

LA POTENCIALIDAD DE LA INDIZACIÓN TEMÁTICA JERARQUIZADA

Catalina Naumis Peña
naumis@servidor.unam.mx
Centro Universitario de Investigaciones Bibliotecológicas
Universidad Nacional Autónoma de México
Piso 12, Torre II de Humanidades, Ciudad Universitaria
Coyoacán, México D. F., México

Resumen: A través del análisis de las diferencias entre la indización y la clasificación y su aplicación en la organización de la información digital, se define una nueva relación entre las dos operaciones. Se observa además la necesidad de aplicar la indización temática jerarquizada promovida a través de diversas herramientas para organizar la información digital. Otro de los aspectos observados en el presente trabajo es la responsabilidad que debe asumir la bibliotecología para clasificar el conocimiento, porque casi no existen propuestas teóricas desde la filosofía que las hizo hasta el siglo XIX.

Introducción

El objetivo de este trabajo es destacar las posibilidades que se ofrecen en el marco de la nueva revolución de la información con la *indización temática jerarquizada* como instrumento mediador para organizar y recuperar información. Mediación porque existe mucha información, pero la recuperación de documentos relevantes y de calidad, requiere de la indización como valor agregado. Nueva revolución porque se presentan cambios en varios órdenes de las esferas sociales y personales tal como ocurrió en la revolución industrial. “La revolución de los medios informáticos afecta a todas las fases de la comunicación, y abarca la captación, la manipulación, el almacenamiento y la distribución, así como afecta también a los medios de todo tipo, ya sean textos, imágenes fijas y en movimiento, sonido o construcciones espaciales” (Manovich, 2005, p. 64)

La Bibliotecología a través de la organización documental tiene un papel trascendente en el proceso social de la revolución informática como puente entre las fuentes de conocimiento y la sociedad. En la medida que ofrezca buenas soluciones de intermediación será piedra fundamental en el acceso a la información para mejorar la calidad de vida. La difusión del conocimiento depende de una organización documental breve, exacta, clara, consistente, uniforme y concisa, usando los adjetivos que proponía Gloria Escamilla (1982, p. 46) para desarrollar los índices de las bibliografías.

La organización documental implica un conjunto de desafíos para adecuarse a los nuevos entornos tecnológicos, que suponen una respuesta oportuna a las necesidades de información de la sociedad. La información se transmite por diferentes medios de comunicación y llega a través de: radio, televisión, periódicos, libros, revistas, internet, discos, videos, i-pod, películas. Tal cantidad de información no asegura que se obtenga conocimiento porque se debe seguir un proceso para ello. La

transformación en conocimiento implica mantener una actitud analítica, sintética, crítica y reflexiva acerca de la información que se recibe y la contrastación con la experiencia anterior y las necesidades actuales. Por ello, la información que proporcione el bibliotecólogo deberá ser motivadora y disparadora de los procesos de conocimiento.

La información no es conocimiento, existe un acto de conocimiento cuando la persona adquiere capacidad de discernimiento, es decir, selecciona y discrimina la información que realmente le interesa, realiza además un ejercicio constante de análisis y síntesis, de co-construcción y reconstrucción de los datos que se tienen, para contrastar con el propio marco referencial que el individuo posee, así la información se transforma en conocimiento útil, para el propio individuo y la sociedad en la cual está inserto (Cisneros, García y Lozano, 1999).

Es decir, la puesta en marcha, desarrollo y mantenimiento de una serie de estrategias hará que la información se convierta en conocimiento. En primer lugar, el ser humano discrimina aquella información relevante y de su interés. Tras haber seleccionado la información, la analiza desde una postura reflexiva, intentando profundizar en cada uno de los elementos, deconstruyendo el mensaje, para coconstruirlo desde su propia realidad. Es decir en el proceso de deconstrucción se desmonta, comprende, entiende las variables, partes, objetivos, elementos, axiomas del mensaje. En el proceso de coconstrucción se realiza el procedimiento inverso. A partir de variables, axiomas, elementos, etc., se vuelve a componer el mensaje, desde la realidad personal, social, histórica, cultural y vital, es decir, desde la perspectiva global del conocimiento y la persona. Sólo, no perdiendo esta perspectiva se puede afrontar y enfrentar la evolución y el progreso de las nuevas tecnologías de tal forma que lleve en un futuro a crear una sociedad más justa donde lo tecnológico y lo humano se integren al igual que los distintos puntos de mira de las diferentes culturas conformando el crisol de la realidad en la que se vive.

En este proceso de conocimiento el bibliotecólogo acompañado del usuario también hará el ejercicio de discriminar la información para involucrarse en un proceso de conocimiento para representar contenidos documentales. Ese rol del bibliotecólogo es difícil y de gran responsabilidad en el mundo actual porque la mediación implica interpretación de contenidos documentales cada vez más especializados, en formatos que exigen lecturas diferentes y llevan consigo además el habitual razonamiento para representar el pensamiento expresado en el documento, es decir, es necesario entender el mensaje que se transmite y convertir la información en conocimiento.

