

# Indización y clasificación: Un problema conceptual y terminológico

## *Indexation and classification: A conceptual and terminologic problem*

Catalina NAUMIS PEÑA

Centro Universitario de Investigaciones Bibliotecológicas  
Universidad Nacional Autónoma de México  
Piso 12, Torre II de Humanidades, Cd. Universitaria  
CP 04510, México D. F.  
Correo electrónico- naumis@servidor.unam.mx

### RESUMEN

Es tarea propia de la Bibliotecología fijar el uso de sus conceptos y términos para la comunicación científica interna y la difusión de sus principios hacia el medio social, por lo tanto debido a la existencia de diferentes conceptos acerca de tesaurus y su relación con la clasificación e indización, se realiza un análisis desde una perspectiva terminológica. Se trata de probar además que una clasificación temática implica una sistematización conceptual del conocimiento de una especialidad y que un tesaurus es un sistema clasificatorio puesto que contiene descriptores contextualizados por categorías semánticas, aclaraciones de las alternancias en los usos lingüísticos y el reconocimiento de propiedades comunes entre ellos.

### PALABRAS CLAVE

Organización del conocimiento tesaurus documental indización Temática Clasificación

### ABSTRACT

It is the proper work of the Library Science to specify its concepts and terms to the internal scientific communication and the diffusion of its principles to the social media, therefore due to different concept about thesaurus and its relation with classification and indexing it is necessary an analysis from the terminology perspective. It deals with that subject indexing involves a knowledge conceptual systematization of a speciality and that a thesaurus is a classification system because it content contextualized descriptors in semantic classes, expound of the alternation to the different linguistic uses and the recognition of common properties between them.

**SUMARIO** 1. El lenguaje de la bibliotecología. 2. Indización en Bibliotecología. 3. La clasificación temática como un sistema de organización del conocimiento. 4 Tesaurus documental como clasificación temática. 5. Consideraciones finales. 6. Referencias bibliográficas.

En la literatura bibliotecológica existe una controversia acerca de la conceptualización de los tesauros, lo que ha dado lugar a la coexistencia de diferentes acepciones de los términos y conceptos implicados. Los diferentes enfoques del tesoro muestran que no existe una única concepción del mismo, y se observa además una definición insuficiente del término clasificación temática en la disciplina bibliotecológica, ya que se presuponen diversas acepciones que es necesario dilucidar y precisar. A este respecto, por ejemplo, Morales (2001: 224) hace la siguiente afirmación... «Considerando que un tesoro no representa un instrumento de clasificación temática, sino una herramienta de control para la indización y recuperación de la información almacenada, JURIVOX plantea una sistematización conceptual que se afianza en la realidad del derecho positivo actual». Esta forma de concebir el tesoro ya no es funcional no sólo porque es contradictoria, dado que su consideración como herramienta de recuperación, involucra la idea de un orden clasificatorio, sino porque es necesario revisar este tipo de conceptos con objeto de facilitar el desarrollo del conocimiento acerca de los problemas que plantea la clasificación documental. Por ello, se parte de la siguiente concepción de tesoro: «Un tesoro se define como una estructura lingüística de clasificación temática o facetada, cuya organización básica está conformada por una relación de descriptores que representan o describen autoridades o contenidos semánticos por medio de unidades lingüísticas extraídas del lenguaje de una disciplina o área específica del conocimiento» (Naumis 2000: 104).

En un primer acercamiento al planteamiento de Morales se aprecia la convivencia de los siguientes pensamientos: 1) el tesoro no es una clasificación temática; 2) el tesoro es una herramienta de control para la indización y la recuperación de información; 3) un tesoro plantea una sistematización conceptual que se afianza en el conocimiento de una disciplina. Subyace en estas palabras la idea que la clasificación temática no está ligada a la indización y que la sistematización conceptual de una disciplina, no es una clasificación del conocimiento.

En este trabajo se pretende probar que en Bibliotecología el tesoro, 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 y un tesoro documental es una clasificación temática. Este análisis se realiza fundamentalmente partiendo de las fórmulas discursivas habituales que vehiculan las relaciones conceptuales o que explicitan propiedades concretas de los conceptos para sistematizar las investigaciones puntuales de los problemas que se plantean, tal como se explica a continuación.

