

應用文件探勘於專利文件之技術分析— 以磁阻性隨機存取記憶體為例

Using Text Mining to Analyze Technology from Patent Documentation: A Study on Magnetic Random Access Memory Technology

王明好¹ Ming-Yeu Wang
真理大學工業管理學系

許旭昇² Shih-Sheng Hsu
真理大學管理科學研究所

¹The Department of Industrial Management, Aletheia University &

²Graduate School of Management Sciences, Aletheia University

(Received April 13, 2005; Final Version September 12, 2005)

摘要：專利文件包含了豐富的技術資訊，公司能透過專利資訊的分析，瞭解某項技術或產品的發展現況。一項技術通常係由許多相關的技術類別所組成，過往多以該技術領域的專家根據專利分類號判斷其所包含的技術類別，然以專家進行技術的分類，不僅耗時且耗成本，且可能會有分類一致性的問題。因此要如何以快速及客觀的方法來取代或輔助人工的分類，找出一項技術所包含的技術類別，為相當重要的議題。文件探勘為近年來興起的新技術，其能從大量的文件中，挖掘隱含且有用的知識，然卻鮮少應用於專利文件上，因此本研究將嘗試以文件探勘的技術，以磁阻性隨機存取記憶體(MRAM)為例，根據相同專利分類號的專利文件，彼此在技術內容的相似度，找出MRAM所包含的技術類別，並以CHI研究公司所提出的專利指標，對MRAM的相關技術類別做進一步的分析。

關鍵詞：專利分析、國際專利分類號、文件探勘、磁阻性隨機存取記憶體、專利指標

Abstract : Patent documents not only provide the technological information, but also reflect the technological development and trend. Companies can discover the technological development after conducting patent analysis. One technology usually comprises several technological fields, so the prior

work at the time of performing a patent analysis is to confirm the technological fields of one technology. In the past, the works to identify the technological fields is often done by experts. Experts make the judgment of the comprised technological fields based on the International Patent Classification used by patent databases. However, this is a time-consuming and costly job and may cause inconsistent problems. In order to solve the problems caused by manual judgment, we introduce an emerging method, called text mining, to clarify the technological fields. The text mining method can discovery the implicit and useful knowledge from a large number of documents. Patents documents related to Magnetic Random Access Memory are selected as an example to demonstrate how to use text mining in clarifying technological fields. The technological fields are classified based on the similarity between patent documents. After classification, patent performances of each technological field is analyzed and compared according to the commonly used patent indicators proposed by CHI Research Corporation. At last we provide suggestions for future research.

Keywords : Patent Analysis, International Patent Classification, Text Mining, Magnetic Random Access Memory, Patent Indicator

1. 結論

專利，為公司最重要的無形資產，其不僅可以保護公司的研發成果，更可以做為競爭市場上相互攻防的工具。專利文件包含了大量有用的資訊，其能用於確認一新技術的發展機會、探測競爭對手的動向及評估研發的趨勢等資訊 (Gupta, 1999)，對於技術的創新及研發上，皆有很重要的影響。專利分析係以統計、歸納及比較等方法，將專利文獻所包含的資訊，轉換為有用的知識情報，透過專利分析的結果，能得知一公司的技術策略及規劃等資訊 (Ashton and Sen, 1988)，故公司能藉由對特定技術的專利分析，來探測此技術及其相關技術的發展狀況。

一項技術通常包含了許多重要及關鍵的技術類別，如微機電系統技術 (Micro Electro Mechanical System) 係包含了半導體製程、電子、機械、材料、控制及光學等重要技術類別 (詹前疆，民90)，而每項技術類別又包含了許多重要的子技術，且各有其對應的專利保護著，分別為不同的專利權人所擁有，因此Brockhoff (1992) 指出在進行特定技術的專利分析時，需先找出其所包含的技術類別，為進行後續分析時的重要前置處理步驟。為了瞭解某技術所包含的技術類別，過往多半係由該技術領域的專家根據國際專利分類號 (International Patent Classification, IPC) 判斷其所包含的技術類別，然專利文件數量眾多，且技術與法律的專有字詞並存 (曾元顯，民93)，如以人工進行專利文件的判讀，不僅耗時且耗力，且可能也會產生分類一致性的問題。因此要如何判斷已依IPC分類完成的專利文件，找出某項技術所包含的技術類別，為一相當重要的議題。

文件探勘 (Text Mining) 為近來年興起的新技術，主要是應用在未結構化或半結構化的文件，其目的為挖掘大量文件資料中的隱含資訊，然卻鮮少應用於專利文件。磁阻性隨機存取記憶體 (Magnetic Random Access Memory, MRAM) 被視為最有可能量產的新世代記憶體，係由於其具備現有三大記憶體的優點，並擁有非揮發性的長期儲存能力、低成本及能源消耗低等特性 (Kim *et al.*, 2004)，因此本研究將以MRAM為例，嘗試應用文件探勘的技術，根據相同IPC的專利文件彼此在技術內容的相似度，找出MRAM所包含的技術類別，取代過往採取人工分類的方法。最後根據技術分類的結果，以CHI研究公司¹所提出的專利指標，對MRAM的相關技術類別做進一步的分析，以期給予產業界參考。

本文共分為五個部份：第一部份為緒論，介紹研究背景、動機及目的；第二部份為回顧關於專利分析及文件探勘的文獻，並對MRAM做一簡介；第三部份為本文應用文件探勘於專利文件的研究方法；第四部份為MRAM技術類別的分類與分析結果；第五部份則對本文的研究結果提出結論與建議。

2. 文獻回顧

2.1 專利分析

專利是國家為了鼓勵發明的創新，而給予發明人對其發明的成果享受短暫的獨占權，而專利的申請與授與中，規定申請者有義務完整公開其發明內容，而授與者則給予申請者法律上的保護 (趙晉枚等，民91)。專利資訊係指從專利局所出版的文件中，所獲得的技術、市場、法律及其他公司的資訊²，廣義而言，專利資訊包括各國專利局所出版的專利公報、分類表、分類索引、統計資料及在審查過程中所有相關的文件；狹義而言，則單指發明、新型及新式樣說明書(黃文儀，民91)。專利資訊能提供關於技術發展及公司策略方面的情報 (Gupta, 1999; Michael, 1997)，並且為唯一能即時辨別及確認技術變化的資訊 (Campbell, 1983)。Ernst (2003) 便根據專利資訊發展公司研發管理的系統，其能用於市場競爭者的監控、技術的評估、研發部門的策略管理、評估外在潛在知識及研發部門的人力資源管理，並將專利資訊所能幫助決策者的應用歸納成七點：評估公司在競爭市場的技術組合、評估公司在特定領域的技術能力、辨別公司在競爭環境的策略、確認與評估外在知識的來源、評估新產業領域的專利情況、評估重要的市場伙伴及一技術領域的人力資源管理。

