

Structured vocabularies for information retrieval — Guide —

Part 2: Thesauri

ICS 01.140.20

Committees responsible for this British Standard

The preparation of this British Standard was entrusted by Technical Committee IDT/2, Information and documentation, to Subcommittee IDT/2/2, Indexes, filing and thesauri, upon which the following bodies were represented:

- British Association for Information and Library Education and Research
- CAB International
- CILIP — Chartered Institute of Library and Information Professionals
- INSPEC
- Society of Indexers
- Stationery Office
- Co-opted members

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 1 November 2005

© BSI 1 November 2005

First published November 2005

Amendments issued since publication

Amd. No.	Date	Comments

The following BSI references relate to the work on this British Standard:
Committee reference IDT/2/2
Draft for comment 04/30094113

ISBN 0 580 46799 6

Contents

	Page
Committees responsible	Inside front cover
Foreword	ii
Introduction	1
1 Scope	1
2 Normative references	1
3 Definitions	1
4 Symbols, abbreviations and other conventions	1
5 Vocabulary control in a thesaurus	2
6 Thesaurus terms: scope, form and selection	2
7 Complex concepts	14
8 Relationships between terms	19
9 Facet analysis	31
10 Presentation and layout	33
11 Thesaurus functions in electronic systems	42
12 Management aspects of thesaurus construction	43
13 Updating	50
14 Requirements of thesaurus management software	52
Bibliography	56
Index (BS 8723-1 and -2)	57
Figure 1 — Paradigmatic and syntagmatic relationships	19
Figure 2 — Relationship that satisfies the “all-and-some test”	24
Figure 3 — Relationship that does not satisfy the “all-and-some test”	24
Figure 4 — Partially expanded classified display of an “industries” class, with node labels indicating a) changes of facet, for example (<i>products</i>) and b) characteristics of division of arrays, for example (<i>milk by source animal</i>)	32
Figure 5 — Alphabetical display of thesaurus terms	35
Figure 6 — Hierarchical display of thesaurus terms arranged by facets	37
Figure 7 — Classified display of thesaurus terms arranged by facets within subject fields	39
Figure 8 — Classified display of thesaurus terms arranged by facets within subject fields, with additional information such as SN, UF and RT	41
Table 1 — Attributes and relationships of a preferred term	46
Table 2 — Attributes and relationships of a non-preferred term	46
Table 3 — Attributes of a node label	47

Foreword

This part of BS 8723 has been prepared by Subcommittee IDT/2/2. Together with BS 8723-1 it supersedes BS 5723:1987, which is withdrawn. When published BS 8723-4 will revise and supersede BS 6723:1985 which will then be withdrawn. The other parts of BS 8723 will cover new scope, not previously published in a British Standard.

The other parts of BS 8723 are as follows:

- *Part 1: Definitions, symbols and abbreviations;*
- *Part 3: Vocabularies other than thesauri¹⁾;*
- *Part 4: Interoperation between multiple vocabularies¹⁾;*
- *Part 5: Interoperation between vocabularies and other components of information storage and retrieval systems¹⁾.*

The index to this standard conforms to BS ISO 999 and BS 0. Location is to numbered clause within the individual part. Part numbers in arabic numerals in square brackets precede the first locator to the part, e.g. [2]12.4.5. A single letter following a numerical locator refers to a numbered bullet point within that locator, e.g. [2]8.3.1b).

Part 1 covers definitions, symbols, abbreviations and other conventions applying to all the parts.

Part 3 covers other types of structured vocabulary, including those for applications such as automatic categorization schemes and free-text search aids.

Part 4 applies to situations in which more than one language or vocabulary is in use, but access to all resources is needed through the one vocabulary chosen by the user.

Part 5 sets out the protocols and formats needed for the exchange of vocabulary data.

BS 8723-2 covers all of the scope of BS 5723, and extends it. In particular, more attention is given to presentation of thesauri using electronic rather than printed media. Recommendations are included on the functionality of software to manage thesaurus data.

