

## ***INTRODUCCIÓN***

El almacenamiento de los documentos en papel se vuelve cada vez más costoso, menos seguro, y la explotación de los mismos más compleja y menos efectiva. Para resolver estos problemas, las tendencias actuales apuntan al desarrollo de *Sistemas de Información Documental*, los cuales consisten en la computarización de la información documental.

Los *Sistemas de Información Documental* ofrecen funcionalidades tales como: almacenamiento de grandes volúmenes de documentos, búsqueda de documentos según su contenido, comunicación de documentos a otros usuarios, visualización y modificación de documentos por parte de usuarios finales. Para ello, según las características de la aplicación, se propone almacenar los documentos en forma de imagen o de texto modificable a través de un procesador de texto.

Las fuentes de datos pueden ser documentos en papel o provenientes de redes globales (ej. Internet)., Los documentos en papel pueden ser digitalizados y almacenados como imágenes. En el caso de textos y formularios puede extraerse la información a través de la aplicación de mecanismos de *Reconocimiento Óptico (OCR-ICR)*. El Sistema de Información Documental resuelve la gestión de estos datos que serán almacenados en diferentes tipos de bases de datos.

Las Bases de Datos de Texto almacenan la información documental en diferentes formatos según las características de los documentos y las necesidades de recuperación. Permiten almacenar textos complejos (ej. leyes) y explotar esta información a través de consultas *Full-Text* en forma eficiente. Las consultas *Full-Text* permiten al usuario recuperar los documentos que contengan palabras o frases dadas en cualquier lugar del documento.

En un *Sistema de Información Documental* también pueden utilizarse bases de datos Relacionales y de imágenes para almacenar descriptores de documentos, relaciones entre descriptores, imágenes, etc. En cuanto a los mecanismos de almacenamiento, se aplican medios magnéticos y/o ópticos.

Los Sistemas de Información Documental no deben quedar aislados dentro de la organización, sino que deben conectarse con otros sistemas para permitir una mejor explotación de la información en forma combinada.

Por ejemplo, un usuario final que debe tomar decisiones debería poder acceder en forma combinada a los Sistemas de Información Documentales, Sistemas de Gestión y Sistemas Geográficos.

### 2.1 DEFINICIÓN:

Una base de datos es un conjunto de información estructurada en registros y almacenada en un soporte electrónico legible desde un ordenador.

En una **base de datos documental** cada registro se corresponde con un documento (publicación impresa, documento audiovisual, gráfico o sonoro, documento de archivo, documento electrónico, etc.).

Una base de datos documental es sinónimo de sistema de almacenamiento y recuperación de información (SARI).

### 2.2 ESTRUCTURA

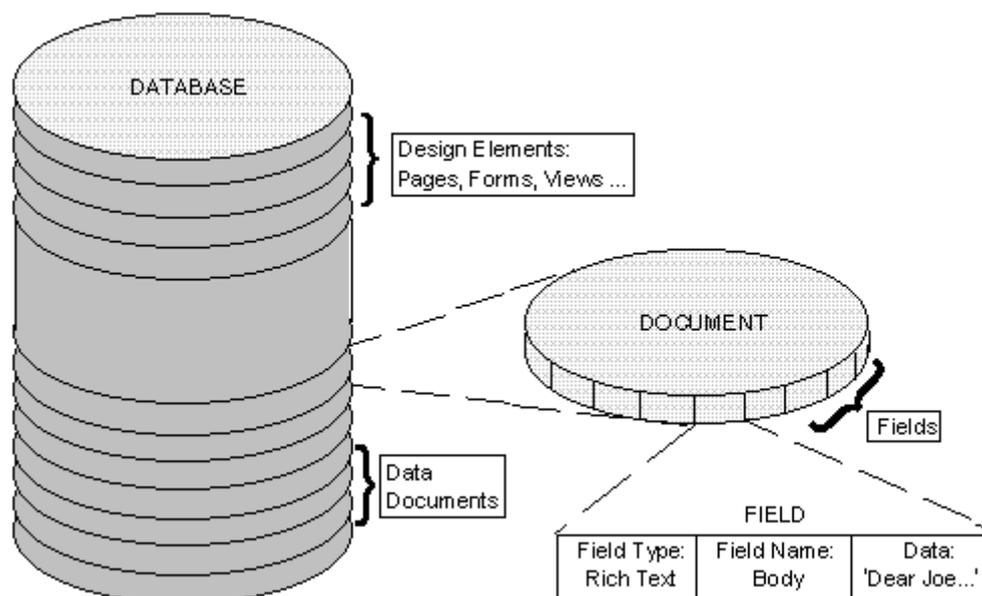


Figura 2.1 Estructura de una base de datos documental

### 2.2.1 CARACTERÍSTICAS DE UN SARI

Las tres características que permiten identificar a un SARI<sup>1</sup> son:

1. Utilización de la técnica del fichero inverso.
2. Aplicación de la lógica booleana como método de recuperación de información.
3. Manipulación de registros de extensión variable o indefinida.

#### 1. La utilización del fichero inverso.

El Fichero invertido es un mecanismo para indexar una colección de documentos obteniendo alta rapidez en la recuperación.

El fichero inverso contiene, ordenadas alfabéticamente, todas y cada una de las palabras o términos con significado semántico, contenidos en los documentos presentes en la base de datos documental. Valdría la imagen de una larga lista de palabras ordenadas, acompañada por una indicación del documento en el que aparece cada una de ellas.

En el momento de realizar una búsqueda o consulta, el SARI no procedería a leer todos y cada uno de los documentos: simplemente, buscaría en el fichero inverso correspondiente todas y cada una de las ocurrencias de los términos buscados, ofreciendo como resultado el número de documentos en el cual aparecen. Dependiendo del sistema utilizado, acto seguido pueden mostrar, o no, los documentos pertinentes. En la mayor parte de las situaciones, el sistema queda a la espera de nuevas órdenes del usuario, sean de ejecutar nuevas búsquedas, sean de visualización de los documentos.

---

<sup>1</sup> SARI: Sistema de Almacenamiento y Recuperación de Información

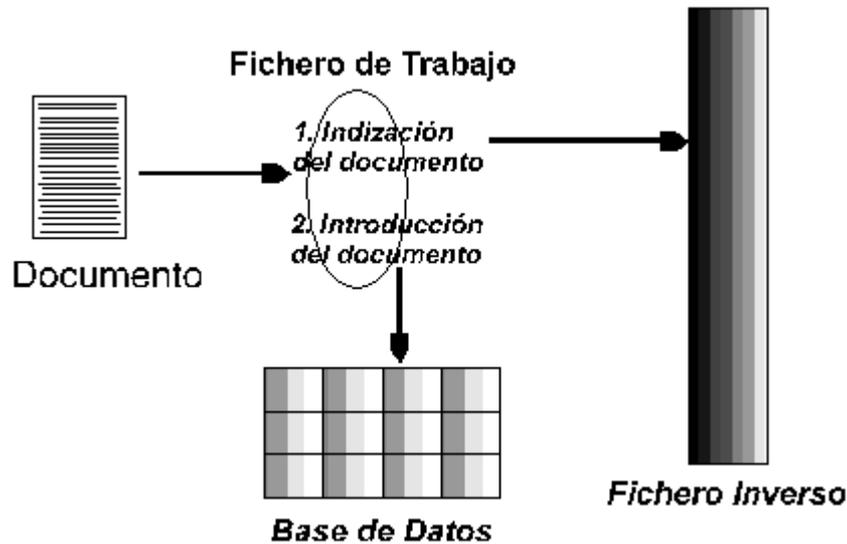


Figura 2.2 Utilización del fichero inverso

### *Tratamiento del documento en un SARI*

En la técnica del fichero inverso se ha introducido progresivamente mejoras y especializaciones. Así, los ficheros inversos no suelen limitar su contenido al término y el documento o documentos en los que aparece. Se han incorporado también los datos referidos al campo y párrafo, dentro del documento en el que aparece, así como la línea y la posición absoluta dentro de la línea, en el modo: "documento 3, campo título, línea 2, palabra 4".