En el nuevo marco social de la información como valor económico, la biblioteca es además de un centro de acopio documental, un centro de distribución de documentos a distancia. Los sistemas de información especializados de texto completo brindan un servicio que por lo costoso asegura cierto rigor en los documentos incluidos. En cambio, los documentos que se distribuyen por Internet no pasan por procesos de validación social como los artículos de revistas o libros indizados por sistemas de información o los sistemas de bibliotecas que operan en cierto modo como filtro de información, a través del apoyo de sus usuarios en las bibliotecas públicas o de los docentes o investigadores de las instituciones educativas, industriales o comerciales para asegurar que los documentos distribuidos mantengan cierto nivel de calidad. Si bien la Internet es dinámica en producción de

información el resultado es incontrolado y esporádico. Los recursos informativos son volátiles y los servidores de las redes que los soportan cambian constantemente.

Relación entre antecedentes y situación actual de la indización temática jerarquizada

La incorporación de la tecnología en los procesos y servicios de las bibliotecas les ha permitido avanzar y enfrentar los desafíos para satisfacer las necesidades que presentan los usuarios tanto presenciales como remotos y promover “la biblioteca sin muros”. La nueva responsabilidad es diseñar instrumentos adecuados a las nuevas tecnologías que potencien la recuperación de información especializada.

En este sentido, el desarrollo de tesauros, taxonomías y ontologías para indizar y recuperar información conformados por estructuras de conocimiento y el lenguaje especializado de diferentes áreas disciplinarias permiten trabajar en un nivel de síntesis de los contenidos documentales en los sistemas de información especializada. Desde hace varios siglos la Bibliografía y la Catalogación usaban mecanismos de ordenación temática de los documentos, jerarquizados para acomodarlos en la estantería, e indizados por la expresión en el lenguaje, de todos los temas que contuviera el documento a través del índice temático o del catálogo. Sin embargo, desde hace medio siglo aproximadamente, la Bibliografía ha tenido un cambio profundo y su conversión en bases de datos, ha llevado a la indización de los contenidos temáticos insertados en un proyecto social más amplio, que se relaciona con el análisis de comportamientos terminológicos en las diferentes lenguas y la clasificación de los temas para su ubicación en el ciberespacio.

Los conceptos usados para comunicarse en la ciencia, el arte y la tecnología, que son las áreas donde más se produce la intermediación de la Bibliotecología requiere de un mayor grado de precisión que el lenguaje natural, aún cuando se parta de la idea que es el producto del mismo proceso de pensamiento acerca de un objeto tanto concreto como abstracto. En este sentido se comparte la preocupación con la Terminología que estudia la naturaleza de las expresiones lingüísticas especializadas, sin entrar en la especulación filosófica. La Bibliotecología y la Terminología comparten la metodología terminológica, aunque la aplican de manera diferente, pero ambas estudian y explican la naturaleza y comportamiento de las expresiones lingüísticas que utiliza el conocimiento especializado en su comunicación, tanto en nuestra lengua como en otras (Lara, 2001: 180)

La Bibliotecología desde el comienzo del uso de la informática la aprovechó para mejorar sus sistemas de comunicación con el usuario. La catalogación fue una de las primeras tareas asistidas por la computadora, para representar características de forma y contenido de los documentos, compartiendo créditos con el préstamo de libros en las grandes bibliotecas. La mediación lingüística se seguía realizando a través de las listas de encabezamientos de materia, pero utilizadas y aprovechadas de manera diferente a los sistemas manuales que habían prevalecido hasta la mitad del

siglo XX, porque se combina con la indización automática de palabras en títulos y/o resúmenes. En estos primeros años también comenzaron los sistemas de recuperación de artículos de revistas en los cuales la mediación se realizaba a través de las palabras significativas de los títulos, pero la Bibliotecología continuaba haciendo propuestas innovadoras como la construcción de tesauros.

Estos son los antecedentes de la indización temática jerarquizada que es ahora el paradigma para la organización de información en la Web Semántica. Sin embargo, antes de entrar en el detalle de la indización jerarquizada es conveniente definir los diferentes instrumentos clasificatorios que se manejan en las comunicaciones especializadas

Nomenclaturas y clasificaciones

Las clasificaciones o taxonomías son sistemas que pretenden ordenar conceptos y objetos según determinados criterios, agrupándolos en categorías o clases con características comunes. Las nomenclaturas son un tipo de terminología aplicadas a cosas naturales u objetos que forman series más o menos homogéneas (animales, plantas, medicamentos, sustancias químicas, microorganismos, accidentes anatómicos, etc.) cuyas denominaciones se crean conforme a reglas uniformes (Díaz Rojo, 2001, p. 40). Es decir, de acuerdo a su origen etimológico del latín nomenclature es asignación de nombres y es una asignación de nombres estandarizados, a diferencia de la clasificación que son los principios implícitos en la organización de objetos en grupos de acuerdo a sus similitudes y diferencias o sus relaciones en un conjunto de criterios. (<http://www.search.eb.com/eb/article-9024237>).

