

The name that is most familiar to the users of the thesaurus should be designated as the preferred term. Preference should be given to the official rather than the popular name. The short form of the official name should be preferred. Standard authorities, such as ISO 3166, should be consulted for the official forms. Reciprocal references should be made between the preferred and non-preferred forms.

EXAMPLE 4

Ireland	
UF	<i>Irish Republic</i>
	<i>Eire</i>
<i>Eire</i>	
USE	Ireland
<i>Irish Republic</i>	
USE	Ireland

6.6.9.3 Names of institutions and persons

Variant names are common and can create problems if they are not controlled. When included in a thesaurus, the form of the names should be selected in accordance with a recognized code of cataloguing practice, such as the *Anglo-American cataloguing rules*³⁾[37]. Care should be taken to ensure the name used as the preferred term is correct, current and sufficiently complete to avoid confusion with other persons or institutions. Alternative forms of the name that are in common use should be established as non-preferred terms.

EXAMPLES

Armstrong, Louis (1901-1971)	UF	Satchmo (1901-1971)
Berners-Lee, Tim		
Otlet, Paul (1868-1944)		
Centre national de la danse (France)	UF	CND (Centre national de la danse)
Intergovernmental Committee for Physical Education and Sport		
Association for Professional Broadcasting Education (U.S.)	UF	Broadcast Education Association (U.S.)

7 Complex concepts

7.1 General

Concepts range from the very simple to the very complex. At the simple end are notions such as "silver" or "people". Greater complexity is found in "Georgian silver teapots" or "human rights campaigners". The latter expression, for example, combines at least three distinct concepts into one more complex concept.

Complex concepts are very often conveyed by compound terms, i.e. terms that can be morphologically split into two or more components. Some of these are multi-word terms, and some consist of just one word.

"Biodegradability", for example, is a single-word term combining the idea of being easy to break down with the notion that living organisms will be the cause of the breakdown. In theory one could synthesize the concept with a combination of "degradability + living organisms".

3) While this set of rules is widely used at the time of preparation of this part of ISO 25964, there are plans to replace it with a new standard known as RDA: Resource Description and Access.

ISO 25964-1:2011(E)

Splitting "biodegradability" into two concepts might seem cumbersome or unsatisfactory, but when the complex concept is usually conveyed by a multi-word term, splitting into simpler concepts becomes a more easily understandable option. For example, "human rights campaigners" can easily be split into "human rights + campaigners", or "camping holidays" could be adequately conveyed using "camping + holidays". Some concepts present many options for synthesis. For example, "human resource management" could be conveyed by "human resources + management" or "people + resource management" or even "people + resources + management".

The availability of so many choices presents the thesaurus editor with a difficult and subjective decision: whether to admit the complex concept or whether to rely on simpler concepts for users to apply in combination. Where the thesaurus allows several options, an important consideration is to guide all users to use the same combination for a given concept. Consistency becomes even harder to achieve if the concept is split into more than two components, as in the example of "people + resources + management".

Another consideration is that the introduction of complex concepts, whether represented by single-word or multi-word terms, tends to increase specificity. Greater specificity helps users to achieve greater discrimination. For example, if "biodegradability" is admitted as a narrower term of "degradability", searchers have an easy way to find relevant documents without unwanted items on degradability by chemical means, the action of ultraviolet light, etc. However, there is a penalty in that the thesaurus becomes larger and sometimes separates like items, making it harder for users to identify the right term.

Most of the examples in Clause 7 are based on multi-word terms. However, similar considerations apply to many complex concepts conveyed by single-word terms. Furthermore, in languages such as German, which uses very many compound words, the option of splitting single-word terms will occur much more commonly than in English and the same guidance applies.

7.2 The nature of compound terms

In the English language, compound terms very often consist of more than one word.

EXAMPLE SET 1

road safety thesaurus management software birds of prey

The parts of most such terms can be distinguished as follows.

- a) The focus or head, i.e. the noun component that identifies the general class of concepts to which the term as a whole refers.

EXAMPLE SET 2 (of term focus)

- | |
|---|
| 1) the noun component "indexes" in the term "printed indexes"
2) the noun "hospitals" in the prepositional phrase "hospitals for children" |
|---|

- b) The difference or modifier, i.e. one or more further components that serve to narrow the scope of the focus and thereby specify one of its subclasses.

