

Los tesauros

Damián Martínez Ferreras

PID_00143963



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundación para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Índice

Objetivos.....	5
1. El tesoro.....	7
2. Estructura del tesoro.....	9
2.1. Unidades léxicas	9
2.2. Relaciones semánticas	10
2.3. Notas y aclaraciones	13
3. Formas de presentación del tesoro.....	15
3.1. Presentación alfabética	15
3.2. Presentación sistemática	16
3.3. Presentación gráfica	17
3.4. Presentación con índices permutados	19
4. Creación y utilización de un tesoro.....	22
4.1. Creación de un tesoro	22
4.2. Utilización de un tesoro	27
4.2.1. Uso del tesoro en los procesos de indización y recuperación de la información	28
4.2.2. Revisión y actualización del tesoro	29
5. Sistemas automatizados de gestión de tesauros. Tesauros en línea.....	31
5.1. Sistemas automatizados de gestión de tesauros	31
5.2. Los tesauros en el mundo digital. Tesauros en línea	32
6. Tesauros y ontologías.....	34
6.1. Los tesauros	34
6.2. Las ontologías	35
6.3. Similitudes y diferencias entre tesauros y ontologías	37
Bibliografía.....	39

Objetivos

Con el estudio de los materiales asociados a este módulo didáctico alcanzaréis los siguientes objetivos:

- 1.** Saber qué es un tesoro, conocer su estructura y los principales elementos que lo componen y definen: construcción, mantenimiento, tipologías, formas de representación y usos del tesoro.
- 2.** Analizar el papel de los tesauros en el proceso de indización y recuperación de la información y su importancia en los sistemas documentales especializados.
- 3.** Conocer el proceso de creación de un tesoro y elaborar un microtesoro.
- 4.** Entender la función del tesoro como una herramienta específica que hace posible el objetivo final de la cadena documental, esto es, la recuperación y difusión de la información.

1. El tesoro

Etimológicamente, el término proviene del latín *thesaurus*, cuyo significado es tesoro y que poco tiene que ver con el concepto que tiene hoy para el profesional de la información.

La norma ISO 2788-1986 define un tesoro como:

“Un vocabulario controlado y dinámico, compuesto por términos que tienen entre ellos relaciones semánticas y genéricas y que se aplica a un dominio particular del conocimiento”.

Por su parte, van Slype (1991) define un tesoro como:

“Una lista estructurada de conceptos 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”.

Podríamos definir un tesoro como una lista de términos empleados para representar los conceptos, temas o contenidos de los documentos y que guardan entre sí relaciones semánticas y genéricas. Se trata de un instrumento de normalización y control del vocabulario cuyo objetivo reside en eliminar la ambigüedad del lenguaje, convirtiendo el lenguaje natural de los documentos en un lenguaje controlado.

Estáramos hablando del tesoro como un **vocabulario controlado de descriptores** con un significado previamente establecido y cuyo fin último sería el de definir de manera unívoca el contenido de un documento. El fin último de este lenguaje es el de ser puente de conexión entre los usuarios y las unidades de información, esto es, de servir tanto para la indización como para la recuperación documental.

A diferencia de otros lenguajes documentales, como las clasificaciones y las listas de encabezamientos de materia, los tesauros suelen ser lenguajes especializados, hacen referencia a una temática concreta y están contruidos para servir a una institución u organización determinada.

Frente a los lenguajes clasificatorios, cuya función es describir el tema de un documento, los términos contenidos en un tesoro responden al análisis del texto o a la materia del mismo. Un tesoro recoge todos los conceptos, y cada uno de ellos constituye un punto de acceso para la recuperación del documento.

La **potencia de un tesoro** radica, además, en la posibilidad de combinar todos esos conceptos, llamados **descriptores**, lo que le convierte en un lenguaje combinatorio mucho más rico que los tradicionales encabezamientos de materias a la hora de realizar los procesos de análisis y recuperación de la información.

Las **principales características** de un tesoro son las siguientes:

- a) Nace por la necesidad de crear nuevos sistemas de indización más flexibles que los tradicionales.
- b) Presenta mayor especificidad de los términos relativos a campos disciplinares concretos, lo que permite un análisis más preciso de los documentos.
- c) A partir de la ordenación alfabética y como complemento a la ordenación sistemática de los descriptores (por temas y significado), el tesoro facilita la consulta del analista y del usuario. Son precisamente estas relaciones semánticas que se establecen entre los descriptores (relaciones de equivalencia, de jerarquía y de asociación) las que definen un tesoro y proporcionan la ayuda necesaria en los procesos de indización y recuperación de información.

Los tesauros pueden ser de los siguientes **tipos**:

- **Generales o especializados**, según su campo de aplicación.
- **Multidisciplinares o monodisciplinares**, según el número de disciplinas o temas que incluyen.
- **Macrotesauros o microtesauros**, según la naturaleza y compatibilidad con otros tesauros. Los macrotesauros, aunque poseen las características de un tesoro en cuanto a estructura y función, y al mismo tiempo, sus diferentes campos semánticos o microdisciplinas, pueden funcionar como si fueran tesauros independientes, con total autonomía. En este sentido, se podría hablar de un conjunto de tesauros más pequeños o microtesauros.
- **Monolingües o multilingües**, según las lenguas utilizadas.
- **Alfabéticos o sistemáticos**, según la presentación utilizada para mostrar los términos y sus relaciones semánticas.

Ved también

Las relaciones semánticas se estudian en el subapartado 2.2 de este módulo didáctico.

Ved también

Los tesauros alfabéticos o sistemáticos los trataremos en el apartado 3 de este módulo didáctico.

2. Estructura del tesoro

La estructura conceptual de los tesauros no es arbitraria en ningún sentido, puesto que existen tanto manuales de elaboración de tesauros como normas internacionales que recomiendan los elementos estructurales a tener en cuenta. La consulta tanto de manuales como de normas con relación a la construcción de tesauros nos lleva a una estructura conceptual que implica relaciones semánticas y en la que se tienen en cuenta los términos como representaciones léxicas de los conceptos.

En consecuencia, podemos basar la estructura de un tesoro en los siguientes elementos: unidades léxicas, relaciones semánticas y notas y aclaraciones.

2.1. Unidades léxicas

Las **unidades léxicas** son grupos de descriptores que agrupan los términos de indización bien por campos (temas), bien por clases de términos (facetas).

En los tesauros, los descriptores representan conceptos o nociones que pueden ser concretos, abstractos, entidades individuales o clases. Para expresar los conceptos, los tesauros contienen unidades léxicas llamadas descriptores y no descriptores.

1) **Descriptores**: términos permitidos. Un descriptor es una palabra o conjunto de palabras del lenguaje natural que responden a un concepto, expresado con un único término y ese término responde a un único concepto. Esta univocidad se consigue seleccionando el descriptor entre sus sinónimos y estableciendo una red de relaciones semánticas con el resto de descriptores del tesoro. Dicho término será utilizado para el análisis y la recuperación del documento. Los descriptores pueden ser:

a) **Términos simples**: se emplean cuando el concepto es claro en sí mismo, sin necesidad de añadir ninguna otra palabra, o cuando se usa como genérico de términos más concretos o específicos, como, por ejemplo, **Coches**.

b) **Términos compuestos**: la mayoría de los términos de un tesoro adquieren su verdadero sentido cuando se convierten en términos compuestos, ya que así se logra una mayor especificidad. La estructura sintáctica de los términos compuestos suele ser la siguiente:

- Sustantivo + adjetivo: **Coches deportivos**.

- Sustantivo + sintagma preposicional: **Coches de época**.
- Sustantivo + sintagma aposicional: **Coches F1**.

2) **No descriptores**: términos prohibidos. Palabras incluidas en el tesauro, que pertenecen a una lista de sinónimos y términos emparentados unidos a los descriptores por una relación de equivalencia semántica, que son susceptibles de aparecer en los documentos o en las preguntas, pero que no son utilizados para formular la interrogación al sistema. La inclusión de estos en un tesauro tiene como finalidad mejorar la coherencia de la representación de los documentos o de una pregunta al enviarnos al término de indización.

Cuando un mismo concepto se puede expresar mediante dos o más sinónimos, uno de ellos tiene que seleccionarse como término preferente o descriptor. Los demás pasan a ser no descriptores.

