

# Gestión electrónica de los documentos

**Resumen:** *La gestión electrónica de los documentos implica una serie de procesos en los que están implicados dispositivos de digitalización como escáneres y cámaras digitales, programas para el tratamiento de ficheros gráficos, aplicaciones para la elaboración y adaptación de documentos electrónicos para su integración en los sistemas de tratamiento y recuperación de información documental como las bases de datos documentales que son en definitiva las que posibilitan la utilización, aprovechamiento y difusión de esos documentos.*

## **Introducción: Gestión documental, porqué, pasos**

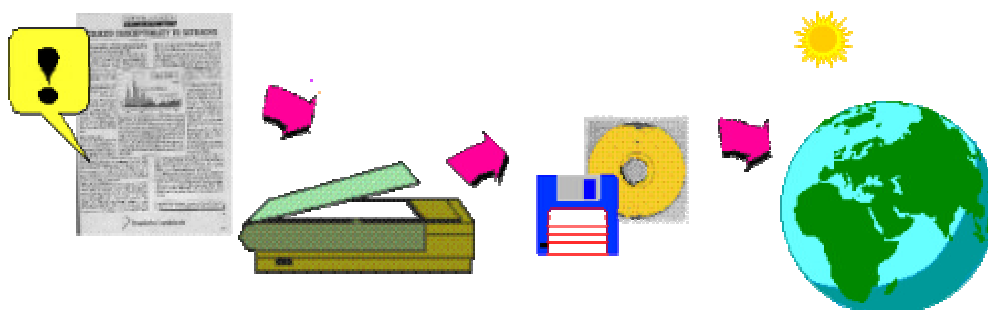
---

La necesidad y el valor de la información ha adquirido una enorme importancia en cualquier ámbito de la vida. El estar informado es indispensable para disponer de criterios que nos permitan tomar decisiones o emitir juicios de valor. Incluso el hecho de tenerla accesible no es suficiente. Hemos de ser capaces de acceder y conseguir la información de forma rápida, ágil y eficaz. La aplicación de la tecnología a la gestión de la información ha dado lugar a disciplinas como la gestión documental, que podemos definir como la gestión de los datos tomando como unidad de información el documento y los procesos necesarios para poder incluir estos documentos en sistemas que permitan su tratamiento informático.

Aunque hoy en día la mayor parte de la producción de información en forma de documentos se realiza por medio de herramientas informáticas (procesadores de texto,...), puede ocurrir, y de hecho será el caso más habitual, que la demanda de información no se restrinja a documentos actuales, ya automatizados, sino que sea interesante y a veces necesario, ya que así lo exigen los usuarios combinarla y/o contrastarla con información más antigua. Incluso puede ocurrir que nos interese disponer sólo de la información antigua (archivos, manuscritos, ...) y nos tengamos que enfrentar a problemas adicionales como el estado o el formato de los documentos.

Para conseguir una gestión eficaz y ágil una vez los documentos gestionados, éstos deben tener una estructura homogénea a la que se le puedan aplicar mecanismos automáticos de tratamiento, gestión y explotación. En los procesos de generación, producción y explotación de toda esta información se utilizan herramientas informáticas que permitan automatizarlos en lo posible.

La generación o creación de un sistema de gestión documental la podemos dividir en cuatro pasos fundamentales:



1. El primero es la recopilación de la información. La gestión documental no implica que los documentos sean realmente textos aunque éste será el caso más habitual. Podemos gestionar, en principio, cualquier tipo de documento; ya sean textos, fotografías, pinturas, mapas, o documentos de tipo mixto.
2. Los documentos en principio podrán estar en cualquier formato. Aunque, como hemos mencionado, si son documentos actuales ya estarán en formato electrónico. Cuando se trata de archivos antiguos, colecciones de pintura, fotografías... es necesario convertirlos a formato electrónico. Este proceso es lo que denominamos digitalización y en general podemos decir que lo que conseguimos es una copia electrónica del documento en cuestión, además de una primera homogenización de los documentos ya que los tenemos todos en formato electrónico.
3. El siguiente paso consiste en la preparación de esta información de forma que nos permita integrarla en un sistema de gestión documental. En el caso de los documentos de texto el proceso de digitalización, definida ésta de forma genérica como la obtención de una copia electrónica del documento, admite un paso más, que es su conversión a formato de texto electrónico. Es lo que se denomina OCR o reconocimiento óptico de caracteres (Optical Character Recognition). Además, la integración de los documentos en sistemas de gestión documental exige normalmente procesos como la extracción de información estructurada, en forma de valores por campos o la homogenización de los formatos de los documentos o la indexación. Todos estos procesos son procesos automatizables en diferentes grados. La indexación por ejemplo es totalmente automatizable mientras que la extracción de datos estructurados y OCR puede exigir cierto trabajo o control manual.

4. La gestión, consulta y explotación de estos sistemas es la parte de usuario final. Normalmente son sistemas basados en Bases de Datos documentales, en algunos casos utilizados exclusivamente para la gestión, en otros para la difusión, y que a veces se combinan con otros productos o desarrollos específicos sobre los sistemas de gestión para facilitar y simplificar su uso. Estos sistemas son capaces de integrar documentos de diferentes tipos.

## **Digitalización**

---

Podemos definir la digitalización como el proceso por medio del cual transformamos el documento en un formato gráfico o escrito ya sea en soporte papel, material fotográfico, etc. en ficheros en formato electrónico (digital) susceptible de ser almacenado, manejado y transformado por medio de ordenadores y programas informáticos.

Lo que en realidad se obtiene del proceso de digitalización, propiamente dicha, es una copia electrónica del documento tratado, que además es una copia gráfica.

En el caso de los documentos de textos el proceso puede tener un paso adicional: el reconocimiento óptico de caracteres (OCR) por medio del cual transformamos la versión gráfica digital en un documento de texto electrónico, con las ventajas que esto conlleva de reducción de tamaño de los ficheros y la posibilidad de aplicarles mecanismos de tratamiento, búsqueda y selección de la información de tipo textual.

La calidad de la versión gráfica del documento depende de la calidad del propio documento. Gracias a la tecnología, la calidad de la copia digital puede ser mejorada. Esta mejora dependerá de la utilidad que se le vaya a dar al documento digitalizado. Normalmente se utilizan programas de retoque fotográfico o tratamiento de imagen para realzar, limpiar de impurezas, clarificar los documentos gráficos obtenidos de la digitalización. Puede ser interesante, por ejemplo, como un proceso anterior al OCR, el eliminar del documento posibles sellos, formateados de preimpresión, y conseguir así aislar la parte de texto del documento que es lo que de verdad nos interesa en un proceso de este tipo.

Aunque el soporte original más usual en los procesos de digitalización es el papel, y normalmente con información escrita, este proceso es susceptible de ser aplicado a cualquier tipo de soporte e información.

Indudablemente, la mecanización del proceso físico de la digitalización, la automatización de los pasos que se siguen con cada documento, será más

sencilla cuanto más homogénea sea la entrada original. Del mismo modo, la automatización será mas abordable cuando los formatos de entrada sean formatos estándar, como el A4 para el papel, ya que en estos casos la disponibilidad de las herramientas a utilizar en estos procesos es mucho mayor.

