

## TEMA 2: EL PROCESO DOCUMENTAL: DEL ANÁLISIS A LA RECUPERACIÓN DE LA INFORMACIÓN

- 2.1. Selección y análisis documental.
  - 2.1.1. Criterios y elementos que intervienen en la selección de información.
  - 2.1.2. Técnicas de análisis documental.
- 2.2. La descripción bibliográfica: la ficha bibliográfica, partes y tipos de ficha.
- 2.3. La catalogación.
- 2.4. El proceso de indización.
  - 2.4.1. Ventajas. Criterios a tener en cuenta.
  - 2.4.2. Los descriptores. Tipos de descriptores.
  - 2.4.3. Los tesauros.
- 2.5. El resumen documental. Metodología para realizar un resumen y utilidades.
- 2.6. Almacenamiento y recuperación de la información.
- 2.7. Lenguajes documentales. Tipología.
- 2.8. Bases de datos. Tipología.
- 2.9. Bases de datos para los profesionales de la comunicación.

### 2.1. Selección y análisis documental.

El primer paso que tiene que dar un documentalista es la selección del material con el que va a trabajar el centro de documentación. A la hora de seleccionar se siguen unas pautas marcadas por el director del centro, para que cuando haya que recuperar la información no se planteen demasiados problemas. También hay que pensar en *la vida* que pueden tener los documentos. La mayoría de los documentos científicos quedan obsoletos cuando pasa un período corto de tiempo. No sucede así con las humanidades.

El constante aumento de documentos dio origen a una nueva ciencia, encuadrada dentro de la documentación y que hay que tener muy en cuenta a la hora de seleccionar documentos: La **BIBLIOMETRÍA**. <http://clio.rediris.es/clionet/articulos/bibliometria.htm>

#### 2.1.1. Criterios y elementos que intervienen en la selección de información.

##### *Criterios*

- ¿Cómo determinados el valor de un documento?
- ¿Qué significa actualidad para la documentación?

##### *Elementos que intervienen en la selección de las informaciones*

- (desde el punto de vista del documentalista)
- (desde el punto de vista del comunicador)

#### 2.1.2. Técnicas de análisis documental.

Análisis Documental: *consiste en extraer de un documento un conjunto de palabras que sirvan para la representación condensada del mismo.*

Los niveles de AD son:

- Nivel formal o externo. Forma (soporte)
- Nivel de contenido o interno. Contenido (información)

El análisis externo se efectúa sobre el continente, en base o soporte documental y el interno se refiere al contenido o mensaje del documento.

### 2.2. La descripción bibliográfica.

Inicialmente, las bibliotecas cumplieron una función de depósito, de tal forma que sólo se controlaban y anotaban unos cuantos datos elementales, como el autor o el título, pero según crecía el número de volúmenes se fueron plasmando más datos, ya que con los elementales no era suficiente. Se necesitó entonces una técnica que recogiera todos esos datos en un **asiento bibliográfico** para posteriormente ordenarlos dentro de un catálogo.

Los códigos de catalogación y las Reglas Angloamericanas han concretado el campo de actuación del análisis formal en dos conceptos: descripción bibliográfica y catalogación.

Descripción bibliográfica: *Es la operación encargada de escoger todos aquellos elementos aparentes y convencionales que posibiliten la identificación precisa y formal de cada documento en una colección determinada. Indica, pues, el número, orden y naturaleza de cada documento.*

Ante la insuficiencia lexicográfica, sobre todo para los nuevos soportes documentales, algunos teóricos prefieren llamar a este concepto "descripción documental".

#### La ficha bibliográfica: partes y tipos de ficha.

Para realizar la descripción bibliográfica es necesario una ficha, bien se trate de un sistema manual, bien de uno automático. La ficha suele ser de cartulina flexible, rectangular y con unas medidas internacionales (12,5 x 7,5 cm) y perforada en el centro, a ocho mm. del borde inferior, para permitir el paso de la varilla que la sujete al fichero, evitando su desorden y deterioro.

#### Partes

1. Encabezamiento: Palabra o palabras que colocadas a la cabeza de los asientos bibliográficos sirven de punto de acceso al documento. Los puntos de acceso pueden ser múltiples (autor, título, materia).
2. Descripción: Zona donde se registran los elementos de la descripción, como título, edición, datos de publicación, serie, ISBN, etc.
3. Registro de fichas secundarias: Ya que un documento puede estar representado por un número ilimitado de fichas, en función de los posibles puntos de acceso. Se utiliza este registro a modo de relación en la que se enumeran los distintos encabezamientos de los asientos secundarios por los que puede ser localizado un documento en el catálogo.
4. Número de registro: Precedido de la abreviatura R, indica el número de orden que corresponde al documento en el libro de registro de entrada en la biblioteca.
5. Signatura: Señala el lugar que ocupa el documento en los estantes de la biblioteca. Se trata de la signatura topográfica.

Para acceder de una forma rápida y universal a la información bibliográfica ha sido necesario un esfuerzo normalizador a nivel internacional (FID, FIAB, ISO, UNESCO), y nacional (AENOR Y AFNOR) para emitir normas y soportes físicos que sean entendidos por la mayoría de los investigadores y usuarios.