Part 1 together with part 2 of BS 8723 corresponds broadly to ISO 2788:1986, since this is equivalent to the withdrawn BS 5723. ISO 2788:1986 remains current as an international standard. Part 1 together with part 4 will correspond to ISO 5964 (equivalent to BS 6723:1985), while extending the scope of the latter considerably. ISO 5964:1985 will remain current as an international standard

BS 8723-2 covers development and maintenance of thesauri rather than how to use them in indexing, which is covered by BS ISO 999.

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

Compliance with a British Standard does not of itself confer immunity from legal obligations.

¹⁾ In preparation.

Summary of pages

This document comprises a front cover, an inside front cover, pages i and ii, pages 1 to 59, and a back cover.

The BSI copyright notice displayed in this document indicates when the document was last issued.

Introduction

Today's thesauri are mostly electronic tools, having moved on from the paper-based era when thesaurus standards were first developed. They are built and maintained with the support of software and need to integrate with other software, such as search engines and content management systems. Whereas in the past thesauri were designed for information professionals trained in indexing and searching, today there is a demand for vocabularies that untrained users will find to be intuitive.

BS 8723 makes the transition needed for the world of electronic information management. However, part 2 retains the assumption that human intellect is involved in the selection of indexing terms and in the selection of search terms. If both the indexer and the searcher are guided to choose the same term for the same concept, then relevant documents will be retrieved. This is the main principle underlying thesaurus design.

Thesauri are typically used in post-coordinate retrieval systems, but may also be applied to hierarchical directories, pre-coordinate indexes and classification systems. Increasingly, thesaurus applications need to mesh with others, such as automatic categorization schemes, free-text search systems, etc. Other parts of BS 8723 describe additional types of structured vocabulary and give recommendations to enable interoperability of the vocabularies at all stages of the information storage and retrieval process.

1 Scope

This part of BS 8723 gives recommendations for the development and maintenance of thesauri intended for information retrieval applications. It applies to vocabularies used for retrieving information from text resources, such as knowledge bases, bibliographic databases or collections of full text documents, and also catalogues of images or artefacts, such as those in a museum collection.

This part of this British Standard applies to multilingual and monolingual thesauri, but does not cover recommendations specific to multilingual situations, which are given in BS 8723-4.

The BS 8723 series does not apply to the preparation of back-of-the-book indexes, although many of its recommendations may be useful for that purpose. It does not cover the structure of authority files of proper names of people, organizations and places, while recognizing these may sometimes interact with vocabularies of subject terms (see also 6.1.1.3).

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.

BS 8723-1:2005, *Structured vocabularies for information retrieval — Guide — Part 1: Definitions, symbols and abbreviations*.

BS EN ISO 28601 (ISO 8601), *Data elements and interchange formats — Information interchange — Representation of dates and times*.

3 Definitions

For the purposes of this British Standard, the definitions given in BS 8723-1 apply.

4 Symbols, abbreviations and other conventions

For the purposes of this British Standard the symbols, abbreviations and conventions given in BS 8723-1:2005, Clause 3 apply.

5 Vocabulary control in a thesaurus

5.1 The fundamental aim of a thesaurus is to guide the indexer and the searcher to choose the same term for the same concept. In order to achieve this, a thesaurus should firstly list all the concepts that might be useful for retrieval purposes in a given domain. The concepts are represented by terms, called preferred terms. Secondly, a thesaurus should present the preferred terms in such a way that people will easily identify the one(s) they need. This is achieved by establishing relationships between the terms and using the relationships to present the preferred terms in a structured display.

For reasons of practicality this standard refers frequently to “terms”. But it should never be forgotten that the purpose of manipulating terms is to support retrieval of the underlying concepts.

5.2 Establishing an appropriate preferred term to represent a particular concept is not always straightforward because one concept can often be expressed in more than one way. Furthermore, in ordinary discourse one term may have more than one meaning. Vocabulary control is therefore essential, and thesauri are used to achieve this by the following two principal means.

a) Terms are deliberately restricted in scope to selected meanings. Unlike the terms in a dictionary, which may be accompanied by a number of different definitions reflecting common usage, each term in a thesaurus is generally restricted to whichever single meaning serves the needs of a retrieval system most effectively. The structure of a thesaurus, notably its display of hierarchical relationships, frequently indicates the intended meaning of a term. If this technique is not sufficiently explicit, a scope note should accompany the term. This note should state the chosen meaning, and it may also indicate other meanings which are recognized in natural language but which have been deliberately excluded for indexing purposes.

b) When the same concept can be expressed by two or more synonyms, one of these terms is usually selected as the preferred term, which is then used consistently in indexing. Reference to the preferred term should be made from any synonym which might also function as a user’s access point.

