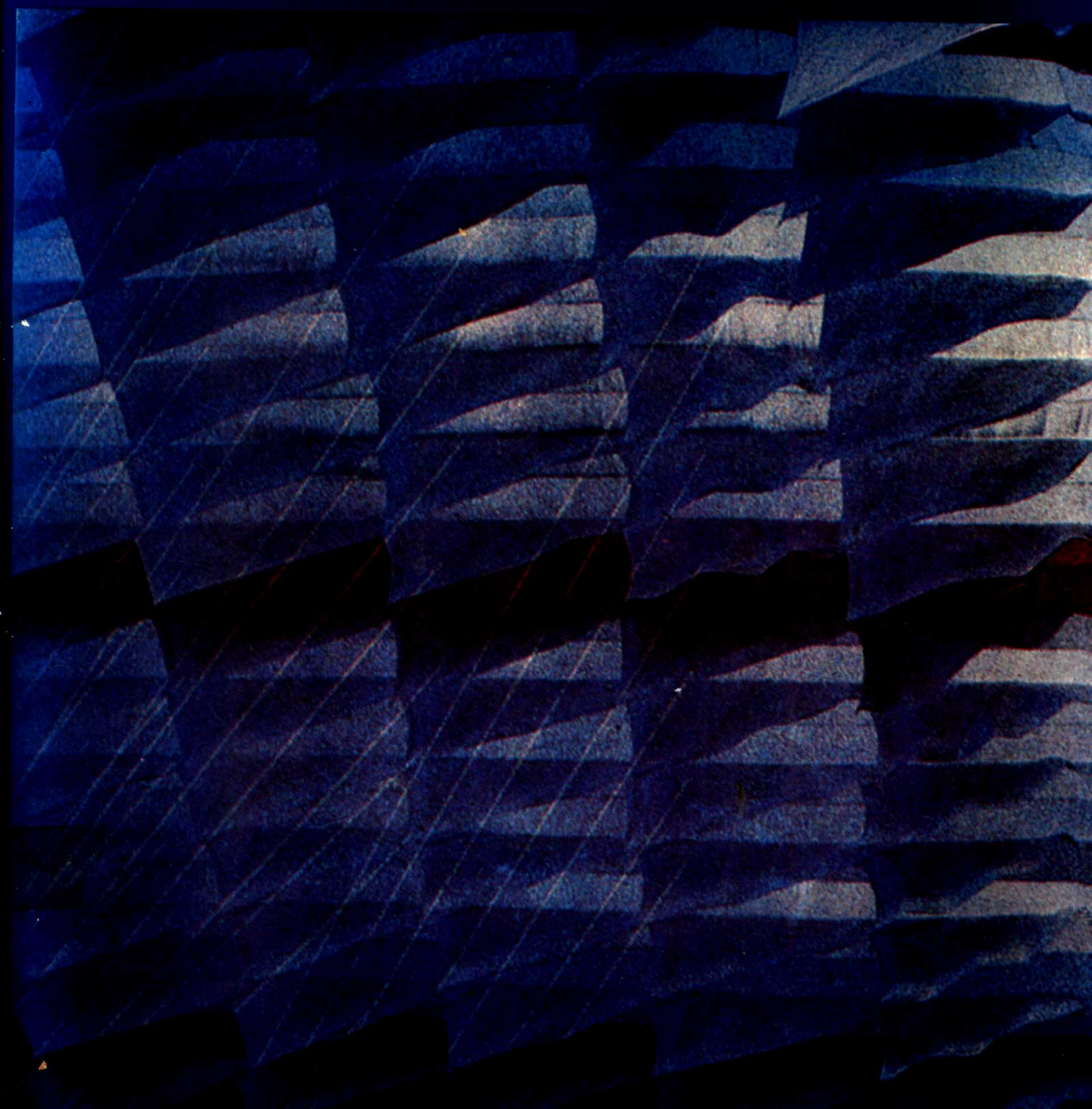


中文電腦資料處理

馬建白・蕭茜寧・陳聰明 編著





中文電腦 資料處理

馬建白·蕭茜寧·陳聰明 編著

中文電腦資料處理

中華民國 74 年 6 月初版

編著者 馬建白·蕭茜寧·陳聰明

發行人 楊 榮 川

發行所 五南圖書出版公司

局版臺業字第 0598 號

臺北市銅山街 1 號

電話：3 9 1 6 5 4 2

郵政劃撥：0 1 0 6 8 9 5-3

印刷所 茂榮印刷事業有限公司

臺北縣三重市重新路五段 632 號

電話：9711628 · 9713227

售 價 300 元

(本書如有缺頁或倒裝，本公司負責換新)

新時代電腦文庫

總 主 編

范 光 陵 博士

美 國 斯 頓 蒙 大 學 企 管 碩 士
美 國 猶 他 州 立 大 學 哲 學 博 士
美 國 哥 倫 比 亞 大 學 超 博 士 研 究 生
美 國 加 州 大 學 院 士

會 任

美 國 IBM 電 腦 公 司 系 統 分 析 師
美 國 猶 他 州 立 大 學 電 腦 研 究 計 劃 主 任
美 國 加 州 大 學 教 授
美 國 立 中 興 大 學 企 管 系 系 主 任
達 甲 大 學 電 腦 系 系 主 任
文 化 大 學 電 腦 研 究 所 所 長
國 立 成 功 大 學 商 學 院 院 長

榮 膺

美 國 傑 出 學 者 獎
美 國 國 際 傑 出 人 物 獎
泰 國 電 腦 成 就 獎

現 任

國 立 中 興 大 學 教 授

新時代電腦文庫

總主編序

