

結合文字探勘與本體論檢索雲端運算專利之研究

黃嘉彥* 黃青輝

國立勤益科技大學 資訊管理研究所 (41170 台中市太平區中山路二段 57 號)

jygiant@ncut.edu.tw

摘要

雲端運算是當今網絡技術的熱門話題，也是許多企業選擇此領域作為未來技術發展的方向。目前有關雲端運算產業之專利分析目前僅僅散見在一些雜誌刊物，鮮有期刊論文之發表。而在相關專利分析的介紹僅出現在研討會、雜誌或是網路上，未見有期刊發表之研究。為正確進行專利分析，需先能檢索出正確的專利文件。有鑑於此，本研究提出以文字探勘為基，運用滾雪球法配合分區組合檢索與本體論應用的專利檢索策略，可有效率且正確地檢索出雲端運算三類商業模式所屬的主要專利類別。此研究之分析結果，將可提供國內相關產業進一步進行雲端運算專利分析及專利佈局之參考。

關鍵字：雲端運算、本體論、專利檢索、文字探勘

壹、緒論

近年來，許多資通產業視雲端運算為未來 10 年的產業發展趨勢，並爭先推出相關的創新應用產品與服務模式。諸如 Microsoft、Google、Amazon、IBM、Yahoo、HP 等國際大廠均已陸續加入雲端運算的戰場，期能搶食大餅(Wang et al. [20])。我國政府已將雲端運算產業作為目前積極推動之四大智慧型新興產業之一，經濟部於 101 年 11 月核定「雲端運算應用與產業發展方案」，規劃推動 15 項雲端運算計畫。由工研院、資策會、電機電子公會、資訊軟體協會及中華電信等機構結合國內 50 多家廠商已共同組成雲端運算產業聯盟。大型企業如鴻海、廣達及趨勢科技等公司亦皆大舉佈局雲端運算不同技術領域。

專利文件內容包含大量重要技術相關的資訊，可用於開發潛在新技術、預測競爭對手動向及研發趨勢。專利為企業技術的代表，內含企業重要資訊，為企業研發之基礎。專利為企業重要的資產，可藉以保護公司的研發成果，亦可積極做為市場上競爭的工具。因雲端運算的發展正處於快速發展與渾沌的階段，導致進行專利檢索時大幅增加其困難度。目前仍少見有關於雲端運算產業專利分析的文獻。欲對雲端運算產業之專利進行正確之分析、分類或分群，需先對其所屬專利完整地檢索。

雲端運算的發展主要是承襲「網格運算」(Grid Computing)和「分散式運算」(Distributed Computing)，而在國際專利分類(IPC, International Patent Classification)定義下，關於網路技術及資料儲存等相關技術屬 G06 類與 H04 類(其中，G06 類係包含電子數位資料處理、數據處理系統或方法，以及資訊儲存記憶體之技術，H04 類係包含電信通信之技術)。因此，在 Xue 等人[21]之研究中，係將專利檢索的範圍限制在 IPC 國際專利分類 G06 類與 H04 類。然而，如此鎖定 G06 類與 H04 類進行檢索專利的作法會有搜尋專利過於廣泛的問題，若干不屬於雲端運算領域的專利也被檢索收集。顯然，依據這些數據進行分析的結果是令人存疑的。

一般而言，該專利文件需深度閱讀其內容，才可明確判斷出該專利之類別，故需耗費大量人力及時間成本。近年來，文字探勘的技術在資訊領域之應用頗有進展，陸續有若干應用此技術於專利分析之研究。張瑞芬等學者[3]主張以先進知識管理技術進行專利與智財分析的重要性，其相關的主題與方法包括：專利自動摘要方法、專利自動分類方法、專利自動分群方法、專利品質評估方法及專利侵害鑑定與分析等。該文章指出，以軟體去發掘並以視覺方法去呈現大量文件的連鎖關聯性，是一般人工逐篇解讀專利文件所難以達到的境界。顯然，以文字探勘技術為基礎去執行的專利分析，對於提高專利分析的效率有很大的幫助。有鑑於雲端運算未來發展前景，若能透過專利分析找出相關技術的發展狀況，將有助於企業分析未來趨勢方向及企業訂定發展略。然而，這些均有賴完整的雲端運算專利資料庫的建立。為此，本研究以文字探勘為基，結合雲端運算知識本體論(ontology)的檢索策略，以獲取完整而正確的雲端專利範圍。

