端資料處理系統，包含了文件字詞剖析、權重計算、分群分類、主題地圖萃取與合併、擷取新聞候選人名詞、擷取新聞文件事時地物詞、事件合併處理、新聞事件重要關鍵字擷取與自動建置 Ontology 功能等。第二部分為前端網頁檢索與瀏覽介面，包含了分類新聞事件內容瀏覽、關鍵字和分類概念檢索、網友對於該新聞事件的評論、相關新聞圖片、新聞多文件摘要、事件脈絡主題地圖及事件相關新聞文件，如圖 11 所示：

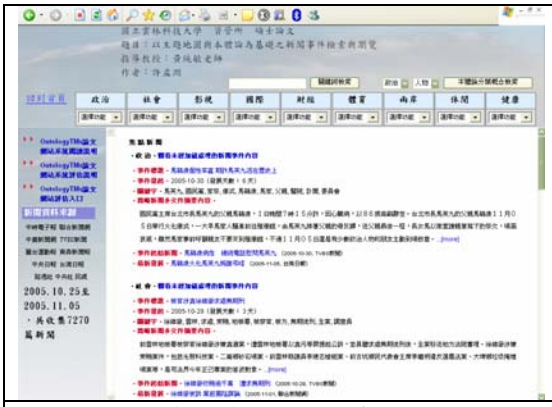

圖 11. 前端系統第一層新聞事件畫面展示圖

圖 12. 前端系統第二層新聞事件畫面展示圖—以政治類別的新聞事件分類瀏覽為例

使用者於畫面上方區塊，可點選不同新聞類別，其中包含三種新聞事件內容呈現功能：「新聞事件分類瀏覽」、「新聞事件主題合併地圖」及「新聞事件合併瀏覽區」;右下方區塊則是資料內容的常駐瀏覽區。網站的三大功能，分別是事件分類新聞瀏覽、主題地圖新聞瀏覽及關鍵詞檢索區(分為關鍵字檢索與分類概念檢索)。如圖 12，讀者可觀看到新聞事件標題、事件發展天數、新聞事件關鍵字、簡略新聞多文件摘要，以及事件起始新聞和最新發展等，讀者可點選其中超連結而進入第三層新聞事件瀏覽畫面。在前端系統第三層顯示畫面左邊地方，除了顯示出該新聞事件重要的詞彙、相關新聞圖片，最特別的地方是本研究加入了網友討論的意見，可見到新聞事件中各方的評論和探討，讓整個事件的呈現更加多元化。在畫面右邊部分，則顯示出新聞事件代表的事件標題、多文件摘要、相關新聞和事件脈絡主題地圖。系統畫面如圖 13，圖 14 則為主題地圖視覺化呈現工具。

圖 13. 前端系統第三層事件內容瀏覽畫面展示圖—以新聞類別「政治」、新聞事件「馬鶴凌個性率直 期許馬英九活在歷史上」為例

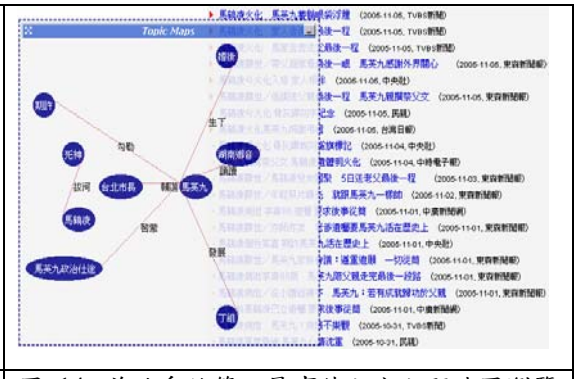

圖 14. 前端系統第三層事件內容主題地圖瀏覽畫面展示圖—以新聞類別「政治」、新聞事件「馬鶴凌個性率直 期許馬英九活在歷史上」為例

4.2. 系統評估

為驗證本研究發展之系統模型確實能改進先前研究(許登傑, 2005)實作出的研究結果，採取讓受測者透過前端瀏覽介面，以線上問卷方式進行評估。評估目標是針對本研究系統運用新聞本體論輔助所產生之新聞事件關鍵字、新聞多文件摘要、主題地圖結果、新聞事件人事時地物字詞辨識程度、新聞事件合併處理、本體論分類概念檢索等項目，期望能取得較先前研究更加進步的成果。