¹ CHI Research Inc.，網址：<http://www.chiresearch.com>。

² 歐洲專利局 (European Patent Office)，「何謂專利資訊」，上網日期：民國94年2月1日，網址：http://www.european-patent-office.org/patinfo/pro/index_moreinfo.shtml。

專利分析係利用統計的方法，將專利文獻所包含的資訊，轉換為有用的知識情報，其可以廣泛地應用於國家、產業、公司及技術領域上 (Pavitt, 1988)。如Jung and Imm (2002) 以專利分析探討台灣與南韓兩國的專利活動及專利管理策略；Pilkington *et al.* (2002) 以專利分析探討電動車輛技術的發展軌跡。從專利除了可瞭解技術的發展之外，過去有許多學者更將其視為經濟或創新的重要指標，如 Martiro (1993) 以相機、打字機及手錶三種產品為例，提出專利能作為產品市佔率的指標；Acs *et al.* (2002) 將專利件數視為創新活動的指標，來衡量一區域的經濟成長與發展狀況；賴奎魁等 (民91) 將專利累積獲准數視為重要的技術指標，來衡量全球、日本及台灣在影像感測器業的技術發展趨勢。國內首篇關於專利指標的研究 (孟憲鈺等，民89) 係以CHI研究公司 (CHI Research) 所提出的專利指標及資料庫為基礎，探討我國產業在專利件數、被引用次數及各項技術指標的表現。CHI研究公司為提供技術、科學及財務指標方面的研究及顧問公司，其根據美國專利庫提出了三類九種指標，分別為專利件數、專利成長比率、公司專利比率、引用次數、當前影響力、技術強度、技術生命週期、科學連結及科學強度。由於指標衡量容易且淺顯易懂，因此被全球各大企業及分析機構所廣泛的應用。

專利分析的結果如以圖形化的方式呈現即稱為專利地圖 (Patent Map)，依不同的製作目的可分為經營管理圖及技術圖兩大類³。前者係藉由各種統計的分析來整理圖表，如專利獲准件數、國家別及競爭公司別等統計圖表，其能作為經營管理的重要參考資訊。後者則是將特定的專利資訊作更進一步的探討，如歸納出每篇專利的技術及功效類別；其圖表隱含豐富的技術及研發資訊，可以得知特定技術的動向，並能預測技術的未來發展趨勢。

現行專利制度在技術上的分類係以IPC為主，IPC為全世界統一採用的專利分類制度，其編排的方式係由8部 (Section)、120類 (Class)、628次類 (Subclass)、約69000個主目 (Group) 加次目 (Subgroup) 五個層次所構成如圖1所示。目前的IPC版本係由世界智慧財產權組織 (World Intellectual Property Organization, WIPO) 於2000年1月1日開始實施的第七版，每五年更新一次，不僅分類簡單且檢索容易，而各IPC所代表的技術說明，可由WIPO所提供的IPC查詢系統⁴查得。

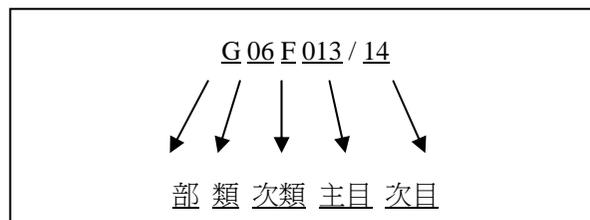


圖1 IPC的組成

³ 電通所法務智權園地，「何謂專利地圖」，民國89年，上網日期：民國94年2月1日，網址：<http://www.ccl.itri.org.tw/products/patent/88005.htm>。

⁴ WIPO線上IPC查詢系統，網址：http://www.wipo.int/classifications/fulltext/new_ipc/index.htm。

2.2 文件探勘

近年來，文件探勘已廣泛地應用於各個領域，過去已有些學者嘗試將文件探勘的技術應用於專利文件，然大部份皆著重於視覺化的呈現專利分析的結果，鮮少應用於專利文件的技術分類，如Fattori *et al.* (2003) 使用文件探勘的分群技術，輔以PackMOLE軟體，繪製出包裝產業與相關技術類別的專利地圖；Yoon and Park (2004) 應用文件探勘的概念，描繪分波多工技術(Wavelength Division Multiplexing) 的專利引用網路圖。文件探勘係由Feldman and Dagan (1995) 根據資料庫之知識探勘 (Knowledge Discovery in Database) 所提出的架構，主要是應用在未結構化或半結構化的文件，沒有明確區分欄位的資料，如網頁、電子郵件及答客問集等，目的為挖掘大量文件中的隱含及有用知識。

Sullivan (2001) 將文件探勘定義為一種編輯、組織及分析大量文件的過程，爲了要提供特定使用者之特定的資訊，以及發現某些特徵及隱含的關聯行爲。文件探勘整合了許多傳統資訊檢索技術，包括了特徵值的擷取、文件分類、自動摘要及文件分群等，其中文件分群係依文件內容的相似度進行分群，使群集具有「群內相似度高及群間相似度低」的特性。一般而言，文件分群的步驟分爲特徵值的擷取與選擇、文件呈現及分群三部份 (Wei *et al.*, 2002)。

(1) 特徵值的擷取及選擇：Frakes and Baezay (1992) 認爲文件中出現頻率高的字詞，與文件的主題有較高的關聯性，能將這些字詞視爲文件的特徵值 (Feature)，然若一字詞在每篇文件皆出現，則此字詞對文件則不具代表性 (Salton and Buckley, 1988)。由於文件分群的結果，會因文件的高向量維度(即特徵值的數目)及資料不足而導致分群成效下降 (Aggrawal and Yu, 2000)，故減少特徵值的數目爲相當重要的處理程序，一般常用的方法即爲對文件進行特徵值的擷取及特徵值的選擇。曾元顯 (民 86) 將特徵值的擷取方法分爲詞庫比對法、語義剖析法及統計分析法三種，分別說明如下。

- 1) 詞庫比對法 (Dictionary Approach)：即利用事先建立的詞庫，將文件中有出現在詞庫中的字詞擷取出來，其優點爲製作簡單、快速及容易實行，而缺點爲詞庫的大小、字詞的相關性及後續的維護爲一個大問題。
- 2) 語義剖析法 (Linguistic Approach)：係運用自然語言處理技術的語義剖析程式，輔以已經建立完成的詞庫，擷取出文件中的字詞，再運用一些方法及準則，過濾掉不適合的字詞，其缺點與詞庫比對法相同。
- 3) 統計分析法 (Statistical Approach)：爲利用統計方法來擷取文件的特徵值，最常用的方法爲字詞頻率，即將頻率落在某特定範圍中的字詞擷取出來。優點爲不用事先建立詞庫及文法也能擷取出重要的字詞，而最大的缺點爲出現頻率較低的關鍵詞無法被擷取出來。

特徵值的選擇係依據某些條件及準則，從擷取出來的若干特徵值中，選擇最具有意義及代表性的特徵值。一份文件中約有90%的字詞不具代表性 (林傑斌等，民91)，即文件中若有100

