

本體論與文字探勘在注意力經濟下之 應用研究與價值

■ 葉乃菁 / 國研院科技政策中心助理研究員

一、注意力經濟（attention economy）

隨著網路應用的普及化，資訊量驟增且快速傳遞，不但擴充了人們接收資訊的廣度，也改變人們搜尋與瀏覽資訊的行為模式。然而，大量的資訊對人們在有限時間與吸收、處理資訊的能力而言，確實也帶來困惑與焦慮，因而衍生出「資訊超載」（information overload）的現象。一般而言，「資訊超載」是因為資訊量太大致使人們無法萃取出所需資訊（Nelson, 2001）。在高呼資訊經濟與知識經濟時代來臨的同時，Goldhaber（1997）更強調應在大量資訊中吸引人們的注意力，進而帶來商業價值的注意力經濟（attention economy）模式。

Davenport & Beck（2002）在『哈佛商業評論』中也提到在可取得大量資訊的狀態下，人們的注意力成為知識經濟中的稀有資源。早期學者針對人們對資訊的注意力區分為兩大類型（Laar, Heskes, & Gielen, 1997），包括由外部資訊提供者所驅動的「外生注意力」（exogenous attention），及來自個人主觀意圖的「內生注意力」（endogenous attention）。因此，企業除了推陳出新期望引起人們注意外，如何根據使用者的偏好提供符合使用者需

求的有價資訊，更是受到學界與業界的重視。然而，不同的領域、學科甚至不同的使用者本體有其特定的本體概念，且其重要的資訊大多以文字的方式儲存與散佈，因此，「本體論」（ontology）與「文字探勘」（text Mining）成為注意力經濟下，根據不同主體與特性從大量資訊中萃取出相對應的關鍵概念與架構，以吸引人們注意的重要理論與技術。

二、本體論（ontology）

「本體論」（ontology）是一個可共享概念化的正規表示形式（Borst, Akkermans, & Top, 1997），亦可稱為用以計算出正式詞彙之偏好意涵的邏輯理論（Guarino, 1998）。Alberts（1993）認為「本體論」是關於特定任務或領域的知識本體，用以提供語義解釋以描述概念的分類。「本體論」大多以物件（object）的型態描述真實世界中的實體，其所採用的描述方法與呈現的形式會依不同系統或領域而有所差異（Chuang & Yang, 2000），而不同領域對其本體論有不同的定義與論述。

在哲學領域中，「本體論」被視為真實世界中實體（entity）的分類，具有決定性的詳盡描述（Smith & Welty, 2001）。而在電腦科學領域，「本體論」的概念則是對於群體共享



之正式且明確表示的概念化形式，透過建立一個可運作的機制，使人們在處理大量資訊時，可以協助人們以規則化的操作步驟，找出所需要的重要知識（Gruber, 1993）。從知識應用的層面，Natalya和Deborah（2001）認為本體論是透過標準的定義達到知識分享與再利用的目的，且可將定義好的領域本體論（domain ontology）參考應用於其他領域。

根據「本體論」的普遍性程度，Guarino（1998）建構出「高層次本體論」（top-level ontology）、「領域本體論」（domain ontology）、「任務本體論」（task ontology）及「應用本體論」（application ontology）四種類型的本體論架構，如圖1所示。其中「高層次本體論」是用以描述一般性的概念，並無限定於特定問題或領域；「領域本體論」與「任務本體論」主要是描述特定領域或任務的詞彙與概念，會將上層「高層次本體論」的詞彙轉換為特定的專有名詞；「應用本體論」則是將

「領域本體論」與「任務本體論」所定義的概念應用在特定事件中描述其概念與關係。

在大量的資料或文字中透過本體分析產生有意義的詞彙，可使知識或概念以更清楚且明確的架構呈現。本體的建立並不容易，透過人工方式處理既耗時又費力，因此許多學者試圖結合資訊科學技術設計自動化本體建立機制。黃育宇（2007）針對INSPEC資料庫50萬篇計算機領域文獻建立計算機科學領域的知識本體，其所設計的系統可以讓試用者輸入關鍵字進行搜尋與瀏覽，搜尋結果更可將使用者導向各種關係資料的來源網站。此外，應用本體論可設計建置自動摘要系統，目的在於將文章拆成數個主題並取得摘要，所產生的本體樹狀圖即是代表各種不同的文章主題，最常應用於新聞事件的搜尋與瀏覽（吳家威、劉昭麟，2002；許孟淵，2006）。

