

BASES DE DATOS MULTIMEDIA

Pedro José Blanco Andrés

ETIS

Àlex Alfonso Minguillón

10 Enero de 2005

A mis queridos padres y hermano y a todos aquellos que se sienten orgullosos de mi esfuerzo, que siempre pensaron que lo conseguiría y que me han animado a no dejarlo a mitad de camino. Y a Olga, mi mujer, por su paciencia conmigo

INDICE

1.- INTRODUCCIÓN	6
2.- DOCUMENTACIÓN AUTOMÁTICA.....	8
2.1- Concepto de Information Retrieval	8
2.2- Information Science.....	8
2.3- Concepto de Documentación Automática.....	9
3.- SISTEMAS DE BASES DE DATOS Y LOS SGBD	12
3.1- Las bases de datos	12
3.2- El modelo de arquitectura de bases de datos.....	13
3.3- Los modelos de datos.....	14
3.4- Los sistemas de gestión de bases de datos.....	15
3.5- Los usuarios.....	16
3.6- La creación de bases de datos.....	17
3.7- Creación de una base de datos: enfoque E/R y transformación relacional.....	18
3.7.1- El enfoque entidad-relación de Chen.....	18
3.7.2- La normalización.....	19
3.8- Propuesta de un método estándar de diseño.....	19
3.9- Sistema de bases de datos distribuida.....	21
3.9.1- Estructura de Base de Datos Distribuidas	21
3.9.2- Consideraciones al distribuir la base de datos.....	21
3.9.3 Ventajas de la distribución de datos.....	21
3.9.4- Desventajas de la distribución de los datos.....	22
4.- RECUPERACIÓN DE INFORMACIÓN	23
4.1- Concepto de recuperación de información.....	23
4.2- El problema de los lenguajes.....	24
4.3- El proceso de recuperación.....	25
4.4- Lenguajes de interrogación y operadores	26
4.4.1- Operadores lógicos (o booleanos)	27
4.4.2- Operadores posicionales.....	27
4.4.3- Operadores de comparación.....	28
4.4.4- Operadores de truncamiento.....	28
4.5- Estrategia de la interrogación	28
4.5.1- Tipos de estrategia.....	29
4.6- La exploración como mecanismo de recuperación	29
4.7- Revisión y análisis de resultados	30
4.8- Recuperación de información y sistemas expertos.....	31
5.- SISTEMAS INFORMATICOS DE TRATAMIENTO Y RECUPERACIÓN DE INFORMACIÓN DOCUMENTAL	32
5.1- La gestión informática de documentos	32
5.2- Tipos de sistemas de tratamiento y recuperación de información documental	32
5.3- Especificidad de los STRID	33

5.3.1- La utilización del fichero inverso.....	34
5.3.2- La lógica booleana.....	34
5.3.3- Manipulación de registros de estructura y extensión variable.....	35
5.4- Tratamiento del texto.....	35
5.5- Las funciones técnicas de un STRID.....	36
5.6- Consultas al STRID: las ecuaciones de búsqueda.....	37
5.6.1- El proceso de búsqueda.....	37
5.6.2- Tipos de operadores utilizados en las ecuaciones de búsqueda.....	38
5.6.3- La recuperación mediante exploración.....	39
5.6.4- Los resultados: análisis y modificación.....	39
6.- HIPERTEXTO E HIPERMEDIA.....	40
6.1- Orígenes del hipertexto.....	40
6.2- Conceptos de hipertexto e hipermedia.....	40
6.2.1- Hipertexto.....	40
6.2.2- Hipermedia.....	41
6.3- Componentes de un sistema de hipertexto/hipermedia.....	41
6.4- Tipos de sistemas de hiperdocumentos.....	42
6.5- Navegación y exploración de un sistema hipertextual.....	42
6.6- La recuperación de información en un sistema de hipertexto o hipermedios.....	43
6.7- La creación y organización de hiperdocumentos.....	43
7.- BASES DE DATOS MULTIMEDIA.....	45
7.1- Multimedia y nuevos documentos.....	45
7.1.1- Aspectos económicos de la industria multimedia: el producto.....	45
7.1.2- Usuario multimedia.....	46
7.1.3- Contenidos multimedia.....	47
7.1.4- Sectores de mercado; servicios y aplicaciones.....	47
7.1.5- Políticas de desarrollo.....	49
7.2- La integración de multimedia en las bases de datos.....	50
7.2.1- Los documentos multimedia.....	50
7.2.2- Los sistemas de gestión de bases de datos multimedia.....	51
7.2.3- Componentes de una base de datos multimedia.....	51
7.2.4- El mercado multimedia en España:.....	52
7.2.5- Formato de los componentes de una base de datos multimedia.....	53
7.3- DISEÑO Y ARQUITECTURA DE UNA BASE DE DATOS MULTIMEDIA.....	58
7.4- La metodología en una base de datos multimedia.....	60
7.5- Documentación y bases de datos multimedia.....	61
7.6- Hipermedia.....	62
8.- TELEDOCUMENTACION.....	63
8.1- Concepto de teledocumentación.....	63
8.2- Los actores de la teledocumentación.....	63
8.2.1- El papel de los soportes ópticos.....	64
8.2.2- Otros servicios de información en línea.....	64
8.3- Herramientas para el usuario final.....	64
8.4- El IRS (Information Retrieval System/Sistema de Recuperación de Información).....	65

8.5- La creación de bases de datos en línea	66
8.6- Clasificación de las bases de datos.....	66
8.7- Fases tradicionales de una búsqueda teledocumental.....	67
8.8- Estructura de los registros existentes en los host.....	67
8.9- Servicios complementarios ofrecidos por los distribuidores	68
8.10- Ventajas e inconvenientes de la teledocumentación	68
9.- LA SEGURIDAD EN EL DISEÑO DE BASES DE DATOS MULTIMEDIA Y SISTEMAS DE INFORMACIÓN.....	69
9.1- Concepto de Seguridad.....	69
9.2- Modelos de seguridad en Bases de Datos.....	72
9.3- Prototipos de Bases de Datos Seguras	74
10.- FUTURO DE LAS BASES DE DATOS	76
11.- SISTEMAS GESTORES DE BASES DE DATOS EXISTENTES EN EL MERCADO	79
11.1- DB2.....	79
11.2- INFORMIX	80
11.3- InterBase	81
11.4- ORACLE.....	82
11.4.1- Arquitectura de Oracle	87
11.4.2- La instancia Oracle	87
11.4.3- El Área Global del Sistema (SGA).....	88
11.4.4- El Área Global de Programas (PGA).....	88
11.4.5- Las Transacciones.....	89
11.4.6- Roles y Responsabilidades del DBA de Oracle	90
11.5- SYBASE	90
11.6- MsACCESS.....	92
11.7- ADABAS (de Software A.G.).....	92
11.8- Conzept 16.....	93
11.9- Lotus Approach.....	93
11.10- Motores de búsqueda: BDE (Borland Database Engine)	94
11.11- Motores de búsqueda: MS Jet (Microsoft)	94
11.12- Estadística de la distribución de los diferentes SGBD en las empresas españolas.....	95
12.- BIBLIOTECA MULTIMEDIA.....	96
12.1- Ejemplo de una base de datos multimedia (Carnegie Mellon University).....	97
12.1.1- Distribución de servidores de la Carnegie Mellon Universit	101
12.2- Conexión segura mediante VPN (Virtual Private Network).....	102
12.2.1 – VPN sobre Internet. Inconvenientes.....	103
12.2.2 – Túneles VPN mediante L2TP	104
13.- GLOSARIO.....	106
14.- BIBLIOGRAFÍA.....	129

1.- INTRODUCCIÓN

Las Ciencias de la Información y la Documentación, cuya formulación moderna es necesario buscar tomando como punto de partida las obras de P. Otlet, particularmente su monumental *Traité de Documentation*, publicado en 1934, han estado íntimamente relacionadas con el desarrollo de los medios técnicos y tecnológicos. Otlet, como han señalado otros trabajos (ESPINOSA et al., 1994) ya señaló en su momento a la tecnología como una disciplina general, amplia, que estudia los medios mediante los cuales el hombre intervenía en su medio ambiente. La Documentación es una disciplina que, correspondientemente, se encarga de intervenir en los procesos informativos desarrollados por el ser humano, y en la que interviene, con un papel fundamental, la tecnología, o aquella parte del conjunto total de la misma que es aplicable de forma útil al trabajo informativo.

De esta forma, Otlet expone los principios y relaciones de la Tecnología documental, insistiendo en un punto clave: que la tecnología documental debe elaborarse por especificación de la tecnología general. Es decir, que el desarrollo, métodos y técnicas de la Tecnología Documental dependerán de los desarrollos de la tecnología general. La rama de la Documentación encargada de la teorización, metodología y aplicación de los medios tecnológicos es identificada bajo los nombres de Bibliología Tecnológica o Documentación Técnica, mostrando una partición cruzada, dentro de la cual es posible identificar conceptos de plena actualidad.

En primer lugar, Otlet identifica los factores que caracterizan la nueva subdisciplina: existe una parte encargada de estudiar los soportes, y otra que se centra en las máquinas que hacen posible la misma: esta vertiente recibe el clarificador nombre de Documentación Automática. Además, estos medios se emplean en un proceso documental completo, desde la producción del documento hasta su difusión, con unos objetivos claros: ofrecer información actualizada, universal, rápida y verdadera, lo que en la actualidad puede englobarse bajo el término "optimizada". En segundo lugar, coloca en un lugar preponderante la importancia de la organización, como criterio indispensable en la Documentación, y por extensión a cualquier parte componente de la misma. Por último, aunque Otlet lo sitúa en su trabajo previamente, también presupone la existencia de una parte de la Documentación Automática encargada de estudiar el aspecto físico del documento, y de parte otra centrada en el estudio de los procesos documentales.

El breve resumen realizado permite identificar en la obra de Otlet los componentes fundamentales del moderno concepto de Documentación Automática (o Automatizada): una teoría sobre la organización, las herramientas y los soportes; una aplicación práctica al proceso documental, y, por último, un objetivo: satisfacer las necesidades informativas del usuario.

Este trabajo intenta ser un acercamiento a la situación actual de las bases de datos multimedia. Para ello se ha optado por un enfoque múltiple que tiene en cuenta diferentes aspectos del producto multimedia: los aspectos económicos, es decir, la situación del mercado: número de bases de datos, volumen de negocio, crecimiento, etc., los aspectos técnicos, es decir, la herramienta utilizada, su evolución y sus limitaciones y/o problemas, y sus aplicaciones y servicios en distintas áreas: con especial interés en el campo de la Documentación y la Formación e Investigación.

Aunque pueda parecer superfluo, es necesario hacer una primera precisión sobre lo que aquí se entiende sobre base de datos multimedia.

*Un sistema de computo que combina medios de texto, gráficos, animación, música, voz y video; puede incluir sonidos estereofónicos como dispositivos de salida.
Cualquier forma de comunicación que usa más de un medio para presentar información.
También se refiere a un programa de computadora que integra texto, gráficos, animación y sonido.*

Es muy habitual, incluso entre profesionales, referirse con este único término a dos productos informáticos distintos: el Sistema de Gestión de Bases de Datos (SGBD), es decir, el programa que me permite la creación, mantenimiento, explotación y gestión de la BD y por otro lado, a la BD en sí misma como conjunto de información estructurada (que tiene la consideración de producto, si es ofrecida en soporte físico unitario, o de servicio, si es accesible de forma telemática). Se incluyen y revisan aquí ambos elementos: el producto multimedia (como bien de consumo) y la herramienta multimedia (los

sistemas de gestión) como parte fundamental del equipamiento lógico, en que se basa la creación del producto multimedia.

El origen de multimedia se produce principalmente en las artes y educación donde se encuentra una tradición de experimentar como se trabaja con la información. El desempeño de multimedia y exhibiciones, material de entretenimiento multimedia, y presentaciones multimedia; todos estos elementos usan varios canales y modos de expresión. Esta tradición existente es ahora usada por un nuevo tipo de multimedia, uno basado en tecnología digital. Ordenadores de escritorio pueden manipular imágenes fotográficas, grabaciones de sonido y videos domésticos en forma digital. Los medios digitales son combinados y procesados y están emergiendo como elementos clave en la moderna tecnología de información.

Desarrolladores de hardware y software nos bombardean con nuevos productos multimedia diariamente. Grupos de entretenimiento y educacionales están explorando nuevas aplicaciones. Nuevos ordenadores, las telecomunicaciones y las industrias electrónicas están desarrollando nueva tecnología. El termino, "multimedia" es frecuentemente usado pero raramente definido. Usuarios de ordenadores tienen presentaciones multimedia, estaciones de trabajo multimedia y bases de datos multimedia, pero la gente involucrada ven el fenómeno multimedia desde diferentes perspectivas. Por ejemplo, presentaciones multimedia, como su nombre lo dice, involucra la presentación de información multimedia, el multimedia de estaciones de trabajo negocia con el procesamiento de información multimedia y bases de datos multimedia. Entonces, la noción de información multimedia se unifica con multimedia digital.

Un flujo de información, variado en estilo y sustancia, fluye por nosotros en el mundo moderno. Recibimos y procesamos esta información en ambos niveles conscientes e inconscientes. Por ejemplo, caminar a través de una tienda con distintos departamentos, observando una atractiva organización de paquetes en los estantes, que ofrecen sus contenidos. Música suave de fondo incrementa nuestra preparación para comprar, mientras los empleados contestan a preguntas y promueven sus artículos.

En las tiendas más grandes, que tienen personal limitado, videos presentan demostraciones de productos. Pantallas con imágenes colocadas estratégicamente y monitores muestran ventajas de un producto a los compradores que pasan. Una voz calmada, apoyada por una melodiosa música, lee un artículo diseñado para anticipar preguntas de los compradores.

Desde que la información es distribuida tan intensamente hoy en día, la efectividad de los componentes multimedia son constantemente estudiados y revisados. Los anuncios constantemente buscan nuevas técnicas para empatar el creciente potencial de consumidores.

La efectividad de la multimedia ha crecido aun más allá incorporando la interacción del usuario. Aquí el usuario es incluido en el proceso, desde que el o ella controlan el camino y el tipo de información que se presenta. Los procedimientos de aprendizaje más exitosos usan este tipo de interacción.

A lo largo del presente trabajo, iremos desarrollando diferentes aspectos del concepto multimedia, tratando especialmente el concepto de Base de Datos Multimedia, que es el principal lugar de almacenamiento de todo este nuevo tipo de información.

2.- DOCUMENTACIÓN AUTOMÁTICA

2.1- Concepto de Information Retrieval

El desarrollo posterior de la Documentación, así como el nacimiento de varias escuelas teóricas dentro del mismo ámbito, y el desarrollo de diferentes enfoques y perspectivas sobre la misma ha sido objeto de concienzudos estudios (LÓPEZ YEPES, 1995). Sin embargo, en todas ellas el aspecto tecnológico se ha considerado como un medio para desarrollar tareas, profundizando sólo en uno de los ámbitos propuestos por Otlet para la Documentación Automática

Hay que esperar al desarrollo del concepto de Information Retrieval, especialmente en el ámbito anglosajón, en la década de 1950, para la aparición de un corpus teórico y experimental en el que las tecnologías de la información ocupan un destacado papel. Independientemente de la acuñación del término, es este entorno científico y social el que marca la aparición del concepto y el recurso continuo a las tecnologías de la información, esencialmente ordenadores y redes de comunicaciones. Las décadas de 1950 y 1960 ven el auge que alcanzan los estudios sobre comunicación, psicología, lógica, cibernética, telecomunicaciones, teoría de sistemas... al tiempo que una imparable multiplicación de los recursos informativos en todos los campos de la actividad humana. Los especialistas se centran en el problema de la búsqueda y recuperación de información, documental o no, lo que les lleva al comienzo del ciclo, a los principios de representación y clasificación de los documentos según su contenido semántico. Information Retrieval era un síntoma de un cambio más profundo, que llegó en 1966-68, con la introducción de la Information Science (Ciencia de la Información), como ciencia integradora de la teoría, proceso y práctica documental, con otras ciencias complementarias, como la cibernética, la informática, la teoría de la información y la comunicación, etc. Biblioteconomía y Documentación serían, desde esta perspectiva, aplicaciones particulares de la Information Science.

Del impacto de la informática (no debe olvidarse que en Norteamérica no se utiliza un término general como informática, sino que se utilizan sus partes como computer science, computer engineering, data processing...) se deduce la presencia de términos que definen a profesionales relacionados con la tecnología, como information specialist o information engineer. El éxito de la Information Science, independientemente de los problemas teóricos y conceptuales pertinentes a toda disciplina científica, facilitaron introducción cada vez más profunda de las tecnologías de la información, en especial la informática, en numerosos aspectos del proceso informativo documental. El impacto favorece que, en ámbito anglosajón se acuñe el término Informatics como la unión entre Computer Sciences e Information Science.

El desarrollo de nuevas teorías ha traído, de la mano de la Ciencia de la Información, la aparición de la disciplina Information Management (Gestión de Información y la Documentación en las Organizaciones), en la que desempeñan un papel fundamental las telecomunicaciones y la informática, íntimamente relacionadas con los sistemas de información, en el marco de redes complejas de información.

La escuela soviética que se acoge al nombre de Informatika también concede un papel importante a las tecnologías de la información. El problema derivado de la utilización del término Informatika, y la dificultad para su aceptación en el ámbito europeo, a causa de la homofonía con el término que engloba las ciencias de los ordenadores y sus aplicaciones (Informática), ha sido la causa de proponer un nuevo término, Informología, con un contenido similar, aunque desde un posicionamiento teórico original, a la Ciencia de la Información, incluyendo sus vertientes tecnológicas.

2.2- Information Science

El desarrollo e implantación de la Information Science han sido claves en la madurez que ha alcanzado la aplicación de tecnologías de la información en el trabajo informativo. La utilización de medios informáticos ha corrido pareja con el desarrollo de nuevas máquinas y herramientas desde la década de 1960. Los diferentes estadios tecnológicos de los medios informáticos han tenido su reflejo en el tratamiento de la información, y, por ende, en los enfoques teóricos que han guiado la misma. De Mey ha trazado una útil evolución del desarrollo del tratamiento de información (DE MEY, 1980):

1. Una etapa monádica en la cual las unidades de información se tratan independientemente unas de otras, como diferentes entidades existentes por sí mismas (por ejemplo, un libro con un término que lo define).
2. Una etapa estructural, en la que la información se ve como una entidad compleja, compuesta de unidades simples de información, relacionadas de algún modo, cuya estructura es objeto de estudio (por ejemplo, las frases o párrafos en un libro, con palabras clave y análisis sintáctico)).
3. Una etapa contextual, en la que al enfoque de la etapa anterior se une un análisis del contexto en el que se sitúa la información, para eliminar o reducir la ambigüedad que puede producirse en el significado del mensaje (por ejemplo, utilizar los valores semánticos)
4. Una etapa cognitiva, en la que la información se ve como un complemento a un sistema conceptual que representa el conocimiento del sistema de procesamiento de información del mundo (por ejemplo, sistemas basados en el conocimiento y sistemas adaptativos).

Para este autor, los actuales sistemas informáticos habrían cubierto completamente las fases 1 y 2, actuarían con apoyo humano en la fase 3, y se encontrarían en un momento experimental en la fase 4.

Esta fase 4 es el objeto principal de investigación de Ingwersen (INGWERSEN , 1992), y ha sido tratado por otros autores, entre los que debe destacarse Ellis (ELLIS, 1990). Para Ingwersen pueden diferenciarse varios enfoques en el tratamiento y recuperación de la información, coincidentes con el desarrollo histórico de la informática y de las fases propuestas por De Mey, que culminan con la utilización de un enfoque cognitivo. De esta forma diferencia un primer enfoque tradicional, centrado en los problemas de representación del texto y recuperación de información mediante técnicas mecanicistas. La superación del anterior se obtiene mediante un enfoque orientado al usuario, potenciando la figura del intermediario humano, y considerando todos los procesos de comunicación y representación de la información. El enfoque cognitivo, por último, pretende superar los problemas de representación y recuperación de la información mediante la creación de mecanismos que reflejan las estructuras cognitivas del usuario, tanto en lo que respecta a modelos como a estados. Puede deducirse, por lo tanto, que en este enfoque cobran especial importancia tanto la semántica del texto y el estudio del lenguaje natural, en lo explícito y en lo implícito, como en la contextualización, en su más amplia acepción, de las necesidades de información del usuario en un momento y situación dados. Para Ingwersen, el desarrollo de nuevos y mejores mecanismos de tratamiento y recuperación de la información debe basarse en la combinación de los mecanismos clásicos de recuperación de información, con métodos y elementos de polirepresentación, contextualización e interacción que son relevantes a manifestaciones de necesidades de información de los usuarios.

2.3- Concepto de Documentación Automática

La construcción de un concepto en el ámbito científico tiene dos vertientes: en primer lugar, la elección del término que lo englobará, y en segundo lugar, la redacción de la definición del término. Problema éste presente en todas las ciencias, por el propio carácter dinámico de las mismas, siempre sometidas a continua revisión, al que no son ajenas las Ciencias de la Información y la Documentación, por su propia riqueza epistemológica. En el ámbito que nos ocupa la numerosa terminología, en ocasiones contradictoria, ha favorecido la aparición de términos que pretenden englobar al objeto de estudio: bibliología tecnológica, informatics, informática documental, documentique, documatique, informatika, documática, tecnologías de la información, information retrieval... cada una de las cuales con su definición correspondiente, en las que se pueden identificar matices y particularidades de diversos tipos y niveles, aunque todos ellos ofrecen un sustrato común.

Ante esta situación, es necesario comenzar por el término. En esta vertiente resulta adecuado inclinarse, al menos desde nuestra perspectiva, por la vuelta a la fuente clásica, el trabajo de Otlet, citado con anterioridad. Sin entrar en disquisiciones sobre el término "automática", resulta evidente que Otlet ofreció con claridad meridiana un término válido en la actualidad, como es Documentación Automática, tomando como punto de partida la situación tecnológica del tiempo que le tocó vivir. La utilización del calificativo "Automática" resulta de suma importancia, ya que está haciendo referencia directa a la aplicación de determinadas tecnologías. Otros términos posteriores pueden dar lugar, por el problema terminológico citado, a cuestiones sobre la automatización o informatización de las tareas que definen (en puridad, puede hablarse de information retrieval sin tratar cuestiones informáticas). Desde otro punto de vista, distinguir entre Biblioteconomía Automática, Archivística Automática, Gestión de Información Automática, etc, etc, parece un ejercicio de división innecesario a un primer nivel, ya que comparten teoría, métodos y herramientas con la Documentación Automática, distinguiéndose sólo por los

componentes contextual y organizativo. Por lo tanto, se prefiere el término Documentación Automática, independientemente de que los términos sean unidos para formular el nuevo término Documática.

Una vez adoptado el término de Documentación Automática/Documática, debe resolverse el problema de la definición, sus componentes y organización. Para establecer un continuum con el criterio de unidad adoptado en lo referente al término, la redacción de la definición debe ofrecer una integración de elementos restrictivos que permitan dotar al concepto de unos límites adecuados, con elementos sincréticos que doten a la definición de validez frente a nuevos desarrollos tecnológicos. Para ello se toman como punto de partida dos definiciones de Information Retrieval que son pertinentes:

- ❖ "Information retrieval is concerned with the processes involved in the representation, storage, searching and finding of information which is relevant to a requirement for information desired by a human user... The objective is to study and understand IR processes in order to design, build and test retrieval systems that may facilitate the effective communication of desired information between human generator and human user." (INGWERSEN, 1992: 49)
- ❖ "The central problem of Information Retrieval is how to represent documents for retrieval... The central task of Information Retrieval research is to understand how documents should be represented for effective retrieval." (BLAIR, 1990: vii).

En ambas definiciones se insiste en la importancia de la representación de documentos, con la finalidad de poder ofrecer respuestas adecuadas a una necesidad de información planteada. La definición de Ingwersen añade dos elementos más de sumo interés. En primer lugar, el diseño de sistemas para la recuperación de información; en segundo lugar, el proceso de comunicación inherente a todo proceso informativo y documental. Sin embargo, ninguna de las dos definiciones hace referencia a los medios tecnológicos que pueden ser de aplicación, ni a criterios de organización relacionados con el contexto en el que se sitúe el sistema. Debe señalarse, además que ambas hacen referencia no a datos, sino a información (information) y a documentos (documents).

Esta última cuestión resulta de suma importancia en nuestro planteamiento, ya que permite diferenciar la disciplina en discusión de otras como el procesado de datos, al que, por norma general se adscribe a la disciplina informática. Esta sutil diferencia, ya señalada por Blair (BLAIR, 1990) va a servir como punto de partida para un análisis del material de trabajo de Biblioteconomía, Archivística y Documentación. A la pregunta de cuál es el material de trabajo de un profesional de estos ámbitos, la respuesta más rápida es "libros, seriadadas, protocolos, documentación administrativa..." Es necesario realizar un análisis previo: cuando un profesional se enfrenta a un libro o a un protocolo, lo analiza, establece sus características físicas, de forma acorde a las normas existentes, para posteriormente realizar una representación de su contenido informativo (mediante la aplicación de lenguajes documentales u otro tipo de métodos). Ahí termina el trabajo con el documento primario. Además, ese trabajo ha consistido en extraer información, que no datos, para generar una representación plenamente informativa del documento primario. A partir de ese momento, todas las acciones del profesional, independientemente del tipo de unidad de información en el que se encuentre, se centran en la información que generó, que utiliza para satisfacer necesidades de usuarios y para los mecanismos de gestión y planificación de la unidad. Sólo se recurre al documento original para su préstamo o consulta temporal por parte del usuario. Esta diferencia es clave, a nuestro juicio, para diferenciar la Documentación Automática dentro del ámbito general de la disciplina Informática.

Desde estas consideraciones, la Documentación Automática /Documática es la disciplina que se ocupa de la investigación y aplicación de las Tecnologías de la Información en todos los ámbitos de las Ciencias de la Documentación, teniendo como material de trabajo la información generada y requerida durante el proceso documental que se trate, en el marco de los sistemas de información, en un contexto que integra los medios automáticos, el ser humano y las interacciones entre ambos.

La Documentación Automática/Documática ofrece dos vertientes; correspondientes a la investigación y desarrollo de nuevas aplicaciones, y a la utilización práctica de las mismas. En la primera participa de la interdisciplinariedad con otras ciencias, como la Informática, las Ciencias Cognitivas o las Ciencias de la Organización y la Gestión. El objetivo es obtener nuevas aplicaciones o herramientas informáticas para el tratamiento y recuperación de la información, así como métodos para la implantación, gestión y control de estas aplicaciones o herramientas dentro de diferentes tipos de organizaciones. La segunda es la encargada de aplicar las herramientas al entorno o contexto que se precise, de forma que se optimicen los

procesos documentales e informativos que se desarrollen en el mismo. Ambas se encuentran relacionadas, ya que la investigación se beneficia y alimenta de las experiencias obtenidas en la aplicación de las herramientas que desarrolla, y a su vez las nuevas herramientas permiten mejorar y adecuar cada vez más los procesos documentales que se desarrollan en las unidades de información, al mismo tiempo que intervienen, en numerosas ocasiones, en los procesos de organización y gestión de las mismas. Junto a esta interacción, no debe olvidarse la integración, tanto vertical como horizontal, de la Documentación Automática/Documática, tanto con la Ciencia de la Documentación, como con la Informática y la Ciencia de la Información, ésta última en su más amplia formulación

3.- SISTEMAS DE BASES DE DATOS Y LOS SGBD

3.1- Las bases de datos

Una de las conclusiones obtenidas en el capítulo anterior era que los recursos de información, y los mecanismos necesarios para su interrogación, resultaban ser uno de los objetivos fundamentales en los sistemas de información que debían hacer frente a grandes cantidades de documentos e información en diferentes formatos y soportes. Y uno de los componentes principales de estos sistemas son las bases de datos, o, más concretamente, los sistemas de bases de datos. Resultará evidente, por otra parte, que la gestión del gran volumen de datos demandará una teoría sobre la organización de esos datos para alcanzar la máxima efectividad posible. En este capítulo se va a efectuar una revisión de los principios que inspiran la arquitectura y el diseño de sistemas de bases de datos, poniendo especial interés en el modelo entidad-relación, por cuanto será el utilizado más adelante para el diseño de las bases de datos con sistemas de gestión de bases de datos relacionales, ya que gran parte de los métodos de modelado conceptual pueden utilizarse igualmente en la construcción de bases de datos documentales.

En el entorno informático, la gestión de bases de datos ha evolucionado desde ser una aplicación más disponible para los computadores, a ocupar un lugar fundamental en los sistemas de información. En la actualidad, un sistema de información será más valioso cuanto de mayor calidad sea la base de datos que lo soporta, la cual resulta a su vez un componente fundamental del mismo, de tal forma que puede llegarse a afirmar que es imposible la existencia de un sistema de información sin una base de datos, que cumple la función de "memoria", en todas sus acepciones posibles, del sistema.

Las bases de datos almacenan, como su nombre dice, datos. Estos datos son representaciones de sucesos y objetos, a diferente nivel, existentes en el mundo real: en su conjunto, representan algún tipo de entidad existente. En el mundo real se tiene percepción sobre las entidades u objetos y sobre los atributos de esos objetos; en el mundo de los datos, hay registros de eventos y datos de eventos. Además, en ambos escenarios se puede incluso distinguir una tercera faceta: aquella que comprende las definiciones de las entidades externas, o bien las definiciones de los registros y de los datos.

La transferencia entre las entidades del mundo real, y sus características, y los registros contenidos en una base de datos, correspondientes a esas entidades, se alcanza tras un proceso lógico de abstracción, conjunto de tareas que suelen englobarse bajo el título de diseño de bases de datos. Sin embargo, es necesario definir, en primer lugar, qué es una base de datos, independientemente de su diseño y/o su orientación. Entre las numerosas definiciones que pueden encontrarse en la bibliografía, pueden escogerse, por su exhaustividad, las siguientes:

- ❖ "Colección de datos correspondientes a las diferentes perspectivas de un sistema de información (de una empresa o institución), existentes en algún soporte de tipo físico (normalmente de acceso directo), agrupados en una organización integrada y centralizada en la que figuran no sólo los datos en sí, sino también las relaciones existentes entre ellos, y de forma que se minimiza la redundancia y se maximiza la independencia de los datos de las aplicaciones que los requieren." (GUILERA, 1993: 377)
- ❖ "Una base de datos es una colección de datos estructurados según un modelo que refleje las relaciones y restricciones existentes en el mundo real. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de éstas, y su definición y descripción han de ser únicas estando almacenadas junto a los mismos. Por último, los tratamientos que sufran estos datos tendrán que conservar la integridad y seguridad de éstos." (MOTA, CELMA y CASAMAYOR, 1994: 9)

La segunda definición añade los objetivos que debe cumplir un sistema de gestión de bases de datos, sobre los cuales se tratará más adelante. Por ahora, basta considerar que deben cumplir los objetivos de independencia de los datos (las aplicaciones no deben verse afectadas por cambios en la estructura de los datos), integridad de los datos (los datos deben cumplir ciertas restricciones que aseguren la correcta introducción, modificación y borrado de los mismos) y seguridad (establecer diferentes niveles de acceso a los datos a diferentes tipos de usuarios).

La entidad existente en el mundo real es objeto de un doble tratamiento, desde el momento en que convierte en objeto de la base de datos. El tratamiento de sus datos se va a realizar en un nivel lógico, por

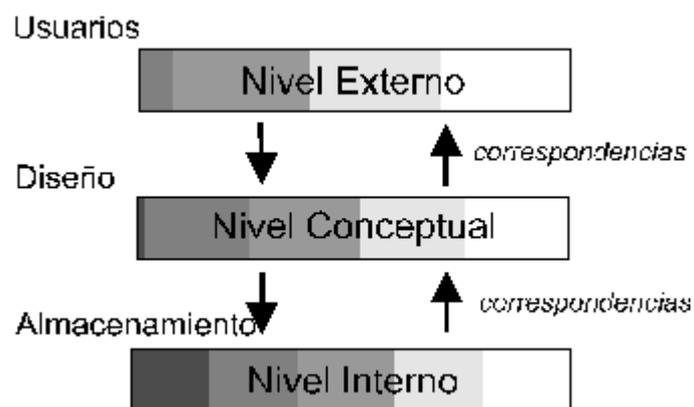
una parte, y en un nivel físico, por otra. En el primero de ellos, el lógico, se va a trabajar en los aspectos referidos a la identificación de las características de la entidad, su descripción y organización, mientras que en el segundo todo lo anterior se va a plasmar en la organización, acceso y almacenamiento de los datos en un soporte físico. Esta división entre un nivel lógico y otro físico se va a reflejar en todos los métodos y conceptos subsiguientes.

3.2- El modelo de arquitectura de bases de datos.

Hasta fecha relativamente cercana, las bases de datos eran el resultado de una compleja programación y de complicados mecanismos de almacenamiento. Con la popularización de la microinformática, la aparición de aplicaciones específicas también trajo con ella la disponibilidad de herramientas de gestión de datos, que acabaron desembocando en los denominados sistemas de gestión de bases de datos, identificados por sus siglas SGBD (DBMS en inglés, siglas de DataBase Management Systems). De esta manera, la gestión de base de datos pudo liberarse de los grandes ordenadores centrales, pudiendo distribuirse según los intereses de los usuarios, y dotando de autonomía en la gestión de información a muchas entidades. Los SGBD permitieron a todo tipo de usuarios crear y mantener sus bases de datos, dotándolos de una herramienta que era capaz de transformar el nivel lógico que éstos diseñaban en un conjunto de datos, representaciones y relaciones, traduciéndolo al nivel físico correspondiente. Para que fuese posible, y para asegurar a los usuarios cierta seguridad en el intercambio de datos entre diferentes sistemas, y en el diseño de ficheros y bases de datos, fue necesario normalizar los esquemas que guiaban la creación de las bases de datos.

Las bases de datos respetan la arquitectura de tres niveles definida, para cualquier tipo de base de datos, por el grupo ANSI/SPARC. En esta arquitectura la base de datos se divide en los niveles externo, conceptual e interno (KORTH y SILBERSCHATZ, 1994:5; MIGUEL y PIATTINI, 1993: 83-107; MOTA, CELMA y CASAMAYOR, 1994: 11-12):

1. Nivel interno: es el nivel más bajo de abstracción, y define cómo se almacenan los datos en el soporte físico, así como los métodos de acceso.
2. Nivel conceptual: es el nivel medio de abstracción. Se trata de la representación de los datos realizada por la organización, que recoge las vistas parciales de los requerimientos de los diferentes usuarios y las aplicaciones posibles. Se configura como visión organizativa total, e incluye la definición de datos y las relaciones entre ellos.
3. Nivel externo: es el nivel de mayor abstracción. A este nivel corresponden las diferentes vistas parciales que tienen de la base de datos los diferentes usuarios. En cierto modo, es la parte del modelo conceptual a la que tienen acceso.



Niveles de la arquitectura de bases de datos.

En ocasiones puede encontrarse el nivel conceptual dividido en dos niveles, conceptual y lógico. El primero de ellos corresponde a la visión del sistema global desde un punto de vista organizativo independiente, no informático. El segundo correspondería a la visión de la base de datos expresada en términos del sistema que se va a implantar con medios informáticos.

El modelo de arquitectura propuesto permite establecer el principio de independencia de los datos. Esta independencia puede ser lógica y física. Por independencia lógica se entiende que los cambios en el esquema lógico no deben afectar a los esquemas externos que no utilicen los datos modificados. Por independencia física se entiende que el esquema lógico no se vea afectado por cambios realizados en el esquema interno, correspondientes a modos de acceso, etc.

3.3- Los modelos de datos.

En el proceso de abstracción que conduce a la creación de una base de datos desempeña una función prioritaria el modelo de datos. El modelo de datos, como abstracción del universo de discurso, es el enfoque utilizado para la representación de las entidades y sus características dentro de la base de datos, y puede ser dividido en tres grandes tipos (KORTH y SILBERSCHATZ, 1993: 6-11):

1. Modelos lógicos basados en objetos: los dos más extendidos son el modelo entidad-relación y el orientado a objetos. El modelo entidad-relación (E-R) se basa en una percepción del mundo compuesta por objetos, llamados entidades, y relaciones entre ellos. Las entidades se diferencian unas de otras a través de atributos. El orientado a objetos también se basa en objetos, los cuales contienen valores y métodos, entendidos como órdenes que actúan sobre los valores, en niveles de anidamiento. Los objetos se agrupan en clases, relacionándose mediante el envío de mensajes. Algunos autores definen estos modelos como "modelos semánticos".
2. Modelos lógicos basados en registros: el más extendido es el relacional, mientras que los otros dos existentes, jerárquico y de red, se encuentran en retroceso. Estos modelos se usan para especificar la estructura lógica global de la base de datos, estructurada en registros de formato fijo de varios tipos. El modelo relacional representa los datos y sus relaciones mediante tablas bidimensionales, que contienen datos tomados de los dominios correspondientes. El modelo de red está formado por colecciones de registros, relacionados mediante punteros o ligas en grafos arbitrarios. el modelo jerárquico es similar al de red, pero los registros se organizan como colecciones de árboles. Algunos autores definen estos modelos como "modelos de datos clásicos".
3. Modelos físicos de datos: muy poco usados, son el modelo unificador y el de memoria de elementos. Algunos autores definen estos modelos como "modelos de datos primitivos".

De lo anterior se deduce que el punto clave en la construcción de la base de datos será el modelo de datos. Se denomina modelo:

"...al instrumento que se aplica a una parcela del mundo real (universo del discurso) para obtener una estructura de datos a la que denominamos esquema. Esta distinción entre el modelo (instrumento) y el esquema (resultado de aplicar el instrumento) es importante... Es importante también distinguir entre mundo real y universo del discurso, ya que este último es la visión que del mundo real tiene el diseñador... podemos definir un modelo de datos como un conjunto de conceptos, reglas y convenciones que nos permiten describir los datos del universo del discurso." (MIGUEL y PIATTINI, 1993: 162)

Los objetivos del modelo de datos son dos:

1. Formalización: definir formalmente las estructuras permitidas y las restricciones a fin de representar los datos de un SI.
2. Diseño: el modelo resultante es un elemento básico para el desarrollo de la metodología de diseño de la base de datos.

Los diferentes modelos de datos comparten, aunque con diferentes nombres y notaciones, unos elementos comunes, componentes básicos de la representación de la realidad que realizan. Estos componentes se identifican gracias a la clasificación, y pueden identificarse conceptos estáticos y conceptos dinámicos. Los conceptos estáticos corresponden a:

1. Objeto: cualquier entidad con existencia independiente sobre el que almacenan datos. Puede ser simples o compuestos.
2. Relación: asociación entre objetos.

3. Restricción estática: propiedad estática del mundo real que no puede expresarse con los anteriores, ya que sólo se da en la base de datos; suele corresponder a valores u ocurrencias, y puede ser sobre atributos, entidades y relaciones.
4. Objeto compuesto: definidos como nuevos objetos dentro de la base de datos, tomando como punto de partida otros existentes, mediante mecanismos de agregación y asociación.
5. Generalización: se trata de relaciones de subclase entre objetos, es decir, parte de las características de diferentes entidades pueden resultar comunes entre ellas.

Por su parte, los conceptos dinámicos responden a:

1. Operación: acción básica sobre objetos o relaciones (crear, modificar, eliminar...).
2. Transacción: conjunto de operaciones que deben ejecutarse en su conjunto obligatoriamente.
3. Restricción dinámica: propiedades del mundo real que restringen la evolución en el tiempo de la base de datos.

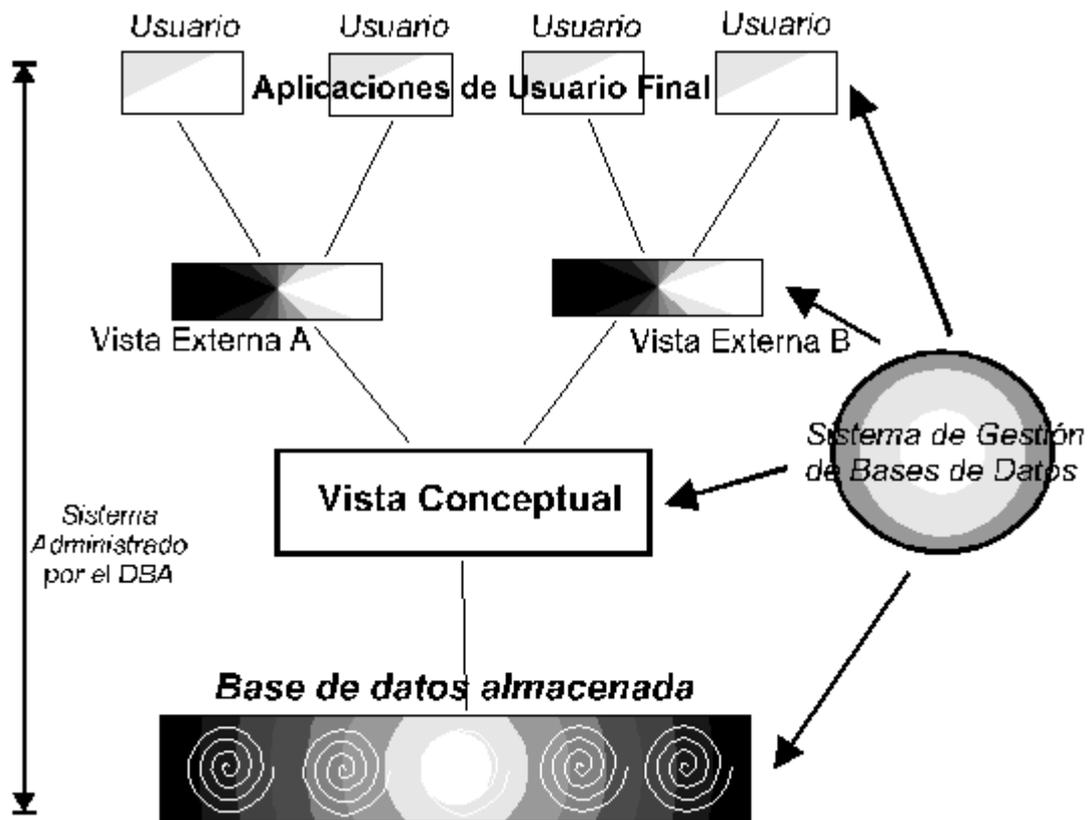
3.4- Los sistemas de gestión de bases de datos.

Para plasmar los tres niveles en el enfoque o modelo de datos seleccionado, es necesaria una aplicación que actúe de interfaz entre el usuario, los modelos y el sistema físico. Esta es la función que desempeñan los SGBD, ya reseñados, y que pueden definirse como un paquete generalizado de software, que se ejecuta en un sistema computacional anfitrión, centralizando los accesos a los datos y actuando de interfaz entre los datos físicos y el usuario. Las principales funciones que debe cumplir un SGBD se relacionan con la creación y mantenimiento de la base de datos, el control de accesos, la manipulación de datos de acuerdo con las necesidades del usuario, el cumplimiento de las normas de tratamiento de datos, evitar redundancias e inconsistencias y mantener la integridad. Se han señalado como componentes de un sistema ideal de gestión de bases de datos los siguientes (FROST, 1989: 90):

1. Un lenguaje de definición de esquema conceptual.
2. Un sistema de diccionario de datos.
3. Un lenguaje de especificación de paquetes de entrada/salida.
4. Un lenguaje de definición de esquemas de base de datos.
5. Una estructura simétrica de almacenamiento de datos.
6. Un módulo de transformación lógica a física.
7. Un subsistema de privacidad de propósito general.
8. Un subsistema de integridad de propósito general.
9. Un subsistema de reserva y recuperación de propósito general.
10. Un generador de programas de aplicación.
11. Un generador de programas de informes.
12. Un lenguaje de consulta de propósito general.

El SGBD incorpora como herramienta fundamental dos lenguajes, para la definición y la manipulación de los datos. El lenguaje de definición de datos (DDL, Data Definition Language) provee de los medios necesarios para definir los datos con precisión, especificando las distintas estructuras. Acorde con el modelo de arquitectura de tres niveles, habrá un lenguaje de definición de la estructura lógica global, otro para la definición de la estructura interna, y un tercero para la definición de las estructuras externas.

El lenguaje de manipulación de datos (DML, Data Manipulation/ Management Language), que es el encargado de facilitar a los usuarios el acceso y manipulación de los datos. Pueden diferenciarse en procedimentales (aquellos que requieren qué datos se necesitan y cómo obtenerlos) y no procedimentales (que datos se necesitan, sin especificar como obtenerlos), y se encargan de la recuperación de los datos almacenados, de la inserción y supresión de datos en la base de datos, y de la modificación de los existentes.



Arquitectura de un Sistema de Bases de Datos.

