

Tesauros y glosarios CINDOC: temática, estructura y modos de acceso

Ángeles Maldonado Martínez
CINDOC-CSIC
Departamento de Ciencias Sociales

1. Introducción

El Centro de Información y Documentación Científica (CINDOC), perteneciente al Consejo Superior de Investigaciones Científicas (CSIC), cuenta en el momento actual con una colección de herramientas terminológicas, tesauros y glosarios especializados, que ha ido elaborando y actualizando a lo largo de su existencia. El presente texto tiene por finalidad hacer una presentación de dichas herramientas haciendo especial hincapié en la metodología de elaboración, estructura y disponibilidad actual de las mismas.

2. Tesauros

2.1. Historia

El interés del ahora CINDOC, nacido en 1992 de la unificación del Centro de Documentación Científica y Técnica (ICYT) y el Centro de Documentación en Ciencias Sociales y Humanidades (ISOC), en la elaboración de tesauros se manifiesta a principio de los años 80 en dos hechos que marcarán el comienzo de esta línea de trabajo que continúa en la actualidad. Por una parte la concesión al ISOC en 1987 por parte de la Comisión Interministerial de Ciencia y Tecnología (CICYT) de un proyecto de investigación para la elaboración de tesauros sectoriales en el campo de las Ciencias Sociales. El otro hecho relevante es el encargo al ICYT por parte de la UNESCO de la traducción y adaptación al español del *Tesaurus Spines*, vocabulario pensado para el control y desarrollo de políticas científicas y tecnológicas.

Con ayuda del proyecto concedido por la CICYT se elaboraron los tesauros de Urbanismo, Economía y Psicología y se avanzaron los trabajos de los de Derecho, Ciencias Políticas y Sociología. Todos estos tesauros nacieron con el propósito de controlar el vocabulario en cada uno de los ficheros de la base ISOC, base de datos bibliográfica que recoge el vaciado de las revistas científicas editadas en España especializadas en Ciencias Sociales y Humanas.

Posteriormente se ha completado y publicado el *Tesaurus de Derecho* (1997), gracias a la colaboración de otras instituciones del ámbito jurídico español. Dentro de las Ciencias Humanas se acaba de publicar el *Tesaurus de Historia Contemporánea de España* (1999). En el campo de las ciencias cabe mencionar el *Tesaurus ICYT de Biología Animal* y el *Tesaurus de Geología*.

Especial mención requiere el *Tesaurus de topónimos*, cuya primera edición data de 1993, actualizado posteriormente, que contiene unos 26.000 términos de entidades político-administrativas, de las cuales 20.000 pertenecen al ámbito de los países de habla española y portuguesa.

A lo anterior hay que unir la creación y la traducción de diversos tesauros realizados por encargo de otras instituciones.

2.2. ¿Por qué tesauros?

Los tesauros realizados por el CINDOC surgen como herramientas de control de vocabulario para ser utilizados en la indización y recuperación de información de sus bases de datos. Se eligió el tesaurus como herramienta de control del vocabulario por pensar que, además de eliminar las ambigüedades del lenguaje en cada disciplina o área temática, presenta otras ventajas adicionales que deben tenerse en cuenta.

La primera de estas ventajas es que, si lo comparamos con un listado de autoridades, un tesaurus añade información complementaria de cada uno de los términos que contiene, pues todos ellos aparecen rodeados de aquellos otros términos relacionados de alguna manera. Un tesaurus siempre nos sugiere nuevos términos a utilizar en la indización y en la recuperación de información.

La otra ventaja fundamental del uso de tesauros en bases de datos documentales está más relacionada con la recuperación de información. A un tesaurus se le saca su máximo rendimiento en una base de datos gestionada con un programa que permita la recuperación de información utilizando la estructura jerárquica del mismo. Esta función permite expandir las búsquedas al presentar la posibilidad de recuperar automáticamente por los términos específicos o genéricos de un término expresado anteriormente. Esta función representa un ahorro considerable de tiempo y añade comodidad a las búsquedas.

2.3. Metodología de elaboración

Prácticamente todos los tesauros fueron coordinados por un documentalista especialista en el tema tratado, participando además en la elaboración expertos en las diferentes materias, quienes en unos casos establecieron las diferentes relaciones entre los términos y en otros dieron el visto bueno a lo anteriormente organizado.

