BRITISH STANDARD

BS 8723-2:2005

Structured vocabularies for information retrieval — Guide —

Part 2: Thesauri

ICS 01.140.20

BSI British Standards

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

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

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 Together with BS 8723-1 it supersedes BS 5723:1987, which is withdrawn. When published BS 8723-4 will revise and supersede BS 6723:1985 which will then be withdrawn.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 thesauri*[1)];

— *Part 4: Interoperation between multiple vocabularies*[1)];

— *Part 5: Interoperation between vocabularies and other components of information storage and retrieval systems*[1)].

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 interoperation 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 Dictionary[2)]. 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

cookery (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

Salvador (country) **USE** El Salvador
The Needles **USE** Needles (Isle of Wight)
Who, The **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	*Not* β-rays
databases	*Not* data-bases
nonfiction	*Not* non-fiction
research and development	*Not* 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)

c) the concept is represented by an established term which is a proper name, or incorporates a proper name;

EXAMPLES

Boolean logic
United Nations

d) the difference in an established term has lost its original meaning;

EXAMPLES

lawn tennis
deck chairs
trade winds

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

EXAMPLES

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;

EXAMPLES

fire escapes (a discussion of fire escapes might say nothing about fires or escaping)
swimming pools (a discussion of swimming pools might have little information on swimming)

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

EXAMPLES

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 **8.2.5** for more details):

rail safety
  USE railways
  + safety

7.5 Retention of constituent concepts

7.5.1 *General*

In the event that a multi-word term is accepted as a preferred term, consideration should be given to inclusion 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 Clause **8**).

7.5.2 *Parts and components*

Confusion can arise when the difference of the multi-word 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.

7.6 Consistency in the treatment of complex concepts

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* [4] 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.

EXAMPLES

matter, states of	**USE** states of matter
prey, birds of	**USE** birds of prey

For adjectival phrases (for example brown bread or double beds) an inverted entry may not be necessary, because the focal noun is usually present as a broader term (see **8.3**). The entry for the broader term (for example bread or beds) provides an entry point from which all its narrower terms may be found.

8 Relationships between terms

8.1 General

8.1.1 The purpose of establishing and displaying inter-term relationships is to guide users to choose the most appropriate term(s) for expressing a given concept. This is achieved by presenting terms that a user might consider instead of, or as well as, the term looked at first. Means for displaying relationships are described in Clause **9**.

Two kinds of relationship are commonly encountered between indexing terms.

a) Relationships which exist only because the terms are used together in the context of a particular document, called syntagmatic relationships.

For example, an indexer dealing with a work on "Computers in banks in Amsterdam" may assign three terms, banks (financial institutions), computers and Amsterdam, to the document. In a post-coordinate system the relationship between these terms is not explicitly indicated, but the document would nonetheless be retrieved if any or all of these terms were used as retrieval keys. In a pre-coordinated index an entry may be provided for the combination of three terms, with a reference to the location where the corresponding document may be found. The terms in this example are not normally associated according to common frames of reference, and their interrelationships can, therefore, be regarded as document-dependent. Inclusion of syntagmatic relationships in a thesaurus is not recommended.

b) Relationships that are valid in almost all contexts, especially when they are inherent in the definitions of the concepts which the terms represent, known as paradigmatic relationships.

Using terms from the previous example, banks (financial institutions) has an inherent relationship with the broader term financial institutions; computers has a strong association with data processing, independent of the subject of the document being indexed; and Amsterdam is inseparably linked with Netherlands. A user looking up any of these linked terms might be interested in information indexed with the former terms. These paradigmatic relationships are independent of any particular document. They are generally recognized and could be established through reference to standard works, such as dictionaries and encyclopaedias. It is useful to show paradigmatic relationships between thesaurus terms, as they will frequently guide users to concepts closely related to the terms they first thought of.

The distinction between these two kinds of inter-term relationships can be displayed as shown in Figure 1.

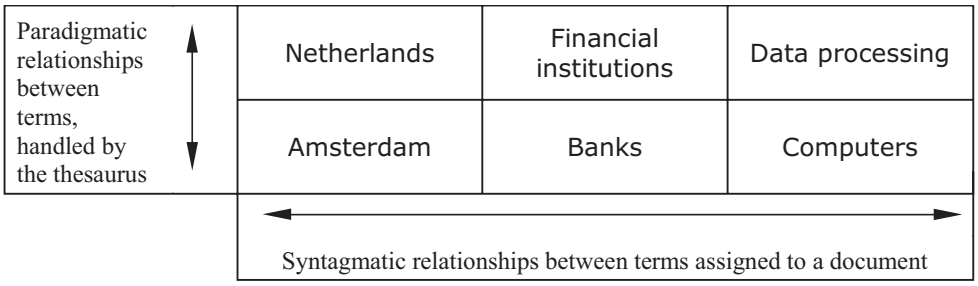


Figure 1 — Paradigmatic and syntagmatic relationships

8.1.2 Three classes of paradigmatic relationship between terms should be established and clearly distinguished:

a) the equivalence relationship (see **8.2**);

b) the hierarchical relationship (see **8.3**);

c) the associative relationship (see **8.4**).

Further subdivisions of each of these classes are described in **8.2** to **8.4**. Each of these relationships should be established reciprocally, and this should be indicated by a system of tags, symbols or abbreviations used to represent relationships in a thesaurus. The conventional tags described in BS 8723-1:2005, Clause **3** are used in the examples that follow.

8.2 The equivalence relationships

8.2.1 *General*

In the context of a monolingual thesaurus, or the terms in any one language within a multilingual thesaurus, the equivalence relationship is the relationship between a preferred term and its corresponding non-preferred term(s).

NOTE The relationship between two corresponding preferred terms in different languages is also called equivalence, but equivalence across languages is discussed in BS 8723-4.

For indexing and searching purposes, the two or more terms are regarded as referring to the same concept. Preferred and non-preferred terms should be distinguished typographically whenever possible, so that the non-preferred status of the non-preferred terms is obvious. Reciprocity is expressed by the following conventions:

USE, written as a prefix to the preferred term;

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

EXAMPLE

greenhouses
UF *glasshouses*

glasshouses
USE greenhouses

Equivalence is established in four general situations:

a) the terms are synonymous;

b) the terms are quasi-synonyms; or

c) a term is regarded as unnecessarily specific and it is represented by another term with broader scope;

d) a term is regarded as unnecessarily specific and it is represented by a combination of two or more terms (known as “compound equivalence”).

8.2.2 *Synonyms*

Synonyms occur more frequently in a controlled vocabulary, where meanings are carefully delimited, than in natural language. Various types of synonym are encountered in practice. The following list is not exhaustive, but it indicates some of the more common classes of synonyms:

a) terms of different linguistic origin;

EXAMPLES

freedom; liberty
sweat; perspiration

b) popular names and scientific names;

EXAMPLES

aspirin; acetylsalicylic acid
rock roses; Cistus

c) common nouns and trade names;

EXAMPLE

vacuum flasks; Thermos® flasks; Thermos® bottles

d) variant names for emergent concepts;

EXAMPLES

hovercraft; air cushion vehicles
laptop computers; notebook computers

e) current or favoured terms versus outdated or deprecated terms;

EXAMPLES

developing countries; underdeveloped countries
radio; wireless

f) variant spellings, including stem variants, inverted word order and irregular plurals;

EXAMPLES

geese; goose
groundwater; ground-water; ground water
paediatrics; pediatrics
radiation, ionizing; ionizing radiation; ionising radiation
Romania; Rumania; Roumania

g) terms originating from different cultures sharing a common language;

EXAMPLES

flats; apartments
lifts; elevators

h) abbreviations or acronyms and full names;

EXAMPLES

FAO; Food and Agriculture Organization
pvc; polyvinyl chloride

i) common nouns and slang or jargon terms;

EXAMPLES

psychiatrists; shrinks
soluble coffee; instant coffee

In these and similar cases, preferred terms should be selected to serve the needs of the majority of users, bearing in mind the recommendations in **6.3**. For the sake of predictability, these criteria should be applied consistently throughout the thesaurus. If, for example, it is decided that popular names rather than scientific names should serve as preferred terms, this decision should be applied consistently except when an appropriate popular name does not exist. Similarly, consistent spelling conventions should be adopted. The editorial conventions and criteria should be noted in an Introduction to the thesaurus (see **10.7**).

Optionally, it is possible to designate which type of equivalence prevails between a pair of terms, by using tags other than USE/UF. The following tags are sometimes used:

SP Spelling variant

AB Abbreviation

FT Full term

Additional tags may be devised to extend this principle to the cases described in **8.2.3** and **8.2.4**. However, customization should only be undertaken with caution (see **8.5**). All non-standard tags should be fully explained in the Introduction. (See **10.7**.)

8.2.3 *Quasi-synonyms*

Quasi-synonyms frequently represent points on a continuum, and can even be antonyms (opposites).

EXAMPLES (1)

consistency, inconsistency
wetness, dryness

In the second example, the concept wanted in the thesaurus is “the amount of moisture in an object or material”, which could be expressed in terms either of wetness or of dryness. Arbitrarily, one of the terms is chosen as the preferred term and the other as the non-preferred term.

In other cases, the concepts are so closely related that discussion of one is likely to interest a user seeking the other.

EXAMPLES (2)

bushes, shrubs
mallets, hammers

The extent to which terms are treated as quasi-synonyms depends to a large extent upon the subject field covered by the thesaurus. In a thesaurus specializing in clothing manufacture, the terms gloves and mittens might both be established as preferred terms; whereas a more general thesaurus for the textile industry as a whole might treat the two terms as quasi-synonyms. The decision should be based on anticipating the degree of discrimination required at the time of searching.

8.2.4 *Specific terms subsumed in a broader concept*

It is sometimes helpful to treat the name of a class, and also the names of its members, as an equivalence set, the broader term functioning as the preferred term.

EXAMPLE

rock
UF *basalt*
granite
slate
etc.

basalt **USE** rock

granite **USE** rock

slate **USE** rock

This technique is sometimes used to reduce the number of preferred terms in a thesaurus. It limits the specificity of indexing and searching that can be achieved, and is beneficial when the collection to be indexed has very little information on the subject area in question. The presence of the non-preferred terms provides extra entry points.

8.2.5 *Representation of complex concepts by a combination of terms*

Where a multi-word term is deemed to be unsuitable as a preferred term, but might be sought by some users, it may be represented by a combination of two or more preferred terms [see **7.3.1** option b)] and an entry of the following type may be admitted in the thesaurus:

EXAMPLES (1)

coal mining
USE coal
\+ mining

ferromagnetic films
USE ferromagnetic materials
\+ films

Reciprocals should be provided, of the following form:

coal
UF+ *coal mining*

mining
UF+ *coal mining*

The three-way relationship appears complex, but the tags **USE** and **+** are differentiated only for ease of legibility. In functional terms, the relationship between coal and *coal mining* is identical to that between mining and *coal mining*, in both directions.

In theory, combinations of terms could be linked with "OR" rather than "AND".

EXAMPLE (2)

pitch
USE audio frequency
OR gradient

However, this type of usage should be avoided, by providing entries that clarify the scope of the alternatives.

EXAMPLE (3)

pitch (sound)
USE audio frequency

pitch (steepness)
USE gradient

8.3 The hierarchical relationship

8.3.1 *General*

The hierarchical relationship should be established between a pair of terms when the scope of one of them falls completely within the scope of the other. It should be based on degrees or levels of superordination and subordination, where the superordinate term represents a class or whole, and subordinate terms refer to its members or parts. The following tags are used, reciprocally:

BT (i.e. broader term), written as a prefix to the superordinate term;

NT (i.e. narrower term), written as a prefix to the subordinate term.

