運用自然語言處理技術建研究領域詞彙架構以Healthcare之詞頻分析為例

Using natural language processing technology to construct vocabulary structure – an example of frequency analysis on healthcare domain

祝國忠[1] 林育澍[2] 葉俊成[3]

國立臺北護理健康大學資訊管理研究所

kcchu@ntunhs.edu.tw blake8872000@gmail.com a5594163@gmail.com

摘 要

一般的研究學者而言，在著手研究文獻探討時，大部份仍以人工方式來蒐集資料，可能導致蒐集的時間較為冗長，也可能限制於某特定領域資料庫，或因人工的方式進行整理，而導致處理上有些許的誤差，每年在各個專業領域之中，因研究學者所提出研究文獻越來越豐富，導致要找出各個領域的核心文獻越來越困難，本研究將運用部份自然語言處理技術（natural language processing technology，NLP），以及詞彙頻率分析法（frequency analysis，FA）擷取出特定領域較為相關的詞彙，並透過關聯規則演算公式，以及本體論分享知識的概念，建構出與特定領域相關的詞彙架構，透過該架構可提供查詢特定領域詞彙的方向。

一、 緒論

1-1 研究背景

要了解一研究領域的知識，需要透過一些描繪知識領域的工具，例如引用文獻分析法、本體論等。引用文獻分析法，可說明知識內容的繼承和利用，其被引文獻和引用文獻之間具有的聯繫可找出特定領域之基本理論結構、研究主題。

1-2 研究動機

本研究針對現有引用文獻分析檢索系統中的摘要、關鍵字檢索等功能提出關聯規則演算公式，強化其檢索的效果，接著透過部份自然語言處理與詞頻分析法以及本體論的概念，從文獻摘要中擷取出與領域較相關的詞彙，並建構出特定領域相關的詞彙架構，以協助研究學者查詢特定研究領域詞彙的方向，以及檢索出與領域相關的文獻。

1-3 研究目的

1. 發展關聯規則演算公式：因每一篇文獻可能包含數個與領域相關的詞彙，因此需先得到文獻關聯演算的結果，才可計算出詞彙關聯係數；其文獻關聯演算公式包含引用文獻次數、引用期刊次數等參數，其結果可得到文獻關聯係數；詞彙關聯演算公式的部份則包含詞彙出現頻率，以及該詞彙所在文獻的文獻關聯係數，最後可得到詞彙關聯係數，其係數越高代表對領域是越相關。
2. 建構特定領域擷取詞彙介面：透過引用文獻分析、詞彙頻率分析法以及關聯規則演算公式的輔助，可擷取出定領域中的較為相關的詞彙，最後參考本體論的概念，建構與領域相關的詞彙架構，提供查詢特定研究領域詞彙的方向。
3. 建構詞彙交叉檢索介面：以目的 1 所建構的詞彙架構為基礎，增加詞彙與詞彙交叉檢索功能，以利研究學者能確實檢索出有意義的文獻。

二、 文獻探討

2-1 自然語言處理

在本研究中主要透過 NLP 中的詞性標記(Part-of-Speech Tagging)、文法分析(Grammar Analysis)來描述論文中詞彙之間的關聯，再透過一些方法擷取出有意義的詞彙，並以此關聯架構定義其階層關係[1]。

壹 詞性標記

詞性標記簡稱 POS，進而瞭解句中所表達的事物。例：對「The book has changed the way we about information.」進行 POS 分析，會得到「The/DT, book/NN, has/VBZ, changed/VBN, the/DT, way/NN, we/PRP, about/IN , information /NNS」上述結果[5]。，透過 POS 可找出句中的關鍵詞彙、文法以及文句架構[2]。

貳 文法分析

主要先經由 POS 所產生的結果為基礎，進一步探討詞彙與詞彙之間的關聯，定義出在文句中的資訊[3]。文件中的關鍵詞彙或與主題相關的詞彙，通常由「名詞-名詞」以及「名詞-動詞」所構成。

參 擷取詞彙方法

Gibson、Pearlmutter 與 Loomis 等學者曾提過辨識一個詞彙的詞性與意義涉及許多主觀的想法，大致上句子結構往往不會脫離名詞與動詞等子句[4]，句子結構中一定有名詞，根據不同子句的組合，可產生出不同的句子結構。因此可針對特定的句子結構進行分析，並發展出專門擷取名詞的方法。除了分析句子結構擷取詞彙的方法外，亦可透過詞彙頻率分

