

HERRAMIENTAS INFORMÁTICAS PARA LA DOCUMENTACIÓN

Tema 4. Tipos y formatos de documentos

1. Información digital

Se denomina información digital o electrónica a aquella que se almacena en formato binario, en un soporte magnético u óptico y que puede consultarse utilizando un ordenador. Este término es muy amplio e incluye documentos de muchos tipos y formatos, tales como libros electrónicos, bases de datos, páginas web, enciclopedias electrónicas, etc. Pudiendo estar todos ellos almacenados por ejemplo en un CD-ROM o accesibles a través Internet.

En este tema comenzaremos por describir las principales características de todo documento digital y citaremos las ventajas e inconvenientes de este tipo de información con respecto a los documentos en papel. Después describiremos los principales tipos de documentos digitales en cuanto a su forma y al tipo de información que presentan. Por último revisaremos los principales tipos de formatos de ficheros informáticos.

1.1. Características de la información digital

Las principales características de todo documento digital son las siguientes:

- Se almacena en **formato binario**. La información digital se almacena en un soporte magnético u óptico en forma de ceros y unos.
- Es **procesable** por ordenador. Dado el tipo de soporte en que se almacena, esta información puede ser accedida mediante programas de ordenador y puede ser por tanto procesada informáticamente. Este tipo de procesamiento facilita llevar a cabo de forma rápida y segura operaciones repetitivas sobre la información que antes era necesario llevar a cabo manualmente. Además, este tipo de formato ha aumentado enormemente la capacidad que tenemos de acceder, manipular y visualizar la información.

Como resultado de esta capacidad de procesamiento de la información:

- a) Es posible obtener nueva información con valor añadido, por ejemplo mediante la extracción de gráficas o estadísticas de una serie de datos numéricos o mediante la obtención de resúmenes y el contraste de distintas informaciones sobre un mismo tema.
- b) Es posible acceder a la información a distancia. A través de las redes de ordenadores podemos acceder de forma casi instantánea a información situada en cualquier parte del mundo.
- c) Es posible crear documentos que integren todo tipo de información textual, visual y sonora: documentos multimedia.

- La información es **reutilizable**. Es fácil duplicar la información y transferirla de un tipo de soporte a otro manteniendo un alto grado de fiabilidad en la copia. Además, varios usuarios pueden acceder a la misma copia al mismo tiempo.
- Es **interactiva**. Algunos documentos digitales pueden ser interactivos, modificándose en función del comportamiento del lector. Una forma básica de interactividad puede ser por ejemplo, la proporcionada a través de formularios de búsqueda y recuperación de almacenes de información, o bien la utilizada por las enciclopedias infantiles en CD para enseñar a los niños conceptos básicos mediante su interacción con una serie de gráficos y animaciones.
- Es **actualizable**. La información digital puede ser actualizada por sus autores de forma sencilla y económica, pudiéndose generar versiones actualizadas de la misma con mucha rapidez. Además, su facilidad de distribución en forma de discos ópticos, y sobre todo a través de Internet, permite que las nuevas versiones de los documentos lleguen a sus destinatarios de forma muy rápida.
- Es **navegable**. El uso de enlaces en documentos hipertexto permite que el usuario defina la forma en la que quiere acceder a los documentos, tanto en cuanto a la parte del documento a la que quiere acceder, como al orden en que quiere consultarlo.
- Es **recuperable**. Es fácil generar almacenes de documentos digitales y utilizar mecanismos de búsqueda y recuperación de los mismos. Los modernos buscadores permiten buscar de forma rápida y relativamente eficaz documentos que cumplan una gran variedad de condiciones en repositorios que pueden contener miles de ellos.

1.2. Documento digital frente a documento en papel

Para comparar los documentos digitales con los documentos en papel vamos a utilizar unos parámetros similares a los propuestos por diversos autores: ergonomía, actualización, densidad, interactividad, durabilidad, autenticidad y visualización. Veremos las ventajas e inconvenientes de cada uno de los dos tipos de documento en relación a estos parámetros.

Con respecto a la **ergonomía**, una vez disponemos del documento en cualquiera de los dos formatos, la versión en papel es mucho más fácil de manipular (coger, mover, ...). Para acceder a los documentos digitales es necesario utilizar un ordenador, con un aparato de lectura y unos programas adecuados. Obviamente no siempre disponemos de estos elementos, y es mucho más fácil coger un libro, o artículo en papel y leerlo en cualquier parte. En la actualidad se está investigando en el desarrollo de nuevos soportes de lectura de documentos digitales de pequeño tamaño y fáciles de transportar (ver

figura 1). Puede que cuando estos aparatos funcionen de forma adecuada y estén disponibles para el gran público, la ergonomía de los documentos digitales iguale a la de los documentos en papel.

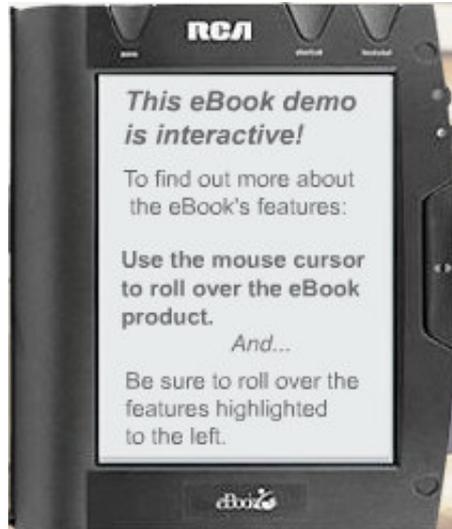


Figura 1. Dispositivo para la lectura de documentos digitales.

La **durabilidad** es una propiedad crítica para garantizar la transmisión de la ciencia y la cultura a largo plazo. Aunque parezca mentira, los modernos sistemas de almacenamiento magnético y óptico son menos durables que los documentos en papel. Se sabe que un documento editado en un buen papel tiene una duración mínima de unos quinientos años, mientras un disco óptico puede durar en condiciones óptimas unos cien. Más aún, se ha comprobado que un uso normal de un disco óptico le garantiza una vida media de entre dos y diez años.

Por otro lado, los documentos digitales están sujetos, no sólo al desgaste físico del soporte, sino a la obsolescencia tecnológica. La evolución de los tipos de soporte y de la tecnología y los programas de acceso a los documentos digitales hace que en pocos años, un soporte que parecía legible a largo plazo deje de ser accesible. Pensemos en lo que ha ocurrido con los disquetes de 5,25 pulgadas y lo que pronto ocurrirá con los de 3,5. En pocos años será difícil encontrar un ordenador que disponga de un lector de disquetes, con lo cual será complicado acceder a toda la información que hayamos guardado en este tipo de soporte. Esta misma duda podía establecerse respecto a la sustitución de los "modernos" CDs, por los "más modernos" DVDs, pero los distintos fabricantes de aparatos lectores, garantizan que los lectores de DVD, también sean capaces de leer CDs.

El tercer parámetro en el que el papel supera al documento digital es la **autenticidad**, referida a su resistencia a la falsificación. Es relativamente complicado falsificar un documento en papel, en cambio, no existen muchos impedimentos para duplicar o modificar un documento digital, en ocasiones, de modo indetectable. Esta circunstancia hace que los documentos digitales no tengan validez jurídica. También en este ámbito la

tecnología y las leyes están evolucionando, y se están estableciendo sistemas de autenticación que permiten identificar de modo inequívoco la originalidad de un documento digital.

En relación a la **visualización** de los documentos, la calidad de la misma es superior en el caso del papel, que en el de los documentos digitales. Las pantallas de ordenador tradicionales suelen funcionar en base a un rayo de luz que actualiza el contenido de la misma varias veces por segundo y proyecta su contenido hacia los ojos del usuario. Este modo de funcionamiento produce un parpadeo, apenas perceptible, pero que fatiga la vista y resta calidad a la visualización. Se sabe que una frecuencia superior a 120 actualizaciones por segundo (120Hz) elimina el parpadeo e iguala la calidad de visualización con el papel en este sentido. En la actualidad la mayoría de las pantallas ofrecen unas frecuencias de refresco entre 60 y 90Hz.

Además de las dificultades relacionadas con el parpadeo de las pantallas, es necesario tener en cuenta la definición con que se visualizan los documentos. Por definición nos referimos a la cantidad de puntos por unidad de superficie con que se presenta el documento en la pantalla. En último término, tanto las imágenes como el texto que vemos en la pantalla está formada por una cantidad dada de puntos de distintos colores. Cuanta más definición tenga el documento en pantalla, mejor se verá. Las actuales pantallas de ordenador todavía tienen una definición bastante inferior que la obtenida a partir de la impresión en papel. Mientras las pantallas habituales ofrecen una definición de entre 75 y 90 puntos por pulgada, el papel puede ofrecer entre 300 y 3000 puntos por pulgada.

Por otro lado, los documentos digitales son superiores al papel en los parámetros de actualización, densidad e interactividad, además de en otras características comentadas en el apartado anterior que les son propias: procesabilidad, navegabilidad y recuperabilidad. En el apartado anterior hemos comentado ya las ventajas de la información digital en cuanto a la actualización y la interactividad. En cuanto a la **densidad**, nos referimos con este concepto a la cantidad de información que podemos almacenar por unidad de volumen. Tal y como comentamos en el tema anterior, los soportes informáticos, tales como los discos ópticos o magnéticos permiten almacenar una enorme cantidad de información en un espacio muy reducido. Por ejemplo, en un disco óptico es posible guardar más de 200.000 páginas de texto. Imaginemos la cantidad de espacio que ocuparían 1.000 libros de 200 páginas cada uno y comparémoslo con lo que ocupa un CD.

