

中文電子文獻系統的現況與展望（稿）

Chinese Electronic Document Systems, present
status and perspectives (draft)

丁之侃，謝清俊

中華民國七十七年五月

中央研究院資訊科學研究所

英文摘要

Using computer to process Chinese text document is a newly developing area in Taiwan, ROC. A survey of developing tools and systems under developing in this area will be presented. Topics discussed will emphasis on the problem of document representation in computer and the problem of document retrieval. Some suggestions on the directions of future development are also presented.

中文摘要

本文之目的在報導目前電腦處理中文文獻之現況，並檢討目前面臨之問題，以及探討今後發展之方向。文中所討論之文獻，主要是以文字表達者；而對問題之討論，則偏重在電腦中文獻表達之方式，以及對文獻尋取的方法。討論之範圍則包括：發展中文文獻系統之工具，中文文獻系統之應用，目前之問題和研究發展的方向，以及對今後發展的一些建議。

一．前言

處理資訊的問題由來已久，它伴隨著人類的文明日益複雜。在最近的四十年，藉著電腦、通訊設備和各式各樣的新技術，它開拓了廣大的領域：從處理的內容來看，由數據的處理（data processing），至資訊的處理（information processing），而至於知識的處理（knowledge processing）；從訊息表達的形態上而言，包括了數、文字、語音、圖形、影像、等等形式；從結構方面來說，由簡而繁，由單純至成熟。以文字表達的訊息作例子，早先是字或詞的問題，進而處理表格形式的資料，再發展為處理非格式化的字串，自然語文中的片語、句子仍至於結構化的文獻。本文談論的，就是目前國內以電腦來處理中文文獻的現況和問題，其內容包括：

- ． 中文電子文獻系統概說。
- ． 發展中文文獻系統的環境和工具。
- ． 中文電子文獻系統的應用情形。
- ． 對中文電子文獻系統未來發展的建議。

由於此領域在國內僅具雛型，對於許多正在進行中，要取得確實且合時宜之資料頗不容易，所以疏漏之處在所難免，然而，若能藉此文引發更多的興趣，加強同好間之溝通或激起更多的迴響，則是作者拋磚引玉的心願。

二．中文電子文獻概說

廣義的文獻是泛指留存在媒體上的記錄。媒體以紙最為普通，記錄的形式則以文字最為普遍。因此，狹義的文獻多指在紙上以文字記錄的資料，然而亦因適用的範圍不同而致在不同的地方其義意有許多細緻的區別。文獻所記載的資訊除直敘性的文字以外，還有許不同的格式，譬如數字、公式、表格、註解等等；亦有非文字的資訊，譬如圖案，影像、等等。在這許多不同的表達方式之中，有些是和語言文字無關的，它不受國家或地理上的限制，然而文字記載的卻和我們使用的語言息息相關。因此，本文所談者，以字記載的文獻為主。對於和語言文字無關者不在本文討論之內。本文不討論它們不是表示它們在領域中不重要，而是因為一則處理它們的技術和中文的關係不是那麼密切，二則是國內目前幾無綜合語文、圖形、影像、聲音等文獻處理的研究。至於語音方面，它雖與我們的語言關係密切，且國內研究成積也很好，然而它的性質和文字表達的資訊相去甚遠，故本文亦不討論語音涉及的問題。

所謂電子文獻是指可以用電子設備處理的文獻。在諸多電子設備中，以電腦的功能最為引人注目，是故本文所指的電子文獻是指電腦系統能處理的文獻為主，而所謂的處理亦以電腦能做的事為限。即就如此界定，文獻的範圍依然相當廣泛；它可以是書信、公文、書本、雜誌、報紙、詩詞等等，在處理方面，我們不談涉及中文語文的基本問題，譬如：輸入、字集、編碼、字形產生、輸出等等，因為這些問題雖然均仍有待研究改進，卻是大多已有產品沿用多年。本文將重點放在電腦中文獻的表達方式（representation）和文獻的尋取的方法上。本文所談的多圍繞著這兩個問題討論。