### *2. La lógica booleana.*

El álgebra booleana es el resultado de la obra de Boole, que estableció las bases de las operaciones a través de operadores lógicos. La combinación de los términos mediante operadores lógicos crea una ecuación, la cual puede ser transformada por medios informáticos. El SARI selecciona y compara, mediante la aplicación de la teoría de conjuntos, los resultados de la consulta al fichero inverso, y selecciona, de acuerdo a los operadores establecidos por el usuario, aquellos que responden a la lógica de la ecuación de búsqueda.

### *3. Manipulación de registros de estructura y extensión variable.*

Por último, la capacidad para representar y manipular registros de estructura y longitud variable es otra de las características propias de los SARI. Como es previsible, no todos los

documentos poseen la misma estructura, ni ofrecen la misma longitud. En los SGBD<sup>2</sup> tradicionales se ha considerado necesario que la estructura y contenido de todos los registros fuese similar. Sin embargo, las características propias de los documentos obligan a que el sistema de bases de datos documental sea capaz de aceptar, almacenar y manipular registros de forma flexible.

### **2.2.2 TIPOS DE SISTEMAS DE TRATAMIENTO Y RECUPERACIÓN DE INFORMACIÓN DOCUMENTAL**

La recuperación de documentos a través de ecuaciones de búsqueda es posible gracias a que existen unas aplicaciones que no sólo las ejecutan, sino que son capaces de gestionar, en el sentido más amplio, las bases de datos documentales. El proceso de especialización que dio lugar a la informática documental ha producido, a su vez, aplicaciones especializadas en diferentes aspectos, que en muchas ocasiones entrecruzan sus características:

1. *Sistemas de gestión de bases de datos documentales*: son sistemas que incorporan todas las características de los SGBD tradicionales, incluyendo la creación y mantenimiento de bases de datos documentales, usuarios, controles de seguridad, e incluso lenguajes propios de programación. Debe citarse BRS/Search, de BRS Information Technologies (uno de los más completos), Inmagic, CDS-Isis y su interfaz WinIsis, Texto.

2. *Sistemas de indización*: son aquellos que crean ficheros de índice, los cuales contienen los términos existentes en diferentes tipos de ficheros, como por ejemplo ASCII o cualquier procesador de textos, y permiten una limitada recuperación a través de aquellos, que suele utilizar operadores booleanos. Uno de los más conocidos es ZyIndex.

3. *Sistemas de exploración o escáneres*: se trata de aplicaciones que, sin necesidad de crear ficheros inversos ni diccionarios, son capaces de acceder a ficheros con diferentes formatos, y buscar dentro de los mismos las cadenas de caracteres que respondan a lo expresado en la ecuación de búsqueda. Pueden encontrarse aplicaciones que combinen la exploración con la indexación, como dtSearch.

4. *Sistemas de gestión bibliográfica*: sistema especializado para la gestión y mantenimiento de bibliografías especializadas, es una aplicación específica de los sistemas de gestión de

---

<sup>2</sup> SGBD: Sistema de Gestión de Bases de Datos

bases documentales, que permiten no sólo el almacenamiento y la recuperación de referencias bibliográficas, sino también la exportación de estas referencias en diferentes formatos de cita bibliográfica a diferentes procesadores de textos, sistemas de gestión de bases de datos, etc. Como ejemplo se pueden citar Reference Manager, ProCite, EndNote...

5. *Sistemas de recuperación de información*: son aplicaciones que se encargan exclusivamente de recuperar información de bases de datos documentales no modificables. Ponen a disposición del usuario potentes herramientas de búsqueda y de apoyo a la búsqueda, pero su funcionalidad queda reducida a la consulta y exportación de documentos. Las bases de datos que vienen en soporte CD-ROM utilizan este tipo de sistemas, de los que debe citarse SPIRS (Silver Platter Information Retrieval System) y WinSpirs (versión para Windows de Spirs), así como los proveedores de bases de datos en línea, como DIALOG.

6. *Sistemas hipertextuales*: en su origen, los hipertextos e hipermedias son una forma de organizar, acceder y explorar documentos de diferentes tipos, que posteriormente se ha popularizado como motor y parte de tutoriales y presentaciones. Actualmente estos sistemas están volviendo a ser considerados como una forma válida y muy avanzada de gestionar documentación.

7. *Sistemas de Gestión Documental o de Gestión Electrónica de Documentos (GED)*: se trata de sistemas que pretenden ofrecer una solución integral para la documentación, especialmente administrativa y de gestión, que se utiliza en una organización dada. Incorporan funciones clásicas de gestión de bases de datos, y utilizan esquemas de obtención de una copia del documento original mediante escáner, almacenamiento óptico o magneto-óptico, y un nivel básico de descripción textual del documento y de su contenido.

8. *Sistemas ó Gestores de información personal (Personal Information Systems / Managers)*: son aquellos que integran, en un único entorno, todos los documentos, ficheros y relaciones entre ellos que son de interés para el trabajo de un usuario. Numerosos sistemas integrados de informatización ofrecen a sus usuarios un acceso homogéneo a los diferentes tipos de documentos y ficheros que manejan en su trabajo diario.

9. *Sistemas compuestos*: se denomina así a aquellos que dan soporte a todas las tareas que se realizan en una unidad informativa, sea ésta un archivo, biblioteca o centro de documentación. Esto significa que cubren tanto la cadena documental como la gestión

administrativa. Sirvan como ejemplo las aplicaciones de automatización de bibliotecas, como Absys o Libertas, o las aplicaciones de automatización de archivos, como la desarrollada para el Archivo de Indias de Sevilla. Normalmente, integran un motor documental, encargado de gestionar las bases de datos documentales que cubren los catálogos, y un motor relacional, que cubre las tareas administrativas.

### 2.2.3 FUNCIONES TÉCNICAS DE UN SARI

Los Sistemas de Almacenamiento y Recuperación de Información Documentales deben ofrecer a sus usuarios, como especializaciones de los SGBD que son, las mismas funciones técnicas de creación, mantenimiento y recuperación de la información que éstos. Esto supone tanto la disponibilidad de funciones para el usuario final, de forma que tenga a su alcance todos los mecanismos necesarios para acceder a los documentos, como la existencia de herramientas y utilidades que aseguren la adecuada administración y mantenimiento del sistema, en sus diferentes niveles, por parte del DBA<sup>3</sup>. Ambos deben ser complementados por nuevas capacidades, más adecuadas a un tratamiento documental e informativo de la base de datos. En esta situación, un SARI debe ser capaz de ofrecer las siguientes funciones:

1.- *Creación de bases de datos documentales*, con una estructura de campos (también llamados párrafos o párrafos) fija. Los diferentes campos, para lograr un correcto tratamiento de los datos, deben ser de longitud variable y, a ser posible, permitir la creación de subcampos o subpárrafos.

