

Eastern
Economy
Edition

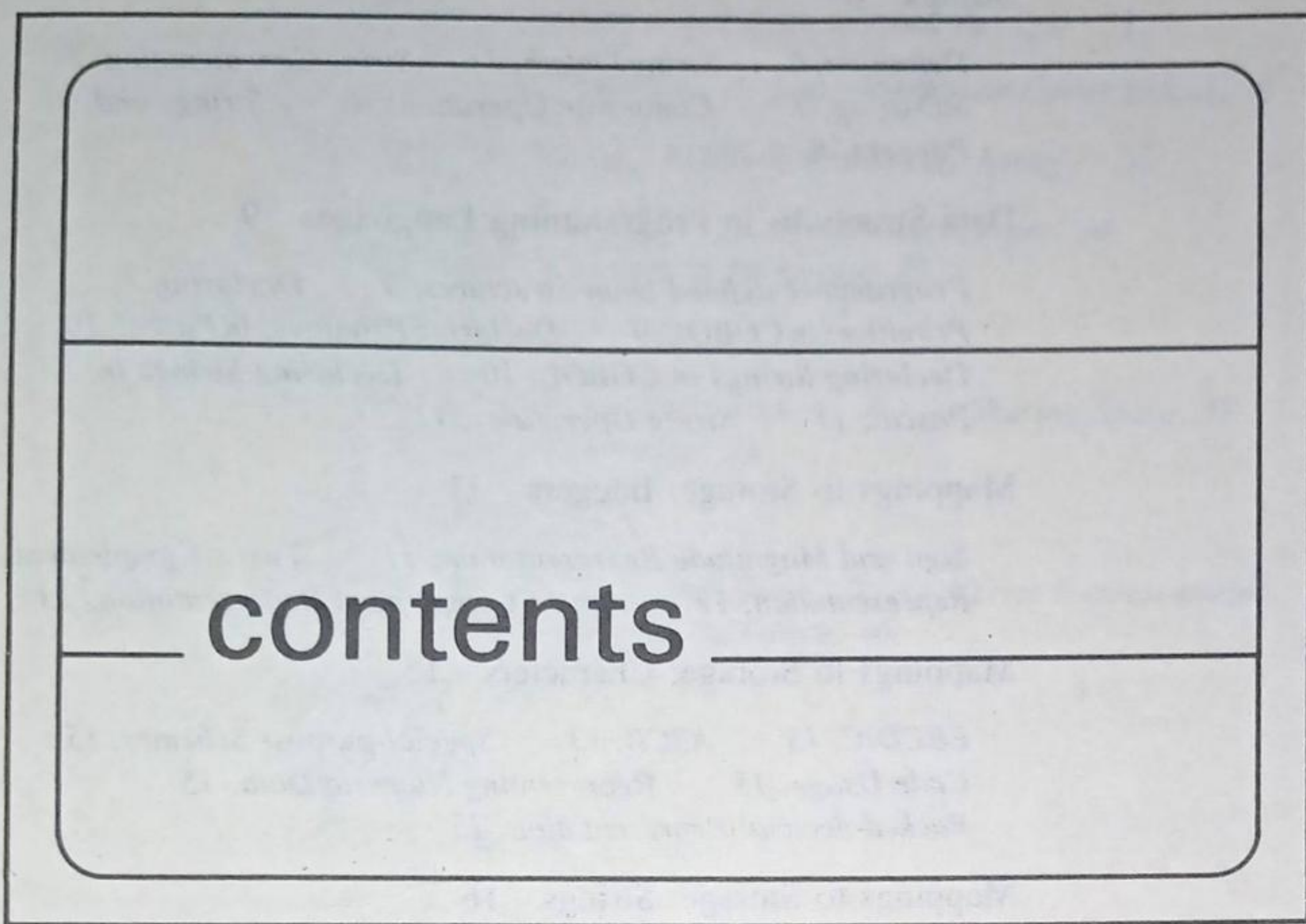
Data Management and File Structures

Second Edition



Mary E.S. Loomis

001.553
c



PREFACE	xxi
CHAPTER 1 INTRODUCTION TO DATA STRUCTURES	1
The Use of Data	1
Sources of Information, 2	
Kinds of Decisions, 2	
Data and Decisions, 2	
Data Management	3
Data Management Objectives, 3	
A Classification of Data Structures	3
Logical Data Structures, 4	
Primitive and Simple Structures, 4	
Linear and Nonlinear Structures, 4	
File Organizations, 4	
Primitives	5
Integers, 5	
Booleans, 5	
Characters, 6	

Strings 6

Definition, 6 *String Length*, 7 *String Concatenation*, 7
Substring, 7 *Composite Operations*, 8 *Strings and Integers*, 8

