整合資訊檢索與臨床醫學本體論之知識管理系統

顏永泰、黃芷瑩、陳博揚、徐青立、徐建業*、王博彥、郝德慧、李友專、劉建財

摘要

臨床醫學資訊分散於各個文獻、資料庫、和病患病歷之中，當臨床醫療人員面臨疾病的預防、診斷、治療和研究工作時，掌握整合的病人資訊將是非常有幫助的，尤其當面對急性傳染病時，臨床醫師需要即時而正確的資訊，以便快速地做出診斷及治療的決策。

爲了縮小知識範疇以利研究的進行，本研究以嚴重急性呼吸道症候群 (Server Acute Respiratory Syndrome, SARS) 爲主題，建置了一個抗 SARS 知識管理系統 (Anti-SARS KMS)，整合和 SARS 相關的電子病歷、胸部 X 光影像、醫院標準作業程序、生物醫學文獻、新聞，以及生物資訊資料庫。並以向量空間模型 (Vector Space Model) 爲基礎，建立一套資訊檢索的程序，以協助臨床醫療人員有效率地檢索到和病人真正相關的資訊。

爲了讓本系統的檢索能力更具彈性與效率，我們使用一套 SARS ontology 來展開使用者的查詢條件，以便檢索到更多相關的資訊。這一套 ontology 以生物醫學、疾病診斷、流行病學、管理與症狀五大分類來涵蓋整個 SARS 相關的知識領域，此 ontology 的建置並不是本研究最主要目的，然而其發展的結果將可套用於本研究之中。

雖然本研究是以 SARS 爲主題，然而運用相同的概念與方法，我們預期可以將研究方法應用到其他的疾病上，如肺結核、禽流感、登革熱。因此，提供一個有效率的方法來整合不同的資料來源，並且依據使用者的查詢條件檢索出相關的資訊，爲本研究的最終目標。

關鍵詞：資訊檢索、Information Retrieval、本體論、Ontology、MEDLINE、MeSH、SARS

壹、前言

臨床醫學資訊分散於各期刊文獻、結構性資料庫與各種病歷資料之中，對於醫療人員來說，尤其是面對急性傳染病，如：嚴重急性呼吸道症候群 (Severe Acute Respiratory Syndrome, SARS)，若能有效率地整合與運用和病人病情相關的資訊，將有助於疾病的預防、診斷、治療與提升病患安全。

本研究以 SARS 爲例，建置一套抗 SARS 知識管理系統 (Anti-SARS KMS)，以整合與管理 SARS 相關的電子病歷、胸部 X 光影像、標準作業程序、文獻、新聞與生物資訊資料庫，並應用資訊檢索中的向量空間模型 (Vector Space Model)，建立檢索系統。

除此之外，爲了檢索到更多相關的資訊，我們應用 SARS 相關的本體論 (Ontology)，擴展使用者的查詢關鍵字，並同時帶入同義詞查詢，最後由檢索系統計算使用者的查詢與文章之間的相似度，依據相似度的大小將查詢結果排序，以幫助臨床醫療人員有效率地檢索到和病人相關的資訊。

貳、研究方法

一、 資料收集與處理

在 SARS 流行期間，總共有 54 位病患於台北市立萬芳醫院被診斷罹患 SARS 並接受治療，因此可以利用這些病患的資料，作爲建置知識管理系統的資料來源。本研究中，包括了電子病歷、胸部 X 光數位影像、標準作業程序、生物醫學文獻、新聞與生物資訊資料庫六項資料來源，詳述如下：

1. 電子病歷：這 54 位 SARS 病患的電子病歷，包括入院病摘及出院病摘，內容描述病患的個人基本資料、醫師的診斷、治療及實驗室檢驗報告。
2. 胸部 X 光影像：由於每一位 SARS 病患皆需

作胸部 X 光檢查，從這 54 位 SARS 病患總共可以獲得 431 張 X 光影像，並可從醫院放射科的 PACS (Picture Archiving and Communication System) 匯出轉成 JPEG 的格式，以利後續影像檢索系統的處理。

3. 標準作業程序：為了因應 SARS 所帶來的衝擊，萬芳醫院制訂了 42 項標準作業程序，包括狀況處理的準則、病患處置的方法等等。
4. 生物醫學文獻：本研究使用由美國 NCBI (National Center for Biotechnology Information) 所開發的 Entrez 檢索系統，並利用 MeSH (Medical Subject Heading) term 分類，從 PubMed 資料庫中收集和 SARS 相關的 MEDLINE 文獻。本研究以 "SARS Virus" MeSH term 和其下的 entry terms，包括：
   (1) SARS-Associated Coronavirus
   (2) SARS-CoV
   (3) Severe Acute Respiratory Syndrome Virus
   (4) SARS Coronavirus
   (5) SARS-Related Coronavirus
   (6) Urbani SARS-Associated Coronavirus

   等作為查詢關鍵字，共檢索出 1233 篇和 SARS 相關的 MEDLINE 文獻。
5. 新聞：台北市政府衛生局在 SARS 流行期間收集了 835 則中文新聞，我們已經合法申請到可以作為學術研究之用。
6. 生物資訊資料庫：由於生物資訊的資料庫分散於世界各地，如美國 NCBI 的基因序列資料庫 GeneBank，日本的 DDBJ，美國的 PDB 蛋白質資料庫等等，為了整合如此分散的資料庫，本研究使用商業化的生物資訊整合平台 SRS，利用它所提供的檢索工具，從各個不同的資料庫中找出和 SARS 相關的資訊。