Establecidos los conceptos de bases de datos, su arquitectura y las características de las aplicaciones que soportan su gestión, es conveniente revisar los pasos o fases que sigue la ejecución de una tarea cualquiera por parte del sistema de gestión de bases de datos (MOTA, CELMA y CASAMAYOR, 1994: 13-14):

1. Petición de la aplicación del usuario.
2. Examen de la petición en el marco del esquema externo del usuario.
3. Transformación del esquema externo al lógico.
4. Transformación del esquema lógico al interno.
5. Interacción con el almacenamiento físico.
6. Envío de los datos a los buffers del SGBD.
7. Transformaciones de los datos entre el esquema lógico y el externo.
8. Transferencia de los datos necesarios al área de trabajo del usuario.

3.5- Los usuarios.

En consonancia con las posibles, y diferentes, vistas externas, se pueden identificar varios tipos de usuarios. En primer lugar, los usuarios finales, que hacen un uso limitado de las capacidades del sistema, normalmente referentes a introducción, manipulación y consulta de los datos. Los usuarios finales pueden ser sofisticados o especializados e ingenuos, dependiendo de su nivel de interacción con el sistema. En segundo lugar hay que citar a los programadores de base de datos, encargados de escribir aplicaciones limitadas, mediante el lenguaje de programación facilitado por el SGBD, normalmente

algún lenguaje de cuarta generación, que faciliten la ejecución de tareas por parte de los usuarios finales. Por último, el administrador de base de datos (DBA, Data Base Administrator) cumple las importantes funciones de crear y almacenar las estructuras de la base de datos, definir las estrategias de respaldo y recuperación, vincularse con los usuarios y responder a sus cambios de requerimientos, y definir los controles de autorización y los procedimientos de validación.

3.6- La creación de bases de datos.

Con los antecedentes señalados, se inicia la creación de las bases de datos. En primer lugar, y acorde con los diferentes niveles de arquitectura de bases de datos reseñados, tiene lugar la construcción del modelo y del esquema conceptual de la base de datos (REINGRUBER y GREGORY, 1994):

- ❖ El esquema conceptual.
El esquema conceptual puede definirse como una descripción abstracta y general de la parte o sector del universo real que el contenido de la base de datos va a representar, llamada en ocasiones "universo del discurso". En este nivel de análisis se está tratando con una descripción de la realidad, no con datos, y suele contener listas de tipos de entidades, de las relaciones existentes entre esas entidades y de las restricciones de integridad que se aplican sobre ellas. El esquema conceptual de la base de datos puede utilizarse para integrar los intereses de los diferentes usuarios, como herramienta de representación y de formación, así como para prever futuras modificaciones del sistema. En el aspecto de la representación, lo más interesante es utilizar algún tipo de especificación formal en sentido matemático, lo que facilita la consistencia y los análisis lógicos de los esquemas propuestos. Del esquema conceptual formalizado pueden derivarse diferentes subesquemas conceptuales, que representan aquellas partes del esquema conceptual de interés para un usuario o grupo de usuarios finales.
- ❖ El esquema de la base de datos.
Una vez construido el esquema conceptual, el diseño de bases de datos obliga a realizar varias tareas previas a la construcción del esquema lógico global del sistema, también llamado esquema de bases de datos. Por el momento, basta saber que el esquema de la base de datos representa la descripción de los datos de la base de datos, mientras que el esquema conceptual representaba a la realidad. La primera de las tareas necesarias es la identificación de los datos requeridos, para obtener como resultado las partes del área de aplicación que deben representarse mediante datos, y en que forma deben presentarse éstos a los usuarios. El siguiente paso es el análisis de datos, consistente en la definición y clasificación de esos datos, su descripción, que suele presentarse en forma de diccionario de datos, como una "metabase de datos". Por último, debe hacerse la especificación de los paquetes de entrada y de salida, correspondientes con los datos que deben introducir y obtener como respuesta los usuarios, según sus necesidades. Las tres tareas habrán permitido obtener tres documentos sobre descripción del área de aplicación, definición y clasificación de los datos y especificación de las características de los diversos paquetes, respectivamente. Tomando como punto de partida estos tres elementos, se construye la especificación de esquema de la base de datos, que responderá al contenido total de la base de datos y las características de las vías de acceso requeridas a través de estos datos. Frente al análisis de datos, que es la definición y clasificación de los datos, el esquema se encarga de la utilización de esos datos.
- ❖ El diccionario de recursos de información (MIGUEL y PIATTINI, 1995).
La gestión efectiva de los datos involucrados en la base de datos implica necesariamente disponer de alguna herramienta que controle las características y funciones de aquéllos. Esta función es cubierta mediante el diccionario de recursos de información (DRI), que asegura la integración de toda la información contenida en el sistema. Se habla entonces de metadatos, como datos que definen y describen los datos existentes en el sistema. En un primer momento, este tipo de cuestiones eran resueltas a través de los diccionarios de datos, que reunían información sobre los datos almacenados, sus descripciones, significados, restricciones, usos, etc., y los directorios de datos, subsistemas del sistema de gestión, encargados de describir dónde y cómo se almacenaban los datos, Actualmente se aplica el concepto de diccionario de recursos de información, que engloban todo lo señalado anteriormente, dando lugar a lo que ha pasado a llamarse "metabases".
- ❖ El enfoque de tratamiento de los datos.
La construcción de los modelos conceptual y lógico de las bases de datos requiere la adopción de un determinado enfoque para la descripción y el tratamiento de los datos. Sin embargo, es necesario insistir en que la modelización de datos se orienta al conocimiento en profundidad de

los datos que va a manejar la organización, para lograr una implantación óptima. La unión del modelo de datos con el sistema de gestión de base de datos dará como resultado la base de datos real. El modelo de datos será una representación gráfica orientada a la obtención de las estructuras de datos de forma metódica y sencilla, agrupando esos datos en entidades identificables e individualizables, y será reflejo del sistema de información en estudio.

3.7- Creación de una base de datos: enfoque E/R y transformación relacional.

3.7.1- El enfoque entidad-relación de Chen.

Por sus características, se ha seleccionado el enfoque entidad-relación propuesto por Chen (CHEN, 1976; MOTA, CELMA Y CASAMAYOR, 1994; KORTH y SILBERSCHATZ, 1993: 25-226; BATINI, CERI y NAVATHE, 1994). Este modelo toma como punto de partida considerar la existencia de entidades, que representan objetos, personas, etc, sobre las que se quiere almacenar información relevante. Las entidades con las mismas características forman un tipo de entidad. A las características necesarias para describir completamente a cada tipo de entidad se les denominará atributo. Posteriormente, las entidades y sus atributos se representan físicamente a través de tablas (transformación en un modelo relacional) en las que los datos se almacenan en dos dimensiones. Las filas de la tabla contienen los atributos de cada una de las entidades, y las columnas el conjunto de atributos del mismo tipo de cada entidad. El grado de la tabla corresponderá al número de columnas de la tabla. En este momento estaremos trasladando el modelo semántico entidad/relación al modelo clásico relacional, se decir, la transformación entre el modelo conceptual y el lógico. El principio fundamental en este modelado, que no puede obviarse de ninguna forma, es que hechos distintos deben almacenarse en objetos distintos.

Uno de los puntos fuertes de este modelo es que prevé que las entidades puedan mantener relaciones entre ellas. En primer lugar, es necesario definir la clave de la entidad. La claves serán el atributo, o conjunto de atributos, perteneciente al mismo tipo de entidad que hacen único el acceso a esa entidad u ocurrencia de la tabla, determinando de esta forma a una única entidad. La presencia de varios atributos que pueden funcionar como clave da lugar a la existencia de claves candidatas, y por otra parte se puede hablar de claves simples (formadas por un único atributo) y claves múltiples, compuestas o concatenadas (formadas por un conjunto de atributos. No hay que obviar tampoco el concepto de clave ajena, aquel atributo de una tabla que puede funcionar como clave en otra. La ocurrencia de entidad será, en este contexto, cada uno de los posibles valores reales que puede tomar la clave de una entidad.

Las relaciones entre tablas, basadas en la conexión de éstas a través de las claves, pueden ofrecer diferentes cardinalidades, entendiendo como tal el número de ocurrencias de una entidad que se relacionan con ocurrencias de la otra entidad. Pueden identificarse tres tipos: (1,1), donde una ocurrencia se relaciona con otra; (1,m), donde una ocurrencia puede relacionarse con varias; y (m,n), donde varias ocurrencias de una entidad pueden relacionarse con varias ocurrencias de la otra entidad. El modelo de Chen es n-ario, lo cual quiere decir que las relaciones pueden establecerse entre una, dos o más entidades. Las entidades pueden ser de dos tipos:

1. Entidad regular: aquella sobre la que se puede definir la clave primario dentro de sus propios atributos.
2. Entidad débil: aquellas que no puede utilizar sus propios atributos como clave, al estar asociada a otra entidad.

La definición del modelo conceptual con la técnica propuesta por Chen propone una secuencia de fases para la obtención del modelo:

1. Identificar las entidades dentro del sistema: para ello, debe conocerse el funcionamiento del sistema en estudio, a través de estudios de usuarios, de necesidades de información, de tipos de información, etc. Como guía puede utilizarse para la definición de las entidades objetos reales, personas, actividades del sistema, objetos abstractos, etc.
2. Determinar las claves o identificadores de entidades: señalar aquellos atributos que identifiquen inequívocamente cada ocurrencia de la entidad, y que no puedan ofrecer valores nulos.
3. Establecer las relaciones entre la entidades, describiendo el grado de las mismas: estudiar las asociaciones entre las entidades, para definir su importancia dentro del contexto del sistema, y obtener su cardinalidad.
4. Dibujar el modelo de datos: representar gráficamente el modelo obtenido.

5. Identificar y describir los atributos de cada entidad: señalar aquellas propiedades de la entidad de interés para el sistema.
6. Verificaciones: eliminación de las relaciones redundantes y que puedan ser obtenidas a través de combinar otras asociaciones.

El modelo obtenido se representa mediante una notación gráfica especializada, a través de diagramas, cuyas normas generales y variantes especializadas pueden encontrarse en la bibliografía pertinente.

3.7.2- La normalización.

El modelo conceptual de datos obtenido mediante la técnica de entidad-relación será refinado y convertido en un modelo lógico relacional, utilizando la normalización, lo que ofrecerá como resultado el conjunto de tablas a implementar en la base de datos (JACKSON, 1990; MIGUEL Y PIATTINI, 1993: 425-674). Su finalidad es reducir las inconsistencias y redundancias de los datos, facilitar el mantenimiento y evitar las anomalías en las manipulaciones de datos. El objetivo será obtener un modelo lógico normalizado que represente las entidades normalizadas y las interrelaciones existentes entre ellas. Para ello, se toma como punto teórico de partida el concepto de dependencia funcional, que dice: "Un atributo B depende funcionalmente de otro atributo A, de la misma entidad si a cada valor de A le corresponde sólo un valor de B." Lo anterior se completa mediante la dependencia funcional completa y la dependencia transitiva.

El procedimiento de normalización consiste en someter a las tablas que representan entidades a un análisis formal para ver si cumplen, o no, las restricciones necesarias que aseguren evitar los problemas citados con anterioridad. A mayor nivel de normalización, mayor calidad en la organización de los datos y menor peligro para la integridad de los datos. Este procedimiento consiste en ir alcanzando formas normales

Todo el proceso se basa en que una primera relación universal plantearía enormes problemas de redundancia, consistencia e integridad de los datos, por lo que es necesario mejorar las relaciones. Estas mejoras deben dar como resultado tablas equivalentes y mejores que sus respectivas originales, y poseer siempre tres propiedades: conservación de la información (de atributos y de tuplas), conservación de dependencias y mínima redundancia de los datos. Las mejoras introducidas obligan a plantear hasta que Forma Normal es necesario llegar, es decir, a que nivel de depuración. Normalmente, es recomendable alcanzar la máxima Forma Normal, aunque luego es muy probable que restricciones existentes, de algún tipo, obliguen a retroceder a un nivel inferior de normalización, o incluso a cierto nivel de "desnormalización".

3.8- Propuesta de un método estándar de diseño.

Con los métodos que se han expuesto, el diseño de una base de datos relacional puede seguir dos caminos. Por una parte, puede crearse tomando como punto de partida la observación del universo en estudio, dando lugar a un conjunto de esquemas de relaciones, que contengan los atributos y sus restricciones. Por otra parte, puede dividirse el diseño en dos fases, la primera de las cuales sería definir el modelo conceptual y su esquema, y la segunda transformar el esquema conceptual en un esquema relacional mediante una transformación realizada de acuerdo a unas reglas dadas.

Sin perjuicio del rigor en el diseño relacional, el diseño de una base de datos no puede limitarse a la aplicación exclusiva de la teoría de la normalización. Del mismo modo que se ha visto la existencia de variadas metodologías en el ámbito de los sistemas de información, se encuentra el mismo panorama en el diseño de bases de datos, aunque aquí tampoco aparece una metodología consagrada. De esta forma, Elmasri y Navathe comparan el ciclo de diseño de los sistemas de información y de las bases de datos, y definen el problema de diseñar una base de datos como:

"Desing the logical and physical structure of one or more databases to accommodate the information needs of the users in an organization for a defined set of applications"(ELMASRI y NAVATHE, 1989: 457)

y señalan la existencia de seis fases en el proceso de diseño de una base de datos:

- ❖ Fase 1: Recopilación y análisis de requerimientos.
En esta fase se trata de conocer las expectativas del usuario sobre la base de datos. Para ello, se identifican los grupos de usuarios reales y posibles y las áreas de aplicación, se revisa la documentación existente, se analiza el entorno operativo y los requerimientos de procesado, y se realizan entrevistas y cuestionarios con los usuarios. Para todo ello existen técnicas formalizadas de especificación de requerimientos.
- ❖ Fase 2: Diseño conceptual de la base de datos
Esta fase se subdivide en otras dos.
 - La Fase 2a corresponde al Diseño del esquema conceptual, esquema de especificación del modelo de datos a alto nivel, independiente de cualquier SGBD, que no puede utilizarse para implementar directamente la estructura de la base de datos. Para obtenerlo puede adoptarse un enfoque de esquema centralizado (en el cual se unen previamente los diferentes requerimientos a la realización del esquema), o un enfoque de integración de vistas (en el cual se unen los esquemas de cada requerimiento en uno global realizado a posteriori).
 - La Fase 2b corresponde al diseño de transacciones, es decir, a aquellas aplicaciones que van a manipular datos contenidos en la base de datos. Se suelen identificar mediante el estudio de las entradas y salidas de datos y su comportamiento funcional. De esta forma se identifican transacciones de recuperación, de actualización y mixtas.
- ❖ Fase 3: Elección de un SGBD.
Se consideran diferentes factores técnicos, económicos y de beneficio, de servicio técnico y formación de usuarios, organizativos de rendimiento, etc. Sin embargo, resulta difícil la medida y cuantificación ponderada de los diferentes factores.
- ❖ Fase 4: Transformación del modelo de datos (o fase de diseño lógico).
En esta fase se crea un esquema conceptual y los esquemas externos necesarios en el modelo de datos del SGBD seleccionado, mediante la transformación de los esquemas de modelo de datos a alto nivel obtenidos en la Fase 2a, al modelo de datos ofrecido por el SGBD.
- ❖ Fase 5: Diseño de la base de datos física.
Consiste en definir las estructuras de almacenamiento y de acceso para alcanzar una rendimiento óptimo de las aplicaciones de la base de datos. Los criterios adoptados suelen ser el tiempo de respuesta, la utilización de espacio y el volumen de transacciones por minuto. Fase 6: Implementación del sistema de base de datos. En esta fase final se hace realidad la base de datos, mediante la creación y la compilación del esquema de bases de datos y de los ficheros de bases de datos, así como de las transacciones, a través de las aplicaciones.

La metodología expuesta, que puede servir como marco de referencia general, puede modificarse según las características del contexto en el que se diseña e implanta el sistema de bases de datos.

En el dinámico entorno de la información almacenada en las bases de datos, las recientes tendencias, derivadas en muchas ocasiones de las propias necesidades, han obligado a completar e incorporar nuevos conceptos y enfoques en el tratamiento de los datos. Por ejemplo, la existencia de relaciones complejas en el mundo real han obligado a la incorporación del modelado semántico, lo que ha dado como resultado la evolución del modelo entidad-relación extendido, con sus conceptos de superclases y subclases, y los procesos de generalización y especialización, así como la importante noción de herencia. También es necesaria la referencia ineludible al paradigma de la orientación a objetos (BERTINO y MARTINO, 1995), enfoque de tratamiento de la información que cobra cada vez mayor auge en aplicaciones comerciales, y que se configura como la opción de mayor futuro en el desarrollo de aplicaciones. La identificación de los datos y sus procesos como objetos individuales, el encapsulamiento y las propiedades de herencia son las características principales del enfoque a objetos. Por último, no puede olvidarse la creciente tendencia entre el enfoque relacional y el modelo de objetos, así como la integración de información referenciada espacialmente en modelos relacionales.

Es innegable que la gestión y la explotación subsiguiente de los registros que contienen datos, y, como consecuencia, información, depende de las herramientas existentes en el campo de la gestión de la información, por una parte, y del cuerpo teórico de la ciencia de la información, por otra. La explotación satisfactoria de esta información, de la misma forma, demanda experiencia en dos áreas de conocimiento: en las técnicas de recuperación de información y en el estudio de las necesidades de los usuarios.

3.9- Sistema de bases de datos distribuida.

En un sistema de base de datos distribuida, los datos se almacenan en varios computadores. Los computadores de un sistema distribuido se comunican entre sí a través de diversos medios de comunicación, tales como cables de alta velocidad o líneas telefónicas. No comparten la memoria principal ni el reloj. Los procesadores de un sistema distribuido pueden variar en cuanto su tamaño y función. Pueden incluir microcomputadores pequeños, estaciones de trabajo y sistemas de computadores grandes de aplicación general. Estos procesadores reciben diferentes nombres, tales como localidades, nodos o computadores.

Un sistema distribuido de bases de datos consiste en un conjunto de localidades, cada uno de las cuales puede participar en la ejecución de transacciones que accedan a datos de una o varias localidades. La diferencia principal entre los sistemas de base de datos centralizados y distribuidos es que, en los primeros, los datos residen en una sola localidad, mientras que, en los últimos, se encuentran en varias localidades.

3.9.1- Estructura de Base de Datos Distribuidas

Un sistema distribuido de base de datos consiste en un conjunto de localidades, cada una de las cuales mantiene un sistema de base de datos local. Cada localidad puede procesar transacciones locales, o bien transacciones globales entre varias localidades, requiriendo para ello comunicación entre ellas. Las localidades pueden conectarse físicamente de diversas formas, las principales son:

- Red totalmente conectada ·
- Red prácticamente conectada ·
- Red con estructura de árbol ·
- Red de estrella ·
- Red de anillo ·

Las diferencias principales entre estas configuraciones son:

- Coste de instalación: El coste de conectar físicamente las localidades del sistema ·
- Coste de comunicación: El coste en tiempo y dinero que implica enviar un mensaje desde la localidad A a la B.
- Fiabilidad: La frecuencia con que falla una línea de comunicación o una localidad.
- Disponibilidad: La posibilidad de acceder a información a pesar de fallos en algunas localidades o líneas de comunicación.

Las localidades pueden estar dispersas, ya sea por un área geográfica extensa (a lo largo de un país), llamadas redes de larga distancia; o en un área reducida (en un mismo edificio), llamadas redes de área local. Para las primeras se utilizan en la comunicación líneas telefónicas, conexiones de microondas y canales de satélites; mientras que para las segundas se utiliza cables coaxiales de banda base o banda ancha y fibra óptica.

3.9.2- Consideraciones al distribuir la base de datos

Existen varias razones para construir sistemas distribuidos de bases de datos que incluyen compartir la información, fiabilidad y disponibilidad y agilizar el procesamiento de las consultas. Pero también tiene sus desventajas, como desarrollos de software más costosos, mayor posibilidad de errores y costos extras de procesamiento.

3.9.3 Ventajas de la distribución de datos.

La principal ventaja de los sistemas distribuidos es la capacidad de compartir y acceder a la información de una forma fiable y eficaz.

- Utilización compartida de los datos y distribución del control: La ventaja principal de compartir los datos por medio de la distribución es que cada localidad pueda controlar hasta cierto punto los datos almacenados localmente. En un sistema centralizado, el administrador de base de datos

de la localidad central controla la base de datos. En un sistema distribuido existe un administrador global de la base de datos que se encarga de todo el sistema. Parte de esta responsabilidad se delega al administrador de base de datos de cada localidad. Dependiendo del diseño del sistema distribuido, cada administrador local podrá tener un grado de autonomía diferente, que se conoce como autonomía local. La posibilidad de contar con autonomía local es en muchos casos una ventaja importante de las bases de datos distribuidas.

- **Fiabilidad y disponibilidad:** Si se produce un fallo en una localidad de un sistema distribuido, es posible que las demás localidades puedan seguir trabajando. En particular, si los datos se repiten en varias localidades, una transacción que requiere un dato específico puede encontrarlo en más de una localidad. Así, el fallo de una localidad no implica necesariamente la desactivación del sistema. El sistema debe detectar cuando falla una localidad y tomar las medidas necesarias para recuperarse del fallo. El sistema no debe seguir utilizando la localidad que falló. Por último, cuando se recupere o repare esta localidad, debe contarse con mecanismos para reintegrarla al sistema con el mínimo de complicaciones. La disponibilidad es fundamental para los sistemas de bases de datos que se utilizan en aplicaciones de tiempo real.
- **Agilización del procesamiento de consultas:** Si una consulta comprende datos de varias localidades, puede ser posible dividir la consulta en varias subconsultas que se ejecuten en paralelo en distintas localidades. Sin embargo, en un sistema distribuido no se comparte la memoria principal, así que no todas las estrategias de intersección se pueden aplicar en estos sistemas. En los casos en que hay repetición de los datos, el sistema puede pasar la consulta a las localidades más ligeras de carga.

3.9.4- Desventajas de la distribución de los datos

La desventaja principal de los sistemas distribuidos es la mayor complejidad que se requiere para garantizar una coordinación adecuada entre localidades.

El aumento de la complejidad se refleja en:

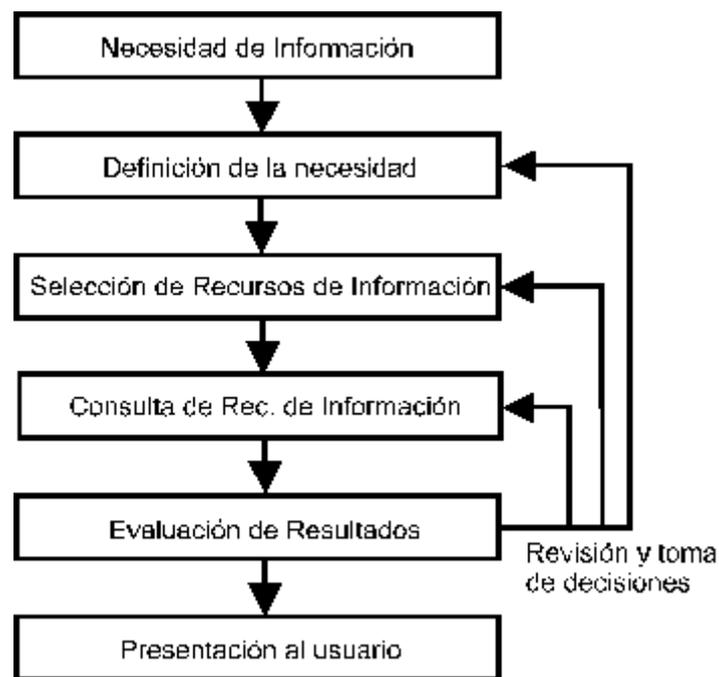
- **Coste del desarrollo de software:** es más difícil estructurar un sistema de bases de datos distribuidos y por tanto su coste es menor.
- **Mayor posibilidad de errores:** puesto que las localidades del sistema distribuido operan en paralelo, es más difícil garantizar que los algoritmos sean correctos.
- **Mayor tiempo extra de procesamiento:** el intercambio de mensajes y los cálculos adicionales son una forma de tiempo extra que no existe en los sistemas centralizados.

4.- RECUPERACIÓN DE INFORMACIÓN

4.1- Concepto de recuperación de información.

Cuando un usuario se plantea la necesidad de obtener nueva información sobre un asunto o materia de su interés, está manifestando una carencia, una situación irregular de sus estructuras mentales y cognitivas. Belkin, Oddy y Crofts han hablado de ASK, o "Anomalous State of Knowledge" (BELKIN, ODDY y CROFTS, 1982), un estado mental de incertidumbre que mueve al individuo a desarrollar una serie de acciones para salir de ese estado. Ingwersen deduce la existencia de un problema personal de espacio, "problem space" (INGWERSEN, 1992), por la diferencia entre el estado actual del conocimiento del usuario, y del estado que sería necesario para solucionar algún tipo de necesidad planteada. La respuesta a este tipo de situaciones es un conjunto de actividades que desarrolla el individuo para salir del estado anómalo, o para solucionar su problema de espacio, actividades que están íntimamente relacionadas con la adquisición de nueva información, y con el proceso comunicativo pertinente.

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. (CROFT, 1987) Todos estos factores ya han sido tratados exhaustivamente por la bibliografía especializada, por lo que no resulta necesario insistir sobre ellos en este lugar. Sin embargo, si es necesario recordar, como se ha señalado en la introducción, que la informatización de estas tareas se engloba dentro del campo de la Documentación Automatizada o Informática Documental.



Proceso genérico de recuperación de información

En principio, la recuperación de información engloba las acciones encaminadas a identificar, seleccionar y acceder a los recursos de información útiles al usuario, sin perjuicio de otras acepciones del concepto, en las que puede profundizarse utilizando la bibliografía correspondiente (ROBERTSON, 1977). Como puede deducirse del capítulo anterior, el objeto documental se ha organizado y representado, utilizando una serie de normas y convenciones, en un soporte informático, mediante el diseño, creación y mantenimiento de bases de datos (RIJSBERGEN, 1979). La siguiente fase lógica es la recuperación del contenido de las mismas, siguiendo unos criterios de pertinencia. En el ámbito de la documentación

automatizada, la recuperación de información adoptará la forma del acceso, selección y explotación de las bases de datos, utilizando diferentes técnicas.

El planteamiento de la recuperación de información (Information Retrieval, IR), en su moderno concepto y discusión, hay que buscarlo en la realización de los test de Cranfield (ELLIS, 1990: 1-17), y en la bibliografía generada desde ese momento y referida a los mecanismos más adecuados para extraer, de un conjunto de documentos, aquellos que fuesen pertinentes a una necesidad informativa dada. Las propias características de las entidades del mundo real, así como del tratamiento al que son sometidas, proveen a la representación de las mismas de un cierto grado de indefinición. Es decir, que el proceso documental, por muy alto nivel de perfección que pueda alcanzar, siempre introduce un factor de distorsión en la representación del documento. Si se considera que el acceso al documento se realiza casi por completo utilizando esta representación como intermediario, puede deducirse que los mecanismos en los que se basa la IR no son perfectos, sino que se verán influenciados por ese factor, independientemente de su validez técnica.

Ha sido Blair (BLAIR, 1990: 2-4) quien ha resumido las diferencias entre "data retrieval" (recuperación de datos, RD) e "information retrieval" (recuperación de información, RI), utilizando como criterios las siguientes cuestiones:

1. Según la forma de responder a la pregunta: en RD se utilizan preguntas altamente formalizadas, cuya respuesta es directamente la información deseada. En RI las preguntas resultan difíciles de trasladar a un lenguaje normalizado, y la respuesta es un conjunto de documentos que pueden contener, sólo probablemente, lo deseado, con un evidente factor de indeterminación.
2. Según la relación entre el requerimiento al sistema y la satisfacción de usuario: en RD la relación es determinística entre la pregunta y la satisfacción. En RI es probabilística, a causa del nivel de incertidumbre presente en la respuesta.
3. Según el criterio de éxito: en RD el criterio a emplear es la corrección y la exactitud, mientras que en RI el único criterio de valor es la satisfacción del usuario, basada en un criterio personal de utilidad.
4. Según la rapidez de respuesta: en RD depende del soporte físico y de la perfección del algoritmo de búsqueda y de los índices. En RI depende de las decisiones y acciones del usuario durante el proceso de interrogación.

Ha señalado este autor la importancia, en ocasiones ignorada, que tiene el factor de predicción. Predicción por parte del usuario, ya que éste debe intuir, en numerosas ocasiones, los términos que han sido utilizados para representar el contenido de los documentos, independientemente de la presencia de mecanismos de control terminológico. Este criterio de predicción es otro de los elementos que desempeñan un papel fundamental en el complejo proceso de la recuperación de información.

4.2- El problema de los lenguajes

Una de las cuestiones a considerar es el problema planteado por la traslación de los conceptos a diferentes tipos de lenguajes, cada uno de ellos con características propias (DESCHATELETS, 1986). En un primer momento, el usuario plantea sus necesidades utilizando el lenguaje humano, principalmente en sus variantes escrita y oral. El mensaje emitido es recibido, en numerosas ocasiones, por el documentalista, que debe asegurarse de reducir al mínimo las posibles diferencias entre lo expresado por el usuario y su propia comprensión, confrontando el significado real de los términos para el usuario final, y ofreciendo al mismo otras posibilidades: eliminación de polisemias, uso de sinonimias... Esta primera transmisión y depuración del mensaje coincide con la fase clásica de entrevista, en un proceso de búsqueda documental.

Una vez establecido claramente el contenido del primer mensaje, en un lenguaje al que podríamos llamar, con las debidas precauciones "clásico", los conceptos delineados y las relaciones entre ellos deben traducirse a un lenguaje documental. Esta es la segunda transformación que se realiza sobre los conceptos, y consiste en adecuar lo expresado por el usuario, conceptos y relaciones, a los términos y relaciones propios del lenguaje documental pertinente para cada caso. Entra en juego un "segundo lenguaje". Se trata, en este momento, de traducir los términos y expresiones utilizados por el usuario, a términos y relaciones entre los términos que estén contemplados en el lenguaje documental, independientemente de su tipo, que haya sido utilizado en la creación y representación de los documentos en la base de datos que se vaya a consultar. Resulta necesario utilizar los mecanismos de control

terminológico establecidos para cada recurso de información. Por supuesto, es posible utilizar directamente los términos y relaciones obtenidos del "primer lenguaje", pero puede imaginarse fácilmente que los resultados no serán los adecuados.

Sin embargo, no será ésta la última transformación "lingüística" a realizar. Una vez obtenidos y validados los términos que deban utilizarse, así como las relaciones entre ellos, las expresiones resultantes deben transformarse a un "tercer lenguaje", el cual cumple la función de interrogar, de una forma consistente y comprensible para el sistema informático y sus aplicaciones, la base de datos, con la finalidad de extraer de ésta aquellos documentos que cumplan los requisitos establecidos. A este tercer lenguaje se le denomina lenguaje de interrogación de bases de datos, y es el marco dentro del cual se deben introducir las expresiones del "segundo lenguaje". En este "tercer lenguaje" los términos y conceptos expresados en los anteriores serán válidos, pero será preciso utilizar las características propias del mismo en el momento de expresar las relaciones entre los conceptos.

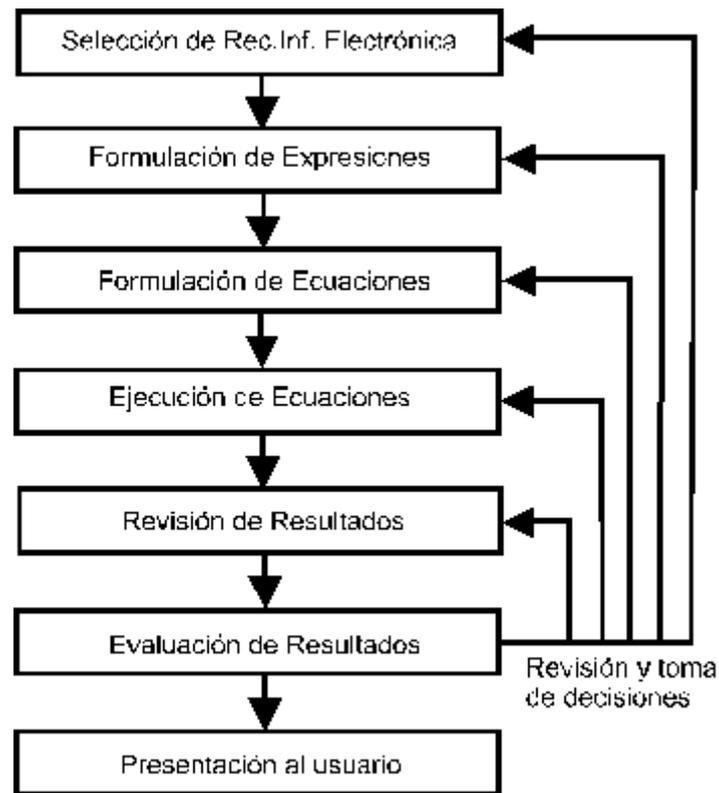
Los procesos de transferencia entre los tres lenguajes suponen uno de los principales problemas en la recuperación de información en bases de datos. Mientras los "terceros lenguajes" sólo ofrecen problemas de tipo técnico, superables mediante la práctica, los procesos que rodean los "primeros y segundos lenguajes" se encuentran con un mayor nivel de azar, lo que repercute en la recuperación de información. Para superar estos problemas se investiga en el desarrollo de nuevas técnicas que permitan superar, en lo posible, las diferencias de lógica entre los esquemas mentales de los usuarios, y los esquemas existentes en los lenguajes documentales y los lenguajes informáticos, más limitados y rigurosos. Se trataría, en este caso, de desarrollar intermediarios que facilitasen las transformaciones, de forma transparente para el usuario.

4.3- El proceso de recuperación

Como se ha señalado anteriormente, el proceso de recuperación de información engloba numerosas tareas, de las que la consulta de recursos de información electrónica resulta ser una más de ellas (SARACEVIC et alii, 1988). No se van a analizar en este lugar los procesos relacionados con la definición de áreas de interés del usuario, ni con la utilización de los lenguajes documentales en la recuperación, aunque se haga referencia a estas funciones a lo largo de la exposición. Un proceso de recuperación, al que podríamos considerar "genérico" (BELKIN y CROFT, 1987) 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 (que se expondrá en el capítulo correspondiente), 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.



Proceso de recuperación en un entorno informático.

Resulta necesario realizar una última aclaración, referida a la utilización del término "estrategia". La bibliografía suele utilizar esta palabra para hacer referencia a las ecuaciones utilizadas en un proceso de recuperación. Sin embargo, y desde una perspectiva tanto semántica como funcional, es más correcto utilizar el término estrategia para todo el proceso, o por lo menos para englobar las directrices generales definidas y utilizadas por el documentalista, usando el término "táctica" para las acciones más prácticas encaminadas a la obtención de los resultados. Desde este enfoque, la utilización de un conjunto de ecuaciones de consulta es más una táctica, siendo la estrategia la planificación de consulta de fuentes, los criterios de selección de las mismas, etc.

4.4- 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. La definición mínima de un lenguaje de interrogación y de sus componentes puede encontrarse en el borrador de norma ISO 8777-1988.

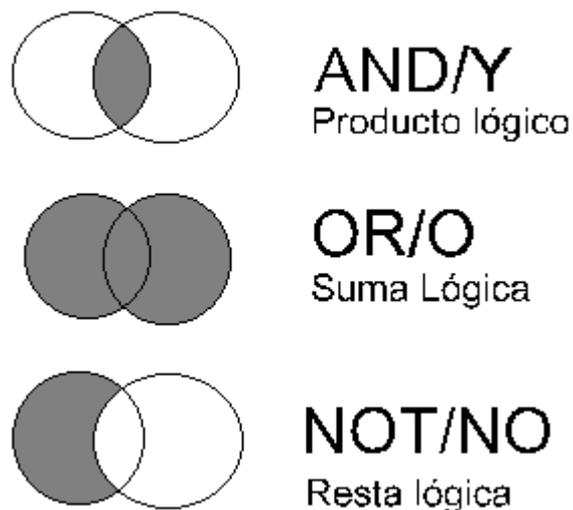
Las normas lógicas que rigen un lenguaje de interrogación responden a cuestiones relacionadas con la coordinación de los elementos, es decir, con la formulación de ecuaciones. Estas normas (a modo de sintaxis) especificarán el orden de los elementos, la disposición de las estructuras, sus posibilidades 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 (Common Command Language) 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 gráficas 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 (SALTON y MCGILL, 1983), que se analizan a continuación.

4.4.1- Operadores lógicos (o booleanos)

Los operadores lógicos, llamados booleanos en honor a George Boole, precursor de la lógica simbólica y del álgebra de conjuntos, 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 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 consideran 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.



Los tres operadores booleano básicos.

4.4.2- 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:

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.

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...).

4.4.3- 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.

4.4.4- 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.

4.5- 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.

4.5.1- 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 darse varios tipos principales de búsqueda, que pueden clasificarse en dos grandes grupos, sin perjuicio de que puedan darse situaciones en las que se combinen (MEADOW, 1992: 243-251; BATES, 1981; TURTLE y FLOOD, 1995):

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 ítems 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.

4.6- La exploración como mecanismo de recuperación

Las limitaciones inherentes al proceso de recuperación mediante ecuaciones han conducido a experimentar otras aproximaciones. Una de las más utilizadas es aquella que utiliza la exploración, es decir, el acceso a los documentos mediante técnicas de visualización de parte de su contenido que puede ser relevante, y la posterior asociación con otros documentos de perfil similar. Para Doyle (ELLIS, 1990: 22-23), esta capacidad de exploración debería ser fundamental en los sistemas. El usuario accede a un listado o enumeración de elementos descriptivos, y mediante un proceso de selección de elementos, va centrando el objetivo de su búsqueda. Los criterios utilizados por el usuario se basan en la deducción y la asociación de conceptos (aproximación ésta similar a la que utiliza un sistema hipertextual -ver infra el capítulo pertinente-) frente a la lógica de conjuntos que se plantea en un sistema de ecuaciones. Este tipo

de representación es más adecuada para reflejar la polirepresentación que un concepto puede tener para un usuario individual. En cambio, la utilización de la exploración suele realizarse en entornos en los cuales el usuario no posee una idea clara de cual debería ser la mejor táctica para aproximarse a la información que precisa. Por lo tanto, la cuestión clave a considerar en un sistema de exploración es combinar las ideas y esquemas del usuario con el esquema de organización de la información que ofrece el sistema (INGWERSEN, 1992: 135-140). Esta es la aproximación que pretenden desarrollar los enfoques cognitivos, poniendo su énfasis en el intermediario que debe existir entre el modelo del usuario y el modelo del sistema.

4.7- Revisión y análisis de resultados

El resultado de la ejecución de una ecuación de búsqueda es un conjunto de documentos que cumplen las condiciones expresadas en la ecuación. Se trata, a su vez, de un subconjunto del conjunto total de documentos existentes en el recurso o fuente de información consultado. Sin embargo, puede darse el caso de que la respuesta sea un número excesivamente elevado de documentos, o un número mínimo. Por otra parte, los documentos resultantes responden a la lógica y a las condiciones expresadas en la ecuación de búsqueda, lo cual no supone, como ya se ha señalado, que sean pertinentes a las necesidades del usuario. En realidad, es posible ejecutar ecuaciones perfectas, desde un punto de vista funcional (operadores, términos...), sin que los documentos resultantes reúnan las características que los harían deseables para el usuario.

Para superar esta posible distorsión en los resultados es necesario valorar y evaluar la respuesta a las ecuaciones planteadas. La primera modificación a realizar en la formulación de las ecuaciones afecta al número de respuestas obtenidas. En el caso de un excesivo número, se utilizan técnicas de restricción, mediante la introducción de términos más específicos, desechar términos generalistas, o limitar los truncamientos. En el caso de un número muy reducido, las acciones a tomar son las contrarias: utilizar términos más generales, incluyendo derivados y relacionados, limitar los operadores más restrictivos, añadir truncamientos, etc. Si se da la situación de ecuaciones correctas funcionalmente, pero sin respuesta adecuada, sería necesario replantear el proceso de recuperación, especialmente en la utilización de los lenguajes documentales y en la selección de fuentes.

Los resultados de una búsqueda se pueden valorar cuantitativamente utilizando dos parámetros, que son la exhaustividad y la precisión. Estos parámetros combinan el número de documentos pertinentes y no pertinentes, y recuperados y no recuperados, según las siguientes fórmulas (HEAPS, 1978: 28):

	Pertinentes	No pertinentes
Extraídos	a	b
No extraídos	c	d

La tasa de exhaustividad responde a la fórmula $a/(a+c)$, y debería situarse entre el 0,6 y el 0,8. La tasa de precisión responde a la fórmula $a/(a+b)$, y se sitúa entre 0,2 y 0,8. El principal problema para calcular la tasa de exhaustividades conocer la variable c, casi imposible de ajustar en situaciones normales de recuperación de información, por lo que se suelen utilizar técnicas estadísticas y de muestreo para obtener valores fiables.

4.8- Recuperación de información y sistemas expertos

Los sistemas expertos, con su capacidad para combinar información y reglas de actuación, han sido vistos como una de las posibles soluciones al tratamiento y recuperación de información, no sólo documental. La década de 1980 fue prolija en investigación y publicaciones sobre experimentos de este orden, interés que continua en la presente década. En resumen, un sistema experto es una aplicación capaz de realizar las tareas propias de un experto humano en un área restringida. Se compone de una base de datos, de una base de reglas y de un motor de inferencia (FROST, 1989). La base de datos almacena el conjunto de datos o documentos sobre los que se desea ejecutar una serie de acciones. La base de reglas contiene un compendio de reglas lógicas que el sistema debe utilizar para desarrollar razonamientos, así como las normas que permiten combinar las reglas, por último, el motor de inferencia es el encargado de ejecutar las órdenes del usuario, utilizando como criterios las reglas, y como material de partida el contenido de la base de datos, hasta alcanzar una conclusión simulando el razonamiento que seguiría el experto humano. El desarrollo posterior de estos sistemas ha traído un conjunto de nuevas aplicaciones a las que se identifica como Sistemas Basados en el Conocimiento (SBC), que incorporan técnicas más sofisticadas como la lógica difusa, razonamiento basado en modelos, etc., y lo que puede ser más interesante para el especialista en información, medios de recuperación de información deductiva (FROST, 1989: 6).

Lo que diferencia a estos sistemas de un sistema tradicional de recuperación de información es que estos últimos sólo son capaces de recuperar lo que existe explícitamente, mientras que un sistema experto debe ser capaz de generar información no explícita razonando con los elementos que se le dan (WORMELL, 1988). Pero la capacidad de los SE y de los SBC en el ámbito de la recuperación de la información no se limita a la recuperación. Pueden utilizarse en ayudas al usuario, en selección de recursos de información, en filtrado de respuestas... (ALBERICCO y MICCO, 1990) Un SE/SBC puede actuar como un intermediario inteligente que guía y apoya el trabajo del usuario final. Para desempeñar de forma adecuada esta tarea, los enfoques centrados en la creación de modelos de las estructuras cognitivas del usuario son los más prometedores (INGWERSEN, 1987).

5.- SISTEMAS INFORMATICOS DE TRATAMIENTO Y RECUPERACIÓN DE INFORMACIÓN DOCUMENTAL

5.1- La gestión informática de documentos

Son corrientes, en la actualidad, numerosas denominaciones como Sistemas de Gestión Documental, Sistemas de Archivo Electrónico, Sistemas de Recuperación de Información, Sistemas de Gestión de Bases de Datos Documentales... todas ellas se refieren a aplicaciones que, en mayor o menor medida, incorporan las funciones típicas de las Ciencias de la Documentación a un entorno informático (WILLET, 1988). Sin embargo, es corriente encontrar todo tipo de variaciones, tanto de estilo como de rango, entre la amplia panoplia de aplicaciones que se encuadran tras estos términos. Una breve discriminación terminológica de todos los términos señalados indica que el más adecuado para cubrir los contenidos de este capítulo, analizando las cuestiones planteadas por Reid y por Deogun y Raghavan (REID, 1990; DEOGUN y RAGHAVAN, 1988) es el concepto de "sistema de recuperación de información y de gestión de bases de datos documentales", ya que el resto resultan más generales, o todavía delimitados de forma somera, mientras que el objeto de estudio, en este caso, se va a centrar en las aplicaciones capaces de crear, mantener e interrogar una base de datos documental, algo que no todas las aplicaciones intituladas como se ha explicado son capaces de realizar.

Los sistemas de recuperación de información (Information Retrieval Systems) son aquellos que ofrecen al usuario mecanismos para acceder a fuentes de información en soporte informático, y recuperar y extraer de las mismas aquellos documentos cuyo contenido sea capaz de responder a una cuestión planteada por el usuario (SOERGEL, 1985). Se trata de terminología anglosajona, que en origen también se utilizaba para englobar sistemas capaces de crear y gestionar fuentes de información, aunque en la actualidad es preferible usar el término para aquellos sistemas que se limitan a la recuperación, como por ejemplo en la consulta de bases de datos en CD-ROM. Otro tipo de sistema que se está extendiendo progresivamente, sobre todo en entornos empresariales y de organismos públicos, es el identificado como sistema de gestión documental, cuya finalidad es establecer un control de toda la documentación recibida y generada por una organización, independientemente de recibir un tratamiento documental completo o no. Este último término resulta excesivamente amplio, en muchos casos, dadas las prestaciones ofrecidas por los sistemas.

