

# LENGUAJES DOCUMENTALES. LOS TESAUROS: CREACIÓN Y MANTENIMIENTO. NORMA ISO 2788.

## I. INTRODUCCIÓN

Aunque los primeros lenguajes documentales (todos ellos de carácter enciclopédico) datan de finales del siglo XIX, el **concepto moderno de lenguaje documental** no se consolidó hasta el siglo XX, con la figura de **Cutter**. Hoy en día, contamos con numerosos ejemplos y tipos, como paso a exponer a continuación:

## II. LOS LENGUAJES DOCUMENTALES

En palabras de **Blanca Gil**, un lenguaje documental es «todo **sistema artificial de signos normalizados**, que facilitan la **representación formalizada** del contenido de los documentos para permitir la **recuperación**, manual o automática, de la información». Así entendido, el lenguaje documental interviene en dos etapas clave del proceso documental: la **descripción del contenido** y la **localización de la información** por parte de los usuarios.

### TIPOS

Existen diversos **criterios** para clasificar los lenguajes documentales, si bien los cuatro más extendidos son: el **control** sobre el vocabulario, la **coordinación** de los términos, la **estructura** organizativa y el grado de **especialización**.

- Atendiendo al **CONTROL**, pueden ser:
  - o **Libres**, aquellos que se van generando a raíz de indizaciones en lenguaje natural llevadas a cabo por los documentalistas durante los procesos de análisis del contenido. Son las **listas de palabras clave o de descriptores libres**.
  - o **Controlados**, aquellos que, previa realización del análisis, ya presentan un vocabulario prefijado con las materias, notaciones o descriptores utilizables de cara a la representación de contenidos. Son los lenguajes documentales mayoritarios y, entre ellos, se encuentran las **listas de encabezamientos de materia**, las **clasificaciones** y los **tesauros**.

- Según la **COORDINACIÓN**, se distingue entre:
  - o **Precoordinados**, aquellos cuyos componentes se combinan conforme a unas reglas sintácticas o semánticas establecidas de antemano por el propio lenguaje, que condicionan tanto las descripciones de contenido como la posterior recuperación de documentos. Son las **clasificaciones** y **listas de encabezamientos de materias**.
  - o **Postcoordinados**, aquellos que permiten combinar los términos en el momento de la búsqueda o recuperación de la información. Dentro de esta modalidad se encuadran las **listas de descriptores libres y de palabras clave** así como los **tesauros**.
  
- Según la **ESTRUCTURA**, se habla fundamentalmente de:
  - o **Jerárquicos**, caracterizados por una estructura arborescente en la que cada concepto depende de uno superior. Es el caso de la mayoría de **clasificaciones sistemáticas**, como por ejemplo la CDU.
  - o **De estructura combinatoria o asociativa**, cuyos términos se relacionan entre sí, permitiendo gran cantidad de combinaciones. A este esquema responden, por ejemplo, las **clasificaciones alfabéticas** y los **tesauros**, si bien es cierto que, en la organización sistemática de estos últimos, también participa la estructura jerárquica.
  
- Según su **ámbito de ALCANCE y especialización**, puede diferenciarse entre:
  - o **Índices generales**, como las **clasificaciones universales**, que intentan abarcar todo el conocimiento humano.
  - o **Índices especializados**, restringidos a una rama determinada del saber, como se observa en el caso de los **tesauros especializados**.

Independientemente de que existan diversas tipologías y matices, todos los lenguajes documentales deben compartir unos **MISMOS PRINCIPIOS BÁSICOS**, como son la **exhaustividad**, **simplicidad** organizativa, **entropía** o economía de signos, **no redundancia**...

De todos los lenguajes mencionados, el tema que nos ocupa demanda nuestra atención hacia los **TESAUROS** y hacia la **normativa UNE e ISO** por la que deben regirse.

### III. LOS TESAUROS

Según la norma **ISO 2788** de 1986, los tesauros pueden definirse desde dos puntos de vista:

Según su **FUNCIÓN**, son **instrumentos de control terminológico** para trasladar a un lenguaje más estricto (lenguaje documental) el lenguaje natural utilizado en los documentos.

