

# HERRAMIENTAS INFORMÁTICAS PARA LA DOCUMENTACIÓN

## Tema 3. Entornos de almacenamiento de información

En temas anteriores hemos visto que las personas y las organizaciones tienen unas determinadas necesidades de información que deben satisfacer para poder cumplir sus objetivos. Los almacenes de información electrónica constituyen buenas fuentes de información, por lo que cada vez son más utilizados para satisfacer estas necesidades. En este tema se estudia el funcionamiento y las ventajas de los almacenes de información electrónica más habitualmente utilizados. Estos son los discos ópticos, las bases de datos, las bibliotecas digitales e Internet.

Antes de empezar hemos de dejar claro que la anterior no constituye una clasificación homogénea en la que los almacenes de información incluidos son del mismo tipo. Así, los discos ópticos son un tipo de soporte físico para almacenar información, las bases de datos son entornos de almacenamiento y acceso a la información que pueden guardarse por ejemplo en discos ópticos o accederse a través de Internet, las bibliotecas digitales son almacenes de documentos accesibles a través de Internet y, finalmente, Internet es un mecanismo de acceso, algunos de cuyos servicios (web, FTP), constituyen excelentes sistemas de almacenamiento de información.

### 1. Bases de datos

Podemos definir una **base de datos** como un conjunto de información estructurada que ha sido almacenada en formato digital y que se refiere a un mismo tema, por ejemplo, a una organización, una materia o un problema determinado.

Además, las bases de datos incorporan un programa que permite la gestión automatizada de la información almacenada. Este tipo de programas se denominan **Sistemas de Gestión de Bases de Datos (SGBD)** y permiten la creación, utilización y consulta de las mismas.

Las bases de datos pueden contener información en forma de texto, información numérica, imágenes, etc. y contener datos sobre personas, organizaciones o empresas, productos, libros, etc.

Cuando decimos que la información contenida en una base de datos se encuentra estructurada, queremos decir que se almacena en **registros** de datos. El registro es la unidad básica de información de la base de datos y almacena la información relacionada con cada uno de los elementos de que consta. Por ejemplo, en una base de datos bibliográfica en la que se almacena la información relativa a los artículos publicados en diversos congresos y conferencias, cada registro almacenará la información sobre un artículo.

Los registros se dividen en **campos**, cada uno de los cuales almacena uno de los datos relativos al mismo. Por ejemplo, en la base de datos bibliográfica antes comentada, cada registro relacionado con un artículo podría contener los siguientes campos: título, autor, resumen, lugar donde se ha publicado, etc. En la figura 1 se muestran los distintos componentes en una base de datos con registros que describen artículos científicos y se muestra un ejemplo del contenido de uno de ellos. Cada registro de esa base de datos tendrá los siete campos que se indican, con los valores correspondientes a un artículo científico diferente.

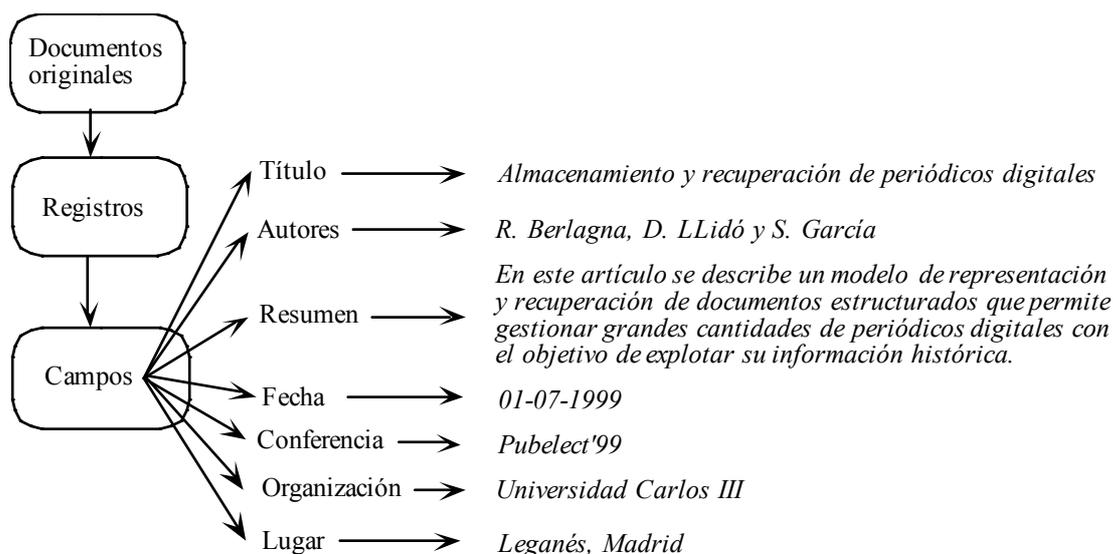


Figura 1. Componentes fundamentales de una base de datos de documentos.