Los sistemas de tratamiento y recuperación de información documental (STRID) manipulan documentos, en su más amplia acepción, frente a los sistemas de bases de datos (SGBD), que manipulan registros (TRAMULLAS y CUBILLO, 1995). Limitándose al contexto de la Documentación, podría decirse que los STRID 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 STRID, tan claro hace pocos años, está difuminando sus límites, al igual que sucede en otros ámbitos de la informática.

5.2- 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 (CODINA, 1990; CODINA Y ABADAL, 1992):

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 aquéllos, 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 (TRAMULLAS, 1996b), es una aplicación específica de los sistemas de gestión de 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 (véase infra el capítulo correspondiente).
 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 (PRAX, 1994; LASSOURY, 1994). 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.

5.3- Especificidad de los STRID

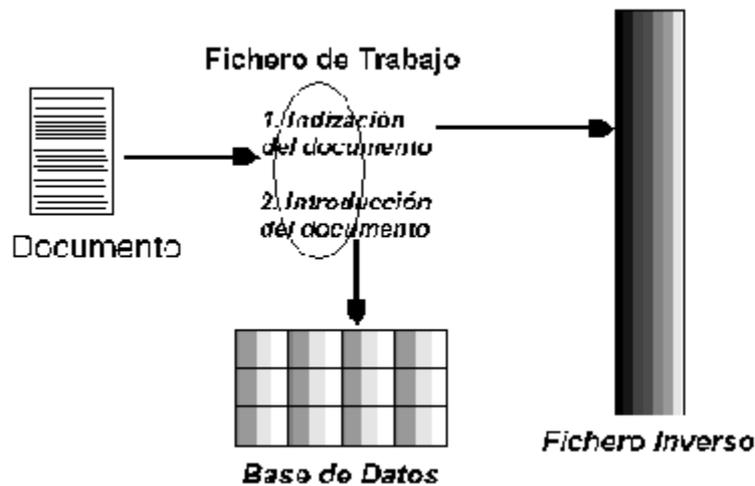
Evidentemente, algún tipo de características debe diferenciar a los sistemas de tratamiento y recuperación de información, como tales, de los sistemas de gestión de bases de datos, en el sentido más tradicional del término, vistos en el capítulo anterior. Como Codina y Abadal han señalado, en primer lugar los SGBD se centrarían en almacenar información correspondiente a procesos de gestión, fácilmente predecibles (aunque sería necesario matizar esta afirmación). Por oposición, los STRID se harían cargo de datos e informaciones poco estructuradas, con un fuerte componente textual, y de más difícil predicción, a causa de su contenido (MEADOW, 1992; ASHFORD y WILLET, 1988; SAFFADY, 1989). Las tres características que permiten identificar a un STRID 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.

5.3.1- La utilización del fichero inverso.

En primer lugar, la utilización de la técnica del fichero inverso (o fichero invertido) es un elemento clásico de los STRID. Dada la gran cantidad de información textual contenida en las bases de datos documentales, los procedimientos clásicos de búsqueda secuencial o de ficheros indexados no son capaces de responder de manera adecuada a los requerimientos de velocidad y exactitud en la respuesta necesarios para satisfacer al usuario. Por esta razón, los STRID utilizan una especialización de los ficheros indexados, creando un nuevo tipo de fichero, al que se conoce como "fichero inverso".

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 STRID 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.



Tratamiento del documento en un STRID

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".

5.3.2- 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, como se ha visto en un capítulo anterior. El STRID 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.

5.3.3- 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 STRID. Como es previsible, no todos los documentos poseen la misma estructura, ni ofrecen la misma longitud. En los SGBD 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 documentales sea capaz de aceptar, almacenar y manipular registros de forma flexible.

5.4- Tratamiento del texto

El factor crucial que determina la calidad y el éxito de un STRID 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 (MEADOW, 1992; SALTON, 1989; LANCASTER, 1986). 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.

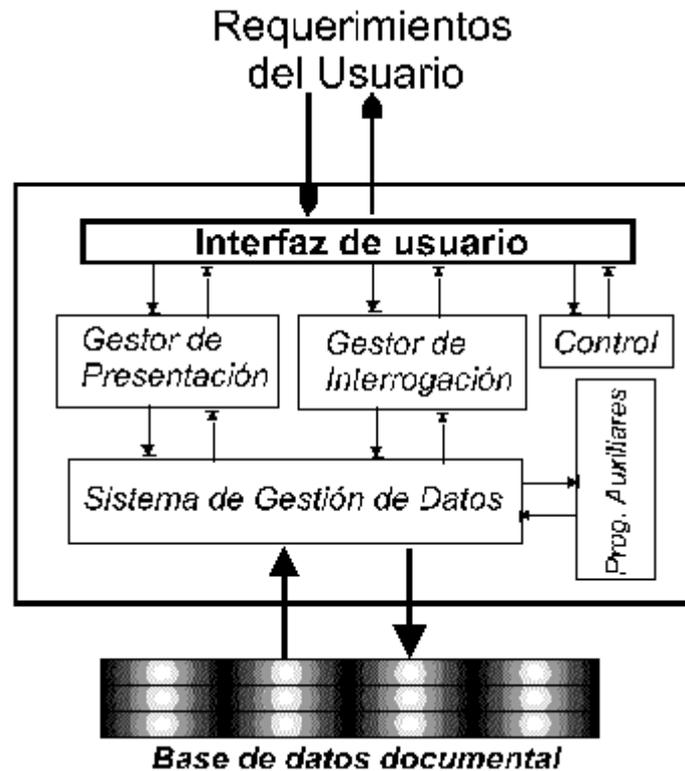
Las técnicas de asociación de palabras o términos intentan buscar y ofrecer al usuario términos relacionados con los que éste expresa en su consulta. Los más conocidos son los diccionarios y los tesauros y mini-tesauros. Los diccionarios ofrecen significados, sinónimos y antónimos del término deseado, 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 coocurrencia. 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.

5.5- Las funciones técnicas de un STRID

Los Sistemas de Tratamiento 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 (TRAMULLAS y CUBILLO, 1995; WEITZMAN y MATTHEW, 1995). 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. Ambos deben ser complementados por nuevas capacidades, más adecuadas a un tratamiento documental e informativo de la base de datos (CONLON y CONLON, 1996). En esta situación, un STRID 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. Todos los sistemas deben respetar la arquitectura de bases de datos explicada en el capítulo correspondiente.
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 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 coocurrencias.
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.



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 STRID (LANCASTER y WARNER, 1993: 15-16):

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.

5.6- Consultas al STRID: 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, como se ha visto en el capítulo correspondiente. 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 (TAGUE SUTCLIFFE, 1996).

5.6.1- 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 (véase el panorama esbozado con anterioridad sobre la recuperación de información), 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. Este segundo tipo de elementos son el objeto de los siguientes párrafos.

5.6.2- 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, en honor de George Boole, que en 1847 publicó un trabajo titulado *The mathematical Analysis of Logic*, en el que sentó las bases de la denominada álgebra lógica o simbólica. 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.

5.6.3- 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 (INGWERSEN, 1992), 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.

5.6.4- 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.

6.- HIPERTEXTO E HIPERMEDIA

6.1- Orígenes del hipertexto

Los términos hipertexto e hipermedia son ahora un lugar común al hablar de las nuevas tecnologías de presentación y acceso a la información, y se están revelando como una herramienta imprescindible en lo respecta tanto a la representación de documentos e información, como herramienta para procesos comunicativos y formativos en todos los ámbitos (CANALS, 1990). Sin embargo, y a pesar de su pretendida novedad, este enfoque de organización y acceso a la información tiene antecedentes desde la década de 1940 (GARCÍA MARCO, 1996). En tan temprana fecha, Vannevar Bush ideó un sistema de control, gestión y acceso a la documentación al que llamó MEMEX, y cuya idea principal era utilizar un principio de asociación de conceptos entre recursos informativos, de tal forma que el usuario pudiese acceder a estos recursos, independientemente de su tipo, simplemente utilizando la asociación de ideas. El sistema no llegó nunca a ser construido, por la dificultad técnica que suponía. Hubo que esperar a la década de 1960 para que Douglas Engelbart, un investigador del Stanford Research Institute, dirigió un proyecto de investigación para desarrollar máquinas, basadas en sistemas informáticos, que permitiesen aumentar la capacidad intelectual humana, y gracias al cual se introdujeron conceptos técnicos sin los cuales no sería posible el hipertexto: sesiones interactivas, dispositivos señaladores, ventanas y escritorios... que permitían gestionar gran cantidad de información según criterios jerárquicos y asociativos. El final del proyecto supuso que gran parte del equipo de trabajo pasó a Xerox, donde desarrollaron los fundamentos de los sistemas operativos de interfaz gráfico tan extendidos en la actualidad (MacOS, Windows, X-Window...).

En la misma década que Engelbart, Theodor Nelson acuñó el término hipertexto, al que definió como "escritura no secuencial". Retomando el concepto de asociación de Bush, por otra parte completamente aceptado por los estudios sobre el comportamiento del pensamiento humano, Nelson trabajaba en un proyecto para introducir, organizar y recuperar toda la información del mundo en un sistema informático, utilizando el principio de la asociación de ideas y documentos, al que llamó Xanadu. En este sistema los documentos se relacionaban utilizando enlaces a fragmentos y citas comunes, en un entorno de red informática en el que podían participar numerosos ordenadores. El proyecto Xanadu, comenzado en la Universidad de Harvard, en la que no cuajó, pasó a ser desarrollado en Australia, y recientemente Nelson ha pasado a trabajar en varios proyectos de investigación en Japón, siempre en el ámbito del diseño y desarrollo de sistemas hipertextuales.

Como consecuencia directa de los proyectos enumerados, 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. Las nuevas generaciones de STRID, así como los últimos desarrollos de SGBD incorporan mecanismos especializados para organizar la información utilizando, si el usuario lo desea, de forma hipertextual, incorporando mecanismos de acceso a diferentes tipos de información (gráfica, sonora...).

6.2- Conceptos de hipertexto e hipermedia

6.2.1- Hipertexto.

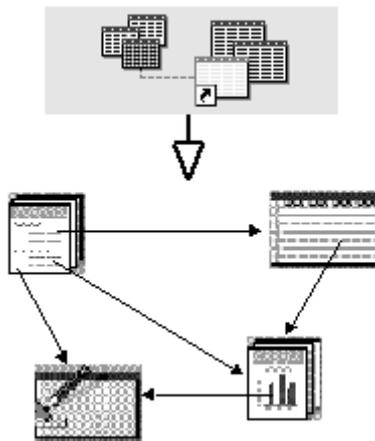
Las ideas expuestas en el bloque anterior deben servir como una primera aproximación al concepto de hipertexto. Como señaló Nelson, el hipertexto se basa en la escritura, y en la lectura, no secuencial de los documentos (NELSON, 1990: 1-3). Esto supone la existencia, y la adopción, del criterio de asociación de ideas y conceptos como principio organizador del conjunto de documentos, del conjunto informativo (LANDOW, 1995). 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, es establecen varios caminos posibles para acceder a la misma información.

6.2.2- Hipermedia.

El panorama esbozado es perfectamente comprensible desde una perspectiva textual, en la que los nodos corresponden a documentos con información escrita. Sin embargo, la información y la documentación actual incorporan gran cantidad de componentes gráficos (estáticos o dinámicos), sonidos u otros complementos (MOSCOSO y CARIDAD, 1991). La evolución de la tecnología informática ha hecho posible que estos documentos no textuales sean incorporados en formato digital a los sistemas informáticos. Como consecuencia, los sistemas hipertextuales han evolucionado hasta disponer de las capacidades necesarias para incluir en los sistemas hipertextuales documentos e información con presentación gráfica y/o sonora. Esta inclusión, o expansión de las funciones de gestión de información a cualquier forma que pueda presentar la misma, manteniendo como criterio la asociación o relación de conceptos, es lo que caracteriza a un sistema hipermedia.

La creación de estructuras informacionales que integran hipertexto e hipermedia dan lugar a un nuevo concepto de documento electrónico dinámico, cuyas características ha resaltado Martin (MARTIN, 1990: xii):

"The term hyperdocument refers to an electronic document which combines hypertext with diagrams, possibly sound, animation, or video, and possibly other types of software... with hyperdocument software we can create "intelligent documents" which guide their users, have built-in artificial intelligence or expertise, and can adapt themselves to user's needs. Many separate documents can be electronically interlinked."



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

6.3- Componentes de un sistema de hipertexto/hipermedia

Para que sea posible una existencia real de los conceptos de hipertexto e hipermedia, deben utilizarse aplicaciones que sean capaces de crear los vínculos y asociaciones entre los documentos. Las aplicaciones ofrecen unos elementos particulares que facilitan la creación y navegación por las estructuras hipertextuales (CONKLIN, 1987; NIELSEN, 1990: 101-114):

1. Un conjunto de ficheros que contienen los documentos relacionados.
2. Ventanas de presentación de los documentos, las cuales son modificables en tamaño y posición.
3. Dispositivos señaladores, que facilitan la selección y el acceso a los documentos mostrados en las ventanas.
4. Punteros o enlaces, que generalmente utilizan una representación gráfica distinta a la del resto del material informativo, en forma de color, iconos, botones...
5. Herramientas de creación de enlaces y anotación de la navegación, lo que da al usuario la posibilidad de crear sus propias asociaciones y documentos.

Estas funcionalidades se integran en una herramienta que en el entorno hipertextual es conocida como "browser", navegador o visualizador. El visualizador actúa como una interfaz, que muestra al usuario el contenido informativo de los documentos que selecciona, mediante la selección de enlaces. Suele completarse con la posibilidad de ejecutar búsquedas en el texto completo que contienen los documentos., y/o búsquedas más rígidas utilizando lenguajes clásicos de interrogación. La interrogación, sea de texto, imágenes o sonidos, suele realizarse a través de la ejecución de patrones, que representan una necesidad dada de información por parte del usuario. Además, una completa aplicación para este ámbito debería ser capaz de generar mapas gráficos de la estructura hipertextual, y utilizar estas representaciones para acceder directamente a los documentos deseados.

La visión que obtiene el usuario mediante el visualizador es una visión transparente, integrada, en la que no resulta complicado navegar de un documento a otro. Esta aparente facilidad no debe ocultar que los documentos pueden encontrarse en diferentes ficheros informáticos, e incluso en diferentes ordenadores, formando lo que se llama repositorio de información, que será tratado con más detalle en el próximo capítulo, por su relación con las bases de datos multimedia.

Los sistemas y estructuras de hipermedia pueden además incorporar inteligencia embebida, es decir, ser capaces de ejecutar otras aplicaciones o de tomar decisiones, acordes con la actividad desarrollada por el usuario, tanto en la utilización de los enlaces como en el acceso a los contenedores.

6.4- Tipos de sistemas de hiperdocumentos

Las aplicaciones de los sistemas hipermedia pueden ser múltiples, y sólo se encuentran limitadas por la capacidad de la aplicación o herramienta elegida, y por los criterios fijados por los creadores y usuarios. Se han establecido algunas grandes categorías que engloban las posibles clases de sistemas (MARTIN, 1990):

1. Macroliterarios: trabajos y artículos relacionados con bibliografías, críticas, comentarios...
2. Colecciones de documentos: información organizada sobre un tema, de gran tamaño (servicios de información industrial y técnica, servicios bibliotecarios...)
3. Herramientas de exploración de problemas: organizar material e ideas desconectadas para estructurar una idea general.
4. Conjunto interrelacionados de información: informaciones diversas que son útiles si se asocian (modelos complejos, información policial...).
5. Documentos largos y complejos: manuales, regulaciones... (productos, metodologías, referencias legales, guías...).
6. Documentos inteligentes: combinado con inteligencia artificial mediante sistemas expertos (Diagnósticos, autoenseñanza, ayuda sensible al contexto...)

La anterior división no es obstáculo para poner en relación, en un sistema informático, una aplicación de hipertexto con otras aplicaciones, de cuya interacción puedan beneficiarse ambas, como pueda ser un sistema de gestión de bases de datos, aplicaciones ofimáticas, sistemas de apoyo a la decisión, ayudas en línea o sistemas expertos.

6.5- Navegación y exploración de un sistema hipertextual

La utilización de un sistema hipertextual por parte de un usuario se basa en la navegación o exploración de los contenidos informativos del mismo. (MCKNIGHT, DILLON y RICHARDSON, 1991: 75-86). Exploración o navegación que ofrece diferencias respecto a la lectura tradicional de un documento, ya que el documento hipertexto posee características distintivas, como la duración temporal, la situación espacial, la variabilidad de la presentación o la reutilización del documento o de sus partes en la generación de nuevos documentos (LYNCH, 1994).

La exploración de un espacio de información como el que configura un sistema hipertexto/hipermedia puede provocar problemas en el usuario, como el citado desbordamiento cognitivo. El propio mecanismo de exploración/navegación se basa en el paso entre conceptos asociados, transición establecida y controlada por el usuario final (COVICE, 1994). Por lo tanto, los sistemas deben ofrecer mecanismo de apoyo a la navegación, en dos áreas:

1. Revisión histórica del proceso de exploración.
2. Situación del usuario dentro del espacio informativo.

Los sistemas deben ofrecer al usuario mecanismos para controlar en todo momento el proceso seguido, desde el comienzo, hasta llegar al documento actual. De esta forma se puede limitar el impacto de la pérdida del objetivo inicial. Una segunda acción es la posibilidad de utilizar mapas de la estructura del sistema de hipertexto, que permitan al usuario conocer su situación en todo momento. De esta forma se complementa el acceso por exploración con el posible acceso directo al documento o nodo deseado. Estos mecanismos de apoyo pueden implementarse directamente, o bien utilizar metáforas (bi o tridimensionales), representativas de las acciones y decisiones del usuario (NELSON, 1990: 127-137).

6.6- La recuperación de información en un sistema de hipertexto o hipermedios

El acceso al documento, en el contexto de los sistemas de recuperación de información y de bases de datos documentales, se realiza, generalmente, mediante la composición y ejecución de expresiones (ecuaciones) de búsqueda. Sin embargo, y como se ha señalado con anterioridad, en principio organizador de las estructuras hipertextuales es la asociación. Esto implica que el mecanismo de acceso y selección de la información relevante es diferente al "tradicional", y que los procesos mentales desarrollados por el usuario serán diferentes en varios aspectos.

En el acceso y la recuperación de información en un sistema de hipertexto o hipermedios se han diferenciado seis grandes categorías de actividad (ELLIS, 1990):

1. Inicio: fase inicial de búsqueda de información.
2. Encadenado: proceso de navegación siguiendo las secuencias entre documentos o conceptos pertinentes, utilizando los enlaces existentes en los documentos.
3. Búsqueda por visualización: también llamada búsqueda semi-directa. Consiste en consultar listados de referencias o enlaces que el sistema debe ofrecer al usuario, previo al acceso al documento.
4. Diferenciación: empleo de diferencias entre varias fuentes, y sus tipos, como criterio de filtro y calidad de los documentos.
5. Monitorización: control del desarrollo de un área de conocimiento utilizando como referencia fuentes seleccionadas.
6. Extracción: uso sistemático de una fuente como herramienta de identificación de información pertinente.

Las categorizaciones establecidas demandan del sistema que soporta la estructura hipertextual avanzadas funcionalidades: recuperación en texto completo, compilación de listados de referencia, mecanismos de creación de mapas de navegación... De ello se puede deducir que la capacidad de recuperación de información en un sistema de estas características está íntimamente ligado con las herramientas de soporte que ofrece, no sólo de organización y de visualización.

6.7- La creación y organización de hiperdocumentos

Los sistemas de hipertexto e hipermedia ofrecen al usuario la posibilidad de navegar, utilizando criterios de asociación, entre gran cantidad de documentos. En numerosas ocasiones, y dependiendo del tipo de sistema, esas asociaciones han sido creadas por terceros, según un conjunto de criterios lógicos y comprensibles, o al menos así lo han intentado. Sin embargo, la bibliografía ha recogido la existencia de problemas cuando un lector se enfrenta a una estructura de información que utiliza el hipertexto como principio organizativo (RODRÍGUEZ DE LAS HERAS, 1991).

Entre esos problemas deben citarse la posible desorientación del usuario, causada por el desconocimiento de la estructura de la información y de los medios existentes para navegar por ésta; el desbordamiento cognitivo, provocado por la dificultad en mantener el mismo nivel de concentración en el proceso de

navegación que en la comprensión de la información a la que se accede; por último, los posibles errores en la interpretación de la relación implícita en los enlaces y nexos que unen los documentos.

Para crear hiperdocumentos es preciso respetar ciertas normas que eviten el amenazador “hipercaos”. El hiperdocumento se divide en unidades de información, cada una de las cuales se encuentra en un envoltorio o contenedor, organizándose mediante un conjunto de criterios que pueden combinar la jerarquía con la asociación, de forma tal que la organización de las unidades sea claramente visible para el usuario. El factor de éxito en la construcción de hiperdocumentos es la división en nodos y los criterios pertinentes adoptados. Sin embargo, no debe olvidarse en ningún momento que el objetivo final es la comprensión activa y provechosa por parte del usuario, en un proceso comunicativo diferente a la lectura secuencial tradicional. Esto exige claridad de estructura, de organización de ideas, de lenguaje, de gráficos e ilustraciones y de mecanismos de navegación, de tal manera que el hiperdocumento resulte ser mucho más visual que el libro tradicional. Todas estas cuestiones hacen inevitable un riguroso estudio del usuario: quién es, porqué lee, qué sabe, que comprende, qué le resulta de valor...

Se han definido varias aproximaciones para la construcción de hiperdocumentos (SEYER, 1991: 93-117; DÍAZ, CATENAZZI y AEDO, 1996: 146-163). De esta forma, el diseñador puede optar por utilizar aproximaciones basadas en tabla de contenidos, jerárquico o de índices (en todos ellos se seleccionan opciones tomadas de menús que se ofrecen al usuario), mediante hipermapas (gráficos que representan la estructura del conocimiento almacenado en el hiperdocumento), visitas guiadas (especialmente para usuarios no expertos), tablas de decisión (según las opciones seleccionadas por el usuario se accede o ejecutan diferentes procesos), o plenamente empírica (basada en la experiencia del usuario y del diseñador, y en la observación del comportamiento del usuario). Por supuesto, estas aproximaciones no son excluyentes, si no que pueden combinarse.

Para terminar, cabe realizar una última, pero no por ello menos importante, consideración. Por sus propias características, un verdadero sistema hipertextual no debería ser un sistema cerrado, sino un sistema abierto y dinámico. Esto supone la necesidad de establecer unas normas de actualización y mantenimiento de la información introducida, de tal forma que se asegure la consistencia del conjunto, independientemente de su variabilidad espacial y temporal.

7.- BASES DE DATOS MULTIMEDIA

7.1- Multimedia y nuevos documentos.

El término multimedia ofrece dos problemas. En primer lugar, el entorno, mercado o ámbito que delimitan a multimedia todavía no ha sido bien definido, y sus límites resultan difusos. En segundo lugar, multimedia es un término que puede significar diferentes cosas, según quién lo utilice. Originalmente, el término multimedia, a finales de la década de 1970, resultaba ser la integración de voz, texto, datos y gráficos, a los que habría que añadir, actualmente, los gráficos interactivos, la imagen en movimiento, la composición de documentos electrónicos y la realidad virtual (KOEGL, 1994).

La aparición de aplicaciones multimedia sólo ha sido posible gracias al desarrollo de entornos gráficos de usuario para el aprovechamiento de los computadores. La aparición de estos entornos gráficos (Macintosh, Windows, X-Window, NextStep...), basados en la presentación de información en paneles o ventanas, posicionables en la pantalla, y en la utilización de representaciones pictóricas (iconos) como indicadores de elementos, así como de menús de opciones rápidas seleccionables por el usuario, ha modificado en profundidad la manera de presentar y de interactuar sobre los elementos informativos.

Uno de los problemas a resolver ha correspondido al almacenamiento y acceso a la información de tipo gráfico. A lo largo de los últimos años han proliferado gran cantidad de formatos de almacenamiento para gráficos (PICT, TIFF, gif, JPEG y otros), que han hecho necesario la creación de filtros y programas de transferencia entre unos y otros. Afortunadamente, se comienzan a establecer ciertos niveles de compatibilidad, con vista a su utilización en bases de datos, a través de metaficheros gráficos, que faciliten el acceso a todos los formatos, independientemente de cual se trate y de las aplicaciones que los manipulen.

Por último, el auge de multimedia debe gran parte de su éxito a la popularización de los medios ópticos y magneto-ópticos de almacenamiento masivo. La gran cantidad de espacio necesaria para el almacenamiento de ficheros de imagen de calidad adecuada, y el acceso y manipulación de éstos, dentro de un tiempo de respuesta prudencial, ha recibido una respuesta adecuada con el desarrollo de las últimas generaciones de CD-ROM y de lectores de alta velocidad. Tampoco puede pasarse por alto que la gran cantidad de espacio de almacenamiento requerida por imágenes y grandes volúmenes de información textual demanda, a pesar de los medios ópticos de almacenamiento, la utilización de técnicas de compresión y descompresión de ficheros de alta velocidad, que en su mayor parte se realizan mediante software específico, de forma que se combine un aceptable nivel de ocupación de recursos con una garantía de calidad elevada, en lo que respecta al contenido.

Los párrafos anteriores cubren cuestiones eminentemente relacionadas con los soportes la información. Sin embargo, un sistema multimedia debe contemplarse desde dos niveles. En primer lugar un nivel intencional, superior, en el que se recogen los conceptos del conocimiento humano que son creados o representados. En segundo lugar, un nivel extensional, inferior, que engloba la organización de los datos. Los conceptos de procesamiento de la información multimedia se distribuyen en ambos niveles. Los nuevos desarrollos en las bases de datos multimedia extienden las concepciones de los sistemas de información y de sus usuarios más allá de la simple integración. Técnicas de inteligencia artificial y sistemas expertos, mecanismos de representación y recuperación de información gráfica, desarrollo de nuevas interfaces de usuario, integración general de hipertexto e hipermedia, aplicaciones en el campo de la educación y la formación, etc., (APERS, BLANKEN y HOUTSMA, 1997): Los STRID que se impondrán en el mercado de pocos años incorporarán al tratamiento documental tradicional todas estas herramientas.

7.1.1- Aspectos económicos de la industria multimedia: el producto

La importancia económica de esta tecnología dentro de la Sociedad de la Información, parece indiscutible. Se trata de un sector en plena expansión y que, según todos los indicios, acaparará una importante cuota de mercado. Pero conviene, ante la euforia general, ser cautelosos a la hora de hacer afirmaciones definitivas sobre su trascendencia o futuros desarrollos.

Hay que decir en primer lugar, que se trata de un mercado en constitución/formación, cuyos actores o elementos participantes son múltiples: productores de hardware y software, operadores de

telecomunicación, editores y medios de comunicación, fabricantes de electrónica de consumo y las propias administraciones públicas y gobiernos, etc. lo que implica una complejidad añadida.

Por otro lado, se están produciendo a nivel mundial una serie de alianzas entre las principales empresas e industrias de los sectores arriba mencionados, y que suponen la aparición de grupos financieros muy poderosos. Dicho con otras palabras, se está produciendo una globalización del mercado y un cambio o incremento de las actividades productivas de los actores, que amplían sus actividades a campos no habituales en su actividad empresarial principal.

Los operadores de telecomunicación invierten en los medios de comunicación, o los productores de contenidos multimedia en la producción de equipamientos, por ejemplo. Hasta que el mercado multimedia no se asiente, resulta muy difícil hacer predicciones fiables o extraer valoraciones definitivas.

Es difícil encontrar cifras y datos fiables en cuanto a la presencia, características y contenidos de las bases de datos multimedia en el conjunto de la Industria de la Información. Esto es debido principalmente a dos factores:

- La corta trayectoria comercial de este tipo de productos. Recordemos que las bases de datos multimedia públicas no aparecen en el mercado hasta mediados de los años ochenta.
- A pesar de la popularización del término multimedia, éste no está reflejado como una categoría individual en los directorios y estudios estadísticos sobre bases de datos.

En una concepción estricta, un contenido multimedia sería la combinación de dos o más medios; en este sentido, las bases de datos orientadas a texto, en especial las de texto completo que incluyen ilustraciones, fotos y gráficos, pudieran ser consideradas multimedia y, por otro lado, bases de datos de imagen fija (archivos fotográficos) no presuponen más que la presencia de un tipo de información, sin poder clasificarse como multimedia.

Si se acude al Gale Directory of Databases 1997, principal fuente de información para obtener una perspectiva internacional, vemos que clasifica las BD según la representación de los datos, en: orientadas a texto, orientadas a imagen (fija y vídeo) y orientadas a sonido (audio).

Los datos cuantitativos más interesantes son los siguientes : En 1996, el total de BD de imagen fija y vídeo es de 846 y representa un 9% de total del mercado. Destaca, también, en esta categoría que un crecimiento notable ya que ha pasado de 14 BD (menos de 1%) , en 1988, a 846 (9%), en 1996. El número de BD se ha incrementado con un factor 11 desde 1990 y se multiplicado por 70 desde 1988. En cuanto a las bases de datos de sonido: pasan de 1 BD (en 1988) a 304 (en 1996) representando un 3% del mercado. Si sumamos ambos porcentajes, podemos decir que las bases de datos multimedia representan aproximadamente de un 12% del mercado actual, aunque esta cifra se antoja excesivamente baja para la oferta actual. El estudio destaca la alta velocidad de crecimiento del sector y una previsible mayor presencia cuando las condiciones técnicas sean óptimas y se hayan superado las limitaciones actuales.

7.1.2- Usuario multimedia

El mercado multimedia tiene como primera necesidad u objetivo satisfacer al usuario (y así lo reconocen el Informe Bangemann y la mayoría de los trabajos consultados), de tal manera, que esta industria es impensable sin la aceptación de los usuarios. Es necesario, dicen, crear una “masa crítica” de usuarios, que sea suficiente para rentabilizar las enormes inversiones en infraestructura de comunicaciones y desarrollo de nuevas tecnologías. El usuario o cliente objetivo de este mercado puede tipificarse en tres categorías:

- Las PYME
- La Administración Pública
- El usuario privado.

Cada uno de estos usuarios debe ser satisfecho en sus necesidades de información y servicios, y esto determinará el desarrollo del producto multimedia más que las posibilidades tecnológicas disponibles en cada momento.

El usuario PYME es el que menos valora el multimedia, sus usos principales son las aplicaciones transaccionales (telebanco) y comunicación interna (entre departamentos y sucursales). Se espera de él, de acuerdo con las políticas y directrices europeas, que asuma un papel activo y la mayor parte de la inversión. El usuario “institucional” se decanta, en sus usos, hacia la comunicación entre profesionales afines y la consulta de información sobre subvenciones y ayudas, principalmente.

Cabe esperar que este usuario sea un promotor decidido del empleo de productos multimedia, y en algunos casos, modelo a imitar, (véanse las aplicaciones creadas por el Instituto Madrileño de Formación, o el INEM), aunque no es su responsabilidad la construcción de un entramado comercial. El usuario particular hace uso del multimedia con el fin principal de obtener información. El perfil del usuario medio resulta cuando menos curioso. Mayoritariamente, es una persona joven, menor de 25 años (70%); en segundo lugar, en la clasificación por grupos de edad aparecen los usuarios entre 25-40 años (23%). El usuario medio tiene estudios de nivel superior (66 %) y no tiene ingresos propios (63 %) y su actividad principal es el estudio (63%). Sus contenidos preferidos son las enciclopedias electrónicas y la formación de adultos.

7.1.3- Contenidos multimedia

Como principal aplicación, figuran las b.d. representadas por 72 productos. En segundo lugar, y casi con el mismo porcentaje, aparece el “libro electrónico”(71 productos), que curiosamente no incluye las “enciclopedias electrónicas” (36 productos), que representan un 16% aprox. del mercado. Merece también mencionarse que los cursos de formación (35 productos) tienen prácticamente la misma presencia porcentual que las enciclopedias.

<i>Tipo de aplicación</i>	<i>Nº de productos</i>	<i>%</i>	<i>Tipo de aplicación</i>	<i>Nº de productos</i>	<i>%</i>
<i>Bases de datos</i>	72	31,8	Vídeo animado	7	3,0
<i>Libro electrónico</i>	71	31,4	Realidad virtual	3	1,3
<i>Juego/entr/oicio</i>	53	23,4	GIS	3	1,3
<i>Enciclopedias</i>	36	15,9	TV Interactiva	3	1,3
<i>Curso de formación</i>	35	15,4	Video-juego	2	0,8
<i>Prensa electrónica</i>	8	3,5	Otros	30	13,2

7.1.4- Sectores de mercado; servicios y aplicaciones

El sector de educación/formación con 138 productos y representando el 61% del mercado, es el mejor atendido. La necesidad de formación y educación adecuadas a las nuevas circunstancias, léase Sociedad de la Información, es considerada por todos los estudiosos como una condición indispensable para el triunfo de ésta.

<i>Sectores</i>	<i>Nº de productos</i>	<i>%</i>	<i>Sectores</i>	<i>Nº de productos</i>	<i>%</i>
<i>Educación/formación</i>	138	61,0	Universidad	11	4,8
<i>Gran público</i>	120	53,0	Bibliotecas/C. Doc.	9	3,9
<i>Información</i>	44	19,4	Puntos de ventas	5	2,2
<i>Museos/cultura</i>	36	15,9	Medicina	5	2,2
<i>Juegos</i>	23	10,2	Edición de vídeo	2	0,8
<i>Puntos de Información</i>	18	7,9	Televisión	2	0,8
<i>Turismo</i>	14	6,2	Teleconferencia	1	0,4
<i>Presentaciones</i>	14	6,2	Otros	5	2,2

Los servicios multimedia básicos que se ofertan al cliente aumentarán previsiblemente. Servicios hasta ahora considerados minoritarios, como la transferencia de ficheros, el correo electrónico o la videoconferencia se popularizarán por sus aplicaciones en el mundo empresarial, principalmente. La aparición de nuevas redes de comunicación - las llamadas “redes de alta velocidad”-, servirán de infraestructura adecuada para superar los problemas de transmisión y gestión de la información multimedia. Proliferarán los servicios de valor añadido: telebanco, telecompra, educación a distancia, televisión a la carta, etc. Los contenidos de alta calidad se revelan ahora como un factor crítico, siendo necesaria la creación de productos específicamente adaptados a esta tecnología y que hagan uso de sus capacidades específicas. (Bajo el acogedor manto del multimedia, se reúnen demasiados productos que no son más que una reedición de libros y documentos tradicionales en un soporte diferente.) Entre las múltiples aplicaciones posibles para los sistemas multimedia, aquí nos interesan:

- **Aplicaciones a la Documentación y Gestión de la información:** En este área, las aplicaciones son numerosas y significativas. Quizá, la mayor popularidad la ostenten las “unidades de información virtuales”, nos referimos a las bibliotecas, museos y archivos virtuales que, gracias a las comunicaciones, ofrecen acceso remoto a enormes volúmenes de información y documentos. Actualmente, la mayoría de los fondos bibliográficos disponibles en estas instituciones se refieren a autores y obras clásicas (libres de derecho de autor) y, en muchos casos, se trata de versiones digitales de materiales manuscritos, raros e incunables que amplían, de esta manera, su disponibilidad efectiva para el estudio y la consulta, al tiempo que se preserva su integridad. Desde un punto de vista de la gestión de información, los multimedia (en especial, sistemas de almacenamiento óptico) permiten un tratamiento integrado de los distintos tipos de documentos o información en un único soporte. Respecto a la recuperación de la información, los sistemas multimedia han ampliado las formas en que el usuario busca y recupera la información que necesita. Además de la recuperación tradicional de las bases de datos, por medio de palabras claves o descriptores, los sistemas multimedia suelen permitir también, y simultáneamente, la búsqueda basada en opciones o sistemas de “menús” preestablecidos, que adoptan la forma de clasificaciones temáticas, de temas generales a más específicos. Y finalmente, los más avanzados, suelen permitir la búsqueda de información mediante sistemas gráficos o mapas. Estos últimos sistemas de representación gráfica son específicos de este producto, y han devenido fundamentales como herramienta para paliar los problemas de desorientación del usuario, inherentes a la exploración (“navegación”) en una estructura hipertexto, con conexiones múltiples y no lineales..
- **Aplicaciones a la formación:** En contra de lo que pudiera parecer, por esa especie de euforia colectiva antes mencionada, los herramientas pedagógicas multimedia no están todavía al alcance de todas las instituciones formativas, ni su empleo es mayoritario en la enseñanza. El empleo del multimedia como herramienta instructiva supone importantes cambios en los métodos de enseñanza y aprendizaje, afectando a pilares básicos de la filosofía educacional: principalmente, al papel del estudiante o aprendiz. Frente a concepciones tradicionales, basadas en la responsabilidad central del profesor -como transmisor de contenidos y supervisor del proceso instructivo - la pedagogía actual plantea el traslado al estudiante del control de aprendizaje: siendo éste el responsable de determinar cómo, cuando, y qué aprender. Este cambio, que el multimedia hace factible desde el punto de vista tecnológico, se encuentra con dos obstáculos primordiales en su implantación: la aceptación por parte del profesorado y por parte de los estudiantes. Los primeros, para quienes supone un evidente esfuerzo de actualización y reciclaje, aceptan fácilmente el multimedia como un instrumento de apoyo complementario, pero son reacios a modificar su metodología tradicional y a aceptar su nuevo rol, ser un guía más que un depositario del saber. Los estudiantes, por su parte, tienen graves dificultades para llevar a cabo un auténtico aprendizaje multimedia, en el sentido de un proceso de exploración individualizado y autoprogramado, porque carecen de formación real y útil sobre cómo se aprende (metacognición) y nunca han sido impulsados a planificar los horarios, contenidos, o los métodos de aprendizaje o autoevaluación.. Sin embargo, el estudiante encuentra el multimedia “intrínsecamente motivante”, es decir, atractivo y lo acepta fácilmente como herramienta de aprendizaje, que le permite la comunicación (con otros estudiantes y con sus profesores o tutores) y el trabajo en cooperación.

En general, podemos decir que la educación basada en el concepto multimedia requiere un estudiante más maduro, activo e implicado en el proceso de aprendizaje que debe ser formado, en este sentido,

desde la educación básica. En el caso concreto de la formación de documentalistas - entendidos como gestores de información - resulta evidente que se trata una herramienta imprescindible para lograr profesionales que se adapten al mercado laboral actual. Los estudios deben incluir no sólo la utilización instrumental de productos multimedia sino también el diseño y gestión de este tipo de bases. El uso, cada vez más generalizado, de Internet como herramienta de formación, facilita a las instituciones públicas un vía de enseñanza multimedia de coste moderado. Pero la Universidad española no debe contentarse con ser un usuario multimedia sino que debe convertirse en creador de contenidos, especialmente cursos de formación y entrenamiento que se adapten a las necesidades de usuarios, permitiendo una progresiva independencia de los horarios de estudio rígidos así como de los escenarios físicos de aprendizaje.

7.1.5- Políticas de desarrollo

Este nuevo mercado y este conjunto de actores requieren un marco legal que se adapte a las nuevas circunstancias. Como regla general, los gobiernos tienden a adoptar medidas de ordenación que, por un lado, liberalicen el mercado y eviten los monopolios, y por otro, sean lo menos entorpecedoras para la evolución de las nuevas industrias y servicios.

El papel de los poderes políticos, especialmente en Europa, es el de proporcionar este corpus legislativo de una manera rápida (sin dilaciones) y eficiente, es decir, que permita el libre intercambio de bienes y servicios, protegiendo las peculiaridades nacionales y potenciando la presencia del producto europeo en el mercado mundial.

Dentro de las medidas políticas, en el ámbito nacional, destaca el Plan Nacional De Banda Ancha o PlanBA. Este plan nacional se desarrolla entre 1992-1995, y tiene como objetivo la puesta en marcha de un demostrador de red BA experimental. Se pretenden obtener, tras su desarrollo, diversos prototipos de elementos de red: terminales, adaptadores, aplicaciones, etc. Funcionará en Modo de Transferencia Asíncrono (MTA), una tecnología de transmisión y conmutación de paquetes muy flexible. Proporcionará velocidades de transmisión por encima de los 2 Mb/s (normalmente a 155 Mb/s). En la tabla adjunta, puede verse una breve descripción de los distintos programas y sus objetivos. Destacamos las aplicaciones en el campo de la Medicina y, por la relación con nuestro campo de trabajo, el proyecto EDUBA, aplicado a la formación y aprendizaje.

PROYECTO	DESCRIPCIÓN	COORDINADOR
TELEMEDICINA	Aplicación para el intercambio de información multimedia entre hospitales	IBERMÁTICA
MEDISAT	Aplicación para telediagnóstico médico	PHILIPS S. MÉDICOS
EDUBA	Aplicación multimedia para entornos de educación	SW DE BASE
PINTUR	Aplicación de informática turística multimedia	IBERMÁTICA
SIMM	Servidor de información multimedia de propósito general	CITAM
TEMA	Terminal multimedia en doble versión: bajo coste y altas prestaciones	TELEFÓNICA I+D
AGRADA	Sistema de gestión de redes privadas	TCP S-I
CRIPTO	Seguridad de acceso y criptografía en redes de banda ancha	CSIC
AFTER	Adaptador flexible de terminales (FDDI y G.703)	ALCATEL SESA
UNICORN	Unidad de interfuncionamiento entre Redes Privadas IEEE802.6 y red MTA	SIEMENS
IRMEN	Unidad de interfuncionamiento entre Redes Públicas IEEE802.6 y red MTA	ALCATEL SESA
TRI-622	Terminador de Red 1 a 622, 02 Mbit/s	TELEFÓNICA I+D
TEST-BA	Instrumentación de medición en el nivel físico de RDSI-BA	ICT ELECTRONICS
IFCO	Equipo terminal a 2,5 Gb/s para línea de fibra óptica con detección coherente	ALCATEL SESA
RAL-MT	Infraestructura de Red de Área Local con tecnología MTA	UPM-ETSIT
PROYECTO DE INTEGRACIÓN	Responsable del interfuncionamiento de todos los proyectos	TELEFÓNICA

De entre las diversas líneas de investigación que se desarrollan en este campo, destacan los trabajos relativos al “interfaz de usuario”, es decir, al medio de comunicación que permite al hombre relacionarse o interactuar con la máquina y el sistema. El objetivo principal de estas investigaciones es la obtención de un dispositivo físico y lógico que disminuya o suprima las barreras existentes. Para el usuario, la comunicación debe ser fácil: asequible a cualquier usuario sin formación previa, atractiva y múltiple. El desarrollo de este área, lógicamente, está directamente relacionado con la vital importancia que el usuario tiene para el triunfo de la tecnología multimedia. En un sistema ideal, el interfaz sería tan transparente y amigable que el usuario acabaría por olvidar que existe y percibiría la relación con la máquina como una comunicación directa, sin necesidad de un canal artificial intermedio. En este sentido, los programas de reconocimiento de voz, que permiten prescindir de teclados o ratones, representan un avance en la supresión de barreras para la comunicación hombre-máquina y una muestra del tipo de comunicación que probablemente tendremos en un futuro cercano.

7.2- La integración de multimedia en las bases de datos.

La aparición y desarrollo de la tecnología multimedia ha revolucionado el concepto tradicional de base de datos, entendidas como elementos de información textual y numérica, los cuales a su vez eran organizados de acuerdo a un conjunto preestablecido de normas. Estos componentes se ven enriquecidos por la posibilidad de completar los objetos a los que representan mediante elementos gráficos y/o sonoros, en un entorno integrado. El usuario obtiene entonces una representación compleja, múltiple, de los objetos del mundo real, y de sus categorías de información, que son representados en la base de datos. El ambiente de un sistema de gestión de bases de datos multimedia integra texto, datos, video, imagen (estática y dinámica) y sonido.

La integración de diferentes tipos de información en un único documento pone de relieve la importancia de las tareas de representación del contenido informativo, especialmente la indización de imágenes y documentos sonoros. Los mecanismos de recuperación de información de estos sistemas siguen utilizando, (excepto aquellos más punteros, todavía en fase de investigación) términos como elemento de formulación de requerimientos. La utilización de complejos mecanismos y normas de indización de estos documentos merece cada vez en mayor volumen la atención de los investigadores (WALKER y JONES, 1994; AZORÍN Y LÓPEZ, 1994), por lo menos hasta que se difundan mecanismos de recuperación basados en patrones gráficos o técnicas similares.

En una base de datos multimedia se está trabajando con un marco que no sólo incorpora los datos. Esos datos pueden tener variabilidad espacial y temporal. Por lo tanto, un documento introducido en una base de datos multimedia es una composición temporal, en la cual hay que introducir los diferentes tipos de datos, tanto como las relaciones de configuración y temporales existentes entre ellos. Los datos deben estar sincronizados, controlando tanto su estado como su comportamiento.

