

Artículos

El tesoro en el ambiente digital

CATALINA NAUMIS PEÑA¹

Centro Universitario de Investigaciones Bibliotecológicas
de la UNAM, 04510, México D.F., Tel: 56-23-03-29
E-Mail: naumis@servidor.unam.mx

RESUMEN

Se explica cómo la estructura de un lenguaje documental, tipo tesoro complementa el ciclo de la organización de la información en el ambiente digital. Las relaciones tesaurales se comportan de la misma forma que la estructura de los datos en el entorno informático y por lo tanto su utilización facilita la clasificación temática para representar aquellos textos e imágenes que contienen conocimientos y saberes, explícitos e implícitos. Así, en un documento digital los descriptores tesaurales sustituyen al lenguaje natural y permiten recuperar en forma precisa el documento o documentos con el mismo contenido informativo lo cual no puede lograrse adecuadamente con el lenguaje natural.

Palabras Clave: Descriptores, Tesoros, Documento Digital.

THE THESAURUS IN A DIGITAL ENVIRONMENT CATALINA NAUMIS-PEÑA

ABSTRACT

The text explains the way in which the structure of a documentary language, of a thesaurus type, is complemented with the cycle of organization of information in a digital environment. The thesaural relationships behave in the same way as the structure of the data in an informatic environment, and therefore its use simplifies the thematic classification in order to represent those texts and images which contain knowledge and learning explicit or implicit. Thus, in a digital document thesaural descriptors substitute the natural language and allow a precise document retrieval with the same informative content, something which cannot be achieved with natural language.

Key Words: Descriptors, Thesaurus, Digital Documents.



- 1 Agradezco la revisión, comentarios y sugerencias sobre los borradores previos de este artículo que tuvieron la paciencia de leer Surya Peniche, Roberto Garduño y Elsa Ramírez

*¿Dónde está la sabiduría que hemos perdido
con el conocimiento?*

*¿Dónde está el conocimiento que hemos perdido
con la información?*

Thomas Stearns Eliot

INTRODUCCIÓN

Este epígrafe encabeza un libro publicado recientemente,² evidencia la pérdida de elementos del conocimiento que ha surgido en el proceso del ciclo social de la información y sostiene el supuesto de que la forma de expresión del pensamiento es el lenguaje y que la forma de expresión del conocimiento es la información.

A través de la información organizada se transmite el conocimiento generado en los diferentes ámbitos del quehacer humano, por lo que idealmente esa organización debe poseer los elementos que permitan una recuperación oportuna y eficiente de la información. Al evolucionar la tecnología se desarrollan nuevos paradigmas que en ocasiones le restan validez a los viejos sistemas de catalogación y ordenación. La intención de este trabajo es valorar el papel que desempeña el tesoro para hacer organizaciones temáticas en el entorno digital y reducir así la pérdida de conocimiento. También busca fundamentar una postura contraria al utilitarismo económico e intelectual que se le otorga al lenguaje natural como una opción para recuperar la información.

Se utilizará aquí como método la definición de elementos, características y relaciones del tesoro como un sistema que se interrelaciona en forma más estrecha que el lenguaje natural, (sin menoscabar a éste como una opción de búsqueda) con el documento digital como otro sistema en sí mismo. El tesoro y el documento digital se interrelacionan en la recuperación de temáticas dentro de un sistema de información para facilitar el acceso y uso al propio documento digital. El presente estudio se restringirá a productos multimedia, en presentación digital para una base de datos de información educativa; es decir información visual, auditiva y textual con un propósito académico.

Los sistemas de información son parte del ciclo productivo del conocimiento, cuyo cometido es la transferencia de éste para el uso social. Los registros de los documentos informativos conforman el sistema de información y están estructurados de modo tal que no sólo se puedan consultar uno por uno, sino también recuperar por puntos de acceso estratégicamente seleccionados, o bien pueden agrupar elementos comunes de varios registros mediante la activación de categorías de

2 Raffaele Simone. *La tercera fase: formas de saber que estamos perdiendo* / Traducción de Susana Gómez López. – Madrid, 2001. Taurus. p. 7.

documentos que contienen los mismos conceptos. Lafuente define las operaciones de categorización de información sobre documentos que presenta como la generalización de la abstracción de un concepto en relación con todos los documentos que lo contengan:

Los resultados del análisis de grupos de documentos se expresan en forma de categorías de documentos, es decir en conceptos abstractos que definen sus propiedades comunes y sus relaciones más generales. Estas categorías son el resultado de una abstracción que generaliza los aspectos particulares o singulares de los documentos que produce y utiliza una comunidad³

Las propiedades de categorización y agrupamiento de los documentos analizados en los sistemas ayudan al usuario a solucionar problemas específicos de información. Los temas comunes que se pueden encontrar en un sistema de información están organizados con una clasificación temática mediante un control del vocabulario que ayuda en la conceptualización de las temáticas contenidas tanto al analista como al usuario.

El paradigma de la base de datos responde a una visión del mundo que consiste de objetos y sucesos que se pueden describir a la perfección a través de un conjunto de frases especiales en el registro bibliográfico. Una base de información, por definición, se tiene que ordenar con algún objetivo en mente y sólo se puede tener acceso a ella a través de un sistema de clasificación que represente ese orden.⁴

Para representar ese orden en relación con la temática de la información contenida en un sistema de información, se utilizan lenguajes documentales que exhiben una estructura vertical y horizontal conformada por un conjunto ordenado de procedimientos. Relacionados entre sí tales procedimientos contribuyen a realizar la indización de los contenidos y elementos constitutivos de la información digital que se prepara para su posterior recuperación.

El lenguaje controlado con mayor nivel de estructuración de los elementos lingüísticos es el tesoro y sirve para darle consistencia a la indización y permitirle mejor acceso del usuario,⁵ sin descuidar la versatilidad de la lengua. Un tesoro se define como una estructura lingüística de clasificación temática o facetada, cuya organización básica la conforma una relación de descriptores que representan o describen autoridades o contenidos semánticos representados por unidades lingüísticas extraídas del lenguaje formal de una disciplina o área específica del conocimiento.⁶

En los tesauros se conectan por su significado descriptores y términos equivalentes y descriptores con descriptores, de modo tal que uno nos conduzca a los otros con rigor en la reciprocidad.

3 Ramiro Lafuente-López. "Sobre el análisis y representación de documentos", 2001. p. 163.

4 Colin Beardon y Suzette Worden. "El conservador virtual: las tecnologías multimedia y el papel de los museos", 1997. p. 98.

5 Catalina Naumis Peña. "Dinámica en la construcción de tesauros y el acceso a la información", 1998. p. 73.

6 Catalina Naumis Peña. "Análisis de la confluencia entre término y descriptor en la elaboración de tesauros", 2000. p. 104.

Los objetivos básicos de un tesoro son:

- 1) Servir de vocabulario oficial para coordinar dos procesos: la indización y la recuperación de los documentos digitales de un sistema.
- 2) Proponer un sistema de símbolos lingüísticos para agrupar información similar relacionada o guiarla hacia grupos más específicos o más generales de una temática
- 3) Lograr una normalización de la terminología del sistema de información para el cual se va a utilizar.
- 4) Proponer un conjunto estructurado de términos sobre la base de un sistema de conceptos aptos para organizar los contenidos de ese sistema.

Algunos de los objetivos de los tesauros coinciden con las necesidades de símbolos lingüísticos para construir relaciones de metadatos en un ambiente digital, porque también buscan claridad y mnemotecnia, características de los descriptores de un tesoro.

