

Lenguajes documentales

Manela Juncà Campdepadrós
Damián Martínez Ferreras

PID_00143955

Material docente de la UOC



Universitat Oberta
de Catalunya

www.uoc.edu

Manela Juncà Campdepadrós

Licenciada en Geografía e Historia por la Universidad de Barcelona y posgrado de Información y Documentación en la empresa ICT. Ha diseñado diversos sistemas de información: en el Comité organizador de los Juegos Olímpicos de Barcelona 1992, en el Servicio de archivos del Departamento de Cultura y en el Archivo Nacional de Catalunya, y en ejercicio libre ha participado en numerosos másters de la Universidad de Girona. Actualmente es consultora en la UOC. Las áreas de interés son los Lenguajes documentales y las Fuentes de información.

Damián Martínez Ferreras

Licenciado en Geografía e Historia por la Universidad de Deusto. Actualmente ocupa el puesto de responsable del área de Research para España y Portugal en la empresa de consultoría Accenture. Experto en la implantación de Sistemas de Gestión del Conocimiento y en Primary Research. Máster en Documentación Digital por la UPF y de Liderazgo y Gestión de Equipos por AEDE Business School. Ha sido profesor asociado de la Universidad Carlos III en el área de Documentación e Información y actualmente es consultor de la UOC.

Primera edición: septiembre 2009

© Manela Juncà Campdepadrós, Damián Martínez Ferreras

Todos los derechos reservados

© de esta edición, FUOC, 2009

Av. Tibidabo, 39-43, 08035 Barcelona

Diseño: Manel Andreu

Realización editorial: Eureka Media, SL

ISBN: 978-84-692-4642-9

Depósito legal: B-33.217-2009



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. Fundació 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>

Contenidos

Módulo didáctico 1

La cadena documental

Manela Juncà Campdepadrós

1. Cadena documental
2. Fase de entrada
3. Fase de análisis y tratamiento
4. Fase de salida
5. Cadena documental y cooperación
6. Los lenguajes documentales dentro de la cadena

Módulo didáctico 2

Análisis de contenido: resumen e indización

Manela Juncà Campdepadrós

1. El análisis de contenido
2. El resumen
3. La indización
4. Los lenguajes documentales
5. Tipología de los lenguajes documentales

Módulo didáctico 3

Sistemas de clasificación documentales

Manela Juncà Campdepadrós

1. Clasificar y sistemas de clasificación
2. La clasificación decimal universal
3. Estructura de la clasificación decimal universal
4. Empezamos a clasificar
5. Prácticas con la CDU
6. Creación de un sistema de clasificación propio

Módulo didáctico 4

Listas de encabezamientos de materia y listas de autoridades

Manela Juncà Campdepadrós

1. Las autoridades en el análisis de contenido
2. ¿Qué es una lista de encabezamientos de materia?
3. Origen y evolución de las listas de encabezamientos
4. Tipologías de las listas de encabezamientos
5. Principios de las listas de encabezamientos
6. Encabezamientos y subencabezamientos
7. Reglas de vocabulario y signos de puntuación de las LEM
8. Términos simples y compuestos
9. Las relaciones semánticas
10. Aprendiendo a indizar con una lista de encabezamientos
11. La lista de autoridades

Módulo didáctico 5

Los tesauros

Damián Martínez Ferreras

1. El tesoro
2. Estructura del tesoro
3. Formas de presentación del tesoro
4. Creación y utilización de un tesoro
5. Sistemas automatizados de gestión de tesauros. Tesauros en línea
6. Tesauros y ontologías

Módulo didáctico 6

Listado de descriptores libres y listado de palabras clave

Manela Juncà Campdepadrós

1. Lenguajes documentales libres
2. Listado de descriptores libres
3. Listado de palabras clave

La cadena documental

Manela Juncà Campdepadrós

PID_00143966



Universitat Oberta
de Catalunya

www.uoc.edu



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. Fundació 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

Introducción	5
Objetivos	6
1. Cadena documental	7
2. Fase de entrada	9
2.1. Selección	9
2.1.1. Tamaño y contenido de una colección y un fondo archivístico	10
2.1.2. Herramientas para seleccionar	11
2.1.3. Evaluación del documento	11
2.1.4. Expurgar documentación	12
2.2. Adquisición	14
2.3. Recepción	15
2.4. Conclusiones	16
3. Fase de análisis y tratamiento	17
3.1. Análisis documental	18
3.1.1. Del documento a la referencia documental	18
3.1.2. Análisis formal	19
3.1.3. Análisis de contenido	20
3.1.4. Intercambio entre SID	21
3.1.5. Referencias bibliográficas	22
3.1.6. Conclusiones	22
3.2. Procesamiento técnico	23
3.2.1. Signatura topográfica	23
3.2.2. Ordenación	23
3.2.3. Almacenamiento y conservación	25
3.2.4. Conclusiones	26
4. Fase de salida	28
4.1. Instrumentos de búsqueda	28
4.1.1. Fuentes bibliográficas para bibliotecas y centros de documentación	28
4.1.2. Documentación de archivo	29
4.1.3. Instrumentos de búsqueda para recursos en línea	30
4.2. Servicios de difusión	31
4.3. Conclusiones	33
5. Cadena documental y cooperación	34

6. Los lenguajes documentales dentro de la cadena.....	37
Actividades.....	39
Glosario.....	40
Bibliografía.....	44

Introducción

Este módulo os introduce en los procesos documentales, que se llevan a cabo en archivos, bibliotecas y centros de documentación. Estos procesos se conocen como cadena documental.

Itinerario de estudio

El módulo empieza con una tabla que describe esta cadena. Acto seguido se analiza cada una de las fases y subfases y se aplican a los tres tipos principales de servicios de información: archivos, bibliotecas y centros de documentación, comentando las particularidades de cada uno. Finalmente se sitúan los lenguajes documentales, que son el cuerpo central de esta asignatura dentro de la cadena.

Conceptos más importantes

Concepto	Ved
Selección	2.1. Selección
Adquisición	2.2. Adquisición
Recepción	2.3. Recepción
Análisis documental	3.1. Análisis documental
Análisis formal	3.1.2. Análisis formal
Análisis de contenido	3.1.3. Análisis de contenido
Procesamiento técnico	3.2. Procesamiento técnico
Instrumentos de búsqueda	4.1. Instrumentos de búsqueda
Servicios de difusión	4.2. Servicios de difusión

Objetivos

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

1. Conocer las fases y subfases de la cadena documental.
2. Aproximarse al desarrollo de la cadena en diferentes servicios de información.
3. Situar los lenguajes documentales en la cadena.

1. Cadena documental

La **cadena documental** es el conjunto de operaciones y técnicas que se aplican a la colección o fondo documental de un servicio de información y documentación (SID¹) y que comprenden las fases de entrada, análisis-tratamiento y difusión.

⁽¹⁾SID es la sigla de *servicio de información y documentación*.

Cuando un usuario consulta un catálogo o una base de datos encuentra perfectamente organizados todos los documentos. Cada uno de ellos ha sufrido un proceso, conocido como cadena documental², que lo ha llevado desde el punto donde se ha generado el documento (editorial, despacho...) hasta las estanterías a punto de ser consultado. El documentalista del SID, de acuerdo a las diferentes políticas de adquisición o de gestión documental definidas en la organización, ha decidido que valía la pena seleccionarlos, los ha registrado, ha analizado tanto los atributos formales como los de contenido de acuerdo con las normativas nacionales e internacionales y finalmente los ha preparado para ser difundidos entre los usuarios.

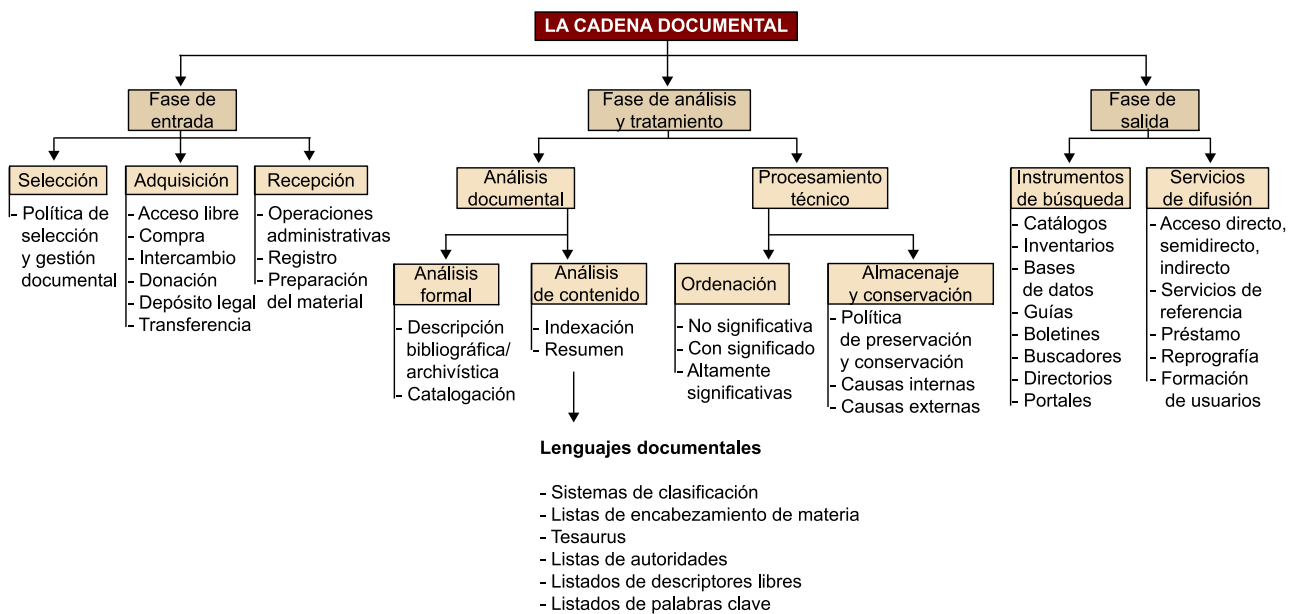
⁽²⁾La cadena también se conoce con el nombre de *proceso documental*.

Los procesos de la cadena se aplican a todo tipo de SID (archivos, bibliotecas y centros de documentación) y de documentos (textos, audiovisuales, tanto en papel como electrónicos). Los procesos no se aplican, sin embargo, de manera uniforme en todos los casos, ya que la documentación de archivo generada en procesos internos se trata de manera diferente a la documentación incluida en la cadena bibliográfica de una biblioteca o de un centro de documentación, ya que se utilizan normas de descripción diferentes y se generan instrumentos de búsqueda también diferentes.

Documentos de archivo y colecciones

- **Documentos de archivo:**
Documentos que se originan en el seno de un archivo, fruto de sus actividades y procesos (memorias, balances, actos, facturas, albaranes, cartas, informes, catálogos de productos, etc.), cuyo tratamiento se hará a partir de las técnicas archivísticas y las políticas de gestión documental.
- **Colecciones:**
Documentos que una biblioteca o un centro de documentación adquieren para llevar a cabo su actividad principal de “gestionar información” (libros, revistas, bases de datos, material audiovisual, etc.), cuyo tratamiento se hará según las normas desarrolladas por la biblioteconomía.

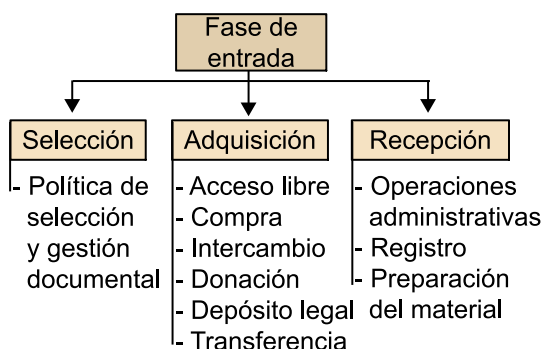
La cadena documental consta de **tres fases**: entrada, análisis y tratamiento, y salida o difusión. Cada una de ellas se divide en diferentes subfases, donde se recogen los diferentes instrumentos relacionados con técnicas tanto archivísticas como bibliográficas, como podéis ver en el esquema siguiente:



El tratamiento detallado de las diferentes partes de la cadena documental es objeto de estudio de otras asignaturas.

2. Fase de entrada

La fase de entrada consta de tres subfases: selección, adquisición y recepción.



2.1. Selección

La **selección** es el proceso en el que se decide qué documentos se añaden o eliminan de la colección o del fondo de archivo.

Es un proceso relativamente nuevo que nace de las siguientes circunstancias:

- El elevado volumen de obras disponibles en el mercado editorial y en la web.
- El crecimiento de las colecciones que determina que algunas obras queden obsoletas.
- Las limitaciones de espacio y de presupuesto de los SID o de los archivos.
- La imposibilidad de llevar a cabo el análisis documental del total de documentos interesantes para el centro.
- El crecimiento de documentos generados diariamente en la actividad cotidiana de las organizaciones.

Todos los SID tienen una política de fondos de archivo y/o de desarrollo de la colección que permite construir un útil y equilibrado fondo bibliográfico o archivístico. Esta política incluye la selección como una de las fases más relevantes, además de la planificación de recursos a compartir, por ejemplo, reemplazar obras perdidas o dañadas, deseleccionar ítems (no renovar suscripciones) y expurgar documentos.



La selección es un proceso complejo en el que intervienen muchos factores. Con respecto a la colección, la adquisición de documentos depende de las necesidades del centro y de sus usuarios, de estadísticas de uso, recursos humanos y recursos económicos, mientras que con respecto a los fondos de archivos, la selección está vinculada al sistema de gestión documental.

La selección y su plasmación real en la fase siguiente, la adquisición, son las fases de más peso económico de toda la cadena para bibliotecas y centros de documentación. La mayor parte del presupuesto de una red o un SID se dedica a la adquisición de fondos. Tema aparte son los archivos, ya que la entrada de documentos está condicionada por la propia actividad de la organización.

2.1.1. Tamaño y contenido de una colección y un fondo archivístico

En las bibliotecas, la selección da respuesta a cuestiones de tipo cuantitativo (se relacionan con el tamaño) y cualitativo (se relacionan con el contenido).

El **tamaño de la colección** responde a la pregunta ¿cuántos documentos hace falta seleccionar?

Decidir el tamaño óptimo es una tarea compleja, ya que una colección grande no es forzosamente la mejor, pero si no contiene un número adecuado de documentos, puede dejar algunas necesidades sin cubrir. Existen normas cuantitativas, editadas por organismos como las asociaciones y federaciones de bibliotecas, que dan pautas del tipo de 2 a X volúmenes por habitante o estudiante, número de volúmenes mínimos de una colección, incremento anual, etc. Se aplican sobre todo en bibliotecas públicas.

Con respecto a los criterios de **contenido de la colección**, responden a: “¿qué tipo de documentos?”

Actualmente las colecciones son muy diversas y no están formadas sólo por un tipo de documento, como tiempo atrás. Los tipos de documentos a seleccionar son muy variados: obras de referencia, monografías, publicaciones periódicas, publicaciones oficiales, bases de datos de revistas, etc., tanto en formato impreso como electrónico. La documentación no es original (excepto algunos ejemplares antiguos y raros), lo que permite que sea reemplazado si se deteriora o se pierde. Este hecho hace de la documentación bibliotecaria una documentación estable y segura.

Ved también

Los factores que intervienen en el proceso de selección son objeto de estudio en las asignaturas *Gestión integral de archivos* y *Gestión documental*.

El recurso de la economía de escala

Dado que la mayor parte del presupuesto se dedica a la adquisición de fondos, los consorcios bibliotecarios dedican muchos esfuerzos a confeccionar economías de escala y conseguir pactar mejores condiciones económicas en los contratos de paquetes documentales.

Obras de referencia:

Enciclopedias, diccionarios, manuales, estadísticas, anuarios, directorios, guías, fuentes geográficas, fuentes bibliográficas, catálogos, repertorios, boletines, ...

En los centros de documentación se debe hacer una mención especial a la literatura gris, es decir, a toda la documentación que no se distribuye por los canales habituales de distribución. Dado su elevado grado de especialización y actualización, es realmente significativa en un centro de documentación.

Literatura gris:

Tesis, informes, normativas, patentes, dossieres de prensa, actas, ...

En los archivos, el tamaño y el contenido del fondo vendrán definidos por la política de gestión documental de cada organización, plasmada en un sistema de gestión documental donde se definirá la elección y eliminación de documentos.

2.1.2. Herramientas para seleccionar

Los documentalistas seleccionan a partir de las herramientas siguientes:

- Catálogos comerciales de editoriales, distribuidoras y algunos librereros de viejo.
- Bases de datos de publicaciones en serie (revistas).
- Contacto con las entidades o personas que producen literatura gris.
- Bibliografías nacionales y temáticas.
- Repertorios bibliográficos como la base de datos del ISBN³ que recoge toda la producción editorial.
- Peticiones de lectores profesionales. Se da especialmente en bibliotecas universitarias, donde los profesores piden las últimas novedades.
- Sistema de gestión documental de la organización, donde se identifican los documentos que son evidencias de las actividades de la organización y que, por lo tanto, forman parte del sistema.

⁽³⁾ISBN es la sigla de *International Standard Book Number*, Número Normalizado Internacional del Libro.

2.1.3. Evaluación del documento

“La selección es una tarea discriminativa que supone separar una cosa de otra mediante su elección o eliminación. Ahora bien, para discriminar es necesario valorar los documentos según el provecho que se espera obtener. Por lo tanto, la valoración es el principal filtro, aunque no el único, de la tarea selectiva.”

M.^a Ángeles López (1996, pág. 145).

Para evaluar la calidad de una obra lo mejor es disponer de un ejemplar y objetivamente valorar aspectos físicos y de contenido (Carrizo, 2000):

a) **Aspectos físicos:** la encuadernación, la facilidad en el manejo, la tipografía.

Lectura complementaria

Podéis ampliar la información sobre la selección leyendo la obra siguiente:

M. A. López Hernández (1996). “La selección documental”. *Revista general de información y documentación*. Madrid.

b) **Aspectos de contenido:** título suficientemente explícito, cobertura adecuada al SID (cobertura temática, cronológica y geográfica). Los elementos más definitivos son los de contenido intelectual: la autoría reconocida del organismo y autor, los elementos constitutivos de la obra (prólogo, introducción, etc.), la organización del contenido (sumarios, índices, cuadros, imágenes, bibliografía, etc.), la exhaustividad y las posibilidades de actualización (suplementos, nuevas ediciones, etc.).

También se seleccionan y evalúan otras fuentes como los recursos web, que aunque no supongan una limitación de espacio, son tan abundantes y de cualidades tan diversas, que hay que hacer una elección. Para evaluar también los recursos web se añaden los parámetros de acceso a la información, ergonomía y facilidad de utilización, luminosidad, ubicuidad y aprendizaje.

La política de selección de recursos webs gratuitos de la biblioteca de la UOC

La biblioteca de la UOC utiliza un criterio de excelencia más que uno de exhaustividad. De los muchos recursos web gratuitos, selecciona un mínimo de 5 y un máximo de 15 por nodo temático. Dentro de cada nodo temático se desarrollan unas formas documentales mínimas: portales temáticos, organismos e instituciones, revistas electrónicas, webs, enciclopedias y diccionarios. Siempre que sea posible, también se intenta desarrollar el máximo de las formas restantes: colecciones de textos, directorios, estadísticas, prensa, publicaciones oficiales, traductores, base de datos, documentos de trabajo, documentos electrónicos, libros electrónicos, normativas y estándares, buscadores, listas de discusión, recursos docentes, tesis, TFC-Prácticums. No hay limitaciones geográficas, se valora el contenido especializado y de licencia libre. Con el fin de valorar la calidad de cada recurso, la UOC utiliza los parámetros de Lluís Codina: autoría/fuente, contenido, acceso a la información, ergonomía y facilidad de utilización y, finalmente, luminosidad, más dos complementarios: ubicuidad y aprendizaje.

Por último, la evaluación del documento de archivo se hace de acuerdo con el sistema de gestión documental definido en la organización, donde están establecidas las tablas de evaluación documental.

2.1.4. Expurgar documentación

Expurgar es hacer una evaluación crítica de la colección y del fondo a partir de criterios materiales e intelectuales que nos permitan retirar de manera temporal o definitiva una parte de la colección o del fondo con el fin de reorganizar los contenidos del SID y ofrecer un mejor servicio.

Según Vall (2006), el expurgo se tiene que entender como un proceso más de la gestión de la colección y el fondo, y por lo tanto los criterios que se seguirán se han de incluir en la política documental de cada centro.

a) Expurgar fuentes de archivos

Lectura complementaria

Podéis encontrar el trabajo de Gloria Carrizo en:

G. Carrizo Sainero et al. (2000). *Manual de fuentes de información*. Madrid: CEGAL.

Ved también

Los parámetros establecidos por Lluís Codina se estudian en la asignatura *Fuentes de información*.

Términos sinónimos

Otros términos sinónimos de expurgar son: descartar, retirar, hacer una selección negativa. UNE-ISO 15489-1:2006. *Información y documentación: gestión de documentos parte 1 Generalidades*. UNE-ISO 15489-2:2006. *Información y Documentación. Gestión de documentos parte 2 directrices*.

Hay que expurgar la documentación archivística porque la documentación sigue un ciclo de vida documental: primero se genera en la actividad diaria de la gestión del centro, en segundo lugar pasa a tener valor administrativo y finalmente puede tener valor histórico.

En muchos países, el expurgo está regulado por ley. En el ámbito del Estado español, hay comunidades autónomas que disponen de comisiones evaluadoras y de normativa específica.

b) Expurgar colecciones de biblioteca

En colecciones de bibliotecas y centros de documentación hay que expurgar para:

- Mantener la pertinencia de la colección, eliminando los documentos no usados. Se calcula a partir de la fecha de la última consulta y circulación del documento. Estos índices se adaptan a los diversos ámbitos, ya que no es lo mismo la pervivencia de los documentos científicos (breve) que la de los documentos de humanidades (mucho más larga, ya que no dependen tanto de la actualización).
- Evitar la duplicidad de formatos (revistas que ahora se han reconvertido en bases de datos, revistas que ahora tienen edición electrónica, etc.).
- Favorecer un mejor manejo de la colección: los documentos en las estanterías son los más pedidos y no hay que buscar entre toda la producción.
- Solucionar problemas de espacio.
- Eliminar ejemplares deteriorados y envejecidos.

En la opinión de Vall (2006), durante mucho tiempo el modelo bibliotecario nacional ha sido preservador y conservador de las colecciones documentales, pero ahora se enfrenta a la fuerte presión del crecimiento constante del fondo. La solución pasa por la expurgación de la colección.

Marco legal y normativo de la expurgación en bibliotecas

A nivel internacional se cuenta con las directrices de la IFLA/UNESCO para bibliotecas públicas (2002) y la guía para la revisión de colecciones de biblioteca de la American Library Association (1999). En España se cuenta con las normas y directrices para bibliotecas universitarias y científicas de REBIUN (1999), y en Cataluña hay dos leyes y un decreto del año 1993 sobre patrimonio, servicios y personal del sistema de lectura pública.

Lectura complementaria

Podéis ampliar la información sobre el tema de la expurgación en:

A. Vall Casa (2006). "Esporgar: per què, com i quan". *BID Textos universitaris de biblioteconomia i documentació* (n.º. 16, junio, pág. 2). Barcelona.

Lectura recomendada

Marc legal de l'esporgament en arxius; Comissió Nacional d'Accés, Avaluació i Tria Documental (CNATD). Generalitat de Catalunya.

El destino de los documentos expurgados puede ser diverso. Algunos se quedan en la biblioteca, otros se retiran, se dan a otras bibliotecas, se hacen donaciones solidarias a ONG, van a la Central de Préstamo del gobierno correspondiente, se guardan en almacenes (como el proyecto GEPA del Consorcio de Bibliotecas Universitarias de Cataluña), se regalan, se venden (opción muy establecida en el mundo anglosajón y no tanto aquí), o se destruyen.

2.2. Adquisición

Los SID adquieren o incorporan volúmenes y documentos a través de diferentes vías, algunas de ellas gratuitas y otras de pago:

1) **Acceso gratuito (*open acces*):** es toda aquella información dispuesta de forma gratuita y universalmente accesible vía Internet.

Open acces es un nuevo modelo de distribución de información académica desarrollado por investigadores y documentalistas para contrarrestar las limitaciones impuestas por los comerciales (suscripciones elevadas), particularmente en el campo de las ciencias de la salud y la tecnología.

2) **Compra:** puede ser directa o por suscripción al editor, la librería, el proveedor. Es la vía más habitual en bibliotecas y centros de documentación. Implica un conocimiento elevado de la producción bibliográfica, los canales de distribución y venta.

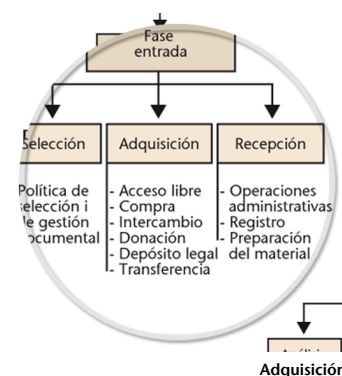
3) **Donación:** es la cesión gratuita de documentos. Puede venir por parte del autor de la obra, por un particular o por la propia dinámica de la institución. Las donaciones pueden ser definitivas o temporales. Se da, sobre todo, en las donaciones de las editoriales, de las administraciones públicas y en los lotes fundacionales que la administración estatal o autonómica remite a las bibliotecas públicas de su titularidad. Entre las donaciones podemos distinguir los legados y los depósitos:

- Los **legados** son donaciones provenientes de testamentos y están formados por colecciones particulares completas o parciales.
- Los **depósitos** son donaciones en las que la propiedad continúa siendo del donante.

4) **Intercambio:** como su nombre indica, consiste en el intercambio de material duplicado (o poco solicitado) entre dos centros. El intercambio se acostumbra a dar más fácilmente en las redes de SID. Se intercambian duplicados, publicaciones propias o de las instituciones a las que están vinculados y literatura gris.

El CBUC

En Cataluña, el Consorci de Biblioteques Universitàries de Catalunya (CBUC) mantiene el programa GEPA (Garantía de Espacio para la Perpetuidad del Acceso), que es un almacén cooperativo donde conservar y preservar los documentos de bajo uso del consorcio.



Open Acces

Ejemplos de información gratuita son los repositorios institucionales, los depósitos de archivos abiertos o los recolectores.

Web recomendada

En esta dirección podéis consultar todas las instituciones españolas que tienen archivos en abierto:

- OpenDOAR (Directory of Open Access Repositories).

Webs recomendadas

En la web de la Biblioteca de Catalunya podemos consultar las condiciones y el listado de donantes.

En la web del Arxiu Nacional de Catalunya, concretamente en el apartado de descripción del fondo, podemos consultar la tipología de fondos que han llegado por donación.

No es un procedimiento totalmente gratuito, ya que requiere una contrapartida.

5) Depósito legal (DL⁴): es la exigencia impuesta por ley de depositar, en una o diversas agencias especificadas, ejemplares de las publicaciones de todo tipo, en cualquier soporte, por cualquier procedimiento de distribución (venta, alquiler, etc.). El primero en instaurarlo fue Francisco I en 1537 en Montpellier. En España lo instauró Felipe III el año 1619. El DL es un sistema de adquisición propio de un reducido número de bibliotecas, concretamente las nacionales.

Los documentos susceptibles de DL son: libros, folletines, hojas impresas, publicaciones periódicas y seriadas, partituras musicales, grabados, mapas y planos, carteles, postales, naipes, diapositivas, producciones cinematográficas. Está estipulado depositar cinco ejemplares de las obras que llevan ISBN y tres para el resto, dos para los audiovisuales y un ejemplar para los guiones de cine.

El papel del DL es mucho más que una simple vía de adquisición, ya que entre sus objetivos destacan formar la colección nacional del país y elaborar la bibliografía nacional.

6) Transferencias: se da sobre todo en archivos. Es el acto por el cual un conjunto de documentos es trasladado de la unidad administrativa que lo ha producido a un servicio de archivos con plenos derechos para su tratamiento y disposición. Este traslado se efectúa mediante la presentación de un instrumento denominado *formulario de transferencia* (Alberch, 1999).

2.3. Recepción

La recepción es el último proceso de la fase de entrada y consta de tareas administrativas, el registro y la preparación del material.

Las tareas administrativas vinculadas a la adquisición son diferentes según la documentación haya llegado por compra, transferencia, donación, etc.

Con respecto a las compras, las tareas son comprobar el estado del pedido, aceptar la factura o, si es el caso, reclamar y gestionar los pagos. Con respecto a los documentos procedentes de la actividad y el proceso de organización, las tareas a realizar en el momento de la recepción son la evaluación y el registro, de acuerdo con los procedimientos definidos en el sistema de gestión documental.

Instituciones

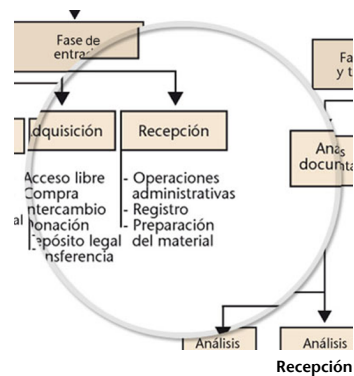
Son ejemplos de intercambios entre instituciones las publicaciones de la diputación o del consejo comarcal.

⁽⁴⁾DL es la sigla de *depósito legal*.

Lectura complementaria

Sobre las transferencias podéis consultar la obra siguiente:

R. Alberch i Fugueras (1999). "El tractament tècnic dels documents". *Gestió integral d'arxius*. Barcelona: Fundació per a la Universitat Oberta de Catalunya.



Finalmente, el proceso acaba con operaciones de preparación de los diferentes tipos de material, como sellar el documento, forrar, reforzar para que no se deterioren fácilmente, magnetizar o etiquetar para el préstamo.

En cuanto a los documentos de archivo, hay operaciones específicas: ser delicado con la documentación histórica, desdoblar documentos como planos o mapas, retirar clips y fundas plastificadas que puedan estropear la documentación a largo plazo, etc.

2.4. Conclusiones

La fase de **entrada** consta de tres subfases:

1) La **selección**: es el proceso en el que se decide qué documentos se añaden o eliminan de la colección o fondo. Es un proceso complejo en el que intervienen muchos factores: las necesidades del centro, las de sus usuarios, estadísticas de uso, recursos humanos y recursos económicos o el sistema de gestión documental si se trata de un archivo.

La selección da respuesta a cuestiones de tipo cuantitativo (se relacionan con el tamaño) y cualitativo (se relacionan con el contenido). Los documentalistas seleccionan usando diversas herramientas, por ejemplo catálogos comerciales, bases de datos, contactos directos, bibliografías, repertorios y desideratas.

El principal criterio a la hora de seleccionar es evaluar la calidad de una obra. Hay que valorar los aspectos físicos y los de contenido. Un tipo especial de selección es la expurgación, que consiste en retirar de manera temporal o definitiva una parte del fondo con el fin de reorganizar los contenidos del SID y ofrecer un mejor servicio.

2) La **adquisición**: las vías de adquisición son el acceso gratuito, la compra, la donación, el intercambio, el depósito legal y la transferencia.

3) La **recepción**: consta de tareas administrativas, el registro y la preparación del material.

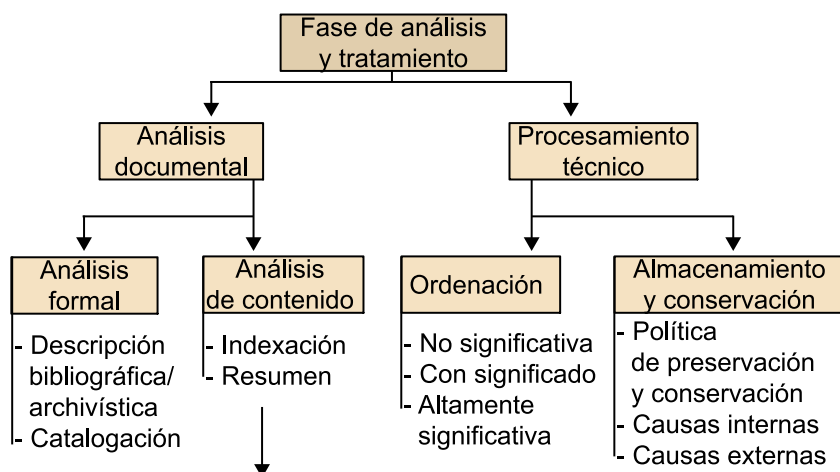
Ved también

Todas estas operaciones ligadas a los documentos de archivo son objeto de estudio en las asignaturas *Gestión integral de archivos* y *Gestión documental*.

3. Fase de análisis y tratamiento

La fase de **análisis y tratamiento** incluye dos operaciones: el análisis documental y el procesamiento técnico. Su función es organizar los documentos seleccionados en la primera fase y producir herramientas que faciliten la recuperación.

El análisis documental transforma el documento original en un registro documental que se introduce en el catálogo o inventario. El procesamiento técnico se ocupa de ordenar y conservar los documentos en condiciones óptimas.



Lenguajes documentales

- Sistemas de clasificación
- Listas de encabezamiento de materia
- Tesauros
- Listas de autoridades
- Listados de descriptores libres
- Listados de palabras clave

3.1. Análisis documental

“El análisis documental es la operación intelectual de aplicar técnicas específicas y normalizadas a un colectivo de documentos con la finalidad de hacerlos controlables y recuperables.”

García Gutiérrez (1984, pág. 77).

3.1.1. Del documento a la referencia documental

Los documentos tienen una serie de atributos formales (autor, título, edición, extensión, etc.) y atributos de contenido (la materia, el resumen); los documentos de archivo tienen, además, el contexto histórico y social en el que fueron generados. El analista vacía estos atributos de acuerdo con normas descriptivas y elabora otro producto documental nombrado **referencia documental**, que contiene todos los datos identificativos y descriptivos del documento. Esta referencia se introduce en el catálogo y se convierte en la herramienta que representa el documento para futuras búsquedas.

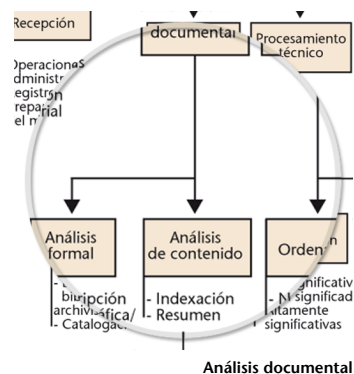
Referencia documental de *Vocabulario básico de la historia medieval, de Pierre Bonnassie*

Ejemplo de referencia documental

Autor	<u>Bonnassie, Pierre</u>
Título	Vocabulario básico de la historia medieval / Pierre Bonnassie; trad. cast. y adapt. de Manuel Sánchez Martínez
Descripción	Barcelona: Crítica, 1994
Edición	4.ª ed.
ISBN	8474232015
Colección	<u>Serie general (Crítica). Estudios y ensayos; 110</u>
Materia	Historia medieval - Terminología

Como se puede apreciar, la referencia documental es la representación condensada del documento/fondo. Puede ir acompañada del sumario y el resumen de la obra.

¿Por qué necesitamos referencias documentales? Porque no es operativo trabajar con los documentos originales, ya que estos sólo pueden estar en un lugar a la vez, y porque a veces no son unidades independientes sino que forman parte de otra obra. Por ejemplo, el documento de un expediente, etc.



Documentos de archivo

Dada la extensión de un registro archivístico, para ver un ejemplo recomendamos la consulta directa a la norma ISAD(G), pág. 99, fondo del organismo *Consejo real de España e Indias*.

Terminología

No hay consenso terminológico sobre el término *referencia documental*. Otros nombres son reseña, ficha catalográfica, registro bibliográfico (usado en catálogos), registro documental (usado en bases de datos documentales).



Portada de *Vocabulario básico de la historia medieval*, de Pierre Bonnassie

Construir la referencia documental es una tarea intelectual, la más intelectual de toda la cadena. El analista tiene a su disposición normativas, reglas y listados para consultar, que le indican qué información tiene que seleccionar, cómo la tiene que introducir, en qué campo y en qué formato.

El análisis documental se divide en dos análisis diferenciados: el análisis formal y el análisis de contenido.

3.1.2. Análisis formal

El objetivo del análisis formal⁵ es describir los atributos estructurales de los documentos, proporcionando una representación del documento que lo describa de manera única, sin ambigüedades.

⁽⁵⁾El análisis formal también se conoce por el nombre de *catalogación y descripción bibliográfica*.

Los objetivos concretos y la normativa usada en cada caso es la siguiente:

1) Obtener parte de los elementos que componen una referencia documental: autor, título, año, edición, lugar de edición, colección, ISBN, etc.

En documentación archivística, el análisis formal se lleva a cabo según una descripción normalizada acordada a nivel internacional: las ISAD(G)⁶ del International Council on Archives.

⁽⁶⁾ISAD(G) es la sigla de *General International Standard Archival Description* (Norma Internacional General de Descripción Archivística).

En Cataluña, la Norma de Descripción Archivística de Catalunya (NODAC⁷) se articula en siete áreas de información descriptiva que reúnen los veintiséis elementos identificados y definidos por la ISAD(G).

⁽⁷⁾NODAC es la sigla de *Norma de Descripción Archivística de Catalunya*.

Para colecciones de biblioteca se utilizan las International Standard for Bibliographic Description (ISBD⁸), publicadas por la International Federation of Library Associations and Institutions (IFLA⁹) y la primera parte de las Anglo-American Cataloguing Rules (AACR2R¹⁰), que nos indican los elementos esenciales de la descripción documental, de qué partes del documento se tienen que sacar, en qué orden hay que ponerlas, y qué signos de puntuación y caracteres gráficos se tienen que usar. Hay una ISBD(G) que es el marco general y abstracto y toda una serie de ISBD para los diferentes tipos de documentos.

⁽⁸⁾ISBD es la sigla de *International Standard for Bibliographic Description*, Norma Internacional para la Descripción Bibliográfica.

⁽⁹⁾IFLA es la sigla de *International Federation of Library Associations and Institutions*, Federación Internacional de Asociaciones e Instituciones de la Biblioteca.

⁽¹⁰⁾AACR2R es la sigla de *Anglo-American Cataloguing Rules*.

Ejemplos de ISBD

- ISBD(A): libros antiguos.
- ISBD(ER): recursos electrónicos.
- ISBD(CM): documentos cartográficos.
- ISBD(M): monografías.
- ISBD(NBM): documentos no libros.
- ISBD(PM): música impresa.
- ISBD(S): publicaciones en serie.

Ved también

Las normas ISBD y AACR2R se estudian en la asignatura *Análisis documental*, y la norma ISAD(G) en la asignatura *Gestión integral de archivos*.

2) Establecer y normalizar los puntos de acceso bibliográficos por autor y título. Los puntos de acceso son nombres, códigos o términos que sirven para localizar y buscar el documento. Los más habituales en este análisis formal son el autor y el título, pero en entornos automatizados cualquier elemento puede ser recuperado (como el lugar y el año de edición, el ISBN).

Para alcanzar este objetivo con colecciones de biblioteca utilizaremos la segunda parte de las AACR2R; 2.ª edición revisada: puntos de acceso por autor y título. En archivos, la norma para el control de autoridades es la Norma Internacional sobre los Registros de Autoridad de archivos relativos a instituciones, personas y familias (ISAAR[CPF]¹¹) del Consejo Internacional de Archivos.

⁽¹¹⁾ISAAR(CPF) es la sigla de *Norma Internacional sobre los Registros de Autoridad de Archivos Relativos a Instituciones*.

Campos propios del análisis formal en la referencia de Pierre Bonnassie

Autor	Bonnassie, Pierre
Título	Vocabulario básico de la historia medieval / Pierre Bonnassie; trad. cast. y adapt. de Manuel Sánchez Martínez
Descripción	Barcelona: Crítica, 1994
Edición	4.ª ed.
ISBN	8474232015
Colección	Serie general (Crítica). Estudios y ensayos, 110

Ved también

El análisis formal se estudia en la asignatura *Análisis documental*.

3.1.3. Análisis de contenido

El objetivo del análisis de contenido es identificar y representar de manera precisa la materia de los documentos, con el objetivo de permitir la recuperación. Esta parte del análisis documental establece los puntos de acceso por materias o contenidos de los documentos.

Reflexión

El material que tenéis en las manos trata precisamente del análisis de contenido de la cadena documental.

Se basa en dos operaciones:

- a) El **resumen**, que es la representación abreviada y precisa del contenido.
- b) La **indización**, que consiste en representar el contenido del documento mediante términos de indización extraídos de **lenguajes documentales**: notaciones, encabezamientos de materias, descriptores, identificadores, palabras clave, unitérminos. Cuando se representa el contenido siguiendo un sistema de clasificación en lugar de una *indización* se conoce como *clasificación*.

Las **normativas** que usamos en esta parte del análisis documental son:

- UNO 50-103-90, preparación de resúmenes.
- UNO 50-121-91, métodos para el análisis de documentos, determinación de su contenido y selección de términos de indización.

- Las normativas propias de cada lenguaje documental: vocabulario, combinaciones, mantenimiento, actualización.

Campos propios del análisis de contenido en la referencia de Pierre Bonnassie: materia y resumen

Campos propios del análisis de contenido en la referencia de Pierre Bonnassie

Materia	Historia medieval - Terminología
Resumen	Este es un libro poco corriente. Ni diccionario ni manual, significa una nueva y eficaz forma de introducción -a la vez analítica y sintética- a los problemas de la historia de la Edad Media. En efecto, a partir del análisis de medio centenar de conceptos fundamentales y de su evolución semántica, el profesor Pierre Bonnassie, de la Universidad de Toulouse, consigue definir, con insólita precisión, las grandes cuestiones que hoy tiene planteadas la historia medieval. El resultado es un texto innovador, de uso obligado para profesores y estudiantes, que encontrarán en él un instrumento de trabajo insustituible.

Los lenguajes documentales usados tradicionalmente en los archivos son cuadros de clasificación contruidos a medida del fondo. El análisis de contenido es sintético, no se analizan los documentos individualmente, sino el fondo en su conjunto, dado que un documento forma parte de una cadena de documentos ordenados (cronológicamente, orgánicamente, funcionalmente) y aislado pierde su contexto. La clasificación puede ser orgánica, funcional (por funciones, por grandes materias) o mixta.

Clasificación orgánica

La clasificación orgánica es el retrato de la estructura orgánica de la entidad que haya generado la documentación.

En bibliotecas y centros de documentación se usan la mayoría de los lenguajes documentales. Los más habituales son los sistemas de clasificación, como la Clasificación Decimal Universal (CDU¹²) o la Clasificación Dewey, los listados de autoridades, las listas de encabezamientos de materia, los tesauros y la indización automática por palabras clave.

⁽¹²⁾CDU es la sigla de *Clasificación Decimal Universal*.

3.1.4. Intercambio entre SID

Con el mismo espíritu cooperativo y de intercambio de las ISBD y AACR, surge el formato Machine Readable Catalog (MARC 21¹³). Este formato permite intercambiar registros entre diferentes bases de datos y sus normas codifican, almacenan e intercambian datos bibliográficos.

⁽¹³⁾MARC 21 es la sigla de *Machine Readable Catalog*.

Consulta del registro de Bonnassie en el catálogo de la Biblioteca de la UOC, en el icono MARC

```

CAPÇALERA 00000nam 2200217 a 4500
008 950531s1994 spc||||| |00| ||spa|c
020 8474232015
035 8474232015|9ES-BaCBU001
035 0959-64760|9ES-BaCBU035
100 1 Bonnassie, Pierre
245 10 Vocabulario básico de la historia medieval /|cPierre
Bonnassie
250 4ª ed.
260 Barcelona :|bCrítica,|c1994
300 246 p.;|c20 cm
650 04 Història medieval|xDiccionaris|xCastellà
650 04 Història medieval|xTerminologia
830 0 Serie general (Crítica) ;|v110
935 0959-64760
998 UDG
998 UOC
998 UVIC
999 VTLSFF5000 990926 0000

```

A nivel de archivo, el intercambio de información en el contexto de sistemas y redes telemáticas está menos adelantado, por el hecho de que la descripción responde a documentos específicos de cada organismo.

Ved también

El formato MARC se estudia en la asignatura *Análisis documental*.

3.1.5. Referencias bibliográficas

No es una tarea de la cadena documental, pero por la similitud de nombres es conveniente evitar las confusiones entre referencia documental y referencia bibliográfica.

Una referencia bibliográfica es la descripción del documento primario al que se remite al lector. En términos especializados, es el documento que ha sido citado expresamente en un trabajo de investigación y al que es obligatorio, ética y textualmente, hacer alusión (Borgoñós, 1999, pág. 22).

Ejemplo

Si en una bibliografía citamos la obra de Pierre Bonnassie, su referencia bibliográfica es:

BONNASSIE, Pierre. *Vocabulario básico de la historia medieval*. Manuel Sánchez Martínez (trad.). Barcelona: Crítica, 1994. 246 p. Serie general (Crítica). Estudios y ensayos; 110. ISBN 8474232015.

Las **normativas** para hacer estas referencias son:

- Para materiales publicados, la norma ISO 690 (traducida al castellano como UNE 50-104).
- Para materiales electrónicos, la ISO 690-2.

Ved también

Estas dos normativas se estudian en la asignatura *Fuentes de Información*.

3.1.6. Conclusiones

A modo de resumen, a continuación tenéis una tabla que recoge los conceptos más importantes de las operaciones y normativas que hacen referencia al análisis documental.

Resumen de operaciones y normativas del análisis documental

	Operación	Normativa con la cual se alcanza
Análisis formal	Obtener parte de los elementos que componen una referencia documental.	ISBD y la primera parte de las AACR2R con respecto al material bibliográfico. ISAD(G) con respecto a la descripción de documentos de archivo.
	Elaborar los puntos de acceso bibliográficos por autor y título.	La segunda parte de las AACR2R. 2ª edición revisada para material bibliográfico.
Análisis de contenido	Resumir	Norma UNE 50-103-90; preparación de resúmenes. ISAD(G); campo: alcance y contenido para documentos de archivo.

	Indizar	Norma UNE 50-121-91; métodos para el análisis de documentos, determinación de su contenido y selección de términos de indización. La normativa propia de cada lenguaje documental sobre vocabulario, combinaciones, mantenimiento, actualización.
En conjunto	Intercambiar formato MARC	MARC 21 para material bibliográfico.

3.2. Procesamiento técnico

En este momento de la cadena, el documento está a punto para ser ordenado, ya sea en una estantería, una carpeta o unos archivadores. En paralelo, el SID tiene que prever que las condiciones del almacén sean óptimas, hecho que implica medidas de conservación y una política de prevención.

Este conjunto de tareas, ordenación, almacén y medidas de conservación, se conoce como **procesamiento técnico**.

3.2.1. Signatura topográfica

La **signatura topográfica** es el código que identifica cada material en las salas y estanterías del SID.

La signatura queda registrada en la referencia documental y también en la obra en cuestión. Está estrechamente relacionada con la gestión del espacio y, en consecuencia, con la ordenación.

Signatura topográfica

Los libros, por ejemplo, llevan la signatura topográfica en el extremo inferior del lomo.

3.2.2. Ordenación

Los SID tipo archivos ordenan su documentación en cajas archivadoras y en muebles compactos que ahorran espacio. Estos muebles no están al alcance de los usuarios. Los documentos de un fondo se ordenan en series y expedientes. Físicamente se ordenan en cajas archivadoras, carpetas o camisas. Los documentos con formatos que impidan su ordenación en cajas, como pergaminos, planos, carteles, material audiovisual, etc., se ponen en cajoneras aparte. En estos casos, se clasifican con el resto de su fondo, pero se les da un topográfico diferente.

Las bibliotecas dan acceso libre y pueden ordenar sus colecciones según diferentes criterios espaciales: por el tipo de documentos (todas las monografías juntas, las revistas juntas, etc.), por temas (apartado para ciencias sociales, ciencias exactas, etc.) o en función del público (sala para adultos, sala infantil).

Combinación de criterios

Estos criterios se pueden combinar en una misma biblioteca. Por ejemplo, una biblioteca pública de una ciudad de 30.000 habitantes puede tener una sala infantil y el resto del edificio para adultos; dentro de cada espacio, los documentos ordenados por temas, y dentro de cada tema, por tipo (dentro de geografía, primero los atlas, después monografías, dvd, etc.), y finalmente una sala o espacio para documentos con algún formato especial.

Una vez los documentos están en los compactos o las estanterías, hay tres tipos de ordenaciones (Maniez, 1993):

a) Ordenaciones no significativas: son aquellas en las que no hay relación con el contenido del documento. Es una ordenación útil en SID que no den acceso libre a las estanterías, y que la recuperación la haga el documentalista. Apropiado para SID que tengan su colección en el depósito y no en la sala de lectura. Es el método más usado en Archivos.

Ordenaciones no significativas

Un ejemplo de ordenación no significativa sería ordenar según el número asignado a la llegada.

b) Ordenaciones con significado limitado: se ordenan por algún criterio. Es la ordenación que encontramos en librerías, bibliotecas personales o bibliotecas pequeñas.

Ordenaciones con significado limitado

Ejemplos de criterios utilizados en este tipo de ordenación serían la lengua del documento, el autor, el tema.

c) Ordenaciones altamente significativas: se ordena a partir de un cuadro de clasificación, de manera que los contenidos afines se colocan uno al lado del otro. Es apropiada para SID de libre acceso, como las bibliotecas públicas y universitarias.

Ordenaciones altamente significativas

Un ejemplo sería ordenar siguiendo las clases de la CDU.

Los tres tipos de ordenaciones se pueden combinar en un mismo SID.

Lectura complementaria

Podéis encontrar más información sobre estos tres tipos de ordenación en la obra siguiente:

J. Maniez (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruijérez.

Combinación de criterios

Estos criterios se pueden combinar en una misma biblioteca. Por ejemplo, una biblioteca pública puede ordenar de la siguiente manera:

- Ordenaciones no significativas: prensa, boletines, revistas, etc.
- Ordenaciones con significado limitado: novelas. Se hacen distinciones entre novela histórica, ciencia-ficción, biografías, etc. Dentro de cada grupo las novelas se ordenan alfabéticamente por autores.
- Ordenaciones altamente significativas: todo el resto de la colección. Es el grueso mayor.

Sea cual sea la ordenación, cada documento tiene que tener un lugar único y el sistema tiene que ser susceptible de ampliación (estanterías parcialmente vacías). En los SID de libre acceso la ordenación tiene que ser sencilla y comprensible para el usuario.

3.2.3. Almacenamiento y conservación

El año 1979 es un hito en la historia de la conservación de las colecciones bibliográficas. En esta fecha, la IFLA, a través de su Sección de Conservación, presentó un documento que se editó el año 1986 con el nombre de *Principios para la preservación y conservación de los materiales bibliográficos*.

Hasta entonces, la conservación documental se relacionaba con dos operaciones: la encuadernación y la restauración de libros y manuscritos antiguos. Estas actividades se consideraban casi un lujo reservado sólo a selectivas bibliotecas y archivos con fondo de extraordinario valor.

El documento de la IFLA amplió el concepto de conservación documental en todo tipo de SID y de fondo. Distingue tres **actividades**: la preservación, la conservación y la restauración.

Las causas que provocan la degradación pueden ser internas o externas al documento (Cid 1999, basado en Borell y Bello 1995):

a) Causas internas. La naturaleza orgánica del papel, los aditivos y las tintas de los documentos los hacen susceptibles de deterioro.

b) Causas externas. Pueden ser:

- Fisicomecánicas: producidas por el uso cotidiano y la manipulación de los documentos. Afecta tanto a los usuarios como al personal técnico del SID. Hace falta mantener la higiene de las manos, no humedecerse los dedos para pasar página, no doblar los márgenes de las páginas, no escribir sobre la documentación, no comer ni beber, etc.

La ordenación de las novelas

¿Por qué las novelas no se ordenan siguiendo la clasificación CDU, es decir de manera altamente significativa? La CDU clasifica las novelas por el idioma original del texto, de manera que una obra de Paul Auster está clasificada como novela norteamericana, y una de Jorge Luis Borges como novela argentina, aunque el SID la tenga en catalán o castellano. Buscar por el país de origen del escritor puede resultar complicado, así que se prefiere un criterio alfabético.

- **Ambientales:** por humedad, temperatura, luz intensa. La humedad y la temperatura son los factores más importantes para la conservación. Se aconseja entre un 50-60% de humedad relativa y entre 18-20° C de temperatura. Estos niveles se tienen que mantener sin oscilaciones bruscas. La luz solar se tiene que matizar porque es un factor muy degradador. La luz artificial es mejor que sea de fluorescente, y se tiene que mantener en torno a los 100 lux en las salas de exposiciones y 50 lux para los materiales delicados.
- **Químicas:** contaminación, polución atmosférica. La solución es mantener las salas ventiladas, instalar filtros en el aire acondicionado.
- **Biológicas:** hongos, bacterias, insectos, roedores, etc. Las salas y depósitos tienen que estar limpios y desinfectados para evitar los microorganismos y roedores. Por otra parte, el mantenimiento de las condiciones mencionadas de temperatura, ventilación e iluminación dificulta el crecimiento de las especies bibliófilas.
- **Causas extraordinarias:** incendios, inundaciones, robos. Se instalan sistemas de detección del fuego y puertas cortafuegos. Los documentos se alejan de los cuadros eléctricos y depósitos de agua. Alarmas conectadas con los sistemas de seguridad de la ciudad.

Las medidas de conservación son muy importantes en la documentación histórica de archivo: digitalización o microfichas de los documentos cuya conservación peligre o sea muy voluminosa, medidas contra los excesos de humedad, temperatura y luz, de control de plagas, talleres de restauración de documentos, etc.

Documentación delicada

Ejemplos de documentación que corre un peligro especial son la documentación de los primeros siglos, los documentos con una tinta poco estable, los documentos realmente muy valiosos, las grandes colecciones, etc.

La preservación de documentos electrónicos, y más concretamente los específicos de archivo, que llevan incorporada la firma electrónica, es otro aspecto que hay que tener en cuenta en esta parte de la cadena.

3.2.4. Conclusiones

La fase de análisis y tratamiento incluye dos subfases:

1) **El análisis documental.** Esta etapa es la operación intelectual de aplicar técnicas específicas y normalizadas a un colectivo de documentos con la finalidad de hacerlos controlables y recuperables. El análisis documental transfor-

Precauciones especiales

Algunos tipos de documentos necesitan unas precauciones especiales, por ejemplo los vídeos, que se tienen que alejar de los campos magnéticos, o los negativos de fotografías, que se tienen que guardar con papel de PH neutro.

Ved también

La preservación de documentos electrónicos se estudia de manera más detallada en la asignatura *Preservación de documentos digitales y gestión documental*.

ma el documento original en un registro documental que se introduce en el catálogo o inventario y es la representación condensada del documento original. El análisis documental se divide en dos análisis diferenciados:

- El **análisis formal**. El objetivo de este análisis es describir los atributos estructurales de los documentos, proporcionando una representación del documento que lo describa de manera única, sin ambigüedades.
- El **análisis de contenido** tiene por objetivo identificar y representar de manera precisa la materia de los documentos, con la finalidad de permitir la recuperación. Esta parte del análisis documental establece los puntos de acceso por materias. Tiene dos operaciones: el resumen y la indización.

El formato MARC permite intercambiar registros entre diferentes bases de datos y sus normas codifican, almacenan e intercambian datos bibliográficos.

2) El **procesamiento técnico**. Esta subfase se ocupa de ordenar y conservar los documentos en condiciones óptimas. Hay tres tipos de ordenaciones:

- Las no significativas.
- Las de significado limitado.
- Las altamente significativas.

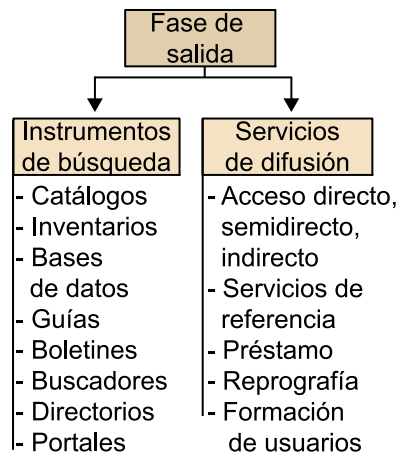
En cuanto al almacenaje y la conservación, la IFLA distingue tres actividades:

- La preservación.
- La conservación.
- La restauración.

Las causas que provocan la degradación pueden ser internas o externas al documento.

4. Fase de salida

La misión de la cadena documental es ofrecer el fondo documental de un SID al usuario. Para hacerlo pone a su disposición un conjunto de instrumento para la busca y unos servicios de difusión.



4.1. Instrumentos de búsqueda

Los SID ponen a disposición de sus usuarios una serie de fuentes bibliográficas (los inventarios, los catálogos, las guías de lectura, las bases de datos de sumarios o los boletines de nuevas adquisiciones) y de recursos de información en red. Todos ellos son fuentes de información secundaria (los buscadores, los directorios y los portales), ya que ofrecen la referencia documental y el resumen.

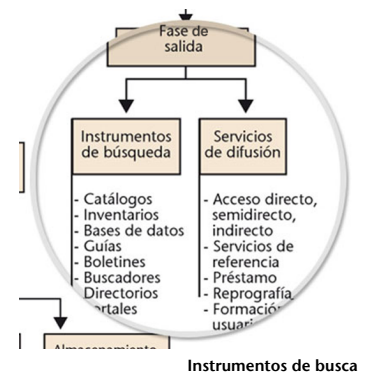
Fuentes primarias y fuentes secundarias

Hay que recordar que las **fuentes primarias** son aquellas que nos dan información nueva u original, es decir, de primera mano. Por otra parte, las **fuentes secundarias** son aquellas que tienen por finalidad indicar qué fuente o documento contiene o nos puede proporcionar la información final.

4.1.1. Fuentes bibliográficas para bibliotecas y centros de documentación

Dentro de las fuentes bibliográficas para bibliotecas y centros de documentación encontramos:

1) **Catálogos:** recopilan el fondo de una biblioteca, archivo o centro de documentación. Su función no es tan sólo identificar las obras, sino también localizarlas. Los catálogos colectivos son la suma de los catálogos de diversas



bibliotecas de la misma temática o red institucional con la finalidad de compartir registros y tareas como la catalogación colectiva y el préstamo interbibliotecario.

Catálogos

Consultad, por ejemplo, el catálogo de la Biblioteca de la UOC y el Catálogo Colectivo de las Universidades de Catalunya, al cual pertenece.

Otro catálogo colectivo es el Catálogo de Bibliotecas Universitarias de Rebiun.

2) **Bases de datos documentales:** los SID contratan en la fase de entrada bases de datos de artículos de revistas internacionales y nacionales, tanto especializadas como de cariz más general. Las bases de datos que se suscriben pueden ser referenciales o de texto completo.

Bases de datos documentales

Algunos ejemplos de bases de datos en documentación son:

- CSIC: ISOC - Biblioteconomía y Documentación.
- Dialnet.
- ISI Web of Knowledge (base de datos de acceso restringido, pero podéis acceder a través de la biblioteca del Campus de la UOC).

3) **Guías de lectura:** son recopilaciones de fuentes y recursos web sobre un tema particular. También se conocen como exposiciones virtuales o exposiciones bibliográficas. Las confeccionan las bibliotecas públicas y universitarias.

Guías de lectura

Algunos ejemplos de guías de lecturas son:

- Las del sistema de bibliotecas municipales de Girona.
- Guías de lectura de la Universidad Autónoma de Barcelona.
- Biblioteca Universitaria de Sabadell

4) **Boletines:** son recopilaciones de últimas informaciones sobre un tema concreto.

Boletines

Por ejemplo, el Observatorio SIC (Servicio de información y conocimiento) de ESADE es un boletín semanal que se envía por correo electrónico a la comunidad académica e investigadora, para informar sobre la actualidad empresarial y económica nacional e internacional.

4.1.2. Documentación de archivo

Dentro de la documentación de archivo encontramos:

1) **Catálogos:** en cuanto a los archivos, el catálogo tiene como objetivo describir la pieza o unidad documental, tanto si se trata de un documento suelto (pergamino, cédula real o fotografía) como de un expediente o una agrupación

Lectura complementaria

Podéis encontrar más información sobre la cuestión de los catálogos en:

R. Alberch i Fugueras (1999). *Gestió integral d'arxius*. Barcelona: Fundació per a la Universitat Oberta de Catalunya.

de documentos referidos al mismo asunto. Por la exhaustividad y la minuciosidad de la información, el catálogo se considera el instrumento culminante de toda tarea de descripción archivística (Alberch, 1999).

2) **Inventarios:** consiste en una descripción global y sistemática de las series documentales, entendidas como un conjunto de documentos relacionados desde el punto de vista de procesos, funcional u orgánico. Permite conocer las series más relevantes, su volumen y el alcance cronológico (Alberch, 1999).

3) **Bases de datos de gestión documental:** específicos para gestionar los documentos asociados a los expedientes resultantes de la actividad y los procesos específicos de la organización donde la base de datos está implementada.

4) **Guías:** la guía sirve para orientar al usuario sobre el conocimiento general del contenido de un archivo o de un conjunto de archivos (Alberch, 1999).

4.1.3. Instrumentos de búsqueda para recursos en línea

En cuanto a los instrumentos de búsqueda para recursos en línea, tenemos los siguientes:

1) **Buscadores:** son programas informáticos con un robot que rastrea la red de forma automática. Su misión es ir de una página en otra usando los hipervínculos y guardar la información en una base de datos. Actualmente hay dos grandes buscadores: Google y Yahoo, ya que han ido absorbiendo el resto de motores anteriores. Están diseñados para indizar y recuperar páginas HTML, pero sus prestaciones son cada vez más amplias, indizando documentos de audio, imagen, pdf, doc, archivos comprimidos y ejecutables, etc.

Términos sinónimos

Otros sinónimos de *buscadores* son motores de búsqueda, *search engines*, arañas (*spiders*), robots.

Los buscadores ofrecen los resultados de la búsqueda en un listado. Los sistemas anteriores a Google combinaban los factores de repetición de la palabra buscada y su posición en la página; a partir de Google y su *Pagerank*, las páginas mejor posicionadas son las que reciben mayor número de enlaces.

2) **Directorios:** como los buscadores y los portales, los directorios son instrumentos que ofrecen enlaces a recursos de Internet seleccionados por el SID, pero su particularidad es que los organizan en una clasificación temática. Los directorios están realizados por un equipo humano, lo que implica una selección de fuentes, identificación del contenido e indización más valiosa que la realizada por un robot, como en el caso de los buscadores. Por eso los directorios se conocen como productos de información de alto valor añadido.

Términos sinónimos

Otros sinónimos de *directorios* son índices temáticos, guías temáticas, *Internet subject gateways*.

Directorios

Ejemplos de directorios son las guías temáticas, como la de la Universitat de Barcelona o la de la Universidad Carlos III de Madrid.

3) Portales: son páginas de entrada a un conjunto de recursos y servicios electrónicos organizados para permitir al usuario acceder a un gran volumen de información. Un portal siempre es una página web, pero no a la inversa: no todas las páginas web son portales. Hay dos tipos de portales: los genéricos y los especializados. Entre estos últimos destacan los portales corporativos.

Portales

Algunos ejemplos de portales son:

- El portal genérico de Telefónica
- El portal especializado en documentación Internet library for librarians.
- El portal corporativo de la UOC

4.2. Servicios de difusión

Los SID ofrecen servicios de información colectivos y servicios personalizados. Algunos servicios requieren la presencia física del usuario y otros son consultables virtualmente.

Los servicios de difusión más destacados son:

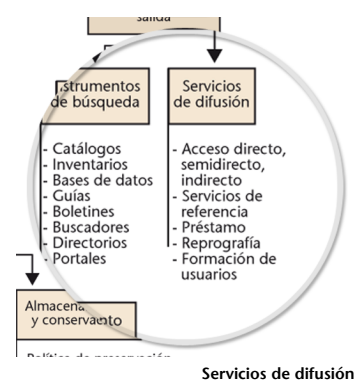
1) Acceso al fondo

Hay dos tipos de acceso al fondo:

a) **Acceso directo.** El usuario puede pasear, hojear y leer los documentos de la sala. Pasear por las salas y estanterías es un derecho adquirido en el ámbito de las bibliotecas públicas y universitarias. Acceder directamente al fondo tiene ventajas, como ver la cantidad de obras que hay de un determinado tema, poder hojear el sumario de la obra que nos llame la atención, consultar obras de temáticas afines que no conocíamos pero que hemos encontrado al lado de una conocida... El acceso libre comporta que el documentalista tiene que tener la colección ordenada de forma significativa, los materiales tienen que estar forrados porque su desgaste es superior y las estanterías tienen que estar debidamente señalizadas.

b) **Acceso indirecto.** El usuario no tiene acceso a los documentos, sólo a la colección de referencia. Tiene que consultar el catálogo o inventario y esperar en la sala de consulta a que le lleven la documentación. Este tipo de acceso se da en archivos. El motivo es la naturaleza delicada de la documentación, su valía o que está sujeta a algún tipo de restricción en la consulta.

2) Servicios de información y de referencia



Ved también

En el subapartado 3.2.2 de este módulo se tratan las diversas maneras de ordenar físicamente un fondo en las estanterías: ordenaciones no significativas, con significado limitado y altamente significativas.

Los servicios de información y de referencia son los servicios que prestan los SID para orientar y ayudar al usuario en la consulta del fondo, en la búsqueda, localización, selección e identificación de las fuentes de información y la información final. Son servicios personalizados. Los servicios más destacados son:

- El mostrador del bibliotecario/archivista al cual nos hemos dirigido para pedir información. En este caso hace falta la presencia física.
- Los servicios llamados de referencia o los iconos del tipo “Póngase en contacto con nosotros”, en los cuales podemos enviar un mensaje electrónico o iniciar un chat pidiendo la información.
- Los servicios de difusión selectiva de información (DSI¹⁴). La filosofía de este servicio es poner a disposición de los usuarios una clasificación de temas y que el usuario marque sus áreas de interés. Cuando el SID recibe una noticia o documento nuevo de este tema, automáticamente envía un mensaje al usuario. Si la información es en línea, el SID facilita la dirección del recurso web. Entre las informaciones ofrecidas hay las últimas novedades adquiridas, los sumarios de revistas previamente seleccionadas por el usuario, las últimas noticias sobre un tema, etc.

Servicios de difusión selectiva de información

La biblioteca de la UOC, por ejemplo, en el apartado Servicios nos ofrece diversos servicios: servicio de referencia (“La biblioteca respon”), servicio de obtención de documentos (SOD), sumarios (“Distribució de sumaris”) y difusión selectiva (“Butlletins de biblioteca”).

3) Préstamo y préstamo interbibliotecario

El **préstamo** es la entrega de un documento por un tiempo determinado y bajo ciertas condiciones. Se presta todo tipo de documentos a excepción de los documentos valiosos (manuscritos), ejemplares únicos de los cuales es difícil encontrar una copia, las obras de referencia y las publicaciones periódicas.

El **préstamo interbibliotecario** es la modalidad de préstamo que se da entre bibliotecas. Junto con los catálogos colectivos, son las modalidades más básicas de cooperación bibliotecaria.

4) Reprografía

La reprografía consiste básicamente en un servicio de fotocopias.

5) Formación de usuarios

Los SID elaboran guías de uso general, de consulta de su centro o sobre el funcionamiento de sus bases de datos.

Términos sinónimos

Otros términos sinónimos de *servicios de referencia* son servicio de referencia virtual (en inglés, *virtual reference services*; véase la Biblioteca Virtual de la UOC), servicio de referencia digital (en inglés, *digital reference services*; véase la sección “Preguntad al bibliotecario” de la Universidad de Barcelona), mostrador digital (*Desktop reference*) y Pregunte: las bibliotecas responden (*Ask-An-Expert*).

⁽¹⁴⁾DSI es la sigla de *Servicios de difusión selectiva de información*.

Obras de referencia

Ejemplos de obras de referencia son los anuarios, los atlas, los diccionarios, las enciclopedias, las guías, las estadísticas, los manuales, los tratados. Las obras de referencia se usan como material de consulta para obtener una información breve, rápida y exacta.

4.3. Conclusiones

La misión de la cadena documental es ofrecer el fondo documental de un SID al usuario. Para hacerlo pone a su disposición un conjunto de instrumentos para la búsqueda y unos servicios de difusión. Los instrumentos de búsqueda son:

- Los inventarios.
- Los catálogos.
- Los buscadores.
- Los directorios.
- Los portales.
- Las guías de lectura.
- Los boletines de sumarios y los boletines de nuevas adquisiciones.

Estos instrumentos son fuentes de información secundarias, ya que ofrecen la referencia documental y el resumen.

Los servicios de difusión más destacados son:

- El acceso al fondo.
- Los servicios de referencia
- El préstamo y el préstamo interbibliotecario.
- La reprografía.
- La formación de usuarios.

Formación de usuarios

- La biblioteca de la UOC, por ejemplo, elabora guías de uso de los diversos programas y bases de datos del centro, como el manual sobre el programa Gestor de referencias bibliográficas Refworks.
- La Universidad de Barcelona elabora tutoriales sobre biblioteconomía y documentación.
- La Universidad de Málaga también hace sus tutoriales.

5. Cadena documental y cooperación

A medida que el uso generalizado de normas para el análisis formal (ISBD, AACR2R, ISAD[G]) y el análisis de contenido (sistemas de clasificación, encabezamientos de materia, tesauros, etc.) se ha ido extendiendo, los SID han podido hablar un lenguaje común que ha hecho posible el intercambio entre ellos.

Hoy en día la normalización se ha convertido en una necesidad, no se concibe hablar de redes, cooperación (de creación de colecciones, de catalogación, de préstamo, de MARC 21), bibliotecas virtuales y accesos a recursos compartidos sin que existan normas que lo sistematicen.

La normalización posibilita la colaboración en forma de redes y consorcios donde las fases de la cadena documental se pueden trabajar de forma compartida.

Redes y consorcios

Algunos ejemplos de redes y consorcios:

- Red de archivos comarcales.
- Consorcio de bibliotecas universitarias catalanas.
- Sistema de lectura de bibliotecas públicas de Cataluña.
- Consorcio de bibliotecas universitarias de Rebiun.

La cooperación entre bibliotecas aparece a finales del siglo XIX en Estados Unidos y en los años noventa en España. Las bibliotecas cooperan entre ellas, pero además, como dice Miquel Térmens (2005, pág. 166) en una etapa más adelantada también se coordinan y colaboran.

Por **coordinación** entendemos que establecen acuerdos comunes sobre la misión y objetivos de los centros.

Por **colaboración** entendemos que llegan a crear estructuras formales de trabajo, incluso de cariz legal, que incluyen alguna forma de autoridad para planificar y operar un conjunto de misiones. Estas estructuras son los consorcios e implican un sentido de cooperación más profundo, siendo la adquisición de documentos una de las piezas clave.

En España el primer consorcio fue el Consorci de biblioteques universitàries de Catalunya (CBUC¹⁵), fundado en 1995. Su misión era crear un catálogo colectivo y hacer catalogación compartida. Más adelante, con el proyecto de la Biblioteca digital de Catalunya, también incorporó las adquisiciones de documentación electrónica.

Lectura complementaria

Sobre el tema de la cooperación de las bibliotecas podéis leer:

M. Térmens Graells (mayo-junio, 2005). "Los consorcios una nueva etapa de la cooperación bibliotecaria". *El profesional de la información* (vol. 14, n.º 3, págs. 166-173).

⁽¹⁵⁾CBUC es la sigla de *Consorci de biblioteques universitàries de Catalunya*.

Siguieron el mismo ejemplo los consorcios siguientes:

- Consorcio de universidades de la comunidad de Madrid y de la UNED para la cooperación bibliotecaria (Madroño).
- Consorcio de Bibliotecas Universitarias de Andalucía (CBUA¹⁶).
- Consorcio de Bibliotecas universitarias de Galicia (Bugalicia).

⁽¹⁶⁾CBUA es la sigla de *Consorcio de Bibliotecas Universitarias de Andalucía*.

Hay fondos, sin embargo, que no se pueden analizar de manera compartida, como es el caso de la documentación de archivo, que es única, y, en las colecciones bibliotecarias, la colección local, que es el fondo bibliográfico de la localidad en la que se encuentra el centro. Es especialmente importante en la biblioteca pública. Está constituido por monografías, diarios y revistas, programas de actos y fiestas, fotografías, etc., de la ciudad y sus habitantes. Es responsabilidad de la biblioteca (y no de la red) hacer la selección y análisis.

La cooperación favorece las tres fases de la cadena documental.

1) En la fase de entrada:

- La **selección**: la cooperación evita duplicidades de esfuerzos en la localización de novedades, así como en el apoyo técnico y en la compra de software de gestión. Cooperando se accede a fondos de financiación, para adquirir fondos o para migrar hacia sistemas electrónicos y en red.
- La **adquisición compartida**: es un procedimiento de adquisición en el que intervienen diversas bibliotecas para comprar de forma coordinada. Las ventajas son: mayor cobertura temática de las colecciones, reducción de gastos y tareas técnicas e intelectuales, mejor servicio al usuario.
- La **compra**: se proponen compras conjuntas para toda la red, pactos de mejora de las condiciones económicas para los contratos de paquetes documentales. Se consigue un cierto ahorro presupuestario gracias a las compras en conjunto y la economía de escala que representa.