Ejemplo

```
"Coche" (no descriptor)  
USE "Automóvil" (descriptor)
```

Si buscamos información de ventas de coches, caeremos en lo que se denomina **silencio documental**, ya que habrá documentos que hablen sobre este tema pero que estén indizados utilizando la palabra “Automóvil”. Asimismo, si buscamos sólo por la palabra “Automóvil”, nos aparecerán exclusivamente documentos que contengan esta palabra, y no, la palabra “Coche”.

Este problema se hubiera evitado si en el tesauro hubiera figurado la relación arriba expresada y, en consecuencia, tanto a la hora de indicar un documento sobre ventas de coches como a la hora de buscar información sobre este tema, no hubiéramos tenido otra opción que utilizar el término pertinente o descriptor, esto es, el término “Automóvil”.

En consecuencia, podemos decir que los no descriptores forman parte del tesauro estableciendo una relación de equivalencia con el descriptor pertinente, pero no cuentan ni para la indización ni para la recuperación documental. Gracias a los no descriptores, descartamos todo aquello que podemos denominar accidentes del lenguaje natural, con el fin de no retener más que un término referencial.

2.2. Relaciones semánticas

Los campos semánticos agrupan aquellos descriptores adscritos al mismo ámbito de conocimiento del tesauro y esta agrupación se realiza por medio de una relación de pertenencia a dichos campos o microdisciplinas. Estas no se

pueden utilizar como descriptores, esto es, no intervienen en la indización y recuperación de los documentos, y sólo sirven para agrupar a los descriptores afines a su temática.

En cuanto a las **relaciones semánticas** propiamente dichas, son todas aquellas relaciones que se establecen entre los términos de un tesoro y tienen como objetivo principal especificar el significado de un descriptor asignándolo a un campo semántico o microdisciplina determinada.

Las relaciones semánticas establecidas en los tesauros son de cuatro tipos: de pertenencia, de equivalencia, de jerarquía y de asociación. A continuación veremos una descripción de cada una de ellas y recomendamos recurrir a los manuales y a las normas citadas en la bibliografía para una definición más detallada:

1) **Relaciones de pertenencia:** se establecen entre un descriptor y el campo semántico al que pertenece.

Notación

No existen siglas o abreviaturas normalizadas para expresar la relación de pertenencia.

2) **Relaciones de equivalencia:** la equivalencia semántica entre descriptores se aplica a los sinónimos, términos que poseen significados iguales en diferentes contextos, y a los cuasi-sinónimos, términos considerados de significado diferente en el uso común pero tratados como si fueran sinónimos en materia de indización.

Notación

Las relaciones de equivalencia se suelen representar mediante el término USE (relación entre el término no preferente y el término preferente, o término a usar más aceptado) y las siglas UP (Usado por), que en inglés son USE y UF.

Notación usada en las relaciones de equivalencia

	Catalán	Castellano	Inglés
Use	EM	USE	USE
Usado por	EP	UP	UF

Descriptor	No descriptor
Automóvil	coche
UP coche	USE Automóvil

Se consigue así evitar la ambigüedad terminológica de la sinonimia del lenguaje natural expresada en varios términos para referirse a un mismo concepto.

3) Relaciones de jerarquía: son las relaciones que se establecen entre los descriptores. En una relación de jerarquía, un término es superior o genérico de otro y este otro término es inferior o específico del anterior. Esta estructura clasificatoria distingue al tesoro de las listas alfabéticas de materias o palabras clave, así como de diccionarios y léxicos. De esta forma se crean campos conceptuales que clasifican los términos y los agrupan en un campo delimitado de significados o facetas que se van agrupando en disciplinas y sub-disciplinas.

Notación

Para la presentación de las relaciones jerárquicas se suelen utilizar determinadas siglas como TG (*término genérico*) y TE (*término específico*); o por su siglas en inglés, BT (*broad term* o *término más amplio*), NT (*narrow term* o *término más específico*).

Notación usada en las relaciones de jerarquía

	Castellano	Inglés	Catalán
Término genérico	TG	BT	TA (término amplio)
Término específico	TE	NT	TE

Descriptor	Descriptor
Productos lácteos	Yogur
TE yogur	TG Productos lácteos

Las relaciones de jerarquía constituyen el rasgo más característico de los tesauros y proporcionan una ayuda importantísima al usuario en los procesos de indización y recuperación documental.

4) Relaciones de asociación: son las relaciones asimétricas establecidas entre descriptores de distintas familias, distintas a las jerárquicas o de equivalencia, y que designan conceptos capaces de evocar mutuamente, por asociación de ideas, relaciones entre ellos.

Notación

Este tipo de relación se suele representar mediante las siglas TR (Término Relacionado) o en inglés RT (Related Term).

Notación usada en las relaciones de asociación

	Castellano	Inglés	Catalán
Término relacionado	TR	RT	TR

Descriptor	Descriptor
Cargos municipales	Administración Central
TR Administración Central	TR Cargos municipales

La mayoría de las relaciones asociativas se establecen entre términos que comparten alguna característica común pero que pertenecen a campos semánticos diferentes.

Ayudan al usuario en las tareas de indización y recuperación documental puesto que proporcionan nuevos descriptores relacionados que facilitan la formulación de dichas tareas.

2.3. Notas y aclaraciones

Cuando el significado de un descriptor es difuso o no está bien precisado por su contexto semántico o léxico, se puede añadir una explicación adicional relacionada con el descriptor, de forma que no se produzca ambigüedad dentro del tesoro.

Estas explicaciones pueden ser notas (explicativas, históricas o de aplicación) o modificadores:

1) **Notas.** Utilizan como notación NA:

Ejemplo

INSTALACIONES DEPORTIVAS

NAÚsese este término para referirse a instalaciones de los distintos deportes desde el punto de vista de sus características y descripción. Para referirse a aspectos económicos y de gestión úsese el EQUIPAMIENTOS DEPORTIVOS

2) **Modificadores.** Añaden la explicación entre paréntesis para evitar la ambigüedad.

Ejemplo

Juegos Olímpicos (Invierno).

3. Formas de presentación del tesoro

Existen diferentes formas de presentar los términos que componen un tesoro y las relaciones que se establecen entre esos términos dentro del tesoro. Los principales tipos de presentaciones son los siguientes: alfabética, sistemática, gráfica y con índices permutados.

3.1. Presentación alfabética

En la presentación alfabética todos los términos del tesoro, descriptores y no descriptores, se encuentran agrupados en una sola secuencia alfabética, acompañados de sus relaciones. Se utiliza principalmente en el momento de la indización de los documentos ya que un simple vistazo al listado alfabético nos permite identificar los descriptores pertinentes.

En la presentación alfabética, las informaciones asociadas a cada descriptor (informaciones que especifican las relaciones semánticas del tesoro) van precedidas de sus respectivas abreviaturas o notaciones de la siguiente manera:

Información asociada y descriptores usados en la presentación alfabética

Descriptor	Información asociada
DOM	Campo semántico o microdisciplina a la que pertenece el descriptor (relación de pertenencia)
NA	Nota de aclaración para evitar la ambigüedad del término
UP	<i>Usado por.</i> El término preferente o descriptor hace una referencia al no descriptor o término no preferente (relación de equivalencia)
USE	El término no preferente o no descriptor hace referencia al descriptor (relación de equivalencia)
TC	Término cabecera. No es de uso obligatorio (relación de jerarquía)
TG	Término genérico que hace referencia al término al cual pertenece el descriptor en cuestión (relación de jerarquía)
TE	Término específico que hace referencia al término que se incluye dentro del término en cuestión (relación de jerarquía)
TR	Término relacionado que hace referencia al término que tiene relación con el descriptor que nos ocupa, pero que no pertenece al mismo grupo del descriptor analizado (relación de asociación)

Ejemplo

Bebidas espirituosas

DOM Industrias

NA Se consideran bebidas espirituosas aquellas bebidas, exceptuando los vinos tradicionales y la cerveza, con contenido alcohólico procedentes de la destilación de materias primas agrícolas.