La descripción bibliográfica se rige por las normas ISBD (International Standard Bibliographic Description-Normas Internacionales de Descripción Bibliográfica), con las que se intenta facilitar la comunicación internacional de información bibliográfica, a través de intercambio de asientos procedentes de diferentes fuentes, facilitando la interpretación de dichos asientos, pese a las barreras lingüísticas y facilitando la conversión de los asientos bibliográficos a formato legible por ordenador.

### 2.3. La catalogación.

Una vez elaboradas las fichas bibliográficas es necesario ordenarlas siguiendo una normativa. Se dispone de muchos datos en la ficha bibliográfica, pero sólo si están bien catalogados se podrá llegar a ellos.

Catalogar es, según Teresa Malo de Molina, diferenciar para localizar los documentos. *"La catalogación de los documentos nos proporciona una representación de los mismos, con todos los datos que nos permiten comparar. Esta representación es a lo que se le llama registro bibliográfico"*.

#### Concepto

*El proceso por el que son transferidos conforme a determinadas reglas ciertos datos técnicos de un documento a un soporte documental.*

Esta operación completa el asiento bibliográfico dotándole de encabezamiento, registro de fichas secundarias, etc. Su fin es facilitar la identificación física de los documentos y su producto final es el catálogo, que actuará de intermediario entre los usuarios y el fondo bibliográfico. El proceso de catalogación se centra en dos fases:

1. Dedicada al estudio y observación de los datos del documento en cuestión, lo que permite determinar el punto de acceso.
2. Concreta la forma de encabezamiento.

**Puntos de acceso** por los que puede ser identificado un documento: autor personal, autor corporativo y título.

*Autor personal:* Toda persona responsable total o parcialmente del contenido de una obra. Hay que distinguir entre personal único y autor múltiple.

1. Un solo autor: Cervantes, Miguel de
2. Un autor y colaboradores: uno de ellos es el responsable principal.
3. Dos o tres autores: ninguno aparece como responsable principal. Se escoge el que figura en primer lugar.
4. Formas del nombre: se tomará el nombre por el que comúnmente se le conoce como autor. Ej. Pablo Neruda (No Neftalí Ricardo Reyes Basoalto)

*Autor corporativo:* Entidades o entes colectivos, es decir, organizaciones, instituciones, gobiernos, etc. Para la entrada se destacará tipográficamente el nombre de la misma. ESPAÑA. Ministerio de Hacienda

Título: Se hará destacando tipográficamente la primera palabra que no sea artículo. POEMA DEL CID.

### 2.4. El proceso de indización.

La indización consiste en extraer los conceptos representativos del contenido de un documento con la ayuda de un lenguaje documental o lenguaje controlado.

Unesco: *"describir y caracterizar un documento con la ayuda de representaciones de los conceptos contenidos en dicho documento"*.

Se considera la actividad más importante de todo el proceso documental.

Se pueden emplear materias, palabras claves o descriptores.

La indización no se limita sólo a detectar los vocablos presentes en el documento, sino también su traducción e interpretación para pasar del lenguaje natural al lenguaje documental.

El indizador, cuando ya tiene el documento original o su expresión condensada, retiene unas cuantas nociones que representan su contenido con la máxima fidelidad.

Una correcta indización no lleva a descubrir lo que dice el documento indizado, pero sí desvela sobre qué trata y esto en el mundo de la documentación es suficiente muchas veces.

Los teóricos de la documentación consideran que la indización se realiza, casi siempre, después de la condensación, ya que si el resumen es correcto y lo ha realizado un experto, resulta más fácil y más operativa la operación de indizar y ésta la puede llevar a cabo un documentalista sin necesidad de ser un especialista en el tema.

Van Slype *diferencia entre clasificación e indización*.

- Considera la Clasificación como la expresión más general del contenido (destaca el tema general) y se lleva a cabo a través de lenguajes de estructura jerárquica.
- La Indización analiza el documento y los conceptos que lo constituyen, asigna descriptores, generalmente extraídos de un Tesoro, para describir el contenido conceptual del documento.

#### **Indización manual y automática.**

##### 1. Humana, inteligente, manual

Exige un gran esfuerzo de síntesis para extraer las palabras claves más significativas del contenido de un texto.

Se trataría de aislar los conceptos más representativos, aquellos que pudieran interesar en el tiempo y en el espacio al usuario.

##### 2. Automática

Trabaja con unitérminos, de modo que va leyendo una a una todas las palabras que componen el texto, a excepción de los contenidos en un tesoro negativo de palabras vacías.

El ordenador las rechaza. Se conocen con el nombre de Stop word list.

#### **2.4.1. Ventajas e inconvenientes.**

1. La humana parece más eficaz en cuanto a captación y desestructuración de contenidos, pero presenta como inconveniente la velocidad, ya que es más lenta.
2. La automática: Es más o menos eficaz en función de los programas. No lee el texto de una forma convencional, sino que lo hace de forma secuencial.