個字詞，則只要選取10個具有代表性的關鍵字，就足以代表此一文件的主題。特徵值選擇的方法很多，而不須事先知道文件類別且較易應用於分群的方法為文件頻率及字詞強度兩種 (Liu *et al.*, 2003)，各別說明如下。

- 1) 文件頻率(Document Frequency)：為一字詞在文件集裡出現的文件數，如字詞低於所設定的門檻值則去除，以留下出現頻率高及具有代表性的字詞，為一相當容易且低成本的方法。
- 2) 字詞強度(Term Strength)：係評估重要字詞與相似文件之間的關係，將大於門檻值的字詞留下。字詞在相似文件中出現的條件機率如(1)式，其中 x 與 y 是相似的文件而 t 為一字詞。評估當 t 在 x 文件出現時，也在 y 文件中出現的條件機率即 $S(t)$ ，最後將 $S(t)$ 大於門檻值的 t 留下，此法的概念與文件頻率法相似。

$$S(t) = P_r(t \in y | t \in x) \quad (1)$$

由於每篇文件中的特徵值出現次數及分佈皆不同，因此必須將所有的文件依特徵值進行權重的計算，常用的方法有以下三種 (Salton, 1989)。

- 1) 布林法(Boolean)：一字詞如有在文件中出現則權重為 1，反之則為 0，為計算特徵值權重中最簡單的方法。
- 2) 字詞頻率 (Term Frequency)：係直接以字詞出現在文件的次數為權重。
- 3) 字詞頻率乘文件頻率之倒數 (Term Frequency-Inverse Document Frequency, TF-IDF)：係由 Salton(1989)所提出的方法如(2)式， TF_{ij} 為特徵值 i 在文件 j 中出現的次數， DF_i 為所有文件中出現特徵值 i 之文件數， N 為全部文件數，最後將這兩數相乘作為特徵值 i 在文件 j 的權重 W_{ij} 。

$$W_{ij} = TF_{ij} * \log(N/DF_i) \quad (2)$$

- (2) 文件呈現：過去研究多以 Salton *et al.*(1975)所提出的向量空間模式 (Vector Space Model) 為主，係以向量的觀點將文件 D_j 視為具有多個維度的向量， f_{ij} 為特徵值 i 在文件 j 的出現次數，每個特徵值代表一個維度即分量，因此每份文件即是由一群分量所組成的向量，投射在多個維度的空間中和圖 2 所示，此為兩個特徵值在文件 j 的出現次數。

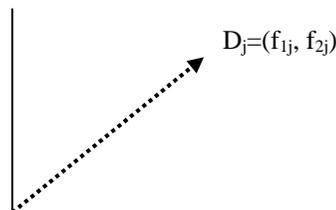


圖 2 向量空間模式

(3) 分群：一般常用的分群方法為階層凝聚式分群 (Hierarchical Agglomerative Clustering) 及非階層分群法中的 K-means 分群 (Steinbach *et al.*, 2000)。前者初始時係將每篇文件視為一群，循序合併相似高的群；後者則將所有的文件視為一群，依其文件相似度執行分割動作，且需要事先決定群組數。在所有分群的方法中，雖然階層凝聚式分群法的運算速度較慢，然分群完的群集品質最好 (Dubes and Jain, 1988)；而在階層凝聚式分群法中，又以華德法 (Ward's Method)及完全聯結法 (Complete Linkage Method) 表現較佳 (Griffith *et al.*, 1986)。

2.3 MRAM 簡介

MRAM的概念早於1972年就已經被提出，然由於還有許多問題待解決，故當時沒有造成極大的迴響。圖3為MRAM的技術發展流程，1996年美國國防先進研究計畫署 (Defense Advanced Research Projects Agency) 首先開始MRAM技術的研發計劃，但延自Baibich *et al.* (1998) 發現巨磁電阻效應 (Giant Magnetoresistance, GMR) 能應用於磁性及非磁性的多層膜後，MRAM技術的開發才開始又引起大家的重視。

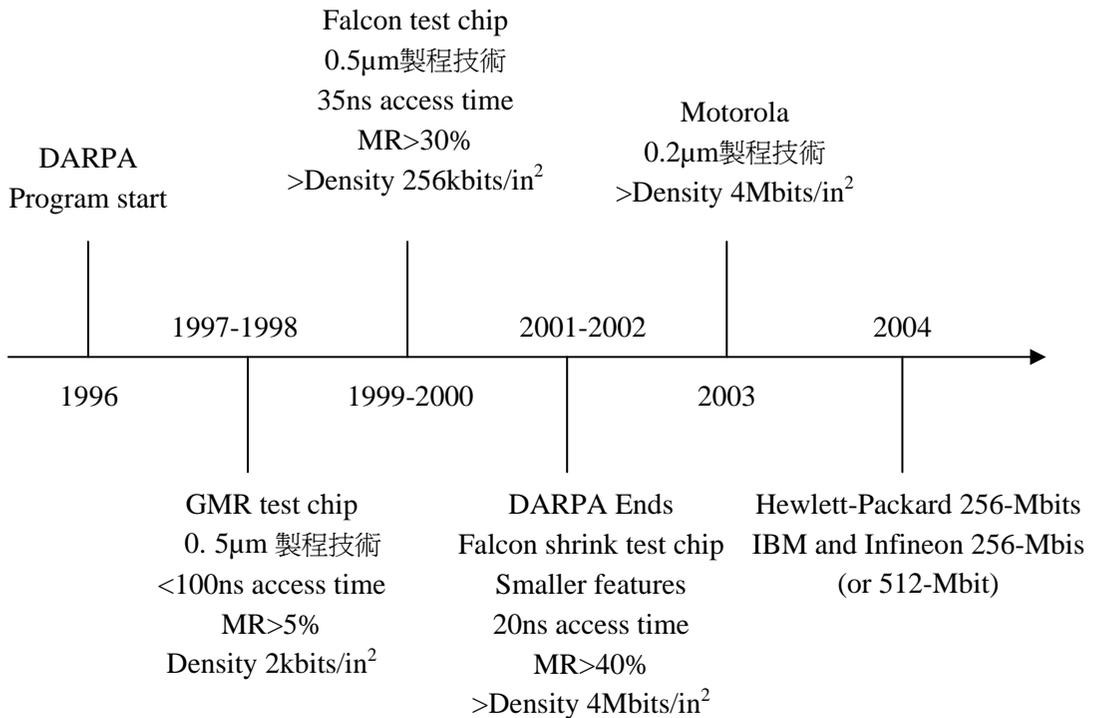


圖3 MRAM技術發展流程

資料來源：葉林秀等 (民93)

MRAM的基本單位係由磁性、非磁性與磁性三層薄膜堆疊所構成如圖4所示，當兩磁性層的磁化方向為順向排列時，因磁阻較低故輸出電壓較低；當兩磁性層的磁化方向為反向排列時，磁阻較高故輸出電壓較高。所以MRAM的運作原理便是根據輸出電壓的高低，來判別0與1的訊號(葉林秀等，民93)。

