

XXIII JORNADAS MEXICANAS DE BIBLIOTECONOMIA
17,18 Y 19 de septiembre de 1992

→ PROPUESTA DE LINEAMIENTOS PARA LA ELABORACION
DE TESAuros EN COMPUTACION



BIBLIOTECA
CENTRO UNIVERSITARIO
DE INVESTIGACIONES
BIBLIOTECOLOGICAS

RESUMEN.

En este trabajo se proponen lineamientos para la elaboración, presentación y difusión de tesauros en el área de la computación. Se expone la finalidad que se persigue al elaborar un tesauro, los problemas que se presentan en su elaboración, la forma en que se debe presentar, algunos recursos para elegir los términos que se pretenda incluir, la estructura del tesauro, la simbología empleada (utilizamos la versión española) para establecer las relaciones básicas del tesauro y la necesidad de actualización del mismo. Finalmente, un comentario a manera de conclusión.

PROPUESTA DE LINEAMIENTOS PARA LA ELABORACION
DE TESAURUS EN COMPUTACION

Revisión

INTRODUCCION.

Con este proyecto estamos proponiendo lineamientos para la elaboración de tesauros monolingües en el área de Computación. (para la elaboración de tesauros multilingües deberán consultarse las Normas ISO 1.3 y la 0000 que propone la UNESCO). Al hacer esta propuesta, no quiere decir que deben seguirse al pie de la letra, sino adecuarse a las necesidades de cada proyecto, tomando en cuenta el volumen de documentos que se tienen, el número de términos que se pretenden incluir y el grupo de personas que reúnan los términos y establezcan sus relaciones (ya sea un indizador o un especialista en el área).

Finalidad del tesoro.

Un tesoro debe servir como guía para la clasificación e indización de los documentos de una disciplina determinada. El análisis conceptual del documento permitirá la extracción de descriptores y debe realizarse por un grupo de expertos del área, para la recuperación de información específica.

Para que la finalidad del tesoro se cumpla, deberá constituirse fundamentalmente de tres partes: enumerativa, clasificatoria y relacional del conocimiento. La delimitación temática del tesoro es a veces complicada y el indizador tiene que precisar el tema y las subdivisiones que no pueden desligarse sistemáticamente.

Aunque es mucho más complicada su estructuración, cuando el área temática cubierta lo permite, la relación establecida puede enmarcarse más flexiblemente con términos relacionados de acuerdo a las necesidades y cantidad de documentos disponibles para el indizador.

Deberán considerarse los siguientes elementos:

- el usuario a quien está destinado el tesoro
- los documentos recopilados
- la terminología incluir
- el carácter general o específico del tesoro y
- el objetivo que se pretenda alcanzar.

Se puede decir que los factores que determinan la necesidad de un control de terminología son:

- el factor ideológico, que se refiere a la conceptualización que representa un término en determinado contexto social, económico y político, ya sea por su creación, introducción, o modificación en un área determinada del conocimiento.

- el factor temporal, que hace a un término representar o no el concepto para el que fué creado, o por el cual ha sido utilizado desvirtuándose a través del tiempo por la introducción de términos nuevos y sus posibles combinaciones.

- el factor de aplicación, que se hace patente en el momento de su utilización y del papel que juegue en el desarrollo de la disciplina a que pertenezca.

Problemas que se presentan en la elaboración del tesoro.

En el caso de la computación, está latente la descalificación del idioma, que se da por la dificultad de representar la terminología muerta con conceptos precisos de la ciencia. No sólo esto perjudica al idioma, sino también la introducción de adaptaciones morfológicas, el uso de sufijos, prefijos, etc. incomprensibles para los iniciados o no conocedores de la materia. Lo que no sucede a colegas de experiencia. Incluso en el manejo de neologismos, y tecnicismos tan específicos para cada área determinada.

Otro problema es la traducción literal de la terminología que dificulta y altera la estructura regular de un idioma. El significado real o tradicional de un vocablo o palabra, crea confusión entre los especialistas y los iniciados.

Uno más es la falta de especialistas que se interesen en el control de la terminología que ellos mismos manejan y desarrollan, así como la falta de conocimiento de un lenguaje propio que permita la creación o modificación morfológico-linguística del idioma. Todo ello es provocado por la falta de interés, recursos, tiempo y, sobre todo, del uso de términos desconocidos.