4.2.1 評估資料回收與分析

本研究系統評估採取網站公開發佈的方式進行測試，評估時期共計五天，回收了 72 份問卷。以下則針對回收的問卷內容進行統計分析，描述評估項目的評估結果。

1. 受測者背景資料分析：男女比例大約是一比一，學歷多集中於 21 至 30 歲年齡層學生，佔受測者 95%；大學生與碩士生比例為 13:7；受測者普遍有電子新聞閱讀習慣。
2. 加值處理新聞內容效益分析：本部份包含四個問項，個別是：「您認為新聞事件內容經加值處理後，是否比原事件內容的呈現，更易於理解事件中所出現的關鍵字、事件多文件摘要內容、主題元素所提示的字詞、關聯元素所表達的主題元素間關聯」之中，非常相關的比例與相關的比例相加，約有七成以上受測者感到滿意！
3. 新聞所含人事時地物詞彙擷取效益評估：本部份包含四個問項，個別是：「您認為新聞事件內容的呈現：(1)人名 (2)事物、地點、時間詞 (3)本體論分類概念檢索找出的人名 (4) 本體論分類概念檢索找出的事物、地點、時間詞，其辨識程度」之中，非常相關與相關的比例相加，約有七成以上的受測者感到滿意；而在事物、地點及時間詞的辨識，則約六成受測者感到滿意。
4. 本體論分類概念檢索效益評估：本部份問項是：「您認為新聞事件合併功能所查詢出來的合併事件標題，與相關的新聞事件編號所顯示的事件內容，彼此間的相關程度為何(亦即標題與事件間的相關程度)」，評估結果顯示，經由事件本體論所萃取出的重要概念(如人物、事物、地點及時間)，與相關的新聞事件所對應的事件內容，有七成左右的受測者認為是滿意的。
5. 新聞事件合併效益評估：本部份問項是：「您認為新聞事件合併功能所查詢出來的合併事件標題，與相關的新聞事件編號，所顯示的事件內容、以及新聞事件間，彼此間的相關程度為何」從評估結果可觀察到，應用事件本體論所產生的新聞事件合併處理，在標題與事件間的相關程度，以及被合併的事件間彼此相關程度，約七成左右受測者給予滿意的高度評價。
6. 新聞事件查詢介面觀感評估：大約八成左右受測者對新聞事件檢索與瀏覽機制，給予好評！

4.2.2 與國內各大新聞網站的比較

本研究的新聞事件系統，採用不少以往新聞網站所沒有的技術與概念，讓讀者在閱讀新聞事件上具有更大的便利性及更完整的新聞內容。茲整理如表 2 所示：

表 2. 本新聞事件系統與國內各大新聞網站採用技術之比較

採用技術 / 新聞網站	視覺化呈現(主題地圖)	Ontology 輔助新聞內容	以事件方式瀏覽新聞	新聞多文件摘要	新聞專輯	網友的事件相關討論	個人化自訂新聞瀏覽	RSS 訂閱	新聞內容關鍵字連結	新聞檢索
本新聞系統	●	●	●	●	●	●			●	●
Yam 蕃薯藤					●	●		●		●
Yahoo！奇摩					●	●	●	●	●	●
自由電子報					●	●				●
中時電子報					●	●	●	●		●
udn.com					●	●		●		●
Google 新聞						●	●	●		●

5. 結論與未來研究方向

5.1. 研究成果

本研究延續先前研究成果(黃純敏 *et al.*, 2004; 黃純敏 *et al.*, 2003)，針對主題地圖中主題關聯式間相關程度不高、多文件摘要的正確性及可讀性、新聞關鍵字與事件的相關程度、資訊檢索較不符合語意等，提出有效改善方法。在本體論的幫助下，解決了導致上述問題產生的原因：「受限於並非是基於新聞文件語意為主的處理方式」。鑑於人名、事物、地點及時間詞，在新聞事件(文件)扮演重要角色，本研究開發一套人事物時地物斷詞辨識處理系統，能有效將新聞文件中潛藏人名、事物、地點及時間擷取出來，做為重要新聞主題；本研究亦改進先前研究(陳雅絹, 2003)中新聞 Domain Ontology 結構，發展出改良新聞本體論架構；藉本體論幫助可有效改善先前研究(許登傑, 2005)成果，另外開發出事件合併處理機制與重要關鍵詞擷取系統，增加事件內容的呈現豐富性。

本研究在資訊檢索部分，鑑於以往讓使用者直接鍵入關鍵字搜尋相關新聞事件的方式，似乎無法讓讀者了解到事件發展全貌，因此加入網友對於該事件的討論、相關新聞圖片以及本體論分類檢索的概念，讀者可藉此看到新聞事件不同討論觀點，有更多元化看法；特別是在本體論分類檢索部分，作者將新聞事件中重要的人、事、時、地、物等概念擷取成新聞事件關鍵詞，讀者可依據感興趣的內容做檢索，如對"張錫銘"這個人感到興趣，點選它之後即可看到和它相關的所有事件內容，以及主題地圖的呈現。此一機制能提升讀者在閱讀新聞事件上的便利，快速掌握事件發展脈絡！

本研究觀察到，當新聞事件是圍繞在某一特定人物的相關報導時，如名模林志玲、政治人物馬英九，在事件群聚效果、新聞多文件摘要、Topic Maps 主題和關聯間語意、事件合併的成效等，會呈現較佳的結果。作者推想，由於報導某一特定人物的相關新聞很單純，會圍繞著人物報導。針對此類型新聞事件，本研究的事件相關處理機制都有很良好的效果！反觀如財經、生活和健康等新聞類別的事件內容呈現，則和上述結論相反，可能會有較差的事件處理結果；作者推想，應該是這些類別的「新聞事件」本身持續報導的機會不大，造成事件呈現的結果不佳。