十年前我在美國哥倫比亞大學作超博士研究，主攻人機模控學——也就是研究人類和機器間，如何彼此模仿及有效控制的一門新科學時，接到德國國際會議中心的電話，要我擔任該會舉辦之國際研討會講座。會議是在西德的柏林市舉行。參加者有世界各國電腦專家多位。擔任講座的有美國、日本、奧國、加拿大、意大利、蘇聯、德國、英國、法國等電腦界人士。

閉幕的前一天晚上，大會執行長華特博士在高聲雲霄的自由之針上的旋轉廳，請全體講座吃德國南部名菜豬蹄，並用巨杯喝慕尼黑啤酒。酒過三巡，歌唱十遍之後，華博士說：「全世界都希望聽聽各位的高見，究竟十年後電腦會如何？廿年後會如何？」各國講座即席紛紛發言，又要我作了一個報告如下：

- 一、十年後快速成長的電腦會又小、又快、又好、又便宜，人人買得起。
- 二、十年後電腦將進入辦公室、進入社會、進入家庭，連兒童也要學電腦——電腦文庫將成必備讀物。
- 三、廿年後電腦將從無思考力變成有思考力。
- 四、廿年後電腦將使不懂電腦的人變成功能性文盲。

今天看起來，第一及第二個預測都已應驗了。而日本、英國、美國更自前年起，紛紛從事第五代電腦的設計及製造。我國有識之士，已於今年起一再研究第五代電腦之問題及發展。而即將來臨之新電腦將是一個具有智慧及思考力的機器。他可以讀書，可以與人類以語言交談；如果配在相關機器上，便近乎一位又聰明又能幹的人類。這種電腦系統預計於十年左右完成，一定更會形成新時代的科技及人文大革命。

第三項預測將在各國的大膽革新下實現，而形成對吾人生活及文化之重大衝擊。就第四項而言在那個時候認得「電」字，也認得「腦」字，而不知「電腦」二字加起來是什麼意思的人，便不再是被「新書香社會」尊敬的人士。

