

第二版

資料結構

與檔案處理

林宏明 編譯



松崗電腦圖書資料有限公司

原 序

本書是專為意圖將資料資源 (Data resources) 做有效而允當管理的人士所寫的。按，資料是所有資訊系統的原動力，也是構成一個電腦系統的各部份所以存在的理由。對於處在各種經濟環境裏，大大小小的公家機關或民間企業，資料資源在各組織的決策過程中，均扮演一個不可或缺的角色。再者，資料的成本是昂貴的；必須先花功夫去搜集，使它具其有效性，並經過組織後，儲存起來，然後應加以維護並把它們聚集及整理出來，最後還要去傳播它並保護它。本書將針對有關資料群體之結構，組織及處理等各方面之問題，予以研究探討以達到使資料具有可用性，並能隨時回應使用者對資訊的需求之目的。

本書共分為兩個主要部份：第一部份（第一至十三章）討論傳統上所稱的資料結構 (Data structures)，第二部份（第十四至廿一章）則著重在檔案的管理與處理 (Management and processing of files)。最後的第廿二章是資料庫管理系統的介紹。讀者將從最基本的資料管理觀念（包括各種堆疊佇列及串列）開始，經由較複雜的資料結構（樹狀結構、B-樹及圖形）之討論，最後被導入有關檔案處理的各種方法（循序檔、相關檔、索引循序檔以及多重鍵檔案組織），並得以瞭解更複雜的資料庫管理系統之構成要素。

對於已具有電腦系統及程式設計基礎知識的讀者來說，本書特別具有重大意義。很多書提到如何建立及操作資料結構與檔案組織，這些程序通常都以 COBOL 或 PASCAL 來表示。COBOL 之所以被選用是因為它在當今資料管理環境的普遍性。而 PASCAL 則是愈來愈流行的區段結構化 (Block-structured) 語言的代表，它具有很好的資料型態的特性 (Nice data-typing features) 更可以很方便地將演算式 (Algorithms) 的邏輯結構表示出來。但

是讀者並不需要先學會這兩種（或其中一種）語言才能開始研讀本書。如果您只是才剛剛學過 FORTRAN（或 BASIC 或 PL/I 或任一種組合語言或……），您只要多花一點點時間察看本書中程式的寫法，則讀完本書後，您不僅能瞭解，資料管理與檔案處理的原則，同時也不知不覺中獲得了有關 COBOL 與 PASCAL 的基礎知識。

本書採用標準 COBOL(ANSI X3.23 - 1974，ISO 1989 - 1978，又稱為 COBOL 74) 及基礎的 PASCAL 架構，附錄包括了描述 COBOL 的語法說明，讀者不必擔心是否具有數學基礎，但是對於解決問題時所應有的基本邏輯思維方式，將有助您瞭解本書。

每章結束前均有一段簡短的摘要（如果每章之前先閱讀，也可當成該章概括性的介紹），專用術語表，包括該章提及的或對了解文章內容特別重要的文字，您可以一面看著這些術語，一面先自我測驗。再精讀附於書後的字彙索引部份（Glossary），以加強學習的效果。

有些練習題只需簡單的回答，其餘的則借助程式及計畫案的方式（Programs and projects）來增強讀者對該章的了解。通常，較簡單的回憶性習題列在前面，再接著較有趣的，具“行動性”的習題。部份問題的答案列在本書字彙索引部份的後面。

本書適合做為下列 ACM 課程委員會，各種建議中所訂課程的輔助教材：

1. “對大學部學生在電子計算機科學課程計畫方面的建議”中所訂的課程。C85（Introduction to File Processing）及 CS7（Data Structures and Algorithm Analysis）中與數學較無關的部份。
2. “對研究生在電子計算機科學計畫方面的建議”中所訂的課程 CS32（Information Storage and Access）。
3. “對研究生在資訊系統專業化計畫方面的建議”中所訂的課程 C1（Information Structures）及部份 C3 的課程（File and Communications Systems）。

4. “對大學部學生在資訊系統方面程式設計的課程建議”中所訂的相關課程。

我在一所企業管理與公共行政管理學院的管理資訊系統課程規劃 (Management Information Systems program) 中採用本書做為研究所一學期，及大學部連續兩學期的授課教材。其中研究所的課程亦列入另一學院的電子計算機科學課程規劃 (Computer Science program) 的一部份。

