

Temario de Ayudante de Biblioteca de la Administración General del Estado

Este temario ha sido elaborado por un opositor, para presentarse al proceso selectivo de Ayudante de Bibliotecas de la Administración General del Estado en la [convocatoria de 2021](#).

Incluye todos los temas, de legislación y específicos de bibliotecas, del programa correspondiente a la convocatoria de la Administración General del Estado para cubrir plazas de Ayudante de Bibliotecas en el Ministerios de Cultura y Deporte, Ministerio de Defensa, Ministerio de Asuntos Exteriores, Unión Europea y Cooperación y Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática. «BOE» núm. 149, de 23 de junio de 2021.

Temario completo disponible en:

<https://www.bibliopos.es/>



Temario de Ayudante de Biblioteca de la Administración General del Estado, cedido por su autor a [Bibliopos.es](https://www.bibliopos.es) para su publicación bajo licencia [Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional License](#).

Bajo esta licencia puedes utilizar libremente el temario para uso personal y compartirlo siempre que [cites la fuente](#) y proporciones un enlace a la [licencia](#). No puedes hacer uso comercial del documento.

B07 Lenguajes documentales

Los tesauros: creación y mantenimiento.

Lenguajes documentales

El objetivo fundamental de todo centro de documentación o biblioteca es facilitar la recuperación de documentos o de la información contenida en ellos. Tradicionalmente, dicha recuperación puede llevarse a cabo por diferentes vías: a partir del nombre del autor, del título o del tema que trate la obra. El procedimiento seguido para conseguir la recuperación con eficacia en base al contenido conceptual del documento se denomina **indización**, que consiste en el análisis y selección de aquellas nociones que representan con mayor fidelidad la información que contiene y su traducción a un lenguaje documental. Por tanto, la principal característica de los lenguajes documentales es el control terminológico, que permite la representación de los conceptos derivados de la indización sin ambigüedades, garantizando el empleo de los mismos términos para expresar los mismos conceptos.

Según Blanca Gil Urdiciain, puede considerarse al **lenguaje documental** como “todo sistema artificial de signos normalizados, que facilitan la representación formalizada del contenido de los documentos para permitir la recuperación manual o automática, de información solicitada por los usuarios”. El lenguaje documental es un lenguaje no natural, aunque utiliza los signos de este. Este tipo de lenguajes se componen de una lista de términos, que le dan configuración de vocabulario, y una serie de relaciones entre ellos, que dinamizan esos vocabularios. Estos signos adquieren valor semántico por medio de su normalización y de las reglas morfosintácticas que lo articulan. Son precisamente esas reglas las que le otorgan la categoría de lenguaje. El lenguaje documental se sirve de las palabras, que representan nuestro conocimiento de la realidad, y en ocasiones, las reemplaza por símbolos cargados de significado preciso, de forma que lleva a cabo la representación del documento en virtud de una correspondencia analógica.

Los lenguajes documentales son lenguajes normalizadores y al mismo tiempo normalizados, ya que influyen a la vez que se crean en el momento de la descripción (completa el proceso técnico de catalogación dotando al soporte de la descripción de puntos de acceso temáticos) y en la recuperación de información. En la fase de descripción, tras la lectura e identificación de los conceptos contenidos en los documentos, se representan mediante un lenguaje documental, que permite traducir los conceptos en términos normalizados. Sin embargo, su objetivo primordial se cumple, en la segunda fase, suministrando el vocabulario de los conceptos de la demanda: el lenguaje documental proporciona una base para efectuar búsquedas al nivel de generalidad o especificidad deseados. En este sentido, se le puede considerar un lenguaje intermedio o metalenguaje en la medida en que sirve de puente entre las informaciones contenidas en los documentos y las informaciones solicitadas por los usuarios.

Las primeras manifestaciones datan de finales del siglo XIX, cuando aparecen las clasificaciones bibliográficas, basadas en la lógica y en los sistemas filosóficos del conocimiento y fundadas en los principios de precoordinación y de carácter enciclopédico. El concepto moderno de lenguaje documental se consolidó en el siglo XX a raíz de *Rules for a printed dictionary catalog* (1876) de Charles Ammi Cutter. Los lenguajes especializados nacieron como respuesta a la dispersión del conocimiento, que obligó a una creciente especialización de los contenidos.