貳、文獻探討

由於雲端運算產業涵蓋領域仍屬混沌，其在網格運算和分散式運算之間的定位未明，導致其專利檢索的困難。目前有關雲端運算產業之專利分析如此重要的議題僅僅散見在一些雜誌刊物，鮮有期刊論文之發表。雲端運算產業的發展方興未艾，運用客觀的資訊來評估企業間競爭對手，以掌握雲端領域技術的發展情形及技術的成熟度，並且進一步以專利技術來分析雲端運算產業的市場，是企業避免盲目投資與降低投資的風險的重要課題。

目前雲端運算商業模式發展，包含「基礎架構即服務」(IaaS, Infrastructure as a Service,)、「平台即服務」(PaaS, Platform as a Service)，以及「軟體即服務」(SaaS, Software as a Service)三部分(Sultan [16], Kulkarni et al., [10]-[12])。其各項模式主要服務如下：

- (一) 基礎架構即服務係直接提供硬體的環境及網路頻寬給企業用戶使用，功能是處理、存儲空間、網路和其它基礎性的運算資源，提供運行任意軟體，包括作業系統和應用程式。不需管理或控制底層的雲端基礎架構，但可配置存儲空間，作業系統和已部署之應用程式控制權限，及網路元件管理控制（如，防火牆等）。
- (二) 平台即服務提供了開發、運行、管理和監控的環境，優良的平台層設計可滿足雲端在擴充性、可用性和安全性等方面的要求。此層的消費者可透過平台供應商提供的程式開發工具來將自身應用建構於雲端架構之上，雖能掌控運作應用程式的環境，但並不掌控作業系統、硬體或運作的網絡基礎架構。
- (三) 軟體即服務提供給使用者網路的軟體應用，使用運行於雲端基礎架構上的應用程式。從不同的客戶端設備透過簡易型客戶端介面，如網頁瀏覽器，即可連接至伺服器進行資源或服務的取得，消費者不需管理或控制底層的雲端架構，如網路、伺服器、作業系統、儲存空間等操作。

雲端運算雖是當今網絡技術的熱門話題，但是相關專利分析之文章僅散見在各類的雜誌，而且其主題大多侷限在探討雲端運算技術發展現況、產業應用、對企業運作之影響、創新商業模式與資安問題等(Brandic and Buyya [7], CaoLei et al. [8])。

國內近年也開始有不少有關於雲端運算技術與管理方面之研究。林逸筑[2]將雲端運算運用於數位學習社群平台，以增進教師與學生間的互動，讓最新教學資訊能夠即時於平台上流通。李芳榮[1]以近年來許多企業競相投以關注焦點的雲端運算服務為例，從資訊技術治理觀點，探討企業在導入新的資訊科技或服務時，所面臨的環境不確定性以及複雜度，考慮組織規模、高階主管支持程度、資訊密度、外部環境競爭者與社會政策、資訊技術成熟度等因子之間的關連。陳伯璋[4]於雲端運算環境下探討界定台灣資訊服務業的產業價值鏈，以策略要素描述與驗證產業的經營策略，以及找出產業關鍵成功因素與策略要素的關係。劉一誠[5]以 SWOT 方法分析雲端運算對台灣 ICT 產業的機會、威脅，及台灣 ICT 產業在雲端運算發展上的優勢與劣勢。這些文獻主要著重在雲端之應用以及雲端對企業運作模式之影響的研究，但是目前尚未見到有關雲端運算領域專利分析方面的研究。

雖然雲端運算的發展如火如荼，但是國內外有關雲端運算方面的專利分析很少。而為進行雲端運算之專利分析，必先正確地檢索出相關之專利。Chiou [9]以文

獻計量分析及生命週期演變的研究觀點，指出雲端技術自 1996 起至 2007 年中為萌芽期，自 2007 年開始成長階段，到了 2012 年為技術生命週期成長階段之轉折點，技術逐漸進入成熟期。預計 2016 年以後，將到達飽和期，其專利件數增加速度逐漸減緩，專利飽和數估計為 27745 件。由此觀之，雲端運算相關技術的研發目前仍處於一個成長的階段。Xue 等人[21]針對發展雲端運算最具知名度之六家目標企業（包括 Google、IBM、Microsoft、Apple、Amazon 及 Intel 公司），研究其雲端運算技術之主流技術領域與專利佈局現況。他們認為雲端運算即網路運算，故將雲端運算專利檢索的範圍限制在 G06 與 H04 兩類國際專利分類。