#### **Criterios a tener en cuenta.**

1. Exhaustividad: Todos los conceptos básicos, nombres propios importantes, lugares geográficos significativos, han de estar representados en la indización.
2. Concreción: Han de evitarse palabras demasiado genéricas, expresiones vagas o ambiguas o cualquier tipo de generalización
3. Pertinencia: Si no puede retenerse todo, ni lo más importante, hay que ser exigentes en la selección de los vocablos más expresivos y mas significativos.
4. Uniformidad: Es el más difícil y el que exige una actitud mas positiva por parte de quien indiza, sobre todo cuando no existen palabras similares a la escogida por el documentalista.
5. De autor: El que realiza la indización ha de tener sus criterios, pero debe aceptar primero los que impone el centro para el que trabaja. No tiene porque coincidir ni con el del documentalista, ni con el del usuario que solicita la información.
6. Interés del usuario: Esta será siempre la principal misión de la indización si se quiere ofrecer un servicio útil.
7. Estadístico: Frecuencia del uso de un término en el texto, es decir, el número de veces que aparece.
8. Especificidad: Se utiliza en el lenguaje controlado, y expresa la exactitud con la cual unos determinados términos representan a ideas concretas.
9. Precisión: Aquello que mide la habilidad o la aptitud de un sistema de información para encontrar, en respuesta a una pregunta formulada correctamente, unos resultados precisos.

#### **2.4.2. Los descriptores.**

Es un término normalizado o controlado que expresa el contenido significativo del documento.

#### **2.4.3. Los TESAUROS:**

El TESAURO es un instrumento de control terminológico utilizado para trasponer a un lenguaje más estricto el idioma natural de los documentos. Por su estructura, es un vocabulario controlado y dinámico de términos que tienen entre ellos relaciones semánticas y genéricas, y que se aplica a un dominio concreto del conocimiento. El componente esencial de los tesauros son los descriptores, palabras o expresiones corrientes escogidas por el constructor del tesoro para designar los conceptos representativos del documento.

- TESAURO DE LA UNESCO
- TESAUROS DEL CINDOC
- TESAUROS EN ESPAÑOL
- TESAUROS EN INTERNET

#### **2.5. El resumen documental.**

El volumen creciente de documentos académicos, científicos, técnicos y otros documentos informativos e instructivos hace que sea cada vez más importante, tanto para los lectores del documento primario como para los usuarios de los servicios secundarios, que el contenido básico del documento sea identificado de la manera más rápida posible.

Esta identificación rápida se facilita si el autor del documento primario (ayudado por los editores) lo encabeza con un título significativo y un resumen elaborado.

Las directrices básicas sirven para la preparación de resúmenes tanto por los autores como por otras personas y se incluyen reglas específicas para su presentación en publicaciones y servicios secundarios.

- Resumir es una operación que permite disminuir considerablemente el volumen de la información primaria y destacar los aspectos que tienen especial interés para el usuario.
  - Un resumen documental debe ser la representación condensada del contenido de un documento.
  - En esta operación se usa el lenguaje natural, si bien éste sufrirá modificaciones.
  - Un resumen no es una simple enumeración de ideas esenciales, realmente se trata de la representación o reconstrucción de ese texto condensado.
- Para ello, es preciso conocer el texto en profundidad (leerlo todo, al completo) y luego alejarse del texto a la hora de resumirlo (no tenerlo delante)
- El resumen es una representación sintética del contenido de un documento. Pero, hay que distinguir entre:
    - Una anotación es un comentario o explicación breve acerca de un documento o de su contenido; o también una descripción muy breve del contenido, a menudo añadida como una nota a continuación de la referencia bibliográfica del documento.
    - Un extracto es una o más partes del documento seleccionadas para representar el todo.
    - Un resumen de conclusiones, si se necesita, es una exposición breve (generalmente colocado al final del documento), de sus principales hallazgos y conclusiones, que intenta completar la orientación del lector que ha estudiado el texto precedente.

### Metodología para efectuar el resumen.

La persona que va a resumir debe conocer el texto a resumir y para ello empleará su propio estilo. Hay que hacer dos lecturas previas y una posterior del texto para analizar en profundidad el texto original.

- Primera lectura. (Llamada también lectura recuperadora)  
Es realizada por el documentalista de una forma rápida, con el fin de localizar en el texto aquellas partes que contengan información importante para el resumen sobre objetivos, alcance, métodos, resultados, conclusiones o recomendaciones. Esta lectura se efectúa de una sola vez, sin pararse, con un número concreto de retrocesos y sin fijaciones. Mientras leemos tenemos que ir anotando mentalmente o en el margen de papel qué partes del material, o del texto, contienen información sobre estos aspectos:
  - Información clave o básica
  - Objetivos, alcance y métodos
  - Resultados y conclusiones o recomendaciones.
- Segunda lectura. (Llamada también lectura creativa)  
Es aquella en la que el analista vuelve a leer el material identificado durante la primera lectura para seleccionar, extraer, organizar y escribir la información pertinente para el resumen.  
En este punto ya hay que tener en cuenta las normas de estilo, pues tras la lectura se procede a la redacción, cuidando especialmente la frase anotativa.
  - La frase anotativa es la primera que se hace.
  - Debe, en lo posible, contener la idea esencial del texto original si no se encuentra expresada en el título del documento.
  - Esta frase será, aconsejablemente, de unas tres líneas y a lo largo de todo el texto es recomendable redactar párrafos que a ser posible no excedan de las seis líneas.
- Tercera lectura. (Llamada también crítica)  
Es aquella en la que el analista lee el resumen escrito cualitativamente con el objeto de corregirlo, atendiendo a su unidad y concisión y para cerciorarse de que siguió las normas de estilo.

La confección de un resumen depende de sus objetivos, de sus contenidos temáticos y de las directrices del organismo para el que se elabora.

El lenguaje adoptado será correcto; legible (breve y simple); la redacción clara -evitando ambigüedades, palabras mal empleadas o términos no específicos-; además de apropiada (en estilo y tono) y concisa.