Sin embargo a pesar de esta utilidad complementaria de los tesauros no se construyen los suficientes en español, ni siquiera para cubrir los requerimientos mínimos de los sistemas de información. Una de las razones principales de esto es el trabajo que supone y su alto costo de desarrollo, ya que hasta ahora se han construido en forma manual. La literatura sobre el tema cuenta con procedimientos para elaborar tesauros automatizados y propone modelos de varios tipos: lingüísticos, estadísticos y cognitivos. En general los resultados han permitido utilizar estos modelos para contribuir al desarrollo del tesoro en forma semiautomática, pero no se ha podido prescindir del ser humano para hacer construcciones tesaurales.⁷

Las estructuras que se han estado utilizando en los buscadores para recuperar información en la red al realizar una búsqueda mantienen una relación de lo general a lo particular que ha probado ser muy arbitraria.⁸ A diferencia de estos resultados, las relaciones tesaurales se obtienen a través de un estudio conceptual de las temáticas, y enfrentándolas a la práctica de la indización y recuperación de información. El tesoro refleja tanto los conocimientos de una disciplina como las necesidades de sus usuarios. En general las estructuras de las redes informáticas actuales (por lo menos en español) son producto de una praxis no contrastada con la polisemia de los lenguajes y otros fenómenos lingüísticos que afectan su efectividad.

El tesoro es el producto de un proyecto que toma en cuenta todos los elementos de un sistema de información y desarrolla el mapa conceptual de conocimientos de

7 En la red se pueden obtener referencias para la consulta de páginas que contienen información sobre tesauros ya construidos e información que ayuda en el tema:

<http://www.willpower.demon.co.uk/thesbibl.htm>

También se puede encontrar un portal que contiene información sobre programas preparados para construir tesauros. La dirección para su consulta es:

<http://www.willpower.demon.co.uk/thessoft.htm>

8 Cory Doctorow. Metacrap: putting the torch to seven straw-men of the meta – utopía: schemas aren't neutral. –, 2001. En línea, versión 1.3. – Disponible en:

<http://www.well.com/~doctorow/metacrap.htm>

una especialidad, partiendo de las categorías principales que tales conocimientos agrupan de acuerdo con sus elementos comunes, los que a la vez separa con base en otros elementos que establecen las diferencias. A su vez estas categorías principales agrupadas en clases se desglosan en las siguientes categorías secundarias, que se van reproduciendo tantas veces como la especificidad del sistema lo requiera.

Para organizar un sistema de información digital los documentos deben ser representados desde el punto de vista del tema o acontecimiento tratado a lo largo del documento, al igual que en los sistemas de información impresa tradicional, pero además deben representar objetos muy puntuales, aunque la aparición sea fugaz (como en el caso de personajes, instituciones o información muy específica) y esté ligada o no con el tema general del documento digital original o principal.⁹

Es importante revisar las formas, categorías gramaticales, accidentes gramaticales y relaciones tesaurales de los descriptores, para comprender como corolario no sólo su adecuación a la estructura en un ambiente digital, sino los cambios que presentan esos elementos a la luz de un sistema que tiene la capacidad de integrar las imágenes y el audio. Los cambios pueden no ser sustanciales, pero se aprecia su existencia al entrar en la discusión de cada uno de sus aspectos por separado.

FORMA DE LOS DESCRIPTORES CONSTITUTIVOS DEL TESAURO

Los descriptores y sus sinónimos (no descriptores) constituyen un subconjunto controlado del lenguaje natural y son representados por términos simples (constituidos por una palabra), cuando el significado es inequívoco, para el sistema de información. Esta claridad está implícita en el término y en su contextualización en la entrada del tesoro. En el medio digital la precisión de elementos simples aumenta la aparición de nombres de cosas, y por lo tanto la ocurrencia de términos simples es mayor que en tesauros del medio impreso. En un tesoro para el medio digital es común encontrar términos simples como los ejemplos siguientes: EPAZOTE, SOFÁ, LÁMPARA. CAMA. TAMBOR, GUITARRA.

La regla en el tesoro es escoger descriptores que representen conceptos simples o unitarios, que no siempre quedan representados por un término simple y se requiere por lo tanto de la composición de términos.

Los términos compuestos son sintagmas terminológicos, es decir contienen un conjunto de varias palabras. A diferencia de los términos simples los compuestos exhiben un mayor nivel de precoordinación cuyo significado es una mayor especificidad para la representación de conocimiento. La combinación de los elementos terminológicos reproduce una suma de los significados semánticos para aumentar la precisión o expresar un concepto simple:

9 Mary A. Burke. *Organization of multimedia resources: principles and practice of information retrieval*, 1999. p. 96.

Ejemplo:

INSEMINACIÓN ARTIFICIAL
CONTROL DE PLAGA
REACCIÓN BIOQUÍMICA

Los términos compuestos se utilizan para aumentar la precisión porque la heterogeneidad que se logra a través de la combinación de dos términos simples no le soluciona al usuario la ubicación de la información que busca en un documento. A veces los documentos contienen dos palabras que el usuario puede asociar, pero que en el documento no representan el concepto que el usuario busca; se trata de dos palabras utilizadas por separado o con conectores que le otorgan un significado muy diferente al buscado.

Ejemplo:

En un documento se pueden encontrar estos dos términos simples:

1. INVESTIGACIÓN
2. TÉCNICA

El documento puede tratar temáticas completamente diferentes, algunas de las cuales estarán muy bien representadas por los dos términos combinados de la siguiente manera:

1. TÉCNICA DE INVESTIGACIÓN
2. INVESTIGACIÓN TÉCNICA

Cuando la base de datos que se consulta es de gran tamaño, la utilización de términos con un mayor grado de precisión ayuda a la exactitud de los documentos recuperados. Los términos compuestos aclaran la información para ubicarla en un contexto cuando los términos simples afectan la comprensión de los usuarios del sistema. En el medio digital en general los términos compuestos son aplicables a las temáticas que tratan los documentos.

La división sintáctica es una técnica que se aplica a términos que pueden ser objeto de un análisis morfológico en componentes separados, cada uno de los cuales puede ser aceptado por sí mismo como término de indización.¹⁰

Cuando se hace efectiva la división sintáctica en un término de indización demasiado largo, la nota de alcance se usa como recordatorio de que algunos términos deben asociarse con otros para culminar la representación de un concepto. Esta nota le recuerda al indizador y también a quien busca información, cómo debe establecer la asociación de términos para indizar o localizar un concepto.

Esta nota se puede utilizar además para aquellos casos en que los términos exhiben perspectivas diferentes desde diversos ámbitos.

10 Grupo de Terminología INFOLAC/SNI. *Guía metodológica para la elaboración de tesauros*. (Colombia), 1992. p. 43.

Ejemplo:

CIRCUITO DE LA COMUNICACIÓN

NA: Cuando utilice este descriptor se pueden agregar otro u otros que especifiquen un enfoque diferente de contenido que el sugerido en el contexto de esta entrada del tesoro. Ej.:

CIRCUITO DE LA COMUNICACIÓN + MEDIOS DE COMUNICACIÓN

TG: LENGUA

TE: ELEMENTO DEL ACTO COMUNICATIVO
INTERFERENCIA DE LA COMUNICACIÓN
PROCESO DE COMUNICACIÓN
SISTEMA DE COMUNICACIÓN

TR: COMUNICACIÓN

Es una verdadera sugerencia tanto al indizador como al buscador de información para que tengan presente esta asociación e incluyan los dos términos que aparecen unidos por un signo de más. En la base de datos los dos términos estarán incluidos como separados.

Los homógrafos son términos con la misma ortografía pero significados diferentes. Este tipo de término no puede incluirse solo, sino que se le deben agregar elementos que aclaren los distintos significados, ya sea un calificador o término entre paréntesis o bien el acompañamiento del término por otro que le agregue la especificación necesaria para ubicarlo en el contexto.

Ejemplo de homógrafos resueltos con calificadores

RADIO Para designar el medio masivo de comunicación
RADIO (APARATO ELÉCTRICO)
RADIO (ELEMENTO QUÍMICO)
RADIO (HUESO DEL BRAZO)

Ejemplo de homógrafo aclarado con el uso de término compuesto

RADIO En Lenguaje y comunicación
RADIO DEL CÍRCULO En Matemáticas

Los términos polisémicos en cambio son aquellos que por sí solos tienen un significado múltiple, en este tipo de términos la creatividad metafórica juega un papel preponderante dado que surgen como una extensión de significado en el uso del término original. Sin embargo para fines del tesoro la solución que se adopta para aclarar el sentido en el que se debe usar el término polisémico es la misma que ocurre con los homónimos.

