I

Indexing Languages and Thesauri: Construction and Maintenance

DAGOBERT SOERGEL
College of Library and Information Services
University of Maryland

A WILEY-BECKER & HAYES SERIES BOOK

-jgj| MELVILLE PUBLISHING COMPANY
Los Angeles, California

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 THESAURI: 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

r-j» MELVILLE PUBLISHING COMPANY
1SI5) Los Angeles, California

rs Copyright © 1974, by John Wiley & Sons, Inc*
I Published by Melville Publishing Company,
S' a Division of John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

No part of this book may be reproduced by any means,
nor transmitted, nor translated into a machine language
without the written permission of the publisher.

Library of Congress Cataloging in Publication Data:

Soergel, Dagobert.
Indexing languages and thesauri.

(Information sciences series)
"A Wiley-Becker & Hayes series book."
Bibliography: p.
1. Indexing. 2. Thesauri. I. Title.

Z695.9.S63 029.5 73-20301
ISBN 0-471-81047-9

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

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-

mann, H. Krieg, E. Lutterbeck, E. Meyer, C. v. Rothkirch-Trach, M. Scheele, H. Schneider, M. Simon, D. Soergel). The German Society for Documentation, which published the book in 1969, gave permission to use the material without restriction.

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 con* struction 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.

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

2. The problems dealt with in this book are highly interrelated. A second reading might be helpful.
3. 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.
4. 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.
5. 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 **1**

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 **15**

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 **181**

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 **323**

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 **469**

K Thesauri as a basis for cooperation in information services 471

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

AO 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

BO 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

CO Introduction 68
Cl Classificatory structure 69
C2 Problems of file organization related to classification. Practical applications of semantic factoring 112
C3 Concept formation in thesaurus building. Definition and scope notes 142
C4 Types of concepts, descriptors, terms to be included in an indexing language or thesaurus 147
C5 The lead-in structure. USE and SEE 155
C6 Synonyms proper versus spelling variants 171
C7 Summary of relationships displayed in a thesaurus 174

Part H Presentation of indexing languages and thesauri 181

D Thesaurus format 183
DO Introduction 183
D1 The different parts of a thesaurus 183
D2 Format of entries in the main part 228
D3 How to display descriptors and their interrelationships (methods for the design of a classified index) 235
D4 Notation 273
D5 Multilingual thesauri 293

E Rules concerning the form of terms and related problems 298
EO Introduction. Difference in requirements between systems using notations and systems using terms 298
E2 Spelling and transliteration 317
E3 Alphabetization 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
FI Collect and record material (concepts, terms, relationships between and among them) 355
F2 Sort into alphabetical order and merge information on identical terms on one card 366

F3 Work out the preliminary structure of the thesaurus: the synonym-homonym structure, the equivalence structure, and the classificatory structure. Select preferred terms 384
F4 Work out first draft of the classified index (schedule) 392
F5 Complete first draft of the thesaurus as a whole 397
F6 Test the thesaurus by indexing and retrieval experiments 411
F7 Duplicate or print the user version of the thesaurus 412
F8 Further remarks concerning the work-flow and modifications of the standard work-flow 413
F9 Use of punched paper tape and punched cards in thesaurus construction 417

Use of computers in thesaurus construction 420
GO Rationale for computer application. Overview 420
G1 Computer assistance in the collection and recording of material 428
02 Computer assistance in sorting into alphabetical order and in merging information on identical terms into one record 429
G3 Computer assistance in working out the preliminary structure of the thesaursus 434
G4 Computer assistance in working out the classified index 441
G5 Computer assistance in completing the first draft of the thesaurus as a whole 442
G7 Printing the final thesaurus by computer 443
G8 Updating a computer-stored thesaurus 443
G9 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 449
HO Introduction 449
HI Definition of units of text and counting methods 450
H2 Identification of descriptor candidates from frequency patterns 451
H3 Detection of terms or concept relationships from co-occurrence patterns 451
H4 Automatic derivation of classification schemes (“global” structures) 455