One consequence of using the measures in a) and b) for vocabulary control is that the resultant language may not correspond to a user’s preferences. The thesaurus has an important role in mediating between terms used in discourse and those that function effectively for information retrieval. To achieve the retrieval benefits, users need to accept a degree of artificiality in the controlled vocabulary.

6 Thesaurus terms: scope, form and selection

6.1 Scope of preferred terms

6.1.1 *Conceptual basis*

6.1.1.1 Each preferred term included in a thesaurus should represent a single concept (or unit of thought). A concept may be expressed by a single-word term or a multi-word term.

6.1.1.2 The concepts represented by preferred terms may be considered as belonging to mutually exclusive categories based on shared characteristics. The following examples are indicative but do not list all the possible categories.

a) Things and their physical parts

EXAMPLES

birds
documents
limbs
microforms
monuments
mountain regions

NOTE Physical parts are also things, and can have their own parts.

b) Materials

EXAMPLES

adhesives
rubber
titanium

c) Activities or processes

EXAMPLES

dressmaking
fertilization
glaciation

d) Events or occurrences

EXAMPLES

birthdays
civil wars
revolutions

e) Properties of persons, things, materials or actions

EXAMPLES

consciousness
elasticity
personality
speed

f) Disciplines or subject fields

EXAMPLES

archaeology
organic chemistry
theology

g) Units of measurement

EXAMPLES

hertz
kilometres

h) Types of people and organizations

EXAMPLES

charities
children
international nongovernmental organizations
nations
poets
visually impaired people

6.1.1.3 Unique entities expressed as proper names may also be included in a subject thesaurus, where they should be used only for records that are about the person, document or object etc. When the entity represented by the proper name is not the subject of a document, but has a different relationship, such as being its creator or title, the proper name should not be recorded together with subject terms, but elsewhere.

Proper names used as subjects also belong to categories such as those listed in **6.1.1.2**.

a) Places

EXAMPLES

Australia
Milky Way
South Kensington
Sri Lanka

b) Specific objects, topographical features and other entities

EXAMPLES

Magna Carta
Mona Lisa
Nelson's Column
Romeo and Juliet
Skylab

c) Individuals, organizational posts and corporate bodies

EXAMPLES

Burns, Robert
United Nations. Secretary-General
World Health Organization

6.1.2 Clarification and disambiguation of preferred terms

6.1.2.1 General

The scope of each preferred term is limited to one meaning within the domain of the thesaurus. This meaning is not always the most common meaning associated with the term in ordinary discourse. In the thesaurus, context is usually provided by the hierarchy of broader and narrower terms linked to the term in question and this helps to elucidate the intended scope. Furthermore the preferred term itself should be formulated in such a way that it conveys the intended scope to any user, avoiding ambiguities wherever possible. A qualifier (see **6.1.2.2**) is often sufficient for disambiguation purposes. However, in cases where these measures are not appropriate or sufficient, or where additional information would help to clarify the meaning and make the usage more consistent, an explicit scope note should be used.

6.1.2.2 Homographs and qualifiers

Homographs (sometimes referred to by the broader term "homonyms") are words with the same spelling but different meanings.

EXAMPLE 1

cranes, which can refer to either birds or lifting equipment

When homographs are needed as thesaurus terms, the meaning of each term should be clarified by adding to it a qualifier in parentheses. The qualifier should be as brief as possible, ideally consisting of one word, and it forms part of the term. The qualifier does not serve as a scope note (see **6.1.2.3**).

EXAMPLE 2

cranes (birds)
cranes (lifting equipment)

A qualifier should be added to each homographic term, even when one of its senses is more common than the other in the domain of the thesaurus. For example, **beams (structures)** is an appropriate preferred term in an engineering thesaurus that also includes **beams (radiation)**.

