BS ISO 25964-2:2013

BSI Standards Publication

Information and documentation — Thesauri and interoperability with other vocabularies

Part 2: Interoperability with other vocabularies

bsi.

...making excellence a habit.™

Document Sharing Hub

Full document details: BS ISO 25964-2:2013

National foreword

This British Standard is the UK implementation of ISO 25964-2:2013. It supersedes BS 8723-3:2007 which is withdrawn, and together with BS ISO 25964-1:2011, it supersedes BS 8723-1:2005 and BS 8723-4:2007, which are withdrawn.

The UK participation in its preparation was entrusted to Technical Committee IDT/2/2, Information description, source identification and indexing.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provi sions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2013.
Published by BSI Standards Limited 2013

ISBN 978 0 580 72688 0
ICS 01.140.20

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 May 2013.

Amendments/corrigenda issued since publication

Date	Text affected

Documents
Document Sharing Hub

Full document details: BS 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

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

Partie 2: Interopérabilité avec d'autres vocabulaires

Reference number
ISO 25964-2:2013(E)

Documents
Document Sharing Hub
© ISO 2013

Full document details: BS 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

Documents
Document Sharing Hub
© ISO 2013 – All rights reserved

Full document details: BS ISO 25964-2:2013

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

Documents
Document Sharing Hub

© ISO 2013 – All rights reserved

Full document details: BS ISO 25964-2:2013

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

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

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.

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

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.

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

INTERNATIONAL STANDARD ISO 25964-2:2013(E)

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

In the following, the sibling concepts outerwear and underwear form an array within the concept clothing.
clothing
outerwear
overcoats
underwear

3.2
associative relationship
relationship between a pair of **concept**s (3.17) that are not related hierarchically but share a strong semantic connection

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

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 **concept**s (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.

3.7
chain index
index to a scheme that represents compound **concept**s (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 **concept**s (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 **concept**s (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 term**s (3.62) or **notation**s (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 **concept**s (3.17)

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

3.10
class
concept (3.17) or group of similar or related **concept**s (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 **class**es (3.10) in a logical and helpful sequence

3.12
classification scheme
schedule (3.67) of **concept**s (3.17) and **pre-coordinated** (3.60) combinations of **concept**s (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 **term**s (3.84) or **mapping** (3.41) between **concept**s (3.17) in which one **term** (3.84) or **concept** (3.17) in one context is represented by two or more **term**s (3.84) or **concept**s (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 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

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 **concept**s (3.17) selected by some specified criterion, such as relevance to a particular subject area

3.19
controlled vocabulary
prescribed list of **term**s (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

3.21
crosswalk
table of **mapping**s (3.41) between the **concept**s (3.17) in two or more **structured vocabularies** (3.74)

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

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.

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

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 **class**es (3.10) available are explicitly listed in the **schedule**s (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 **term**s (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 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

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 **concept**s (3.17) are analysed into their constituent **facet**s (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 **concept**s (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

In English: The word “bank” could refer to a financial institution or the side of a river.
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.

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

3.36
indexing
intellectual analysis of the **subject** (3.75) matter of a **document** (3.24) to identify the **concept**s (3.17) represented in it, and allocation of the corresponding **index term**s (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 **concept**s (3.17) of one vocabulary and those of another

3.40
mapping, gerund (verbal noun)
process of establishing relationships between the **concept**s (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 **concept**s (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 **mapping**s (3.41) between the **concept**s (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)

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

3.44
markup language
set of encoding conventions that can be used to provide instructions for the interpretation of a text, by the use of annotations embedded in the text itself

NOTE The interpretation often concerns issues such as content, structure or rendering of the text. Widely used examples include HTML (Hypertext Markup Language), which is largely concerned with presentation, and XML (Extensible Markup Language), which addresses the structure of text.

3.45
metadata
data that identify attributes of a **document** (3.24) typically used to support functions such as location, discovery, documentation, evaluation and/or selection

NOTE Preferred terms or notations selected during the indexing process are commonly applied as metadata values.

3.46
microthesaurus
designated subset of a **thesaurus** (3.86) that is capable of functioning as a complete **thesaurus** (3.86)

3.47
monohierarchical structure
hierarchical arrangement of **concept**s (3.17), in a **thesaurus** (3.86) or **classification scheme** (3.12), in which each **concept** (3.17) can have only one broader **concept** (3.17) at the level immediately above

cf. **polyhierarchical structure** (3.58)

EXAMPLE

In a monohierarchical structure, the concept of pianos cannot be listed under keyboard instruments as well as under stringed instruments; a choice has to be made of one of these concepts to determine its placing.

3.48
multilingual thesaurus
thesaurus (3.86) in which **term**s (3.84) and relational structures are available in two or more natural languages

3.49
multi-word term
term (3.84) consisting of more than one word

cf. **compound term** (3.16)

EXAMPLE

cost benefit analysis

3.50
name authority list
controlled vocabulary (3.19) for use in naming particular entities consistently

NOTE The entities in question are unique individuals, such as Benjamin Disraeli, Kilimanjaro or the Bayeux Tapestry, rather than classes such as politicians, mountains or embroideries. A name authority list may also be known as a name authority file. In this part of ISO 25964, a name authority list is sometimes referred to simply as an authority list.

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

3.51
narrower term
preferred term (3.62) representing a **concept** (3.17) which is narrower 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 NT. For more explanation see ISO 25964-1:2011, 10.2.1.

3.52
node label
label inserted into a hierarchical or **classified** (3.11) display to show how the **term**s (3.84) have been arranged

NOTE A node label is neither a preferred term nor a non-preferred term. It contains one of two different types of information: either a) the name of a facet to which following terms belong; or b) the attribute or characteristic of division by which an array of sibling concepts has been sorted or grouped. See examples in ISO 25964-1:2011, Clause 11.