OVARIO En Ciencias de la Salud
OVARIO DE LA FLOR En Botánica

CATEGORÍAS GRAMATICALES DE LOS TÉRMINOS TESAURALES

Tanto los descriptores como los no descriptores deben ser sustantivos o la forma sustantiva equivalente:

Ejemplos:

Minimizar costos	deberá incluirse como	MINIMIZACIÓN DE COSTOS
Entretener	“ “ “	ENTRETENIMIENTO
Envenenar	“ “ “	ENVENENAMIENTO
Invertir	“ “ “	INVERSIÓN
Expandir	“ “ “	EXPANSIÓN

Los adjetivos acompañan a los sustantivos para calificarlos cuando así lo requiera el sistema de información.

Ejemplo:

FORMA MUSICAL
ENSAYO TEÓRICO

En un tesoro que debe reflejar contenidos digitales con imágenes es necesario mantener adjetivos que califiquen situaciones que deberán especificarse para aclarar una imagen, a diferencia de lo que ocurre en los tesauros tradicionales:

Ejemplos:

VERDE
BLANCO

También se usan como adjetivos sustantivados los participios de los verbos, como se describe en el apartado de verbos.

Los adverbios, al igual que en los sistemas tradicionales, no se usan como términos de indización simples, pero pueden formar parte de un descriptor compuesto cuando exista un término de indización que lo incluya y esté validado por el argot de alguna disciplina, y en general calificando a un adjetivo u otro adverbio.

Los verbos en general se usan como formas sustantivadas o adjetivadas tal como se explicó en el apartado de sustantivos. Las acciones no están representadas en el tesoro, más bien lo que interesa es el resultado de la acción.

Ejemplo de participios:

ARTESONADO
DORADO
ENCHAPADO

Para la calificación de un documento con imágenes, por ejemplo para una programación de televisión, puede ser necesario representar acciones, pero como tales el verbo sería incluido en infinitivo. La inclusión de verbos conjugados es otro nivel

de estructura gramatical que haría que el descriptor dejara de ser un sintagma terminológico para convertirse en oración.

Se usan preposiciones y artículos en medio de los descriptores para conectar términos. Este tipo de conector juega un papel fundamental en la recuperación de información mediante indización por tesoro y hace la gran diferencia con los términos extraídos para la indización automatizada, que los suprimen a través de los diccionarios de términos no significativos. Su presencia les otorga un significado diferente a los términos compuestos que unen.

Ejemplos:

DIBUJO A ESCALA
ANIMACIÓN POR COMPUTADORA
ESTRUCTURA DE LA HOJA

ACCIDENTES GRAMATICALES EN EL TESAURO

Las normas ISO 2788 especifican que se use la forma singular para sustantivos no cuantificables, nombres de conceptos abstractos como procesos específicos, propiedades (atributos), sistemas de creencias, actividades y disciplinas, y la forma plural para sustantivos cuantificables.

Para determinar el uso de la forma plural se sugiere la pregunta al candidato descriptor sustantivo ¿cuántos/as?. La forma singular se usará cuando el sustantivo responda a la pregunta ¿cuánto/a?

Las mismas reglas proponen como ejemplo de singular el sustantivo PINTURA. Esto será válido en un sistema de información especializado en arte, pero en un sistema de información para educación, PINTURA también responde a la pregunta de ¿cuántas? Porque existen muchos tipos de pinturas desde rupestres hasta los diferentes tipos de pintura para paredes.

Una decisión que ya ha sido adoptada en tesauros en español y que es más adaptable para un ambiente digital dada la mayor cantidad de descriptores que deben ser incluidos para representar los diversos atributos,¹¹ es utilizar la forma singular tal como se hace en la terminología, criterio también adoptado en la traducción al español del Tesoro de la UNESCO.¹²

Sin embargo, en los casos en que las formas singular y plural de un término designen diferentes conceptos, las dos deben ser incluidas en el tesoro.

11 Hsin-Liang Chen y Edie M. Rasmussen. "Intellectual access to images" -, 1999. p. 295.

12 UNESCO *Thesaurus = Tesauros de l'UNESCO = Tesoro de la UNESCO: lista estructurada de descriptores para la indización y la recuperación bibliográficas en las esferas de la educación, la ciencia, las ciencias sociales y humanas, la cultura, la comunicación y la información.* — París, Francia: UNESCO, 1995. — xxxix, 705 p.

Ejemplo:

FOTOGRAFÍA
FOTOGRAFÍAS

Para designar a la técnica o arte
Para designar a los productos

RELACIONES TESAURALES

Las relaciones tesaurales se dividen en relaciones de equivalencia que enlazan términos sinónimos o equivalentes con descriptores; relaciones jerárquicas, que enlazan términos generales y específicos (hipónimos e hiperónimos), y relaciones asociativas, que enlazan términos relacionados y notas de alcance para acotar el significado del término en un sistema de información.

El primer caso, los enlaces de términos sinónimos o equivalentes con descriptores, permite descartar el uso de diferentes formas para denominar un mismo concepto, pero mediante su relación se conoce cuál es el descriptor validado para la indicación y la recuperación de información. Los sinónimos son los términos con igual significado pero la sinonimia absoluta es rara, lo más frecuente son los cuasi-sinónimos, expresiones que son más o menos semejantes en significado pero no idénticas, como es el caso de los sinónimos absolutos.¹³ Se distinguen además los sinónimos parciales que son otra categoría diferente a los absolutos y los cuasi-sinónimos. Es decir existen diferentes grados de sinonimia y por tanto es importante considerar que el tratamiento que se le da en el tesoro es el mismo para todos los casos.

Los sinónimos o términos equivalentes del tesoro son escogidos por el constructor del tesoro con la finalidad de agrupar información para la indicación (es decir con fines prácticos), y por tanto no siempre se incluyen los sinónimos aceptados por la gramática tradicional.

Así, se pueden hacer equivalencias: de las abreviaturas a su nombre completo; de términos con variaciones ortográficas; de términos no incluidos en diccionarios integrantes de la jerga de usuarios del sistema, y de expresiones compuestas, sinónimos verdaderos y cuasi-sinónimos.

Los criterios para seleccionar los términos equivalentes deben tomarse de acuerdo con las NECESIDADES de la comunidad de usuarios a la cual se destina el sistema.

CONTRABAJO
UP: **Violón**

DESCRIPTOR
TÉRMINO EQUIVALENTE O
SINÓNIMO (NO SE USAPARA LA
INDIZACIÓN)

Si un indizador o usuario busca en el tesoro el término equivalente no autorizado encontrará la relación contraria:

13 John Lyons. *Lenguaje, significado y contexto*, 1981. p. 56.

Violón

USE CONTRABAJO

Si un indizador o usuario busca en el tesoro por cualquiera de los términos equivalentes no autorizados, que pueden ser varios, encontrará las relaciones contrarias para cada uno de ellos. Esto permite una concentración y agrupación de información que optimiza los contenidos digitales.

Las relaciones jerárquicas comienzan en el campo semántico más general, para llegar al más específico. Los términos descriptores se dividen entonces en genéricos y específicos.

Para ello los términos son analizados de modo tal que se realiza la descomposición del sentido en sus elementos constituyentes. El sentido de cada palabra debe ser representado como el producto de varios factores, y cada uno de esos factores es a su vez descompuesto para ir conformando los diferentes niveles de lo más general a lo más especializado.

La relación de hiponimia se ejemplifica mediante parejas de expresiones del tipo de *perro* y *animal*, de las que el primero es un hipónimo del segundo: el sentido de *perro* incluye el de *animal*.¹⁴

Por lo tanto las relaciones jerárquicas se establecen con base en las relaciones semánticas de los términos y niveles, y van de lo más general a lo particular. En el tesoro se debe escoger la agrupación jerárquica según características comunes inherentes e invariables y esto parecería muy claro, pero hay términos cuya naturaleza no es determinante para escoger su ubicación en una categoría u otra. No siempre la decisión es fácil porque existen diferentes doctrinas y pensamientos acerca de la organización de cada ciencia. En el tesoro, como herramienta lingüística, se asumen las decisiones con fines prácticos para ofrecer una consistencia que permita organizar la información.

