



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

POSGRADO EN BIBLIOTECOLOGÍA
Y ESTUDIOS DE LA INFORMACIÓN

**ANÁLISIS DEL LENGUAJE RELACIONADO CON
EL AGUA:
NECESIDAD DE UN TESAURO PARA MÉXICO**

T E S I S

QUE PARA OPTAR POR EL GRADO DE

MAESTRA EN BIBLIOTECOLOGÍA

Y ESTUDIOS DE LA INFORMACIÓN

P R E S E N T A

VERÓNICA EUGENIA VARGAS SUÁREZ

DIRECTORA DE TESIS:
DRA. CATALINA NAUMIS PEÑA

MÉXICO, D.F.

ABRIL DE 2007



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mis padres, que han tenido la sabiduría de soportarme
en las búsquedas e inseguridades propias de los viajes.

A Paquita, compañera de viaje, por sus enseñanzas y alegría de vida;
por el milagro de crear y criar juntas.

A Mateo y Paula. A Nicolás, Benita, Tita y Mischa.
A todos ellos, por seguir educándome con sus ocurrencias.

Agradecimientos

A la Dra. Catalina Naumis, por compartir conmigo sus experiencias y conocimientos.

A la Dra. Elsa Barberena, por confiar en mí y permitirme adentrarme en el mundo de la Bibliotecología, y por sus observaciones a este trabajo. Así como a mis demás sinodales, los doctores Filiberto Felipe Martínez, Salvador Gorbea y Araceli Torres.

A la Dra. Elsa Ramírez, por todo su apoyo y entusiasmo.

Al Dr. Adolfo Rodríguez Gallardo, por sus consejos y orientaciones.

A la Lic. Gabriela Olguín, por su paciencia e invaluable ayuda en el transitar administrativo.

A mis maestros del Centro Universitario de Investigaciones Bibliotecológicas, por la formación recibida.

Tabla de contenido

Introducción	1
Capítulo I. Marco teórico-conceptual	8
1. Vocabularios controlados	8
1.1 Definición	8
1.2 Características	12
1.3 Utilidad	15
2. La indización y recuperación de la información	19
2.1 Indización	19
2.1.1 Definición	19
2.1.2 Metodología	21
2.1.3 Evaluación	23
2.2 Recuperación de la información	25
2.2.1 Sistemas de recuperación de información	25
2.2.2 Evaluación de los sistemas de recuperación	28
2.3 La relevancia	31
3. El tesaurus	34
3.1 Definición	34
3.2 Estructura y presentación	36
3.3 Elaboración y mantenimiento	40
3.4 Los tesaurus y el entorno digital	43
Capítulo II. Metodología	46
1.1 Evaluación de tesaurus relacionados con el agua	46
1.2 Criterios de evaluación	46
1.3 Procedimiento de evaluación	51
1.4 Fuentes de estudio	53
Agrovoc	53
General Multilingual Environmental Thesaurus	58
OECD Macrothesaurus	64
Tesaurus de Ingeniería Hidráulica	70
Tesaurus de Ingeniería Sanitaria y Ambiental	73
Thésaurus Eau	77
Capítulo III. Análisis de resultados	81
Consideraciones finales	90
Obras consultadas	95
Anexo	

Tabla de contenido

Introducción

Capítulo I. Marco teórico-conceptual

1. Vocabularios controlados

1.1 Definición

1.2 Características

1.3 Utilidad

2. La indización y recuperación de la información

2.1 Indización

2.1.1 Definición

2.1.2 Metodología

2.1.3 Evaluación

2.2 Recuperación de la información

2.2.1 Sistemas de recuperación de información

2.2.2 Evaluación de los sistemas de recuperación

2.3 La relevancia

3. El tesoro

3.1 Definición

3.2 Estructura y presentación

3.3 Elaboración y mantenimiento

3.4 Los tesauros y el entorno digital

Capítulo II. Metodología

1.1 Evaluación de tesauros relacionados con el agua

1.2 Criterios de evaluación

1.3 Procedimiento de evaluación

1.4 Fuentes de estudio

Agrovoc

General Multilingual Environmental Thesaurus

OECD Macrothesaurus

Tesoro de Ingeniería Hidráulica

Tesoro de Ingeniería Sanitaria y Ambiental

Thésaurus Eau

Capítulo III. Análisis de resultados

Consideraciones finales

Obras consultadas

Anexo

*El agua como conflicto, cuando sobra o cuando falta,
cuando se desborda o cuando desaparece.
El agua como rehén de las ganancias o los dispendios,
de los escamoteos o los desperdicios.*
Roberto Zamarripa
(Reforma, 13.03.06)

INTRODUCCION

El objetivo principal de esta investigación es probar que es necesario contar con una herramienta terminológica relacionada con la temática hídrica y determinar si se puede adoptar y/o adaptar algún tesoro vigente a la realidad mexicana. Lo anterior, debido a que los tesauros existentes, bajo las condiciones actuales, no permiten una mayor precisión y el rigor necesario para garantizar la calidad en los procesos de indizar y recuperar la información relacionada con el agua.

Para alcanzar este objetivo general, se han planteado tres objetivos específicos: 1) analizar descriptivamente seis tesauros; 2) examinar de manera general el manejo de los términos, así como los temas contemplados y 3) establecer las características que debe cumplir un tesoro sobre el tema hídrico a partir de los análisis llevados a cabo.

Tomando en cuenta todos los recursos tanto humanos como financieros que intervienen en la construcción de un tesoro, razón por la cual los estudiosos recomiendan plantearse la elaboración de uno nuevo solamente cuando se hayan agotado las posibilidades de adoptar y/o adaptar alguno ya existente, se llevó a cabo un análisis descriptivo de seis tesauros relacionados con el tema hídrico.

Pero, ¿por qué un tesoro del agua para México? Amén de que en otras lenguas existan tesauros sobre el recurso hídrico, la falta de una normalización del lenguaje en español empleado en nuestro país y relacionado con la temática hídrica ha ocasionado una importante inconsistencia en el proceso de indización que, sin duda, ha dado lugar a la dispersión de la información generada sobre el tema. Se ha de tomar en cuenta que la normalización del lenguaje facilita la identificación y la recuperación de la información, evitando las confusiones a que

se pueden prestar los homónimos, sinónimos o la variedad de nombres con los que puede ser denominado un concepto.

Por otro lado, además de ser un recurso vital para todo ser vivo, el agua es indispensable para el desarrollo social y económico de cualquier comunidad. El problema de la escasez del recurso hídrico en nuestro país, y a nivel mundial, se percibe cada vez más grave y es más común y cercano. Actualmente México enfrenta problemas tanto de escasez como de contaminación del recurso hídrico.

Según la Comisión Nacional del Agua (Conagua), el crecimiento de la población y de las actividades económicas ha provocado una disminución sustancial en la disponibilidad *per cápita* de agua. En el lapso comprendido entre los años 2000 y 2005 la disponibilidad se redujo 5.5% al pasar de 4,841 a 4,573 m³ por habitante al año¹. La misma Conagua señala que a partir de la década de 1970 ha venido aumentando considerablemente el número de acuíferos sobreexplotados, de 32 en 1975 a 104 en 2005. Cabe resaltar que de los acuíferos se extrae casi el 60% del agua para todos los usos².

Por otra parte, según datos de la Secretaría de Medio Ambiente y Recursos Naturales, la calidad del agua superficial, que es la única que se mide sistemáticamente, indica que 66% es excelente o aceptable mientras que el resto requiere tratamiento o se encuentra severamente contaminada³.

El segundo informe de las Naciones Unidas sobre el desarrollo de los recursos hídricos en el mundo, *El agua: una responsabilidad compartida* (2006), coincide al señalar que, a nivel mundial, el problema de la disponibilidad efectiva del agua es aun mayor por los desequilibrios hidráulicos que ocasiona el constante crecimiento de la demanda, la ineficiencia de su uso y el aumento de los niveles de contaminación ocasionados por prácticas inadecuadas en esquemas de producción y consumo. Además menciona que debido al crecimiento de la

¹ Comisión Nacional del Agua. *Estadísticas del agua en México 2006*, p. 39

² *Ibid*, p. 52

³ Jiménez, Blanca *et al.* *El agua en México vista desde la academia*. México : Academia Mexicana de Ciencias, 2005, p.11.

población urbana, muchas grandes ciudades se han visto obligadas a importar agua de cuencas cada vez más lejanas, ya que las fuentes locales de aguas superficiales y subterráneas han dejado de satisfacer la demanda de agua por agotamiento o contaminación.

Evidentemente es esencial sensibilizar a la sociedad ante esta situación. No hay que olvidar que la falta de una gestión sustentable del agua parte del desconocimiento de lo que se quiere proteger, cuidar y usar. Como lo advierte Corral Verdugo (citado por Vargas Pasaye, 2005), “La responsabilidad fundamental de la academia es la investigación y la difusión de los hallazgos científicos al respecto de las causas y las consecuencias de la crisis del agua. Les corresponde a los políticos, incluidos los universitarios (autoridades), hacer llegar a la sociedad la información y la tecnología (tanto física como social) derivada de este conocimiento”⁴.

Indiscutiblemente una sociedad no informada es una sociedad no participativa, por lo que es trascendental que la información relacionada con el recurso hídrico esté a disposición de todos. Si partimos de la base de que la participación ciudadana es básica para la solución de la problemática del agua, entonces es necesario que la información fluya hacia la sociedad con la finalidad de que ésta tome conciencia sobre el valor ambiental, económico y social del recurso. Ciertamente es que el estar informado sobre la problemática real permite la sensibilización y crea el espacio y los elementos necesarios a través de los cuales los patrones de consumo, uso y conservación se modifican.

En síntesis, la falta de uniformidad en los términos empleados y la inexistencia de un vocabulario controlado en lengua castellana sobre el tema del agua *per se*, así como la problemática del agua en México, fueron las razones principales que motivaron esta investigación, cuyo objeto ha sido probar la necesidad de contar con una herramienta terminológica y determinar si es factible adoptar y/o adaptar

⁴ Vargas Pasaye, Rafael G. (2005) “Los académicos alertan sobre el problema del agua”. *Campus suplemento universitario*. México : UNAM, septiembre, No. 143. <<http://www.campusmilenio.com/>> Consulta: 04.09.05

algún tesoro existente, a través del análisis descriptivo de seis tesauros que de alguna manera u otra están relacionados con el tema del agua.

Es necesario enfatizar que la gestión del recurso hídrico no es tarea de una sola disciplina, es multi e interdisciplinaria. Además, la naturaleza heterogénea de la gestión del recurso hídrico provoca la generación de una cantidad considerable de información, razón por la cual es necesario estudiar el ámbito de conocimiento en que ocurren los fenómenos de transmisión de información y las relaciones entre ellos, para poder representarlo con mayor éxito de recuperación de documentos en el sistema de información.

Hipótesis.

El supuesto del que partimos es que los tesauros existentes, relacionados con la temática hídrica, no permiten indizar ni recuperar la información de manera precisa, pues en su mayoría no incluyen la terminología empleada en México.

Metodología.

Sin duda, la carencia de un vocabulario controlado en español sobre la temática hídrica, tratada holísticamente, ha provocado que el bibliotecólogo en este campo no cuente con una herramienta que le permita una mayor precisión y el rigor necesario para garantizar la calidad en los procesos de indizar y recuperar la información. Es decir, la experiencia nos ha mostrado que existe una grave inconsistencia en los términos utilizados en la literatura en castellano, y como es sabido las inconsistencias en la terminología pueden dar lugar a la dispersión de documentos sobre un mismo tema.

Como muestra de lo anterior podemos citar el siguiente ejemplo: el término “gestión integrada de los recursos hídricos”⁵ también es empleado como “manejo

⁵ Vale aclarar que la *Ley de Aguas Nacionales* (2004) en su título primero, capítulo único, artículo 3, fracción XXIX define la gestión integrada de los recursos hídricos como el “...proceso que promueve la gestión y desarrollo coordinado del agua, la tierra, los recursos relacionados con éstos y el ambiente, con el fin de maximizar el bienestar social y económico equitativamente sin comprometer la sustentabilidad de los ecosistemas vitales. Dicha gestión está íntimamente vinculada con el desarrollo sustentable”. Es decir, los

integrado de los recursos hídricos”, “administración integrada de los recursos hídricos”, “gestión integral de los recursos hídricos”, “manejamiento integrado de los recursos hídricos”, “gestión integral del recurso hídrico”, “gestión integral del agua”, “gestión integrada del agua”, “gestión integrada de recursos hidráulicos”, “ordenación integrada de los recursos hídricos”, “gestionamiento integral de los recursos hídricos”, “gestión sostenible del agua”, por mencionar sólo algunos.

De igual manera, encontramos términos como “desertización” vs. “desertificación”, “desarrollo sostenible” vs. “desarrollo sustentable”, “resolución de problemas” vs. “solución de conflictos”, “xeripaisajismo” vs. “jardinería conservacionista”, etcétera.

Como se comentó anteriormente, la elaboración de un tesoro involucra el trabajo de un gran número de especialistas y la inversión de tiempo y de recursos financieros, por lo que se aconseja agotar las posibilidades de adoptar y/o adaptar alguno ya existente. En palabras de Fernández Sevilla (1974: 56), “El tesoro es una magna compilación -idealmente exhaustiva- de todo el léxico de una lengua. Lo que se pretende es inventariar el léxico de una ciencia utilizando todas las manifestaciones escritas y a ser posible orales. Un proyecto de tal importancia precisa de una minuciosa y complicada planificación, así como de un mantenimiento, que puede resultar algo costoso, porque -todavía hoy- las máquinas no han conseguido funcionar sin el ser humano”.

Por lo anterior, hemos emprendido la tarea de analizar descriptivamente seis tesauros⁶: *Agrovoc*, *General Multilingual Environmental Thesaurus*, *OECD Macrothesaurus*, *Tesoro de Ingeniería Hidráulica*, *Tesoro de Ingeniería Sanitaria y Ambiental* y *Thésaurus Eau*. Si bien el número de tesauros seleccionados no pretende ser de alguna manera exhaustivo, dicha selección se encuentra guiada

recursos hídricos se refieren a los recursos naturales asociados al agua como el suelo y la vegetación, por lo que no podemos referirnos únicamente con gestión o administración integrada del agua al recurso hídrico como lo hacen algunos términos.

⁶ Cabe señalar que el *Water Resources Thesaurus*, del Departamento del Interior de los Estados Unidos, no se incluyó, ya que su última edición es de 1980 y se encuentra fuera de circulación. Por su parte, el *InterWATER Thesaurus*, elaborado por el Centro Internacional de Agua y Saneamiento (mejor conocido como IRC), con sede en Holanda, estuvo disponible una vez que ya habíamos concluido nuestro análisis. <http://www.thesaurus.watsan.net/> [Consulta: 06.11.06]

por las características que a nuestro criterio son relevantes para poder contemplar la posibilidad de adoptar y/o adaptar alguno ya existente a la realidad mexicana.

Para realizar dicha tarea se ha tomado en cuenta la inclusión del tema del recurso hídrico, desde todos los enfoques posibles; la institución que respalda al tesoro; y la disponibilidad y accesibilidad, condiciones indispensables para poder llevar a cabo el estudio correspondiente. Al no contar con una metodología sistemática y probada en el medio para el análisis de este tipo de herramientas terminológicas, se tomó la propuesta de Naumis (2002), a la que se le añadieron algunos elementos.

Asimismo, consideramos pertinente definir ciertos términos, pues un concepto no es algo simple que se puede captar o no, sino que es algo complejo que sólo puede aprenderse poco a poco. En otras palabras, al entender más proposiciones se comprende más el concepto, el cual nos ayuda a entender la realidad y no a manipularla.

Organización del trabajo.

Tomando en cuenta lo anteriormente expresado, en la primera parte de esta investigación hemos optado por abocarnos a la tarea de discernir sobre algunos conceptos que nos permitieran comprender con mayor profundidad a los vocabularios controlados, su uso y el papel que juegan dentro de los procesos de indizar y recuperar la información. De igual manera, para poder llevar a cabo un análisis objetivo de los tesauros seleccionados, evidentemente consideramos indispensable discurrir sobre la conceptualización del tesoro.

Asimismo, considerando que la información que ha de recuperar el usuario para satisfacer sus necesidades de información debe ser relevante, pues es la relevancia precisamente la esencia de la indización, consideramos apropiado comentar sobre la relevancia y la pertinencia, ya que ambos conceptos están en la base del resto de las medidas que tradicionalmente se vienen aplicando en los sistemas de recuperación de información.

En la segunda parte se incluye el análisis de los tesauros arriba mencionados. Dicho análisis se llevó a cabo considerando tres elementos principales: 1) el contenido, que alude a la cobertura temática, a su nivel de estructuración y a su mantenimiento y actualización; 2) la presentación, que se refiere a las características distintivas de cada tesoro y a la forma como éstas son tratadas; y 3) la consulta, la cual hace mención a la facilidad de manejo del tesoro. Es decir, al tipo de soporte, así como a su disponibilidad y accesibilidad, y a la claridad con la que se presenta su guía de uso.

El análisis también incluye un examen general semántico de los términos empleados en cada tesoro, además de los temas contemplados. Nuestro propósito se limitó a marcar ciertas similitudes y diferencias en el uso y acepción de los términos, por lo que consideramos pertinente mencionar algunos ejemplos.

En la tercera parte exponemos el análisis de los resultados que nos han permitido inferir la factibilidad de adoptar y/o adaptar alguno de los tesauros analizados y demostrar que es necesario contar con una herramienta terminológica que refleje la realidad actual mexicana. Asimismo, establecemos las características que a nuestro parecer debe cumplir un tesoro sobre el tema hídrico para México, a partir de los análisis llevados a cabo.

Para terminar, presentamos las consideraciones finales a modo de recapitulación.

El trabajo incluye un anexo, el cual describe detalladamente los elementos que se consideraron para el análisis de los tesauros consultados.

Por último, es preciso mencionar que toda traducción es traición y siempre tendrá en contra cierta pérdida de ritmo y de significaciones inherentes a la lengua en que se pensaron y escribieron, razón por la cual no se tradujeron ciertas definiciones y aseveraciones y se han escrito tal cual se han leído.

*Concepts are mental constructs with unclear boundaries
and to use them as a basis for decision making invites inconsistency.*
E. Svenonius

Capítulo I

MARCO TEORICO-CONCEPTUAL

En el marco de la investigación que nos ocupa no consideramos pertinente debatir sobre la naturaleza de los lenguajes documentales, pues esta tarea nos llevaría a multitud de cuestiones que quedan lejos de nuestros objetivos. Sin embargo, creemos necesario definir ciertos conceptos que nos permitirán entender la utilidad e importancia de los vocabularios controlados dentro de los procesos de indización y de recuperación de la información, y sobre todo comprender el papel que el tesauro, como vocabulario controlado que es, juega dentro de estas actividades.

1. Vocabularios controlados

1.1 Definición

Con el propósito de obtener una comprensión más amplia, a continuación presentamos diversas acepciones sobre el concepto de vocabulario controlado, pero antes de entrar en materia queremos mencionar que durante la revisión de la literatura bibliotecológica pudimos percatarnos de cierta falta de uniformidad en el uso de los términos “lenguaje documental” y “vocabulario controlado”. Observamos que en español diversos autores emplean indistintamente los términos “lenguaje documental”, “lenguaje documentario”, “lenguaje artificial”, “lenguaje controlado”, “lenguaje estructurado”, “lenguaje descriptor”, “lenguaje de indización”, “lenguaje de indización controlado”, “lenguaje terminológico”, “lenguaje normalizado”, “lenguaje de búsqueda de información”, “lenguaje de recuperación”, por mencionar algunos¹, para referirse a un lenguaje intermediario que sirve de puente entre las informaciones contenidas en los documentos y las informaciones solicitadas por los usuarios.

También observamos que rara vez en español y en francés se emplea el término “vocabulario controlado”. Sin embargo no ocurre lo mismo en inglés, que por lo

¹ Es interesante lo que señala Soergel: “The use of a different term is symptomatic of the lack of communication between scientific communities”.

general se emplea el término “controlled vocabularies”, aunque el *UK Archival Thesaurus* utiliza como término preferente “controlled languages”. Al respecto, Svenonius (2003: 822) hace notar que

“The term *controlled vocabulary* is often used interchangeably with *controlled language*. There is imprecision in this usage insofar as a language, whether it be a natural language like English or an artificial language like a retrieval language, can be regarded as consisting not only of a vocabulary (i.e., set of terms), but also of a syntax and a semantics. On the other hand, the confusion is perhaps understandable since a controlled vocabulary incorporates the relational and referential semantics of a retrieval language as well as some of the syntax of the language”.

Es decir, para la autora la confusión es hasta cierto punto comprensible pues el vocabulario controlado incorpora las relaciones semánticas y de referencia del lenguaje artificial así como algo de la sintaxis del lenguaje natural.

Si bien el término “lenguaje documental” es el que se emplea con más frecuencia en la literatura bibliotecológica en castellano, cabe aclarar que para los fines de este trabajo, y para no caer en las inconsistencias ya mencionadas y tampoco alterar los conceptos manejados por los autores consultados, consideraremos el término lenguaje documental como expresión equivalente de vocabulario controlado.

Entrando en materia y hablando en términos bibliotecológicos², el *Diccionario enciclopédico de ciencias de la documentación* (2004), define el vocabulario controlado como una

“Lista de términos que deben ser utilizados para representar el contenido de los documentos en la indización y en la recuperación. Esta lista se establece *a priori* y es cerrada y nominativa”.

² Es de destacar que Domingo Buonocuore no incluye en su *Diccionario de bibliotecología* (1976) definición alguna para las locuciones “lenguaje documental” y “vocabulario controlado”, sólo proporciona una definición para “lenguaje” y “lenguaje mecánico”. Por su parte, Martínez de Sousa, en su *Diccionario de bibliología y ciencias afines* (1989), tampoco ofrece una definición para “vocabulario controlado” ni para “lenguaje documental”, a este último lo considera como sinónimo de “lenguaje artificial”.

El mismo diccionario brinda la siguiente definición para lenguaje documental:

“Lenguaje formalizado que permite representar el contenido de los documentos y de las peticiones de búsqueda”.

Y aclara,

“No existe un criterio unánime en la denominación de este concepto, aunque las dos expresiones más utilizadas son las de lenguaje documental y lenguaje de indización. Mientras que para unos autores son términos sinónimos, para otros el lenguaje de indización sería un tipo de lenguaje documental”.

Por su parte, Lancaster (1986: 8) explica que

“A controlled vocabulary need be nothing than a limited set of terms that must be used by both the indexer and the searcher”.

En otras palabras, tanto el lenguaje del indizador como del usuario deben coincidir. El autor continúa explicando que para los términos deberá contemplarse alguna estructura de forma tal que sean vinculados aquellos cuyo significado esté relacionado. Esto permitirá, tanto al indizador como al usuario, una mejor selección de los términos que representen un tema en particular. Y más importante aun, ayudará al usuario a identificar todos los términos que se requieren para llevar a cabo una búsqueda exhaustiva.

Van Slype (1991: 21) entiende que

“un lenguaje documental será entonces todo sistema de signos que permita representar el contenido de los documentos con el fin de recuperar los documentos pertinentes en respuesta a consultas que tratan sobre ese contenido”.

Cleveland y Cleveland (2001: 35) explican que un vocabulario controlado es un lenguaje artificial que sitúa al indizador entre el texto y el usuario. Es decir, los vocabularios controlados involucran de manera particular al indizador con objeto de mejorar la relación conceptual entre éste y el usuario durante la recuperación de la información, ya que el indizador examina un documento, lo analiza y posteriormente elige los términos de un vocabulario controlado que representan

los conceptos apropiados y las relaciones como éste las interpreta; el usuario realiza la búsqueda utilizando términos y conceptos propios. La función del lenguaje documental como mecanismo de control es guiar tanto al indizador como al usuario al mismo punto.

Por último, la Library and Archives Canada³ señala que

“A controlled vocabulary is an established list of standardized terminology for use in indexing and retrieval of information... A controlled vocabulary ensures that a subject will be described using the same preferred term each time it is indexed and this will make it easier to find all information about a specific topic during the search process”.

A partir de las características de las definiciones anteriores, en adelante cuando hablemos de vocabularios controlados nos referiremos a todos aquellos conjuntos normalizados y normativos de términos, organizados según una metodología en la cual es posible especificar las relaciones entre conceptos de forma semántica, con el propósito de optimizar el proceso de recuperación de información reduciendo la incertidumbre entre las palabras que elige el indizador para representar un texto y las que emplea el usuario para recuperarlo. Los vocabularios controlados son usados como una especie de filtro entre el lenguaje utilizado por el autor y la terminología⁴ del área y también pueden ser considerados como asistentes de investigación ayudando al usuario a refinar, expandir o enriquecer sus investigaciones proporcionando resultados más objetivos.

1.2 Características

Existe una gran variedad de vocabularios controlados que se distinguen por su riqueza, su extensión, su organización o la manera de utilizarlos. Ya se trate de

³ <http://www.collectionscanada.ca/index-e.html> [Consulta: 02.12.05]

⁴ Fernández Sevilla (1974: 117) explica que “...las terminologías tienen carácter de nomenclaturas, son enumerativas y se adaptan a las diferencias y delimitaciones de los objetos. Dicho con otras palabras, las terminologías responden a la realidad y a sus características, mientras que las voces del lenguaje ordinario responden a la *visión* que de la realidad se ha formado cada cultura”. Por otro lado, Cabré (1999: 238) señala que “la terminología es una forma de representar la realidad especializada”.

tesauros, de encabezamientos de materia, de listas de autoridad, de ontologías, etc., todos estos vocabularios pertenecen a la misma familia, tienen el mismo fin y presentan numerosas características comunes.