析，發展一詞頻加權公式，擷取出較高頻率以及對於研究領域較有意義的詞彙;Robertson 等學者曾提過相關領域詞頻加權公式[5]：

$$w_j(\bar{d},C)=\frac{(k'_1+1)tf'_j}{k'_1\left((1-b)+b\frac{dl'}{avdl'}\right)+tf'_j}\log\frac{N-df_j+0.5}{df_j+0.5}$$

該公式包含詞彙頻率、出現詞彙的文獻頻率以及文獻內文長度等多種因素，給予詞彙加權參數，進而分析該詞彙是否與特定領域相關。

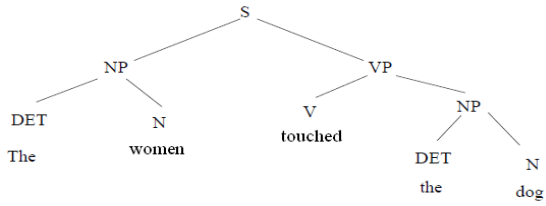


圖 1 句子結構(NP 表示名詞子句,VP 表示動詞子句)

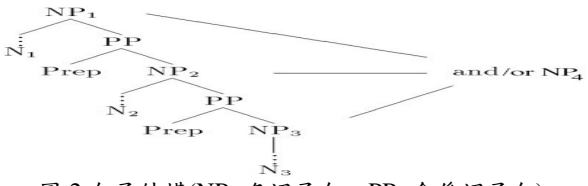


圖 2 句子結構(NP=名詞子句，PP=介係詞子句)

2-2 論文檢索系統

本研究主要使用現有的論文檢索系統作為檢索論文資訊之工具，該系統本身有提供引文分析功能、檢索論文功能以及下載論文全文功能，詳細將透過以下章節進行介紹。

壹　引用文獻分析 Citation analysis

其方法主要是數量的計算，純以計量方式來統計分析文獻、作者、期刊被引用的次數。從引用文獻的網狀系統進行分析引用文獻系統是一種以文獻的引用和被引用所特有的相互關係所構成的文獻資訊體，引用與被引用關係形成了鏈、樹、網型結構。引用文獻網路系統大致可分為以下幾種[6]:(1)時序網路 (2)書目耦合(3)共被引 (4)共同被引作者 (5)共同引用詞彙。

2-3 本體論

Gruber 於 1993 年提出「本體論可以將共享(share)的概念(conceptualization)明確(explicit)的形式化(formal)。」當中說明本體論的四層涵義，包含(1)共享：本體論是被群體共同認可的知識，不是個體所有，而是屬於群體的；(2)概念：指本體論是從真實世界的現象中得到一個概念的模式；(3)形式化：指本體論可以被電腦所解讀；(4)明確性：指概念的使用，概念必須被明確的定義及表達其限制[7][8]。

壹　本體建構方法

在知識本體的建構及使用上，並沒有標準的方法出現，但目前已有許多針對知識本體建構的研究被提出，主要可分成人工建構、半自動建構以及自動建構等方式，

貳　人工建構本體論

Noy 及 McHuinness 所提出知識本體的建構屬於人工建構方式，大致分成以下幾個步驟[26]：1. 決定本體論的領域及範圍：2. 是否有現存的本體論可以採用或延伸使用本體論:3. 列舉領域中的關鍵詞彙：4. 定義領域類別及其階層關係：5. 定義類別的屬性(Property)：6. 建立實體(Instance)：

參　自動化建構本體論法

目前學者所提出自動化建構本體架構的方式[9]：1. 文字分群為主：2. 以關聯規則為主：3. 以字典為主：自動化的部份則是透過關聯規則計算出詞彙關聯係數，以及找出詞彙相關的文獻。由於本研究每經過一段時間，便會更新 Pubmed Central 所釋出的 XML 資訊。

肆　半自動化建構法

透過本體學習(ontology learning)的方式建構本體架構，透過現有的領域本體架構，運用人工的方式將現有資訊進行分析，步驟包含選擇資訊來源、概念學習、確定領域焦點、關聯學習以及評估等步驟。

