• ORGANIZING INFORMATION
Principles of Data Base and Retrieval Systems
LIBRARY AND INFORMATION SCIENCE

consulting Editors: Harold Borko and G. Edward Evans
GRADUATE SCHOOL OF LIBRARY SCIENCE
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Thomas H. Mott, Jr., Susan Artandi, and Leny Struminger
Introduction to PL/I Programming for Library and Information Science

Karen Sparck Jones and Martin Kay
Linguistics and Information Science

Manfred Kochen (Ed.)
Information for Action: From Knowledge to Wisdom

Harold Borko and Charles L. Bernier
Abstracting Concepts and Methods

F. W. Lancaster
Toward Paperless Information Systems

H. 5. Heaps
Information Retrieval: Computational and Theoretical Aspects

Harold Borko and Charles L Bernier
Indexing Concepts and Methods

Gerald jahoda and Judith Schiek Braunagel
The Librarian and Reference Queries: A Systematic Approach

Charles H. Busha and Stephen P. Harter
Research Methods in Librarianship: Techniques and Interpretation

Diana M. Thomas, Ann T. Hinckley, and Elizabeth R. Eisenbach
The Effective Reference Librarian

James Cabeceiras
The Multimedia Library, Second Edition: Materials Selection and Use

C. Edward Evans
Management Techniques for Librarians, Second Edition

Irene P. Codden (Ed.)
Library Technical Services: Operations and Management

Jessica L Milstead
Subject Access Systems: Alternatives in Design

Dagobert Soergel
Organizing Information: Principles of Data Base and Retrieval Systems
# Contents

Preface xiii

I THE SYSTEMS APPROACH TO INFORMATION TRANSFER

1 INFORMATION SYSTEMS FOR PROBLEM SOLVING 3

THE NATURE OF INFORMATION 9

Objectives 9

2.1 The Role of the Image 9

2.2 Image Formation by a Solitary Individual 11

2.3 Image Formation through Communication 12

2.4 A Model of Information Transfer and Use 14

2.5 Data, Information, and Knowledge 16

2.6 Classification of Information Systems by Services Delivered 18

2.7 Summary and Evolving Principles 20

3 THE STRUCTURE OF INFORMATION 21

Objective 21

Introduction 21

3.1 The University Data Base: An Example 21

3.2 Elements of Information Structure 23

3.3 Analyzing Reference Tools: An Example 32

4 THE INFORMATION TRANSFER NETWORK 33

Objectives 33

Introduction 33

4.1 Transactions in the Information Transfer Network 34

4.2 Configurations of Transactions 35

4.3 Characteristics of Transactions 37
## 5 THE STRUCTURE OF INFORMATION SYSTEMS

<table>
<thead>
<tr>
<th>Objective</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 The Overall Structure of Information Systems</td>
<td>41</td>
</tr>
<tr>
<td>5.1.1 Identifying the Needs of Specific Users</td>
<td>41</td>
</tr>
<tr>
<td>5.1.2 Acquiring Entities or Information about Them</td>
<td>44</td>
</tr>
<tr>
<td>5.1.3 The ISAR System</td>
<td>45</td>
</tr>
<tr>
<td>5.1.4 Making Entities or Information Available to the User</td>
<td>45</td>
</tr>
<tr>
<td>5.1.5 Further Processing of Information</td>
<td>46</td>
</tr>
<tr>
<td>5.1.6 Identifying Needs in General: The Needs Directory</td>
<td>46</td>
</tr>
<tr>
<td>5.1.7 Public Relations</td>
<td>49</td>
</tr>
<tr>
<td>5.1.8 Functional versus Organizational Breakdown of a System</td>
<td>50</td>
</tr>
<tr>
<td>5.2 The Retrieval Problem: A View from Scratch</td>
<td>50</td>
</tr>
<tr>
<td>5.2.1 The Structure of Index Languages and Files: A Preview</td>
<td>56</td>
</tr>
<tr>
<td>5.3 The Structure of an ISAR System</td>
<td>57</td>
</tr>
<tr>
<td>5.3.1 The ISAR System as a Whole</td>
<td>59</td>
</tr>
<tr>
<td>5.3.2 System Rules and Conventions. The Conceptual Schema</td>
<td>59</td>
</tr>
<tr>
<td>5.3.3 Notes on the Other Components of an ISAR System</td>
<td>60</td>
</tr>
<tr>
<td>5.4 Definitions</td>
<td>62</td>
</tr>
<tr>
<td>5.4.1 Descriptor, Lead-in Term, Index Language, Lead-in Vocabulary, Thesaurus</td>
<td>62</td>
</tr>
<tr>
<td>5.4.2 Search Request, Query, Query Statement, Query Formulation</td>
<td>62</td>
</tr>
<tr>
<td>5.4.3 Indexing, Cataloging, Coding</td>
<td>63</td>
</tr>
</tbody>
</table>