No hay que olvidar que el vocabulario controlado se emplea cuando se realizan las operaciones de entrada en el almacenamiento y búsqueda de información y las operaciones de salida y de difusión. Por tanto, hay que tener en cuenta que se trata principalmente de herramientas destinadas a realizar operaciones determinadas, en circunstancias concretas y con el fin de dar respuestas a necesidades definidas. Estas condiciones funcionales son esenciales y tan importantes como las consideraciones de orden lingüístico.

Es decir, los vocabularios controlados deben tener en cuenta las relaciones entre las palabras del lenguaje natural, sus distintos valores y eliminar, desde el principio, las ambigüedades que reducen normalmente el contexto y el diálogo en la comunicación en este tipo de lenguaje, condiciones que no existen en la comunicación documental, en la que cada palabra sólo se considera en sí misma. Además, los tipos de vocabularios controlados se distinguen unos de otros de acuerdo con distintos criterios: el principio de clasificación o de construcción; la extensión del campo que abarcan; las clases de palabras que se emplean; los tipos de relaciones entre las palabras; el tipo de disposición; el número de idiomas naturales que controlan; y el tipo de utilización.

Por otro lado, Van Slype (1991: 21) explica que la tipología de los vocabularios controlados se basa esencialmente en el nivel de normalización⁵ de su terminología, que se distingue entre los lenguajes libres, los lenguajes controlados y los lenguajes codificados. Asimismo, señala que existen dos tipos principales de lenguajes documentales: 1) los lenguajes de indización o combinatorios, que permiten representar el contenido de los documentos y de las consultas de forma

⁵ En este trabajo entendemos por normalización el proceso de convertir o convertirse en norma, entendida como patrón, estándar o forma de referencia.

analítica; y 2) los lenguajes de clasificación⁶, utilizados principalmente para representar este contenido de manera sintética.

Por su parte, Vizcaya Alonso (1995: 34) señala que los vocabularios controlados se pueden dividir en *precoordinados* y *postcoordinados*, a partir del criterio de coordinación que se establece entre los términos que lo componen. Siendo los *precoordinados* los que se estructuran tanto en esquemas y sistemas jerárquicos como en esquemas alfabéticos de materias y los *postcoordinados* constituyen “un lenguaje documentario de tipo más balanceado en cuanto al nivel de precisión que proporciona su uso, dado fundamentalmente a través de los términos que pueden ser combinados con posterioridad a su introducción en el sistema y pueden ser localizados a través de su índice alfabético sin perder la relación que dentro del sistema temático tenga ese término con otros a través de su artículo léxico y de su índice sistemático”.

Para Cleveland y Cleveland (2001: 36) un vocabulario controlado se caracteriza por:

1. Representar la estructura conceptual general de un área temática y presentar una guía tanto para el indizador como para el usuario que busca recuperar la información.
2. Reflejar de una manera precisa el vocabulario de una disciplina y el uso técnico que el propio usuario le da, pues los términos se derivan del vocabulario de uso.
3. Emplear un número considerable de frases pre-coordinadas para reducir al mínimo los errores de sintaxis, que pudieran provocar el significado contrario.
4. Ofrecer un vocabulario normalizado al controlar los sinónimos y cuasi-sinónimos con el propósito de aumentar la consistencia. Siendo ésta una función principal, ya que brinda un mecanismo que asegura que solamente un término de una lista de términos similares sea utilizado al indizar un concepto dado.
5. Definir términos ambiguos cuando es necesario.

⁶ Para Gil Urdiciain (1997: 92) el lenguaje de clasificación se define “...como aquél fundado en la representación estructurada de una o varias áreas del conocimiento, en clases en las que los conceptos y las posibles relaciones existentes entre ellos se representan abreviadamente por medio de signos alfabéticos, numéricos o alfanuméricos”. Por su parte Maniez (1992: 19) señala que “Los lenguajes de clasificación (o *clasificaciones documentales*, o también *clasificaciones bibliográficas*) ...permiten representar de manera sintética el contenido de un documento y reunir los libros en los estantes por la afinidad de su contenido”.

6. Mostar relaciones horizontales y verticales entre los términos, a través de referencias cruzadas.

Por otro lado, los vocabularios controlados, en razón de su propia naturaleza, no pueden ser redundantes. La eficacia de un vocabulario controlado aumenta a medida que la reiteración de términos es eliminada o controlada; además, han de ser unívocos, por ello se llaman controlados, pues no se pueden permitir la ambigüedad del lenguaje natural. Cabe destacar que una de las cualidades del vocabulario controlado es su control sobre la homonimia, la cual está relacionada con la precisión y el ruido, y la sinonimia, que está relacionada con el factor de acierto⁷ y el silencio.

La tasa de acierto se refiere al porcentaje de documentos relevantes recuperados durante la búsqueda en relación al número total de documentos relevantes incluidos en la base de datos. Por tasa de precisión debemos entender el porcentaje de material relevante recuperado durante la búsqueda, en relación al número total de documentos incluidos en la base de datos. El ruido se refiere a los documentos que se recuperan y que no son relevantes a la búsqueda que realiza el usuario y el silencio a los documentos que existen en la base de datos, que interesarían al usuario que busca información y que, sin embargo, no se recuperan porque no ha habido coincidencia entre los términos que empleó el indizador y los que utiliza el usuario durante la búsqueda de información.

Chowdhury (2004: 248) propone las siguientes fórmulas para el cálculo de la tasa de acierto y la tasa de precisión:

$$\text{Acierto} = \frac{\text{Número de documentos relevantes recuperados}}{\text{Número total de documentos relevantes contenidos en la colección}} \times 100$$

$$\text{Precisión} = \frac{\text{Número de documentos relevantes recuperados}}{\text{Número total de documentos recuperados}} \times 100$$

⁷ Si bien algunos autores lo interpretan como “exhaustividad” o “llamado”, para la traducción del término “recall” hemos recurrido al *Terminology of Documentation*, que maneja el término “recall ratio” o “recall factor”, y lo traduce como “tasa de acierto”.

En síntesis, los vocabularios controlados se caracterizan por realizar una condensación y una simplificación del lenguaje natural, manteniendo en lo posible la riqueza de la información original en su propio contenido y sus posibilidades de asociación, además de ser de fácil manejo, tanto para los indizadores como para los usuarios.

1.3 Utilidad

La conveniencia de los vocabularios controlados radica en que éstos expresan conceptos empleando un conjunto normalizado de términos. Sin embargo, como lo señala Kline (2002: 253), los usuarios a menudo tienen problemas, por un lado, para identificar el término apropiado asignado a un concepto y, por otro, para discernir cómo ese concepto está relacionado con otros. Para contrarrestar esta dificultad Lee-Smeltzer, (2000: 213) menciona que los bibliotecólogos, los catalogadores específicamente, han desarrollado desde hace tiempo la habilidad de organizar la información almacenada en medios físicos para ponerla a disposición de una manera coherente y consistente por medio del catálogo. Según la autora, la coherencia y la consistencia se alcanzan a través del establecimiento de reglas de catalogación y el uso de un vocabulario controlado.

Por su parte Gil Urdiciain (1998) concluye que “el lenguaje controlado consigue más bajos niveles de ruido que el libre al tiempo que puede llegar a aportar tan altos índices de exhaustividad⁸ como aquél; igualmente, ...se puede concluir que el lenguaje controlado neutraliza las deficiencias del libre y viceversa y, por tanto, ambos son complementarios”. La misma autora argumenta que el control del vocabulario es factor determinante del éxito en el proceso de recuperación de información para lograr altos niveles de precisión, por lo que el vocabulario controlado consigue más bajos niveles de ruido que el libre, particularmente cuando los títulos de los documentos no son representativos de sus contenidos. Razón por lo cual podemos deducir que uno de los aspectos más importantes

⁸ Según la autora, la exhaustividad mide la capacidad del sistema para recuperar los documentos útiles (relevantes para el usuario), mientras que la precisión mide la habilidad para rechazar el material no relevante. Por lo tanto, y tomando en cuenta estos índices, la flexibilidad de un sistema se podrá valorar en función de su capacidad para adaptarse a las necesidades de exhaustividad y precisión de los usuarios.

concierno a cómo están representados los documentos en un sistema documental, pues es de suponer que ni las más sofisticadas técnicas de búsqueda pueden superar los problemas derivados de una pobre representación, de naturaleza lingüística o temática, de los documentos.

Al respecto, Martínez Arellano (2001) comenta que una indización efectiva involucra a los vocabularios controlados, pues considera que éstos son herramientas importantes de recuperación temática. De igual manera, hace notar que el uso de palabras libres dentro del proceso de búsqueda conlleva a una recuperación de una gran cantidad de información pero su precisión disminuye. Es decir, los lenguajes libres provocan que en ocasiones el usuario obtenga una gran cantidad de información, pero sin duda mucha de ella le es irrelevante. La solución a este problema es el uso de vocabularios controlados. Así lo apunta el mismo autor cuando menciona que “Aunque el uso de los lenguajes libres es una opción fácil y barata para indizar recursos electrónicos, las bibliotecas y otras unidades de información tendrán que utilizar lenguajes controlados para almacenar y recuperar la información precisa que satisfaga las necesidades de los usuarios. Los lenguajes controlados continuarán jugando un importante papel en la organización de la información”.

De igual manera, Cleveland y Cleveland (2001: 35) señalan que el utilizar un conjunto de palabras o frases de manera incontrolada para indizar y buscar información nos lleva a un esfuerzo inútil y a una búsqueda muy poco fructífera debido a la enorme posibilidad de elección de términos. Añaden que el uso inconsistente de palabras es el mejor camino para el fracaso en la búsqueda de información por la simple razón que la posibilidad de que los usuarios empleen el mismo término utilizado por el indizador o el autor del documento es mínima, sin tomar en cuenta que en muchos casos el usuario puede elegir los términos usados por el indizador pero con una comprensión diferente de su significado.

Para evitar estas confusiones existen ciertos principios para la selección de términos a incluir en un vocabulario controlado. Svenonius (2003: 824) hace notar

que todo vocabulario controlado debe contar con una garantía literaria, (*literary warrant*), una garantía del usuario (*user warrant* o *common usage*) y una justificación de los expertos en el tema (*scholarly usage*).

El American National Standards Institute, a través de la National Information Standards Organization⁹ (ANSI/NISO Z39.19-2003: 7) define *literary warrant* como

“words and phrases drawn from the literature of the field should determine the formulation of descriptors. When two or more variants have literary warrant the most frequently used term should be selected as the descriptor”.

En otras palabras, el principio señala que el vocabulario controlado debe derivar de la literatura que contiene el vocabulario a ser controlado. En ocasiones la garantía literaria se rige por la frecuencia con la que se presenta dicho término.

La misma norma (2003: 38) define *user warrant* como

“justification for the representation of a concept in an indexing language or for the selection of a preferred term because of frequent requests for information on the concept or free-text searches on the term by users of an information storage and retrieval system”.

Es decir, es recomendable incluir el término que con más frecuencia emplean los usuarios en la búsqueda de información.

Por otro lado, Svenonius (2003: 824) señala que al construir un vocabulario controlado la opinión de los expertos en la materia (*scholarly usage*) con frecuencia se toma más en cuenta que las garantías arriba mencionadas, lo que no es reprochable siempre y cuando también se considere a los usuarios que realmente utilizarán este vocabulario controlado, ya sea tesauros, listas de autoridad, encabezamientos de materia, etc., para realizar sus búsquedas. La misma autora explica que un vocabulario controlado que ignora las expectativas lingüísticas puede resultar muy caro y difícil de usar.

⁹ La NISO es una institución que desarrolla y publica normas técnicas en el ámbito de la gestión de información para Estados Unidos.

Sin duda la utilidad de los lenguajes documentales ha sido probada en diversos estudios¹⁰ y para Svenonius (2003: 822) no es aventurado concluir que un sistema que emplea un vocabulario controlado es mejor que uno que no lo utiliza. Además menciona que existen situaciones en donde el acceso a través de una simple tecla al texto completo puede representar la alternativa más rentable. Así como también hace notar que el propósito de todo vocabulario controlado es optimizar la capacidad de precisión y acierto del lenguaje de indización. Es decir, la tasa de precisión y la tasa de acierto son los objetivos principales de cualquier lenguaje documental.

En cuanto a esta última afirmación, Chu (2003: 191) señala que además de las tasas de acierto y de precisión se deben tomar en cuenta factores como la interoperabilidad, el costo y la velocidad de recuperación.

A modo de síntesis podemos decir que los vocabularios controlados deben ser una fuente de terminología normalizada consensuada, que permite organizar y recuperar la información de la forma más amplia posible. El analista, al indizar un documento, sirve de intermediario entre el autor de dicho documento y sus lectores eventuales. Como Maniez lo señala, la razón de ser de los vocabularios controlados es recuperar fácil y rápidamente la información.

Sin olvidar que el vocabulario controlado no deja lugar para la reiteración de términos. Su eficacia aumenta a medida que la sinonimia (silencio) y la homonimia (ruido) son eliminadas o controladas. Es decir, el vocabulario controlado ha de ser unívoco, pues no puede permitirse la ambigüedad del lenguaje natural. Por otro lado, será necesario darle a conocer al usuario los términos del vocabulario controlado con objeto de que utilice estos términos en su búsqueda y pueda recuperar los documentos en el menor tiempo posible, eliminando al máximo el ruido y el silencio.

¹⁰ Véanse por ejemplo los trabajos de Raya Fidel (1992), Borlund e Ingwersen (1997), Gil Urdiciain (1998) o Harrison (1998).

2. La indización y recuperación de la información.

2.1 Indización

2.1.1 Definición

Antes de entrar en materia, consideramos pertinente aclarar que si bien muchos autores manejan la indización y la elaboración de resúmenes como actividades indisociables, para fines de este trabajo sólo nos enfocaremos a la indización. Por otro lado, cuando hablemos de indizar no nos referiremos a la elaboración de índices¹¹, nos referiremos, como lo señala Naves (2002: 119) al “proceso de identificar y organizar los asuntos de los documentos es conocido como indización, a la que se considera un proceso básico de la recuperación de información”. A este respecto, Lancaster (2003) coincide y puntualiza que la indización es el factor central para el acceso a la información. Es decir, la indización considerada como el proceso intelectual fundamental para toda recuperación de la información. Siendo el indizador el responsable del procesamiento técnico del acervo de documentos.

Chaumier (1988: 22-23), siguiendo los principios del Sistema Mundial de Información Científica y Tecnológica de la UNESCO (UNISIST), define a la indización, por su proceso, como una operación que consiste en describir un documento con la ayuda de la representación de los conceptos que contiene el propio documento. Es decir, los conceptos son traducidos empleando el vocabulario controlado después de haberlos extraído del documento por medio de un análisis. Señala, además, que la traducción en lenguaje documental se hace gracias a las herramientas de indización como son los tesauros, los encabezamientos de materia, etc. Asimismo, destaca que la indización interviene

¹¹ La *International Encyclopedia of Information and Library Science* (1997), define al índice como 1) una lista alfabética de temas, nombres de personas, lugares, etc., mencionados en un libro o series de libros, indicando en qué lugar o lugares aparecen en él (los) documento(s) fuente(s), generalmente por número de página (algunas veces con un símbolo adicional denotando una posición en la página) y algunas veces por sección o número de entrada. 2) En el uso cotidiano, un índice se refiere a una herramienta para localizar los documentos en una biblioteca. Se utiliza en ocasiones como sinónimo de catálogo. Sin embargo, aunque los principios de análisis empleados son muy similares, una entrada en un índice simplemente ubica un tema, mientras que una entrada en un catálogo también incluye una especificación descriptiva del documento que tiene que ver con el tema.

no solamente en la fase de entrada de la información en el sistema documental¹², sino también en la salida con la formulación de las preguntas para la búsqueda y recuperación de la información, por lo que su finalidad es permitir una búsqueda eficaz.

Para Cabré (1999: 237) la indización

“es en el fondo una operación terminológica, es una actividad de detección de la terminología más indicativa del contenido documental, por la que se atribuye un valor de contenido a determinadas unidades y expresiones del texto, que son las que servirán de anclaje en la recuperación de la información”

La norma ANSI/NISO Z39.19 (2003: 36) define a la indización como

“An operation intended to represent the results of the content analysis of a document by means of a controlled indexing language or by natural language”.

Según la misma norma, por lenguaje de indización (*indexing language*) debemos entender que es un vocabulario controlado o un sistema de clasificación tomando en cuenta las reglas para su aplicación. Es decir, un lenguaje de indización se utiliza para representar los conceptos de los documentos, lo que, a través de un sistema de recuperación, facilita recuperar dichos documentos de un sistema de almacenamiento de información.

Por su parte, la Organización Internacional del Trabajo (OIT), en sus *Técnicas de indización de la OIT – Módulo de entrenamiento*¹³, en donde se explican las técnicas utilizadas en la preparación de registros bibliográficos para la base de datos de este organismo internacional, señala que el objetivo de la indización es el de proporcionar un tema, en donde cada término o descriptor es utilizado sin ambigüedad para representar un contenido determinado, con objeto de acceder al conjunto de información y dar la posibilidad al usuario de recuperar documentos sobre temas específicos. Por lo tanto, “la eficacia en la obtención de la información depende de la calidad y de la consistencia en la indización”.

¹² El sistema documental es un proceso de soporte permanente, en donde el documento satisface una necesidad de información específica.

¹³ <http://www.ilo.org/public/spanish/support/lib/indexati/unit2/unit2.htm#unit2>. [Consulta: 09.01.06]

A partir de las características que hemos resaltado de las definiciones anteriores, podemos decir que la indización es un proceso de traducción de las expresiones de un autor hacia el público en la ausencia total de interacción con este último. En otras palabras, la traducción del lenguaje natural a través del vocabulario controlado comienza por un trabajo de neutralización de las palabras extraídas del lenguaje original. Sin olvidar que la indización participa a la vez en el tratamiento del documento y en la transmisión de la información que contiene. Es decir, la indización juega sobre un plan doble, de entrada y salida, dentro de un sistema documental en donde se condiciona el valor y la eficacia.

Dado que el objeto de estudio de este trabajo es el tesauro, creemos pertinente añadir aquellas características del lenguaje científico y técnico señaladas por Rotislav Kocourek (1991, mencionado por Holzem: 139)

“Elle [la langue de spécialité] a tendance à définir ses concepts, à contrôler la polysémie et l’homonymie, à supprimer les synonymes, à simplifier et à mieux délimiter les moyens syntaxiques, à neutraliser l’affectivité, la subjectivité et les fonctions conatives [...]. La langue de spécialité peut, en conséquence être l’instrument de sa propre formation, de son propre fonctionnement, de sa précision et de son évolution”.

En otras palabras, el texto especializado exige un grado de conocimiento sobre el tema por parte del receptor y está orientado a la comunidad científica que conoce el tema y la terminología científica y tecnológica. Razón por la cual los vocabularios controlados, como lo señala Van Slype (1991: 24), “deben evolucionar rápidamente para adaptarse al desarrollo de la terminología científica y técnica”.

2.1.2 Metodología

La indización requiere de un procedimiento determinado, como lo señalan diversos autores. Para Hudon (1998: 78) la indización de un documento se hace en dos etapas. En la primera, análisis del contenido, se lleva a cabo la identificación y la

selección de los conceptos¹⁴ que serán representados. En la segunda etapa se realiza la traducción en lenguaje de indización. Es decir, un símbolo, una palabra, un término o una expresión son escogidos para representar de manera clara y concisa cada uno de los conceptos seleccionados en la etapa del análisis. Para esta autora, los términos de indización pueden ser controlados, presentados en una forma más o menos distante del lenguaje natural o simplemente ser palabras sacadas directamente de la terminología de una disciplina.

Por su parte, la OIT señala que el método de indización consiste básicamente en cuatro pasos:

1. Examinar el documento para determinar el tema contenido, pues es esencial adquirir una buena comprensión del tema principal.
2. Anotar los conceptos clave, tomando en cuenta que la importancia relativa de cada concepto debe ser juzgada de manera tal que no se dé mayor énfasis a partes del documento que no tienen relevancia. Cabe recordar que una tendencia a “sobreindizar” resulta generalmente en la obtención de un mayor número de documentos, pero al mismo tiempo disminuye el número de documentos relevantes para la búsqueda del usuario (tasa de precisión y tasa de acierto).
3. Convertir los conceptos en descriptores, utilizando el vocabulario controlado, ya que éstos facilitarán al usuario la recuperación de la información.

A modo de paréntesis, consideramos pertinente recalcar lo que menciona la OIT en esta etapa

“El uso de un vocabulario controlado es necesario. Los términos utilizados en lenguaje normal pueden tener distintos significados o connotaciones diferentes de acuerdo al contexto en el que se encuentran. Para la indización solamente un significado es aceptable para cada término. Utilice siempre el descriptor más específico para expresar un determinado concepto. Cuando no se encuentra un descriptor apropiado para el concepto, será necesario buscar sinónimos y expresiones más generales. Por otra parte, cuando no se encuentre un descriptor

¹⁴ El concepto es un elemento del pensamiento, una representación mental de seres o de cosas, de cualidades, de acciones, de relaciones, de situaciones, de propiedades, etc. Todo concepto posee un cierto número de características esenciales que lo permiten definir. El concepto existe en el abstracto, pero se exterioriza en un término que permite hacerle referencia.

que corresponda al tema principal del documento, es recomendable emplear el o los términos disponibles que obviamente sean más apropiados”.

4. Redacción de la cadena de descriptores, ordenándolos en descriptores primarios, descriptores secundarios y descriptores de forma¹⁵.

Para resumir, podemos mencionar las etapas para la indización que proponen Guinchat y Menou (1990: 179):

“el conocimiento previo del documento; la determinación de su tema principal; la identificación de los elementos del contenido que deben describirse y la extracción de los términos correspondientes; la verificación de la pertinencia de los términos elegidos; la traducción de los términos del lenguaje natural a los términos correspondientes del lenguaje documental; la verificación de la pertinencia de esta descripción; la formalización de la descripción”.

2.1.3 Evaluación de la indización

Los mismos autores señalan que la calidad de una indización se aprecia según varios criterios (183):

“La exhaustividad¹⁶, o sea, que todos los temas, objetos y conceptos que encierra el documento estén bien determinados en la indización; la selectividad, es decir, que sólo se retienen las informaciones que interesan a los usuarios; la especificidad, según la cual el indizador debe buscar y utilizar una descripción que traduzca fielmente el contenido del documento, rechazando los descriptores demasiado generales o demasiado particulares con relación a los conceptos que expresa el documento; la uniformidad (u homogeneidad), consistente en que todos los indizadores y usuarios deben describir el mismo documento o documentos sobre el mismo tema de la misma manera”.

¹⁵ Los descriptores primarios son aquellos que reflejan el contenido temático más importante del documento como un todo; los secundarios se emplean para reflejar temas de menor importancia o que han sido tratados en algunas secciones del documento; los de forma se utilizan cuando se trata de publicaciones de naturaleza muy general, como son los materiales de referencia de amplia cobertura temática.

¹⁶ Chaumier (1988: 24) se refiere a la exhaustividad como sinónimo de profundidad de la indización, misma que define como el número promedio de descriptores asignados para cada documento, el cual permite controlar la calidad de la indización. Asimismo, señala que la profundidad aumentará conforme se enriquezca el vocabulario controlado y los indizadores tengan un mejor manejo de dicho vocabulario.

Por último, mencionan que la verificación puede hacerse comparando la indización con el documento original o con los términos que se extrajeron primero, haciendo algunas preguntas de ensayo para ver si el documento es fácil de encontrar, buscando informaciones sobre algunos de los temas que trata y comparando, si se considera oportuno, varias indizaciones de documentos semejantes. Resaltan que la revisión de la indización permite asegurarse de que ésta es de buena calidad y conforme a las reglas previamente establecidas, recordando que desde el punto de vista de la forma, será necesario verificar si las reglas de escritura se han tenido en cuenta, si los descriptores utilizados existen en realidad y si están escritos correctamente.

Cleveland y Cleveland (2001:179) también destacan la importancia de la evaluación de la indización. Consideran que una indización de mala calidad trae como consecuencia un fracaso en la recuperación y, para ponerlo en sus propias palabras, “unretrieved information is the same as nonexistent information”.

Para estos autores una buena indización permite al usuario recuperar la información sin dificultad y con la mayor tasa de acierto y precisión. Es decir, en todos los casos la indización debe permitir recuperar sólo aquellos documentos relevantes para el usuario, razón por la cual la evaluación de la indización comienza con la definición de los usuarios. Por lo tanto, para llevar a cabo la evaluación de la indización se deben establecer, en una primera instancia, ciertos criterios sobre las necesidades y demandas de los usuarios. Recordando que el objetivo de una buena indización es ayudar al usuario a encontrar la información que requiere en el menor tiempo y esfuerzo posibles, pero sobretodo, la información que ha de recuperar debe ser relevante, de ahí que la relevancia sea considerada la esencia de la indización.

En resumen, la indización es la operación más significativa de todo sistema documental y se refiere al proceso de descripción y representación del contenido de un documento, mediante un número limitado de conceptos extraídos del texto del propio documento a través del empleo de vocabularios controlados. Esos

conceptos deben ser los que más apropiadamente caractericen al documento, de modo que facilite su futura recuperación temática.

2.2 Recuperación de la información

El término “recuperación de información” fue acuñado por Calvin Mooers a finales de la década de 1940 y se refiere a la búsqueda de información ya sea en documentos, bases de datos, hipertextos, etcétera.

2.2.1 Sistemas de recuperación de información

En la sociedad actual es sabido que la gran abundancia de información y su representación en diferentes formatos pueden hacer que las tareas de recuperación sean más complejas, lo que refleja que la recuperación de la información sea un campo interdisciplinar, en donde es necesaria una colaboración entre los diferentes enfoques para demostrar su utilidad y eficacia. Sin duda son muchas las disciplinas directamente relacionadas (Bibliotecología, Lingüística, Informática, Lógica, Psicología, por mencionar sólo algunas) y todas comparten el mismo objetivo: facilitar la búsqueda y obtención de la información relevante que satisfaga las necesidades de información de los usuarios.

