

NOTA IMPORTANTE:

La entidad sólo puede hacer uso de esta norma para si misma, por lo que este documento NO puede ser reproducido, ni almacenado, ni transmitido, en forma electrónica, fotocopia, grabación o cualquier otra tecnología, fuera de su propio marco.

ININ/ Oficina Nacional de Normalización

METODOS PARA EL ANALISIS DE DOCUMENTOS, DETERMINACION DE SU CONTENIDO Y SELECCION DE LOS TERMINOS DE INDIZACION (ISO 5963: 2000 IDT)

Information and documentation. Methods for examining documents,
determining their subjects and selecting indexing terms

Descriptores: Documento; Determinación; Selección;
Terminología; Indización.

1. Edición Diciembre 2000

ICS: 01.140.20; 01.040.01

REPRODUCCION PROHIBIDA

Oficina Nacional de Normalización (NC) Calle E No. 261 Vedado, Ciudad de La Habana.
Teléf.: 30-0835 Fax: (537) 33-8048 E-mail: ncnorma@ceniai.inf.cu

Prefacio

La Oficina Nacional de Normalización (NC), es el Organismo Nacional de Normalización de la República de Cuba que representa al país ante las Organizaciones Internacionales y Regionales de Normalización.

La preparación de las Normas Cubanas se realiza generalmente a través de los Comités Técnicos de Normalización. La aprobación de las Normas Cubanas es competencia de la Oficina Nacional de Normalización y se basa en evidencias de consenso.

Esta Norma Cubana:

- Ha sido elaborada por el Comité Técnico de Normalización NC/CTN 15 de Información y Documentación en el que están representadas las instituciones siguientes:

Instituto de Información Científica y Tecnológica (IDICT) - CITMA
Centro de Tecnología y Calidad-SIME
Biblioteca Nacional "José Martí"-MINCULT
Sistema de Información-MINFAR
Centro Nacional de Información de Ciencias Médicas-MINSAP
Centro de Información-MICONS
Instituto de Investigaciones en Normalización (ININ) - ONN
Oficina Nacional de Normalización

- Es una adopción idéntica por el método de traducción de la Norma Internacional ISO 5963:1985 Information and documentation – Methods for examining documents, determining their subjects and selecting indexing terms
- Sustituye a la NC 39-22: 1984 “Indización manual de documentos”
- Consta del Anexo A, informativo.

© **NC, 2000**

Todos los derechos reservados. A menos que se especifique, ninguna parte de esta publicación podrá ser reproducida o utilizada por alguna forma o medios electrónicos o mecánicos, incluyendo las fotocopias o microfilmes, sin el permiso previo escrito de:

**Oficina Nacional de Normalización (NC).
Calle E No. 261 Ciudad de La Habana, Habana 3. Cuba.**

Impreso en Cuba

METODOS PARA EL ANALISIS DE DOCUMENTOS, DETERMINACION DE SU CONTENIDO Y SELECCION DE LOS TERMINOS DE INDIZACION

1 Alcance

1.1 Esta norma recomienda procedimientos para el análisis de documentos, determinación de su contenido y selección de los términos de indización. Se imita a las primeras etapas de la indización y es independiente de la práctica del sistema de indización, ya sea pre-coordinado o post-coordinado. Describe métodos generales de análisis de documentos que deben aplicarse en cualquier situación. Sin embargo, estos métodos están destinados especialmente a los sistemas de indización en los que las materias de los documentos se expresan de forma abreviada con ayuda de los términos de un lenguaje de indización controlado. En este contexto, el lenguaje controlado está constituido generalmente por un subconjunto de términos extraídos del lenguaje natural y estructurado, por ejemplo, mediante un tesaurus. Estos métodos se pueden aplicar, normalmente, a sistemas en los que, con fines de recuperación, los conceptos se representan por símbolos seleccionados del esquema de clasificación.

1.2 Las técnicas descritas en esta norma se pueden utilizar por cualquier organismo en el que se empleen indizadores humanos para analizar temáticamente el documento y expresar el contenido en forma de términos de indización. No se aplican en instituciones que utilizan técnicas de indización automáticas en las que los términos existentes en los textos se organizan en conjuntos o clases según criterios que pueden aplicarse mediante un ordenador, por ejemplo, por frecuencia de aparición y/o adyacencia en el texto, aunque la finalidad de estos sistemas sea la misma.

1.3 Esta norma debe, en primer lugar, servir de guía a los indizadores para las etapas de análisis de los documentos e identificación de los conceptos. Puede aplicarse también a la búsqueda documental para transformar las peticiones de los usuarios en términos de indización controlados. También puede servir como guía para la elaboración de resúmenes analíticos, teniendo presente, sin embargo, que estas tareas, aunque análogas, no son idénticas.

