

¿Desalación o desalinización? ¿Agua dulce o agua fresca?... o la necesidad de contar con un vocabulario controlado relacionado con el tema hídrico.

Verónica Vargas Suárez
Patricia Navarro Suástegui

Instituto Mexicano de Tecnología del Agua, Jiutepec, México

Resumen

El propósito de este trabajo es compartir las experiencias que se han vivido durante la indización de material documental relacionado con los recursos hídricos, éstos vistos de manera holística, es decir abarcando disciplinas de las ciencias exactas y de las ciencias sociales. Se inicia este trabajo con un somero análisis sobre los vocabularios controlados, destacando su importancia como instrumento eficiente para indizar y recuperar la información. Posteriormente presentamos la problemática a la que nos enfrentamos en nuestra labor cotidiana al emplear diversos tesauros para el proceso de indización, por carecer de una única herramienta terminológica. En seguida, exponemos la solución que hemos encontrado a esta laguna y la manera cómo hemos ido creando nuestro catálogo de autoridades de materia.

Introducción

Para entrar en contexto hemos de mencionar que laboramos en un centro de información especializado en el recurso hídrico. La experiencia nos ha mostrado que la carencia de un vocabulario controlado en español sobre la temática hídrica, tratada holísticamente, ha provocado que no podamos garantizar la calidad idónea en los procesos de indización, y por ende, de recuperación de información. Y por si esto fuera poco, la práctica también nos ha enseñado 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 término “gestión integrada de recursos hídricos”, del cual hemos encontrado hasta doce variaciones y así podríamos citar un sinfín de vocablos más, a pesar de que “una de las características más importantes de un texto especializado es la presencia de unidades terminológicas, cuanto más elevado sea el nivel de especialización de un texto, más alta será su densidad terminológica. Normalmente, un texto con un alto nivel de especialización es preciso, conciso y sistemático; la terminología que se utiliza en este texto tiene tendencia a la monosemia y a la univocidad.” (Cabré, 2002).

Antes de entrar en materia, consideramos pertinente mencionar algunos aspectos sobre los vocabularios controlados para valorar su importancia en el proceso de indización y, por consiguiente, de recuperación de información. Pues como lo menciona Gorman (citado por Martínez Arellano, 2001) “*The reason is that effective cataloguing involves controlled*

vocabularies...”, y lo confirma Martínez Arellano al mencionar que los vocabularios controlados son una herramienta importante de recuperación de información y, por lo tanto, juegan un papel relevante al organizar el conocimiento.

Como todos sabemos, los vocabularios controlados son usados como una especie de filtro entre el lenguaje utilizado por el autor y la terminología⁴⁴ 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. En síntesis, el propósito de todo vocabulario controlado es optimizar la capacidad de precisión y acierto del lenguaje de indización.

Hay que añadir que en razón de su propia naturaleza, los vocabularios controlados 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 ruido⁴⁵, y la sinonimia, que está relacionada con el factor de acierto y el silencio.

Como podremos apreciar más adelante, los vocabularios controlados son importantes por diversas razones, pues tanto en el proceso de indización como en el proceso de búsqueda de información nos proporcionan consistencia, precisión, automatización, simplificación, interoperabilidad y nos permiten un uso eficiente del tiempo (Bermúdez *et al.*, 2011).

Para fines de este trabajo entendemos por vocabulario controlado una lista normalizada y organizada de términos, administrada de tal manera que incluye una selección cuidadosa y un mantenimiento constante, es decir la adición, el descarte y la actualización, con el propósito de optimizar el proceso de recuperación de información reduciendo la incertidumbre entre las palabras que elige el catalogador para representar un texto y las que emplea el usuario para recuperarlo.

Por último, hemos de mencionar que la indización 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, sin olvidar que esos conceptos deben ser lo que más apropiadamente caractericen al documento, de modo que facilite al usuario su recuperación. En síntesis, la indización es el proceso intelectual fundamental para toda recuperación de información.

A continuación explicamos las dificultades a las que nos enfrentamos en nuestro trabajo como catalogadoras, en lo que a la indización de material documental especializado en el recurso hídrico se refiere; asimismo mostramos la alternativa que hemos encontrado

⁴⁴ Como lo señala Cabré, “la terminología es una forma de representar la realidad especializada”.

⁴⁵ 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 y que sin embargo no se recuperan porque no ha habido coincidencia entre los términos que empleó el catalogador y los que utiliza el usuario durante la búsqueda de información.

para facilitar y darle mayor calidad a nuestro trabajo de indización, sabiendo las repercusiones que tiene para el usuario en la recuperación de información.

