

### 3. Almacenamiento y recuperación de información bibliográfica

3.1- Almacenamiento de la información

3.2- Centros bibliotecarios y centros de documentación. Los sistemas de teledocumentación.

3.3- Recuperación de la información. Estrategias de búsqueda.

#### 3.1- Almacenamiento de la información

La documentación se puede encontrar sobre una gran diversidad de soportes. Históricamente la **escritura** ha sido el testimonio si no único sí fundamental sobre el que se ha levantado nuestro conocimiento del pasado. Sin embargo, en el presente junto a los textos manuscritos o impresos en papel encontramos **microfilmes, microfichas, cintas de audio y vídeo, videodiscos, película fotográfica y soportes magnético**. Todos son documentos, todos contienen información, pero la diferente naturaleza del material sobre el que se mantiene exige, por sus cualidades físicas, un *diferente almacenamiento* y también un método de descripción diferente:

<u>SOPORTE</u>	<u>LOCALIZACION</u>
Manuscritos e impresos	Archivos
Libros y folletos	Bibliotecas
Revistas y periódicos	Hemerotecas
Informáticos	Bases y bancos de datos
Audiovisuales	Fonotecas, videotecas, filmotecas

Hasta del desarrollo de los **sistemas automatizados**, los repertorios impresos, bibliográficos o de resúmenes, eran los que actuaban como depósitos de información. La aplicación de la **informática** en las tareas de documentación científica ha dejado progresivamente en desuso esos sistemas.

Para estudiar como se almacena la información y como se puede recuperar, debemos distinguir en primer lugar los *tipos de centros de información y documentación*:

- \* Las **bibliotecas** son sistemas de información documental cuyo objeto es permitir a sus usuarios acceder directamente o por sí mismos a documentos primarios (libros, revistas, obras de consulta, etc.)
- \* Los **centros de documentación** son sistemas de información documental que tienen como objetivo principal poner a disposición de los usuarios documentos secundarios (ficheros, boletines bibliográficos, índices, etc.) que les permitan descubrir los documentos primarios que les interesen.
- \* Los **sistemas de teledocumentación**, las bases y bancos de datos, constituyen un sistema de información documental que se caracteriza por utilizar *medios informáticos* para el almacenamiento y el suministro a los clientes de una amplia gama de servicios informativos.

Todos los centros de documentación e información tienen en común el ser **organismos intermedios** en la labor de documentación, en cuanto a que su finalidad fundamental es acumular documentos informativos que después se ponen a disposición del público.

### **3.2- Centros bibliotecarios y centros de documentación. Los sistemas de teledocumentación.**

#### **\* Centros bibliotecarios**

Según Nuria Amat (1982), **biblioteca** es “toda colección organizada de libros, publicaciones periódicas u otros documentos cualesquiera (en especial gráficos y audiovisuales), así como los servicios del personal, que facilita a los usuarios la utilización de estos documentos con fines informativos, de investigación, de educación o recreativos”.

Las bibliotecas constituyen, pues, sistemas de documentación en los que las **entradas** son los *documentos* que constituyen el **fondo bibliográfico de la biblioteca**, y las **salidas** los *servicios* que las bibliotecas prestan al público, para facilitar la información bibliográfica y el acceso a los documentos primarios.

A pesar de la aparición moderna de los centros de documentación y de los bancos de datos, las bibliotecas siguen siendo sistemas de documentación de **obligada utilización** en la labor de localización y consulta de diversos tipos de fuentes utilizados en la elaboración de trabajos científicos, y de modo especial en los estudios retrospectivos y de tipo histórico referidos a épocas no abarcadas en los centros de documentación.

Destacar en este punto que son cada vez más las bibliotecas que están conectadas a redes informáticas, lo que permite al investigador consultar los **catálogos** sin la presencia física en dicha biblioteca. Esto da acceso a un número enorme de bibliotecas de todo el mundo (ej. HYTELNET).

#### **\* Los centros de documentación**

La multiplicidad de fuentes de información que existe y se produce en las sociedades hizo necesaria, ya desde muy antiguo, la existencia de centros especializados para su recogida, almacenamiento y puesta a disposición de los usuarios. De ellos son buenos ejemplos las bibliotecas.

La revolución actual en la información y las comunicaciones informativas, con la multiplicidad de los medios utilizados y la amplitud enorme de material informativo que continuamente se está produciendo en el mundo moderno, ha hecho que las *bibliotecas ya no sean suficientes*. Ello dio lugar al **nacimiento**, alrededor de los años 50, de los **centros de documentación y sus técnicas**, orientados a recoger, ordenar, clasificar y proporcionar a los usuarios toda esa masa enorme de información que en las sociedades actuales se encuentra desperdigada en un incontable número de documentos de todo tipo.