Según su **ESTRUCTURA**, se trata de **vocabularios controlados y dinámicos**, aplicables a un área concreta del conocimiento y cuyos términos mantienen relaciones semánticas y genéricas entre sí.

Además, atendiendo a los criterios tipológicos anteriormente estudiados, un tesoro podría definirse como un lenguaje documental **controlado, postcoordinado, de estructura eminentemente combinatoria o asociativa y especializado**.

La razón de existir de los tesauros es la **INDIZACIÓN**, operación documental básica que, junto a la clasificación y el resumen, posibilita el análisis de contenido. En los centros de documentación, el lenguaje utilizado para indizar o, lo que es lo mismo, para describir brevemente el contenido intelectual de sus fondos, es el **tesauro**. Este mismo lenguaje es, a su vez, el que debería ser consultado por el usuario para desarrollar sus estrategias de búsqueda y recuperar la información satisfactoriamente.

#### FUNCIONES

Como todo lenguaje documental, sus **tres funciones básicas** son:

1. La **NORMALIZACIÓN del vocabulario**, mediante un léxico apropiado capaz de reflejar con precisión la esencia conceptual de los documentos. La **filosofía de todo tesoro** es huir de la riqueza del lenguaje natural y de los accidentes lingüísticos que le caracterizan: sinonimia, polisemia, homonimia, ambigüedad, etc.
2. La **INDUCCIÓN**, ya que el establecimiento de referencias cruzadas entre sus términos anima a los usuarios a consultar temas que, de otro modo, ni siquiera habrían considerado.
3. La **REPRESENTACIÓN** de los conceptos más significativos extraídos de los documentos analizados.

Así pues, el tesoro sirve de punto de encuentro entre el documento y el usuario y, en este proceso de mediación, el papel del documentalista es clave.

## ELEMENTOS

Las unidades léxicas más importantes de los tesauros son los **descriptores**, en oposición a los no descriptores:

Los **descriptores** o términos preferentes son aquellos escogidos de entre un conjunto de palabras equivalentes para representar, sin ambigüedades, una noción determinada.

Los términos no preferentes, es decir, los descartados de entre ese conjunto de palabras equivalentes reciben el nombre de **no descriptores**.

Desde el punto de vista sintáctico, un descriptor es **simple** cuando está formado por una sola palabra y **compuesto** cuando está formado por más de una palabra. Los descriptores compuestos se utilizan cuando es necesario recurrir a un sintagma nominal o preposicional para representar un concepto.

En cuanto a su tipología, la mayoría de descriptores son **temáticos**, aunque también puede haber algunos **onomásticos** (nombre de persona o institución), **geográficos** o **cronológicos**.

## ESTRUCTURA

Los tesauros pueden organizar sus términos, y las relaciones entre ellos, de diversas maneras. Lo normal es que coexistan varias de ellas dentro de un mismo tesauro, de ahí que muchos presenten distintos tipos de índices. Los **tres más frecuentes** son:

- **ÍNDICE ALFABÉTICO**. Es el apartado referencial más importante, ya que puede constituir un tesauro por sí solo. Dispone los **descriptores y no descriptores en orden alfabético**, especificando las relaciones entre ellos. Estas últimas son las que distinguen a un tesauro de un glosario o un léxico, por ejemplo, y pueden ser:
  - **Relaciones de equivalencia**: Controlan la sinonimia del lenguaje natural. Todos los términos considerados equivalentes y, por ende, no descriptores remiten al término válido o descriptor mediante la notación **USE**. El descriptor, a su vez, muestra cuáles son sus sinónimos o cuasisinónimos mediante la indicación **UP** (*Usado Por*).
  - **Relaciones de jerarquía**: Expresan el matiz de superioridad o inferioridad semántica entre los conceptos mediante las iniciales **TG** (*Término Genérico*) y **TE** (*Término Específico*).

- **Relaciones de asociación:** Como su propio nombre indica, asocian términos de forma recíproca cuando las materias representadas se encuentran ligadas conceptualmente de alguna manera. Esta conexión se expresa mediante la abreviatura **TR** (*Término Relacionado*).