Problemática en la indización de material documental relacionado con el agua

Consideramos oportuno mencionar que a diferencia de otros vocabularios controlados, los tesauros suelen ser lenguajes especializados, circunscritos a una ciencia o conjunto de ciencias en particular. Sin duda, en el sector hídrico y con referencia al idioma español, se han elaborado tesauros de carácter general con énfasis en determinadas disciplinas, como el *Tesaurus de Ingeniería Sanitaria y Ambiental* o el *Tesaurus de Ingeniería Hidráulica*, los cuales, por su propia naturaleza, atienden la demanda de información de un grupo determinado de usuarios.

Como se ha aludido anteriormente, la mayoría de la literatura que indizamos está especializada en el recurso hídrico y para ello empleábamos los siguientes tesauros: *Tesaurus de Ingeniería Sanitaria y Ambiental*, de la Organización Panamericana de la Salud; *Tesaurus de Ingeniería Hidráulica*, del Ministerio de Agricultura, Alimentación y Medio Ambiente de España; *AGROVOC Thesaurus*, de la FAO; *InterWater Tesaurus*, del Centro Internacional de Agua Potable y Saneamiento; y *General Multilingual Environmental Thesaurus* (GeMET), de la Agencia Europea de Medio Ambiente.

Es necesario señalar que existen otros dos tesauros que también podrían ser de utilidad: *Thésaurus Eau* (última versión 2006), de la Oficina Internacional del Agua y *EnVoc Multilingual Thesaurus of Environmental Terms* (última versión 1997), del Programa de Naciones Unidas para el Medio Ambiente (PNUMA). Sin embargo estos tesauros no se encuentran en línea, por lo que su consulta nos resulta imposible, amén de mencionar la falta de actualización del tesoro del PNUMA.

A medida que crece un sistema de información sin duda aumenta la necesidad de un control más estricto de los términos que traduzcan el contenido de la materia indizada y, como mostraremos más adelante, el emplear diversos tesauros nos resultó poco práctico. Antes, queremos destacar que gran parte del material documental que se recibe en el centro de información está en inglés. Por otro lado, es necesario enfatizar que la gestión del recurso hídrico no es tarea de una sola disciplina, es multi e interdisciplinaria, por lo que la indización de este material nos obliga a contar con una herramienta terminológica que abarque el tema de una manera holística.

Somos conscientes que cada tesoro presenta sus propios criterios de elaboración, su propia extensión, organización y riqueza, así como su funcionalidad. Sin embargo el tener que hacer uso de varios tesauros a la vez representó un mayor grado de dificultad, pues los descriptores extraídos pueden presentar inexactitudes semánticas o discrepancias en el uso de singulares y plurales o en las traducciones, por mencionar algunas diferencias. Lo anterior repercute sin duda en la recuperación de información por parte del usuario, por lo

que teníamos que ser sumamente cuidadosas al asignar un descriptor, para no cometer el error de incluir un término equivalente a uno ya registrado.

Como ejemplo podemos mencionar algunos términos que aparecen de manera recurrente en la literatura, como “Gestión del agua”. No todos los tesauros antes mencionados lo incluyen, es más, algunos lo validan como no descriptor o tiene una connotación diferente a la que se conoce en México. Otro ejemplo es la expresión “Tratamiento aerobio”, la cual en algunos tesauros aparece como “Tratamiento anaeróbico”. También nos encontramos que en los tesauros se incluye un término, pero en la literatura especializada mexicana se emplea de manera distinta, como es el caso de la palabra “desalinización”, que en los tesauros el descriptor a emplear es “Desalación”.

Indudablemente estas situaciones nos llevaron a elegir alguno de ellos corriendo el riesgo de cometer una imprecisión en relación con la connotación dada en el documento que se está indizando. Es decir, estas diferencias, que uno pudiera considerar como sutiles, tienen varias implicaciones durante el proceso de indización y, por consiguiente, en el de recuperación de la información, pues, sin lugar a dudas, los usuarios podrían tener problemas, por un lado, para identificar el término apropiado asignado a un concepto y, por el otro, para discernir cómo ese concepto está relacionado con otros.

De igual manera, podemos referirnos a las traducciones. Cabe recordar que una buena parte del material documental que se indiza está en inglés, por lo que el uso de tesauros bilingües fue de gran ayuda. Empero, como lo señala Turuguet (1994) “Las traducciones crean problemas adicionales; a menudo un término del tesauro original da lugar a dos o más traducidos, es difícil tener equivalencias perfectas y quizá algunos términos sean intraducibles.”