Como se verá más adelante el enlace jerárquico es el más cercano a la estructura del documento digital interactivo para el cual se pueden encontrar incluso varios niveles de relaciones de metadatos hacia temas más puntuales que los presentados en el documento principal, y esto es válido tanto para una página WEB hasta un artículo científico, porque ambos se organizan de lo general a lo particular.

Utilizamos relaciones asociativas para organizar las relaciones en el nivel horizontal, donde no se puede establecer una jerarquía pero sí relaciones semánticas para conectar los términos entre ellos y recordarle al indizador o a quien busca información, que además del término utilizado existen otros términos bajo los cuales puede agrupar información o, en su caso, localizarla.

14 John Lyons. *Lenguaje, significado y contexto*, 1981. p. 97.

Las relaciones asociativas juegan un papel muy importante porque se utilizan para conectar términos que dependiendo del punto de vista pertenecen a varias categorías.

En la guía preparada por el grupo de Infolac¹⁵ se establece que la relación asociativa puede unir dos tipos de términos:

- ❖ DESCRIPTORES PERTENECIENTES A LA MISMA CATEGORÍA
- ❖ DESCRIPTORES PERTENECIENTES A CATEGORÍAS DIFERENTES

La nota de alcance explica el uso de un descriptor en un sistema de información, es decir el sentido en el que se entiende y el que debe ser usado para indizar y recuperar información, pero a diferencia del diccionario no presenta todas las acepciones de un término. En un sistema de información el significado del término es mucho más acotado al intentar darle consistencia tanto a la indización como a la recuperación de información en el sistema.

La nota de alcance no es el único modo de reducir el significado del término, porque a través de las relaciones con otros términos se asigna significado adicional, pero este tipo de nota es determinante porque la presencia de una frase ayuda a clarificar y limitar el concepto.

INDIZAR PARA RECUPERAR

Los elementos, características y relaciones constituyen al tesauro, pero las etapas siguientes son las que le otorgan su verdadero sentido: la indización y la recuperación. La indización es la utilización de un signo o símbolo autorizado para representar o significar un concepto determinado, después de hacer un análisis de contenido del documento y por eso, como se dijo con anterioridad, constituye un elemento más del registro documental.

En griego, el antónimo exacto del símbolo es el diablo; el que separa, Dia-bólico es todo lo que divide, sim-bólico todo lo que acerca.¹⁶

La operación de indización puede ser el resultado de un análisis humano que utiliza el tesauro como un sistema de signos autorizado o la puesta en marcha de un análisis automático a partir de determinados campos que se eligen para representar la información, tales como el autor o el título de la obra.

El tesauro como sistema de signos se estructura en una clasificación del conocimiento para asignarle a las diferentes piezas de información un lugar en esa clasificación, pero es necesario precisar esta operación de indizado sobre una base temática con respecto a la operación de clasificación antes de asignar una signatura tipográfica.

15 Grupo de Terminología INFOLAC/SNI (Colombia), *Guía metodológica para la elaboración de tesauros*, 1992. – p. 79.

16 Regis Debray. *Vida y muerte de la imagen: historia de la mirada en Occidente*. –, 1992. p. 53.

El tesoro se utiliza para indizar y la diferencia que presenta respecto de la clasificación tiene que ver con la propia naturaleza de los lenguajes que se utilizan para una y otra operación. Un “esquema de clasificación” le asigna al documento su lugar en una sola clase; a diferencia de eso el tesoro permite asignar varias clases que se pueden combinar para recuperar el mismo documento en un sistema de información, lo cual otorga mayor flexibilidad y convierte el sistema en más sensible para ampliar o limitar la respuesta que requiere un usuario.

Otro elemento de diferenciación que tienen los sistemas clasificatorios es el nivel de precoordinación y postcoordinación. La relación en un sistema precoordinado incluye una expresión sintagmática de muchos elementos al momento de clasificar y esos mismos elementos deben utilizarse en la búsqueda de recuperación de información. El sistema postcoordinado incluye elementos sintagmáticos aislados al momento de la indización, que se conjuntan al momento de localizar la información.

Un lenguaje postcoordinado implica que se pueden realizar mayor cantidad de combinaciones de información al momento de salir ésta del sistema, al revés de lo que ocurre con la utilización de una precoordinación, que restringe las combinaciones pero le otorga mayor precisión al sistema, como ocurre en el tesoro.

Sistema precoordinado

Ejemplo:

Término de indización: CONVENIOS SOBRE PRODUCTOS
BÁSICOS

(En la ordenación clasificatoria de una biblioteca se juntan todos los CONVENIOS SOBRE PRODUCTOS BÁSICOS y una búsqueda simplemente por CONVENIOS soslaya en sus resultados los CONVENIOS SOBRE PRODUCTOS BÁSICOS.

Sistema postcoordinado

Ejemplo:

Término de indización CONVENIOS

Término de indización PRODUCTOS BÁSICOS

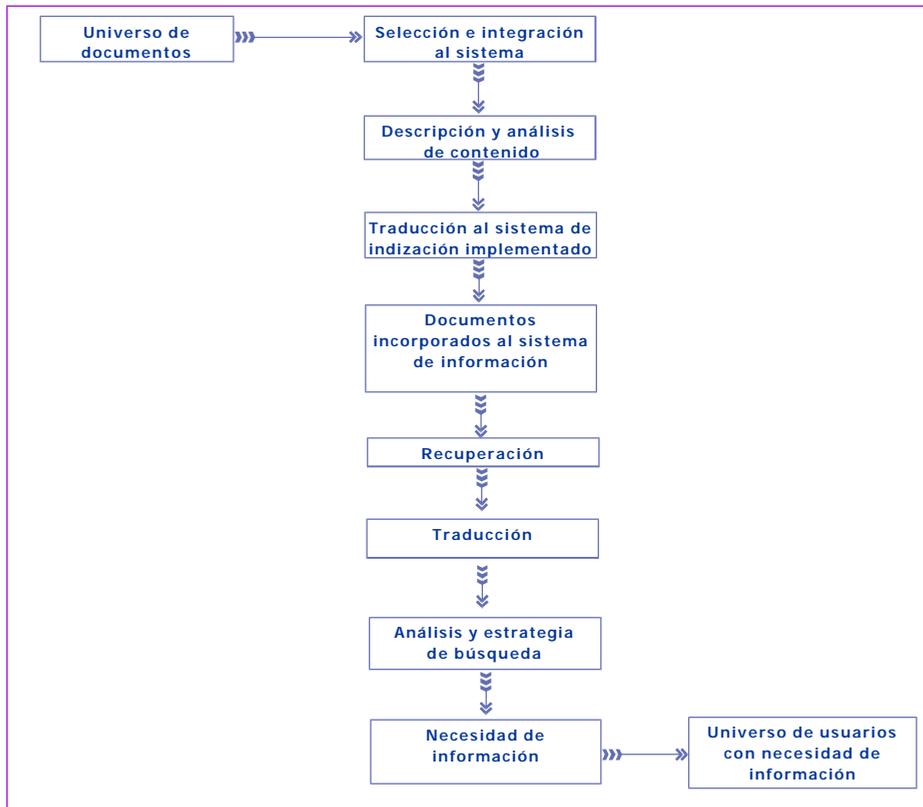
(En un sistema de información existen documentos agrupados bajo CONVENIOS y documentos agrupados bajo PRODUCTOS BÁSICOS, la búsqueda permitirá localizar documentos que traten acerca de modelos de convenios u obtener un listado de productos básicos o solucionar cualquier otra pregunta relacionada con cualquiera de los temas aislados o combinados entre sí, como podría ser CONVENIOS DE PRODUCTOS BÁSICOS o asociar esto además con otros elementos, como podría ser CONVENIOS DE PRODUCTOS BÁSICOS EN MÉXICO)

Otra diferenciación que conviene dejar asentada es la indización automática y la indización humana:

La indización automática es la operación que consiste en que el ordenador reconoce los términos que figuran dentro del título, del resumen, del texto completo (si éste ha sido almacenado junto con la descripción documental) y a veces también dentro de la indización humana, y emplea estos términos, o bien tal cual, o bien después de transformarlos en otros términos, equivalentes o conceptualmente próximos, con el fin de convertirlos en elementos que se incorporan al fichero de búsqueda y quedan disponibles para recuperar el documento.¹⁷

Casi todas las operaciones de indización realizadas para un sistema de información de calidad o individualmente en un centro de documentación, se apoyan en un vocabulario controlado, como un tesoro.

