

NTP 106: Documentación: indización

Definition of the term "indexing", as well as of its modalities and results
 Definition de l'indexation et de ses modalités ainsi que de ses résultats

Las NTP son guías de buenas prácticas. Sus indicaciones no son obligatorias salvo que estén recogidas en una disposición normativa vigente. A efectos de valorar la pertinencia de las recomendaciones contenidas en una NTP concreta es conveniente tener en cuenta su fecha de edición.

Redactor:

CENTRO NACIONAL DE FORMACIÓN Y DOCUMENTACIÓN - BARCELONA

Objetivo

Definición de indización, de sus distintas modalidades y de los resultados de la operación de indización y la expresión de la misma según el sistema empleado.

Definiciones

Cualquier tipo de documento presenta unas características formales y otras de contenido, la expresión de las cuales permite la identificación completa del documento.

En esta nota se describen las características de contenido; las formales serán objeto de otra nota.

Se definen los siguientes términos:

- Análisis documental.
- Indización.
- Término de indización
- Lenguaje de indización.
- Descripción bibliográfica.
- Catalogación.
- Clasificación documental: clasificación decimal Dewey; clasificación decimal universal.
- Palabra clave.
- Descriptor.
- Vocabulario controlado.
- Sinónimo.
- Homógrafo.
- Término polisémico.
- Tesoro: término amplio, término específico, término relacionado, nota explicativa.

Análisis documental

Examen de un documento por el que se extraen aquellos conceptos significativos que expresan su contenido.

Indización

Representación del contenido de los documentos mediante símbolos (frases, palabras, números, caracteres alfanuméricos, letras, etc.)

Término de indización

Aquel símbolo (palabra, frase, número, etc.) que expresa el resultado de la indización.

Lenguaje de indización

El conjunto de indicadores léxicos y sintácticos utilizados para expresar el contenido de los documentos científicos y técnicos, con el fin de almacenar y recuperar la información contenida en ellos.

Los indicadores léxicos son equivalentes a los términos de indización, mientras los indicadores sintácticos son símbolos o relaciones propias de cada lenguaje y que en muchos casos quedan reducidos a la coma y a la conjunción "y" que indican simplemente yuxtaposición de los términos de indización.

Ejemplos:

En la CDU, se indica la relación entre dos términos representados por cifras mediante el signo (:), así 534.84 representa acústica y 725.822 teatros de ópera, mientras 534.84 _: 725.822 indicará acústica de los teatros de ópera.

614.8 indica seguridad y 621.3 electricidad, de manera que 614.8 _: 621.3 representará seguridad eléctrica.

En determinados vocabularios empleados en indización de documentos de la industria química un mismo término puede representar una materia prima o un producto final y si previamente consideramos que (1) es materia prima y (2) producto final, los términos benceno (1) y benceno (2) indicarán respectivamente benceno como materia prima de un proceso y benceno como producto final de otro.

Descripción bibliográfica

Conjunto de datos elementales que se utilizan para la referencia de los documentos particulares en un archivo.

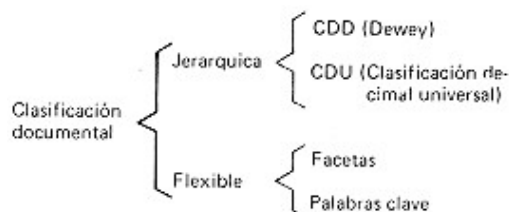
Catalogación

Registro de descripciones bibliográficas u otras formas de identificación de los documentos, bajo una forma normalizada.

Clasificación documental

Ordenación de documentos mediante un criterio dado en divisiones y subdivisiones (clasificación jerárquica) o según otros criterios de mayor flexibilidad (Clasificación no jerárquica) del contenido de los documentos.

Atendiendo a su mayor o menor jerarquización, se podrían dividir los distintos tipos de clasificaciones documentales en:



Clasificación decimal Dewey

Ordenación de los documentos atendiendo a su contenido en divisiones decimales. Fue ideada en 1876 por el bibliotecario norteamericano Melvey Dewey para colocar, los libros en estanterías y aún hoy se usa en muchos países americanos.

Clasificación decimal universal

Adaptación de la clasificación decimal de Dewey, con la introducción de indicadores sintácticos que le proporciona cierta flexibilidad.

Fue introducida por P. Otlet y H. Lafontaine en 1892, en Bruselas.

Esta clasificación es la más extendida y por abarcar todo el conocimiento humano se califica de universal.

Sus grandes apartados clasificatorios son los siguientes:

0. Generalidades.
1. Filosofía. Psicología.
2. Religión. Teología.
3. Ciencias Sociales. Derecho. Educación.
4. Filosofía y lingüística.
5. Ciencias puras.
6. Ciencias aplicadas. Medicina. Tecnología.
7. Artes, juegos y deportes.
8. Literatura.
9. Geografía. Biografía. Historia.