以上的資料除了胸部 X 光影像和生物資訊資料庫外，其餘的資料皆需事先轉換為純文字格式，以利後續的 Vector Space Model 計算 term weighting。

二、 Term weighting 的計算

本研究從電子病歷、標準作業程序、生物醫學文獻和新聞，建立四個 corpus，並應用 Porter Stemming Algorithm [9] 處理英文字的 morphological 問題 (如：動詞變化、名詞單複數等等) 以減少 corpus 中 vocabulary 的數量。之後應用 Vector Space Model 去計算 query 和 document 之間的 similarity 並依此將查詢的結果作排序。本研究應用 Vector Space Model [4][9]的過程詳述如後：

Vector Space Model 的符號定義：

N： 所有文章的數量。

k_i ： 在詞典(lexicon)中第 i 個索引詞(index term)。

$W_{i,j}$： 在文章 d_j中，索引詞 k_i 的 weighting。

$W_{i,q}$： 索引詞 k_i 在使用者查詢 q 中的 weighting。

$\vec{d_j}$： 文章向量($W_{1,j}, W_{2,j}, \ldots, W_{i,j}$)。

$W_{1,j}$ 受到兩個因素影響：

1. *tf* (term frequency): 索引詞 k_i 在文章 d_j 中出現的頻率(intra-document).
2. *idf* (inverse document frequency): 包含索引詞 k_i 的文章　，在所有文章中出現的頻率(inter-documents)。

因此索引詞 k_i 的 weighting 可以定義成：

$$W_{i,j} = tf_{i,j} \times idf_i \tag{1}$$

（一） Term frequency 的 normalization

n_i 是包含索引詞 k_i 的文章數目且 $freq_{i,j}$ 是 k_i 在文章 d_j 中的原始出現頻率(raw frequency)，因此 k_i 在文章 d_j 中的 normalized term frequency 為

$$f_{i,j} = \frac{freq_{i,j}}{\max\ freq_{l,j}} \tag{2}$$

idf_i 為 k_i 的 the inverse document frequency：

$$idf_i = \log \frac{N}{n_i} \tag{3}$$

因此 k_i 的 term weighting 可以用下列的方式表示：

$$W_{i,j} = f_{i,j} \times \log \frac{N}{n_i} \tag{4}$$

為了運算的便利性，(4)式可以改寫成：

$$W_{i,j} = \left(1 + \ln(f_{i,j})\right) \cdot \ln\left((N+1)/n_i\right) \tag{5}$$

（二） 使用者查詢和文章的相似度

Query Q 可以用上述的方式，以特徵向量 $\vec{q_i}$ 的集合來表示，因此對於每一個查詢的索引詞，皆可計算出一個 cosine 的測量值，來估計使用者的查詢和文章的相似度：

$$R_j(\vec{q}_j,\vec{d}_j) = \left(\vec{q}_j \bullet \vec{d}_j\right) \Big/ \left(\left\|\vec{q}_j\right\| \cdot \left\|\vec{d}_j\right\|\right) \tag{6}$$

最後再乘以每一個索引詞相對應的 weighting 並加總起來，便可以計算出整個使用者查詢和文章之間的相似度：

$$R(Q,D) = \sum_j w_j \cdot R_j(\vec{q}_j,\vec{d}_j) \tag{7}$$

三、 應用 ontology 展開使用者查詢

在大多數 keyword-based 的檢索系統中，最常發生的問題是，當系統的辭典沒有包含使用者的查詢關鍵字時，就檢索不到任何結果。爲了解決這樣的問題，本研究依據已經建立好的 ontology 去展開使用者的查詢關鍵字，並計算與調整每一個查詢關鍵字的 weighting。

本研究裡所用到的 SARS ontology，是由一位胸腔科專科醫師，徐青立醫師所建立的 [10]，總共有五個分類，包括生物醫學、疾病診斷、流行病學、管理和症狀。以下詳細描述應用 ontology tree 展開使用者查詢關鍵字的流程(圖 1)：

1. 同義詞的擷取：透過一個事先定義好的同義詞對照表，找出使用者的查詢關鍵字，所有可能的同義詞。
2. Ontology term 的擷取：在 ontology tree 裡檢查是否含有該查詢關鍵字及其同義詞，並找出其 sibling、上一層的 parent、下一層的 child node 的 term，再對這些 term 做同義詞的展開，最後對所展開的查詢關鍵字做 weighting 的計算與調整。
3. Weighting 的計算：使用者查詢關鍵字 weighting 的計算是依據表 1 的原則，使用者的原始查詢關鍵字的 weighting 定爲 1，在 ontology term 裡的 parent node term 爲 0.4，sibling node term 爲 0.2，child node term 爲 0.6，各自同義詞的 weighting 爲本身 weighting 的一半。這些 weighting 的設定可以依據使用者的需求而做調整。當使用者的查詢關鍵字被展開後，會計算出一個 total weighting (W_{total})

$$W_{total} = \sum_i (Wt_i \cdot k + \sum_j Ws_{ij}) \tag{8}$$

Wt 和 Ws 爲 ontology 裡，term class 和同義詞的 weighting；k 爲查詢關鍵字在 ontology 裡，所屬的 term class 數目，例如某一個關鍵字在 ontology 裡有兩個 child node，則 $k = 2$。