EXAMPLE 1

animals
NT mammals

mammals
BT animals

The hierarchical relationship covers three logically different situations distinguished as follows:

a) the generic relationship;

b) the hierarchical whole-part relationship;

c) the instance relationship.

Each of these leads to hierarchies which are amenable to a logical test through reference to the basic types of concept represented by the terms, for example those listed in **6.1.1.2**. Every subordinate term should refer to the same basic kind of concept as its superordinate term, i.e. both the broader and narrower term should represent a thing, or an action, or a property, etc. (See additional discussion of facet analysis in Clause **9**.)

EXAMPLE 2

a) "metals" (a class of materials) and "casting" (an action) represent different types of concept, and, therefore, cannot be related hierarchically;

b) "metals" and "gold" both represent materials, and could, therefore, be hierarchically related.

The main function of hierarchical relationships is to help users choose the correct level of specificity. A search may be broadened or narrowed by moving up or down, respectively, in the hierarchy. A common technique for performing exhaustive searches is called "search explosion", in which the search is extended to include all of the narrower terms of the preferred term first selected (see also Clause **11**). Careful adherence to the guidelines in **8.3.2** to **8.3.5** ensures that exploded searches retrieve only items that fall within the scope of the superordinate concept.

8.3.2 *The generic relationship*

8.3.2.1 The generic relationship identifies the link between a class or category and its members or species. In addition to the test for validity described in **8.3.1**, this relationship is also amenable to a logical "all-and-some test", as shown in Figure 2.

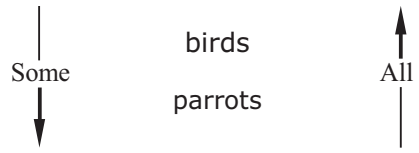


Figure 2 — Relationship that satisfies the "all-and-some test"

Figure 2 indicates that some members of the class "birds" are known as "parrots", and all "parrots", by definition and irrespective of context, are regarded as "birds". This test usually ensures that a term such as "parrots" is not subordinated to a class such as "pets", as not all parrots are pets. Figure 3 shows the relationship between the latter pair of terms.

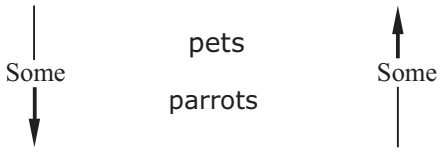


Figure 3 — Relationship that does not satisfy the "all-and-some test"

In Figure 3, some members of the class "pets" are "parrots", and only some parrots are regarded as pets. These terms should, therefore, not be given a BT/NT linkage. When indexing a work on "parrots as pets", both preferred terms should be assigned.

8.3.2.2 This argument might not apply in the context of a specialist thesaurus devoted to domestic animals, in which the only parrots in the frame of reference are pets. In such a case, "parrots" could be subordinated to "pets" in the same hierarchy. Such approximations should be applied with caution, however, especially in networked environments in which the records from one system may be mixed with those of another. When interoperability is required, the relationships established should be universally acceptable.

8.3.2.3 The tags BT/NT are normally adequate to identify the generic relationship, but optionally, the following tags may be used:

BTG: Broader term (generic)

NTG: Narrower term (generic)

EXAMPLE

rats
BTG rodents

rodents
NTG rats

8.3.3 *The hierarchical whole-part relationships*

8.3.3.1 The hierarchical whole-part relationship covers a limited range of situations in which the name of a part implies the name of its possessing whole. This applies to four main classes of terms:

a) systems and organs of the body;

EXAMPLE

cardiovascular system
  blood vessels
    arteries
    veins

b) geographical locations;

EXAMPLE

Canada
  Ontario
    Ottawa
    Toronto

c) disciplines or fields of discourse;

EXAMPLE

science
  biology
    botany
    zoology

d) hierarchical social structures;

EXAMPLE

armies
  corps
    divisions (army)
      battalions
        regiments

Most other cases of the whole-part relationship are not eligible for a hierarchical linkage because the part could belong to more than one whole. For example, a BT/NT relationship should not be established between bicycles and wheels because a wheel may be part of a motor car or a wheelbarrow or many other artefacts. An exploded search for bicycles would retrieve much unwanted material if it were extended to all types of wheel. It is sometimes the case, however, that the parts of an artefact are unique to that artefact, at least in the field of application of the thesaurus. It might be appropriate to establish a BT/NT link between fireplaces and hearths, or between bows and bowstrings, for example. This would usually cause no confusion and would help with exploded searches under the broader term. This practice is not recommended, however, in the case of complex machines and their components, for which a more appropriate solution is to create a broader term such as engine components, with the several components listed as narrower terms (see also **7.5.2**).

8.3.3.2 The tags BT/NT are normally adequate to identify the partitive relationship, but optionally, the following tags may be used:

BTP = broader term (partitive)

NTP = narrower term (partitive)

EXAMPLE

central nervous system
BTP nervous system

nervous system
NTP central nervous system

8.3.4 *The instance relationship*

8.3.4.1 The instance relationship identifies the link between a general class of things or events, expressed by a common noun, and an individual instance of that class, often represented by a proper name.

EXAMPLE

mountain regions — class

Alps, Himalayas } instances

In this example, the "Alps" and "Himalayas" are assigned to subordinate positions in a hierarchy. However, they are neither kinds nor parts of "mountain regions", but represent individual instances.

8.3.4.2 The tags BT/NT are normally adequate to identify the instance relationship, but optionally, the following tags may be used:

BTI = broader term (instantial)

NTI = narrower term (instantial)

EXAMPLES

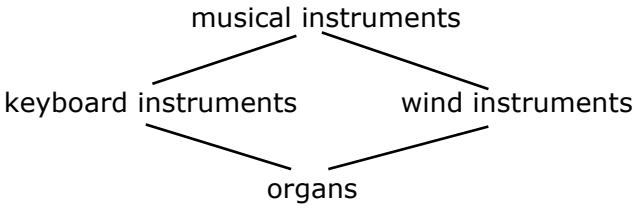
Paris
BTI capital cities

capital cities
NTI Paris

8.3.5 *Polyhierarchical relationships*

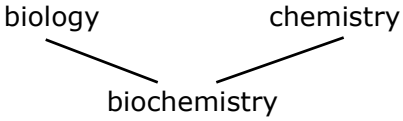
Some concepts can belong, on logical grounds, to more than one category at the same time, in which case they are said to possess polyhierarchical relationships.

EXAMPLE 1



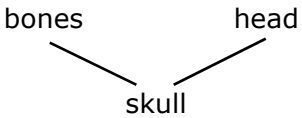
In this example, the term "organs" is assigned to subordinate positions on the basis of its generic relationship to two broader terms. In other cases, polyhierarchical links may be based upon whole-part relationships.

EXAMPLE 2



In some cases, polyhierarchical links can be based upon logically different relationships.

EXAMPLE 3



In this example, the link between "bones" and "skull" is based upon the generic relationship (the skull is a kind of bone), whereas the link between "head" and "skull" is based upon the hierarchical whole-part relationship (the skull is a part of the head).

This type of situation causes no problems in the thesaurus display, as two or more BT relationships may be assigned to one preferred term.

EXAMPLE 4

skull
BT bones
head

8.4 The associative relationship

8.4.1 *General*

The associative relationship covers associations between pairs of preferred terms which are not related hierarchically or by equivalence, but the terms are semantically or conceptually associated to such an extent that the link between them needs to be made explicit in the thesaurus, on the grounds that it may suggest additional or alternative terms for use in indexing or retrieval. The relationship is indicated by the tag "**RT**" (related term) and it is always used reciprocally.

EXAMPLE

birds
RT ornithology

ornithology
RT birds

A general guideline for associative relationships is that whenever one term is used, the other should always be implied within the common frames of reference shared by the users of the thesaurus. Moreover, one of the terms is often a necessary component in any explanation or definition of the other; the term birds, for example, forms a necessary part of the explanation of ornithology.

It is particularly important to establish an associative relationship between preferred terms whose meanings overlap.

8.4.2 *Preferred terms with overlapping meanings*

8.4.2.1 A pair of terms can often be used interchangeably in some contexts but not in others. For example, the terms ships and boats may both be established as preferred terms, because while they have much in common, neither of them adequately covers the scope of the other. It is essential that an associative relationship between the two terms is provided to remind searchers that they may need to try both preferred terms to be sure of finding all the relevant material.

EXAMPLE

boats
RT ships

ships
RT boats

8.4.2.2 A pair of preferred terms with overlapping meanings may often be siblings, i.e. they share a common broader term. However, it is not necessary to interrelate all sibling terms in this way. For example, there is no need to associate terms such as "horses" and "donkeys" on the grounds that they share a common broader term, i.e. "equines", since the meanings of the terms do not overlap in this case.

8.4.3 *Other cases of associative linkage*

Many grounds can be established for associating terms meeting the general requirement that one of the terms should be strongly implied by the other. The following groups are offered only as representative examples of typical relational situations encountered in practice:

a) a discipline or field of study and the objects or phenomena studied;

EXAMPLES

1) forestry
RT forests

forests
RT forestry

2) neurology
RT nervous system

nervous system
RT neurology

b) an operation or process and its agent or instrument;

EXAMPLES

1) temperature control
RT thermostats

thermostats
RT temperature control

2) crime investigation
RT detectives

detectives
RT crime investigation

c) an action and the product of the action;

EXAMPLES

1) weaving
RT cloth

cloth
RT weaving

2) ploughing
RT furrows

furrows
RT ploughing

d) an action and its patient or target;

EXAMPLES

1) harvesting
RT crops

crops
RT harvesting

2) imprisonment
RT prisoners

prisoners
RT imprisonment

e) objects or materials and their defining properties;

EXAMPLES

1) poisons
RT toxicity

toxicity
RT poisons

2) magnets
RT ferromagnetism

ferromagnetism
RT magnets

f) an artefact and its parts, if they do not qualify for the hierarchical whole-part relationship;

EXAMPLE

optical instruments
RT lenses

lenses
RT optical instruments

g) concepts linked by causal dependence;

EXAMPLES

1) bereavement
RT death

death
RT bereavement

2) diseases
RT pathogens

pathogens
RT diseases

h) an object or process and its counter agent;

EXAMPLES

1) plants
RT herbicides

herbicides
RT plants

2) inflammation
RT anti-inflammatory agents

anti-inflammatory agents
RT inflammation

i) a concept and its unit of measurement;

EXAMPLE

electric current
RT amperes

amperes
RT electric current

j) a multi-word term and the noun which is its focus, if the two do not have a true hierarchical relationship;

EXAMPLES

1) model ships
RT ships

ships
RT model ships

2) fossil reptiles
RT reptiles

reptiles
RT fossil reptiles

k) an organism or substance bred or derived from another;

EXAMPLES

1) mules
RT donkeys

donkeys
RT mules

2) brass
RT copper

copper
RT brass

8.5 Customized relationships

The equivalence, hierarchical and associative relationship types described in **8.2**, **8.3** and **8.4** are well established, widely used, and generally found to be adequate in contexts in which a human operator views a display of broader, narrower or related terms before choosing how to express a search query. However, there may sometimes be a need for additional or more closely specified relationship types.

In the same way that the hierarchical relationship may optionally be subdivided into generic, partitive or instantial, the equivalence and associative relationships may optionally be subdivided for a particular application. For example, abbreviations and/or acronyms may be designated separately from all the other equivalence relationships. Tags distinct from USE/UF should be assigned, for example FT/AB for Full Term and Abbreviation. Similarly, for the associative relationship, it is possible to subdivide into types such as cause and effect, designated with tags such as CAUSE/EFFECT.