Estos centros pueden tener:

- \* carácter nacional o internacional
- \* carácter general o especializado, limitado por tanto a un determinado campo de la ciencia o de la técnica
- \* carácter no autónomo, dependiente de empresas, centros de investigación, asociaciones profesionales, lo que no obsta para que puedan suministrar también informaciones particulares.

De todas formas, los centros de documentación no suelen ser generales (como frecuentemente son las bibliotecas), sino *especializados*. La mayoría cumplen todos o algunos de los siguientes **cometidos** respecto a los documentos: su identificación, registro, organización, almacenamiento, recuperación, conversión en formas útiles, su síntesis y su diseminación entre los usuarios, preferentemente en forma de resúmenes de los documentos y listas de referencias bibliográficas y documentales.

En cuanto a su **definición**, según la UNESCO, los centros de documentación “*están constituidos por grupos de personas que se dedican de lleno al procesamiento de documentación relativa a un cierto campo o tema, a fin de proveer regularmente información abreviada a una comunidad de usuarios especializados*”.

Las **formas de difusión** de la información utilizadas por los centros de documentación son múltiples. Entre ellas, las más habituales son:

- \* Notas bibliográficas
- \* Informaciones sobre obras recibidas
- \* Resúmenes analíticos
- \* Informes y resúmenes temáticos
- \* Conferencias o extractos de conferencias, conclusiones de congresos, etc...
- \* Filmes, emisiones de radio y televisión
- \* Publicaciones, manuales técnicos, traducciones
- \* Artículos

Desde el punto de vista de la documentación científica en el *trabajo de investigación*, los **servicios** que se pueden obtener de los centros de documentación son:

- \* Proporcionar recopilaciones de documentos ya publicados acerca de un determinado tema. Estas recopilaciones suelen denominarse **búsquedas retrospectivas**, dado que recogen información documental fechada en tiempos pasados y no se ocupan de lo que pueda aparecer con posterioridad.
- \* Mantener una información continuada al día sobre los documentos que van apareciendo acerca de un determinado tema. Los “**sistemas de alerta**” mediante los que se realiza esta **distribución selectiva de la información** se ajusta a un **perfil** consistente en una serie de palabras claves o descriptores que, separadamente o por combinaciones entre sí, describen con la mayor precisión posible el tema de que se trata. Los perfiles pueden ser elaborados por los clientes o bien el mismo centro de documentación puede ofrecer una serie de ellos para que elijan aquellos.
- \* Suministrar **reproducciones y traducciones**, en su caso, de los documentos que interesen al cliente.

\* Los centros de documentación importantes que poseen bases de datos suelen facilitar las **cintas magnéticas** cuyo contenido encierra la información almacenada por el centro. Permiten el establecimiento de terminales con el ordenador de su base de datos. Además, suelen publicar periódicamente boletines, revistas de resúmenes e índices.

*\*Centros españoles y extranjeros, características y servicios\*\*\*Transparencia\**

### **\* Los sistemas de teledocumentación. Bases de datos y bancos de datos.**

Los sistemas de **teledocumentación** aplican los *ordenadores* y las *redes de telecomunicación* para realizar las operaciones las operaciones y prestar los servicios propios de los sistemas de documentación.

La extensión creciente de la aplicación de la teledocumentación encuentra su fundamento y explicación en la concurrencia de dos factores, uno de carácter documental y otro técnico:

\* La primera es la incapacidad de los sistemas tradicionales para abarcar la explosión informativa típica de nuestro tiempo y satisfacer las crecientes necesidades de información en todos los campos de la vida.

\* La circunstancia técnica es la extraordinaria capacidad que muestran los modernos procedimientos informáticos para almacenar y difundir la información documental.

Los **sistemas teledocumentales**, constituyen grandes redes de comunicación que facilitan el acceso de terminales situados en centros de información y documentales, tales como universidades bibliotecas, etc., a las bases de datos documentales almacenadas en grandes ordenadores. Un rasgo típico a destacar de un sistema de este tipo es que, mientras los centros de documentación y las bibliotecas constituyen sistemas aislados o independientes, la telecomunicación proporciona la oportunidad de conexión entre diversos centros, lo que da lugar a que se formen redes documentales de ámbito internacional.

#### Elementos de los sistemas de teledocumentación

Los elementos que forman un sistema de teledocumentación se pueden reunir en diferentes grupos:

##### \* Elementos materiales

- *Memorias.*- Están formadas por unidades de almacenamiento de datos, sobre discos o memorias de acceso rápido.