Since the qualifiers make the terms slightly cumbersome to apply, their use in preferred terms should be avoided if another means of avoiding ambiguity can be found.

Parenthetical qualifiers should not be used to provide a preferred term in inverted form, for example **cookery (fish)** and **pipes (concrete)** are incorrect. **Concrete** is used in the latter example to indicate a type of pipe rather than to disambiguate the word “pipe” (see also 7.6 on the order of words in compound terms). Appropriate uses of qualifiers with the term “pipes” are: **pipes (musical instruments)** and **pipes (smoking implements)**.

6.1.2.3 Scope notes

A scope note is used to restrict or expand the meaning of a preferred term, to distinguish between preferred terms that have overlapping meanings in natural language, or to provide other advice on term usage to either the indexer or the searcher. A scope note need not be a full definition but should clarify the intended use of a term within the thesaurus. Unlike the qualifiers considered in 6.1.2.2, a scope note is not regarded as forming part of the term to which it is attached.

EXAMPLE 1

microwave frequencies
SN 1 GHz to 300 GHz

EXAMPLE 2

illuminations
SN Includes both the ornamental decoration and the illustrations in manuscripts, as well as in some early printed books, if done by hand

6.1.2.4 Reciprocal scope notes

When reference is made to other preferred terms in a scope note, a reciprocal scope note should generally be provided for each preferred term mentioned.

EXAMPLE

food products

SN Use only for products intended for human consumption. For products for animals, see **petfoods** or **feeds**

petfoods

SN Food products for animals maintained as domestic pets. For products for human use, see **food products**. For products intended for non-domestic animals, see **feeds**

feeds

SN Products intended for non-domestic animals. For products for domestic pets, see **petfoods**. For products for human consumption, see **food products**

Even where the scope of only one of the terms requires clarification, it is useful to make an editorial note (see 12.3.1) in the term record for the second preferred term, of where it has been cited. The purpose of the reciprocal reference is to ensure that when a change is made to one of the preferred terms, or it is deleted, the effect on the other preferred term is considered.

6.1.2.5 Definitions

A full dictionary definition is not usually required to clarify the way in which a preferred term should be used. However, if a definition is required for another reason, a separate note field should be established for the definitions so that they do not become confused with any scope notes. The source of each definition should be recorded alongside the definition itself.

EXAMPLE

illuminations

DEF The designs, etc., employed in the embellishment of a letter or writing with colours, etc. (OED)

NOTE “OED” refers to the Oxford English Dictionary²⁾. Any acronyms used in thesaurus definitions or scope notes should be explained in the introduction to the thesaurus. See 10.7.

²⁾ Oxford English Dictionary, www.oed.com

6.1.2.6 *History notes*

A history note should be used when a new preferred term is added to the thesaurus or a change is made to an existing preferred term that affects its scope in different periods of application. While it is possible to include such information in the scope note, a separate history note is preferable. A history note may record the date of introduction, or it may give more complex advice on how to search for the concept in earlier or later times.

EXAMPLE 1

microwave ovens

HN Introduced 1985

EXAMPLE 2

notebook computers

HN Introduced 1999; prior to that use "laptop computers"

6.2 Grammatical form of terms

6.2.1 *Nouns and noun phrases*

A preferred term should preferably consist of a noun or a noun phrase. This includes gerunds, which are verbal nouns (see 6.2.4). Noun phrases occur in two forms:

a) **adjectival phrases**

EXAMPLES

cold fusion

tropical diseases

b) **prepositional phrases**

EXAMPLES

accessories after the fact

hospitals for children

prisoners of war

As prepositions can unnecessarily add to the length and clumsiness of a term, they should be avoided if possible. For example, use **carbohydrate metabolism** rather than **metabolism of carbohydrates**.