2. Documentos digitales

En este apartado vamos a tratar de clasificar los distintos tipos de documentos digitales existentes y a revisar sus principales características

2.1. Tipos de documentos

En primer lugar, podemos clasificar los documentos electrónicos en función del tipo de información que almacenan. En este sentido podemos diferenciar cuatro grandes tipos de documentos: textuales, no textuales, multimedia e hipertexto. Revisemos brevemente en qué consiste cada uno de estos tipos de documentos:

- **Documentos textuales.** Este tipo de documentos contienen solamente texto y suelen ser bastante habituales. Por ejemplo, la mayoría de los libros pueden almacenarse en este formato, pero también numerosos artículos en prensa o en revistas, legislación, etc. Dos de las principales ventajas de este formato son la posibilidad de visualizarlos fácilmente sin problemas de formato al cambiar de ordenador o de programa, y la gran cantidad de información textual que podemos almacenar en un espacio muy reducido, p.e. un disquete o un disco óptico. Además, existen una gran cantidad de programas que nos permiten visualizar y modificar los documentos de este tipo.
- **Documentos no textuales.** Son documentos que contienen otro tipo de información distinta del texto. Por ejemplo, podemos incluir en este tipo de documentos aquellos que almacenan imágenes estáticas, sonidos, películas o programas ejecutables.
- **Documentos multimedia.** Son documentos electrónicos que combinan textos, gráficos, sonidos, imágenes y programas ejecutables. Estos documentos permiten la interacción del usuario, y se componen de muchos elementos de información diferentes cuyo contenido puede variar en algunos casos como resultado de la interacción con el usuario. Como ejemplos destacables de este tipo de documentos, podemos citar las páginas web o las enciclopedias interactivas en CD.
- **Hipertextos.** Son documentos que permiten definir una estructura compleja compuesta de elementos con diferentes tipos de información multimedia unidos mediante enlaces lógicos. Se trata de documentos que pueden ser hojeados rápidamente y cuya visualización no tiene que ser secuencial. La manera de interaccionar con estos documentos consiste en ir recorriendo los elementos de información (textos, imágenes, animaciones, ...) en el orden preferido utilizando para ello los enlaces disponibles. De esta manera se puede ir descartando cómodamente aquella información que se considere irrelevante, al tiempo que se profundiza en los aspectos que se consideren oportunos. El ejemplo más conocido de este tipo de documentos son las páginas web.

En segundo lugar, podemos clasificar los documentos en función de su contenido y de su finalidad. En este sentido, podemos diferenciar entre los documentos secundarios y los primarios.

Los documentos **secundarios** son aquellos que contienen referencias a otros documentos, remitiéndonos a ellos si queremos obtener la información deseada. Este tipo de documentos incluyen por ejemplo, las bibliografías, los catálogos de tiendas o bibliotecas, los directorios de recursos en Internet, etc.

Los documentos **primarios** son aquellos que no son secundarios, es decir, aquellos que nos dan la información de primera mano sin remitirnos a otros para conseguirla. Podemos incluir en este tipo de documentos:

- Las obras de creación, desde una obra literaria hasta una fotografía artística.
- Las obras científico-técnicas, incluyendo tesis doctorales, artículos en revistas científicas o manuales de usuario de un electrodoméstico o de un programa de ordenador.
- Los documentos administrativos o legales, incluyendo normativas y legislación, documentación de gestión, como facturas, nóminas, etc.

En los siguientes apartados vamos a revisar las características básicas de algunos de los principales tipos de documentos, tanto primarios como secundarios, que podemos encontrar en formato digital.

2.2. Libros

El libro es considerado el medio por excelencia para la difusión de la cultura. Su implantación y arraigo es muy grande y van muy unidos al papel, el soporte material más utilizado hasta el momento. No obstante, cada vez más, podemos encontrar los libros en formato digital: son los denominados **libros electrónicos** o *e-books*. Existen diversas formas de distribución de los libros electrónicos:

- En disco óptico. En la actualidad se están distribuyendo en este soporte fundamentalmente obras de referencia, incluyendo diccionarios y enciclopedias, aunque también es posible encontrar por ejemplo libros infantiles interactivos.
- Consulta en línea. En este caso se accede al libro por Internet y suele poder leerse de forma gratuita o previo pago de una cantidad determinada. Cada día aumentan los repositorios en Internet en los cuales es posible acceder a una gran cantidad de libros, incluyendo fundamentalmente literatura clásica en diversos idiomas que no posee restricciones en cuanto a derechos de copia o distribución.

Por ejemplo, la *biblioteca Cervantes Virtual* (www.cervantesvirtual.com), nos permite acceder en formato de texto a una enorme cantidad de libros de autores clásicos en español, incluyendo por ejemplo a Cervantes, Góngora, Garcilaso, etc. Otros repositorios de este mismo tipo como *The on-line books page* (onlinebooks.library.upenn.edu) dan acceso a más de 10.000 libros en inglés de autores como Julio Verne, Charles Dickens, Edgar Allan Poe, entre otros muchos, además de incluir una gran cantidad de enlaces a otros repositorios de libros electrónicos. La figura 2 muestra la página principal de este último repositorio.

The Online Books Page

onlinebooks.library.upenn.edu

EST. 1993 - UPDATED SEPTEMBER 18, 2002 - FREE

BOOKS ONLINE

[Search our 17,000+ Listings](#) -- [New Listings](#) -- [Authors](#) -- [Titles](#) -- [Subjects](#) -- [Serials](#)

NEWS

[Supreme Court to hear public domain appeal](#) -- [New collections add thousands of 19th-century titles](#) -- [Latest Book Listings](#)

FEATURES

[A Celebration of Women Writers](#) -- [Banned Books Online](#) -- [Prize Winners Online](#)

ARCHIVES

[General](#) -- [Foreign Language](#) -- [Specialty](#)

THE INSIDE STORY

[About Us](#) -- [FAQ](#) -- [Get Involved!](#) -- [Books In Progress/Requested](#) -- [More Book Links](#)

Figura 2. Repositorio de libros electrónicos en Internet.

- Descarga a un ordenador personal. En ocasiones, podemos encontrar en Internet, libros almacenados en ficheros que están pensados para ser descargados en un ordenador personal y ser leídos o impresos posteriormente en el mismo. Existen editoriales como *Edicions Virtuals UPC* que ofrecen libros mediante este sistema (ver figura 3). También es posible acceder a *netLibrary* (www.netlibrary.com), donde se ofrecen más de 20.000 libros digitales a particulares y organizaciones, unos 4.000 de los cuales son gratuitos.
- Descarga a un lector de libros digitales. Tal y como hemos indicado en un apartado anterior, existen ya en el mercado dispositivos de lectura de libros digitales de tamaño reducido. Es posible conectar estos dispositivos a un ordenador y copiar los libros a los mismos para poder leerlos posteriormente. Los principales problemas de estos dispositivos siguen siendo la baja calidad

de visualización en relación al papel y la inexistencia de un formato estándar de almacenamiento de libros digitales.

La mayoría de los libros que podemos encontrar en Internet, se encuentran almacenados como texto puro, esto es, no incluyen ningún tipo de formato, tal como letras en negrita o cursiva o con distintos tamaños y fuentes, sangrado y justificación de párrafos, etc. No obstante, existen otros formatos de almacenamiento que permiten incorporar estas características, junto con la posibilidad de incluir imágenes y gráficos. Entre ellos podemos destacar el formato PDF (*Portable Document Format*) o el más reciente formato LIT de Microsoft. En un apartado posterior revisaremos los principales formatos de ficheros que nos permiten almacenar textos electrónicos.

Análisis meteorológico en la mar

Información General
Sinopsis

Índice:

Presentación
Observaciones marítimas
Codificación de las observaciones
Análisis
Análisis de masas de aire y de frentes
El proceso de análisis
Ciclones tropicales

Título:
Análisis meteorológico en la mar
Autores:
Conesa Prieto, Gerard
Año:
1994
Materia:
Náutica
Idioma:
Castellano

pdf
Comprar libro entero

Figura 3. Página de descarga de un libro electrónico en Edicions Virtuals UPC.

Por ahora, los libros digitales son un instrumento muy útil para realizar consultas de forma rápida y sencilla, como las que se pueden realizar a un diccionario. Ahora bien, los libros dirigidos al gran público, con tiradas elevadas, siguen editándose mediante el sistema tradicional y en papel. No obstante, ya existen experiencias en la distribución de algunos superventas, a través de la web. Baste citar el ejemplo del libro de Arturo Pérez Reverte *El oro del rey*, que pudo accederse durante un mes por Internet en formato PDF, antes de editarse a un precio muy superior en su versión tradicional. También cabe citar el libro *Riding the bullet*, de Stephen King, que se distribuyó de forma exclusiva por Internet y que se vendió a más de 500.000 lectores durante las dos primeras semanas, antes de tener problemas con las copias piratas del mismo.

En un futuro no muy lejano, cuando mejore notablemente la calidad de los dispositivos portátiles de lectura o cuando se llegue a un mínimo acuerdo sobre el formato de almacenamiento y distribución, muchos libros se editarán en formato digital y podrán

ser vendidos a través de Internet, con todas las ventajas que ello comporta. En 1999 se fundó el *Open eBook Forum* (www.openebook.org) que reúne a compañías de hardware y software, editores, autores y usuarios de libros electrónicos y cuyo objetivo es establecer las especificaciones y estándares para la publicación digital (ver figura 4).