Las nomenclaturas más conocidas están relacionadas con las ciencias médicas y son: la botánica y zoológica de Linneo (*Species plantarum*, 1753; *Sistema naturae*, 1758); de química de Guyton de Morveau y Lavoisier (*Méthode de nomenclaturas chimique*, 1787); la anatómica, aceptada en 1895, por medio de la *Nomina anatómica* de Basilea; la clasificación de enfermedades, propuesta por el médico Francois Sauvages; la bioquímica, la bacteriología; la farmacología; los procedimientos médicos, etc. Estas nomenclaturas han sido aprobadas por autoridades científicas para lograr la uniformidad de conceptos y lograr una comunicación aceptable en el proceso de la producción de conocimiento médico.

Con respecto a la clasificación una de las primeras propuestas la realizó Aristóteles en el *Organon*, a través de la definición de reglas para crear categorías. Las categorías generales que propuso fueron: sustancia, cantidad, cualidad, relación, lugar, tiempo, situación, posesión, acción y pasión (Broughton, 2004, p. 14). A partir de esta definición de categorías, Porfirio propuso disponerlas en forma de árbol y describe cómo las cualidades atribuidas a las cosas pueden ser clasificadas, rompiendo con el concepto filosófico de sustancia como una relación de genero/especie. En los libros de texto medievales el *Arbor porphyriana* (Árbol de Porfirio) ilustra su clasificación lógica de la sustancia. Para Porfirio los conceptos se subordinan partiendo de los más generales a los más simples. El *Arbor porphyrian* inició el Nominalismo que se podría ver como una especie de antecesor de las modernas clasificaciones taxonómicas. Un ejemplo del *Arbor porphyrian* sería este:

- Substancia - Puede ser corporal o incorporeal
 - Cuerpo- Puede ser animado o inanimado
 - Viviente - Puede ser sensible o insensible
 - Animal - Puede ser racional o irracional
 - Racional – El hombre

En este despliegue ya existe un orden jerárquico siguiendo sólo una de las dos líneas entre las que se puede optar desde la categoría principal, a las clases y subclases que continúan los niveles jerárquicos. En 1605, Francis Bacon publica “The Advancement of Learning” donde introduce su clasificación que luego fue usada por los enciclopedistas franceses y desarrollada posteriormente por Melvin Dewey en 1876 como base de los esquemas de clasificación de bibliotecas. Otros filósofos acometieron el estudio de la clasificación de las ciencias como Charles Sanders Peirce a principios del siglo XX, pero como bien decía Hjørland en 2005 ([http://en.wikipedia.org/wiki/Classification_of_the_sciences_\(Peirce\)](http://en.wikipedia.org/wiki/Classification_of_the_sciences_(Peirce))), no existía ningún programa de investigación de clasificación de las ciencias. Hoy sólo se puede mencionar el estudio que realiza Chaim Zins en Israel (<http://hw.haifa.ac.il/human/hebrew/ISMapWeb.htm>)

Las nomenclaturas y las clasificaciones tienen un origen común, pero se prefiere la primer denominación para la asignación de nombres a elementos clasificatorios normalizados por una ciencia y las clasificaciones para una organización más general, porque establecen la división en categorías en cualquier área. A pesar de la necesidad de clasificar las ciencias, en el entorno actual, no existen esfuerzos más que aislados por hacer propuestas teóricas sobre el problema.

Indización y clasificación

En la Bibliotecología tradicional, en las primeras seis decenas del siglo XX, la indización y la clasificación son conceptos relacionados entre sí y son además actividades complementarias entre ellas, una clasificación temática implica una sistematización conceptual del conocimiento de una especialidad en base a un esquema clasificatorio para ordenar documentos en un lugar físico y la indización es la asignación de todos los temas que contiene un documento para recuperarlo en un sistema de información que contiene datos sobre los documentos. Al comenzar la utilización de la computadora para transmitir información, en los últimos cuarenta años del siglo XX y en los primeros años del presente siglo, los conceptos de indización y clasificación comienzan a entremezclarse.

En el lenguaje natural la indización es el registro ordenado de datos e información para elaborar un índice y así aparece en el *Diccionario de la Lengua Española*. En el diccionario especializado de Martínez de Sousa se incluyen tres acepciones: 1) acción y efecto de indizar, 2) confección de índices y 3) extracción de conceptos del texto de un documento para expresarlos con la ayuda de un lenguaje, como palabras clave, descriptores o índices de una clasificación. La tercera y

última es la connotación especial que se le otorga en el campo de la Bibliotecología, que muy frecuentemente se designa también como indización temática, término compuesto que exhibe menor ambigüedad para un tipo de indización que se refiere únicamente al contenido semántico.

Langridge (1992: 63) explica como el término “indización” se usa de tres modos diferentes: 1) sinónimo de organización del conocimiento en las bibliotecas e incluye tanto el índice de autor, título, descripción del documento y todos los temas tratados en el documento 2) acto de registro del contenido de una colección (en contraste con el acto de búsqueda de la información requerida) 3) en sentido restringido como llave alfabética de un orden sistemático, el índice alfabético para el contenido de libros arreglados temáticamente o el índice alfabético a un catálogo arreglado sistemáticamente. En el segundo apartado se debería agregar la ordenación de las palabras significativas del texto completo de un documento, que en realidad forman parte de las tareas que realizan las máquinas automáticamente.

