

異質資料源之新聞查詢與事件追蹤系統

A News Query and Tracking System for intermediary heterogeneous data source

陳志達

南台科技大學資訊管理所副教授

andypony@mail.stut.edu.tw

邱俊銘

南台科技大學資訊管理所研究生

m9590236@webmail.stut.edu.tw

摘要

網路迅速發展之下，資料量日益增多，使得搜尋結果中常常找到許多並非符合使用者需求或網頁鏈結失效的資訊；因此，本研究為了克服資訊來源失效與提升資料搜尋之準確度，故以 RSS 和語意網路技術為基礎，提供一個自然語言查詢和新聞事件關聯與追蹤模組，以建置一個異質資料元處理的中介系統，希望提供高準確性的搜尋技術和完整的資訊整合。首先收集目前存在於網際網路上的 RSS 文件並加以整理分類，再利用本體論對資源描述的能力進行推論，得到相關資訊並產生語意註解的文件，而使得搜尋上更能提升其準確性；另一方面，自然語言和新聞事件能夠讓使用者提升其便利性，並且在使用者介面上利用 Ajax 技術，達到即時的互動，並有效的整合功能，提供使用者真正需要且有用的資訊，提升系統整體搜尋的品質與效率。

關鍵字：語意網路、自然語言查詢、事件追蹤、本體論、註解。

Abstract

With the rapid development of Internet, Information is more and more abundant than ever. This causes the problems that search results on Internet often can't meet users' demand and failure of website link. To solve the mentioned problems, we adapt the technology of RSS and Semantic Web. In this paper, the goal is to build an intermediary heterogeneous data processing system that provides high accuracy and integrity of the Internet search results. First, the RSS document will be collected and be filtered to some classification. Second, the ontology of information description will be used to complete the machine reasoning. We also use the annotation of documents to maintain the information what users really want to get. On the other hand, the design of news event tracking and query by design natural language provide more efficient user interaction and searching results.

Keywords : Semantic Web, Ontology, Annotation, event tracking, Natural language.

一、前言

第一代的全球資訊網起初只是為了分享資料，但隨著網際網路的蓬勃發展，資料量不斷的增加，以致於想要在廣大的網路中，找尋到符合使用者需求的資料，往往不是那麼容易的，因此，為了提高資料搜尋的準確性，有學者提出當網路的資源漸趨龐大的同時，需要有一個可以理解網路上資訊的機器。

近年來的另一個概念「語意式網站」(Semantic Web)也被提出，稱為第二代全球資訊網。Tim Berners-Lee 定義語意式網站為「一個可以被機器所理解的網站」，同時也是一個資訊的集合體。為了達到語意網的目的，採用的方式為使用本體論定義不同領域所用到的知識，這些知識包含字彙和關係，本體論以 XML based 的方式表達以方便網路資源存取。語意網[17]中的本體論(ontology)運用在網路的資訊表達，可達到兩個功用：分類(taxonomy)和推論(reasoning)，分類為了將不同類別的資訊作區分，並可將之視作階層化的表示，而推論知識結合類別與階層性的關係，將隱性知識發掘出來。

由於網際網路的資訊量相當龐大，維護不易，容易造成鏈結失效或結果無法滿足使用者需求的問題。因此，本論文之研究目的如下：

1. 利用 RSS(Really Simple Syndication)自動的蒐集資料，並且加以分類，讓資料更有組織性。
2. 透過 Semantic Web 和 Annotation 去建置檢索機制，以提升資訊搜尋的準確度並且提供於自然語言的查詢。

3. 匯總所有新聞事件，讓使用者在檢閱新聞時，可以針對單一新聞事件進行檢索，如“911 事件”的相關資訊和重要時間，使用者皆可以完整的閱覽，增加其閱讀新聞之便利性。
4. Web 應用在 Internet 中愈來愈多元化，而許多技術的加入，使得伺服器負擔過重，因此使用 Ajax 技術之非同步性以降低伺服器的負擔。並能與使用者更緊密的互動。

二、相關研究

以下將針對與本研究相關領域之重點部份進行探討，包含 Semantic Web、語意註記。

2.1 Semantic Web

語意網是呈現知識本體語言中之一種標準語言，這種形式可以讓電腦理解其中的語意。目前所制定規範的規格有 RDF(S)[14]、DAML + OIL[18]、OWL[19] 等等。

2.1.1 資源描述架構 (RDF(S))

W3C 發展出 RDF(Resource Description Framework)，使能夠在 Web 上交換 metadata，強化關係描述與資料處理的能力。RDF 是一個三元架構，其內容包含三種物件型態：資源(Resource)、屬性(Property)、敘述(Statements)。