UP Licores y aguardientes

TC Industria alimentaria

TG Industria vinícola

TE Brandy
Ginebra

Ron

Vodka

Whisky

TR Alcoholismo

En cuanto a las otras lenguas, a continuación se presenta una tabla con las equivalencias de las notaciones:

Descriptor	Castellano	Inglés	Catalán
DOM	No es obligatorio	No es obligatorio	No es obligatorio
NA	NA	SC	NA/NE
UP/USE	UP/USE	UF/USE	EP/EM
TC	TC	TT	TC
TG	TG	BT	TA
TE	TE	NT	TE
TR	TR	RT	TR

3.2. Presentación sistemática

La presentación sistemática, también llamada jerárquica, muestra los descriptores agrupados por campos semánticos o microdisciplinas y, dentro de estos, por cadenas jerárquicas, en función de las relaciones jerárquicas de los términos (término genérico y/o específico). En la presentación sistemática los títulos de los campos semánticos encabezan la jerarquía y a continuación aparecen por orden jerárquico el conjunto de términos genéricos y específicos.

La ordenación de los descriptores que pertenecen al mismo campo semántico se realiza por sangrados (una sangría a la derecha por cada nivel jerárquico) y por una ordenación alfabética de los descriptores que pertenecen al mismo nivel jerárquico y que, por tanto, pertenecen al mismo descriptor genérico.

Ejemplo

Ponemos en negrita los términos empleados al tratar la representación alfabética:

```
Industrias  
  Industria alimentaria  
    Industria cárnica  
    Industria conservera  
    Alimentos congelados  
    Industria del aceite  
      Aceite de girasol  
      Aceite de oliva  
  Industria vinícola  
    Bebidas alcohólicas  
      Bebidas espirituosas  
        Brandy  
        Ginebra  
        Ron  
        Vodka  
        Whisky  
      Licores y aguardientes  
        Cava  
        Vino
```

3.3. Presentación gráfica

La representación gráfica permite visualizar las relaciones entre los términos y su estructura jerárquica de forma muy efectiva. Las relaciones se disponen como una figura que permite al usuario asociar los términos que se encuentran relacionados.

Sin embargo, necesitan ir acompañadas de un índice alfabético, ya que es muy difícil representar aquellas relaciones que no aparecen en el gráfico, como pueden ser los no descriptores o las notas aclaratorias. Se da así la circunstancia de que la sección auxiliar del índice contiene más información que la parte en teoría principal, lo que convierte a esta parte gráfica en una presentación de apoyo.

Existen dos tipos de representación gráfica: la estructura arborescente y el diagrama de flechas.

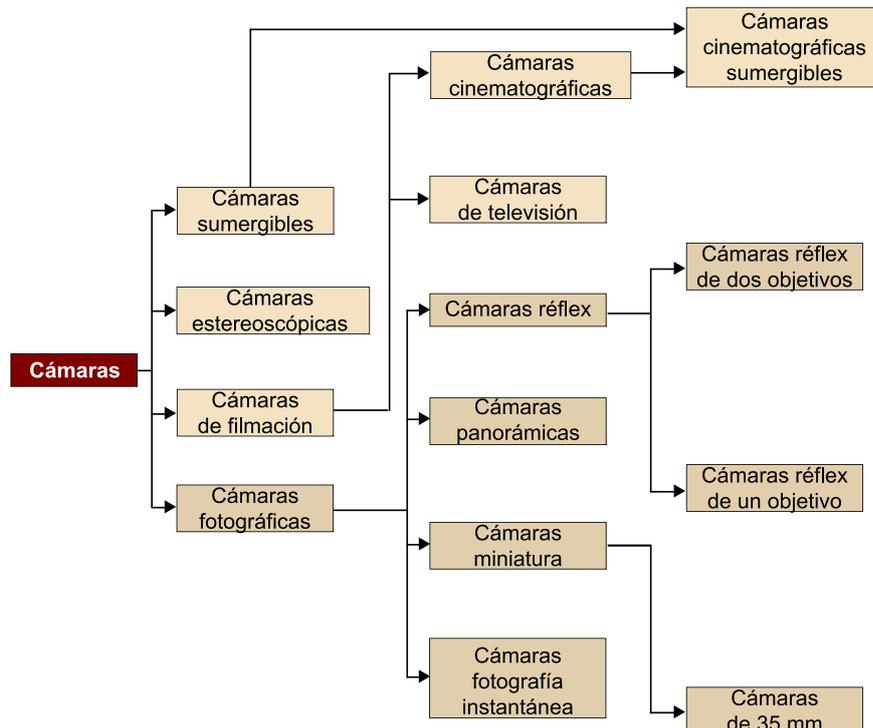
1) Estructura arborescente

Esta representación se caracteriza por la forma de árbol, y en ella el término principal se sitúa en una posición destacada del gráfico. A continuación los términos específicos se sitúan en función de los diferentes grados de jerarquía que poseen.

Este tipo de presentación sólo permite visualizar las relaciones jerárquicas siendo necesario el índice alfabético para entender el resto de relaciones que se dan entre los términos (de equivalencia, asociación y notas de aplicación).

Ejemplo

Figura 1. Ejemplo de estructura arborescente



Fuente: Norma UNE 50-UNE-50-106-90 (pág. 44).

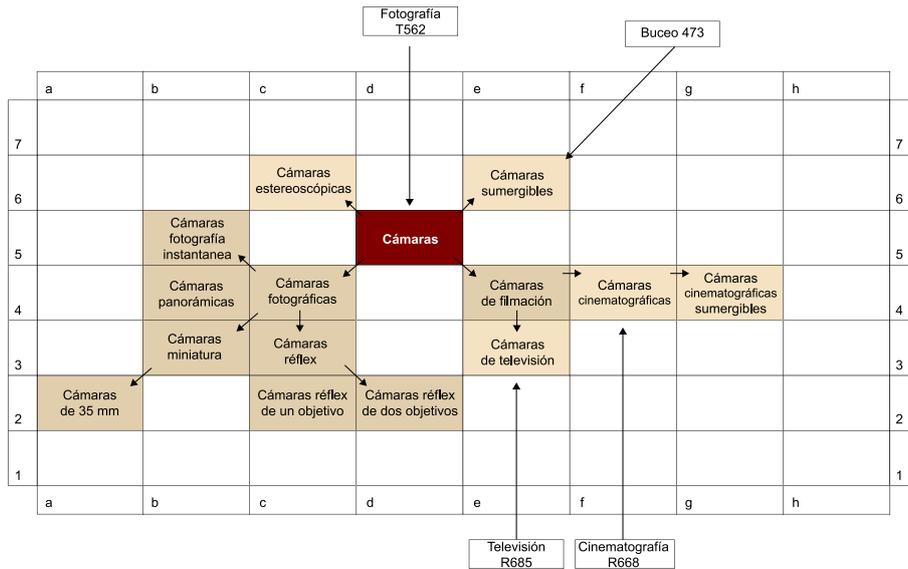
2) Diagrama de flechas

En esta presentación cada campo semántico aparece por separado y tiene una forma de presentación rectangular:

- En el centro del rectángulo se encuentra el campo semántico o término principal y a su alrededor gravitan los descriptores que pertenecen a dicho campo semántico, así como las relaciones jerárquicas y asociativas que se establecen entre ellos. Las relaciones jerárquicas se indican por medio de flechas que van de los términos más genéricos hacia los más específicos y las relaciones asociativas se fijan a través de líneas rectas que conectan con los descriptores relacionados.
- Fuera del rectángulo se encuentran los descriptores que pertenecen a otros campos semánticos pero que están relacionados con los descriptores situados en el interior del rectángulo. Al igual que en el caso anterior, se utilizan flechas o líneas rectas para enlazar con los descriptores de dentro del rectángulo.

Como ocurre con la representación arborescente, la representación mediante diagrama de flechas necesita el índice alfabético para entender el resto de relaciones (de equivalencia, asociación y notas de aplicación) que se dan entre los términos.

Figura 2. Ejemplo de diagrama de flechas



Fuente: Norma UNE 50-UNE-50-106-90 (pág. 46).

3.4. Presentación con índices permutados

Un índice es una serie ordenada de puntos de acceso que nos remiten de un término conocido por el usuario a los conceptos contenidos en los documentos. Su objetivo se centra, por tanto, en la recuperación documental. La automatización de los sistemas documentales ha posibilitado la realización de índices más complejos, con permutaciones de los términos.