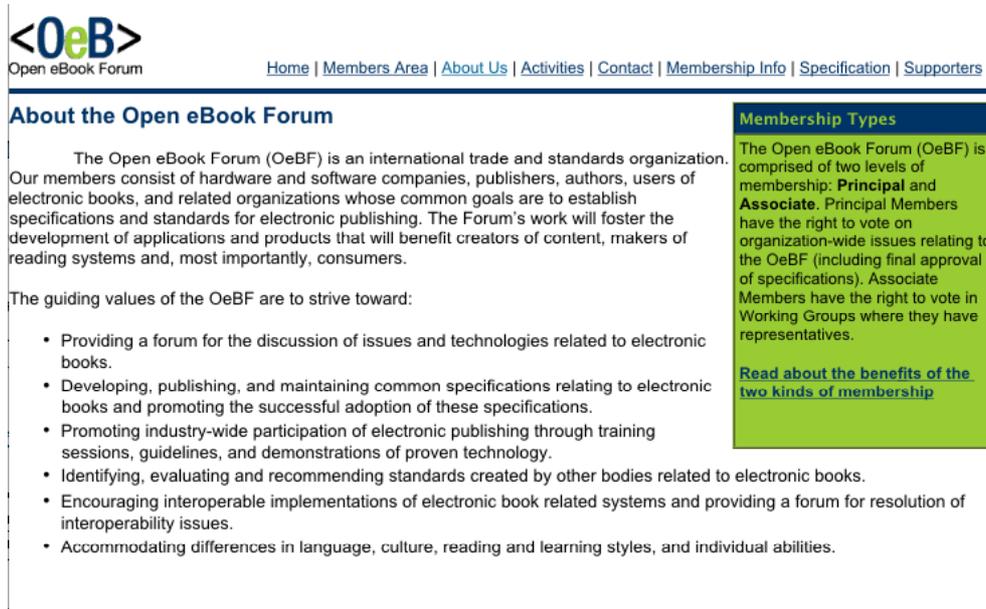


Figura 4. Página del Open eBook Forum.

2.3. Revistas especializadas

Uno de los tipos de publicación que más se ha beneficiado de la extensión de Internet y de los discos ópticos son las revistas especializadas. Existen una gran cantidad de revistas científico-técnicas que cada vez más se distribuyen en disco óptico o a través de Internet. Dos de los principales problemas de las versiones en papel de las revistas especializadas son la piratería de los derechos de autor (las fotocopias) y el excesivo tiempo transcurrido entre el envío de los artículos originales y su publicación.

Para solucionar este segundo problema, numerosas editoriales especializadas en este tipo de revistas, tales como *Academic Press*, *Elsevier* o *Springer-Verlag*, ofrecen a sus suscriptores la posibilidad de acceder a las versiones electrónicas en PDF de sus títulos a través de Internet (ver figura 5). Además, muchos de los congresos científicos que publican las comunicaciones de sus participantes en papel, también están añadiendo últimamente versiones de los mismos en disco óptico.



WILEY
InterScience®

PERSONAL HOME PAGE JOURNAL FINDER BOOK FINDER SEARCH WILEY INTERSCIENCE REFERENCE WORKS HELP CONTACT US

ALL JOURNALS ONLINE SAMPLE COPY PREVIOUS ISSUE NEXT ISSUE ACRONYM FINDER

Table of Contents

CONCURRENCY AND COMPUTATION PRACTICE & EXPERIENCE

Online ISSN: 1532-0634 Print ISSN: 1532-0626
Concurrency and Computation: Practice and Experience
Volume 14, Issue 12, 2002.

Copyright © 2002 John Wiley & Sons, Ltd.

Research Article

957-985 **Dynamically adapting to system load and program behavior in multiprogrammed multiprocessor systems**
Iffat H. Kazi, David J. Lilja
[Abstract](#) [PDF Full Text](#) (Size: 311K)
Published Online: 12 Sep 2002
[DOI](#) 10.1002/cpe.699

987-1008 **Simulating multiple inheritance in Java**
Douglas Lyon
[Abstract](#) [PDF Full Text](#) (Size: 230K)
Published Online: 12 Sep 2002
[DOI](#) 10.1002/cpe.703

Figura 5. Acceso a la versión electrónica en PDF de un artículo en una revista especializada.

Además, también es posible encontrar en Internet un número cada vez mayor de boletines y revistas electrónicas sin su equivalente publicación en papel, que tratan los temas más diversos: astronomía, turismo, química, arte, medicina, filosofía, etc. (ver figura 6).

Algunos de los motivos que explican el auge de las versiones electrónicas de las revistas especializadas son los siguientes:

- Se dirigen a un tipo de público especialmente preparado. Las características de sus lectores suelen incluir un nivel de formación alto, conocimientos de idiomas y acceso y entrenamiento en el uso de las nuevas tecnologías.
- Los costes de producción son pequeños. Las publicaciones académicas se dirigen a un público especializado y, por lo tanto, sus tiradas son notablemente reducidas. Al no utilizar los procesos habituales de impresión y distribución, tienen unos costes materiales menores.
- La distribución es más rápida y barata. El acceso a la publicación en formato electrónico es prácticamente inmediato y reduce enormemente los costes de distribución (envíos por correo, gestión de devoluciones, ...).
- Es posible incluir documentos complementarios. La publicación electrónica de los artículos permite adjuntar a los mismos material complementario, incluyendo imágenes en movimiento, sonido o programas informáticos, entre otros.



Figura 6. Revista electrónica publicada tan sólo en Internet.

Quedan algunos problemas por resolver en relación a las revistas electrónicas. Por ejemplo, el prestigio de las publicaciones en papel suele ser muy superior que el de las versiones electrónicas sin este respaldo, aunque como hemos comentado anteriormente, las grandes editoriales comienzan a editar versiones electrónicas de las revistas más prestigiosas.

2.4. Prensa

La prensa es uno de los sectores que está utilizando con mayor profusión y éxito la distribución digital de sus contenidos. La tendencia actual se orienta a la distribución a través de la web de los periódicos, tomando como base las mismas noticias que los periodistas han elaborado y que se han publicado en la edición impresa, a las cuales se suelen añadir diversas singularidades que comentaremos a continuación.

Casi todos los grandes periódicos del mundo disponen de versiones en línea a través de Internet. En España, y según datos de la cuarta encuesta a usuarios de Internet (abril-junio de 2001) realizada por la Asociación de Investigación de los Medios de Comunicación (www.aimc.es), de las primeras 15 webs que habían visitado recientemente los usuarios, tres corresponden a prensa escrita, concretamente: *El País*, *El Grupo Recoletos* (periódicos económicos como *Expansión* y deportivos como *Marca*) y *El Mundo*.

Las características comunes a las ediciones digitales de los periódicos son las siguientes:

- Proporcionan acceso al texto íntegro y a informaciones complementarias, ampliando las noticias o añadiendo sonidos, imágenes o pequeñas películas.

- Se actualizan diversas veces al día. Pueden consultarse las últimas noticias casi en el instante en que se producen, lo que constituye una de las principales ventajas de las versiones digitales de los periódicos con respecto a las versiones impresas.
- Disponen de servicios complementarios. Tal y como puede verse en la figura 7, las páginas web de los periódicos, no sólo incluyen la versión electrónica de las noticias, sino que se acerca al concepto de portal descrito en el tema 2. En este sentido, los principales periódicos en Internet suelen incluir foros de debate, diccionarios, envío de noticias por correo, tiendas.



Figura 7. Versión electrónica del periódico El País incluyendo servicios complementarios.

- Permiten la consultar y recuperar ediciones anteriores. Los periódicos suelen incluir buscadores que permiten acceder a ejemplares anteriores o buscar en su contenido en base al texto completo de las noticias, (ver figura 8).
- El acceso es parcialmente gratuito. En su origen, todo el contenido de la mayoría de los periódicos en la web podía consultarse de forma gratuita. Sin embargo, la tendencia actual es a que sea necesario suscribirse y pagar una cuota para poder acceder a la mayor parte del contenido. Como mucho, algunos periódicos permiten acceder a sus titulares y a parte de la información que contienen de modo gratuito.
- La consulta se realiza mediante un sistema de navegación. Las noticias de los periódicos se presentan como páginas web, con las ventajas que ello implica en cuanto a la inclusión de información multimedia y a la navegación no secuencial por su contenido basada en el uso de enlaces.



Figura 8. Buscador de ediciones anteriores del periódico El Mundo.

En cuanto a la distribución de periódicos en soporte óptico, se utiliza exclusivamente para ediciones retrospectivas. Periódicos como *El País* o *El Mundo*, editan discos ópticos que contienen los ejemplares del último año y facilitan la consulta de su contenido.

Por otro lado, cabe citar que, aunque menos numerosos, también existen periódicos accesibles tan sólo por Internet y sin una versión impresa. Estos son los verdaderos periódicos digitales e incluyen cabeceras como *La Estrella Digital* (www.estrelladigital.es) o *Vilaweb* (www.vilaweb.com).



Figura 9. Buscador de noticias en prensa.

Por último, y también en relación con la prensa en Internet, hemos de destacar que prácticamente todos los portales de Internet incluyen un sistema de noticias que permite acceder a las últimas noticias publicadas en prensa o distribuidas por las agencias.

Además, estas páginas incluyen sus propios buscadores de noticias en prensa. Adicionalmente es posible encontrar buscadores especializados en encontrar este tipo de información, tales como *Prensa Digital* (www.prensadigital.com), cuya página principal podemos ver en la figura 9.

2.5. Obras de referencia

Otro de los tipos de documentos que suelen publicarse en versión electrónica en disco óptico o a través de Internet son las obras de referencia. Este tipo de documentos condensan y facilitan un acceso rápido a la información. Debido a su estructura facilitan la localización de información de forma rápida y precisa y están pensados para su consulta y no para su lectura completa y ordenada. Incluyen documentos tales como los diccionarios, las enciclopedias o las publicaciones oficiales.

Podemos distinguir las obras de referencia primarias y las secundarias:

Las obras de referencia primarias incluyen la información que se quiere consultar, por ejemplo, la definición de un término, el contenido de una determinada ley o el mapa de un país dado.

Las obras de referencia secundarias remiten a otros libros o documentos para poder encontrar la información buscada. Un ejemplo de este tipo de obra de referencia sería un catálogo de biblioteca.

