

行政院國家科學委員會補助專題研究計畫成果報告

企業網路使用記錄之資料發掘

計畫類別：個別型計畫

計畫編號：NSC - 90 - 2416 - H - 004 - 029

執行期間： 90年 8月 1日至 91年 7月 31日

計畫主持人：楊亨利 教授

計畫參與人員：賴冠龍、郭展勝

執行單位：國立政治大學 資訊管理學系

中 華 民 國 91年 10月 11日

行政院國家科學委員會專題研究計畫成果報告

企業網路使用記錄之資料發掘

計畫編號：NSC - 90 - 2416 - H - 004 - 029

執行期限：90年8月1日至91年7月31日

主持人：楊亨利 教授 國立政治大學 資訊管理學系

計畫參與人員：賴冠龍、郭展勝

一、中文摘要

以往對全球資訊網的資料發掘主要侷限於網際網路上，而發掘的規則也分別是有關網頁內容或使用紀錄者，沒有看到兼具兩者的知識。然而，隨著企業內網路的運用日益廣泛，有關其資料發掘的研究也日益。本研究把「資料內容」與「使用者紀錄」加以結合，提出一個具有學術價值及實務可行的資料發掘架構，來發掘企業網路中所隱含的兼具「特性規則」及「關聯規則」之使用行為知識。另一方面也考慮到發掘之前置處理工作。同時以 JAVA 建置一個雛形來驗證其可行性。

關鍵詞：企業內網路，使用記錄、關聯規則、特性規則、資料發掘

Abstract

Most of past researchers have only paid attentions to data mining on Internet. Their mining rules are either related to Web content or Web Usage, but not to both. Since the applications on Intranet have become widespread, its related data mining is now important. This research has combined "Web content" and "Web usage" data, and proposed a data mining framework, which should be academically valuable and practically feasible. This framework includes the pre-processing phase of data mining. In addition, we will implement a prototype in JAVA to test the feasibility of the framework.

Keywords: Intranet, Usage Log, Association Rule, Characteristic Rule, Data Mining

二、計畫緣由與目的

在現代社會中，能快速有效的掌握資訊與知識就是贏家。資料發掘 (Data

Mining)，或稱知識發現 (Knowledge Discovery) 課題便是在此種背景下應運而生。各種針對現行資料庫的資料發掘方法與技巧也紛紛提出，而廣泛地在國內外為人所討論 (如 Frawley, Piatetsky-Shapiro and Matheus, 1991; Han, Cai, and Cercone, 1991; Yoon and Kerschberg, 1993; Grupe and Owrang, 1995; Fayyad, 1996; 周立平, 民 84 年; 薛如芳, 民 84 年; 宋振裕, 民 85 年; 陳博智, 民 86 年; 顏秀珍, 民 86 年; 黃明正與苑守慈, 民 88 年)。

隨著這幾年全球資訊網在網際網路 (Internet) 上的蓬勃發展，主要以 HTML 格式分散存在於網際網路上各公開或私有網站中的資料，也逐漸累積成為一個內容豐富而不可忽視的資料來源。但由於全球資訊網在與傳統資料庫相較之下，又增加了較非結構化、資料分散、內容廣闊與複雜、使用者難掌握其意圖等特性 (Han, Zaine and Fu, 1995)。使其中的知識之發掘變更加不易 (Schwartz, et al., 1992; Bowman, et al., 1994; Etzioni, 1996)。

隨著網際網路的發展，存在防火牆之後的企業內網路 (Intranet) 也運用日益廣泛。雖然企業內網路被定位在企業內部，但隨著許多企業組織在地理上擴展到全球的現實，企業內網路也有可能成為橫跨國際分布的巨大網路系統。企業網路上的資料也形成一個知識發掘的來源。

不管是 Internet 或是 Intranet，其網頁伺服器均會留存 Web log 的交易紀錄 Log 本身其實隱含相當多人類使用資訊科技或是網路的行為知識。傳統 Log 的分析常常是以統計為主要的分析工具，但是這種做法，只能適用於在網路基礎建設作業的資訊分析上，對於紀錄檔中可能隱含的有效資訊，往往無法進一步加以擷取，因而使得網路紀錄檔的應用侷限於伺服器的平衡