La indización se retroalimenta de las consultas recibidas en el sistema porque el indizador debe hacer confluír la necesidad de información de un usuario con la respuesta que el sistema le brinda a través de los documentos que contiene:



17 Georges Van Slype. *Lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*, 1991. p. 125.

Ante la solicitud de información de un usuario el proceso de indización deberá hacerse a la inversa:

- ❖ Indizar la solicitud y formular la prescripción de búsqueda.
- ❖ Trazar la estrategia de búsqueda
- ❖ Confrontar los términos índices de la prescripción de búsqueda con los descriptores utilizados para la indización de la colección.
- ❖ Recuperar el documento adecuado

La tarea de indización es de gran importancia porque de las decisiones tomadas por ella para representar los documentos en el sistema dependerá su posterior recuperación.

DOCUMENTO DIGITAL

Los sistemas de información se incrementan cada vez más por la presencia de documentos digitales que se evalúan, seleccionan, organizan y son puestos en bibliotecas digitales para ofrecer una serie de servicios complementarios a los ofrecidos por la biblioteca tradicional, lo que se logra a través de un acceso basado en comunicaciones remotas y una mayor diversidad en la integración de recursos informativos como el texto, la imagen y el audio. El concepto de biblioteca digital está constituido por una serie de elementos abordados en fechas recientes por diferentes autores, que se resumen para los fines del presente trabajo a través de la opinión de Torres Vargas.¹⁸

[...] la biblioteca supera su papel de agente intermediario de documentos impresos y electrónicos, y asume el de productor de información y navegador en la red, digitalizando documentos de su colección y creando su página de presentación en Internet, como instrumento principal, y el acceso remoto a los OPAC (On-line public access catalogue) a través de la Web, para hacer búsquedas y pedidos de documentos. Además la biblioteca les ofrece a sus usuarios un conjunto de enlaces de interés con fuentes electrónicas, y se encuentra en disposición de ofrecer una combinación de los servicios que presta su personal *in situ* así como los servicios electrónicos a los que puede acceder cualquier persona desde su oficina o casa.¹⁹

En este nuevo ambiente digital el bibliotecólogo se transforma en una especie de cartógrafo que ayuda a navegar por el ciberespacio. Y como en los mapas geográficos, la información en los sistemas de información debe estar organizada en varios niveles de categorías de elementos comunes según el nivel de generalidad y especificidad de información que se requiera para acceder tanto a los temas generales como a los específicos.

18 Georgina Araceli Torres Vargas. *La biblioteca virtual ¿que es y que promete?*. – 2000 p. 27-35.

19 *Ibid* p. 33.

Otro elemento nuevo en los documentos digitales es la integración de documentos escritos complementados con imágenes y audio que constituyen una nueva responsabilidad para los analistas de información, quienes deben organizar contenidos muy diversos en documentos completos o parciales sin omitir datos que pueden ser relevantes, y también relacionar los conocimientos incluidos en ellos.

Por tanto la expresión “documento digital como objeto” implica una variedad de elementos y características que deberemos precisar para entender las operaciones de indizado y recuperación de información que intentan facilitar el análisis de sus contenidos temáticos. Las computadoras han evolucionado respecto a su primer papel de procesadores numéricos con interfases textuales, y se han convertido en procesadores de alta velocidad que tienen la capacidad de enlazar al texto con imágenes a todo color y sonido.

El documento digital que se usará como base para el estudio en este trabajo es aquel cuyo elemento principal es la imagen enlazada con texto hablado y sonido, pero que además ha pasado por mecanismos de validación de la calidad de la información que justifican la inversión para su indización y recuperación. Tampoco se consideran documentos digitalizados sin interactividad, ni documentos analógicos, pero es requisito imprescindible para sus fines precisar los elementos que lo componen y sus características. Este tipo de documento digital es el multimedia, expresión que incluye una gran variedad de unidades de información dentro de su composición.

En primer lugar se tratará de dilucidar el concepto de documento multimedia, ya que los elementos que lo componen son los mismos que los del documento digital multimedia, en tanto que los formatos que se utilizan exhiben características comunes y diferencias que tienden a desaparecer con el desarrollo tecnológico. Los elementos comunes tanto del concepto de documento digital como del documento multimedia están ligados con el manejo interactivo de diferentes tipos de información: visual, auditiva, y discurso escrito y hablado. Mary Burke después de analizar varias definiciones de “documento multimedia” lo presenta como sinónimo del concepto “documento digital”, aunque no lo manifiesta explícitamente.

[...]almacenamiento integrado, recuperación y despliegue de palabras, números, imágenes y sonido por una computadora.²⁰

El propio concepto de documento u objeto de información es complejo, está abierto a muchas interpretaciones e incluye la posibilidad para combinar varios componentes en un conjunto de sistema multimedia que le agrega elementos a esa complejidad. Se dan todos los conceptos de objetos de información y de relaciones entre ellos. Según otros autores los objetos físicos corresponden a cintas o discos de video o audio o formatos equivalentes, y los objetos lógicos corresponden a programas o exhibiciones, colecciones, series, partes y piezas que corresponden a series de

20 Mary A. Burke. *Organization of multimedia resources: principles and practice of information retrieval*. – 1999 p. 7.

programas, el programa de un día específico, antes sin su correspondiente programa y secciones de ese ente.

[...]podemos definir como documento multimedia a “aquel capaz de presentar información textual, sonora y audiovisual de modo integrado y coordinado: gráficos, fotos, secuencias animadas de video, gráficos animados, sonidos y voces, texto[...].”²¹

De acuerdo con esta definición el documento multimedia difiere del digital porque éste no incluye la integración de medios analógicos o gráficos, y porque todos los elementos del documento digital son digitales, aunque es importante reconocer, con Mary Burke, que cada vez con mayor frecuencia el ambiente multimedia es digital por la facilidad que ofrece para integrar físicamente diferentes objetos lógicos. La producción documental académica tiende cada vez más al uso de estándares abiertos como el SGML, HTML y XML (desarrollados como lenguajes de marcado) para la estructura de documentos digitales.²²

El concepto de documento digital a su vez está ligado al hipertexto, porque no es posible concebir un documento digital con las características de un documento tradicional como una sola obra terminada. El reto principal de un documento digital es cómo escribir y desarrollar la información en una forma no lineal para que el usuario obtenga una visión lineal de la misma. El texto escrito a través de íconos es el que se encarga de conectar las formas lineales y no lineales de presentación, para crear, ver y manipular la información.

En este sentido, la World Wide Web permite la organización de documentos digitales en modo de hipertexto. El hipertexto se sustenta en métodos de organización de la información a través de los cuales los diferentes elementos del contenido pueden ser ligados mediante los elementos del propio contenido. Lo anterior determina la intervención del bibliotecario en la fase de organización normalizada de los documentos digitales existentes en Web, y esto lo remite a una tarea novedosa en el sentido de que tiene que manejar documentos en modo de hipertexto, lo cual implica información almacenada en forma digital con vínculos de referencias hacia sitios Web; avanzar de un recurso informativo a otro; navegar por el texto, y obtener información con base en los intereses específicos del usuario final de la información.²³

Para facilitar el acceso a los recursos de información digital se plantea organizar ésta en metadatos presentando los datos que permiten conocer el documento al que se busca llegar, entre los cuales ocupan un lugar preponderante los términos de indicación que resumen el contenido temático del documento.

21 (Bartolomé, 1994) citado por José Sánchez Rodríguez y Ángel Ramón Romance García. “Multimedia”. 2000 p. 202.

22 Gary Heba. “Digital architectures: a rhetoric of electronic document structures” 1997 p. 275

23 Ramiro Lafuente López y Roberto Garduño Vera. *Lenguajes de marcado de documentos digitales de carácter bibliográfico*. – 2001 p. 107.

