

## Problemática de la informatización de un corpus de filigranas



*Por J.C. Lázaro Obensa*

Numeración Decimal:  
676.014.01

Ref.: FILIGRANAS  
ARCHIVOS  
INFORMATIZACIÓN  
HISTORIA DEL PAPEL

**RESUMEN.**—La rápida revolución producida en el campo de la informática, tanto en el avance de los dispositivos físicos como en los programas y soporte, nos ha llevado al concepto de ordenador personal. Las prestaciones y características que han adquirido estos sistemas durante los últimos años nos permite abordar problemas hasta hace poco impensables de resolver mediante un sistema personal. Un ejemplo de este tipo de aplicaciones podría ser la que nos ocupa en esta presentación: la informatización de un corpus de filigranas. Un sistema de este estilo permite que un investigador interesado en la historia del papel, puede disponer de una herramienta de gran ayuda, fácil manejo y coste relativamente reducido. Tras una introducción que pone de manifiesto esa evolución tecnológica, se comentan los aspectos particulares que caracterizan el problema. Posteriormente se realiza una breve discusión sobre las formas de aprovechar estas características para la realización óptima de una solución completa. Por último se comentan las enormes posibilidades de las redes informáticas como medio efectivo para intercambiar información entre instituciones e investigadores de distintos centros de todo el mundo.

**RESUMÉ.**—La rapide révolution produite dans le domaine de l'informatique même dans l'avance des dispositifs physiques que dans les programmes et le support, nous ont amené au concept de l'ordinateur personnel. Les prestations et les caractéristiques que ces systèmes ont acquis pendant les dernières années nous permettent d'aborder les problèmes jusqu'à il ne fait pas long temps inopinés de résoudre avec un système personnel. Un exemple de ce type d'applications pouvait être celle qui nous occupe dans cette présentation: l'informatisation d'un corpus de filigranes. Un système de ce style permet qu'un investigateur intéressé dans l'histoire du papier, puisse disposer d'un outil de grande aide, d'un facile mode d'emploi et un coût non très grand. Après d'une introduction qui met en évidence cette évolution technologique, on comment les aspects particuliers que caractérisent le problème. Postérieurement on fait une brève discussion sur les formes de profiter ces caractéristiques pour la réalisation optimale d'une solution complète. Comme

fin, on comment les grandes possibilités des réseaux informatiques comme un moyen effectif pour interchanger information entre institutions et investigateurs de différents centres de tout le monde.

**SUMMARY.**—The swift revolution in hardware and software, has driven the idea of personal computer. These systems have acquired features and characteristics that allow to solve problems not possible up to now. A sample of these type of applications is the computerization of a corpus of rubber marks. This type of systems would allow to a researcher to use a very efficient, easy to use and low price tool. The author begins with an introduction about this technological revolution, comments on some specific details, and a brief discussion on the ways to get the best benefit from this performance. At the end, the local networks chances as an effective way to interchange information between institutions and researchers are commented.

**ZUSAMMENFASSUNG.**—Die schnelle Entwicklung auf dem Gebiet der Informatik und bei den Fortschritten in der Hard- und Software hat uns zum PC geführt. Die Merkmale und Leistungen dieser Systeme erlauben es uns jetzt, Probleme zu anzugehen, die bisher mit einem auf Personen basierenden Systems unmöglich zu lösen waren. Ein uns in diesem Zusammenhang interessierendes Anwendungsbeispiel wäre die Informatisierung von Wasserzeichen. Ein System dieser Art ist für einen an der Papiergeschichte interessierten Forscher ein sehr hilfreiches, leicht zu handhabendes und relativ billiges Werkzeug. Nach einer Einführung in die technologische Entwicklung werden besondere Aspekte dieses Problems kommentiert. Danach folgen Kommentare über Nutzungsformen dieser Merkmale für die beste Durchführung einer vollständigen Lösung. Zu Schluss werden noch die enormen Möglichkeiten von Informatiknetzen als wirksames Mittel für den Austausch zwischen Institutionen und Forschern verschiedener Zentren in aller Welt erörtert.

## INTRODUCCIÓN

Durante los últimos años, hemos asistido a lo que podríamos denominar como la explosión de la informática personal. Hace unos años, el ordenador doméstico era prácticamente desconocido y quedaba reducido a pequeñas máquinas dotadas de un microprocesador de ocho bits que ni tan siquiera disponía de monitor propio, teniendo que hacer uso de los aparatos convencionales receptores de televisión, para visualizar la información. Este ejemplo típico de uso de un ordenador personal se correspondía con la inexistencia de equipos similares en el ámbito profesional. Las posibilidades de

estas máquinas en los entornos profesionales eran bastante limitadas pues no pasaban de ser un caro juguete de última generación.

Las máquinas empleadas en entornos profesionales, eran de gran tamaño, con precios exorbitantes, con elevadas exigencias de mantenimiento y tremendamente complicadas de manejar, lo que obligaba a disponer de personal altamente cualificado y escaso, con lo que los costos de explotación resultaban enormes. El concepto de ordenador personal no existía; a lo sumo una máquina podría ser gobernada desde varios terminales individuales pero los recursos del sistema eran compartidos por varios usuarios, mientras una o varias personas especializadas estaban encargadas de la gestión de todo el conjunto. Además con el concepto de ordenador personal no sólo se hace referencia a que los recursos vayan a ser utilizados de forma individual, sino que más bien la gestión de todo el sistema es realizada por el usuario final. Es decir no se precisa un personal específico para la gestión del sistema al margen del usuario final, sino que una única persona es la encargada de gestionar el sistema y además de realizar la explotación del mismo. Es una característica adicional el hecho de que esta única persona puede ser cualquier individuo sin conocimientos informáticos previos. Es decir, un ordenador personal es aquél que puede ser completamente gestionado por un único individuo sin cualificar en temas informáticos.

