

Essential thesaurus construction

Vanda Broughton



facet publishing

© Vanda Broughton 2006

Published by
Facet Publishing
7 Ridgmount Street
London WC1E 7AE

Facet Publishing is wholly owned by CILIP: the Chartered Institute of Library and Information Professionals.

Vanda Broughton has asserted her right under the Copyright, Designs and Patents Act, 1988 to be identified as author of this work.

Except as otherwise permitted under the Copyright, Designs and Patents Act, 1988 this publication may only be reproduced, stored or transmitted in any form or by any means, with the prior permission of the publisher, or, in the case of reprographic reproduction, in accordance with the terms of a licence issued by The Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to Facet Publishing, 7 Ridgmount Street, London WC1E 7AE.

First published 2006

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

ISBN-13: 978-1-85604-565-0

ISBN-10: 1-85604-565-X

Typeset in 11/14pt Aldine 721 and Humanist by Facet Publishing.

Printed and made in Great Britain by MPG Books Ltd, Bodmin, Cornwall.

Contents

Acknowledgements vii

1 **Introduction** **1**

2 **What is a thesaurus?** **3**

The reference thesaurus – the information retrieval thesaurus – appearance of the thesaurus – thesaurus applications – origin and development

3 **Tools for subject access and retrieval** **13**

Classification schemes – taxonomies – concept maps, topic maps and ontologies – keyword lists – subject heading lists – the thesaurus proper

4 **What a thesaurus is used for** **26**

The thesaurus as an indexing tool – as a source of metadata – as a search tool – query formulation and expansion – as a browse and navigation tool

5 **Why use a thesaurus?** **38**

Disadvantages of natural language – advantages of controlled language

6 **Types of thesaurus** **40**

General and special thesauri – guides and directories of thesauri – thesauri for non-textual resources – thesauri for local collections

7 **The format of a thesaurus** **49**

The systematic display – the alphabetic display – elements in the display

8 **Building a thesaurus I: vocabulary collection** **58**

Existing vocabularies – encyclopaedias and reference sources – dictionaries, glossaries, and word lists – published literature – other sources – a model thesaurus

IV ESSENTIAL THESAURUS CONSTRUCTION

9 Vocabulary control 1: selection of terms 70

Synonym control – preferred and non-preferred terms – guidelines for selection – names – disambiguation

10 Vocabulary control 2: form of entry 81

Grammatical forms – spelling – punctuation and other symbols – plural and singular – compound terms – factoring – semantic factoring – syntactical factoring – factoring rules – word order

11 Building a thesaurus 2: term extraction from document titles 99

Significant terms – simple vocabulary control – compound terms

12 Building a thesaurus 3: vocabulary analysis 107

Facet analysis – fundamental categories – categories versus clustering – categories and facets

13 The thesaural relationships 117

USE and UF – equivalence relationships – broader terms and narrower terms – hierarchical relationships – polyhierarchy – related (associative) terms

14 Building a thesaurus 4: introducing internal structure 131

Selection of terms – facets and arrays – order within arrays – order between arrays – exhausting the principle of division

15 Building a thesaurus 5: imposing hierarchy 143

Hierarchical structure – representation of hierarchy – authorities – hierarchies in different facets

16 Building a thesaurus 6: compound subjects and citation order 152

Order between facets – filing order of facets – citation order – standard citation order – combining and locating terms – nature of compound terms – notation

17	Building a thesaurus 7: conversion of the taxonomy to alphabetical format	170
	Editing the schedule for conversion – selecting the preferred terms – identifying hierarchical relationships – related terms – terms in the same array – coordinate terms – compound terms and BT/RT relationships	
18	Building a thesaurus 8: creating the thesaurus records	186
	Management of the thesaurus data – creating the thesaurus records – presentation of the entries – final editing	
19	Managing and maintaining the thesaurus: thesaurus software	197
	Updating the thesaurus – automating the thesaurus – sources of thesaurus software – directories and evaluative sources – features of the software	
20	Conclusion	206
	Glossary	208
	Bibliography	226
	Appendix 1: Sample titles for thesaurus vocabulary	228
	Appendix 2: Sample terms for the thesaurus	236
	Appendix 3: Facets at stage 1 of analysis	241
	Appendix 4: Facets at stage 2 of analysis	249
	Appendix 5: Completed systematic display	256
	Appendix 6: Thesaurus entries for sample page	277
	Index	281

Acknowledgements

I should like to thank Jean Aitchison for reading large sections of the text, for making some pertinent and useful suggestions, and for her invaluable help and advice in the area of all things thesaural. Her contribution to the development of the thesaurus over the last forty years cannot be overestimated, and I'm honoured to follow in that tradition.

I must also thank my colleague John Bowman for his keen eye for textual errors, and for his kindness in pointing these out diplomatically. He also made a number of helpful suggestions when my treatment of the more intractable parts of the subject had crossed the border into unintelligibility.

Needless to say, any errors that remain are my own.

Thanks are also due to those students of the School of Library, Archive & Information Studies at University College London who have undertaken the module in advanced cataloguing and classification over recent years. Their response to the teaching material has been an essential part of the development of the course on which this book is largely based. They have been a delight to teach, and their ingenuity in devising thesauri for the most unlikely of subjects is truly inspiring.

Vanda Broughton

