東海大學資訊管理研究所

碩士學位論文

從研討會論文徵稿關鍵字了解學術發展趨勢

-以計算機領域為例

Understanding the Trend of Academic Researches from Call for Papers – Taking the Domain of Computer Science as Example

指導教授：林正偉 博士

研 究 生：林祐陞 撰

中華民國 104 年 7 月

東海大學資訊管理學系碩士學位

考試委員審定書

資訊管理學系研究所 林祐陞 君所提之論文

從研討會論文徵稿關鍵字了解學術發展趨勢

-以計算機領域為例

經本考試委員會審查，符合碩士資格標準。

學位考試委員會 召集人：_____ (簽章)

委　員：_____

_____

_____

_____

中 華 民 國　104　年　7　月　2　日

誌　　謝

總算抵達寫這一段話的時刻了，回想起研究生活，好像都只剩下愉快的回憶，人總會自動忽略痛苦的記憶。當初金山老師在研究方法提到，曾有研究生報完一篇 paper 就會跑去吃燒肉，想想這好像就是我在這期間體重一直上升的原因。

在這學生生涯的最後，有許多需要感謝的人，感謝家人這段時間的支持，感謝正偉老師不厭其煩的指導我，以及各位口委在口試給予的建議與指正。謝謝各位學長，剛進 lab 時指導我的阿鑫、承穎哥、小白，感謝在 lab 一起宅的同學，育瑋、蓬蓬、耀民與肯德基(超帥)，還有一起八卦的艾波，旻臻。也謝謝在這段時間一直幫我的學弟妹，侑侑、允妍、會計哥、炯文，進皓和爾廷。還要感謝我的搭檔凱元大大，這期間忍受我的吐槽，並阻止我在誌謝放上三熊圖。最後再次的讓我說，謝謝你們!!

論文名稱：從研討會論文徵稿關鍵字了解學術發展趨勢-以計算機領域為例

校所名稱：東海大學資訊管理學系研究所

畢業時間：104 年 7 月

研究生：林祐陞　　指導教授：林正偉

論文摘要：對於研究人員來說，了解研究發展趨勢是很重要的。有許多系統可以協助我們從已經刊登的論文著手，也有許多研究報告討論這種方式的優缺點。本研究提出另一種觀察策略，從研討會徵稿公告(Call For Papers)著手，觀察未來幾年學術研究的發展趨勢，並可作為擬定研究方向的參考。以計算機領域的國際研討會為例，我們蒐集了 248 個研討會的 Call For Papers，整理出 12419 個關鍵字，輔以 ACM (Association for Computing Machinery)計算機分類系統進行關鍵字的分析。最後，使用視覺化的方式幫助我們解讀這些資料。實驗結果顯示，這種觀察策略有助於研究人員了解未來的研究發展趨勢。

關鍵詞: 文字探勘、研究趨勢分析、資料視覺化

Title of Thesis：Understanding the Trend of Academic Researches from Call for Papers – Taking the Domain of Computer Science as Example
Name of Institute: Tunghai University, Graduate Institute of Information Management
Graduation Time：07/2015
Student Name：You-Sheng Lin Advisor Name： Jeng-Wei Lin

Abstract：It is always important for researchers to understand the trend of academic researches. A wildly adopted strategy is to analyze published articles. Many systems had been developed to help researchers doing this and a lot of studies on this strategy had been reported. In this paper, we proposed a new strategy: analyzing call for papers (CFP) of academic conferences. In the experiment, CFPs of 248 international conferences on computer science were collected, and 12,419 keywords were extracted. The keywords were further structured according to the 2012 ACM (Association for Computing Machinery) computing classification system. Visualization technologies were used to monitor the change of keyword frequency in different views. The experiment result had shown that our system can help researchers understanding the trend of computer science research in the near future.

Keywords：Text Mining, Research Trend Analysis, Data Visualization

目　　錄

圖　次

圖　　次

表　　次

第一章 緒論

1.1 研究背景與動機

對於研究人員來說，了解研究發展趨勢是很重要的。一種常見的方式是從已經刊登的論文著手，有許多系統可以協助我們，如 SCI (Science Citation Index)索引資料庫等等，也有許多研究報告討論這種方式的優缺點。這種策略的中心思想是從已發生過的事情來推敲未來的發展方向。

近年來隨著資訊進步，知識的更新與學科發展越趨龐大。從全國博碩士論文書目資料收錄範圍(2015)的資料顯示 100 學年度收錄了 63539 筆、101 學年度收錄 63186 筆、102 學年度收錄 60530 筆，單就每年在台灣的碩博士論文收錄成長就如此龐大，IEEE 的資料庫數位圖書館的收錄更超過 360 萬項目(IEEE Xplore digital Library,2015)。每一日都有新的論文發表，學科不斷的增長加上各種跨領域的結合，研究人員在研究方向的選擇上成為一個問題。研究人員在尋找研究資訊多透過電子資料庫進行索引，在數量龐大的資料庫尋找研究方向是廢時的。

為了解決此問題，出現了許多相關研究。除了資訊檢索(Information Retrieval)的發展外，趨勢分析與主題預測的研究也不斷出現。從過去趨勢分析的研究中，學術趨勢多以研討會與期刊文獻做為其主題探測及趨勢追蹤。如林宜瑩（2010）在期刊使用老化理論的概念，探討特定領域的研究主題趨勢及消長。王宏德(2013)從臺灣博碩士論文知識加值系統內，收錄的碩博士論文探討學術研究趨勢。Cai & Card(2008)使用 ACM (Association for Computing Machinery)的分類系統，將總數 691 篇，屬於軟體工程領域的頂級期刊論文與研討會論文進行主題分類。一些研究也針對研討會文獻與期刊文獻的先後關係觀察對趨勢預測的影響(許育聞,2008;王京盛,2012)。Tu & Seng(2009)在關於資訊檢索的研究中，探討研討會與期刊在時間上的關係性，從前後領導檢測新的發展趨勢。相關的研究從單一資料集的文獻探討，發展到期刊與研討會的趨勢分析，到觀察期刊與研討會的期刊關係性，其目標都是著重在趨勢發展的時間上，這也說明了時間選擇對於趨勢發展的重要性。

1.2 研究目的

圖 1-1 為研究的時間軸，一個研究的過程會歷經研究方向的選擇，接著發表在研討會，最後刊登至期刊上。過去的研究探討中，研討會文獻與期刊文獻皆是研究的成果發表，從最早的研究方向選擇到期刊的刊登也許會歷時二至三年以上。

過去的研究觀察，我們可以從已發表的研討會文獻預測未來的期刊主題趨勢。在此遇到的問題是，已發表的研討會文獻或者期刊文獻，在時間上皆為過去的研究，等同使用過去數年的研究預測未來趨勢。

本研究提出另一種策略，透過觀察學術研討會的論文徵稿文件的變化，探討研討的發展趨勢。參加學術研討會是研究員發表研究成果、交換心得、進行討論的重要學術活動之一。為了讓討論的主題能夠聚焦，吸引具有共同研究興趣的人員參與，不同領域的研討會一般都會設定各自關注的焦點。此外，研討會的徵稿方向也會隨著當前的熱門研究議題變動。此一策略與第一種策略最大的不同點在於觀察的時間點，研討會的徵稿文件多半是在會議舉辦前半年、一年、甚至更早就張貼，因此，它代表著主辦學術會議的議程委員們對於未來半年、一至二年內研究議題的預測，是目前可能正在進行的研究，而已經發表的論文呈現的是過去進行的研究。利用徵稿關鍵字在時間上的優勢進行觀察，可幫助研究人員尋找研究的方向並觀察未來的研究趨勢。科技日新月異，常出現改朝換代的現象，本論文選擇計算機相關領域的研討會作為觀察對象。在此本研究使用研討會論文的徵稿(Call For Papers)關鍵字進行分類，並以資料視覺化方法分析，期望能從近年的研討會徵稿方向觀察出研究領域的興衰。

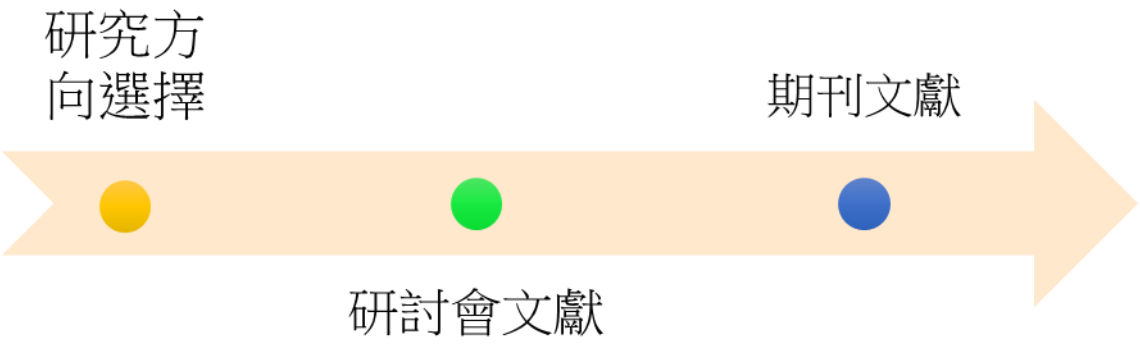

圖 1-1 研究的時間歷程

1.3 研究架構與流程

本研究的流程如圖 1-2 所示，詳述如下:

一、緒論-研究主題與目的

本研究主要提出一新的觀察策略，探討未來研討會的發展趨勢並藉此提供學術研究者一個參考依據，與舊有使用文獻的觀察策略不同，我們的觀察策略提供了一個時間上的優勢，並提出一個方法幫助我們完成此工作。

二、文獻回顧

此章節探討一些過去的做法，從過去的學術趨勢研究比較與我們觀察策略的差異，接著對常見的文字探勘與相關領域的應用做介紹。最後說明視覺化與資料分析的結合。

三、研究方法

首先說明過去的研究與新觀察策略的差異，解答如何找趨勢圖，探討常見技術的可行性與不可行性，接著提出我們的方法並詳述步驟與做法，在做法中定義資料的表達，最後說明如何視覺化。

四、實驗設計與結果

經由上述提出的研究方法，提出實驗的設計與模組。說明資料來源與研究限制，詳述資料的儲存結構。提出一些例子並與 Google Trends 比較，說明研究的觀察現象。

五、結論

分析本研究的實驗結果做出結論，並藉此提出建議和未來研究方向。

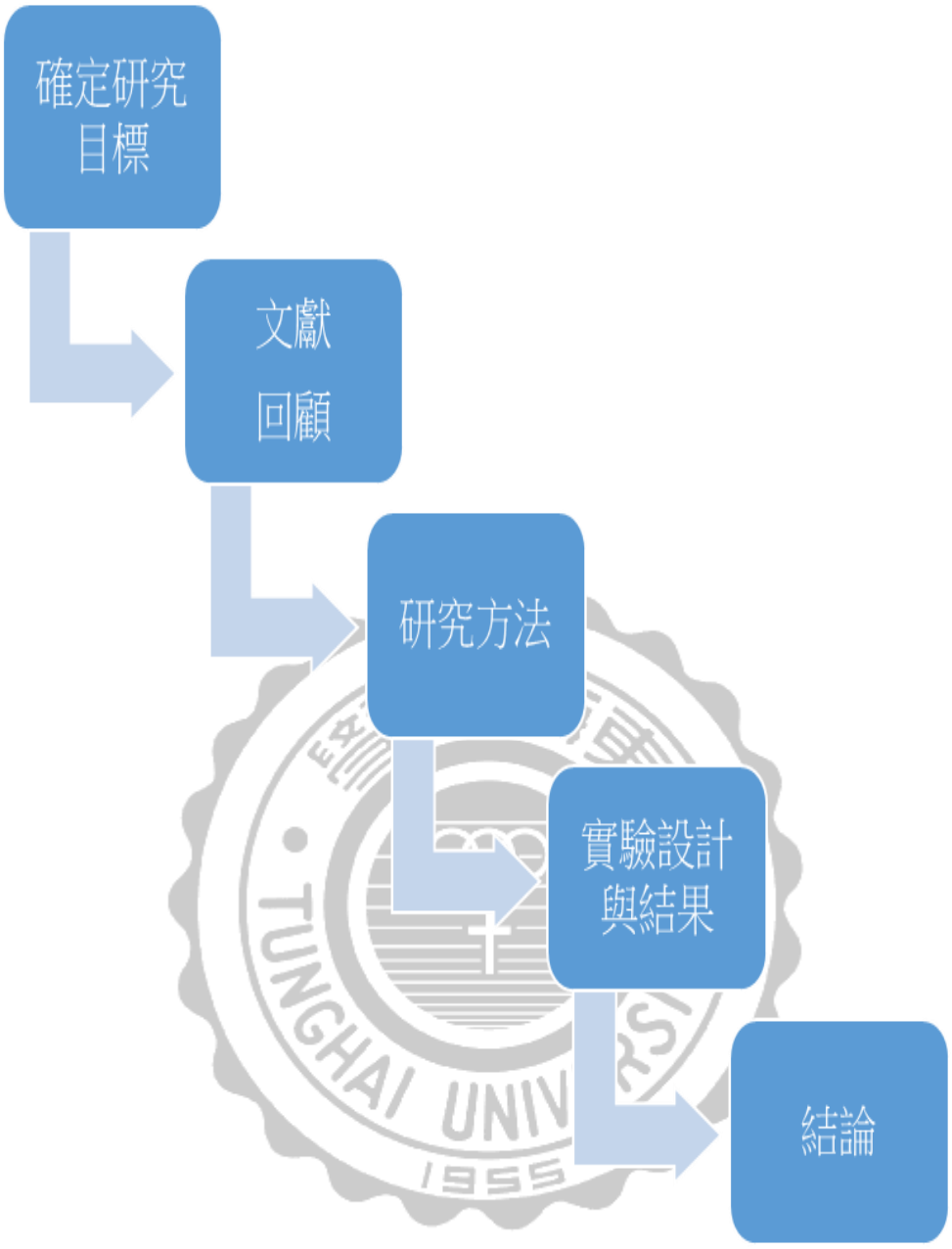

圖 1-2研究流程

第二章 文獻探討

本章節分為四個部份逐步探討，前二部分將幫助我們了解相關領域的研究發展。第一部份是趨勢分析與學術的相關研究，包括其使用的方法；第二部份為文件探勘常見的技術回顧;其中會介紹 ACM 的分類系統；第三部分我們討論視覺化對趨勢分析與資料分析所扮演的角色。最後，小結部分將探討過去的做法與新策略的相關性。