Los índices permutados muestran la relación alfabética de todos los términos significativos de un tesoro mediante la agrupación de los descriptores afines desde el punto de vista semántico, facilitando de esta forma la recuperación de los descriptores compuestos independientemente de su situación dentro del descriptor. En definitiva, se recurre a una agrupación de descriptores afines desde el punto de vista semántico que en otro tipo de representación aparecerían más distanciados.

La aplicación de estos tipos de índices a los descriptores de un tesoro ha generado nuevos tipos de presentaciones de sus términos. Entre estos tipos de índices producidos por el ordenador, destacan los índices KWIC, los KWOC y los KWAC.

1) Índices KWIC

Los índices KWIC (*key words in context*) muestran una relación de descriptores en la cual las palabras significativas se ordenan alfabéticamente en una columna central. El resto del texto se cita antes y después de dichas palabras.

El KWIC crea el índice de tal manera que una columna central ordenada de forma alfabética muestra la palabra clave con una tipografía destacada (esta representación se hace para todas la palabras claves del documento, es decir,

se produce una permutación de términos para dejar siempre en la columna central el término significativo). Un mismo documento aparecerá representado de este modo en el sistema tantas veces como términos indizados posea en el índice KWIC. La **notación** del índice es **numérica**.

Ejemplo de índice KWIC

reajuste del funcionamiento de **COMPRESORES** centrífugos. Posibilidades de químicos - La selección de **COMPRESORES** para las industrias de procesos velocidad. Particularidades de **COMPRESORES** de procesos con rotor de gran de pirolisis - Obtención de **COQUE** electródico a partir de productos líquidos Consideraciones críticas sobre **COSMÉTICOS** para la piel

2) Índices KWOC

En los índices KWOC (*key word out context*), los descriptores se enumeran poniendo todas las palabras importantes en orden alfabético, seguidas por el nombre entero del descriptor.

Ejemplo de índice KWOC

COMPRESORES

Posibilidades de reajuste del funcionamiento de **compresores** centrífugos
La selección de **compresores** para las industrias de procesos químicos
Particularidades de **compresores** de procesos con rotor de gran velocidad

COQUE

Obtención de **coque** electródico a partir de productos líquidos de pirolisis

COSMÉTICOS

Consideraciones críticas sobre **cosméticos** para la piel

3) Índices KWAC

El índice KWAC (*key word and context*) es una mezcla del KWIC y el KWOC.

Ejemplo de índice KWAC**COMPRESORES**

reajuste del funcionamiento de **COMPRESORES** centrífugos. Posibilidades de químicos - La selección de **COMPRESORES** para las industrias de procesos velocidad. Particularidades de **COMPRESORES** de procesos con rotor de gran

COQUE

de pirolisis - Obtención de **COQUE** electródico a partir de productos líquidos

COSMÉTICOS

Consideraciones críticas sobre **COSMÉTICOS** para la piel

4. Creación y utilización de un tesoro

Antes de iniciar la construcción de un nuevo tesoro debemos considerar las siguientes cuestiones:

a) Comprobar si ya existe algún tesoro sobre la temática que nos interesa y si dicho tesoro puede servir para nuestro sistema documental.

b) Si decidimos iniciar la construcción de uno nuevo, deberemos establecer las características que tendrá, esto es, temas que tiene que tratar, tamaño del tesoro, si será monolingüe o multilingüe, tipo de relaciones entre los descriptores y otra serie de cuestiones formales relativas a la tipografía de los términos, longitud de los mismos, cuestiones relativas a su género, forma, etc.

4.1. Creación de un tesoro

La construcción del tesoro comprende ocho o nueve etapas, dependiendo de si es monolingüe o multilingüe (Normas UNE-50-106 y UNE 50-125):

- Recolección del vocabulario en lenguaje natural dentro de los dominios que cubra el tesoro.
- Subdivisión del conjunto de los dominios que se van a cubrir en una serie de microdisciplinas.
- Transformación del vocabulario libre en un lenguaje controlado, estableciendo las relaciones de pertenencia, de equivalencia semántica, de jerarquía, y redactando las notas explicativas.
- Búsqueda de las equivalencias interlingüísticas (si se trata de un tesoro multilingüe).
- Enriquecimiento del tesoro por medio de relaciones asociativas.
- Elaboración del borrador del tesoro.
- Formación de los indicadores.
- Test del tesoro.
- Revisión final y primera edición.

1) Recolección de los términos

Esta fase consiste en buscar y registrar todas las palabras y expresiones significativas que intervienen en la disciplina o disciplinas cubiertas por el tesoro.

Pueden utilizarse algunas **fuentes terminológicas alternativas** o complementarias:

- Un lenguaje documental ya existente en el servicio de documentación: sistemas de clasificación, listados de palabras clave o de descriptores libres, listas de autoridades.
- Tesoros ajenos al servicio de documentación, pero dedicados a los mismos dominios o a dominios próximos, descubiertos gracias a alguno de los repertorios de tesoro disponibles en el mercado. Un tesoro ya existente resulta útil porque aporta una colección organizada de conceptos de la que se podrá extraer lo que convenga, pero rara vez se adoptará en su totalidad para organizar el propio patrimonio informativo. En Internet se pueden localizar multitud de tesoros disponibles para su consulta.
- Terminología utilizada en los tratados, manuales, léxicos especializados, reglamentaciones técnicas publicadas o específicas del organismo. Esas fuentes tienen la gran ventaja de ofrecer una terminología generalmente admitida y estructurada por sus autores.
- Bases de datos terminológicas.
- Indización intelectual, en lenguaje natural, por medio de descriptores libres, de una muestra de varios cientos de documentos representativos para cada uno de los grandes dominios que va a cubrir el tesoro. Se obtiene de esta manera un inventario no ordenado de la terminología efectivamente utilizada por los autores para designar los conceptos que ellos manejan habitualmente, puede que en varias lenguas, si se trata de una muestra multilingüe para crear un tesoro multilingüe.
- Indización automática, en lenguaje natural, por medio de palabras clave, de esos mismos documentos. Este método es poco eficaz porque la lista obtenida sólo incluye palabras aisladas (unitérminos), que sólo en algunos casos, cuando están esclarecidas por su contexto, sirven para encontrar los conceptos. Además, aunque se utilice una gran lista de palabras vacías, el ordenador proporciona un porcentaje extremadamente elevado de términos sin ningún interés discriminante.
- Consulta a especialistas, vinculados con el organismo al que pertenece el sistema documental. Esta fuente es indispensable para obtener la terminología propia de las diversas prácticas profesionales dentro del organismo, así como una explicitación de la acepción específica de ciertos términos para estos especialistas.

2) Creación de los campos semánticos o microdisciplinas

Con todo el conjunto de términos recogidos durante la fase anterior podemos establecer un listado provisional de microdisciplinas que completarán el cuerpo del tesoro.

La **lista de las microdisciplinas** tiene como única finalidad preparar un sistema de agrupación de los descriptores que sea aceptable para el usuario y cómodo de consultar. De esta forma la lista así dispuesta será utilizada directamente durante la siguiente etapa y se revisará continuamente a lo largo de todo el proceso de construcción del tesoro.

3) Reducción a un lenguaje documental

En esta fase se establecen las relaciones entre los descriptores definidos y se introducen las notas aclaratorias precisas:

a) Relaciones de pertenencia: lo primero que debemos conseguir es la distribución de todos los términos del lenguaje natural recopilados durante la primera fase en las diferentes microdisciplinas que hemos establecido en la segunda. De esta manera, todos los términos quedan adscritos de forma provisional a un campo semántico o una microdisciplina.

b) Relaciones de equivalencia: una vez que tenemos todos los descriptores distribuidos entre las diferentes microdisciplinas, debemos examinarlas una por una con el fin de establecer entre sus conceptos el listado de términos sinónimos o cuasi-sinónimos. De esta manera, empezamos a seleccionar los términos que pasarán a ser descriptores y los que serán no descriptores, y que tendremos que enlazar por medio de una relación de equivalencia semántica.

