



SLOVENSKI STANDARD

SIST ISO 25964-2:2013

01-julij-2013

Informatika in dokumentacija - Tezavri in interoperabilnost z drugimi slovarji - 2. del: Interoperabilnost z drugimi slovarji

Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies

iTeh STANDARD PREVIEW

Information et documentation -- (Standards.iTeh.ai) Thésaurus et interopérabilité avec d'autres vocabulaires -- Partie 2: Interopérabilité avec d'autres vocabulaires

[SIST ISO 25964-2:2013](https://standards.iTeh.ai/catalog/standards/sist/05da009d-e0b1-4cd1-9eb9-514871cad084/sist-iso-25964-2-2013)

Ta slovenski standard je istoveten z: [ISO 25964-2:2013](https://standards.iTeh.ai/catalog/standards/sist/05da009d-e0b1-4cd1-9eb9-514871cad084/sist-iso-25964-2-2013)

ICS:

01.140.20 Informacijske vede Information sciences

SIST ISO 25964-2:2013

en,fr,de

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[SIST ISO 25964-2:2013](#)

<https://standards.iteh.ai/catalog/standards/sist/05da009d-e0b1-4cd1-9eb9-314871ead084/sist-iso-25964-2-2013>

INTERNATIONAL STANDARD

ISO
25964-2

First edition
2013-03-15

Information and documentation — Thesauri and interoperability with other vocabularies —

Part 2: Interoperability with other vocabularies

iTeh STANDARD PREVIEW
(standards.iteh.ai)
*Information et documentation — Thésaurus et interopérabilité avec
d'autres vocabulaires —
Partie 2: Interopérabilité avec d'autres vocabulaires*

[SIST ISO 25964-2:2013](https://standards.iteh.ai/catalog/standards/sist/05da009d-e0b1-4cd1-9eb9-314871ead084/sist-iso-25964-2-2013)

<https://standards.iteh.ai/catalog/standards/sist/05da009d-e0b1-4cd1-9eb9-314871ead084/sist-iso-25964-2-2013>



Reference number
ISO 25964-2:2013(E)

© ISO 2013

iTeh STANDARD PREVIEW (standards.iteh.ai)

SIST ISO 25964-2:2013

<https://standards.iteh.ai/catalog/standards/sist/05da009d-e0b1-4cd1-9eb9-314871ead084/sist-iso-25964-2-2013>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2013

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

Contents

Page

Foreword	v
Introduction.....	vi
1 Scope	1
2 Normative references.....	1
3 Terms and definitions	1
4 Symbols, abbreviations and other conventions	15
5 Objectives and identification	16
5.1 Objectives of interoperability and mapping	16
5.2 Identifying the elements to be mapped.....	16
6 Structural models for mapping across vocabularies	16
6.1 General	16
6.2 Model 1: Structural unity	17
6.3 Model 2: Direct-linked	17
6.4 Model 3: Hub structure	17
6.5 Selective mapping	19
6.6 Choosing among the options.....	20
7 Types of mapping.....	20
7.1 General	20
8 Equivalence mappings.....	21
8.1 General	21
8.2 Simple equivalence	21
8.3 Compound equivalence.....	22
9 Hierarchical mappings.....	25
10 Associative mappings.....	25
11 Exact, inexact and partial equivalence.....	26
11.1 General	26
11.2 Exact equivalence	26
11.3 Inexact equivalence.....	27
11.4 Partial equivalence	27
12 Use of mappings in information retrieval	28
12.1 General	28
12.2 Human mediation	28
12.3 Examples of mapping implementation.....	29
12.4 Conclusions and recommendations	30
13 Handling pre-coordination	31
13.1 General	31
13.2 Mapping between a thesaurus and a scheme with context-dependent captions	32
13.3 Mapping more complex classes	37
14 Techniques for identifying candidate mappings	38
14.1 General procedure.....	38
14.2 Computer assisted direct matching	40
14.3 Co-occurrence mapping	40
14.4 Other methods	40
15 Managing the data	41
15.1 Types of data to be recorded	41
15.2 Storing the data	43
15.3 Maintaining the mappings data.....	43

ISO 25964-2:2013(E)

