

---

# 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.

# Contents

	Page
<b>0</b> Introduction .....	1
<b>1</b> Scope and field of application .....	1
<b>2</b> References .....	2
<b>3</b> Definitions .....	2
<b>4</b> Abbreviations and symbols .....	3
<b>5</b> Vocabulary control .....	4
<b>6</b> Indexing terms .....	4
<b>6.1</b> General .....	4
<b>6.2</b> Forms of terms .....	5
<b>6.3</b> Choice of singular or plural forms .....	6
<b>6.4</b> Homographs or polysemes .....	7
<b>6.5</b> Choice of terms .....	7
<b>6.6</b> Scope notes and definitions .....	9
<b>7</b> Compound terms .....	9
<b>7.1</b> General .....	9
<b>7.2</b> Terms that should be retained as compounds .....	10
<b>7.3</b> Terms that should be syntactically factored .....	11
<b>7.4</b> Order of words in compound terms .....	13
<b>8</b> Basic relationships in a thesaurus .....	13
<b>8.1</b> General .....	13
<b>8.2</b> The equivalence relationship .....	13
<b>8.3</b> The hierarchical relationship .....	15
<b>8.4</b> The associative relationship .....	17
<b>9</b> Display of terms and their relationships .....	19
<b>9.1</b> General .....	19
<b>9.2</b> Alphabetical display .....	20
<b>9.3</b> Systematic display .....	20
<b>9.4</b> Graphic display .....	24
<b>10</b> Management aspects of thesaurus construction .....	29
<b>10.1</b> Methods of compilation .....	29
<b>10.2</b> Recording of terms .....	30
<b>10.3</b> Term verification .....	30
<b>10.4</b> Specificity .....	30
<b>10.5</b> Admission and deletion of terms .....	30
<b>10.6</b> The use of automatic data processing equipment .....	30
<b>10.7</b> Form and contents of a thesaurus .....	31
<b>10.8</b> Other editorial matters .....	31
<b>Annex</b> — Symbolization of thesaural relationships .....	32

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.<sup>1)</sup>

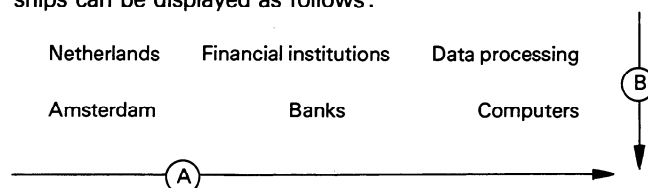
The first of these tasks is described separately in ISO 5963. The second and third tasks concern not only the indexer but also the user of the index. This International Standard deals with some aspects of term selection, since it contains recommended procedures for vocabulary control, but it is particularly 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".

The remainder of this document  
is available for purchase online at

**[www.saiglobal.com/shop](http://www.saiglobal.com/shop)**

SAI Global also carries a wide range of publications from a wide variety of Standards Publishers:



**SAI GLOBAL**



Click on the logos to search the database online.