En la elaboración del texto final habrá que tomar en consideración las necesidades de los lectores y reflejar la estructura del documento original, manteniendo una postura objetiva.

### Utilidades de un resumen.

1. Sirve de anticipo del documento original, permitiendo a los usuarios decidir sobre la conveniencia o no de consultar el texto original.
2. Actúa a veces como sustituto del documento original, siempre y cuando la información que aporte sea satisfactoria para el receptor.
3. Actualiza los conocimientos del especialista sobre los desarrollos habidos en su campo teórico, ahorrándole tiempo y esfuerzo.
4. Contribuye a la superación de las barreras del lenguaje.
5. Ayuda en la tareas de búsqueda retrospectiva y recuperación de información, cumpliendo un papel importante en la

estructura de los sistemas automatizados, ya que muchas bases de datos incluyen, junto a las referencias bibliográficas, resúmenes que permiten la localización y selección del texto completo o documento original.

6. Facilita la indización, ya que concreta la materia indizable y elimina los problemas del lenguaje, por eso, en muchos Centros de Documentación e Información se han empleado los resúmenes como base para la confección de índices, al contener, por regla general, la macroestructura del texto y la información esencial de su contenido.

## 2.6. Almacenamiento y recuperación de la información.

La documentación tiene como fin último guardar, conservar y almacenar los documentos para que luego se puedan recuperar fácilmente.

La tarea fundamental de la documentación, especialmente en los medios de comunicación, es la recuperación de la información. Los centros de documentación llevan a cabo el proceso documental pensando en el usuario-redactor. Por tanto, desde que llega el documento hasta que lo recibe el redactor se prepara el documento pensando en la *necesidad informativa*.

No siempre coinciden los documentos con los que trabaja el centro de documentación con las necesidades reales del redactor. Para ello, el documentalista deberá conocer la *realidad informativa* y anticiparse a los hechos noticiosos. En ocasiones, implica un riesgo que ha de ser asumido por el director del centro de documentación.

Otra parte de los documentalistas trabaja con documentos no tan urgentes, pero importantes, pues no todas las informaciones son de actualidad; es más, en ocasiones, ni siquiera necesitan apoyos documentales. Sin embargo, estas pequeñas noticias que abundan en casi todas las secciones de los medios terminan por constituir entes informativos y documentales importantes. Así, si un robo de unas características especiales se lleva a cabo con demasiada frecuencia y en determinadas zonas, aquellas informaciones que tenían un escaso valor documental pasan a ser con el tiempo documentos determinantes.

Es responsabilidad del documentalista encargado de la sección, actualizar y considerar válidas documentalmente informaciones que no tienen inicialmente un contenido documental.

En resumen: *A la hora de recuperar la información es importante conocer la realidad informativa presente e intuir qué valor documental tienen ciertas informaciones.*

Algunas tienen una continuidad informativa clara, pues no terminan de manera inmediata. El centro hace el seguimiento documental diario, pero puede esperar a que finalice para tratar documentalmente la información. Por ejemplo, una guerra.

### Búsqueda documental.

Es necesario seguir un proceso para llegar a la información pertinente que el usuario necesita. El documentalista se convierte así en gestor de la información, para atender con garantías de éxito los requerimientos de los usuarios.

### Etapas de la búsqueda documental.

- El primer paso es la petición de información por parte del usuario, si bien algunos centros elaboran dossiers sobre aspectos generales.
- El usuario y el documentalista se entrevistan con el fin de precisar al máximo la petición.
- Entonces, se formula la pregunta en lenguaje documental y se determinan las estrategias y ecuaciones de búsqueda.
- Para ello, hay que traducir el documento a un lenguaje de indización. La expresión que resulta de este proceso se conoce como *modelo de búsqueda del documento*.
- A continuación, se hace la petición de información en el mismo lenguaje o en un lenguaje compatible. Se obtiene así el *perfil de búsqueda*.

El perfil de búsqueda es una ecuación de búsqueda permanente de información formulada por un documentalista con la ayuda de descriptores o palabras claves que se encuentran en el lenguaje del sistema hacia el que se dirige. Se consigue a través de una entrevista usuario-documentalista.

Tres tipos de perfiles en los centros de documentación:

1. Perfil personal o individualizado: el que se adapta exactamente al tema preciso que demanda el usuario.
2. Perfil normalizado o estándar: que recoge aquellos temas de interés seleccionados por el propio centro de documentación que en forma de listas propone a sus usuarios.
3. Perfil de grupo: el que se realiza para un grupo de usuarios que necesitan similares documentos.

- Concretado el perfil de búsqueda y formulada la pregunta se hace una selección de las referencias más pertinentes en función de las especificaciones de la demanda y sus características.
- A continuación se comunica al usuario los resultados de búsqueda.
- La última etapa es la *verificación* por parte del usuario de la validez de la respuesta, y si es necesario, formular de nuevo una petición y volver a seguir los mismos pasos.

En ocasiones, se presenta al usuario un formulario para que conteste al grado de acierto en la búsqueda realizada por el documentalista.

Lo mismo que en la recuperación se puede BUSCAR:

Búsqueda en texto completo: posibilidad de buscar palabras o combinaciones de palabras de un texto completo almacenado en la memoria del ordenador, utilizando a éste como herramienta.