## 1. El lenguaje de la bibliotecología

El lenguaje usado para comunicarse en la ciencia y la tecnología 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. Es necesario mantener una actitud crítica de la terminología utilizada en la disciplina y realizar un esfuerzo por comprender la naturaleza de sus expresiones lingüísticas, sin entrar en la especu-

lación filosófica, únicamente desde el enfoque de la propia Bibliotecología. Para hacer esto partimos de una metodología terminológica, cuyo objetivo es estudiar y explicar la naturaleza y comportamiento de los vocablos que utiliza el conocimiento especializado en su comunicación, tanto en nuestra lengua como en otras (Lara 2001: 180).

En opinión de Lara (2001: 185) el ejercicio y la operación con los objetos no se da en contextos aislados de la vida social, la lengua de la comunidad que rodea al individuo puede o debe tener algún papel en la formación de los esquemas de conocimiento. «La lengua es la que hace inteligibles los conceptos que cada quien elabora (tanto para el mismo, como para los demás), pues su validez y su operatividad no se cristalizan mientras no se los contrasta con la experiencia común, si no los significa con signos, que los alumbran y los ponen en cuestión» (Lara 2001: 188). En consecuencia, según la teoría terminológica la designación de un objeto, no se da nada más a través del término, sino que constituye el instrumento comunitario, mediante el cual se delimita el conocimiento de algo y se confirma. La expresión «concepto» que se utiliza en este trabajo está enmarcada por este pensamiento.

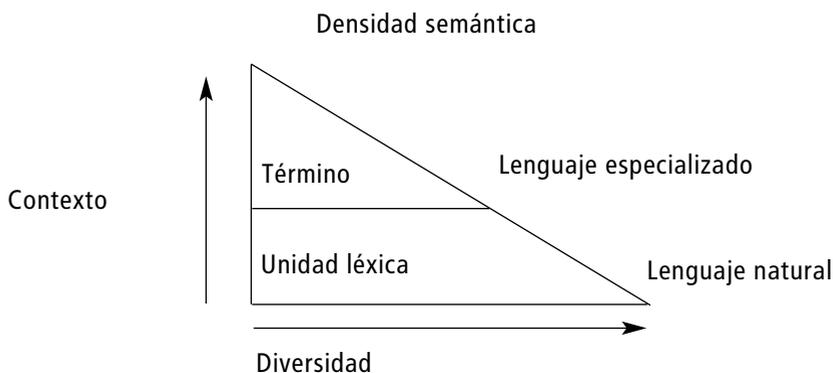
En la última definición de concepto propuesta por el Comité ISO/TC37 (Pozzi 1999: 16) (expresada en la ISO 1087-1:2000, *Terminology work – Vocabulary – Part 1: Theory and application*) se establece al concepto como una unidad de conocimiento creada por una combinación única de características. Se explica además en la norma que el objetivo del concepto-término asignado en un lenguaje especializado dado es asegurar que un término dado es asignado a sólo un concepto y un concepto dado está representado por un solo término. La norma mencionada además de aclarar un poco más el uso del término concepto para el presente trabajo, ayuda a comprender que debe existir una relación biunívoca entre conceptos y términos en cada campo especializado del conocimiento y que la ambigüedad debe ser eliminada, con la finalidad de lograr una comunicación fluida.

La Bibliotecología debe fijar sus conceptos y sus términos, a través de definiciones que describan los conceptos y estructuren el significado de los términos, así como sus limitaciones para asegurar por este camino, la comunicación eficaz y eficiente entre los miembros de su comunidad. Los términos que componen un lenguaje bibliotecológico no tienen un significado opuesto al lenguaje natural, se trata más bien de definir el alcance semántico que tienen los términos en el contexto de la especialidad.

La especificidad y claridad de los términos bibliotecológicos se irá adquiriendo a través de su especialización, agregándoles una densidad semántica relacionada con la propia disciplina y contenido semántico único, con respecto al lenguaje natural. La representación gráfica de este pensamiento se puede observar a través del triángulo que aparece a parece en la página siguiente.

## 2. Indización en Bibliotecología

Antes de realizar el análisis de la relación que mantiene la indización con la clasificación temática, se abordará la investigación terminológica de la indización en la discipli-



na bibliotecológica, tal como se especifica en el nombre asignado al apartado y como primer paso en la búsqueda de la desambiguación de significados del problema que se estudia.

