

---

# 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 monolingües*

First edition — 1974-08-15

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO 2788:1974

<https://standards.iteh.ai/catalog/standards/sist/e4f77c74-b5b7-4568-a355-4c48da04aac1/iso-2788-1974>

---

UDC 001.815 : 025.3

Ref. No. ISO 2788-1974 (E)

**Descriptors :** documentation, information retrieval, thesauri, preparation.

Price based on 13 pages

## FOREWORD

ISO (the International Organization for Standardization) is a worldwide federation of national standards institutes (ISO Member Bodies). The work of developing International Standards is carried out through ISO Technical Committees. Every Member Body interested in a subject for which a Technical Committee has been set up 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.

International Standard ISO 2788 was drawn up by Technical Committee ISO/TC 46, *Documentation*, and circulated to the Member Bodies in October 1972.

It has been approved by the Member Bodies of the following countries :

Australia	Germany	Poland
Belgium	Hungary	Romania
Brazil	India	South Africa, Rep. of
Canada	Iran	Sweden
Czechoslovakia	Ireland	Thailand
Egypt, Arab Rep. of	Israel	Turkey
Finland	Italy	United Kingdom
France	New Zealand	U.S.A.

The Member Body of the following country expressed disapproval of the document on technical grounds :

Austria

## CONTENTS

	Page
<b>0 Introduction</b> . . . . .	1
<b>1 Scope and field of application</b> . . . . .	1
<b>2 Definitions</b> . . . . .	1
2.1 Definition of thesaurus . . . . .	1
2.2 Relations to other documentation languages . . . . .	2
<b>3 Thesaurus structure</b> . . . . .	2
3.1 Descriptors . . . . .	2
3.2 Formal requirements . . . . .	2
3.3 Methods of avoiding ambiguity . . . . .	5
3.4 Descriptor interrelationships . . . . .	5
<b>4 Presentation and arrangement of the thesaurus</b> . . . . .	8
4.1 Introduction to the thesaurus . . . . .	8
4.2 Main part of the thesaurus . . . . .	8
4.3 Auxiliary parts of the thesaurus . . . . .	8
4.4 Collating sequence (alphabetization) . . . . .	9
<b>5 The process of the thesaurus construction</b> . . . . .	10
5.1 Avoidance of duplicative work . . . . .	10
5.2 Notification of intent . . . . .	10
5.3 General principles . . . . .	10
5.4 Different approaches . . . . .	10
5.5 Selection of descriptors . . . . .	11
5.6 Recording procedures . . . . .	11
5.7 Pilot run . . . . .	12
<b>6 Maintenance</b> . . . . .	12
6.1 Periodic verification . . . . .	12
6.2 Elimination of new descriptors . . . . .	12
6.3 Choice of new descriptors . . . . .	12
<b>7 References</b> . . . . .	13

iTeh STANDARD PREVIEW  
(standards.iteh.ai)

ISO 2788:1974

<https://standards.iteh.ai/catalog/standards/sst/c417/c/4-65b7-4568-a555->

ISO 2788:1974

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO 2788:1974

<https://standards.iteh.ai/catalog/standards/sist/e4f77c74-b5b7-4568-a355-4c48da04aac1/iso-2788-1974>

# Documentation – Guidelines for the establishment and development of monolingual thesauri

## 0 INTRODUCTION

In information and documentation centres there is a need for practical methods of representing concepts simply and clearly and of ordering them by clarifying their interrelationships. By applying these methods and the terminological control of concepts, it is possible to introduce an effective and cost-saving system of analysis, classification and retrieval in documentation systems.

Agencies in many countries have either compiled or are about to compile thesauri for their own purposes. It is obvious that the transfer of information requires a standard set of rules for the construction of thesauri. The guidelines set out in this International Standard will be useful:

- for a single agency, in order to facilitate construction and development of a thesaurus;
- to facilitate the transfer of information between agencies.

A list of publications related to these guidelines is given in clause 7.

Examples have been chosen in accordance with the language of the text.