En un congreso de naturaleza histórica como éste, se habla de la evolución de determinados aspectos del conocimiento a través del tiempo. En este caso, el tema que nos ocupa es el de la historia del papel y sus filigranas, cómo son y cómo han sido a lo largo de los siglos. Otras actividades históricas conllevan un estudio que se extiende a lo largo de miles de años. Esto contrasta fuertemente con la historia de la evolución informática.

El principio de toda historia, se sitúa en el punto en el que se comienzan a tener documentos sobre un tema particular. Estos momentos, se sitúan por regla general hace cientos o miles de años. Sin embargo, podemos suponer que la historia de los equipos informáticos comienza con la segunda guerra mundial. Es decir, hace tan sólo unas décadas. Éste es el momento en el que podemos encontrar las primeras máquinas de funcionamiento electromecánico, pero el nacimiento de la Informática como ciencia es posterior.

El hecho de que tengamos un lapso de tiempo tan pequeño no quiere decir que la evolución haya sido escasa. Más bien todo lo contrario.

En este período de tiempo, hemos pasado de disponer de máquinas de funcionamiento electromecánico a los modernos microprocesadores de funcionamiento completamente electrónico y con tecnología VLSI (Very Large Scale Integration).

Quizá el aspecto más sobresaliente sea el hecho de que durante varias décadas la velocidad de proceso se ha ido multiplicando por un ¡¡ factor mil !! cada diez años. Los tiempos medios para realizar una operación elemental del procesador han pasado de milisegundos en los primeros años

de la década de los cincuenta a microsegundos en los sesenta-setenta, nanosegundos en los ochenta e incluso a décimas y centésimas de nanosegundo con los procesadores de última generación. Y esta evolución que ha multiplicado aproximadamente por un factor  $10^7=10.000.000$  la velocidad de proceso, se ha producido en apenas medio siglo.

Con toda seguridad, ningún otro aspecto de la actividad humana, haya sufrido en tan poco tiempo un cambio de siete órdenes de magnitud, como ha sucedido con la velocidad de procesamiento. Esta evolución, puesta de manifiesto de forma bien palpable con la velocidad de procesamiento, se hace extensiva a casi todas las áreas relacionadas: peso, dimensiones, consumo, facilidad de uso, flexibilidad, adaptabilidad a distintas tareas, costos de adquisición y mantenimiento, etc...

Esta evolución ha puesto al alcance de cualquiera la posibilidad de disponer de un equipo para la realización de las tareas más variadas. Se ha pasado del ordenador corporativo o departamental al ordenador personal. Cualquier investigador o profesional puede encontrar actualmente una gran ayuda en su trabajo con un desembolso relativamente bajo y realizar algunas cosas impensables hace tan sólo unos años.

Ése puede ser el caso de la informatización de un corpus de filigranas que vamos a analizar en el presente trabajo. Actualmente, la facilidad de manejo, los reducidos costes de los dispositivos de almacenamiento de información y la velocidad que podemos encontrar en los ordenadores personales, los convierte en herramientas valiosísimas y muy adecuadas para el almacenamiento de un corpus de filigranas permitiendo un uso individualizado aunque no por esto queda cerrado a la comunicación con otros entornos.

## ACERCAMIENTO AL PROBLEMA

El problema de la informatización de un corpus de filigranas, puede asimilarse con el concepto general de bases de datos. En esencia, cuando tratamos de informatizar un corpus de filigranas, disponemos de una gran cantidad de información y deseamos dotarla de unos medios y métodos que permitan un acceso rápido y cómodo a la misma.

Una base de datos es una colección de datos interrelacionados y estructurados de forma que puedan ser manejados por un sistema de gestión de bases de datos. Estos dos conceptos a menudo se entremezclan y llegan a emplearse indistintamente debiendo en ese caso atender al contexto para distinguir a cuál de los dos se hace referencia.

Un sistema de gestión de bases de datos (S.G.B.D.) es un concepto más amplio y complejo y consiste de una colección de datos interrelacionados y un conjunto de programas para acceder a esos datos. El objetivo primordial de un S.G.B.D. es proporcionar un entorno que sea a la vez conveniente y eficiente para ser utilizado al extraer y almacenar información de la

base de datos. Los sistemas de bases de datos están diseñados para gestionar grandes bloques de información. La gestión de datos implica tanto la definición de estructuras para el almacenamiento de información como la provisión de mecanismos para la gestión eficiente de la misma.

Otras cuestiones auxiliares al estricto tratamiento de la información, pero no por ello menos importantes deben también ser tenidas en cuenta a la hora del diseño del sistema global. Estas cuestiones auxiliares engloban aspectos tales como el control de los accesos a la información para evitar consultas o modificaciones no autorizadas, posibles fallos en la alimentación, averías de los dispositivos físicos, el control de accesos simultáneos por distintos usuarios, etc...

Estos tipos de cuestiones son generales a la mayoría de los sistemas de bases de datos y no presentan una problemática especial dentro del marco de la informatización de un corpus de filigranas, por lo que no serán considerando en esta presentación.

Según lo comentado, el problema de la informatización de un corpus de filigranas se enmarca dentro del problema general de bases de datos ya que se trata de diseñar una estructura de datos para almacenar la información referente a las filigranas y los programas necesarios para una gestión eficiente de la misma.