## II OBJECTIVES OF ISAR SYSTEMS

<table>
<thead>
<tr>
<th>6 SYSTEMS ANALYSIS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives</td>
<td>69</td>
</tr>
<tr>
<td>Introduction</td>
<td>69</td>
</tr>
<tr>
<td>6.1 Approaches to Decision Making</td>
<td>70</td>
</tr>
<tr>
<td>6.2 Functions in the Systems Analysis Process</td>
<td>71</td>
</tr>
<tr>
<td>6.3 Phases in Systems Analysis: System Life Cycle</td>
<td>79</td>
</tr>
<tr>
<td>6.4 Information and Data Collection in Systems Analysis</td>
<td>84</td>
</tr>
<tr>
<td>6.5 Selection Decisions</td>
<td>86</td>
</tr>
<tr>
<td>6.6 Resource-Oriented versus Procedure-Oriented Systems Analysis</td>
<td>89</td>
</tr>
<tr>
<td>6.7 Performance versus Impact of Information Services</td>
<td>90</td>
</tr>
</tbody>
</table>

## 7 ASSESSMENT OF USERS’ PROBLEMS AND NEEDS

<table>
<thead>
<tr>
<th>Objective</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 User Studies as a Basis for System Design</td>
<td>93</td>
</tr>
<tr>
<td>7.2 Principles for the Study of Needs</td>
<td>93</td>
</tr>
<tr>
<td>Introduction</td>
<td>93</td>
</tr>
<tr>
<td>7.2.1 Setting Priorities</td>
<td>97</td>
</tr>
<tr>
<td>7.2.2 Shared Responsibility—User and Information Professional</td>
<td>98</td>
</tr>
<tr>
<td>7.2.3 Need, Want, Demand or Recognized Need, Use, and Impact</td>
<td>98</td>
</tr>
<tr>
<td>7.2.4 Unencumbered Assessment of Needs</td>
<td>100</td>
</tr>
</tbody>
</table>
7.3 Approaches to Studying Needs
7.3.1 Problem Analysis Based on Records 102
7.3.2 Analysis of Requests and Searching Behavior 103
7.3.3 Querying (Potential) Users about Their Needs 104

8 OBJECTIVES OF ISAR SYSTEMS 109
Objectives 109
Introduction 109
8.1 A Measure of Answer Quality (Local Performance) 114
8.1.1 Measures for Individual Aspects of Answer Quality 114
8.1.2 A Single Composite Measure of Answer Quality 123
8.2 A Measure of Global ISAR System Performance 125
8.3 Testing versus Evaluation 126
8.4 Relevance and Relevance Judgments 127
8.5 Implications for ISAR System Design and Operation 129
8.5.1 Implications for System Design: Innovative Features 129
8.5.2 Implications for System Operation 131
8.5.3 Relevance Judgments and Professionalism 132