- (1) 資源(Resources)：事情可以由 RDF 描述就稱為來源，亦為資料來源。
- (2) 屬性(Properties)：屬性可以是來源的觀點、特徵或是關係。每個屬性都有一個意義，定義許可的值。描述來源

的型別和其他屬性的關係。

- (3) 敘述(Statements)：敘述的語句以RDF的格式表示。敘述被分成三個部分，分別是Subject、Predicate 和Object。Subject 可以表示來源，Predicate 可以表示屬性，Object可以是文字或是其他的來源。

RDF Schema 提出了抽象世界中主要關係，而建立了類別系統，從而支援了從客觀世界到抽象世界的映射，也成了知識分享的基礎。屬性就是資源間的關係，其中包含類別、限制的概念、屬性從屬的概念，這樣就可以基於RDF Schema為某個領域定義出自己的Schema。然而RDF 對於知識的表示能力上仍有所限制，缺乏支援變數(Variables)、一般性定量(General Quantification)和規則(Rules)的功能。

2.1.2 DAML +OIL 與 OWL

近年來，語意網路語言不斷演化，其語意定義更加豐富完整，DAML + OIL 由歐洲與美洲的聯合委員會所發展的語意網路標準語言，其定義描述的語法是承續RDF(S)的基礎之上，再更進一步制定更豐富更嚴謹的標記定義。

OWL 是 W3C 所公佈網路本體描述語言，除此之外，OWL 的語法是以 DAML+OIL 為基礎，主要的功能是用來定義 Web Ontologies 以及 Ontologies 的相關知識基礎。OWL 的應該是無限領域的，而且允許資源可明確的描述反推(negative information)。目前 OWL 有三種規格：OWL Lite、OWL DL、OWL Full。

2.2 語意註記(Annotation)

網際網路發展至今，網頁的數量已經到了難以估計的量，語意網路更是往後發展的趨勢所在，然而註解系統的發展是為了建立一個網路文件語意化的世界，讓透過網路能夠準確取得正確的資料。

Swoogle[11]是一個在語意網路上搜尋與註記的搜尋引擎。它從每個文件上擷取出一些註記，並且計算每個文件的關聯。它認為語意網路跟一般網頁(HTML)是平行發展的，語意網路只是在一般的文件上給予語意的註記。

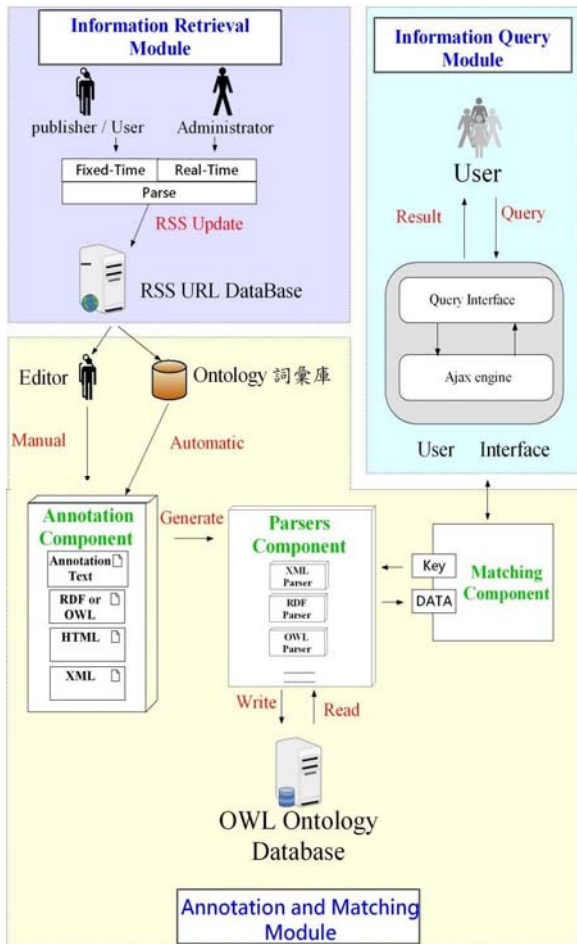
SMORE(Semantic Markup, Ontology, and RDF Editor)[4]是一套資源註記的工具，它提供使用者對於網頁、電子信箱和其他線上內容的整合環境。SMORE 允許使用者在圖片的地方使用 SVG 去註記。