3.53
non-preferred term
non-descriptor
term (3.84) that is not assigned to **document**s (3.24) but is provided as an entry point in a **thesaurus** (3.86) or index

cf. **entry term** (3.25)

EXAMPLE

hounds
USE dogs

NOTE A non-preferred term is followed by a reference to the appropriate preferred term or preferred terms.

3.54
notation
class code
class number
classmark
set of symbols representing a **concept** (3.17) or **class** (3.10) in a **structured vocabulary** (3.74), especially a **classification scheme** (3.12)

EXAMPLES

Notation	Source vocabulary	Concept
07.04.4	ILO Thesaurus	Fishery policy and development
622.342 2	Dewey Decimal Classification	gold mining
373.3.016:51	Universal Decimal Classification	mathematics curriculum in primary schools
SBS XEJ B	Bliss Bibliographic Classification	endangered species law
H40-H42	International Statistical Classification of Diseases and Related Health Problems	Glaucoma

NOTE Notation is sometimes used to sort and/or locate concepts in a pre-determined systematic order and, optionally, to display how the components of complex concepts have been structured and grouped. A notation can provide the link between alphabetical and systematic lists in a thesaurus. In the context of classification schemes, “concepts” are often known as “subjects”, especially when they are complex as in the examples above.

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

3.55
one-to-many mapping
mapping (3.41) in which a single **concept** (3.17) in one vocabulary is **map**ped (3.39) to a combination of two or more **concept**s (3.17) in another vocabulary

NOTE This situation is different from one in which a concept has two or more independent mappings to concepts in the other vocabulary.

3.56
one-to-one mapping
mapping (3.41) in which a single **concep**t (3.17) in one vocabulary is **map**ped (3.39) to a single **concept** (3.17) in another vocabulary

NOTE 1 The terms or notations used to label the concepts in the two vocabularies may or may not be identical.

NOTE 2 It is possible for a single concept to have two or more one-to-one mappings, if these are independent of each other.

3.57
ontology
a formal, explicit specification of a shared conceptualization

NOTE This definition is attributable to Studer et al.[31], extending an earlier definition by Gruber[18], and is adopted in this part of ISO 25964 because it is widely accepted in the ontology development community. An ontology typically includes definitions of concepts and specified relationships between them, set out in a formal way so that a machine can use them for reasoning. This definition excludes thesauri, classification schemes and other structured vocabularies described in this part of ISO 25964, even though these are sometimes described as “lightweight ontologies”.

3.58
polyhierarchical structure
hierarchical arrangement of **concept**s (3.17), in a **thesaurus** (3.86) or **classification scheme** (3.12), in which each **concept** (3.17) can have more than one broader **concept** (3.17)

cf. **monohierarchical structure** (3.47)

EXAMPLE

In a polyhierarchical structure, organs (musical instruments) could be listed under keyboard instruments as well as under wind instruments.

NOTE In a polyhierarchical structure, a single concept can occur at more than one place in the hierarchical structure of the thesaurus. Its attributes and relationships, and specifically its narrower and related terms, are the same wherever it occurs.

3.59
post-coordination
combination of **preferred term**s (3.62) of a **controlled vocabulary** (3.19) at the time of searching

cf. **pre-coordination** (3.60)

EXAMPLE

The post-coordinated search expression “microwaves AND radiation” can be used to retrieve documents on microwave radiation, when these have been indexed under the separate terms “microwaves” and “radiation” rather than a compound term.

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

3.60
pre-coordination
combination of **concept**s (3.17), **class**es (3.10) or **term**s (3.84) of a **controlled vocabulary** (3.19) at the time of its construction or at the time of using it for **indexing** (3.36) or **classification** (3.11)

cf. **post-coordination** (3.59)

EXAMPLE 1

The class “general theory”, when placed within the broader class “music”, refers only to the pre-coordinated subject “theory of music” and not to theory in general.

EXAMPLE 2

The pre-coordinated string “cardboard — recycling” might appear in a subject heading scheme, or if not enumerated there, it might be synthesized by an indexer when needed for a particular document.

3.61
precision
measure of retrieval performance defined by R/T, where R is the number of relevant items retrieved and T is the total number of items retrieved from the same **collection** (3.13)

3.62
preferred term
descriptor
term (3.84) used to represent a **concept** (3.17) when **indexing** (3.36)

cf. **non-preferred term** (3.53)

NOTE A preferred term is usually a noun or noun phrase.

3.63
protocol
convention that defines the syntax, semantics and synchronization of the communication process between two computers in order to enable a particular service

3.64
quasi-synonym
near-synonym
one of two or more **term**s (3.84) whose meanings are generally regarded as different in ordinary usage but which may be treated as labels for the same **concept** (3.17), in a given **controlled vocabulary** (3.19)

EXAMPLES

diseases, disorders
earthquakes, earth tremors

3.65
recall
measure of retrieval performance defined by R/N, where R is the number of relevant items retrieved and N is the total number of relevant items in the **collection** (3.13)

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

3.66
related term
preferred term (3.62) representing a **concept** (3.17) which has an **associative relationship** (3.2) with the one in question

NOTE The relationship between related terms is commonly indicated with the tag RT. For more explanation, see ISO 25964-1:2011, 10.3.