El lenguaje documental no deja lugar para la reiteración de términos, su eficacia aumenta a medida que la redundancia es eliminada. Por todo ello, este lenguaje debe regirse por el **principio de entropía** que, como concepto opuesto a la redundancia, se corresponde con la mayor cantidad de información aportada por un solo signo o mensaje. Precisamente el carácter entrópico del lenguaje documental es lo que le diferencia del lenguaje natural, en donde es posible accidentes lingüísticos (tanto la homonimia como la sinonimia). Derivado de estos accidentes lingüísticos surge la

ambigüedad, que crea dificultades estrechamente relacionadas con la aplicación de un lenguaje artificial a la recuperación. Se trata del silencio y ruido documentales, que se oponen a la relevancia de la indización: en el primer caso, no se recuperan documentos que contienen la información deseada; mientras que en el segundo caso, se recuperan documentos sin haber sido solicitados (o en el mejor de los casos, responden sólo parcialmente al perfil de búsqueda).

Existen diversos criterios de tipificación de los lenguajes documentales, los más generalizados son, según Gil Urdiciain:

1. Control del vocabulario

- a) *Lenguajes libres*. Los términos proceden del texto original, formados por los mismos términos en los que está escrito el texto. Se componen de un vocabulario no predefinido que se va generando a partir de la realización de procesos de indización. Su ventaja radica en que son términos muy específicos y actuales (permite seguir de cerca la evolución terminológica), así como la rapidez de la elaboración y su bajo coste, pero tienen el inconveniente de la ambigüedad semántica (homonimia y polisemia).
- b) *Lenguajes controlados*. Mediante un análisis previo de los documentos, se obtiene un vocabulario, que pasan por un proceso realizado por documentalistas con el fin de evitar la ambigüedad existente en el lenguaje natural. Se logra, así, una representación unívoca y normalizada, y un lenguaje artificial, creados para la indización y la recuperación de la información.

2. Grado de coordinación: en función del momento en que se combinan los elementos que los componen.

- a) *Lenguajes precoordinados*. Coordinan los diferentes conceptos de una materia cuando se elabora el lenguaje o en el momento de la descripción, antes de la recuperación. Los términos están unidos de acuerdo con unas reglas sintácticas establecidas por el propio lenguaje. Con ellos se obtiene una mayor precisión en la búsqueda.
- b) *Lenguajes postcoordinados*. Permiten la coordinación de conceptos en el momento de la recuperación. Se coordinan los conceptos con posterioridad al momento en que se han creado los registros bibliográficos y sus respectivos índices. Carecen de sintaxis; salvo excepciones, su vocabulario consiste en términos simples o unitérminos que se pueden conectar en el momento de la recuperación por medio de una sintaxis booleana.

3. Criterio estructural. Es el más utilizado al clasificar los lenguajes documentales.

- a) *Lenguajes de estructura combinatoria (o asociativa)*. Los términos (descriptores) se combinan sin que estén organizados en una estructura determinada de antemano. Incluye a los léxicos documentales y a los tesauros.
- b) *Lenguajes de estructura jerárquica*. Son las clasificaciones enciclopédicas, facetadas y especializadas, que organizan sistemáticamente el conocimiento, es decir, agrupan las materias en áreas categoriales con una estructura arborescente: un orden previsto en una serie de clases, que integran los conceptos más específicos dentro de los más generales.