Before embarking on an exercise to provide more specific relationship types, the thesaurus developer should first check that the refined relationships are really necessary, and not simply an intellectually attractive construct. The purpose of the thesaurus is to serve a set of users, in a given context. The benefits of the extra relationships could be outweighed by the extra complications perceived by the users. However, there may be other justifications for the extra work, if the thesaurus is wanted to interoperate with or function as an ontology. (Ontologies usually provide more specific and closely defined relationships.)

NOTE See BS 8723-3 and -4 for recommendations on these issues.

In the event that customization is undertaken, care should be taken to ensure that users understand the practice, and that confusion does not arise when the customized thesaurus is handled together with a conventional one. The introduction of non-standard tags or symbols carries the risk of incompatibility in circumstances in which the thesaurus in question interoperates with others. The risks may be reduced by ensuring that the introduced relationships are true subdivisions of the established relationship types.

9 Facet analysis

Facet analysis is useful in generating hierarchies that conform to the rules for hierarchical relationships as set out in **8.3**, because these relationships are valid only for terms belonging to the same general category.

The choice of facets may vary depending on the subject field, but at the highest levels it is usual to use fundamental categories such as **objects**, **materials**, **agents**, **actions**, **places**, **times**, etc. These fundamental facets may be analysed into subfacets where it is helpful to do so, down to the level required; for example, actions may be subdivided into intransitive[3] ***processes*** such as ripening or deterioration and transitive ***operations*** such as cutting or repairing.

An example of applying facet analysis in a classified arrangement appears in Figure 4. This shows some terms from the general subject of industries. Directly under this term are the two narrower terms agricultural industries and engineering industries. One of the facets shown under agricultural industries is ***products***. Within this facet, only two levels of hierarchy are shown, except for the term milk, which has been expanded to show how its narrower terms have been grouped into arrays, each preceded by a node label showing the characteristic of division. These arrays each represent a different way of subdividing the concept of milk, by fat content, by source animal, etc., and the word "by" occurs in each of these node labels. The node labels that introduce new facets, such as (***people***) and (***products***) do not contain the word "by".

It is important to note the pattern of hierarchical relationships around the node labels. Where the label shows the characteristic of division of the superordinate term, all the terms in the array following it are true narrower terms of the superordinate term. Thus, in Figure 4, whole milk, buffalo milk and sterilized milk are all narrower terms of milk. In contrast, where the label introduces a new facet, the terms that follow are typically not narrower terms of the preceding term: farm managers and cereal products, etc., are not narrower terms of agricultural industries.

Node labels are not preferred terms or non-preferred terms. They are present only for the purposes of systematic display, and they do not qualify for any of the inter-term relationships described in Clause **8**. Node labels should be distinguished typographically from the preferred terms, to avoid confusion. Generally italics and parentheses or angle brackets are used, as shown in Figure 4.

The terms in an array may be arranged either alphabetically or systematically. Alphabetical sequence should be used when there is no other obvious way to arrange a group of concepts. Systematic sequence should be used when it is likely to be familiar to most users, or when the arrangement helps to clarify the scope of the terms. In the example for electromagnetic radiation, the types of radiation are presented in order of increasing wavelength, as this might help some indexers in selecting the correct term(s).

NOTE This example uses slightly different but equally acceptable conventions in the node label: angle brackets instead of round brackets, and the parent term is not spelt out at the start of the node label. The presence of the word "by", however, indicates that the node label specifies the characteristic of division by which the types of radiation are differentiated.

EXAMPLE

electromagnetic radiation
  <by wavelength>
  ultraviolet radiation
  visible radiation
  infrared radiation
  microwave radiation
  radio waves

[3] An action is said to be intransitive when the "actor" does not act upon any object; whereas the actor of a transitive action does act upon an object. Thus an apple ripens by an internal process, but in cutting an apple, a knife acts upon an external object.

industries

agricultural industries

(people)
farm managers
dairy personnel
shepherds

(products)
cereal products
dairy products
butter
cheese
cream
ice cream
milk
(milk by fat content)
whole milk
low fat milk
skim milk
(milk by form)
dried milk
liquid milk
(milk by source animal)
buffalo milk
cow milk
goat milk
sheep milk
(milk by treatment type)
condensed milk
evaporated milk
homogenized milk
pasteurized milk
sterilized milk

engineering industries
(people)
engineers

(products)
bolts
wheels
etc

Figure 4 — Partially expanded classified display of an "industries" class, with node labels indicating a) changes of facet, for example (*products*) and b) characteristics of division of arrays, for example (*milk by source animal*)

10 Presentation and layout

10.1 General

10.1.1 *Use of a database for maintaining presentation information*

Within a computer system a thesaurus will normally be held in a database structure, in which each individual term and each link between a pair of terms is stored once only. This maintains the consistency of the relationships and allows terms to be extracted and displayed in different ways.

When presented for human use, on a screen or in printed form, an indexing vocabulary can be displayed in the following ways:

a) a *single record* is the most elementary form of display, showing the preferred term or non-preferred term itself and (optionally) any or all of the relationships, codes and notes that attach to it;

b) an *alphabetical* arrangement allows access to concepts from the words in which they are initially expressed by the user — in a printed thesaurus it acts as an index, and in a computer display it can supplement a direct search function;

c) a *hierarchical* sequence based on BT/NT relationships helps in expanding or refining the concept being indexed or sought;

d) a *classified* sequence allows browsing in a subject area, drawing attention to related concepts;

e) a *graphical* display shows terms and their relationships laid out pictorially;

f) a *permuted* display helps to find words embedded in multi-word terms.

These different arrangements complement each other. Alphabetical access is essential and should be provided by direct search or by an alphabetical sequence as described in b). At least one of c) and d) should be provided to give an overview of the systematic arrangement. Additional information about each concept, such as scope notes and relationships to other concepts, may be shown in any of these sequences. Thesauri can differ widely in their approaches to the relative importance and functions of the different sequences, and also in the arrangement and the kind of relational information provided in each. A graphical presentation can sometimes give a useful overview of a field, but it is complex to create and update and the structure is obscured if it contains too much additional information. A permuted display serves as an index to all the words in all the terms (preferred and non-preferred), using either KWIC (key word in context) or KWOC (key word out of context) format; this is useful in a printed thesaurus, but not usually necessary in a computerized system in which searches for character strings are possible.

10.1.2 *Notation and links between sequences*

When presented in printed form, the alphabetical sequence should act as an index to show where terms are located in the other sequences, using either line numbers, a system of notation, or, if the hierarchies are not too extensive, by showing the top term of each hierarchy in which the term of interest will be found. To provide links to a graphical display from the alphabetical index it may be necessary to use a system of coordinates. Explicit linking symbols may not be necessary when a thesaurus is displayed on a computer screen, as automatically generated hyperlinks should allow easy switching from one sequence to another, while keeping the focus on the term of interest.

Notation in a classified sequence may, but need not, be expressive of the classified structure and may be designed to allow synthesis of notational symbols to create pre-coordinated class numbers for complex concepts. Any system of notation should provide for the insertion of new concepts at any point in the classified sequence.

The basic types of display are described in **10.2** to **10.6** and each is illustrated by a typical arrangement of a common set of terms related to “cameras” and related topics. In these arrangements, a simple line-number system of notation has been used.

10.2 Single record display

When the record for a preferred term is viewed individually the other fields are customarily presented after the term itself in the following sequence:

a) **CC** classification code or notation, locating the term in a classified sequence;

b) **SN** scope note;

c) **UF** references to non-preferred equivalent term(s);

d) **TT** references to top terms;

e) **BT** references to broader terms;

f) **NT** references to narrower terms;

g) **RT** references to related terms;

h) **DEF** dictionary definition of the term (see **6.1.2.5**);

i) **HN** history note (see **6.1.2.6**).

In this arrangement, the **SN** and **UF** are given near the beginning of the list because they clarify the meaning of the concept. They are followed by relationships with other terms. **DEF** and **HN** are given last because they are viewed as administrative fields, used more by the thesaurus editor than the user. Optionally, these last fields may be shown after the scope note, bringing all text fields together. However, it is important to ensure that **SN** and **DEF** are not confused when both are present.

The sequence for a non-preferred term record places its preferred equivalent(s) on the line(s) after the non-preferred term, and then any history note if present.

For particular applications the selection and sequence of fields in a) to j) may be varied. For example, the thesaurus editor may wish to include some housekeeping fields (see **12.3.1**) in a display for his/her own use.

10.3 Alphabetical display

10.3.1 In an alphabetical display all terms, whether preferred or non-preferred, are organized as a single alphabetical sequence. For some purposes a simple list of the terms alone may be sufficient. More commonly the complete record for each term is included, as in **10.2**.

10.3.2 An example of this form of display is shown in Figure 5. Apart from single record displays, this is probably the easiest type of display to construct and reproduce. The most common convention is to show immediate broader and narrower terms only, in which case space is saved but any additional levels of hierarchy are not obvious.

NOTE The colons shown in Figure 5 are optional.

35 mm cameras
CC: H012
BT: film cameras

aqualungs
CC: D002
BT: diving equipment

camera accessories
CC: H002
BT: photographic equipment
NT: flash guns
light meters
tripods
RT: cameras

camera components
CC: H006
BT: cameras and camera components
NT: camera lenses
camera viewfinders

camera lenses
CC: H007
BT: camera components

camera viewfinders
CC: H008
BT: camera components

cameras
CC: H009
BT: cameras and camera components
NT: digital cameras
film cameras
instant picture cameras
plate cameras
reflex cameras
special-purpose cameras
RT: camera accessories
photography

cameras and camera components
CC: H005
BT: photographic equipment
NT: camera components
cameras

colour
CC: A204
BT: optical properties

contrast
CC: A205
BT: optical properties

digital cameras
CC: H010
UF+: underwater digital cameras
BT: cameras

divers
CC: D008
BT: people
RT: diving

diving
CC: D000
BT: fields of work
RT: divers
diving equipment

diving equipment
CC: D001
BT: equipment
NT: aqualungs
diving suits
face masks
swimming fins
underwater cameras
RT: diving

diving suits
CC: D003
BT: diving equipment
NT: dry suits
wet suits

dry suits
CC: D004
BT: diving suits

equipment
CC: A001
NT: diving equipment
electrically-powered equipment
fixed equipment
human-powered equipment
photographic equipment
physics equipment
portable equipment

exposure meters
USE: light meters

face masks
CC: D006
BT: diving equipment

fields of work
CC: A300
NT: diving
photography
physics

film cameras
CC: H011
BT: cameras
NT: 35 mm cameras
medium format cameras
miniature cameras

instant picture cameras
CC: H015
SN: Cameras which produce a finished print directly
UF: Polaroid® cameras
BT: cameras

people
CC: A100
NT: adults
children
divers
infants
models (people)
photographers
physicists

photographic equipment
CC: H001
BT: equipment
NT: camera accessories
cameras and camera components
RT: photography

photography
CC: H000
BT: fields of work
RT: cameras
photographers
photographic equipment

physical properties
CC: A202
BT: properties
NT: pressure
temperature

physicists
CC: P005
BT: people
RT: physics

physics
CC: P000
BT: fields of work
RT: physicists

Polaroid® cameras
USE: instant picture cameras

pressure
CC: A208
BT: physical properties

properties
CC: A200
NT: physical properties

reflex cameras
CC: H017
SN: Cameras in which the image is reflected on to a glass screen for composing and focusing.
BT: cameras
NT: single lens reflex cameras
twin lens reflex cameras

single lens reflex cameras
CC: H018
UF: SLR cameras
BT: reflex cameras

SLR cameras
USE: single lens reflex cameras

special-purpose cameras
CC: H021
BT: cameras
NT: stereo cameras
underwater cameras

stereo cameras
CC: H022
BT: special-purpose cameras

swimming fins
CC: D007
BT: diving equipment

temperature
CC: A209
BT: physical properties

tripods
CC: H004
BT: camera accessories

twin lens reflex cameras
CC: H019
BT: reflex cameras

underwater cameras
CC: D028, H023
UF+: underwater digital cameras
BT: special purpose cameras
diving equipment
RT: diving

underwater digital cameras
USE: digital cameras
+ underwater cameras

Figure 5 — Alphabetical display of thesaurus terms (some terms omitted to save space)

10.3.3 Alternatively, it may be useful to indicate more than one level of subordination and/or superordination by different levels of indentation and/or by numbering the levels as shown in the example. If this convention is used to show all the levels of hierarchy for every term, it can produce a lengthy display.