La metodología seguida para la elaboración de los tesauros mencionados fue básicamente la misma. Las fases de elaboración fueron las siguientes:

a) Selección de fuentes

Los términos de indización procedentes de las bases de datos ISOC e ICYT formaron el núcleo fundamental que sirvió de base para la elaboración de los diferentes tesauros. Sin embargo, este conjunto de términos no fue suficiente y hubo que completar cada una de las terminologías especializadas con términos extraídos de otras fuentes: tesauros de igual temática o temáticas próximas existentes en otros idiomas, diccionarios y léxicos especializados.

b) Extracción de términos

Una vez seleccionadas las fuentes se procedió a la extracción sistemática de los términos aparecidos en ellas.

c) Normalización de términos

En cuanto a la forma gramatical, género y número de los términos, se siguieron las especificaciones contempladas en la norma ISO 2788-1986 versión española en el caso de los tesauros elaborados con anterioridad a la aprobación de la norma UNE 50106:1990 sobre *Directrices para el establecimiento y desarrollo de tesauros monolingües*. Una vez aprobada esta norma se siguieron sus especificaciones.

d) Determinación de grandes áreas y distribución de términos

Paralelamente a la extracción de términos se pensó en la estructura jerárquica básica de cada uno de los tesauros, determinando las grandes áreas temáticas o familias semánticas en las que más tarde se distribuirían los términos. A partir de ese momento se trabajó individualmente con cada familia en el establecimiento de relaciones semánticas.

e) Establecimiento de relaciones de sinonimia

Agrupados los términos por áreas temáticas, el paso siguiente fue identificar los términos sinónimos dentro de las mismas. Detectadas las equivalencias semánticas, se procedió a elegir uno de los términos como representativo del concepto, adjudicándole la categoría de descriptor o término principal. El término o términos equivalentes se remitirían al principal.

f) Establecimiento de la estructura jerárquica

Una vez elegidos los descriptores, se comenzó a construir la jerarquía con los mismos. Paulatinamente se establecieron los diferentes niveles jerárquicos hasta crear una estructura piramidal donde quedaron encajados todos los términos principales ocupando el nivel que les correspondiera según el grado de especificidad - generalidad de sus significados en relación con el resto.

g) Establecimiento de relaciones asociativas y notas de aplicación

Creada la estructura jerárquica, se pasó a establecer las relaciones de asociación entre términos que, aun no manteniendo una vinculación jerárquica directa, tuvieran sin embargo una relación de vecindad semántica. Finalmente con las notas de aplicación, que definen y aclaran el sentido de los términos con más de una significación, se eliminó la polisemia.

h) Informatización

Una vez establecidas todas las relaciones posibles entre los términos, hubo que informatizar el resultado con el fin de conseguir una presentación adecuada para la edición. Para ello se utilizó el módulo de creación de tesauros del gestor de bases de datos estándar BASIS, capaz de generar los listados básicos en los que se difunde un tesauro: alfabético, jerárquico y permutado. En algunos casos también se utilizó el programa de creación de tesauros para microordenadores conocido como CAT (Construcción Automática de Tesauros), elaborado por los informáticos del propio CINDOC.

i) Edición provisional

Finalizados todos los trabajos anteriormente relacionados se procedió a una edición provisional, para hacerla circular con el fin de recoger opiniones externas al equipo de trabajo.

j) Revisión y edición definitiva

La edición provisional sirvió para poner a prueba los tesauros. Una vez editados, pudieron ser utilizados en la indización de documentos y en la elaboración de estrategias de búsqueda, comprobándose así su adecuación al tema desarrollado. Después de su uso, en la mayoría de los casos se incluyeron nuevos términos, se eliminaron otros y se modificaron las relaciones en los casos en los que se creyó conveniente.

2.4. Estructura y presentación de los tesauros

La presentación de los tesauros mencionados responde al mismo esquema. Todos ellos presentan los términos que contienen en tres índices diferentes: alfabético, jerárquico y permutado. A la hora de utilizar un tesoro, lo aconsejable es consultar los tres índices, ya que cada uno ofrece una información diferente sobre un término dado. A continuación se presentan las características de cada uno de los índices.

a) Índice alfabético

En el índice alfabético están presentes todos los términos contenidos en el tesoro, descriptores o términos principales y no descriptores o sinónimos. Ambas categorías se diferencian gráficamente, pues los primeros suelen ir en letras mayúsculas y los segundos, en minúsculas. Cada uno de los términos principales se presenta con todos sus términos relacionados, mencionando la naturaleza de la relación. Los no descriptores tan sólo remiten a los términos principales.

b) Índice jerárquico

En este listado aparecen representadas las cadenas jerárquicas completas. Este índice da una visión completa de cada cadena jerárquica y proporciona el encuadre de los términos dentro de ellas. La representación de la cadena jerárquica se hace a través de sangrados y puntos. Los listados alfabético y permutado remiten a este índice jerárquico a través de los códigos que aparecen en todos los términos principales o descriptores.

c) Índice permutado KWIC

En el índice aparecen todos los términos del tesoro, descriptores y no descriptores, alfabetizados por todas las palabras significativas de que están compuestos. Este índice reúne todos los términos que tienen una misma palabra o al menos una misma raíz.