El rango de los tipos de información a los cuales podemos aplicar la digitalización es prácticamente ilimitado; así la entrada original de información podrían ser microfichas de fondos documentales antiguos, archivos, manuscritos, etc. sobre los cuales aplicaríamos estos procesos de digitalización.

### **Porqué digitalizar los documentos**

Podemos ver algunos planteamientos que podrían llevarnos a pensar que una determinada organización necesita digitalizar sus documentos. Estos serían algunos de ellos:

- ✂ ✂ Las microfichas son difíciles de manejar. Busco alternativas para guardar un volumen importante de documentos.
- ✂ ✂ En mi organización se generan multitud de documentos diferentes. Estamos estudiando escanearlos pero nos gustaría clasificarlos automáticamente.
- ✂ ✂ Mi empresa genera un montón de listados, albaranes, facturas, etc. Además de distribuir una copia a nuestros clientes, nos quedamos una copia para nosotros. Esta copia que nos quedamos debemos archivarla convenientemente, con el consiguiente coste. Los listados deben enviarse a todas nuestras sucursales para su consulta. Nuestro consumo de papel se ha disparado !!!
- ✂ ✂ Tengo miles de documentos con la misma estructura, son hojas de pedidos, formularios, etc. Deseo conservarlos un tiempo y además me cuesta mucho esfuerzo introducir bs datos contenidos en ellos en el ordenador.
- ✂ ✂ Tenemos multitud de planos en papel. Nos ocupan espacio y nunca encontramos el que buscamos. ¿Qué se puede hacer? Ahora trabajamos con AutoCAD. Junto con los planos hay otra información que deseamos guardar como son presupuestos, fotos, anexos técnicos, etc.
- ✂ ✂ Desearíamos archivar de una forma estructurada toda la documentación que se genera en ofimática. Cartas de Word, Hojas de cálculo, etc. El problema principal es localizar lo que se busca. Nunca me acuerdo del nombre del fichero. ¿Hay alguna técnica especial?
- ✂ ✂ Tengo recortes de prensa, que quiero archivar y localizar con facilidad. ¿Cómo puedo manejarlos?
- ✂ ✂ Recibimos multitud de faxes cada día. Al final del año llenamos carpetas y carpetas. Ya nos compramos un fax de chorro de tinta pues el papel

térmico era un desastre; a los dos meses los faxes se habían borrado. Ahora ya no se borran pero hay tantos que cada vez que buscamos uno perdemos un cuarto de hora....

- ✂✂ Mi empresa comercializa multitud de artículos. Nos gustaría crear un catálogo electrónico con la documentación de todos ellos. ¿Qué herramienta podemos usar para crear este catálogo? ¿Puedo actualizarlo con facilidad?

Estos planteamientos aportan una visión más general que la simple digitalización. Algunos de ellos nos plantean además la necesidad de sistemas de gestión documental.

La digitalización, además de ser un proceso necesario para conseguir la homogenización de los documentos, nos da las siguientes ventajas o posibilidades:

- ✂✂ El hecho de digitalizar, de tener una versión digital, permite preservar la información de forma que permita utilizarla, reproducirla o recuperarla tantas veces como queramos. Por ejemplo en caso de posible degradación de los originales el trabajar con copias digitales es una buena alternativa.
- ✂✂ Evidentemente el cambiar de soporte, por ejemplo de papel a CD-ROM , va a reducir enormemente el espacio que ocupa la información.
- ✂✂ Además, este aumento de la manejabilidad nos va a facilitar la distribución de la información.
- ✂✂ Muy relacionado con la distribución están la publicación y difusión de la información. El disponer de grandes volúmenes de información en formato digital posibilita el incluirlos en sistemas de publicación tipo web que darán a esta información un carácter prácticamente universal sin apenas esfuerzo.

Toda información puede publicarse como tal pero admite tratamientos que la dotan de valor añadido, Estos tratamientos pueden automatizarse gracias al carácter electrónico de la información, que facilita al mismo tiempo su gestión y utilización.

En definitiva, la digitalización de la información va a proporcionarnos la posibilidad de gestionar y acceder a grandes volúmenes de información de forma rápida, eficiente y eficaz.

## **Conceptos sobre el color digital**

---

Como paso previo a la profundización en los conceptos relativos a la digitalización, necesario para su comprensión vamos a analizar una serie de puntos relativos a la teoría del color digital.

Al descomponer la luz encontramos que está compuesto por siete colores (existen otros muchos más pero que no son detectados por el ojo humano).

Básicamente los colores se descomponen en dos tipos: primarios o colores luz y complementarios o colores pigmento.

Los colores primarios o colores luz, son el rojo, verde y el azul (en inglés RGB) y permiten construir el resto de los colores como combinación de diferentes cantidades de ellos. Al superponerlos reducen o cambian su color, y al superponerlos todos producen el color blanco o ausencia total de color. Por eso también se llaman sustractivos.

Los colores complementarios son el cian, magenta y amarillo (CMY en inglés), se generan a partir de los primario y su suma produce el negro.

Cuando se trabaja con la luz, en la captación, generación de imágenes se utilizan esquemas de color basados en los colores primarios para conseguir el resto de colores. En cambio, cuando el resultado es físico, se imprime en papel, etc. se usan esquemas de color basados en los colores complementarios.

Es en este tipo de esquemas en los que se basarán los diferentes modelos de color que se utilizan en la definición de las imágenes digitales.

### **Tipos de imágenes digitales**

En el mundo de la imagen digital hay que diferenciar dos tipos básicos de imágenes, las imágenes bitmap (no confundir con el formato bitmap) y las imágenes vectoriales. Aunque comentaremos ambos tipos nos centraremos en las imágenes bitmap ya que son las que se obtienen como resultado de los procesos de digitalización por medio de escáneres y cámaras. Las imágenes vectoriales, aunque se pueden conseguir a través de procesos aplicados sobre imágenes bitmap, normalmente son generadas a partir de cero.

#### Imágenes vectoriales

Las imágenes se reducen a formas básicas: puntos, líneas, curvas, polígonos, representados por medio de formulas matemáticas y colores y tramas de relleno y trazado con las que se configura la imagen final. Es lo que se conoce como imagen vectorial.

La definición esquemática de la imagen ofrece una serie de ventajas como:

- ✂ El tamaño de las imágenes es muy ajustado y reducido. En lugar de pesadas imágenes, difíciles de manejar, tenemos unas pocas fórmulas matemáticas que apenas ocupan espacio.

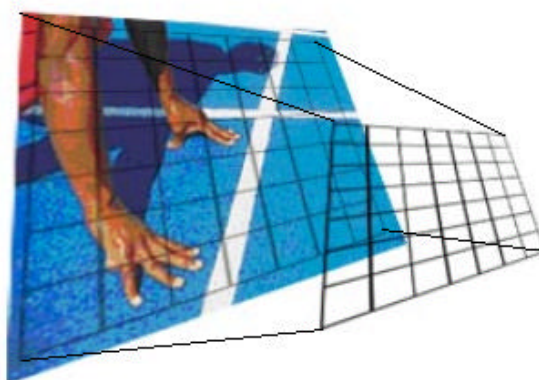
☞ Además, un dibujo o ilustración de este tipo puede redimensionarse, deformarse, descomponerse en sus distintos elementos y volverse a reconstruir sin ninguna pérdida de calidad. La escalabilidad de la imagen no depende de la resolución. Es más, en las imágenes vectoriales el concepto de resolución no tiene sentido. La calidad de la representación depende del dispositivo sobre el que la representemos o visualicemos y no de la imagen en sí..