De acuerdo al *Diccionario enciclopédico de ciencias de la documentación*, los sistemas de recuperación de información (SRI) son “sistemas diseñados para almacenar y buscar información”. También menciona que si bien han existido desde siempre sistemas manuales destinados a estas funciones, generalmente se utiliza este término para “designar a los sistemas automatizados de recuperación de información soportados por sistemas informáticos”.

García Marco (1997: 250) apunta que “recuperar” significa volver a conseguir algo y, según este autor, para lograrlo se requiere de seis condiciones: 1) necesidad de recuperar alguna información; 2) conocimiento de su existencia (identificación); 3) descripción del documento que se desea recuperar; 4) conocimiento de su localización; 5) representación ordenada del espacio de recuperación; y 6) la recuperación propiamente dicha. En otras palabras, la recuperación no implica

solamente el conocimiento del documento que se quiere recobrar, significa también el conocimiento del espacio (ya sea físico o intelectual) donde se realizará la búsqueda.

El mismo autor señala que los SRI son conjuntos de informaciones que han sido convertidos en espacios de recuperación ordenados de tal forma que facilitan la búsqueda; además están dotados de memoria y de procedimientos de selección, a través de los cuales el usuario expresa lingüísticamente una petición concreta, en los términos y relaciones admitidos por el vocabulario controlado, y solicita al sistema que realice la búsqueda. Hace notar que, en teoría, el usuario debería recuperar todos los documentos contenidos en el SRI que sean capaces de satisfacer su necesidad de información. Sin embargo, se sabe que en la práctica esto no ocurre así debido al ruido y al silencio producidos, porque la petición o los propios documentos no han sido adecuadamente representados en la indización o en la búsqueda.

Por su parte, Cleveland y Cleveland (2001: 23-24) mencionan que el objetivo común de todos los SRI es proporcionar la coincidencia posible más cercana entre la descripción que hace el usuario de un tema y la descripción del mismo tema en la base de datos del sistema.

Los mismos autores también apuntan que actualmente un área de investigación en recuperación de la información está enfocada en hacer los ajustes necesarios a los SRI utilizando la retroalimentación del usuario. Por lo tanto, para ellos (2001: 28) “Feedback is the crucial and central element in an information retrieval system”¹⁷. A este respecto López Huertas (1998) coincide cuando menciona que “el diseño de un lenguaje documental que no tenga en cuenta al usuario va a estar incompleto y dará origen a un lenguaje deficiente en sus prestaciones a los SRI”.

¹⁷ Como hacen notar Mandala *et al.* (2000: 363) “past research has verified the effectiveness of relevance feedback, but it puts a burden on users to a certain extent. Furthermore, if a user is not familiar with the vocabulary of a document collection, it is difficult to obtain good expansion terms, unless the system can suggest terms to the user”.

En efecto, si bien los diversos experimentos llevados a cabo a partir de la década de 1940, como por ejemplo los proyectos Cranfield, han sido un punto de partida para el avance de los vocabularios controlados, López Huertas (1998) hace notar que al dejar de lado al usuario y seguir el modelo de comunicación de Shannon y Weaver¹⁸ los SRI se concibieron al margen del usuario por considerarlo un elemento pasivo y sin interés para el diseño de los vocabularios controlados. Sin lugar a dudas esta tendencia incide negativamente en el diseño y en la construcción de estos vocabularios porque, como lo apunta la autora, “supone ignorar uno de los pilares básicos de los SRI que son los usuarios”. Si además tomamos en cuenta que los usuarios son la razón de ser de estos sistemas, todavía se hace más evidente la necesidad de su presencia, tanto en su elaboración como en su evaluación.

Por lo tanto, podemos suponer que si no se representa en los SRI lo que es relevante para el usuario, difícilmente se podrán satisfacer sus necesidades de información con pertinencia y exhaustividad. López Huertas (1998) explica al respecto que cualquier método que persiga la representación del usuario en los vocabularios controlados y en los SRI debe responder a la siguiente pregunta: ¿Qué aspectos son susceptibles de ser representados? Y para la autora son dos los aspectos básicos a considerar: 1) los subjetivos, es decir la intencionalidad, (propósitos concretos que llevan al usuario a buscar un determinado tipo de información); y 2) los objetivos, que se identifican con el tema de una determinada necesidad de información del usuario.

Tomando en cuenta estas consideraciones podemos decir que es necesario establecer una dinámica en la que el usuario se reconozca en sus necesidades de información, lo que lo lleva a interactuar mejor con el sistema y a establecer una

¹⁸ En este modelo, eminentemente tecnológico, se introducen elementos que no se habían abordado con anterioridad como son el transmisor, los ruidos, el receptor. Este modelo surge en el campo de la ingeniería de las comunicaciones con el principal objetivo de optimizar la transmisión de mensajes aumentando la precisión con que se pueden transmitir los símbolos de la comunicación sin perder información y disminuyendo posibles distorsiones. La comunicación, que llega a identificarse con la transmisión del mensaje, se considera un éxito en la medida en que el mensaje que recibe el destinatario se asemeja al mensaje que la fuente transmite. Para lo cual, será necesario minimizar los posibles ruidos e interferencias. Shannon introduce la idea de redundancia, que él concibe como la forma de luchar contra los ruidos.

comunicación sistema-usuario más flexible y transparente, aumentando los niveles de relevancia en la recuperación.

2.2.2 Evaluación de los SRI

Para Salvador y Arquero (2004) la recuperación de la información es un campo donde se realiza una actividad importante de práctica profesional y de investigación científica. Esta última ha estado dirigida hacia el diseño de SRI más eficaces, dando lugar a diversas teorías, modelos y experimentos en los que la evaluación ha ocupado un papel central. Como lo señala Blair (1990, citado por Martínez y Rodríguez, 2004: 154), “la propia naturaleza de los SRI propicia su necesidad crítica de evaluación, justo como cualquier otro campo de trabajo que aspire a ser clasificado como campo científico”.

Como se mencionó anteriormente, son muchos los experimentos, pruebas e investigaciones llevadas a cabo en el campo de la recuperación de la información y también muchas las críticas y discusiones sobre los resultados obtenidos, derivadas, en su mayoría, por la diferente metodología utilizada. Al respecto Martínez y Rodríguez (2004) mencionan que si bien un SRI puede ser evaluado mediante diversos criterios, éstos deben tomar en cuenta la eficacia en la ejecución (tiempo para realizar una operación); la eficiencia del almacenamiento; y la efectividad de la recuperación, basada en la relevancia de los documentos recuperados.

Salvador y Arquero (2004) proponen que considerar a los usuarios como punto de partida en la evaluación implica una concentración sobre el significado, semántica y aspectos pragmáticos de la información, existiendo dos áreas principales de interés: por un lado, el estudio del significado que los usuarios dan a la información en el contexto de búsqueda y uso de la información; por el otro, las implicaciones en el diseño de los sistemas de información.

Chowdhury (2004: 243) menciona a la eficacia y a la eficiencia como los dos parámetros básicos para evaluar el desempeño de cualquier SRI. Por eficacia se

refiere al nivel que alcanza el sistema en relación a los objetivos establecidos. Es decir, la eficacia de un SRI se puede medir por la cantidad de documentos relevantes recuperados. La eficiencia está ligada a la economía del sistema para alcanzar sus objetivos y se puede medir a través del costo mínimo requerido para que el SRI funcione eficientemente, debiendo calcularse este costo indirectamente y el cual comprende factores como el tiempo de respuesta, el esfuerzo del usuario, el gasto financiero por búsqueda, entre otros.

Cyril Cleverdon (1966, citado por Chowdhury, 2004: 246) identifica seis criterios para la evaluación de los SRI, los cuales comprenden:

1. *acierto*, la habilidad del sistema para presentar todos los documentos relevantes;
2. *precisión*, la habilidad del sistema para presentar sólo aquellos documentos que son relevantes;
3. *tiempo de respuesta*, el intervalo promedio entre la solicitud de búsqueda y la respuesta;
4. *esfuerzo*, tanto intelectual como físico, del usuario;
5. *forma de presentación*, de los resultados de la búsqueda, lo que puede repercutir en la habilidad del usuario para hacer uso de la información recuperada;
6. *cobertura de la colección*, el alcance del sistema para incluir información relevante.

Por su parte Salvador y Arquero señalan que las variables que se han investigado en el proceso de recuperación de información, se pueden agrupar en: bases de datos; representación de la información; tipo de usuarios; consultas y sentencias de búsqueda (entendiéndose por “consulta” la expresión, escrita u oral, de la necesidad de información del usuario, y por “sentencia de búsqueda” la traducción de la consulta al lenguaje del sistema); proceso de búsqueda; y evaluación de la recuperación (siendo las medidas tradicionales el cálculo de la tasa de acierto y la tasa de precisión).

No cabe duda que uno de los principales problemas a los que se enfrenta cualquier usuario es cómo localizar información relevante ante el exceso de información existente. Por lo tanto, la investigación y la evaluación en SRI deben ir dirigidas a dar respuesta a este problema, tomando en cuenta que: a) el acceso a la información se ha generalizado y es el propio usuario quien busca la información, por lo que es necesario diseñar interfaces de búsqueda más amigables, además de que resulta indispensable dar cabida a la terminología utilizada por el usuario en la búsqueda de la información; y b) que el concepto de información se amplía y ya no sólo se trata de documentos impresos, sino también de imágenes y sonido, por lo que debe de investigarse cómo recuperar estas formas de información (recuperación multimedia)¹⁹.

En síntesis y por todo lo anterior, se deduce la importancia que reviste la calidad del vocabulario controlado empleado, la calidad de la indización y, por último, y no por eso menos importante, la capacidad de seleccionar los temas adecuados para formular una estrategia de búsqueda. Sin olvidar que entre las ventajas que se pueden reconocer en el uso de un vocabulario controlado, está su capacidad para adaptarse con facilidad a las necesidades de información más o menos específicas del usuario potencial.

Por otro lado, dado que el indizador debe tener siempre en mente que el propósito de la indización es la recuperación de la información y que los descriptores que se asignan a un documento deben reflejar en forma puntual los temas principales expresados, a fin de facilitar la identificación rápida y precisa de las referencias *relevantes* para un asunto determinado, consideramos oportuno abrir un paréntesis para disertar sobre la relevancia, pues consideramos importante definir este concepto, ya que está en la base del resto de las medidas que tradicionalmente se vienen aplicando en los SRI.

¹⁹ Un buen ejemplo es *Aqualingua*, diccionario multilingüe de palabras y expresiones técnicas relacionadas con los recursos hídricos. Para conocer más sobre este proyecto se puede consultar el sitio <http://scripts.oieau.fr/aqualingua/pres/> [Consulta: 19.08.05].

2.3 La relevancia

Algunos autores la consideran un elemento fundamental para la recuperación de información²⁰. Sin embargo, vale resaltar que el término “relevancia” se emplea a menudo como sinónimo de “pertinencia”, como lo hace notar Gil Urdiciain (1998)²¹. Si para precisar ambos conceptos se acude al *Diccionario enciclopédico de ciencias de la documentación* (2004) encontraremos que relevancia es definida como la

“Propiedad de una información resultado de una búsqueda que denota en qué grado es considerado adecuado para resolver una determinada necesidad de información desde un punto de vista objetivo, por contraposición a la pertinencia, que es juzgada por el propio sujeto interesado. En la evaluación de resultados de búsqueda, se mide mediante la tasa de precisión”.

Y pertinencia es la

“Propiedad de una información que denota hasta qué punto ésta resuelve, desde el punto de vista del usuario final, una determinada necesidad de información”.

Es decir, en la relevancia el juicio sobre el valor o utilidad de la información es emitido por personas desinteresadas aunque expertas, y en la pertinencia es el propio usuario quien juzga el valor de la información recuperada para sus necesidades de información.

Por otro lado, Froehlich (1994, citado por Salvador y Arquero) destaca una serie de cuestiones comunes en los análisis sobre la relevancia, que resulta interesante mencionar: 1) la incapacidad para definirla²²; 2) lo inadecuado que resulta utilizar el tema de los documentos como base para valorar la relevancia; 3) la diversidad

²⁰ Véanse por ejemplo los trabajos de T. Saracevic (1975) o S. Mizzaro (1997).

²¹ La autora señala que la relevancia es la valoración que realizan una o varias personas en relación con una determinada solicitud de información, encontrando coincidencia entre pregunta y respuesta. Pertinencia es la valoración que hace el usuario de una respuesta dada por un SRI a una necesidad concreta de información formulada por el mismo usuario. Es decir, relevancia es la medida de cómo una pregunta se ajusta a un documento y pertinencia es la medida de cómo un documento se ajusta a una necesidad de información.

²² Los autores indican que la discusión y falta de acuerdo vienen determinadas, entre otras razones, porque no existe una única relevancia. Es el concepto y la naturaleza de la relevancia, los diferentes tipos que existen y los numerosos factores que influyen en los juicios de relevancia, muchos de ellos subjetivos, lo que impide una definición formal y una de las razones de que no exista una teoría global e integradora en recuperación de la información.

de criterios en torno al usuario y que no están relacionados con el tema que afectan al juicio de la relevancia; 4) el carácter fluido y dinámico del comportamiento en la búsqueda de información; y 5) la necesidad de modelos más sólidos y eficientes en el diseño y evaluación de sistemas.

Considerando lo anterior, Martínez y Rodríguez (2004: 155) proponen que “un documento será relevante cuando el contenido del mismo posea alguna significación o importancia en relación con la pregunta realizada por el usuario, es decir, con su necesidad de información”. Sin embargo, señalan que a la hora de determinar con exactitud cuándo un documento puede ser considerado relevante o no, se presentan algunos problemas estrechamente relacionados con la naturaleza cognitiva de este proceso, el cual resumen de la siguiente manera:

1. “un mismo documento puede ser considerado relevante o no relevante por dos personas en función de su necesidad de información o grado de conocimiento de la materia. Sin contar que un mismo documento puede parecer relevante o no a la misma persona en momentos diferentes de tiempo;
2. es complicado establecer *a priori* los criterios que permitan determinar cuándo un documento es relevante o no e incluso resulta problemático explicitarlo de manera clara y concisa, siendo más fácil proceder a la determinación de la “relevancia” que explicar cómo ésta se lleva a cabo;
3. es muy aventurado calificar categóricamente un documento como relevante o no relevante, con un tema. En la realidad lo normal es encontrarnos con documentos relevantes con una materia determinada en alguno de sus apartados, pero no en el resto de sus contenidos”.

Podemos resumir lo expresado por estos autores señalando que un usuario tendrá problemas a la hora de definir qué es relevante y qué no lo es, pero tendrá pocos problemas a la hora de decidir si el documento le parece o no útil.

Por su parte, Park (1997: 339-40) establece que “Relevance is the key abstract concept in information retrieval. For effective information retrieval the system needs to retrieve documents based on the user’s notion of relevance”. De igual manera menciona que en la recuperación de la información, la relevancia de un documento es juzgada durante la solución del problema, ya que “...the role of an

information retrieval system is to make preliminary relevance judgments for the user”.

El mismo autor señala que en la recuperación de información existen tres clases de juicios de relevancia: 1) el juicio del usuario sobre lo que es relevante para él dentro de la colección de documentos; 2) el juicio del sistema sobre lo que es relevante con la representación de los documentos; y 3) el juicio del usuario sobre lo que es relevante con las representaciones de los documentos que son considerados relevantes por el sistema.

De lo anterior podemos deducir que si el SRI emplea una noción de relevancia diferente a la del usuario, la efectividad del sistema será casi nula. Por lo tanto, para que un SRI sea eficaz se requiere que tanto la representación del documento como la de la búsqueda y la noción de relevancia tengan cierta similitud²³.

Profundizando un poco más sobre el tema, Chu (2003: 189-190) menciona que la relevancia es esencialmente un juicio subjetivo dictado por diversos factores interrelacionados. Es decir, la relevancia es un concepto que no puede ser descrito explícitamente ni tampoco fácil de cuantificar. “Simply put, relevance is a property that reflects the relationship between a document and the user’s query”. Por su parte, Regazzi (citado por Saracevic, 1975) sugiere que además de tomar en cuenta la relevancia, los documentos deben ser valorados a través de atributos más específicos, incluyendo exactitud, contenido, cobertura, claridad, actualidad y estar abiertos a la aceptación de sugerencias.

A modo de conclusión podríamos pensar que, a pesar de todo, la relevancia no refleja el grado de éxito conseguido en encontrar la información que requiere el usuario, ya que sólo relaciona los resultados de la búsqueda con la petición y será solamente el usuario quien podrá juzgar si la información obtenida en respuesta a una petición es pertinente, ya que únicamente él puede conocer su necesidad de información. Sin olvidar que su juicio puede variar según la situación.

²³ Saracevic (1975: 154) menciona que “A system can only answer questions. It can only guess what the information need is”.

3. El tesoro

3.1 Definición²⁴

El término tesoro tiene su origen etimológico en la palabra latina *thesaurus*, que, a su vez, tiene su origen en la palabra griega *θησαυρός*. En ambos casos, el significado era tesoro o repositorio de palabras. Los tesauros, como herramientas lingüísticas para indizar documentos, aparecieron en los últimos años de la década de 1950.

La Organización Internacional para la Normalización, en su *ISO 2788-1986 (E)* define al tesoro como

“The vocabulary of a controlled indexing language, formally organized so that the *a priori* relationships between concepts (for example as “broader” and “narrower”) are made explicit”.

Martínez de Sousa (1989) proporciona varias definiciones, entre las que destacan

“... 2. Lista alfabética de palabras estándares utilizadas para la clasificación de la documentación. 3. Diccionario que muestra la equivalencia entre términos del lenguaje natural y términos normalizados y preferentes del lenguaje documental, así como las relaciones semánticas que existen entre otros términos. ... 5. Vocabulario, generalmente especializado, coherente y limitado de palabras, dotado de sus correspondencias semánticas, elegido de manera que pueda representar todas las nociones que figuran en un texto dado y así servir a la informática y al establecimiento de índices”.

Para Van Slype (1991: 23) un tesoro

“es una *lista estructurada de conceptos*²⁵, destinados a representar *de manera unívoca* el contenido de los documentos y de las consultas dentro de un sistema documental determinado, y a ayudar al usuario en la indización de los documentos y de las consultas; los conceptos son extraídos de una lista finita, establecida *a priori*; sólo los términos que figuran en esta lista pueden ser utilizados para indizar

²⁴ La citación de definiciones no es exhaustiva, y dicha selección se encuentra guiada por las características que a nuestro criterio son relevantes en el concepto de tesoro como herramienta documental.

²⁵ Las cursivas son del autor.

los documentos y las consultas; la ayuda al usuario la proporciona la estructura semántica del tesoro: fundamentalmente las relaciones de equivalencia, de jerarquía y de asociación”.

Por último, la NISO (ANSI/NISO Z39.19-2003 : 38) explica que

“A thesaurus is a controlled vocabulary arranged in a known order and structured so that equivalence, homographic, hierarchical, and associative relationships among terms are displayed clearly and identified by standardized relationship indicators that are employed reciprocally”.

La misma norma especifica que los propósitos de un tesoro son:

1. *Traducción*: proveer un medio para traducir el lenguaje natural de los autores, indizadores y usuarios en un vocabulario controlado utilizado para indizar y recuperar la información.
2. *Consistencia*: promover la consistencia en la asignación de términos.
3. *Indicación de relaciones*: indicar relaciones semánticas entre los términos.
4. *Recuperación*: servir como una ayuda de búsqueda en la recuperación de documentos.

Chaumier (1988: 71) hace notar que desde el punto de vista de su función, un tesoro es un instrumento de control terminológico utilizado para convertir a un lenguaje documental el idioma natural empleado en los documentos. Y desde el punto de vista de su estructura, es un vocabulario controlado y dinámico de términos que guardan entre sí relaciones semánticas y genéricas, y por lo general se aplica a un dominio particular del conocimiento. Y como el mismo autor menciona, el tesoro es una herramienta terminológica y no un fin en sí.

De todo lo anterior podemos resumir que un tesoro es un vocabulario estructurado de términos que guardan entre sí relaciones semánticas de equivalencia, jerárquicas y asociativas, cuya finalidad es convertir el lenguaje natural de los documentos en un lenguaje controlado que represente de manera unívoca el contenido de los mismos, garantizando así la calidad en los procesos de indización y recuperación de la información. Cabe destacar que a diferencia de

otros vocabularios controlados, los tesauros suelen ser lenguajes especializados, circunscritos a una ciencia o conjunto de ciencias en particular.

3.2 Estructura y presentación

Si bien cada tesoro tiene sus propias características, existen ciertas normas que se deben contemplar en su elaboración y presentación, como por ejemplo la *ISO 2788-1986(E) Documentation - Guidelines for the establishment and development of monolingual thesauri* o la *UNE 50106: 1990 Directrices para el establecimiento y desarrollo de tesauros monolingües*, elaborada por la Asociación Española de Normalización y Certificación (AENOR).

Esto es, la estructura de los tesauros está controlada por estándares internacionales que, sin lugar a dudas, han tenido que ir adaptándose a las tecnologías de información y comunicación. Como ejemplo podemos citar la norma ANSI/NISO Z39.19:2003, equivalente a la ISO-2788, misma que en el año 2003 fue revisada con el fin de adecuarla a los nuevos requerimientos electrónicos, como el incluir métodos adicionales de búsqueda y navegación, búsqueda por palabras clave, uso de metadatos²⁶, formatos de presentación, aplicaciones en Internet, etc. Dando como resultado de esta revisión una norma que permite una mayor interoperabilidad entre distintos sistemas, pues muestra cómo formular descriptores, establecer relaciones entre los términos y presentar la información ya sea impresa o en línea.

Como herramienta documental todo tesoro debe incluir una serie de elementos que sirvan para garantizar la calidad en los procesos de indización y recuperación de la información. Como se puede observar por las definiciones citadas, los elementos constitutivos de un tesoro son, por un lado, las unidades léxicas, que

²⁶ Debido a la gran diversidad y volumen de las fuentes y recursos en Internet, se hizo necesario establecer un mecanismo para catalogar, clasificar y describir los recursos presentes en la *www* con el fin de facilitar la búsqueda y recuperación de la información. Este mecanismo lo constituyen los llamados metadatos, que son datos estructurados sobre la información. Los metadatos en el contexto de la *www*, son datos que se pueden guardar, intercambiar y procesar y están estructurados de tal forma que permiten ayudar a la identificación, descripción, clasificación y localización del contenido de un documento o recurso en línea y que, por lo tanto, también sirven para su recuperación.

son los descriptores (términos preferentes) y los no descriptores (términos no preferentes), y por el otro, las relaciones semánticas, es decir las relaciones que se establecen entre los términos que son, generalmente, de equivalencia, jerárquicas y asociativas.

Los descriptores son los términos permitidos. Es decir un descriptor es una palabra o conjunto de palabras que responde a un concepto representativo de un documento y que será utilizado para la indización y la recuperación de la información. Cabe recordar que un concepto se expresa con un único término y ese término corresponde a un único concepto.

Los términos pueden ser simples, como por ejemplo *Agua*, o compuestos, como *Agua dulce*. Vale mencionar que la mayoría de los descriptores de un tesoro adquieren su verdadero sentido cuando se convierten en términos compuestos, ya que así se consigue una mayor especificidad.

La estructura sintáctica de los términos compuestos suele ser:

Sustantivo + adjetivo = *Agua potable*

Sustantivo + sintagma preposicional = *Agua de lluvia*

Sustantivo + sintagma aposicional = *Interfase hielo-agua*

Un no descriptor es un cuasi-sinónimo del descriptor o término que designa el lenguaje natural como concepto semejante al que cubre el descriptor. Al elegir un único término para eliminar la sinonimia del lenguaje natural, se rechazan los términos restantes, que serán precisamente los no descriptores. Es decir, los no descriptores no pueden ser utilizados para la indización del documento, pero cada uno de ellos reenvía a uno o varios descriptores para representar los conceptos correspondientes. La inclusión de los no descriptores en un tesoro tiene como finalidad mejorar la coherencia de la representación de los documentos o de una pregunta al enviarnos al término de indización.

Por otra parte, las normas no aconsejan la utilización de formas gramaticales como adjetivos aislados, adverbios o formas verbales no sustantivadas. Respecto

a las siglas y abreviaturas, recomiendan su uso solamente cuando éstas sean mejor conocidas que la forma completa. Por ejemplo, es recomendable utilizar FAO en lugar de Organización de las Naciones Unidas para la Agricultura y la Alimentación.

En cuanto a la utilización de singulares y plurales, las normas indican que:

- ♦ deben ir en plural los objetos contables y grupos;
- ♦ deben ir en singular nombres abstractos y fenómenos, así como disciplinas, ideologías y creencias, procesos y actividades, y objetos concretos no contables (como el agua).

En cuanto a las relaciones semánticas, destacan las siguientes:

§ *Relaciones jerárquicas*: son las relaciones que van de lo general a lo particular.

Para la representación de este tipo de relaciones se utilizan las siglas *TG* (término genérico, señala el término más amplio al que pertenece el descriptor) y *TE* (término específico, indica los términos derivados del descriptor base).

§ *Relaciones asociativas*: son las relaciones establecidas entre descriptores de distintas familias y que designan conceptos capaces de evocar mutuamente, por asociación de ideas, relaciones entre ellos.

Este tipo de relaciones se representan por lo general mediante las siglas *TR* (término relacionado).

