ISO 25964-1:2011-08 (E)

Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval

Contents

Page

Foreword v

Introduction vi

1 Scope 1

2 Terms and definitions 1

3 Symbols, abbreviated terms and other conventions 12

4 Thesaurus overview and objectives 15
4.1 Overall objective 15
4.2 Vocabulary control and its purpose 16
4.3 Paradigmatic versus syntagmatic relationships 16
4.4 Types of paradigmatic relationship 17

5 Concepts and their scope in a thesaurus 18
5.1 Conceptual basis 18
5.2 Scope notes 20
5.3 Reciprocal scope notes 21

6 Thesaurus terms 21
6.1 Form of terms 21
6.2 Clarification and disambiguation of thesaurus terms 21
6.3 Grammatical form of terms 23
6.4 Capitalization, punctuation and special characters 26
6.5 Singular or plural forms 27
6.6 Selection of the preferred form 30

7 Complex concepts 37
7.1 General 37
7.2 The nature of compound terms 38
7.3 Deciding whether or not to admit a complex concept 39
7.4 How to split a complex concept 43
7.5 Retention of constituent concepts 43
7.6 Consistency in the treatment of complex concepts 44
7.7 Order of words in multi-word terms 44

8 The equivalence relationship, in a monolingual context 44
8.1 General 44
8.2 Synonyms 45
8.3 Quasi-synonyms 48
8.4 Specific terms subsumed in a broader concept 48
8.5 Representation of complex concepts by a combination of terms 49

9 Equivalence across languages 50
9.1 General 50
9.2 Degrees of equivalence 51
9.3 Typical problems and solutions 52
9.4 Representation of cross-language equivalence between preferred terms 57
9.5 Cross-language equivalence between non-preferred terms 57

10 Relationships between concepts ... 57
10.1 Introduction ... 57
10.2 The hierarchical relationship ... 58
10.3 The associative relationship ... 63
10.4 Customized relationships ... 67

11 Facet analysis ... 68

12 Presentation and layout ... 70
12.1 General ... 70
12.2 Alternative display styles ... 71
12.3 Presentation and layout of multilingual thesauri ... 80
12.4 Language and character encoding issues ... 85

13 Managing thesaurus construction and maintenance ... 88
13.1 Planning a thesaurus ... 88
13.2 Early stages of compilation ... 90
13.3 Construction ... 91
13.4 Introduction to the thesaurus ... 93
13.5 Dissemination ... 93
13.6 Updating ... 95

14 Guidelines for thesaurus management software ... 98
14.1 General ... 98
14.2 Size and character limitations ... 98
14.3 Relationships between terms and between concepts ... 99
14.4 Notes applying to terms or concepts ... 100
14.5 Codes and notation ... 100
14.6 Node labels ... 100
14.7 Status of languages ... 100
14.8 Data import/export ... 101
14.9 Editorial navigation and support ... 102
14.10 Editorial safeguards ... 102
14.11 Housekeeping tools ... 103

15 Data model ... 103
15.1 General ... 103
15.2 Notes on the model ... 105
15.3 Tabular presentation ... 109

16 Integration of thesauri with applications ... 115
16.1 Introduction ... 115
16.2 Interoperability needs for thesauri ... 116
16.3 Integration with indexing and searching applications ... 116

17 Exchange formats ... 118

18 Protocols ... 119
18.1 General ... 119
18.2 Purposes and use cases ... 119
18.3 Application environment and architecture ... 120
18.4 Thesaurus-specific protocols ... 120
18.5 General-purpose web database protocols used with thesauri ... 120

Annex A (informative) Examples of displays found in published thesauri ... 122

Annex B (informative) XML Schema for data exchange ... 139

Bibliography ... 140

Index ... 144

Table 1 -- Symbols and abbreviations 13

Table 2 -- English language tags and their equivalents in other languages 14

Table A.1 -- Tags used in Inspec Thesaurus alphabetical display 122

Figure 1 -- Paradigmatic and syntagmatic relationships 17