

電子計算機化之中文編排系統

Computerized System for Chinese Text Editing and Typesetting

蔡中川 Jong-Chuang Tsay

Department of Computer Science, N. C. T. U.

(Received July 22, 1976)

ABSTRACT — This paper describes an automatic-processing system for Chinese text. The description of the system is divided into two parts: text-editing subsystem and typesetting subsystem. The system is implemented on PDP 11/40 minicomputer with a 32 K words memory. The principal result of the system is illustrated by utilizing this system to help the user to type, edit, typeset, and print this paper.

摘要——此篇文章敘述一套可應用於實際中文編排的計算機系統。此系統主要包含有中文編輯系統及中文排印系統。本文將討論此兩系統，並談到如何藉以從事本文之編排。

一、引言

一篇中文文章由作者寫成草稿到付印總要經過很長的時間才能完成。文章由作者寫成後即送到出版商處，出版商接到作者的文章草稿即抄寫然後排版及付印。此付印出來的資料當然非為最後所需者。必須由出版商及作者校正兩次，然後才印成書本、雜誌或報紙。由此傳統式的處理過程可以看出來一本書的出版需要相當長的時間；而且出版商不希望作者在校正之時修改太多，作者就是想要將自己的文章安排或修改得更完美也只好放棄；又有時由於資料之過時而須加以更新，其更新工作不易；再者此出版過程甚為費力而且容易產生誤差。為了避免上述缺點而有自動編排系統之產生。此自動編排系統利用電子計算機以幫助作者及出版商解決出書過程之兩大問題——編輯及排印。編輯系統之主要工作為幫助作者修改或更新文章，而排印系統則將編好之後的文章以適當格式印出。

欲設計一套計算機化之中文編排系統不是一件單純的問題，其所牽涉的問題甚多。其一為必須建立約有八千個字的中文檔案於計算機內，其二為必須解決中文字輸出及輸入之問題，其三為設計印製中文表格之系統（即排印系統），其四為編輯系統之設計。這些問題的研究可參考第1篇至第8篇文章。

本文包括五章。第二章將以寫作者的觀點來看此系統。第三章敘述編輯系統，主要為說明各種編輯命令（EDITING COMMAND）。第四章則提到排印系統及本文之排印。第五章為結語。

二、系統之綜觀

此計算機化之編排系統是設計在PDP-11/40計算機上。以寫作者的觀點來看此系統的話，其含有一中文輸入鍵盤，一 TEKTRONIX 4010 顯示器（CRT DISPLAY），一 VERSATEC 1600A 靜電印刷機，及兩個 RK05 磁碟讀寫機。使用此系統時，作者可以利用中文輸入鍵盤按入中文字。每一中文字之輸入仍由輸入鍵盤按入相對應的字根串及字間以完成之。此時作者可以看到其相對應的中文字顯示在 4010 顯示器上。簡單的修改，如去掉最後一個字或字根，去掉最後一行，可以在鍵盤按入適當的鍵以完成之。如此作者可以在鍵盤上打入一篇中文文章並藉編輯系統之助建立相對應之中文檔案。此建立出來的中文檔以固定式二進制

檔(FORMATTED BINARY FILE)之形式存於磁碟內。其含有不等長之記錄(RECORDS)，每一記錄代表一行，一行最長為64字。此檔案僅存入文章的初稿而已。此初稿可以由靜電印刷機印出，然後作者可以在此印出來的紙上從事修改的工作。此事做完後，作者可以開始利用各種編輯命令從事修改已建立好的中文檔。在編輯時，作者可由中文鍵盤打入適當的編輯命令從事現行、存行、去字、加字、尋字、等工作，一切動作之結果皆可在顯示器上看出。編輯命令之設計原則上務期能做到使作者想修改某一處即可很方便的借助於編輯命令完成之。檔案修改後即可利用排印系統以宋體字印出(在此步驟之前，顯示在顯示器上或由靜電印刷機印出的初稿為一種較簡單的組成字字體[8]，由此可增加此系統編輯速度)。排印之格式可由存於中文檔內之控制碼或由排印系統之格式檔來控制。此將於第四章內敘述。

三、編輯系統

中文檔案建立後，作者可由中文鍵盤按入適當的編輯命令促使計算機內的編輯系統去執行之。編輯時此系統存有一指標(POINTER)，此指標隨時皆指到中文檔的某一個字或是指到中文檔的最後一個字的後面。編輯開始時指標指到中文檔的第一個字。作者可利用編輯命令將此指標前後移動，唯往前移動時，最遠僅能移到同行(現在指標所指的那一行)的第一個字。假設\$字代表正整數，而#字代表正或負整數，則各編輯命令之形式及其作用為