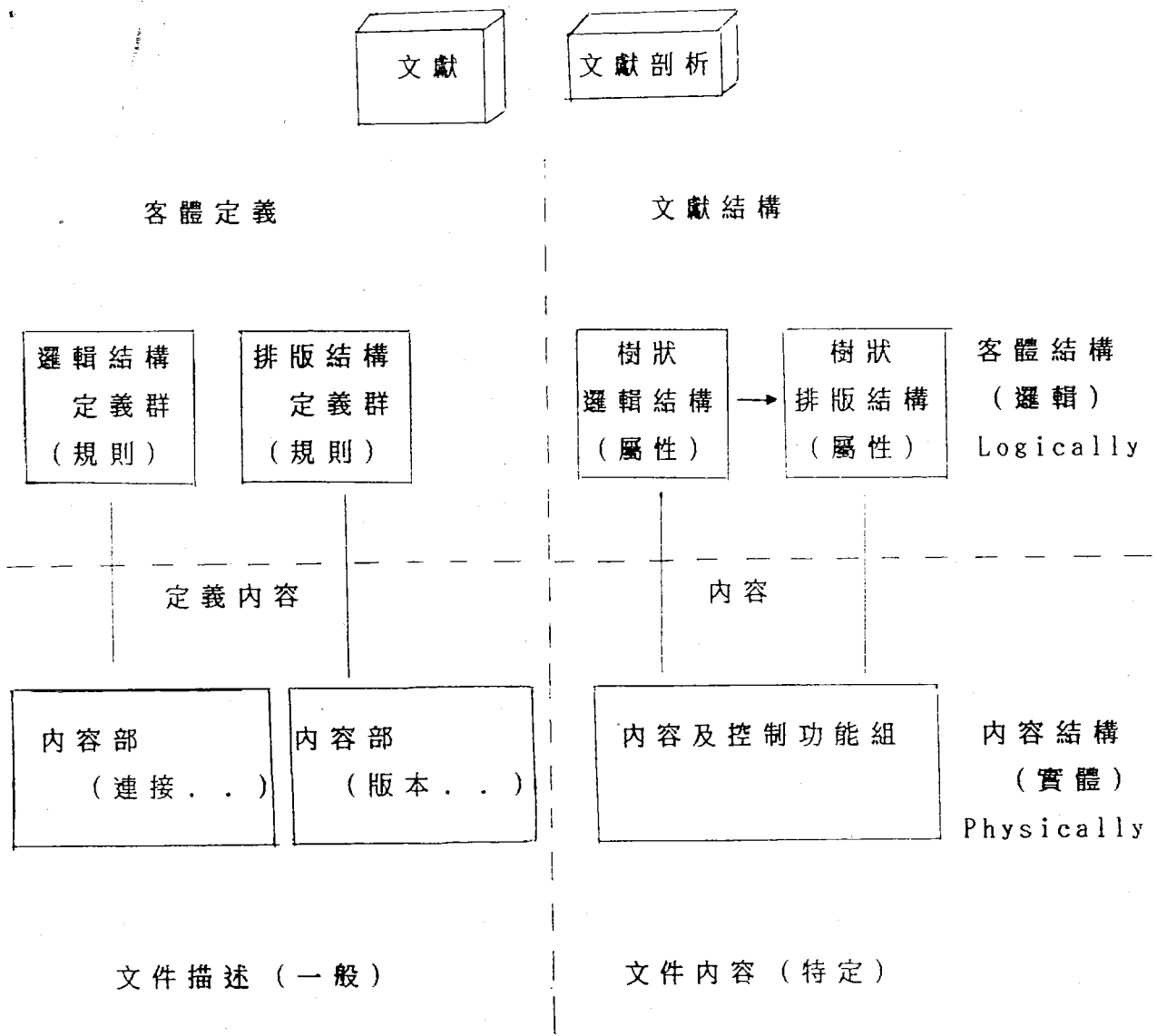
(一) . 電子文獻的表達方式

文獻在結構上，並無中外語文上的差異。因此中文文獻在電腦中的表達方式，可以參考國際標準組織（ISO）所發表的資料。在圖一中所表示的文獻表達模式是中1985年ISO DP8613的標準：Information Proocersing-Text preparation and Interchange中的文獻結構部份所引來的。

根據這個模式，一個文獻可由文獻結構（document structure），客體定義（object definition），和文獻描述（document profile）三個部份表達。所謂文獻結構是指文獻的抽象組織，一般而言，文獻多呈樹狀的結構，然而亦有序列的清單（orderd list），非序列的清單（unordered list），以及一些主文外的說明文字，例如註解，等屬從結構的變化。文獻的抽象結構不只一種。例如文獻的目錄是一種結構，而文獻的編排是另一個結構，而且此二者之間還有關連，可用一些規則來表達。

