
Los tesauros como herramienta fundamental en la descripción y recuperación de documentos electrónicos

Ángeles Maldonado Martínez
CCHS-CSIC

INTRODUCCIÓN

Los tesauros se han venido utilizando en las bases de datos documentales desde la década de los 50 del siglo pasado, pero tienen su auge entre los años 70 y 90, cuando las bases de datos bibliográficas internacionales de mayor relevancia elaboran el suyo propio. Un tesoro es una herramienta muy útil para gestionar la información, pero a la vez su elaboración es compleja y costosa, por lo que su uso no se ha generalizado lo suficiente. Con la euforia de Internet, los usuarios quedaban contentos al encontrar siempre algo de información de forma muy sencilla, decayó el interés por este tipo de herramientas, que se pensaban complicadas y obsoletas. Ahora de nuevo vuelven a revalorizarse. Este trabajo trata de reivindicar el uso de tesauros en la recuperación documental en todo tipo de sistemas de información, justificando de nuevo su valor. Analiza el uso tradicional que se ha hecho de esta herramienta y propone formas de uso actualizadas acordes con las posibilidades del nuevo entorno electrónico.

1. DEFINICIÓN Y UTILIDADES DEL TESAURO

Un tesoro es una herramienta terminológica compuesta por un conjunto de términos relacionados entre sí sobre un tema o materia determinada. Los tesauros se utilizan en bases de datos o sistemas de información automatizados para controlar el vocabulario y para sugerir nuevos términos tanto en la indización de documentos como en la recuperación de información.

2. EL TESAURO COMO HERRAMIENTA DE CONTROL DEL VOCABULARIO

En Documentación es fundamental el control del vocabulario para no perder información en la recuperación de información. Controlar el vocabulario es filtrar por un lenguaje controlado los términos provenientes del lenguaje natural. El objetivo fundamental de cualquier tipo de lenguaje controlado es identificar términos de igual significación y establecer una relación de equivalencia entre ellos. Listados de autoridades y tesauros, son los tipos fundamentales de lenguajes controlados. La diferencia básica que existe entre ellos es que el tesoro, además de la relación de equivalencia, contempla también otros tipos de relaciones entre sus términos constitutivos. En este apartado se hablará de tesauros para simplificar, pero se recuerda que desde el punto de vista del control de vocabulario los listados de autoridades cumplen el mismo objetivo, funcionan y se utilizan de la misma manera que estos.

El tesoro, al identificar y relacionar términos sinónimos o equivalentes, contribuye a evitar la pérdida de información en las búsquedas, circunstancia que se produce cuando en una base de datos se utiliza lenguaje natural en lugar de un lenguaje controlado en el análisis de contenido de los documentos. En Documentación el lenguaje natural se opone al lenguaje controlado.

2.1 El lenguaje natural y el fenómeno de la sinonimia

Se llama lenguaje libre o natural a aquel que se utiliza cotidianamente en la comunicación tanto oral como escrita. El lenguaje natural se caracteriza por su riqueza y espontaneidad en el uso del vocabulario. Muchos conceptos pueden ser expresados por más de un término. Cuando esto ocurre, nos encontramos frente al fenómeno de la sinonimia o equivalencia lingüística. El lenguaje natural está repleto de sinónimos o términos equivalentes. Al hablar se utilizan las palabras que surgen de forma natural en la mente, generalmente sin ningún tipo de control sobre las mismas. Al escribir, función que se suele realizar de forma más meditada, se evita ser repetitivo con las palabras, recurriendo a menudo a sinónimos para hacer menos "pesados" los textos. En ambos casos se están utilizando continuamente términos equivalentes sin apenas percibirlo.

A continuación se enumeran diferentes tipos de sinonimia o equivalencia lingüística, para dar una idea del alcance del problema.

- Sinonimia semántica, es aquella que se da entre términos del mismo idioma con orígenes lingüísticos históricos diferentes.
 - Nevera / Frigorífico
 - Tren / Ferrocarril
- Variantes ortográficas
 - Quiosco / Kiosco
 - Derbi / Derby

- Traducción de palabras extranjeras
Software / Programa de ordenador
Parking / Estacionamiento / Aparcamiento
- Denominación popular y científica
Pingüino / Esfenisciforme
Senderuela / Marasmius Oreades
- Nombres comerciales y científicos
Aspirina / Ácido acetilsalicílico
- Nombres comerciales y genéricos
Spontex / Ballea
Dodotis / Pañales
- Términos actuales y arcaísmos
Viejos / Ancianos / Mayores / Tercera edad
Chabolas / Infraviviendas
- Siglas y su desarrollo
IVA / Impuesto sobre el Valor Añadido
CAM / Comunidad Autónoma de Madrid
- Términos de un mismo idioma originarios de diferentes culturas o regiones
Hormigón / Concreto
Boquerón / Bocarte
- Sinonimia gramatical o formal, es aquella que se da entre un término formado por “un sustantivo + un adjetivo”, y el mismo término donde el adjetivo es sustituido por la “preposición de + un sustantivo”
Ahorro energético / Ahorro de energía
Estación ferroviaria / Estación de ferrocarril

El uso de sinónimos, muchas veces aconsejable desde el punto de vista lingüístico en el discurso oral y escrito, es un inconveniente grave para el análisis y la recuperación de documentos, ya que produce pérdida de información en las búsquedas, debido a la dispersión que provoca en la indización y en la interrogación.