## 1 SCOPE AND FIELD OF APPLICATION

**1.1** This International Standard sets out guidelines intended to facilitate the preparation and development of thesauri regardless of whether they are administered mechanically or manually. These guidelines were, therefore, prepared as an attempt to lay the basis for compatibility, both at present and in the future, of thesauri that are being elaborated in most of the disciplines of science, basic as well as applied.

**1.2** These guidelines are applicable in the sector of information and documentation for the preparation and development of monolingual thesauri for information storage and retrieval, irrespective of the subject field being dealt with. No provision has been made for mathematical and structural chemical formulae. The guidelines are drafted for general application. They should be completed in accordance with the characteristics of each particular language.

## 2 DEFINITIONS

### 2.1 Definition of thesaurus

A thesaurus may be defined either in terms of its function or its structure.

In terms of function, a thesaurus is a terminological control device used in translating from the natural language of documents, indexers, or users into a more constrained "system language" (documentation language, information language).

In terms of structure, a thesaurus is a controlled and dynamic vocabulary of semantically and generically related terms which covers a specific domain of knowledge.

As a structured subset of natural language it describes the subject content of documents, objects, or collections of data. Terms needed to describe the formal characteristics of the document or object – non-subject terms such as names of authors, corporate authors, journals, museums, locations, etc. (bibliographical description) – need not be a part of a thesaurus though such terms are usually necessary for identification or location of that which is indexed. They may, in fact, comprise an adjunct to a thesaurus.

A particular thesaurus should accurately reflect the information content of the body of documents or other items in a collection to which the thesaurus applies. It should contain terms and cross-references appropriate to the subject matter, taking into consideration both the language of the document collection and the language and the information needs of the users.

Based on the nature of the terminological control, two major types of thesauri are currently being developed:

- a) thesauri which perform terminological control by preferred terms, i.e. thesauri in which only one of the terms denoting a concept is permitted for indexing and retrieval, and
- b) thesauri which perform terminological control by allowing all terms denoting a concept to be used for indexing and retrieval, but which assign these terms to another unambiguous representation of the concept (e.g. concept number, notation).<sup>1)</sup> This type of thesaurus makes it easier to cancel earlier synonyms and concept relationships if a term changes its meaning.

1) While thesauri using preferred terms can be maintained manually, thesauri not using preferred terms require machine maintenance and retrieval.

## 2.2 Relations to other documentation languages

In contrast to a dictionary, which provides definitions for given words or terms, a thesaurus provides words or terms to express meanings that are implied by the term relationships given in the thesaurus.

A thesaurus may be arranged like an alphabetical index and the terms in a thesaurus may be used to construct an index. However, the thesaurus itself is not an index. An index to a collection must have addresses or locators for items in the collection associated with each term, but a thesaurus contains the terms only, without the addresses or locators of an index. A thesaurus classifies terms by arranging them in hierarchical classes. As a "term classification system" a thesaurus has some similarities with subject matter classification systems, as represented, for example, by the Universal Decimal Classification.

But whereas hierarchical subject classification systems try to show the whole system of hierarchical relations, the thesaurus shows relations necessary for indexing and retrieval according to the body of documents and the information needs of users.

A thesaurus is one kind of authority list, that is, the preferred terms in a particular thesaurus are required indexing and retrieval terms for a given information and documentation system. There are other kinds of natural-language-based authority lists, such as subject heading lists. In general, however, these do not have the hierarchical structure of the thesaurus.

## 3 THESAURUS STRUCTURE

### 3.1 Descriptors

The internal form of individual entries and the arrangement of the various entries in relation to one another constitute the structure of a thesaurus. Cross-references in a thesaurus make explicit the ways in which entries relate to each other in a network of concepts.

The terms permitted by a thesaurus for use in indexing are called descriptors. In thesauri using preferred terms, the preferred terms are the descriptors. The descriptor can be characterized as an authorized and formalized term or symbol in a thesaurus, used to represent unambiguously the concepts of documents and queries.

With regard to the aim of international cooperation, two levels of descriptors may be distinguished :

- more general descriptors, which could be the subject of multinational agreement after translation into several languages;
- more specific descriptors, which could be the subject of agreement in several specific fields.