Búsqueda en texto libre: términos extraídos directamente de un texto –lenguaje natural– por un analista humano, que serán luego almacenados y recuperados.

De gran utilidad es la posibilidad de realizar búsquedas con fragmentos de palabras: *truncamientos*.

### Factores que favorecen la búsqueda.

3.2.1 Extensión: Número de puntos de acceso.

3.2.2 Redundancia: Será usualmente mayor en texto libre, aumentando las posibilidades de encontrar un ítem específico. No

obstante, la gran variedad de maneras en que un tema puede representarse en una base de datos dificulta la obtención de aciertos.

3.2.3 Especificidad: texto libre usualmente será más específico, favoreciendo la precisión. La diversidad de maneras en que los conceptos son representados, mientras tanto, dificulta mucho la obtención de un alto acierto en búsquedas conceptuales generales.

3.2.4 Actualidad: Las representaciones en texto libre serán siempre más actuales. Para encontrar un tema nuevo en un sistema de vocabulario controlado, quien hace la búsqueda tal vez tenga que experimentar con diversos términos.

3.2.5 Familiaridad. Los especialistas en formación lo usarán más eficazmente que aquellos que no lo están. El usuario final puede manejarse mejor con el lenguaje natural que encuentra en los documentos de su campo temático.

#### **Problemas de la búsqueda documental: SILENCIO y RUIDO documental**

- A veces, aparecen elementos que pueden producir distorsiones en la demanda, acarreado respuestas no pertinentes o inexactas.
  - Desconocimiento del usuario de sus necesidades reales de información y por consiguiente mala presentación y formulación.
  - Ignorancia del usuario ante las posibilidades de la unidad de información o de las fuentes documentales almacenadas.
  - Falta de entendimiento entre el usuario y el analista.
  - Mala utilización de las demandas por los documentalistas, sea al traducirlas a un lenguaje documental o al corregir y precisar cada uno de los términos de la pregunta.
  - Uso deficiente de los operadores lógicos en el momento de estructurar la búsqueda.
  - Desconocimiento de las posibilidades del programa gestor de base de datos, que podría derivar en una subutilización del mismo y como consecuencia una búsqueda no exhaustiva.

#### **Modelos de recuperación.**

Existen varios modelos para recuperar la información, aunque todo depende del análisis efectuado al documento. El proceso de referencia es fundamental para hacer **búsquedas de información** y sirve para hacer cualquier búsqueda en un centro de documentación.

Los pasos a seguir son los siguientes:

- a. presentar la consulta que se quiere resolver.
- b. determinar la materia o el tema sobre el que se quiere hacer la petición.
- c. identificar el tipo de información que estamos buscando; precisar un poco más.
- d. seleccionar lo que puede responder mejor a nuestra pregunta; de las diferentes posibilidades que tenemos hay que ir eligiendo para ir determinar la más adecuada.
- e. localización de las respuestas en las principales fuentes de información.

En los medios de comunicación, la independencia con la que trabajan unos y otros impide que alguno de los modelos utilizados pueda servir. Ni siquiera en periódicos que forman parte de una cadena se trabaja con el mismo método y se recupera con idénticos planteamientos.

La búsqueda documental puede realizarse sobre la base del texto completo del documento almacenado, sobre los descriptores establecidos por los documentalistas al efecto o mediante la utilización de ambos sistemas simultáneamente.

#### **Sistema de recuperación mediante descriptores.**

- Se basa en el trabajo previo sobre el documento destinado a traducir el contenido de descriptores.
- Posteriormente en la fase de búsqueda, el sistema sólo reconoce como válidos en la interrogación a estos términos.

#### **Sistema de texto completo.**

- En este sistema, el documento se ingresa como elemento de búsqueda y se elimina todo un trabajo previo de determinación de descriptores por los documentalistas.
- El ordenador utiliza el texto completo de los documentos almacenados en la base de datos, para buscar en ellos las palabras que integran la consulta del usuario al sistema.
- Resulta muchos menos costoso que el sistema de descriptores, ya que no requiere un análisis previo –salvo la definición de las palabras vacías, o sea, de las que se han determinado como carentes de significado o trascendencia.

Sin embargo, en el momento de la recuperación este sistema presenta poca precisión en los resultados obtenidos, debido a:

- la ambigüedad del lenguaje natural
- La ausencia de definición semántica de los términos empleados para expresar la búsqueda
- Las referencias vacías empleadas por el autor del documento.

#### **Sistema mixto.**

- Este sistema combina los métodos anteriores, permite la búsqueda basada en los descriptores y mantiene además como elemento de recuperación documental cada palabra o sintagma interesante del texto del documento almacenado.

## **2.7. Lenguajes documentales.**

Los lenguajes documentales aparecen como una necesidad de estructurar el pensamiento, agrupando y asociando cada documento a una lista clasificatoria, o bien mostrando el contenido, sobre todo a través de resúmenes o palabras claves.

En el comienzo, los lenguajes documentales se presentaron como sistemas clasificatorios. Algunos teóricos de la documentación consideran que **el primer sistema clasificatorio** fue el realizado por Brunet en 1804, aunque terminó por imponerse **la Clasificación Decimal de Dewey**.

Melvin Dewey, en 1876, clasificó y dividió el pensamiento en varias clases y a cada una de estas las volvió a dividir y subdividir sucesivamente, dando origen de esta manera a una Clasificación Decimal.