以 W_{total} 的倒數爲 unit weighting，來 normalize 每一個查詢關鍵字的 weighting，使得所有展開後的查詢關鍵字，其 weighting 總和爲 1。

$$W_{unit} = \frac{1}{W_{total}} \tag{9}$$

每一個展開後的查詢關鍵字，其 weighting 爲所屬的 term class 和同義詞的 weighting，乘以 unit weighting 後的總和。

$$W_{iquery} = \begin{cases} Wt_i \times W_{unit} \\ Ws_i \times W_{unit} \end{cases} \tag{10}$$

因此使用者的原始查詢關鍵字經過 ontology 展開，在重新調整 term weighting 後，便可套用 Vector Space Model 來計算查詢與文章之間的相似度。

四、胸部 X 光的影像檢索

本研究使用一套 open source 的軟體 GIFT (GNU Image-Finding Tool) [2][7]，它是一個 content-based 的影像檢索系統，主要的功能是從影像中擷取顏色、材質等的特徵值，建立特徵值的 dictionary 和 inverted file 來進行 query by example (QBE) 方式的檢索。因此可以用一張或多張的影像對整個影像資料庫做相似度的計算。

爲了減少系統上線後的檢索時間，本研究利用 GIFT 的 QBE 檢索方式，事先將每一張的胸部 X 光影像和所有 431 張影像的相似度計算出來，並將排序的結果儲存爲文字檔，作爲系統檢索的 index。

五、系統建置

Anti-SARS KMS 的使用者介面以 *JSP* 和 *Macromedia Flash* 來完成，檢索系統的 Vector Space Model 以 *Java* 來撰寫，整個系統建置在 Mandrake Linux 10 搭配 Apache、Tomcat 和 MySQL。

此外，本系統的資料來源除了生物醫學文獻爲純英文之外，其餘的電子病歷、新聞、以及標準作業程序，內容均爲中英文參雜，爲了讓本系統能適用於中英文檢索，在計算 term weighting 時，以事先制訂好的中英文辭典斷詞，不論中文或英文字皆可計算出其在文章中的 weighting。並且在同義詞對照表內，已將相對應的中英文定義爲同義詞，因此不管使用者的查詢關鍵字爲中文或英文，經過同義詞展開後，皆可將相對應的中英文加入爲查詢關鍵字。如此便讓中英文的查詢關鍵字均能檢索到相關的資訊。

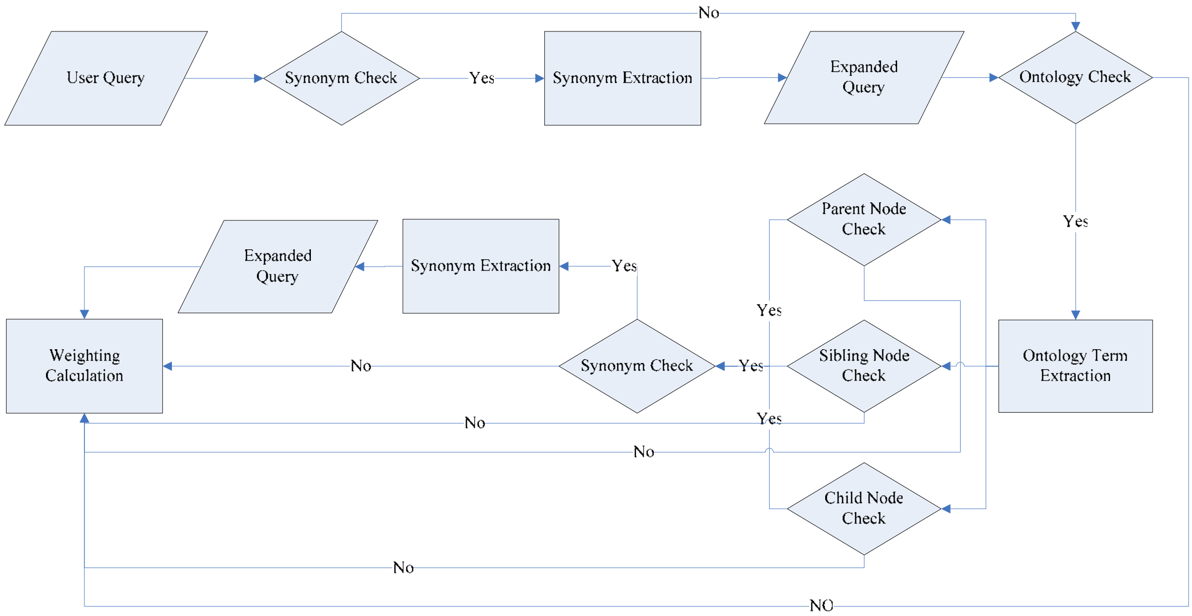

圖 1：應用 ontology 展開使用者查詢關鍵字流程。1. 透過同義詞對照表，擷取出使用者查詢關鍵字的同義詞。2. 利用 ontology tree 展開使用者的查詢關鍵字及其同義詞，擷取出其在 sibling、上一層的 parent、下一層的 child node 的 term。3. 展開後的查詢關鍵字再做一次同義詞的擷取。4. 計算查詢字串中，每一個關鍵字的 weighting。

