

30  
B )  
當前語文問題學術研討會（國立台灣大學文學院主辦）

語文工作與資訊發展 ——  
從電子文件的發展談對語文研究的期盼

謝清俊

中央研究院 資訊科學研究所 研究員  
台灣大學 文學院圖書館學系 兼任教授

中華民國八十三年六月廿六日

# 語文工作與資訊發展 —— 從電子文件的發展談對語文研究的期盼

## 壹、前言

### 一、與世界對話

1994年6月10日至12日，中央大學中國文學系主辦的『與世界對話——甲戌年世界紅學會議』有項創舉：會中所發表的四十四篇論文全都放在學術網路上。誠如主辦單位宣稱，此舉「突破了會場空間的限制」，使得沒有到達現場與會的人士，無論在世界那個角落，隨時可以透過電腦網路看到所有的論文，也可以把想要的論文列印下來，或者乾脆把全部會議的議程、參與人員名單、以及論文等電子檔案抄錄在小小一片磁片中以備後用。這是電子文件展現的魅力。

其實，電子文件的魅力尚不止於此。例如：主辦單位不需勞民傷財地印製論文集，而電子形式的論文集，卻可無窮無盡地供應給任何一位真正需要它的人。又如，電子文件和資訊技術的運用，改變了傳統學術會議的溝通方式，使參與會議的員額幾乎無所限制、降低了會中發問討論的約束、省下了差旅、食宿、交通……等事務。這些方便之處，幾可任憑想像難以盡舉。如果說它能改變了傳統的知識傳播和分配的形態，也不過份。

此外，在發表的論文方面，有利用計算機對紅樓夢的內容作計量分析者，歸納出許多有趣的發現；有利用計算機作內容檢索者；有利用超文件（hyper-text）和多媒體（multi-media）技術製作綜合性的電子紅樓夢系統者。這些事實說明了中文電子文件的環境業已越來越成熟。【註一】

### 二、由來同一夢

用計算機處理浩瀚的文獻一事，一直是人類的夢想。早在1960年代就有利用計算機處理法律和新聞檔案之事。然而，要能充份利用計算機來處理文獻，有其客觀條件。首先，文獻的數量是一個大負擔，計算機的儲存資料空間和處理資料速度若未達到能夠應付大量文獻的水準，那麼一切免談。其次，計算機若是沒有累積相當的語言文字處理能力來做文件表達、檢索和查詢的工作，那麼，也只能望文興歎了。此外，文獻索引和擷取的技術，各行業使用文獻的專業要求，使用人和機器間溝通的問題，以及文件共享和傳輸機制之建立等等，無一不是必須克服的障礙。再者，計算機對各種媒體的處理能力，如圖表、影像、語音之類，亦與電子文件的發展有增益的關係。

到1980年，上述的各種困難均經研究而獲紓解。從1983年開始，市面上陸續出現了全文資料庫，正式啓開了計算機普遍地處理大量文獻的紀元。能有這樣的成就，在此之前廿多年計算語言學的研究是幕後最大功臣。

過去十年中的發展更是一日千里。目前，計算機的功能和價格已經不再是問題。電腦網路、主從（client/server）式的資料伺服結構、超文件、多媒體、光碟、視窗、智慧型資料庫、智慧型人機介面、文件表達的標準……琳瑯滿目的發展，造成了近年來電子文件大量出爐的盛況，僅從出版的各種光碟文獻來看，就已令人目不暇給而歎嘆為觀止了。

電子文件（Electronic Document）簡單地說，即是以電子媒體所表達的文件。較之於書本，似乎僅是媒體之改變，然而其影響之廣大深遠，遠比畢昇活字版印刷技術之發明，實有過之而無不及。它是目前「資訊發展」中重要之主流，對人文、社會、乃至於整個文化的發展關係至鉅。電子文件之所以能如此發展，除了靠電子媒體特殊的性質之外，實多有倚重「語文工作」基本研究之支持。此間關係即本文所欲申述之重點。