La tercera acepción de Landgrigde sin el calificativo de sentido restringido, sino como representación del concepto indización aparece con frecuencia ligado a la indización de materias, como se puede percibir en autores de la categoría de Vizcaya Alonso (1997: 152) “se puede considerar por tanto, la indización como un proceso de descripción más exhaustivo y profundo que el proceso de clasificación. Ello implica la necesidad de prestar atención a todos los elementos presentes en el contenido semántico del documento que puedan ser de interés a los destinatarios finales del servicio y no sólo al contenido central de la fuente procesada como ocurría en el caso de la clasificación”.

Cleveland & Cleveland (1990: 17) definen un índice como una lista de información bibliográfica o citas hacia un cuerpo literario, usualmente arreglados en orden alfabético y basado en algunos datos específicos, tales como autor, tema o palabras claves. El mismo autor cuando se refiere a la asignación de términos para representar el contenido temático lo adjetiva para acotar el ámbito de acción y usa la expresión “índice temático”.

En la acción de indizar por temas se distingue la indización humana y la automática, la primera trabaja en base a los conceptos tratados en un documento y la segunda en base a los términos que aparecen mencionados. La indización humana reconoce los elementos constitutivos del documento a través del análisis de su lectura y garantiza que si un concepto tratado en el mismo, no aparece mencionado o se usa un término no habitual, el analista lo incluirá con el término en uso en el sistema, para recuperar el concepto. En la indización automática, la máquina separa cadenas de caracteres ya sea en el título, en el resumen, descartando únicamente las llamadas palabras vacías o reconociendo en el texto completo los sintagmas nominales (Amar, 2000: 19). En medio de estos dos tipos de indización existen clasificaciones intermedias, como la asignación humana de palabras significativas en los textos, sin someterlas al proceso de control que supone la utilización de una herramienta de normalización del vocabulario o la asignación automática de palabras previamente definidas.

La indización es un proceso de ordenación insertado para buscar y recuperar información en los sistemas. De este modo, los usos del término “indización” detallados por Landgridge se pueden resumir en que es uno de los elementos del proceso general al que se someten los conocimientos y datos asentados en los documentos para su comunicación al medio social, sin tener que revisar cada uno de los documentos contenidos en el sistema de información. La indización temática tal como lo expresa Cleveland & Cleveland es el análisis de contenido y la traducción de los términos que sintetizan la temática del documento a términos del lenguaje del sistema de información en el cual serán insertados los documentos. La indización automática incluye no sólo la temática sin intervención humana, sino además la organización de los índices de datos por puntos de acceso que contiene cada uno de los registros bibliográficos del sistema o palabras significativas del texto completo de los documentos.

Sin entrar al aspecto filosófico desde el punto de vista de su tratamiento terminológico, la clasificación aparece para algunos filósofos como Bronowsky (1993: 11) en el propio uso del lenguaje humano, quien lo define como un elemento para clasificar el mundo “La naturaleza y el poder del lenguaje humano como instrumento creativo para ordenar y estructurar la experiencia humana”

La primera acepción de clasificar que incluye el diccionario de Martínez De Sousa es la acción y efecto de clasificar y de ahí continúan las diferentes interpretaciones que se le asignan al término en Bibliotecología y Archivística en los cuales está presente siempre el ordenamiento y la estructuración que significa la acción de clasificar, aunque también aparece en este inventario lexicográfico especializado, una designación para el repertorio que se utiliza para clasificar y una notación con códigos alfabéticos, numéricos o alfanuméricos que guían e indican la posición relativa que ocupa un código con respecto a otros.

En la literatura bibliotecológica la clasificación está ligada a la descripción de contenido como forma de agrupación del conocimiento de acuerdo a la temática implícita en cada uno de los documentos que integran un sistema de información, pero los propios expertos de la disciplina insertan la acción en un ámbito más global. “Clasificar, en términos generales, -afirma Chan (Citado por Gil Urdiciain, 1981: 209) - es el acto de organizar el universo del conocimiento en algún orden sistemático. Ha sido considerada la actividad más fundamental de la mente humana. El acto de clasificar consiste en el dicotómico proceso de distinguir cosas u objetos que poseen cierta característica de aquellos que no la tienen, y agrupar en una clase cosas u objetos que tienen la propiedad o característica en común”.

Chan habla de universo de conocimientos, pero no ligado a un esquema clasificatorio que abarca todo, en cambio Kedrov (1974: 7) se refiere a la aplicación de esquemas clasificatorios de tipo general que son los utilizados para la clasificación bibliográfica en una biblioteca: “la unificación de todos los conocimientos en un sistema único, en el cual se reflejan la lógica del objeto de estudio y las concepciones generales sobre el mundo y su conocimiento por el hombre”.