7.2.1- Los documentos multimedia.

Los documentos multimedia se forman mediante la agregación de objetos de diferentes tipos, objetos que pueden ser de gran tamaño, y llegan a mostrar gran variedad en lo referido a estructura y representación. La representación del documento multimedia busca comunicar de forma efectiva no sólo el documento y su contenido, sino también la semántica, los conceptos subyacentes y la relación entre ellos. Una comunicación de información que siga estas normas tiene asegurado un adecuado nivel de calidad. Por lo tanto, resulta ineludible fijar como uno de los enfoques básicos el análisis de las características de los usuarios humanos del sistema, para lo cual deben analizarse los patrones de comportamiento del usuario en el acceso y comprensión de las estructuras informativas (BIANCHI et alii, 1996).

Sirva el párrafo anterior como introducción para establecer una primera conclusión: en el momento en el que se comienza a diseñar e implantar una base de datos multimedia, el objeto de trabajo ya no resulta ser el registro tradicional: el usuario está tratando con un documento electrónico. Este ha sido creado procesando los diferentes tipos de información (texto, gráficos, imágenes, datos, imágenes animadas, sonidos) presentes en un almacén o repositorio de información. A esto hay que añadir que tras largos años de implantación y desarrollo de las tecnologías de bases de datos, las organizaciones y empresas disponen de grandes volúmenes de información almacenadas en bases de datos, generalmente en diferentes plataformas y aplicaciones, distribuidas en diferentes localizaciones geográficas. Estas circunstancias exigen que la implantación y explotación de bases de datos multimedia exija dos elementos:

1. La disponibilidad de bases de datos avanzadas, en un entorno de fácil utilización por parte del usuario.
2. El establecimiento de redes de telecomunicaciones de alta velocidad.

7.2.2- Los sistemas de gestión de bases de datos multimedia.

De hecho, es la heterogeneidad de los tipos de información que son necesarios en la actualidad unas las razones que ha favorecido, por parte de la industria y los usuarios, el desarrollo de sistemas de gestión de bases de datos multimedia, que han sido llamados también "gestores de información hipermedial" (DÍAZ, CATENAZZI y AEDO, 1996: 174). Las aplicaciones SGBD tradicionales ofrecían limitaciones en aspectos como el acceso complejo a datos, la transferencia de datos con otros sistemas, o la inexistencia de adecuados interfaces de usuario. Como respuesta, se tiende a diseñar e implementar nuevos SGBD que sean capaces de utilizar "inteligentemente" los datos disponibles, e integrar las viejas y las antiguas aplicaciones de forma no traumática. Una base de información hipermedial tiene varios componentes:

1. Base de presentación: parámetros a aplicar para mostrar la información al usuario.
2. Base de estructura: visión lógica del hiperdocumento, según un modelo.
3. Base de contenido: conjunto de documentos que se integran en el hiperdocumento.
4. Base de utilización: información sobre hábitos y comportamiento de cada usuario.

En este mismo sentido, la concepción de una base de datos multimedia, en su modelo conceptual (correspondiente al esquema conceptual definido por ANSI/X3/SPARC), debe cumplir dos fases (CHORAFAS, 1994: 312):

1. Cognición, centrado en cómo reconocer el mundo real, sus entidades y relaciones.
2. Modelización, centrado en cómo representar los conocimientos obtenidos en la fase anterior, de manera que sean manipulables por la máquina.

A pesar de ser la integración de bases de datos heterogéneas una de las razones del desarrollo de sistemas de gestión de bases de datos multimedia, la industria todavía no ha establecido todavía estándares de formato entre ellas, por lo que se repite la misma situación que en momentos anteriores, entre ficheros de base de datos correspondientes a aplicaciones como dBASE, Paradox, etc. Aunque las aplicaciones tradicionales ya ofrecen "puentes" para compartir sus bases de datos, por el momento esto no es posible en lo que respecta a las bases de datos multimedia. Sin embargo, están comenzando a aparecer en el mercado nuevas aplicaciones que, sin cumplir los requerimientos clásicos de un SGBD, pueden generar documentos multimedia, tomando como base información y datos contenidos en otros tipos de fichero. Los límites entre la base de datos "real", y el documento multimedia resultante "virtual", son difíciles de establecer en el estado actual de la tecnología. Si puede decirse que es posible establecerse un algoritmo que define el proceso de formateo y composición del documento, en dos niveles:

1. Jerarquía lógica del contenido del documento.
2. Jerarquía física del documento formateado.

Esta es la premisa que define la norma ODA (Open Document Architecture) de ISO. En lo que respecta a otros estándares, la norma ISO que define el EDI (Electronic Document Interchange, y su versión EDIFACT), y el protocolo ANSI X.12 están más orientados a la estructura del documento que a sus posibles aspectos multimedia, lo que por el momento dificulta su aplicación real.

7.2.3- Componentes de una base de datos multimedia.

En la actualidad nos encontramos con la necesidad de manejar una gran cantidad de material digitalizado; texto, video, imágenes, sonidos son las principales formas en las cuales podemos encontrar dicho material. Todos estos datos multimedia, deben estar almacenados en una Base de Datos para poder ser manipulados convenientemente. Tal y como se ha comentado a lo largo del trabajo, una base de datos es una colección de datos que es gestionada y organizada por un software específico; el DBMS (DataBase Management System, Sistema de Gestión de Base de Datos). Un DBMS es, sustancialmente, un software que se coloca entre el usuario y los datos como tales.

Pero sabemos que la mayoría de las bases de datos relacionales actuales no soportan gran volumen de información, como la de almacenar un video, que puede llegar a ocupar hasta varios gigabytes.

Hay que tener en cuenta varios problemas a la hora de guardar datos multimedia en una base de datos:

- La base de datos debe soportar objetos grandes, dado que los datos multimedia como los vídeos pueden ocupar hasta gigabytes de almacenamiento.
- En muchas aplicaciones de datos multimedia se necesita de la recuperación basada en la semejanza. Por ejemplo en una base de datos que guarda imágenes de huellas dactilares se proporciona la imagen de una huella digital para consulta y se deben recuperar las huellas dactilares de la base de datos que sean parecidas.
- La recuperación de algunos tipos de datos, como el sonido y el vídeo, tiene el requisito de que el suministro de datos debe realizarse a una velocidad constante garantizada. Por ejemplo si los datos de sonido no se proporcionan a tiempo, habrá interrupciones en el sonido. Si los datos se proporcionan demasiado rápido, la memoria interna del sistema pueden desbordarse lo que daría lugar a la pérdida de datos.
- Además deben tener operaciones específicas para trabajar con estos tipos de datos multimedia.

Multimedia es una nueva plataforma donde se integran imágenes, videos y sonidos para hacer ciertas tareas que proporcionan a los usuarios nuevas oportunidades de trabajo y acceso a nuevas tecnologías. Multimedia permite utilizar el texto para interactuar con los sistemas de información. Permite mostrar impresionantes imágenes de gran colorido y excelente resolución, animación y vídeo real. Brinda una mejora significativa en la efectividad de la comunicación. La riqueza de los elementos audiovisuales, combinados con el poder del computador, añaden interés, realismo y utilidad al proceso de comunicación.

Una base de datos multimedia está compuesta, por lo tanto, de objetos. Los objetos multimedia representan una abstracción de la vida real, es decir, es la manera de representar a través del computador momentos y situaciones de la vida cotidiana. Los objetos multimedia suelen tener atributos de descripción, como los que indica el momento en que se crearon, la persona que los creó y la categoría a la que pertenecen. Entre los diferentes tipos de objetos multimedia, se encuentran los siguientes:

- Video
- Texto
- Imágenes
- Audio

7.2.4- El mercado multimedia en España:

La investigación de la situación actual del sector multimedia en nuestro país adolece de una cierta escasez de estudios y trabajos en profundidad. En el mercado español, una de las fuentes es el Catálogo de servicios españoles de información electrónica ASCII, elaborado por FUINCA. En su estudio estadístico sobre 224 servicios españoles, la tipología de BD usada es:

- Servicios referenciales
- Servicios Fuente (entre los que aparece la subclase Icónicos, que incluye imágenes, gráficos, mapas, etc.)
- Servicios mixtos (combinación de algunos de los anteriores).

Es decir, no refleja en ningún momento el multimedia como una categoría aislada. Por otro lado, las cifras y datos adolecen de una cierta antigüedad. Las cifras que destacamos y que sirven como indicadores generales son las siguientes: Los servicios españoles de información han aumentado su oferta en 365% entre 1984 y 1993. Dentro de éstos, los servicios mixtos (los que presuponemos más cercanos al multimedia) son 50 (21%), en 1993. Aumenta la presencia del productor privado, frente al tradicional predominio del productor público en el mercado español, pasando de un 4% a 20%. Al mismo tiempo, la oferta de productos institucionales reduce su importancia en los contenidos, pasando de representar un 96% a un 66% del total.

En cuanto a la forma de acceso, la mitad de los usuarios aproximadamente (46%) lo hace en línea, aunque es una tendencia a la baja. El soporte multimedia supone unos ingresos “off-line” de 847 millones y son los más altos entre los distintos tipos de soportes. La segunda fuente utilizada para el estudio del mercado nacional es el informe Multimedia 1996 -Tendencias, de la serie Informes anuales de Fundesco. En este estudio, se recogen y analizan 226 productos dentro del mercado español y se incluye un útil inventario de los mismos. Según sus datos, las características del multimedia español son: · Soporte y acceso: El CD-ROM es el soporte más utilizado (el 45,8% de los usuarios) en España, aunque hay que destacar que más de la mitad de los usuarios disponen de acceso telemático. La conexión se establece principalmente vía módem, y tiene como objetivo Internet, con una frecuencia de un uso diario. En cuanto a los nuevos servicios, destaca la débil penetración de la videoconferencia (8%) y la realidad virtual (2%).

<i>Soportes</i>	<i>Nº de productos</i>	<i>%</i>	<i>Soportes</i>	<i>Nº de productos</i>	<i>%</i>
<i>CD-ROM</i>	157	69,4	DV/Indeo	4	1,7
<i>PC</i>	79	34,9	Realidad Virtual	2	0,8
<i>MAC</i>	35	15,4	EB, PDA	1	0,4
<i>CD-I</i>	18	7,9	TV Interactiva	1	0,4
<i>CD-ROM XA</i>	13	5,7	Video CD	1	0,4
<i>LaserDisc</i>	6	2,6	Otros	22	9,7

7.2.5- Formato de los componentes de una base de datos multimedia.

Video

Es evidente el auge que han tomado hoy en día los videos dentro de los datos que manejan los computadores, especialmente a través de Internet. La necesidad de acceder los datos en formato de video se presenta en una gama inmensa de aplicaciones, y típicamente los modelos de acceso varían considerablemente de una aplicación a otra. Algunos posibles modelos de acceso a videos son los siguientes:

- Consultas en una base de datos de vídeo: Al realizar una consulta en una base de datos de vídeo, estamos principalmente interesados en los tipos siguientes de consultas:
 - Recuperación de Segmentos: Encuentre todos los segmentos, de uno o más videos en la Base de Datos que satisface una condición dada.
 - Recuperación de Objetos: Dado un vídeo y un segmento del vídeo [s,e], encuentre todos los objetos que aparecen entre s y e.
 - Recuperaciones de actividades: Dado un vídeo y un segmento[s,e], encuentre todas las actividades que aparecen entre s y e.
 - Recuperaciones basados en propiedades: En estas consultas nos dicen que encontremos todos los videos y/o segmentos de vídeo en el que un objeto o una actividad con ciertas propiedades aparecen.

Existen diversos formatos de archivos de video dependiendo de la plataforma hardware y el software seleccionado- El video QuickTime de Apple para línea Macintosh, siendo uno de los formatos de archivos de video más ampliamente utilizados. Video for Windows de Microsoft es el competidor más próximo de QuickTime. Ambos ofrecen a los usuarios, la posibilidad de capturar video, audio, animación e imágenes, todo ello digitalizado; cortar y pegar en aplicaciones; y almacenarlos, recuperarlos y editarlos. La siguiente Tabla resume los distintos formatos de archivos de video:

FORMATO DE ARCHIVO.	NOMBRE COMPLETO	EXTENSIÓN DE ARCHIVOS DE PC	ESTÁNDAR O PATENTADO
AVI	AudioVisual intercalado	.avi	Patentado (Intel)
DVI	Digital Video Interactivo	.dvi	Patentado (Intel)
Indeo	Intel Video Technology	N/A	Patentado (Intel)
MME	Multimedia Extensions	N/A	Patentado (Microsoft)
QuickTime	QuickTime	.mov	Patentado (APPLE)
MPEG	Moving Picture Experts Group	.mpeg	
Vivo		.viv	
DVD	Digital Versatil Disc	.	

AVI (AudioVisual intercalado)

Uno de los más populares formatos de presentación de video es el formato AVI desarrollado por Microsoft. Este formato fue desarrollado para ejecutar videos en la plataforma Windows.

<i>Extensiones</i>	<i>Plataforma(s)</i>	<i>Medio</i>
.avi	AVI es por defecto el formato de video para ambientes Windows, pero existen players para algunas plataformas tales como Amiga y Macintosh.	Video

Características

- AVI posee un amplio rango de cualidades de video.
- Este puede aceptar 256 millones de colores codificados así como soportar sonido desde 5Khz Mono hasta calidad de sonido estéreo CD.
- De acuerdo a las necesidades que se tengan en términos de calidad y espacio en disco, AVI puede comprimirse.

DVI (Digital Video Interactive)

El video digital interactivo (DVI) es un hardware patentado y una solución software para suministrar video de movimiento total, imágenes y audio en las computadoras personales. Los chips de hardware especializados de Intel permiten compresión y descompresión de alta velocidad y visualización de video de alta velocidad. Con la tecnología DVI, un usuario puede editar, operar sobre archivos, recuperar y enviar cualquier tipo de multimedia como si fuese un simple texto.

Indeo

Desarrollado, por Intel es un formato para convertir señales de video analógicas estándar NTSC (National Television System Commettee) en señales de video digital que pueden ser almacenadas y reproducidas en una computadora personal. El video se convierte vía una tarjeta de PC y se almacena en la unidad de almacenamiento secundario.

Extensiones Multimedia (MME)

Microsoft añadió soporte para multimedia a Windows en la versión 3.0. Las extensiones multimedia (MME) permiten a un usuario combinar video, audio, imágenes y animación en un PC multimedia. MME también proporciona soporte a formatos de archivos, bibliotecas enlazadas y la interfaz de control de medios (MCI). MCI es una interfaz independiente de la plataforma que permite que aplicaciones multimedia controlen la interfaz de diversos dispositivos electrónicos y archivos de computadora.

Quicktime

Quicktime es un formato de video digital desarrollado y soportado primordialmente por computadoras Apple. Este fue uno de los primeros formatos de video digital para PC y es muy común.

<i>Extensiones</i>	<i>Plataforma(s)</i>	<i>Medio</i>
<i>Mov, movie, qt.</i>	Cualquier PC ó estación de trabajo.	Video

Características:

- Excelente calidad de video.
- Fácil de capturar y editar.
- Una gran capacidad de compresión de sonido si es usado solamente sobre Macintosh.

MPEG (Moving Picture Experts Group)

Es un estándar de compresión de audio, video y datos establecido por la Unión Internacional de Telecomunicaciones. Originariamente había 4 tipos diferentes MPEG-1, 2, 3 y 4 que se diferencian en la calidad y ancho de banda usado. Una cadena MPEG se compone de tres capas: audio, video y una capa a nivel del sistema. Entre sus múltiples ventajas destacan su compatibilidad mundial, una gran compresión de datos y su poca degradación de la imagen.

- MPEG-1: La calidad es similar al VHS. Se usa para videoconferencia, el CD-i, etc. Si es usado a mayor velocidad, es capaz de dar más calidad.
- MPEG-2: En banda, proporciona 720x486 pixels de resolución, es decir, calidad TV. Ofrece compatibilidad con MPEG-1.
- MPEG-3: Fue una propuesta de estándar para la TV de alta resolución, pero como se ha demostrado que MPEG-2 con mayor ancho de banda cumple con éste cometido, se ha abandonado.
- MPEG-4: Está en discusión. Se trata de un formato de muy bajo ancho de banda y resolución de 176x144, pensado para videoconferencia sobre Internet, etc.

Video Vivo

El formato de video vivo es relativamente nuevo con respecto a los formatos ya existentes. Los beneficios de este formato son dos:

- Comprime video y sonido a muy altos ratios
- Se puede obtener videos vivo sobre conexiones regulares a 28.8 Kbps sin necesidad de contar con servidores especiales para despacho de video (streaming).

<i>Extensiones</i>	<i>Plataforma(s)</i>	<i>Medio</i>
<i>.viv</i>	La plataforma de producción es Microsoft Windows y Macintosh.	Video

Características

- El mayor beneficio de estos videos es la tasa de compresión en ratios.

DVD (Digital Versatil Disc)

Es el nuevo formato de distribución de películas de vídeo (y datos en general). Tiene el mismo tamaño que un CD-ROM pero su capacidad de almacenamiento va de los 4.5GB hasta los 17GB dependiendo de si se usan una o dos capas/ una o dos caras. El sistema de compresión de las imágenes es el MPEG-2. Entre sus principales ventajas destacan las siguientes:

- Soporta en un solo disco 8 idiomas con 32 subtítulos
- Pueden llegar a caber hasta 8 horas de audio y video
- El tipo de sonido puede llegar a ser envolvente Dolby Surround
- Posee una espléndida calidad de video.

Texto

Una base de datos de texto es simplemente una colección de strings puestos en un índice de una manera conveniente. Cuando un usuario quiere encontrar un documento relacionado a un tema T, el programa de la búsqueda intenta encontrar los documentos en la base de datos de texto que contengan este string T, haciendo match entre el string y las palabras encontradas en el documento. Desgraciadamente, este método ha demostrado ser bastante ineficiente por dos razones fundamentales:

- Sinónimos: Puede darse el caso de que el usuario este buscando un tema particular a través de una palabra muy precisa pero el documento aún cuando este muy ligado al tema no contenga la palabra introducida por el usuario.
- Homónimos: Otro problema fundamental son los llamados Homónimos, palabras que aún cuando se escriban exactamente igual pueden tener múltiples significados.

Para poder solventar estas contrariedades se han introducido dos nuevos conceptos:

- Latent Semantic Indexing LSI (Posicionamiento Semántico Latente) para acceder a bases de datos de texto con textos de volúmenes muy grandes. Esta técnica nos permite eliminar las palabras y frases que no pertenezcan al contexto buscado y también identificar las palabras similares a través de un diccionario de sinónimos.
- Telescoping Vector Tree TV-Tree (Árbol de Vectores Telescopicos). Supongamos que D es un conjunto finito de documentos, un algoritmo de búsqueda lo que hace es lo siguiente: Supongamos que queremos hallar información referente a un tópico T el algoritmo intuitivo retornaría A(t), donde A(t) son todos los documentos del universo D que contienen información referente al tópico T.

Imagen

Una imagen de computadora es un cuadro completo capturado por medio de un digitalizador o un dispositivo facsímile. El usuario final crea un dibujo de computadora poco a poco, una línea, un círculo o una curva cada vez. Las imágenes y los dibujos se almacenan digitalmente en el computador, bien sea con un formato reticular o con un formato vectorial. (Francis E. Chea 1996).

Las imágenes consisten en data de la imagen (Bits Digitalizados) y atributos que describen y caracterizan la data de la imagen. Las aplicaciones que trabajan con estas imágenes, a veces asocian información específica de la aplicación, como por ejemplo incluyendo el nombre de una persona retratada en una fotografía, descripción de la imagen, fecha de realización, etc. Guardándola en una columna o atributo de la tabla en la Base de Datos.

La data de la imagen (Pixel) puede tener profundidad variada (bits por pixel) dependiendo de cómo fue capturada la imagen, y puede ser organizada de varias maneras. La organización de la data de la imagen es conocida como formato de la data. InterMedia Image puede almacenar y extraer data de la imagen de cualquier formato.

El espacio de almacenamiento requerido para imágenes digitalizadas puede ser grande en comparación a los atributos de datos populares tales como texto y números, ya que deben ser convertidos a ceros y unos (Números Binarios) para ser almacenados en la base de datos como objetos binarios largos (BLOB – Binary Large Object).

Durante las últimas décadas las empresas han ido aumentando considerablemente su colección de Fotografías y demás datos con formato de imágenes. Por Ejemplo: NASA's EOSDIS recopila y colecciona enorme cantidades de imágenes de la tierra. Los Hospitales generalmente coleccionan imágenes de Rayos X. En los registros de pasaporte se coleccionan fotografías de las diferentes personas que allí solicitaron sus pasaportes. Es sólo recientemente que gracias a los nuevos equipos de digitalización de imágenes, se pueden digitalizar estas imágenes y así convertir lo que anteriormente eran archivos enormes en datos digitales, lo que significa ahorro en espacio, y mayor velocidad y facilidad de búsqueda.

En un ambiente relacional las búsquedas o consultas son específicas sobre campos específicos, por ejemplo si se quiere saber en una lista de personas cuales de estas nacieron en un determinado año, se realiza una consulta comparando este valor con el campo Año de Nacimiento de cada registro, sin

embargo en las bases de datos de imágenes se introduce el concepto de "Imprecise Match" (Coincidencia Imprecisa). Por ejemplo en un registro criminológico de fotografías de delincuentes se puede hacer una comparación con una fotografía de un sospechoso y pedirle a la Base de Datos que elabore un listado de todas las fotografías del registro que sean similares a la del sospechoso.

Otro uso común de las Bases de Datos para Imágenes es para la comparación de huellas dactilares, se imagina lo difícil que sería comparar manualmente una huella dactilar con un registro de millones de huellas? Esta labor tan tediosa puede realizarse en sólo algunas horas en un sistema de Bases de Datos de Imágenes.

Los formatos estándar de archivos gráficos permiten que los datos gráficos sean transportados entre distintos sistemas de computadora. Los formatos de archivos gráficos más comunes son BMP, CGM, EPS, PCX, TIFF, GIF, JPEG, PNG y WMF. Estos formatos y sus características se definen en la tabla que sigue a continuación:

<i>Formato de archivos gráficos</i>	<i>Nombre completo</i>	<i>Extensión de archivo</i>	<i>Estándar o patentado</i>
BMP	Mapa de Bits	.BMP	Estándar
CGM	Metarchivo de gráficos de computadora	.CGM	Estándar
DIB	Independiente del dispositivo mapa de bits	.DIB	Estándar
DRW	Diseñador Micrografx	.DRW	Patentado (Micrografx)
EPS	PostScript encapsulado	.EPS	Patentado (Adobe)
GIF	Formato de Intercambio de gráficos	.GIF	Patentado (CompuServe)
MS53	Estándar bitonal ANSI		Estándar
PICT y PICT2	Apple picture	.PIC	Patentado (Apple)
PCT	PC Paint	.PCT	Patentado
PCX	PC paintbrush	.PCX	Patentado
TIFF	Formato de Archivo de imágenes etiquetadas	.TIF	No estándar
TGA	Adaptador de gráficos TrueVision Advanced Raster (TARGA)	.TGA	Patentado (TrueVision)
WMF	Metaarchivo de Windows	.WMF	Patentado (Microsoft)
JPEG	Joint Photographic Expert Group	.JPEG	
PNG	Portable Network Graphic	.PNG	

- **BMP (Mapa de Bits):** Los archivos BMP son archivos de retículos de mapas de bits y contienen un bit por cada punto de la imagen almacenada en la memoria de la computadora. Cada bit, también denominado elemento de imagen o pixel, corresponde a un punto de la imagen. Un archivo de mapa de bits contiene filas y columnas de bits. Los formatos de imágenes de retículos BMP son generados por programas de dibujo por computadora como MacPaint, Microsoft PaintBrush y PC Paint.
- **TIFF (Formato de Archivo de Imágenes Etiquetadas):** El formato permite que las imágenes digitalizadas se integren en los programas de aplicación de las marcas de computadoras más importantes. TIFF soporta esquemas de 8 y 24 bits, así como los de color por escala de grises y de blanco y negro. Los archivos TIFF son bastante utilizados en las máquinas facsímil, los programas de diseño asistido por computadora (CAD) y las imágenes médicas.
- **GIF (Formato Gráfico de Intercambio):** El GIF es el formato frecuentemente utilizado en el Web para imágenes. Este no posee un método de compresión y usa un máximo de 256 (8-bit) paleta de colores. GIF usa el LZW (Lempel-Ziv Welch) esquema de compresión para reducir la dimensión del gráfico. LZW es para colores que no cambian dentro de una imagen. Existen dos tipos diferentes de archivos de formato GIF: GIF87a. y GIF89a. GIF87a es el básico, más comúnmente usado como formato GIF por soportar interenlaces. GIF89a, por otro lado, soporta transparencia y animación en adición al interenlace.

- **JPEG (Joint Photographic Experts Group):** Un formato en JPEG, creado para las imágenes fotográficas, asegura que la imagen mantiene la misma o comparable calidad cuando está online. Se diseñó la compresión de JPEG por las fotografías y continuo tono de imágenes. A diferencia GIF se limita a imágenes de 8-bit (256 colores), JPEG apoya 24-bit imágenes (sobre los 16.7 millones de colores). El formato de JPEG debe usarse para fotografía, imágenes de continuo-tono que tienen cambios coloridos sutiles. JPEG, sin embargo, no es apropiado para las imágenes con áreas grandes de color llano y ilustraciones de arte de línea.
- **PNG (Portable Network Graphic):** PNG es un nuevo formato del archivo para imágenes de Web que sostienen mucha promesa para el uso futuro. El formato del archivo no posee un método de compresión que permita imágenes manteniendo su calidad original. Con PNG, usted puede salvar gráficos con 8, 24 y 32 bit de profundidad de color (es decir, millones de colores).

Audio

Como en el caso de imágenes y vídeo, el volumen de los datos Audio puede caracterizarse de dos maneras básicas (eso puede llevar después a los híbridos): usando metadata que explica el volumen de un archivo Audio o extrayendo rasgos pertinentes de los datos de Audio procesando a través de diversas técnicas.

Metadata para representar el contenido de Audio. Considere un solo archivo conteniendo Audio, digamos, una Ópera de Wagner, En este caso, así como nosotros hicimos antes con los datos videos, nosotros podemos asociar un juego de segmentos de la línea (cada uno que se refiere a un periodo de tiempo) cubriendo la ópera entera, y nosotros podemos asociar con cada segmento de la línea un juego de actividades que ocurren en los periodo de tiempo denotado por aquellas líneas o segmentos. Cada actividad puede tener un juego de campos asociado.

1. Cantantes: Éste puede ser un campo fijo que contiene archivos que tienen un papel, tipo del cantante, y campo de nombre de cantante. Por ejemplo, si la tripleta (Lohengrin, Tenor, René Kollo) aparece como un miembro de los cantantes asociado con un segmento [5,9], entonces René Kollo, un tenor, está cantando el papel de Lohengrin durante el segmento de tiempo [5,9] de la ópera.
2. Score: Éste puede ser un campo de música que contiene apuntes relevantes de la música asociados con el segmento de tiempo [5,9].
3. Transcripción: Éste puede ser un campo del tipo documento que contiene apuntes de la parte pertinente del libreto (letra de canciones) durante el segmento de tiempo.

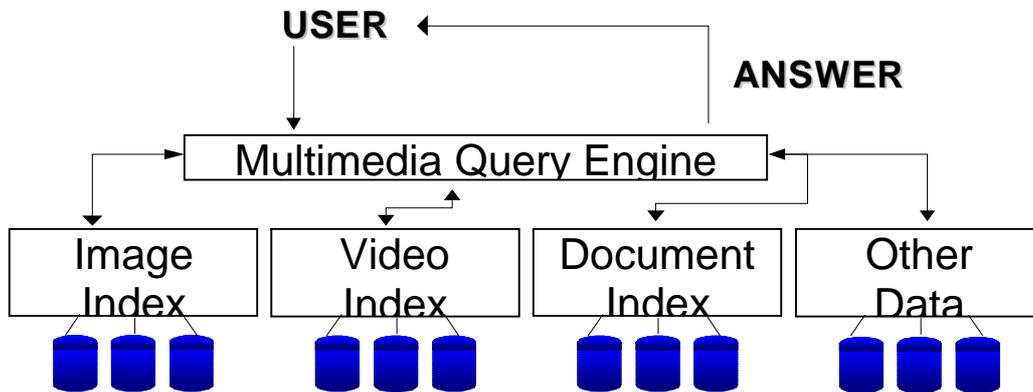
En general, la metadata representa el contenido del Audio que puede verse como un juego de capas de los objetos sobre una línea de tiempo, exactamente como en el vídeo. Los objetos, individuos, rasgos, y actividades que ocurren exactamente en Audio como en el vídeo. la única diferencia es que en Audio ellos se oyen pero no se ven, mientras en el vídeo ellos son tanto vistos como escuchados. Así, nosotros podemos poner en un índice el metadata asociado con Audio exactamente de la misma manera que en un índice de vídeo, y las mismas técnicas de procesamiento de consultas pueden usarse de nuevo en Audio.

7.3- DISEÑO Y ARQUITECTURA DE UNA BASE DE DATOS MULTIMEDIA

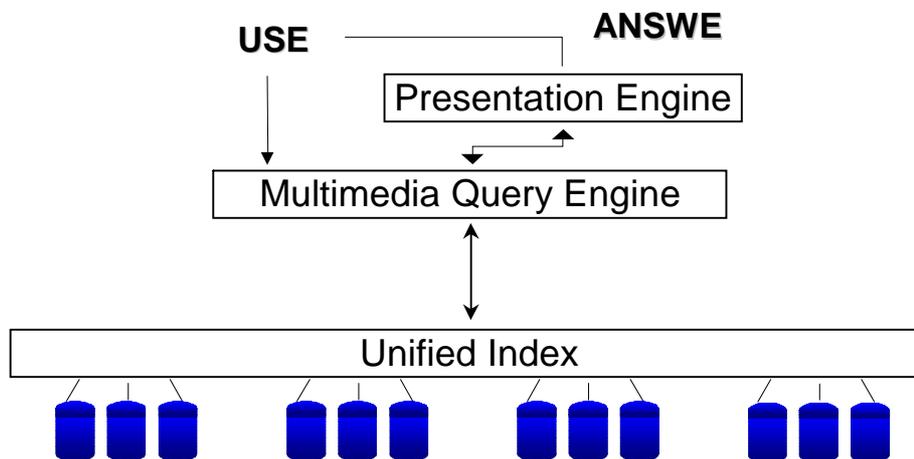
Cuando diseñamos un sistema de base de datos multimedia que representa una gran variedad de tipos, nosotros estamos forzados a confrontar severas cuestiones concernientes a la organización del contenido de los datos medios en los dispositivos de almacenamiento así como también de la organización física de los datos en los dispositivos de almacenamiento.

Se consideran tres arquitecturas para la organización del contenido de los sistemas de base de datos multimedia:

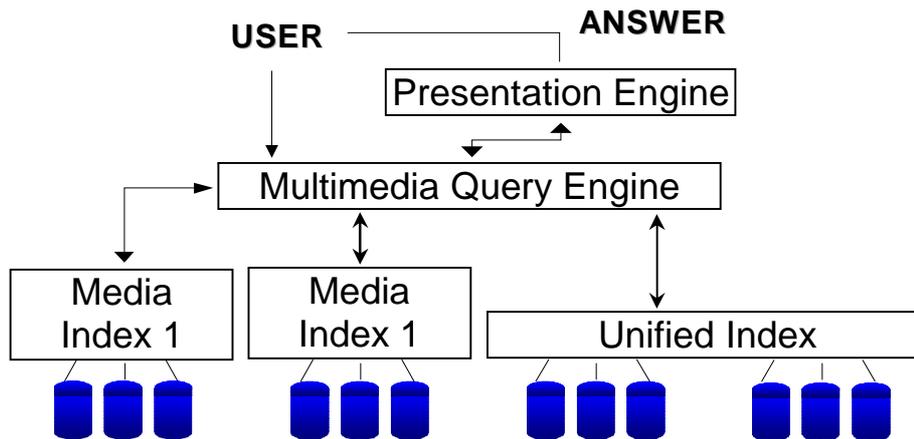
1.-El principio de autonomía. Bajo este principio se busca que los tiempos de acceso y respectiva captura de los datos se realicen de la forma más eficiente posible, esto significa invertir una cantidad de tiempo ínfimo en dicho proceso. Para lograr esto se organizan de manera correcta los datos según sea su tipo para a continuación asignar mediante una estructura auxiliar índices, los cuales dado que se requiera un dato de un tipo determinado, sea suficiente con navegar por la estructura del tipo asociado la cual me dará el enlace al dato dentro de la base de datos.



2.-El principio de uniformidad. Alternativamente, nosotros deberíamos intentar encontrar buscar una estructura abstracta que pueda ser usada para indexar todos los tipos media y pueda ser usado para crear índices unificados y poder acceder a objetos media. En otras palabras, podemos representar el contenido de todos los diferentes objetos media (imagen, vídeo, documento, audio, etc.) dentro de una única estructura de datos y desarrollar algoritmos para consultar esta estructura de datos. En la siguiente figura se muestra un diagrama conceptual de diferentes tipos de datos media, organizados acorde del principio de uniformidad.



3.-El principio de organización híbrida: Una tercera posibilidad es usar un híbrido de los dos principios anteriores. En efecto, acorde a este principio, ciertos tipos media usan sus propios índices, mientras otros usan índices unificados. Exactamente que tipo de medios usan que tipos de índices dependen de varias características que serán discutidas en breve. En la siguiente figura se muestra un diagrama conceptual de diferentes tipos de datos media organizados acorde al principio híbrido.



La siguiente tabla resume los conceptos anteriormente comentados.

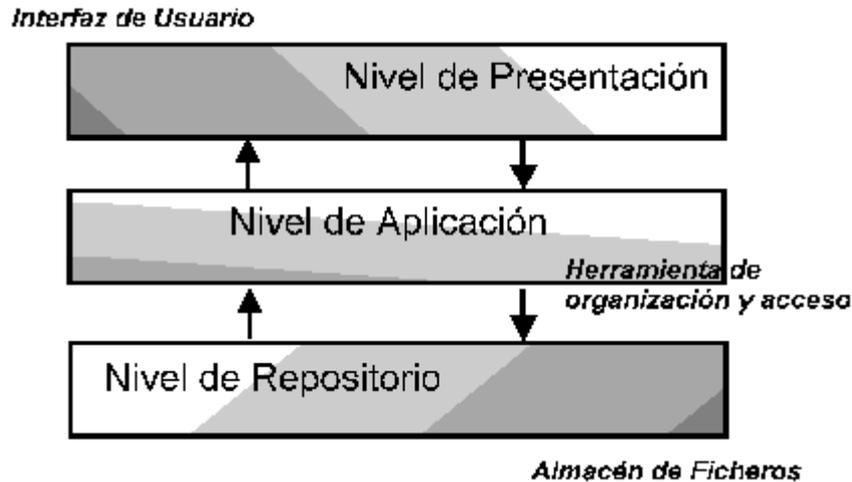
<i>Principios</i>	<i>Ventajas</i>	<i>Desventajas</i>
<i>Autonomía</i>	Tiempos de consultas relativamente rápidos.	Requiere la creación de algoritmos y estructuras de datos para cada tipo de media individual.
<i>Uniformidad</i>	Fácil de implementar, y los algoritmos resultantes son frecuentemente más rápidos.	Se puede perder información si el lenguaje para las anotaciones no es suficientemente expresivo para capturar todos los aspectos del contenido.
<i>Híbrido</i>	Se aprovecha de índices ya existentes y del código de software.	

Por lo tanto, la generación de base de datos multimedia, nos permite almacenar diferentes tipos de datos multimedia, con una gran capacidad de almacenamiento y ancho de banda requerido (en el caso de .bases de datos multimedia basadas en web o on-line), el uso de técnicas de compresión y descompresión de datos, que pueden ser utilizadas en una gran diversidad de sistemas. Y estamos seguros, que cuanto más avance la tecnología , el auge del concepto multimedia, será más importante todavía.

7.4- La metodología en una base de datos multimedia.

La creación e implantación de un sistema de bases de datos multimedia no puede separarse de la adopción de una metodología adecuada para ello. De la misma forma que en los sistemas "tradicionales" de bases de datos, el enfoque a utilizar es el basado en un sistema integrado de información:

1. Definición de las tareas y objetivos a cumplir por la unidad informativa.
2. Desarrollo de un metamodelo y de una metodología para el desarrollo de un sistema de información estratégica.
3. Inventario de los elementos de tecnologías de la información disponibles.
4. Inventario de las herramientas y de los repositorios de información existentes.
5. Creación de un diccionario de metadatos y de su metodología.
6. Consolidación de las conclusiones obtenidas, a través de metamodelos.



Los tres niveles de un sistema de bases de datos multimedia.

La utilización de este tipo de bases de datos, en lo que se incluyen representaciones complejas de la realidad, hace necesario la utilización de dos niveles de organización y de descripción. En primer lugar, una nivel de conocimiento, de metadatos, conceptual y difuso. En segundo lugar, un nivel de datos, concreto y analítico. El concepto clave de las nuevas bases de datos, a partir de la presente década, será "significado", superando el clásico "dato" o "información".

7.5- Documentación y bases de datos multimedia.

En lo que respecta al ámbito documental, las bases de datos multimedia facilitarán el enriquecimiento de la representación de la información (LÓPEZ YEPES, 1993). El documento no se reducirá a la utilización de unas categorías descriptivas, complementadas con la aplicación de un lenguaje documental. La inclusión de nuevos elementos descriptivos de los documentos, como una imagen, gráficos, o complementos sonoros, requerirán la experimentación, el desarrollo y la utilización de nuevos mecanismos de recuperación de información, de los cuales ya se están utilizando, a determinados niveles, algunos de ellos, como el reconocimiento e identificación de imágenes según esbozos de líneas maestras. La aplicación de representaciones semánticas de la información, seguramente basada en el enfoque orientado a objetos, que incluye tanto los datos como las acciones a ejecutar sobre los mismos o sobre terceros, será uno de los paradigmas a emplear en este entorno.

Una cuestión clave, en los nuevos sistemas de bases de datos multimedia, es el tratamiento de la imagen de los documentos, tanto en lo referido a su almacenamiento, como en lo referido a su procesamiento. De hecho, esta es una de las mayores áreas de expansión de la informática documental en los próximos años, sirviendo como indicador el auge que están alcanzado los llamados SGD (Sistemas de Gestión Documental). Los documentos, en cuyo origen suelen encontrarse en soporte papel, son introducidos en el sistema a través de algún mecanismo de captura, generalmente un escáner. Se almacenan como gráficos (en cuyo caso se pierde la posibilidad de efectuar búsqueda a texto completo), o como texto, mediante la utilización de una aplicación de tipo OCR (Optical Character Recognition), que permite integrarlo como un texto perfectamente manipulable y accesible, desde una perspectiva documental. Evidentemente, una aplicación que incorpore la segunda posibilidad, o ambas, posee una importante ventaja sobre los meros almacenes de imágenes de documentos.

Esta última cuestión pone de manifiesto que un adecuado sistema de gestión de bases de datos multimedia debería ofrecer herramientas para buscar, recuperar, manipular, ordenar y organizar los textos y las imágenes en un entorno de formato libre, de tal forma que pueda utilizar documentos de muy diverso formato, tamaño y disposición (NICHOLLS y RIDLEY, 1996; OZKARAHAN, 1995). El sistema de gestión de bases de datos multimedia debería ofrecer lenguajes de descripción de documentos avanzados, permitir el diseño de estructuras de datos muy flexibles, y ofrecer mecanismos de búsqueda altamente efectivos. Como corolario, un sistema de este tipo debería (CHORAFAS, 1994: 338-339):

1. Encontrar rápidamente la información multimedia, buscando cualquier número de objetos referidos a un contenido de información, en un contexto de texto completo (y, a poder ser, en un contexto totalizador de la información)
2. Conectar documentos según su contenido informativo, relacionando extractos de información relacionada de varios documentos, en uno nuevo.
3. Facilitar el acceso instantáneo a los ficheros pertinentes, incluyendo los mecanismos necesarios de compresión y descompresión para la manipulación de aquellos.
4. Conocer y manipular, por parte del usuario, la estructura misma de la información.
5. Crear relaciones entre grupos de elementos informativos.

Los documentos pueden encontrarse en una única localización, en un computador local, o bien ser el resultado de la integración de datos y representaciones dispuestos en ordenadores dispersos espacialmente. Esta última situación obliga a tratar el asunto de las bases de datos distribuidas. Se está hablando de una única base de datos a nivel lógico, pero de diferentes bases de datos a nivel físico. Esto supone que las aplicaciones deben acceder a diferentes tipos de información, en diferentes estructuras, a través de redes de ordenadores, en entornos sumamente heterogéneos. La necesidad que se deriva del panorama esbozado es la presencia de un diccionario de datos, repositorio general a través del cual se disponen las definiciones estándares de los objetos presentes en toda la extensión de la base de datos.

7.6- Hipermedia.

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

Una hiperbase debería cumplir los siguientes requisitos (DÍAZ, CATENAZZI y AEDO, 1996: 185-191):

1. Toda información es enlazable.
2. El hiperdocumento es dinámicamente modificable.
3. Es necesario guardar información sobre la presentación física de los contenidos.
4. Se debe dar soporte a todos los tipos de acceso a los sistemas hipermediales.
5. El hiperdocumento debe ser abierto y puede estar físicamente distribuido.
6. El sistema debe incluir mecanismos de recuperación ante fallos.
7. Debe permitirse el acceso concurrente.
8. Hay que preservar la seguridad de la información.
9. Hay que posibilitar el trabajo en cooperación.
10. El hiperdocumento puede estructurarse siguiendo una determinada jerarquía.
11. Existen contenidos virtuales enlazados dinámicamente.

8.- TELEDOCUMENTACION

8.1- Concepto de 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 (WALKER y JONES, 1993). 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.

8.2- Los actores de la teledocumentación

La teoría tradicional sobre teledocumentación (LANCASTER y WARNER, 1993) 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.

8.2.1- El papel de los soportes ópticos.

La teledocumentación ha sido puesta en entredicho en el cambio entre la década de 1980 y la década de 1990. La popularización de las bases de datos documentales en soportes ópticos, especialmente en CD-ROM, a un precio moderado, hizo pensar en ocasiones en una decadencia de los distribuidores tradicionales. Sin embargo, unos pocos años han bastado para poner las cosas en su justo sitio. No hay que ver ambas posiciones como enfrentadas, sino como complementarias. El CD-ROM ofrece un acceso local, a un coste módico, a bases de datos muy completas, que sin embargo ofrecen el problema de la actualización. No es posible una actualización inmediata de las bases de datos en CD-ROM, lo que sí es ofrecido por la teledocumentación. Como puede imaginarse, en áreas donde la actualización y la novedad en el estado de la información son decisivas, poco puede aportar el CD-ROM, como en el caso de una agencia de noticias. Sin embargo, en lo que se refiere a archivo, el CD-ROM posee una ventaja innegable frente a la teledocumentación (LIZASOAIN, 1992). Ambos deben considerarse bajo la óptica de las necesidades del usuario y del tipo de información, más que sobre las características de la tecnología utilizada.

8.2.2- Otros servicios de información en línea.

El mundo de los servicios de información en línea resulta ser sumamente dinámico y cambiante. En épocas recientes se ha visto una rápida aparición y caída de algunos sistemas, como el teletexto (CHAMORRO y GONZÁLEZ, 1992), o la generalización de sistemas de información y comunicación interpersonales, como los BBS (Bulletin Board Systems) (ABOBA, 1993). El teletexto, una iniciativa europea principalmente apoyada por Francia, ha tenido una vida y resultados efímeros, excepto en el caso del Minitel galo. Un insuficiente nivel de definición de la imagen, la lentitud en el acceso a la información y el elevado coste han sido definitivos en su fracaso, a pesar de ser una opción válida, en principio, para popularizar el acceso a la información mediante redes de telecomunicaciones. Muy diferente ha sido el caso de los BBS, los cuales se han desarrollado como foros electrónicos de acceso e intercambio de información, principalmente en los Estados Unidos de América, con empresas señeras como Compuserve o American On Line (AOL). Compuserve ofrece acceso directo a gran cantidad de bases de datos sobre todo tipo de cuestiones, incluyendo revistas con artículos a texto completo de variados ámbitos, información económica y comercial... La facilidad de acceso, la gran cantidad de información de todo tipo contenida en ellos, la posibilidad de intercambiar información con otros usuarios, los sistemas de mensajería electrónica, y la incorporación de algunos servicios de acceso a bases de datos, a través de las propias BBS, y el bajo coste, similar al de una llamada telefónica, así como la utilización como terminal de cualquier ordenador equipado con modem, han sido factores de éxito que han favorecido, casi sin transición, la entrada en el mundo de Internet.