3.67
schedule
terms (3.84), **notation**s (3.54), **caption**s (3.4), cross-references and **scope note**s (3.68) set out to exhibit the content and structure of a **structured vocabulary** (3.74)

3.68
scope note
note that defines or clarifies the semantic boundaries of a **concept** (3.17) as it is used in the **structured vocabulary** (3.74)

NOTE A term used to label a concept can have several meanings in normal usage. A scope note is used to restrict the concept to only one of those meanings, and where necessary refers to other concepts that are included or excluded from the scope of the concept being clarified.

3.69
search term
term (3.84) forming all or part of a search query

NOTE In the context of ISO 25964, search terms are usually drawn from a controlled vocabulary.

3.70
search thesaurus
vocabulary intended to assist searching even though it has not been used to **index** (3.36) the **document**s (3.24) being searched

NOTE Search thesauri are designed to facilitate choice of terms and/or expansion of search expressions to include terms for broader, narrower or related concepts, as well as synonyms. Optionally, a thesaurus complying with this part of ISO 25964 can be used as a search thesaurus.

3.71
source language
language serving as a starting point in translation or in a search for **term** (3.84) equivalents

3.72
source vocabulary
vocabulary that serves as a starting point when seeking a corresponding **term** (3.84) or **concept** (3.17) in another vocabulary

3.73
specificity
capability of a **structured vocabulary** (3.74) to express a **subject** (3.75) in depth and in detail

NOTE For more explanation see the discussion of specificity in ISO 25964-1:2011, 8.4, and other places.

3.74
structured vocabulary
organized set of **term**s (3.84), headings or codes representing **concept**s (3.17) and their inter-relationships, which can be used to support **information retrieval** (3.37)

NOTE A structured vocabulary can also be used for other purposes. In the context of information retrieval, the vocabulary needs to be accompanied by rules for how to apply the terms. Various types of structured vocabulary are discussed in this part of ISO 25964, including classification schemes, subject heading schemes, etc.

Documents
Document Sharing Hub
© ISO 2013 – All rights reserved
Full document details: BS ISO 25964-2:2013

3.75
subject
concept (3.17) or combination of **concept**s (3.17) dealt with in a **document** (3.24) or occurring in discourse

3.76
subject heading
term (3.84) or pre-coordinated string of **term**s (3.84), drawn from a **subject heading scheme** (3.77)

3.77
subject heading scheme
subject heading language
subject heading list
SHL
structured vocabulary (3.74) comprising **term**s (3.84) available for subject **indexing** (3.36), plus rules for combining them into pre-coordinated strings of **term**s (3.84) where necessary

3.78
synonym
one of two or more **term**s (3.84) denoting the same **concept** (3.17)

EXAMPLES

In English: guarantees, warranties heart attack, myocardial infarction HIV, human immunodeficiency virus
In French: schiste, phyllade VIH, virus de l'immunodéficience humaine crise cardiaque, infarctus du myocarde

NOTE Abbreviations and their full forms can be treated as synonyms.

3.79
synonym ring
set of synonymous or almost synonymous **term**s (3.84), any of which can be used to refer to a particular **concept** (3.17)

EXAMPLE

stream; river; brook; beck; burn

3.80
synthetic classification scheme
classification scheme (3.12) in which users can synthesize **notation** (3.54) for complex **class**es (3.10) from lists of simpler **class**es (3.10)

cf. **enumerative classification scheme** (3.26)

3.81
target language
language providing a translation or an equivalent for a **term** (3.84) existing in a **source language** (3.71)

3.82
target vocabulary
vocabulary in which a **term** (3.84) or **concept** (3.17) is sought corresponding to an existing **term** (3.84) or **concept** (3.17) in a **source vocabulary** (3.72)

AVA Documents
Document Sharing Hub

© ISO 2013 – All rights reserved

Full document details: BS ISO 25964-2:2013

3.83
taxonomy
scheme of **categories** (3.5) and subcategories that can be used to sort and otherwise organize items of knowledge or information

NOTE Taxonomies range from the very simple to the very complex. In the simplest taxonomies the categories are not necessarily divided into subcategories, while in the complex ones multiple hierarchical levels can be found. Other features too can be present, such as all the thesaurus features described in ISO 25964-1, and/or the features commonly found in classification schemes. Outside of this International Standard, the term is often used loosely to refer to any type of structured vocabulary.

3.84
term
word or phrase used to label a **concept** (3.17)

EXAMPLES

schools
school uniform
costs of schooling
teaching

NOTE Thesaurus terms can be either preferred terms or non-preferred terms.

3.85
terminology
set of designations belonging to one special language

[ISO 1087-1:2000, definition 3.5.1]

NOTE The term “special language” is defined in ISO 1087 to mean a “language used in a subject field and characterized by the use of specific linguistic means of expression”; while a “designation” is defined therein as a “representation of a concept by a sign which denotes it”.

3.86
thesaurus
controlled (3.19) and **structured vocabulary** (3.74) in which **concept**s (3.17) are represented by **term**s (3.84), organized so that relationships between **concept**s (3.17) are made explicit, and **preferred term**s (3.62) are accompanied by lead-in entries for **synonym**s (3.78) or **quasi-synonym**s (3.64)

NOTE The purpose of a thesaurus is to guide both the indexer and the searcher to select the same preferred term or combination of preferred terms to represent a given subject. For this reason a thesaurus is optimized for human navigability and terminological coverage of a domain.

3.87
top term
preferred term (3.62) representing a **concept** (3.17) that has no broader **concept** (3.17) in the **thesaurus** (3.86)

NOTE Sometimes indicated by the abbreviation TT.