Al margen de éstas, la notación **NA** significa **Nota de Alcance** y sirve para aclarar o precisar el uso de un término en un contexto concreto.

En algunos tesauros, estas abreviaturas pueden aparecer en **inglés**, por lo que conviene familiarizarse con ellas para manejar con soltura las diferentes herramientas a nuestro alcance (**BT** = *Broader Term*, **NT** = *Narrower Term*, **RT** = *Related Term*, **SN** = *Scope Note*...).

- **ÍNDICE SISTEMÁTICO.** Estructura los descriptores en función de las jerarquías establecidas en el tesoro. Despliega las **familias de términos**, situando los descriptores junto a sus conceptos más próximos. El término con mayor carga semántica dentro de cada familia se conoce como **macrodescriptor**, aunque no puede realizar las funciones de descriptor. Cada uno de estos macrodescriptores y su descendencia conforman un **microtesoro**.
- **ÍNDICE PERMUTADO.** Ordena los descriptores alfabéticamente para facilitar su búsqueda, con la particularidad de que aquellos formados por más de un término figuran en el índice tantas veces como elementos los componen.

Los tesauros suelen incluir, además, una **INTRODUCCIÓN** para explicar su objetivo, estructura, autoría, campo semántico e idiomas abarcados, instrucciones de uso, previsiones de actualización, etc.

## **CREACIÓN**

Para la elaboración de un tesoro, lo primero que ha de tenerse en cuenta es la **NORMATIVA** existente al respecto. En España, las directrices para la construcción de tesauros se hallan en dos normas: la mencionada **ISO 2788** y su versión española **UNE 50-106:1990**.

Una vez consultada la normativa, hay que comprobar la existencia o no de otros tesauros previos sobre el mismo campo, para valorar su utilidad y su posible adaptación a las nuevas necesidades. Si estuvieran en otro idioma, habría, además, que traducirlos.

Pues bien, para adaptar o confeccionar un tesoro completamente nuevo, es necesario seguir una serie de **PASOS**, que podrían sintetizarse en:

1. **Delimitar bien el CAMPO TEMÁTICO**, ya que, aunque cada centro de documentación esté especializado en un área concreta, siempre existen ramificaciones y conexiones con otras materias relacionadas descartables. Son los llamados **temas marginales**. En esta fase, habrá que plantearse el grado de precisión con el que se desea trabajar.

2. **Recopilar los TÉRMINOS**, para lo que se aconseja elaborar una lista con las palabras más utilizadas y significativas en ese campo del saber, anotando su frecuencia de aparición en los documentos analizados. Pueden extraerse de muy distintas fuentes, entre ellas diccionarios, léxicos especializados, bases de datos, libros, estudios sobre vocablos utilizados por los usuarios para realizar sus consultas, etc. Los términos más usados y actuales se convertirán en los descriptores, que, conforme a la **norma ISO 2788**: a) deberán utilizarse preferiblemente en su **forma sustantiva**, con la ortografía y sintaxis admitida; b) se diferenciarán del resto de términos mediante **tipografía negrita o mayúscula**; c) irán en **plural** si son contables y en **singular** si son incontables; y d) **no** incluirán **abreviaturas, siglas o signos de puntuación**.

3. **Establecer las RELACIONES entre los términos**, que, como se ha detallado, podrán ser de equivalencia, jerarquía y asociación. En la práctica, esta tarea y la anterior podrán simultanearse.

4. **Comprobar el FUNCIONAMIENTO (test del tesoro)**, para evaluar la eficacia del vocabulario escogido, ya sea mediante situaciones reales o ficticias. Lo ideal es ir haciendo pruebas durante el propio proceso de elaboración, para detectar posibles errores a tiempo y comprobar que la estructura elegida responde a las necesidades planteadas.

En todo caso, los cambios y ajustes realizados durante la fase de la elaboración no serán los últimos que experimentará el tesoro, ya que éste **deberá ser revisado y actualizado a lo largo de toda su vida útil**, siempre que se estime conveniente y sin periodicidad fija.