EXAMPLE SET 3 (of term difference)

- | |
|--|
| 1) the adjective "printed" in the term "printed indexes"
2) the preposition-plus-noun combination "for children" in the term "hospitals for children" |
|--|

A similar analysis can sometimes be applied to single-word compound terms.

EXAMPLE SET 4 (of single-word compound terms)

Term	Focus	Difference
agroindustry	industry	agro (derived from "agriculture")
biosecurity	security	bio (prefix associated with biology)

7.3 Deciding whether or not to admit a complex concept

7.3.1 The options in outline

Taking a straightforward example, a concept such as "road safety" could be represented using a combination of the simpler concepts "roads" and "safety". As the complexity increases, more choices have to be considered. Thus, thesaurus management software could be represented by "thesauri" + "management" + "software" or by "thesaurus management" + "software".

The following five main options should be considered for handling a proposed complex concept.

- a) Admit the concept, representing it with a single preferred term.

EXAMPLE SET 1

camping holidays road safety thesaurus management software
--

- b) Admit the concept, representing it as a combination of two or more terms. In this case, a non-preferred term for the whole concept should be entered, pointing to the appropriate combination of simpler preferred terms. This option is often referred to as "splitting" the concept.

EXAMPLE SET 2

camping holidays USE camping + holidays thesaurus management software USE thesaurus management + software
--

- c) Admit the concept as a candidate, for review after a trial period.
- d) Reject the complex concept, as too infrequently required, irrelevant or inappropriate.
- e) Reject the concept for direct inclusion, but retain a broader concept and admit a non-preferred term pointing to the broader preferred term.

EXAMPLE SET 3

road safety USE safety thesaurus management software USE software
--

ISO 25964-1:2011(E)

7.3.2 Factors to consider

The decision on admission is often difficult and subjective. It may in part be guided by analysis of any transaction log files that are available (see 13.1.3.3 and 13.2.2). The thesaurus developer should also consider the balance of advantages in the retrieval situations that are likely to occur. Factors such as the following (which are not listed in order of priority) should be borne in mind.

- a) Frequency of use. Some provision should be made for a candidate complex concept if the concept is frequently sought or if it is of growing importance to the user community. This is all the more straightforward if the term representing it is widely used and understood by the audience. Option a) of 7.3.1 should usually be adopted but options b) and c) may also be considered.
- b) Degree of discrimination needed. If there is a large anticipated volume of relevant documents to be indexed, it may be important to discriminate between the complex concept and similar concepts. Where "road safety" occurs, for example, the same document collection may also deal with rail safety and/or transport safety more generally, as well as fire safety, safety in the home, etc. Many users might want to discriminate between these different aspects of safety. Admitting all these terms gives an increased level of specificity, allowing greater precision in both indexing and searching.
- c) Multiplicity of search options, leading to incomplete retrieval. The availability of a number of closely-related preferred terms complicates the choice of options. For example, if the thesaurus has all of the terms "safety", "transport safety", "passenger safety", "pedestrian safety", "vehicle safety" as well as "road safety", the searcher might need to try all these options (and also combinations of "safety" with terms such as "roads", "vehicles", "cars", "road traffic", etc.) to be sure of finding all the relevant material. If the total number of documents on safety is small, searching might be easier and more effective if none of the complex concepts is accepted as a preferred term and a post-coordinated search expression such as "roads + safety" is the only retrieval option.
- d) Compound terms in which one focus has more than one difference. In the preceding examples, relatively simple terms were used in which the focus "safety" was qualified by only one difference, for example "transport". Greater complexity occurs with terms such as "vehicle fire safety", for which the focus is qualified simultaneously by two differences. Terms like these multiply still further the search options and impede straightforward retrieval, and so they should be avoided. For example, the combination of "vehicle safety" + "fire safety" adequately represents the needed concept and causes no confusion.
- e) Expectations of the anticipated users. A term such as "pre-coordinate indexes" could be acceptable and useful in a thesaurus to serve information retrieval specialists, but could be confusing if the thesaurus is intended for a more general audience. In the latter case the best solution might be not to split the term into "pre-coordination" + "indexes", but to choose option e) of 7.3.1, making it a lead-in term to the broader term "indexes" (see 8.4).
- f) Avoiding ambiguity. Sometimes inclusion of a complex concept is needed to avoid ambiguity if the component concepts could be combined in different ways, having different meanings. For example, a combination of "libraries" and "science" could be used to represent "library science" or "science libraries". A preferred term for one or both of the complex concepts may, therefore, be admitted to avoid the retrieval of unwanted items.
- g) Multilingual complications (see Clause 9). In the case of multilingual thesauri, constraints or particular needs found in any one of the languages may influence inclusion or exclusion of a complex concept across all the languages.
- h) Anticipated search environment. If the search software offers sophisticated capabilities such as clustering of search results, customization to reflect the interests of particular users or user communities, etc., the level of specificity needed in the thesaurus can be affected. This factor should be balanced against the probability that the software and other aspects of the environment might change in future.