El Instituto Internacional de Bibliografía adoptó este sistema, publicando en 1905 la primera edición internacional. Con posterioridad, la federación Internacional de Documentación realizó varias modificaciones hasta que se constituyó la CLASIFICACIÓN DECIMAL UNIVERSAL.

### Concepto.

El lenguaje documental es un sistema convencional que utiliza una unidad de información para describir el contenido de los documentos, con miras a su almacenamiento y recuperación.

Por regla general, un documento trata de más de una noción, más de un contenido.

*El lenguaje documental actúa como vehículo de comunicación entre el contenido del documento y el usuario y con él se pretende reducir y, a ser posible, evitar la multiplicidad de sentidos que tiene el lenguaje natural.*

### Diferencias entre lenguajes documentales y lenguajes naturales.

Si el *lenguaje natural* se usa para la comunicación inmediata, esta que realizamos ahora mismo; el *lenguaje documental* se emplea para conseguir una comunicación, que es primordialmente un medio, un código unívoco y estereotipado, controlado y no libre, normalizado y no arbitrario.

En el *lenguaje natural* coexisten diferentes significados para un solo significante o diversos significantes sinónimos.

Por el contrario, el *lenguaje documental* ejerce un control léxico que impide la utilización de distintos significantes libres para un mismo significado con el objeto de sobrevivir dentro de un código normalizado.

### Funciones del lenguaje documental.

El lenguaje documental sirve como elemento aglutinador de todos los trabajos que se lleven a cabo en el proceso documental. Así, puede considerarse la expresión lenguaje natural como sinónimo de discurso común, el lenguaje normalmente usado en la escritura y la conversación. Pero, en el contexto de la recuperación de la información, la expresión usualmente se refiere a las palabras que ocurren en los textos impresos y "texto libre" hay que considerarlo como sinónimo.

El lenguaje documental ha de precisar cada uno de los diferentes encuadres que presenta el documento para facilitar la recuperación. Se trata de un lenguaje que recoge los elementos más importantes cuando se está elaborando el análisis y los aplica para hacer efectiva la recuperación.

Sin embargo, existen una serie de problemas:

- Las palabras contienen varios significados
- Las traducciones de otros idiomas
- Los nuevos términos que se introducen
- Las especificaciones técnicas resultan para la mayoría de los usuarios incomprensibles.
- La forma en que se presentan los nuevos documentos, avalados por unas tecnologías que no disponen de aspectos comunes adaptados internacionalmente
- El desacuerdo entre los documentalistas a la hora de aplicar los diferentes lenguajes.

El objetivo básico de un lenguaje documental es suministrar los conceptos que aporta cada palabra, una vez efectuado el análisis. Se trata de reducir los términos que aporta el lenguaje documental, como precisa Blanca Gil: "*El lenguaje documental reduce considerablemente el volumen de términos del lenguaje natural no tomando en consideración más que los sustantivos o los sintagmas nominales*".

### Tipología de lenguajes documentales.

Hay diversas formas de tipificar los lenguajes documentales; sin embargo, las más usadas son:

Lenguajes controlados (Clasificaciones, tesauros.etc): aquellos que han establecido una lista de descriptores antes de proceder al análisis documental. Esta es cerrada y nominativa. Define todos los términos y únicamente aquellos que se pueden utilizar para presentar el contenido de un documento. La búsqueda y recuperación es más rápida y eficaz.

Lenguajes libres (Listas de descriptores libres). Al contrario, cuando se trabaja con un vocabulario o *lenguaje libre* no se conocen a priori listas de términos autorizados. Basta con extraer de los documentos los términos más apropiados. La búsqueda es más lenta y menos eficaz.

De todas formas, hay que señalar que ningún lenguaje es completamente puro, ninguno es pre o poscoordinado, libre o controlado.

Lenguajes de estructura jerárquica o clasificatoria (clasificaciones jerárquicas) aquellos que siguen un orden lógico que agrupa y aproxima los conceptos más sencillos o específicos dentro de los conceptos más generales. Este tipo de lenguaje se puede emplear para localizar un documento, pero no para indizar con profundidad.

Se establece una clasificación sistemática lineal, en la cual los conceptos se encuentran ordenados siguiendo una jerarquía natural, definida por el estado de los conocimientos en el momento en que ha sido elaborada.

Cada concepto de estas estructuras jerárquicas se halla representado por un símbolo numérico, alfabético o alfanumérico que indica la situación de cada materia.

Los lenguajes de estructura jerárquica se dividen en:

#### **Clasificaciones enciclopédicas:**

- Permiten la organización de documentos que tratan sobre cualquier materia: son de ámbito universal y multidisciplinario.
- Presentan dos inconvenientes: el objetivo de su universalidad limita la descripción de un documento especializado y su rigidez dificulta una puesta al día ágil y rápida.

Clasificación de Dewey (1876): divide el conjunto de los conocimientos en 9 clases principales, designadas en números arábigos del 1 al 9, reservando el 0 para las generalidades. Cada clase se subdivide sucesivamente en 10 subclases y así sucesivamente, con números que se dividen en grupos de tres por medio de puntos para hacer más fácil la lectura.