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.

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 Landrigde 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 pro-

fundo 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.

El uso del término indización en el propio campo bibliotecológico exhibe otro uso que agrega un elemento de discusión, ya que en inglés para designar los lenguajes documentales se usa el término «indexing language» y se traduce al español como «lenguaje de indización» en forma literal. Así la herramienta utilizada para indizar por temas incluye la palabra indización. En los cuadros que se presentan más abajo se plantea una diferencia en el uso de la expresión lenguajes de indización para diferentes autores, además de algunos otros que no lo utilizan.

#### Georges Van Slype (1991: 21-30)

		Lenguaje libre
	Lenguajes de indización	
		Lenguaje controlado
Lenguajes documentales		
	Lenguajes de clasificación	

**María Texia Iglesias (1992: 18-24, 88-100)**

		Concordancias
		Índices permutados
	Indización automática	Índice Kwic
		Índice Kwoc
		Indización por citas
		Esquemas de clasificación
	Lenguajes de indización	Tesauros
		Listas de enc. de materia
		Listados de palabras clave

**Blanca Gil Urdiciain (1996: 17-28)**

	Lenguajes libres	Listas de descriptores libres
Lenguajes documentales		
	Lenguajes controlados	Clasificaciones
	Tesauros	

**Jacques Chaumier (1986: 7-12)**

	Lenguajes evolucionados	Transforma los términos del texto
	Lenguajes de descripción de datos	
Lenguaje documental	Lenguajes controlados	Listas de «vedettes» Tesauros
	Clasificaciones	Clasificaciones de tipo decimal Clasificaciones de facetas

Mientras Jacques Chaumier no utiliza el término lenguajes de indización, al igual que Gil Urdiciain sino lenguajes documentales como sinónimo de lenguajes de indización (aunque no mencionan esta otra forma de designación) con la distinción en lenguajes libres y controlados para Gil Urdiciain y con una variedad mayor para Chaumier. Iglesias hace la división entre lenguajes donde no se requiere un proceso intelectual y utiliza la designación lenguajes de indización para la indización humana apoyada en diferentes herramientas de normalización.

Georges Van Slype en cambio utiliza lenguajes controlados, para englobar lenguajes de indización y lenguajes de clasificación. Como se puede observar el término lenguajes de indización comparte créditos con términos sinónimos, pero su calidad de término compuesto ayuda a clarificar su uso.

La indización en todos los casos vistos es un proceso de ordenación insertado para la búsqueda y recuperación de 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 que contiene cada uno de los registros bibliográficos del sistema o palabras significativas del texto completo de los documentos. En cuanto a la designación lenguaje de indización como la herramienta que se utiliza para indizar requiere un análisis particular y más profundo, pero en principio no causa problemas terminológicos para el presente estudio.

### 3. La clasificación temática como un sistema de organización del conocimiento

En primer lugar será abordado el punto de vista terminológico acerca de la clasificación y la clasificación temática para estudiar el problema conceptual de acuerdo a su tratamiento en el dominio bibliotecológico, la relación estrecha que mantiene con el concepto indización y con el tesoro como una herramienta de clasificación temática. Como en el apartado anterior para tejer el argumento cognitivo se parte del estudio de textos especializados que contienen información lexicográfica, terminológica y conceptual sobre aspectos que se consideran relevantes.

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.

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 mis-

mo 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 ahora 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 tesauro aseguran que exista un índice de las temáticas contenidas en el fondo documental.

No se podría decir que el tesauro es una indización temática porque este concepto representa el acto de adjudicar los temas a los documentos, el concepto clasificación temática está más relacionado con las características intrínsecas del tesauro como herramienta donde se presenta una clasificación para indizar. En el próximo apartado se abordará el estudio de la definición de tesauro y de su estructura como clasificación temática.

