Allen Kent:
INFORMATION ANALYSIS AND RETRIEVAL

Robert S. Taylor:
THE MAKING OF A LIBRARY

Herman M. Weisman:
INFORMATION SYSTEMS, SERVICES, AND CENTERS

Jesse H. Shera:
THE FOUNDATIONS OF EDUCATION FOR LIBRARIANSHIP

Charles T. Meadow:
THE ANALYSIS OF INFORMATION SYSTEMS, Second Edition

Stanley L. Swihart and Beryl F. Hefley:
COMPUTER SYSTEMS IN THE LIBRARY

F. W. Lancaster and E. G. Fayen:
INFORMATION RETRIEVAL ON-LINE

Richard A. Kaimann:
STRUCTURED INFORMATION FILES

Dagobert Soergel:
INDEXING LANGUAGES AND THESAURUS: CONSTRUCTION AND MAINTENANCE

Thelma Freides:
LITERATURE AND BIBLIOGRAPHY OF THE SOCIAL SCIENCES

Manfred Kochen:
PRINCIPLES OF INFORMATION RETRIEVAL
Indexing Languages and Thesauri: Construction and Maintenance

DAGOBERT SOERGEL
College of Library and Information Services
University of Maryland

A WILEY-BECKER & HAYES SERIES BOOK

MELVILLE PUBLISHING COMPANY
1SI5) Los Angeles, California
Information Sciences Series

Information is the essential ingredient in decision making. The need for improved information systems in recent years has been made critical by the steady growth in size and complexity of organizations and data.

This series is designed to include books that are concerned with various aspects of communicating, utilizing, and storing digital and graphic information. It will embrace a broad spectrum of topics, such as information system theory and design, man-machine relationships, language data processing, artificial intelligence, mechanization of library processes, nonnumerical applications of digital computers, storage and retrieval, automatic publishing, command and control, information display, and so on.

Information science may someday be a profession in its own right. The aim of this series is to bring together the interdisciplinary core of knowledge that is apt to form its foundation. Through this consolidation, it is expected that the series will grow to become the focal point for professional education in this field.
Preface

This book has two objectives. First, to reassess thoroughly the functions of an indexing language or thesaurus in an information storage and retrieval system and in the light of this reassessment to analyze the structure of indexing languages and thesauri. Most importantly, this reassessment is based on a unified view of indexing languages (classification schemes) and thesauri as used in traditional libraries on the one hand and in modern (mechanized) information storage and retrieval systems on the other. It results in general principles that are applicable to a wide range of situations.

The second objective is to give a comprehensive overview of the state of the art of the display and the construction and maintenance of indexing languages and thesauri.

The first objective is a prerequisite for the second. A separate textbook on information storage and retrieval would perhaps be a more appropriate form to deal with it, and there are indeed plans for such a textbook incorporating much of the material presented in chapters B and C.

Information from many sources has been evaluated and synthesized to compile the state of the art of thesaurus construction and display as completely as possible. Some sources—for example, the rules used for the TEST thesaurus—have been referenced in detail in the footnotes. For other sources this was not possible.

I wish to give my acknowledgements for numerous examples that have been taken from Thesaurofacet, from Mandersloot et al. 1970, and from Thomas et al. 1953. In all other cases the sources of examples are given in the notes.

The table of contents has an unusual three-level format. This is to illustrate the display of a classification scheme with “summaries” or “synopses” on several levels.

I wish to thank the many people who contributed to the completion of this book. First of all, the book is based on the German “Klassifikationssysteme und Thesauri” which I wrote on behalf of the Committee for Thesaurus Research of the German Society for Documentation, using materials prepared by the committee and with the benefit of the review and comments of the committee members, especially Ingetraut Dahlberg and Alwin Diemer, Chairman (the other members of the committee were: R. Fugman, G. Heinz-

William Kurmey of the Faculty of Library Science of the University of Toronto thoroughly reviewed the whole manuscript and made many valuable suggestions to improve the content and the clarity of presentation. His contribution should enhance considerably the usefulness of the book. Katherine Packer, a member of the Faculty of Library Science of the University of Toronto and also a Ph.D. candidate at the University of Maryland, and my wife Lissa both read the whole manuscript for style and clarity and spent many hours with me discussing individual formulations.

Tom Wilson of the Graduate Library School of the University of Sheffield contributed various ideas, especially for chapter K, “Thesauri as a basis for cooperation in information services”. The acronym ISAR (or, as he prefers, isar) for Information Storage And Retrieval was first used by him.

The excellent work of Faith Bange, who diligently typed and retyped the many versions and expertly interpreted my handwriting, was a great help in finishing the manuscript.

Above all, my thanks go to Calvin Mooers who, back in 1962, encouraged my interest in the field and started a process of thought without which this book would not have been possible.

College Park, Maryland

DAGOBERT SOERGEL
HOW TO READ THIS BOOK

1*. This book is a handbook as well as a textbook; not every section is for every reader. Nor should every section be read in the first reading. Therefore, a number of sections have been marked as follows:

Technical—The information contained in these sections is not important for a general understanding of the problems and the procedure or for the overall planning of a thesaurus development project. The information is needed only as one comes to the step in question. These sections can therefore be omitted in the first reading.

Special topic—These sections deal with problems that occur only in special situations. They can therefore be omitted without any loss in understanding of other sections. An example is Section D5, “Multi-lingual thesauri”.

Advanced—These sections are meant only for the reader who is interested in depth.

A reader with sufficient background in the structure of indexing languages and thesauri might turn immediately to chapter F, “Flow of work in the construction of indexing languages and thesauri”, and return to previous sections as the need arises.

Readers who are interested only in a general orientation and those who have the task of constructing a small indexing language/thesaurus, need only read the sections of the book given in the following guide.

A reader interested only in a general orientation about indexing languages and thesauri and their role in an information storage and retrieval system should read the following sections:

Al;
B;
C through Cl.3;
possibly C2 through C2.5 (if interested in conventional systems like subject headings and shelving classification);
D through Dl.3.2
(omitting, of course, sections labeled “advanced”, “special topic”, or “technical”.)

A reader who has the task of constructing a small indexing language/thesaurus need not concern himself with the details and ramifications important for large systems. The following sections should provide sufficient information.
How to Read this Book

1. A;
2. B;
3. C through C4, C5, possibly C5.0 (but not C5.1-C5.3), C7;
4. D1-D3, D4.0, D4.1, D4.3.3, D4.4;
5. EO, ELO, E.1.1, E1.5, E1.6, E1.8;
6. F;
7. J;
8. K0, K1 through K1.2.1, K1.2.3.

The problems dealt with in this book are highly interrelated. A second reading might be helpful.

Chapters B and C provide a limited background in classification theory, they are not intended to give a full treatment of this topic. A number of good books are available (see the first note to chapter C), and it is strongly recommended that the reader who does not have the background in information storage and retrieval and classification consult one of these books first.

All notes are at the end of the book. They are identified by section number and are formulated in such a way that the particular point referred to is readily clear. This procedure made it possible to omit any numbers referring to notes from the main text and thus improve readability. The reader interested in further references and other background material for a given section should simply look in the back under the section number (and possibly under a broader section number).