[...]De manera general se puede señalar que los metadatos representan datos acerca de recursos informativos disponibles en redes por una parte, y que conciernen, por la otra, los elementos útiles para facilitar la identificación y localización de recursos digitales.²⁴

Los términos de indización son seleccionados como tales a partir de los conceptos de los documentos y las necesidades de los usuarios mediante un proceso de normalización para luego integrar un lenguaje controlado. En las bibliotecas digitales las necesidades de los usuarios son más difíciles de determinar porque al existir mayor interactividad el espectro de uso se abre no sólo a especialistas en una temática, sino también a especialistas en temáticas relacionadas o usuarios con conocimientos más generales. Sin embargo también es más fácil recoger datos que permitan conocer los intereses de los usuarios porque se pueden incluir en los portales preguntas para evaluar y retroalimentar el sistema, como es habitual encontrar en los portales de orientación informativa,²⁵ los cuales facilitan además la retroalimentación de los lenguajes controlados..

El lenguaje controlado es utilizado al clasificar el documento digital para normalizar los contenidos de la información, tanto a través de la descripción como de su análisis de contenido.

DESCRIPCIÓN DE FORMA Y CONTENIDO EN EL DOCUMENTO DIGITAL

Todo documento digital debe ser analizado y representado desde dos puntos de vista: su forma o física y su contenido, de manera tal que el documento original sea sustituido por un nuevo documento que concisa y claramente exprese los elementos que permitan la recuperación del documento original, y que además sea consistente con la masa de nuevos documentos producidos a partir del análisis y la representación de los documentos originales.

La descripción física se refiere a los datos que presenta el documento y que el catalogador o indizador transfiere al registro de datos bibliográficos.²⁶ La representación física es el resultado de un proceso de catalogación y descripción bibliográfica que debe ubicar y explicar además:

- ❖ EL SOPORTE FÍSICO QUE CONTIENE LA INFORMACIÓN
- ❖ La UNIDAD LÓGICA DE INFORMACIÓN

24 *Ibid.* – p. 63.

25 Kristine Schultz. “An overview of readers’ advisory service with evaluations of related websites”. – p. 21-33. – En *Acquisitions Librarian*. – No. 23, 2000.

26 Información complementaria al registro bibliográfico para documentos documentales se recomienda consultar: Ramiro Lafuente López y Roberto Garduño. *Lenguajes de marcado de documentos digitales de carácter bibliográfico*. — 2001 p. 27.

No siempre un documento digital responde a un objeto informativo, en este tipo de soporte es muy clara la distinción entre el soporte físico que contiene la información –que puede ser un disco, varios discos o un portal de un sistema de información– y la unidad lógica de información que también puede estar contenida en una parte de un portal, un disco, o distribuida en varios documentos digitales. La descripción física de cada una de las unidades lógicas de información en un sistema documental, constituye una información básica para llegar a la unidad física que la contiene.

La descripción de contenido es el conjunto de operaciones que describen el tema y los elementos que tienen que ver con el significado de una obra intelectual y que son importantes para un sistema de información.

En general el objetivo de la descripción de contenido es precisar el campo de conocimiento; clasificar los diferentes temas que lo componen; expresar éstos en términos concretos y sin ambigüedades, y traducir dichos términos a las expresiones más apropiadas del lenguaje, todo con la finalidad de:

- ❖ Reconocer el documento objeto del análisis en una masa de documentos.
- ❖ Representar el conocimiento y saberes implícitos en el documento a través de una clasificación.
- ❖ Sustituir el documento.

Las operaciones factibles de realizar para describir un documento desde el punto de vista de su contenido son: signatura topográfica (número de clasificación de acuerdo con un esquema elaborado de antemano para su ordenamiento físico en una estantería), indización mediante palabras claves, términos de indización o descriptores, y resumen del documento que lo represente en unas pocas frases. Por tanto la finalidad de representación de su contenido y la principal actividad en cada una de sus operaciones es la siguiente:

SIGNATURA TOPOGRÁFICA	ORDENAR
INDIZACIÓN	AGRUPAR
RESUMEN	CONDENSAR Y ABREVIAR

Así una de las tareas del conjunto de operaciones que se deben realizar para organizar documentos es la indización mediante descriptores, cuya función es agrupar información que le es común a diferentes entes informativos. La indización por descriptores permite localizar, por temas, información que puede no quedar claramente definida por un título ambigüo o por palabras claves recogidas del documento, las cuales no siempre logran abarcar el contenido real de una unidad lógica de información porque no ha sido representada por un término.

INDIZACIÓN POR DESCRIPTORES

La indización por descriptores es una operación costosa y su incorporación a un ambiente digital debe estar justificada. En la literatura sobre organización de la información este tipo de indización está integrada como un lenguaje controlado y opuesto a la indización mediante la extracción de términos del lenguaje natural utilizados en bases de datos y redes de información. Se presenta a continuación un cuadro que resume las ventajas del lenguaje controlado sobre el lenguaje natural:

LENGUAJE NATURAL	LENGUAJE CONTROLADO
<i>Desventajas</i>	<i>Ventajas</i>
Diferente denominación para el mismo concepto	Mayor uniformidad y constancia en la denominación de conceptos
Dispersión en la información como consecuencia de las diferentes denominaciones	Agrupamiento de información como consecuencia de la uniformidad y constancia en la denominación de conceptos.
Aislamiento de la información obtenida	Relaciona estructuras jerárquicas y de asociación y explica el uso de un término para el sistema de información
La temática puede no estar expresada en palabras en el documento para que sean utilizadas para la recuperación	La temática será captada en la indización y expresada en términos para su recuperación.
Posibilidad de establecer relaciones incorrectas entre los términos.	Las relaciones entre los términos se mantienen más estables con respecto a los documentos porque hay coherencia entre los sistemas de clasificación y búsqueda
Dificultad para la consulta a los no especialistas en la materia.	Ayuda para conocer el lenguaje de la disciplina a los no expertos
Aumenta los tiempos de búsqueda	Disminuye los tiempos de búsqueda
Incrementa la información no relevante	Precisión en la recuperación

A diferencia de la indización por términos extraídos del lenguaje natural o palabras claves, la información indizada por medio de descriptores es más precisa para la recuperación porque guía los términos de indización no usados hacia los descriptores. Es decir la indización apoyada en un tesauro marca una ruta hacia términos normalizados y validados para cualquier temática, expresada o no, que pueda contener un documento.

Sin embargo, es necesario acotar que la indización en documentos de diferentes tipos considera la inclusión de los temas que se discuten en ellos con suficiente profundidad para luego ser recuperados. En los documentos digitales la situación es diferente y no sólo se deben expresar las temáticas en términos de indización sino

representar elementos de contenido que no necesariamente son temas y que no implican un tratamiento profundo de ellos; a veces la presencia de un objeto en un segundo plano de una imagen es importante para el sistema de información.

LA INDIZACIÓN DE DOCUMENTOS DIGITALES

En los problemas de recuperación de documentos digitales intervienen variables propias que ya han sido mencionadas, como el no estar constituidos únicamente por documentos escritos para los cuales han sido implementados los sistemas tradicionales de indización y recuperación, sino además incluir imágenes, sonidos y audio. El sonido no es una variable que se estudie en el presente trabajo y el audio será considerado como el discurso hablado que acompaña la imagen, pero no se puede dejar de mencionar la imagen como nuevo elemento de aprendizaje o ilustración y el gran impulso que ha obtenido a partir de la tecnología de cómputo para el procesamiento de la imagen.²⁷ Todo ello permite el uso de una imagen en movimiento con una resolución cada vez de mayor calidad que ocupa cada vez menos espacio y se hace más manejable para acompañar los textos del ámbito científico y educativo destinados a apoyar los desarrollos más actuales de los sistemas de información.

Hasta hace pocos años el paradigma dominante era la comunicación lingüística para la información y la enseñanza e incluso para el entretenimiento,²⁸ pero en los documentos digitales multimedia la imagen constituye el elemento central del discurso, como ocurre por ejemplo con una operación grabada para enseñanza de los estudiantes de medicina o la exploración interna del cuerpo humano. Las imágenes son un lenguaje análogo a cualquier otro de los que maneja el ser humano, tales como el verbal o el musical; sin embargo la imagen debe ir acompañada de una explicación de lo que se está viendo para guiar al cerebro y a la vista.