EXAMPLE

Bovidae
  BT1 ruminants
    BT2 mammals
      BT3 vertebrates
  NT1 buffaloes
  NT1 cattle
    NT2 beef cattle
    NT2 dairy cattle
  NT1 yaks

10.4 Hierarchical display

A hierarchical display is constructed by indenting each array of narrower terms under its parent broader term. Other relationships are not normally shown in this form of display as they would make the structure more difficult to follow. Many hierarchies may result, each headed by a distinct top term (a term with no broader term relationships). Alternatively, if facet analysis is applied at the top level of the thesaurus, all terms from the same facet may be grouped into a single hierarchy headed by a node label containing the facet name, as shown in Figure 6.

When facet analysis is applied in this way, terms are organized into facets according to the basic kinds of concept they represent, with no initial regard for the field or fields with which a given concept is usually associated.

Comparing this approach with that described in **10.5.2**, where the primary organization is by subject and facets are applied within these, primary organization of terms into facets offers certain advantages:

a) the need for a major revision is less likely to occur if a given concept changes the field with which it is usually associated;

b) a higher level of agreement between different documentation centres could be expected.

It is also necessary, however, to note the following disadvantages.

1) Organization by facets tends to scatter the concepts usually associated with a given field or discipline.

2) The basis on which the thesaurus is organized will be generally less self-evident to both indexers and searchers.

3) If the thesaurus covers all subjects, or a broad group of subjects, bringing the entities and activities of different disciplines together is unlikely to produce a useful and intelligible sequence of terms. This type of display is, therefore, particularly useful in a thesaurus devoted to a single subject.

(objects)
equipment
. *<equipment by portability>*
. . fixed equipment
. . portable equipment
. *<equipment by power source>*
. . electrically-powered equipment
. . human-powered equipment
. *<equipment by application>*
. . diving equipment
. . . aqualungs
. . . diving suits
. . . . dry suits
. . . . wet suits
. . . face masks
. . . swimming fins
. . . underwater cameras
. . photographic equipment
. . . camera accessories
. . . . flash guns
. . . . light meters
. . . . tripods
. . . cameras and camera components
. . . . camera components
. . . . . camera lenses
. . . . . camera view finders
. . . . cameras
. . . . . *<cameras by imaging technique>*
. . . . . digital cameras
. . . . . film cameras
. . . . . *<film cameras by film size>*
. . . . . 35 mm cameras
. . . . . medium format cameras
. . . . . miniature cameras
. . . . . instant picture cameras
. . . . . plate cameras
. . . . . *<cameras by viewing method>*
. . . . . reflex cameras
. . . . . single lens reflex cameras
. . . . . twin lens reflex cameras
. . . . . viewfinder cameras
. . . . . special-purpose cameras
. . . . . stereo cameras
. . . . . underwater cameras

(objects) (cont.)
equipment *(cont.)*
. *<equipment by application> (cont.)*
. . physics equipment
. . . optical instruments
. . . . light meters
. . . . microscopes

(fields of work)
fields of work
. diving
. photography
. physics

(people)
people
. *<people by age>*
. . infants
. . children
. . adults
. *<people by profession>*
. . divers
. . models (people)
. . photographers
. . physicists

(properties)
properties
. physical properties
. . optical properties
. . . colour
. . . contrast
. . . luminance
. . . luminosity
. . pressure
. . temperature

Figure 6 — Hierarchical display of thesaurus terms arranged by facets

10.5 Classified display

10.5.1 *Thesaurus terms arranged by subject fields*

The approach of arranging thesaurus terms by subject fields is similar to that adopted by the developer of a library classification who begins by organizing the universe of concepts into main classes or disciplines. When applied to a thesaurus, categories of concepts should be grouped initially to reflect the various fields of interest of its users. This technique applies to a thesaurus covering a range of different disciplines. Terms which are usually associated with a given field, such as "art", are brought together, and are also effectively separated from those belonging to different areas of interest, such as "economics" or "physics".

This approach can also be applied to a thesaurus which is restricted to a single field. A thesaurus devoted to "medicine", for example, might be organized initially into subfields such as "surgery", "physiology", "anatomy", etc. To complement the core terms, terms from different disciplines may also be used, for example "management", "law" and "data processing", which should be separated from the medical terms.

The subject approach has the advantage of organizing concepts into groups which generally correspond to the modes of thinking of the users. One difficulty is that some concepts are used habitually in several different fields, for example if a thesaurus covers forestry as well as agronomy, there will be many concepts, such as planting, which occur in both. Such concepts need to be either repeated in each field in which they occur or given in a separate list of general concepts not limited to a specific field.

Grouping terms by subject field, with sibling terms arranged either alphabetically or systematically, and with indentation of some under others, produces displays which look like hierarchies but in which the subordinate terms may not follow the rules for hierarchical relationships described in **8.3**. Node labels giving the facet names should, therefore, be inserted into such displays to indicate where different facets occur and where the relationships are not hierarchical, as discussed in Clause **9**.

10.5.2 *Thesaurus terms arranged by facets within subject fields*

In practice, the different approaches described in **10.4** and **10.5.1** are frequently combined, for example when a thesaurus organized primarily into subject fields is further subdivided by facets.

Figure 7 illustrates the combined approach using the example of a thesaurus that takes in the fields of diving, photography and physics. Only the terms relevant to a particular field are shown within that field. Some terms common to all three fields are presented in a section of general terms. After the primary organization by subject, the next subdivision is by facet. In Figure 7 the facets "objects", "people" and "properties" have been applied to each subject section in turn.

In printed displays, it is often necessary to limit the total size of the compilation by choosing one principal location for each term. The term may be repeated in other locations but space is saved by avoiding repetition of all its narrower terms. The user is expected to refer to the principal location to find the complete hierarchy. Thus in Figure 7, the "principal location" of properties is at A200, where chemical properties, physical properties and all their narrower terms are also displayed. In the diving section, the only physical properties to be mentioned are pressure and temperature; in the photography section, only the optical properties are fully spelt out.

Figure 7 also illustrates the use of notation for locating terms in a classified display. In this example, a term has only one notation, corresponding to its principal location. Where the term is repeated in other locations, it is followed by the notation in brackets, helping the user to refer to the principal location. Some thesauri allow a term to have subsidiary notations, so that every repetition can easily be located.

The notion of "principal location" is an artifice, convenient for restricting the size of a printed thesaurus, but adding extra steps to the process of thoroughly researching a concept. Computer displays can make the look-up process much easier if they allow users to expand or collapse portions of hierarchy, showing as much detail as users require, only when they want to see it.

A000 general terms

(objects)
A001 . equipment
. . *<equipment by portability>*
A002 . . . fixed equipment
A003 . . . portable equipment
. . *<equipment by power source>*
A004 . . . electrically-powered equipment
A005 . . . human-powered equipment
. . *<equipment by application>*
. . . diving equipment (D001)
. . . photographic equipment (H001)
. . . physics equipment (P001)

(people)
A100 . people
. . *<people by age>*
A101 . . . infants
A102 . . . children
A103 . . . adults
. . *<people by profession>*
. . . divers (D008)
. . . models (people) (H025)
. . . photographers (H026)
. . . physicists (P005)

(properties)
A200 . properties
A201 . . chemical properties
A202 . . physical properties
A203 . . . optical properties
A204 . . . . colour
A205 . . . . contrast
A206 . . . . luminance
A207 . . . . luminosity
A208 . . . pressure
A209 . . . temperature

(fields of work)
A300 . fields of work
. . diving (D000)
. . photography (H000)
. . physics (P000)

D000 diving

(objects)
. equipment (A001)
D001 . . diving equipment
D002 . . . aqualungs
D003 . . . diving suits
D004 . . . . dry suits
D005 . . . . wet suits
D006 . . . face masks
D007 . . . swimming fins
. . . underwater cameras (H023)

(people)
. people (A100)
D008 . . divers

(properties)
. properties (A200)
. . physical properties (A202)
. . . pressure (A208)
. . . temperature (A209)

H000 photography

(objects)
. equipment (A001)
H001 . . photographic equipment
H002 . . . camera accessories
H003 . . . . flash guns
. . . . light meters (P003)
H004 . . . . tripods
H005 . . . cameras and camera components
H006 . . . . camera components
H007 . . . . . camera lenses
H008 . . . . . camera viewfinders
H009 . . . . cameras
. . . . . *<cameras by imaging technique>*
H010 . . . . . digital cameras
H011 . . . . . film cameras
. . . . . *<film cameras by film size>*
H012 . . . . . 35 mm cameras
H013 . . . . . medium format cameras
H014 . . . . . miniature cameras
H015 . . . . . instant picture cameras
H016 . . . . . plate cameras
. . . . . *<cameras by viewing method>*
H017 . . . . . reflex cameras
H018 . . . . . single lens reflex cameras
H019 . . . . . twin lens reflex cameras
H020 . . . . . viewfinder cameras
H021 . . . . . special-purpose cameras
H022 . . . . . stereo cameras
H023 . . . . . underwater cameras
H024 . . . . . high pressure underwater cameras

(people)
. people (A100)
H025 . . models (people)
H026 . . photographers

(properties)
. properties (A200)
. . physical properties (A202)
. . . optical properties (A203)
. . . . colour (A204)
. . . . contrast (A205)
. . . . luminance (A206
. . . . luminosity (A207)

P000 physics

(objects)
. equipment (A100)
P001 . . physics equipment
P002 . . . optical instruments
P003 . . . . light meters
P004 . . . . microscopes

(people)
. people (A100)
P005 . . physicists

(properties)
. properties (A200)
. . physical properties (A202)

Figure 7 — Classified display of thesaurus terms arranged by facets within subject fields

10.5.3 *Expanded classified display*

It is possible to add to the type of display shown in Figure 7, to the extent of presenting, for every term in every one of its locations, all of the data described in the single record display as in **10.2**. Although Figure 8 does not go to this length it illustrates the inclusion of additional information such as scope notes, non-preferred terms and related terms.

NOTE See, for example, underwater cameras at H023.

The inclusion of extra data adds to length. To counteract at least some of this, Figure 8 adopts a slightly different convention regarding repetition of hierarchies. Comparing Figure 8 with Figure 7, it can be seen that the mention of pressure and temperature in the diving section of Figure 8 does not include their broader terms physical properties and properties as shown at the corresponding location in Figure 7. To find these, the user needs to follow the notation links to A208 and A209, where the full hierarchies are visible. Apart from saving some space, this approach avoids giving a false impression of the scope of a broad term such as properties or physical properties, when appearing in its secondary locations.

To conclude the comparison of the merits of Figure 6, Figure 7 and Figure 8, it should be clarified that no one of these is being recommended as the best for all situations. Each of the approaches has pros and cons, the importance of which varies according to the context of application. Further variations are also possible. However, it is essential that at least one form of systematic display, hierarchical or classified, should be available to users in addition to the customary alphabetical display.

10.6 Displaying polyhierarchical relationships (see also **8.3.5**)

10.6.1 The standard treatment of polyhierarchical relationships is demonstrated in Figure 5, in which the term underwater cameras has the two broader terms special purpose cameras and diving equipment. It, therefore, appears at two positions in the *equipment* facet of Figure 6. If it had narrower terms as well, these would be shown in both positions.