三、研究方法

3-1 研究步驟

本研究使用 PubMed Central 之論文資料集，擴增現有論文檢索系統的功能，並導入演算公式建構特定領域相關的詞彙架構，以協助研究學者能檢索出有意義的文獻。在開發系統前必須先分析資料集格式，經過許多資料處理程序後，開始建置系統、開發模組，系統建置流程分為七大步驟，(1)解析文獻 XML 檔案、(2)以 WEB 呈現論文資訊、(3)建置詞彙資訊資料庫、更新論文資訊資料庫、(4)擴增論文檢索功能、(5) 開發擷取特定領域詞彙模組、(6)建構特定領域詞彙架構、(7)建構推薦查詢詞彙功能。

3-2 資料蒐集

本研究進行系統開發之資料來源，主要來自存取『PubMed Central』美國國家衛生研究院國家生物科技研究中心所發展與維護的資料庫，可取得全文論文。在文獻資料的取得權限上，PMC 有許多可公開存取之期刊論文，而本研究使用 PMC 提供之『論文資料集』,專門給需要資料挖掘的 XML 檔案而不需要論文 PDF 檔、圖檔或補充資料之研究學者，而此資料集之 XML 格式已遵循 DTD 檔案規範，並且依照期刊名稱整理與排序，每一篇論文都有 PMC_ID，適合用來做資料分析。

壹　資料來源

PubMed Central 是一個開放式的文獻資料庫，提供了四種可存取資料的方法，本研究使用第四種 FTP 方法，下載已封裝好的壓縮檔 article.tar.gz，檔案大小約 6.17GB，於 2013 年 07 月 15 日取得，解壓縮後的大小約 10GB，所有期刊以資料夾方式區隔保存，資料夾名稱以期刊簡寫命名並排序，

貳　資料狀態

本研究所取得之原始資料，共有 571,890 篇文

獻分佈在 1,874 本期刊中，為了解文獻授權開放狀態，將各期刊作數量統計，以觀察此生物醫學文獻檔案之期刊比重分佈。

3-3 前置處理

為了便於導入論文、詞彙關聯係數演算法，因此必須建立相關資料庫以供演算法數據分析，故本研究使用 MySQL 當作系統存取之資料庫，將論文所屬期刊、作者、引用文獻以及論文摘要中的詞彙擷取出來，並存入相關資料庫中，包含期刊、作者、引用文獻以及詞彙分析等資料表。本節將描述相關資料庫建置過程所進行之步驟。

壹 資料表設計

在 MySQL 建置一資料庫為『PMC』，將資料存放於表格『article』、『author』、『reference_paper』、『journal』以及『 vocabulary_list』中，根據『article』所儲存的欄位，依照相似名稱建立資料表欄位，並設定其資料形態。

3-4 關聯係數演算法步驟

本研究參考文獻探討中的自然語言處理技術之擷取知識分析法，另外發展一關聯係數演算公式，藉以分析特定領域中關聯性較高的知識，關聯係數演算虛擬碼如下：

```
PLn  = 所有論文資訊集合[期刊資訊、參考文獻資訊] ; /* JCT(期刊被引用次數)、LJT(目前論文清單期刊的總數量)、PCT(文獻被引用次數)、LPT(目前論文清單文獻的總數)  */
for(i = 1; i <= n ; i++) /* i(目前選取論文索引值)、n(論文總數)  */
PRSi = PRSi-1 + JCTi / LJTi + PCTi / LPTi); /* PRS(文獻關聯係數)  */
VLm  = 所有詞彙資訊集合;  /* TN(詞彙出現次數)、LTT(目前詞彙總數)  */
for( j = 1 ; j <= m ; j++ )  /*  j(目前選取詞彙索引值)、m(詞彙總數)  */
TRSj = TRSj-1 + ( TN j / LTTj ) * PRSi;  /* TRS(文獻關聯係數)  */
x = j ; /*  x 用於排序詞彙關聯係數之索引值  */
while (TRSj >0.02 && x≠0 ) /*開始取得較高關聯詞彙，由高至低排序*/
for ( z = 1 ; z <= x ; z++ )
for( w=z+1 ; w <= x ; w++ )
if( TRSz < TRSw)
VRLz+1= VLz ; /* VRL 用於存入較高關聯詞彙的集合 */
VRLz = VLw;
end
next
next
do( x=x-1 )
next
return  VRL = 較高關聯詞彙的集合
```