大體而言，雲端運算的專利檢索的範圍確實是可限制在 IPC 國際專利分類 G06 類與 H04 類，此乃因為雲端運算即網路運算，而網路運算技術在國際專利分類上屬於有關網路技術及資料儲存等的 G06 類與 H04 類。但是，若以僅僅以 G06 與 H04 當作雲端運算專利搜尋的唯一依據，顯然其搜尋範圍太大。換言之，許多 G06 與 H04 的專利並非屬於雲端運算專利。

由於在專利申請核准過程，專利審查權責單位的各領域專家依據申請專利內容決定其歸屬的分類項目。但是，雲端運算產業範圍廣大而渾沌，故不易直接以 IPC 進行雲端運算相關專利的搜尋。近年來，以自動化方式協助專利工程師與決策者進行專利分析之需求日增，Tseng 等人[19]發展以文字探勘進行奈米碳管的專利地圖分析的技術。其主要專利分析的步驟包括：文字區隔、結論萃取、特徵選擇、語幹關聯、集群產生、主題辨識與專利地圖等。Taghaboni-Dutta 等人[18]以文字探勘萃取 RFID 專利之關鍵詞，以研究 Intermec Technologies 與 RFID Patent Pool 之專利策略。有鑑於以文字探勘為基的專利檢索與專利分析漸受重視，本研究提出運用滾雪球法配合分區(block)組合檢索的方式，使用 Patent Guider 初步收集與雲端運算相關之專利，再利用本體論關鍵字篩選出屬於雲端運算的專利。

參、研究方法

一般在進行專利檢索時，往往需藉由反覆的修改檢索策略來取得適當的檢索範圍。專利文件之檢索是一件困難與耗時的工作，其檢索範圍正確與否，攸關後續專利分析的正確性。本研究所提的專利檢索策略如圖 1 所示，茲說明如下。

首先，在專利檢索第一階段所採用滾雪球法的作法，係指事先掌握一篇或數篇相關專利，利用這些相關專利尋找更多相關的專利。為此，本研究採用連穎科技公司的 Patent Guider 軟體，在 Title、Abstract 與 Claim 欄位中以 "Cloud Computing" 一詞進行檢索。將初步檢索後得到專利，以人工方式閱讀專利摘要欄位，以篩選出確定屬於雲端運算領域的專利。此步驟之目的在於找出所謂的「珍珠」，以便續依此搜尋出更多屬於雲端運算領域的專利。

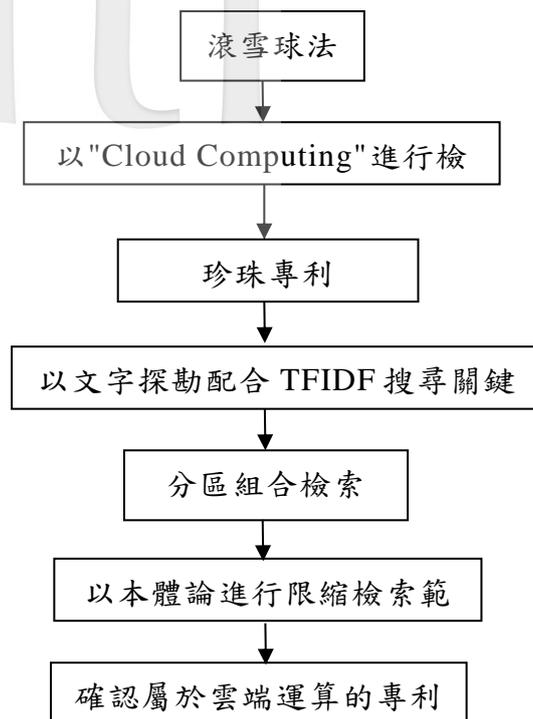


圖 1 雲端運算專利檢索流程

本研究續以文字探勘技術針對珍珠專利進行分析，以搜尋出雲端運算領域主要關鍵字。為能更精確掌握雲端運算之專利，本文根據上述三類商業模式之珍珠專利檢索結果依據 TF-IDF (詞頻-反轉文件頻率) 的概念，找出權重值較高之關鍵詞。Salton and Buckley[15] 提出的 TF-IDF 意謂若是一個詞項出現在該文件中的頻率越高而且反轉換文件頻率越低，則該詞項就越具有代表性，重要程度也相對高。計算公式如下式所示：