1.4 Esta norma se destina a promover la utilización de una práctica normalizada

- a) en una institución o una red de centros o instituciones;
- b) en diferentes servicios de indización, en especial en aquellos que intercambian registros bibliográficos.

2 Términos y definiciones

Para los fines de esta norma, se aplican las definiciones siguientes:

2.1 Documento

Cualquier fuente de información, impresa o no, que se pueda catalogar o indizar.

NOTA: Esta definición se refiere no sólo a los materiales escritos e impresos en papel o microforma (por ejemplo, libros, revistas especializadas, diagramas, mapas, etc.), sino también a me-

dios no impresos (por ejemplo, registros legibles por ordenador, películas y grabaciones sonoras) ya objetos de colección.

2.2 Noción o concepto

Una unidad de pensamiento. El contenido semántico de un concepto puede ser expresado por una combinación de otros diferentes que pueden variar de un idioma a otro.

2.3 Materia

Cualquier concepto o combinación de conceptos que representa el tema de un documento.

2.4 Término de indización

La representación de un concepto en forma de:

- un término derivado del lenguaje natural, preferiblemente un sustantivo simple o compuesto.
- un código de clasificación.

NOTA: Un término de indización puede constar de más de una palabra. En un lenguaje de indización controlado, un término se designa como descriptor o no descriptor.

2.5 Descriptor

Término usado siempre, en la indización, para representar un concepto dado, conocido también como "término preferente".

2.6 No-descriptor

Sinónimo o cuasi sinónimo de un descriptor. Los no-descriptores no se asignan a los documentos pero pueden servir como puntos de entrada en un Índice, dirigiendo al usuario mediante una instrucción (por ejemplo: ver o véase) al descriptor. Se llaman, también "términos no preferentes".

2.7 Índice

Lista alfabética y sistemática de materias que señala el lugar en que se encuentra cada materia en un documento o en una colección de documentos'.

2.8 Indización

Acción de describir o identificar un documento en relación con su contenido.

3 Proceso de indización

3.1 La indización no concierne a la descripción de un documento como entidad física (por ejemplo, no indica la forma, editor, fecha, etc.), aunque estos factores pueden estar incluidos en un índice de materias si esta información puede permitir a un usuario determinar, de forma más precisa, si un documento dado es relevante para su necesidad de información.

3.2 Durante la indización los conceptos se extraen del documento mediante un proceso de análisis intelectual y después se transforman en términos de indización. Tanto el análisis como la transcripción deben realizarse con ayuda de herramientas de indización, como tesauros y sistemas de clasificación.

3.3 La indización consiste esencialmente en tres etapas, que tienden a solaparse en la práctica:

- a) examen del documento y determinación de su contenido;
- b) identificación y selección de los conceptos principales del contenido;
- c) selección de los términos de indización.

Cada una de estas etapas, junto con un capítulo sobre el control de calidad, son tratadas en los capítulos 4a 7.

4 Examen del documento

4.1 La precisión con que se puede examinar un documento depende en gran manera de su forma física. Se pueden distinguir dos casos diferentes: documentos impresos y documentos no impresos.

4.2 Los documentos impresos constituyen el material habitual de las bibliotecas y centros de documentación cuyo fondo consiste principalmente en libros, revistas, informes, actas de congresos, etc. De forma ideal la comprensión completa de estos documentos requiere su lectura detallada. Sin embargo, una lectura completa es a menudo impracticable y no siempre necesaria, pero el indizador debe asegurarse de que no se ha descuidado ninguna información útil. Las partes importantes del texto deben examinarse cuidadosamente y se debe prestar especial atención a las siguientes:

- a) título;
- b) resumen, si lo tiene;
- c) sumario o tabla de contenido;
- d) introducción, párrafos iniciales de los distintos capítulos o apartados y conclusiones;
- e) ilustraciones, diagramas, tablas y su leyenda o explicación;
- f) palabras o frases que están destacadas mediante una tipografía diferente o subrayadas.

Todos estos elementos deben examinarse cuidadosamente por el indizador durante el estudio del documento. No se recomienda la indización a partir del título solamente, y el resumen si existe, no se debe considerar como un sustituto del examen del texto. Hay títulos que pueden inducir a errores; existen resúmenes que son insuficientes y ni los unos ni los otros constituyen una fuente segura del tipo de información que necesita el indizador.

4.3 Los documentos no impresos, tales como medios audiovisuales, visuales y sonoros, requieren procedimientos diferentes. En la práctica no siempre es posible examinarlos en su totalidad. La indización hay que realizarla en estos casos a partir del título y de la sinopsis o reseña. Sin embargo si estos son inadecuados o insuficientes, el indizador debe visualizar o escuchar el documento.