2.1 學術、趨勢分析與相關研究

學術論文一般被分為研討會論文與期刊論文，研討會是特定領域學者的討論會，學者們會基於研討會的核心主題，發表他們的最新研究。鑒於研討會舉辦有其規律，許多會議是定期性的重新召開，論文的審查較為迅速，且研究議題是當下較為新穎與波動的。研討會論文的接受，普遍取決於其主題的新穎性，而不是研究的嚴謹性。相較研討會論文，發表在期刊的論文受到更嚴格的評估程序。研究人員首先需經過總編的評估，隨後必須不斷修改他們的論文，直到符合評估委員的建議，這些委員都是研究領域的專家，整個過程所花費的時間通常超過一年，甚至高達兩至三年或更久。因此,期刊論文缺乏研討會論文的波動性與新穎性(Tu & Seng, 2009)。

許多研究人員首先會在研討會論文發表自己的新研究，接受建議後進一步投稿期刊，再經過一系列的評估流程後被接受並發表於期刊上，Tu & Seng 將會議論文與期刊論文關係細分為四類，如表 2-1 所示:

表 2-1期刊與研討會關係

	領導	
跟隨	會議論文	期刊論文
會議論文(C)	C→ C	J→ C
期刊論文(J)	C→ J	J→ J

資料來源: Tu & Seng, 2009

會議論文與期刊論文之間的關係如表 2-1，其中本次的研討會論文引導研討會論文稱為 C→ C 、本次研討會的論文引導期刊論文稱為 C→ J、該雜誌的論文領導研討會論文稱為 J→ C，該期刊論文引導期刊論文則稱為 J→ J。Tu & Seng 的研究結果表示，會議論文可以幫助研究人員發現新的主題趨勢，該年的研討會論文題目，會影響同年及隨後兩年的期刊論文題目。

研討會論文與期刊論文相關研究中，許育聞(2008)蒐集了資訊檢索領域具代表性的 SIGIR 會議文獻、及五種期刊以資訊檢索為主題之文獻，將其收錄文獻已分類好之關鍵字詞彙與主題，比較趨勢預測的差異，結果發現會議與期刊文獻的詞彙與關注主題有所差異、主題涵蓋範圍越大，整體趨勢走向越為一致、而大部分的主題出現時間上，會議文獻會早於期刊文獻。在王京盛(2012)的研究中，額外提取了論文本身的被引用次數，增加特徵選取的效率，使用會議與期刊的先後影響幫助進行主題的偵測與追蹤，並藉此分析研究主題與趨勢的走向。

綜合上述研究，我們可以確立趨勢變化的時間軸如圖 2-1，圖中，T1 是研究開始的方向選擇，T2 與 T3 分別為研討會文獻與期刊文獻。我們無法預期 T3 出現的時間，但可以從 T2 預期 T3 可能的發展趨勢，而研討會的 CFP 發佈時間更早於 T2 研討會文獻，從圖 2-1 表示，在時間 T1 即可觀測到 T2 與 T3。因此，我們預期使用研討會 CFP 的公告做為趨勢預測，相較研討會文獻將更有時間上的優勢。

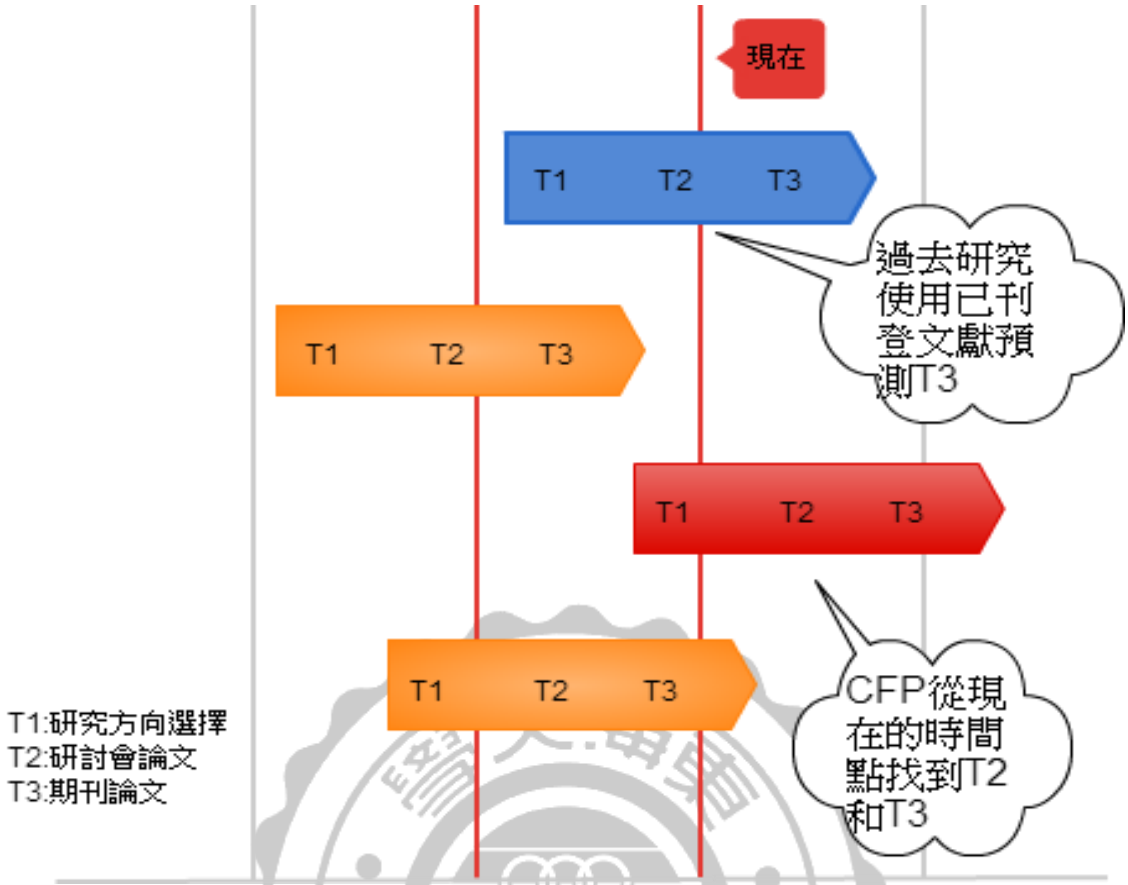

圖 2-1 研究歷程時間軸

在過去的研究可以看出，趨勢分析上需確定方法，如使用分類好的詞彙，選定的期刊範圍等，參考學者陳光華(2010)提出的應用，其提出學術會議資訊檢索與擷取的自動程序，運用學門分類表做為學科主題關鍵字，透過 Google 搜尋，建立一個會議的檢索系統。另一個相關研究，從國際研討會通告網站所發佈的會議名稱、地點、日期等進行擷取，其中使用機器學習的方式幫助擷取資訊(胡妹涵 2005)。

過去的趨勢研究主要使用文獻進行分析，但是從已發表的論文只能看出研究人員過去的研究。因此，本研究選擇研討會的 CFP (Call For Papers)，從各研討會徵稿關鍵字之變動，期望可以看出對趨勢需求的變化。在資訊的擷取上，除了關鍵字的蒐集外，相關的時間、地點等資訊，也能幫助我們對趨勢的解讀。

2.2 文字探勘

圖 2-2 是常見的研討會網頁，左右各為不同的研討會，其網頁公告的格式都不相同，因此是種半結構化（semi-structured）或非結構化（unstructured）的格式來源，如同 Big Data 提及的 4V 的 Variety，如何將這些資料結構化並找出有用的訊息，成為我們在趨勢分析的最大關卡。因此，進行分析文件，從非結構化的文件找出有用的資訊的文字探勘(Simoudis,1996)成為最適合的選擇。如圖 2-3，Tan(1999)提出的文字探勘架構，將非結構化的文件經過萃取轉成中介格式，再知識淬煉出有用的訊息。

圖 2-2 研討會公告網頁

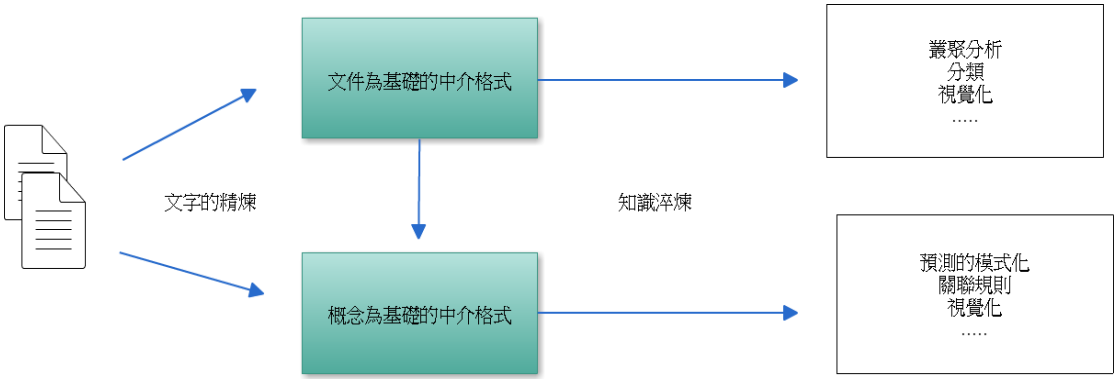

圖 2-3 文字探勘架構

資料來源: Tan(1999)

資料探勘(Data Mining)技術主要用於處理結構化的資料，如規律結構的資料庫或表格。而文字的資料通常是以半結構化（semi-structured）或非結構化（unstructured）的型式儲存，如隨寫的文章，電子郵件，網頁，或者論文文獻等資料類型，為分析這些有用的資料並萃取知識，回顧圖 2-3，文字探勘包含了文字精煉與知識淬煉兩部分，文字精煉先將不定格式轉成半結構或結構化的中介格式，再由中介格式推理出範式或知識(Tan,1999）。因此，文字探勘(Text Mining)的目的即是處理非結構化資料並找出有用資訊。丁一賢(2006)在文字探勘的相關工作列出以下幾項:

- 文字的分類（Text categorization）
- 文件的叢聚（Document clustering）
- 文字規則的探勘（Rule mining from text）
- 文字中探勘概念、分類、關係(Concept/Taxonomy/Relationship mining from text）
- 資訊整合（Information integration）
- 文件結構的探勘與分段（Structuring mining and text segmentation）
- 文件摘要（Text summarization）
- 文件的瀏覽與視覺化（Text navigation and visualization）

隨著巨量資料的發展，適合處理半/非結構化的文字探勘，也越趨重要。在社群網路(Social Networking Service)、資訊檢索(Information Retrieval)與事件偵測領域都有相關的研究。例如 Ozer Ozdikis et al. (2012)將 TF-IDF（term frequency–inverse document frequency）使用在推特(Tweet)的事件偵測，應用文件的相似性和群集演算法提高事件的檢測。以下小節將介紹一些常見的相關技術。

2.2.1 TF-IDF

TF-IDF(term frequency–inverse document frequency)是種常用在文件探勘的技術，使用統計方法評估單字在文件上或文件集合中的重要性。詞彙頻率(term frequency)考量某個詞，在單一文件出現頻率越高，那可以判斷該詞彙在該單一文件的重要性越高 Salton and McGill(1983)。

逆向文件頻率(inverse document frequency)的概念為，若某單詞出現在文件集裡的文件數量越少，則代表該詞彙對文件的代表性越高，一些常見的單詞如「的」、「我」、「是」屬於大多文件都會出現的單詞，則沒有鑑別性。

TF-IDF=詞彙頻率*逆向文件頻率，因此 TF-IDF 的中心概念為一單詞高頻率出現在一文件中，且在其他文件很少出現，那可以認為該一單詞有很好的判別性。TF-IDF 可幫助找出重要的單詞，並過濾掉常見的單詞，在資料庫裡有 N 篇文章，而某個關鍵字出現在一篇文章或多篇文章中，那這關鍵字是否能代表某篇文章或某群文章。

2.2.2 WordNet

WordNet (Christiane Fellbaum 1998)為英文的的詞彙字典，常見於處理語義相關的研究，將名詞、動詞、形容詞與副詞組織成語義字典。WordNet 以同義概念構成關係結構，其中包含了幾種關係性。最常見的語義是從屬關係(hyperonymy)，例如類別 furniture 包含了 bed 與 bunkbed，而 bed 與 bunkbed 又構成了 furniture 類別，其上下位關係是遞移的，上位包含了下位，而下位又屬於上位關係。Meronymy 關係是一種部分與整體(holonyms)關係，例如 brim 與 hat 是一種 holonym 的關係，paper 與 book 是一種 Meronymy 的關係，其他關係如形容詞與動詞同義、反義關係等。

Prantik Bhattacharyya et al. (2011)發表的研究中，透過 WordNet 字典的關係性將關鍵字建立森林模型，構築底層語義關係。概念如圖 2-4，在研究中透過森林的結構，如果關鍵字為同一棵子樹，即計算兩個節點之間的距離，如果關鍵字在不同的子樹則判定關鍵字之間沒有關係性。藉此森林模型計算 FaceBook 使用者在興趣上的相似性。WordNet 在文字處理上，也可使用在字根還原，篩選名詞等。

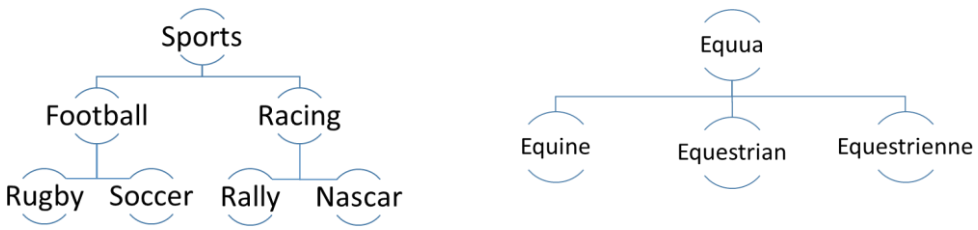

圖 2-4 森林模型範例

2.2.3 本體論(Ontology)與分類系統(Classification System)