El ahorro es relativo

El ahorro es relativo; no es que colaborando ahorren, sino que un SID que actúa en consorcio consigue mucho más por el mismo precio que actuando en solitario.

2) En la fase de análisis y tratamiento:

La catalogación compartida sólo se da en bibliotecas y centros de documentación. La cooperación entre bibliotecas permite que la catalogación se pueda llevar a cabo de forma centralizada y cooperativa:

- **Catalogación centralizada**: el proceso de catalogación lo realiza una biblioteca que distribuye las referencias documentales al resto de bibliotecas de la red. Implica que hay una autoridad central que asume la responsabilidad.

- **Catalogación cooperativa:** consiste en la aportación de referencias documentales por parte de diferentes bibliotecas que juntas crean un catálogo colectivo. En este caso, hay que corregir las desigualdades en la descripción bibliográfica.

3) En la fase de salida:

- **Instrumentos de búsqueda:** catálogos colectivos y repositorios institucionales, que multiplican exponencialmente el acceso a la documentación de otros SID. Permite pasar del concepto de documentación *en propiedad* al de *acceso* a la documentación.
- **Servicios de difusión:** préstamo interbibliotecario, ofrecer servicios de referencia en línea (*ask-an-expert*), más visibilidad social, llegan a más ciudadanos, amplían la cobertura informativa en diversos apoyos y el mantenimiento de la colección es más seguro.

6. Los lenguajes documentales dentro de la cadena

Esta asignatura se llama *Lenguajes documentales* y trata de los vocabularios de conceptos, libres o controlados, que usamos para indizar, es decir, para responder a la pregunta: ¿cuál es el tema de este documento? Además de la indicación hay otra operación que nos permite representar el contenido: son los resúmenes.

Durante el curso dedicaremos mucho más tiempo a la indicación que al resumen. Y es que para indizar hacen falta los lenguajes documentales.

En color se destaca el “itinerario conceptual” que sitúa los resúmenes y los lenguajes documentales dentro de la cadena documental:

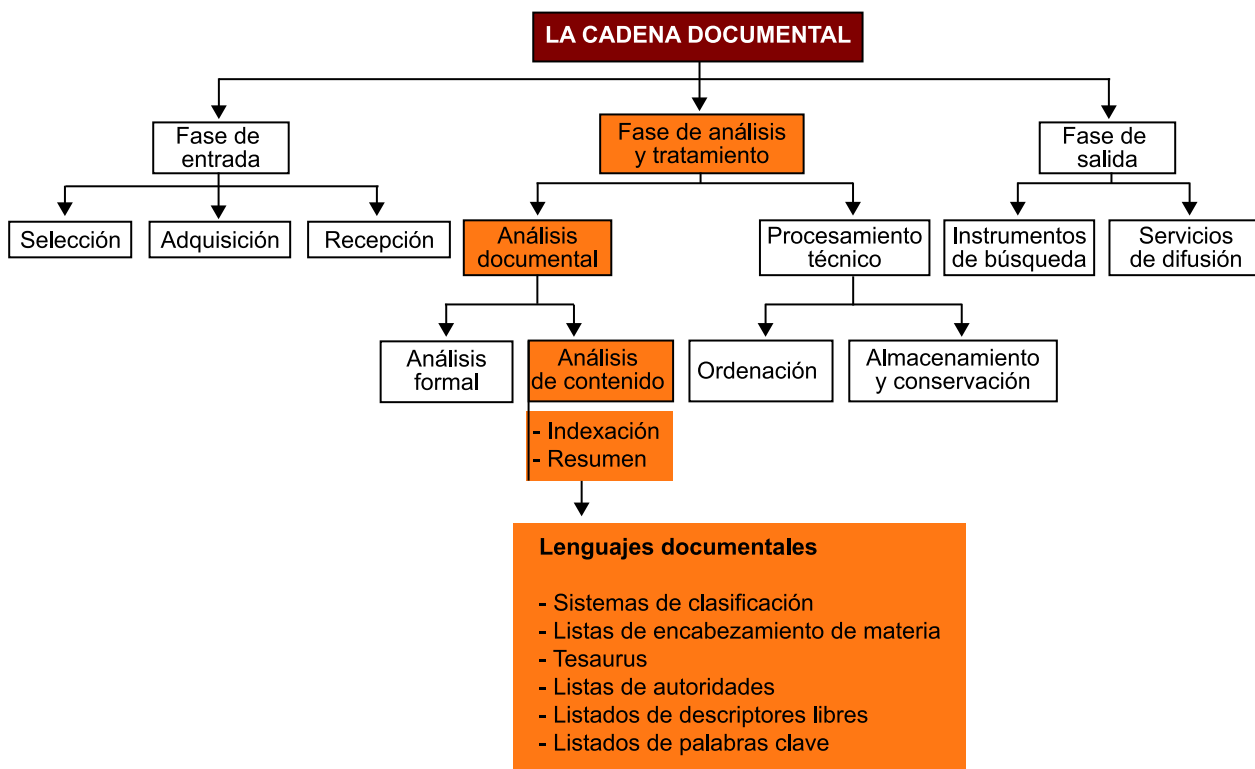
Ved también

La indicación se trata en los módulos “Sistemas de clasificación documental”, “Listas de encabezamientos de materia y listas de autoridades” y “Los tesauros”, mientras que el resumen se trata sólo en el módulo “Análisis de contenido: resumen e indicación”.

Observación

Estadísticamente este material dedica un 25% de los contenidos al marco general de la cadena y la fase de análisis de contenido; un 2% al resumen y el 73% a la indicación.

Localización de los lenguajes documentales en la cadena documental



Todas las operaciones y normativas se desarrollan en los módulos siguientes.

Actividades

Proponemos las siguientes actividades para practicar los conocimientos de este módulo:

1. Consultad la base de datos Temaria y localizad artículos donde se describa el proceso de la cadena documental en cualquier tipo de SID.
2. Imaginad que tenéis que redactar un artículo científico sobre el SID en el cual trabajáis. Redactad un esquema de la cadena documental como guión para describir las actividades de vuestro SID.
3. Localizad datos sobre la política de selección en webs de los archivos y bibliotecas de vuestro entorno.
4. Consultad las webs de la BC y la BNE y localizad información sobre los lenguajes documentales que usan en el análisis documental.
5. Consultad la web de una biblioteca pública o un archivo municipal. Observad los instrumentos de búsqueda disponibles y los servicios de difusión a vuestro alcance.
6. Consultad la base de datos CSIC-ISOC de Documentació. Observad las posibilidades de búsqueda.

Reflexión

Las soluciones de las actividades se tratarán en el tablón de la asignatura.

Glosario

acceso directo *m* Modalidad de acceso al fondo de la colección de un SID. El usuario puede pasear, hojear y leer los documentos de la sala. Pasear por las salas y estanterías es un derecho adquirido en el ámbito de las bibliotecas públicas y universitarias.

acceso indirecto *m* Modalidad de acceso al fondo de la colección de un SID. EL usuario no tiene acceso a los documentos, sólo a la colección de referencia. Tiene que consultar el catálogo o inventario y esperar en la sala de consulta a que le lleven la documentación. Este tipo de acceso se da en los archivos.

adquisición *f* Acción de incorporar documentos nuevos al SID. Las vías de adquisición son: gratuidad (*open access*), compra, donación, intercambio y depósito legal.

adquisición compartida *f* Procedimiento de adquisición en el que intervienen diversas bibliotecas para comprar de forma coordinada. Las ventajas son mayor cobertura temática de las colecciones, reducción de gastos y tareas técnicas e intelectuales, mejor servicio al usuario.

almacenamiento y conservación *m* Operaciones de la segunda fase de la cadena que procuran un espacio y unas medidas de protección al documento. La IFLA distingue tres actividades: la preservación, la conservación y la restauración.

análisis de contenido *m* Operaciones de análisis que identifican y representan de manera precisa la materia de los documentos, con el objetivo de permitir la recuperación. Las operaciones son dos: el resumen y la indización. Esta parte del análisis documental establece los puntos de acceso por materias.

análisis documental *m* Operación que analiza los atributos formales y de contenido de un documento, con técnicas específicas y normalizadas con la finalidad de hacerlos controlables y recuperables.

análisis formal *m* Descripción de los atributos estructurales de los documentos (autor, título, edición, colección, etc.), proporcionando una representación del documento que lo describa de manera única, sin ambigüedades. Esta parte del análisis documental establece los puntos de acceso por autor y por título.

análisis y tratamiento *m* Segunda fase de la cadena documental. Incluye dos operaciones: el análisis documental y el procesamiento técnico. Su función es organizar los documentos seleccionados en la primera fase y producir herramientas que faciliten la recuperación.

arañas *f* Ved Buscadores.

buscadores *m* Programas informáticos con un robot que rastrea de forma automática la red. Su misión es ir de una página a otra usando los hiperenlaces y guardar la información en una base de datos. Actualmente hay dos grandes buscadores, Google y Yahoo, ya que han ido absorbiendo el resto de motores anteriores.

cadena documental *f* Conjunto de operaciones y técnicas que comprenden las fases de entrada, análisis y tratamiento y difusión de los documentos, que se llevan a cabo en un centro de documentación o en una unidad administrativa o de trabajo.

catalogación *f* Ved Análisis formal.

catalogación centralizada *f* Catalogación realizada por una biblioteca que distribuye las referencias documentales al resto de bibliotecas de la red. Implica que hay una autoridad central que asume la responsabilidad.

catalogación cooperativa *f* Catalogación consistente en la aportación de referencias documentales por parte de diferentes bibliotecas que juntas crean un catálogo colectivo. En este caso, hay que corregir las desigualdades en la descripción bibliográfica.

catálogos *m* Fuente de información o instrumento que recopila el fondo de una biblioteca, archivo o centro de documentación. Su función no es tan sólo identificar las obras, sino también localizarlas. Los catálogos colectivos son la suma de los catálogos de diversas bibliotecas de la misma temática o red institucional con la finalidad de compartir registros y tareas como la catalogación colectiva y el préstamo interbibliotecario.

compra *f* Modalidad de adquisición. La compra puede ser directa o por suscripción al editor, la librería, el proveedor. Es la vía más habitual en bibliotecas y centros de documentación. Implica un conocimiento elevado de la producción bibliográfica, los canales de distribución y venta.

depósito legal (DL) *m* Modalidad de adquisición. El depósito legal (DL) es la exigencia impuesta por ley de depositar, en una o diversas agencias especificadas, ejemplares de las publicaciones de todo tipo, en cualquier soporte, por cualquier procedimiento de distribución (venta, alquiler...). El DL es un sistema de adquisición propio de un reducido número de bibliotecas, las nacionales. Los documentos susceptibles de DL son: libros, folletines, hojas impresas, publicaciones periódicas y seriadas, partituras musicales, grabados, mapas y planos, carteles, postales, naipes, diapositivas, producciones cinematográficas.

descartar *v* Ved expurgar.

descripción bibliográfica *f* Ved Análisis formal.

difusión selectiva de información (DSI) *f* Información que periódicamente se facilita a los usuarios y que previamente ha sido pactada; también se facilita bajo demanda.

directorios *m* Fuentes de información o instrumento que ofrecen enlaces a recursos de Internet seleccionados por el SID organizados en una clasificación temática. Los directorios están realizados por un equipo humano, lo cual implica una selección de fuentes, identificación del contenido e indización más valiosa que la realizada por un robot, como en el caso de los buscadores. Es por eso que los directorios se conocen como productos de información de alto valor añadido.

donación *f* Modalidad de adquisición. Es la cesión gratuita de documentos. Puede venir por parte del autor de la obra, por un particular o por la propia dinámica de la institución. Las donaciones pueden ser definitivas o temporales. Entre las donaciones podemos distinguir los legados y los depósitos. Los legados son donaciones provenientes de testamentos, están formados por colecciones particulares completas o parciales. Los depósitos son donaciones en las que la propiedad continúa siendo del donante.

expurgar *v* Evaluación crítica de la colección a partir de criterios materiales e intelectuales que nos permitan retirar de manera temporal o definitiva una parte del fondo con el fin de reorganizar los contenidos del SID y ofrecer un mejor servicio.

formación de usuarios *f* Conjunto de acciones que los SID llevan a cabo para ayudar a los usuarios en su búsqueda, consistentes en la elaboración de guías de uso general, de consulta del centro o sobre el funcionamiento de sus bases de datos.

guías de lectura *f* Recopilaciones de fuentes y recursos web sobre un tema particular. También se conocen como exposiciones virtuales o exposiciones bibliográficas. Las confeccionan las bibliotecas públicas y universitarias.

instrumentos de búsqueda *m* Fuentes de información secundaria que ofrecen la referencia y el resumen del documento original. Son instrumentos los inventarios, los catálogos, los buscadores, los directorios, los portales, las guías de lectura, los boletines de sumarios y los boletines de nuevas adquisiciones.

intercambio *m* Modalidad de adquisición. Como su nombre indica, consiste en el intercambio de material duplicado (o poco solicitado) entre dos centros.

International Standard Bibliographic Description (ISBD) *m* Normas internacionalmente admitidas para la descripción en el análisis formal.

inventarios *m* Fuentes de información e instrumentos de búsqueda. Descripción global y sistemática de las series documentales de archivo.

ISAD(G) *m* Norma internacional general de descripción archivística.

Machine Readable Catalog (MARC 21) *m* Formato que permite intercambiar registros entre diferentes bases de datos; sus normas codifican, almacenan e intercambian datos bibliográficos.

motores de búsqueda *m* Ved buscadores.

NODAC *f* Norma de Descripción Archivística de Cataluña.

Open Access Modalidad de adquisición. Se trata de toda aquella información dispuesta de forma gratuita y universalmente accesible vía Internet, como los repositorios institucionales, depósitos de archivos abiertos o los recolectores.

ordenaciones altamente significativas *f* Criterio de ordenación. Se ordena a partir de un cuadro de clasificación, de manera que los contenidos afines se colocan uno al lado del

otro. Es apropiado para SID de libre acceso, como las bibliotecas públicas y universitarias. Por ejemplo, útil para ordenar siguiendo las clases de la CDU (Clasificación decimal universal).

ordenaciones con significado limitado *f* Criterio de ordenación. Se ordenan, por ejemplo, por la lengua del documento, el autor, o el tema. Es la ordenación que encontramos en librerías, bibliotecas personales o bibliotecas pequeñas.

ordenaciones no significativas *f* Criterio de ordenación. Son aquellas en las que no hay relación con el contenido del documento. Por ejemplo, ordenar según el número asignado a la llegada. Es una ordenación útil en SID que no den acceso libre a las estanterías, y donde la recuperación la haga el documentalista. Apropiado para SID que tengan su colección en el depósito y no en la sala de lectura. Es el método más usado en archivos.

portales *m* Fuentes de información o instrumento. Los portales son páginas de entrada a un conjunto de recursos y servicios electrónicos organizados para permitir al usuario acceder a un gran volumen de información. Un portal siempre es una página web, pero no a la inversa: no todas las páginas web son portales.

préstamo *m* *Ved* préstamo interbibliotecario.

préstamo interbibliotecario *m* Entrega de un documento por un tiempo determinado y bajo ciertas condiciones. Se presta todo tipo de documentos a excepción de los documentos valiosos (manuscritos), ejemplares únicos de los cuales es difícil encontrar una copia, las obras de referencia y las publicaciones periódicas.

procesamiento técnico *m* Conjunto de tareas de ordenación, almacenamiento y medidas de conservación.

proceso documental *m* *Ved* cadena documental.

recepción *f* Proceso de la fase de entrada, que consta de tareas administrativas, el registro y la preparación del material.

referencia documental *f* Producto documental que contiene todos los datos identificativos y descriptivos del documento. La referencia documental es la representación condensada del documento original. Esta referencia se introduce en el catálogo y se convierte en la herramienta que representa el documento para futuras búsquedas.

referencias bibliográficas *f* Descripción del documento primario al que se remite al lector. En términos especializados es el documento que ha sido citado expresamente en un trabajo de investigación y es obligatorio, ética y textualmente, hacer alusión a él.

reprografía *f* Servicio de fotocopias.

robots *m* *Ved* buscadores.

search engine *m* *Ved* buscadores.

selección *f* Proceso en el que se decide qué documentos se añaden o eliminan de la colección. La selección es un proceso complejo en el que intervienen muchos factores: las necesidades del centro, las de sus usuarios, estadísticas de uso, recursos humanos y recursos económicos.

selección negativa *f* *Ved* expurgar.

servicios de difusión *pl* Servicios de la fase de salida puestos a disposición de los usuarios para hacer difusión de la colección, por ejemplo el acceso al fondo, servicios de referencia, préstamo, reprografía y formación de usuarios.

servicios de referencia *m* Servicios que prestan los SID para orientar y ayudar al usuario en la consulta del fondo, en la búsqueda, la localización, la selección, la identificación de las fuentes de información y la información final.

SID *m* Servicio de información y documentación.

signatura topográfica *f* Código que identifica cada material en las salas y estanterías del SID. Es el conjunto de números, letras y símbolos que localiza y ordena físicamente el documento en la estantería.

salida *f* Última fase de la cadena que pone a disposición de los usuarios un conjunto de instrumentos para la búsqueda y unos servicios de difusión.

spiders *fpl* *Ved* buscadores.

transferencia *f* Traslado del conjunto de documentos de la unidad administrativa que los ha producido a un servicio de archivos con plenos derechos para su tratamiento y disposición.

Bibliografia

Alberch i Fugueras, R. (1999). *Gestió integral d'arxius*. Barcelona: Fundació per a la Universitat Oberta de Catalunya.

Allo Manero, A. (2002). "Conservación de la colección". En: Luisa Orera Orera (ed.). *Manual de Biblioteconomía*. Madrid: Síntesis (pág. 181-200).

Asociación Española de Normalización y Certificación (1994). *Referencias bibliográficas. Contenido, forma y estructura*. UNE 50-104:1994. Madrid: AENOR.

Ayuso García, M.ª D. y Martínez Navarro, V. (2006). "Evaluación de calidad de fuentes y recursos digitales: guía de buenas prácticas". *Anales de Documentación*, núm. 9, pág. 17-42.

Biblioteca UOC (2005). *Criteris de selecció de recursos web gratuïts. Document de treball*. Barcelona (revisión marzo del 2006).

Borgoños, M.ª D. (1999). *Análisis de redacción de referencias y citas bibliográficas en literatura científica, a través del estudio de la normalización bibliográfica*. Tesis doctoral. Directora Celia Chaín Navarro. Universidad de Murcia, Departamento de literatura española y teoría de la literatura.

Carrizo Sainero, G. et al. (2000). *Manual de fuentes de información*. Madrid: CEGAL.

Cid Leal, P. (1999). *Fonaments de llenguatges documentals*. Barcelona: Fundació per la Universitat Oberta de Catalunya.

Codina, Ll. (2000). "Evaluación de recursos digitales en línea: conceptos, indicadores y métodos". *Revista Española de Documentación Científica*, vol. 23, pág. 9-44.

Fuentes, J. J. (2005). "Bibliotecas nacionales y consorcios bibliotecarios". *Jornadas sobre bibliotecas nacionales. Las bibliotecas nacionales del siglo XXI*. Valencia.

García Gutiérrez, A. L. (1984). *Lingüística documental*. Barcelona: Mitre.

García Rodríguez, A. (2002). "Servicios de lectura en sala y préstamo". En: Luisa Orera Orera (ed.). *Manual de Biblioteconomía*. Madrid: Síntesis (pág. 247-263).

García Valenzuela, H. (2002). "El soporte de la biblioteca". En: Luisa Orera Orera (ed.). *Manual de Biblioteconomía*. Madrid: Síntesis (pág. 201-214).

ISAD(G): Norma internacional general de descripció arxivística: adoptada per la Comissió Ad Hoc de Normes de Descripció, Suècia, 21-23 de gener de 1993: versió final aprovada pel CIA. Barcelona: Associació d'Arxivers de Catalunya: Departament de Cultura de la Generalitat de Catalunya, 1995. ed. orig.: Ottawa: CIA, 1994. ISBN 84-605-4625-X.

López Hernández, M. A. (1996). "La selección documental". *Revista general de información y documentación*. Madrid.

Maniez, J. (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.

Melero, R.; Abad, M. F. (2008). "Revistes d'accés obert: característiques, models i tendències". *BID Textos universitaris de biblioteconomia i documentació* (juny). Barcelona.

Orera Orera, L. (2002). "La colección, formación, desarrollo y mantenimiento" En: Luisa Orera Orera (ed.). *Manual de Biblioteconomía*. Madrid: Síntesis (pág. 92-112).

Pacios Lozano, R. (1995). "El control de la gestión de las adquisiciones". *Revista general de información y documentación*, vol 5, Núm. 2. Madrid.

Termens Graells, M. (2005). "Los consorcios una nueva etapa de la cooperación bibliotecaria". *El profesional de la información* (mayo-junio), vol. 14, núm. 3, pág. 166-173.

Turiel, J. (2006). "Recursos web per a les adquisicions". *BID Textos universitaris de biblioteconomia i documentació*, núm. 16 (juny). Barcelona.

Vall Casa, A. (2006). "Esporgar: per què, com i quan". *BID Textos universitaris de biblioteconomia i documentació*, núm. 16 (juny). Barcelona.

Vilchez Pardo, J. (2002). "Tratamiento y ubicación de la colección" En: Luisa Orera Orera (ed.). *Manual de Biblioteconomía*. Madrid: Síntesis (pág. 113-135).

Análisis de contenido: resumen e indización

Manela Juncà Campdepadrós

PID_00143964



Universitat Oberta
de Catalunya

www.uoc.edu



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. Fundació 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

Introducción	5
Objetivos	7
1. El análisis de contenido	9
2. El resumen	11
2.1. Tipos de resúmenes	14
2.2. Resumen automático	16
3. La indización	20
3.1. Lenguaje natural y lenguaje documental	20
3.1.1. Número de términos	21
3.1.2. Control de las formas	22
3.1.3. Control del significado	22
3.1.4. Relaciones de significado de los términos	24
3.2. ¿Cómo se indiza?	26
3.3. Calidad y coherencia de la indización	31
4. Los lenguajes documentales	33
4.1. Los términos de indización	33
4.2. Evolución histórica de los lenguajes documentales	35
4.3. ¿Cuándo son necesarios los lenguajes documentales?	39
4.4. Complementariedad de los lenguajes documentales	42
5. Tipología de los lenguajes documentales	44
5.1. Naturaleza del término: codificado o natural	44
5.2. Nivel de control: libre o controlado	45
5.3. Nivel de coordinación: precoordinado o postcoordinado	46
5.4. Estructura: jerárquica o combinatoria	48
5.5. Nivel de análisis: materias, conceptos, palabras clave	50
5.6. Conclusiones	52
Actividades	53
Glosario	54
Bibliografía	58

Introducción

Este módulo os introduce en los procesos documentales de la segunda fase de la cadena documental, llamada **análisis de contenido**, formada por el resumen y la indización.

Itinerario de estudio

El módulo empieza con un capítulo dedicado al análisis de contenido, para situar al estudiante en las dos operaciones mencionadas, el resumen y la indización.

El apartado dedicado al **resumen** está diseñado para responder a las preguntas de qué es un resumen, quién lo redacta, qué utilidades tiene y cuántos tipos de resúmenes hay. Finalmente, se presentan los resúmenes automáticos, explicando su evolución y funcionamiento.

La **indización** es el grueso de esta asignatura y en este módulo tiene tres apartados. El primero de ellos trata de dar respuesta a las preguntas de qué es indizar, quién indiza, por qué hacen falta los lenguajes documentales y cómo se indiza. El apartado titulado “Lenguajes documentales” responde a las preguntas de qué son los lenguajes, cuántos hay, qué son los términos de indización, cómo han evolucionado, cuándo son necesarios y cuál es su uso en solitario o combinados. El último apartado, titulado “Tipología”, trata de los diferentes criterios usados para clasificar los lenguajes.

Este es un módulo básico para el aprendizaje de la terminología que se usará en el resto de módulos.

Conceptos más importantes

Concepto	Ved
Resumen informativo Resumen indicativo Resumen selectivo Resumen automático	1. El resumen
Ambigüedad Lenguaje natural Exhaustividad Especificidad Traducción Univocidad	2. La indización

Concepto	Ved
Lenguaje documental Sistemas de clasificación Listados de encabezamientos de materia Listados de autoridades Tesauros Listados de descriptores libres Listados de palabras clave Notación Encabezamiento Descriptor Identificador o autoridad Palabra clave	4. Los lenguajes documentales
Codificado Natural Libre Controlado Precoordinado Postcoordinado Jerárquico Combinatorio Materias Conceptos Palabras clave	5. Tipología de los lenguajes documentales

Objetivos

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

En cuanto al **resumen**:

1. Aprender a hacer resúmenes de manera intelectual: resúmenes informativos, indicativos y selectivos.
2. Aprender a hacer resúmenes con programas de resúmenes automáticos.

En cuanto a la **indización**:

1. Analizar los factores necesarios para que haya una buena comunicación documental: entender los problemas del lenguaje natural y la función de los lenguajes documentales dentro de esta comunicación.
2. Conocer los procesos de indización: examen del documento, selección y traducción.

En cuanto a los **lenguajes documentales**:

1. Conocer las características principales de los lenguajes documentales.
2. Conocer la evolución histórica de los lenguajes documentales.
3. Aprender a distinguir y saber utilizar la diferente tipología de los lenguajes documentales: sintéticos-analíticos, precoordinados-postcoordinados, controlados-libres, jerárquicos-combinatorios, materias-conceptos-palabras clave.

1. El análisis de contenido

El **análisis de contenido** se sitúa en la segunda fase de la cadena documental y reúne todo el conjunto de operaciones destinadas a representar la materia de los documentos para una posterior recuperación.

Son tareas de cariz intelectual en las que la formación y la habilidad del analista juegan un papel importante.

“Representar la materia” o “describir el contenido” es responder a la pregunta: “¿cuál es el tema de un documento?”.

Para representar el contenido de un documento el analista tiene que llevar a cabo dos operaciones:

- 1) El **resumen**, que condensa el contenido en un texto más breve y manejable.
- 2) La **indización**, que identifica los conceptos o temas principales. También se conoce como descripción característica.

Estas dos operaciones admiten una elaboración humana o automática. Por lo tanto, habrá resúmenes elaborados por documentalistas y resúmenes elaborados por programas, y también indizaciones hechas por analistas e indizaciones elaboradas por un software.

Operaciones humanas y automatizadas

	Humano	Automatizado
Resumen	Resumen informativo Resumen indicativo Resumen selectivo	Resumen automático
Indización	Sistemas de clasificación Listas de encabezamientos de materia Listados de autoridades Tesauros Listados de descriptores libres	Listado de palabras clave

Los dos sistemas tienen ventajas e inconvenientes. La calidad y coherencia que aporta un documentalista supera en estos momentos la que ofrecen los programas informáticos, pero en cambio los sistemas automáticos son instantáneos, baratos y capaces de asumir ingentes cantidades de documentos.

La rama científica que estudia cómo emular el conocimiento humano, en cuanto a la identificación de los conceptos y las frases con contenido relevante para el resumen y la indización, es el procesamiento en lenguaje natural.

El procesamiento en lenguaje natural (PLN¹) es una rama de la inteligencia artificial y de la lingüística computacional que estudia los lenguajes que usan los humanos para interactuar con los ordenadores en contextos escritos y orales.

A modo de conclusión

Para representar o describir el contenido de un documento el analista tiene que llevar a cabo dos operaciones:

- El resumen, que condensa el contenido en un texto más breve y manejable.
- La indización, que identifica los conceptos o temas principales. También se conoce como descripción característica.

Las dos operaciones se pueden llevar a cabo de manera humana o automática.

Ved también

Trataremos el procesamiento del lenguaje natural en el subapartado 2.4 y en el apartado 3.

⁽¹⁾PLN es la sigla de *procesamiento en lenguaje natural*.

Lectura complementaria

I. Gil Leiva; J. V. Rodríguez Muñoz (1996). "El procesamiento del lenguaje natural aplicado al análisis del contenido de los documentos". *Revista general de información y documentación* (vol. 6, núm. 2, pág. 205-218).

2. El resumen

Según la norma UNE 50-103-90 *Preparación de resúmenes*, un **resumen** es la presentación abreviada y precisa de un documento, sin interpretación ni crítica y sin mención expresa del autor del resumen.

Ved también

Encontraréis la norma UNO 50-103-90 en el espacio "Materiales y fuentes" de las aulas.

Cuando decimos documento nos estamos refiriendo a todo tipo de documento, sea cual sea su soporte material. Podemos resumir un texto, la imagen de una fotografía, un vídeo, audios, información en línea o hipertextos.

Los resúmenes, como la indización, pueden ser de elaboración humana o automática. En el primer caso hay cuatro tipos de personas que pueden redactar un resumen. En el caso de los resúmenes automáticos, se trata de un software.

1) Resumen humano:

a) El **autor** del documento. Los resúmenes elaborados por los propios autores son muy habituales en el mundo de las comunicaciones científicas y tecnológicas.

b) Un **especialista** en la materia de la que trata el documento.

c) La **editorial**. Son los resúmenes que aparecen en la contraportada de los libros impresos y que tienen una función claramente publicitaria.

d) Un **profesional de la documentación**. Aporta su conocimiento sobre la redacción de buenos resúmenes y los elabora pensando en las utilidades futuras.

2) **Resumen automático**: los programas se conocen como programas resumidores de textos o *Automatic Text Summarizer*.

La norma internacional ISO 214:1976, traducida por AENOR como norma UNE 50-103-90 *Preparación de resúmenes*, establece las directrices que se tienen que seguir para presentar los resúmenes en los documentos. Pone especial énfasis en la preparación de resúmenes por parte de los autores de los documentos primarios y en la misma publicación.

Resúmenes para revistas

Las revistas acostumbran a dar directrices a sus autores para la elaboración de resúmenes. Ved, por ejemplo, la revista *EPI* en su apartado "Instrucciones para los autores".

Programas resumidores de textos

Un ejemplo de programas resumidores de textos es *Swe-sum*, que hace un análisis estadístico del texto y elabora el resumen con los fragmentos que contienen las palabras más ponderadas (más repetidas pero con significado).

Redactar un resumen es fácil. Lo difícil es redactar un buen resumen. El punto de inflexión es la calidad del resumen, que lo hará más o menos útil en un sistema documental. Un resumen propagandístico no aportará muchos conceptos principales para indizar, aunque haya sido un buen reclamo para las ventas.

Ejemplo de resumen elaborado por la editorial con finalidad publicitaria

SAGAN, Carl. *Cosmos*. Traducció: Albert Santamaria i Martínez; pròleg: Ricard Guerrero. Barcelona: Publicacions i Edicions de la Universitat de Barcelona: Omnis Cellula, cop. 2006.

“He aquí una de las obras más destacadas de la literatura internacional de divulgación científica, publicada por primera vez en catalán. Una obra imprescindible de uno de los grandes maestros de la divulgación, que nos introduce en los grandes enigmas que la humanidad ha tratado de entender y explicar desde tiempos inmemoriales, y por los cuales ha nacido lo que llamamos ciencia.

Desde la infinitud del Universo hasta el mundo invisible de los átomos, desde el nacimiento de las estrellas hasta la aparición de la vida, Carl Sagan consigue transmitir los conocimientos de la ciencia actual de una manera clara y apasionante.”

Para un analista sólo tendría utilidad el último párrafo, en qué aparecen términos como *universo, átomos, estrellas, vida*.

El resumen es útil en dos fases de la cadena, en los procesos de selección y adquisición que se da en la primera fase de la cadena y en la fase de salida, donde es un excelente instrumento de recuperación, ya que el resumen ofrece más datos que la simple referencia documental. La principal utilidad del resumen es la de difundir la información.

Difundir la información

Cada vez más bases de datos referenciales ofrecen el resumen de sus monografías y revistas, como por ejemplo Ebsco, Dialnet, Compludoc, CBUC, Eric database o *ISI current contents connect*. También lo hacen las bases de datos de novedades editoriales, por ejemplo la editorial Trea (recomendamos el acceso desde la biblioteca de la UOC).

En todos los casos es indudable el valor informativo que aporta el resumen para difundir el contenido del documento de la colección. Pero además, el resumen tiene otras utilidades, tal como dice la norma UNE 50-103-90:

a) Determinar la pertenencia: un resumen bien elaborado capacita a los lectores para identificar de forma rápida y precisa el contenido de un documento y decidir si hay que leerlo en su totalidad.

b) Evitar la lectura del texto completo en documentos de interés secundario. Un resumen bien elaborado proporciona suficiente información sobre temas que no sean de interés principal para el lector. Ahorra tiempo al usuario.

c) Ayudar en la búsqueda automatizada. Los resúmenes automatizados incorporados en los catálogos son muy útiles para:

- Extraer términos de indización de su texto, es decir, indizar a partir del resumen.
- Hacer búsquedas de palabras clave que no se encuentran en el título.

- Servir de control bibliométrico, al comparar los términos usados en una ecuación de búsqueda con los términos que aparecen en un resumen y así establecer la pertinencia de la recuperación.
- Ayudar a la difusión desde los servicios de alerta.

Según María Pinto (1992), las **características de un resumen** son las siguientes:

- Brevedad. Se tienen que omitir datos preliminares o temas del conocimiento común.
- Pertinencia. El resumen se tiene que adecuar al mensaje principal del documento, sin obviar o interpretar los datos.
- Claridad y coherencia. Frases completas, dotadas de coherencia lineal y global.
- Profundidad. Varía en función del tipo de resumen o de los diferentes niveles de detalle que se persigan.
- Consistencia lingüística. Un resumen se tiene que adaptar a las pautas lingüísticas en uso y tiene que tener en cuenta las reglas morfológicas y sintácticas correspondientes.
- Proximidad cronológica entre las ediciones del documento original y el resumen. Es importante que el tiempo transcurrido entre la publicación del original y el resumen no sea excesivo, especialmente en ámbitos científicos y técnicos.

A modo de conclusión

- El resumen es la presentación abreviada y precisa de un documento, sin interpretación ni crítica y sin mención expresa del autor del resumen.
- El resumen puede ser redactado por el autor del documento, un especialista en la materia, la editorial, un documentalista o un programa informático.
- El resumen es útil en dos fases de la cadena: en los procesos de selección y adquisición que se da en la primera fase de la cadena y en la fase de salida, donde es un excelente instrumento de recuperación.
- La principal utilidad del resumen es la de difundir la información, pero además, el resumen tiene otras utilidades, como determinar la pertinencia, evitar la lectura del texto completo en documentos marginales y ayudar a la búsqueda automatizada.
- Los resúmenes automatizados incorporados en los catálogos son muy útiles para extraer términos de indización del texto, para hacer búsquedas de palabras clave que no se encuentran en el título, para servir de control bibliométrico y ayudar a la difusión a través de los servicios de alerta.

Lectura complementaria

Podéis ampliar la información sobre el resumen leyendo la obra siguiente:

M. Pinto Batanea (1992). *El resumen documental: principios y métodos*. Madrid: Pirámide/Fundación Germán Sánchez Ruijérez (Biblioteca del Libro, Y).

2.1. Tipos de resúmenes

Hay diversos tipos de resúmenes, según el tamaño, los usuarios y la profundización en el contenido. Los tipos más habituales son los resúmenes informativos, indicativos y selectivos.

1) Resumen informativo

Redactaremos el tema central, temas adicionales, naturaleza y objetivo del documento, metodología, resultados, conclusiones y anexos. La idea de fondo es que un resumen informativo puede sustituir en ocasiones la lectura del documento original. La norma UNE 50-103-90 recomienda que el esquema a seguir sea el de:

objetivo + metodología + resultados (o conclusiones)

Sin embargo, no hay que seguir forzosamente este orden, ya que hay entornos, como el técnico científico, donde se prefieren los resúmenes orientados a los resultados (para que la discriminación sea más rápida).

En cuanto al tamaño del resumen, la norma da pautas pero advirtiéndole que el contenido del documento es más significativo que las pautas para determinar la extensión del resumen. De todas maneras la norma nos sugiere:

- Monografías, informes, tesis: 500 palabras.
- Artículos de revista, capítulos de monografías: 250 palabras.
- Comunicaciones breves: 100 palabras.

Ejemplo de resumen informativo

CONSUEGRA FERNÁNDEZ, Jesús: "El Ajedrez: evolución y claves de un juego milenario". En *Mundo antiguo*. Madrid: 2002. n.º 3-4, año 1, p. 60-61.

"Artículo divulgativo sobre el juego del ajedrez, estructurado según sus orígenes, antigüedad, expansión, variantes y simbolismo.

El origen del ajedrez es hindú y el primer representante conocido es el Ghaturanga, aparecido entre el 3000 y el 2000 a.C. en Sri Lanka, aunque no aparece documentado hasta el siglo VII d.C.

Del Ghaturanga proceden en cascada las diferentes variantes del ajedrez: de la India viajó a Persia en el siglo VI d.C., donde pasó de los 4 jugadores originales a 2 en la versión persa Shatranj. Desde Persia se extendió hacia Occidente y hacia Oriente.

Hacia Occidente: paralela a la expansión árabe, el juego llega a la Península Ibérica durante la Alta Edad Media, y desde aquí se expande al resto de Europa y al resto del mundo en la época de las colonizaciones.

Hacia Oriente: en la China, en el s. VII d.C., el ajedrez toma la forma del ajedrez chino Xiang qi; en el Japón, el Shogi; en Indochina, el ajedrez birmano y tailandés. Tanto en Oriente como en Occidente, el ajedrez presenta innumerables variaciones locales.

El tablero y las fichas parecen poseer un significado simbólico. El tablero, con la alternancia de casillas blancas y negras, forma un mandala. El simbolismo de las fichas es menos esotérico y ha ido cambiando según los tiempos: obispos, elefantes, etc.

El autor concluye que el ajedrez, además de un juego, es una herramienta educativa de primer orden, casi una ciencia.”

Como podéis comprobar, este resumen tiene 237 palabras.

2) Resumen indicativo

Redactaremos sólo las ideas centrales del documento. Su lectura no puede sustituir la lectura del original. Como su nombre sugiere, el resumen indicativo presenta de forma abreviada y muy sintética el contenido o la tipología del documento. Su extensión puede oscilar entre una frase o 4 líneas de texto.

Ejemplo de resumen indicativo

CONSUEGRA FERNÁNDEZ, Jesús: “El Ajedrez: evolución y claves de un juego milenario”. En *Mundo antiguo*. Madrid: 2002. n° 3-4, año 1, p. 60-61.

“Artículo divulgativo sobre el juego del ajedrez, trata de su origen hindú, antigüedad, expansión histórica tanto en Oriente como en Occidente, variantes nacionales y simbolismo del tablero y las fichas.”

3) Resumen selectivo

Redactaremos sólo una parte concreta del documento. El más habitual es el resumen de conclusiones, pero también hay otros tipos, como la reseña (*review*), que es un análisis del documento con elementos críticos. Este tipo de resumen se adapta muy bien a las necesidades de los usuarios, por ejemplo investigadores o técnicos que necesitan un dato muy concreto sobre el objetivo del documento o las conclusiones a las que llega.

Ejemplo de resumen selectivo

CONSUEGRA FERNÁNDEZ, Jesús: “El Ajedrez: evolución y claves de un juego milenario”. En *Mundo antiguo*. Madrid: 2002. n° 3-4, año 1, p. 60-61.

“El ajedrez, además de un juego, es una herramienta educativa de primer orden, casi una ciencia.”

A modo de conclusión

Los resúmenes más habituales son el resumen informativo, el indicativo y el selectivo:

- El **resumen informativo** consigna el tema central, temas adicionales, naturaleza y objetivo del documento, metodología, resultados, conclusiones y anexos. La idea de fondo es que un resumen informativo puede sustituir en ocasiones a la lectura del documento original.
- El **resumen indicativo** consigna sólo las ideas centrales del documento. Su lectura no puede sustituir a la lectura del original.
- El **resumen selectivo** consigna sólo una parte concreta del documento. El más habitual es el resumen de conclusiones, pero también hay otros tipos, como la reseña (*review*).

2.2. Resumen automático

Una de las necesidades más perentorias ante el aumento de información digital debido al crecimiento exponencial de Internet es manejar y filtrar el gran volumen de información. Una de las soluciones aportadas por el PLN han sido los programas de resumen automático, que actúan sobre textos, imágenes, webs y correo electrónico.

Los primeros en trabajar en el campo de la automatización de los resúmenes fueron Hans Peter Luhn en el año 1958 y Edmundson en 1969, que aplicaron técnicas como la frecuencia de las palabras, o la posición de una frase dentro de un documento para redactar resúmenes sin intervención humana.

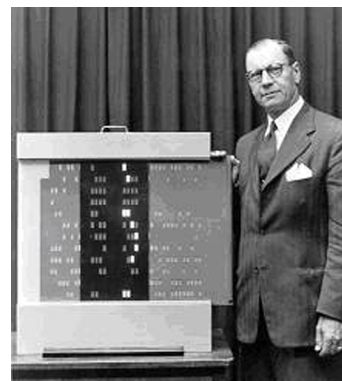
A partir de estas primeras investigaciones se han perfeccionado muchas técnicas diferentes basadas en conocimiento y recursos lingüísticos (como las de Lin y Hovy, 2002; Gotti *et al.*, 2007) o basadas en métodos estadísticos y de aprendizaje automático (Hirao *et al.*, 2002; Svore, 2007) (autores citados en Lloret *et al.*, 2008; y Mateo *et al.*, 2003).

Últimamente las investigaciones giran en torno al resumen multidocumento, es decir, resumir más de un documento (Goldstein *et al.*, 2000; Qiu, 2007; Huo y Chen, 2008) de contenidos afines o redundantes (autores citados en Lloret *et al.*, 2008; y Mateo *et al.*, 2003).

Los resúmenes automáticos se conocen también como *extracts*. La terminología anglosajona diferencia así los *extracts* y los *abstracts*. Los *extracts* son los resúmenes formados a partir de la extracción de algunas frases del texto previamente seleccionadas por un programa, mientras que los *abstracts* son los resúmenes elaborados por una persona.

La base de todas las técnicas de funcionamiento de un programa de resúmenes automático es el cómputo de la frecuencia de las palabras.

Hay diversas herramientas para hacer estos cálculos, por ejemplo WVTool. Se trata de contar cuántas veces sale una palabra no vacía en el texto.



Hans Peter Luhn

Lecturas complementarias

Podéis consultar los resultados de las investigaciones de estos autores en los artículos siguientes:

E. Lloret; O. Ferrández; R. Muñoz; M. Palomar (2008). "Integración del reconocimiento de la impliación textual en tareas automáticas de resúmenes de textos". *Procesamiento del lenguaje natural*, n.º. 41, pág. 183-190.

P. L. Mateo; J. C. González; J. Villena; J. L. Martínez (2003). Un sistema para resumen automático de textos en castellano.

Ved también

Encontraréis una explicación detallada sobre las palabras vacías en el módulo "Indización automática y descriptores libres".

Ejemplo de funcionamiento de un programa de resúmenes automático (extraído de Lloret *et al.*, 2008)

“Tropical storm Gilbert formed in the eastern Caribbean and strengthened into a hurricane Saturday night. There were no reports of casualties.”

Oración 1:	Tropical (2) storm (6) Gilbert (7) formed (1) in (0) the (0) eastern (1) Caribbean (1) and (0) strengthened (1) into (0) a (0) hurricane (7) Saturday (4) night (2).
Oración 2:	There (0) were (0) no (0) reports (1) of (0) casualties (1).

Lo primero que vemos es que las palabras vacías, es decir, las palabras que no tienen significado (preposiciones, artículos, verbos) no se computan.

Al lado de cada palabra con significado vemos el número de veces que sale en todo el texto. Se suman los valores, de manera que la oración 1 tiene 3,2 puntos y la oración 2, 0,2. El programa seleccionará la frase 1 como más representativa para el resumen automático.

Este sistema de resumir a partir de las frases con las palabras más significativas en el texto parece simplista pero tiene cierta justificación. Según Kupiec *et al.* (1995) aproximadamente el 80% de las frases en resúmenes humanos están copiadas literalmente o con pequeñas modificaciones del texto original.

A partir de esta base estadística se incorporan otras técnicas para dotar al programa de más conocimiento y paliar la escasa coherencia del resultado, como puede ser, por ejemplo, la resolución de la anáfora o aplicar programas (por ejemplo, WordNet) que proporcionen relaciones como las de sinonimia o hiperonimia, o mecanismos para detectar y eliminar la redundancia.

Definimos brevemente qué son las anáforas y la hiperonimia:

a) Las **anáforas** son la relación de referencia entre un elemento lingüístico y otro anterior en el discurso.

b) Decimos que una palabra es **hiperónima** cuando tiene un campo significativo que incluye otro de menor extensión.

Los expertos consideran que la tecnología actual no tiene problemas para detectar las frases con más significado, pero sí para ordenarlas según su importancia.

Los programas funcionan a grandes rasgos de la siguiente manera: se copia el texto a resumir o bien se escribe la dirección del documento. Se escoge el tipo de documento (académico, periodístico, etc.) y el tanto por ciento de reducción del texto.

A continuación tenéis unos cuantos programas de los más conocidos:

Anáfora

“El Salón del Hobby ha tenido más de 60.000 visitantes este año. Este salón se ha convertido en la feria de ocio familiar más visitada”.

En este ejemplo, la anáfora se da en “este salón”, que hace referencia al Salón del Hobby, expresado en la frase anterior. Como se puede comprobar, si en el resumen automático aparece sólo la segunda frase, el lector no sabrá a qué salón hace referencia.

Hiperonimia

Color es un hiperónimo. Su contrario es hipónimo: *amarillo*, *naranja*, *verde* son hipónimos.

- Connexor
- Daedalus
- Extractor
- FociSum
- InTEXT (Dynamic Summarizing)
- Inxight Summarizer
- IslandInText
- K-Site de Daedalus
- Pertinence Summarizer
- Sinope Summarizer
- Summarizer
- SweSum²
- System Q
- TextAnalyst
- Trestle

⁽²⁾Podéis practicar con el programa Swesum, que es gratuito y traduce al español.

El programa K-Site de Daedalus

De entre los programas de resumen automático mencionados, veamos el funcionamiento del programa K-Site de Daedalus. Este programa tiene cinco módulos:

- **Módulo 1: Análisis morfosintáctico.** En este módulo se determina la categoría léxica de cada palabra: sustantivo, verbo, adjetivo, artículo, preposición, etc. También se determina el lema. Estas operaciones permiten distinguir las palabras con significado (sustantivos, adjetivos, verbos) de las vacías (artículos, preposiciones, pronombres, etc.). El lema permite agrupar todas las palabras que son flexiones de otra (info/informar/información/informador/informacional/etc.). El producto final es un listado con las palabras puntuadas y un listado de frases candidatas.
- **Módulo 2: Ponderación de frases.** Este módulo recibe las palabras etiquetadas por el módulo anterior, y su función es escoger entre todas las frases candidatas. Para hacerlo se ayuda de diversos submódulos que ponderan las frases según los parámetros siguientes: la frecuencia, la presencia de palabras indicativas (buscan palabras como *importante, esencial, conclusiones*, etc.), buscan frases que contengan palabras que aparezcan en el título, o que tengan nombres propios, o que la tipografía sea destacada (negritas, cursivas, tamaño superior, etc.) y seleccionan frases que aparezcan en posiciones destacadas en el texto (al principio de cada párrafo, al final a modo de conclusión).
- **Módulo 3: Detección de anáforas.** Una vez tiene las frases seleccionadas, puede ser que se dé el caso de anáforas mal resueltas (una frase contiene una anáfora que se encontraba en la frase previa y que no ha sido seleccionada). El programa busca las anáforas (especialmente los demostrativos pronominales o pronombres personales, por ejemplo *este, aquel, lo que, eso*) y su posición en la frase: al principio, entre las seis primeras palabras, en otras posiciones.
- **Módulo 4: Selección de frases.** Este módulo computa toda la información recogida en las fases anteriores: frases candidatas, puntuaciones, detección de anáforas. Selecciona las frases candidatas de puntuación más alta hasta llegar al tanto por ciento pedido por el usuario. Si entre estas frases hay alguna que contenga una anáfora, se selecciona la frase anterior (que contiene la palabra a la cual se está haciendo referencia) siempre y cuando forme parte de las frases candidatas y no sobrepase la longitud del resumen.
- **Módulo 5: Postprocesado del extracto.** Su función es detectar expresiones que conectan partes del texto, ya sea para mostrar causalidad, contraposición, etc. Son expresiones del tipo *por lo tanto, en contra*, etc. Como en el caso de las anáforas, si forman parte de una frase seleccionada, se procura incluir en el resumen la frase con la cual están relacionadas.

Por último, debemos recordar que algunos procesadores de textos, como Microsoft Word, también ofrecen esta opción (*Autosummarize* o Auto-resumen).

A modo de conclusión

- Los resúmenes automáticos (*extracts*) son una de las soluciones aportadas por el PLN para hacer frente al manejo de grandes volúmenes de información en línea.
- Los primeros en trabajar en el campo de la automatización de los resúmenes fueron Hans Peter Luhn en el año 1958 y Edmundson en 1969.
- Las técnicas han evolucionado de los primeros cálculos sobre la frecuencia de las palabras, o la posición de una frase dentro de un documento, a las técnicas basadas en conocimiento y recursos lingüísticos o en métodos estadísticos y de aprendizaje automático.
- La base de todas las técnicas es el cálculo de la frecuencia de las palabras. A partir de esta base estadística, se incorporan otras técnicas para dotar al programa de más conocimiento y paliar la escasa coherencia del resultado, por ejemplo la resolución de la anáfora o se aplican programas que proporcionen relaciones como las de sinonimia o hiperonimia o mecanismos para detectar y eliminar la redundancia.
- Los expertos consideran que la tecnología actual no tiene problemas para detectar las frases con más significado, pero sí para ordenarlas según su importancia.

3. La indización

“Indizar es la acción de describir o identificar un documento con relación a su contenido.”

Norma UNE 50-121-91.

Indizar es el resultado de examinar el documento, seleccionar los conceptos y almacenarlos en una base de datos.

Esta definición implica tres acciones, de las cuales la más significativa es la selección de los conceptos y su traducción al lenguaje documental.

Al igual que se ha tratado en el resumen, la indización la puede realizar una persona o un programa.

Si la indización es intelectual, es decir, la llevan a cabo personas, estas personas pueden ser:

- **Profesionales** (documentalistas), que llevan a cabo la tarea de indización de manera individual o en equipo. A su vez, los equipos pueden indizar de manera centralizada o coordinada.
- **Amateurs** (usuarios de Internet que indizan de manera social o *tagging* –por ejemplo, en Delicious–).

El elemento humano permite un análisis más rico del documento, captando conceptos y matices que un programa no llegaría a detectar, pero tiene el inconveniente del tiempo que se tiene que dedicar y la coherencia entre indizadores.

La indización automática se realiza a través de un programa informático. Su funcionamiento es muy sencillo: extrae del título, resumen o texto completo las palabras más significativas. Es un método económico y muy rápido.

3.1. Lenguaje natural y lenguaje documental

Para indizar necesitamos los lenguajes documentales. ¿Qué diferencia hay entre el lenguaje natural y el documental?

Ved también

La indización se estudia en los módulos “Sistemas de clasificación documentales”, “Listas de encabezamientos y listados de autoridades”, “Los tesauros” y “Listado de descriptores libres y listado de palabras clave”.

Ved también

La forma de indizar de los equipos se trata en el apartado 5 del módulo “La cadena documental” de esta asignatura.

Ved también

La indización automática se estudia en el módulo “Listado de descriptores libres y listado de palabras clave”.

Por **lenguaje natural** entendemos el lenguaje que usamos de forma cotidiana: catalán, castellano, vasco, gallego, francés, etc.

Por **lenguaje documental** entendemos el listado o vocabulario de términos que usamos para indizar y que puede estar en formato libre o controlado.

¿Y por qué hay que controlar los términos del lenguaje natural? Porque el lenguaje natural es ambiguo, los conceptos se pueden representar de formas diversas, dando lugar a problemas de recuperación. El lenguaje natural es rico en terminología, en formas (plurales y singulares), tiempos verbales, acrónimos, sinónimos, polisemias, etc.

La principal diferencia entre el lenguaje natural y el documental controlado es precisamente el control terminológico, que permite representar los conceptos de forma unívoca, sin ambigüedades.

Para ser más concretos, las diferencias se dan en el número de términos del vocabulario, el control de las formas, el control del significado y las relaciones de significado entre términos.

3.1.1. Número de términos

Los lenguajes documentales son entrópicos (Blanca Gil, 2004, pág. 20), es decir, tienden a la selección, a la restricción del vocabulario. Es el proceso contrario del lenguaje natural, que tiende a la abundancia, a la reiteración de conceptos, a la sinonimia en beneficio de una expresión más rica.

Los lenguajes documentales reducen considerablemente el número de términos del lenguaje natural, ya que sólo tienen en consideración los sustantivos y algunos sintagmas nominales, pero no adjetivos, preposiciones, conjunciones, adverbios, verbos, etc. Además, entre todos los sustantivos, escogen uno que representará al resto cuando el significado sea el mismo. Y entre diversas formas aceptadas por el mismo término, sólo una será la aceptada, como es el caso de las siglas.

Los lenguajes documentales son en esencia sencillos, su eficacia aumenta a medida que las reiteraciones y la redundancia son controladas en una única forma que reúne conceptos afines.

La riqueza del lenguaje natural

- Ejemplos de sinónimos del mismo concepto: Cosmos / Universo / Infinito / Firmamento / Cielo.
- Ejemplo del mismo concepto en formas diferentes, siglas o frases, y en idiomas diferentes: OTAN / NATO / Organizació del Tractat de l'Atlàntic Nord / Organización del Tratado del Atlántico Norte / North Atlantic Treaty Organization.
- Ejemplo de polisemia: Banco / Planta / Carta / Sierra / Estrella / Lengua / Capital.

Univocidad

La univocidad consiste en representar un concepto con un único término.

3.1.2. Control de las formas

Los lenguajes documentales controlan las formas plural/singular, el uso de acrónimos y siglas y la construcción de las frases, y de esta manera establecen unos modelos.

Ejemplo

Modelo	Ejemplo
Sustantivo	Pintura
Sustantivo + adjetivo	Pintura medieval
Sustantivo + preposición + sustantivo	Pintores de vitrales

Estas reglas gramaticales y sintácticas unifican las palabras seleccionadas y las frases.

Ejemplos en las listas de encabezamientos de materia

- Se acostumbra a usar el singular para expresar conceptos abstractos. Así, por ejemplo, es *solidaridad* y no *solidaridades*.
- No se permite el uso de siglas; se prefiere la expresión entera del concepto y en la lengua del servicio de información y documentación (SID³). Por ejemplo, Organización del Tratado del Atlántico Norte.
- Es preferible la expresión natural del concepto compuesto, y no su forma inversa. Es correcto *Objetos de arte*, y no *Arte, objetos de*.

Ved también

Los mejores ejemplos se ven en los módulos “Listas de encabezamientos y listados de autoridades” y “Los tesauros”.

⁽³⁾SID es la sigla de *servicio de información y documentación*.

3.1.3. Control del significado

Los problemas más importantes en cuanto al significado son la sinonimia y la polisemia.

a) **Sinonimia:** decimos que las palabras son sinónimas cuando tienen el mismo significado. En un sistema documental, si no se controlan y se usan indiscriminadamente, comportan silencio documental. En el caso de “alimento, nutriente, comida, provisión”, el usuario puede estar buscando por “alimento” y no recuperar documentos porque se encuentran indizados con otras formas, como “nutriente”. La solución de los lenguajes controlados es recoger todos los términos sinónimos y seleccionar uno para representar a todo el conjunto de términos que tienen el mismo significado, porque dos sinónimos son sustituibles el uno por el otro en cualquier contexto.

Ejemplo

Una lista de encabezamientos de materia como la del Consejo Superior de Investigaciones Científicas (CSIC) recoge todos estos sinónimos:

- Hispanoamericanos.
- Iberoamericanos.
- Latinoamericanos.
- Sudamericanos.

Pero sólo da como término aceptado “Latinoamericanos”. Si al SID⁴ llegara un documento titulado “Los sudamericanos del siglo XX”, el analista lo indizaría como **Latinoamericanos**, ya que es el término aceptado.

⁽⁴⁾A partir de ahora denotamos *servicios de información y documentación* con la sigla SID.

b) Polisemia: decimos que dos palabras son polisémicas cuando el mismo signo lingüístico, palabra o sonido tiene más de un significado. Habitualmente el contexto de la conversación o lectura donde está insertada la palabra deshace los problemas de ambigüedad, pero una palabra polisémica introducida en un sistema documental, sin el contexto, puede dar lugar a ruido documental.

Ejemplo

Un usuario puede estar buscando sobre columnas en arquitectura y recuperar datos sobre columnas tipográficas de diarios. Los lenguajes documentales controlan la polisemia diferenciando cada significado con paréntesis, usando el plural o el singular, adjetivando, etc.

Un tipo de polisemia es la homonimia. La diferencia entre ellas radica en la etimología de la palabra. Si la etimología de las dos palabras es la misma, hablamos de polisemia; si la etimología es diferente, hablamos de homonimia.

Ejemplos de polisemia y homonimia

Misma etimología = polisemia

La polisemia se da cuando una palabra tiene un único origen etimológico y acaba teniendo significados diferentes sin cambiar su categoría gramatical: por ejemplo, no pasa de sustantivo a verbo, como pasa en castellano entre el vino (bebida) y el vino (verbo venir). Es una palabra que con el tiempo ha ido adquiriendo diferentes significados, pero aun así, todos guardan entre sí una relación de significado; por ejemplo, en catalán y castellano fulla/hoja, que viene del latín *folia*, tiene diversos significados, como hoja de una planta, hoja de metal de una herramienta, página de un libro, cada una de las partes de una puerta doble o ventana, etc. Y en todos los significados lleva implícita la idea de una lámina.

Si queremos saber si una palabra es gramaticalmente polisémica, basta con consultar un diccionario etimológico y ver si proviene de un mismo origen. Encontraremos la palabra, un único origen y una lista de diferentes significados. En castellano podemos consultar el *Diccionario de la Real Academia*.

Más ejemplos de polisemia:

- *Servicio*, del latín *servitium*, que ha dado lugar a oficios religiosos, lavabos, misiones militares, cubiertos para comer y, en deportes, poner la pelota en juego. Y en todos ellos permanece la idea de ser útil.
- *Crucero*, del latín *crux*, significando ‘cruz’, intersección entre las dos naves de una iglesia, encargado de llevar la cruz a la cabeza de una procesión, viaje de placer por el mar, etc. En estos significados la idea es la de la forma de cruz, el cruzar como ir de un extremo a otro.
- *Columna*, del latín *columna*, que usamos para referirnos a los pilares arquitectónicos, las partes verticales de una página impresa de un diario, en física la forma que adoptan

algunos fluidos, como “columnas de humo”, en el ámbito militar, la formación de barcos o soldados. Y la idea que permanece es la de verticalidad.

Diferente etimología = homonimia

La homonimia se da cuando dos conceptos han llegado a tener el mismo nombre, la misma forma, pero vienen de orígenes diferentes y, por lo tanto, tienen etimologías diferentes.

Por ejemplo, *metro* puede ser el transporte urbano, una unidad de medida o el utensilio para medir. Pero el origen etimológico entre el transporte y los otros dos significados es evidente: el primero es una abreviación de la palabra inglesa *metropolitan*, y en el segundo caso viene del griego μέτρον y significa medida.

Otro ejemplo: la palabra castellana *botín* puede venir del latín *bota* y significará ‘calzado hasta el tobillo’, o puede venir del alemán *bytin* y significará ‘premio de una conquista’.

En castellano y catalán este fenómeno es menos frecuente que en otras lenguas, como el inglés o el francés, en las que abundan las palabras homónimas que dan mucho juego en los chistes.

Dentro de la homonimia podemos diferenciar las palabras que escribiéndose igual tienen significados diferentes, llamadas homógrafas, como las anteriores *metro* o *botín*, de las palabras que sonando igual también tienen significados diferentes, conocidas como palabras homófonas: *vell/bell* en catalán, o *tubo/tuvo* en castellano.

En resumidas cuentas, la sinonimia provoca silencio documental y la polisemia y variantes provocan ruido documental. El control terminológico del vocabulario garantiza el criterio de univocidad que tienen que tener los lenguajes documentales controlados, según el cual un concepto se representa con un término y un término sólo puede tener un significado.

3.1.4. Relaciones de significado de los términos

Por **relaciones de significado** entendemos la relación de genérico, específico o relacionado que puede tener un término con respecto a otro.

En el lenguaje natural estas relaciones son implícitas. Por ejemplo, cuando hablamos de manzanas todos entendemos que se trata de una fruta fresca y que las Fuji y las Golden son variedades concretas. Es decir, situamos el término “manzana” dentro de una jerarquía de términos conceptualmente más genéricos (fruta) y más específicos (Golden, Fuji). Incluso podemos relacionar por asociación de ideas la manzana con otras frutas, como la naranja o el plátano. Pero en un lenguaje documental hay que definir estas relaciones, agrupando y relacionando los términos afines.

La estructura que relaciona los términos es implícita en el lenguaje natural, pero en los lenguajes documentales hay que hacerla explícita. Eso se puede hacer de dos maneras:

a) En una secuencia jerárquica, donde la propia posición del concepto ya define sus términos genéricos y específicos. También deshace problemas de significado.

Ejemplo de la pesca

Ved el ejemplo de la pesca extraído de la *Clasificación Decimal Universal* (CDU). El concepto *pesca* puede ser la actividad económica o la pesca como deporte. Si nos fijamos en la cadena jerárquica vemos que cada uno cuelga de una clase diferente:

```
6 Ciencias aplicadas. Medicina. Tecnología
 63 Agricultura y ciencias relacionadas
   639 Caza. Pesca

7 Bellas artes. Juegos. Deportes
 79 Diversiones. Espectáculos. Juegos
   799 Caza deportiva. Pesca deportiva.
```

b) En una presentación alfabética donde cada término se acompaña de todos sus términos relacionados, ya sean equivalentes, genéricos, específicos o relacionados.

El tesoro del CSIC

En el tesoro de Psicología del CSIC, consultamos “Sueños” y encontramos:

Sueños

TG Dinámica de la personalidad

TE Contenido del sueño

TE Pesadilla

TR Déjà vu

TR Interpretación de los sueños

TR Sueño fisiológico

TR Sueño REM

TR Trastornos de conciencia

Las siglas nos informan del tipo de relación que establecen: TG significa término genérico (por encima de “Sueños” el tesoro tiene “Dinámica de la personalidad”), TE son los términos específicos (son términos específicos de “Sueños”: Contenido del sueño, Pesadilla) y los TR son los términos relacionados (se relacionan con “Sueño”, “Déjà vu”, la “Interpretación de los sueños”, el Sueño REM”, etc.).

Finalmente, las principales ventajas e inconvenientes del lenguaje natural y el documental controlado son:

Ventajas e inconvenientes de los lenguajes documentales

	Ventajas	Inconvenientes
Lenguaje natural	Amigable Actualizado Económico	Dificulta la búsqueda Poco preciso
Lenguaje documental controlado	Unívoco Facilita la búsqueda	Caro Poco actualizado

A modo de conclusión

Indizar es la acción de describir o identificar un documento en relación con su contenido.

La indización la puede realizar una persona (de forma centralizada o de forma coordinada) o un programa.

Por lenguaje natural entendemos el lenguaje que usamos de forma cotidiana (catalán, castellano, vasco), y por lenguaje documental entendemos el listado o vocabulario de términos que usamos para indizar y que puede estar en formato libre o controlado. La principal diferencia entre el lenguaje natural y el documental controlado es el control terminológico:

- El control del número de términos del vocabulario: los lenguajes documentales son entrópicos, tienden a la selección, a la restricción del vocabulario.
- El control de las formas: los lenguajes controlados, controlan las formas plural/singular, el uso de acrónimos y siglas y la construcción de las frases.
- El control del significado: los lenguajes controlados controlan la sinonimia y la polisemia. Decimos que las palabras son sinónimas cuando tienen el mismo significado. Decimos que dos palabras son polisémicas cuando el mismo signo lingüístico tiene más de un significado. La sinonimia provoca silencio documental y la polisemia y variantes provocan ruido documental. El control terminológico del vocabulario garantiza el criterio de univocidad que tienen que tener los lenguajes documentales controlados, según el cual un concepto se representa con un término y un término sólo puede tener un significado.
- Las relaciones de significado entre los términos son las relaciones de genérico, específico o relacionado que puede tener un término con respecto a otro. En el lenguaje natural estas relaciones son implícitas pero en los lenguajes documentales hay que hacerlas explícitas a través de una secuencia jerárquica o una presentación alfabética.

3.2. ¿Cómo se indiza?

Ahora que ya hemos visto la necesidad de contar con lenguajes documentales para paliar la ambigüedad del lenguaje natural, estamos en condiciones de preguntarnos por el proceso de indización que lleva a cabo un analista.

A continuación presentamos las **fases** que proponen diversos autores antes de llegar a la que nos servirá como marco de referencia en este subapartado:

- Dos fases: análisis del texto y traducción (Chaumier, 1988; Fidel, 1994).
- Tres fases: análisis del texto, identificación de conceptos y traducción (Amat, 1989; Norma UNE 50-121-91).
- Cuatro fases: análisis del texto, identificación de conceptos, traducción y establecer enlaces sintácticos entre descriptores (Slype, 1991).
- Cinco fases: registro de datos, análisis del texto, identificación de conceptos, traducción y examen de la indización.

En este módulo seguiremos la **norma UNE 50-121-91** y sus tres etapas:

- 1) Examinar el documento para identificar su contenido.
- 2) Seleccionar los conceptos principales del contenido.
- 3) Traducir a un lenguaje documental.

Norma UNE 50-121-91

UNE 50-121-91. *Métodos para el análisis de documentos, determinación de su contenido y selección de términos de indización.*

Ejemplo

Examinamos un libro titulado *Mitos de antiguas civilizaciones*. Leemos el título, el resumen, el sumario, etc.

En una segunda etapa seleccionamos como conceptos principales: Mitos, Grecia, Roma, India, Japón, Indios norteamericanos.

En la tercera etapa indizamos. Si indizamos con un lenguaje libre podemos escribir el término como deseamos o como salga en el texto. Por ejemplo:

Mitología india americana.

En cambio, si indizamos con un lenguaje controlado tendremos que traducir estos conceptos a una forma controlada. Pongamos por ejemplo que pensáramos indizar Mitología india americana. Veamos cómo quedaría en tres lenguajes documentales diferentes:

CDU	259.2
LEMAC	Mitología ameríndia
LEM del CSIC	Indios de América - Religión y mitología

A continuación se detalla cada parte del proceso.

1) Examen del documento e identificación de los conceptos

El analista tiene que examinar con precisión el documento. La lectura completa es, a menudo, impracticable, pero sí que tiene que prestar atención al título, resumen, sumario, introducción, ilustraciones y palabras o frases destacadas en una tipografía diferente.

No se recomienda la indización sólo a partir del título, ya que hay títulos que llevan a error, y tampoco confiar en que el resumen sea un sustituto del texto, ya que no todos los resúmenes están bien elaborados.

Ejemplo de títulos y resúmenes que no aportan datos significativos para la indización

- CHESNEAUX, Jean. *¿Hacemos tabla rasa del pasado?* México: Siglo XXI Editores 1981. Su materia es *Historia, historiadores, historiografía*. En el catálogo de la Biblioteca Nacional de España (BNE⁵) lo encontramos indizado como Historia.
- MALLOL, Tomas. *Si la memòria no em falla*. Girona: CCG Ediciones 2005. Su materia es *Memorias, cine, coleccionismo*. En la Biblioteca de Catalunya (BC⁶) lo encontramos indizado como Cine amateur.

Si recordamos el resumen del libro de Carl Sagan, *Cosmos*, nos daremos cuenta de que no era suficiente para indizar el contenido de la obra. Por estos motivos se recomienda una lectura ágil del resto de partes significativas del documento.

⁽⁵⁾BNE es la sigla de *Biblioteca Nacional de España*.

⁽⁶⁾BC es la sigla de *Biblioteca de Catalunya*.

Ved también

Recordad que el ejemplo del resumen del libro de Carl Sagan, *Cosmos*, salía en el apartado 2 de este módulo.

2) Selección de los términos de indización

Tal como dice la norma UNE, el analista tiene que identificar las nociones que son elementos esenciales de la descripción del contenido. Si la indización es compartida, la institución que la patrocina tiene que establecer claramente los factores que considera importantes.

Para seleccionar los conceptos del documento, el analista tiene que ser consciente del número de conceptos (criterio de exhaustividad) y de la exactitud de los mismos (criterio de especificidad).

a) Exhaustividad

A medida que el analista va leyendo, tiene que ir tomando nota de los conceptos interesantes del documento.

Una buena praxis es la que identifica los conceptos relevantes sobre:

- El tema.
- Los nombres personales que puedan ser interesantes de indizar.
- Los nombres geográficos.
- Las fechas cronológicas.
- La forma en que se presenta el documento: artículo, estadística, formulario o divulgación, científico, etc.

La exhaustividad es un criterio relacionado con el número de conceptos que se tienen en cuenta para caracterizar el contenido entero de un documento. El principal criterio de selección es el valor potencial del concepto para los usuarios de su SID.

Podemos distinguir entre una exhaustividad baja, media y alta en función del número de descriptores. Es en este entorno donde la norma UNE 50-121-91 da sus recomendaciones en cuanto a la exhaustividad. Los criterios que el indizador tiene que tener en cuenta son:

- El tipo de SID y perfil de usuario. No es lo mismo indizar para una base de datos genérica que para una específica.
- El tipo de documento. No se indiza con el mismo número de descriptores una monografía que un artículo de revista, una tesis, etc.

Tal como recomienda la norma UNE, no es conveniente ser estrictos con el número de términos, no se tiene que limitar el número de forma arbitraria, tipo “para una monografía dos términos de indización”, ya que puede conducir a una pérdida de objetividad y a una deformación de la información. Es preferible sugerir un baremo, entre tantos y tantos términos para cada tipo documental y SID y ser flexibles, ya que los criterios que tienen que regir son el propio contenido del documento y su posterior recuperación.

Ejemplo

Cuervo Herrero, C.; Fernández González, A.: “Objetos celestes erróneos”. *Tribuna de Astronomía y Universo. Revista de Astronomía, Astrofísica y Ciencias del espacio*. 2000. II Época, n° 16 – octubre. p. 36-40.

A partir del siguiente resumen informativo, elaboraremos tres tipos de indizaciones sugiriendo un baremo (para esta asignatura y sus prácticas) y una finalidad:

“Análisis y descripción de los errores más frecuentes que cometen los profesionales y aficionados a la fotografía astronómica mientras intentan descubrir nuevos objetos celestes todavía no identificados.

Estos errores son debidos a cuatro causas: errores en el proceso de positivado de la copia como consecuencia de la presencia de partículas de polvo en los negativos o en las lentes del equipo de laboratorio; errores en el negativo debidos a defectos de lavado, deficiencias en la emulsión, rayas y rasguños o por el uso de películas de color destinadas a ser forzadas, y errores en las lentes de los objetivos, debidos a efectos de distorsión y a alteraciones en la refracción. Finalmente se describen otras causas: reflejos de la luz del sol sobre las antenas de satélites artificiales Iridium, retoques digitales o de fotocopiadoras y duplicadoras, uso de objetivos sencillos y poco potentes para captar imágenes de cielo profundo y, en último término, oscilaciones del condensador de luz del microscopio.

Todos estos errores pueden dar lugar a imágenes falseadas: objetos inéditos, diámetros erróneos, efectos de redondeo, alineaciones planetarias erróneas, etc. El artículo facilita imágenes de estos errores fotográficos.

Los autores concluyen que hace falta ser cauteloso y hacer las oportunas comprobaciones antes de dar a conocer el descubrimiento de un nuevo objeto celeste a las sociedades astronómicas.”

Ejemplo de los tres grados de exhaustividad

Exhaustividad baja	Exhaustividad media	Exhaustividad alta
Baremo 1-3	Baremo 4-6	Baremo 7 ...
Ejemplo de uso: catálogo de una biblioteca pública	Ejemplo de uso: bases de datos de una biblioteca especializada en astronomía	Ejemplo de uso: bases de datos de una biblioteca especializada en astrofotografía
<ul style="list-style-type: none"> • Errores fotográficos • Fotografía astronómica 	<ul style="list-style-type: none"> • Astrofotografía • Errores fotográficos • Descubrimientos • Identificación de objetos celestes • Objetos erróneos 	<ul style="list-style-type: none"> • Alineaciones planetarias • Defectos de lavado • Deficiencias de la emulsión • Diámetros erróneos • Efectos de redondeo • Errores en el negativo • Errores en el positivado • Errores en las lentes • Objetos inéditos • Objetivos • Oscilaciones del microscopio • Partículas de polvo • Rayadas • Reflejos del sol • Retoques digitales

b) Especificidad

La especificidad está relacionada con la exactitud en que un concepto particular que aparece en un documento está representado por un término de indización.