III DATA SCHEMAS AND DATA STRUCTURES

9 DATA SCHEMAS AND FORMATS 137
Objectives 137
Introduction 137
9.1 Designing a Conceptual Schema 138
9.1.1 Rules and Conventions for the Form of Entity Identifiers 142
9.1.2 General Rules for Establishing Relationships 143
9.2 Record Formats: General Considerations 144
9.2.1 Functions of Records in Data Base Processes 144
9.2.2 Structure as a Key Concept 146
9.2.3 Fixed Field and Variable Field Records 148
9.3 Criteria for the Design and Evaluation of Data Schemas 150
9.4 Input Formats 152
9.4.1 Design Considerations for Input Formats 155
9.5 Output Formats 157
Appendix 9.1 Examples of Record Formats 159

10 ELEMENTARY QUERY FORMULATION 165
Objectives 165
10.1 Logical NOT> 165
10.2 Logical OR 166
10.3 Logical AND with Logical OR 167
10.4 The Ambiguity of Natural Language “and” 168
10.5 Levels of Parentheses 169
10.6 Logical NOT and Its Pitfalls 170
## 11 DATA STRUCTURES AND ACCESS

**Objectives**

11.1 Exploration of Data Structures ................................................. 173
11.2 Functions and Characteristics of Data Structures ..................... 195
11.3 Main File and Index File(s) as a Data Structure ......................... 196
   11.3.1 Notes on Terminology ..................................................... 196
   11.3.2 Some Usage and Design Considerations ............................... 197
   11.3.3 A Main File of Entities as Part of the Data Structure .......... 198
11.4 The Concept of Order .......................................................... 199
11.5 The Two-Dimensional Continuum of File Types ......................... 202
11.6 Trade-offs between Data Base Costs and Searching Costs ............ 203
   11.6.1 Amount of Information .................................................. 203
   11.6.2 Degree of Order .......................................................... 204
   11.6.3 Design Considerations .................................................. 204
11.7 Definitions ............................................................................. 206

---

## IV INDEX LANGUAGE FUNCTIONS AND STRUCTURE

## 12 TERMINOLOGICAL CONTROL

**Objectives**

Introduction: The Problem of Terminological Control ..................... 213
12.1 Concepts versus Terms: The Synonym-Homonym Structure ............ 217
12.2 Grouping Closely Related Concepts: The Equivalence Structure .... 219
12.3 Classificatory Structure ....................................................... 220
12.4 Index Language ...................................................................... 221
12.5 Thesaurus .............................................................................. 222

## 13 INDEX LANGUAGE FUNCTIONS

**Objective**

13.1 Review: The Information Retrieval Problem ............................ 225
13.2 The Role of the Index Language in Indexing ............................. 227
   13.2.1 Disadvantages of Entity-Oriented Indexing ......................... 227
   13.2.2 Request-Oriented Indexing: General Approach .................... 230
   13.2.3 Request-Oriented Indexing: Implementation ....................... 233
   13.2.4 Supplementary Entity-Oriented Indexing ............................. 236
13.3 The Role of the Index Language in Searching ........................... 239
   13.3.1 The Checklist Technique Applied to Query Formulation ......... 239
   13.3.2 Compensating for the Lack of Request-Oriented Indexing ......... 240
   13.4 The Role of the Index Language in Data Base Organization ....... 240
13.5 Choosing the Best Indexing Approach ..................................... 244
   13.5.1 Cost of Indexing ............................................................. 244
13.5.2 Quality of Indexing 245
13.5.3 Cost and Quality of Searching 246
13.6 The Functions of Hierarchy: A Summary 246
13.7 A Philosophy of Indexing and Classification 247

14 INDEX LANGUAGE STRUCTURE 1: CONCEPTUAL 251
Objectives 251
introduction 251
14.1 Hierarchy 252
14.2 Concept Combination and Semantic Factoring.
Facet Analysis 256
14.3 Interaction between Concept Combination and Hierarchy 261
14.4 Application and Illustration: Searching 272
14.5 Conceptual Analysis, Facet Analysis: Elaboration 278
14.5.1 Developing a Scheme of Facets 278
14.5.2 Recognizing General Concepts 278
14.5.3 Subfacets 280
14.5.4 Facet Analysis and Relationships among Precombined
Descriptors 280
14.5.5 Advantages of Semantic Factoring and Facet Analysis 280
14.6 Hierarchy: Elaboration 281
14.6.1 Hierarchical versus Associative Relationships 281
14.6.2 Types of Hierarchical Relationships 282
14.6.3 Introducing New Broader Concepts 283
14.7 Concept Formation in Thesaurus Building 285