The notes for each chapter or major subsection can also be considered as a bibliographical supplement that can be read separately.

Documents mentioned in the notes are cited by author and date of publication, e.g., Lancaster 1969.4. The bibliography is arranged by author.
# Content: Overview

## Introduction

A General overview of the functions and structure of a thesaurus. Major tasks to be performed and resources and work required for the construction of a thesaurus 3

## Part I Conceptual structure of indexing languages and thesauri

B Concepts and terms. Indexing language and thesaurus and their functions in an ISAR system 17

C The structure of indexing languages and thesauri 68

## Part II Presentation of indexing languages and thesauri

D Thesaurus format 183

E Rules concerning the form of terms and related problems 298

## Part III Procedures for the construction and maintenance of indexing languages and thesauri

F Flow of work in the construction of indexing languages and thesauri 325

G Use of computers in thesaurus construction 420

H Automatic methods in the construction of indexing languages and thesauri, starting from the texts of documents and/or search requests. Automatic classification 449

J Updating and maintenance of indexing languages and thesauri 457

## Part IV Thesauri as a basis for cooperation in information services

K Thesauri as a basis for cooperation in information services 471

## Appendices

Appendices 521

Chapter Notes 535

Bibliography 559

Index 609
Condensed Table of Contents

Introduction 1

A General overview of the functions and structure of a thesaurus.
   Major tasks to be performed in and resources and work
   required for the construction of a thesaurus 3
A0 Introduction 3
   A1 Overview of functions and structure of a thesaurus in an
      Information Storage And Retrieval (ISRA) system 3
   A2 Administrative considerations. Resources and work
      required for the development of a thesaurus 10

Part I Conceptual structure of indexing languages and thesauri 15

B Concepts and terms. Indexing language and thesaurus and
   their functions in an ISAR system 17
B0 Introduction 17
   B1 Plane of concepts versus plane of terms; the synonym-
      homonym structure 17
   B2 Treatment of nearly related concepts: the equivalence
      structure 22
   B3 Indexing language 26
   B4 Thesaurus. Summary of and further remarks on the
      definition of preferred term, descriptor, concept, and
      indexing language 29
   B5 The functions of the indexing language within an ISAR
      system. A preliminary overview of the structure of indexing
      languages in relation to their functions in an ISAR system 39
   B6 The functions of the lead-in vocabulary in an ISAR system 61
   B7 “User's” or “author's” vocabulary versus logical structure
      and request-oriented indexing as implemented through
      the checklist technique 66

C The structure of indexing languages and thesauri 68

xiii
### Condensed Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0</td>
<td>Introduction</td>
<td>68</td>
</tr>
<tr>
<td>C1</td>
<td>Classificatory structure</td>
<td>69</td>
</tr>
<tr>
<td>C2</td>
<td>Problems of file organization related to classification.</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>Practical applications of semantic factoring</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>Concept formation in thesaurus building. Definition and scope notes</td>
<td>142</td>
</tr>
<tr>
<td>C4</td>
<td>Types of concepts, descriptors, terms to be included in an indexing language or thesaurus</td>
<td>147</td>
</tr>
<tr>
<td>C5</td>
<td>The lead-in structure. USE and SEE</td>
<td>155</td>
</tr>
<tr>
<td>C6</td>
<td>Synonyms proper versus spelling variants</td>
<td>171</td>
</tr>
<tr>
<td>C7</td>
<td>Summary of relationships displayed in a thesaurus</td>
<td>174</td>
</tr>
<tr>
<td><strong>Part H</strong></td>
<td>Presentation of indexing languages and thesauri</td>
<td><strong>181</strong></td>
</tr>
<tr>
<td>D</td>
<td>Thesaurus format</td>
<td>183</td>
</tr>
<tr>
<td>DO</td>
<td>Introduction</td>
<td>183</td>
</tr>
<tr>
<td>D1</td>
<td>The different parts of a thesaurus</td>
<td>183</td>
</tr>
<tr>
<td>D2</td>
<td>Format of entries in the main part</td>
<td>228</td>
</tr>
<tr>
<td>D3</td>
<td>How to display descriptors and their interrelationships (methods for the design of a classified index)</td>
<td>235</td>
</tr>
<tr>
<td>D4</td>
<td>Notation</td>
<td>273</td>
</tr>
<tr>
<td>D5</td>
<td>Multilingual thesauri</td>
<td>293</td>
</tr>
<tr>
<td><strong>E</strong></td>
<td>Rules concerning the form of terms and related problems</td>
<td>298</td>
</tr>
<tr>
<td>EO</td>
<td>Introduction. Difference in requirements between systems using notations and systems using terms</td>
<td>298</td>
</tr>
<tr>
<td>E2</td>
<td>Spelling and transliteration</td>
<td>317</td>
</tr>
<tr>
<td>E3</td>
<td>Alphabetization</td>
<td>320</td>
</tr>
<tr>
<td><strong>Part III</strong></td>
<td>Procedures for the construction and maintenance of indexing languages and thesauri</td>
<td><strong>323</strong></td>
</tr>
<tr>
<td>F</td>
<td>Flow of work in the construction of indexing languages and thesauri</td>
<td>325</td>
</tr>
<tr>
<td>F0</td>
<td>Overview and general problems</td>
<td>325</td>
</tr>
<tr>
<td>F1</td>
<td>Collect and record material (concepts, terms, relationships between and among them)</td>
<td>355</td>
</tr>
<tr>
<td>F2</td>
<td>Sort into alphabetical order and merge information on identical terms on one card</td>
<td>366</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>Condensed Table of Contents</strong></td>
<td>xv</td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>Work out the preliminary structure of the thesaurus: the synonym-homonym structure, the equivalence structure, and the classificatory structure. Select preferred terms</td>
<td>384</td>
</tr>
<tr>
<td>F4</td>
<td>Work out first draft of the classified index (schedule)</td>
<td>392</td>
</tr>
<tr>
<td>F5</td>
<td>Complete first draft of the thesaurus as a whole</td>
<td>397</td>
</tr>
<tr>
<td>F6</td>
<td>Test the thesaurus by indexing and retrieval experiments</td>
<td>411</td>
</tr>
<tr>
<td>F7</td>
<td>Duplicate or print the user version of the thesaurus</td>
<td>412</td>
</tr>
<tr>
<td>F8</td>
<td>Further remarks concerning the work-flow and modifications of the standard work-flow</td>
<td>413</td>
</tr>
<tr>
<td>F9</td>
<td>Use of punched paper tape and punched cards in thesaurus construction</td>
<td>417</td>
</tr>
<tr>
<td></td>
<td>Use of computers in thesaurus construction</td>
<td>420</td>
</tr>
<tr>
<td>GO</td>
<td>Rationale for computer application. Overview</td>
<td>420</td>
</tr>
<tr>
<td>G1</td>
<td>Computer assistance in the collection and recording of material</td>
<td>428</td>
</tr>
<tr>
<td>G2</td>
<td>Computer assistance in sorting into alphabetical order and in merging information on identical terms into one record</td>
<td>429</td>
</tr>
<tr>
<td>G3</td>
<td>Computer assistance in working out the preliminary structure of the thesaurus</td>
<td>434</td>
</tr>
<tr>
<td>G4</td>
<td>Computer assistance in working out the classified index</td>
<td>441</td>
</tr>
<tr>
<td>G5</td>
<td>Computer assistance in completing the first draft of the thesaurus as a whole</td>
<td>442</td>
</tr>
<tr>
<td>G7</td>
<td>Printing the final thesaurus by computer</td>
<td>443</td>
</tr>
<tr>
<td>G8</td>
<td>Updating a computer-stored thesaurus</td>
<td>443</td>
</tr>
<tr>
<td>G9</td>
<td>Devices for the input (keying) of thesaurus data</td>
<td>447</td>
</tr>
<tr>
<td></td>
<td>Automatic methods in the construction of indexing languages and thesauri, starting from the texts of documents and/or search requests. Automatic classification</td>
<td>449</td>
</tr>
<tr>
<td>HO</td>
<td>Introduction</td>
<td>449</td>
</tr>
<tr>
<td>HI</td>
<td>Definition of units of text and counting methods</td>
<td>450</td>
</tr>
<tr>
<td>H2</td>
<td>Identification of descriptor candidates from frequency patterns</td>
<td>451</td>
</tr>
<tr>
<td>H3</td>
<td>Detection of terms or concept relationships from co-occurrence patterns</td>
<td>451</td>
</tr>
<tr>
<td>H4</td>
<td>Automatic derivation of classification schemes (“global” structures)</td>
<td>455</td>
</tr>
</tbody>
</table>
J Updating and maintenance of indexing languages and thesauri 457