10.6.2 A different approach may be adopted in the classified or hierarchical section of a printed thesaurus in which space is strictly limited and an extensive entry is needed for the polyhierarchical term. It might have a large hierarchy of narrower terms, and possibly also a long scope note, and/or several non-preferred synonyms and related terms. As in Figure 4 and Figure 5, one "principal location" might be chosen for the polyhierarchical term, and in other locations a less complete display might be shown. See, for example, the treatment of underwater cameras in Figure 8.

Computer displays should offer users the facility to see all the relationships for a polyhierarchical term, wherever it occurs.

A000 general terms

<*objects*>
A001 . equipment
. . <*equipment by portability*>
A002 . . . fixed equipment
A003 . . . portable equipment
. . <*equipment by power source*>
A004 . . . electrically-powered equipment
A005 . . . human-powered equipment
. . <*equipment by application*>
. . . diving equipment [D001]
. . . photographic equipment [H001]
. . . physics equipment [P001]

<*people*>
A100 . people
. . <*people by age*>
A101 . . . infants
A102 . . . children
A103 . . . adults
. . <*people by profession*>
. . . divers [D008]
. . . models (people) [H025]
. . . photographers [H026]
. . . physicists [P005]

<*properties*>
A200 . properties
A201 . . chemical properties
A202 . . physical properties
A203 . . . optical properties
A204 . . . . colour
A205 . . . . contrast
A206 . . . . luminance
RT: light meters [P003]
A207 . . . . luminosity
A208 . . . pressure
A209 . . . temperature

<*fields of work*>
A300 . fields of work
. . diving [D000]
. . photography [H000]
. . physics [P000]

D000 diving

<*objects*>
D001 . diving equipment
BT: equipment [A001]
D002 . . aqualungs
D003 . . diving suits
D004 . . . dry suits
D005 . . . wet suits
D006 . . face masks
D007 . . swimming fins
. . underwater cameras [H023]

<*people*>
D008 . divers
BT: people [A100]

<*properties*>
. pressure [A208]
. temperature [A209]

H000 photography

<*objects*>
H001 . photographic equipment
BT: equipment [A001]
H002 . . camera accessories
H003 . . . flash guns
. . . light meters [P003]
H004 . . . tripods
H005 . . cameras and camera components
H006 . . . camera components
H007 . . . . camera lenses
H008 . . . . camera viewfinders

photography *(cont.)*

<*objects*> (cont.)
. photographic equipment *(cont.)*
. . cameras and camera components *(cont.)*
H009 . . . cameras
. . . . <*cameras by imaging technique*>
H010 . . . . . digital cameras
UF+: underwater digital cameras
H011 . . . . . film cameras
. . . . . <*film cameras by film size*>
H012 . . . . . 35 mm cameras
H013 . . . . . medium format cameras
H014 . . . . . miniature cameras
H015 . . . . . instant picture cameras
SN: Cameras which produce a finished print directly
UF: Polaroid® cameras
H016 . . . . . plate cameras
. . . . <*cameras by viewing method*>
H017 . . . . . reflex cameras
SN: Cameras in which the image is reflected on to a glass screen for composing and focusing
H018 . . . . . single lens reflex cameras
UF: SLR cameras
H019 . . . . . twin lens reflex cameras
H020 . . . . . viewfinder cameras
H021 . . . . special-purpose cameras
H022 . . . . . stereo cameras
H023 . . . . . underwater cameras
SN: cameras for taking pictures under water
UF+: underwater digital cameras
BT: diving equipment [D001]
H024 . . . . . high pressure underwater cameras

<*people*>
. people [A100]
H025 . . models (people)
H026 . . photographers

<*properties*>
. optical properties [A203]
. . colour [A204]
. . contrast [A205]
. . luminance [A206
. . luminosity [A207]

P000 physics

<*objects*>
P001 . physics equipment
BT: equipment [A100]
P002 . optical instruments
P003 . . light meters
UF: exposure meters
RT: luminance [A206]
P004 . . microscopes

<*people*>
P005 . physicists
BT: people [A100]

<*properties*>
. physical properties [A202]
. . optical properties [A203]
. . . colour [A204]
. . . contrast [A205]
. . . luminance [A206]
. . . luminosity [A207]
. . pressure [A208]
. . temperature [A209]

Figure 8 — Classified display of thesaurus terms arranged by facets within subject fields, with additional information such as SN, UF and RT

10.7 Introduction to the thesaurus

All thesauri should contain a comprehensive introduction which clearly states the purpose of the thesaurus:

a) the subject field(s) covered, with core and fringe areas identified;

b) the meaning of all conventions, abbreviations and any punctuation marks used in non-standard ways;

c) the total number of terms, with separate totals of preferred and non-preferred terms;

d) the rules adopted in selecting the forms of preferred terms, including a reference to any style manual that is followed, and the rules for deciding their interrelationships;

e) the filing rules employed, following (and quoting) an appropriate national or international standard where possible;

f) a statement on the updating policy, and the name and address of the responsible agency to whom comments and suggestions should be sent;

g) the date of the last change;

h) references to, and acknowledgements of, sources used in compiling and reviewing the thesaurus.

Whenever possible, these points should be illustrated by examples.

If the thesaurus was designed for a particular system or application, the introduction should also explain how to use it in that context.

11 Thesaurus functions in electronic systems

11.1 If a thesaurus is to be used for information retrieval, it should be capable of integration with the systems in which indexing and searching take place. For a printed thesaurus, integration is achieved by the process of a person looking through the pages, selecting terms and re-entering them into the information retrieval system, with the correct syntax. With an electronic thesaurus, these functions should be made as simple as possible.

For all types of electronic application, the first requirement is to export the thesaurus data from the thesaurus management system into the information retrieval system. Formats and protocols for this purpose are described in BS 8273-5.

11.2 For indexing and searching applications, facilities for browsing and searching the thesaurus are fundamental. These should include the following.

a) The ability to search for terms containing any word or character string.

b) The ability to move from the record for one term to that of any of its NTs, BTs or RTs, generally achieved by setting up hypertext links. This facility may remove the need for a notation to link the alphabetical and classified or hierarchical displays.

c) The ability to expand and contract levels when viewing a classified or hierarchical display. To gain an initial overview of the structure it should be possible to start with a view of the top level, then expand progressively and selectively downwards. It should also be possible to expand or contract fully at one stroke.

d) The ability to switch from one style of display to another, for example from the classified to the alphabetical display. Preferably it should be possible to view them simultaneously.

11.3 When integrated with a searchable document collection, the displays should also show the number of postings against each preferred term. While browsing the thesaurus, a user should be able to move directly to the searching phase. The following facilities should be provided.

a) The ability to select one or more terms from a thesaurus display to be used in a search.

b) The ability to build a search statement by copying terms from the thesaurus without retyping (for example by double-click or drag-and-drop). Help with applying valid search syntax (for example selecting Boolean operators) should ideally be provided.

c) The ability to select subtrees from the thesaurus to be used in a search, i.e. to search for a given term and all its narrower terms combined with Boolean OR, sometimes called the “explode” function.

d) Selection of a non-preferred term should lead to substitution with the corresponding preferred term in the search statement (ensuring that the searcher is notified of what has been done). In the event that the user enters a non-preferred term into a search statement without first browsing the thesaurus, he/she should be offered the option of converting it to the corresponding preferred term.

e) During the process of search refinement, the numbers of postings in any of the thesaurus displays should be adjusted to show the numbers that apply in the subset of the collection already selected.

11.4 When integrated with an indexing or cataloguing system, the following functions should be provided.

a) The ability to copy terms directly into an indexing form without retyping.

b) Validation of terms entered directly, against the thesaurus. In the event that a non-preferred term is entered, this needs to be converted automatically to the corresponding preferred term, with a notification to the indexer. When a string is entered that corresponds to the start of one or more thesaurus terms, the indexer should be offered all the matching preferred terms from which to choose.

c) A facility to submit candidate terms to the thesaurus editor (see **13.2**).

11.5 If the system is for automatic indexing, rather than for human indexers to use, a means of linking each preferred term with the algorithms or rules that drive its selection should also be provided.

12 Management aspects of thesaurus construction

12.1 Planning a thesaurus

12.1.1 *Determining objectives*

Building a thesaurus is a labour-intensive job, requiring commitment for many years if it is to prove worthwhile. It is vital to clarify objectives, based on considerations such as the following, before significant time and money are spent:

a) what the thesaurus is to be used for, and by whom;

b) whether it will be limited by the constraints of existing software with which it is to be used;

c) how knowledgeable the users will be in the subject area of the thesaurus and in thesaurus use.

Different stakeholders (such as senior managerial staff, IT support staff and the prospective users) have different expectations about what a thesaurus is and how it will perform. Agreeing a statement of objectives is a key step in procuring effective teamwork.

12.1.2 *Determining thesaurus features*

When knowledge of objectives and of the planned uses and user community has been obtained, the scope and style of the thesaurus should be defined.

The following features should be determined:

a) whether the thesaurus will be used in print or electronic format, or both;

b) the style of display required;

c) whether special formats are needed for integration with other systems, such as searching or indexing systems;

d) what format is envisaged for updates, and how frequently these will be required.

Decisions should be taken on optional features, such as whether to subdivide the hierarchical relationship as described in **8.1** to **8.3.4**, whether to introduce customized relationships (see **8.5**), whether to show node labels in output displays and whether it is important to store term definitions, sources, and date information.

At a more detailed level decisions should be taken on the conventions to be used for singular/plural, spelling and capitalization. It may be useful to consider practical constraints such as the maximum length of term that may be introduced, any diacritics or other special characters needed, and whether or not there needs to be any limit to the number of levels of hierarchy.

12.1.3 *Taking stock of resources*

12.1.3.1 *General*

During the planning phases of a thesaurus project, the key resources, such as people, a budget, software tools and vocabulary resources, should be determined.

12.1.3.2 *Human resources*

Human resources should include the following:

a) a champion at the senior management level, and preferably corporate commitment to the thesaurus project;

b) interest and support from users, particularly from any with an excellent knowledge of the terminology in one or more areas of the subject field;

c) information professionals with aptitude and preferably training in thesaurus work, to take on the main task of thesaurus construction and maintenance;

d) support from IT professionals who can help with installation/adaptation of thesaurus management software, any programming required to integrate thesaurus outputs with the systems in which the thesaurus will be applied, and the day-to-day running of normal desktop tools.

If more than one thesaurus editor is involved, careful planning is needed to avoid simultaneous, potentially conflicting, inputs to the same portion of the thesaurus.

12.1.3.3 *Vocabulary resources*

It is rare to start building a thesaurus without some existing vocabulary resources, including any or all of the following.

a) Existing thesauri or classification schemes with a scope that partially or wholly matches the requirement. In the extreme case, it may be cost-effective simply to adopt an existing thesaurus unchanged. More commonly, portions of an existing vocabulary may be taken and/or adapted, subject to respect of copyright. Existing works are a source of ideas for terms, or structure, or both.

b) Collections of terminology or frequently asked queries that colleagues may have assembled in the course of their work.

c) The indexes of existing databases or other relevant reference tools, showing useful entry points to the subject matter. If an existing relevant database has been indexed with a controlled vocabulary, it may be possible to obtain from it a report of all the index terms used, and their frequency of usage.

d) The transaction log of a relevant website, showing the search terms most frequently employed by users.

e) Standard reference works such as dictionaries and encyclopaedias are useful, for verification of a term's meaning rather than as a guide to selection of terms.

12.1.4 *Establishing responsibilities*

At an early stage an editor should be appointed, to take responsibility for managing the project through all of its stages, possibly including continuing maintenance as well as compilation of the first edition. Depending on the size of the project, the editor may be supported by an editorial team, and possibly by an editorial review board. There are some advantages in combining thesaurus editorial responsibility with quality control of the subject indexing of resources using the thesaurus.