Como se observa las opiniones coinciden al relacionar la clasificación con el conocimiento que está en el contenido de documentos publicados, en ello está implícito además que han pasado

por controles sociales que les otorgan valor para ser comunicados a su medio como conocimiento. “La actividad ordenadora de documentos cae dentro del terreno del conocimiento, puesto que al tratar de hacer inteligible un orden documental, se requiere de la interpretación, la exégesis, la relación y la conceptualización, en un esfuerzo por establecer relaciones o conexiones especialmente pertinentes entre los documentos, los datos, la información en una forma coherente, y poder explicar de que trata un orden documental” (Lafuente, 1993: 11)

En el mismo sentido, Landgridge (1992: 70) opina que la clasificación forma parte de un proceso de comunicación porque la finalidad de la misma es la comunicación de información entre el emisor y el receptor de un conocimiento, que es un proceso diferente a la simple transferencia de información. El objetivo de una clasificación es la organización del conocimiento de una especialidad o conjunto de especialidades y la asignación de un lugar en esa estructura a cada uno de los documentos que ingresan a un sistema de información. Esta concepción supone el manejo de una estructura de conocimiento como base de la clasificación documental en el área cubierta por el sistema de información y la capacidad de relacionar cada una de las temáticas de los documentos dentro de esa estructura.

Cleveland & Cleveland (1990: 17) manifiestan que para identificar entre millones de documentos los adecuados a nuestra necesidad de información se deben organizar de modo tal que muestren las relaciones existentes entre ellos y este será el objetivo tanto de la clasificación bibliográfica como la indización temática, en su calidad de conceptos similares. Agregan además literalmente “En efecto, la indización es un acto de la clasificación”

Entre los principios de la indización Lancaster (1996: 15) estudia en un apartado “la indización como clasificación” para explicar como en Bibliotecología el término “clasificar” es utilizado en la asignación de números de clasificación extraídos de esquemas especializados en ello, para ordenar los libros o documentos en estantes o archivos. Sin embargo, para este mismo autor también se clasifica cuando se asigna un término, que dentro de un índice de materias adopta la forma de un esquema de clasificación. Es decir, la existencia de un esquema clasificatorio como base de la asignación de temas constituye un acto de clasificación para otorgarle un lugar al documento en el índice temático del sistema de información.

La aseveración de que la clasificación no es una indización aparece en la obra de Collard (1995: 223) porque distingue entre la clasificación, donde el documento puede tener cabida en diferentes lugares, pero se tiene que escoger una, al contrario de la indización que tiene la función de paliar las deficiencias de la clasificación, ya que se le pueden asignar varios términos de indización a un documento para su recuperación. Esta distinción no parece ser válida, ya que está basando la diferencia entre clasificación e indización en la selección del tema o lugar que se le asigna a un documento y no existe una objetivación de la sustancia verdadera de la operación de clasificación e indización, sino una referencia al modo de desarrollar la actividad y el uso del instrumento codificado utilizado. Además la posibilidad de asignar una o más posiciones a una clasificación bibliográfica depende también de la clasificación utilizada, ya que por ejemplo la Clasificación Decimal Universal es

una clasificación flexible que permite la ordenación de cada uno de los elementos temáticos incluidos en el documento.

Lancaster (1996: 16) adjudica la confusión terminológica que existe en torno a las operaciones de indizar y clasificar, a la falta de distinción de las etapas de la indización. El análisis de contenido que es previo a la indización implica que un documento es colocado en una clase conceptual y conlleva la preparación de una representación del contenido temático de un documento. En este mismo sentido Esteban Navarro (1996: 28) insiste que “si se sigue la perspectiva de atender al carácter del proceso intelectual que les sirve de fundamento – es decir, las operaciones mentales que se producen y los recursos cognitivos que se ponen en acción durante su ejercicio-, se descubre que son actividades complementarias que aparecen en todo proceso de representación y organización documental, con independencia del instrumento empleado”

Aunque Esteban Navarro plantea la indización y clasificación como productos de operaciones complementarias, separa una actividad de la otra, a diferencia de Lancaster. Esta complementariedad que le adjudica Esteban Navarro se debe al hecho que el mismo plantea el lenguaje controlado sin una estructura de clases, a diferencia de las clasificaciones bibliográficas y se maneja con la concepción tradicional de clasificación e indización: con la clasificación se sitúa un documento en el nivel más sintético posible y la indización analítica (utilización del término en el mismo sentido de indización temática) se mueve en una franja más ancha en cuanto a límites de asignación de contenido temático, pero sin integrar el resultado en una estructura de clases.

Sin embargo para Lancaster la asignación de una temática a un documento, aún en el caso de que no incluya una estructura de clases significa una operación de clasificación porque escoge entre una variedad de posibilidades y en palabras textuales agrega “Una mayor confusión ocurre cuando uno comprende que la indización de materias puede incluir el empleo de un esquema de clasificación o que un índice de materias en forma impresa puede adoptar la secuencia de algún esquema de clasificación” (Lancaster, 1996: 16) Es decir, que la indización por materias puede o no estar basada en un sistema clasificatorio, pero cuando lo está el grado de confusión entre indización y clasificación aumenta.