本體是指一種形式化的，對於共享概念體系的明確又詳盡的說明 (Studer et al, 1998)，本體論具有一個分類體系。在人工智慧領域與資訊領域裡，本體論是一種詳盡和結構化的概念，它專指資源特徵的集合和它們之間的關係，因此本體論除了提供概念上的資訊，也提供了從屬關係。

為了促進知識共享，學科會發展分類系統，並使用共同的術語進行通訊，通常分類系統會基於主題，採用在社團、日記、系統等被用來做為期刊的術語索引或關鍵字(Iris Vesseya et al. 2005)。分類和本體論的差異在於資訊的豐富度。兩者都提供了觀念的結構或分類的項目。分類用於標示項目，而本體提供許多相關概念的資訊，包括從屬關係。如果運用分類進行資訊的分類，等同將資料放在有標記的盒子，分類著重在建構知識資源，本體則是一種詳盡和結構化的概念，它專指資源特徵的集合和它們之間的關係。

綜合上述本體論與分類系統的定義，共同都有結構化的特性。分類系統可以幫助知識共享與索引，本體論本身俱備描述概念的資訊。在羅濟群等人(2012)的研究中中，提出使用國際疾病分類，建立一個動態本體論(Dynamic Ontology)藉此建立疾病與飲食的關聯，並使用自組織映射圖網路(Self-Organizing Map,SOM)進行推薦。

2.3.4 ACM Computing Classification System

計算機領域的分類系統，Henri Barki et al.(1988)在 Management Information Systems Quarterly (MISQ)發表的研究，其特別處在於分類頂層有參考學科，使用參考學科是新興學科的象徵，information Systems 依賴於更成熟的學科來輔助理論，如管理學、社會學等(Richard L. Baskerville , Michael D. Myers,2002)。

ACM 是最大的計算機科學領域的學術組織，ACM 提供了用戶一個交流資訊和創新的平台。而 ACMCCS(ACM Computing Classification System)是透過許多專家與歷年來的論文所完成，從 1964 的第一版共經歷了 4 個版本，最新版為 2012。圖 2-5 是 ACM 分類系統的目錄，ACM 的分類系統底下有 14 大類，且分類系統是多層次的，如表 2-2 與圖 2-6，從概念節點向下發展成一顆樹狀的分類。ACM 將

此分類應用在他們的圖書系統裡，它依賴語義作為類別和概念，反映了計算機科學藝術的來源。在 ACM 網站的說明裡 (ACM Computing Classification System toc , 2012)，給予 2012 版本分類系統一個定義:

The 2012 ACM Computing Classification System has been developed as a poly-hierarchical ontology that can be utilized in semantic web applications.

ACMCCS 是一種兼具本體論特性的分類系統，在 ACM 網站上提供了 SKOS、HTML 等格式供研究和教育目的。ACM 提供的 SKOS(Simple Knowledge Organization System)格式可以表達詞彙的結構與概念，ACM 提供的格式有幾項特徵，其特徵說明如下:

- rdf:資源描述框架，rdf 做為唯一值標籤，用來代表每個唯一的概念。
- prefLabel:首選標籤，關鍵字的主要名詞，在此與 rdf 同樣是唯一值，在 ACMCCS 網站提供的 HTML 格式裡，顯示的即為首選標籤。
- altLabel:非首選標籤，在概念和語義上為同義，但可能是不同的單字或組合、縮寫，例如 prefLabel 是 network 但 altLabel 則有 data Communication，與 computer network。
- broader:直接上位概念關係，節點在分類結構裡的上層或父母節點。
- narrower:直接下位概念關係，為節點所擁有的下層或子節點。
- inScheme:在概念體系中，這裡的概念體系是 ACM 的分類系統。

表 2-3 是 ACMCCS 的 SKOS 範例，可以看出一個節點只會有一個 rdf 與一個 prefLabel，而 altLabel 非首選會有多個。

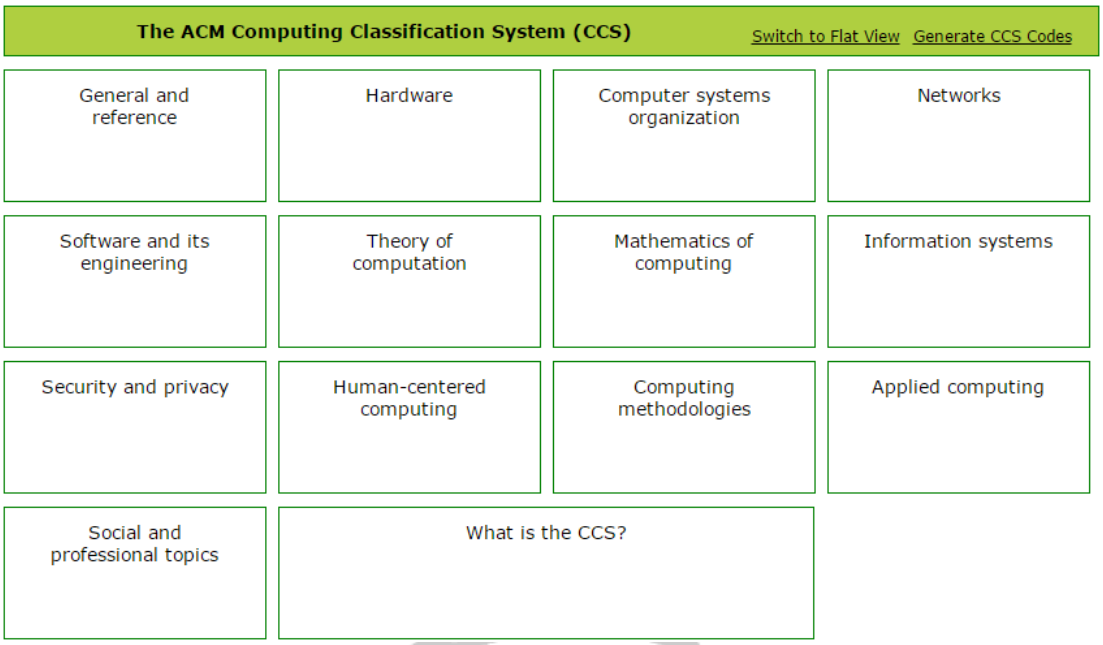

圖 2-5 ACMCCS visual display format

資料來源:Association for Computing Machinery

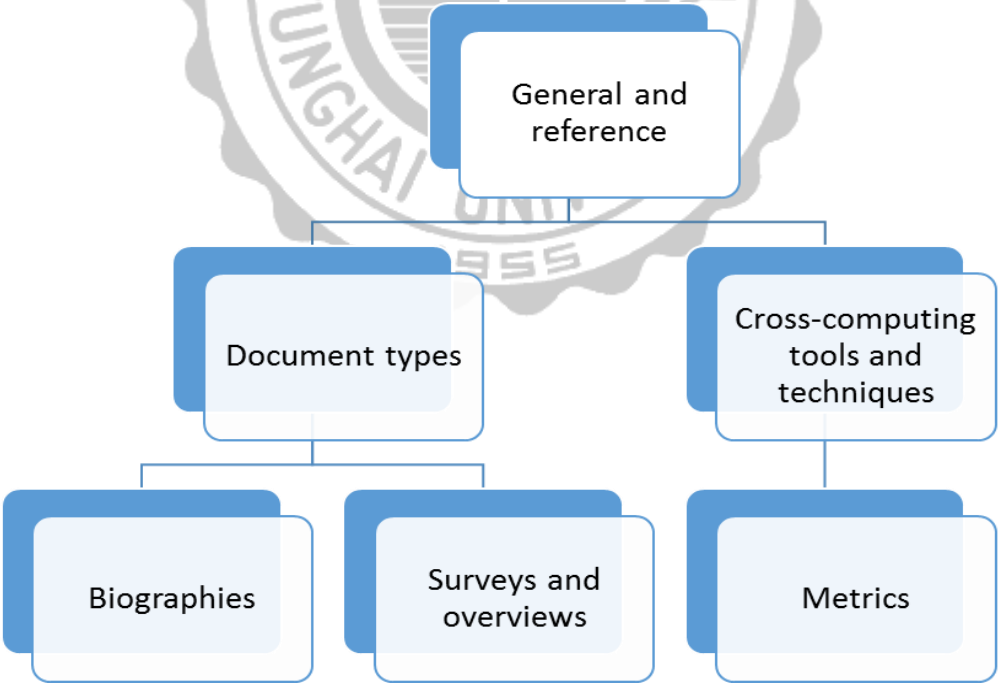

圖 2-6 ACM 分類系統的分層結構

表 2-2 ACM 分類系統樹狀結構

- General and reference
  - Document types
    - Surveys and overviews
    - Reference works
    - General conference proceedings
    - Biographies
    - General literature
    - Computing standards, RFCs and guidelines
  - Cross-computing tools and techniques
    - Reliability
    - Empirical studies
    - Measurement
    - Metrics
    - Evaluation
    - Experimentation
    - Estimation
    - Design
    - Performance
    - Validation
    - Verification

資料來源: Association for Computing Machinery

表 2-3 ACM 分類系統的 SKOS 格式

```
<skos:Concept rdf:about="#10002958" xml:lang="en">
<skos:prefLabel xml:lang="en">Semi-structured data</skos:prefLabel>
<skos:altLabel xml:lang="en">semistructured data</skos:altLabel>
<skos:altLabel xml:lang="en">semi structured data</skos:altLabel>
<skos:altLabel xml:lang="en">semi-structured information</skos:altLabel>
<skos:altLabel xml:lang="en">semi-structured knowledge</skos:altLabel>
<skos:altLabel xml:lang="en">semistructured information</skos:altLabel>
<skos:altLabel xml:lang="en">semi structured information</skos:altLabel>
<skos:altLabel xml:lang="en">semistructured database</skos:altLabel>
<skos:altLabel xml:lang="en">semi structured knowledge</skos:altLabel>
<skos:altLabel xml:lang="en">semistructured knowledge</skos:altLabel>
<skos:altLabel xml:lang="en">semistructured databases</skos:altLabel>
<skos:altLabel xml:lang="en">semi structured databases</skos:altLabel>
<skos:altLabel xml:lang="en">semi-structured database</skos:altLabel>
<skos:altLabel xml:lang="en">semi structured database</skos:altLabel>
<skos:inScheme rdf:resource="http://totem.semedica.com/taxonomy/The
ACM Computing Classification System (CCS)"/>
<skos:broader rdf:resource="#10010820"/>
</skos:Concept>
```

資料來源:Association for Computing Machinery

2.3 Google Trends

Google Trends 是 Google 的一項服務，可以藉著選擇位置與主題探索趨勢與故事，每個趨勢故事都是由 Google 策劃，使用演算法透過搜尋引擎、新聞及 YouTube 三個平台檢測。Google 趨勢也提供探索主題性的趨勢，如工商業、房地產、美容與健身等主題，其它還有搜尋趨勢、熱搜排行榜、YouTube 熱門搜尋等服務(Trends-Google,2015)。其最大優勢是擁有長年來自世界各地的搜尋資料，越多人搜尋的關鍵字或者主題代表越多人感興趣。圖為 Google Trends 的探索主題畫面，在使用 Google 提供的各種趨勢服務時，可以發現資料的視覺化呈現，在各種趨勢觀察裡，主要由折線圖與地理位置作為一個趨勢的呈現方式，這方面 Google 也提供多項關鍵字的同時比較。

2.4 視覺化

電腦的進步助長了視覺化效果，不斷成長的視覺化檢視，我們可以從日常的許多地方看到，如:圖表、報紙、雜誌、工作流程、網站、地圖等。一些資訊的表達可以使用視覺化的呈現，視覺化並不需要對資料進行分析，它是一種溝通語說故事後的結果，如休閒資訊的視覺化即使不分析也可以提供認知、幫助對社會映射的見解(Z. Pousman, J. T Stasko and M. Mateas. 2007)。

視覺化對資料解讀的幫助，G. Judelman(2004)將知識的視覺化從條件、陳述、程序型知識來區分。Luca Masud et al.(2010)指出將視覺化分為分析形視覺化、溝通型視覺化、形成型視覺化可以更方便幫助設計者依據不同資料進行設計。在視覺化的應用，Felix et al.(2005)使用資訊視覺化，做為資訊檢索的介面，找出不同領域的文件之間隱藏的關係。

綜述以上文獻，資料的視覺化可以幫助人們更方便的發現資訊，如何呈現視覺化，如何藉由資料的儲存格式幫助視覺化。一個簡單的例子，部分的社交網路中的友誼與地域無關(David Liben-Nowell et al. 2005)， 如圖 2-7，將研討會與ACMCCS 結合，發現新的訊息。

圖 2-7 視覺化的基本特徵

2.5 小結

在上述的文獻探討裡我們回顧了一些文字探勘的技術與應用，在學術的發展中，找出趨勢線是本研究的重點。圖 2-8，一個關鍵字出現次數，經過多年的變動會出現一條趨勢線，如何找出這條趨勢線，是我們將要解決的問題。TF-IDF 適用全文檢索，但在 CFP 的結構上並不適合。CFP 的資料是半結構化的，有多個關鍵字的組合，一些關鍵字可能在單一研討會只出現一次，卻出現在所有研討會上，而一些較上層概念的關鍵字可能會頻繁在同一研討會出現，且出現在多個研討會上，如表 2-4，左右各為不同的研討會徵稿關鍵字詞頻，列出的是前 10 名最高頻率的關鍵字，Computing 與 System 等常見且在各個組合裡容易出現的關鍵字，這會造成 TF-IDF 的區別度過低。如果取 TF 為特徵值，這部分雖適用於判定該研討會特別注重哪些領域，或幫助研討會索引與分類，但在尋找趨勢線方面是不利的，因此我們捨棄 TF-IDF 的方法。

WordNet 是常見的應用，但語料庫並非學術專門，且大多針對單字做處理，在多個單字的組合上並不適用，我們使用 WordNet 進行測試如圖 2-10，抽 50 個關鍵字進行比對，只有 27 個能找出結果，如關鍵字 Parallel 與 Computing，各自時可以在字庫裡找到，但組合後的 Parallel Computing 在字庫裡並無結果，Prantik Bhattacharyya et al. (2011)利用 WordNet 的字詞關係性建立起樹狀結構並計算相關性，而在 CFP 的情境下卻會失去其特性，但我們參考此概念，圖 2-9 顯示出每個關鍵字的變動底下，是否有較底層的關鍵字在影響大範圍的趨勢走向，圖中橫軸是上層的關鍵字 A~E，Z 軸是個關鍵字底下的節點，在此圖例假設每個關鍵字底下各有 3 個子節點。因此，ACM 分類系統本身的關係性可以幫助我們此項工作。實際的例子 ACM 分類系統的頂層節點 Hardware，假設該類別年年出現大成長趨勢，但真正呈現成長趨勢的是底下的哪個關鍵字，在此藉由 ACM 分類系統的階層特性，期望幫助觀察這類現象。

