

# 6. La edición digital

JOSÉ MANUEL BARRUECO



## Capítulo 6

<b>La edición digital</b> (José Manuel Barrueco)	285
La edición digital: Introducción	285
Desde lo analógico a lo digital: guía de digitalización de documentos	286
La planificación del proyecto de digitalización	287
Equipamiento	288
Selección de materiales	289
El escaneado	290
El OCR ( <i>Optical Character Recognition</i> )	291
Control de calidad	291
Metadatos	292
Presentación y publicación	293
Gestión	294
Desde el archivo digital al soporte digital	295
El diseño de la estructura del documento	296
Documentos de texto no estructurados	297
Documentos estructurados	301
Los formatos de imagen	307
Documentos multilingües	311
El diseño de la presentación	313
Hojas de estilo	313
Tipos de visualización	315
Uso de buscadores	317
La adición de valor: metadatos	318
Estándares de accesibilidad y usabilidad	320

## LA EDICIÓN DIGITAL: INTRODUCCIÓN

Durante los últimos diez años en el área de la edición científica, los contenidos digitales han experimentado un fuerte crecimiento, motivado por la generalización del uso y el aumento de la velocidad de acceso a Internet. Cualquier tipo de documento, que tradicionalmente se ha publicado en formato impreso, puede tener una representación digital y estar al alcance de una audiencia potencial de millones de usuarios con una inversión reducida.

En paralelo, el desarrollo de aplicaciones informáticas multimedia ha permitido la aparición de nuevas tipologías documentales como aplicaciones interactivas, animaciones o simulaciones en 3D. No obstante, este incremento no ha estado exento de problemas ni se ha dado por igual en todos los tipos documentales ni disciplinas científicas. Mientras en casos como las obras de referencia o las revistas científicas las versiones digitales están relegando a las impresas, el desarrollo de libros electrónicos sigue teniendo una aceptación muy limitada.

La edición científica de contenidos digitales tiene ventajas e inconvenientes. Entre los principales inconvenientes<sup>1</sup> tenemos:

- ❑ Control de los derechos de copia. Es prácticamente imposible evitar la copia de un documento electrónico.
- ❑ Falta de normalización. Existen multitud de formatos y estándares que están compitiendo para hacerse un hueco en el mercado de la edición digital.

Como puntos a favor cabría destacar:

- ❑ Portabilidad. Una vez que un documento está en formato digital es fácil reproducirlo en diferentes formas en función de las necesidades de la audiencia (Braille, impresión bajo demanda, contenido web, etc.).
- ❑ Renovación. El concepto de «obra» agotada podría desaparecer ya que aquellos libros agotados podrían volver a la vida si se digitalizan y distribuyen electrónicamente.
- ❑ Mejoras. Los documentos electrónicos tienen características nuevas no disponibles en el papel como, por ejemplo, los enlaces hipertextuales o las animaciones. Además, es mucho más barato producir documentos electrónicos en color que documentos impresos en color.
- ❑ Difusión. Si se colocan en la red la audiencia potencial de los documentos es mayor que en cualquier documento impreso.

En la actualidad estamos en una época de transición del medio impreso al digital que llevará aún cierto tiempo completar. Durante la misma, ambas tecnologías coexistirán en un entorno híbrido. En general, cada servicio de publicaciones tendrá que tomar sus propias decisiones y definir la política a seguir en cuanto a edición digital. En cualquier caso, está claro que la opción que se plantea no es publicar en digital o no, sino que la verdadera cuestión será «cuándo» comenzar. La opción de la edición digital, a medio plazo, se convertirá en inevitable. Cuando esto suceda, será útil realizar un acercamiento pragmático y cuidadosamente organizado.

El presente capítulo está destinado a ayudar a localizar todos los recursos técnicos necesarios para comenzar a publicar en formato digital. Para ello se ofrecen una serie de recomendaciones a seguir para la creación de documentos digitales. Se han diferenciado dos tipos de documentos digitales en función de su origen. En primer lugar tratamos las representaciones digitales de documentos impresos, esto es, los documentos digitalizados. En segundo lugar, los que han sido diseñados y creados en formato digital desde su origen.

## DESDE LO ANALÓGICO A LO DIGITAL: GUÍA DE DIGITALIZACIÓN DE DOCUMENTOS

La digitalización es aquel proceso que comprende todos los pasos necesarios para hacer disponibles en formato digital documentos que se encuentran en formato impreso o en cualquier otro formato analógico (audio, vídeo, etc.). En esta guía nos centramos exclusivamente en documentos impresos. A grandes rasgos la cadena de digitalización se concreta en los siguientes pasos: redacción de un plan del proyecto, selección del equipamiento necesario, selección de los materiales a digitalizar, escaneado de los mismos y opcionalmente conversión a texto a través de un proceso de OCR (*Optical Characters Recognition*), control de calidad de los resultados, adición de metadatos, presentación y publicación.

La digitalización del patrimonio documental es una de las prioridades de la política cultural de la Unión Europea en relación con la sociedad de la información<sup>2</sup>. Esta preocupación por poner a disposición del público versiones electrónicas de documentos, tanto históricos (archivos) como clásicos o fondo antiguo (bibliotecas) es compartida también por gobiernos a nivel internacional así como por universidades y otros agentes culturales. Esto hace que se hayan realizado múltiples guías de buenas prácticas y recomendaciones sobre digitalización. Una lista de las mismas ha sido recopilada por el proyecto Minerva<sup>3</sup> de la Unión Europea.

Para la elaboración de este capítulo hemos utilizado como base algunas de estas obras. Entre ellas:

- ❑ *Good Practices Handbook*, realizado por Minerva Europe<sup>4</sup>. Este documento es el resultado del grupo de trabajo sobre buenas prácticas dentro del proyecto Minerva financiado por la Unión Europea. Presenta un manual práctico para el establecimiento, ejecución y gestión de proyectos de digitalización. No es un documento técnico dado que todas las cuestiones técnicas se han especificado en un informe titulado *Technical Guidelines for Digital Cultural Content Creation Programmes*<sup>5</sup>.
- ❑ Las especificaciones de la IFLA (*International Federation of Libraries Associations*)<sup>6</sup>. Elaboradas por un grupo de trabajo conjunto de la IFLA y el Consejo Internacional de Archivos (CIA) y financiado por la UNESCO. Identifican y discuten los aspectos claves implicados en la conceptualización, planificación e implementación de cualquier proyecto de digitalización.
- ❑ La guía de digitalización de la biblioteca digital de Georgia (USA)<sup>7</sup> está centrada en la digitalización de documentos históricos, fotografías, etc. Tiene un componente local muy importante.
- ❑ La guía de digitalización del proyecto COINE (*Cultural Objects in Networked Environments*) de la Unión Europea<sup>8</sup>. Centrada en aspectos técnicos y con recomen-

daciones adicionales sobre la digitalización de material de audio y vídeo que en este documento no hemos incluido.

❑ *Moving Theory into Practice: Digital Imaging Tutorial*<sup>9</sup>.

## LA PLANIFICACIÓN DEL PROYECTO DE DIGITALIZACIÓN

Digitalizar es caro debido a los recursos humanos, técnicos y económicos necesarios, por lo que la digitalización debería ser el último recurso para publicar contenidos en la Red.

Cada proyecto de digitalización tendrá unos motivos concretos para llevarse a cabo. En nuestro campo de actuación, los servicios de publicaciones de las universidades españolas, la principal motivación será proporcionar acceso al material impreso agotado. No es recomendable utilizar de forma indiscriminada la digitalización para colocar en la Red documentos que han sido diseñados y producidos para ser distribuidos en formato impreso. Los gastos necesarios para la digitalización de este material no compensarían los beneficios obtenidos, a no ser que la audiencia potencial o el interés académico lo justificaran. En estos momentos no se justifica editar en formato impreso exclusivamente, sino que se debe tener en cuenta que cualquier obra puede acabar existiendo en formato digital en el futuro por lo que nos ahorrará tiempo y esfuerzo diseñar los nuevos documentos con esa posibilidad y en formatos que nos permitan obtener productos impresos o digitales indistintamente a partir de un documento fuente. Tampoco es recomendable utilizar la digitalización por motivos de conservación. No es función de los servicios de publicaciones garantizar la permanencia de los documentos en el tiempo. Esta función la cumplen las bibliotecas.

No obstante, cualesquiera que sean las motivaciones para llevar a cabo un proyecto de digitalización en un servicio de publicaciones, la redacción de un plan debe ser el primer paso a seguir. El tiempo que se gaste en esta etapa se plasmará en una más fácil gestión y ejecución del proyecto en el futuro. Según las recomendaciones del proyecto Minerva, en el plan deben ser respondidas al menos las siguientes cuestiones:

- ❑ ¿Qué se va a digitalizar?
- ❑ ¿Quién lo va a hacer?
- ❑ ¿Dónde se va a realizar?
- ❑ ¿Cuándo tendrá lugar?
- ❑ ¿Cómo se va a realizar?

En general, un proyecto de digitalización debería tener claramente especificados los objetivos que se pretenden alcanzar. Igualmente, se debería recopilar información sobre experiencias similares llevadas a cabo por otras instituciones, de forma que se pueda aprender de sus aciertos y errores.

El manual del proyecto Minerva hace las siguientes sugerencias:

- ❑ El proyecto debe tener unos objetivos concretos que deben estar documentados.
- ❑ Los objetivos deben ser realistas en relación a los recursos materiales, financieros y humanos disponibles.
- ❑ Todos los pasos del proyecto deben ser validados para asegurar que el trabajo que se está llevando a cabo contribuye a alcanzar los objetivos.
- ❑ Se deben especificar los beneficios que se prevé que el proyecto proporcione a la institución.

Igual que sucede en otras áreas los recursos humanos son fundamentales para el buen funcionamiento del proyecto. Antes de que éste comience es importante asegurarse de que estén disponibles los recursos humanos necesarios. Muchas instituciones no cuentan con personal con tiempo disponible para llevar a cabo la digitalización además de sus tareas habituales. Se debería dotar con el personal adecuado, con unos conocimientos apropiados, así como prever un plan de formación para proporcionar una experiencia adicional en caso de que el proyecto así lo requiera. Se sugiere que:

- Se asegure el personal suficiente para llevar a cabo el proyecto de forma satisfactoria.
- Se asigne un personal determinado a cada tarea que se especifique en el plan del proyecto.
- Se identifiquen las necesidades de formación del personal.
- Se proporcione formación en el uso del hardware y software seleccionados para llevar a cabo el proyecto antes de que éste comience.

## EQUIPAMIENTO

Se debe preparar un entorno apropiado para la digitalización. Los elementos de ese entorno incluyen hardware para el proceso de digitalización propiamente dicho, una infraestructura informática a la cual conectar el hardware, software para la captura y procesamiento de las imágenes, para incluir metadatos y para llevar a cabo el control de calidad de los resultados.

El hardware básico serán equipos para la captura de imágenes (cámaras digitales o escáneres para documentos impresos). Si bien es difícil dar indicaciones concretas sobre los equipos a utilizar dado que el mercado está cambiando constantemente, se sugiere de forma general que:

- El hardware apropiado debe estar instalado y probado antes de que el proyecto comience. Se deben hacer las pruebas necesarias con documentos tipo para valorar la calidad de los dispositivos de captura de imágenes.
- En caso de utilizar un escáner debería ser al menos del tamaño del mayor documento que se vaya a escanear.
- La captura de imágenes debería realizarse a una resolución razonable. Si bien esto se traducirá en ficheros muy grandes, se podrían obtener ficheros más pequeños a partir de esta copia master para propósitos como la distribución a través de Internet. Por el contrario, el proceso inverso no sería posible ya que una imagen de alta calidad no puede nunca obtenerse a partir de una de poca resolución.
- La definición de «resolución razonable» dependerá de la naturaleza del material que se está escaneando y de los usos a los que se va a dedicar. Por ejemplo, si las imágenes escaneadas se van a utilizar únicamente como miniaturas, esto puede conseguirse con una baja resolución. Igualmente, la resolución debe capturar los detalles más significativos del documento. Si el escaneado a una alta resolución no aportase ninguna información adicional que si se hiciera a una resolución menor, la primera opción es difícil de justificar.
- La imagen capturada debería almacenarse en un fichero con formato «sin pérdida», por ejemplo, no comprimido. Normalmente, para las copias master se utiliza un formato TIFF (*Tagged Image File Format*).
- Se debería contar con un sistema informático con suficiente capacidad de almacenamiento.
- De este sistema deberían hacerse copias de seguridad de forma regular.

Una vez que obtenemos la imagen digital del documento, el fichero resultante requerirá algún tipo de procesamiento utilizando un software de tratamiento de imágenes antes de poder ser usado. Por ejemplo, un caso frecuente es que el fichero sea demasiado grande y se requiera alguna versión comprimida (por ejemplo, una miniatura de una portada para ser utilizada en una página web). El software seleccionado debe adecuarse a los objetivos y a la tipología documental que se está tratando. Si bien los equipos de digitalización vienen con software incluido, normalmente éste no tiene suficiente potencia ni flexibilidad.

El proyecto debe adquirir el más apropiado y potente software que pueda afrontar dado su presupuesto. La evaluación de diferentes paquetes puede salvar mucho tiempo y trabajo a largo plazo. Como un mínimo absoluto, el software elegido debe permitir:

- Abrir ficheros de imágenes de gran tamaño.
- Modificar la resolución y la profundidad de color.
- Salvar versiones diferentes, con diferentes tamaños de fichero.
- Seleccionar y copiar una parte de una imagen y guardar ésta como un fichero diferente.
- Exportar imágenes en diferentes formatos incluidos los estándares del Web: JPEG y GIF.

En el caso de que la digitalización vaya acompañada de un OCR (*Optical Character Recognition*), la elección del software apropiado es crítica. Todo OCR lleva implícita una cierta cantidad de trabajo manual de edición y corrección; la forma en que esto esté soportado por el software puede tener un efecto decisivo sobre el tiempo y el esfuerzo requerido por el proyecto. Los mejores paquetes de OCR deben permitir corregir y visualizar en una sola pantalla, sugerir posibles correcciones para las palabras mal interpretadas, ofrecer soporte de texto en varias columnas, etc.

## SELECCIÓN DE MATERIALES

La selección del material para ser digitalizado es una decisión importante en cualquier proyecto. Los criterios diferirán dependiendo de los objetivos establecidos y estarán limitados por factores adicionales, como pueden ser restricciones legales, políticas institucionales, dificultad técnica de la digitalización, copias digitales de los documentos ya existentes, etc. Se sugiere establecer criterios para la selección del material y reflejarlos en el plan del proyecto. Al menos los siguientes puntos deben tenerse en cuenta:

- Demanda prevista de los documentos digitalizados.
  - Contenido. De una forma u otra la selección del material estará orientada por el contenido. La pregunta que se debe responder sería: ¿el contenido intelectual de este documento justifica los esfuerzos, costes y recursos que se van a necesitar? Cada institución deberá tener sus propias definiciones de valor basadas en los objetivos que se pretenden obtener. Durante los últimos años se ha popularizado la creación de bibliotecas digitales o colecciones virtuales. Los servicios de publicaciones pueden crear este tipo de colecciones con los materiales impresos, pero en la actualidad es mucho más rentable planear que cualquier documento acabará existiendo en formato electrónico de una forma u otra, por lo tanto, es mejor tener en cuenta este hecho desde el momento mismo de su creación.
  - Derechos de copia. Este es un aspecto crucial en la selección. Se debe asegurar

que la institución ostenta los derechos de copia de los documentos que se deben digitalizar o alternativamente que tiene el permiso escrito de los autores de los mismos.