En el caso de las obras de referencia las principales ventajas de las versiones digitales con respecto a las versiones en papel son, la facilidad y rapidez en la búsqueda, la capacidad de navegar en su contenido conectando temas afines y la velocidad con la que pueden actualizarse sus contenidos.

Revisaremos a continuación los principales tipos de obras de referencia existentes, dando algunos ejemplos accesibles a través de Internet.

2.5.1. Diccionarios

Los diccionarios, presentan la información por orden alfabético y explican el significado de términos o conceptos, o nos muestran su traducción a otro idioma.

Existen una gran cantidad de tipos de diccionarios, incluyendo los generales, los normativos, los de uso, los especializados. Algunos de ellos se distribuyen en CD, y también podemos encontrarlos en Internet. Un ejemplo reciente de incorporación de este tipo de diccionarios a Internet es el caso del Diccionario de la Real Academia de la Lengua (www.rae.es), cuya página principal puede verse en la figura 10. Hasta la fecha venía publicándose tan sólo su versión en papel cada 10 años, con lo cual el avance continuo de la lengua lo superaba en poco tiempo. El objetivo de la Real Academia es actualizar la versión por Internet de modo continuo, interaccionar con los usuarios y editar una versión en papel cada dos o tres años.

The screenshot shows the header of the Real Academia Española website with the date 'Jueves, 19 de septiembre de 2002'. Below the header is a search bar with the text 'Escriba la palabra que desea consultar:' and the word 'diccionario' entered. To the right of the search bar are buttons for 'Ayuda', 'Consultar', and 'Salir'. A list of navigation links is on the left, including 'Presentación', 'El Diccionario en cifras', 'Americanismos', 'Diccionarios académicos', '¿Cómo es y cómo se hace el Diccionario?', 'Consultas y sugerencias acerca del Diccionario', 'Adiciones y enmiendas', 'Añada el Diccionario a su navegador', and 'Adquiera la nueva edición'. The search results for 'diccionario' are shown in a box, including the definition: '(Del b. lat. *diccionarium*). 1. m. Libro en el que se recogen y explican de forma ordenada voces de una o más lenguas, de una ciencia o de una materia determinada. 2. m. Catálogo numeroso de noticias importantes de un mismo género, ordenado alfabéticamente. *Diccionario bibliográfico, biográfico, geográfico.*' and a link to 'mostrar formato DRAE'.

Figura 10. Resultado de una búsqueda en la página del Diccionario de la Real Academia Española.

No obstante, el tipo de diccionarios más comunes en Internet son los bilingües, que nos permiten traducir palabras de un idioma a otro. A través de páginas como *Your Dictionary* (www.yourdictionary.com) es posible acceder a diccionarios bilingües en varios cientos de idiomas, incluyendo desde el inglés y el alemán, hasta el chino o el esperanto. Esta misma página también permite el acceso a decenas de diccionarios especializados en temas tales como los negocios, la informática, el derecho, la medicina o los deportes.

2.5.2. Enciclopedias

Las enciclopedias son obras de referencia que reúnen el conocimiento y la experiencia de una determinada sociedad, y que la presentan en un orden preestablecido.

The screenshot shows the Britannica Online website. The header includes the Britannica logo and the text 'ENCYCLOPEDIA Britannica Online' and 'ENCYCLOPEDIA Britannica | Premium Service'. Below the header, there are two columns of text. The left column is for 'College, School, Library, and Business Users' and includes links for 'Log In Here', 'Subscription Information', 'Frequently Asked Questions', and 'A Quick Tour: An Introduction to the site for new users. Requires Adobe Acrobat.' The right column is for 'Individual and Family Users' and includes links for 'Log In Here', 'Sign Up for a Free Trial', 'Subscription Information', and 'Frequently Asked Questions'. At the bottom, there is a 'Visit the Britannica Store' link and an advertisement for the 'Encyclopaedia Britannica 2002 Print Set' with a price of USD \$1,295.

Figura 11. Página principal de la versión en línea de la Enciclopedia Británica.

Aunque Internet en sí misma constituye la mayor enciclopedia del mundo, al permitimos acceder a información sobre todos los temas imaginables, las versiones en

línea de las enciclopedias tradicionales son más bien escasas. Existen algunas enciclopedias como la Enciclopedia Británica (www.eb.com) que, además de editar una versión en CD y DVD, permiten el acceso gratuito durante unos días a su versión en línea y el acceso continuado previa suscripción. En la figura 11 podemos ver la página principal de esta enciclopedia.

2.5.3. Directorios

El directorio es una obra de referencia que contiene información presentada en orden alfabético o cronológico que permite la identificación o localización de personas, organismos o empresas. Si se trata de directorios de personas, pueden incluir su nombre, actividad, función, dirección, teléfono o correo electrónico.

Internet incorpora un servicio de directorio denominado X.500, o el más reciente LDAP (*Lightweight Directory Access Protocol*), que en principio debería permitir encontrar las direcciones postales, números de teléfono y direcciones de correo electrónico de personas que forman parte de organizaciones que están conectadas a Internet. No obstante este sistema funciona muy mal y tan sólo permite acceder en ocasiones a datos sobre personas que trabajan en la misma empresa u organización. Por ejemplo, la UJI dispone de un sistema de directorio basado en el servicio X.500, además de estar incorporando otro basado en el LDAP (ver figura 12).

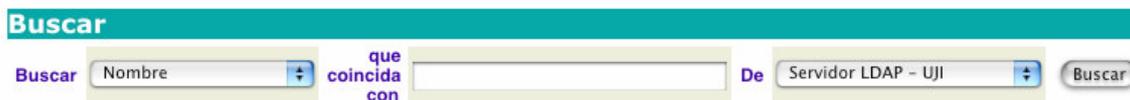


Figura 12. Buscador de personas en el servicio de directorio (LDAP) de la UJI.

Además, existen en Internet buscadores especializados en encontrar información sobre personas, que suelen agruparse bajo la categoría de páginas blancas. Entre ellos podemos citar por ejemplo *Yahoo! People Search* (people.yahoo.com), cuyo formulario de búsqueda avanzada podemos ver en la figura 13, o *Infospace* (www.infospace.com). No obstante, por diversas razones, los servicios centrados en la búsqueda de personas en Internet son bastante limitados, ya que permiten encontrar a un porcentaje bastante reducido de personas que disponen de una cuenta en algún proveedor de Internet o de dirección de correo. Entre las razones que justifican esta limitación se encuentra el que los proveedores no suelen proporcionar los datos de sus usuarios.

Entre los directorios de empresas podemos destacar por ejemplo las páginas amarillas (www.paginasamarillas.es) que funcionan de modo similar a la versión en papel, pero abarcan un mayor ámbito geográfico que las que recibimos en nuestra casa.

Figura 13. Formulario de búsqueda avanzada de *Yahoo People Search*.

2.5.4. Fuentes geográficas

Las fuentes geográficas nos permiten encontrar lugares mediante mapas o planos e incluyen atlas generales, o especializados (como los atlas históricos o económicos), y guías y callejeros de ciudades.



Figura 14. Ejemplo de mapa interactivo de España que podemos encontrar en Internet.

Por ejemplo, en *encarta.msn.com* podemos acceder a la versión de demostración en línea interactiva del Atlas incluido en la Enciclopedia Encarta. En ella podemos acceder a una mapa sensible (mapa que responde a las pulsaciones del usuario) que nos permite ver mapas de distintas regiones o países, e incluso acceder a información sobre ciudades concretas sin más que pulsar con el ratón en la zona deseada (ver figura 13). Otro ejemplo de atlas electrónico puede ser *Atlapedia* (www.atlapedia.com) nos permite acceder a mapas físicos y políticos de distintos países, así como a información sobre su clima, lengua, historia, etc.

Como ejemplo de callejero que nos puede permitir ver planos de zonas de distintas ciudades de España, podemos acceder al incluido en el periódico *El Mundo* (www.elmundo.es/callejero).

2.5.5. Publicaciones oficiales

Uno de los tipos de publicación que se ha venido difundiendo en versión electrónica desde hace más tiempo son los repertorios de legislación y las publicaciones oficiales. Desde hace muchos años, editoriales especializadas como Aranzadi vienen distribuyendo en CD las disposiciones legales incluidas en el BOE o dictadas por los gobiernos de las comunidades autónomas o de la Unión Europea.

Asimismo, y desde hace menos tiempo, podemos acceder a través de Internet a los contenidos de los últimos números de los boletines oficiales, tanto a nivel provincial y autonómico (DOGV - www.pre.gva.es/dogvweb/cindex.htm), como nacionales (BOE - www.boe.es) o de la Unión Europea (DOCE - europa.eu.int/eur-lex/es/o).



Figura 15. Página del Diario Oficial de la Generalitat Valenciana.

También es posible acceder en Internet buscadores especializados en encontrar legislación publicada en diversos boletines como *Toda la ley* (www.todalaley.com).

2.5.6. Obras de referencia secundaria

Entre las obras de referencia cuyos resultados nos remiten a otros documentos para encontrar la información final, podemos destacar los catálogos de bibliotecas y las bibliografías.

Como ya comentamos en el tema 3, la mayoría de las grandes bibliotecas ofrecen sus servicios a través de páginas web. Uno de los servicios fundamentales a los que puede accederse es la búsqueda de libros en los catálogos de la biblioteca. Este tipo de catálogos accesibles en línea se denominan OPAC (*On-line Public Acces Catalogue*).

Figura 16. Catálogo en Internet de la Biblioteca Nacional de España.

Los catálogos de biblioteca son un inventario de los documentos localizados en ella. Cada documento viene descrito siguiendo unas directrices internacionales por parámetros tales como el autor, título, lugar y año de edición, etc. Además, en los catálogos consta la información sobre los ejemplares concretos que posee la biblioteca, su disponibilidad y su localización. En la figura 16 podemos ver el buscador del catálogo de la Biblioteca Nacional Española (www.bne.es).