在客體定義上，則是把抽象的結構賦序對某特定文獻實物上的定義，例如樹狀結構可以表示章回，亦可表示章節段落的結構。此二者之間，亦可以規則（rules）表達相互之關係或之轉換（transformation）。在一個通用的抽象結構下，其客體結構可能是一本書，一篇報告，一個備忘錄，甚至一封信、一個合約。這些實際的文獻都在客體定義上表達——只要它們有共同的抽象結構。

關於文獻描述部份是指此文獻相關的一些訊息，舉例說，若文獻是一本書，這本書的書目記錄就可作它的文獻描述。事實上，此部份還可包括文獻之間的關係，如分類，參考文獻，等等。



圖一．文獻結構模式

這個 ISO DP8613 的標準，部份被歐州電腦製造聯盟 (European Computer Manufacturers Association ECMA) 採用在 ECMA101 標準中。

另一個相關的標準是 ISO 8879-1986 (E)，這是一個通用的標誌語言標準，Standard Generalized Markup Language，在這個標準裡，訂定了標誌各種文獻中構成元素的規則，經由這個標誌語言的解析器 (parser)，可以讓電腦認出每一個文獻元素，以供後繼的各種處理工作之用。這些標準的訂定是為了在各種系統中分享資訊，譬如，它可用在文字處理 (Word processing)，編輯 (editors)，列印及排版 (typesetting)，以及各種資訊檢索系統 (information retrieval systems) 等等均可以用統一的標誌語言可以達到資訊分享的效果。

(二) . 中文文獻的檢索方法

在處理欄位性質 (格式化) 的資料時，偶爾就已經觸及了文獻處理的問題。例如，書目紀錄中的書名欄，或是摘要欄，戶籍資料中的記事欄、等等皆是。這些欄位是一些長短不同的文字敘述，具一般文獻的性質。然而由於它們的資料量比較少，裁截成固定長度後，約略可視為欄位化資料處理，可是卻須配以特殊的功能，譬如字串比對 (string matching)，或是用人工作詞的索引 (word index) 來協助檢索。早期的文獻檢索多沿用此法，雖然整個文獻都已變成機器的形式 (機器可以處理的形態)，可是在檢索上卻仍需借重欄位化的工具—將檢索的詞彙作成欄位化的索引，來作為檢索的途徑。

除上述兩種全文檢索的方法以外，在國外還有利用文獻特性所發展的 Signature 方法和 Hashing 方法，這些方法多多少少和語文有關，換言之，它們無法直接用於中文的文獻檢索。此外利用一群文獻間關係的 Clustering 方法和 Browsing 方法等等亦可提供更多的檢索選擇。

若是根據前一節的文獻表達模式，文獻結構本身就可成為一種檢索方法，而文獻描述的資料亦能提供另一個檢索途徑。

三．發展中文電子文獻系統的環境與工具

現行的工具在國內有王安電腦公司 (WANG INDUSTRIAL CO.) 引進 Computer Power 公司的 status 全文檢索軟體，可以解決一部份文獻系統的需求；此外王安公司利用 PACE 關聯式資料庫組合成 "王安整合影像系統 (WIIS)" 可以貯存圖像資料 (Image Text)。美商漢維公司 (Bull) 代理 MISTRAL 全文檢索軟體，可以處理非結構化的文獻 (Unstructured Text)，並與結構化的資料 (如索引典 Thesaurus) 相結合；圖像資料可離線式地存在工作站中，視需要而叫出顯示。寶來公司 (Burroughs) 提供 TEXT-TRIEVE Document Retrieval System 在其 B5000/6000/7000 系統的主機上，提供信件，備忘錄，產品說明，合約，法規等不同的文件型態應用。DATA POINT 公司也有 VISTA-FINDER 軟體提文件內文的找尋及相關的動作如列印，複製等等。VERITY 公司的 TOPIC 全文查詢系統可以與傳統的關聯式資料庫管理系統連用，具備自由詞 (free-text) 查詢等基本功能。國際商業機器公司 (IBM) 也有一個在美國全文使用很多的系統 STAIRS，具有 Automatic word indexing 功能。