我國的知識水準一向不錯，一般說來大專程度以上者約有百分之五，中學程度百分之四十，初中以下百分之五十五。而臺北市之大專程度更高達百分之十四，「文盲」實在很少。在這樣漂亮的統計數字下，我們的「電腦文盲」是多是少呢？

做電腦文盲並不可怕，只要你有「三念」原則就不怕了——那就是要有「念」頭來學習電腦；學後必須要能改變舊觀「念」成為新觀「念」。

這個時代更是「電腦兒童」時代，他們生在電腦時代，所以愛電腦，不怕電腦，電腦可成為他們生活的一部份，他們與電腦在一起覺得很自然；正如許多生在農村社會的人，愛毆子、愛竹馬一樣的自然。這個時代的兒童不會成電腦文盲，也不應該成為電腦文盲；尤其是我們中國的兒童們，他們出生在電腦時代，他們將在國

際商場上為國家作一名鬥士。我們都希望子女成龍成鳳，為什麼不早讓他們學這一個最重要的工具及文化呢？我們都知道練武功要從小開始，學芭蕾舞，學鋼琴要從小開始，而且越早越好；為什麼學電腦不能從小開始？為什麼不准他們玩他們自己新時代的電腦，偏要他們玩「舊時代」的毽子和竹馬？

如果我是一位「電腦文盲」，我會自己先從事「新識字運動」；而不把「上古史」硬拿來束縛住「電腦兒童」及「電腦文盲」們的手腦。須知新時代已迅速而堅決地來了——現在是「鯢鳥」飛上枝頭變「鳳凰」的最後機會。

我國電腦資訊的急速發展有目共睹；在發展及成長過程中，陣痛是免不了的，但如不能懷有「臨事而懼，好謀以成」的心理，則美國奧斯邦電腦公司、德州儀器公司及阿他雷公司、富蘭克林公司等，在電腦發展上的失敗，便是殷鑒不遠。所以我們有必要提出檢討，提出改進方法，因為自「1984」年起不過十年左右，「有思考力」的電腦便將誕生了。

要有效促進我國電腦成長，吾人必須積極從事十個新方向：

- 一、輸出要重點突破，不可兼容並包——吾人有較廉價之技術人才、聰明苦幹之知識份子，但限於國力資源及學識，還是抓住幾個重點發展為佳。
- 二、要注意「顧客為主」原則在開拓市場上之意義及價值，不可把生金蛋的鵝趕走。
- 三、造成容許發展之電腦環境及市場，不可朝令夕改；不可因噎廢食；不可過份干涉；應多獎勵學習。

四、電腦成長要以「行銷導向」不可以「生產導向」。

五、全國修訂不合時宜之法令解釋，行政管轄權及書刊，並引進新知識，以配合新時代之新需要。

六、由政府及民間合作成立全國性公正而客觀之電腦資訊委員會，以求統一意見，教育及導引各界，事先準備，迎接新時代。

七、用新人行新政——須知在電腦時代，善意的無知為害之烈勝於惡人——因惡人易為人知而加以防範。

八、要學習以新管理方法來管理電腦資訊之成長——要學習如何來管理電腦資訊之成長，要重視電腦成長戰略，而不可用「農業波」或「工業波」時代之舊觀念，來管理「電腦資訊波」時代之新成長。

九、要把握市場、原料及知識來源——不可俯仰由人，靠天吃飯，要研讀先機、未雨綢繆。

十、發動全民力量加入發展電腦之通盤策劃及推廣——須知以全國之力，公私合作，仍不見得能容易應付的挑戰，怎可以有限的人力挑上太重的擔子？！

所以，五南圖書出版公司發行人楊榮川先生開拓「新時代電腦文庫」的魄力與努力，是配合全民發展電腦資訊運動中，堅定而有力的一步。新時代電腦文庫將敦請最好的人才來著述及翻譯最新的學問及出版物。凡是與電腦有關，且有重要性或實用性的新知，均在網羅之列，希望「新時代電腦文庫」，將成為中國電腦發展史上，又一個新的里程碑。而個人才疏學淺，得以參與此一新時代新工作；其惶恐，其愉快，又豈這一篇序文所能表達。