§ *Relaciones de equivalencia*: es la relación entre los términos preferentes y los no preferentes que se refieren al mismo concepto. En otras palabras, permite descartar el uso de diferentes formas para nombrar un mismo concepto, es decir la sinonimia. Cabe señalar que dado que la sinonimia absoluta es rara, manejaremos preferentemente la cuasi-sinonimia.

Este tipo de relaciones se suelen representar mediante el término *USE* y las siglas *UP* (Usado por).

Como ya hemos mencionado, cada descriptor tiene un significado preciso. El contenido semántico de cada descriptor está indicado por medio de sus equivalencias y sus relaciones asociativas y jerárquicas. Sin embargo, se emplea una *Nota de Alcance* (NA), ya sea para precisar y/o restringir el significado de un término o para proporcionar la definición usual del mismo. Generalmente se utiliza cuando el término puede ser interpretado de modo ambiguo o equívoco. La nota de alcance también puede utilizarse para precisar el sentido del descriptor para fines de indización, por ejemplo “Úsese como descriptor para postcoordinación”.

Las formas más comunes de presentación del tesoro son la temática, la alfabética y la permutada; en ocasiones también se utiliza la gráfica. Generalmente coexisten varias presentaciones, dado que son complementarias, y se recomienda que cada una de estas partes contenga toda la información relativa a cada término, es decir, el conjunto de relaciones que correspondan al ámbito semántico del término.

La presentación temática, que corresponde generalmente a la parte principal del tesoro, supone ordenar los términos en clases y aparecen con sus relaciones genéricas-específicas. Se visualizan los diferentes niveles de la jerarquía mediante un descenso de los términos hacia la derecha (sangrado) o por una disposición arborescente.

La presentación alfabética muestra cada descriptor y no descriptor en orden alfabético y le siguen las indicaciones de las relaciones con los otros términos de acuerdo a este orden: Nota de alcance (NA); Cuasi-sinónimos (UP); Términos genéricos (TG); Términos específicos (TE) y Términos relacionados (TR).

El índice permutado contiene todos los términos del tesoro, tanto descriptores como no descriptores, alfabetizados por todas las palabras significativas que contienen. Este índice constituye un complemento de la sección alfabética y permite la ubicación de un descriptor compuesto haciendo búsqueda por cualquiera de sus elementos. La mayor ventaja de este índice es evitar que algún

descriptor pase inadvertido, especialmente en los casos en que la primera de las palabras que lo conforman es menos significativa que las demás. Los términos siempre aparecen completos, de tal manera que en la columna central aparece la palabra alfabetizada.

En la presentación gráfica los descriptores se presentan en grupos semánticos. La presentación más corriente consiste en que cada grupo está dotado de una rejilla y cada descriptor se halla asignado a un lugar determinado de acuerdo con unas coordenadas verticales y horizontales. Las relaciones entre los descriptores se manifiestan por: disposición gráfica (coordenadas, círculos concéntricos, etc.); flechas entre descriptores o colocación en los márgenes.

Por último mencionaremos que los tesauros pueden clasificarse atendiendo a múltiples puntos de vista. Entre los más usuales encontramos, según su ámbito: institucionales, nacionales, internacionales; según su área de conocimiento; y según su lengua: monolingües, bilingües o multilingües

3.3 Elaboración y mantenimiento

En la literatura consultada sobre construcción de tesauros (NISO 2003; Aitchison, Gilchrist y Bawden, 2000; Currás, 1998; ISO 2788-1986) se pueden distinguir algunas etapas comunes en las propuestas metodológicas, como son la definición del campo de trabajo, la recolección de términos, la normalización de la terminología, la organización de los descriptores, la evaluación del tesoro, así como su mantenimiento y actualización.

Los autores mencionados coinciden en señalar que en una primera instancia se debe analizar detalladamente el entorno de uso del tesoro, precisar la temática, definir el tipo de tesoro y dar a conocer la intención de su elaboración, con objeto de contar con el apoyo y aportación de diversos especialistas, y para asegurar la inexistencia de trabajos paralelos y así evitar la duplicidad. También mencionan la necesidad de revisar tesauros ya elaborados, pues como señala Currás (1998: 75)

“Se debe advertir que es muy difícil que no exista algún tesoro sobre el tema en que se trabaja. Hoy en día están todos los campos del saber cubiertos”.

Si bien un tesoro existente puede adaptarse o servir de guía, la misma autora señala que “...se tendrá que actuar como si se tratase de construir un tesoro nuevo”. Debido a todos los recursos tanto humanos como financieros que intervienen en la construcción de un tesoro, los autores recomiendan plantearse la elaboración de uno nuevo solamente cuando se hayan agotado las posibilidades de adoptar o adaptar alguno ya existente.

De acuerdo a estos autores un tesoro se puede elaborar a través de dos métodos: el deductivo y el inductivo. Vale destacar que todos ellos coinciden en proponer la utilización de ambos métodos en la elaboración del tesoro. El método deductivo consiste en extraer los términos de los documentos, sin necesidad de controlar el vocabulario ni determinar las relaciones entre ellos. Ya que se tenga un número suficiente de términos, un grupo de especialistas procederá a su revisión, identificando aquellos que representen las clases más genéricas y posteriormente deberán agruparse los términos restantes dentro de estas clases sobre la base de una relación lógica, de tal forma que las jerarquías tiendan a establecerse de lo general a lo particular. El control del vocabulario se determina entonces una vez que se hayan definido las jerarquías.

En el método inductivo, de igual manera, se extraen los términos de los documentos pero el control del vocabulario se establece desde el principio. Cada término es asignado a una o más clases genéricas, que fueron establecidas con anterioridad, y así el tesoro se va elaborando de lo particular a lo general. También se recomienda la colaboración de especialistas en el tema, que pueden actuar como miembros de un comité editorial.

Al respecto, Cleveland y Cleveland (2001: 41) señalan que “a thesaurus should be based on what is relevant to users”. También mencionan las características que Bates (1989) alude como indispensables para tesauros orientados a los usuarios:

1. Comprende una lista de todos los términos que se encuentran en la base de datos.
2. Distingue cuidadosamente los términos que actualmente están en uso de los que ya no se utilizan.
3. Brinda notas de alcance para problemas que posiblemente se puedan encontrar los usuarios finales.
4. Emplea nombres que se explican por sí mismos para términos o relaciones.
5. Incluye un amplio vocabulario, enfocado a los requerimientos del usuario final.

Vale recalcar que es imprescindible que durante el proceso de elaboración de un tesoro se lleven a cabo diversas pruebas para comprobar su utilidad, de acuerdo a las necesidades de la función documental, tomando en cuenta si la estructura propuesta va de acuerdo con las necesidades tanto del indizador como del usuario; si los términos incluidos son útiles para la indización y la recuperación de la información; si las relaciones son correctas y suficientes, sólo por mencionar algunos aspectos.

Para verificar su utilidad debe consultarse a especialistas en el tema y a los usuarios, y se debe tener en cuenta la continua evolución del lenguaje científico. Como señala Naumis (1998: 80) "La evaluación debe hacerse para la detección de nuevos términos, la supresión de algunos y la adecuación de otros que requieren las comunidades de información, una vez que está en uso el tesoro".

Actualmente existen diversos programas que permiten una actualización permanente. Amén de mencionar que algunas instituciones invitan a los usuarios a proponer términos para su evaluación y posible admisión en el tesoro. Sin olvidar que es necesaria la consulta con especialistas en el tema, incluyendo a lingüistas especializados en terminología, con el fin de asegurar la inclusión de los descriptores más pertinentes.

La actualización del tesoro debe hacerse tanto para incorporar la terminología derivada del desarrollo de la ciencia o materia a la que se dedica, como para cubrir lagunas o fallas detectadas durante su utilización, así como para adaptarlo a las necesidades de recuperación manifestadas por los usuarios a través de sus búsquedas. Este segundo proceso de corrección y adaptación puede ejecutarse a partir del estudio de las consultas realizadas por los usuarios al recuperar documentos del sistema.

A modo de conclusión se puede decir que un tesoro debe reflejar las evoluciones científicas y tecnológicas del área del conocimiento en el que se utiliza, razón por la cual podemos afirmar, sin temor a equivocarnos, que un tesoro nunca está terminado, por lo que es necesario tener presente su mantenimiento y actualización permanente.

3.4 Los tesauros y el entorno digital

En los últimos tiempos, el ámbito de la representación y recuperación de información en la Bibliotecología ha tenido que asumir el impacto de Internet y sus tecnologías asociadas, en especial la *World Wide Web* (*www*). Los cambios tecnológicos derivados están conduciendo a una progresiva digitalización del ámbito de la representación y recuperación de información que afecta por igual a los recursos de información, las herramientas de representación y recuperación, y el comportamiento de los usuarios en cuanto a sus requerimientos de recuperación de información. Esta influencia, sin duda, abre nuevas posibilidades a las pautas de diseño, elaboración, gestión y uso de las herramientas utilizadas para representar y recuperar la información.

Desde esta perspectiva López-Huertas (1997), De la Cueva Martín (2000), Shiri y Revie (2000), entre muchos otros, han comenzado a destacar los aportes del entorno digital a los tesauros, que podemos resumir de la siguiente manera:

1. El enriquecimiento de la funcionalidad de la estructura de los tesauros a partir de la hipertextualidad. La *www* es un medio idóneo para desarrollar la estructura hipertextual de la red semántica de un tesoro, con enlaces entre

los términos que mantienen una relación de equivalencia, jerárquica o asociativa y entre las diferentes secciones del tesoro (alfabética, temática y permutada) y la posibilidad de “navegar” entre ellas. Cabe recordar que pueden establecerse enlaces con imágenes y sonidos.

2. La reducción de costos en el mantenimiento y la actualización. La creciente sistematización en los procesos de elaboración de tesauros y el progresivo abandono del soporte impreso para la publicación de estas herramientas, hace viable una reducción de costos. Sin olvidar que el entorno digital permite simplificar la estructura de los tesauros, facilitando así su manejo, con interfaces adecuados, respecto a las versiones impresas²⁷.
3. La integración del usuario al proceso de creación, gestión y optimización de los tesauros a través de pruebas de utilidad, uso de técnicas de modelado de usuario, etc., lo cual permite elaborar herramientas que tengan en cuenta los requerimientos de los usuarios y descartar su construcción como simples estructuras teóricas. Además, el entorno digital permite disponer de herramientas terminológicas, de acceso universal, que facilitan la recuperación en distintas bases de datos y como fuente para establecer bases del conocimiento.
4. La posibilidad de aplicar medidas de reutilización e interoperabilidad en el momento de planificar y construir los tesauros. Con ello se posibilita el aprovechamiento y enriquecimiento de la información conceptual y lingüística ya generada para otros recursos. De igual manera, puede contribuir al desarrollo de tesauros multilingües y multidisciplinarios, con equipos de trabajo de diversos países.

Por otro lado, De la Cueva hace notar que aunque muchos sistemas siguen utilizando tesauros impresos para indizar y recuperar información, cada vez son más comunes las versiones electrónicas. Sin embargo, “los tesauros en línea son aun escasos, en comparación con el vasto número de bases de datos disponibles a través de este medio”.

Al respecto, la explicación dada por Dubois (1984, mencionado por De la Cueva) posiblemente siga siendo válida. Este autor señala que “la construcción de un tesoro requiere tiempo y dinero y el rendimiento inadecuado de la recuperación

²⁷ Al respecto, Lee-Smeltzer (2000: 213) coincide al señalar que “The advantages of an online thesaurus are that it can be modified and updated more easily and automatically. The results of the mapping can either be transparent to the searcher or a ranked list of choices of broader, narrower, or related terms for the searcher to further refine the search. The thesaurus may have the capability of mapping the natural language search terms to more than one controlled vocabulary scheme as well as linking among different controlled vocabularies schema.

en una base de datos en línea es generalmente invisible o relativo para el usuario, para quien no es fácil juzgar el resultado de la búsqueda, sobretodo si queda satisfecho con el mismo”. Desde este punto de vista, De la Cueva explica que el sistema puede ahorrarse la inversión en técnicas para mejorar la recuperación, especialmente cuando tales técnicas pueden conducir a un menor uso global de la base de datos.

Por su parte, Shiri y Revie (2000: 278) hacen notar que “Web browsers are not thesaurus-aware and existing metadata formats make little use of thesauri. The primary concern should be with interoperability rather than with construction or display”. De igual manera, estos autores mencionan que la falta de normas para la publicación de tesauros en línea ocasionan problemas relacionados con la utilidad, interoperabilidad y la forma de compartir el tesoro.

Sin duda, el entorno digital ha permitido que los tesauros, que antes eran propiedad exclusiva del indizador, se conviertan en herramientas documentales útiles para recuperar información para todos aquellos usuarios en línea, siempre y cuando se les proporcione a éstos la mayor cantidad de información posible para facilitar la elección de los términos de búsqueda. Es necesario que el tesoro contribuya a enriquecer la comunicación entre un sistema de información y sus usuarios, eliminando la ambigüedad y la vaguedad de los términos.

En síntesis, la evolución del tesoro tradicional, hacia una estructura de gran capacidad para la representación conceptual y para la identificación de numerosas relaciones conceptuales, apunta a la posibilidad de poder convertirse en una base de conocimiento que facilite de forma decisiva la recuperación de la información en un dominio temático determinado.

En el siguiente capítulo se presenta el análisis realizado de los seis tesauros seleccionados. Cabe señalar que para llevar a cabo un estudio más reflexivo se han tomado en cuenta los conceptos aquí mencionados.

Capítulo II

METODOLOGÍA

1.1 Evaluación de tesauros relacionados con el agua

Como se mencionó en el capítulo anterior, antes de emprender la tarea de elaborar un nuevo tesoro es indispensable llevar a cabo una valoración de los recursos existentes para asegurarnos de que el trabajo de comenzar de cero sea verdaderamente necesario. Razón por la cual, para sustentar los elementos que servirán de base para determinar si es posible adoptar y/o adaptar alguno ya existente, hemos realizado el análisis de seis tesauros relacionados con la temática hídrica, sin que el número de tesauros seleccionados pretenda ser de alguna manera exhaustivo.

A continuación se presentan los resultados del análisis realizado. Antecede una revisión de la literatura relacionada con los criterios para evaluar tesauros, mismos que han servido de base para llevar a cabo dicho análisis.

1.2 Criterios de evaluación

El proceso de evaluación de un tesoro puede aumentar el valor de éste en términos de uso, alcance, precisión y acierto. La evaluación implica poner en práctica un conjunto de medidas que ayudarán a determinar si un tesoro contiene los términos suficientes y las relaciones semánticas adecuadas que permitan representar las materias de los documentos en la indización y en las peticiones de búsqueda. Cabe destacar que dependiendo del propósito principal del tesoro, diferentes criterios pueden ser utilizados para guiar la revisión y el proceso de evaluación¹.

¹ Al respecto Davis y Rush (1979: 87) opinan que “The troublesome aspect of evaluation is that there are too many different criteria upon which to base evaluative procedures and none of these criteria has, or will likely receive, wide acceptance. This is because evaluation always reduces, ultimately, to a personal value judgement”.

De una manera sencilla, Davis y Rush (1979: 22) proponen los siguientes criterios para evaluar tesauros:

Terminología: ¿Va de acuerdo con la disciplina? ¿Está actualizada?
¿Es precisa?

Alcance: ¿Es demasiado amplio o reducido para abarcar la disciplina?

Subdivisiones: ¿Son razonables?

Definiciones y notas: ¿Se incluyen suficientes para evitar confusiones?

Referencias: ¿Son las adecuadas en número y forma?

Formato: ¿Es legible?

Números de clasificación: ¿La lista está relacionada con algún esquema de clasificación?

Ya más detalladamente, Lancaster (1986: 155) hace notar que un tesoro puede ser evaluado de manera superficial tomando en cuenta si incluye una introducción² que informe de su ámbito de aplicación y una explicación sobre sus características distintivas, como las notaciones empleadas. Ya en un nivel superior habrá que tomar en cuenta si las relaciones entre los términos generales, específicos y relacionados son las apropiadas, y si los términos ambiguos o poco usuales son aclarados a través de notas de alcance, además de valorar que todas las relaciones tengan sus recíprocos.

El autor continúa comentando que, de igual modo, es posible comprobar si el tesoro cumple con las normas internacionales en cuanto a las convenciones singular/plural, formatos de palabras, entradas directas y otros aspectos de consistencia. Además, pueden considerarse aspectos estéticos de composición y tipografía. Sin embargo, señala que a pesar de las pruebas y evaluaciones que se realicen, "...it is impossible to evaluate a thesaurus except under conditions of actual use. The acid test is whether the terms are sufficiently specific to adequately

² Al respecto Foskett (1997: 133) menciona que "The thesaurus can take several different forms. In every case, it will require an introductory section which explains how it can best be used. If this is not provided, the unlucky users will have to produce one for themselves. An introduction should therefore aim to cover all the sections: the limits of the field or fields covered, the choice and forms of terms, the network of relations, the form of presentation, and the use of special features like the graphical displays".

represent the subject matter of documents and requests, how much syntactic ambiguity exists in the data base, and how helpful the thesaurus is in suggesting all terms needed to adequately perform a particular search” (1986: 157).

Por otro lado, Naumis (1998: 80-81) propone, con el objeto de aportar los elementos que mejor conduzcan a una evaluación desde varios ángulos, cuatro formas desde las cuales se puede realizar una valoración: la estructural, la relacional, la de eficiencia y la integral. La autora hace notar que “las condicionantes que propiciaron esta clasificación están dadas por los elementos que intervienen en la construcción, la utilidad en el proceso de indización y el resultado para la recuperación de información”, y explica que

“...en la forma estructural se evalúan los aspectos lógico-lingüísticos e intervienen sólo aquellos elementos concernientes a la organización interna del tesoro: los términos, sus relaciones y las asociaciones derivadas de ellos.

“En la valoración relacional se busca la relación de la estructura de conocimientos de la disciplina implicada con la estructura temática del tesoro, esto es, la conjugación de términos y contenidos que garanticen la utilización de un lenguaje actualizado acorde con el empleado por los generadores de información en el área temática implicada.

“Por lo regular, la eficiencia de un tesoro u otro lenguaje de búsqueda de información se mide por la capacidad de recuperación de información relevante, pero los factores que intervienen en este resultado van más allá de la capacidad del lenguaje utilizado; aquí intervienen elementos tales como las características del fondo, de la colección o del recurso de información para el cual fue diseñado el lenguaje, la calidad del proceso de indización y hasta la estrategia de búsqueda que se emplee en la recuperación de la información. La evaluación de la eficiencia en la recuperación de documentos relevantes en el sistema de información permitirá disponer de otros elementos para juzgar la calidad.

“De las tres formas anteriores se deriva la cuarta, considerada por su nivel de complejidad como integral. Como se observa, cada una de las tres primeras no resulta excluyente, todo lo contrario, pues son etapas que van de lo particular a lo

general, mediante el incremento de variables a considerar, en el que se conjugan elementos internos del lenguaje de búsqueda, así como aquellos que forman parte de su entorno o medio ambiente”.

Enfocados más en las relaciones semánticas, Cleveland y Cleveland (2001: 47) hacen notar que aunque la satisfacción del usuario (o la falta de ella) puede ayudar a determinar la calidad de un tesoro, es fundamental que se establezcan de antemano ciertos criterios de evaluación, teniendo siempre en mente que un tesoro ha sido diseñado como un instrumento terminológico que permite controlar la indización para facilitar la recuperación. No hay que olvidar que uno de los propósitos del tesoro es establecer un vínculo entre los conceptos y los términos agrupando palabras del lenguaje natural en grupos de palabras similares y seleccionando los términos preferentes. Pero dada la ambigüedad del lenguaje y de su uso, esto rara vez puede ser logrado de manera absoluta; sin embargo, los autores señalan que el grado de correspondencia es un indicador de la calidad del tesoro.

De manera similar a Davis y Rush, Cleveland y Cleveland sugieren las siguientes preguntas como parámetros para la evaluación:

1. ¿Qué tan acertada es la cobertura de la disciplina? ¿Los conceptos incluidos permiten de manera adecuada indizar y recuperar la información?
2. ¿Las relaciones semánticas están establecidas adecuadamente?
3. ¿Qué tan pertinente es el despliegue del tesoro? ¿Es fácil de consultar y comprender? ¿Permite una indización y recuperación de la información eficiente y efectiva?

Por último, otro aspecto que pocos autores mencionan y que Cleveland y Cleveland hacen notar es el costo en tiempo y dinero que representa evaluar un tesoro.

Por otro lado, Owens y Cochrane (2004 : 91) señalan que si bien las normas técnicas para la elaboración de tesoros no establecen estándares para la

interoperabilidad, éstas pueden servir de base para evaluar la estructura y el rigor de un tesoro existente. Las autoras proponen cuatro técnicas de evaluación:

- 1) *evaluación de la forma*: se lleva a cabo antes de que el tesoro sea divulgado y, además del análisis que pueden realizar diversos expertos, se toma en cuenta la opinión de los usuarios potenciales, y también se acude a la retroalimentación en cuanto a la presentación y facilidad en su uso.
- 2) *evaluación de la estructura*: las normas para la elaboración de tesauros pueden usarse para evaluar la estructura y el rigor del tesoro. Al igual que Lancaster, las autoras señalan que a través de esta evaluación es posible confirmar si el tesoro se apega a los criterios que establecen las normas en términos del uso del singular y del plural, al establecimiento de relaciones semánticas, y otros temas que tienen que ver con la consistencia. Además este tipo de evaluación podría combinarse con una evaluación subjetiva realizada por expertos para confirmar la efectividad del vocabulario controlado en relación con el nivel de satisfacción del usuario y sus necesidades de información.
- 3) *evaluación comparativa*: se realiza con el propósito de determinar la mejor audiencia para el tesoro y para proponer sugerencias específicas para mejorarlo.
- 4) *evaluación por observación*: una vez que el tesoro ya está bien establecido se evalúa por los usuarios en situaciones reales.

En síntesis, podemos mencionar que cuando un tesoro está siendo analizado por expertos, valorado por los usuarios, sus características comparadas con normas y estándares internacionales o si se realizan pruebas de acceso y recuperación, ya sea en un sistema o en línea, el tesoro está siendo evaluado. Sin olvidar que cada uno de estos procesos implica diferentes metodologías.

Sin embargo, hemos de coincidir con Correa Uribe (1999: 145) cuando señala que “Los atributos globales de planeación, estructura, organización y actualización son fundamentales al desarrollar tesauros para lograr el objetivo deseado. Pero realmente, como prueba definitiva, la mejor manera de evaluar un tesoro, después de cumplir con los requisitos mínimos, es ponerlo en contacto con los usuarios directos, quienes a través del uso cotidiano, su conocimiento y dominio de las materias, determinen la verdadera efectividad de este tipo de herramientas”.

Es decir, la eficacia real de un tesoro la determinará su uso en la práctica por parte de los indizadores y los usuarios; los primeros, al encontrar términos suficientemente específicos que les permitan representar las materias y los segundos, al constatar si la estructura de las relaciones semánticas les ayuda a escoger el término apropiado para recuperar la información que requieren.

1.3 Procedimiento de evaluación

Para fines de este trabajo se consultaron los siguientes seis tesauros:

- 1) Agrovoc
- 2) General Multilingual Environmental Thesaurus
- 3) OECD Macrothesaurus
- 4) Tesoro de Ingeniería Hidráulica
- 5) Tesoro de Ingeniería Sanitaria y Ambiental
- 6) Thésaurus Eau

Los cinco primeros están disponibles en línea y el sexto sólo se pudo consultar en formato impreso. Los criterios para seleccionar dichos tesauros fueron:

- ♣ inclusión del tema del recurso hídrico, no importando el enfoque con el que se aborda;
- ♣ institución que respalda al tesoro, tomando en cuenta su experiencia y el prestigio que goza dentro del medio científico y académico;
- ♣ disponibilidad y accesibilidad, condición *sine qua non* para acceder al tesoro y poder llevar a cabo el análisis correspondiente.

Nuestro propósito es doble. Por una parte, realizar una revisión en cuanto al contenido, la presentación, y la facilidad de consulta. Por otra, llevar a cabo un análisis de los términos incluidos que tengan que ver con los recursos hídricos. Todo lo anterior con objeto de determinar cómo estos tesauros pueden adoptarse y/o adaptarse, para en un futuro elaborar un tesoro del agua para México.

En relación con la terminología, el análisis se hizo en función de conocer los recursos disponibles y aprovecharlos para definir un proyecto de tesoro acorde a las necesidades de los sistemas de información relacionados con el recurso hídrico en México. Tomando en cuenta que es necesario mostrar las relaciones que existen entre los diversos conceptos de una disciplina para facilitar la obtención de una visión integrada acerca de la dinámica del lenguaje y, por consiguiente, del movimiento cultural que este representa.

Por otro lado, al no disponer de una metodología sistemática y probada en el medio para la evaluación de este tipo de herramientas terminológicas, se tomó la propuesta de Naumis (2002), a la que se le añadieron algunos elementos para el análisis llevado a cabo (ver Anexo), que la que suscribe validó y aplicó a los tesauros seleccionados. Dicha propuesta contempla tres elementos principales: 1) la presentación; 2) la consulta; y 3) el contenido.

La presentación se refiere a las características distintivas de cada tesoro y a la forma como éstas son tratadas. Es decir, si el nombre refleja realmente la temática abordada, cual es su campo temático o dominio de aplicación, si se incluye una introducción, cual es la estructura del tesoro, cuales son las jerarquías empleadas, cual ha sido el año de última edición, si es monolingüe o multilingüe, entre otros aspectos.

La consulta hace mención a la facilidad de manejo del tesoro. Esto es, al tipo de soporte, así como a su disponibilidad y accesibilidad y a la claridad con la que se presenta su guía de uso.