2.- *Introducción de datos*, tanto de manera interactiva, como de ficheros preexistentes, con posibilidad de conversión de formatos de bases de datos externas. Las funciones de introducción de datos deben permitir controles de entrada, validación, etc, de cara a asegurar, en lo posible, la calidad de los datos introducidos.

3.- *Recuperación de la información* contenida en los documentos de la base de datos a través de cualquier término existente en ella, mediante la formulación de ecuaciones de búsqueda que permitan combinar los términos según diferentes criterios. Los sistemas ofrecen la posibilidad de ejecutar las consultas sobre una o varias bases de datos

---

<sup>3</sup> DBA: Administrador de Base de Datos

simultáneamente. Los documentos resultantes se agrupan en sets o conjuntos, susceptibles de combinación posterior.

4.- *Salida de la información*, mediante edición en pantalla, impresión y redirección a ficheros de los documentos de interés para el usuario. Las órdenes de salida de información deben ofrecer la posibilidad de enviar ésta a diferentes destinos, así como los formatos de presentación de los datos a utilizar (tamaño, campos...). Deben incluirse aquí las capacidades para ordenar, según diferentes criterios, los documentos resultantes. Otra función a considerar es la posibilidad de crear nuevas bases de datos, tomando como base los documentos recuperados en una búsqueda previa.

5.- *Análisis de la respuesta*, utilizando herramientas de análisis de frecuencias de los términos o de concurrencias.

6.- *Creación de ficheros* que definan el perfil de búsqueda de los usuarios, así como de las ecuaciones que ejecuten. A los ficheros que contienen ecuaciones de búsqueda previamente grabadas, y que pueden ejecutarse en cualquier momento, se les llama normalmente "macros". Estas macros son ficheros susceptibles de edición y modificación, lo que facilita la recuperación de información con un mínimo esfuerzo de tiempo y coste.

7.- *Gestión de mecanismos de control terminológico*, tanto para la entrada de datos como para su recuperación. Puede tratarse de un tesoro, de un glosario o de un diccionario terminológico.

8.- *Ayuda al usuario en todo momento*, a través de mensajes y líneas de estado, especialmente durante el proceso de interrogación (interrogación asistida). En sistemas de recuperación en línea (teledocumentación), el sistema informa al usuario del tiempo de conexión, tareas ejecutadas, coste de la sesión, etc. Los mecanismos de ayuda al usuario, especialmente aquellos referidos a la evaluación y refinamiento de las búsquedas, son una de las principales áreas de investigación.

9.- *Dependiendo de la configuración del sistema*, éste puede ofrecer opciones de acceso multiusuario, niveles de seguridad, reorganización y recuperación de ficheros, etc.

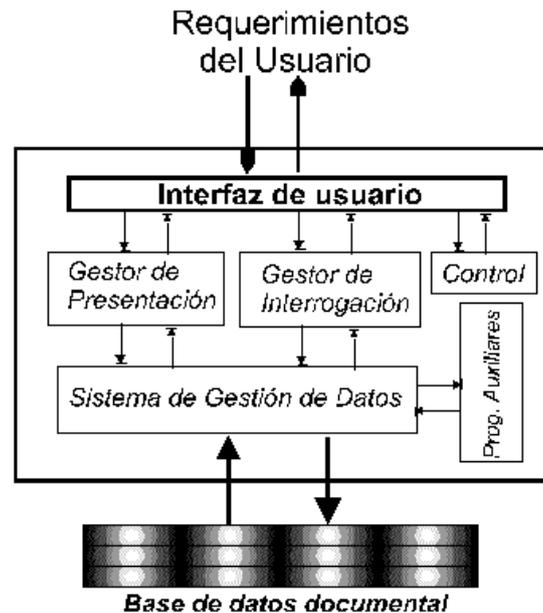


Figura 2.3. Esquema de un IRS (Information Retrieval System)

En general, las funciones anteriores se engloban en alguno de los seis subsistemas genéricos que debe ofrecer un SARI:

- 1.- Subsistema de selección de documentos.
- 2.- Subsistema de indización.
- 3.- Subsistema de vocabulario.
- 4.- Subsistema de búsqueda.
- 5.- Subsistema de interacción hombre-máquina.
- 6.- Subsistema de comparación.

#### 2.2.4 CONSULTAS AL SARI - LAS ECUACIONES DE BÚSQUEDA

La búsqueda se realiza sobre un conjunto de términos introducido en el sistema por el usuario. Sin embargo, esta es la fase final de un proceso que comienza por el análisis de los documentos que se van a introducir, los datos que los conforman, y cómo estos forman una

estructura de datos. Existen múltiples metodologías de diseño de estructuras de datos, cada una con sus virtudes y sus defectos. La estructura de datos se plasma en el sistema informático en una estructura de base de datos, a partir de la cual se introducen los documentos. La organización de la información con vistas a su recuperación dentro de un sistema documental ha producido abundante bibliografía en la que pueden analizarse múltiples aspectos.

### *1. El proceso de búsqueda.*

El proceso de búsqueda de la información es más complejo de lo que pueda parecer en un primer momento, y tiene, con posibles variaciones, las siguientes fases:

- 1.- Definición del tema de búsqueda.
- 2.- Selección de términos descriptivos sobre el tema de búsqueda. Consulta de tesauros o diccionarios, etc.
- 3.- Selección de las bases de datos sobre las que realizar la búsqueda.
- 4.- Formulación y ejecución de ecuaciones de búsqueda.
- 5.- Evaluación de los resultados. Replanteamiento, si procede, de las ecuaciones para ajustar la búsqueda.

Se debe considerar, además, que la definición del tema y de los términos para la búsqueda se hace en lenguaje natural, lenguaje que debe traducirse a la terminología empleada en la base de datos, y que los términos empleados en las ecuaciones son la traslación al lenguaje de interrogación de la base de datos de los términos originales, combinados mediante operadores que aplican lógica matemática.

Tampoco hay que olvidar que la localización de esos términos dentro de la base de datos está en muchos casos sujeta a la estructura de campos existente. Lo anterior hace ineludible la utilización de lenguajes documentales para la correcta explotación de las bases de datos documentales. En principio, debe aceptarse que la búsqueda incluye una serie de términos significativos, los cuales describen el área de interés para la recuperación de información. Estos términos mantienen entre ellos algún tipo de relación semántica.

Esta relación debe reflejarse a través de la utilización de operadores que la reflejen, y por esta causa una ecuación de búsqueda se compone de dos tipos de elementos. En primer lugar, los términos que representan el tema de interés, y en segundo lugar, un conjunto de operadores que expresan la relación que deben mantener los términos entre sí.

### *2. Tipos de operadores utilizados en las ecuaciones de búsqueda.*

La combinación de los términos necesarios para la recuperación de los documentos adecuados se basa en la utilización de unos operadores, especialmente los denominados booleanos. La combinación de los términos mediante los operadores crea una forma matemática, una ecuación, susceptible de ser tratada por medios informáticos.

Los operadores booleanos establecen relaciones entre los términos, y son O (operador de unión), Y (operador de intersección) y NO (operador de exclusión). Trabajan sobre el conjunto de los documentos, estableciendo subconjuntos con aquellos documentos que se ajusten a las condiciones fijadas en la ecuación.