5.2. 未來研究方向

本研究尚存在許多可再精益求精之處；以人名辨識而言，如原住民名字"瓦歷斯．貝林"，受限於原住民名字取法沒有固定規則，成為辨識上限制；再者，由於本研究架構龐大，在本體論開發方法部分，是沿用較舊的關聯法則法，而非熱門的 Formal Concept Analysis(FCA)，後續研究可考慮採用，也許能得到更佳的建構結果。本研究另一限制，在於沒有實作語意推理機制；為增添本體論實用程度，強烈建議未來研究可朝此方向發展。現今本體論發展遇到不少問題。如新聞領域本體論的建構方式與理念，並沒有可依循規範；若能發展出一套可遵循的新聞本體論架構，應能提升新聞本體論間互相交流與應用程度。此外本研究系統，需經長時間後置處理，使得新聞本體論無法做到即時更新。以上問題描述，除了會產生新聞事件瀏覽無法做到一般新聞網站的動態更新外，會引發一個有趣議題：新聞事件發展有其時間性，某些重要概念會隨著事件發展，逐漸降低其重要性；例如以影視圈話題女王許純美舉例，男友從原先的林宗一到後續的邱品叡，為了反應出人名「許純美」此事件本體論的「現今」重要概念，應該將人名「邱品叡」特別加強其權重，而人名「林宗一」重要性應隨著降低，應能得到更符合「當時情況」的新聞本體論。另外為因應讀者閱讀新聞需求，應加入英文新聞文件剖析與處理機制。有些受測者反應本研究的新聞事件資訊量雖然豐富，但對於某些讀者，也許是另類的「資訊過載」，造成閱讀上負擔；可以考慮專注在新聞呈現的質與量如何達成某個平衡點。作者認為，後續研究可從上述方向進行改善。

6. 參考文獻

1. Berners-Lee, T., & Fischetti, M. (1999). *Weaving the web：the original design and ultimate destiny of the world wide web by its inventor* (1st edition ed.). San Francisco: Harper Business.
2. Chang, J. S., Chen, S. D., Ker, S. J., Chen, Y., & Liu, J. S. (1994, June 1994). *A multiple-corpus approach to recognition of proper names in chinese texts.* Paper presented at the Computer Processing of Chinese and Oriental Languages.
3. Chen, H. H., Ding, Y. W., & Tsai, S. C. (1998). Named entity extraction for information retrieval. *Computer Processing of Oriental Languages, 24*, 75-85.
4. Finin, T., Ding, L., Dornbush, S., Doshi, V. C., Java, A., Kolari, P., et al. (2004). Swoogle semantic web search engine. 3.1. Retrieved 08/08, 2006, from http://swoogle.umbc.edu/
5. Lee, C. S., Liao, J. X., & Kuo, Y. H. (2002). *A semantic-based concept clustering mechanism for chinese news ontology construction.* Paper presented at the International Computer Symposium, Taiwan.
6. Lindesay. (2003), V. Schemaweb - rdf schemas directory. Retrieved 08/08, 2006, from http://www.schemaweb.info/default.aspx
7. Miller, D., Schwartz, R., Weischedel, R., & Stone, R. (1999). *Named entity extraction for broadcast news.* Paper presented at the Proceedings of DARPA Broadcast News Workshop.
8. Radev, D. R., & McKeown, K. R. (1998). *Generating natural language summaries from multiple on-line source.* Paper presented at the Computational Linguistics.
9. Salton, G., & McGill, M. J. (1983). *Introduction o modern information retrieval*. New York: McGrw-Hill Co.
10. 吳家威, & 劉昭麟. (2002). 應用本體論設計與建置摘要系統, *民生電子研討會論文集*.
11. 李振昌, 李御璽, & 陳信希. (1994). *中文文本人名辨識問題之研究.* Paper presented at the Proceedings of ROCLING VII.
12. 翁頌舜, & 許正欣. (2004). 於語意網上自動化建構本體論之研究. Paper presented at the 2004 臺灣商管與資訊研討會論文集(光碟片).
13. 許登傑. (2005). 應用主題地圖理論建置知識索引研究. Paper presented at the 2005「開放原始碼」技術與應用研討會, 成功大學數位生活科技研究中心.
14. 陳雅絹. (2003). 基於 ontology 之模糊代理人於中文新聞文件摘要技術之研究. In 國科會.
15. 黃純敏, 郭家良, & 楊顓溥. (2004). *新聞知識管理系統之建構與評估.* Paper presented at the 第十屆資訊管理暨實務研討會.
16. 黃純敏, 戴尚學, & 郭家良. (2003). 新聞事件自動偵測與追踪及多文件摘要系統研究, *中華民國九十二年全國計算機會議*: 教育部.
17. 楊昌樺, & 陳信希. (2004). *以語法分析為輔建立新聞名詞知識庫.* Paper presented at the The Association for Computational Linguistics and Chinese Language Processing.
18. 龔俊杰. (2000). *具物件導向式 ontology 自動建構能力之個人化 xml 資訊服務系統.* 國立成功大學.