A la hora de representarlos en el ordenador es necesario un proceso de rasterización de las imágenes vectoriales, hay que convertirlas en imágenes bitmap ya que las imágenes que vemos en pantalla o impresas son en última instancia un mapa de bits.

Es el sistema preferido para la ilustración y el diseño. Se utiliza en la definición de gráficos, logos.... Los logos son un ejemplo claro de la utilidad de los gráficos vectoriales, ya que una única versión de la imagen nos sirve para representarla en diferentes tamaños, formatos, ...Por ejemplo, de un fichero de 200kb podríamos obtener un póster de 12 con una nitidez y una calidad perfectas.

### Imágenes bitmap o mapa de bits

Una imagen bitmap (no confundir con formato bitmap) o de mapa de bits es el nombre que se ha hecho popular para el tipo de imágenes digitales que consisten en un conjunto, o *mapa*, de puntos ó *píxeles*, cada uno de ellos con su propio valor de color / luminosidad. Es también el resultado de la digitalización de una imagen.



Los píxeles o puntos de imagen suelen ser cuadrados. Como consecuencia, las imágenes digitales son siempre cuadradas o rectangulares y se caracterizan por tener cuatro características básicas: resolución, dimensiones, profundidad de bits y modelo cromático.

Cuando se habla de digitalización de imágenes de estas cuatro características hay que tener en cuenta básicamente dos de ellas. La primera es la profundidad de color, o dicho de otra forma, el número máximo de colores que el dispositivo digitalizador es capaz de reconocer en la imagen original.

Como ya hemos visto, en una imagen tipo bitmap la única información que cada uno de los píxel nos da es la correspondiente al color. Esta información que se traduce en bits es lo denominamos profundidad de color. Sabemos también que los colores se descomponen en sus componentes, según el esquema de color que se utiliza, y tenemos que guardar la información de todas las componentes.

Obviamente, a mayor valor en la profundidad de color obtendremos una mayor fidelidad en los colores. Así, tenemos las imágenes con una profundidad de color de un bit que se corresponden con las imágenes en blanco y negro en las que el bit con dos posibles valores, 1 o 0, nos dice si el punto es blanco o negro. En color podemos usar un byte (8 bits) que nos permite una gama de 256 valores para cada componente que multiplicado por tres nos da una calidad de color de 24 bits que se traduce en algo más de 16 millones de colores, que es más o menos lo que el ojo humano es capaz de discriminar y que es lo que se llama "color real". El modelo RGB es una de las formas habituales de color de 24 bits.

La segunda de las características principales de las imágenes digitalizadas es la resolución da idea de la cuadrícula colocada sobre la imagen y que nos descompone ésta en los puntos de color (píxel) que la forman. El número de columnas y filas de esta cuadrícula es lo que se denomina resolución, en definitiva es el número de puntos que conforman el mapa de bits. Para un mismo tamaño de imagen, cuantos más puntos de imagen tengamos mejor será la imagen. O lo que es lo mismo, a mayor resolución mayor calidad de imagen.

Se mide en puntos por pulgada (ppp o ppi, points per inch en inglés). A la hora de especificar la calidad la imagen de la que disponemos podemos especificarlo indicando el número de píxeles que la componen (600x800 píxeles) o dando el tamaño de la imagen y la resolución por unidad de longitud, (10x15 pulgadas y 72ppi)

### **Modelos de color**

Otra de las características de las imágenes digitales es el modelo cromático o modelo de color que utiliza para codificar la información de color. Con esta denominación nos referimos a las formas en que podemos representar el color de la naturaleza por medios artificiales. En la definición del color ya hemos mencionado dos de los más conocidos el RGB y el CMY que se basan en la separación en de tres componentes de color. Lo que se guarda como información son los valores de esos tres colores para cada píxel.

Sin embargo, este último sistema es bastante imperfecto, y a pesar que de la suma, a partes iguales, de CMY debiera resultar el color negro, sólo se logra un gris oscuro de aspecto sucio. Para corregir este defecto se añadió un nuevo componente de color de tonalidad negra (lo que compone el modelo CMYK) que persigue que los colores sean más nítidos y el negro tenga la tonalidad adecuada.

Pero tanto uno como otro tienen sus inconvenientes, por lo cual se han creado otros modelos de color que utilizan codificaciones de color diferentes, siendo uno de los más usados es el  $L^*a^*b$ . A grandes rasgos, es un modelo de color propuesto por la Commission Internationale d'Eclairage, encargada de realizar un estándar de codificación del color, de forma que cada color se componga de tres canales de información: uno de luminosidad y dos de color, variando el canal «a» de verde a rojo, y el «b» de azul a amarillo. El gran logro de este modelo de color es que se utiliza como estándar independiente del



dispositivo y no está ligado a pantallas, como RGB, o a dispositivos de impresión, como CMYK.

Otro de modelos de codificación del color usado con frecuencia es el HSB, el cual se caracteriza por almacenar tres parámetros por color: uno indica la tonalidad, otro la saturación o cromatismo y, por último, otro el brillo.

Otro modelo de tres canales, que soporta la misma cantidad de información y, por tanto, proporcionan la misma calidad que el RGB, desarrollada por Kodak para su sistema PhotoCd es el YCC.

Por último, hay otros modelos de color, que son menos importantes para nosotros: Indexado, bitmap o duotono. Ninguno de ellos podemos admitirlo como entrada de imágenes directa, aunque en ciertos casos, podrían usarse incluidos en documentos de programas vectoriales. Es más, para ciertos efectos son imprescindibles.

No hay que olvidar que estas consideraciones sobre los modelos de color se aplican, no a un determinado tipo de ficheros, sino a la manera en que se almacena la información de color en los documentos, sean estos bitmap o vectoriales. Existe otra manera de hacer esto: consiste en utilizar modelos estándar, de los cuales el más conocido, especialmente en España, es el Pantone.

### **Tamaño de los ficheros**

Teniendo en cuenta lo anteriormente expuesto, el modo habitual de capturar cualquier imagen sería llevado a cabo a la máxima resolución y a la máxima profundidad de color posible. Este procedimiento, aunque el ideal, no siempre resulta factible y aconsejable, debido a la ingente cantidad de información que se genera en cada una de las digitalizaciones. Así, debido a las limitaciones impuestas por el hardware hay que buscar un equilibrio entre el tamaño y la calidad en función de las necesidades y la utilidad que vayamos a dar a la imagen.