3.88
undifferentiated mapping
mapping (3.40) methodology that does not distinguish between different types of **mapping** (3.40), nor indicate different degrees of equivalence

cf. **differentiated mapping** (3.23)

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

3.89
vocabulary control
management of a vocabulary in order to disambiguate and constrain the form of the **term**s (3.84) and limit the number of **concept**s (3.17) and **term**s (3.84) available for **indexing** (3.36)

NOTE Control is achieved by distinguishing between homographs, so that each one has just one meaning, and by picking out from a set of synonyms or quasi-synonyms the one which is to be preferred for use in indexing. The purpose of these restrictions is to increase the likelihood of indexers and searchers choosing the same term to label a particular concept.

4 Symbols, abbreviations and other conventions

For the purposes of this part of ISO 25964, the symbols, abbreviations and conventions in ISO 25964-1:2011, Clause 3 apply. Additional tags and symbols are set out in Table 1. The alpha-2 language codes from ISO 639-1 are also used in this International Standard.

In some examples, labels such as “VOC1” and “VOC2” are used to signify “Vocabulary 1”, “Vocabulary 2”, etc.

To indicate the direction of a mapping statement, one option is to use an arrow explicitly. Alternatively, in languages that are customarily read from left to right, the source vocabulary concept should be shown on the left, followed by the appropriate tag(s) from Table 1, followed by the target concept(s).

Table 1 — Additional abbreviations and symbols used in mappings

Symbol	Meaning
EQ	Equivalence; the term that follows the tag is the preferred term in a target vocabulary that is closest in meaning to the preferred term preceding the tag, from the source vocabulary.
=	This symbol (an equals sign, Unicode character U+003D) should be used in conjunction with a mapping abbreviation to indicate that the mapping is exact. Specifically, “=EQ” signifies exact equivalence.
~	This symbol (a tilde, Unicode character U+007E) should be used in conjunction with a mapping abbreviation to indicate that the mapping is inexact. Specifically, “~EQ” signifies inexact equivalence.
BM	Broader mapping; the term that follows the tag represents a concept having a wider meaning.
NM	Narrower mapping; the term that follows the tag refers to a concept with a more specific meaning.
RM	Related mapping; the term that follows the tag labels an associated concept, but is not a synonym, a quasi-synonym, a broader term or a narrower term.
|	This symbol, (a vertical bar or “pipe line”, Unicode character U+007C), is shown between two or more preferred terms from a target vocabulary, the scopes of which, in sum, best cover the scope of a broader preferred term in the source vocabulary. Each of the target concepts represents part of the scope of the source concept. When converting index terms, all the target vocabulary preferred terms should be applied to a record indexed with the source vocabulary term. When converting search statements, the target vocabulary preferred terms should be combined with Boolean OR. EXAMPLE **Source vocabulary** — **Target vocabulary** Livestock — sheep | cattle | pigs | poultry
+	This symbol (a plus sign, Unicode character U+002B) is shown between two or more preferred terms from a target vocabulary that are used in conjunction to represent a compound concept in the source vocabulary. Each of the target concepts represents an aspect of the source concept. When converting index terms, all the target vocabulary preferred terms should be applied to a record indexed with the source vocabulary term. When converting search statements, the target vocabulary preferred terms should be combined with Boolean AND. EXAMPLE **Source vocabulary** — **Target vocabulary** women executives — women + executives

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

5 Objectives and identification

5.1 Objectives of interoperability and mapping

During information retrieval, the principal aim of interoperability between vocabularies is to enable an expression formulated using one vocabulary to be converted to (or supplemented by) a corresponding expression in one or more other vocabularies. This applies whether or not the vocabularies use the same natural language. The expression in question may be either a search query or part of the metadata associated with a document. In both cases, mapping is the key step. If each of the concepts in Vocabulary A has been mapped to the corresponding concept(s) in Vocabulary B, it becomes possible to interchange (or augment) the terms or identifiers representing the concept in each of the vocabularies. At the working level, interoperability is enabled by establishing inter-concept mappings, particularly equivalence, for which guidance and recommendations are given in Clauses 7 to 13.

NOTE Another aspect of interoperability is enhancement of the vocabulary tools through activities such as merging multiple vocabularies or using portions of one to extend another. This International Standard should not be interpreted as restricting novel forms of interoperability that could emerge.

5.2 Identifying the elements to be mapped

For all types of vocabulary, mappings aim to show relationships between concepts in different vocabularies, although the concepts are generally known as “classes” in the context of a classification scheme, and “categories” in the case of a taxonomy. This International Standard often uses the term “concept” to take in all of these.

Table 2 shows the principal elements that are used to represent concepts in different vocabulary types. Statements of mappings should use these elements when intended for human readers. But when mappings are intended for manipulation by computer, the concepts should instead be represented by unique, persistent identifiers.

Table 2 — Human-readable elements used to represent concepts in mapping statements

Vocabulary type	Concepts represented by:
Thesaurus	Preferred terms
Classification scheme	Notations
Taxonomy	Category labels or notations (see Note)
Subject heading scheme	Headings
Name authority list	Names
Ontology	Labels
Terminology	Terms or other types of designation
NOTE Different styles of concept representation are used by different taxonomies. Where there is no notation or unique identifier and the category labels are not unique, it is usually necessary to spell out the whole hierarchical path to specify a given concept uniquely.	

6 Structural models for mapping across vocabularies

6.1 General

This clause deals with the overall models within which equivalence and other mappings should be managed. Three basic models are described in 6.2, 6.3 and 6.4, and an alternative approach in 6.5. Management of the mappings data is discussed in Clause 15.

