

TESAUROS : CONCEPTO, ELABORACION Y MANTENIMIENTO

CONCEPTO

Etimológicamente viene del latín y su significación era tesoro. Definición que tiene poco que ver con el concepto que tiene hoy para un documentalista, así la **norma ISO 2788 / TC 46** "Principes directeurs pour létablissement et le developpement de thesaurus monolingues", nos dice que estos se pueden definir según su función y según su estructura :

- Por su función. como un instrumento de control terminológico utilizado para trasponer a un lenguaje más estricto el idioma natural empleado en los documentos y por los indizadores

-Por su estructura. Es un vocabulario controlado y dinámico de términos que tienen entre ellos relaciones semánticas y genéricas y que se aplica a un dominio particular del conocimiento.

P. **Leverly**, es el puente entre el lenguaje del informado (documentalista) y el lenguaje del no informado (usuario).

ELEMENTOS DE UN TESAURO

- Unidades lexicales.

Estas pueden comportar las siguientes categorías: grupos de descriptores, que agrupan los términos de indización bien por campos (temas) o por clases de términos o "facetas"

- los descriptores.

Son palabras o expresiones del lenguaje corriente retenidas por el constructor del tesoro para designar los conceptos representativos del documento y las preguntas utilizadas por el

indizador.

- No descriptores.

Son sinónimos o cuasi-sinónimos de los descriptores o términos que designan en el lenguaje de uso conceptos afines a los que cubren los descriptores. Los no descriptores no pueden ser utilizados para la indización del documento o de las preguntas, pero cada uno de ellos reenvía a uno o dos descriptores para representar los conceptos correspondientes

Comerciantes
U.p. Mercaderes

Mercaderes
Use Comerciantes

TIPOLOGIA

Pueden concebirse desde dos enfoques:

- Por facetas, que según la norma ISO 2788 sobre la construcción de tesauros monolingües, es una noción más abstracta que el concepto de campo, lo cual requiere una disciplina mental más rigurosa para el compilador. Los documentos se agrupan por clases o por puntos de vista, según un tipo básico de conceptos.

- Por campos, que pone en relación los conceptos con los cuales generalmente se asocia un concepto.

DESCRIPTORES

¿Que es un descriptor? Es el término o símbolo autorizado y formalizado que figura en un tesoro, que se utiliza para representar sin ambigüedad los conceptos contenidos en los documentos y en las peticiones de recuperación de la información.

Los descriptores pueden ser simples o compuestos, y la AFNOR los define como palabra o grupo de palabras retenidas en tesoro y elegidas entre un grupo de términos equivalentes para representar sin ambigüedad una noción contenida en un documento o en una condición de búsqueda documental.

Modos de ambigüedad.

Debido a la univocidad de los descriptores, este designa un solo concepto y un concepto es designado por un solo descriptor. Esto es asegurado por diversas modalidades que son las formas de evitar la ambigüedad.

La sinonimia es eliminada gracias a las relaciones de equivalencia. La polisemia se

elimina gracias a ciertas relaciones semánticas, tanto por la pertenencia a un grupo semántico como por las relaciones jerárquicas.

Por una parte nos inclinaremos más hacia a la elección de una expresión más que de un término aislado. Así será preferible escoger "Menor de edad" que simplemente "Menor", que puede inducir a equívocos.

Otra de las formas de evitar esa ambigüedad del lenguaje natural, es la adición de una expresión al descriptor cuando su sentido no es suficiente, es un cualificador que aparece entre paréntesis "Menor (álgebra)", esta técnica se utiliza cuando no es posible retener una expresión significativa. Por medio del cualificador a la hora de hacerla búsqueda distinguiremos entre dos conceptos de la misma grafía y diferente significado a fin de que cuando queramos documentos sobre "Mercurio (mineral)", no nos aparezcan los indizados por "Mercurio (planeta)", por lo tanto evitaremos el ruido documental. Se concreta el descriptor homónimo mediante un símbolo que lo califica.

Las notas de alcance, se usan para limitar el uso de los descriptores, para escribir un posible sentido del término cuando este tiene diferentes significados, pues es usado en diferentes disciplinas, no forma parte del descriptor, aunque aparezca en el tesoro. "Seno: Usado solo para Matemáticas". Las notas de alcance pueden ser explicativas, de definición, de aplicación o históricas.