$$w_{ij} = tf_{ij} \times \log(N / df_i) \quad (1)$$

其中， w_{ij} 代表在專利文件 j 中關鍵詞 i 的權重， tf_{ij} 關鍵詞 i 在文件 j 中出現的頻率， N 是專利文件總數， df_i 是包含關鍵詞 i 的文件總數。本研究將三類服務模式之 TF-IDF 權重值較高之前五十筆關鍵詞進行比對，將有交集之關鍵詞予以刪除，據此即獲致分屬於三類商業模式的關鍵詞。

在專利檢索第二階段主要是進行分區組合檢索。所謂分區組合檢索策略是將一個檢索問題分為數個主題層面 (facets)，確定這些主題層面間的關係 (通常多為 "and" 的關係)，然後再將這些主題層面以布林運算元作結合。再找出代表每個主題層面之所有詞彙，並將該主題層面下的所有詞彙以布林運算元 "OR" 作聯集，以求完整性。

本研究在第一區檢索係以不限定欄位方式，使用 "Cloud Computing" 關鍵字搜尋。其次，第二區檢索為運用雲端運算各模式經篩選後之關鍵詞為檢索依據。第一

區與第二區再以”and”連結後進行檢索。以 PaaS 檢索為例，我們在第一區檢索係以”Cloud Computing”關鍵字為搜尋之設定條件。再於第二區以前述所得之關鍵詞為搜尋之設定條件。然後將第一區與第二區以”and”連結後進行檢索，以找出屬於 PaaS 服務模式之專利。同樣地，我們可獲得 IaaS 服務模式專利與 SaaS 服務模式專利。

依分區組合檢索擴大檢索範圍之結果，很可能還是會有檢索範圍過大的問題，因此本研究以本體論進一步限縮。換言之，初步檢索出來的專利，必須再經文字探勘解構，將關鍵字與本體論比對才可確知所搜尋出來的專利是否與雲端運算技術有關。文字探勘是將非結構化的文字經萃取與轉換成結構化的資料，使分散的文字組合成可用的資訊與規則。常用的技術有文字分群、概念萃取、文件摘要與關聯規則等。其優點是可從文件中抽取出特定用語間的關聯性，而這些用語不需事先定義，因此可應用在各個不同的領域，大幅降低專家參與的必要性。

Androcec 等人[6]透過分析24篇重要的文獻，以系統化回顧的方法，建立一個全方位角度的雲端運算的知識本體論及其應用。它們將雲端運算在科學文獻中主要聚焦與應用之範圍分為四類：雲端資源與服務(將現有服務及定價模式分類或定義新型態的雲端服務)、雲端安全性、雲端讓軟硬體在多種品牌機器上能有意義的溝通、以及如何發現雲端與選用雲端服務。Reshma與 Belasis[14]認為雲端知識本體論係提供一個介於人與電腦代理間溝通領域的支援分享平台，其聚焦在源於雲端軟體環境發展的應用，以及使用雲端技術與來其他的雲端服務組成一項新的服務。雲端知識本體論涵蓋一組概念與關係，可用於使用者查詢的資訊檢索之用。該研究使用雲端知識本體論技術，企圖使雲端服務發掘系統變得更有效率。當使用者提出詢問時(給定名稱與需求)，系統會到雲端知識本體論去比較其相似度。Han 與 Sim [13] 提出一個雲端服務發現系統 (CSDS)，該系統可配合雲端本體論來決定各式雲端服務之間的相似性。該文是目前唯一可查到繪出雲端運算本體論的文獻，因此，本研究採用該文所提之雲端運算知識本體論研究為比較基準。

肆、研究結果與討論

一、數據分析

本節根據專利檢索之結果，進行關鍵詞擷取、篩選與文件向量化作業。本研究使用 Patent Guider 檢索資料美國專利商標局(USPTO)資料庫。首先，以”Cloud Computing”作為檢索的關鍵字，其目的是儘量找出確實屬於雲端運算的專利。因此本研究限定在專利名稱、摘要及申請專利範圍三個欄位內進行檢索，其指令為：(TTL/”Cloud Computing” OR ABST/”Cloud Computing” OR ACLM/”Cloud Computing”)，共得到 104 筆專利。