Los criterios para elegir el término descriptor entre sus sinónimos se deben fundamentar en la selección del menos ambiguo o del más utilizado en la profesión o por los usuarios del sistema documental. Probablemente, deberemos consultar obras terminológicas para asegurar la corrección de la ortografía de estas palabras.

c) Notas explicativas: si el término seleccionado como descriptor todavía puede provocar alguna ambigüedad dentro del tesoro, tendremos que redactar la nota explicativa correspondiente (de definición, histórica o de aplicación).

d) Relaciones jerárquicas: el paso siguiente consiste en establecer las cadenas jerárquicas entre descriptores genéricos y específicos dentro de cada campo semántico o microdisciplina.

Partimos del descriptor más genérico (término cabecera), y desarrollamos en forma arborescente las diferentes ramificaciones hasta llegar a los descriptores más específicos. La ordenación de los términos dentro de un mismo nivel jerárquico puede hacerse por facetas o temas.

4) Búsqueda de equivalencias interlingüísticas

En el caso de los tesauros multilingües, una vez que se ha fijado el contenido semántico de los descriptores (gracias a las relaciones de pertenencia, equivalencia y jerarquía y a las notas explicativas), podemos empezar a establecer las equivalencias lingüísticas del término en los diferentes idiomas definidos en los tesauros.

5) Establecimiento de relaciones de asociación

El paso siguiente en el proceso de construcción del tesoro es el establecimiento de las relaciones de asociación entre los descriptores de sus diferentes campos semánticos. Es sin duda la fase más laboriosa, ya que se debe examinar cada uno de los descriptores del tesoro con el fin de encontrar términos afines temáticamente entre el resto de los descriptores.

También se deberán revisar los descriptores localizados en otros campos semánticos con los que el descriptor que se está examinando pueda tener una relación de asociación. Si esto se produce, las relaciones de asociación podrán ser creadas así mismo entre descriptores que pertenezcan a cadenas jerárquicas diferentes dentro de la misma disciplina.

En ningún caso se crearán relaciones asociativas:

- Descriptores de la misma cadena jerárquica.
- Descriptores y no descriptores.

6) Elaboración del borrador del tesoro

La elaboración de todo el proceso anterior dará lugar a una edición piloto o experimental del tesoro. Se debe procurar hacer, por lo menos, una presentación alfabética y otra jerárquica.

Esta edición, o edición 0, está destinada a la formación de los indizadores o analistas y a la prueba de funcionamiento del tesoro.

El ejercicio dura varios días y consiste en:

- Presentar el borrador del tesoro.
- Hacer que los indizadores o analistas indiquen los mismos documentos.

Tras la indización de cada documento:

- Se dialoga sobre los puntos de vista adoptados por cada indizador, con el fin de hacer que progresivamente coincidan.
- Se calcula la tasa de coherencia (ratio entre el número de descriptores comunes y el número total de descriptores distintos utilizados por dos personas o dos grupos para indizar el mismo documento); al principio del ejercicio la tasa será escasa (entre el 20 y el 30 % según van Slype), pero al terminar la formación tenderá hacia su valor ideal (entre el 50 y el 80 %).

7) Test del tesoro

El test del tesoro tiene como finalidad la identificación y resolución de los problemas que presenta el tesoro como herramienta para la indización de un sistema documental determinado. Todas las anomalías detectadas por los indizadores durante la indización de una muestra significativa de documentos se señalan en un informe en forma de propuestas para:

- Añadir descriptores y no-descriptores olvidados durante la construcción.
- Modificar las estructuras jerárquica y asociativa.
- Explicitar por medio de notas explicativas descriptores todavía ambiguos.

Para organizar este test:

- a) Se seleccionan varios cientos de documentos y, si es posible, de consultas.
- b) Se hace que los documentalistas los indiquen usando el nuevo tesoro.
- c) Se pide a los documentalistas que preparen un informe especial en el que aparezcan identificados los tipos de propuestas descritos más arriba.
- d) Se realiza, por último, una estadística de estas propuestas, que serán valoradas por los responsables de elaborar el tesoro.

8) Revisión final y primera edición

Una vez que se han añadido y revisado las propuestas de los indizadores, estamos en condiciones de preparar la edición definitiva del tesoro.

La edición de un tesoro tiene que incluir, en forma de prefacio o introducción, las siguientes informaciones:

- Nombres de los autores y del servicio o centro responsable de la edición.
- El objetivo que se quiere alcanzar con la creación del tesoro.

- El listado de las microdisciplinas que contiene.
- Las fuentes terminológicas utilizadas.
- El proceso de construcción del tesoro.
- Las características del tesoro: lenguas, número de descriptores, de no descriptores, de microdisciplinas, de términos específicos, de términos relacionados, etc.
- Algunos ejemplos de las presentaciones del tesoro: relaciones incluidas, abreviaturas o signos utilizados.
- Instrucciones para la indización y la formulación de consultas.
- Información sobre el mantenimiento del tesoro: futuras perspectivas, uso de sistemas automatizados, etc.

Una vez redactado el prefacio con esta información, podemos iniciar la impresión o edición del tesoro.

4.2. Utilización de un tesoro

El tesoro es una herramienta viva y dinámica que se nutre de nuevos descriptores, gracias a la indización de nuevos documentos y a los cambios cualitativos que el mismo proceso de construcción genera. Por consiguiente, el mantenimiento y la actualización periódica del tesoro se convierten en una necesidad.

Se debe actualizar periódicamente, tanto para corregir errores y omisiones como para adaptarlo a la evolución científica de la disciplina que abarca. El proceso de control debe seguir los siguientes pasos:

1) Control de la frecuencia de uso de los descriptores

Para llevar esto a cabo, se debe confeccionar periódicamente un listado de los descriptores con el número de ocurrencias, es decir, la cantidad de veces que han sido utilizados en la indización de los documentos.

La elaboración de estos listados, normalmente facilitada por el *software* de gestión documental, permite detectar:

- Descriptores que se han utilizado muy poco, lo que indica que el descriptor es problemático.
- Descriptores que se han utilizado mucho, lo que también puede indicar problemas.

Descriptores poco utilizados

Esto puede suceder, por ejemplo, porque el descriptor es demasiado específico, no se entiende su significado, etc.

2) Detección de ausencia de descriptores

Con el uso del tesoro en el proceso de indización se pueden detectar nuevos términos que aparecen en los documentos, pero que no están incluidos en el tesoro.

Generalmente, la mayoría de los sistemas documentales, especialmente las bases de datos documentales, ya han previsto esta incidencia con la inclusión de un campo denominado descriptores candidatos en la base de datos. Todos los conceptos que no están en el tesoro se incluyen en este campo.

Periódicamente, los responsables del tesoro elaboran un listado de estos descriptores candidatos y estudian su posible inclusión en el tesoro.

Descriptores demasiado utilizados

Esto puede suceder porque el descriptor es demasiado genérico y en consecuencia hay que crear descriptores más específicos, etc.

3) Problemas de uso del tesoro

Con el uso del tesoro también se empiezan a detectar diferentes problemáticas:

- Descriptores con un significado muy próximo y que, por tanto, provocan la duda del indizador en el momento de seleccionarlos.
- Relaciones semánticas cuya ausencia se note. Por ejemplo, la de un no descriptor que ayudaría a localizar el descriptor admitido.
- Divergencias entre las diferentes presentaciones del tesoro: un término aparece adscrito a un campo semántico determinado en la presentación jerárquica, y a otro en la alfabética.

La detección de estos problemas se debe llevar a cabo con la colaboración estrecha de los analistas, que son las personas que los encuentran durante el análisis del documento.

4.2.1. Uso del tesoro en los procesos de indización y recuperación de la información

El uso del tesoro en el proceso de indización de un documento tiene que ver fundamentalmente con la traducción de los conceptos extraídos del lenguaje natural a un lenguaje controlado, es decir, con la conversión de los términos del lenguaje natural en descriptores del tesoro.

Ahora bien, al tratarse de un lenguaje documental muy flexible, a la hora de indizar con un tesoro tenemos que seleccionar los conceptos relevantes del documento de la manera más precisa posible, ya que la estructura semántica del tesoro garantiza la coherencia del significado de todos los descriptores. Esta circunstancia influirá de forma definitiva en un eficaz proceso de recuperación de la información.

En este último proceso, esto es, en el momento de proceder a la consulta y la recuperación pertinente de la información, el tesoro dirige al usuario a la hora de interrogar a la base datos o sistema documental concreto puesto que traduce el lenguaje natural del usuario y le dirige de los conceptos o descriptores no aceptados a los descriptores aptos para proceder a una interrogación efectiva del sistema.