JO Introduction 457
J1 Types of changes 457
J2 Sources for new terms, concepts and relationships to be included in the thesaurus 458
J3 Procedures for regular updating 460
J4 Revision of the indexing language or the thesaurus at longer intervals 463
J5 Remarks on the flexibility of structured indexing languages (classification schemes) 464
J6 Problems of re-indexing (re-classification) 465
J7 Thesaurus updating and thesaurus compatibility: common problems 467

Part IV Thesauri as a basis for cooperation in information services 469

KO Introduction 471
K1 Cooperation in the construction of indexing languages and thesauri 472
K2 Cooperation through sharing the results of subject indexing 493
K3 The idea of a Universal Source Thesaurus (UST) 516

Appendices 521

Chapter Notes 535

Bibliography 559

Index 609
# Contents

Preface vii  
How to Read this Book ix  
Content: Overview xi  
Condensed Table of Contents xiii  
Table of Contents xvii  
List of Figures xli

**Introduction** 1  
A General overview of the functions and structure of a thesaurus. 3  
Major tasks to be performed in and resources and work required for the construction of a thesaurus 3  
A0 Introduction 3  
A1 Overview of functions and structure of a thesaurus in an Information Storage and Retrieval (ISAR) system 3  
A1.1 Requirements for an ISAR system: conceptual structure and terminological control 3  
A1.2 Thesaurus 4  
A1.3 The thesaurus in the context of an ISAR system as a whole 5  
A1.3.1 Parameters determining thesaurus size 6  
A1.3.2 Estimating parameters; dangers of a thesaurus of inappropriate size or quality 7  
A1.4 Use of a thesaurus for improving indexes 8  
A1.5 Use of a classification scheme or a thesaurus for purposes other than ISAR 8  
A1.6 Intellectual problems in the development of a thesaurus 9  
A1.7 Criteria for the evaluation of a thesaurus 9  
A1.8 Concluding remarks 9  
A2 Administrative considerations. Resources and work required for the development of a thesaurus 10  
A2.1 Justifying the creation of a new thesaurus 10  
A2.2 Staff needed for the development of a thesaurus 10  

xvii
xviii Contents

A2.3 Time-frame for the development of a thesaurus 12
A2.4 Necessity of continuous updating 13

Part I Conceptual structure of indexing languages and thesauri

B Concepts and terms. Indexing language and thesaurus and their functions in an ISAR system 15
B0 Introduction 17
B1 Plane of concepts versus plane of terms: the synonym-homonym structure 17
B1.1 Homonyms and homographs (advanced) 20
B2 Treatment of nearly related concepts: the equivalence structure 22
B2.1 Equivalent concepts (equivalent terms) 22
B2.2 Classificatory structure 23
B2.3 Summary of B1 and B2 23
B2.4 A more realistic but less practical model 26
B3 Indexing language 26
B3.0 Definition of “descriptor 1 (retrieval cue)”, “descriptor 2 (subject descriptor)”, and “indexing language” 26
B3.1 Remark on terminological control 29
B4 Thesaurus. Summary of and further remarks on the definition of preferred term, descriptor, concept, and indexing language 29
B4.1 Simple definition of “thesaurus”. Use of the lead-in structure in indexing 29
B4.2 Summary of and further remarks on the definition of preferred term, descriptor, concept, indexing language, and thesaurus 30
B4.2.0 Summary of definitions 31
B4.2.1 Preferred term and descriptor 31
B4.2.2 Concept, preferred term, descriptor 34
B4.2.3 Indexing language, system vocabulary, classification scheme 34
B4.2.4 Subject access vocabulary, thesaurus, indexing language, and classification scheme 35
B4.3 Complex thesaurus structures. Use of the lead-in structure for terminological control in searching (advanced) 36
B4.4 Formal definition of “thesaurus” (advanced) 38
B5 The functions of the indexing language within an ISAR system. A preliminary overview of the structure of indexing languages in relation to their functions in an ISAR system 39
B5.0 “Indexing” versus “grouping of documents” I. Solutions to the retrieval problem 39
xx Contents

B6.1 Thesaurus as store of intellectual decisions made in day-to-day indexing and search request formulation 62

B6.1.2 Gradual development of a thesaurus over time 63

B6.2 Mechanization of indexing or search request formulation (special topic) 63

B6.2.1 Semi-mechanized versus fully mechanized indexing and search request formulation: description of methods 63

B6.2.2 Problems and implications for thesaurus building 64

B6.2.3 Discussion: advantages and disadvantages of mechanized indexing 65

B6.3 Thesauri for terminological control in the searching stage 66

B7 “User’s” or “author’s” vocabulary versus logical structure and request-oriented indexing as implemented through the checklist technique 66

C The structure of indexing languages and thesauri 68

CO Introduction 68

C1 Classificatory structure 69

C1.0 Introduction: “representation” versus “grouping of documents” II (continuation of Section B5.0) 69

C1.1 Decomposition of concepts into semantic factors—concept combination (concept coordination) 74

C1.1.0 Foundations of semantic factoring and concept combination 74

C1.1.1 Advantages of semantic factoring 77

C1.2 Polyhierarchical structure. Definition of hierarchy 78

C1.3 Interaction of hierarchy and concept combination 83

C1.3.1 Limitations of the model for the generation of hierarchical structures 91

C1.3.2 Application of the model to hierarchy construction. Facet analysis 91

C1.4 Further topics in hierarchy and its use in indexing and searching 95

C1.4.1 Further considerations on pragmatic hierarchy building 95

C1.4.1.1 Extending the definition of concepts or introducing new broader concepts 95