8.3- Herramientas para el usuario final

El documentalista, como principal usuario final, e intermediario de otros posibles usuarios, debe disponer de un equipo que le permita acceder al distribuidor con las adecuadas garantías. Deberá disponer de un ordenador, que cumple las funciones de terminal del host, en el momento en el que entre en comunicación con éste. Sin embargo, los actuales ordenadores no vienen preparados, en su mayor parte, para la conexión a una red física de comunicaciones, por lo que resulta necesario adecuar el equipo a tal fin.

Para utilizar los sistemas de información electrónica disponibles a través de redes de comunicaciones hay que completar el ordenador con dos tipos de componentes, tanto de hardware como de software (DVORAK y ANIS, 1992). En primer lugar, es necesario instalar y configurar un modem (modulador/demodulador). La función del modem es la de actuar como interlocutor entre el ordenador y la red de comunicación telefónica. Por lo tanto, deberá convertir la información del ordenador (digital) en información capaz de viajar por la red de comunicaciones (analógica), y viceversa. En ocasiones el ordenador puede estar conectado a algún tipo de red de comunicaciones entre ordenadores que hará

innecesario este tipo de adminículo, en cuyo caso deberá ser sustituido por una tarjeta de red, que cumplirá el mismo cometido, aunque sin la necesidad de traducir entre analógico y digital.

Para acceder al host será necesario instalar un software o aplicación capaz de interactuar con el modem y presentar la información en nuestra máquina local de forma inteligible, para que cumpla su función de terminal. Se le suele denominar software de comunicaciones y/o terminal. Este software es el encargado de establecer la comunicación con el host, representar en la pantalla del usuario la información que se le envía, y enviar a su vez al host, las órdenes emanadas del usuario. Las características de este software pueden variar enormemente, dependiendo del gusto de usuario, de las facilidades de uso, de la recomendación por parte del host, o de la adecuación del mismo a las funcionalidades ofrecidas por el propio host. Incluso alguno de ellos facilita con el contrato un software de comunicaciones y terminal especialmente diseñado para la consulta de sus bases de datos.

8.4- El IRS (Information Retrieval System/Sistema de Recuperación de Información)

Una vez conectado el usuario al servicio, y validada su autorización, se encuentra en situación de utilizar el software de interrogación y recuperación que pone a su disposición el host. A este tipo de sistemas informáticos se les engloba bajo el término IRS (Information Retrieval System, Sistema de Recuperación de Información) (MEADOW, 1992). 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 (HAWKINS et alii, 1985) 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 tesoro: 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 (analizadas en el capítulo dedicado a las bases de datos documentales).
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.

```

CORDIS Main Menu Version 3.1
-----
Community Research and Development Information Service
NOTE : Menu Version 3.1 changes in database menus released on 1995-10-05:
       for details, type '00'.
What would you like to see?
  1+ A CORDIS database
  2  Information on CORDIS and on CORDIS databases
  3  Complementary RTD-related information
  5  General instructions for new users
Type option number, then press ENTER
  8 Help      9 Exit      88 Menu Version 3.1 Readme File

```

Una sesión en el host ECHO.

8.5- La creación de bases de datos en línea

La teledocumentación es, como se ha apuntado, un negocio. Como tal, la existencia de una base de datos consultable en línea, mediante un servicio comercial, responde a una decisión tomada en tal sentido por parte del productor. El factor decisivo es la existencia de un volumen considerable de usuarios, dispuestos a pagar por acceder a la información que pueda contener la futura base de datos. Una vez tomada la decisión inicial de crear la base de datos, es necesario adoptar otras decisiones complementarias que atañen a los siguientes aspectos:

1. Cobertura de la base de datos: área o campo que va a cubrir la base de datos.
2. Diseño de los registros individuales: determinación de las categorías de datos que permitirán representar a las entidades.
3. Selección de las fuentes primarias: selección de las fuentes y de los documentos de los que se recopilará información para ser introducida en la base de datos, de la cobertura temporal y espacial de los mismos, etc.
4. Creación del contenido de los registros: introducción de los datos pertinentes, tomando en consideración las reglas de descripción y de uso del lenguaje, de catalogación, de composición de consultas...
5. Mecanismos de introducción, manipulación y actualización de los datos.
6. Control de calidad: mecanismos que aseguren la adecuación del contenido de los registros a la realidad, y la satisfacción del usuario.

Puede apreciarse que las cuestiones señaladas resultan similares a las vistas en el capítulo dedicado a los SGBD y a la creación de bases de datos. El diseño y desarrollo de los recursos de información electrónica, que tienen como soporte principal bases de datos, siguen las normas y principios generales enunciados al tratar las bases de datos.

8.6- Clasificación de las bases de datos

Los distribuidores pueden ofrecer bases de datos de contenido temático muy variado: medicina, política, derecho, cultura... 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 su vez son representaciones de documentos primarios. Dentro de éstas distinguen, as 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.

8.7- Fases tradicionales de una búsqueda teledocumental

La bibliografía ha señalado como clásicas las siguientes fases en la ejecución de una búsqueda teledocumental (BATES, 1987):

1. Conexión con el host, identificación del usuario y autorización de acceso.
2. Selección del fichero o ficheros de datos a consultar.
3. Estudio de los descriptores a través de los tesauros.
4. Agrupación de los descriptores sinónimos en diferentes grupos.
5. Establecimiento de las relaciones existentes entre los conjuntos.
6. Combinación de los descriptores que representan a los conjuntos.
7. Formulación y ejecución de la ecuación de búsqueda.
8. Visualización de los registros resultantes.
9. Impresión o importación de los registros.

Este esquema tradicional resulta inadecuado en la actualidad (BASCH, 1993). Una búsqueda de este tipo consume gran cantidad de recursos económicos, tanto de comunicaciones como de uso del host, que son facturadas al usuario, lo que repercute en un aumento del precio final de acceso a la información. Más correcta resulta la ejecución de los pasos 2 a 6 en modo local. Es decir, utilizar los recursos facilitados por el distribuidor con el contrato como herramientas locales para agilizar estas tareas, de tal forma que en el momento de conexión al host se sabe positivamente cuales son los ficheros a utilizar, los descriptores pertinentes, y las relaciones entre ellos. El esquema propuesto obvia, además, la casi necesaria revisión de las ecuaciones, por lo que será recomendable efectuar un somero análisis de los documentos resultantes de una búsqueda entre los puntos 7 y 8, antes de ejecutar órdenes que puedan suponer una sobrecarga para el usuario.

Otra cuestión clave a considerar es la presencia de diferentes IRS en los hosts disponibles, lo que significa, a su vez, la necesidad de conocer y manejar varios lenguajes de interrogación (ARMSTRONG y LARGE, 1992). El usuario se ve obligado a conocer las características de los diferentes lenguajes, así como las particularidades de cada sistema, lo que supone un inconveniente para la optimización de la recuperación de información mediante teledocumentación

8.8- Estructura de los registros existentes en los host

Los registros que se incluyen en las bases de datos ofrecidas por los host suelen ofrecer una estructura muy similar, especialmente los contenidos en las bases de datos bibliográficas referenciales (FUNDESCO, 1986). Las categorías de datos contenidas responden a:

1. Número de identificación del registro.
2. Título.
3. Autor (en ocasiones se diferencia entre el primer autor y el resto).
4. Centro de trabajo (investigación, etc.) en el que se localiza el autor del trabajo.
5. Cita bibliográfica completa (obra, número, año, páginas, etc.)

6. Resumen.
7. Descriptores o encabezamientos.
8. Identificadores.

De esta forma combinan la necesaria estructuración de la información, con las categorías de datos que permiten describen cada uno de los recursos de información. La utilización de técnicas de manipulación y recuperación de grandes extensiones de texto es similar a la analizada al tratar los sistemas de gestión de bases de datos documentales

8.9- Servicios complementarios ofrecidos por los distribuidores

Los distribuidores o host no suelen limitar sus servicios a la puesta en línea para su consulta de las bases de datos. Junto a esas consultas, ofrecen otros tipos de servicios. Entre estos merecen destacarse los referidos a búsquedas retrospectivas y a las copias de documentos primarios.

Las características de las bases de datos y de los lenguajes de interrogación permiten realizar búsquedas retrospectivas por parte del usuario, facilitar la difusión selectiva de la información, y establecer servicios de alerta sobre los temas de interés. Hay que destacar que los servicios de alerta, junto a la creación de periódicos o boletines personalizados de información, son una de las áreas de mayor expansión e investigación en el mercado de la información estadounidense, que suele marcar las pautas en estas cuestiones.

La obtención de copias del documento original es otro de los servicios complementarios ofrecidos por los distribuidores, que pueden hacerlos llegar al usuario por diferentes vías, con criterios de coste y oportunidad sumamente variados.

Otro de estos servicios es la teledescarga, entendiéndolo como tal la obtención de una copia del documento o registro, sea referencial o factual, y su almacenamiento en el ordenador local. Es imposible controlar la teledescarga de registros por parte de los usuarios, por lo que los host han optado por aplicar diferentes medidas, que van desde la liberalización de la misma, hasta la carga de cuotas por cada documento que es visualizado por el usuario, y que por lo tanto puede ser susceptible de teledescarga.

8.10- Ventajas e inconvenientes de la teledocumentación

Al igual que todo servicio de acceso a la información, la teledocumentación ofrece ventajas e inconvenientes. Como ventajas deben señalarse el acceso inmediato a gran volumen de información, exhaustiva y muy completa, además de actualizada casi instantáneamente. Se dispone además de conocimiento sobre literatura gris o de difícil acceso. La obtención de los documentos pertinentes puede realizarse a través de varias vías, mediante la modificación inmediata de las ecuaciones de búsqueda, o el replanteamiento de los conceptos utilizados.

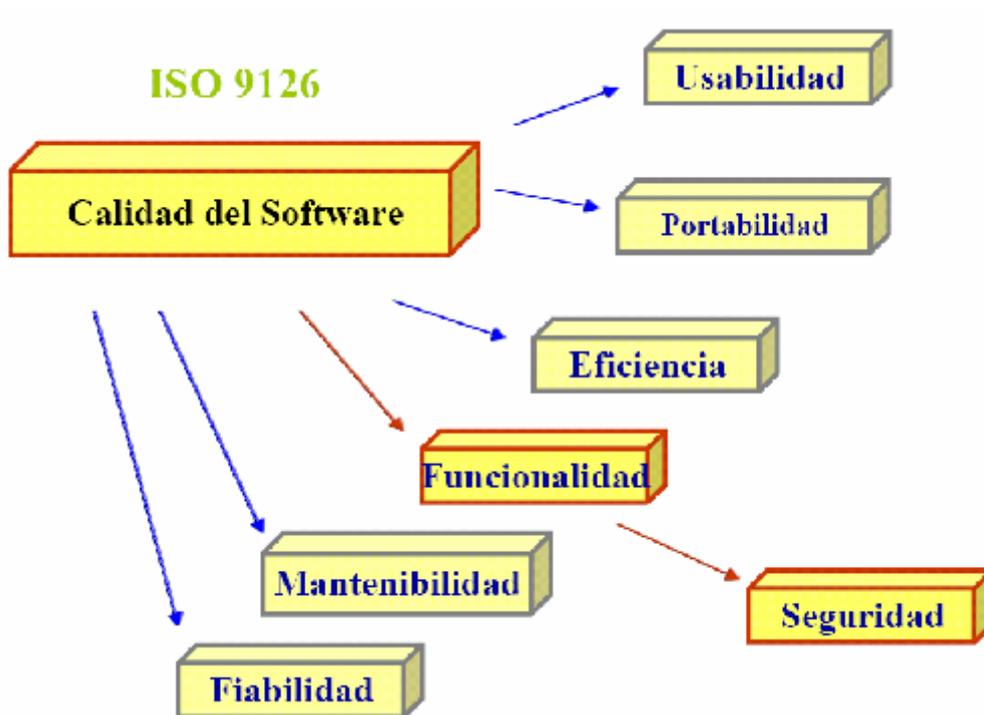
Como limitaciones deben señalarse la ausencia, en numerosas ocasiones, de información antigua, anterior a la creación de la base de datos, y la falta de bases de datos en áreas consideradas económicamente poco rentables.

No existe un único esquema de tarifas para la teledocumentación. La aplicación de tarifas puede variar entre los diferentes distribuidores, con motivo de lanzamiento de bases de datos o nuevos servicios, competitividad, etc. Incluso en ocasiones se realizan "ofertas" de conexión y utilización de bases de datos. Independientemente de esto, hay que tener en cuenta que los hosts suelen cargar en la factura del usuario como conceptos tiempo de conexión, número de ecuaciones ejecutadas, bases de datos utilizadas, número de documentos visualizados, términos consultados en los tesauros, uso de correo electrónico interno o espacio utilizado para el almacenamiento de ecuaciones de búsqueda. Esto supone que el documentalista debe controlar rigurosamente la utilización de los servicios del hosts, de forma que el coste de la búsqueda se adecue lo más posible a los resultados de información obtenidos.

9.- LA SEGURIDAD EN EL DISEÑO DE BASES DE DATOS MULTIMEDIA Y SISTEMAS DE INFORMACIÓN

9.1- Concepto de Seguridad.

Las bases de datos almacenan información importante en todos los ámbitos: comerciales, militares, médicos, administrativos, judiciales, etc. La información ha pasado a ser el activo más importante de las organizaciones, por encima incluso de los activos tradicionales (económicos, humanos y materias primas). Además, hay datos especiales (personales) que están protegidos por leyes en casi todos los países.



Seguridad es la capacidad de un producto software de proteger los datos y la información para que personas no autorizadas no puedan leerlos o modificarlos, y que el acceso no sea denegado a personal autorizado (ISO/IEC 1999b).

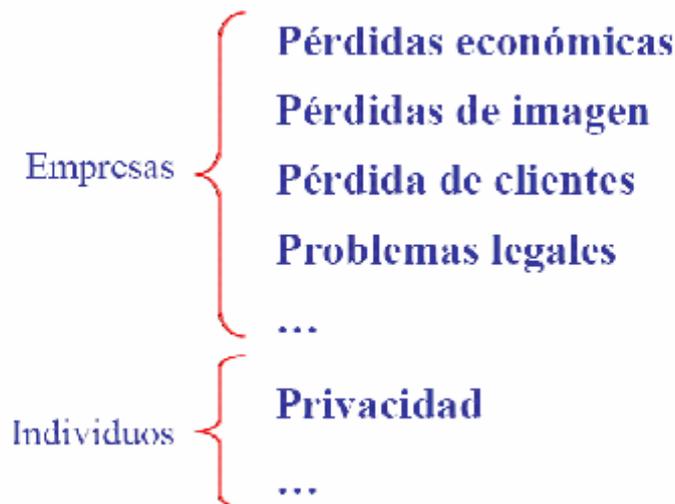
- **Confidencialidad:** prevenir / detectar / impedir el descubrimiento de información. En general la Confidencialidad se refiere a la protección de datos implicados en entornos altamente protegidos, como entornos militares, comerciales, etc. Privacidad se refiere a información sobre individuos. En la mayoría de los países la Privacidad está protegida por las leyes.
- **Integridad:** prevenir / detectar / impedir la modificación inadecuada de información. En un entorno militar, el mando responsable de un misil no debe ser modificado inadecuadamente. En un entorno comercial, la integridad de los datos es especialmente relevante, puesto que el éxito de una organización depende de lo correctas que son las operaciones que se llevan a cabo y la coherencia en los datos.
 - **Integridad semántica:** Respeto en todo momento de las reglas de integridad definida en la base de datos.
 - **Integridad Operacional:** Garantizar la consistencia de la base de datos con respecto al uso concurrente de la misma.
- **Disponibilidad:** prevenir / detectar / impedir la denegación inadecuada del acceso a servicios ofrecidos por el sistema. Por ejemplo, en un entorno militar, cuando el mando correspondiente da la orden de lanzar el misil, el misil es disparado. En entornos comerciales, las órdenes de

pago deben ser hechas en el momento. También relacionada disponibilidad con los mecanismos de recuperación de la base de datos ante caídas del sistema.

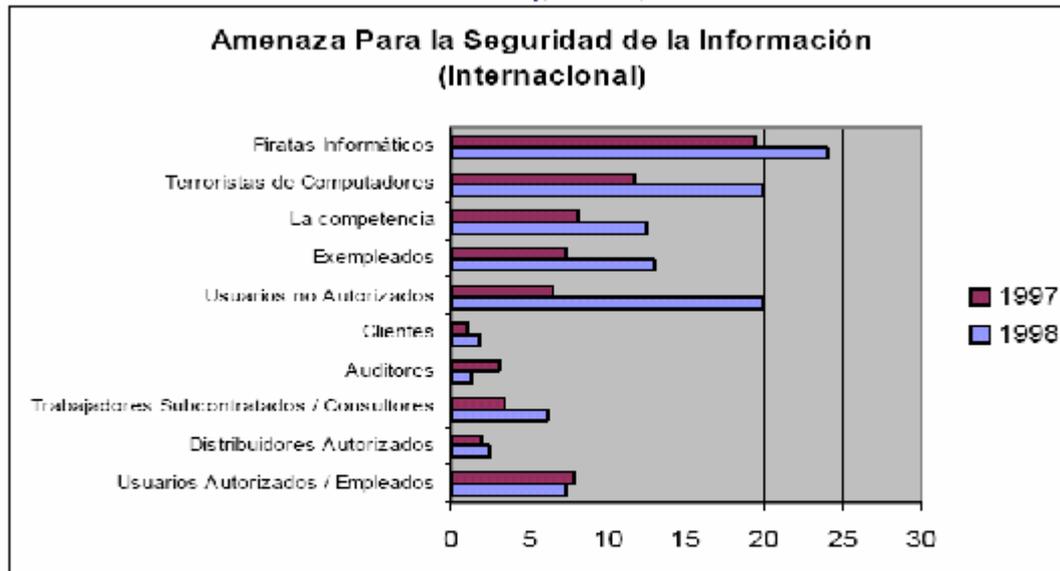
La protección de bases de datos puede obtenerse a través de medidas de seguridad para:

- **Control del flujo de información:** Regula la distribución de información entre objetos accesibles. Un flujo entre el objeto X y el objeto Y se produce cuando un sujeto lee el valor de X y escribe el valor en Y. Si el flujo de información no se controla puede dar lugar a situaciones de pérdida de confidencialidad o privacidad. El problema se acentúa con políticas de control de acceso discrecionales y obligatorias.
- **Control de inferencia:** Trata de proteger los datos de descubrimientos indirectos de información. Ocurre cuando un conjunto de items de datos X puede ser leído por un usuario, y puede ser usado para obtener el conjunto de datos Y de la forma $Y=f(X)$. Un ‘canal de inferencia’ es un canal donde los usuarios pueden encontrar un item X y entonces usarlo para derivar Y.
 - El problema de la inferencia estadística Implica la deducción de datos. Las bases de datos estadísticas permiten la consulta de datos estadísticos sobre grupos de individuos. En esas bases de datos, el acceso a datos sobre individuos individuales no está permitido, puesto que los datos son sólo accesibles a través de funciones estadísticas. Sin embargo, analizando datos estadísticos se pueden obtener datos sobre individuos concretos. Para hacer frente a este tipo de problemas se suelen utilizar dos tipos de controles:
 - **Control de consultas:** Restringen las consultas estadísticas que podrían revelar al usuario información confidencial
 - **Perturbaciones de datos.** Introducen algunos tipos de modificaciones durante el procesamiento de la consulta.
- **Control de acceso a la información:** Es el responsable de asegurar que todos los accesos directos a objetos del sistema se producen exclusivamente de acuerdo a los modos y reglas definidos por las políticas de protección definidas.

Para estos controles, las técnicas criptográficas pueden ayudar, pero no resuelven el problema. Las deficiencias en la seguridad pueden generar diferentes tipos de problemas:



En el siguiente gráfico, puede observarse alguna de las amenazas contra la seguridad en la información (ámbito internacional) (Ernst & Young, 1998)



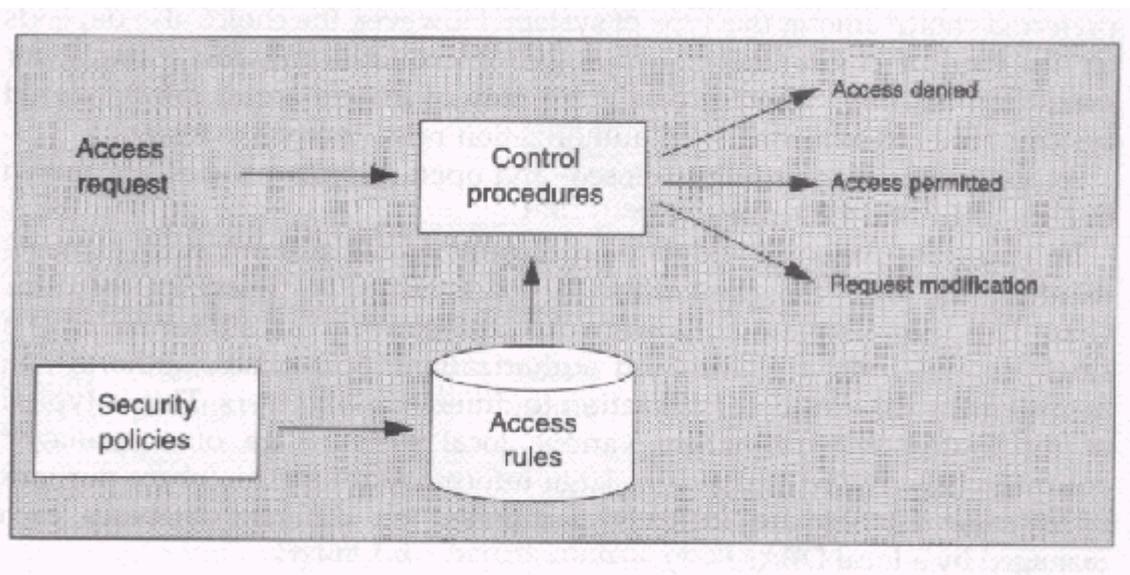
Por tanto, es labor de todos, usuarios y empresas y una debida legislación, para que la información contenida en cualquier base de datos, sea utilizada única y exclusivamente para el fin que fue recopilada y no con fines especulativos, económicos, etc.



Diversos autores (en varios foros internacionales) reclaman la integración de seguridad en el proceso de desarrollo de software (IEEE Software, Vol 19, N° 1, 2002) La seguridad está integrada ya por alguna de las más importantes SGBD's (Oracle9i Label Security).



Una buena manera de proteger una base de datos es mediante una correcta política de control de acceso, que es el mecanismo a través del cual intentamos asegurar que solamente los sujetos autorizados pueden acceder a los recursos del sistema de información.



9.2- Modelos de seguridad en Bases de Datos.

El objetivo de los modelos de seguridad es producir un modelo conceptual de alto nivel, independiente del software que describa las necesidades de protección del sistema. Un modelo de seguridad proporciona lo siguiente:

- Una representación semántica de las propiedades de seguridad del sistema.
- La facilidad a los desarrolladores de dar una definición de alto nivel de los requisitos de protección y las políticas del sistema, describiendo de manera precisa el comportamiento del sistema
- Demostrar que el sistema satisface algunas propiedades de seguridad.

Han sido propuestos un gran número de modelos de seguridad, y en general se podrían clasificar en dos categorías:

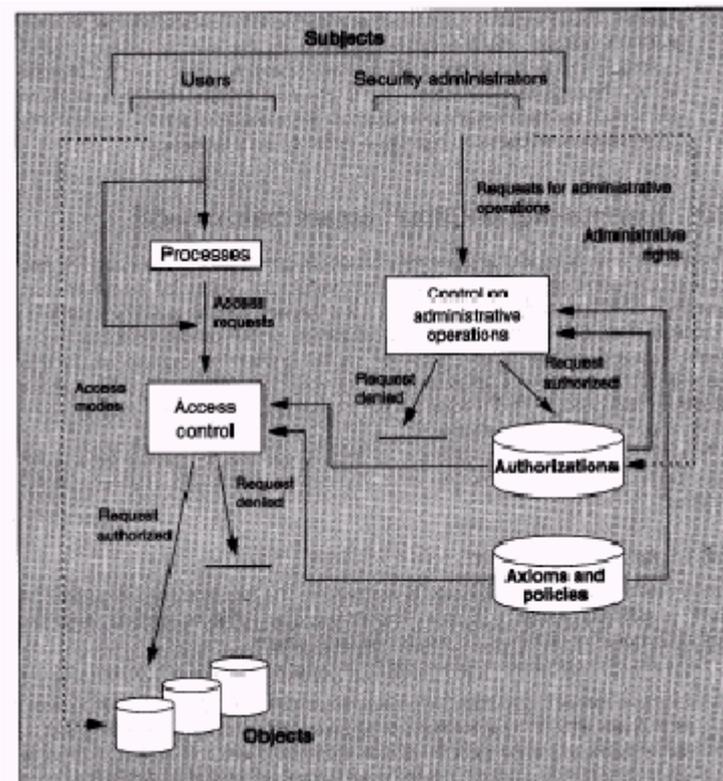
- Discrecionales
- Obligatorios

Todos estos modelos, tanto discrecionales como obligatorios se pueden subdividir atendiendo a los siguientes criterios:

- Elementos que se desean proteger (BD, SO, etc.)
- Tipo de política (Discrecional, Obligatoria)
- Enfoque de la política (confidencialidad / integridad)
- Tipo de control (acceso directo /acceso indirecto de información)

Al describir un modelo de seguridad, siempre aparecen los siguientes conceptos:

- **Sujetos:** Son las entidades activas del sistema que piden accesos sobre objetos.
- **Objetos:** Son entidades pasivas del sistema, que contienen información que debe ser protegida ante accesos o modificaciones no autorizados.
- **Modos de acceso:** Son los tipos de acceso que los sujetos pueden ejercitar sobre los objetos. Siempre que se produce un acceso, causa un flujo de información desde el objeto al sujeto y/o desde el sujeto al objeto.
- **Políticas:** Son las leyes que orquestan el control de acceso. Generalmente de alto nivel.
- **Autorizaciones:** Definen qué accesos podrán realizar los usuarios sobre los objetos y qué otros no.
- **Permisos administrativos** (implementadas por primitivas como 'grant', 'revoke', 'own'): permiten la modificación de autorizaciones: En sistemas discrecionales, la posesión de privilegios de administración en los sujetos, indica que pueden otorgar y revocar privilegios (autorizaciones) a otros sujetos.
- **Axiomas:** Son propiedades que deben ser satisfechas por el sistema.



9.3- Prototipos de Bases de Datos Seguras

Casi todas las marcas comerciales han desarrollado algún prototipo para implementar bases de datos seguras, pero sin tener ninguna repercusión en el mercado. Por ejemplo, alguno de estos prototipos son:

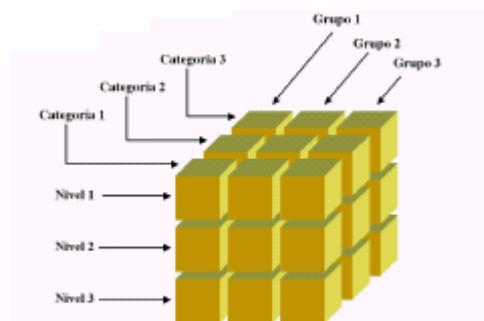
- TRUDATA
- Secure Sybase
- Trusted Oracle
- Trusted Informix

Recientemente ha aparecido un nuevo producto de Oracle en su versión 9 que parece prometedora: Oracle9i Label Security (OLS). Esta base:

- Implementa bases de datos multinivel
- Control de acceso basado en unas etiquetas para datos y usuarios.
- Control de acceso a nivel de fila.
- Capa virtual que gestiona las etiquetas y asegura el cumplimiento de las políticas de seguridad OLS.
- El control de acceso es mixto:
 - Discrecional; Mediante política de privilegios
 - Obligatoria; Mediante control de etiquetas

Para que un usuario tenga acceso sobre una determinada fila de una tabla, deben cumplirse las siguientes condiciones:

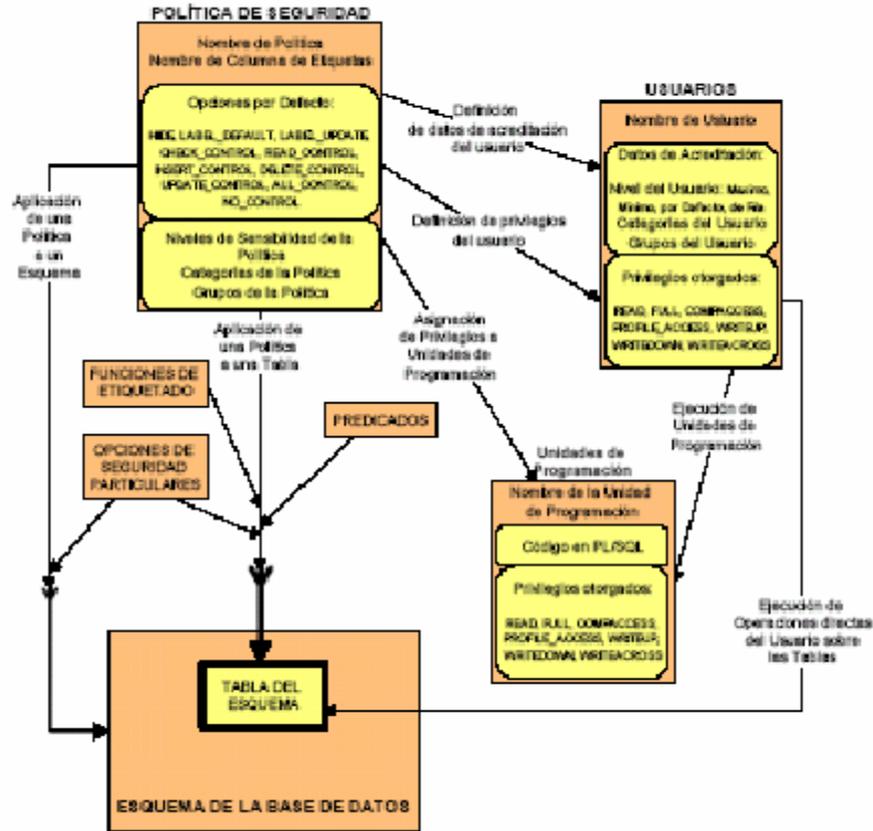
- Que el usuario esté autenticado por la base de datos.
- Que el usuario tenga los privilegios sobre la fila (CA discrecional).
- Que el usuario cumpla las reglas de acceso en función del nivel de sensibilidad del dato y del nivel de habilitación del usuario.



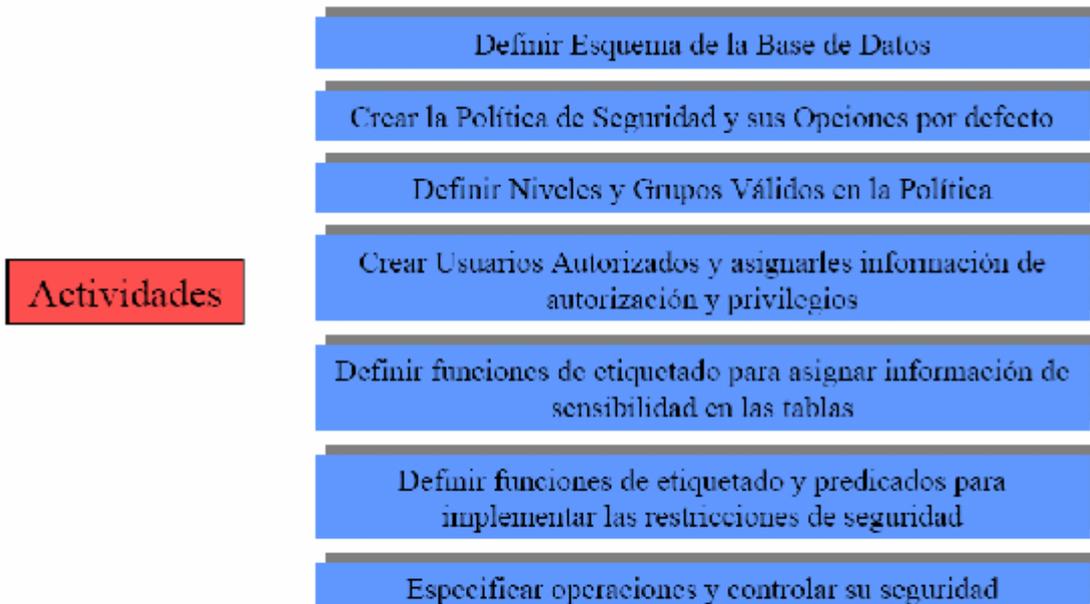
Componentes de las Etiquetas de Datos.

- Información del nivel de sensibilidad de la información.
- Distintas categorías organizacionales de información.
- Distintos grupos jerárquicos de información.

Toda esta información se almacena codificada en una columna de datos en cada tabla. El nivel de seguridad es un elemento que permite denotar la sensibilidad de la información etiquetada. La información será más sensible a medida que el nivel de seguridad sea más alto.



En la siguiente figura vemos reflejado el diseño lógico de una de las bases de datos más extendidas mundialmente (Oracle9i Label Security (OLS)).



10.- FUTURO DE LAS BASES DE DATOS

Las bases de datos se han convertido en un producto estratégico de primer orden debido a:

- Excepcional productividad
- Impresionante impacto económico
 - 8.000M \$ y 20% de crecimiento
- Grandes inversiones
 - Aplicaciones desarrolladas con SGBD
 - Investigación y Desarrollo
 - Marketing

A lo largo de la joven historia de los computadores, podemos establecer básicamente las siguientes generaciones de bases de datos:

- 1ª SGBD EN RED
- 2ª SGBD RELACIONALES
- 3ª SGBD INTELIGENTES, ACTIVOS,

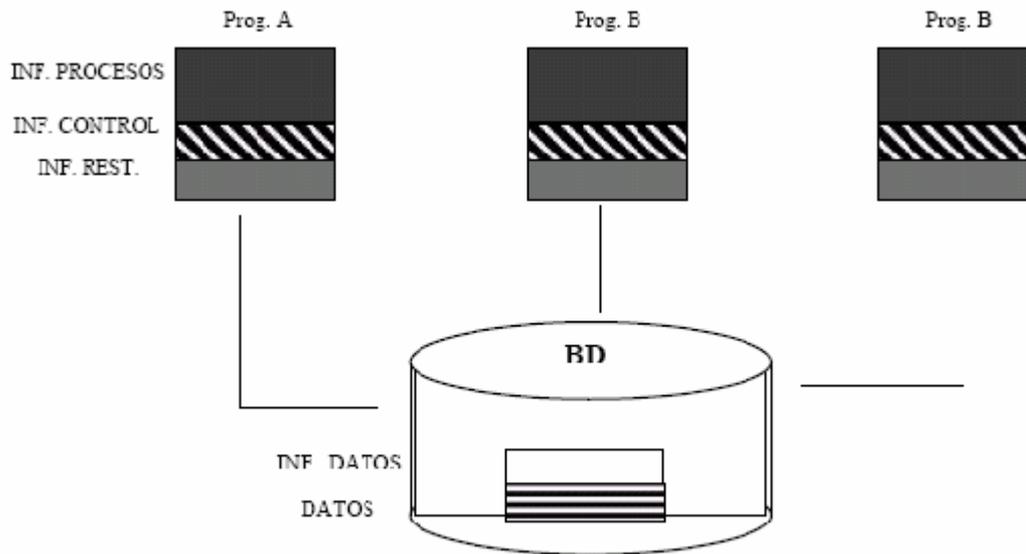
Esta tercera generación de bases de datos vista por dos autores diferentes nos dice que:

- **Cattell (1991):** “Proporciona capacidades de gestión de datos al igual que sus predecesoras, permitiendo que grandes cantidades de datos persistentes sean compartidos por muchos usuarios. También proporcionan gestión de objetos, permitiendo tipos de datos muchos más complejos, objetos multimedia, datos derivados, encapsulamiento de la semántica de los datos, así como otras nuevas capacidades. Algunos proporcionan incluso gestión de conocimiento, soportando un gran número de reglas complejas para inferencia automática de información y mantener las restricciones de integridad entre datos”
- **Khoshafian et al. (1990):** “Gestiona información de una manera natural, haciendo esta información fácil de almacenar, acceder y utilizar. Es necesario:
 - Un fuerte soporte para texto, imagen, voz, animación y vídeo
 - Un modelo de bases de datos orientado al objeto
 - Un soporte de reglas declarativas para expresar las interrelaciones semánticas entre objetos”

Según Buchmann (1996) y Zachman (1996) todavía queda mucho por mejorar en el campo de las bases de datos, ya que:

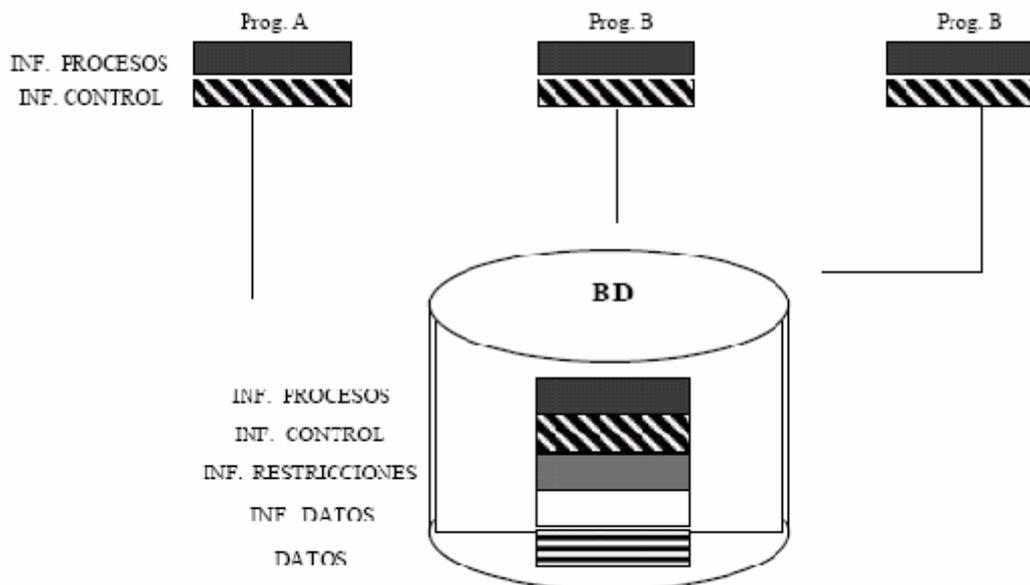
- Los SGBD son monolíticos
- Existen más datos en hojas de cálculo que en SGBD
- El 50% de los datos en producción están en sistemas heredados (legacy systems)
- Muchas aplicaciones no necesitan más que ficheros
- Los servicios de réplica no escalan por encima de los 10.000 nodos
- No se puede combinar datos estructurados y no estructurados
- El 70% de los informes se utilizan para volver a introducir datos en bases de datos
- 80/90% del tiempo se dedica al mantenimiento
- 60/70% se dedica a mantener interfaces y transformar datos de una base a otra
- 100M\$ a “parchear” roas. Por errores al pasar datos
- 70% de las líneas de código sólo mueven datos
- 20/40% de toda la mano de obra de EEUU se dedica a recoger, almacenar, recuperar, ... información

La siguiente figura muestra el modo de gestión de la información de una base de datos clásica.



Bases de datos clásicas

Y en la siguiente figura, vemos como se gestionan los datos en las bases de datos orientadas a objetos (cuyo principal exponente son las bases de datos multimedia).



Bases de datos orientadas a objetos

Según sea el tipo de objetivo que se busca obtener, al diseñar y utilizar una base de datos, la siguiente clasificación nos muestra donde podemos ubicar las diferentes bases de datos actuales.

- RENDIMIENTO
 - BD PARALELAS
 - BD EN TIEMPO REAL
 - BD EN MEMORIA PRINCIPAL
- DISTRIBUCIÓN
 - BD DISTRIBUIDAS
 - BD FEDERADAS
 - MULTIBASES DE DATOS
 - BD MÓVILES
 - BD Y “WEB”
- INTELIGENCIA
 - BD ACTIVAS
 - BD DEDUCTIVAS
 - BD ORIENTADAS A OBJETOS
 - BD MULTIMEDIA
 - BD TEMPORALES
 - BD SEGURAS
 - BD DIFUSAS
 - ALMACENES DE DATOS

Entre las principales aplicaciones prácticas de las bases de datos actuales y en el futuro encontramos las siguientes:

- CASE
- CAD/CAM/CIM
- SIG
- APLICACIONES CIENTÍFICAS
- SISTEMAS MÉDICOS
- PUBLICACIÓN DIGITAL
- EDUCACIÓN
- SISTEMAS ESTADÍSTICOS
- COMERCIO ELECTRÓNICO

11.- SISTEMAS GESTORES DE BASES DE DATOS EXISTENTES EN EL MERCADO

Para superar los límites de los sistemas de ficheros surgieron los denominados Sistemas Gestores de Bases de Datos (SGBD). Estas aplicaciones daban cabida a múltiples sistemas reales de datos (múltiples bases de datos). Esto se debe a que no solo almacenaban los datos es sí, sino que además almacenaban la definición lógica de la base de datos (esquemas y subesquemas), así como la definición de los usuarios y sus respectivos permisos de acceso.

En el ámbito físico presentaban la ventaja esencial de acceso simultáneo a varios ficheros, lo que supone una gran mejora respecto a los tiempos de búsqueda y respuesta. A nivel lógico, la nueva y estructurada organización de estos sistemas, apoyados en el modelo relacional de bases de datos, permiten que su gestión sea posible mediante lenguajes de muy alto nivel y de cuarta generación.

En la práctica los SGBD se presentan en compactos entornos de desarrollo que facilitan enormemente el diseño y creación de base de datos. Podemos dividir el grueso de estos programas en dos grandes bloques:

- SGBD Corporativos: Están orientados a la implementación de un gran volumen de información y a proveer accesos a ésta a través de redes (tanto de área extensa como locales). Son aplicaciones muy extensas que requieren plataformas específicas y dedicadas, lo cual las hace exclusivas de grandes instituciones.
- SGBD Personales: Están destinados a ofrecer solución a las necesidades domésticas de almacenamiento de información, así como gestión para Pymes. En la actualidad estos sistemas están muy extendidos, por ser muy flexibles y fáciles de utilizar.

A continuación vamos a comentar algunos de los sistemas gestores de bases de datos más utilizados en la actualidad por los diferentes tipos de usuarios del mercado.

11.1- DB2

Este gestor de base de datos ha evolucionado de manera considerable. Podríamos decir que esta base de datos relacional es de corte moderno, con una interfaz bastante intuitiva. Entre las mejoras de este producto tenemos algunas cuantas generalidades respecto a rendimiento, capacidad y memoria. Pero lo más llamativo es la posibilidad de aprovechar multiprocesadores simétricos (SMP) compartiendo recursos entre ellos.

De esta forma las consultas SQL pueden ser ejecutadas más rápido. El lenguaje SQL que soporta es el standard pero con algunas mejoras. Podemos destacar el OLAP (proceso analítico en línea) ayudado de la posibilidad de creación de supergrupos. También se pueden realizar uniones externas, manipular datos, disponer de un soporte adicional de actualización, y definir ciertas restricciones de accesibilidad en tablas.

En este programa también se incorpora un nuevo tipo de dato llamado datalink capaz de referencia a un objeto que se encuentra fuera de la base de datos. Una adecuada gestión de campos permite el almacenamiento de datos de gran tamaño soporte directo de objetos multimedia (imagen, vídeo, etc)

La escalabilidad se obtiene gracias a la posibilidad de particionar la base de datos a lo largo de múltiples sistemas independientes conectados a través de un LAN. Internet es siempre la gran estrella, con DB2 es posible acceder a los datos usando JDBC (tan potente como escribir directamente C contra la base de datos), Java y SQL (tanto el SQL estático, como complementa el SQL dinámico).

También es posible efectuar las tareas de administración a través de Internet y ver los manuales. Igualmente proporciona una interfaz de programación para PERL, para ciertas plataformas. El programa encargado de las conexiones recibe el nombre de DB2- Connect, que tradicionalmente ha utilizado el protocolo SNA, para conectar con los servidores, aunque ahora se ha añadido el TCP/IP. De esta forma se ha popularizado mucho más la herramienta, disminuyendo los requerimientos de prestaciones de los clientes.

Los servidores DB2 soportan cualquier comunicación basadas en los protocolos APPC, IPX/SPX, NetBios, TCP/IP, Pipes con nombres. Las plataformas que admite el producto son en principio OS/2, Windows NT y UNIX. Esta última versión del software del servidor, ha sido adaptado para funcionar también con OS/2, WARP, HP- UX, AIX, SCO- UNIX UE7, Windows 98, NT y Solaris.