事實上，本體的建立不僅限於領域學科的知識或大量的新聞事件，亦可針對個人的偏好與興趣建立出個人本體。吳政勳（2007）採用正規概念分析（formal concept analysis, FCA）將本體論應用在文件推薦系統，並配合使用者對文件的評分機制，提升推薦的準確度。另外應用本體論建置的推薦系統亦可應用於挖掘顧客的對產品的潛在偏好，透過建立群組偏好樹（group preference tree, GPT）推論顧客對產品的喜好強度，再根據顧客屬性分群並進行協同推薦（劉雅茹，2007）。然而在許多研究針對不同應用範疇建立本體論的同時，但卻鮮少研究著重本體論的學習機制。

Maedche & Staab（2001）針對本體論的學習提出一個四階段架構，如圖2所示。第一階段為「引進與再利用」（import and reuse），主

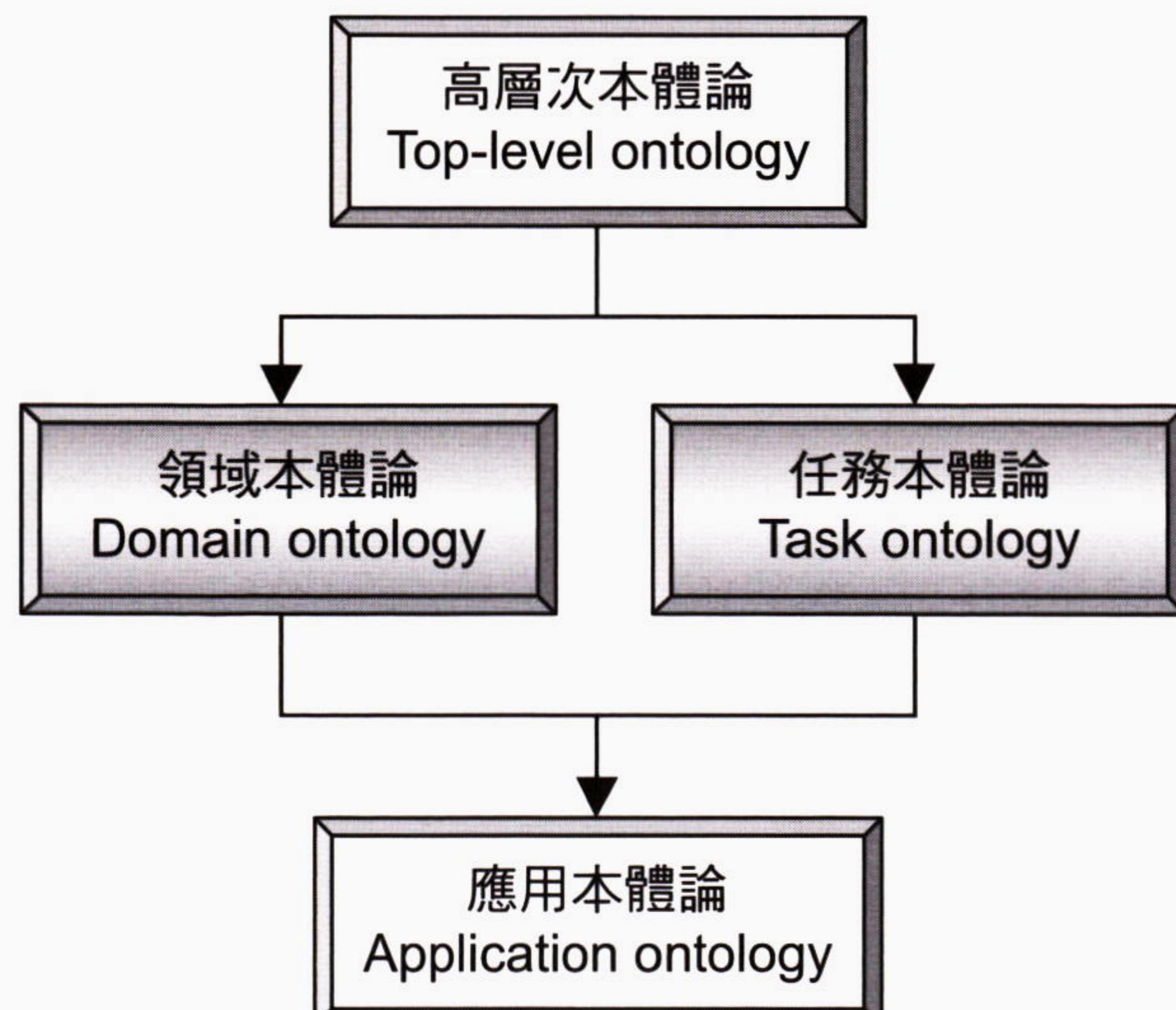


圖1 本體論類型架構

資料來源：Guarino, 1998。



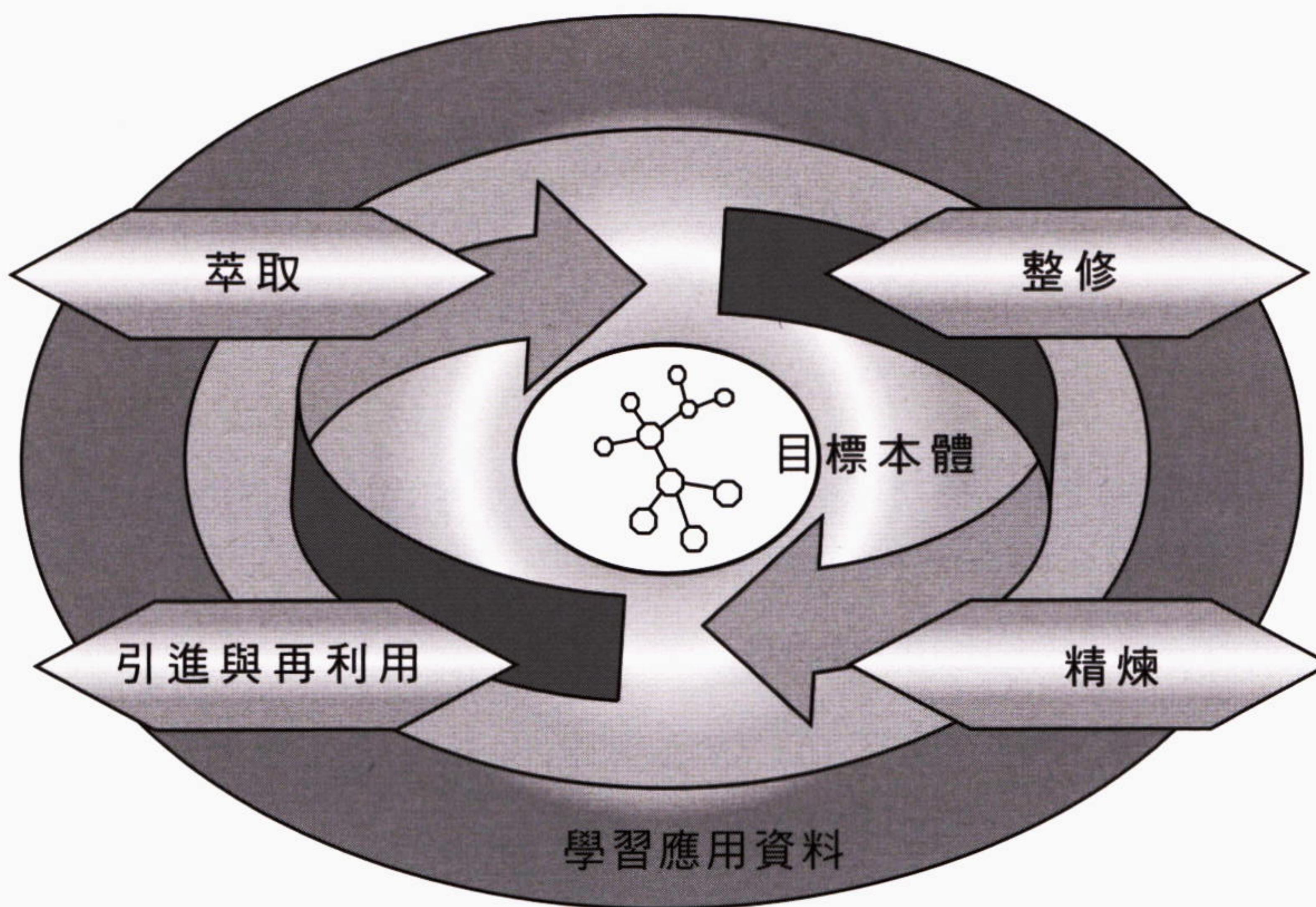


圖2 本體論學習架構

資料來源：Maedche & Staab, 2001。

要是針對已經存在的本體論進行結構合併或定義映對（mapping）的規則，例如定義出本體的分類規則；第二階段是「萃取」（extract），利用Web的資源支援本體論的學習，進而建構出目標本體論的初步架構，例如根據本體的分類規則進行分類；第三階段是「修整」（prune），針對前兩階段結果進行目標本體論的修正，例如針對分類結果之確認分類架構之合適性並進行修正；第四階段則是「精練」（refine），使修正過的目標本體論更精細與目標主體更適配，例如針對修正過的分類架構建構更準確的分類結果。