- *Ordenador central.*- Permite la búsqueda de las informaciones en las unidades de almacenamiento, mediante la utilización de programas de interrogación.

Las memorias y el ordenador constituyen el **servidor o "host"**

- *Módem.*- Modulador-demodulador. Transforma órdenes dadas al ordenador por el teclado de la terminal y las respuestas del ordenador en impulsos eléctricos transportables sobre las redes.

- *Red telefónica.*- Trabaja en tiempo compartido con el ordenador, prestando servicios a múltiples clientes.

- *Teclado.*- Permite hacer las preguntas al ordenador y dialogar con este, mediante la utilización del lenguaje.

- *Impresora.*- Imprime mecanográficamente las ordenes dadas al ordenador y las respuestas de este.

- *Pantalla.*- Permite visualizar sobre ella lo que se imprime en la impresora.

El teclado, la impresora y la pantalla forman en conjunto la **terminal**.

Este sistema clásico, aunque sigue funcionando, ha evolucionado a un sistema más personalizado. En la actualidad los *usuarios* pueden conectarse desde su propio ordenador personal al ordenador central a través de proveedores de red (INTERNET) o como veremos más adelante ciertas instituciones (como universidades o bibliotecas) permiten el acceso a bases de datos que reciben en formato CD-ROM. Dicho de otra manera, la teledocumentación no solo se limita a los grandes centros de información o instituciones.

#### \* Elementos organizativos

A la complejidad técnica de los sistemas de teledocumentación corresponde una **complejidad administrativa**, con relación a los demás sistemas de documentación. Mientras que en un sistema clásico, como una biblioteca, una sola organización engloba sus diferentes servicios y departamentos, en los sistemas de teledocumentación los diversos *servicios* que implica suelen estar a *cargo de organizaciones o centros independientes*.

Como elementos documentales organizativos de un sistema de teledocumentación se pueden distinguir los siguientes:

#### **- Centros productores de bases y bancos de datos.**

Los sistemas de teledocumentación, como los demás sistemas de documentación se fundan en la recogida, tratamiento y almacenamiento de documentos. Cuando en estos sistemas se habla de *centros productores de bases de datos*, se hace referencia a centros, que pueden ser bibliotecas o centros de documentación, que realizan esta tarea de recoger, tratar y almacenar documentos bibliográficos (**bases de datos**) o datos numéricos (**bancos de datos**), pero con la particularidad de que disponen de ordenadores o soportes electrónicos para almacenar los documentos y datos.

Se pueden resumir sus **funciones** en:

- \* Adquisición de la información, que incluye la localización de datos relevantes y de información auxiliar y la organización de esta adquisición, reproducción, duplicación y difusión subsiguiente.
- \* Proceso de la información preparándola para su distribución y análisis (creación de bases y bancos de datos).
- \* Mantenimiento de la información, asegurando la continuidad de los datos y la documentación en orden a prevenir irreparables pérdidas de información.

#### **- Bases y bancos de datos.**

Están constituidos por el conjunto de ficheros electrónicos en los que los centros productores de bases y bancos de datos almacenan la información documental que han recogido y procesado. Todo documento procesado por uno de estos centros, da lugar a un **registro** ("record"), compuesto por diversos elementos o **campos** (autor, título, editorial, etc.).

Un conjunto de registros referentes a la misma disciplina o asunto, da lugar a un **fichero** ("file"). Finalmente, el conjunto de ficheros sobre distintos temas y

disciplinas de que dispone el centro productor, recibe el nombre de **base de datos** o **banco de datos**.

Los términos de *base de datos* y *banco de datos* se emplean a veces indistintamente, pero se tiende a utilizar la expresión de base de datos para designar un conjunto de ficheros de un centro de carácter bibliográfico o textual, y la de banco de datos para nombrar un conjunto de ficheros cuando contienen datos de carácter numérico.

#### - **Centros proveedores o distribuidores de bases y bancos de datos.**

Los productores de información tienen la posibilidad de poner a disposición de una clientela restringida o del gran público un determinado número de informaciones, pero no todos tienen los medios informáticos para apoyar esta distribución. Es por ello que recurren a *sociedades especializadas en el suministro de servicios informatizados*. Estas sociedades son los **proveedores** de bases y bancos de datos. Son empresas públicas o privadas que, con el apoyo logístico de los ordenadores, ofrecen comercialmente la posibilidad de un acceso a un cierto número de datos (propios o creados por otros). Estos centros reciben *cintas magnéticas* de los creadores de bases de datos y las reproducen en soporte informático (*discos magnéticos*) que permite el acceso directo a la información precisa.