Veamos como funciona un sistema de información o base de datos que no controle el vocabulario, o lo que es lo mismo que trabaje con lenguaje natural. Pensemos en dos términos sinónimos como por ejemplo “contaminación del aire” (A) y “contaminación atmosférica” (B). Ahora imaginemos un texto que trate este fenómeno, donde el autor utilice indistintamente uno u otro término. El indizador al analizar este texto habrá de elegir entre el término A o el término B como palabra-clave en su indización. Imaginemos que elige el término A (contaminación del aire). Con el paso del tiempo el analista indizará más textos donde se hable del

mismo fenómeno. Como indiza en lenguaje libre, pues no utiliza ninguna herramienta de control del vocabulario en sus indizaciones, en algunos casos volverá a utilizar el término A, pero posiblemente en otras utilizará el término B. Pasado un tiempo, pongamos por caso, en la base de datos habrá 7 documentos indizados con “contaminación del aire” y 25 con “contaminación atmosférica”.

Examinemos ahora la situación desde el punto de vista de los usuarios de esta base de datos. El usuario X, a la hora de realizar su consulta decide, utilizar como término de búsqueda “contaminación del aire”, con lo que obtiene como resultado los 7 documentos indizados con este término. Por el contrario el usuario Y utiliza “contaminación atmosférica”, y recupera 25 documentos. En ambos casos se habrá producido pérdida de información, pues en la base de datos en cuestión en realidad existen 32 documentos que tratan del mismo fenómeno representado por dos términos equivalentes, contaminación del aire (7) y contaminación atmosférica (25). Si se hubiese utilizado un tesoro o cualquier otra herramienta de control de vocabulario, la pérdida de documentos se habría evitado, pues los usuarios X e Y hubiesen obtenido el mismo resultado, al verse obligados ambos a utilizar el mismo término.

En un tesoro los sinónimos o términos equivalentes aparecen identificados y relacionados entre sí, lo que permite utilizar esta información para evitar la pérdida de información.

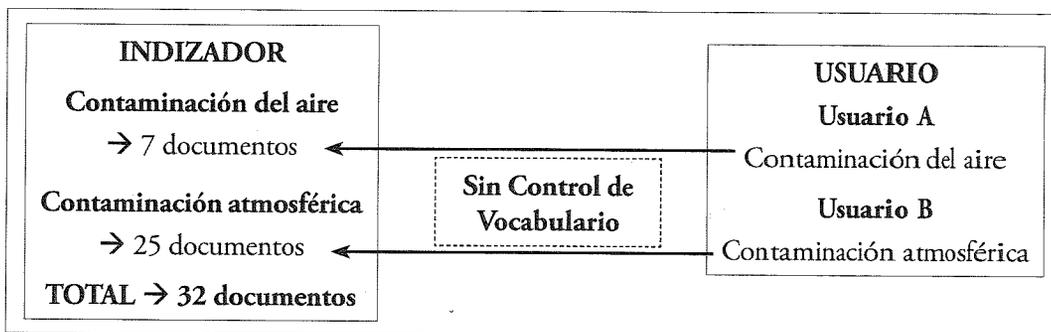


Gráfico 1. Indización y Recuperación de Información sin Control de Vocabulario

2.2 Uso tradicional de los lenguajes controlados

En una base de datos tradicionalmente el tesoro se ha venido utilizando para controlar el vocabulario en los campos de indización. Los campos de indización son aquellos donde se encuentra el conjunto de términos descriptivos del contenido temático de un documento. El tesoro se ha utilizado para controlar el vocabulario filtrando los términos procedentes del lenguaje natural preseleccionados por indizadores y usuarios para ser utilizados como términos de indización por los primeros y como términos de búsqueda por los segundos.

En esta concepción tradicional un tesoro ante dos o más términos de igual significación elige uno de ellos como representativo del concepto en cuestión, remitiendo el resto al término elegido, con el objetivo de que indizador y usuario coincidan y utilicen siempre un único término referido a un concepto.

En un tesoro se da entrada a todos los términos sinónimos o equivalentes representativos de un concepto, pero, como ya se ha dicho anteriormente, se elige uno de ellos como representativo de dicho concepto. El término elegido se denomina “preferente”, y el resto de términos equivalentes reciben el nombre de “no preferentes”. La relación entre términos equivalentes se indica con las expresiones “use/usado por”. Los términos “no preferentes” aparecen acompañados de “use” seguidos a continuación del término “preferente”. Por el contrario al término “no preferente” le seguirá la expresión “usado por” y será acompañado por el resto de sinónimos. Según las normas actuales ISO, NISO y UNE de elaboración y uso de los tesauros, tanto indizadores como usuarios deben utilizar exclusivamente términos “preferentes”.

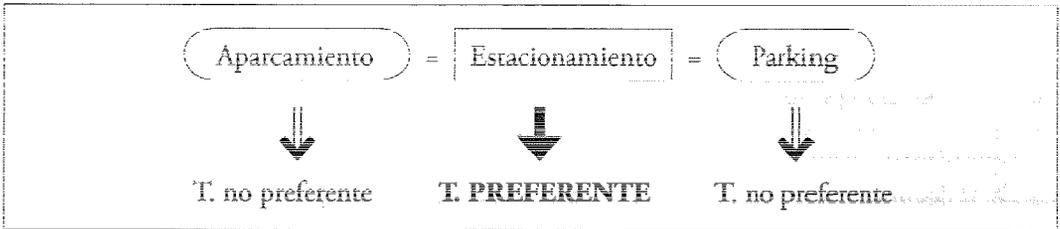


Gráfico 2. Relación de equivalencia. Términos preferentes y no preferentes

En las bases de datos y sistemas de información donde se controla el vocabulario, indizadores y usuarios utilizan la misma herramienta de control para filtrar los términos preseleccionados, y finalmente elegir términos “preferentes” desechando los “no preferentes”. Con este modo de proceder se elimina la dispersión en la selección de términos sinónimos y se garantiza que ante dos o más términos equivalentes, unos y otros coincidirán y seleccionarán el mismo.

