

TESAUROS

Blanca R. Tamborenea

Un tesoro es una herramienta de trabajo que se usa por lo general en centros de documentación para indizar y recuperar documentos así como en bibliotecas se usan listas para asignar encabezamientos de materia.

Mientras que los encabezamientos de materia clasifican a una obra bajo el tema o los temas que ésta abarca, los descriptores extraídos del tesoro señalan los conceptos que contiene el tema. Cuando la información se maneja en el campo de lo general (por ej. en una biblioteca pública), el documento predominante suele ser el libro y para ubicarlo en la estructura temática de la institución se recurre a las clasificaciones enciclopédicas sistemáticas o a los encabezamientos de materia. De igual manera se procede con el resto de los documentos. Cuando la información que se procesa está circunscripta a un sector especial del conocimiento y dirigida a usuarios con exigencias más específicas y definidas, el documento que prevalece es el artículo de publicación periódica, el informe, la 'patente. Lo que interesa al sistema, en este caso, no es sólo qué tema trata el documento sino también qué aspectos, conceptos, enfoques lo integran. Es decir, en el segundo caso, el análisis del contenido del documento es más profundo. A ese proceso de análisis en profundidad del contenido de un documento se lo denomina indización.¹

Según el vocabulario de Wersig y Nevcling (9), indizar es "asignar términos de indización a documentos u objetos con el fin de poder recuperar más adelante los documentos u objetos de acuerdo con los conceptos seleccionados que designan los términos de indización".

Para indizar se puede recurrir al lenguaje natural o al lenguaje controlado. Indizar, en lengua natural significa extraer los términos que usó el autor para indicar los elementos que integran el contenido. Los términos se pueden extraer del título, del resumen del documento o del documento mismo en su totalidad. El ejemplo más representativo de indización en lenguaje natural lo constituyen los índices permutados (KWIC, KWOC, etc.).

¹ Algunos autores llaman indización a toda operación que signifique ubicar documentos por tema. Clasificar un libro por CDU es indizar. En nuestro medio es más común referirse a clasificar cuando se trata de agrupar documentos según el tema o la materia que tratan e indizar cuando el análisis del contenido llega a indicar los conceptos incluidos en el tema.

La necesidad de evitar al ruido en la transmisión de información llevó a la elaboración de lenguajes controlados. Indizar en lenguaje controlado significa asignar a cada concepto un término ya establecido de mutuo acuerdo aún cuando no sea éste precisamente el usado por el autor para expresarse. Es decir que se debe representar el contenido de los documentos utilizando términos extraídos de un conjunto de términos autorizados. Siguiendo a Lancaster (4) se debe ubicar a los encabezamientos de materia dentro de los lenguajes controlados no manipulativos, precoordinados, mientras que los descriptores son el elemento fundamental de un lenguaje controlado manipulativo, postcoordinado y sintético. Atomiza los temas en sus conceptos constitutivos en el momento de la entrada permitiendo manipular ese concepto y, con menor cantidad de elementos, postcoordinándolos (es decir, coordinándolos según se necesite en el momento de la búsqueda), logra una mayor cantidad de combinaciones. Cada combinación significa una posibilidad de acceso a la información almacenada. El conjunto de los términos autorizados a que recurre un sistema para indicar el contenido de documentos se presenta generalmente en forma de listas. A esas listas de términos autorizados se las llama tesauros.²

Continuando con las definiciones de Wersig y Neveling (9) un tesauro es "un *lenguaje documentario* controlado y dinámico que contiene *términos* relacionados semántica y genéricamente y que cubre comprensivamente un dominio específico del conocimiento".

Se entiende por *lenguaje documentario*³ un lenguaje artificial usado por los sistemas de documentación con propósitos de indización, almacenamiento y recuperación" y por *término* "una palabra o frase usada para denotar un concepto".

Se lo llama dinámico porque ese lenguaje está en continua evolución y debe ser perfeccionado y actualizado permanentemente.

Para facilitar la precisión en la adjudicación de términos del vocabulario se indican las relaciones que existen entre ellos. Éstas conducen: de los términos no usados, a los usados para denotar igual concepto (y así evitar la sinonimia); de un concepto dado, a aquél o aquéllos más generales que lo contienen, a los más específicos en los que está contenido o a otros con los que se relaciona por tener aspectos en común. Las relaciones son semánticas porque se refieren al significado, y genéricas cuando establecen

² La lista no es la única forma de presentación, pero es la más frecuente y accesible.

³ Para Wersig y Neveling *Lenguaje artificial* es el opuesto (OT) a *lenguaje natural* y uno de sus tres términos específicos (NT) es *lenguaje documentario*. *Lenguaje controlado* no figura.

relaciones jerárquicas. Con esto se consigue poner de manifiesto la estructura que subyace en el lenguaje.

Cuando se comparó el tesoro con la lista de encabezamientos se dijo que el primero se usaba en el ámbito de la información especializada. Por ello la definición concluye refiriéndose a la cobertura comprensiva de un campo específico. No existe un tesoro que cubra todos los dominios del conocimiento. Cada sector del conocimiento que necesita organizar la información que produce y consume elabora su propio conjunto de términos autorizados⁴.