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

6.2 Model 1: Structural unity

In the structural unity model, all of the participating vocabularies share exactly the same structure of hierarchical and associative relationships between concepts. The structure may be represented or expressed in any number of different languages, notations or coding systems. The data model in ISO 25964-1:2011, Clause 15, illustrates structural unity as applied to a symmetrical multilingual thesaurus and discussed in ISO 25964-1, 9.1. The structural simplicity of this model makes it both feasible and desirable to manage all the concepts, terms, notations, captions and relationships between them within one system.

NOTE Model 1 does not usually involve mapping, but is included for comprehensiveness and for comparison with the other models. Some applications call for Model 1 to be applied in combination with other models.

6.3 Model 2: Direct-linked

The direct-linked model addresses linkages between two or more vocabularies that do not share the same structure. As well as differing in scope, language and structure, the vocabularies may include other types of vocabulary (classification scheme, name authority list, etc.) as well as one or more thesauri. Direct mappings should be established between the concepts of each vocabulary and those of each other vocabulary. This model may be extended to any number of vocabularies, by establishing direct mappings from each vocabulary to each other one. Any one of the vocabularies may then be used to search any collection indexed with any other one of the vocabularies.

Figure 1 illustrates the mappings that are needed to handle four vocabularies, using the direct-linked model. The boxes represent the concepts in the four vocabularies, respectively. The mappings are represented by arrows between the concepts. Each arrow has a direction. For simplicity, Figure 1 shows all the mappings as double-headed arrows, indicating that the mappings are intended to work in both directions. To describe this more accurately, each double-headed arrow represents a pair of mappings, one in each direction. Thus Figure 1 shows a total of 12 sets of mappings, presented in six pairs.

To reduce the amount of mapping work, sometimes mappings in one direction only are established, enabling conversion of index terms or search terms in one direction only. One-way mappings would be represented by single-headed arrows in the same basic model.

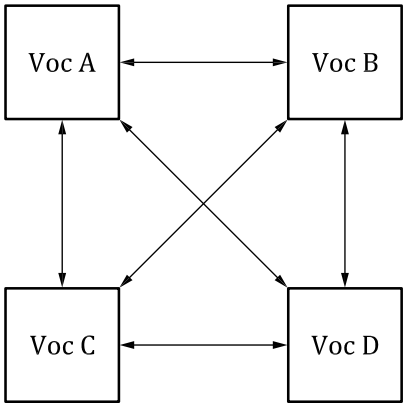

Figure 1 — Model 2 (direct-linked) as applied to four vocabularies

6.4 Model 3: Hub structure

Instead of establishing direct mappings between each pair of vocabularies, it is often convenient to designate one vocabulary as a “hub”, or comprehensive structure to which each of the other vocabularies is mapped. Figure 2 shows Model 3, in which vocabulary B serves as hub and the others as “satellites”.

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

Using this model, each concept in the hub vocabulary should be mapped to the corresponding concept(s) in the other vocabularies, and vice versa. In the same style as for Model 2, double-headed arrows indicate mappings in both directions.

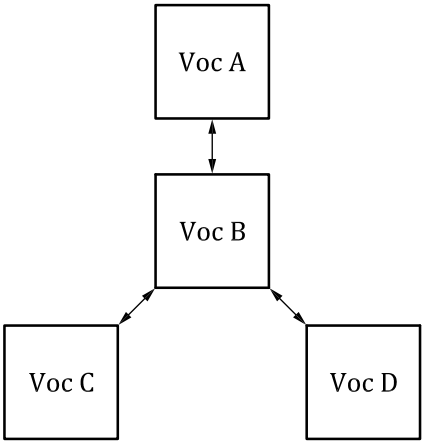

Figure 2 — Model 3 (hub model) as applied to four vocabularies

This model enables the hub vocabulary to be used for searching any of the resources indexed with any of the satellite vocabularies, either by conversion of search queries or by conversion of index terms. Similarly, an index term or a search term drawn from any of the satellite vocabularies can be converted to the corresponding term(s) in the hub vocabulary. A third possibility is to apply two conversions sequentially, for example from Voc A to Voc B, and then from Voc B to Voc D. When both steps involve exact, simple equivalence (see Clause 8 and 11.2), sequential mapping gives entirely acceptable results. But if the first step involves compound equivalence, or if two inexact mappings are applied in sequence, the quality of the result is unpredictable.

To achieve good quality mappings for the entire scope of the satellite vocabularies, the hub vocabulary needs to incorporate all the concepts present in all the satellites. When there is not a vocabulary of sufficient scope and specificity among those to be mapped, it may be necessary to construct a new hub vocabulary to meet the need. Alternatively, one of the original vocabularies may serve as hub if it is first extended to match the combined scope and depth of all the others.

For some applications, two-way mappings are not necessary. Figure 3 shows two variations on Model 3, in which mappings apply in one direction only.

— The variation on the left would allow a user of Vocabulary Q to search resources indexed with Vocabularies P, R or S. Alternatively, if the mappings in it were used to convert the metadata of resources indexed with Vocabulary Q, users of Vocabularies P, R and S would be enabled to search those resources.

— The variation on the right would allow users of Vocabularies W, Y and Z to search resources indexed with Vocabulary X. Alternatively, if the mappings in it were used to convert the metadata of resources indexed with Vocabularies W, Y or Z, users of Vocabulary X would be enabled to search those resources.

In neither case is it possible to convert terms from any one of the satellite vocabularies to any of the others.