La información de que se dispone puede ser de naturaleza variada. Pueden ser datos numéricos como dimensiones o fechas tanto del documento como de la filigrana en sí, gráfica, como la representación del dibujo, etc... Sin embargo, desde el punto de vista informático, toda ella puede ser tratada de una forma global. Basta recordar, que la informática se puede definir en pocas palabras como la ciencia que estudia el tratamiento sistemático de la información principalmente por máquinas de funcionamiento electrónico. Desde el punto de vista informático, toda información no es más que una secuencia de caracteres de un alfabeto dado. La información tiene por tanto un carácter formal y sintáctico y está desprovista de cualquier tipo de contenido semántico. Este último aspecto establece la diferencia entre información y conocimiento. El conocimiento es el contenido semántico, la interpretación y el significado de la información. Cuando queremos expresar el conocimiento, por ejemplo en forma de libro, ha de usarse una representación adecuada, normalmente en forma de cadena de caracteres. Hay una diferencia esencial entre las ideas y conceptos que queremos guardar o transmitir y las cadenas de caracteres que empleamos en la práctica. La información está constituida por las cadenas de caracteres, no por las ideas y conceptos a los que representa. La información se puede expresar en una gran variedad de formas, bien pueden ser unos símbolos dibujados sobre un papel, una fotografía de determinado objeto o situación, unos agujeros en una tarjeta, la orientación de la polarización de las partículas de una superficie magnética, etc...

Si consideramos el ejemplo de expresar determinadas ideas o conceptos en forma de libro, elegiríamos como se comentó, una representación

adecuada. Para esta representación emplearíamos un idioma con su correspondiente alfabeto de caracteres y sus reglas sintácticas. Una vez hecho esto tendríamos una secuencia de caracteres más o menos larga que constituye la información. Esa información podría ser tratada de forma automática (fundamentalmente por máquinas de funcionamiento electrónico) de muchas maneras distintas. Podría ser codificada o cifrada, corregida ortográficamente, etc... Sin embargo ningún sistema automático podría extraer ningún concepto ni pensamiento abstracto de la misma. Es decir, las máquinas únicamente son capaces de procesar información pero no conocimiento.

Ahora bien, cualquier tipo de conocimiento es susceptible de ser expresado de forma más o menos precisa mediante algún tipo de representación adecuada. Hay algunos tipos de información que son fácilmente convertibles a cadena de caracteres, como un nombre de autor, el título de un documento, un lugar o una determinada fecha. Sin embargo para otros tipos de información como un dibujo, o una fotografía esta conversión no resulta tan inmediata. Sin embargo también puede realizarse una transformación de este tipo. Para esta transformación se emplea lo que se conoce como formatos gráficos. Al igual que hay numerosos idiomas distintos para representar una misma información, también hay numerosos formatos gráficos distintos con distintas posibilidades. Mediante el algoritmo adecuado y estos formatos gráficos podemos obtener una secuencia de caracteres a partir de una imagen. Una característica de especial relevancia cuando se trabaja con imágenes gráficas es el espacio que ocupan, es decir, las secuencias de caracteres obtenidas son por lo general excesivamente largas. Algunos de estos formatos aportan técnicas de compresión que aprovechando las redundancias que presenta la imagen, permiten reducir el tamaño de las secuencias de caracteres y por lo tanto el tamaño del fichero asociado.

Dentro de la informatización de un corpus de filigranas pretendemos almacenar una serie de datos que repiten su estructura (que no su contenido) según un determinado patrón. Este patrón viene determinado por el objeto filigrana. En un corpus de filigranas disponemos de una gran colección de éstas que puede llegar a varias decenas de miles. Para todas ellas podemos establecer una misma serie de atributos, tales como tipo de documento en que aparece, título, dimensiones, lugar, características del papel sobre el que aparece, el dibujo de la filigrana en sí, etc...

Los distintos valores que toman cada uno de estos atributos va a ser lo que nos permita distinguir una filigrana de otra. Realmente, la filigrana queda caracterizada por el propio dibujo que recoge, pero la verdadera utilidad del estudio de las filigranas es el poder relacionarlas tanto con el entorno en que fueron realizadas como aquél en el que se encuentran. Por esto, además del dibujo deben recogerse datos tales como naturaleza del documento, tipo de papel, procedencia, fabricante del papel, objetos representados en el dibujo, etc...

En un documento del I.P.H. (International Association of Paper Historians) (1) se establece un estándar para el registro de filigranas realmen-

te completo. El objetivo de este documento es el de fijar un formato común para facilitar el intercambio de información entre distintos investigadores. En este documento, se establecen hasta 82 atributos distintos, repartidos en distintos grupos como datos referentes a la hoja, al fabricante, datos bibliográficos del documento, etc... Aparte de esta colección de 82 atributos se establece una forma para clasificar las filigranas en base a los objetos o elementos que en ella aparecen. Todos estos datos permiten catalogar la filigrana con gran precisión. Este conjunto de atributos definen el patrón al que se hacía referencia anteriormente y que se repite para cada filigrana. Todos estos atributos pueden ser representados como una cadena de caracteres y otros como valores numéricos bien sean enteros o reales.

## **ASPECTOS PARTICULARES DE LA INFORMATIZACIÓN DE UN CORPUS DE FILIGRANAS**

Como se ha comentado, el estándar propuesto por el I.P.H. establece un gran número de atributos que pueden codificarse, o bien como cadenas de caracteres de algún alfabeto convencional, o bien como valores numéricos. Sin embargo, aunque no lo contempla explícitamente, sería de gran utilidad el incluir en nuestra base de datos la imagen de la filigrana. A este respecto el estándar propuesto se limita a incluir una descripción estructurada del contenido y apariencia de los elementos de la filigrana.