6.2.2 *Adjectives*

Adjectives may be useful as components of noun phrases, but if used standing alone they are a potential cause of problems. For example, if a document discussing "the use of red lights as warning signals for low bridges" were indexed with the terms **red**, **lights**, **warning signals**, **low**, **bridges**, it would be retrieved by searches for **low AND lights**, or **low AND warning signals** or **red AND bridges**, etc. For this reason the use of adjectives as preferred terms should be avoided. Exceptions may be made where there are strong reasons, for example in a thesaurus for indexing images or other non-text collections, where adjectives describing the appearance may be very important.

6.2.3 *Adverbs*

Adverbs such as "very" or "highly" should not be used as thesaurus terms. A phrase beginning with an adverb should not be accepted as a preferred term unless it has acquired a special meaning.

EXAMPLES

very high frequency

very large scale integration

6.2.4 Verbs

Verbs expressed as infinitives or participles should not be used alone as thesaurus terms, unless they are expressed as gerunds (for example **weaving**, **broadcasting**). Activities should be represented by nouns or verbal nouns.

EXAMPLES

cooking (not “cook”, “to cook”, “cooked”, etc.)
distillation (not “distil”)
swimming (not “swim”)

6.2.5 Initial articles

6.2.5.1 Omission

The use of initial articles in preferred terms should be avoided. If necessary, a parenthetical qualifier should be used.

EXAMPLES

arts *rather than* the arts
state (political entity) *rather than* the state

6.2.5.2 Retention

If the initial article is an integral part of a proper name, and needs to be searchable, it should be included in the preferred term in direct order. Otherwise, the article should be omitted or the term inverted. If this causes ambiguity, it may be necessary to add a qualifier. Whether or not the article is considered an integral part of the name depends on the language and the context. In the following examples, English is the assumed language of the thesaurus.

EXAMPLES

El Niño
Le Havre
Los Angeles
Needles (Isle of Wight)
The Who (rock music group)

Where a term might be sought with or without the article, a reference should be made from the non-preferred form.

EXAMPLES

<i>Salvador (country)</i>	USE El Salvador
<i>The Needles</i>	USE Needles (Isle of Wight)
<i>Who, The</i>	USE The Who (rock music group)

6.3 Capitalization, punctuation and special characters

6.3.1 Capitalization

A consistent style should be used for the presentation of preferred terms and non-preferred terms. This standard uses lower case throughout, except where capitals are required for proper names. As computers are no longer restricted to using upper case only, either lower case or lower case with initial capitals should be used. Exceptions may be made for abbreviations, acronyms, proper names, or terms that are conventionally written in a special style. In these cases, the style that is most widely accepted among the anticipated users of the thesaurus should be used.

EXAMPLES

British Airways Plc
ActiveX
DNA
NPK fertilizers
pH
photocopies

6.3.2 Non-alphabetic characters

The use of punctuation marks, diacritics and other special characters can cause problems for entering terms, sorting and processing in search applications. Apostrophes, parentheses and numerical characters can complicate a search expression. The use of such characters should be minimized. However, as their use cannot be avoided completely, they should be retained wherever the terminology would otherwise be ambiguous, ungrammatical or unacceptable to the user community of the thesaurus.

EXAMPLES (avoidance)

beta rays	<i>Not</i> β -rays
databases	<i>Not</i> data-bases
nonfiction	<i>Not</i> non-fiction
research and development	<i>Not</i> research & development

EXAMPLES (retention)

2,4-D
bis(tributyltin) oxide
Boyle's law
Burkitt's lymphoma
X-rays

6.4 Singular or plural forms

6.4.1 Cultural factors

Different traditions exist in different languages concerning the use of singulars or plurals. Indexers in some language communities, for example French and German, tend to prefer the singular form so that the user can approach the thesaurus or index in the same way as a dictionary. In English-speaking countries, however, it is usual to base the choice on whether a particular term is a count noun or a non-count noun. The latter convention helps to distinguish between a process such as **painting**, which can only be expressed in the singular, and the product of the same process, in this case **paintings**.

6.4.2 Count nouns

Count nouns are names of countable entities that are subject to the question "How many?" but not "How much?". They should be expressed as plurals.

EXAMPLES (of count nouns)

documents
penguins
political parties
windows

One exception is the names of parts of the body, which are usually expressed in the singular.