3. Glosarios especializados multilingües

Otra línea de trabajo relacionada con la terminología desarrollada en el CINDOC es la creación de glosarios especializados multilingües. Estos trabajos tienen como principal finalidad recoger y sistematizar la terminología específica de un campo científico determinado y el establecimiento de equivalencias en otros idiomas para propiciar el intercambio entre profesionales y expertos.

En el momento presente existen cuatro glosarios publicados de estas características: *Diccionario de alimentos*, *Terminología multilingüe de drogas*, *Terminología multilingüe de máquinas-herramienta* y *Glosario de términos de Acuicultura*. Los tres primeros han sido realizados por los componentes del grupo TermEsp (Terminología española) que desarrolla sus actividades en el CINDOC desde 1985. El cuarto ha sido desarrollado por los componentes del Centro de Documentación en Acuicultura, todos ellos asimismo pertenecientes al CINDOC.

Estos glosarios tienen todos una estructura similar. Constan de un índice alfabético general donde los términos en castellano van acompañados de sus equivalentes en alemán, inglés y francés (en el caso del *Glosario de términos en Acuicultura* también aparece el equivalente en italiano). En los casos que procede también figuran los sinónimos en español. Además, algunos términos van acompañados de una breve nota aclaratoria que precisa su significado. Este índice alfabético general se acompaña de un índice permutado y otro alfabético de los términos en alemán, inglés y francés.

4. Acceso a tesauros y glosarios

Los soportes de edición de los tesauros y glosarios mencionados son diversos. De prácticamente todos, salvo del *Tesauro de Geología*, existe una edición impresa en soporte papel que puede ser adquirida en el propio CINDOC por todo aquel que esté interesado.

El acceso a través de Internet es más reciente. Todos los tesauros, con la excepción del de Derecho y del de Historia Contemporánea de España, y también los glosarios, pueden ser consultados a través de Internet (<http://www.cindoc.csic.es>). Las ediciones web de los tesauros y glosarios son una reproducción de las versiones impresas, pues se puede acceder al contenido de los mismos a través de los índices alfabético sistemático, jerárquico y permutado. La ventaja que trae consigo la consulta a través de Internet es el poder hacer uso de los hiperenlaces, lo que permite saltar automáticamente de un índice a otro. El uso de los tesauros y los glosarios a través de Internet es gratuito y está pensado para realizar consultas puntuales.

Lo más novedoso en el tema del acceso es la próxima aparición en CD-ROM de casi la totalidad de los tesauros en formato HTML para que puedan ser incorporados a los sistemas de información de aquellas organizaciones que estén interesadas en utilizarlos en la indización y recuperación de información de sus bases de datos. Para el uso de los tesauros en esta modalidad habrá que pagar los derechos correspondientes.

Tesauros y glosarios del CINDOC

Tesauros	Última edición	Idiomas	Términos			Índices			Edición		
			Total	Descrip.	No Descrip.	Alfab.	Jerárq.	Permut.	Papel	CD-ROM	Internet
<i>Tesoro Spmas</i>	1988	Spa, Eng, Fra	10.382	7.993	2.839	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ
<i>Tesoro I.SOC Urbanismo</i>	1992	Spa	4.422	3.652	770	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ
<i>Tesoro I.SOC Economía</i>	1995	Spa, Eng, Fra	6.792	5.464	1.328	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ
<i>Tesoro I.SOC Psicología</i>	1995	Spa	4.400	2.788	1.612	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ
<i>Tesoro ICYT Biología Animal</i>	1995	Spa	3.781	2.782	999	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ
<i>Tesoro I.SOC de Topónimos</i>	1996	Spa	26.000			SÍ	SÍ	SÍ	SÍ	SÍ	SÍ
<i>Tesoro de Derecho</i>	1987	Spa	19.494	14.179	5.315	SÍ	SÍ	SÍ	SÍ	NO	NO
<i>Tesoro de Geología</i>		Spa	2.108	2.055	43	SÍ	SÍ	SÍ	NO	NO	SÍ
<i>Tesoro Historia Contemporánea Española</i>	1989	Spa	4.139	3.776	350	SÍ	SÍ	SÍ	SÍ	NO	NO
Glosarios	Última edición	Idiomas	Términos			Índices			Edición		
			Total	Descrip.	No Descrip.	Alfab.	Jerárq.	Permut.	Papel	CD-ROM	Internet
<i>Diccionario de Alimentos</i>	1994	Spa, Eng, Fra, Ger	3.814			SÍ	NO	SÍ	SÍ		SÍ
<i>Glosario de Términos de Acuicultura</i>	1996	Spa, Eng, Fra, Ger	2.200			SÍ	NO	SÍ	SÍ		SÍ
<i>Term. Multilingüe Drogas</i>	1997	Spa, Eng, Fra, Ger	2.000			SÍ	NO	SÍ	SÍ		SÍ
<i>Term. Multilingüe Máquinas Herramientas</i>	1999	Spa, Eng, Fra, Ger	1.800			SÍ		SÍ	SÍ		SÍ