Por otro lado, en Internet podemos encontrar diversas recopilaciones de bibliografía sobre todo tipo de temas, así como buscadores especializados en encontrar referencias bibliográficas. Este tipo de buscadores van dirigidos a profesionales e investigadores especializados y permiten el acceso a referencias sobre su ámbito de trabajo. Funcionan de modo similar a un catálogo de biblioteca, pero ofrecen como resultado referencias a artículos científico-técnicos aparecidos en revistas, actas de conferencias o informes técnicos. La información relacionada con cada artículo suele incluir los autores, título, revista o conferencia en que se publicó, lugar y fecha de publicación, etc. En algunas bibliografías se incluye también un resumen del contenido del artículo. Más aún, algunos buscadores bibliográficos ofrecen enlaces que permiten acceder a las fuentes primarias de algunos de los artículos que contienen. Así, podríamos acceder a una versión electrónica de las referencias obtenidas, bien en la página relacionada con alguna conferencia, en algún repositorio FTP e incluso en la página personal de alguno de los autores.

Entre los buscadores de bibliografía podemos citar *ResearchIndex* (citeseer.nj.nec.com/cs), especializado en literatura científica (ver figura 17), o *The collection of computer science bibliographies* (iinwww.ira.uka.de/bibliography) especializado en informática.

The screenshot shows the CiteSeer search interface. At the top, the CiteSeer logo is on the left, followed by a search bar containing 'Toeplitz'. To the right of the search bar are two buttons: 'Documents' and 'Citations'. Below the search bar, the text reads: 'Searching for **toeplitz**. Restrict to: [Header](#) [Title](#) Order by: [Citations](#) [Hubs](#) [Usage](#) [Date](#) Try: [Amazon](#) [B&N](#) [Google \(RI\)](#) [Google \(Web\)](#) [CSB](#) [DBLP](#) 1136 documents found. **Only retrieving 500 documents (System busy - maximum reduced).** Retrieving documents... **Order: citations weighted by year.**

Two search results are visible:

- [Spatio-Temporal Coding for Wireless Communication - Raleigh, Cioffi \(1998\)](#) (Correct) (35 citations)
 . with each subblock possessing the well-known **Toeplitz** form. To clearly illustrate the effect of over multipath components 3 .2) where the **Toeplitz** pulse shaping matrix is . one-dimensional vector expression with a block **Toeplitz** channel matrix. As we will see, this www-isl.stanford.edu/~cioffi/dsm/wlpap/mimocioffi98.pdf
- [Toeplitz and Circulant Matrices: A review - Gray \(2001\)](#) (Correct) (5 citations)
Toeplitz and Circulant Matrices: A review # (pdf) file at ee.stanford.edu/gray/toeplitz.pdf c Robert M. Gray, 1971, 1977, 1993, 1997, inverses, and products of "finite section" **Toeplitz** matrices and **Toeplitz** matrices with absolutely ee.stanford.edu/~gray/toeplitz.pdf

Figura 17. Resultado obtenido en el buscador de bibliografía *ResearchIndex*.

3. Tipos de software

En este apartado vamos a revisar brevemente los principales tipos de programas que es posible encontrar en Internet.

En primera instancia podemos diferenciar los **programas comerciales**, desarrollados por fabricantes de software y de pago, de los **programas de libre distribución**, pensados para su difusión por Internet, en muchos casos de forma totalmente gratuita.

Los formatos más comunes de software de libre distribución son los siguientes:

Shareware. Como su propio nombre indica, están pensados para ser compartidos y distribuidos. La idea de este tipo de software es que el usuario pruebe su funcionamiento durante un tiempo, y si le resulta de utilidad y va a seguir usándolo, pague una pequeña cantidad a su autor.

Freeware. Su finalidad es similar a la del tipo anterior, pero en este caso se trata de software totalmente gratuito que puede usarse y copiarse sin ninguna restricción.

Además de los dos formatos anteriores, en Internet podemos encontrar otros tipos de programas:

Addware. Se trata de versiones de programas que podemos descargar de Internet y utilizar de modo totalmente gratuito. Sin embargo, estos programas insertan en alguna de sus ventanas publicidad.

Versiones de demostración. Muchos fabricantes de software distribuyen versiones de demostración de sus programas. Estas versiones suelen tener una funcionalidad limitada respecto a la versión de pago y pueden dejar de funcionar al cabo de unos días.

Actualizaciones. Se trata de pequeños programas que actualizan las versiones de otros que tenemos instalados a unas más nuevas. En muchas ocasiones la compra de un programa nos da derecho a recibir actualizaciones periódicas del mismo de modo gratuito o a un precio más reducido.

4. Formatos de documento

En el ámbito de la informática, todos los documentos, y en general toda la información que es manejada mediante ordenadores se almacena en forma de ficheros. Los ficheros son el elemento básico mediante el cual los sistemas operativos almacenan la información. Sistemas Operativos como las distintas versiones de Windows o de UNIX almacenan toda la información en ficheros y la organizan en los dispositivos de almacenamiento (discos duros, disquetes, discos ópticos, etc.) mediante un sistema jerárquico de ficheros y directorios. Windows representa gráficamente estos elementos mediante iconos correspondientes a ficheros y carpetas donde se almacenan, tanto los ficheros como otras carpetas.

Los ficheros guardan la información en un determinado formato que, en sistemas operativos de tipo Windows, se identifica mediante una extensión de 2 a 4 letras añadidas a su nombre y mediante un tipo particular de icono con el que se representa en pantalla.

Cuando hablamos de **formato** de un fichero informático nos referimos a la forma en que se ha codificado la información en el mismo, esto es, en último término el formato de un fichero es el modo en que se ha traducido a ceros y unos la información que almacena. Este formato debe ser reconocido e interpretado de forma adecuada por los programas que lo manejan para poder visualizar y manipular la información que contienen.

Existen formatos propios de determinadas aplicaciones que corresponden a la forma en que las mismas almacenan la información por defecto. Por ejemplo, los documentos almacenados mediante *Microsoft Word* suelen tener la extensión *.doc*, mientras los almacenados con *WordPerfect* usan la extensión *.wpf*. Por otro lado, existen formatos de fichero más genéricos que corresponden a un tipo particular de información y que pueden venir o no dados por un estándar. Por ejemplo, la mayoría de las imágenes que se manejan en Internet se almacenan en formato GIF o JPEG. En el caso de los formatos genéricos, existen distintos programas que pueden leerlos y visualizarlos y que nos pueden permitir modificar el contenido de los ficheros. Así, numerosas aplicaciones gráficas nos permiten acceder a los formatos GIF y JPEG o guardar los gráficos que hemos generado o modificado con ellas en estos formatos. Lo anterior no significa que no podamos intentar abrir cualquier tipo de fichero con la mayoría de las aplicaciones. Lo que ocurre simplemente es que cuando las aplicaciones no reconocen un determinado formato, pueden producir un error o bien mostrarnos la información más extraña imaginable en pantalla, tal y como puede constatarse en la figura 18.

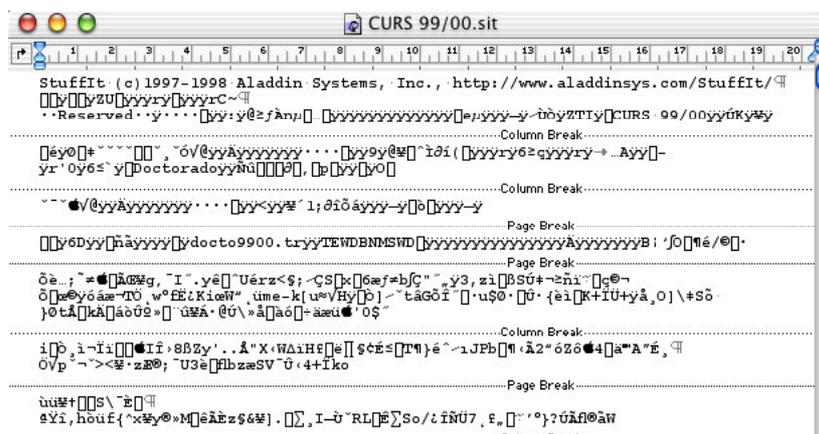


Figura 18. Ejemplo de resultado en pantalla al abrir un fichero no reconocido por el procesador Word.

En este apartado vamos a revisar algunos de los principales formatos informáticos en los que pueden guardarse distintos tipos de información. Intentaremos centrarnos en los formatos genéricos de mayor difusión y citaremos algunas aplicaciones que nos pueden permitir visualizar y, en algunos casos, modificar su contenido.

4.1. Textos electrónicos

Uno de los tipos de información más difundida son los textos electrónicos. En un primer nivel podemos clasificar estos en dos tipos básicos:

El **texto puro** o texto sin formato corresponde a ficheros que sólo almacenan información textual, incluyendo letras, números, signos de puntuación y otros símbolos como paréntesis, interrogantes, etc. Algunos de estos ficheros restringen tanto los caracteres que pueden almacenar que no permiten por ejemplo guardar texto con acentos o la letra ñ.

El **texto con formato** es aquel que no sólo contienen la información textual, sino que además dota al texto de determinadas características que mejoran la presentación global del documento. En este tipo de documentos puede por ejemplo incluirse texto utilizando distintas fuentes (tipo de letra), podemos utilizar texto en negrita o en cursiva, o añadir características tales como el sangrado o la justificación a los párrafos. Más aún, aunque estrictamente hablando no se trata de texto, podemos incluir en la categoría de textos electrónicos con formato aquellos que incorporan imágenes, gráficos o tablas y que pueden ubicarlos en determinadas posiciones del documento.

Algunas de las condiciones deseables en los ficheros que guardan texto con formato son las siguientes:

- Que puedan ser visualizables con un visor universal, fácil de conseguir para distintos sistemas operativos y a ser posible gratuito.
- Que conserven el color, formatos y fuentes con los que fueron generados.
- Que puedan incluir imágenes.