16	Display of mapped vocabularies.....	45
16.1	General.....	45
16.2	Single record display	46
16.3	Complete displays based on one of the vocabularies.....	48
16.4	Crosswalks	49
17	Classification schemes	50
17.1	Key characteristics and background.....	50
17.2	Semantic components and relationships compared with thesaurus components	53
17.3	Recommendations for mapping between a thesaurus and a classification scheme	55
18	Classification schemes used for records management.....	56
18.1	Key characteristics and background.....	56
18.2	Semantic components and relationships compared with thesaurus components	57
18.3	Recommendations for interoperability with a thesaurus	58
19	Taxonomies	59
19.1	Key characteristics and background.....	59
19.2	Types of taxonomies	60
19.3	Semantic components and relationships	61
19.4	Mapping between a thesaurus and a taxonomy.....	63
20	Subject heading schemes.....	67
20.1	Key characteristics and background.....	67
20.2	Semantic components and relationships	68
20.3	Mapping subject headings to and from thesaurus concepts	69
21	Ontologies	72
21.1	Key characteristics and background.....	72
21.2	Semantic components and relationships	73
21.3	Structural comparison between thesauri and ontologies	76
21.4	Interoperability with thesauri.....	77
22	Terminologies	78
22.1	Key characteristics and background.....	78
22.2	Semantic components and relationships compared with those of a thesaurus	79
22.3	Interoperability with thesauri.....	81
23	Name authority lists	81
23.1	Key characteristics and background.....	81
23.2	Semantic components and relationships	83
23.3	Mapping between a thesaurus and a name authority list.....	85
24	Synonym rings	87
24.1	Key characteristics and background.....	87
24.2	Semantic components and relationships	88
24.3	Interoperability with thesauri.....	89
Annex A (informative) Management of terminological data in support of interoperability		90
Bibliography.....		93
Index.....		96

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 25964-2 was prepared by Technical Committee ISO/TC 46, *Information and documentation*, Subcommittee SC 9, *Identification and description*.

ISO 25964-1:2011, Clauses 1 to 13, correspond broadly to the content of ISO 2788:1986 and ISO 5964:1985. The remaining clauses of ISO 25964-1, and the whole of this part of ISO 25964, cover new ground not previously covered in any International Standard.

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*
- *Part 2: Interoperability with other vocabularies*

ISO 25964-1 covers the development and maintenance of thesauri, both monolingual and multilingual, including formats and protocols for data exchange.

This part of ISO 25964 covers interoperability between different thesauri and with other types of structured vocabulary.

ISO 25964 covers development and maintenance of thesauri rather than how to use them in indexing, which is covered by ISO 999.

Introduction

The ability to identify and locate relevant information among vast collections and other resources is a major and pressing challenge today, hence the need for semantic interoperability. To support this need, a variety of web services and other tools are under active development, including (but not limited to):

- registries of vocabularies and metadata schemas,
- repositories of vocabularies and metadata schemas, and
- crosswalk services.

While registries and repositories are outside the scope, this International Standard provides guidance fundamental to the establishment of crosswalk services. Typically these rely on mapping between different schemas and vocabularies. Accordingly, inter-vocabulary mapping will be the principal focus of this part of ISO 25964.

The foundation for discussion of interoperability is laid in ISO 25964-1, which describes the key characteristics of thesauri and provides guidance on best practice. Unfortunately, however, a comparable International Standard does not exist for the other types of vocabulary with which a thesaurus needs to interoperate. For this reason, this second part of ISO 25964 provides some elementary description of other vocabularies such as classification schemes, subject heading schemes, etc., before making recommendations on mapping between these and thesauri. It does not provide guidance on the construction of vocabularies other than thesauri.

Clauses 1 to 12 and Clauses 14 to 16 of this part of ISO 25964 deal with the principles and practicalities of interoperability, especially mapping, that apply to most vocabularies and especially thesauri. Clause 13 provides additional guidance for handling the pre-coordinated classes that occur in classification schemes and other vocabularies using the classification approach.

Clauses 17 to 24 are each dedicated to a different type of vocabulary. First priority is given to vocabularies that are commonly used for classifying or indexing resources, namely classification schemes (including those used for records management), taxonomies, subject heading schemes and name authority lists. Notwithstanding their different purposes, terminologies, ontologies and synonym rings are also included. Each clause provides a brief informative description of the vocabulary's key characteristics, contrasting its semantic components with those of a thesaurus, so as to provide context for the interoperability requirements and recommendations that follow.