5 Identificación de los conceptos

5.1 Después de examinar el documento el indizador debe identificar las nociones que son elementos esenciales de la descripción del contenido. Las instituciones que patrocinan la realización del índice deben establecer los factores que se consideran importantes en el campo temático cubierto por el índice.

Algunas cuestiones que ilustran los ejemplos de criterios a retener son:

- a) ¿Trata el documento de algún objeto sometido a una acción?
- b) ¿Contiene algún concepto activo? (por ejemplo, una acción, un procedimiento, etc.)
- c) ¿Se ve afectado el objeto por la acción identificada?
- d) ¿Trata del agente causante de la acción?
- e) ¿Se describen los medios para llevar a cabo la acción? (por ejemplo, instrumentos, técnicas o métodos especiales)
- f) ¿Existen factores considerados en un medio o lugar particular?
- g) ¿Se identifican variables dependientes o independientes?
- h) ¿Se trata el tema desde un punto de vista particular no asociado normalmente a ese campo? (por ejemplo, estudio de la religión desde un punto de vista sociológico).

Estos son ejemplos de criterios susceptibles de aplicación en muchos campos; en disciplinas particulares puede ser necesario formular otras cuestiones.

5.2 El indizador no tiene necesariamente que utilizar como términos de indización todos los conceptos identificados durante el examen del documento. La selección o el rechazo de conceptos depende de la finalidad con que se van a utilizar los términos de indización, que pueden variar desde la producción de Índices alfabéticos impresos hasta la creación de una base de datos bibliográfica informatizada. La identificación de conceptos puede también estar influida, como se indicó anteriormente, por el documento a indizar. Por ejemplo la indización de libros puede diferir de la de artículos de revistas.

Las dos características de un Índice más afectadas por la selección de los términos de indización son la exhaustividad y la especificidad.

5.3 La exhaustividad esta relacionada con el número de conceptos que se tienen en cuenta, y que caracterizan el contenido íntegro de un documento.

5.3.1 Un indizador que sigue los procedimientos indicados antes, debe poder identificar todos los conceptos de un documento, que tienen valor potencial para los usuarios de un sistema de información. En algunos casos, en un mismo documento, se presentan independientemente dos o más temas dentro del campo cubierto por la indización. En ese caso los temas deben tratarse de forma separada y, si es necesario, por diferentes especialistas.

5.3.2 La cobertura de la indización no debe interpretarse de una forma demasiado estricta. Hay que tener en cuenta qué términos de indización creados inicialmente para un grupo de usuarios (por ejemplo, científicos y técnicos) pueden utilizarse por otros grupos (por ejemplo, economistas). Se aconseja que los indizadores de literatura científica y técnica tengan presentes Otros aspectos del tema, en particular, los sociales y económicos.

5.3.3 El principal criterio de selección de conceptos debe ser su valor potencial como elemento de expresión del tema del documento para su recuperación.

En la selección de conceptos, el indizador debe tener en mente las preguntas que se pueden hacer al sistema de información, en la medida en que dichas preguntas se pueden conocer. En efecto, este criterio constituye la principal función de la indización. Dentro de este contexto el indizador debe:

- a) elegir las nociones mas apropiadas para un grupo de usuarios dado, sin perder de vista el objetivo de la indización;
- b) modificar, si es necesario, tanto las herramientas como el procedimiento de indización, como resultado de las preguntas hechas al sistema. Dichas modificaciones no deben producir distorsión de la estructura o de la lógica del lenguaje de indización.

5.3.4 El número de términos o descriptores que se pueden asignar a un documento no debe limitarse de forma arbitraria. Debe determinarse enteramente por la cantidad de información contenida en el documento en relación con las necesidades supuestas de los usuarios a que va destinado el índice. Imponer un límite arbitrario puede conducir a una pérdida de objetividad en la indización y a una deformación de la información que se podrá utilizar en la recuperación.

Si por imposiciones establecidas es necesario limitar el número de términos, la selección de conceptos debe ser guiada por el juicio del indizador sobre el papel de cada término para expresar el contenido total del documento.

5.4 La especificidad está relacionada con la exactitud con que un concepto particular que aparece en un documento está representada por un término de indización. Se produce una pérdida de especificidad cuando un concepto particular está representado por un término que tiene un significado más general.