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

K Thesauri as a basis for cooperation in information services 471
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.
Major tasks to be performed in and resources and work required for the construction of a thesaurus 3

AO 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

A 1.3.1 Parameters determining thesaurus size 6

Al.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

A 1.7 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

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 15

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

BO Introduction 17

B1 Plane of concepts versus plane of terms: the synonym-homonym structure 17

Bl.l Homonyms and homographs (advanced) 20

B2 Treatment of nearly related concepts: the equivalence structure 22

B2.Ö Equivalent concepts (equivalent terms) 22
B2.1 Forming ISAR concepts and naming them. Preferred terms 23
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

B5.0.1 Elementary solution: search whole (un-indexed) collection 40
B5.0.2 First economy measure: batch search requests 40
B5.0.3 Second economy measure: prepare abstracts 41
B5.0.4 Third economy measure: anticipated search requests: collect anticipated search requests and analyze documents in advance 41
B5.0.5 Fourth economy measure: provide a retrieval mechanism 42
B5.0.6 Concluding remarks 43
B5.1 A preliminary view of the structure of indexing languages (classification schemes) 44
B5.2 Request-oriented indexing and the checklist technique 45
B5.2.0 Introduction 45
B5.2.1 Disadvantages of the method of indexing commonly used (document-oriented indexing) 45
B5.2.2 Implementation of request-oriented indexing; the checklist technique described 46
B5.2.3 Summary: request-oriented indexing and the checklist technique 49
B5.2.4 The checklist technique for search request formulation 49
B5.2.5 Request-oriented indexing and cost-benefit considerations 50
B5.3 Adding descriptors through supplementary document-oriented indexing 50
B5.4 Representation versus filing of documents or catalog cards 51
B5.4.1 Use of relatively broad descriptors in filing arrangement 51
B5.4.2 Use of relatively broad descriptors for peek-a-boo cards or other indexes 55
B5.4.3 Use of additional descriptors beyond those used for filing or manual indexes 56
B5.4.4 The use of free (open*ended) terms supplementing the descriptors from the indexing language 57
B5.4.5 Note on “descriptor” and multi-purpose systems 58
B5.5 Summary of the functions of hierarchy and classified arrangement 59
B5.6 Indicative versus informative representation of documents 60
B5.7 Cost-benefit consideration for the design of an indexing language 61
B6 The functions of the lead-in vocabulary in an ISAR system 61
B6.1 Advantages of a lead-in vocabulary for human indexing and search request formulation 62
B6.1.1 Alphabetical index to the indexing language is more effective 62

B6.1.2 Thesaurus as store of intellectual decisions made in day-to-day indexing and search request formulation 62

,1 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

Cl Classificatory structure 69

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

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

Cl. 1.0 Foundations of semantic factoring and concept combination 74

Cl. 1.1 Advantages of semantic factoring 77

Cl.2 Polyhierarchical structure. Definition of hierarchy 78

Cl .3 Interaction of hierarchy and concept combination 83

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

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

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

Cl .4.1 Further considerations on pragmatic hierarchy building 95

,1 Extending the definition of concepts or introducing new broader concepts 95

,2 Introduction of additional broader concepts for searching 95

,3 Introduction of a broader concept to replace a number of specific concepts 97

,4 Introduction of new broader concepts to serve as headings ("organizational headings") 97

,5 Antonyms 97

,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
C1.4.3 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 .6 Transitions between the synonym-homonym structure, the equivalence structure, and the classificatory structure 110
Cl .7 Psychological dimensions of relationships 112

C2 Problems of file organization related to classification. Practical applications of semantic factoring 112

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

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 organization: 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 13 7

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*l 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