No obstante, ningún lenguaje documental tiene valor de forma aislada, así los conceptos de coordinación y control pueden combinarse, dando lugar, según Courier, a la siguiente tipología:

- *Lenguajes documentales postcoordinado con vocabulario libre*: son las listas de palabras clave (en inglés *keyword*) y listado de descriptores libres. Las **palabras clave** son las palabras o grupo de palabras seleccionadas de manera automática del título, resumen o texto de un documento que representan el contenido y permite la recuperación. Se componen en listas con ordenación alfabética, entre las que se encuentran los índices permutados, ideados por Luhn, que se componen de palabras clave presentes en el título o texto de los documentos, desde las que puede recuperarse la información. Los dos tipos más representativos son: los índices KWIC (*keyword in context*), donde las palabras significativas quedan alfabetizadas dentro del texto y los índices KWOC (*keyword out of context*), donde la palabra que alfabetiza se encuentra fuera. Los **listados de descriptores libres** son un listado ordenado alfabéticamente de conceptos destacados, mediante un

proceso intelectual, a partir de los documentos registrados dentro de un sistema documental determinado. Estos conceptos se expresan mediante palabras o expresiones extraídas de los documentos, o bien los proponen los documentalistas. Estos listados no limitan la incorporación de nuevos conceptos.

- *Lenguajes documentales postcoordinado con vocabulario controlado*: son los **listados de autoridades** y los **tesauros**. Las listas de autoridades son listados de las formas normalizadas de nombres propios y títulos de las entradas de un catálogo bibliográfico.
- *Lenguajes documentales precoordinado con vocabulario libre*: el **sistema Precis** (Preserved Context Indexing System). Es un lenguaje documental de estructura sintáctica e indización articulada, que posibilitaba una búsqueda no secuencial (que permitía recuperar por cualquiera de los conceptos que figuraban en la cadena del encabezamiento).
- *Lenguajes documentales precoordinado con vocabulario controlado*: los sistemas de clasificación (jerárquicas y facetadas) y los encabezamientos de materia. Las **clasificaciones jerárquicas**, también denominadas analítico-enumerativas, poseen una serie de clases y subclases que se organizan en forma arborescente, dando lugar una serie de divisiones y subdivisiones dependientes unas de otras. Las **clasificaciones facetadas** permiten construir áreas de conocimiento mediante un sistema analítico-sintético, lo cual significa que primero se analizan las distintas facetas relativas al documento (considera cada término clasificable desde determinados puntos de vista, lo que posibilita la agrupación por afinidades) y después, mediante síntesis, se llega a materializar su correspondiente término de clasificación. Los **encabezamientos de materia** tienen su origen en la obra de Cutter sobre las reglas para un catálogo diccionario. Se basan en dos principios: el de la especificidad (un término indizado sólo puede describir una sola materia, es decir, los conceptos deben utilizarse bajo su nombre más específico y no deben englobarse en uno más general); y el de entrada directa (los encabezamientos compuestos se presentarán como en el lenguaje natural, sin invertir los términos). Se componen de encabezamientos propiamente dichos y de subencabezamientos. Los primeros, una o varias palabras, representan los conceptos. Los segundos se utilizan, a veces, para precisar el significado de los primeros (nunca constituyen una entrada principal) y pueden ser: topográficos, cronológicos, de forma (tipo de documento concreto: diccionario, etc.), de materia o punto de vista (historia, filosofía, etc.) bajo el que se estudia la materia principal.

Los tesauros: creación y mantenimiento

Un **tesauro** es un lenguaje documental de estructura combinatoria, de carácter especializado (en torno a un tema, área concreta del conocimiento o bien se realizan para servir al desarrollo temático o profesional de una institución determinada). Se trata de un vocabulario controlado (instrumento de control terminológico) y estructurado al que se llega mediante la selección de términos del lenguaje natural empleado en los documentos y por los usuarios. Está formado por listas de palabras (descriptores) estructuradas de forma que unas se relacionan semántica y jerárquicamente con otras que también están en el tesauro.

Etimológicamente, proviene del término griego *thésauros*, que significa *tesoro*; por lo que, figurativamente, hace referencia a acumular bienes intelectuales. A menudo, se empleó para referirse a los diccionarios especializados en un área de conocimiento.

Los tesauros permiten cumplir tres **funciones** fundamentales del proceso documental: normalización (unificar todo el vocabulario para controlar todas las posibles entradas), inducción (debe indicar constantemente todas las alternativas posibles para entrar a recuperar, remitiendo al usuario hacia la consulta de temas que no había pensado) y representación (correspondencia entre los términos que componen sus índices y los conceptos presentes en los documentos).