### 1.1. Tipos de bases de datos

Desde el punto de vista del tipo de información almacenada, podemos diferenciar dos tipos básicos de bases de datos: referenciales y fuente.

Las **bases de datos referenciales** contienen información secundaria sobre los elementos que almacenan, es decir que hacen referencia a los elementos y no los contienen. Por ejemplo, una base de información bibliográfica es de este tipo, dado que almacena información sobre libros o artículos, pero no almacena el contenido de los mismos.

Dentro de las bases de datos referenciales podemos distinguir dos subtipos:

Las **bases de datos bibliográficas** contienen referencias documentales que remiten a libros, artículos en revistas, informes u otros tipos de documentos. Los campos típicos de estos registros son título, autor, fecha, lugar de publicación, y posiblemente un resumen y unas palabras claves que describen su contenido.

Los **directorios** contienen datos domiciliarios, biográficos o curriculares que remiten a organizaciones, instituciones, empresas o personas, o descripciones de objetos tales como obras de arte.

Las **bases de datos fuente** almacenan el contenido completo de los documentos a los que se refieren o la información final a la que se va a acceder. Tienen la ventaja con respecto a las referenciales de que no se restringen a indicar cómo localizar el documento, sino que nos proporcionan su contenido completo. A su vez se distinguen tres subtipos de bases de datos fuente:

Las **bases de datos textuales** contienen los textos completos de documentos, con la incorporación en muchos casos de recursos gráficos para la presentación de la información.

Las **bases de datos numéricas** contienen datos estadísticos o factuales sobre alguna materia. Por ejemplo, la figura 2 muestra los resultados, tanto numéricos como en forma de gráfica, obtenidos al consultar la evolución del IPC en la base de datos estadística Tempus que puede encontrarse en la página del Instituto Nacional de Estadística ([www.ine.es](http://www.ine.es)).

Las **bases de datos multimedia** contienen documentos en los que podemos encontrar no sólo el texto, sino también sonido, imágenes fijas o en movimiento.

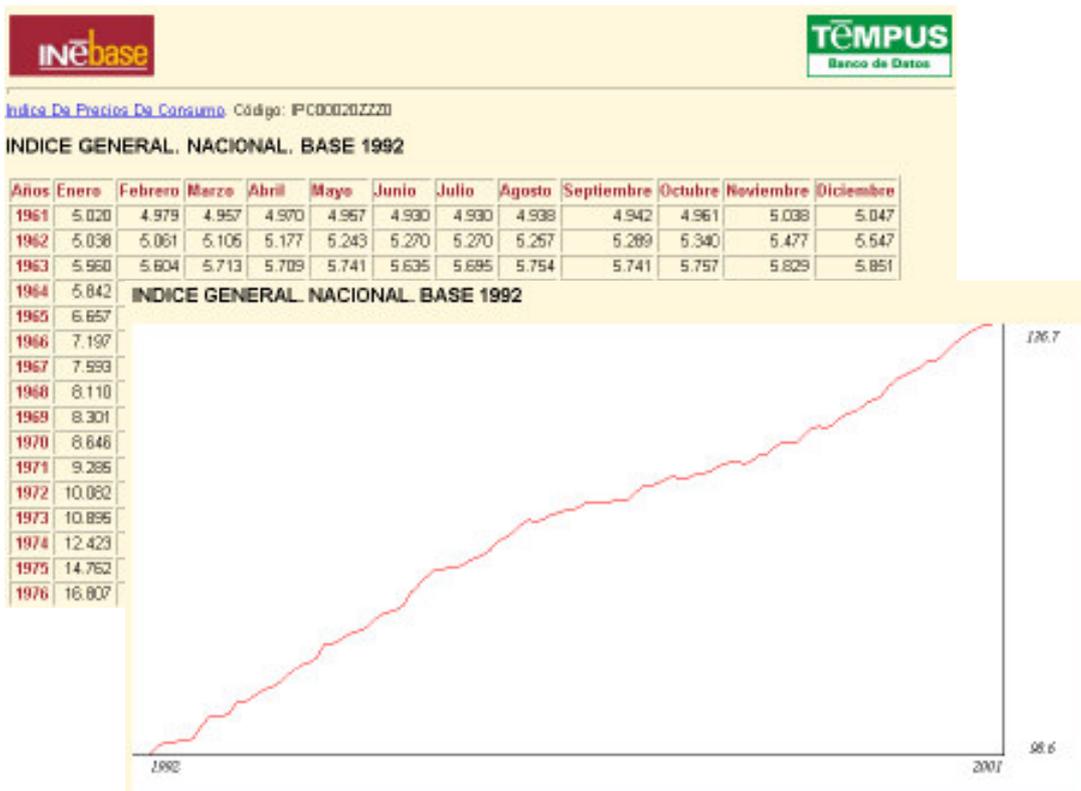


Figura 2. Base de datos estadística Tempus del Instituto Nacional de Estadística.

## 1.2. Formas de acceso a las bases de datos

Podemos distinguir dos formas fundamentales de acceso a las bases de datos: en soporte digital o a distancia.