La complejidad que rodea a los mecanismos de recuperación de la información ha hecho necesario aplicar otros operadores, que complementen a los booleanos. En primer lugar, los operadores de localización, referidos a la posición de los términos dentro de los documentos. Se basan en una hipótesis según la cual la cercanía entre dos términos puede significar una estrecha relación entre ellos. Pueden ser absolutos, cuando se establece que un término debe aparecer en un campo determinado, independientemente del resto de los términos incluidos en la ecuación, o relativos, cuando se establece que un término debe tener una posición referida a otro término incluido en la ecuación, por ejemplo en la misma línea, la misma frase, separados por "x" palabras, etc. Ambos pueden combinarse.

También se encuentran los operadores de rango o intervalo, que establecen un intervalo, dentro del cual deben encontrarse los términos a recuperar, especialmente utilizados en la recuperación de datos numéricos y de fechas.

Las dificultades que se desprenden de las características intrínsecas de los términos también afectan a la recuperación de la información. Se pueden plantear problemas con el uso de plurales de los términos, géneros o prefijos y raíces. Para intentar reducirlos, se emplean símbolos de truncamiento, cuya finalidad es indicarle al sistema de recuperación que se está buscando una cadena de caracteres dentro de un término, no un término en sí mismo.

Todos los operadores señalados pueden combinarse entre sí para crear ecuaciones complejas, considerando que siempre se establecen unas limitaciones por parte de los operadores. Normalmente, se ejecutan primero aquellas operaciones con los operadores más restrictivos, y su resultado se combina con los términos relacionados con operadores menos restrictivos. Los sistemas de gestión de bases de datos documentales incluyen en su documentación tanto los operadores disponibles y su uso, como las limitaciones y restricciones establecidas.

### *3. La recuperación mediante exploración.*

La utilización de ecuaciones de búsqueda como principal herramienta en la recuperación de información plantea problemas, que los enfoques actuales de investigación pretenden resolver mediante la creación de mejores interfaces de usuario, que reflejen los mapas cognitivos de los usuarios, y permitan generar imágenes de los mismos que puedan ser utilizadas para interrogar los recursos de información. Interrogación que debería verse complementada por la capacidad de los sistemas de añadir funciones de exploración de los contenidos informativos, durante el proceso de recuperación.

El fundamento de los procesos de exploración es la revisión y análisis de documentos por sus contenido, de forma que el sistema sea capaz de llevar al usuario a otras informaciones similares o relacionadas, sin necesidad de formular nuevas ecuaciones de búsqueda.

### *4. Los resultados: análisis y modificación.*

Una vez considerado todo lo expuesto, y realizada una búsqueda, es de rigor analizar su pertinencia. Una alta pertinencia se caracteriza por la recuperación de los documentos adecuados a la información que se deseaba obtener, evitando la introducción de otros que puedan distorsionar el conjunto, y cumplir los requisitos de una correcta recuperación de la información

En sistemas de tratamiento y recuperación de información documental, ésta debe ser exacta, exhaustiva, precisa, oportuna, íntegra y significativa. Por contra, el silencio (falta de documentos) y el ruido (exceso de documentos, muchos de ellos no significativos) caracterizan a un bajo nivel de pertinencia.

Como resultado del análisis anterior, cabe plantearse la adecuación de la respuesta obtenida a lo esperado. Pueden darse dos situaciones, ante las cuales deben adoptarse diferentes acciones. En un primer caso, puede suceder que el resultado de la ecuación sea

demasiado escaso. Entonces procede ampliarla, lo cual puede hacerse con la utilización de términos más genéricos, sinónimos, o ampliación de truncamientos. En un segundo caso, si el resultado obtenido es excesivamente amplio, deben utilizarse medidas contrarias: utilización de términos más específicos, reducción de truncamientos, etc.

### 2.3 ALMACENAMIENTO DE DATOS

El factor crucial que determina la calidad y el éxito de un SARI es su capacidad para manipular la información textual, tanto en lo que se refiere a su adquisición y tratamiento, con vista a su posterior aprovechamiento, como en lo que respecta a la recuperación de la información contenida en el mismo. Mayor importancia alcanza este factor cuando el usuario debe aceptar que la presencia de un término en un documento no asegura la pertinencia del mismo al objetivo de la búsqueda (por ejemplo, la frase "este documento no versa sobre bibliotecas").

En gran parte de las ocasiones, el texto que se introduce en la base de datos no es sometido a un riguroso control sintáctico y terminológico. Esto obliga a disponer de sistemas que no sólo se centren en el término, sino que sean capaces de reconocer patrones de texto, para superar estas limitaciones.

La primera acción ejecutada por el sistema cuando recibe un nuevo documento o registro es su indización o indexación. El sistema trata de identificar individualmente a cada uno de los términos presentes en el documento que poseen significado propio, eliminando aquellos cuya frecuencia suele ser tan alta que no son significativos, y que corresponden a lo que se engloba bajo el término "palabras vacías" (artículos, conjunciones...).

Para facilitar el trabajo, los sistemas utilizan directamente una lista preexistente con esos términos, lo que agiliza la tarea. Sin embargo, la indización no permite calibrar el peso específico de un término en un documento, lo que podrá ser llevado a cabo por el sistema con posterioridad, tomando como punto de inicio el propio índice.

La información obtenida en el proceso de indización, que es introducida en el fichero inverso como se ha señalado anteriormente, debe ser lo más completa posible. No debe limitarse a la presencia o ausencia del término, sino que debe ofrecer datos sobre la posición absoluta del término en el contexto del documento. Esta información va a ser la que permita efectuar búsquedas sobre el texto que utilicen criterios de presencia, posición absoluta y proximidad.

Sin embargo, es necesario disponer de otras técnicas que completen los mecanismos de recuperación de información, y que se engloban en técnicas de asociación de términos y técnicas de asociación de documentos.

Las técnicas de asociación de palabras o términos intentan buscar y ofrecer al usuario términos relacionados con los que éste expresa en su consulta. Los más conocidos son los diccionarios y los tesauros y mini-tesauros. Los diccionarios ofrecen significados, sinónimos y antónimos del término deseado, mientras los tesauros ofrecen listas de términos relacionados, según varios criterios, con el término en cuestión. Junto a éstos deben citarse las estadísticas de concurrencia, que establecen cuando unos términos aparecen asociados a otros en los documentos de la base de datos, y los truncamientos de raíz, que permiten asociar una raíz semántica con todas las variantes de la misma presentes en la base de datos.

Las técnicas de asociación de registros o documentos facilitan recuperar un registro o documento, y ejecutar una serie de cálculos estadísticos para ofrecer al usuario aquellos cuyo contenido es similar matemáticamente, a través del análisis de las estadísticas de concurrencia. Deben citarse las medidas de similaridad, que se aplica a dos documentos analizando la frecuencia de las palabras en ambos; el *clustering*, utiliza la técnica de agrupación de registros similares en grupos; el *análisis de marcas*, que transforma los términos en valores numéricos, formando una representación del documento y agrupa los similares; y los *métodos discriminantes*, que utilizan un conjunto limitado de términos para incluir o excluir los documentos de los grupos.

## 2.4 RECUPERACIÓN DE INFORMACION

La recuperación de información es el conjunto de tareas mediante las cuales el usuario localiza y accede a los recursos de información que son pertinentes para la resolución del problema planteado. En estas tareas desempeñan un papel fundamental los lenguajes documentales, las técnicas de resumen, la descripción del objeto documental, etc., es necesario recordar, que la informatización de estas tareas se engloba dentro del campo de la Documentación Automatizada o Informática Documental.