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

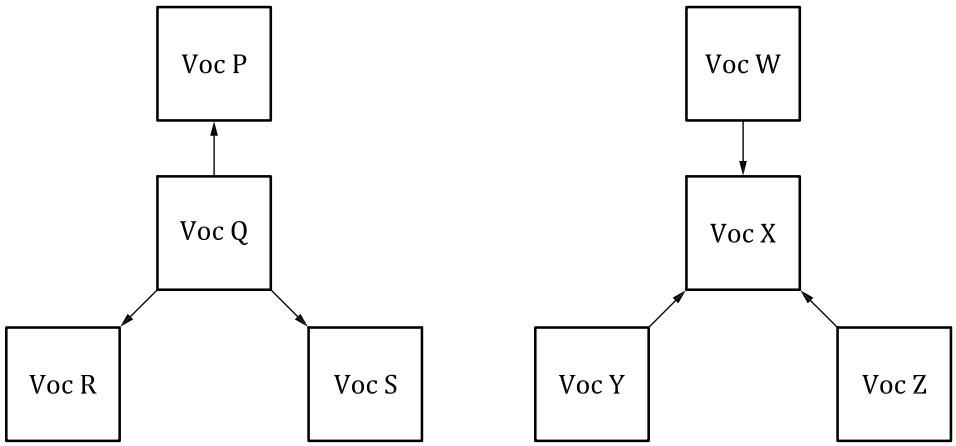

Figure 3 — Two variations on Model 3, with one-way mappings

6.5 Selective mapping

Models 1, 2 and 3 all require significant work to build and maintain. For some applications, however, it is unnecessary to map the vocabularies comprehensively. One alternative is to establish mappings only for the concepts that have been used or are likely to be used within the application in question. Two cases can be distinguished:

a) When overlap is small

If there are relatively few concepts common to two or more vocabularies, only a limited number of mappings can and should be established. See Figure 4.

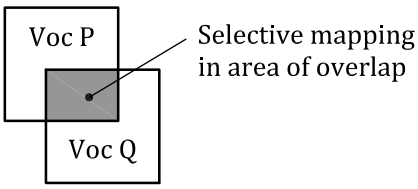

Figure 4 — Selective mapping where vocabulary overlap is small

b) Mapping from an index or catalogue

An approach that is often found convenient is to establish mappings for the headings (or terms or notations) that occur in an index or catalogue of the collection(s) involved in the application. Using this approach, the index or catalogue is treated as a sub-vocabulary derived from the full version of the vocabulary. This reduces the initial mapping effort, but can increase updating maintenance when changes are made in the collection. Figure 5 illustrates the approach when applied to a catalogue.

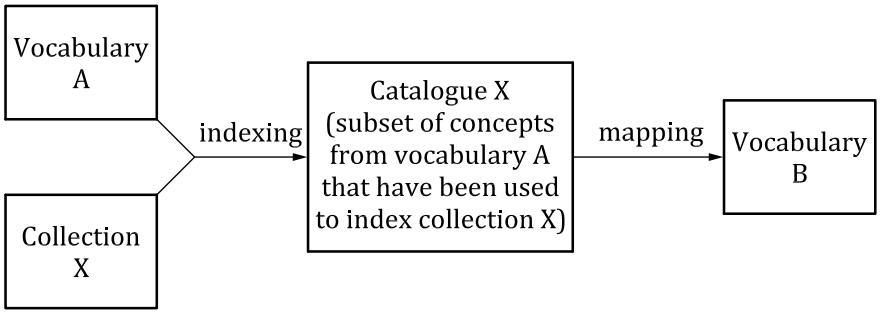

Figure 5 — Mapping between catalogue entries and a vocabulary

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

When a catalogue or index is used as the basis of mapping to more than one other vocabulary, it may be treated as the hub vocabulary in Model 3. Typically the mappings will enable conversion in one direction only, from the hub to the satellites.

6.6 Choosing among the options

The three structural models have been enumerated separately but in real applications combinations of them often occur and the boundaries might be blurred. At the start of any mapping project, it is important to clarify which model or combination of models will be used, with particular attention to the direction of the mappings.

In choosing among the options, the following factors should be borne in mind.

a) The structural unity model (Model 1) is most applicable when a new thesaurus is being prepared for use in more than one language, as described in ISO 25964-1. The absence of an existing indexed collection makes it easier to give each language equal status.

b) Models 2 and 3 are appropriate for the reconciliation of vocabularies that have been independently developed and/or have already been applied to collections. Model 2 is applicable particularly when there are only two or three such vocabularies.

c) Model 2 is more powerful than Model 3 if direct interoperability between all pairs of vocabularies is needed because it avoids sequential mapping. However, direct interoperability between the satellite vocabularies is not always needed. And the maintenance costs for Model 2 tend to be higher.

d) Between any pair of vocabularies, the mapping quality that can be achieved is best when the target vocabulary has equal specificity as well as the same breadth of coverage as the source vocabulary. (The mapping examples in 16.2 illustrate the difficulty when a target vocabulary is sparse in the subject area required.)

e) As a consequence of d), a vocabulary chosen or built as hub should be rich in the required subject area(s), at all levels. (That is to say, it needs to have at least equal specificity as well as the same breadth of coverage, as all of the satellite vocabularies combined.)

f) Among a group of vocabularies to be mapped, if one of them has significantly less specificity than the others, relatively few equivalence mappings will be possible. The more specific concepts in the other vocabularies can only be mapped hierarchically or associatively to the relatively generic concepts in the first. (See Clauses 7 to 10 for a discussion of the different types of mapping.)

g) If a multilingual thesaurus includes some non-symmetrical structures, a hybrid approach may be adopted. Model 1 should be used wherever possible, but Model 2 may be adopted when it proves impossible to establish one-to-one equivalence across languages.

h) Selective mapping based on an index or catalogue, as described in 6.5, has limited application. It may be less expensive to implement, but supports mappings for a particular application in one direction only. Since the index lacks some of the concepts and relationships in the vocabulary from which it was derived, the semantic structure of the vocabulary becomes incomplete and thus unavailable as a support for the mapping work or for the application of the mappings.