Por último y no por eso menos importante, el contenido alude a la cobertura temática, a las características cuantitativas, a la consistencia de los descriptores, a la notación utilizada para designar las relaciones semánticas, a su nivel de estructuración y a su mantenimiento y actualización. Vale mencionar que las características cuantitativas advierten sobre la cuantificación de temas, descriptores, no descriptores y jerarquías incluidos.

1.4 Fuentes de estudio

Agrovoc (FAO)³

Desarrollado a principios de la década de 1980 por la Organización de las Naciones Unidas para la Agricultura y la Alimentación (mejor conocida como FAO, por sus siglas en inglés) y la Comisión de las Comunidades Europeas (hoy en día Unión Europea), *Agrovoc* es un tesoro agrícola multilingüe que abarca la terminología de todos los ámbitos de la agricultura, la silvicultura, la pesca y aquellos temas relacionados con los alimentos, como el agua y el medio ambiente.

El objetivo principal del *Agrovoc* es la indización y búsqueda de datos de los sistemas de información agrícola, especialmente los sistemas internacionales cooperativos AGRIS (sistema internacional de información sobre ciencias y tecnología agrícolas) y CARIS (sistema de información sobre investigaciones agronómicas en curso), ambos coordinados por la FAO.

Aunque no se ofrece una introducción propiamente dicha, navegando por el sitio del *Agrovoc* fácilmente se puede obtener información sobre las características del mismo, así como información adicional sobre diversas cuestiones, como por ejemplo la relacionada con los derechos de autor, las últimas actualizaciones, el uso de ontologías, etcétera.

Agrovoc está disponible en los cinco idiomas oficiales de la FAO, que son inglés, francés, español, chino y árabe. Sin embargo, nunca se hace mención sobre cuál es el idioma original del tesoro. También está disponible en checo, portugués y tailandés, y actualmente está siendo traducido o revisado a otras lenguas como el alemán, italiano, coreano, japonés, húngaro y eslovaco.

Tomando en cuenta que una imagen vale más que mil palabras, consideramos pertinente mostrar la página principal del sitio del tesoro en lengua castellana (Figura No. 1), donde se puede apreciar toda la gama de información que ofrece.

³ http://www.fao.org/aims/ag_intro.htm [Consulta: 07.08.06]

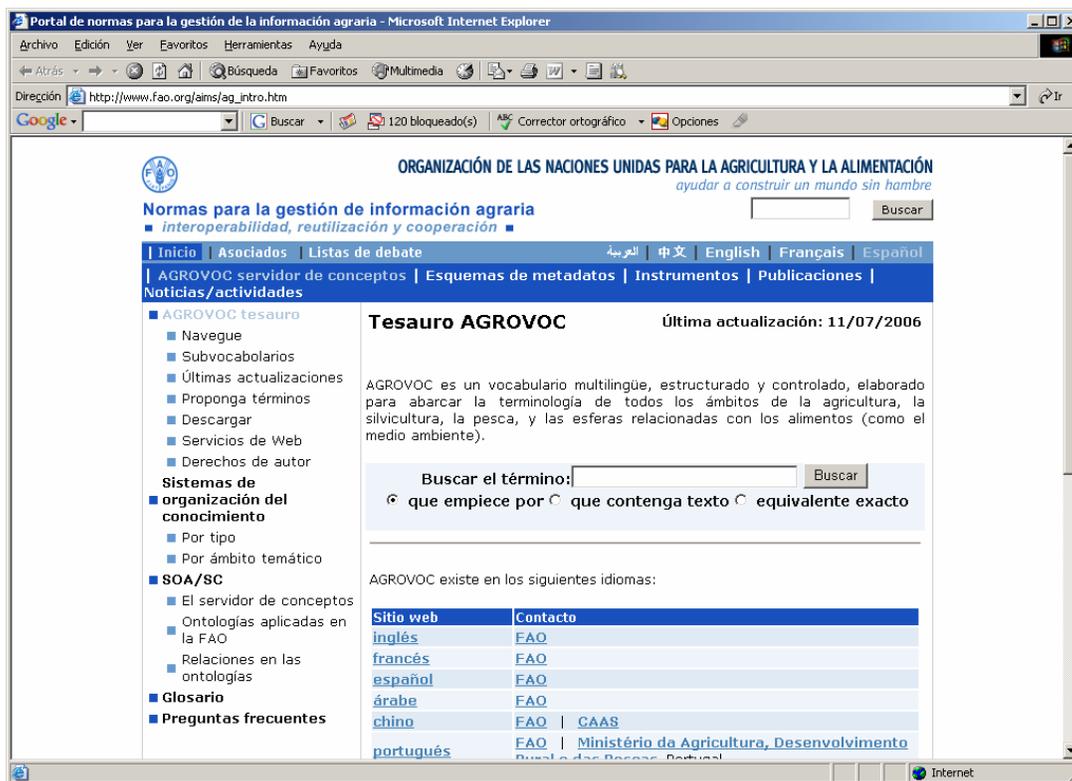


Figura No. 1

Como se observa, además del tesoro, en el sitio se encuentra información diversa. Sin embargo, nos enfocaremos en aquella relacionada con los lenguajes documentales como por ejemplo, los *Sistemas de organización del conocimiento* en donde existe la posibilidad de encontrar diferentes modalidades de vocabularios controlados (opción “Por tipo”), como tesauros, ontologías, listas de autoridad, encabezamientos de materia, etc., de los temas relacionados con la agricultura, la silvicultura, la pesca y el medio ambiente. Dentro de la opción “Por ámbito temático”, se localiza información relacionada con los once temas que comprende *Agrovoc*, mismos que se mencionan más adelante dentro de este análisis.

Es pertinente resaltar que también se encuentra el *Servidor de conceptos del Servicio de Ontología Agrícola (SOA/SC)*, del cual brevemente mencionaremos que bajo la premisa de que la ontología lleva a un paso más adelante al usuario al reestructurar más formalmente los términos y proporcionar relaciones más

complejas entre los conceptos, en comparación con los tesauros actuales, el SOA/SC busca ofrecer un marco rico y moderno para hacer modelos, dar servicio y para la gestión de la terminología agraria, mediante la definición formal de los términos y con una descripción más precisa de las relaciones entre éstos⁴.

En cuanto a su disponibilidad y accesibilidad, consideramos que es eficaz, ya que *Agrovoc* puede descargarse sin costo en los formatos MySQL, TagText, ISO 2709, XML, y Access de Microsoft, siempre y cuando sea con objetivos didácticos o de otro tipo que no tengan fines comerciales. La descarga está autorizada por los titulares de los derechos de autor, sin autorización previa por escrito, siempre que se dé crédito a la fuente. Sin embargo, está prohibida la reproducción de cualquier material relacionado con el tesauro, para venta o con cualquier otro propósito comercial sin autorización por escrito de los titulares de los derechos de autor, los cuales en los idiomas oficiales pertenecen a la FAO. Los derechos de autor en las otras lenguas corresponden a las instituciones responsables de producir las versiones nacionales del tesauro.

En cada uno de los idiomas, *Agrovoc* ofrece dos posibilidades de búsqueda. La primera a través de escribir un término (como se puede apreciar en la Figura No. 1) eligiendo una de tres opciones: que empiece por la palabra seleccionada, que sea el término exacto o que esté contenido en el texto. Al final de cada bloque de términos se dan los equivalentes en las otras lenguas, siempre para los descriptores (en mayúsculas y en negritas) y opcionalmente para los no descriptores (en minúsculas y en cursivas).

Como se puede apreciar en la figura No. 2, la segunda posibilidad brinda la opción de escoger por orden alfabético los términos. También se puede mostrar la lista completa.

⁴ Además se señala que "Al integrarse con instrumentos de búsqueda en Internet, facilitará enormemente la recuperación de recursos, no sólo al proporcionar acceso a los documentos específicos que se busquen, sino también proponiendo otros recursos relacionados que pueden ser pertinentes al tema de interés. Esta funcionalidad adicional no sólo incrementará espectacularmente el alcance de los motores de búsqueda en Internet, sino que revolucionará la forma en que el usuario interesado en recursos agrícolas interactuará con la red".

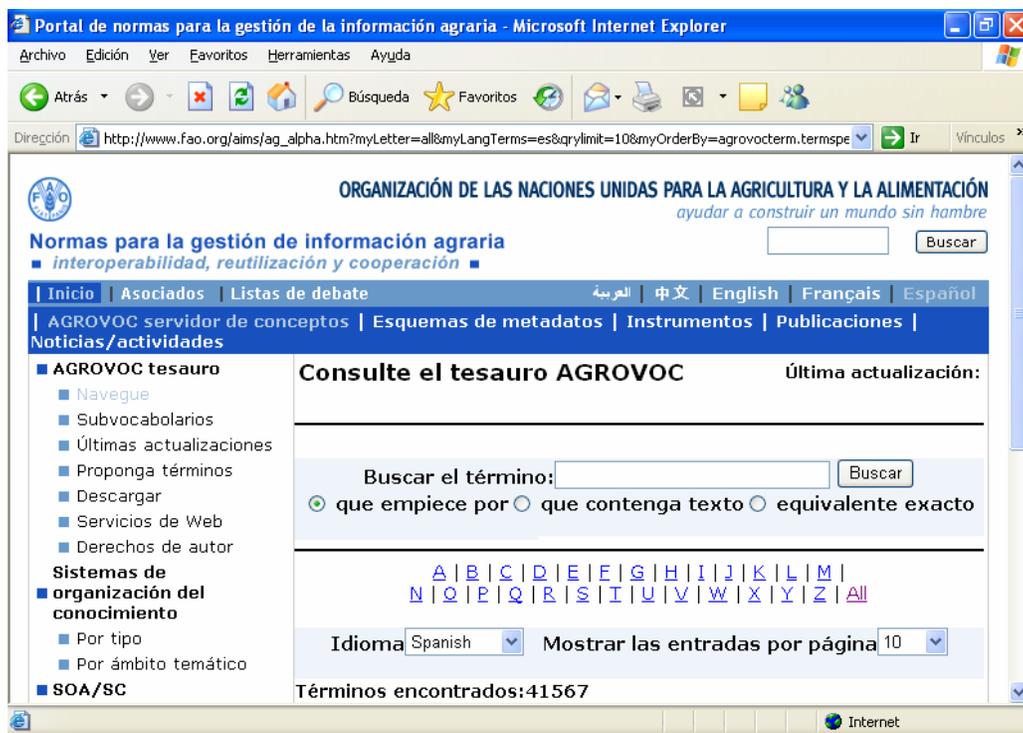


Figura No. 2

En este tesoro se establecen las relaciones indicadas por las normas técnicas, es decir las jerárquicas, las asociativas, y las de equivalencia, las cuales abarcan todos los casos de cuasi-sinonimia, incluidas las variantes ortográficas y las siglas *versus* nombres completos. En cuanto a la notación empleada para designar las relaciones semánticas se han utilizado las siguientes: *Término específico (NT)*, *Término genérico (BT)*, *Referencia de Nota de Alcance (SNR)*, *Se refiere a nota de alcance (SNX)*, *Término relativo (RT)*, *Usado para (UF)*, *Usado para + (UF+)*, *Usar (USE)*, *Ver (SEE)*, *Visto Para (SF)*. Cabe aclarar que no se ofrece una descripción que explique el significado de cada una de estas notaciones.

Se menciona que las notas de alcance se han empleado “con profusión” para precisar el sentido tanto de los descriptores como de los no descriptores, además de contener definiciones de términos. De igual manera, para los términos que son polisémicos se han utilizado remisiones a los descriptores que representan los demás conceptos que no son considerados para este término y limitaciones (o ampliaciones) del significado del término. Sin embargo, cabe hacer notar que durante la búsqueda se encontraron pocas notas de alcance.

En cuanto a la consistencia de los descriptores, podemos señalar que el uso de singular y plural se apega a lo establecido por las normas. Por otro lado, se incluyen identificadores geográficos y nombres de instituciones, más no de personas ni periodos. Referente a sus características cuantitativas, *Agrovoc*, en su versión en español, contiene actualmente 28,190 descriptores y 13,377 no descriptores incluidos en los siguientes 11 temas:

1. Agricultura - general
2. Educación y especialización
3. Ingeniería, tecnología e investigación
4. Pesca y acuicultura
5. Nutrición humana e inocuidad de los alimentos
6. Silvicultura
7. Información geográfica y regional
8. Gobierno, administración y legislación
9. Gestión de la información
10. Recursos naturales y medio ambiente
11. Producción y protección vegetal

La primera edición del *Agrovoc* fue publicada en 1982 y la última actualización, según se señala en el sitio, fue en julio de 2006. También se menciona que mantiene una actualización continua según las propuestas enviadas por los centros participantes AGRIS y CARIS, y que “un tesoro es un lenguaje evolutivo que debe reflejar las evoluciones científicas y tecnológicas en el sector de la agricultura. Por lo tanto, serán bien acogidas las sugerencias para introducir nuevos conceptos o para mejorar los que aparecen en esta edición”, lo que resulta posible a través de la opción “Proponga términos”, que aparece en la página principal del *Agrovoc* (figura No. 1).

Realizando un somero análisis terminológico, encontramos que dentro de la temática hídrica la mayoría de los términos incluidos son similares a los términos empleados en México, a excepción de algunos, obviamente, como por ejemplo

“Enfermedades transmitidas por agua”, que en nuestro país se emplea por lo regular “Enfermedades de origen hídrico” o “Agotamiento del agua” para referirse a la “Escasez de agua”, así como “Legislación medioambiental”, que en México se conoce como “Legislación ambiental”.

Sin embargo, consideramos necesario tomar en cuenta que algunas de las traducciones son literales del inglés, lo que nos hace suponer que la versión original, que si bien no se menciona, es en este idioma. Para citar sólo un ejemplo, para referirse a los datos cuantitativos del tesoro se muestra la instrucción “Vea algunas ilustraciones”, ya que en la versión en inglés está como “See some illustrations” (que en realidad debería de ser “figures” en lugar de “illustrations”).

En síntesis, consideramos que *Agrovoc* es una buena opción a considerar, pues su estructura va más allá de permitir la mera consulta del lenguaje. Y si bien su temática se centra en la agricultura, no hay que olvidar que ésta está estrechamente relacionada con los usos y conservación del recurso hídrico, ya que la agricultura consume a nivel mundial un porcentaje mayoritario del agua dulce disponible en el mundo.

General Multilingual Environmental Thesaurus⁵ (Unión Europea)

La parte introductoria del *Tesoro Ambiental General Multilingüe* (*GeMET*, por sus siglas en inglés) explica que éste ha sido desarrollado por la Agencia Europea de Medio Ambiente y la Agencia de Protección Ambiental de los Estados Unidos (mejor conocida como EPA, por sus siglas en inglés), de acuerdo a las normas ISO para tesauros monolingües y multilingües. Fue concebido como una herramienta terminológica para el *Catálogo de Fuentes de Datos* con objeto de facilitar la indización, la recuperación y el control de la información relacionada con el medio ambiente.

⁵ <http://www.eionet.europa.eu/gemet> [Consulta: 23.08.05]

El año de creación y de última edición no se mencionan, como tampoco se hace referencia a los derechos de autor. Sin embargo, se explica que la idea fundamental para el desarrollo del *GeMET* fue utilizar “lo mejor de los tesauros multilingües disponibles”, con el objeto de ahorrar tiempo y dinero, y evitar la duplicidad de tareas. Además se señala que el *GeMET*, concebido como un tesoro general, tiene como propósito definir un lenguaje común dentro de la temática ambiental, siendo el inglés británico el idioma elegido, aunque los equivalentes en inglés norteamericano se han añadido con la colaboración de la EPA. Actualmente la versión 2.0 del *GeMET* se presenta en 21 lenguas y no requiere de la instalación de algún programa de cómputo en particular.

De igual manera, se comenta que los términos que componen este tesoro han sido seleccionados de documentos multilingües, generados, en su gran mayoría, por organismos europeos relacionados con el medio ambiente. Presuntamente a través de un código se relaciona el término con el documento de donde se obtuvo. También se menciona que, como en otros tesauros multilingües, una notación neutral alfanumérica permite la identificación de un concepto independientemente del idioma del usuario. Al respecto, es necesario señalar que de todos los términos que se consultaron no se encontró notación alguna, excepto la que se muestra más adelante.

Se explica además que las locuciones resultantes han sido organizadas en un esquema de clasificación compuesto por 3 grupos principales, 30 grupos intermedios y 5 grupos instrumentales accesorios. Cada descriptor ha sido ordenado en una estructura jerárquica encabezada por un “*Top Term*”. La parte introductoria también señala que el grado de polijerarquía se ha tratado de reducir al mínimo, aunque para permitir una recuperación de los términos relacionados temáticamente pero distribuidos en diferentes grupos, se ha acordado establecer una serie de 40 temas dentro de los cuales cada descriptor ha sido asignado a tantos temas como ha sido necesario. Esto con objeto de acceder a los descriptores a través de la lista jerárquica, temática y alfabética.

Como se puede ver a continuación el esquema de clasificación difiere un poco de lo señalado anteriormente, pues son cuatro grupos principales y no tres. En cuanto a los grupos intermedios efectivamente son 30 y los cinco grupos instrumentales accesorios no se muestran en el tesoro en línea.

ACTIVIDADES Y PRODUCTOS HUMANOS, EFECTOS SOBRE EL MEDIO AMBIENTE

AGRICULTURA, SILVICULTURA, CRÍA DE ANIMALES, PESCA
ASPECTOS FÍSICOS, RUIDO VIBRACIONES, RADIACIONES
COMERCIO, SERVICIOS
EFECTOS, IMPACTOS
ENERGÍA
INDUSTRIA, ARTESANÍA, TECNOLOGÍA, EQUIPAMIENTOS
PRODUCTOS MATERIALES
QUÍMICA, SUSTANCIAS, PROCESOS
RECREACIÓN, TURISMO
RECURSOS (uso de los recursos)
RESIDUOS, CONTAMINANTES, CONTAMINACIÓN
TRÁFICO, TRANSPORTE

MEDIO NATURAL, MEDIO ANTRÓPICO

ANTROPOSFERA (medio edificado, asentamientos humanos, ordenación del territorio)
ATMOSFERA (aire, clima)
BIOSFERA (organismos, ecosistemas)
ESPACIO
HIDROSFERA (agua dulce, agua marina, aguas)
LITOSFERA (suelo, procesos geológicos)
MEDIO AMBIENTE (medio natural, medio antrópico)
TERRITORIO (paisaje, geografía)
TIME (chronology) [Translation not available] (sic)

ASPECTOS SOCIALES, POLÍTICA AMBIENTAL, MEDIDAS

ADMINISTRACIÓN, GESTIÓN, POLÍTICA, INSTITUCIONES, PLANIFICACIÓN
ECONOMÍA, FINANZAS
INFORMACIÓN, EDUCACIÓN, CULTURA, CONCIENCIA AMBIENTAL
INVESTIGACIÓN, CIENCIAS
LEGISLACIÓN, NORMAS, CONVENCIONES
POLÍTICA AMBIENTAL
RIESGOS, SEGURIDAD
SALUD, NUTRICIÓN
SOCIEDAD

LISTAS ADICIONALES

TERMINOS FUNCIONALES
TÉRMINOS GENERALES

Por su parte, la lista temática presenta los siguientes 40 temas:

administración	general
agricultura	geografía
agua	industria
aire	información

alimento, agua potable	investigación
áreas naturales, paisaje, ecosistemas	legislación
aspectos militares	materiales
aspectos sociales, población	medio ambiente urbano, tensión urbana ("stress urbano")
biología	pesca
clima	política ambiental
comercio, servicios	química
contaminación	radiaciones
cría de animales	recursos
desastres, accidentes, riesgo	residuo
dinámica natural	ruido, vibraciones
economía	salud humana
edificación	silvicultura
energía	suelo
espacio	transportes
física	turismo

En la versión 2.0 del GeMET se indica que éste contiene 5,298 términos, incluyendo 4,034 descriptores y 1,264 no descriptores, además de 4,000 definiciones disponibles, que suponen asegurar una coherencia interna sistemática y lingüística. Como se mencionó anteriormente, no se hace referencia alguna a la notación utilizada para designar las relaciones semánticas. Al respecto, consideramos pertinente mostrar algunos ejemplos elegidos al azar que hablan por sí mismos. Nuestra intención es hacer notar la presentación y la traducción de los vocablos. Hemos de señalar que de entre todos los términos que se revisaron no se encontró ningún no descriptor.

agua de marea

Concept definition: *

Any water whose level changes periodically due to tidal action. (Source: MGH)

* **Note:** Definition is not available for the current language

broader terms

 agua salada

narrower terms

 alta de marea agua

related terms

 mareas

Scope note:

scope note is not available

Themes:

agua
dinámica natural

Groups:

HIDROSFERA (agua dulce, agua marina, aguas)

tratamiento de residuos

Concept definition: *

Any process or combination of processes that changes the chemical, physical or biological composition or character of any waste or reduces or removes its harmful properties or characteristics for any purpose. (Source: HMD)

* **Note:** Definition is not available for the current language

 **broader terms**

 Sobrecarga

 **narrower terms**

 compactación

 depuración

 desgasificación de residuos

 digestión (aguas residuales)

 Estancamiento

 técnicas de conversión de desechos

 tratamiento de aguas residuales

 tratamiento de lodos

 tratamiento de residuos biológicos (sic)

 tratamiento químico de residuos

 **related terms**

 Compostificación

 gestión de residuos

 plantas de tratamiento de residuos

 proceso termoselectivo

 trituradora

 vitrificación

Scope note:

scope note is not available

Themes:

industria
residuo

Groups:

RESIDUOS, CONTAMINANTES, CONTAMINACIÓN

Como se puede observar, la simbología utilizada difiere considerablemente de la que comúnmente se sugiere en las normas técnicas para la elaboración de tesauros. Por otro lado, hemos de hacer notar que tanto la definición del concepto y las relaciones semánticas se presentan en inglés, aun cuando se haya revisado la versión en castellano. Además, de todos los términos que se consultaron, ninguno tenía disponible la nota de alcance. En cuanto a la consistencia de descriptores podemos comentar que el uso de singular y plural se apega a lo establecido en las normas y sí se emplean nombres geográficos y de instituciones. También hemos de señalar que en relación al mantenimiento del *GeMET*, no se menciona ni la periodicidad ni el método de actualización.

Dentro del análisis terminológico podemos decir que si bien algunos vocablos coinciden con los empleados en México, como “tratamiento de aguas residuales”, existen otros como “gestión de residuos”, que en nuestro país es más conocido como “manejo de residuos”. También hay otros cuya traducción es poco comprensible, como “alta de marea agua” u otro caso sería la de aquellos términos con una connotación diferente a la que se le da en México, como por ejemplo “water pricing”, cuya traducción en este tesoro es “valoración del agua”, cuando en México sería “tarifas”, pues “valoración del agua” va mucho más allá de la valoración económica y comprende también tanto la dimensión ambiental (la función del agua en el ecosistema) como la social (la utilización del agua para producir alimentos, por ejemplo).

Por último, debemos mencionar que los términos incluidos dentro del tema “agua” están más enfocados a las ciencias naturales y en menor medida a las ciencias sociales. Sin embargo, existen algunos temas que podrían tomarse en cuenta, pues el tema del medio ambiente está contemplado de una manera holística. Tampoco se deben descartar algunas definiciones y traducciones.

OECD Macrothesaurus⁶ (Organisation for Economic Co-operation and Development)

Dado que este tesaurus presenta una característica particular, pues aparentemente la Organización para la Cooperación y el Desarrollo Económico (OCDE) no se hace responsable de la versión disponible en línea, aun cuando cuenta con los derechos de autor, consideramos pertinente presentar a continuación la hoja principal del sitio (Figura No. 3) en donde se puede tener acceso a su macrotesaurus.

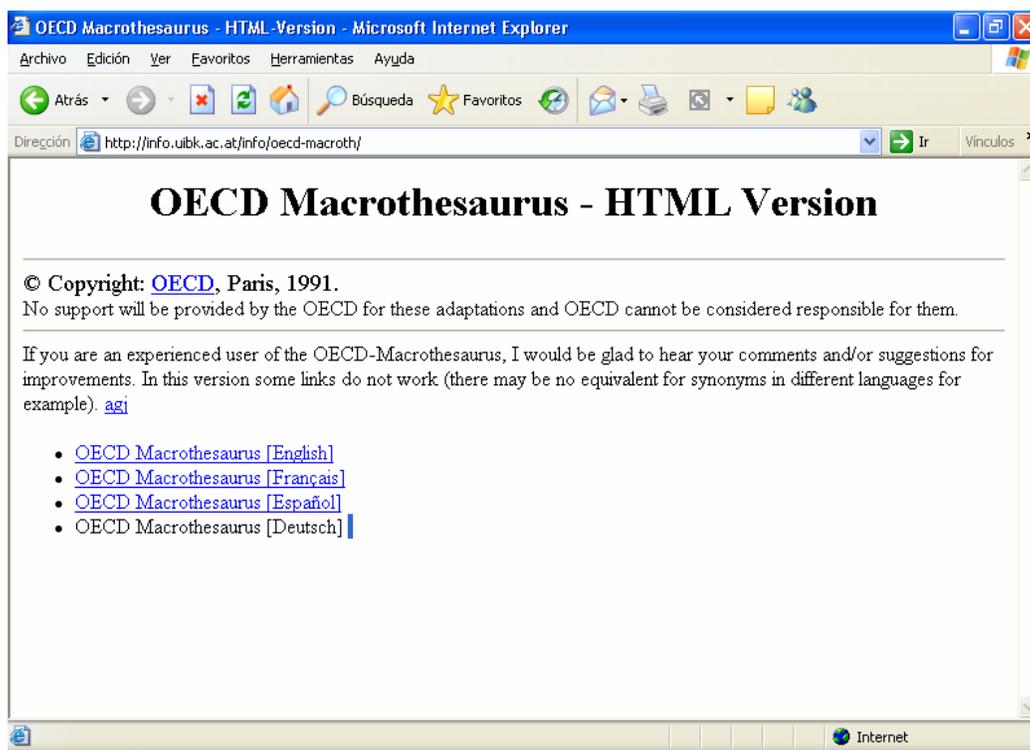


Figura No. 3

Como se puede observar, el macrotesaurus no incluye introducción ni guía de uso. A continuación presentamos la pantalla que se muestra al elegir el idioma deseado (Figura No. 4), en la cual se despliegan los 19 temas principales con sus respectivos subtemas. De igual manera, en esta pantalla se tiene acceso a los términos por orden alfabético.