以上工具，除了王安利用中文編碼及特殊終端機外，都有中文應用上的問題，比如說中文終端顯示採用二位元組碼會與系統控制碼衝撞，且大部分對文獻結構上的描述都不夠；對於某種特殊文獻必需外加以程式設計。嚴格地說，對於中文電子文獻系統而言皆非理想的工具。

特別一提的是，國人自資的虹成公司自行在個人電腦上為中文環境上設計了類似全文查詢工具如巧書僮，自動調檔系統等，爭取速度快，容量不大的使用需求。中央研究院計算中心與歷史語言研究所的「史籍自動化」計劃中，計算中心設計了一個系統 CTP (Chinese Text Processoe)，嘗試用 ISO 標準的精神，架構出一個通用的中文文獻處理系統，也是國內唯一一套將文獻中邏輯結構與排版結構分開處理的精神發揮出來的系統，也是一套具有標準規範標誌語言 (Generalized Mark-up Language) 處理能力的系統。

分析項目 名稱	使用字碼	使用系統 (硬, 軟體)	檔案架構	查詢語言 及HOST LANGUAGE	適用範圍	採用標準 ISO 8613
TUS (王安)	中, 英文 (王安碼)	VS series VS/OS	Character Inverted with space	Custom-designed	管理性 敘述性 文件	NO
TRAL	英, 法文	DPS 7000 GCOS 7	Fully Inverted +Formatted	COBOL	書目性 摘要性 文件	NO
T-TRIEVE	英文	B5000 6000 7000 MCP3.2	Inverted by Assignment	INTERACTIVE	敘述性 摘要性 文件	NO
TA-FINDER	英文	RMS	Inverted	INTERACTIVE	敘述性 摘要性 文件	NO
IC VERITY	英文	MS-DOS Link with RDBMS	Inverted +Formatted	INTERACTIVE COBOL, C	格式化 摘要性 文件	NO
AIRS	英文	Mainframe 43 30 VM/OS	Auto-indexed ISAM	COBOL, ASSEMBLY	格式化 書目性 文件	NO
會全文檢索軟	中, 英文	PC XT/AT DOS	WORD Inverted CHARACTER BASED	INTERACTIVE	條文化 文件	
(中研院, 計 算中心)	中, 英文	UNIX 5.0	HYPER-TREE	INTERACTIVE QUERY LANGUAGE	全文性 敘述性 文件	部份採用

表一中文電子文獻系統發展工具

四．中文電子文獻系統的應用

在國外，利用已有的工具或自行發展的文獻處理系統非常多，如法國羅浮宮採用MISTRAL系統管理其院藏，利用鍵語安排搜尋，必要時可將預先已貯存在光碟內的圖像資料輔助性的顯示出來。DIALOG中的全文資料庫，BRS，ORBIT中也有相似的資料庫，有些是採用STATUS，另一些則是採用STAIRS。參考資料中列有詳細的資料庫內容供參考。

在國內，華興資訊的COIRES套裝軟體是採用IBM STAIRS的原則而發展的。鼎盛公司則利用PICK OS的特性及McDonnell Douglas的「資料庫機」也具有初步處理中文文獻的功能。

資訊工業策進會為法規會設計了一套法律文獻的全文檢索系統，技術上採用單字倒轉檔的方法（Inverted file），建立條文的索引。查詢時鍵入所查的法規語辭後，系統即將所鍵入語詞分解成單字，利用所建的索引先找出具有所查每一個字的法條。再剔除語辭不合者，顯示出符合查詢條件的法條。