實際上，有效運用資訊科學技術可以建立本體論的學習機制，例如：資料探勘（data mining）、機器學習（machine learning）、文字探勘（text mining）等。由於網路應用的普及化，許多知識與重要內容大多以文字的形式透過電子檔進行傳遞與散佈，因此，如能有效運用文字探勘技術建立特定事件、特定議題的本

體論，對於知識閱聽者而言可更有效掌握有價資訊。

三、文字探勘（text mining）

「文字探勘」（text mining）是資料探勘（data mining）、知識發現（knowledge discovery）的延伸應用，其以文件內容為分析主體，目的在於從非結構的文件中精鍊出有意義、有價值的範型或知識（Simoudis, 1996）。由於大部分的資訊多以文字的方式進行儲存，且過去研究指出有80-98%的重要公司資訊儲存在文件中，例如：E-mail、Office文件或PDF文件等（Cheung, Lee, & Wang, 2005），與資料探勘相較之下文字探勘更具潛在的商業價值（Tan, 1999）。在Tan（1999）所提出的文字探勘運作架構中主要包含「文本萃取」（text refining）及「知識淨化」（knowledge distillation）兩個重要步驟，其中「文本萃取」主要將各種形式的文件轉換成中介主體；而



「知識淨化」則是從以中介形式呈現的文字內容中歸納出範型或知識。初始文件所轉換的中介主體可以文件中介形式（document-based）或概念中介形式（concept-based）呈現，文件中介形式主要以一份文件表示一個分析實體，可以在文件中歸納出範型或彼此間的關係，例如集群分析、分類分析或視覺化呈現等；而概念中介形式則是以一個物件或特定領域的概念表示一個分析實體，分析結果包括預測性模型、整合性發現及視覺化呈現等，亦可以根據物件或特定領域概念對文字中介形式進行資訊萃取，轉換成概念中介形式，如圖3所示。

四、文字探勘之應用研究

隨著資訊超載的現象日益嚴重，文字探勘技術也愈來愈受到重視。許多學者針對文字探勘的演算法與計算模型進行研究，以提升分析結果的精確性；也有學者結合本體論的概念針對文字探勘技術衍伸出不同領域的應用範疇。以下從管理層面針對近幾年相關應用研究歸納出顧客關係管理、主題分析管理、網站內容管理及趨勢預測管理等四種類型並加以說明。

4.1 顧客關係管理

許多受歡迎的網站每天會接收到數以百計甚至千計的顧客意見，對網站客服單位而言如何從大量的訊息中得到重要的訊息並提出立即性的回應是相當重要的，例如技術障礙問題或是顧客的申訴問題。Lo (2008) 以文字探勘及「支援向量機」（support vector machine）技術為基礎建立一個自動分類機制，可針對顧客意見進行過濾，以提升客服單位的服務效能與顧客滿意度。此外，Chang, Lin, & Wang (2009) 採用內容分析及決策樹分析方法針對顧客相關資料進行文字探勘分析，將顧客進行分類提供相對應的服務以提供更有效率的市場顧客關係管理策略。