- Acceso al material que de otra forma no estaría disponible o con disponibilidad limitada.
  - Ampliar y facilitar el acceso a un material popular.
  - Coste de la digitalización.
  - Disponibilidad de otras versiones digitales ya existentes.
  - Adecuación de los documentos fuente para su visualización en línea.
- Los criterios de selección deberían ser explícitos y discutidos por todas las partes involucradas en el proceso.
  - Los criterios deberán estar completamente documentados, de forma que las razones para digitalizar o no estén claras a lo largo de todo el proceso.

## EL ESCANEADO

El escaneado es el proceso a través del cual obtenemos una imagen digital de un documento impreso. Los documentos fuente son transformados en imágenes digitales utilizando un escáner o una cámara digital. A través del proceso de captura de imagen los documentos son leídos o escaneados con una resolución y profundidad de bit predefinida. Los ficheros resultantes, conteniendo los bits por cada píxel, son entonces formateados y enmarcados en una forma que haga fácil para un sistema informático almacenarlos y recuperarlos. A partir de esos ficheros el sistema puede reproducir representaciones analógicas para mostrarlas en pantalla o imprimirlas. Dado que los ficheros que contienen imágenes de alta resolución son muy grandes, puede ser necesario reducir su tamaño mediante técnicas de compresión.

Cuando un documento ha sido escaneado, todos los datos son convertidos a un formato de fichero particular para ser almacenado. Hay varios formatos de imágenes en el mercado que son ampliamente utilizados.

La calidad de la imagen obtenida en este proceso determinará la calidad final del producto. De esta forma se recomienda:

- Asegurarse de que el soporte de cristal del escáner está completamente limpio en todo momento.
- Establecer una convención para nombrar los ficheros obtenidos, por ejemplo, usando el sistema de identificación existente o dándoles nombres significativos:
  - El nombre del fichero debería permitir la asociación entre el fichero y el documento fuente.
  - Si el documento tiene un identificador o código asignado (por ejemplo un ISBN o ISSN) es recomendable utilizarlo.
  - Utilizar solamente minúsculas.
  - Por compatibilidad con diferentes sistemas se recomienda utilizar un nombre de ocho caracteres con una extensión de tres siempre que sea posible.
  - Si es necesario establecer un sistema nuevo, una buena opción sería utilizar una o dos letras seguidas de al menos tres números, de la forma: gr001.tif, gr002.tif, etc.
  - Es conveniente mantener siempre el mismo número de caracteres en el nombre el fichero y no variar la longitud. Esto hace el procesamiento automático mucho más fácil.

- ❑ Escanear con la mayor resolución y profundidad de bit que sea posible según los objetivos que se persigan, las limitaciones de los equipos y las condiciones de almacenamiento de los datos.
- ❑ Realizar copias de seguridad del sistema en que se almacenen los datos con una periodicidad diaria.
- ❑ Es importante el control de la calidad de la imagen y de los metadatos.

## **EL OCR (*OPTICAL CHARACTER RECOGNITION*)**

Los documentos escaneados pueden presentarse al lector en forma de imágenes exclusivamente o bien convertirse en texto como parte del proceso de digitalización. Esto se hace a través de un proceso de OCR. El texto resultante podrá ser presentado en una página web, almacenado como XML/SGML o incluido en una base de datos, de tal forma que puedan realizarse búsquedas, conversión de formatos o cualquier otro tipo de procesamiento. Si se decide utilizar OCR se recomienda:

- ❑ Tener en cuenta que el OCR funciona mejor con documentos que se encuentren en buenas condiciones físicas y de conservación.
- ❑ Evaluar múltiples paquetes de software antes de seleccionar un producto en particular. En muchos casos, los programas de OCR están incluidos con el escáner, pero suelen ser de poca calidad. Es mejor buscar paquetes vendidos por separado.
- ❑ Un software que proporcione una interfaz amigable para llevar a cabo la identificación y corrección manual de errores, ambigüedades y partes del texto que no se hayan podido procesar.
- ❑ Se debería evaluar la existencia o no de diccionarios del lenguaje en que se encuentren los documentos fuente como parte del software de OCR.

Entre los productos disponibles en el mercado, los más relevantes serían: OmniPage, TextBridge y Adobe Capture.

## **CONTROL DE CALIDAD**

El control de calidad es un componente importante de todas y cada una de las etapas de un programa de digitalización. Sin esta actividad es imposible garantizar la integridad y consistencia de los ficheros de imágenes.

Aunque el control de calidad es un factor crucial para asegurar los mejores resultados, no hay una forma normalizada de asegurar una cierta calidad de imagen en el momento de la captura. Las condiciones para la calidad de las imágenes digitales pueden ser descritas con tres pasos:

- ❑ Identificación del producto final esperado y los objetivos de producción. Decidir qué será producido y evaluado.
- ❑ Establecimiento de estándares. Definir los niveles de calidad de las imágenes basados tanto en los atributos de los documentos originales como en la capacidad del equipo digitalizador que se esté usando.
- ❑ Establecimiento de unas pautas de referencia. Decidir con qué se va a comparar el resultado de la digitalización.

Dependiendo de estas decisiones el siguiente paso será determinar qué deberían incluir los programas de control: el total de la colección o una muestra. En la mayor parte de los casos es suficiente con establecer un plan de muestreo cubriendo, por ejemplo, el 10% de

todas las imágenes producidas por cada escáner durante un cierto período de tiempo (día, semana...). Si en un porcentaje significativo de las imágenes examinadas se han encontrado errores, entonces la colección completa deberá someterse a examen.

Por otra parte hay que decidir el método de evaluación de las imágenes digitales utilizando bien la pantalla bien las versiones impresas. En cualquier caso habrá que tener en cuenta ciertas limitaciones técnicas que pueden afectar al proceso de evaluación como, por ejemplo, disponer de impresoras en color de calidad. Si la evaluación se hace sobre pantalla se recomienda:

- Visualizar las imágenes a una escala de 1:1.
- Usar valores de referencia para evaluar la reproducción de colores o escala de grises.
- Utilizar valores de referencia de resolución e histogramas para evaluar la resolución espacial y la reproducción tonal.

En caso de evaluar versiones impresas se recomienda comparar las imágenes impresas con los documentos originales para comprobar que se cumplen los objetivos.

## **METADATOS**

Una vez que el fondo de documentos escaneados comience a crecer, se hará imprescindible alguna herramienta de búsqueda y localización de información que nos permita acceder de forma rápida a los contenidos. Para describir las imágenes digitales son necesarios los metadatos, esto es, datos estructurados sobre los propios datos. Su uso está relacionado de cerca con las reglas para la descripción y catalogación de publicaciones impresas. La diferencia reside en que en el mundo digital se necesitan categorías adicionales de metadatos para soportar la navegación y la gestión de los ficheros.

Los metadatos que describen imágenes digitales pueden contener información de diferentes tipos. La Biblioteca del Congreso de los Estados Unidos identificó tres categorías:

- Descriptivos: para la descripción e identificación de fuentes de información.
- Estructurales: para facilitar la navegación entre las distintas partes de un documento y la presentación del mismo.
- Administrativos: para gestión y proceso.

El nivel de metadatos descriptivos siempre determina el nivel de posibilidades de recuperación de la información. Es de crucial importancia en el momento de planificar el proyecto, decidir hasta qué nivel de profundidad se podrían buscar las imágenes.

Se debe tener en cuenta que la creación e implementación de metadatos son procesos que consumen muchos recursos. Habrá que hacer un balance de costes y beneficios, teniendo en consideración las necesidades de los usuarios actuales y futuros. Los requerimientos de metadatos deberían estar íntimamente unidos a las funciones que deben soportar: gestión de derechos, preservación a largo plazo, localización de información, etc.

Habrà que considerar los siguientes puntos:

- El uso de metadatos apropiados es crucial para permitir la búsqueda y recuperación de material de una colección digital. Esto es mucho más importante cuando se plantean búsquedas en más de una colección.

- ❑ Ya existen múltiples modelos de metadatos. Cada proyecto deberá elegir el más adecuado a sus objetivos, evitando en la medida de lo posible crear uno nuevo.
- ❑ Entre los esquemas de metadatos más utilizados en la actualidad está el *Dublin Core*. Este formato fue creado a finales de los años noventa a partir de una iniciativa de la OCLC (*Online Computer Library Center*) y la NCSA (*National Center for Supercomputing Applications*). Esta iniciativa presentaba un formato compuesto por 15 elementos de metadatos descriptivos. Los objetivos de este formato eran ser simple, internacional y que funcionara en varios sectores. El *Dublin Core* es uno de los estándares de metadatos más ampliamente aceptados en el mundo.
- ❑ Aunque algunos metadatos son estáticos (fecha de creación, resolución utilizada, etc.) algunos elementos pueden continuar evolucionando y requiriendo actualizaciones y mantenimiento.
- ❑ La creación y gestión de los metadatos puede ser manual (creación de un registro en *Dublin Core*) y automatizada (generar un índice a partir del resultado del OCR).
- ❑ Los metadatos pueden ser internos (cabeceras de ficheros TIFF) o externos (índices y bases de datos asociados). El factor clave para tomar una decisión es evaluar si su localización soporta la funcionalidad y la gestión de los recursos. Por ejemplo, las cabeceras de los ficheros TIFF contienen metadatos pero estos se perderán si son convertidos a otros formatos como JPEG.

## PRESENTACIÓN Y PUBLICACIÓN

Hasta este momento se ha creado y almacenado una copia máster de los materiales. Hemos identificado un esquema de metadatos apropiado, y se han creado los metadatos adecuados para cada documento.

La preparación para la publicación implica un cierto procesamiento del material creado. Normalmente, por publicación entendemos alguna forma de visualización de los contenidos a través de Internet y procesamiento significa reducir el tamaño de los ficheros, ajustar la calidad y elegir los formatos de descarga más apropiados para adaptarse a las características de Internet.

Los ficheros TIFF, creados durante el proceso de escaneado, suelen ser demasiado grandes (varios Mb) lo que los hace inapropiados para su publicación en Internet debido al elevado tiempo que el usuario final necesitaría para completar una descarga.

Sugerencias prácticas:

- ❑ Crear versiones apropiadas para ser distribuidas en Internet a partir del máster. Como mínimo, debe haber una versión para descargar de Internet.
- ❑ Las versiones para descarga se crean abriendo los ficheros TIFF en un paquete de procesamiento de imágenes y exportándolos a formato JPEG, PNG, etc. Aunque existen programas gratuitos de tratamiento de imágenes como GIMP, Paint o Paintshop, dado el volumen de documentos que vamos a tratar es conveniente invertir en productos comerciales como Photoshop, etc.
- ❑ Normalmente, la profundidad del color puede reducirse a 256 colores, a no ser que se aprecie una pérdida de calidad, lo cual exigirá usar una resolución mayor.
- ❑ Una imagen creada a 72 ppi se visualizará a aproximadamente su tamaño original en muchos monitores. Esto hace la resolución de 72 ppi una resolución razonable para muchas imágenes que han de visualizarse en pantalla.

- ❑ Elegir el formato de los ficheros, la resolución de color y la resolución de píxel implica decidir cuál es una aceptable calidad para las imágenes. Se debe encontrar un balance entre calidad y tamaño de los ficheros.
- ❑ En general, el número total de ficheros de imágenes en una página web no debería exceder de los 100 kb. Aunque imágenes mayores puedan publicarse, se sugiere acceder a ellas a través de un enlace en la página, más que incluirlas en la misma.

Muchos proyectos de digitalización llevan a la creación de recursos accesibles en línea, normalmente a través de páginas web. Van desde los sitios más simples, con páginas HTML estáticas, hasta los portales más elaborados acompañados de bases de datos, etc. Existe multitud de documentación sobre la creación y el mantenimiento de páginas web, por lo que no consideramos oportuno profundizar en el tema en este capítulo.

## GESTIÓN

El éxito de cualquier proyecto, incluyendo los de digitalización, está influenciado por la gestión del mismo. Vamos a proporcionar ahora una serie de recomendaciones dirigidas a cuatro aspectos fundamentales del proceso: gestión propiamente dicha, recursos humanos, externalización de procesos y costes.

- ❑ Dado que el proyecto de digitalización implicará la digitalización de cientos de documentos, es importante establecer un diagrama de flujos para facilitar el trabajo del equipo.
- ❑ Establecer y documentar cada uno de los pasos a través de los cuales debe pasar un ítem.
- ❑ Desarrollar una base de conocimiento que permita rastrear un determinado ítem a través de todo el proceso de digitalización, así como permitir revisar el estado del proyecto en cualquier momento. Esta base de conocimiento puede tomar la forma de una base de datos, una hoja de cálculo o, simplemente, una colección de documentos e informes. Lo importante no es el formato sino el proceso que asegure el registro de las acciones que se están llevando a cabo.
- ❑ El nombre, identificador y otra información relevante de cada documento debería ser introducida en la base de conocimiento desde el mismo momento que ha sido seleccionado para ser digitalizado.
- ❑ Establecer un manual de procedimientos.
- ❑ Siempre que sea posible incluir, al menos, una persona con formación adecuada en el área de las nuevas tecnologías de la información.
- ❑ Determinar, antes de empezar el proyecto y registrarlo en la fase de planificación, los conocimientos que tiene el personal que va a trabajar en el mismo sobre tecnologías de la información e indentificar las necesidades de formación.
- ❑ A veces es apropiado para el proyecto de digitalización contar con los servicios de terceras partes. Los servicios que son externalizados habitualmente incluyen la digitalización en sí misma, la gestión del proyecto, el desarrollo de software, etc.
- ❑ Las relaciones con las terceras partes deberían estar claramente reguladas mediante un contrato escrito. Se debería acordar una completa especificación de los servicios o productos que han de ser proporcionados antes de que el trabajo comience.
- ❑ El trabajo desarrollado debería ser revisado regularmente para asegurar que lo que se está suministrando es en realidad lo que el proyecto quiere o necesita.
- ❑ Si bien la utilización de terceras partes puede ser conveniente, hay que tener en cuenta que toda la experiencia ganada durante la ejecución del trabajo externalizado será perdido para nuestra institución al final del proyecto.

- ❑ Además de la externalización de tareas, se puede recurrir a la colaboración con otras instituciones similares. Un ejemplo sería la participación en proyectos de investigación financiados por la Unión Europea, donde se podría contar con socios de varios países. En este tipo de colaboración, cada socio debería tener un papel claro en el proyecto. Si, en algún caso, el papel no está claro hay que revisar si el socio es de verdad necesario.
- ❑ Asegurarse de que todos los socios están al corriente y han soportado sus papeles y responsabilidades dentro del proyecto. Refrescar este conocimiento regularmente.
- ❑ Establecer un modo de comunicación entre todos los socios y asegurarse de que cada uno recibe la información que está destinada a él. El correo electrónico y las listas de distribución son ideales para este propósito.

## DESDE EL ARCHIVO DIGITAL AL SOPORTE DIGITAL

En este apartado vamos a tratar la creación de documentos digitales propiamente dichos, es decir, aquellos que existen en formato digital desde su origen. Un documento digital puede diferenciarse de uno tradicional, además de por su formato, en al menos tres aspectos<sup>10</sup>:

- ❑ Permanencia: los documentos digitales son por principio fluidos, cambiantes y pueden evolucionar constantemente. Si bien los analógicos también pueden serlo, por ejemplo las diferentes ediciones de una obra, en el caso de los digitales no son fijados por el medio de difusión. De hecho un documento digital puede no existir nunca en un formato tangible.
- ❑ Apariencia: en los documentos tradicionales su importancia se refleja en la apariencia física de los mismos. Así, el tamaño y la presentación de un libro lo distinguen como más importante que un informe técnico fotocopiado. En el caso de los documentos digitales se pueden usar las mismas técnicas y obtener los mismos resultados independientemente de la importancia del documento.
- ❑ Tamaño: el documento digital no está limitado por cuestiones de espacio.