El conjunto de problemas arriba expuestos entorpece el diseño de los medios que nos permiten el acceso real a la documentación : un obstáculo a los lenguajes técnicos, científicos, y especializados de cualquier área.

La solución del problema requiere de un amplio periodo de tiempo, la obtención de material de referencia y un trabajo conjunto de especialistas del área con un indizador o documentalista profesional y comprometido.

Dada la dificultad que presenta el recopilar términos confiables de computación, el compilador se enfrenta a la tarea de:

- revisar la literatura existente
- determinar la temática de los documentos
- dar de alta los vocablos empleados comúnmente en el área y
- seleccionar los términos que funcionarán como descriptores

Este último lo recomienda la Norma ISO (0) en su versión española.

Cuando el indizador se dispone a revisar la literatura existente, debe tener en cuenta el volumen o cantidad de documentos que pueda recopilar y analizar, así como las características del mismo, es decir, si se trata de literatura en español o en otro idioma, publicaciones primarias y secundarias, manuales, libros de texto, determinar las fechas de publicación de los documentos, etc.

El indizador debe reconocer los términos de uso corriente en el área, así como considerar neologismos, extranjerismos, sinónimos, siglas, acrónimos, y otras equivalencias de los términos, que ocasionalmente llegan a utilizarse y provocan confusiones.

Para determinar la temática de los documentos proponemos revisar los índices analíticos de las obras que se hayan compilado en la materia, debiendo el compilador tener la habilidad para distinguir los documentos de autoridad que otorguen los términos necesarios y además validen el uso de términos en el área. Pero se debe de tener cuidado ya que estos documentos no son el reflejo fiel de la temática discutida en el documento.

Para dar de alta un término, deberá apoyarse en la revisión de fuentes de consulta especializadas que determinarán la garantía literaria de los términos. La garantía de uso será evaluada con la consulta a especialistas que juzguen si el término es aceptado o no.

La selección de los descriptores, que son los términos de entrada para la búsqueda de información en un vocabulario controlado (tesauro), debe realizarse en base a un estudio lingüístico amplio y cuidadoso.

El constante cambio en la terminología utilizada en el área, el uso de términos de reciente inclusión -que no están conceptualizados de acuerdo al contexto real del lugar donde se emplean-, el manejo arbitrario de los términos y la falta de trabajos de especialistas que validen y acepten conceptos del área, es decir, trabajos que estén realmente fundamentados en la práctica y sean útiles, hacen que la tarea de elaborar tesauros de computación sea ruda o casi nula.

De ahí que se proponga el establecimiento de relaciones a priori o tesaurizadas como el mecanismo más adecuado para realizar una lista de términos que permitan la elaboración de vocabularios controlados para su mejor manejo en el área.

El método de indización que proponemos es el post coordinado, de acuerdo a:

- a) la presentación y organización de los términos
- b) la elección de términos preferentes
- c) la aplicación de acuerdo a la lógica de la elección de los términos preferentes.

Para una mejor interpretación de la presente propuesta, definimos los conceptos a que hacemos referencia:

Descriptor: Palabra o término seleccionado y normalizado, de un conjunto de términos que son equivalentes en su significado, a fin de evitar ambigüedades o/y equívocos.

Documento: Toda base de conocimiento expresado en un soporte material y susceptible de ser utilizada para consultas, estudios o pruebas y que puede ser catalogado, clasificado, o bien, indizado.

Índices analíticos: Listado de términos o descriptores incluidos en el tesoro, ordenados alfabéticamente.

Indización: Lista de términos pertenecientes a un vocabulario controlado, con un fin determinado: la recuperación oportuna y relevante de información.

Publicaciones primarias: Son las que contienen información original, no abreviada: libros, folletos, monografías, artículos de revistas, manuscritos. Se les llama también fuentes o publicaciones de primera mano.

Publicaciones secundarias: Son los documentos derivados de fuentes primarias (sumarios, bibliografías, etc.), es decir, los documentos reelaborados para sintetizar y/o hacer referencia.

Tesoro: Lista de descriptores, indicadores e identificadores temáticos de un documento, con sus relaciones semánticas y genéricas claramente especificadas, para utilizarlos

en la tarea de indización, análisis, registro y clasificación del conocimiento.