#### 4. Tesauro documental como clasificación temática

La palabra tesauro, en su origen *thesaurus* o *thesauros*, era el nombre dado en las casas de época griega a la cámara donde se guardaba el *tesoro*. Posteriormente se usó para designar ciertos diccionarios, índices o repertorios de términos (Martínez de Sousa 1993: 776), con la misma connotación de lugar donde se guarda lo más importante, que en este caso es menos corpóreo porque se trata del saber humano. A lo largo de la historia, el término tesauro está asociado a obras de estructuras muy variadas tanto diccionarios como repertorios terminológicos o vocabularios. De hecho, la voz *thesaurus* en latín fue, durante años, el término más utilizado para designar tanto un diccionario muy exhaustivo (de lenguas muertas) como un vocabulario que contuviera todas sus unidades léxicas:

« Existe una gran variedad de opiniones sobre los antecedentes, tanto etimológicos como conceptuales, del tesauro. La mayoría de los autores se limitan a dar cuenta de su raíz semántica griega, como *tesoro*, bien material o del conocimiento. No cabe duda que los precedentes más antiguos podrían ser los léxicos (*Thesaurus*) de autores griegos y latinos y diccionarios de la lengua que desde finales del siglo XV y durante los siglos XVI, XVII y XVIII proliferan en Europa, bajo la denominación de «*Tesoros*» (Rubio, 1999: 57)

El primero que usó la palabra con este sentido fue el florentino Brunetto Latini en su enciclopedia *Livre dou Trésor*, compuesta en francés, —pese a ser italiano su autor— entre los años 1262 y 1268 (Martínez de Sousa 1993: 776). Era la época en que se empezaba a escribir en lenguas vulgares.

En España fue empleada por Vidal de Canyelles (o Canelles), prelado y jurista catalán (muerto en 1252) en su obra *Compilatio Maior o In excelsis Dei Thesaurus*, obra en la que se compila y comenta el derecho aragonés.

En el siglo XV el término aparece en el *Thesaurus adversus haereticus* de Cirilo de Alejandría, copia del original (sin fecha), pero sin duda bastante anterior. Durante el Renacimiento aparecen los diccionarios de una sola lengua, como el *Thesaurus linguae latinae*, de Robert Estienne en 1532 y el *Thesaurus graecae linguae*, de Henri Estienne, fechado en 1572.

El primer diccionario de la lengua española fue el *Thesoro de la Lengua Castellana*, publicado en 1611 por don Sebastián de Covarrubias (o Covarrubias), reimpresso en 1672 por Gabriel de León. Este diccionario antecede por un año a la aparición del *Vocabolario degli Accademici della*

*Crusca*, del italiano. El fundamento de su aparición es «dar noticia a los extranjeros del lenguaje español, y de su propiedad y elegancia, que es muy grande honor de la Nación Española» (Lara 1997: 35). El objetivo de este diccionario era la etimología, «encaminada a descubrir la causa del nombre y, con ello, dar a conocer la realidad de la cosa designada» (Seco 1987:114), como correspondía al planteamiento medieval manifestado, entre otros, por San Isidoro de Sevilla. «En la etimología de cada vocablo –dice Covarrubias– está encerrado el ser de la cosa, sus cualidades, su uso, su materia, su forma, y de alguna de ellas toma nombre» (Lara 1997: 35)

Sin embargo, algunos autores le otorgan la calidad de primer diccionario español al *Diccionario de la Lengua Castellana* (hoy conocido como *Diccionario de Autoridades*)..., «en que se explica el verdadero sentido de las voces, su naturaleza y calidad, con las frases o modos de hablar, los proverbios o refranes, y otras cosas convenientes al uso de la lengua» compuesto por la Real Academia Española en 1726 (Fernández 1989: 167), trece años después de su fundación.

El término se popularizó a partir del *Thesaurus of English Words and Phrases*, un diccionario ideológico recopilado por Peter Mark Roget en 1852 y reeditado unas noventa veces. En la introducción de este diccionario se podía leer la siguiente definición de tesoro: «Un conjunto... clasificado, no en el orden alfabético como lo están (las palabras) en un diccionario, sino según las ideas que expresan... El objetivo consiste, una vez dada la idea, en encontrar la palabra o las palabras a través de las cuales puede expresarse conveniente y justamente esta idea» (Chaumier 1986: 12). Rubio Liniers menciona que las ideas representadas a través de conceptos como cantidad, espacio, tiempo o materia, significan también un claro antecedente de las clasificaciones facetadas que tanta incidencia tendrán en el desarrollo de los tesauros.