,1 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
Dl.l Thesaurus format: the Roget-Soergel model 184
Dl. 1.0 Rationale 184
,1 Classified index required for the checklist technique of indexing and search request formulation 184
,2 A descriptor should always be seen in its place in the overall structure before it is used in indexing and searching 192
,3 Not too much information should be given in the alphabetical index 193
,4 Roget-Soergel model appropriate for systems using notation 193
,5 Parts of the thesaurus 193
Dl.l. 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
Dl.l.3 Alphabetical index 195
D1.2 Thesaurus format: the TEST model 196
Dl .2.0 Parts of TEST 196
D 1.2.1 Classified listings in TEST 196
,1 Subject category index 196
,2 Hierarchical index 196
Dl.2.2 Thesaurus of terms (main part) 197
D 1.2.3 Alphabetical index 197
Dl .3 The look-up problem and how to arrange the entries in the main part 198
D 1.3.1 Where the look-up problem occurs 198
Dl.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
Dl.3.4 Necessity of notation 202
D1.4 What Broader and Narrower Terms should be listed in the fields BT or NT, respectively? 202

Dl .4.1 Inverse cross-references to broad descriptors of general application (advanced) 205
D 1.4.2 Listing coordinate terms (advanced) 207

D1.5 Alphabetical index 207

Dl.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
Dl.7.2 FR Thesaurus (problems of developing countries) 211
Dl.7.3 UDC, DDC, LCC 211
D1.7.4 Thesaurofacet 212
Dl.7.5 Medical Subject Headings (MeSH) 216
D 1.7.6 Euratom-Thesaurus 220
D 1.7.7 Thesaurus of education terms 221
Dl.7.8 American Petroleum Institute (API) Subject Authority List 221
Dl .7.9 ERIC thesaurus 223
D1.7.10 Library of Congress Subject Headings (LCSH) 223
Dl.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

D2A Typographical design of entries 235

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

El Rules for the form of terms 299

E1.0 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

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
FO.3.3 Discussion sessions for review and/or decisions on difficult problems 335
FO.3.4 Inter-disciplinary approach 336
FO.3.5 Briefing of subject experts on thesaurus functions 336
FO.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

FO.4.2 Criteria for the selection of a preferred term from a class of synonyms and quasi-synonyms (arranged according to decreasing priority) 337
FO.4.3 Criteria for the selection of descriptors 338
FO.4.4 The use of frequency data in the selection of descriptors (technical) 340

,0 Introduction 340
,1 Gathering of frequency and co-occurrence data 341
,2 Use of frequency data in descriptor selection 343

FO.4.5 Central area versus peripheral areas 345

F0.5 Use of a thesaurus form and related problems 345

F0.5.1 Instructions on how to use the thesaurus form (technical) 345
FO.5.2 Reasons for having an index card for each term 347
F0.5.3 Reasons for having a form rather than blank cards 347
FO.5.4 Size 348
FO.5.5 Width of lines 348
FO.5.6 Sequence of data fields 348

F0.6 Working file and user version 348
F0.7 Source indications for data elements entered in the thesaurus 349

FO.7,1 Why source indications? 349

,1 Use of the source indications for the elaboration of the thesaurus 349
,2 Why source indications in the user version of the thesaurus? 349

FO.7.2 Keeping track of the sources in the working file (technical) 350
F0.7.3 Experts and lexicographers as sources (technical) 352
FO.7.4 Keeping track of deletions (technical) 352

F0.8 Keeping track of decisions and dates 352

F0.8.1 Keeping track of decisions and dates in working file (technical) 353

,1 Keeping track of decisions made 353
,2 Keeping track of decisions still to be made 354
,3 Keeping track of why decisions have been made 354

FO.8.2 Giving dates in the user version of the thesaurus (technical) 354

FI Collect and record material (concepts, terms, relationships between and among them) 355

FI. 1 Kinds of sources. Criteria for selection of sources 355

Fl. 1.1 Sources in which terms are already arranged according to some principle (prearranged sources) 355
FI. 1.2 Sources in which terms are not ordered or from which terms must first be derived (open-ended sources) 355
FI. 1.3 Selection of the sources to be used 356
FI.1.4 Term-association lists (special topic) 358

FI.2 Technical procedures for the recording of terms, etc. 358

FI.2.0 Introduction 358
FI.2.1 Preparation of sources (technical) 359

,0 Source identification codes 359
,1 Preparation of prearranged sources 359

,1.1 Adding an auxiliary notation 360

,2 Preparation of open-ended sources: mark terms to be transferred 360
,3 Pre-processing of open-ended sources 362