4.2 主題分析管理

對學術研究者而言，檢視研究議題的相關文獻是不可或缺的重要工作，然而在大量的期刊文獻資料庫中要找到與議題相符的文獻需要耗費相當多的時間與精力。雖然透過電子期刊資料庫可以對搜尋標的下關鍵字，但是往往會得到成千上百筆的搜尋結果，若使用人工判讀

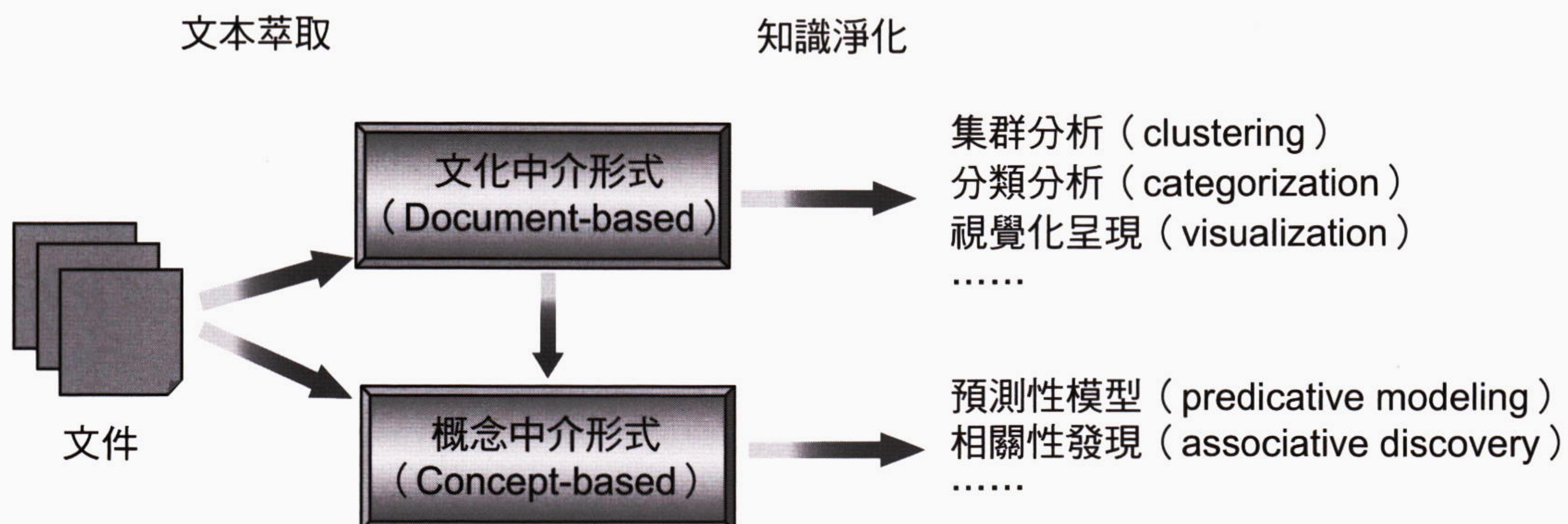


圖3 文字探勘運作架構

資料來源：Tan, 1999。

仍相當費時。Delen & Crossland (2008) 採用文字探勘技術針對標的期刊的論文摘要進行分析，以分析出各期刊在時間區間中最頻繁出現的關鍵字，並可比較同屬性標的期刊在時間區段中重視議題的差異性，對使用者而言可以快速了解其資訊標的為何。此外，Pons-Porrata, Berlanga-Llavori, & Ruiz-Shulcloper (2007) 提出一個新的遞增階層集群演算法（incremental hierarchical clustering algorithm），可以針對大量新聞事件依照主題性進行分群，可有效縮短使用者進行主題式搜尋的時間。國內也有學者應用文字探勘技術建構特定主題之知識地圖，可以充分建構領域知識的概念性架構（陳文華、施人英、吳壽山，2004；Yeh, 2007）。Hung, Chi, & Chen (2009) 更採用自組織映射圖技術分析使用者搜尋行為，將使用者經常使用的關鍵字儲存於使用者電腦中的興趣庫（interest base）中，且該興趣庫會動態更新。透過文字探勘技術以非侵入的方式可以建構出使用者興趣地圖（user interest map），當使用者透過搜尋引擎（例如：Google）下關鍵字尋找資訊標的時，會針對使用者的興趣地圖對搜尋引擎所列出的搜尋結果重新排序，使搜尋結果更符合使用者的需求。