En el caso del **acceso a la base de datos en soporte digital**, el usuario dispone de la base de datos en algún soporte magnético u óptico tal como un CD-ROM o un DVD, y puede utilizar un ordenador para acceder a la misma. En este caso la base de datos ha sido adquirida por el usuario, tal y como ocurre cuando compramos una enciclopedia en CD. En algunos casos, los datos almacenados se actualizan con bastante frecuencia y el usuario paga una suscripción periódica que le da derecho a recibir las distintas actualizaciones que van produciéndose.



Figura 3. Acceso a las bases de datos del BOE.

En el caso del **acceso a distancia**, también denominado en línea, el usuario consulta la información almacenada en la base de datos a través de una red conectándose a algún ordenador en el que se encuentra la misma. Por ejemplo, la biblioteca de la UJI dispone de una serie de ordenadores desde los cuales se puede acceder a la información contenida en la base de datos jurídica Aranzadi. Por otro lado, existen diversas bases de datos que pueden consultarse a través de Internet accediendo a alguna página web determinada. En algunos casos, el acceso a los datos es gratuito, pero en la mayoría de ellos es necesario pagar una cuota de suscripción. El pago de esta cuota da derecho a acceder a la información durante un tiempo dado mediante el uso de una cuenta con un nombre de usuario y una contraseña asociados (ver figura 3). Un ejemplo de base de datos de acceso gratuito es INDIBOE, que nos permite acceder al contenido del Boletín Oficial del Estado (BOE) de los últimos años. La figura 4 nos muestra el formulario de búsqueda asociado a esta base de datos que podemos encontrar en *www.boe.es*.

## Búsqueda en Sumarios (INDIBOE)

Búsqueda en la Sección:

Escriba en cada campo la palabra o frase de búsqueda que desee. Puede utilizar el comodín \* al final de una palabra.

Ej: tribu\* de justicia

Título   
 Y Departamento   
 Y Número de BOE   
 Y Ref. (aaaa-nnnn)   
 Y Fecha de Publicación De:  a:   
 (dd/mm/aaaa)

Núm. resultados por pág.

Ordenación de resultados por:

Figura 4. Formulario de búsqueda de la base de datos INDIBOE.

Cuando se accede a una base de datos el usuario dispone de los medios necesarios para recuperar un conjunto de registros de interés de entre la gran cantidad que almacenan. Dependiendo del tipo de base de datos, a veces también es posible recuperar los documentos completos junto con sus registros.

Estos medios de recuperación consisten principalmente en un **formulario de consulta** (ver figura 3), donde el usuario especifica las condiciones que tienen que cumplirse sobre los campos de los registros que desea seleccionar. Por ejemplo, sobre bases de datos referenciales se podría indicar el autor de los documentos que se buscan, o su fecha de publicación. El SGBD ejecuta la consulta y devuelve al usuario todos los registros y documentos que cumplen las condiciones especificadas. Alternativamente, hay bases de datos que proporcionan **lenguajes de consulta** en los que el usuario puede especificar condiciones de recuperación mucho más complicadas que con un formulario. Sin embargo, para aprender a manejar estos lenguajes hay que hacer un entrenamiento previo, ya que pueden llegar a ser bastante complicados. Un ejemplo sencillo de consulta a una base de datos utilizando el lenguaje de recuperación SQL es el siguiente:

```
SELECT codcli, nombre, telefono
FROM clientes
WHERE codcli BETWEEN 100 AND 200;
```

En este caso estaríamos recuperando los campos codcli (código de cliente), nombre y telefono de una base de datos de clientes, para aquellos clientes cuyo código se encuentre entre 100 y 200.

Algunos de los SGBDs más conocidos son *Oracle* e *Informix* para organizaciones y entornos empresariales y *Microsoft Access* para usuarios particulares o pequeñas empresas.

## 2. Discos ópticos

Los discos ópticos que fueron inventados a principios de los años 80 y que comenzaron utilizándose básicamente para el almacenamiento de música, han pasado a ser la tecnología actual más avanzada para almacenar, gestionar y distribuir grandes volúmenes de información multimedia (texto, gráficos, imágenes y sonido). Sus principales características se pueden resumir como sigue:

- Puesto que son discos extraíbles, no hay límite en la cantidad de información que se puede almacenar utilizando tantos discos como sea necesario.
- La densidad de grabación es tan alta que en un solo disco óptico se puede almacenar una gran enciclopedia con cientos de imágenes.
- Se pueden producir a muy bajo coste, ya que están hechos de materiales muy baratos.
- A diferencia de los discos magnéticos, los discos ópticos son inmunes a campos magnéticos y al polvo, presentando una gran resistencia a golpes y rozaduras.