La finalidad de su estructura (relaciones de pertenencia y jerarquía) y de su red de relaciones (relaciones de equivalencia y asociación) permite ayudar al usuario a construir la mejor estrategia posible en función de sus necesidades de información, lo que facilita una alta precisión y exhaustividad en la recuperación.

4.2.2. Revisión y actualización del tesoro

Después del seguimiento del uso del tesoro y la detección de sus principales problemas, es preciso iniciar su revisión y actualización. Conviene que esta revisión la lleven a cabo la persona o las personas administradoras del tesoro, ya que se trata de la única forma de garantizar el mantenimiento de su coherencia.

La periodicidad de esta revisión depende de las características del centro y del tesoro; puede ser semestral, anual o bianual.

También es importante mantener una nota histórica en los descriptores que han sufrido modificaciones, que incluya la fecha de la introducción del cambio y el descriptor que se utilizaba antes de su incorporación o modificación. Parte del *software* de gestión de tesauros incluye entre sus prestaciones la posibilidad de mantener esta nota histórica o fichero LOG como una información interna para los administradores.

Durante este proceso de revisión se deben tomar decisiones para solucionar los problemas que pueda presentar el tesoro:

- a) Con relación a los descriptores que se han utilizado con mucha frecuencia, se pueden mantener tal cual o crear otros descriptores más específicos.
- b) Con relación a los descriptores que se han utilizado con muy poca frecuencia, se pueden mantener tal cual o suprimirlos del tesoro.

c) Con relación a los descriptores candidatos, se pueden tomar las siguientes opciones:

- Si sólo se han requerido en muy pocas ocasiones, no sería necesario incluirlos ya que podemos utilizar el descriptor más genérico. Otra solución es incluirlos como no descriptores.
- Si la petición de los descriptores es frecuente, podemos incluirlos, estableciendo previamente las correspondientes relaciones de pertenencia, jerarquía, equivalencia y asociación.

d) Otro problema adicional se deriva de la eliminación de descriptores que hacen referencia a términos obsoletos que ya no se usan. El problema de esta operación consiste en que si ya tenemos documentos indizados con estos descriptores, es necesario que los volvamos a indizar, con el fin de mantener la coherencia entre el tesoro y el sistema de consulta.

5. Sistemas automatizados de gestión de tesauros. Tesauros en línea

5.1. Sistemas automatizados de gestión de tesauros

Los sistemas automatizados de gestión de tesauros facilitan la gestión y edición de un tesoro de manera automatizada, permitiendo de esta forma reducir el tiempo y los costes a la hora de su elaboración.

Las prestaciones de dichos sistemas son variadas en función del *software* que se utilice, pero en líneas generales su utilización permite:

a) Incluir términos y registrar automáticamente las relaciones invertidas con otros términos del tesoro.

b) Editar diferentes presentaciones del tesoro (alfabéticas, sistemáticas, índices permutados) y seleccionar el tipo de relación que queremos visualizar en el listado (con términos relacionados, notas explicativas, términos genéricos, etc. o sin ellos).

c) Utilizarse directamente en la consulta. Si el software forma parte del sistema de gestión de la base de datos documental, puede permitir la selección de los términos desde el tesoro, la ejecución inmediata de la búsqueda y la recuperación de los documentos pertinentes.

d) Contar con prestaciones para los administradores del tesoro como las siguientes:

- Posibilidad de mantener un fichero histórico donde se reflejen las modificaciones y los cambios que se hacen en los diferentes descriptores.
- Facilidades a la hora de eliminar un descriptor del tesoro (cuando este desaparece, se eliminan todas las relaciones que tenía con el resto de los términos).
- Elaboración automática de estadísticas (de frecuencia de uso de los descriptores, del número de relaciones existentes en el tesoro, etc.).

Existen dos grandes tipos de sistemas de gestión de tesauros:

1) Los más simples están asociados a un sistema de gestión documental concreto. Aunque contienen las prestaciones básicas para la gestión de tesauros, presentan también carencias en cuanto a la identificación correcta de las relaciones entre los descriptores.

2) Los más complejos son programas específicos de gestión de tesauros. Incorporan funcionalidades diversas, como la de validar automáticamente los términos y las relaciones entre ellos, lo cual ahorra tiempo en la verificación de dichas relaciones y evita cometer errores.

Ejemplos de *software* de gestión de tesauros

Recomendamos consultar la bibliografía, pero por poner algún ejemplo podemos señalar los siguientes según usemos el idioma inglés o el español:

- **Multites** (en inglés): permite la conversión y generación de ficheros en lenguaje HTML y por ende la inclusión del tesoro en la red Internet.
- **TemaTres** (en español): es una aplicación web para la gestión de lenguajes documentales. Se encuentra orientada especialmente al desarrollo de tesauros jerárquicos, pero también puede utilizarse para desarrollar estructuras de navegación web, o como complemento de un gestor de contenidos en bibliotecas digitales o en una biblioteca tradicional.

5.2. Los tesauros en el mundo digital. Tesauros en línea

En los últimos tiempos, el ámbito de la representación y recuperación de información ha tenido que asumir el impacto de Internet y sus tecnologías asociadas. Las consecuencias de estos cambios tecnológicos 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 los requerimientos de los usuarios.

En este sentido se pueden destacar los siguientes hechos:

1) Incremento del número de recursos de información que requieren descripción, identificación y elementos de localización, es decir, un análisis formal y de contenido.

2) Con relación al análisis de contenido, se hace necesario un replanteamiento en cuanto a las herramientas de gestión y organización de información entre las cuales se encuentran los vocabularios controlados.

3) Es cada vez más importante proporcionar a los propios usuarios unas estructuras de conocimiento que les permitan un acceso rápido y amigable a la información. En este sentido, el tesoro, como vocabulario controlado, puede ejercer un papel protagonista: los usuarios requieren herramientas concep-

tuales y semánticas destinadas a una efectiva organización de la información digital, y los tesauros son una de las estructuras que pueden tener una participación activa en este campo.

Por otra parte, los tesauros se ven significativamente beneficiados por el entorno digital dado que favorece el desarrollo de los siguientes aspectos:

- Nuevas posibilidades en cuanto a las pautas de diseño, elaboración, gestión y uso de las herramientas utilizadas para la representación y recuperación de información.
- Desarrollo de una estructura hipertextual donde se establezcan hipervínculos entre todos los elementos estructurales (descriptor, no descriptor, notas de alcance, etc.), y también entre las distintas partes del tesoro.
- Simplificación de la estructura a través de la hipertextualidad, facilitando por tanto su manejo a partir de interfaces adecuadas.
- Reducción de los costes de actualización.
- Potencia el desarrollo de tesauros multilingües y multidisciplinarios fomentando la cooperación internacional, y, por lo tanto, impulsando la reutilización e interoperabilidad de recursos.
- Acceso universal a herramientas terminológicas que favorecen la recuperación de información.
- Integración del usuario en el proceso de creación, gestión y optimización de las herramientas conceptuales documentales.

Enlace de interés

Recomendamos visitar la página de Willpower, empresa de consultoría independiente especializada en todos los temas relacionados con la gestión de la información y que proporciona exhaustiva y abundante información relativa a la construcción, gestión, mantenimiento y utilización de tesauros, así como multitud de referencias a tesauros de todo el mundo, normativa, instituciones y gestión de tesauros.

6. Tesauros y ontologías

En relación con el proceso señalado en el apartado anterior, relativo a la creciente digitalización del ámbito de la representación y recuperación de información según Arano (2005), son varias las herramientas que han sido objeto de estudio, al igual que lo han sido los campos de conocimiento donde estas herramientas se han originado: la lingüística, la inteligencia artificial, la documentación, la ingeniería lingüística, etc.

En relación con nuestro objeto de estudio, esto es, la representación y recuperación de información, estas herramientas son, entre otras, las siguientes:

- Sistemas de clasificación.
- Listas de encabezamiento.
- Bases de datos léxicas.
- Bases de conocimiento.
- Taxonomías.
- Tesauros.
- Mapas conceptuales.
- Ontologías.
- Redes semánticas.