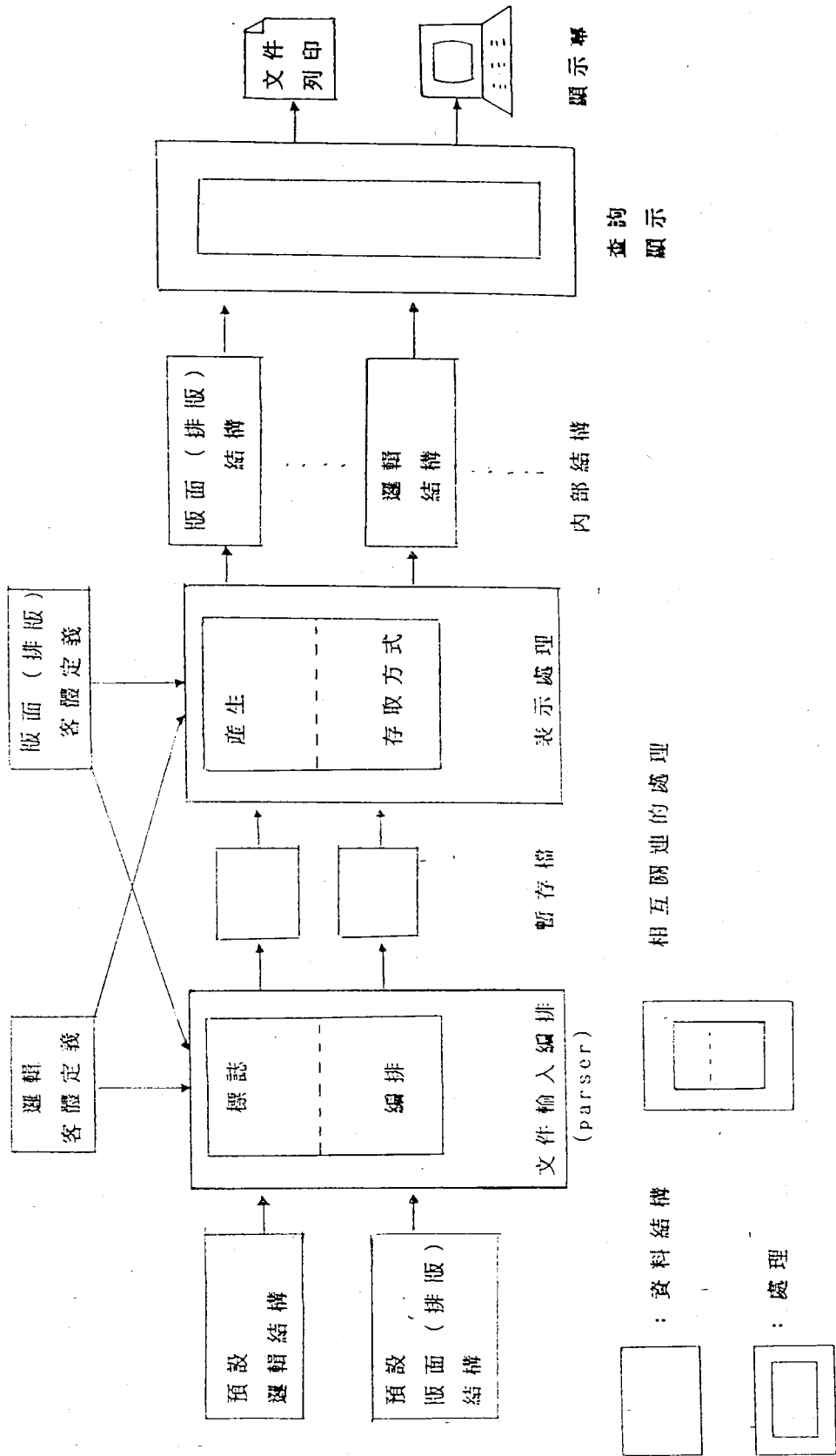
表二是以上系統的對照表。除了這些系統此外，市面上還有些電子辭典、唐詩、宋詞、聖經．．．等等商品在賣，這些系統的資料量表列的系統大，從應用的觀點來看，這是一個頗值得開發的領域。從技術觀點而，若資料量不大，以目前已有的軟體工具，如資料庫管理系統，檔案系統語言編譯器及螢幕畫面管理軟體等等中文屬理能力，要將一文獻自動化，已是許多公司可以做到的事情。

功能特性 系統名稱	輸入情況	結構要求	存取方式	查詢方式 或資料結構	資料內容及現況 (Retrival)
行政院法規會 法規查詢系統	現成Editor 如PE2, edlin 批次, 線上建檔		反轉檔型式	分類, 自由 詞查面導向	法規會3000種可 擴充至非條列式 資料
立法院 立法諮詢系統	簡單Mark 《》控制詞彙 批次建檔	編, 章, 節, 項 款, 目, 版 不須排版	書目檔型式 (反轉檔)	條文要旨查詢 自由查詢 自由查詢 自由查詢 自由查詢	六法全書內 55萬字及 系統建置 (招標時)
台北市議會 中文法規資訊 查詢系統	批次, 線上建檔	半結構化	書目檔型式	自由詞查詢 索引及關鍵語 分類	台北市單行法規 (44萬字) 台北市各行政命令 (120萬字) 台北市議會內規 (1.6萬字) 系統簽約中
中央研究院 史籍自動化	中文標誌 語言分析 批次, 線上建檔	邏輯結構 排版結構	全文搜尋法 元搜尋法 元特性法	書面指引導向 查詢語言查詢 自由查詢 自由查詢	正史 (史記、漢書、後漢書、 三國志) 及專業經濟史 (擴 644萬字) 77.6使用現擴 充中