Data Structures in Programming Languages 9

Programmer-defined Data Structures, 9 *Declaring Primitives in COBOL*, 9 *Declaring Primitives in Pascal*, 10
Declaring Strings in COBOL, 10 *Declaring Strings in Pascal*, 11 *String Operations*, 11

Mappings to Storage: Integers 11

Sign and Magnitude Representation, 11 *Two's Complement Representation*, 12 *One's Complement Representation*, 13

Mappings to Storage: Characters 13

EBCDIC, 13 *ASCII*, 13 *Special-purpose Schemes*, 15
Code Usage, 15 *Representing Numeric Data*, 15
Packed-decimal Representation, 15

Mappings to Storage: Strings 16

Storage Alternatives, 16 *Packed-string Representation*, 17
Unpacked-string Representation, 18

Selection of Appropriate Mappings 18

Summary 19

Terms 20

Selected References 20

Review Exercises 21

CHAPTER 2 ARRAYS

23

One-Dimensional Arrays 23

Subscripts, 24 *Definition*, 24 *Examples*, 24

Multi-Dimensional Arrays 25

Subscripts, 25 *Definitions*, 25 *Examples*, 26 *Cross-section*, 26
Transpose, 27 *Extensions to More Dimensions*, 27

Arrays in COBOL and Pascal 29

One-dimensional Arrays, 29 *Two-dimensional Arrays*, 29
More Dimensions, 30 *Operations on Array Elements*, 30
Operations on Arrays, 31

Mappings to Storage: One-dimensional Arrays	31
<i>Lower Bound: One, 31</i>	
<i>Generalizing the Lower Bound, 32</i>	
Mappings to Storage: Multi-dimensional Arrays	32
<i>Row-major Order, 32</i>	
<i>Column-major Order, 34</i>	
<i>Selecting a Linearization Technique, 35</i>	
Triangular Arrays	37
<i>Definitions, 37</i>	
<i>Linearization, 37</i>	
<i>Sharing Space, 37</i>	
<i>Accommodating Change, 38</i>	
✓ Sparse Arrays	38
<i>Definition, 38</i>	
<i>Linearization, 39</i>	
<i>Vector Representation, 39</i>	
<i>Linked-list Representation, 40</i>	
Summary	40
Terms	40
Selected References	41
Review Exercises	41
CHAPTER 3 RECORDS	44
Definitions	44
<i>Forming Records, 45</i>	
<i>Examples, 45</i>	
<i>Identifying Keys, 46</i>	
<i>Files, 46</i>	
Records in COBOL and Pascal	46
<i>COBOL, 46</i>	
<i>Pascal, 48</i>	
<i>Qualified Names, 48</i>	
<i>Traffic Flow Example, 49</i>	
Mappings to Storage	51
Summary	51
Terms	51
Selected References	52
Review Exercises	52
CHAPTER 4 STACKS	53
Definitions	53
<i>Linear List, 53</i>	
<i>Stack, 54</i>	
<i>Examples, 54</i>	
<i>Stack Operations, 55</i>	
<i>Example, 56</i>	

Stacks in COBOL and Pascal 57

Housing Stacks in Arrays, 57 *Declaring Stacks, 57*
Stack Operations, 58

Example Applications of Stacks 59

Matching Parentheses, 59 *Recursion, 61* *Postfix*
Notation, 63

Mappings to Storage 68

Sharing Space, 68

Summary 70

Terms 70

Selected References 70

Review Exercises 71

CHAPTER 5 QUEUES

Definitions 73

Queue Operations, 74 *Example, 76*

Queues in COBOL and Pascal 77

Housing Queues in Arrays, 77 *Declaring Queues, 78*
Queue Operations, 78 *Moving Through Storage, 79*

Circular Queues 80

Using Circular Queues, 83 *Variations, 85* *An*
Alternative Representation, 86 *Summary, 89*

Performance of Queues 89

Performance Parameters, 90 *Observation, 90*
Simulation, 90 *Queueing Theory, 91*

Summary 91

Terms 92

Review Exercises 92

CHAPTER 6 LINKED LISTS

Linked List Representation 94

Problems with Sequential Representation, 94 *Nonsequential*
Representation, 95 *Basic Concepts, 95* *Advantages, 95*
Costs, 96