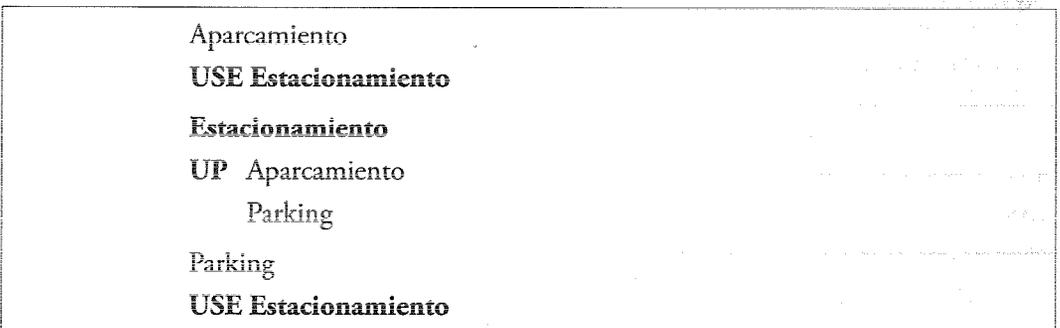


Gráfico 3. Relación de equivalencia. Listado alfabético de un tesoro

Retomaremos el ejemplo anterior. Si en la misma base de datos donde antes no se controlaba el vocabulario, ahora se utiliza un tesoro, uno de los términos sinónimos “contaminación del aire” o “contaminación atmosférica”, aparecerá en él como “preferente”. El indizador, independientemente de cualquiera que fuera de los dos términos el que apareciera en el texto y fuera preseleccionado en primera instancia por el mismo, siempre utilizará como término de indización el señalado como “preferente” en el tesoro.

Pongamos que el elegido como “preferente” es “contaminación atmosférica”. En ese caso “contaminación atmosférica” aparecerá como término de indización en todos los documentos que traten este fenómeno, ya que el indizador, siguiendo las normas de utilización del tesoro, se verá en la obligación de utilizarlo finalmente, aunque previamente hubiera preseleccionado “contaminación del aire”. Por otro lado, un usuario que quisiera hacer una consulta en la misma base sobre el tema y quisiera controlar el vocabulario, antes de lanzar la consulta, visualizaría el tesoro, y seleccionaría el término señalado como “preferente” como término de búsqueda, que en este caso sería “contaminación atmosférica”. Procediendo ambos de esta manera no habría pérdida de información, pues tanto indizador como usuario utilizarían “contaminación atmosférica”.

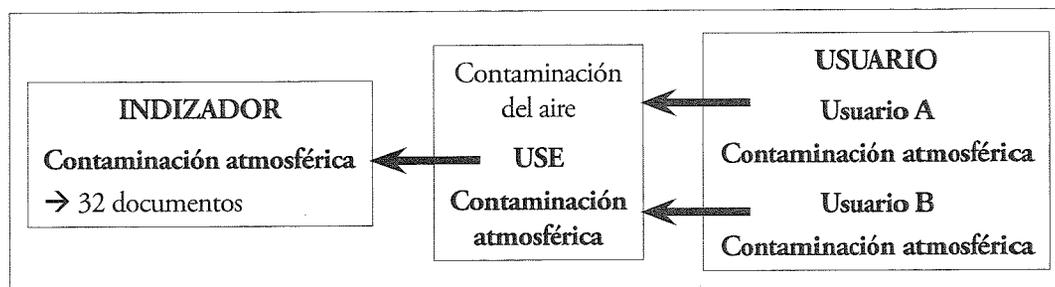


Gráfico 4.- Uso tradicional del Tesoro

Para que este sistema, eficaz a nivel teórico, sea efectivo en la práctica, tanto indizadores como usuarios deben de proceder correctamente. Por la parte de los indizadores no hay problema, pues al ser profesionales de la documentación saben qué es un tesoro, cómo se maneja y el objetivo que se persigue con su uso. Además, los sistemas de gestión de bases de datos suelen poder rechazar los términos “no preferentes” utilizados en la indización. Sin embargo, por parte de los usuarios el procedimiento no se suele seguir de forma correcta, a no ser que se trate de usuarios expertos que conozcan el proceso de búsqueda en lenguaje controlado. Si el usuario no ejecuta la búsqueda de forma adecuada y no sabe manejar el tesoro, el control del vocabulario realizado por los indizadores no reporta ningún beneficio en sus búsquedas.

El usuario debe saber como proceder correctamente en una base de datos donde se controla el vocabulario para beneficiarse de dicho control y no obtener resultados erróneos. El usuario debe conocer cual es el procedimiento de búsqueda en el momento presente en este tipo de bases de datos. El usuario debe saber fundamentalmente lo siguiente:

- 1.- Qué las búsquedas se realizan desde la visualización del propio tesoro.
- 2.- Qué el tesoro sólo actúa en los campos de indización. Las equivalencias no funcionan ni en los títulos ni en los resúmenes, ni por su puesto en los textos completos, donde siempre se utiliza lenguaje natural.
- 3.- Qué si hace una búsqueda en los campos de indización utilizando términos “no preferentes”, el resultado será cero, pues el indizador habrá utilizado siempre en su análisis de contenido el término “preferente”.

Este uso tradicional de los tesauros que se ha venido haciendo en las bases de datos, está pensado para que sean los profesionales de la información los que introduzcan y recuperen la información en las mismas. La complejidad del proceso dificulta que los usuarios no experimentados hagan un uso adecuado de esta herramienta, perdiendo las ventajas que supone su aplicación.