在此本研究使用 ACM 的分類系統，藉由其本體論的特性幫助我們分類，優勢在於此分類系統有完善的資源描述，且 ACM 分類系統是由多位專家歷時建立而成，是目前最完善的計算機分類系統，我們可以透過將關鍵字分類進 ACM 分類系統裡，藉此將關鍵字結構化以方便計算出關鍵字的出現次數，幫助我們找出一條趨勢線。

表 2-4 CFP 的詞頻表

關鍵字	詞頻	關鍵字	詞頻
Computer	11	Systems	21
Computing	8	Processing	9
systems	7	Engineering	8
data	6	Computing	8
processing	5	Software	7
engineering	4	information	7
networks	3	architecture	6
security	3	modelling	6
digital	3	design	5
distributed	5	data	3

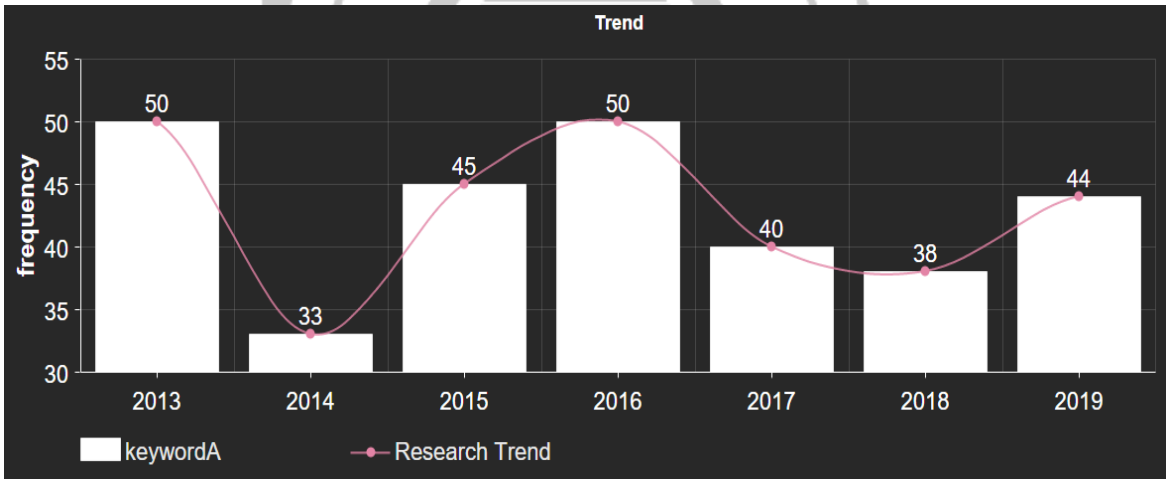

圖 2-8 關鍵字變動所產生的趨勢線

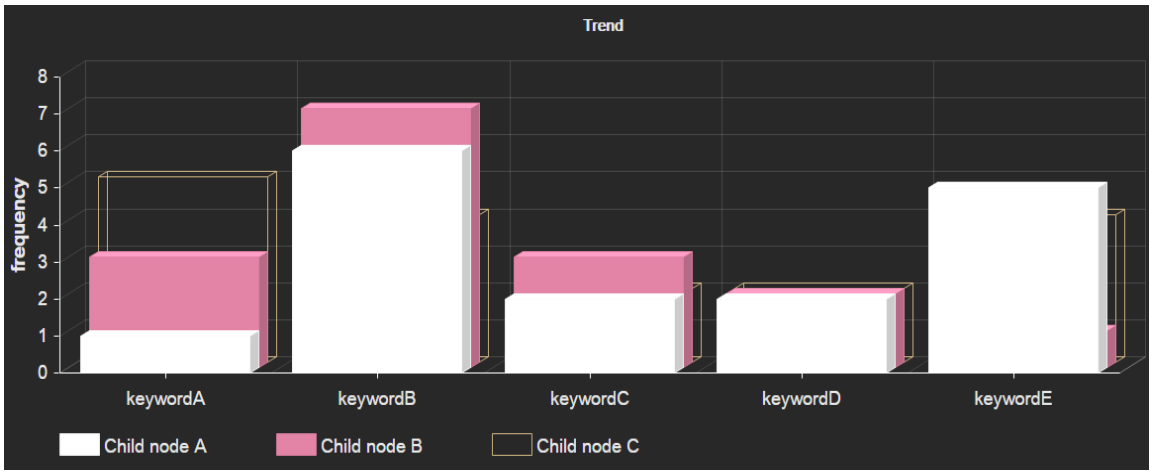

圖 2-9 關鍵字底下子節點的變動

```
[Synset('algorithm.n.01')]
[]
[]
[Synset('computer.n.01'), Synset('calculator.n.01')]
[Synset('architecture.n.01'), Synset('architecture.n.02'), Synset('architecture.n.03'), Synset('computer_architecture.n.02')]
[]
[Synset('real_number.n.01'), Synset('real.n.02'), Synset('real.n.03'), Synset('real.a.01'), Synset('real.a.02'), Synset('real.
s.03'), Synset('real.s.04'), Synset('actual.s.03'), Synset('real.a.06'), Synset('substantial.a.03'), Synset('real.s.08'), Syns
et('veridical.s.01'), Synset('very.r.01')]
[Synset('time.n.01'), Synset('time.n.02'), Synset('time.n.03'), Synset('time.n.04'), Synset('time.n.05'), Synset('time.n.06'),
 Synset('clock_time.n.01'), Synset('fourth_dimension.n.01'), Synset('meter.n.04'), Synset('prison_term.n.01'), Synset('clock.v
.01'), Synset('time.v.02'), Synset('time.v.03'), Synset('time.v.04'), Synset('time.v.05')]
[Synset('system.n.01'), Synset('system.n.02'), Synset('system.n.03'), Synset('system.n.04'), Synset('arrangement.n.03'), Synse
t('system.n.06'), Synset('system.n.07'), Synset('system.n.08'), Synset('organization.n.05')]
[Synset('database.n.01')]
[]
[Synset('data.n.01'), Synset('datum.n.01')]
[Synset('mining.n.01'), Synset('mining.n.02'), Synset('mine.v.01'), Synset('mine.v.02')]
[]
[Synset('reliable.a.01'), Synset('dependable.s.02'), Synset('authentic.s.01')]
[]
[Synset('autonomic.s.01')]
[Synset('computer_science.n.01'), Synset('calculation.n.01'), Synset('calculate.v.01')]
[Synset('distribute.v.01'), Synset('spread.v.01'), Synset('distribute.v.03'), Synset('distribute.v.04'), Synset('circulate.v.0
3'), Synset('circulate.v.02'), Synset('distribute.v.07'), Synset('distribute.v.08'), Synset('distribute.v.09'), Synset('stagge
r.v.03'), Synset('distributed.a.01')]
[]
[Synset('analogue.n.01'), Synset('latitude.n.03'), Synset('parallel.n.03'), Synset('parallel.v.01'), Synset('parallel.v.02'),
Synset('twin.v.01'), Synset('parallel.a.01'), Synset('parallel.s.02')]
[Synset('system.n.01'), Synset('system.n.02'), Synset('system.n.03'), Synset('system.n.04'), Synset('arrangement.n.03'), Syns
t('system.n.06'), Synset('system.n.07'), Synset('system.n.08'), Synset('organization.n.05')]
[]
[Synset('algorithm.n.01')]
[]
```

圖 2-10 WordNet 測試

第三章 研究方法

本研究在此提出的觀察策略以及尋找趨勢方法的核心如下:

1. 使用研討會的 CFP 做為趨勢研究的資料集，其優勢在於和已發表的文獻不同，具有時間上的優勢。

2. 使用 ACMCCS 做為分類準則，藉由已完成的分類架構幫助我們分類，最後依照分類好的關鍵字給予其他特徵比對，藉此找出新的趨勢圖。

本研究在流程上可分為三大階段，第一階段為資料的蒐集，第二階段分為前處理與分類，最後階段是進行資料視覺化與分析，三階段的說明如下。

A. 資料的蒐集:在這階段裡，我們透過國外的研討會通知網站蒐集相關的訊息。

B. 資料的處理:在資料的處理上分為前處理與分類，最先遇到的問題，各個網站的 CFP 結構不同。故在蒐集完關鍵字後，需要處理各種非結構化的文字描述，每個研討會的徵稿敘述皆不同，在此階段我們將進行斷字處理以方便我們進行之後的工作。前處理完成後，使用 ACM 分類系統來幫助資料的結構化，回顧表 2-3 的 SKOS 格式，rdf 是分類系統的唯一值，我們將關鍵字透過比對的方式，對應到這些唯一的特徵值。

C. 資料視覺化:最後我們將分類完成的關鍵字，結合研討會本身的特徵進行視覺化，藉此幫助我們解讀這些資料背後的含意

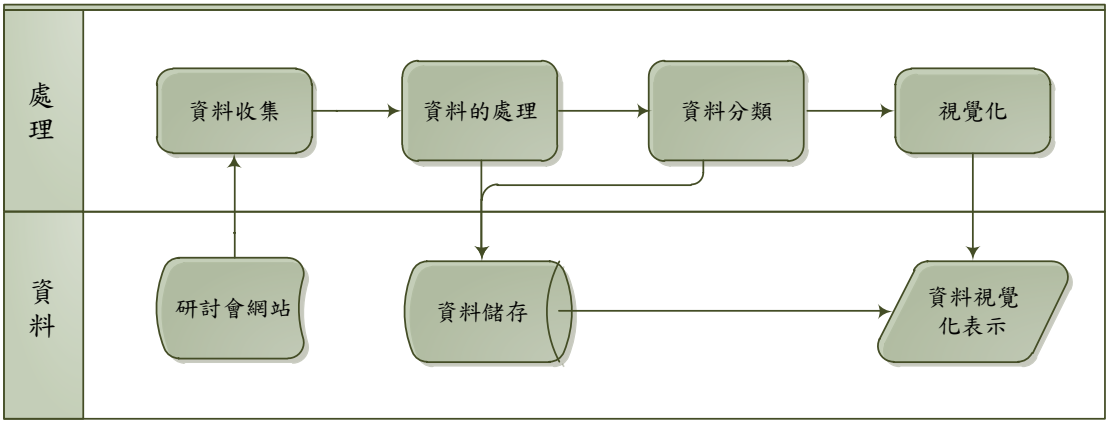

圖 3-1 研究流程圖

如圖 3-1 為本研究的流程，依照流程架構可以分為以下主要步驟，再後續章節將針對各步驟進行說明。

Step1:研討會特徵蒐集

Step2:資料前處理

Step3:使用 ACMCCS 進行分類

Step4:視覺化

3.1 研討會特徵蒐集

研討會 CFP 的蒐集是透過 Conference Alerts 網站的整理去搜尋。在此蒐集了 2013~2015 的研討會相關訊息。其訊息特徵包括了 CFP 的關鍵字，研討會名稱，日期，地點等，這些資訊將有助於現象的觀察。研討會特徵表示如下:

$C_i=\{N_i,T_i,L_i,W_i,K_{ij}\}$

C_i 為研討會 i 的集合，其中的 N_i 為研討會名稱，T_i 是研討會的舉辦時間，L_i 為研討會的舉辦地點，W_i 是研討會網址，K_{ij} 為研討會 i 裡的第 j 個關鍵字。

3.2 資料的前處理

在資料蒐集完後，接著進行資料的處理，考慮到各研討會的 CFP 的結構皆不相同。因此，我們將各個關鍵字給予結構化。在此之前，為了讓關鍵字能與 ACM 配對，必須先將各個描述類型的關鍵字斷字處理。

各個研討會的關鍵字的結構各不相同，一些研討會會用句子去描述需求內容，每個研討會有自己獨特的描述風格，都有非結構化的共通點。如: Big data and smart computing Applications (health care, medicine, finance, business, law, education, transportation, telecommunication, science, engineering, ecosystem, etc.)。在本節，我們將描述如何處理這些資料。

經過斷字處理後的 C 如下:

$SC_i=\{N_i,T_i,L_i,W_i,SK_{ij},SF_{ij}\}$

其中 SK_{ij} 為經過斷字處理後的關鍵字，SF_{ij} 為統計 K_{ij} 重複出現的次數，這新的特徵在之後可以幫助我們分析。ACMCCS 的集合，我們取 ACMCCS 的五大特徵，定義為:

$L=\{R_i,P_i,A_{ij},B_i,N_{ij}\}$

R_i 為第 i 個關鍵字的 rdf，P_i 是 prefLabel，在此 P_i 與 R_i 皆是唯一值。A_{ij} 是第 i 個關鍵字的第 j 個 altLabel，B_i 與 N_{ij} 分別是 broader 及 narrower。在此我們讓 C_i 的 $K_{ij} \in L$ 的 P_i 或 A_{ij}。因此，在斷字的處理上，斷字時同步將 C_i 的 Kij 透過 SQL 與字串比對處理，比對的目標為 ACMCCS 的 P_i 與 A_{ij}，如果符合則該關鍵字則直接紀錄成 SK_{ij}，如無法直接配對則進行斷字工作，斷字的規則如表 3-1。

- 不相關或者獨立出現的字眼:experiences、innovative、application 等，這些字眼如果獨立出現會無法分類，而一些抽象字眼或者形容詞、副詞等也會刪除。
- 標點符號:刪除標點符號並將受符號分隔的關鍵字分開當成獨立字。
- 組合類:常見的有 A and B、A and B C、A B and C，在這裡會先經過 SQL 去搜尋是否有節點，如果沒有則會轉換成 AB、AC、BC 等去做搜尋，都沒有節點則會拆成 A、B、C 各自獨立，交由分類階段去處理。

- 無介係詞組合類:AB 等一般組合如 network security，先做搜尋，如沒有節點則拆成獨立單字 A、B。
- 介係詞類:常見的有 for、in、on，這種情況處理類似 3，許多情況會直接刪除介係詞，將介係詞前後的關鍵字各自獨立。
- 描述類型: 人工判斷取出計算機領域關鍵字