為確保珍珠專利確實屬於雲端的專利，本文將此 104 筆專利進行人工閱讀，經篩選非雲端運算專利文件後，剩下 62 筆珍珠專利。續將 62 筆珍珠以人工閱讀方式，依

各專利所屬商業服務模式特性進行分類，其中，IaaS 共得到 9 筆、PaaS 共得到 21 筆以及 SaaS 共得到 45 筆。依此，本研究將各類專利文件進行關鍵詞擷取，並依據 Salton and Buckley[15]所提 TF-IDF 的概念，找出權重值較高之關鍵詞。為能更精確掌握雲端運算各服務模式之專利，將三類服務模式之 TF-IDF 權重值較高之前五十筆關鍵詞進行比對，將有交集之關鍵詞予以刪除。篩選後我們可獲得屬於各服務模式的關鍵詞，其中，IaaS 剩下 7 項關鍵詞；PaaS 剩下 5 項關鍵詞以及 SaaS 剩下 10 項關鍵詞。

其次，本研究發展分區組合檢索策略，將一個檢索問題分為數個主題層面，再確定這些主題層面的關係，彼此之間以布林邏輯“AND”運算元作連結，找出可能代表該主題層面的所有相關詞。首先，在第一區檢索係以不限定欄位方式，使用“Cloud Computing”關鍵字搜尋。其次，第二區檢索為運用各模式經篩選後之關鍵詞。第一區與第二區係以“and”連結後進行檢索。以 PaaS 檢索句為例：我們在第一區檢索係以“Cloud Computing”關鍵字為搜尋之設定條件。再於第二區以 ("application template" or nodegroup or "operational nodes" or query or "private text")關鍵詞為搜尋之設定條件。再將第一區與第二區以“and”連結後進行檢索，以找出屬於 PaaS 服務模式之專利。經過搜尋結果，共取得 IaaS 服務模式專利計 167 筆、PaaS 服務模式專利計 393 筆以及 SaaS 服務模式專利計 650 筆，將其結果輸出至 Excel 以便後續文字探勘分析。

依上述分區組合檢索擴大檢索範圍之結果，必須再以本體論進一步限縮。本研究採用 Han 與 Sim [13]所提之雲端運算知識本體論關鍵字為篩選基準。經比對本研究萃取之關鍵字與本體論之關鍵字，本研究可確認三類商業模式所屬的主要關鍵字，如表 1 所示。

經此本體論篩選專利文件的步驟，本研究共刪 345 專利，所剩餘之專利皆為雲端運算技術有關的專利，其中，IaaS 服務模式專利共計 261 筆、PaaS 服務模式專利 103 筆以及 SaaS 服務模式專利 109 筆。

Xue 等人在 2011 年之研究提出雲端的 IPC 國際專利分類是 G06 與 H04。但是此分類顯然過於廣泛，因此並無法提供專利分析者足夠的分析線索。本研究根據前述所檢索之專利，可進一步探勘出屬於雲端運算更細分類的 IPC 國際專利分類。限於文章空間，本文就僅將數量較多的專利類別列出。三類商業模式之 IPC 分類分別如圖 2~圖 4 所示。其中，數量較多的包括 G06F 15/173、G06F 15/16 與 H04L 29/06 所對應之商業活動將於下一小節中說明。

表 1 三類商業模式所屬的主要關鍵字

IaaS	PaaS	SaaS
database	worker node	proxy
application servers	http	access control
software applications	backend	performance metrics
virtual machine	interface	security
bandwidth	web application	database
memory	scalable	dns
device	security	software
security	hosting	graphical user interface
networks	language	networks
infrastructure	platform	communication