12.1.5 *Choosing management software*

Although it is possible to compile and produce a thesaurus entirely without software support, this procedure is not recommended. Purpose-built thesaurus management software should be used for efficiency and to avoid clerical errors.

When choosing software a list of requirements should be drawn up, based on the features outlined in Clause **14**, and any additional requirements. Existing software products on the market should be evaluated. If none of them meets all of the requirements, some compromises may be needed. Alternatively one of the existing software products could be customized, or a new one could be developed.

12.2 Early stages of compilation

12.2.1 *General: how and when to start*

Assuming that the thesaurus is needed to support the indexing and searching of a database or other resource, the thesaurus should be completed before database compilation begins in order to ensure that indexing is conducted systematically. However, it is difficult to decide which terms will be needed in the thesaurus until a substantial part of the database has been built. To avoid a later need for re-indexing, at least part of the thesaurus should be prepared in advance, but the need for substantial continuing work on the thesaurus should be anticipated.

Software tools are available which analyse a corpus of text in the relevant subject field(s) and extract significant words or phrases, together with their frequency of use. Some packages also organize the terms into a structured format. While an automatically created thesaurus is unlikely to match the quality of one prepared intellectually, as a source of ideas it can speed up the intellectual process.

12.2.2 *Collecting terms*

As stressed in **6.1.1**, the terms appearing in a thesaurus are only representations of the underlying concepts on which the thesaurus is based. During construction of the thesaurus, the overall aim is to collect and organize concepts, but in practice this is achieved by collecting terms and arranging them so that they adequately represent the required concepts.

Prime sources of terms are listed in a) to d) of **12.1.3.3**. The full text of literature central to the required scope is also useful. As described in **12.2.1**, automatic techniques may be used to extract key words and phrases from the literature. While it is relatively easy to study the material to be input, at least as much emphasis should be given to any indications of what it is that users want to search for, such as lists of queries registered at an enquiry desk. The users themselves may be invited to suggest useful terms and concepts.

At this early stage, it is not necessary to make decisions as to the best representation of a required concept. The initial task is to collect all the options available, noting the source of each and the frequency of occurrence.

12.2.3 *Analysing terms*

The terms collected should be sorted into systematic order before they can be studied properly and entered into the thesaurus. Standard desktop tools such as spreadsheets may be useful for this preliminary sorting out. The process is made easy and efficient if the terms are obtained from a machine-readable source. Once the collected terms are in a spreadsheet, together with their source and frequency, they should be sorted into subject areas or facets (see Clause **9**) so that like terms with all their variants and synonyms are brought together. Although the classification or facet structure employed at this stage may be rudimentary, and need not determine the eventual structure of the thesaurus, it may help in developing ideas for the latter. The main product of this analytical stage should be a list or set of small lists of terms, grouped by subject and/or facet, with indications of the source and frequency of use of each term. Armed with these data, the editor may then begin systematic construction.

12.3 Managing construction

12.3.1 *Data management*

12.3.1.1 *General*

Thesaurus management software is generally designed to record all the attributes and relationships associated with the terms that are entered, support maintenance by the editor and deliver outputs as required. Various levels of sophistication are possible.

12.3.1.2 *Management at the simplest level*

Two types of entity or information object need to be manipulated, namely preferred and non-preferred terms. Attributes and relationships for these entities are shown in Table 1 and Table 2.

Table 1 — Attributes and relationships of a preferred term

Element	Occurs	Essential or not
Attributes		
Preferred term	1 time	Essential
Notation or class code [**10.1.2**]	1 or more times	Optional
Scope note [**6.1.2.3**]	0 or 1 time	Essential
History note [**6.1.2.6**]	0 or 1 time	Optional
Source	0, 1 or more times	Optional
Definition [**6.1.2.5**]	0 or 1 time	Optional
Status	1 time	Optional
Editorial note	0, 1 or more times	Optional
Date of input	1 time	Essential
Date last modified	0 or 1 time	Essential
Term identifier	1 time	Essential
Relationships		
Corresponding non-preferred term [**8.2**]	0, 1 or more times	Essential
Corresponding non-print non-preferred term [**6.5.2**]	0, 1 or more times	Optional
Corresponding partial non-preferred term (for UF+ terms) [**8.2.5**]	0, 1 or more times	Optional
Broader term [**8.3**]	0, 1 or more times	Essential
Narrower term [**8.3**]	0, 1 or more times	Essential
Related term [**8.4**]	0, 1 or more times	Essential

In Table 1, Table 2 and Table 3, the second column indicates the number of times an element may occur for a particular entity; the third column indicates whether it is essential or optional to have the capability for entering that element. For example, Status may or may not be used in a thesaurus management system, but if it is used, then a Status must occur once and only once for each term.

Table 2 — Attributes and relationships of a non-preferred term

Element	Occurs	Essential or not
Attributes		
Non-preferred term	1 time	Essential
History note [**6.1.2.6**]	0 or 1 time	Optional
Source	0 or 1 time	Optional
Definition [**6.1.2.5**]	0 or 1 time	Optional
Status	1 time	Optional
Editorial note	0, 1 or more times	Optional
Date of input	1 time	Essential
Date last modified	0 or 1 time	Essential
Term identifier	1 time	Recommended
Relationships		
Corresponding preferred term [**8.2**]	1 time, unless a combination of preferred terms is present	Essential
Preferred term to be used in combination with others (in case of USE A + B) [**8.2.5**]	2 or more times, unless a single preferred term is present	Optional

Most of the elements are described in other clauses of this standard, as indicated by the cross-references in square brackets in the first column. The others are used as follows.

a) Source may be used to note the reference work or individual who contributed the term in question.

b) Status records whether the term is approved or still a candidate (see **13.2**).

c) Editorial note is useful for entries such as "Review this term after the company merger complete" or "This term is mentioned in the scope note of Term X" (see **6.1.2.4**) or "Check spelling with expert A".

d) The Date attributes and Term identifier are essential for the provision of a good updating service, e.g. if the spelling of a term changes, a constant Term identifier facilitates continuity during successive updates; the Term identifier has to be unique.

Several of the attributes are more useful for housekeeping than for user consultation. A choice may be made whether or not to allow users to view them.

12.3.1.3 *Managing concepts*

Each concept in the thesaurus has at least one preferred term, and any number of non-preferred terms. The notation, scope note, and broader/narrower/related term relationships apply to the concept as a whole, not just to its preferred term.

In some systems, the concept is identified only by its preferred term, but this has disadvantages if the spelling of the term changes. Alternatively, the term identifier of the preferred term may be used to identify the concept, or a unique concept identifier may be assigned. The unique concept identifier applies to the concept itself, and to all its preferred and non-preferred terms.

12.3.1.4 *Node labels*

The entities and elements already described are sufficient to support most types of display. But if node labels are used, they have also to be managed.

Table 3 — Attributes of a node label

Attribute	Occurs	Essential or not
Node label	1 time	Essential
Status	1 time	Optional
Notation or class code [**10.1.2**]	1 time	Optional

Node labels are very different from thesaurus terms, in that they do not need to be unique. The same form of words may be used for many different labels. In this event the separate instances are distinguished by having a unique notation or sort code.

If the thesaurus uses an expressive notation, then applying it to node labels will allow them to be shown in the correct place in classified displays.

If there is no expressive notation, some other means must be found of outputting the node labels correctly in the display, such as a sort code attribute that is not displayed to users. In this event, the same attribute will be required at the display level for preferred terms.

Some systems organize the display by assigning broader and narrower terms to the node labels, but this is done only for display purposes and not for the normal functions of broader and narrower term relationships. Only a preferred term can have true broader and narrower terms.

12.3.1.5 *Managing additional attributes and relationships*

A good management software package will allow additional attributes and relationships to be created in the event of customization. (See **8.5**.)

12.3.2 *Input*

When terms are entered manually, they should be handled in subject/facet groups or hierarchies. Groups of synonyms and near-synonyms should then be considered together and preferences and equivalence relationships established. At this stage, the terms and their hierarchical relationships should be selected carefully, taking into account potential duplications, overlaps, omissions and the degree of specificity required. Scope notes may also be considered and entered at the same time. If desired, source and definition may also be entered. Date of input and Term identifier are best added automatically by the software, rather than by manual keying.

If the initial stocktaking exercise has revealed an existing thesaurus of which a substantial portion should be adopted, it may be possible to import this directly into the chosen management software. This procedure can save time, particularly if the software permits import of structured data with scope notes and relationships as well as the terms. However, it is essential that the permission of the copyright holder is obtained.

Unstructured lists of terms may also be imported directly into some software products. This does not save as much time since relationships still need to be entered and careful inspection is needed to ensure that spelling inconsistencies do not enter unnoticed.

12.3.3 *Construction sequence*

When working with hierarchies, the question arises of whether to build from the top downwards or from the bottom upwards. A view of the top layers is useful in ensuring that the thesaurus has the appropriate coverage and balance to meet the desired scope. It is also useful when inviting contributions from experts and other users to be able to place their advice in the context of the overall structure.

On the other hand, work at the more specific level tends to throw up anomalies which may affect structure at the higher levels. The availability of low-level portions of hierarchy, revealing many concepts that need to be accommodated somewhere, tends to inform development of a sound superstructure.

Both strategies are useful and should be pursued in combination. A sketch of the high-level structure should be prepared. Then work should proceed on lower-level sections of hierarchy, one at a time. As the work proceeds, the high-level sketch should be modified to meet emerging needs.

It is efficient to work with groups of hierarchically clustered terms, entering them together with equivalence and hierarchical relationships. However, associative relationships cannot be entered as easily at this stage, because the terms at the other end of the association are usually in another hierarchy, still to be developed. Associative relationships are usually entered at a later stage, when the bulk of the terms are present in the system.

12.3.4 *Involving experts and stakeholders*

The following benefits may be obtained from involving other people at various stages of the work.

a) Specialist terminology can be difficult to grasp and have surprising nuances of meaning. Someone who knows a specialized area well can quickly check whether the terminology chosen is appropriate and relationships correctly appointed. If terms are applied in ways that are unacceptable to experts within the user community, the thesaurus will not easily gain acceptance. In principle, therefore, it is a good idea to involve experts in checking as work proceeds. However, terms typically have more than one meaning, and the meaning favoured by one expert may not meet the approval of another. Furthermore, as the concept most useful for information retrieval purposes may be slightly broader or narrower than the term preferred for representing it, some experts find it difficult to accept the perceived "inaccuracy". The experts chosen to assist should be well briefed in the purposes and conventions of a controlled vocabulary (see Clause **5**). The thesaurus developer should be prepared to guide and mediate all contributions.

b) It is difficult for one person to make all the correct choices at the first attempt. A great deal may be gained from discussing difficult terms among an editorial team, with a degree of lateral thinking. If the team consists of one editor, it may be beneficial to set up a committee or panel to review sections of the work in progress, or to consult selected colleagues ad hoc. Optionally, the editorial review board may be asked to approve all elements of the thesaurus before they are issued.

c) The occasional involvement of senior management and/or other stakeholders may be an asset in securing continued funding, acceptance and implementation of the finished product.

d) Involvement of the users may greatly enhance buy-in. People with a stake in the project are much more likely to use it.

12.4 Dissemination

12.4.1 *Integration with an electronic system*

For people using a thesaurus, either for indexing or for searching, access is most convenient if the thesaurus with all its terms, relationships and scope notes is fully integrated with the indexing/browsing/searching functions, as discussed in Clause **11**.