Information and documentation — Thesauri and interoperability with other vocabularies —

Part 2: Interoperability with other vocabularies

1 Scope

This part of ISO 25964 is applicable to thesauri and other types of vocabulary that are commonly used for information retrieval. It describes, compares and contrasts the elements and features of these vocabularies that are implicated when interoperability is needed. It gives recommendations for the establishment and maintenance of mappings between multiple thesauri, or between thesauri and other types of vocabularies.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 25964-1:2011, *Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1

array

group of sibling **concepts** (3.17)

EXAMPLE

<p>In the following, the sibling concepts outerwear and underwear form an array within the concept clothing.</p> <p>clothing</p> <ul style="list-style-type: none">outerwear<ul style="list-style-type: none">overcoatsunderwear
--

3.2

associative relationship

relationship between a pair of **concepts** (3.17) that are not related hierarchically but share a strong semantic connection

ISO 25964-2:2013(E)

3.3

broader term

preferred term (3.62) representing a **concept** (3.17) that is broader than the one in question

NOTE The scope of the narrower concept falls completely within the scope of the broader. The relationship between the two is commonly indicated with the tag BT. For more explanation see ISO 25964-1:2011, 10.2.1.

3.4

caption

text label of the **class** (3.10) represented by a **notation** (3.54) in a **classification scheme** (3.12)

NOTE A caption is read in conjunction with its hierarchical context. It does not need to be as complete or as self-contained as a scope note or even a preferred term in a thesaurus. Captions are sometimes known as class headings or class names.

3.5

category

concept (3.17) or group of similar or related **concepts** (3.17) used as a division or subdivision in a **taxonomy** (3.83)

NOTE 1 In a classification scheme, such a group is normally called a class.

NOTE 2 This definition of category should not be confused with “fundamental categories”, discussed in ISO 25964-1:2011 (for example in Clause 12) in the context of faceted classification.

3.6

category label

text label representing a **category** (3.5) in a **taxonomy** (3.83)

NOTE The category label in a taxonomy is comparable to the caption in a classification scheme, and although a category is often referred to as a node, a category label should not be confused with a node label.

[SIST ISO 25964-2:2013](https://standards.iteh.ai/catalog/standards/sist/05da009d-e0b1-4cd1-9eb9-314871ead084/sist-iso-25964-2-2013)

3.7

chain index

index to a scheme that represents compound **concepts** (3.17) by a string of **pre-coordinated** (3.60) **terms** (3.84), such as a **classification scheme** (3.12), in which entries are generated by successive left truncation of strings of **terms** (3.84) representing complex **concepts** (3.17)

NOTE See the example in 17.2.4, Figure 12.

3.8

characteristic of division

attribute by which a **concept** (3.17) can be subdivided into an **array** (3.1) of narrower **concepts** (3.17) each having a distinct value of that attribute

cf. **node label** (3.52)

EXAMPLE

In the following, age group is the characteristic of division applied to the concept of people.

people
 (people by age group)
 children
 youths
 adults

3.9

citation order

order in which **preferred terms** (3.62) or **notations** (3.54) are combined in a **pre-coordinate** (3.60) **indexing** (3.36) system or a **classification scheme** (3.12) to form strings representing complex **concepts** (3.17)

3.10**class**

concept (3.17) or group of similar or related **concepts** (3.17) used as a division or subdivision in a **classification scheme** (3.12)

NOTE Classes are the basic units of which a classification scheme is constructed. In a taxonomy, although this is a type of classification scheme, they are generally known as categories (see 3.5). Occasionally they are described as “nodes”, although they should not be confused with node labels. The term “class” has a different meaning in the context of ontologies. See 21.2.2.

3.11**classification
classifying**

activity involving the components of grouping similar or related things together, separating dissimilar or unrelated things, and arranging the resulting **classes** (3.10) in a logical and helpful sequence

3.12**classification scheme**

schedule (3.67) of **concepts** (3.17) and **pre-coordinated** (3.60) combinations of **concepts** (3.17), arranged by **classification** (3.11)

NOTE A classification scheme often also includes an index.

3.13**collection**

set of information resources that can or could be accessed by a **structured vocabulary** (3.74), whether the items in the set are collected in one place or distributed over a network

3.14**compound equivalence**

relationship between **terms** (3.84) or **mapping** (3.41) between **concepts** (3.17) in which one **term** (3.84) or **concept** (3.17) in one context is represented by two or more **terms** (3.84) or **concepts** (3.17) in another

3.15**compound heading**

pre-coordinated heading

subject heading (3.76) formed by **pre-coordination** (3.60) of more than one **term** (3.84) into a string

EXAMPLE

The single terms Buddhism, Mass media and Religious aspects may be combined to form the compound heading Mass media — Religious aspects — Buddhism.