Presentación del tesoro

La última parte a considerar en la elaboración de un tesoro es la de su presentación, que debe corresponder a su contenido.

Se deben incluir los siguientes apartados:

- Portada : en que debe figurar el lugar en donde se elaboró el tesoro, y el año de su publicación, la edición si se trata de una modificación o actualización, el título propiamente dicho y los nombres de los responsables de su elaboración.

- Sumario o tabla de contenido: en que se den a conocer los apartados a tratar en el tesoro y sus diferentes puntos.

- Introducción : en que se definan los objetivos que pretende alcanzar el realizador del tesoro, la exposición delimitativa del tema a tratar, el tipo de simbología empleada, el ordenamiento alfabético, la información sobre actualización, los criterios generales de inclusión de terminología, etc.

Es importante delimitar el periodo de tiempo en que se realizó la compilación de documentos indizados, así como las características de los documentos contenidos, ya que un tesoro con terminología vigente no serviría o no apoyaría una búsqueda retrospectiva por el desfase del nivel de conocimiento en que se encontrara la disciplina. En el caso de la computación, el desfase sería muy amplio y complejo.

- Sección clasificatoria : que permite la organización y sistematización de grandes o pequeñas cantidades de información contenida en documentos, y nos muestra a la vez el grado específico de desarrollo del área. Los cambios y especificidad de cada disciplina, en cuanto a conceptualización y teorización. El sistema de clasificación puede ser adoptado, pero debe tenerse cuidado ya que un tesoro siendo más apropiado para ramas especializadas del conocimiento va desarrollándose y ampliándose, por lo que debe ser a la vez rígido y flexible, para permitir una ampliación clasificatoria dentro de una estructura dada.

Si creamos o adaptamos un sistema de clasificación debemos cuidar nuestros criterios de inclusión, "pues la interpretación corriente de una característica esencial de una cosa como aquella sin la cual la cosa no sería lo que es, obviamente calificaría toda característica de una cosa como esencial y así convertiría trivial el concepto" (1). De ahí que un conjunto de objetos o términos elegidos para conformar un grupo o subdivisión temática de una disciplina deberá ser previamente teorizada para especificar las clases a incluir y la manera en que se desarrollarán.

Las áreas o divisiones deben ser desarrolladas, de preferencia, por especialistas del área de estudio y pueden corresponder a especialidades, tipos de actividades, líneas de investigación, relaciones con otras disciplinas, etc., de acuerdo a su propio desarrollo. Así, la elección de términos preferentes deberá caer en la estructura divisoria de la discusión temática particular.

- Sección sistemática : en que se incluye el esquema utilizado de clasificación, señalando los términos preferentes y sus subordinaciones (respetando una sangría para destacar qué está subordinado a qué). Aquí se encontrará que la terminología, en su conjunto, es el reflejo de la disciplina, puesto que pueden encontrarse los conceptos, palabras, vocablos, neologismos, extranjerismos, adaptaciones morfológicas, palabras truncadas o cortadas, que expresen lo obsoleto o vigente de la disciplina que representan.

- Sección alfabética : que consiste en el ordenamiento, palabra por palabra, de todos los términos incluidos en el tesoro, sean descriptores o términos no preferentes. Es la herramienta principal de apoyo al usuario, pues en él encuentra las diversas representaciones del lenguaje y sus relaciones. Esta parte del tesoro puede ser elaborada con apoyo en sistemas automatizados. Es aquí donde se hace referencia a la simbología utilizada que acompañará a los términos de indización y a los no preferentes. Esta información auxiliar debe presentarse en un orden lógico tal, que presente la generalidad y especificidad del conocimiento, para que el usuario pueda elegir de una manera más específica el término que se adecúe a sus necesidades de información.

La elección de los términos.

Un recurso para normalizar la elección de términos y establecer posteriormente las relaciones jerárquicas entre ellos, es el manejo de las variantes de los conceptos que representan.

Estos recursos son:

- entes concretos: se refiere a todos los términos que expresan características cuantificables, por ejemplo:

sistema periférico
salida
multiprocesadora
minicomputadora

1. HEMPEL, Carl G. "Clasificación". En : Fundamentos de la formación de conceptos en ciencia.. -- p. 81. -- México : [s.n.], 198-?