然而，電子文件究竟有多大的影響力呢？電子文件的發展和語文研究工作究竟有怎麼樣的關係呢？要說明白這些事情的緣由，且讓我們在下文中從電子文件的基本性質談起。

## 貳、電子文件的性質

### 一、文件的現代風貌

電子文件最明顯的特色是易於複製，它幾乎可算是不消耗資源、亦無污染的生產行為。於是，文件被製成電子形式後，似乎變成了「取之不盡，用之不竭」的資源：不僅可供自愉悅，亦堪持贈君。這是因為以複製品贈人後自己並無損失的緣故。

復次，透過電腦網路，電子文件可以高速傳輸，無遠弗近轉瞬即至。如果把文件比作貨物（例：書報、雜誌），那麼電子文件透過網路的運輸就像是不花錢又不花時間而能迅速、確實送達目的地的方法。

由於電子文件有這兩個史無前例的特點，再加上它體積小，重量輕，可長久保存等等有利的性質，使得電子媒體賦予文件嶄新的活力，讓文件能活出現代風貌，並提供徹底改變知識分配和擴散形態的能力，以及開啓真正人人平等共享知識環境的契機。

當然，要達到上述的境界，一些基本的建設是必須的，像是四通八達高容量和高速度的電腦網路，以及無所不在的終端設備（如：個人電腦）等。正因為這個緣故，世界各國莫不競相開始建設「國家的基本資訊設施」（National Information Infrastructure, NII），我國亦然。這些基本設施的建設均由國家統籌規劃，相信明天會比今日更好。而事實上，國內的學術網路目前已經可以做很多事了，所以在此暫且不討論這個以工程為主的問題。【註二】

## 二、內容與形式

借用傳播學的觀點，文件 (document) 是作者把他想傳遞的概念 (concept) 用語言文字表達。文件的內容 (content) 即指作者將抽象欲傳遞的概念，讓讀者透過文字去瞭解，這也是電子文件應設法能夠處理的。【註三】 語言文字只是表達這抽象概念的媒介，文件則是此概念映射在媒體上的形式 (form)。於是，欲了解文件的內容，必須經過以此形式來詮釋其內容的過程。然而這個詮釋的過程計算機並不會做，這是因為計算機本質上只是一個能處理形式的設備而已，對於抽象的內容而言，它是無能為力。於是，如何設法使計算機能夠盡可能輾轉地處理電子文件的內容，就成為今日研究發展的主要課題。

要解決上述的問題，最好的途徑是儘量找出文件的內容與形式之間的對應關係，此關係若能以形式表達出來，那麼計算機就可以用處理形式的方式間接地來處理內容了。

在人文社會的研究領域中，對於文件內容的研究一直是一個主要的課題，且其中不乏量化的形式可供計算機直接處理，然而也有許許多多頗有系統的研究成果，並沒有能夠表達為計算機能夠處理的形式，這也許是以往的研究成果主要是給人用的而不是給機器處理的緣故吧！在對文件內容的研究方面，語言學的方法是不可或缺的，這正是語言學的研究對電子文件的發展最能立見成效的著力之處。

一旦文件中一些形式和內容的映射關係釐清了，在電子文件中常常將此映射關係用標誌 (markup) 的形式表達在計算機中。換言之，電子文件目前已經有標誌這種有效的工具，來表達文件內容和形式之間的映射關係了。如果想要用計算機處理文件，了解標誌工具是必要的步驟。

了解電子文件的內容和形式之辨，就能知道計算機能為電子文件做些什麼事。由此我們也可知道怎麼樣改變原有研究習性，然後可以配合電子文件的性質，進而利用資訊技術來配合我們的工作。

### 三、分割與合併

文件在計算機中通常會被分割，形成一些適用於計算機處理的大小單元。【註四】 雖然電子文件還保留原件的結構和版面關係，然而經此切割卻衍生出電子文件的特殊結構生態。

首先，這些單元之間的關係不再僅限於傳統文件中的線性關係。在計算機中，這些單元很容易就產生多度空間狀的立體關係。從好的方面說，這些立體關係可以突破傳統的限制，更深刻地表達一些文件的內容，例如：可連接各單元間相關的內容構成超文件，以利使用者檢索、分析及學習；甚至可構成特殊的知識結構和認知網路作更多的用途。【註五】