3.16**compound term**

term (3.84) that can be split morphologically into separate components

EXAMPLES

In English: “copper mines” can be split into “copper” and “mines” “lawnmowers” can be split into “lawns” and “mowers”
In French: “mine de cuivre” can be split into “mine” and “cuivre” “biodiversité” can be split into “biologie” and “diversité”

NOTE Compound terms can be multi-word terms, or can consist of only one word.

ISO 25964-2:2013(E)

3.17

concept

unit of thought

NOTE Concepts can be expressed in a variety of different ways. They exist in the mind as abstract entities independent of terms used to express them. They range from the very simple, e.g. “child”, to the very complex, e.g. “child protection legislation”.

3.18

concept group

group of **concepts** (3.17) selected by some specified criterion, such as relevance to a particular subject area

3.19

controlled vocabulary

prescribed list of **terms** (3.84), headings or codes, each representing a **concept** (3.17)

NOTE Controlled vocabularies are designed for applications in which it is useful to identify each concept with one consistent label, for example when classifying documents, indexing them and/or searching them. Thesauri, subject heading schemes and name authority lists are examples of controlled vocabularies.

cf. **terminology** (3.85)

3.20

cross-language equivalence

equivalence relationship (3.28) between **terms** (3.84) representing the same **concept** (3.17) in different languages

iTeh STANDARD PREVIEW
(standards.iteh.ai)

3.21

crosswalk

table of **mappings** (3.41) between the **concepts** (3.17) in two or more **structured vocabularies** (3.74)

SIST ISO 25964-2:2013

NOTE Crosswalks support the ability of search engines to search effectively across heterogeneous databases.

314871ead084/sist-iso-25964-2-2013

3.22

data model

abstract model that describes how data is represented and used

NOTE The data model in ISO 25964-1 provides a generic definition of thesaurus structure and semantics. It can be used as the basis for defining a database model or an exchange format for thesauri.

3.23

differentiated mapping

mapping (3.40) methodology that aims to distinguish between different types and qualities of **mapping** (3.41)

NOTE Types of mapping that can be distinguished include equivalence, associative and hierarchical; equivalence can be further subdivided into simple or compound, and the degree of equivalence can be marked to support further distinctions at the time of use.

cf. **undifferentiated mapping** (3.88)

3.24

document

any resource that can be **classified** (3.11) or indexed in order that the data or information in it may be retrieved

NOTE This definition refers not only to written and printed materials in paper or microform versions (for example, conventional books, journals, diagrams, maps), but also to non-printed media such as machine-readable and digitized records, internet and intranet resources, films, sound recordings, people and organizations as knowledge resources, buildings, sites, monuments, three-dimensional objects or realia; and collections of such items or parts of such items.

3.25**entry term**

lead-in term

term (3.84) provided in a **controlled vocabulary** (3.19), not for direct use in **metadata** (3.45), but for the purpose of guiding the user to another **term** (3.84) which may be used as a **category label** (3.6), **subject heading** (3.76) or **preferred term** (3.62)

NOTE Entry terms occurring in a thesaurus are generally known as non-preferred terms.

3.26**enumerative classification scheme**

classification scheme (3.12) in which all the **classes** (3.10) available are explicitly listed in the **schedules** (3.67)

cf. **synthetic classification scheme** (3.80)

3.27**equivalence mapping**

mapping (3.41) that states that the **concept** (3.17) in the **target vocabulary** (3.82) is considered identical in scope to the **concept** (3.17) in the **source vocabulary** (3.72)

cf. **equivalence relationship** (3.28)

3.28**equivalence relationship**

relationship between two **terms** (3.84) in a **thesaurus** (3.86) that both represent the same **concept** (3.17)

NOTE In ordinary discourse, terms that are quasi-synonyms may represent slightly different concepts. After inclusion in the thesaurus, however, the equivalence relationship clarifies that both are regarded as representing the same concept. When two or more such terms are in the same language within a monolingual or multilingual thesaurus, one of them is designated a preferred term and the other(s) as non-preferred term(s); when two or more such terms are in the different languages of a multilingual thesaurus, each of them may be a preferred term in its own language respectively, and the relationship is known as cross-language equivalence.

3.29**exchange format**

machine-readable format for representing information that is intended to facilitate exchange of the information between different applications

NOTE The exchange format for a thesaurus often uses a markup language based on a standard such as XML (Extensible Markup Language), and is based on a data model for thesauri. While the data model provides a generic description of thesaurus structure and semantics, the exchange format expresses this in a formal language for the purpose of exchanging thesauri.