負載、異常活動、無法成功的呼叫、網站的開發及管理用途上。以往對全球資訊網資訊發掘研究分類有如內容、鏈結與使用三類 (Zaine, 1999)，各類有其發掘出的規則。但即使如 Han 等的研究，也尚未看到某一規則兼具兩類觀點者 (如內容與使用)。而對使用狀況的資訊發掘也均侷限於網際網路 (Internet)，挖掘的規則種類以網頁瀏覽之關聯規則為主 (如 Mobasher, et al., 1996; Chen, Park and Yu, 1998; 陳仕昇等, 民 88 年; Pei, et al., 2000; Chen, Chu and Yang, 民 89 年; 陳榮靜與沈慧宇, 民 89 年)，其運用也多偏於產品促銷組合、網頁佈置方式等行銷方面 (如 Allen, Kania, and Yaeckel, 1998; Lai and Yang, 2000; 楊子青, 民 89 年)。也有試圖來發掘其隱含的時間序列，以分析網路流量特性 (Zaine, Xin, and Han, 1998)。這些方面當然很重要，但是否還有其他種類的規則隱含於 Log 中呢？另一方面，由於瀏覽者來自各處，無法確實辨識其身份，即使採用會員登入制，對其基本資料的掌握仍相當有限，甚至，其基本資料本身也不正確。若未經瀏覽者同意的擷取其個人資料或使用 Cookie，又會侵犯其隱私權。這就形成其資料發掘之網際網路基本限制。

相較於網際網路，我們對於企業網路之使用者其實有更多的背景知識。只要，在工作契約上載明，我們對其使用企業網路可以留存更多的紀錄與使用 Cookie，而無隱私權的顧慮。所以，相對於網際網路留存了更多的資料。這對於在講求知識管裡的今日企業，將是把企業員工日常運作所隱含的知識外顯化的重要課題。本研究目的在提出一個完整的企業網路知識發掘架構 (包含其前置作業)，並建置一雛形來發掘結合對網頁「內容」及「使用者紀錄」的特性、區別、關聯或混合規則。

三、研究方法、結果與討論

(一) 研究架構

a. 本研究所提出企業內網路知識發掘之包含前置作業的整體架構，如圖一所示。其中「AP Database」儲存了企業運作所需要的知識或資料，譬如人事、業務績效等資料。「資訊定義檔」乃參考我們

以往研究 (楊亨利、金士俊, 民 90 年) 所建議的各文件的 Meta Data 定義。每一份 Intranet 文件，都會包含「資訊特徵辨識碼」、「目的」、「作者」、「URL」、「格式」、「頁數」、「領域」等資料。資訊特徵辨識碼為此文件的主鍵，根據文件內容的 ASCII 碼，使用 MD5 或 SHA-1 演算法計算其特徵值。當企業內部使用者平日瀏覽存取 Intranet 文件時，「傳送資訊編碼器」將每次傳輸給 Intranet 使用者的資訊，動態擷取特徵進行編碼，並將其存於「資訊存取特徵日誌」的紀錄中。此紀錄主要包含數個欄位：連線序號、伺服器狀態、方法、URL、IP、存取時間及經由傳送資訊編碼器計算的資訊特徵辨識碼。

「資料前置處理模組」乃在資料進行探勘前，做初期工作。清洗工作為去除空值或修正一些錯誤的雜訊處理。轉換工作為將一些資訊作格式的轉換 (例如時間轉換成格林威治時間；將 IP 轉會成 Domain Name；有時必須把一份文件之各子文件 Part1、2、3... 等聚合成文件)。去除工作為去除一些資料探勘時期不需要或是多餘的工作紀錄 (如去除訊密度較低或已包含在文字資訊中之多餘資料記錄，如公司 Logo 等圖檔)。整合工作為將多種不同的資料來源 (績效、人事、交易紀錄等)，整合成為一個具有完整意義的資訊來源。

不管是 Internet 或 Intranet，凡任何存取 Web 資源的動作，其網頁伺服器均會留存 Web Log 的交易紀錄。在網路伺服器軟體問世之初，美國國家超級電腦應用中心 (NCSA) 曾訂定其網路之「通用紀錄格式」 (Common Log Format)，而後微軟有其專屬的 IIS Log，國際的 W3C 組織也提出延伸的 Log 格式 (Extended Log) (Hallman-Baker and Behlendorf, 2000)。以往網頁的 Log 紀錄發掘大多是以 Web 伺服器所產生的 Log 為依據，進行 Mining。但面對企業環境彈性多變，往往需要高度的應變能力，以及對知識的渴望，單純靜態網頁的內容，已經沒辦法滿足員工對企業內的知識的需求。企業常必須利用動態網頁的能力，讓使用者可隨意查詢企業內部的資訊。