7.3.3 Circumstances that favour splitting a complex concept

The following circumstances favour splitting.

- a) Splitting may be considered when the concept is quite specific and falls outside the core scope of the thesaurus. Inclusion of a large number of peripheral terms increases the bulk and complexity of the vocabulary without giving much retrieval benefit.
- b) If very few documents are likely to be indexed with the proposed term, inclusion as a preferred term might not be worthwhile, and splitting is an option to consider.
- c) The complex concept should be split if the focus is qualified by more than one difference. For example, "underwater cine cameras" should be split into "underwater cameras" and "cine cameras".
- d) The complex concept should usually be split if the focus represents a property, part or component of the difference.

EXAMPLES

aircraft engines hospital floors instrument reliability soil acidity

However, exceptions can occur if the concept is well-defined and distinctive, for example "lamp shades" or "body temperature". References to body temperature are likely to differ significantly from references to the temperature of other inanimate objects or substances that might be discussed in the same document collection.

7.3.4 Circumstances that rule against splitting a complex concept

A complex concept should not be split if the following conditions apply.

- a) A term for the concept has become so familiar in common use, or in the field covered by the thesaurus, that its expression as separate elements would hinder comprehension.

EXAMPLE SET 1

data processing gross domestic product

- b) Splitting the concept into its parts would lead to a loss of meaning, or ambiguity.

EXAMPLE SET 2

In English: plant food (“food” + “plants” could represent either “food” for “plants” or “plants” as “food”.)
In French: marque de voiture (“marque” + “voiture” could represent either a “voiture” used as a “marque” or a “marque” of a “voiture”.) société de construction (“société” + “construction” could represent either a “société” for “construction” or the “construction” of a “société”.)

ISO 25964-1:2011(E)

- c) The concept is represented by an established term that is a proper name, or incorporates a proper name.

EXAMPLE SET 3

Boolean logic United Nations

- d) The difference in an established term has lost its original meaning.

EXAMPLE SET 4

In English: lawn tennis deck chairs trade winds
In French: pistolet à température maison mère police d'assurance

- e) The concept is represented by a term containing a difference suggesting a resemblance, as a metaphor, to an unrelated thing or event.

EXAMPLE SET 5

tree structures wing nuts

- f) The parts of the concept when separated do not convey the overall concept. Thus, it would be misleading to index the corresponding documents with terms representing the part concepts.

EXAMPLE SET 6

In English: fire escapes swimming pools	(a discussion of fire escapes might say nothing about fires or escaping) (a discussion of swimming pools might have little information on swimming)
In French: établissement de santé classe de neige	(a discussion of "établissement de santé" might have little information on "santé") (a discussion of "classe de neige" might have little information on "neige")

- g) The concept is represented by a term in which the focus has another meaning in the absence of the difference.

EXAMPLE SET 7

artificial flowers chocolate eggs	(It would be misleading to use the term "eggs" to index items dealing with chocolate eggs, since the chocolate ones are a type of confectionery rather than a type of egg.)
--------------------------------------	---

7.4 How to split a complex concept

In simple cases, for example that of road safety, the concept may be adequately expressed using the words in the multi-word term, i.e. "roads" + "safety". The justification for doing this in a particular thesaurus might be that any document dealing with road safety is likely to contain information relevant to both of the constituent terms and that somebody looking for information about roads is likely to judge an article on road safety as relevant.