1. \$A: 將指標往後移動到\$行之第一個字。
2. #J: 將指標移動#字(#如為負數則往前移動)。#如大於指標到此行最後一個字之長度時則自動移到下一行。
3. #D: 由指標開始去掉#字(#如為負數則代表往前去字)。#如大於指標到此行最後一字之長度時則自動去下一行的字。
4. @D: 此命令也是去字。其去掉由一個指標到另外一個指標間的字。此另外一指標稱為標誌(MARK POINTER)其可以由#M命令來移動他。所去的字串必在同一行。
5. \$K: 去掉\$行字。
6. \$L: 由指標處往後顯示\$行。\$如大於1則指標自動移到顯示的最後行的第一個字。
7. #V: 不管#之大小將指標所指的那一行顯示出(指標可能指到此行後的空字)。
8. #|/.../; 如#不為0則不管#之大小將在兩界線符號/間的字加到指標之後面(界線符號不一定用/可以用任何其他字)。如#為0則兩界線間的字必須為一正整數，此數如小於或等於2047則將其加16400八進位數；此數如大於2047且小於或等於6143則將其減去2048再加上15000八進位數；所得的結果即代表某一中文字之內用碼(INTERNAL CODE)。此內用碼可以用之代表編輯系統所沒有而宋體字檔內已有的字。此種內用碼亦可利用中文鍵盤上之第三階(LEVEL)來建立。按此種字時先按第三階然後按一整數再按一字間即可。此字在顯示器上皆以大口字(鍵 033)表示。
9. #I 跟著按CR鍵：此時可加入一行一行的中文資料到指標的後面，直到J字按下為止。#無作用。
10. #M: 將指標之值存到標誌內。#如為0的話，其將指標所指的字其相對應的內用碼以16二進位碼輸出在電動打字機上。不等於0之#值則無作用。
11. \$S: 將\$行資料暫存起來，並且將指標移到第\$+1行之第一個字。參考U命令。
12. #U: 此命令利用\$S命令所存的資料整個加到指標的後面。#無作用。
13. #P: 如#為正值則將#行資料由靜電印刷機印出。#如為1則指標不變。#如為負值的話則由現在指標所指的那一行印到檔案末。資料由印刷機印出時自動加上0到9的行數(SEQUENCE NUMBER)。
14. \$H: 此命令執行時，作者可將顯示在顯示器上之垂直細線(X CROSS HAIR) 移動到欲將指標移到的那個位置上然後在4010之鍵盤上按任意一個鍵，則指標自動移到那個字上。\$表示Y方向之位置如1H, 2H命令。執行此命令前通常先執行V命令。
15. #T: 編輯命令雖然分開討論，實際可成串執行如AV3J1/12/ 即代表同時執行四個編輯命令 A, V, J, I。T命令之作用可如下例所示，如編輯命令成串為 XXXX3 TYY5T 則其代表執行編輯命令串XXXX四次而且執行YY六次。#為負數的話相等於正無窮大。
16. \$G/.../: 此為尋找中文檔內某字串之編輯命令。此欲尋找之字串為在兩界線符號/間之字串。此命令由指標開始尋找第\$個所欲尋找之字串然後將指標移到此字串後，如某中文

檔內，其由指標後沒有\$個此種字串，則編輯系統會顯示出 SEARCH FAILURE?，同時將指標移到檔案之末端。若所欲尋找的字串中有大口字時，則表示其與任何字皆可匹配(MATCH)。

17. #R：執行此命令時，作者須將另外一中文檔案名打入，則編輯系統會將此檔案加到正在修改之檔案中。新的檔案是加到指標後面。#無作用。

18. #Q：此命令表示剛才所修改的皆無效，重新由頭開始修改。#無作用。

19. #E：此命令表示檔案修改完成。可以重新再修改。#無作用。

20. #去：此命令去掉正在修改的整個檔案。#無作用。

執行編輯時除了須了解上述之編輯命令外，有幾點須加以說明的。在上述之編輯命令中如\$或#不打入時則代表其值為1。指標之往前移最多祇能往前移到指標所指的那一行的第一個字，如欲修改此行前面的資料時，唯有先執行Q或E命令。OA, OK, OL, OS 之作用皆為將指標移到正在修改那一行的第一個字。S命令所儲存之資料一直保持到S或P命令執行時為止，如此可以將檔案之數行資料暫存起來然後將他搬到另外一檔案內。