Ejemplo

Si en el texto que estamos indizando aparece el concepto *Diplomacia*, y este término aparece en el lenguaje documental controlado, tenemos que indizar “Diplomacia”. Si indizamos “Relaciones internacionales” o “Embajadores” no estaremos siendo específicos, como podéis ver en la tabla siguiente:

Ejemplo de especificidad

Materia	Correcto, y por lo tanto:	Incorrecto por:	
	Específico	Genérico	Demasiado específico
Diplomacia	Diplomacia	Relaciones internacionales	Embajadores

Los conceptos se tienen que identificar de la manera más específica posible, pero en determinados casos se pueden preferir nociones más genéricas:

- Cuando el indizador considere que un exceso de especificidad puede ser negativa en la recuperación; por ejemplo, puede decidir que un modelo muy específico de una máquina se indice con el nombre más genérico de este tipo de máquinas.
- Cuando la idea no esté plenamente desarrollada en el documento, o sólo se haga alusión a ella.
- Cuando se esté a la espera de validar el término más específico.

3) Traducción a un lenguaje documental controlado

Para traducir el concepto inicial escrito en lenguaje natural a un lenguaje documental, el indizador tiene que consultar las listas del lenguaje buscando la forma correcta de introducir el concepto.

Ejemplos

Concepto tal como sale en el texto	Traducción	Lenguaje documental utilizado
Tragicomèdia	791.221.28	Classificación Decimal Universal (CDU)
Eolític	Edat de la pedra	Lista de encabezamientos de materia en catalán
Matriz	Útero	Lista de encabezamientos del CSIC
Monarquía absoluta	Absolutismo	Tesoro de Historia contemporánea del CSIC

Cuando el analista procede a traducir el concepto del texto se puede encontrar en las siguientes situaciones:

a) Encuentra el concepto, solo o repartido por las tablas:

- Consulta el lenguaje y encuentra el concepto a la primera. Entonces indiza con este término de indización. Por ejemplo, buscaba “Eolítico” y encuentra que tiene que indizar “Absolutismo”.
- Consulta el lenguaje y encuentra el concepto o las partes del concepto repartidos por el lenguaje. Entonces tiene que conocer las reglas de combinación de las partes integrantes del término de indización. Ejemplos:
 - Una notación con CDU como 391.91(961.3) “Tatuajes de la isla de Samoa” está formada por 2 elementos, tatuajes + Samoa. Estos elementos van colocados en un orden determinado por las reglas de precoordinación de la CDU (primero la clase principal + auxiliar).
 - Un encabezamiento construido con la LEM del CSIC como Agua-Aspectos económicos está formado por dos partes: Agua + Aspectos económicos, que es un encabezamiento y un subencabezamiento respectivamente y van en este orden.

Con los lenguajes tesauros y listado de autoridades no hay una sintaxis de combinación.

b) No encuentra el concepto:

- Consulta el lenguaje y no encuentra el concepto. Entonces el indizador tiene que conocer las obras de referencia que su SID considera como autoridades reconocidas en la materia. Estas obras de referencia son diccionarios, enciclopedias, otros lenguajes documentales (especialmente los tesauros construidos de acuerdo con las normas ISO y UNE 50-106 y UNE 50-125), atlas, etc.
- Hay lenguajes, como tesauros, donde el indizador tiene que proponer el término nuevo como descriptor candidato y esperar a que la dirección del tesoro lo valide como descriptor. Mientras tanto indiza con un término más genérico.

3.3. Calidad y coherencia de la indización

La **calidad** y la **coherencia** de la indización dependen de factores como la competencia del indizador y la calidad de los instrumentos o lenguajes documentales. La coherencia es un factor importante en el comportamiento de un sistema de indización, especialmente cuando forma parte de una red de centros y la información se tiene que intercambiar entre ellos.

La coherencia se calcula de la siguiente manera: dos analistas indizan el mismo documento, con un lenguaje de descriptores como un tesoro. Se cuentan separadamente el número de descriptores idénticos entre los dos analistas sobre el total de descriptores.

Ejemplo

Como ejemplifica van Slype:

- El documentalista 1 ha asignado los descriptores A, B, C, D, E, F.
- El documentalista 2 ha asignado los descriptores A, C, D, F, G, H.
- Hay 4 descriptores idénticos A, C, D, F y un total de 8 descriptores diferentes. Tasa de coherencia = $4/8 = 50\%$ (van Slype, 1991, p. 123).

La consistencia en la indización suele oscilar entre el 20% de mínima y el 60% de máxima (Isidoro Gil, 2001).

A modo de conclusión

La norma UNE 50-121-91 *Métodos para el análisis de documentos, determinación de su contenido y selección de términos de indización* establece tres fases:

Lectures complementary

Podéis ampliar la información sobre la coherencia en la indización leyendo las obras siguientes:

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

I. Gil Leiva (2001).

- Examinar el documento para identificar su contenido: el analista tiene que examinar con precisión el documento. La lectura completa es a menudo impracticable, pero sí que tiene que prestar atención al título, resumen, sumario, introducción, ilustraciones y palabras o frases destacadas en una tipografía diferente.
- Seleccionar los conceptos principales de los contenidos: el analista tiene que identificar las nociones que son elementos esenciales de la descripción del contenido, tiene que ser consciente del número de conceptos (criterio de exhaustividad) y la exactitud de los mismos (criterio de especificidad).
- Traducir a un lenguaje documental: para traducir el concepto inicial escrito en lenguaje natural a un lenguaje documental hay que consultar el listado del lenguaje buscando la forma aceptada.

4. Los lenguajes documentales

Un **lenguaje documental** es un vocabulario de términos en lenguaje natural o un sistema artificial de signos normalizados que facilitan la representación del contenido de los documentos.

Sus funciones principales son indizar el contenido de los documentos y permitir la recuperación a partir del campo materia.

Hay seis lenguajes documentales:

- Los sistemas de clasificación.
- Los listados de encabezamientos de materia.
- Los listados de autoridades.
- Los tesauros.
- Los listados de descriptores libres.
- Los listados de palabras clave.

En teoría todos los documentos se pueden indizar con cualquiera de estos seis lenguajes, pero en la práctica la tipología del SID (si es archivo, biblioteca o centro de documentación) y el tipo de usuario (general o especializado) condicionan que un SID indice y recupere con uno u otro lenguaje. En líneas generales:

- las bibliotecas indizan con sistemas de clasificación + listados de encabezamientos de materia + listados de autoridades;
- los centros de documentación indizan con tesauros + listados de palabras clave;
- los archivos, con sistemas de clasificación y/o tesauros.

Como podéis observar, los SID pueden trabajar con un solo lenguaje o con una combinación de lenguajes.

4.1. Los términos de indización

Llamamos **término de indización** a la representación de un concepto en lenguaje natural o un código de clasificación.

Los términos de indización pueden estar formados por una palabra o más de una.

La parte más pequeña con significado de un término de indización se conoce como **unitérmino**.

La norma UNE 50-113-92/1 define unitérmino como:

“El elemento significativo más pequeño de un lenguaje documental utilizado para representar un concepto específico en un sistema de indización coordinado; no se debe confundir con palabra clave o descriptor”.

UNE 50-113-92/1.

Cada lenguaje documental da un nombre diferente a su término de indización. Esta es la terminología que usaremos en esta asignatura:

Términos de indización

Lenguaje documental	Su término de indización se conoce como
Sistemas de clasificación	Notación o símbolo de clase
Listados de encabezamientos de materia	Encabezamiento
Listados de autoridades	Autoridad, identificador o descriptor
Tesauros	Descriptor
Listados de descriptores libres	Descriptor
Listados de palabras clave	Palabra clave

La norma UNE 50-113-92/1 define estos conceptos de la siguiente manera:

- “Notación/Símbolo de clase: es la representación de una clase mediante la notación de un sistema de clasificación.
- Identificador: nombre utilizado como descriptor.
- Descriptor: términos de indización asignados por el analista fruto de alguna de las operaciones intelectuales que implica el proceso de indización.
- Palabra clave: una palabra o grupo de palabras seleccionadas de manera automática del título, resumen o texto de un documento del que representan su contenido y permiten la recuperación.”

Norma UNE 50-113-92/1. *Documentación e información. Vocabulario. Parte 1. Conceptos fundamentales*.

A modo de conclusión

Un lenguaje documental es un vocabulario de términos en lenguaje natural o un sistema artificial de signos normalizados que facilitan la representación del contenido de los documentos. Sus funciones principales son indizar el contenido de los documentos y permitir la recuperación a partir del campo materia.

Hay seis lenguajes documentales:

- Los sistemas de clasificación.
- Los listados de encabezamientos de materia.
- Los listados de autoridades.
- Los tesauros.
- Los listados de descriptores libres.

Ejemplo

Ejemplos de términos de indización:

- De una palabra: “Bosques”.
- De más de una palabra: “Fuente de información”.

Ejemplo

El descriptor “Fuente de información” está formado por dos unitérminos: “Fuente” e “Información”. La preposición “de” no se indiza.

Lectura recomendada

Para cuestiones de terminología recomendamos la consulta de la norma UNE 50-113-92/1. *Documentación e información. Vocabulario. Parte 1. Conceptos fundamentales*. En: *Documentación: Normas fundamentales*. Madrid: AENOR, 1994.

- Los listados de palabras clave.

Llamamos término de indización a la representación de un concepto en lenguaje natural o un código de clasificación. Los términos de indización pueden estar formados por una palabra o más de una.

4.2. Evolución histórica de los lenguajes documentales

Los primeros analistas mesopotámicos, egipcios o romanos leían el documento, copiaban las primeras líneas del texto o seleccionaban los conceptos que mejor representaban el contenido y los escribían en la tablilla, *pinake*, *cartela* o ficha correspondiente. Poco a poco estas materias fueron conformando un listado de temas. En la Edad Media sabemos de la existencia de catálogos de algunas grandes bibliotecas, como la de Lorsh en Alemania, que tenía 600 títulos clasificados en 63 materias.

Edad contemporánea

Ahora bien, para muchos autores la historia de los lenguajes documentales empieza en las bibliotecas del siglo XIX con los sistemas de clasificación, ya que fueron el primer intento serio de controlar las materias de los documentos.

Los sistemas de clasificación empezaron a ser considerados propiamente lenguajes en el siglo XIX con las **clasificaciones bibliográficas** de Brunet, Harris, Dewey, Cutter o la de la Library of Congress. Eran cuadros de clasificación jerárquicos, de cariz enciclopédico y sus clases se combinaban de una forma definida con anterioridad, es decir, precoordinada. Los conceptos se representaban con códigos, no palabras. Por ejemplo, el concepto “Fotografía” era el código 77 (ejemplo extraído de la CDU).



Library of Congress

El siguiente paso en la evolución de los lenguajes lo formuló Charles Ammi Cutter en 1876, creando un listado de materias escritas en lenguaje natural. Ya no se usaba un código, sino que se expresaba el concepto (como “Fotografía”) con todas las letras. Estos listados, llamados **listas de encabezamientos de materia**, eran alfabéticos y se basaban en los principios de especificidad (hay que indizar con el término específico, no el genérico) y el de entrada directa (hay que respetar el orden natural de las expresiones y no optar por formas inversas del tipo “Electrónico, comercio”).



Charles Ammi Cutter

Las colecciones bibliotecarias estaban cubiertas con estos dos lenguajes documentales: sistemas de clasificación + listados de encabezamientos de materia. Los listados de autoridades controlaban el resto de autoridades. Además, se combinaban en los registros bibliográficos con el fin de minimizar el inconveniente de la codificación, ya que no era de fácil comprensión para los usuarios. La indización era sintética, sumaria, dos o tres entradas por el campo materia, ya que tenemos que ser conscientes de que nacieron en sistemas no automatizados.

A medida que la producción científica iba generando cada vez más volumen de información, surgió la necesidad de indizar de una forma más analítica, con más conceptos. Se crean **centros de documentación** con una vocación más especializada que las bibliotecas. El uso de tecnología informática facilitaba el acceso a un documento a través de diversos puntos de acceso. Nacen los lenguajes especializados por excelencia, los **tesauros**. Se aplican a los centros de documentación y a algunos archivos históricos y administrativos.

Los tesauros recogen todo lo mejor de sus antecesores: la estructura arborescente de los sistemas de clasificación, que aplican a su presentación jerárquica, y la estructura combinatoria de las listas de encabezamientos de materia, que aplican a su presentación alfabética. Además incluyen nuevas estructuras de presentación, como la gráfica y la de índices permutados.

Los tesauros se automatizan y desde mediados de los años setenta el crecimiento de la industria de las bases de datos posibilita la consulta en línea de muchas publicaciones seriadas. Nace el último lenguaje documental, el **listado de palabras clave o indización automática**.

Internet

La última gran etapa la marca **Internet**. La globalización de la red a partir de la década de los años 1990 impulsa el acceso a la información, ya no hace falta que los SID dispongan en propiedad del documento, ya que la red permite acceder a la información alojada en cualquier otro centro de información. La cooperación impulsa a todos los lenguajes documentales a automatizarse y formar parte de proyectos colectivos (catálogos colectivos, consorcios, redes). En el mismo sentido se buscan pasarelas entre los diferentes lenguajes para solucionar problemas idiomáticos entre países.

Surge la necesidad de indizar la abundante producción de recursos electrónicos, como por ejemplo con el uso de metadatos para definir e intercambiar datos entre sistemas informáticos (etiquetas del tipo <subject>,<keywords>) y explotar la indización automática en los potentes robots de los buscadores. También los usuarios pueden indizar los recursos gracias a iniciativas de indización social o *tagging*.

Los expertos opinan que en la actualidad el problema principal no es tanto indizar o recuperar, sino presentar los resultados en algún orden significativo, lo que implica el uso de algoritmos que valoren los resultados.

A continuación reproducimos algunas de las fechas más significativas, extraídas de la cronología de Isidoro Gil (2008) sobre las listas de encabezamiento de materia, los tesauros y la indización automática.

Lectura complementaria

Podéis encontrar esta cronología en la obra siguiente:

I. Gil Leiva (2008). *Manual de indización. Teoría y práctica*. Gijón: Ediciones Trea (Bibliotecología y Administración cultural, 193), pág. 110-114.

Cronología de la evolución de los lenguajes documentales

Fechas	Concepto	Breve explicación
30.000 a.C.	Etiquetas de barro	Los antiguos escribas mesopotámicos guardaban las tablillas de barro (documentos) en cestas de mimbre. Por fuera, la cesta llevaba otra tablilla de barro con el contenido.
Egipto	Las <i>cartelas</i> de Egipto	Los egipcios introducen el papiro como soporte documental. El papiro se enrollaba en torno a una varita de madera o metal. Para no desplegar completamente el rollo, ponían las primeras frases del documento en una etiqueta o <i>cartela</i> en un extremo.
1876	Charles A. Cutter Rules for a dictionary catalog	
1895	List of subject headings for use in dictionary catalogs	Publicada por la American Library Association (ALA) para bibliotecas medias y pequeñas, con fondos no especializados.
1909	Library of Congress Subject Headings	Nace a partir de la lista de ALA y las reglas de Cutter. A partir de aquí esta lista se convierte en el referente de todas las listas de encabezamientos de materia del mundo.
1923	List of subject headings for small libraries	Minnie Earl Sears es la autora de esta lista conocida como SEARS. Es una versión reducida de la LCSH para bibliotecas pequeñas.
1934	Guía para los encabezamientos de materia	Juan Manrique Lara publica la primera lista de encabezamientos en castellano en México. Era una traducción de la Library of Congress Subject Headings (LCSH), el ALA y la SEARS.
1946	Répertoire de vedettes-matière RVM	Primera lista de encabezamientos en francés (Universidad de Laval Canadá).
1951	Descriptor	Calvin Mooers acuña el término.
1952	Unitérmino	Mortimer Taube acuña el término.
1957	Indización automática	Hans Meter Luhn empieza a trabajar en indización automática aplicando el método de la frecuencia.
1960	Compatibilidad	En la década de los 60 se inician los primeros proyectos para hacer compatibles los diferentes lenguajes documentales mediante tablas de equivalencia.
1961	Sistema SMART	Gerald Staton desarrolla el sistema SMART de análisis automático de textos.
1967	Guidelines for the development of information retrieval thesauri	Directrices para elaborar tesauros confeccionadas por el US Federal Council for Science and Technology de Washington
1967	Lista de encabezamientos de materia para bibliotecas	Lista compilada por Carmen Rovira y Jorge Aguayo en español para la Unión Panamericana.
1974	Norma ISO 2788:1974 Guidelines for the establishment and development of monolingual thesauri	1ª edición de la norma ISO para la confección de tesauros monolingües.
1980	Répertoire d'autorité-matière encyclopédique et alphabétique unifié RAMEAU	Primera lista de encabezamientos de materia de la Biblioteca Nacional de Francia. Se basaron en la RVM y la LCSH.
1983	Bilindex	Lista de encabezamientos bilingüe en inglés y castellano. Es equivalente a la LCSH. En el año 2007 se editó la 15 ed.
1985	Norma ISO 5963:1985 Methods for examining documents	Norma ISO que no sería traducida a norma UNE hasta 1991 con el número UNE 50-121-91.

Fechas	Concepto	Breve explicación
1985	Norma ISO 5964:1985 Guidelines for the establishment and development of multilingual thesauri	1ª edición de la norma ISO para la confección de tesauros multilingües.
1986	Abandono de los símbolos tradicionales de las listas de encabezamientos por los propios de los thesaurus	La LCSH, en su 10 edición, abandona los símbolos de x, see, xx, v, a por los propios de los tesauros Use, BT, NT, RT. Las demás listas mundiales también los adoptan.
1986	Unified medical language system	El sistema unificado de lenguajes en medicina es un proyecto para integrar los diferentes vocabularios de ciencias de la salud. Es un proyecto de la Biblioteca Nacional de Medicina de EE.UU. (actualmente coordina el MESH, Medical Subject Headings).
1995	Universalización de Internet	Internet ha difundido y popularizado conceptos, técnicas y prácticas propias de documentalistas.
1995	Metadatos	Uso de metadatos para definir e intercambiar datos entre sistemas informáticos. Los lenguajes de marcaje tienen etiquetas para el resultado de la indización del tipo <subject>, <keywords>.
1997	Proyecto MACS	Iniciativa de la Conference of European National Libraries CENL para hacer compatibles tres listas de encabezamientos de materia, la alemana SWD, la RAMEAU francesa y la LCSH usada en Gran Bretaña y Suiza.

A modo de conclusión

Para muchos autores la historia de los lenguajes documentales empieza en las bibliotecas del siglo XIX con los sistemas de clasificación, ya que fueron el primer intento serio de controlar las materias de los documentos.

El siguiente paso en la evolución de los lenguajes lo formuló Charles Ammi Cutter en 1876, creando una lista de materias escritas en lenguaje natural.

A medida que la producción científica iba generando cada vez más volumen de información, surgió la necesidad de indizar de una forma más analítica, con más conceptos. Se crean centros de documentación con una vocación más especializada que las bibliotecas. Nacen los lenguajes especializados por excelencia, los tesauros.

Desde mediados de los años setenta el crecimiento de la industria de las bases de datos posibilita la consulta en línea de muchas publicaciones seriadas. Nace el último lenguaje documental, el listado de palabras clave o indización automática.

La última gran etapa la marca Internet. La globalización de la red a partir de los años 1990 impulsa el acceso a la información. La cooperación impulsa a todos los lenguajes documentales a automatizarse y formar parte de proyectos colectivos (catálogos colectivos, consorcios, redes). En el mismo sentido se buscan pasarelas entre los diferentes lenguajes para solucionar problemas idiomáticos entre países.

Surge la necesidad de indizar la abundante producción de recursos electrónicos, como por ejemplo el uso de metadatos para definir e intercambiar datos entre sistemas informáticos (etiquetas del tipo <subject>, <keywords>) y explotar la indización automática en los potentes robots de los buscadores. También los usuarios pueden indizar los recursos gracias a iniciativas de indización social o *tagging*.

4.3. ¿Cuándo son necesarios los lenguajes documentales?

Los lenguajes documentales son necesarios en dos momentos de la cadena documental:

- La fase de análisis y tratamiento > Análisis documental > Análisis de contenido > Indización.
- La fase de salida > Instrumentos de recuperación.

Tanto en la fase de indización como en la fase recuperación, el proceso de análisis-selección-traducción de conceptos es el mismo. En el momento de la indización el analista lee el documento, extrae conceptos y si hace falta los traduce a un lenguaje controlado para almacenarlos en el sistema. En el momento de la recuperación, el analista tiene que trabajar con la consulta del usuario, extraer los conceptos y traducirlos. Si se trata de un lenguaje post-coordinado, además tendrá que saber cómo convertir los descriptores a una ecuación de búsqueda.

Ejemplo de la fase de recuperación

- **Usuario:** "Necesito información sobre las instalaciones deportivas de hockey hierba que se construyeron en la ciudad de Terrassa con motivo de la celebración de los Juegos Olímpicos de 1992."
- **Analista:** selecciona los conceptos más relevantes para la búsqueda: instalaciones deportivas, hockey hierba, Terrassa, Juegos Olímpicos. El próximo paso es traducir los conceptos a un lenguaje documental, en el ejemplo, el "Tesauro d'Història local de Catalunya". Como se puede apreciar entre la expresión en lenguaje natural del usuario y los descriptores aceptados del tesauro hay ciertas diferencias:

En la expresión del usuario:	Traducido al tesauro:
Instalaciones deportivas	Equipamientos deportivos (Equipaments esportius)
Hockey hierba	Hockey (Hoquei)
Terrassa	Terrassa
Olimpiadas	Juegos Olímpicos 1992 (Jocs Olímpics 1992)

Traducido a una ecuación de búsqueda: Equipamientos deportivos AND Hockey AND Terrassa AND Juegos Olímpicos 1992.

G. van Slype (1991, pág. 161) considera que los lenguajes documentales pueden intervenir, como máximo, hasta en seis momentos diferentes en la recuperación:

- 1) Selección de los sistemas documentales que se interrogarán: qué catálogos, qué bases de datos, etc.
- 2) Selección de los conceptos expresados por el usuario en su enunciado.
- 3) Traducción a un lenguaje documental controlado.
- 4) Formulación de la ecuación de búsqueda.
- 5) Extensión asistida por ordenador.
- 6) Evaluación final de la pertinencia de los resultados obtenidos.

Hay una tercera función dentro de la cadena documental, pero sólo afecta a un lenguaje documental concreto, que son los sistemas de clasificación:

- La fase de Análisis y Tratamiento > Procesamiento técnico > Ordenación.

Los códigos numéricos de los sistemas de clasificación jerárquicos, como la CDU, son la herramienta para ordenar los documentos en las estanterías de acuerdo a un orden secuencial de las materias (ordenación altamente significativa).

En teoría todo documento se podría indizar con cualquiera de los seis lenguajes. En la práctica cada tipología de SID tiende a utilizar un lenguaje o combinación de lenguajes concreta.

Lectura complementaria

Podéis ampliar la información sobre los lenguajes documentales en la obra siguiente:

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

Ejemplo: un documento y seis indizaciones

Veamos cómo sería el resultado de indizar el mismo documento con cada uno de los seis lenguajes documentales:

El mercado del tabaco en España durante el siglo XVIII: fiscalidad y consumo / Santiago de Luxán Meléndez, Sergio Solbes Ferri, Juan José Laforet (ed.). Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria, Servicio de Publicaciones, 2000.

Resumen:

“En este libro se ha querido poner el énfasis en un tema hasta ahora poco tratado como es el consumo de tabaco en España durante el siglo XVIII.

No obstante también se atienden otros aspectos como los fiscales. La obra se ha estructurado en tres partes: la primera se ocupa de la fiscalidad, la segunda atiende el área del monopolio y la tercera analiza los mercados regionales de Canarias y Navarra. El libro se cierra con un apartado dedicado al cultivo del tabaco.”

Ejemplo de un único documento y seis indizaciones

Sistema de clasificación: CDU	Listas de encabezamientos de materia: LEMAC	Listado de autoridades: Gran Enciclopèdia Catalana
336.226(460)“17”:663.97	Industria tabaquera- España- Historia - s. XVIII Tabaco - impuestos - España - Historia - s. XVIII	Canarias España Navarra
Tesaurus: Tesaurus d’Història local de Catalunya (UAB)	Listado de descriptores libres: Consultores de la asignatura	Listado de palabras clave: programa Swesum
Tabaco Consumo Historia Impuesto de consumos Cultivos Monopolios Siglo XVIII	Canarias Cultivo Consumo España Fiscalidad Monopolio Navarra Siglo XVIII Tabaco	libro tabaco

Aunque en este momento el estudiante no conozca el funcionamiento de estos lenguajes, sí que está en disposición de observar algunos rasgos característicos de cada uno:

- El sistema de clasificación ha indizado un código, no son palabras. Es un código construido a base de números y símbolos, incomprensible a primera vista para un profano.
- La lista de encabezamientos de materia ha indizado dos términos en lenguaje natural, que están formados por diversas palabras separadas con guiones.
- La lista de autoridades ha indizado sólo nombres geográficos y ha prescindido del resto de conceptos. También ha usado el lenguaje natural.

- El tesoro ha indizado unos cuantos descriptores en lenguaje natural, poniendo un término bajo el otro.
- El listado de descriptores libres no se diferencia a simple vista de la indización con tesoro. En cambio, la diferencia es fundamental ya que el tesoro es controlado y los descriptores libres son libres.
- En el listado de palabras clave, la indización la ha realizado un programa informático, que ha seleccionado las palabras *libro* y *tabaco* porque salen dos veces en el texto, son las palabras más repetidas.

A modo de conclusión

Los lenguajes documentales son necesarios en dos momentos de la cadena documental:

- La fase de análisis y tratamiento > Análisis documental > Análisis de contenido > Indización.
- La fase de salida > Instrumentos de recuperación.

Los sistemas de clasificación también son útiles en:

- La fase de análisis y tratamiento > Procesamiento técnico > Ordenación.

4.4. Complementariedad de los lenguajes documentales

Indizar con más de un lenguaje documental al mismo tiempo es muy conveniente, porque así se suman las ventajas y minimizan los inconvenientes de los diferentes sistemas. Significa un esfuerzo añadido en el momento de la indización pero permite recuperar de manera más precisa. Es decir, combinamos lenguajes para recuperar mejor.

Algunas de las combinaciones posibles son las siguientes:

- Sistema de clasificación + listas de encabezamientos + listados de autoridades.
- Sistema de clasificación + listas de encabezamientos + listados de autoridades + palabras clave.
- Sistemas de clasificación + tesoro.
- Tesoro + listados de autoridades + palabras clave.

Ejemplo de combinación de lenguajes

Ejemplo de una captura de un registro del catálogo de la Biblioteca Nacional de España donde vemos un campo para la notación con CDU y otro para un encabezamiento de materia.

Ved también

Todos estos temas serán desarrollados en los módulos siguientes, dedicados a cada uno de los lenguajes documentales.

Historia de las ideas políticas [Texto impreso]
Touchard, Jean
CDU: 32(091)
Autor personal: [Touchard, Jean](#)
Título uniforme: [\[Histoire des idées politiques Español\]](#)
Título: [Historia de las ideas políticas \[Texto impreso\] / Jean Touchard ; con la colaboración de Louis Bodin ... \[et al. ; traducción de J. Pradera\]](#)
Edición: 6ª ed.
Publicación: Madrid : Tecnos, 2006
Descripción física: 659 p. ; 24 cm
Serie: [\(Colección de ciencias sociales. Serie de ciencia política\)](#)
Nota al título y mención: Traducción de: Histoire des idées politiques
Encabez. materia: [Política -- Historia](#)
N. depósito leg.: M 7115-2006

A modo de conclusión

En teoría todos los documentos se podrían indizar con cualquiera de los seis lenguajes. En la práctica cada tipología de SID tiende a utilizar un lenguaje o combinación de lenguajes concreta.

Indizar con más de un lenguaje documental al mismo tiempo es muy conveniente porque se suman las ventajas y minimizan los inconvenientes de los diferentes sistemas. Significa un esfuerzo añadido en el momento de la indización pero permite recuperar de manera más precisa. Es decir, se combinan lenguajes para recuperar mejor.

5. Tipología de los lenguajes documentales

Podemos clasificar los seis lenguajes documentales a partir de unas características o tipologías que los describen. Concretamente los lenguajes se tipifican según la naturaleza de sus términos, el nivel de control, el nivel de coordinación, la estructura y el nivel de análisis:

Tipología de los lenguajes documentales

Tesaurus

Un lenguaje es la suma de diversas características. Así, por ejemplo, un tesaurus es natural, controlado, postcoordinado, jerárquico y combinatorio e indiza por conceptos.

		Sistemas de clasificación	Listados de encabezamientos de materia	Listados de autoridades	Tesaurus	Listado de descriptores libres	Listado de palabras clave
Según la naturaleza de los términos	Codificado	X					
	Natural		X	X	X	X	X
Según el nivel de control sobre los términos	Libre					X	X
	Controlado	X	X	X	X		
Según el nivel de coordinación de los términos	Precoordinado	X	X				
	Postcoordinado			X	X	X	X
Según la forma de agrupar los términos o estructura	Jerárquico o Sistemático	X			X		
	Combinatorio		X	X	X	X	X
Según el nivel de análisis	Por materias	X	X				
	Por conceptos			X	X	X	
	Por palabras clave						X

A continuación, vamos a ver estas características.

5.1. Naturaleza del término: codificado o natural

Los términos pueden expresarse en lenguajes codificados o naturales:

a) **Lenguajes codificados.** Entendemos por *codificado* el uso de un código artificial compuesto de números, letras y símbolos que traducen un concepto. Por ejemplo, el Sol, en un lenguaje como la CDU, sería 523.9.

Los lenguajes codificados son lenguajes sintéticos, muy usados en bibliotecas, ya que, además de clasificar el contenido del fondo documental, son operativos en cualquier idioma y permiten la ordenación de los fondos. Por otra parte, tienen el inconveniente de ser poco comprensibles por parte de los usuarios.

Sólo hay un tipo de lenguaje codificado: son los **sistemas de clasificación.**

b) **Lenguajes naturales.** Entendemos por *natural* el uso de palabras del lenguaje usual, habitual, no códigos. Es mucho más próximo al usuario, más amigable. Hay cinco lenguajes documentales naturales:

- Las listas de encabezamientos de materia.
- Los listados de autoridades.
- Los tesauros.
- Los listados de descriptores libres.
- Los listados de palabras clave.

5.2. Nivel de control: libre o controlado

Hace referencia al control del vocabulario, es decir, si las palabras seleccionadas para indizar corresponden al lenguaje natural o a un lenguaje artificial construido para garantizar la indización y recuperación:

a) **Lenguajes libres.** Son listas de términos extraídos del lenguaje natural sin sufrir ningún tipo de control. Normalmente los lenguajes libres se utilizan en sistemas automatizados donde hay un fichero inverso o diccionario de la base de datos. Tienen muchas ventajas en la indización, como el gasto mínimo de construcción, la actualización inmediata, coherencia máxima y la riqueza terminológica. Pero presentan inconvenientes en la recuperación, ya que al trabajar con lenguaje natural, arrastran todos los problemas derivados de la ambigüedad (sinonimia, polisemia, homonimia).

Los lenguajes libres son dos:

- Los listados de descriptores libres.
- El listado de palabras clave.

b) **Lenguajes controlados.** Consideramos lenguajes controlados aquellos que están redactados previamente en forma de listas o listados de términos que se consideran aceptados y unívocos para la indización. Sólo los términos de la lista se pueden usar para indizar.

Algunos lenguajes codificados

Son ejemplos de lenguajes codificados la Clasificación Decimal Universal (CDU), la Clasificación Dewey (DDC), la Clasificación de la Library of Congress (LCC) o la Clasificación Colon (CC).

Ved también

Los sistemas de clasificación se estudian con más profundidad en el módulo "Sistemas de clasificación documentales" de esta asignatura.

Son términos seleccionados tanto en su forma (plural, singular, sintagma nominal, adjetivado, siglas, etc.), como en su contenido (de todos los sinónimos se escoge uno, los homónimos se diferencian entre ellos, etc.) y como en sus relaciones de jerarquía y asociación (términos conceptualmente más genéricos o específicos y términos que se evocan mutuamente). Requieren unos gastos de construcción elevados, tanto en personal cualificado como en tiempo. Para muchos autores son los verdaderos lenguajes documentales. También se conocen por el nombre de **lenguajes artificiales**.

Su función documental es la de representar un concepto con un único término y que sólo haya un término por concepto, lo que se conoce como univocidad.

Los lenguajes controlados son cuatro:

- Los sistemas de clasificación.
- Las listas de encabezamientos.
- Los listados de autoridades.
- Los tesauros.

5.3. Nivel de coordinación: precoordinado o postcoordinado

a) **Precoordinación.** La precoordinación consiste en determinar a priori cómo se combinan los términos, tanto sea a la hora de construir el lenguaje como a la hora de indizar el documento o a la de recuperarlo.

Lenguajes precoordinados

Un ejemplo de construcción con un lenguaje precoordinado como la “Lista de encabezamientos de materia” como la del CSIC prevé que la materia *Construcción de viviendas* se represente como:

Viviendas - Construcción

Es decir, por este orden y separados con un guión.

Un ejemplo de indización con un lenguaje precoordinado, por ejemplo, de una materia compuesta por tres elementos como *Enciclopedia de los perros pastores europeos* se representa como:

Perros Pastores - Europa - Enciclopedias

El encabezamiento se hace en este orden concreto, y las reglas sintácticas del lenguaje evitan la posibilidad de otras combinaciones.

La precoordinación tiene dos grande ventajas:

- Agrupa en proximidad todos los documentos que tienen una temática afín, de manera que si consultamos el catálogo de una biblioteca *Viviendas*, también veremos otros documentos como:

Viviendas - Alumbrado

Viviendas - Arrendamiento

Viviendas - Calefacción y ventilación

- Un solo término de indización reúne los elementos principales para la búsqueda.

La precoordinación era una auténtica necesidad en el entorno de las bibliotecas manuales, ya que no se podía buscar por una combinación de dos o más términos.

b) Postcoordinación. La postcoordinación consiste en combinar los términos de indización en el momento de la recuperación. Permite combinar múltiples términos de indización siguiendo la lógica de los operadores booleanos y de esta manera profundizar en el análisis de contenido. No tienen sintaxis en el momento de la indización. Cada término indizado es un punto de acceso al documento; cuantos más términos indizamos, más posibilidad tenemos de recuperarlo.

Lenguajes postcoordinados

Un lenguaje postcoordinado, como un tesoro, representaría el documento anterior sobre perros pastores como:

Perros pastores
Europa
Enciclopedia

que sería recuperado siguiendo la lógica de los operadores booleanos:

Perros Pastores AND Europa

Los lenguajes postcoordinados sólo tienen sentido en sistemas documentales automatizados que dispongan de un fichero inverso. El fichero inverso está donde se almacenan todos los descriptores que el analista va indizando, se sitúan uno detrás del otro de forma secuencial y asociados al documento al que hacen referencia.

Los lenguajes postcoordinados son cuatro:

- Listados de autoridades.
- Tesoro.
- Listados de descriptores libres.
- Listados de palabras clave.

Ved también

El tema de la precoordinación se trata sobradamente en los módulos dedicados a los dos lenguajes precoordinados: "Sistemas de clasificación documentales" y "Listas de encabezamientos y listados de autoridades".

Ejemplo de fichero inverso

Fichero inverso

Documento	Fichero inverso: concepto y n°. de documento
Documento 1	Alimentación (2) Enciclopedia (1,3)
Perros pastores	Entrenamiento (2)
Europa	Europa (1)
Enciclopedia	Perros Pastores (1,2) Química orgánica (3)
Documento 2	
Perros pastores	
Alimentación	
Entrenamiento	
Documento 3	
Química orgánica	
Enciclopedia	

5.4. Estructura: jerárquica o combinatoria

El vocabulario de los lenguajes documentales se organiza en dos estructuras básicas, en forma jerárquica o en forma combinatoria:

a) **Jerárquica:** en la estructura jerárquica o arborescente, el vocabulario se presenta en forma de cadena, con términos genéricos que agrupan términos más específicos. Todos los términos dependen de un término superior y de significado más genérico. Esta estructura permite agrupar los conceptos por temas y también situarlos en contexto, ya que la secuencia jerárquica nos informa de cuál es el campo temático al que está adscrito el concepto.

Ejemplo

Pongamos un ejemplo extraído de la CDU:

```
37 Educación
  371 Organización de la educación
  372 Contenido. Materias
  373 Tipo de escuelas
  374 Enseñanza extraescolar
  376 Escuelas especiales
  377 Formación profesional
  378 Universidades
```

Así, el concepto "Universidades" depende del concepto "37 Enseñanza", por lo tanto hace referencia a la educación que se imparte en la universidad y no a la arquitectura de las universidades (que estaría dentro de "72 Arquitectura").

Los lenguajes jerárquicos son dos:

- Los sistemas de clasificación.
- Los tesauros (en la parte de presentación sistemática o jerárquica).

b) Combinatoria: en la estructura combinatoria, los términos no forman cadena, están listados por orden alfabético. Este tipo de estructura surgió como reacción a la rigidez de la estructura jerárquica, que no era fácil de actualizar.

Ejemplo extraído de la Lista de encabezamientos del CSIC

[Peaies](#)
[Pearcea](#)
[Pearl Harbor, Ataque a, 1941](#)
[Pecado](#)
[Pecado \(Islam\)](#)
[Pecado original](#)
[Pecados](#)
[Pecados capitales](#)
[Pecaris](#)

La estructura combinatoria permite la inclusión de nuevos términos y la eliminación de los obsoletos sin afectar al resto de la estructura del lenguaje. La facilidad para actualizar el vocabulario los convierte en lenguajes adecuados para todo tipo de entornos: enciclopédicos, científicos y técnicos.

Los lenguajes de estructura combinatoria son cinco:

- Listas de encabezamientos de materia.
- Listados de autoridades.
- Tesoros.
- Listados de descriptores libres.
- Listados de palabras clave.

Como se puede observar, los tesauros participan de las dos estructuras: tienen una presentación sistemática en forma jerárquica y una presentación alfabética en forma combinatoria.

El descriptor “Còmic”

Veamos el descriptor “Còmic” tanto en una presentación como en la otra (extraído del Tesauro d’història local de Catalunya).

Presentación jerárquica (izquierda) y alfabética (derecha)

Fuentes primarias	Comics
UP Documentos primarios	TG Fuentes primarias
	TR Dibujos
TG Fuentes de información	TR Ilustraciones
	TR Láminas
TE Almanagues	
TE Atlas	
TE Biografías	
TE Comics	
TE Cronologías	
TE Diccionarios	
TE Directorios	
TE Enciclopedias	

5.5. Nivel de análisis: materias, conceptos, palabras clave

Los lenguajes pueden indizar más o menos conceptos, de manera que podemos establecer una última tipología según la cantidad de información que transmiten cada uno. En el punto más sintético, con uno o dos términos de indización, tenemos los lenguajes que indizan por materias; en el punto medio, los lenguajes de conceptos, también llamados de descriptores, y en el punto más analítico, los lenguajes de palabras clave.

Indizar por materias, conceptos y palabras clave está en relación directa con los dos paradigmas de búsqueda. La indización por materias es adecuada para sistemas de *browsing* (o de navegación o directorio). En cambio, las indizaciones por conceptos y palabras clave se adaptan mejor a los sistemas de interrogación en buscadores.

a) Por materias: responden a la pregunta: “¿cuál es el tema de este documento?”. Los lenguajes que indizan por materias son dos:

- Los sistemas de clasificación.
- Las listas de encabezamientos de materia.

b) Por conceptos: indizar por conceptos significa indizar las ideas y nociones del texto sin reducirlo a un tema principal. Responden a la pregunta: “¿cuáles son los conceptos de este documento?”. Van ligados necesariamente a sistemas automatizados, ya que no es factible elaborar tantas fichas de cartulina como conceptos se van a indizar.

Los lenguajes que indizan por conceptos son tres:

- Listados de autoridades.
- Tesauro.
- Listados de descriptores libres.

c) Por palabras clave: indizar por palabras clave significa indizar todas las palabras con significado del texto. Es el proceso más analítico que existe. No es una tarea de indización humana, sino automática. Los programas que indizan por palabras clave seleccionan sólo las palabras que tienen significado (preferentemente sustantivos).

Sólo hay un lenguaje por palabras clave, y es evidentemente el único lenguaje automático: el listado de palabras clave.

Ejemplo de indización con los tres niveles de análisis

Indizaremos con los tres niveles de análisis el siguiente resumen indicativo:

MUÑOZ CRUZ, Valle. El papel del gestor de la información en las organizaciones a las puertas del siglo XXI. A. *Los sistemas de información al servicio de la sociedad: actas de las jornadas*. Valencia: FESABID, 1998, vol. 2, p. 649-660.

“Artículo sobre el papel y funciones del gestor de la información, un nuevo profesional de la documentación, en las organizaciones del siglo XXI. Describe el panorama laboral español, analizando la Administración pública y la empresa privada. Propone desarrollar una política nacional de información y una formación adaptada a las necesidades organizativas de las instituciones.”

Ejemplo de niveles de análisis

Por materias	Por conceptos	Por palabras clave	
Gestor de información	Gestor de información Documentación Administración pública Empresa privada Política de información	Adaptada Administración Artículo Documentación Empresa Español Formación Funciones Gestor Información Instituciones Laboral Nacional	Necesidades Nueve Organizaciones Organizativas Panorama Papel Política Privada Profesional Pública Siglo XXI

A modo de conclusión

Los lenguajes documentales se tipifican según:

- **La naturaleza de los términos:** los términos pueden expresarse en lenguaje codificado o natural. Entendemos por codificado el uso de un código artificial compuesto de números, letras y símbolos que traducen un concepto. Entendemos por natural el uso de palabras del lenguaje usual, habitual, no códigos.
- **El nivel de control del vocabulario:** los lenguajes pueden ser libres o controlados. Los lenguajes libres son listas de términos extraídos del lenguaje natural. Consideramos lenguajes controlados aquellos que están redactados previamente en forma de listas o listados de términos que se consideran aceptados y unívocos para la indización. Sólo los términos de la lista se pueden usar para indizar.
- **El nivel de coordinación:** precoordinado o postcoordinado. La precoordinación consiste en determinar a priori cómo se combinan los términos, ya sea a la hora de construir el lenguaje, a la hora de indizar el documento o a la hora de recuperarlo. La postcoordinación consiste en no establecer reglas a la hora de la indización y combinar los términos de indización en el momento de la recuperación siguiendo la lógica de los operadores booleanos.
- **La estructura:** el vocabulario de los lenguajes documentales se organiza en dos estructuras: jerárquica o combinatoria. En la estructura jerárquica o arborescente, el vocabulario se presenta en forma de cadena, con términos genéricos que agrupan términos más específicos. En la estructura combinatoria, los términos no forman cadena, están listados por orden alfabético.
- **El nivel de análisis:** materias, conceptos, palabras clave. Indizar por materias consiste en indizar la materia principal del documento. Indizar por conceptos significa indizar las ideas y nociones del texto. Indizar por palabras clave significa indizar todas las palabras con significado del texto. Es el proceso más analítico que existe. No es una tarea de indización humana, sino automática.

5.6. Conclusiones

El estudio de las tipologías de los lenguajes documentales permite elaborar la ficha descriptiva de cada uno.

Fichas descriptivas de cada lenguaje documental

Sistemas de clasificación	Listado de encabezamiento de materias	Listado de autoridades
<ul style="list-style-type: none"> • Sintético por materias • Símbolos de clase o notaciones • Humana • Codificado • Controlado • Precoordinado • Jerárquico 	<ul style="list-style-type: none"> • Sintético por materias • Encabezamientos • Humana • Natural • Controlado • Precoordinado • Combinatorio 	<ul style="list-style-type: none"> • Analítico por conceptos • Identificadores y descriptores • Humana • Natural • Controlado • Postcoordinado • Combinatorio
Tesaurus	Listado de descriptores libres	Listado de palabras clave
<ul style="list-style-type: none"> • Analítico por conceptos • Descriptores • Humana • Natural • Controlado • Postcoordinado • Jerárquico • Combinatorio 	<ul style="list-style-type: none"> • Analítico por conceptos • Descriptores • Humana • Natural • Libre • Postcoordinado • Combinatorio 	<ul style="list-style-type: none"> • Analítico por palabras clave • Palabras clave • Automática • Natural • Libre • Postcoordinado • Combinatorio

Actividades

1. A partir del siguiente artículo elaborad un resumen informativo, uno indicativo, uno selectivo de conclusiones y uno automático que tenga una extensión parecida al informativo.

VALLEZ, M; PEDRAZA-JIMÉNEZ, R. "El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines" [en línea en <http://www.hipertext.net/web/pag277.htm>]. *Hipertext.net*, núm. 5, 2007. ISSN 1695-5498.

2. Indizad el mismo artículo con los tres niveles de exhaustividad. Argumentad en qué tipo de base de datos y SID podría ser útil cada uno.

3. Proponed dos títulos de documentos, reales o inventados, donde la materia se exprese a través de dos sinónimos.

4. Imaginad dos títulos más donde aparezcan dos polisémicos y proponed una manera de diferenciarlos. Buscad el origen etimológico de las palabras y decid si son polisémicas u homónimas.

5. Responded las siguientes afirmaciones justificando la solución:

- a) ¿Todo lenguaje controlado es codificado?
- b) ¿Todo lenguaje precoordinado es controlado?
- c) ¿Todo lenguaje libre es natural?
- d) ¿El lenguaje que tiene la tasa de coherencia más elevada es el Listado de palabras clave?

6. El siguiente texto es un compendio de errores y medias verdades. ¿Sabrías localizarlas y argumentar por qué no son correctas?

Usar lenguajes naturales en la indización y recuperación permite una buena comunicación documental. Los sistemas de clasificación representan la materia de los documentos a través de múltiples notaciones. Los lenguajes que indizan por materias son los tesauros y las listas de encabezamientos de materia. Para recuperar de manera precisa tenemos que utilizar sistemas de clasificación y listados de encabezamientos de materia. Los lenguajes controlados son muy amigables para el analista y el usuario. Los lenguajes precoordinados permiten ordenar los documentos en las estanterías.

Glosario

abstract *m* Terminología anglosajona para los resúmenes redactados por personas.

anáforas *f* Relación de referencia entre un elemento lingüístico y uno anterior en el discurso.

análisis de contenido *m* Operaciones de análisis que identifican y representan de manera precisa la materia de los documentos, con el objetivo de permitir la recuperación. Las operaciones son dos: el resumen y la indización. Esta parte del análisis documental establece el punto de acceso por materias.

análisis morfosintáctico *m* Análisis que determina la categoría léxica de cada palabra: sustantivo, verbo, adjetivo, artículo, preposición, etc. También determina el lema. Estas operaciones permiten distinguir las palabras con significado (sustantivos, adjetivos, verbos) de las vacías (artículos, preposiciones, pronombres, etc.). El lema permite agrupar todas las palabras que son flexiones de otra (info/informar/información/informador/informacional/etc.).

autoridad *f* Término de indización propio del lenguaje documental Listado de Autoridades. También se conocen con el nombre de identificadores y descriptores.

codificado *adj.* Tipología de lenguaje documental consistente en el uso de un código artificial compuesto de números, letras y símbolos que traducen un concepto. Sólo hay un tipo de lenguaje codificado, son los sistemas de clasificación.

combinatoria *f* Tipología de lenguaje documental consistente en estructurar los términos de indización por orden alfabético. La estructura combinatoria permite la inclusión de nuevos términos y la eliminación de los obsoletos sin afectar al resto de la estructura del lenguaje. Los lenguajes de estructura combinatoria son cinco: las listas de encabezamientos de materia, los listados de autoridades, los tesauros, el listado de descriptores libres y el listado de palabras clave.

controlado *adj.* Tipología de lenguaje documental consistente en listas de términos seleccionados tanto en su forma (plural, singular, sintagma nominal, adjetivado, siglas, etc.) como en su contenido (de todos los sinónimos se escoge uno, los homónimos se diferencian entre ellos, etc.) y como en sus relaciones de jerarquía y asociación (términos conceptualmente más genéricos o específicos y términos que se evocan mutuamente). Requieren unos gastos de construcción elevados, tanto en personal cualificado como en tiempo. Son los verdaderos lenguajes documentales. También se conocen por el nombre de lenguajes artificiales. Su función documental es la de representar un concepto con un único término y que sólo haya un término por concepto, lo que se conoce como univocidad. Los lenguajes controlados son cuatro: los sistemas de clasificación, las listas de encabezamientos, listados de autoridades y tesauros.

describir el contenido *loc. v.* *Ved* representar el contenido

descripción característica *f* *Ved* indización.

descriptor *m* Término de indización propio de tres lenguajes documentales: listado de autoridades, tesauros, listado de descriptores libres.

encabezamiento *m* Término de indización propio del lenguaje documental de las listas de encabezamientos de materia.

entropía *f* Calidad aplicable a los lenguajes documentales que tienden a la selección, a la restricción del vocabulario. Es el proceso contrario al lenguaje natural que tiende a la abundancia, a la reiteración de conceptos, a la sinonimia en beneficio de una expresión más rica.

especificidad *f* Criterio relacionado con la exactitud en que un concepto particular que aparece en un documento está representado por un término de indización.

estructura *f* Tipología de los lenguajes documentales que los clasifica en jerárquicos o combinatorios.

examen del documento *m* Primera fase del proceso de indización consistente en la lectura del título, resumen, sumario, introducción, ilustraciones y palabras o frases destacadas en una tipografía diferente.

exhaustividad *f* Criterio relacionado con el número de conceptos que se tienen en cuenta para caracterizar el contenido entero de un documento. El principal criterio de selección es el

valor potencial del concepto para los usuarios de su SID. Podemos distinguir entre exhaustividad baja, media y alta en función del número de descriptores.

extract *m* Terminología anglosajona para los resúmenes automáticos. Los *extracts* son los resúmenes formados a partir de la extracción de algunas frases del texto previamente seleccionadas por un programa.

fichero inverso *m* Fichero donde se almacenan todos los términos de indización. Estos se sitúan uno detrás del otro de forma secuencial y asociados al documento al que hacen referencia.

hiperónimo *adj.* Decimos que una palabra es hiperónima cuando tiene un campo significativo que incluye otro de menor extensión. Ejemplo: color es un hiperónimo con respecto a amarillo, naranja, verde...

hipónimo *adj.* Decimos que una palabra es hipónima cuando tiene un campo significativo que queda incluido en otro de mayor extensión. Ejemplo: amarillo, naranja, verde son hipónimos ya que pertenecen al término color.

homonimia *f* Tipo de polisemia. Se da cuando dos conceptos diferentes han llegado a tener el mismo nombre, la misma forma, pero vienen de orígenes diferentes y por lo tanto tienen etimologías diferentes.

identificador *m* Término de indización propio del lenguaje documental Listado de autoridades. También se conocen con el nombre de autoridad y descriptores.

indización *f* Acción de describir o identificar un documento en relación a su contenido. Norma UNE 50-121-91. Indizar es el resultado de examinar el documento, seleccionar los conceptos y almacenarlos en una base de datos. Esta definición implica tres acciones, de las cuales la más significativa es la selección de los conceptos y su traducción al lenguaje documental.

indización por conceptos *loc. v.* Indización de las ideas y nociones del texto, sin reducirlo a un tema principal. Responden a la pregunta “¿cuáles son los conceptos de este documento?”, van ligados necesariamente a sistemas automatizados. Los lenguajes que indizan por conceptos son tres: listados de autoridades, tesauros, listados de descriptores libres.

indización por materias *loc. v.* Indización sintética. Responden a la pregunta “¿cuál es el tema de este documento?”. Los lenguajes que indizan por materias son dos, los sistemas de clasificación y las listas de encabezamientos de materia.

indización por palabras clave *loc. v.* Indización de todas las palabras con significado del texto. Es el proceso más analítico que existe. No es una tarea de indización humana, sino automática. Los programas que indizan por palabras clave seleccionan sólo las palabras que tienen significado (preferentemente sustantivos). Sólo hay un lenguaje por palabras clave, y es evidentemente el único lenguaje automático, el listado de palabras clave.

ISO 214: 1976 *f* Norma internacional, traducida por AENOR como norma UNE 50-103-90 *Preparación de resúmenes.*

jerárquica *adj.* Tipología de lenguaje documental consistente en estructurar los términos de indización de forma arborescente. El vocabulario se presenta en forma de cadena, con términos genéricos que agrupan términos más específicos. Todos los términos dependen de un término superior y de significado más genérico. Esta estructura permite agrupar los conceptos por temas.

lenguaje artificial *m* *Ved* controlado.

lenguaje documental *m* Vocabulario de términos en lenguaje natural o un sistema artificial de signos normalizados que facilitan la representación del contenido de los documentos. Sus funciones principales son indizar el contenido de los documentos y permitir la recuperación a partir del campo materia.

lenguaje natural *m* Lenguaje que usamos de forma cotidiana: catalán, castellano, vasco, gallego, francés, etc.

libre *adj.* Tipología de lenguaje documental consistente en listas de términos extraídos del lenguaje natural sin formar parte de ningún listado establecido a priori, ni haber pasado un proceso de control de su vocabulario. Los lenguajes libres son dos: los listados de descriptores libres y el listado de palabras clave.

listado de autoridades *m* Lenguaje documental. Analítico por conceptos, natural, controlado, postcoordinado y combinatorio. Su término de indexación se conoce como identificador, autoridad o descriptor.

listado de descriptores libres *m* Lenguaje documental. Analítico por conceptos, natural, libre, postcoordinado y combinatorio. Su término de indexación se conoce como descriptor.

listado de encabezamientos de materia *m* Lenguaje documental. Sintético por materias, natural, controlado, precoordinado y combinatorio. Su término de indexación se conoce como encabezamiento.

listado de palabras clave *m* Lenguaje documental. Analítico por palabras clave, natural, libre, postcoordinado y combinatorio. Su término de indexación se conoce como palabra clave.

natural *adj.* Tipología de lenguaje documental consistente en el uso de palabras del lenguaje usual, habitual, no códigos. Hay cinco lenguajes documentales naturales: las listas de encabezamientos de materia, los listados de autoridades, los tesauros, los listados de descriptores libres y los listados de palabras clave.

naturaleza de los lenguajes *f* Tipología de los lenguajes documentales que los clasifica en codificados o naturales.

nivel de análisis *m* Tipología de los lenguajes documentales que los clasifica en lenguajes de materias, conceptos y palabras clave.

nivel de control *m* Tipología de los lenguajes documentales que los clasifica en libres o controlados.

nivel de coordinación *m* Tipología de los lenguajes documentales que los clasifica en precoordinados o postcoordinados.

notación *f* Término de indexación propio del lenguaje documental de los sistemas de clasificación.

palabra clave *f* Término de indexación propio del lenguaje documental de las palabras clave o indexación automática. Palabra o grupo de palabras seleccionadas de manera automática del título, resumen o texto de un documento que representan el contenido y permiten la recuperación.

palabra vacía *f* Palabra sin significado en las operaciones de indexación y resumen. Son preposiciones, artículos, verbos, adverbios, etc.

polisemia *f* Propiedad de un signo lingüístico de tener más de un significado. Decimos que dos palabras son polisémicas cuando el mismo signo lingüístico, palabra o sonido, tiene más de un significado. La palabra tiene un único origen etimológico y acaba teniendo significados diferentes sin cambiar su categoría gramatical.

ponderación (de frases, de palabras) *f* Método que evalúa las frases y las palabras de un texto en función de parámetros como la frecuencia, la presencia de palabras indicativas (buscan palabras como *importante*, *esencial*, *conclusiones*, etc.), la aparición en lugares destacados, por ejemplo el título: al principio de cada párrafo, al final a modo de conclusiones, etc.

postcoordinación *f* Tipología de lenguaje documental consistente en combinar los términos de indexación en el momento de la recuperación. Los lenguajes postcoordinados sólo tienen sentido en sistemas documentales automatizados que dispongan de un fichero inverso. Los lenguajes postcoordinados son cuatro: listados de autoridades, tesauros, listados de descriptores libres y listados de palabras clave.

precoordinación *f* Tipología de lenguaje documental consistente en determinar a priori cómo se combinan los términos, ya sea a la hora de construir el lenguaje, de indexar el documento, o de recuperarlo. Los dos lenguajes precoordinados son los sistemas de clasificación y las listas de encabezamientos de materia.

procesamiento en lenguaje natural (PLN) *m* Rama de la inteligencia artificial y de la lingüística computacional que estudia los lenguajes que usan los humanos para interactuar con los ordenadores en contextos escritos y orales. EL PLN estudia cómo emular el conocimiento humano, en cuanto a la identificación de los conceptos y frases con contenido relevante.

relación de significado *f* *Ved* relación semántica.

relación semántica *f* Relaciones de significado de las palabras. Las relaciones pueden ser de tipo genérico, específico o relacionado de un término con respecto a otro. En lenguaje natural estas relaciones son implícitas pero en un lenguaje documental hay que definir estas relaciones, agrupando y relacionando los términos afines.

representar el contenido *loc. v.* Expresión que significa describir el tema o los temas de un documento.

resumen *m* Presentación abreviada y precisa de un documento, sin interpretación ni crítica y sin mención expresa del autor del resumen. Norma UNE 50-103-90 *Preparación de resúmenes*.

resumen indicativo *m* Resumen que consigna sólo las ideas centrales del documento. Su lectura no puede sustituir la lectura del original.

resumen informativo *m* Resumen que consigna el tema central, temas adicionales, naturaleza y objetivo del documento, metodología, resultados, conclusiones y anexos. La idea de fondo es que un resumen informativo puede sustituir en ocasiones la lectura del documento original.

resumen selectivo *m* Resumen que consigna sólo una parte concreta del documento. El más habitual es el resumen de conclusiones, pero también hay otros tipos, como la reseña (*review*).

selección de los términos de indización *f* Segunda fase en el proceso de indización consistente en identificar las nociones que son elementos esenciales de la descripción del contenido. Los criterios de selección son el número de conceptos (criterio de exhaustividad) y la exactitud de los mismos (criterio de especificidad).

símbolo de clase *m* Ved Notación.

sinonimia *f* Palabras que tienen el mismo significado. Ejemplo: alimento, nutriente, comida, provisión. En un sistema documental, si no se controlan y se usan indiscriminadamente, comportan silencio documental.

sistema de clasificación *m* Lenguaje documental. Sintético por materias, codificado, controlado, precoordinado y jerárquico. Su término de indización se conoce como notación o símbolo de clase.

término de indización *m* Representación de un concepto en lenguaje natural o un código de clasificación. Los términos de indización pueden estar formados por una palabra o más de una.

tesauro *m* Lenguaje documental. Analítico por conceptos, natural, controlado, postcoordinado, jerárquico y combinatorio. Su término de indización se conoce como descriptor.

traducción a un lenguaje documental controlado *f* Buscar un concepto expresado en lenguaje natural en el listado de términos de un lenguaje documental controlado. Utilizar el término controlado para indizar y recuperar.

UNE 50-103-90 Preparación de resúmenes *f* Norma española que establece las directrices que se tienen que seguir para presentar los resúmenes en los documentos. Pone especial énfasis en la preparación de resúmenes por parte de los autores de los documentos primarios y en la misma publicación.

UNE 50-113-92/1 *f* Norma española titulada Documentación e información. Vocabulario. Parte 1. Conceptos fundamentales. En: *Documentación: Normas fundamentales*. Madrid: AENOR, 1994.

UNE 50-121-91 *f* Norma española titulada *Métodos para el análisis de documentos, determinación de su contenido y selección de términos de indización*. Baza el proceso de indización en tres fases: examinar el documento para identificar su contenido, seleccionar los conceptos principales de los contenidos y traducir a un lenguaje documental.

unitérmino *m* La parte más pequeña con significado de un término de indización. La norma UNE 50-113-92/1 define los unitérminos como el elemento significativo más pequeño de un lenguaje documental utilizado para representar un concepto específico en un sistema de indización coordinado; no se debe confundir con palabra clave o descriptor.

univocidad *f* Representar un concepto con un único término.

Bibliografía

Bibliografía sobre el resumen

AENOR (1990). *Documentación. Preparación de resúmenes. UNE 50 103 90*. Madrid: AENOR.

Climent, Salvador. "Sistemes de resum automàtic de documents". *Digit. Hum. Revista digital d'humanitats*. ISSN 1575-2275.

Lloret, E.; Ferrández, O.; Muñoz, R.; Palomar, M. (2008). "Integración del reconocimiento de la implicación textual en tareas automáticas de resúmenes de textos". *Procesamiento del lenguaje natural*, núm. 41, pág. 183-190.

Mateo, P. L.; González, J. C.; Villena, J.; Martínez, J. L. (2003). Un sistema para resumen automático de textos en castellano.

Pinto Molina, M. (1992). *El resumen documental: principios y métodos*. Madrid: Pirámide/Fundación Germán Sánchez Ruipérez (Biblioteca del Libro, Y).

Bibliografía sobre la indización

Abadal, E.; Codina, L. (2005). "Recuperación de Información". En: *Bases de Datos Documentales: Características, funciones y método* (cap. 2. p. 29-92). Madrid: Síntesis.

AENOR (1997). *Métodos para el análisis de los documentos, determinación de su contenido y selección de los términos de indización. Norma UNE 50-121-91*. Madrid: AENOR.

AENOR (1997). "Documentación e información. Vocabulario. Parte 6: lenguajes documentales". *Revista Española de Documentación Científica*, Norma UNE-50-113/6 (ISO 5127/6), vol. 20, núm. 4, pág. 417-436.

Cid, P.; Cuadrado, M.; Aguiriano, C. (1999). *Fonaments de llenguatges documentals*. [Documento electrònic]. Barcelona: UOC.

Codina, L. (1994). "El papel del lenguaje natural en los sistemas multimedia: una reflexión sobre la tecno-simpleza y la ciber-ingenuidad". *Cuadernos de documentación multimedia*, núm. 3 (junio).

Gil Leiva, I. (2008). *Manual de indización. Teoría y práctica*. Gijón: Ediciones Trea (Biblioteconomía y Administración cultural, 193).

Gil, I.; Rodríguez Muñoz, J. V. (1996). "El Procesamiento del lenguaje natural aplicado al análisis del contenido de los documentos". *Revista general de información y documentación*, vol. 6, núm. 2, pág. 205-218.

Gil Urdiciain, B. (1992). "Función de los lenguajes documentales en el tratamiento de la información en las organizaciones". *Revista general de información y documentación*, vol. 2, Núm. 2, pág. 195-200.

Gil Urdiciain, B. (2004). *Manual de lenguajes documentales*. Gijón: Ediciones Trea (Biblioteconomía y Administración cultural, 106).

Norma UNE 50-113-92/1. *Documentación e información. Vocabulario. Parte 1. Conceptos fundamentales* (1994). En: *Documentación: Normas fundamentales*. Madrid: AENOR.

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

Sistemas de clasificación documentales

Manela Juncà Campdepadrós

PID_00143962



Universitat Oberta
de Catalunya

www.uoc.edu



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. Fundació 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

Introducción	5
Objetivos	8
1. Clasificar y sistemas de clasificación	9
1.1. Clasificaciones bibliográficas o documentales	10
1.1.1. ¿Qué son las clasificaciones bibliográficas o documentales?	10
1.1.2. ¿Dónde se aplican?	10
1.1.3. Funciones	11
1.1.4. ¿Cuántas clasificaciones hay?	12
1.1.5. Conclusiones	12
1.2. Clasificaciones documentales: tipología	14
1.2.1. Por alcance de contenido	14
1.2.2. Por estructura	14
1.2.3. Conclusiones	16
1.3. Breve historia de las clasificaciones bibliográficas	17
1.3.1. La clasificación decimal de Dewey, de 1885	19
1.3.2. La clasificación decimal universal, de 1905	21
1.3.3. La clasificación Colon, de 1933	22
1.3.4. Otras clasificaciones documentales de los siglos XIX y XX.....	24
1.3.5. Los sistemas de clasificación en España y Cataluña	25
1.3.6. Conclusiones	27
2. La clasificación decimal universal	29
2.1. Características de la clasificación decimal universal	29
2.2. Ediciones de la CDU	31
2.3. Gestión de la CDU	33
2.3.1. ¿Quién es el responsable de la gestión de la CDU?	33
2.3.2. ¿Se actualizan las clases de la CDU?	34
2.3.3. Nuevas ediciones de la CDU y reclasificación de los fondos	34
2.3.4. Conclusiones	35
3. Estructura de la clasificación decimal universal	36
3.1. Clases y notaciones	37
3.1.1. Clases	37
3.1.2. Notación	39
3.1.3. Conclusiones	40
3.2. Tablas principales	41

3.2.1. Tablas principales de la edición de bolsillo CDU 2004	41
3.3. Tablas auxiliares de la clasificación decimal universal (CDU)	54
3.3.1. Tablas de auxiliares comunes	56
3.3.2. Tablas de auxiliares especiales	65
3.4. Signos con función combinatoria y signos con función instructiva	70
3.4.1. La importancia de los signos	70
3.4.2. Los signos con función combinatoria	72
3.4.3. Los signos puramente instructivos	73
3.4.4. Conclusiones	73
4. Empezamos a clasificar.....	75
5. Prácticas con la CDU.....	78
5.1. Ejercicios de nivel bajo	78
5.2. Ejercicios de nivel medio	80
5.3. Ejercicios de nivel alto	81
6. Creación de un sistema de clasificación propio.....	83
6.1. Estudio previo	83
6.2. Metodología	84
6.3. Proceso de creación	84
6.4. Mantenimiento	86
Glosario.....	90
Bibliografía.....	94

Introducción

El módulo “Sistemas de clasificación” está pensado para estudiantes que no tengan conocimientos previos sobre el mundo de las clasificaciones documentales. Se trata, pues, de un módulo introductorio al estudio de los diferentes sistemas de clasificación documentales y, de manera especial, de la clasificación decimal universal (CDU), que es el sistema más aplicado en nuestro ámbito cultural europeo.

Itinerario de estudio

El módulo está concebido como una aproximación del concepto más general al más concreto. La secuencia es:

Sistemas de clasificación > Clasificaciones bibliográficas o documentales > Clasificación decimal universal.

Completaremos la visión del tema aportando las técnicas más interesantes de Jacques Maniez (1993) sobre la construcción de un sistema nuevo, muy adecuado para aquellos documentalistas que tengan que crear un sistema a medida.

El módulo es teórico, normativo y en mayor medida, práctico, ya que uno de los objetivos que se proponen es conseguir una cierta habilidad en la clasificación de documentos con la CDU.

Conceptos más importantes

Concepto	Ved
Clasificación documental	1.1. Clasificaciones bibliográficas 1.2. Clasificaciones bibliográficas: tipología 1.3. Breve historia de las clasificaciones bibliográficas
Clases	3.1. Clases y notación 3.2. Tablas principales 3.3. Tablas auxiliares de la CDU
Notación	3.1. Clases y notación 3.4. Signos: combinatorios e instructivos
Auxiliares comunes	3.3.1. Tablas de auxiliares comunes
Auxiliares especiales	3.3.2. Tablas de auxiliares especiales
Clasificar con CDU	4. Empezamos a clasificar 5. Prácticas con la CDU

Ved también

Para construir un cuadro de clasificación, ved el apartado 6 de este módulo didáctico.

Bibliografía y recursos web

Presentamos la bibliografía en dos secciones:

- una selección de fuentes de información indispensables para los estudiantes que se introducen en el estudio de las clasificaciones.
- la lista entera a partir de la cual se ha hecho la selección anterior.

Selección bibliográfica

Es indispensable la consulta de la clasificación decimal universal –en edición de bolsillo o en CD.

La tabla siguiente muestra que las obras de María Pinto (ed.), Jacques Maniez e I. C. McIlwaine son muy indicadas con respecto a la teoría de las clasificaciones; y que obras como las Miguel Benito, Carmen Díez-Carrera e I. McIlwaine lo son para las prácticas.

Obras recomendadas

Bibliografía	Teoría/historia de las clasificaciones	Estructura de la CDU	Ejercicios de CDU con solución
<i>Clasificación Decimal Universal (CDU) de bolsillo</i> (2004). Madrid: AENOR.		Sí	Sí
Benito, M. (1999). <i>El sistema de clasificación decimal universal: manual de aprendizaje</i> . Madrid: Taranco.		Sí	Sí
Díez Carrera, C. (1999). <i>Técnicas y régimen de uso de la CDU (Clasificación Decimal Universal)</i> (pág. 134). Gijón: Trea ("Biblioteconomía y Administración Cultural", 26).		Sí	Sí
Maniez, J. (1992). <i>Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales</i> . Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.	Sí		
McIlwaine, I. C. (2003). <i>Clasificación Decimal Universal. Guía para uso de la CDU</i> . Madrid: AENOR.		Sí	Sí
Pinto Molina, M. (ed.) (1997). <i>Manual de clasificación documental</i> . Madrid: Síntesis ("Biblioteconomía y Documentación").	Sí	Sí	Sí

Bibliografía	Teoría/historia de las clasificaciones	Estructura de la CDU	Ejercicios de CDU con solución
San Segundo Manuel, R. (1997). "Principales sistemas de clasificación". A: M. Pinto (ed.). <i>Manual de clasificación documental</i> (pág. 67-89). Madrid: Síntesis "Biblioteconomía y Documentación".	Sí		

Selección recursos web

Por lo que respecta a los recursos web, no podemos dejar de seleccionar las webs de las sedes oficiales de los sistemas de clasificación más aplicados en el mundo –La CDU y la Dewey– y los tutoriales sobre la CDU.

Clasificación decimal universal (web oficial, blog):

- *Universal Decimal Classification Consortium Homepage*. La Haya: UDC Consortium Act. 2002-08-01 [en línea]. [Consulta: 10 de octubre de 2008].
- *Universal Decimal Classification Blog*. UDC Consortium [en línea]. [Consulta: 10 de octubre de 2008].

Clasificación Dewey (actualizaciones, clases, blog):

- OCLC Online Computer Library Center. *Dewey services, Dewey decimal classification for use with OCLC's online cataloging services*. [Consulta: 10 de octubre de 2008]. [En línea].
- *025.431: The Dewey blog* [Consulta: 10 de octubre de 2008]. [En línea].
- **Public libraries using spanish**. *Spanish version of the Dewey Decimal Classification System based on the 18th edition of DDC* [Consulta: 10 de octubre de 2008]. [En línea].
Cabe destacar las páginas personales y tutoriales de Miguel Benito, Ferran Burguillos y Enric Mestre para sus exposiciones pedagógicas:
- **Benito, Miguel**. *El sistema de Clasificación Decimal Universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008]. O, también, en la URL Instituto de Taranco [Fecha de consulta: 10 de octubre de 2008].
- **Burguillos Martínez, Ferran**. *Tutorial de la Clasificación decimal universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008].
- **Mestre i Ribera, Enric**. *Classificació decimal universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008].

Objetivos

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

1. Definir los sistemas de clasificación documental y su tipología.
2. Presentar la evolución histórica de estos lenguajes documentales.
3. Conocer la clasificación decimal universal: características, estructura y sintaxis.
4. Adquirir una cierta habilidad en la clasificación de documentos con la CDU.
5. Introducirse en la creación de un sistema de clasificación nuevo.

1. Clasificar y sistemas de clasificación

Clasificar es distribuir en clases o categorías.

Hay tres tipos de sistemas de clasificación (Maniez, 1992, pág. 23).

1) Las **clasificaciones filosófico-científicas** que clasifican el conocimiento humano. Catalogan las leyes universales y los fenómenos naturales. Landgridge las califica de formas primarias, ya que observan directamente el mundo y la naturaleza del conocimiento.

Clasificaciones filosófico-científicas

Sistemas de Porfirio, Platón, Aristóteles, F. Bacon, A. Comte, Leibniz.

2) Las **clasificaciones biológicas o taxonomías** clasifican organismos vivos e inertes en una jerarquía. Actualmente, el término *taxonomía* se aplica también en el mundo de la información y las fuentes disponibles en el WWW.

Clasificaciones biológicas o taxonomías

Elementos químicos, plantas, animales...

3) Las **clasificaciones bibliográficas o documentales** clasifican el fondo de un servicio de información y documentación (SID¹). Para Landgridge, tratan del conocimiento cuando ya se ha fijado en los documentos; por lo tanto, no todo el conocimiento, sino sólo aquel que se ha fijado en un soporte material.

A grandes rasgos, las diferencias más importantes entre las clasificaciones filosófico-científicas y las documentales son que las primeras buscan organizar el conocimiento, mientras que las segundas quieren representar el contenido de los documentos y, además, las primeras son teóricas y las segundas prácticas (María Pinto; Carmen Gálvez, 1997, pág. 35).