#### - **Empresas de redes de telecomunicación.**

Son las empresas privadas o públicas que gestionan las líneas telefónicas nacionales e internacionales las que unen entre sí los demás elementos del sistema de teledocumentación. Estas redes son redes dedicadas a la transmisión de datos, pero en la actualidad todo queda interconectado por **INTERNET**, una **red de redes** que comunica prácticamente todo.

#### - **Centros de acceso a bases y bancos de datos.**

Son entidades públicas o privadas que disponen de terminales, las cuales permiten pedir información a las bases y bancos de datos. En este caso volvemos a resaltar la importancia de INTERNET en este tipo de conexiones por su universalidad.

#### - **Redes de bibliotecas y centros de documentación.**

La proliferación de bibliotecas y centros de documentación desconectados entre sí ha dado lugar a la creación de redes para coordinar su labor y evitar la dispersión y duplicidad de tareas.

#### - **Las bases de datos en CD-ROM**

Otro gran avance en el campo de la telecomunicación es el que representa el CD-ROM siglas de "compact disk, read only memory" (disco compacto, con memoria de sólo lectura).

Se trata de discos susceptibles de lectura óptica por rayos láser digitalizada, que tiene una enorme capacidad de almacenamiento de información, recuperable simplemente mediante un ordenador personal.

Los CD-ROM suponen la posibilidad de disponer de una base de datos en el propio domicilio del usuario. La **recuperación personal** de la información contenida en el CD-ROM solo precisa de un ordenador y un lector de CD-ROM.

No todo son ventajas, pues no puede olvidarse que el CD-ROM solo permite acceder a una información cuya actualización depende de nuevas publicaciones de discos, esto no es tan rápido como el acceso “en línea”.

Destacar en este punto el comienzo del DVD, de tecnología digital similar al CD-ROM pero con un incremento enorme de capacidad (de unos 650 MB de un CD, a los 2,6 a 5,2 GB de un DVD)

### 3.3- Recuperación de la información. Estrategias de búsqueda.

El proceso de recuperación de información consiste esencialmente en extraer de una colección de documentos aquellos que se ajustan a las especificaciones de una petición determinada. Se trata pues de una comparación sistemática entre los documentos o sus representaciones y la petición o demanda de información. Podemos descomponer el proceso en tres fases:

- \* Traducción del documento en un lenguaje de indexación. La expresión que resulta de este proceso se llama **modelo de búsqueda** del documento. Se trata de representar el documento por una serie de descriptores que lo determinen lo más directamente posible para su posterior búsqueda.
- \* Expresión de la petición de información en el mismo lenguaje del modelo de búsqueda. Se obtiene el denominado **perfil de búsqueda**.
- \* Comparación sistemática de los modelos de búsqueda de los documentos con el perfil de búsqueda, a fin de seleccionar los que se ajusten a este.

La operación resultaría relativamente sencilla si los lenguajes de indexación y de búsqueda coincidieran exactamente. Ello, sin embargo, no es así en la mayoría de los casos, ya que los productores de bases de datos y repertorios bibliográficos suelen indexar en lenguaje libre o semilibre, lo que resulta mucho más fácil y económico para sus fines. Por tanto, al seleccionar **los descriptores para la búsqueda**, habrá que tener en cuenta *todas las posibles formas de expresión de un concepto en la indexación de los documentos* (sinónimos, conceptos más generales y más específicos, etc.) ya que de otro modo podrían perderse cantidades importantes de información. Por otra parte, si la selección es demasiado amplia, se obtendrán documentos carentes de interés. De ahí la importancia de preparar adecuadamente el perfil de búsqueda, operación que resulta así la más importante en el proceso de recuperación de información.



## Preparación de perfiles de búsqueda

El perfil de búsqueda parte de una petición de información en lenguaje natural, y consta esencialmente de tres elementos:

- \* Identificación de los conceptos **(1)**
- \* Desarrollo y expansión de los conceptos, mediante una colección de términos **(2)**
- \* Expresión de las relaciones entre los términos, mediante operadores lógicos (AND, OR, NOT) **(3)**

Vamos a referirnos a la elaboración de un perfil para una búsqueda temática (también es posible realizar búsquedas por autores, revistas, idiomas, etc.). Tomemos un ejemplo sencillo:

\*\* Petición de información sobre "pinturas anticorrosivas de alto contenido en cinc" \*\*