4.3 網站內容管理

WWW是資訊檢索與知識發現的主要來源，知識以各種形式在網際網路中散佈與流傳，為了便利使用者搜尋資訊標的，衍伸出許多資訊檢索機制，其中建構「事先定義的網站字典」（predefined web directories）以貼近使用者搜尋標的是廣泛採用的資訊檢索機制之一。Yang & Lee (2004) 採用自組織映射圖技

術分析網站內容，自動建構出階層式的網站字典，提供使用者根據分析出的網站關鍵字連結到資訊標的。此外，亦有學者從社群經營與傳播效應的觀點，對部落格文章內容進行分析，探討部落格之網路媒體特性，研究結果發現部落格成功關鍵因素在於文章內容所蘊含的意識與深度（施百俊、施如齡，2006）。此外，最近更興起「行為定向廣告」（behavior target AD）相關應用，透過收集使用者曾經查詢過的關鍵字，從歷史紀錄分析使用者的偏好，可提供客製化廣告內容，以求更貼近使用者對產品的需求。

4.4 趨勢預測管理

股票投資分析最常採用基本面與技術面分析，其內涵主要在於評估公司的營運與財務狀況以預測未來營收，並參考歷史股價的波動預測股價可能的漲跌趨勢。然而，除了基本面與技術面之外，消息面對短期股價的衝擊影響也是不容忽視的。鍾任明、李維平與吳澤民（2007）整合歷史股價交易資料並針對中文財經新聞內容進行文字探勘，建構出台股個股日內股價漲跌的預測模型，該模型達81.48%的預測正確率，對進行短期股票投資者而言具有相當參考價值。

應用在科技發展趨勢預測方面，Santo, Coelho, Santos, & Filho (2006) 針對國際級科技論文、專利文獻進行文字探勘分析，萃取出奈米重點科技關鍵字並做統計分析，並進行國際奈米科技發展趨勢與重點議題之比較，其研究成果可作為提供政府部分對奈米科技投入研發經費之參考，以提升巴西在相關領域的國際競爭力。

五、本體論與文字探勘之應用價值

本體論所描述的本體可以小至個人、特定事件，或大至領域、學科甚至國家，其所能應用的範疇相當廣泛，為能有效從本體中發掘有意義的規則、議題重點或主題趨勢分析等，可結合文字探勘技術進行深度剖析。而學者對於文字探勘技術更是日益精進，除了不斷改進分析技術的精確度與穩定性外，更積極發展視覺化工具，可以更人性化的方式呈現本體概念。未來學者可積極發展各種不同層面的應用，例如應用在政府政策方面，可針對各部會新聞、政策白皮書及計畫等內容進行文字探勘分析，以建構政策地圖洞悉政策發展脈絡。此外，亦可針對各國科技相關領域發展之相關新聞進行文字探勘分析，並進行國際比較以了解各國科技發展動向。奠基於可輔助管理者從龐大的資訊量中快速了解本體意涵並有效擬定決策，足見本體論與文字探勘之應用價值是不容小覷。

參考文獻

- 黃育宇（2007）。以電腦科學文獻自動建立本體論之架構。未出版之碩士論文，國立雲林科技大學資訊工程研究所，雲林縣。
- 吳家威、劉昭麟（2002）。應用本體論設計與建置摘要系統。*2002民生電子研討會論文集*（頁41-46）。新竹：2002民生電子研討會。
- 許孟淵（2006）。以本體論為基礎之新聞事件檢索與瀏覽。未出版之碩士論文。國立雲林科技大學資訊管理研究所，雲林縣。
- 吳政勳（2007）。以本體論文基礎之文件推薦系統開發與實證。未出版之碩士論文。國立屏東科技大學資訊管理研究所，屏東縣。
- 劉雅茹（2007）。利用本體論建構群組偏好樹