由於 MRAM 具有非揮發性的長期儲存能力、低成本及能源消耗低等特性 (Kim *et al.*, 2004)，國際半導體技術藍圖 (International Technology Roadmap for Semiconductors) 將 MRAM 列為最有可能量產的新一代記憶體技術。我國的工研院與台積電，自 2002 年便開始合作開發 MRAM 的相關技術，並於 2004 年時成功開發出 MRAM 的雛型，且後續的研究仍然在持續中，顯示了我國對 MRAM 技術的重視。時至今日，MRAM 已成為世界許多家大廠如國際商業機器 (IBM)、摩托羅拉 (Motorola) 及東芝 (Toshiba) 等爭相投入的重要技術。

3. 研究方法

本研究的流程分為專利文件的蒐集、文件探勘及專利分析三部份如圖 5 所示，詳細說明如下。

(1) 專利文件的蒐集：本研究係以 "Magnetic Random Access Memory" 為全文檢索的關鍵字，蒐集美國專利資料庫中，專利公告日為 2005 年 1 月 1 日之前的專利文件。由於美國為研發 MRAM 技術的領導國家，且許多重要的技術皆會在美國申請專利 (Mogee, 1991)，故本研究選擇美國專利暨商標局 (United States Patent and Trademark Office, USPTO) 所提供的線上專利資料庫⁵，做為 MRAM 專利文件蒐集的來源。

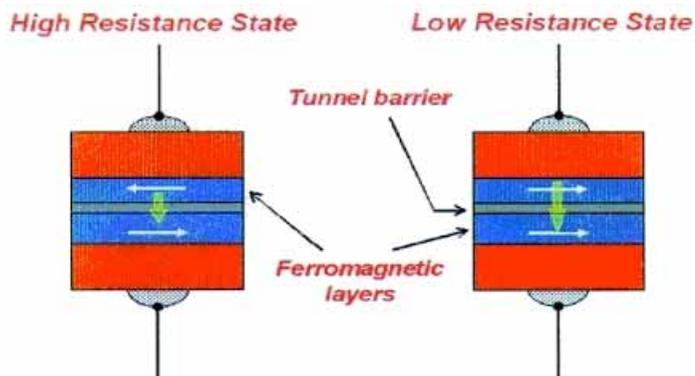


圖4 MRAM的結構及原理
資料來源：Ditizio *et al.* (2004)

⁵ USPTO，網址：<http://www.uspto.gov>。

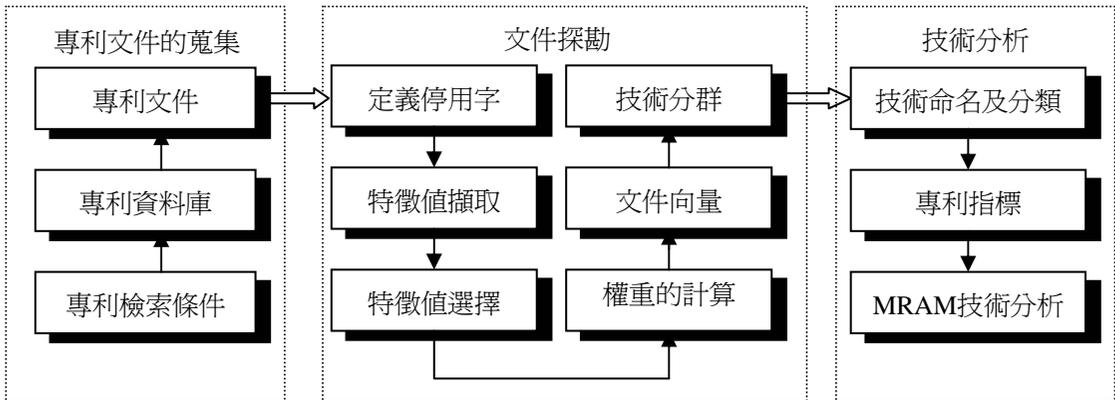


圖 5 研究流程

(2) 文件探勘：此部份分為六個步驟，各別說明如下。

- 1) 定義停用字：本研究以 USPTO 所定義的停用字⁶，做為文件特徵值擷取的停用字。
- 2) 特徵值擷取：在特徵值擷取的部份，本研究係採用 Megaputer 公司⁷所發行的 TextAnalyst 軟體，計算全部專利文件中，各個字詞出現的頻率。TextAnalyst 為結合語義學 (Semantic) 及類神經網路 (Neural Network) 的人工智慧軟體，其主要功能係能從大量的原始文件資料中，快速及準確地挖掘出各文件之間的關聯，並能統計所有字詞的頻率並給予權重。
- 3) 特徵值選擇：本研究在特徵值選擇的部份，係將 TextAnalyst 所擷取出來的特徵值，計算各特徵值的熵值 (Entropy)。熵源自於熱力學，係評估整個熱力環境中，分子的混亂活動程度，故又可稱為混亂函數。Shannon (1948) 將其概念應用於資訊理論 (Information Theory)，其計算公式如(3)式，其中 $P(x)$ 為某字詞在文件出現的比率，為評估與衡量資料中各字詞的分佈程度。

$$H(x) = - \sum_{x \in X} P(x) \log P(x) \quad (3)$$

如圖 6 所示，特徵值經過上式熵值的處理後，若一字詞均勻的分佈在各文件中則熵值較高，出現過多及過少的字詞反而熵值較低，之後根據熵值的大小給予排序，選擇前幾個最具有代表的字詞做為此份文件集的特徵值。

⁶ USPTO，停用字列表，網址：<http://www.uspto.gov/patft/stopword.htm>。

⁷ Megaputer Intelligence Inc.，網址：<http://www.megaputer.com/>。

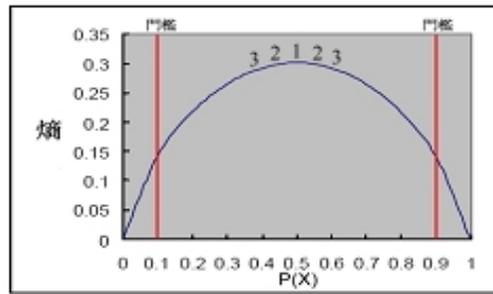


圖6 特徵值的選擇

註：圖中1, 2等數字依指將特徵值依熵值大小的排序結果；

1指熵值最高的特徵值，其餘依此類推。

本研究係根據相同IPC的專利文件彼此在技術內容的相似度，找出MRAM所包含的技術類別，其中IPC係由部、主類、次類、主目及次目五個層次所構成，本研究係取到IPC的主目為止，即將相同主目之專利文件集合起來視為一分析單位，如IPC為G06F 017的相關專利包括了G06F 017/10、G06F 017/21及G06F 017/23...等專利文件如圖7所示。

