SLOVENSKI STANDARD
SIST ISO 2788:1996

01-april-1996

Dokumentacija - Smernice za zasnovo in razvoj enojezičnih tezavrov

Documentation -- Guidelines for the establishment and development of monolingual thesauri

Documentation -- Principes directeurs pour l'établissement et le développement de thésaurus monolingues

Ta slovenski standard je istoveten z: **ISO 2788:1986**

ICS:

01.140.20 Informacijske vede Information sciences

SIST ISO 2788:1996 **en**

International Standard 2788

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION•МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ•ORGANISATION INTERNATIONALE DE NORMALISATION

Documentation — Guidelines for the establishment and development of monolingual thesauri

Documentation — Principes directeurs pour l'établissement et le développement de thesaurus monolingues

Second edition — 1986-11-15

UDC 025.48 **Ref. No. ISO 2788-1986 (E)**

Descriptors : documentation, subject indexing, information retrieval, thesauri, monolingual thesauri, preparation, rules (instructions).

Price based on 32 pages

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council. They are approved in accordance with ISO procedures requiring at least 75 % approval by the member bodies voting.

International Standard ISO 2788 was prepared by Technical Committee ISO/TC 46, *Documentation*.

This second edition cancels and replaces the first edition (ISO 2788-1974), of which it constitutes a technical revision.

Users should note that all International Standards undergo revision from time to time and that any reference made herein to any other International Standard implies its latest edition, unless otherwise stated.

Printed in Switzerland

Contents

This page intentionally left blank

Documentation — Guidelines for the establishment and development of monolingual thesauri

0 Introduction

The effectiveness of a subject index as a means for identifying and retrieving documents depends upon a well-constructed indexing language. This applies to any system where the selection of indexing terms calls for human intellectual decisions, including those systems in which a computer is used to store and manipulate terms or to identify documents associated with terms or combinations of terms assigned by an indexer.

The compiler of a subject index faces three main tasks

a) determining the subject matter of documents;

b) selecting the terms which together summarize the subject;

c) indicating relationships between the concepts represented by these terms.[1)]

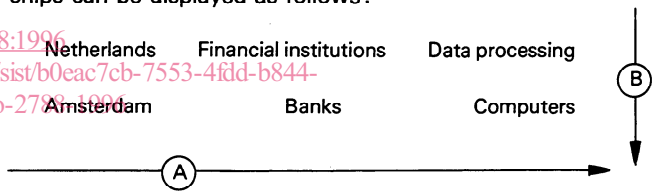
The first of these tasks is described separately in ISO 5963. The second and third tasks concern not only the indexer but also the user of the index. This International Standard deals with some aspects of term selection, since it contains recommended procedures for vocabulary control, but it is particularly concerned with means for establishing and displaying certain kinds of relationships between indexing terms.

Two kinds of inter-term relationships can be distinguished:

a) syntactical or *a posteriori* relationships between the terms which together summarize the subject of a document. For example, an indexer dealing with a work on "Computers in banks in Amsterdam" may assign three terms, "Banks", "Computers" and "Amsterdam", to the document. In a post-coordinate system the relationship between these terms is not explicitly indicated, and the document would be retrieved if any or all of these terms were used as retrieval keys. In a pre-coordinated index the relationships between the terms may be conveyed in various ways, for example by symbols which express specific relationships, the positions of terms within entries, their typography and/or their accompanying punctuation. 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;

b) those *a priori* or thesaural relationships between terms assigned to documents and other terms which, because they form part of common and shared frames of reference, are present by implication. In the example above, "Banks" would imply a broader term such as "Financial institutions"; "Computers" is mentally associated with "Data processing"; and "Amsterdam" implies the wider location "Netherlands". Any of these mentally-associated terms might serve as a user's approach to the subject index. These relationships are document-independent, since they are generally recognized and could be established through reference to standard works, such as dictionaries and encyclopaedias.

The distinction between these two kinds of inter-term relationships can be displayed as follows:



(A) = *a posteriori* relationships between indexing terms assigned to a document

(B) = *a priori* relationships handled by the thesaurus

This International Standard is especially concerned with those *a priori* relationships which can be displayed in a thesaurus, where they then, in effect, add a second dimension to an indexing language, as shown above.

1 Scope and field of application

1.1 The recommendations set out in this International Standard are intended to ensure consistent practice within a single indexing agency, or between different agencies (for example members of a network). They should not be regarded, however, as mandatory instructions. Optional procedures are described in many cases, for example for the display of inter-term relationships, without indicating one of these approaches as the preferred technique. The choice of procedure will vary from one indexing agency to another, depending on management decisions that fall outside the scope of this International

1) For practical purposes, "term" and "concept" are sometimes used interchangeably.