[...] Y sin embargo, la imagen no es la lengua hablada de nuestros niños, pues no tiene ni sintaxis ni gramática. Una imagen no es ni verdadera ni falsa, ni contradictoria ni imposible. En cuanto que no es argumentación no es refutable. Los códigos que puede o no puede movilizar son sólo lecturas e interpretaciones.²⁹

La imagen debe ser representada en sus contenidos para localizarla entre un conjunto de documentos con imágenes y sólo puede ser representada por términos de indización; en cambio del documento escrito o hablado se pueden extraer palabras claves para la indización. Si en un documento escrito se discuten las columnas griegas, por ejemplo, seguramente con un término que represente el tema será suficiente. En un sistema de información que contenga la imagen de una columna corintia,

27 Hsin-Liang Chen. "An analysis of image queries in the field of Art History". 2001 – p. 260.

28 P.G. B. Enser. "Progress in documentation pictorial information retrieval". 1995 – p. 127.

29 Régis Debray. *Vida y muerte de la imagen: historia de la mirada en Occidente*. 1994 – p. 53.

aunque el texto hablado que lo acompaña sea general, si las políticas del sistema son la precisión de elementos artísticos deberá describirse la columna de la imagen y hasta es probable que se requiera agregar el tipo de hoja que adorna su dintel.

En un documento se pueden expresar ideas más abstractas aun cuando se refiera a situaciones más concretas representadas en una imagen. En la imagen es importante también el tema,³⁰ la forma, el color, la época, los detalles, y para ejemplificar se puede mencionar un documento escrito sobre el paisaje de una zona geográfica y una imagen que representa ese paisaje. En un sistema de información el documento escrito quedará bien representado por un descriptor que mencione la zona y el tipo de paisaje, pero un sistema de información de imagen deberá describir la zona, el tipo de paisaje, la estación del año, la hora del día (puede ser el amanecer o atardecer) las especies de árboles más notorias, etcétera.

La información digital incluye construcciones en imágenes que muchas veces deben ser reutilizadas por su alto costo de obtención o construcción: unas veces para grabarlas porque son acontecimientos o personajes que ya no podrían ser vueltos a grabar, o por su lejanía con el entorno físico o temporal, o la inmediatez de la necesidad; otras debido a la dificultad para construirlas y grabarlas, pues aun las imágenes construidas como objetos digitales a veces insumen horas de trabajo de edición.

Además de la necesidad de usar tesauros para indizar y recuperar información contenida en los documentos digitales existe otra necesidad que es la propia organización del documento digital hipertextual. Para organizar los documentos digitales se requieren los lenguajes de marcado, que obligan al autor intelectual a utilizar términos precisos que tengan calidad de representación del contenido como iconos que guían hacia el documento creado. El lenguaje usado en el documento digital es menos redundante que el de un documento escrito en forma tradicional porque existe la posibilidad tecnológica para ligar en forma múltiple un documento con otros temáticamente semejantes. Es decir, se empieza a vislumbrar la necesidad de contar con un lenguaje preciso para redactar el documento digital porque los iconos funcionan como términos de indización al cumplir una función de marcadores de información para una posterior recuperación.

Entre los trabajos incluidos en una reciente recopilación de investigaciones del Instituto Tecnológico de Massachusetts llama la atención el apartado de una en especial que se ocupa de los retos que enfrentan los investigadores en multimedia que se ocupan de la interpretación de la cultura. El investigador se pregunta “¿Cómo empezamos a reunir fragmentos de información en diferentes medios de forma que tengan sentido para nosotros y los otros?” Y menciona a su vez la siguiente cita:

[...] el problema de gestionar la información multimedia no se resuelve con una construcción más contundente de bases de datos, sino que la resolución del problema pasa por la construcción de herramientas para

30 Hsin-Liang Chen y Edie M. Rasmussen. “Intellectual access to images” – 1999 p. 295.

conceptualizar visualmente las relaciones entre la información (Horney, 1994).”³¹

La estructura terminológica que se requiere en los documentos digitales exige símbolos lingüísticos que representen salidas hacia otros elementos del documento que contienen diferentes niveles de jerarquización de los conocimientos y saberes implícitos y explícitos, así como información complementaria en un nivel horizontal, por lo que se asemeja a la estructura de indizado que proporciona el tesoro para representar el contenido de los documentos en el sistema de información. Esta estructura del tesoro lo constituye entonces en una herramienta lingüística para conceptualizar las relaciones de metadatos.

En un trabajo de hace ya algunos años se mencionaba la inversión millonaria que se hacía en personal ayudado por equipos, computadoras, programas de cómputo y redes, y velocidad en el aumento de la producción y transferencia de documentos, en contraste con una fuerte inclinación a no invertir, organizar, describir documentos y aplicar términos de indización basados en una estructura tesaural que los documentos puedan ser identificados y recuperados en el futuro.³²

CONSIDERACIONES FINALES

Tanto en la descripción como en la representación de contenidos, como es el caso de la clasificación temática, está implícita una inexorable reducción de elementos que se sacrifican en virtud de alcanzar las exigencias del análisis y síntesis que se ejerce para traducir el conocimiento que se genera como información.

En apariencia la utilización que hace el autor del lenguaje natural en que se presentan los documentos supone la no-reducción de elementos y por lo tanto la pérdida sería menor, pero aquí intervienen otras variables. La tecnología ha permitido la construcción y difusión masiva de documentos que pretendidamente contienen conocimientos representados por palabras que sirven para su recuperación.

La búsqueda directa por lenguaje natural como único método de recuperación es mucho más reduccionista porque la guía son los términos que se incluyen en el título o en el resumen, los cuales no siempre reflejan la temática y además pueden tener diversos significados que sólo se podrán entender mediante la lectura directa del documento y la indización por palabras del lenguaje natural, y por lo tanto resulta indicativa en términos muy generales.

La manipulación del lenguaje natural por medios automáticos en buena medida ha sido promovida por empresas que venden productos bibliográficos porque la operación de construcción y mantenimiento de un tesoro, así como el indizado y la

31 Ricki Goldman-Segall. “ La desconstrucción del mito de Humpty Dumpty: juntar las piezas para crear significado cultural” 1997 p. 53 - 54.

32 “Document indexing and profiling”. 1995 p. 58.

recuperación mediante el uso del tesoro, son de un alto costo en investigación y desarrollo y no convienen a los intereses de una empresa comercial. En este caso el costo lo absorbe el usuario, quien debe discriminar de entre una cantidad insólita de documentos aquellos que pueden ser útiles para satisfacer su necesidad de información, e invertir tiempo y recursos para ello.

En consecuencia los sectores bibliotecarios están comprometidos y tienen mucho que aportar para incrementar la investigación relacionada con la construcción de tesoros orientados a recuperar documentos en el medio digital, a partir de la experiencia acumulada a través de las últimas tres décadas;

La inversión en documentos digitales es muy alta, y organizarlos para su recuperación posterior constituye una necesidad que a la larga reeditará un mayor aprovechamiento de la información. La indización temática por medio de descriptores es un filtro que asegura el acceso a la información relevante asentada en documentos.

Hay que tener presente que la agrupación de una clasificación actúa como un índice directo para un conjunto informacional cada vez mayor y que esa clasificación temática constituye también una parte del lenguaje natural que incluye las siguientes propiedades:³³

- ❖ Al igual que todos los lenguajes son un sistema
- ❖ Al igual que todos los lenguajes son un sistema de signos
- ❖ Al igual que todos los lenguajes son un sistema de signos cuya estructura se expresa en niveles jerárquicos y relaciones

La clasificación temática asentada en un tesoro ofrece una organización jerárquica y asociada que permite recuperaciones en el plano conceptual a través de las relaciones generales o particulares o del plano horizontal para las asociaciones. Sin duda esta clasificación temática es válida para los sistemas de información en ambiente digital porque les permite representar las temáticas de los documentos y formar una red de significados entre los documentos incluidos.