表 1： Ontology term class weighting 定義表。使用者的原始查詢關鍵字的 weighting 定為 1，在 ontology term 裡的 parent node term 為 0.4，sibling node term 為 0.2，child node term 為 0.6，各自同義詞的 weighting 為本身 weighting 的一半。這些 weighting 的設定可以依據使用者的需求而做調整。

Term Class	Term Weighting	Synonym Weighting
Original Query Term	1	0.5
Parent Node Term	0.4	0.2
Sibling Node Term	0.2	0.1
Child Node Term	0.6	0.3

參、研究結果

目前 Anti-SARS KMS 已經建置在萬芳醫院，並且與醫院資訊系統的急診、門診和住院醫令結合，以利臨床醫師做即時的查詢。

自 2003 年 7 月系統於萬芳醫院的內科門診、急診、加護病房、負壓隔離病房上線後，將近有 121 名醫師與護理人員使用本系統，每月平均有 300 次的查詢，且院外有疾病管制局與衛生署國際合作處兩個單位使用本系統。另外，本系統除了有 SARS 相關的資料外，並已加入禽流感、肺結核和登革熱相關的生物醫學文獻與新聞。

為了能幫助臨床醫師快速地檢索和病患有關的資訊，一些常用的選項，如症狀、病史、實驗室檢驗項目、胸部 X 光的觀察情形等，已經設計在使用者介面上(圖 2)。

在本研究中有兩種使用者介面：診斷評估表(Encounter form)，用來篩選新的 SARS 病例(圖 2)；追蹤治療表(Follow-up form)，用來追蹤 SARS 病患或疑似 SARS 病患(圖 3)，兩個介面的不同之處在於：常用選項的內容與只有追蹤治療表可以從病歷中檢索到胸部 X 光影像。

圖 2：診斷評估表(Encounter form)。藉由勾選主述症狀、過去病史、檢驗項目、胸部 X 光所見等項目，提供臨床醫師相關的資訊來篩選新的 SARS 病患。

圖 3：追蹤治療表(Follow-up form)。藉由此介面所檢索出來的資訊，幫助臨床醫師追蹤 SARS 病患或疑似 SARS 病患。

Anti-SARS KMS 的系統流程如圖 4，並詳述如下：

1. 使用者的查詢關鍵字經過 ontology search 和同義詞對照表被展開爲更多相關的關鍵字，再透過表 1 中 ontology term class weighting 的定義去計算每一個查詢關鍵字的 weighting。
2. 展開後的查詢字串，透過 Vector Space Model，計算和其與文章之間的相似度，並依據相似度分數的大小，將結果予以排序。由於本研究有四個不同的資料來源(電子病例、標準作業程序、生物醫學文獻、新聞)，因此一次的查詢會得到四個排序的結果。
3. 四個排序的結果(電子病例、標準作業程序、生物醫學文獻、新聞)會顯示在同一個網頁的四個獨立分割區域(圖 5)。每個排序的結果只有前三筆記錄會顯示在這一個網頁上，並提供顯示全部查詢記錄的連結。此外生物資訊資料庫的檢索結果會顯示在 SRS 的介面上，此介面會以鑲嵌的方式，顯示在檢索結果網頁的另一個分割區域。
4. 胸部 X 光影像只會顯示在追蹤治療表的電子病例檢索結果上(圖 6)。臨床醫師可以挑選有興趣的影像，透過 GIFT 產生的 QBE 結果，進行相似影像的檢索，如圖 7。