Los principales inconvenientes de los discos ópticos como sistema de almacenamiento de información son:

- Su rápida obsolescencia técnica. Probablemente, mucho antes que cualquier dispositivo óptico resulte afectado por el paso del tiempo, se volverá ilegible porque ya no existirán en el mercado máquinas que puedan leerlo. Recordemos por ejemplo el caso de los disquetes de 5,25 pulgadas, que se utilizaron de forma masiva en los PC durante la década de los años 80 y que en la actualidad son muy difíciles de leer por la dificultad de acceder a un ordenador con una disquetera de ese tamaño.
- La gran cantidad de formatos de grabación de información hace que sea complicado disponer de lector y de los programas adecuados para leer todos los posibles tipos de información almacenada.
- La necesidad de disponer de un dispositivo especial para su lectura. A diferencia de los libros o las revistas en papel, para poder acceder a la información almacenada en un disco óptico es necesario utilizar un ordenador con un lector y unos programas adecuados.

A pesar de los inconvenientes citados, la producción de discos ópticos no ha parado de aumentar de manera espectacular desde su creación. La figura 5 presenta el número de títulos publicados en este formato según el directorio *New media titles*. Como sucede con todos los directorios, las cifras que se ofrecen quedan muy superadas por la oferta real que puede encontrarse en el mercado.

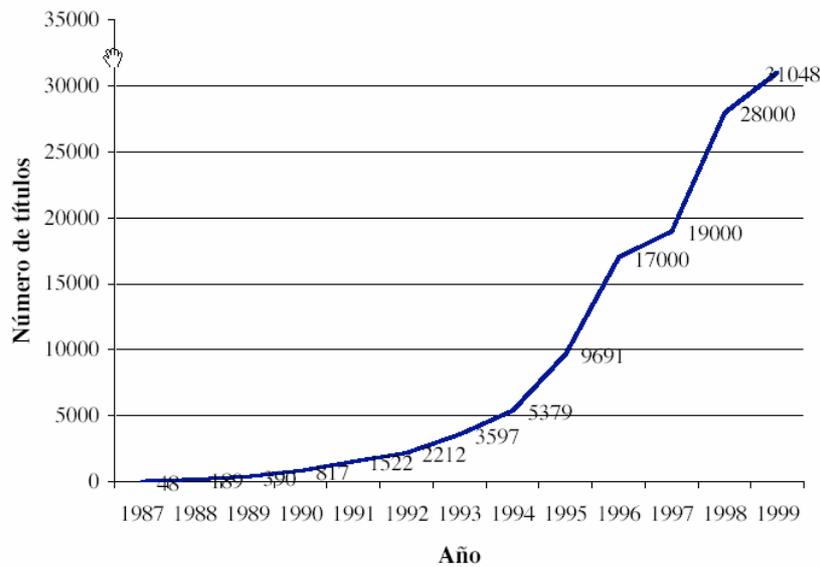


Figura 5. Evolución en el número de títulos de CD publicados en el mundo.

## 2.1 Almacenamiento de Información

Para almacenar la información dentro de un disco óptico, previamente debe ser codificada y digitalizada por medio de un programa de ordenador adecuado. **Codificar** la información consiste en asignar a cada elemento de información un código numérico distinto. El sistema de codificación a seguir dependerá del tipo de información de que se trate, así como de los programas que se utilicen para su manipulación. Por ejemplo, utilizando el sistema de codificación para información textual ASCII, la letra 'A' se codifica con el número 65, que en binario corresponde a 01000001. Una vez ya se encuentra toda la información codificada, se tiene que traducir a código binario para poder ser grabado, es decir el código inicial se traduce a otro de ceros y unos. Todo este proceso se denomina **digitalización de la información**.

Hoy en día, cualquier tipo de información se puede digitalizar, y por lo tanto puede ser manipulada y almacenada dentro de la memoria del ordenador, o dentro de un disco óptico o magnético. Con estos sistemas de codificación, y gracias a la gran densidad de grabación que soportan los discos ópticos actuales, su capacidad es suficiente para poder almacenar cientos de miles de páginas o varias horas de música.

Los discos ópticos consisten en una superficie circular de plástico transparente de 12 cm. de diámetro, sobre la que se disponen diversas capas de materiales diferentes en función de si el disco es de producción industrial, si es un disco grabable (CD-R) o un disco regrabable (CD-RW). El principio subyacente en todos los casos es el mismo, el disco contiene una pista espiral sobre la que se crean zonas que reflejan la luz hacia un sensor de lectura, mientras otras no la reflejan o lo hacen en un ángulo distinto. Cuando la luz reflejada vuelve al sensor de lectura se produce un impulso eléctrico y se

considera que la información almacenada es un 1, y en caso contrario se ha grabado un 0.

Por ejemplo, los discos grabables (CD-R) constan de tres capas (ver figura 6), una primera capa de material reflectante (normalmente aluminio), una segunda capa de material semitransparente y una capa de mayor grosor de material plástico transparente. Esta última capa es la que solemos ver como irisada cuando movemos el disco. A la hora de grabar la información, se hace pasar el láser a través de la capa de plástico para incidir en la capa semitransparente. El láser quema las zonas correspondientes a los ceros haciéndolas totalmente opacas, mientras el resto de zonas, correspondientes a los unos, dejan visible la capa reflectante inferior.