Al respecto, nos encontramos con traducciones literales como “Multietapa agita el separador”, refiriéndose a “*Multi-stage swirl separator*”, o esta otra “*Boom del petróleo*” (*Oil boom*), aludiendo a la barrera de contención que se emplea para controlar los derrames de petróleo, y no al auge petrolero, como la mayoría de los mexicanos comprendemos la expresión “Boom”.

Por último, hemos de señalar las omisiones en los tesauros. Hoy en día existen algunos términos muy recurrentes en la literatura especializada como “Cambio climático”, “Gobernanza del agua”, por mencionar sólo unos cuantos. Sin embargo notamos que la mayoría de los tesauros no incluyen estas locuciones.

Indudablemente todas estas situaciones complican la tarea de indizar y, en consecuencia, pueden dificultar la recuperación de la información por parte del usuario. Razón por la cual nos hemos dado a la tarea de ir construyendo nuestro propio catálogo de autoridades de materia, como explicamos a continuación.

Catálogo de autoridades de materia.

Como respuesta a la problemática mencionada anteriormente, en 2011 comenzamos a construir una lista de términos, basándonos principalmente en la revisión de las palabras clave incluidas en publicaciones periódicas mexicanas relacionadas con el tema del agua y que la misma experiencia nos ayudó en su selección; la validación de dichas palabras clave en tesauros y diccionarios especializados; y la consulta a especialistas de las diferentes áreas involucradas.

Inicialmente la lista incluía básicamente términos validados y algunos envíos. Con el tiempo, esta lista se ha ido convirtiendo en nuestro catálogo de autoridades de materia. Es decir, los términos de la lista se han ido registrando en el sistema que se utiliza en la biblioteca para la automatización de sus actividades, lo que nos ha facilitado la indización al tener una validación automática. La siguiente figura nos muestra un ejemplo de un término registrado en el catálogo de autoridades de materia.



Somos conscientes que es necesario comprender el proceso del desarrollo y administración de un vocabulario controlado, el cual incluye (Isenor, 2011):

1. Definir claramente la necesidad para desarrollar un nuevo vocabulario controlado y determinar sus requerimientos específicos.
2. Aprovechar la experiencia de especialistas para evaluar la inclusión de cada término (cabe preguntarse ¿Es el término ampliamente utilizado? ¿Tiene un significado apropiado para la comunidad científica?).
3. Revisar meticulosamente dicho vocabulario y darle un formato adecuado.

4. Registrar cada término con una organización apropiada.
5. Evaluar el vocabulario controlado en comunidades científicas y tomar en cuenta la retroalimentación con el fin de mejorarlo.

Sin duda, la creación de un registro de autoridad de materia ha implicado un proceso de investigación sobre la autoridad que se quiere validar, para asegurarnos de su consistencia y unificación. Es decir se corrobora que el término que se va a incluir esté perfectamente identificado y reúna todas las variantes que pueda presentar; se verifica que el término no pueda ser confundido con otro y, por último, se recopila toda la información que sirva para identificarla. Dos buenos ejemplos son “calentamiento global” y “cambio climático” o “desarrollo sostenible” y “desarrollo sustentable”, que en ocasiones se emplean como si fueran sinónimos, cuando cada uno de estos términos tiene su propio significado.

Por otro lado, hemos utilizado las *Directrices para los registros de autoridad y referencia de materia*, de la Federación Internacional de Asociaciones e Instituciones Bibliotecarias (IFLA, por sus siglas en inglés) y el formato MARC 21 para registros de autoridad, como instrumentos normativos para llevar a cabo el control de autoridades. Las etiquetas de MARC 21 utilizadas son:

- 008** Información codificada de longitud fija.
- 150** Encabezamiento aceptado de materia general.
- 450** Menciones de referencia de “véase”-Materia general. Contiene un envío de una forma no autorizada o establecida que se encuentra en el campo 150. Esta etiqueta es repetible.
- 550** Menciones de “véase también”-Materia general. Establece un vínculo entre dos términos. Esta etiqueta es repetible.
- 667** Nota no pública. Tiene el propósito de delimitar el alcance del término.
- 670** Fuente donde se localizó la información.

materia sea consistente, cumpliendo los principios de encabezamiento único, controlando la sinonimia, homonimia y la sintaxis, porque todo lo anterior va a condicionar tanto el acceso a los registros por parte de los usuarios como el trabajo de indización. No hay que olvidar que el control de la consistencia permite, por un lado, comprobar que los registros de autoridad cumplan las condiciones de ser únicos e inconfundibles y, por el otro, posibilita verificar que las referencias de “véase” y “véase también” sean adecuadas y cumplan el principio de reciprocidad.