Esta nueva versión persigue inclinar la balanza hacia un producto que ha ido ganado adeptos de forma espectacular desde aquella primera versión con tecnología de Sybase. Sobre todo, por que a parte de unas consideraciones técnicas tiene todo el empuje comercial y de soporte de Microsoft. Se podría decir que en cierta medida pretende ser el servidor de bases de datos genérico para Windows. No tanto por que la causa de desarrolla sea la misma, ni siquiera porque el SQL Server, a diferencia de otros servidores solo trabaja bajo Windows, sino porque Microsoft promete integración con todos los productos suyos (por ejemplo MsOffice 2000, ya que Access 2000 traerá consigo un nuevo MSDE- DATA- Engine, como alternativa al existente y compatible con SQL Server). También será posible llamar a SQL Server desde MsAccess. Otras características son los servicios OLAP integrados en el Paquete.

La escalabilidad es total. No es necesario decir que el producto puede funcionar en un Servidor NT multiprocesador de elevadas prestaciones, pero si lo es decir que también puede funcionar en un portátil con Windows 95/98. Las características de SQL Server son impresionantes. Tenemos soporte de transacciones OLPT, una maquinaria de búsqueda de textos completos que permite localizar información a lo largo de una tabla (lo que hace las delicias si el proyecto a considerar es para Internet). Posibilidad de ejecutar consultas en paralelo, así como homogéneas y distribuidas. Bloqueos dinámicos a nivel de filas, optimizador de consultas, estadísticas Automáticas, Unicode Nativo, Replicación Avanzada, Replicación dinámica de datos y un largo etcétera. Pero lo mejor de todo es la sencillez de comprensión y navegación entre procesos así como lo intuitivo de la herramienta.

Alrededor de 30 asistentes pueden conseguir que un usuario sin ninguna experiencia pueda realizar multitud de opciones, o facilitar la vida de los experimentados. Unas herramientas auxiliares como el Query Profiler o Query Analyzer, habilitan colecciones y análisis de bases de datos, junto con la actividad de las consultas. Los protocolos de red soportados con garantía d funcionamiento son TCP/IP, Pipes con nombre, IPX/SPX, AppleTalk ADSP y Banyanvines.

11.2- INFORMIX

Al igual que Oracle, la presencia de Informix se deja sentir en una cantidad de plataformas, concretamente Windows NT y UNIX. Una de las características de este sistema es un completo conjunto de herramientas gráficas que permiten asistir tanto en los procesos de instalación como administración del servidor, sin necesidad de que el administrador tenga grandes conocimientos. Existe la posibilidad de gestionar múltiples bases de datos remotas de una única y centralizada consola donde se muestran gráficamente tanto la base de datos como los objetos que contiene (tablas índices, procedimientos, etc...) También podrá establecerse un calendario de tareas a ejecutar en cualquier objeto o grupo de objetos. Adicionalmente Informix proporciona tablas que forma el SMI (interfaz de monitorización del sistema).

Con una arquitectura que denominan DSA (Dynamic Scalable Architecture) que afecta a todas las plataformas se tiene como objetivo una avanzadas escalabilidad en las bases de datos y un alto rendimiento en cualquier entorno. Una de las características que se encuentra implementada para WinNT son:

- La implementación de procesadores virtuales
- Manipulación directa de acceso a disco
- Uso de la multitarea preventible propia del sistema
- Implementación como si fuera un servicio más
- Integración con el registro del sistema
- Utilización del servicio de eventos NT para alertas, etc...

Esta arquitectura es el núcleo de Informix, y tiene la capacidad de optimizar el rendimiento de una forma inteligente gracias a la ejecución en paralelo de las principales operaciones de la base de datos (particularmente, consulta o compactación de índices, y también salvaguarda y recuperación de datos).

Múltiples hilos de ejecución y un pool de procesadores virtuales atienden en paralelo a los usuarios concurrentes, consiguiendo que todo el rendimiento (sobre todo si tiene más de un procesador) se aproveche repartiéndolos según prioridades. Estas propiedades son entre las mismas tareas. La idea es conseguir una escalabilidad total y lineal. De forma que ante más y más usuarios podamos añadir más recursos de procesamiento, sin grandes cambios. Igualmente si crece el tamaño de la base de datos o de los sites donde se ubica, también podremos obtener rápidas y eficientes soluciones.

En este sentido Informix tiene la capacidad de relación de datos en múltiples lugares físicos (dentro de la citada arquitectura DSA), suponiendo un eficaz sistema de distribución y compartición de datos dentro de un mismo sistema organizativo. Para ello utiliza una técnica de publicar y suscribir donde el propietario de la información publica la misma siendo republicado por los subscriptores de múltiples sites. La idea es dar suficiente flexibilidad para que puedan definir el conjunto de datos que desean capturar, en vez de recoger siempre la totalidad de los mismos.

Internet no ha sido olvidada e Informix incluye de forma integrada el servidor Netscape FaxTrap, para soporte y conectividad de aplicaciones Web e Intranet. Como características esenciales tenemos servicios escalables Web/Inet de alto rendimiento, entorno de desarrollo basado en Java/JavaScript seguridad basada en SSL con autenticación. Encriptación e integridad de mensajes, control de accesos, soporte de certificados del lado del cliente y asistentes para la publicación en Internet.

Además de los expuestos Informix ofrece una componente denominado Web Integration Option que facilita la generación de Aplicaciones orientadas a Internet. Se trata de un Runtime que enlaza al servidor Web y el servidor de base de datos de forma que las aplicaciones puedan incorporar sencilla los datos suministrados dinámicamente suministrados por Informix. También es posible incorporar una consulta SQL dentro de una página Web. El WIO incluye al WebDB Publisher cuya misión es la de generar informes mediante asistentes, es decir una forma automática para generar páginas a partir de datos de una base de datos.

Los bloqueos se pueden efectuar a todos los niveles posibles. Los niveles de aislamiento son completos, soportando DirtyMan, CommittedRead, CursorStability and RepeatableTableRead. El lenguaje sigue las normas del standard ANSI SQL-92, al igual que también se cumple el standard X-Open. La seguridad e integridad de los datos es uno de los objetivos de Informix, incluyendo duplicación, espejo de disco, volcado de seguridad y restauración de alta velocidad transacciones en línea y un diseño de seguridad conforme a la clasificación C2, OLTP y DataWarehousing están disponibles.

11.3- InterBase

Sencillez y potencia son las características de Interbase que reúnen muchas características de los productos que ya hemos visto. Su arquitectura multigeneracional permite las transacciones críticas manteniendo concurrencia en los mismos cuando se hace una operación mixta, como puede ser la consulta y actualización de un entorno determinado. Esto se consigue creando múltiples versiones de cada registro de datos, dentro de una transacción, un usuario puede estar modificando un registro y otro leyendo simultáneamente, ya que se trata de versiones distintas.

Para procesos críticos se pueden usar transacciones que aíslan un grupo de cambios de otros. El bloqueo aplicado cuando un cliente actualiza datos sin utilizar transacciones se hace a nivel de registro exclusivamente. Esto permite que varios clientes alteren la información de una misma tabla, a excepción de cuando coincidan en un mismo punto, donde se produce una serialización.

La herramienta ISQL (Interactive SQL) utiliza una interfaz gráfica para introducir sentencias SQL y de forma sencilla. La sentencia introducida es enviada al servidor, preparada y ejecutada, recibiendo la posible respuesta del mismo. Windows SQL puede manejar transacciones, mostrar meta-datos y producir y ejecutar fragmentos de código que contengan sentencias SQL.

En lo que respecta a SQL, el lenguaje está basado en el estándar ANSI-SQL-92. Es posible definir funciones a medida UDF. La gestión del servidor está basada en un elemento llamado Guardian. Normalmente es un proceso silencioso que arranca automáticamente y ofrece sus servicios, pero dispone de opciones manuales.

Para poder efectuar las conexiones cliente servidor, además del necesario TCP/IP, Interbase soporta el protocolo NetBEUI cuando se trata de servidores NT y clientes Windows, así como un modo de conexión local que no precisa de interfaces de red. La versión 5.5 puede soportar IPX/SPX.

Interbase proporciona dos elementos específicos para Internet. Internet Client es un driver de red para desarrollos en Java, que incorpora un protocolo remoto JDBC, para intercambio de datos entre cliente y servidor. Esto significa que el navegador del cliente no precisa de librerías preinstaladas para poder acceder a través a los datos de Interbase. Por el otro lado tenemos Inserver, bajo NT o UNIX que es el que opera del lado del servidor que comunica con Intertclient y permite las operaciones a Interbase.

Con el fin de acceder de forma más Exhaustiva a las funcionalidad de Interbase es posible acceder a una colección de funciones (o API). Mediante estas funciones se pueden enviar sentencias SQL a la maquinaria de base de datos, recuperando el resultado. Este proceso puede parecer muy similar al empleado integrando SQL en aplicaciones, pero en realidad tiene una serie de ventajas, entre las que destaca el control sobre la gestión de memoria, el acceso a los mensaje de error y a los manipuladores de transacciones y opciones. Para acceder a este API hemos de usar un compilador de C, como el Visual C++ de Microsoft.

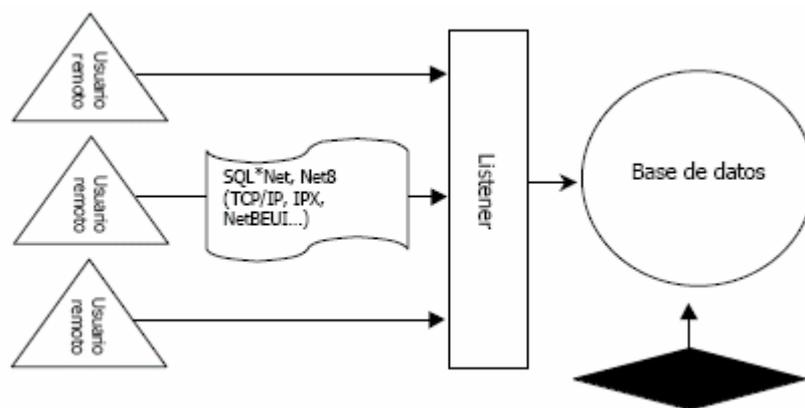
Interbase soporta handles como opción a los campos GLOB, así como matrices de tipos de datos incluso multidimensionales. Es posible definir funciones a medida, de campos BLOB, llamadas UDF, de forma que se realicen tareas específicas en formato binario, ilegibles por el sistema.

Otra de las características añadidas a Interbase son las tablas estadísticas relativas a los accesos de Datos.

11.4- ORACLE

A pesar de llevar ya algún tiempo en el mercado la versión 8 de Oracle, sigue junto con SQL Server, liderando el mercado NT. Además de Windows, el servidor de Oracle puede funcionar en una gran cantidad de sistemas operativos y diversidad de Hardware. Prácticamente tenemos a todas las familias de UNIX, MVS, VM, HPMPEXL, Siemens ICL, Novell Netware y OS/2.

Además el funcionamiento está optimizado para ajustarse a las peculiaridades de cada sistema operativo, Oracle se ajusta al Modelo de ejecución de Hilos o procesos que aporta el sistema. Dentro de la idea de Oracle de potenciar los grupos de trabajo distribuido se necesita un acceso a los datos de dicho grupo.



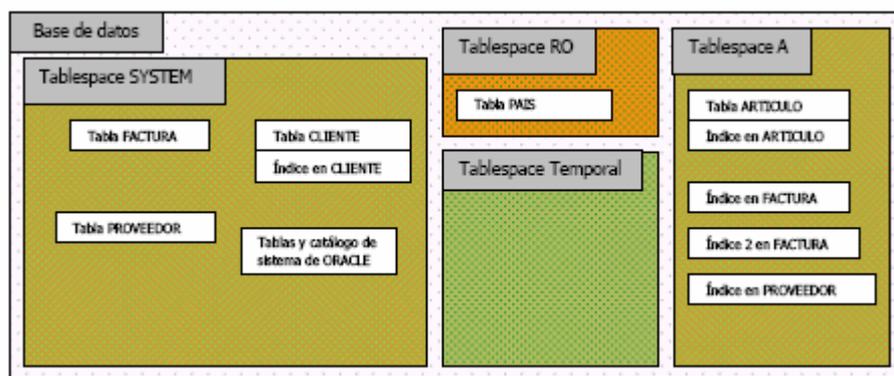
Estructura básica de toda base de Datos.

En este sentido se posibilitan consulta y actualizaciones distribuidas permitiendo la compartición de datos a través de múltiples servidores, con una consistencia de los datos en un comit de dos fases (proceso que resulta transparente al usuario). Igualmente se proporciona replicación de los datos con las ventajas anteriormente enunciadas y enlaces a Bases de Datos, de forma que permiten a los datos remotos ser definidos y usados como si fueran locales.

La versión 8 de la base de datos Oracle incluye una herramienta de administración gráfica que es mucho más intuitiva y cómoda de utilizar. Se emplea en forma alternativa a los comandos de línea de texto que se usan para efectuar administración. Por lo tanto, no es necesario disponer de esta herramienta en forma obligatoria porque siempre es posible administrar una base de datos desde la línea de comandos, sin embargo, es mucho más recomendable por la facilidad de uso y rapidez para efectuar la mayoría de los comandos.

La base de datos de Oracle tiene una capa lógica y otra física. La capa física consiste de archivos que residen en el disco y los componentes de la capa lógica son estructuras que mapean los datos hacia estos componentes físicos.

- La Capa Física: Consta de archivos físicos que se encuentran en los discos. Estos pueden ser de tres tipos diferentes:
 - Uno o más datafiles: Los datafiles almacenan toda la información ingresada en una base de datos. Se pueden tener sólo uno o cientos de ellos. Muchos objetos (tablas, índices) pueden compartir varios datafiles. El número máximo de datafiles que pueden ser configurados está limitado por el parámetro de sistema MAXDATAFILES.
 - Dos o más archivos redo log (de deshacer): Los archivos del tipo redo log almacenan información que se utiliza para la recuperación de una base de datos en caso de falla. Estos archivos almacenan la historia de cambios efectuados sobre la base de datos y son particularmente útiles cuando se necesita corroborar si los cambios que la base de datos ya ha confirmado se han efectuado realmente en los datafiles.
 - Uno o más control files: Estos archivos contienen información que se utiliza cuando se levanta una instancia, tal como la información de dónde se encuentran ubicados los datafiles y los archivos redo log. Estos archivos de control deben encontrarse siempre protegidos.
- La Capa Lógica: La capa lógica de una base de datos consta de los siguientes elementos:
 - Uno o más tablespaces: El esquema de la base de datos (schema), el cual consiste de objetos como tablas, clusters, índices, vistas, procedimientos almacenados, triggers, secuencias y otros.
 - Los Tablespaces y los Datafiles: Una base de datos se encuentra dividida en una o más piezas lógicas llamadas tablespaces, que son utilizados para separar la información en grupos y así simplificar la administración de los datos. Los tablespaces pueden ocupar uno o más datafiles. Si se decide que utilice varios datafiles, el administrador del sistema puede gestionar que éstos queden localizados en discos diferentes, lo que aumentará el rendimiento del sistema, principalmente por la mejora en la distribución de la carga de entrada / salida.



En este esquema podemos ver que, por ejemplo, la tabla ARTICULO se almacena dentro del tablespace A, y que por lo tanto tendrá todas las propiedades del tablespace A que pueden ser:

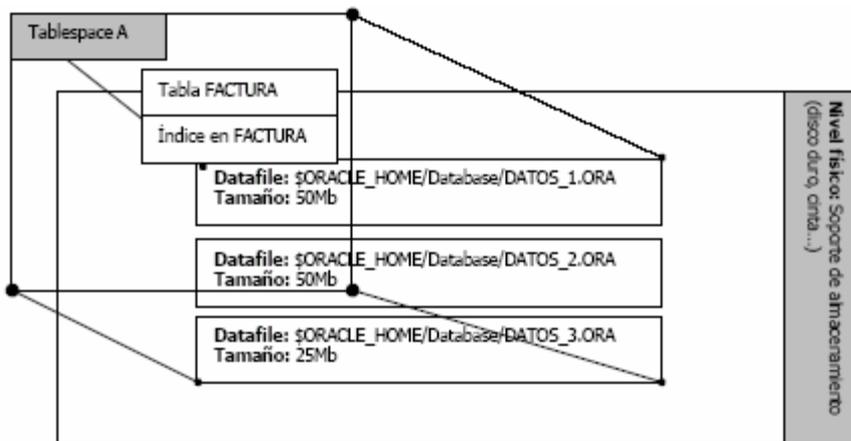
- Sus ficheros de datos están en \$ORACLE_HOME/datos/datos_tablespace_A
- Los objetos no pueden ocupar más de 10Mb de espacio de base de datos.
- En cualquier momento se puede poner fuera de línea todos los objeto de un cierto tablespace.
- Se pueden hacer copiar de seguridad sólo de ciertos tablespaces.

Si nos fijamos, se puede apreciar que es posible tener una tabla en un tablespace, y los índices de esa tabla en otro tablespace. Esto es debido a que los índices no son más que objetos independientes dentro de la base de datos, como lo son las tablas. Y al ser objetos independientes, pueden ir en tablespaces independientes.

El tablespace SYSTEM es un tablespace por defecto en todas las bases de datos Oracle. En él se almacenan todos los datos de sistema, el catálogo y todo el código fuente y compilado de procedimientos PL/SQL. También es posible utilizar el mismo tablespace para guardar datos de usuario.

En el esquema también vemos que hay un tablespace Temporal (en gris oscuro). Este tablespace representa las propiedades que tendrán los objetos que la base de datos cree temporalmente para sus cálculos internos (normalmente para ordenaciones y agrupaciones). La creación de un tablespace temporal difiere en una de sus cláusulas de creación.

El tablespace RO (en gris claro) difiere de los demás en que es un tablespace de solo lectura (Read Only), y que por lo tanto todos los objetos en él contenidos pueden recibir órdenes de consulta de datos, pero no de modificación de datos. Estos tablespaces puede residir en soportes de sólo lectura, como pueden ser CDROMs, DVDs, etc. Cuando se crea un tablespace, éste se crea de lectura/escritura. Después se puede modificar para que sea de solo lectura.



En el esquema podemos ver como el Tablespace A está compuesto (físicamente) por tres datafiles (DATOS_1.ORA, DATOS_2.ORA y DATOS_3.ORA). Estos tres datafiles son los ficheros físicos que soportan los objetos contenidos dentro del tablespace A.

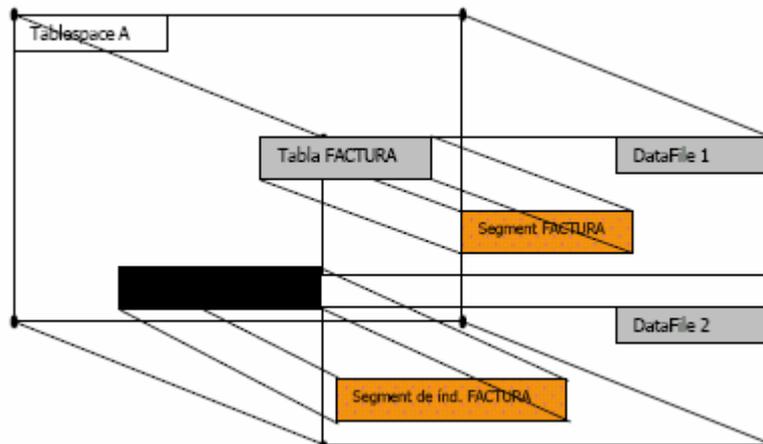
Aunque siempre se dice que los objetos están dentro del tablespace, en realidad las tablas están dentro del datafile, pero tienen las propiedades asociadas al tablespace. Cada uno de los datafiles utilizados está ocupando su tamaño en disco (50 Mb los dos primeros y 25 Mb el último) aunque en realidad sólo contengan dos objetos y estos objetos no llenen el espacio que está asignado para los datafiles.

Los datafiles tienen una propiedad llamada AUTOEXTEND, que si está activa se encarga de que el datafile crezca automáticamente (según un tamaño indicado) cada vez que se necesite espacio y no exista.

Dentro de los tablespaces y datafiles, el espacio utilizado para almacenar datos es controlado por el uso de ciertas estructuras; éstas son las siguientes:

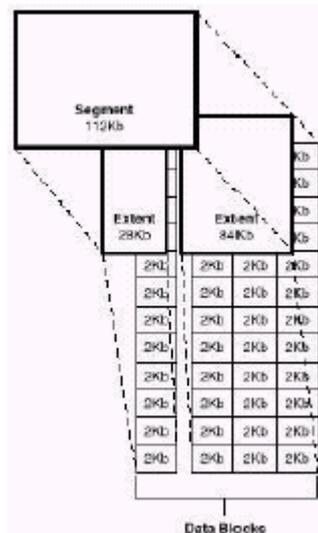
- Bloques: Un bloque es la unidad de almacenamiento más pequeña en una base de datos Oracle. Contiene una pequeña porción de información (header) referente al bloque en sí y el resto a los datos que guarda. Generalmente, un bloque de datos ocupará aprox. 2 KB de espacio físico en el disco (asignación típica).
- Extensiones: Es un grupo de bloques de datos. Se establecen en un tamaño fijo y crecen a medida que van almacenando más datos. También se pueden redimensionar para aprovechar mejor el espacio de almacenamiento.

- Segmentos: Es un grupo de extensiones utilizados para almacenar un tipo particular de datos. Existen 4 tipos de segmentos: datos, índices, rollback y temporales.



Podemos ver como el espacio que realmente se ocupa dentro del datafile es el segment y que cada segment pertenece a un objeto. Existen tres tipos de segments (principalmente):

- Segmentos de tipo TABLE: aquellos que contienen tablas
- Segmentos de tipo INDEX: aquellos que contienen índices
- Segmentos de tipo ROLLBACK: aquellos se usan para almacenar información de la transacción activa.



El esquema muestra toda la cadena de almacenamiento de Oracle. Desde el nivel más físico al más lógico:

- Unidades de asignación del SO (El más físico. No depende de Oracle)
- Data blocks de Oracle
- Extents
- Segments
- DataFiles
- Tablespaces (El más lógico)

El tamaño de las unidades de asignación del SO se define durante el particionado del disco duro (FDISK, FIPS...), y el espacio de los data blocks de Oracle se define durante la instalación y no puede ser cambiado.

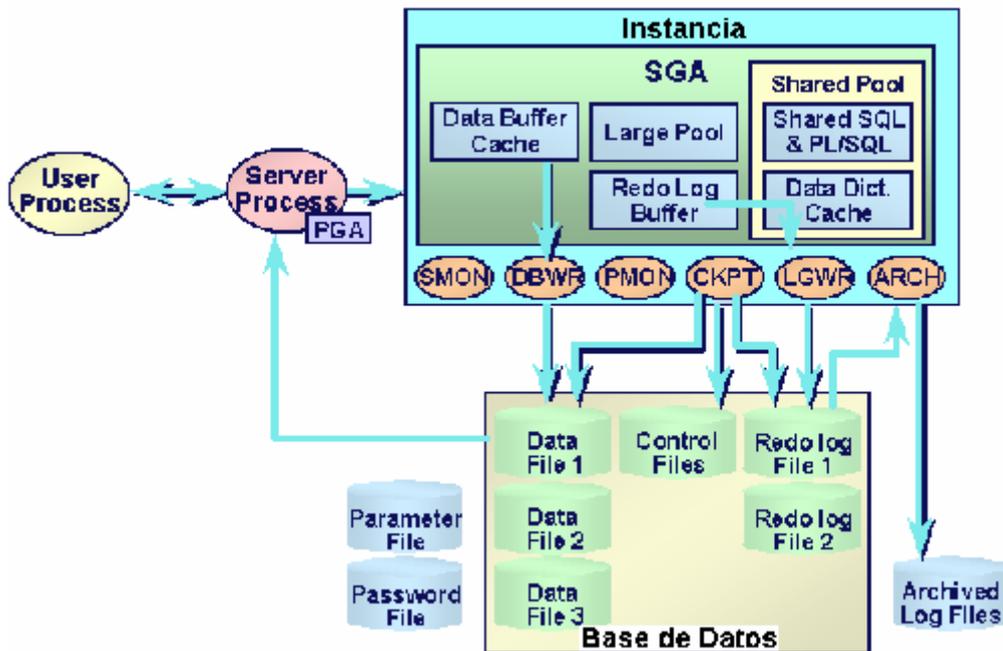
Como es lógico, el tamaño de un data block tiene que ser múltiplo del tamaño de una unidad de asignación, es decir, si cada unidad de asignación ocupa 4 K, los data blocks pueden ser de 4K, 8K, 12K... para que en el SO ocupen 1, 2, 3... unidades de asignación.

Un esquema es una colección de objetos lógicos, utilizados para organizar de manera más comprensible la información y conocidos como objetos del esquema. Una breve descripción de los objetos que lo componen es la siguiente:

- **Tabla:** Es la unidad lógica básica de almacenamiento. Contiene filas y columnas (como una matriz) y se identifica por un nombre. Las columnas también tienen un nombre y deben especificar un tipo de datos. Una tabla se guarda dentro de un tablespace (o varios, en el caso de las tablas particionadas).
- **Cluster:** Un cluster es un grupo de tablas almacenadas en conjunto físicamente como una sola tabla que comparten una columna en común. Si a menudo se necesita recuperar datos de dos o más tablas basado en un valor de la columna que tienen en común, entonces es más eficiente organizarlas como un cluster, ya que la información podrá ser recuperada en una menor cantidad de operaciones de lectura realizadas sobre el disco.
- **Índice:** Un índice es una estructura creada para ayudar a recuperar datos de una manera más rápida y eficiente. Un índice se crea sobre una o varias columnas de una misma tabla. De esta manera, cuando se solicita recuperar datos de ella mediante alguna condición de búsqueda (cláusula where de la sentencia), ésta se puede acelerar si se dispone de algún índice sobre las columnas-objetivo.
- **Vista:** Una vista implementa una selección de varias columnas de una o diferentes tablas. Una vista no almacena datos; sólo los presenta en forma dinámica. Se utilizan para simplificar la visión del usuario sobre un conjunto de tablas, haciendo transparente para él la forma de obtención de los datos.
- **Proced. Almacenado:** Son programas que permiten independizar el manejo de datos desde una aplicación y efectuarla directamente desde el motor de base de datos, disminuyendo así el tráfico de información a través de la red y mejorando el rendimiento de los procesos implementados mediante estos programas.
- **Trigger:** Un trigger es un procedimiento que se ejecuta en forma inmediata cuando ocurre un evento especial. Estos eventos sólo pueden ser la inserción, actualización o eliminación de datos de una tabla.
- **Secuencias:** El generador de secuencias de Oracle se utiliza para generar números únicos y utilizarlos, por ejemplo, como claves de tablas. La principal ventaja es que libera al programador de obtener números secuenciales que no se repitan con los que pueda generar otro usuario en un instante determinado.

11.4.1- Arquitectura de Oracle

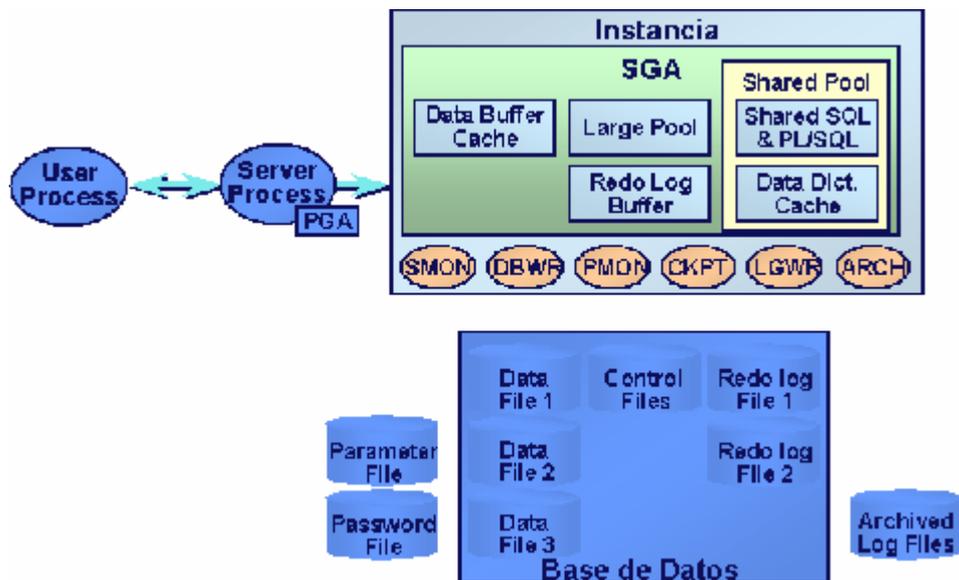
En la siguiente figura observamos como está jerarquizada la estructura de la base de datos Oracle.



La Arquitectura general de Oracle consiste de varios procesos corriendo en la máquina donde reside la instancia, más los espacios de memoria dedicados a ejecutar procesos específicos o al almacenaje de información de cada proceso y la base de datos física propiamente tal, con sus archivos de control, de datos y de transacciones.

11.4.2- La instancia Oracle

Una instancia de Oracle está conformada por varios procesos y espacios de memoria compartida que son necesarios para acceder a la información contenida en la base de datos. La instancia está conformada por procesos del usuario, procesos que se ejecutan en el background de Oracle y los espacios de memoria que comparten estos procesos.



Según lo que se advierte en la figura, los procesos que se implementan en una instancia de Oracle y su función principal son los siguientes:

- DBWR (database writer): Es el responsable de la escritura en disco de toda la información almacenada en los buffers de bloques que no se han actualizado.
- LGWR (log writer): Es el responsable de escribir información desde el buffer de log hacia el archivo redo log.
- CKPT (checkpoint): Es el responsable de advertir al proceso DBWR de efectuar un proceso de actualización en el disco de los datos mantenidos en memoria, incluyendo los datafiles y control files (para registrar el checkpoint). Este proceso es opcional, si no está presente, es el proceso LGWR quien asume la responsabilidad de la tarea.
- PMON (process monitor): Su misión es monitorizar los procesos del servidor y tomar acciones correctivas cuando alguno de ellos se interrumpe en forma abrupta, limpiando la caché y liberando los posibles recursos que pudieran estar asignados en ese momento. También es responsable por el restablecimiento de aquel proceso que se ha interrumpido bruscamente.
- SMON (system monitor): Levanta una instancia cuando se le da la instrucción de partida (al comienzo del trabajo, encontrándose previamente en shutdown). Enseguida limpia los segmentos temporales y recupera las transacciones que pudieran haberse interrumpido debido a una falla del sistema. Además disminuye la fragmentación del sistema agrupando aquellas extensiones libres que existen dentro de la base de datos.
- ARCH (archiver): La función de este proceso es la de respaldar la información almacenada en los archivos redo log cuando éstos se llenan. Este proceso está siempre activo cuando se ha establecido el modo ARCHIVELOG. Si el sistema no está operando en este modo se hace más difícil recuperar el sistema sin problemas luego de una falla general.

11.4.3- El Área Global del Sistema (SGA)

El SGA es un área de memoria compartida que se utiliza para almacenar información de control y de datos de la instancia. Se crea cuando la instancia es levantada y se borra cuando ésta se deja de usar (cuando se hace shutdown). La información que se almacena en esta área consiste de los siguientes elementos, cada uno de ellos con un tamaño fijo:

- El buffer de caché (database buffer cache): Almacena los bloques de datos utilizados recientemente (se hayan o no confirmado sus cambios en el disco). Al utilizarse este buffer se reducen las operaciones de entrada y salida y por esto se mejora el rendimiento.
- El buffer de redo log: Guarda los cambios efectuados en la base de datos. Estos buffers escriben en el archivo físico de redo log tan rápido como se pueda sin perder eficiencia. Este último archivo se utiliza para recuperar la base de datos ante eventuales fallas del sistema.
- El área shared pool: Esta sola área almacena estructuras de memoria compartida, tales como las áreas de código SQL compartido e información interna del diccionario. Una cantidad insuficiente de espacio asignado a esta área podría redundar en problemas de rendimiento. En resumen, contiene las áreas del caché de biblioteca y del caché del diccionario de datos.
 - El caché de biblioteca se utiliza para almacenar código SQL compartido. Aquí se manejan los árboles de parsing y el plan de ejecución de las queries. Si varias aplicaciones utilizan la misma sentencia SQL, esta área compartida garantiza el acceso por parte de cualquiera de ellas en cualquier instante.
 - El caché del diccionario de datos está conformado por un grupo de tablas y vistas que se identifican la base de datos. La información que se almacena aquí guarda relación con la estructura lógica y física de la base de datos. El diccionario de datos contiene información tal como los privilegios de los usuarios, restricciones de integridad definidas para algunas tablas, nombres y tipos de datos de todas las columnas y otra información acerca del espacio asignado y utilizado por los objetos de un esquema.

11.4.4- El Área Global de Programas (PGA)

Esta área de memoria contiene datos e información de control para los procesos que se ejecutan en el servidor de Oracle (relacionados con la base de datos, por supuesto). El tamaño y contenido de la PGA depende de las opciones del servidor que se hayan instalado.

11.4.5- Las Transacciones

El término transacción describe a una unidad lógica de trabajo que está compuesta de una o más sentencias SQL, que deben terminar con una instrucción commit o rollback. En ese instante, una nueva transacción dará comienzo y estará activa hasta que se ejecute alguno de esos dos comandos otra vez. Cabe destacar que una transacción no se considera confirmada hasta que ésta se termina de escribir en el archivo de redo log.

Oracle Enterprise Manager proporciona la posibilidad de gestionar múltiples grupos de trabajo remoto desde una única y centralizada consola. El administrador verá una representación gráfica de todos los objetos importantes. También podrá establecer un calendario de tareas a ejecutar en un objeto o grupo de objetos, para lo cual se verá ayudado por el Intelligent Agent, que obedece las ordenes ejecutadas desde la consola central.

Otro cometido de este agente es detectar problemas que puedan ir surgiendo e informar de los mismos. En lo referente a Internet las aplicaciones web pueden acceder a los datos almacenados en las bases de datos de Oracle así como presentar documentos HTML generados dinámicamente a partir de un modelo de una consulta. Sin embargo para conseguir una óptima conectividad basada en Java o soporte procedimental de este es necesario acudir a la reciente aparición de Oracle 8i.

Oracle soporta paralelismo dentro de una consulta lo que proporciona un incremento notable en su ejecución. También soporta procesos de transacciones On- Line y DataWareHousing.

Atendiendo a las características de manejabilidad escalabilidad rendimiento y soporte entre plataformas tenemos una arquitectura de servidor con ejecución multihilo y rendimiento de multiprocesadores simétricos (SMP). En cuanto a los bloqueos se refiere, los admite a nivel de filas sin restricciones. Las bases de datos pueden crecer hasta límites que en la práctica pueden considerarse inalcanzables y gracias a un rediseño respecto a las versiones se han eliminado ciertos cuellos de botella.

Las características más relevantes en lo referente a la escalabilidad residen en las tablas particionadas que son aquellas que pueden dividirse en múltiples dispositivos de almacenamiento de acuerdo con valores preestablecidos y el multiplexado y pooling de conexiones. Precisamente la gestión de conexiones se realiza mediante cierto software específico. Concretamente el de red, que antes se llamaba SQL NET, ahora se denomina Net, que a parte de cambiar de nombre ha avanzado tanto en las características (interface de programación, navegador de Internet con soporte Java y soporte de red adicionales) y rendimiento, como en facilidad.

El objetivo fundamental de Net 8 es establecer sesiones y transferir datos entre un cliente y un servidor o entre dos servidores. No obstante para obtener multiplexado en conexiones (se combinan varias conexiones de clientes entrantes hacia una única conexión de salida hacia el servidor), es necesario un componente adicional denominado Connection Manager.

Los protocolos de red soportados por Oracle son los siguientes:

- Net 8
- TCP/IP
- IPX/SPX
- Pipes con nombre
- DECNET
- DCE
- NDS
- LU 6.2 (APPC)

Entre las características de SQL podemos destacar una optimización independiente de la sintaxis, unas consultas y una generalización de estadísticas de tablas (ANALIZE). Pero lo que realmente incrementa drásticamente el rendimiento es el soporte de referencia a objetos que son rápidos punteros de tablas los cuales sustituyen a las lentas y traicioneras unidades relacionales.

Este sistema sabiamente utilizado puede cambiar el modo en que las bases de datos relacionales son diseñadas. Existen ciertos métodos de extensibilidad basados en un soporte limitado con llamadas en C. Por desgracia no son tan brillante en DB2 o Informix Universal Server. Un componente denominado ServeTuner es el encargado de optimizar el rendimiento de la base de datos de forma automática. Para ello analiza los recursos de los que dispone el sistema y los requerimientos de la aplicación mientras la base de datos esta en funcionamiento, recomendando los valores más adecuados y aplicándolos en el momento más adecuado.

11.4.6- Roles y Responsabilidades del DBA de Oracle

El administrador de la base de datos de una empresa es siempre considerado como la persona con más experiencia en el área de bases de datos. Por lo anterior, es conveniente tener muy claras las expectativas que se generan en torno a su trabajo y cuáles son los principales roles que debe asumir dentro del marco corporativo o de un proyecto. Entre sus tareas básicas, destacaríamos:

- Instalación de nuevos componentes del software: Una de las tareas principales del DBA consiste en la instalación periódica de nuevas actualizaciones de software de Oracle, tanto en lo referente a programas de aplicaciones como a herramientas administrativas. También es recomendable que el propio DBA y otros usuarios de Oracle prueben la instalación y nuevas configuraciones antes de migrarlas a los ambientes de producción.
- Interacción con el administrador del sistema: En la mayoría de los casos los programas sólo pueden ser instalados o accedidos por el administrador del sistema. En este caso, el DBA debe trabajar siempre muy bien coordinado con él para garantizar que tanto la instalación y configuración de software como de hardware permita un adecuado funcionamiento del motor de base de datos y de las aplicaciones.
- Garantizar la seguridad del sistema: El DBA debe siempre monitorear y administrar la seguridad del sistema. Esto involucra la incorporación y eliminación de usuarios, administración de espacios de disco (cuotas), auditorías y una revisión periódica para detectar probables problemas de seguridad.
- Monitorización: El DBA debe monitorear continuamente el rendimiento del sistema y estar preparado para efectuar ajustes de sintonización de éste. En ciertas oportunidades esto involucra cambiar sólo algunos parámetros y otras veces reconstruir índices o reestructurar tablas.
- Respaldos: Debido a que la tarea más importante del DBA es proteger la integridad de los datos, se deberá desarrollar una estrategia efectiva de respaldos y recuperación de datos para mantener la estabilidad de toda la información guardada. Las frecuencias de estos respaldos deberán decidirse dependiendo de la cantidad de procesos que alteran los datos a través del tiempo.
- Prevención de riesgos: Otra tarea del DBA es la de calendarizar mantenciones a las bases de datos (archivos lógicos) o cooperar en el mantenimiento de las máquinas al administrador del sistema. El DBA debe fortalecer sus esfuerzos en orden a eliminar problemas o situaciones potencialmente peligrosas.
- Analizar datos y efectuar recomendaciones concernientes a mejorar el rendimiento y la eficiencia en el manejo de aquellos datos que se encuentran almacenados.
- Apoyar en el diseño y optimización de modelos de datos.
- Asistir a los desarrolladores con sus conocimientos de SQL y de construcción de procedimientos almacenados y triggers, entre otros.
- Apoyar en la definición de estándares de diseño y nomenclatura de objetos.
- Documentar y mantener un registro periódico de las mantenciones, actualizaciones de hardware y software, cambios en las aplicaciones y, en general, todos aquellos eventos relacionados con cambios en el entorno de utilización de una base de datos.

11.5- SYBASE

Con la evolución del producto se ha ido mejorando la interfaz sustancialmente y en la versión NT cada vez son más las pantallas gráficas y los asistentes.

El rendimiento de Sybase, al igual que otros productos evaluados, conforme crece el hardware del servidor ya que también admite configuración específica para multiprocesadores simétricos (SMP). La idea es que consulta, ordenación de dato y otros procesos puedan ser ejecutados en paralelos. Según los

datos ofrecidos por Sybase las consultas distribuidas en paralelo, pueden proporcionar respuestas hasta 15 veces más rápidamente (en situaciones muy optimas)

Sybase en su versión Adaptive Server está constituido por una selección de componentes Adaptive Server (RDBMS). Backup Server se utiliza para las copias de seguridad de datos. Monitor Server o Historial Server se encargan de capturar, mostrar y evaluar datos de rendimiento, así como ajustar el Component Integration Service, se encargan de extender la funcionalidad presentando una vista de los datos de forma uniforme en las aplicaciones clientes.

Server Config es la herramienta de configuración y SQL remote proporciona replicación basada en mensajes entre una base de datos central y un conjunto de bases de datos sitas en ordenadores de sobremesa en la misma ubicación u otra diferente. Por la parte del cliente existen más componentes especialmente útiles. Si no tenemos ninguna aplicación propia que manipule los datos. Así tenemos el Net Impact Dynamo, que es un conjunto de herramientas cuya finalidad es construir y manipular websites enlazados a bases de datos de Sybase. Se trata de diseñar de paginas HTML que sirvan como modelo e incluir SQL en las mismas, pero utilizando un asistente que reduce enormemente el tiempo de diseño. Se incluye un superconjunto de JavaScript. También tenemos en el paquete al popular PowerSoft Infomakerr, herramienta diseñadora multifunción ideal para general formularios, informes, gráficos, tablas, etcétera.

Pero tal vez la herramienta más útil para el administrador sea SQL Modeler, con la cual es posible diseñar estructuras de bases de datos de forma optima, incluyendo eventos y procedimientos integrados. Esto puede ser muy interesante que también es posible realizar procesos de ingeniería inversa interpretando la estructura de las bases de datos de otros sistemas (Access, DB2, SQL Server...) Una vez interpretada se construye la equivalente de Sybase, que se modifica con las herramientas del entorno.

Una de las posibilidades que ofrece el producto es la replicación y se logra gracias al componente SQL remote. Además ofrece un Replication Server, cuando los servidores de bases de datos son un pequeño número y el tiempo característico de actualización es en orden de segundos. Un ejemplo de uso característico de SQL Remote, unido al Adaptive Server, es la creación de una Intranet dónde se puedan añadir puestos móviles.

Las plataformas que admite el producto son:

- WinNT
- Digital Unix
- DCR System
- SCO- UnixWare
- Silicon Graphics
- HPUX
- Aix,
- Solaris
- Linux.

Se soportan los siguientes protocolos:

- APPC
- IPX/SPX
- NetBios
- TCP/IP
- Pipes con nombres

Finalmente, decir que la seguridad de integridad de los datos es uno de los objetivos de Sybase, con una certificación ISO 9001 DBMS. Incluye Duplicación/Espejo de Disco, Volcado de seguridad y restauración de alta velocidad, transacciones en línea con rollback automático.

11.6- MsACCESS

MsAccess es un programa comercial de la empresa Microsoft Corporation. Se trata de un sistema gestor de bases de datos diseñado para ser utilizado en computadoras personales tipo PC. Aunque no está considerado como un sistema serio, lo cierto es que permite realizar las mismas tareas que los tradicionalmente considerados sistemas serios, como puede ser Oracle. A diferencia de éste, Access puede funcionar en sistemas de bajo coste, y es mucho más asequible. Además, dispone de un entorno muy amigable (al estilo Windows 95). Por esta razón, cada vez más empresas apuestan por este sistema.

Por todas estas razones, estudiar un sistema como Access es una buena apuesta, ya que abre las puertas a un mercado cada vez más importante, sin que ello suponga dejar de conocer otros sistemas muy extendidos.

Los programadores no pueden ser ajenos a los nuevos modelos de diseño de bases de datos, tanto en soluciones de escritorio como en los grandes servidores. Es difícil que en una aplicación no se utilice una base de datos. Incluso en aquellas aplicaciones que no están destinadas al mundo empresarial, la base de datos es algo fundamental, pues siempre hay algo que almacenar.

Evidentemente, dada la cantidad de situaciones distintas en las que se puede utilizar una base de datos, es lógico que existan múltiples tipos, así como muchas formas distintas de acceder a ellas.

Las bases de datos actuales son todas relacionales, con un conjunto de tablas que se encuentra dividido en filas y columnas. Para acceder a las bases de datos relacionales, IBM desarrolló el lenguaje SQL. Pese a que la gran mayoría de las bases de datos fueron implementando este lenguaje, la forma de controlarlo desde un programa era diferente en función del fabricante. Para evitar estos problemas se desarrollaron controladores de acceso, siendo uno de los más famosos el ODBC, que proporcionan al programador un conjunto de funciones (API) estándar para acceder al motor. Cada fabricante suele desarrollar junto a la base de datos el controlador correspondiente, de igual forma que cada fabricante de impresoras fabrica el controlador de la misma. Con esto se consiguió que los programadores pudiéramos acceder de forma estándar a cualquier base de datos que incorporase el correspondiente controlador. A este tipo de software intermedio se le suele denominar middleware (palabra utilizada siempre que existe software intermedio).

Sin embargo, con el fin de facilitar aún más las cosas, algunos fabricantes de herramientas de desarrollo dotaron a éstas con una nueva capa intermedia de software (motor de base de datos) intercalada entre el código fuente del programa y el controlador ODBC, o directamente entre el programa y la base de datos. En la actualidad los dos motores de base de datos más comunes son Microsoft Jet y BDE (Borland Database Engine).