A su vez el documento digital incluye las siguientes propiedades que lo relacionan con la estructura del tesoro:

- ❖ Es un subsistema en el marco de un sistema de información
- ❖ Como todos los sistemas sus elementos pueden ser representados por signos
- ❖ La estructura del sistema presenta tanto el nivel jerárquico como el de relaciones horizontales

La organización temática del tesoro lo vuelve útil no sólo para indizar sino también para ser utilizado como vocabulario por los autores para precisar el lenguaje del documento digital cuando éste representa salidas con íconos hacia otros documentos digitales. El costo de construcción de un tesoro podría entonces ser compartido en el ámbito digital académico y habría un mayor beneficio para la organización de la información contenida en ellos, que es de un alto valor económico.

33 Dolores Vizcaya Alonso. *Lenguajes documentarios*. Rosario, 1997. Nuevo Paradigma p.90

REFERENCIAS BIBLIOGRÁFICAS

- BASILI, Carla. "Subject searching for information : what does it mean in today's Internet environment?". – p. 459 – 466. – En *The Electronic Library*. – Vol. 13, no. 5, (October 1995)
- BEAN, Carol A. y Rebecca Green. "Relevance relationships". – p. 115 – 132. – En *Relationships in the organization of knowledge* / edited by Carol A. Bean and Rebecca Green. – Dordrecht: Kluwer, 2001. – (Information science and knowledge management; vol. 2)
- BEARDON, Colin y Suzette Worden "El conservador virtual: las tecnologías multimedia y el papel de los museos". – p. 93 – 117. – En *Medios contextuales en la práctica cultural: la construcción social del conocimiento* / Edward Barret y Marie Redmond, compiladores; traducción de Laura Trafi. – Barcelona: Paidós, 1997.
- BEGHTOL, Clare. "Relationships in classificatory structure and meaning". – p. 99 – 114. – En *Relationships in the organization of knowledge* / edited by Carol A. Bean and Rebecca Green. – Dordrecht: Kluwer, 2001. – (Information science and knowledge management; vol. 2)
- BERGER, F. C. y P. Van Bommel. "Augmenting a characterization network with semantic information". – p. 453 - 479. – En *Information Processing & Management*. – Vol. 33, no. 4 (July, 1997)
- BUCKLAND, Michael K. "Information as thing". – p. 351 – 360. – En *Journal of the American Society for Information Science*. – Vol. 42, no. 5 (1991)
- . "Information retrieval of more than text". – p. 586 – 588. – En *Journal of the American Society for Information Science*. – Vol. 42, no. 8 (1991)
- BURKE, Mary A. *Organization of multimedia resources: principles and practice of information retrieval*. – Hampshire: Gower, 1999. – xii, 224 p.
- CHEN, Hsin-Liang. "An analysis of image queries in the field of Art History". – p. 260 – 273. – En *Journal of the American Society for Information Science and Technology*. – Vol. 52, no. 3, 2001.
- y Edie M. Rasmussen. "Intellectual access to images" – p. 291-302. – En *Library Trends*. – Vol. 48, no. 2, Fall 1999.
- DEBRAY, Régis. *Vida y muerte de la imagen: historia de la mirada en Occidente* / traducción de Ramón Hervás. – Barcelona: Paidós, 1994. – 317 p. – (Paidós; 58)
- DOCTOROW, Cory. *Metacrap: putting the torch to seven straw-men of the meta-utopía: schemas aren't neutral*. – En línea, versión 1.3. – Disponible en <http://www.well.com/~doctorow/metacrap.htm> – Consultado el 28 de agosto de 2001

“Document indexing and profiling”. – p. 58-66. – En *Document management and imaging products and services* / comp. Roger Broadhurst. – Hertfordshire: Cimtech Limited, 1995

ENSER, P. G. B. “Progress in documentation pictorial information retrieval”. – p. 126 – 170. – En *Journal of Documentation*. – Vol. 51, no. 2, June 1995.

GOLDMAN-SEGALL, Ricki. “ La desconstrucción del mito de Humpty Dumpty: juntar las piezas para crear significado cultural”. – p. 51 – 79. – En *Medios contextuales en la práctica cultural: la construcción social del conocimiento* / Edward Barret y Marie Redmond, compiladores; traducción de Laura Trafi. – Barcelona: Paidós, 1997.

GRUPO DE TERMINOLOGÍA INFOLAC/SNI (COLOMBIA) *Guía metodológica para la elaboración de tesauros*. – Santafé de Bogotá: ICES, 1992. – 103 p.: il.

HEBA, Gary. “Digital architectures: a rhetoric of electronic document structures”. – p. 275 – 283. – En *IEEE Transactions on professional communication*. – Vol 40, no. 4, December 1997.

IGLESIAS MATURANA, María Texia. *Análisis y recuperación de información documental*. – Santiago de Chile: Ediciones IPS, 1992. – xvi, 237 p.

LAFUENTE LÓPEZ, Ramiro. *Biblioteca digital y orden documental*. – México: UNAM.CUIB, 1999. – xxi, 100 p. – (Monografías; 27)

– – –. “Sobre el análisis y representación de documentos”. – p. 163-193. – En *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*. – Vol. 15, no. 30, enero-junio 2001.

– – – y Roberto Garduño Vera. *Lenguajes de marcado de documentos digitales de carácter bibliográfico*. – México: UNAM, CUIB, 2001. – xiii, 280 p. – (Sistematización de la información documental)

LYONS, John. *Lenguaje, significado y contexto* / traducción de Santiago Alcobá; supervisión de Fernando Huerta. – Barcelona: Paidós, 1983, 1995. – 263 p. – (Paidós comunicación; 6)

MILSTEAD, Jessica L. “Standards for relationships between subject indexing terms”. – p. 53 – 66. – En *Relationships in the organization of knowledge* / edited by Carol A. Bean and Rebecca Green. – Dordrecht: Kluwer, 2001. – (Information science and knowledge management; vol. 2)

NAUMIS PEÑA, Catalina. “Análisis de la confluencia entre término y descriptor en la elaboración de tesauros” p. 95-113. – En *Investigación Bibliotecológica: archivonomía, bibliotecología e información*. – Vol. 14, no. 29, julio/diciembre de 2000.

- NAUMIS PEÑA, Catalina. "Dinámica en la construcción de lenguajes documentales y el acceso a la información". – Vol. 1: p. 69-90. – En *La información en el inicio de la era electrónica : organización del conocimiento y sistemas de información*. – México D.F.: UNAM. CUIB, 1998
- SÁNCHEZ RODRÍGUEZ, José y Ángel Ramón Romance García. "Multimedia". – p. 201-211. – En *Nuevas tecnología de la información y de la comunicación aplicadas a la educación* / José Manuel Ríos Ariza y Manuel Cebrián de la Serna. – Málaga: Ed. Aljibe, 2000. – 242 p
- SCHULTZ, Kristine. "An overview of readers´ advisory service with evaluations of related websites". – p. 21-33. – En *Acquisitions Librarian*. – No. 23, 2000
- SIMONE, Raffaele. *La tercera fase: formas de saber que estamos perdiendo* / traducción de Susana Gómez López. – Madrid: Taurus, 2001. – 165 p.
- SLYPE, Georges Van. *Los lenguajes de indización : concepción, construcción y utilización en los sistemas documentales*. – Madrid : Fundación Germán Sánchez Ruipérez, 1991. – 200 p.
- TORRES VARGAS, Georgina Araceli. *La biblioteca virtual: ¿qué es y que promete?*. – México D.F.: UNAM.CUIB, 2000. – xii, 127 p.
- UNESCO Thesaurus = *Tesaurus de l'UNESCO = Tesoro de la UNESCO: lista estructurada de descriptores para la indización y la recuperación bibliográficas en las esferas de la educación, la ciencia, las ciencias sociales y humanas, la cultura, la comunicación y la información*. – París, Francia: UNESCO, 1995. – xxxix, 705p.
- VIZCAYA ALONSO, Dolores *Información: procesamiento de contenido*. – Rosario: Parhadigma, 1997. – 187 p