表 3-1 斷字規則表

情境	規則
不相關或者獨立出現的字眼	刪除
標點符號	SQL 搜尋，無節點則刪除標點符號並將單字獨立
組合類:A and B	SQL 搜尋後，無節點則拆成各組合去比搜尋
介係詞類:for、in、on	SQL 搜尋後，無節點則刪除介係詞並取各自單字 A、B
無介係詞組合類	QL 搜尋後，無節點則取 A、B
描述類	人工判斷取出計算機領域關鍵字

3.3 使用 ACMCCS 進行分類

本階段的工作，將斷字處理後的關鍵字資料配對進 ACM 的分類系統裡。透過與 ACMCCS 的比對後，可以得到結構化的關鍵字，這些研討會的關鍵字都會有共同的結構。

將 SC_i 與 L 進行配對，可以得到 LC_i，LC_i 的集合表示式如下。

$$LC_{i=\{}\ LK_{ij},LF_{ij}\ |LK \in R\ \}$$

在關鍵字比對裡，LK_{ij} 為第 i 個研討會的第 j 個關鍵字的 rdf，其中$|R|$是 ACMCCS 分類系統 L 裡的 rdf 的集合，比對出來的 LK_{ij} 紀錄為 rdf，LF_{ij} 為 LK 的重複次數。

比對進分類樹的規則是先使用 SQL 過濾掉可以直接配對進樹的關鍵字，剩下的則透過人工處理，在此的人工處理需透過情境的方式，我們必須觀察 CFP 裡的上下文出現狀況。例如 Security 這個關鍵字單獨出現，在分類樹裡頂層概念有 Security and Privacy 底下有數百個相關的應用，如果將字直接分類到頂層那可能會影響最後的結果。因此，考量情境，如上一個關鍵字是 Networking 下一個關鍵字是 Network Reliability，我們可以判斷單獨出現的 Security 屬於 Network Security 節點的機率很大，透過上下文的判斷分類到對應節點，做法差別如圖 3-2、3-3。取層結果將影響最後的趨勢圖。

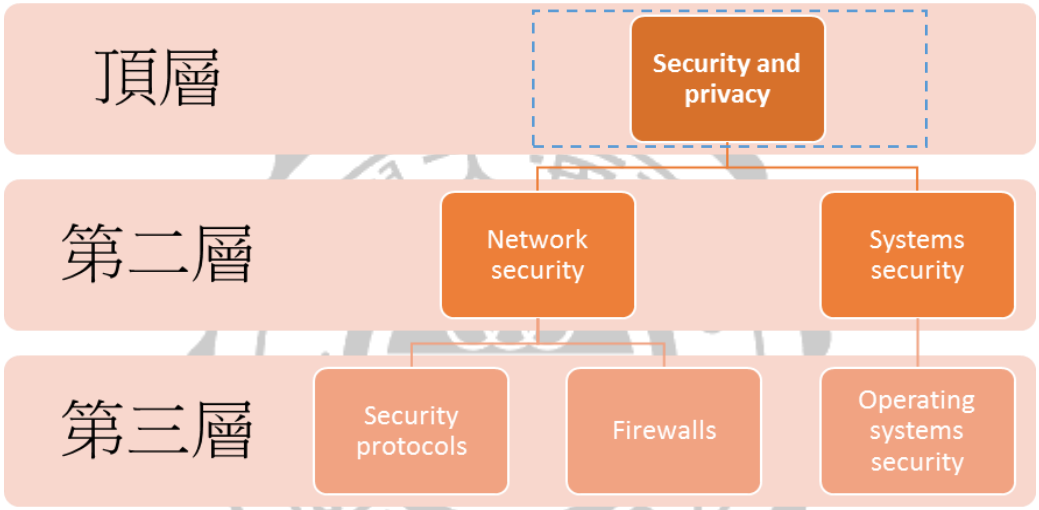

圖 3-2 分類直接取層

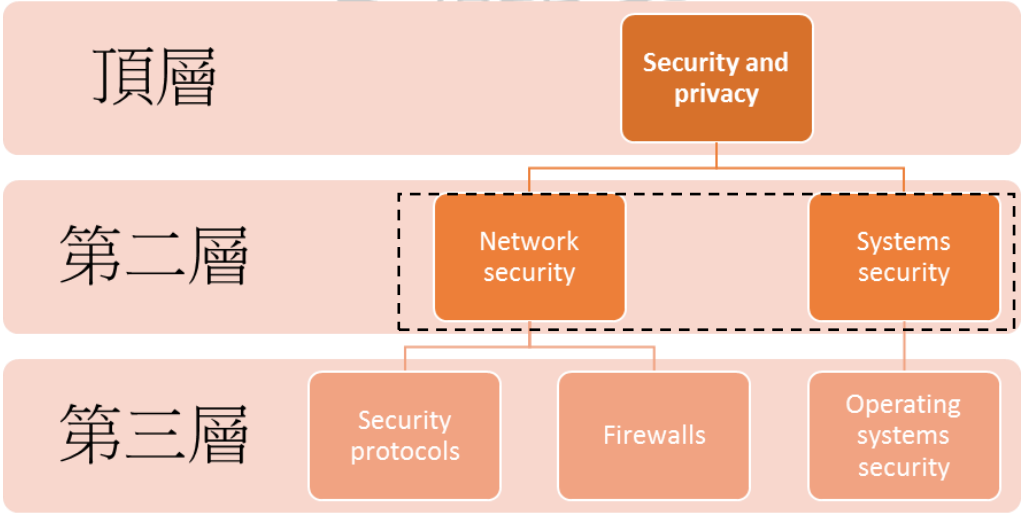

圖 3-3 分類情境取層

有些關鍵字找不到節點的情況下，我們只能藉此往上層分類或者尋找相似概念的節點，例如在各個研討會有一些領域如: Space-time application 關鍵字分類，透過觀察分類樹的節點，在 Spatial-temporal systems 這節點底下有 Location based services、Geographic information systems、Global positioning systems 等關鍵字，使用相似概念去分類到 Spatial-temporal system 節點。

另一個例子如學科的應用或工程。 Accounting 關鍵字在分類系統中沒有節點，在 Applied computing 這個頂層節點下有許多學科的應用，比如 Aerospace、Military、Arts and humanities，此顆樹是各領域的應用，關鍵字沒有節點的情況下，我們可以透過這種觀察將 Accounting 這關鍵字歸類至 Applied computing 母節點。此階段的工作就是將所有關鍵字分類到與 ACMCCS 系統對應的節點，分類完成後我們看到的就不再是非結構的資料，而是一種結構化的、有依據的資料。一些新興技術的問題，如 Big data、物聯網等，在此我們不破壞分類樹的本體結構，處理方式是獨立提出，直接計算出現頻率。圖 3-4 是資料處理的流程。

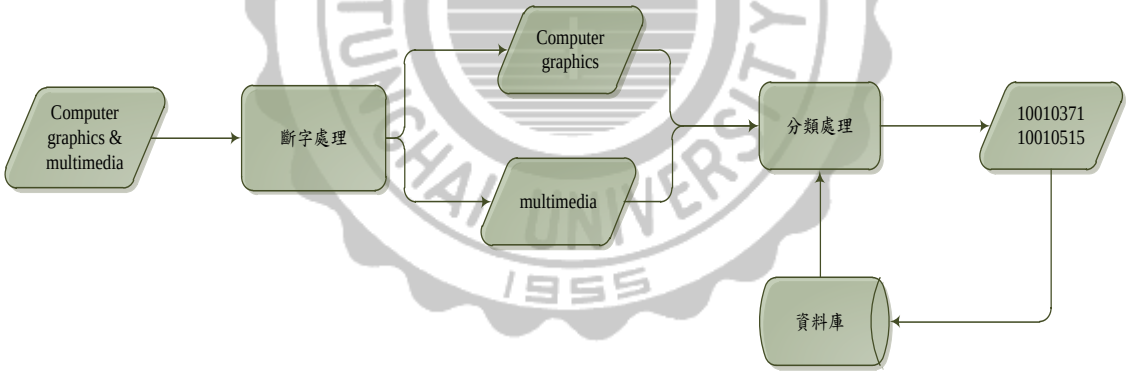

圖 3-4 關鍵字處理流程

3.4 視覺化

圖 3-5 是資料處理的流程，從起點開始，原始關鍵字經過斷字處理後，進入分類階段，在分類處理後我們可以得到 rdf，rdf 是我們最終結構化的資料，只要透過資料庫的比對即可得到 rdf 所對應的關鍵字。在此我們藉由 rdf 做為結構化的索引，當每個關鍵字對應到符合的 rdf，透過 ACMCCS 的分類結構特性，在視覺化下呈現下進行下鑽或者統計節點，如圖 3-5，此結構可以結合研討會本身特徵進行視覺化呈現。

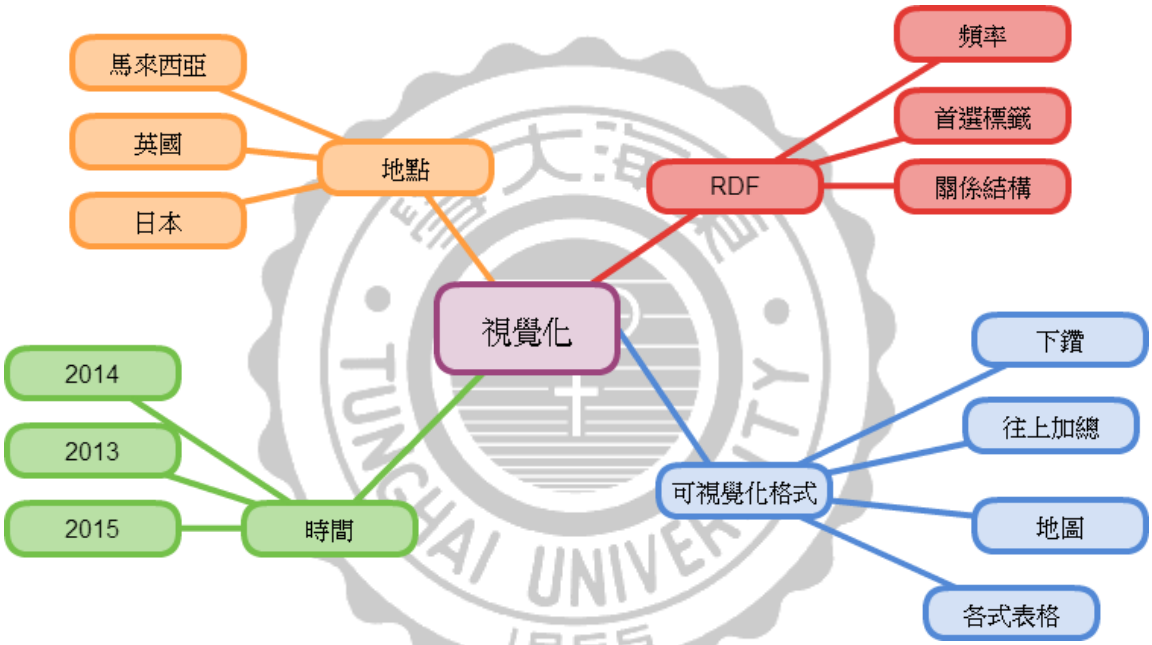

圖 3-5 視覺化特徵

第四章 實驗設計與結果

本研究提出的分類機制會經歷四個步驟，於四個步驟結束後再進行趨勢的呈現。在這章節首先會針對流程做一個說明，其中包含研究的限制，資料細節，儲存的結構與視覺化。最後我們提出部分例子來說明我們的觀察策略所觀察到的現象。本研究的實驗處理共分三個部分，資料蒐集;資料處理;視覺化，資料處理又分為前處理-斷字與分類，如圖 4-1。

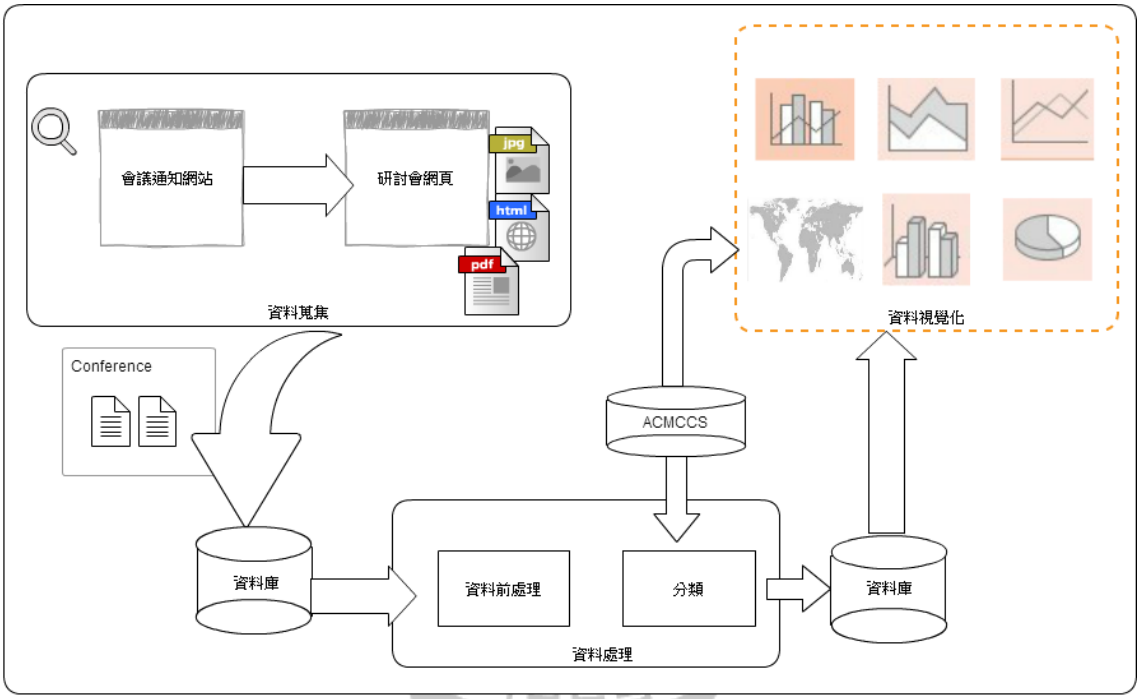

圖 4-1 實驗架構

4.1 時間跨度與研究限制

本研究的資料透過會議通知網站與研討會網站蒐集，蒐集了 248 個研討會的 Call For Papers，整理出 12419 個關鍵字，資料篩選皆為計算機領域。在研究上我們將遇到一些限制，趨勢分析需要一段時間的觀察，研討會通知網站的資料庫不會保存已過期的研討會，而研討會本身的網站也可能會隨著時間關站，或者新一屆的資訊會蓋過前一屆。我們蒐集了 2013 開始至 2015 年 9 月的研討會資訊。在蒐集的數量上，可能也會影響到趨勢的準確，在此我們的選擇以國際研討會為主。在時間選擇上，本研究將時間區塊分成 2013、2014、2015 三年，Google Trends 的時間設為 2013 年至 2015 年 8 月。