Descriptors may be :

- a) terms denoting concepts or concept combinations;

b) terms denoting individual entities. These terms are also called proper names (or identifiers). Proper names may be :

- project names;
- nomenclatures;
- identification numbers or symbols;
- geographical or geopolitical names;
- trade marks;
- names of persons and organizations;
- abbreviations and acronyms;
- other proper names (e.g. programming systems).

It is advisable to use proper names in the same way as other descriptors, i.e. to interrelate them. The same may apply when internationally agreed nomenclatures are integrated into the thesaurus.

In thesauri not using preferred terms, all terms included in the thesaurus may, in principle, be descriptors.

The concept-denoting terms not permitted in indexing must be regarded as unauthorized terms. They are called non-descriptors.

In most cases it would be helpful to provide the possibility of formal distinction between descriptors and non-descriptors. This can be achieved by using a special type for descriptors in print-out or by using special symbols marking the beginning and the end of a descriptor.

### 3.2 Formal requirements

#### 3.2.1 Compound expressions

A descriptor may consist of one or more words. As a general rule, the descriptor should reflect the terminology of the subject, irrespective of the number of words required to denote the concept, but it is desirable that the descriptor should contain as few words as possible, and preferably only one. It has to be borne in mind, however, that in the course of abbreviation a term may lose something of its clarity. The words of compound descriptors should be entered preferably in their natural word order (e.g. electrical engineering); i.e. not artificially inverted. It may be helpful to include the inverted forms as non-descriptors preferentially related to the descriptors.

When the inverted form is chosen for the entry, i.e. the descriptor, the inclusion of the non-inverted form as a synonym is necessary.

#### 3.2.2 Representation of concepts by several descriptors

To keep the number of descriptors within limits, it may sometimes be useful to represent concepts or combinations of concepts by a combination of descriptors.

### 3.2.2.1 FACTORING (SPLITTING) OF CONCEPTS OR WORD COMBINATIONS

As a general rule there are two possibilities :

#### a) MORPHOLOGICAL (LEXICOLOGICAL) FACTORING

If the concept is represented by a compound or a combination of words, it is often possible to break it down into its component parts. When the words thus obtained are combined, however, they only represent the original concept if the morphological factoring coincides with the semantic factoring. Only in this case can a correct result be obtained by morphological factoring.

#### b) SEMANTIC FACTORING

In semantic factoring a concept is split into components which when combined reproduce the original concept. The descriptors used are not necessarily contained in the terms.

Example of a)

ANIMAL PSYCHOLOGY =  
ANIMALS + PSYCHOLOGY

Example of b)

TRADE WIND = METEOROLOGY + GEOGRAPHY

Negative example of a) ( $\neq$  b)

TRADE WIND  $\neq$  TRADE + WIND

### 3.2.2.2 WORD COMBINATIONS OR CONCEPTS

Terms may be combined (or precoordinated) before they enter the system and hence must be retrieved as such, or combined during search (postcoordinated) to represent a sought concept. Decisions for pre- or postcoordination may be taken according to the following considerations :

#### a) POSTCOORDINATION

This should be confined to cases where the simple descriptors are not used too frequently on their own and/or the precombined descriptor is used very seldom. Care must be taken that the combination of simple descriptors really represents the same concept as the precombined descriptor and results in an unambiguous representation of the concept.

Negative example :

TRANSFORMERS + OIL = TRANSFORMER OIL  
or OIL TRANSFORMER

#### b) PRECOORDINATION

Precombined descriptors should always be used when

- the meanings of the simple descriptors on their own differ from their meaning in the precombined descriptor.

Example : COLLARBONES

- the simple descriptors are used in hierarchical connections other than the precombined descriptors.

Example : HOUSEWIVES

- the precombined descriptor is a proper name, or doubt exists as to whether the combination of simple descriptors reflects the conceptual content exactly and exclusively.

In most cases it is recommended to enter terms which are broader terms to other descriptors as precombined descriptors : if a concept is represented by the combination of simple descriptors, this should be expressed by the "USE" reference. Under the simple descriptors as specified "UF" reference (e.g. UFC) has to be made to the unused precombined term.