Basic Operations on a Linked List	96
<i>Notation</i> , 96	
<i>Removing a Node</i> , 98	
<i>Inserting a Node</i> , 99	
Managing Available Space	100
<i>The Storage Pool</i> , 100	
<i>Getting a New Node</i> , 101	
<i>Freeing a Node</i> , 102	
Linked Lists in Pascal, Using Pointer Variables	103
<i>Defining Linked Lists</i> , 103	
<i>Managing Space</i> , 104	
<i>Removing a Node</i> , 104	
<i>Inserting a Node</i> , 104	
Linked Lists in COBOL and Pascal, Without Using Pointer Variables	105
<i>Using an Array</i> , 105	
<i>Defining a Linked List</i> , 106	
<i>Removing a Node</i> , 107	
<i>Inserting a Node</i> , 107	
Further Manipulations of Singly Linked Lists	108
<i>Finding a Particular Node</i> , 108	
<i>Inserting at the End of a List</i> , 109	
<i>Reversing a List</i> , 111	
Circularly Linked Lists and Head Nodes	112
<i>The Josephus Problem</i> , 113	
<i>Header Nodes</i> , 114	
<i>Removing a Particular Node</i> , 115	
Doubly Linked Lists	117
<i>Basic Concepts</i> , 117	
<i>Defining a Doubly Linked List</i> , 117	
<i>Removing a Node</i> , 118	
<i>Inserting a Node</i> , 120	
Example Applications of Linked Lists	121
<i>Polynomials</i> , 122	
<i>A Simple Multi-linked List</i> , 126	
<i>Sparse Arrays</i> , 126	
Summary	128
Terms	129
Selected References	129
Review Exercises	129

CHAPTER 7 GRAPHS**133**

Definitions	133
<i>Paths</i> , 135	
<i>Cycles</i> , 135	
<i>Directed Graphs</i> , 136	
Graphs in Programs	137

The Partition-Exchange Sort (Quicksort) 221

Performance, 223

The Heapsort 224

Heap Structure, 225
the Heap, 228*Creating a Heap*, 226*Processing*

The Tournament Sort 231

Performance, 236

Summary 237

Terms 239

Selected References 239

Review Exercises 240

CHAPTER 10 FILE SYSTEMS

242

Files 242

Classifying Files by Function, 243
Files, 245*Modes of Accessing*

File Organizations 246

File Operations 247

Creating a File, 248
from a File, 248*Updating a File*, 248
Maintaining a File, 249*Retrieving*

File Systems 250

File Directories 251

Device Control 252

Channels, 252
Devices, 253*Kinds of Channels*, 253
Channel Activities, 254
Blocking Records, 255*Kinds of Processing a*

Managing Buffers 256

Single Buffering, On Demand, 256
256 *Buffering with Blocks*, 257
Three Buffers, 260*Anticipatory Buffering, Double Buffering*, 258

Opening and Closing Files 261

Database Systems 262

Summary 264

Terms 265

Selected References 266

Review Exercises 266

CHAPTER 11 SEQUENTIAL FILE ORGANIZATION 268

Definitions 268

Example, 269 Ordering Records, 270 Processing, 270
Advantages and Disadvantages, 271

Storing Sequential Files 271

Magnetic Tape 272

Data Representation and Density, 273 Parity and Error
Control, 274 Blocking, 275 Tape Marks and Labels,
277 Tape Usage, 278

Declaring Sequential Files 278

COBOL, 278 Pascal, 281

Creating a Sequential File 282

Editing Transactions, 282 Intelligent Data Entry, 284
Writing Records, 284 Report Files, 285

Retrieving from Sequential Files 287

Updating Sequential Files 289

How Often to Do It?, 289 Generation Files, 290 Kinds
of Updates, 291 Handling Errors, 292 Update Logic,
292

Performance of Sequential Files 294

Blocking Factor, 294 File Length, 294 Selecting the
Key, 296

Summary 296

Terms 297

Selected References 298

Review Exercises 298

CHAPTER 12 SORTING AND MERGING FILES 301

Introduction to File Sorting and Merging 301

Merge Logic, 302 Phases, 302 Sort/Merge
Performance, 303 The Sort, 304 The Merge, 304

Natural Merges	304		
<i>Example, 304</i>		<i>Performance, 305</i>	
Balanced Merges	306		
<i>Example, 306</i>		<i>Performance, 310</i>	
Polyphase Merges	310		
<i>Example, 310</i>			
Cascade Merges	312		
<i>Example, 312</i>			
Sorting/Merging with Utilities	313		
<i>Specifying the Process, 314</i>		<i>Example, 314</i>	<i>Another</i>
<i>Example, 315</i>		<i>Customizing the Process, 315</i>	
Sorting/Merging in COBOL Programs	316		
<i>First Example, 317</i>		<i>Second Example, 318</i>	<i>Merging</i>
<i>Sorted Files, 319</i>			
Sort/Merge Performance	319		
Summary	320		
Terms	321		
Selected References	321		
Review Exercises	322		