De esta forma obtendríamos algunas ventajas adicionales, que escapan a la mera descripción de la filigrana, por muy detallada que ésta sea. Se puede aplicar en este caso la célebre frase que afirma que más vale una imagen que mil palabras. De esta forma podríamos eliminar posibles interpretaciones subjetivas que podrían introducirse al realizar la descripción, a la vez que se pueden apreciar las dimensiones y proporciones relativas de los elementos que aparecen en el dibujo. Tendríamos también la posibilidad de obtener una copia impresa a tamaño real para realizar comparaciones con un determinado documento. Esto podría realizarse en el centro en el que se encuentra el documento, ya que actualmente una combinación de ordenador portátil e impresora puede transportarse cómodamente en un maletín de reducidas dimensiones. Otro aspecto interesante del hecho de incluir la imagen es el de emplear un lenguaje universal, como es el de la imagen visual al no particularizarse la descripción para un idioma específico.

Como puede comprobarse, estas posibilidades son realmente apetecibles pero conllevan un esfuerzo extra, tanto para los programas de gestión como para los dispositivos de almacenamiento por el elevado espacio físico que puede ocupar una imagen gráfica.

Se ha afirmado que vale más una imagen que mil palabras, pero desde el punto de vista de los dispositivos de almacenamiento, posiblemente también sea cierto que en la mayoría de los casos esa imagen ocupe más espacio que el requerido por la descripción en el sentido propuesto por el mencionado documento del I.P.H. Como era de esperar, nada es gratis, y la

capacidad de los sistemas de almacenamiento deberá ser bastante mayor, sobre todo si por compatibilidad añadimos también la descripción estándar, con lo que el costo asociado a los dispositivos de almacenamiento será mayor.

No obstante, como se comentará con detalle más adelante, gracias a los medios actuales tanto en dispositivos de almacenamiento como en técnicas de compresión, no resulta un problema especialmente grave. Más importante, sin embargo, son los requerimientos adicionales que tendremos que exigir a los programas de gestión de la base de datos.

Actualmente, los sistemas personales disponen de entornos de trabajo amigables, de fácil manejo y con enormes posibilidades, y para cualquiera de estos entornos podemos encontrar potentes sistemas de gestión de bases de datos. Además, debido a que estos entornos han evolucionado incorporando posibilidades gráficas, los sistemas de gestión de bases de datos disponibles, permiten en su mayoría añadir algún campo de tipo imagen que es de gran ayuda en la mayoría de las aplicaciones. Sin embargo, cuando el objeto de interés es la propia imagen, todos estos sistemas resultan un tanto limitados. Éste es un aspecto especialmente importante ya que las imágenes de las filigranas presentan una características muy concretas, y comunes a todas ellas, que pueden permitir el uso de técnicas de compresión adecuadas e incluso técnicas de almacenamiento de la imagen diseñadas a medida.

Otro problema que presenta el empleo de un S.G.B.D. de propósito general, es el de realizar la captura de la imagen. Aún suponiendo que la imagen ya ha sido recogida mediante alguno de los procedimientos convencionales como el calco o la radiación beta, no podemos incluirla directamente en nuestra base de datos. Necesitaríamos un escáner de sobremesa, de blanco y negro o color, y el correspondiente "software" de control que permita realizar la captura de la imagen proporcionándonos ésta en un fichero con un formato adecuado y compatible con nuestro sistema de gestión de datos.

Sería también interesante el disponer de algún tipo de programa de retoque fotográfico con un algoritmo de binarización adecuado, si el escáner no dispone ya de uno. También sería interesante el disponer de alguna herramienta de tratamiento morfológico de imágenes para generar el esqueleto de la filigrana reduciendo la cantidad de puntos negros sin distorsionar la imagen y obteniendo, por tanto, una imagen más fácilmente comparable y con un tamaño menor. Este aspecto es importante ya que la imagen correspondiente a cualquier filigrana puede ser expresada empleando únicamente 2 colores (blanco y negro) sin pérdida en la calidad de imagen. Al hacer esto, conseguimos reducir el espacio que ocupará la imagen de forma que ocupe un bit por píxel. Esto no quita el que se puedan emplear técnicas de compresión adicionales que reduzcan aun más el espacio necesario.

Todo este proceso de inclusión de la imagen en la base de datos conlleva, por tanto, el uso de uno o varios programas ajenos al propio de gestión de bases de datos, lo que hace que el conjunto total resulte pesado de manejar, sin contar con el coste adicional que supone la adquisición de

estos programas auxiliares y el espacio extra que van a ocupar en el dispositivo de almacenamiento y que podría ser mejor empleado para almacenar la información relativa a numerosas filigranas.

Un tema de especial importancia a la hora de implementar una base de datos es el tratamiento de los índices. Un índice de un archivo de base de datos, funciona de manera similar a un catálogo en una biblioteca. Si estamos buscando un libro por un autor determinado, buscamos en el catálogo de autores y tras encontrar la ficha correspondiente, ésta nos dice dónde encontrar el libro. Para facilitar la búsqueda en el catálogo, las fichas correspondientes a los distintos libros se almacenan en orden alfabético, de forma que no tenemos que recorrer y comprobar cada una de las tarjetas hasta encontrar la que queramos. También podemos contar con más de un índice al igual que en una biblioteca real hay un catálogo de autores, otro de títulos, otro de materias, etc... En las situaciones reales, es posible que los índices de este tipo sean excesivamente grandes y resulten poco eficientes. Sin embargo pueden utilizarse técnicas de indexación más sofisticadas como las tablas de asociatividad o tablas "hash".