**(1)** En esta petición identificamos como conceptos:

- pintura
- corrosión y su prevención
- cinc

**(2)** La segunda fase, o expansión de conceptos, es imprescindible debido a que, para efectuar la búsqueda, el ordenador compara los distintos términos del perfil con los que contienen los modelos de búsqueda de los documentos carácter a carácter y sólo da como aciertos los que coinciden exactamente. Así, un documento sobre "recubrimientos anticorrosivos a base de cinc" no se registraría como acierto. Es, pues, necesario desarrollar cada concepto teniendo en cuenta los sinónimos, palabras más genéricas, más específicas, etc., para lo cual será apreciable la ayuda de un *thesaurus* (conjunto de términos de una bases de datos). En nuestro ejemplo, y sin ser muy exhaustivos, el desarrollo de los conceptos conduciría al siguiente cuadro:

A	B	C
Pintura	Corrosión	Cinc
Recubrimiento	Anticorrosivos	Zinc
Barniz	Degradación	Zn
Laca	Desgaste	
Esmalte	Incrustación	

Como se puede ver, no sólo se tienen en cuenta palabras relacionadas sino también las diversas formas ortográficas. En este sentido, ocurre con frecuencia que, para agotar todas las posibilidades, es preciso considerar palabras derivadas de la que se ha utilizado como término. Por ejemplo, si se ha elegido como término "polimerización" será frecuente que interesen también términos como "polímero", "polimerizado", "copolímero", etc. Para tenerlos en cuenta se hace uso del **truncado**, artificio mediante el cual el ordenador considerará aciertos a todos los términos que contengan determinados fragmentos de palabras, sean cualesquiera las letras que se hallen antes o después de los mismos.

Existen tres tipos de *truncado*:

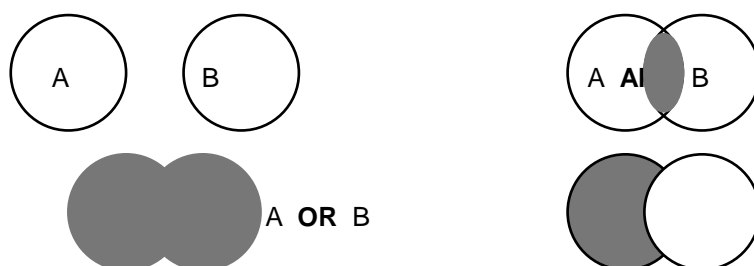
- De  **sufijo**: polimer\*, cubriría polimer-o, polimer-os, polimer-izado, polimer-ización,...
- De  **prefijo**: \*polímero, cubriría polímero, co-polímero, homo-polímero,...



- De **infijo**: \*polimer\*, cubriría todas las posibilidades apuntadas.

De todas formas no se debe abusar del truncado porque puede conducir a la recuperación de muchos documentos irrelevantes.

(3) La tercera fase, relaciona los términos y conceptos mediante los operadores lógicos.



\* El operador **OR** permite obtener un nuevo conjunto formado por los documentos que contienen indistintamente el término A o el B (*operación de unión*)

\* El *operador de intersección*, **AND**, permite obtener un conjunto formado por los documentos que contienen simultáneamente los términos A y B.

\* El operador de *exclusión*, **NOT**, permite formar conjuntos de documentos que contienen el término A, pero no el B.

En general, se relacionarán con la lógica OR los términos correspondientes a un mismo concepto; con la lógica AND los conceptos que deben estar presentes simultáneamente y con la lógica NOT, aquellos que se desee excluir. En nuestro ejemplo, el perfil resultante sería:

**(pintura OR recubrimiento OR barniz OR laca OR esmalte)**

**AND**

**(corrosión OR anticorrosivos OR degradación OR desgaste OR incrustación)**

**AND**

**(cinc OR zinc OR Zn)**

Aquí no hemos utilizado el operador NOT. Si se desea, por ejemplo, recuperar información sobre “pinturas anticorrosivas, excepto las que contengan cinc”, bastaría sustituir, en el perfil anterior, el último operador AND por el operador NOT.

Una vez preparado el perfil de búsqueda, se efectuará la comparación con los modelos de búsqueda de los documentos, para obtener los que se ajustan al perfil. Finalmente habrá que comprobar si los documentos obtenidos satisfacen los requisitos del petionario, es decir, la **relevancia** de dichos documentos para la búsqueda solicitada. Muy a menudo, la información que en verdad se necesita no coincide exactamente con lo que se pide. Ello se debe, normalmente, a que el usuario no conoce con precisión sus necesidades, o no es capaz de expresarlas de forma adecuada.