三、系統架構

本系統功能大概可分為三大部份：

- (1) 資訊擷取模組：主機定時或及時去擷取資訊，剖析所需要之資料，並加入 RSS URL Database。
- (2) 註解與比對模組：將擷取蒐集來的資料，加入語意和註解的技術，讓搜尋引擎增加其準確性，並且將其文件透過剖析器分析後，存入其 OWL Ontology Database。在比對方面，必須確保使用者語意上是否和系統取出資料後一致，再回傳給使用者。
- (3) 使用者查詢：使用者在檢索或搜尋自己想要的資料，透過使用者界面，並且在此界面加入 AJAX 技術，使得檢索上更加快速且方便，並且降低伺服

器負擔；另一方面，使用者所送出的關鍵字透過比對模組去比對，找出最符合使用者需求之資訊。圖一為系統架構圖。



圖一 系統架構圖

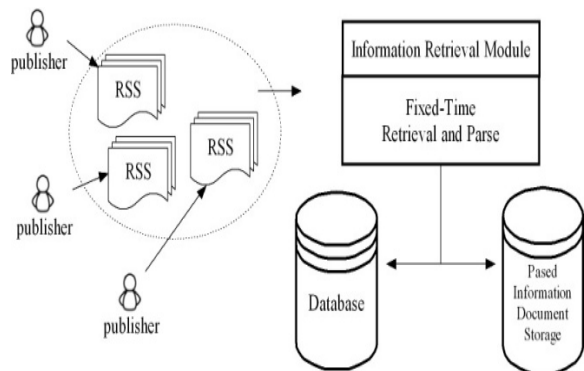
3.1 資料擷取模組(Information Retrieval Module)

資訊擷取模組主要有三項功能，以下將分別對這三項功能作詳細說明。

3.1.1 定時擷取(Fixed-Time)

系統將會預設每 30 分鐘(可自行修改時間間隔)擷取 RSS URL Database 內所有 RSS 資料進行剖析與儲存。此功能主要是與比對模組和整合模組配合，由於比對或

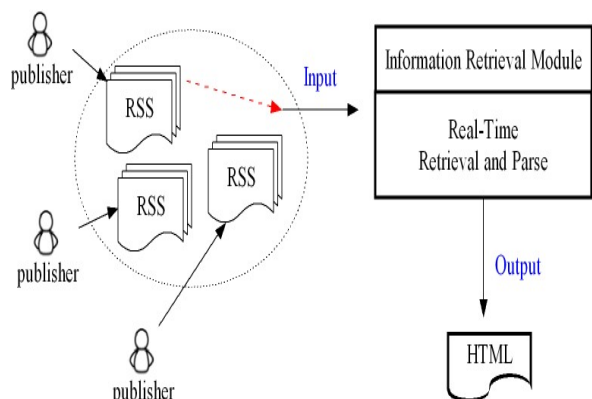
整合可能必須要對 Parsed Information Document Storage 內的許多相關文件進行搜尋與比對，若每次比對或整合必須再一次到網路上擷取 RSS 文件，其搜尋效率勢必會大打折扣，因此系統規劃定時擷取功能，定時至網際網路上擷取並儲存 RSS 文件。圖二為定時擷取功能資料模型。



圖二 定時擷取功能資料模型

3.1.2 即時擷取(Real-Time)

使用者可透過此功能，檢視 RSS URL Database 內目前所儲存的 RSS 資訊，直接擷取資訊提供者發佈在網際網路上最新的那一份 RSS 文件，亦可將此功能視為一個只可讀取不能訂閱的簡易型 RSS 閱讀器。圖三為即時擷取功能的資料模型。



圖三 即時擷取功能資料模型

3.1.3 文件剖析(Parse)

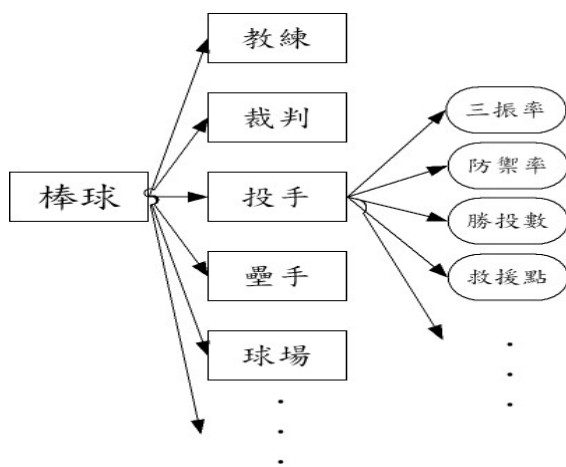
主要功能是剖析擷取後的RSS文件，判斷該文件是否需要更新資訊，並依照分類、日期再進行文件重新命名並存放的動作，避免文件過於龐大，造成執行效率降低。

3.2 語意註解與比對模組(Annotation and Matching Module)