C1.4.1.2 Introduction of additional broader concepts for searching 95

C1.4.1.3 Introduction of a broader concept to replace a number of specific concepts 97

C1.4.1.4 Introduction of new broader concepts to serve as headings ("organizational headings") 97

C1.4.1.5 Antonyms 97
Contents xxii

6 Hierarchical relationships versus associative relationships in indexing and searching 98
7 Introduction of new broader concepts as a creative activity 99

Cl.4.2 Kinds of hierarchical relationships (advanced) 99
Cl.4.3 Special cases of hierarchical structure 102
1 Coarse hierarchy: subdivision of the preferred terms into subject fields 102
2 Facets 103
Cl.4.4 “General descriptors” and “Other descriptors” as a special type of heading applicable throughout the classification scheme 103
Cl.4.5 Implementation of inclusive searching. Generic posting and the POST TO instruction 103
Cl.4.6 Descriptors “... inclusive” and “... general references” 105
1 Descriptor usage depends on hierarchy 106
Cl.4.7 Descriptors ... other” 106

Cl.5 Associative relationships between concepts 107
Cl.5.1 Concepts similar in meaning 108
Cl.5.2 Concepts connected empirically (“contextual contiguity”) 108
1 Contiguity based on definition 108
2 Contiguity based on empirical knowledge 109
3 Contiguity and frequency of combination 109
0.5.3 Instructional scope note 109
Cl.5.6 Transitions between the synonym-homonym structure, the equivalence structure, and the classificatory structure 110
Cl.5.7 Psychological dimensions of relationships 112

C2 Problems of file organization related to classification. Practical applications of semantic factoring

C2.0 Introduction 112
C2.1 On the relationship between conceptual structure and file organization in classification theory 113
C2.2 The problem defined 114
C2.3 Principal solutions: post-combination versus pre-combination—a quantitative view 115

C2.3.1 Post-combination and pre-combination (post-coordination and pre-coordination) defined 115
C2.3.2 Conceptual indexes (auxiliary ISAR systems) I (special topic, for systems using pre-combination only) 119
xxii Contents

C2.3.3 Summary 121
C2.4 Selection and arrangement of descriptors with particular reference
to ISAR systems using pre-combination (special topic) 122
C2.5 A unified classification scheme for different kinds of file organi-
ization: core classification and extended classification (partly
special topic) 126
C2.5.1 Special problems arising in the implementation of this
proposal (technical) 128
 ,1 Multiple entry versus entry under a precombined
descriptor 128
 ,2 Multiple entry using a faceted classification 129
C2.6 Summary: strategies for the application of semantic factoring 129
C2.6.1 Considerations in the choice of a strategy 130
C2.7 Rules for the use of precombined descriptors in an intermediary
strategy not using roles and links 131
C2.7.1 What compound concepts should be used as precombined
descriptors? 132
C2.7.2 What compound concepts should be represented by a
combination of descriptors rather than by a precombined
descriptor? 134
C2.8 Optimization of an indexing language with a constraint as to the
number of descriptors (advanced) 134
C2.8.1 Semantic super-imposed coding 135
C2.8.2 Considerations to be taken into account in reducing the
number of concepts used as descriptors 137
C3 Concept formation in thesaurus building. Definition and scope notes 142
C3.1 Concept formation in thesaurus building 142
C3.2 Definition and scope note 145
C3.2.0 Introduction 145
C3.2.1 Formal definition 145
C3.2.2 Scope notes for the thesaurus user 146
C4 Types of concepts, descriptors, terms to be included in an indexing
language or a thesaurus 147
C4.0 Typology of concepts, descriptors, terms 148
C4.0*1 Remarks on descriptive versus subject indexing 149
C4.1 The treatment of proper names used as subject descriptors 151
C4.2 Treatment of elements of nomenclatures 152
C4.2.1 Nomenclatures as adjunct thesauri 152
C4.2.2 Alternative possibility: inclusion of selected elements of
nomenclatures into the thesaurus proper 153
C4.3 Concepts of general application (common attributes, common isolates) 153

C5 The lead-in structure. USE and SEE 155

C5.0 Introduction 156

C5.0.1 The lead-in problem: alphabetical index method versus main part method 156
C5.0.2 The lead-in problem: the crude form and the detailed form 158

Further illustrative examples of the crude lead-in form 160

C5.0.3 Use of the lead-in structure 161

C5.1 The detailed lead-in form 162

C5.2 Alternative lead-in forms 165

C5.2.1 Simpler forms 165
C5.2.2 More detailed form: expressing the equivalence structure (advanced) 166

C5.3 OR-type USE instructions 167

C5.3.1 Homonymous lead-in terms 168
C5.3.2 Broad lead-in terms 168
C5.3.3 Leads to related terms 169
C5.3.4 OR-combination of descriptors as semantic factor 170

C5.4 Other matters related to USE instructions 171

C5.4.1 Qualified USE instructions (special topic) 171

C6 Synonyms proper versus spelling variants 171

C6.1 Synonyms proper 171
C6.2 Spelling variants 172

C6.2.1 Distinction between synonyms proper and spelling variants 172
C6.2.2 Number of spelling variants to be included in the thesaurus 173
C6.2.3 Where to store the spelling variants 174

C7 Summary of relationships displayed in a thesaurus 174

C7.1 Cross-references and inverse cross-references 180

Part II Presentation of indexing languages and thesauri 181

D Thesaurus format 183

DO Introduction 183
D1 The different parts of a thesaurus 183
D1.0 Introduction 183
D1.1 Thesaurus format: the Roget-Soergel model 184
D1.0.1 Rationale 184
1. A descriptor should always be seen in its place in the overall structure before it is used in indexing and searching 192
2. Not too much information should be given in the alphabetical index 193
3. Roget-Soergel model appropriate for systems using notation 193
4. Parts of the thesaurus 193
D1.1.1 Classified index (the schedule) 193
0. Summary (overview, synopsis) of the main subject fields 193
1. Display of the checklist descriptors and the important relationships among them 193
2. Classified index 194
3. Note on 1 and 2 194
D1.1.2 Main part of the thesaurus 194
D1.1.3 Alphabetical index 195
D1.2 Thesaurus format: the TEST model 196
D1.2.0 Parts of TEST 196
D1.2.1 Classified listings in TEST 196
1. Subject category index 196
2. Hierarchical index 196
D1.2.2 Thesaurus of terms (main part) 197
D1.2.3 Alphabetical index 197
D1.3 The look-up problem and how to arrange the entries in the main part 198
D1.3.1 Where the look-up problem occurs 198
D1.3.2 Finding the appropriate descriptor for a term that comes to mind 198
D1.3.3 Variations of the TEST model 202
1. Inclusion of spelling variants in the main part 202
2. Always look in the alphabetical index first 202
D1.3.4 Necessity of notation 202
D1.4 What Broader and Narrower Terms should be listed in the fields BT or NT, respectively? 202
## Contents xxv