Finalmente, con la introducción de herramientas de desarrollo rápido (RAD) han proliferado los controles y los objetos de acceso a datos. Estos establecen un puente entre la interfaz del usuario y el motor de acceso a datos, por lo que la tarea del programador, tal y como se ve, queda reducida a la mínima expresión. Actualmente nos encontramos en esta etapa, donde los controles y los objetos de acceso a datos gobiernan las nuevas aplicaciones.

11.7- ADABAS (de Software A.G.)

Existen otras aplicaciones análogas a las comentadas anteriormente que normalmente están dedicadas a miniordenadores, siendo frecuente su uso en ámbito gubernativo o de gran empresa. Pero también existen otros que pueden funcionar en los PC y cuya introducción en el mercado crece día a día. Este es el caso de Adabas (de Software A.G.). El sistema de gestión de bases de datos ha evolucionado desde la primera versión hasta el momento actual con la versión 6, para ofrecer una familia de servidores optimizados para las principales arquitecturas del mercado (esta disponible en entornos Main-Frame MVS, y entornos departamentales como Unix OpenVMS, BS2000 y Windows NT) y para soportar aplicaciones de misión crítica en entornos corporativos.

Entre las características de Adabas podemos destacar a un servidor existente de datos multimedia a escala corporativa, que permiten la gestión de todo tipo de estructuras de datos y objetos complejos, así como arquitectura distribuida o replicada con funcionalidad SQL- Ansi ISO y accesibilidad via ODBC y JDBC.

La capacidad de conexión y acceso al Web permite la generación dinámica de páginas HTML, admitiendo la combinación con lenguajes script lo que facilita una sencilla interacción de nuevos servicios de información en línea. La adaptación de nuevas estructuras organizativas workflow vía Internet o la puesta en marcha de aplicaciones de Comercio electrónico.

Concretamente en la Versión 6, Adabas comenta la extensibilidad del sistema, por ejemplo, aumenta el número de bases de datos por nodos, el número de ficheros, y el número de registro por ficheros, al tiempo que mejora el rendimiento con su nueva gestión de hilos de ejecución. Perfecciona alguna de sus características, como los triggers (basados en Natural) y procedimientos almacenados.

Otra aplicación de alto nivel, una RDBMS orientada a DataWareHousing que corre en Unix y Windows NT (desde 1998). Como su nombre hace sospechar tiene una escalabilidad prácticamente ilimitada, diseñada para trabajar con bases de datos enormes gracias a una tecnología llamada DSS (Decision Support Scalability). De hecho su punto más fuerte es el contar con un sector de mercado donde los tamaños de las bases de datos superan los 500 Tbytes. La versión 2.0.1 para NT dispone de herramientas con interfaces avanzadas dentro del entorno Windows, así como drivers JDBC para acceso a Java desde Internet.

11.8- Conzept 16

Distribuido en España por Mastertelecom Corp., Conzept 16 es un gestor de Bases de Datos Multiplataforma. Este sistema es un servidor No SQL, con características que lo hacen resaltar entre los demás de su categoría. Entre los protocolos que soporta ésta son TCP/IP, IPX/SP. Está diseñado para trabajar en Red, con una arquitectura cliente- servidor. Se distribuye con documentación en castellano, acompañado de una Guía de Aprendizaje y una Guía de Consulta.

Entre las características destacables de Conzept 16 se encuentra una sofisticada gestión de memoria pudiendo utilizar cualquiera de los sistemas BIOS, XMS, EMX, VCPI o DPML.

El lenguaje de programación utiliza palabras reservadas en Inglés y Alemán, y tiene utilidades y asistentes para realizar las tareas simples y procesos por lotes. Incorpora generadores de listados y consultas integrados en el mismo producto.

Es otro producto de interés al ofrecer una base de datos integrada de prestaciones moderadas pero suficientes para muchas aplicaciones. En su nueva versión se ofrece como un servicio más de NT, que ofrece sus prestaciones de forma silenciosa. Admite la definición de funciones externas (triggers y procedimientos) permitiendo el uso de las DLL que se encuentren en el sistema (el acceso a estructura de datos complejas y dispositivos especiales son algunas de las utilidades).

También dispone de un driver JDBC (nivel 4) para acceso a Java desde Internet y un soporte mejorado de ODBC, compatible con la versión 3.

11.9- Lotus Approach

La solución doméstica de Lotus para las bases de datos es Approach. Se distribuye junto con su paquete integrado SmartSuite. Se trata de un gestor de bases de datos relacionales que incorpora las funciones básicas para crear, mantener pequeños sistemas de datos, así como facilitar las tareas de acceso a los mismos. Su principal virtud es la total integración con el Sistema Lotus Notes para trabajo en grupo, que proporciona una gestión eficiente para la gestión de usuarios en LANs, siendo ideal para Pymes.

Incorpora un interfaz único representado como una aplicación individual desde la cual podemos crear la base de datos mediante sencillos pasos y menús gráficos. Una vez creadas las tablas y relaciones, es posible definir índices vistas y consultas. Dispone de compatibilidad SQL y con un generador de instrucciones de este lenguaje, muy simple y limitado.

Genera Gráficos cuantitativos a partir de tablas, aspecto muy atractivo para la creación de informes de utilidad en el ámbito profesional. Dispone un lenguaje de script llamado Lotus- Script (propio de SmartSuite), con el que se pueden crear potentes macros.

11.10- Motores de búsqueda: BDE (Borland Database Engine)

BDE (Borland Database Engine) nació como solución al acceso a las bases de datos desde los lenguajes clásicos. Inicialmente únicamente proporcionaba acceso a unas cuantas bases de datos, pero en la actualidad se ha convertido en una potente herramienta para el programador que desee acceder a prácticamente cualquier BD. El API que proporciona a los programadores se denomina Integrated Database Application Program Interface (IDAPI) y, entre otras, cuenta con las características que detallamos a continuación.

En primer lugar proporciona un acceso unificado a bases de datos como Paradox, dBASE, InterBase, Oracle, Sybase e Informix, además de poder acceder a cualquier base de datos que posea el correspondiente controlador ODBC. La última versión incluida en Delphi 3 incorpora también conectividad nativa con MS Access, FoxPro, DB2 y AS/400 en la versión Delphi/400 y C++Builder/400.

BDE está formado por el ya mencionado IDAPI, un motor (engine) de consultas en SQL y QBE (Query By Example), tres controladores directos para acceder a Paradox, dBASE y ficheros de texto, controladores nativos opcionales para Oracle, Sybase, Informix e Interbase y conexión ODBC. Al tratarse de un software orientado a componentes, BDE contiene una serie de servicios compartidos, con el fin de facilitar el desarrollo y maximizar la reutilización. Por ejemplo, un administrador permite a todos los controladores compartir el mismo buffer, de forma que los buffers propietarios de algunos controladores pueden ser ubicados dentro del mismo buffer general. Pese a que los controladores no tienen por qué compartir este buffer general, si se hace, los recursos del sistema se maximizan. Con BDE se consigue una alta independencia del sistema operativo, pues éste queda aislado, controlando de forma directa la entrada y salida de archivos, accesos a la red y ubicación de memoria.

Con el fin de evitar los típicos cachés que hay que realizar para trabajar con BLOBs, BDE posee su propio servicio de caché. Éste es accesible desde todos los controladores y, además, es posible abrir múltiples BLOBs de forma simultánea. Una característica muy interesante es que el caché de BLOBs permite un acceso aleatorio, evitando así tener que transferir cada uno de ellos a un archivo. Esta es una característica que BDE implementa cuando la fuente de datos no lo permite hacer de forma directa.

Mediante un servicio de reestructuración que actualmente está disponible para cualquier base de datos, se proporciona al desarrollador la posibilidad de añadir o modificar campos o cualquier aspecto estructural de una tabla. Este mecanismo crea una nueva tabla que refleja las modificaciones en la estructura de la original y transfiere todos los registros adaptándolos de forma automática. Mediante la herramienta SQL Explorer permite realizar modificaciones de la estructura de cualquier tabla utilizando SQL.

Otra característica interesante de BDE es que permite la utilización de cursores enlazados dando posibilidad de establecer relaciones de una a muchas entre dos tablas. Mediante esta característica se permite la creación de complejas aplicaciones multitabla sin demasiada programación.

11.11- Motores de búsqueda: MS Jet (Microsoft)

MS Jet es el motor de bases de datos que Microsoft proporciona en la mayoría de sus productos. Se trata de un producto que es independiente de la aplicación que lo utilice y, por este motivo, puede ser utilizado desde los clásicos Excel o Word hasta las herramientas de programación más avanzadas, como Visual C++ o Visual Basic.

Mediante los objetos de acceso a datos (DAO) el programador tiene a su alcance un control directo sobre la base de datos. Microsoft Jet puede controlar directamente ODBC, aunque no se reduce únicamente a él, ya que puede acceder, mediante ISAM, a un número limitado de bases de datos, como FoxPro, Paradox, dBASE, Lotus 123, Excel y ficheros de texto.

Las aplicaciones o herramientas de programación que utilicen Microsoft Jet, pueden almacenar sus propios objetos en la base de datos, así por ejemplo, en Access se almacenan dentro de la misma base de datos formularios, macros y módulos de programa. Dentro de la base de datos los registros son organizados en páginas. Cada página puede contener más de un registro, pero nunca un registro puede superar el tamaño de una página. Algunos campos particulares, como Memos y objetos OLE, se suelen almacenar en páginas separadas del registro que los contiene. Otra cualidad es que los registros contenidos dentro de una página no tienen por qué tener la misma longitud.

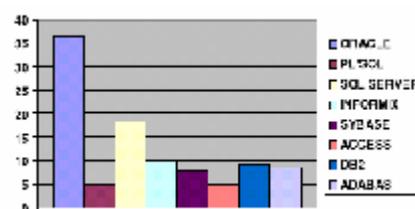
La ingeniería de consulta (query engine) de Microsoft Jet soporta la recuperación y manipulación de datos desde SQL, aunque también se pueden utilizar DAO para programar el acceso a los datos de las tablas. Dentro del terreno de la seguridad, Microsoft Jet permite definir usuarios y grupos, así como fijar los permisos de cada objeto de la base de datos para los usuarios y grupos definidos. Tanto los grupos como los usuarios y sus claves de acceso se manejan dentro de un entorno seguro proporcionado por el Administrador de Grupos de Trabajo (Workgroup Administrator).

Mediante Microsoft Jet es posible dar acceso simultáneo a varios usuarios a la base de datos. Para ello, Jet cuenta con dos sistemas de bloqueo: pesimista y optimista. Mediante el bloqueo pesimista dos usuarios no pueden modificar al mismo tiempo el mismo registro, ya que se bloquea la página que lo contiene en el momento que comienza la edición del mismo por parte de algún usuario. Sin embargo, mediante el bloqueo optimista, Jet permite la edición del mismo registro a varios usuarios, puesto que la página que lo contiene es bloqueada únicamente cuando se actualiza.

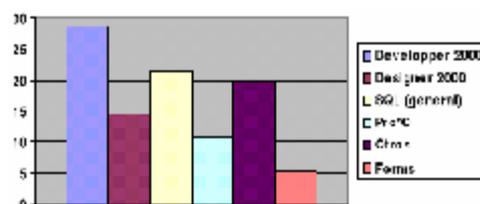
Dentro de las funciones típicas de mantenimiento, Microsoft Jet proporciona funciones que permiten la compactación, reparación y conversión de la base de datos. La compactación reestructura la base de datos fragmentada como consecuencia de la eliminación e inserción de registros. Cuando la base de datos se daña por cualquier motivo, como por ejemplo un corte de fluido eléctrico, se deberá acudir a las funciones de reparación que, si bien no funcionan el 100% de las veces, sí resuelven los problemas más típicos.

11.12- Estadística de la distribución de los diferentes SGBD en las empresas españolas.

Como para los restantes tipos de Software, el mercado de los Sistemas Gestores de Bases de Datos está muy diversificado. Existen infinidad de ellos (sólo he tratado los más relevantes). Como siempre la elección, en el caso de encontrarse ante la necesidad de adquirir uno de ellos, irá en función de las necesidades que tengamos. Siempre será muy importante evaluar éstas necesidades antes de tomar cualquier decisión.



Porcentaje de aparición en el mercado español



Porcentaje de aparición en el mercado español

12.- BIBLIOTECA MULTIMEDIA

Muchas veces se presenta la necesidad de almacenar cierto tipo de información como imágenes, sonidos documentos de texto antiguos o hechos en procesadores de texto (como Microsoft Word) , o hecho con hojas de cálculo o muchos datos, correspondientes a una base de datos, diapositivas e incluso archivos de video de pequeño o gran tamaño, en un lugar que no sea el sistema de archivos del disco de un servidor de archivos, ya sea porque se carece de espacio en disco o porque es indispensable cumplir con un requerimiento de negocio o por el mero hecho de la consulta.

No son solo estas circunstancias las que nos hacen abrir los ojos para diferenciar que existe información de carácter estructurado e información no estructurada. Los términos "estructurada" o "no estructurada" son han sido usados a lo largo de este trabajo y corresponden a información simple e información compleja. En términos de base de datos y almacenamiento, podemos llamar información simple a un campo en una tabla que guarda el nombre de un empleado y para la cual existe un tipo de dato básico como una cadena de caracteres. Así en los términos anteriores el tipo de información compleja sería la correspondiente a la almacenada en un campo que guarda la fotografía del empleado y para la cual existen tipos de datos especiales implementados por las distintas base de datos del mercado ya sea de escritorio como las corporativas.

Pero, ¿De qué me serviría tener este tipo de información como archivos JPEG, MP3, archivos de MS Word, archivos de MS Excel, archivos de imágenes MPEG e incluso videos de pequeño o gran tamaño en formato AVI, almacenados en una tabla de una base de datos relacional? Pues como toda solución para satisfacer los requerimientos de las personas a quienes nos debemos, es decir nuestros amigos usuarios, tiene sus beneficios.

Uno de ellos es que tanto información "estructurada" como "no estructurada" que estén relacionadas lógicamente puedan estar almacenadas en un mismo repositorio y formando parte de un registro en una base de datos, así permanecerán siempre juntas aun si la base de datos cambiara de lugar o de servidor. Un ejemplo es la posibilidad de tener en una tabla llamada Empleado información como el nombre, apellidos, dirección, etc. y la fotografía del mismo en la misma tabla, incluso su hoja de vida originalmente en un archivo de MS Word almacenada junto con toda la información antes mencionada.

Otro beneficio es la posibilidad de almacenar documentos de carácter oficial como Normas, Resoluciones, etc. en su versión definitiva para asegurar la unicidad de los mismos y para que no estén expuestos a cambios o dañen su integridad. Como se ha visto ya anteriormente, la seguridad de la información también entra en este aspecto. Para estos casos las bases de datos corporativas manejan mecanismos de protección que permiten autorizar o denegar el cambio de ciertos campos a determinados usuarios de base de datos, el tema de seguridad cubre varios niveles por lo que también se pueden implementar mecanismos de seguridad en el nivel de la aplicación web.

Habiendo enumerado algunos de los beneficios de esta solución, ahora toca ver el otro lado de la moneda y es que la principal desventaja de esta solución es el consumo de espacio en la base de datos por lo que es recomendable no abusar en almacenar tanta información "no estructurada" ya que generalmente esta se almacena como flujos de bytes en campos de tipo binarios para los cuales se mantiene metida en una proporción mayor a la usada para los tipos de datos simples.

Afortunadamente, el avance tecnológico actual, nos permite almacenar gran cantidad de datos, entendiendo como datos cualquier tipo de imágenes, sonidos, videos, etc, que pueden ser archivados digitalmente, en cada vez más soportes diferentes (discos duros de Gb, CD-ROM, DVD, etc), aumentando también la velocidad de acceso a los mismos. En definitiva, que la creación de una biblioteca multimedia, con todos los avances tecnológicos existentes en la actualidad, supone no sólo una fuente inagotable de posibilidades, sino que es una forma de ofrecer a todos los usuarios una manera rápida y eficaz de consultar cualquier tipo de contenido.

Esta revolución en el área de la información, que dicho sea de paso, también pretende mayores beneficios, permite un acceso directo a la información, con lo que tiene de beneficioso si hablamos de tiempo de trabajo, pero requiere una infraestructura tecnológica adecuada. Equipos informáticos de gran capacidad

de almacenaje y procesamiento muy veloz, así como una rápida conexión a Internet, son absolutamente necesarios para este proyecto.

La realidad virtual que nos permite una biblioteca multimedia, se refiere a la representación de los recursos y servicios bibliotecarios por medio de una interficie visual y espacial, que requiere efectos de luz e imágenes de aquello constituye a la biblioteca en particular. Por ejemplo, viajar a través de la biblioteca Británica conociendo sus colecciones, edificios y cada uno de los detalles que la componen, y todo ello, sin movernos del puesto de consulta o trabajo.

Podemos establecer, a lo largo de la historia, la siguiente clasificación de los tipos de bibliotecas, tal y como los conocemos en la actualidad:

- Biblioteca Tradicional: desde su origen hasta el inicio de la automatización. Desde la aparición de las primeras bibliotecas en el siglo XII hasta las primeras décadas del siglo XX.
- Biblioteca Moderna o Automatizada: s. Aparece hacia la mitad del siglo XX. Intenta reducir el trabajo manual a su máxima expresión. Aparecen otros enfoques diferentes a una biblioteca como tal (se almacenan videos, CD, DVD, discos de música, etc).
- Biblioteca Virtual o del futuro: Aparece en la década de los 80. Intenta la accesibilidad universal de la información por medio de las telecomunicaciones, en tiempo real. Tiene las siguientes subdivisiones:
 - Biblioteca Electrónica: Implica su acceso en línea. Los documentos se ofrecen en formato electrónico
 - Biblioteca Digital: Colección de textos en soporte electrónico ó digital, disponibles para el usuario final
 - Biblioteca Virtual: Reemplaza los servicios de bibliotecas tradicionales, independientemente de donde se encuentre el usuario. Permite el acceso a los recursos de información de otras instituciones

Una definición de biblioteca virtual dada por Progressive Technology Federal Systems, Inc. es:

La Biblioteca Virtual provee acceso instantáneo y conexión electrónica a bibliotecas, personas, instituciones, y empresas de todo el mundo. La BV provee acceso a un amplio rango de recursos intelectuales, no limitado a los textos tradicionales, disponibles en los sistemas de información, como también a bases de datos de texto e imagen, objetos multimedia, a través de una interfaz interactiva. La BV provee un único punto de acceso desde el escritorio del usuario, a los recursos y proveedores de información

12.1- Ejemplo de una base de datos multimedia (Carnegie Mellon University)

A continuación se muestra un pequeño ejemplo de tratamiento de un tema (concretamente el fenómeno climatológico conocido como "El niño") mediante el uso de diferentes fuentes de información puestas a disposición del usuario, como noticia, hecho histórico o científico. Se ha respetado el formato original del proyecto, para que sea analizado tal y como fue creado en su origen. Para cualquier consulta o ampliación de conocimientos, puede visitarse la web de Carnegie Mellon University en la siguiente dirección.

<http://www.informedia.cs.cmu.edu/>

En las figuras se observa como el tratamiento que se da a la noticia es amplio y con gran profusión de imágenes, videos y sonido. Se entrelazan imágenes en tiempo real (por ejemplo las noticias en directo), con videos (los grados por los reporteros gráficos en el lugar del suceso) e imágenes fijas (por ejemplo los mapas de la zona).

Project Description

Vast collections of video and audio recordings which have captured events of the last century remain a largely untapped resource of historical and scientific value. The Informedia Project at Carnegie Mellon University has pioneered new approaches for

automated video and audio indexing, navigation, visualization, search and retrieval and embedded them in a system for use in education, information and entertainment environments. This project was initiated in 1994 as one of six Digital Library Initiative (DLI) projects funded jointly by NSF, DARPA and NASA, and is the only one focused on the video field. We continue the fundamental goal of enabling for video all the functionality and capability existing for textual information retrieval, while leveraging its temporal and visual qualities for richer information delivery. Informedia-II establishes an era focused for the user as we introduce new paradigms for video information access and understanding. We will create and integrate video content on-demand to enable summarization and visualization that provides responses to queries in a useful broader context, perhaps with historic or geographic information.

Background

The Informedia system provides full-content search and retrieval of current and past TV and radio news and documentary broadcasts. The system implements a fully automated process to enable daily content capture, information extraction and storage in on-line archives by applying artificial intelligence and advanced systems technology. The current library consists of a 1,500 hour, one terabyte library of daily news captured over the last two years and documentaries produced for public information and government agencies. This prototype database allows for efficient retrieval of individual video paragraphs which satisfy an arbitrary spoken or typed subject area query based on the words in the soundtrack, closed-captioning or text overlaid on the screen. There is also a capability for matching of similar faces and images.

Our approach uniquely combines speech recognition, image understanding and natural language processing technology to automatically transcribe, segment and index the linear video. These same tools are applied to accomplish intelligent video search, navigation and selective retrieval. The process automatically generates various summaries for each video segment: headlines, filmstrip thumbnails and video-skims. Figure 1 illustrates a typical query and result set display.



Figure 1. IDVL interface showing 12 documents returned for “El Niño” query along with different multimedia abstractions from certain documents.

The Informedia-II Project improves both speed and accuracy of the underlying information extraction, now including interpretation of name, place, date and time information, and adding the challenges of dynamic video segmentation, speaker voice and face identification, and video event characterization and similarity matching. The

performance goals will be real-time processing for will be to enable contemporaneous incorporation into an active library, and interoperability across distributed proprietary video archives.

Summaries rather than documents become the units of discourse, as shown in Figure 2. will be sources can be viewed in the context of these summaries, showing how events unfold over time and across geographic boundaries, allowing visualizations that emphasize time and space will be cial l .

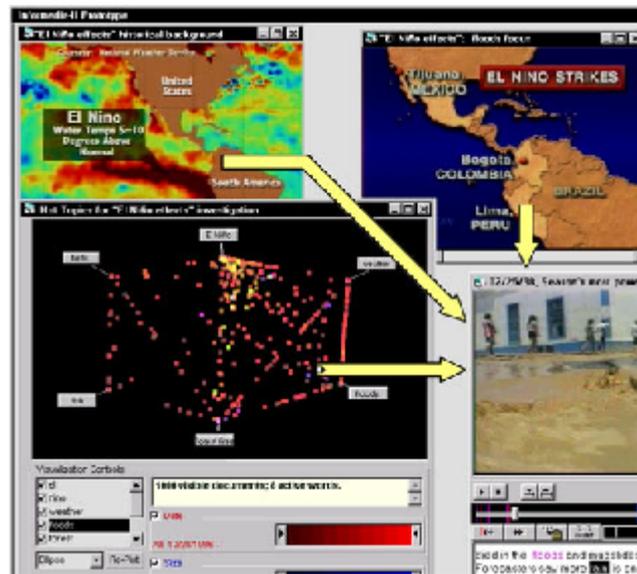


Figure 2. Additional views provided by the Informedia-II interface for an will be into “El Niño effects”.

Spatial and Temporal Analysis

Keywords were used successfully for information retrieval within Informedia, but were less effective for cases will be as distinguishing the person “Prince William” from the location “Prince William Sound”. The Informedia-II Project will be automatically extract will be cial to named will be cial from the video material, i.e., names, places and times. The interface will be then build from the correlations and aggregations of will be named will be cial. For example, a video report about Prince William Sound can be associated with Alaskan waterways through geographical thesauri. A user searching for oil spills could be shown a map interface with highlighted “hot spots” where each hot spot is a cluster of relevant documents in a particular geographic region; one will be cluster would be on this Alaskan waterway.

Whereas Informedia merely will be c results by recency (or will be cial), extracted temporal information in Informedia-II will be provide the ability to analyze search results over time, minimizing redundancy and exposing will be c and developments. Given a collection of stories covering an evolving event, the system will be return a time-sequenced will be c document segments derived from multiple sources with a visual timeline that summarizes the information. The sources may be text as well as audio or video, and may be will be cial from multiple, distributed content providers.

Summarization

Informedia offered static text summaries of individual video paragraphs, static filmstrip summaries of individual video paragraphs, and precompiled static video-skims of the

content in a video paragraph. For *Informedia-II* we build text, image and video abstractions that are dynamically computed based on the content and history of the user queries and interactions. Moreover, summaries can extend across video documents to represent events across time and space, illustrated in Figure 2. Instead of using a single paragraph as the unit of *informedia*, we will create synthetic video documents that reflect the distillation of information across multiple video paragraphs.

Beyond the video *informedia*, we propose to create collages that *informedia* documents from text corpora, images, audio and video in one single abstraction. Given multiple, near-duplicate or overlapping units of information, the *informedia* work will combine them into one unit, a synthetic *informedia* or video magazine that summarizes all the salient information.

Multimodal Exploration

We will allow multidimensional queries that may combine image elements, video clips, text and speech. New visualization techniques will be designed to help the user obtain an overview of the temporal and spatial evolution given large sets of documents in multiple media. Distribution of the video library's contents across time, space, topics and *informedia* will be will *informedia* for efficient user examination and action. Details can be selectively shown in the context of these dimensions, i.e., the visualizations are active objects supporting direct manipulation for zooming into specific areas. Figure 2 shows summaries and visualizations for "El Niño effects". By zooming in on a contributing document for the summary, a video on Peruvian flooding is shown.

Impact

The *Informedia-II* Project has the *informedia* goal to unlock the barriers to productive video information access. Via automated processing, the contents of *informedia* video libraries can be more fully *informedia* where *informedia* the only window to that data was through the titles and perhaps a few manually entered keywords. Videos of past reviews, training and teleconferences can become *informedia* corporate assets. Outreach for *informedia* archives and public service video collections can be enhanced. We envision that the legacy for this project will enable:

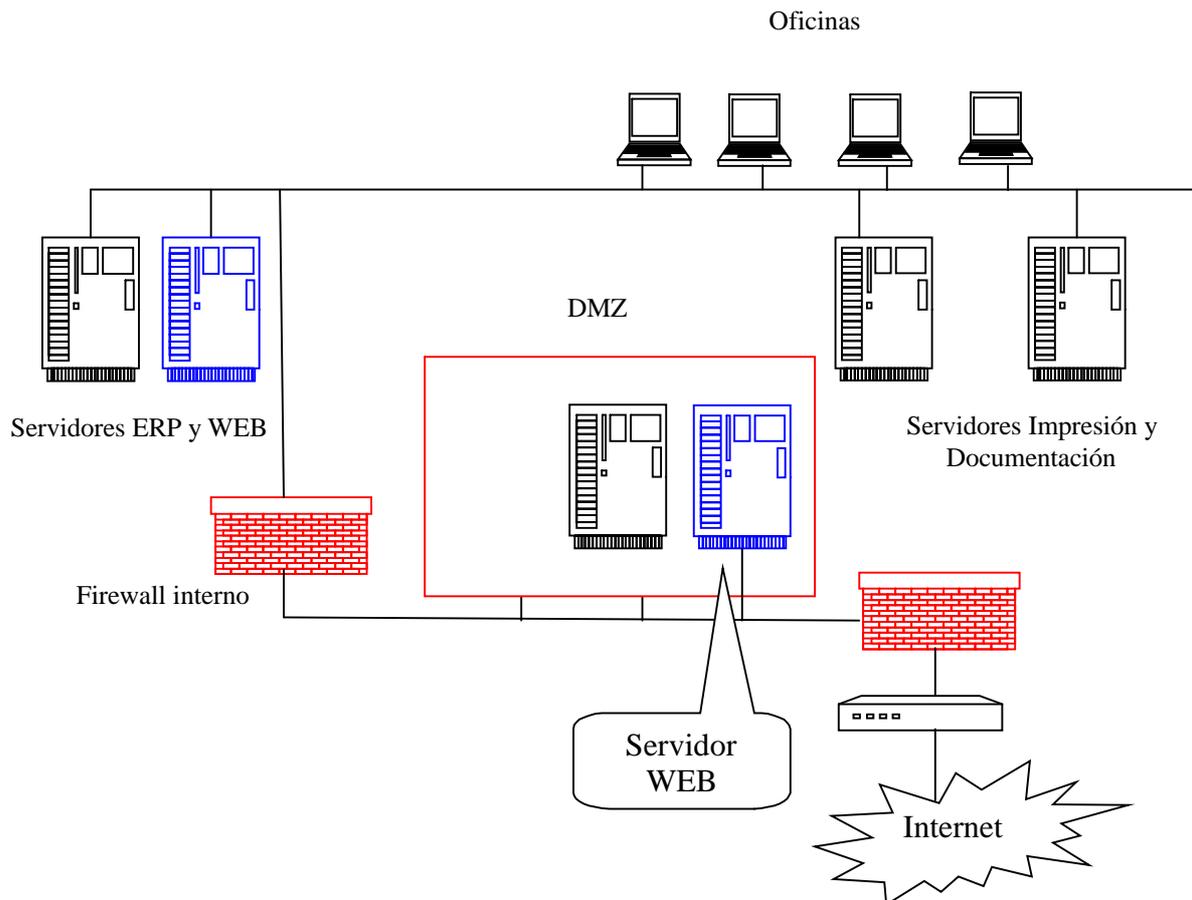
- Broader accessibility to the historical video record being created across the globe.
- Pervasive societal impact as we provide a *informedia* new capabilities to understand how events evolve and are correlated over time and geographically.
- Increased use of video in domains previously dominated by text or speech.

We will work with content providers to make their materials more *informedia*, and to study patterns of use of their video by appropriate communities of users. Our goal for unlocking the information embedded in video for easy access by the student, teacher, journalist, scientist or home user has the *informedia* goal to create video resources with significant educational and *informedia* value. Our approach of automatically processing large libraries will stress current limits on network and i/o bandwidth, disk space and processor speed. Our focus on video as a searchable resource has broad implications for information gathering and dissemination.

Carnegie Mellon University
 School of Computer Science
 5000 Forbes Avenue
 Pittsburgh, PA 15213
 informedia@cs.cmu.edu

12.1.1- Distribución de servidores de la Carnegie Mellon University

En capítulos anteriores se ha comentado el tema de la seguridad. Seguridad no sólo en el aspecto de diseño de los SGBD, sino en cuanto a la protección de los datos que estos contienen o pueden contener. Si se va a permitir el acceso a los ficheros de datos a todo tipo de usuarios, ya sea “in situ”, ya sea “on-line”, debemos estar completamente seguros que estos accesos no representan ningún inconveniente para el resto de usuarios del sistema (a nivel de hardware), ni para los datos personales de estos usuarios o los datos de la empresa que permite el acceso.



En la siguiente figura observamos una posible distribución que podría tener la Carnegie Mellon University, con el fin de salvaguardar los datos internos de los empleados, así como los de la propia empresa, permitiendo sin embargo el acceso a su servidor WEB (donde se supone alojada la información referente al proyecto en sí).

Vemos como en la DMZ (Zona Desmilitarizada) se encuentran alojados los servidores a los que se tendrá acceso desde el exterior, vía Internet, mediante el router correspondiente y atravesando el cortafuegos externo. Este cortafuegos, cuyas reglas serán controladas por el administrador de sistemas, sólo permitirá el acceso para todas aquellas operaciones que haya sido habilitado. Habitualmente, los usuarios externos podrán visitar y descargar imágenes, sonidos, videos, texto, etc desde el servidor de la aplicación, pero tendrán denegado el acceso para el resto de operaciones (principalmente, dejar archivos).

Estos servidores de la DMZ, se encuentran duplicados en la Intranet (red interna de la empresa), tras el cortafuegos interno. De este modo, se intenta evitar que cualquier usuario ajeno a la empresa, pueda tener acceso al resto de servidores de la misma (servidores ERP, de correo, etc), y pueda apropiarse o espiar

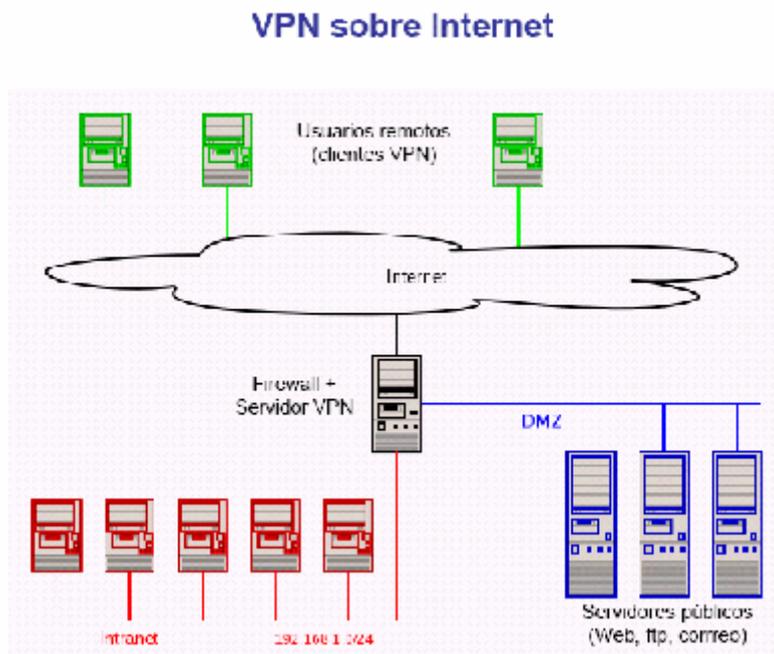
documentación interna de la misma. El administrador informático de la empresa puede decidir incluso, que trabajadores de la misma pueden tener acceso al exterior y que trabajadores no lo tienen.

Así, de una manera tan sencilla, todos los accesos de consulta de información que se realicen sobre los servidores de la Carnegie Mellon University, tanto los internos como los externos, estarán sometidos a unos controles de autenticación (mediante contraseñas, por ejemplo) y se evitarán, en la medida de lo posible, la aparición de “intrusos” en el sistema. La proliferación de virus (troyanos, gusanos, etc) y de hackers (persona que accede on-line a sitios sin autorización, con diferentes fines), provocan cada año, miles de millones de € en pérdidas de horas de trabajo y de documentación.

12.2- Conexión segura mediante VPN (Virtual Private Network)?

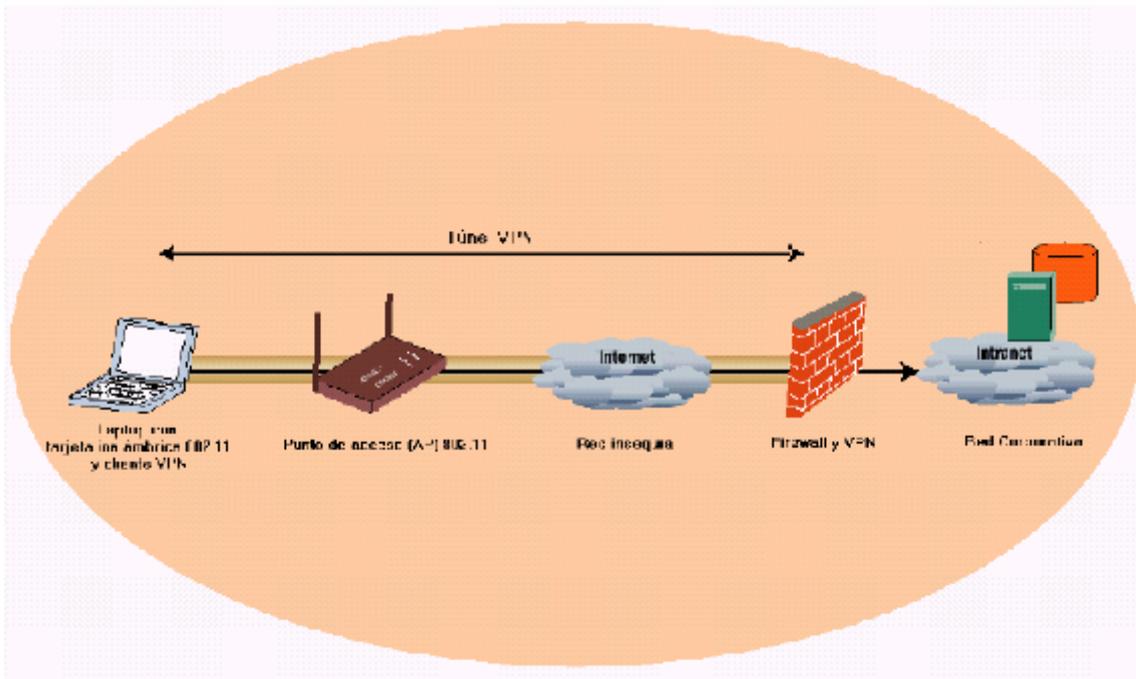
Es una red que usa recursos de transmisión y conmutación “públicos” para llevar tráfico privado en forma segura. Los recursos son públicos en el sentido que son compartidos por muchos usuarios, como es el caso de Internet. Las redes FrameRelay/ATM de los proveedor de servicios de telecomunicaciones también son compartidas, aunque no son tan “públicas” como Internet.

La esencia de VPN reside en que el tráfico privado viaja protegido como si se tratara de una red dedicada, por eso se habla de una red “privada virtual”, ya que no es una red “privada real”. En la siguiente figura podemos ver un pequeño ejemplo de conexión de una VPN.

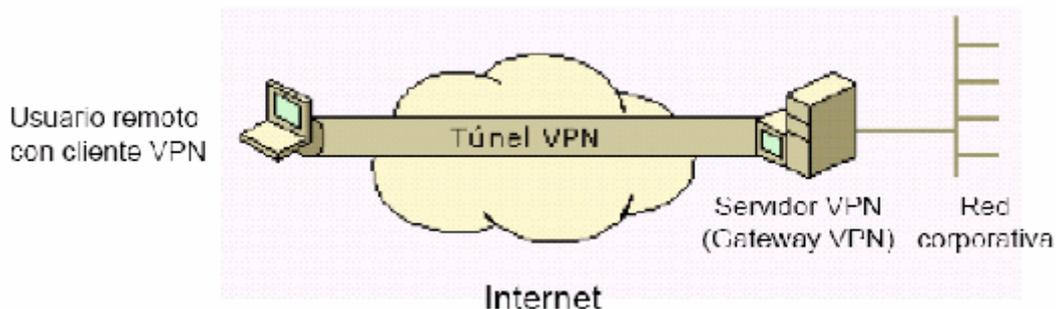


VPN sobre Internet posee ciertas ventajas sobre una red “privada real”. Por ejemplo, VPN ofrece una manera más práctica y económica, para que los tele-trabajadores y comerciales (Ej. ejecutivos de ventas) se puedan conectar en forma segura a la red corporativa.

VPN también es atractiva para conectar en forma segura la sede principal de una organización con sus sucursales remotas. VPN también se usa en redes inalámbricas, que son más inseguras que las redes cableadas convencionales, tal y como se muestra en la figura.



Mediante la tecnología VPN se establecen canales de comunicación seguros a través de Internet (llamados túneles VPN) para usuarios y sedes remotas. Mientras viaja por el túnel, la información está “blindada”, es decir se garantiza su integridad y confidencialidad.



Por todo lo visto hasta ahora, vemos que la tecnología VPN y sus diferentes protocolos, es la más indicada para la conexión de tres de los cuatro tipos del enunciado, quedando descartado el uso de esta tecnología entre las redes internas de la sede central, puesto que el intercambio de información no se realiza a través de una “zona pública”, como en el caso de las otras tres opciones, que siempre deben hacerlo a través de Internet. Más concretamente, para el personal comercial (viajantes), el sistema VPN tunel mediante L2TP/IPSec es el más indicado. A continuación se desarrolla un poco más este tipo de conexión.

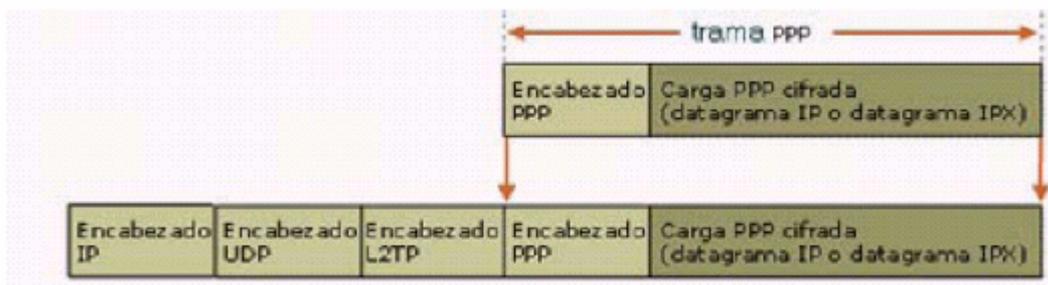
12.2.1- VPN sobre Internet. Inconvenientes

Internet no fue prevista para ofrecer la calidad de servicio que se requiere en muchas aplicaciones, por ejemplo, telefonía sobre IP. Si se transmite información sensible al retardo o urgente, VPN sobre Internet puede no ser la mejor solución, porque puede encontrar problemas de desempeño debido a los retardos.

VPN sobre Internet tampoco garantiza disponibilidad y protección contra ataques de negación de servicio (DoS). Estos aspectos se deben tomar muy en cuenta cuando se vayan a enviar datos importantes a través de una red como Internet en la cual nadie tiene prácticamente el control.

12.2.2- Túneles VPN mediante L2TP (Layer2 Tunneling Protocol)

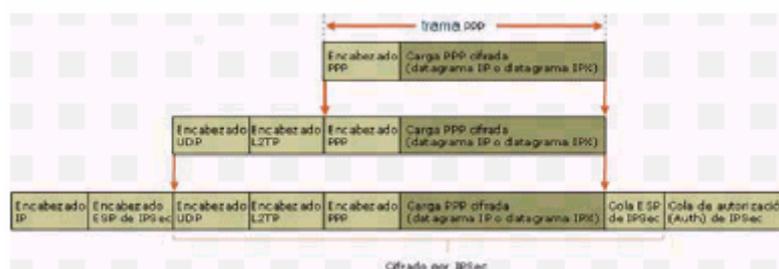
L2TP (Layer2 Tunneling Protocol) es un protocolo de capa 2 que permite la creación de túneles a través de una gran variedad de tipos de redes (IP, SDH, ATM) para el transporte de tráfico PPP (Point to Point Tunnel Protocol).



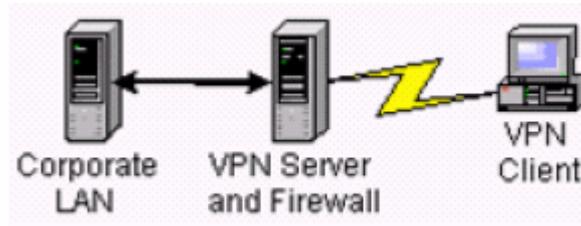
Estructura de un paquete L2TP

Como L2TP no brinda suficiente seguridad, se combina con IPSec para proporcionar un modo sencillo y eficaz de construir el túnel y proteger la información a través de la red. Un requisito habitual es proteger las comunicaciones entre los clientes de acceso remoto y la red corporativa a través de Internet. Puede ser el caso de un ejecutivo de ventas que pasa la mayor parte del tiempo viajando o un empleado que trabaja desde casa.

L2TP encapsula los paquetes originales, primero en una trama PPP (con compresión, si es posible) y después en un mensaje UDP con el puerto 1701. Como el mensaje UDP es una carga de IP, L2TP utiliza IPSec para proteger el túnel.



El firewall debe configurarse para que permita pasar el tráfico VPN. L2TP usa el puerto UDP 1701. IPSec utiliza el identificador de protocolo 51 en el paquete IP para AH y el identificador 50 para ESP. IKE utiliza el puerto UDP 500. Otro aspecto a tener en cuenta es la dificultad de acceso al servidor VPN si se encuentra detrás de un NAT (Network Address Translator), ya que el servidor VPN debe tener una dirección IP válida.



13.- GLOSARIO

A continuación se muestran algunos de los términos o expresiones más utilizados en Internet, computadores, bases de datos, etc.

A

@:arroba (en inglés significa "at" [en]).

En las direcciones de e-mail, es el símbolo que separa el nombre del usuario del nombre de su proveedor de correo electrónico. Por ejemplo: pepe@hotmail.com.

3D tridimensional

Imagen construida con tres dimensiones: largo, ancho y profundidad. Cuando el usuario puede interactuar con imágenes 3D sintiéndose involucrado en la escena, la experiencia se llama realidad virtual.

AAAAA

American Association Against Acronym Abuse: Asociación Americana Contra el Abuso de las Siglas.

Acceso directo

Es un icono que permite abrir más fácilmente un determinado programa o archivo.

Acrobat

Programa de Adobe que permite capturar documentos y verlos en su apariencia original. Acrobat trabaja con archivos PDF.

Administrador de una Web

Operador o persona que se encarga del mantenimiento del sistema de un sitio Web.

ADSL

Asymmetric Digital Subscriber Line. Tecnología para transmitir información digital a elevados anchos de banda. A diferencia del servicio dial up, ADSL provee una conexión permanente y de gran velocidad. Esta tecnología utiliza la mayor parte del canal para enviar información al usuario, y sólo una pequeña parte para recibir información del usuario.