步驟 1 主要擷取論文資訊以及計算基本參數，包含期刊被引用次數、目前論文清單期刊的總數量、文獻被引用次數、目前論文清單文獻的總數，步驟 2 到步驟 3 主要計算文獻關聯係數，步驟 4 到步驟 6 主要計算詞彙關聯係數，步驟 8 到步驟 18 主要紀錄較高關聯係數與排序功能，最後步驟 19 將較高關聯詞彙且排序完成的集合回傳至系統，以便於建構領域相關詞彙架構。其公式如下列所示：JS = JCT / LJT – (1)、CS = PCT / LPT – (2)、PRS = Σ(JS1+JS2+…+JSi) + Σ (CS1+CS2+…CSj) – (4)文獻關聯係數(PRS)為論文期刊關聯係數(JS)加上文獻關聯係數(CS)總和。TRS = (TN / LTT) * [Σ(PRS1+PRS2+…+PRSn)] – (5)詞彙關聯係數(TRS)則是將該詞彙所在論文的文獻關聯係數加總，接著與該詞彙出現次數除以目前詞彙總數的結果進行相乘。

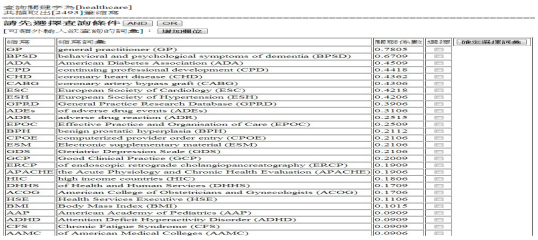
四、 實例分析

4-1Healthcare 分析實例

透過關鍵字 Healthcare 檢索摘要中含有 Healthcare 的論文，取得 8031 篇文獻，詞彙共 2493 個；透過部份自然語言處理分析、詞頻分析方法，再加上關聯係數演算法擷取出關聯係數大於 0.02 的概念，最後建構出領域相關的詞彙架構，並利用 Flash 建構出 Healthcare 研究領域相關的詞彙架構，第一層是與 healthcare 直接相關的縮寫詞，第二層為縮寫詞的全部詞彙，部份架構如下圖所示。

圖 3 Healthcare 領域相關的部份架構

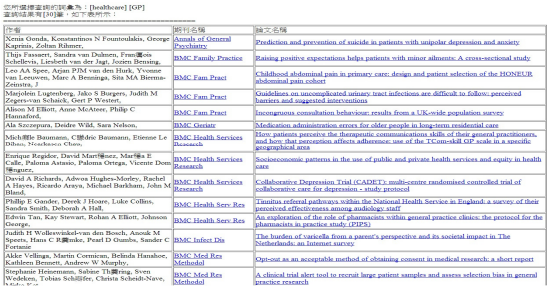
經過系統分析後，Healthcare 研究領域相關的詞彙包含 general practitioner (GP)、instrumental activities of daily living (IADL)、behavioral and psychological symptoms of dementia (BPSD)、american diabetes association (ADA)、continuing professional development (CPD)等詞彙，如下圖所示，參考系統所建構的領域相關詞彙架構，可協助研究學者得知領域之關鍵詞彙的查詢方向，提高檢索相關文獻的效率。本研究分別以 GP、IADL 與 BPSD 等詞彙來進行說明。



GP	general practitioner (GP)	0.7805
BPSD	behavioral and psychological symptoms of dementia (BPSD)	0.6709
ADA	American Diabetes Association (ADA)	0.4509
CPD	continuing professional development (CPD)	0.4418
ET	Emergency Treatment (ET)	0.4382
CABG	coronary artery bypass graft (CABG)	0.4308
ESC	European Society of Cardiology (ESC)	0.4218
ESH	European Society of Hypertension (ESH)	0.4206
GPRD	General Practice Research Database (GPRD)	0.3906
ADEs	of adverse drug events (ADEs)	0.3106
ADR	adverse drug reaction (ADR)	0.2918
EPOC	Effective Practice and Organisation of Care (EPOC)	0.2509
BPH	benign prostatic hyperplasia (BPH)	0.2112
CPOE	computerized provider order entry (CPOE)	0.2106
ESM	Electronic supplementary material (ESM)	0.2106
GDS	Geriatric Depression Scale (GDS)	0.2106
GCP	Good Clinical Practice (GCP)	0.2009
ERCP	of endoscopic retrograde cholangiopancreatography (ERCP)	0.1909
APACHE	the Acute Physiology and Chronic Health Evaluation (APACHE)	0.1906
HIC	high income countries (HIC)	0.1806
DHHS	of Health and Human Services (DHHS)	0.1709
ACOG	American College of Obstetricians and Gynecologists (ACOG)	0.1706
HSE	Health Services Executive (HSE)	0.1106
BMI	Body Mass Index (BMI)	0.1015
AAP	American Academy of Pediatrics (AAP)	0.0909
ADHD	Attention Deficit Hyperactivity Disorder (ADHD)	0.0909
CFS	Chronic Fatigue Syndrome (CFS)	0.0909
AAMC	of American Medical Colleges (AAMC)	0.0906