發 行 者 言

范光陵博士被稱為中國電腦之父。他首先在國內揭開了電腦啓蒙運動；他舉辦了中國第一屆人造智慧會議；寫了整個中國第一本電腦書「電腦和你」——是海內外千千萬萬中國人看過的第一本電腦書，他創造了「中文電腦化」，「電腦中文化」的新觀念；舉辦了中國第一屆中文電腦會議；他和有志之士共同創辦了中國第一個全國性電腦團體，也擔任過十次國際電腦資訊會議主席；中國第一任電腦研究所所長，第一任電腦科主任及第一任電子計算機系主任，又主持過中國第一次電視電腦節目。在中國電腦史上他創造了許多第一，也使得新時代的其他新人物，更進一步創造了許多第一。

新時代電腦文庫能由范光陵博士擔任總主編，實在是一件很榮幸的事，相信在他的策劃主編之下，配合碩士級以上的電腦編譯人才，必定能夠達到「不是好書不出版，出版的都是好書」的嚴格要求，共同為中國電腦化，盡一份心力。

楊 榮 川

目次

0

緒論..... 1

0-1 中華民國臺灣中文電腦發展簡介 3

0-2 中文字辨認系統簡介 11

1

中文處理系統..... 17

1-1 中文電腦系統開發的研究和實踐 19

1-2 中文資料的處理基本筆劃方案的科學原理及中文電腦打字機的鍵盤設計 36

1-3 富士通中文資料處理系統(CEF) 50

1-4 中文電腦處理系統 CWP 的設計 63

1-5 中文圖形資料的冗餘度分析—中文發生器設計理論之一 71

1-6 對電腦系統擴充中文處理功能和中文軟體工具的研究 79

1-7 高檔小型電腦中文處理系統的研究 92

1-8 在微電腦上實現中文操作系統和資料庫系統 110

1-9 多用戶中文資料處理的人機對話 122

1-10 中文資料與資料字典 131

1-11 中文檔案管理系統的軟體總體設計 143

1-12 中文報表程式的設計—一種簡易自然語言描述報表過程的方案 155

1-13 中文資料五因次模型和分詞、祈句、辨義的算法研究 165

1-14 中文COBOL 語言的設計 174

1-15 嵌套結構中文字庫訊息壓縮 187

2 中文電腦輸入輸出技術的摘要 197

- 2-1 拼寫電腦系統 199
- 2-2 關於聯想式半合成中文輸入鍵盤的說明 211
- 2-3 從同時做不同動作時發生打擾的角度來考察打字作業 216
- 2-4 微電腦中文輸出和中文報表語言 227
- 2-5 廉價的磁觸式中文整字輸入鍵盤 235
- 2-6 手寫中文輸入識別系統及在 Z80 微電腦上的實現 244
- 2-7 RPH-2 型多功能壓感式中西文輸入中文大鍵盤 254
- 2-8 梳齒狀並行針打機 262

3 中文字識別之摘要 271

- 3-1 TRS-80 (I) B 型文字處理系統 273
- 3-2 電腦系統擴充中文處理功能的通用途徑 282
- 3-3 中文句子的多標記、多叉樹形圖分析法 291
- 3-4 中頓句與無頓句的頻度分析 313
- 3-5 微電腦採用中文時實現中西文共通的一種方案 317
- 3-6 電腦進行中文字詞頻率統計的問題和方法 322
- 3-7 印刷體中文識別的粗分類方法 330
- 3-8 高分辨率中文字形在電腦中的壓縮及字形點陣的復原設備 349