作者寫本書的主要目的是想以簡明、易懂的表達方式來建立資料管理及檔案處理原則的架構，並整合兩者的處理原則。這目的是想保留主題中較嚴謹的數學處理本質，而不會遺漏太多細節。理論觀念與實際應用的結合一直引導著本書寫作的方向。

讀完本書後，有程式設計能力的讀者對一個特別的問題就能分析出如何取舍資料處理的需求，選擇適當的資料結構或檔案組織，建立結構，擷取所要的資料，更新及維護結構，且了解整個事情的進展。本書將使您在複雜的資訊管理領域裏，得到一穩固的基礎，而能與日新月異不斷發展的新技術一起成長。職故，當讀者繼續鑽研有關資料庫管理技巧時，將已有所準備了！

MARY E. S. LOOMIS

序

如何將存放於主記憶體內的“資料結構”與儲存在輔助記憶體中的“檔案組織”兩種觀念結合起來，以運用於各種資訊系統（包括作業系統）的設計，一直是電腦從業人員必備的知識及不斷研究的主題。本人有感於此，特選榮列1984年全美資訊名著排行榜的“Data Management and File Processing” by Mary E. S. Loomis（Englewood Cliffs, N. J. : Prentice-Hall, 1983）乙書，改以“資料結構與檔案處理”為書名，以易懂易記的形式，將之編譯為中文，以饗國內讀者。

事實上，資料管理（Data Management）不僅是由傳統的資料結構（Data Structures）所延伸出的新名詞，它更代表了近年來資訊科學發展上的深一層意義；它是檔案處理（File Processing）及資料庫管理系統（Database Management Systems）執行之基礎。原書作者特別強調了資料結構與檔案處理的整合性，使理論觀念與實務應用合而為一，是一本邏輯結構完整，可讀性極高的好書。目前採用該書作為“資料結構”及“檔案處理”課程之教科書的國內外大專院校已相當多，足見學界對於一本好書的熱烈迴響。

全書共分二十二章，以循序漸進的方式，從傳統的、基本型的資料結構，推展到較複雜的，各類型的樹狀結構，進而討論到各種型式的檔案組織之技巧及其應用，最後則貫以資料庫管理系統的基本觀念作為結論。不僅引導了讀者逐步去品味此一資訊科學的精髓部份，並為往後去研究資料庫及作業系統提供了堅實的基礎。

本書行文，力求流暢，每一章節悉以最簡潔有力的文字及圖表敘述表達。至於文中較不易為初學者理解的地方，本人特以多年教學之經驗，以淺顯易懂的文字加以闡述，使讀者更能體會原意。本人歷年來除了擔任大專院校電腦及資訊科學系教席外，一直在國內外企業界從事着電腦化的實務工作，認為本

書具有相當的實用性。本書乃適於研習過任何一種電腦語言的學生，或社會上正從事電腦工作的人士研讀。

本書之出版，除了要感謝電腦界諸位先進好友提供的寶貴意見外，也要感謝健生公司莊總經理健宗先生，莊副總經理銘國先生的鼓勵。同時也謝謝松崗公司全體同仁的協助，以及本人所授課之學生們反映的意見。謹此致最大的謝意。

林 宏 明 謹識

中華民國 74 年 6 月

再版序言

在資訊科學的領域裏，有許多新的構想不斷地被提出探討並加以實現。尤其是資料結構與檔案組織的觀念，不僅對於系統之設計，具有深遠的影響，對於資料處理之效率與可靠性，更具意義。本書之出版即基於此一目的。

本書初版發行後，承蒙工商界，各電腦公司，及各大專院校惠予採用，迄今已銷售一空。根據各方的反應，本書循序漸進的編排方式，確能引導讀者產生融會貫通的觀念，並由於此書之邏輯結構相當完整，對於準備參加各類考試的讀者而言，亦極有助益。

於今乃將初版時疏漏之處加以訂正，予以再版。期待本書能成為研究、學習資料結構與檔案處理之人士的最佳指南。

林 宏 明 謹識

中華民國 75 年 3 月

目 錄

第一章 資料結構簡介	1
1-1 資料	2
1-2 資料管理	4
1-3 資料結構	5
1-4 資料型態	6
1-5 資料結構在程式語言中的宣稱	11
1-6 整數存放在儲存體中的對映關係	14
1-7 字元在儲存體中的對映關係	17
1-8 字串在儲存體中的對映關係	19
1-9 選擇適當的對映	23
1-10 摘要	24
1-11 名詞	25
1-12 習題	26
第二章 陣 列	31
2-1 定義：一維陣列	32
2-2 定義：多維陣列	33
2-3 利用 COBOL 和 PASCAL 宣告陣列	38
2-4 一維陣列的儲存對映	41
2-5 多維陣列的儲存對映	42
2-6 三角陣列	46
2-7 稀疏陣列	48