- entes abstractos: se refiere a las propiedades intangibles de lo que representa el vocablo, por ejemplo:

Programación
entrada

- entes individuales: expresados por nombres propios, por ejemplo:

Pascal
Babbage, Charles
Cobol

- siglas y acrónimos: se refieren al uso de términos que son utilizados generalmente en su forma simplificada, por ejemplo:

IBM
ACM
RAM
CAD

Deberá considerarse el término cuyo uso y aceptación sea generalizado, además deberán indicarse reciprocamente, las relaciones entre los términos aceptados y no aceptados, mediante envíos de "véase", por ejemplo:

IBM
véase
International Bussiness Machinery

ACM
véase
Association Computer Machinery

RAM
véase
Random access memory

CAD
véase
Computer aided desing

*es valido en
trabajar*

- extranjerismos: se refiere a términos tomados de otra lengua. Debe considerarse el término reconocido o de aceptación generalizada, ya sea el extranjerismo o el término traducido, por ejemplo:

chip
byte
diskette
software
hardware

Cabe destacar que la computación es una de las áreas que acepta más el uso de extranjerismos que el uso de las equivalencias o traducciones de los términos utilizados, porque es un área nueva que no ha logrado establecer un lenguaje definido por el constante cambio que sufre y la aplicación inmediata que se hace de ellos.

- Términos compuestos que no pueden dividirse: se refiere a los descriptores que no deben separarse, puesto que de su fraccionamiento pueden resultar ambigüedades o pérdidas de significado, como: los nombres propios, frases sustantivadas, frases preposicionales y frases adjetivadas, por ejemplo:

bases de datos
sistemas de cómputo
computadora personal
cuarta generación
lenguajes de segundo nivel
proceso de datos

La forma en que se introducirá el término en el tesoro deberá ser preferentemente aquella que se use en el lenguaje natural y sea la forma más reconocida por el usuario. Es recomendable no hacer inversiones ya que los envíos de términos no aceptados le causan confusión por no corresponder a su lógica de búsqueda.

Estructura del Tesoro.

Simbología usada para indicar relaciones.

Las relaciones entre los términos preferentes y no preferentes se establecen de acuerdo a la elección de la simbología que presentamos a continuación, estas dos variantes han sido propuestas por la International Standard Organization y su adaptación en España.

Tipo de relación	ISO	Versión española
Usado por	#	U
Término general	<	TG
Término específico	>	TE
Término relacionado	-	TR
Nota de alcance		NA

La elección de uno de los sistemas o de ambos, está a consideración del indizador o del equipo de trabajo que se encargue de establecer las relaciones entre los términos, así como el número de símbolos a utilizar y el que consideren de mayor valor mnemotécnico, relacionándolo con el área de estudio.

Otra característica mnemotécnica, en éste caso para reconocer los términos preferentes y no preferentes, es destacar tipográficamente el término preferente en mayúsculas y el no preferente en minúsculas, salvo en los casos que se crean confusos.

Un recurso para normalizar el uso adecuado de los términos que se indizarán es el uso de la:

Nota de alcance. Se utiliza cuando el término aceptado es muy amplio y reconocido en diversas áreas del conocimiento, para delimitarlo en un significado y restringir su uso a los términos del área.

La nota o aclaración que se añada al término preferente debe establecer el significado único, el significado elegido con el que se deberá utilizar la palabra, que sirve también para el establecimiento de las relaciones entre términos, o sea, de los términos del lenguaje natural del usuario al lenguaje controlado o utilizado por el indizador.

Las relaciones básicas del tesoro.

Se concretan por el uso convencional de los tres tipos siguientes:

- relación de equivalencia
- relación jerárquica
- relación de afinidad

La primera se refiere a términos de diferente origen lingüístico, sinónimos y cuasisinónimos, nombres populares y nombres científicos, nombres comunes y nombres comerciales, términos vigentes y obsoletos, abreviaturas y nombres completos, formas simples y formas compuestas, que servirán para señalar el término equivalente y aceptado en el tesoro. Ejemplos:

ordenadores vs. computadoras
BIT vs. binary digit
entrada/salida vs. input/output
PC vs. computadora personal

La segunda se refiere a las relaciones que se dan entre las categorías aceptadas y sus subcategorías o supracategorías, conectando así conceptos individuales del mismo tipo o clase de acuerdo a su subordinación género-especie o todo-parte. Ejemplo:

Lenguajes de programación
Lenguaje de nivel cero
Lenguaje de primer nivel
Lenguaje de segundo nivel
Lenguaje de tercer nivel
Basic
Pascal
Cobol
Fortran
Algol

La última se refiere a las relaciones de términos del mismo o diferente nivel jerárquico, por ejemplo de diferente origen lingüístico, de procedencia popular o científica, de escritura diferente, de moda, regionalismos, abreviaturas, neologismos, extranjerismos y conceptos confusos o de polisemia. Por ejemplo:

computadoras analógicas vs. computadoras digitales
bases de datos vs. banco de datos
hardware vs. software



Es importante mencionar que debe señalarse siempre la correspondencia entre cada uno de los términos, pues de no ser así habrá aislamiento de datos que pueden ser útiles al usuario para la recuperación de información relevante.

Actualización y vigencia del tesoro.

El tesoro debe actualizarse y enriquecerse constantemente con la introducción de términos nuevos, de delimitaciones precisas hechas por un equipo de especialistas para que las modificaciones sean válidas y útiles a la disciplina.

La actualización y vigencia debe planearse para su realización periódica, o bien, para cuando se considere necesaria. Cuando los términos encontrados ya no se apeguen a una estrategia de búsqueda real, o mejor, cuando los resultados no sean útiles al desarrollo de un proyecto de investigación.

La vigencia de los términos la dará el desarrollo de la propia disciplina, que detectará el indizador por medio de un análisis conceptual y/o contextual del término indizado o descriptor que se apegue a las necesidades del investigador.

Conclusión

Por el cambio constante de nuestra sociedad, de la cultura mundial, del desarrollo de la tecnología y la agilización mecánica de procesos caseros, tecnológicos e industriales; un tesoro es el reflejo fiel de una especialidad que permite el uso de la documentación y el manejo de los términos en cada especialidad, éste acorde a la época o no y a las necesidades de recuperación de información específica.

Se comprende que en la preparación de un tesoro se deben plantear de manera concreta y sin ambigüedades cada una de las diferentes etapas que lo forman, desde su planeamiento, su presentación y su publicación posterior.

La cantidad de información que se produce en cada área específica del conocimiento, la diversidad de ideologías que conceptualizan un término de diferente manera, la garantía literaria que valida la terminología, la falta de control en cuanto a la creación de vocabularios, la proliferación de tecnicismos y términos populares, así como de siglas, acrónimos, sistemas de clasificación y codificación de la información, manifiestan cada vez más la necesidad de un control estricto de los términos que se introduzcan en un tesoro.

Obras consultadas.

- BUONOCORE, Domingo. Diccionario de bibliotecología. -- 2a. ed. -- Buenos Aires : Marymar, 1976. 465 p.
- COMITE TECNICO CONSULTIVO DE CENTROS DE DOCUMENTACION. "Información sobre ISO : preparación de resúmenes". En : Boletín del sistema de información documental. -- (1981) junio, año 1, no. 4. -- 8 h.
- CHRISTIE, Linda Gail. Enciclopedia de términos de microcomputación. -- México : Prentice Hall Hispanoamericana, 1986.
- HEMPEL, Carl G. "Clasificación". En : Fundamentos de la formación de conceptos en ciencia.. -- p. 81-84 : [s.n.], 198-?
- HUNT, Roger. Computadoras y sentido común. -- Englewood Cliffs : Prentice, 1987. -- 159 p.
- IGUINIZ, Juan B. Léxico bibliográfico. -- México : UNAM, 1987. -- 129 p.
- ISO. "Directrices para el establecimiento y desarrollo de tesauros monolingües : parte I y parte II". En : Revista española de documentación científica. -- (1989-90) vol. 12-13, no. 1 y 4.
- LEATHERDALE, Donald y María José Galrao. "La construcción de vocabularios controlados". En : Revista AIBDA (1980) vol 1, p. 1-16.
- MONTERO HERNANDEZ, Esther. Manual de organización y registro de publicaciones periódicas. -- México : UNAM, 1987. -- 197 p.
- PRESSER, León. Ciencias de la computación. -- México : Limusa, 1990.
- SIPPL, Charles. Microcomputadoras : diccionario de términos. -- 2a. ed. -- Madrid : McGraw Hill, 1991. 750 p.

No relevantes