由於熵值的計算係以各別文件為單位，故將原公式更改成(4)式，以利計算各特徵值的熵值並依照大小排序，其中 $P(x)$ 為特徵值在相同IPC之專利文件中出現的平均次數。

$$H(x) = -P(x) \log P(x) \tag{4}$$

- 4) 權重的計算：Aas and Eikvil(1999)指出資訊檢索領域中最廣泛及常用的權重計算方法為TF-IDF，由於本研究係視IPC為一分析單位，因此將原式略做修改如(5)式，其中N為全部MRAM專利文件中所涵蓋之IPC的個數， TF_{ij} 為特徵值i在 IPC_j 出現的平均次數， DF_i 為特徵值i在不同IPC類別中出現的類別數。

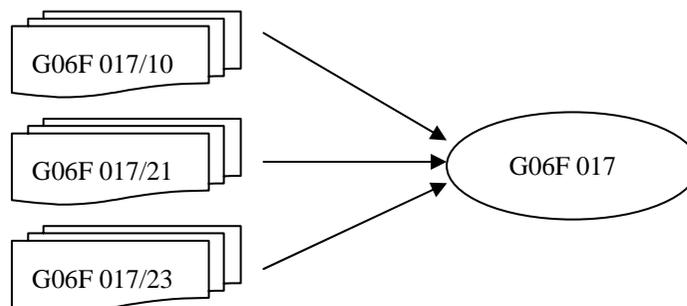


圖7 IPC分析單位

$$W_{ij}=TF_{ij}*\log (N/DF_i) \tag{5}$$

- 5) 向量空間：本研究依據 Salton *et al.* (1975) 的概念，將第三步驟所得到的特徵值根據式(4)，將各個 IPC_j 依特徵值 i 的權重 W_{ij}，轉換成 IPC 向量模式如圖 8 所示，如 IPC₁ 原為 IPC₁=(t₁₁, t₁₂, t₁₃,.....t_{1i})係由 i 個特徵值 t 所組成的文字向量，經過特徵值權重處理後變成 IPC₁=(W₁₁, W₂₁, W₃₁,.....,W_{1i})。
- 6) 技術分群：本研究選擇分群成效品質較佳的華德法進行 IPC 文件的分群。華德法又稱為最小變異數法(Minimum Variance Method)係由 Ward(1963)所提出，其將每一樣本資料視為一個群集，以歐幾里德直線距離的平方如(6)式，衡量各文件之間的相似度，然後依序將群集合併。

$$Sim(x,y)=\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + + (x_i - y_i)^2} \tag{6}$$

(3) 專利分析：本研究在 MRAM 技術的分類上，係依據相同 IPC 的專利文件彼此在技術內容的相似度，並參考各群 IPC 及特徵值的分佈，找出 MRAM 所包含的技術類別，取代過往的人工歸類方式，並根據技術分類的結果，以專利指標對 MRAM 所包含之技術類別做進一步的分析。對技術而言，本研究考量其投入時間為相當重要的衡量指標，而 CHI 所提出的九項專利指標中，沒有能衡量技術投入時間的指標，因此本研究在技術分析的部份，加入技術活動年限來衡量各技術類別的投入時間。此外，當前影響指標、技術強度及科學強度係以公司立場所訂定的衡量指標，故沒有將上述指標納入技術分析的部份。因此本研究係以專利件數、專利成長率、專利被引用數、技術影響力、技術生命週期、科學連結及技術活動年限七個專利指標為主，各指標的說明如下。

- 1) 專利件數：一技術類別的專利件數，為衡量技術的研發產出狀況。
- 2) 專利成長率：係將今年所獲得的專利與前一年獲得的專利相比較，計算今年較前年專利件數增減的幅度，主要用來衡量技術活動的變化。

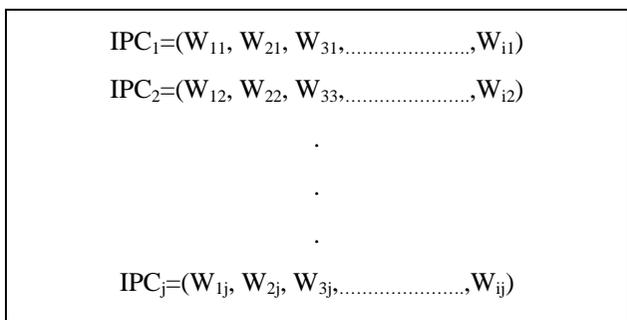


圖8 IPC向量模式

- 3) 專利被引用數：為一技術類別的專利被往後專利引用的平均次數，為衡量技術的重要性及關鍵性。
- 4) 技術影響力：技術影響力是指在過去五年內，一技術類別的專利被目前專利引用的平均次數，相對於整體專利被引用的平均次數，為衡量技術的影響力及重要性。
- 5) 技術生命週期⁸：為一技術類別的專利，引用先前專利之平均年齡中位數，主要用來衡量技術的研發速度及替換時間。
- 6) 科學連結：為一技術類別的專利引用到科學研究文獻的次數，為衡量技術與科學研究的關係。
- 7) 技術活動年限：係以本年為基期，將各技術類別中申請第一件專利的年份，減去基期所得之值，為衡量技術的研發投入時間。

4. 研究結果

4.1 MRAM 的技術類別

本研究由 USPTO 蒐集到 853 件關於 MRAM 技術的專利文件，其中專利文件共分佈在 133 個 IPC，圖 9 為近十年來 MRAM 技術的專利成長情形，第一件關於 MRAM 技術的專利係由 Xeron 於 1979 年所申請，由此圖得知自 1998 年 Baibich 等人發現 GMR 後，之後幾年的專利有顯著增加的趨勢，且目前還在持續的增加當中。由於 MRAM 相關技術類別的發展情況，無法由整體專利成長趨勢得知，因此需確定 MRAM 所包含的技術類別，進而對各技術類別加以分析及探討。

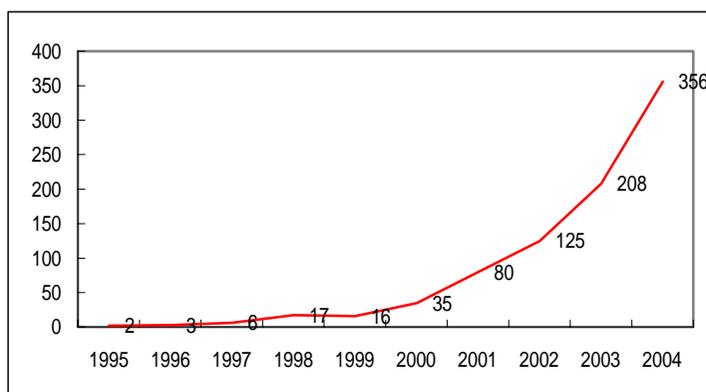


圖 9 近十年的專利件數

⁸ 假如有一專利在 1994 年被核准，其引用的專利有三件，這三件專利的公告年份分別為 1986、1988 及 1990 年，故這三件專利的中位數為 1988，所以此專利的技術生命週期即為 6 年 (1994-1988=6)。