然而，從不好的方面說，有兩個明顯的後遺症：其一、分割後的單元，就其內容而言很可能會失去一些資料，而不再是原先渾然一體的文件所表達的內容，這是由於內容之於形式的本性所引發的；其二、當各單元重新組合而不依照原來關係的次序呈現時，可能會產生些新的內容，而這些內容卻不一定是應該有的。換言之，可能產生一些雜訊、干擾、甚至於錯誤的信息。這些問題亦可視為文件中前後文（context）重新排列所造成的後果。若將許多文件置於一處，則各文件單元間彼此的干擾就會更明顯。【註六】

舉例來看，當文件以單元方式呈現時，代名詞或者是具有代名詞作用的詞彙，就可能產生不知所指或是誤指的理解。又如，即使是討論同樣主題的兩個單元，若原來的兩個文件寫作方式或理念不同，當它們放在前後相互參照時，可能會產生可笑的後果。這種情形在多媒體的情境下情況較為嚴重，這是因為各種媒體製作和詮釋的方式有明顯差異的緣故。

要解決本節所談電子文件分割與合併所帶來的問題，從語文學下手是必須的途徑。事實上，這些現象已影響到文件的寫作，也許電子文件的確須要比傳統更複雜的寫作技巧來忠實地維持電子文件的內容，而綜合認知心理學、傳播學（包括廣告學）和語言學的新寫作學似乎是一條解決的道路。

## 四、知識的增益

資料的匯集往往可導致新知的發現。例如：兩個沒有機密的普通文件湊在一起，就有可能產生機密的信息。計算機是匯集和整理資料的能手，文件若是轉換為電子形式，即可利用計算機來匯集和整理。事實上，經資料匯集整理而導致的新知並不僅限於發現一些現象而已，經常會導致新的理論和新的模式的產生。計算語言學之於語言學就是一個好例子，類似的例子還有許多，像是：計算物理學，或甚至如計算科學 (computational science) 這樣的通稱。【註七】

如果從文件可能涵蓋的知識內容來看，幾乎可以包容所有的學科。當這些文件匯集在相同的資訊設備架構下時，無形中縮短了彼此的距離，他們不僅使用相同的檢索語言、操作程序和機器設備，且相互提供了較以往更方便的擷取環境 (access)，更有甚者，在此環境下更易於建立彼此相關參照的關係。所以說這是在人文與科學久分之後，出現最可能使彼此將鴻溝消除的徵兆。

其實，電子文件本身就是人文和科技合作的產物。電子文件的處理必須有語言學的基礎也就是一種人文和科學的結合。日後電子文件普遍的使用下，語言學勢必將受重視，用以協助解決各行業間的資料匯集、整理和溝通等問題。

## 參、電子文件的檢索和處理

雖然電子文件的檢索和處理，目前仍是各行其道，然而由於大量電子文件的出現和許多新的應用如雨後春筍似地冒出來，在文件共享的理念下，造成了整合檢索和處理的大勢。在本節中，我們將從電子文件的檢索問題出發，來探討這整合的大勢，並將探討語文研究在此中可能扮演的角色。

### 一、電子文件的檢索環境

在1990年前後的電子文件檢索系統的設計、製作、使用的環境，約略可如圖一所示。由圖中可知，此系統的製作過程中涉及許多專業不同的人員。由於目前絕大多數系統檢索功能的設計都是以詞彙為檢索的鑰匙，例如：文件書目中之欄位、索引典、主題詞表、檢索詞表等均是，這些各行業的人員難免因對詞彙理解不同，使得整個系統成為一個相當複雜的語意情境。因此，由於語意差異而引起的檢索誤差，就成為這種系統效能的上限。比方說：文件分類的錯誤可能是由於編目人員對分類詞彙的認知，與制定圖書分類表的專家，在認知

上不同所引起的。又如：兩個不同的使用者若使用相同的詞彙作檢索時，他們心目中想找的文件卻極可能不相同。根據分析，使用者對於有限範圍的檢索詞表（controlled vocabulary）滿意程度經常在40%以下，這不是說該表品質不好，而是使用者對詞彙的理解常與專家不同的緣故。【註八】