此部份又可分為三個部分，以下將分別對這三項功能作詳細說明。

3.2.1 Annotation Component

註解組件(Annotation Component)，以下我們簡稱為AC，AC所負責的工作就是將資料加入語意和註解，在此分為二個部份，第一個部份便是人工手動(Manual)加入註解，將一些基本的資訊或是難以自動化註解工作，透過人工的方式來完成；第二個部份則是利用 Ontology 詞彙庫去產生自動(Automatic)註解，透過一些基本簡單的詞彙和語意方式去推論產生註解，如：新聞中有“王建民”的名字，王建民便是投手中的一個實例，在推論時，便會依照投手類別應有的屬性進行可能的推論，來產生更多的註解。圖四為棒球之本體論描述。



圖四 棒球之本體論描述

人工所加入註解或是自動註解所以產生的資料送到 AC，以建立一份文件(Annotation Text、RDF or OWL、XML、HTML..)，之後轉至 Parser Module 去進行剖析的工作。

3.2.2 Matching Component

比對組件(Matching Component)，以下我們簡稱為MC，為了提高使用者搜尋的準確性，此部分主要是協調使用者和系統之間的工作，首先，必須先進行剖析使用者查詢所下關鍵字之含意，以提高資料搜尋之準確性，我們在此採用詞性剖析的方法來瞭解使用者可能包含的語意，其剖析的過程中，若所下關鍵字只是一般字詞(如：王建民)，則只需要將此字詞相關資料取出即可，但關鍵字如果是為句子，則便視為自然語言剖析，必須清楚句子中的每個詞性，所以在判斷詞性上，我們採用中央研究院詞庫小組所研發的一套系統：“CKIP 中文斷詞系統[2]”來為我們進行剖析，如“2007年王建民的勝投數？”，透過中文斷詞後如下：“2007年(Nd) 王建民(Nb) 的(DE) 勝投(Na) 數(Neu) 【註1】”，我們先在此句中找出重要的名詞，原因為名詞比其他詞性含有更多的語意[13]，進行剖析，其中取出較固定的名詞：“2007年”、“王建民”和“勝投”，並將此關鍵字丟至 Parsers Component，Parsers Component 從 Ontology Tree 中發現，“王建民”是投手類別中的實例，“勝投”便是投手類別中的屬性，“2007年”亦是屬性之一，便從 Tree 中取出與之相關的資料回傳至 MC，MC 在收到資料後，必須和使用者所下查詢關鍵字進行比對查證，看是否取出資料有錯誤或偏差，在合理範圍內之下，MC 再回傳相關資料給使用者。

註 1：Nd：時間詞 Nb：專有名詞
DE：的，之，得，地
Na：普通名詞 Neu：數詞定詞

3.2.3 Parsers Component

剖析組件(Parsers Component)，以下我們簡稱為 PC，此模組所負責的工作也分為二個部份：

第一個部份是剖析 AC 所送來的文件，剖析文件格式且判斷是否正確，並依檔案類型存入至 OWL Ontology Database。

另一個部份則是負責等待 MC 所需的資料，MC 會根據使用者關鍵字中，透過 Ontology Tree 判斷，最後從 OWL Ontology Database 中取出資料，最後回送至 MC 進行比對的工作。

3.3 資訊查詢模組 (Information Query Module)

在此模組中，使用者介面透過 AJAX 技術降低使用者等待時間，並且在搜尋上採用如 Google Suggest 上的技術，利用下拉式清單回傳可能性的結果數，當關鍵字有改變，將會不斷的搜尋和回傳，讓使用者在搜尋所需的資料更加容易。在檢索方面，除了一般最基礎的基本檢索之外，還有進階搜尋、歷史事件關聯檢索和自然語言檢索。

3.3.1 基本和進階搜尋

使用者依關鍵字或字串去進行資料搜尋，找出使用者所需的資料；進階查詢則可選擇年月份、資料型態、類別...等等去進行查詢。

3.3.2 歷史事件關聯檢索

在網際網路盛行之下，新聞每日發行的數量數以千計，以致於使用者在檢索新聞時，當看到新聞中的一些專有名詞 (如人名、地名、新名詞、未知詞...等)或某單一事件(如 911 事件)時，這些訊息對使用者來說可能不易去瞭解，若想進一步深入瞭解，往往需要花一段時間去做資訊搜尋和蒐集，除此之外，使用者若想持續追蹤這些訊息，亦需要再花費相當多的時間去搜尋資訊，因此，本論文發展了一個模組來解決諸如此類的問題，我們且稱它為“新聞事件關聯與追蹤”(News Event Relation and Tracker; NERT)，此模組主要負責鎖定某一事情、人物、地名進行相關資料搜尋和持續追蹤，分析新聞事件所要表達的語意，並且過濾掉不需要及重複的資訊；在完成每一個新聞事件之後，再將這些新聞事件彼此互相關聯，讓整個模組更加完整，使用者檢索更為方便；此模組所提出的方法和詳細內容我們將在第四章加以說明。