## EL DISEÑO DE LA ESTRUCTURA DEL DOCUMENTO

En general cuando se habla de diseño de documentos se tiene en mente un diseño basado en la presentación o apariencia final del documento. Una idea que se ve potenciada por los editores de texto modernos que se basan en WYSIWYG.

El desarrollo de software para la preparación de documentos ha tenido una importancia notable en la industria informática ya desde los años sesenta. Todos los sistemas desarrollados han seguido básicamente dos paradigmas. El primero especifica el documento en términos de su apariencia esperada cuando sea impreso. Este diseño procede de la idea, acarreada desde el mundo impreso, de que los documentos son fijos y estables, tanto si van a ser visualizados en papel como en una pantalla y de que serán vistos por el mismo tipo de dispositivo de visualización por todos los lectores. Una representación de esta tendencia son los editores WYSIWYG (*what you see is what you get*), es decir, lo que se ve en la pantalla es exactamente lo que obtendrá el lector si imprime el documento o lo visualiza en el mismo editor que usó el escritor. El usuario debe centrarse en la apariencia del texto que está escribiendo, una apariencia que pretenderá ser fija y estable a lo largo de

la vida del documento. Cualquier transformación o cambio de formato exigirá enormes esfuerzos para revisar precisamente el formato de presentación.

Con la introducción por parte de IBM del *Generalized Markup Language* (GML) apareció un segundo tipo de representación basada en la estructura lógica de los documentos. Los documentos tienen una estructura jerárquica intrínseca<sup>11</sup>: los capítulos se componen de secciones, las secciones de subsecciones y así sucesivamente hasta llegar a las frases, palabras y letras. Si las secciones de un documento están marcadas de alguna forma entonces será posible operaciones de procesamiento como mostrar solamente los títulos de las secciones, identificar el resumen, etc. Incluso se podrían mover componentes de un documento de una localización a otra del mismo sin necesidad de reajustar cada uno de los títulos.

Los mecanismos para marcar la estructura de un documento han variado considerablemente y van desde los más permisivos como LaTeX, a los más estrictos, como SGML.

Los sistemas de documentos estructurados tienen ventajas e inconvenientes. En general son más robustos porque incrementan la vida del documento y de sus componentes. Además, dado que el marcado sigue una norma y está separado de la aplicación particular que se esté dando al documento, su mantenimiento se mejora y se hace posible la reutilización de los componentes en diferentes documentos. En el lado opuesto, su preparación requiere un mayor esfuerzo en su especificación inicial. Por ello es percibida por los autores como más difícil y compleja porque parece contradecir el concepto de manipulación directa del documento. No obstante tomando el conjunto de sus ventajas contrarresta fuertemente el de sus debilidades, especialmente cuando hay una necesidad de robustez, reutilización y exactitud.

Generalmente hay dos papeles implicados en la creación de un documento: especificación de su contenido y especificación de su apariencia<sup>12</sup>. La tarea de especificación del contenido debe ser obra del autor, mientras que la de apariencia debe ser la obra de un diseñador de estilo. En muchos casos una única persona puede cumplir los dos papeles, pero normalmente se puede ganar eficacia separándolas ya que las técnicas necesarias en ambos casos son diferentes.

Se ha definido una gran variedad de representaciones lógicas que intentan establecer un equilibrio entre estructuras complejas y acotadas por un lado, y estructuras simples y no acotadas por otro. Podríamos establecer tres niveles básicos de complejidad. En el punto más simple estarían sistemas como Microsoft Word donde la apariencia de la estructura del documento formateado no está directamente reflejada en la codificación interna del documento. El documento está definido como una secuencia de trozos de contenido para cada uno de los cuales se define una transformación particular que define su apariencia física. En segundo lugar, en sistemas como TeX/LaTeX que utiliza un lenguaje de marcas, se puede especificar una estructura básica y predefinida dependiendo del tipo de documento que tratemos. Finalmente, en último lugar tendríamos SGML donde las representaciones jerárquicas han sido llevadas a su punto más complejo.

A continuación vamos a describir de forma somera algunas de las distintas representaciones de documentos que existen, separando aquellas que tienen en cuenta la estructura de los documentos de aquellas que no la tienen.

Los editores, a menudo, se plantean la opción entre mantener las cosas simples (ahorrando tiempo y dinero) o hacerlas más ricas (pero más complicadas y caras)<sup>13</sup>. Los editores que se

plantean sólo la creación de documentos impresos tienden a optar por la simplicidad y utilizar el formato con el que más a gusto trabajan. Por otro lado, los editores que necesitan presentar versiones electrónicas e impresas al mismo tiempo, deberían optar por trabajar con una combinación de SGML o XML más XSLT.

## DOCUMENTOS DE TEXTO NO ESTRUCTURADOS

❑ **Texto plano.** El documento no estructurado más básico es aquel que no tiene absolutamente ningún tipo de formato, solo contiene una sucesión de caracteres y en algunos casos saltos de línea para hacerlo más legible. Todos los procesadores de texto tienen la opción de grabar un documento en texto plano.

En algunas ocasiones se habla de texto plano como sinónimo de ASCII lo cual es incorrecto. ASCII es el acrónimo de *American Standard Code for Information Interchange*. Se trata de un código para representar en formato digital caracteres (letras, números, etc.). Al texto codificado usando ASCII se le denomina texto ASCII. Por otro lado, texto plano se refiere a texto sin ningún tipo de formato, independientemente del código utilizado para representarlo. Se debe tener en cuenta que cuando alguien se refiere a texto ASCII de lo que realmente está hablando es de texto plano.

El uso más frecuente del texto plano es para la documentación de software informático. En general, es una buena opción cuando hay una incertidumbre total sobre las capacidades tecnológicas disponibles por la audiencia a la que va destinado el documento, ya que cualquier editor o procesador de texto es capaz de abrir un fichero en texto plano. No obstante en estos momentos y a no ser en casos muy extremos, es difícil justificar la utilización del texto plano como formato de publicación de documentos.

Entre los ejemplos de instituciones que utilizan la edición en texto plano podemos señalar la *Internet Engineering Task Force*<sup>14</sup>. Esta organización se encarga de la elaboración de los estándares que regulan la arquitectura de Internet. Tales estándares, que se publican con el nombre de RFC (*Request For Comments*), están elaborados utilizando exclusivamente texto plano con saltos de línea.

❑ **HTML.** Son las siglas de *Hypertext Markup Language*. Se trata, por lo tanto, de un lenguaje de marcas utilizado para representar tanto la estructura como la apariencia de un documento. Es el lenguaje utilizado para realizar las páginas web.

Lo hemos colocado en el apartado de formatos no estructurados porque sus capacidades para representar la estructura son muy limitadas y no pasan de describir aspectos generales como establecer cuál es el título o los párrafos de un documento. El principal desarrollo del HTML se ha llevado a cabo en el área de la presentación o visualización de contenidos debido al empuje de la industria, la publicidad y el comercio electrónico que necesitan de aplicaciones con un impacto visual importante.

Los documentos en HTML son documentos en texto plano a los que se les han añadido marcas codificadas usando los caracteres < y >. Tales ficheros se pueden crear con cualquier editor o procesador de texto tal como Bloc de Notas de MS Windows. No obstante, si se plantea la creación de sitios web, es mejor contar con la ayuda de paquetes especializados como FrontPage de Microsoft o Dreamweaver de Macromedia.

Existen cuatro tipos de marcado definidos en HTML:

■ **Estructural:** describe la estructura del texto. Por ejemplo, <h1>Saludo</h1> indica a los visualizadores que Saludo debe tratarse como un encabezamiento de primer nivel.

- **Presentación:** describe la presentación visual del texto sin tener en cuenta su estructura. Por ejemplo, se puede usar `<b>Negrita</b>` para colocar un texto en negrita. No obstante, estas marcas están en desuso y se aconseja a los autores utilizar hojas de estilo CSS (*Cascading Style Sheets*).
- **Hipertexto.** Dado que HTML es un lenguaje hipertextual, la marca más importante es el enlace entre varios objetos. Un enlace remite a un lector de una parte a otra de un documento o desde un documento a otro disponible en el mismo sistema o en cualquier localización de Internet. Por lo tanto un documento en HTML es un hipertexto que no presenta una lectura lineal sino que el lector es libre de seguir el camino que desee entre las múltiples referencias o posibilidades que le ha abierto el escritor.
- **Elementos destinados a crear objetos,** por ejemplo botones y listas.

Si un documento HTML contiene objetos adicionales, por ejemplo imágenes, éstos son almacenados fuera del documento y enlazados utilizando las marcas oportunas. Por lo tanto, un documento HTML consiste del fichero HTML propiamente dicho y, opcionalmente, de una serie de material adicional como referencias, gráficos o documentos multimedia. Todos estos objetos deben estar disponibles cuando el usuario abra el documento.

Del desarrollo del HTML se encarga el *World Wide Web Consortium* o W3C<sup>15</sup>. En paralelo, las empresas fabricantes de navegadores web como Microsoft han añadido y modificado características de la norma con objeto de mejorar sus productos. Una breve historia de las especificaciones desarrolladas por el W3C es:

- **HTML 2.0.** Definido por la IETF en 1996 basado en un trabajo de 1994. Está descrita en la RFC 1866<sup>16</sup>.
- **HTML 4.0.** Hecho público por el W3C en 1997 con una revisión posterior en 1998.
- **HTML 4.01.** Publicado a finales de 1999 para fijar errores en la especificación anterior<sup>17</sup> Es la especificación estable en estos momento y la que debería seguirse para la elaboración de todo tipo de documentos en HTML.

El seguimiento de estas normas varía considerablemente. Dagfinn Parnas<sup>18</sup>, después de revisar en su tesis casi dos millones y medio de páginas, llegó a la conclusión de que la mayoría de ellas no se ajustaba a la norma del W3C. Igual que sucede con el lenguaje natural, se pueden crear convenciones al margen de la norma que hacen uso del carácter abierto de la misma.

Actualmente, el W3C está trabajando en el desarrollo del XHTML (la versión 1.0 ya está disponible) que viene a ser una refundición de HTML y XML. XHTML está pensado para ser lo suficientemente similar al HTML como para que los desarrolladores puedan hacer una transición fácil de HTML a XML. La adopción de este nuevo estándar está impulsada por el interés, tanto comercial como del consumidor, de contenidos de herramientas más poderosas que sobrepasen las limitaciones del HTML. No obstante, su generalización llevará aún cierto tiempo. Para una introducción a este nuevo lenguaje puede consultarse el artículo de Tom Lieb<sup>19</sup>.

En el campo de la edición científica, el HTML no es un lenguaje apropiado por varias razones: en primer lugar por el escaso control que el diseñador tiene sobre la forma en que el lector visualizará un documento. En esta visualización juegan un papel importante tanto las características técnicas del equipo donde se visualice el documento (tamaño del monitor, capacidad de resolución, etc.), como del software utilizado. Existen multitud de visualizadores web que, en mayor o menor medida, aceptan la especificación W3C así como características adicionales incluidas por estas empresas. Finalmente, la configuración particular de cada visualizador puede afectar a tamaños de letras, etc. En muchos casos, los documentos científicos serán impresos en lugar de ser leídos en pantalla. La calidad de la impresión

de un documento HTML es mucho más pobre que en otros formatos como PDF. Existen importantes limitaciones para la inclusión de fórmulas matemáticas o químicas si no es con la adición de imágenes.

No obstante estas limitaciones, son muchas las editoriales que ofrecen versiones en HTML de los documentos que publican, si bien nunca de forma exclusiva, sino acompañadas de versiones en formatos más aptos para ser impresos como el PDF.

Así por ejemplo:

- *HighWire Press*<sup>20</sup>, uno de los grandes editores de revistas de acceso libre ofrece dos alternativas para cada artículo: HTML para visualización en pantalla y PDF para impresión.

- *ScienceDirect*<sup>21</sup>, el portal de revistas electrónicas de Elsevier sigue la misma línea.

- *SpringerLink*<sup>22</sup>.

□ **PDF.** Ninguno de los formatos de marcas que hemos visto puede proporcionar una representación facsímil de una página impresa, utilizando, por ejemplo, los mismos colores, fuentes, etc. y no puede incluir gráficos a no ser que sean convertidos en imágenes. En muchos casos esta fidelidad al documento original no es necesaria, las variaciones en fuentes y proporciones son de menor importancia cuando se trata de ofrecer un acceso rápido a los contenidos. No obstante, cuando éste no es el caso o cuando se quiere ofrecer al lector una versión de calidad del documento para ser impreso, el formato recomendable es el PDF (*Portable Document Format*)<sup>23, 24</sup>.

PDF es un formato propietario desarrollado por la empresa Adobe Systems pero que se está convirtiendo en un estándar de facto para la industria editorial, no sólo en el ámbito científico sino también empresarial. PDF permite presentar una publicación sobre pantalla o sobre papel con la misma apariencia y calidad que si fuera realizada en una imprenta. Lo que hace este formato es congelar la publicación original convirtiéndola en una imagen (que dependiendo de la longitud del documento puede ir desde una a cientos de páginas).

PDF viene a solucionar el problema de la portabilidad de los ficheros. En muchos casos, para poder leer un documento escrito en un determinado procesador de textos, es necesario que el lector disponga del software adecuado. Si no es así puede recurrir a algún tipo de conversión entre formatos que, por definición, implicará alguna pérdida de información. Un documento escrito en PDF puede ser visualizado en cualquier plataforma informática con el único requisito de disponer de un lector que está disponible gratuitamente en la Red. Además del lector desarrollado por Adobe, llamado Acrobat Reader, existen multitud de lectores creados por terceras partes y disponibles gratuitamente en Internet.

PDF es el sucesor del lenguaje PostScript también desarrollado por la misma empresa y muy popular en las áreas científicas y técnicas durante los años ochenta y noventa. PostScript es un lenguaje de descripción que representa cada página de un documento incluyendo todos los elementos que aparecen sobre ella, como emplazamiento de gráficos, fuentes, etc.

Dado que se trata de una imagen, su tamaño es mayor y es más lento de descargar que la versión en HTML del mismo documento. Por lo tanto, es recomendable que documentos muy grandes se dividan en partes que puedan descargarse por separado. Un versión completa puede ser útil también para usuarios con buena conectividad.

PDF es un formato muy atractivo para editores particularmente si se utiliza en conjunción con XML. Ahora es posible incluir hipertexto en documentos PDF así como extraer el texto de los mismos para ser indizado por los motores de búsqueda. No

obstante, PDF no es apropiado para todos los mercados y, además, los sistemas utilizados por deficientes visuales no trasladan documentos en PDF en sonido.

Como ejemplos de editores que actualmente están usando este formato se podrían señalar los mismos que en el apartado anterior pues generalmente se acompañan dos versiones del mismo documento, una en HTML para visualización y otra en PDF para impresión.

- ❑ **Macromedia Director<sup>25</sup> y Flash<sup>26</sup>.** Ambos productos son software de la empresa Macromedia. Están orientados hacia la publicación en web, y soportan gráficos de vectores animados.

Flash es una herramienta para crear animaciones que consuman poco ancho de banda, presentaciones y otras formas de contenidos dinámicos. Ofrece una interfaz de programación para la creación de aplicaciones, interfaces web, etc. *Macromedia Director Shockwave Studio* es una herramienta similar en cuanto a su objeto pero mucho más poderosa y sofisticada. Está orientada a la creación de aplicaciones de ocio en 3D, demostraciones interactivas de productos, aplicaciones de e-learning y otros tipos de contenidos multimedia destinados tanto a la Web como a ser publicados en CD o DVD.