## **MANTENIMIENTO**

Según la normativa ISO, **la elaboración de un tesoro no termina jamás**, ya que éste es una herramienta viva, dinámica y en constante evolución. Su uso diario es el que irá dando las pautas para futuras modificaciones y ampliaciones.

Existen muchas razones para revisar un tesoro, entre ellas la **evolución misma del lenguaje y de la ciencia**. En este sentido, habrá que prestar mucha atención a la aparición de nuevos términos así como a la inevitable caída en desuso de algunos inicialmente incluidos. Al mismo tiempo, habrá que detectar relaciones mal estructuradas o fallos conceptuales, que, de no ser corregidos, se seguirán arrastrando en las indizaciones de los documentos.

El **añadir o suprimir términos del tesoro** implica una tarea cuidadosa y reflexiva, por lo que el mantenimiento del tesoro, al igual que su elaboración previa, siempre deberá dejarse en manos de **expertos**, preferentemente de dos tipos: **especialistas en la materia central del tesoro**, por un lado, y **documentalistas**, por otro.

Por último, conviene recordar que las modificaciones introducidas también afectarán, de un modo u otro, a los documentos indizados con anterioridad. Por eso, será aconsejable realizar búsquedas retrospectivas para **armonizar los términos descriptivos y de consulta**. Algunas de estas tareas ya pueden realizarse automáticamente.

#### IV. CONCLUSIÓN

En definitiva, los tesoros son los **instrumentos de control terminológico por excelencia en los centros de documentación**. Aunque tradicionalmente se han creado y manejado en **soporte papel**, cada vez resulta más habitual trabajar con ellos en **formato electrónico** e incluso **en línea**. Y es que, en la actualidad, la inestimable ayuda de la informática ha simplificado mucho el trabajo de construcción, actualización y uso de tesoros.

## I. INTRODUCCIÓN

## II. LOS LENGUAJES DOCUMENTALES

### TIPOS

- Atendiendo al **CONTROL** sobre el vocabulario:

- o Libres
- o Controlados

- Atendiendo a la **COORDINACIÓN** de los términos:

- o Precoordinados
- o Postcoordinados

- Atendiendo a la **ESTRUCTURA** organizativa:

- o De estructura jerárquica
- o De estructura combinatoria o asociativa

- Atendiendo al **GRADO DE ESPECIALIZACIÓN**:

- o Índices generales
- o Índices especializados

## III. LOS TESAUROS

### FUNCIONES

1. La **normalización del vocabulario**
2. La **inducción**
3. La **representación** de los conceptos más significativos.

### ELEMENTOS

Descriptores o términos preferentes

Términos no preferentes o no descriptores

### ESTRUCTURA

#### 1. **ÍNDICE ALFABÉTICO**

- o Relaciones de equivalencia
- o Relaciones de jerarquía
- o Relaciones de asociación

#### 2. **ÍNDICE SISTEMÁTICO**

#### 3. **ÍNDICE PERMUTADO**

### CREACIÓN

1. **Delimitar bien el campo temático**
2. **Recopilar los términos**
3. **Establecer las relaciones entre los términos**
4. **Comprobar el funcionamiento**

### MANTENIMIENTO

## IV. CONCLUSIÓN



## LENGUAJES DOCUMENTALES. LOS TESAUROS: CREACIÓN Y MANTENIMIENTO. NORMA ISO 2788.

### I. INTRODUCCIÓN

- 1<sup>os</sup> lenguajes documentales → f. s. XIX, pero **CONSOLIDACIÓN concepto moderno** s. XX con Charles **Cutter**.

### II. LOS LENGUAJES DOCUMENTALES

- Def. **Blanca Gil Urdiciain** («todo **sistema artificial de signos normalizados**»).

#### TIPOS

1. Según el **CONTROL del vocabulario**, pueden ser:

- a) **Libres**. Listas de palabras clave y de descriptores libres.
- b) **Controlados**. Clasificaciones, LEM y tesauros.

2. Según la **COORDINACIÓN de los términos**, se distingue entre:

- a) **Precoordinados**. Clasificaciones y LEM.
- b) **Postcoordinados**. Listas de palabras clave y de descriptores libres, tesauros.