El tamaño del fichero correspondiente a la imagen digitalizada dependerá de la resolución y la profundidad del color en principio, aunque también del modelo de color utilizado. Suponiendo que utilizamos un byte para codificar la información de cada una de las componentes de color de un píxel, la siguiente fórmula nos dará el tamaño del fichero de imagen:

*(res. x an.) x (res. x al.) x nº de canales de color = tamaño del fichero en bytes*

El trabajo con imágenes de calidad, conlleva inevitablemente grandes volúmenes de información cuya problemática de manipulación y almacenamiento hay que solventar con las técnicas de compresión que utilizan algunos de los formatos de ficheros gráficos más comunes.

### **Compresión**

Por medio de la utilización de algoritmos de compresión se consigue reducir el tamaño del archivo a costa del tiempo invertido en su compresión y

descompresión. Métodos de compresión hay muchos, aunque los más usados son el algoritmo LZW, propiedad de Unisys y empleado por los formatos GIF y en algunos TIFF, y el algoritmo de compresión con pérdidas que incorpora el formato JPEG.

Las diferencias existentes entre unos y otros radican en su propia capacidad de compresión y en la pérdida de calidad. El algoritmo de compresión utilizado por el formato JPEG es mejor que el del LZW, aunque este último conserva la imagen intacta mientras que el sistema JPEG, según el nivel de compresión seleccionado, ofrece una mayor o menor pérdida de calidad.

LZW, se basa en el análisis de la imagen buscando zonas de color homogéneo, o detalles que se repiten, etc. De esta forma, dependiendo de las características del fichero, pueden obtenerse ratios de compresión muy importantes sin pérdida alguna de información.

JPEG se basan en que se puede prescindir de una cierta cantidad de información sin que sea percibido por los sentidos humanos. Utiliza para ello la DCT (Transformada Digital del Coseno).

En este caso la utilización de uno u otro tipo de compresión de depende si la pérdida de información compensa la manejabilidad y tamaño ajustado.

### **Formatos de ficheros**

Todas estas posibilidades derivan en la proliferación de formatos de fichero para imágenes digitales, de los cuales hay unos pocos que son en definitiva los mas utilizados que enumeramos a continuación:

| <u>Vectoriales</u>      |  |
|-------------------------|--|
| WMF (Windows Meta File) |  |
| EPS                     | El estándar en cuanto a imágenes vectoriales |
| <u>Mapas de bits</u>    |  |
| BMP:                    | Formato estándar                             |
| TIFF:                   | Compresión LZW y RLE (sin pérdidas)          |
| GIF                     | Compresión LZW. Uso HTML y gráficos en color |
| JPEG                    | Compresión JPEG                              |
| PHOTO CD                | Desarrolla do por Kodak                      |

## Herramientas para la digitalización

---

Una vez repasada la teoría y conceptos más importantes que definen la imagen digital, nos será mucho más fácil comprender los mecanismos que intervienen en el proceso de digitalización de una imagen por parte de los escáneres y cámaras digitales. El principio de funcionamiento de estos dispositivos es la digitalización, es decir, la conversión de una información analógica en datos comprensibles por cualquier ordenador; y para lograr este objetivo, se valen de una serie de componentes internos destinados a realizar dicho cometido. Estos componentes aunque básicamente tienen que hacer lo mismo son sensiblemente diferentes ya que su utilización esta orientada hacia ámbitos que si no contrapuestos, sí son complementarios.

### Escáneres

Los escáneres son dispositivos de digitalización en principio 'fijos'. Aunque existen productos de pequeño tamaño, que podemos utilizar como si se tratase de un lápiz. En general son equipos de gran tamaño que se utilizan en procesos de digitalización de documentos, a veces de forma masiva y mecánica. Pueden llegar a ofrecer un calidad de digitalización muy alta, en parte propiciado por ese carácter de 'fijo'.

#### Cómo funciona un escáner.

En primer lugar, antes de iniciar cualquier operación habrá que tener en cuenta la naturaleza del soporte sobre el cual se encuentra la imagen original. Obviamente, no será lo mismo capturar una imagen de una fotografía impresa en el típico papel fotográfico, que hacer este mismo proceso partiendo de una transparencia o negativo, o capturar la información de documentos escritos. Así pues, el procedimiento para llevar a cabo la correspondiente exploración de la imagen será diferente en función del tipo de original. No obstante, las diferencias entre un procedimiento y otro son mínimas, por lo cual una explicación general será más que suficiente para entender el funcionamiento básico de un escáner.

Los dispositivos de digitalización de imágenes basan su funcionamiento en la tecnología CCD (Charge Couple Device), un conjunto de elementos fotoeléctricos o fotodiodos orientados linealmente cuyo número (o más bien densidad lineal) determina la resolución horizontal de lectura, que se mide en puntos por pulgada (ppp). La resolución vertical de la imagen es establecida por la velocidad relativa que alcanza el tractor de avance de los diodos (en el caso de los escáneres planos).

Sin entrar en demasiados pormenores, el CCD es un elemento formado por un conjunto numeroso de transistores sensibles a la luz, los cuales reaccionan eléctricamente ante las diferentes intensidades de luz que reciban. El proceso se basa en la adecuada iluminación de aquella cara del papel de la cual se quiere obtener una imagen digitalizada. De este modo, la luz que refleja el propio papel es dirigida a través de una o varias lentes al sistema de sensores de luz conocido (CCD).

Los datos recogidos son, por tanto, una señal analógica puesto que cada elemento óptico proporciona un voltaje proporcional a la cantidad de luz que cae sobre él cuando el documento es iluminado por la luz fluorescente. Un

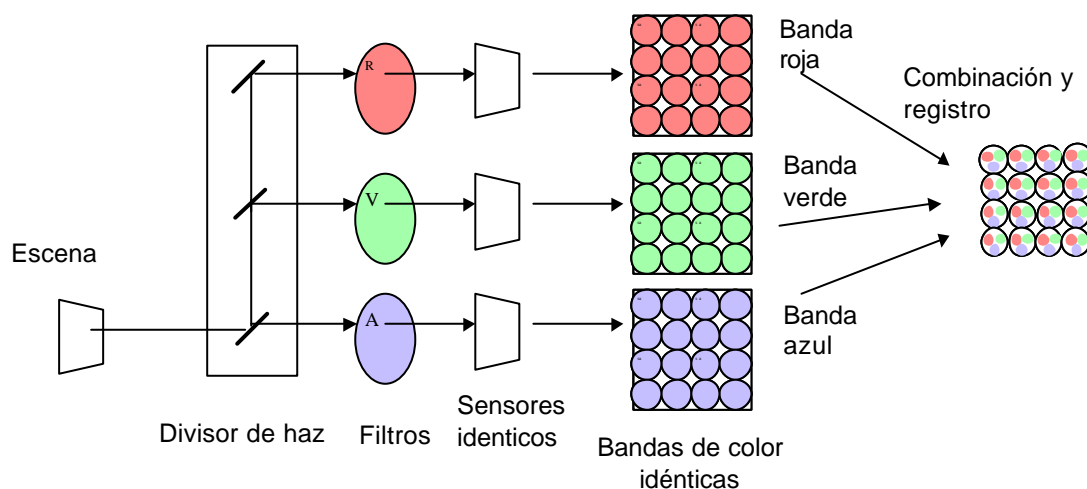
punto negro del documento absorbe la mayor parte de la luz, permitiendo que muy poca se refleje en el CCD (...) Un punto blanco refleja la mayor parte de la luz, dando como resultado una salida de alto voltaje.