Standard. As far as possible the techniques described in this International Standard are based upon general principles which apply to any subject field. It is recognized, however, that an indexer working within a limited subject field may sometimes need to depart from these general recommendations, and this is noted where appropriate.

1.2 As far as possible the techniques described in this International Standard are not limited to a particular method of indexing, whether post-coordinate or pre-coordinate. This International Standard is, however, subject to the following restrictions:

a) it deals with the display and organization of terms that form a controlled subset of natural language. It does not suggest procedures for organizing and displaying mathematical and chemical formulae;

b) it is generally based on the concept of "preferred terms" (see 3.5);

c) its application is limited to agencies in which human indexers are used to analyse documents and express their subjects in the terms of a controlled indexing language. It is not applicable to these agencies which apply entirely automatic indexing techniques, where terms occurring in texts are organized into sets according to criteria which can be established by a computer, for example their frequency of occurrence and/or adjacency in the text. It is considered, however, that a well-constructed monolingual thesaurus can serve as a useful aid when searching such a free-text system;

d) it deals mainly with procedures for indexing collections of documents listed in catalogues or bibliographies. It is not concerned with the preparation of "back-of-the-book" indexes, although many of its recommended procedures may be useful for that purpose.

1.3 The recommendations contained in this International Standard relate to monolingual thesauri, without reference to the special requirements of multilingual thesauri, i.e. those thesauri in which conceptual equivalences are expressed in terms selected from more than one natural language. The construction and maintenance of a multilingual thesaurus is dealt with separately in ISO 5964. Since the principles on which this International Standard is based can be regarded as both language-independent and culture-independent, they have also been accepted as the basis for the multilingual standard. Consequently, general principles and procedures which apply equally to both kinds of thesauri are fully described only in this International Standard, i.e. they are not repeated in ISO 5964.

2 References

ISO 5963, *Documentation — Methods for examining documents, determining their subjects and selecting indexing terms.*

ISO 5964, *Documentation — Guidelines for the establishment and development of multilingual thesauri.*

3 Definitions

For the purposes of this International Standard, the following definitions apply.

3.1 document: Any item, printed or otherwise, that is amenable to cataloguing and indexing.

NOTE — This definition refers not only to written and printed materials in paper or microform versions (for example books, journals, diagrams, maps), but also to non-printed media (for example machine-readable records, films, sound recordings, etc.), and three-dimensional objects or realia used as specimens.

3.2 indexing language: A controlled set of terms selected from natural language and used to represent, in summary form, the subjects of documents.

3.3 thesaurus: The vocabulary of a controlled *indexing language* (see 3.2), formally organized so that the *a priori* relationships between concepts (for example as "broader" and "narrower") are made explicit.

3.4 indexing term: The representation of a concept, preferably in the form of a noun or noun phrase.

NOTE — An indexing term can consist of more than one word, and is then known as a *compound term* (see 3.7). In a controlled indexing vocabulary, a term is designated either as a *preferred term* or as a *non-preferred term*.

3.5 preferred term: A term used consistently when indexing to represent a given concept; sometimes known as "descriptor".

3.6 non-preferred term: The synonym or quasi-synonym of a preferred term. A non-preferred term is not assigned to documents, but is provided as an entry point in a thesaurus or alphabetical index, the user being directed by an instruction (for example USE or SEE) to the appropriate preferred term; sometimes known as "non-descriptor".

3.7 compound term: *An indexing term* (see 3.4) which can be factored morphologically into separate components, each of which could be expressed, or re-expressed, as a noun that is capable of serving independently as an indexing term.

NOTE — The parts of the great majority of compound 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 compound 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 compound term "printed indexes".

2) the preposition-plus-noun combination "for children" in the compound term "hospitals for children".

The focus and its difference(s) may be written as separate words, as in "dining rooms" and "soup spoons", or they may be concatenated into single words, as in "bedrooms" and "teaspoons".

3.8 node label: A "dummy" term not assigned to documents when indexing, but inserted into the systematic section of some types of thesauri to indicate the logical basis on which a category has been divided; sometimes known as a "facet indicator".

Examples:

By occupation

By purpose

Parts

NOTE — See 8.3.3 for a further description of node labels.