於產品推薦之研究。未出版之碩士論文。

朝陽科技大學資訊管理研究所，台中縣。

陳文華、施人英、吳壽山（2004）。探討文字採掘技術在管理者知識地圖之應用。*中山管理評論*, 12 (6), 35-64。

施百俊、施如齡（2006）。以文字探勘技術探究部落格之網路媒體特性。*淡江人文社會學刊*, 28, 95-122。

鍾任明、李維平、吳澤民（2007）。運用文字探勘於日內股價漲跌趨勢預測之研究。*中華管理評論國際學報*, 10 (1), 1-30。

Alberts, L. K. (1993). *YMIR: An ontology for engineering design*. Unpublished doctoral dissertation, University of Twente, Netherlands.

Borst, P., Akkermans, H., & Top, J. (1997). Engineering ontologies. *International Journal of Human-Computer Studies*, 46(2-3), 365-406.

Chang, C.-W., Lin, C.-T., & Wang, L.-Q. (2009). Mining the text information to optimizing the customer relationship management. *Expert Systems with Application*, 36, 1433-1443.

Cheung, C. F., Lee, W. B., & Wang, Y. (2005). A multi-facet taxonomy system with applications in unstructured knowledge management. *Journal of Knowledge Management*, 9(6), 76-91.

Chuang, W. T., & Yang, J. (2000). Extracting sentence segment for text summarization: A machine learning approach. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development*

- in information retrieval* (pp. 152-159). NY: ACM.
- Davenport, T. H., & Beck, J. C. (2002). *The attention economy: Understanding the new currency of business*. MA: Harvard Business School Press.
- Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Application*, 34, 1707-1720.
- Goldhaber, M. H. (1997). The attention economy and the net. *First Monday*, 2(4). Retrieved April 20, 2009, from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/519/440>
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5, 199-220.
- Guarino, N. (1998, June). Formal ontology and information system. In N. Guarino (Chair), *INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS*, Trento, Italy.
- Hung, C., Chi, Y. L., & Chen, T. Y. (2009). An attentive self-organizing neural model for text mining. *Expert Systems with Application*, 36, 7064-7071.
- Laar, P. V. D., Heskes, L., & Gielen, S. (1997). Task-dependent learning of attention. *Neural Networks*, 10(6), 981-992.
- Lo, S. (2008). Web service quality control based on text mining using support vector machine. *Expert Systems with Application*, 34, 603-610.
- Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent Systems*, Mar./Apr., 72-79.
- Natalya, F. N., & Deborah, M. (2001). *Ontology development 101: A guide to creating your first ontology*. Stanford Knowledge Systems Laboratory Technical Report, KSL-01-01, and Stanford Medical Informatics Technical Report, SMI-2001-0880.
- Nelson, M. R. (2001). *We have the information you want, but getting it will cost you: Being held hostage by information overload*. Retrieved April 20, 2009, from <http://www.acm.org/crossroads/xrds1-1/mnelson.html>
- Pons-Porrata, A., Berlanga-Llavori, R., & Ruiz-Shulcloper, J. (2007). Topic discovery based on text mining techniques. *Information Processing & Management*, 43(3), 752-768.
- Santo, M. M., Coelho, G. M., Santos, D. M., & Filho, L. F. (2006). Text mining as a valuable tool in foresight exercises: A study on nanotechnology. *Technological Forecasting and Social Change*, 73, 1013-1027.
- Simoudis, E. (1996). Reality check for data mining. *IEEE Expert*, 11(5), 26-33.
- Smith, B., & Welty, C. (2001). Ontology: Towards a new synthesis. In *Proceedings of the International Conference on formal ontology in Information system* (pp. 3-9). Ogunquit, Maine: ACM.
- Tan, A.-H. (1999). Text mining: The state of the art and the challenges. In *Proceedings of PAKDD Workshop on Knowledge discovery from Advanced Databases* (pp. 71-76). Beijing:

Springer.

Yang, H.-C., & Lee, C.-H. (2004). A text mining approach on automatic generation of web directories and hierarchies. *Expert Systems with Applications*, 27, 645-663.

Yeh, Y. J. (2007). Apply growing hierarchical self-organizing map in constructing topic maps. *ICL Technical Journal*, 119, 83-90.