但是，這往往也造成網頁伺服器在網頁瀏覽紀錄的問題：現有的網頁伺服器只

能針對網址作一個紀錄，並不能紀錄使用者看過的內容。在以往的研究中只有針對存取紀錄作資料探勘，這種做法忽略了下面幾種的狀況。(a)同一個 Intranet 使用者在看不同網址的網頁時，看到相同的內容。(b)不同的 Intranet 使用者在看相同網址的網頁時，看到不同的內容。(c)同一個 Intranet 使用者在不同的時間點，看相同網址時，看到不同的內容。

本研究嚐試以資料內容替代原始 Log 的網址再進行資料探勘，這種做法可以使探勘的資料更為精確，讓發掘的知識更能正確的反應使用者在 Intranet 上的行為。為了要解決上述的問題，我們可以有多種的做法。方法一使用 Web ISAPI 延伸規格，缺點是 Web 伺服器的 ISAPI 規格並不統一，甚至有些 Web Server (如 Tomcat) 不支援。方法二使用特製的 Web Browser 加以紀錄；但使用者常會有個人的偏好，難以強迫其使用。方法三使用 Web Proxy 所謂的 Web Proxy 即是透明的 Web Browser 代理人，當使用者欲瀏覽網頁資訊時，所提出的 Request 實際上並不是直接送入 Web 伺服器，而是先被 Proxy 代理人攔截；由 Proxy 代替原有的 Browser 向 Web 伺服器提出 Request，當伺服器把資料回傳時，由 Proxy 加以紀錄，並把結果送回 Browser。本研究使用方法三，並解決幾個問題(1)關閉 Browser 的快取，以免擷取到舊的文件；(2)實作 Thread Pool 的管理功能，限制最大可以連接的 Connection，以解決連線效能問題；(3)透過 RSA MD5 或 SHA-1 演算法，計算網頁的內容的特徵為固定長度的雜湊值，方便比對內容是否一致，以解決動態文件內容的問題。讓資訊探勘的工作由原始 Log 中的路徑資料紀錄的層次提升到以資訊為探勘對象的層次。

在演算法方面，我們主要是採用 Han 的概念樹導向歸納學習法 (Han, Cai and Cerocne, 1992) 來找特性規則與區別規則，略修改 Apriori 之關連資料分析法 (在找 Candidate 前，增加對項目的排序) 來找關聯規則 (Agrawal and Skikant, 1995; Houtsma and Swani, 1995a, 1995b)。除此外，本研究嚐試混合了尋找關聯規則與特

性規則兩種演算法。混合資料發掘演算法後所產生的結果常可以產生更有意義的解釋。在關聯規則中，只看到 A->B 這種規則，例如購買牛奶的人，有 75% 的機率會購買麵包。在特性規則中，往往只有某一種群資料的特性描述，例如：研究生有 75% 的機率是加拿大人，得到極佳的 GPA 或是有 25% 的機率是外國學生主修 science 而得到不錯的 GPA。倘若混合了這兩種的演算法，我們可以得到另外一種更精準且有趣的知識的描述。例如購買牛奶的人有 75% 的機率會購買麵包，而具有這種規則的人往往具有以下的特徵，75% 是女性、已婚、養育著多個小孩，25% 是男性、未婚、沒有小孩。

但在使用混合演算法時，必須要指定何種的演算法先執行。若先執行特性規則再執行關聯規則，可以針對有興趣的資料，找出一些特性規則，並從所有的資料中滿足任一特性規則之樣本為起使表格，再找所有這些樣本之行為的關聯樣式。倘若先執行關聯規則再執行特性規則，則會先找出有興趣資料在行為上的各種關聯樣式；每種關聯樣式隱含一種樣本的集群，對每一集群找出其特性規則。

(二)雛形實作

本雛形所使用的硬體設備為二台 Pentium III 500、256MB 記憶體。Windows 2000 之作業平台，分別模擬伺服器端環境及資料發掘用戶端環境的作業環境。開發語言為 Java。企業內部伺服器端的模擬採用微軟的 IIS Web 伺服器、SQL Server 2000 資料庫系統、以及 Serv-U FTP Server，在連接資料庫的部分，則是採用微軟官方的 JDBC Type 4 Drive。