A diferencia del HTML es esencialmente independiente del visualizador web que se utilice. La mayor parte de los usuarios con versiones actualizadas de sus navegadores verán versiones idénticas de los documentos. Por el contrario, a diferencia de HTML se trata de un software cerrado y para crear una animación es necesario disponer del software distribuido por Macromedia y para ver los documentos se necesitan los *plug-in* adecuados, que sí son distribuidos gratuitamente. Cuando los productos se distribuyen en CD o DVD el software de visualización va integrado y no requiere mayores requisitos. No obstante en estos casos los requerimientos de hardware pueden ser elevados.

A la hora de trabajar con animaciones y vídeo hay que actuar con cautela. Es recomendable repasar las cuestiones de accesibilidad y usabilidad de las páginas web. En general, aunque ambas posibilidades son posibles y tienen su lugar en la Web, los diseñadores tienden a despreocuparse de cuestiones como navegación o paciencia de los usuarios.

Aunque Flash puede ser utilizado para diseñar sitios web completos, esto limitará la accesibilidad considerablemente por lo que solamente se recomienda para realizar animaciones allí donde sea estrictamente necesario y nunca como un fin en sí mismo. Igualmente, los productos desarrollados con Director pueden ser impactantes pero sólo funcionarán bien en ordenadores multimedia equipados con tarjetas de vídeo y sonido y potentes procesadores.

Macromedia mantiene una página<sup>27</sup> donde se hace una revisión de las empresas e instituciones que están utilizando sus productos.

- ❑ **RTF.** RTF son las siglas de *Rich Text Format*. Se trata de una especificación<sup>28</sup> desarrollada por Microsoft para describir la representación interna de documentos. Los ficheros RTF son ficheros de texto con una serie de comandos embebidos que proporcionan a un ordenador información básica sobre fuentes, márgenes, etc. RTF es útil para asegurar que cuando se intercambien documentos entre procesadores de texto diferentes permanezca un formato básico. Igualmente se utiliza para compatibilidad con antiguas versiones de procesadores de texto. Así mismo, en edición científica los documentos en forma manuscrita son a menudo remitidos a los editores en RTF para asegurar la compatibilidad.
- ❑ **Microsoft Word.** Microsoft Word<sup>29</sup> es el procesador de texto más popular. Los documentos creados con Word no podrán ser abiertos con otros procesadores de texto,

a no ser que se lleve a cabo algún tipo de conversión entre ellos. La utilización de este formato solamente se justificaría en entornos cerrados, por ejemplo la Intranet de una institución, donde el editor tenga la absoluta seguridad de que todos los lectores tienen acceso a este editor y quiera hacer uso de sus facilidades para incluir imágenes, tablas o gráficos. En cualquier otro caso, sobre todo si el medio de distribución va a ser Internet, es absolutamente desaconsejable.

- ❑ **Microsoft Reader.** Microsoft Reader<sup>30</sup> es el software de Microsoft destinado a crear y distribuir libros electrónicos (e-books). Una de las principales desventajas del medio electrónico frente al papel, que hace que el comercio de e-books no despegue, es la dificultad para leer sobre una superficie luminosa como es el monitor y con una resolución mucho menor que las páginas impresas. MS Reader es el producto desarrollado por Microsoft para permitir la lectura sobre pantalla que se acerque a la conveniencia y la calidad del papel. Es el primer producto que incluye tecnología ClearType, que mejora la resolución sobre pantallas de cristal líquido hasta en un 300% para suministrar un producto lo más parecido posible al papel.

En cierta forma, MS Reader es un competidor de PDF, con la salvedad que PDF está diseñado para producir documentos impresos y Reader lo está para leer sobre una pantalla de ordenador. Los documentos PDF son estáticos en el sentido de que no se adaptan a los diferentes equipos de visualización, por ejemplo, al tamaño de la pantalla. Por el contrario, Reader es dinámico, ya que es capaz de construir cada página del libro de acuerdo a las preferencias de cada usuario y con el equipo que está usando.

Las herramientas para visualizar y generar documentos en este formato son gratuitas y pueden ser descargadas de la página web de Microsoft. No obstante, este es el principal inconveniente de MS Reader, el ser un software propietario que solamente existe para MS Windows sin que la empresa piense desarrollar versiones para otros sistemas operativos.

Para generar un libro en formato Reader es necesario partir de un documento en HTML que cumpla con las especificaciones de la *Open eBook Foundation*. Adicionalmente, se requieren las imágenes de la portada del libro. Se puede generar el documento a través de un *add-in* para MS Word que se puede descargar gratuitamente; utilizando el servicio *e-Book Express* vía Internet (esta aplicación recibe un documento en una amplia variedad de formatos y devuelve el e-book); o bien utilizando herramientas profesionales como *ReaderWorks Publisher*, que permite generar automáticamente una tabla de contenidos o índices a partir del documento HTML, ordenar archivos de capítulos en el orden lógico, etc.

Si bien desde nuestro punto de vista no es un formato recomendable por los inconvenientes señalados, se trata de un formato que se está popularizando gracias, en parte, a la fuerza del gigante de la informática. Así editoriales como Random House o Mondadori, y el distribuidor Amazon utilizan MS Reader. Microsoft mantiene una lista<sup>31</sup> de editoriales que utilizan sus productos.

## DOCUMENTOS ESTRUCTURADOS

- ❑ **SGML.** Son las siglas de *Standard Generalized Markup Language*, o, lo que es lo mismo, la norma ISO 8879 de 1986 que define un lenguaje de marcas general para delimitar los elementos que componen un documento de texto sin describir su apariencia. Para ello se separa estructura de presentación y SGML se centra sólo en la primera de forma, que si un documento cambia de un sistema a otro o partes de un documento son utilizadas en otro, no necesiten ser recodificadas. Utiliza texto plano

para representar tanto el contenido como las marcas que lo describen. No tiene ningún código propietario, por el contrario cada usuario (o grupo de usuarios) puede crear cualquier código que necesite.

La clave para interpretar los códigos en un documento SGML es la DTD (*Document Type Definition*). De hecho, un documento SGML no está completo sin tres partes distintas: declaración (que da al usuario información del tipo de lenguaje y codificación que se está usando, por ejemplo inglés y ASCII), la DTD (que detalla todos los códigos y reglas que se pueden utilizar para describir un particular tipo de documento) y finalmente la instancia (el texto que va a ser publicado), marcado con los códigos descritos en la DTD.

Una DTD es una descripción formal de un particular tipo de documento. Especifica qué nombres se van a utilizar para los distintos elementos, en qué situación pueden aplicarse y cómo se relacionan con el resto de elementos. Por ejemplo si queremos que un documento pueda contener listas compuestas de ítems se podría hacer la siguiente definición:

```
<!ELEMENT List (Item)+>
<!ELEMENT Item (#PCDATA)>
```

Esto define una lista como un elemento compuesto de uno o más ítems y cada ítem se define como un elemento que contiene nada más que texto plano. Así, en el documento SGML se podría especificar lo siguiente:

```
<List><Item>Chocolat</Item><Item>Music</Item><Item>Surfi
ng</Item></List>
```

Aunque SGML no debe ser complicado, a menudo lo es dado que definir una nueva DTD no es tarea fácil. Cuanto más amplia sea la variedad de documentos que van a ser descritos con una DTD mayor será la complejidad de su definición. En este sentido la ISO ha definido en su norma 12083<sup>32</sup> una DTD para el marcado de libros. Una DTD puede incluir no sólo descripciones de los elementos sino también definiciones, es decir, reglas que determinan cómo usar cada elemento. Así, es posible especificar qué elementos se quiere que aparezcan en un contexto pero no en otro. Por ejemplo, el DTD para una revista puede distinguir entre los elementos que el editor quiera distribuir gratuitamente y aquellos para los que se necesitará pagar.

A medida que el número y la complejidad de sus publicaciones electrónicas crecen, los editores necesitan alguna forma de organizarlas y gestionarlas. Dadas sus poderosas y flexibles capacidades para describir la estructura, así como su capacidad para capturar y organizar información sobre las publicaciones (metadatos), SGML es una solución efectiva.

Las revistas son candidatas ideales para representarlas en SGML dado que son documentos altamente estructurados y repetitivos.

Para una lista de instituciones que actualmente están usando SGML puede consultarse la compilación que hace CoverPages<sup>33</sup>.

- ❑ **XML.** Son las siglas de *Extensible Markup Language*. Se trata de un lenguaje de marcas diseñado para llevar un paso más adelante todas las características que ofrecía el HTML. De hecho, los dos derivan de SGML (*Standard Generalized Markup Language*) del que hemos hablado anteriormente. El responsable de su desarrollo es el *World Wide Web Consortium* por lo que se trata de un estándar abierto. La versión actual<sup>34</sup> es la tercera edición de la número 1.0 de febrero de 2004.

La principal diferencia de XML frente a HTML es que el primero describe la estructura de un documento no su apariencia. La relevancia del XML en la actualidad procede del hecho de que el HTML básico no sería capaz de proporcionar un entorno adecuado para la siguiente generación de edición basada en la red, donde los datos tendrán que ser mostrados en un creciente número de entornos y buscados en formas más sofisticadas.

En cuanto a su relación con SGML podríamos decir que ésta es la lengua madre, que ha sido utilizada para describir miles de tipos de documentos en múltiples campos de la actividad humana, desde la edición electrónica hasta registros clínicos o notaciones musicales. SGML es demasiado grande y compleja lo que hace difícil su implementación. XML es una versión abreviada de SGML para hacer más fácil a los autores crear tipos de documentos y a los programadores desarrollar aplicaciones que puedan trabajar con ellos.

La variedad de documentos XML es muy grande y puede ir desde los más simples con marcas diseñadas por el propio autor como podría ser:

```
<?xml version="1.0" standalone="yes"?>
<conversacion>
  <saludo>Hola, ¿cómo estas?</saludo>
  <respuesta> Encantado de verte!</respuesta>
</conversacion>
```

En otros casos pueden ser más complicados con un DTD o esquema asociado que delimita el vocabulario y elementos a utilizar. Un caso podría ser:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE titlepage SYSTEM "http://www.foo.bar/dtds/typo.dtd"
[<!ENTITY % active.links "INCLUDE">]>
<titlepage id="BG12273624">
  <white-space type="vertical" amount="36"/>
  <title font="Baskerville" size="24/30"
    alignment="centered">Hello, world!</title>
  <white-space type="vertical" amount="12"/>
  <!-- In some copies the following decoration is
hand-colored, presumably by the author -->
  <image location="http://www.foo.bar/fleuron.eps"
    type="URL" alignment="centered"/>
  <white-space type="vertical" amount="24"/>
  <author font="Baskerville" size="18/22"
    style="italic">Vitam capias</author>
  <white-space type="vertical" class="filler"/>
</titlepage>
```

No se deben confundir DTD con esquemas XML. Una DTD sirve para describir la estructura de un documento: asigna nombres a los elementos, atributos y entidades que se pueden utilizar. En principio, están pensadas para ser utilizadas con documentos textuales, sin datos tabulados. De esta forma si se necesita especificar rangos de números o definir limitaciones sobre el uso del contenido texto, el DTD no puede hacer nada. Es aquí donde interviene el esquema XML que proporciona la posibilidad de validar el contenido de los elementos en función de cada tipo de documento y ofrecer a sus diseñadores criterios para saber si ese contenido es correcto o no.

La visualización de un documento XML requiere la existencia de una hoja de estilo asociada. En HTML la forma en que se interpretan las marcas y elementos está predefinida en los visualizadores web dado que el número de marcas es limitado. En XML cada autor puede crear sus propios elementos lo que hace que los visualizadores no puedan prever cómo esos elementos puedan ser utilizados. Los visualizadores que admiten XML, probablemente, aceptarán también las hojas de estilo CSS (*Cascading Style Sheets*). No obstante, también se puede utilizar un lenguaje mucho más robusto como es XSLT<sup>35</sup> para transformar XML en HTML que ya puede ser entendido por cualquier visualizador.

Los prefijos de nombres de espacio distinguen nombres idénticos que estén sacados de diferentes vocabularios XML. Una declaración de espacio de nombres en un documento XML asocia un prefijo de espacio de nombres con un único URI (*Uniform Resource Identifier*). El prefijo puede entonces ser utilizado sobre elementos o nombres de atributos en el documento.

En resumen, no habrá una forma mejor de describir la estructura de los documentos que usando XML. Se trata de una norma internacional, no propietaria, lo cual libera los documentos de procesos tediosos y costosos de conversión entre sistemas. No requiere ningún software o hardware especial. Aunque cada vez hay más sistemas basados en XML es posible crear un documento XML utilizando cualquier editor de texto. Finalmente, libera el documento de las restricciones de cualquier método de presentación, adaptándolo a cualquier forma de visualización, impresa o electrónica, conocida o por descubrir.

- **TEI.** Son las siglas de *Text Encoding Initiative*<sup>36</sup>. Inicialmente lanzado en 1987, TEI es un estándar internacional e interdisciplinar destinado a ayudar a bibliotecas, museos, universidades e incluso a investigadores a representar todo tipo de textos científicos o literarios, utilizando un sistema de codificación que sea lo más expresivo y menos obsolecente posible. TEI fue creado, inicialmente, por tres sociedades científicas: *Association for Computers and the Humanities*, *Association for Computational Linguistics* y *Association for Literary and Linguistic Computing*, pero hoy se ha convertido en un consorcio internacional compuesto por editoriales, bibliotecas, universidades, etc.

Las recomendaciones para la codificación de todo tipo de textos en todo tipo de lenguajes se materializan en una guía titulada *Guidelines for Electronic Text Encoding and Interchange*<sup>37</sup>. La primera edición apareció en 1994 y la última, conocida como P4, en 2002. En estos momentos se está trabajando en una nueva que se anuncia aparecerá en 2004. Esta guía no proporciona sólo un DTD sino también un entorno para la creación de múltiples DTDs adaptados a las necesidades de cada institución. Uno de estos DTDs se conoce como *TEI Lite*. Este define una selección manejable de todos los elementos que forman TEI. Básicamente, es un conjunto de elementos que cualquiera, trabajando con TEI, debería conocer.

El principal objetivo de P4 fue implementar un soporte en XML adecuado al mismo tiempo que aseguraba que los documentos producidos con versiones anteriores de TEI permanecieran utilizables con la nueva. Las versiones anteriores se habían basado en SGML. Ambos lenguajes tienen en común la definición de textos en términos de elementos y atributos y reglas que rigen su apariencia en el texto. El uso de XML en TEI es básicamente el mismo que en cualquier otra aplicación por lo que cualquier software de carácter general capaz de trabajar con XML debería ser capaz de procesar los documentos marcados como TEI.

El lanzamiento de una versión compatible con XML es un avance considerable que coloca la iniciativa TEI en la corriente de los más actuales desarrollos de las biblio-

tecas digitales y de la *World Wide Web*. En resumen, la guía define un lenguaje para describir como los textos son construidos y propone nombres para sus componentes. Hay muchos vocabularios similares en la industria (banca, aeronáutica o modelos químicos por ejemplo). El logro alcanzado por TEI ha sido exportarlo al mundo de los documentos textuales y datos lingüísticos.

En la página web del proyecto<sup>38</sup> se puede consultar una lista de instituciones que están trabajando con TEI, así como una serie de tutoriales<sup>39</sup> para comenzar a utilizarlo.

- ❑ **La estructura de edición del *Open e-book Forum*.** El *Open e-book Forum*<sup>40</sup> es una organización internacional dedicada al desarrollo y promoción de la edición electrónica a través del desarrollo de estándares para sistemas de libros electrónicos de los cuales se beneficiarán los creadores de contenidos, los desarrolladores de sistemas de lectura digital y los consumidores. En ella participan compañías de software y hardware, editores, autores, usuarios y organizaciones relacionadas.

Un libro electrónico es un documento compuesto de tres partes: publicación o contenido digital que se lee; dispositivo de lectura o componente físico utilizado para acceder al contenido (puede ser un lector de libros electrónicos dedicado, un PC, un asistente personal PDA, etc.); y un sistema de lectura, que es la combinación de hardware y software que procesa la publicación en un dispositivo de lectura determinado y la presenta al usuario.

Estas especificaciones se materializan en la *Open e-book Publication Structure* (OEBPS)<sup>41</sup>. Es una especificación basada en XML para representar el contenido, la estructura y la presentación de los libros electrónicos. En especial está orientada a ofrecer a los proveedores de contenidos, autores y distribuidores unas guías mínimas que aseguren la fidelidad, exactitud, accesibilidad y adecuada presentación de libros electrónicos sobre varias plataformas. El objetivo es definir una norma de descripción de contenido para que los productores puedan ofrecer sus productos en una variedad de sistemas de lectura.

Está desarrollada por un grupo de trabajo dentro del Forum. Su primera versión es de 1999 y la actualmente vigente, la número 1.2 es de 2002. Esta versión no cambia sustancialmente la primera y se ha asegurado la compatibilidad hacia atrás para que todos los documentos creados con la anterior continúen cumpliendo la nueva. Entre los cambios más notables están el incremento del número de elementos, la eliminación de todos aquellos cuyo uso no se recomendaba, la mejora del soporte de CSS, etc.

OEBPS está basada en XML por la simplicidad de este lenguaje y porque es previsible que los documentos en XML se adapten bien a futuras tecnologías y usos. Además, XML proporciona reglas bien definidas para la sintaxis de los documentos, lo cual reduce el coste de implementación y la incompatibilidad entre sistemas. Por otra parte, XML es extensible, puesto que no está limitado a ningún conjunto particular de elementos de descripción.

Otros estándares con los que está relacionado son: CSS (*Cascading Style Sheet*), *Dublin Core* y Unicode. CSS<sup>42</sup> es un lenguaje de estilos utilizado para definir la apariencia con que se presentará al usuario final un documento en formato XML. El cumplimiento de esta norma no es exhaustivo ya que no están soportadas todas las características de CSS y además se han incluido otras nuevas para limitar los problemas de implementación que pudieran tener los desarrolladores de sistemas de lectura. Por su parte *Dublin Core* es un formato de metadatos para describir recursos y documentos electrónicos. Unicode define un conjunto de códigos caracteres

para poder representar caracteres en todas las lenguas conocidas y que aparezcan en el futuro.

En la página web del Forum puede consultarse una lista de organizaciones participantes entre las que podemos destacar Oxford University Press, Gale Group, Elsevier o Springer-Verlag. Igualmente se puede descargar un ejemplo de libro<sup>43</sup> codificado según OEBPS. Se ofrece tanto la fuente en formato XML, como las hojas de estilo para su visualización.

- **TeX/LaTeX.** Es un sistema de composición de textos orientado al mundo científico que utiliza un sistema de marcado similar en cierta forma a SGML o XML. Cada documento en TeX/LaTeX se compone de texto plano acompañado de marcas que definen tanto la estructura general del documento como cada uno de los elementos que lo forman. Hay que tener presente que fue diseñado originalmente para trabajar con documentos impresos y que su adaptación a documentos web o multimedia no es perfecta.

TeX fue desarrollado por Donald Knuth en el año 1978, a petición de la *American Mathematical Society*, para poder componer texto y fórmulas matemáticas con la calidad de los sistemas de fotocomposición de entonces, pero sin el penoso esfuerzo que esos sistemas requerían.

TeX/LaTeX es un sistema abierto del que existen versiones para casi todas las plataformas informáticas. Además existen aplicaciones para convertir los documentos fuente en otros formatos como PDF o HTML. Es un lenguaje especialmente apropiado para el tratamiento de fórmulas matemáticas.

TeX/LaTeX cuenta con una gran comunidad de usuarios. Gracias a ello, dispone de mucha ayuda y recursos on line, y también de una gran cantidad extensiones para campos concretos (que van desde la música o la química molecular hasta los circuitos eléctricos o el ajedrez), o que aportan nuevas funcionalidades (documentos con un número variable de columnas, transparencias, colores...).

Un documento TeX/LaTeX se divide en dos partes: el preámbulo y el contenido del documento. El preámbulo se define como la parte del documento que va antes del contenido. Se usa para incluir información sobre la naturaleza del documento, los comandos que definimos nosotros mismos y para incluir los paquetes opcionales que queramos usar en nuestro diseño.