本研究係以文件探勘的分群技術，首先將停用字剔除，接著以 TextAnalyst 軟體對全部專利文件進行特徵值的擷取，根據式(3)計算各個特徵值的 Entropy 如圖 10 所示， $P(x)$ 係介於 0~1 之間，當特徵值出現頻率過高及過低其 Entropy 愈小，當特徵值均勻的分佈在各 IPC 中其 Entropy 值愈大，之後將特徵值依 Entropy 由高至低排序。在特徵值的選擇上，本研究選擇排名前 100 名的字詞，並以人工剔除無意義的字詞如 *et al.*，最後選擇的特徵值結果如附錄所示，作為 MRAM 技術之專利文件的特徵集。

本研究係以相同 IPC 的專利文件為一衡量單位，找出各個 IPC 中特徵值平均出現頻率，並以 TF-IDF 計算各特徵值在 IPC 中的權重。在分群的部份，本研究係以分群成效較佳的華德法對 133 個 IPC 作分群，群數的選擇係由凝聚係數的變化幅度來判斷，當凝聚係數增幅變大時，表示組內的誤差增加，應馬上停止凝聚的程序。在實務的應用上，最適群組數大多介於 3~10 群之間 (周文賢，2002)，由表 1 得知當群組由 5 群變為 4 群時，凝聚係數增加的幅度最高，故本研究決定選擇 5 個群組數，進行 MRAM 技術類別的分類及命名，最後各群 IPC 的數目依序為 85、25、2、19 與 2 個。

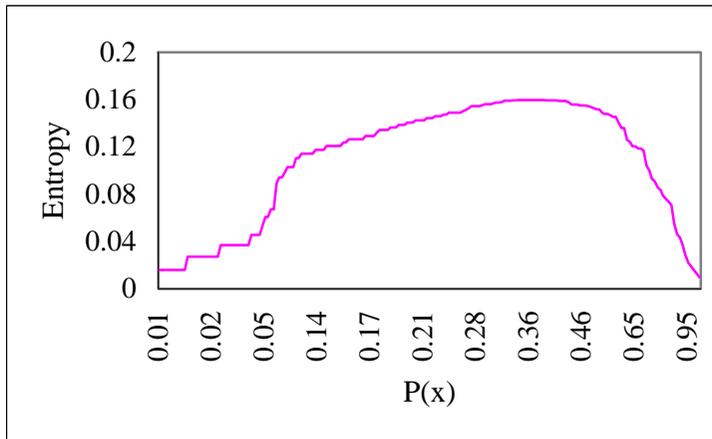


圖 10 Entropy 計算結果

表 1 凝聚係數表

群數	係數遞增量	增加幅度	群數	係數遞增量	增加幅度
3	413.7705	0.0367	7	347.2324	0.0355
4	407.3672	0.0375	8	333.8545	0.0353
5	378.6143	0.0361	9	319.668	0.035
6	347.2617	0.0343	10	280.3105	0.0317

本研究在 MRAM 技術類別的命名原則，係以擁有最多專利件數的 IPC 為主，並參考各群出現次數最多的前十個特徵值，確定 MRAM 所包含的技術類別，如 TF1 大部份的專利集中在 G06F 的 IPC，其所代表的技術類別為電子數位資料的處理技術，並參考特徵值的分佈如 system、device 與 information 等，對此群組進行技術類別的命名。最後的命名結果如表 2 所示，顯示 MRAM 技術共包含了「數位資料的處理及元件」、「資料儲存的原理及設備」、「元件材質的分析及選擇」、「磁感測存取資料的原理」及「資料的傳遞技術」五種技術類別 (Technological Field, TF)。

4.2 MRAM 技術類別的分析

確定 MRAM 技術所包含的技術類別後，接著以七個專利指標，對各技術類別做進一步的分析如表 3 所示，各技術類別的表現說明如下。

- (1) 數位資料的處理及元件：TF1 在 1979 年便開始申請第一件的專利，為 MRAM 的技術類別中最早投入的技術，且其成長率的表現也是最高的，顯示此技術仍在持續的發展當中。但由技術生命週期及科學連結來看，TF1 為 MRAM 的技術類別中次高的，顯示此技術創新及研發速度較慢，需要許多先前科學研究的成果及發現，才能發展出來的技術，故技術的替代時間長，可能為 MRAM 的基礎及重要技術。
- (2) 資料儲存的原理及設備：TF2 的技術生命週期為 MRAM 相關技術中最低的，顯示其技術的創新及替代速度較快。TF2 的第一件申請的專利是在 1993 年開始，而由專利件數、技術影響力、專利成長率及科學連結得知，此技術仍然有專利上的產出，即有許多家公司仍然有投入 TF2 的研發，顯示此技術在 MRAM 技術領域裡具有一定的重要性，甚至影響其他相關技術的發展。

表 2 各群 IPC 及特徵值分佈

TF	IPC	特徵值
1 數位資料的處理及元件	G06F 013 G06F 009 G06F 017	system、memory、device、method、information、logic、communications、plurality、connection、application
2 資料儲存的原理及設備	G11B 005 H01L 031 H01L 029	material、memory、device、magnetizations、method、current、direction、substrate、thickness、regions
3 元件材質的分析及選擇	G01N 027 H01L 047	Pt、current、electrode、material、system、density、Ta、diffusion、substrate、antiferromagnetic
4 磁感測存取資料的原理	G11C 017 G11C 011 G11C 005	memory、current、voltage、devices、direction、plurality、magnetization、output、amplifier、storage
5 資料的傳遞技術	H04B 017 H04Q 007	nodes、system、devices、measurement、analyzer、information、threshold、configuration、dialog、communications

表 3 專利指標分析結果

專利指標	TF1	TF2	TF3	TF4	TF5
專利件數	226	228	3	394	2
專利成長率	0.97	0.49	-0.50	0.84	0
專利被引用數	1.77	5.12	2.67	5.9	0.50
技術影響力	0.56	1.09	1.75	1.07	0.33
技術生命週期	7.35	5.83	12.83	5.9	6
科學連結	3.06	2.33	3.67	1.05	0
技術活動年限	26	12	6	13	6