A pesar de todo, no podemos negar que las primeras influyen en las últimas y que las clasificaciones documentales están íntimamente influidas por las sucesivas concepciones del conocimiento. De hecho, las clasificaciones son fruto de una sociedad y de un tiempo determinado.

La influencia de Francis Bacon

El sistema filosófico de Francis Bacon influye en la clasificación bibliográfica de Brunet (Francia, 1804) y en la clasificación bibliográfica de Harris de 1870.

Lectura complementaria

Sobre los tres tipos de clasificaciones mencionados aquí, podéis consultar la obra siguiente:

J. Maniez (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.

⁽¹⁾SID es la sigla de servicio de información y documentación.

A partir de este momento, el módulo se centrará en las clasificaciones de tipo bibliográfico o documental.

A modo de conclusión

Clasificar es distribuir en clases o categorías. Hay tres tipos de sistemas de clasificación: los filosófico-científicos, que clasifican el conocimiento humano de manera genérica; los taxonómicos, que clasifican organismos vivos e inertes; y los bibliográficos o documentales, que clasifican el fondo documental de un SID.

Bibliografía

B. Gil Urdiciain (1997). "Los lenguajes de clasificación documental". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 91-108). Madrid: Síntesis ("Biblioteconomía y Documentación").

J. Maniez (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.

M. Pinto Molina (ed.) (1997). *Manual de clasificación documental*. Madrid: Síntesis ("Biblioteconomía y Documentación").

1.1. Clasificaciones bibliográficas o documentales

Primero definiremos qué son, dónde se aplican, qué funciones tienen y cuántas hay en el mundo. En segundo lugar, trataremos la diversa tipología de las clasificaciones bibliográficas. En tercer y último lugar, haremos un pequeño viaje por la historia de las clasificaciones bibliográficas. ¡Empezamos!

1.1.1. ¿Qué son las clasificaciones bibliográficas o documentales?

Las clasificaciones bibliográficas o documentales son un lenguaje documental que divide el conocimiento (enciclopédico o específico de un área temática) en una serie ordenada de clases y subclases basadas en las semejanzas y diferencias.

La estructura jerárquica o facetada de estas clasificaciones refleja las relaciones y el contexto del concepto.

Turismo económico o turismo de ocio

Si el concepto *turismo* depende jerárquicamente de la clase *economía*, entenderemos que hace referencia al turismo como factor económico. Si, en cambio, depende jerárquicamente de la clase *tiempo libre*, entenderemos que se refiere al turismo como un elemento de ocio y descanso. En los dos casos, la estructura refleja el contexto del concepto, a diferencia de una lista simplemente ordenada de términos de indexación.

1.1.2. ¿Dónde se aplican?

Todos los servicios de información y documentación, SID², (archivos, bibliotecas y centros de documentación) utilizan algún sistema de clasificación, pero difieren bastante unos de otros. En palabras de Miguel Ángel Esteban:

⁽²⁾A partir de ahora denotamos servicios de información y documentación con la sigla SID.

1) Uso de los sistemas de clasificación en los archivos

“En los archivos se presenta como la identificación de agrupaciones documentales naturales mediante el auxilio de un sistema de clasificación que reproduce la estructura orgánica o el entramado funcional de la institución generadora del fondo como resultado de su actividad natural para el cumplimiento de sus fines.”

2) Uso de los sistemas de clasificación en bibliotecas

“La clasificación se identifica en las bibliotecas con la agrupación intelectual e incluso física de los documentos en clases que mantengan entre sí relaciones de superioridad y dependencia, a partir de la extracción y representación codificada de su contenido temático principal (materia) de acuerdo con las reglas de un sistema de clasificación jerárquico o facetado.”

Uso de los sistemas de clasificación en los centros de documentación

“En los centros de documentación se considera como una actividad innecesaria habitualmente porque tiene prioridad el análisis conceptual del contenido de los documentos, si bien puede estar presente como una función auxiliar para el control de la diversidad de la colección y la reducción del área de busca con el uso de los niveles superiores de un tesauro.”

Añadiremos el uso de los sistemas de clasificación para organizar el conocimiento del WWW: tal como dice I. McIlwaine, los sistemas como la CDU pueden tener dos utilidades:

- Organizar y visualizar recursos de información dentro de los servicios de información de calidad.
- Clasificar metadatos de documentos electrónicos.

Los cuadros de clasificación utilizados en los archivos son por esencia originales y no extrapolables a otros fondos documentales, ya que responden a los procesos propios de la organización y se estudian en otras asignaturas. En consecuencia, este módulo se centrará en los sistemas utilizados en las bibliotecas.

1.1.3. Funciones

Los sistemas de clasificación cumplen las tres funciones siguientes:

- Representar el contenido de un documento de manera sintética y codificada.
- Ordenar físicamente los documentos con un grado altamente significativo.

Observación

Recordemos que podemos distinguir tres tipos de ordenaciones (Maniez, 1993, pág. 21):

Lectura complementaria

Sobre los sistemas de clasificación en los servicios de información y documentación según Miguel Ángel Esteban, podéis consultar la obra siguiente:

María Pinto Batanea (ed.) (1997). *Manual de clasificación documental* (pág. 20). Madrid: Síntesis (“Biblioteconomía y Documentación”).

Uso del sistema CDU

La CDU se utiliza en el portal de las Guías temáticas electrónicas del CBUC (Consortio de bibliotecas universitarias de Cataluña) [en línea].

Lectura recomendada

Para todos los interesados en las aplicaciones de la CDU en Internet, podéis consultar la obra siguiente:

I. C. McIlwaine (2003). *Guía para uso de la CDU* (pág. 258-272).

- 1) Ordenaciones no significativas, sin ninguna relación con el contenido de los documentos, por ejemplo, ordenaciones numéricas a partir del número de registro.
 - 2) Ordenaciones con significado limitado: por lengua, país, autor, disciplina, etc.
 - 3) Ordenaciones altamente significativas: ordenación por materias, de manera que los documentos que traten de contenidos afines se coloquen cerca.
- Generar subproductos del catálogo, sobre la base de su materia. Por ejemplo, listas bibliográficas, últimas adquisiciones ordenadas por temas...

1.1.4. ¿Cuántas clasificaciones hay?

En estos momentos, hay al menos nueve sistemas de clasificación documental vigentes e implementados en las redes de bibliotecas universitarias y/o públicas del mundo:

- 1) La clasificación decimal universal (CDU).
- 2) La clasificación Dewey (CD).
- 3) La clasificación de la Library of Congress (LC).
- 4) La clasificación china.
- 5) La clasificación japonesa.
- 6) La clasificación rusa (LBC, antigua BBK).
- 7) La clasificación Colon (CC).
- 8) La clasificación Bliss (CB).
- 9) La clasificación Brown.

Como dice Miguel Ángel Esteban (1997, pág. 26):

“No existe un sistema de clasificación esencial, natural o único que sea adecuado *a priori* a la naturaleza de la realidad. En consecuencia no tiene razón juzgar en abstracto si una clasificación es mejor que otra, sino que habrá que prestar atención al contexto en el cual se construyó y al propósito que persigue. Como máximo podemos mantener que un sistema clasificatorio es superior a otro porque puede aplicarse a un número mayor de contextos, porque revela un número mayor de relaciones entre los hechos clasificados y porque tiene una estructura más rigurosa.”

1.1.5. Conclusiones

Las clasificaciones bibliográficas son un lenguaje documental que representa el contenido, ordena físicamente los documentos (ordenaciones altamente significativas) y genera subproductos de catálogo (como las listas bibliográficas o las últimas adquisiciones ordenadas por temas).

Se usan en todo tipo de SID: archivos, bibliotecas, centros de documentación y servicios de la red.

A escala mundial, hay diversos sistemas de clasificación documental: clasificación decimal universal (CDU), clasificación Dewey (CD), clasificación de la Library of Congress (LC), clasificación china, clasificación japonesa, clasificación rusa (LBC, antigua BBK), clasificación Colon (CC), clasificación Bliss (CB), clasificación Brown.

En estos momentos de la lectura del módulo, os habéis de haber situado en la secuencia:

Clasificar > Sistemas de clasificación > Sistemas de clasificación documentales > Aplicación en las bibliotecas.

Más información

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

M. A. Esteban (1997). "Fundamentos epistemológicos de la clasificación documental". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 19-32). Madrid: Síntesis ("Biblioteconomía y Documentación").

M. Granados; A. Nicolau (2005, 6-8 de julio). "La recuperació de la informació en els catàlegs en línia: l'ús de la Classificació decimal universal i la seva implicació en la indexació" [en línea]. 7è. Congrés del Capítol Espanyol d'ISKO. Barcelona: Departament de Biblioteconomia i Documentació de la Universitat de Barcelona. [Consulta: 1 octubre de 2008].

Library of Congress Classification system [en línea]. [Fecha de consulta: 1 de octubre de 2008].

J. Maniez (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

Clasificación Dewey (actualizaciones, clases, blog)

OCLC Online Computer Library Center. *Dewey services, Dewey decimal classification for use with OCLC's online cataloging services* [en línea]. [Consulta: 10 de octubre de 2008].

M. Pinto Molina (ed.) (1997). *Manual de clasificación documental*. Madrid: Síntesis ("Biblioteconomía y Documentación").

R. San Segundo Manuel (1997). "Principales sistemas de clasificación". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 67-89). Madrid: Síntesis ("Biblioteconomía y Documentación").

025.431: *The Dewey blog* [en línea]. [Consulta: 10 de octubre de 2008].

Universal Decimal Classification Blog [en línea]. UDC Consortium. [Consulta: 10 de octubre de 2008].

Universal Decimal Classification Consortium Homepage. La Haya: UDC Consortium. Act. 2002-08-01 [en línea]. [Consulta: 10 de octubre de 2008].

1.2. Clasificaciones documentales: tipología

Las clasificaciones documentales se pueden agrupar según el alcance del contenido o la estructura.

Tipología de clasificaciones

Por alcance de contenido	Por estructura
<ul style="list-style-type: none"> • Clasificaciones enciclopédicas • Clasificaciones especializadas o sectoriales 	<ul style="list-style-type: none"> • Clasificaciones enumerativas o jerárquicas • Clasificaciones facetadas • Clasificaciones híbridas o mixtas

Esta tipología no es excluyente, ya que una clasificación es el resultado de la combinación de los elementos enciclopédica/sectorial con jerárquica/facetada/híbrida.

Tipología no excluyente
La CDU es una clasificación enciclopédica e híbrida.

1.2.1. Por alcance de contenido

Las clasificaciones enciclopédicas, como su nombre indica, recogen todas las disciplinas universales. Se usan principalmente en bibliotecas nacionales, universitarias y públicas.

Las tablas principales nos informan de qué tratan todas las áreas de conocimiento: filosofía, religión, ciencias sociales, ciencias exactas y naturales, tecnología, arte, lenguas y literatura, geografía e historia.

Clasificaciones enciclopédicas
Son ejemplos de clasificaciones enciclopédicas la clasificación decimal universal, la clasificación Dewey, la Library of Congress, la clasificación Bliss.

Las clasificaciones especializadas, como indica su nombre, hacen referencia sólo a una disciplina. También se las conoce por el nombre de clasificaciones sectoriales. Por su capacidad de especificación, se usan en centros de documentación.

Clasificaciones especializadas
Clasificaciones especializadas en el ámbito médico: clasificación internacional de enfermedades (CIE, *clasificación internacional de enfermedades*), *medical subject subheadings* –por la base de datos Medline.

1.2.2. Por estructura

Las clasificaciones tienen dos tipos de estructura: la **enumerativa** y la **facetada**. Las dos tienen ventajas e inconvenientes; por eso, nace una tercera estructura que es la combinación de las dos anteriores, con el objetivo de sumar ventajas y minimizar los inconvenientes, y se conoce con el nombre de **híbrida** o **mixta**.

Las clasificaciones enumerativas o jerárquicas estructuran las clases en forma de arborescencia.

Cada clase se va dividiendo en subclases tantas veces como sea necesario para representar el dominio entero o todos los supuestos posibles a la hora de clasificar un fondo. Las clases derivan de un punto común y se puede recorrer el camino desde el objeto más específico al más general, siguiendo uno solo. Cada concepto tiene un lugar, y sólo uno; por ello, también se las conoce como **clasificaciones monojerárquicas**.

Son enumerativas, es decir, enumeran de arriba abajo y de izquierda a derecha todas las materias previstas. Tienen una estructura lógica, van de general a específico y las clases son disjuntas, es decir, son conjuntos mutuamente excluyentes.

Son las más utilizadas porque son de fácil construcción y uso para el analista. También se adecuan muy bien en entornos enciclopédicos y universales.

Las **clasificaciones facetadas** están formadas por diversas clasificaciones jerárquicas autónomas a partir de una faceta (el periodo artístico, la técnica, etc.) que, cuando se combinan, describen la materia del documento de manera analítica. Son clasificaciones autónomas pero articuladas, combinables.

Las clasificaciones facetadas son más flexibles que las enumerativas. El principal teórico de las clasificaciones facetadas es Ranganathan, el autor de la clasificación Colon.

Las **facetadas** son principios de división basados en las características que las materias tienen en común. Se trata de agrupar los conceptos *según una característica concreta* que comparte con otras clases. Hay facetadas de tipo universal, aplicables a todos los campos del saber, y las hay propias de una materia.

Algunos ejemplos de facetadas

Facetas universales son el espacio, el tiempo.

Algunas facetadas propias de una materia son, para la materia Artes: *Periodo artístico, Técnica artística, Tema representado...* Así, *Románico, Gótico, Barroco, Modernismo*, serían clases de la faceta *Periodo artístico* que podríamos aplicar a cualquier tipo de arte. Por ejemplo: escultura modernista, pintura modernista, arquitectura modernista...

Las **clasificaciones híbridas** o **mixtas** son la combinación de las dos estructuras: la enumerativa con la facetada.

En este tipo de clasificaciones, las clases tienen estructura jerárquica y están formadas por:

La clasificación de la Library of Congress

La clasificación de la Library of Congress es un buen ejemplo de clasificación jerárquica y enumerativa.

Ved también

En el subapartado 1.3.3 se profundiza en el estudio de las facetadas universales o fórmula PMEST de Ranganathan.

- Clases principales para las grandes materias.
- Tablas auxiliares para facetas de tiempo, lugar, forma, lengua...
- Pequeñas tablas auxiliares especiales dentro de las clases, que se corresponden a las facetas propias de una materia. Ejemplo: facetas propias para astronomía.
- Signos clasificadores que permiten combinar más de una materia expresando la relación.

La CDU: una clasificación híbrida

La clasificación decimal universal (CDU) combina la estructura jerárquica de las tablas con los elementos facetados para representar conceptos más complejos y conceptos nuevos a partir de la relación de dos conceptos preexistentes.

17:57 Bioética → construido a partir de 17 Ética y 57 Biología

Para acabar, presentamos las ventajas y los inconvenientes que señalan B. Gil (1997, pág. 107) y J. Maniez (1993, pág. 40):

Ventajas e inconvenientes de las clasificaciones jerárquicas y facetadas

	Enumerativas o jerárquicas	Facetadas
Ventajas	La estructuración de un dominio en términos de más genérico a específico proporciona una visión completa, intuitiva y fácil para el clasificador, y también para el usuario que pasea por las estanterías en acceso libre.	Son analíticas y, por lo tanto, permiten describir con precisión el contenido de un documento. Son flexibles y no quedan desfasadas rápidamente. Son fáciles de automatizar, ya que los documentos se pueden buscar en conjunto o para cada faceta (tipo base de datos relacional).
Inconvenientes	Son poco flexibles a la hora de actualizar y de combinar diversas clases.	Su aplicación es compleja. Hay muchas materias que no se pueden representar fácilmente con facetas (conceptos de tipo mental, no objetos). No todos los documentos tienen todas las facetas (por ejemplo, la fórmula PMEST de Ranganathan), lo cual hace que la notación no sea homogénea. La notación es poco comprensible y memorizable.

1.2.3. Conclusiones

Las clasificaciones documentales admiten dos tipos de categorías:

- Por alcance de contenido: enciclopédicas o sectoriales.
- Por estructura: enumerativas o jerárquicas, facetadas, híbridas o mixtas.

Las clasificaciones jerárquicas estructuran las clases en forma de arborescencia. Cada clase se va dividiendo en subclases tantas veces como sea necesario para representar el dominio entero.

Las clasificaciones facetadas estructuran las clases, primero dominios de conocimiento, dentro de cada dominio por facetas, y en cada faceta de manera jerárquica.

Las clasificaciones híbridas participan de las características de las dos anteriores, combinando la estructura arborescente propia de la jerárquica con la posibilidad de expresar facetas con tablas auxiliares comunes y especiales y signos de valor clasificatorio.

Más información

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

B. Gil Urdiciain (1997). "Los lenguajes de clasificación documental". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 91-108). Madrid: Síntesis ("Biblioteconomía y Documentación").

Library of Congress Classification outline [en línea]. [Consulta: 1 de octubre de 2008].

J. Maniez (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.

Online Computer Library Center. *Dewey services, Dewey decimal classification for use with OCLC's online cataloging services* [en línea]. [Consulta: 10 de octubre de 2008].

R. San Segundo Manuel (1997). "Principales sistemas de clasificación". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 67-89). Madrid: Síntesis ("Biblioteconomía y Documentación").

Universal Decimal Classification Blog. UDC Consortium [en línea]. [Consulta: 10 de octubre de 2008].

Universal Decimal Classification Consortium Homepage. La Haya: UDC Consortium. Act. 2002-08-01 [en línea]. [Consulta: 10 de octubre de 2008].

1.3. Breve historia de las clasificaciones bibliográficas

Si hiciéramos un rápido viaje por el tiempo para conocer la historia de las clasificaciones documentales, veríamos que se fueron construyendo influidas por los sistemas filosóficos contemporáneos, y a medida que iban surgiendo las necesidades. Estas necesidades han sido tradicionalmente las siguientes:

a) Compilar el conocimiento documental en general. Es el caso de la CDU, que no nació para clasificar una biblioteca, sino para crear una bibliografía enciclopédica y mundial, el proyecto RBU (Repertorio Bibliográfico Universal).

b) Crear un instrumento práctico para clasificar y ordenar el fondo de una biblioteca concreta. Es el caso de la clasificación de la Biblioteca del Congreso y su fondo de dos millones de libros.

c) Facilitar el acceso libre a las estanterías.

d) Crear/adaptar el sistema de clasificación según la ideología imperante del momento, como por ejemplo, la influencia del marxismo y del leninismo en las clasificaciones de la antigua URSS y China.

En la historia de las clasificaciones documentales, también observamos que los cuadros de clasificación se basan unos en otros. No hay auténticas revoluciones, sino evoluciones.

Como dice Jacques Maniez (1987, pág. 151):

“En clasificación, como en cualquier disciplina, es inútil reinventar la rueda.”

Precisamente Aristóteles será el primer pensador que encontraremos en nuestro viaje, con la formulación de la lógica aristotélica. Es él y el árbol dicotómico de Porfirio.

Aristóteles y la lógica aristotélica

“Aristóteles y la lógica aristotélica: la lógica (que él llamaba *analítica*) no es una ciencia sustantiva, sino un instrumento (en griego *organon*). No es una simple lógica formal, sino toda una teoría de la ciencia. Distingue la deducción de la inducción y desarrolla una teoría bastante completa de una forma de razonamiento deductivo: el **silogismo**. Aunque la lógica aristotélica tiene limitaciones importantes (desconoce las proposiciones relacionales; considera sólo las predicativas, y reduce la deducción al silogismo), ha sido la aportación fundamental a esta ciencia hasta el siglo XIX.

Las ciencias, según Aristóteles, pueden ser de tres tipos:

- Teóricas: física, matemáticas y filosofía primera o teología.
- Prácticas: moral y política.
- Productivas: ciencias aplicadas.”

Gran Enciclopèdia Catalana

Porfirio y la dicotomía

“División en dos. Método de clasificación en que las divisiones y las subdivisiones son binarias.”

Gran Enciclopèdia Catalana

A continuación visitamos la Biblioteca de Alejandría (Egipto), fundada por Ptolomeo aproximadamente en el año 300 a. C., que catalogaba su colección de manuscritos en *pinakes* (fichas con información sobre el nombre del autor y el sumario del contenido).

En la Edad Media, encontramos bibliotecas en los monasterios y en las universidades. Los fondos documentales eran religiosos o de las materias propias del Trivium (gramática, retórica, dialéctica) y Quatrivium (aritmética, geometría, música, astronomía). En aquellos momentos, las clasificaciones bibliográficas estaban vinculadas a la colección que tenían que clasificar, no a una voluntad enciclopedista, criterio que no surgirá hasta el siglo XVIII.

No hay revoluciones, sino evoluciones

La clasificación filosófica científica de Francis Bacon sirve de base a Brunet y a Harris, que a la vez es la base en la que se inspira Dewey, que a su vez es copiado por Otlet y La Fontaine. Cutter inspira la Library of Congress, Ranganathan comparte cuerpo teórico con Aristóteles...

En la Edad Moderna, y gracias a las teorías de Francis Bacon (1561-1626), se acaban las concepciones medievalistas de la clasificación de las ciencias. Su filosofía se basa en la sustitución de la lógica deductiva medieval por un nuevo método, experimental e inductivo. De esta época es famosa la clasificación Brunet (Francia, 1804), que siguiendo a F. Bacon clasificaba el saber en cinco clases básicas: teología, jurisprudencia, historia, filosofía y literatura.

Y llegamos a la Edad Contemporánea, pero aquí reducimos el ritmo para conocer a cuatro documentalistas que han afirmado las bases de la clasificación moderna. Y para hacerlo, nos situaremos aproximadamente entre 1870 y 1940.

La clasificación de Harris

En 1870, un bibliotecario norteamericano llamado Harris publica su sistema de clasificación. Agrupa las clases en ciencias, artes e historia, un esquema que sigue directamente el estilo de las tres facultades de Bacon.

A caballo entre los siglos XIX y XX, se da una circunstancia clave que da gran importancia a la teoría de las clasificaciones: el incremento de las colecciones de las bibliotecas.

Una vez situados en el tiempo, nos ubicaremos geográficamente: primero iremos a Estados Unidos, a visitar a un joven de 22 años, Melvil Dewey, de Massachusetts, que tiene el honor de ser el creador de la clasificación más implementada por todo el mundo actualmente; después, a Bélgica, a conocer a Paul Otlet y Henry la Fontaine, considerados padres de la documentación moderna y creadores de la CDU; y, finalmente, a Madrás, en la India, para conocer a Shiyali R. Ranganathan, bibliotecario y profesor de matemáticas que revolucionó la teoría de las clasificaciones.

Finalmente, haremos un breve resumen del resto de clasificaciones documentales surgidas durante los siglos XIX y XX.

1.3.1. La clasificación decimal de Dewey, de 1885

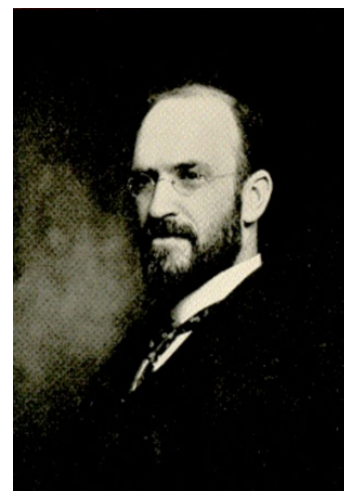
Melvil Dewey era estudiante y ayudante de bibliotecario del Amherst College de Massachusetts. En 1873, preparó como trabajo académico (tenía 22 años) la clasificación decimal, basándose en la clasificación de Harris.

La **clasificación decimal de Dewey** se publicó en 1876 con el nombre de *A classification and subject index for cataloguing and arranging the books and pamphlets of a Library* y en 1885 se editó como *Decimal classification and relatix index* (título que incluye las dos grandes aportaciones de Dewey). Dewey buscaba una clasificación bibliotecaria práctica que permitiera solucionar dos problemas:

a) De **catalogación**: teniendo en cuenta el aumento de las colecciones, había que reducir tiempo y gastos a la hora de catalogar.

El método de Francis Bacon

Bacon propone un nuevo método para el estudio de las ciencias, basado en: observar – inducir – hacer hipótesis – experimentar – demostrar o no la hipótesis – y concluir. Para Bacon, todo el saber se reducía a tres facultades humanas: la memoria (que genera la historia), la imaginación (que crea la poesía y las artes) y la razón (que produce la filosofía y las ciencias).



Melvil Dewey

b) De **acceso**: para facilitar el acceso directo a los usuarios sin tener que consultar el catálogo.

Dewey estudió las clasificaciones anteriores y adaptó el sistema de división decimal al sistema jerárquico de clasificación y añadió un índice.

Este sistema tiene las ventajas siguientes:

a) Las clases se enumeran de forma decimal, de manera que se pueden añadir clases nuevas a las previas con tendencia al infinito.

b) Los números arábigos tienen la ventaja de ser casi universales (a diferencia de las palabras o letras).

c) Las 10 clases representan una concepción moderna (F. Bacon) del conocimiento.

d) Los documentos se pueden agrupar por materia.

e) Se acompaña de un índice alfabético de materias e instrucciones para el catalogador.

Las tablas de la DDC³ son 10:

```
000 Computer science, information & general works
100 Philosophy & psychology
200 Religion
300 Social sciences
400 Language
500 Science
600 Technology
700 Arts & recreation
800 Literature
900 History & geography
```

Aplicación del sistema decimal

10 clases principales que se pueden subdividir en 10 clases más, y así sucesivamente.

⁽³⁾DDC es al sigla de *Dewey decimal classification*.

Actualmente, la Library of Congress de Washington tiene creada una división especial para su mantenimiento, la *decimal classification division*. Se aplica en bibliotecas de 135 países. La última edición es la vigésimo segunda. Y se considera una clasificación puntera, que siempre tenemos que ver como modelo, porque está actualizada.

Más información

OCLC Online Computer Library Center. *Dewey services, Dewey decimal classification for use with OCLC's online cataloging services* [en línea]. [Consulta: 10 de octubre de 2008].

025.431: *The Dewey blog* [en línea]. [Consulta: 10 de octubre de 2008].

Public libraries using Spanish. *Spanish version of the Dewey Decimal Classification System based on the 18th edition of DDC* [en línea]. [Consulta: 10 de octubre de 2008].

1.3.2. La clasificación decimal universal, de 1905

Dos abogados belgas, Paul Otlet y Henry la Fontaine, recibieron una copia de la clasificación decimal de Dewey en 1894. La encontraron tan magnífica y apropiada a sus intereses que pidieron permiso a M. Dewey para traducirla al francés y adaptarla a las bibliotecas europeas. Así nació la clasificación decimal universal (CDU⁴).

¿Cuáles eran los intereses de estos dos abogados? Desde 1895, fecha en que crearon el Instituto Internacional de Bibliografía, trabajaban para realizar el repertorio bibliográfico universal (RBU⁵), una bibliografía/catálogo que recogiera todo el saber científico del momento. Una tarea de gran magnitud que tecnológicamente tenía como soporte fichas de cartulina, doce millones de fichas de cartulina...

Recursos disponibles

Nos podemos hacer una idea de la magnitud del proyecto a partir de los siguientes recursos:

- El Mundaneum (1910), sede del Instituto de Documentación creado por Otlet y La Fontaine, destruido en la Segunda Guerra Mundial por las tropas nazis y trasladado a Mons en 1998 como museo y centro de documentación de cariz universal.
- Documental *Biografía de Paul Otlet* narrado por su biógrafo W. Boyd Rayward [en línea].
- Extractos del documental de sugerente título *The man who want to classify the world* de Françoise Levie, colgados en YouTube [en línea].
- La RBU no era el único proyecto de estas características. Otro proyecto contemporáneo internacionalista y utópico fue la creación de la lengua esperanto, por el médico polaco L. L. Zamenhof en 1887.

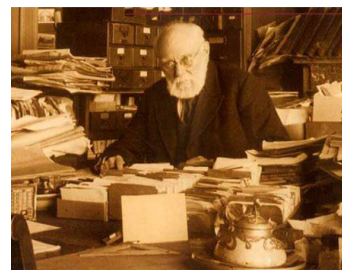
Con el fin de clasificar la materia de tantos documentos, necesitaban un sistema que fuera flexible, que se pudiera ampliar. Una vez autorizados a adaptar la Dewey (que entonces iba por la quinta edición) la tradujeron al francés e incorporaron novedades en ella.

En palabras de Otlet, en su tratado de documentación y en referencia a la clasificación Dewey:

“Han pasado 36 años desde que una afortunada casualidad puso a nuestra disposición una copia de su magnífico sistema de clasificación, cuya idea, tan simple como práctica, abrió para nosotros, más como bibliógrafos que como bibliotecarios, un camino hacia el hito que deseábamos, la construcción de un catálogo universal de todo el conjunto de ideas que los hombres durante siglos intentaron perpetuar por escrito [...] Nuestras necesidades bibliográficas precisaban un sistema clasificatorio que se pudiera ampliar sin límites.”

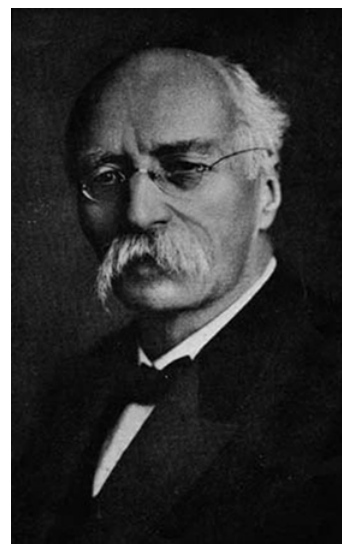
En 1905, publicaban la primera edición de las tablas de la clasificación decimal universal con el nombre *Manuel abrégé du repertoire bibliographique universel*; dos años más tarde, editaron la versión completa, que tenía 33.000 subdivisiones (actualmente, el *Master Reference File* tiene 62.000). Aunque la finalidad

⁽⁴⁾CDU es la sigla de *clasificación decimal universal*.



Paul Otlet

⁽⁵⁾RBU es la sigla de *repertorio bibliográfico universal*.



Henry la Fontaine



El proyecto repertorio bibliográfico universal

inicial era la RBU y la CDU sólo era la herramienta que permitía clasificarlo, la interrupción que supuso el conflicto armado de la Primera Guerra Mundial hizo que la clasificación fuera ganando terreno al proyecto monumental, universal y utópico, y que, en la segunda edición de la CDU de 1927-1933, el objetivo fuera la propia clasificación y no la RBU.

Clases principales de la clasificación decimal universal:

```
0 Generalidades. Información y Documentación. Informática
1 Filosofía. Psicología
2 Religión. Teología
3 Ciencias sociales
4 sin ocupar de momento
5 Ciencias puras. Ciencias exactas y naturales
6 Ciencias aplicadas. Medicina. Técnica
7 Bellas Artes. Artes aplicadas. Diversiones. Juegos. Deportes
8 Lenguaje. Lingüística. Literatura
9 Arqueología. Geografía. Biografía. Historia
```

La CDU mantiene respecto de la Dewey:

- La estructura decimal.
- El alcance temático universal.
- Las Tablas auxiliares (en la Dewey los auxiliares se conocen como tablas y las tablas principales son Schedules).
- El índice alfabético final.

Asimismo, incorpora algunas novedades con respecto a la Dewey:

- Deja una tabla principal vacía, la 4, mientras que la Dewey las utiliza todas.
- Permite combinar 2 materias o más materias con el Colon (;) y doble Colon (::) y corchetes ([]) para subagrupamientos.
- Se hacen subdivisiones cada tres dígitos.
- Las subdivisiones sueñan más específicas que en la Dewey.
- Mientras la Dewey nace para clasificar libros, la CDU nace con la voluntad de clasificar información, más concreta, más específica.

En palabras de Eric Grolier, las innovaciones de Otlet son con respecto a la teoría de las taxonomías superiores a las aportaciones de Dewey.

1.3.3. La clasificación Colon, de 1933

El uso generalizado de esta clasificación sólo se ha dado en la India, pero su influencia es internacional, gracias al resurgimiento de los debates sobre la teoría de las clasificaciones que provocó el concepto de las facetas.

Reflexión

La CDU tiene una gran implantación en Europa y se estudiará de manera exhaustiva en este mismo módulo.

La primera edición es de 1933, pero no será hasta la edición de 1944 en que aparecerán las facetas. Actualmente, la clasificación Colon (CC⁶) va por la séptima edición (1971).

⁶CC es la sigla de *clasificación Colon*.

Shiyali R. Ranganathan, bibliotecario y matemático indio, propuso una nueva estructura para los sistemas de clasificación (recordemos que el modelo básico era el jerárquico o enumerativo). Esta nueva estructura se basaba en el análisis facetado, que es aquel que estudia las características que permiten definir una materia y que, sin embargo, se puede subdividir en subclases. Una auténtica evolución para el concepto de clasificación jerárquica en que cada concepto tiene un lugar inamovible dentro del esquema general. En cambio, la CC proponía una clasificación analítico-sintética, en la que cada documento recibe una notación creada para él a partir de combinaciones y relaciones. Uno de los signos que se utilizan para unir las diferentes facetas que representarán un concepto son los dos puntos: que en inglés se llaman, precisamente, *Colon*.



Shiyali R. Ranganathan

La CC es una clasificación basada en las facetas, es decir, representa los diversos aspectos de una materia identificando las características/categorías básicas que se pueden subdividir en clases.

Ranganathan identifica unas 30 clases principales o dominios (del tipo matemáticas, botánica, educación...), que a su vez pueden ser caracterizadas por cinco grandes categorías o facetas: personalidad, materia, energía, espacio, tiempo o fórmula PMEST. Algunas clases concretas tienen, también, facetas propias.

La fórmula PMEST y las categorías gramaticales

Jacques Maniez (Maniez, pág. 46) compara la fórmula PMEST con las categorías gramaticales, como se puede observar en el siguiente ejemplo:

La fórmula PMEST y las categorías gramaticales

Documento: robo de coches a mano armada el sábado en la ciudad de Vicennes		
Facetas de Ranganathan:	Paralelismo con la proposición gramatical:	Ejemplo de J. Maniez:
Personalidad	Objeto	coche
Materia	Manera	a mano armada
Energía	Verbo	robo
Espacio	Complemento de lugar	Vicennes
Tiempo	Complemento de tiempo	sábado

Ejemplo de clasificación con la Colon

Para poner un ejemplo real de clasificación con la Colon, tomemos el caso presentado en el libro de Rosa San Segundo (1997, pág. 81-82):

Una monografía sobre *La clasificación de libros en las bibliotecas infantiles españolas*

Clase (BC)	Biblioteconomía	2
Personalidad [P]	Biblioteca infantil	,61
Materia [M]	Libros	;43
Energía [E]	Clasificación (operación)	:51
Espacio [S]	España	.541
Tiempo [T]	Siglo XX	eN

Fuente: Rosa San Segundo Manuel, "Principales sistemas de clasificación", 1997.

La notación resultante es: 261;43:51.541'N. La notación es alfabética, y utiliza letras de los alfabetos latinos, griegos y cifras arábigas.

La notación facetada necesita una sintaxis para traducirla, un orden en las partes que la forman que va desde la categoría con más carga significativa hasta la más circunstancial. A diferencia de las clasificaciones monojerárquicas, en las que hay una lista entera de clases, en las facetadas, cada notación se crea específicamente para cada documento.

Categorías de Ranganathan y Aristóteles

Las categorías de Ranganathan y Aristóteles son parecidas: la personalidad y la materia de Ranganathan es la sustancia de Aristóteles; el espacio es el lugar aristotélico; la energía es la acción, y el tiempo se corresponde con el tiempo.

1.3.4. Otras clasificaciones documentales de los siglos XIX y XX

Hay otros sistemas de clasificación importantes a los cuales, desgraciadamente, no podemos dedicar la misma extensión. Son los siguientes:

- **Clasificación Brunet:** Francia, 1804. Tuvo mucho éxito a lo largo del siglo XIX. Era heredera del sistema de Prosper Marchand y clasificaba las materias en: teología, jurisprudencia, historia, filosofía y literatura (secuencia de Francis Bacon).
- **Clasificación expansiva de Cutter:** Boston, 1891. Charles Cutter consideraba que la Dewey no era aplicable a las bibliotecas pequeñas; por eso, ideó un sistema original que se adaptaba al volumen de los fondos de la biblioteca: la clasificación expansiva. Consiste en pasar de una tabla a otra de manera progresiva. Influyó en la clasificación de la Library of Congress.
- **Clasificación de la Library of Congress:** Washington, 1899-1920. Basada en las ideas de Cutter. No es una clasificación conjunta; está formada por clasificaciones especializadas en cada materia e independientes unas de otras, hasta el punto de que no se interrelacionan entre sí. Tiene veintiún grupos temáticos. Se usa actualmente en la Biblioteca del Congreso y en las bibliotecas universitarias norteamericanas.

- **Clasificación Bliss:** Nueva York, 1912. Clasificación facetada. Cada materia se puede definir según las facetas de punto de vista filosófico, teórico, histórico y práctico.
- **Clasificación decimal japonesa:** Mori-Kiyosi, 1929, para la Biblioteca Imperial de Tokio. Influida por la CDU y Cutter. Tiene diez clases principales y pocos auxiliares especiales (sólo los de lugar y forma).
- **Clasificación rusa (BBK, actualmente LBC):** Moscú, 1924. Está fuertemente impregnada de ideología marxista leninista (la primera clase es marxismo y leninismo, la segunda ciencias naturales...). Es el sistema de clasificación nacional de la antigua URSS, que significa que tiene una magnitud sólo comparable a la de la Library of Congress (hablamos de volumen de libros). La CDU también se usa en los países de la ex URSS, especialmente en las bibliotecas públicas, gracias a la tarea que el VINITI lleva a cabo. Recordemos que VINITI, como la española AENOR, forma parte actualmente del Consorcio de la CDU, entre otras.
- **Clasificación documental china (CDC):** Beijing, 1975. Es el sistema nacional de clasificación en China. Después de la revolución, la China popular propulsó la creación de un nuevo sistema de clasificación adecuado al cambio ideológico. Es una clasificación mixta o híbrida, es decir, jerárquica, pero con posibilidad de expresar facetas. Tiene cinco clases principales (la primera es el marxismo leninismo y el pensamiento de Mao Zedong).

Mao Zedong, bibliotecario

Como curiosidad, apuntamos que Mao Zedong o Mao Tse-Tung fue ayudante de bibliotecario en su juventud, antes de iniciar su carrera política.

1.3.5. Los sistemas de clasificación en España y Cataluña

En el siglo XIX, la clasificación más utilizada era la Brunet, procedente de Francia y editada por primera vez en 1804. Estuvo vigente hasta la década de 1930.

Sin embargo, nuestros bibliotecarios estaban al corriente de las novedades que venían de Bruselas:

- En 1908, tres años después de la primera edición de la CDU de 1905, el capitán Leopoldo Giménez la aplicó en la Biblioteca de Ingenieros Militares.
- El 1911, Ricardo Codorniu publicó una monografía de temática forestal, de cuya causa era un defensor adaptando las clases de la CDU.
- En 1920, en Cataluña, Jordi Rubió Balaguer publicó la adaptación de la clasificación de Bruselas, que –recordémoslo– es la CDU basada en la Dewey, al catalán. En la década de los 30, esta edición se agotó y hubo que pensar en reeditarla, pero el estallido de la Guerra Civil y la posguerra aplazó las ediciones.

Ricardo Codorniu Stárico

Ricardo Codorniu Stárico (Cartagena, 1846-Murcia, 1923). Fue el ingeniero forestal que llevó a cabo la repoblación de la sierra de Espuña. Se considera un hombre visionario que, además de ver las posibilidades que ofrecía la CDU, también introdujo el esperanto en España y fundó la Sociedad Murciana del Esperanto en 1902. Como podemos observar, son dos de los proyectos universalistas del momento.

Jordi Rubió Balaguer

Jordi Rubió Balaguer (Barcelona, 1887–1982). Trabajó de profesor, bibliotecario e investigador. Fue director de la Biblioteca de Cataluña y de la Escuela de Biblioteconomía. Entre 1936 y 1939, trasladó la Biblioteca de Cataluña al antiguo Hospital de la Santa Cruz y contribuyó decisivamente a la salvación del tesoro bibliográfico de Cataluña. Durante estos años, dirigió también el servicio de bibliotecas del frente. En 1939, cesó en sus cargos oficiales y entró a trabajar en una editorial. Participó, sin embargo, desde el primer momento en las actividades culturales catalanas que se retomaban después de la guerra dentro de la más estricta privacidad.

Durante la Segunda República (1931-1939), las políticas bibliotecarias impulsan la implementación de la CDU en las bibliotecas nacionales, universitarias, municipales..., y una vez acabada la Guerra Civil, el ministro de Franco, Javier Lasso de la Vega, regula el uso de la clasificación decimal de Melvil Dewey, modificada por el Instituto Internacional de Bibliografía de Bruselas con las correcciones introducidas por Berlín.

Esta clasificación oficial fue abolida en 1989, pero no fue sustituida por ninguna otra recomendación; por lo tanto, se produjo un vacío legal. Cada comunidad autónoma tenía transferida la competencia de elección del sistema de clasificación.

En Cataluña, el hijo de Jordi Rubió Balaguer, Jordi Rubió Lois, continuó la tarea de su padre, primero como editor en la editorial Teide de las nuevas ediciones al catalán de la clasificación decimal y, finalmente, en 1982, como responsable de la cuarta y última edición hasta la actualidad.

Prestad atención al autor y coautor de la obra: *Clasificación decimal: adaptación para las bibliotecas catalanas*; dirigida por Jordi Rubió Balaguer, 4.^a ed. preparada por Jordi Rubió Lois. Barcelona: Teide.

Clasificación decimal frente a la CDU

Las ediciones catalanas de la clasificación decimal se titulan *Clasificación decimal* y no *CDU*, porque la adaptación catalana propone en ciertos casos dobles soluciones de clasificación: la de la CDU y la de Dewey.

En palabras de Teresa Rovira (1989, pág. 11):

“Por otro lado, el tronco americano y la adaptación belga siguen una evolución diferente: la primera orientada [...] a la ordenación de libros en las estanterías y de fichas en los catálogos; la segunda hacia la clasificación de bibliografías. Esto hace que en esta tercera edición catalana se diera en diversos casos «cuando hay razones que lo hacen aconsejable» las soluciones dobles de la CDU y la DC. Además, no se suprime la clase 4 de lenguaje como había hecho la CDU [...] tampoco dudó en otros casos, como en las tablas de psicología, en dar soluciones alternativas desde la Dewey, mucho más puestas al día. Por todo eso, la adaptación de 1976 no lleva ya el título de clasificación decimal de Bruselas, sino simplemente el de clasificación decimal, que será, también, la de la cuarta y última edición, publicada en 1982, de la cual ya es plenamente responsable Jordi Rubió Lois.”

En toda España, podemos encontrar redes que catalogan con diferentes sistemas de clasificación, aunque el más usado es la CDU, y eso favorece la catalogación cooperativa y el préstamo interbibliotecario.

La elección alemana

Cabe anotar el hecho de la elección –a raíz de las relaciones diplomáticas que España mantenía con Alemania– de la edición internacional alemana que iba por la tercera edición, frente a otras ediciones internacionales como la inglesa, que iba por la cuarta, o la francesa, que ya iba por la quinta.

En Cataluña, se utilizan la edición de Jordi Rubió Lois, la CDU, la Library of Congress (aplicada en la Biblioteca de la Universidad Pompeu Fabra de Barcelona).

1.3.6. Conclusiones

Las clasificaciones documentales han evolucionado, influidas por los sistemas filosóficos y científicos contemporáneos. Entre ellos, destaca Francis Bacon, que supuso el final de las concepciones medievalistas en la clasificación de las ciencias.

Las necesidades que han impulsado la redacción de nuevos cuadros de clasificación han sido las de querer compilar el conocimiento documental en general, o las de crear un instrumento práctico que sirviera para clasificar y ordenar el fondo de una biblioteca concreta, sin olvidar la importancia de cuestiones de cariz ideológico.

En la historia de las clasificaciones documentales, también podemos observar que los cuadros de clasificación se basan unos en otros.

Grandes clasificaciones del siglo XIX: Brunet, Dewey, Cutter.

Grandes clasificaciones del siglo XX: CDU, Colon, Library of Congress, Bliss.

En España y en Cataluña, en el siglo XX el uso que está más generalizado es el de la CDU.

En este punto del módulo, el estudiante tendría que situarse en la secuencia:

Clasificaciones documentales > para bibliotecas > CDU.

Bibliografía

025.431: *The Dewey blog* [en línea]. [Consulta: 10 de octubre de 2008].

AENOR. *Asociación española de normalización y certificación* [en línea].

Pilar, Arnau Riverd (2005). *Documentación: hitos históricos, precedentes - Dewey - Otlet - FID*. (pág. 96-105 y 248-255). Madrid: Mundarnau.

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

B. Gil Urdiciain (1997). "Los lenguajes de clasificación documental". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 91-108). Madrid: Síntesis ("Biblioteconomía y Documentación").

M. D. Gili Martínez; J. Vives Gràcia (1996, julio-diciembre). "La nova edició espanyola de la CDU: arriscar-se al canvi?". Ítem: *Revista de Biblioteconomia i Documentació* (núm. 19, pág. 94-121).

Jack Glazier (2003). "Las raíces culturales de la clasificación modern". En: *IV Coloquio Internacional de Ciencias de la Documentación*. Salamanca: Congreso del Capítulo Español de ISKO. VI.

Eric Grolier (1976, noviembre-diciembre). "La clasificación cien años después de Dewey". *Boletín Unesco* (vol. XXX, núm. 6, pág. 342-350).

Library of Congress Classification outline [en línea]. [Consulta: 1 de octubre de 2008].

José López Yepes (1995). *La documentación como disciplina: teoría e historia* (2.ª ed., pág. 90-106). Pamplona: Eunsa.

J. Maniez (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.

OCLC Online Computer Library Center. *Dewey services, Dewey decimal classification for use with OCLC's online cataloging services* [en línea]. [Consulta: 10 de octubre de 2008].

P. Otlet (1996). *El tratado de documentación: el libro sobre el libro, teoría y práctica* (trad. De M. Dolores Ayuso). Murcia: Universidad de Murcia.

M. Pinto Molina (ed.) (1997). *Manual de clasificación documental*. Madrid: Síntesis ("Biblioteconomía y Documentación").

Public libraries using Spanish. *Spanish version of the Dewey Decimal Classification System based on the 18th edition of DDC* [en línea]. [Consulta: 10 de octubre de 2008].

Luca Rosati (2005). "Clasificación facetada para los servicios comunitarios mediante el uso de las categorías estándares del Classification Research Group". *Congreso del capítulo español de ISKO: La dimensión del conocimiento* (núm. VII). Barcelona.

T. Rovira (1989, enero-junio). "Jordi Rubió i Loís: la seva influencia en el món bibliotecari de Catalunya". *Revista Ítem* (núm. 4).

Sagredo, Félix (2001). "Tras las huellas euro-norteamericanas del movimiento documental contemporáneo I" [en línea]. Madrid: *Revista Documentación de las Ciencias de la Información* (núm. 24, pág. 35-43). [Fecha de consulta: 1 de octubre de 2008].

R. San Segundo Manuel (1997). "Principales sistemas de clasificación". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 67-89). Madrid: Síntesis ("Biblioteconomía y Documentación").

R. San Segundo Manuel (2001). *Teoría e historia de la clasificación bibliotecaria en España, siglos XIX y XX*. Madrid: Universidad Complutense.

R. San Segundo Manuel (1999). "Futura implantación de los distintos sistemas de clasificación en las bibliotecas del estado español" [en línea]. [Consulta: 1 de octubre de 2008]. *7es. Jornades Catalanes de Documentació*. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya.

R. San Segundo. *Use of UDC classification system in Spain: implementation, application, teaching and research* [en línea]. Universidad Carlos III. Madrid. (Presentación) [Consulta: 1 de octubre de 2008].

Universal Decimal Classification Consortium Homepage. La Haya: UDC Consortium. Act. 2002-08-01 [en línea]. [Consulta: 10 de octubre de 2008].

2. La clasificación decimal universal

El grueso de este módulo está dedicado a la clasificación decimal universal, ya que es la clasificación más usada en nuestro ámbito cultural más próximo.

En las páginas siguientes, se tratan los temas siguientes:

1) Características que definen la CDU, basadas en las tipologías estudiadas en el módulo 3, capítulos dedicados a las tipologías de lenguajes documentales: sintético frente a analítico / precoordinados frente a postcoordinados / controlados frente a libres / jerárquicos frente a combinatorios frente a sintácticos.

2) Las ediciones de la CDU.

3) La gestión de la CDU.

4) La estructura de la CDU:

a) Clases principales. Clases y subclases. Notación.

b) Tablas principales.

c) Tablas auxiliares:

- Auxiliares comunes independientes.
- Auxiliares comunes dependientes.
- Auxiliares especiales.

d) Signos: combinatorios e instructivos.

5) Finalmente, el bloque dedicado a ejercicios prácticos de clasificación.

a) Empezamos a clasificar.

b) Prácticas con la CDU.

Entre los ejercicios solucionados que ilustran la teoría y los ejercicios propuestos como actividades de autoevaluación, el estudiante resolverá aproximadamente 200 casos de clasificación de documentos. Se ha procurado que abarquen todas las áreas de conocimiento y se han organizado por niveles de dificultad –baja, media y alta. Aunque la mayoría de ejercicios son nuevos, no queremos dejar de hacer referencia a los ejemplos proporcionados por la propia CDU y la Guía de Aenor de I. McIlwaine.

2.1. Características de la clasificación decimal universal

La clasificación decimal universal presenta las características siguientes:

a) La clasificación decimal universal es un lenguaje documental sintético, porque el objetivo es sintetizar el contenido de los documentos en un único término de indexación, en este caso una notación. Por ejemplo: *378 documento sobre enseñanza universitaria*. Aunque el uso de elementos facetados (auxiliares y signos clasificatorios) la convierte en analítico-sintética, es decir, primero se analizan los elementos que forman la materia del documento para después sintetizarlos en una notación.

b) Es un lenguaje **codificado** (no natural) porque la indexación de los documentos se basa en el uso de un sistema de notación o código artificial formado por símbolos (numéricos y alfanuméricos) que traducen las materias de una clasificación bibliográfica a índices ordenados. Estos códigos, además, sirven para facilitar la ordenación física de los documentos. Y son internacionales; no les hace falta traducción. Por ejemplo: 378.

c) Es un lenguaje **controlado** (no libre), porque tiene como objetivo la representación unívoca del contenido de los documentos y de las consultas: un único concepto por término y un único término por concepto (eliminando la sinonimia, polisemia y homonimia). En la indexación, sólo se pueden utilizar los códigos que figuran en las páginas de los sistemas de clasificación.

d) Es un lenguaje **precoordinado** (no postcoordinado), porque la coordinación de los diferentes conceptos que forman una materia se produce en el momento de la indexación, es decir, antes de almacenarlos o incorporarlos a un catálogo.

e) Es un lenguaje de estructura **híbrida** o **mixta**, es decir, jerárquica con elementos de tipo facetados (no sintáctica) porque sigue un orden lógico que aproxima los conceptos más específicos y los agrupa dentro de los más generales. Esta estructura hace que los conceptos estén ordenados siguiendo una jerarquía: cada término está representado por un símbolo numérico, alfabético o alfanumérico que indica su situación dentro de la cadena jerárquica. La CDU presenta una estructura jerárquica, pero tiene la posibilidad de expresar facetas, lo cual la convierte en híbrida o mixta.

f) Es el lenguaje que permite **clasificar/recuperar** la materia del documento y al mismo tiempo **ordenarla** físicamente, agrupando las obras por contenidos afines, y también generar **subproductos** del catálogo, como boletines de últimas novedades o sumarios organizados temáticamente.

g) Es **enciclopédico** porque incorpora todos los campos del conocimiento humano.

h) Es **decimal**, ya que las tablas están ordenadas por el principio de los números decimales. Sus cifras tienen el valor de las fracciones decimales situadas detrás de un 0 (cero coma) que siempre se supone. Esta estructura posibilita que un número determinado se pueda subdividir indefinidamente por sucesivas agrupaciones, sin que la de partida llegue al inmediatamente superior.

A modo de conclusión

La CDU es un lenguaje documental analítico-sintético, codificado, controlado, precoordinado, de estructura híbrida, enciclopédico y decimal.

Más información

F. J. Álvarez García (1997). "Introducción al uso práctico de la CDU". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 219-243). Madrid: Síntesis ("Biblioteconomía y Documentación").

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000; (incluye las modificaciones de la Norma UNE 50001:2004/1M. Trad. del Master Reference File realizada por el Centro de Información y Documentación Científica (CINDOC) (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR,.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

B. Gil Urdiciain (1997). "Los lenguajes de clasificación documental". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 91-108). Madrid: Síntesis ("Biblioteconomía y Documentación").

M. Granados; A. Nicolau (2005, 6-8 de julio). "La recuperació de la informació en els catàlegs en línia: l'ús de la Classificació decimal universal i la seva implicació en la indexació" [en línea]. 7è. Congrés del Capítol Espanyol d'ISKO. Barcelona: Departament de Biblioteconomia i Documentació de la Universitat de Barcelona. [Consulta: 1 de octubre de 2008].

M. J. López-Huertas (1997). "Estructura de la clasificación decimal universal". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 177-217). Madrid: Síntesis ("Biblioteconomía y Documentación").

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

L. M. Moreno Fernández; M. D. Borgoños Martínez (2002). *Teoría y práctica de la Clasificación Decimal Universal* (2.ª ed. revisada, corregida y aumentada). Gijón: Trea.

2.2. Ediciones de la CDU

La edición es un dato importante, ya que la CDU tiene diversos formatos. El tronco común es el *master reference file* (MRF) o fichero básico de referencia, que contiene cerca de 62.000 entradas, pero del cual se hacen diversas presentaciones:

a) **Entera / abreviada / de bolsillo**: físicamente, la edición entera tiene 3 volúmenes, la abreviada tiene 1, de 800 páginas, y la de bolsillo también tiene 1, de 300 páginas aproximadamente. Su funcionamiento es idéntico, pero varía la cantidad de entradas (más específica en la entera, más general en la abreviada, sintética en la de bolsillo). Cada una de las tres presentaciones tiene una finalidad: la entera permite clasificar documentos con un alto grado de especificidad, es útil en bibliotecas especializadas. La abreviada es de implantación general por su buen equilibrio entre entradas y especificidad. La presentación

de bolsillo está pensada para los no iniciados (ya que es mucho más sintética) y para la clasificación de pequeñas bibliotecas como las escolares, infantiles, hospitalarias, de pequeños municipios, etc.

b) En papel / CD-ROM

c) **Años 1995/2000/2004.** Las ediciones anteriores a 1992 dependían de la Federación Internacional de Documentación (FID), pero a partir de 1992 dependen del Consorcio de la CDU. Aunque el funcionamiento es idéntico, sí que pueden cambiar los dígitos de una clase, desaparecer conceptos, crearse nuevos, subdividir un concepto para especificarlo más, etc.

Ved también

En el espacio de "Recursos del aula" os damos acceso a la edición digital de la CDU del año 2000, *Programa de la CDU*.

Cambios en las últimas ediciones de la CDU

Tipos de cambios que se suceden en las diferentes ediciones

Cambios de la edición 2000 (en relación con la de 1995)	Cambios de la edición 2004 (en relación con la de 2000)
<p>Se amplían considerablemente las entradas (físicamente se pasa de dos volúmenes a tres, en la edición entera). Edición digital en CD. Se potencia la facetación de las tablas de auxiliares comunes independientes de lugar y lengua. Subclases, como informática, pasan a clase.</p>	<p>Inclusión de una nueva tabla de auxiliares comunes de propiedad. Reestructuración total de la tabla 2 de religión. Cambios a: cine, bienestar social, ciencias medioambientales, veterinaria.</p>

En España, las últimas publicaciones han sido:

- Entera de 2000.
- Abreviada de 2004.
- Bolsillo de 2004.
- Versión electrónica de 2000.
- En todos los casos, el responsable en España es AENOR.

AENOR

AENOR es miembro del Consorcio de la CDU, con sede en La Haya, como principal editor en España y responsable de la edición en castellano.

Es importante remarcar que la técnica de clasificación es la misma independientemente de la edición que se consulte.

A pesar de las diversas presentaciones y ediciones, una vez se sabe clasificar con una edición concreta, por ejemplo la abreviada de 1991, también se sabe clasificar con la última edición de 2004, porque los cambios afectan a las numeraciones de las clases, pero no a su construcción y funcionamiento, que mantiene la sintaxis propia del lenguaje.

Para este curso de Fundamentos de lenguajes documentales, escogemos la CDU en **formato bolsillo** o en **formato CD**.

Más información

AENOR. *Asociación española de normalización y certificación* [en línea]. [Consulta: 10 de octubre de 2008].

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

Gerhard J. A. Riesthuis (1998, 16-21 de agosto). "The UDC Master Reference File" [en línea]. *64th IFLA General Conference*. Amsterdam. [Consulta: 1 de octubre de 2008].

C. Soria. Cambios introducidos en la última edición (nov. 2000) de la CLASIFICACIÓN DECIMAL UNIVERSAL [en línea]. [Consulta: 1 de octubre de 2008].

A modo de conclusión

El tronco común a todas las presentaciones de la CDU es el *master reference file* (MFR), el cual contiene cerca de 62.000 entradas.

La CDU española tiene diversas presentaciones: entera, abreviada y de bolsillo, en formato papel o digital y de diversas ediciones.

La técnica de clasificación es la misma independientemente de la edición que se consulte.

2.3. Gestión de la CDU

2.3.1. ¿Quién es el responsable de la gestión de la CDU?

El responsable de la gestión de la CDU es el Consorcio de la CDU con sede en la Biblioteca Real de La Haya (accesible en línea).

Este organismo se creó en 1992, y absorbió las responsabilidades que hasta entonces tenía la FID⁷ en materia de CDU.

⁽⁷⁾FID es la sigla de Federación Internacional de Documentación.

Breve historia de la FID

La I Conferencia Internacional de Bibliografía organizada por Otlet y La Fontaine (1895) creó en 1895 el Instituto Internacional de Bibliografía (IIB) encargado de la RBU. Durante la Primera Guerra Mundial, las actividades del IIB se pararon y se reanudaron a partir de 1920. El año siguiente, en 1921, pasa a llamarse *Instituto Internacional de Documentación* (IID). En 1938, con toda una serie de cambios en los objetivos, medios y estructura, se convirtió en la Federación Internacional de Documentación (FID).

La CDU nace y crece en el seno de estas organizaciones, hasta que en 1992 se crea el Consorcio de la CDU como un organismo con entidad propia para velar por su mantenimiento.

En el Consorcio de la CDU, o UDC Consortium, participan diversas agencias de normalización y editores como la española AENOR, la rusa VINITI o la inglesa BSI. El Consorcio es el responsable de los cambios, y cada agencia los publica en su idioma.

Representación española

Dentro del Consorcio, con respecto a la representación española, hay una representante de AENOR y la profesora Rosa San Segundo, responsable de la adaptación de las ediciones españolas de la CDU de AENOR y autora de diversa bibliografía recomendada en este módulo.

2.3.2. ¿Se actualizan las clases de la CDU?

Las clases de la CDU se actualizan anualmente. La voluntad de los gestores de la CDU es ir adaptando las tablas a la realidad actual. A pesar de todo, los cambios no se suceden con la rapidez que sería deseable.

El Consorcio publica cada mes de noviembre los cambios en el documento *Extensions and corrections to the UDC* y, electrónicamente, edita el uno de enero siguiente el *master reference file* o fichero básico de referencia. De forma informática, su soporte es el software CD-ISIS de la UNESCO.

Actualización de la CDU

Pondremos dos ejemplos sobre la –lenta– actualización de la CDU: los temas de religión y de informática.

1) **Religión.** La tabla *Religión* ha sido, desde las primeras ediciones de la CDU, muy desarrollada (en comparación con otros campos científicos o culturales) pero estaba dedicada, casi en exclusiva, a la Iglesia cristiana, que tenía ocho de las nueve clases en que se dividía la tabla y que sólo dejaba una única clase para el resto de religiones. No ha sido hasta la edición de 2004 que se ha dado entrada propia a cada una de las religiones:

```
24 Budismo
26 Judaísmo
27 Cristianismo
28 Islam
```

2) **Informática.** Hasta la edición de 1995, la informática se clasificaba en la clase:

```
68 Industrias varias
 681 Mecánica de precisión
   681.3 Equipo para el proceso de datos. Calculadoras.
       Ordenadores.
```

Como podemos apreciar, era una subclase de 4 dígitos que colgaba de una genérica Industrias diversas, mientras que en la edición del año 2000 pasa a tener consideración de clase, sustancialmente más relevante, con una notación de tres dígitos, la 004, dedicada íntegramente a la informática:

```
004 Ciencia y tecnología de los ordenadores. Informática
```

2.3.3. Nuevas ediciones de la CDU y reclasificación de los fondos

¿Las bibliotecas tienen que reclasificar sus fondos cada vez que aparece una nueva edición de la CDU? No. Los documentos no se reclasifican, sino que se espera que el tiempo acabe haciendo desaparecer los viejos.

En palabras de Miguel Benito (2008):

“Todo depende del tipo de biblioteca. Una forma práctica, sobre todo para bibliotecas en que las colecciones son de actualidad, es la de no preocuparse por los cambios de las obras que se hayan adquirido hasta hoy. El tiempo es un factor importante en nuestro trabajo. De aquí a unos años las obras adquiridas hasta hoy a duras penas existirán en la colección. Hay bastante con hacer los cambios para las adquisiciones hechas a partir de hoy.”

La solución es poner una referencia al catálogo, que vincule la notación nueva con la antigua y el topográfico. Esta opción significa que se modifica el registro bibliográfico del catálogo, pero no la signatura topográfica del documento, y permite recuperarlo, tanto si se hace la busca por la notación antigua como por la nueva.

2.3.4. Conclusiones

La CDU evoluciona, lentamente.

El responsable de las modificaciones es el Consorcio de la CDU con sede en La Haya, que edita anualmente el documento *Extensions and corrections to the UDC* y, electrónicamente, edita el uno de enero siguiente el *master reference file* o fichero básico de referencia.

No se suele reclasificar. Una solución es poner un punto de enlace en el catálogo remitiendo la notación antigua y nueva con la signatura topográfica.

Bibliografía

AENOR. *Asociación española de normalización y certificación* [en línea]. [Consulta: 10 de octubre de 2008].

Miguel Benito. *El sistema de Clasificación Decimal Universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008]. O, también, en la URL Instituto de Taranco. [Fecha de consulta: 10 de octubre de 2008].

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

T. Guardans(1989, enero-junio). "Les altres religions a la classificació decimal universal". *Revista Ítem* (núm. 4).

I. C. McIlwaine (2003). *Clasificación decimal universal. Guía para uso de la CDU*. Madrid: AENOR.

Gerhard J. A. Riesthuis (1998, 16-21 de agosto). "The UDC Master Reference File" [en línea]. *64th IFLA General Conference. Amsterdam*. [Consulta: 1 de octubre de 2008].

3. Estructura de la clasificación decimal universal

Si tuviéramos la CDU en las manos, veríamos que se estructura en tres partes diferenciadas:

Introducción	Trata de tres aspectos: <ul style="list-style-type: none"> • Naturaleza de la clasificación • Estructura de la CDU • Gestión de la CDU 	
Parte sistemática	Tablas principales:0 Generalidades 1 Filosofía. Psicología 2 Religión. Teología 3 Ciencias sociales 4 <i>sin ocupar de momento</i> 5 Ciencias puras. Ciencias exactas y naturales 6 Ciencias aplicadas. Medicina. Técnica 7 Bellas Artes. Artes aplicadas. Diversiones. Juegos. Deportes 8 Lenguaje. Lingüística. Literatura 9 Arqueología. Geografía. Biografía. Historia	Tablas auxiliares comunes: <ul style="list-style-type: none"> • la Coordinación. Extensión • Ib Relación • Ic Lengua • Id Forma • Ie Lugar • If Razas • Ig Tiempo • Ih Notaciones ajenas a la CDU • Ik Características generales
Índice alfabético	Es un índice que contiene parte de los conceptos utilizados en las tablas.	

Visualmente, sería:

Introducción Ejemplo, pág. 1.	1 NATURALEZA DE LA CLASIFICACIÓN La clasificación decimal universal (CDU) es un sistema de clasificación bibliográfica en el que tiene cabida y lugar todo el conocimiento humano, tanto en el plano material como conceptual. Su notación, basada en números arábigos (...)
Parte sistemática Ejemplo de la clase 0. Generalidades	Tablas principales 0 Generalidades 00 Prolegómenos 01 Bibliografía. Catálogos 02 Biblioteconomía 030 Obras de referencia... 050 Publicaciones seriadas 06 Organizaciones 070 Periódicos 08 Poligrafías 09 Manuscritos
Índice alfabético Ejemplo de la letra A	Abanicos 688.5 Abastecimiento de agua 628.1 Abdomen 617.55 Abogados defensores 364-48 Abolición de la esclavitud 326.8 (...)

Este es el orden en que aparecen en la CDU, pero en beneficio de una mejor comprensión, empezaremos hablando de las clases y notaciones, en segundo lugar de las tablas principales, en tercer lugar de las tablas auxiliares, y en cuarto lugar de los signos clasificatorios, que permiten la unión de diversas clases.

3.1. Clases y notaciones

3.1.1. Clases

Una **clase**, en palabras de McIlwaine (2003, pág. 286), es un concepto simple dentro de una clasificación. Las características principales de las clases es que son conceptos generales y que son disjuntas.

Las clases se pueden analizar desde diversos puntos de vista:

a) Según el alcance conceptual:

- Clases principales.
- Clases y subclases.

b) Según el principio de división:

- Clases dispuestas en cadena.
- Clases dispuestas en serie.

c) Según la ordenación:

- De general a particular (de psicología a pedagogía).
- De la disciplina inicial a la que depende de ella: las matemáticas antes de la física.
- De lo que es simple a lo que es complejo.
- De la teoría a la práctica.
- En orden cronológico: nacimiento, vida, muerte.
- En orden topográfico, partiendo de una visión europeísta y occidental: Europa en primer lugar, después Asia, África, América, Australia y Oceanía.

Según el alcance conceptual

Las **clases principales** son las clases situadas en el primer nivel de la estructura.

Conjuntos disjuntos

Dos conjuntos son disjuntos cuando son mutuamente excluyentes. Por ejemplo: o es escultura, o es pintura o es grabado.

730 Escultura
75 Pintura
76 Grabado

Según Maniez (pág. 37), no conviene que haya ni muchas ni demasiado pocas. La base decimal nos permite llegar a tener 10 clases (la CDU, por ejemplo, tiene 10 clases principales). Si utilizamos alfabetos, podemos ampliarlas más, ya que además de las letras, tenemos las combinaciones de mayúsculas y minúsculas y la de diferentes alfabetos como el latín o el griego.

Las clases principales tienen que contener todo el dominio. Una buena construcción prevé tener una clase de generalidades comunes a todas las ciencias y otra de diversos/otros o simplemente vacía. Así, la CDU tiene la clase 0 de generalidades y la 4 vacía. Esta estructura flexibiliza el sistema de cara a las posibles actualizaciones.

Clases y subclases

Las clases y subclases son todas aquellas clases no principales que conforman la secuencia jerárquica de la clasificación.

La característica principal de las clases y subclases es que se agrupan en una sola arborescencia, lo cual significa que sólo hay un camino para conectarse con la raíz. De igual forma que las clases principales, se construyen siguiendo el método de la división lógica para crear clases disjuntas.

Según el principio de división

Una **clasificación** refleja las relaciones entre los conceptos, no sólo los identifica, ya que entonces es una ordenación y no una clasificación.

Ejemplos de ordenación y de clasificación

- **Ordenación:** hacer una lista de títulos de películas de cine ordenados alfabéticamente.
- **Clasificación:** agrupar las películas de cine según su género: musicales, comedia, tragedia, de suspense, acción...

Las relaciones lógicas entre las materias son básicamente de dos tipos: de subordinación y de coordinación. En la CDU se conocen con el nombre de *cadena* y *serie*, respectivamente.

Clases dispuestas en cadena

Las **clases dispuestas en cadena** son clases subordinadas sucesivamente de genérico a específico.

2 Religión

25 Religiones de la Antigüedad

252 Religiones de Mesopotamia
252.1 Religión de los sumerios

Clases dispuestas en serie

Las **clases dispuestas en serie** consisten en series de clases coordinadas. Es un grupo de clases mutuamente excluyentes, derivadas de la aplicación de una característica.

34 Derecho

341 Derecho internacional	342 Derecho constitucional	343 Derecho penal	346 Derecho económico y financiero	347 Derecho civil
---------------------------	----------------------------	-------------------	------------------------------------	-------------------

Según la ordenación de las clases

Como podemos observar hojeando la CDU, las clases se encuentran ordenadas:

- De general a particular: ciencias naturales > zoología > invertebrados > protozoos.
- De la disciplina inicial a la disciplina de la cual depende: las matemáticas antes de la física.
- De la teoría a la práctica: primero las clases de la tabla 5 y después sus aplicaciones prácticas, en la clase 6.
- En orden cronológico: nacimiento > vida > muerte.
- En orden topográfico (la CDU parte de una visión europeísta y occidental): Europa en primer lugar, después Asia, África, América, Australia y zonas polares.

3.1.2. Notación

La **notación** es el término de indexación propio de los sistemas de clasificación. Es un código artificial, no un lenguaje natural, hecho a partir de los números, letras y símbolos que acompañan las rúbricas. Representa la materia del documento; también se usa para su ordenación física.

La longitud de la notación puede representar el grado de especificidad de la materia de un documento (cuanto más larga, más específica), pero también puede estar traduciendo una arborescencia con unas pocas clases principales. Jacques Maniez calcula que una notación de hasta 7 dígitos es comprensible, más ya no. Por ejemplo, 524.42 es una notación.

La notación será diferente en una clasificación puramente jerárquica que en una facetada.

a) Notación en una clasificación puramente jerárquica. J. Maniez (pág. 162), citando a Dobrowolski, nos dice que una notación tiene que ser:

- Breve.
- Lógica.
- Expresiva.
- Hospitalaria (que permita ampliaciones).
- Fácil de leer.
- Fácil de memorizar.
- Evolutiva (que no cierre la evolución de los conocimientos).
- Ha de utilizar símbolos, cuyo orden tiene que ser universalmente admitido.

b) Notación en una clasificación híbrida o facetada. Desde el momento que se usan facetas, hace falta una sintaxis. Hay que fijar el orden de sucesión en función de cómo queramos agrupar los documentos en las estanterías y prever una notación que distinga claramente las facetas de las otras clases. En la CDU hay una sintaxis pensada para el orden de las diferentes partes de una notación y cada elemento facetado se distingue por un signo concreto. Por ejemplo, el símbolo igual = introduce un auxiliar de lengua:

```
=361 significa Lengua vasca
```

```
72=361 Arquitectura escrito en vasco
```

3.1.3. Conclusiones

Las clases tienen el rol central de la estructura de una clasificación. Podemos analizar las clases en función de su alcance conceptual (clases principales, clases y subclases), en función del principio de división (clases en cadena o en serie) o en función de su ordenación.

La notación es el término de indexación de los lenguajes de clasificación. Es un código artificial que acompaña las rúbricas. Representa la materia y ordena el fondo documental. Diferenciamos entre la notación de un sistema jerárquico de la de un sistema híbrido o facetado porque, en este caso, se necesita una sintaxis.

Bibliografía

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

J. Maniez (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.

Joan Reitz. *Odlis Online dictionary of library and information science*. Western Connecticut State University [en línea]. [Consulta: 10 de octubre de 2008].