En cuanto a la primera vez que el término tesoro fue utilizado como un tipo de instrumento lingüístico para la indización de los documentos, Chaumier presenta en su libro la opinión de Vickery y Lancaster. Según Vickery, fue Helen Brownson la primera que lo empleó en la conferencia sobre clasificación en el año 1957. En cambio, Lancaster le otorga el mérito a Hans Peter Luhn, quien lo empleó en un texto del mismo año 1957 (Chaumier 1986: 11). Los años cincuenta fueron fructíferos para la recuperación de información, a través de la temática de los documentos y así Calvin Mooers es el primero en usar la palabra descriptor para designar los términos de los vocabularios controlados, que ahora es el elemento que conforma y caracteriza el tesoro (Rice 1999: 18).

En la actualidad, la voz de tesoro se usa para designar el instrumento de indización en el ámbito de la Bibliotecología y la Documentación, aunque aparece en algunos diccionarios y enciclopedias usado en el sentido de diccionario, catálogo o antología (El pequeño Larousse Ilustrado 1996) y algunos diccionarios sobre todo etimológicos también utilizan esta designación. Para hacer particular el uso en la Bibliotecología y la Documentación se puede utilizar un término que lo califique y lo distinga, como tesoro documental.

La designación «tesoro» para el concepto manejado en este trabajo es la más común, sin embargo habría que estudiar más a fondo los usos de otros términos que se comienzan a vis-

lumbrar en el ámbito del procesamiento de la información, para asumir una defensa terminológica que rescate significados más o menos permanentes a través del tiempo. La Red Educativa de la Comunidad Europea presentó la versión definitiva de su lenguaje controlado a fines de marzo del 2002 y el nombre que le otorgaron llama la atención porque rescata el término tesoro que le sirvió de origen y utiliza además el término buscador que proviene del medio informático: *European Treasury Browser: Thesaurus*. Parecería que no se consideró clara la utilización única del término tesoro y el nombre adoptado resulta redundante, porque repite características inherentes al concepto tesoro y a la historia del término.

Además de los nombres que se adjudican a los tesauros existen otros casos de estudio como las etiquetas de «control de autoridades» del formato MARC cuya estructura se asemeja a la del tesoro documental y mientras la Library of Congress de los Estados Unidos de Norteamérica habla del uso de encabezamientos de materia tesaurizados para libros y publica el *Thesaurus for graphic materials: access terms*, que no presenta la división en campos o facetas que acompaña siempre una estructura clasificatoria; la British National Library trabaja descriptores en el mismo módulo.

Peter Van Dijk en el anuncio de los trabajos para la versión 2 del XFML (eXchangeable Faceted Metadata Language), contempla la posibilidad de soportar las características de un vocabulario controlado como el tesoro. El XFML es uno de los lenguajes de comunicación contruidos para soportar la norma ISO/IEC 13250 Topic Map, norma elaborada por la International Organization for Standardization y la International Electrotechnical Commission, para la descripción de documentos y procesamiento del lenguaje, mediante una estructura para interrelacionar recursos de información y que adopta el formato de un tesoro basado en una clasificación de tipo facetas. En realidad, la adecuación de un sistema que mantenga la estructura del mapa de tópicos a un sistema de tesauros no requiere de mucho trabajo adicional, por las características comunes de uno y otro.

En el tesoro documental como clasificación temática para concentrar los documentos que contengan una misma temática se construyen agrupaciones jerárquicas, asociativas y equivalentes, es decir estructuras en series horizontales y verticales, unidas por las relaciones de significado entre sí, conformando un verdadero sistema de términos que remiten a conceptos. (La ISO 1087-1: 2000 define «el sistema de conceptos» como «un conjunto estructurado de conceptos, establecido acorde a las relaciones existentes entre los mismos, en el cual cada concepto queda determinado por su posición dentro del sistema de conceptos»). Currás desarrolla un enfoque sistémico de la clasificación documental y explica un sistema de información y un lenguaje documental como expresiones de la misma idea, consideradas la primera desde un prisma de la lógica y la segunda desde la perspectiva de la terminología (Currás 1999: 50).