7 Types of mapping

7.1 General

When relationships are established across vocabularies, they are usually known as mappings. In principle, any type of relationship may be defined and applied as a mapping. In practice, the mappings likely to be useful are determined by the types of vocabulary to be linked. Between thesauri, the three main types of mapping to consider are equivalence, hierarchical and associative (exactly analogous to the relationships used internally within any thesaurus). Of these, equivalence is the type most commonly needed. Additional relationship types may be considered if another vocabulary type is involved, especially an ontology.

Documents
Document Sharing Hub
© ISO 2013 – All rights reserved
Full document details: BS ISO 25964-2:2013

The three main mapping types, and their subdivisions, are described in Clauses 8 to 10. Clause 12 shows how they are used when mapping between thesauri. Clause 13 deals with the more complex case where one of the vocabularies to be mapped is a pre-coordinate scheme such as a classification scheme or subject heading scheme.

8 Equivalence mappings

8.1 General

Equivalence should be established when matching concepts are found in two or more different vocabularies. Unlike the establishment of equivalence between two terms in a single monolingual thesaurus, where one of them is designated the preferred term while the other is designated a non-preferred term (see ISO 25964-1:2011, Clause 9), in the context of cross-vocabulary equivalence the mapping is between concepts and there is no difference of status between the concepts or between the preferred terms or notations representing them. The tag “EQ” should be used to indicate equivalence in mapping statements.

Typically the vocabularies include different selections of concepts and develop these to different levels of specificity. Hence several different equivalence situations can arise, with different solutions as described in subclauses 8.2 to 11.

8.2 Simple equivalence

In the ideal situation, the target vocabulary contains a concept identical in scope to the concept in the source vocabulary. A one-to-one equivalence mapping may immediately be established between the concepts.

EXAMPLES

Vocabulary 1	**Vocabulary 2**
mobile phones	cell phones
aubergines	egg-plants

The mapping is conventionally expressed in the following format: “mobile phones EQ cell phones”.

If more than two vocabularies are under consideration, the mapping statement may need to distinguish between them. Clause 16 shows how this should be done in displays for human readers.

Very often the preferred terms for the matching concepts are identical, in which case a mapping statement such as: “dishwashers EQ dishwashers” is appropriate.

The converse does not always apply. Identical preferred terms in two different vocabularies should not be accepted as equivalents without checking the underlying concepts. For example, the preferred term “operations” could have different meanings in a military thesaurus and a medical thesaurus.

Even when the contexts are similar, there can be subtle differences of scope. For example, the concept labelled “teachers” in one vocabulary might include teaching staff at universities as well as schools, whereas another vocabulary might limit the scope of “teachers” to schoolteachers, and provide a separate concept labelled “lecturers” for teaching staff at universities. Similarly the term “public schools” has a very different meaning for American and British contexts, because the educational systems of these countries are differently organized. Where identical terms are found in different vocabularies, an equivalence mapping should be established only if the underlying concepts are judged equivalent.

NOTE Further examples of equivalence, including situations where different degrees of equivalence apply, are given in Clause 11.

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

8.3 Compound equivalence

8.3.1 General

A complex concept that is included in one vocabulary with just one preferred term may be represented in another vocabulary by a combination of two or more concepts/terms.

EXAMPLES

Vocabulary 1	Vocabulary 2
genetically modified wheat	genetic modification wheat
inland waterways	rivers canals
fossil fuels	coal petroleum oil natural gas

In such cases a compound equivalence mapping (also known as one-to-many equivalence) may be established between the concepts. The mapping generally applies in one direction only.

Two distinct types of compound equivalence are commonly found, known as “intersecting” and “cumulative” respectively.

8.3.2 Intersecting compound equivalence (EQ +)

The intersecting compound equivalence mapping type gets its name because it can be represented in set theory as the intersection of two or more sets. Figure 6 illustrates the first of the following examples.

EXAMPLES

Vocabulary 1	Vocabulary 2
women executives	women executives
genetically modified wheat	genetic modification wheat
abused children	children abuse
rail passenger safety	railways passengers safety

Intersecting compound equivalence mappings are not reversible. For example, if “women” and “executives” coexist in the metadata of a single document, the combination should not be mapped automatically to “women executives”, since the document might instead deal with relationships between male executives and other women.

The use of some examples of compound equivalence is illustrated in 12.3.

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

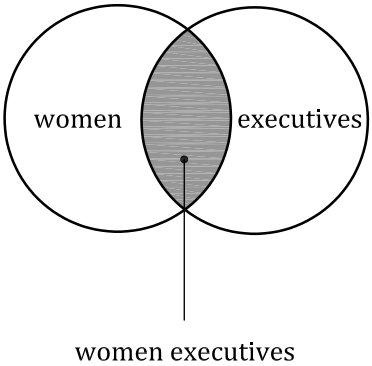

Figure 6 — The set of women who are also executives is the subset shown at the intersection of the two larger sets

A statement of intersecting compound equivalence should be conventionally expressed in the following format: “women executives EQ women + executives”

If the mappings will be used only for conversion of search expressions, the symbol “+” in mapping statements may be replaced by the Boolean operator “AND”. Boolean operators do not apply in the context of mapping index terms. See 12.3 for examples of how mappings are used in information retrieval.

8.3.3 Cumulative compound equivalence (EQ |)

The cumulative compound equivalence mapping type applies when a complex concept in one vocabulary corresponds to the sum of two or more concepts in another. Figure 7 shows its representation in set theory. Note that the simpler concepts do not necessarily overlap, but they will add up to the whole of the complex concept.