本研究實作所使用的 Log 紀錄與比起前述架構中的 Web 存取紀錄，多了 FTP 的存取紀錄。Web 伺服器紀錄大多是屬於一個檔案的存取便是一筆紀錄，然而在 FTP 上，一個檔案存取的動作會再分為更多的細部動作，因此，會產生多筆的紀錄。本研究之 Web 存取監控模組只是一個 Java 指令，由執行 java Jar 檔來啟動、停止 Proxy Server。FTP 並沒有採用 Proxy 的紀錄方式，主要是考量 FTP 協定並非採用單純的 Request-Response 模式，若以 Proxy

的做法實作，雛型則會過於複雜。因此，FTP 的 Log 處理採用的是透過 FTP 匯入模組，分析 Serv-U 目錄底下的 Log Files¹，解讀每一個 FTP Connection 的 Log，產生本研究所需要的 Log 格式記錄，再以檔案匯入的方式，存入資料庫。無論是 FTP 或 Web HTTP 兩者都可以在一個 Socket Connection 中傳遞多個檔案。在 Http1.1 協定中，一個 Connection 可以同時傳遞數個檔案，每一次傳遞稱之為一個 Hit，FTP 雖然沒有這個名稱，但在傳輸過程也必須要維持同一個 Connection，也具有在同一個 Connection 傳遞檔案的能力。所以，Log 在資料庫的 Schema 有兩張表格，一個是 Connection Table，一個是 HitTable。另外，在本系統中，每棵概念樹是由一組 XML 所定義的。

本研究已實作下列幾類資料發掘 (1) 特徵規則：看公司業務網頁的人有何特性、具有某種人格特質的員工所看的文件有何特性、或容易導致網路塞車的網頁特性；(2) 區別規則：高績效與低績效員工所看的文件有何差異、或業務部門及非業務部門所看的文件有何差異；(3) 有頻率考量之特徵規則：常被檢索異動的文件有何特徵？(4) 關聯規則：那些文件常會被同時瀏覽；(5) 混合規則(先特徵再關聯)有某些特徵銷售金額為公司前 70% 的人員的員工的瀏覽文件行為規則；(6) 混合規則(先關聯再特徵)：有某些瀏覽行為的員工群組有何特性；(7) 特殊的規則：不同部門間文件的關係，即找出甲部門建立的文件常常被乙部門去修改

在介面設計上，如圖二，本研究將視窗分為數個區域，各有不同的功能。為讓使用者能夠快速產生所需要的 SQL 指令，在操作的動線上，採用 Z 型動線的方式。這種設計方法可以讓使用者以由左而有由上而下的方式進行操作，當然，使用者也可以隨時反向退回任一個步驟，或是由隨時由任何一個步驟切入。在輸入限制條件，採用樹狀結果搭配中序 (In-order) 追蹤的方式限制式優先權的關係，可表達

¹ 但需要另一個檔案系統監視模組每隔一段時間，即監控是否檔案有因 FTP 異動。若有，需即時計算其資訊特徵辨識碼，與 FTP Log 比對。

如 A and (B or C) and D、或 (A and B) or (C and D)。本系統內含一個 SQL 產生器，但也容許使用者在調整視窗預覽，並額外調整。這一個輔助畫面可以分為兩區：命令區及預覽區。命令區針對每一種屬性獨立設定，有兩列：第一列是指派以頻率或是順序的高低排序，第二列是指派參數。參數有兩種格式的輸入，一種是絕對值、一種是百分比。舉例來說，假設使用者想篩選出「常常使用 Intranet 員工」的存取紀錄，則使用者可以存取紀錄的 User 欄位中選頻率由高到低的命令，並在參數區的地方輸入 20%。

四、結論與建議

本研究在學術上的貢獻有：(1) 提出一個可以挖掘企業內網頁瀏覽紀錄的架構，(2) 提出一個可以線上記錄動態網頁知識變化的 Log 方法。(3) 以新的角度整合兩種目的完全截然不同的演算法進行資料發掘，產生出更有價值的知識。在實務應用上的貢獻有：(1) 實作一個系統，該系統具備擴充性及獨立性²。(2) 提供具有彈性的 SQL 產生器使用者介面，讓使用者容易下達其需求。(3) 若應用本系統，對於企業一般使用者，本雛形可使其了解 Intranet 上的文件，是依何種知識或規則的方式存在。企業主可以藉由本雛型找出知識及規則，進而應用此規則到員工身上，提昇員工的作業績效。公司的 MIS 部門也可以依此照這些規則，最佳化公司文件的配置，讓員工在 Intranet 上能夠以最容易、最快速的方式存取員工有興趣的文件。