2.3. Uso avanzado de los lenguajes controlados

La generalización del uso de Internet hace que cada vez haya más usuarios que realizan directamente sus búsquedas en bases de datos y sistemas de información sin bibliotecarios o documentalistas como intermediarios. Estos usuarios en la mayoría de los casos desconocen el funcionamiento de una base de datos donde se controla el vocabulario.

Normalmente los usuarios realizan sus búsquedas haciendo uso de los formularios, ya sean estos muy simples y tener una única ventana o más complejos y tener más de una donde poder combinar diferentes campos. Suelen escribir directamente los términos y lanzar la búsqueda sin pensar en nada más. Al proceder de esta forma, en muchos de los casos los términos que se utilizan, pueden tener sinónimos, hecho que pasa desapercibido al usuario, lo que hace perder registros en la recuperación.

A continuación se consideran las diferentes situaciones que se pueden producir en una base de datos donde se indizan los documentos con lenguaje controlado y se añade un resumen del contenido de los mismos, cuando en una consulta un usuario utiliza un término con sinónimos,

1. Situación A

Uso del formulario simple de una única ventana o del formulario de búsqueda avanzada y lanzamiento de la consulta al registro completo o a todos los campos de contenido (títulos, resumen y términos de indización). El resultado obtenido será diferente en función del tipo de término utilizado.

- Utilización de un término preferente
 - Recuperación de los registros donde aparece dicho término en los campos de indización y/o en el título/resumen
 - Pérdida de los registros donde no aparece en los campos de indización, pero aparece(n) su(s) sinónimo(s) en título/resumen
- Utilización de un término no preferente o sinónimo
 - Recuperación de los registros donde aparece en título/resumen
 - Pérdida de los registros donde aparece(n) su(s) sinónimo(s) en título/resumen
 - Pérdida de los registros donde no aparece en título/resumen, pero sí aparece su sinónimo en los campos de indización como término preferente

2. Situación B

Uso de los campos de indización del formulario de búsqueda avanzada

- Utilización de un término preferente
 - Recuperación de los registros donde aparece como término de indización
 - Pérdida de los registros donde el término no aparece en los campos de indización, pero sí en título/resumen
 - Pérdida de los registros donde el término no aparece ni en los campos de indización ni en el título/resumen, pero sí aparece(n) su(s) sinónimo(s) en el título/resumen
- Utilización de un término no preferente
 - Resultado cero, pues no habrá ningún registro donde aparezca dicho término en los campos de indización

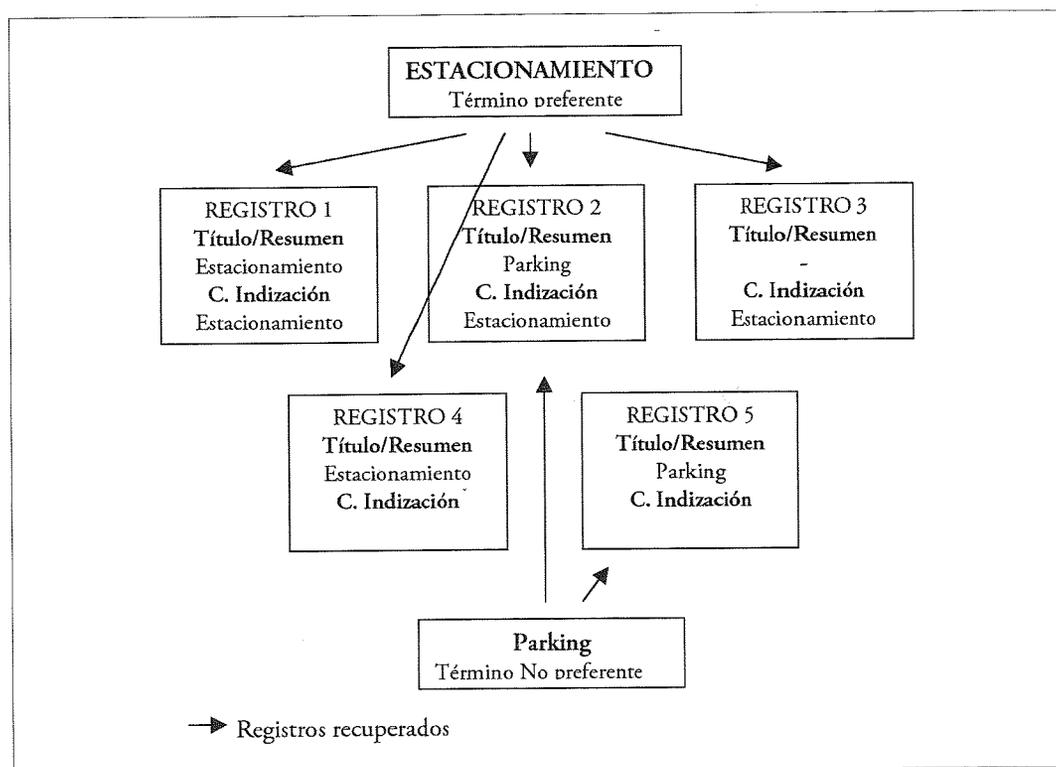


Gráfico 5. Uso tradicional del Tesauro en la recuperación de información

En el Gráfico 5 se muestra que realizando una consulta en registro completo con el término "preferente" ESTACIONAMIENTO, el usuario recupera 4 registros de los 5 posibles. Los 4 registros son aquellos donde aparece este término (Estacionamiento) o su sinónimo (Parking). Sin embargo, al utilizar el término "no preferente" Parking, de los 5 posibles tan sólo se recuperan 2 registros, que son en los que aparece expresamente este término. En el caso de que la

pregunta se realice exclusivamente en los campos de indización, se recuperarán los 3 registros donde aparece el término “preferente” (Estacionamiento) como término de indización.