3.3.3 自然語言檢索

在一般最常見的查詢便是關鍵字查詢，但是要選定一個良好且適合的關鍵字就並非這麼容易，許多人容易選了不適當的關鍵字，往往使得搜尋結果不是那麼理想，並且浪費了相當多的時間，因此，自然語言的檢索便是在搜尋領域的目標之一，為了達到這個目標，我們在此建立出 Ontology，希望系統在搜尋上能夠提升其準確性，並且採用 CKIP 中文斷詞系統，判斷出每個字的詞性和意義，以利於進行搜尋和推論。

在推論上，我們必須建立一些規則，並且依照這些規則來建立 Ontology Tree，並將資料區分成每個不一樣的類別，如“大象”是一個動物的類別，就不會出現

在植物的類別上，又如“蘋果”是一種水果，但也有“蘋果電腦”，容易在搜尋上搞混，若在此我們先建立起類別之後，當我們要找尋“蘋果電腦”的時候，就不易找到其他不相關的資訊；在類別建立上，必須建立起“同義”、“互斥”之類需注意的規則，如“本季”、“當季”、“第四季”...等等它們是否同義，以避免搜尋上的誤差。

四、系統實作與 NERT 模組說明

本研究使用語意網中本體論的概念，建立領域本體論描述知識；我們使用 OWL 為網頁資源之描述語言，並且使用 Protégé[12]來建構 Web Ontology 之工具，而在於使用者查詢界面使用 PHP 和 Ajax 技術作為網頁開發工具。

4.1 資訊擷取模組

由於本系統是以 RSS 文件為資訊來源，因此如何從網路上擷取 RSS 文件並加以分類處理，便成了本系統實作時的首要工作。倘若文件內容過大或是文件數量過多，系統又需歷經開檔、讀檔與關檔的過程，則其執行效率勢必會大打折扣，所以，在此除了保留 RSS 文件原本格式與資訊之外，還另存了一份資訊到資料庫之中，以便使用者比對或整合時使用。圖五為擷取模組之實作界面。



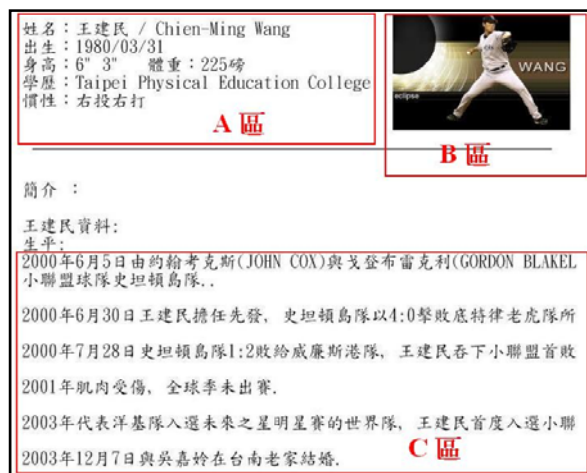
圖五．擷取程式執行畫面

4.2 新聞事件關聯與追蹤

新聞每日發行的數量數以千計，想要找出某一主題的新聞，往往重複性太高或準確性太低；並且，在檢索新聞有時看到專有名詞或未知詞，欲想知道其意義，通常相當費時費力，所以我們提出了一個方法：NERT (News Event Relation and Tracker；新聞事件關聯與追蹤)，使用者在檢索新聞事件時，更增加其“便利性”。