### 2.4.1 Lenguajes de interrogación y operadores

Se puede definir a un lenguaje de interrogación como un conjunto de órdenes, operadores y estructuras que, organizados según unas normas lógicas, permiten la consulta de fuentes y recursos de información electrónica. El resultado de la combinación de estos elementos, siguiendo las normas establecidas, es una expresión, a la que se identifica con el nombre "ecuación", capaz de interrogar el contenido de la fuente de información.

Las normas lógicas que rigen un lenguaje de interrogación responden a cuestiones relacionadas con la coordinación de los elementos, es decir, con la formulación de ecuaciones. Estas normas (a modo de sintaxis) especificarán el orden de los elementos, la disposición de las estructuras, sus posibilidades combinatorias, las prioridades en la ejecución, y todo tipo de posibles funciones.

Las órdenes serán aquellas palabras o abreviaturas que le indicarán al sistema las acciones a ejecutar (buscar la expresión, mostrar los documentos o registros resultantes, consultar el tesoro o los ficheros inversos, ejecutar un perfil de usuario...).

Sin embargo, no todos los lenguajes de interrogación utilizan las mismas palabras como órdenes, aunque las órdenes ejecuten las mismas funciones. Existen intentos para homogeneizar la interrogación de las bases de datos, como el lenguaje CCL<sup>4</sup> promovido por la Unión Europea, que aún no han alcanzado el objetivo para el que fueron desarrollados. A este panorama se une la proliferación de interfaces gráficos de usuario, que sustituyen a las órdenes y la sintaxis tradicional, dejando al usuario (si éste lo desea) sólo la labor de introducir los términos y los operadores que expresan las relaciones existentes entre ellos.

En un lenguaje de interrogación, los operadores son los encargados de expresar las relaciones que mantienen entre sí los términos que definen (más adecuado sería decir que pueden definir) las necesidades informativas del usuario. Pueden distinguirse diferentes tipos de operadores, que se analizan a continuación.

#### 2.4.1.1. Operadores lógicos (o booleanos)

Los operadores lógicos, llamados booleanos, son los más utilizados en numerosos sistemas. El principio que rige la utilización de este tipo de operadores es que las relaciones entre conceptos pueden expresarse como relaciones entre conjuntos. Las ecuaciones de

---

<sup>4</sup> CCL: Common Command Language

búsqueda pueden transformarse en ecuaciones matemáticas, que ejecutan operaciones sobre los conjuntos, lo que da como resultado otro conjunto. Los tres operadores básicos son el operador suma/unión (generalmente identificado como O/OR), el operador producto/intersección (identificado como Y/AND), y el operador resta/negación (identificado como NO/NOT). A su vez estos operadores pueden combinarse entre si, generando operaciones más complejas, como el O exclusivo (elimina la intersección), etc.

No deben obviarse los problemas que plantean los operadores booleanos, independientemente de su potencia. En primer lugar, siempre se plantean en términos de absoluto (presente/ausente), sin considerar el peso específico del término en el contexto. En segundo lugar, exigen un alto valor de precisión en los términos utilizados. Por último requieren claridad en la composición de las expresiones a buscar.

#### *2.4.1.2. Operadores posicionales.*

La utilización de operadores posicionales pretende superar algunas de las limitaciones que ofrecen los operadores booleanos. Toman como punto de partida la consideración del valor del término dentro del contexto, es decir, que la posición de ese término en relación con otros, o dentro del propio registro, es significativa para valorar su pertinencia a los objetivos buscados. Los operadores posicionales pueden dividirse en dos tipos:

##### *2.4.1.2.1. Posicionales absolutos.*

Son aquellos que permiten buscar un término en un lugar dado del documento o registro. Por regla general, son operadores de campo, es decir, permiten al usuario fijar en que campo o campos presentes en la estructura de base de datos debe aparecer el término buscado. La presencia del término en un campo dado (por ejemplo, en el campo título), puede ser una garantía de la adecuación del documento a los objetivos, en la mayor parte de las situaciones.

##### *2.4.1.2.2. Posicionales relativos.*

También llamados de proximidad, se trata de operadores que permiten establecer la posición de un término respecto a otro dado. Se considera que la cercanía entre los dos términos puede reflejar una íntima relación entre los conceptos reflejados por los mismos. Estos operadores permiten definir el nivel de proximidad entre los términos (mismo campo, línea, frase, número de términos significativos que los separa...).

#### *2.4.1.3. Operadores de comparación.*

Especifican el rango de búsqueda, fijando unos límites para la misma. Estos límites pueden ser tanto numéricos como alfabéticos, correspondiendo los operadores a formas del tipo "mayor que", "menor o igual que". Se utilizan principalmente en documentos que pueden contener datos numéricos.

#### *2.4.1.4. Operadores de truncamiento.*

Pueden darse situaciones en las cuales sea necesario utilizar no un término simple, sino también sus derivados, fijados por prefijación o sufijación, mínimas variantes léxicas, etc. Para facilitar la búsqueda de este tipo se han introducido operadores de truncamiento, a los que también se llama máscaras. Se trata de operadores (normalmente símbolos como \*, \$), cuya presencia puede sustituir a un carácter o a un conjunto de caracteres, situados a la izquierda, dentro o a la derecha del término en cuestión.

En los actuales sistemas de recuperación de información es posible encontrar todos estos tipos de operadores, que pueden combinarse entre sí, permitiendo crear ecuaciones complejas que reflejan con bastante precisión los conceptos y sus relaciones. La combinación de los operadores debe respetar un conjunto de reglas, básicas en todos los sistemas, que establecen las prioridades y formas de ejecución de ecuaciones complejas, cuando éstas combinan más de dos conceptos.

En primer lugar, los sistemas tienden a resolver, a ejecutar en primer lugar, aquellas expresiones que se relacionan utilizando el operador más restrictivo o prioritario. Por ejemplo, un operador posicional absoluto posee un nivel de restricción (una prioridad) mayor que un operador booleano, lo que significa que el sistema ejecutará antes la expresión cuyo operador es el posicional absoluto, combinando posteriormente el resultado con el operador booleano y su término relacionado.

Sin embargo, pueden darse expresiones en las cuales sea necesario variar estas prioridades, y ordenar al sistema que ejecute en primer lugar expresiones con operadores de menor nivel de restricción, relacionando luego su resultado con términos a través de operadores más restrictivos. Para estas situaciones, se utilizan paréntesis, los cuales engloban a las expresiones que deben ejecutarse en primer lugar, independientemente de las prioridades fijadas por el sistema. La utilización de expresiones entre paréntesis hace posible, por ejemplo, que el resultado de una expresión con un operador booleano pueda

ser combinada con un operador posicional absoluto. Además, los paréntesis pueden anidarse, resolviéndose las ecuaciones planteadas desde dentro hacia fuera, de la misma forma que las igualdades y polinomios matemáticos.

#### **2.4.2 El proceso de recuperación**

Un proceso de recuperación, al que podríamos considerar "genérico" seguiría las siguientes fases:

1. Definición de las necesidades informativas del usuario.
2. Selección y ordenación de las fuentes a utilizar.
3. Traslación de las necesidades del usuario al lenguaje documental propio de la fuente a utilizar en cada caso. Es posible, además, encontrar fuentes en las que no se utilice ningún tipo de vocabulario controlado, en cuyo caso resultará necesario afinar el trabajo terminológico.
4. Traducción de la expresión de lenguaje documental al lenguaje de interrogación propio de cada sistema.
5. Ejecución de las expresiones del lenguaje de interrogación obtenidas.
6. Consulta de las respuestas obtenidas, para analizar su pertinencia o no a la cuestión planteada.
7. Replanteamiento, si procede, de las expresiones utilizadas, si los resultados obtenidos no son pertinentes.
8. Selección y obtención de los documentos que respondan a las necesidades manifestadas por el usuario.
9. Transmisión del resultado, preparado adecuadamente, al usuario.