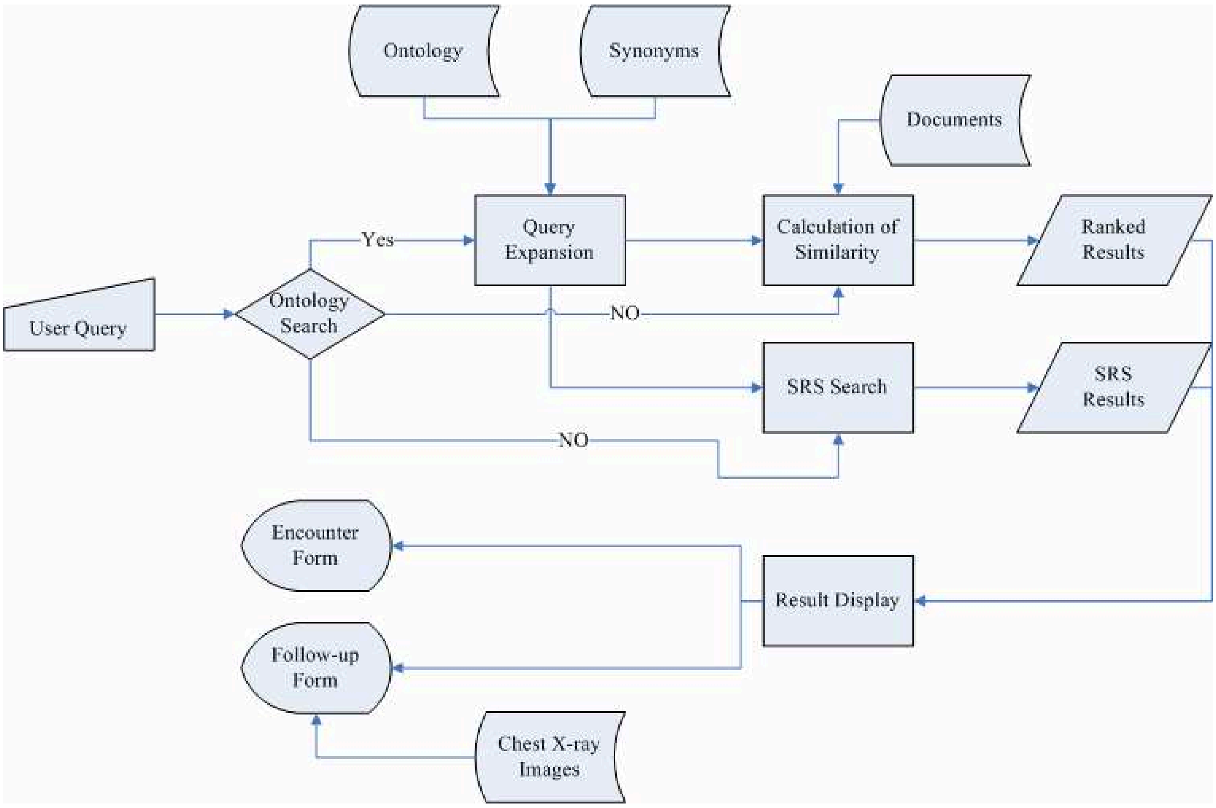

圖 4：Anti-SARS KMS 系統流程。1. 使用者的查詢關鍵字經過 ontology search 和同義詞對照表被展開爲更多相關的關鍵字。2. 展開後的查詢字串，透過 Vector Space Model，計算和其與文章之間的相似度，並依據相似度分數的大小，將結果予以排序。3. 透過 SRS 檢索生物資訊資料庫。4. 胸部 X 光影像只會顯示在追蹤治療表的電子病例檢索結果上。

圖 5：檢索結果顯示介面。以關鍵字「diarrhea」為例，經過同義詞及 Ontology 展開後為 loose stool、flu like syndrome、fever above 38 degree、nausea vomiting、sputum、inspiratory crackle、sob、腹瀉、有痰、吸氣期囉音、喘、氣促、呼吸困難、呼吸急促、呼吸短促等等，以展開後的查詢字串從四個資料來源(電子病例、標準作業程序、生物醫學文獻、新聞) 檢索和 SARS 相關的資料，其檢索結果顯示於同一個網頁的四個獨立分割區域。