Podemos establecer las funciones de indexación y asociatividad de forma que se adecuen al máximo a las necesidades particulares de la base de datos que estemos diseñando. Hay múltiples técnicas y cada una de ellas puede evaluarse en función de un conjunto de parámetros. Los parámetros principales para evaluar una técnica de indexación serán:

**Tiempo de acceso:** El tiempo que se tarda en encontrar un determinado elemento.

**Tiempo de inserción:** El tiempo que se tarda en añadir un dato nuevo. Esto incluye tanto el tiempo de encontrar el lugar correcto y realizar la inserción física del dato, así como el tiempo necesario para modificar la estructura de indexación.

**Tiempo de eliminación:** El tiempo que se tarda en eliminar un dato concreto. Esto incluye tanto el tiempo de encontrar el elemento en cuestión como la modificación del índice o índices necesarios.

**Espacio extra:** El espacio extra que ocupa la estructura de indexación.

Dada la importancia del tema, todos los S.G.B.D. disponibles en el mercado contemplan la generación automática y potente de índices, pero el conocimiento específico del tipo de información que debemos manejar nos permite una elección de la estructura de índices más adecuada.

En el caso concreto que nos ocupa, pretendemos obtener un sistema de bases de datos, que sirva de apoyo a la investigación sobre la historia del papel y sus filigranas. Por este motivo, el parámetro de optimización más importante es el de permitir un acceso rápido a los datos para efectuar

consultas, aunque esto introduzca un empeoramiento de los tiempos de inserción o eliminación, ya que estas tareas resultarán menos frecuentes. Téngase en cuenta que una vez introducida la información raramente será eliminada o modificada, al contrario de lo que puede suceder por ejemplo en una base de datos de facturas emitidas y recibidas por una determinada empresa o la gestión del inventario de un almacén. En ambos casos, la inserción, eliminación y modificación de registros va a ser una estructura de índices de propósito general, como los que incluyen los sistemas convencionales, pueden ser superados por estructuras más orientadas al problema concreto que se pretende resolver.

Una característica importante a la hora de trabajar con un corpus de filigranas, es el hecho de que la mayoría de las consultas se hacen en base a los elementos que la filigrana representa. Es decir, van a ser más frecuentes las consultas que busquen filigranas que contengan por ejemplo una cruz encima de un jinete y que incluyan determinadas letras, que consultas que busquen un tipo de documento o un lugar de origen del documento. Será pues en este tipo de consultas en el que habrá que poner mayor énfasis a la hora de diseñar la estructura de indexación.

En este aspecto, podemos aprovechar favorablemente la clasificación que establece el I.P.H. referente a las clases y subclases en que pueden catalogarse los elementos que aparecen en la filigrana.

En el documento de estandarización propuesto por el I.P.H. (1) se establecen 25 clases principales. Cada una de éstas se subdivide a su vez en otro conjunto de subclases o categorías secundarias y cada una de éstas en otras más y así sucesivamente hasta un total de cuatro niveles de clasificación. Esto nos proporciona directamente una estructura de tipo arborescente que resulta especialmente interesante en temas de búsqueda, ya que da lugar a algoritmos tremendamente rápidos, optimizando aun más las consultas a la base de datos.

## **Herramientas disponibles**

El tipo de herramientas a evaluar a la hora de realizar una base de datos como la mencionada de las filigranas, lo podemos dividir en tres clases principales: dispositivos físicos, recursos y herramientas "software" y técnicas de optimización. A continuación se hace un breve comentario de cada uno de estos tipos.

Quizá el tipo de herramienta que menos dudas genere sea el de los dispositivos físicos. En parte porque ya desde el principio hemos establecido como sistema base un ordenador personal, lo que restringe bastante el conjunto de sistemas a contemplar.

Para centrar ideas, tomamos como referencia la combinación "hardware"- "software" más extendida en los sistemas personales como es el en-

torno MS-Windows v.3.x de Microsoft sobre una arquitectura x86 de Intel. En otros sistemas distintos, por cuestiones de libre competencia, podemos encontrar posibilidades similares dentro del mismo margen de precios.

No obstante dado el tipo de aplicación a que se va a destinar, debe ser un ordenador personal de gama alta que proporcione un nivel de recursos satisfactorio. En este sentido, deberíamos pensar en un procesador i486 DX2-66 o superior con al menos 8 Mb de memoria RAM, aunque sería mejor trabajar con 16Mb. En cuanto al sistema de presentación gráfica (combinación tarjeta gráfica-monitor) no se requieren especificaciones especiales y basta con una tarjeta SVGA corriente con una resolución de 800x600 puntos y su monitor correspondiente. El coste de un sistema de este estilo y sin incluir los sistemas de almacenamiento podría variar entre las 150.000 y 200.000 pesetas, aproximadamente.

Si deseamos introducir las imágenes de las filigranas, algo que ya se comentó sería de gran interés, precisaríamos un escáner corriente, aunque sería necesario que permitiese la digitalización de una página completa para poder introducir de una sola pasada y sin uniones, filigranas de grandes dimensiones a tamaño real. En cuanto a resolución, dado el tipo de imagen a digitalizar es suficiente con los 300 ppp (puntos por pulgada) que ofrecen como mínimo casi todos los equipos de sobremesa. El precio de un dispositivo de estas características puede oscilar entre las 70.000 y las 150.000 pesetas.

Por último, el aspecto más importante en todo sistema de bases de datos, es el tema de los dispositivos de almacenamiento. En este apartado podemos distinguir tres tipos de almacenamiento. En primer lugar, el almacenamiento en disco duro, es el más tradicional y también el más caro pero por contra ofrece las ventajas de ser el más extendido (todos los sistemas personales disponen de uno de mayor o menor capacidad), el sistema de acceso más rápido y el que permite una mayor facilidad de actualización. El costo del megabyte de almacenamiento puede variar entre las 45 pts. para los discos IDE de gran capacidad (1 Gb) a las 100 pts/Mb para los discos con interface SCSI. En este último caso hay que añadir también el precio de una controladora SCSI que puede ser de unas 30.000 pesetas.