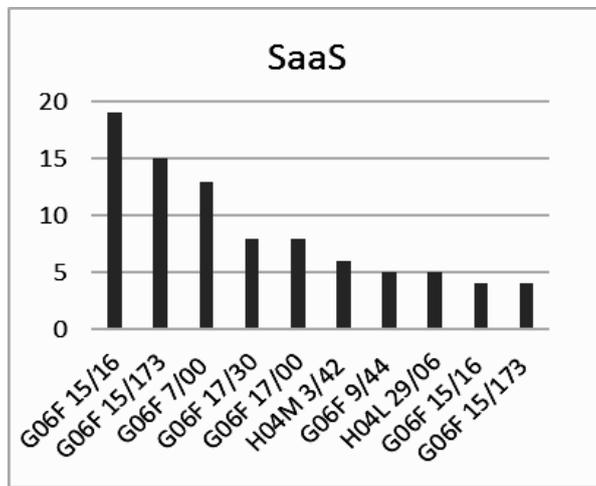


圖 2 SaaS 商業模式所屬的 IPC 分類與其統計數量

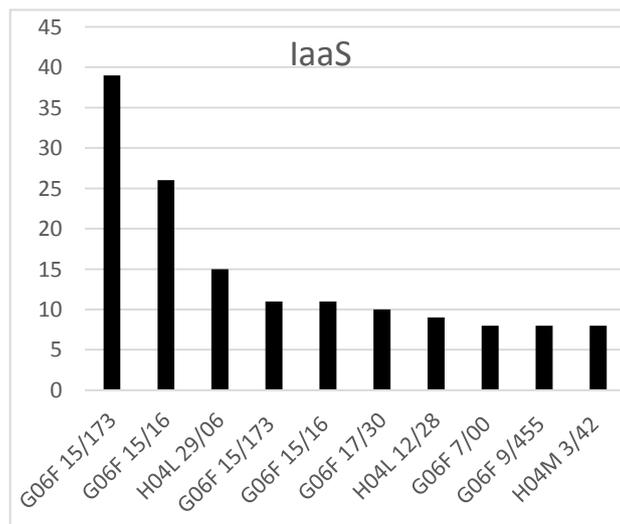


圖 3 IaaS 商業模式所屬的 IPC 分類與其統計數量

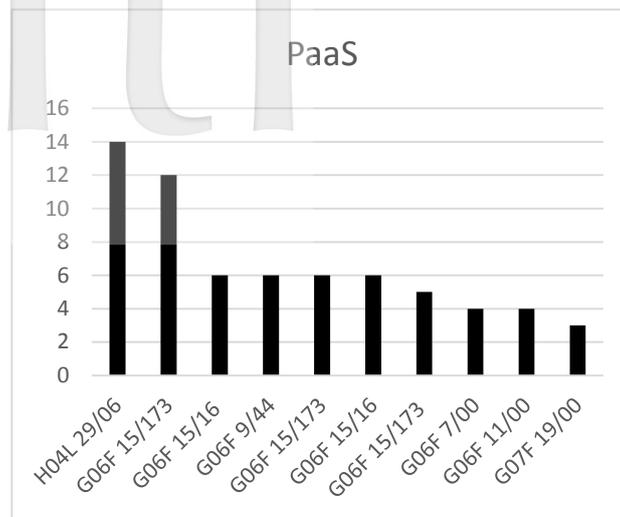


圖 4 PaaS 商業模式所屬的 IPC 分類與其統計數量

二、討論

圖 2~圖 4 中，G06F15/173 分類佔最大部分，對照到 IPC 分類之定義，此分類主要技術內容是：“SECTION G PHYSICS > COMPUTING; CALCULATING; COUNTING > ELECTRIC DIGITAL DATA PROCESSING > Digital computers in general; Data processing equipment in general > Combinations of two or more digital computers each having at least an arithmetic unit, a programme unit and a register, e.g. for a simultaneous processing of several programmes > Interprocessor communication > using an interconnection network, e.g. matrix, shuffle, pyramid, star or snowflake”。其次的 G06F15/16 分類主要技術內容是：“SECTION G PHYSICS > COMPUTING; CALCULATING; COUNTING > ELECTRIC DIGITAL DATA PROCESSING > Digital computers in general; Data processing equipment in general > Combinations of two or more digital computers each having at least an arithmetic unit, a programme unit and a register, e.g. for a simultaneous processing of several programmes”。數量佔第三的 H04L 29/06 分類主要技術內容是：“SECTION H ELECTRICITY > ELECTRIC COMMUNICATION TECHNIQUE > TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION > Arrangements, apparatus, circuits or systems, not covered by a single one of groups > Communication control; Communication processing > characterised by a protocol”。