ACMCCS 雖然經歷多年與許多專家所建構，但仍有其限制在，首先目前最新版為 2012 年版本，之後較熱門或新興的關鍵字可能沒有節點。其次為分類樹本身的完整性，同第一點並非所有的節點皆可在分類系統找到。最後的問題為重複節點，部分關鍵字可能會出現在不同子樹，也就是說一個節點會出現多個父母節點的現象。其第一、二點的替代方案為往上層取，只要沒有節點可以透過人工判斷往上層取，第三點的解決方式為如出現重複節點的關鍵字，各個子樹下皆計算。雖然可以透過一些替代方案解決限制，但仍有可能會對結果造成影響。

4.2 資料的索引結構

索引結構如圖 4-2 所示，有三個主要的資料表，分別儲存的是研討會資訊、研討會關鍵字、ACMCCS，其中研討會資訊儲存了時間地點等特徵，研討會關鍵字儲存了各研討會斷字後的關鍵字，node 欄位則是對應到 rdf。最後的 ACMCCS 儲存了 ACM 分類系統的 SKOS 資料，做為我們在資料視覺化時的索引，如表 4-1。

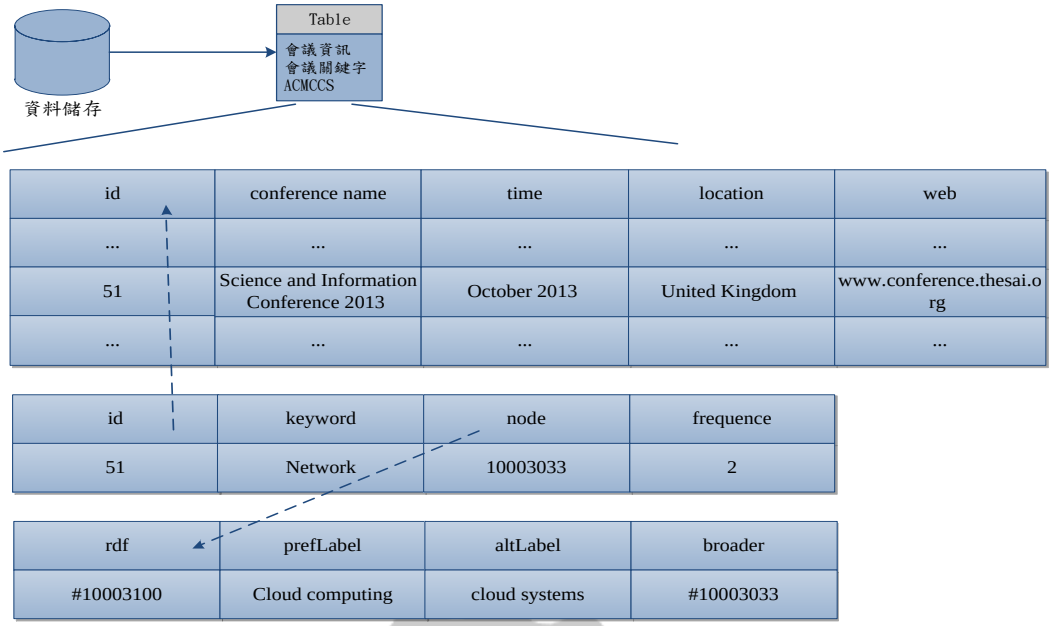

圖 4-2資料庫的索引結構

表 4-1 分類完的索引結構

conference	ACMCCS	
keyword	rdf	broader
10003145	10002959	10002953
10011006	10002971	10002952
10003465	10002974	10003450
10011004	10002975	10003451
10011076	10002975	10010031
10003465	10002976	10003451
10011114	10002979	10002978
10010769	10002980	10002979
10010220	10002981	10002979
10010972	10002982	10002979
10011096	10002983	10002979
10011004	10002984	10002979
10003514	10002985	10002979
10011007	10002986	10002978
10003138	10002987	10002986
10003543	10002988	10002986
10011082	10002989	10002986
10011094	10002990	10002986
10011124	10002990	10003790
	10002991	10002978

4.3 資料視覺化

在這章節裡我們提出一些大方向的趨勢，由於關鍵字的節點眾多，必須將節點計算到上層以方便觀察，而對特定領域有興趣時可往下展開。並將結果與 Google Trends 進行比較。

首先將所有關鍵字計算到頂層節點，圖 4-3 是研討會感興趣領域比例圖，我們可以從圖 4-3 看出研討會普遍對 Information systems 與 Computing methodologies 的興趣最高，再來是 Networks 與 Applied computing 領域。圖 4-4 是挑選出五個數量最多領域的時間序列，從圖上來看 Applied computing 領域有開始降溫的趨勢，而 Information systems 從 2013 年到 2014 後有出現成長的狀態，Networks 領域在 2013 到 2014 間有成長趨勢，嗣後就開始呈現弱勢，Computing methodologies 則出現了持續成長趨勢。從圖 4-3 與圖 4-4 觀察，Information systems 在總數佔了 17%的比例，且三年來的趨勢都呈現穩定，因此可以判斷該領域非常受到關注。雖然 Applied computing 在比例上佔了 13%，從分析上來看三年來感興趣的研討會則有減少趨勢。圖 4-5 是圖 4-4 的 Google Trends 版本，上圖的趨勢線為 Network 的走勢，因計算搜尋比例機制，其它關鍵字在搜尋比例下被視為 0，從這裡可以看出學術領域與一般使用者的搜尋不同，相同的部分為在學術或一般使用者在 Network 關鍵字的關注程度都維持在一定比例，並不會有大幅度的變動。4-5 的下圖是去掉 Network 後的趨勢圖，藍色為 Information System 紅色是 Applied computing，趨勢表示出一般使用者對於 Applied computing 的興趣高於 Information System。

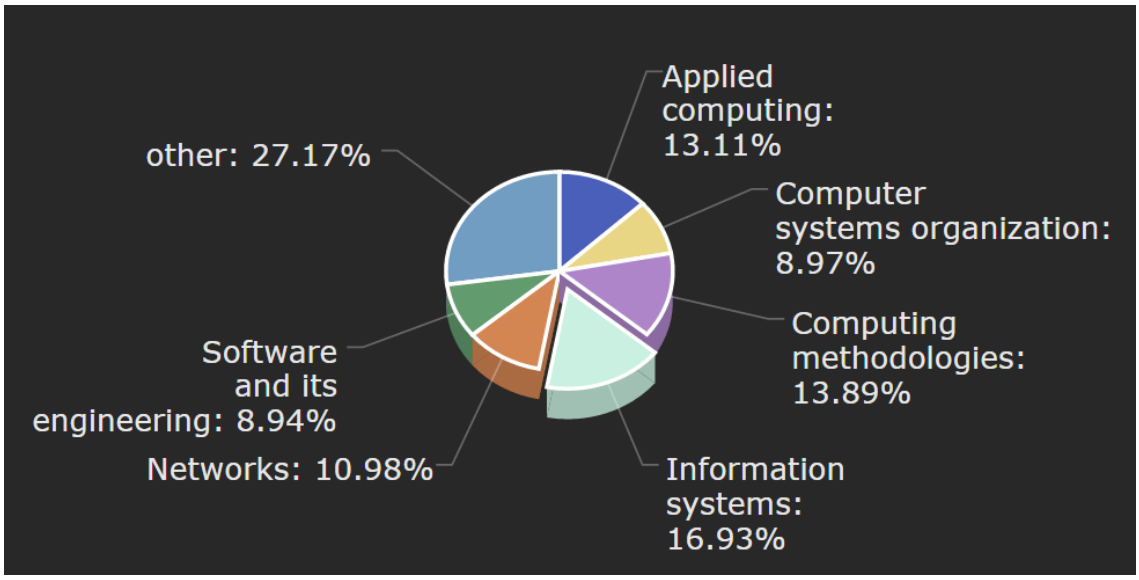

圖 4-3 各領域出現比例

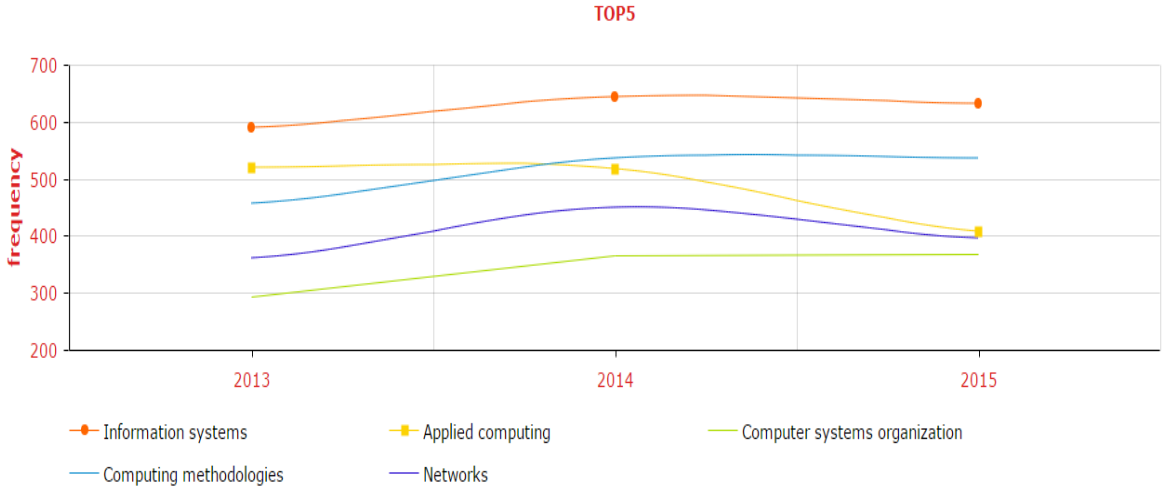

圖 4-4 前五高領域趨勢圖

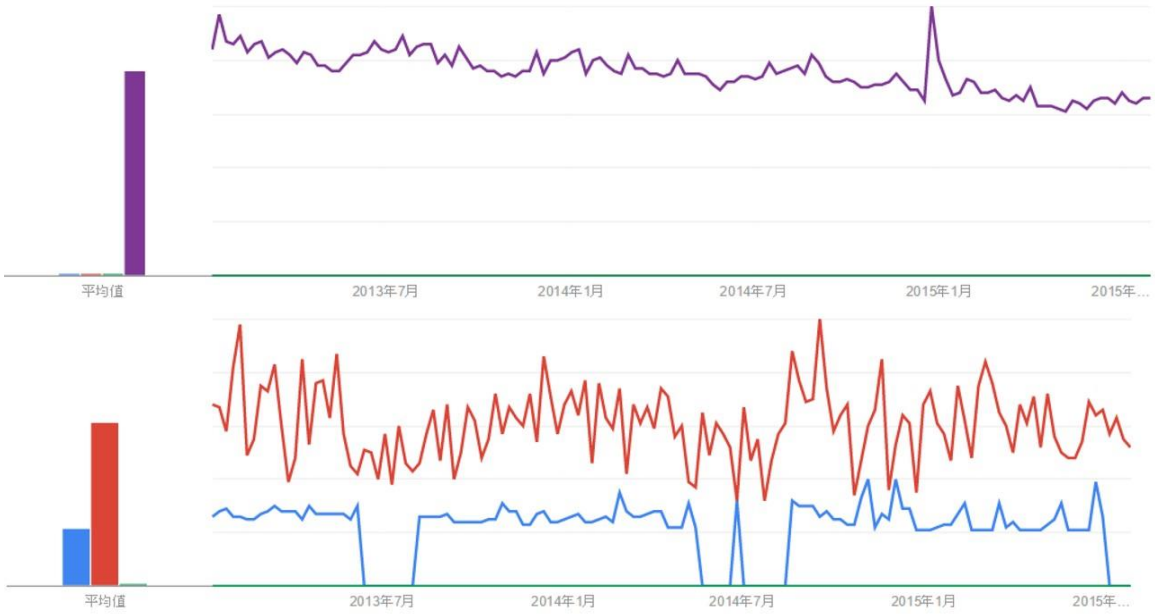

圖 4-5 Google Trends-前五高領域趨勢圖

在計算機領域常常會出現新興的應用，故我們選擇近年較熱門的 Big Data 做為觀察對象。圖 4-6 是 Big Data 近年的變化，從圖 4-6 可以明顯的看出 2 年的出現頻率有明顯的增加，走勢有熱潮的現象出現。雲端的趨勢呈現較高且有成長趨勢，從學術的趨勢線來看雲端的熱門程度仍高於 Big Data。圖 4-7 是 Google Trends 版本的巨量資料與雲端趨勢，與學術不同的是 Big Data 的搜尋次數在 2014 年超過雲端，相似的是兩項趨勢的上下浮動有同步的現象。參考 Google Trends 的地區搜尋熱門度，將趨勢結合地圖觀察，如圖 4-8 與圖 4-9 將 Big Data 的出現地點視覺化，這裡地圖所表示的是關鍵字在各地出現的頻率，顏色越深代表次數越多。圖 4-10 的 Google 搜尋地區熱門度，與學術的出現地點相似。