- (3) 元件材質的分析及選擇：TF3 為近幾年來才有成果的技術，係由於 MRAM 技術的突破是在 1998 年，故 1999 年才始有 TF3 專利的申請。TF3 的技術影響力、科學連結及技術生命週期為 MRAM 五個技術類別中最高的，其中技術力影響力愈高，代表 TF3 的專利在近五年來一直被廣泛地所引用，顯示其為相當重要的技術。科學連結愈高，代表 TF3 係奠基於相當多的基礎科學文獻，才能開發出來的技術。技術生命週期愈高，代表 TF3 係從很早前就有投入研究，然專利件數僅有三件且專利成長率為負的，顯示此技術的研發難度較高，需要不斷的嘗試及研究，為 MRAM 技術領域裡最重要的發現及突破。
- (4) 磁感測存取資料的原理：TF4 的專利件數為 MRAM 相關技術類別中最多的，為最多家公司投入研發的技術，雖然在 1992 年就開始有專利的產出，然專利成長率僅次 TF2，顯示此技術在這幾年來仍在不斷的發展當中，為各公司研發的重心之一。而 TF4 的專利引用次數最高，顯示其為 MRAM 技術裡最重要及最關鍵的技術。
- (5) 資料的傳遞技術：TF5 與 TF3 同為 1999 年才開始第一件的申請專利案件，由各項專利指標皆沒有明顯的表現來看，其可能有二種原因：第一種為此技術類別的專利件數過少，因此在各項指標的表現較差。第二種為此技術類別的研發難度過高，目前可能還處於摸索的階段，因此仍無法看見其立即的成果。此外，本研究比較 TF1 與 TF5 兩者的 IPC 及特徵值發現，TF5 可能為 TF1 引申出來的技術，然目前研發重心還在 TF1，因此 TF5 無法吸引眾多公司的注意，故各公司的投入並不積極，造成專利成長率、專利被引用數及技術影響力皆沒有明顯的表現。

4.3 小結

本研究在技術分析的部份，係以專利指標的觀點來說明 MRAM 相關技術的發展狀況，目前在 MRAM 技術的研發係以 TF1 及 TF4 為主流，雖然 TF1 為最早開始投入的技術，然近幾年還在持續的發展當中，而 TF4 的專利件數最多且為重要的關鍵技術，故自 1998 年便開始吸引多

家公司投入研發。在其他 MRAM 相關的技術類別中，研發創新速度較快且易被取代的技術為 TF2，故以此技術為研發重心的公司，必需不斷的投入研發，以保持本身在此技術領域的地位。TF3 由於是 MRAM 技術的新發現，對日後發展 MRAM 技術的影響力最高，然此技術的進入障礙高，目前僅有少數幾家公司擁有 TF3 的技術。TF5 在專利指標中的表現最差，仍有待更多的公司投入此技術類別的開發。

5. 結論

文件探勘的優點為能同時處理大量的文件資料，並從中萃取所需要的知識及資訊，並能對大量的文件進行分類、摘要處理及分群等工作，然卻鮮少應用於專利文件上，係由於專利文件數量眾多，且技術與法律的專有字詞並存，目前仍有許多問題待解決。過往使用專利資料進行某技術的分析，係以專家根據專利分類號的分佈，判斷一技術所包含的技術類別，其分析的過程不僅耗時且耗力，更可能會無法保有客觀的標準。因此本研究嘗試將文件探勘的分群技術，依據各 IPC 彼此間在技術內容的相似度，找出 MRAM 所包含的技術類別，不僅省去人工審閱的時間及成本，並且也提高分類的客觀性及一致性。將來如能進一步的將分類技術應用於專利文件上，做為專利審查員的輔助工具，必能節省更多的人力及審查時間。

對實務上而言，當公司想跨足一技術領域的市場，可以藉由該技術之專利分析，來瞭解目前技術的發展現況，以及各技術類別之競爭廠商的分佈，選擇進入門檻較低且對公司有利的技術投入研發。甚至以代工為主要的公司，可以藉由技術分析的結果，瞭解整體技術的發展流程及廠商定位，對於公司日後欲向上或向下發展，能給予重要的參考資訊。在原技術領域的公司也能藉由技術分析，規劃公司未來的研發計劃、瞭解各競爭者的研發動向及監測技術的變化。最後對於未來後續的研究提出下列四點建議。

- (1) 技術的分群、選擇及分析：由於 IPC 為全世界廣泛採用的專利分類制度，故本研究在技術分群的部份，係將具有相同 IPC 的專利文件視為一分析單位，以找出 MRAM 所包含的技術類別。未來如能以各別專利文件進行技術的歸類，必能提供更精確的分析結果。在技術選擇的部份，由於 MRAM 為一新興技術，因此目前的專利文件並不多，將來可嘗試以較成熟且擁有較多專利的技術，進行專利文件的分類及技術的分析。此外，在技術分析的部份，如能結合該技術領域的專家意見，必能提供更完整的技術分析結果。
- (2) 欄位的選擇：本研究係以整份專利文件進行特徵值的擷取，未來可嘗試選擇個別不同的欄位，如專利文件的標題 (Title)、摘要 (Abstract)、專利保護範圍 (Claims) 及技術的描述 (Description) 等特定欄位進行特徵值的擷取，並比較不同的分析結果，找出適合專利文件的欄位分析組合，以增進分析的成效。
- (3) 特徵值的選擇：本研究係以 USPTO 所定義的停用字，輔以軟體進行特徵值的擷取，並以

Entropy 進行特徵值的選擇，然最後的特徵集仍會出現無意義的特徵值，故仍需以人工剔除無意義的特徵值。未來如能輔以該技術領域的專家，對所選擇的特徵集進行最後的刪選及確認工作，並與研究員共同建立完整的停用字集與特徵集，必能提供更深入的分析結果。

- (4) 特徵值的擷取：在特徵值擷取的部份，由於 TextAnalyst 軟體所擷取的字詞會產生二個字詞以上的特徵值，而透過本研究的實驗發現，非單一字詞之特徵值的出現頻率很低，故在特徵值選擇的部份，應將非單一字詞的關鍵值刪除，以提高分析的速度及維持分析結果的有效性。

附錄：MRAM 特徵集

Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature
1	density	31	operator	61	series	91	magnet
2	presence	32	circuitry	62	height	92	MOS
3	version	33	driver	63	interactions	93	polarity
4	improvement	34	potentials	64	matrix	94	Pt
5	substrate	35	magnetization	65	development	95	CVD
6	magnitude	36	capability	66	priority	96	DSL
7	analysis	37	sequence	67	adjustment	97	integrator
8	equation	38	voltage	68	deposition	98	Mn
9	interval	39	converter	69	formation	99	nitrides
10	resolution	40	failure	70	CMOS	100	oxidation
11	domains	41	strength	71	comparator		
12	office	42	transmitter	72	Cu		
13	amplifier	43	relationship	73	termination		
14	basis	44	detection	74	anisotropy		
15	equipment	45	ratio	75	session		
16	threshold	46	reliability	76	stability		
17	energy	47	selection	77	region		
18	node	48	Transition	78	ion		
19	relation	49	diode	79	magnetoresistance		
20	limitations	50	IC	80	Ta		
21	distribution	51	parameters	81	analyzer		
22	measurements	52	activation	82	electrode		
23	property	53	categories	83	Si		
24	carrier	54	IR	84	characteristics		
25	width	55	magneto	85	variety		
26	database	56	magnetoresistive	86	diffusion		
27	thickness	57	CA	87	polarization		
28	generator	58	materials	88	television		
29	consumption	59	decoders	89	translation		
30	fabrication	60	modulation	90	verification		