En el extremo opuesto en cuanto a costo tenemos el almacenamiento en CD-ROM. El coste de grabar un CD-ROM de 660 Mb es de unas 5000 pesetas, lo que conlleva un costo de unas 10 pts/Mb, a lo que hay que añadir el precio de la unidad lectora que puede encontrarse en el mercado a partir de las 20.000 pesetas. El principal inconveniente de este sistema es que una vez introducidos los datos, no se pueden modificar. La forma de trabajar con este sistema sería el de generar la base de datos sobre alguno de los otros sistemas y una vez que la información fuese correcta, grabarla en CD-ROM. Esta labor debería encargarse a una empresa especializada, ya que el precio de una grabadora de CD-ROM (>300.000 pts.) es todavía excesivamente alto.

Otros problemas asociados con este sistema de almacenamiento son la velocidad, menor que la de un disco duro moderno y la limitación a 660

Mb de capacidad aunque esto no es fácilmente alcanzable en un principio y además existen unidades que permiten trabajar con varios discos simultáneamente. Como aspectos más atractivos están su bajo coste, su facilidad de intercambio y transporte y su enorme fiabilidad y resistencia de uso.

Por último nos queda un último tipo de sistemas que son los discos removibles. En este apartado cabe señalar dos tecnologías distintas: los discos duros removibles tipo Syquest de tecnología puramente magnética y los sistemas magneto-ópticos. Los primeros permiten un máximo de 270 Mb en un disco de unas 9500 pesetas (35 pts/Mb) y los segundos permiten hasta 230 Mb en un disco de unas 5000 (22 pts/Mb) o hasta 1.3Gb en un disco de unas 12.000 pesetas (9.2 pts/Mb). Los precios de las unidades de lectura son aproximadamente de unas 70.000 pesetas para el caso de los discos Syquest, 100.000 pesetas en el caso de los magneto-ópticos y de unas 300.000 pesetas para los de 1.3Gb. En la mayoría de los casos habrá que añadir la correspondiente controladora SCSI si fuese necesario.

De todos estos comentarios se deduce que el sistema más potente y que mejores prestaciones ofrece es el disco duro tradicional. Hay que tener en cuenta que aunque el coste por Mb de información almacenado es el más elevado de todos, en los otros hay que añadir el precio de la unidad lectora que sólo quedará compensado cuando el número de discos a manejar y, por lo tanto, la capacidad total sea elevado.

La principal ventaja de los discos removibles es también su principal inconveniente a la hora de trabajar con una única base de datos, ya que distintas partes de la misma deberían almacenarse en discos separados. Esto acarrearía incomodidades en el manejo y un tratamiento especial por parte de los programas de gestión. No obstante, estos sistemas son muy válidos para el intercambio de elevadas cantidades de información entre distintos sistemas que no estén físicamente conectados entre sí. El sistema de CD-ROM sería especialmente interesante para la publicación de una recopilación de filigranas, ya que podría ser distribuido por bajo precio, no puede ser modificado y el lector necesario es muy económico y muy pronto todos los sistemas personales incluirán uno en su configuración más básica.

Un aspecto más delicado es el de los recursos "software" de que disponemos para la implementación de la base de datos de filigranas. En un principio tenemos dos posibilidades distintas de atacar el problema. Por un lado podemos emplear un S.G.B.D. comercial que nos va a permitir una creación y manejo sencillos de la base de datos. Por otra parte podemos emplear un lenguaje de programación tradicional como C, Pascal, Basic, etc...

Cada sistema tiene sus ventajas y sus inconvenientes. Dentro del primer grupo podemos destacar productos como Microsoft Access, Lotus Approach, dBase, FileMaker, FoxPro, o Borland Paradox por señalar los más representativos del entorno MS-Windows 3.x. Para una descripción detallada y comparativa de estos y otros productos puede consultarse (2).

La mayor parte de estos sistemas permite la inclusión de imágenes dentro de uno de sus campos pero el soporte que dan al respecto es bastante limitado en cuanto a formatos entendibles, optimización de la compresión, etc... No disponemos de posibilidades de adquirir la imagen directamente por carecer de un control de dispositivos de digitalización.

La principal ventaja de este tipo de sistemas está en su facilidad de uso y su elevada especialización en el tema de bases de datos, lo que los hace difícilmente superables. Estas características permiten realizar el diseño de la base de datos en un período que puede variar entre varias horas y unos pocos días, dependiendo de la experiencia que se tenga con el paquete y la formación previa, sin ser necesarios en ningún caso unos elevados conocimientos informáticos previos.

Por contra, un desarrollo realizado, a partir de un lenguaje de programación tradicional, requiere unos conocimientos y una experiencia en programación considerables y puede llevar un tiempo de desarrollo que oscila entre varias semanas a unos pocos meses. Las ventajas de esta segunda opción están en la mayor flexibilidad y potencia que aportan para realizar cualquier tipo de tarea. Esto permite crear las rutinas necesarias para controlar el escáner, emplear una técnica de compresión especialmente optimizada para las imágenes de las filigranas y una total libertad a la hora de diseñar los ficheros y la estructura de índices.

El resultado es un programa compacto que realiza todas las operaciones necesarias de una forma fácil, cómoda y rápida dentro de un mismo entorno. Además, una vez compilado el código, se obtiene un programa que ocupará muy poco espacio en disco en comparación con los programas anteriormente mencionados.