表二 現有系統功能特性表

五. 中文電子文獻系統的發展

電子文獻處理方式如圖二所示，圖二的資料亦取材自 ISO DP 8613。根據圖二的模式，文獻處理相關的作業，可以分為文件輸入編排，內部結構表示及查詢顯示三種主要的處理過程，本節將從這三個方向來探求現在的研究發展方向及情況。



圖二 電子文獻處理方式

符號 (mark-up symbol) 指示出來，這些符號之間的關連則是標誌語法 (grammar rules)。ISO 8879 提出了標誌語言應是一種描述語言 (descriptive type language) 有別於一般程式設計所用的程序語言 (procedure type language)，它的訂定應視處理文獻的特性，搜集整理後，儘量設計成一般化 (generalized) 以適應大多數不同類別的文獻。

在國外，由於不同種類的文獻經由不同系統如 UNIX 的 TROFF, TEX, DOS 下的 postscript 編排輸入後，格式不統一，交換文獻產生了困難，遂有 ISO 8879 的想法，這是個自然的演變及趨勢，國外各業界也向文獻處理的第一關文獻輸入編排的統一標準邁進。

在國內，文件輸入編排處理方式類以英文文書處理的有 Wordstar (中文)，排版處理的有“雅墨”等，它們採用的大都是 procedure-type 的標誌符號。中央研究院嘗試以古代典籍編排方式及結構，歸納出一個中文標誌語言 (Chinese Markup Language, 簡稱 CML)，這個語言分為宣稱部及程序部二大部分；試圖將條列式文獻，書卷式文獻，單篇式文獻，公文式文獻 (固定型態一三段式) 等訂定出一個一致的文獻輸入編排。

相關的研究有此種語言的解析翻譯器如何構成 (parser, compiler)，因為這種翻譯器結構很顯然地與一般程式語言翻譯器 (ex. C-compiler) 有極大的不同，在這方面國內外相關報告還未出現。

(一) 文件輸入編排

在 ISO 8613 及 8879 二個標準都提到了標誌語言 (mark-up language) 的重要性，它可以將文獻中必需處理的語義 (semantic) 連接，結構的安排，特殊處理的效果等藉由標誌

(二) 內部結構表示

內部結構表示包含文獻表達的型式 (model) 及存取方式 (access method) 前者很顯然 (自然) 地應是階層式 (hierarchical) 的形態但有別於傳統資料庫管理系統中 hierarchical model。國外報告顯示出文獻結構型式朝向人工智慧型式 (如 semantic net, entity-relationship) 方向發展, 另一方向則是朝向混合文件型式 (hypertext) 發展。在國內這方面沒有太大的研究。唯有中央研究院計算中心提出一個基本的樹狀結構, 配合文獻基本的型式及 UNIX 系統的樹狀檔案結構而構成。

關於存取方式一般而言據 FALOUTSOS 氏在 1985 的調查報告可分為反轉串列法 (inverted-List), 全文搜尋法 (full-text scanning), 特性法 (signature or hashing) 及叢聚法 (cluster) 四大類。各有各的優缺點及適用範圍。

反轉串列法, 這是一種最簡單的方法, 將要被查詢的文獻依其控制詞, 自由詞 (free-text), 甚至於是單字依某種順序 (內碼, 出現頻率. . .) 排列及串列即可。王安是逐字反轉 (word inverted), 資策會的法則系統也類似; 中研院早期的食貨志查詢也是將控制詞 (專業標記) 反轉串列而成。這個方法是簡單, 快速且可以有 automatic indexed 的功能; 所以一般的工具 (STATUS, MISTRAL) 或最初接觸文獻處理的系統或計劃都會選擇此種方式, 唯此法在串列的維護 (加減資料), 空間存量及文獻階層結構的特性表現上, 出現問題。此外此法在下一小節的查詢顯示中有些基本功能必須再次處理 (post processing) 才能適應要求。參考資料沒有列出是因為資料太多且散見在各資料結構的教科書及其相關文獻。