造成這種情況的原因之一是在設計系統時，考慮不周和引用的知識不夠。到1990年時，通常被用來協助文件檢索的知識約可歸為三類，即是：詞彙上的（morphological），句法上的（syntactical）以及文件結構上的。【註九】句法上的知識在分析句子語法結構的功能上，常用在查詢句的解析，在檢索上則少有這樣的用法。句法的知識用於檢索時，最常用於判別詞彙的分類和解析構詞上的一些問題。文件結構的知識是用來標誌文件構成的元素，例如：篇、章、節、段等，以及標誌版面的訊息，諸如：頁次、欄次、行數等。由此可知，這種系統並沒有真正利用語言學的知識，以了解文件內容的方式來協助作文件檢索。事實上這種系統在設計時有一個相當強烈的假設，那就是：詞彙的意義是和文件的內容是相當一致的。因此，找到文件中含有某種詞彙的這一種形式，即表示找到了含有該內容的文件。這種假設該是多麼粗糙！

另一個設計上的問題是：工程師喜歡用量化的方式來做檢索或擷取方法的設計，但是，卻又不能詳細正確地說明，所用的數學計算在檢索上有什麼實際的意義。筆者贊成用數學方法，但是卻非常不同意在沒能了解其數學算式的實際意義之前，便冒然嘗試。這麼做有許多毛病，首先是這種系統無法明確地告訴使用者，找到的文獻是在什麼樣的情形下找出來；也沒有辦法說明，為什麼還有多少文件找不出來。在這些原因不明情形下，想要改善這種系統自然也就不容易。其次是這樣找出的文獻和文獻後繼的處理工作無法配合，因為文獻處理所需要的文件內容標誌，往往與為這種檢索所做的標誌是兩回事。換言之，也正因為這種檢索方法的設計忽略了對文件內容與形式之間的考量，以致於對其檢索中運用之計算無法作使用者可以了解的詮釋。例如：有許多檢索系統在檢索過程中，會經由計算將文件分為若干群聚（clusters），然而卻無法說明此類群聚的特徵和意義，像這種文件的群聚性質是無法用之於後繼文件處理上的。【註十】這類的例子在目前檢索系統中比比皆是。

為了要發展更成熟的文件檢索系統，就必須要克服上述的缺點。克服的辦法無他，只有多運用點能表示文件內容的知識。

## 二、專業知識之利用

前文已經談過，人文社會學科時常做文件的內容分析（content analysis）或是做文件主題和論點分析的（discourse analysis）。這些知識都是在1990年後發展文件檢索系統時極欲利用的知識。這些知識約可從語言學、心理學或認知科學，以及社會科學的三個角度來考量。【註十一】較之於1990年前引用的知

識，其範圍擴大了許多。從心理或認知的角度來考量時，其思考的方向偏重於文件在計算機中表達的方式。換言之，當一個文件表達於計算機中時，它的結構已不再僅僅是原來的結構了。這是一個文件知識的再建構(reconstruction)問題（請參照前文『分割與合併』一節）。在計算機中再建構的文件，涉及到使用者對文件之認知，也給文件的檢索設下既有的框框或限制，此中涉及的問題正是心理或認知科學方面要探討的要點。

從社會科學的角度來考量時，其重點放在系統涉及之人的因素上，尤其是針對使用者而發展。例如：使用者性別、使用的語言、教育程度、職業等所造成之差異。這兩方面的研究亦涉及語言學，諸如研究中常引用到心理語言學和社會語言學等。從語言學的角度考量此問題時，固然傳統的語言學仍有其發揮的空間，但是涉及應用語言學的份量卻與日俱增。例如：對寫作學、文獻學等知識的應用已日益受到重視。

這些從不同專業角度出發，把文件內容知識的詮釋表達在計算機中時，不僅對該專業處理文件甚有用處，也提供了嶄新的檢索機制。同時，這種做法可以使文件相關之知識，在計算機中是既可用之於檢索，又可用之於該專業對文件之進一步處理。換言之，這種做法把文件檢索的系統的功能擴大了，使檢索和文獻處理合而為一。這正是未來檢索系統發展的大勢所趨。