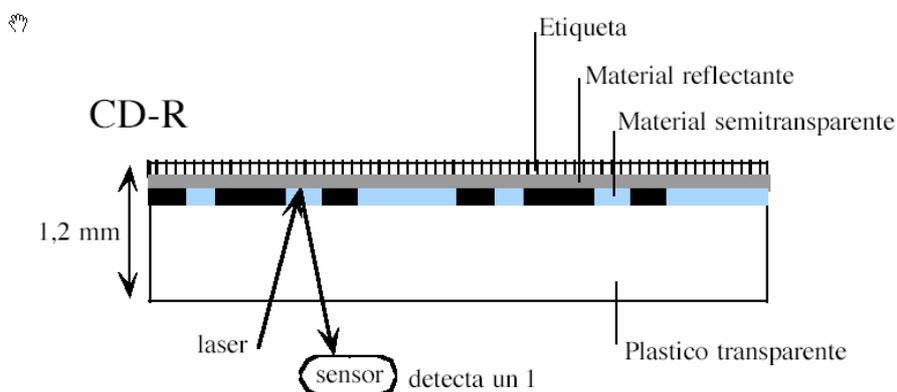


Figura 6. Estructura de un CD-R grabado.

### 2.3 Tipos de disco óptico

Es difícil ofrecer un panorama exhaustivo de los distintos tipos de discos ópticos existentes debido a la amplia variedad y a la continua aparición de nuevos formatos. No obstante vamos a utilizar una serie de criterios básicos para establecer una posible clasificación. En primer lugar distinguiremos los discos en función de que el usuario pueda o no grabar información en ellos. En este sentido distinguiremos tres tipos de discos fundamentales:

- Discos que sólo permiten la lectura, ROM (*Read Only Memory*).
- Discos que pueden ser grabados una sola vez por el usuario, WORM, (*Write Once Read Many*), también denominados R (*Recordable*).
- Discos regrabables, RW (*Rewritable*), que permiten reescribir la información una gran cantidad de veces borrando la anterior.

A su vez, dentro de cada una de las tres clases anteriores, existen múltiples variedades de formato que se diferencian por las dimensiones del disco, por su capacidad de almacenamiento o por el sistema de grabación utilizado.

Por ejemplo, en función del sistema de grabación utilizado podemos diferenciar dos grandes familias: los discos compactos (CD), que cumplen el estándar ISO 9660, y los discos de vídeo digital (DVD) que se basan en otro estándar.

### 2.3.1. Discos compactos (CD)

Los discos compactos fueron desarrollados a partir de 1976 como consecuencia de la colaboración de dos empresas: Philips y Sony. Su primer uso fue la grabación de música, lo que dio lugar a los conocidos discos compactos de audio (CD-A) que sustituyeron rápidamente a los clásicos LPs de vinilo. La siguiente tabla refleja los principales tipos de discos compactos existentes en la actualidad.

Nombre	Año	Características
CD-A	1982	Información sonora equivalente a 74 minutos de sonido de alta calidad.
CD-ROM	1985	Información textual y gráfica (imagen fija).
CD XA	1988	<i>eXtended Architecture</i> . Información textual, sonido, imagen fija y fragmentos de imagen animada.
CD-I	1992	Interactivo. Información multimedia (estándar desarrollado por Philips) que puede leerse desde un ordenador o un monitor de televisión.
Foto CD	1992	Información gráfica (fotografías) de alta calidad (estándar desarrollado por Kodak). Lectura desde un ordenador o un monitor de televisión.
CD-R	1992	<i>Recordable</i> . Puede ser grabado por los usuarios y leído con cualquier lector convencional.
CD-RW	1998	Regrabable.

A continuación se comenzaron a utilizar los discos compactos para la grabación de otros tipos de información, tales como textos, imágenes estáticas o películas. Para ello se estableció el estándar ISO 9660, que define, entre otras cosas, el diámetro de los discos (12 cm.) y el formato en el que se almacena la información (secuencialmente mediante marcas de tamaños definidos a lo largo de una pista espiral de unos 5 kms. de largo). A partir de esta norma surgieron diferentes tipos de disco compacto en función del tipo de información almacenada y de cómo se almacena esta exactamente.

En todo caso, la mayoría de los discos compactos tienen una misma capacidad de almacenamiento, aunque también en este aspecto existen diversos tipos. Esta capacidad corresponde a 650 Mbytes de información, lo que es equivalente a:

- 400 disquetes de 1,4Mbytes cada uno.
- 250.000 páginas de texto con unos 2.000 caracteres por páginas.
- 70 millones de palabras.
- 5.000 fotografías.
- 74 minutos de música de alta calidad.