⁶ <http://info.uibk.ac.at/info/oezd-macroth/> [Consulta: 21.08.05]

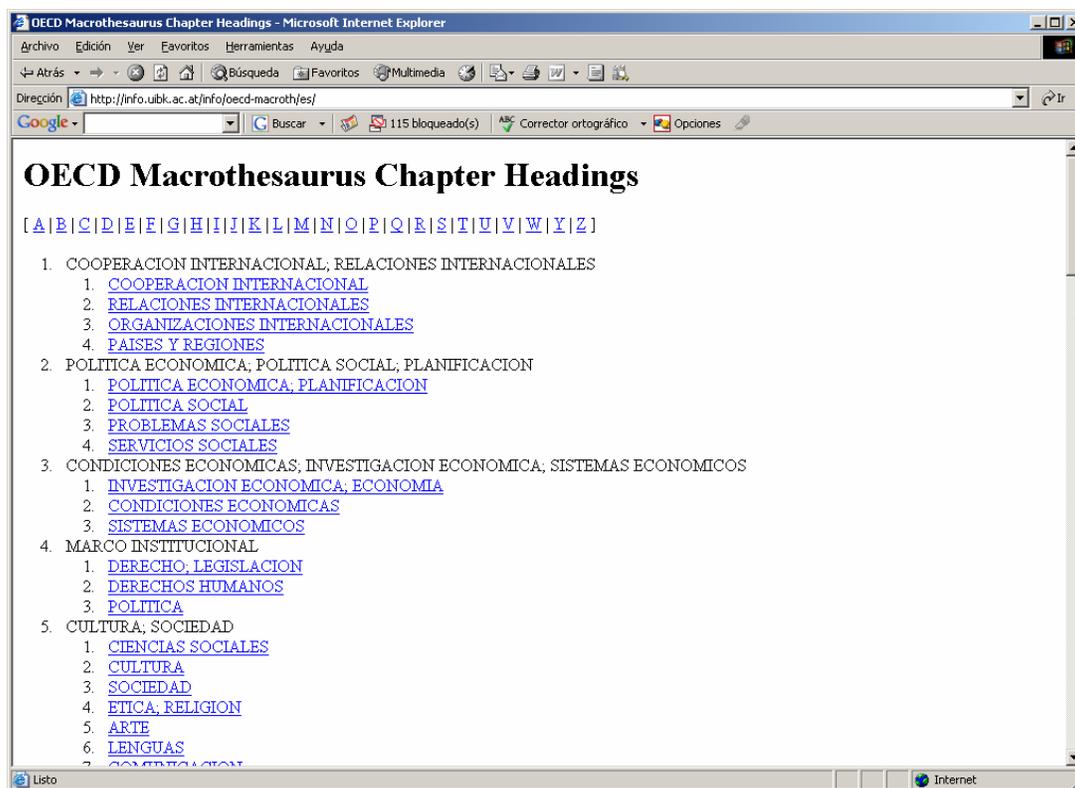


Figura No. 4

Vale mencionar que en el sitio de la OCDE⁷ hacen la siguiente aclaración en relación al macrotesauro: “We regret to inform you that this collection was interrupted in 1998. The last and fifth edition of the *Macrothesaurus for Information Processing in the Field of Economic and Social Development* is available at the Online Bookshop”. También señalan que la preparación de esta quinta edición estuvo a cargo de un comité formado por representantes del Centro Internacional de Investigaciones para el Desarrollo (Canadá), del Departamento de Asuntos Económicos y Sociales de Naciones Unidas, y del Centro de Desarrollo de la OCDE con sede en París.

El área de especialización del macrotesauro es el desarrollo social y económico. En la versión en línea no se menciona el idioma original, aunque como se puede apreciar en la página principal (Figura No. 3), se tiene acceso a la versión en inglés, francés y español. Tampoco se señala el año de creación. A este respecto en el sitio de la OCDE se menciona que la quinta edición del macrotesauro ha sido

⁷ <http://www.oecd.org> [Consulta: 21.08.05]

el resultado del esfuerzo de diversas organizaciones durante casi 30 años para crear un vocabulario común que facilite la indización, la recuperación y el intercambio de información relacionada con el desarrollo económico y social. También se señala que “Efforts have been made to improve the user-friendliness and flexibility of the *Macrothesaurus* by increasing the number of non-descriptors (i.e. cross-references) and scope notes in this edition”.

En cuanto a su facilidad de manejo, y en particular a la disponibilidad y accesibilidad, la consulta en línea no requiere de la instalación de algún programa de cómputo especial. La posibilidad de poder acceder a los términos por orden alfabético sin duda facilita y agiliza la búsqueda.

Al carecer de una introducción, es difícil saber con exactitud su alcance y las características cuantitativas del macrotesauro. Sin embargo, podemos decir que su estructura semántica consta de 19 temas principales (los cuales incluyen 125 subtemas):

OECD Macrothesaurus Chapter Headings:

1. Cooperación internacional; Relaciones internacionales
2. Política económica; Política social, Planificación
3. Condiciones económicas; Investigación económica; Sistemas económicos
4. Marco institucional
5. Cultura; Sociedad
6. Educación; Capacitación
7. Agricultura
8. Industria
9. Comercio
10. Transporte
11. Finanzas públicas; Banca; Relaciones monetarias internacionales
12. Administración; Productividad
13. Trabajo
14. Demografía; Población

15. Biología; Alimentación; Salud
16. Medio ambiente; Recursos naturales
17. Ciencias de la Tierra; Ciencias del espacio
18. Ciencia; Investigación; Metodología
19. Información; Documentación

En relación a la consistencia de los descriptores, en general el uso de singular y plural es el aconsejado por las normas técnicas para la elaboración de tesauros, aunque encontramos algunas inconsistencias como en los descriptores “Agua subterránea” y “Aguas superficiales”, en donde no se aplica de igual manera el criterio del uso del plural. Por otro lado, no hay diferencias tipográficas entre descriptores y no descriptores, pues todos están en mayúscula. De igual modo es de destacar la notación utilizada, pues si bien se emplean las siglas sugeridas en las normas y se incluyen notas de alcance, encontramos algunas no tan usuales, como se muestra en los siguientes ejemplos de un no descriptor y su respectivo envío y de un descriptor.

No descriptor:

OECD Macrothesaurus - RECURSOS HIDRAULICOS

ID: 6337

SI: RECURSOS HIDRAULICOS

UTL: RECURSOS HIDRICOS

OECD Macrothesaurus - RECURSOS HIDRICOS

ID: 2236

PC: RECURSOS HIDRICOS

TM: RECURSOS NATURALES

TE: AGUA DULCE

TE: AGUA POTABLE

TE: AGUA SALADA

TE: AGUA SUBTERRÁNEA

TE: AGUAS SUPERFICIALES

TE: AGUAS TERMALES

TE: RECURSOS HIDRICOS COMPARTIDOS
TR: AGUA
TR: BALANCE HIDRICO
TR: CONSERVACION DEL AGUA
TR: CONTAMINACION DEL AGUA
TR: DERECHOS SOBRE LAS AGUAS
TR: ENERGIA HIDRAULICA
TR: RECURSOS AGRICOLAS
UP: RECURSOS HIDRAULICOS
FA: 16.02.02

Descriptor:

OECD Macrothesaurus - AGUA

ID: 2372

PC: AGUA

TM: FUENTES DE ENERGIA RENOVABLES
TR: ABASTECIMIENTO DE AGUA
TR: ALMACENAMIENTO DE AGUA
TR: CALIDAD DEL AGUA
TR: CONSERVACION DEL AGUA
TR: CONSUMO DE AGUA
TR: CONTAMINACION DEL AGUA
TR: CONTAMINANTES DEL AGUA
TR: DERECHOS SOBRE LAS AGUAS
TR: DISTRIBUCION DEL AGUA
TR: ENERGIA HIDRAULICA
TR: HIDROLOGIA
TR: HIELO
TR: NECESIDADES DE RECURSOS HIDRICOS
TR: ORDENAMIENTO DE LAS AGUAS
TR: QUIMICA DEL AGUA
TR: RECURSOS HIDRICOS
TR: TRATAMIENTO DEL AGUA
TR: UTILIZACION DEL AGUA
FA: 17.05.05

Como lo aconsejan algunas normas, se han empleado los nombres de instituciones de forma abreviada y también se han incluido nombres de países, pero no de personas ni periodos.

En cuanto a su mantenimiento, no se menciona ni la periodicidad ni el método de actualización, aunque por lo señalado en el sitio de la OCDE es de suponer que un comité de asesoramiento formado por diversas organizaciones es quien se ha encargado de actualizar el macrotesauro en su última y quinta edición.

En un plano más terminológico, podemos advertir que este tesauro emplea términos similares a los usados en el vocabulario técnico mexicano, aunque existen algunas diferencias como “Utilización del agua”, que en nuestro país se conoce comúnmente como “Usos del agua”, y como éste existen diversos ejemplos que no viene al caso citar.

Por último, podemos decir que si bien el enfoque del macrotesauro de la OCDE es el desarrollo social y económico, el tema del agua está tratado desde un punto de vista principalmente hidrológico y muy limitado.

En conclusión, si bien esta versión en línea no ofrece una gran cantidad de términos relacionados con los recursos hídricos, consideramos que puede ser un punto de referencia factible de adoptar y adaptar, pues el vocabulario empleado es similar al mexicano y los aspectos sociales y económicos incluidos en algunos temas podrían ser de ayuda para el enfoque holístico que pretendemos tenga el tesauro mexicano; además de aprovechar las equivalencias tanto en inglés como en francés.

Tesaurus de Ingeniería Hidráulica⁸ (España).

Fue elaborado por el Sistema Español de Información sobre el Agua (Hispagua), dentro del convenio de colaboración entre el Centro de Estudios y Experimentación de Obras Públicas (Cedex), la Dirección General de Obras Hidráulicas, del extinto Ministerio de Obras Públicas, Transportes y Medio Ambiente, y personal del Centro de Información y Documentación Científica (CINDOC) del Consejo Superior de Investigaciones Científicas (CSIC).

En cuanto a los elementos considerados para el análisis de la presentación, podemos mencionar que este tesaurus, netamente monolingüe, pues comprende solamente términos en castellano, tiene como área de especialización, como su nombre lo indica, la ingeniería hidráulica. En el sitio no se menciona año de creación ni de última edición, ni tampoco se alude a los derechos de autor. Incluye, sin embargo, un resumen, en el cual se especifica el público al que va dirigido y los beneficios que conlleva el contar con una herramienta terminológica.

Asimismo, comprende una introducción en donde se hace notar la carencia de tesaurus en lengua castellana en temas de ingeniería hidráulica. Además, incluye el alcance, que menciona los campos semánticos comprendidos; su estructura, en donde se especifican las características cuantitativas del tesaurus, así como el método que se llevó a cabo para la selección de términos. De igual manera, se explica detalladamente el proceso de construcción, en donde se indica que se consultó la norma *UNE 50106: 1990*, con objeto de unificar criterios.

También se describe la presentación del tesaurus, el cual aparentemente consta de tres partes diferenciadas: la alfabética, la jerárquica y el índice permutado. Vale mencionar que se ofrece una explicación puntual de lo que es cada parte y se muestra un ejemplo. Sin embargo, en línea sólo se tiene acceso al “Tesaurus jerárquico” y al “Tesaurus conceptual”. El primero está dividido en doce grandes temas; el segundo incluye los términos por orden alfabético con sus relaciones semánticas.

⁸ http://hispagua.cedex.es/documentacion/tesaurus/Tes_hidro/tesaurus.php?localizacion=Tesaurus%20Ingenier%EDa%20Civil
[Consulta: 21.08.05]

En cuanto a su facilidad de manejo, y en particular a la disponibilidad y accesibilidad, la consulta en línea no requiere de la instalación de algún programa de cómputo especial. Por otro lado, si bien su guía de uso es clara y concisa, en la práctica no es así, pues los doce grandes temas no se pueden apreciar en conjunto, ya que para pasar de un tema a otro es necesario recorrer todos los términos incluidos en el tema anterior. Lo mismo sucede con los términos, que están organizados por orden alfabético, pero no se cuenta con una opción que nos permita ir directamente a la letra solicitada. Es decir, si se quieren consultar los términos de la letra “f” hay que recorrer de la letra “a” hasta llegar a la “f”.

En relación con los elementos para el análisis del contenido, en la parte introductoria se menciona que el tesoro tiene un total de 3,105 términos, de los cuales 2,932 son descriptores (indicados con mayúsculas) y 173 no descriptores (indicados con minúsculas); no se hace alusión a la cantidad de notas de alcance o aplicación. En cuanto a la consistencia de los descriptores, podemos señalar que el uso de singular y plural se apega efectivamente a lo establecido por la norma *UNE 50106: 1990*. Por otro lado, si bien no se encontraron identificadores de nombres geográficos, ni de instituciones, ni de personas, ni periodos, hay descriptores como “Agitador Mammoth Rotor Aerator”⁹, que es un nombre comercial.

De su estructura semántica, podemos decir que el esquema de campos temáticos principales que constituye el alcance es el siguiente:

- 01 Abastecimiento y distribución del agua
- 02 Análisis y calidad del agua
- 03 Centrales hidroeléctricas y de bombeo
- 04 Conducciones de agua
- 05 Depuración de aguas residuales
- 06 Gestión y administración del agua

⁹ Mammoth Rotor es un sistema de aeración para el tratamiento de aguas residuales, patentado por la división Zimpro de la empresa US Filter (una compañía de Siemens).

- 07 Hidrogeología - captaciones subterráneas
- 08 Hidrología e hidrografía
- 09 Ingeniería de regadíos
- 10 Ingeniería fluvial
- 11 Presas y embalses
- 12 Saneamiento y alcantarillado

En este tesoro se establecen las relaciones indicadas por las normas técnicas, es decir las jerárquicas, las asociativas y las de equivalencia. En cuanto a las primeras, se menciona que se ha evitado la multijerarquía, de modo que cada descriptor está asignado exclusivamente a un área temática, con lo cual cada término tiene un único término genérico con excepción de los doce términos principales.

De igual forma, se indica que las siglas usadas en esta relación son TG y su inversa TE y para las relaciones asociativas se ha empleado TR. Los no descriptores (señalados con las siglas UP) son relacionados con un término preferido (descriptor) mediante la instrucción USE. “En caso necesario se ha añadido una nota de aplicación (NA) que puede ser de definición o de utilización”. Sin embargo, la simbología utilizada discrepa considerablemente de lo antes mencionado, pues se han usado símbolos como LT, <, .<, ..<, >, .>, ..>, = y =>. En cuanto a las notas de aplicación mencionadas en la introducción, no se encontraron reflejadas en el texto, ya que no se encontró alguna.

En lo que a la periodicidad en su mantenimiento se refiere, no se especifica. Sin embargo, en cuanto a su actualización se menciona que

“Como todo Tesoro, y especialmente los tesauros técnicos o científicos, requieren ser actualizados permanentemente para poderse adaptar al natural nacimiento de nuevos vocablos, fruto del continuo progreso humano, incorporando los nuevos términos y relaciones que se vayan acuñando con el devenir de los tiempos. En este sentido se ruega al lector que nos sugiera cuantas observaciones o comentarios estime oportunos con objeto de mantener y mejorar esta herramienta en revisiones posteriores”.

Llevando a cabo un somero análisis terminológico encontramos algunos vocablos que no se utilizan en el lenguaje técnico en México, como “Faederas” o “Soleras”. Asimismo, se incluye el término “Aguas blancas” para referirse al “agua de lluvia”, locución que se emplea generalmente en nuestro país; lo mismo ocurre con el término “Aguas negras”, que en México se utiliza de manera un tanto cuanto coloquial y ya en un plano más técnico se emplea “aguas residuales”. Y como éstos hay un sin fin de ejemplos, que sería ocioso citar.

En síntesis, este tesoro podría aportar ciertos elementos que servirían de base para la propuesta del tesoro mexicano, como el esquema de campos temáticos principales, aunque hay que tomar en cuenta que el tema se centra principalmente en la ingeniería hidráulica. De igual manera, si bien algunos términos difieren del vocabulario empleado en México, otros sin duda podrían adoptarse.

Tesoro de Ingeniería Sanitaria y Ambiental¹⁰ (América Latina)

Ha sido elaborado por el Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente (CEPIS), mismo que fue fundado en 1968 y tiene su sede en Lima, Perú. El CEPIS forma parte del Área de Desarrollo Sostenible y Salud Ambiental de la Organización Panamericana de la Salud (OPS) y desarrolla sus actividades con el apoyo de la Oficina Regional para las Américas de la Organización Mundial de la Salud (OMS).

El tesoro fue editado por primera vez en 1979 y sus posteriores revisiones han considerado la experiencia lograda con la base de datos bibliográfica de la Red Panamericana de Información en Salud Ambiental (REPIDISCA¹¹), creada en 1982. Como su nombre lo indica, la especialización del tesoro comprende la ingeniería sanitaria y las ciencias ambientales.

¹⁰ <http://www.cepis.ops-oms.org/bvsair/e/manuales/tesa/teses.pdf> [Consulta: 16.07.06]

¹¹ Estas siglas respondían al nombre original de Red Panamericana de Información y Documentación en Ingeniería Sanitaria y Ciencias del Ambiente. Sin embargo, aun cuando se haya adoptado el nombre definitivo de Red Panamericana de Información en Salud Ambiental, conserva las siglas del nombre inicial.

El tesoro ofrece una introducción y una nota de uso. En la primera se señala que para la inclusión de nuevos descriptores se ha tomado en cuenta la terminología de la literatura técnica y, en particular, los términos propuestos por los Centros Cooperantes¹² de la REPIDISCA. También se menciona que para la eliminación de descriptores se ha verificado si el uso de los mismos ha sido nulo, o poco frecuente, y que igualmente se han realizado correcciones en la estructura jerárquica de los descriptores e incluido sinónimos y notas de alcance. “La sección semántica ha sido revisada y presenta modificaciones que permiten un ordenamiento más lógico y funcional para los fines de clasificación”. Cabe resaltar que no se hace alusión alguna a los derechos de autor.

También se indica que el tesoro ha sido compatibilizado con el DECS (Descriptores en Ciencias de la Salud), elaborado por el Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud (BIREME¹³), lo que ha requerido modificaciones y adaptaciones de algunos descriptores. La terminología de salud ambiental está incorporada en el esquema jerárquico de BIREME bajo la categoría SP4.

La versión disponible en línea es la 17ª (2004), y en su introducción se señala que el tesoro incluye 2,082 descriptores y 2,742 no descriptores. Al respecto se menciona que “El incremento sustantivo de sinónimos o no descriptores, se debe a la necesidad de aproximar el lenguaje controlado al lenguaje natural, dado que con la existencia de la Biblioteca Virtual de Salud y Ambiente (BSVA) los usuarios realizan sus búsquedas de información directamente en las bases de datos”. El idioma original es el español, sin embargo, se señala que existe en la base de datos de la REPIDISCA la traducción al portugués, inglés, alemán y francés.

¹² Actualmente existen 352 centros cooperantes distribuidos en Argentina, Bolivia, Brasil, Colombia, Costa Rica, Cuba, Chile, Ecuador, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, Puerto Rico, República Dominicana, Santa Lucía, Uruguay y Venezuela.

¹³ BIREME es un centro especializado de la OPS, establecido en Brasil desde 1967, y trabaja en colaboración con el Ministerio de Salud, el Ministerio de Educación, la Secretaría de Salud del Estado de São Paulo y la Universidad Federal de São Paulo.

En cuanto a la estructura del tesoro, éste presenta una sección semántica y otra alfabética. La primera, que comprende el ordenamiento de descriptores por temas, se ha elaborado tomando como base el esquema de clasificación utilizado en la REPIDISCA y divide el campo de la ingeniería sanitaria y ciencias del ambiente en diez grandes grupos o categorías de materia:

- 0 Salud Ambiental y Saneamiento
- 1 Ingeniería Sanitaria y Ambiental
- 2 Ciencias Aplicadas y del Ambiente
- 3 Recursos Hídricos y Contaminación del Agua
- 4 Abastecimiento de Agua
- 5 Aguas Residuales
- 6 Residuos Sólidos
- 7 Contaminación del Suelo
- 8 Contaminación del Aire
- 9 Salud Ocupacional

Estas diez categorías se encuentran a su vez subdivididas en temas o áreas más específicas, cada una de las cuales se representa bajo códigos de cuatro dígitos. Se menciona que “la estructura de la parte semántica le facilita al indizador la ubicación más precisa del tema que está analizando, a la vez que ofrece en conjunto los distintos descriptores que se adscriben bajo determinado tema, permitiendo la elección del descriptor pertinente”. Es decir, bajo un determinado código se localiza el tema y en orden alfabético los descriptores que le corresponden.

La sección alfabética comprende el listado de los descriptores y sus relaciones, y cada descriptor base va acompañado de un código que lo ubica en la sección semántica. En cuanto al uso del singular y plural, así como los símbolos empleados para representar las relaciones semánticas son las que se sugieren en la normas técnicas para la elaboración de tesoros. Por otro lado, esta edición no incluye nombres geográficos, ni de personas ni periodos, aunque si de

instituciones (curiosamente el único es “REPIDISCA”). Los descriptores se presentan en negritas y los no descriptores en letra cursiva.

En relación a su consulta, no se requiere de la instalación de ningún programa en particular y su presentación en formato *pdf* (portable document format) lo hace práctico, pues también se puede almacenar en un archivo, a diferencia de los tesauros que están disponibles en línea. A este respecto, cabe resaltar que el tesoro en red y el tesoro en formato *pdf* (equivalente a la versión impresa) no tienen la misma funcionalidad.

En cuanto a su mantenimiento, si bien no se indica la periodicidad sí se señala que “El tesoro se encuentra en un periodo de crecimiento, buscando convertirse en una herramienta eficaz para la BVSA, por lo que toda sugerencia será bien recibida”. Vale mencionar que se tuvo también acceso a una versión impresa (8ª, 1994), en la cual se indica que “La actualización del Tesoro requiere del esfuerzo de los Centros Cooperantes de la REPIDISCA, los cuales, con sus sugerencias y aportes sobre la calidad de la terminología, permitirán que el Tesoro se mantenga como un instrumento de trabajo vigente y dinámico. Se espera que los Centros Cooperantes continúen y refuercen permanentemente la actualización de este vocabulario técnico controlado”.

Ya en un plano más terminológico, podemos advertir que este tesoro es de los que más se apega al léxico técnico mexicano, sin duda porque existen algunos Centros Cooperantes de la REPIDISCA en México. Sin embargo, encontramos términos como “represas” para referirse a lo que en nuestro país se conoce como “presas” (muro construido a través de un río, con objeto de regular su caudal o embalsar agua para aprovecharla en el riego o la producción de fuerza hidráulica) o “jacintos”, mejor conocidos en nuestro país como “lirios acuáticos”, sólo por mencionar algunos ejemplos.

En síntesis, consideramos que este tesoro tiene cualidades que no se deben ignorar, pues su estructura es de fácil manejo y su alcance abarca un amplio

espectro dentro de la ingeniería sanitaria y ambiental. Por otro lado, el vocabulario empleado es similar al utilizado en nuestro país, así como sus connotaciones.

Thésaurus Eau (Francia)

Las seis Agencias del Agua¹⁴, la Dirección del Agua del Ministerio de Ecología y del Desarrollo Sustentable (MEDD, por sus siglas en francés), y la Oficina Internacional del Agua¹⁵ (OIEau), que en conjunto forman el “Réseau Fontaine”, han elaborado las diferentes versiones del *Thésaurus Eau* (tesauro del agua), desde 1997. Antes de seguir adelante cabe precisar que actualmente está disponible la versión 2006; sin embargo, la versión a la cual se tuvo acceso para hacer el análisis es la 2001 y sólo a la lista temática, la cual carecía de una introducción. La información que se presenta a continuación se obtuvo del sitio de la OIEau¹⁶, donde se promueve la venta de este documento.

La temática del tesauro, como su nombre lo indica, es el agua. El propósito de desarrollar y elaborar un vocabulario controlado relacionado con los recursos hídricos fue el de contar con una lista de términos normalizados y estructurados, que sirviera como herramienta de indización y de recuperación de información principalmente a los centros de documentación galos relacionados con el medio ambiente y en particular con el agua. El idioma original es el francés y no ofrece ninguna traducción.

El tesauro consta de tres volúmenes: *Liste alphabétique globale* (lista alfabética global), presenta los términos con sus relaciones genéricas y específicas, los reenvíos o términos asociados y sus notas de alcance; *Liste thématique* (lista temática), incluye los términos organizados por temas, en donde un término solamente está ligado a un campo semántico; y *Liste permutée* (índice

¹⁴ Las Agencias del Agua son los equivalentes a los Consejos de Cuenca en México. Para conocer más sobre las Agencias se puede consultar el sitio <http://www.lesagencesdeleau.fr/>

¹⁵ Con sede en Limoges, la OIEau es una asociación sin fines de lucro, cuyo objetivo es crear una red formada por diversos organismos públicos y privados involucrados en la gestión y protección de los recursos hídricos. Se puede consultar su sitio en <http://www.oieau.fr/> [Consulta: 14.07.05]

¹⁶ En este sitio se ofrece información sobre las características principales del tesauro, como son: autores, número de descriptores y precio <http://www.oieau.fr/eaudoc/thesaurus.htm> [Consulta: 07.07.06]

permutado), que contiene todos los términos del tesoro, tanto descriptores como no descriptores, alfabetizados por todas las palabras significativas y presentados sin relación ni reenvío a los campos semánticos.

