Contents

	Preface	*v*
Chapter 1	**Automatic Information Systems**	1
	1-1 Introduction	1
	1-2 Information Dissemination	4
	1-3 Information Search and Retrieval	7
	1-4 Automatic Content Analysis	9
	References	20
Chapter 2	**Information Analysis and Dictionary Construction**	21
	2-1 Introduction	21
	2-2 Language Analysis	22
	2-3 Dictionary Construction	25
	The Synonym Dictionary (Thesaurus)	25
	The Word Stem Thesaurus and Suffix List	30
	The Phrase Dictionaries	33
	The Concept Hierarchy	38

2-4 Dictionary Performance 40
The Stem Thesaurus 41
The Regular Thesaurus 44
The Phrase Dictionary 47
2-5 Automatic Thesaurus Construction 49
Fully Automatic Methods 50
Semiautomatic Methods 51
Sample Thesaurus Generation 54
2-6 Automatic Hierarchy Formation 57
References 64

Chapter 3 **Dictionary Operations** 66
3-1 Introduction 66
3-2 Structure Representation and Information Search 67
3-3 Search Algorithms 70
Sequential Scan 70
Controlled Scanning Methods 71
Chaining Methods 74
3-4 Thesaurus Operations 79
Thesaurus Setup 79
Thesaurus Look-up System 85
3-5 Statistical Phrase Processing 93
3-6 Processing of the Concept Hierarchy 96
Internal List Storage 97
Hierarchy Updating and Setup Operations 99
External Tape Storage 101
Hierarchy Look-up and Expansion Operations 105
References 108

Chapter 4 **The Statistical Operations** 110
4-1 Introduction 110
4-2 Statistical Term Associations 113
Term-Document Mapping 113
Statistical Associations 116
Linear Associative Retrieval 119
4-3 Implementation of Associative Retrieval 122
4-4 Evaluation of Concept Associations 129
4-5 Automatic Classification 133
Eigenvalue and Factor Analysis 135
Overlapping Clustering 139
Term-Sentence Clustering 144
References 148

Chapter 5 The Syntactic Operations 151
5-1 Introduction 151
5-2 Automatic Content Analysis 152
5-3 Syntax and Semantics 156
Phrase Structure Grammars 156
Automatic Phrase Structure Recognition 158
Transformational Grammars 162
Semantics 164
5-4 Criterion Phrases 166
Specification 166
Phrase Formats 171
Criterion Phrase Processing 176
5-5 Syntactic Tree Matching 179
General Topological Structure Matching 179
Graph Matching under Restricted Conditions 184
Node-by-node Tree Matching 191
5-6 Evaluation 196
References 199

Chapter 6 Retrieval Models 202
6-1 Introduction 202
6-2 Elementary Set Theory 203
6-3 Inclusive Information Retrieval 210
6-4 Systems Based on Classification 215
6-5 The Use of Negation 223
6-6 Tree and Graph Models of Retrieval 228
References 233

Chapter 7 The Retrieval Process 235
7-1 Introduction 235
7-2 Association Coefficients for Term and Document Vectors 236
7-3 Search Strategies and File Organization 243
Direct File Organization 244
Inverted File Organization 245
The Combined File System 250
The Multilist System 252
Multilevel Searching 254
7-4 Iterative Search Procedures Using Feedback 266
7-5 Adaptive Real-time Information Retrieval 275
References 278

Chapter 8 The Evaluation of Computer-based Retrieval Systems 280
8-1 Introduction 280
8-2 Evaluation Environment 282

8-3 Measures Based on Recall and Precision 283
8-4 Methods for Determining the Recall Value 293
8-5 The Presentation of Results 296
Standard Procedures 296
Programmed Evaluation 301
Output Summary 304
Significance Computations 310
8-6 The Design of Automatic Information Systems 315
Test Environment 316
Document Length 319
Matching Functions and Term Weights 322
Language Normalization—The Suffix Process 328
Synonym Recognition 330
Phrase Recognition 334
Hierarchical Expansion 341
Manual Indexing 343
Iterative Searching 345
8-7 Concluding Comments 346
References 348

Chapter 9 Auxiliary Information Services 350
9-1 Introduction 350
9-2 Special-purpose Equipment 351
9-3 Source Data Automation 354
Input Conversion 354
Text Editing Systems 355
Automatic Publication Methods 357
9-4 Text Transformations 359
Item Significance 359
Extracting, Abstracting, and Paraphrasing Systems 362
9-5 Index and Glossary Production 364
Index Types 364
Term Oriented Indexing 367
Document Oriented and Special-purpose Indexes 376
9-6 Citation Indexing 378
9-7 Selective Dissemination of Information 381
References 383

Chapter 10 Data Base Retrieval Systems 387
10-1 Introduction 387
10-2 Manipulation of Simple Data Files 388
10-3 Retrieval Languages 390

10-4 Automatic Question Answering Systems 392
Introduction 392
Basic Retrieval Framework 394
Organization of the Data Base 397
Semantic Interpretation 401
Extensions of the Data Base 405
Outlook 412
10-5 On-line Information Retrieval Systems 413
System Characteristics 413
A Sample Conversational System 416
References 419

Appendix A Smart Systems Organization 422
A-1 Introduction 422
A-2 Processing Summary 425
A-3 Basic Operating System 429
A-4 Processing Specifications 434
Dictionary Look-up 440
Concept Associations 442
Hierarchical Expansions 443
Vector Formation 444
Request-document Correlation 445
Document-Document Expansion 446
Author, Journal, Citation, and Cluster Indexing 447
Miscellaneous Run Specifications 448
Cluster Searching 449
Relevance Feedback 449
A-5 Data Preparation 450
Data Input 450
Dictionary Input 455
Auxiliary Programs 464
A-6 Sample Input Deck 465
A-7 Typical Processing Sequence 467
References 483

Appendix B Selective Bibliography in Information Organization and Retrieval 485

Name Index 499

Subject Index 505