Todas las fases son susceptibles de tratamiento informático, aunque éste queda claramente resaltado en las fases 5, 6 y 8. La perspectiva tradicional de la teledocumentación, ampliamente expuesta en numerosas obras de referencia, ha servido como base a la

estructura de fases propuesta, aunque es necesario puntualizar que la expansión y la aparición de nuevas técnicas informáticas pueden modificar tanto el planteamiento como la ejecución de las acciones encaminadas a acceder a la información.

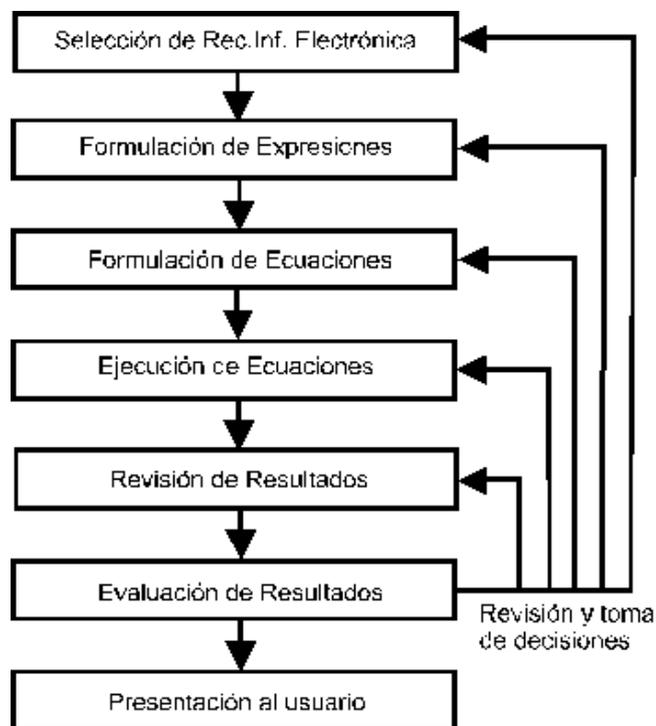


Figura 2.4. Proceso de recuperación en un entorno informático.

### 2.4.3 Estrategia de la interrogación

Los lenguajes, sus órdenes y operadores son utilizados dentro del proceso de recuperación de información, la cual se encuentra almacenada en un repositorio, que suele ofrecer la forma de base de datos. La base de datos es consultada mediante la ejecución de búsquedas, expresiones que reúnen los elementos citados con anterioridad, y cuya resolución da como resultado aquellos elementos que responden a la lógica expresada en la búsqueda. El término "estrategia", en lo que se refiere a la consulta de bases de datos, ha servido para identificar diferentes enfoques y conceptos, que engloban desde la visión general del proceso hasta la formulación de ecuaciones individuales.

La estrategia debe ser un plan ideal de interrogación de la base de datos que incluya el objetivo de la búsqueda, el plan general y el plan específico de operación. El objetivo de la búsqueda se obtiene identificando que tipo de información se necesita y sus características.

Una vez definido el objetivo, debe establecerse un plan general de operación, que incluya una selección de la base o bases de datos a consultar, las primeras aproximaciones a los términos a utilizar en las ecuaciones, así como las posibles relaciones lógicas.

El plan específico de operación se pone en marcha una vez obtenidos los resultados del anterior, y debe formular ecuaciones y utilizar términos con el mayor grado de precisión, establecer una secuencia lógica con todo ello, y redefinirlo si es preciso. Independientemente de ambos planes, resulta necesario conocer con anterioridad la respuesta a varias cuestiones que afectan a la interrogación de la base de datos, tales como el contenido y alcance de la base de datos, coste de consulta, lenguaje y operadores a utilizar durante las consultas, límites preestablecidos (por el usuario o el sistema)... todas ellas afectan y modifican el enfoque del interrogador.

#### *Tipos de estrategia.*

En el momento actual, parece más adecuado utilizar el término para identificar el plan general de búsqueda. No existe una única ni perfecta aproximación a las estrategias de interrogación de bases de datos. En la mayor parte de las ocasiones depende de la experiencia del usuario y de la calidad del contenido de los registros existentes en la base de datos, especialmente en lo que corresponde a su control terminológico. La estrategia depende, en gran manera, de la formación, intuición y experiencia del usuario. Tomando en consideración la intención del interrogador, la bibliografía señala que pueden varios tipos principales de búsqueda, que pueden clasificarse en dos grandes grupos, sin perjuicio de que puedan darse situaciones en las que se combinen:

#### **1. Categorización por objetivo:**

- Búsqueda de elemento conocido: se trata de búsquedas en las cuales el interrogador sabe cual será la respuesta, cuyo contenido, por ejemplo, utiliza para completar una referencia bibliográfica, o utiliza un número de inventario para comprobar el contenido de un registro dado.
  
- Búsqueda de información específica: el interrogador busca una información específica dada, generalmente sobre un tema concreto y limitado, como trabajos publicados en un año o por un autor.

- Búsqueda de información general: intenta buscar la información sobre una materia o asunto, de forma general, que obtenga una visión global del estado de la misma.
- Exploración de la base de datos: se trata de conocer que tipos de información y/o documentos se encuentran almacenados en la base de datos, a qué pueden responder y cómo pueden utilizarse.

## **2. Categorización por plan de operación:**

- Búsqueda directa: se trata de una aproximación expeditiva, en la que se intenta resolver el problema con la formulación de una única consulta. Como puede deducirse, resulta difícil obtener buenos resultados con la misma.
- Búsqueda "breve": es una evolución de la anterior, en la que se trata de recuperar unos items significativos entre un gran número obtenido tras una sola ecuación.
- Ampliación: comienza con ecuaciones muy restrictivas, que ofrezcan documentos pertinentes. Tras analizar la respuesta, el usuario puede ampliar o expandir las ecuaciones de búsqueda hasta recuperar toda la información existente. Puede ofrecer problemas si la ecuación inicial no es adecuada.
- Restricción: opuesta a la anterior, formula ecuaciones que ofrecen resultados muy amplios, para posteriormente utilizar ecuaciones más restrictivas, hasta delimitar los documentos pertinentes.
- Construcción de bloques: intenta establecer bloques de información que respondan a los componentes de su lógica, para combinarlos entre sí posteriormente, hasta encontrar una combinación que responda a las necesidades planteadas.

### **2.5 .- COMPARACIÓN CON OTRAS BASES DE DATOS**

Los sistemas de almacenamiento y recuperación de información documental (SARI) manipulan documentos, en su más amplia acepción, frente a los sistemas de bases de datos (SGBD), que manipulan registros.

Limitándose al contexto de la Documentación, podría decirse que los SARI están trabajando con construcciones cognitivas que representan creaciones humanas diferentes a las que

representan los registros presentes en los SGBD, más estáticos y sujetos a normas fijas. Esto no es óbice para poder utilizar un SGBD para una finalidad documental, ya que los fundamentos teóricos de las bases de datos, como se ha visto anteriormente, son iguales para ambos, pero si debe considerarse que las técnicas de manipulación de unos y otros serán más o menos adecuadas para tal fin.