3.2. Tablas principales

3.2.1. Tablas principales de la edición de bolsillo CDU 2004

- Tabla 0 Generalidades⁸

(8)

```
Tabla 0 Generalidades
00 Prolegómenos. Fundamentos de la ciencia y la cultura
01 Bibliografía y bibliografías. Catálogos. Listas de libros
02 Biblioteconomía. Bibliotecología
030 Obras de referencia general. Enciclopedias
050 Publicaciones periódicas y publicaciones seriadas
06 Organizaciones y colectividades. Asociaciones. Congresos.
    Exposiciones. Museos
070 Diarios. Prensa. Periodismo
08 Poligrafías. Colecciones. Series
09 Manuscritos. Libros preciosos y raros
```

- Tabla 1 Filosofía. Psicología⁹

(9)

```
Tabla 1 Filosofía. Psicología
101 Naturaleza, esencia y método de la filosofía
11 Metafísica
13 Filosofía de la mente y del espíritu
14 Posiciones doctrinales. Sistemas
159.9 Psicología
16 Lógica. Epistemología. Teoría del conocimiento
17 Moral. Ética. Filosofía práctica
```

- Tabla 2 Religión. Teología¹⁰

(10)

```
Tabla 2 Religión. Teología
21/29 Sistemas religiosos. Religiones y creencias
21 Religiones primitivas y prehistóricas
22 Religiones del lejano Oriente
23 Religiones del subcontinente indio. Religiones hindúes
24 Budismo
25 Religiones de la Antigüedad. Religiones precolombinas
26 Judaísmo
27 Cristianismo
28 Islam
29 Movimientos espirituales modernos
```

- Tabla 3 Ciencias sociales. Sociología. Estadística. Política. Economía. Derecho. Administración pública. Arte y ciencia militares. Asistencia social. Seguros. Educación. Etnología¹¹

(11)

Tabla 3 Ciencias sociales. Sociología. Estadística. Política. Economía. Derecho. Administración pública. Arte y ciencia militares. Asistencia social. Seguros. Educación. Etnología
 30 Teoría, metodología y métodos en las ciencias sociales
 31 Estadística. Demografía. Sociología
 32 Política
 33 Economía. Economía política. Ciencia económica
 34 Derecho. Jurisprudencia
 35 Administración pública. Gobierno. Asuntos militares
 36 Protección de las necesidades materiales y espirituales
 Consumismo. Seguros
 37 Educación. Enseñanza. Usos o costumbres. Vida social
 38
 39 Etnología. Etnografía. Usos y costumbres. Vida social. Folclore. Tiempo libre

- Tabla 4 sin ocupar de momento¹²

(12)

Tabla 4 sin ocupar por el momento

- Tabla 5 Ciencias puras. Ciencias exactas y naturales¹³

(13)

Tabla 5 Ciencias puras. Ciencias exactas y naturales
 502 Estudio de la naturaleza y su conservación.
 Ciencias naturales
 504 Ciencias del medio ambiente. Ambientología
 51 Matemáticas
 52 Astronomía. Astrofísica. Investigación espacial. Geodesia
 53 Física
 54 Química. Cristalografía. Ciencias mineralógicas
 55 Ciencias de la Tierra. Geología. Meteorología, etc.
 56 Paleontología
 57 Ciencias biológicas
 58 Botánica
 59 Zoología

- Tabla 6 Ciencias aplicadas. Medicina. Técnica¹⁴

(14)

Tabla 6 Ciencias aplicadas. Medicina. Técnica
 60 Cuestiones generales
 61 Medicina
 62 Ingeniería
 620 Ensayo de materiales. Materiales comerciales
 621 Ingeniería mecánica en general. Tecnología nuclear.
 Ingeniería eléctrica. Maquinaria
 622 Minería
 623 Ingeniería militar
 624 Ingeniería civil y estructural
 625 Ingeniería del transporte terrestre
 625.7/.8 Ingeniería de carreteras
 626 Ingeniería hidráulica
 627 Ingeniería de puertos y costas
 628 Ingeniería de la salud pública. Agua. Sanidad.
 Ingeniería de la eliminación
 629 Ingeniería de los vehículos de transporte
 63 Ciencias agrícolas
 64 Economía doméstica
 65 Dirección y organización de la industria, comercio y
 comunicación
 66 Tecnología química
 67 Industrias, oficios y profesiones
 68 Industrias, oficios y comercio de artículos acabados
 69 Construcción

- Tabla 7 Bellas Artes. Juegos. Espectáculos. Deportes¹⁵

(15)

Tabla 7 Bellas Artes. Juegos. Espectáculos. Deportes
 71 Urbanismo. Planificación física. Ordenación del territorio.
 Planificación regional, urbana y rural.
 Arquitectura de jardines. Configuración del paisaje
 72 Arquitectura
 73 Artes plásticas
 74 Dibujo. Diseño. Artes aplicadas y artesanía
 75 Pintura
 76 Artes gráficas
 77 Fotografía y procesos similares
 78 Música
 79 Diversión. Entretenimiento. Juegos. Deportes

- Tabla 8 Lenguaje. Lingüística. Literatura¹⁶

(16)

Tabla 8 Lenguaje. Lingüística. Literatura
 80 Cuestiones generales relativas a la lingüística y literatura.
 Filología
 81 Lingüística y lenguas
 82 Literatura

- Tabla 9 Arqueología. Geografía. Biografía. Historia¹⁷

(17)

Tabla 9 Arqueología. Geografía. Biografía. Historia
 902 Arqueología
 903 Prehistoria. Restos prehistóricos. Utensilios prehistóricos
 904 Restos culturales de tiempos históricos
 908 Descripción geográfica e histórica de un territorio
 91 Geografía. Exploración de la Tierra y países concretos.
 Viajes. Geografía regional
 929 Estudios biográficos y relacionados
 93 Historia
 94 Historia general

Son diez tablas, nueve ocupadas y una vacante. Recordemos que tener una clase principal vacía es un excelente método para incluir nuevos conceptos sin alterar el esquema inicial.

El sumario de las tablas y clases pertenece a la edición CDU de bolsillo 2004; las ediciones abreviada y entera son más amplias. Reproducimos las clases hasta el segundo nivel de clasificación, de dos dígitos, y en algunos casos, como la tabla 6, hasta tres dígitos, igual que en la edición original de la CDU de bolsillo.

Notaciones con la misma raíz

Ved el ejemplo siguiente, en el que se muestra que la notación tiene la misma raíz, pero es más específica en un caso que en el otro.

Edición abreviada 2000 o programa CDU del aula	Edición de bolsillo 2004
671.1 Joyería 671.15 Trabajo de las piedras preciosas 671.152 Talla del diamante	671.1 Joyería

A continuación, se explica brevemente el contenido de cada tabla y se detallan las disciplinas enumeradas en cada clase. No se pretende hacer un análisis detallado de las tablas, sino introducir al estudiante en su estructura. Finalmente, cada tabla va acompañada de diversos ejercicios con soluciones incluidas.

Más información

Todos los ejercicios se pueden resolver con la edición de bolsillo 2004, pero en determinadas soluciones, y en beneficio de una mejor comprensión, se ha escogido la notación más extensa y específica procedente de la edición abreviada 2000 (o del programa CDU que tenéis en el aula). Por regla general, si no encontráis la notación en la edición de bolsillo, la encontraréis en el programa CDU.

F. J. Álvarez García (1997). "Introducción al uso práctico de la CDU". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 219-243). Madrid: Síntesis ("Biblioteconomía y Documentación").

M. Benito. (1999). *El sistema de clasificación decimal universal: manual de aprendizaje*. Madrid: Taranco.

M. Benito. *El sistema de Clasificación Decimal Universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008]. O, también, en la URL Instituto de Taranco. [Fecha de consulta: 10 de octubre de 2008].

Classificació decimal: adaptació per a les biblioteques catalanes (1982) (4.ª ed. preparada por Jordi Rubió Lois). Barcelona: Teide.

Reflexión

Muchos de los ejercicios son nuevos, pero no queremos dejar de hacer constar que hemos incluido ejercicios extraídos de la propia CDU o bien de la guía oficial de AENOR, de McIlwaine.

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

C. Díez Carrera (1999). *Técnicas y régimen de uso de la CDU (Clasificación Decimal Universal)*. Gijón: Trea ("Biblioteconomía y administración cultural", 26).

T. Guardans (1989, enero-junio). "Les altres religions a la classificació decimal universal". *Revista Ítem* (núm. 4).

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

L. M. Moreno Fernández; M. D. Borgoños Martínez (2002). *Teoría y práctica de la Clasificación Decimal Universal* (2.ª ed. revisada, corregida y aumentada). Gijón: Trea.

Tabla 0 Generalidades

```

0 Generalidades
  00 Prolegómenos. Fundamentos más generales de la ciencia y la cultura
  004 Informática
  01 Bibliografía y bibliografías. Catálogos. Listas de libros
  02 Biblioteconomía. Bibliotecología
  030 Obras de referencia general. Enciclopedias. Diccionarios.
      Manuales, etc.
  050 Publicaciones periódicas de información general y publicaciones seriadas
  06 Organizaciones y colectividades de cualquier tipo. Asociaciones.
      Congresos. Exposiciones. Museos
  070 Diarios. Prensa. Periodismo
  08 poligrafías. Colecciones. Series
  09 Manuscritos. Libros preciosos y raros
  
```

A primera vista, el nombre de la tabla no resulta muy claro, ya que *Generalidades* no aclara su contenido, pero en realidad agrupa temas comunes a todas las disciplinas: los principios básicos de la ciencia y las materias que son comunes a la actividad intelectual (conocimiento, informática, documentación, organizaciones...).

A grandes rasgos, podemos encontrar las siguientes áreas:

- Fundamentos de la ciencia y el conocimiento (incluye la organización del trabajo, escritura, informática, normativas...). Recordemos que 004 Informática es una clase de reciente aparición (año 2000).
- Fuentes de información: bibliografías, catálogos, obras de referencia, publicaciones periódicas, poligrafías, manuscritos...
- Biblioteconomía y Documentación: para la CDU las cuestiones referidas a las bibliotecas se clasifican aquí, mientras que las referentes a archivística se clasifican en el 930, dentro de Historia.

- Organizaciones, asociaciones, congresos, exposiciones, museos.

Actividad

1. Clasificad los siguientes documentos:

- Documento sobre Thesaurus.
- Documento sobre enciclopedias.
- Documento sobre anuarios.
- Documento sobre museos.

Hallaréis la solución en el solucionario.

Tabla 1 Filosofía. Psicología

```

1 Filosofía
  101 Naturaleza, esencia y método de la filosofía
  11 Metafísica
  122 Metafísica especial
  13 Filosofía de la mente y del espíritu. Metafísica de la vida espiritual
  14 Posiciones doctrinales. Sistemas. Puntos de vista metafísicos y ontológicos
  159.9 Psicología
  16 Lógica. Epistemología. Teoría del conocimiento. Metodología lógica
  17 Moral. Ética. Filosofía práctica

```

Esta tabla sigue la tradicional división de la filosofía en metafísica, lógica y ética. E incluye la psicología, como una parte de la filosofía. También aspectos de la ética y la moral. Es una tabla bastante corta (las clases 15, 18 y 19 están vacías) y de comprensión sencilla.

Como dice López-Huertas (1997):

“el contenido de esta clase denota la marcada orientación occidental aquí puesta de manifiesto por el poco protagonismo concedido a la filosofía no occidental”.

Psicología

Las clases de psicología 159.9 acostumbran a combinarse con -05... (tabla Ik) auxiliares de personas. Por ejemplo:

```
159.922.7-053.4 Psicología infantil de la edad preescolar
```

Actividad

2. Clasificad los siguientes documentos:

- Documento sobre el neoplatonismo.
- Documento sobre las profecías.
- Documento sobre la naturaleza del conocimiento.

Hallaréis la solución en el solucionario.

Tabla 2 Religión. Teología

Hasta la edición de 2004, esta clase era básicamente de religión cristiana. El resto de religiones (todas) se habían de clasificar en una sola clase, la 29. Este hecho nos ilustra a la perfección dos características de los sistemas de clasifi-

cación en general y de la CDU en particular: la traducción del mundo según el prisma propio de una sociedad y de un tiempo (perspectiva cristiana occidental en la CDU) y la lenta actualización.

En la edición de 2000:

```

2 Religión. Teología

21 Teología natural. Teodicea. De Dios. Teología nacional.
    Filosofía religiosa
22 La Biblia. Escrituras bíblicas.
    Organizaciones y colectividades de cualquier tipo.
    Asociaciones. Congresos. Exposiciones. Museos
23 Teología dogmática
24 Teología práctica
25 Teología pastoral
26 Iglesia cristiana
27 Historia general de la Iglesia cristiana
28 Iglesias cristianas. Comunidades y sectas
29 Religiones no cristianas. Mitología. Cultos. Religión comparada

```

Religión en la CDU

Para más información sobre el tema de la religión en la CDU, ved Guardans (1989). Este artículo analiza la dificultad de indexar un fondo religioso con la CDU del momento, por dos motivos: a) la enorme variedad del fenómeno religioso, y b) la perspectiva cristiano-occidental que regía la CDU.

La CDU, en aquellos momentos, juntaba temas muy diversos bajo una misma clase, en palabras de la autora del artículo:

“Para hacerlo un poco más gráfico: si un clasificador chino hiciera con el cristianismo lo que aquí se hace con el zen, recogería bajo un mismo número un estudio sobre el Palmar de Troya, el Evangelio de san Mateo, *Marcelino, pan y vino* de Sánchez Silva, el *Cántico* de san Juan y los documentos del Vaticano II.”

Guardans (1989, pág. 22-23).

En cambio, en la revisión del 2004, se empezaron a equiparar las grandes religiones. El cambio es notorio.

```

2 Religión. Teología

21/29 Sistemas religiosos. Religiones y creencias
21 Religiones primitivas y prehistóricas
22 Religiones de lejano Oriente
23 Religiones del subcontinente indio. Religiones hindúes
24 Budismo
25 Religiones de la Antigüedad. Religiones precolombinas
26 Judaísmo
27 Cristianismo
28 Islam

```

29 Movimientos espirituales modernos

Actividad

3. Clasificad los siguientes documentos:

- Documento sobre el judaísmo sefardí.
- Documento sobre el budismo.
- Documento sobre la cienciología.
- Documento sobre la religión de los sumerios.

Hallaréis la solución en el solucionario.

Tabla 3 Ciencias sociales

3 Ciencias sociales. Sociología. Estadística. Política. Economía. Derecho.
 Administración pública. Arte y ciencia militares. Asistencia social.
 Seguros. Educación. Etnología

30 Teoría, metodología y métodos en las Ciencias sociales en general.
 Sociografía.

31 Estadística. Demografía. Sociología

32 Política

33 Economía. Economía política. Ciencia económica

34 Derecho. Jurisprudencia

35 Administración pública. Gobierno. Asuntos militares

36 Protección de las necesidades materiales y espirituales de la vida.
 Trabajo social. Vivienda. Consumismo. Seguros

37 Educación. Formación. Enseñanza. Usos o costumbres. Vida social.

38

39 Etnología. Etnografía. Usos y costumbres. Vida social. Folclore.
 Tiempo libre

Recordemos que las ciencias sociales son las que estudian la estructura y organización de la sociedad y que son: estadística, demografía, sociología, política, economía, derecho, administración y gobierno, educación, etnología. En definitiva, las ciencias que estudian al hombre en relación con su comunidad.

Como se puede apreciar, son muchas disciplinas bajo una misma clase; en consecuencia, es una de las tablas más subdivididas y en que las notaciones de temas importantes son más largas.

Otra característica de esta clase es la abundancia:

- de referencias de una clase a otra, por ejemplo de economía a política, ya que hay muchos temas que pueden ser tratados desde las dos perspectivas. De hecho, las ciencias sociales están muy interrelacionadas.
- de elementos facetados, sobre todo de las tablas de tiempo, lugar y persona, necesarias para representar el momento histórico de la sociedad, su ubicación geográfica...

Ejemplos de elementos facetados

Tiempo: para diferenciar la sociedad de 1800 de la actual

Lugar: para diferenciar la sociedad francesa de la rusa

Persona: para determinar que el objeto es una categoría concreta de la sociedad: adultos, empresarios...

Las disciplinas más extensas son economía y derecho.

Algunas particularidades son:

a) **Demografía, sociología, estadística:** la estadística sirve para clasificar documentos que tratan sobre estadística, pero no estadísticas concretas, que se clasificarán con la materia principal y la auxiliar de forma (083.41).

b) **Política:** formas de gobierno, política interior y exterior, elecciones, parlamentos, migraciones, partidos políticos. Es una clase que remite a otras clases que están relacionadas, como la sociología, economía, derecho. Por ejemplo, hay un paralelismo entre esta clase y la de derecho, por temas como elecciones y parlamentos, que se pueden clasificar desde los dos puntos de vista.

32 Política 324 Elecciones	34 Derecho 342.8 Sistemas electorales
-------------------------------	--

c) **Derecho:** sigue la división tradicional de derecho internacional, constitucional, penal, civil, eclesiástico y otras ramas (del trabajo, social, urbanístico, del medio ambiente, nuclear).

Actividad

4. Clasificad los siguientes documentos:

- Documento sobre la estadística como ciencia.
- Documento sobre las finanzas públicas.
- Documento sobre el derecho constitucional.
- Documento sobre las facultades universitarias.
- Documento sobre la etiqueta y el protocolo.

Hallaréis la solución en el solucionario.

Tabla 4 sin ocupar de momento

Está vacante desde 1964, momento en que el comité de la FID decidió agrupar las clases 4 y en 8 en la actual clase 8 de Lengua y literatura. Desde 1964 hasta ahora, no se ha encontrado un uso adecuado. Recordemos que en Cataluña, además de la CDU, existe la adaptación al catalán de Jordi Rubió Balaguer, que mantiene la clase 4 de Llengües.

Ved también

Ved el apartado dedicado a la historia de las clasificaciones documentales en España y Cataluña.

Tabla 5 Ciencias puras. Ciencias exactas y naturales

5 Ciencias puras. Ciencias exactas y naturales
502 Estudio de la naturaleza y su conservación. Ciencias naturales
504 Ciencias del medio ambiente. Ambientología

51 Matemáticas
 52 Astronomía. Astrofísica. Investigación espacial. Geodesia
 53 Física
 54 Química. Cristalografía. Ciencias mineralógicas
 55 Ciencias de la Tierra. Geología. Meteorología, etc.
 56 Paleontología
 57 Ciencias biológicas
 58 Botánica
 59 Zoología

La clase 5 representa las ciencias inanimadas o físicas, mientras que la clase siguiente, la 6, representa las ciencias animadas o de la vida.

Aquí encontramos las matemáticas y las ciencias naturales (astronomía, física, química, ciencias de la tierra, paleontología, biología, botánica y zoología). Son materias fáciles de identificar ya que son lo bastante disjuntas, pero requieren un cierto conocimiento de la materia a la hora de clasificar los documentos. Si no se tienen conocimientos previos en estas materias, se recomienda la consulta de obras de referencia, tipos diccionarios y enciclopedias, que ayuden a diferenciar, por ejemplo, el álgebra de la geometría, la química orgánica de la inorgánica, la geomorfología de la petrología, etc. Esta recomendación también es válida para las clases de la tabla 6.

Una particularidad de esta tabla es que se encuentra en paralelo a la tabla 6, ya que algunas ciencias de la clase 5 tienen sus aplicaciones prácticas en la 6.

Ejemplo

Los diamantes se encuentran clasificados en la clase 5, como química inorgánica 546 o como piedra preciosa 553, mientras que en la clase 6 aparecen en sus diversas aplicaciones:

- En ingeniería mecánica como un abrasivo en el 621.
- En minería como taladradoras de punta de diamante en el 622.23.
- En joyería como joyas en el 671.

Clase 5	Clase 6
546.26 Derivados orgánicos del carbono 553.8 Yacimientos de diamantes	621.921.34 Diamantes como abrasivos y productos para pulir 622.23 Taladros de diamantes 671.152 Talla del diamante

Actividad

5. Clasificad los siguientes documentos:

- Documento sobre aritmética.
- Documento sobre la galaxia.
- Documento sobre climatología.
- Documento sobre ammonitas (fósiles cefalópodos).

Hallaréis la solución en el solucionario.

Tabla 6 Ciencias aplicadas. Medicina. Técnica

6 Ciencias aplicadas. Medicina. Tecnología
60 Cuestiones generales
61 Medicina
62 Ingeniería
620 Ensayo de materiales. Materiales comerciales
621 Ingeniería mecánica en general. Tecnología nuclear. Ingeniería eléctrica. Maquinaria
622 Minería
623 Ingeniería militar
624 Ingeniería civil y estructural
625 Ingeniería del transporte terrestre
625.7/.8 Ingeniería de carreteras
626 Ingeniería hidráulica
627 Ingeniería de puertos y costas
628 Ingeniería de la salud pública. Agua. Sanidad. Ingeniería de la eliminación
629 Ingeniería de los vehículos de transporte
63 Ciencias agrícolas
64 Economía doméstica
65 Dirección y organización de la industria, comercio y comunicación
66 Tecnología química
67 Industrias, oficios y profesiones
68 Industrias, oficios y comercio de artículos acabados
69 Construcción

Es otra tabla con muchas materias de peso que acaban teniendo notaciones de cuatro o más cifras. Y tiene el honor de tener la clase más extensa, la 62, de Ingenieros y tecnología. La causa es que es el campo científico que más se ha desarrollado en los últimos años. De hecho, el desarrollo de las clases en la CDU es desigual.

En la clase 6 entran todas las ciencias aplicadas, como por ejemplo, ciencias agrícolas, economía doméstica, oficios y profesiones, industria del libro (editoriales, librerías).

Actividad

6. Clasificad los siguientes documentos:

- Documento sobre otorrinolaringología.
- Documento sobre la forma de preparar comidas rápidas para llevarse.
- Documento sobre editoriales (de libros).
- Documento sobre acabados especiales del cuero.

Hallaréis la solución en el solucionario.

Observación

Miguel Benito (1999, pág. 20) pone de manifiesto que, como la clase 62 es tan larga, es posible que dos notaciones con el mismo número de dígitos representen dos niveles muy diferentes con respecto a la especificidad del tema. Y pone de ejemplos como que la notación de 4 cifras 621.3 de Ingeniería electrónica sea igual a la 637.4, que es Huevos y productos a base de huevos. Si hacemos la prueba, veremos que la clase 621.3 se abre a muchas clases, mientras que la 637.4 no lo hace.

Tabla 7 Bellas artes. Artes aplicadas. Diversiones. Juegos. Deportes

```

7 art. Artes industriales. Fotografía. Música. Juegos. Deportes
  71 Urbanismo. Planificación física. Ordenación del territorio.
      Planificación regional, urbana y rural. Arquitectura de jardines.
      Configuración del paisaje
  72 Arquitectura
  73 Artes plásticas
  74 Dibujo. Diseño. Artes aplicadas y artesanía
  75 Pintura
  76 Artes gráficas
  77 Fotografía y procesos similares
  78 Música
  79 Diversión. Entretenimiento. Juegos. Deportes

```

Esta clase, aunque contiene diversas disciplinas, es de las más fáciles de entender para un profano, ya que todas las disciplinas son bastantes conocidas. Recordemos que es la clase que engloba todo tipo de artes, diversiones y deportes.

Si un artista es polifacético, se le clasifica directamente en el 7. Si, en cambio, queremos clasificar un arte concreto, vamos a la clase correspondiente.

En el inicio de la tabla 7 encontramos unos extensos auxiliares especiales que nos ayudarán a clasificar los aspectos de técnica artística (herramientas, materiales, copias, restauraciones...), periodos artísticos (prehistoria, culturas de la Antigüedad, románico, gótico... hasta el postmodernismo), temas representados (figura humana, fauna, pasajes históricos, paisajes...), etc.

Extraído de Carmen Díez (1999, pág. 108)

Documento sobre la obra de Leonardo da Vinci:
 7 Leonardo da Vinci
 Documento sobre la pintura de Leonardo da Vinci:
 75 Leonardo da Vinci

Actividad

7. Clasificad los siguientes documentos:
- Documento sobre parques privados.
 - Documento sobre catedrales.
 - Documento sobre óptica de las cámaras fotográficas.
 - Documento sobre música para orquesta sinfónica.
 - Documento sobre carreras de karts.

Hallaréis la solución en el solucionario.

Tabla 8 Lenguaje. Lingüística. Literatura

```

8 Lenguas. Lingüística. Literatura
  80 Cuestiones generales relativas a la lingüística y literatura. Filología
  81 Lingüística y lenguas
  82 Literatura

```

Hay dos grandes áreas, la 80 y 81 para Lingüística, y la 82 para Literatura. Aparentemente es una clase corta, pero es rica en auxiliares especiales y en divisiones paralelas.

Hay algunos ejemplos resueltos de lenguas y literatura de diversos países, pero no están representadas todas las posibilidades. La CDU nos indica que podemos construir todas las clases que necesitamos uniendo el 811 de lengua o el 821 de literatura a un **auxiliar de lengua**, sacando el signo igual y cambiándolo por un punto.

Ejemplo: 821.111

El 821 es la clase para literatura

El 111 proviene de: =111 aux. de lengua para el idioma inglés, cuando se combina para crear “literatura inglesa” pierde el signo inicial (=) y se cambia por un punto.

Para las obras de autores individuales, usaremos las especificaciones del 82....A/Z, que consisten en poner en primer lugar la literatura que escribe el autor; por eso tendremos que buscar el idioma original en que escribe, y si se trata de inglés, francés o castellano se indicará también la nacionalidad del escritor. Finalmente, el nombre del autor.

Ejemplo: 821.134.2(82)BORG

821 es Literatura en las distintas lenguas

134.2 proviene de: =134.2 aux. de lengua para el idioma español; cuando se combina para crear “literatura española” pierde el signo inicial del igual (=) y se cambia por un punto.

(82) es el auxiliar de lugar para Argentina, que añadimos a la literatura en español para diferenciar que se trata de literatura argentina en español.

BORG para el apellido del autor

Actividad

8. Clasificad los siguientes documentos:

- Documento sobre cuentos de hadas
- Documento sobre literatura hebrea

Hallaréis la solución en el solucionario.

Recordatorio importante

Como los auxiliares comunes de lengua no se explicarán hasta más adelante, recomendamos volver a este punto, una vez se hayan entendido estos auxiliares para resolver los ejercicios que se plantean a continuación.

Tabla 9 Arqueología. Geografía. Biografía. Historia

```

9 Geografía. Biografía. Historia
  902 Arqueología
  903 Prehistoria. Restos prehistóricos. Utensilios prehistóricos
  904 Restos culturales de tiempos históricos
  908 Descripción geográfica e histórica de un territorio
91 Geografía. Exploración de la Tierra y países concretos.
  Viajes. Geografía regional
929 Estudios biográficos y relacionados
93 Historia
94 Historia general

```

En esta tabla, encontramos el mismo caso que en la clase anterior: no hay una clase para la geografía de cada país ni para la historia. Tenemos que combinar la notación de la clase 913 o 94 con los **auxiliares de lugar y los auxiliares de tiempo**.

Ejemplo

Para representar historia de Irlanda se hará con el 94 de historia y con el auxiliar de lugar para Irlanda, que es (415).

Resultado:

94(415) Historia de Irlanda

Actividad

9. Clasificad los siguientes documentos:

- Documento sobre la geografía de África.
- Documento sobre datación con carbono.
- Documento sobre banderas nacionales.
- Documento sobre paleografía.
- Documento sobre historia de Australia.

Hallaréis la solución en el solucionario.

Recordatorio importante

Como los auxiliares comunes de lengua no se explicarán hasta más adelante, recomendamos volver a este punto, una vez se hayan entendido estos auxiliares para resolver los ejercicios que se plantean a continuación.

3.3. Tablas auxiliares de la clasificación decimal universal (CDU)

Cuando uno piensa en un cuadro de clasificación, se imagina páginas y páginas de códigos y conceptos. Cree que sólo ha de ir leyendo hasta que encuentre la clase más apropiada. Es cierto, pero todavía hay más cosas... La CDU incorpora elementos de tipo facetado, que son clases diseñadas para ser combinadas con otras clases.

Cuando estas facetas se pueden combinar con –todas– las nueve tablas, se conocen como *auxiliares comunes*. Para poner un ejemplo sencillo: los nombres de lugares geográficos, o los nombres de lenguas, sólo salen una vez en la CDU, pero se pueden combinar con cualquier clase de las nueve tablas, con lo cual las combinaciones son infinitas.

Cuando estas facetas sólo se aplican a una materia determinada, se conocen como *auxiliares especiales*. Hay algunos casos, por ejemplo la clase 53 de Física, que tienen auxiliares especiales, los cuales permiten precisar aspectos del tipo teoría y naturaleza de los fenómenos, efectos de los fenómenos, observación y registro, aparatos para su estudio..., que se pueden combinar con las materias propias de esta clase, la mecánica, la óptica, la termodinámica, la electricidad... Y eso sólo con la clase 53.

Función de los auxiliares

La función de los auxiliares es la de matizar y precisar el significado de las clases.

Concepto solo	Matizado
666.5 Porcelana	666.5(510)"14" Porcelana china del siglo XV

Tipo de auxiliares

La CDU tiene dos tipos de auxiliares:

- Auxiliares **comunes** (se abren en independientes y dependientes): se encuentran en las tablas auxiliares.
- Auxiliares **especiales**: no se encuentran reunidos en ninguna tabla, sino que se encuentran repartidos a lo largo de todas las tablas, tanto de las auxiliares como de las principales. Son un tipo de auxiliares previstos sólo para precisar el contenido de ciertas materias, no de todas.

Aunque la clasificación Dewey ya tenía auxiliares, la CDU los ha ampliado mucho.

Cuadro guía

En la tabla siguiente, anotamos las características principales que se irán explicando a lo largo de este apartado dedicado a la estructura de los auxiliares de la CDU:

Guía de los auxiliares comunes y especiales

	Auxiliares comunes		Auxiliares especiales
	Independientes	Dependientes	
Cuáles son	Ic Lengua Id Forma Ie Lugar If Razas, pueblos y nacionalidades Ig Tiempo	Ih Notaciones ajenas a la CDU Ik Características generales	Dependen de la materia que precisan. Un ejemplo: Periodos artísticos 7.03 dentro de la clase principal 7
Combinables con	Pueden ir solos o combinarse con todas las clases	Tienen que combinarse con las clases forzosamente.	Sólo con las clases para las cuales han estado previstos
Se encuentran a	Tablas propias llamadas <i>Tablas de auxiliares comunes independientes</i>	Tablas propias llamadas <i>Tablas de auxiliares comunes dependientes</i>	Repartidos en todas las clases que los necesitan: clases principales y tablas de auxiliares comunes como los de lugar y forma.
Los reconocemos por	Lengua = Forma (0..) Lugar (1/9) Razas, pueblos y nacionalidades (= ...) Tiempo "..."	Notaciones ajenas a la CDU A/Z Características generales de: Propiedades-02 Materiales-03 Persones-05	Por qué empiezan por.0/.9 0/-9 -00/-009, porque tienen una línea negra vertical o porque están encabezados con el nombre de auxiliares especiales.

A modo de conclusión

La función de los auxiliares es matizar y precisar el significado de las clases.

Hay dos tipos de auxiliares:

- Auxiliares **comunes** (se abren en independientes y dependientes): se encuentran en las tablas auxiliares.
- Auxiliares **especiales**: no se encuentran reunidos en ninguna tabla, sino que se encuentran repartidos a lo largo de todas las tablas, tanto de las auxiliares como de las principales. Son un tipo de auxiliares previstos sólo para precisar el contenido de ciertas materias, no de todas.

3.3.1. Tablas de auxiliares comunes

Independientes	Dependientes
Tabla de auxiliares comunes de Lengua	Tabla de auxiliares comunes. Notaciones ajenas a la CDU
=00 /03 Multilingüe =1/=2 Lenguas indoeuropeas =3... Lenguas caucásicas (vasco...) =4... Lenguas afroasiáticas =5... Lenguas uroaltaicas (japonés, coreano...) =6... Lenguas austroasiáticas =7... Lenguas indopacíficas, australianas =8... Lenguas indias americanas =9... Lenguas artificiales	* Números y códigos A/Z Nombres en orden alfabético
Tabla de auxiliares comunes de forma	Tabla de auxiliares comunes de características generales
(0.0...) Características físicas (01) Bibliografía (02) Libros en general (03) Obras de referencia (04) Separatas (05) Publicaciones en serie (06) Publicaciones de entidades (07) Documentos para enseñanza (08) Poligrafías (09) Presentación en forma histórica	De propiedades: -021 Propiedades de existencia -022 Propiedades de magnitud. Cantidad -023 Propiedades de forma. Dimensiones -024 Propiedades de estructura -025 Propiedades de disposición -026 Propiedades de acción y movimiento -027 Propiedades operacionales -028 Propiedades de estilo y presentación -029 Propiedades derivadas de otras clases principales De materiales: -032 Materiales minerales -033 Materiales manufacturados -034 Metales -035 Materiales orgánicos -036 Materiales macromoleculares -037 Textiles, fibras. -039 Otros materiales De Personas: -051 Personas como agentes, sujetos activos. -052 Personas como objeto, sujetos pasivos -053 Personas según la edad -054 Personas según etnia, nacionalidad... -055 Personas según sexo y parentesco -057 Personas según ocupación, trabajo -058 Personas según clase social y estado civil
Tabla de auxiliares comunes de lugar	
(1) Lugar y espacio en general. Localización. (2) Designación fisiográfica (3) Lugares del mundo antiguo (4) Europa (5) Asia (6) África (7) América del Norte y central (8) América del Sur (9) Australia. Ártico. Antártico	
Tabla de auxiliares comunes de razas, pueblos y nacionalidades	
(=081/=088) Razas y pueblos primitivos. Razas mixtas (=1.2/.9) Pueblos de regiones naturales (paralela a la tabla de lugar) (=11/=8) Pueblos y grupos lingüísticos y culturales diversos (paralela a la tabla de lengua)	
Tabla de auxiliares comunes de tiempo	
"0/2" Fechas y periodo de tiempo "3" Divisiones y subdivisiones del tiempo convencional "4" Periodos, edades "5" Frecuencia "6" Divisiones del tiempo geológicas, arqueológicas, culturales "7" Fenomenología del tiempo	

Tablas de auxiliares comunes independientes

Los auxiliares independientes son: lengua, forma, lugar, razas, pueblos y nacionalidades, tiempo. Como tienen un símbolo de inicio y uno de final pueden ir en cualquier lugar de la notación (la excepción es el de lengua, que sólo tiene símbolo en el inicio, pero la CDU permite añadirle detrás dos puntos).

Si en un mismo documento hay que representar más de un concepto con auxiliares independientes, la CDU recomienda el orden siguiente:

Clase + lugar + tiempo + forma + lengua + raza

A continuación, explicamos el campo de aplicación y los símbolos de cada auxiliar. Se proporcionan ejemplos en que sólo cambia el auxiliar para una rápida comprensión y aparecen ordenados según la lógica decimal de la CDU. Cada bloque tiene ejercicios, con la solución incluida, para practicar.

a) Tabla de auxiliares comunes de lengua

```
=00 /03 Multilingüe  
=1/=2 Lenguas indoeuropeas  
=3... Lenguas caucásicas (vasco...)  
=4... Lenguas afroasiáticas  
=5... Lenguas uroaltaicas (japonés, coreano...)  
=6... Lenguas austroasiáticas  
=7... Lenguas indopacíficas, australianas  
=8... Lenguas indias americanas  
=9... Lenguas artificiales
```

La lengua del documento se indica con el símbolo =1 / =9, igual a igual uno a 9. En teoría se puede aplicar a cualquier documento, indicando el idioma en que está escrito, pero en la práctica sólo es útil cuando hay que distinguir la lengua de los documentos. Ejemplo: ordenando todos los documentos en francés, otro conjunto en inglés, etc.

Documento sobre setas comestibles

```
635.8=133.1 Setas comestibles , escrito en francés  
635.8=134.1 Setas comestibles , escrito en catalán  
635.8=134.2 Setas comestibles , escrito en castellano
```

También se usan para crear las clases de Lengua y literaturas combinándolos con las clases de la tabla 8. Se combinan con el 811 para representar la lengua y con el 821 para representar literatura. Cuando se combinan, pierden el signo inicial del igual = por un punto.

Ejemplo

821.111 Literatura inglesa

El 821 es la clase para la literatura

Y el 111 proviene de: =111 auxiliar de lengua para el idioma inglés, cuando se combina para crear literatura inglesa, pierde el signo inicial del igual = por un punto.

Ejemplos de uso para crear clases de lengua y literatura:

811.133.1 Lengua francesa
 811.134.1 Lengua catalana
 811.134.2 Lengua castellana

821.133.1 Literatura francesa
 821.134.1 Literatura catalana
 821.134.2 Literatura castellana

Recordatorio importante

Ahora estáis en disposición de resolver los ejercicios planteados en el apartado anterior sobre las clases principales > clase 8 de Lengua y literatura.

b) Tabla de auxiliares comunes de forma

(0.0...) Características físicas
 (01) Bibliografía
 (02) Libros en general
 (03) Obras de referencia
 (04) Separatas
 (05) Publicaciones en serie
 (06) Publicaciones de entidades
 (07) Documentos para enseñanza
 (08) Poligrafías
 (09) Presentación en forma histórica

Indican la forma o presentación del documento (no la materia). Su símbolo es el (0...) paréntesis cero.

Hay que recordar que hay un paralelismo entre la tabla principal 0 y estos auxiliares de forma, ya que un concepto como el de “diccionarios” puede ser:

Una materia en sí misma –el estudio de los diccionarios– y se representaría con la clase 030.

Una forma –diccionario de francés– y se representaría con un auxiliar de forma (038)=133.1

Paralelismo con la tabla de auxiliares de forma

Paralelismo con la tabla de auxiliares de forma, con respecto a ciertas disciplinas como las enciclopedias, el diarios, el anuarios... que pueden ser traducidos al lenguaje documental como materia o como forma.

Concepto	Como materia, por lo tanto, clasificaremos con la tabla 0		Como forma, por lo tanto, clasificaremos con la tabla de auxiliares de forma	
Enciclopedias	030(460.23)	Las enciclopedias catalanas	(031)GEC	La <i>Gran Enciclopèdia Catalana</i>
Diarios	070	Los diarios	(054)EL PAIS	Diario <i>El País</i>

Ejemplo

77(031) Enciclopedia de fotografía
 77(035) Manual de fotografía
 77(051) Revistas de fotografía

Los auxiliares de forma se pueden colocar en cualquier lugar de la notación. A efectos prácticos, puede ser conveniente agrupar los documentos por forma, y así ordenar, por ejemplo, todos los diccionarios o todos los diarios juntos, etc.

Ejemplo

(038)=111 Diccionarios de inglés
 (038)= 124'02 Diccionarios de latín clásico
 (038)=133.1 Diccionarios de francés

Esta tabla tiene también auxiliares especiales para expresar características físicas, de producción etc.

Ved también

Ved el subpartado siguiente, dedicado a los auxiliares especiales, concretamente "Los auxiliares especiales de los auxiliares comunes".

c) Tabla de auxiliares comunes de lugar

(1) Lugar y espacio en general. Localización
 (2) Designación fisiográfica
 (3) Lugares del mundo antiguo
 (4) Europa
 (5) Asia
 (6) África
 (7) América del norte y central
 (8) América del sur
 (9) Australia. Ártico. Antártico

Indican el espacio geográfico, ya sea continente, país, nación, localidad u otras referencias de espacio (lugares físicos, lugares del mundo antiguo...). Se introducen entre paréntesis (1/9).

Ejemplo

599.742.71(355) Leones de la antigua Persia
 599.742.71(540) Leones de la India
 599.742.71(6) Leones africanos

El Estado español está representado por las comunidades autónomas y, dentro de cada una de ellas, por provincias.

(4) Europa
 (46) Península Ibérica
 (460) España
 (460.23) Comunidad Autónoma de Cataluña
 (460.231) Provincia de Lleida
 (460.232) Provincia de Girona
 (460.233) Provincia de Barcelona

```
(460.234) Provincia de Tarragona
(...)
(460.27) Comunidad Autónoma de Madrid
(...)
(460.35) Comunidad Autónoma de Andalucía
(460.351) Provincia de Córdoba
(460.352) Provincia de Jaén
(460.354) Provincia de Sevilla
(460.355) Provincia de Huelva
(460.356) Provincia de Cádiz
(460.357) Provincia de Málaga
(460.358) Provincia de Granada
(460.359) Provincia de Almería
```

También se usan para crear las clases de Geografía e Historia, combinándolos con las clases de la tabla 9.

Por ejemplo, se combinan con el 913/919 para representar la geografía y con el 94 para representar la historia. Cuando se combinan, no pierden el signo ().

Ejemplos

Ejemplo en que el (71) es el auxiliar de lugar para Canadá

```
913(71) Geografía de Canadá
94(71) Historia de Canadá
```

Ejemplos de uso para crear clases de Geografía e Historia

```
913(460.32) Geografía de las Islas Baleares
913(599) Geografía de Filipinas
913(82) Geografía de Argentina

94(460.32) Historia de las Islas Baleares
94(599) Historia de Filipinas
94(82) Historia de Argentina
```

Esta tabla tiene también auxiliares especiales para expresar características específicas de los lugares geográficos, como zonas, puntos cardinales, unidades políticas, etc.

d) Tabla de auxiliares comunes de razas, pueblos y nacionalidades

```
(=081/=088) Razas y pueblos primitivos. Razas mixtas
(=1.2/.9) Pueblos de regiones naturales
(paralela a la tabla de lugar)
(=11/=8) Pueblos y grupos lingüísticos y culturales diversos
(paralela a la tabla de lengua)
```

Ved también

Ved el subapartado siguiente dedicado a los auxiliares especiales, concretamente "Los auxiliares especiales de los auxiliares comunes".

Indican la nacionalidad o los aspectos étnicos de una materia. Su signo es el paréntesis igual (= ...), de hecho derivan principalmente de los auxiliares de lengua.

Ejemplos de construcción de auxiliares de raza desde los auxiliares de lengua

=214.58	Lengua romaní (gitana)
(=214.58)	Pueblo o nacionalidad gitana
=411.21	Lengua árabe
(=411.21)	Pueblo o nacionalidad árabe
=411.16	Lengua hebrea o judía
(=411.16)	Pueblo o nacionalidad judía

Estos auxiliares de razas, pueblos y nacionalidades son útiles para expresar grupos de países con algún rasgo común de raza o nacionalidad (por ejemplo, países de habla inglesa) o países que no se encuentran delimitados geográficamente (como el caso de los gitanos y los hebreos).

Ejemplo

Ved la diferencia de representación entre lo que sería el país Israel y los pueblos judíos:

(569.4)	Israel es un auxiliar de lugar.
(=411.16)	Judíos -de cualquier país del mundo, no sólo Israel- es un auxiliar de razas.

Ved la diferencia de representación entre lo que sería el país y el conjunto de países que hablan la misma lengua:

(41)	Gran Bretaña es un auxiliar de lugar.
(=111)	Países de habla inglesa -de cualquier país del mundo, no sólo de Gran Bretaña- es un auxiliar de razas.

e) Tabla de auxiliares comunes de tiempo

"0/2"	Fechas y periodo de tiempo
"3"	Divisiones y subdivisiones del tiempo convencional
"4"	Periodos, edades
"5"	Frecuencia
"6"	Divisiones del tiempo geológicas, arqueológicas, culturales
"7"	Fenomenología del tiempo

Indican la fecha, punto en el tiempo o intervalo de tiempo de una materia. Su signo son las comillas dobles "... " -no sirven para indicar la fecha de publicación del documento, que es tarea del análisis documental y no del análisis de contenido.

- Las fechas: se indican en la secuencia *año.mes.día*, con 4 dígitos para el año y dos para el mes y el día.

Recordatorio importante

Ahora estáis en disposición de resolver los ejercicios planteados en el capítulo 7.2 de las Clases principales > Clase 9 de Geografía e Historia.

Ved también

Ved la diferencia entre el análisis documental y el análisis de contenido en el módulo 1 de esta asignatura.

- Los siglos: se indican con dos dígitos.
- Los intervalos de tiempo: se indican poniendo entre la fecha inicial y la final una /

A modo de conclusión

Los auxiliares independientes son: lengua, forma, lugar, razas, pueblos y nacionalidades, tiempo:

- **Lengua:** indican la lengua del documento, con el símbolo =1 / =9, igual a uno igual a 9. También se usan para crear las clases de lengua y literatura combinándolos con las clases de la tabla 8.
- **Forma:** indican la forma o presentación del documento (no la materia). Su símbolo es el paréntesis cero (0...).
- **Lugar:** indican el espacio geográfico, ya sea continente, país, nación, localidad o bien otras referencias de espacio (lugares físicos, lugares del mundo antiguo...). Se introducen por (1/9) paréntesis uno a nueve. También se usan para crear las clases de Geografía e Historia combinándolos con las clases de la tabla 9.
- **Razas y pueblos:** indican la nacionalidad o los aspectos étnicos de una materia. Su signo es el paréntesis igual (=...); de hecho, derivan principalmente de los auxiliares de lengua.
- **Tiempo:** indican la fecha, punto en el tiempo o intervalo de tiempo de una materia. Su signo son las comillas "...".

Como tienen un símbolo de inicio y uno de final, pueden ir a cualquier sitio de la notación.

Si en un mismo documento hay que representar más de un concepto con auxiliares independientes, la CDU recomienda el orden siguiente:

Clase + lugar + tiempo + forma + lengua + raza

Bibliografía

F. J. Álvarez García (1997). "Introducción al uso práctico de la CDU". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 219-243). Madrid: Síntesis ("Biblioteconomía y Documentación").

M. Benito (1999). *El sistema de clasificación decimal universal: manual de aprendizaje*. Madrid: [en línea].

M. Benito. *El sistema de Clasificación Decimal Universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008]. O, también, en la URL Instituto de Taranco [en línea]. [Fecha de consulta: 10 de octubre de 2008].

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

C. Díez Carrera. (1999). *Técnicas y régimen de uso de la CDU (Clasificación Decimal Universal)* (pág. 134). Gijón: Trea. ("Biblioteconomía y Administración Cultural", 26).

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

L. M. Moreno Fernández; M. D. Borgoños Martínez (2002). *Teoría y práctica de la clasificación decimal universal* (2.ª ed. revisada, corregida y aumentada). Gijón: Trea.

Fechas

"2000.01.01" es el uno de enero de 2000.
Si la fecha es anterior a Cristo, se indica con un guión en frente:
"-0350" es el año 350 antes de Cristo.

Siglos

Fijaos en que la numeración se corresponde a la centuria y no a la expresión habitual en números romanos.
"06" siglo VII, es decir, los años 600 -
"14" siglo XV, es decir los años 1400 -
"19" siglo XX, es decir los años 1900 -

Ejemplo de intervalos de tiempo

"1910/1970": del año 1910 a 1970
"11/18": del siglo XII al XIX

Tablas de auxiliares comunes dependientes

Los auxiliares dependientes son dos tablas: notaciones ajenas a la CDU y características generales (de propiedades, materiales y personas). Como su nombre indica, son auxiliares que sólo se pueden usar como sufijos; no se han de usar independientemente ni en el primer lugar de la notación. Han de ir siempre al final.

a) Tabla de auxiliares comunes de notaciones ajenas a la CDU

* Números y códigos
A/Z Nombres en orden alfabético

El asterisco (*) marca el código que no es propio de la CDU, pero que se utiliza en su notación. Suelen ser temas que la CDU no trata y que aparecen, en cambio, en otras clasificaciones o listas. Hay que anotar en el fichero de autoridades cuáles son estas otras fuentes.

Ejemplo

(460*17491) Peralada Font: código postal correos.

La localidad de Peralada, en el Alto Ampurdán, tiene por código postal el 17491.

El A/Z nos permite poner nombres propios y geográficos, siglas y abreviaturas a números de la CDU. Cuando acompañan auxiliares de lugar, van entre paréntesis.

Ejemplo

(460.233 San Cugat) Ciudad de San Cugat en la provincia de
Barcelona
821.133.1 Verne Literatura francesa de Jules Verne
929:791.44 Newman Biografía de Paul Newman -la notación
expresa biografías + cine + Paul Newman-

Observación

Los nombres se pueden abreviar con las tres o cuatro primeras letras, tipo BARC / VERN / NEWM.

b) Tabla de auxiliares comunes de características generales: propiedades, materiales y personas

- De propiedades:

-021 Propiedades de existencia
-022 Propiedades de magnitud. Cantidad
-023 Propiedades de forma. Dimensiones
-024 Propiedades de estructura
-025 Propiedades de disposición
-026 Propiedades de acción y movimiento
-027 Propiedades operacionales
-028 Propiedades de estilo y presentación
-029 Propiedades derivadas de otras clases principales

Los auxiliares de propiedades: derivan de la antigua tabla de auxiliares de punto de vista (cancelada en 1998) y denotan las propiedades generales o atributos.

Ejemplo

656.23.03-027.567 Precios subvencionados de ferrocarriles.

- De **materiales**:

```
-032 Materiales minerales
-033 Materiales manufacturados
-034 Metales
-035 Materiales orgánicos
-036 Materiales macromoleculares
-037 Textiles, fibras.
-039 Otros materiales
```

Los auxiliares de materiales señalan los materiales o componentes de los objetos.

Ejemplo

A partir de -035. 3 madera y 685.341.3 zapatos, según el material de fabricación, construimos zapatos de madera: 685.341.353 Zuecos

La fabricación y procesamiento de los materiales se encuentra en la clase principal 6, clases 66/67.

- De **personas**:

```
-051 Personas como agentes, sujetos activos
-052 Personas como objetos, sujetos pasivos
-053 Personas según la edad
-054 Personas según etnia, nacionalidad...
-055 Personas según género y parentesco
-057 Personas según ocupación, trabajo
-058 Personas según clase social y estado civil
```

Los auxiliares de personas denotan a las personas a las cuales hace referencia el tema clasificado.

Roles y combinaciones de auxiliares

Observad el rol activo (clase -051) o pasivo (clase -052) de la persona, y la combinación de dos auxiliares de persona en la misma notación (-055. 2 sexo femenino). Ejemplo de McIlwaine (2003, pág. 103):

```
75.041.5 Retrato
75.041.5-051 Retratista
75.041.5-052 Persona que ha sido retratada
75.041.5-052-055.2 Mujeres retratadas
```


A modo de conclusión

Los auxiliares dependientes son:

- **Notaciones ajenas:** * números y códigos y A/Z Nombres en orden alfabético. El asterisco * marca el código que no es propio de la CDU, pero que se utiliza en la notación. Suelen ser temas que la CDU no trata y que aparecen, en cambio, en otras clasificaciones o listas. Nos permite poner nombres propios y geográficos, siglas y abreviaturas a números de la CDU.
- **Características generales:** pueden ser características de propiedades, materiales o personas. Como su nombre indica, son auxiliares que sólo se pueden usar como sufijos; no se han de usar independientemente ni en primer lugar, y van siempre al final de la notación.

Más información

F. J. Álvarez García (1997). "Introducción al uso práctico de la CDU". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 219-243). Madrid: Síntesis ("Biblioteconomía y Documentación").

M. Benito (1999). *El sistema de clasificación decimal universal: manual de aprendizaje*. Madrid: Taranco.

M. Benito. *El sistema de Clasificación Decimal Universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008]. O, también, en la URL Instituto de Taranco. [Fecha de consulta: 10 de octubre de 2008].

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

C. Díez Carrera (1999). *Técnicas y régimen de uso de la CDU (Clasificación Decimal Universal)* (pág. 134). Gijón: Trea ("Biblioteconomía y administración cultural", 26).

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

L. M. Moreno Fernández; M. D. Borgoños Martínez (2002). *Teoría y práctica de la Clasificación Decimal Universal* (2.ª ed. revisada, corregida y aumentada). Gijón: Trea.

3.3.2. Tablas de auxiliares especiales

Los auxiliares especiales ayudan a precisar el contenido de algunas materias, pero no de todas.

Al contrario que los auxiliares comunes, que se encuentran reunidos en una sola tabla, los auxiliares especiales se encuentran extendidos, repartidos por todas las clases principales y algunas tablas de auxiliares.

A continuación, mencionamos algunas observaciones que hay que tener en cuenta con respecto a los auxiliares especiales.

a) ¿Todas las clases tienen auxiliares especiales?

No, no todas las clases tienen auxiliares especiales, ya que no todas las temáticas son susceptibles de tener elementos de matización comunes.

b) ¿Cómo se localizan?

La manera en que podemos localizar a los auxiliares especiales dependerá de si consultamos una edición en papel (como la edición de bolsillo de la CDU) o una edición digital (como la que tenéis en el aula). Sin embargo, los reconoceréis porque siempre empiezan por las secuencias siguientes:

- Punto cero a punto nueve **.0/-9**: sirven para crear conjuntos y subconjuntos de conceptos.
- Guión cero a guión nueve **-0/-9**: tienen una función analítica.
- Apóstrofo cero a apóstrofo nueve **>'0/'9**: son más específicas que las de guión cero y tienen una función sintetizadora.

En las **ediciones en papel**, estos auxiliares especiales se marcan con una línea negra vertical.

Auxiliares especiales en ediciones en papel

La clase 02 Técnica teatral se ve claramente que es un auxiliar especial porque tiene la línea negra y empieza por punto cero.

792 Teatro, ciencia del teatro
 → 725,82; 782; 82-2

	.02 Técnica teatral
	.022 Tramoya. Luminotecnia. Talleres teatrales
	.023 Attrezzo. Decorados
	.024 Vestuario teatral
	.026 Libretos. Partituras
	.027 Dirección artística

En las versiones digitales, los auxiliares especiales se reconocen porque habremos hecho clic en un enlace que dice “auxiliares especiales”, o bien porque empiezan de una de estas maneras:

```
.0/.9 punto cero a punto 9
-0/-9 guión cero a guión 9
'0/'9 apóstrofo cero a apóstrofo nueve
```

Auxiliares especiales en versiones digitales

Si vemos el mismo caso en la versión digital: primero hacemos clic donde dice auxiliares especiales, y una vez dentro, vemos que se introducen por punto cero.

c) ¿Misma numeración, mismo significado?

Los auxiliares especiales representan las facetas particulares de las materias y sólo tienen sentido de aplicación en su ámbito temático. En la CDU, un mismo auxiliar como -4 significa conceptos bien diferentes según la clase en la que se aplique:

```
54-4 Productos químicos, reactivos
62-4 Estado de los materiales en ingeniería
82-4 Ensayos, literatura
```

d) ¿Cómo se combinan?

Tenemos que ver en qué clase están adscritos los auxiliares, ya que sólo se podrán combinar con esta o con otras clases que empiecen de la misma manera.

Ejemplos de auxiliares especiales

Auxiliar especial introducido por un guión:

Notación final:	244.82-23 Libros sagrados del budismo zen
Construida a partir de	24-23 Libros sagrados 244.82 Budismo zen
Explicación	En primer lugar, ponemos la clase para el budismo zen y añadimos el auxiliar especial. Prestamos atención al hecho de que, cuando se combinan, desaparece el 24 del auxiliar. Otras combinaciones: siempre que la clase empiece por 24 242-23: Libros sagrados del budismo Mahayana

El 24, que desaparece cuando se combina, nos indica que esta posición puede ser ocupada por cualquier clase que empiece por 24 (como la 244 de Budismo zen o la 242 de Budismo Mahayana), pero no por una clase como la 28 del Islam (28 no es 24).

Auxiliar especial introducido por punto:

Notación final:	749.035.93 Mobiliario modernista
Construida a partir de:	7.035.93 Modernismo 749 Mobiliario artístico
Explicación:	Ponemos en primer lugar la clase para mobiliario (punto de vista de las artes aplicadas) y le añadimos el auxiliar especial que hace referencia al periodo artístico. Vemos que, cuando se combinan, desaparece el 7 del auxiliar. Otras combinaciones: siempre que la clase empiece por 7.

Los auxiliares especiales de la clase 7 son extensos y se aplican a todas las clases que empiecen por 7, a excepción de la clase 77, es decir, las clases 71, 72, 73, 74, 75, 76, 78, 79, lo cual los convierte en uno de los auxiliares más combinables de las tablas.

```
7.035.93  Arte modernista
72.035.93 Arquitectura modernista
73.035.93 Escultura modernista
74.035.93 Dibujo modernista
75.035.93 Pintura modernista
76.035.93 Grabado modernista
(...)
```

e) Los auxiliares especiales de los auxiliares comunes

La CDU es rica y potente y ha previsto que incluso algunos auxiliares comunes –como los de lengua, forma y lugar– puedan tener sus propios elementos facetados. Haciendo un juego de palabras, podríamos decir que son “facetas de facetas”.

Con un ejemplo se entenderá rápidamente. Para representar el concepto sur de Italia, la notación es la siguiente, donde la (450) es Italia y el (1-13) es sur (este es el auxiliar especial, concretamente de la tabla de lugar):

(450-13) Sur de Italia

Una vez combinados el 1 ha desaparecido, porque si no fuera así, habría significado en Italia y en el sur en general.

(450-13) Sur de Italia

(450)(1-13) Italia y el sur

Los auxiliares especiales que tienen las tablas de auxiliares comunes son:

- En la tabla de lengua: por orígenes y periodos del tipo (antiguo, clásico, medieval, moderno...). Por ejemplo:

=14'04 Griego medieval o bizantino

En este ejemplo, =14 es *Griego* > y el ' auxiliar especial es =...'04 (hay que llenar los puntos sucesivos con la clase de la lengua que estamos caracterizando).

- En la tabla de forma: por características físicas o de producción, del tipo miniatura, de bolsillo, en soportes magnéticos... Por ejemplo:

025.45.027.5 CDU edición bolsillo

Donde 025.45 es clasificaciones decimales y el auxiliar especial es (0.027.5) para ediciones de bolsillo –el primer cero desaparece cuando se combina con la clase 025.45.

- En la tabla de lugar: para fronteras, puntos cardinales (norte, sur, este, oeste y combinaciones), unidades políticas o administrativas (municipios, comunidades, capitales, imperios, colonias, grupos de Estados...). Por ejemplo:

(4-662) Países capitalistas europeos

Donde (4) es Europa y el auxiliar especial es (1-662) para países capitalistas –el primer uno desaparece cuando se combina con la clase 4.

A modo de conclusión

Los auxiliares especiales ayudan a precisar el contenido de algunas materias, pero no de todas.

No hay una tabla de auxiliares especiales, sino que se encuentran repartidos por todas las clases principales y las tablas auxiliares que disponen. Los reconocemos porque empiezan por:

- Punto cero a punto nueve .0/.9 sirven para crear conjuntos y subconjuntos de conceptos.
- Guión cero a guión nueve -0/-9 tienen una función analítica.
- Apóstrofo cero a apóstrofo nueve '0/'9 son más específicas que las de guión cero y tienen una función sintetizadora.

Más información

F. J. Álvarez García (1997). "Introducción al uso práctico de la CDU". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 219-243). Madrid: Síntesis ("Biblioteconomía y Documentación").

M. Benito (1999). *El sistema de clasificación decimal universal: manual de aprendizaje*. Madrid: Taranco.

M. Benito. *El sistema de Clasificación Decimal Universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008]. O, también, en la URL Instituto de Taranco. [Fecha de consulta: 10 de octubre de 2008].

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

C. Díez Carrera. (1999). *Técnicas y régimen de uso de la CDU (Clasificación Decimal Universal)* (pág. 134). Gijón: Trea ("Biblioteconomía y Administración Cultural", 26).

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

L. M. Moreno Fernández; M. D. Borgoños Martínez (2002). *Teoría y práctica de la Clasificación Decimal Universal* (2.ª ed. revisada, corregida y aumentada). Gijón: Trea.

3.4. Signos con función combinatoria y signos con función instructiva

Las clases se pueden combinar entre sí, siguiendo la sintaxis que marca la misma CDU, y por este motivo son necesarios los signos clasificatorios. O pueden remitir unas a otras, en beneficio de una mejor clasificación, y por este motivo necesitaremos saber interpretar los signos no clasificatorios que acompañan las rúbricas.

En la CDU hay dos tipos de signos:

- 1) Los que tienen una función combinatoria.
- 2) Los que tienen una función puramente instructiva.

3.4.1. La importancia de los signos

Los signos son importantes porque forman parte de la sintaxis de la notación: los signos permiten diferenciar elementos y ordenar los documentos.

Por ejemplo:

```
=581 lengua china es un auxiliar de lengua  
(581) Afganistán es un auxiliar de lugar  
"0581" Año 581 dC es un auxiliar de tiempo  
-058. 1 Personas según clase social es una auxiliar de características personales
```

581 Botánica es una clase de las tablas principales

Como se puede apreciar, son conceptos bien diferentes. Y la diferencia radica en los símbolos que los caracterizan: el igual, el paréntesis, las comillas, el guión o la ausencia de símbolos (como en las tablas principales).

Para una mejor comprensión, y vista la diversidad de signos, funciones y localizaciones dentro de las tablas, hemos decidido agrupar todos los signos de la CDU en este subapartado.

Función	Signo	Se lee	Función	¿Dónde se encuentran?
Combinatoria	+	Más	Suma	Tabla aux. comunes la
	/	Barra	Extensión	Tabla aux. comunes la
	:	Dos puntos o Colon	Relación	Tabla aux. comunes lb
	::	Doble Colon	Nexo indisoluble de relación	Tabla aux. comunes lb
	[...]	Corchetes	Agrupar notaciones	Tabla aux. comunes lb
	=...	Igual	Lengua	Tabla aux. comunes lc
	(0...)	Paréntesis cero	Forma	Tabla aux. comunes ld
	(...)	Paréntesis	Lugar	Tabla aux. comunes le
	(=...)	Paréntesis igual	Razas y pueblos	Tabla aux. comunes lf
	" "	Comillas	Tiempo	Tabla aux. comunes lg
	*	Asterisco	Combinación con otros sistemas externos a la CDU	Tabla aux. comunes lh
	A/Z	A a la Z	Ordenación alfabética	Tabla aux. comunes lh
	-0...	Guión cero	Características generales	Tabla aux. comunes lk
	-...	Guión	Aux. especial, analítico	Repartido por las tablas
	.0...	Punto cero	Aux. especial, conjuntos	Repartido por las tablas
e...	apóstrofo	Aux. especial, sintetizador	Repartido por las tablas	
Puramente instructiva	.	Punto	Separar grupos de tres cifras	Repartido por las tablas
	...	Puntos suspensivos	Sustitución de posibles cifras	Repartido por las tablas
	→	Ved	Indica que hay otros números a consultar	Repartido por las tablas
	@	Subdividid como	Subdivisión paralela	Repartido por las tablas

3.4.2. Los signos con función combinatoria

Los signos con función combinatoria son los signos que permiten unir dos o más clases de la CDU para formar un concepto que no estaba recogido previamente.

Los signos con función combinatoria tienen una función sintetizadora; permiten la expresión de frases documentales compuestas.

Son los signos siguientes:

- + Se usa para representar un concepto compuesto para el cual no existe un número simple previsto. Une dos o más clases no consecutivas de las tablas.

(450 + 494) Italia y Suiza

- / Se usa para representar un concepto que no aparece en una única clase en el sistema, sino que aparece en diferentes clases consecutivas.

73/75 documento sobre artes plásticas, dibujo y pintura que equivale a las clases 73, 74 y 75

- : Se usa para representar dos conceptos relacionados. Las relaciones pueden ser diversas: de relación en general, de comparación, de diferencia, de influencia.

91:93 Relación entre Geografía e Historia

- :: Si cuando se colocan los números de las dos clases relacionadas cambia el significado, entonces se usa el doble Colon. Se considera el primer número como el tema principal y el segundo como dependiente. Sólo se entra en el catálogo una vez.

070::32 Periodismo político

- [...] Se usa para agrupar números y clarificar las relaciones entre las partes de la notación resultante.

[72+75](450) Arquitectura y pintura italianas

Colon o doble Colon

Si utilizáramos el Colon 070:32 significaría relación entre periodismo y política; en cambio, nos interesa remarcar que es periodismo político y no otra combinación, como por ejemplo, la política del periodismo.

El resto de signos van asociados a los auxiliares comunes y especiales de la CDU, explicados en los subapartados anteriores.

3.4.3. Los signos puramente instructivos

Los signos puramente instructivos son los signos sin valor clasificatorio.

Observación

Mientras que los signos + y / unen conceptos que aparecen separados o de manera consecutiva en las tablas, los signos : y :: unen conceptos y denotan claramente la relación que hay entre ellos.

Los signos puramente instructivos son los siguientes:

- . (**punto**). El punto separa las cifras de la notación de tres en tres, para facilitar la lectura, pero no tiene valor clasificatorio.

- \cong (**dividir como**). Encontramos este signo cuando una parte de las tablas se subdivide igual que otra. Por ejemplo: 675.025 \cong 675.055 significa que la primera clase se subdivide como la segunda. Así: 675.025.35 Pulir el cuero proviene de:

675.025 \cong 675.055

675.025 Tratamiento del cuero

675.055.35 Máquinas para pulir el cuero

Esta práctica se va abandonando poco a poco, tal como dice McIlwaine y, en cambio, se introducen los dos puntos.

“Esto se debe al hecho de que estas notaciones no tienen la apariencia de haber sido construidas por la combinación de dos números separados y, en consecuencia, el ordenador es incapaz de detectar en qué casos se trata de un concepto compuesto y de recuperar la información requerida.”

McIlwaine (pág. 61)

Este signo puede indicar que una clase se subdivide como otra clase o que se subdivide como una tabla de auxiliares. Por ejemplo, 616-05 \cong -05 tabla Ik significa que los auxiliares especiales de la clase 616-05 de patología se abren igual que los auxiliares de la tabla de características de personas:

616-056.25 Enfermedades de personas desnutridas,
donde -056.25 es el estado de nutrimento de las personas.

- \rightarrow (**Ved también**). Este signo llama la atención sobre ámbitos temáticos parecidos o afines con respecto a un tema en particular. Así, 027.7 Bibliotecas universitarias \rightarrow 378.16 significa que si estamos consultando la clase 027.7 sobre bibliotecas universitarias también nos puede interesar consultar la clase 378.16, de campus, edificios e instalaciones universitarias.

3.4.4. Conclusiones

En la CDU, hay dos tipos de signos:

Los que tienen una **función combinatoria**. Son los signos que permiten unir dos o más clases de la CDU para formar un concepto de lo que estaba recogido previamente. Tienen una función sintetizadora; permiten la expresión de frases documentales compuestas. Son los signos: + /::: []

Los que tienen una **función puramente instructiva**. Son los signos sin valor clasificatorio, pero que tienen una función instructiva. Son los signos . (punto), ≅ (dividir como) y → (ved también).

Más información

F. J. Álvarez García (1997). "Introducción al uso práctico de la CDU". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 219-243). Madrid: Síntesis ("Biblioteconomía y Documentación").

M. Benito (1999). *El sistema de clasificación decimal universal: manual de aprendizaje*. Madrid: Taranco.

M. Benito. *El sistema de Clasificación Decimal Universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008]. O, también, en la URL Instituto de Taranco. [Fecha de consulta: 10 de octubre de 2008].

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

C. Díez Carrera (1999). *Técnicas y régimen de uso de la CDU (Clasificación Decimal Universal)* (pág. 134). Gijón: Trea ("Biblioteconomía y Administración Cultural", 26).

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

L. M. Moreno Fernández; M. D. Borgoños Martínez (2002). *Teoría y práctica de la Clasificación Decimal Universal* (2.^a ed. revisada, corregida y aumentada). Gijón: Trea.

4. Empezamos a clasificar

Para empezar a clasificar, se trata de seguir la estrategia siguiente:

- 1) Analizar el contenido del documento y decidir la clase o las clases. Es especialmente importante cuando el tema tiene diversos enfoques o el contenido es múltiple.
- 2) Consultar la tabla o las tablas principales, prestando atención a si tiene auxiliares especiales. Hay que consultar los auxiliares comunes si son necesarios.
- 3) Comprobar la clase en el índice alfabético o el buscador del programa CDU. Es muy importante que el estudiante sea consciente de que los índices y buscadores son ayudas adicionales y no el medio principal para localizar la clase. Si la materia es polisémica o tiene diversos enfoques, el índice o buscador nos puede llevar a error.

Algunos ejemplos

Ejemplo de término con polisemia: documento sobre los caballitos de feria

```
595.7 Ácidos dicarboxílicos de ciclohexanos
597.5 Teleósteos
688.774 Juguetes. Artículos de diversión e ilusionismo
```

Nos interesa la última clase, la 688.774; las otras dos serían un error (la primera hace referencia a los insectos caballitos del diablo, la de teleósteos hace referencia a los caballitos de mar).

Ejemplo de término con diversos enfoques: documento sobre la poligamia

```
173.2 Poligamia (desde el punto de vista de la ética familiar)
314.545.4 Poligamia (desde el punto de vista de la demografía,
    estudios de población según nupcialidad)
392.544 Poligamia (desde el punto de vista de la etnología,
    usos y costumbres de la vida privada)
613.89 Poligamia (desde el punto de vista de la higiene en
    general y la sexual en particular)
```

Si nuestro documento tuviera un carácter antropológico, lo clasificaríamos en 392.544.

Una vez tenemos el documento enfrente, nos podemos encontrar en las situaciones siguientes:

1. Encontramos una clase a la primera y de manera unívoca

El documento tiene una materia que se corresponde con un solo índice.

61 Medicina

2. El documento tiene más de una materia, pero la CDU las recoge juntas

332.8 Economía de la vivienda

3. Tenemos que combinar dos o más clases de las tablas principales porque el concepto no existe *a priori*

3.1. Cuando queramos unir dos materias independientes, pero que están relacionadas (para formar un nuevo concepto o para precisarlo), utilizaremos el signo:

17:61

61:17 Medicina y ética -como materias relacionadas-

En este caso, la CDU permite introducir la notación dos veces en el catálogo para poder recuperarlo tanto si se hace la busca por ética como por medicina.

3.2. Cuando queramos unir dos materias dependientes y relacionadas en un orden fijo, utilizaremos el signo :: (en este caso, para diferenciarlo de la economía de las asociaciones):

061::33 Asociaciones económicas

3.3. Cuando queramos unir dos materias, con una relación simple de adicción, utilizaremos el signo +.

77 + 791 Fotografía y cine

3.4. Cuando queramos unir dos materias, expresadas consecutivamente en las tablas, utilizaremos el signo /.

58/59 Botánica y zoología

4. Tenemos que matizar una clase de las tablas principales con algún tipo de auxiliar porque el concepto no existe *a priori*.

4.1. Cuando la faceta sea de tipo universal en todas las tablas como la de lengua, forma, lugar, de razas y de tiempo.

En primer lugar, el número de la clase que representa el contenido mayoritario del documento y, acto seguido, el auxiliar común. Si hubiera más de un auxiliar común, la CDU recomienda el orden siguiente:

Clase + lugar + tiempo + forma + lengua + raza.

Ejemplo

398.332.34(495.2)“17” Fiestas de la vendimia en el Peloponeso en el siglo XVIII.

Como estos auxiliares son independientes, también podemos ponerlos en el inicio de la notación por tal de agrupar todos los documentos por forma, idioma...

(495.2)398.332.34“17” Fiestas de la vendimia en el Peloponeso en el siglo XVIII, teniendo en cuenta que ordenamos todos los documentos del Peloponeso juntos.

4.2. Cuando queramos expresar una faceta de tipo particular, de las que sólo afectan a una materia, pondremos en primer lugar el número de la clase que representa el contenido mayoritario del documento y, acto seguido, el auxiliar especial.

611.3.01 Anatomía del aparato digestivo

4.3. Si el documento tuviera al mismo tiempo un auxiliar especial y uno común, el orden sería:

Clase + auxiliar especial + auxiliar común.

Ejemplo

28-57(532) Documento sobre la peregrinación musulmana a La Meca

28 es Islam, 2-57 es el auxiliar especial para el concepto peregrinación y (532) es el auxiliar de lugar para la Arabia Saudí.

A modo de conclusión

Cuando empezamos a clasificar un documento hemos de analizar el tema o los temas, consultar las tablas principales y las de auxiliares comunes y comprobar si disponen de auxiliares especiales.

No se recomienda usar el índice alfabético o el buscador del programa CDU, porque si la materia es polisémica o tiene diversos enfoques nos puede llevar a error.

Si hay que construir una notación, puede ser el resultado de las siguientes combinaciones:

- Una clase principal + clase principal (unidas con + /:).
- Una clase principal + auxiliar común (lugar o tiempo o forma o lengua o raza).
- Una clase principal + auxiliar especial.
- Una clase principal + auxiliar especial + auxiliar común.

Más información

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Díez Carrera, C. (1999). *Técnicas y régimen de uso de la CDU (Clasificación Decimal Universal)* (pág. 134). Gijón: Trea ("Biblioteconomía y Administración Cultural", 26).

5. Prácticas con la CDU

A continuación, se propone una serie de ejercicios ordenados según el grado de dificultad para un estudiante que se introduce en este lenguaje documental. No se presentan todas las casuísticas que ofrece este lenguaje documental, pero sí algunas de las más habituales.

Más información

M. Benito (1999). *El sistema de clasificación decimal universal: manual de aprendizaje*. Madrid: Taranco.

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

C. Díez Carrera (1999). *Técnicas y régimen de uso de la CDU (Clasificación Decimal Universal)* (pág. 134). Gijón: Trea ("Biblioteconomía y Administración Cultural", 26).

I. C. McIlwaine (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

5.1. Ejercicios de nivel bajo

Clasificad los documentos siguientes:

Clase 0

- Documento sobre bibliografías de autor:
- Documento sobre catálogos por materias:
- Documento sobre las cartas al director de un diario:
- Documento sobre ex-libris:

Clase 1

- Documento sobre métodos de filosofar:
- Documento sobre estados emocionales:
- Documento sobre lógica proposicional:
- Documento sobre ética profesional:

Clase 2

- Documento sobre la Iglesia católica romana:
- Documento sobre el Islam chiíta:
- Documento sobre el agnosticismo:
- Documento sobre el confucianismo (religión china):

Objetivos

Estos ejercicios os tienen que permitir alcanzar los objetivos siguientes:

- Aprender a ubicar conceptos en las tablas principales. Suelen ser conceptos que tienen entrada propia en la CDU (no hay que construir la notación). Para facilitar la tarea, se presentan ordenados por tablas.
- Aprender a situar un concepto según el enfoque adecuado.