EXAMPLES (of exceptions)

digestive system
eye
head

Another exception is the names of organisms. Many species, such as *Escherichia coli* or *Euonymus fortunei*, do not have a common name and are known only by their scientific name, which by convention is expressed in Latin, in the singular. In the interest of consistency, when Latin names and names accepted as English coexist in one compilation, the singular may be applied throughout. Use of the singular is not mandatory, however, and thesaurus editors may prefer the plural, for example zebras, whales or daffodils. Once a convention has been chosen, it should be applied consistently.

When a controlled vocabulary is used for naming museum objects, the singular forms of terms are commonly used. In general, however, concepts represent categories of objects rather than individual objects, and the plural form is more appropriate. A single item might be named a **chair**, but the *category* likely to be sought by an enquirer is **chairs**. Using the plural form of preferred term in these cases can add consistency so that the same thesaurus may be used for museum objects as for other types of information resource.

6.4.3 *Non-count nouns*

Non-count nouns are names of concepts such as materials or substances which are subject to the question “How much?” but not “How many?”. They should usually be expressed as singulars.

EXAMPLES (1)

paint
quartz
steam

However, if the community of users served by the thesaurus regards a given substance or material as a class with more than one member, the class should be expressed in the plural.

EXAMPLES (2)

grasses
poisons
steels

The names of abstract phenomena, properties, systems of belief, activities and disciplines, are often non-count nouns that should be expressed in their singular forms.

EXAMPLES (3)

Abstract phenomena: personality; winter
Properties: brittleness; opacity; solubility
Systems of belief: Catholicism; Shintoism; communism
Activities or processes: cutting; immigration; shrinkage
Disciplines: astronomy; sociology

However, when an abstract concept is regarded as a class with more than one member, the term representing the class should be expressed in the plural.

EXAMPLES (4)

chemical reactions
intelligence tests
physical sciences

6.4.4 *Coexistence of singular and plural*

Where the singular and plural forms of a term refer to different concepts both should be entered in the thesaurus. The distinction between them should be reinforced by adding a scope note and if possible a qualifying term or phrase.

EXAMPLES

wood (material)
woods (areas of woodland)

The added qualifier becomes an integral part of the term; it does not constitute a scope note (see also 6.1.2.2).

Where singular and plural denote the same concept, but their spelling differs to such an extent that the terms would be separated by unrelated terms when filed alphabetically, a reference should be made from the non-preferred form.

EXAMPLE

mouse **USE** mice

6.5 Selection of the preferred form

6.5.1 *General*

Where there is a choice between synonymous forms of expression, the preferences of the community to be served by the thesaurus should be adopted (subject to the avoidance of ambiguity or language that might be offensive to some users).

6.5.2 Spellings

The most widely accepted spelling of words should be adopted. If variant spellings exist and are commonly recognized, each should be entered in the thesaurus, and a reference should be made from the non-preferred to the preferred form.

EXAMPLES (1)

Roumania **USE** Romania
Rumania **USE** Romania

Where possible, spelling should follow the practice of a well-established dictionary or glossary. If a choice between spellings is made for reasons of dialect (for example between American English and British English), the chosen source should be stated in the introduction to the thesaurus, and the choice should be adhered to consistently throughout. However, proper names may be exceptions, as exemplified below.

EXAMPLES (2)

colour **UF** *color*
defence **UF** *defense*
Department of Defense (a US government department)
Ministry of Defence (a UK government department)

Misspelt words are not acceptable in preferred terms. However, some misspellings are so common that they can provide useful entry points, especially in electronic thesauri.

EXAMPLES (3)

abattoirs
UF *abatoirs*
 abbatoirs
 abbattoirs

Misspelt entry points should be provided only where:

- a) the misspelling is common enough to be worthwhile; and
- b) there is no danger that the misspelling relates to a term other than the preferred term indicated.

The entries should be marked in some way to clarify that the terms are misspelt. One option is to substitute MS (meaning MisSpelling) for UF as the introductory tag.

When misspelt words are entered as non-preferred terms, a choice may be made to suppress the misspelt entries in printed versions, where they add nothing to a user's understanding of the concept, but retain them in electronic versions used to interpret search terms entered by users.