D1.4.1 Inverse cross-references to broad descriptors of general application (advanced) 205
D1.4.2 Listing coordinate terms (advanced) 207
D1.5 Alphabetical index 207
D1.5.1 General considerations 207
D1.5.2 KWIC or KWOC format 209
D1.6 Guidance devices to facilitate look-up 209
D1.7 Description of selected thesauri 209
D1.7.1 Thesaurus of the Vision Information Center, Harvard Medical School 211
D1.7.2 FR Thesaurus (problems of developing countries) 211
D1.7.3 UDC, DDC, LCC 211
D1.7.4 Thesaurusfacet 212
D1.7.5 Medical Subject Headings (MeSH) 216
D1.7.6 Euratom-Thesaurus 220
D1.7.7 Thesaurus of education terms 221
D1.7.8 American Petroleum Institute (API) Subject Authority List 221
D1.7.9 ERIC thesaurus 223
D1.7.10 Library of Congress Subject Headings (LCSH) 223
D1.7.11 Other thesauri 225
D1.8 Introduction to the thesaurus 225
D2 Format of entries in the main part 228
D2.1 Information given for each term 228
D2.2 Rationale for the sequence of data fields (cross-reference types) (advanced) 228
D2.3 Arrangement of terms within one data field (technical) 232
D2.3.0 General 232
D2.3.1 Synonyms and equivalent terms (quasi-synonyms) 232
D2.3.2 BT, NT, RT 233
   , 1 What Broader Terms and Narrower Terms to list 233
   , 2 Broader Terms 233
   , 2.1 Broader Terms and Semantic Factors 233
   , 2.2 USE instructions containing Broader Terms 233
   , 2.3 Upward hierarchical chains 233
   , 3 Narrower Terms 234
   , 4 Display of different kinds of hierarchical relationships 234
   , 5 Related Terms 234
   , 6 Arrangement by notation 234
   , 7 How descriptors are entered in the data fields BT, NT, RT 234
Contents

D3 How to display descriptors and their inter-relationships (methods for the design of a classified index) 235
  D3.0 Relational displays vs. classification principles 236
    D3.0.1 Alternate classified index (advanced) 237
  D3.1 Displays for hierarchical relationships 237
    D3.1.1 Linear arrangement of descriptors (and possibly other preferred terms) in classified order with cross-references 237
      ,1 Preferred monohierarchical structure and cross-references 237
      ,2 Sequence of descriptors on the same level 239
        ,2.1 How to achieve helpful arrangement (technical) 239
      ,3 Details of presentation (technical) 241
    D3.1.2 Graphical display of hierarchical relationships 243
      ,1 Usual tree display 243
      ,2 Tree display with horizontal arrangement of hierarchical levels 246
      ,3 Circular display of hierarchical relationships 246
  D3.2 Network structures for the combined display of hierarchical and associative relationships 249
  D3.3 Comparison of different methods 255
  D3.4 Use of different type fonts (technical) 255
  D3.5 Methods for compressing the display of checklist descriptors 263
  D3.6 Auxiliary ISAR systems (“conceptual indexes”) II 263
    D3.6.0 Introduction and rationale 263
    D3.6.1 Implementation of ISAR systems for descriptors (auxiliary ISAR systems) 265
      ,1 Mechanized auxiliary ISAR systems 265
      ,2 Combinatorial indexes 265
  D3.7 On-line display of thesauri (special topic) 272

D4 Notation 273
  D4.0 Definition 273
  D4.1 Purpose of notation 275
    D4.1.1 Changes in notations (advanced) 275
    D4.1.2 Notation and machine-internal code in computerized ISAR systems (special topic) 276
  D4.2 The fallacy of overstressing notation 277
D4.3 Design of a notation (technical, especially Sections D4.3.3-D4.3.6) 277
  D4.3.1 Design criteria 278
  D4.3.2 Types of notation 278
    ,1 Expressive notation 278
    ,2 Purely ordinal notation 280
      ,2,1 A special device for intercalating new serial numbers 282
      ,3 Example 282
  D4.3.3 A partly expressive, partly ordinal system of notation (mixed notation) 282
    ,1 Mixed notation 1: small indexing languages/classification schemes (less than 1,000 descriptors) 282
    ,2 Mixed notation 2: large indexing languages/classification schemes 283
    ,3 Notations for compound concepts 285
  D4.3.4 An easy-to-produce expressive notation 286
  D4.3.5 Notation for precombined descriptors 286
  D4.3.6 Notations for preferred terms that are not descriptors 288
D4.4 Specific problems in notation (technical) 289
  D4.4.1 “Incorporating” standard classification schemes 289
  D4.4.2 Descriptors with “data field” 291
    ,1 Numerical data field 291
    ,2 Data field proper name 292
  D4.4.3 The UDC method of handling time, modified 292
D5 Multilingual thesauri (special topic) 293
  D5.0 Definitions 293
  D5.1 Format of a type-1 multilingual thesaurus (lead-in only) 294
  D5.2 Format of a type-2 multilingual thesaurus (indexing language in different languages) 294
    D5.2.1 Separate editions for each language (recommended) 294
    D5.2.2 One all-language edition (not recommended) 295
  D5.3 Production of a type-2 multilingual thesaurus 295
  D5.4 Production of an English thesaurus that contains translations in other languages 296
  D5.5 Interlingual thesauri 296

E Rules concerning the form of terms and related problems 298
  E0 Introduction. Difference in requirements between systems using notations and systems using terms 298
xxviii Contents

El Rules for the form of terms 299

El.0 Preliminary remarks 299
   El.0.1 Selection of rules 299
   El.0.2 When to apply the rules for the form of terms in the process
       of thesaurus building 299
   El.0.3 The application of the rules in the alphabetical index 300
   El.0.4 Preview 300

El. 1 Formulating terms more precisely 300
   El. 1.1 Disambiguation of homonyms through parenthetical
       qualifiers 301
   El. 1.2 Homonymous multiword or composite terms 301
   El.1.3 Omission of parenthetical qualifiers in classified listings
       (technical) 302
   El. 1.4 Artificial homonyms 303

El. 2 Rules on what parts of speech (nouns, adjectives, verbs) are
   allowed (technical) 303
   El.2.1 Permit-all rule 303
   El.2.2 Prefer-nouns rule 304
   El.2.3 Grammatical form to be used for each part of speech 304

El. 3 Designation of actions and processes, on the one hand, and of
   their results on the other (technical) 305
   El.3.0 The problem 305
   El.3.1 Rules 305
       ,1 Verb-noun rule 305
       ,2 “-ing”- “-ation” rule 305
       ,3 Explicit disambiguation 306
       ,4 Recommended rules 306

El.4 Singular vs. plural (technical) 306
   El.4.0 When rules are necessary 306
   El.4.1 Terms that are used in singular or in plural only 307
       ,1 Terms that are used in singular only 307
       ,2 Terms that are used in plural only (pluralia
           tantum) 307
   El.4.2 Simple rules 308
   El.4.3 More complicated rules: rules used in the TEST
       thesaurus 308

El.5 Sequence of words in multiword or composite terms
   (technical) 308
   El.5.1 Direct entry (TEST) 309
   El.5.2 Inverted entry 312
Contents xxix

El.6 Terms formed as strings of terms, interpreted as OR combination 313
El .7 Symbols, especially numerals, as components of terms 313
El.8 Acronyms and abbreviations 314
   El.8.1 Commonly used acronyms 314
   El.8.2 Use of abbreviations to save space 315
   El.8.3 Standardized abbreviations for descriptors 315
El.9 Term length 316
El.10 Terms in foreign languages 316
El.11 Proper names and trademarks 316