Reflexiones finales

Somos conscientes que desde hace algún tiempo en el ámbito bibliotecario la tendencia ha sido la de realizar proyectos de control de autoridades a través de la cooperación entre varias bibliotecas, ya sea a nivel nacional o internacional, y coincidimos con Herrero Pascual (1999: 133) cuando señala que, “Este sistema reduce los costos y las tareas de control porque todas las bibliotecas utilizan la información disponible y no se duplica el trabajo.” Desafortunadamente en nuestro caso esto no ha sido posible, pues la experiencia al intentar formar un catálogo de autoridades de materia colectivo no tuvo éxito, al no poder unificar criterios en la forma y uso de vocabularios controlados. Por un lado, unos centros utilizaban encabezamientos de materia de la Biblioteca del Congreso de Estados Unidos, mientras que otros utilizamos descriptores. Y como lo señala Cabré (2002) “Dentro de cualquier proceso de normalización, la uniformización del conocimiento se lleva a cabo mediante consenso”.

Consideramos que en un futuro este catálogo de autoridades de materia podría convertirse en un tesoro, pues éste admite una mayor especificidad de los términos relativos a campos disciplinares determinados, lo que implica un análisis más preciso de los documentos, pues como menciona la *Special Libraries Association*, los profesionales de la información requieren, entre otras cosas, sentido de compromiso con la excelencia del servicio que se traduce en la buena organización del conocimiento; capacidad para enfrentar y buscar el reto que le permita desarrollar todas sus habilidades y destrezas al máximo; visión amplia para obtener de su entorno las herramientas que faciliten y mejoren su desempeño; ser flexibles y tener una actitud positiva ante los continuos cambios (Alejos, 2008: 8).

Bibliografía:

ALEJOS ARANDA, Ruth Soledad (2008). Principios éticos y de calidad: buenas prácticas en la organización del conocimiento. *Biblios*, 33-34: 1-15.

BERMUDEZ, Luis, *et al.* (2011). The importance of controlled vocabularies. *The MMI Guides: Navigating the World of Marine Metadata*. Disponible en línea: <http://marinemetadata.org/guides/vocabs/vocimportance>.

- CABRÉ, María Teresa (2002). Terminología y lingüística: la teoría de las puertas. *Estudios de Lingüística en Español*, número 16. Disponible en línea: <http://elies.rediris.es/elies16/Cabre.html>
- ESPAÑA. Ministerio de Agricultura, Alimentación y Medio Ambiente. *Tesaurus de Ingeniería Hidráulica*. Disponible en línea: <http://hispagua.cedex.es/documentacion/tesauro>
- EUROPEAN Environment Agency. *General Multilingual Environmental Thesaurus*. Disponible en línea: http://www.eionet.europa.eu/gemet/index_html?langcode=es
- FAO. *AGROVOC Thesaurus*. Disponible en línea: <http://aims.fao.org/website/AGROVOC-Thesaurus/sub>
- HERRERO PASCUAL, Cristina (1999). El control de autoridades. *Anales de documentación*, No. 2, pp. 121-136. Disponible en línea: <http://redalyc.uaemex.mx/pdf/635/63500208.pdf>
- IRC Centro Internacional de Agua Potable y Saneamiento. *InterWater Tesaurus*. Disponible en línea: <http://thesaurus.watsan.net/es/index.html>
- ISENOR, Anthony (2011). A last resort: Developing a local vocabulary. *The MMI Guides: Navigating the World of Marine Metadata*. Disponible en línea: <http://marinemetadata.org/guides/vocabs/cvdev>.
- MARTÍNEZ ARELLANO, Filiberto F. (2001). Teaching of subject access and retrieval at Mexican LIS schools. *67th IFLA Council and General Conference*. Disponible en línea: <http://www.findthatfile.com/search-23931651-hPDF/download-documents-026-142e.pdf.htm>
- ORGANIZACIÓN PANAMERICANA DE LA SALUD. Área de Desarrollo Sostenible y Salud Ambiental (2004). *Tesaurus de Ingeniería Sanitaria y Ambiental*.
- SOLER MONREAL, Concha; Gil Leiva, Isidoro (2010). Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías. *Revista Interamericana de Bibliotecología*, 33(2): 361-377.
- TURUGUET, Domènec (1994). Consideraciones sobre los tesauros. *El profesional de la información*, número 21, febrero. [rev. 21 de mayo, 2012]. Disponible en línea: http://www.elprofesionaldelainformacion.com/contenidos/1994/febrero/consideraciones_sobre_los_tesauros.html
- ZENG, Marcia Lei (2005). *Construction of controlled vocabularies, a primer* (based on Z39.19). Disponible en línea: <http://www.slis.kent.edu/~mzeng/Z3919/index.htm>