圖 4 Healthcare 領域詞彙分析表

根據詞彙分析表顯示 GP 詞彙關聯程度最高，關聯係數為 0.7805，其次為 BPSD，關聯係數為 0.6709，接下來為 ADA，其關聯係數為 0.4509，以此類推表示上圖所列出的詞彙為目前 Healthcare 領域較為相關的詞彙。



作者	期刊名稱	論文名稱
Xenia Gonda, Konstantinos N Fountoulakis, George Kaprinis, Zoltan Rihmer,	Annals of General Psychiatry	Prediction and prevention of suicide in patients with unipolar depression and anxiety
Thijs Passarel, Sandra van Dulmen, Frans Kas Schellevis, Liesbeth van der Jagt, Jozien Bensing,	BMC Family Practice	Raising positive expectations helps patients with minor ailments: A cross-sectional study
Leo AA Spee, Arjan PJM van den Hurk, Yvonne van Leeuwen, Marc A Benninga, Sita MA Bierma-Zeinstra, J	BMC Fam Pract	Childhood abdominal pain in primary care: design and patient selection of the HONEUR abdominal pain cohort
Marjolein Lugtenberg, Jako S Burgers, Judith M Zegers-van Schaick, Geri P Westert,	BMC Fam Pract	Guidelines on uncomplicated urinary tract infections are difficult to follow: perceived barriers and suggested interventions
Alison M Elliott, Anne McAteer, Philip C Hannaford,	BMC Fam Pract	Incongruous consultation behaviour: results from a UK-wide population survey
Ala Szczepura, Deidre Wild, Sara Nelson,	BMC Geriatr	Medication administration errors for older people in long-term residential care
Mich?le Baumann, C?dric Baumann, Etienne Le Pihuo, Noelsoo Oleo,	BMC Health Services Research	How patients perceive the therapeutic communications skills of their general practitioners, and how that perception affects adherence: use of the TCom-skill GP scale in a specific geographical area
Enrique Regidor, David Martí?nez, Mar?a E Calle, Paloma Astasio, Paloma Ortega, Vicente Dom?nguez,	BMC Health Services Research	Socioeconomic patterns in the use of public and private health services and equity in health care
David A Richards, Adwoa Hughes-Morley, Rachel A Hayes, Ricardo Araya, Michael Barkham, John M Bland,	BMC Health Services Research	Collaborative Depression Trial (CADET): multi-centre randomised controlled trial of collaborative care for depression - study protocol
Phillip E Gander, Derek J Hoare, Loke Collins, Sandra Smith, Deborah A Hall,	BMC Health Serv Res	Tinnitus referral pathways within the National Health Service in England: a survey of their perceived effectiveness among audiology staff
Edwin Tan, Kay Stewart, Rohan A Elliott, Johnson George,	BMC Health Serv Res	An exploration of the role of pharmacists within general practice clinics: the protocol for the pharmacists in practice study (PIPS)
Judith H Volkwijnkel-van den Bosch, Anouk M Speets, Hans C P?mmke, Pearl D Dumbs, Sander C Fortanie	BMC Infect Dis	The burden of varicella from a parent's perspective and its societal impact in The Netherlands: an Internet survey
Akke Vellinga, Martin Cormican, Belinda Hanahoe, Kathleen Bennett, Andrew W Murphy,	BMC Med Res Methodol	Opt-out as an acceptable method of obtaining consent in medical research: a short report
Stephanie Heinemann, Sabine Th?ring, Sven Wedeken, Tobias Schilffer, Christa Scheidt-Nave,	BMC Med Res Methodol	A clinical trial alert tool to recruit large patient samples and assess selection bias in general practice research