全文搜尋法，這是一個唯一沒有經先處理（preprocessing）過程的存取方式，文獻的內容存在貯存體內，視查詢語句及範圍利用機器的速度逐一比對（matching）。關於對比之實務一般皆用Aho氏的有限狀態自動機（finite state automata），此種方法空間最省，唯查詢速度與查詢量的大小及查詢詞的多少成正比。在國外，此種方式會與其他方式相結合做為小範圍，再次處理的一種方式，或直接用硬體來處理，如美國專利局的計劃。在國內此種方式僅有中研院的史籍自動化在研究及使用。

特性法，在國外由FALOUTSOS氏倡議此用，此法的特點在於預先將文獻的特性（signature）如字碼的安排，字碼的結構等利用數學計算（hashing function）做成特性檔（signature file）。欲查詢的的詢問語也經由同樣的數學計算與特性檔匹配而得。它完全靠調合精確率（precision ratio）叫回率（recall ratio）來找到最佳組合（fine turned）。在國外利用英文的特性已有數十種ranking的技巧來達到預設的效率。相關文獻可參考費及RO文。國內相關研究不多，僅有中大劉寶鈞及中研院謝清俊等二個小組在研究中文的文字，文件組成特性。

叢聚法，此法由SALTON氏大力提倡，原意在於能否從文獻中，相關段落間找出較具代表性的含義作檢索：想這麼做的理由主要是圖書館界對文獻分類時發覺有一個叢聚現象：那就是相關文獻它們所用的文字差異應不大，所以可歸為一類以利檢索，它與特性法最不同的一點是：特性法是儘量找到文件的單一性（unification），而提高其鑑別率（discrimination ratio）；而叢聚法則是儘量找到相關文獻而提高其叫回率（recall ratio）。國外叢聚法應用在書目性的文獻較多。國內農資中心的農學索引典補助工具可勉強算是類似的研究了。中研院詞庫計劃中有關詞類分析，整理所得的「電子辭典」構架也部分採用此法的精神。

(三) 查詢顯示

文獻查詢語言的設計仍是一個值得研究的問題。主要是它不像傳統式資料庫查詢語言 (query language) 有一套嚴謹的邏輯式表示 (AND, OR, SELECT . . .)。在數學上可以證明邏輯式是完整的 (completed) 所以根據其設計出的查詢語言也是完整的。而文獻的查詢語言的性質就不是這樣，現階段無法範圍的限定，字詞的限定 (wild card...) 等。因此，文獻查詢語言不能由傳統資料庫的查詢語言來代替使用。

DIALOG-base 的查詢語言是以指令行 (command line) 的形式為主；當然使用者 (end-user) 仍喜歡友善型 (friendly) 的交談式語言。對系統使用者而言，仿效 SQL 訂定一套文獻式的 query language 仍是國內外一可追逐的目標。

在國內外，人機介面交談式查詢設計皆朝向使用者需要而發展，也大都滿足使用者需求，更何況假以時日及投入人力，這些查詢需求只要不是變動內部結構表示的話，總是可以完成的。

在國內，文獻式 query 以 command 的形式為主，然而中央研究院計算中心林晰嘗試用 prefix-type query language 來描述使用者的邏輯查詢程序，並將其轉成內部存取模式。這個方法在史籍自動化計劃所發展的25史全文資料庫中應用得很成功，使用的人都認為它是相當「友善」的系統。

在顯示部分，由於中文有直列印的現實需要，故利用終端機控制碼來控制垂直顯示的效果不錯，至於底線，加強背景，閃爍提示等功能，國內外皆有一致的水準。

文獻中若有非文字型 (non-character type) 的資料，如圖表，國外的做法之一是混合顯示，如美馬里蘭大學的 Hypertext on Hypertext。它利用終端機的 graphic mode 來做大略的顯示。另一種是分別顯示，圖表用較高顯像度的顯示幕顯示出來，如 MISTRAL, WIIS。至於更高精密的影像處理則牽涉到電腦圖學 (computer graphics) 及影像處理 (image processing) 超出了本文涵蓋的範圍。