6.5.3 Loan terms and translations of loan terms

Loan terms may be used as preferred terms if they are well-established.

EXAMPLES (1)

bouquets
gestalt therapy
ombudsmen

Occasionally a loan term and a putative translation coexist. If the loan term is more widely accepted, it should be treated as the preferred term, but the translation may be preferred if it becomes well-established. Reciprocal references should be made between the two terms.

EXAMPLES (2)

coiffeurs
USE hairdressers
hairdressers
UF *coiffeurs*
abattoirs
UF *slaughterhouses*
slaughterhouses
USE abattoirs

6.5.4 Transliteration

When transliterating terms from languages with different alphabets, a recognized scheme should be followed. The scheme used should be specified in the introduction to the thesaurus. For details of some relevant schemes see the standards publications listed in the Bibliography and *ALA-LC romanization tables: transliteration schemes for non-roman scripts* [1].

6.5.5 Neologisms, slang terms and jargon

The established names of concepts should be used instead of slang or jargon terms, but slang or jargon entries may be needed in circumstances such as the following.

- a) A newly-emerging concept is expressed by a term originating within a particular subculture or social group, and no widely accepted alternative exists. The slang or jargon term should then be accepted as an indexing term.

EXAMPLES

hippies
hypertext

- b) Where a slang or jargon term is widely used in place of the well-established term, many users would find it a useful entry point. A lead-in entry should be provided from the slang or jargon to the preferred term;

EXAMPLE

association football
UF soccer
soccer
USE association football

6.5.6 Common names and trade names

A product is frequently known by a recognized trade name. Where a suitable common name also exists, this should be adopted as the preferred term, and the trade name should be admitted as a non-preferred term only if it is likely to serve as a user's access point. As registered trade marks enjoy legal protection, the symbol "®" should be added to the term to avoid legal problems.

EXAMPLE

soluble coffee
UF Nescafé®
Nescafé®
USE soluble coffee

6.5.7 Popular names and scientific names

If a popular and a scientific name refer to the same concept, preference should be given to the form more likely to be sought by users of the thesaurus. For example, **penguins** might be chosen as the preferred term in a general thesaurus, but the scientific equivalent, **Sphenisciformes**, may be preferred in a zoological thesaurus. Reciprocal references should be made in these cases.

6.5.8 Abbreviations and acronyms

Abbreviations and acronyms should not be used as preferred terms except when they are widely used, unambiguous and readily understood within the field covered by the thesaurus. As acronyms and abbreviations can refer to more than one concept, the full form of the name should function as the preferred term, with a reciprocal reference from the abbreviated form.

EXAMPLES (1)

World Health Organization

UF *WHO*

WHO

USE World Health Organization

direct current

UF *DC (direct current)*

DC (direct current)

USE direct current

national insurance

UF *NI (national insurance)*

NI (national insurance)

USE national insurance

Abbreviations and acronyms may function as preferred terms if they have become so well-established that the full form of the name is rarely used or is generally ignored. Reciprocal references should still be made between the full term and its abbreviation.

EXAMPLES (2)

UNICEF

UF *United Nations International Children's Emergency Fund*

United Nations International Children's Emergency Fund

USE UNICEF

HIV

UF *human immunodeficiency virus*

human immunodeficiency virus

USE HIV

6.5.9 Proper names

6.5.9.1 General

Proper names are often useful in a thesaurus but are sometimes excluded if there are too many candidates, especially if their forms are controlled by another authority list, or by a set of rules such as the *Anglo-American cataloguing rules* [2]. However, they should be included if the thesaurus is the only means of validating index terms. Inclusion has the additional benefit that hierarchical or associative links (see Clause 8) can be established between subject terms and relevant proper names, for example when the latter represent instances of the former.