EXAMPLES

Vocabulary 1	**Vocabulary 2**
inland waterways	rivers canals
hosiery	stockings socks
woody plants	trees shrubs
fossil fuels	coal petroleum oil natural gas

Cumulative compound equivalence mappings should not be used in the reverse direction without checking whether another mapping would be more suitable. For example, in Vocabulary 1 the concept “inland waterways” might have narrower concepts “rivers” and “canals”. In this case, “rivers” and “canals” in Vocabulary 2 should be mapped to their simple equivalents, in preference to the broader concept.

The use of some examples of compound equivalence is illustrated in 12.3.

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

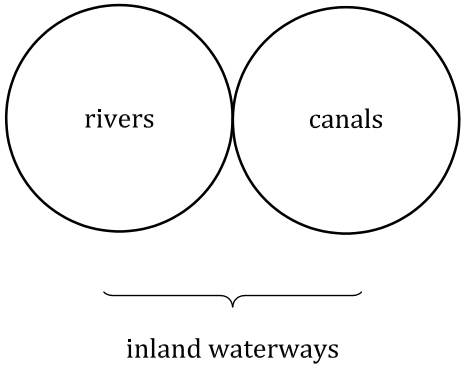

Figure 7 — The set of inland waterways is the superset composed of the set of rivers plus the set of canals

NOTE This example works best in contexts and regions where all the rivers are navigable, and there are no lakes or other bodies of water used as inland waterways. See 11.4 c) for advice on the situation where the narrower concepts available do not add up to the whole complex concept.

A statement of cumulative compound equivalence should be conventionally expressed in the following format: "inland waterways EQ rivers | canals"

If the mappings will be used only for conversion of search expressions, the symbol "|" in mapping statements may be replaced by the Boolean operator "OR". Boolean operators do not apply in the context of mapping index terms.

8.3.4 Compound equivalence involving a combination of target vocabularies

A special case of compound equivalence can arise when the target vocabulary is inadequate to convey the concepts in a source vocabulary, unless it is accompanied by a "qualifier vocabulary", that is to say a vocabulary composed of concepts that can be used to qualify the concepts in the main target vocabulary. In this unusual circumstance, a single concept in the source vocabulary may be mapped to a combination in which one concept is drawn from the main target vocabulary and the other(s) is/are selected from the qualifier vocabulary/ies.

EXAMPLE

A source vocabulary for languages includes a comprehensive listing of regional dialects and other sublanguages as well as main languages. In this vocabulary, the English spoken in Britain has preferred term "en-GB", while that spoken in the United States of America has preferred term "en-US".

For a particular application, mappings are required to a target vocabulary (VocL) which lists the main languages, but has no provision for dialects. It is accompanied by another "qualifying" target vocabulary (VocG) with a comprehensive listing of countries and regions. Compound equivalence mappings in the following style may be established:

en-GB EQ VocL:English + VocG:United Kingdom

en-US EQ VocL:English + VocG:United States of America

Combinations as described here should be used with caution, since they may have limited validity outside the particular application for which they are intended. The index term "United Kingdom", for example, is normally used only for documents about this country, and could prove misleading if applied to situations where the country is not involved at all.

© ISO 2013 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013

9 Hierarchical mappings

Hierarchical mapping may be established between concepts when one is clearly broader than the other. Just as for hierarchical relationships within a single thesaurus, described in ISO 25964-1:2011, 11.2, the relationship between the concepts can be generic or instantial. The whole-part situation may also justify establishment of a hierarchical mapping, in the limited circumstances described in ISO 25964-1:2011, 11.2.3.

EXAMPLES

Vocabulary 1	**Vocabulary 2**
school staff	schoolteachers
roads	streets
international organizations	United Nations
armies	battalions
United Kingdom	England

A hierarchical mapping from the narrower to the broader concept should be expressed in the following format: “streets BM roads”. The reciprocal hierarchical mapping, from the broader to the narrower concept, should be expressed in the following format: “roads NM streets”.

NOTE The tags BM and NM, short for Broader Mapping and Narrower Mapping respectively, are analogous to the BT and NT tags used within a single thesaurus. Use of a slightly different tag avoids confusion in displays that include cross-vocabulary mappings as well as internal relationships.

Optionally, it is possible to make a formal distinction between generic, instantial and whole-part hierarchical mappings, just as may be done for generic, instantial and whole-part hierarchical relationships within a single thesaurus. Instead of the tags BTG/NTG, BTI/NTI and BTP/NTP described in ISO 25964-1:2011, 10.2, the tags BMG/NMG, BMI/NMI and BMP/NMP should be used respectively, as illustrated in Table 3 below.

Table 3 — Mapping statements that distinguish between subtypes of hierarchical mapping

Subtype	Mapping statement example	Reciprocal example
Generic	rats BMG rodents	rodents NMG rats
Instantial	Paris BMI capital cities	capital cities NMI Paris
Whole-part	fingers BMP hands	hands NMP fingers

10 Associative mappings

The associative mapping type may be established between concepts when they do not qualify for equivalence or hierarchical mappings, but are semantically associated to such an extent that documents indexed with the one are likely to be relevant in a search for the other.

The dividing line between an associative mapping and inexact equivalence (see 11.3) is ill-defined and subjective, but may be made pragmatically according to the context in which the mappings are to be used (taking into account, for example, the interests of the users, the volume of relevant resources to be searched, and the capability of the retrieval and display systems to make fine distinctions).

© ISO 2013 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25964-2:2013