在國內，顯示功能有求於硬體設備的配合，在這方面沒有進一步的研究成果。

六 展望與建議

由以上的情形看來，在國內對電子文獻系統的需求會日益殷切。在辦公室自動化方面和在資料庫的應用方面將是電子文獻系統的兩大應用領域，由於電子文獻系統應用的領域如此廣大，制度共享資訊的相關標準是國內資訊界應有的共識，由 ISO 制定 ISO DP 8613 和 ISO 8879 等標準的情形來看亦可說明制定這方面標準的必須性，可資國內借鏡。這些標準和其他的標準不同，因為它涉及我們的語言文字，所以不能全部移植國內，我們必須以自己的學識智慧，訂定相關的標準。

由日本近兩年文獻處理相關的軟體來看，他們推動文獻表達標準已有相當成效，這是連歐美都不如的。這些標準不僅給廠商帶來方便和市場，更給使用人莫大的好處，他們不須學許許多多不同的操作來使用各種文獻處理的軟體，因為這些軟體均用同一的表達模式和標誌語言，甚至連操作的鍵也頗一致。這些也是我們可以學習的。

其實，發展影響大軟體系統，必須要有國家級統一策略的機構推動。譬如，為推動資料庫的應用，日本有政府的白皮書說明資料庫在日本應用和發展的狀況，以及政府推動的政策。對於電子文獻系統而言，其重要性不亞於資料庫，由上述之 ISO 標準及日本的實施情形而言，就是前車之鑑。反觀國內，這種事由誰來做？似乎還沒有人關心。

從技術的角度來看，國內從事此方面研究發展的人口太少，留有許多問題待解決，譬如：

- 對中文文獻在電腦中表達方法的研究
- 對中文文獻描述語言的研究及編譯器之設計
- 對中文文獻存取方法的研究和效能評估
- 對中文文獻查詢語言的設計與實務
- 對於新貯存媒體，譬如光碟之用於文字、聲音，以及影像之貯存。
- 對於多媒體電子文獻系統的研究
- 對於電子文獻系統的應用方面，諸如文獻資料庫的開發，運作，使用等等涉及的管理問題亦待探討。

此外，文獻處理和自然語文處理的結合是未來必然的趨勢，也是真正結合文字、語言、語意等等處理的基礎，亦為智識表達的途徑之一，因此，往遠處看，發展中文的計算語言學（Chinese computational linguistics）也是未來發展中文電子文獻系統必須奠定的基礎。

一般參考資料

1. W. Horak, "Office Document Architecture and office Document Interchange Formats : Current Status of International Standardization", Computer, Oct, 1985
PP. 50~60
2. ISO DP 8613, "Information Processing-Text Preparation and Interchange" V01 1,2,3, (Part 1~6) , 1985。
3. WANG LAB, "STATUS-Text Information storage and Retrieval System", Technical Report , 1985
4. Computer Power, "STATUS-General Information Guide", Technical Report, 1983

5. BULL, "MISTRAL: Information Retrieval System"
Reference Guide, 1986
6. Burroughs, "TEXT-TRIEVE document retrieval system
Capabilities Manual, Aug, 1982
7. Data Point, "Using VISTA-finder", user's Guide, 1985
8. VERITY, "TOPIC Technical Qverview ", 1988
9. SALTON, G. "Dynamic Information and Library Prolessing
Prolessing", Prentice-Hall, Englewood Cliffs, 1975
10. 國科會科學技術中心, "科技性全國資訊網路", 操作使用手冊
, STIC-UM-011-0002, DEC, 28, 1988, 台北
11. 陳悅表, "我國法律全文檢索系統的設計與開發", 資訊與電腦
雜誌, Jan, 1988, 台北
12. 沙燕琪, "全文檢索統比較", 資訊傳真系統版, May, 1988,
台北
13. 謝清俊等, "全文資料庫專輯", 科學月刊, Vol, 19, NO.4
April, 1988, 台北
14. ISO DIS 8879, "Information Processing-TSXT and office
System-standard Generalized Markup Language (SGML) ",
1986

15. 謝清俊等, "中文全文處理系統", 中央研究院計算中心, 技術報告, 1986, 台北南港
16. Faloutsos, C. "Access Methods for text", ACM Computing Surveys 17, 1 (March) , PP49-74, 1985
17. Aho and Corasick, "Efficient String Matching: An Aid to Bibliographic Search, "CACM 18, 6 (June) , PP333-340, 1975
18. RO, "An Evaluation of the Applicability of Ranking Algorithms of improve the effectiveness of Full-text Retrieval", I & II JASIS 39 (2) (3) ,1988