"Política Agrícola Común"

N.A. Política agrícola de los países miembros de la CEE. (explicativa)

Las notas de alcance explicativas están destinadas a precisar el ámbito de aplicación de un descriptor

"Construcción"

N.A. Utilizar sólo para la acción de construir, para designar lo que se construye utilizar el descriptor "Edificio"

Las notas históricas están destinadas a explicar desde cuando se usa el descriptor

"Telemática"

N.A. para antes de 1982 utilizar: "Telecomunicación" e "Informática"

La notación para las notas de alcance según la norma ISO 2788 será:

NE (Note explicative)

SN (Scope note)
NA (Nota de alcance)

Indican el sentido limitado del término, solamente con fines de indización.

NO DESCRIPTORES

Palabras incluidas en el tesoro, que pertenecen a una lista de sinónimos o cuasi-sinónimos y términos emparentados unidos a los descriptores por una relación de equivalencia semántica, que son susceptibles de aparecer en los documentos o en las preguntas, pero que no son utilizados para formular la interrogación al sistema.

La inclusión de estos en un tesoro tiene como finalidad mejorar la coherencia de la representación de los documentos o de una pregunta al enviarnos al término de indización.

"Coche" (no admitido)
USE "Automóvil"(admitido)

Por medio de ellos el tesoro descarta todo aquello que podemos denominar accidentes del lenguaje natural, con el fin de no retener más que un término referencial.

La notación para señalar los términos admitidos y no admitidos es:

EM (Employer)
EP (Employer pour)
USE (Use)
UF (Use for)
USE
UP (Usado por)

RELACIONES ENTRE LOS TERMINOS DE UN TESAURO

- Relaciones de equivalencia
- Relaciones jerárquicas
- Relaciones asociativas

Relaciones de equivalencia

Es la relación entre descriptores o términos preferidos y los no descriptores o términos no usados en la indización referentes a un mismo concepto

USE
UP

Esta relación cubre dos tipos de términos los sinónimos y los cuasi-sinónimos. Los

sinónimos son términos cuyo significado puede considerarse igual, por lo que son intercambiables, y tienen diferente origen lingüístico

- Términos populares y científicos
- Variantes para conceptos modernos
- Términos de uso actual y anticuados
"Países en vías de desarrollo"
UP "Países subdesarrollados"
- Formas variantes de un término
"Sicología/Psicología"
- Términos originales de distintas culturas
- Abreviaturas y nombres desarrollados
- Forma dividida y no dividida
"Carbón" + "Minería"/ "Minería de carbón"

Se suelen considerar términos preferidos aquellos que más corrientemente son utilizados por los usuarios

Los Cuasi-sinónimos son términos cuyo significado es considerado diferente en el uso común, pero son tratados como sinónimos en la indización. por ejemplo puntos de un continuo "Humedad y Sequedad", o simplemente el término más genérico

"Ascensión vertical"
Up. Ascensor
Up. Montacargas

RELACIONES JERARQUICAS

Es la relación vertical entre todos los descriptores de una misma clase, expresada en términos de subordinación de los conceptos en un doble sentido:

- Genero/especie
- Todo/parte

Basado en grados de superioridad y de subordinación; donde el término superior representa una clase o un todo y los subordinados se refieren a las partes o especies de estos.

La reciprocidad se expresa mediante la notación:

TG (Término genérico)
TE (Término específico)

BT (Broad term)
NT (Narrow Term)

Líneas aéreas
TG Servicios de transporte

Servicios de Transporte
TE Líneas aéreas

Por lo tanto podemos hablar de una relación asimétrica, en la que si A es inferior a B, necesariamente B es superior a A. Cuyo descriptor cabeza de jerarquía será el "Top term". Si un descriptor tiene un solo descriptor por encima de él, este es monojerárquico, si tiene dos o varios es polijerárquico. Como regla general puede apuntarse que cada término subordinado debe de referirse al mismo tipo de concepto que el término superior, que puede ser un objeto del mismo, una acción, una propiedad

Metales (clase de material)
TE Fundición (acción)

Relación genero/especie

Se aplica a acciones, propiedades y agentes

"Vertebrados"
TE Anfibios
TE Aves
TE Peces
TE Reptiles, etc.

Relación jerárquica todo/parte

En la asociación de ideas el nombre de una parte implica y evoca la del todo poseedor. De modo que en esta relación la parte del todo funciona como término superior, mientras que la parte lo hace como término subordinado.

