

This is a preview of "ISO 25964-1:2011". [Click here to purchase the full version from the ANSI store.](#)

First edition
2011-08-15

Information and documentation — Thesauri and interoperability with other vocabularies —

Part 1: Thesauri for information retrieval

*Information et documentation — Thésaurus et interopérabilité avec
d'autres vocabulaires —*

Partie 1: Thésaurus pour la recherche documentaire



Reference number
ISO 25964-1:2011(E)

© ISO 2011

This is a preview of "ISO 25964-1:2011". [Click here to purchase the full version from the ANSI store.](#)



COPYRIGHT PROTECTED DOCUMENT

© ISO 2011

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

This is a preview of "ISO 25964-1:2011". Click here to purchase the full version from the ANSI store.

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

This is a preview of "ISO 25964-1:2011". [Click here to purchase the full version from the ANSI store.](#)

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

This is a preview of "ISO 25964-1:2011". [Click here to purchase the full version from the ANSI store.](#)

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 25964-1 was prepared by Technical Committee ISO/TC 46, *Information and documentation*, Subcommittee SC 9, *Identification and description*.

This first edition of ISO 25964-1 cancels and replaces ISO 2788:1986 and ISO 5964:1985, which have been merged and technically revised. Clauses 1 to 13 of this part of ISO 25964 correspond broadly to the content of ISO 2788:1986 and ISO 5964:1985. The remaining clauses cover new material.

ISO 25964 consists of the following parts, under the general title *Information and documentation — Thesauri and interoperability with other vocabularies*:

— *Part 1: Thesauri for information retrieval*

The following parts are under preparation:

— *Part 2: Interoperability with other vocabularies*

This part of ISO 25964 covers the development and maintenance of thesauri, both monolingual and multilingual, including formats and protocols for data exchange.

ISO 25964-2 will cover interoperability between different thesauri and with other types of structured vocabulary, such as classification schemes, name authority lists, ontologies, etc., not previously covered in any International Standard.

Introduction

Today's thesauri are mostly electronic tools, having moved on from the paper-based era when thesaurus standards were first developed. They are built and maintained with the support of software and need to integrate with other software, such as search engines and content management systems. (For example, data from the thesaurus database might need to be presented in combination with the number of postings found by a search application.) Whereas in the past thesauri were designed for information professionals trained in indexing and searching, today there is a demand for vocabularies that untrained users will find to be intuitive, and for vocabularies that enable inferencing by machines.

ISO 25964 makes the transition that is needed in order to be compatible with the world of electronic information management. However, this part of ISO 25964 retains the assumption that human intellect is usually involved in the selection of indexing terms and in the selection of search terms. If both the indexer and the searcher are guided to choose the same term for the same concept, then relevant documents will be retrieved. This is the main principle underlying thesaurus design, even though a thesaurus may also be applied in situations where computers make the choices.

Efficient exchange of data is a vital component of thesaurus management and exploitation. This part of ISO 25964 therefore includes recommendations for exchange formats and protocols. Adoption of these will facilitate interoperability between thesaurus management systems and other computer applications, such as indexing and retrieval systems, that will utilize the data.

This part of ISO 25964 covers development and maintenance of thesauri rather than how to use them in indexing. Where multilingual issues and examples are addressed, efforts have been made to cover as wide a selection of languages as possible, consistent with clarity and comprehensibility.

Thesauri are typically used in post-coordinate retrieval systems, but may also be applied to hierarchical directories, pre-coordinate indexes and classification systems. Increasingly, thesaurus applications need to mesh with others, such as automatic categorization schemes, free-text search systems, etc. ISO 25964-2 will address additional types of structured vocabulary (such as classification schemes, name authority lists, ontologies, etc.) and give recommendations to enable interoperation of the vocabularies at all stages of the information storage and retrieval process.