### 三、標誌的應用

在『內容與形式』一節裡，我們已經談到，標誌工具是將文件知識的內容和在文件中文字表達的形式之間，連繫映射關係的工具。現在，讓我們來說明如何運用標誌工具，將內容和形式在計算機中連繫在一起。【註十二】

要言之，工作的步驟如下：

1. 將欲利用的專業知識有系統地整理出它的形式。
2. 找出計算機中相對應的抽象知識結構。
3. 將1中之知識以2之結構形式表達在計算機中，成為此專業之通類知識 (generic knowledge)，可長久儲存備用。

以上三個步驟是將專業知識建於計算機中。當我們要分析文件的內容時，便可隨下列步驟用標誌的方式來完成。

4. 取出機讀文件檔案，並讀出欲利用之專業通類知識結構。
5. 用標誌工具，將文件中的相關文字和通類知識的對應項目連繫起來。
6. 重複5，直至所有欲標誌者做完為止。



通常在設計電子文件的檢索系統時，應考慮儘量用通類知識，這是希望所設計的系統能夠通用於某一類專業文獻的緣故。如果所用的知識太狹隘，譬如取之於某一本書中特有的知識，那麼所作的軟體就很可能就只能適用於這一本書而無法用之於其他的文件。目前，這種「只能用於某文件的軟體」的例子不是沒有，而是很多，實應引以為戒，不足以取。通類知識的例子很多，例如：圖書分類表、索引典、字書等等。

在人工智慧領域中，常將知識的結構形式用數學或邏輯的方式來表達，以便計算機能處理之。在此我們簡稱之為抽象知識結構，例如：樹（tree），紀錄（record），網（graph, net）及推理規則（inference rules）均是。在各專業中，對文件內容的知識，若能以抽象知識結構來表達，那麼計算機就可以處理與這知識相關的事務。如何將我們習以為常，在文件中隱而不顯的知識用抽象知識結構表達，的確是不太容易做的事。通常，體例的蒐集和整理是解決此問題的一個好方法。例如：俞樾在《古書疑義舉例》中將校書之經驗歸納出37條，可稱為通例，此即是表列式的通類知識，將之納入計算機中並非難事。

在上述的工作程序中，涉及語文的問題很多，例如：體例之整理就涉及句法及語意問題。又如：詞彙在不同專業領域中有語意之同異問題。凡此種種均待語言學家之努力。

今年四月，筆者曾將本節中所述之方法用於處理佛經中不同版本問題，欲觀此例者，請洽筆者。【註十三】

#### 四、寫作與檢索的整合

為了檢索文件的內容，我們希望把與文件內容相關的知識儘量整理出來利用。這些知識中許多是和寫作有關的，諸如在寫作時可能用到的大綱、細目、體裁、修辭、乃至於文章結構學和寫作學等等。美國的科技傳播學會（Society for Technical Communication）在1992年起就鼓勵會員用能作SGML標誌的文書軟體來寫作。其實，依文章的內容作檢索的這件事正好是寫作的反向程序，因此，寫作與檢索之間可以共用的知識甚多。如果設計寫作軟體工具（authoring tool）時，能同時考慮到日後文章檢索的需要，那麼對雙方都會有益。

可惜的是目前寫作軟體還不太成熟，所能做的事大部份都是協助作者編輯文句（editing）而已，真正和寫作相關的知識少得可憐，更不用說考慮到檢索之需了。但是，我們深信兩者相合是必然的趨勢，也是今日應該重視的方向。語文與寫作的關係密不可分，這個方向正是語言學可大力著手處，詳細關係實不須細說。