FI.2.2 Transfer to terms to cards (thesaurus forms) (technical) 362

, 1 Entering Synonymous, Broader, Narrower, and Related Terms 362
,2 Entering the source indication 365
,3 Transfer of terms and other information with manual procedures 365

FI.2.3 An alternative procedure 366

F2 Sort into alphabetical order and merge information on identical terms on one card 366

F2.1 Sort into alphabetical order. Rules for preliminary alphabetical sorting 366
F2.2 First round of merging: merge information for identical terms 367

F2.2.1 Procedure for merging cards and keeping track of sources (technical) 367
F2.2.2 Steps after the first round of merging 368
F2.2.3 “Pulling” information from additional sources (match and merge) 374

,1 Procedure for “pulling” (technical) 374

F2.3 Second round of merging: merge information for terms in the same concept class (advanced and technical) 376

F2.3.1 The procedure (algorithm) 376
F2.3.2 Treatment of terms that consist of a string of Synonymous Terms 380
F2.3.3 Editing during or prior to the second round of merging 380

F2.3.4 Concluding remark 381

F2.4 Remarks regarding both rounds of merging 381

F2.4.1 Spelling and morphological variants 381
F2.4.2 Homonyms 384

F3 Work out the preliminary structure of the thesaurus: the synonym-homonym structure, the equivalence structure, and the classificatory structure. Select preferred terms 384

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) 392

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 397

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

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

,1 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
,2 Use of the decomposition of compound concepts in working out the classificatory structure (Step F3.3.2) 439

G3.4.3 Computer assistance in semantic factoring 440

G4 Computer assistance in working out the classified index 441
05 Computer assistance in completing the first draft of the thesaurus as a whole 442

G5.1 Computer assistance in revising entries in the working file 442
G5.3 Check of inverse cross-references by computer 442

G7 Printing the final thesaurus by computer 443
G8 Updating a computer-stored thesaurus 443

G8.1 Types of changes 443
G8.2 Input of updating information 445

G8.2.1 Line-oriented input of updating information 445
G8.2.2 Term-oriented input of updating information 446
G8.2.3 Comparison of the two methods 446

G9 Devices for the input (keying) of thesaurus data 447

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

HO Introduction 449
HI Definition of units of text and counting methods 450
H2 Identification of descriptor candidates from frequency patterns 451
H3 Detection of term or concept relationships from co-occurrence patterns 451

H3.0 Nearness measures 451
H3.1 Interpretation of high association between two terms A and B 452
H3.2 Second-order associations for the detection of definitional relationships 453
H3.3 The use of inconsistent association profiles for the detection of homonyms 453
H3.4 Detection of hierarchical relationships 454
H3.5 Combined application of different methods 454

H4 Automatic derivation of classification schemes ("global" structures) 455

H4.1 Automatic derivation of classification schemes by clustering methods 455

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

Kl. 1 Cooperation in material collection and merging only 472
Kl.2 Cooperation in the development of the terminological and classificatory structure 473

Kl.2.1 Cooperation between two (or a few) institutions 473
Kl.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

Kl.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
Kl.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

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 **521**

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**

Chapter Notes 535

Bibliography 559

Index 609

List of Figures

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

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* (Dl.3.2) 199
Figure 24. *Look-up in the TEST-model* (Dl.3.2) 200
Figure 25. *Alphabetical index in KWIC and KWOC format* (Dl.5.2) 210
Figure 26. *Thesaurofacet: sample page of the classified index* (Dl.7.4) 214
Figure 27. *Thesaurofacet: sample page of the main part* (Dl.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* (Dl.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,l) 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 indention* (D3.2) 260
Figure 45. *Network display based on a sequence of processes and their results* (D3.2) 261

Figure 46. *Network display following EURATOM 2: Group 05 Blood system (EURATOM l, 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.Ö;** *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* **(FI.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* **(Kl.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,l)** 506
Figure 72. *Indexing convertibility categories (conversion from B to A)* **(K2 2.1,2)** 507