E2 Spelling and transliteration (technical) 317
   E2.1 Authorities 317
   E2.2 Punctuation 317
   E2.3 Capitalization 318
   E2.4 Character set available 319
   E2.5 Transliteration 319

E3 Alphabetization (technical) 320

Part III Procedures for the construction and maintenance of indexing languages and thesauri 323

F Flow of work in the construction of indexing languages and thesauri 325
   F0 Overview and general problems 325
      F0.1 The major steps 325
      F0.2 Cooperative thesaurus development 326
      F0.3 Collaboration of experts from different subject areas 326
         F0.3.0 Necessity of full-time staff and collaboration of subject experts 326
         F0.3.1 Supply of material 334
         F0.3.2 Answering questions on single problems that come up during the work on the thesaurus 334
         F0.3.3 Discussion sessions for review and/or decisions on difficult problems 335
         F0.3.4 Inter-disciplinary approach 336
         F0.3.5 Briefing of subject experts on thesaurus functions 336
         F0.3.6 Source codes for subject experts and panels 336
      F0.4 Criteria for the selection of terms and descriptors 336
         F0.4.1 Criteria for the selection of terms (whether nonpreferred lead-in terms, preferred lead-in terms, or descriptors) to be included in the thesaurus 337
F3 Work out the preliminary structure of the thesaurus: the synonym-homonym structure, the equivalence structure, and the classificatory structure. Select preferred terms
   F3.1 Define broad subject fields and sort terms into these broad fields 385
   F3.2 Define subfields within each subject field and sort terms accordingly 386
   F3.3 Work out detailed thesaurus structure. Select preferred terms.
      Merge information for terms in the same concept class 386
      F3.3.1 Work out the synonym-homonym structure and the equivalence structure 388
      F3.3.2 Work out the classificatory structure 389
      F3.3.3 Use of judgment and creative thinking in processing the information collected from different sources 390
      F3.3.4 Introducing more specific concepts 391
      F3.3.5 Scope notes and definitions 391
      F3.3.6 Preliminary selection of descriptors from among the preferred terms 391
      F3.3.7 Some suggestions for the technique to be used (technical) 392

F4 Work out first draft of the classified index (schedule)
   F4.0 Classified index and cross-references in BT, NT, and RT 392
   F4.1 Type preliminary classified index. Amend working file 393
   F4.2 Improve the classificatory structure 394
   F4.3 Type improved classified index and amend working file 395
   F4.4 Discuss classified index with subject experts. Select descriptors and checklist descriptors 395
   F4.5 Assign notational symbols 397
   F4.6 Make a systematic search for additional cross-references 397

F5 Complete first draft of the thesaurus as a whole
   F5.0 Introduction 397
      F5.0.1 Special problems of smaller projects not using computer assistance (special topic) 398
   F5.1 Revise entries in the working file 398
   F5.2 Produce the main part of the thesaurus in list form 402
   F5.3 Check inverse cross-references and insert where necessary 402
   F5.4 Duplicate preliminary version of the thesaurus 403
   F5.5 Review the whole thesaurus. Consult with subject experts 403
   F5.6 Enter modifications in the master copy 404
Contents xxxiii

F5.7 Production of the alphabetical index (technical) 404
  F5.7.1 Production of a KWIC index 404
  F5.7.2 Manual production of the alphabetical index 406
  F5.7.3 TEST model: produce alphabetical main part and alphabetical index 406
  F5.7.4 Remark 408

F5.8 Check homonyms and improve cross-reference structure using the alphabetical index 408

F5.9 Reproduce test version of the thesaurus 409

F5.10 Remarks on some technical problems arising in F5, F6, and F7 (technical) 409
  F5.10.1 Use of notations as “shorthand” for descriptors 409
  F5.10.2 Technical considerations as to the production of the main part of the thesaurus in smaller projects without computer assistance 410

F6 Test the thesaurus by indexing and retrieval experiments 411

F7 Duplicate or print the user version of the thesaurus 412
  F7.1 Duplication or printing of main part and the alphabetical index 412
  F7.2 Duplication or printing of the classified index 412
  F7.3 Proofreading 413

F8 Further remarks concerning the work-flow and modifications of the standard work-flow 413
  F8.0 Introduction 413
  F8.1 Sequence of the Steps F3, “Work out the preliminary structure of the thesaurus” and F4, “Work out the first draft of the classified index” 413
  F8.2 When should the notation be introduced? 414
  F8.3 When should the main part be typed (smaller projects without computer assistance)? 415
  F8.4 Drawing up and using a “core classification” consisting of elemental concepts early in the process 416
  F8.5 Extending the collection of conceptual relationships, especially for cooperative information services 417

F9 Use of punched paper tape and punched cards in thesaurus construction (special topic, in part technical) 417
  F9.1 Use of punched-paper-tape typewriters in thesaurus construction 417
    F9.1.1 Modifications in the flow of work 417
    F9.1.2 Conversion of punched paper tape to punched cards 418
  F9.2 Use of conventional punched card equipment 419
F9.2.1 Punched-card-controlled typewriters (for example, the IBM 870 Document Writing System) 419
F9.2.2 Keypunch and unit-record equipment 419

G Use of computers in thesaurus construction (advanced; technical with the exception of Sections G0.1 and G0.2) 420

GO Rationale for computer application. Overview 420

GO. 1 Rationale for computer application 420

GO. 1.1 Performing routine operations 420
GO. 1.2 Continuous modification of data base 421

G0.2 Overview. “Entry points” for computer processing. Modifications in work flow 421

G0.3 Record organization in the computer 424

G0.3.1 Complete summary of the organization of cross-reference subrecords 425

G1 Computer assistance in the collection and recording of material 428

G1.2.2 Recording the data from the sources in machine-readable form 428

G2 Computer assistance in sorting into alphabetical order and in merging information on identical terms into one record 429

G2.2.3 Computer assistance in “pulling” information from big thesauri by computer 432

G2.3 Second round of merging by computer 433
G2.4 Standardization of spelling variants by computer 433
G2.5 Miscellaneous problems 433

G2.5.1 Cross-references given using notations 433
G2.5.2 Record identification 434
G2.5.3 Substituting numbers for terms to save storage space 434

G3 Computer assistance in working out the preliminary structure of the thesaurus 434

G3.3 Computer assistance in clerical tasks to be performed in F3.3, “Work out the detailed structure of the thesaurus” 435

G3.3.1 Merging information for each class of synonyms 436
G3.3.2 Rearranging the working file in classified order 436

G3.4 Computer assistance for intellectual tasks in working out the detailed thesaurus structure 436

G3.4.1 Computer assistance in hierarchy construction 436
G3.4.2 Use of the decomposition of compound concepts into elemental concepts in working out the preliminary structure of the thesaurus 438
Use of the decomposition of compound concepts in sorting terms into subject fields and subfields, and in forming groups of synonyms (Steps F3.1, F3.2, and F3.3.1) 439

Use of the decomposition of compound concepts in working out the classificatory structure (Step F3.3.2) 439