我們主要的做法是：透過本體論 (Ontology) 和建立註解 (Annotation) 去過濾其新聞，以達到較高的準確性去建立新聞事件關聯。

NERT 蒐集的事件主要可分為三類：人物、事物、地方；再依照這些蒐集的類型來完成單一新聞事件。圖六為人物新聞事件範本。



圖六 人物新聞事件範本

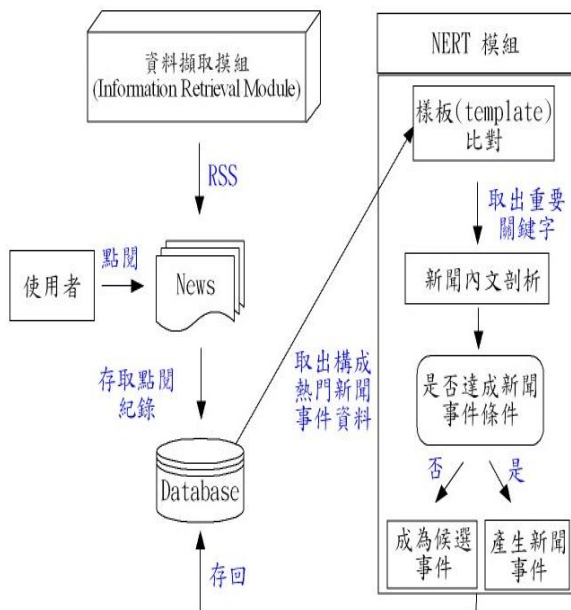
在上圖中，A 區為使用者或管理者建立樣板時，自行輸入的基本資料，好方便讓自己掌握此新聞事件的基本訊息，而 B 區圖片的方面，系統會自動發送“事件名稱”去 Google 或 Yahoo 搜尋圖片，搜尋

之後會列出相關圖片直接讓使用者點選，使用者依喜好程度去選擇圖片；C 區新聞的部份則是利用 RSS 搜集回來，經過系統之權重分析取出一個代表性新聞，每一則新聞都可以經過鏈結到網頁看全文。

而主要單一新聞事件又分成二大部份來進行產生：(1)熱門新聞事件、(2)客製化新聞新事件。

(1) 熱門新聞事件：

由系統模組自動去記錄點閱率夠高的新聞，進而去過濾其新聞內文，看是否達到可以產生“新聞事件”的條件，若條件達成，則從語意庫取出相關資料來完成此事件。圖七為熱門事件流程圖。

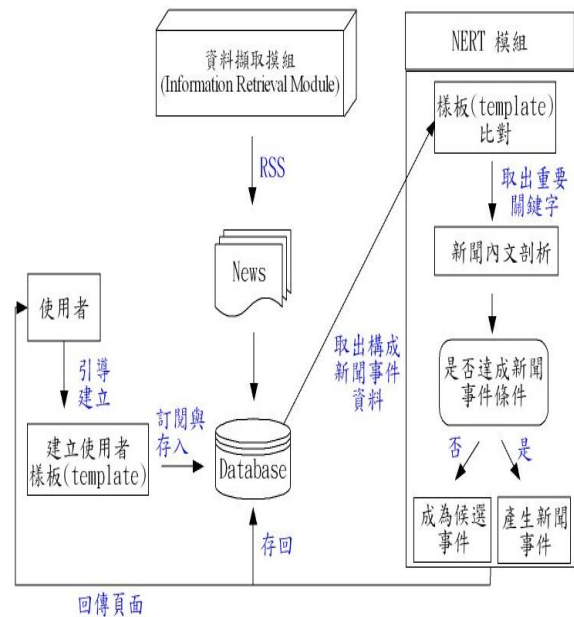


圖七 熱門事件流程圖

(2) 客製化新聞事件：

客製化新聞事件和熱門新聞事件最大的不同是由使用者自行產生範本，使用者透過模組精靈引導輸入一些重要資訊，來

產生新聞事件；在使用者建立好樣板之後，會先存入資料庫，並且送出一份訂閱資訊到 NERT 模組，系統會自動搜尋出相關新聞資訊，再經由演算法編譯整理成新聞事件，最後回傳給使用者。圖八為客製化新聞事件流程圖。



圖八 客製化新聞事件流程圖

NERT 模組說明：

在數量相當多的新聞中，要找出與使用者所需的資訊，其方法大致可分為二類：

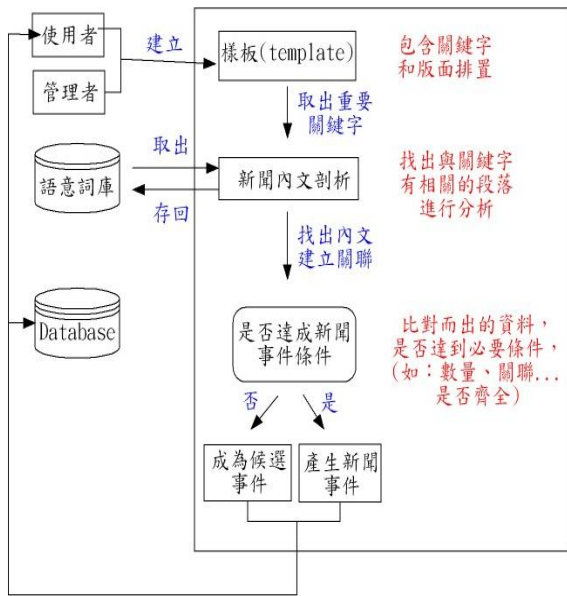
(1) 使用專業領域知識(Domain Knowledge)