Dentro de este amplio espectro de herramientas de representación y recuperación de información, son los tesauros y las ontologías los que con mayor frecuencia son objeto de vinculación en la bibliografía aunque provengan de ámbitos disciplinarios distintos.

6.1. Los tesauros

Como ya hemos visto anteriormente, un tesoro es un tipo de lenguaje documental que representa la estructuración conceptual de un determinado campo del conocimiento y se convierte en una herramienta de control terminológico, puesto que la estructuración del tesoro se basa en conceptos.

Los conceptos son representados por términos seleccionados que evidencian ese control terminológico, ya que con dicho control se busca mejorar la precisión de la indización y recuperación de la información, ambas funciones básicas de los tesauros.

El uso de información conceptual y lingüística almacenada en otros tipos de recursos (por ejemplo, en una ontología) permite el enriquecimiento a nivel estructural de los tesauros, y posibilita el incremento de la amigabilidad de estas herramientas para su utilización por parte de un usuario final no especializado.

6.2. Las ontologías

Las ontologías han sido un ámbito de interés de la inteligencia artificial. Una ontología está englobada dentro de la definición de lo que es un lenguaje controlado, ya que proporciona una representación formal del conocimiento donde los conceptos, las relaciones y las restricciones conceptuales son explicitadas mediante formalismos en un determinado dominio.

Son muchas las definiciones existentes en el ámbito de la inteligencia artificial. La más conocida es la de Gruber (1993), quien define *ontología* como una descripción formal en el dominio de un discurso.

Arano (2005), a partir de las definiciones existentes, propone la siguiente.

“Representación explícita y formal de una conceptualización compartida que implica una perspectiva sobre una cierta realidad y que se constituye en la estructura conceptual de una base de conocimiento. Su objetivo final es compartir el conocimiento que representa.”

Las ontologías definen conceptos y relaciones de algún dominio, de forma compartida y consensuada, y esta conceptualización debe ser representada de una manera formal, legible y utilizable por los ordenadores.

Las ontologías constan de términos, sus definiciones y los axiomas que los relacionan con otros términos, que están organizados en una taxonomía. El fin último de una taxonomía es presentar un sistema de clasificación que agrupe todos los conceptos en diferentes niveles jerárquicos.

A continuación, vamos a analizar las características principales de las ontologías (Lozano, 2000; Gruber, 1993).

Las ontologías tienen los siguientes **componentes**:

a) Conceptos. Son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.

b) Relaciones. Representan la interacción y enlace entre los conceptos del dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, etc.

c) **Funciones.** Son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como categorizar-clase, asignar-fecha, etc.

d) **Instancias.** Se utilizan para representar objetos determinados de un concepto.

e) **Axiomas.** Son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: "Si A y B son de la clase C, entonces A no es subclase de B".

Las **características** más representativas de las ontologías son las siguientes:

- **Pueden existir ontologías múltiples.** El propósito de una ontología es hacer explícito algún punto de vista, por lo que a veces será preciso combinar dos o más ontologías.
- **Podemos identificar niveles de abstracción de las ontologías.** Estos niveles de generalización o abstracción nos dan una tipología de ontologías. La idea es caracterizar una red de ontologías usando multiplicidad y abstracción, y puesto que no podemos aspirar a tener una descripción completa del mundo, podemos pensar en una estrategia de construcción gradual de abajo arriba.
- **Multiplicidad de la representación.** Un concepto puede ser representado de muchas formas, por lo que pueden coexistir múltiples representaciones de un mismo concepto.
- **Mapeo de ontologías.** Establecer relaciones entre los elementos de una o más ontologías, para establecer conexiones, especializaciones, generalizaciones, etc.

En términos prácticos, el **desarrollo de una ontología** incluye:

- Definir *clases* en la ontología, esto es, los conceptos y sus relaciones (clases, subclases).
- Colocar las clases en una *jerarquía de taxonomías* (subclase-superclase).
- Definir atributos y describir los valores permitidos para esos atributos.
- Rellenar los valores de los atributos con ejemplos.

Algunos ejemplos de desarrollos de ontologías

Figura 3. Ejemplo de ontología (Mulholland, 1999)

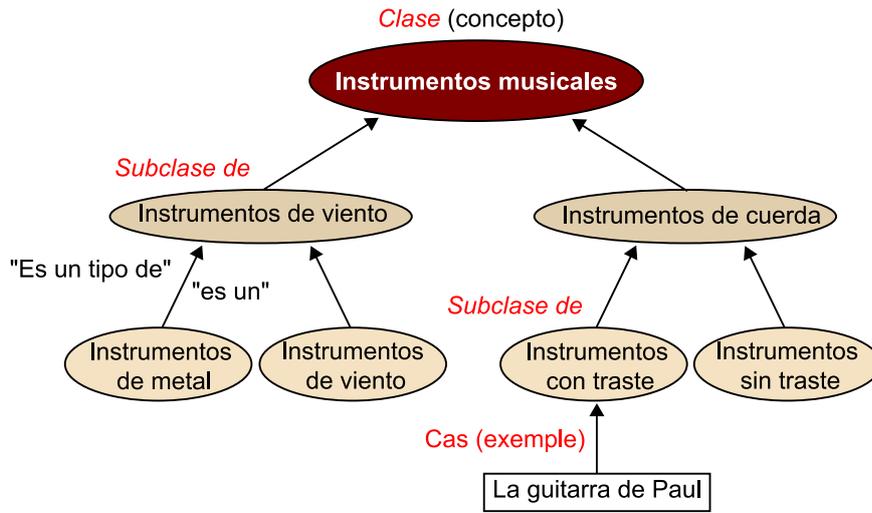
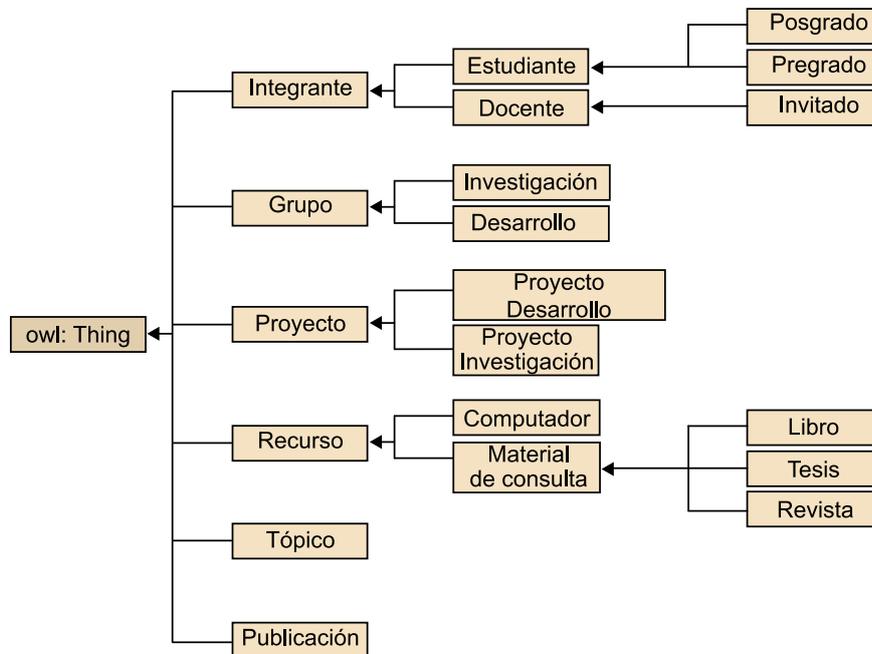


Figura 4. Ejemplo de ontología para laboratorios de investigación y desarrollo



Fuente: Hurtado y Segnera (2006)

6.3. Similitudes y diferencias entre tesauros y ontologías

Antonio García Jiménez (2004) menciona algunas de las diferencias y similitudes entre tesauros y ontologías. Éstas se relatan a continuación:

a) Similitudes

- Se engloban dentro de los lenguajes controlados, ya que utilizan el vocabulario de un dominio específico a partir de un número de términos básicos y de las relaciones que se establecen entre los mencionados términos.