Los esquemas de clasificación bibliográficos para el ordenamiento documental en la estantería brindan servicios que coinciden con la indización temática que se pretende realizar con los lenguajes controlados basados en un esquema clasificatorio. Los esquemas de clasificación bibliográficos numérico o numérico alfabético tienen tablas sistemáticas que establecen la equivalencia de estos códigos numéricos o alfabéticos al término de indización que representan. Lafuente (1993: 83) expresa y resume este pensamiento dominante, acerca de la clasificación... “La representación de contenidos documentales, en el ámbito bibliotecológico, tiene la intención de poner de manifiesto un orden determinado ante la presencia del público. Esta representación se expone a través de palabras o símbolos, y tiene como objetivo sustituir un objeto (libros, materiales hemerográficos, videos, etcétera), o bien, los datos provenientes que se dan como producto de la descripción de los documentos y sus contenidos (resúmenes, fechas, editores, etcétera). Al suplir un objeto por un

símbolo construido específicamente para figurar en lugar de éste, se pretende que represente a un objeto en particular, pero además que el símbolo tenga características que hagan factible asociar un símbolo con otro para formar un orden, que puede ser lógico, alfabético o de cualquier otro tipo, dependiendo de los propósitos que se persigan con la organización documental”

En el proceso de clasificación bibliográfica cuando se asigna un lugar a un libro en la estantería se está creando un índice de las temáticas que se incluyen en la colección. En el proceso comúnmente denominado indización temática se le asignan a veces varios lugares en el índice de un fondo documental para cada pieza documental, en cualquiera de los dos casos se crea un índice de los temas tratados en los documentos incluidos en el sistema de información. En los dos casos se puede acceder a los documentos que contienen una temática de interés: en el primero el tratamiento de la temática comprende todo el documento y en el segundo puede ser que se localice menos información sobre el mismo tema en cada documento por separado, pero uno y otro proceso están asegurando que el tema asignado está reflejado para la recuperación.

Otra de las diferencias es en la recuperación, ya que la clasificación bibliográfica mediante el recorrido por la estantería se accede al tema, en la indización temática de los sistemas de información bibliográfica, también se accede a los documentos digitales que los contienen de forma bastante inmediata y cada vez es mayor la rapidez de obtención de cualquier documento en cualquier parte del mundo. La designación de un proceso como clasificación bibliográfica y al otro como indización temática, no los convierten en conceptos diferentes, lo cierto es que tanto los esquemas de clasificación bibliográfica, como los tesauros están basados en una estructura del conocimiento y los dos por lo tanto deben ser considerados sistemas de clasificación. Es frecuente que los tesauros incluyan códigos numéricos o alfa-numéricos para representar las clases principales, los esquemas de clasificación bibliográfica mantienen índices con los temas representados en los esquemas. El resultado de la aplicación tanto del esquema de clasificación bibliográfica, como del tesoro aseguran que exista un índice de las temáticas contenidas en el fondo documental.

El tesoro es un instrumento que permite la indización temática jerarquizada para adjudicar los temas a los documentos y recuperarlos desde lo general a lo específico o al revés. Los sistemas de información y la búsqueda de transmitir un significado conceptual de los documentos contenidos en los mismos y una inserción en el área de conocimiento correspondiente hace de la indización temática jerarquizada una herramienta imprescindible que conjunta los servicios que brindan la clasificación y la indización de la clasificación tradicional a la organización en el medio digital. En el próximo apartado se abordará el estudio y comparación del tesoro con otros instrumentos que se han estado desarrollando en la Web para clasificar los documentos digitales y recuperarlos por sus contenidos temáticos, pero siempre pensando en una clasificación que relacione los contenidos en un marco más amplio.

Tesauros, taxonomías y ontologías

La indización puede ser realizada con el modelo de los tesauros, las taxonomías y las ontologías dependiendo del ámbito de conocimiento a representar desde el punto de vista temático. La diferencia del tesoro con la taxonomía y la ontología es la estructura que se usa en uno y otro.

Los tres están basados en mapas de conocimientos, el tesoro está enfocado a los temas de los contenidos documentales, la taxonomía a la estructura de una empresa, industria o institución y la ontología cumple funciones de diccionario para etiquetar con precisión el lenguaje de la Web, establecer sus propiedades y sus relaciones con otros conceptos y transformarlos en términos para que cuando un software de búsqueda encuentre una palabra pueda interpretar su significado (Moreira González, 2004, p. 215).

Estas herramientas lingüísticas han despertado un interés creciente porque apoyan la organización sistemática de la información mediante estructuras categorizadas de conocimientos y recopilan términos representativos de ámbitos especializados (Gilchrist y Kivi, 2000; Hill y Koch, 2001; Hodge, 2000; Taylor, 2004; Tudhope y Koch, 2004; Williamson y Beghtol, 2003), (Roe y Thomas, 2004), citados por Caminotti y Martínez (2006, p. 75).

Una buena organización del recurso bibliográfico digital es instrumentada por los metadatos que agregan características genéricas y abstractas para recuperar los contenidos documentales. Es decir los metadatos se agregan al contenido del documento sin formar parte del documento pero permiten manipularlo y relacionarlo con los otros documentos del sistema. Entre los diferentes metadatos que deben marcarse en un documento, se destacan los que se refieren al contenido temático que es el aspecto tratado en este trabajo. Los metadatos permitirán relacionar los documentos que contengan los mismos temas en forma inmediata, por lo cual los temas comunes a los documentos deben estar definidos de antemano para recuperarlos.