Example :

SHIPBUILDING	USE	SHIPS + BUILDING
VESSELS	USE	SHIPS
SHIPS	UF	VESSELS
	UFC	SHIPBUILDING
BUILDING	UFC	SHIPBUILDING

The combination of simple descriptors must be included in the systematic sections of the thesaurus, and the unused precombined descriptor in the alphabetical sections as non-descriptor.

### 3.2.3 Word form

Once it has been decided to include a given term in the thesaurus, care should be taken to ensure that, taken together with the relations represented, it conveys the intended meaning as accurately as possible.

#### a) SPELLING

The most widely accepted spelling of the word should be adopted. In cases where, due to varying usage, more than one spelling of a word is accepted, both spellings should be included in the thesaurus and reference made from one to the other.

Example :

SULFUR  
SULPHUR

Alternatively, a well-established dictionary can be chosen to act as arbitrator whenever this problem arises.

#### b) TRANSLATION

Many current technical terms have arisen by translation from other languages, but sometimes a modern foreign language, Latin or Greek term is incorporated into the specialized vocabulary for a particular subject. When both the foreign language term and its putative translation coexist with the same meaning, both should be included in the thesaurus and reference made from one to the other.

Example :

BRAKING RADIATION  
BREMSTRALHUNG



c) TRANSLITERATION

The problem is further complicated when the foreign language in question is written in a different alphabet. This is particularly true in the case of identifiers. The transliteration standards recommended by the International Organization for Standardization should be used whenever applicable. Wherever a choice exists, the transliteration which does not employ diacritical marks should be selected.

Example :  
 SATELLITES  
 SPUTNIKS

3.2.4 Noun form

Descriptors should be preferably in the form of a noun (or a noun phrase) or that form of the verb which is grammatically equivalent.

Example :  
 "DEMOCRACY" instead of "DEMOCRATIC"  
 "EXECUTION" instead of "EXECUTE"

3.2.5 Number

The use of the singular or the plural form of descriptors should be decided in accordance with the usage in the language of the thesaurus. It is necessary to establish and follow national standards for this decision. Where no national standard exists, consideration should be given to the following of the lexicographical practice within this language. Sometimes the singular and plural forms of a word denote different concepts; in this case both should be entered.

Example :  
 WOOD  
 WOODS

In English, in general, the plural form should be used for descriptors, particularly when generic terms are involved (i.e. when the descriptor denotes classes of things). The singular form is used for specific material or property terms (attributes), process terms, proper names and disciplinary areas.

Examples :

Processes	Properties	Classes of things
ACIDIFICATION	CONDUCTIVITY	TEETH
CALENDERING	OPACITY	STARS
CURING	TEXTURE	PAINTINGS

3.2.6 Adjectives

There are, of course, a certain number of cases where only adjectives or other non-noun forms can be used.

Example :  
 SOCIAL  
 INTERNATIONAL

A small proportion of single-word terms in adjectival form may be useful as modifiers (continuous, horizontal). Since adjectives can be precoordinated with nouns and entered as compound descriptors, the choice to enter adjectives singly should be dictated by considerations of practicability and flexibility. Precoordination is recommended whenever a modifier appears very frequently in combination with another particular term.

3.2.7 Abbreviations and acronyms

In general, abbreviated forms of terms should be avoided because their use may not be general enough, their meaning may be dependent on context, or their recognition may be dependent on capitalization and periods which become constraints if computer printers or other electronic data processing equipment is used in conjunction with the thesaurus. Therefore they should be used only when their meanings are well established within the group of users concerned or their meaning is internationally established and when significant gains in practicality can be demonstrated.

Abbreviated and unabbreviated forms of a given term should be treated as synonyms and cross-referenced accordingly.

Abbreviations with several meanings are to be treated as homonyms (homographs).

Sometimes the necessity of limiting the length of the descriptor entails the use of less well established abbreviations. In all these cases a scope note should be appended. Well established acronyms are acceptable as descriptors.

Examples : RADAR, LASERS

3.2.8 Use of characters

a) CHARACTER SET

The eventual use of electronic data processing equipment may entail

- the use of only the upper-case format for the descriptors;
- avoidance of diacritical marks;
- limitation of the number of characters that a descriptor may have.