EXAMPLE

mountains

NT Ben Nevis

Mount Everest

6.5.9.2 *Place names*

Names of countries and geographical regions sometimes vary from language to language. Variant terms referring to the same place also occur within a single language community for reasons such as the following:

- a) a “popular” and an “official” name are both in common use;

EXAMPLE 1

Holland
Netherlands

- b) the vernacular form of the name differs from the usual form of the name in the language of the thesaurus;

EXAMPLE 2

Leghorn
Livorno

- c) in a country with more than one official language, two forms may commonly be used;

EXAMPLE 3

Gand
Gent

The name which 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 should be consulted for the official forms; two examples of such authorities are ISO 3166 and *Geographical names and information* [3] published by the UK Foreign and Commonwealth Office. Reciprocal references should be made between the preferred and non-preferred versions.

EXAMPLE 4

Netherlands
UF *Holland*
The Netherlands
Holland
USE Netherlands
The Netherlands
USE Netherlands

6.5.9.3 *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* [2]. 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. Non-preferred term entries should be provided for alternative forms of the name that are in common use.

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 concept, for example, combines at least three distinct concepts into one that is conveniently expressed in the English language using a term made up of three words. Between the extremes are moderately complex concepts, which are very often conveyed by multi-word terms, but can sometimes be expressed using just one term.

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**.

Splitting **biodegradability** into two concepts might seem cumbersome or unsatisfactory, but when the complex concept is usually conveyed by a multi-word term, splitting 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 used 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.

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 users may find it harder to identify the right term.

For pragmatic reasons the discussion in this clause deals mostly with multi-word terms. However, similar considerations apply to many complex concepts conveyed by single-word terms. Furthermore, it should be remembered that in languages such as German, which uses very many compound words, the option of splitting single-word terms will occur much more commonly.

7.2 The nature of multi-word terms

A multi-word term is one consisting of more than one word.

EXAMPLES

road safety
thesaurus management software
birds of prey

The parts of most multi-word terms can be distinguished as follows:

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

EXAMPLES

- 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 which serve to narrow the extension of the focus and so specify one of its subclasses;

EXAMPLES

- 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 terms representing complex concepts.

EXAMPLES

Term	Focus	Difference
lawnmowers	mowers	lawn
bathrooms	rooms	bath

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 four main options should be considered for handling a proposed complex concept.

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

EXAMPLES

camping holidays
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.

EXAMPLES

camping holidays **USE** camping + holidays
thesaurus management software **USE** thesaurus management + software

- c) Reject the concept as a preferred term, but admit a non-preferred term pointing to a single broader preferred term.

EXAMPLES

camping holidays **USE** camping
thesaurus management software **USE** software

- d) Reject the complex concept, as too infrequently required, irrelevant or inappropriate to merit a preferred or non-preferred term.

7.3.2 *Factors to consider*

The decision on admission is often difficult and subjective. The thesaurus developer should consider the balance of advantages in the retrieval situations that are likely to occur. The following factors should be borne in mind.

If the concept is frequently sought, and especially if the term representing it is widely used and understood by the audience, then some provision should be made for it, probably as in option a) of 7.3.1 but options b) and c) may also be considered.

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 may 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.

On the other hand, 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 may 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 may be easier and more effective if none of the complex concepts are accepted as preferred terms and a post-coordinated search expression such as **roads + safety** is the only retrieval option.

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.

Another important factor is the expectations of the anticipated users. A term such as **precoordinate 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 **precoordination + indexes**, but to choose option c), making it a lead-in term to the broader term **indexes** (see 8.2.4).

Sometimes a preferred term for a complex concept is needed to avoid ambiguity if the component concepts could be combined in ways with 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.

7.3.3 *Circumstances in favour of splitting*

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 may not be worthwhile, and splitting is an option to consider.
- c) The complex concept should be split if the term’s 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 term’s focus represents a property, part or component of the difference.

EXAMPLES

aircraft engines
hospital floors
instrument reliability
soil acidity

However, exceptions may occur if the concept is well-defined and distinctive, for example **lamp shades** or **body temperature**. References to body temperature are likely to be significantly different from those to the temperature of any of the inanimate objects or substances whose temperature might be discussed in the same document collection.

7.3.4 *Circumstances that rule against splitting*

A complex concept term 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;

EXAMPLES

data processing
gross domestic product

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

EXAMPLE

plant food
(food + plants could represent either food for plants, or plants as food)