Existen procedimientos para aplicar el tesoro como herramienta de control de vocabulario de forma imperceptible para el usuario. En estos casos, el tesoro funciona de forma automática en la búsqueda de términos equivalentes.

El uso de las bases de datos directamente por los usuarios debe de empujar a los profesionales de información a simplificar los procesos de búsqueda sin que estos pierdan eficacia. El control del vocabulario en un sistema de información es esencial y no se debe de renunciar a él en aras de la simplificación. Lo que se debe de simplificar son los procesos de recuperación de información.

El activar automáticamente el uso del tesoro desde los formularios de búsqueda simplificaría mucho el proceso de recuperación de información sin renunciar a la eficacia. En el caso del control del vocabulario, todos los términos sinónimos referidos a un mismo concepto tendrían el mismo rango, pues no se establecerían diferencias entre términos “preferentes” y “no preferentes”, y el sistema intentaría localizarlos a todos de forma automática. Con esta forma de proceder ante dos o más términos sinónimos, el usuario tendría libertad para utilizar cualquiera de ellos en su consulta, obteniendo siempre el mismo resultado.

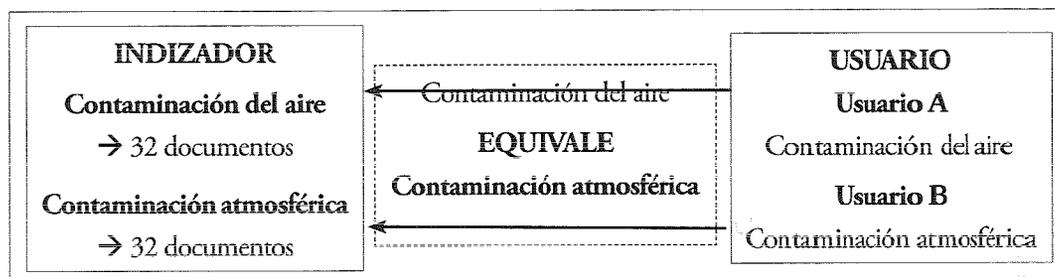


Gráfico 6. Uso avanzado del Tesoro

A continuación se explica como es la recuperación de información en una base de datos en la que se hace un uso automático de un tesoro en el que todos los sinónimos referidos a un mismo concepto se consideran términos equivalentes de igual rango.

1. Situación A

Uso del formulario simple de una única ventana o del formulario de búsqueda avanzada y lanzamiento de la consulta al registro completo o a todos los campos de contenido (títulos, resumen y términos de indización).

- Utilización de cualquiera de los términos equivalentes
- Recuperación de los registros donde alguno de ellos aparece en los campos de indización y/o en el título/resumen

2. Situación B

Uso de los campos de indización del formulario de búsqueda avanzada

- Utilización de cualquiera de los términos equivalentes
 - Recuperación de los registros donde aparece alguno de ellos como término de indización
 - Pérdida de los registros donde ninguno aparece en los campos de indización, pero sí en título/resumen

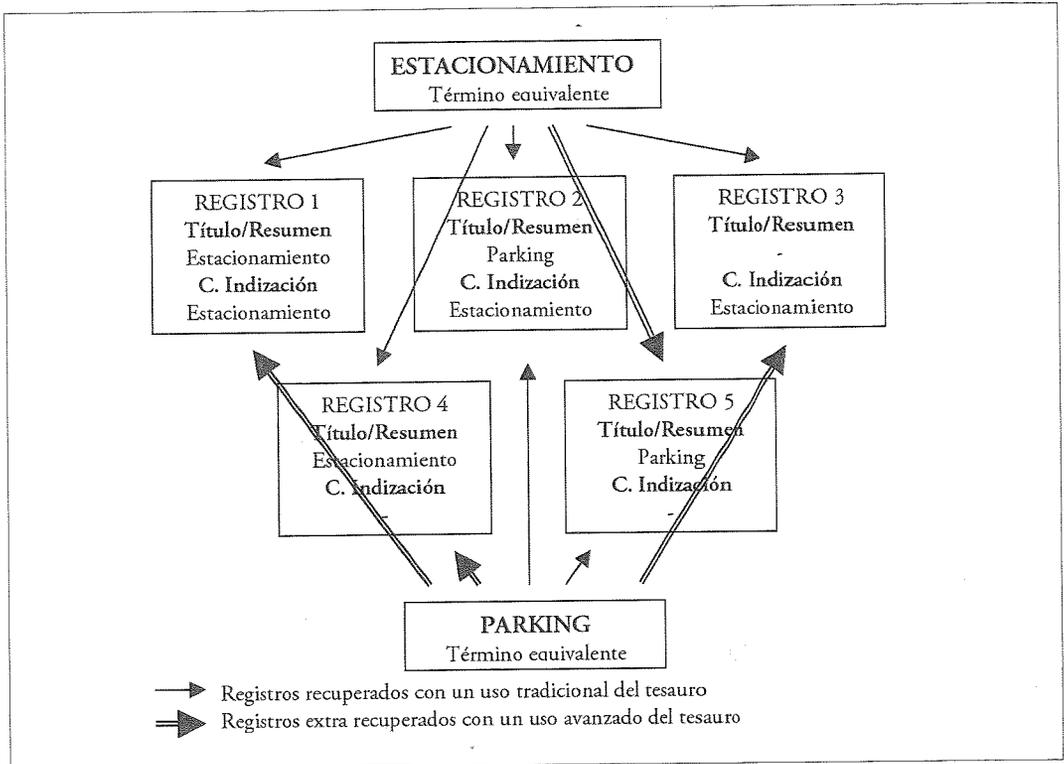


Gráfico 7. Uso avanzado del Tesoro en la recuperación de información

En el gráfico 7 se observa que en una consulta realizada sobre el registro completo al utilizar Estacionamiento o Parking, términos considerados equivalentes de igual rango, el resultado que se obtiene en cualquiera de los dos casos es el mismo: 5 registros de los 5 posibles. En caso de que la consulta se realice exclusivamente en los campos de indización, el resultado será menor, pues sólo se recuperaran aquellos registros (3) donde aparezca alguno de los términos equivalentes como término de indización.