4 Abbreviations and symbols

4.1 The following abbreviations, which are used throughout this International Standard, are printed as prefixes to terms, etc. Each abbreviation indicates the relationship or function of the term or note which follows, as explained below:

SN Scope note; a note attached to a term to indicate its meaning within an indexing language

USE The term that follows the symbol is the preferred term when a choice between synonyms or quasi-synonyms exists

UF Use for; the term that follows the symbol is a non-preferred synonym or quasi-synonym

TT Top term; the term that follows the symbol is the name of the broadest class to which the specific concept belongs; sometimes used in the alphabetical section of a thesaurus

BT Broader term; the term that follows the symbol represents a concept having a wider meaning

BTG Broader term (generic)

BTP Broader term (partitive)

NT Narrower term; the term that follows the symbol refers to a concept with a more specific meaning

NTG Narrower term (generic)

NTP Narrower term (partitive)

RT Related term; the term that follows the symbol is associated, but is not a synonym, a quasi-synonym, a broader term or a narrower term

4.2 Abbreviations with equivalent meanings also occur in thesauri in other languages.

Examples:

French

NE Note explicative

EM Employer

EP Employé pour

MV Nom de la classe la plus générale

TG Terme générique

TGG Terme générique générique

TGP Terme générique partitif

TS Terme spécifique

TSG Terme spécifique générique

TSP Terme spécifique partitif

VA Voir aussi

German

D Definition

BS Benutzen

BF Benutzt für

SB Spitzenbegriff

OB Oberbegriff

OA Oberbegriff (Abstraktionsrelation)

SP Verbandsbegriff (Bestandsrelation)

UB Unterbegriff

UA Unterbegriff (Abstraktionsrelation)

TP Teilbegriff (Bestandsrelation)

VB Verwandter Begriff

4.3 The abbreviations listed in 4.1 and 4.2 have acquired status as generally recognized conventions, and they occur in many published thesauri. They have obvious mnemonic value, yet it is realized that they are also language-dependent. If this characteristic is regarded as sufficiently important to justify using a "neutral" system, an agency can adopt the language-independent symbols developed by ISO and given in the annex.

4.4 The following conventions are also used in examples throughout this International Standard:

a) preferred terms are printed in upper case throughout:

Examples:

CARS

ANIMALS

b) non-preferred terms are printed in lower case except when the non-preferred term is a proper name requiring an upper case initial, or an abbreviation or acronym which should be printed throughout in upper case:

Examples:

CARS
UF automobiles

ANIMALS
UF fauna

WORLD HEALTH ORGANIZATION
UF WHO

4.5 A compound term is sometimes factored morphologically into components, and these are assigned independently as indexing terms (see clause 7). If the unfactored form is likely to be sought by users, a reference should be made from the compound term as a whole to the separate terms used in combination.

Example:

coal mining **USE** COAL + MINING

5 Vocabulary control

5.1 Two principal means for achieving vocabulary control are employed in thesauri:

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 an indexing 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 definition or scope note should be appended to 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 (see 3.5) 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.

5.2 Further means for achieving vocabulary control are reviewed in the following clauses. These deal with matters such as the choice of singular or plural forms, the selection of the preferred term when synonyms are encountered, and the extent to which a compound term should either be retained in its pre-coordinated form or factored into separate components, each expressed as a noun and used independently as an indexing term.

6 Indexing terms

6.1 General

6.1.1 The concepts represented by indexing terms belong to general categories such as the following:

a) **Concrete entities**

1) Things and their physical parts

Examples:

BIRDS

LIMBS

MICROFORMS

MOUNTAIN REGIONS

2) Materials

Examples:

ADHESIVES

RUBBER

TITANIUM

b) **Abstract entities**

1) Actions and events

Examples:

GLACIATION

GOLF

MARKETING

2) Abstract entities, and properties of things, materials or actions

Examples:

ELASTICITY

NEWS

PERSONALITY

SPEED

3) Disciplines or sciences

Examples:

ARCHAEOLOGY

CHEMISTRY

4) Units of measurement

Examples:

HERTZ

KILOMETRES

c) **Individual entities, or "classes-of-one", expressed as proper nouns**

Examples:

SRI LANKA

WORLD HEALTH ORGANIZATION

6.1.2 The compiler of a thesaurus needs to be aware of these classes, since they are likely to affect some of the procedures considered in later clauses, for example the choice of singular or plural forms, and applying a test for the validity of a hierarchy.

6.2 Forms of terms

6.2.1 Nouns and noun phrases

An indexing term should preferably consist of a noun or a noun phrase. Noun phrases belong to the category of compound terms, and occur in two forms:

a) **adjectival phrases**

Example:

MARINE BIRDS

This class also includes those single-word compounds which can be factored morphologically into a noun plus a modifying difference having an adjectival function:

Examples:

FOOTBALL

MOTORWAYS

b) **prepositional phrases**

Example:

HOSPITALS FOR CHILDREN

Those parts of a compound term which function as differences [see item b) of the note to 3.7] should be considered as potential sources of extra terms in a thesaurus. When a difference consists of an adjective, the noun from which the adjective was derived should be preferred as the extra candidate term. If these terms are accepted, the thesaurus should display reciprocal relationships between the extra term and the compound term as a whole.