值得注意的倒是 TEI 計劃所付出的努力。TEI 是 Text Encoding Initiative 的簡稱，其意為文章編碼先鋒，它是計算機與人文協會(ACH, The Association for Computers and Humanities)，計算語言學協會(ACL, The Association for Computational Linguistics) 和文學與語言計算協會(ALLC, The Association for Literary and Linguistic Computing) 等三大組織聯合全球力量推動的計劃，迄今已有六年多了。在1990年，TEI 就發表了第一版的《電子文章編碼和交換規則》(Guidelines for Electronic Text Encoding and Interchange 亦簡稱 TEI)。1992年起，此文件的第二版文稿，TEI P2，就放在 INTERNET 網路上，由作者們共同寫作，隨時更新。這是一種新型聯合寫作的方式，文稿是動態的，隨時可能會添新資料；讀者也不必等到完稿、排版、印刷之後才能看到文件，而是隨時可上網路瀏覽已發表的部份文稿。

TEI P2 的主要工作是根據 SGML 標準，對各種文章體裁的結構和內容作標誌。其目的是促成電子文件的應用和共享，其應用中當然包括了文章寫作、檢索、排版...等等。1993年9月20日時，TEI P2 的工作內容如表一所示。由其中可知，像詩歌、散文、劇本、字典、信札等等都已是 TEI 既定的工作標的。相信 TEI 計劃的推展，對於整合文獻的寫作、檢索和處理有極重要的意義。

## 肆、結語

在這篇小文章裡，我們以電子文件為引子，以文件（或知識）共享的理想為軸心，來討論資訊技術的發展和語文在研究、應用之間的關係。電子文件之所以是個好引子，是因為它真能帶來了文件的革命，能讓典籍活出多彩多姿活潑潑的現代風貌。如果文件（或典籍）真正能做到人人平等地充份共享，那麼我們社會中知識的擁有、分配、散播、以及利用等的結構也會隨之改變，到那時，新的社會意識形態將會形成，而新的社會——資訊社會(information society) 才會真正到來。

由此可知，電子文件是時下資訊發展中重要的項目，而文件的表達是以語文為主要媒介，其間的關係密切不言可喻。由於電子媒體有其非常特殊的性質，是故會引發一些新的觀念、理論、做法和新的問題。在本文中，筆者嚐試說明這些觀念、做法和問題，然而，一來、茲事體大；二來、筆者才疏學淺，豈能盡道其詳？謬誤之處在所難免，尚祈指正。

綜合本文所述之重點可知，電子文件的發展對文章的結構、理解、寫作、檢索等各方面都會帶來改變，其主要的原因是電子文件帶來新的知識表達方式和表達的工具。在本文中，筆者嚐試著用古老的哲學命題：形式和內容的關係來說明知識結構與知識表達相關的問題，雖然所談的並不深入，並不嚴謹，卻希望藉此淺談，能說明這是一種很值得參考的思考方式。

由本文敘述之各端可知，在資訊發展的過程中，語言學的介入越來越深。在資訊發展初期，用一些簡單的語文計量或統計，就足以使資訊發展提升品質不少。之後，語言學幾乎全然投入到資訊發展的洪流中，遂有計算語言學之蓬勃發展。在近期的趨勢則是：凡是語言學應用面涉及的知識，亦逐一捲進資訊發展中，進而形成一股分久必合的大勢。在文中所談到的，諸如：寫作與檢索之結合，以及文章內容分析知識的綜合運用於檢索等，都指出了這一大勢已經存在。今後，應用語言學之發展，以及語言學和其他學科越來越密切結合的趨勢，應有更廣闊的空間和前途，是可預期的。

利用標誌工具，將與文章內容相關的通類知識和文章中的文句關係標誌出來，是把內容和形式相互映射關係表達出來的極好手段。它不僅簡單易行，適用於各種應用，並且是促成文件共享的必要手段。TEI的例子是個好榜樣，我們是不是也該發動相似的計劃呢？

電子媒體已經賦予文章新的形式，超文件和多媒體就是現成的例子。文章的結構已經不再只是停留於線性的階段，它已步入了多度空間的形式。文稿不再是靜態的了，也可以是動態的。處此情境，我們不禁會想到，電子媒體會帶來新文體嗎？新的寫作技巧嗎？新的文學理論嗎？……似乎有問不完的問題，而這種問題中，語言學將扮演什麼樣的角色呢？目前，我們似乎只能正視，關心這些問題，沒有辦法可以給個滿意的解答。