Synergytek [17] 透過 Relecura 專利分析平台進行統計在 2013 年期間核准約有 30 多萬件專利。由其所分析之前 20 大專利權人的專利的主要技術分類，可發現此 20 大專利技術分類主要是包含在 G06 與 H04。而本研究所得之 G06F15/173、G06F15/16 與 H04L 29/06 均隸屬其中。這些均是與目前當紅的海量資料、新一代

無線網路、雲端運算以及智慧裝置的指紋及觸控介面設計議題相關的技術。由此亦可看出本研究檢索資料之正確性。

本研究並不直接採用本體論之關鍵字到 Patent Guider 去搜尋雲端專利之原因有二：首先是因為在實務上 Patent Guider 軟體並無法一次輸入太多關鍵字。其次，我們先採取與 Han and Sim [13]本體論不同的檢索方向，最後再將兩者加以比較，取其交集的集合為最終的雲端專利範圍。此種作法反而能更客觀地決定適合的雲端運算專利範圍。

珍珠專利因為篇數有限，以人工閱讀即可確定其商業模式歸屬。但是當以分區組合檢索擴展時，難免會有納入其他商業模式的情況，因此藉由與本體論的比對，可以刪除不屬於雲端的專利外，還可確實地將專利進行分類。因此，雖然珍珠到分區組合檢索的程序都已經將所檢索之專利分類，而最終還要再以本體論關鍵字分類，這個過程不但並不矛盾，反而更凸顯出本研究之重要貢獻。

本研究所擷取關鍵字與本體論關鍵字之交集的關鍵字具有非常重要之用途，包括：(1).用以判定所檢索之專利是否屬於雲端運算專利。本研究將分區組合檢索所得之專利進行文字探勘所擷取的關鍵字與本體論之關鍵字比對，若是專利文件中不包含本體論中的關鍵字，則表示該文件並不屬於雲端運算之專利，因此加以刪除。另外，由該專利包含之關鍵字可判定該專利最終歸屬哪一種商業模式。(2).未來運用這些具有分類區別度的關鍵字，可用於開發關聯分類器，以建立雲端運算產業專利文件自動分類系統。

伍、結論

雲端運算產業的發展方興未艾，加上雲端運算在網格運算和分散式運算之間存有重疊涵蓋領域之因素，導致其專利檢索的困難。傳統上使用專利分類號進行專利分析有其限制，近年來，以文字探勘進行專利分析的研究方式已成主流。本研究在學術上的貢獻在於，提出以文字探勘為基，配合滾雪球法、分區組合檢索與本體論應用的專利檢索策略，以有效率檢索出屬於雲端運算三類商業模式的專利文件，以利於工作者方便進一步地進行其專利的分析。

先前有關雲端運算專利檢索的文獻僅鎖定G06類與H04類為目標，因此會收集到許多不屬於雲端運算領域的專利。而本研究藉由所提之檢索策略，可收集到正確的專利，故可揭櫫真正雲端運算所屬的主要專利類別。由數據可發現，三類商業模式有相重疊的專利類別(G06F015)，也有各自不同的類別，此項結果與一般認知雲端三類商業模式有部分技術重疊的現實狀況是一致的。因此，本研究在實務上排除了先前文獻之限制與錯誤。本研究所建立雲端運算專利資料庫，是未來進一步進行專利分析與建立專利文件自動分類系統之重要基礎。也便利於往後開發關聯分類器，建立雲端運算產業專利文件自動分類系統，國內外進行雲端運算產業相關分析