4 中文語音識別 369

- 4-1 拼音中文一音調制自然語言 371
- 4-2 指揮自動化與通信拼音字 388

5 中文輸入輸出編碼 395

- 5-1 新式“字形”中文編碼法 397
- 5-2 中文字形編碼的原理和實踐：WBZX 中文編碼方案中文基本構件

- 實用頻率表 418
- 5-3 ZN9 中文 26 鍵輸入碼方案 439
- 5-4 關於中文輸入編碼評測規則的建議 447
- 5-5 海曼公式的一般分析 453
- 5-6 中文編碼的現狀與發展 463
- 5-7 一種實用的編碼方案：雙音編碼 484
- 5-8 論中文輸入編碼的民族性與國際性 486
- 5-9 系統碼 492
- 5-10 中文字根編碼的探討 506
- 5-11 海曼公式與中文編碼時間 512
- 5-12 中文編碼多道輸入系統 524

6

中文電腦的理論與應用..... 535

- 6-1 人工智慧革命—介紹一個建立在高智慧字母系統上的拼音字：電腦的自然語言 537
- 6-2 中文語言的理解系統 557
- 6-3 結合上下文輔助分詞的教學系統 570
- 6-4 電腦輔助科技英語詞匯的統計結果的初步分析 579
- 6-5 語言智能與智能與電腦輔助語言智能兼談電腦輔助中文教學系統 592
- 6-6 中文機械翻譯：FY-IV 型機器翻譯系統設計綱要 625
- 6-7 自然語言的機器翻譯 641
- 6-8 關於應用型機譯系統的幾點考慮 648
- 6-9 機器理解中文的一些探索和設想 655
- 6-10 中文資料處理大有可為—兼論語言工程產業的開發 661
- 6-11 漢語的遞迴正則模糊語言模型 674



緒 論

0—1 中華民國台灣中文電腦發展簡介

繼各種中文電腦輸入法之後，電腦博士范光陵又發表以「0」、「1」、「2」、「3」代表中文各種筆法的「中文電腦簡筆輸入法」。

這種輸入法的方式是：

- 中文之「、」「、」「／」不明確筆法一律定為「0」。
- 中文之直形，不論長短，一律定長「1」。
- 凡橫形，不論長短，一律定為「2」。
- 以上三種形狀之任何連筆，均定為「3」。

范光陵博士指出，這種輸入法是現有輸入法中最簡便的一種，可使兒童及家庭婦女她能自行使用電腦，並可使中文電腦除中華民國外，聯合國、新加坡、香港、韓國及日本所採用的漢字都能通用自如。

一位電腦應用專家指出，范光陵博士發明的輸入法僅需使用四個鍵，每分鐘可輸入五十個字，而倉頡輸入法每分鐘雖可輸入高達八十個字，但輸入按鍵為二十四鍵，顯然前者輸入法較為簡單，比較易學。

資訊工業策進會的調查指出，今年中文電腦國內營業額可達新台幣二十一億七千萬元，和去年十四億元比較，成長率達百分五十五。

調查顯示，今年中文電腦二十一億七千萬元的市場中，硬體佔二十一億一千萬元，較去年成長百分之五十三點三六，軟體佔一億六千萬元，較去年成長百分之七十八點七五。由成長趨勢看來，軟體重要性已被市場逐漸肯定。

這項統計指出，中文電腦外銷金額僅佔百分之一，而且大多數是提供國外中文教學使用。資訊會建議，拓展世界相類似文字市場或圖形文字市場是中文電腦進一步突破的可行途徑，否則廠商將面臨市場狹窄的困局。

中華民國的宏碁電腦公司（Multitech）創造了一種偏旁綜合法，它在標準字母鍵盤上的每一個鍵代表了一個特定偏旁。Multi-tech 的鍵盤輸入系統對於每一個字只需要三到四個鍵，而且超過每分鐘 60 個字。