15 INDEX LANGUAGE STRUCTURE 2:
DATA BASE ORGANIZATION 289
Objectives 289
Introduction 289
15.1 The Problem 290
15.2 Grouping Entities, Searching in Grouped Files 291
15.2.1 The Idea of Grouping and Precombined Descriptors 291
15.2.2 From Ideal to Reality: Limited Precombination 299
15.2.3 Access Advantages of Grouped Files 303
15.3 Grouping versus Description of Entities 303
15.4 Postcombination and Precombination 305
15.4.1 Postcombination versus Precombination as a Matter
of Degree 305
15.4.2 Deciding on the Overall Degree of Precombination 307
15.4.3 Deciding on Individual Precombined Descriptors 308
15.4.4 Precombined Descriptors in Indexing and Searching 310
15.5 Organizing an Index Language for Access 312
15.5.1 Descriptor-Find Indexes 313
15.5.2 Arrangement and Designation of Descriptors 317
15.6 A Unified Index Language for Different Search Mechanisms 322
V
ISAR SYSTEMS OPERATION AND DESIGN

16 INDEXING SPECIFICITY AND EXHAUSTIVITY 327
   Objectives 327
   16.1 Importance of Indexing for System Performance 327
   16.2 Definition of Exhaustivity and Specificity of Indexing 328
      16.2.1 Definition of Exhaustivity 328
      16.2.2 Definition of Specificity 330
   16.3 Effects of Exhaustivity and Specificity of Indexing on Retrieval Performance 331
      16.3.1 Effects of Exhaustivity 332
      16.3.2 Effects of Specificity 336
      16.3.3 Misconceptions about the Effects of Exhaustivity and Specificity 337
      16.3.4 Summary 338
   16.4 Designing Indexing Rules and Procedures 338
      16.4.1 Factors Influencing Exhaustivity 338
      16.4.2 Factors Influencing Specificity 339
      16.4.3 Cost Considerations 340

17 SEARCHING 343
   Objective 343
   Introduction 343
   17.1 Recognize and State the Need. State Search Requirements 346
      17.1.1 Recognize the Existence of a Need 346
      17.1.2 Develop the Query Statement 347
      17.1.3 Determine Specific Search Requirements 350
   17.2 Develop the Search Strategy 350
      17.2.1 Formulate the Query Conceptually 351
      17.2.2 Select Sources and Arrange Them in a Search Sequence 359
      17.2.3 Translate the Conceptual Query Formulation into the Language of Each Source 362
      17.2.4 Free-Text Searching 366
      17.2.5 The Interplay between Conceptual and Source-Specific Query Formulation 368
   17.3 Execute the Search Strategy 368
   17.4 Review Search Results and Revise Search 369
   17.5 Edit Search Results and Send Them to the User 370
   17.6 Check Whether the Answer Was Helpful 371
   17.7 Interaction 371
   17.8 Monitor the Search Process and Assess Results 376

18 DESIGN AND EVALUATION OF INFORMATION SYSTEMS 379
   Objective 379
   Introduction 379
18.1 Determine User Requirements and Abilities 382
  18.1.1 Determine the Scope of the Information System 382
  18.1.2 Obtain Search Requests 383
18.2 Develop the Collection and Obtain Relevance Judgments 384
  18.2.1 Develop the Collection of Entities 384
  18.2.2 Obtain Presearch Relevance Judgments 384
18.3 Design and Construct the ISAR System 385
18.4 Operate the ISAR System 387
  18.4.1 Index Entities 387
  18.4.2 Formulate Queries 387
  18.4.3 Retrieve Entities and Judge Their Relevance 389
18.5 Evaluate ISAR System Performance 389
  18.5.1 Macroanalysis: Performance Measures 389
  18.5.2 Microanalysis: Retrieval Successes and Failures 390
18.6 Retrieval Testing and System Design and Operation 390
18.7 Cost-benefit Analysis as a Design Principle 392
18.8 Problem-Orientation as a Design Principle 395