Clase 3

- Documento sobre teorías de la población:
- Documento sobre campañas electorales:
- Documento sobre el Ministerio de Hacienda:
- Documento sobre el poder adquisitivo del consumidor:

Clase 4

Clase 5

- Documento sobre ciencias del medio ambiente:
- Documento sobre la capacidad calorífica de los líquidos:
- Documento sobre hidrocarburos saturados:
- Documento sobre silicatos básicos:
- Documento sobre biología:
- Documento sobre botánica - las magnolias:
- Documento sobre la ballena azul:

Clase 6

- Documento sobre psicoterapia:
- Documento sobre drenaje de minas:
- Documento sobre enfermedades por hongos de las plantas:
- Documento sobre instalaciones de agua potable (en las viviendas):
- Documento sobre producción de sosa y potasa:

Clase 7

- Documento sobre escultura:
- Documento sobre dibujo anatómico:
- Documento sobre diseño de carteles:

Clase 8

- Documento sobre la lengua coreana:
- Documento sobre literatura griega:

Clase 9

- Documento sobre pinturas rupestres:
- Documento sobre viajes de descubrimientos geográficos:
- Documento sobre órdenes de caballería:
- Documento sobre la tercera guerra carlista:

Diversos enfoques de las materias:

- Documento sobre el ateísmo, desde el punto de vista filosófico:
- Documento sobre el ateísmo, desde el punto de vista religioso:
- Documento sobre las elecciones políticas, desde el punto de vista político:
- Documento sobre las elecciones políticas, desde el punto de vista del derecho:
- Documento sobre los delitos contra los derechos humanos, desde el punto de vista del derecho penal:
- Documento sobre los delitos contra los derechos humanos, desde el punto de vista del derecho internacional:
- Documento sobre los delitos contra los derechos humanos, desde el punto de vista del derecho constitucional:

5.2. Ejercicios de nivel medio

Clasificad los siguientes documentos

Clase 0

- Documento sobre las piezas prehistóricas de un museo:
- Documento sobre la clasificación decimal Dewey:
- Documento sobre bibliografía nacional francesa:
- Documento sobre thesaurus de psicología:

Clase 1

- Documento sobre la filosofía de Arthur Schopenhauer:
- Documento sobre el esnobismo inglés:

Clase 2

- Documento sobre los santos de la religión cristiana:
- Documento sobre el chamanismo en las religiones de los indios de América del Norte:
- Documento sobre teología postmoderna:
- Documento sobre el culto a la luna:

Clase 3

- Documento sobre las escuelas privadas:
- Documento sobre los partidos conservadores franceses:
- Documento sobre la mujer en la política:
- Documento sobre los impuestos municipales:
- Documento sobre las funciones del Parlamento catalán:

Clase 5

Objetivos

Estos ejercicios os tienen que permitir alcanzar los objetivos siguientes:

- Construir notaciones combinando tablas principales con tablas de auxiliares comunes y auxiliares especiales.
- Construir notaciones utilizando los signos clasificatorios de la CDU, es decir, expresando la relación entre las materias unidas.

- Documento sobre los cráteres y mares de la Luna:

Clase 6

- Documento sobre la otitis:
- Documento sobre los libros de cocina:

Clase 7

- Documento sobre la pintura de Velázquez:
- Documento sobre los artistas alemanes postimpresionistas:
- Documento sobre las mujeres glamourosas en el cine de suspense:
- Documento sobre las competiciones de judo:
- Documento sobre el himno nacional de Estados Unidos de América del Norte:

Clase 8

- Documento sobre literatura hebrea:
- Documento sobre la poesía de Antonio Machado:
- Documento sobre las novelas de Harry Potter:

Clase 9

- Documento sobre la geografía de la India:
- Documento sobre la historia de la casa de los Borbones en España:
- Documento sobre geografía política:
- Documento sobre geografía de la Polinesia Francesa:

5.3. Ejercicios de nivel alto

Clasificad los documentos siguientes:

- Documento sobre Barcelona ciudad como fondo argumental de películas de cine:
- Documento sobre el psicoanálisis como terapia médica:
- Documento sobre el turismo como elemento del tiempo libre:
- Documento sobre el turismo como factor económico:
- Documento sobre los diarios franceses:
- Documento sobre los ritos y ceremonias budistas:
- Documento sobre las unidades calorimétricas como medida de cantidad de calor:
- Documento sobre la bibliografía de las publicaciones periódicas de zología:
- Documento sobre las catedrales góticas francesas:
- Documento sobre herramientas para hacer esculturas:
- Documento sobre los héroes en las películas de ciencia-ficción:

Objetivos

Estos ejercicios os tienen que permitir alcanzar los objetivos siguientes:

- Construir notaciones combinando todos los elementos antes descritos, sin mencionar expresamente las partes que no integran la notación ni la clase a la cual pertenecen.

- Documento sobre historia de España bajo los romanos:
- Documento sobre la Biblia por los testimonios de Jehová:
- Documento sobre la Gran Enciclopedia Catalana, en CD-ROM:
- Documento sobre los suelos de arenisca:
- Documento sobre órdenes religiosas femeninas católicas:
- Documento sobre el psicoanálisis como proceso psicológico:
- Documento sobre los cuentos fantásticos de Julio Cortázar:

6. Creación de un sistema de clasificación propio

Hay circunstancias en las cuales el documentalista se ve impulsado a crear su propio sistema de clasificación: colecciones muy especializadas, fondos documentales históricos... son casos en los que es difícil, si no imposible, encontrar un sistema de clasificación ya existente. Lo más común es que el documentalista pueda aprovechar partes, pero no la totalidad.

Del conocimiento que se ha ido desprendiendo de la historia de las clasificaciones y de la estructura y funcionamiento de modelos clasificaciones, como la CDU, podemos extraer las líneas básicas para crear un sistema de clasificación nuevo.

Premisas

Lo primero que el documentalista tendría que sopesar es si realmente un sistema de clasificación es el lenguaje documental apropiado para su fondo.

Recordemos que hay otros lenguajes: la lista de palabras clave (o indexación automática), la lista de descriptores libres, la lista de autoridades, las listas de encabezamientos de materia y los thesaurus, todos ellos estudiados en los módulos de esta asignatura.

A continuación, exponemos las fases de creación de un sistema de clasificación nuevo, según Jacques Maniez, convenientemente sintetizadas:

- Estudio previo: de usuarios, clasificaciones anteriores, estructura, lengua y otros
- Metodología
- Proceso de creación
- Mantenimiento

6.1. Estudio previo

La fase de estudio previo presenta las posibilidades siguientes, según la orientación que se quiera dar al lenguaje documental de creación nueva:

1) **De usuarios:** si los usuarios tienen libre acceso a las estanterías, hace falta una cuidada ordenación por materias; si los usuarios buscan la referencia en el catálogo en línea, la ordenación por materias no es tan definitoria.

Tipo de ordenaciones

Recordemos que podemos distinguir entre tres tipos de ordenaciones (Maniez, 1993, pág. 21):

- 1) Ordenaciones no significativas, sin ninguna relación con el contenido de los documentos, por ejemplo, ordenaciones numéricas a partir del número de registro.
- 2) Ordenaciones con significado limitado: por lengua, país, autor, disciplina, etc.

Lectura complementaria

El proceso de creación de un lenguaje documental que se seguirá en los subapartados siguientes sigue el punto de vista que se expone en la obra siguiente:

J. Maniez (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.

3) Ordenaciones altamente significativas: ordenación por materias, de manera que los documentos que traten de contenidos afines se coloquen unos cerca de otros.

2) **De clasificaciones anteriores:** hace falta optimizar recursos y consultar cuadros previos para inspirarse o adoptarlos directamente. Ejemplo: CDU, DDC, clasificaciones sectoriales, recopilaciones de thesaurus.

3) **De estructura:** podemos escoger entre tres tipos (jerárquica, facetada o mixta).

4) **De lengua y nivel de especificidad.**

5) **Otros: gastos, calendario, funciones.**

Ved también

Los thesaurus tienen diversas presentaciones, entre las cuales la jerárquica, que se corresponde a un cuadro de clasificación. Se estudiará más detenidamente en el módulo "Listas de encabezamientos de materia" de la asignatura.

6.2. Metodología

La **metodología de trabajo** consiste en una combinación de dos métodos, el inductivo y el deductivo. El primero se basa en la realización de un cuadro según las materias que tengamos en nuestro SID, el segundo se basa en el dominio conceptual y su subdivisión siguiendo las reglas de división lógica.

Metodologías inductivas y deductivas

En una metodología inductiva, reunimos todos los documentos sobre derecho de manera intuitiva, agrupándolos por afinidad de contenido. Después, relacionamos las diversas clases que hayan salido en una más general, y así sucesivamente.

En una metodología deductiva, estructuramos el dominio derecho en derecho civil, derecho constitucional, derecho penal, derecho internacional... (división tradicional académica).

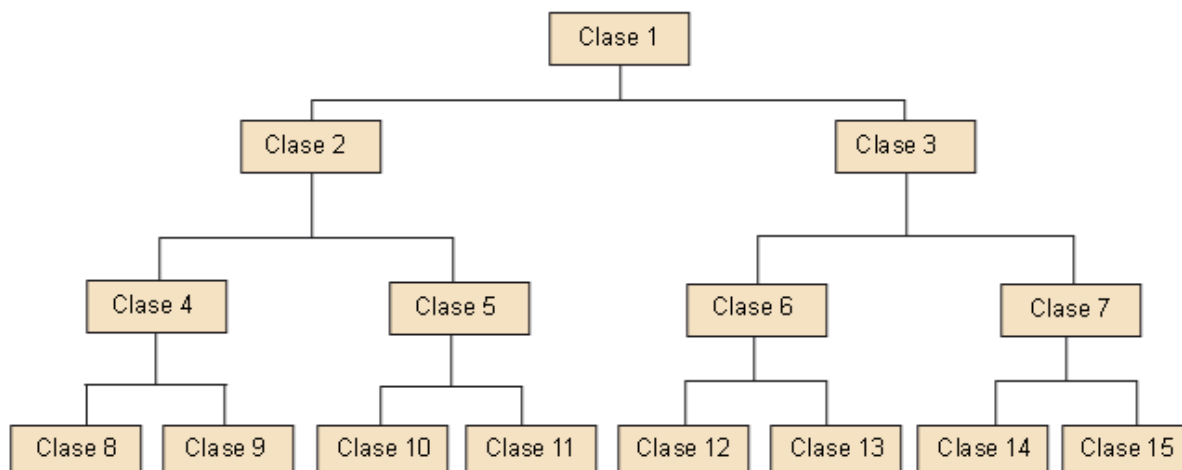
Con respecto a los **instrumentos de trabajo**, hay diversos tipos de programas de gestión documental, pero el documentalista también necesitará procesadores de textos y algún tipo de base de datos que le permita relacionar rúbricas con sinónimos y otras anotaciones de uso, como las propias del fichero de autoridades.

6.3. Proceso de creación

En el proceso de creación de un lenguaje documental se distinguen las etapas siguientes:

1) **Elección de la forma global de la arborescencia.** El *quid* reside en decidir cuántas clases pondremos en el primer nivel, qué intervalo de rango usaremos y hasta qué nivel consideramos que la notación será comprensible.

Tal como se puede ver en el esquema siguiente, si un cuadro de clasificación parte de una clase principal y rango dos, en el tercer nivel se consiguen 15 clases y la notación tendrá cuatro dígitos (por ejemplo, 111.1).



Una clase principal que se va abriendo en dos clases cada vez (rango 2) da 15 clases en el tercer nivel.

Decimos que hay 15 clases, y no sólo las 8 del nivel inferior, porque todas son clases susceptibles de ser utilizadas para clasificar. Por ejemplo, en la clase marcada como número 2, podemos clasificar obras sobre Religión católica; en la clase 4 clasificar obras sobre la Biblia; en la clase 8, obras sobre el Antiguo Testamento.

Como se puede comprobar, harían falta muchas subdivisiones para llegar, por ejemplo, a 200 clases. ¿Cómo se soluciona? Aumentando el número de clases principales y/o aumentando el rango hasta encontrar el equilibrio que se necesita.

Podemos conseguir un cuadro de 200 clases con diversas opciones:

- 5 clases principales, a rango 3, en el tercer nivel tendría 200 clases.
- 10 clases principales a rango 4, en el segundo nivel tendría 210 clases y la notación sería más corta.

La recomendación de J. Maniez (pág. 160) a la hora de ampliar el número de clases es la siguiente:

“Por nuestra parte, aconsejamos forzar el número de clases principales, sin dudar ante la necesidad de superar la decena (lo que no permite una notación decimal en el primer nivel) [...] Por consiguiente, es razonable una media de 4 clases por rango.”

J. Maniez, *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales* (pág. 160)

2) **Lista y ordenación de las clases principales.** Las clases principales tienen que ser disjuntas. Conviene reservar una clase para *Diversos*. Conviene reservar otra para generalidades. Y conviene crear una para auxiliares de tipo facetado (forma, lugar, tiempo). Recordemos que las clases se pueden ordenar de diversas maneras:

- a) de general a particular.
- b) de la disciplina inicial en la que depende.
- c) según el nivel de complejidad (de simple en complejo).
- d) por orden cronológico.
- e) por orden topográfico.

3) **Elección de la notación.** Se trata de traducir el lenguaje natural a un código artificial, que puede ser numérico, alfabético o alfanumérico. Si además hemos previsto facetas, hará falta prever una sintaxis (fijar el orden de las partes de la notación).

4) **Edición experimental.** Se pone en práctica una prueba experimental del nuevo lenguaje.

5) **Formación de los clasificadores y test de validez.** A continuación, se forman los clasificadores y se hace la prueba sobre un conjunto limitado de documentos. Después, se comentan las dificultades o facilidades del sistema y se toman las medidas correctivas necesarias.

6) **Revisión y primera edición.** Finalmente, se hace la revisión completa y se publica la primera edición de este nuevo lenguaje documental de creación nueva.

6.4. Mantenimiento

Hace falta un mantenimiento regular. Si el sistema está automatizado, podemos hacer el seguimiento de las notaciones y saber la frecuencia de utilización de las rúbricas y códigos, y así se pueden tomar decisiones (ampliar, modificar o suprimir clases).

A modo de conclusión

Las fases de creación de un cuadro de clasificación, según Jacques Maniez, son:

- **Estudio previo:** de usuarios, clasificaciones anteriores, estructura, lengua y otros.
- **Metodología:** combinación de dos métodos: el inductivo y el deductivo.
- **Proceso de creación:** elección de la forma global de la arborescencia y ordenación de las clases.
- **Mantenimiento.**

Actividad

Cread un cuadro de clasificación para vuestra biblioteca personal.

Solucionario

1.
 - Documento sobre Thesaurus: **025.43**
 - Documento sobre enciclopedias: **030**
 - Documento sobre anuarios: **050.8**
 - Documento sobre museos: **069**
2.
 - Documento sobre el neoplatonismo: **141.131**
 - Documento sobre las profecías: **159.961.3**
 - Documento sobre la naturaleza del conocimiento: **165.2**
3.
 - Documento sobre el judaísmo sefardí: **264**
 - Documento sobre el budismo: **24**
 - Documento sobre la ciencia: **299.93**
 - Documento sobre la religión de los sumerios: **252.1**
4.
 - Documento sobre la estadística como ciencia: **311**
 - Documento sobre las finanzas públicas: **336.1**
 - Documento sobre el derecho constitucional: **342**
 - Documento sobre las facultades universitarias: **378.096**
 - Documento sobre la etiqueta y el protocolo: **395.6**
5.
 - Documento sobre aritmética: **511.1**
 - Documento sobre la galaxia: **524.6**
 - Documento sobre climatología: **551.58**
 - Documento sobre ammonitas (fósiles cefalópodos): **564.53**
6.
 - Documento sobre otorrinolaringología: **616.21**
 - Documento sobre la forma de preparar comidas rápidas para llevar: **641.557**
 - Documento sobre editoriales (de libros): **655.41**
 - Documento sobre acabados especiales del cuero: **675.26**
7.
 - Documento sobre parques privados: **712.26**
 - Documento sobre catedrales: **726.6**
 - Documento sobre óptica de las cámaras fotográficas: **771.35**
 - Documento sobre música para orquesta sinfónica: **785.11**
 - Documento sobre carreras de karts: **796.714**
8.
 - Documento sobre cuentos de hadas: **82-343**
 - Documento sobre literatura hebrea: **821.411.16**
9.
 - Documento sobre la geografía de África: **913(6)**
 - Documento sobre datación con carbono: **902.65**
 - Documento sobre banderas nacionales: **929.921**
 - Documento sobre paleografía: **930.272**
 - Documento sobre historia de Australia: **94(94)**

Soluciones de los ejercicios de nivel bajo

Clase 0

- Documento sobre bibliografías de autor: **012**
- Documento sobre catálogos por materias: **017**
- Documento sobre las cartas al director de un diario: **070.45**
- Documento sobre ex-libris: **097**

Clase 1

- Documento sobre métodos de filosofar: **101.8**
- Documento sobre estados emocionales: **159.942.5**
- Documento sobre lógica proposicional: **164.1**
- Documento sobre ética profesional: **174**

Clase 2

- Documento sobre la Iglesia católica romana: **272**
- Documento sobre el Islam chiíta: **284**
- Documento sobre el agnosticismo: **299.3**
- Documento sobre el confucianismo (religión china): **221.7**

Clase 3

- Documento sobre teorías de la población: **314.12**
- Documento sobre campañas electorales: **324**
- Documento sobre el Ministerio de Hacienda: **354.21**
- Documento sobre el poder adquisitivo del consumidor: **366.2**

Clase 4**Clase 5**

- Documento sobre ciencias del medio ambiente: 502
- Documento sobre la capacidad calorífica de los líquidos: 536.632
- Documento sobre hidrocarburos saturados: 547.21
- Documento sobre silicatos básicos: 549.614
- Documento sobre biología: 57
- Documento sobre botánica - las magnolias: 582.677.1
- Documento sobre la ballena azul: 599.511

Clase 6

- Documento sobre psicoterapia: 615.851
- Documento sobre drenaje de minas: 622.5
- Documento sobre enfermedades por hongos de las plantas: 632.4
- Documento sobre instalaciones de agua potable (en las viviendas): 696.11
- Documento sobre producción de sosa y potasa: 661.33

Clase 7

- Documento sobre escultura: 730
- Documento sobre dibujo anatómico: 743
- Documento sobre diseño de carteles: 766

Clase 8

- Documento sobre la lengua coreana: 811.531
- Documento sobre literatura griega: 821.14

Clase 9

- Documento sobre pinturas rupestres: 903.27
- Documento sobre viajes de descubrimientos geográficos: 910.4
- Documento sobre órdenes de caballería: 929.71
- Documento sobre la tercera guerra carlista: 94(460).073.1

Diversos enfoques de las materias:

- Documento sobre el ateísmo, desde el punto de vista filosófico: 141.45
- Documento sobre el ateísmo, desde el punto de vista religioso: 299.2
- Documento sobre las elecciones políticas, desde el punto de vista político: 324
- Documento sobre las elecciones políticas, desde el punto de vista del derecho: 342.8
- Documento sobre los delitos contra los derechos humanos, desde el punto de vista del derecho penal: 343.4
- Documento sobre los delitos contra los derechos humanos, desde el punto de vista del derecho internacional: 341.231.14
- Documento sobre los delitos contra los derechos humanos, desde el punto de vista del derecho constitucional: 342.72/.73

Soluciones de los ejercicios de nivel medio**Clase 0**

- Documento sobre las piezas prehistóricas de un museo: 069.51:903
- Documento sobre la clasificación decimal Dewey: 025.45CD
- Documento sobre bibliografía nacional francesa: 015(44)
- Documento sobre thesaurus de psicología: 025.43:159.9

Clase 1

- Documento sobre la filosofía de Arthur Schopenhauer 1(091) SCHOPENHAUER o 1(430)18"SCHOPENHAUER
- Documento sobre el esnobismo inglés: 177.5(410)

Clase 2

- Documento sobre el santos de la religión cristiana: 27-36
- Documento sobre el chamanismo en las religiones de los indios de América del Norte: 259.2-38
- Documento sobre teología postmoderna: 2-11:141.78
- Documento sobre el culto a la Luna: 2-137:523.3

Clase 3

- Documento sobre las escuelas privadas: 37.058
- Documento sobre los partidos conservadores franceses: 329.11(44)
- Documento sobre la mujer en la política: 32-055.2
- Documento sobre los impuestos municipales: 336.22:352
- Documento sobre las funciones del Parlamento catalán: 328.3(460.23)

Clase 5

- Documento sobre los cráteres y mares de la Luna: 523.3-8

Clase 6

- Documento sobre la otitis: 616.284-002
- Documento sobre los libros de cocina: 641.55(083.12)

Clase 7

- Documento sobre la pintura de Velázquez: 75Velázquez
- Documento sobre los artistas alemanes postimpresionistas: 7.036.4(430)
- Documento sobre las mujeres glamorosas en el cine de suspense: 791.221.5-56
- Documento sobre las competiciones de judo: 796.853.23.09

- Documento sobre el himno nacional de Estados Unidos de América del Norte: **784.71(73)**

Clase 8

- Documento sobre la literatura hebrea: **821.411.16**
- Documento sobre la poesía de Antonio Machado: **821.134.2-1MACHADO**
- Documento sobre las novelas de Harry Potter: **821.111-312.9ROWLING**

Clase 9

- Documento sobre la geografía de la India: **913(540)**
- Documento sobre la historia de la casa de los Borbones en España: **94(460).05**
- Documento sobre geografía política: **911.3:32**
- Documento sobre geografía de la Polinesia Francesa: **913(963)**

Soluciones de los ejercicios de nivel alto

- Documento sobre Barcelona ciudad como fondo argumental de películas de cine: **791.237.6(460.235-25)**
- Documento sobre el psicoanálisis como terapia médica: **615.851**
- Documento sobre el turismo como elemento del tiempo libre: **379.85**
- Documento sobre el turismo como factor económico: **338.48**
- Documento sobre los diarios franceses: **070(44)**
- Documento sobre los ritos y ceremonias budistas: **24-5**
- Documento sobre las unidades calorimétricas como medida de cantidad de calor: **536.6.081**
- Documento sobre la bibliografía de las publicaciones periódicas de zoología: **016:59(05)**
- Documento sobre las catedrales góticas francesas: **726.6.033.5(44)**
- Documento sobre herramientas para hacer esculturas: **730.022**
- Documento sobre los héroes en las películas de ciencia-ficción: **791.221.8-51**
- Documento sobre historia de España bajo los romanos: **94(460).013**
- Documento sobre la Biblia por los testimonios de Jehová: **279.17-23**
- Documento sobre la *Gran Enciclopèdia Catalana*, en CD-ROM: **(031.034)GEC**
- Documento sobre los suelos de arenisca: **645.1-033.6**
- Documento sobre órdenes religiosas femeninas católicas: **27-788-055.2**
- Documento sobre el psicoanálisis como proceso psicológico: **159.964.2**
- Documento sobre cuentos fantásticos de Julio Cortázar: **821.134.2-32(82)**

Glosario

Los términos del glosario de este módulo están en orden alfabético. Remiten unos a otros, y muestran las relaciones semánticas, como un lenguaje controlado.

- Si hay dos conceptos sinónimos, los términos no aceptados se remiten al aceptado con un ved. Ejemplo: código numérico, ved notación.
Significa que el término por lo común aceptado es notación. En el glosario, se mantienen las dos entradas porque, a veces, hay poco consenso terminológico en la bibliografía.
- Si el término tiene relación conceptual (genérica, específica o de asociación) con otros términos de este glosario, se indica con un ved también. Ejemplo: Signatura topográfica, ved también ordenación.
Significa que el término *Signatura topográfica* está relacionado con el de ordenación.

Los ejemplos de este glosario hacen referencia a la clase 52 de astronomía de la CDU, edición de bolsillo 2004; hay que recordar que otras ediciones, como la abreviada o el fichero básico de referencia, tienen más clases.

			Tabla principal / Clase principal
		5 Matemáticas. Ciencias naturales	
Clase principal		52 Astronomía. Astrofísica. Investigación espacial. Geodesia	
Auxiliares especiales	52-1	Investigación inicial	
	52-3	Propiedades y fenómenos (especialmente geométricos)	
	52-3	Procesos relativos a cuerpos y sistemas	
	52-6	Procesos de radiación	
	52-7	Carácter de la radiación	
	52-8	Partes y características de sistemas concretos	
		520 Instrumentos y técnicas astronómicas	
	520.1	Observatorios	
		7272.912	
Índice / Número / Código		521 Astronomía teórica	
		523 Sistema solar	
	523.2	Propiedades generales	
	523.3	Lunas	
	523.4.	Planetas	
	523.41	Mercurio	
	523.42	Venus	
Rúbrica	523.43	Marte	
	523.44	Planetas menores	
	523.45	Júpiter	
	523.46	Saturno	
	523.47	Urano	
	523.481	Neptuno	
	523.482	Plutón	
	523.483	Planetas hipotéticos	
			Clases

auxiliar común dependiente (tabla) m Las tablas son dos: notaciones ajenas a la CDU y características generales (de propiedades, materiales y personas). Como su nombre indica, son auxiliares que sólo se pueden usar como sufijos; no se han de utilizar independientemente ni en primer lugar en la notación.

auxiliar común independiente (tabla) m Las tablas son: lengua, forma, lugar, razas, pueblos y nacionalidades, tiempo. Ya que tienen un símbolo de inicio y uno final pueden ir en cualquier lugar de la notación, a excepción del de la lengua, que sólo tiene símbolo en el inicio, aunque la CDU permite añadirle detrás dos puntos.

Si en un mismo documento hay que representar más de un concepto con auxiliares independientes, la CDU recomienda el orden siguiente: Clase + lugar + tiempo + forma + lengua + raza.

auxiliar especial (tabla) m Los auxiliares especiales ayudan a precisar el contenido de algunas materias, no de todas, ya sea de alguna de las nueve clases principales, como de algunos auxiliares comunes.

Ya que hay auxiliares especiales repartidos por todas las clases principales, no los encontramos reunidos en una sola tabla como pasa con los auxiliares comunes. En el ejemplo de astronomía, vemos que las clases marcadas con la línea vertical negra son las auxiliares especiales previstas para la clase 52 (y sólo la clase 52).

cadena f Clases subordinadas sucesivamente (en contraposición a las clases coordinadas que se llaman series). En cada nivel, se escoge una característica para ir subdividiendo. En el ejemplo: de todas las ciencias naturales (física, química, botánica...) escogemos astronomía. De todas las materias estudiadas en astronomía (teoría astronómica, estrellas, universo), escogemos el sistema solar. De todos los elementos que forman el sistema solar (propiedades y leyes, lunas, meteoritos...), escogemos los planetas. De todos los planetas, escogemos Mercurio. (*Ved también:* jerarquía.). Por ejemplo:

5 Ciencias Naturales

52 Astronomía. Astrofísica. Investigación espacial. Geodesia

523 Sistema Solar

523. 4 Planetas

524.41 Mercurio

clase f Concepto, materia o disciplina de una clasificación. Las clases se basan en características, atributos, propiedades o cualidades comunes. Encontramos clases en todos los niveles de la cadena jerárquica. Las clases de primer nivel son las clases principales. Cada clase se divide en subclases, y así sucesivamente.

Ejemplo: desde la clase 5 hasta la última clase, todas son clases. Las dos primeras, la 5 y la 52, se consideran clases principales; el resto, clases.

Ved también clases principales.

clase principal f Las clases principales son las que ocupan el primer nivel en la cadena jerárquica. Se equiparan a las disciplinas. Por ejemplo: 52 Astronomía.

Ved también disciplinas.

clasificación bibliográfica f Ved clasificación documental.

clasificación documental f Lenguaje documental que divide el conocimiento enciclopédico, o el de un área, en una serie ordenada de clases y subclases basadas en las similitudes y diferencias. Se utiliza para clasificar la materia de los documentos y ordenar físicamente los documentos originales. Son lenguajes sintéticos, precoordinados, codificados, controlados y de estructura jerárquica o híbrida con posibilidad de expresar facetas. Se llama documental para diferenciarlo de las clasificaciones filosóficas, científicas y de las clasificaciones o taxonomías biológicas. Por ejemplo: la clasificación Dewey, la clasificación decimal universal, la clasificación Colon, La clasificación de la Library of Congress. *Ved también:* **clasificación filosófica y científica; taxonomía.**

clasificación enumerativa f Ved clasificación monojerárquica.

clasificación filosófica y científica f Sistema filosófico y de clasificación del conocimiento. A diferencia de las clasificaciones documentales, se interesa por todo el conocimiento humano, y no sólo por el conocimiento depositado en la colección del SID. Por ejemplo: la clasificación filosófica que más trascendencia ha tenido, desde la Edad Moderna, es la de Francis Bacon.

clasificación jerárquica f Estructura que enumera las clases de arriba abajo y de izquierda a derecha. Cada tema se encuentra en un solo lugar de la cadena. Son las más numerosas por su facilidad de construcción y uso. Se adecuan muy bien en entornos enciclopédicos. Ejemplos: la clasificación de la Biblioteca del Congreso (LC).

clasificación por facetas f Estructura que agrupa los conceptos según las características y los atributos. La materia no se encuentra representada en el cuadro de clasificación, sino que el

analista la ha de crear combinando las facetas. La clasificación parte de dominios principales (en el estilo de las disciplinas) que no se estructuran jerárquicamente, sino de forma facetada. Dentro de cada faceta, la clasificación vuelve a ser jerárquica. Por ejemplo: la clasificación Colon, la clasificación Bliss.

clasificar *v tr* Proceso de división de objetos o conceptos en clases que, a la vez, se dividen en subclases. Cada división se basa en las características comunes que permiten agrupar los conceptos y en los elementos que las distinguen, de manera que se creen clases disjuntas. Como ejemplo de característica común y diferencial: la clase 52 agrupa los conceptos que tienen en común el tema astronomía y se subdivide, por ejemplo, en las clases 523.3 lunas y 523.4 planetas para crear estas dos clases disjuntas, de manera que no haya duda a la hora de clasificar un documento (o es una luna o es un planeta).

código *m* Ved notación.

Consortio de la CDU (UDC Consortium) *m* Organismo responsable de la gestión de la CDU, con sede en la Biblioteca Real de La Haya. Se creó en 1992, absorbiendo las responsabilidades que hasta entonces tenía la FID en materia de CDU. En el Consorcio, participan diversas agencias de normalización y editores como la española AENOR, la rusa VINITI o la inglesa BSI. El Consorcio es el responsable de los cambios y cada agencia los publica en el idioma propio.

decimal (número) *m* Las diez clases principales se pueden subdividir en diez clases más y así sucesivamente. Los números decimales de la CDU tienen el valor de las fracciones decimales situadas detrás de un 0, (cero coma) que se presupone. Este sistema permite subdividir las clases de manera casi indefinida. Por ejemplo: la clase 3 y la clase 33, serían la 0,3 y la 0,33.

dicotomía *f* Método de clasificación en el que las divisiones y las subdivisiones son binarias.

disciplina *f* ámbitos temáticos. Son campos de estudio que tienen existencia independiente, identificable y coherente. Ejemplos: filosofía, matemática, ciencias sociales, historia, religión... En la CDU, se corresponderían las clases principales. En las clases principales que agrupan más de una materia, como la 3,5, 6 y la 7, también llamamos disciplinas a las clases de 2 dígitos, como 52 astronomía (*ved también* clases principales.). Por ejemplo:

5 Ciencias Naturales
52 Astronomía

ediciones de la CDU *f pl* La CDU española tiene diversas ediciones: entera, abreviada, bolsillo, en formato papel o digital. Desde la creación del Consorcio de la CDU, UDC Consortium, se impulsa el Master Reference File, o fichero básico de referencia, de 62.000 entradas como única edición autorizada a partir de la cual se pueden hacer diversas ediciones.

extensiones y correcciones de la CDU *f pl* Publicación anual del Consorcio de la CDU en que se publican las nuevas tablas revisadas y aprobadas. Es la base sobre la cual el fichero básico de referencia es revisado cada año (las extensiones y correcciones se publican cada mes de noviembre y el MRF sale el uno de enero). (*Ved también* fichero básico de referencia y Consorcio de la CDU.)

en extensions and corrections UDC

faceta *f* Características o atributos de los objetos y conceptos que permite agruparlos. Hay facetas universales y facetas particulares. Por ejemplo:

- Facetas universales: la fórmula PMEST de la clasificación Colon: personalidad, materia, energía, espacio, tiempo.
- Facetas particulares: en el ejemplo de astronomía, son los auxiliares especiales: investigación, propiedades, procesos, partes...

fichero básico de referencia (master reference file o MRF) *m* Base de datos que contiene la versión oficial de la CDU; tiene unas 62.000 entradas. La mantiene el Consorcio de la CDU con software de la UNESCO CD-ISIS.

híbrido -a *adj* Dicho del tipo de estructura de clasificación documental que combina la estructura monojerárquica con la posibilidad de expresar facetas. Por ejemplo: la clasificación decimal universal (CDU). En este ejemplo, son elementos facetados los auxiliares especiales marcados con la línea vertical.

índice *m* Los códigos que preceden la rúbrica en las tablas. Una vez introducida en el catálogo, se llama notación. Por ejemplo: 52
Ved código y notación.

jerarquía *f* La subordinación de las clases, dividiéndolas en sus componentes lógicos. Se aplican sucesivamente principios de división genéricos y principios de todo-parte.
Ved también cadena.

mixto -a *adj* *Ved* híbrido -a.

notación *f* Es el término de indexación propio de los sistemas de clasificación. Es un código artificial, no un lenguaje natural, hecho con números, letras y símbolos que acompañan las rúbricas. Representa la materia del documento; también se usa para la ordenación física de los documentos.

La longitud de la notación puede representar el grado de especificidad de la materia de un documento (cuanto más larga, más específica), pero también puede estar traduciendo una arborescencia a partir de pocas clases principales. Maniez calcula que una notación de hasta 7 dígitos es comprensible, más no. Por ejemplo: 524.42 es una notación.

Ved también signatura topográfica y rúbrica.

ordenación *f* En documentación, es la operación de colocar los libros de forma ordenada en los estantes, siguiendo la signatura topográfica. Según Maniez, hay tres tipos de ordenaciones: 1) Ordenaciones no significativas, sin ninguna relación con el contenido de los documentos, por ejemplo, ordenaciones numéricas a partir del número de registro. 2) Ordenaciones con significado limitado: por lengua, país, autor, disciplina, etc. 3) Ordenaciones altamente significativas: ordenación por materias, de manera que los documentos que traten de contenidos afines se coloquen unos cerca de otros.

orden de las partes de una notación *m* En primer lugar, la clase de la materia principal del documento + auxiliares especiales (si los hay) + auxiliares comunes. A continuación, el orden dentro de los auxiliares comunes (también conocido como orden de secuencia) no está fijo, pero la CDU recomienda ir de general a particular, y propone el orden siguiente: lugar + tiempo + forma + idioma.

rúbrica *f* Nombre de la materia expresado en lenguaje natural. Por ejemplo: Venus.

serie *f* Serie de clases coordinadas (en contraposición a clases subordinadas, que serían cadena o jerarquía). Es un grupo de clases mutuamente excluyentes, derivadas de la aplicación de una característica.

Ved también faceta.

signatura topográfica *f* Conjunto de números, letras y símbolos que localizan y ordenan, físicamente, el documento en la estantería. A diferencia de la notación, que indica la materia y puede estar repetida (2 o más obras de la misma temática), la signatura topográfica indica su ubicación física y tiene que ser única.

signos combinatorios o clasificatorios *m pl* Son : :: + / []. Estos signos tienen una función sintética (de unir), ya que permiten expresar conceptos compuestos que no se encuentran previstos en el cuadro de clasificación. Por ejemplo: documento sobre el sistema solar (desde las propiedades, a las lunas y planetas) 523.2/.4

signos instructivos o no clasificatorios *m pl* Son . (punto)≅(dividir como) y → (ved también). No se usan para clasificar, sino para instruir sobre los usos del cuadro de clasificación. Por ejemplo:

520.1 Observatorios

→ 727.912

Significa que, si estamos consultando la clase 520.1 *Observatorios* nos puede interesar también consultar o clasificar el documento bajo la clase 727.912 que es *Arquitectura > Edificios con finalidades educativas y científicas > Observatorios*. Daos cuenta de la diferencia entre clasificar el documento bajo la clase 52 astronomía o bajo la clase 727 arquitectura.

tabla principal *f* *Ved* clase principal.

taxonomía *f* Sistemas que clasifican organismos vivos o inertes en una jerarquía. Actualmente, el término taxonomía se aplica también en el mundo de la información y las fuentes disponibles en la WWW.

Bibliografía

Maniez, J. (1992). *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide / Fundación Germán Sánchez Ruipérez.

McIlwaine, I. C. (2003). *Clasificación Decimal Universal. Guía para uso de la CDU*. Madrid: AENOR.

Reitz, Joan. *Odlis Online dictionary of library and information science* [en línea]. Western Connecticut State University. [Fecha de consulta: 10 de octubre de 2008].

Referencias bibliográficas

025.431: *The Dewey blog* [en línea]. [Fecha de consulta: 10 de octubre de 2008].

Álvarez García, F. J. (1997). "Introducción al uso práctico de la CDU". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 219-243). Madrid: Síntesis ("Biblioteconomía y Documentación").

AENOR. *Asociación española de normalización y certificación* [en línea]. [Fecha de consulta: 10 de octubre de 2008].

Arnau Riverd, Pilar (2005). *Documentación: hitos históricos, precedentes - Dewey - Otlet - FID* (pág. 96-105 y 248-255). Madrid: Mundarnau.

Benito, M. (1999). *El sistema de clasificación decimal universal: manual de aprendizaje*. Madrid: Taranco.

Benito, Miguel. *El sistema de Clasificación Decimal Universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008]. O, también, en la URL Instituto de Taranco. [Fecha de consulta: 10 de octubre de 2008].

Burguillos Martínez, Ferran. *Tutorial de la Clasificación decimal universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008].

CDU: Clasificación decimal universal: norma UNE 50 001:2000 (2000) (3 vols.). Madrid: AENOR.

Classificació decimal: adaptació per a les biblioteques catalanes (1982) (4.^a ed. preparada por Jordi Rubió i Lois). Barcelona: Teide.

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000 (2001) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU): edición abreviada de la norma UNE 50001: 2000; (incluye las modificaciones de la Norma UNE 50001:2004/1M. Trad. del Master Reference File realizada por el Centro de Información y Documentación Científica (CINDOC) (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Clasificación Decimal Universal (CDU) de bolsillo (2004) (adaptada por Rosa San Segundo Manuel). Madrid: AENOR.

Díez Carrera, C. (1999). *Técnicas y régimen de uso de la CDU (Clasificación Decimal Universal)* (pág. 134). Gijón: Trea. ("Biblioteconomía y Administración Cultural", 26).

Esteban, M. A. (1997). "Fundamentos epistemológicos de la clasificación documental". En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 19-32). Madrid: Síntesis ("Biblioteconomía y Documentación").

Gil Urdiciain, B. "Los lenguajes de clasificación documental". En: M. Pinto (ed.) (1997). *Manual de clasificación documental* (pág. 91-108). Madrid: Síntesis ("Biblioteconomía y Documentación").

Gili Martínez, M. D.; Vives i Gràcia, J. (1996, julio-diciembre). "La nova edició espanyola de la CDU: arriscar-se al canvi?». Ítem: *Revista de Biblioteconomia i Documentació* (núm. 19, pág. 94-121).

Glazier, Jack (2003). "Las raíces culturales de la clasificación moderna". En: *IV Coloquio Internacional de Ciencias de la Documentación*. Salamanca.

Glazier, Jack (2003). *VI Congreso del Capítulo Español de ISKO*. Salamanca.

Gran enciclopèdia catalana [en línea]. [Fecha de consulta: 10 de octubre de 2008].

Granados, M.; Nicolau, Anna. “La recuperació de la informació als catàlegs en línia: l'ús de la Classificació decimal universal i la seva implicació en la indexació” [en línea]. 7è Congrés del Capítol Espanyol d'ISKO. Barcelona, 6-8 de julio de 2005. Barcelona: Universitat de Barcelona, Departament de Biblioteconomia i Documentació [Consulta: 1 de octubre de 2008].

Grolier, E. (1976, noviembre-diciembre). “La clasificación cien años después de Dewey”. *Boletín Unesco* (vol. XXX, núm. 6, pág. 342-350).

Guardans, T. (1989, enero-junio). “Les altres religions a la Classificació decimal universal”. *Revista Ítem* (núm. 4).

La Clasificación Decimal Universal en los currículos escolares: instrucciones para la organización temática de los fondos bibliográficos de las bibliotecas educativas no universitarias [en línea]. Pamplona: Gobierno de Navarra, Departamento de Educación y Cultura. [Consulta: 1 de octubre de 2008].

Library of Congress Classification outline [en línea]. [Fecha de consulta: 1 de octubre de 2008].

López-Huertas, M. J. (1997). “Estructura de la clasificación decimal universal”. En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 177-217). Madrid: Síntesis (“Biblioteconomía y Documentación”).

López Yepes, José (1995). *La documentación como disciplina: teoría e historia* (2.ª ed., pág. 90-106). Pamplona: Eunsa.

McIlwaine, I. C. (1998, 16-21 de agosto). “The UDC at the present time” [en línea]. *64th IFLA General Conference. Amsterdam*. [Consulta: 1 de octubre de 2008: 1 de octubre de 2004].

Mestre i Ribera, Enric. *Classificació decimal universal* [en línea]. [Fecha de consulta: 10 de octubre de 2008].

Moreno Fernández, L. M.; Borgoños Martínez, M. D. (2002). *Teoría y práctica de la Clasificación Decimal Universal* (2.ª ed. revisada, corregida y aumentada). Gijón: Trea.

Online Computer Library Center. *Dewey services, Dewey decimal classification for use with OCLC's online cataloging services* [en línea]. [Consulta: 10 de octubre de 2008].

Otlet, P. (1996). *El tratado de documentación: el libro sobre el libro, teoría y práctica* (trad. de M. Dolores Ayuso). Murcia: Universidad de Murcia.

Perpinyà i Morera, R. (1995). “Los lenguajes documentales”. En: M. E. Fuentes i Pujol (ed.). *Manual de documentación periodística* (pág. 111-132). Madrid: Síntesis.

Pinto Molina, M. (ed.) (1997). *Manual de clasificación documental*. Madrid: Síntesis (“Biblioteconomía y Documentación”).

Public libraries using Spanish. *Spanish version of the Dewey Decimal Classification System based on the 18th edition of DDC* [en línea]. [Consulta: 10 de octubre de 2008].

Riesthuis, Gerhard J. A. (1998, agosto) “The UDC Master Reference File” [en línea]. *64th IFLA General Conference. Amsterdam*. [Consulta: 1 de octubre de 2008].

Rosati, Luca (2005). “Clasificación facetada para los servicios comunitarios mediante el uso de las categorías estándares del Classification Research Group”. *VII Congreso del Capítulo Español de ISKO: La dimensión del conocimiento*. Barcelona.

Rovira, T. (1989, enero-junio). “Jordi Rubió i Lois: la seva influencia en el món bibliotecari de Catalunya”. En: *Revista Ítem* (núm. 4).

Sagredo, Félix (2001). “Tras las huellas euro-norteamericanas del movimiento documental contemporáneo I” [en línea]. Madrid: *Revista Documentación de las Ciencias de la Información* (núm. 24, pág. 35-43). [Fecha de consulta: 1 de octubre de 2008].

San Segundo Manuel, R. (1997). “Principales sistemas de clasificación”. En: M. Pinto (ed.). *Manual de clasificación documental* (pág. 67-89). Madrid: Síntesis (“Biblioteconomía y Documentación”).

San Segundo Manuel, R. (2001). *Teoría e historia de la clasificación bibliotecaria en España, siglos XIX y XX*. Madrid: Universidad Complutense.

San Segundo Manuel, R. (1999). "Futura implantación de los distintos sistemas de clasificación en las bibliotecas del estado español" [en línea]. *7es Jornades Catalanes de Documentació*. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya. [Consulta: 1 de octubre de 2008].

San Segundo, R. *Use of UDC classification system in Spain: implementation, application, teaching and research* [en línea]. (Presentación) Madrid: Universidad Carlos III. [Consulta: 1 de octubre de 2008].

Soria, C. (2000, noviembre). *Cambios introducidos en la última edición (nov. 2000) de la CLASIFICACIÓN DECIMAL UNIVERSAL* [en línea]. [Consulta: 1 de octubre de 2008].

Tabla de Materias CDU [en línea]. Agencia Española del ISBN. [Consulta: 1 de octubre de 2008].

Turiel, J. (2001, junio). *Recursos web per a la classificació* [en línea]. *BiD: textos universitaris de biblioteconomia i documentació* (núm. 6). [Consulta: 1 de octubre de 2008].

Universal Decimal Classification Blog [en línea]. UDC Consortium. [Consulta: 10 de octubre de 2008].

Universal Decimal Classification Consortium Homepage (2002, 1 de agosto). [En línea]. La Haya: UDC Consortium. Act. 2002-08-01. [Consulta: 10 de octubre de 2008].

Listas de encabezamientos de materia y listas de autoridades

Manela Juncà Campdepadrós

PID_00143965



Universitat Oberta
de Catalunya

www.uoc.edu



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. Fundació 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

Introducción	5
Objetivos	6
1. Las autoridades en el análisis de contenido	7
2. ¿Qué es una lista de encabezamientos de materia?	9
3. Origen y evolución de las listas de encabezamientos	12
4. Tipologías de las listas de encabezamientos	16
5. Principios de las listas de encabezamientos	19
6. Encabezamientos y subencabezamientos	23
6.1. Encabezamientos	23
6.2. Subencabezamientos	23
7. Reglas de vocabulario y signos de puntuación de las LEM	28
8. Términos simples y compuestos	30
8.1. Encabezamientos simples	30
8.2. Encabezamientos compuestos	30
8.3. Orden de los encabezamientos	32
9. Las relaciones semánticas	34
9.1. Relación de equivalencia	34
9.2. Relación de jerarquía	36
9.3. Relación asociativa	37
9.4. Símbolos utilizados para designar los tres tipos de relaciones	38
10. Aprendiendo a indizar con una lista de encabezamientos	40
11. La lista de autoridades	44
11.1. Las listas de autoridades en Cataluña y España	44
11.2. Tipología	47
11.3. Indizar con una lista de autoridades	49
Actividades	51
Glosario	52

Bibliografía..... 56

Introducción

Este módulo os introduce en el uso de las autoridades en la indización por materias. Comprende dos lenguajes: las listas de encabezamientos de materia y los listados de autoridades.

Itinerario de estudio

El módulo empieza con la definición de las autoridades y su uso en el análisis de contenido. A continuación se describen las listas de encabezamientos de materia y los elementos comunes a cualquier lista de encabezamientos, sea cual sea el idioma. El último apartado se dedica al resto de autoridades.

Conceptos más importantes

Concepto	Ved
Autoridad	1. Las autoridades en el análisis de contenido
Encabezamiento y subencabezamiento de materia	5. Principios de las listas de encabezamientos 6. Encabezamientos y subencabezamientos 7. Reglas de vocabulario y signos de puntuación en las LEM 8. Términos simples y compuestos
Relaciones semánticas: de equivalencia, jerarquía y asociación	9. Las relaciones semánticas
Principio de especificidad Principio de economía	10. Aprendiendo a indizar con una lista de encabezamientos
Nombre, título, título uniforme, nombre geográfico	11. Listas de autoridades

Objetivos

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

En cuanto a las **listas de encabezamientos de materia**:

1. Saber definir las listas de encabezamientos de materia y su tipología.
2. Ser capaz de presentar la evolución histórica de este lenguaje documental.
3. Conocer las principales listas de encabezamientos.
4. Conocer los elementos, relaciones y sintaxis de las listas de encabezamientos.
5. Adquirir una cierta habilidad en la indización de documentos con las listas LEMAC (Lista de encabezamientos de materia en catalán) y AM BNE (Autoridades de materia del catálogo de autoridades de la Biblioteca Nacional de España) y AM CSIC (Autoridades de materia de la red de bibliotecas del CSIC).

En cuanto a los **listados de autoridades**:

1. Conocer el uso de las autoridades en el análisis de contenido.
2. Aprender a indizar materias que incluyan algún tipo de autoridad.

1. Las autoridades en el análisis de contenido

Este módulo está dedicado a los dos lenguajes documentales que tratan las autoridades: los listados de encabezamientos de materia y los listados de autoridades.

Las **autoridades** son las formas normalizadas de las entradas de un catálogo bibliográfico o inventario de archivo.

Concretamente estamos hablando de las formas normalizadas para:

- Nombres propios
- Títulos
- Nombres propios y títulos
- Materias
- Entidades
- Nombres geográficos

Para evitar problemas en la recuperación (de ruido y silencio documental) debidos a la ambigüedad del lenguaje natural hay que establecer formas normalizadas en cada caso de acuerdo con las reglas internacionales de descripción: reglas norteamericanas de catalogación (para las bibliotecas) y de las Normas Internacionales sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias (ISAAR[CPF]¹) (para archivos).

⁽¹⁾ISAAR(CPF) es la sigla de *International Standard Archival Authority Records for Corporate Bodies, Persons and Families*, Norma Internacional sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias.

Estas formas normalizadas se registran en el fichero de autoridades, que anota la forma aceptada, los términos no seleccionados, notas de aplicación, relaciones semánticas si es oportuno y la fuente en la que se ha basado. Este fichero de autoridades es el listado que el analista tiene que consultar antes de indizar para comprobar la forma aceptada.

Para analizar el contenido de un documento, el analista tiene que saber consultar las listas de encabezamientos de materia, pero también tiene que saber consultar el resto de autoridades, ya que la materia de un documento también puede ser:

- Un nombre propio: documento sobre la vida de Doris Lessing.
- Un nombre de institución presente o histórica: documento sobre el consejo de guerra.
- Un título uniforme: interpretaciones de la obra *El retrato de Dorian Gray*.
- Un lugar geográfico: documento sobre la India.

- Un nombre de empresa, entidad, etc.: documento sobre la cooperativa Eroski.

Estos términos (Doris Lessing, consejo de guerra, retrato de Dorian Gray, India, Eroski) no aparecen en las listas de encabezamientos de materia porque tienen entrada como nombres personales, de título, geográficos, etc. En consecuencia, el analista tiene que saber consultar todo tipo de autoridades.

Entre todas las autoridades, las de materia destacan por su singularidad, por este motivo este módulo se divide en dos grandes bloques:

- Las autoridades por materia: listados de encabezamientos de materia.
- El resto de autoridades: listado de autoridades.

Ejemplo de nombre propio que actúa también como materia

Este ejemplo muestra cómo la misma autoridad introducida como nombre propio **Dalí, Salvador, 1904-1989** puede ser Autor o Materia en cada caso:

- Documento a): Salvador Dalí autor de la obra *El mito trágico del Angelus de Millet*.
- Documento b): Ignasi Puig escribe una obra titulada *El Dalí esotérico*.
- Documento c): Salvador Dalí escribe *Vida secreta* (es su autobiografía).

	Autor	Materia
a)	Dalí, Salvador, 1904-1989	Millet, Jean-François. Ángelus
b)	Puig Alemán, Ignasi	Dalí, Salvador, 1904-1989
c)	Dalí, Salvador, 1904-1989	Dalí, Salvador, 1904-1989

Soluciones extraídas del catálogo de la Biblioteca de Catalunya

Como se puede observar, autor y materia coinciden en el caso de las autobiografías.

Títulos uniformes

Un título uniforme es un título ficticio dado por el analista para agrupar todas las entradas en un catálogo de una obra. El título uniforme se crea para controlar las diferentes ediciones, traducciones, etc. Un ejemplo extraído del listado LENOTI: el *Pájaro de fuego* de Igor Stravinski tiene como título uniforme "Stravinski, Igor. *Jar ptitsa*", título que engloba cualquier otra traducción tipo *El pájaro de fuego*, *L'oiseau de feu*, etc.

2. ¿Qué es una lista de encabezamientos de materia?

Una **lista de encabezamientos de materia**² es un lenguaje documental controlado que recoge ordenadas alfabéticamente todas las materias o temas susceptibles de representar el contenido de un fondo documental. Es una lista, un léxico de palabras y frases uniformes que se usan para indizar la materia del documento.

⁽²⁾Las listas de encabezamientos de materia también se conocen como *autoridades de materia*.

Cada una de las entradas de esta lista se conoce como encabezamiento.

Las funciones de las listas de encabezamientos son:

a) En la indización:

- Describen el contenido de los documentos.
- Agrupan en el catálogo todos los documentos de temáticas afines.
- Normalizan la terminología. Cada concepto se expresa en un término y sólo en este, evitando la ambigüedad del lenguaje natural.

b) En la recuperación:

- Permiten recuperar los documentos por el campo materia.
- Ayudan a localizar otros documentos a partir de las relaciones semánticas entre términos.

Las listas de encabezamientos se pueden clasificar por idioma, por tema y por tipo de biblioteca donde se aplican:

1) Por idioma

Hay listas de encabezamientos en diversos idiomas: catalán, inglés, francés, castellano, italiano, portugués, etc.

Listas de encabezamientos en catalán, castellano, inglés y francés

Lengua	Acrónimo	Nombre de la Lista
Catalán	LEMAC	Lista de encabezamientos de materia en catalán
Castellano	AM BNE	Autoridades de materia del catálogo de autoridades de la Biblioteca Nacional de España

Lengua	Acrónimo	Nombre de la Lista
	ME BUS	Encabezamientos de materia de la Biblioteca Universitaria de Sevilla
	AM CSIC	Autoridades de materia de la red de bibliotecas del CSIC
	LEM BP	Lista de encabezamientos de materia para bibliotecas públicas
Inglés	LCSH por materias	Library of Congress Subject Headings
	SEARS	Lista Sears (proviene del nombre de Minnie Earl Sears)
	BILINDEX	BILINDEX (inglés - castellano)
Francés	LAVAL	Lista Laval de la Universidad Laval (Canadá)
	RAMEAU	Répertoire d'autorité matière encyclopédique et alphabétique unifié

2) Por tema

Hay listas de ámbito enciclopédico (todas las ejemplarizadas en la tabla anterior) y listas especializadas en un tema.

3) Por tipo de biblioteca

Podemos distinguir las listas especializadas de aplicación en bibliotecas nacionales (AM BNE, LEMAC), en bibliotecas universitarias (ME BUS, LEMAC), en bibliotecas públicas (LEM BP) y en bibliotecas o centros especializados (AM CSIC). Algunos ejemplos son:

- Lista especializada en medicina: Medical Subject Headings (MESH)
- Lista especializada en música: Universidad de Yale.

Nosotros, en esta asignatura, trabajaremos con las listas siguientes:

- El “Catálogo de autoridades de materia de la Biblioteca Nacional de España”.
- El “Catálogo de autoridades de materia de la red de bibliotecas del CSIC”.
- La “Lista de encabezamientos de materia en catalán”.
- Y consultaremos la Library of Congress Subject Headings (LCSH).

Los criterios para escoger estas listas sobre el resto son:

- El conocimiento del idioma: para indizar es necesario un cierto conocimiento de la lengua de cada lista, por lo que en este módulo trabajaremos básicamente las listas en catalán (LEMAC) y en castellano (AM BNE y AM

A partir de ahora:

AM BNE = Autoridades de materia del catálogo de autoridades de la Biblioteca Nacional de España.
 LEMAC = Lista de encabezamientos de materia en catalán.
 ME BUS = Encabezamientos de materia de la Biblioteca Universitaria de Sevilla.
 LEM BP = Lista de encabezamientos de materia para bibliotecas públicas.
 AM CSIC = Autoridades de materia de la red de bibliotecas del CSIC.

⁽³⁾A partir de ahora denotamos Library of Congress Subject Headings con la sigla LCSH.

CSIC). También se facilitarán ejemplos de la LCSH³, para facilitar la desenvolvatura en su consulta y contribuir a la competencia lingüística.

- El alcance conceptual: las listas se adaptan a la realidad de la biblioteca, por eso escogemos dos listas de alcance enciclopédico (como las de la AM BNE y la LEMAC) y una especializada (como la del AM CSIC).
- La facilidad en el acceso: las listas están disponibles en red y de forma gratuita.

Utilizaremos el acrónimo LEM⁴ (Listas de encabezamientos de materia) cuando nos refiramos a principios comunes a todas las listas, sea cual sea su nacionalidad e idioma. Cuando se haga referencia a una lista concreta, se usarán sus acrónimos respectivos.

Listas en red

- AM BNE [accesible en línea].
- AM CSIC [accesible en línea].
- LEMAC [accesible en línea].
- LCSH [accesible en línea] (tiene algunas restricciones, por ejemplo que no se visualizan los términos genéricos).
- La lista de encabezamientos de materia para bibliotecas públicas del Ministerio de Cultura también se encuentra en la web, en formato listado.

⁽⁴⁾LEM es el acrónimo de *listas de encabezamientos de materia*.

3. Origen y evolución de las listas de encabezamientos

El origen de este tipo de lenguaje documental lo encontramos en Estados Unidos de América. Aparece por primera vez en 1876 de manos de Charles A. Cutter.

Charles A. Cutter era el bibliotecario en jefe del Ateneo de Boston cuando el gobierno de Estados Unidos le encargó un estudio sobre la catalogación en las bibliotecas del país. El informe tenía cuatro volúmenes, el último de los cuales contenía las *Rules for a printed dictionary catalogue*, donde expone las reglas para encontrar e indizar un documento según su materia. Aparece ya el nombre de *headings* (encabezamientos) para designar los términos de indización.

Lecturas complementarias

Para más información sobre el origen y la evolución de las listas de encabezamientos ved las obras siguientes:

I. Gil Leiva (2008). *Manual de indización. Teoría y práctica*. Gijón: Trea.

B. Gil Urdiciain (2004). *Manual de lenguajes documentales*. Gijón: Trea.

Charles A. Cutter elaboró los principios que rigen las LEM, como el principio de especificidad y el de entrada directa; todos ellos continúan vigentes en la actualidad. También formuló el principio de economía, las referencias entre términos aceptados y no aceptados, ya sean sinónimos u homónimos, los encabezamientos simples y compuestos, los signos de puntuación, la sintaxis a la hora de construir encabezamientos compuestos, etc.

Charles A. Cutter era miembro de la American Library Association (ALA⁵). En el año 1895, nueve años después de la primera edición de Cutter, la asociación publicó la *List of Subject Headings for use in a dictionary catalog* como lista de encabezamientos para indizar en colecciones medias y pequeñas. Tres años más tarde (1898) la lista fue ampliada por el departamento de catalogación de la Library of Congress o Biblioteca del Congreso.

1) Listas de encabezamientos de materia en inglés

En la tabla siguiente se recogen las listas más importantes de encabezamientos de materia en inglés:



Charles Ammi Cutter. Recordemos que Cutter fue el creador de la *Clasificación expansiva de Cutter* (Boston, 1891).

Ved también

Estos principios se explican en el apartado 5 de este módulo didáctico.

⁽⁵⁾ALA es la sigla de *American Library Association*.

Library of Congress Subject Headings 1909-hasta la actualidad	<p>En el año 1909 aparece la primera edición de la <i>Subject headings used in the dictionary catalogs of the Library of Congress</i>. Fue redactada para hacer frente a las necesidades reales de la biblioteca. Poco después la lista ya se utilizaba en bibliotecas públicas y académicas y la Biblioteca del Congreso editaba manuales y guías explicativas. En 1975, coincidiendo con la octava edición, pasó a llamarse Library of Congress Subject Headings (LCSH) que es el nombre y acrónimo que ha conservado hasta hoy en día.</p> <p>La versión entera tiene unos treinta mil términos y la reducida unos cuatro mil. Se actualiza cada año (se calculan unos siete mil cambios anuales entre incorporaciones y modificaciones) y los cambios se publican en el <i>Cataloguing Service Bulletin</i>. Los manuales para los usuarios se titulan <i>Subject cataloguing Manual: subject headings</i>, son normas para la correcta aplicación de encabezamientos, subencabezamientos y relaciones semánticas.</p> <p>Está implementada por todo el mundo. Isidoro Gil (2008, pág. 403) hace una recopilación de las bibliotecas nacionales del mundo que la usan o han adaptado y recoge aproximadamente treinta repartidas por los cinco continentes.</p> <p>En 1986 cambió las siglas que utilizaba (<i>sa, xx, x, see</i>) por las siglas de los tesauros⁶, contribuyendo a la normalización en la descripción de las relaciones semánticas de los términos.</p> <p>Actualmente la mayoría de las listas de encabezamientos de materia del mundo se basan en ella o son directamente traducciones de ella. Por ejemplo, si tomamos el concepto “Mar⁷” y miramos cómo está introducido en las diversas listas, podemos constatar el hecho de que unas se basan en las otras y que el rol predominante lo juega la LCSH, que está presente como fuente original en todo el resto.</p>
Bilindex	<p>Bilindex es una lista basada en la LCSH, en la LEM de la Organización de Estados Americanos (OEA) y en la LEM de la Universidad Autónoma de México (UNAM). Es una lista bilingüe, inglés-castellano. Se creó para ayudar a los analistas que no dominaban el castellano para indizar documentos en esta lengua. Se aplica en bibliotecas públicas de ciudades de menos de 100.000 habitantes y escolares, pero su grado de especialización y la cobertura que da a las publicaciones hispanas la hicieron también útil en centros especializados y bibliotecas técnicas.</p>
Sears 1923	<p>Es una LEM creada en Estados Unidos como versión abreviada de la LCSH, para bibliotecas pequeñas y escolares. La creó Minnie Earl Sears el año 1923 con el título <i>List of Subject headings for small libraries</i> basándose en la experiencia de nueve pequeñas bibliotecas y siguiendo las directrices de la Library of Congress. Actualmente se usa en este tipo de bibliotecas norteamericanas y también en las de Gran Bretaña y Australia.</p>

OEA es la sigla de *Organización de Estados Americanos*.

UNAM es la sigla de *Universidad Autónoma de México*.

Las listas Bilindex y Sears no están en abierto en la red pero podemos consultar ejemplos en: B. Gil Urdiciain (2004), *Manual de lenguajes documentales* (págs. 44 a 63).

(6)

Siglas que utiliza la LCSH

	LCSH hasta 1986	A partir de 1986
Equivalencia	X	UF
Ved	See	Use
Genérico	xx	Bt broader term
Específico	s.a	Nt narrow term
Términos relación asociativa	Va	Rt related term

(7)

El concepto "Mar"

AM BNE	AM CSIC	LEMAC	LCSH	RAMEAU
Mares y océanos Fuentes: LCSH (Ocean) CSIC (Mares y océanos) ME BUS (Océano)	Mares y océanos Fuente LCSH (Ocean) Térm.mat. LCSH/BNF <u>Ocean</u>	Mar Fuente LCSH: (enc.: Ocean)	Ocean.	Mer Equív. LCSH: Ocean

2) Listas de encabezamientos de materia en francés

En la tabla siguiente se recogen las listas más importantes de encabezamientos de materia en francés:

Laval Répertoire de vedettes-matière (RVM) de la Universidad Laval	El Répertoire de vedettes-matière es una lista de encabezamientos de materia en francés, parcialmente bilingüe francés-inglés. Contiene materias y nombres geográficos. Sus autores son de la Universidad canadiense de Laval. Es una adaptación de diversas listas, como la LCSH, la Canadian Subject Headings (CSH*) y la lista de la Biblioteca Nacional de Canadá. Es una lista desarrollada en el respeto por la lengua francesa y conceptos propios de Canadá que no estaban recogidos en la LCSH. Actualmente también incorpora términos provenientes del Medical Subject Headings de la Biblioteca Nacional de Medicina de EE.UU. (MESH**) y del tesoro AAT*** (Art, Architecture Thesaurus) de la fundación J. Paul Getty.
RAMEAU	RAMEAU son las siglas de Répertoire d'Autorité Matière Encyclopédique et Alphabétique Unifié, la lista de encabezamientos usada en Francia. Está basada en el Répertoire de vedettes-matière de la Universidad de Laval. RAMEAU se usa en la red de bibliotecas públicas francesas y en la Biblioteca Nacional Francesa (BNF) cubriendo de esta manera la mayoría de la red bibliotecaria del país. Desde 1987 se gestiona desde la BNF.

* CSH es la sigla de *Canadian Subject Headings*.

** MESH es la sigla de *Medical Subject Headings*.

*** AAT es la sigla de *Art, Architecture Thesaurus*.

3) Listas de encabezamientos de materia en castellano

a) En América:

En la tabla siguiente se recogen las listas más importantes de encabezamientos de materia en castellano en América:

Lista de encabezamientos de la Organización de Estados Americanos (OEA)	La lista de encabezamientos de la OEA es la ampliación de la lista elaborada por Carmen Rovira y Jorge Aguayo para bibliotecas nacionales, universitarias, especializadas y públicas de Latinoamérica. Está basada en parte en la LCSH. Es una lista que incluye todas las ramas del conocimiento y abarca desde términos generales a términos muy específicos de un ámbito temático.
--	---

b) En el Estado español:

En España la primera noticia que tenemos de una lista de encabezamientos es de 1935, de Juan Vicens Lavalle, pero no llegó a ser utilizada. La siguiente noticia es del año 1939, cuando Lasso de la Vega y Cesáreo Goicoechea publican

Reglas para la redacción y ordenación de los catálogos diccionarios, seguidas de una *Lista de encabezamientos de materia y números apéndices*. Diez años más tarde, en 1949, Concepció Guarro publica una lista de encabezamientos para el término Cervantes (para bibliotecas que tienen colecciones cervantinas, como la BC) y otra más amplia en 1952.

En el año 1965 aparecen, de manos del CSIC, los *Encabezamientos de materia para bibliotecas* y en 1986 la lista de encabezamientos de materia para bibliotecas públicas, del Ministerio de Cultura. Con respecto a las bibliotecas universitarias, las de Sevilla y la Complutense fueron las primeras en redactar una lista de materias para sus fondos.

En la tabla siguiente se recogen las listas más importantes:

AM BNE	Los encabezamientos de materia se empiezan a utilizar en la Biblioteca Nacional de España en el año 1958, cuando se establece la "ficha única" para todas las bibliotecas públicas del Estado y se crea un primer fichero de autoridades de materia con los elementos básicos de los registros de autoridad. Este fichero de materias sirvió de base para la futura <i>Lista de encabezamientos de materia para bibliotecas públicas</i> , editada por primera vez por la Dirección General del Libro del Ministerio de Cultura en el año 1986.
ME BUS	La Universidad de Sevilla se inspiró en la LCSH y en la Laval. Es una lista de encabezamientos de tipo enciclopédico, ya que tiene que dar cobertura a todas las bibliotecas universitarias de la ciudad. En un volumen aparte se facilitan las traducciones de los epígrafes al inglés y francés.
AM CSIC	La primera lista data de 1965, pero la primera editada es de 1987. Actualmente contiene aproximadamente treinta y ocho mil términos y en la versión impresa cada encabezamiento viene acompañado de su traducción al inglés. Es la lista que sirve para indizar las bases de datos del CSIC. Dada la temática de estas bases de datos, la lista recoge términos especializados.
LEM BP	En el año 1981 se creó una comisión de bibliotecarios para redactar una LEM especial para los fondos de una biblioteca pública, es decir, para colecciones generales de unos treinta mil ejemplares. En el año 1986 el Ministerio de Cultura la editó por primera vez. Es una lista muy implementada en el Estado español. En 1991, el Ministerio, junto con la Biblioteca Nacional de España, redactó unas normas para su aplicación y redacción de nuevos epígrafes.

4) Lista de encabezamientos de materia en catalán

En la tabla siguiente se recoge la lista de referencia más utilizada en la lengua catalana:

LEMAC	La primera lista en catalán fue elaborada en el año 1983 por el Institut Català de Bibliografia. Se decidió traducir y adaptar la Lista de encabezamientos de materias para bibliotecas de 1967 de C. Rovira y J. Aguayo y en paralelo se iba comprobando con el original de la LCSH. La tarea actualmente corre a cargo del Servei de Normalització de la Biblioteca de Catalunya y las fuentes se han ampliado con las adaptaciones de la Laval y el Rameau. Se actualiza constantemente, en teoría el periodo es semanal. La LEMAC se usa en la Biblioteca de Catalunya y en el Consorci de Biblioteques Universitàries de Catalunya (CBUC ⁸).
--------------	---

⁽⁸⁾CBUC es la sigla de *Consorci de Biblioteques Universitàries de Catalunya*.

4. Tipologías de las listas de encabezamientos

Las listas de encabezamientos de materia son un lenguaje documental que tiene las siguientes características:

a) **Sintético**. Se trata de un lenguaje documental sintético porque tiene por objetivo sintetizar el contenido de los documentos en un único término de indización o encabezamiento.

Ejemplo

La obra siguiente contiene muchos temas, pero las LEM lo sintetizan en un solo encabezamiento.

BALAGUER, Víctor. *Los Frailes y sus conventos: su historia, su descripción, sus tradiciones, sus costumbres, su importancia / por Victor Balaguer; obra acompañada de hermosas láminas grabadas sobre acero por Antonio Roca, dibujadas por J. Puiggarí*. Barcelona [etc.]: Llorens Hermanos, -1851.

Materia con AM CSIC: **Vida religiosa y monástica**

Materia con LEMAC: **Vida religiosa i monàstica**

b) **Natural**. Es un lenguaje natural porque la indización se basa en el uso de palabras y/o expresiones que pertenecen al lenguaje o discurso común y que se encuentran en los mismos documentos (título, resumen, texto, etc.) o son propias de la misma disciplina.

Ejemplo

En una LEM el encabezamiento es **Pintura**, mientras que en un sistema de clasificación como la CDU, que es codificado, es **75**.

c) **Controlado**. Es un lenguaje controlado porque tiene como objetivo la representación unívoca del contenido de los documentos y de las consultas: un único concepto por término y un único término por concepto (eliminando la sinonimia y homonimia). En la indización y en la recuperación sólo se pueden utilizar los términos o encabezamientos que figuran en el vocabulario finito que es la lista de encabezamientos.

Ved también

Recordad que estas características se trataron en el módulo "Análisis de contenidos: resumen e indización".

Ejemplo de control de la sinonimia

Un documento sobre los chicos escoltas podría ser indizado de todas las formas siguientes, pero una LEM sólo valida un término como término aceptado.

Sinónimos	Término aceptado (LEMAC)
Boy Scouts Escoltas Exploradores (Boy Scouts) Minyons de Muntanya Noies guies Nois guies	Escoltes

O un documento sobre el hambre en el mundo:

Sinónimos	Término aceptado (AM CSIC)
Escasez de alimentos Hambre colectiva Hambre en el mundo Hambruna	Hambre

Ejemplo de control de polisémicos

Un documento trata de la masa de harina. El término “Masa” es polisémico, es decir, tiene el mismo origen etimológico y se escribe y pronuncia igual que otro término con significado diferente. Las LEM diferencian los polisémicos y los homónimos de diversas formas, por ejemplo poniendo entre paréntesis el ámbito temático.

Polisémicos	Término aceptado (AM CSIC)
Masa	Masa (Cocina) Masa (Física)

d) Precoordinado. Es un lenguaje precoordinado porque la coordinación de los diferentes conceptos que conforman una materia se produce en el momento de la indización, es decir, antes de almacenarlo o de incorporarlo a un catálogo según unas reglas y principios propios.

Ejemplo de lenguaje precoordinado

Una guía sobre los palafitos construidos a lo largo del río Mekong se indiza:

Incorrecto	Correcto
Obras de referencia - Derecho	Derecho - Obras de referencia

Obras de referencia lleva una nota de aplicación que dice:

Véase además el subencabezamiento *Obras de referencia* bajo determinadas materias *Arte-Obras de referencia*

Por ello, en primer lugar debemos poner la materia, en este caso *Derecho* y en segundo lugar *Obras de referencia*.

Breve explicación:

Cuando consultamos *palafitos* vemos que nos remite al término aceptado que es *Viviendas lacustres*, que puede llevar subdivisión geográfica, por lo tanto podemos añadir el nombre del río Mekong. Finalmente las listas de encabezamientos tienen subdivisiones de forma, como es el caso de las guías, que se pueden añadir en última posición. Por lo tanto el orden del encabezamiento es el propuesto y no ninguno de los incorrectos. La precoordinación asegura que todos los indizadores indizan de la misma manera y forma.

e) **Estructura combinatoria o asociativa.** Es un lenguaje de estructura combinatoria en que los términos de indización se organizan alfabéticamente y se combinan entre sí (encabezamientos y subdivisiones) en el momento de la indización.

Ejemplo de combinación de un encabezamiento + un subencabezamiento

AM CSIC: Vidrio - Propiedades eléctricas

5. Principios de las listas de encabezamientos

En 1990, la Sección de Clasificación e Indización de la International Federation of Library Associations (IFLA⁹) encargó un proyecto para estudiar las diferentes listas de encabezamientos usadas en las bibliotecas nacionales del mundo. Los países participantes fueron Canadá, Alemania, Irán, Portugal, EE.UU., Francia, Noruega, Polonia y España.