圖 5 Healthcare 領域 GP 檢索結果

針對 GP 詞彙進行檢索後，其結果表示 healthcare、GP 詞彙相關的文獻包含 30 筆，其中 Prediction and prevention of suicide in patients with unipolar depression and anxiety 文獻較為相關，而期刊為 Annals of General Psychiatry 較相關，其結果如上圖所示。

作者	期刊名稱	論文名稱
Elizabeth C Hersch, Sharon Falzgraf,	Clinical Interventions in Aging	Management of the behavioral and psychological symptoms of dementia
Katie Palmer, Massimo Musicco, Carlo Caltagirone,	Int J Alzheimers Dis	Are Guidelines Needed for the Diagnosis and Management of Incipient Alzheimer's Disease and Mild Cognitive Impairment?
Jacob HG Grand, Sienna Caspar, Stuart WS MacDonald,	J Multidiscip Healthc	Clinical features and multidisciplinary approaches to dementia care

圖 6 Healthcare 領域 BPSD 檢索結果

針對 BPSD 詞彙進行檢索後，其結果表示 healthcare、BPSD 詞彙相關的文獻包含 3 筆，其中 Management of the behavioral and psychological symptoms of dementia 文獻較為相關，而期刊為 Clinical Interventions in Aging 較相關，其結果如上圖所示。

作者	期刊名稱	論文名稱
Hodigere SJ Ramesh, Tom Boase, Riccardo A Audisio,	Clinical Interventions in Aging	Risk assessment for cancer surgery in elderly patients
Jacob HG Grand, Sienna Caspar, Stuart WS MacDonald,	J Multidiscip Healthc	Clinical features and multidisciplinary approaches to dementia care

圖 7 Healthcare 領域 IADL 分析表子詞彙分析表

針對 IADL 詞彙進行檢索後，其結果表示 healthcare、IADL 詞彙相關的文獻包含 2 筆，其中 Risk assessment for cancer surgery in elderly patients 文獻較為相關，較為相關的期刊為 Clinical Interventions in Aging，其結果如上圖所示。

五、 結論與討論

本研究利用健康照護 Healthcare 關鍵字為引文分析的實作範例，研究結果發現 Healthcare 研究領域相關的詞彙包含 general practitioner (GP)、instrumental activities of daily living (IADL)、behavioral and psychological symptoms of dementia (BPSD)、american diabetes association (ADA)、continuing professional development (CPD)等詞彙，其中 GP 為最相關詞彙，其次包含 IADL 等詞彙，說明了 Healthcare 領域檢索詞彙的方向。以下針對本研究的限制與討論做詳述。

5-1 系統評估

本節針對本研究所開發的研究領域詞彙分析系統進行評估，與 PubMed Central 文獻資料庫進行數據及內容比較，並說明本系統之貢獻所在。

壹　比較搜尋結果之數量

本研究使用 Healthcare 關鍵字作為分析實例，可取得 8031 篇文獻結果，根據演算法結果可取得 2493 較有意義的詞彙，同時在 PubMed Central 使用 Healthcare 關鍵字進行檢索，包含未釋出的期刊 2503 篇，共取得 10534 篇、在 ScienceDirect 中取得 9,082 篇、在 Web of Science 中取得 57,192 篇、在 PubMed 中取得最多的文獻 842,515 篇。

貳　比較檢索結果

為了評估本系統所分析的準確性是否正確，將相關研究分析的結果與 PubMed Central 的分析結果作比較，即使用 healthcare、general practitioner (GP) 詞彙進行檢索，並排除 PubMed Central 未釋出期刊名稱，其檢索出 30 篇文獻，結果如圖所示，與本系統所分析的結果一致，如圖所示，證明本系統所推薦的詞彙具有準確性。

參　研究貢獻

本研究最大的貢獻是結合了部份自然語言處理技術、詞頻分析法、關聯規則以及部份本體論概念，提供『研究領域詞彙分析』、『推薦檢索詞彙』、『詞彙與詞彙交叉檢索』等功能，若 PubMed Central 完全將期刊釋出，本系統『推薦文獻』與『推薦檢索詞彙』的功能將會更加完善。