### **2.3.2. Discos de vídeo digital (DVD)**

El disco de vídeo digital (DVD), también llamados disco versátil digital, empezó a desarrollarse para superar las deficiencias y limitaciones de los discos compactos a la hora de almacenar información multimedia, especialmente imágenes en movimiento. Se espera que a corto plazo sustituyan a los discos compactos y también a las cintas de vídeo.

Este tipo de discos fueron desarrollados entre 1994 y 1995 por dos grandes consorcios empresariales que acabaron llegando a un acuerdo sobre el estándar a utilizar. Una de las principales características de este estándar es su compatibilidad con los discos compactos actuales, es decir, que los lectores de DVD también son capaces de leer CDs.

Las principales diferencias con los discos compactos se encuentran en el formato con el que se almacena la información y en el tamaño de las marcas utilizadas para la grabación. Al utilizarse marcas de menor tamaño que en los discos compactos, y al poderse grabar en ambas caras, la capacidad de este tipo de discos es de 4,7 Gbytes (siete veces más que los actuales discos compactos). Esto permite la grabación de unos 133 minutos de vídeo de alta calidad comprimido, además de su doblaje a cinco idiomas y los subtítulos en otros 30. Se han previsto varios tipos de DVDs en cuanto a su capacidad de grabación, pudiendo alcanzar hasta 17 Gbytes aquellos que graben a doble cara y con una doble capa por cara.

Por otro lado, y en función del tipo de información almacenada y del modo en que esta se graba, podemos diferenciar tres tipos de DVD:

Nombre	Año	Características
DVD-ROM	1997	Información multimedia. Entre 4,7 y 17 Gbytes. Leído desde un dispositivo específico instalado en un ordenador.
DVD-vídeo	1997	Películas visualizadas en un televisor y que acabarán sustituyendo a las cintas de video.
DVD-RW	2000	Permiten grabar 2,6 Gbytes por cara y pueden ser visualizados tanto con lectores de DVD-ROM como de DVD-vídeo.

A finales del año 2000 salieron al mercado los primeros DVD regrabables, aunque existen todavía problemas de compatibilidad y precio a solventar. Actualmente se están comercializando tres sistemas de grabación de DVD diferentes:

- DVD-RAM: Panasonic, Toshiba y otros. Al igual que los disquetes tradicionales suelen venir dentro de un cartucho de plástico rígido (*cartridge*), lo que hace que no puedan leerse con los lectores de DVD-CD normales y necesiten un dispositivo especial (ver figura 7). De todas formas, existe una variante en los que el disco se encuentra al descubierto y pueden ser leídos con algunos dispositivos DVD-CD que admitan este nuevo formato.
- DVD-RW: Pioneer y Sharp.
- DVD+RW: Sony, Philips y otros.



Figura 7. Disco DVD-RAM.

### 3. Bibliotecas digitales

Las bibliotecas tradicionales, vistas como grandes edificios en los que es posible acceder a libros y revistas, han experimentado un enorme cambio con la aparición de los ordenadores y más recientemente con la extensión de Internet.

Originalmente en las bibliotecas tan sólo era posible acceder a documentación en papel y para ello era necesario desplazarse físicamente a la misma. Cada vez más, la información se almacena en formato digital y las bibliotecas deben dar cabida a este tipo de documentos, tanto a la hora de almacenarlos como de proporcionar los medios para acceder a la información que contienen.

El primer paso hacia la informatización de las bibliotecas fue equiparse con ordenadores que permitirán acceder a sus catálogos, e incluso a la información original almacenada en formatos digitales, fundamentalmente constituida por bases de datos.

A partir de la aparición de Internet, muchas bibliotecas comenzaron a ofrecer sus servicios a través de la red, de modo que era posible consultar los catálogos a distancia (en línea). Este tipo de catálogos se denominan **OPAC** (On-line Public Access Catalog), es decir, Catálogos de acceso público en línea. Originalmente este acceso se realizaba a través de una conexión remota, telnet (ver figura 8), pero con la aparición de la web, la mayoría de los catálogos son accesibles mediante formularios situados en páginas web (ver figura 9).



Figura 8. Catálogo telnet de la biblioteca de la Universidad de Málaga.

Figura 9. Catálogo web (OPAC) de la biblioteca de la UJI.

El tercer paso en la evolución de las bibliotecas tradicionales es la aparición de las denominadas bibliotecas digitales en las que la información se proporciona en formato digital y que pueden incluso carecer de una ubicación física y ser sólo accesibles a través de Internet.