Ahora bien, ¿cuál es el sustento teórico de estas estructuras de conocimiento que ayudan a mantener una agrupación y por lo tanto un ordenamiento de los documentos? García Marco distingue cuatro tipos de clasificaciones: institucionales, especulativas, idiosincrásicas y documentales. Las clasificaciones institucionales como su nombre lo indica se refieren a una

estructura que refleja el pensamiento de una comunidad científica o educativa en particular. Las clasificaciones especulativas del conocimiento son aquellas construidas por los científicos, especialistas o filósofos como una propuesta de la categorización del pensamiento en un área del conocimiento o en su totalidad. Las clasificaciones idiosincráticas son las adoptadas por individuos o grupos acerca de la división del conocimiento para ser utilizada en una realidad dada y se puede partir de conocimientos comunes o científicos. Por último, las clasificaciones documentales que son las adoptadas como resultado del conocimiento de las anteriores y sirven para organizar colecciones de documentos para luego recuperarlos.

Estos tipos de clasificaciones que se distinguen reflejan el pensamiento que guía al especialista que compone un esquema de organización en un área del conocimiento para construir una clasificación bibliográfica o cualquier otra herramienta para clasificar. El constructor de la herramienta debe utilizar como base de la propuesta lo que aparece designado en la literatura como mapas conceptuales u ontologías (Soergel 1999: 1119) (Vickery 1997: 280), que operan como un esquema clasificatorio documental o una taxonomía que propone una definición del campo de actividad después de un proceso de discusión científica.

El tesoro como sistema de conceptos entrelazados tiene como función la indización de los documentos y como objetivo, la recuperación en los sistemas de información en los cuales se inserta. La misma tecnología que se está generando en el medio de la Informática exige que el tesoro mantenga una estructura cognitiva para que los términos incluidos mantengan una relación entre ellos, como elementos aglutinantes, para que el contexto ayude a entender el término aislado. En un lenguaje especializado la connotación actúa mediante una limitación previa de los atributos que posee para un sistema de conocimientos (Naumis 2000: 98). La opción léxica del tesoro exige desambiguar lo que puede ser polisémico para elaborar la red nocional que exige el instrumento. Es importante aclarar que el descriptor representa conceptos, pero no como elementos aislados, cada concepto individual es parte integrante de un sistema conceptual que a la vez refleja la ordenación y la estructura de algún campo del conocimiento, a diferencia del término que representa conceptos aislados. Un tesoro bien elaborado brinda precisión, claridad, confiabilidad e integridad en los elementos que lo componen, debido a que opera como una estructura de base para la organización de la información (Naumis 2000: 104).

A diferencia de las clasificaciones bibliográficas que son generadas en la lejanía de la experiencia, la construcción de la estructura de un tesoro está más ligada a la realidad del sistema de información al que está destinado. Dice García Marco (1999: 267)... «que las clasificaciones documentales universales, por su propia pretensión de globalización, tienden a alejarse de las culturas institucionales y científicas de los usuarios concretos». La experiencia con los tesoros es que se construyen para un área de conocimiento con una estructura jerarquizada validada por el mismo medio científico, educativo o artístico donde se aplica.

Sin embargo, una estructura sugerida por el propio medio científico, no siempre es clara para la organización de documentos en una especialidad y el documentalista deberá hurgar con

acuciosidad científica en los documentos generados o sistemas implementados para definir el tipo de clasificación temática que le servirá de base.

En un apartado sobre la necesidad de la clasificación en los entornos automatizados García Marco (1999: 249) presenta los tesauros como una de los modelos teóricos de moderna clasificación analítico-sintética para complementar la organización sistemática con el acceso alfabético y dice «los tesauros en su más elevado estado de desarrollo, cuando superan la fase durante la cual no son más que agregados de campos léxicos, terminan constituyendo elaboradas clasificaciones, a las cuales sólo les falta un mecanismo sintético de precoordinación».

La tendencia en la actualidad es dotar a los términos descriptores de un mayor grado de pre-coordinación, con lo cual se les otorga la síntesis de temáticas muy específicas. La conformación del sistema conceptual de los descriptores tesaurales, se realiza a través de las relaciones jerárquicas, asociativas y equivalentes, con los siguientes resultados de la contextualización: desambiguación de los términos polisémicos, asistencia para la navegación al ubicar temas generales y específicos, efectividad en las estrategias de recuperación de información y posibilidades de aplicación en mecanismos de inteligencia artificial.