當然，本研究的研究在實作上仍有限制如下：(1) 伺服器端模組並無圖形的操作介面；(2) 在 SQL 設定篩選限制式時無法接受概念樹非葉節點之概念；(3) 操作者仍必須對資料庫查詢路徑有一定程度

² 伺服器端或是用戶端皆以元件的形式加以開發，每一個元件都可針對特定目的獨立運作。在用戶端，所有的資料發掘演算法，都是 Design Pattern (Gamma, Helm, Johnson, Vlissides 1998) 中具物件導向特性 Strategy Pattern 進行實作，因此即使未來演算法有所變動，都可依需要，新增或修改。此外，可擴充領域範圍，容許使用者在 Meta Data 項目中新加入的網際網路應用領域屬性。也可以利用 XML 快速加入新的概念樹定義。

的了解，系統不能將表格間的關係自動建立；(4)對於資料庫的處理，仍有使用微軟 SQL SERVER 所獨有的 T-SQL 專屬語法。

在後續研究上，未來可行的發展方向包含：(1)可以繼續定義更周全的企業 Meta Data 項目與屬性，並找尋其概念樹；(2)可以收集更多的 Log 資料，以更詳盡的方式來作演算法正確性的驗證；(3)增加 SQL 產生器產生巢式的查詢子句功能；(4)使用其他演算法，以期滿足實務上高效率需求。

五、計畫成果自評

研究內容與原計畫甚為相符，由上述結論可以看出達成之預期目標包含：創新模式架構之提出及實驗雛形之建立，並在此過程中培育人才。學術價值高，也有應用價值，可供企業參考，相關之成果正在改寫成論文，期待發表於國內外之期刊。

六、參考文獻 (因篇幅限制，僅列出部分文獻)

- 【1】 Agrawal, R. and Srikant, R., " Mining Sequential Patterns, " *IEEE 11th International Conference on Data Engineering*, Taipei, Taiwan, March 1995.
- 【2】 Allen, C., Kania, D., , and Yaeckel, B., *Internet World Guide to One-to-One Web Marketing*, Wiley, New York, 1998.
- 【3】 Bowman, M., Danzig, P. B., Manber, U. and Schwartz, M., " A Scalable, Customizable Discovery and Access System, " Technical Report CU-CS-732-94, Department of CS, University of Colorado, Boulder, July 1994.
- 【4】 Chen, M. S., Park, J.S. and Yu, P. S. " Efficient Data Mining for Path Traversal Patterns, " *IEEE Transaction on Knowledge and Data Engineering*, April 1998, pp:209-221
- 【5】 Chen, Tzung-Shi, Chu, Chih-Hao, and Yang, Chia-Ching, " Mining Traversal Patterns on the Internet, " *第六屆資訊管理研究與實務研討會*, 民 89 年。
- 【6】 Etzioni, O., " The World-Wide Web: Quagmire or Gold Mine? " *Communications of ACM*, Vol.39, No.11, November 1996, pp.65-68.
- 【7】 Fayyad, U. M., " Data Mining and Knowledge Discovery : Making Sense out of Data, " *IEEE Expert*, Vol.11, No.5, October 1996, pp.926-938.
- 【8】 Frawley, W. J., Piatetsky-Shapiro, G. and Matheus C. J., " Knowledge Discovery in Databases : An Overview , " *Knowledge Discovery in Databases*, California, Edited by Piatetsky-Shapiro, G. and Frawley, W. J., AAAI/MIT Express, 1991, pp.1-30.
- 【9】 Gamma, R. Helm, R. Johnson, and Vlissides, J, *Design Patterns*, Addison-Wesley 1998.
- 【10】 Grupe, F. H. and Owrang, M. H. " Data Base Mining Discovering New Knowledge and Cooperative Advantage, " *Information System Management*, Vol.12, No. 4, Fall 1995, pp.26-31
- 【11】 Hallman-Baker, P.M. and Behlendorf, " Extended Log File Format, " 2000, available from <http://www3.org/pub/WWW/TR/WD-logfile.html>.
- 【12】 Han, J., Cai, Y. and Cercone, N., " Attribute-Oriented Induction in Relational Databases, " in G. Piatetsky-Shapiro and W. J. Frawley (eds.), *Knowledge Discovery in Databases*, AAAI/MIT Press, 1991, pp. 213-228
- 【13】 Han, J., Cai, Y. and Cercone, N., " Knowledge Discovery in Databases : An Attribute-Oriented Approach, " *Proceeding of the 18th VLDB Conference*, Canada, August, 1992, pp. 547-549.
- 【14】 Han, J., Zaine, O. R., and Fu, Y., " Resource and Knowledge Discovery in Global Information Systems: A Scalable Multiple Layered Database Approach, " *Proc. of a Forum on Research and Technology Advances in Digital Libraries (ADL'95)*, McLean, Virginia, May 1995.
- 【15】 Houtsma, M. and Swami, A., " Set-Oriented Data Mining in Relational Databases, " *Data & Knowledge Engineering*, Vol.17, No.3, September 1995a, pp.245~162.
- 【16】 Houtsma, M. and Swami, A., " Set-Oriented Data Mining for Association Rules in Relational Databases, " *IEEE 11th International Conference on Data Engineering* , Taipei, Taiwan, March 1995 b, pp.25-33.
- 【17】 Lai, H. and Yang, T-C, " A Group-based Inference Approach to Customized Marketing on the Web-Integrating Clustering and Association Rules Techniques, " *Proceedings fo the 33rd Hawaii International Conference on System Sciences*, 2000.
- 【18】 Mobasher, B., Jain, N., Han, E., and Srivastava, J., " Web Mining: Pattern Discovery from World Wide Web Transactions, " *Technical Report TR86-050*, Department of Computer Science, University of Minnesota, 1996.
- 【19】 Pei, J., Han, J. Mortazavi-asl, B., and Zhu, H., " Mining Access Patterns Efficiently