```
\documentclass[a4paper,12pt]{article}
\usepackage[latin1]{inputenc}
\usepackage[T1]{fontenc}
\usepackage[spanish]{babel}
\title{Título del texto}
\author{Autor del texto}
```

En este ejemplo hacemos lo siguiente:

1. Se define el tipo de documento (artículo, tipo de letra base de 12 puntos y hojas A4).
2. Se incorpora el paquete *inputenc* para que acepte símbolos como letras con tilde, eñes, ...
3. Se incorpora el paquete *fontenc* para que se codifiquen los símbolos especiales con los códigos correspondientes y no con combinaciones de caracteres.
4. Se carga el paquete *babel* que modifica el comportamiento de TeX/LaTeX para adaptarlo a idiomas distintos del inglés.
5. Se define el título del documento.
6. Se define el autor del documento.

El resto del documento (el contenido junto con las diferentes marcas estructurales) se introducirá dentro del entorno *document*, en el ejemplo incluimos la orden *maketitle* para que TeX/LaTeX construya una portada de nuestro documento.

```
\begin{document}
\maketitle
Texto a continuación...
\end{document}
```

Para escribir textos en TeX/LaTeX solamente se necesita un editor de texto. Ahora bien esos ficheros deben ser compilados para conseguir las versiones utilizables por el usuario (PDF, PostScript, HTML, etc). La compilación se realiza a través de aplicaciones que son gratuitas y que pueden descargarse de Internet. En la página web del grupo de usuarios de TeX en castellano, CervanTeX<sup>44</sup>, puede encontrarse más información sobre cómo componer textos utilizando este sistema. Además existen algunos manuales básicos<sup>45, 46</sup>.

## LOS FORMATOS DE IMAGEN

□ **Imágenes estáticas.** Los formatos digitales de imágenes se dividen en vectoriales y mapas de bits. El formato vectorial para la Web de preferencia es SVG. Su soporte por parte de los navegadores no es muy amplio todavía. Una imagen digital de mapa de bits es una imagen que ha pasado por un proceso de conversión, para que pueda ser almacenada en forma de bits en un ordenador. Entre los formatos de imagen de este tipo más usados están JPG, GIFF y PNG.

Existen multitud de formatos en los que presentar un archivo de imagen digital estática. Los más conocidos son:

- BMP, Microsoft Windows Bitmap file.
- CUR, Microsoft Windows Cursor file.
- EPS, Encapsulated PostScript.
- GIF, CompuServe Graphics Image Format file.
- HDF, Hierarchical Data Format file.
- ICO, Microsoft Windows Icon file.
- JPG, Joint Photographic Experts Group.
- WMF, Window Meta File.
- PBM, Portable Bitmap file.
- PGM, Portable Grayscale Map file.
- PIC, PIXAR Picture file.
- PCX, PC Paintbrush.
- PICT, SoftImage PICT file.
- PIX, Alias Pixel image file.
- PNG, Portable Network Graphic.
- PPM, Portable Pixel Map file.
- PS, PostScript.
- RAS, Sun Raster file.
- RGB, Silicon Graphics RGB image file.
- RGBa, 4-component Silicon Graphics image file.
- RGBA, 4-component Silicon Graphics image file with generated alpha.
- RLA, Wavefront raster image file.
- RLE, Utah Runlength-encoded image file.
- RPBM, Raw Portable Bitmap file.

RPGM, Raw Portable Grayscale Map file.  
RPNM, Raw Portable any Map file.  
RPPM, Raw Portable Pixel Map file.  
SYNU, Synu Image file.  
TGA, Truevision Targa Image file.  
TIFF, Tagged Image File.  
VIFF, Khoros Visualization Image File Format.  
X, Stardent AVS X Image file.  
XBM, X11 Bit Map file.  
XWD, X Window Dump Image file.

Hay muchos programas gráficos que nos permiten salvar la imagen creada a la mayoría de los formatos relacionados en el cuadro anterior, pero hoy por hoy para insertar en nuestras páginas web sólo nos resultan útiles cuatro de los anteriores formatos, ya que los demás no son reconocidos por los navegadores. Estos formatos posibles son: GIF, JPG, BMP y PNG. Vamos a detallar las principales características de estos formatos y también del formato TIFF. Los demás nos hemos limitado a reseñarlos.

■ **Formato JPEG.** Es un formato bastante flexible que fue desarrollado por el *Joint Photograph Experts Group* al intentar crear un formato para intercambio en la red que soportara gráficos de gran calidad, muchos colores (24 bits de profundidad de color) y un nivel de compresión elevado y configurable, es decir, utiliza una distribución de 24 bits/píxel, cada imagen contiene alrededor de 16,777,216 colores, el algoritmo de compresión es de tipo «con pérdida». Los archivos de este tipo al ser comprimidos resultan más pequeños que los de tipo GIF.

En un principio costó que este formato se abriera camino ya que no estaba muy bien valorado por los diseñadores gráficos, sin embargo poco a poco fue siendo más apreciado llegando a ser en la actualidad el más utilizado. Su sistema de compresión se basa en eliminar de la imagen aquellos colores que el ojo humano no puede distinguir y en agrupar gamas de colores similares. A mayor nivel de compresión, menor calidad en la imagen resultante. Se trata de encontrar un equilibrio entre el peso del archivo y la calidad de la imagen resultante. Entre sus ventajas podemos destacar:

- Su sistema de compresión es complejo y graduable, lo que permite una amplia gama de compresiones diferentes para una misma imagen. El tamaño de la imagen no depende del número de colores que contenga sino de su nivel de compresión.
- Maneja 16 millones de colores.
- Permite crear imágenes progresivas. Un JPG progresivo es la equivalencia en este formato de los GIFs entrelazados, permite dividir una imagen en franjas de forma que ésta va apareciendo ante los ojos conforme aumenta la definición de la pantalla, y el usuario la va viendo poco a poco, sin tener que esperar a que se cargue completa en la memoria.

Entre sus defectos destacan:

- Aunque permite grandes niveles de compresión, si la imagen es de gran resolución, el fichero sigue teniendo mucho peso.
- No permite la creación de imágenes de fondo transparente ni imágenes animadas.
- Con pocos colores y gradaciones da archivos más pesados que el formato GIF, sin mejora apreciable de la calidad.

- **Formato GIF (*Graphic Interchange Format*)**. Formato creado por CompuServe en 1987 (su primera versión) aunque la patente del mismo pertenece a la empresa Unisys. En 1989, CompuServe amplió las propiedades de este formato para que se pudieran crear con él imágenes animadas, transparencias e imágenes entrelazadas, dando lugar a la segunda versión de este estándar, conocida como gif 89a. Usa un sistema de compresión LZW modificado.

Los ficheros gráficos GIF son los más comunes en las páginas web, debido sobre todo a su facilidad de construcción y a su poco peso, además de ser soportados por todos los navegadores.

Con este formato, cuantos menos colores tenga la imagen y más extensión total ocupe cada uno de ellos dentro de la misma, mejor compresión obtendremos.

Dentro de sus ventajas debemos destacar:

- Con este formato se pueden crear gráficos de fondo transparente.
- Se puede guardar una imagen asociándola con su propia paleta de colores. De esta forma disminuye el tiempo de carga y las imágenes son más compatibles con sistemas operativos y navegadores.
- Podemos construir imágenes animadas, creando un gráfico formado por varias escenas que se unen y forman una secuencia con un tiempo establecido entre una imagen y otra, como en una película.
- Permite construir imágenes entrelazadas, tal como hemos descrito en el formato JPG.

Entre sus desventajas debemos destacar:

- Sólo nos permite trabajar con 256 colores.
- Debido a esta limitación en los colores, no es adecuado para representar escenas con muchos colores ni con una gradación entre estos, como es el caso de una fotografía. En estos casos es mejor usar otro tipo de formato, como el JPG.

- **Formato PNG (*Portable Network Graphic*)**. Es uno de los últimos formatos de compresión que aparecen en la escena web, exactamente en 1994. Conjuga lo mejor de los formatos que habitualmente se han venido utilizado, que son GIF (*Graphics Interchange Format*) y JPG (*Joint Photographer's Experts Group*). Con él se ha pretendido obtener un sistema de compresión como el que usa el formato GIF pero de código libre. Es el formato oficial de la suite Microsoft Office. Combina la posibilidad de asociar a un fichero gráfico una paleta de color, ventaja que se obtenía con el formato GIF a la vez que permite utilizar millones de colores (color real), ventaja que se obtenía con JPG. El formato de compresión que utiliza es de alta calidad, lo que permite reducir los tiempos de descarga. El sistema de creación de imágenes entrelazadas que utiliza es más óptimo que con GIF porque las procesa con mayor velocidad. El efecto de transparencia fondo/figura es mejor, ya que no se basa en un color de fondo (caso del GIF) sino en la máscara creada por alguno de los canales alfa que hayamos seleccionado desde Photoshop. PNG es el mejor formato a utilizar cuando trabajamos con imágenes transparentes y en los casos en que queramos introducir capas en un documento web.

Sus principales desventajas son que no soporta las imágenes animadas y que su uso no está aún muy extendido, aunque los principales programas gráficos permiten trabajar con él.

- **Formato BMP**. Desarrollado e impulsado por Microsoft, BMP es una abreviatura de *Windows BitMaP* (Mapa de Bits de Windows), y en realidad una imagen en este formato es la sucesión de puntos coloreados guardados cada uno independientemente. Al ser un formato gráfico propiedad de Microsoft sólo el navegador Internet Explorer lo soporta. Netscape no soporta el tipo MIME BMP en ninguna de sus

versiones. Entre sus ventajas está el ser un formato extremadamente simple, y entre sus desventajas el ser sólo compatible con Internet Explorer y el gran peso de los ficheros de este formato.

- **Formato TIFF.** *Tagged Image File Format* es un formato que permite el almacenamiento e intercambio de informaciones gráficas provenientes de escáneres, *frame grabbers* y de programas de retoque y edición de fotos. Puede ser utilizado en varias plataformas como PC (IBM), o Macintosh y estaciones de trabajo UNIX.

Sus principales ventajas son:

- Su estructura es conocida por una amplia variedad de aplicaciones.
- Es independiente de sistemas operativos y plataformas gráficas.
- Los archivos con este formato se pueden ajustar a las características de un escáner, monitor o impresora.

Una de sus principales desventajas es que un usuario puede indicar nuevos atributos (tags) en una imagen, imposibilitando la interpretación correcta de los datos por otro sistema de lectura de imágenes.

La estructura de un archivo de formato TIFF se define por una secuencia de 8 bits (1 byte), donde los bytes son numerados de 0 a N. Este archivo puede alcanzar 232 bytes de extensión.

En conclusión, los principales formatos de imagen comprimida son: GIF (*Graphics Interchange Format*), recomendable para archivos de imágenes con pocos colores como logotipos, dibujos, etc.; JPG o JPEG (*Joint Photographic Experts Group*), recomendable para fotografías, texturas, etc.; TIF; y TIFF comprimidos.

Ya que la compresión de imágenes reduce su calidad se recomienda guardar una copia de la imagen en formato de mayor calidad (TIF sin comprimir) antes de utilizar formatos GIF o JPG. Los archivos gráficos en formato TIFF conservan muy bien los detalles de la imagen pero son enormemente voluminosos (otro tanto se puede decir del formato BMP). Los archivos en formato JPG (o JPEG), en cambio, ocupan menos espacio a cambio de perder un poco de definición que se aprecia esencialmente cuando ampliamos la imagen.

El grado de compresión de un JPG puede ser variable: cuanta más compresión el archivo resultante será más pequeño a costa de perder detalles y calidad.

- **Imágenes dinámicas.** Un archivo de vídeo es una mezcla de formato de imagen vectorial e imagen de mapa de bits.

Dentro de los formatos que sirven para crear imágenes dinámicas o vídeo nos centraremos en tres: MOV, AVI y MPEG.

- **Formato MOV** (*Macintosh QuickTime* .mov, .qt). QuickTime es un formato de archivo desarrollado por Apple Computer para crear, modificar, publicar y ver archivos multimedia. QuickTime admite vídeo, animación, gráficos, 3D y realidad virtual (VR).

- **Formato AVI (PC).** El formato AVI es el nativo de Windows (el formato MOV es el nativo de Macintosh). Un vídeo será estándar, entendiéndose por estándar que se reproducirá en cualquier ordenador con sistema operativo Windows (o capaz de leer archivos AVI), siempre y cuando no apliquemos ninguna compresión al vídeo. Con el vídeo en formato AVI sin comprimir sucede lo mismo que con los archivos BMP: ocupa demasiado, casi 30 GB para una hora a un tamaño de pantalla (resolución) de 352x288, el usado para el VCD, VHS y/o Video-8. Por tanto, lo normal es que, a excepción de en la captura, se le aplique una compresión al vídeo AVI. Hay una infinidad de formatos de compresión. A estos compresores se les conoce como codificadores de vídeo. Su gran variedad supone un problema