La primera norma sobre tesauros fue la *ANSI Z39.19-1974 Thesaurus structure, construction and use*, que fue publicada por el American National Standards Institute (ANSI) en 1974 y revisada

en 1980. La norma *ISO 2788:1986 para tesauros monolingües* fue la norma a nivel internacional que se mantuvo durante 25 años, hasta la aparición de la *ISO 25964:2011*, aplicable también a vocabularios utilizados comúnmente para la recuperación de información. Consta de dos partes: la *ISO 25964-1:2011 Parte 1: Tesauros para la recuperación de la información (UNE-ISO 25964-1:2014)* y la *ISO 25964-2:2013 Parte 2: Interoperabilidad con otros vocabularios (UNE-ISO 25964-2:2016)*. Surge como revisión de dichas normas, que ya no respondían a las prácticas actuales de procesos de indización y recuperación de información, ni cubrían la demanda de aplicaciones de red que quieren utilizar recursos heterogéneos.

Un tesauro se compone de una base léxica estructurada en relaciones jerárquicas, asociativas y de equivalencia. Un buen tesauro debe dejar muy claro qué significado cubre un término, mediante los términos no válidos y las relaciones semánticas. Tienen dos tipos de **términos**:

- **Descriptor.** Palabra o grupo de palabras incluidas en un tesauro y escogidas de entre un conjunto de términos equivalentes para representar sin ambigüedad una noción contenida en un documento o en una petición de búsqueda documental. Pueden ser temáticos, onomásticos, geográficos o cronológicos.
- **No descriptor (término equivalente o no preferente).** Son sinónimos o cuasisinónimos de los descriptores. No pueden ser utilizados para indizar documentos ni para formular consultas, pero cada uno de ellos reenvía a uno o más descriptores, por lo que aumenta la tasa de precisión y pertinencia.

Los términos que componen el tesauro y sus relaciones se pueden representar de varias **formas de presentación**, siendo las más comunes: la presentación alfabética (los descriptores y no descriptores se encuentran agrupados en una sola secuencia alfabética, acompañados de sus relaciones semánticas), la presentación sistemática (contiene a los descriptores estructurados en función de jerarquías o categorías, en base a campos temáticos o facetas) y la presentación gráfica (exponen la estructura semántica de cada campo mediante una estructura arborescente, diagramas de flechas o terminogramas). Para que un tesauro constituya una precisa herramienta de recuperación de información, es menester que en su estructura estén presentes, al menos dos de las formas de presentación: un índice alfabético, que permita una primera toma de contacto con la información, que remita mediante un código a la parte sistemática o gráfica. Además, existe la posibilidad de incluir **índices auxiliares** que faciliten el acceso a la información contenida en los índices. Al menos es aconsejable elaborar un índice permutado general en donde estarán todos los descriptores alfabetizados tantas veces como términos los compongan.

Su **elaboración** supone el trabajo continuo de un grupo de especialistas en la materia durante un periodo de tiempo bastante amplio, que dependerá del volumen del tesauro, de los campos a cubrir y de la profundidad de la indización. Además, en la actualidad existen ayudas automáticas (como Beat, TCS o MultiTes) para la construcción y mantenimiento de tesauros, especialmente en el proceso de establecimiento de relaciones asociativas, jerárquicas y preferenciales entre los términos. Antes de construir un tesauro conviene tener en cuenta algunos parámetros que delimiten la función que debe cumplir o su nivel de especificidad: conocimiento del tipo de usuario y su nivel de cualificación, volumen y naturaleza de la documentación a tratar, planteamiento de la estructura de la terminología, etc. Analizadas las cuestiones previas, se pasaría a la elaboración del tesauro, realizado en las siguientes etapas:

1. Delimitación de la profundidad y extensión del campo terminológico
2. Selección de la terminología. La compilación de los términos puede hacerse en base al método analítico, al sintético o combinando ambos procedimientos. En ambos casos, la selección se hace por estudios de estadísticas de frecuencia: número de veces que aparece el término en la literatura que se maneja y frecuencia de uso. Posteriormente se lleva a cabo siempre una normalización de los términos, aunque sea mínima: los descriptores deben expresarse en mayúsculas y los no descriptores en minúsculas, con preferencia por la forma singular, masculina, forma sustantiva y escritura directa en locuciones, es preferible utilizar la forma desarrollada a las siglas o acrónimos (salvo si se han sustantivado o se conocen en

- cualquier idioma).
3. Determinación de campos temáticos y distribución de descriptores. Es la parte más importante y más difícil de realizar. Se trata de crear la única estructura jerárquica del tesoro, que presenta sistemática y sintéticamente toda la información que hay en el mismo. Consiste en dividir en áreas temáticas de afinidad, constituyendo la estructura básica en la que se distribuirán los descriptores del fichero alfabético. Cada campo tiene un código numérico y todos los componentes del campo llevarán la misma codificación, que debe aparecer en el índice.
 4. Redacción de notas aclaratorias o de aplicación. Se busca contextualizar cada término definiendo con precisión su significado artificial. Se introducen por el operador *NA* y ejercen una función de normalización pues restringen todos los sentidos posibles de un descriptor a uno solo, o para desarrollar siglas.
 5. Establecimiento de las relaciones semánticas preferenciales o de equivalencia entre los descriptores y los no descriptores. Es la agrupación de términos preferenciales a través de la localización de sinónimos en la terminología, que no deben aparecer simultáneamente como descriptores autorizados. Detectadas las equivalencias semánticas, se elige uno de los términos como representativo del concepto, dándole la categoría de término preferente o descriptor. Los operadores más habituales para representar esto son los siguientes: *UP* (Úsase por) y *USE*. Un término no descriptor se conecta normalmente al descriptor correspondiente mediante la referencia *USE*. La referencia correspondiente en la dirección opuesta es *UP*. También se emplean los signos de igual (=) o una flecha (→).
 6. Establecimiento de relaciones jerárquicas. Permiten reunir aquellos descriptores que, unos respecto a otros, son más generales o más específicos en niveles superiores o inferiores, formando así un conjunto jerárquico preciso. Puede ser de amplitud y especificidad (género/especie) y partitivas (parte-todo). Las relaciones jerárquicas se transcriben por medio de los operadores *TG* (Término Genérico) y *TE* (Término Específico), o bien los signos de menor (<) y mayor (>).
 7. Establecimiento de relaciones asociativas. Se trata de una relación de afinidad semántica (por asociación de ideas). El operador que lo representa es *TR* (Términos Relacionado), o bien empleando una raya (-).
 8. Redacción de una introducción, en la que se explique cuál es la intencionalidad, profundidad, extensión, estructura, fuentes, los códigos y su explicación, etc.
 9. Evaluación del tesoro. Se indiza una muestra de documentos para conocer el grado de coherencia de la estructura y su capacidad de respuesta real. Todo ello implicará modificaciones en la estructura jerárquica o la incorporación o eliminación de términos.

Periódicamente (cada seis meses en uno de nueva creación, o cada dos o tres años en aquellos que se han utilizado durante determinado tiempo) se procede a su **revisión**. Por tanto, es un proceso continuo que permite conocer el uso que se hace de la terminología tanto en la operación de indización como en la de recuperación de información (las consultas realizadas en lenguaje natural por los usuarios aportan la terminología real del centro). Se tendrá en cuenta el grado de evolución del lenguaje, la opinión de los especialistas en la materia, las consultas de los usuarios y la utilización de los descriptores en el proceso de indización. La actualización ha de hacerse para cubrir lagunas y fallos detectados en la práctica de su aplicación: inclusión de nuevos términos, eliminación de un descriptor, subdivisión de un descriptor presente en el tesoro en una serie de descriptores de menor nivel jerárquico, inclusión de descriptores de un grupo jerárquico en otro, cambio de términos sinónimos (empleo de un antiguo no descriptor), desambiguación en el caso de detectar términos sinónimos y la incorporación o eliminación de relaciones semánticas.