中華民國目前中文輸入法有三、四十種。而各種輸入方案，互相批評攻擊。如戰國時期互相撻伐，用戶選擇時也無所適從，任關係推銷，用戶倒霉，以台灣原已做小的現在人力、物力、財力在互相競爭下重複研究的浪費，實在令人可惜。因此政府與民間結合，正採用公正的評估，測驗方式，來選擇最好的中文輸入法，發佈評估報告，供廠方、用戶選擇，可見現狀若組織聯合開發，綜合研究是有困難的。但大家還是要群策群力，分工合作，以期對中文電腦化，電腦中文化有所貢獻。

(一)突破電腦中文化的瓶頸

目前中文資訊似乎存有許多的問題，使得電腦普遍中文化的目標，未有突破性的發展，深入去研究中文資訊的發展過程，有三大瓶頸需要國人合力去突破，這三大瓶頸如歸結如下：

1. 中文輸入的問題：

亦即如何讓電腦接受中文資料，並讓電腦用中文顯示出其處理結果。

2. 電腦中文化的問題：

其中還可分為兩個層次，一為個別電腦中文化的問題，亦即如何讓世界所有具標準規格電腦都能處理中文資訊。

3. 電腦中文化的推廣問題：

亦即如何讓電腦中文化的觀念與意義，讓大家接受，使中文資訊處理為大家所採用，以提高作業效率。這是電腦中文化的最終目標，是吾人所應全力追求的。

(二)輸入編碼方案

台灣目前輸入編碼方案有三、四十種，正採用通過評估訂標準優選予以解決。各方案互相批評攻擊，如戰國時期互相打伐，用戶選擇無所適從，憑關係推銷。民國七十一年四月成立了資訊工業策進會，八三年二月六日公佈了占台灣產品百分之三十的四家公司的輸入方案，評估結果是：

CDC（康大）公司 中文首次尾三碼檢字法 56.7 字／分鐘

宏碁電腦公司 倉頡字母輸入法 43.7 字／分鐘

王安電腦公司 三角號碼 42.2 字／分鐘

IBM 公司 大字鍵盤法 14.9 字／分鐘

台灣傳技公司推出的旅美學人王一中博士設計的“中華一號”用小鍵盤字根輸入中文字、日文及南朝鮮文，“中華一號”具有通信功能，是在 1980 年為國防部研究的。

台灣製定了“中文資訊交換碼”，十三個單位整理出電腦用字 16,132 字，圖書館用字 31,000 餘字。在文教系統公佈了常用國家標準字體表為 4,808 個，乙表為次常用字 7,895 個，丙表為異體字 2,845 個。

我們的方針是“中文電腦，立足應用”。努力找尋中文電腦的瓶頸突破，找尋理想的輸入編碼方案。把易學、易用、好記、輸入速率高、設備成本低作為奮鬥目標。

(三)發展內碼系統：通用漢字交換碼

資訊會計劃中心主任黃台陽形容說：如果要害一個人叫他去做訂標準的工作。

大部分業者則說：「一動不如一靜，那麼多標準怎麼跟？還是觀望吧！」

學界為了「全漢字」與通用漢字各執一詞，都認為自己的東西才有資格作為國家標準。目前除了政府明令公佈試用兩年的「全漢字標準交換碼」及「通用漢字標準交換碼」之外，電報明碼也有廠家正在用，因而形成三大主流三個碼的情況。

如果仔細的去了解這種混亂狀況，可以發現主要原因在於沒有一套交換碼是十全十美，各有其優點，也各有不少缺點。

套一句台北市電腦公會理事長施振榮的話：「等不到一百分的；……七十分的也好啊！」。其實目前的態勢看，「通用漢字標準交換碼」已隱然而成為國家標準的趨勢，資策會聯合十三家軟、硬體廠商，依據「通用漢字標準交換碼」來發展 BIG 一五內碼，應用在國產十六位元個人電腦上，來發展五大系統。

(四)電腦輸入方式的突破性

由於電腦輸入方式已有革命性的突破，使資訊時代目前「鍵盤文化」的型態