在強調知識共享及工作上合作的強烈整合大勢下，不禁使我們又聯想到人文和科學這「兩種文化」對立的老問題。筆者以為，現今的大勢正是消解對立的契機。也許是應驗了「解鈴還需繫鈴人」的諺語吧！人文與科學的對立仍須以科學工具來協助解決。事實上，廣義地說，這種次文化的衝突並不限於人文與科學之間，不同的學科之間，甚至於不同的媒體間(如文學之於電視)不是一樣有衝突嗎？解決這些衝突的唯一辦法是多多溝通，多多彼此了解，而電子文件帶來的新環境不是正好適合做這些事嗎？

在本文結束之前，請不要忘了，計算機只是一個邏輯的工具，資訊科技也不過只是處理形式的利器而已；然而，人文社會學科中，是不能不涉及感情、感覺和美學等這感性的一面的，電子文件對這等事將何以自處呢？坦白說，我不知道，只好請您想想吧！

## 誌謝

這篇小文章寫得有些匆忙，定稿的時間急迫，然而在助理張翠玲小姐的努力協助下，得以順利完成編打排印的工作，並且維持良好的品質，殊為不易，謹在此表達感謝之忱。在趕稿期間，家人亦頗體諒能給我一段清靜的期間，亦在此一併誌謝。

## 註 釋

- 【註一】請參閱該會議中：王三慶、彭昆侖、陳郁夫、施鐵民、羅鳳珠與朱四明等各位之論文。
- 【註二】自從美國副總統高爾提出 NII 計劃以來，各國爭相仿效，相關文件隨時可見，在 internet 上即可取得全部計劃之文件。在我國，行政院科技顧問組亦於今年六月十四日向行政院提出初步規劃草案，並將此議題列入今年科技顧問會議中討論。
- 【註三】此處之概念一詞是相當廣義的，凡是人的知覺意識能知之者均屬之。在人工智慧系統中，籠統地說，亦有泛稱之為知識者。
- 【註四】Gerard Salton, James Allan, Chris Buckley, Amit Singhal, 〈Automatic Analysis, Theme Generation, and Summarization of Machine-Readable Texts〉, Science Vol.264, June 1994, pp 1431-1426
- 【註五】Douglas C. Smith & Sandra J. Nelson, 〈Hypertext, an emerging and important medium of business and technical communication〉, Journal of Business and Technical Communication, vol.8, No.2, April 1994, pp 231-243
- 【註六】Geoffrey A. Cross, 〈Recontextualizing Writing, roles of written texts in multiple media communications〉, Journal of Business and Technical Communication, Vol.8, No.2, April 1994, pp 212-230
- 【註七】Computational Science 一詞最早由諾貝爾獎得主 Ken Wilson 使用，意指一切以計算機為主要研究工具的學門。另可參考：Abmed Sameb & John Riganati, 〈Computational Science and Engineering〉 Computer, October 1993, pp 8-12
- 【註八】William Saffady, 《Text Storage and Retrieval Systems》Mackler, 1989
- 【註九】Gerard Salton, 《Automatic Text Processing》Addison-Wesley, 1989
- 【註十】Faxli Can & Esen A. Ozkarahan, 〈Computation of Term/Document Discrimination Values by Use of the Cover Coefficient Concept〉, J. ASIS Vol.38, No.3, 1987, pp 171-183
- 【註十一】此處的這三個角度，是依據【註五】中的說法。事實上這種考量的角度是可再議的。譬如，它貶低了寫作學和文獻學等的影響。
- 【註十二】此處所指的標誌，是依據國際標準，〈ISO 8879, Standard Generalized Markup Language, SGML〉，1986年中所訂定者。此外TEI P2（如下文），以及 HTML（Hyper Text Markup Language，在 Internet 的 Gopher 上即可找到），亦值得參考。
- 【註十三】謝清俊、陳昭珍、莊德明、周亞民〈電子佛典中處理版本的方法〉，中央研究院資訊所，1994年4月