Computer assistance in semantic factoring 440

Computer assistance in working out the classified index 441

Computer assistance in completing the first draft of the thesaurus as a whole 442

Computer assistance in revising entries in the working file 442

Check of inverse cross-references by computer 442

Printing the final thesaurus by computer 443

Updating a computer-stored thesaurus 443

Types of changes 443

Input of updating information 445

Line-oriented input of updating information 445

Term-oriented input of updating information 446

Comparison of the two methods 446

Devices for the input (keying) of thesaurus data 447

Automatic methods in the construction of indexing languages and thesauri, starting from the texts of documents and/or search requests*

Automatic classification (advanced) 449

Introduction 449

Definition of units of text and counting methods 450

Identification of descriptor candidates from frequency patterns 451

Detection of term or concept relationships from co-occurrence patterns 451

Nearness measures 451

Interpretation of high association between two terms A and B 452

Second-order associations for the detection of definitional relationships 453

The use of inconsistent association profiles for the detection of homonyms 453

Detection of hierarchical relationships 454

Combined application of different methods 454

Automatic derivation of classification schemes (“global” structures) 455

Automatic derivation of classification schemes by clustering methods 455
### Contents

**H4.2 Automatic derivation of classification schemes by graph-theoretical methods** 455

**J Updating and maintenance of indexing languages and thesauri** 457

**JO Introduction** 457

**XI Types of changes** 457

**J2 Sources for new terms, concepts, and relationships to be included in the thesaurus** 458

**J2.1 Sources within the ISAR system** 458

**J2.1.1 Search request statements, search request formulations, and search performance** 458

**J2.1.2 Documents and indexing of documents** 459

**J2.1.3 Collection of updating information from sources within the ISAR system** 459

**J2.2 Sources outside the ISAR system** 460

**J2.2.1 Information on changes in user needs** 460

**J2.2.2 Information on new developments in the subject fields of the ISAR system** 460

**J3 Procedures for regular updating** 460

**J3.1 Use of thesaurus forms in updating** 460

**J3.2 Processing of updating information** 461

**J3.3 Issuing supplements and/or revised versions** 461

**J3.3.1 Time schedule for updating** 461

**J3.3.2 Physical form of supplements** 462

**J3.3.3 Listing of changes made** 462

**J3.4 Organization for and decision-making in thesaurus updating** 462

**J3.5 “Interactive” updating of thesauri** 463

**J4 Revision of the indexing language or the thesaurus at longer intervals** 463

**J5 Remarks on the flexibility of structured indexing languages (classification schemes)** 464

**J6 Problems of re-indexing (re-classification)** 465

**J6.1 Re-indexing problems due to introduction of new descriptors** 465

**J6.2 Re-indexing problems due to changes in descriptor usage** 466

**J7 Thesaurus updating and thesaurus compatibility: common problems (advanced)** 467

**Part IV Thesauri as a basis for cooperation in information services** 469

**K Thesauri as a basis for cooperation in information services** 471

**K0 Introduction** 471

**K1 Cooperation in the construction of indexing languages and thesauri** 472
K1.1 Cooperation in material collection and merging only 472
K1.2 Cooperation in the development of the terminological and
classificatory structure 473
   K1.2.1 Cooperation between two (or a few) institutions 473
   K1.2.2 Generalized cooperation: the concept of a source
       thesaurus (advanced) 473
           , 1 The structure of a source thesaurus 475
           ,2 Extraction of indexing languages or thesauri for
               special applications from a source thesaurus 477
               ,2.1 Specific extraction of indexing languages from
                   a source thesaurus 477
                   ,2.1.1 Format for “extraction specifications”
                       (technical) 478
               ,2.2 General extraction of indexing languages from
                   a source thesaurus 479
               ,2.3 Use of a source thesaurus in the revision of
                   existing indexing languages and thesauri 479
   K1.2.3 Adjunct thesauri 481
K1.3 The concept of a cumulative thesaurus (advanced) 485
   K1.3.0 Definition and use 485
   K1.3.1 Record organization for a cumulative thesaurus
       (technical) 486
       ,1 Treatment of the recommended structure and of
           source indications in a cumulative thesaurus 486
       ,1.1 Treatment of the recommended structure 486
       ,1.2 Group of data fields F-L 486
       ,1.3 Use of the data fields Cl, K, and L for increasing
           the precision of source indications 487
       ,1.4 Further refinements 489
       ,2 Keeping track of decisions and dates in a cumulative
           thesaurus 489
   K1.3.2 Development of a cumulative thesaurus 489
   K1.3.3 Display of a cumulative thesaurus 490
K1.4 Incorporation of an additional thesaurus into the cumulative
    thesaurus and/or analysis and improvement of that thesaurus
    using a cumulative thesaurus (advanced and technical) 491
K2 Cooperation through sharing the results of subject indexing
    (special topic) 493
   K2.1 Introduction. Statement of the problem. Searching conversion
       versus indexing conversion 494
Contents

K2.1.1 Multilateral shared subject indexing using a “switching language” 501
K2.2 Framework for the comparison of two indexing languages or thesauri 502
  K2.2.1 Convertibility categories (advanced) 503
    ,1 Searching convertibility categories 503
    ,2 Indexing convertibility categories 504
K2.3 Production of conversion tables 510
  K2.3.1 Ideal situation: the indexing languages of the cooperating institutions are still to be built 511
    ,1 The development of a total thesaurus through parallel development of constituent thesauri 512
    ,2 Alphabetical index for the total thesaurus 512
  K2.3.2 Usual situation: each of the cooperating institutions already has its own indexing language long in use 513
    ,1 The local approach and the global approach to the construction of conversion tables 513
  K2.3.3 Updating of the individual indexing languages or thesauri 514
K2.4 Compatibility on a general level. The concept of an umbrella classification 514
  K2.4.1 Shared subject indexing on a general level 514
  K2.4.2 The concept of an umbrella classification 515
K3 The idea of a Universal Source Thesaurus (UST) (special topic) 516
  K3.0 Universal Source Thesaurus versus universal classification 516
  K3.1 The structure of UST 517
  K3.2 Neutrality of UST with regard to classification principles 518
  K3.3 UST as a framework for “semi-universal” indexing languages for shared subject indexing 519
  K3.4 Implementation of a Universal Source Thesaurus 519