3.30**facet**

grouping of **concepts** (3.17) of the same inherent category

EXAMPLE 1

Animals, mice, daffodils and bacteria could all be members of a living organisms facet.

EXAMPLE 2

Digging, writing and cooking could all be members of an actions facet.

EXAMPLE 3

Paris, the United Kingdom and the Alps could all be members of a places facet.

ISO 25964-2:2013(E)

NOTE Examples of high-level categories that can be used for grouping concepts into facets are: objects, materials, agents, actions, places and times.

cf. **node label** (3.52)

3.31**faceted classification scheme**

classification scheme in which **concepts** (3.17) are analysed into their constituent **facets** (3.30)

NOTE Schedules are compiled for each facet, and terms or notations from these are then combined according to prescribed rules to express a complex concept. Some complex concepts are often enumerated explicitly in the schedules; others can be synthesized by the user.

3.32**hierarchical relationship**

relationship between a pair of **concepts** (3.17) of which one has a scope falling completely within the scope of the other

cf. **broader term** (3.3), **narrower term** (3.51)

NOTE Several different types of hierarchical relationship exist. For more explanation see ISO 25964-1:2011, 10.2.

3.33**homograph**

one of two or more words that are written in the same way, but have different meanings

EXAMPLE

iTeh STANDARD PREVIEW
(standards.iteh.ai)

In English:	
The word “bank” could refer to a financial institution or the side of a river.	https://standards.iteh.ai/catalog/standards/sist/05da009d-e0b1-4cd1-9eb9-314871ead084/sist-iso-25964-2-2013
In French:	
The word “avocat” could refer to a lawyer or to a fruit.	

NOTE Homographs are sometimes referred to as homonyms, although the latter term applies more broadly, as it also includes pairs of terms such as “weights” and “waits” in English or “mer” and “mère” in French, which sound the same although they are spelt differently.

3.34**identifier**

set of symbols, usually alphanumeric, designating a **concept** (3.17) or a **term** (3.84) or another entity for purposes of unique identification within a determined context or resource, especially in a computer system or network

NOTE A notation is sometimes used as an identifier.

3.35**index term**

term (3.84) assigned to a **document** (3.24) in the process of **indexing** (3.36)

NOTE Sometimes index terms are referred to as indexing terms, as keywords or as tags, but the latter terms have other meanings too. Preferred terms from a thesaurus are very often used as index terms.

3.36**indexing**

intellectual analysis of the **subject** (3.75) matter of a **document** (3.24) to identify the **concepts** (3.17) represented in it, and allocation of the corresponding **index terms** (3.35) to allow the information to be retrieved

NOTE The term “subject indexing” is often used for this activity, but as ISO 25964 does not deal with the indexing of other elements such as authors or dates, “indexing” is sufficient. Indexing can be carried out by human users or by automated agents.

3.37**information retrieval**

all the techniques and processes used to identify **documents** (3.24) relevant to an information need, from a **collection** (3.13) or network of information resources

NOTE Selection and inclusion of items in the collection are included in this definition; likewise browsing and other forms of information seeking.

3.38**interoperability**

ability of two or more systems or components to exchange information and to use the information that has been exchanged

NOTE Vocabularies can support interoperability by including mappings to other vocabularies, by presenting data in standard formats and by using systems that support common computer protocols.

3.39**map, verb**

establish relationships between the **concepts** (3.17) of one vocabulary and those of another

3.40**mapping, gerund (verbal noun)**

process of establishing relationships between the **concepts** (3.17) of one vocabulary and those of another

3.41**mapping, noun**

⟨product of mapping process⟩ relationship between a **concept** (3.17) in one vocabulary and one or more **concepts** (3.17) in another

NOTE 1 A mapping generally has a direction, as discussed in Clause 6.

NOTE 2 Exceptionally, a mapping may involve a combination of two or more target vocabularies, where one or more of them act(s) as qualifier to the other (see 8.3.4).

3.42**mapping cluster**

coordinated set of **mappings** (3.41) between the **concepts** (3.17) of three or more vocabularies

NOTE See examples of mapping clusters in 6.3 and 6.4. A mapping cluster is generally maintained and published with a particular publishing or application objective. For example, a cluster of mappings between four different thesauri might be maintained so that the user of any one of them can easily search document collections indexed with any of the four.

3.43**markup**

annotations or other type of encoding embedded in text, in conformity with a **markup language** (3.44)