- Que se impriman con fidelidad al original.
- Que permitan añadir condiciones de seguridad para impedir su alteración, e incluso su impresión.
- Que ocupen poco espacio en disco.

En los siguientes apartados vamos a revisar la codificación utilizada en los ficheros de texto puro, los principales tipos de textos con formato que podemos encontrar y sus ventajas e inconvenientes con respecto a las características anteriores. Entre los textos con formato vamos a diferenciar tres tipos básicos:

- El generado mediante procesadores de texto.
- Los formatos de visualización e impresión.
- Los lenguajes de formato

4.1.1. Texto puro (ASCII)

Para guardar información textual es necesario acordar una codificación que haga corresponder cada carácter almacenado con una combinación de ceros y unos. Existen diversos estándares internacionales que establecen esta correspondencia, el más extendido de los cuales es el código ASCII (*American Standard Code for Information Interchange*).

La versión básica de este código incluye un máximo de 128 caracteres, de los cuales los 31 primeros no corresponden a caracteres representables en pantalla, sino que se utilizan para almacenar información de control, por ejemplo, para indicar cuando se producen los cambios de párrafo. En este código los números del 0 al 9 se corresponden con los códigos del 48 al 57 y las letras minúsculas quedan codificadas con los valores entre el 97 y el 122. En la versión básica del código no están representadas por ejemplo las letras con acento, la letra ñ, o los símbolos de apertura de interrogación y exclamación (¡, ¿), ni diversos caracteres especiales utilizados por idiomas distintos del inglés. Debido a que esto representa un claro problema fuera de los países del ámbito anglosajón, existe la versión extendida del código ASCII, que permite almacenar hasta 256 caracteres distintos y que incluye todos los de la versión básica en la misma posición y añade los que hemos comentado (p.e. la ñ corresponde al código 164). En la tabla 1 podemos ver los caracteres del código ASCII extendido representables.

00100000 32	00100001 33 !	00100010 34 "	00100011 35 #	00100011 36 \$	00100101 37 %	00100110 38 &	00100111 39 '	00101000 40 (00101001 41)	00101010 42 *	00101011 43 +
00101100 44 ,	00101101 45 -	00101110 46 .	00101111 47 /	00110000 48 0	00110001 49 1	00110010 50 2	00110011 51 3	00110100 52 4	00110101 53 5	00110110 54 6	00110111 55 7
00111000 56 8	00111001 57 9	00111010 58 :	00111011 59 ;	00111100 60 <	00111101 61 =	00111110 62 >	00111111 63 ?	00100000 64 @	01000001 65 A	01000010 66 B	01000011 67 C
01000100 68 D	01000101 69 E	01000110 70 F	01000111 71 G	01001000 72 H	01001001 73 I	01001010 74 J	01001011 75 K	01001100 76 L	01001101 77 M	01001110 78 N	01001111 79 O
01010000 80 P	01010001 81 Q	01010010 82 R	01010011 83 S	01010100 84 T	01010101 85 U	01010110 86 V	01010111 87 W	01011000 88 X	01011001 89 Y	01011010 90 Z	01011011 91 [
01011100 92 \]	01011100 93]	01011110 94 ^	01011111 95 _	01100000 96 ,	01100001 97 a	01100010 98 b	01100011 99 c	01100100 100 d	01100101 101 e	01100110 102 f	01100111 103 g
01101000 104 h	01101001 105 i	01101010 106 j	01101011 107 k	01101100 108 l	01101101 109 m	01101110 110 n	01101111 111 o	01110000 112 p	01110001 113 q	01110010 114 r	01110011 115 s
01110100 116 t	01110101 117 u	01110110 118 v	01110111 119 w	01111000 120 x	01111001 121 y	01111010 122 z	01111011 123 {	01111100 124 	01111101 125 }	01111110 126 ~	01111111 127 DEL
10000000 128 Ç	10000001 129 ü	10000010 130 é	10000011 131 â	10000100 132 ä	10000101 133 à	10000110 134 á	10000111 135 ç	10001000 136 ê	10001001 137 ë	10001010 138 è	10001011 139 ï
10001100 140 î	10001101 141 ì	10001110 142 Ë	10001111 143 Ä	10010000 144 É	10010001 145 æ	10010010 146 Æ	10010011 147 ô	10010100 148 ö	10010101 149 ò	10010110 150 û	10010111 151 ù
10011000 152 ÿ	10011001 153 Ö	10011010 154 Ü	10011011 155 ø	10011100 156 £	10011101 157 Ø	10011110 158 x	10011111 159 f	10100000 160 á	10100001 161 í	10100010 162 ó	10100011 163 ú
10100100 164 ñ	10100101 165 Ñ	10100110 166 ª	10100111 167 º	10101000 168 ¿	10101001 169 ®	10101010 170 ¬	10101011 171 _	10101100 172 _	10101101 173 i	10101110 174 «	10101111 175 »
10110000 176 _	10110001 177 _	10110010 178 _	10110011 179 	10110100 180 	10110101 181 Á	10110110 182 Â	10110111 183 Ã	10111000 184 ©	10111001 185 	10111010 186 	10111011 187 +
10111100 188 +	10111101 189 ¢	10111110 190 ¥	10111111 191 +	11000000 192 +	11000001 193 -	11000010 194 -	11000011 195 +	11000100 196 -	11000101 197 +	11000110 198 ã	11000111 199 Ä
11001000 200 +	11001001 201 +	11001010 202 -	11001011 203 -	11001100 204 	11001101 205 -	11001110 206 +	11001111 207 _	11010000 208 _	11010001 209 _	11010010 210 Ê	11010011 211 Ë
11010100 212 È	11010110 213 _	11010111 214 Í	11011000 215 Î	11011001 216 Ï	11011010 217 +	11011011 218 +	11011100 219 _	11011100 220 _	11011101 221 	11011110 222 î	11011111 223 _
11100000 224 Ó	11100001 225 ß	11100010 226 Ô	11100011 227 Ò	11100100 228 ø	11100101 229 Ï	11100110 230 μ	11100111 231 _	11101000 232 _	11101001 233 Ú	11101010 234 Û	11101011 235 Ü
11101100 236 y	11101101 237 Y	11101110 238 -	11101111 239 ,	11110000 240 _	11110001 241 ±	11110010 242 _	11110011 243 _	11110100 244 ¶	11110101 245 §	11110110 246 ÷	11110111 247 ,
11111000 248 °	11111001 249 ..	11111010 250 .	11111011 251 1	11111100 252 3	11111101 253 2	11111110 254 _	11111111 255 !				

Tabla 1. Código ASCII extendido. En cada celda: código binario, decimal y carácter correspondiente.

Con respecto a las ventajas e inconvenientes de los documentos de texto puro:

La principal ventaja del texto puro es que utiliza una codificación estándar de cada símbolo. De este modo, numerosas aplicaciones pueden leer el contenido de este tipo de documentos y nos permiten su modificación. Así, todos los procesadores y editores de texto pueden leer documentos en este formato y la mayoría de ellos ofrecen la posibilidad de guardar los documentos utilizándolo. Esto hace que con el fin de facilitar la máxima difusión de la información y simplificar su acceso, muchos de los documentos que podemos encontrar en Internet se encuentren en formato de texto puro. Así, muchos de los libros electrónicos que podemos encontrar en los principales repositorios de la red, están almacenados en este formato.

Por su propia definición los documentos en texto puro no conservan color, formatos o fuentes porque no permiten incorporarlos, y tampoco permiten incluir imágenes.

Al tratarse de un formato tan simple, la impresión del texto puro conserva una alta fidelidad con el original.

En cuanto a la seguridad, no es posible añadir condiciones para evitar la alteración de los documentos en texto puro.

Finalmente, y como otra ventaja de este tipo de formato, los documentos guardados en el mismo ocupan un espacio muy reducido, dado que sólo guardan los caracteres que contienen y ninguna información adicional sobre su formato.

4.1.2. Generado con procesadores de texto

Una parte importante de los documentos de texto que solemos manejar con los ordenadores han sido generados utilizando algún procesador de textos. El objetivo de este tipo de aplicaciones es precisamente la creación o modificación de texto con formato. Alguno de los procesadores de textos más conocidos son *Microsoft Word* y *WordPerfect*.

Los procesadores de textos más extendidos son comerciales y guardan por defecto la información en un formato que les es propio. Esto significa que tan sólo ese mismo procesador es capaz de leer estos documentos. No obstante, dada la importancia del fácil intercambio de información, todos los procesadores permiten almacenar los documentos en otros formatos que no poseen la restricción anterior. Por ejemplo, podemos almacenar la información como texto puro o en formato RTF (*Rich Text Format*). El problema de guardar los documentos con un procesador en un formato distinto del que le es propio, es que en muchos casos se pierden parte de sus características. Por ejemplo, si guardamos un documento Word como texto puro, perdemos todo el formato que contiene, incluyendo obviamente cualquier imagen o gráfico.

Otro problema que tienen los formatos propios de los procesadores de texto es la poca consistencia de los mismos entre distintas versiones del programa. Así, distintas

versiones de un mismo procesador pueden no reconocer un documento generado por otras o cambiar notablemente su apariencia. Es muy normal que versiones antiguas de un mismo procesador no sean capaces de abrir documentos generados con versiones más recientes. Por ejemplo, *Word 6* puede tener bastantes dificultades en abrir y visualizar correctamente un documento creado y guardado con *Word 2000*. No sólo eso, una misma versión de un procesador, pero utilizada sobre dos sistemas operativos distintos (p.e. *Windows XP* y *MacOS X*) puede dar lugar a visualizaciones bastante diferentes del mismo documento.

En definitiva, este tipo de documentos no disponen de un visor universal y conservan, aunque no totalmente el color, formato y fuentes del original.