3. EL TESAURO COMO RECURSO DE ENRIQUECIMIENTO TERMINOLÓGICO

Otra de las utilidades que tiene un tesoro, además de la de controlar el vocabulario, es la de sugerir términos en la indización y en la recuperación de información. Un tesoro

funciona como fuente terminológica para indizadores y usuarios, de donde pueden extraer términos complementarios para realizar el análisis de contenido de los documentos, en el caso de los primeros, y para formular sus consultas, en el caso de los segundos. La sugerencia de nuevos términos es consecuencia del establecimiento de relaciones jerárquicas y asociativas entre términos. En los tesauros, cada término aparece acompañado de otros términos próximos a él con los que guarda algún tipo de relación semántica. A continuación se definen las relaciones jerárquica y asociativa y se describen sus principales tipos.

3.1 Relaciones jerárquicas y asociativas

La relación de jerarquía (genérico / específico) expresa la subordinación entre dos términos en función de su significación. Se da una relación jerárquica cuando un término (genérico) engloba conceptualmente a otro (específico). La significación del término genérico siempre engloba la del término específico. Existen dos tipos básicos de relación jerárquica:

- Genérica, que es aquella que se da entre una clase o categoría y sus miembros o especies. A continuación se ponen diversos ejemplos:

Frutas / Naranjas

Desastres Naturales / Terremotos

Clases sociales / Burguesía

- Partitiva, que es aquella que se da entre el todo y sus partes. El todo puede ser un objeto/ente o una organización social, política, administrativa, etc. A continuación se ponen varios ejemplos:

Viviendas / Dormitorios

Fuerzas armadas / Ejército del Aire

Extremadura / Cáceres

Cuando en la relación jerárquica un término genérico tiene más de un término específico, nos encontramos ante un fenómeno que se da con cierta frecuencia y que recibe la denominación de polijerarquía. Como ejemplo se puede citar el término “iglesia”, que tiene al menos dos términos genéricos como son: “arquitectura religiosa” y “lugar de culto”.

Tradicionalmente a la hora de elaborar un tesauro el mayor esfuerzo se realizaba a la hora de construir una jerarquía completa, donde todos y cada uno de los términos que componían el tesauro debía encajar perfectamente dentro de la misma.

La relación asociativa es aquella relación que se establece entre términos próximos entre sí, que representan conceptos que pueden evocarse mutuamente por asociación de ideas, pero que no son ni equivalentes ni existe entre ellos una relación de jerarquía. Es esta una relación que puede tener muy diversa naturaleza al estar basada en la asociación de ideas. Se pueden asociar entes con procesos, procesos con objetos, objetos con propiedades, etc. La asociativa es un tipo de relación muy abierta. A continuación se enumeran algunos de estas asociaciones que dan origen a relaciones entre términos, advirtiendo que existen otras mu-

chas más que no aparecen en esta relación. Cada asociación se acompañada de un ejemplo ilustrativo.

- Acción / agente
Asistencia social / Trabajadores sociales
- Proceso / instrumento
Tráfico / Vehículo
- Fenómeno / consecuencia
Lluvia torrencial / Inundaciones
- Lugar / agentes relacionado con el lugar
Restaurante / Camarero
- Causa / efecto
Emisión de humos / Contaminación atmosférica
- Ente / propiedad
Fuego / Calor
- Materia prima / producto derivado
Petróleo / Gasolina
- Términos opuestos (antónimos)
Casado / Divorciado

En los tesauros tradicionales la relación de asociación ha sido menos considerada que la relación de jerarquía. En la mayoría de los casos se ha dedicado más esfuerzo a la construcción de la estructura jerárquica y al encaje de cada uno de los términos en dicha estructura que al establecimiento de relaciones de asociación. Un término cuantos más términos relacionados tenga mejor, siempre y cuando sean relaciones consistentes. No se debe de ser restrictivo a la hora de establecer relaciones asociativas. En la elaboración de un tesoro se debe de hacer un esfuerzo en el establecimiento de relaciones de asociación, pues ayudan a ampliar el número de relaciones por término, lo que enriquece la información sobre el mismo.

3.2 Enriquecimiento de la indización

Los tesauros tradicionalmente han sido utilizados por los indizadores como un recurso para enriquecer su indización. Repasemos cual es la manera de proceder de estos en la selección de términos a utilizar en la indización de documentos. Al indizar un documento, en primera instancia se preseleccionan una serie de términos extraídos del texto analizado. Posteriormente se añaden otros, que, aunque no estén presentes en el documento, están relacionados con el tema y ayudan a completar la descripción del contenido del mismo. Es este un paso aconsejable, pues al enriquecer la indización de un documentos con nuevos términos se amplían las posibilidades de recuperación del mismo por parte de los usua-

rios, pues se ofrecen más puntos de acceso al mismo. Cuantos más términos de indización contenga un documento, más posibilidades habrá de recuperarlo. Ni que decir tiene, que todos los términos de indización seleccionados tienen que ser “pertinentes”, es decir representativos del contenido del documento, pues si no se estaría introduciendo “ruido” en el sistema. En indización hay que ser todo lo exhaustivo que se pueda, pero sin dejar de ser preciso.