AGP

Puerto acelerador de gráficos. Permite correr velozmente archivos gráficos tridimensionales.

Algoritmo

Conjunto de reglas bien definidas para la resolución de un problema. Un programa de software es la transcripción, en lenguaje de programación, de un algoritmo.

Ancho de banda (bandwidth)

Medida de capacidad de comunicación o velocidad de transmisión de datos de un circuito o canal.

ANSI

American National Standards Institute. Organización que promueve el desarrollo de estándares en los Estados Unidos. Es miembro de la ISO (International Organization for Standardization).

Antivirus

Programa que busca y eventualmente elimina los virus informáticos que pueden haber infectado un disco rígido o disquete.

AOL: America Online

Proveedor de servicios de Internet de los Estados Unidos.

Apache

Servidor web de distribución libre. Fue desarrollado en 1995 y ha llegado a ser el más usado de Internet.

Applet

Mini programa en lenguaje de programación Java integrado en una página web.

Árbol (tree)

Estructura de datos en la cual los registros son almacenados de manera jerárquica.

Archivo adjunto (Attachement)

Archivo que acompaña un mensaje de e-mail. Es apropiado para el envío de imágenes, sonidos, programas y otros archivos grandes.

ARPA

Advanced Research Projects Agency: Agencia del Departamento de Defensa de los Estados Unidos que creó ARPAnet, la red que dio origen a Internet.

ARPANet

Advanced Research Projects Agency Network: Red de comunicación desarrollada por ARPA a fines de la década de los 60. Se la considera el origen de la actual Internet.

Arrastrar y colocar

Concepto de GUI (Interfaz gráfica de usuario) que permite seleccionar un objeto de la pantalla y pasarlo como entrada a otro objeto (icono).

ASCII

American Standard Code of Information Interchange: Código normalizado estadounidense para el intercambio de la información. Código que permite definir caracteres alfanuméricos; usado para lograr compatibilidad entre diversos procesadores de texto.

AutoCad

Programa de dibujo técnico.

AVI

Formato de Microsoft para archivos de audio y video.

B

Backbone

Red de banda ancha para conexiones entre conmutadores. Ruta/circuito de comunicaciones de capacidad media. Suele indicar una velocidad de 64000 bps a 1544 Mbps.

Back-end processor

Procesador que se utiliza para determinada función muy especializada, como por ejemplo, administrar una base de datos.

Backup

Copia de seguridad. Se hace para prevenir una posible pérdida de información.

Banda ancha

Ruta/circuito de comunicaciones de gran capacidad. Normalmente implica una velocidad superior a 1544 Mbps.

Bandeja de entrada

Buzón de entrada de una utilidad de correo electrónico.

Banner

Gráfico, generalmente rectangular, que se inserta en una página web. Puede tener carácter publicitario.

Base de datos

Conjunto de información para varios usuarios. Suele admitir la selección de acceso aleatorio y múltiples "vistas" o niveles de abstracción de los datos subyacentes.

BASIC

Beginner's All-Purpose Symbolic Instruction Code: Código de Instrucción Simbólica Multipropósito para Principiantes. Lenguaje de programación, creado en 1963, sencillo y muy difundido.

Baudio

Número de elementos de señalización que pueden transmitirse por segundo en un circuito. (Término antiguo que se está reemplazando por bps - bits por segundo).

Benchmark

Programa especialmente diseñado para evaluar el rendimiento de un sistema, de software o de hardware.

BIOS

Basic Input/Output System: Sistema básico de ingreso/salida de datos. Conjunto de procedimientos que controla el flujo de datos entre el sistema operativo y dispositivos tales como el disco rígido, la placa de video, el teclado, el mouse y la impresora.

Bit

Abreviatura de binary digit (dígito binario). Cantidad de información más pequeña que puede transmitirse. Una combinación de bits puede indicar un carácter alfabético, un dígito, una señal, un modificador u otras funciones.

Bookmark

Anotación, en el navegador, de una dirección de Internet que se almacena para agilizar su uso posterior. En el programa Internet Explorer, se llama "Favoritos".

Boot (butear)

Cargar el sistema operativo de una computadora.

BPS

Bits por segundo. Medida de velocidad de un módem.

Buffer

Área de la memoria que se utiliza para almacenar datos temporalmente durante una sesión de trabajo.

Bug

Bicho, insecto. Error de programación que genera problemas en las operaciones de una computadora.

Bus

Enlace común; conductor común; vía de interconexión. Método de interconexión de dispositivos mediante una sola línea compartida.

Bus serial

Método de transmisión de un BIT por vez sobre una sola línea.

Buscador (motor de búsqueda)

Search engine. Es un programa, ubicado en un sitio de Internet, que recibe un pedido de búsqueda, lo compara con las entradas de su base de datos y devuelve el resultado. Algunos de los más conocidos: Yahoo, Altavista, Lycos, Infoseek.

Byte

Unidad de información utilizada por las computadoras. Cada byte está compuesto por ocho bits.

C

Cable coaxial

Es el tipo de cable usado por las compañías de televisión por cable para establecer la conexión entre la central emisora y el usuario. También se lo utiliza mucho en las conexiones de redes de área local (LAN). Según el tipo de tecnología que se use, se lo puede reemplazar por fibra óptica.

Caché de disco

Pequeña porción de memoria RAM que almacena datos recientemente leídos, con lo cual agiliza el acceso futuro a los mismos datos.

CAD (Computer Aided Design)

Diseño Asistido por Computadora. Software que permite crear dibujos de precisión, bidimensionales y tridimensionales. Lo usan principalmente arquitectos e ingenieros.

Capacidad

La mayor velocidad de transmisión posible (fiable) que puede darse en un canal, un circuito o una pieza de equipo. La capacidad puede expresarse como la velocidad bruta o como el rendimiento neto. Término aplicado también para medir la cantidad de información que puede almacenarse en un dispositivo.

Carácter

Número, letra o símbolo en la computadora, conformado por un byte.

CD-ROM (Compact Disk - Read Only Memory)

Disco compacto de sólo lectura. Tiene una capacidad de almacenamiento de hasta 650 megabytes, mucho mayor que la de un disquete.

Celeron

Microprocesador de la familia Intel, de menor costo que el Pentium II.

CGI (Interfaz de gateway común)

Interfaz para programadores que crean archivos de comandos o aplicaciones que se ejecutan internamente en un servidor de Web. Estos archivos de comandos pueden generar texto y otros tipos de datos de forma inmediata, en respuesta a una entrada del usuario, o bien tomando la información de una base de datos.

Chip

Abreviatura de "microchip". Circuito muy pequeño, compuesto por miles a millones de transistores impresos sobre una oblea de silicio.

Ciberespacio

Se refiere al campo colectivo de la comunicación asistida mediante equipos informáticos.

Cluster

Grupo; racimo; agrupamiento. En la tecnología de las computadoras, un cluster es la unidad de almacenamiento en el disco rígido. Un archivo está compuesto por varios clusters, que pueden estar almacenados en diversos lugares del disco.

Comando (command)

Instrucción que un usuario da al sistema operativo de la computadora para realizar determinada tarea.

Compresión/Descompresión

Método para reducir el tamaño de un archivo para ahorrar espacio o para transmitirlo a mayor velocidad. Consiste en cifrar/descifrar señales que permite transmitir (o almacenar) más información de la que, de otro modo, podría aceptar el soporte. Uno de los programas de compresión más populares de Windows es WinZip.

Cookie

Pequeño archivo de texto que un sitio web coloca en el disco rígido de una computadora que lo visita. Al mismo tiempo, recoge información sobre el usuario. Agiliza la navegación en el sitio. Su uso es controvertido, porque pone en riesgo la privacidad de los usuarios.

CPU

Central Processing Unit. Unidad central de procesamiento. Es el procesador que contiene los circuitos lógicos que realizan las instrucciones de la computadora.

Cuello de botella

Límite en la capacidad del sistema que puede reducir el tráfico en condiciones de sobrecarga.

Cursor

Símbolo en pantalla que indica la posición activa: por ejemplo, la posición en que aparecerá el próximo carácter que entre.

D

Data

Datos, información.

Database

Base de datos.

Debugging

Depuración, corrección de errores o bugs.

Delete

Borrar; eliminar; anular.

Digital

Dispositivo o método que utiliza variaciones discretas en voltaje, frecuencia, amplitud, ubicación, etc. para cifrar, procesar o transportar señales binarias (0 o 1) para datos informáticos, sonido, vídeo u otra información.

Dirección

Código exclusivo asignado a la ubicación de un archivo almacenado, un dispositivo en un sistema o red, o cualquier origen de datos de una red.

Dirección IP

Dirección de 32 bits del protocolo Internet asignada a un host. La dirección IP tiene un componente del host y un componente de la red.

Dirección URL (Uniform Resource Locator)

Formato de las direcciones de sitios que muestra el nombre del servidor en el que se almacenan los archivos del sitio, la ruta de acceso al directorio del archivo y su nombre.

Directorio (directory)

Grupo de archivos relacionados entre sí que se guardan bajo un nombre.

Disco rígido

Soporte giratorio de almacenamiento en forma de placa circular revestida por una película magnética. Los datos se graban en pistas concéntricas en la película.

Display

Unidad de visualización; monitor; pantalla.

Download

Descargar, bajar. Transferencia de información desde Internet a una computadora.

Dpi

(Dots per inch). Puntos por pulgada. En las impresoras, la calidad de la imagen sobre el papel se expresa en dpi.

DSL

Digital Subscriber Line: Línea Digital de Suscripción. Tecnología que permite enviar mucha información a gran velocidad a través de líneas telefónicas.

DVD

Digital Versatile Disc: Disco Versátil Digital. Disco que posee gran capacidad de almacenamiento, hasta 4,7 Gb y sirve también para almacenar películas.

Dynamic HTML

Variante del HTML (Hyper TextMark-up Language) que permite crear páginas web más animadas.

E

e-mail

Correo electrónico.

Emulación

(Emulation). Proceso de compatibilización entre computadoras mediante un software.

Encriptar

Proteger archivos expresando su contenido en un lenguaje cifrado. Los lenguajes cifrados simples consisten, por ejemplo, en la sustitución de letras por números.

Ergonomía

Consideración del elemento humano en el diseño de ingeniería. Estudio del diseño de dispositivos (por ejemplo, teclados), con el objetivo de que las personas los usen de manera saludable, cómoda y eficiente.

Explorador

Programa de aplicación que proporciona una interfaz gráfica interactiva para buscar localizar, ver y administrar la información a través de una red.

Extranet

Parte de una intranet de acceso disponible a clientes y otros usuarios ajenos a la compañía.

F

FAQ

Frequently asked questions. Las preguntas más frecuentes (y sus respuestas) sobre el tema principal de un sitio web.

Fibra óptica

Tecnología para transmitir información como pulsos luminosos a través de un conducto de fibra de vidrio. La fibra óptica transporta mucha más información que el cable de cobre convencional. La mayoría de las líneas de larga distancia de las compañías telefónicas utilizan la fibra óptica.

Finger

Protocolo que permite localizar información sobre los usuarios en la red del host. Algunas redes no permiten su uso desde un sistema externo, y otras no lo permiten en absoluto.

Firewall

Mecanismo de seguridad que impide el acceso a una red.

FTP (Protocolo de transferencia de archivos)

Protocolo utilizado para transferir archivos a través de una amplia variedad de sistemas.

Fuente

Variedad completa de caracteres de imprenta de un determinado estilo.

G

Gateway

Puerta; acceso; pasarela Conversor de protocolos. Nodo específico de la aplicación que conecta redes que de otra forma serían incompatibles. Convierte códigos de datos y protocolos de transmisión que permiten la interoperatividad.

GIF

Formato de intercambio de gráficos. GIF es un formato estándar para archivos de imágenes en WWW. El formato de archivos GIF es muy común, ya que utiliza un método de compresión para reducir los archivos.

GIF animado

Variante del formato GIF. Se usa en la World Wide Web para dar movimiento a íconos y banners.

Giga Gigabyte (GB)

Prefijo que indica un múltiplo de 1.000 millones, o sea 10⁹. Cuando se emplea el sistema binario, como ocurre en informática, significa un múltiplo de 2³⁰, o sea 1 gigabyte = 1024 megabytes = 1.073.741.824.

Gigaflop

Medida de velocidad de una computadora equivalente a 1.000 millones de operaciones de coma flotante por segundo.

GNU

Conjunto de programas desarrollados por la Free Software Foundation (Fundación por el Software Libre); es de uso libre.

Gopher

Programa de búsqueda y exploración de bases de datos públicas en Internet.

GPS

Global Positioning System. Sistema de localización global compuesto por 24 satélites. Se usa, por ejemplo, en automóviles, para indicarle al conductor dónde se encuentra y sugerirle rutas posibles.

Grid

Cuadrícula para representar conjuntos de datos en forma de tabla.

GUI

Interfaz gráfica de usuario.

Gusano

Programa que se copia a sí mismo hasta ocupar toda la memoria. Es un virus que suele llegar a través del correo electrónico, en forma de archivo adjunto.

H

Hard disk

Disco rígido. Disco Duro.

Hardware

Todos los componentes físicos de la computadora y sus periféricos.

Hertz

Hercio. Unidad de frecuencia electromagnética. Equivale a un ciclo por segundo.

Hipermedio

Método para presentar información en unidades discretas, o nodos, que están conectados mediante vínculos. La información puede presentarse utilizando distintos medios, como documentación ejecutable, de texto, gráficos, audio, vídeo, animación o imagen.

Hipertexto

Textos enlazados entre sí. Haciendo clic con el mouse el usuario pasa de un texto a otro, vinculado con el anterior. Describe un tipo de funcionalidad de exploración en línea interactiva. Los vínculos (direcciones URL) incrustados en palabras o frases permiten al usuario seleccionar texto (p. ej. haciendo clic con el mouse) y mostrar inmediatamente información relacionada y material multimedia.

Hipervínculo : link.

Conexiones entre una información y otra.

Hosting

Alojamiento. Servicio ofrecido por algunos proveedores, que brindan a sus clientes (individuos o empresas) un espacio en su servidor para alojar un sitio web.

HTML (Hypertext Markup Language)

Lenguaje de "etiquetas" en el que se asigna formato a las páginas de Web y se distribuye la información.

HTTP (Protocolo de transferencia de hipertexto)

Método mediante el que se transfieren documentos desde el sistema host o servidor a los exploradores y usuarios individuales. Es un protocolo que permite transferir información en archivos de texto, gráficos, de video, de audio y otros recursos multimedia.

Ícono

Imagen que representa un programa u otro recurso; generalmente conduce a abrir un programa.

IEEE 1394

Tecnología para la transmisión de datos a alta velocidad. Fue desarrollada en 1986 por Apple Computer, con el nombre de FireWire. En 1995 fue adoptada como estándar. Es similar al USB, pero más veloz. Permite conectar múltiples dispositivos.

IEEE

Institute of Electrical and Electronics Engineers: importante asociación de técnicos y profesionales, con sede en los Estados Unidos. Fue fundada en 1884 y en 1998 tenía aproximadamente 320.000 miembros en 147 países. Favorece la investigación en campos diversos, como la tecnología aeroespacial, la computación, las comunicaciones y la tecnología biomédica. Promueve la estandarización de normas.

Image map

Imagen de una página web que permite clicar en diferentes áreas para acceder a diferentes destinos.

Importar

Incorporar un objeto desde otro programa.

Impresora de chorro de tinta

Impresora que trabaja pulverizando la tinta sobre el papel.

Impresora de matriz de puntos

Impresora que trabaja por medio de un cabezal que presiona una cinta entintada contra el papel.

Impresora láser

Impresora veloz y de alta resolución que utiliza la tecnología de rayos láser. Cuando el rayo toca el papel, forma una imagen electrostática que atrae la tinta seca.

Impresora matricial

impresora de matriz de puntos

Impresora

dispositivo periférico que reproduce textos e imágenes en papel. Los principales tipos son: de matriz de puntos, de chorro de tinta y láser.

Inteligencia artificial

Simulación de los procesos de la inteligencia humana, por medio de sistemas de computación.

Internet

Red de redes. Sistema mundial de redes de computadoras interconectadas. Fue concebida a fines de la década de 1960 por el Departamento de Defensa de los Estados Unidos; más precisamente, por la ARPA. Se la llamó primero ARPAnet y fue pensada para cumplir funciones de investigación. Su uso se popularizó a partir de la creación de la World Wide Web. Actualmente es un espacio público utilizado por millones de personas en todo el mundo como herramienta de comunicación e información.

Internet2

Proyecto de interconexión de más de 100 universidades estadounidenses. El objetivo es desarrollar una red de altísima velocidad para la educación y la investigación.

Intranet

Red de redes de una empresa. Su aspecto es similar al de las páginas de Internet.

IP (Protocolo Internet)

Define la unidad de información enviada entre sistemas, que proporciona un servicio de entrega de paquetes básico.

IrDA (Infrared Data Association)

Organización fundada para crear las normas internacionales para el hardware y el software usados en enlaces de comunicación por rayos infrarrojos. La tecnología de rayos infrarrojos juega un importante papel en las comunicaciones inalámbricas.

ISDN Integrated Services Digital Network

(Red digital de servicios integrados) También llamada RDSI. Juego de normas de la transmisión a gran velocidad de información simultánea de voz, datos e información a través de menos canales de los que serían necesarios de otro modo, mediante el uso de la señalización fuera de banda.

ISO International Organization for Standardization.

Fundada en 1946, es una federación internacional que unifica normas en unos cien países. Una de ellas es la norma OSI, modelo de referencia universal para protocolos de comunicación.

ISP

Internet Service Provider. Proveedor de servicios de Internet.

J

Jitter

Variación en la cantidad de latencia entre paquetes de datos recibidos.

Joystick

Dispositivo para manejar ciertas funciones de las computadoras, especialmente en juegos.

JPEG Joint Photographic Experts Group

Nombre del comité que diseñó un estándar para la compresión de imágenes. Se trata de un conocido método para comprimir imágenes fotográficas. Muchos exploradores del Web aceptan imágenes JPEG como el formato de archivo estándar para la visualización

K

Kernel

Núcleo o parte esencial de un sistema operativo. Provee los servicios básicos del resto del sistema.

Keyboard

Teclado.

Kilobyte (KB)

Unidad de medida de una memoria. 1 kilobyte = 1024 bytes.

L

LAN Manager

Sistema operativo de red.

LAN Local Area Network

Red de Área Local. Red de computadoras interconectadas en un área reducida, por ejemplo, una empresa.

Laptop

Computadora portátil del tamaño aproximado de un portafolio. Hay otras más pequeñas: palmtop y handheld, por ejemplo.

Latencia

Lapso necesario para que un paquete de información viaje desde la fuente hasta su destino. La latencia y el ancho de banda, juntos, definen la capacidad y la velocidad de una red.

LCD

Liquid Crystal Display. Pantalla de cristal líquido, usada generalmente en las notebooks y otras computadoras pequeñas.

lenguaje de programación

Sistema de escritura para la descripción precisa de algoritmos o programas informáticos.

Línea dedicada

Línea privada alquilada a una empresa de telecomunicaciones.

Link

Enlace. Imagen o texto destacado, mediante subrayado o color, que lleva a otro sector del documento o a otra página web.

Linux

Sistema operativo gratuito para computadoras personales derivado de Unix.

login name

Nombre de identificación del usuario en un sistema online.

LPT Line Print Terminal

Conexión entre una computadora personal y una impresora u otro dispositivo. Es un puerto paralelo y es más veloz que un puerto serial.

M

Mac OS

Sistema operativo de las computadoras personales y las workstations de Macintosh.

Macintosh

Computadora que Apple empezó a fabricar en 1984. Fue la primera computadora personal que incorporó una interfase gráfica, con el propósito de facilitar un uso más intuitivo de la máquina. Tiene su propio sistema operativo, llamado Mac OS. El uso de la Macintosh está muy difundido entre diseñadores gráficos, artistas visuales y músicos.

Mailing list

Lista de correo. Grupo de personas suscriptas a una discusión periódica por e-mail sobre determinado tema.

Mainframe

Estructura principal. Computadora de gran tamaño de tipo multiusuario, utilizada en empresas.

Megabyte (MB)

Unidad de medida de una memoria. 1 megabyte = 1024 kilobytes = 1.048.576 bytes.

Megahertz (MHz)

Un millón de hertz o hercios.

Memoria flash

Tipo de memoria que puede ser borrada y reprogramada en unidades de memoria llamadas "bloques". Su nombre se debe a que el microchip permite borrar fragmentos de memoria en una sola acción, o "flash". Se utiliza en teléfonos celulares, cámaras digitales y otros dispositivos.

Mensajería instantánea

Sistema que permite enviar mensajes de manera inmediata en ciertos programas de chat, como ICQ y PowWow.

Microprocesador (microprocessor)

Es el chip más importante de una computadora. Su velocidad se mide en MHz (Megahertz).

Milisegundo

Milésima parte de un segundo.

Mini disc

Pequeño disco diseñado por Sony para almacenar video o música en un dispositivo portátil.

Mini DV

Pequeño cassette de cinta para video digital.

Mirror site

Sitio espejo. Sitio web copiado a otro servidor con el propósito de facilitar el acceso a sus contenidos desde el lugar más cercano o más conveniente para el usuario.

MMX: MultiMedia eXtension.

Microprocesador Pentium diseñado para dar mayor velocidad a aplicaciones multimedia.

Módem

Modulador-demodulador. Dispositivo periférico que conecta la computadora a la línea telefónica. permite transmitir datos digitales a través de dispositivos de transmisión analógicos, como las líneas telefónicas

Módem de 56 K V.90

módem que responde a la norma V.90, el estándar que la ITU (International Telecommunications Union) fijó para los módem de 56 K.

Mosaic

Fue el primer navegador que utilizó una interfase gráfica. Fue creado en 1993 por Marc Andreessen. Contribuyó a la enorme difusión de la World Wide Web.

Movie player

Tecnología que combina texto, sonido, animación y video en un solo archivo. Por ejemplo, el programa Quick Time de la empresa Apple. MP1,2,3 y 4: tecnología y formatos para comprimir audio y video con alta calidad de emisión.

MPEG 1,2,3 y 4

Tecnología y formatos para comprimir audio y video con alta calidad de emisión.

MPEG Moving Pictures Expert Group

Desarrolla estándares para video digital y compresión de audio. Se trata de un modo estándar de comprimir vídeo de imágenes en movimiento. Tiene el auspicio de la ISO.

MS-DOS

Microsoft Disk Operating System: Sistema operativo del Disco Microsoft.

Multicast

Comunicación entre un solo emisor y múltiples receptores en una red.

Multimedia

Sistemas informáticos que integran audio, vídeo y datos.

N

Nano

Prefijo que significa una milmillonésima parte.

Nanosegundo

Una milmillonésima de segundo. Es una medida común del tiempo de acceso a la memoria RAM.

Navegador

Programa para recorrer la World Wide Web. Algunos de los más conocidos son Netscape Navigator, Microsoft Explorer, Opera y Neoplanet.

NCSA

El National Center for Supercomputing Applications es una institución de formación. El explorador de Web Mosaic se creó aquí. <http://www.ncsa.uiuc.edu/>

Netscape

Navegador desarrollado en 1995 por un equipo liderado por Marc Andreessen, el creador de Mosaic.

Network

Red.

Newsgroup

Grupo de discusión sobre determinado tema, en Internet u otras redes.

Ñ

Ninguna entrada

O

Office

Suite de Microsoft para trabajo de oficina; incluye procesador de texto, base de datos y planilla de cálculo.

Online

En línea, conectado. Estado en que se encuentra una computadora cuando se conecta directamente con la red a través de un dispositivo, por ejemplo, un módem.

OSI

Open Systems Interconnection: Interconexión de Sistemas Abiertos. Norma universal para protocolos de comunicación.

P

Página

Documento de hipermedia en el Web.

Página web

Una de las páginas que componen un sitio de la World Wide Web. Un sitio web agrupa un conjunto de páginas afines. A la página de inicio se la llama "home page".

Paquete (packet)

La parte de un mensaje que se transmite por una red. Antes de ser enviada a través de Internet, la información se divide en paquetes.

Password

Contraseña.

Pay per view

Pagar para ver. En Direct TV, o televisión satelital, es un sistema por el cual se paga para ver ciertos programas. Es preciso tener un decodificador o un receptor de televisión por satélite. En Internet también se accede a determinadas informaciones por el sistema "pay per view".

PCMCIA Personal Computer Memory Card International Association

Tarjetas de expansión de memoria que aumentan la capacidad de almacenamiento.

PDA Personal Digital Assistant.

Pequeña agenda personal electrónica.

PDF

Portable Document Format. Formato de archivo que captura un documento impreso y lo reproduce en su apariencia original. Los archivos PDF se crean con el programa Acrobat.

Performance

Desempeño, rendimiento.

Periférico

Todo dispositivo que se conecta a la computadora. Por ejemplo: teclado, monitor, mouse, impresora, escáner, etcétera.

Píxel

Combinación de "picture" y "element". Elemento gráfico mínimo con el que se componen las imágenes en la pantalla de una computadora.

PKZIP

PKZIP es una utilidad compartida de compresión para PCs. Para descomprimir estos archivos se utiliza un programa llamado PKUNZIP.

Placa aceleradora de gráficos

Circuito que se agrega a una computadora para mejorar los recursos gráficos y darles más velocidad.

Placa aceleradora

Circuito que se agrega a una computadora para aumentar su velocidad.

Placa de sonido

La que proporciona sonido a una computadora. Una de las más conocidas es Sound Blaster.

Placa Ethernet

Placa que se inserta en una computadora para conectarla en red con otras a través de un cable.

Placa madre

Motherboard. Placa electrónica donde se aloja el microprocesador y el resto de placas funcionales de un computador.

Placa

Tarjeta que se inserta en un slot de la motherboard para expandir la capacidad de una computadora.

Player

Programa que permite escuchar archivos de sonido.

Plug & play

Significa "enchufar y usar". Reconocimiento inmediato de un dispositivo por parte de la computadora, sin necesidad de instrucciones del usuario.

Plug-in

Programa que puede ser instalado y usado como parte del navegador. Un ejemplo es Macromedia's Shockwave, que permite reproducir sonidos y animaciones.

PoP

Point of Presence. Punto de acceso a Internet.

POP3 (Post Office Protocol 3)

Protocolo 3 de Correo. Es un protocolo estándar para recibir e-mail. Conexión de acceso telefónico de los proveedores de servicios de Internet para usuarios de módem, que se utiliza principalmente para describir conexiones locales, de forma que los usuarios no tengan que hacer llamadas de larga distancia. Por ejemplo, un determinado ISP puede tener su base en San Jose, pero tener "POP" en Los Angeles y Nueva York.

Portal

Sitio web que sirve de punto de partida para navegar por Internet. Los portales ofrecen una gran diversidad de servicios: listado de sitios web, buscador, noticias, e-mail, información meteorológica, chat, newsgroups (grupos de discusión) y comercio electrónico. En muchos casos el usuario puede personalizar la presentación del portal. Algunos de los más conocidos son Altavista, Yahoo!, Netscape y Microsoft.

PPP (Protocolo punto a punto)

Conexión a Internet de acceso telefónico que utiliza el protocolo TCP/IP; algo más rápido que SLIP.

Pretty Good Privacy

Programa usado para encriptar y desencriptar correo electrónico, a fin de proteger la privacidad. También se puede usar para archivos de otro tipo.

Privilegios de acceso

Privilegio para tener acceso a carpetas y hacer cambios en ellas.

Procesador (processor)

Conjunto de circuitos lógicos que procesa las instrucciones básicas de una computadora.

Protocolo

Lenguaje que utilizan dos computadoras para comunicarse entre sí.

Puerto infrarrojo IrDA

Puerto para comunicación inalámbrica que usa el standard IrDa.

Puerto paralelo

Conexión por medio de la cual se envían datos a través de varios conductos. Una computadora suele tener un puerto paralelo llamado LPT1.

Puerto serie

Conexión por medio de la cual se envían datos a través de un solo conducto. Por ejemplo, el mouse se conecta a un puerto serie. Las computadoras tienen dos puertos series: COM1 y COM2.

Puerto

En una computadora, es el lugar específico de conexión con otro dispositivo, generalmente mediante un enchufe. Puede tratarse de un puerto serie o de un puerto paralelo.

Puntero

Dirección URL incrustada en los datos que especifica su ubicación en otro registro o archivo. El hipervínculo es un ejemplo de puntero.

Q

QBL Query By Example

Consulta por ejemplo. Método de consulta para la base de datos.

QL Query Language

Lenguaje de consulta.

Quality of Service

Calidad de servicio. En Internet y otras redes, designa la posibilidad de medir, mejorar y, en alguna medida, garantizar por adelantado los índices de transmisión y error. Es importante para la transmisión fluida de información multimedia: por ejemplo, para los usos académicos de Internet2.

Query

Consulta. Búsqueda en una base de datos.

R

RAM Random Acces Memory

Memoria de acceso aleatorio. Memoria donde la computadora almacena datos que le permiten al procesador acceder rápidamente al sistema operativo, las aplicaciones y los datos en uso. Tiene estrecha relación con la velocidad de la computadora. Se mide en megabytes.

Realidad virtual

Simulación de un medio ambiente real o imaginario que se puede experimentar visualmente en tres dimensiones. La realidad virtual puede además proporcionar una experiencia interactiva de percepción táctil, sonora y de movimiento.

Reconocimiento de voz

Capacidad de un programa para interpretar palabras emitidas en voz alta o ejecutar un comando verbal.

Red

Sistema de elementos interrelacionados que se conectan mediante un vínculo dedicado o conmutado para proporcionar una comunicación local o remota (de voz, vídeo, datos, etc.) y facilitar el intercambio de información entre usuarios con intereses comunes. En tecnología de la información, una red es un conjunto de dos o más computadoras interconectadas.

Red de área local

LAN.

Resolución

Número máximo de píxeles que se ven en una pantalla. Dos ejemplos: 800 x 600 y 640 x 480. / En una impresora, la resolución es la calidad de la imagen reproducida y se mide en dpi.

Ripper

Programa que permite copiar sonido desde un CD al disco rígido, como archivo con extensión wav. Para convertir un .wav en un MP3, es preciso usar un programa conocido como encoder. Para escucharlos, es necesario un programa de tipo player. Existen suites que ofrecen los tres programas.

Roaming

Acceso a Internet desde diversos lugares del mundo, al precio de una llamada local.

Robot

Los "Robots" suelen mencionarse en el contexto de WWW como programas que se mueven por el Web buscando información; por ejemplo, para crear índices en dispositivos de búsqueda o localizar errores en sitios Web o funciones similares.

ROM Read Only Memory

Memoria de sólo lectura. Memoria incorporada que contiene datos que no pueden ser modificados. Permite a la computadora arrancar. A diferencia de la RAM, los datos de la memoria ROM no se pierden al *-apagar el equipo.

Router

Ruteador. Sistema constituido por hardware y software para la transmisión de datos en Internet. El emisor y el receptor deben utilizar el mismo protocolo.

S

Salvapantalla

Protector de pantalla.

Scanner

Escáner.

SDRAM

Memoria muy rápida, de gran capacidad, para servidores y estaciones de trabajo.

Search engine

Buscador, motor de búsqueda.

Seguridad

Mecanismos de control que evitan el uso no autorizado de recursos.

Semiconductor

Se llama así a las sustancias aislantes, como el germanio y el silicio, que se transforman en conductores por la adición de determinadas impurezas. Los semiconductores tienen enorme importancia en electrónica.

Señal

Cambio de estado orientado a eventos (p. ej. un tono, cambio de frecuencia, valor binario, alarma, mensaje, etc.).

ScanDisk

Programa de Windows que revisa un disco, detecta errores y los corrige.

Servidor

En una red, computadora central de un sistema de red que provee servicios y programas a otras computadoras conectadas.

Servidor de archivos

Sistema informático que permite a usuarios remotos (clientes) tener acceso a archivos.

SGML

Standard Generalized Markup Language. Lenguaje para la descripción de otros lenguajes de documentos estructurales basados en etiquetas. Por ejemplo, el HTML está definido mediante el SGML.

Shareware

Software distribuido en calidad de prueba. Al cabo de cierto tiempo de uso (generalmente 30 días) el usuario tiene la opción de comprarlo.

Sistema operativo

Programa que administra los demás programas en una computadora.

Sitio

Ubicación de la dirección de un servidor en Internet. véase página web

SLIP (Protocolo Internet de línea en serie)

Conexión de acceso telefónico a Internet que utiliza el protocolo TCP/IP.

Slot

Ranura de la motherboard que permite expandir la capacidad de una computadora insertándole placas.

SMS Short Message Service

Servicio de mensajería para teléfonos móviles. Permite enviar a un celular un mensaje de hasta 160 caracteres. Este servicio fue habilitado inicialmente en Europa. Hay varios sitios web desde los cuales se puede enviar un "sms".

SMTP Simple Mail Transfer Protocol.

Es un protocolo estándar para enviar e-mail.

SNA System Network Architecture

Arquitectura de red para mainframes, desarrollada por IBM.

Sniffer

Programa que monitorea y analiza el tráfico de una red para detectar problemas o cuellos de botella. Su objetivo es mantener la eficiencia del tráfico de datos. Pero también puede ser usado ilegítimamente para capturar datos en una red.

Software

Término general que designa los diversos tipos de programas usados en computación.

Spam

Correo electrónico no solicitado. Se lo considera poco ético, ya que el receptor paga por estar conectado a Internet.

Socket

(Soporte). Conector eléctrico, toma de corriente, enchufe. / Un socket es el punto final de una conexión. / Método de comunicación entre un programa cliente y un programa servidor en una red (véase cliente / servidor).

Soporte

Formato de distribución y almacenamiento de información (p. ej. cinta de vídeo, disquete, disco óptico, impresora, etc.). Una ampliación de la capacidad de comunicación de la humanidad. Es el mensaje.

SQL Structured Query Language.

Lenguaje de programación que se utiliza para recuperar y actualizar la información contenida en una base de datos. Fue desarrollado en los años 70 por IBM. Se ha convertido en un estándar ISO y ANSI.

SSL

Secure Sockets Layer. Protocolo diseñado por la empresa Netscape para proveer comunicaciones encriptadas (transacciones seguras) en Internet.

Suite

Serie, conjunto. Conjunto de programas que se comercializan en un solo paquete. socket: (soporte) conector eléctrico, toma de corriente, enchufe. / Un socket es el punto final de una conexión. / Método de comunicación entre un programa cliente y un programa servidor en una red (véase cliente/servidor).

Superautopista de la información

Una palabra de moda (véase también media, hiper...). Hace referencia al plan de la administración de Clinton/Gore para liberalizar los servicios de comunicación, permitiendo la integración de todos los aspectos de Internet, CATV, teléfono, empresas, ocio, proveedores de información, educación, etc.

Superdisk

Disquetera en la cual se pueden usar los disquetes de 120 MB y también los anteriores, de 1.44 MB. Hay internas y externas (estas últimas, sobre puerto paralelo, USB y PCMCIA).

Supervideo

Tecnología para transmitir señales de video dividiéndolas en dos: color y luminosidad.

SVGA Super Video Graphic Array.

Monitor superior en definición al VGA. Según las características de la computadora, puede soportar una paleta de hasta 16.777.216 colores.

T

Tarjeta CompactFlash

es la tarjeta de memoria flash más pequeña, más duradera y de mayor capacidad (octubre de 1999). Tiene controlador incorporado y viene en tamaños desde 2 MB hasta 64 MB.

TCP/IP Transfer Control Protocol / Internet Protocol

Es el protocolo que se utiliza en Internet. Protocolo de control de transmisiones/Protocolo Internet. Es el protocolo estándar de comunicaciones en red utilizado para conectar sistemas informáticos a través de Internet.

TLD Top Level Domain

Véase DNS.

Telemática

Combinación de las palabras "telecomunicaciones" e "informática". Disciplina que asocia las telecomunicaciones con los recursos de la informática.

Telnet

Programa de red que ofrece una forma de conectarse y trabajar desde otro equipo. Al conectarse a otro sistema, los usuarios pueden tener acceso a servicios de Internet que quizás no tengan en sus propios equipos.

Tiempo real

Rápida transmisión y proceso de datos orientados a eventos y transacciones a medida que se producen, en contraposición a almacenarse y retransmitirse o procesarse por lotes.

Touch pad

Pequeña superficie sensible al tacto, incorporada al teclado de una computadora. Cumple las mismas funciones que el mouse.

Touch screen

Pantalla sensible al tacto. Se basa en la utilización de rayos infrarrojos. Cuando el usuario toca la pantalla, genera una señal electrónica; el software interpreta la señal y realiza la operación solicitada.

Transferir

Trasladar programas o datos de equipos informáticos a dispositivos conectados, normalmente de servidores a PCs.

Troyano (Trojan horse; caballo de Troya)

Programa que contiene un código dañino dentro de datos aparentemente inofensivos. Puede arruinar parte del disco rígido.

U

Unix

Sistema operativo multiusuario libre, fue muy importante en el desarrollo de Internet.

Unplugged

Desenchufado; que funciona sin cables.

UNZIP

Descomprimir o expandir un archivo que se había reducido mediante una utilidad de compresión.

URL

Uniform Resource Locator. Véase DNS.

USB (Universal Serial Bus)

Es una interfase de tipo plug & play entre una computadora y ciertos dispositivos, por ejemplo, teclados, teléfonos, escáners e impresoras.

Usenet (USEer NETwork)

Grupos de debate de Internet. Hay miles de estos foros de debate, y cualquier usuario puede crear uno nuevo. En algunos casos los foros tienen un moderador que filtra, edita y envía los mensajes. Uno de los primeros formatos de "correo electrónico colectivo". Actualmente hay unos 10000 grupos de debate diferentes.

V

VHS

Video Home System. La norma más usada en grabación y reproducción de video hogareño.

Videoconferencia

Conversación entre dos o más personas que se encuentran en lugares diferentes pero pueden verse y oírse. Las videoconferencias que se realizan fuera de Internet requieren que en cada lugar donde se encuentran los participantes se disponga de una videocámara especial y de dispositivos para presentación de documentos. En la Web, productos como CU-SeeMe permiten hacer chat con video.

Vínculo

Véase Hipervínculo.

Vínculo de comunicaciones

Sistema de equipo y programas que conecta a dos usuarios finales.

Virus

Pequeño programa que "infecta" una computadora; puede causar efectos indeseables y hasta daños irreparables.

Voice chat

Chat de voz. Conversación a través de la Internet. Para llevarla a cabo es preciso que los dos usuarios intercomunicados tengan micrófono, altoparlantes y placa de sonido y que usen el mismo programa de voice chat.

VRML - Virtual Reality Modeling Language

Lenguaje de "etiquetas" en el que las páginas de Web están formateadas para permitir gráficos en 3D y la exploración espacial interactiva.

W

WAIS (Wide Area Information Server)

Potente sistema para buscar grandes cantidades de información muy rápidamente en Internet.

WAP (Wireless Application Protocol)

Norma internacional para aplicaciones que utilizan la comunicación inalámbrica, por ejemplo el acceso a Internet desde un teléfono móvil.

Wav

Extensión de un archivo de sonido llamado "wave", creado por Microsoft. Se ha convertido en un estándar de formato de audio para PC. Se puede usar también en Macintosh y otros sistemas operativos.

web page

Página web.

web site

Sitio web.

Web

World Wide Web.

Webcam

Videocámara que registra imágenes a las cuales se puede acceder desde un sitio web.

Webmail

Servicio que ofrecen ciertos sitios web para crear una cuenta gratuita de e-mail. Mediante el webmail el correo electrónico se revisa con el navegador. Se puede acceder a él desde cualquier computadora situada en cualquier lugar.

Webmaster

Persona responsable de la creación, administración, programación y control técnico de un sitio web.

WebTV

Tecnología en la que convergen la televisión y la World Wide Web. Un televisor permite recorrer las páginas web usando un navegador y un control remoto. La señal llega a través de un módem conectado a una línea telefónica.

Windows 2000

Versión del sistema operativo Windows, cuyo lanzamiento ha sido anunciado por Microsoft para el año 1999.

Windows 95

Sistema operativo lanzado por Microsoft en agosto de 1995.

Windows 98

Sistema operativo lanzado por Microsoft en 1998, como sucesor de Windows 95. Una de las más visibles diferencias con el anterior consiste en la integración del sistema operativo con el navegador Internet Explorer. Esta característica dio pie a un juicio por monopolio.

Windows CE

Sistema operativo basado en Windows. Fue diseñado para dispositivos móviles o pequeños. Viene incorporado en varias marcas de handheld.

Windows NT Server

Windows NT diseñado para máquinas que proveen servicios a computadoras conectadas a una LAN.

Windows NT Workstation

Windows NT diseñado especialmente para empresas, se lo considera más seguro y estable que Windows 95 y 98.

Windows NT

Sistema operativo Windows de Microsoft diseñado para usuarios avanzados y empresas. En realidad se trata de dos productos: Windows NT Workstation y Windows NT Server.

WINZIP

Winzip es una utilidad de compresión que permite a los usuarios de Windows 95, 3.1 y NT reducir el tamaño de sus archivos para transferirlos más rápidamente a través de Internet. Esta utilidad también descomprime archivos comprimidos con formatos PKZIP o TAR.

<http://www.winzip.com/winzip/>

Wireless

Inalámbrico.

Workstation

Estación de trabajo. Computadora personal conectada a una LAN. Puede ser usada independientemente de la mainframe, dado que tiene sus propias aplicaciones y su propio disco rígido.

World Wide Web

Red mundial; telaraña mundial. Es la parte multimedia de Internet. Es decir, los recursos creados en HTML y sus derivados. Sistema de información global desarrollado en 1990 por Robert Cailliau y Tim Berners-Lee en el CERN (Consejo Europeo para la Investigación Nuclear). Con la incorporación de recursos gráficos e hipertextos, fue la base para la explosiva popularización de Internet a partir de 1993.

X

XDSL

El término se refiere a las diferentes variaciones de DSL, tales como ADSL, HDSL y RADSL.

XGA eXtended Graphic Array.

Monitor que soporta hasta 65.536 colores

Y

Ninguna entrada

Z

ZIP

Al "comprimir" (es decir, hacer una copia de menor tamaño, pero igual) un archivo mediante "pkzip" o "winzip", el archivo resultante se denomina un archivo "zip". Suele terminar con la extensión ".zip".

zip drive

Periférico para almacenamiento de datos. Cada zip drive puede contener hasta 100 MB (mega bites) o el equivalente a 70 disquetes.

Zippear

Comprimir.

14.- BIBLIOGRAFÍA

- Castano, S., Fugini, M., Martella, G. y Samarati, P. (1994). Database Security. Addison-Wesley.
- Fernández-Medina, E., Piattini, M. y Serrano, M. A. (2001). Seguridad en Bases de Datos. Fundación Dintel, Madrid.
- Fernández-Medina, E. y Piattini, M. (2002). Una Metodología para Diseñar Bases de Datos Seguras Implementadas en Oracle9i Label Security. Cuore, Vivat Academia. Nº 3. Noviembre.
- Fernández-Medina, E., Moya, R. y Piattini, M. (2003). Seguridad en TI. La Construcción para una Sociedad Conectada. AENOR. Madrid.
- Piattini, M. y Fernández-Medina, E. (2001). Specification of Security Constraints in UML. Actas del 35th Annual 2001 IEEE International Carnahan Conference on Security Technology (ICCST 2001), páginas 163-171. Octubre de 2001. Londres (Reino Unido).
- Informe sobre Europa y la sociedad global de la Información (Grupo Bangemann). en: Crecimiento, competitividad, empleo: seguimiento del libro Blanco. Luxemburgo: Comisión Europea (ISBN 92-826-8543-8)
- Gale Directory of Databases 1997. (En especial, el estudio preliminar "The State of Databases Today: 1997" de Marta Williams, de la Universidad de Illinois).
- RUIZ GONZÁLEZ, B (ed.) Catálogo de servicios españoles de información electrónica (7ª edición). Madrid, Fuinca, 1997.
- Multimedia 1996. Tendencias . Madrid, Fuinca, 1997.
- PLANBA: Plan de Acción Nacional para la I+D en Comunicaciones Integradas de Banda Ancha. Madrid: Ministerio de Obras Públicas, Transportes y Medio Ambiente, 1995 (NIPO 161-95-122-3)
- SZPRUWICZ, Bohan O. Multimedia Networking. New York, McGraw-Hill, 1995.
- Apuntes Bases de Dades I y Bases de Dades II de la UOC
- Apuntes Enginyeria Programari I y Enginyeria de Programari II de la UOC
- Apuntes Seguretat en Xarxes de Computadors de la UOC
- Apuntes Administració de Xarxes y Sistemes Operatius de la UOC
- Buscador de páginas de Internet <http://www.google.es>
- Página principal de Oracle, una de la mejores Bases de Datos <http://www.oracle.com/es>
- Página principal de SQL, la Base de Datos (Microsoft), <http://www.microsoft.com/es>
- Información del proyecto Informedia de la Carnegie Mellon University <http://www.informedia.cs.cmu.edu/>