## 5. Consideraciones finales

La Bibliotecología como disciplina necesita asignar significados precisos a sus términos. Los esfuerzos individuales de interpretación proponiendo usos fijos para los significados de los términos permitirán mejorar la claridad y precisión de la comunicación.

El definir procesos e ideas involucradas en la clasificación y la indización y relacionarlos con su expresión lingüística es reconciliar diferentes caminos para referirse a técnicas, conceptos y principios similares. Es importante no perderse en los modos de aplicación de las técnicas y prestar atención a los fundamentos y objetivos de los procesos a los que se debe someter la información para su recuperación.

El diseño de sistemas temáticos para agrupar documentos y de esta forma organizarlos para su recuperación, descansa en fundamentos intelectuales del conocimiento adquirido por el hombre. En el proceso de asignación de un lugar en un esquema cognitivo intervienen tanto la clasificación como la indización. La indización temática supone la ubicación de un documento o parte de un documento en un orden clasificatorio, por lo tanto la indización temática es una expresión de la clasificación apropiada para organizar información y en consecuencia si existe un proceso de asignación temática existe una clasificación sobre la cual se sustenta.

Una clasificación exhibe la relación temática que existe entre los conceptos incluidos en ella y es independiente de los códigos que se utilicen para representarlos. Los términos de un sistema de clasificación para el ámbito del procesamiento documental están contextualizados, se designan descriptores y suponen la existencia de categorías semánticas, aclaración de las alterancias en los usos lingüísticos reflejados en el sistema y el reconocimiento de propiedades comunes entre ellos.

El tesoro documental en el ambiente digital que prevalece en la actualidad es una herramienta de sistematización conceptual de una disciplina mediante una clasificación de sus significados armados con una estructura teórica o pragmática que coadyuve a la univocidad de los términos, en un entorno que propicie la comprensión de los mismos desde una perspectiva interdisciplinaria.

## 6. Referencias bibliográficas

AMAR, Muriel

- 2000 *Les fondements théoriques de l'indexation: une approche linguistique*. Paris: Association des professionnels de l'information et de la documentation. 355 pp. (Collection Sciences de l'information, série Recherches et documents).

BRONOWSKI, Jacob

- 1993 *Los orígenes del conocimiento y la imaginación*. Traducción de Enrique Lynch, 2.<sup>a</sup> ed., Barcelona, Editorial Gedisa, 152 pp. (Colección hombre y sociedad: Filosofía).

CHAUMIER, J.

- 1986 *Análisis y lenguajes documentales: el tratamiento lingüístico de la información documental*, traduc. L. Porta, Barcelona, Mitre, 173 pp.

CLEVELAND, Donald y Ana D. CLEVELAND

- 1990 *Introduction to indexing and abstracting*, 2nd ed., Englewood, Col., Libraries Unlimited, xiii, 329 pp.

COLLARD, Claude, Isabelle GIANNATTASIO y Michel MELOT

- 1995 *Les images dans les bibliothèques*, Paris, Éditions du Cercle de la Librairie, 390 pp., il., (Collection Bibliothèques).

CURRÁS, Emilia

- 1999 Enfoque sistémico de la clasificación documental, pp. 49-63, en *Manual de clasificación documental*, editora María Pinto, Madrid, Editorial Síntesis (Biblioteconomía y documentación).

ESTEBAN NAVARRO, Miguel Ángel

- 1999 «Fundamentos epistemológicos de la clasificación documental», pp. 19-32, en *Manual de clasificación documental*, editora María Pinto, Madrid, Editorial Síntesis, (Biblioteconomía y documentación).

FERNÁNDEZ GONZÁLEZ, Ángel Raimundo, Salvador HERVÁS y Valero BÁEZ

- 1989 *Introducción a la semántica*, 5.<sup>a</sup> ed., Madrid, Cátedra, 252 pp.

GARCÍA MARCO, Francisco Javier

- 1999 «Clasificación y recuperación de información», pp. 247-285, en *Manual de clasificación documental*, editora María Pinto, Madrid, Editorial Síntesis (Biblioteconomía y documentación).