Es en este momento de enriquecimiento de la indización, cuando el tesoro es un recurso muy útil, ya que puede sugerir términos nuevos que en un principio no estaban en el texto y en los que no ha reparado el indizador. En esta fase, ante la ausencia de un tesoro, el indizador sólo puede recurrir a sus capacidades relacionadas tanto con su conocimiento del tema como con su estado de inspiración, por lo que el resultado de una indización puede ser muy variable. Sin embargo con la existencia de un tesoro el margen de variabilidad disminuye, pues el indizador puede recurrir a una herramienta estable donde las relaciones entre términos están preestablecidas, lo que le ayuda a seleccionar otros términos en los que podría no haber pensado en un primer momento.

3.3 Expansión de la consulta

Tradicionalmente los que suelen utilizar los tesauros como recurso de enriquecimiento terminológico son los profesionales, pues son los que conocen la mecánica de uso de los mismos. Los usuarios, a no ser que sean expertos en el manejo de bases de datos, no sacan a los tesauros el rendimiento debido, pues en la mayoría de los casos no saben para qué sirven exactamente ni cómo funcionan.

En la formulación de una consulta, la elección de los términos de búsqueda es muy importante, pues de ello depende el éxito o el fracaso de la misma. El elegir un único término para una consulta es muy arriesgado y disminuye las posibilidades de respuesta. Es frecuente que en documentos sobre un mismo tema se utilicen diferentes términos para describir el contenido. Hay que recordar que indizar tiene una componente de subjetividad grande. Ante un mismo documento analizado por dos indizadores diferentes, o incluso por el mismo indizador en dos momentos distintos, los términos seleccionados no coincidirán totalmente en todos los casos. Por esto es aconsejable utilizar en las consultas términos relacionados para aumentar las posibilidades de éxito.

Para respaldar esta afirmación se presenta el siguiente ejemplo. Como puede observarse en el Gráfico 8, ante tres documentos de temática similar con una indización no coincidente totalmente, sólo si se utiliza el término A en solitario, se recuperarán los tres documentos posibles, pero si se utiliza cualquiera de los otros tres términos (B o C o D) por separado, se perderá información, pues en los tres casos se recuperarán dos documentos de los tres posibles. Únicamente se volverán a recuperar los 3 documentos posibles, si se realiza una combinación entre los mencionados términos.

INDIZACION			CONSULTA	
Documento 1	Documento 2	Documento 3	Términos Utilizados	Documentos Recuperados
Término A	Término A	Término A	A	1, 2, 3
Término B	Término B	Término C	B	1, 2
Término C	Término D	Término D	C	1, 3
			D	2, 3
			B + C	1, 2, 3
			B + D	1, 2, 3
			C + D	1, 2, 3

Gráfico 8. Resultados de una consulta según los términos utilizados

En las bases de datos tradicionales el uso del tesoro está separado de los formularios de búsqueda. Para utilizar el tesoro hay que trasladarse a una opción propia e independiente de visualización del mismo, buscar allí los términos, seleccionar los adecuados y finalmente lanzar la consulta en la base de datos propiamente dicha. Actualmente el uso del tesoro en una base de datos es opcional. Si el usuario quiere utilizar el tesoro, lo utiliza, pero no está obligado a hacerlo.

El tesoro debería servir de ayuda en la expansión de las consultas de los usuarios. Sería mucho más eficiente que la consulta al tesoro se produjera de forma automática. El tesoro debería activarse automáticamente desde los formularios de búsqueda, de tal forma que al escribir una palabra o término sobre cualquier tipo de formulario, simple o avanzado, el sistema automáticamente buscara dicha palabra o término en el tesoro. El usuario a continuación visualizaría el término solicitado acompañado de todos aquellos otros términos relacionados con él, y a la vista de lo cual seleccionaría definitivamente los términos de su búsqueda. Una vez hecho esto, el sistema interrogaría la base de datos documental con los términos seleccionados.

TÉRMINO SOLICITADO: Submarinismo		
RESPUESTA OBTENIDA:		
Término Genérico	Términos Específicos	Términos Asociados
• Deportes acuáticos	• Espeleobuceo	• Actividades submarinas
	• Pesca submarina	• Aletas de buceo
	• Submarinismo en pecios	• Barotraumatismo ocular
	• Submarinismo nocturno	• Chalecos hidrostáticos
		• Cursos de submarinismo
		• Deportes subacuáticos
		• Enfermedades por descompresión
	• Inmersiones	
	• Nitrox	
	• Trajes de neopreno	

Gráfico 9. Respuesta de una consulta al tesoro

Otro de las circunstancias que se dan en el uso actual de los tesoros, es que, como en el caso de los términos equivalentes, es que actúa exclusivamente en los campos de indización. Una vez seleccionados los términos, los sistemas de recuperación lanzan la búsqueda exclusivamente en dichos

campos que es donde se controla el vocabulario, no ofreciendo la opción de lanzarla también en otros campos de lenguaje libre, como el título o el resumen o incluso en el propio texto del documento. Es útil ofrecer esta posibilidad, porque en ocasiones el indizador no es todo lo exhaustivo que debiera ser, y no utiliza todos los términos pertinentes que debiera, pero en algunos casos dichos términos, aunque no hayan sido seleccionados como términos de indización, pueden encontrarse en el título y/o en el resumen o en el caso de que exista incluso en el texto completo.