Sin embargo, hay que concluir que el desarrollo de la tecnología de bases de datos es continuo, y que su propia dinámica está favoreciendo, cada vez en mayor medida, la aparición de aplicaciones mixtas, y la adopción de técnicas de unas en otros, por lo que el panorama de los SARI, tan claro hace pocos años, está difuminando sus límites, al igual que sucede en otros ámbitos de la informática.

## **2.6 DESARROLLO DE APLICACIONES CON BASES DE DATOS DOCUMENTALES**

Este tipo de Bases de Datos tiene las siguientes alternativas prácticas:

*Bibliotecas virtuales*, la gran mayoría de las bibliotecas existentes en el mundo ofrecen su catálogo de documentos en línea, lo que facilita la búsqueda y abre nuevas fuentes de información para los usuarios.

*Sitios con material de estudio*, no son bibliotecas formales, sin embargo también ofrecen documentos de estudio relacionados con trabajos realizados por alumnos universitarios de distintas partes del mundo. La Base de Datos Documental, está compuesta por trabajos y documentos facilitados por los propios usuarios, con lo que la oferta siempre va en aumento y es bastante variada. La ventaja de estos sitios es que se pueden publicar anuncios sobre un tema en particular del que se necesite información y si algún usuario o navegante tiene dicha información la hace llegar por correo electrónico.

*Organismos de Gobierno*, estos sitios también cuentan con bases de datos documentales haciendo completamente accesibles los documentos y estadísticas a la población, que si no tendrían que acceder a ellos a través de una biblioteca bastante documentada.

Además de Internet, este tipo de Base de Datos tiene , también, aplicaciones en las empresas. Muchas empresas mantienen carpetas de archivos con información sobre los proyectos o negocios realizados. Pues bien, estos podrían ser guardados y tratados bajo el

concepto de Bases de Datos Documentales, de esta forma pueden ser consultados y analizados por múltiples usuarios que necesiten tener acceso a dicha información. La ventaja sería la disminución del tiempo de consulta buscando entre papeles, evitar el deterioro de éstos y permitir el análisis simultáneo de los documentos. Por ejemplo, una empresa constructora que participa en licitaciones de proyectos , ahorraría tiempo en formular su propuesta si consultara a su Base de Datos Documental sobre un proyecto o tema afín.

### **2.6.1 HIPERTEXTO**

En la segunda mitad de la década de 1980 comenzaron a aparecer en el mercado gran cantidad de aplicaciones y herramientas para ordenadores personales, que facilitaban la creación de sistemas hipertextuales a nivel personal, y que en los últimos años incorporan capacidades multimedia, generando de esta forma sistemas hipermedia.

#### **Concepto de hipertexto**

El hipertexto se basa en la escritura, y en la lectura, no secuencial de los documentos. Esto supone la existencia, y la adopción, del criterio de asociación de ideas y conceptos como principio organizador del conjunto de documentos, del conjunto informativo. Este tipo de organización no puede reflejarse adecuadamente en material impreso, ya que en éste predomina la linealidad. Sin embargo, existe la posibilidad de superar esa linealidad, utilizando como herramientas sistemas informáticos capaces de establecer relaciones entre ficheros.

Dos elementos clave en estos sistemas son los de enlace y nodo. Un nodo es un documento o elemento informativo, en soporte informático. Entre nodos pueden establecerse relaciones y asociaciones a través de enlaces. De esta forma, dos documentos que contengan conceptos relacionados pueden ofrecer acceso directo uno a otro, o a la parte de los mismos que sea pertinente. Por otra parte, se establecen varios caminos posibles para acceder a la misma información.

### 2.6.2 HIPERMEDIA.

Un desarrollo de las bases de datos multimedia distribuidas, combinadas con el hipertexto, es el servicio hipermedia.

Hipermedia es un enfoque flexible y muy eficiente, para la gestión de información, que permite la navegación a través de la información distribuida. Su principio básico es la posibilidad de navegar, de forma interactiva, en una red formada por nodos conectados entre sí.

El hipertexto sería un subconjunto del hipermedia, principalmente referido a información textual. La navegación se realiza saltando por la información, cualquiera que sea su tipo, de forma no secuencial, a través de links o punteros, establecidos según ciertos principios inteligentes, e inteligibles.

De esta forma, un sistema hipermedia mostraría todo tipo de información almacenada en diferentes bases de datos y en diferentes formatos, de manera completamente transparente para el usuario.

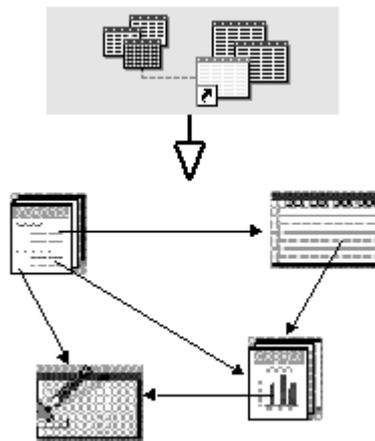


Figura 2.5. Elementos de un sistema Hipermedia

*El hipertexto/hipermedia se basa en el enlace entre informaciones de diferente tipo.*

### 2.6.3 TELEDOCUMENTACIÓN

La teledocumentación, que consiste en el acceso y la consulta de bases de datos existentes en una localización remota, lejana al usuario, mediante la utilización de un ordenador conectado a una red de telecomunicaciones, ha sido considerada durante largo tiempo como el punto de referencia inexcusable en la informática documental.

Sin embargo, resulta necesario situar en su justo lugar la teledocumentación. Esta no sería posible sin la previa tarea de diseño y creación de bases de datos documentales, y sin el desarrollo de diferentes técnicas de recuperación de información.

Es conveniente, en este punto, establecer dos hipótesis de trabajo sobre la teledocumentación. En primer lugar, una de las bases de la teledocumentación es la limitación física del espacio. Tras el nacimiento de numerosas organizaciones gubernamentales, a diferentes escalas, desde inicios del presente siglo, así como la proliferación de la información científica, la cantidad de documentos generados por los mismos, que pueden ser de interés para un usuario en un momento dado, es enorme.

Difícilmente una biblioteca, archivo o centro de documentación pueden ofrecer acceso directo al documento, ya que les es imposible almacenarlos todos físicamente. Y en segundo lugar, aunque fuese posible, el coste económico que supondría la adquisición y mantenimiento de la documentación resultaría inasequible. Para satisfacer a sus usuarios debe ofrecer algún tipo de mecanismo que ofrezca acceso a ese gran volumen de documentación, a un coste razonable.

La solución al problema pasa por que un tercero ponga a disposición de otros muchos usuarios los documentos, y algún mecanismo para facilitar la localización y acceso a los que sean de interés para el usuario. Lo más probable es que ese repositorio de información se encuentre en alguna situación lejana, lo que implica la necesidad de redes de comunicaciones para facilitar el acceso al mismo, así como la utilización de algún tipo de mecanismo que, utilizando esas redes, consulte los contenidos del fondo. De esta manera se llega a la definición clásica de teledocumentación, considerada como la unión de informática, telecomunicaciones y documentación.