2-8	摘要	49
2-9	名詞	50
2-10	習題	51
第三章 資料錄		55
3-1	資料錄的定義	56
3-2	資料錄在 COBOL 與 PASCAL 中的宣稱	58
3-3	儲存體之對映	63
3-4	摘要	63
3-5	名詞	64
3-6	習題	64
第四章 堆 疊		67
4-1	定義：堆疊	68
4-2	COBOL 及 PASCAL 對 STACK 之宣稱	73
4-3	堆疊例子的應用	75
4-3-1	例 1：括號的配對	75
4-3-2	例 2：遞迴	77
4-3-3	例 3：後序表記法	80
4-4	儲存體的對映	84
4-5	摘要	86
4-6	名詞	87
4-7	習題	88
第五章 佇 列		91
5-1	佇列的定義	92

5-2	以 COBOL 和 PASCAL 語言宣稱佇列	97
5-3	佇列的應用實例	109
5-4	摘要	111
5-5	名詞	112
5-6	習題	112
第六章	鏈接串列	115
6-1	鏈串列表示法	116
6-2	鏈接串列上的基本運算	119
6-3	可用空間的管理	123
6-4	在 PASCAL 中使用指標變數宣稱鏈接串列	126
6-5	在 COBOL 及 PASCAL 中，不用指標變數之鏈串列 宣稱	128
6-6	單向鏈串列的進一步處理技巧	131
6-7	環狀的鏈串列和開頭節點	134
6-8	雙向鏈串列	137
6-9	鏈串列的應用實例	143
6-9-1	多項式	143
6-9-2	一個簡單的多重鏈串列	147
6-9-3	稀疏陣列	148
6-10	摘要	150
6-11	名詞	151
6-12	習題	151
第七章	循序搜尋與 $O(N^2)$ 排序技巧	157
7-1	循序搜尋	158

7-2	內部排序方法的介紹	163
7-3	選擇排序法	164
7-4	插入排序法	167
7-5	交換排序法：氣泡式排序法	170
7-6	摘要	173
7-7	名詞	174
7-8	習題	174
第八章	一般樹與二分樹	177
8-1	定義：圖形	178
8-2	定義：一般樹	182
8-3	定義：二分樹	185
8-4	二分樹的表示法	188
8-5	一般樹的二分樹的表示法	190
8-6	二分樹追蹤	193
8-7	串穿引線之二分樹	197
8-8	二分樹的運算	204
8-9	樹的應用	207
8-10	摘要	207
8-11	名詞	208
8-12	習題	209
第九章	二分法搜尋樹	215
9-1	二分法搜尋樹的定義	216
9-2	二分法搜尋樹之運算	218
9-3	二分法搜尋樹之評估	221

9-4	高度—平衡 (AVL) 樹	226
9-5	限界—平衡樹	228
9-6	摘要	229
9-7	名詞	230
9-8	習題	230
第十章	二分法搜尋與 $\sigma(N \cdot \log_2 N)$ 排序技巧	233
10-1	二分搜尋法	234
10-2	區隔—交換排序法 (快速排序法)	237
10-3	堆積排序法	241
10-4	賽程排序法	249
10-5	摘要	257
10-6	名詞	258
10-7	習題	259
第十一章	M - 向搜尋樹與 B - 樹	261
11-1	以二分法搜尋樹作為索引	261
11-2	定義：M- 向搜尋樹	263
11-3	M-way 搜尋樹的搜尋	266
11-4	定義：B- 樹	267
11-5	B - 樹的操作	269
11-6	摘要	276
11-7	名詞	278
11-8	習題	279