En el caso de imágenes en color, el proceso se repite tres veces con la interposición antes de los CCD de unos filtros de colores que separan la luz en sus tres componentes de color (rojo, verde y azul) para determinar la cantidad de esos colores primarios requerida.

En un segundo proceso, tanto para imágenes en grises o en color, las distintas intensidades eléctricas de carácter analógico (variaciones continuas de intensidad) generadas en el CCD se transforman mediante un conversor analógico a digital o DAC a su correspondiente formato digital, una trama de bits, con el objeto de que cualquier sistema informático sea capaz de interpretarlas y comprenderlas.

Este proceso se repite para todas y cada una de las líneas que conforman la imagen o el documento a capturar, de modo que el carro móvil se irá desplazando con una velocidad lineal constante para proceder a digitalizar la siguiente línea. La posición y el tipo de movimiento permite clasificar los escáneres en diferentes tipos.

En la siguiente figura podemos ver esquemáticamente los pasos de la digitalización de una escena en un escáner a color.



Cuando la imagen a digitalizar está contenida en un soporte transparente o translúcido, se da la circunstancia de que el necesario fenómeno de la reflexión de la luz no se produce adecuadamente. En tal caso, el soporte que contenga la imagen a digitalizar deberá ser iluminado por su otra cara (por arriba en los escáneres planos). Se utilizan para ello adaptadores de transparencias. De este modo, la luz emitida por el adaptador de transparencias atraviesa la diapositiva o negativo, con lo que los cambios de luminosidad serán recogidos

convenientemente por el mismo sensor CCD. Existen así mismo como veremos dispositivos específicos para la captura de imágenes en este tipo de soporte.

Según se van digitalizando las sucesivas líneas, la información digital extraída suele almacenarse en una pequeña caché de memoria incluida en el propio escáner. Naturalmente, también existe un elemental microcontrolador encargado de gestionar el periférico, pudiendo realizar algunos ajustes sobre la imagen a medida que se digitaliza, como ocurre en el caso de los escáneres que realizan de forma automática ajustes en la gama de colores.

### La resolución

Aunque ya se ha mencionado, consideramos necesario hacer hincapié en la principal característica de un escáner, la resolución. Hay que diferenciar los tres tipos de resolución que podremos encontrar referenciadas en las especificaciones de un escáner.

La resolución óptica es una de las características más relevantes en un escáner, y está asociada al valor de la resolución horizontal. La determina el número de sensores integrados en el CCD. Nos da idea de la cantidad de sensores individuales que se utilizan para capturar una línea.

En cambio, la resolución mecánica o resolución vertical, viene determinada por la distancia recorrida por el módulo de reproducción durante la exposición de una línea a la siguiente y nos da idea del número de pasos o líneas que el escáner es capaz de reproducir por unidad de longitud.

Como ejemplo, cuando vemos especificada una resolución real de 600 x 1.200 ppp., valores correspondientes a un escáner de sobremesa normal, estas cifras nos vienen a decir que el primer valor de 600 ppp. es la resolución máxima óptica del CCD, mientras que el segundo número refleja el número máximo de pasos por pulgada que puede dar el escáner en el plano vertical.

La mayoría de los fabricantes de escáneres ofrecen valores de resolución muy por encima de la resolución óptica y mecánica del dispositivo. Para lograr estas espectaculares cifras, los distintos fabricantes han desarrollado métodos y técnicas especiales para incrementar la resolución mediante la interpolación por software o hardware. Es lo que se denomina resolución máxima interpolada. Esta mejora se consigue por medio de técnicas que incrementan la resolución real de una imagen rellenando los espacios existentes entre los puntos en función del color de los puntos adyacentes. De forma sencilla podemos decir que lo que hace el escáner, ya que no es capaz de darnos esa resolución real, es intercalar por su cuenta líneas que luego rellena con puntos de color que se adecuen a los de su alrededor que son los que el escáner si ha capturado. Conseguimos así imágenes con más líneas por pulgada pero algunas de las cuales han sido "calculadas" por el escáner. Este proceso normalmente realizado por el propio escáner afecta en gran medida a la velocidad de exploración. Hay herramientas de software para el tratamiento y retoque fotográfico que permiten realizar esta misma operación sobre imágenes ya digitalizadas y que consiguen resultados muy similares.

### Tipos de escáneres

Existen varios tipos de escáneres cuyo uso depende del destino de los documentos a digitalizar, del soporte de estos y de la calidad que queramos conseguir.

- ✍ En primer lugar, los digitalizadores planos o flatbeds son muy similares a una fotocopidora. Dentro de esta clasificación entran los escáneres de sobremesa de uso personal. En ellos, la cabeza lectora con los diodos, así como la lámpara de iluminación se mueven a lo largo del documento a digitalizar. En este grupo la gama es muy amplia pudiendo encontrar escáneres de 1 bit por píxel hasta 24 bits por píxel, con diferentes resoluciones.
- ✍ Los escáneres de tambor son aparatos muy complejos y de uso prácticamente exclusivo en los servicios de fotomecánica. La resolución que son capaces de conseguir estos dispositivos supera a la de cualquier otro tipo de escáner, pudiendo alcanzar sin dificultad resoluciones de miles de puntos por pulgada. En este casos la cabeza lectora o de exploración es fija, y es el original que fijado a un tambor gira a su alrededor a gran velocidad.
- ✍ Como ya hemos mencionado, existen aparatos especializados mediante los cuales es posible la digitalización directa de película impresionada -negativo o diapositivas- sin necesidad de positivarla previamente. Son los digitalizadores de diapositivas. Lógicamente estos dispositivos trabajan a resoluciones bastante altas debido al pequeño formato del soporte original.
- ✍ Los escáneres de exploración superior o cenitales son un caso especial de dispositivo de digitalización al que se da el uso de un escáner a cámaras de digitalización. Se asemejan a aparatos de retroproyección en los que la lente de proyección se ha sustituido por una cámara digital. Se utilizan por ejemplo en procesos de digitalización de documentos delicados que se colocan en un soporte que en algunos casos es capaz de adaptarse al mismo documento para minimizar al máximo los problemas que se puedan derivar de la manipulación de estos.

Merece la pena hacer una ultima consideración respecto a los escáneres y mencionar la existencia de un estándar definido por empresas dedicadas a la fabricación de escáneres y empresas relacionadas que desarrollaban aplicaciones que los utilizaban denominado Twain.

El estándar Twain permite utilizar cualquier escáner desde cualquier programa de edición o retoque de imágenes, o de OCR. El usuario puede incluir en su programa una imagen capturada desde un escáner sin tener que utilizar ningún otro programa.

Son los escáneres de sobremesa los que ofrecen una mayor versatilidad en sus controladores Twain.

### **Cámaras digitales**

El segundo dispositivo de digitalización que hemos mencionado es la cámara digital. Si bien, su aplicación al ámbito de la gestión de la información

es mucho más limitada que la de los escáneres, sí merece un breve comentario.