porque para poder reproducir un vídeo comprimido con un codificador concreto es necesario tenerlo instalado en el sistema.

AVI significa *Audio Video Interleave* o Audio y Vídeo Intercalado lo que significa que en un fichero AVI los datos de audio y vídeo son almacenados consecutivamente en capas (un segmento de datos de vídeo es seguido inmediatamente por otro de audio). Es el formato más extendido para el manejo de datos de audio/vídeo en un PC.

Los archivos AVI son un caso especial de archivos RIFF (*Resource Interchange File Format* o Formato de Archivos para el Intercambio de Recursos) un formato de propósito general para el intercambio de datos multimedia que fue definido por Microsoft e IBM tiempo atrás.

Sólo existen dos tipos generales de AVI: los basados en *Video for Windows* (los primeros en aparecer) y los basados en *DirectShow* (originalmente *ActiveMovie*). Los formatos de AVI basados en *Video for Windows* son los que ahora forman el núcleo de los denominados AVI DV Tipo-2, mientras que los basados en *DirectShow* (y por extensión en *DirectX*) son los denominados AVI DV tipo-1.

■ **Formato MPEG.** Un vídeo no es más que una sucesión de imágenes en movimiento. Si comprimimos todas esas imágenes (las de un vídeo) en formato JPEG obtendríamos el formato MJPEG, o Motion JPG. Con este formato ya se logra una buena compresión con respecto al original. Partiendo del MJPEG se llegó al formato MPEG (*Moving Picture Experts Group* o Grupo de Expertos de Imágenes en Movimiento) La compresión MPEG supone un avance importante con respecto a la compresión MJPEG al incluir un análisis de cambios entre una imagen clave, o cuadro clave, y un número determinado (suele ser 14) de imágenes posteriores.

Los principales estándares de MPEG son:

- **MPEG-1.** Este estándar se diseñó para permitir la codificación de vídeo progresivo a una velocidad de transmisión de aproximadamente 1,5 millones de bits por segundo (bps). Se diseñó específicamente para medios de vídeo CD y CD-i. Las implementaciones más comunes del estándar MPEG-1 proporcionan una resolución de vídeo de 352x240 a 30 imágenes por segundo (fps). Cuando utiliza este estándar, obtiene una calidad de vídeo ligeramente inferior a la de los vídeos VCR convencionales.
- **MPEG-4.** Se trata del estándar MPEG más reciente que deberá ser aprobado por la Organización Internacional de Normalización (ISO). Microsoft ha creado la primera implementación de este estándar para tecnologías de Windows Media en Estados Unidos con la publicación de la versión 3 del codificador de vídeo Microsoft MPEG-4. Este estándar fue desarrollado específicamente para la codificación eficaz de contenido multimedia con diversas velocidades de bits, que incluyen desde velocidades lentas en Internet hasta velocidades de reproducción de imagen completa con calidad de televisión. El codificador de vídeo MPEG-4 admite intrínsecamente la transmisión de multimedia ya que permite que existan varias secuencias en una sola secuencia de datos codificada. Este estándar también dispone de un algoritmo avanzado de estimación del movimiento que ofrece mayor compresión.

## DOCUMENTOS MULTILINGÜES

Los documentos intercambiados en el medio de Internet muchas veces precisan ser una mezcla multilingüe, es decir, deben ser presentados utilizando más de un lenguaje en un mismo párrafo, página o documento. Un ejemplo típico es una lista de partici-

pantes de un encuentro o reunión en línea, donde cada participante debe ser mencionado en su lenguaje/código.

La mayoría de los navegadores de hoy día tienen soporte «multilingüe». Se accede a sus funciones «multilingües» al seleccionar un lenguaje adecuado en el menú. Por esto, deberían ser denominadas de multi-localización más que multilingües, ya que no pueden presentar adecuadamente un documento que contenga partes de lenguajes diferentes en una misma página o en el mismo documento. Para salvarlo se ha de utilizar una codificación de caracteres determinada en nuestro documento.

Una codificación de caracteres es un código que empareja un conjunto de caracteres de un lenguaje natural determinado con un conjunto de números o impulsos eléctricos. Ejemplos típicos incluyen el código Morse que codifica las letras como impulsos eléctricos cortos o largos o el código ASCII<sup>47</sup> que codifica las letras, números y otros caracteres como números enteros o su representación binaria en 7 bits. ASCII se utiliza desde los primeros días de la informática ya que fue normalizado por la *American Standards Association* (luego ANSI) en 1963. Con 7 bits el sistema es capaz de codificar hasta 127 caracteres, que solamente es suficiente para la lengua inglesa.

Esta limitación se intentó solucionar con el desarrollo de la norma ISO/IEC 8859 que utiliza una codificación de 8 bits. Esta norma tiene en la actualidad 15 partes numeradas correlativamente: 8859-1, 8859-2... cada una de las cuales describe caracteres de una lengua o de un grupo de lenguas. Así, por ejemplo 8859-1, también llamada Latín-1 permite describir todos los caracteres de las lenguas de la Unión Europea. Otras partes son por ejemplo, 8859-7 para el griego, 8859-8 para el hebreo o 8859-16 para las lenguas del sureste de Europa (croata, albanés, húngaro, etc.).

El principal inconveniente de esta norma es que cada parte es independiente, lo que hace imposible utilizar en un mismo fichero o documento caracteres procedentes de varias lenguas, por ejemplo mezclar griego con español. Un paso más en la codificación se dio con la aparición de *Unicode* que utiliza una codificación de 16 bits y permite representar más de 100.000 caracteres. Los primeros 128 caracteres son idénticos a ASCII y los 256 restantes lo son a ISO 8859-1. La primera versión de esta norma la publicó el *Unicode Consortium*<sup>48</sup> en 1991. La última es la número 4.0.1 de 2003<sup>49</sup>. La Organización Internacional de Normalización ha desarrollado la norma ISO 10646 que define un repertorio de caracteres idéntico a *Unicode*.

Su objetivo es permitir la representación de cualquier carácter utilizado por cualquier persona en cualquier parte del mundo y utilizando cualquier lengua. En la actualidad se ha convertido en el sistema de codificación más utilizado y está soportado por los principales sistemas operativos como todas las versiones de Windows NT y sucesores, Linux, BSD, etc. No obstante no está exenta de críticas principalmente procedentes de Japón y otras lenguas orientales alegando problemas técnicos y políticos derivados del hecho de que los principales impulsores de la norma son empresas estadounidenses como Microsoft. De hecho existe en estos países iniciativas para crear alternativas a *Unicode*, entre ellas UTF-2000<sup>50</sup> o *Giga Character Set*<sup>51</sup>.

En resumen, *Unicode* proporciona un código numérico para cada carácter de cualquier lengua, el cómo esos números son almacenados en los sistemas de tratamiento de textos es otro problema. Los problemas resultan del hecho de que la mayor parte del software utilizado fue escrito para tratar con códigos de 8 bits y el soporte para 16 bits se ha ido incluyendo progresivamente en los últimos años.

La conversión entre números *Unicode* y bits se realizan a través de un sistema llamado UTF (*Unicode Transformation Format*) del que existen las siguientes variantes: UTF-32, UTF-16 y UTF-8.

UTF-8 es una codificación de *Unicode* en 8 bits. Los primeros 128 códigos son idénticos a ASCII con lo cual son sistemas compatibles. Por otro lado UTF-8 es incompatible con ISO 8859-1 después del código 127. UTF-16 por su parte utiliza 16 bits. Es compatible hacia atrás tanto con ASCII como con ISO 8859-1. Además puede añadir hasta un millón de caracteres utilizando un mecanismo conocido como pares sustitutos (dos unidades de 16 bits por cada carácter).

## EL DISEÑO DE LA PRESENTACIÓN

A lo largo de la exposición venimos haciendo hincapié en la distinción entre contenidos de un documento y presentación de los mismos. Hemos visto formatos de documentos como XML o SGML que se centraban en describir la estructura de los documentos dejando la presentación para otras aplicaciones. En esta sección tratamos precisamente las formas de presentación de la información. En el primer apartado describimos las hojas de estilo utilizadas tanto para visualizar en la Web documentos XML como para definir la apariencia de los documentos HTML. En segundo lugar hacemos un repaso a las principales formas de organización de la información en un sitio web y terminamos con una serie de recomendaciones sobre la utilización de buscadores.

### HOJAS DE ESTILO

Las hojas de estilo sirven para controlar la apariencia de los documentos web tanto si se utiliza HTML como XML. Nielsen<sup>52</sup> hace una serie de recomendaciones generales sobre la utilización de hojas de estilo:

- Utilizar una única hoja de estilo para todas las páginas (o un conjunto coordinado de hojas si se trata de páginas con necesidades muy diferentes). Una de las principales ventajas de las hojas es asegurar la continuidad visual a medida que el usuario navega por el sitio web.
- Utilizar siempre hojas de estilo enlazadas con los documentos y nunca embebidas en los mismos. Así, sólo será necesario mantener un único fichero y se aligerarán las páginas de información redundante que solamente ralentizarán su descarga. No obstante, se debe permitir a los autores incluir estilos adicionales embebidos en las páginas cuando sea necesario.
- Para cada sitio es conveniente que todas las hojas de estilo las diseñe la misma persona o grupo de diseñadores. Esto se debe a dos razones, el trabajo centralizado es la única forma de asegurar un diseño consistente para todo el sitio, y, además, el creador de contenidos no siempre tiene que ser un buen creador de hojas de estilo.
- Las páginas deben continuar funcionando cuando las hojas de estilo estén deshabilitadas.
- En cuanto a la implementación de hojas:
  - No se deben usar más de dos fuentes diferentes en cada página.
  - No usar tamaños de fuentes absolutos.
  - Si se utilizan múltiples hojas habrá que asegurar que se asignan los mismos nombres a los mismos conceptos en todas ellas.

A continuación detallamos dos tipos de hojas de estilo, CSS (*Cascading Style Sheets*) y XSLT (*Extensible Stylesheet Language Transformations*) aplicables a documentos HTML y XML, respectivamente.

CSS es una recomendación<sup>53</sup> del W3C. Existen dos versiones de esta norma: CSS1 apareció en 1996 y actualmente está soportada en prácticamente todos los visualizadores web como Internet Explorer, Netscape u Opera. CSS2 apareció en 1998 para ampliar las funcionalidades de la primera. No obstante, los requerimientos que establecía eran tan altos que su implementación fue prácticamente nula. Para intentar solucionarlo en 2003 se hizo una revisión que quitó todos los aspectos de alto nivel que nunca iban a ser implementados por los navegadores.

En la práctica, CSS se materializa en un fichero de texto que debe tener la extensión .css y que debe estar escrito siguiendo la especificación CSS. Un ejemplo sencillo sería:

```
BODY
{ font-family: Verdana, "Minion Web", Helvetica, sans-serif; font-size: 1em; text-align: justify}
H1
{ font-family: Verdana, sans-serif; font-size: 1.3em }
```

En él se define la apariencia de los elementos <body> y <h1> de un documento HTML. Por cada elemento se especifica la fuente que se va a utilizar, su tamaño y la alineación del texto.

Además de cómo ficheros .css independientes, el estilo se puede embeber en cada documento particular dentro del elemento <head> de HTML.

En los recursos mencionados sobre CSS que están disponibles en las páginas del W3C, se puede consultar un manual muy completo y con numerosos ejemplos elaborado por la consultora Westciv<sup>54</sup>.

XSLT, por su parte, también es una recomendación del W3C aprobada en 1999<sup>55</sup>. Su objeto es distinto de CSS en el sentido de que está pensada para transformar un documento XML en otro con formato HTML o en cualquier otro formato que use texto plano. Por lo tanto, más que una hoja de estilo, XSLT es un verdadero lenguaje de programación mucho más robusto que CSS pero también mucho más complicado. Dependiendo de las necesidades se elegirá uno u otro. Si únicamente se necesita añadir elementos de estilo a un documento HTML, la opción CSS será suficiente. Si estamos trabajando con XML entonces deberemos emplear XSLT.

¿Cómo funciona XSLT? Supongamos que disponemos de un documento XML sencillo del tipo:

```
<source>
<title>XSLT</title>
<author>John Smith</author>
</source>
```

a través de una hoja de estilo XSLT como la siguiente:

```
<xsl:stylesheet version = '1.0'
```

```

xmlns:xsl='http://www.w3.org/1999/XSL/Transform'>
<xsl:template match="/">
<h2>
<xsl:value-of select="//author"/>
</h2>
<h1>
<xsl:value-of select="//title"/>
</h1>
</xsl:template>
</xsl:stylesheet>

```

podríamos convertirlo en el siguiente documento HTML:

```

<h2>John Smith</h2>
<h1>XSL</h1>

```

Una breve descripción del contenido de la hoja de estilo sería:

- ❑ La primera instrucción es obligatoria e indica la versión y el espacio de nombres de XML que estamos usando.
- ❑ A continuación, el elemento *match* nos indica qué parte del documento XML fuente vamos a procesar. En este caso se indica la raíz del documento con /, por lo que se procesa en su totalidad, pero podríamos haber procesado sólo determinados elementos.
- ❑ Por cada elemento que encuentre en el documento fuente, si se trata de un *author* lo pondrá entre las etiquetas <h2></h2> en el documento HTML resultante. Si se trata de un *title* lo pondrá entre <h1></h1>.

Para más información sobre XSLT se pueden consultar dos manuales básicos de Michael Kay<sup>56</sup> y Sal Mangano<sup>57</sup>. Por otro lado, el W3C mantiene una recopilación exhaustiva de recursos accesibles en línea<sup>58</sup> sobre el tema.

## TIPOS DE VISUALIZACIÓN

El potencial expresivo que permiten los enlaces en un hipertexto es peligroso. La teoría hipertextual se ha dedicado, en gran medida, a establecer modelos conceptuales de las posibilidades de orientación del usuario en función, entre otros aspectos, de la estructura de los enlaces. La tecnología web, con todas sus limitaciones, no permite elaborar estructuras demasiado complicadas. A continuación se detallan las cuatro estructuras básicas más habituales en el diseño de páginas web<sup>59</sup>:

- ❑ Secuencial. La manera más simple de organizar la información es colocarla secuencialmente. La ordenación puede ser cronológica, alfabética, temática, etc. Las secuencias son la forma más apropiada de organización en documentos didácticos o de formación, donde el lector ha de conocer por fuerza una serie de datos antes de avanzar en la lectura. Otros sitios web más complejos pueden estar organizados secuencialmente también, de manera que cada página de la secuencia principal puede contener enlaces a páginas de explicaciones, información parentética (susceptible de ir entre paréntesis, como este comentario) o información sobre otros webs.

- ❑ **Rejillas.** Las rejillas son una buena forma de correlacionar variables. Por ejemplo, una rejilla puede estructurarse según la relación que existe entre una línea de evolución temporal (siglos, años, etc.) por un lado y un conjunto de información histórica clasificada por temas (arte, cultura, hechos históricos, etc.) por el otro. Para que una rejilla sea efectiva las unidades básicas de información deben compartir una estructura de temas y subtemas muy uniforme y los lectores han de comprender la naturaleza de la estructura general y el significado de las variables correlacionadas. Si no sucede así, se perderán fácilmente. Por ello, esta estructura es adecuada sobre todo para público experto, con conocimientos del tema objeto del documento y su articulación lógica. Los mapas conceptuales pueden ser muy útiles para facilitar la orientación al lector si se utiliza esta estructura.
- ❑ **Jerarquías.** Las jerarquías son la mejor forma de organizar corpus complejos de información. Dado que los sitios web están organizados, habitualmente, en torno a una única página principal, los esquemas jerárquicos son adecuados para este tipo de documentos; son fáciles de entender y muy habituales en todos los aspectos de la vida por lo que los usuarios están acostumbrados a trabajar con ellos. Una organización jerárquica impone también una disciplina útil en la aproximación analítica del diseñador de contenidos, ya que las jerarquías sólo son útiles sobre material muy bien organizado.
- ❑ **Mallas.** La estructura en malla impone muy pocas restricciones. El objetivo es imitar el pensamiento asociativo y el flujo libre de ideas. Así, los lectores pueden seguir sus propios intereses. Irónicamente, los esquemas de organización asociativos son a menudo la estructura más impracticable para un sitio web, porque suelen ser difíciles de entender y de predecir. No obstante, son útiles para webs pequeños que estén compuestas fundamentalmente por listas de enlaces, o bien para sitios dirigidos a usuarios muy expertos en el tema, que necesiten una información muy avanzada o complementaria y no una mera presentación de los aspectos básicos.

Además de la organización del sitio es de vital importancia disponer de un sistema de navegación apropiado. El web es, básicamente, un sistema de navegación. La interacción del usuario consiste en hacer clic en los vínculos de hipertexto para moverse por un amplio espacio de información. Dado que ese espacio es tan grande, la navegación se hace difícil y es necesario ofrecer algún tipo de soporte. Las interfaces de navegación han de ayudar al usuario a responder a tres preguntas fundamentales:

- ❑ **¿Dónde estoy?** La ubicación del usuario se ha de mostrar respecto de la Web como un todo, para lo cual es necesario que en todas las páginas se identifique el sitio en el que nos encontramos (a través de logos, etc.). Y también en relación a la estructura del propio sitio web. La ubicación relativa se suele dar mostrando esquemas del sitio donde está resaltado el lugar donde se encuentra el usuario. También es importante tener un título claro en la página que se ha de utilizar para darle un nombre significativo, de forma que los usuarios puedan localizarlas fácilmente en su lista de marcadores.
- ❑ **¿Dónde he estado?** Dada la tecnología web actual esta pregunta es difícil de responder si no es acudiendo a la utilización de *cookies* o alguna otra medida de control. Afortunadamente los mecanismos de desplazamiento de los navegadores sirven para aclarar dudas al respecto. Además disponen de un historial donde se incluye una lista de las páginas visitadas recientemente.
- ❑ **¿Dónde puedo ir?** A esta pregunta responden las opciones de navegación y todos los demás vínculos que puede haber en la página. En una página se pueden utilizar dos tipos de vínculos:

- Los asociativos constituyen el texto subrayado normal que indica que hay más cosas sobre un tema que se menciona en el cuerpo del texto. En particular se pueden utilizar para dar al usuario pistas de «véase además» sobre páginas que les puedan interesar también por estar relacionadas con la página actual.
- Los estructurales son vínculos que señalan sistemáticamente otros lugares de la estructura del sitio, además de los iguales o secundarios dentro de una jerarquía. Es importante tener los mismos vínculos estructurales en todas las páginas de forma que el usuario entienda qué opciones de navegación estructural puede esperar.

## USO DE BUSCADORES

Una forma adicional de ayudar a localizar información en un sitio web es la utilización de algún tipo de motor de búsqueda. Los usuarios acceden al sistema de búsquedas entrando una o más palabras clave en un formulario. El motor de búsqueda entonces selecciona y muestra las páginas que contienen dichas palabras.

La opción de búsqueda debería proporcionarse sólo allí donde sea útil. Lo será en sitios muy ricos en contenidos y ofrecerá poco valor añadido en el resto. Los diseñadores de páginas deberían usar esta opción con cautela. Un motor de búsqueda nunca puede ser un sustituto de una buena organización de los contenidos y no siempre mejoran la efectividad de la búsqueda por parte de los usuarios. Antes de incluir un sistema de este tipo se deberían evaluar detenidamente sus ventajas e inconvenientes.

Si se decide ir adelante con un buscador, cada página del sitio debería contener un formulario de búsqueda. Normalmente un formulario básico es suficiente para llevar a cabo búsquedas sencillas. Si se requieren búsquedas más complejas, es conveniente proporcionar una búsqueda avanzada en una página específica. Cuando existan varias posibilidades de búsqueda habrá que advertírsele al usuario.

Cuando múltiples usuarios tienden a ejecutar búsquedas similares suele ser útil ofrecer patrones de búsqueda.

Los formularios deberían ser fáciles de utilizar. Es un hecho demostrado que muchos usuarios tienden a usar estrategias de búsqueda simples, empleando muy pocas palabras. Sólo excepcionalmente utilizan operadores booleanos, truncamientos u operadores de proximidad. Es conveniente proporcionar unas instrucciones simples y ejemplos que sirvan de guía.

Es fundamental que el motor de búsqueda no haga distinciones entre mayúsculas y minúsculas. En caso contrario habrá que advertirlo claramente.

El sistema debe buscar en todo el sitio y, si no es este el caso, advertir claramente en qué partes son las que se está buscando. Esto no siempre es fácil, puesto que lo que un diseñador entiende por «todo el sitio» no tiene que coincidir necesariamente con lo que considera el usuario.

Por ejemplo, muchos sitios grandes tienen diferentes secciones, cada una mantenida por diseñadores distintos, por lo que el usuario puede interpretar que solamente hay un sitio cuando un diseñador puede interpretar que hay muchos.

Habrá que asegurar que los resultados de la búsqueda proporcionen la información precisa que los usuarios necesitan. Cuando esto no sucede así los usuarios quedan confundidos y frustrados.

## LA ADICIÓN DE VALOR: METADATOS

A medida que la colección de documentos electrónicos crece se hace necesario mantener un catálogo con sus descripciones bibliográficas que permita tanto al servicio como a los usuarios externos hacer consultas sobre el fondo. Utilizar algún sistema de metadatos permitirá mantener esos datos en un formato normalizado para poder intercambiarlos con terceras partes, como por ejemplo, otras editoriales o motores de búsqueda en Internet.

Los metadatos han estado con nosotros desde que los primeros bibliotecarios compilaron las primeras listas de contenidos de sus depósitos. En su acepción más general metadatos son datos sobre datos. En la era de Internet hace referencia a lo que los bibliotecarios han colocado en sus catálogos durante años con el nombre de información bibliográfica y que normalmente se refiere a información descriptiva sobre recursos web. Un registro de metadatos se compone de un conjunto de elementos necesarios para describir un recurso particular.

El enlace entre un registro de metadatos y el recurso o documento que describe puede hacerse bien en forma de registro de una base de datos externa al recurso o bien embebido dentro del mismo recurso. Mientras la primera opción es la más utilizada por las bibliotecas para construir bases de datos de recursos electrónicos, la segunda es la más apropiada para las editoriales y otros proveedores de contenidos. De hecho, las editoriales vienen proporcionando metadatos embebidos dentro de las publicaciones impresas en la forma de CIP (*Catalogación en Publicación*)<sup>60</sup> con el objeto de proporcionar a los bibliotecarios unos datos bibliográficos básicos para incluir en los catálogos. En el ámbito electrónico la iniciativa TEI permite embeber metadatos en el encabezamiento de un texto electrónico.

En la actualidad, existen multitud de estándares de metadatos. Por ejemplo la comunidad de bibliotecarios utilizan de forma generalizada el formato MARC (*Machine Readable Cataloguing*)<sup>61</sup>. Por otro lado, en el ámbito de la descripción de recursos electrónicos se ha generalizado el uso del *Dublin Core*<sup>62</sup>.

*Dublin Core* es un conjunto simple pero efectivo para describir un amplio rango de recursos disponibles en la red. Incluye dos niveles, simple y cualificado. El *Dublin Core* simple se compone de quince elementos de descripción, mientras que el cualificado añade uno más (audiencia) así como una serie de limitadores (cualificadores) que refinan la semántica de los elementos de alguna forma que facilite la localización de los recursos.

Los objetivos de la iniciativa son básicamente cuatro:

- Crear un sistema de metadatos simple tanto para la creación de registros como para su mantenimiento.
- Una semántica de los elementos común para toda la comunidad de usuarios de Internet. Por ejemplo, mientras en el área científico-técnica se puede estar interesado en localizar artículos de un particular autor, en museos se puede necesitar saber

qué obras pertenecen a un determinado artista. Así ambas comunidades pueden ponerse de acuerdo en utilizar un elemento de descripción denominado «creador».

- ❑ Un alcance internacional. Aunque fue desarrollado originalmente en inglés a partir de una iniciativa de OCLC (*Online Computer Library Center*) y NCSA (*National Center for Supercomputing Applications*) a mediados de los años noventa, se han elaborado versiones en otros idiomas, entre ellos el español<sup>63</sup> y catalán<sup>64</sup>.
- ❑ Una «extensibilidad». Aunque simple, se reconoce la posibilidad de que en el futuro se necesiten nuevos elementos para describir nuevos recursos o para mejorar la recuperación de los mismos.

A continuación se detallan los quince elementos básicos de *Dublin Core*. Hay que tener en cuenta que todos ellos pueden matizarse con algún cualificador<sup>65</sup>:

- ❑ DC.Title. El nombre dado a un recurso, usualmente por el autor.
- ❑ DC.Creator. La persona u organización responsable de la creación del contenido intelectual del recurso. Por ejemplo, los autores en el caso de documentos escritos, artistas, fotógrafos e ilustradores en el caso de recursos visuales.
- ❑ DC.Subject. Los tópicos del recurso. Típicamente, Subject expresará las claves o frases que describen el título o el contenido del recurso. Se fomentará el uso de vocabularios controlados y de sistemas de clasificación formales.
- ❑ DC.Description. Una descripción textual del recurso, tal como un resumen en el caso de un documento o una descripción del contenido en el caso de un documento visual.
- ❑ DC.Publisher. La entidad responsable de hacer que el recurso se encuentre disponible en la red en su formato actual, por ejemplo la empresa editora, un departamento universitario u otro tipo de organización.
- ❑ DC.Contributor. Una persona u organización que haya tenido una contribución intelectual significativa en la creación del recurso pero cuyas contribuciones son secundarias en comparación a las de las personas u organizaciones especificadas en el elemento *Creator* (por ejemplo, editor, ilustrador y traductor).
- ❑ DC.Date. Una fecha en la que el recurso se puso a disposición del usuario en su forma actual. Esta fecha no ha de confundirse con la que pertenece al elemento Coverage, que sería asociada con el recurso sólo en la medida en que el contenido intelectual está de algún modo relacionado con esa fecha.
- ❑ DC.Type. La categoría del recurso, por ejemplo, página personal, romance, poema, minuta, diccionario. Para asegurar la interoperabilidad, *Type* debería ser seleccionado de entre una lista de valores que actualmente se encuentra bajo desarrollo en un grupo de trabajo.
- ❑ DC.Format. El formato de datos de un recurso, usado para identificar el software y posiblemente, el hardware que se necesitaría para mostrar el recurso. Para asegurar la interoperabilidad, los valores de *Format* deberían ser seleccionados de entre una lista de valores que actualmente se encuentra bajo desarrollo en un grupo de trabajo.
- ❑ DC.Identifier. Secuencia de caracteres usados para identificar unívocamente un recurso. Ejemplos para recursos en línea pueden ser URLs y URNs (cuando estén implementados). Para otros recursos pueden ser usados otros formatos de identificadores, como por ejemplo ISBN (*International Standard Book Number*, Número Internacional Normalizado para Libros).
- ❑ DC.Source. Secuencia de caracteres utilizado para identificar unívocamente un trabajo a partir del cual proviene el recurso actual.

- ❑ DC.Language. Lengua/s del contenido intelectual del recurso. Prácticamente el contenido de este campo debería coincidir con los de la RFC 1766<sup>66</sup>; por ejemplo: en, es, de, fi, ja y zh.
- ❑ DC.Relation. Un identificador de un segundo recurso y su relación con el recurso actual. Este elemento permite enlazar los recursos relacionados y las descripciones de los recursos. Por ejemplo:
  - IsVersionOf. Incluye la edición de un trabajo.
  - IsBasedOn. La traducción de un trabajo.
  - IsPartOf. Un capítulo de un libro.
  - IsFormatOf. Un mecanismo de transformación de un conjunto de datos en una imagen.

Para asegurar la interoperabilidad, las relaciones deberían ser seleccionadas de una lista de elementos que actualmente se encuentra bajo desarrollo en un grupo de trabajo.

- ❑ DC.Coverage. La característica de cobertura espacial y/o temporal del contenido intelectual del recurso. La cobertura espacial se refiere a una región física, uso de coordenadas (por ejemplo, longitud y latitud) o nombres de lugares extraídos de una lista controlada. La cobertura temporal se refiere al contenido del recurso en vez de a cuando fue creado o puesto accesible ya que este último pertenece al elemento *Date*.