Una **biblioteca digital** es aquella que proporciona los contenidos de sus documentos en formato digital a través de *Internet* u otro medio. Aunque se trata de un proyecto muy ambicioso, por todo el mundo hay bibliotecas que ya han comenzado la tarea de crear copias digitales de los libros, imágenes y grabaciones que almacenan. En nuestro país, la Universidad de Alicante lleva a cabo un proyecto pionero de crear una biblioteca digital con las principales obras de la literatura clásica en castellano. Su nombre es Biblioteca Virtual Miguel de Cervantes y se puede acceder desde <http://cervantesvirtual.com>. Las ventajas de digitalizar toda esta información pueden resumirse como sigue:

- Los objetos raros y frágiles se pueden preservar de manera segura sin necesidad de prohibir el acceso a quienes deseen estudiarlos. Es decir, los investigadores pueden utilizar las copias digitales idénticas a las originales, las cuales se encuentran en un lugar seguro.
- La información digital es más cómoda de utilizar, ya que se puede leer a través de *Internet*, sin necesidad de desplazarse hasta la biblioteca. Además, varias personas pueden acceder al mismo documento simultáneamente. En otras palabras, los documentos están mucho más accesibles y disponibles.
- Las copias digitales ocupan mucho menos espacio que las originales, lo cual reduce sus costes de mantenimiento.

Podemos diferenciar dos formas fundamentales mediante las que las bibliotecas digitales pueden obtener y guardar los documentos:

- En formato de imágenes digitales obtenidas mediante exploración óptica de los originales.
- En formato de texto electrónico. A su vez el texto electrónico puede obtenerse de tres formas básicas:
  - \* Mediante la introducción del documento por teclado.
  - \* Mediante la conversión automática de las imágenes digitales.
  - \* A partir de los documentos originales en formato digital proporcionados por los autores o editores.

El procedimiento más barato y rápido de obtener copias digitales de los documentos de las bibliotecas es por medio de su **exploración óptica** (procesamiento con un escáner). Esta técnica de procesamiento es también una opción muy adecuada con miras a la conservación, ya que capta el aspecto original del libro, y las notas manuscritas al margen. Además, los elementos no textuales, aunque se hallen repletos de detalles minuciosos, pueden convertirse en imágenes de alta resolución. En la figura 10 vemos una imagen de la portada de una antigua edición de “El Quijote”, en concreto, la primera edición de Juan de la Cuesta (1605).

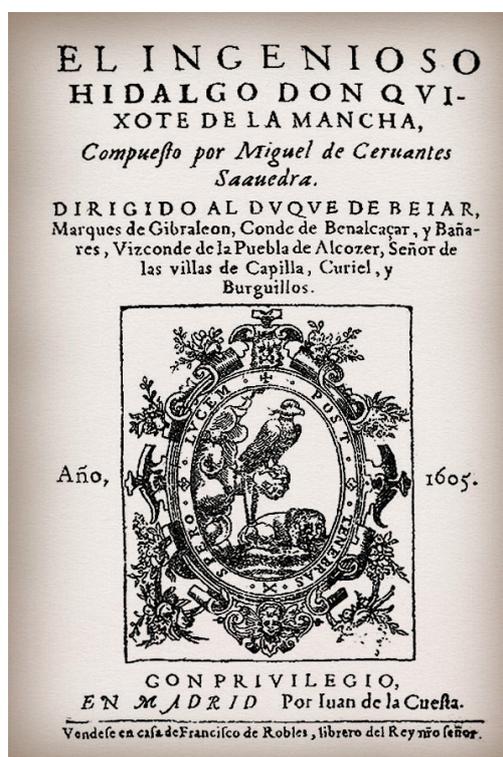


Figura 10. Imagen digital de la portada de El Quijote de 1605.

Un ejemplo de almacén de información con imágenes digitales de los documentos es el BOE. En la página de Internet del Boletín Oficial del Estado ([www.boe.es](http://www.boe.es)) es posible

acceder a la versión completa de los últimos números del mismo. Los resúmenes de las distintas resoluciones se encuentran en formato de texto, pero las páginas completas se han almacenado en un formato gráfico tras ser procesadas mediante un escáner, con lo que para leerlas será necesario utilizar un programa que permita visualizar este formato en particular (TIFF).

El principal inconveniente del barrido óptico por medio de un escáner es que sólo produce imágenes digitales, las cuales, como las páginas impresas de los libros, son adecuadas exclusivamente para ser vistas o leídas. Sin embargo, no existen todavía programas de ordenador que permitan localizar una cita concreta o un pasaje enterrado en la imagen digital de un libro. Además, en función de la precisión con que queramos guardar la información (fundamentalmente las fotografías o gráficos de alta resolución), el tamaño de los documentos puede ser muy grande.

El **texto electrónico** presenta ventajas con respecto a la imagen digital, ya que ocupa menos espacio, puede ser indizado para su recuperación, o manipulado con procesadores de texto. Para convertir una imagen digital en texto electrónico se utilizan los programas de reconocimiento óptico de caracteres (ROC), los cuales ya son capaces de reconocer prácticamente cualquier texto escrito a partir de su imagen digital.