去分析文章中的人、事、時、地、物等要素。

(2) 以統計分析(Statistical Analysis)方法直接從原文判斷語句的重要性。

本論文以資訊擷取技術 (Information Retrieval) 為基礎，導入語意網路(Semantic Web)的概念，以找出符合我們所需的文章

段落。圖九為 NERT 模組流程圖。



圖九 NERT 模組流程圖

NERT 模組流程說明：

系統在抓取資料來完成新聞事件，必須先有個依據做為擷取資料的標準，而這個依據便是我們 NERT 中的樣板，所以我們首先便是建立樣板：

(1) 建立樣板：由使用者或管理者建立樣板，使得資料蒐集之後，可以依照個人需求來呈現使用者介面，其樣本的組合包含：關鍵字、排版樣式、權重、基本資料。

A 關鍵字：依每一個事件所專注的重點，建立出關鍵字，如以“棒球”為例，所建立的樣本可能就會以“勝投”、“三振”、“全壘打”……等等來建立其關鍵字，這些字眼對於“棒球新聞事件”來說，就顯的格外重要。

B 排版樣式：在新聞事件搜尋完成之後，可依每位使用者的喜好程度進行排版。

C 權重：管理者和使用者建立樣板時，需建立一些重要的關鍵字，並且設定權重，表示此關鍵字在此新聞事

件中的重要程度，讓系統有個基準去抓取所需的資料。

D 基本資料：建立新聞事件時，輸入的一些基本資料，如事件發起人、成立時間、活動主題……等等。

(2) 新聞內文剖析：

當決定一個新聞事件主題之後，系統會自動找出相關新聞，進而取出所建立樣板(Template)進行比對。在比對時，在同一日期的文章優先進行分析，並且以四個階段去處理新聞文章：取出樣本關鍵字與權重、初步篩選、進階篩選、匯總。

在這四個階段中，我們以“王建民新聞事件”來進行說明：

A 取出樣本關鍵字與權重：

- 王建民 50%
- 大聯盟 25%
- 勝投 15%
- 洋基 10%

B 初步篩選：

(A) 找出所有與關鍵字相關新聞，再以同一日期去進行處理。

(B) 計算權重值，並取出權重值高的二十筆資料。此目的為了篩選出文章中與使用者關鍵字相關性較高的資料，公式如下：

$$BNW = \sum_{i=1}^n (a_i * w_i)$$

其中：

BNW：基本新聞權重值

a：關鍵字有無出現，有出現則為 1，反之則 0

w：個別關鍵字的權重值

n：全部關鍵字個數

例子說明：



圖十 王建民新聞

如圖十，4 個關鍵字皆有出現，故式子如下：

$$\begin{aligned}
 BNW &= \sum_{i=1}^4 (a_i * w_i) \\
 BNW &= (1*0.5) + (1*0.25) + \\
 &\quad (1*0.15) + (1*0.1) \\
 BNW &= 1
 \end{aligned}$$

C 進階篩選

(A) 進而再計算權重值最高的新聞，代表資訊量最豐富。

$$\begin{aligned}
 (B) ANW &= T * \sum_{i=1}^n (tf_i * w_i) + \\
 &\quad (1-T) * \sum_{i=1}^n (f_i * w_i)
 \end{aligned}$$

其中：

ANW：進階新聞權重值

T：標題權重值

tf：標題中的關鍵字出現次數

f：內文中的關鍵字出現次數

例子說明：

T 代表標題的權重值，如果使用者認為標題的重要性大於內文，其將可把值設為高一點，我們在此假設 T 為 60%。

其式子我們可發現分為二個部分，標題和內文，故我們先看標題的部份，如上圖十，標題關鍵字只有出現王建民和勝投各一次，故 $tf_1=1, tf_3=1$ ，所以：

$$T * \sum_{i=1}^n (tf_i * w_i) = 0.6 * (0.5+0.15) = 0.39。$$

而內文部份，王建民和洋基各出現二次，而勝投和大聯盟也各出現出一次，故 $f_1、f_4$ 皆為 2， $f_2、f_3$ 為 1，所以：

$$\begin{aligned}
 (1-T) * \sum_{i=1}^n (f_i * w_i) &\text{ 為 } 40\% * \\
 &\{ (0.5*2) + (0.1*2) + (0.25*1) + (0.15*1) \} \\
 &= 1.6；\text{ 故 ANW 為 } \\
 &1.99(0.39+1.6)。
 \end{aligned}$$