## GIL URDICIAIN, Blanca

- 1996 *Manual de lenguajes documentales*, Madrid, Editorial Noesis, 269 pp.
- 1999 «Los lenguajes de clasificación documental», pp. 91-108, en *Manual de clasificación documental*, editora María Pinto, Madrid, Editorial Síntesis (Biblioteconomía y documentación).

## IGLESIAS MATURANA, María Texia

- 1992 *Análisis y recuperación de información documental*, Santiago de Chile, Ediciones I.P.S., xvi, 237 pp.
- 2000 ISO 1087-1: 2000 *Terminology work Vocabulary Part 1: theory application*. Ginebra, International Standardization Organization.
- 1999 ISO/IEC 13250-1999 *Topic Maps: Information Technology: Document Description and Processing Languages*. Ginebra, International Standardization Organization, International Electrotechnical Commission, 47 pp.

## KEDROV, Bonifatti Mikhailovich

- 1974 *Clasificación de las ciencias*, Moscú, Editorial Progreso, 2 v.

## LAFUENTE LÓPEZ, Ramiro

- 1993 *Los sistemas bibliotecológicos de clasificación*, México, UNAM. CUIB, 112 pp. (Serie monográfica; 15).

## LANCASTER, F. W.

- 1996 *Indización y resúmenes: teoría y práctica*, traducción de Elsa Barber, Buenos Aires, EB Publicaciones, xii, 337 pp.

## LANGRIDGE, D. W.

- 1992 *Classification: its kinds, systems and application*, Londres, Bowker Saur, x, 84 pp. (Topics in library and information studies).

## LERAT, Pierre

- 1997 *Las lenguas especializadas*, traduc. de Albert Ribas, Barcelona, Editorial Ariel, 221 pp. (Ariel Lingüística).

## LARA, Luis Fernando

- 2001 *Ensayos de teoría semántica: lengua natural y lenguajes científicos*, México, El Colegio de México, 261 pp. (Jornadas; 135).

## MORALES CAMPOS, Estela

- 1999 Reseña de Torres Llosa, Enrique y Violeta Angulo Morales. JURIVOX: Macrotesauro de Derecho (primera parte), Buenos Aires, Alfabrama-Universidad Femenina del Sagrado Corazón, pp. 223-225, en *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, vol. 15, n.º 31 (jul.-dic., 2001), ISSN 0187-358X.

NAUMIS PEÑA, Catalina

- 2000 «Análisis de la confluencia entre término y descriptor en la elaboración de tesauros», pp. 95-113, en *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, vol. 14, n.º 29 (jul.-dic.), ISSN 0187-358X.

PINTO, María y Carmen GÁLVEZ

- 1999 Hacia una teoría integradora de la clasificación documental, pp. 33-48, en *Manual de clasificación documental*, editora María Pinto, Madrid, Editorial Síntesis (Biblioteconomía y documentación).

POZZI, María

- 1999 «The concept of "concept" in terminology: a need for a new approach», pp. 28-42, en *Terminology and Knowledge Engineering*, ed. Peter Sandrini, Vienna, Termnet.

RICE, Stanley

- «Autospec thematics: keywords and minimum vocabularies in concept filtering of images», <http://www.cruzio.com/~autospec/access2.htm>, fecha de consulta 16/04/2003.

RUBIO LINIERS, María Cruz

- 1999 *Tesauro de historia contemporánea de España*, Madrid, Consejo Superior de Investigaciones Científicas, 326 pp.

SLYPE, Georges Van

- 1991 *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*, Madrid, Fundación Germán Sánchez Ruipérez, 200 pp.

SOERGEL, Dagobert

- 1999 «The rise of ontologies or the reinvention of classification», pp. 1119-1120, *Journal of the American Society for Information Science*, vol. 50, n.º 12.

SVENONIUS, Elaine

- 2000 *The intellectual foundation of information organization*, Cambridge, MA, Massachusetts Institute of Technology, xiv, 255 pp. (Digital libraries and electronic publishing).

VICKERY, B. C

- 1997 «Ontologies», pp. 277-286, *Journal of Information Science*, vol. 23, n.º 4.

VIZCAYA ALONSO, Dolores

- 1997 *Información: procesamiento de contenido*, Rosario, Parhadigma Ediciones, 187 pp.