	本系統	WOS	PMC	PubMed	SDOL
推薦文獻功能	●				
推薦檢索詞彙功能	●				
詞彙與詞彙交叉檢索	●	◐	◐	◐	◐
引文分析圖表	●	◐			
引用文獻分析	●	●			
研究領域議題分析	●	●	◐		
文獻檢索功能	●	●	●	●	●

表 1　本系統與其他文獻資料庫功能比較

5-2 研究限制

壹　資料匯入程序繁雜

本研究延續原有自動匯入系統，將 XML 檔案資訊批次匯入 MySQL 資料庫，由於原始檔案接壓縮後，就依照期刊名稱建立資料夾，共 1874 個資料夾，系統自動抓取資料夾後，提供點選清單給使用者匯入，而一本期刊必須點選一次匯入的動作，共要點選 1874 次，次數過多；由於主機記憶體容量的限制，篇數超過 500 筆或文獻檔案大小超過

100MB 的期刊必須分批匯入，因此篇數一萬筆的期刊，也要點選 20 次，而延長了資料庫建置時間。

貳　擷取文獻摘要的問題

當文獻統計模組存取 XML 檔案，擷取文獻摘要內容時，其中 XML 標籤的元素< abstract >在少數 XML 檔案中是不存在的，因此系統將排除 XML 標籤無元素< abstract >的 XML 檔案，以避免造成演算法出現錯誤。

參　網頁執行效率

從使用者輸入關鍵字到最後分析，系統都計算了程式執行秒數，以觀察程式運算的效率，各階段所花費的時間都不同。

肆　年份過久無法連結

其中的小功能提供超連結依據 PMC_ID 連結至 PubMed Central 官方頁面，當資訊對稱時，本系統與官方網站的內容是一樣的，但部分文獻由於年份過久，連結到 PubMed 時會出現無此文獻的錯誤訊息，透過標題的搜尋也無法得知。

參考文獻

[1] Kietz, J. U., Maedche, A., & Volz, R. , “A method for semi-automatic ontology acquisition from a corporate intranet.”, Proceedings of the EKAW’2000 workshop on ontologies and texts, 2000.

[2] E. Garfield, "Can citation indexing be automated?", Essasy of an Information Scientist, vol. 1, pp. 84-90, 1964.

[3] 劉佳宗, "利用機器學習摘要概念為基礎之文件摘要自動建立方法", 國立成功大學資訊管理系碩士論文, 2005.

[4] Julie A., “A syntactic predictor to enhance communication for disabled users.”, Department of Computer and Information Sciences, University of Delaware Newark, 1991.

[5] 殷蜀梅,張智雄,吳振新, "一種從醫學文本中實現自動關鍵詞抽取和篩選的技術方法", 現代圖書情報技術, vol. 8, pp.31-36, 2008.

[6] 陳光華, 江玉婷, 莊雅蓁,許雅淑, "引文分析研究發展現況", 書府, vol. 18,19, pp. 15-47, 1998.

[7] T.R.Gruber,“A translation approach to portable ontology specifications. ”, Knowledge Accquisition, Vol.5, No.2, pp. 199-220, 1993.

[8] Gurber, T., SRKB mailing list. 1994(cited from Uschold, M., Gruninger, M., " Ontologies: principles, methods and applications.", The knowledge engineering review 11(2), 1996.

[9] Harith Alani, Sanghee Kim, David E. Millard, Mark J. Weal Wendy Hall, Paul H. Lewis and Nigel R. Shadbolt, “Automatic Ontology-Based Knowledge Extraction from Web Documents.”, IEEE Intelligence System, Vol. 18, Issue: 1, PP.14-21, 2003.

[10] Guarino, N., “Understanding, building and using ontologies.”, International journal of human and computer studies 46(3/4), 219-310, 1997.

[11] Inna Novalija., Dunja Mladenic., Luka Brade ko., “OntoPlus: Text-driven ontology extension using ontology content, structure and co-occurrence information.”, Knowledge-Based Systems 24,1261–1276, 2011.

[12] Weng, S. S., Tsai, H. J., Liu, S. C., & Hsu, C. H. , “Ontology construction for information classification.”, Expert Systems with Applications, 31, 1-12, 2006.