Algunos apartados que corresponden a la prevención son los siguientes:

Estadísticas biomédicas 61.311

Epidemiología 616-036.22
Fisiología 612
Ergonomía 658.3.053
Psicología 616.89
Trauma sonoro 616.28
Toxicología 615.9
Higiene industrial 613.6
Seguridad 614.8
Incendios 614.84
Explosiones 614.83
Electricidad 614.8 -: 621.3
Equipos y prot. person. 614.89

Clasificación por facetas

Ordenación de los documentos atendiendo a unas características o puntos de vista (facetas) que pueden ser asignadas libremente por el indizador, aplicado a un campo concreto de la ciencia o de la técnica.

Teóricamente fue introducida por el profesor indio Ranganathan y aplicada al campo de la ciencia y de la técnica por el inglés Vickery.

Palabra clave.

Término de indización formado por palabras o frases cortas, que representa, según el indizador, el contenido de los documentos. Se asigna libremente y para reflejar el contenido de ciertos documentos se necesitan en general, varias.

Acostumbra a ser un número adecuado el de 3 a 6. Su primer antecedente en documentación se halla en los "Uniterm" de Mortimer Taube, introducidos en 1953.

Descriptor

Aquella palabra clave que se usa preferentemente para designar un concepto dado.

También se ha definido como aquella palabra clave contenida en el vocabulario controlado o en el tesauro empleado en la indización de una colección de documentos pertenecientes a un campo específico de la Ciencia.

Vocabulario controlado

Relación de descriptores usados en la indización de una colección de documentos de un campo temático concreto y usado por una institución o equipo de investigación o técnico.

Sinónimo

Cada una de las palabras que designa el mismo concepto o idea.

Ejemplos: Asbesto y amianto. Benceno y benzol

Homógrafo

Aquella palabra que puede representar conceptos distintos y, en general, con base semántica diferente.

Ejemplos: Cálculo (renal) y Cálculo (aritmético)

Término polisémico

Aquella palabra que puede representar conceptos distintos, pero, en general, relacionados semánticamente.

Ejemplos: Puente (ingeniería) y Puente (odontología)

La diferencia entre términos homógrafos y polisémicos es muchas veces difícilmente distinguible.

Tesauro

Herramienta usada en la indización de documentos para el almacenamiento y recuperación de la información contenida en los mismos, perteneciente a un campo específico de la ciencia o de la técnica, que consta de una relación de términos de indización, con indicación expresa de sus interrelaciones semánticas y sintácticas (paradigmáticas y sintagmáticas).

Todo tesauro consta de las siguientes partes: introducción, clasificación general, diccionario de términos con indicación de sus interrelaciones y un diccionario de los términos contenidos en el tesauro.

Las principales interrelaciones que se expresan entre los términos de un tesoro son las siguientes:

Sinónimo: Expresión de las distintas acepciones de un término de indización. El término aceptado o descriptor se relaciona con los demás sinónimos mediante el indicador UP (útese por), traducción del inglés UF (use for).

Término amplio: Aquel término que engloba jerárquicamente o bien por relación todo-parte a otros varios. Se indica por TA (término amplio) traducción del inglés BT (broader term).

Término específico: Aquel término que se halla comprendido dentro de otro de mayor amplitud o alcance (TA). Se representa por TE (término específico), traducción del inglés NT (narrower term).

Término relacionado: Aquel que no mantiene relaciones jerárquicas ni de parte-todo con otros términos, si no de otro tipo como puede ser de uso o aplicación, similitud de significado (no sinónimo), etc. Se indica mediante TR (término relacionado), traducción del inglés RT (related term).

Nota explicativa: Usada en los tesauros para definir y acotar un término dentro del significado específico que se le atribuye en un tesoro dado.

Bibliografía

(1) AMAT NOGUERA, N

Técnicas documentales y fuentes de información

Barcelona, Bibliograph. S.A., 1978

(2) CHAUMIER, J.

Lo traitement linguistique de l'information documentaire: L'analyse documentaire

Paris, Entreprise Moderne d'Édition, 1972

(3) INSHT

Curso general de técnicas documentarias

Barcelona, CIAT, 1982

(4) SOERGEL, D.

Indexing Languages & Thesauri: Constructions and Maintenance

Los Angeles, Melville Publishing Co., 1974

(5) TURUGUET, D.

Temario del seminario sobre tesauros. (ITB/263581)

Barcelona, INSHT, 1981

(6) UNESCO

UNISIT. Etude sur la réalisation d'un système mondial d'information scientifique

Paris, UNESCO, 1971

(7) VICKERY, B.C.

Techniques of information retrieval

London, Butherworths, 1971