International Standard ISO 5963

Documentation — Methods for examining documents, determining their subjects, and selecting indexing terms

Documentation — Méthodes pour l'analyse des documents, la détermination de leur contenu et la sélection des termes d'indexation

First edition — 1985-12-01

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 5963 was prepared by Technical Committee ISO/TC 46, *Documentation*.

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

This page intentionally left blank

Documentation — Methods for examining documents, determining their subjects, and selecting indexing terms

1 Scope and field of application

1.1 This International Standard describes recommended procedures for examining documents, determining their subjects, and selecting appropriate indexing terms. It is restricted to these preliminary stages of indexing, and does not deal with the practices of any particular kind of indexing system, whether pre-coordinated or post-coordinated. It also describes general techniques for document analysis which should apply in all indexing situations. These methods are, however, especially intended for indexing systems in which the subjects of documents are expressed in summary form, and where concepts are recorded in the terms of a controlled indexing language. In this context, a controlled language usually refers to a subset of terms selected from natural language, and regulated, for example, by a thesaurus. These methods would apply, however, to systems in which concepts are represented for retrieval purposes by symbols chosen from the schedules of a classification scheme.

1.2 The techniques described in this International Standard can be employed by any agency in which human indexers analyse the subjects of documents and express these subjects in indexing terms. They do not apply to agencies which employ those automatic indexing techniques in which terms occurring in texts are organized into sets or classes according to criteria which can be established by a computer, for example frequency of occurrence and/or adjacency in the text, although the aims of these systems are the same.

1.3 This International Standard is intended primarily as a guide to indexers during the stages of document analysis and concept identification. It may also be helpful for the analysis of users' enquiries and their translation, for retrieval purposes, into the controlled terms of an indexing language, and it could function as guidance to abstractors during the preparation of abstracts. It should be borne in mind, however, that although these tasks are analogous they are not identical.

1.4 This International Standard is intended to promote standard practice

a) within an agency or network of agencies;

b) between different indexing agencies, especially those which exchange bibliographic records.

2 References

ISO 2788, *Documentation — Guidelines for the establishment and development of monolingual thesauri.*

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, which is amenable to cataloguing or 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-print media (for example machine-readable records, films, sound recordings), and three-dimensional objects or realia used as specimens.

3.2 concept: A unit of thought.

The semantic content of a concept can be re-expressed by a combination of other and different concepts, which may vary from one language or culture to another.

3.3 subject: Any concept or combination of concepts representing a theme in a document.

3.4 indexing term: The representation of a concept in the form of either

— a term derived from natural language, preferably a noun or noun phrase, or

— a classification symbol.

NOTE — An indexing term can consist of more than one word. In a controlled indexing language, 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 an index, the user being directed by an instruction (for example USE or SEE) to the preferred term; sometimes known as "non-descriptor".

3.7 index (plural "indexes"): An alphabetical or systematic listing of subjects which refers to the position of each subject in a document or collection of documents.

3.8 indexing: The act of describing or identifying a document in terms of its subject content.

4 Operation and purpose of indexing

4.1 Indexing is not concerned with the description of a document as a physical entity (for example by stating the form, publisher, date, etc.), although these factors may be included in a subject index if this information would enable a user to determine more accurately whether or not a given document is likely to be relevant to his enquiry.

4.2 During indexing, concepts are extracted from documents by a process of intellectual analysis, then transcribed into indexing terms. Both analysis and transcription should be performed with the aid of indexing tools such as thesauri and classification schemes.

4.3 Essentially, indexing consists of the following three stages, although these tend to overlap in practice:

a) examining the document and establishing its subject content;

b) identifying the principal concepts present in the subject;

c) expressing these concepts in the terms of the indexing language.

Each of these stages, together with a section on quality control, is considered below in clauses 5 to 8.

5 Examining the document

5.1 The thoroughness with which a document can be examined depends to a large extent upon its physical form. Two different cases can be distinguished, i.e. printed and non-print documents.

5.2 Printed documents represent the usual case in libraries and information centres where the stock consists largely of monographs, journals, reports, conference proceedings, etc. Ideally, full understanding of these documents depends upon an extensive reading of the texts. A complete reading is often impracticable, nor is it always necessary, but the indexer should ensure that no useful information has been overlooked. Important parts of the text need to be considered carefully, and particular attention should be paid to the following:

a) the title;

b) the abstract, if provided;

c) the list of contents;

d) the introduction, the opening phrases of chapters and paragraphs, and the conclusion;

e) illustrations, diagrams, tables and their captions;

f) words or groups of words which are underlined or printed in an unusual typeface.

All these elements should be scanned and assessed by the indexer during his study of the document. Indexing from the title alone is not recommended, and an abstract, if available, should not be regarded as a satisfactory substitute for an examination of the text. Titles may be misleading; both titles and abstracts may be inadequate; in many cases neither is a reliable source of the kind of information needed by an indexer.

5.3 Non-print documents, such as audio-visual, visual and sound media, including realia, call for different procedures. It is not always possible in practice to examine a record in its entirety (for example by running a film). Indexing is then usually carried out from a title and/or synopsis, though the indexer should be allowed to view or hear a performance of the medium if the written description is inadequate or appears to be inaccurate.

6 Identification of concepts

6.1 After examining the document, the indexer should follow a systematic approach to the identification of those concepts which are essential elements in a description of its subject. Agencies should establish check lists of those factors which are recognized as important in the field covered by the index.

The questions listed below illustrate general factors which such a check-list should establish:

a) Does the document deal with the object affected by the activity?

b) Does the subject contain an active concept (for example an action, an operation, a process, etc.)?

c) Is the object affected by the activity identified?

d) Does the document deal with the agent of this action?

e) Does it refer to particular means for accomplishing the action (for example special instruments, techniques or methods)?

f) Were these factors considered in the context of a particular location or environment?

g) Are any dependent or independent variables identified?

h) Was the subject considered from a special viewpoint not normally associated with that field of study (for example a sociological study of religion)?

These are offered as examples of general factors which are likely to apply in any subject field. Other questions may need to be formulated within a special discipline.