Se podría decir que la diferencia básica entre un escáner y una cámara digital estriba en su modo de uso. Una cámara digital lo que nos permite es capturar cualquier imagen, escena o documento desde cualquier posición; mientras que los escáneres, como ya hemos dicho, son dispositivos 'fijos' que ofrecen unas calidades de digitalización muy superiores, como contraprestación a la pérdida de flexibilidad de uso que ofrecen respecto a las cámaras digitales.

Hace relativamente poco tiempo el disponer de una cámara digital era algo impensable para la gran mayoría de la gente. La calidad de la fotografía convencional estaba muy por encima de la digital. Los avances de la tecnología, en parte impulsados por el boom de internet y la necesidad de producción de imágenes digitales generada por éste, han convertido a este tipo de cámaras en objetos de consumo casi habitual, y que en productos de gama alta llega alcanzar niveles de calidad comparables a los que ofrece las cámaras convencionales.

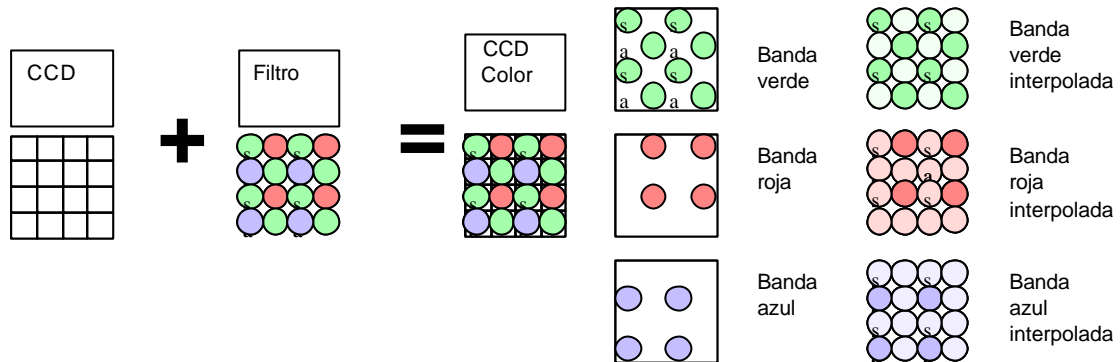
Las ventajas que ofrece la propia digitalización, por ejemplo el aumento de la perdurabilidad de cualquier fotografía que tomemos, o la inmediatez de la producción de imágenes digitales para publicación electrónica, y las ventajas que ofrece la cámara, como la posibilidad de reutilizar el carrete que es sustituido por un dispositivo de almacenamiento temporal que podemos utilizar tantas veces como queramos, o de ver cómo estamos componiendo la fotografía en el pequeño LCD que la mayoría de las cámaras tienen, o la capacidad de la cámara de adaptarse por sí misma al medio en el que se encuentra el usuario liberando a éste de la preocupación de configurarla, equilibran la comparación entre cámara digital y convencional, y hacen que el planteamiento de optar por una u otra sea más complicado.

La estructura de una cámara digital, es fundamentalmente equivalente a una cámara réflex convencional, y se diferencian en la forma de plasmar la imagen definitiva. La luz es conducida a través del objetivo hasta el dispositivo encargado de generar la imagen digital que almacena la cámara.

El mecanismo de generación de la imagen digital de una cámara es básicamente el mismo que el de los escáneres, ya que también están basados en los mismos elementos de captura de información: los CCD. La diferencia radica en la forma que se hace llegar la luz hasta los sensores del CCD, ya que se dirige por medio de lentes a través del objetivo hasta el CCD, en que es la propia cámara la que almacena la imagen digitalizada, y por último y más importante en el propio CCD.

El CCD de las cámaras digitales es bidimensional y no lineal. La imagen se compone de una vez y no por líneas. Esto tiene una importante implicación en la captura de imágenes en color. Si en los escáneres se hacían tres 'pasadas' para tomar la información de las tres componentes de color y se utilizaban todos los sensores para cada una de las componentes, el capturar la imagen de una vez obliga a repartir los sensores del CCD que se utilizan para cada componente de color y que la filtran individualmente. La combinación de las componentes se hace de forma similar, pero la cantidad de información de la que se dispone para cada componente es menor, y por consiguiente también la

resolución. Esta es la causa de la diferencia de calidad entre escáneres y cámaras. Se puede llegar a 3000x2000 píxeles pero lo normal es 2240x1680 píxeles. En la siguiente figura podemos ver esquemáticamente ese mecanismo.



Existen nuevas tecnologías como la CMOS que compiten con los CCD y que subsanan estos y algunos otros inconvenientes como el consumo, coste, ...

En cuanto a almacenamiento sólo mencionar algunas de las diferentes tecnologías que han surgido de la lógica evolución de estos dispositivos. Actualmente, aunque se está trabajando con muchos otros sistemas, los que están adquiriendo una mayor aceptación por parte de los fabricantes son las tarjetas flash, como SmartMedia y CompactFlash. Aunque ahora éstos son los dispositivos más utilizados, muchas compañías ya están probando nuevas tecnologías, como el dispositivo Klik! de Iomega, cuya filosofía es similar a la de los conocidos discos extraíbles Zip de esta misma compañía, o el Microdrive de IBM. Existe otras formas de almacenamiento, algunas más convencionales como el CD-R o el disquete de 3.5 que ofrece Sony o el Memory Stick también de Sony.

Los intentos de estandarización para posibilitar el intercambio de información entre distintos fabricantes se están dando en los sistemas de almacenamiento, pero también en ámbitos como los sistemas operativos para cámaras digitales como Digita o especificaciones para la definición de formatos de impresión integrados con las imágenes como el DPOF (Digital Print Order Format).

Hay otra serie de características más relacionadas con su uso que también tenemos que tener en cuenta al analizar una cámara digital, como son el tipo de alimentación, el consumo, la resolución que ofrece, la memoria, la ergonomía, el software incluido para su utilización, la conexión con el ordenador, y el burst mode que es el tiempo que hay que dejar pasar entre captura y captura.



## **Conexión al ordenador**

Tanto para escáneres como para cámaras la forma de conectarlos al ordenador, que es en definitiva el destino final de los documentos, suele estar basada en tecnologías comunes. Las más normales son

- ✂ SCSI: la más común y garantizada con los S.O. más modernos. Ofrece una gran velocidad de transferencia. Permite la conexión en cascada de varios dispositivos.
- ✂ USB: facilita la conectividad de periféricos ya que los ordenadores disponen de puertos USB sin necesidad de instalar ninguna tarjeta especial. Es el sustituto de los típicos puertos serie y paralelo. Permite conexión de dispositivos en cascada.
- ✂ IEEE 1394 – Wirefire: Combina las ventajas del bus SCSI y la sencillez y coste del USB. No todos los S.O. Ofrecen el soporte necesario.
- ✂ Video: tiende a ser desplazado ya que requiere la utilización de tarjetas de interfaz propietarias. Se responsabiliza el fabricante de la conectividad a un determinado S.O. Este tipo de conexión aparece en las cámaras digitales, más que en escáneres.

## **Tratamiento de las imágenes**

---

El resultado de la digitalización puede tener la calidad que nosotros hayamos especificado ya que prácticamente no hay limitaciones técnicas, pero aun así, los documentos digitalizados pueden ser objeto de una modificación o tratamiento posterior.