Las nociones deben identificarse de la forma más específica posible. Sin embargo, pueden preferirse nociones más generales en los casos siguientes:

- a) cuando el indizador considere que un exceso de especificidad puede actuar de forma negativa sobre el sistema de indización. (Por ejemplo, puede decidir que un modelo muy específico de una máquina se represente por el término más genérico de ese tipo de máquinas, en especial cuando esas nociones aparecen sólo en áreas muy restringidas del campo temático cubierto por el índice);
- b) cuando se trate de una idea no completamente desarrollada, o de la que se hace sólo una alusión por el autor, estará justificada la indización a un nivel más general.

6 Selección de los términos de indización

6.1 Cuando los conceptos se traducen en términos de indización, el indizador debe observar las reglas siguientes:

- a) Los conceptos ya presentes en el lenguaje de indización deben retenerse como descriptores.
- b) Los términos que representan nuevos conceptos deben comprobarse, en cuanto a su exactitud y su aceptación, con ayuda de obras de referencia tales como:
 - diccionarios y enciclopedias, de autoridad reconocida en la materia en cuestión;
 - tesauros
 - clasificaciones temáticas.

Se puede también consultar a especialistas en la materia, prefiriéndose aquellos que tienen conocimientos de indización y documentación.

6.2 El indizador debe estar familiarizado con estas obras y con las limitaciones que presentan, por ejemplo una lista de encabezamientos de materia o un esquema de clasificación puede no permitir la representación exacta de un concepto encontrado en un documento. Si los conceptos están representados por códigos de clasificación, necesita saber que estos códigos designan generalmente un contexto más amplio o más restringido, que puede no ser completamente apropiado para el documento estudiado.

6.3 Si un lenguaje de indización incorpora un tesoro, el número de términos asignados al documento y la multiplicidad de entradas, pueden reducirse sin pérdidas, ya que los términos generales y otras relaciones pueden establecerse con el propio tesoro.

Cuando se utiliza un tesoro debe seleccionarse el término más específico existente para representar un concepto dado.

6.4 Algunos sistemas de indización utilizan indicadores de función de enlace, de ponderación, etc. El indizador debe estar familiarizado con todas las reglas asociadas con el uso de estos mecanismos.

6.5 En la práctica el indizador encontrará con frecuencia conceptos que no existen en ningún tesoro o esquema de clasificación. Según el sistema utilizado dichos conceptos deberán tratarse de diferentes formas, por ejemplo:

- a) expresarlos por términos o descriptores y añadirlos inmediatamente al lenguaje de indización.
- b) representarlos temporalmente por términos más generales, y proponerlos como candidatos para una adición posterior.

7 Control de calidad de la indización

7.1 La calidad y la coherencia de la indización dependen de factores tales como:

- a) la competencia del indizador;
- b) la calidad de los instrumentos de indización.

En una situación ideal, los términos de indización asignados a un documento y el nivel de exhaustividad conseguido son idénticos con cualquier indizador. Para un mismo sistema de indización estos factores deben mantenerse relativamente estables en el tiempo. La coherencia es un factor importante en el comportamiento de un sistema de indización, en especial cuando la información se va a intercambiar entre diferentes centros de una red documental.

7.2 La imparcialidad total del indizador es un factor necesario para conseguir la consistencia de la indización. Un juicio subjetivo en la identificación de los conceptos y en la elección de los términos de indización, afectarán inevitablemente al comportamiento del sistema de indización. La consistencia es más difícil de conseguir con un equipo de indización formado por muchos miembros o cuando la indización se lleva a cabo por equipos de indizadores que trabajan en lugares diferentes, por ejemplo en un sistema descentralizado. En estas situaciones, se recomienda una etapa de comprobación centralizada con devolución a los indizadores.

7.3 El indizador debe tener un buen conocimiento del campo de que tratan los documentos a indizar. Debe comprender los términos que se encuentran en los documentos y las reglas y procedimientos del lenguaje de indización específico. Los centros que manejan documentos en lenguas extranjeras deben disponer de especialistas en esas lenguas.

7.4 La calidad de la indización se podrá conseguir de manera más efectiva si los indizadores tienen contacto directo con los usuarios. Estos podrían, por ejemplo, determinar si ciertos términos o descriptores son susceptibles de producir combinaciones falsas, dando lugar a salidas no pertinentes.

7.5 La calidad de la indización depende también de la posibilidad de poner al día el lenguaje de indización. Es esencial que el sistema permita la introducción de nuevos términos en el lenguaje o cambios en la terminología que respondan a nuevas necesidades de los usuarios.

7.6 Cuando sea posible, debe comprobarse la calidad de la indización, analizando los resultados de la recuperación de documentos, por ejemplo, calculando los porcentajes de exhaustividad y de precisión.

Anexo A
(informativo)

Organigrama del proceso de indización utilizando un tesaurus