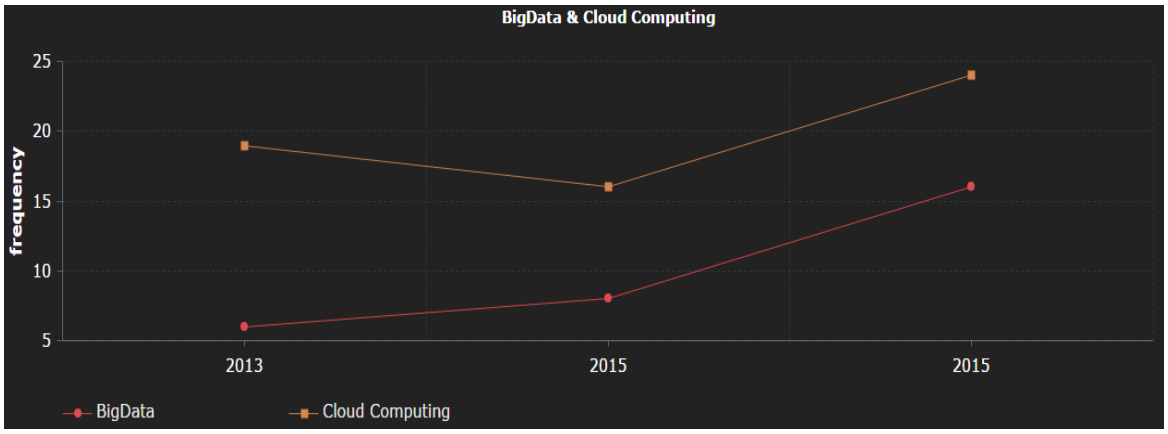

圖 4-6 Big Data 與 Cloud Computing 趨勢圖

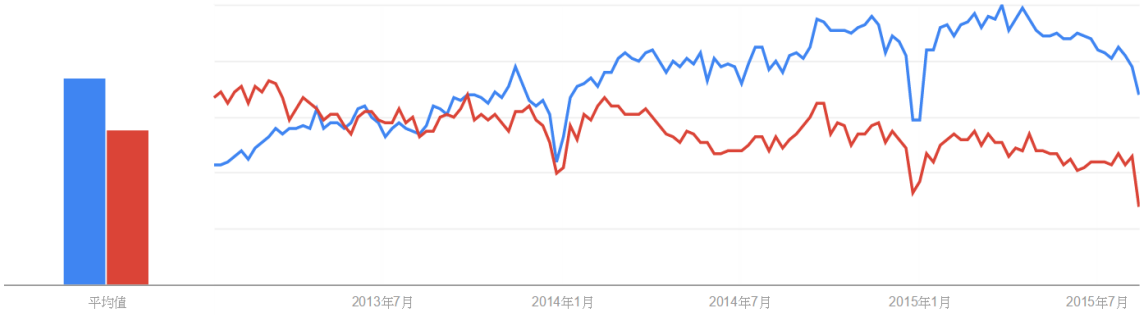

圖 4-7 Google Trends -Big Data 與 Cloud Computing 趨勢圖

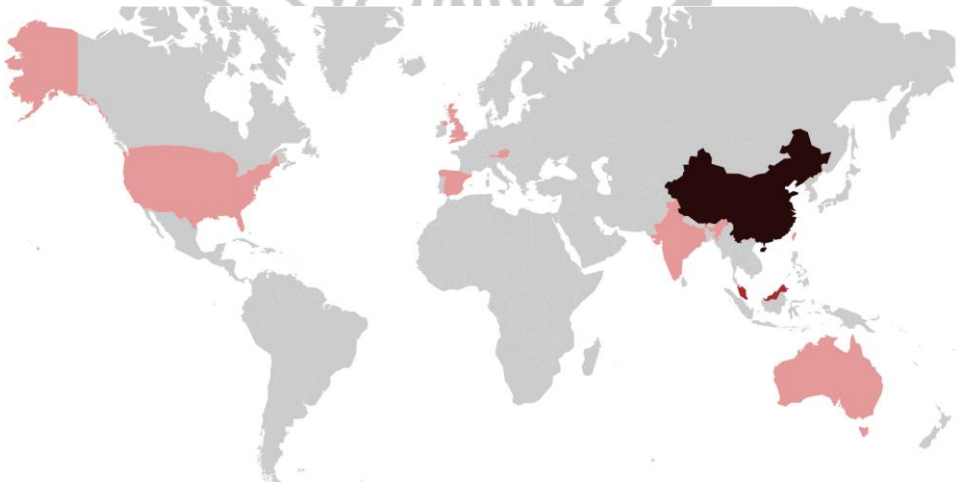

圖 4-8 Big Data 在第 1 年的出現地區圖

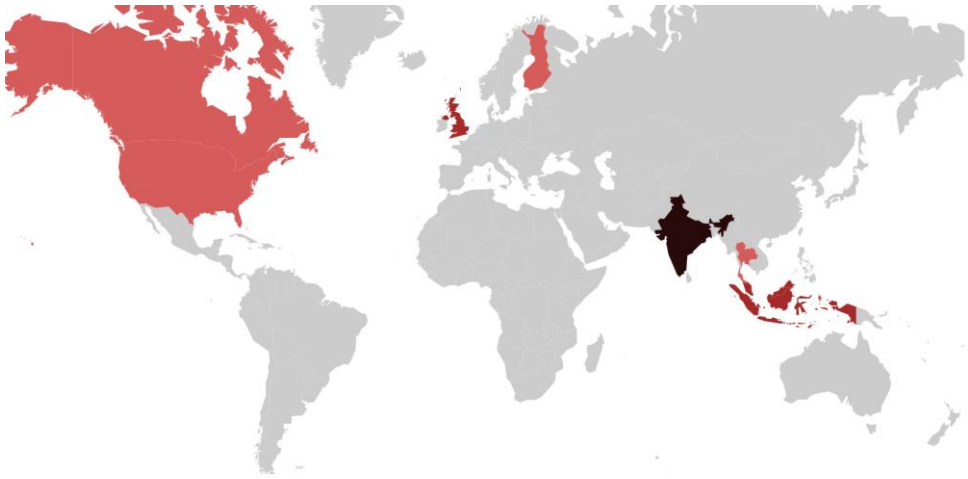

圖 4-9 Big Data 在第 2 年的出現地區圖

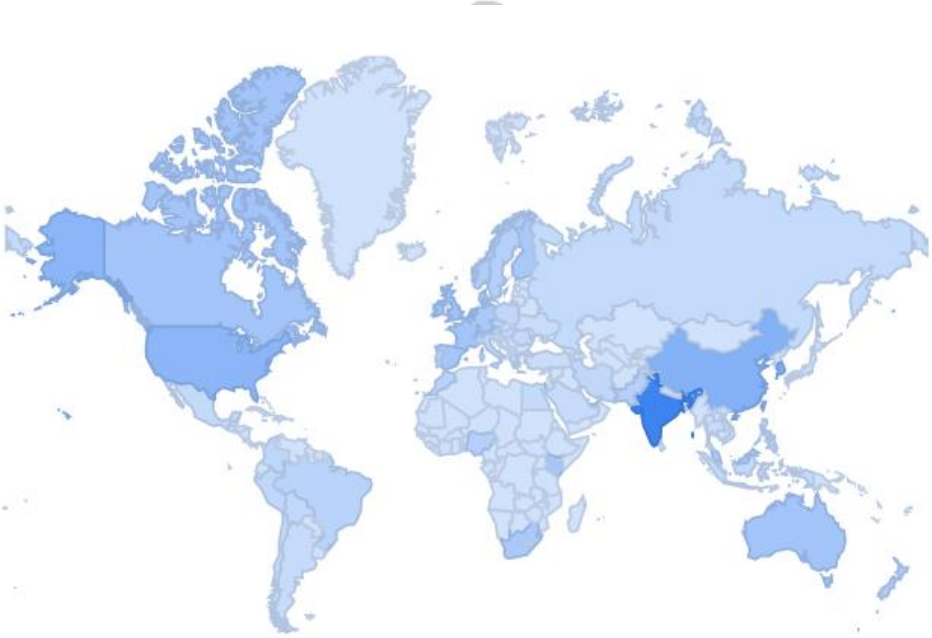

圖 4-10 Google Trends-Big Data 三年間的地區搜尋熱門度

Big Data 持續的成長是一個研究方向，而當一個新興領域熱過頭時是否會出現回饋，我們觀察前幾年很熱門的 RFID，從圖 4-11 可以看到 RFID 出現的頻率低，而三年來的出現次數又持續的下降。物聯網在搜集的資料裡則顯示出 2014 與 2015 沒有變動。圖 4-12 的 Google Trends 版本，藍線的物聯網有明顯成長趨勢，紅線的 RFID 在近一年則沒有明顯變動。

Information systems 是佔總比例最高的領域，在此我們將領域往下展開到第 2 層，藉此觀察這領域在 Big Data 的成長下是否有變化，如圖 4-13。從第 2 層的節

點觀察，Data management systems 與 Information retrieval 在 3 年裡都持續的成長，這現象符合 Big Data 的成長趨勢，圖 4-14 的黃線為 world wide web，藍線是 information systems applications 與紅線 data management systems，從圖可以看出 world wide web 的搜尋比例相比其它關鍵字高出許多。