Se aplica a objetos, a lugares geográficos, estructuras sociales, disciplinas y sus partes de estudio

Argentina
TGP Buenos Aires
Entrerios

Ejercito
TGP Cuerpos
TGP Divisiones
TGP Batallones
TGP Regimientos

Notación : TGP (Término genérico partitivo)
TEP (Término específico partitivo)

Relaciones de ejemplo

Es la conexión entre una categoría general de objetos expresada por un sustantivo común y un ejemplo individual, que suele ser un nombre propio que sirve para representar a otros nombres propios, que no se incluyen en el tesoro para no sobrecargar las categorías.

Regiones montañosas
TE Alpes
Himalaya, etc.

Relaciones polijerárquicas

Un concepto sobre bases lógicas puede pertenecer a más de una categoría, por lo tanto posee relaciones polijerárquicas

Instrumentos de viento
TE Órganos

Instrumentos de teclado
TE Órganos

Como conclusión podemos decir que las relaciones jerárquicas se utilizan durante la búsqueda para enriquecer la formulación de la pregunta, añadiendo a la búsqueda uno o más descriptores superiores o inferiores para concretar la búsqueda.

RELACIONES ASOCIATIVAS

Indican relación o uniones en la significación de los descriptores. Son relaciones simétricas entre dos descriptores, que son susceptibles de evocarse mutuamente por asociación de ideas. Y simétrica pues si A se asocia a B, B se relaciona con A.

La notación utilizada es:

TR (Término relacionado)

RT (Related term)

VB (alemán)

Pueden reunir términos de la misma categoría o de categorías diferentes, los referentes a las mismas son términos con significados superpuestos, tales como "barcos" y "buques", donde cada uno de los términos puede ser objeto de una definición exacta. Sin embargo a veces se les utiliza separadamente y de forma indistinta.

También pueden asociarse términos que tengan una relación de familia o derivativa

Felinos
TE Tigres
TR León

Para términos pertenecientes a diferentes categorías se puede dar los casos:

- Una disciplina y el campo de estudio

Selvicultura
TR Bosques
Bosques
TR selvicultura

Maquetas de barcos
TR Barcos

Ayuda a la búsqueda por asociación de ideas y distintos puntos de vista.

ELABORACION Y MANTENIMIENTO

La elaboración de un tesoro es una operación intelectual, en la cual la automatización sólo puede aportar una ayuda para las tareas de carácter más rutinario y repetitivo. Su elaboración supone el trabajo continuo de un grupo de especialistas en la materia durante un periodo de tiempo bastante amplio, que dependerá del volumen del tesoro, de los campos a cubrir y de la profundidad de la indización.

Para la compilación existen dos métodos:

- Inductivo. Parte de cierto número de términos de los dominios considerados a través de enciclopedias, diccionarios técnicos, índices bibliográficos, otros tesoros

- Deductivo. Cuyo vocabulario se extrae de los documentos, por indización libre de cierto número de documentos, mediante el recuento de la frecuencia de aparición.

Ambos métodos se utilizan conjuntamente, existiendo siempre una normalización, aunque sea mínima con preferencia por la forma singular, masculino, forma sustantiva, escritura directa.

Obtenido el vocabulario se procederá a su reducción por la eliminación de los términos no significativos, y el establecimiento de las relaciones entre los descriptores y entre estos y los no descriptores.

Una vez concluido el vocabulario se procederá a la evaluación del mismo. Con lo cual se procederá a su edición. Pero el tesoro no es algo definitivo y absoluto, este se debe ajustar a

los progresos y a los propios usuarios y cambios en el sistema, por lo que cuando sea necesario debemos de recurrir a su mantenimiento, por medio de estadísticas de utilización de los descriptores.

PRESENTACION DEL TESAURO

Presentación alfabética

Los descriptores y no descriptores se encuentran agrupados en una sola secuencia alfabética, acompañados de sus relaciones, no es la forma más habitual de encontrarlos.

Presentación sistemática

Consta de dos partes, una de ellas son las categorías o jerarquías y por otra un índice alfabético, que dirige a los usuarios a la sección semántica a la que el concepto pertenece. Considerándose esta última sección la parte principal del tesauro y el índice como auxiliar.

Presentación gráfica

Dispuestos como una figura que permite al usuario asociar los términos que se encuentran relacionados. Con dos tipos de presentación:

- Árboles
- Flechados

También debe de constar de un índice alfabético, pues es difícil representar los no descriptores, notas de alcance. Dándose el caso de que la sección auxiliar del índice contiene más información que la parte en teoría principal, siendo esta gráfica de apoyo.