Clasificación Decimal Universal (CDU 1905): es una ampliación de la clasificación decimal de Melvil Dewey (se basa en su quinta edición). Es numérica hasta cierto punto, precoordinada, universal, multidimensional y arborescente. Está agrupada en 10 clases, reducidas a 9 por la fusión de lingüística, filología y literatura, dejando la clase 4 vacía. Cada una de ellas se subdivide. 0 generalidades. 1 Filosofía. 2 Religión. Teología. 3. Ciencias Sociales. Estadística. Política. Economía. Derecho. Administración. Asistencia social. Seguros. Educación. Etnología. 4 sin ocupar. 5 Ciencias puras. Ciencias exactas y naturales. 6 Ciencias aplicadas. Medicina. Técnica. 7 Arte. Artes industriales. Fotografía. Música. Juegos. Deportes. 8 Lingüística. Filología. Literatura. Crítica literaria. 9 Arqueología. Prehistoria. Geografía. Biografía. Genealogía. Historia.

La tarea de actualizar la CDU está encomendada a la FID (Federación Internacional de Documentación), si bien como señala Inocencia Soria: *"En 1992 se constituyó el consorcio CDU, que asumió las responsabilidades que antes tuviera la FID sobre su edición, actualización, versiones, etc. Este consorcio, cuyos socios fundadores son Bélgica, España, Holanda, Japón, Reino Unido y la propia FID, se comprometió a organizar y mantener la CDU y sus esfuerzos ya han dado algunos importantes frutos: se ha creado un fichero informático con más de 60.000 entradas que está sirviendo de base para facilitar su manejo y actualización"*.

Clasificación expansiva de Cutter (1891): se compone de 7 tablas o esquemas, cada una de los cuales incluye la totalidad de los conocimientos, pero con una complejidad progresiva.

Clasificación de la Library of Congress (LCC 1897): cuenta con 21 clases principales, tomadas del sistema de Cutter, que designa con otras tantas letras mayúsculas, dejando las restantes para futuras ampliaciones.

Clasificación bibliográfica de Bliss (1935): está formada por 4 esquemas generales que son filosofía, ciencias zoológicas, físico y social que se dividen en un total de 26 clases principales. Es muy similar a las facetadas.

**Clasificaciones de facetas**: Son de origen enciclopédico, pero su organización permite construir áreas concretas de los conocimientos, ya que faceta es cada uno de los aspectos o puntos de vista que pueden incluirse en un área concreta. Se basan en:

Clasificación colonada de S.R. Ranganathan (1933):

#### **2.8. Bases de datos.**

Coll-Vinent: *una serie de ficheros dispuestos de modo que sirvan para la exhibición y recuperación de documentos.*

#### **Partes de una base de datos.**

Fichero: se trata de una colección de datos almacenados, de modo que sean legibles por el ordenador, lo cual supone una estructuración que haga posible dicha lectura.

Creación de ficheros: El primer paso que hay que dar a la hora de diseñar una base de datos es crear un fichero o un



conjunto de ellos.

Hay que diseñar el fichero principal, es decir, aquel en el que figuran debidamente numerados los registros que han de constituir la base y que en su momento han de ser introducidos en la memoria del ordenador.

No debe ser complicado hacerlo, ya que se trata de una simple yuxtaposición de documentos o registros, a medida que van entrando.

El orden viene dado por el mismo registro, empezando por el número uno y así sucesivamente.

**Registro:** es una unidad completa de información sobre una persona, producto, libro, etc. Es también la información contenida en un fichero, relacionada con un objeto particular, un acontecimiento, una persona concreta, un documento...

Un registro bibliográfico, por ejemplo, es un registro que se refiere a un libro o a alguna otra publicación. Todo registro contiene un mismo género de información particular sobre un documento determinado.

**Campos:** es una subdivisión del registro dedicado a un sólo aspecto de la información en él contenida. Por ejemplo, el nombre de un autor, el título, la fecha, etc.

**Formato:** es la manera en que los registros de un fichero son divididos en campos. Hay multitud de formatos standard, como por ejemplo, el MARC, que se emplea en la catalogación de bibliotecas, ya que una estructura estandarizada es muy útil para el intercambio de registros entre distintas empresas que operan en un mismo campo de actividades.

Los registros de un fichero bien estructurado tendrán uno o más formatos.

- Si se trata de un **formato fijo**, tanto los registros como los campos, dentro de cada uno de ellos tienen una longitud igual, puesto que fueron diseñados para contener un número de caracteres invariables, y en tal caso todos los registros tienen el mismo número de campos y cada campo empieza con la misma posición en cada registro.
- Cuando el **formato es variable**, los registros y los campos dentro de cada registro pueden tener una longitud distinta.

Y, por tanto, no es necesario que cada registro incluya todos los campos posibles. Y a la inversa, los campos pueden repetirse tantas veces como sea necesario.

Para que el ordenador identifique y procese los datos dentro de los campos, cada uno de estos va precedido de un título o etiqueta y cada etiqueta señala el comienzo de un nuevo campo e identifica la clase de campo.

### **Organización y funcionamiento de una base de datos.**

No hay dos bases de datos que contengan un mismo tipo de documentos, pero aún en el caso de que la cobertura informativa fuese la misma, existirían entre ellas diferencias que justifiquen decidirse por una u otra a la hora de acceder a sus fondos.