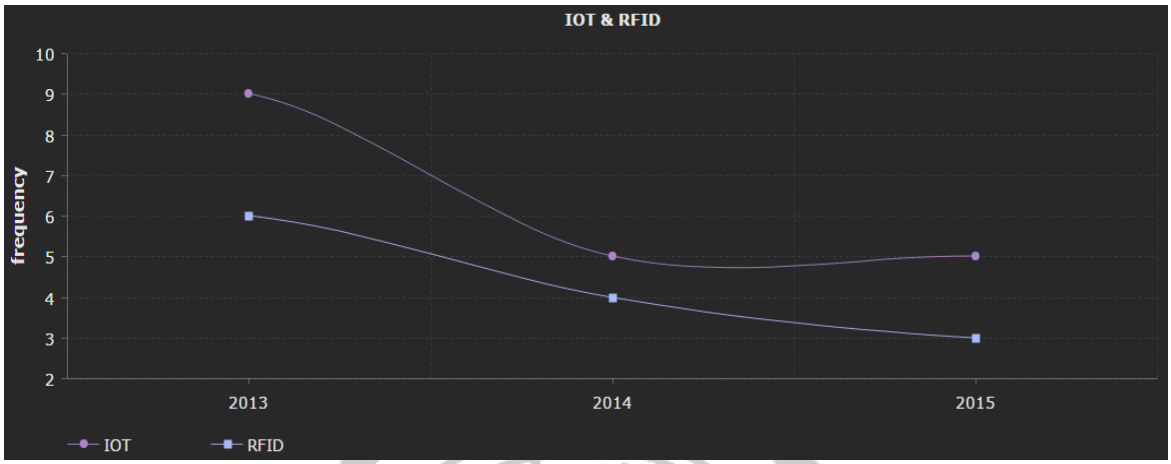

圖 4-11 物聯網與 RFID 趨勢圖

圖 4-12 Google Trends-物聯網與 RFID 趨勢圖

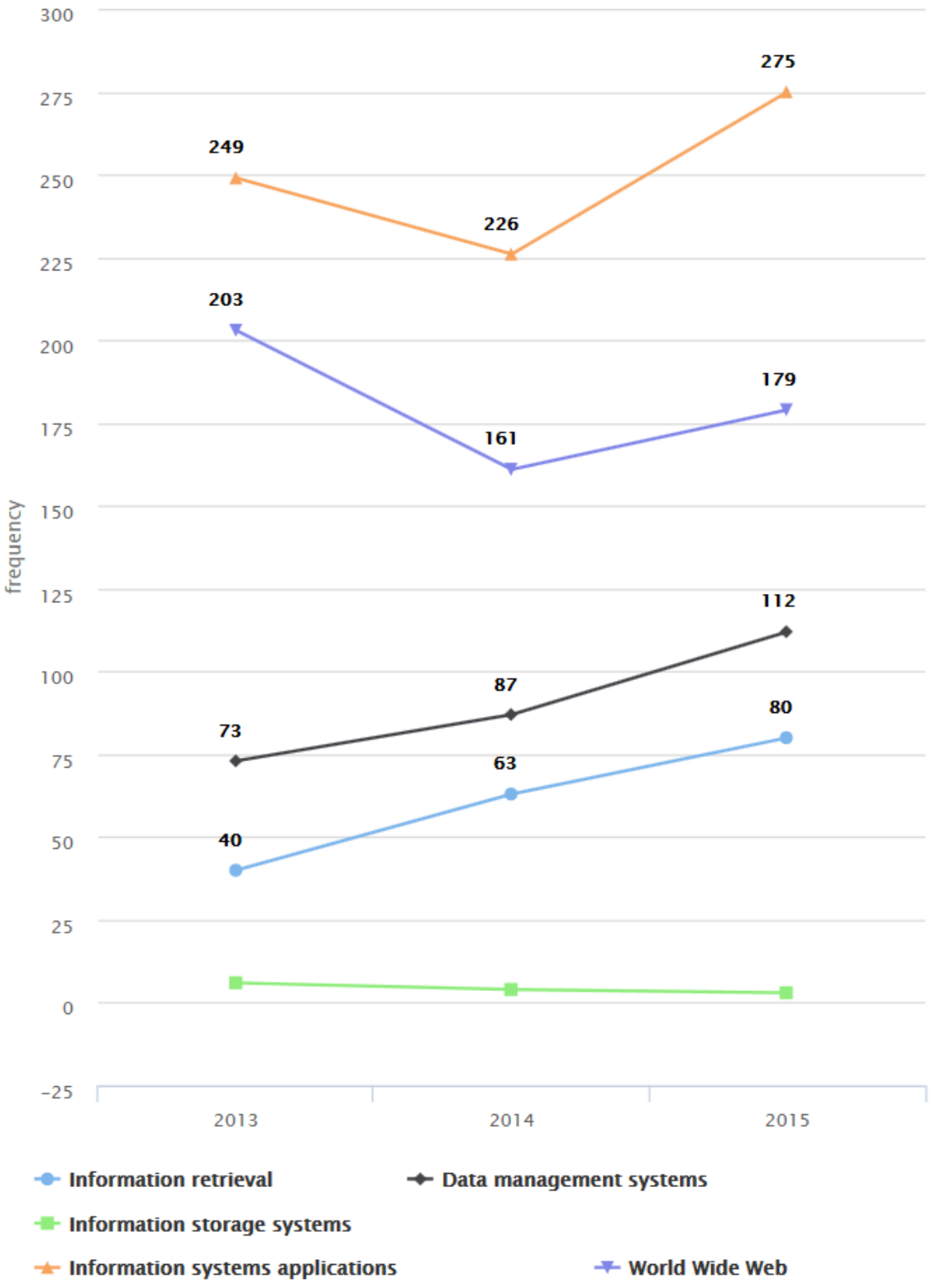

圖 4-13 Information systems 領域底層趨勢圖

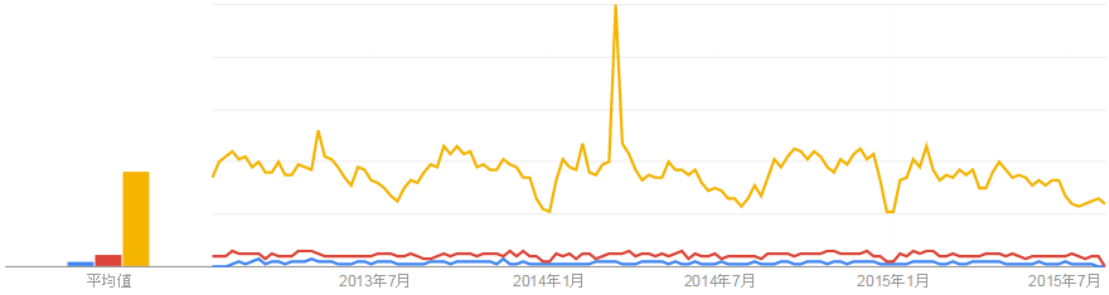

圖 4-14 Google Trends-Information systems 領域底層趨勢圖

圖 4-15 為人工智慧與知識表達式及自然語言處理的趨勢圖，此圖的例子是做為人工智慧與其下層節點關係，將下層節點加總後，觀察子節點在母節點下的比例，並可直接觀察該子節點是否與母節點的走勢相同。圖 4-16 為新興領域圖，在 ACM 的分類底下有新興技術節點，BigData 與 IOT 在 2012 年版本的分類系統中還未將此二節點加入，因此在圖比較中將這二節點歸類在新興技術，other 項目為新興技術節點包含其子節點的加總，雲端計算是近年出現的熱門領域，雖然在 ACM 分類系統已有節點，象徵著已非新興的領域，但仍在此與幾項新興技術做比較。從圖 4-16 的比例圖可以看出，雲端計算在新興技術中是特別熱門的，佔了 41.55%，圖 4-17 的搜尋熱門度則表示出黃色的 BigData 為最高，次之為紅色的雲端運算，最後是藍色的物聯網。

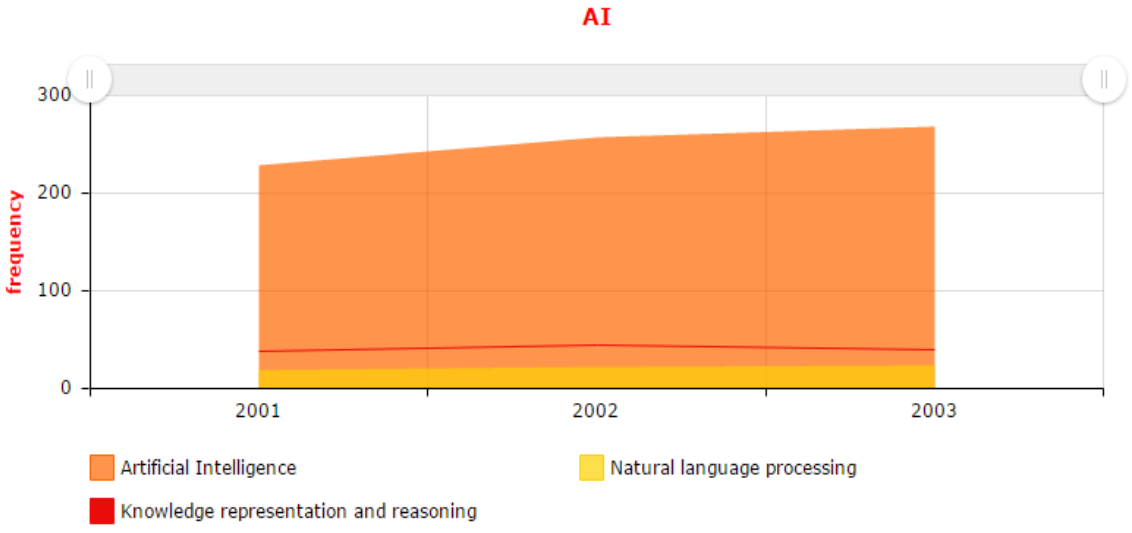

圖 4-15 AI 趨勢圖

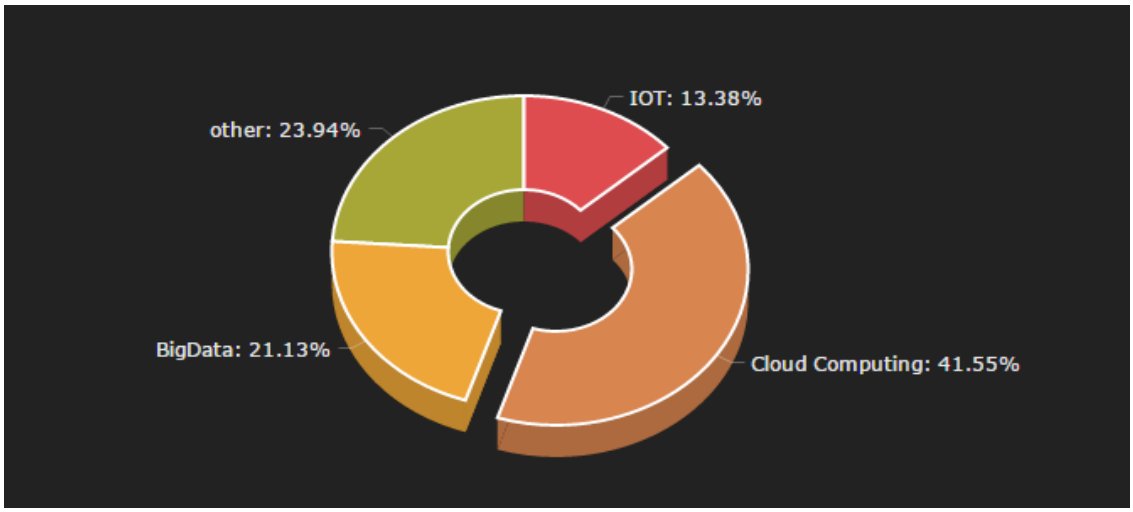

圖 4-16 新興領域比例圖

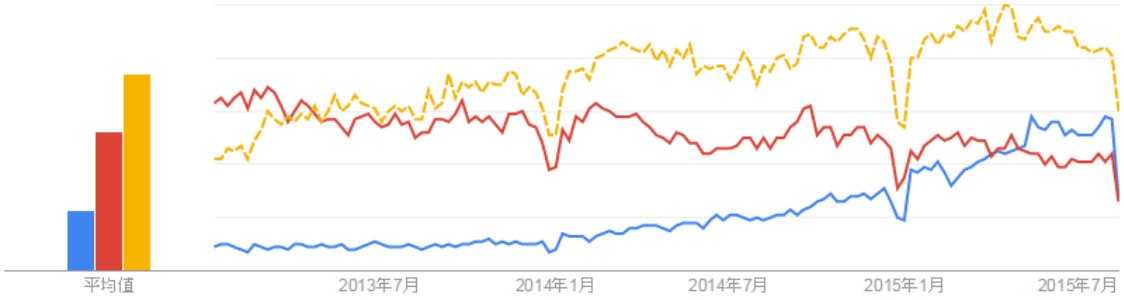

圖 4-17 Google Trends-新興領域比例圖

第五章 結論

本研究提出一方法觀察近年來計算機學科領域的發展趨勢，我們選擇了研討會的 CFP 做為關鍵字，彌補使用過去文獻做為時間點的不足。透過 ACM 分類系統解決了 CFP 的結構化問題，再以視覺化的呈現幫助了解未來的研究發展趨勢。本研究藉由 ACM 的分類系統處理資料視覺化的問題，以 rdf 做為索引，展示出如圖 4-4 與 4-13 的階層概念，從不同的維度觀察趨勢的走向。

使用相同的關鍵字從本研究之方法與 Google Trends 的熱門搜尋做比較，可以發現 Google Trends 無法有效的呈現較冷門關鍵字，在多個關鍵字比較時會受高熱門度關鍵字影響，Google 搜尋的使用者多為一般使用者，在與學術研討會的關鍵字頻率比較時，熱門關鍵字的浮動大致相同，遇到一般使用者較少見的關鍵字，則容易出現搜尋比例過少問題。本研究之方法可以有效處理 Google Trends 的不足，對於過少的關鍵字可以透過 ACM 分類系統的概念往上取層，趨勢的變動則可以往下觀察主要變動的關鍵字，因此本方法可幫助學術領域人員了解研究發展趨勢。

針對本論文的不足地方，我們提出以下幾點未來研究方向。

(1) 從趨勢研究的角度上，蒐集的年份與資料量尚不足，包含時間跨度中遺漏的研討會，未來需持續蒐集研討會資訊，已達到更佳的趨勢分析。

(2) ACMCCS 的限制，許多情況下是領域與技術或技術與技術的組合，ACM 的分類系統無法直接處理這類問題，許多節點都往上層取的情況下，會影響較下層節點的觀察結果。一些較為特殊或者新興的關鍵字，在 2012 版的分類系統尚未有節點而被獨立計算，也因此這類關鍵字將失去分類系統可階層觀察的特性，建議未來可以朝分類系統的改善或者自動分類的方向研究。

參考文獻

中文文獻

1. 丁一賢、陳牧言(2006)，《資料探勘》滄海書局。
2. 王京盛(2012)，《考量語意及引用分析之研究主題趨勢分析方法》，國立成功大學資訊管理研究所碩士論文，未出版。
3. 王宏德(2013)，學術研究趨勢之分析與探討－以 100 學年度臺灣學位論文為例，國家圖書館館刊一〇二年第一期，第 75-98 頁。
4. 林宜瑩(2010），《利用時間因子與名詞片語之文獻主題追蹤法》，國立成功大學資訊管理研究所碩士論文，未出版。
5. 胡姝涵(2006)，《會議公告網站資訊擷取之研究》，國立中央大學資訊工程學系碩士在職專班碩士論文，未出版。
6. 陳光華(2010 年 9 月)，學術會議資訊之擷取及其應用，中文計算語言學期刊，15:3-4 期，第 237-262 頁。
7. 許育聞(2008)，《會議與期刊文獻對預測主題趨勢之比較研究—以「資訊檢索」領域為例》，國立臺灣師範大學圖書資訊學研究所碩士論文，未出版。
8. 羅濟群、陳志華、呂志健、程鼎元(2012 年 9 月)，飲食保健推薦機制之設計與實作--以中國飲食療法為例，電子商務學報，第 513-548 頁。
9. 全國博碩士論文書目資料收錄範圍，
   http://ndltd.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi/ccd=ej252f/aboutnclcdr

英文文獻

10. Chin, A. ,Xu, Bin , Yin, Fangxi , Xia Wang .(2012). Using Proximity and Homophily to Connect Conference Attendees in a Mobile Social Network , Distributed Computing Systems Workshops (ICDCSW) 2012 32nd International Conference,79 – 87.
11. David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, Andrew Tomkins.(2005). Geographic routing in social networks,Proceedings of The National Academy of Sciences - PNAS(102:33),11623–11628.
12. Fellbaum C (1998) WordNet: an electronic lexical database. Bradford Books, Bradford.
13. M.A. Felix, V.Q. Benjamin, C.R. Zaida, C.A. Elena, H.S. Victor, J. M.F. Francisco.(2005). Domain analysis and information retrieval through the construction of heliocentric maps based on ISI-JCR category cocitation ,Information Processing and Management, 41 (6) , 1521–1533.
14. G. Judelman.(2004). Knowledge Visualization. Problems and Principles for Mapping the Knowledge Space, Collective Intelligence Laboratory Fellow University of Ottawa, Canada.
15. G. Salton and M. J. McGill.(1983).Introduction to modern information retrieval, McGraw-Hill, Inc. New York, NY, USA.
16. Henri Barki , Suzanne Rivard , Jean Talbot. (1998). An information systems keyword classification scheme,MIS Quarterly(12:2), 299-322.
17. Iris Vesseya,V. Ramesha, Robert L. Glassb. (2005) .A unified classification system for research in the computing disciplines,Information and Software Technology(47:4),245–255.
18. Kai-Yuan Cai,David Card.(0008).An analysis of research topics in software engineering - 2006,Journal of Systems and Software Volume 81, Issue 6, Jun,1051–1058.
19. Luca Masud , Francesca Valsecchi , Paolo Ciuccarelli , Donato Ricci , Giorgio Caviglia. (2010).From Data to Knowledge - Visualizations as Transformation Processes within the Data-Information-Knowledge Continuum, Proceedings of the 2010 14th International Conference Information Visualisation,445-449.

20. Ozer Ozdikis,Pinar Senkul,Halit Oguztuzun.(2012). Semantic Expansion of Tweet Contents for Enhanced Event Detection in Twitter,"Advances in Social Networks Analysis and Mining (ASONAM 2012) ,20-24.
21. Prantik Bhattacharyya, Ankush Garg ,Shyhtsun Felix Wu .(2011). Analysis of user keyword similarity in online social networks,Social Network Analysis and Mining(1:3) , 143-158.
22. Richard L. Baskerville , Michael D. Myers. (2002). Information systems as a reference discipline, MIS Quarterly(26:1),1-14.
23. R. Studer, R. Benjamins, and D. Fensel.(1998). Knowledge engineering: Principles and methods, Data & Knowledge Engineering, 25(1–2):161–198.
24. Salton G. and McGill M.(1986). Introduction to modern information retrieval,McGraw-Hill, New York.
25. Simoudis, E. (1996). Reality check for data mining, IEEE Expert, 11(5), 26-33.
26. Tu, Y.-N., & Seng, J.-L. (2009). Research intelligence involving information retrieval – An example of conferences and journals, Expert Systems with Applications, 36(10), 12151-12166.
27. 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,71-76.
28. Z. Pousman, J. T Stasko and M. Mateas. (2007).Casual Information Visualization: Depictions of Data in Everyday Life,IEEE Transactions on Visualization and Computer Graphics(13:6), 1145-1152 .
29. IEEE Xplore digital Library，http://ieeexplore.ieee.org/Xplore/home.jsp
30. ACM Computing Classification System，
http://www.acm.org/about/class/class/2012.
31. Google teend，https://www.google.com.tw/trends/