Appendices

Appendix 1: Thesaurus guidelines and thesaurus books 523
Appendix 2: Bibliographies of subject access vocabularies and dictionaries* Specific subject access vocabularies and documents on specific subject access vocabularies included in the bibliography 529
<table>
<thead>
<tr>
<th>Contents xxxix</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter Notes</td>
<td>535</td>
</tr>
<tr>
<td>Bibliography</td>
<td>559</td>
</tr>
<tr>
<td>Index</td>
<td>609</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Examples of relationships displayed in a thesaurus (A1.2)</td>
<td>5</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Flowchart: Considerations for the construction of a thesaurus (A2)</td>
<td>11</td>
</tr>
<tr>
<td>Figure 3</td>
<td>The structure of an information system (BO)</td>
<td>18</td>
</tr>
<tr>
<td>Figure 4</td>
<td>The structure of an ISAR (Information Storage and Retrieval) system (BO)</td>
<td>19</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Examples of synonyms, quasi-synonyms, and homographs (B2.3)</td>
<td>25</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Summary of definitions and further illustrations (B4.2)</td>
<td>32</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Examples of semantic factoring (Cl.1.0)</td>
<td>75</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Some questions that might be useful for semantic factoring (Cl.1.0)</td>
<td>76</td>
</tr>
<tr>
<td>Figure 9a</td>
<td>Example of (poly-) hierarchical relationships in tree-like representation (Cl.2)</td>
<td>82</td>
</tr>
<tr>
<td>Figure 9b</td>
<td>Representation of the hierarchical structure of Figure 9a in linear arrangement with cross-references (Cl.2)</td>
<td>82</td>
</tr>
<tr>
<td>Figure 10a</td>
<td>Hierarchical structure generated by two facets, no within-facet combinations, no hierarchy within facets (Cl.3)</td>
<td>86</td>
</tr>
<tr>
<td>Figure 10b</td>
<td>Different possible linear arrangements of the concepts given in Figure 10a (Cl.3, D3.1.1)</td>
<td>87</td>
</tr>
<tr>
<td>Figure 10c</td>
<td>Same as Figure 10a, but different semantic content (Cl.3)</td>
<td>88</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Hierarchical structure generated by two facets, no within-facet combinations, hierarchy within facets (Cl.3)</td>
<td>89</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Hierarchical structure generated by three generating concepts without hierarchical relationships among generating concepts (Cl.3)</td>
<td>92</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Hierarchical structure generated by five generating concepts with hierarchical relationship among generating concepts (Cl.3)</td>
<td>93</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Examples of different kinds of hierarchical relationships (Cl.4.2)</td>
<td>100</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Transitions between the synonym-homonym structure, the equivalence structure, and the classificatory structure (Cl.6)</td>
<td>111</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Document representations in different file systems (C2.3.1)</td>
<td>117</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Typology of international organizations (C3.1)</td>
<td>144</td>
</tr>
<tr>
<td>Figure 18</td>
<td>Example illustrating the detailed lead-in form (C5.1)</td>
<td>163</td>
</tr>
<tr>
<td>Figure 19</td>
<td>Example illustrating the detailed lead-in form (C5.1)</td>
<td>164</td>
</tr>
</tbody>
</table>
xli! List of Figures

Figure 20. Treatment of spelling variants (C6.2) 173
Figure 21. Types of cross-references and other data elements given in the entry for a term (C1) 176
Figure 22. Example worked out according to the Roget-Soergel model and the TEST model (DO) 185
Figure 23. Look-up in the Roget-Soergel model (D1.3.2) 199
Figure 24. Look-up in the TEST-model (D1.3.2) 200
Figure 25. Alphabetical index in KWIC and KWOC format (D1.5.2) 210
Figure 26. Thesaurofacet: sample page of the classified index (D1.7.4) 214
Figure 27. Thesaurofacet: sample page of the main part (D1.7.4) 215
Figure 28a. Medical Subject Headings: sample page of the classified index (subject category listing) (D 1.7.5) 216
Figure 28b. Medical Subject Headings: sample page of the classified index (tree structures) (D 1.7.5) 217
Figure 29. Medical Subject Headings: sample page of the main part (D1.7.5) 219
Figure 30. Arrangement of types of cross-references and other data elements within a record or entry in the user version of the main part of the thesaurus (D2.1 and D2.2) 229
Figure 31. Sample main party entry: Roget Soergel model (D2.1 and D2.2) 229
Figure 32. Sample main part entry: TEST (D2.2) 231
Figure 33. Sample main part entry: BASF (D2.2) 232
Figure 34. Alphabetical versus meaningful sequence of descriptors on the same level (D3.1.1,2) 240
Figure 35. Segment of the classified index of the FR thesaurus (D3.1.1,3) 242
Figure 36. Display of a large classified index with summaries (D3.1.1,3) 244
Figure 37. Simple tree display (following SYNTOL) (D3.1.2,1) 246
Figure 38. Tree display using space-saving devices. Also example of how to show a part of a big tree (following SYNTOL) (D3.1.2,1) 247
Figure 39. Tree display using space-saving devices (D3.1.2,1) 248
Figure 40. Tree display with horizontal arrangement of hierarchical levels (D3.1.2,2) 250
Figure 41. Circular display: hierarchical levels arranged in concentric circles (following TDCK) (D3.1.2,3) 256
Figure 42. Network display of conceptual relationships following EURATOM 1: Group 15 Anatomy (EURATOM 2, see Figure 46) (D3.2) 258
Figure 43. Network display within a coordinate grid (D3.2) 259
Figure 44. The descriptors and their relationships from Figure 43 displayed in a linear sequence with indentation (D3.2) 260
Figure 45. Network display based on a sequence of processes and their results (D3.2) 261
List of Figures xliii

Figure 46. Network display following EURATOM 2: Group 05 Blood system (EURATOM I, see Figure 42) (D3.2) 262
Figure 47* Illustration of a combinatorial index to both LC Subject Headings and LC Classification (D3.6.1,2) 266
Figure 48. Example showing two types of notation (D4.0; also used for D 4.3) 274
Figure 49. An easy-to-produce expressive notation (D4.3.4) 286
Figure 50. “Relative” alphabetical index to DDC (El. 1.4) 304
Figure 51. Guidelines to singular—plural usage (El.4.3) 310
Figure 52. Flow of work in thesaurus construction: overview flowchart (F0.1) 327
Figure 53. Flow of work in thesaurus construction: detailed flowchart (F0.1) 328
Figure 54. Thesaurus form (F0.5) 346
Figure 55. Example of filled-in thesaurus form (F1.2.2) 363
Figure 56. Merging of data elements from different cards for the same term (F2.2) 369
Figure 57. Further examples to illustrate merging in the first round (F2.2) 370
Figure 58. Example of result of merging in the first round on a thesaurus form (F2.2) 372
Figure 59. Sample file for the second round of merging (F2.3.1) 376
Figure 60. Flowchart for the second round of merging (identifying classes of synonyms) (F2.3.1) 378
Figure 61. Examples illustrating the second round of merging (F2.3.1) 382
Figure 62. “Road map” for the analysis of terms (F3.2) 387
Figure 63. Example of revisions in the working file (F5.1) 399
Figure 64. Construction of a hierarchy by “chaining” hierarchical cross-references (G3.4.1) 437
Figure 65. Example of second-order association (H3.2) 453
Figure 66. Sample guidance classification scheme of a source thesaurus (K1.2,2.2.1.1) 480
Figure 67. Simple extraction specification and resulting classified index (Kl.2.2, 2.1.1) 481
Figure 68. More elaborate extraction specification and resulting classified index (Kl.2.2, 2.1.1) 482
Figure 69. Searching conversion and indexing conversion (K2.1) 498
Figure 70. Two sample indexing languages for the illustration of convertibility categories (K2.2.1) 505
Figure 71. Searching convertibility categories (conversion from A to B) (K2.2.1.1) 506
Figure 72. Indexing convertibility categories (conversion from B to A) (K2.1.2) 507