When a concept is split, the constituent concepts may be represented by preferred terms that differ from the words of the multi-word term. Often the constituent words need to be modified to convey the correct concept. For example, "rail safety" should not be split into "rails" + "safety", but might be expressed as "railways" + "safety". For "human resources" the combination of "people" + "resources" is probably more acceptable to users than "humans" + "resources".

For complex concepts that do not justify a single preferred term but are likely to be sought by users, thesaurus entries of the following form should be provided (see 8.5 for more details).

EXAMPLE

rail safety USE railways + safety

7.5 Retention of constituent concepts

7.5.1 General

In the event that a compound term is accepted as a preferred term, consideration should be given to inclusion of preferred terms representing each of its constituent concepts if they are not already present in the thesaurus. For example, if "road safety" is accepted, the thesaurus should usually also contain the preferred terms "roads" and "safety". The latter term would be established as a broader term of "road safety", and "roads" as a related term (see 10.3).

7.5.2 Parts and components

Confusion can arise when the difference of the compound term is a machine or other complex assembly of which the focus refers to a component. For example, if "aircraft engines" is split into "aircraft" and "engines", a search for the term "aircraft" on its own will return records of all aircraft components as well as complete aircraft.

This might be acceptable for some collections, but in others will result in an overloading of the term "aircraft" and provide no means of isolating the documents dealing with whole aircraft. One solution is to create a preferred term such as "aircraft components". This term may be used in combination with "engines", or "instrumentation" or "under-carriages" or any other type of component.

ISO 25964-1:2011(E)

7.6 Consistency in the treatment of complex concepts

In general, consistency of practice is a useful aim as it encourages consistent application by users. However, absolute consistency in the admission of complex concepts is difficult to achieve and is not always necessary. As the specificity needed in core areas of the thesaurus is usually greater than that for peripheral areas, some inconsistencies are inevitable. This type of inconsistency does not have a negative effect on retrieval performance as long as there are enough clear and helpful entries in the thesaurus, and so it may be disregarded.

In some circumstances, however, consistent rules are helpful for thesaurus editors. For example, when a thesaurus is shared among several organizations, using it for different document collections and user groups, judgements about the level of specificity are harder for one person to make. Similarly, if thesaurus compilation is a networked project, in which suggestions are invited from hundreds of contributors around the world, some guidance is necessary.

To achieve consistency, a set of criteria appropriate to the subject field should be established. For example, the *Art & Architecture Thesaurus*^[26] has a set of rules for when to split compound terms, with an emphasis on the needs in cataloguing objects in museums and galleries. One such rule is to split a compound if the difference is a style or period term, thus the concept "baroque gilding" would be represented by the combination "baroque" + "gilding".

7.7 Order of words in multi-word terms

Preferred terms that are either adjectival or prepositional noun phrases should be entered in the thesaurus in natural language order, not as inverted terms. However, the inverted form of a prepositional phrase may be entered as a non-preferred term. This is especially useful if the thesaurus is used in printed form; it might not be necessary for an electronic thesaurus that can be searched for any word embedded in a term.

EXAMPLES

<i>matter, states of</i>	USE	states of matter
<i>prey, birds of</i>	USE	birds of prey

For adjectival phrases (for example "brown bread" or "double beds") an inverted entry might not be necessary because the focal noun is usually present as a broader term (see 10.2). The entry for the broader term (for example "bread" or "beds") provides an entry point from which all its narrower terms can be found.

8 The equivalence relationship, in a monolingual context

8.1 General

The equivalence relationship is the relationship between a preferred term and its corresponding non-preferred term(s) in the same natural language. As shown in the data model in Clause 15, this relationship applies between terms rather than between concepts.

NOTE The equivalence relationship between terms representing the same concept in different natural languages, often referred to as cross-language equivalence, is discussed in Clause 9.

The relationship is reciprocal, and reciprocity between preferred and non-preferred terms should be expressed by the following conventions:

USE, written as a prefix to the preferred term;

UF (use for or used for), written as a prefix to the non-preferred term.