今試舉一例表明此編輯系統所能做的事。假設欲找出某檔案出現「電子」之處並將其全部改為「ELEC」則可由下述命令申執行之

G/電子/V-2DI/ELEC/-T

四、排印系統

中文檔建立及修改完成後即可開始排印。此編排系統之印刷機主要是利用一種印製中文報表的語言以控制之。欲了解此種語言之讀者可參考[7]。控制排印系統之命令仍由各種基本控制碼(CONTROL CODE或BASIC FORMAT)所組成，此種控制碼主要存於另外格式檔內，不過中文檔內也可以夾雜著控制碼，前者很容易利用PDP-11 計算機之編輯程式建立之，而後者則在建立中文檔時可由中文鍵盤按入。基本控制碼之使用方法及格式檔內存控制碼之形式可由參考[7]查得，此文不再贅述。至於中文檔內其控制碼之形式則如下所述(控制碼仍由中文鍵盤按入，每一鍵對應於三個十六進位數)*：1. 0CB鍵表示空字控制碼。2. 0D4鍵表示往後移一行。3. 0C5鍵表示往後移一頁。4. 06A鍵表示跳開現今格式檔內所發之A命令繼續由格式檔內取下一個控制碼以執行其命。5. 026鍵表示下兩個中文字仍等於一串字，如第一字之宋體碼為1，第二字宋體碼為J，則此串字代表由|到J之所有相對應之中文字。6. 1C2鍵表示下一個字為重複數N(此數可在中文鍵盤第三階時按入。下述之數字假設皆為此種情況按下的。欲得到負值則按入大於3000之數目字，如3024代表-24)，再下一個字為中文字，此相當於有N個相同的中文字。7. 103鍵表示其下一個字為一數字代表中文字欲放大之倍數。8. 027, 028, 029鍵表示此控制碼用到以下一個，二個，三個字。第一個字為一英文碼(設為P)，而第二及第三個則為數字(設為3及5)則此控制命令相當於執行由格式檔內所發出之控制命令(此例為3P5)及執行06A鍵之命令。9. 0AD鍵為一特殊標誌字。在排印開始前使用者可以將控制參數16改為不等於0之數目(設為N)，則表示此系統將由中文檔內的第N個標誌後開始排印。10. 07A鍵表示自動跳開格式檔內之控制命令申，直到下一個平衡的(BALANCED)右括號為止，然後跳開此右括號繼續執行格式檔所發出的下一個控制命令。

排印中文時除了利用上述之一些控制碼外，尚須利用到自動跳頁之控制參數(參數17)。此參數可在排印前改變其值。此值表示欲排印之文章其每一頁之長度。如此當排印系統排完一頁時則自動跳到下一頁排印。

排印時首先建立格式檔，然後在中文檔內之適當位置插入各種控制碼，再來則決定每頁之大小長度。如此即可開始排印。

五、結語

此篇文章敘述電子計算機在中文編排上之實際應用。此電子計算機化之中文編排系統主要含有兩系統。其一為中文編輯系統，其二為中文排印系統。欲了解此編輯系統可由本文得知，而排印系統除

*如1C2鍵表示第一階及在中文鍵盤上坐標為C2之鍵。

了看此篇文章外尚須了解印製中文報表語言(參考[7])。爲了了解此計算機化中文編排系統之實際應用情形，本文之編輯及排印完全用此系統爲之。

圖閣——本系統所使用之計算機爲PDP 11/40，此機仍屬台灣電腦公司。該公司將其借與交大，使此研究得以進行，謹此誌謝。感謝清大所點的宋體字使得本文得以印出。

此系統內之子程式有些是利用到交大數位教員及研究生們所發展完成之程式：杜敏文教授主持下所設計之組程式，謝清俊教授主持下所設計之中文鍵盤輸入控制程式，陳永建同學所設計之中文字顯示程式，中文鍵盤之輸入處理程式，靜電印刷機組成字印字程式等。蔡新民講師之移中文字到靜電印刷機緩衝器之程式。戚樹紅同學之顯示器控制程式。梁恆盛同學之靜電印刷機組成字之控制程式。謹此誌謝他們。

本校計算機中心張仲陶主任之支持及中心人員之協助也是筆者所要感謝的。

此篇文章承蒙交大計算機中心楊美嬌小姐由中文鍵盤打入，謹此致謝。

參考文獻

1. S. K. CHANG, "AN INTERACTIVE SYSTEM FOR CHINESE CHARACTER GENERATION AND TEXT EDITING," IBM RESEARCH REPORT, JULY, 1972.
2. A. VAN DAM AND D. E. RICE, "AN INTRODUCTION TO INFORMATION STRUCTURES AND PAGING CONSIDERATION FOR ON-LINE TEXT EDITING SYSTEMS," ADVANCES IN INFORMATION SYSTEM SCIENCE, VOL. 4, J. T. TOU(ED.) PLENUM PRESS, 1972.
3. J. C. TSAY, "AUTOMATIC PROCESSING SYSTEM FOR CHINESE CHARACTER GENERATION, TEXT EDITING, AND TYPESETTING," NCTU PH. D. THESIS, JULY, 1974.
4. 謝清俊「中文字根之貯存及中文字之合成」交大學刊第六卷第一期，民國六十二年。
5. M. W. DU, J. C. TSAY, AND C. C. HSIEH, "THE DESIGN OF THE CHIPS," PROCEEDINGS OF ICSS, VOL. 1, NCTU, 1975.
6. 蔡中川「中文資料處理系統中之編輯及檔案處理」國科會中文研究報告，民國六十四年。
7. J. C. TSAY, "A NEW LANGUAGE FOR TABULATION-CHITAL," THE JOURNAL OF NCTU, VOL. 1 APR., 1976.
8. 杜敏文「交大中文資料處理系統設計，檔案建立及中文字之組合」國科會中文研究報告，民國六十四年。