第十二章	B* 樹, B⁺ 樹及 TRIES	281
12-1	B* 樹: 特殊的 B 樹	281
12-2	B ⁺ 樹: 特殊的 B 樹	286
12-3	TRIES: 特殊的一般樹	291
12-4	摘要	294
12-5	名詞	295
12-6	習題	296
第十三章	圖 形	299
13-1	鄰近區域性矩陣表示法	301
13-2	串連表示法	306
13-2-1	節點目錄表示法	307
13-2-2	多重串列表示法	310
13-3	圖形的追蹤	312
13-3-1	寬度優先追蹤	313
13-3-2	深度優先追蹤	314
13-3-3	比較次數	315
13-4	可到達性與最短路徑	316
13-5	臨界路線	319
13-6	跨距樹	321
13-7	摘要	324
13-8	名詞	325
13-9	習題	327

第十四章	檔案組織與管理的簡介	329
14-1	檔案	330
14-2	檔案組織	334
14-3	檔案的運作	335
14-4	執行績效上的考慮	339
14-5	摘要	340
14-6	名詞	340
14-7	習題	342
第十五章	檔案的儲存裝置	343
15-1	檔案儲存裝置的基本特性	343
15-2	磁帶的儲存	345
15-2-1	資料表示法與磁帶密度	346
15-2-2	磁帶中的同位及錯誤控制	347
15-2-3	磁帶的區段	348
15-2-4	磁帶的標記標籤	352
15-2-5	磁帶的用法	353
15-2-6	打孔卡片	353
15-2-7	紙帶	354
15-3	磁碟	355
15-3-1	磁碟的實體特性	356
15-3-2	磁碟上的資料表示法與定址方式	357
15-3-3	可移動磁頭式的磁碟存取	359
15-3-4	固定磁頭式磁碟的存取	361
15-3-5	每磁軌具讀寫頭式磁碟機與密閉式磁碟機	362

15-4	其他直接存取的儲存裝置	363
15-4-1	軟式磁碟儲存體	363
15-4-2	小型磁碟盒	364
15-4-3	大量儲存器	365
15-5	摘要	365
15-6	名詞	366
15-7	習題	369
第十六章	循序檔組織	371
16-1	定義與應用：循序組織檔	372
16-2	循序組織檔案之儲存	375
16-3	程式中循序檔之宣告	376
16-3-1	COBOL 中之循序檔	376
16-3-2	Pascal 中的循序檔	380
16-4	循序檔之建立	382
16-5	循序檔之存取	388
16-6	循序檔之更新	391
16-7	循序檔的設計	396
16-8	摘要	399
16-9	名詞	400
16-10	習題	401
第十七章	輸入／輸出控制之簡介	403
17-1	輸入／輸出控制之定義及需求	404
17-2	檔案目錄與控制資訊	405
17-3	裝置的控制	407

17-4	通道管理	410
17-5	緩衝器之管理	413
17-5-1	(A)單一緩衝	413
17-5-2	(B)雙緩衝	416
17-6	開檔與閉檔	420
17-7	摘要	421
17-8	名詞	422
17-9	習題	424
第十八章	檔案之排序與合併	427
18-1	檔案之合併與排序簡介	428
18-2	自然合併法	431
18-3	平衡合併法	435
18-4	多相合併法	437
18-5	串級合併法	440
18-6	引用公用程式來排序 / 合併	442
18-7	程式中的排序 / 合併	444
18-8	排序 / 合併之執行上的績效	448
18-9	摘要	449
18-10	名詞	451
18-11	習題	452
第十九章	相關性檔案組織	453
19-1	相關檔的定義和應用	454
19-2	相關性檔案的直接對映技巧	457
19-2-1	絕對定址方式	457

19-2-2	相關定址方式	458
19-3	相關性檔案的目錄查詢	460
19-4	相關性檔案的位址換算技巧	463
19-4-1	除 / 餘數查位法	466
19-4-2	平方中值查位法	468
19-4-3	摺疊查位法	469
19-4-4	各種查位函數的比較	470
19-5	解決碰撞問題的各種方式	473
19-5-1	線性探測法	474
19-5-2	二次查位法	475
19-5-3	線性探測法與二次查位法之比較	477
19-5-4	同義字鏈接法	479
19-5-5	承儲桶定址法	480
19-6	相關性檔案在程式上的使用	486
19-7	相關檔在 COBOL 中之應用	488
19-7-1	檔案之初建	490
19-7-2	資料錄之擷取	492
19-7-3	檔案之更新	494
19-8	相關檔之設計	496
19-9	摘要	498
19-10	名詞	499
19-11	習題	501
第二十章	索引循序檔案組織	503
20-1	索引循序檔的定義與應用	504
20-2	索引循序檔案組織之執行方式	506