Entre estos dos sistemas se abre un abismo de posibilidades y requerimientos, pero por fortuna podemos encontrar un equilibrio en el punto medio. Este punto medio lo proporcionan los lenguajes de programación con un motor de bases de datos embebido, o una librería auxiliar con las funciones necesarias a la hora de trabajar con bases de datos. Estos productos ofrecen toda la potencia y flexibilidad de un lenguaje de programación, pero eliminan el trabajo necesario para crear las funciones y procedimientos que trabajen directamente con los datos. Incorporan funciones para introducir datos, presentarlos en forma de tablas o, con ayuda de formularios, realizar búsquedas, etc... Productos de este tipo pueden ser VisualBasic, Borland Database Tools for C++, SuperBase o el recientemente aparecido entorno Delphi.

Este tipo de productos los convierte en ideales para la realización completa de todo el proceso de informatización de un corpus de filigranas.

En cuanto a las técnicas de optimización podemos aprovechar las características específicas de las filigranas para conseguir un sistema más eficiente. Por una parte, el ya comentado tema de la clasificación del contenido de la filigrana que permite el diseño de una estructura de índices arbores-

cente que proporciona una base ideal para la implementación de un algoritmo de búsqueda rápido.

Por otra parte está el tema del tratamiento de la imagen para su almacenamiento. En la figura 1 se muestra el espacio necesario para almacenar una filigrana corriente de un tamaño de unos 10x10 cm digitalizada, con una resolución de 100 ppp lo que genera una imagen de 382x382 puntos = 145.924 puntos empleando distintos formatos. Obsérvese el elevado número de puntos a que da lugar una filigrana de reducidas dimensiones digitalizada en baja resolución (es corriente el empleo de resoluciones de 300 ppp e incluso mucho mayores para trabajos de calidad). Hay otros formatos distintos, pero éstos son algunos de los más habituales en un sistema personal.

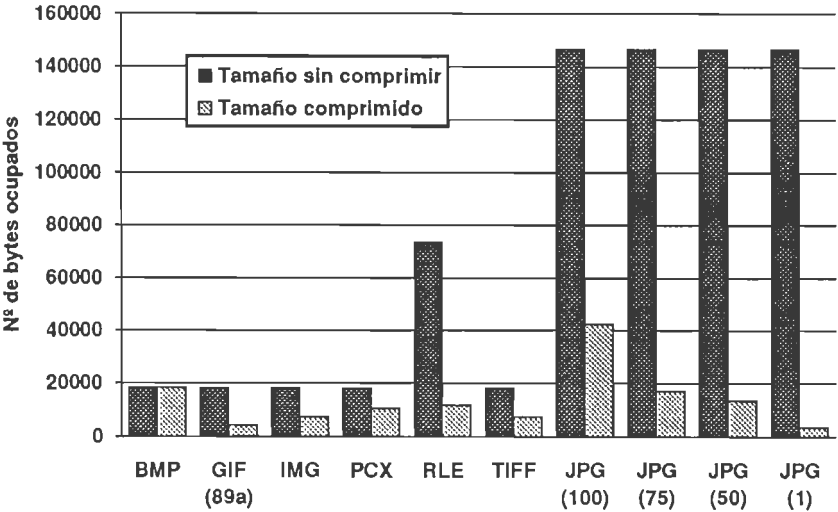
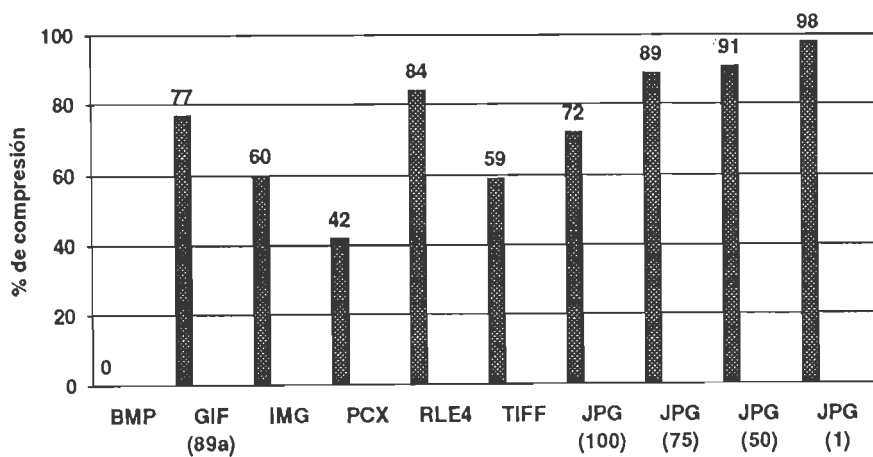


Fig. 1.—Espacios requeridos por los distintos formatos para codificar una imagen antes y después de realizar la compresión.

Para cada formato se muestran dos barras: la primera indica el número de bytes que contempla el formato y la de la derecha el número de bytes a que queda reducido después de realizar la compresión. La relación que haya entre ambas barras da una medida de la relación de compresión que proporciona cada formato. Este último dato se recoge en la figura 2. En esta última se muestran los porcentajes de compresión obtenidos con cada uno de los formatos. Se ha tomado como referencia el formato BMP, que es el formato nativo del entorno Windows y en su concepción básica es un formato no comprimido. Su versión comprimida corresponde al formato RLE. Los seis

primeros formatos mantienen íntegra la imagen. Sin embargo el formato JPEG permite ajustar el grado de conservación de la imagen variando un parámetro del algoritmo de compresión, que recibe el nombre de factor de calidad. El valor 100 indica que no hay pérdida de información y por lo tanto la compresión es menor aunque la imagen se mantiene íntegra. Por el contrario, el valor 1 indica que se realiza una compresión máxima, con lo que se produce una gran pérdida de información y la imagen queda parcialmente desvirtuada.