Some indexing and/or searching systems have an integrated thesaurus maintenance module. If this is used for the thesaurus development, no import or export is required, and no special preparations are needed. However, if the integrated system has no mechanism for exporting the thesaurus in a standard format, difficulties will occur if the need arises to change systems or to make the same thesaurus available for additional applications.

Whether or not routine dissemination is needed, it is essential that a capability for exporting the thesaurus in a standard format is provided. Indexing and searching systems that use the thesaurus should be capable of importing the thesaurus in the same format. Guidance on formats is given in BS 8723-5.

12.4.2 *Electronic publication*

The thesaurus might also be made available in stand-alone mode, i.e. not linked to any particular application. An electronic thesaurus may be distributed on CD-ROM, or on a website on an intranet or the Internet. The applicable browse functions are described in **11.2** and **11.3**.

Even when a thesaurus is published on a network without a direct linkage to any particular application, some search applications on the same network (an intranet or the Internet) can use it for real-time interrogation while a search is in progress. To make this easier, the thesaurus should be made available in XML format, accessible as an XML "namespace".

Certain characters have restricted use in XML (e.g. < > & ‘ “). These characters should be avoided where possible as described in **6.3.2**. Where, however, they are needed for user acceptability (e.g. Boyle's Law, women's rights) they should be encoded as set out in the XML standard [5], by the thesaurus management software when generating the XML output.

12.4.3 *Limited distribution of hard copy*

The thesaurus might be made available in hard copy, for limited distribution under the following circumstances:

a) a full-scale print run would be too expensive;

b) most users prefer an electronic version;

c) the thesaurus is strictly an in-house product;

d) the thesaurus editorial team find a printout convenient for display and/or note-taking.

Realistically, this option is a simple extension of the routine practice of obtaining reports from the thesaurus database on demand. It offers the flexibility to tailor-make the report(s) for ad hoc needs, for example to include or exclude the "housekeeping fields", to issue a subset for a particular user group, to choose between the various styles of display, etc. Such reports may be distributed electronically, leaving individual users to print them out as required.

12.4.4 *Conventional publication*

Printed displays have some advantages over electronic ones, particularly in locations where access to a computer is limited. Conventional publication may be essential if the thesaurus is designed for widespread use rather than one specific application. Furthermore it may bring the thesaurus to the notice of a different, important audience.

When an organization has decided to develop and publish a new thesaurus, notification of intent should be announced in a professional journal in the relevant field(s), as well as through relevant electronic discussion forums.

12.4.5 *Deposit with a clearinghouse*

A copy of the first and any subsequent edition of a published thesaurus should be deposited with an appropriate national depository library. If the thesaurus is in English, or includes English as one of its languages, a copy should also be deposited with the international clearinghouse in Toronto[4)].

NOTE The website <http://www.fis.utoronto.ca/inforum/sas.htm> provides more information about the clearinghouse, and its library catalogue can be searched.

12.4.6 *Website directories*

Lists of thesauri are shown on a number of websites available through the Internet. Some of these sites publicize thesaurus projects, in advance of publication. On or before publication, the thesaurus developer should send full details of the thesaurus, whether electronic or printed to the website administrator. In the case of an electronic thesaurus available through the Internet, the website might provide a direct link to it.

13 Updating

13.1 General

Thesaurus maintenance is essential to provide for changes, both in the natural evolution of terminology and in the scope of the databases or other resources that the thesaurus serves. The need for updating begins the day a thesaurus is issued, if not before. Suggestions for change may be particularly frequent in the first few months after issue, as that is the time when the vocabulary is undergoing its first serious testing. Editorial procedures should be in place from the start.

While change is essential to cope with emerging requirements, it is important to recognize that every change potentially impairs retrieval performance for the "backfile" (i.e. for the resources already indexed with the original vocabulary). When a new term is introduced, it is available for searching the database from that date on, but relevant resources already in the system will not have used the term in their indexing. If the database management software allows automatic or semi-automatic correction of backfile records, changes to existing preferred terms may be implemented. However, other types of changes may be less susceptible to retrospective fixing. A compromise may, therefore, be required between meeting new needs and maintaining stability.

13.2 Suggestion procedure

An easy mechanism for suggesting changes should be provided for all users of the thesaurus, in particular indexers. The mechanism provided may be a form, either printed or electronic. The form should invite people to suggest a new term or other change that is required, and provide as much background information as possible. Useful background information includes a definition of the concept required, any references to its occurrence in the literature, any synonyms or alternative forms of expression, any relationships with terms already present in the thesaurus, and a justification for why the change is needed.

Indexers often discover the need for a new term while processing documents for input to a database. As there may not be time to consult the thesaurus editor before completing the indexing, the proposed term should be applied as a "candidate preferred term". The candidate terms become available for searching, but are not required to pass thesaurus validation. The thesaurus editor regularly reviews all the candidates proposed (see **13.3**) and may admit them to the thesaurus management system with status "candidate". As and when a candidate is approved, the appropriate adjustment should be made to the status of the term. Candidates not approved should be deleted or corrected as appropriate.

13.3 Review procedure

At regular intervals the thesaurus editor should review all the suggestions received and candidate terms proposed. Furthermore the editor should monitor the usage frequency of preferred terms in the database(s) served. Any that are used too often or too little are candidates for deletion or amendment.

User queries may also be used as an aid in maintenance. Terms used in unsuccessful searches can be examined to see if they are valid concepts for the thesaurus. Heavy use or non-use of a term in queries can indicate that it should be considered for division or deletion, respectively.

[4)] Thesaurus Clearinghouse, The Library, Faculty of Information Studies, University of Toronto, 140 St George Street Toronto, Ontario M5S 3G6 Canada.

The suggestions and candidates should be researched thoroughly. Careful study can show that the suggested concept was already present in the thesaurus under another name. As the most common proposals are for new preferred terms, they rarely give consideration to all the alternative or complementary actions required, such as adding a lead-in entry, changing a scope note, adjusting hierarchical or associative relationships, etc. They may also add unacceptably to the level of specificity. The editor should flesh out each proposal to a fully coherent and consistent set of amendments before approval can be given.

In a small establishment, the editor may need to make all the decisions. Consultation with users is advisable, particularly if the thesaurus is used for several applications. A note of all the proposed changes may be circulated to users, with an invitation to comment or a review board, with representation from each major application, may be consulted. The board should meet regularly or organize a regular electronic forum. After fully researching and fleshing out the proposals, the editor should present them to the board. The implications for each application should be taken into account before a decision is reached.

13.4 The nature of changes

While the main focus of updates tends to be adding new terms, other types of change should not be neglected. Any of the following may occur:

a) add a new preferred term;

b) add a new non-preferred term as a lead-in entry;

c) amend a preferred term;

d) delete a preferred term;

e) delete a non-preferred term;

f) reverse the preference between a preferred and a non-preferred term;

g) add or change a scope note;

h) alter the hierarchical structure;

i) move a branch of the hierarchy from one place to another;

j) add or remove associative relationships;

k) almost any combination of a) to i).

It is essential that the same editorial criteria are continued as applied in the original compilation, with the added criterion of respecting the convenience of existing users. Deletions are of particular concern, especially if the term has been used in the past. Instead of deleting a preferred term, the term should be demoted to a non-preferred term that points to the appropriate preferred term, with a history note showing until what date it was valid. The corresponding preferred term should also have a history note pointing out the change.

While the editor should craft an appropriate history note when terms are changed significantly, the date of input and date last modified should be filled automatically by the software. Similarly, the identifier for new terms should be added automatically.

13.5 Dissemination of updates

13.5.1 *General*

The mechanism by which updates are distributed should be planned before the thesaurus is first issued. Communication of changes can be complex, particularly when the attributes and relationships of one term have changed in several ways. It is straightforward to send out a list of new terms, but it is difficult if the users need to know where the terms fit into the hierarchical structure, particularly if several different types of display need to be distributed. Assuming that a simple list of new terms is acceptable to users, the likelihood is that the new terms will be ignored at the time of using the thesaurus because they have not been fully integrated into the look-up procedure.

The simplest way of disseminating changes is to distribute a new edition of the thesaurus. If this is carried out electronically, costs are kept to a minimum. Furthermore, it avoids the difficulty of communicating the nature of the changes, and it ensures that users all have access to the same vocabulary authority when selecting terms. Additionally, there is a need to communicate the details of the changes both to human users and to computer systems that will import the update.

13.5.2 *Reader-friendly updates*

For human users, notification may take the form of a list in alphabetical order, possibly complemented by one in classified order or a categorized list, showing all the new preferred terms added, the preferred terms altered, new lead-in entries, etc. Feedback should be invited on the format. Depending on the views expressed, the list may be made more or less elaborate. Some user groups might want to see the full details of every new or amended term and relationship. All users should be encouraged to note the most important changes, such as new terms and deletions, in their printed copy if they use it. This may be facilitated if printed versions are produced in loose-leaf form, and amendments take the form of complete pages that can replace outdated ones.

13.5.3 *Updates for computers*

The easiest option for updating computer systems is if another thesaurus management system or a retrieval system accepts a new thesaurus with the changes implemented. However, this may not be acceptable if the target system links thesaurus terms with other data, such as postings to bibliographic records or mappings to another vocabulary. For computer systems such as these, reports or files of all the entities, attributes and relationships that have been deleted, added or modified should be provided, as a minimum.

To support these requirements, the thesaurus management system should be capable of recording the date on which the entities, attributes or relationships were last modified, and using this date to extract a report of the modified data. Alternatively, an external log file may record all additions, deletions and changes to entities, attributes and relationships.

For importing systems, some of the changes will be insignificant, while others will be important. For this reason, some thesaurus management systems record the nature of the change as well as the change itself. This allows the importing computer system to separate relatively straightforward changes (for example spelling changes) from complex changes (for example splits of a term into two terms). As a standard categorization of the possible changes does not exist the procedures for achieving this should be developed ad hoc.

13.5.4 *Housekeeping implications*

The importance of the date last modified field has already been noted, in connection with updates for computers. It is sometimes also needed for reader-friendly updates, as is the date of input. The date should be entered in the standard format yyyy-mm-dd (e.g. 2004-03-17) in accordance with BS EN ISO 8601. The term identifier is also vital, in case the spelling of a term is changed. (Without the identifier, a computer has no way of finding out that a given term is intended to replace another with a different spelling.) All of the attributes mentioned should be filled in automatically, to avoid human error.

Another vital housekeeping aid is a note of all terms deleted. However, few thesaurus management packages have a mechanism for retaining a list of terms deleted and providing reports of them. Therefore, the editor should keep a note, recording the identifier as well as the term itself.

14 Requirements of thesaurus management software

14.1 General

Thesaurus management software should be capable of managing the data as described in **12.3.1**. It should be chosen on the basis of the requirements of the project in hand (see **12.1.5**). The characteristics and functions outlined in **14.2** to **14.10** are commonly needed, but exceptions can occur, and additional features are often required.

NOTE No guidance is given in this British Standard on the criteria applied to software selection in general, such as the availability of good documentation, training and other forms of support, the general user-friendliness of the interface and an acceptable price. Considerations applicable to multilingual thesauri and other systems in which one vocabulary is mapped to another are also excluded as these are covered in BS 8723-4.

14.2 Size and character limitations

The software should not impose limitations as follows.

a) There should be no limitations on the number of terms in the vocabulary which prevent it expanding to the size needed.

b) There should preferably be no limitations on the length of terms. While relatively few terms exceed 40 characters, sometimes they need 100 characters or more.

c) There should be no limitation on the number of hierarchical levels admissible.

d) The software should be capable of handling the special characters sometimes found in terms, such as parentheses, apostrophe, ampersand, etc. For scope notes, all the common punctuation marks should be admissible. For languages other than English, diacritical marks and non-ASCII characters may also be needed.

e) The editor should be able to choose upper or lower case characters as appropriate.