Este tipo de documentos permiten incluir imágenes.

Al igual que ocurre con la visualización, no siempre es posible imprimir este tipo de documentos con fidelidad. El resultado depende en gran medida de la versión del procesador que usemos para imprimirlo, de los tipos de caracteres disponibles en el sistema operativo y de la impresora que utilicemos.

Los principales procesadores de textos suelen incorporar algún mecanismo para proteger los documentos contra su posterior alteración. Principalmente el uso de contraseñas.

Otro de los inconvenientes de este tipo de formato es que los ficheros que lo utilizan suelen ocupar mucho espacio. De hecho, bastante más del que parece lógico para incluir información sobre su contenido y formato.

4.1.3. Formatos de visualización e impresión

Existen formatos especialmente creados para la visualización e impresión de calidad de textos con formato. Entre ellos podemos destacar dos muy extendidos: El formato PDF y el formato PostScript.

El formato PDF (*Portable Document Format*) [*.pdf*] es un formato muy difundido creado por Adobe y pensado especialmente para la distribución de textos electrónicos con formato. Incorpora todas las características deseables para este tipo de formatos:

- Existe un visor universal de libre distribución (gratuito) que permite visualizar este tipo de documentos en los sistemas operativos más extendidos: el *Acrobat Reader*.
- El formato está pensado para conservar las fuentes, formatos y colores de los documentos.
- Permite la ubicación precisa de imágenes y los mantiene ante un cambio de sistema operativo.
- Aunque no es estrictamente un formato de impresión, mantiene un alto grado de fidelidad al original en las versiones impresas.

- Las aplicaciones para la generación de documentos en PDF permiten incorporar condiciones de seguridad a los documentos que impiden su posterior alteración o impresión.
- Los documentos en PDF suelen ocupar bastante menos espacio que los equivalentes generados con los procesadores de textos más extendidos.

El principal inconveniente de los documentos en formato PDF es que no pueden editarse y modificarse, y resulta complicado extraer o copiar parte del texto o de las imágenes que contienen. Dado que se trata de un formato pensado para el intercambio y la visualización de documentos, no existen programas pensados para su modificación.

Podemos crear ficheros con este formato a partir de procesadores de texto y otras aplicaciones como editores de gráficos u hojas de cálculo. Para ello es necesario disponer de la aplicación comercial (de pago) *Adobe Acrobat*. Esta aplicación permite imprimir los documentos sobre una impresora especial que no genera una copia en papel, sino un fichero en formato PDF. Además, permite incorporar a aplicaciones como Word la posibilidad de guardar directamente en este formato mediante una de sus opciones de menú.

El formato PostScript [.ps] fue creado específicamente para permitir la impresión de documentos con alta calidad y fidelidad al original. Este tipo de ficheros poseen unas características muy similares al formato PDF.

- Existen visores bastante difundidos que permiten visualizar los documentos PostScript. Sin embargo, no se trata de un formato pensado para una visualización de calidad, sino para su impresión, con lo que no siempre podemos esperar buenos resultados en pantalla. El visor más difundido de este tipo de formato es *GhostView*.
- Permite conservar las fuentes, formatos y colores de los documentos en su versión impresa.
- Permite la ubicación precisa de imágenes.
- Conserva con total precisión el formato al imprimir el documento, ya que es su principal finalidad. No obstante para poder imprimir documentos de este tipo es necesario utilizar impresoras que reconozcan este formato. Este tipo de impresoras se denominan impresoras PostScript y no suelen ser las más vendidas entre usuarios particulares.
- El formato PostScript no incorpora mecanismos para impedir su modificación, aunque, tal y como puede verse en la figura 20, resulta cuanto menos difícil lograrlo accediendo directamente a la información textual almacenada.

- Los documentos en PostScript suelen ser de pequeño tamaño salvo cuando incluyen imágenes o gráficos que ocupan mucha memoria.

```

%!PS-Adobe-2.0
%%Creator: dvips(k) 5.86d Copyright 1999 Radical Eye Software
%%Title: introprogr.dvi
%%Pages: 3
%%PageOrder: Ascend
%%BoundingBox: 0 0 596 842
%%EndComments
%DVIPSWebPage: (www.radicaleye.com)
%DVIPSCommandLine: dvips introprogr
%DVIPSParameters: dpi=600, compressed
%DVIPSSource: TeX output 2001.05.23:1542
%%BeginProcSet: texc.pro
%!
/TeXDict 300 dict def TeXDict begin/N{def}def/B{bind def}N/S{exch}N/X{S
N}B/A{dup}B/TR{translate}N/isls false N/vsize 11 72 mul N/hsize 8.5 72
mul N/landplus90{false}def/@origin{isls{[0 landplus90{1 -1}{-1 1}ifelse 0
0 0]concat}if 72 Resolution div 72 VResolution div neg scale isls{
landplus90{VResolution 72 div vsize mul 0 exch}{Resolution -72 div hsize
mul 0}ifelse TR}if Resolution VResolution vsize -72 div 1 add mul TR[
matrix currentmatrix{A A round sub abs 0.00001 lt{round}if}forall round
exch round exch]setmatrix}N/@landscape{/isls true N}B/@manualfeed{
statusdict/manualfeed true put}B/@copies{/#copies X}B/FMat[1 0 0 -1 0 0]
N/FBB[0 0 0 0]N/nn 0 N/IEn 0 N/ctr 0 N/dE-tail{/nn 8 dict N nn begin
/FontType 3 N/FontMatrix /ctr N/FontBox /FBB N/string/base X array
/BitMaps X/BuildChar{CharBuilder}N/Encoding IEn N end A{/foo setfont}2
array copy cvx N load 0 nn put/ctr 0 N}B/sf 0 N/dF{/sf 1 N/FontType 3

```

Figura 19. Contenido de un fichero en formato PostScript.

Al igual que ocurre con el formato PDF, los ficheros en PostScript no pueden modificarse, ni pueden extraerse o copiarse parte de los mismos, tan sólo pueden visualizarse y, fundamentalmente, imprimirse.

Los ficheros PostScript suelen generarse a partir de cualquier aplicación que genere un documento susceptible de ser impreso. Si disponemos de una impresora PostScript y del software adecuado para imprimir en la misma, las propias aplicaciones permitirán imprimir sobre un fichero PostScript en lugar de sobre papel.

4.1.4. Lenguajes de formato

Los lenguajes de formato están pensados para describir el contenido y formato de documentos. Mediante una sintaxis preestablecida nos permiten describir todas las características relativas al formato del documento y del texto que contienen, pudiendo incluir aspectos tales como los márgenes, las cabeceras, las características asociadas a los caracteres, a los párrafos, etc. Además, nos permiten describir tablas y ubicar imágenes. La idea es poder describir textos electrónicos con formato totalmente independiente del entorno en el que estemos visualizándolos, de modo que su apariencia no varíe aunque cambiemos de ordenador o de sistema operativo.

El más difundido de estos lenguajes de formato en la actualidad es el HTML (*HyperText Mark-Up Language*), debido a que se usa para la creación de páginas web. Este lenguaje de formato tiene su origen en el SGML (*Standard Generalized Mark-Up Language*) que constituye un estándar internacional para definir lenguajes de formato basados en etiquetas. El lenguaje HTML se basa en las normas establecidas por el SGML para dar formato a textos e incorporar enlaces, es decir, para crear documentos hipertexto con formato.

El que el lenguaje esté basado en el uso de etiquetas significa que existen unos textos especiales (etiquetas, o *tags* en inglés) que se incorporan al documento y que permiten definir las partes de que consta y todo lo relativo al formato de las mismas.

Dado que las páginas web han ido aumentando su complejidad para responder a las demandas de los usuarios de la web, el lenguaje utilizado para poder describirlas ha ido evolucionando. De este modo se han definido varias versiones del lenguaje HTML. En la actualidad la versión más reciente es el HTML 4.01. No obstante, este lenguaje sigue teniendo serias limitaciones para adaptarse a las nuevas tecnologías y medios, tales como la telefonía móvil. Lo que se ha hecho es definir una nueva versión de las normas dadas por el SGML denominada XML (*eXtensible Mark-up Language*), a partir de la cual será posible definir nuevas versiones del HTML, denominadas XHTML.

Además de los lenguajes de formato relacionados con el HTML y con las páginas web, existe otro lenguaje de formato muy extendido entre los científicos e investigadores, ya que está especialmente pensado para la descripción de artículos científico-técnicos con un formato sofisticado, incluyendo gráficos y fórmulas matemáticas complejas. Este lenguaje se denomina *LaTeX* (ver figura 21) y existen programas que permiten la traducción de los documentos escritos en el mismo a formatos visualizables o imprimibles, tales como el PostScript.

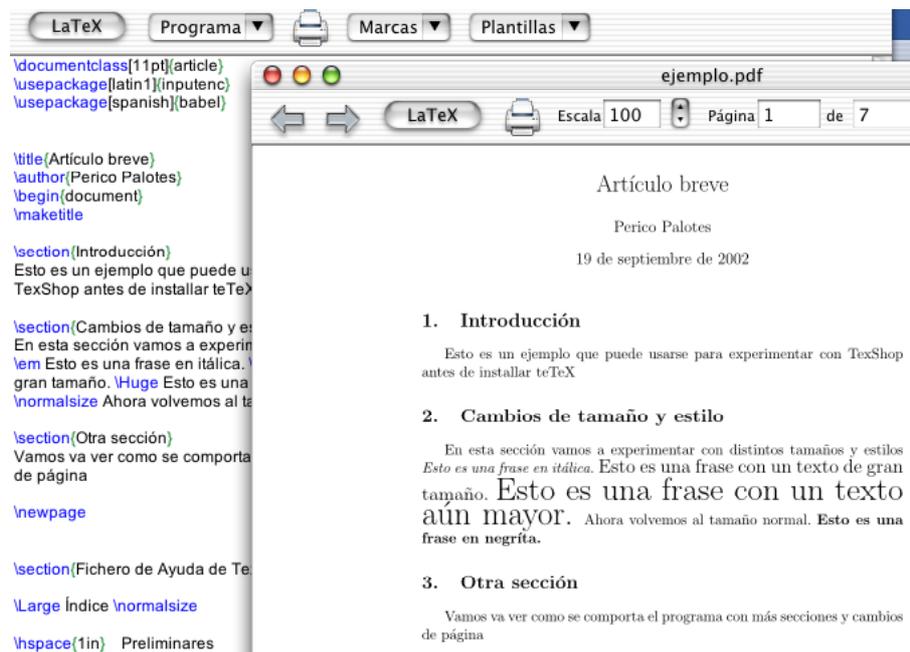


Figura 20. Ejemplo de documento en LaTeX (detrás) y de su resultado una vez interpretado (delante).