Podemos decir que la disponibilidad y accesibilidad a este tesoro son limitados, pues no se tiene acceso en línea y si bien se ofrece en versión impresa y en *pdf*, para acceder a ésta es necesario comprar la versión en papel. Según se menciona en el sitio antes citado, los usuarios del tesoro cuentan con ayuda telefónica.

En su versión 2001 el *Thésaurus Eau* incluye 2,452 descriptores (en mayúsculas y negritas) y 914 no descriptores (en mayúsculas y cursivas) contenidos en los siguientes 45 campos temáticos¹⁷, los cuales se han dejado en su versión original. La traducción entre paréntesis es de la que suscribe:

1. Agriculture (agricultura)
2. Aménagement du territoire – Paysage (ordenamiento del territorio – paisaje)
3. Aménagement des eaux (ordenamiento de las aguas)
4. Analyses et tests (análisis y pruebas)
5. Biochimie – Chimie (bioquímica – química)
6. Boues (lodos)
7. Climatologie (climatología)
8. Contrôle qualité – Entreprise (control de la calidad – empresa)
9. Déchets (residuos)
10. Droit (derecho)
11. Elevage – Aquaculture (cría de animales – acuicultura)
12. Energie (energía)
13. Environnement (ambiente)
14. Faune (fauna)
15. Finances – Économie (finanzas – economía)
16. Flore (flora)
17. Génie civil (ingeniería civil)
18. Géologie (geología)
19. Hydraulique – Hydrologie (hidráulica – hidrología)
20. Hydrogéologie (hidrogeología)

¹⁷ La versión 2006 incluye 2,661 descriptores y 946 no descriptores distribuidos en 44 campos temáticos.

21. Industrie (industria)
22. Information – Informatique (información – informática)
23. Mesures et instrumentation (medidas e instrumentación)
24. Méthode et statistiques (método y estadísticas)
25. Microorganismes non pathogènes – Plancton (microorganismos no patógenos – plancton)
26. Milieu naturel (ambiente natural)
27. Paramètres caractéristiques des eaux et des boues (parámetros característicos de aguas y lodos)
28. Politique (política)
29. Politique de l'eau et gestion de l'eau (política hídrica y gestión del agua)
30. Polluants (agentes contaminantes)
31. Pollutions nuisances et prévention (contaminación ambiental y prevención)
32. Procédés communs épuration et traitement (procedimientos comunes de purificación y tratamiento)
33. Procédés d'épuration stricts (procedimientos estrictos de purificación)
34. Procédés de traitement stricts (eau potable) (procedimientos estrictos de tratamiento) (agua potable)
35. Qualité de l'eau (calidad del agua)
36. Réseaux canalisations (redes de canalización)
37. Ressource captage et distribution de l'eau (captación y distribución del agua)
38. Risques (riesgos)
39. Santé – Hygiène – Microorganisme pathogène (salud – higiene- microorganismo patógeno)
40. Sciences exactes (ciencias exactas)
41. Sciences humaines (ciencias humanas)
42. Tourisme – Sport – Loisirs (turismo – deportes – recreación)
43. Transport (transporte)
44. Typologie des eaux (tipología de las aguas)
45. Mot outil (palabras gramaticales)

En relación a la consistencia de los descriptores, en el uso del singular y plural se ha optado por el singular. No se han incluido identificadores geográficos, ni de personas, ni periodos, más sí de instituciones. Además han sido admitidos nombres de convenios internacionales como “Convention de Ramsar” o “Protocole d'Athènes”. También existen términos de programas o acciones locales, representados por sus siglas, como por ejemplo: “SDAU” o “DUP”, el primero se refiere a *Schéma Directeur d'Aménagement et d'Urbanisme* y el segundo a *Déclaration d'Utilité Publique*.

En cuanto a la estructura semántica del tesoro, desconocemos la notación utilizada al no contar con más información que la proporcionada en el sitio de la OIEau. Sin embargo, podemos decir que los signos diacríticos se han omitido.

Referente al mantenimiento se menciona, en el sitio antes citado, que la actualización del tesoro se lleva a cabo anualmente por un grupo de trabajo formado por profesionales de la información de las Agencias del Agua, de la Dirección del Agua del MEDD y de la OIEau. Estos términos son validados por los especialistas de las Agencias y la red de usuarios de la OIEau.

Haciendo un somero análisis terminológico, podemos mencionar que, en general, los términos empleados son similares a los utilizados en México, por ejemplo: "Auscultation de barrage", que en español mexicano se conoce como "inspección de presas". Con la salvedad de que en francés los términos, en su mayoría, están en singular, cuando en español son utilizados en plural.

Si bien el *Thésaurus Eau* ofrece algunos elementos significativos a adoptar, consideramos que el empleo de términos tan precisos, como nombres de programas, acciones o conceptos de uso regional, limitan el uso de este tesoro en nuestro país. Podemos citar como ejemplos: "Directive SEVESO", "écocitoyenneté", "Pavillon Bleu", entre muchos otros.

Sin embargo, es de destacar la lista de campos semánticos. Indudablemente su amplio espectro sirve de base para el enfoque holístico que se quiere dar al tesoro para México, pues abarca el tema del agua desde diferentes perspectivas como la hidráulica, la hidrología, la economía, la sociología, entre otras. Aunque esta lista podría reducirse, pues consideramos que algunos temas podrían ser englobados en uno solo.

En el siguiente y último capítulo analizamos de manera general los seis tesauros. Nuestra intención no es hacer una comparación sino resaltar lo más significativo de cada uno.

Capítulo III

ANÁLISIS DE RESULTADOS

Como se mencionó en la parte introductoria, el objetivo principal de este trabajo estriba en llevar a cabo un análisis descriptivo de seis tesauros relacionados con el tema hídrico, con el fin de probar la necesidad de contar con una herramienta terminológica y determinar si se puede adoptar y/o adaptar algún tesoro existente a la realidad actual mexicana.

Al respecto, basándonos en las características que hemos resaltado de los tesauros analizados en el capítulo anterior, podemos inferir, en una primera instancia, que no es factible adoptar y/o adaptar algún tesoro en particular; no obstante, resulta posible tomar lo más relevante de estos tesauros y adaptarlo a la realidad mexicana.

Si bien es cierto que hay tesauros enfocados de manera general al agua, como el *Thésaurus Eau*, hay que tener presente que éstos, en su mayoría, han sido concebidos con el fin de satisfacer necesidades regionales y una parte de su terminología queda lejos de la realidad de nuestro país. Sin embargo, es innegable que la estructura de los temas del tesoro mencionado nos puede dar la pauta que nos lleve a desarrollar un mapa conceptual como un primer paso en la elaboración de un tesoro del agua para México. Pues como se mencionó en el capítulo anterior, es el único que abarca de una manera holística la temática hídrica, ya que incorpora la hidráulica, la hidrología, la economía, la ecología, la sociología, entre otros temas.

Por otro lado, y antes de llevar a cabo nuestro tercer objetivo específico: establecer las características que debe cumplir un tesoro sobre agua para México a partir de los análisis realizados, consideramos oportuno nombrar algunas particularidades que observamos durante nuestro estudio.

- ◆ El incremento significativo de no descriptores de una edición a otra en algunos tesauros, como en el *Tesaurus de Ingeniería Sanitaria y Ambiental*, refleja la necesidad de aproximar el lenguaje natural al lenguaje controlado. Una razón válida puede ser que en la actualidad la mayoría de los usuarios realizan sus búsquedas de información directamente en las bases de datos. Otra razón podría ser el surgimiento de modismos¹, que en ocasiones están estrechamente relacionados con el cambio del discurso político.
- ◆ Por los temas incluidos en los seis tesauros nos percatamos que en su mayoría el recurso hídrico se contempla básicamente desde el punto de vista de las ciencias exactas, dándole poca importancia a los aspectos sociales y económicos. El mismo Macrotesauro de la OCDE, cuya área de especialización es el desarrollo social y económico, considera al recurso desde la perspectiva de la ingeniería hidráulica.

Al respecto, cabe señalar que aunque las soluciones a los problemas del agua deben incluir aspectos de ingeniería y tecnologías físicas es claro que el del agua, al ser un problema con un origen y una solución sociales, sólo podrá enfrentarse de manera efectiva empleando también la ingeniería social, desde el punto de vista de las ciencias políticas. Es decir, se debe contemplar de igual modo el conocimiento aplicado a la solución de problemas ambientales desde una perspectiva tanto de las ciencias sociales y humanas, como de las consideradas ciencias duras.

- ◆ La convocatoria a los usuarios para proponer nuevos términos demuestra el interés por parte de ciertas instituciones de que el usuario se reconozca en sus necesidades de información. Lo podemos constatar

¹ A buzzword (also known as a fashion word or vogue word) is an idiom, often a neologism, commonly used in managerial, technical, administrative, and sometimes political environments. Buzzwords appear ubiquitously but their actual meanings often remain unclear. Buzzwords are typically intended to create the impression of knowledge for a wide audience. Buzzwords differ from jargon in that they have the function of impressing or of obscuring meaning, while jargon (ideally) has a well defined technical meaning, if only to specialists. However, the hype surrounding new technologies often turns technical terms into buzzwords.
<http://en.wikipedia.org/wiki/Main_Page> [Consulta: 10.07.06]

en el *Agrovoc* y en los tesauros de Ingeniería Sanitaria y Ambiental y de Ingeniería Hidráulica. Pues como se señaló en la parte teórica, el diseño de un lenguaje documental que no tenga en cuenta al usuario estará incompleto y dará origen a un lenguaje deficiente en sus prestaciones a los sistemas de recuperación de información.

- ◆ La falta de una introducción y guía de uso sin duda dificulta y limita el uso del tesoro. Como se ha mencionado anteriormente, la introducción es una parte sustancial de todo tesoro, pues ésta nos informa de su ámbito de aplicación y nos ofrece una explicación sobre sus características distintivas. De los tesauros analizados solamente la mitad de ellos incluye una parte introductoria. Vale hacer notar que en el *Tesoro de Ingeniería Hidráulica* y en el *GeMET* algunos aspectos no coinciden con lo que se menciona en la introducción y lo que se muestra cuando se consulta el tesoro, como por ejemplo las notaciones.
- ◆ El uso de extranjerismos o traducciones deficientes² puede provocar confusiones en el empleo de los términos, lo cual sin duda empobrece la calidad del tesoro y debilita su credibilidad. Un claro ejemplo es el *GeMET*, que incluye algunas traducciones poco comprensibles. De igual manera, dentro de este mismo tesoro encontramos la falta de traducción de algunos términos como “TIME (chronology) [Translation not available]”, que sí cuentan con una equivalencia en castellano.
- ◆ El índice permutado no fue considerado por aquellos tesauros que están disponibles en formato electrónico, al contrario de las listas temáticas y alfabéticas que sí están incluidas en todos los tesauros analizados. Creemos que esto se debe a que en el formato electrónico el buscador de palabras hace las veces de índice, pues dicho buscador permite hacer consultas por cualquiera de las palabras. Cabe recordar que el

² Cuando la aparición de un extranjerismo crea una situación de hecho, Fernández Sevilla (1974: 134) propone dos posibles soluciones: 1) adaptar las características de la lengua al término foráneo; 2) traducirlo.

índice permutado contiene todos los términos del tesoro, tanto descriptores como no descriptores, alfabetizados por todas las palabras significativas que contienen.

- ◆ De los seis tesauros analizados, cuatro hacen mención sobre su mantenimiento y actualización, lo que deja ver la importancia de tomar en cuenta la evolución del lenguaje científico y técnico. Los únicos que no hacen referencia a estas variables son el *GeMET* y el Macrotesauro de la OCDE. Vale mencionar que la actualización del tesoro debe hacerse tanto para incorporar la terminología derivada del desarrollo de la ciencia o disciplina a la que se dedica, como para cubrir lagunas o fallas detectadas durante su utilización, así como para adaptarlo a las necesidades de recuperación manifestadas por los usuarios a través de sus búsquedas.
- ◆ La creciente consolidación y expansión del entorno digital evidencia cada vez más la necesidad de una mayor disponibilidad y accesibilidad a los tesauros en línea, por lo que actualmente es difícil concebir un tesoro sólo en versión impresa.

En este sentido, habrá que recordar que el acceso a la información se ha generalizado y hoy en día es el propio usuario quien busca la información, por lo que debemos insistir en que es necesario diseñar interfaces de búsqueda más amigables, además de que resulta esencial dar cabida a la terminología utilizada por el usuario en su búsqueda de información, razón por la cual cabe explorar nuevas vías para que la descripción documental, a través de la indización, pueda ser sistemática y la recuperación de información sea mucho más flexible y adaptada al usuario. Como señala McCulloch (2005: 406), "...web display of the thesaurus is largely dependent on user requirement".

- ◆ Por último, hemos de mencionar que la falta de normas para publicar tesauros en línea, ya que como se pudo apreciar se pueden encontrar un sin fin de variedades y estructuras, sin duda ocasiona problemas relacionados con la utilidad, interoperabilidad y la forma de compartir el tesoro. Por lo tanto, consideramos esencial una revisión de los formatos y los estándares empleados en los tesauros en línea, con objeto de encontrar una manera de poder armonizar e integrar todos estos aspectos.

Retomando nuestro tercer objetivo y considerando lo expuesto anteriormente, creemos que un tesoro sobre agua para México debe contemplar como mínimo los siguientes aspectos, sin olvidar que será necesaria la participación de especialistas conocedores de la temática hídrica en México, tanto de instituciones de educación superior como de la iniciativa privada y del sector público, así como de bibliotecólogos, lingüistas e informáticos.

- ◆ *Introducción:* tomando en cuenta su importancia, es esencial incluir una introducción en donde se mencione brevemente cual es la intencionalidad del tesoro, instituciones que intervienen en su elaboración, su alcance, número de descriptores y no descriptores, notaciones empleadas, explicación de las relaciones semánticas, método de actualización, mantenimiento, bibliografía, entre otros aspectos. Es muy importante que se incluya una guía de uso claramente detallada, con objeto de que el usuario comprenda el objetivo del tesoro y lo aproveche de manera cabal. En otras palabras, la introducción deberá contener la justificación del tesoro.
- ◆ *Temas:* como se mencionó inicialmente, nuestra intención es que el tema del agua sea tratado de una manera holística, por lo que se deberán tomar en cuenta tanto las ciencias exactas como las sociales y las humanidades, pues la temática hídrica debe abordarse desde una perspectiva multi e interdisciplinaria. Es decir, es necesario reconocer las diferentes perspectivas desde las que se debe considerar al agua, que ha pasado de compartimentos fragmentados de investigaciones científicas a un enfoque

social más global e integrado. Como se señaló en la parte introductoria, la presión cada vez mayor a la que está sometido en México y en el mundo el recurso hídrico, en una situación de cambios antropogénicos y climáticos, exige enfocar de manera multidisciplinaria, integrada y dinámica los problemas científicos y sociales que plantean los recursos hídricos, como lo demuestra la literatura que se ha venido generando en los últimos tiempos.

Conviene insistir en una de las características de los tesauros en general: la división del conjunto de descriptores en campos temáticos es, en cierta medida, arbitraria. Es evidente que algunos descriptores podrían pertenecer a dos o varios campos temáticos. Ahora bien, se admite comúnmente que la correcta gestión de un tesoro y el razonable acotamiento de su volumen imponen una limitación del fenómeno de la polijerarquía, es decir, la inclusión sistemática de cada descriptor en todos los campos a los que podría pertenecer. Por ello, los descriptores que podrían incluirse en dos o varios campos se pueden asignar a uno solo de ellos, que suele ser el que la mayoría de los usuarios considera más natural.

- *Terminología*: el trabajo terminológico a realizar deberá entenderse como una serie de operaciones encadenadas que incluye la detección, recopilación y edición de los términos empleados en México. Sin olvidar que todo trabajo terminológico debe partir de una selección y análisis de la documentación especializada del tema del que trata, pues la terminología cumple sistemáticamente dos funciones: la función de representar el conocimiento especializado, y la de transmitirlo (Cabré, 1999). De igual manera, será necesario tener presente que un tesoro se ofrece como una herramienta documental que, gracias a su característica de normalización del lenguaje específico, permite mejorar la explotación de las bases de datos a través de una doble función: la indización y la recuperación de la información.

Al respecto, cabe recalcar que el usuario que accede a una base de datos buscando información debe preparar una estrategia de consulta documental, por lo que necesita también un traductor entre su propio lenguaje y el vocabulario controlado de la base de datos. La utilización de unos términos específicos, y no otros, para expresar los conceptos de su búsqueda, simplifica el proceso de interrogación. La selección de tales términos se facilita mediante la propia consulta al tesoro, cuya estructura de relaciones va propiciando la traducción al lenguaje documental las necesidades de los usuarios.

- *Evaluación, mantenimiento y actualización:* para verificar su utilidad deberá consultarse a especialistas en el tema y a los usuarios, y se deberá tener en cuenta la continua evolución del lenguaje científico y la aparición de modismos. No hay que olvidar que la efectividad real del tesoro la determinará su uso en la práctica por parte de los indizadores y los usuarios.

El trabajo de mantenimiento y actualización del tesoro debe ser responsabilidad de un comité integrado por un equipo de especialistas, que deberán revisar las sugerencias para tomar la decisión de incorporar o descartar cualquier término. Habrá que tomar en cuenta que el rápido avance del conocimiento demanda una revisión constante, ya que la especialidad de la materia requiere profundidad de indización, por lo que se vuelve obsoleta más rápidamente y se hace necesaria su puesta al día. El objetivo es hacer evolucionar el tesoro de modo que satisfaga al mayor número de usuarios, coordinar las diferentes propuestas y valorar la repercusión de las actualizaciones.

Al igual que en el proceso de elaboración, en la actualización se deberá tomar en cuenta: el grado de evolución del lenguaje; la opinión de los especialistas en la materia; las consultas a los usuarios; y la utilización de

los descriptores en los procesos de indización y recuperación de información.

- *Monolingüe o multilingüe*: el tesoro deberá ser multilingüe, pues en una sociedad globalizada, un tesoro multilingüe ofrece diversas ventajas, como las correspondencias entre conceptos idénticos expresados en diferentes lenguas, pues dichas correspondencias permiten interrogar un sistema documental en la propia lengua del usuario y encontrar los documentos indizados en cualquiera de las lenguas del tesoro. Sin embargo, deberá tenerse especial cuidado con las traducciones deficientes.
- *Accesibilidad y disponibilidad*: como se mencionó anteriormente, en la actualidad es incomprensible concebir un tesoro sólo en versión impresa, por lo que el tesoro deberá también presentarse en versión electrónica. No hay que olvidar que una forma de propiciar el uso de la información en la sociedad es favorecer su disponibilidad y fácil acceso. Es decir, el espíritu de la disponibilidad de información es ofrecer a un usuario todo lo que él requiere para satisfacer sus necesidades, independientemente del lugar donde se encuentre. Disponibilidad que debe estar apoyada en acciones que permitan conocer qué hay y dónde está. Como lo señala Morales (1999), “La participación de la sociedad es definitiva en el proceso de generación y aprovechamiento de la información; ella es la que asigna a ésta valor y función”.

En resumen, consideramos que no es factible adoptar y/o adaptar algún tesoro en particular. Sin embargo, creemos que es posible tomar lo más relevante de los tesoros analizados como punto de partida para elaborar un tesoro que refleje la realidad mexicana.

Por otro lado, la experiencia nos ha mostrado que la literatura actual no sólo considera que el agua dulce es un recurso finito, esencial para todas las actividades humanas y vital para sostener la vida en la Tierra. Además, aparte de sus funciones geofísica, química y biológica en el ciclo hidrológico, cada vez se

genera mayor información que destaca los diversos valores sociales, económicos y ecológicos que tiene el agua y cómo dichos valores están interrelacionados y respaldados mutuamente.

Por lo ya expuesto, consideramos indispensable que el tesoro comprenda igualmente temas que cubran los factores socioeconómicos, ya que el desarrollo humano y económico es sencillamente imposible si no existe un abastecimiento de agua seguro y estable. Asimismo, se deberá tener presente que el agua también tiene un potencial de destrucción. Acontecimientos extremos pueden influir no sólo en la sociedad humana, sino incluso en los medios acuático y terrestre.

De igual manera, la literatura ha hecho hincapié en los crecientes conflictos entre usuarios que compiten por el agua. El despilfarro de los recursos hídricos y su pésima gestión han provocado a menudo el agotamiento de las existencias como la sobreexplotación de las aguas subterráneas y la baja en la disponibilidad de las aguas superficiales. Al mismo tiempo, la contaminación del agua, debida fundamentalmente a actividades humanas, es cada vez más frecuente y más generalizada.

Tampoco podemos pasar por alto las actividades agropecuarias, ya que la agricultura de riego, por ejemplo, es el consumidor más grande de agua en el mundo, con cerca de 70% de todas las extracciones de agua dulce. Razón por la cual en la literatura actual cada vez es más notoria la importancia que se da al uso del agua para la producción de alimentos destinados a una población en constante crecimiento.

En síntesis, consideramos que temas como los arriba mencionados, que podríamos juzgar como estratégicos o prioritarios, reúnen y circunscriben algunas de las temáticas más importantes del mundo del agua. Cabe mencionar que dichos temas deberán ir complementándose y deberán atravesar todas las áreas de trabajo involucradas con el recurso hídrico. Creemos que estos temas podrían servir como punto de partida para la creación de un mapa conceptual con el fin de elaborar un tesoro del agua para México.

*Quand le dernier arbre sera abattu,
la dernière rivière empoisonnée,
le dernier poisson pêché,
alors vous découvrirez que l'argent ne se mange pas.*
Proverbio de los indios Cree (Canadá)

CONSIDERACIONES FINALES

Esta investigación surgió de una serie de reflexiones sobre la necesidad de contar con una herramienta terminológica que dote de coherencia a los procesos de indizar y recuperar información relacionada con el recurso hídrico. La experiencia nos ha enseñado que la carencia de un vocabulario controlado en México sobre el tema del agua, tratado holísticamente, ha provocado que el lenguaje del bibliotecólogo en este campo adolezca de una falta de precisión y del rigor necesario para garantizar la calidad en los procesos de indización y recuperación de información.

Consideramos pertinente, como primera instancia, establecer un marco de referencia para partir de una base que nos permitiera comprender en su totalidad las ventajas que conlleva contar con una herramienta terminológica como es el tesaurus. Dado que ciertos conceptos dentro de la literatura bibliotecológica con frecuencia se entremezclan y confunden, nos dimos a la tarea de precisarlos, lo cual nos permitió llevar a cabo un análisis más meticuloso de los tesauros seleccionados.

De acuerdo a las sugerencias de los estudiosos, nos propusimos analizar seis tesauros relacionados con el agua con objeto de considerar la posibilidad de adoptar y/o adaptar alguno ya existente a la realidad mexicana. Observamos que la mayoría contempla el recurso hídrico básicamente desde un punto de vista de las ciencias exactas, dándole poca importancia a los aspectos sociales y económicos.

Al respecto, no hay que olvidar que la gestión actual del recurso hídrico nos ha llevado a nuevos enfoques de racionalidad hidrológica. Por un lado, en coherencia

con el paradigma de sustentabilidad, lo que ha significado la aportación de términos como “gestión del agua con enfoque de género”, “conflictos por agua”, “gestión de la escasez”, “gobernabilidad”, “gestión ecosistémica”, etc. Por otro, la globalización vigente viene generando un modelo de desarrollo en pro de la mercantilización del agua como recurso y la privatización de los servicios de abastecimiento y saneamiento, lo que ha generado términos como “*stakeholders*”, “bancos de agua”, “mercados del agua”, “derechos de agua”, “privatización del agua”, “economía del agua”, por mencionar sólo algunos.

Sin duda, ciertos términos han surgido como modas y todavía no se ha alcanzado un consenso sobre su definición; otros han sido difíciles de traducir. Un claro ejemplo es el término “*stakeholders*”, cuyo origen en inglés nace en el Derecho y que dentro de la literatura relacionada con la temática hídrica está vinculada con la Administración, refiriéndose a todos los actores sociales que tienen un interés legítimo en la toma de decisiones para la gestión del recurso hídrico. Hemos encontrado que en la literatura en castellano se emplea frecuentemente el término sin traducir.

Vale mencionar que respecto a la adopción de los términos científicos o técnicos de origen extranjero, que el desarrollo de las diversas disciplinas exige poner en circulación, es necesario tomar en cuenta que el término técnico debe ser siempre unívoco y debe eliminarse toda posibilidad de ambigüedad. De igual manera, ha de propiciarse, en lo posible, el fomento de la universalidad del lenguaje científico y técnico, lo cual no se antoja difícil, pues se puede aprovechar esa gran reserva que para la mayoría de las lenguas representan las lenguas clásicas.

Por otro lado, es fácil comprender que la creación de un vocabulario técnico está en relación directa con la creación y el desarrollo de una determinada disciplina, por lo que las nuevas realidades exigen términos que las designen y, a su vez, los nuevos términos tratan de demarcar esta nueva realidad. No hay que olvidar que una disciplina, además de sus fundamentos teóricos y filosóficos, necesita de una

terminología inequívoca entre los pares, para avanzar en sus pretensiones científicas.