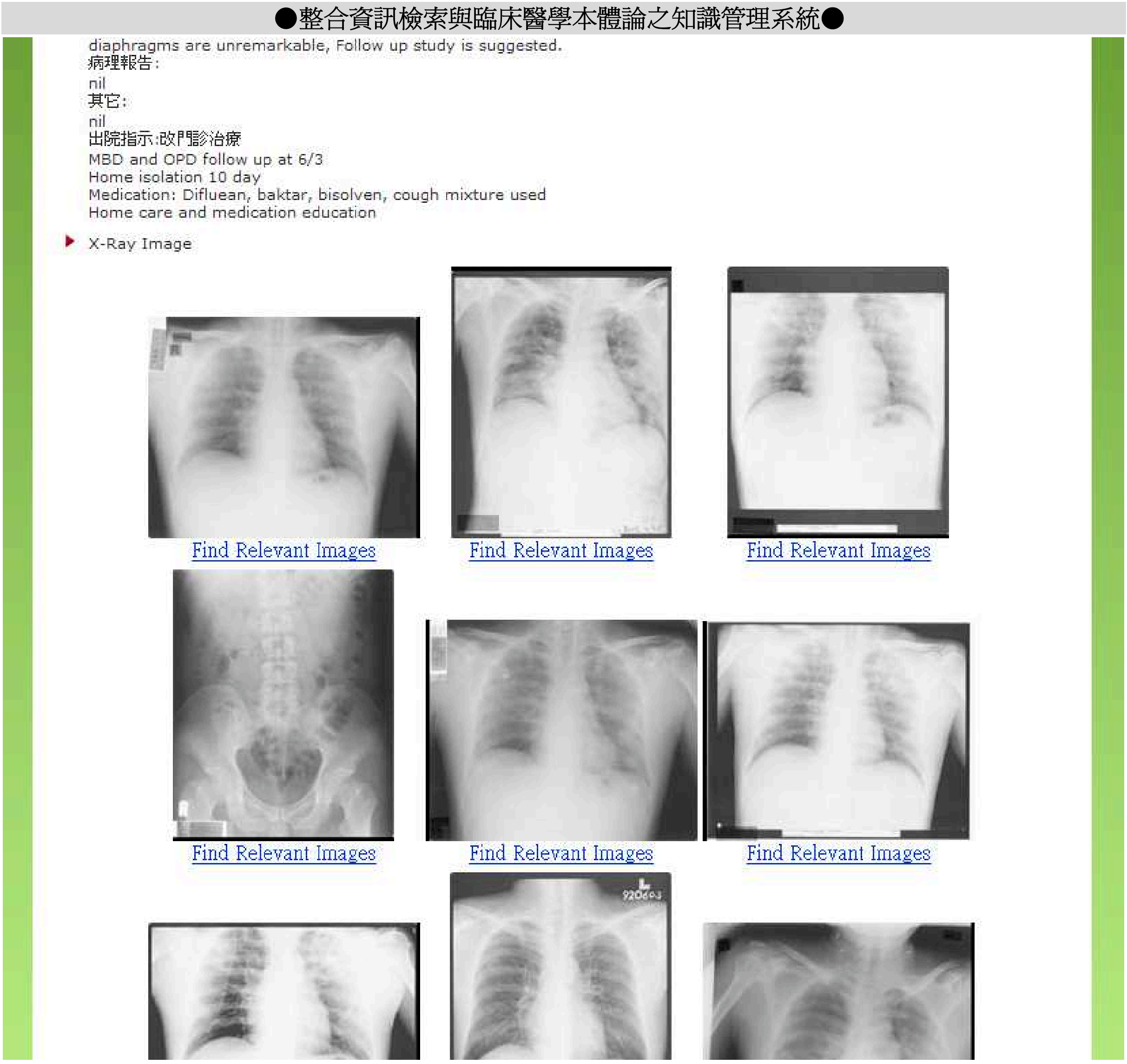

圖 6：追蹤治療表的電子病例檢索結果。內容包括了病患基本資料、醫師的診斷與處置、實驗室檢驗報告以及與胸部 X 光影像。每一張影像皆可透過 GIFT 產生的 QBE 結果，檢索到相似的影像。

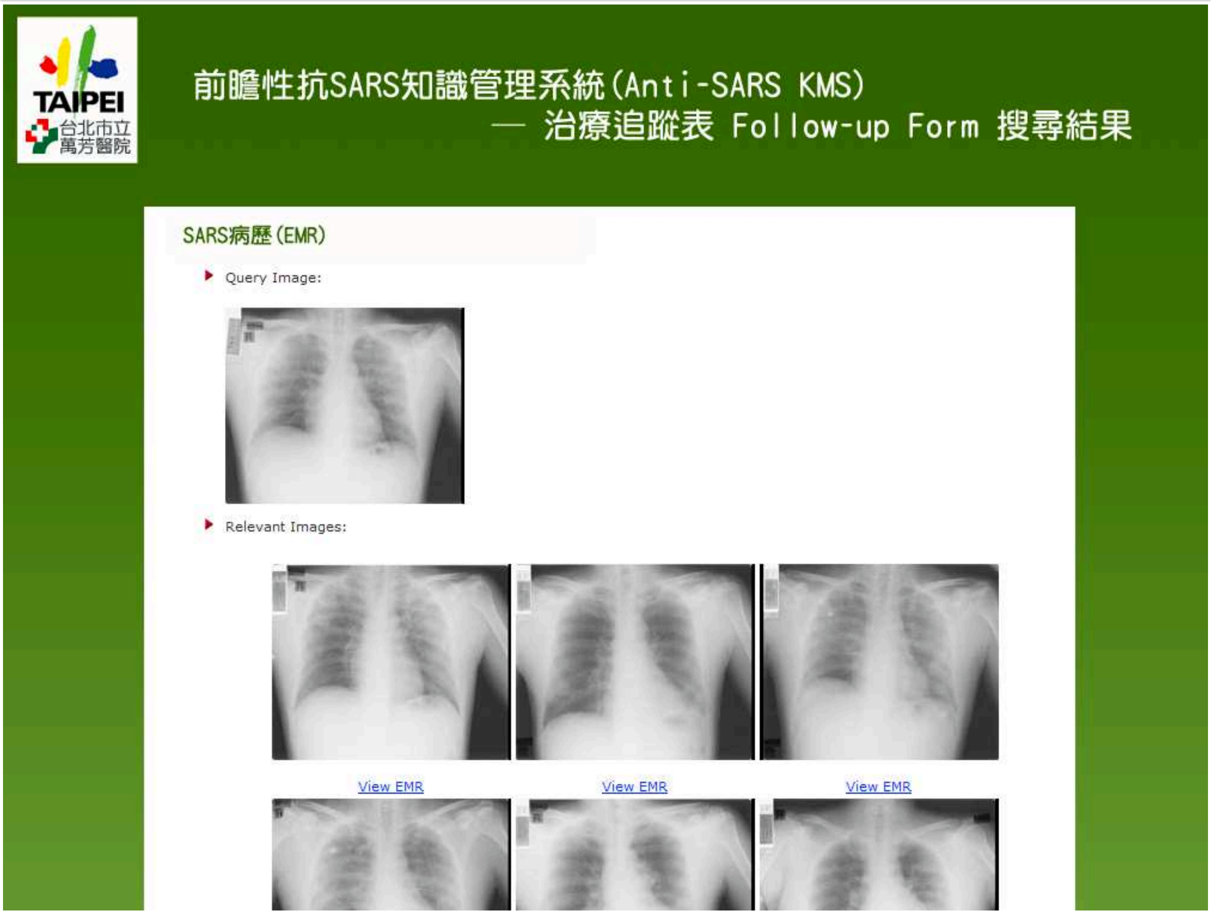

圖 7：胸部 X 光影像檢索結果。以圖 6 中其中一張影像當作查詢影像(Query Image)，透過 GIFT 產生的 QBE 結果，檢索出相似的影像(Relevant Images)。每一張檢索出來的影像，皆可連結到所屬的電子病歷。