Hemos diferenciado dos tipos de tratamiento en el proceso general que se sigue para integrar los documentos en un sistema que permita su explotación final, el tratamiento de la imagen propiamente dicho y el OCR o reconocimiento óptico de caracteres.

### **Tratamiento digital de la imagen.**

Los programas de tratamiento de imagen nos van a permitir por un lado mejorar, o cambiar las propiedades de la imagen: brillo, contraste, pero también permite cambiar el tamaño, el formato, la resolución. Por ejemplo, utilizaremos las capacidades de estos programas para retocar las versiones gráficas de los documentos antes de aplicarles el OCR y así podremos eliminar partes del documento que no nos interesan pasar por el OCR ya que producirían más errores que resultados útiles. La eliminación de logos, sellos, u otro tipo de elementos procedentes de la preimpresión de documentos, o suciedad en el papel original son el tipo de problemas que se intentan resolver por medio de los programas de tratamiento de imagen. Photoshop de Adobe es quizá uno de los programas de retoque más utilizado.

### **Reconocimiento Óptico de Caracteres (OCR)**

El reconocimiento óptico de caracteres es el proceso mediante el cual, a partir de la imagen de un documento se reconocen los caracteres en él contenidos y que se almacenan en un fichero con el texto

El proceso de OCR no siempre es capaz de leer la totalidad del contenido de los documentos. Algunas palabras pueden estar mal interpretadas o algunos caracteres pueden ser erróneos. La tasa de errores dependerá de la calidad y el tipo de original que se maneje. No obstante el OCR puede ahorrar muchísimo trabajo de introducción de datos. En algunas aplicaciones será necesario añadir un post-procesado al OCR para corregir los fallos del mismo, en otras aplicaciones no es necesario.

Para obtener el óptimo resultado de un proceso de OCR es necesario disponer de diccionarios especializados, que ayudarán al motor de reconocimiento. Así mismo estos diccionarios serán necesarios en el post-procesado de la información, para ayudar a realizar las correcciones. El uso de estos programas y métodos permite "obtener" información y almacenarla para poder ponerla a disposición del usuario y cada día es más usado en: empresas, bibliotecas, para la producción de libros digitales, etc.

Omnipage, TextBridge, u Open Book son productos para el reconocimiento óptico de caracteres que podemos encontrar en el mercado. Siendo quizá el primero el más popular.

## **Sistemas para el tratamiento y recuperación de la información documental**

---

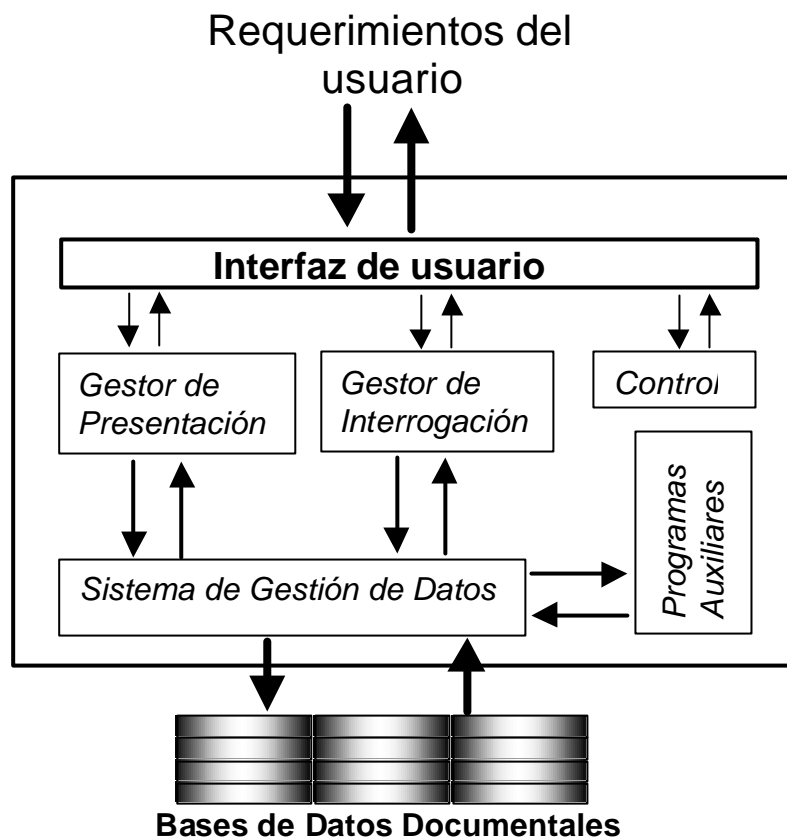
Una vez dispongamos de los documentos digitalizados necesitamos un sistema que permita su tratamiento y la recuperación en ellos contenida.

Las funciones que debería cumplir serían las siguientes:

- ✍ Creación de bases de datos documentales, con una estructura de campos
- ✍ 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.
- ✍ Recuperación de la información contenida en los documentos de la base de datos a través de cualquier término existente en ella, mediante ecuaciones de búsqueda sobre una o varias bases de datos simultáneamente. Los documentos resultantes se agrupan en sets o conjuntos, susceptibles de combinación posterior.

- ✂ Salida de la información, mediante edición en pantalla, impresión y redirección a ficheros de los documentos de interés para el usuario.
- ✂ Análisis de la respuesta, utilizando herramientas de análisis de frecuencias de los términos o de ocurrencias.
- ✂ Creación de ficheros que definan el perfil de búsqueda de los usuarios, así como de las ecuaciones que ejecuten
- ✂ Gestión de mecanismos de control terminológico (tesauros,...)
- ✂ Ayuda al usuario en todo momento.
- ✂ Dependiendo de la configuración del sistema, éste puede ofrecer opciones de acceso multiusuario, niveles de seguridad, reorganización y recuperación de ficheros, etc.

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. Aunque este tipo de aplicaciones ofrecen la mayoría de las funcionalidades que se le exigen a un sistema para el tratamiento y recuperación de información documental, se ha producido un proceso de especialización en la informática documental que ha producido, a su vez, aplicaciones específicas para diferentes aspectos, que en muchas ocasiones entrecruzan sus características.



Como elemento central de este tipo de aplicaciones están los 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...

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

Los sistemas de exploración o escáneres son 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

Cuando nos ceñimos exclusivamente a la de recuperación de información: tenemos aplicaciones que utilizan para ello 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.

Hay además otras aplicaciones específicas orientadas a entornos concretos, administrativo o personal, y sistemas compuestos que combinan la gestión administrativa y la documental en un mismo producto como Absys.

## Conclusión

---

La digitalización de la información va a proporcionarnos la posibilidad de gestionar y acceder a grandes volúmenes de información de forma rápida, eficiente y eficaz.

Las ventajas que así obtenemos hay que contrastarla con la gran cantidad de trabajo que a veces es necesario para conseguir los documentos digitalizados generada por el propio proceso de digitalización y otros derivados de la adaptación del producto de la digitalización como el OCR o la extracción de información estructurada.