Los metadatos han coadyuvado en la tarea de organizar la información en la Web, a través de descripciones genéricas y la utilización de lenguajes apropiados para que las computadoras puedan indizar los documentos. Sin embargo, las etiquetas que se generan no proporcionan mayor descripción de los elementos mismos, más allá de datos que la computadora puede procesar pero no puede comprender. Con el uso de los metadatos y de los tesauros se pueden obtener jerarquías temáticas de los contenidos pero no es posible, de forma automatizada, que las computadoras puedan crear relaciones entre contenidos, así como tampoco se puede delimitar el dominio de conocimiento que comprenden. Por esto se utilizan las ontologías, puesto que pueden basarse en la estructura de un tesoro, pero forman su propia red entre los documentos en el medio digital, para relacionar contenidos y recuperarlos. Para lograr que los contenidos tengan un significado explícito para las máquinas, la Web Semántica utiliza XML como lenguaje para proveer una sintaxis; el RDF (Resource Description Framework) para modelar las relaciones; y las ontologías para las descripciones y los vocabularios.

Dicen los informáticos que por arriba de los tesauros y de los metadatos, las ontologías garantizan consistencia en dominios de conocimiento, hacen explícitos vocabularios, pero principalmente lo que aportan son las reglas que permiten a las máquinas interpretar las relaciones

entre contenidos. Los tesauros y los metadatos estructuran contenidos, las ontologías estructuran la semántica de los contenidos (López Guzmán y García Peñalvo, 2007, p. 5) Esto parece un juego de palabras, pero hay que entenderlo en el marco de la arquitectura documental para recuperar información en el medio digital. Sin embargo se pueden construir ontologías categorizadas a partir de los términos establecidos en un tesoro.

La aparición de las herramientas comentadas aquí y otras más que se producen para organizar información digital parten de la importancia de contextualizar el lenguaje a través de relaciones de términos que suplanten la ausencia de la articulación del lenguaje natural y hacer más expedita la recuperación de información.

Consideraciones finales

Los recursos bibliográficos digitales son ceros y unos que exigen asegurar la recuperación del documento original en su formato y con sus propias características a través de elementos que notifiquen al usuario de su contenido. La sociedad exige con mayor fuerza la organización de los documentos digitales y esto es algo que el bibliotecólogo deberá hacer.

El punto de acceso temático a la información siempre ha estado presente en los procesos bibliotecológicos para recuperar los contenidos documentales. El acceso temático al documento digital se plantea relacionando los diferentes aspectos que lo componen, con un ámbito más genérico o más específico para contextualizar la información que ofrecen.

El significado es un fenómeno mental que remite a una capacidad cognitiva humana, adquirida genéticamente por nuestra especie a lo largo de millones de años de selección evolutiva. Por lo tanto, la semántica y el acceso al significado tienen una base biológica ligada a la actividad intelectual de un organismo vivo que los procesos automatizados aún no han podido resolver.

Existe una demanda creciente de programas informáticos que simulen la inteligencia de un ser humano competente en la capacidad de generar mapas conceptuales jerarquizados de la información relevante contenida en un cuerpo de conocimiento determinado.

Los programas disponibles en el mercado que resumen textos de manera automática suelen operar aplicando índices de frecuencias estadísticas de aparición de palabras (que, por otra parte, es la tarea que, hoy por hoy, los ordenadores realizan más rápidamente y mejor). Para que la computadora arroje datos significativos en este nivel de funcionamiento, los programas excluyen del recuento las "palabras vacías", invariables y sin contenido proposicional inventariable léxico-semánticamente (artículos, pronombres, conjunciones, preposiciones, marcadores, conectores, etc.), ya que, se supone (en nuestro planteamiento, creemos que erróneamente) que no aportan información demasiado interesante al texto. Cuando se pone en marcha, el programa así diseñado identifica las cadenas de caracteres que más se repiten como las palabras claves que supuestamente representan las ideas principales del texto. Las palabras sin las uniones entre ellas, muchas veces cambian el sentido de los contenidos documentales.

Si bien se trata de un planteamiento en apariencia útil, este tipo de diseño condena al programa a un funcionamiento en muchas ocasiones extraño, porque los textos bien contruidos poseen un elevado grado de variación estilística, sinonimia y elisión, y un bajo nivel de repetición. Estas características lingüísticas elusivas del tratamiento estadístico provocan que el programa desatine continuamente sus decisiones de sumarización. (Bonilla, 2007)

La organización de la información de los contenidos documentales digitales se propone a través de jerarquías y relaciones para contextualizar los temas en ámbitos de conocimiento para evitar la polisemia y la sinonimia de los términos aislados.

Las relaciones disciplinares para organizar información digital se realizan con base en la práctica y el uso social y no existen estudios teóricos que avalen una jerarquización de las ciencias.

La asignación de temas en el documento digital experimenta una simbiosis entre la clasificación bibliográfica y la indización temática de la bibliotecología tradicional.