肆、結論與討論

近年來，有許多研究應用資訊檢索的方法，從生物醫學文獻擷取出有用的資訊。Dr. Muller 等開發了一套 ontology-based information retrieval and extraction system [6]，從和 C. elegans 相關的文獻中，找出和基因的相關資訊。另外在 Dr. Ono 等的研究中 [8]，他利用 Vector Space Model 從病人的出院病摘中，擷取出和診斷相關的資訊，以便進行疾病的分類。

另外，有許多研究把焦點放在如何從文獻中擷取出基因和基因之間的關係。Homayouni R.等的研究是利用 Latent Semantic Indexing 從 MEDLINE 文獻的標題和摘要中，分析基因之間的關係 [3]。然而爲了要增加 Vector Space Model 檢索的準確性，Dr. Mao 等將概念性的 phrase 取代原本的 index term [5]，因此查詢和文章之間的相似度由原本單一個 term 的計算方式擴展至由多個 term 所組成的 phrase 去做運算。

此外，Dr. Abasolo 等人指出 [1]，使用者通常花相當多的時間去修改他們的查詢以獲得更好的查詢結果，爲了解決這個問題，他們發展了一個 ontology-based information retrieval agent，利用由專家所建立的 medical ontology 來產生查詢，並藉由查詢結果的評估來重新修正查詢，以獲得更好的結果。

爲了比較 ontology-based search 和 keyword-based search，對於臨床醫療人員使用上的檢索效果，本研究以四種不同的情境問題：

1. SARS 患者的胸部 X 光病灶之分布如何？
2. 我的病人有嘔吐現象，我想知道有關嘔吐的意義。
3. 目前有關 SARS 之檢驗知識如何？
4. SARS 防護的口罩有多種層級，我想知道有關其資訊。

進行系統測試後的問卷調查，總共回收 115 份有效問卷，結果如表 2。除了情境問題二，有 61%的臨床醫療人員認爲 ontology-based search 的方式比較差之外，其餘三個情境問題平均有 62%的臨床醫療人員認爲 ontology-based search 的方式比 keyword-based search 爲佳。

表 2：以四種情境問題請臨床醫療人員進行系統測試，並以問卷調查的方式，比較 ontology-based search 與 keyword-based search 的檢索效果。除了情境問題二之外，臨床醫療人員認爲 ontology-based search 比 keyword-based search 的效果好或兩者相當。

情境問題	Ont.S. 顯著較優	Ont.S. 略優	二者相當	Ont.S.較劣
1	37%	42%	11%	37%
2	7%	11%	21%	61%
3	7%	43%	50%	0%
4	14%	44%	42%	0%

關於 Anti-SARS KMS 的系統效能，本研究以所收集到的 1233 篇 SARS 相關的 MEDLINE 文獻爲 evaluation set，10 個關於 SARS 的生物醫學、診斷、管理、流行病學問題爲 query set，並且請六位醫學專家制訂這 10 個問題 relevant answer set，進行系統效能的評估。以這 10 個問題作爲系統的查詢關鍵字，分別進行 keyword-based 和 ontology-based 的檢索，其結果依據relevant answer set 計算在 11 個 recall level 下的 interpolated precisions ，如：圖 8，其 mean average precision 分別爲：0.569 和 0.472。

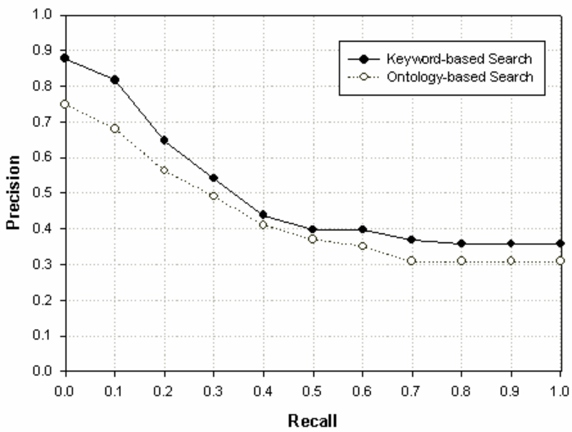

圖 8：Keyword-based 以及 ontology-based search 在 11 個 recall level 下的 interpolated precisions。

由於 ontology-based search 是一種 query expansion 的檢索方式，當查詢關鍵字增加時，可以檢索到更多的資訊，然而 precision 會相對地下降。因此當使用者非常瞭解應該用怎樣的關鍵字檢索時，用 keyword-based search 可以找到較爲準確的資訊；反之使用 ontology-based search，可以幫助不清楚該用何種關鍵字檢索的使用者，尤其是面對新興的疾病時，從所檢索到相似的文章中，找到相關的資訊。

應用資訊檢索的方法，的確可以從文獻中擷取出有用的資訊，本研究運用相同的概念，把資訊檢索的模型建置於知識管理系統當中，來幫助臨床醫療人員，從各個不同的資料來源中，檢索出相關的資訊。另一方面，本研究應用 ontology 的概念，將使用者的查詢自動展開爲更多相關的關鍵字，以幫助他們找到更多相關的資訊。

伍、參考文獻

[1] Abasolo J.M., Gmez M. : MELISA: An Ontology-based Agent for Information Retrieval in Medicine, Proceedings of ECDL 2000 Workshop on the Semantic Web, Lisbon, Portugal (Session 3; 21 September 2000).