El concepto de teledocumentación se ha englobado, por influencia anglosajona, bajo los términos "online searching", término que podría ser válido en la década de 1970. En el

momento actual, con la expansión de las redes de telecomunicaciones y de las bases de datos en soporte CD-ROM, el proceso de búsqueda resulta ser en línea, "on line", el 90% de las ocasiones, lo que hace preferible utilizar el término teledocumentación, con unas características ya definidas, frente al término "online", basado en consideraciones técnicas ya superadas.

### **Los actores de la teledocumentación**

La teoría tradicional sobre teledocumentación establece la existencia de cuatro actores en el proceso teledocumental:

1. *Productores de bases de datos*: son los diseñadores y creadores de las bases de datos, que posteriormente son accesibles al usuario. Estos creadores desarrollan un producto, según diferentes estudios de mercado y de necesidades de información, seleccionado el área de cobertura, los fondos incluidos, y dan soporte al desarrollo y crecimiento posterior de la base de datos.
2. *Distribuidores o "host"*: este segundo actor es el encargado de situar la base de datos a disposición de los usuarios interesados, actuando como intermediario entre el productor y el usuario. Establece una serie de convenios con los productores para ofrecer sus bases de datos, a través de contratos con los usuarios. Normalmente suelen disponer de una compleja estructura informática para hacer frente a numerosas consultas simultáneas por parte de los usuarios. En numerosas ocasiones, los host son también productores de bases de datos, lo que puede abaratar sus servicios. Por contra, los productores suelen licenciar sus bases de datos a varios distribuidores, con la finalidad de favorecer la competencia y obtener mejores resultados.
3. *Redes de comunicaciones*: las redes de comunicaciones son la infraestructura técnica que permite el establecimiento de comunicación entre el distribuidor o host y los usuarios finales. Dependiendo de la situación y de la normativa existente en cada país, la tipología puede ser muy variada, tanto en el ámbito económico como en el ámbito técnico.
4. *Usuarios finales*: son los usuarios y demandantes de la información almacenada en las bases de datos mantenidas por los host. Disponen de una terminal informática conectada a

la red de comunicaciones pertinente, lo que les permite acceder a las bases de datos ofrecidas por el distribuidor, tras el pertinente contrato de ambos servicios.

#### **2.6.4 EL IRS (INFORMATION RETRIEVAL SYSTEM/SISTEMA DE RECUPERACIÓN DE INFORMACIÓN)**

Los IRS son aplicaciones diseñadas ex-profeso para la consulta de bases de datos, por lo que sus funcionalidades se centran en facilitar al usuario la ejecución de consultas, la revisión y la manipulación de los resultados, y la comunicación de los mismos. Carecen, por lo tanto, de las características relacionadas con la creación y administración de bases de datos. Los IRS se encuentran situados en el ordenador remoto o host, en el cual ejecutan las órdenes remitidas por el usuario, a través del software de comunicaciones y/o terminal.

Las funciones más comunes desempeñadas por los IRS corresponden a:

1. *Acceso y notificación*: funciones relacionadas con la conexión y proceso de autenticación de un usuario, así como notificación de las modificaciones, novedades, etc. del servicio que se le presta.
2. *Selección de bases de datos*: permite al usuario seleccionar cual de las bases de datos existentes va a consultar. Casi todos los sistemas ofrecen la posibilidad de buscar en varias al mismo tiempo.
3. *Búsqueda por fichero inverso o tesaurus*: el usuario consulta el contenido de estos ficheros para comprobar la existencia de los términos a utilizar en ecuaciones posteriores, las relaciones con otros términos, etc. Este tipo de consulta no supone la recuperación de registros.
4. *Creación de subconjuntos*: la respuesta a una ecuación crea un subconjunto virtual de documentos, sobre los que es posible ejecutar combinaciones con otros subconjuntos, o acciones de manipulación.
5. *Búsqueda de cadenas de caracteres*: funciones de búsqueda de texto.

6. *Análisis de subconjuntos*: ofrecen funciones que permiten un somero análisis numérico del número de ocurrencias, frecuencias de años de publicación, etc.

7. *Ordenación, visualización y formato de registros*: funciones relacionadas con la presentación por pantalla de los registros, según criterios de ordenación, uso de diferentes formatos de visualización, etc.

8. *Impresión, descarga y solicitud*: tanto en el ordenador local como en el host, se incluye la posibilidad de enviar los registros por correo electrónico. Además, es posible, en algunos casos, obtener copias de los registros para uso local, en formato electrónico, mediante teledescarga. Por último, ofrecen la posibilidad de solicitar una copia del documento primario.

9. *Archivo, ejecución y edición de ecuaciones*: los host ofrecen a sus usuarios la posibilidad de almacenar las ecuaciones de búsqueda de su interés, para ejecutarlas en sesiones posteriores, modificándolas si es preciso. De esta forma pueden incluso crear perfiles de usuario, que faciliten las búsquedas retrospectivas, y la difusión selectiva de la información.

10. *Control de sesión*: información al usuario de los tiempos de conexión, ecuaciones ejecutadas, documentos obtenidos y los costes derivados de todo ello.

## 2.7 CLASIFICACIÓN DE LAS BASES DE DATOS DOCUMENTALES

Los distribuidores pueden ofrecer bases de datos de contenido temático muy variado: medicina, política, derecho, cultura, etc. , por esta razón algunos autores han adoptado una clasificación basada en los trabajos de Fundesco (FUNDESCO, 1986) y de otras fuentes, que tienden a clasificar las bases de datos según el siguiente esquema:

1. *Bases de datos referenciales*: son aquellas que ofrecen registros que a su vez son representaciones de documentos primarios. Dentro de éstas se distinguen, a su vez, entre:

1.1. Bibliográficas: aquellas cuyo contenido son registros de tipo bibliográfico.

1.2. Directorios: aquellas cuyo contenido está referido a la descripción de otros recursos de información, como por ejemplo un directorio de bases de datos.

2. *Bases de datos fuente*: serían aquellas bases de datos que ofrecen el documento completo, no una representación del mismo. También distinguen entre:

2.1. Numéricas: contienen información de tipo numérico, como un censo o indicadores cuantitativos.

2.2. Textuales: ofrecen el texto completo de un documento.

2.3. Mixtas: combinan ambos tipos de información, como por ejemplo informes económicos de empresas, datos geoeconómicos, etc.

De esta clasificación puede deducirse, evidentemente, que la información obtenida podrá ser referencial o factual. Sin embargo, cabe plantear que las modificaciones que están siendo introducidas en el ámbito de las bases de datos, tanto a escala tecnológica como de contenido, como por ejemplo las bases de datos multimedia, introducirán variaciones en esta clasificación, que deberá adoptar criterios más flexibles.

### **El estandarte de internet: WORLD WIDE WEB**

Una breve reflexión sobre el tema lleva a considerar que, en realidad, el ámbito World Wide Web puede ser visto como una gigantesca estructura documental distribuida, en la que los documentos y páginas web corresponden a los registros o documentos clásicos desde el enfoque de una base de datos documental. Aún más, resulta que esos documentos responden a muchas de las características que ofrecen los documentos hipermedia construidos gracias a bases de datos multimedia y repositorios de información. Y por último, la recuperación de información se realiza a través de la formulación de expresiones o

ecuaciones de búsqueda, o a la utilización de índices y catálogos temáticos, en la más pura tradición de la "Information Retrieval".

Se puede establecer que el usuario del ámbito Web encuentra combinados en éste las ideas junto a las más modernas tendencias teóricas de las ciencias de la documentación y la información así como también la aplicación práctica de las mismas en la publicación electrónica y la búsqueda y recuperación de información.