OBRAS CONSULTADAS

- Amar, M. (2000) *Les fondements théoriques de l'indexation: une approche linguistique*. París: Association des professionnels de l'information et de la documentation. (Collection Sciences de l'information, série Recherches et documents)
- Bonilla, S. (2007). Web Semántica y Agentes Metarrepresentacionales basados en Marcadores Discursivos [on line]. "*Hipertext.net*", (5) <<http://www.hipertext.net>> [Consulta: 26/06/2007]. ISSN 1695-5498
- Bronowski, J. (1993) *Los orígenes del conocimiento y la imaginación*. / traducción de E. Lynch. 2ª ed. Barcelona: Editorial Gedisa. (Colección hombre y sociedad: Filosofía)
- Broughton, V. (2004) *Esencial classification*. New York: Neal-Schuman.
- Caminotti, M. L. & Martínez, A. M., (2006) Fútbol, tesauros y taxonomías WEB: desafíos del control del vocabulario. *Información, Cultura y Sociedad*, (14), pp. 73-81
- Chan, L. M. & Zeng, M. L., (2002) Ensuring interoperability among Subject Vocabularies and Knowledge Organization Schemes: a methodological analysis, En IFLA COUNCIL AND GENERAL CONFERENCE (68: agosto 18-24, 2002: Glasgow),
- Cisneros Rodríguez, I., García Doctor, C. & Lozano Jurado, I. M., (2000) *¿Sociedad de la información vs Sociedad del conocimiento?: La educación como mediadora*. Recuperado Agosto 30, 2007 de <http://tecnologiaedu.us.es/edutec/paginas/43.html>
- Classification of the sciences (Pierce)*. Recuperado en Noviembre 5, de 2007 en [http://en.wikipedia.org/wiki/Classification_of_the_sciences_\(Pierce\)](http://en.wikipedia.org/wiki/Classification_of_the_sciences_(Pierce))
- Cleveland, D. y Cleveland A. D. (1990) *Introduction to indexing and abstracting*. 2nd ed. Englewood, Col.: Libraries Unlimited.

- Collard, C., Giannattasio, I. y Melot, M. (1995) *Les images dans les bibliothèques*. París: Éditions du Cercle de la Librarie. (Collection Bibliothèques)
- Díaz Rojo, J. A. (2001), La terminología médica: diversidad, norma y uso. *Panacea*, 2 (4), (pp. 40-46)
- Encyclopaedia Británica (2006) Nomenclatura. Recuperado en Noviembre 5, de 2007 de <http://www.search.eb.com/eb/article-9024237>
- Esteban Navarro, M. A. (1999) Fundamentos epistemológicos de la clasificación documental. pp. 19-32. En *Manual de clasificación documental*. Madrid: Editorial Síntesis. (Biblioteconomía y documentación)
- Escamilla González, G., (1982) *Manual de metodología y técnicas bibliográficas*. México: UNAM, Instituto de Investigaciones Bibliográficas.
- Gil Urdiciain, B. (1996) *Manual de lenguajes documentales*. Madrid: Editorial Noesis.
- Kedrov, B. M. (1974). *Clasificación de las ciencias*. Moscú: Editorial Progreso. 2 v.
- Lafuente López, R. (1993) *Los sistemas bibliotecológicos de clasificación*. México: UNAM. CUIB. (Serie monográfica; 15)
- Lancaster, F. W. (1996) *Indización y resúmenes: teoría y práctica / traducción de E. Barber*. Buenos Aires: EB Publicaciones.
- Langridge, D. W. (1992) *Classification: its kinds, systems and application*. Londres: Bowker Saur. (Topics in library and information studies)
- Lara, L. F. (2001) *Ensayos de teoría semántica: lengua natural y lenguajes científicos*. México: El Colegio de México. (Jornadas; 135).
- López Guzmán, C. & García Peñalvo, F. (2007, agosto) Ontologías para relacionar contenidos educativos digitales. Ponencia presentada en el I Seminario de Organización del Conocimiento: Bibliotecología y Terminología, celebrado en México D. F., México.
- Manovich, L. (2005) *El lenguaje de los nuevos medios de comunicación: la imagen en la era digital*. Barcelona: Paidós. (Paidós Comunicación; 163)
- Martínez de Sousa, J. (1993) *Diccionario de bibliología y ciencias afines*. 2ª ed. aum. y act. Madrid: Fundación Germán Sánchez Ruipérez. (Biblioteca del Libro).
- Moreiro González, J. A. (2004) *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*, Ediciones Trea, Gijón, (Biblioteconomía y administración cultural; 97)
- Nomenclatura (2007) In *Encyclopaedia Británica Online*. Recuperada en Noviembre 5, de 2007 de <http://www.search.eb.com/eb/article-9024237>
- Vizcaya Alonso, D. (1997). Información: procesamiento de contenido. Rosario: Paradigma Ediciones.
- Zins, Chaim (2007) *Knowledge map of information science: issues, principles, implications*. Recuperada en Noviembre 5, de 2007 de <http://hw.haifa.ac.il/human/hebrew/ISMapWeb.htm>