[2] Chen S.: Building A Content Based Image Retrieval System of Medical Image on Health Grid, Graduate Institute of Medical Informatics, Taipei Medical University in Taiwan, Master Thesis, 2004.

[3] Homayouni R., Heinrich K., Wei L., Berry M.W.: Gene Clustering by Latent Semantic Indexing of MEDLINE Abstracts, Bioinformatics. 2004:Aug 12.

[4] K. Ng, Survey of Approaches to Information Retrieval of Speech Messages, Draft, SLSG, MIT, 1996.

[5] Mao W., Chu W.W.: Free-text Medical Document Retrieval via Phrase-based Vector Space Model, Proc AMIA Symp. 2002:489-93.

[6] Muller H.M., Kenny E.E, Sternberg P.W.: Textpresso: An Ontology-Based Information Retrieval and Extraction System for Biological Literature, PLoS Biology 2004: Sep 21;2(11):E309.

[7] Müller W., Design and Implementation of A Flexible Content Based Image Retrieval Framework, The GNU Image Finding Tool, Department of Information, University of Geneva, PhD. Thesis 2001.

[8] Ono H., Takabayashi K., Suzuki T., Yokoi H., Imiya A. , Satomura Y. : Extraction of Diagnosis Related Terminological Information from Discharge Summary, Medinfo. 2004.

[9] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval, Addison Wesley Longman, 1999.

[10] Shyu C.L., Hsu C.Y., Hwa K.Y., Li Y.C.: Ontology-Driven Knowledge Management of the Severe Acute Respiratory Syndrome-Related Knowledge Domain—An Example of Medical Informatics Applied on New and Re-emerging Infectious Diseases, APSR 2004.

● An IR-based Knowledge Management System Integrated with Ontology Structure for Clinical Medical Information Retrieval ●

作者簡介

顏永泰 Yung-Tai Yen：臺北醫學大學醫學資訊所碩士班。

黃芷瑩 Chih-Ying Huang：清華大學生命科學所碩士，現任臺北醫學大學醫學資訊所研究助理。

陳博揚 Po-Yang Chen：臺北醫學大學醫學資訊所碩士班。

徐青立 Ching-Lih Shyu：臺北醫學大學醫學資訊所碩士班，行政院衛生署苗栗醫院胸腔科主治醫師。

徐建業 Chien-Yeh Hsu：美國俄亥俄州州立大學資訊博士，現任臺北醫學大學醫學資訊所研究副教授。

王博彥 Po-Yeng Wang：臺北醫學大學醫學資訊所碩士。

郝德慧 Te-Hui Hao：臺北醫學大學醫學資訊所碩士，現任臺北醫學大學護理學院護理資訊兼任講師。

李友專 Yu-Chuan Li：美國猶他大學醫學資訊學博士，現任臺北醫學大學醫學資訊所所長、台北市立萬芳醫院資訊副院長。

劉建財 Chien-Tsai Liu： 臺北醫學大學醫學資訊研究所副教授。

通訊作者：徐建業，台北市吳興街 250 號 臺北醫學大學醫學資訊所，電話：02-23776730 ext. 209，傳真：02-27339049，Email: cyhsu@tmu.edu.tw。

An IR-based Knowledge Management System Integrated with Ontology Structure for Clinical Medical Information Retrieval

Yung-Tai Yen, Chih-Ying Huang, Po-Yang Chen, Ching-Lih Shyu,
Chien-Yeh Hsu*, Po-Yeng Wang, Te-Hui Hao, Yu-Chuan Li, Chien-Tsai Liu

Abstract

Clinical medical information exists in scientific literature, structure database, and patient data such as patient record, medical image and laboratory test report, etc. Consequently, it will be helpful for clinical personnel to have integrated patient information in disease prevention, diagnosis, treatment, and research issues. In addition, when facing an acute infectious disease, it is very important to retrieve relevant patient information for making a correct decision in a short time. Therefore a knowledge management system which can integrate diverse patient data and have efficient information retrieval ability has been developed in this research.

In order to narrow down the knowledge scope, we focused on Severe Acute Respiratory Syndrome (SARS) as an example to build the knowledge management system, Anti-SARS KMS. In this research, we have integrated SARS related data including electronic medical record (EMR), chest x-ray image, standard operating procedures (SOP) in a public hospital, literature, news, and bioinformatics database into the system and developed an information retrieval procedure based on the Vector Space Model (VSM) to locate the relevant information.

To make the search engine more flexible and efficient, the SARS ontology has been developed and applied in this research to expand the user's query for more relevant information. The ontology has five category branches which contain biomedicine, diagnostics, epidemiology, management, and syndrome to cover the SARS related field. This effort is not the main purpose of this research, but the result has been adopted in this research.

Although the Anti-SARS KMS focuses on the SARS topic, we expect the same concept and procedure can also be applied on other diseases. Therefore the main purpose of this research is to provide an efficient way to integrate different data source and retrieve relevant information according to user's query.

Keywords: Information Retrieval, Ontology, MEDLINE, MeSH, SARS
*Correspondent: Chien-Yeh Hsu (E-mail: cyhsu@tmu.edu.tw)