La presentación de este conjunto de términos es importante porque condiciona su correcta utilización. Hasta ahora lo más frecuente es una lista de términos de entrada en un solo orden alfabético o bien listas alfabéticas dentro de los campos temáticos que integran el conjunto. Los términos autorizados para indizar, almacenar y recuperar son los descriptores. Para favorecer la correcta adjudicación de los descriptores se agregan otros elementos. Notas de alcance para incluir aclaraciones necesarias y evitar la polisemia:

ej.: PENALIDADES

N.A.: Sanciones impuestas por la ley u ordenanzas. Para el castigo impuesto por la autoridad legítima véase PENAS.

PENAS

N.A.: Castigo impuesto por la autoridad legítima. Para las sanciones impuestas por la ley véase PENALIDADES.

términos más amplios,

ej.: MERCADOS

TG: ABASTECIMIENTO

términos más específicos, para ofrecer otras posibilidades en relación con el descriptor elegido en primera instancia,

⁴ Existen macrotesauros, como el de OCDE (6), para desarrollo económico y social, que cubren un campo más amplio. Pero no es tesoro propiamente dicho. Su objetivo es servir de base común para la confección de tesauros en áreas más específicas de ese campo.

ej.: DISTRIBUCIÓN DE LA RENTA

TR: REDISTRIBUCIÓN DE LA RENTA

términos no aceptados por el sistema, que, como ya se dijo, tienden a evitar la sinonimia,

ej.: BASURA

u. RESIDUOS

POLUCIÓN

u. CONTAMINACIÓN

Sea cual fuere la forma de presentación del tesoro ésta debe estar complementada con los índices. Cada uno de ellos constituirá una vía diferente de acceso al vocabulario. Por ejemplo, si los términos se presentan en un solo orden alfabético, deberá ofrecer, además de la introducción e instrucciones para el manejo, el orden sistemático a que pertenecen esos descriptores, un índice jerárquico, uno permutado y, según el campo que cubra el tesoro, otros índices complementarios: geográfico, de fórmulas, de indentificadores, de nuevos términos incluidos, de términos desechados, etc. Tal vez el ejemplo más reseratico a nuestro alcance de este tipo de presentación continúe siendo el TEST (Thesaurus of Engineering and Scientifie Terms del Engineers Joint Council).(3)

Existen otras formas, como las de presentación gráfica, de las que hay excelentes ejemplos en tesoros ya confeccionados. Se cuenta además con literatura sobre ellos. Son útiles en determinados casos, cuando el vocabulario no es muy extenso ni las relaciones incluyen muchos niveles. Generalmente han sido elaborados por equipos muy especializados y pertenecen a instituciones muy solventes.

Pero si un servicio de información en nuestro medio no cuenta con un tesoro ya confeccionado que responda a sus necesidades y debe confeccionar uno, lo más probable es que le convenga comenzar por elaborar una forma más o menos simplificada del ejemplo anterior.

Para localizar la existencia de tesoros en centros de información y documentación y bibliotecas especializadas en nuestro país, el CAYCIT (Centro Argentino de Información en Ciencia y Técnica) está confeccionando un catálogo que será de suma utilidad.

Para confeccionar tesauros la Unesco ha elaborado guías. Una de ellas para elaboración y desarrollo de tesauros monolingües, otra para plurilingües (7) (8). Hay una traducción al castellano de una de ellas. También en castellano se puede consultar el trabajo publicado por CLADES en Chile (5) y del manual de Aitchison y Gilchrist (1) existe una traducción al portugués (2).

En nuestro medio, varios centros de documentación, servicios de información y bibliotecas especializadas han encarado la elaboración de su propio tesoro, debido en gran medida a la escasez de esta herramienta en nuestro idioma.

Esto, unido al incremento del uso de la computadora y la integración en redes de los sistemas de información, hace prever un futuro con mayor experiencia local. Es de esperar que esa experiencia pueda ser intercambiada, enriquecida y aprovechada por todos los que tienen o sienten la necesidad de insertarse en esta etapa del manejo de la información.

BIBLIOGRAFÍA

(1) Thesaurus construction : a practical manual / J. Aitchison and A. Gilchrist. - London : Ashb, 1972.

(2) Manual para construo de tesauros / J. Aitchison and A. Gilchrist; trad. de Helena Medeiros Pereira Braga. - Río de Janeiro : 1979.

(3) Thesaurus of engineering and scientific terms / Engineers Joint Council. - New York : 1967.

(4) Vocabulary control for information retrieval / F. W. Lancaster. - Washington : Information Resources Press, 1972.

(5) Uso de descriptores y tesauros / Naciones Unidas, Consejo Latinoamericano de Documentación Económica y Social. - Santiago, Chile : 1971.

(6) Macrothesaurus para el tratamiento de la información relativa al desarrollo económico y social / Organización para la Cooperación y Desarrollo Económico. - Nueva edición española / preparada por Jean Viet. - París : OCDE. 1979.

(7) Guidelines for the establishment and development of monolingual thesauri. - París : Unesco, 1976.

(8) Guidelines for the establishment and development of multilingual thesauri. - París : Unesco, 1976.

(9) Terminología de la documentación / G. Wersig y U. Neveling. - París : Unesco, 1976.