需求的研究人員，可透過本研究之分析之結果，充分掌握專利資訊，進行市場與技術趨勢之分析。

參考文獻

1. 李芳榮，(2009)，「資訊技術治理對企業雲端運算服務導入意願的影響及其干擾項」，銘傳大學資管系碩士在職專班碩士論文。
2. 林逸筑，(2010)，「結合雲端運算於開發線上數位教材編輯系統之研究」，世新大學資訊傳播學研究所碩士論文。
3. 張瑞芬、張力元、吳俊逸、樊晉源，(2012)，專利分析與智慧財產管理 - 以資訊技術與知識管理方法為手段，華泰文化出版，台北。
4. 陳伯璋，(2009)，「在雲端運算環境下資訊服務業經營策略之研究-以系統整合業為個案」，銘傳大學經濟學系碩士在職專班碩士論文。
5. 劉一誠，(2009)，「雲端運算趨勢對台灣 ICT 產業發展之契機探討」，交通大學高階主管管理碩士學程，碩士論文。
6. Androcec, D., Vrcek, N., Seva, J., (2012), Cloud Computing Ontologies: A Systematic Review, MOPAS 2012: The Third International Conference on Models and Ontology-based Design of Protocols, Architectures and Services. April 29 - May 4.
7. Brandic, I. and Buyya, R., (2012), Special section: Recent advances in utility and cloud computing, Future Generation Computer Systems, Vol.28, Iss.1, pp.36-38.
8. CaoLei, Q., Dengsheng, L., (2012), The thinking of Cloud computing in the digital construction of the oil companies, Physics Procedia, Vol.24, Part A, pp.640-644.
9. Chiou F.C., (2010), Technology Life of Cloud Computing Using Literature Analysis, Master Dissertation, National Dong-hwa University, Hualien, Taiwan.
10. Kulkarni, G., Gambhir, J., Palwe, R., (2012), Cloud Computing-Software as Service, *International Journal of Cloud Computing and Services Science*, Vol.1, No.1, pp.11-16.
11. Kulkarni, G., Khatawkar, P., Gambhir, J., (2011), Cloud Computing- Platform as Service, *International Journal of Engineering and Advanced Technology*, Vol.1, No.2, pp.115-120.
12. Kulkarni, G., Sutar, R., Gambhir, (2012), J., Cloud computing-infrastructure as service-Amazon EC2, *International Journal of Engineering Research and Applications*, Vol. 2, Iss.1, pp.117-125.
13. Han, T. and Sim, K.M., (2010), An ontology-enhanced Cloud service discovery system. *Lecture Notes in Engineering and Computer Science*: Vol.2180, Iss.1, pp.644-649.

14. Reshma, V.K., Saravana Belasis B., (2012), Cloud Service Publication and Discovery Using Ontology. *International Journal of Scientific & Engineering Research* ,Vol. 3, Iss.5, pp.1-5.
15. Salton, G., and Buckley, C., (1988), “Term-Weighting Approaches in Automatic Text Retrieval,” *Journal of Information Processing and Management*, Vol. 24, No.5, pp. 513-523.
16. Sultan, N., (2010), Cloud computing for education: A new dawn?, *International Journal of Information Management*, Vol.30, pp.109-116.
17. Synergytek, “2013 Top20 USPTO Assignee 主要取得那些技術領域的專利呢?”, 參考網址 <http://synergytek.com.tw/blog/2014/03/10/>
18. Taghaboni-Dutta, F., Trappey A., Trappey C., Wu H.Y., An exploratory RFID patent analysis, *Management Research News*, Vol.32, Iss.12, pp.1163 – 1176, 2009.
19. Tseng, Y. H., Lin C.J., Lin Y.I., Text Mining Techniques for Patent Analysis, *Information Processing & Management*, Vol.43, No.5, pp.1216-1247, 2007.
20. Wang, L., Laszewski, G., Kunze, M. and Tao, (2010), J., Cloud computing: a perspective study, *J New Generation Computing*, pp.1-11.
21. Xue, K.Y., Liou, C.L., Huang J.Y., (2011) “Patent portfolio of cloud computing industry : a case study of Google”, *Proceeding of 1th Cross-Strait Academic and Practice Conference on Trade and Business Management*, Feng Chia University, Taichung, Taiwan.

A Research on Combining Ontology with Text-mining for Patent Retrieval of Cloud Computing

Jia-Yen Huang*, Ching-Hui Huang

National Chin-Yi University of Technology, Department of Information Management

jygiant@ncut.edu.tw

Abstract

Cloud computing is now regarded as an emerging business opportunity for network technology after the emergence of Web2.0; however, there are few academic works in this area and many research gaps are still to be investigated. The formation of commercial services of any emerging technology must rely on the support of a huge number of patents, and cloud computing is no exception. Since cloud computing is in a rapid development period, and its context contains a large number of specific terminologies, patent engineers normally cannot efficiently collect and conduct patent analysis. Therefore, this study proposes a patent retrieval policy which is a combination of snowball technique, block building patent retrieval and ontology applications. The text mining based retrieval policy can efficiently and correctly retrieve patent documents of three cloud computing business model; moreover, the major classes of cloud computing patent can be revealed. The analysis process and the results of this study can provide a valuable reference for patent-related researchers intending to conduct patent analysis of cloud computing.