- from Web Logs, ” *PAKDD'00 (Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Kyoto, Japan, April 2000.
- 【20】 Schwartz, M. F., Emtage, A., Kahle, B. and Neuman, B. C., “ A Comparison of Internet Resource Discovery Approaches, ” *Comput. Syst.*, No.5, Fall, 1992, pp. 461-493
- 【21】 Weibel, S., “ A Proposed Convention for Embedding Metadata in HTML, ” 1996, available from <http://www.oclc.org:5046/~weibel/html-meta.html/>.
- 【22】 Yoon, J. P. and Kerschberg, L., “ A Framework for Knowledge Discovery and Evolution in Databases, ” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5, No.6, December, 1993, pp.973-979.
- 【23】 Zaine, O. R., *Resource and Knowledge Discovery from the Internet and Multimedia Repositories*, Ph.D. Thesis, Simon Fraser University of Computer Science, 1999
- 【24】 Zaine, O. R., Xin, M., and Han, J., “ Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs, ” *Proc. Advances in Digital Libraries Conf. (ADL'98)*, Santa Barbara, CA, April 1998, pp. 19-29.
- 【25】 宋振裕, 一個漸進式資料庫知識擷取的方法, 交通大學資訊工程研究所碩士論文, 民國 85 年。
- 【26】 周立平, 從資料庫中發現法則：用學生修課的資料作分析, 淡江大學資訊工程研究所碩士論文, 民國 84 年。
- 【27】 陳博智, 歸納邏輯程式在知識庫修正及知識發掘之運用, 交通大學資訊工程研究所博士論文, 民國 86 年。
- 【28】 陳榮靜、沈慧宇, 「一個新的網站瀏覽路徑資料探勘之研究」, 第六屆資訊管理研究與實務研討會, 民 89 年。
- 【29】 陳仕昇、許秉瑜、陳彥良, 「以可重複序列挖掘網路瀏覽規則之研究」, *資管評論*, 第九期, 民 88 年 12 月, 第 53 至 71 頁。
- 【30】 黃明正、苑守慈, 「知識探勘與競爭者分析之研究」, 第五屆資訊管理研究與實務研討會, 民 88 年, 第 247 至 254 頁。
- 【31】 楊子青, 購物車資料在網站行銷溝通決策之應用, 中山大學資訊管理系博士論文, 民 89 年。
- 【32】 楊亨利、金士俊, 「一個全球資訊網資料發掘的架構 以英語教學為例」, *資訊管理學報*, 已被接受, 即將刊出於第 7 卷第 2 期, 民 90 年 1 月。
- 【33】 薛如芳, 以歸納學習法自關聯式資料庫中發掘知識, 交通大學資訊工程研究所碩士論文, 民國 84 年。
- 【34】 顏秀珍, 從大型資料庫中有效發掘之方法探討, 清華大學資訊科學研究所博士論文, 民 86 年。