Examples:

a) MARINE BIRDS
RT SEAS

SEAS
RT MARINE BIRDS

b) SCHOOLS FOR HANDICAPPED CHILDREN
RT HANDICAPPED CHILDREN

HANDICAPPED CHILDREN
RT SCHOOLS FOR HANDICAPPED CHILDREN

6.2.2 Adjectives

6.2.2.1 Adjectives used alone may occur in an indexing language in special circumstances considered below, but their use should be avoided as far as possible.

6.2.2.2 Adjectives may be accepted as single words in an index or thesaurus in situations such as the following:

a) when working in a language where adjectives generally precede the nouns they modify, the user may be directed, on economic grounds, from a noun to an adjective which serves as the first component of several compound terms. For example, a reference could be made from "France" (the noun) to "French" (the adjective) if the indexing language contains a number of terms such as "French art", "French language", "French literature", "French wines". This would apply especially when the adjective and the noun from which it was derived differ widely in their spellings, for example France/French, Sea/Marine;

b) in languages where adjectives follow the nouns they determine, a reference may be made from an adjective to one or more noun phrases containing the adjective. An index in French, for example, could contain reference from an adjective such as "Pasteurisé" to compound terms such as "Crème pasteurisée", "Lait pasteurisé", and "Produits pasteurisés".

6.2.3 Adverbs

Adverbs such as "Very" or "Highly" should not be used alone as indexing terms. A phrase beginning with an adverb should not be accepted as an indexing term except when it has acquired a special meaning within a jargon.

Example:

VERY HIGH FREQUENCY

6.2.4 Verbs

Verbs expressed as infinitives or participles should not be used alone as indexing terms. Activities should be represented by nouns or verbal nouns.

Examples:

COOKERY (not "cook")

DISTILLATION (not "distil")

6.2.5 Abbreviations and acronyms

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

Example:

WORLD HEALTH ORGANIZATION
UF WHO

WHO **USE** WORLD HEALTH ORGANIZATION

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.

Example:

UNICEF
UF United Nations International Children's Fund

United Nations International Children's Fund
USE UNICEF

6.3 Choice of singular or plural forms

6.3.1 In those languages where a distinction between singulars and plurals can be made, the decision to adopt singular or plural forms as indexing terms is likely to be affected by factors such as the following:

a) post-coordinate or pre-coordinate indexing

In a pre-coordinate index, terms selected from a thesaurus are organized into index entries in such a way that the entry as a whole expresses a subject in summary form. Relationships between terms may be conveyed in various ways, for example by word order and/or the choice of special typography and punctuation. In some systems, terms may be organized into phrases linked by prepositions or other adjuncts. In these circumstances, the meaning or comprehensibility of the index entry as a whole may be affected be the use of singular or plural forms. This does not apply to a post-coordinate system, where terms are assigned to documents as independent retrieval keys, without indicating their interrelationships.

b) cultural factors

Agencies within different countries tend to work within different traditions concerning the use of singulars or plurals. In English-speaking countries, for example, terms may be expressed either as singulars or plurals depending upon factors considered below (see 6.3.2). Indexers in other language communities, for example French and German, tend to prefer the singular form where possible, so that the user can approach the thesaurus or index in the same way as a dictionary. In these cases, however, exceptions to the general preference are likely to occur for pragmatic reasons depending, for example, upon the type of indexing system used [see a) above], or an occasional need to avoid ambiguity where the singular form can refer to more than one concept, and these could be distinguished by expressing one of them as a plural.

6.3.2 In agencies where either the singular or plural form of a term may be adopted, the choice of the preferred term is generally related to the kind of concept to which the term refers. As noted earlier (see 6.1.1) terms can be divided into those that represent concrete entities, and those that refer to abstract concepts. These two classes are reviewed separately below.

6.3.2.1 Nouns that represent concrete entities can be divided into two further categories:

a) count nouns, i.e. names of countable objects that are subject to the question "How many?" but not "How much?". These should be expressed as plurals.

Examples:

DOCUMENTS

PENGUINS

POLITICAL PARTIES

WINDOWS

Special treatment is usually given to the names of parts of the body. These should be expressed as plurals when more than one occurs in a fully formed organism, but in the singular if only one is present.

Examples:

EARS	*but*	DIGESTIVE SYSTEM
HANDS		HEAD
LUNGS		NOSE