Lo que se ha explicado hasta aquí es el uso de tesauros en bases de datos estructuradas y con control del vocabulario, pero los tesauros también pueden ser muy útiles en aquellas otras bases sin estructurar y donde no se controla el vocabulario. El procedimiento de búsqueda sería el mismo que en las bases de datos estructuradas: al introducir el término en el formulario de búsqueda se visualiza el tesoro, apareciendo dicho término acompañado de otros términos relacionados; a partir de ahí el usuario selecciona los términos que crea oportuno, pero en este caso al tratarse de una base de datos sin estructura, el sistema lanza la búsqueda a la totalidad del documento. Si se selecciona un término con equivalentes, el sistema busca en el documento el término seleccionado junto con todos sus equivalentes, para que no se pierdan los documentos donde aparecen exclusivamente estos últimos.

4. CONCLUSIONES Y RECOMENDACIONES FINALES

En este trabajo se ha intentado argumentar la utilidad del tesoro en la recuperación de información y su vigencia en la actualidad. Pero también se ha tratado de evidenciar la necesidad de introducir cambios en su elaboración y uso, para poder beneficiarse en mayor medida de sus potencialidades.

A continuación, y a modo de resumen, se enumeran cinco recomendaciones mencionadas en el texto, unas de forma implícita y otras de manera directa. Estas recomendaciones tienen por objetivo aumentar la eficiencia del tesoro en la recuperación de información en bases de datos documentales.

1. En cuanto a la elaboración de un tesoro se recomienda trabajar en profundidad en el establecimiento de relaciones de equivalencia y asociación, pues la existencia de ambas ayudará a evitar la pérdida de información. La relación de equivalencia ayudará a controlar más estrictamente el vocabulario y la de asociación sugerirá nuevos términos que añadir a la consulta.
2. Eliminar la categorización entre los términos sinónimos (preferentes y no preferentes) y considerarlos todos equivalentes de igual rango.
3. Automatizar la búsqueda de términos equivalentes, dando libertad al usuario para utilizar en una consulta cualquiera de ellos obteniendo siempre el mismo resultado.
4. Automatizar el uso del tesoro desde los formularios de búsqueda, obteniendo como primer resultado de una consulta los términos relacionados del término solicitado, y a partir de dicha información conformar la interrogación definitiva.
5. Posibilitar el uso del tesoro en otros campos (título, resumen, texto completo), además del de indización, para ampliar la búsqueda si así se desea.

BIBLIOGRAFÍA

- AENOR (1990). UNE 50106:1990: Directrices para el establecimiento y desarrollo de tesauros monolingües. . Madrid, AENOR.
- AITCHISON, Jean; GILCHRIST, Alan y BAWDEN, David. (2000). Thesaurus construction and use: a practical manual. 4th ed, London, ASLIB, 2000, 218 p.
- ARANO, S. y CODINA, Ll: La estructura conceptual de los tesauros en el entorno digital: ¿nuevas esperanzas para viejos problemas?. En 9es Jornades Catalanes d'Informació i Documentació / Barcelona, 25-26 de novembre 2004. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 200, 422 p.
- CURRAS PUENTE, Emilia. (2005). Ontologías, taxonomía y tesauros: manual de construcción y uso. Gijón, Trea, 337 p.
- CHAUMIER, J. (1986). Análisis y lenguajes documentales. El tratamiento lingüístico de la información documental. Mitre, Barcelona 1986, 170 p.
- GIL URDICIAIN, B. (1999). Manual de lenguajes documentales. Madrid, Noesis, 269 p.
- GREENBERG, J. (2001). Automatic query expansion via lexical-semantic relationships. Journal of the American Society for Information Science. 2001; 52(5), pp. 402-415.
- GREENBERG, J. (2001). Optimal query expansion (QE) processing methods with semantically encoded structured thesauri terminology. Journal of the American Society for Information Science and Technology, 2001, 52 (6), pp. 487-498.
- ISO (1986). ISO 2788:1986 Documentation -- Guidelines for the establishment and development of monolingual thesauri. 2nd ed. Geneva: ISO.
- LANCASTER, F. W. (1995). El control del vocabulario en la recuperación de información. Valencia, Universitat, 1995, 286 p.
- MANIEZ, J. (1993). Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales. Madrid, Fundación Germán Sánchez Ruipérez, 1993, 240 p.
- NAUMIS PEÑA, C. (2007). Los tesauros documentales y su aplicación en la información impresa, digital y multimedia. Buenos Aires: Alfagrama, 2007, 288 p.
- NISO (2003). ANSI/NISO Z39.19 - 2003 Guidelines for the Construction, Format, and Management of Monolingual Thesauri. [en línea] Bethesda: NISO.
- PEÑAS, F. VERDEJO, J. GONZALO (2002). Terminology Retrieval: towards a synergy between thesaurus and free text searching. Advances in Artificial Intelligence - IBERAMIA 2002, LNAI 2527.
- PÉREZ AGÜERA, JOSÉ RAMÓN (2004). Automatización de tesauros y su utilización en la web semántica. BiD: textos universitarios de biblioteconomía i documentació, nº 13.

ROE, Sandra K. y THOMAS, Alan R. (2004). *The thesaurus: review, renaissance, and revision*. New York [etc.]: Haworth Information Press, 209 p.

SHIRI, ALI Y REVIE, CRAWFORD (2006). Query expansion behavior within a thesaurus-enhanced search environment: A user centered evaluation. *Journal of the American Society for Information Science and Technology*, 57 (4), pp. 463-478.

SLYPE, GEORGES VAN. (1991). *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez, 1991, 200 p.

TUDHOPE, D.; ALANI, H.; JONES, C. (2000). Associative and Spatial relationships in thesaurus-based retrieval; Research and advanced technology for digital libraries: Lisbon, 18-20 September 2000. *Lecture Notes in Computer Science*, (1923), pp. 45-58.

TUDHOPE, D.; ALANI, H.; JONES, C. (2001). Augmenting Thesaurus Relationships: Possibilities for Retrieval. *Journal of Digital Information*, 2001, vol. 1, issue 8.