Bibliography 399
Author Index 423
Subject Index 429
This book gives a theoretical base and a perspective for the analysis, design, and operation of information systems, particularly their information storage and retrieval (ISAR) component, whether mechanized or manual. Information systems deal with many types of entities: events, persons, documents, business transactions, museum objects, research projects, and technical parts, to name a few. Among the purposes they serve are to inform the public, to support managers, researchers, and engineers, and to provide a knowledge base for an artificial intelligence program. The principles discussed in this book apply to all these contexts. The book achieves this generality by drawing on ideas from two conceptually overlapping areas—data base management and the organization and use of knowledge in libraries—and by integrating these ideas into a coherent framework. The principles discussed apply to the design of new systems and, more importantly, to the analysis of existing systems in order to exploit their capabilities better, to circumvent their shortcomings, and to introduce modifications where feasible.

This book is intended for use in an introductory course on organizing and retrieving information (called, for example, “Introduction to the Organization of Information,” “Introduction to Information Storage and Retrieval,” or “Introduction to Information Science”) offered in a school of library and information science, a business school, or a more broadly based information studies or information management program. Beyond that, it is meant to inspire, a modernization and integration of the library/information science curriculum. The book can be used for a broadly based course that teaches the general principles of ISAR and treats cataloging and reference service as specific areas of application. Such a course not only overcomes the artificial separation of cataloging and reference but also gives students wide flexibility in choosing their first position and a sound base from which to strike out in many directions in the further development of their careers. It can be offered as a package of designated sections of the cataloging and reference course, without changing any course numbers. Such a course can be extended to include students from business infor-
mation systems, journalism, and cognate areas: The theoretical base is common to all, but the application areas are different. This book is also suitable for self-study by practitioners who are looking for a sounder theoretical base for their daily work. A workbook with exercises and discussion of additional examples is in preparation; a draft is available from the author.

Information studies is a nascent field. It shows considerable confusion in its terminology, partly due to the lack of a prevalent conceptual framework: The same term is used for different concepts; different terms are used for the same concept. This book follows the terminology of major writers in the field but sometimes introduces a new term for a new concept or to replace an existing term that reflects a faulty concept analysis.

Throughout the book the development of ideas proceeds from the point of view that information specialists are professionals who cooperate with the user in determining information needs and who use their knowledge to design systems or to do searches to meet these needs, as opposed to merely looking for what the user thinks is needed.

Organization of the Book. Part I places information systems in context; it discusses the nature and structure of information and lays out the overall structure of an information system. Part II provides the basis for considering the design and use of ISAR systems in light of the objectives to be achieved, allowing for a discussion of the merits of design alternatives in Parts III through V. Part III deals with the logical representation of data and with structures for providing access to these data. It deals on a general level with the rules and conventions necessary in an ISAR system. Part IV focuses attention on subject retrieval (but many of the principles have more general application). It discusses the nature of index languages—terminological control, basic functions, and conceptual structure. Part V discusses indexing and searching and, in conclusion, testing and design of the system.

Acknowledgments. This book was developed from lectures given at the University of Maryland; the students’ many questions forced me to sharpen my thinking, and their comments on successive versions of the manuscript were extremely useful. Norman Roberts, Harold Borko, and Raya Fidel all gave good advice, which, among other things, was instrumental in reducing this book to a manageable size. Jane Bergling and Marie Somers typed the manuscript many times over, somehow managing to interpret my scribbled revisions. My wife Lissa was always ready to examine and discuss ideas and to make suggestions concerning both content and form; she also spent hours editing and proofreading. I owe an intellectual debt to many in the field but above all to the pioneering spirit of Calvin Mooers.