D·匯總：將權重值最高之新聞加入新聞事件中。

(3) 是否達到新聞事件：

在剖析內文之後，將會找出一篇新聞是比較有代表性的，其必須再審核是否達到新聞事件的標準，如果符合標準，我們則建立新聞事件，反之則存入候選新聞事件，我們從二個部份進行檢查：

A·必要關鍵字：在關鍵字中，可依管理者或使用者去設定必要條件關鍵字，如以上例：我們設定王建民為必要條件；則若搜集而已來的新聞皆沒有王建民等關鍵字，則不加入新聞事件。

B·數量：可設定最低數量，如全數搜集而來的新聞數量不到一定程度，則不加入新聞事件。其意義為可能其重要性不夠高。

(4) 存入資料庫與回送給使用者

完成新聞事件後，我們則必須將結

果先存入資料庫中，並且將結果發送給使用者；若是為候選事件，則一樣存入資料庫，但不發送給使用者，其目的在於後續若有相關文章可以使候選事件達到標準，即可轉換為新聞事件。

五、結論

在資訊量如此豐富的時代，能夠準確找到所需求的資訊顯得格外重要，因此，本研究使用語意網中本體論的概念，建立領域本體論描述知識，以提高系統對於資料的可讀性，來增加資訊的準確性；除此之外，為了增加使用者的便利性，本研究亦提出了新聞事件追蹤與關聯模組，希望使用者更能夠降低搜尋資料的時間就能擁有豐富且完整的資訊，同時利用 AJAX 降低使用者等待時間和伺服器的負擔，希望提供一個高效能、效率的一個資訊平台。

參考文獻

- [1] 陳志達、蔡利國，“以 RSS 為基礎之異質資料源中介系統”，TANET 2007，南台科技大學，2007 年 9 月。
- [2] 馬偉雲、謝佑明、楊昌樺、陳克健，”中文語料庫構建及管理系統設計”，中央研究院資訊科學研究學，<http://ckipsvr.iis.sinica.edu.tw/>。
- [3] 劉艾華、余建昇，“以功能性語意註記協助網際網路搜尋之研究”，淡江大學資訊管理系碩士論文，淡江大學，2006 年 6 月。
- [4] A.Kalyanpur, J.Hendler, B.Parsia, and J.Golbeck, “SMORE - Semantic Markup, Ontology, and RDF Editor”, University of Maryland, College Park A. V. Williams Building, 2005.
- [5] A.Ankolekar, M.krotzsch, Thanh Tran, Denny Vrandecic, “The Two Cultures – Mashing up Web 2.0 and the Semantic Web”, ACM, May 2007.
- [6] B.Medjahed, A.Bouguettaya, Ahmed K. Elmagarmid, ”Composing Web services on the Semantic Web”, The VLDB Journal, pp.333-351, 2003.
- [7] D.Jiang, C.Chi, L.Liu, ”An RDF-Based Annotation Framework for Providing Web Resources Integrity”, IEEE, 2006.
- [8] D.Quan, David R. Karger, ” David R. Karger”ACM,17-22,May.
- [9] F.Xie, H.Gong, D.Deng, S.Wang, George T. Wang, J.Hu, P.C-Y Sheu, ” INTEGRATING SEMANTIC WEB SERVICES FOR DECLARATIVE ACCESSES IN NATURAL LANGUAGE”, IEEE, 2006.
- [10] Haihin Wang, Yan-Qing Zhang, Rajshekhar Sunderraman, “Soft Semantic Web Services Agent”, IEEE 2004.
- [11] L.Ding, T.Finin, A.Joshi, R.Pan, R.Scott Cost, Y.Peng, P.Reddivari, V.Doshi, J. Sachs, “Swoogle : A Search and Metadata Engine for the Semantic Web”,CIKM’04,November 8-13,2004.
- [12] Protégé, <http://protege.stanford.edu/>
- [13] R.Angheluta, R.De Busser, M.F.Moens, “The Use of Topic Segmentation for Automatic Summarization”, ACL-2002.
- [14] Resource Description Framework(RDF), W3C, <http://www.w3.org/RDF/>
- [15] R.Guha, R.McCool, E.Miller, “Semantic Serch “, ACM,May 2003.
- [16] Rui G. Pereira, Mario M. Freire, “SWedt: A Semantic Web Editor Integrating Ontologies and Semantic Annotations with Resource Description Framework”, IEEE, 2006
- [17] Semantic Web, W3C, <http://www.w3.org/2001/sw/>
- [18] The DARPA Agent Markup Language Homepage, <http://www.daml.org/>
- [19] Web Ontology Language(OWL), W3C, <http://www.w3.org/2004/OWL/>