- Están relacionados con la terminología empleada para representar los conceptos de un dominio específico.
- Ambos utilizan jerarquías para agrupar términos en categorías y subcategorías y pueden utilizarse para catalogar y organizar recursos de información.

b) Diferencias

- Si bien los tesauros están destinados a ser usados por humanos, las ontologías pueden ser usadas por humanos para compartir conocimiento y por agentes de software para procesarlo.
- En los tesauros, los descriptores se enlazan mediante tres tipos principales de relaciones muy simples, relaciones enfocadas principalmente para la recuperación pertinente de información. En cambio, las ontologías tienen una mayor variedad de relaciones entre conceptos. La descripción de los conceptos y sus relaciones pueden ser empleadas de un modo más general que en los actuales tesauros de manera que se pueden establecer un mayor número de relaciones diferentes, y tener definidas cada una de ellas.
- En los tesauros, la descripción de los conceptos es limitada y no formalizada, mientras que en las ontologías es explícita y formalizada.

Bibliografía

Bibliografía básica

Aitchison, J.; Gilchrist, A.; Bawden, D. (2000). *Thesaurus construction and use: a practical manual* (4.ª ed.). Chicago: Fitzroy Dearborn.

Arano, S. (2005). "Los tesauros y las ontologías en la Biblioteconomía y la Documentación". *Hipertext.net* (núm. 3, 2005). [Consultado en septiembre de 2009.]

Chaumier, J. (1986). *Análisis y lenguajes documentales: el tratamiento lingüístico de la información documental*. Barcelona: Mitre (Ciencias de la Comunicación).

Craven, T. C. (1997). *Introductory tutorial on thesaurus construction*. Londres (Ontario, Canada). University of Western Ontario (actualizado en 2008). [Consultado en septiembre de 2009.]

Currás, E. (1998). *Ontologías, taxonomía y tesauros: manual de construcción y uso*. Gijón: Ediciones Trea.

García Jiménez, A. (2004). "Instrumentos de representación del conocimiento: tesauros versus ontologías". *Anales de documentación* (núm. 7, 2004, pág. 79-95).

Gruber, T. R. (1993). "A Translation Approach to Portable Ontology Specifications". *Knowledge Acquisition* (5(2), pág. 199-220). [Consultado en septiembre de 2009.]

Hurtado, D; Sequeda, J. (2006). "Propuesta del Uso de Ontologías para la Búsqueda de Semántica en Laboratorios de Investigación y Desarrollo: OLID".

NISO Z39.19-2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*.

NISO Z39.19-2003. *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*.

Lancaster, F. W. (1995). "El control del vocabulario en la recuperación de la información". *Educación. Materiales* (núm. 12). Valencia: Universidad de Valencia.

Lozano Tello, A. (2000). "Ontologías en la Web semántica". I.ª Jornadas de Ingeniería Web'01. [Consultado en septiembre de 2009.]

Mulholland, P. (1999). *Introduction to Ontologies*. Knowledge Media Institute, The Open University (Walton Hall) Milton Keynes, UK.

Slype, G. van (1991). *Los lenguajes de indización. Concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Sánchez Ruipérez (Biblioteca del Libro).

Tomás López, M. (2004). "Diseño de un tesoro conceptual de folklore de la Región de Murcia". *Scire* (vol. 10, núm. 1, pág. 65-76).

UNE-50-106 (ISO 2788-1986). *Documentación: Directrices para el establecimiento y desarrollo de tesauros monolingües* (1990). Asociación Española de Normalización y Certificación (AENOR).

UNE-50-125 (ISO 5964-1985) *Documentación: Directrices para la creación y desarrollo de tesauros multilingües* (1997). Asociación Española de Normalización y Certificación (AENOR).

Willpower Information (1999). *Thesaurus principles and practice*. [Consultado en septiembre de 2009.]

Recopilación de enlaces a recursos de información sobre tesauros en línea

American Society of Indexers. *Thesauri online*. Wheat Ridge [Colorado]: American Society of Indexers. [Consultado en septiembre de 2009.]

Arano, S; Codina, L. (2004). "La estructura conceptual de los tesauros en el entorno digital: ¿nuevas esperanzas para viejos problemas?". 9.ª Jornadas Catalanas de información y Documentación. Barcelona 24-26 Nov 2004. [Consultado en septiembre de 2009.]

Centre for Digital Library Research. BUBL Link/5:15 Thesauri / BUBL Information Service. Glasgow: University of Strathclyde. [Consultado en septiembre de 2009.]

CINDOC. *Tesauros y Glosarios IEDCYT en línea*. Madrid: Centro de Información y Documentación Científica. [Consultado en septiembre de 2009.]

Koch, T. Controlled vocabularies, thesauri and classification systems available in the WWW. Lund: Lund University. [Consultado en septiembre de 2009.]

Library of Congress Thesauri. Washington: Library of Congress. [Consultado en septiembre de 2009.]

McKiernan, G. (2001). Controlled vocabularies: Beyond bookmarks: schemes for organizing the web. Ames: Iowa State University. [Consultado en septiembre de 2009.]

Middleton, M. "Thesaurus online sites. Controlled vocabularies: resource guide. Brisbane: School of Information Systems, QUT." [Consultado en septiembre de 2009.]

Recopilación de enlaces a recursos sobre software de gestión de tesauros

Jessica Milstead. "Thesaurus management software". Wheat Ridge: American Society of Indexers. [Consultado en septiembre de 2009.]

Michael Middleton. "Thesaurus management software: Controlled vocabularies: resource guide. Brisbane: School of Information Systems." [Consultado en septiembre de 2009.]

Willpower Information. "Software for building and editing thesauri". [Consultado en septiembre de 2009.]

Recopilación de enlaces a recursos de información generales sobre ontologías

a) Páginas web

"Ontologías" (Juan Carlos Cámara, 2002) . *Learning Metadata Standards*. [Consultado en septiembre de 2009.]

Ontology (Tom Gruber). [Consultado en septiembre de 2009.]

Semanticweb. [Consultado en septiembre de 2009.]

"Sites Relevant to Ontologies and Knowledge Sharing" (Richard Fikes). [Consultado en septiembre de 2009.]

Guía Breve de la Web Semántica (Consortio World Wide Web, W3C). [Consultado en septiembre de 2009.]

b) Artículos

Arano, Silvia (2005). "Los tesauros y las ontologías en la Biblioteconomía y la Documentación". *Hipertext.net* (núm. 3, 2005). [Consultado en septiembre de 2009.]

Arano, Silvia; Codina, Lluís (2004). "La estructura conceptual de los tesauros en el entorno digital: ¿nuevas esperanzas para viejos problemas?". 9.ª Jornadas Catalanas de Información y Documentación. Barcelona 24-26, nov. 2004. [Consultado en septiembre de 2009.]

Caldera-Serrano, J.; Sanchez-Jiménez, R. (2008). "Ontología para el control y recuperación de información onomástica en televisión". *El Profesional de la Información* (vol. 17, núm. 1, enero-diciembre, pág. 86-91)

García Jiménez, A. (2004). "Instrumentos de representación del conocimiento: tesauros versus ontologías". *Anales de documentación* (núm. 7, 2004, pág. 79-95).

Hurtado Bustamante, D. P.; Sequeda Sanclemente, J. F. (2006). "Propuesta del Uso de Ontologías para la Búsqueda Semántica en Laboratorios de Investigación y Desarrollo: OLID". [Consultado en septiembre de 2009.]

Legg, C. (2007). "Ontologies on the Semantic Web". *Annual Review of Information Science and Technology* (vol. 41, núm. 1, 2007).

Lozano Tello, A. (2000). "Ontologías en la Web semántica". I.ª Jornadas de Ingeniería Web'01. [Consultat a setembre de 2009.]

Pedraza Jiménez, R.; Codina, L.; Rovira, C. (2007). "Web semántica y ontologías en el procesamiento de la información documental". *El Profesional de la Información* (vol. 16, núm. 6, noviembre-diciembre, pág. 569-578).

Sanchez Jiménez, R.; Gil-Urdiciain, B. (2007). "Lenguajes Documentales y ontologías". *El Profesional de la Información* (vol. 16, núm. 6, noviembre-diciembre, pág. 551-560).

Reiss, M.; Moal, M.; Barnard, Y.; Ramu, J. P.; Froger, A. (2006). "Using Ontologies to Conceptualize the Aeronautical Domain". A: *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics* (pág. 56-63).