CHAPTER 13 RELATIVE FILE ORGANIZATION

323

Definitions	323		
<i>Processing, 324</i>		<i>Example, 325</i>	<i>Advantages and</i>
<i>Disadvantages, 326</i>			
Magnetic Disk Storage	326		
<i>Physical Characteristics of Magnetic Disks, 326</i>		<i>Data</i>	
<i>Representation and Addressing, 327</i>		<i>Movable-head Disk</i>	
<i>Access, 329</i>		<i>Fixed-head Disk Access, 330</i>	<i>Head-per-</i>
<i>track Drives and Winchester Disks, 331</i>		<i>Flexible Disk</i>	
<i>Storage, 331</i>			
Addressing Techniques	332		
Direct Mapping Techniques	332		
<i>Absolute Addressing, 332</i>		<i>Relative Addressing, 333</i>	

Directory Lookup Techniques	334
<i>Directory Structure</i> , 334	<i>Storing Records</i> , 335
<i>Advantages</i> , 336	<i>Performance</i> , 336
Address Calculation Techniques	337
<i>Division-Remainder Hashing</i> , 339	<i>Mid-square Hashing</i> , 341
<i>Hashing by Folding</i> , 342	<i>Comparison of Hash Functions</i> , 343
Approaches to the Problem of Collisions	344
<i>Linear Probing</i> , 344	<i>Double Hashing</i> , 345
<i>Comparison of Linear Probing and Double Hashing</i> , 346	<i>Synonym Chaining</i> , 348
<i>Bucket Addressing</i> , 349	
Using Relative Files	352
<i>Declaring a Relative File</i> , 353	<i>Creating a Relative File</i> , 353
<i>Retrieving from Relative Files</i> , 353	<i>Updating Relative Files</i> , 354
<i>Programmer Responsibilities</i> , 354	
Relative Files in COBOL	354
<i>Declaring a Relative File</i> , 355	<i>Creating a File</i> , 356
<i>Retrieving Records</i> , 357	<i>Updating a File</i> , 359
Performance of Relative Files	360
Summary	362
Terms	362
Selected References	363
Review Exercises	366
CHAPTER 14 INDEX STRUCTURES	368
Binary Search Trees as Indexes	368
M-way Search Trees	370
<i>Example</i> , 371	<i>M-way Search Trees as Indexes</i> , 371
<i>Searching an M-way Search Tree</i> , 372	<i>Performance</i> , 373
✓ B-trees	373
<i>Example</i> , 374	<i>Searching B-tree</i> , 374
Inserting into a B-tree	376
<i>Example</i> , 377	<i>Performance</i> , 379
Deleting from a B-tree	381
<i>Example</i> , 381	

B*-Trees 383

Inserting into a B-tree, 384*

Tries 388

Example, 389

Searching a Trie, 389

Maintaining a

Trie, 391

Trie Variations, 391

Summary 393

Terms 394

Selected References 394

Review Exercises 396

CHAPTER 15 INDEXED SEQUENTIAL FILE ORGANIZATION

398

Definitions 398

Examples, 399

Applications 399

B+-tree Structures 400

Manipulating a B+-tree, 402

Searching a B+-tree, 402

Example 403

Direct Access, 403

Sequential Access, 403

Inserting

Records, 403

Physical-Layout Indexes 406

Accessing Data Records, 407

Inserting Records, 408

Wrinkles, 410

Deleting Records, 411

Indexed Sequential Files in COBOL 411

Declaring a File, 411

Creating a File, 413

Creating a

File Using a Utility, 413

Retrieving Records, 414

Updating a File, 415

Design of Indexed Sequential Files 416

Summary 417

Terms 418

Selected References 418

Review Exercises 419

CHAPTER 16 MULTI-KEY FILE ORGANIZATION	422
Multi-Key Access 422	
<i>Example: The Need for Multiple Access Paths, 423</i>	<i>Support</i>
<i>by Replicating Data, 424</i>	<i>Difficulties Caused by Replication,</i>
<i>424</i>	<i>Support by Adding Indexes, 424</i>
Inverted File Organization 425	
<i>Basic Concepts, 425</i>	<i>Example, 425</i>
<i>More Definitions, 426</i>	<i>Variations, 425</i>
<i>Addresses, 428</i>	<i>Indexing with Indirect</i>
<i>Queries, 430</i>	<i>Non-unique Key Values, 429</i>
	<i>Existence</i>
Multi-list File Organization 430	
<i>Basic Concepts, 430</i>	<i>Example, 430</i>
<i>Variations, 434</i>	<i>Processing, 433</i>
Alternate-key Indexed Sequential Files 434	
<i>Example, 434</i>	<i>Declaring a File in COBOL, 435</i>
<i>Creating a File, 435</i>	<i>Retrieving Records Sequentially, 436</i>
<i>Retrieving Records Directly, 437</i>	<i>Updating Records, 438</i>
Comparisons and Trade-offs 438	
File Design Summary 439	
Closing Words 441	
Terms 441	
Selected References 442	
Review Exercises 443	
GLOSSARY	444
ANSWERS TO SELECTED REVIEW EXERCISES	460
APPENDIX: META-LANGUAGE FOR COBOL	475
INDEX	477