A medida que el trabajo cotidiano se integre en los sistemas informáticos, el trabajo de digitalización tenderá desaparecer y será sustituido por procesos orientados a dotar de valor añadido a la información para cuya integración y gestión las bases de datos documentales se configuran como la herramienta ideal que además van a permitir al usuario final acceder a la información tanto de forma directa, on-line, a través de la publicación web, o como productos en formato compacto para distribución masiva como los cd-rom o dvd.

## Bibliografía de interés

---

- ≠≠ Apolonio, Laura Tratamiento electrónico de la imagen / [coordinación general, Javier Cobo Insausti, Javier Jiménez Ortiz] Granada : Ediciones la Montaña, 1997-1998. 3 v. : il. col. ; 22 x 23 cm
- ≠≠ ASHFORD, J.H., y WILLET, P. (1988): *Text retrieval and document databases*. London: Chartwell Bratt.
- ≠≠ BATES, M.J. (1981): "Search techniques." *ARIST*, 16, pp. 139-169
- ≠≠ BUSON, R. (1993): "Base de datos documentales." *PC World*, Enero, pp.120-126.

- ✂ BUSTELO, C. (1994): "El papel de los expertos en documentación en la implantación de tecnologías de gestión de documentos." *Actas de las IV Jornadas Españolas de Documentación Automatizada DOCUMAT 94*, Gijón: Universidad de Oviedo, pp.357-362.
- ✂ CANALS CABIRÓ, I. (1990): "El concepto de hipertexto y el futuro de la documentación." *Terceras Jornadas Españolas de Documentación Automatizada Documat 90*. Palma de Mallorca: Univ. de les Illes Balears, pp.49-76.
- ✂ CARIDAD, M., y MOSCOSO, P. (1991): *Los sistemas de hipertexto e hipermedias. Una nueva aplicación en informática documental*. Madrid: Fundación Germán Sánchez Ruipérez.
- ✂ CHAUMIER, J. (1993): *Técnicas de documentación y archivo*. Barcelona: oikos-tau.
- ✂ CODINA, L. (1990): "Bases de datos documentales para microordenadores." *Actas de las 3ª Jornadas Españolas de Documentación Automatizada DOCUMAT 90*, Palma de Mallorca: Universitat de les Illes Balears, pp. 618-627.
- ✂ Cohen, Luanne S. *Técnicas para el tratamiento de imágenes / Luanne Seymour Cohen, Russell Brown, Tanya Wendling ; [traductor, María del Pilar Toldos Bayle]* Madrid : Anaya Multimedia, [1998]VI, 215 p. : il. col. ; 24 cm
- ✂ COX, J. (1991): *On-line and CD-ROM Database Searching*. London: Mansell.
- ✂ ESPINOSA, Blanca (et al.) *Tecnologías ópticas en documentación e información. "Cuadernos E.U.B.D. Complutense"*, vol 1, nº 1, 1991. Madrid.
- ✂ ESPINOSA, M<sup>a</sup>. B., et al. (1994): *Tecnologías documentales*. Madrid: Tecnidoc.
- ✂ FIDEL, R. et alii (eds.) (1994): *Challenges in Indexing Electronics Text and Images*. Medfor: Learned Information.
- ✂ Gradias, Michael [Schnell Anleitung Scannen. Español] *Escáneres / Michael Gradias ; [traducción, Núria González]* Barcelona : Marcombo, [1999] 159 p. ; 21 cm
- ✂ Gradias, Michael. *La digitalización de imágenes / Michael Gradias ; [traducción, Virginia Pérez, Nuria González]* Barcelona : Marcombo, [1998]285 p. ; 24 cm
- ✂ GREEN, W.B. (1993): *Introduction to Electronic Documents Management*. Boston: Academic Press.
- ✂ Jiménez Sender, Javier. *Introducción al tratamiento de imágenes*. Javier Jiménez Seder, Narciso García Santos [1ª ed.] Madrid : Universidad Autónoma, Centro de Investigación, 1982. 113 p. : il. ; 28 cm.
- ✂ LITC, (1994): *European Directory of Text Retrieval Software*. Aldershot: Gower.
- ✂ LIZASOAIN, L. (1992): *Bases de datos en CD-ROM*. Madrid: Paraninfo.
- ✂ LÓPEZ YEPES, A. (1993): *Documentación Multimedia*. Salamanca: Publicaciones de la Universidad Pontificia.
- ✂ LÓPEZ YEPES, J. (1995): *La Documentación como disciplina. Teoría e Historia*. Pamplona: EUNSA.
- ✂ LÓPEZ YEPES, J. (COORD.) (1996): *Manual de Información y Documentación*. Madrid: Pirámide.
- ✂ Martín Marcos, Alfonso L. *Compresión de imágenes : norma JPEG : Ciencia 3*, [1999] 256 p. : gráf., il. ; 24 cm
- ✂ OZKARAHAN, E. (1995): "Multimedia Document Retrieval." *Information Processing & Management*, 31, 1, pp.113ss.
- ✂ RECODER, M.J., ABADAL, E., y CODINA, L. (1991): *Información electrónica y nuevas tecnologías*. Barcelona: PPU.
- ✂ SAGREDO, F. *Automatización y tecnologías ópticas en información y documentación. Documentación del seminario "Documática hoy: sistemas de archivo óptico"*. Madrid, 16-XII-92.
- ✂ SALTON, G. (1989): *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer*. Reading: Addison-Wesley.
- ✂ SALTON, G., y MCGILL, J.M. (1983): *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.

- ≍≍ SEYER, P. (1991): *Hypertext. Concepts and Applications*. Blue Ridge Summit: Windcrest Books.
- ≍≍ SIEWERT, M.E.C. (1996): "Full-Text Information Retrieval: Introduction." *JASIS*, 47, 5, pp. 261-262.
- ≍≍ TRAMULLAS SAZ, J. (1995): "Una introducción a la Informática Documental." *Apuntes del Centro de Cálculo de la Universidad de Zaragoza*, 8, pp. 6-10.
- ≍≍ TRAMULLAS SAZ, J. (1996a): *Apuntes de Informática Documental*. Zaragoza: Kronos.
- ≍≍ TRAMULLAS SAZ, J. (1996b): "Investigación en sistemas bibliográficos personales: una revisión de aplicaciones *freeware* y *shareware*." Tramullas, J. (Ed.) *Tendencias de investigación en Documentación*. Zaragoza: Universidad de Zaragoza, pp.51-73.
- ≍≍ TRAMULLAS SAZ, J. (en prensa): "Recuperación de información en WWW: planteamiento, herramientas y perspectivas." *Anuari de SOCADI*, Barcelona: SOCADI.
- ≍≍ TRAMULLAS SAZ, J., y CUBILLO, A. (1995): *BRS/Search: Introducción a los sistemas de gestión de bases de datos documentales*. Zaragoza: Kronos.
- ≍≍ TURTLE, H.R. y CROFT, W.B. (1992): "A Comparison of Text Retrieval Models." *The Computer Journal* , 35, 3, pp. 279-290.
- ≍≍ WILLET, P. (ed.) (1988): *Document Retrieval Systems (The Foundations of Information Science, vol. 3)*. London: Taylor Graham.