14.3 Inter-term relationships

The software should conform to the following requirements.

a) It should not permit the coexistence of duplicate terms. Any duplicates entered should be submitted to the editor for correction, amalgamation, the addition of a qualifier, or other remedial action. Typographical differences such as italics or capitalization should be ignored for the purposes of duplicate detection.

b) It should support the standard relationships BT/NT, RT/RT, USE/UF.

c) It should support the standard reciprocity requirements, i.e. if term A has a BT relationship with term B, then term B should have an NT relationship with term A and vice versa; if term C has an RT relationship with term D, then term D should have an RT relationship with term C; if term E has a USE relationship with term F, then term F should have a UF relationship with term E and vice versa. Preferably, the software should make the reciprocal relationships available automatically in response to the editor's insertion one way round, but as a minimum the software should issue a warning if any non-reciprocal relationship is present.

d) When a term is amended or deleted, the change should be propagated automatically in all places in which that term appears related to another (whether as BT, NT, RT, USE or UF). In the event of term deletion, all relationships to and from that term should be deleted. However, if deletion of a term or relationship would leave any preferred term without broader terms, or any non-preferred term without a preferred term, the editor should be warned.

e) There should be no limitation on the number of relationships a given preferred term may have with others. Thus one term may have any number of BTs, NTs, RTs, UFs.

NOTE Certain combinations are inadmissible.

f) It should be possible to set up user-defined reciprocal relationships, for example to distinguish between different types of BT/NT or different types of associative relationship.

g) Validation checks should prevent the entry of inadmissible relationship combinations. If two terms already have one of the standard relationships, no other standard relationship between the same terms is admissible. If term A has BT term B, none of the terms in the BT hierarchy above term B should be admissible as BT, NT or RT of term A. Non-preferred terms (i.e. any term with a USE relationship to another term) may not have any BT, NT, RT or UF relationships.

h) No relationship from a term to itself is admissible.

14.4 Term notes

The software should support term notes as follows.

a) It should support entry of a scope note, associated with any term.

NOTE Scope notes may be of any length.

b) If a scope note makes reference to another term in the vocabulary, the software should preferably support addition of a marker to the record for that term.

c) It should support the setting up of user-defined notes associated with any term, for example history notes, editorial notes etc.

14.5 Codes and notation

The software should support classification codes and notation as follows.

a) It should be possible to associate at least one class code, number or other type of notation with any term. Preferably more than one type should be supported.

b) It should be possible to assign a unique identifier to each term. Preferably, the assignment of the identifier should be automatic whenever a new term is entered, in such a way as never to duplicate any of the existing identifiers or identifiers of terms previously deleted. The identifier should not change when term attributes or relationships are modified in any way.

c) It should be possible to output the vocabulary using the sequence of each type of notation or identifier.

14.6 Node labels

The software should support node labels as follows.

a) Node labels are not regarded as thesaurus terms, and, therefore, are not subject to the relationship strictures under **14.3**. Furthermore, they do not need to be unique.

b) The software should have means for locating a node label in displays in the correct position relative to any parent term and to the highest level terms that come within the corresponding facet or array.

14.7 Data import/export

It is essential to be able to export and import data, so that the thesaurus content is not restricted to specific software. The following mechanisms should be provided.

— Bulk import of datasets from existing vocabularies, comprising terms, scope notes and standard relationships between the terms.

• All of the mentioned features should be retained after import.

— Producing reports or exporting the vocabulary, including all terms, scope notes, notation and standard relationships between the terms.

• It should also be possible to export editor-defined subsets, such as non-preferred terms only, or preferred terms with their scope notes and NTs only, or preferred terms that have no hierarchical or associative relationships, etc.

— Batch edit/delete facility.

• It should be possible to edit batches of records in the same way, preferably using a facility native to the software, but if this is not possible an option may be to export the selected records, edit them externally and re-import. This technique may be used, for example, to add the same history note or relationship to a batch of terms, or perhaps a mapping to terms in another vocabulary. Caution is needed, however, to apply the customary validations on import.

— Exporting all terms that have been amended after a given date, with or without their full details.

• The option should be available of selecting only certain types of amendment, for example only new terms, or of including all terms that have undergone any change of attributes or relationships. It should also be possible to report all terms that have been deleted from a certain date.

— Outputting thesaurus displays, either in hard copy or on the screen.

• It should be possible to choose between a variety of sequences and layouts for the display, of which the alphabetical display is essential and either a hierarchical or a classified display highly recommended. See Clause **10** for a description of the various types of display.

— When sibling terms in alphabetical, hierarchical or classified displays are presented in alphabetical order, either the word-by-word or the letter-by-letter convention may be used, and the same convention should be applied throughout the thesaurus. (An explanation of the conventions[5)] may be found in BS ISO 999.) However, it should be possible to force a different sequence when appropriate. (See Clause **9**.)

— Recommended import/export formats are given in BS 8723-5.

[5)] EXAMPLES:

Word-by-word sequence	**Letter-by-letter sequence**
national insurance	national insurance
national parks	nationality
nationality	national parks

14.8 Editorial navigation and support

The software should support the following editorial functions.

a) The editor should be able to locate the details of any term by direct entry of that term, preferably only needing to enter a portion of the term, with the software assisting the selection by identifying all terms that contain the portion.

b) The editor should be able to locate the full details for any term by browsing via the term relationships. The browse facilities should preferably enable him/her to start off from a list of top terms, but in the absence of this he/she should at least be able to navigate from any term to the record for any of its NTs, BTs, RTs, USEs or UFs.

c) The editorial interface should allow viewing of a term's full hierarchical context, at the same time as the term itself, with all its notes, codes and relationships presented for editing.

d) It should be possible to move a term (together with all its NTs, NTs of those NTs, etc.) easily from one hierarchical position to another, preferably using drag-and-drop. This requires breaking one set of BT/NT relationships and making another. Any RT relationships thus invalidated should be drawn to the editor's attention.

e) Editing capabilities should include standard word processing facilities, such as the ability to add, modify and delete characters without rekeying an entire field.

f) When an editor takes steps to delete a term, the software should first seek confirmation that the deletion of that term is intended, before completing the action.

g) When a relationship is to be established between two terms already present in the vocabulary, it should be possible to achieve this by navigation and selection rather than by having to key in the whole of a term already known to the system. This is important to prevent errors, as well as to support efficiency.

14.9 Editorial safeguards

The software should embody the following safeguards.

a) Editorial changes should be made in the first instance to a master database from which outputs are derived periodically for downstream processes such as indexing or resource discovery applications.

b) If more than one person is editing the master database simultaneously, an inbuilt mechanism should prevent simultaneous write access to the same records.

c) Security/password controls should prevent editorial changes from unauthorized persons.

d) Preferably, the software should allow junior staff to make provisional changes, which are not finally admitted until approved by the principal editor.

e) For each editor, a roll-back (undo) function should allow progressive reversal of the most recent editorial changes he/she has entered.

f) A log should be kept so that the database can be restored from an earlier version.

14.10 Housekeeping tools

It should be possible to obtain reports of the numbers of terms with particular characteristics, particularly the total number of preferred and non-preferred terms.

Bibliography

Standards publications

BS 0 (all parts), *A standard for standards.*

BS 2979:1958, *Transliteration of Cyrillic and Greek characters* [ISO 9 Not Equivalent].

BS 4280:1968, *Transliteration of Arabic characters.*

BS 4812:1972, *Specification for the romanization of Japanese.*

BS 7014:1989, *Guide to the romanization of Chinese.*

BS ISO 999:1996, *Information and documentation — Guidelines for the content, organization and presentation of indexes.*

ISO 9:1995, *Information and documentation — Transliteration of Cyrillic characters into Latin characters — Slavic and non-Slavic languages.*

ISO 233 (all parts), *Documentation — Transliteration of Arabic characters into Latin characters.*

ISO 259 (all parts), *Documentation — Transliteration of Hebrew characters into Latin characters.*

ISO 3166 (all parts), *Codes for the representation of names of countries and their subdivisions.*

ISO 9984:1996, *Information and documentation — Transliteration of Georgian characters into Latin characters.*

ISO 9985:1996, *Information and documentation — Transliteration of Armenian characters into Latin characters.*

ISO 11940:1998, *Information and documentation — Transliteration of Thai.*

ISO/TR 11941:1996, *Information and documentation — Transliteration of Korean script into Latin characters.*

Other documents

[1] *ALA-LC romanization tables: transliteration schemes for non-roman scripts* / compiled and edited by Randall K. Barry, Network Development and MARC Standards Office, Library of Congress. — Washington, D.C.: Library of Congress, 1997. 239 p. — ISBN 0-8444-0940-5. — Also available on the Internet at: http://www.loc.gov/catdir/cpso/roman.html

[2] *Anglo-American cataloguing rules* / prepared under the direction of the Joint Steering Committee for Revision of AACR, a committee of the American Library Association [et al.]. — 2nd ed., 2002 revision with 2004 update. — London: Facet Publishing for Chartered Institute of Library and Information Professionals, 2004. — ISBN 0-85604-469-6 (loose-leaf without binder).

[3] *Geographical names and information* / Foreign and Commonwealth Office, Library and Information Services. — London: Foreign and Commonwealth Office, 2004. — Also available on the Internet at: http://www.fco.gov.uk/Files/kfile/GeographicalNamesDec2004,0.pdf

[4] *Art & architecture thesaurus* / Toni Petersen, director. — 2nd ed. — New York; Oxford: Oxford University Press on behalf of the Getty Art History Information Program, 1994. — 5 vols. — Also available on the Internet at: http://www.getty.edu/research/conducting_research/vocabularies/aat

[5] *Extensible Markup Language (XML) 1.0 (Third Edition)*. W3C Recommendation 04 February 2004. Edited by Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau. Available at http://www.w3.org/TR/REC-xml/

Index (BS 8723-1 and -2)

BSI — British Standards Institution

BSI is the independent national body responsible for preparing British Standards. It presents the UK view on standards in Europe and at the international level. It is incorporated by Royal Charter.

Revisions

British Standards are updated by amendment or revision. Users of British Standards should make sure that they possess the latest amendments or editions.

It is the constant aim of BSI to improve the quality of our products and services. We would be grateful if anyone finding an inaccuracy or ambiguity while using this British Standard would inform the Secretary of the technical committee responsible, the identity of which can be found on the inside front cover. Tel: +44 (0)20 8996 9000. Fax: +44 (0)20 8996 7400.

BSI offers members an individual updating service called PLUS which ensures that subscribers automatically receive the latest editions of standards.

Buying standards

Orders for all BSI, international and foreign standards publications should be addressed to Customer Services. Tel: +44 (0)20 8996 9001. Fax: +44 (0)20 8996 7001. Email: orders@bsi-global.com. Standards are also available from the BSI website at http://www.bsi-global.com.

In response to orders for international standards, it is BSI policy to supply the BSI implementation of those that have been published as British Standards, unless otherwise requested.

Information on standards

BSI provides a wide range of information on national, European and international standards through its Library and its Technical Help to Exporters Service. Various BSI electronic information services are also available which give details on all its products and services. Contact the Information Centre. Tel: +44 (0)20 8996 7111. Fax: +44 (0)20 8996 7048. Email: info@bsi-global.com.

Subscribing members of BSI are kept up to date with standards developments and receive substantial discounts on the purchase price of standards. For details of these and other benefits contact Membership Administration. Tel: +44 (0)20 8996 7002. Fax: +44 (0)20 8996 7001. Email: membership@bsi-global.com.

Information regarding online access to British Standards via British Standards Online can be found at http://www.bsi-global.com/bsonline.

Further information about BSI is available on the BSI website at http://www.bsi-global.com.

Copyright

BSI
389 Chiswick High Road
London
W4 4AL