⁽⁹⁾IFLA es la sigla de la *International Federation of Library Associations*.

El panorama era poco homogéneo entre las diferentes listas, en contraposición a los sistemas de clasificación cuya implementación era superior, quizá por tradición y por la inexistencia de barreras idiomáticas (recordamos que los sistemas de clasificación son codificados).

El grupo de trabajo editó un documento llamado *Principles underlying subject headings languages* o Principios fundamentales en los lenguajes de encabezamientos de materia, donde exponían los principios primordiales de toda lista de encabezamientos fuera cual fuera su lengua. Dividían los principios en: principios de construcción y principios de aplicación.

Principios fundamentales en los lenguajes de encabezamientos de materia

Principios de construcción	Principios de aplicación
Principio de encabezamiento uniforme Principio de sinonimia y principio de homonimia Principio de semántica Principio de sintaxis Principio de consistencia Principio de denominación Principio lingüístico Principio del usuario	Principio de especificidad Principio de economía

En este apartado trataremos los principios de construcción y en el capítulo titulado “Aprendiendo a indexar con una lista de encabezamientos” trataremos los dos principios de aplicación.

a) Principio de encabezamiento uniforme: cada materia tiene que ser representada siempre con un único encabezamiento y cada encabezamiento ha de representar una única materia. Para conseguir este principio hay que aplicar el control sobre el vocabulario.

Ejemplo AM BNC

Términos afines como *Sinopsis* y *Conferencias* siempre se indizarán como **Congresos y asambleas**.

b) Principio de sinonimia y principio de homonimia: las LEM tienen que controlar la sinonimia y las posibles variantes de polisemia, como la homonimia.

Los sinónimos son las palabras o expresiones que tienen el mismo significado y que dentro de un sistema documental pueden comportar silencio documental.

Las LEM controlan los sinónimos mediante la referencia de *Usad* o *Use*.

Ejemplos de sinonimia

Términos	AN BNE
Conceptos semejantes	Edificación USE Construcción
Variantes ortográficas	Kimbanda USE Quimbanda
Variedades lingüísticas	Ananas comosus USE Piñas
Términos obsoletos	Aeronaves USE Aviones

En cuanto a los polisémicos, decimos que dos palabras son polisémicas cuando el mismo signo lingüístico, palabra o sonido tiene más de un significado pero proviene de un mismo origen etimológico, guardando una cierta relación conceptual a pesar de los diferentes usos. Los homónimos son un tipo de polisemia caracterizada por tener dos o más orígenes etimológicos diferentes.

Tanto los polisémicos como los homónimos comportan ruido documental dentro de un sistema documental.

Generalmente, el contexto en el que está insertada la palabra ya evita los problemas de ambigüedad en la comunicación habitual de conversación o lectura, pero un polisémico introducido en un sistema documental, sin el contexto, puede dar lugar a ruido documental. Las LEM controlan los polisémicos y homónimos de diversas maneras: adjetivando, poniendo paréntesis con el nombre de la disciplina, haciendo frases, jugando con el singular y el plural, etc.

Ejemplos de indización de polisémicos y homónimos

Polisémicos	AM CSIC	LEMAC
Matriz: ginecología Matriz: matemáticas	Matriz Matrices (Matemáticas)	Úter Matrius (Álgebra)
Homónimos	AM CSIC	LEMAC
Metro: transporte Metro: medida	Ferrocarriles metropolitanos Metro (Unidad de longitud)	Metro Metre (Unitat de longitud)

Nota

En catalán el término *Metro* sólo tiene un significado, el del transporte, ya que la unidad de medida es "Metre".

c) **Principio de semántica:** las LEM organizan los encabezamientos en una estructura semántica, es decir, en relaciones de equivalencia, jerarquía y asociación.

Ejemplo de un registro de autoridad de la AM BNE

Aviones

Véase además los nombres de los distintos aviones y clases de aviones
Mirage (Avión) Aviones militares

Usado por:

Aeronaves

Aeroplanos

Término específico:

Autogiros

Aviones a reacción

Aviones de entrenamiento

Aviones de transporte

Aviones militares

Helicópteros

Hidroaviones

Ultraligeros

Término relacionado:

Aeronáutica

Aviones

Modelos

Industria aeronáutica

Transportes aéreos

Accidentes aéreos

Aviación

Secuestro de aviones

Fuentes:

LCSH (Airplanes)

LEM (Aviones)

Nº Registro:

BNE19900956268

d) **Principio de sintaxis:** las LEM tienen que establecer las reglas de combinación de los encabezamientos con los subencabezamientos.

Ejemplo

Encabezamiento aceptado

Atlas

Nota de alcance

Este encabezamiento, que sirve para agrupar los atlas geográficos mundiales, se puede utilizar únicamente seguido de subdivisión o del adjetivo que designa el país de publicación, ex.: *Atlas catalans* para los atlas geográficos mundiales publicados en Cataluña. Las obras que tratan de los atlas se encuentran bajo el encabezamiento *Atlas - Historia*

e) **Principio de consistencia:** los encabezamientos nuevos que se incorporen a una LEM tienen que mantener la coherencia con los existentes: de forma y estructura sintáctica.

Ejemplo

Supongamos que tenemos que incorporar un encabezamiento para la materia Matemáticas. Como las demás materias académicas se han introducido en singular, tenemos que escribir **Matemática**.

f) **Principio de denominación:** los nombres propios, de persona y de lugar tienen que mantener también la coherencia con los ya existentes.

Ejemplo

Si hay que introducir un nombre y no existe, tiene que seguir las normas de las reglas angloamericanas de catalogación (AACR2r).

g) **Principio lingüístico:** los términos de la lista tienen que estar en el idioma del catálogo.

Se tiene que respetar el orden natural de las expresiones (lo que Cutter llamaba principio de entrada directa). Las listas española y catalana sólo permiten invertir el orden natural de la expresión cuando la palabra significativa no es la primera.

Ejemplo de excepción

Juegos, Teoría de los

h) **Principio del usuario:** el vocabulario de las LEM se tiene que adecuar al usuario al cual va destinada la lista. El lenguaje utilizado en los encabezamientos tiene que reflejar las características de los usuarios.

Las expresiones específicas o muy académicas se reservan para listas especializadas.

Ejemplos

En una lista enciclopédica, tipo la de la AM BNE y LEMAC:

Quercus ilex USE **Encina**
Biocatalizadores USE **Enzimas**

Si un mismo concepto se puede expresar de dos formas, se escoge la menos vulgar:

Tripas USE **Intestinos**

6. Encabezamientos y subencabezamientos

Los elementos que componen una lista de encabezamientos son los encabezamientos y los subencabezamientos

6.1. Encabezamientos

“Son una o diversas palabras que representan conceptos.”

Blanca Gil (2004, pág. 30).

Los **encabezamientos** son palabras o frases específicas que describen la materia. Están redactados en lenguaje natural pero han sido controlados, es decir, son términos unívocos que han controlado la ambigüedad derivada de la sinonimia y homonimia. Son puntos de acceso por materias.

Ejemplos de encabezamientos

AM BNE	Periodismo Ajedrez
AM CSIC	Macromoléculas Teoría crítica
LEMAC	Fotografia Disseny decoratiu
LCSH	Magic and poetry Trees

6.2. Subencabezamientos

Los **subencabezamientos** son los epígrafes o palabras que van detrás de un encabezamiento a efectos de matizar y precisar el significado. Son encabezamientos situados detrás de otro encabezamiento.

Ejemplo según AM CSIC

En el ejemplo siguiente, *Bacterias* es el encabezamiento y *Crecimiento* es el subencabezamiento.

AM CSIC: **Bacterias - Crecimiento**

El signo que separa los encabezamientos de los subencabezamientos es el guión, pero también hay LEM que utilizan los dos puntos (como LAVAL) y la coma.

Algunas listas ponen los subencabezamientos en listas anexas (LEM de bibliotecas públicas) y otras los listan alfabéticamente en el cuerpo central (LEMAC).

Hay subencabezamientos de aplicación general (como los de forma) y otros que hay que combinar bajo un tipo concreto de encabezamiento. En este último caso la lista nos informa con una nota explicativa.

Ejemplo según AM BNE

Impuestos Véase además los nombres de los distintos impuestos y clases de impuestos Sisa (Impuesto) Impuesto sobre el valor añadido Impuestos municipales y el subencabezamiento Impuestos bajo determinadas materias Automoviles-Impuestos

Significa que podemos añadir *Impuestos* detrás de un encabezamiento de tipo *Tabaco* y hacer:

Tabaco - Impuestos

Hay cuatro tipos de subencabezamientos: los de materia, los geográficos, los cronológicos y los de forma.

1) Subencabezamientos de materia

Se usan para precisar el contenido del encabezamiento principal.

Ejemplos

Acero - Aleaciones
Acero - Electrometalurgia
Acero - Fundición
Acero - Metalografía
Acero - Metalurgia

También se usan para representar la forma en que se presenta el contenido, o el punto de vista.

Ejemplo

Arte - Historiografía

Muchos encabezamientos + subencabezamiento de materia ya vienen contruidos en la lista; otros los tenemos que combinar nosotros siguiendo una nota de aplicación que dice que:

a) El encabezamiento x también puede funcionar como subencabezamiento o subdivisión.

Ejemplo

El encabezamiento *Argot* también puede funcionar como subencabezamiento, según dice la nota.

Ejemplo de encabezamiento (izquierda) y de encabezamiento que funciona como subencabezamiento o subdivisión (derecha)

Enciclopedias

Úsase para obras sobre el estudio de las enciclopedias así como para las propias enciclopedias

Véase además el subencabezamiento Enciclopedias bajo determinadas materias Medicina-Enciclopedias

Usado por: Enciclopedias y diccionarios

Término genérico: Obras de referencia

Término específico: Enciclopedias escolares

Enciclopedias femeninas

Enciclopedias infantiles

Término relacionado: Diccionarios

Fuentes: LCSH (Encyclopedias and dictionaries)

LEM (Enciclopedias)

UNESCO (Enciclopedias)

Nº Registro: BNE19900961589

Enciclopedias

Úsase para obras sobre el estudio de las enciclopedias así como para las propias enciclopedias Véase además el subencabezamiento Enciclopedias bajo determinadas materias Medicina-Enciclopedias

b) Subencabezamientos creados únicamente para ir detrás de algún término. La nota de aplicación nos dirá si hace falta que vayan bajo nombre de persona, tema, guerra, sustancia, nombre geográfico, lengua etc.

Ejemplo

La AM BNE dice en la nota de alcance que el término *Prisioneros* puede funcionar como subdivisión, pero con la condición de ir detrás de un nombre de guerra u organización militar. De manera que podemos hacer *Guerra Mundial, 1939 - 1945*.

Prisioneros de guerra Véase además el encabezamiento *Prisioneros de guerra* seguido de adjetivo que indique nacionalidad *Prisioneros de guerra alemanes* y el subencabezamiento *Prisioneros* bajo los nombres de las distintas guerras *Guerra mundial, 1939-1945-Prisioneros*

La precoordinación evita que algún analista pueda hacer un encabezamiento como:

Prisioneros - Guerra Mundial, 1939 - 1945

2) Subencabezamientos de lugar o topográficos

Son nombres de lugares geográficos que precisan el encabezamiento principal. Se usan en todo tipo de temas (*Economía, Arte, Legislación, Historia, etc.*).

Algunas LEM indican qué encabezamientos pueden llevar subencabezamientos de lugar:

- en la LEMAC sale como *Subd. Geog.*;

- en la LCSH sale (*may subdivisions geographical*) y también su contrario (*Not Subd. Geog.*);
- en la LAVAL: *nom de lieu*.

Otros LEM no especifican nada.

Ejemplos de subencabezamientos de lugar

AM CSIC	Fiestas religiosas -[lugar]	Fiestas religiosas - Galicia
LEMAC	Pintura ¹⁰ (Subd. Geog.)	Pintura - França
LCSH	Balinese drama (May Subdv. Geog)	Balinese drama - Denpasar

⁽¹⁰⁾Nota: Los paréntesis son informativos, la subdivisión debe separarse con un guión.

Hay encabezamientos que permiten adjetivar el nombre del país en lugar de hacer la forma compuesta.

Ejemplo según AM BNE

Escultura

Úsase tanto para obras teóricas como para colecciones, exposiciones, concursos, etc. Véase además el encabezamiento Escultura seguido de adjetivo que indique estilo, época o nacionalidad Escultura renacentista Escultura española

De esta manera, en lugar de hacer Escultura - Francia, nos permite hacer Escultura francesa.

Y como comenta Blanca Gil (2004, pág. 36), los constructores de la lista tienen que procurar que el significado no cambie: Pintura - España (pintura que se encuentra en España) no es Pintura española (sólo de pintores españoles).

3) Subencabezamientos de tiempo o cronológicos

Son periodos, fechas concretas, siglos. Cada lista nos indica la manera de introducir este tipo de subencabezamientos.

Ejemplo

El siglo XIX puede introducirse de diversas maneras: 1800-1899 o 19.

Los periodos se separan por barra: 1939/1975.

4) Subencabezamientos de forma

Son la forma en que se presenta el documento, son formas del tipo: enciclopedias, estadísticas, anuarios, biografías, etc. Algunas listas, como la del LEM BP, incluyen un anexo con estos encabezamientos de forma.

Ejemplo

Medicina - Enciclopedias
 Medicina - Estadísticas
 Medicina - Revistas

Ejemplo

En el módulo “La cadena documental” se pone un ejemplo sobre la obra *Vocabulario básico de la historia medieval* de Pierre Bonnassie y su referencia documental. Pues bien, en este caso el encabezamiento lleva un subencabezamiento de forma:

Historia medieval - Terminología

Ved también

Podéis encontrar el ejemplo sobre la obra *Vocabulario básico de la historia medieval* de Pierre Bonnassie y su referencia documental al apartado 3 del módulo “La cadena documental”.

5) Orden de los subencabezamientos

Si la LEM no indica lo contrario, el orden es el siguiente:

Materia principal + subencabezamiento de materia + de lugar + de tiempo + de forma

Ejemplo

Dientes - Cuidado e higiene - España - S. XIX - Estadísticas

Si el primer lugar lo ocupa un nombre geográfico, entonces el orden es:

Lugar + materia + tiempo + forma

Ejemplo

Barcelona - Política y gobierno - 1939/1975 - Revistas

7. Reglas de vocabulario y signos de puntuación de las LEM

Las reglas forman parte del control del vocabulario que todo lenguaje documental controlado debe tener presente. Son convenciones propias de cada lista. Aunque nunca tengamos que construir una LEM entera, sí que nos podemos encontrar en la situación de tener que incorporar un encabezamiento nuevo y hay que saber las reglas:

1) Regla del singular/plural

Culturalmente, las listas inglesas y españolas prefieren el plural, mientras que las francesas y las italianas se decantan por el singular, si bien siempre hay excepciones debidas a particularidades de la lengua, costumbres, etc.

No hay una normativa clara para el uso del singular/plural. Algunos autores como Blanca Gil apuntan que el plural se usa para temas concretos y contables (enciclopedias, museos) y el singular para temas abstractos e incontables (belleza, positivismo, etc.) Blanca Gil (1996, pág. 337).

En España, la Biblioteca Nacional redactó en el año 1991 unas normas de aplicación del singular/plural. A modo de ejemplo, éstas son algunas de las reglas redactadas por la Biblioteca Nacional:

Ejemplos de reglas singular/plural

En singular	En plural
Conceptos abstractos: <i>caridad</i> Fenómenos de la naturaleza: <i>nieve</i> Propiedades: <i>transparencia</i> Procesos biológicos: <i>respiración</i> Sistemas de creencias: <i>budismo</i> Áreas temáticas, disciplinas: <i>física</i> Objetos no cuantificables: <i>arena</i> Órganos únicos: <i>corazón</i>	Grupos: <i>museos, delfines, italianos</i> Órganos múltiples: <i>huesos, pelos</i> Epígrafes de forma: <i>portales, diccionarios</i>

Si la palabra es polisémica, algunas listas utilizan el singular y el plural a la vez para diferenciar.

Ejemplo

En la lista del AM CSIC se usa Teatro (obra) y Teatros (lugar).

2) Regla de las siglas

Se prefiere la expresión completa a las siglas o acrónimos.

Los sustantivos no contables

En castellano y catalán usamos el singular para los sustantivos no contables, como *arroz* y *café*, pero hacemos excepciones con *garbanzos* y *judías*.

Lectura complementaria

Podéis encontrar las observaciones de Gil en:

B. Gil Urdiciain, (1996) (artículo)

Ejemplo de la regla de las siglas (AM CSIC)

Correcto	Incorrecto
Descripción Bibliográfica Normalizada Internacional	ISBD

3) Regla de mayúsculas/minúsculas

Se recomienda que la primera palabra del encabezamiento, del subencabezamiento, del encabezamiento cambiado y la que va dentro de paréntesis comience con mayúscula.

Ejemplo con la AM CSIC

Lámparas
 Género de punto - Industrias
 Montecarlo, Método de
 Intervalo de energía (Física)

4) Regla de los signos de puntuación

Las LEM utilizan diversos signos y entre ellas no hay una unidad de criterio, por lo tanto se hace necesario consultar las introducciones y manuales adjuntos a cada lista.

Los signos utilizados y sus usos más habituales son los siguientes:

Signos poco habituales

Signo	Utilidad	Ejemplos
Guión	Separa encabezamientos de los subencabezamientos	Alimentos - Calorías África - Historia - 1884/ 1918
Coma	Para expresiones inversas	Gaita, Música para
Paréntesis	<ul style="list-style-type: none"> Diferencia homónimos Expresa el contexto, la situación geográfica, el ámbito temático 	Mercurio (Divinidad romana) Mercurio (Planeta) Capitolio (Italia)

Hay otros signos que no son de uso común en las listas de encabezamientos que estamos utilizando a lo largo de esta asignatura, pero que es bueno que sepáis que existen. Mostramos algunos ejemplos de los mismos extraídos de la obra de Gil (2008).

Lectura complementaria

I. Gil Leiva (2008). *Manual de indexación* (pág. 136). *Teoría y práctica*. Gijón: Trea.

Signo	Utilidad	Ejemplos
Dos puntos	Separa encabezamientos de los subencabezamientos.	Emploi: Paris
Comillas	Para obras literarias y personajes de ficción	Cervantes, Miguel de, "Sancho Panza"
Doble barra	Expresa dos materias relacionadas (al estilo del sistema Colon de la CDU), hay que hacer la inversa	Art//Economía Economía//Art

8. Términos simples y compuestos

Los encabezamientos y subencabezamientos pueden ser simples o compuestos.

8.1. Encabezamientos simples

En los **encabezamientos simples** el tema se expresa en una sola palabra, generalmente un sustantivo, pero también pueden ser adjetivos, verbos sustantivados o nombres geográficos. Tanto pueden ser términos en singular como en plural.

Ved también

Recordad que el uso del singular/plural se explica en el apartado 7 de este módulo didáctico.

Ejemplos de encabezamientos simples

	Singular	Plural
AM BNE	Ética	Extremeños
AM CSIC	Carnitina	Animales
LEMAC	Grècia	Cireres
LCSH	Toponymy	Museums

8.2. Encabezamientos compuestos

En los **encabezamientos compuestos** la materia se expresa con más de una palabra.

Ejemplo

Árboles frutales
Topographical surveying

Los encabezamientos compuestos pueden estar formados por:

1) Dos sustantivos unidos por una preposición

AM BNE	Cantos de pájaros Música para acordeón
AM CSIC	Láseres en cirugía Riego por aspersión
LEMAC	Gèneres de punt Animació per ordinador
LCSH	Packing for shipment

2) Dos sustantivos unidos por conjunción: los temas acostumbran a ser correlativos o bien contrarios.

Industria y Comercio

Industria y Universidades

En estos casos (como en el sistema Colon de la CDU) es conveniente entrar dos veces el encabezamiento en el catálogo.

Industria y Comercio

Comercio e Industria

AM BNE	Máximas y aforismos Cine y política
AM CSIC	Emigración e inmigración Pintura e ilustración de animales
LEMAC	Germans i germanes Internet i adolescents
LCSH	Rites and ceremonies

3) Dos o más sustantivos separados por coma y acabados en etc.

AM CSIC	Tratados, manuales, etc. Restaurantes, cafeterías, etc.
LEMAC	Acadèmies, societats culturals, etc. Situació legal, lleis, etc.

4) Un sustantivo calificado con un paréntesis

AM BNE	Vela (Deporte)
AM CSIC	Cambio (Economía) Cambio (Filosofía) Cambio (Psicología) Cambio (Sociología)
LEMAC	Radi (Metall) Radi (Os)
LCSH	Canarias (Ship) Canarias (Islands)

5) Un sustantivo adjetivado

AM BNE	Industria forestal Escultura italiana
AM CSIC	Medicina tropical Matemáticas difusas

LEMAC	Febre groga Cuina francesa
LCSH	Modern Jazz Giants

6) Frases hechas

AM BNE	Viajes alrededor del mundo
AM CSIC	Pueblos indígenas - Primer contacto con los occidentales
LEMAC	Amor platònic en la literatura
LCSH	Out of the blue series

7) **Frases donde se resalta el concepto al inicio del epígrafe separado con comas:** consiste en alterar el orden de la expresión natural, colocando en primer lugar el término con más carga significativa y separado del resto por una coma.

Ejemplo

Ángeles, Nuestra Señora de los

No se usan mucho porque se prefiere el principio de entrada directa. Son de aplicación para conceptos del tipo *Teoría de, Ley de, Mar de, Tests de, Nuestra Señora de.*

AM BNE	Bender, Test de
AM CSIC	Addison, Enfermedad de
LEMAC	Acció, Teoria de l' Albera, Serra de l'
LCSH	Absolute, The Art, Romanesque

8.3. Orden de los encabezamientos

Las listas están ordenadas alfabéticamente por los términos aceptados y no aceptados.

Ejemplo de la letra B

AM CSIC	LEMAC
Baal (Personaje mitológico) Baamarani Babel, Torre de Babesiasis Babesiasis bovina Babesiosis Babies Babilonio Babismo Bable	Babel, Torre de Babilonia - Civilización Babilonia (Ciudad antigua) Babismo Bable Bacalao Bacalao - Industria y comercio Bacalao - Pesca comercial Bacalao - Industria del Bacanales

Mayoritariamente siguen el orden iniciado por la LCSH:

- Encabezamiento simple/compuesto.
- Encabezamiento simple/compuesto - Subencabezamiento.
- Encabezamiento compuesto cambiado **Acción, Teoría de la**.
- Encabezamiento con calificador **Efecto invernadero (Meteorología)**.
- Encabezamiento formando una frase.

9. Las relaciones semánticas

Cada encabezamiento forma un registro de autoridad. En este registro observamos en primer lugar el encabezamiento aceptado y toda una serie de términos conceptualmente equivalentes, más genéricos, más específicos y términos que evocan una relación.

Este apartado trata los temas siguientes:

a) Relación de equivalencia.

b) Relación de jerarquía:

- Genérica.
- Específica.

c) Relación de asociación.

d) Símbolos utilizados para designar los tres tipos de relaciones.

Ejemplo de registro de autoridad extraído de la AM CSIC

Encabezamiento	● Cartas
Nota acceso público	Usese para las colecciones de cartas literarias. Para las obras sobre las epístolas que se cantan en misa, úsese: Epistolarios
Usado por	● Cartas literarias ● Correspondencia
Térm. genérico	Biografía
Térm. genérico	Literatura
Térm. específico	Novela epistolar
Véase además	Cartas -- Redacción
R.Compleja (Mat.)	Usese además el subencabezamiento: Correspondencia , detrás de nombres de persona o grupos de personas, y el subencabezamiento: Documentos y correspondencia , detrás de tipos de industrias, organizaciones, etc. Puede adjetivarse además por lengua o nacionalidad. Ej.: Médicos-Correspondencia, Cartas latinas
Fuente	LCSH (Letters)
Térm.mat. LCSH/BNF	● Letters

Nota

Estas tres relaciones también se dan en tesauros.

9.1. Relación de equivalencia

En la lista hay términos aceptados y términos no aceptados.

Los **términos aceptados** son los encabezamientos que se pueden utilizar para indizar.

Acostumbran a ir en negrita. En los ejemplos, es el término **Cartas**.

Los **términos no aceptados** son sinónimos, siglas, términos demasiado especializados, barbarismos, formas incorrectas, etc.

Estos términos se encuentran recogidos en la lista pero no están admitidos para indizar documentos. Su función es controlar todas las posibles variantes y ofrecer una forma de encabezamiento única para indizar.

Ejemplo

Los términos no aceptados de **Cartas** son:

Cartas literarias
Correspondencia

La relación de equivalencia une los dos términos. Si el analista está buscando el concepto "Correspondencia", se encontrará en la lista con una referencia de "USE" que le indicará que tiene que indizar **Cartas**. De esta manera los analistas siempre indizarán **Cartas** y se evitarán entradas dobles por el mismo concepto.

La función de estas relaciones de equivalencia es hacer de puente entre las palabras utilizadas en lenguaje natural y las utilizadas en un lenguaje controlado, tanto en la indización como en la recuperación. En la indización este puente permite mantener la coherencia y la univocidad en los indizadores y en la recuperación permite a los usuarios recuperar mejor los documentos, ya que no se produce silencio ni ruido documental. No es posible recuperar un término no aceptado, ya que el mismo programa que mantiene la lista hará la remisión al término aceptado.

Ejemplo

Si un usuario busca en las bases de datos del CSIC documentos sobre los "Genios", el mismo programa lo remitirá a **Superdotados**. Si busca en un catálogo indizado con la LEMAC por el término *Infància*, lo remitirá a **Infants**.

Las siglas utilizadas para expresar la relación de equivalencia entre ambos términos son:

	Catalán	Castellano	Anglès
Empreu	EM empreu	USE	USE
Emprat per	EP emprat per	UP usado por	UF use for

Más ejemplos de un término no aceptado en el término aceptado:

AM BNE	Formación del carácter USE Desarrollo de la personalidad Inversores USE Capitalistas y financieros Biología animal USE Zoología
---------------	---

Ved también

Para más información sobre sinónimos, ved el apartado 3 del módulo "Análisis de contenidos: resumen e indización" de esta signatura.

AM CSIC	Genios USE Superdotados Fortuna USE Patrimonio Casualidad USE Azar
LEMAC	Albins EM Albinisme Vegetals EM Plantes Llibres d'oracions EM Devocionaris
LCSH	IFLA USE International Federation of Library Associations

Los términos aceptados nos informan acerca de cuáles han sido los términos no escogidos para indizar el mismo concepto.

Ejemplos de términos aceptados y no aceptados

AM BNE	Fotografía USADO POR Fotos
AM CSIC	Sermones fúnebres USADO POR Oraciones fúnebres
LEMAC	Hel·lenisme EMPRAT PER Civilització hel·lènica
LCSH	Island animals USE FOR Insular animals

9.2. Relación de jerarquía

Las relaciones de jerarquía son de dos tipos: genéricas y específicas. Los encabezamientos pueden tener términos genéricos y términos específicos.

Los **términos genéricos** son aquellos que conceptualmente son más amplios que el encabezamiento aceptado.

Ejemplo

Cartas vemos que tiene como términos genéricos:

Ejemplo de términos genèrics

Cartas	Biografía Literatura
---------------	---------------------------------------

En cambio, los **términos específicos**, como su nombre indica, tienen un alcance conceptual más concreto que el del encabezamiento aceptado.

Ejemplo

Ejemplo de términos específicos

Cartas	Novela espistolar
---------------	--------------------------

Los encabezamientos tienen encabezamientos genéricos y específicos en función de su materia. Hay temas que por propio significado no tienen algunas de estas relaciones.

Ejemplo

El encabezamiento *Níquel* de la LEMAC tiene un solo término genérico, ninguno equivalente y ninguno específico.

Níquel Término genérico:
Metales de transición **Fuentes:**
LCSH (Nickel)
CSIC (Níquel)
TEI (Níquel)
Nº Registro: BNE19913257085

Habitualmente cada LEM marca los términos con el nombre completo de “genérico” o “específico”, pero a veces se utilizan siglas (las mismas que en los tesauros):

Términos genéricos, específicos y sus siglas

AM BNE	TG / TE	Término genérico / Término específico
AM CSIC	TG / TE	Térm. genérico / Térm. específico
LEMAC	TG / TE	Termes genèrics / Termes específics
LCSH	BT / NT	Broader terms / Narrower terms

Las relaciones de jerarquía son útiles en la indización porque muestran otras términos por encima y por debajo del encabezamiento que estamos consultando, ayudando al analista a indizar con el término más específico que convenga. En la recuperación son útiles porque los términos genéricos y específicos enriquecen la búsqueda.

Ejemplos

- En la indización: el analista tiene que indizar un documento sobre cartas noveladas y cree que el encabezamiento adecuado es *Cartas*, pero al consultar la LEM del CSIC se da cuenta de que el término **Novela epistolar** es más exacto.
- En la recuperación: el usuario busca información sobre *Cartas escritas por personajes famosos* y ve que también puede buscar por el nombre del personaje seguido de *Correspondencia*.

9.3. Relación asociativa

La estructura jerárquica que acabamos de ver es ideal para situar un concepto en su contexto de significado, la jerarquía lo sitúa de forma vertical (más grande que, menor que). Ahora bien, en la realidad los conceptos también se relacionan de manera horizontal, es por eso que existen las relaciones asociativas.

Lo que hacen las relaciones asociativas es poner en contacto dos términos o más que por cuestiones de significado se evocan uno a otro. No pertenecen a la misma secuencia jerárquica pero están relacionados. Mentalmente los asociamos. La relación es simétrica y consiguientemente recíproca.

Ejemplo

Los términos relacionados de *Cartas* son:

Cartas	Cartas - Redacción
---------------	---------------------------

La manera de indicar el tipo de relación es:

AM BNE	Término relacionado
AM CSIC	Véase además
LEMAC	Termes relacionats
LCSH	Related term / See also

Las funciones de las relaciones de asociación son sobre todo de ayuda en la recuperación, ya que asocian otros conceptos al encabezamiento de partida que pueden ser muy interesantes en la búsqueda.

Ejemplo

El encabezamiento **Maquillaje** tiene como dos términos relacionados **Belleza personal** y **Cosméticos**, ya que de alguna manera están relacionados con él aunque no pertenezcan a la misma cadena jerárquica.

9.4. Símbolos utilizados para designar los tres tipos de relaciones

Actualmente la mayoría de LEM que están en línea inscriben cada término dentro de la categoría que le corresponde y ésta es expresada en lenguaje natural siguiendo la terminología propia de los tesauros, adoptada por la LCSH en 1986. Pero no siempre ha sido así, antes se usaban unas siglas muy características que acto seguido reproducimos.

Símbolos utilizados para designar los tres tipos de relaciones

Relaciones de equivalencia	Antes de 1986	Después de 1986
a) De un término aceptado a uno no aceptado	x	EP (<i>emprat per</i>) UP (<i>usado por</i>) UF (<i>use for</i>)
b) De un término no aceptado a uno aceptado	v	ME (<i>empreu</i>) USE (<i>usar</i>) USE (<i>use</i>)

Relaciones de jerarquía	Antes de 1986	Después de 1986
a) Generales	xx	TA (<i>terme ampli</i>) TG (término general) BT (<i>broader term</i>)
b) Específicas	vt	TÉ (<i>terme específic</i>) TÉ (término específico) NT (<i>narrower term</i>)
Relaciones de asociación	Antes de 1986	Después de 1986
	Repetidos en xx y vt	TR (<i>terme relacionat</i>) TR (término relaciona- do) RT (<i>related term</i>)

10. Aprendiendo a indizar con una lista de encabezamientos

Como en todos los lenguajes, el primer proceso es el de **análisis del texto** para determinar la/las materias que trata, así como otros elementos decisivos como el periodo cronológico, el alcance geográfico y la forma bibliográfica en la que se presenta el documento.

Ved también

Podéis consultar el proceso de análisis del texto en el apartado 3 del módulo “Análisis de contenidos: resumen e indización” de esta asignatura.

Para indizar correctamente con una LEM hay que tener presente:

- Los principios de especificidad y economía.
- El orden de los subencabezamientos o subdivisiones.
- Las notas de alcance.
- Las obras de referencia: no todos los conceptos están representados en las listas, algunos hay que incorporarlos de alguna fuente reputada.
- Las reglas de construcción, que hemos visto anteriormente, sobre el uso del singular/plural, siglas, mayúsculas y minúsculas por si tenemos que introducir un término nuevo.

1) Principios de especificidad y economía

a) Principio de especificidad. Hay que indizar el encabezamiento más específico que encontremos. La especificidad está relacionada con la exactitud con que el encabezamiento describe el contenido. Si nos quedamos por encima (indizando con un término genérico) o por debajo (con un término específico) no estaremos siendo exactos.

Ejemplos del principio de especificidad

	Materia	Correcto	Incorrecto por	
			Genérico	Demasiado específico
AM BNE	Medicinas	Medicamentos	Productos sanitarios	Analgésicos
AM CSIC	Energías fósiles	Combustibles fósiles	Combustibles	Petróleo
LEMAC	El latín como lengua	Llatí	Lenguas clásicas	Manuscritos latinos

No podemos indizar con el término genérico y el específico al mismo tiempo.

Ejemplo

Combustibles fósiles / Petróleo

Llengües clàssiques / Llatí

Una buena manera de ser específicos es jugar con los subencabezamientos o subdivisiones.

Bibliotecas públicas - Libros seleccionados

Este principio de especificidad se puede adaptar a la situación del servicio de información y documentación (SID¹¹): si tiene pocos documentos sobre un tema, se recomienda indizar genéricamente; en cambio, si tiene muchos documentos, se recomienda indizar específicamente.

⁽¹¹⁾SID es la sigla de servicio de información y documentación.

b) Principio de economía: el encabezamiento tiene que sintetizar el contenido del documento. Un documento = Uno a tres encabezamientos.

El motivo de esta economía en las descripciones de la materia radica en los orígenes históricos del lenguaje. Tenemos que recordar que las listas de encabezamientos, como los sistemas de clasificación, nacieron en entornos manuales (no informatizados), donde cada documento se describía en fichas de cartulina a mano o con la máquina de escribir. Un documento tenía un mínimo de tres fichas de cartulina: una para el catálogo de autores, una para el de títulos y una para el de materias. Evidentemente, si el documento tenía más de una materia, había que hacer una cuarta ficha de cartulina. Para economizar esfuerzos, incluso, no se indizaba la materia de las novelas, ya que se acostumbraba a recuperar por el autor y título. De esta época manual, que duró hasta la década de 1980, se desprendió una política de actuación que para las bibliotecas públicas recomendaba entre una y tres entradas por materia como mucho.

Actualmente, y gracias a la informatización de los catálogos, el criterio predominante es la obra que se está indizando y cómo es de relevante para la búsqueda indizar uno, dos o más encabezamientos.

Si el documento trata un tema, no hay que indizar subtema por subtema.

Ejemplo

En una obra sobre cocina, no hay que indizar *Bocadillos*, *Entremeses*, *Salsas*, etc.; indizaremos *Cocina*, que es el epígrafe que lo engloba todo.

Si el documento trata temas múltiples que la LEM no recoge juntos, podemos indizar varios encabezamientos.

Ejemplo

Un documento sobre la monarquía hispánica en tiempos del Quijote se puede indizar con tres encabezamientos como los siguientes:

Cervantes Saavedra, Miguel de, 1547-1616. Don Quijote de la Mancha
Derecho - En la literatura
Derecho laboral

Si indizamos más de un encabezamiento, hay que ponerlos por orden de importancia en el documento.

2) Orden de los subencabezamientos o subdivisiones

Recordemos que muchos encabezamientos con subencabezamientos se encuentran contruidos en la misma lista, pero en caso de que haya que combinar alguna subdivisión, el orden es el siguiente:

Materia principal + subencabezamiento de materia + de lugar + de tiempo + de forma

Si el primer lugar es un nombre geográfico, entonces el orden es:

Lugar + materia + tiempo + forma

3) Las notas de alcance

Cuando el sentido de un encabezamiento no está suficientemente claro, las listas añaden una explicación asociada que puede ser de diversos tipos:

a) Nota histórica

Explicación del CSIC para el encabezamiento *Yugoslavia*

Úsese para las obras que tratan sobre los territorios que formaban la **Yugoslavia** entre 1918 y 1992 (entre 1946 y 1991 compuesta por las repúblicas federadas de Bosnia-Herzegovina, Croacia, Macedonia, Eslovenia, Montenegro y Serbia), o los de la República Federal de **Yugoslavia**, reducida a Serbia y a Montenegro entre 1992 y 2003.

b) Definición

Nota de la AM CSIC para la definición de *Waqf*

Waqf

Úsese para las obras tanto las que que tratan del waqf, institución del derechos islámico con fondos inmovilizados dedicados a una obra piadosa o de interés general, como para las realizaciones urbanísticas, artísticas o sociales debidas a estas fundaciones.

c) Nota de aplicación

Nota

Si el servicio de información y documentación (SID) necesita una indización más profunda, se recomienda indizar con un tesoro, que es otro lenguaje documental natural pero postcoordinado y que tiene más prestaciones para hacer análisis que las LEM.

Nota del CSIC para explicar en qué casos se puede indizar con el encabezamiento *Recursos energéticos*

Úsese para las obras sobre todas las formas posibles de **energía** mecánica. Para los estudios sobre la **energía** desde los puntos de vista físico o de la ingeniería úsese: **Energía** mecánica.

d) Nota informando otros términos y subdivisiones.

Nota de la AM CSIC explicando otros encabezamientos más específicos a partir de *Filosofía*

Pueden usarse además los nombres de las distintas partes de la filosofía, así como los nombres de las escuelas, movimientos, etc. Úsese además como subencabezamiento detrás de nombres de personas, excepto filósofos, y de materias a propósito de la teoría fundamental de estas materias

4) Las obras de referencia

No todos los conceptos están representados en las listas. Algunos hay que incorporarlos de alguna fuente reputada:

- Otras listas de encabezamientos, como la LCSH o RAMEAU.
- Diccionarios y enciclopedias considerados como autoridad dentro del dominio en cuestión. La LEMAC considera autoridades la Enciclopèdia y el Atlas del grupo Enciclopèdia Catalana, el Diccionari de la llengua catalana de l'Institut d'Estudis Catalans y el Wordreference como traductor.
- Organismos normalizadores como el Termcat.
- Tesoros, particularmente los contruidos conforme la norma ISO 2788 Directrices para el establecimiento y desarrollo de tesauros monolingües e ISO 5964 Directrices para el establecimiento y desarrollo de tesauros multilingües.
- Sistema de clasificación. El que se use en el mismo SID: CDU o LCC.
- Especialistas en el tema.

5) Las reglas de construcción

Las reglas que hemos visto anteriormente sobre el uso del singular/plural, siglas, mayúsculas y minúsculas por si tenemos que introducir un término nuevo.

Ejemplo

Supongamos que tenemos que indizar OCDE y no existe en la LEM. Como sabemos que se prefiere el nombre entero a las siglas, indizaremos **Organización de Cooperación y Desarrollo Económico** y lo haremos en la lengua del catálogo.

11. La lista de autoridades

Una **lista de autoridades** es un lenguaje documental controlado que recoge ordenados alfabéticamente todos los registros de autoridades presentes en el catálogo o en la base de datos bibliográfica de la unidad de información.

Ved también

Las reglas de construcción de cada registro de autoridad se estudian en la asignatura *Análisis documental*.

Las autoridades recogidas en este apartado son:

- Nombres propios
- Congresos
- Títulos
- Nombres propios y títulos
- Entidades
- Nombres geográficos

La descripción de cada autoridad, con los términos descartados, referencias, las notas de aplicación y la fuente se conoce como **registro de autoridad**. La suma de todos los registros se llama **fichero de autoridades** o **lista de autoridades**. Si este fichero se encuentra vinculado al catálogo bibliográfico, se conoce como **catálogo de autoridades**. Las autoridades también son conocidas como encabezamientos e identificadores.

La principal función de las listas de autoridades es conseguir la univocidad de los nombres propios, títulos y demás autoridades y garantizar una buena recuperación documental. Las listas de autoridades cuidan de los índices que se usan en los catálogos bibliográficos. Concretamente:

- Garantizan la normalización y la univocidad de los puntos de acceso.
- Facilitan la consistencia en su aplicación en el catálogo bibliográfico.
- Generan la red de relaciones de *ved* y *ved también*.
- Asisten al usuario en las búsquedas.
- Permiten el intercambio entre los registros bibliográficos de los diversos catálogos.

11.1. Las listas de autoridades en Cataluña y España

Todas las **bibliotecas y centros de documentación** indizan con un catálogo de autoridades para controlar los nombres propios, geográficos, títulos, etc. Forma parte del compromiso de representar los documentos de manera unívoca.

Al igual que los encabezamientos de materia vistos anteriormente, estas autoridades se encuentran en la misma dirección. Sólo hay que cambiar en el desplegable la opción *Materia* por las otras autoridades.

Listas de autoridades

Catalán	LENOTI	Lista de encabezamientos de nombres y títulos de la Biblioteca Nacional de Catalunya
Castellano	CA BNE CA CSIC	Catálogo de autoridades de la Biblioteca Nacional de España Autoridades de la red de bibliotecas del CSIC
Inglés	LC Authorities	Library of Congress Authorities
Francés	LAVAL RAMEAU	Lista Laval de la Universidad Laval (Canadá) Autorités RAMEAU

Las autoridades en **documentación archivística**, en cambio, hacen referencia a entidades: instituciones, personas y familias.

La norma utilizada para elaborar los registros de autoridades es la ISAAR(CPF) *Norma internacional sobre los registros de autoridad de archivos relativos a instituciones, personas y familias* del Consejo Internacional de Archivos. Esta norma ISAAR(CPF) es complementaria a la norma ISAD(G)¹² *Norma Internacional General de Descripción Archivística*.

Las autoridades en archivos son específicas, ya que son nombres que no se encuentran fácilmente en otros catálogos de autoridades.

El control de las autoridades en archivos permite relacionar los fondos de un mismo productor (persona o entidad) que se encuentre en diversos archivos y también vincular estos nombres con otros recursos bibliográficos o museográficos.

Los registros de autoridad de archivos son semejantes a las autoridades de bibliotecas, ya que los dos permiten la creación de puntos de acceso normalizados en la descripción. Pero los de archivos tienen que responder a unos requisitos adicionales consistentes en documentar toda la información disponible sobre el productor y el contexto de producción del documento. Los registros de autoridad archivísticos acostumbran a tener más información que los bibliotecarios.

En esta asignatura trabajaremos con ficheros de autoridades ya construidos, no con normas para construirlos.

Al igual que en las listas de encabezamientos de materia, trabajaremos con las listas de autoridades que por idioma nos son más próximas: LENOTI, el catálogo de autoridades de la Biblioteca Nacional de España (CA BNE) y el catálogo de las bases de datos del CSIC (CA CSIC).

⁽¹²⁾ISAD(G) es la sigla de *General International Standard Archival Description*, Norma Internacional General de Descripción Archivística.

Autoridades específicas

Son ejemplos de autoridades específicas la familia de los Marqueses de Castellodorsius (cuyo fondo histórico está en el Archivo Nacional de Cataluña), instituciones históricas o el nombre del director general de la empresa X.

Los archivos de los registros de autoridad

Ved el registro de autoridad para la institución "Consejo de guerra", ejemplo disponible en la norma ISAAR pág. 45-52.

Las autoridades que ofrecen en cada desplegable son:

1) Biblioteca de Cataluña

Listas de autoridades de la BC

CANTIC ¹³	Nombres de personas y de familias Nombres de entidad Nombres de congreso Títulos uniformes Combinaciones de nombre-título Se excluyen: Encabezamientos temáticos Nombres geográficos Términos de género/forma Nombres con subdivisión de materia
LENOTI ¹⁴	Nombre Nombre-título Título
LEMAC	Materia

⁽¹³⁾CANTIC es un catálogo de autoridades cooperativo iniciado en 2007 que se realiza en el Catálogo Colectivo de las Universidades de Cataluña (CCUC) y está liderado por la Biblioteca de Cataluña.

⁽¹⁴⁾Lista de encabezamientos de nombres y títulos de la Biblioteca de Cataluña (LENOTI).

⁽¹⁵⁾En Cataluña las obras de referencia son de la editorial Enciclopèdia Catalana: la *Gran Enciclopèdia Catalana* y el *Atles*.

Como se puede observar faltan las autoridades para nombres geográficos. Los nombres geográficos jurisdiccionales se construyen según el capítulo 23 de las AACR2R. Los no jurisdiccionales se sacan de alguna lista de autoridades reputada¹⁵ por el SID.

2) Biblioteca Nacional de España

Listas de autoridades de la BNE

Catálogo de autoridades	Persona Título / Título mapa Congreso Entidades Materias Geográficos Subencabezamientos de materia
--------------------------------	--

3) Bibliotecas del CSIC

Listas de autoridades del CSIC

Catálogo CSIC Autoridades	Autor personal Institución Congreso Materia Término geográfico Título uniforme Subencabezamiento
----------------------------------	--

4) Autoridades en inglés y francés

Consultaremos también autoridades en otras lenguas, como la inglesa o francesa, para facilitar la búsqueda en catálogos internacionales.

- LC Authorities.

- Lista Laval de la Universidad Laval, Canadá (LAVAL)
- Autorités RAMEAU

11.2. Tipología

Las listas de autoridades son un lenguaje documental que tiene las siguientes características: es analítico, natural, controlado y postcoordinado.

1) **Analítico.** Se trata de un lenguaje documental analítico porque permite, en la indización y también en la recuperación, el uso de más de una “autoridad” (o identificador).

El contenido de un documento se puede representar por tantos descriptores controlados como sean pertinentes y necesarios para establecer un buen diálogo documental con el sistema.

Ejemplo con LENOTI

Un documento del tipo “Gironins il·lustres” que tratase de Salvador Dalí, Narcís Monturiol, Pep Ventura, Josep Pla y Xavier Cugat se podría indizar con las autoridades de estos cinco personajes.

Dalí, Salvador, 1904-1989
Monturiol, Narcís
Ventura, Pep
Pla, Josep, 1897-1981
Cugat, Xavier

2) **Natural.** Se trata de un lenguaje documental natural porque la indización se basa en el uso de palabras y/o expresiones libres que pertenecen al lenguaje o discurso común y que se encuentran en los mismos documentos (título, resumen, texto, etc.) y en los conceptos de la consulta del usuario (nombres de personajes, poblaciones, territorios, entidades, etc.). Los nombres no se traducen en ningún código.

3) **Controlado.** Se trata de un lenguaje controlado porque los términos que lo componen han sido depurados de las ambigüedades del lenguaje natural y están destinados a representar de manera unívoca el contenido de un documento y de las consultas dentro de un sistema documental.

Es un lenguaje documental con voluntad de univocidad, que busca eliminar los fenómenos de sinonimia, polisemia y homonimia; por eso exige establecer el control sobre todos los términos que reúne.

Ejemplo en LENOTI

Hay dos Josep Pla con las dos autoridades correspondientes; una corresponde al escritor y la otra al compositor. La polisemia del nombre se evita con las fechas de nacimiento y muerte del personaje:

Pla, Josep, 1728-1762: oboísta y compositor catalán.
Pla, Josep, 1897-1981: el escritor y periodista ampurdanés.

Ved también

Estas características están basadas en las tipologías que vimos en el apartado 5 del módulo “Análisis de contenidos: resumen e indización”.

En el caso de los sinónimos, las autoridades de nombre, nombre geográfico y título siguen las mismas pautas que los encabezamientos de materia vistos anteriormente, ya que el punto de partida son las Reglas angloamericanas de catalogación.

Ejemplo

Si buscamos *URSS* en LENOTI, encontraremos que la autoridad aceptada es:

```
Encapçalament acceptat
Unió de Repúbliques Socialistes Soviètiques
Emprat per
URSS
```

4) Postcoordinado. Se trata de un lenguaje documental postcoordinado ya que permite la coordinación de conceptos en el momento de la recuperación y la utilización de un gran número de términos de indización o puntos de acceso. La combinación precisa de estos diferentes descriptores en la recuperación es la que nos remite al documento buscado.

Ejemplo

Siguiendo con el ejemplo anterior, “Gerundenses ilustres”, la coordinación de términos se produciría en el momento de la recuperación, integrando en la busca 2 o más de estas autoridades (en este caso, personajes):

Ir a *Cerca avanzada* y pedir:

```
Matèria: Monturiol, Narcís
And
Matèria: Ventura, Pep
```

5) Combinatorio. Se trata de un tipo de lenguaje de estructura combinatoria o asociativa, donde los conceptos o descriptores se organizan de manera independiente y sólo se produce la combinación o intersección en las operaciones de indización y recuperación.

Se trata de un lenguaje en el que los términos que lo componen se organizan en una lista por orden alfabético sin ningún otro tipo de estructura que sitúe a cada descriptor, identificador o autoridad en una cadena lógica, jerárquica y estructurada de conceptos.

Ejemplo de autoridades de la CA BNE, letra A

Aaberg, Nils 1888-1957
Aaker, David A.
Aaliyah 1979-2001
Aalto, Alvar 1898-1976
Aalto, Hugo Alvar Henrik 1898-1976
Aapro, M.
Aapro, Matti
A. A. P. y G. 1745-1792
Aarnio, Aulis
Aaron 1964-
Aaron, Bob
Aaron, Jack
Aaron, Jane
Aaron, Javier
Aaron, Lee 1962-

11.3. Indizar con una lista de autoridades

Indizar con una lista de autoridades es parecido a indizar con una lista de encabezamientos de materia: hay que consultar las listas y saber usar las referencias que nos vinculan un término no aceptado con el aceptado.

Pero a diferencia de las LEM, al ser un lenguaje analítico y postcoordinado, nos permite indizar más de una autoridad por documento.

Cuando la autoridad se combina con un encabezamiento de materia, tenemos que leer las notas de aplicación que nos indicarán si podemos hacer el encabezamiento.

Ejemplo

Supongamos un documento sobre las diversas películas que se han hecho sobre *El retrato de Dorian Gray*, la obra de Oscar Wilde. Para indizar este documento consultaremos dos listados:

- 1) Uno de encabezamientos de materias para el concepto películas sobre obras literarias. En LEMAC sería el encabezamiento **Adaptaciones cinematográficas**.
- 2) Uno de títulos uniformes para la obra "Retrato de Dorian Gray". En CA BNE es **The picture of Dorian Gray**.

Dado que Adaptacions cinematogràfiques lleva una nota de alcance que dice que hay que combinarlo detrás del nombre de un autor u obra, haremos el siguiente encabezamiento:

The picture of Dorian Gray - Adaptaciones cinematográficas

Como lenguaje controlado, las listas de autoridades tienen ventajas e inconvenientes en la indización y en la recuperación.

1) En la indización

a) Ventajas:

- Son listados finitos, controlados y elaborados a priori. En principio, cada nombre se expresa con un único descriptor; se búsqueda eliminar la ambigüedad presente en el lenguaje natural. Es unívoco.
- Se trata de un lenguaje natural, no codificado, de dominio común y constituido por elementos léxicos del lenguaje usual.
- Permite un elevado grado de especialización y, por lo tanto, de especificidad en la indización.

b) Inconvenientes:

- El analista tiene que traducir del lenguaje natural al lenguaje documental.
- Más caro económica y humanamente de construcción y mantenimiento que un listado de palabras clave y que un listado de descriptores libres.
- Limitado al vocabulario recogido en el lenguaje.
- Hace falta haber creado este lenguaje antes de empezar la indización.

2) En la recuperación**a) Ventajas:**

- La postcoordinación permite utilizar más de un término de acceso al documento, facilitando así esa tarea.
- Es unívoco.
- Las estrategias de búsqueda suelen ser muy simples.

b) Inconvenientes:

- Se pueden dar casos de falsa coordinación.
- El grado de actualización del vocabulario es limitado.

Actividades

Proponemos las siguientes actividades para practicar los conocimientos de este módulo:

1. Crea el encabezamiento para los documentos siguientes con las listas de encabezamientos de materia y autoridades de la Biblioteca de Cataluña, de la Biblioteca Nacional de España o del CSIC.
 - a) Abecedario griego.
 - b) Aire acondicionado de los coches.
 - c) Antología de poesía siciliana.
 - d) Argot ruso.
 - e) Becas de estudios para aprender italiano.
 - f) Bibliografía sobre judíos sefardíes.
 - g) Bibliografía sobre gobierno municipal.
 - h) Bosques de la URSS.
 - i) Ciclámenes.
 - j) Congresos sobre áreas suburbanas.
 - k) Conservación del papel.
 - l) Contaminación ambiental por níquel.
 - m) Declinaciones latinas.
 - n) Documento sobre becas de investigación sobre los efectos fisiológicos del alcohol.
 - o) Documento sobre el Siglo de Oro de la literatura española.
 - p) Documento sobre la obra de García Lorca *La casa de Bernarda Alba*
 - q) Enseñanza del latín en las universidades.
 - r) Ikebana.
 - s) Novelas de Paul Auster.
 - t) Predicción de la suerte en el siglo XX.

Reflexión

Las soluciones de las actividades se tratarán en el tablón de la asignatura.

Glosario

AM BNE *f* Acrónimo usado en este material para las Autoridades de materia de la Biblioteca Nacional de España.

AM CSIC *f* Acrónimo usado en este material para las Autoridades de materia de la red de bibliotecas del CSIC.

autoridad de materia *f* Ved **encabezamiento**.

autoridades *f* Formas normalizadas de las entradas de un catálogo bibliográfico o inventario de archivo para: nombres propios, títulos, nombres propios y títulos, materias, entidades, nombres geográficos.

BT Inglés. Siglas de *Broad term*, usadas en LEM y tesauros, para designar los términos genéricos en la relación de jerarquía. Equivalencias en otras lenguas: catalán TA, castellano TG. Ved también **relación de jerarquía**.

CA BNE *m* Acrónimo de Catálogo de autoridades de la Biblioteca Nacional de España.

CA CSIC *f* Acrónimo de Catálogo de autoridades de la red de bibliotecas del CSIC.

CANTIC *m* Catálogo de autoridades cooperativo iniciado en el año 2007 que se realiza en el Catálogo Colectivo de las Universidades de Cataluña (CCUC) y está liderado por la Biblioteca de Cataluña.

Cutter, Charles Ammi *m* Bibliotecario jefe del Ateneo de Boston. Autor de las *Rules for a printed dictionary catalogue* donde se exponen por primera vez las reglas para encontrar e indizar un documento según su temática usando encabezamientos de materia. Cutter elaboró los principios que rigen las LEM como el principio de especificidad y el de entrada directa, todos ellos vigentes en la actualidad. También formuló el principio de economía, las referencias entre términos aceptados y no aceptados, ya sean sinónimos u homónimos, los encabezamientos simples y compuestos, los signos de puntuación, la sintaxis a la hora de construir encabezamientos compuestos etc.

EM Catalán. Siglas de EMPREU usadas en LEM y tesauros para designar la relación de equivalencia entre un término no aceptado y otro aceptado. Equivalencias en otras lenguas: castellano USE, inglés USE. Ved también **relación de equivalencia**.

EM BUS Acrónimo de Encabezamientos de materia de la Universidad de Sevilla.

encabezamiento *m* Cada una de las entradas o autoridades de una lista de encabezamientos de materia. Término de indización propio del lenguaje documental LEM. Palabras o frases específicas que describen la materia y representan conceptos. Están redactados en lenguaje natural pero son términos controlados, es decir, han controlado la ambigüedad derivada de la sinonimia y homonimia.

encabezamientos compuestos *m* Conceptos expresados con más de una palabra.

encabezamientos simples *m* Conceptos expresados con una sola palabra, generalmente un sustantivo, aunque también pueden ser adjetivos, verbos sustantivados o nombres geográficos. Tanto pueden ser términos en singular como en plural.

EP Catalán. Siglas de EMPRAT PER usadas en LEM y tesauros para designar la relación de equivalencia entre un término aceptado y otro no aceptado. Equivalencias en otras lenguas: castellano USADO POR, inglés USE FOR. Ved también **relación de equivalencia**.

fichero de autoridades *m* Fichero que recoge todos los registros de autoridades. Si este fichero se encuentra vinculado al catálogo bibliográfico, se conoce como catálogo de autoridades.

ISAAR(CPF) *f* Norma para la elaboración de autoridades en archivos. *Norma internacional sobre los registros de autoridad de archivos relativos a instituciones, personas y familias* del Consejo internacional de archivos.

LC Authorities *f* Acrónimo de Library of Congress Authorities.

LCSH Acrónimo de Library of Congress Subject Headings.

LEMAC *f* Acrónimo de Lista de encabezamientos de materia en catalán.

LEM BP *f* Acrónimo de la Lista de encabezamientos de materia para bibliotecas públicas.

LENOTI *f* Acrónimo de la Lista de encabezamientos de nombres y títulos de la Biblioteca Nacional de Cataluña.

lista de autoridades *m* Lenguaje documental controlado que recoge ordenados alfabéticamente todos los registros de autoridades presentes en el catálogo o base de datos bibliográfica de la unidad de información. Es una lista de frases uniformes que se usan para indizar el autor y el título del documento. Es un lenguaje analítico, natural, controlado, postcoordinado y combinatorio.

lista de encabezamientos de materia *f* Lenguaje documental controlado que recoge ordenadas alfabéticamente todas las materias o temas susceptibles de representar el contenido de un fondo documental. Es una lista, un léxico de palabras y frases uniformes que se usan para indizar la materia del documento. Es un lenguaje sintético, natural, controlado, precoordinado y combinatorio.

NT Inglés. Siglas de Narrow Term, usadas en LEM y tesauros para designar los términos específicos en la relación de jerarquía. Equivalencias en otras lenguas: catalán TE, castellano TE. Ved también **relación de jerarquía**.

obras de referencia *f* Fuentes de información que se usan como material de consulta, para obtener una información breve, rápida y exacta. En el contexto de los lenguajes documentales se consultan cuando un concepto no está representado en la lista de encabezamientos y hay que incorporarlo de nuevo. Son obras de referencia otras listas de encabezamientos, diccionarios y enciclopedias (considerados como autoridad dentro del dominio en cuestión), organismos normalizadores, tesauros, sistema de clasificación y especialistas en el tema.

principio de consistencia *m* Principio de construcción por el cual los encabezamientos nuevos que se incorporen a una LEM tienen que mantener la coherencia con los existentes, en la forma y estructura sintáctica.

principio de denominación *m* Principio de construcción por el cual los nombres propios, de persona y de lugar tienen que mantener también la coherencia con los ya existentes.

principio de economía *m* Principio de aplicación que recomienda sintetizar el contenido del documento en uno, dos o tres encabezamientos.

principio de encabezamiento uniforme *m* Principio de construcción por el cual cada materia tiene que ser representada siempre con un único encabezamiento y cada encabezamiento tiene que representar una única materia.

principio de especificidad *m* Principio de aplicación que recomienda indizar con el encabezamiento más específico que encontremos. La especificidad está relacionada con la exactitud con que el encabezamiento describe el contenido.

principio del usuario *m* Principio de construcción por el cual el vocabulario de las LEM se tiene que adecuar al usuario al que va destinada la lista. El lenguaje utilizado en los encabezamientos tiene que reflejar las características de los usuarios. Las expresiones específicas o muy académicas se reservan para listas especializadas.

principio lingüístico *m* Principio de construcción por el cual los términos de la lista tienen que estar en el idioma del catálogo.

principio de semántica *m* Principio de construcción por el cual los encabezamientos se organizan en una estructura semántica, es decir, en relaciones de equivalencia, jerarquía y asociación.

principio de sinonimia y principio de homonimia *m* Principio de construcción por el cual se controlan la sinonimia y la homonimia propias del lenguaje natural. Las LEM controlan los sinónimos mediante la referencia de *Empreu* o *Use*. Las LEM controlan los polisémicos y homónimos de diversas maneras: adjetivando, poniendo paréntesis con el nombre de la disciplina, haciendo frases, jugando con el singular y el plural, etc.

principio de sintaxis *m* Principio de construcción por el cual las LEM establecen las reglas de combinación de los encabezamientos con los subencabezamientos.

RAMEAU *m* Acrónimo de *Répertoire d'autorité matière encyclopédique et alphabétique unifié*.

registro de autoridad *m* Descripción de cada autoridad. Registro que anota la forma aceptada, los términos no seleccionados, notas de aplicación, relaciones semánticas si es necesario y la fuente en la que se ha basado.

regla de las siglas *f* Regla de control del vocabulario por la cual se prefiere la expresión completa a las siglas o acrónimos de un concepto o entidad.

regla del singular/plural *f* Regla de control del vocabulario. No hay una normativa clara para el uso del singular/plural. Algunos autores apuntan que el plural se usa para temas concretos y contables (enciclopedias, museos) y el singular para temas abstractos e incontables (belleza, positivismo, etc.).

regla de los signos de puntuación *f* Regla de control del vocabulario. Las LEM utilizan diversos signos y entre ellas no hay una unidad de criterio, por lo tanto se hace necesario consultar las introducciones y manuales adjuntos a cada lista.

regla de mayúsculas/minúsculas *f* Regla de control del vocabulario por la cual se recomienda que la primera palabra del encabezamiento, del subencabezamiento, del encabezamiento cambiado y la que va dentro de paréntesis comience por mayúscula.

relación asociativa *f* Relación semántica que pone en contacto dos términos o más que por cuestiones de significado se evocan uno a otro. No pertenecen a la misma secuencia jerárquica pero están relacionados. Mentalmente los asociamos. La relación es simétrica y por consiguiente recíproca. Las funciones de las relaciones de asociación son sobre todo de ayuda en la recuperación, ya que asocian otros conceptos al encabezamiento de partida que pueden ser muy interesantes en la búsqueda.

relación de equivalencia *f* Relación semántica que une los términos aceptados con los no aceptados. La función de estas relaciones es hacer de puente entre las palabras utilizadas en lenguaje natural y las utilizadas en un lenguaje controlado, tanto en la indización como en la recuperación. En la indización este puente permite mantener la coherencia y la univocidad en los indizadores y en la recuperación permite a los usuarios recuperar mejor los documentos, ya que no se produce silencio ni ruido documental. No es posible recuperar por un término no aceptado, ya que el mismo programa que mantiene la lista hará la remisión al término aceptado gracias a estas referencias. Toman la forma de EMPREU/USE.

relación de jerarquía *f* Relación semántica que dispone los encabezamientos en una secuencia lógica de términos genéricos (más grande que) y específicos (menor que). Las relaciones de jerarquía son útiles en la indización porque muestran otros términos por encima y por debajo del encabezamiento que estamos consultando, ayudando al analista a indizar con el término más específico que convenga. En la recuperación son útiles porque los términos genéricos y específicos enriquecen la busca.

RT Inglés. Siglas de Related Term, usadas en LEM y tesauros para designar la relación de asociación entre dos encabezamientos. Equivalencias en otras lenguas: catalán TR, castellano TR. Ved también **relación de asociación**.

RVM *m* Acrónimo de **Répertoire de vedettes-matière de la Universidad de Laval**.

subencabezamientos *m pl* Palabras que van detrás de un encabezamiento a efectos de matizar y precisar el significado. Son encabezamientos situados detrás de otro encabezamiento. Hay 4 tipos de subencabezamientos: los de materia, los geográficos, los cronológicos y los de forma.

subencabezamientos de forma *m pl* Subencabezamientos que se usan para representar la forma en que se presenta el documento, son formas del tipo enciclopedias, estadísticas, anuarios, biografías, etc.

subencabezamientos de lugar *m pl* Subencabezamientos que se usan para representar nombres de lugares geográficos y precisar el encabezamiento principal.

subencabezamientos de materia *m pl* Subencabezamientos que se usan para precisar el contenido del encabezamiento principal y para representar la forma en que se presenta el contenido o el punto de vista.

Subencabezamientos de tiempo *m* Subencabezamientos que se usan para representar periodos, fechas concretas, siglos.

TA Catalán. Siglas de *terme ampli*, usadas en LEM y tesauros para designar los términos genéricos en la relación de jerarquía. Equivalencias en otras lenguas: castellano TG, inglés BT. Ved también **relación de jerarquía**.

TE Catalán y castellano. Siglas de *terme específic / término específico*, usadas en LEM y tesauros para designar los términos específicos en la relación de jerarquía. Equivalencias en otras lenguas: inglés NT.

Ved también **relación de jerarquía**.

términos aceptados *m pl* Encabezamientos que se pueden utilizar para indizar. Acostumbran a ir en negrita.

términos específicos *m pl* Encabezamientos que conceptualmente son más concretos y específicos que el encabezamiento aceptado que estamos consultando. En la cadena jerárquica se encuentran situados por debajo del encabezamiento.

términos genéricos *m pl* Encabezamientos que conceptualmente son más amplios que el encabezamiento aceptado que estamos consultando. En la cadena jerárquica se encuentran situados por encima del encabezamiento.

términos no aceptados *m pl* Términos no admitidos para indizar. Son sinónimos, siglas, términos demasiado especializados, barbarismos, formas incorrectas, etc. Su función es controlar todas las posibles variantes y ofrecer una forma de encabezamiento única para indizar que será el término aceptado.

TG Castellano. Siglas de Término genérico, usadas en LEM y tesauros para designar los términos genéricos en la relación de jerarquía. Equivalencias en otras lenguas: catalán TA, inglés BT.

Ved también **relación de jerarquía**.

título uniforme *m* Título ficticio dado por el analista para agrupar todas las entradas en un catálogo de una obra. El título uniforme se crea para controlar las diferentes ediciones, traducciones, etc.

TR Catalán y castellano. Siglas de *terme relacionat / término relacionado*, usadas en LEM y tesauros para designar la relación de asociación entre dos encabezamientos. Equivalencias en otras lenguas: inglés RT.

Ved también **relación de asociación**.

UF Inglés. Siglas de USE FOR, usadas en LEM y tesauros para designar la relación de equivalencia entre un término aceptado y otro no aceptado. Equivalencias en otras lenguas: catalán EMPRAT PER, castellano USADO POR.

Ved también **relación de equivalencia**

UP Castellano. Siglas de USADO POR, usadas en LEM y tesauros para designar la relación de equivalencia entre un término aceptado y otro no aceptado. Equivalencias en otras lenguas: catalán EMPRAT PER, inglés USE FOR. Ved también **relación de equivalencia**.

USE Castellano e Inglés. Instrucción USE (o úsese) utilizadas en LEM y tesauros para designar la relación de equivalencia entre un término no aceptado y uno aceptado. Equivalencias en otras lenguas: catalán EMPREU. Ved también **relación de equivalencia**.

v Símbolo usado en las LEM, antes de la adopción internacional de las siglas de tesauros, para hacer la relación de equivalencia entre un término no aceptado y otro aceptado. Actualmente equivale a EMPREU/USE/USE. Ved también **relación de equivalencia**.

Vt Símbolo usado en las LEM, antes de la adopción internacional de las siglas de tesauros, para designar los términos específicos. Actualmente equivale a TE/TE/NT. Ved también **relación de jerarquía**.

x Símbolo usado en las LEM, antes de la adopción internacional de las siglas de tesauros, para designar los términos no aceptados. Actualmente equivale a EMPREU PER / USAR POR / USE FOR. Ved también **relación de equivalencia**.

Xx Símbolo usado en las LEM, antes de la adopción internacional de las siglas de tesauros, para designar los términos genéricos. Actualmente equivale a TA/TG/BT. Ved también **relación de jerarquía**.

Bibliografía

Biblioteca de Catalunya. *Llista d'encapçalaments de matèria en català* [disponible en línea (consulta 01-09-2009)].

<http://www.bnc.cat/lemac/principis>

Caro, C.; Travieso, C. "Encabezamientos de materia en las bibliotecas españolas: perspectiva histórica y situación actual". En: *Tendencias de investigación en organización del conocimiento*.

Concrecions a les AACR2 (Part II, encabezamientos, títulos uniformes, referencias) (2002). Recopiladas por el Grup de Treball para las Concrecions a les AACR2 con la coordinación del Servei de Normalització Bibliogràfica. Barcelona: Biblioteca de Catalunya. (Actualizació 2003.)

Chan, Louis Mai (1978). *Library of Congress: Subject Headings: principles and application*. Littleton, Colorado: Libraries Unlimited.

Gil Leiva, I. (2008). *Manual de indización. Teoría y práctica*. Gijón: Trea.

Gil Urdiciain, B. (2004). *Manual de lenguajes documentales*. Gijón: Trea.

Informe per a la creació de la llista d'autoritats de Catalunya de noms i títols: requeriments i funcionalitats (Octubre de 2003). Redactado por la Subcomissió Tècnica per a l'Elaboració dels Requeriments i Funcionalitats de la Llista d'Autoritats de Catalunya per a la Comissió Assesora de Catalogació.

MARC 21 concise authority format (2002). Washington, D.C.: Library of Congress.

Regles Angloamericanes de Catalogació (1996). Barcelona, Biblioteca de Catalunya: Eumo.

Universidad de Laval (2005). *Répertoire des vedettes-matière*. [S.L]: Universidad de Laval. Biblioteca Universitaria.

Los tesauros

Damián Martínez Ferreras

PID_00143963



Universitat Oberta
de Catalunya

www.uoc.edu



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. Fundació 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

NA Úsese 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.

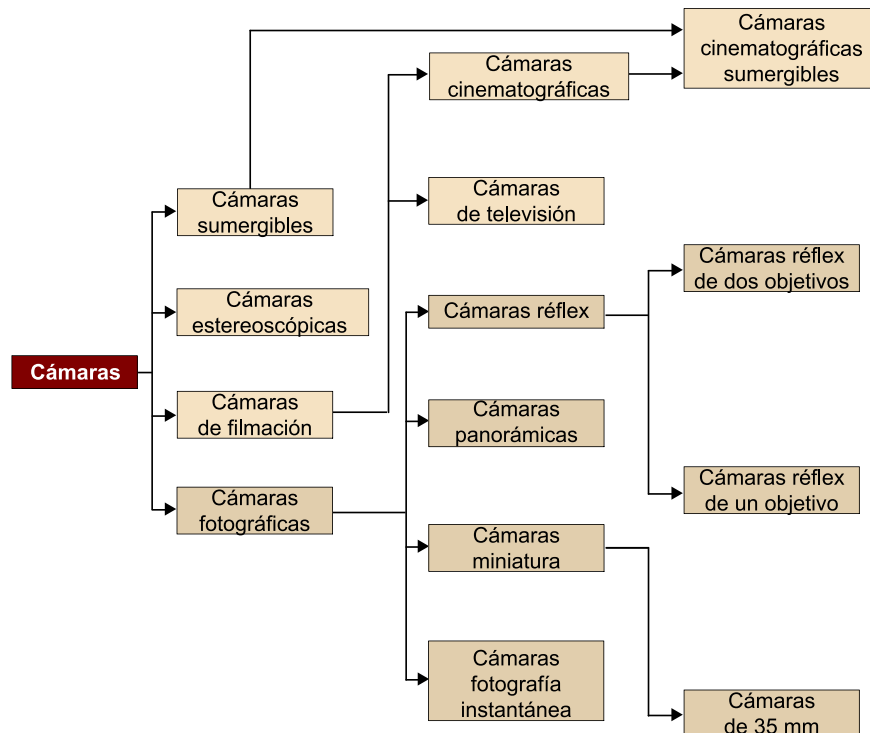
1) Estructura arborescente

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).