**Fig. 2.**—Porcentajes de compresión obtenidos con algunos de los formatos gráficos más frecuentes

Obsérvese que los distintos formatos parten de tamaños descomprimidos distintos. Esto es debido a la forma que tienen de codificar cada punto de información. El formato JPEG emplea un byte (8 bits) por píxel, lo que hace que para una imagen de 382x382 puntos se empleen 145.924 bytes. El formato RLE-4 emplea 4 bits por píxel, con lo que el número de bytes necesarios será justo la mitad. El resto de formatos emplean un único bit por píxel, lo que produce un tamaño de  $(382 \times 382) / 8 = 18241$  bytes. Esto es posible gracias a que al ser la filigrana una imagen negra sobre fondo blanco, sin tonalidades de gris ni colores, puede binarizarse de forma que basta con distinguir entre 2 colores: blanco y negro. Como éstos son mutuamente exclusivos se puede emplear un único bit para almacenar el color correspondiente a cada punto.

En la segunda figura, se muestra que las mayores relaciones de compresión se obtienen con el formato JPEG, pero hay que tener en cuenta

que toma como punto de partida una imagen que tiene mayor redundancia al emplear 1 byte por punto de la imagen. Según esta gráfica, el formato más favorable es el GIF, pero una evaluación correcta que permitiese determinar el formato más adecuado para las imágenes de las filigranas requeriría un estudio específico mucho más extenso.

Un aspecto que conviene destacar a la hora de hablar de los formatos gráficos, es que todos incluyen al comienzo una pequeña cabecera que indica cómo deben interpretarse los datos que vienen a continuación. Puesto que las filigranas presentan gran cantidad de características comunes, esta cabecera, que estaría incluida con cada imagen, podría reducirse a una expresión mínima. Hay que tener en cuenta que si contemplamos varios miles de filigranas, toda información sobrante que sea eliminada de la cabecera de las imágenes, liberará gran cantidad de espacio, ya que afecta a todos y cada uno de los registros.

## NUEVAS POSIBILIDADES

Quizá el aspecto que más ha evolucionado en los últimos años dentro del mundo de la informática sea el de las comunicaciones y las redes de ordenadores. Actualmente, con un "software" sencillo y con una tarjeta de red, podemos conectarnos a una red departamental o corporativa y compartir recursos como discos o impresoras. Con ayuda de un módem podemos emplear la red telefónica para intercambiar datos con un ordenador remoto en cualquier punto del planeta. Sin embargo, existe una red concreta que está creciendo vertiginosamente y que puede considerarse como La Red. Esta red es Internet. De su progresivo y espectacular crecimiento se han hecho eco recientemente no ya las revistas especializadas sino también los semanarios de información general y los periódicos diarios.

El interés que despierta esta red está basado en varios factores. Por una parte, es la más universal de las redes (se habla de unos tres millones de sistemas de todo tipo conectados y unos treinta millones de usuarios en todo el mundo). Por otra parte, es de acceso gratuito para la mayor parte de los usuarios finales y permite el acceso libre a una gran cantidad de información que puede cifrarse en miles de gigabytes.

En la actualidad, la práctica totalidad de las universidades y centros de investigación de todo el mundo ofrecen a sus miembros acceso directo a esta red. También numerosas empresas están conectadas directamente a Internet. El origen de Internet viene de una serie de redes de ordenadores desarrollados en la década de los años 70. Comenzó con una red denominada Arpanet, desarrollada por el Departamento de Defensa de los Estados Unidos para conectar entre sí los sistemas informáticos de los distintos centros.

¿Qué relación puede tener esta red con el tema de las filigranas? Pues exactamente la misma que puede tener con cualquier otro elemento del conocimiento humano. A lo largo de toda la Internet hay miles de servidores

que proporcionan acceso a numerosas bases de datos con la información más variada. Hay servidores que proporcionan las fotografías del satélite Meteosat actualizadas cada hora, fragmentos de películas de Walt Disney, los últimos avances realizados en la lucha contra el cáncer, imágenes de algunas obras de arte del museo del Louvre, la Enciclopedia Británica, y en general información sobre las disciplinas científicas más variadas, dado que como se comentó mantiene conectados la práctica totalidad de centros de investigación del mundo entero. Pues bien, toda esta información está accesible a través de Internet, desde la mesa de trabajo de cualquier investigador que tenga un sistema conectado a Internet. No importa en qué centro de investigación ni en qué punto del globo está el servidor.

El hecho de establecer un enlace a Internet, permitiría que un investigador interesado en una determinada filigrana, pudiera buscarla a través de distintos servidores públicos situados en bases de datos de distintos centros situados en países lejanos, sin salir del despacho y en tan sólo un par de horas. Realmente las posibilidades de intercambio de conocimiento que ofrece Internet son enormes y escapan a cualquier intento de descripción, pero en cualquier caso, las posibilidades que ofrece son realmente enormes.

## REFERENCIAS Y BIBLIOGRAFÍA

- (1) "International Standard for the Registration of Watermarks". International Association of Paper Historians.
- (2) Rais Busom. "Bases de datos para Windows". PCWorld Nº105. Diciembre 1994.
- (3) Ramez Elmasri, Shamkant B. Navate "Fundamentals of database systems" The Benjamin-Cummings Redwood City, 1989.
- (4) Henry F. Korth, Abraham Silberchatz "Fundamentos de bases de datos" Ed. McGraw-Hill. Madrid 1993.
- (5) Harley Hahn, Rick Stout "Internet. Manual de referencia" Ed. McGraw-Hill Madrid 1994.