參考文獻

- 周文賢，多變量統計分析: SAS/STAT使用方法，台北：智勝書局，民國91年。
- 孟憲鈺、黃明居、張東森、郭光輝，「產業創新指標：專利引用分析與專利指標」，科技管理學刊，第五卷第一期，民國89年，31-49頁。
- 林傑斌、劉明德、陳湘，資料採掘與OLAP理論與實務，台北：文魁書局，民國91年。
- 曾元顯，「專利文件之知識探勘：技術與挑戰」，現代資訊組織與檢索研討會，台北：淡江大學，民國93年，111-123頁。
- 曾元顯，「關鍵詞自動擷取技術之探討」，中國圖書館學會會訊，第五卷第三期，民國86年，26-29頁。
- 黃文儀，專利實務，台北：三民書局，民國91年。
- 葉林秀、李佳謀、徐明豐、吳德和，「磁阻式隨機存取記憶技術的發展-現在與未來」，物理雙月刊，第26卷第四期，民國93年，607-619頁。
- 詹前疆，「微機電系統」，產業調查與技術，第138期，民國90年，87-102頁。
- 趙晉枚、蔡坤財、周慧芳、謝銘洋、張凱娜，智慧財產權入門，台北：元照書局，民國91年。
- 賴奎魁、鄭伶如、張智翔，「運用專利資料探討接觸式影像感測器業技術發展趨勢之研究」，科技管理學刊，第七卷第一期，民國91年，117-136頁。
- Aas, K. and Eikvil, L., "Text Categorization: A Survey," Technical Report 941, Norwegian Computing Center, 1999.
- Acs, Z. J., Anselin, L. and Varga, A., "Patents and innovation counts as measures of regional production of new knowledge," *Research policy*, Vol. 31, No. 7, 2002, pp. 1069-1085.
- Aggrawal, C. C., and Yu, P. S., "Finding Generalized Projected Clusters in High Dimensional Spaces," In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, Dallas, New York: ACM Press, 2000, pp. 70-81.
- Ashton, W. B. and Sen, R. K., "Using Patent Information in Technology Business Planning-I," *Research Technology Management*, Vol. 31, No. 6, 1988, pp. 42-46.
- Baibich, M. N., Broto, J. M., Fert, A., Nguyen V. D. F., Petroff, F., Etienne, P. Creuzet, G., Friederich, A. and Chazelas, J., "Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices," *Physical Rrvies Letters*, Vol. 61, 1998, pp. 2472-2475.
- Brockhoff, K., "Instruments for Patent Data Analyses in Business Firms", *Technovation*, Vol. 12, No. 1, 1992, pp. 41-58.
- Campbell, R. S., "Patent Trends as a Technological Forecasting Tool," *World Patent Information*, Vol. 5, No. 3, 1983, pp. 137-143.

- Ditizio, R., Werbaneth, P., Corporation, T. and Zhu, J. G., "Cell Shape and Patterning Considerations for Magnetic Random Access Memory (MRAM) Fabrication," *Semiconductor Manufacturing*, Vol. 5, No. 1, 2004, pp. 90-96.
- Dubes, R. C. and Jain, A. K., *Algorithms for Clustering Data*, Englewood Cliffs, NJ: Prentice Hall, 1988.
- Ernst, H., "Patent Information for Strategic Technology Management. *World Patent Information*, Vol. 25, No. 3, 2003, pp. 233-242.
- Fattori, M., Pedrazzi, G. and Turra, R., "Text Mining to Patent Mapping: A Practical Business Case," *World Patent Information*, Vol. 25, No. 4, 2003, pp. 335-342.
- Feldman, R. and Dagan, I., "Knowledge Discovery in Textual Databases (KDT)," In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, Montreal, CA: AAAI Press, 1995, pp. 112~117.
- Frakes, W. B. and Baezay, R., *Information Retrieval: Data Structures and Algorithms*, Englewood Cliffs, NJ: Prentice-Hall, 1992.
- Griffith, A., Luckhurst, H. C. and Willet, P., "Using Inter-Document Similarity Information in Document Retrieval Systems," *Journal of the American Society for Information Sciences*, Vol. 37, No. 1, 1986, pp. 3-11.
- Gupta, V. K., "Technological Trends in the Area of Fullerenes Using Bibliometric Analysis of Patents," *Scientometrics*, Vol. 44, No. 1, 1999, pp. 17-31.
- Jung, S. and Imm, K. Y., "The Patent Activities of Korea and Taiwan: a Comparative Case of Patent Statistics," *World Patent Information*, Vol. 24, No. 4, 2002, pp. 303-311.
- Kim, T., Kim, Y. K. and Park, W., "Technological Issues for High-Density MRAM Development," *Journal of Magnetism and Magnetic Materials*, Vol. 282, 2004, pp. 232-236.
- Liu, T., Liu, S. and Chen, Z., "An Evaluation on Feature Selection for Text Clustering," In *Proceedings of the Twentieth International Conference on Machine Learning*, Washington, CA: AAAI Press, 2003, pp. 488-495.
- Martino, J. P., *Technological forecasting for decision making*, 3rd ed., New York: McGraw-Hill. 1993.
- Michael, P., "Patent Trend Analysis: Incorporate Current Year Data," *World Patent Information*, Vol. 19, No. 4, 1997, pp. 243-249.
- Mogee, M. E., "Using Patent Data for Technology Analysis and Planning," *Research Technology Management*, Vol. 34, No. 4, 1991, pp. 43-49.
- Pavitt, K., "Uses and Abuses of Patent Statistics," In A. F. J. van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology*, Amsterdam: Elsevier Science, 1988, pp. 509-536.

- Pilkington, A., Romano, D. and Omid, T., "The Electric Vehicle: Patent Data as Indicators of Technological Development," *World Patent Information*, Vol. 24, No. 1, 2002, pp. 5-12.
- Salton, G. and Buckley, C., "Term Weighting Approaches in Automatic Information Retrieval," *Journal of Information Proceeding and Management*, Vol. 24, No. 5, 1988, pp. 513-524.
- Salton, G., *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, New York: Addison-Wesley, 1989.
- Salton, G., Wong, A. and Yang, C. S., "A Vector Space Model for Automatic Indexing," *Communications of the ACM*, Vol. 18, No. 11, 1975, pp. 613-620.
- Shannon, C. E., "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, 1948, pp. 379-423, 623-656.
- Steinbach, M., Karypis, G. and Kumar, V., "A Comparison of Document Clustering Techniques," Technical Report 00-034, Computer Science and Engineering, University of Minnesota, 2000.
- Sullivan, D., *Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing and Sales*, New York: John Wiley & Sons, 2001.
- Ward, J. H., "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, Vol. 58, No. 301, 1963, pp. 236-244.
- Wei, C. P., Hu P. J. and Dong, Y. X., "Managing Document Categories in E-Commerce Environments: an Evolution-Based Approach," *European Journal of Information Systems*, Vol. 11, No. 3, 2002, pp. 208-222.
- Yoon, B. and Park, Y., "A Text-Mining-Based Patent Network: Analytical Tool for High-Technology Trend," *The Journal of High Technology Management Research*, Vol. 15, No. 1, 2004, pp. 37-50.