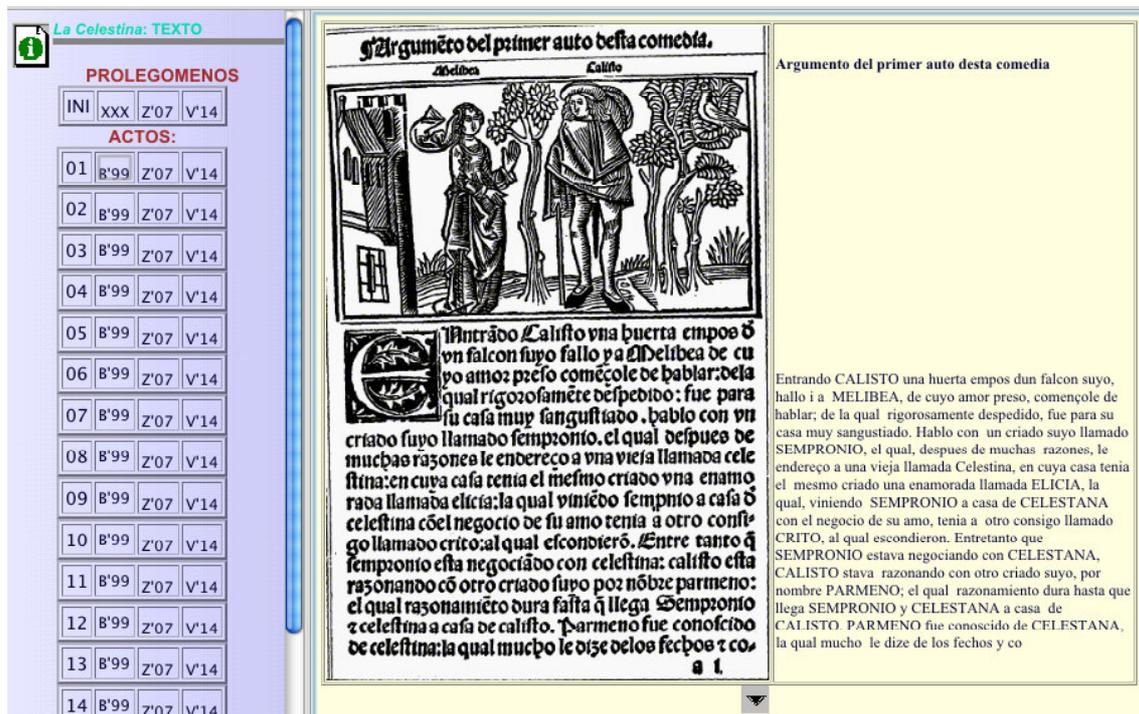


Figura 11. Texto electrónico con imagen integrada de una página de "La Celestina".

Lógicamente, todo archivo que pretenda reemplazar a las publicaciones en papel ha de incluir los dibujos y fotografías que figuran en las obras originales. Los programas ROC, también son capaces de identificar automáticamente las imágenes, lo que permite preservarlas e insertarlas en los lugares adecuados del texto electrónico. En la figura 11

vemos el texto original de “La Celestina” en formato electrónico junto con una imagen digital del documento original.

Una alternativa a utilizar un escáner podría ser la de teclear el texto original del documento. Dado el tiempo y el trabajo que conlleva, ésta técnica resulta demasiado cara, por lo que sólo se utiliza en casos muy especiales. Sin embargo, actualmente la mayoría de los libros que se publican tienen su origen en un fichero de ordenador, por lo que no es necesaria su conversión al medio electrónico. Las propias editoriales podrían, bajo licencia, proporcionar la versión electrónica a las bibliotecas que lo soliciten.

A pesar del gran esfuerzo que se va a necesitar, nos encontramos ante la situación que de aquí a unos años, las bibliotecas digitales estarán accesibles para todos, proporcionando versiones electrónicas de prácticamente cualquier documento publicado. Asimismo, las organizaciones privadas o de ámbito cerrado también contarán con bibliotecas digitales de uso restringido a sus miembros. De esta manera tendremos la información más fácilmente accesible que nunca.

#### **4. Internet**

En una definición estricta Internet no es un sistema de almacenamiento de información, sino un mecanismo para poder acceder a ella e intercambiarla. Sin embargo, en último término sí que podemos decir que el conjunto de la información accesible por Internet constituye un inmenso almacén de información. El uso de los distintos servicios de Internet, tales como páginas web o repositorios FTP que nos permite la publicación y puesta a disposición de información, constituye la fuente de información más importante de la actualidad. Y no sólo eso, sino que a través de Internet podemos acceder de modo remoto a los catálogos de una gran cantidad de bibliotecas y centros de documentación, a un gran número de bases de datos de distintos tipos, a información actualizada de los mercados financieros, a las versiones electrónicas de los artículos publicados en distintas revistas, a miles de textos electrónicos incluyendo libros, legislación o informes técnicos, etc.

Además, el acceso a la información digital a través de Internet tiene múltiples ventajas que veremos con mayor detalle en el tema 4, pero entre las que ya podemos destacar, la existencia de herramientas que facilitan la búsqueda y el acceso, la rapidez con que podemos acceder a la misma o la velocidad con que pueden actualizarse sus contenidos.