3. Según la **ESTRUCTURA organizativa**, se habla de:

- a) **Jerárquicos**. Clasificaciones sistemáticas.
- b) **De estructura combinatoria y asociativa**. Clasificaciones alfabéticas y tesauros.

4. Según su **ESPECIALIZACIÓN o alcance**:

- a) **Índices generales**. Clasificaciones universales como la CDU.
- b) **Índices especializados**. Tesauros.

- **MISMOS PRINCIPIOS BÁSICOS**: exhaustividad, pertinencia, simplicidad organizativa, entropía, no redundancia...
- Atención del tema hacia tesauros y normativa UNE e ISO por la que se rigen.

### III. LOS TESAUROS

- De conformidad con la **norma ISO 2788** de 1986, def. desde 2 puntos de vista:
  - Según su **FUNCIÓN** → instrumentos de control terminológico (lenguaje natural a lenguaje más estricto).
  - Según su **ESTRUCTURA** → vocabularios controlados y dinámicos (aplicables a un área concreta, relaciones).
  - Def. atendiendo a criterios tipológicos estudiados (controlado – postcoordinado – estructura combinatoria – especializado).
- Razón de existir de los tesauros → **INDIZACIÓN**.

#### FUNCIONES

1. La **normalización del vocabulario** (filosofía del tesoro).
2. La **inducción**.
3. La **representación** de los conceptos más significativos extraídos de los documentos analizados.

- Tesoro → punte entre documento y usuario (papel clave del documentalista).

## ELEMENTOS

- Unidades léxicas más importantes del tesoro → DESCRIPTORES, en oposición a los no descriptores:
- **Descriptores** o términos preferentes.
- Términos no preferentes = **no descriptores**.
- Desde el punto de vista sintáctico → descriptores **simples** y **compuestos**.
- En cuanto a su tipología → **temáticos**, onomásticos, geográficos y cronológicos.

## ESTRUCTURA

1. **ÍNDICE ALFABÉTICO** (tesoro por sí solo, orden alfabético de descriptores y no descriptores). Relaciones:

- a) **Relaciones de equivalencia** (controlan la sinonimia). **USE** y **UP**.
- b) **Relaciones de jerarquía** (matiz de superioridad o inferioridad semántica). **TG** y **TE**.
- c) **Relaciones de asociación** (reciprocidad). **TR**.

\* **NA** / Abreviaturas en inglés.

2. **ÍNDICE SISTEMÁTICO** (despliega familias de términos). **Macrodescriptor** → microtesoros.

3. **ÍNDICE PERMUTADO** (descriptores en orden alfabético, cada elemento de los descriptores compuestos).

\***INTRODUCCIÓN** con objetivo, autoría, campo semántico, idiomas, instrucciones de uso, previsión de actualizaciones...

## CREACIÓN

- **NORMATIVA** existente al respecto (España → norma **ISO 2788** de 1986 y **UNE 50-106:1990**).
- Existencia o no de algún tesoro previo sobre el mismo campo → posible adaptación o traducción.
- PASOS:
  1. **Delimitar el campo temático**. Temas marginales, grado de precisión.
  2. **Recopilar los términos**. Términos más usados y actuales = descriptores (características conforme a **norma ISO 2788**).
  3. **Establecer las relaciones entre los términos**, de equivalencia, jerarquía o asociación.
  4. **Comprobar el funcionamiento** (test del tesoro).
- **Debe ser revisado y actualizado durante toda su vida útil.**

## MANTENIMIENTO

- ISO → el **proceso de creación de un tesoro no termina jamás** (las pautas de modificación las va dando el día a día).
- **Evolución de la ciencia y del propio lenguaje**.
- Añadir y suprimir términos = tarea cuidadosa y reflexiva → debe dejarse en manos de **EXPERTOS**.
- Repercusión de las modificaciones en los documentos indizados previamente → búsquedas retrospectivas para **armonización de los términos**.

## **IV. CONCLUSIÓN**

- Tesoros = instrumentos de control terminológico por excelencia en centro de doc. / Nuevos soportes / Ayuda informática.