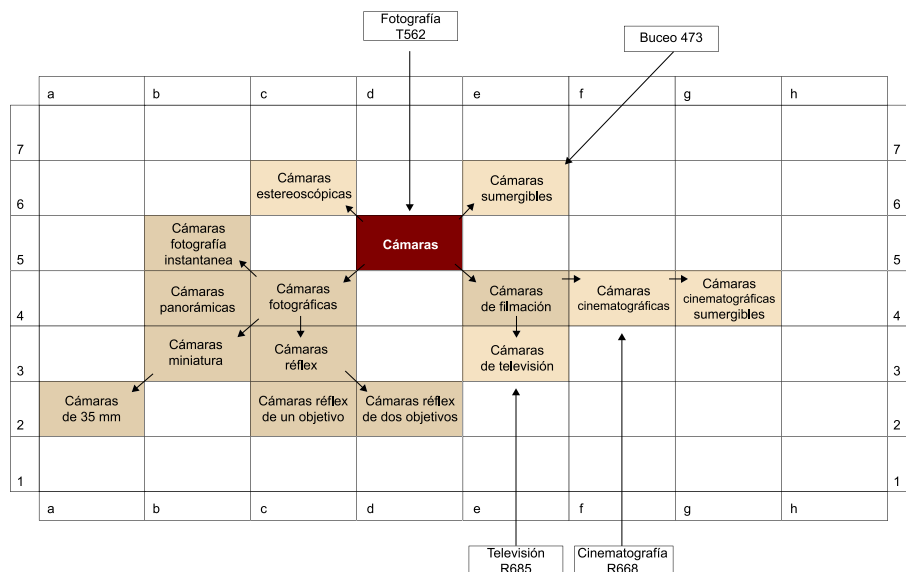
2) Diagrama de flechas

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 tesauro 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 tesauro 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** electrodico 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** electrodico 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 (descriptores, no descriptores, 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 multidisciplinares 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)

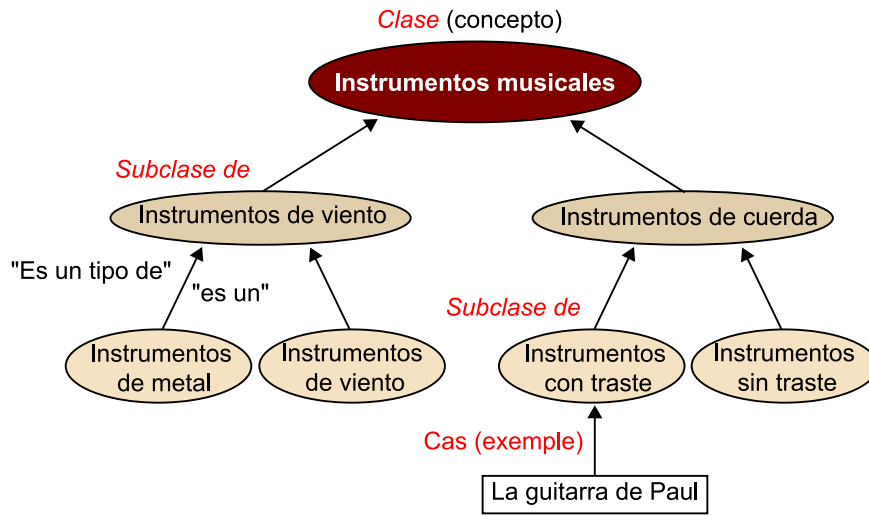
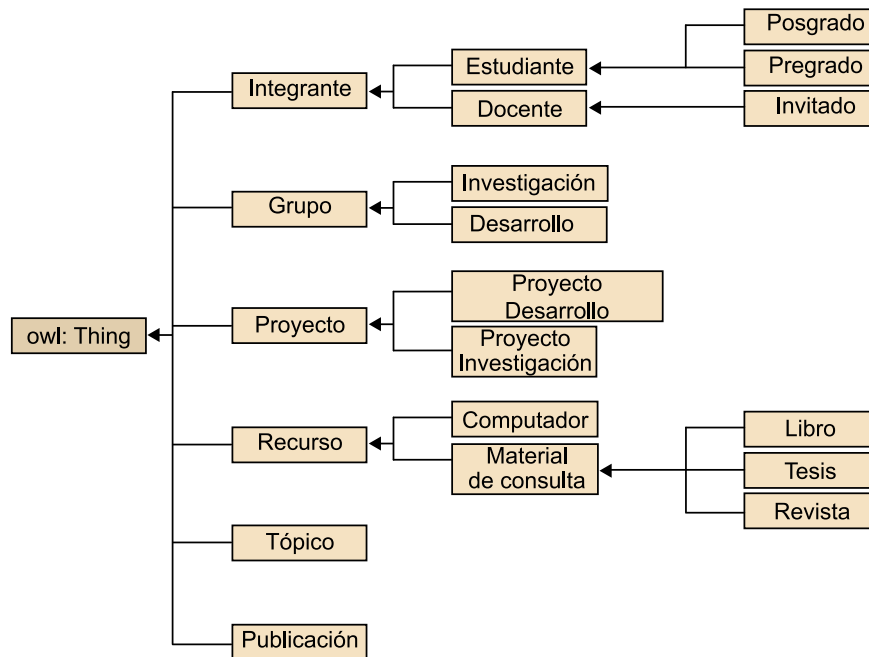


Figura 4. Ejemplo de ontología para laboratorios de investigación y desarrollo



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.⁵ 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?". 9.⁵ 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).

Listado de descriptores libres y listado de palabras clave

Manela Juncà Campdepadrós

PID_00144350



Universitat Oberta
de Catalunya

www.uoc.edu



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. Fundació 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

Introducción	5
Objetivos	6
1. Lenguajes documentales libres	7
2. Listado de descriptores libres	9
2.1. Tipología	9
2.2. Listado de descriptores libres en la indización	10
2.2.1. Creación de un listado de descriptores libres	11
2.2.2. Indización	11
2.2.3. Aplicaciones del listado de descriptores libres	14
2.3. Listado de descriptores libres en la recuperación	15
3. Listado de palabras clave	17
3.1. La indización automática	17
3.1.1. Ciencias implicadas en la indización automática	18
3.1.2. Funcionamiento y evolución de los programas de indización automática	19
3.2. El listado de palabras clave en la indización	30
3.3. El listado de palabras clave en la recuperación	31
Actividades	35
Glosario	37
Bibliografía	39

Introducción

Este módulo os introduce en el uso de los lenguajes libres. Comprende dos lenguajes: los listados de descriptores libres y los listados de palabras clave.

Itinerario de estudio

El módulo empieza con las características de los lenguajes libres, ya que los dos últimos lenguajes del curso comparten esta tipología. A continuación, se describen las listas de descriptores libres, su creación y el proceso de indización y recuperación. Finalmente, el curso acaba con las listas de palabras clave y la indización automática.

Tabla. Conceptos más importantes

Concepto	Ved
Lenguaje libre	1. Lenguajes documentales libres
Descriptor libre	2. Listado de descriptores libres
Palabra clave	3. Listado de palabras clave
Palabra vacía	3.1.2. Funcionamiento y evolución de los programas de indización automática
Cálculo de frecuencia	3.1.2. Funcionamiento y evolución de los programas de indización automática
Frecuencia inversa	3.1.2. Funcionamiento y evolución de los programas de indización automática
Discriminación	3.1.2. Funcionamiento y evolución de los programas de indización automática
Análisis morfológico	3.1.2. Funcionamiento y evolución de los programas de indización automática
Análisis sintáctico	3.1.2. Funcionamiento y evolución de los programas de indización automática
Análisis semántico	3.1.2. Funcionamiento y evolución de los programas de indización automática

Objetivos

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

1. Conocer la aplicación de los lenguajes libres al análisis de contenido.
2. Valorar las ventajas de los lenguajes libres en la indización.
3. Valorar los inconvenientes de los lenguajes libres en la recuperación.
4. En cuanto a los **listados de descriptores libres**:
 - Definir los listados de descriptores libres y sus tipologías.
 - Conocer el proceso de creación del lenguaje.
 - Adquirir una cierta habilidad en la indización de documentos.
5. Por lo que respecta a los **listados de palabras clave**:
 - Introducirse en los mecanismos de la indización automática.
 - Reconocer las aplicaciones estadísticas y lingüísticas aplicadas a la indización automática.
 - Saber simular/entender el proceso de selección y elección de palabras clave.

1. Lenguajes documentales libres

Este último módulo está dedicado a los dos lenguajes documentales libres: los listados de descriptores libres y los listados de palabras clave.

A diferencia de los lenguajes controlados, que neutralizaban las carencias de los lenguajes naturales y requerían conocimientos elevados para ser aplicados, los lenguajes libres no ofrecen tantas garantías en la recuperación, pero son amigables. Son lenguajes muy usados, ya que son:

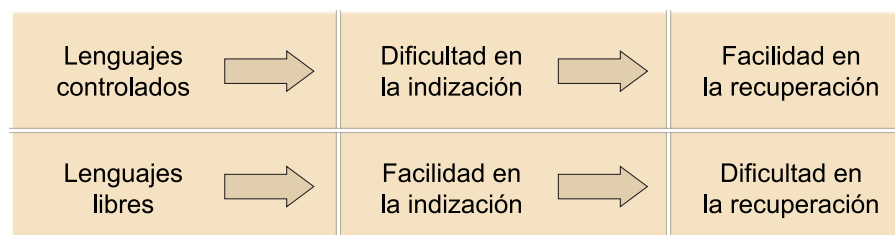
- **Baratos:** los gastos de construcción son mínimos. Los elabora una persona o un programa informático.
- **Rápidos:** tanto en la construcción, que es inmediata, como en la actualización de su vocabulario, que se va incorporando al listado a medida que va ingresando en el fondo documental.
- **Terminológicamente ricos:** de resultados de la actualización inmediata de sus listas.
- **Fáciles de usar:** no necesitan conocimientos previos sobre lenguajes documentales, control de vocabulario, o precoordinación, entre otros.
- **Coherentes:** ofrecen coherencia máxima, de manera que dos servicios de información y documentación (SID¹) con el mismo documento y software de indexación automática llegarán a una indexación idéntica (esta característica sólo es aplicable al listado de palabras clave).

Sin embargo, los lenguajes libres presentan inconvenientes en la recuperación, ya que al trabajar con lenguaje natural libre arrastran todos los problemas derivados de la ambigüedad (sinonimia, polisemia, homonimia), lo que provoca silencio y ruido documentales y presentan un número excesivo de resultados para cada petición. Es el proceso inverso de los lenguajes controlados, en los que la dificultad radicaba en el proceso de indexación y aseguraba, en cambio, una recuperación unívoca.

Tipología según el nivel de control

Recordemos que la tipología según el nivel de control clasifica los lenguajes documentales en libres o controlados, en función de si los términos de indexación corresponden a un lenguaje natural o a un lenguaje artificial construido para garantizar la indexación y recuperación.

⁽¹⁾SID es la sigla de servicio de información y documentación.



Actualmente muchas bases de datos combinan la utilización de los lenguajes controlados con los libres, de manera que podemos buscar y recuperar por diversas vías. Es habitual que los catálogos bibliográficos y otras bases de datos ofrezcan la posibilidad de buscar la materia de dos maneras:

- a) Por el **campo *materia***: indizando con algún lenguaje controlado (tipo lista de encabezamientos de materia, tesauros o una lista de autoridades).

- b) Por el **campo *palabra clave***: indizando con un programa de indización automática o lista de palabras clave.

2. Listado de descriptores libres

Un **listado de descriptores libres** es un vocabulario monolingüe de términos de indización ordenados alfabéticamente.

Estos términos son escogidos por el analista, sin verificar si existen ni cómo se introducen en una lista previamente establecida. Por lo tanto, no es un lenguaje controlado, sino libre, y por eso recibe este nombre.

2.1. Tipología

Las listas de descriptores libres son un lenguaje documental que tiene las **características** siguientes:

1) Se trata de un tipo de **lenguaje analítico** (frente a los sintéticos). En el proceso de indización se identifican los conceptos que conforman el contenido del documento y se representan mediante una serie de descriptores. Por esta razón, su aplicación está ligada a la existencia de sistemas automatizados que permiten la recuperación de la información mediante la combinación de estos descriptores.

Ejemplo

Ponemos como ejemplo el resumen siguiente:

Dalmau Ribalta, A. (2002). *Los cátaros*. Editorial UOC (“Biblioteca Lectus Universitaria”).

El catarismo es un movimiento cristiano disidente de la Baja Edad Media. Se extendió por diversos puntos de Europa, con una incidencia especial en el Languedoc. Esta obra da a conocer qué fue exactamente el catarismo y su trayectoria.

La obra *Los cátaros* procura distinguir la llamada *Iglesia de Dios* de toda una literatura esotérica y legendaria que la ha deformado de manera significativa y que, al mismo tiempo, se aleja de una concepción historiográfica tradicional, ya superada, que la consideraba como una simple prolongación del antiguo maniqueísmo.

A lo largo de esta obra se explica en qué consistía la liturgia, los ritos y la doctrina religiosa de esta Iglesia, considerada herética y perseguida por la Iglesia católica. Resumid su trágica historia con hechos tan relevantes como la llamada *cruzada albigense* y el nacimiento de la Inquisición.

El analista indiza más de un término²:

Baja Edad Media
cátaros
cristianismo
disidencia religiosa
Europa
liturgia
Languedoc

Ved también

Podéis ver la tipología de los lenguajes documentales en el apartado 5 del módulo “Análisis de contenido: resumen e indización” de esta asignatura.

⁽²⁾En contraposición al LEMAC, que es un lenguaje sintético y sería un único término:

LEMAC: Càtars-Història

2) Se trata de un tipo de **lenguaje natural** porque la indización se basa en el uso de palabras o expresiones libres que pertenecen al lenguaje o discurso común y que se encuentran en los mismos documentos (como el título, el resumen o el texto) y a los conceptos de la consulta del usuario. Es decir, en ningún momento se utiliza un código (numérico, alfanumérico) que represente este contenido. El lenguaje natural presenta el problema de la falta de univocidad, es ambiguo y posibilita los fenómenos de la sinonimia, la polisemia y la homonimia.

Ejemplo

Listado de descriptores libres³:

Càtars

⁽³⁾En contraposición, indizado con un sistema de clasificación como la CDU, que es el único lenguaje documental codificado, sería:

CDU año 2004 27-789.67.

3) Se trata de un **lenguaje libre** porque está formado por términos de indización extraídos del lenguaje natural y que se utilizan o se sugieren en el mismo documento. El listado de descriptores resultante está constituido por una colección no ordenada de conceptos (sólo por orden alfabético). Estos conceptos están expresados por palabras o expresiones extraídas de los documentos, o propuestos por los mismos documentalistas, sin verificar si están en una lista establecida a priori (es también, por lo tanto, un lenguaje construido a posteriori).

4) Se trata de un tipo de **lenguaje postcoordinado**, ya que permite la coordinación de conceptos en el momento de la recuperación y la utilización de un gran número de términos de indización o puntos de acceso. La combinación precisa de estos diferentes descriptores, en la recuperación, se tiene que remitir al documento buscado.

5) Se trata de un tipo de **lenguaje de estructura combinatoria o asociativa**, en el que los conceptos o descriptores se organizan de una manera independiente y sólo se produce la combinación o intersección en las operaciones de indización y recuperación. Por lo tanto, se trata de un lenguaje en el que los términos que lo componen se organizan en una lista por orden alfabético sin respetar ningún otro tipo de estructura que sitúe cada descriptor en una cadena lógica, jerárquica o estructurada de conceptos.

Ejemplo

En este caso, el indizador hace:

Càtars AND Europa

2.2. Listado de descriptores libres en la indización

La lista de descriptores libres es un lenguaje documental que da libertad total al analista, por lo que lo convierte en muy amigable, rápido y fácil de usar.

Todos nosotros en algún momento de nuestra vida hemos ordenado fotos en álbumes, hemos clasificado los libros de la biblioteca o hemos ordenado artículos y fotocopias bajo algún tema o subtema. No éramos conscientes de ello, pero estábamos indizando con un listado de descriptores libres.

2.2.1. Creación de un listado de descriptores libres

El **proceso de creación** de una lista de descriptores libres es muy sencillo: el analista va leyendo los documentos que tiene que indizar y va tomando nota de los términos que son interesantes según su parecer. Los va anotando y los clasifica en orden alfabético. Les da un mínimo de forma y procura eliminar duplicidades de número y género (singular-plural, masculino-femenino), usa una única lengua y elimina algunos sinónimos.

Los descriptores libres tienen un gasto de construcción casi nulo. Son inmediatos, permiten un vocabulario actualizado, y se adaptan a la realidad del SID.

En algunos casos el listado de descriptores libres suele ser el primer paso en la tarea de crear un lenguaje documental. Los analistas empiezan por redactar un listado de descriptores libres con la intención de controlarlo y avanzar hacia un lenguaje documental realmente controlado. Si el analista quiere ejercer un poco de control sobre sus descriptores, sin llegar a trabajar con un vocabulario por completo controlado, como serían las listas de encabezamientos o los tesauros, tiene que tomar algunas medidas, como por ejemplo:

- Establecer una lengua o idioma de la lista.
- Controlar las formas singular/plural.
- Establecer los sintagmas nominales preferentes:
 - Sustantivo + adjetivo.
 - Sustantivo + preposición + sustantivo.
- Controlar los sinónimos.
- Utilizar formas usuales frente a otras muy técnicas.

Ved también

Las medidas de control de vocabulario detalladas en el apartado dedicado a las listas de encabezamientos de materia del módulo “Listas de encabezamientos de materia y listas de autoridades” son adecuadas también en este caso.

2.2.2. Indización

El analista lee el documento, determina la materia del documento y acto seguido propone un descriptor. Representándolo verbalmente, el analista diría “este documento trata de tal tema”. No consulta ninguna lista, no comprueba si el término que piensa está aceptado o no, no tiene que seguir las reglas de combinación de ningún lenguaje; tiene libertad total a la hora de indizar.

El hecho de no necesitar una traducción del lenguaje natural a uno artificial controlado tiene ventajas y comporta inconvenientes. El primero, y el más evidente es la disparidad de indizaciones: ante el mismo documento, tres analistas pueden llegar a tres resultados muy diferentes.

Disparidad de indizaciones

Siguiendo con el ejemplo sobre el libro de los cátaros, vemos el resultado de la indización hecha por tres analistas diferentes:

Disparidad de indizaciones

Analista A	Analista B	Analista C
Baja Edad Media cátaros Francia movimientos cristianos disidentes	catarismo herejías órdenes monásticas religión	albigenses cismas historia medieval Languedoc

Como consecuencia de esto, la tasa de coherencia es la más baja de todos los lenguajes documentales.

Ejemplo

En el ejemplo, el uso de sinónimos y términos genéricos ha alterado tanto el resultado que no han coincidido en ningún descriptor (la tasa es del 0%).

Un segundo aspecto que hay que destacar es el grado de exhaustividad. Recordemos que el grado de exhaustividad puede ser alto, medio o bajo y que afecta a los lenguajes postcoordinados como éste, en el que el analista puede escoger el número de términos para representar un documento.

De esta manera, con una lista de descriptores libres podemos indizar el libro de los cátaros con tres grados diferentes:

Exhaustividad profunda o alta	Exhaustividad intermedia o media	Exhaustividad genérica o baja
Baja Edad Media catarismo cátaros cátaros, historia cristianismo cruzada albigense disidencia religiosa doctrina religiosa Iglesia de Dios Europa Inquisición liturgia Languedoc ritos religiosos	Baja Edad Media cátaros cristianismo disidencia religiosa Europa liturgia Languedoc	Baja Edad Media cátaros Languedoc

a) Las **ventajas** de este lenguaje en la indización giran en torno a la facilidad, la simplicidad y la libertad de uso:

- No necesita traducción de los conceptos del lenguaje natural de los documentos a un lenguaje artificial.

Ejemplo

Si en el texto de los cátaros que estamos usando como ejemplo se hablaba de *literatura esotérica y legendaria*, el analista puede indizar **leyendas** o **literatura esotérica** sin comprobar la forma en ninguna lista, si es aceptado o no, si se prefiere la forma singular/plural, si se puede adjetivar o no, si tiene relaciones semánticas que ayuden a la indización.

- Se trata de un tipo de lenguaje rápido y fácil de actualizar. Esta ventaja es muy apreciada en entornos tecnológicos en los que la terminología va más adelantada que su normalización en listas controladas. Recordemos que entidades como el Termcat son la fuente adecuada para consultar la forma correcta de escribir el término.

Ejemplo

Si el texto que hay que indizar habla del *Maniqueísmo* o de cualquier otro tema, no hay problema; se añade a la lista alfabética de descriptores **Maniqueísmo**.

- Se adapta perfectamente al nivel de usuarios y al tipo de SID, ya que es un lenguaje hecho a la medida de su centro.

Ejemplo

Si un SID de tipo genérico recibe este documento lo indizará con algún término del tipo **Religión**; en cambio, si es un SID especializado extraerá términos más concretos como, por ejemplo, **Cruzada albigense**.

- No hace falta una formación previa de los analistas. Precisamente la ausencia de reglas y principios hace innecesaria la formación.

b) Los inconvenientes giran en torno a la elección del término de indización entre todos los términos posibles. Un segundo problema derivado es prever de qué manera el usuario hará la búsqueda y así coincidir.

Ejemplo

En el ejemplo de los cátaros, el analista puede decidir indizar:

- Por el marco geográfico: Languedoc, Francia, Europa.
- Por el marco temporal: Baja Edad Media, Edad Media, historia medieval.
- Por el concepto *cátaros*: cátaros, catarismo, albigense.
- Por el concepto *disidencia*: disidencia, herejía, cisma.

Algunos analistas optan por indizar el término tal y como aparece en el documento (por ejemplo, Languedoc) y otros realizan la acción de escoger términos que creen que tienen más posibilidades de ser recuperados (por ejemplo, Francia).

2.2.3. Aplicaciones del listado de descriptores libres

Hemos comentado con anterioridad el uso generalizado de descriptores libres en el ámbito doméstico y que también suele ser un primer paso en el camino hacia la creación de un lenguaje controlado, pero también encuentra **aplicaciones** en los ámbitos siguientes:

1) En la **indización de información temporal**: cuando los ítems de información tienen una vida determinada ligada a un acontecimiento (político, deportivo o cultural) concreto que hace demasiado costoso invertir en un lenguaje controlado y su formación posterior.

2) En la **indización de documentos audiovisuales (imagen fija o en movimiento y audio) de la web**: los sistemas de indización automática de la web no pueden indizar, de momento, documentos que no contengan texto, como las fotografías o los vídeos de banda ancha, tan numerosos en la red. La solución pasa por indizarlos manualmente con la colaboración de los internautas.

3) En la **indización social** los internautas asignan etiquetas con descriptores a los recursos web. Mathes (2004) habla de *metadatos generados por el usuario*. Pueden indizar fotografías⁴, webs⁵, blogs⁶ y compartir sus descriptores para colaborar en la indización de todo tipo de contenidos en el espacio web compartido y abierto. Cada recurso (por ejemplo, la web de la Wikipedia) es indizado por un grupo de usuarios que pueden coincidir o no con las etiquetas: uno puede indizar *wiki*, otro *enciclopedia*, otro *obra de referencia*, por ejemplo.

La asignación de las etiquetas se hace sin ánimo de lucro. Los internautas no buscan un beneficio económico, sino beneficiarse de búsquedas mejores. Los descriptores escogidos no se supervisan ni tienen ninguna estructura semántica.

Sinónimos

Otros nombres que recibe el fenómeno de la indización social son *tagging*, *etiquetado colaborativo*, *clasificación social*, *etiquetado social* o *folksonomies*. El origen del neologismo *folksonomía* lo debemos a Thomas van der Wal, que fusionó *folk* (gente, popular) y *taxonomía* (gestión de la clasificación), lo que dio como resultado una “indización gestionada popularmente”.

La indización social participa de las características de las listas de descriptores libres en la filosofía de la indización, ya que cada participante indiza unos descriptores libres seleccionados por un proceso intelectual a partir del examen del recurso, sin verificar si los descriptores propuestos están o no en una lista establecida.

Los intentos por mejorar la recuperación van en la línea de aplicar algoritmos de ponderación y eliminación de etiquetas vacías que no aportan significado relevante para la comunidad. Estos mecanismos son totalmente invisibles para

⁽⁴⁾<http://www.flickr.com/>

⁽⁵⁾<http://delicious.com/>
<http://www.mister-wong.es/users/500000111/>

⁽⁶⁾<http://technorati.com/>

Lectura complementaria

Si queréis ampliar la información sobre los metadatos generados por el usuario, podéis consultar la página web de Adam Mathes:

- Folksonomies

el usuario. Algunos autores proponen alfabetizar al usuario dándole instrucciones para indizar, pero se duda de su eficacia, ya que una de las razones del éxito del *tagging* es la libertad que da al indizador internauta.

2.3. Listado de descriptores libres en la recuperación

a) Las **ventajas** que tienen los listados de descriptores libres en la recuperación son las siguientes:

- Como lenguaje analítico, permite la coordinación de diversos términos en la consulta, lo que permite búsquedas precisas.
- El lenguaje se actualiza continuamente.
- Facilidad de uso en la realización de consultas, ya que no hay que dominar previamente ningún lenguaje especializado, sofisticado ni artificial.

b) No obstante, también tienen **inconvenientes**. Si en la indización todo eran facilidades, en la recuperación encontramos todo tipo de problemas derivados del silencio y del ruido documentales.

El principal problema de este lenguaje documental se encuentra precisamente en este punto, la recuperación, ya que la abundancia de sinónimos y homónimos hace difícil predecir con qué término buscará al usuario y si coincidirá con el escogido por el indizador.

Las dificultades más notables se presentan a la hora de formular la petición:

- No elimina la polisemia del lenguaje natural, lo que provoca ruido documental.
- Presencia de sinónimos, lo que genera silencio documental e imposibilita recuperar los documentos que tratan sobre los conceptos de la consulta si éstos han sido indizados por los analistas con formas diferentes.

Ejemplo de sinónimos

Un ejemplo de búsqueda con todo tipo de sinónimos sería: CEE OR UE OR Unión Europea.

- Variantes ortográficas del mismo concepto (disco, disquete, diskete).
- Variantes idiomáticas (bafles y altavoces).
- Falta de una estructura semántica que ayude a la localización de términos. Esta falta provoca vacíos cuando pedimos un término genérico, con el cual no ha sido indizado el documento.

Ejemplo de falta de estructura semántica

Buscamos, por ejemplo, sobre lenguajes documentales y no recuperamos un documento sobre tesoro porque está indizado sólo como tesoro, no con un término conceptual-

mente más genérico como puede ser *lenguaje documental*. En un lenguaje controlado, esta vinculación se haría evidente (con siglas del tipo TG o TA) y ayudaría a la recuperación.

- Coordinación falsa. Entendemos por *coordinación falsa* recuperaciones inesperadas y erróneas que, a pesar de contener los elementos de la búsqueda, corresponden a temas diferentes.

Ejemplo de coordinación falsa

Buscamos, por ejemplo, sobre pintura catalana y pedimos Pintura AND Cataluña y recuperamos:

- Pintores catalanes (por ejemplo, M. Fortuny).
- Cataluña en la pintura (por ejemplo, la visión de J. Sorolla sobre el litoral catalán).
- Pintura en Cataluña (todos aquellos pintores que han pintado en Cataluña).
- Industriales de la pintura catalanes (pintores de paredes).

Las soluciones a estos inconvenientes pasan por recopilar toda la lista que seamos capaces de elaborar sobre el concepto: términos genéricos y específicos, polisémicos, sinónimos, siglas, variantes idiomáticas, cambios de género y número. Las estrategias de búsqueda suelen ser complejas.

3. Listado de palabras clave

El **listado de palabras clave** es un vocabulario ordenado alfabéticamente de los términos con carga significativa extraídos de un documento mediante un programa informático.

Listados de palabras clave

Consideramos *listados de palabras clave* los listados resultantes de la indización automática.

Una lista de palabras clave está constituida por una colección dispuesta en orden alfabético de las palabras significativas, llamadas también *no vacías* (es decir, todas las palabras que no son artículos, conjunciones, pronombres, preposiciones, numerales y ciertos verbos y adverbios), extraídas de una manera automática por el ordenador a partir del título, del resumen y cada vez más a menudo del texto completo de los documentos (van Slype). Acostumbran a ser listas monolingües.

Lectura complementaria

Si queréis ampliar vuestro conocimiento sobre la información dada por van Slype, podéis leer la obra siguiente.

G. van Slype (1991). *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales* (pág. 23). Madrid: Pirámide / Fundación Germán Sánchez Ruipérez (Biblioteca del Libro).

Los listados de palabras clave tienen la misma tipología que los listados de descriptores libres. Los dos lenguajes son:

- analíticos,
- naturales,
- libres,
- postcoordinados,
- de estructura combinatoria.

Las **diferencias** entre ellos, que son diferencias grandes, residen en que:

- Los listados de descriptores libres se basan en la indización humana y los listados de palabras clave, en la indización automática.
- Los listados de descriptores libres indizan conceptos y los listados de palabras clave indizan en su mayoría unitérminos.

3.1. La indización automática

La **indización automática** es el método por el que un ordenador aplica un algoritmo (o programa) al título, resumen o texto completo del documento para identificar los términos que puedan representar la materia y ser usados como términos de indización y recuperación en un índice o lista.

Hay dos **tipos** de indización automática:

a) **Totalmente automatizada** (en inglés, *automatic indexing*).

b) **Semiautomática** (en inglés, *machine-aided indexing*): el programa selecciona posibles descriptores procedentes de un tesoro o de una lista controlada y un documentalista acepta o rechaza la propuesta.

Los factores que hacen posible la indización automática son la elevada inversión de tiempo que conlleva la indización intelectual, el aumento de la documentación electrónica y a texto completo, y la extensión de sistemas de gestión documental en las instituciones y empresas, así como la evolución del procesamiento del lenguaje natural o PLN⁷ (Méndez y Moreiro, 1999).

Aunque podemos encontrar aplicaciones dedicadas exclusivamente al resumen y a la indización automática, lo más habitual es que estos programas formen parte del mismo programa de gestión documental del centro.

3.1.1. Ciencias implicadas en la indización automática

En la indización automática intervienen diversas disciplinas científicas, la informática, la lingüística y la estadística:

1) La **informática** proporciona el algoritmo que representa, procesa, almacena y recupera las palabras clave.

2) La **lingüística** asigna etiquetas de tipo morfológico y lleva a cabo análisis sintácticos. Los programas de indización automática más evolucionados incorporan también análisis de tipo semántico:

- **Morfología:** trabaja en el ámbito de la palabra. Asigna la categoría gramatical de las palabras y la raíz común (o lema). Recordemos que las categorías gramaticales son sustantivo, verbo, adjetivo, adverbio, entre otras, y que morfológicamente pueden ser de diversos géneros, números, tiempo y modo. La lematización permite contar todas las veces que un concepto aparece en el texto aunque se exprese con formas diferentes.
- **Sintaxis:** su campo de trabajo es la frase. Identifica los componentes de la frase, que son el sujeto, el verbo y los objetos, y su concordancia. Gracias a este análisis puede desambiguar gramaticalmente los términos no resueltos en la etapa morfológica.
- **Semántica:** se ocupa del texto. Desambigua términos polisémicos, ya que el texto permite optar por un significado u otro.

3) La **estadística** tiene diversas aplicaciones en documentación, pero en indización automática interviene especialmente en el cálculo de la frecuencia relativa de las palabras en un texto. Esta técnica calcula cuántas veces aparece una palabra en un texto y en función de unos baremos (ni las palabras más

⁽⁷⁾PLN es la sigla de procesamiento del lenguaje natural. A partir de ahora, denotamos solamente PLN.

Lectura complementaria

Para ampliar la información sobre la indización automática, podéis leer la obra siguiente:

E. Méndez; J. A. Moreiro (1999). "Lenguaje natural e indización automatizada". *Ciencias de la Información* (pág. 11-24).

Ejemplo de desambiguación

Por ejemplo, *columna* en un texto de arquitectura o *columna* en un texto de periodismo.

repetidas, ni las menos) se seleccionan las palabras clave del documento. Los teóricos de este método son G. K. Zipf (1949), H. P. Luhn (1957), Sparck Jones (1972) y G. Salton (1989).

Trataremos con más detalle estos procesos en el próximo subapartado.

3.1.2. Funcionamiento y evolución de los programas de indización automática

Los primeros estudios en indización automática datan de las décadas de 1950-1960. Se basaban en principios estadísticos y probabilísticos. En la década de 1970, G. Salton incorpora el modelo de valor de discriminación y de relevancia de los términos. En la década de 1980 aparecen los criterios lingüísticos y en la de 1990 se plantea la indización automática de información multimedia (imágenes y sonido). Desde la década de 1980 hasta la actualidad, los programas de indización automática combinan el cálculo estadístico con el etiquetado morfológico y el análisis sintáctico.

Evolución de los programas de indización automática

1950-1960	1970	1980	1980-actualidad
Cálculos estadísticos y probabilísticos	Relevancia de los términos	Análisis lingüístico	Modelo mixto: estadístico + lingüístico

Los programas informáticos de indización automática buscan emular el proceso mental humano. La cuestión central que nos interesa es “¿cómo pueden detectar las palabras con más carga semántica, con más significado del texto?”. Con este objetivo, suelen seguir los siguiente pasos:

1) Lectura de los documentos

El primer paso es leer el texto. Para hacerlo el documento se debe encontrar en formato electrónico. Actualmente, muchos documentos ya se encuentran en soporte electrónico. También son fácilmente accesibles los títulos, resúmenes, sumarios o bien el texto completo en las bases de datos de publicaciones periódicas y catálogos bibliográficos. Si se trata de documentación en soporte papel, primero hay que escanearla y aplicar un programa de tipo OCR (reconocimiento óptico de caracteres) para transformar la imagen de la página escaneada en texto electrónico.

Programas OCR

Estos programas OCR están implementados en la mayoría de escáneres domésticos. Recomendamos hacer la prueba de escanear una página y convertirla en texto.

¿Cómo puede un programa detectar una palabra? El programa detecta la palabra porque la considera una cadena de caracteres entre espacio en blanco y espacio en blanco.

2) Diseño

El equipo de documentalistas y programadores debe decidir:

- Las partes del documento que serán indizables. Son obligatorios el título y el resumen y es deseable el texto entero, sobre todo en artículos.
- El tratamiento que darán a las cifras, los signos de puntuación, los guiones, las mayúsculas/minúsculas y los acentos. Habitualmente son caracteres que no aportan significado, pero en determinados contextos pueden ser determinantes.

Una vez tenemos el documento en formato electrónico y el programa ha leído las partes seleccionadas, el paso siguiente consiste en distinguir las palabras que tienen significado de las que no lo tienen y, entre las que lo tienen, seleccionar las de más relevancia en la indización y su recuperación posterior. Este paso se resuelve aplicando diversos métodos estadísticos y lingüísticos, ya sea en un orden secuencial o bien alternando los métodos.

3) Palabras con contenido y palabras vacías

En el lenguaje natural hay palabras que tienen mucho significado (como los sustantivos y los adjetivos) y otras que no tanto (como los artículos o las conjunciones). Estas últimas se conocen como *palabras vacías*.

Las **palabras vacías** son palabras sin significado, como los artículos, los pronombres, las preposiciones, las conjunciones o los adverbios, que son filtradas antes o después del procesamiento del texto.

Las palabras vacías son muy frecuentes pero aportan poco valor de contenido semántico al texto y poca ayuda a la recuperación, ya que el usuario no las utiliza en la búsqueda y tampoco son recogidas en el fichero inverso de la base de datos. Se conocen como *listas de detención* en castellano y como *stop word list* en inglés (Hans Peter Luhn fue el creador del término y del concepto *stop word*).

Se elabora una lista de estas palabras vacías y se introducen en el programa, que va leyendo el texto palabra por palabra y las contrasta con el fichero de palabras vacías. Si el programa las encuentra en la lista, es que son vacías y no las indiza. Si no aparecen en la lista, es que tienen significado.

Palabras vacías

Artículos	el, la, uno, una
Pronombres	mío, nuestro
Verbos	comer, ser, reír (en todas las formas verbales)
Adverbios	tranquilamente, eficazmente
Preposiciones	a, con, de, sin

Ejemplo

Hay veces que estos caracteres sí tienen importancia:

- Cifras: TV3, 1492.
- Puntos, guiones, signos: www.uoc.edu, Canal+, e-mail.
- Acentos (útiles para diferenciar diacríticos): deu/Déu en catalán, te/té en castellano.

Numerales	cuarto, octavo
Adjetivos	alto, bajo, grande
Conjunciones	y, por lo tanto, pero, porque

Hay sistemas en los que la lista de palabras vacías:

a) Viene **predeterminada**: el sistema dispone, ya desde el principio, de la lista de palabras vacías de su idioma o idiomas. De hecho, su elaboración es fácil, ya que sólo hay que añadir las categorías vacías de una base de datos de terminología al idioma que se quiera. Los artículos siempre son los mismos, las conjunciones también, incluso los verbos se pueden llegar a contabilizar y flexionar en todos los tiempos verbales.

b) Se **evita expresamente** para permitir al sistema la búsqueda por frases y sintagmas. Los sistemas que los evitan disponen de otras herramientas para reducir significativamente el número de palabras indizadas, como técnicas de *stemming* o lematización que veremos más adelante.

c) Está **contextualizada** (*stop word context-dependent*). Cada sistema elabora la lista de palabras vacías según su ámbito temático. Contextualizar el listado permite evitar dos inconvenientes graves:

- Palabras con significado que se convierten en palabras vacías.

Ejemplo

Si un usuario busca en el catálogo de una biblioteca especializada en astronomía, no buscará por el concepto "Astronomía", ya que toda la colección de la biblioteca hará referencia a este concepto. El concepto *Astronomía* deviene vacío en este contexto.

- palabras vacías que se convierten en importantes en la indización.

Ejemplo

La lista de palabras vacías se puede evitar, por ejemplo, para recuperar un concepto como el diario *El País*, en el que el artículo tiene un papel importante.

Ejemplo

En un texto de historia los números (como 1914, 1936), los numerales (Jaime I) y los adjetivos pueden tener mucha carga significativa (Alta y Baja Edad Media; sobrenombres adjetivados de personajes como Jaime I el Conquistador y acontecimientos como la guerra fría). En el ejemplo siguiente, las palabras subrayadas son palabras vacías:

Palabras castellanas que son vacías o que no dependen del ámbito temático

Palabra	En el texto	Economía	Derecho	Educación	Multidisciplinario
ESO	"Eso es lo importante..." "La reforma de la ESO"	<u>ESO</u>	<u>ESO</u>	ESO	<u>ESO</u>
Cabo	"Se precisa llevar a cabo..." "La composición rocosa del cabo..."	<u>cabo</u>	<u>cabo</u>	<u>cabo</u>	cabo (geografía)
Tuya	"Yo traeré la mía pero tú trae la tuya." "La tuya es un árbol perenne."	<u>tuya</u>	<u>tuya</u>	<u>tuya</u>	tuya (árbol)

Fuente: extracto de I. Gil Leiva (2008). *Manual de indización. Teoría y práctica* (pág. 333). Gijón: Trea.

En el caso de la contextualización, las palabras vacías pueden ser introducidas manualmente o bien ser el resultado de algún cómputo estadístico de frecuencia.

Ejemplo

Supongamos el texto “Osiris es el primero de todos los humanos y dioses que vive una segunda vez”.

Descompongamos ahora el texto como lo haría un programa de indización automática. Partimos del supuesto de que el programa tiene una lista de palabras vacías predeterminada que incluye preposiciones, conjunciones, numerales, verbos y artículos.

Texto	Comprobación en la lista de palabras vacías		Resultado de la indización
Osiris	no está	es un nombre propio	Osiris
es	es una palabra vacía	es una forma verbal	-
el	es una palabra vacía	es un artículo	-
primero	es una palabra vacía	es un numeral	-
de	es una palabra vacía	es una preposición	-
todos	es una palabra vacía		-
los	es una palabra vacía	es un artículo	-
humanos	no está	es un sustantivo	humanos
y	es una palabra vacía	es una conjunción	-
dioses	no está	es un sustantivo	dioses
que	es una palabra vacía	es una conjunción	-
vive	es una palabra vacía	es una forma verbal	-
una	es una palabra vacía	es un artículo	-
segunda	es una palabra vacía	es un numeral	-
vez	no está	es un sustantivo	vez

El texto queda indizado con estas palabras clave:

```
Osiris
humanos
dioses
vez
```

¿Son muy numerosas las palabras vacías en un texto? La presencia de palabras vacías no es en absoluto desdeñable. Los estudios estadísticos calculan que el 50% de las palabras de un texto son vacías. Así, un artículo de 8.000 palabras tiene unas 4.000 palabras vacías y otras 4.000 con significado. En el ejemplo de Osiris, las quince palabras de la frase han quedado reducidas a cuatro palabras clave (en este caso, el ejemplo buscaba pretendidamente muchas palabras vacías para ilustrar este subapartado).

El procedimiento en PLN para distinguir las palabras vacías de las palabras con significado consiste en la aplicación de una lista previa de palabras vacías o de una combinación de métodos estadísticos (frecuencia) y lingüísticos (análisis

morfológico, sintáctico y semántico). Desde que H. P. Hans Meter Luhn en 1957 eliminó las palabras vacías, esta técnica ha sido una constante en los posteriores programas de indización automática.

4) Métodos estadísticos

Los métodos estadísticos han sido la primera aproximación a la indización automática y todavía hoy en día son una parte consustancial. La teoría de fondo es el cálculo del peso (ponderación) de las palabras: ni las palabras más repetidas (por vacías) ni las menos repetidas (por específicas) son adecuadas para ser seleccionadas.

Los métodos estadísticos aplicados en PLN son de tres tipos: frecuencia, frecuencia inversa y discriminación; y se pueden usar solos o en combinación:

a) Frecuencia de aparición (ley de Zipf): el psicólogo y psicolingüista estadounidense George Kingsley Zipf (1949) propuso que el número de veces que un término aparece en un texto (frecuencia) y la posición que ocupa en una lista de palabras más o menos frecuentes (rango) es constante. Para Zipf los escritores usan un abanico reducido de palabras, preferentemente cortas, y su longitud es inversamente proporcional a la frecuencia.

Ejemplo

En la web hay disponible una copia de las palabras francesas y su frecuencia. En esta lista vemos qué palabras de uso cotidiano tienen una frecuencia más alta y qué palabras menos cotidianas a la inversa.

Frecuencia de aparición de las palabras francesas

aliments	13,97	grand	674,68
aller	234,58	laisse	120,97
ami	95,16	les	16.011,00
avec	3.019,71	neutre	11,52
collège	24,42	permutable	0,03
comité	73,71	pour	5.332,48
être	1.986,23	translateur	0,06
goût	98,77	vitamines	3,84

Fuente: extraído de http://www.lexique.org/listes/liste_mots.txt

Hans Meter Luhn (1957) aplica la ley de Zipf en el campo de la indización automática. Luhn ve que la frecuencia con la que se repite una palabra puede ser un buen indicador para seleccionar las palabras clave con más significado. Las que tienen una frecuencia muy alta no acostumbran a tener carga significativa (en el ejemplo anterior serían *avec*, *les* y *pour*, pero recordemos que en

función del contexto también podría serlo *aliments*) y llevan a la recuperación de muchos documentos. Las que tienen una frecuencia baja suelen ser palabras muy específicas en el texto y, por lo tanto, con carga significativa, pero en la recuperación presentan pocos resultados. Para Luhn lo mejor para indizar eran los términos con una frecuencia media.

Luhn propone los pasos siguientes:

- Calcular la frecuencia de todas las palabras del texto o colección.
- Clasificarlas en orden decreciente.
- Eliminar las de frecuencia más alta.
- Eliminar las de frecuencia más baja.
- Indizar con el resto de palabras.

Luhn aplica esta técnica para eliminar las palabras vacías. A partir de él, todos los sistemas utilizan esta técnica para crear la lista de palabras vacías.

b) Frecuencia inversa: Sparck Jones (1972) puso de manifiesto la capacidad de discriminación de un término frente a otro. Esta discriminación tiene que ser vista en el conjunto de la colección, no en un solo documento. Hay que comparar las palabras clave entre los documentos del fondo para detectar cuáles son realmente discriminatorias.

Ejemplo

Imaginemos un fondo de cuatro documentos.

- Documento 1: “Visitad el Museo Picasso de Barcelona en el barrio del Borne”.
- Documento 2: “En Figueres también podéis visitar el Museo del Juguete”.
- Documento 3: “El Museo de Figueres y el Museo Picasso son los más famosos”.
- Documento 4: “En Figueres no dejéis de visitar el Museo Dalí”.

Eliminemos las palabras vacías *el, los, a, al, del, de, los, más, no* y todas las formas del verbo *visitar* (*visitad, visitar*). Nos queda:

Doc.	Museo	Picasso	Barcelona	Barrio	Borne	Figueras	Juguete	Famosos	Dalí
1	1	1	1	1	1	0	0	0	0
2	1	0	0	0	0	1	1	0	0
3	2	1	0	0	0	1	0	1	0
4	1	0	0	0	0	1	0	0	1

Lectura recomendada

Para la resolución del cálculo ved:

Francisco Javier Martínez Méndez. *Recuperación de información en la red.*

Siguiendo el ejemplo, la palabra *museo* que aparece en cada documento no tiene mucha utilidad para discriminar el documento 1 del resto, en cambio Picasso, Barcelona y todas las palabras que aparecen poco serían discriminatorias. Para medir este valor de discriminación se propone la *frecuencia inversa*. El peso de una palabra aumenta si aparece pocas veces en un documento y disminuye si aparece a menudo en el resto de documentos.

Pesos de cada palabra

Museo	0		Borne	0,602
Picasso	0,301		Juguete	0,602
Figueras	0,301		Famosos	0,602
Barcelona	0,602		Dalí	0,602
Barrio	0,602			

c) **Valor de discriminación del término:** G. Salton (1989), a partir de la idea de que los vocablos de un texto se clasifican según su capacidad para discriminar unos documentos de los demás en una colección, ideó un sistema de indicación, conocido como el **modelo de valor de discriminación**, que atribuye el peso o valor más alto a aquellos términos que causan la máxima separación posible entre los documentos de una colección. Es decir, el valor de un término depende de cómo varía la separación media entre los documentos cuando a un término se fija una identificación de contenido. Por lo tanto, las mejores palabras son aquellas que consiguen la distancia mayor. El análisis del **valor de discriminación** consigna una función específica en el análisis de contenido a las palabras simples, a las yuxtapuestas, a las frases y a los grupos de palabras.

El valor de discriminación de un término se define como la medida de los cambios en la separación espacial, que se manifiesta cuando una palabra cualquiera es asignada a una colección como término de indización para representar mejor las diferencias que pueda haber entre los documentos. Precisamente la asignación reduce la densidad espacial y, por el contrario, un discriminador pobre incrementa la densidad espacial. De este modo, si primero se calculan las densidades espaciales y se atribuyen a cada término, es posible especificar los términos en orden decreciente según sus valores de discriminación (Gil y Rodríguez, 1996).

Lectura complementaria

Si queréis ampliar la información sobre el sistema de indización ideado por Salton, podéis leer la obra siguiente:

I. Gil Leiva; J. V. Rodríguez Muñoz (1996). "Tendencias en los sistemas de indización automática. Estudio evolutivo". *Revista Española de Documentación Científica* (vol. 19, n.º 3, pág. 273-291).

5) Métodos lingüísticos

Los primeros analizadores lingüísticos son de la década de 1960-1970. Su aportación al análisis del contenido es primordial, ya que permiten analizar el texto en tres niveles de profundidad: palabra, frase y texto.

Cada uno de estos niveles es analizado por módulos del programa basados en diferentes disciplinas:

- Palabra: morfología.
- Palabra dentro de la frase: sintaxis.
- Palabra dentro del texto: semántica.

a) Análisis morfológico: es la rama de la gramática tradicional que estudia la forma de las palabras, independientemente de sus relaciones o funciones dentro de la frase. La forma de las palabras se distribuye en diversas categorías llamadas *partes de la oración*, por ejemplo nombres, adjetivos y verbos. También estudia las variaciones que pueden experimentar por motivo de los accidentes gramaticales como el género, el número, la declinación o la conjugación.

El análisis morfológico en indización automática se basa en reglas gramaticales y diccionarios de sufijos y afijos.

Su intervención en la indización automática es muy relevante en la formación de la lista de palabras vacías y una primera agrupación de las palabras, ya que:

- Asigna la categoría gramatical de cada una de las palabras del texto (como nombre común, verbo, adjetivo, adverbio o artículo), según cuál sea el género, el número, el tiempo y el modo. Con ello se reduce el abanico de palabras susceptibles de contener información, ya que detecta las palabras vacías y las flexiones derivadas de los accidentes gramaticales (como el género y el número). En la red hay aplicaciones gratuitas que analizan morfológicamente un texto; los ejemplos más comunes son Daedalus en castellano o Thera en castellano y catalán.

Ejemplo

Analizaremos un texto sobre fertilidad con Daedalus:

Etiquetado morfosintáctico de texto

Idioma: Español

Texto para procesar:

Fértil fértiles Fertilidad Fertilización fertilizaciones fertilizado fertilizada Fértilmente Fertilizar

Intentar analizar las palabras desconocidas

Desambiguación morfológica

Aceptar

Resultado

Fértil	Adjetivo masculino singular postnominal (APMS--NN3) <input type="checkbox"/> <i>Lema fértil</i> Adjetivo femenino singular postnominal (APFS--NN3) <input type="checkbox"/> <i>Lema fértil</i>
fértiles	Adjetivo masculino plural postnominal (APMP--NN3) <input type="checkbox"/> <i>Lema fértil</i> Adjetivo femenino plural postnominal (APFP--NN3) <input type="checkbox"/> <i>Lema fértil</i>
Fertilidad	Nombre común femenino singular (NCFS--N-N3) <input type="checkbox"/> <i>Lema fertilidad</i>
Fertilización	Nombre común femenino singular (NCFS--N-N2) <input type="checkbox"/> <i>Lema fertilización</i>
fertilizaciones	Nombre común femenino plural (NCFP--N-N2) <input type="checkbox"/> <i>Lema fertilización</i>
fertilizado	Verbo léxico transitivo singular participio masculino (VPMS--TL-N2) <input type="checkbox"/> <i>Lema fertilizar</i>
fertilizada	Verbo léxico transitivo singular participio femenino (VPFS--TL-N2) <input type="checkbox"/> <i>Lema fertilizar</i>
Fértilmente	Adverbio (E--X3) <input type="checkbox"/> <i>Lema fértil</i> <input type="checkbox"/> <i>Capitalización canónica: fértilmente</i>
Fertilizar	Verbo léxico transitivo infinitivo (VN---OTL-N2) <input type="checkbox"/> <i>Lema fertilizar</i>

(final de frase)

- Asigna la raíz común de diversas palabras. El mismo análisis morfológico procede a la lematización (*stemming*, en inglés), que consiste en localizar la raíz común de un grupo de palabras (fértil, fertilidad, fertilizar, fertilización). La lematización permite contar todas las veces que un concepto aparece en el texto, aunque se exprese con formas diferentes.

Ejemplo

En la recuperación, cuando el usuario busque por *fértil* recuperará todos los demás conceptos, ya que se encuentran indizados conjuntamente.

Sin lematizador	Con lematizador
fértil: un resultado fértiles: un resultado fértilmente: un resultado	fértil: tres resultados

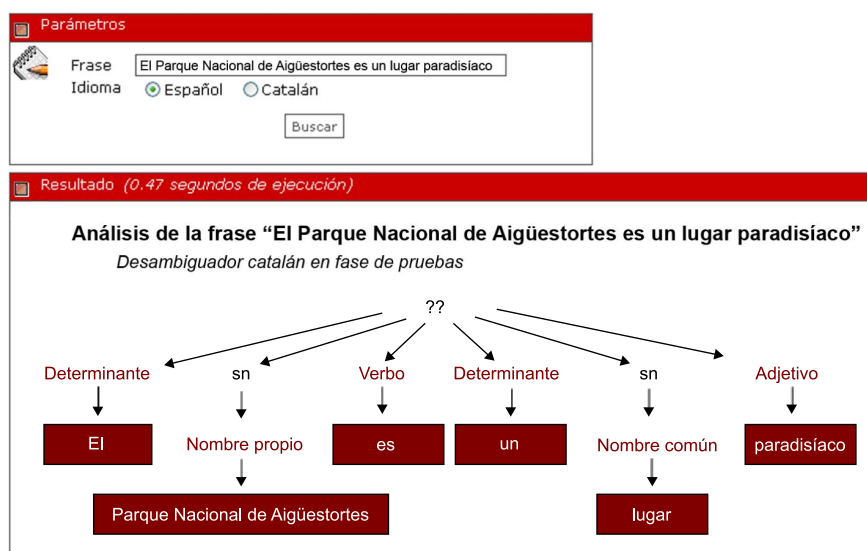
Gracias a este análisis, los cálculos estadísticos se pueden elaborar sobre un conjunto ya seleccionado de términos.

b) Análisis sintáctico: estudia la función o las asociaciones de las palabras dentro de la frase, lo que se llama *estructura sintáctica*. Identifica los componentes de la frase: sujeto, predicado, complementos u objetos y su concordancia. El programa de análisis sintáctico toma la frase e intenta reconocer la estructura con la que se ha construido dentro de un conjunto de análisis posibles conocidos como *gramática*.

Gracias a este análisis, puede desambiguar gramaticalmente los términos no resueltos en la etapa morfológica. Los analizadores sintácticos (*parsing*) se basan en analizadores morfológicos de la etapa anterior, gramáticas y diccionarios.

Ejemplo

Análisis sintáctico de la frase “El Parque Nacional d'Aigüestortes es un lugar paradisíaco” con Thera:



El análisis sintáctico es útil para los polisémicos o alguna palabra que por su similitud podría ser considerada vacía por el programa.

Ejemplo

Analizamos la frase: “Esta nota para mañana”⁸, en la que morfológicamente cada palabra tiene diversas opciones, pero que el análisis sintáctico puede aclarar. Por ejemplo, *nota* puede ser un nombre común femenino (la nota escrita) o una forma verbal (3.^a persona del singular del presente de indicativo o 2.^a persona del singular del imperativo del verbo *notar*), pero dentro de la frase, por su posición, sólo puede ser considerada como nombre común y, en consecuencia, sería una palabra que habría que indizar, no una palabra vacía. Sin embargo, cabe decir que actualmente no todos los analizadores sintácticos pueden llegar a estas prestaciones y cometen errores.

⁽⁸⁾El ejemplo está extraído de I. Gil (2008). *Manual de indización. Teoría y práctica*. Gijón: Trea.

c) **Análisis semántico.** La semántica es la rama de la lingüística que estudia el significado de las palabras. Su utilidad reside en desambiguar sinónimos, polisémicos y vincular anáforas con el término al que hacen referencia. Sin embargo, esta parte del análisis es la más compleja para el software actual. Los analizadores semánticos se basan en diccionarios léxicos, redes semánticas, tesauros y ontologías.

Para la indización automática son muy interesantes las bases de datos del tipo WordNet, que dan información léxica y semántica de los términos. Un algoritmo vinculado a la WordNet puede resolver casos de sinonimia o hiperonimia, ya que la base de datos proporciona relaciones básicas de significado.

WordNet y Euro WordNet son bases de datos multilingües en diversos idiomas (Euro WordNet comprende el neerlandés, el italiano, el castellano, el alemán, el francés, el checo y el estonio). Cada idioma diseña su WordNet pero mantiene la misma estructura que la inicial de Princeton en cuanto a los términos sinónimos con relaciones semánticas de relación de equivalencia entre ellos. Las diferentes WordNet están conectadas entre ellas y proporcionan acceso a una ontología compartida. En acceso gratuito sólo se encuentra la Wordnet de Princeton.

Los tesauros participan en el análisis semántico en que aportan las relaciones semánticas de equivalencia, jerarquía y asociación de cada palabra clave seleccionada. Las ontologías aportan, además, el detalle en cada tipo de relación. Un ejemplo es todo tipo de relaciones de asociación entre dos descriptores.

3.2. El listado de palabras clave en la indización

El programa de indización automática reconoce palabras, para ser más exactos, cadenas de caracteres entre un espacio y el siguiente. Por lo tanto, reconoce palabras, no expresiones formadas por más de una palabra. Decimos, pues, que el listado de palabras clave indiza unitérminos.

Las **ventajas** que tienen las listas de palabras clave en la indización son las siguientes:

- Indización inmediata, en cuestión de segundos.
- Coherencia de indización exacta entre diversos SID. Dos SID con el mismo programa indizador llegarán al mismo resultado.
- Indización exhaustiva y tan específica como lo sea el contenido del documento indizado.
- Actualización rápida de la base de datos.
- Riqueza terminológica extraordinaria: evolución automática de la terminología, paralela a la misma evolución del conocimiento y de la ciencia.
- Costes mínimos de construcción y mantenimiento.

Las redes semánticas

Estas redes son bases de datos que representan el conocimiento de una lengua o ámbito de una manera estructurada.

Unitérmino

Recordemos que los unitérminos son el elemento con significado más pequeño de un término de indización. *Lluvia* es un unitérmino. *Lluvia de estrellas* podría ser un descriptor de un tesoro formado por dos unitérminos: *lluvia* y *estrellas*.

Uno de los **inconvenientes** que tienen las listas de palabras clave en la indicación es que las palabras con significado se indizarán tal como salgan en el texto. Por lo tanto, puede indizar:

- El mismo término en diferentes idiomas, por ejemplo: *fuentes de información, fonts d'informació, information resources*, en función de la lengua del texto. Y no tiene relaciones de equivalencia entre los términos, como tienen los tesauros multilingües.
- Los errores ortográficos (*tesauris, thesaurus*).
- Singular, plural, masculino, femenino, entre otros (encabezamiento/encabezamientos, lista/listado).
- Siglas y conceptos sin distinción (SID/servicio de información y documentación; UE/Unión Europea).
- Sinónimos (lista de palabras vacías/*stop word list*).
- Homónimos (*Hierro*, y no sabemos si hace referencia a la isla o al elemento químico).

Como no hay control sobre el vocabulario, no controla la ambigüedad del lenguaje natural.

3.3. El listado de palabras clave en la recuperación

La indización automática sustenta los buscadores; en consecuencia, los usuarios de Google y Yahoo conocen de una manera implícita las ventajas y los inconvenientes de la recuperación de documentos con este lenguaje. Tal como dice Isidoro Gil (2008, pág. 108), el uso de Internet está convirtiendo a cada usuario en un *paradocumentalista* en potencia. Los usuarios son conscientes de que hay que escoger con cuidado las palabras clave, que deben ser cuanto más específicas mejor, poner el texto entre comillas, restringir la búsqueda con operadores booleanos, especificar fechas o ir añadiendo términos, entre otros, si queremos evitar el ruido documental.

Las **ventajas** que tienen las listas de palabras clave en la recuperación son las siguientes:

- Las búsquedas pueden ser muy precisas (todos los términos significativos están presentes).
- Son muy fáciles de usar para el usuario.
- Es el lenguaje más actualizado, evoluciona automáticamente al mismo tiempo que la terminología del documento. Podemos buscar *iPhone, teléfonos 3G, iPod, spam, spyware*, entre otros, términos que no constarían en ningún otro lenguaje documental, dado que el proceso de actualización es lento.

- La inmediatez: el documento llega y, en cuestión de segundos, está indizado y disponible para la búsqueda.

En cuanto a los **inconvenientes**, la indización automática se caracteriza por la posibilidad de indizar, si se desea, el texto entero del documento, lo que genera muchos puntos de acceso por materia, muchas palabras clave. Cuando realizamos una búsqueda, el buscador nos puede devolver miles de registros, lo que es un problema molesto (ruido documental). Sin embargo, esta característica también minimiza problemas potencialmente peligrosos (el silencio documental), como la pérdida de conceptos y la sinonimia.

Los principales inconvenientes son los siguientes:

- Pérdida de conceptos. Como el programa indiza unitérminos, los conceptos expresados de una manera compuesta se separan en dos unidades y pierden el sentido.
- No reconoce la sinonimia. El programa no puede reconocer los sinónimos del mismo concepto. Si en el texto aparecen todos estos sinónimos, el programa no detectará ninguna semejanza entre ellos y los indizará todos: *cautivar, encantar, hipnotizar, sugerir*.
- No reconoce la anáfora y la elipsis.
- No permite la univocidad.
- Aparece ruido y silencio documentales producidos por la ambigüedad propia del lenguaje natural (fenómenos de sinonimia, polisemia y homonimia).
- No aparece ningún tipo de control sobre variantes ortográficas, sobre plurales, adjetivos, verbos, siglas, entre otros, ni se ejerce ningún control sobre los errores ortográficos o de impresión que pueden aparecer en los documentos.
- Hay complejidad en el momento de plantear la estrategia de búsqueda; se necesita una agrupación de sinónimos completa, el uso de truncamientos y de operadores de proximidad.

Precisamente para minimizar los inconvenientes en la recuperación, los SID acostumbran a complementarse con otros lenguajes documentales preferentemente controlados, como un tesoro o una lista de encabezamientos de materia.

Actualmente la mayoría de catálogos en línea de bibliotecas universitarias nos permiten realizar búsquedas por diversos campos, de los cuales destacamos especialmente el de *palabra clave de materia*. Esta opción descompone el texto

Ejemplo de pérdida de conceptos

Universidad de Barcelona, por ejemplo, se separa en universidad y Barcelona. Separar conceptos puede dar lugar a coordinación falsa en la recuperación como en este caso, en el que recuperamos todas las universidades de la ciudad de Barcelona: UB (Universidad de Barcelona), UPF (Universidad Pompeu Fabra), URL (Universidad Ramon Llull), entre otras.

Ejemplo de estrategia de búsqueda

Para buscar documentos sobre la Unión Europea, por ejemplo, habría que pensar en todos sus sinónimos y siglas como UE / Unión Europea / European Union / CEE / Comunidad Económica Europea.

(incluso el encabezamiento construido con una lista de encabezamiento de materia, es decir, con un lenguaje controlado y humano) en una sucesión de palabras clave, es decir, en un lenguaje libre y automático.

Actividades

Listado de descriptores libres

1. Indizad con un listado de descriptores libres el resumen siguiente con los tres grados de exhaustividad.

López Alonso, C.; Séré, A. (eds.) (2003). *Nuevos géneros discursivos: los textos electrónicos*. Madrid: Biblioteca Nueva.

Esta obra describe la transformación de determinados discursos sociales ante la revolución de las nuevas tecnologías de la información y de la comunicación (TIC). En efecto, Internet forma parte de nuestra vida diaria y ya no es fácil prescindir de este medio tecnológico. Esta obra, que consta de tres partes, analiza estos nuevos géneros discursivos:

En la primera parte, se describen los distintos textos electrónicos –el correo, los chats, los foros– y otros escritos en la red, como prensa, newsletters, E-zinc, etc.

En la segunda, se introduce al lector en la génesis y evolución de los lenguajes informáticos y sus diferentes tipos, especialmente los lenguajes de marcado y los modelos hipermedia.

La tercera, finalmente, se centra en los nuevos productos hipertextuales en el campo del discurso de transmisión de conocimientos con muy variadas utilizaciones: la enseñanza a distancia o textos de consulta como los diccionarios.

2. Indizad el resumen anterior para dos bases de datos diferentes: una genérica y la otra especializada en documentación.

3. Analizad detenidamente el ejemplo de los tres analistas que indizan el documento sobre los cataros y decid por qué motivos son tan diversos los resultados.

Analista A	Analista B	Analista C
Baja Edad Media cátaros Francia movimientos cristianos disidentes	catarismo herejías órdenes monásticas religión	albigense cismas historia medieval Languedoc

4. Elaborad una lista con las operaciones que hay que llevar a cabo sobre el léxico para transformar una lista de descriptores libres en un lenguaje controlado no codificado.

Listado de palabras clave

1. Indizad el texto siguiente simulando el funcionamiento de una lista de palabras clave:

Piero de Benedetto dei Franceschi (Sansepolcro, Toscana, 1416-1492) fue un pintor del *quattrocento* italiano conocido por su alias Piero della Francesca. También fue geómetra y matemático.

Es uno de los personajes principales y fundamentales del Renacimiento. Piero della Francesca es un personaje itinerante: es una figura que encontraremos en Ferrara entre 1447 y 1448 trabajando con Lionello d'Este; en Rimini para Sigismondo Malatesta en torno al 1450; en Roma para el papa Pío II entre 1458 y 1459, y en Urbino para Federico de Montefeltro en diversas ocasiones. Su actitud itinerante se suele comparar con la de Leon Batista Alberti. El autor nace en Borgo San Sepolcro (al norte de Florencia, en la zona de la Umbría) en 1416. El autor proviene de una familia de mercaderes y, por eso, sabía aritmética, cálculo, álgebra, geometría y contar con el ábaco. Después de la formación que tuvo para llevar el negocio familiar se formó con un maestro local que se llamaba Antonio di Anghiari para ser pintor. En aquellos momentos, había muchos estilos: todavía estaba vigente el linealismo y el lirismo de Fra Angelico, Panozzo Gonzolo o Filippo Lipi y, por otra parte, estaba el realismo geométrico de Paolo Uccello.

Tiene un estilo pictórico muy particular y, por lo tanto, es fácil de identificar. Utiliza una luz muy diurna y unos colores muy brillantes, que le provienen de su contacto con Domenico de Veneziano. En su obra siempre está presente el peso de la geometría, que implica, por una parte, utilizar la perspectiva lineal y, por la otra parte, reducir las figuras a la esencia. Las figuras de Piero della Francesca son muy estáticas, poco nerviosas, lo que va al revés que en el resto de pinturas renacentistas de Florencia, que a medida que avanzamos en el tiempo cada vez son más dinámicas. También vemos que sus figuras son poco expresivas y monolíticas. Longhi, cuando habla de Piero, dice que sus figuras son *columnes*.

"Piero della Francesca", *Wikipedia*

- a) Diseñad las opciones del programa que simularéis (partes del texto, signos y símbolos).
- b) ¿Cuántas palabras vacías habéis localizado?
- c) ¿Qué inconvenientes planteará en la recuperación?

Glosario

análisis morfológico *m* Rama de la gramática tradicional que estudia la forma de las palabras, independientemente de sus relaciones o funciones dentro de la frase. La forma de las palabras se distribuye en diversas categorías llamadas *partes de la oración*, que son, entre otros, nombres, adjetivos y verbos. También estudia las variaciones que pueden experimentar por motivos de los accidentes gramaticales como el género, el número, la declinación o la conjugación.

análisis semántico *m* Rama de la lingüística que estudia el significado de las palabras. Su utilidad reside en desambiguar sinónimos, polisémicos y vincular anáforas con el término al que hacen referencia.

análisis sintáctico *m* Estudio de la función o de las asociaciones de las palabras dentro de la frase, lo que se llama *estructura sintáctica*. Identifica los componentes de la frase: sujeto, predicado, complementos u objetos y su concordancia. El programa de análisis sintáctico toma la frase e intenta hacer evidente la estructura con la que se ha construido dentro de un conjunto de análisis posibles conocidos como *gramática*.

automatic indexing *f* Ved **indización automática**.

clasificación social *f* Ved **indización social**.

coordinación falsa *f* Recuperación inesperada (y errónea) que a pesar de contener los elementos de la búsqueda corresponden a temas diferentes.

descriptor libre *m* Término de indización propio del lenguaje documental. Lista de descriptores libres.

etiquetaje colaborativo *m* Ved **indización social**.

etiquetaje social *m* Ved **indización social**.

folksonomía *f* Neologismo (Thomas van der Wal) resultado de la fusión de *folk* (gente, popular) y *taxonomía* (gestión de la clasificación), lo que da como resultado una “indización gestionada popularmente”.
sin. **indización social**.

frecuencia de aparición *f* Ley de Zipf. El psicólogo y psicolingüista estadounidense George Kingsley Zipf (1949) propuso que el número de veces que un término aparece en un texto (frecuencia) y la posición que ocupa en una lista de palabras más o menos frecuentes (rango) es constante. Para Zipf los escritores usan un abanico reducido de palabras, preferentemente cortas y su longitud es inversamente proporcional a la frecuencia.

frecuencia inversa *f* Sparck Jones (1972) puso de manifiesto la capacidad de discriminación de un término frente a otro. Esta discriminación tiene que ser vista en el conjunto de la colección, no en un solo documento. Hay que comparar las palabras clave entre los documentos del fondo para detectar cuáles son realmente discriminatorias.

indización automática *f* Método por el que un ordenador aplica un algoritmo (o programa) al título, resumen o texto completo del documento con el fin de identificar los términos que puedan representar la materia y usar como términos de indización y recuperación en un índice o lista.
en *automatic indexing*.

indización social *f* Tipo de indización en red en el que los internautas asignan etiquetas con descriptores a los recursos web. La asignación de las etiquetas se hace sin ánimo de lucro, no se busca un beneficio económico sino beneficiarse de búsquedas mejores. Los descriptores escogidos no se supervisan ni tienen ninguna estructura semántica.

sin. **clasificación social, etiquetado colaborativo, etiquetado social, folksonomía**
en *tagging*.

lematización *f* Localización de la raíz común de un grupo de palabras.
en *stemming*.

listado de descriptores libres *f* Lenguaje documental. Vocabulario monolingüe de términos de indización ordenados alfabéticamente. Estos términos son escogidos por el analista, sin verificar si existen o cómo se introducen en una lista previamente establecida.

listado de palabras clave *f* Lenguaje documental. Vocabulario ordenado alfabéticamente de los términos con carga significativa extraídos de un documento mediante un programa informático.

machine-aided indexing *m* Ved **método de indización semiautomática**.

método de indización semiautomática *m* El programa selecciona posibles descriptores procedentes de un tesoro o de una lista controlada y un documentalista acepta o deniega la propuesta.
en *machine-aided indexing*.

método estadístico *m* Primera aproximación a la indización automática. Calcula el peso (ponderación) de las palabras: ni las palabras más repetidas (por vacías) ni las menos repetidas (por específicas) son adecuadas para ser seleccionadas. Los métodos estadísticos aplicados son de tres tipos: frecuencia, frecuencia inversa y discriminación. Se pueden usar solos o en combinación.

método lingüístico *m* Método que permite analizar el texto en tres niveles de profundidad: palabra (análisis morfológico), frase (análisis sintáctico) y texto (análisis semántico).

OCR *m* Ved **reconocimiento óptico de caracteres**.

palabra vacía *f* Palabra sin significado como, por ejemplo, artículos, pronombres, preposiciones o conjunciones, que es filtrada antes o después del procesamiento del texto. Son palabras muy frecuentes pero que aportan poco valor de contenido semántico al texto y también poca ayuda en la recuperación, ya que el usuario no las utiliza en la búsqueda y tampoco son recogidas en el fichero inverso de la base de datos.
sin. **lista de detención**
en *stop word list*.

palabra clave *f* Término de indización propio del lenguaje documental. Lista de palabras clave.

reconocimiento óptico de caracteres *m* Programa que transforma la imagen de la página escaneada en texto electrónico (de la expresión inglesa *optical character recognition*, OCR).

stemming *m* Ved **lematización**.

stop word list *f* Ved **palabra vacía**.

tagging *m* Ved **indización social**.

valor de discriminación del término *m* Método estadístico aplicado a la indización automática, desarrollado por Salton (1985). La medida de los cambios en la separación espacial que se manifiesta cuando una palabra cualquiera es asignada en una colección como término de indización para representar mejor las diferencias que pueda haber entre los documentos. La asignación reduce la densidad espacial y, por el contrario, un discriminador pobre incrementa la densidad espacial. De este modo, si primero se calculan las densidades espaciales y se atribuyen a cada término, es posible especificar los términos en orden decreciente según sus valores de discriminación.

Bibliografía

Alonso Berrocal, J. L.; Figuerola, C. G.; Zazo Rodríguez, A. (2009). *Recuperación avanzada de la información*. Salamanca. [Fecha de consulta: 31 de julio de 2009].

Cid, P.; Cuadrado, M.; Aguiriano, C. (1999). *Fonaments de llenguatges documentals* [documento electrónico]. Barcelona: UOC.

Gil Leiva, I. (2008). *Manual de indización. Teoría y práctica*. Gijón: Trea.

Gil Leiva, I.; Rodríguez Muñoz, J. V. (1996). "Tendencias en los sistemas de indización automática. Estudio evolutivo". *Revista Española de Documentación Científica* (vol. 19, n.º 3, pág. 273-291).

Hassan Montero, Y. (2006). "Indización social y recuperación de información". *No Solo Usabilidad* (n.º 5). ISSN 1886-8592.

Méndez, E.; Moreiro, J. A. (1999). "Lenguaje natural e indización automatizada". *Ciencias de la Información* (pág. 11-24).

Quintarelli, E. (2005). "Folksonomies: power to the people". ISKO Italy-UniMIB meeting: Milán. [Fecha de consulta: 31 de julio de 2009].

Salton, G. (1988). *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Reading: Addison-Wesley.

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

Páginas web interesantes

Cibermetría

Cybermetrics. Recursos presentados por el departamento de CINDOC dentro del Consejo Superior de Investigaciones Científicas; artículos, seminarios, etc.

Blog spot. Acceso a algunos artículos sobre cibermetría.

Cybermetrics. Software disponible relativo a la cibermetría.

Webometrics and organizations. Enlaces hacia publicaciones y grupos de trabajo sobre cibermetría.

Applications of informetrics to information retrieval research. Artículo que relaciona la cibermetría con la recuperación de información.

The diameter of the world wide web. Artículo sobre la medición de la web realizada por los autores.

Inteligencia artificial

World Scientific: Connecting great minds. Vínculos a varias publicaciones periódicas sobre campos especializados como inteligencia artificial o sistemas neuronales.

AI international. Vínculos a entidades investigadoras sobre el tema como universidades y laboratorios, así como vínculos a congresos que versen sobre el mismo como la AAI o las PRICAI.

Otras páginas web de interés

The Open Archives Initiative.

GOOGLE-WATCH. Página crítica sobre el sistema de recuperación de Google (el aparente uso comercial del orden de la recuperación).

RDF: modelo de metadatos flexible. Página sobre metadatos, con recursos web relacionados.

DUBLINCORE: Metadata Initiative (DCMI). Iniciativa de metadatos para páginas web, conversión de formatos... Proyectos, software, grupos de trabajo.

Pew Internet & american life. Web dependiente del Pew Research Center, con acceso a publicaciones de actualidad sobre internet.

z39.50 International Standard Maintenance Agency. Página sobre el mantenimiento de este protocolo de búsqueda e intercambio de información.