These restrictions will disappear when the electronic data processing equipment is adapted to the needs of information.

b) PUNCTUATION

Punctuation marks in descriptors should be minimized. Except for specialized nomenclature, only parentheses and the hyphen are needed in descriptors, as specified in the following guidelines.

iTeh STANDARD PREVIEW  
 (standards.iteh.ai)

ISO 2788:1974

<https://standards.iteh.ai/catalog/standards/sist/ea77c74-1307-560-a339-4c48da04aac1/iso-2788-1974>



Full stops<sup>1)</sup> should be allowed only when, due to a limit on the descriptor, a word has to be truncated. Hyphens should be used only when their omission would alter the intended meaning of the descriptor. Commas, colons and apostrophes should be excluded when they are not necessary to convey the meaning of the terms. In case of inversion the use of commas may be allowed. Where punctuation marks are omitted, it is advisable to include them in full in the scope notes.

### c) SPECIAL CHARACTERS AND NUMERALS

When it is considered necessary to use special characters other than hyphens and parentheses in descriptors, their meanings should be clearly defined. Special characters other than those mentioned above may be used in scope notes, definitions and other forms of additional information. If the descriptors contain numerical elements, arabic numerals should be used. The position of the numerals should follow normal usage. Rules must be established for the treatment of diacritics, subscripts and superscript numerals.

## 3.3 Methods of avoiding ambiguity

### 3.3.1 Homonyms (homographs)

The different meanings of homonyms (homographs) must be marked and distinguished by specifying symbols or terms (qualifiers) which should be placed between parentheses immediately after the homonym as part of the descriptor. They can be specifying terms, being themselves not homonyms, or other usable signs. Homonym and bracketed qualifiers form a compound descriptor.

Example :

BEAMS (ELECTROMAGNETIC)  
BEAMS (STRUCTURAL)

When a term defined in an internationally or nationally standardized technical vocabulary is selected as a descriptor, it should be written without substantial change, including those parts of the term shown between brackets in the corresponding entry of the vocabulary.

### 3.3.2 Scope notes

A scope note is a brief explanation of the intended use of a descriptor. It may accompany the descriptor in the main part of the thesaurus, but does not form part of the descriptor.

Scope notes may be used

- to restrict the usage of a descriptor;
- to explain abbreviations and acronyms;
- to exclude a possible meaning from a term, especially for terms which are in common use in different disciplines;

- to date addition and deletion of terms and to record changes in the handling of terms.

Scope notes should be indicated by special characters and clearly distinguished from qualifiers.

Examples :

COPPER ALLOYS  
(alloys in which copper is the principal constituent)  
MICROWAVE FREQUENCIES  
(1 to 300 GHz)  
MAMMALS  
(only the species on land)

### 3.3.3 Definitions

The conceptual content of a descriptor in a thesaurus is indicated mainly by the represented relations to other thesaurus words. Whenever there is doubt regarding the unique interpretation of a descriptor, a definition should be added, specifying the exact conceptual content, which may accompany the descriptor in the main part of the thesaurus, but does not form part of the descriptor.

Example :

DOCUMENTATION  
(the process of storing and retrieving information in all fields of learning)

### 3.3.4 Translations

In many cases it will be helpful to show the equivalent terms in other languages to ensure that the descriptor is correctly used in the analysis of foreign language texts. Where the meaning is not entirely equivalent, attention can be drawn to this in the form of an explanatory note. Translations of single descriptors should be treated as synonyms or quasi-synonyms.<sup>2)</sup>

### 3.3.5 Source of information

Information on the source of a descriptor or a definition can be very important for the further development of the thesaurus. The source information could therefore be collected together with the descriptors but need not be included in the printed main part of the thesaurus.<sup>3)</sup>

## 3.4 Descriptor interrelationships

### 3.4.1 General

By definition, an indispensable function of a thesaurus is to represent the interrelationships between concepts by representing the interrelationships between the words used to denote them. The network of relations of one descriptor to other descriptors thus provides a kind of definition by placing the descriptor into the semantic space.

1) It may be worthwhile to apply another sign depending on the machine used.

2) The problem of translation of thesauri will be treated in special guidelines (in preparation).

3) ISO 3166, Code for the representation of names of countries. (At present at the stage of draft.)