De igual manera, los términos técnicos desempeñan un importante papel en el enriquecimiento del vocabulario; casi siempre es por la vía de los vocabularios técnicos y de las lenguas de especialidad por donde se efectúan los préstamos lingüísticos. Por otra parte, es sabido que la renovación del vocabulario está en gran medida determinada por la evolución de las diversas disciplinas.

Tomando en cuenta que por su propia naturaleza el tesoro es una herramienta dinámica, observamos que en un contexto automatizado el uso de un vocabulario controlado resulta adecuado debido a la literalidad con la que los programas informáticos almacenan, organizan y recuperan la información, y porque el tesoro permite combinar los términos en la recuperación. Además, nos percatamos que en una sociedad global de la información, donde la comunicación se manifiesta cada vez más de manera intercultural, el tesoro multilingüe se convierte en una herramienta fundamental para indizar y recuperar la información.

También cabe enfatizar que resulta esencial tomar en cuenta a los usuarios para la inclusión de términos, pues es necesario establecer una dinámica en la que el usuario se reconozca en sus necesidades de información, lo que lo lleva a interactuar mejor con los SRI y a establecer una comunicación sistema-usuario más flexible y transparente, aumentando los niveles de relevancia y pertinencia en la recuperación.

Asimismo, no hay que olvidar que los tesauros son herramientas terminológicas destinadas a ser usadas por usuarios reales con diferentes necesidades y en las más diversas circunstancias, por lo que cuando se ofrece el tesoro en línea no se les debe ignorar en el proceso de diseño del portal, con objeto de facilitar y hacer más amigable la navegación en el sitio.

Por otro lado, observamos que los tesauros deben revisarse regularmente a través de un procedimiento bien organizado y normalizado. De esta manera, la inclusión

de términos está estrictamente controlada y documentada y se evita así cualquier conflicto con la estructura del tesoro y los términos existentes, lo cual permite mantener un control de calidad en la indización y, por ende, en la recuperación de información.

Esta investigación también nos dejó ver que el tesoro admite una mayor especificidad de los términos relativos a campos disciplinares determinados, lo que implica un análisis más preciso de los documentos. De igual modo, permite una incorporación de relaciones asociativas que favorecen la versatilidad y ayudan a crear un universo real e interdisciplinar como es el lenguaje científico.

Es decir, la formulación de relaciones entre los términos que constituyen el lenguaje especializado de una determinada disciplina, conlleva un proceso de interpretación que implica, por un lado, a organizarlos en torno a categorías y, por el otro, a dotar a estas categorías de significados apropiados con el fin de crear un orden lógico que sea capaz de expresar la diversidad de relaciones existentes entre los significados que integran el cuerpo de conocimientos de una o varias disciplinas.

De igual manera, la creciente consolidación y expansión del entorno digital evidencia cada vez más la interrelación de los distintos ámbitos del conocimiento. En el caso de la Bibliotecología, y específicamente en el campo de la representación y recuperación de la información, es indudable la importancia de los aportes de la Lingüística y la Informática. Consideramos que esta vinculación posibilita un redimensionamiento del tesoro como herramienta terminológica para indizar y recuperar la información, con énfasis en la coherencia y riqueza relacional de la estructuración conceptual.

Nuestra reflexión debe partir de una reutilización de los recursos, donde los esfuerzos en recopilar, organizar y sistematizar grandes volúmenes de información relacionada con el agua no se limite a la construcción de un recurso aislado, el tesoro, sino que esta herramienta terminológica sea estructurada de tal forma que pueda ser aprovechada para generar o enriquecer otros recursos. Es en este

sentido que se dirige esta investigación. Pues como se mencionó en la parte introductoria, es necesario sensibilizar a la población sobre la problemática hídrica, ya que la participación ciudadana es básica para la solución de este problema, por lo que es necesario que la información organizada fluya hacia la sociedad con la finalidad de que ésta, en todos sus niveles, tenga un mayor acceso a la información.

En este trabajo solamente hemos ejemplificado una porción de los tesauros que pueden servir de base para la formalización en la elaboración de un tesoro del agua para México. Un campo de investigación futuro será la profundización para concretar el desarrollo de un mapa conceptual que permita ejemplificar los temas a considerar, así como los descriptores y no descriptores, y las relaciones semánticas, propias de un determinado ámbito del conocimiento.

Obras consultadas

- Aguado, Amelia *et al.* (1999). "Definición: nuevas tendencias en tesauros y otros vocabularios controlados". *Representación y organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información* / María José López-Huertas, Juan Carlos Fernández Molina (editores). Actas del IV Congreso ISKO-España EOCOSID'99. (Abril 22-24, 1999 : Granada)., pp. 127-131.
- Aitchison, Jean; Alan Gilchrist, David Bawden. *Thesaurus construction and use: A practical manual*. Chicago : Fitzroy Dearborn, 2000, 218 p.
- Aitchison, Jean; Stella Dextre Clarke (2004). "The thesaurus: a historical viewpoint, with a look to the future". *Cataloging & Classification Quarterly*. 37(3/4) : 5-21.
- Arano Poggi, Silvia (2005). "Los tesauros y las ontologías en la Biblioteconomía y la Documentación". *Hipertext.net*, núm. 3. <http://www.hipertext.net> [Consulta: 20.03.2006]
- Bates, Marcia J. (2002). "After the dot-bomb: getting web information retrieval right this time". http://firstmonday.dk/issues/issue7_7/bates/index.html [Consulta: 13.01.2004]
- Borlund, Pia; Peter Ingwersen (1997). "The development of a method for the evaluation of interactive information retrieval systems". *Journal of Documentation*. 53(3) : 225-250.
- Borri, Dino; Domenico Camarda, Laura Grassini (2005). "Complex knowledge in the environmental domain: building intelligent architectures for water management". *IEA/AIE LNAI 3533* : 762-772.
- Buonocore, Domingo. *Diccionario de bibliotecología*. 2ª ed. aumentada, Buenos Aires : Marymar, 1976, 452 p.
- Cabré, María Teresa. "Terminología y documentación". *La terminología: representación y documentación*. Barcelona : IULA, 1999, pp. 231-247.
- "Importancia de la terminología en la fijación de la lengua: la planificación terminológica". *La terminología: representación y documentación*. Barcelona : IULA, 1999, pp. 295-313.
- Cardero García, Ana María. *Terminología y procesamiento*. México : UNAM, ENEP Acatlán, 2003, 280 p.
- Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente. *Tesauro de Ingeniería Sanitaria y Ambiental*. 8ª ed., Lima : CEPIS, 1994, 152 p.
- *Tesauro de Ingeniería Sanitaria y Ambiental*. 17ª ed., 2004.
<http://www.cepis.ops-oms.org/bvsair/e/manuales/tesa/teses.pdf> [Consulta: 16.07.06]
- Cleveland, Donald B.; Ana D. Cleveland. *Introduction to indexing and abstracting*. 3rd ed., Englewood : Libraries Unlimited, 2001, 283 p.
- Correa Uribe, Guillermo (1999). "Un modelo para la evaluación integral de tesauros". *Revista Interamericana de Bibliotecología*. 22(1) : 139-145.

- Cueva Martín, Alejandro de la. *Acceso y utilización de tesauros en Internet*. <http://docdigital.upf.es/digital/aula2000/aula.htm> [Consulta: 31.10.05]
- Currás, Emilia. *Tesauros: manual de construcción y uso*. Madrid : Kaher II, 1998, 209 p.
- Chaumier, Jacques. *Le traitement linguistique de l'information*. 3^{ème} ed., Paris : Entreprise moderne d'édition, 1988, 186 p.
- Chaumier, Jacques ; Martine Dejean (1990). « L'indexation documentaire : de l'analyse conceptuelle humaine à l'analyse automatique morphosyntaxique ». *Documentaliste – Sciences de l'information*. 27(6) : 275-279.
- Chowdury, Gobinda G. *Introduction to modern information retrieval*. 2nd ed., London : Facet, 2004, 474 p.
- Chu, Heting. *Information representation and retrieval in the digital age*. Medford : American Society for Information Science and Technology, 2003, 248 p. (ASIST Monograph Series).
- Davis, Charles H.; James E. Rush. *Guidance to information science*. Westport : Greenwood, 1979, 305 p.
- European Environment Agency. *General Multilingual Environmental Thesaurus* <http://www.eionet.europa.eu/gemet> [Consulta: 23.08.05]
- Fernández Sevilla, Julio. "El vocabulario científico y técnico". *Problemas de lexicografía actual*. Bogotá : Instituto Caro y Cuervo, 1974, pp. 115-155 (Series Minor; XIX).
- Fidel, Raya (1992). "Who needs controlled vocabulary?". *Special Libraries*. 83(1) : 1-9.
- Foskett, Douglas J. "Thesaurus". *Readings in information retrieval* / Karen Sparck Jones, Peter Willet (editores). San Francisco : Morgan Kaufmann, 1997, pp. 111-134.
- García Jiménez, Antonio (2004). "Instrumentos de representación del conocimiento: tesauros versus ontologías". *Anales de Documentación*. 7 : 79-95.
- García Marco, Francisco Javier. "Clasificación y recuperación de información". *Manual de clasificación documental* / María Pinto (editora). Madrid : Síntesis, 1997, pp. 247-285.
- Gil Urdiciain, Blanca (1998). "Evaluación del rendimiento de tesauros españoles en sistemas de recuperación de información". *Revista Española de Documentación Científica*. 21(3) : 286-302.
- "Los lenguajes de clasificación documental". *Manual de clasificación documental* / María Pinto (editora). Madrid : Síntesis, 1997, pp. 91-108.
- Gilchrist, Alan (2003). "Thesauri, taxonomies and ontologies – an etymological note". *Journal of Documentation*. 59(1) : 7-18.

Guinchat, Claire; Michel Menou. *Introducción general a las ciencias y técnicas de la información y documentación*. 2ª ed. corregida y aumentada por Marie-France Blanquet, Madrid : UNESCO, 1990, 555 P.

Harrison, Lauren D. *The impact of thesauri on information retrieval*. New Brunswick : L.D. Harrison, 1998. Dissertation (Doctor of Philosophy) – Rutgers, The State University of New Jersey, 102 p.

Holzem, Maryvonne. *Terminologie et documentation: pour une meilleure circulation des savoirs*. Paris : ADBS, 1999, 292 p. (Recherches et documents).

Hudon, Michèle (1997-1998). « Indexation et langages documentaires dans les milieux archivistiques à l'ère des nouvelles technologies de l'information ». *Archives*. 29(1) : 75-98.

Iglesias Maturana, María Texia (1997). "Beneficios de la utilización de lenguajes controlados en el análisis y recuperación de la información". *Investigación Bibliotecológica: archivonomía, bibliotecología e información*. 11(23) : 32-45.

International Organization for Standardization. *ISO 2788-1986(E) Documentation – Guidelines for the establishment and development of monolingual thesauri*. 2nd ed., Geneva : ISO, 1986, 32 p.

Jorna, Kerstin; Sylvie Davies (2001). "Multilingual thesauri for the modern world – no ideal solution?". *Journal of Documentation*. 57(2) : 284-295.

Kharraz, Jauad El. "Criteria for the selection of an inland water thesaurus and glossary water thesaurus". Système euro-méditerranéen d'information sur les savoir-faire dans le domaine de l'eau (SEMIDE).
http://www.semide.org/documents/pdf/20041124_JeK_thesaurus.pdf
[Consulta: 30.08.05]

Kline, Vickie (2002). "Missing links: the quest for better search tools" *Online Information Review*. 26(4) : 252-255.

Lafuente López, Ramiro (2000). "El estudio del lenguaje de la bibliotecología y disciplinas afines". *Investigación Bibliotecológica: archivonomía, bibliotecología e información*. 14(29) : 130-150.

Lancaster, Frederick W. *Vocabulary control for information retrieval*. 2nd ed., Arlington : Information Resources Press, 1986, 270 p.

--- (2003). "Do indexing and abstracting have a future?" *Anales de Documentación*. 6 : 137-144.

Lee-Smeltzer, Kuang-Hwei (2000). "Finding the needle: controlled vocabularies, resource discovery, and Dublin Core". *Library Collections, Acquisitions & Technical Services*. 24(2) : 205-215.

Library and Archives Canada
<http://www.collectionscanada.ca/index-e.html> [Consulta: 02.12.05]

López Alonso, Miguel Ángel (1999). "Los tesauros conceptuales como herramientas de precisión en los sistemas de organización científica". *Revista Interamericana de Bibliotecología*. 22(1) : 21-35.

López-Huertas Pérez, María José (1998). "La representación del usuario en la recuperación de la información". VI Jornadas Españolas de Documentación. http://fesabid98.florida-uni.es/Comunicaciones/mj_lopez/mj_lopez.htm
[Consulta: 27.02.04]

----- (1997). "Thesaurus structure design: A conceptual approach for improved interaction". *Journal of Documentation*. 53(2) : 139-177.

----- "Potencialidad evolutiva del tesoro: hacia una base de conocimiento experto". *Representación y organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información* / María José López-Huertas, Juan Carlos Fernández Molina (editores). Actas del IV Congreso ISKO-España EOCONSID'99. (Abril 22-24, 1999 : Granada)., pp. 133-144.

López Yepes, José (editor). *Diccionario enciclopédico de ciencias de la documentación*. Madrid : Síntesis, 2004.

Maniez, Jacques. *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid : Fundación Germán Sánchez Ruipérez, 1992, 230 p. (Biblioteca del libro; 56).

Mandala, Rila; Takenobu Tokunaga, Hozumi Tanaka (2000). "Query expansion using heterogeneous thesauri". *Information Processing and Management*. 36 : 361-378.

Martínez Arellano, Filiberto Felipe (2001). "Enseñanza en la recuperación y acceso por tema en las escuelas de bibliotecología mexicanas". *IFLA Council and General Conference* (67 : August 16-25 : Boston).

Martínez de Sousa, José. *Diccionario de bibliología y ciencias afines*. Madrid : Fundación Germán Sánchez Ruipérez, 1989, 852 p. (Biblioteca del libro).

Martínez Méndez, Francisco Javier; José Vicente Rodríguez Muñoz (2004). "Reflexiones sobre la evaluación de los sistemas de recuperación de información: necesidad, utilidad y viabilidad". *Anales de Documentación*. 7 : 153-170.

McCulloch, Emma (2005). "Thesauri: Practical guidance for construction". *Library Review*. 54(7) : 403-409.

Miller, Uri. "Thesaurus construction". *Encyclopedia of Library and Information Science*. New York : Marcel Dekker, 2003, pp. 2800-2810.

----- "Thesaurus and new information environment". *Encyclopedia of Library and Information Science*. New York : Marcel Dekker, 2003, pp. 2811-2819.

Morales Campos, Estela. *El derecho a la información y las políticas de información en América Latina* (65th IFLA Council and General Conference – Thailand, 1999). <http://www.ifla.org/IV/ifla65/papers/056-137s.htm>. [Consulta: 02.12.04].

Naumis Peña, Catalina. “Dinámica en la construcción de tesauros y el acceso a la información”. *La información en el inicio de la era electrónica: organización del conocimiento y sistemas de información*. México : UNAM. CUIB, 1998, pp. 69-90.

---- (2000). “Análisis de la confluencia entre término y descriptor en la elaboración de tesauros”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*. 14(29) : 95-113.

---- Modelo de construcción de tesauros documentales multimedia. Aplicaciones a los contenidos educativos en televisión. Madrid : Universidad Complutense de Madrid, 2002.

Naumis Peña, Catalina; María Texia Iglesias Maturana. “Un lenguaje para la organización del conocimiento bibliotecológico en América Latina”. *Representación y organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información* / María José López-Huertas, Juan Carlos Fernández Molina (editores). Actas del IV Congreso ISKO-España EOCONSID'99. (Abril 22-24, 1999 : Granada)., pp. 81-86.

National Information Standards Organization. *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*. Bethesda : NISO, 2003, 69 p. (ANSI/NISO Z39.19).

Naves Martins Lopes, Madalena (2002). “El trabajo de los indizadores: factores que afectan el análisis de contenido”. *SCIRE*. 8(1) : 119-130.

OECD Macrothesaurus
<http://info.uibk.ac.at/info/oecd-macroth/> [Consulta: 21.08.05]

Office International de l'Eau. *Thésaurus Eau*. 2001.

Organización de las Naciones Unidas para la Agricultura y la Alimentación. *Agrovoc*. http://www.fao.org/aims/ag_intro.htm [Consulta: 07.08.06]

Owens, Leslie Ann; Pauline Atherton Cochrane (2004). “Thesaurus evaluation”. *Cataloging & Classification Quarterly*. 37(3/4) : 87-102.

Park, Hongseok (1997). “Relevance of science information: origins and dimensions of relevance and their implications to information retrieval”. *Information Processing and Management*. 33(3) : 339-352.

Pérez Agüera, José Ramón (2004). “Automatización de tesauros y su utilización en la web semántica”. *Textos universitarios de biblioteconomía i documentació*. 13
http://www2.ub.es/bid/consulta_articulos.php?fichero=13perez2.htm [Consulta: 06.10.05]

Portal del @gua UNAM.
<http://www.agua.unam.mx/default.aspx> [Consulta: 25.08.06]

Rodríguez Muñoz, Vicente (1997). "Un modelo de datos para la construcción de tesauros". *Investigación Bibliotecológica: archivonomía, bibliotecología e información*. 11(22) : 39-50.

Salvador Oliván, José Antonio; Rosario Arquero Avilés (2004). "La investigación en recuperación de información: revisión de tendencias actuales y críticas". *Cuadernos de documentación multimedia*. No. 15.
<http://multidoc.rediris.es/cdm/> [Consulta: 23.12.05]

Saracevic, Tefko (1975). "Relevance: A review of and a framework for the thinking on the notion in Information Science". *Readings in information retrieval* / Karen Sparck Jones, Peter Willet (editores). San Francisco : Morgan Kaufmann, 1997, pp. 143-165.

Shearer, James R. (2004). "A practical exercise in building a thesaurus". *Cataloging & Classification Quarterly*. 37(3/4) : 35-56.

Shiri, Ali Asghar; Crawford Revie (2000). "Thesauri on the Web: current developments and trends". *Online Information Review*. 24(4) : 273-279.

Sihvonen, Anne; Perti Vakkari (2004). "Subject knowledge improves interactive query expansion assisted by a thesaurus". *Journal of Documentation*. 60(6) : 673-690.

Sistema Español de Información sobre el Agua. *Tesauro de Ingeniería Hidráulica*.
http://hispagua.cedex.es/documentacion/tesauro/Tes_hidro/tesauro.php?localizacion=Tesauro%20Ingenier%EDa%20Civil [Consulta: 21.08.05]

Slype, Georges van. *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid : Fundación Germán Sánchez Ruipérez, 1991, 200 p. (Biblioteca del libro).

Svenonius, Elaine. "Design of controlled vocabularies". *Encyclopedia of Library and Information Science*. New York : Marcel Dekker, 2003, pp. 822-838.

Thomas, Alan R. (2004). "Teach yourself thesaurus: exercises, readings, resources". *Cataloging & Classification Quarterly*. 37(3/4) : 23-34.

Tudhope, Douglas; Harit Alani, Christopher Jones (2001). "Augmenting thesaurus relationships: Possibilities for retrieval". *Journal of Digital Information*. Volume 1, Issue 8, Article No. 41, 25 p.
<http://jodi.tamu.edu/Articles/v01/Tudhope> [Consulta: 16.11.05]

Vizcaya Alonso, Dolores (1995). "La terminología: disciplina indispensable en la formación de los especialistas en lenguajes de indización". *Ciencias de la información*. 26(1) : 33-34.

Water Librarian's Home Page

<http://www.interleaves.org/~rteeter/waterlib.html> [Consulta: 27.09.05]

ANEXO

Elementos considerados para el análisis de los tesauros seleccionados

1. Elementos para el análisis de **contenido**. Alude a la cobertura de la temática y a su nivel de estructuración.
 - 1.1. Características cuantitativas del tesoro: advierte sobre la cuantificación de temas, descriptores, no descriptores y jerarquías incluidos.
 - 1.1.1. Cantidad de campos temáticos: determina el número de temas en los que se divide el tesoro.
 - 1.1.2. Cantidad total de términos: se refiere a los términos como unidades lingüísticas que pueden ser examinadas lingüística y gramaticalmente, y que por su función se pueden agrupar en simples, compuestos, identificadores e indicadores de función.
 - 1.1.3. Cantidad de descriptores: refiere a la cantidad de términos preferentes incluidos y si éstos son diferenciados de alguna manera (mayúsculas, cursivas, etc.) de los no descriptores.
 - 1.1.4. Cantidad de no-descriptores: refiere a la cantidad de términos equivalentes de carácter no preferente y, al igual que en el inciso anterior, si éstos se diferencian de alguna forma (minúsculas, negritas, etc.).
 - 1.1.5. Cantidad de notas de alcance o de aplicación: alude a la inclusión de aclaraciones sobre el sentido del descriptor (nota de definición) y se orienta sobre cómo emplearlo en la indización y en la recuperación de la información (nota de aplicación).
 - 1.1.6. Lista jerárquica: alude al número máximo de jerarquías contempladas.
 - 1.2. Consistencia de los descriptores: refiere a la coherencia en la elección de los términos que integran el tesoro.
 - 1.2.1. Relaciones recíprocas: refiere a la relación que los descriptores deben de mantener con los demás, ya que si no existe reciprocidad disminuyen las posibilidades de que se seleccionen los descriptores más adecuados, tanto en la indización como en la recuperación de la información.

- 1.2.2. Uso de singular / plural: advierte su uso de acuerdo a las normas establecidas.
 - 1.2.3. Nombres geográficos, de instituciones, de personas y fechas: refiere a la inclusión de identificadores geográficos, instituciones, personas o fechas relevantes y pertinentes a la temática abordada.
 - 1.2.4. Traducción a otro(s) idioma(s): advierte la exactitud de la equivalencia de términos en otras lenguas y también implica reconsiderar el uso de singular sobre el plural o viceversa.
- 1.3. Estructura semántica del tesoro: alude a la división temática del tesoro.
- 1.3.1. Campos semánticos incluidos: refiere al ordenamiento de los descriptores por temas y se analiza la disección y organización de la terminología.
 - 1.3.2. Notación utilizada para designar las relaciones semánticas (relaciones jerárquicas, asociativas y de equivalencia): identifica las abreviaturas empleadas para designar las relaciones existentes entre descriptores y describe si son abreviaturas o signos.
- 1.4. Mantenimiento: alude a la periodicidad y forma de actualización del tesoro.
- 1.4.1. Periodicidad: indica cada cuándo se actualiza.
 - 1.4.2. Método de actualización: analiza los criterios utilizados para añadir o sustraer descriptores (consulta en diccionarios especializados, a especialistas y aceptación de propuestas de los usuarios).
2. Elementos para el análisis de la **presentación**. Se refiere a las características distintivas de cada tesoro y a la forma como éstas son tratadas.
- 2.1. Nombre del tesoro: advierte si el nombre refleja realmente la temática abordada.
 - 2.2. Nombre de la institución(es) responsable(s) de la elaboración: respalda la labor de sus creadores.
 - 2.3. Año de creación: se refiere a la fecha de la primera edición del tesoro.
 - 2.4. Año de última edición: se refiere a la versión que se está revisando, con objeto de considerar la actualidad de los términos.

- 2.5. Área de especialización: alude al campo temático o dominio de aplicación.
- 2.6. Introducción: explica de manera sintética los antecedentes del tesoro, así como su alcance, el empleo de normas técnicas reconocidas, el proceso de construcción, su estructura, guía de uso, actualización, etc.
- 2.7. Origen de creación (a priori / a posteriori – método deductivo / método inductivo): identifica si la concepción del tesoro fue previa a su aplicación a un ámbito documental determinado o fue creado sobre la base de dicho ámbito y cómo fue el proceso de elaboración.
- 2.8. Derechos de autor: refiere al permiso necesario para usar el tesoro.
- 2.9. Estructura del tesoro: identifica la inclusión de las partes básicas de todo tesoro:
 - 2.9.1. Lista alfabética: se determina si los descriptores cuentan con relaciones completas (jerárquicas, asociativas y de equivalencia) o parciales.
 - 2.9.2. Lista temática (o jerárquica): se determina si esta parte del tesoro cuenta con un esquema lógico de agrupación y criterios establecidos para las relaciones entre los descriptores.
 - 2.9.3. Índice permutado: se determina si incluye todos los términos del tesoro, tanto descriptores como no descriptores, alfabetizados por todas las palabras significativas que contienen y si los descriptores aparecen acompañados de un código que los relaciona con la parte jerárquica, mientras que los no descriptores carecen de dicho código.
- 2.10. Tipo de jerarquías: refiere si las jerarquías empleadas son facetadas o por clases.
 - 2.10.1. Presentación jerárquica: describe si la presentación es por sangrías, grupos, flechas o combinaciones de estas formas.
- 2.11. Monolingüe o multilingüe: advierte sobre la conveniencia de poder recuperar los documentos indizados en cualquiera de las lenguas del tesoro.
 - 2.11.1. Idioma original: considera que el idioma fuente no debe dominar sobre los otros.

2.11.2. Monolingüe con traducción: indica que los términos en el idioma fuente sólo están traducidos a otro idioma, pero no incluye una sección en esa lengua.

3. Elementos para el análisis de la **consulta** del tesoro. Considera la facilidad de manejo.

3.1. Tipo de soporte: refiere al medio en el que se encuentra disponible el tesoro.

3.2. Disponibilidad y accesibilidad: analiza si el tesoro está disponible en línea o sólo en versión impresa o ambas, si el acceso tiene un costo o es gratuito, etc. Tomando en cuenta que la accesibilidad puede estar limitada a contar con los requerimientos propios de cada programa que permitan visualizar su contenido.

3.3. Claridad de la guía de uso: advierte sobre las explicaciones de la estructura del tesoro y de las notaciones utilizadas.