Las características generales de los documentos definidos mediante lenguajes de formato son las siguientes:

- Al estar pensados para ser visualizados en cualquier entorno, existen programas para distintos sistemas operativos que son capaces de traducir el lenguaje a un documento con formato. Por ejemplo, los navegadores web

están pensados para visualizar documentos escritos en HTML, y las versiones más recientes de los principales procesadores de textos también permiten su visualización.

- Permiten la inclusión de fuentes, formatos y colores en el documento, aunque la variedad de los mismos depende del lenguaje en concreto. Así, el lenguaje HTML es bastante limitado en este sentido, mientras el LaTeX es mucho más completo.
- Los lenguajes de formato suelen permitir la ubicación de imágenes en el texto, pero la precisión del resultado obtenido depende de cada lenguaje concreto.
- La calidad de la impresión obtenida depende mucho del lenguaje utilizado. Mientras HTML no garantiza una impresión fidedigna y depende mucho del programa visor utilizado para imprimir, el lenguaje LaTeX está pensado para generar versiones imprimibles que no varíen al cambiar el entorno de trabajo.
- Los lenguajes de formato no están pensados para incorporar mecanismos de protección de los documentos obtenidos.
- Los ficheros con lenguaje de formato suelen ser de tamaño reducido, ya que contienen tan sólo texto puro. No obstante hay que tener en cuenta que no incorporan las imágenes o gráficos y que estas deben ser tratadas como ficheros adicionales cuyo tamaño debería sumarse al del fichero con el lenguaje de formato.

4.2. Imágenes

Otro de los tipos de información más extendidos en el ámbito de la informática y de las comunicaciones son las imágenes. Existen muchos formatos para almacenar imágenes con distintas características en cuanto a la precisión de las mismas y el tamaño de los ficheros resultantes.

Cuando se guarda un documento de texto puro, los caracteres que lo componen son codificados mediante un estándar como ASCII, y sus códigos numéricos son almacenados en forma de ceros y unos en el fichero. En el caso de las imágenes sus componentes básicas no son los caracteres, sino los puntos de distintos colores o con distintos tonos de gris que la conforman. Estos puntos se denominan *pixels*. Una forma simple de almacenar una imagen es codificar cada uno de los colores distintos de que puede constar la misma como un código numérico único. Cada imagen se guardará como una sucesión de números (traducidos a ceros y unos) correspondientes a cada uno de sus puntos (*pixels*). Este es el mecanismo utilizado por el formato GIF (*Graphics Interchange Format*), que codifica un total de 256 colores distintos. Este es, junto con el formato JPEG uno de los dos sistemas de almacenamiento de imágenes que más se

utiliza en Internet. La mayor parte de las imágenes incluidas en páginas web siguen uno de estos dos formatos y cualquier navegador web puede visualizarlos sin problemas.

El formato JPEG (*Join Photographers Expert Group*), permite guardar las imágenes en un formato comprimido. De esta forma, una misma imagen suele ocupar menos espacio en este formato que en formato GIF. No obstante, la precisión con que se guarda la imagen en formato JPEG es inferior.

En general, para acceder a las imágenes y poder visualizarlas, crearlas o modificarlas son necesarios programas especiales denominados editores de imágenes. Existen toda una serie de formatos de amplia difusión que permiten el intercambio de imágenes entre distintos editores, mientras otros son propios de algún editor en concreto y no pueden ser visualizados por otros. Entre los formatos transferibles cabe destacar el TIFF (*Tagged Image File Format*), o el EPS (*Encapsulated PostScript*) pensado para la inserción de imágenes en documentos PostScript.

4.3. Audio

Menos extendidos que los ficheros de texto o imágenes podemos encontrar los ficheros de audio. El formato para el almacenamiento de audio más extendido en el mundo de la informática es *WAVE File Format* [.wav], dado que es el formato por defecto de Windows.

No obstante, la aparición de Internet y la posibilidad de intercambiar canciones por este medio ha dado lugar a un nuevo formato ampliamente extendido. Se trata del formato MP3. Este formato mantiene la calidad digital del sonido de los CDs originales pero permite su almacenamiento en un espacio mucho más reducido. Así, mientras una canción en formato WAV puede ocupar unos 40 Mbytes, la misma canción almacenada en formato MP3 suele ocupar entre 3 y 4 Mbytes. Esto permite almacenar más de 130 canciones en un único CD. El reducido tamaño de los ficheros de audio en este nuevo formato, hace que sea el formato imperante para el rápido intercambio de canciones a través de Internet y para el almacenamiento de canciones en pequeños reproductores portátiles que pueden almacenar varias horas de música. En Internet es posible encontrar multitud de páginas y repositorios donde es posible obtener, de una forma más o menos legal, una gran cantidad de canciones en formato mp3. Asimismo, existen programas especializados en la creación de grandes repositorios para el intercambio de este tipo de ficheros. El más conocido de este tipo de programas es *Napster* (www.napster.com).

Existen muchos programas de ordenador que permiten escuchar música en formato mp3. El más conocido de ellos quizás sea *Winamp*. Además, también existen programas que permiten la traducción de las canciones incluidas en CDs de música (CD-A) a este formato.

4.4. Vídeo

En informática se manejan un número relativamente reducido de formatos de vídeo. Los formatos más comunes son los siguientes:

MPEG (*Moving Pictures Experts Groups*) - [.mpg]. Es el formato más difundido para las pequeñas películas en Internet, lo cual significa que existen una gran cantidad de programas para reproducir este formato en los tres sistemas operativos más extendidos (Windows, MacOS y UNIX). Usa un esquema de compresión para reducir el tamaño de los ficheros.

AVI (*Audio Video Interleave*) - [.avi]. Es un formato de vídeo definido por Microsoft y es el más común en el mundo del PC y de Windows. Pueden reproducirse por ejemplo con el programa *Windows Media Player*.

QuickTime [.mov]. Es un formato de vídeo digital desarrollado por Apple y puede considerarse el equivalente del formato AVI en el mundo de los Macintosh, aunque existen reproductores de este tipo de ficheros para distintos sistemas operativos. Pueden reproducirse con el programa *QuickTime Player*.

Real media [.rm]. Se trata de un formato de vídeo especialmente útil para Internet, ya que permite iniciar la reproducción del mismo antes de estar completamente descargado. Para reproducir este tipo de vídeos debe usarse el programa *RealPlayer*.

4.5. Ficheros ejecutables

Los ficheros ejecutables son aquellos cuyo contenido indica al ordenador que debe realizar una serie de acciones más o menos complejas. Un fichero ejecutable puede simplemente sumar dos números introducidos por teclado y mostrar el resultado por pantalla, puede tratarse de un complejo programa de edición de video 3D para realizar animaciones, o de un procesador de textos. Las extensiones más comunes de los ficheros ejecutables en Windows son .exe y .com.

El formato de los ficheros ejecutables depende totalmente del sistema operativo. De este modo, un fichero ejecutable en Windows, no lo será en otros sistemas como UNIX o MacOS X y viceversa. Incluso suele ocurrir que un fichero ejecutable en una versión de un sistema operativo (p.e. Windows XP) no funcione en otra (p.e. Windows 95).

4.6. Ficheros comprimidos

Una de las características deseables en los ficheros es un tamaño reducido, ya que esto permite el almacenamiento de mucha información en los distintos soportes informáticos, y reduce el tiempo de transferencia de los mismos a través de Internet.

Algunos de los formatos de fichero que hemos revisado incorporan su propio sistema de compresión de la información para lograr ficheros reducidos. Por ejemplo: las imágenes JPEG, los ficheros de audio MP3 y prácticamente todos los formatos de vídeo. No

obstante existen numerosos programas que nos permiten comprimir cualquier tipo de fichero y se han definido diversos formatos de compresión de información.

Además, en muchas ocasiones interesa agrupar varios ficheros en uno sólo para facilitar su almacenamiento o transmisión. Muchos de los programas de compresión permiten agrupar varios ficheros como uno sólo y comprimir el resultado. De este modo, cuando se descomprime el fichero resultante, se vuelven a obtener los ficheros originales.

Los formatos más comunes son los siguientes:

Zip [*.zip*]. Se trata del formato de compresión más común en el mundo del Windows. Puede ser descomprimido con programas como *PKZip* o *WinZip*, aunque existen muchos más de libre distribución.

GZip [*.gz*]. Se trata de un formato de compresión muy común en UNIX, aunque también puede ser manejando con programas de compresión de Windows como *WinZip*.

Sit [*.sit*]. Es un formato de compresión muy utilizado en los ordenadores Macintosh. Se maneja fundamentalmente con el programa *Stuffit*.

En ocasiones podemos traernos un fichero comprimido y no disponer del programa de descompresión adecuado. Por ello, existen ficheros comprimidos capaces de autodescomprimirse. Este tipo de ficheros son ficheros con la extensión *.exe*, que al ser ejecutados se descomprimen.