Antes de enfrentarnos con una base de datos y una vez asesorados por el documentalista, debemos conocer estos aspectos:

1. Cobertura de la base: es decir, el contenido, la extensión de los mismos y si se solapan o no con otras BD.
2. *Actualidad y actualización de sus documentos, cada cuánto tiempo se introducen nuevos datos y desde qué fecha han aparecido en las publicaciones primarias extractadas por la base.*
3. Facilidad de acceso y experiencia personal del usuario en el manejo de esa base.
4. Lenguaje de indización; natural, controlado o ambos; profundidad y grado de especificación en la indización, si van incluidos o no resúmenes en los registros, si se puede disponer del texto íntegro, etc.
5. Documentación complementaria que refuerce los resultados de la búsqueda o que aumente su eficacia, tales como índices, impresos, clasificaciones, tesauros...
6. Los costes del servicio.
7. *La empresa distribuidora a través de la cual la base es accesible y las facilidades que proporciona para la búsqueda.*

Las BD no bibliográficas son producidas por tres distintos colectivos:

- Organizaciones de carácter gubernamental
- Agrupaciones profesionales
- Organizaciones comerciales.

Una base de datos, producida y diseñada por un PRODUCTOR no es informatizada hasta que ha sido introducida en la memoria del ordenador.

El término *host* no indica la traducción del cometido que desempeñan las empresa distribuidoras, sino el tipo de ordenador de grandes dimensiones, en el que han sido montadas muchas bases de datos. Un *host* permite al usuario llegar más fácilmente al contenido de una base de datos, ya que su función es transformar los datos, adquiridos normalmente en soporte papel, en cintas magnéticas, etc.

Un sólo DISTRIBUIDOR puede ofrecer más de una base de datos a sus usuarios, de hecho hay algunos que distribuyen varios cientos.

### **Tipología de las bases de datos.**

Mercedes Caridad hace una división entre:

**Bases de datos referenciales:** nos remiten a otras fuentes (documentos, organizaciones, personas, etc)

Se dividen en:

Bibliográficas: contienen referencias de literatura impresa (libros, artículos de revistas, etc).

Directorios: contienen referencias de informaciones relativas a organismos, individuos, etc.

**Bases de datos fuente:** proporcionan el dato original o el texto completo de la fuente primaria de información.

Se dividen en:

1. Numéricas: datos de encuestas o estadísticas que se presentan generalmente en forma de series temporales.
2. Textual-numéricas: contienen en sus registros campos de información, textos y otros datos numéricos.
3. Textuales: contienen los datos completos de un ítem informativo (noticias de prensa, informes, etc).

## 2.9. BASES DE DATOS PARA LOS PROFESIONALES DE LA COMUNICACIÓN

La recuperación de información aparecida en prensa ha sido y es una necesidad constante para cualquier profesional de la Comunicación.

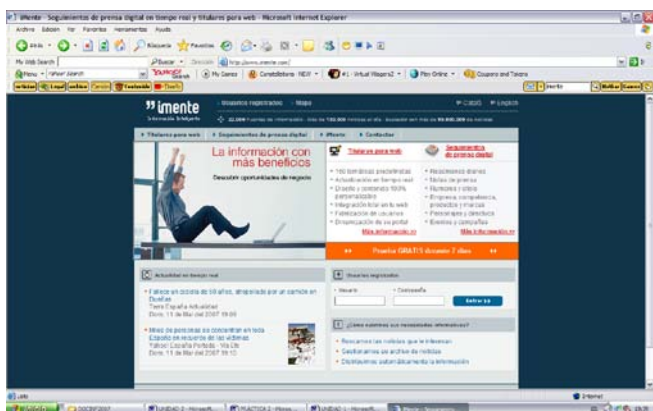
Durante la década de los noventa varios diarios españoles (ABC, El Periódico de Catalunya, EL Mundo, etc.) decidieron comercializar sus bases de datos en soporte CD-ROM. Éstas contenían las referencias bibliográficas de noticias, entrevistas, reportajes, crónicas, etc., pero no permitían el acceso al texto completo.

Durante las décadas de los ochenta y de los noventa tan sólo hubo una base de datos española **BARATZ**, que recogía información procedente de distintos periódicos nacionales. Era referencial y estaba especializada en Economía, Política y Sociedad, por lo que analizaba cualquier tipo de artículo, columna, etc. que apareciera en la prensa española sobre estos temas.

La desaparición de Baratz (en torno a 2002) fue paralela al desarrollo de dos servicios que han venido a cubrir un vacío importante en el acceso a la información periodística: las **HEMEROTECAS DIGITALES** y los **AGREGADORES DE PRENSA**.

En la actualidad las ediciones digitales de los periódicos incorporan **hemerotecas digitales** que permiten a cualquier usuario consultar los fondos documentales de ese medio de comunicación (a veces hace falta ser suscriptor del diario).

Los **agregadores de prensa** son empresas encargadas de desarrollar motores de búsqueda que exploran los distintos archivos de los diarios, revistas u otras fuentes de información para recuperar documentos interesantes para el usuario. Algunos de los agregadores de prensa más conocidos son MYNEWS ([www.mynewsonline.com](http://www.mynewsonline.com)), ICONOCE ([www.iconoce.com](http://www.iconoce.com)) e IMENTE ([www.imente.com](http://www.imente.com))



## BASES DE DATOS DE AGENCIAS DE NOTICIAS

Son un referente obligado para cualquier profesional de la comunicación, en especial para los periodistas. Dos de las más frecuentes en los medios de comunicación españoles son **EFEDATA** y **FACTIVA**.