- ❑ DC.Rights. Una referencia (URL, por ejemplo) para una nota sobre derechos de autor, para un servicio de gestión de derechos o para un servicio que dará información sobre términos y condiciones de acceso a un recurso. Una especificación formal del elemento *Rights* se encuentra actualmente en discusión y, por lo tanto, su uso se considera experimental.

En cuanto a la sintaxis, la norma no especifica ninguna en particular y dependerá de la forma en que se utilice el *Dublin Core*. Ya hemos dicho que en el ámbito de esta guía lo más acertado sería embeber los elementos en los propios documentos, así podríamos tener diferentes sintaxis dependiendo que usemos HTML<sup>67</sup> o XML<sup>68</sup>.

## ESTÁNDARES DE ACCESIBILIDAD Y USABILIDAD

A la hora de diseñar un documento y, en general, un sitio web debemos hacernos la pregunta: ¿cómo es de accesible? En la búsqueda de la mayor audiencia posible debemos pensar que puede ser visitado por personas con problemas de audición, de visión u otro tipo de discapacidad. Algunos estudios sugieren que el 4,5% de los visitantes tienen algún tipo de problema con la percepción de los colores<sup>69</sup>. Además, hay que tener en cuenta que no todo el mundo dispone del mismo visualizador, el mismo ordenador o el mismo tipo de conexión a Internet. El W3C advierte a los desarrolladores que muchos usuarios estarán funcionando desde contextos muy diferentes, por ejemplo, pueden:

- ❑ No ser capaces de ver, oír, moverse, o ser incapaces de procesar algún tipo de información.
- ❑ Disponer solamente de una pantalla en modo texto, una pantalla pequeña o una conexión muy lenta.
- ❑ Tener dificultades para leer el texto.
- ❑ No hablar o entender perfectamente el lenguaje en que se ha escrito el documento.

- No disponer o no ser capaces de usar un teclado o un ratón.
- Estar en una situación donde sus ojos, oídos o manos estén ocupados.
- Tener una versión antigua de visualizador, diferentes modelos de visualizador incluyendo visualizadores basados en voz, o disponer de un sistema operativo totalmente diferente.

Teniendo en cuenta toda esta casuística, el W3C<sup>70</sup> hace las siguientes recomendaciones para garantizar la accesibilidad a un sitio web:

- Proporcionar alternativas cuando se utilice un contenido audiovisual. En general, los documentos audiovisuales necesitan unos requisitos de hardware y software que no están al alcance de todos los usuarios, por ello es necesario proporcionar la información equivalente en texto.
- Asegurar que texto y gráficos son legibles cuando se visualicen sin colores.
- Utilizar las marcas de HTML para identificar la estructura del texto de forma apropiada. Igualmente hay que controlar la apariencia del texto utilizando hojas de estilo en lugar de elementos de presentación propios de HTML.
- Utilizar marcas que faciliten la pronunciación o interpretación de textos abreviados o escritos en otras lenguas. Por ejemplo utilizando las marcas «xml:lang» en XML o utilizando los elementos ABBR y ACRONYM de HTML.
- Asegurarse de que las tablas utilizan un marcado apropiado que permita visualizarlas en todo tipo de plataformas.
- Asegurarse de que las páginas sean accesibles incluso cuando las más modernas tecnologías no estén soportadas o estén deshabilitadas en el sistema del usuario.
- Asegurarse de que los objetos con movimiento o que se actualizan automáticamente pueden ser pausados o parados por el usuario.
- Diseñar páginas que sean independientes de los periféricos. Por ejemplo, si un control sólo puede ser activado con un ratón, alguien que esté usando la página con otro dispositivo de puntero o desde un sistema adaptado para invidentes, o simplemente usando un teclado, no sería capaz de utilizarlo.

Además de accesible el sitio web debe ser fácilmente utilizable. El concepto de usabilidad se ha popularizado en los últimos años. A grandes rasgos define todas aquellas características que hacen que los usuarios que acuden a nuestro sitio web se sientan cómodos en él y encuentren la información que necesitan de la forma más rápida posible. A no ser que un sitio web responda a las necesidades de los usuarios a los que va dirigido, tampoco cumplirá con las necesidades de la organización que lo mantiene.

Que un sitio sea usable es hoy en día uno de los principales requerimientos para la aceptación por parte de los usuarios. Por ejemplo, en comercio electrónico, un usuario insatisfecho con nuestro sitio por ser poco amigable, seguramente se convertirá en un cliente de la competencia.

Entre los errores que pueden hacer poco utilizable un sitio web y que cabría evitar desde el principio, tenemos<sup>71</sup>:

- Reproducir en el sitio los intereses y la estructura interna de la organización en lugar de orientarlos a solucionar las necesidades de los usuarios.
- Ofrecer materiales no adaptados para ser presentados en la Web.
- No someter el proceso de creación de la página web a los criterios de calidad que

se utilizan para otros tipos de publicaciones. Así la publicación en web proporciona una oportunidad única para que proveedores de contenidos con poca experiencia en este campo creen una nueva generación de sistemas de difícil utilización. Para crear páginas web es necesario la combinación de expertos en la materia a tratar, HTML/XML, diseño gráfico y arquitectura y usabilidad de la información.

Un buen diseño, junto con el establecimiento de alguna metodología de evaluación ayudarán a crear un sitio web usable<sup>72, 73</sup>. Existen multitud de guías<sup>74, 75</sup> de usabilidad de sitios web. Es de obligada referencia la página de Nielsen: useit.com<sup>76</sup>, quien es autor también de varios libros sobre el tema. A la hora de proporcionar unas pautas básicas que todo sitio debería seguir podemos seguir las *Research-Based web design and usability guidelines*<sup>77</sup>:

- Durante el proceso de diseño y evaluación se deben establecer unos objetivos a cumplir por el sitio web, proporcionar un contenido útil que entienda y responda a las necesidades de los usuarios.
- Optimizar el tiempo que el usuario pase en el sitio web a través de:
  - Mostrar la información en un formato directamente utilizable, es decir, sin que el usuario necesite instalar aplicaciones adicionales.
  - No mostrar ventanas o gráficos innecesarios.
  - Proporcionar ayuda a los usuarios y hacerlo en un lenguaje comprensible para ellos.
  - Minimizar el tiempo de descarga de las páginas.
  - Reducir la carga de trabajo y cognitiva del usuario.
  - Informar sobre los tiempos de descarga estimados cuando se trate de ficheros muy grandes.
  - Desarrollar páginas que se puedan imprimir correctamente.
- En cuanto al hardware y software, se debe diseñar para:
  - Los visualizadores más comunes y teniendo presente también los más minoritarios.
  - Los sistemas operativos más habituales.
  - Las velocidades de conexión más frecuentes.
  - Las resoluciones de pantalla más comunes.
- En el diseño de las páginas se debe:
  - Establecer la longitud de la página y de las líneas más apropiada.
  - Establecer niveles de importancia, colocando la información más importante en el área superior de la pantalla.
- Navegación:
  - Mantener al usuario informado en todo momento del lugar donde se encuentra.
  - En caso de páginas muy largas se debe crear una lista de contenidos.
  - No crear páginas sin opciones de navegación (callejones sin salida).
  - Utilizar mapas del sitio.
  - Utilizar entradas apropiadas en los menús que de verdad respondan a lo que el usuario va a encontrar en los enlaces.
- Apariencia del texto:
  - Utilizar texto en negro sobre un fondo plano y de alto contraste.
  - Asegurar la consistencia visual.
  - Usar fuentes familiares al usuario y de, al menos, 12 puntos.
- Gráficos e imágenes:
  - Utilizar vídeo, animaciones y audio con mucha cautela.

- Limitar el uso de imágenes de gran tamaño.
- Añadir etiquetas a las imágenes que escondan algún enlace.
- Contenidos:
  - Usar acrónimos y abreviaturas sólo si es imprescindible.
  - Usar palabras familiares para el usuario evitando el lenguaje coloquial.
  - Organizar los contenidos de forma clara y precisa.
  - Colocar la información más importante al comienzo de la página.
  - Agrupar los elementos que estén relacionados.
  - Mostrar sólo la información que sea útil y necesaria.
  - Formatear la información teniendo en cuenta distintos tipos de audiencias.

## GLOSARIO

**Imagen digital.** Una imagen digital es una «fotografía electrónica» compuesta por un número determinado de puntos, llamados píxeles, que forman una matriz con filas y columnas. Cuanto mayor sea el número de filas y columnas mayor será el detalle de la imagen y mayor también su tamaño de archivo. Cada píxel de una imagen almacena información de su tono o luminosidad, donde el tono negro es el valor 0 y el blanco el valor más alto, normalmente 255 en la escala de grises, pero representados en formato binario. El formato binario es una notación numérica que utiliza solamente dos dígitos (cero y uno) para representar las cifras.

**Píxel.** Normalmente las cifras binarias se componen de un número total de ceros y unos que es potencia de 2, como 8, 16, 32 etc. A cada uno de estos unos o ceros se les llama bit y a un conjunto de 8 bits, byte. En un byte, el primer bit tiene un valor de 1, el segundo de 2, el tercero 4, el cuarto 8 y así sucesivamente. El número más bajo que puede contener un byte es 0 (00000000) y el más alto 255 (11111111). El número de bits utilizado para representar un píxel en una imagen determina el número de colores o variantes de gris que pueden ser representados en una imagen digital. Esto es lo que se denomina profundidad de color o de bit.

**Profundidad de color.** En una imagen en escala de grises cada punto se almacena en un byte, donde su valor numérico representa su tono que puede oscilar entre el blanco (255) y el negro (0). Esto quiere decir que es una imagen donde existen 256 tonos de gris. Es decir, la profundidad de color es el número de bits que definen cada píxel, que determinan el máximo número de colores que pueden tener. Si cada píxel viene determinado por 2 bytes (= 16 bits) existirán 65.536 tonos de gris y así sucesivamente. Una imagen digital en color se genera con sus componentes R, G y B (del inglés *red*, *green* y *blue*). Así pues se compondría de tres imágenes, una por cada color. La suma de las tres hace la imagen final. Cada píxel va definido por 3 bytes, uno por cada color. Si cada uno de ellos tiene una gama de color de 255 tonos, en la imagen final habrá una gama de 16.777.216 colores (256 x 256 x 256) que se suele abreviar como «16 millones de colores». Es lo que se llama una imagen con una profundidad de color de 24 bits (8 por cada color).

**Tamaño de imagen.** Se define como las dimensiones en píxeles de la matriz o cuadrícula. Si una imagen está formada por una matriz de 800 columnas por 500 filas, tiene entonces un tamaño de 800 x 500 píxeles.

**Tamaño de archivo.** Es la cantidad de información que contiene, medida en bits o en alguno de sus múltiplos: bytes, Megabytes, etc.

**Resolución.** Es la medida de la cantidad de píxeles por unidad de longitud, comúnmente píxeles por pulgada (una pulgada equivale a 2,54 cm de longitud). Se suele abreviar como ppp o dpi (puntos por pulgada o *dots per inch*). De esta definición se deduce que a mayor resolución, mayor número de puntos de imagen en el mismo espacio, y por tanto mayor definición. Por ejemplo, si una imagen tiene unas dimensiones en píxeles de 548 x 366 y se imprime o se está mostrando en pantalla con unas dimensiones de 2 x 3 pulgadas, entonces tiene una resolución de 72 dpi. Debe quedar claro que la resolución es la relación entre las medidas digitales (medidas en píxeles) y las físicas (las que tiene una vez impresa, o en el dispositivo que se está visualizando). La resolución no es una medida de la calidad de una imagen digital, aunque a menudo se utilice para ello. Es una medida de nitidez o definición, de forma que cuanto más alta sea, mayor definición y viceversa.

**Compresión.** La compresión se utiliza para reducir el tamaño de los ficheros de imágenes para ser procesados más fácilmente, almacenados o comunicados a través de Internet. Los métodos utilizados se basan, generalmente, en abreviar información repetida o eliminar información que el ojo humano tiene dificultades para apreciar. Por lo tanto, la calidad de una imagen puede verse afectada por las técnicas de compresión utilizadas o por el nivel de compresión aplicado. Las técnicas de compresión pueden ser «sin pérdida» que significa que cuando la imagen descomprimida será idéntica a su estado inicial ya que no se ha eliminado ninguna información o «con pérdida» cuando sí que se elimine alguna información.

---

<sup>1</sup> *Electronic publishing: guide to best practices for canadian publishers* (en línea). Natinal Library of Canada, 2001. <<http://www.collectionscanada.ca/9/13/index-e.html>> (Consulta: 28 de julio de 2004).

<sup>2</sup> *DigiCULT Home Page* (en línea). <[http://www.cordis.lu/ist/directorate\\_e/digi-cult/index.htm](http://www.cordis.lu/ist/directorate_e/digi-cult/index.htm)> (Consulta: 28 de julio de 2004).

<sup>3</sup> MINERVA EUROPE. *Digitisation Guidelines: a selected list* (en línea). Minerva Europe, 2003. <<http://www.minervaeurope.org/guidelines.htm>> (Consulta: 28 de julio de 2004).

<sup>4</sup> MINERVA EUROPE. *Good Practices Handbook version 1.3* (en línea). Minerva Europe, 2004. <[http://www.minervaeurope.org/structure/workinggroups/goodpract/document/good-practices1\\_3.pdf](http://www.minervaeurope.org/structure/workinggroups/goodpract/document/good-practices1_3.pdf)> (Consulta: 28 de julio de 2004).

<sup>5</sup> MINERVA EUROPE. *Technical Guidelines for Digital Cultural Content Creation Programmes* (en línea). Minerva Europe, 2004. <[http://www.minervaeurope.org/structure/workinggroups/servprov/documents/techguid1\\_0.pdf](http://www.minervaeurope.org/structure/workinggroups/servprov/documents/techguid1_0.pdf)> (Consulta: 28 de julio de 2004).

<sup>6</sup> INTERNATIONAL FEDERATION OF LIBRARIES ASSOCIATIONS. *Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives* (en línea). IFLA, 2002. <<http://www.ifla.org/VII/s19/pubs/digit-guide.pdf>> (Consulta: 28 de julio de 2004).

<sup>7</sup> *Digital Library of Georgia* (en línea). Digital Library of Georgia, 2001. <<http://dlg.galileo.usg.edu/guide.html>> (Consulta: 28 de julio de 2004).

<sup>8</sup> COINE. *Guide to digitization* (en línea). COINE, 2002. <[http://www.uoc.edu/in3/coine/eng/deliverables/Guide\\_to\\_digitization\\_Task3\\_6.pdf](http://www.uoc.edu/in3/coine/eng/deliverables/Guide_to_digitization_Task3_6.pdf)> (Consulta: 28 de julio de 2004).

<sup>9</sup> CORNELL UNIVERSITY LIBRARY. *Moving Theory into Practice: Digital Imaging Tutorial* (en línea). Cornell University Library, 2003. <<http://www.library.cornell.edu/preservation/tutorial/contents.html>> (Consulta: 28 de julio de 2004).

<sup>10</sup> FUTURA, Richard. *Defining and using structure in digital documents* (en línea), *Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, June 19-21, 1994 - College Station, Texas, USA*. <<http://www.cSDL.tamu.edu/DL94/paper/furuta.html>> (Consulta: 28 de julio de 2004).

<sup>11</sup> COOMBS, James H. ; RENEAR, Allen H. ; DEROSE, Steven J. *Markup systems and the future of scholarly text processing*, *Communications of the ACM*, 30, 1987, págs. 933-947.

<sup>12</sup> FUTURA, Richard. Op. Cit.

<sup>13</sup> KASDORF, Bill. *SGML and PDF: why we need both* (en línea), *Journal of Electronic Publishing*, 1998. Vol. 3, núm 4. <<http://www.press.umich.edu/jep/03-04/kasdorf.html>> (Consulta: 28 de Julio de 2004).

<sup>14</sup> *Internet Engineering Task Force* (en línea) <<http://www.ietf.org>> (Consulta: 28 de julio de 2004).

<sup>15</sup> *World Wide Web Consortium* (en línea) <<http://w3c.org>> (Consulta: 28 de julio de 2004).

<sup>16</sup> BERNERS-LEE, T. ; CONNOLLY, D. *Hypertext Markup Language - 2.0* (en línea) *Internet Engineering Task Force*, 1996. <<http://www.ietf.org/rfc/rfc1866.txt?number=1866>> (Consulta: 28 de julio de 2004).

<sup>17</sup> *HyperText Markup Language (HTML) Home Page* (en línea). <<http://www.w3.org/Mark-Up/>> (Consulta: 28 de julio de 2004).

- <sup>18</sup> PARNAS, Dagfinn. *How to cope with incorrect HTML* (en línea). Tesis de licenciatura, Universidad de Bergen, Departamento de Informática, 2001. <<http://www.ub.uib.no/el-pub/2001/h/413001/Hovedoppgave.pdf>> (Consulta: 28 de julio de 2004).
- <sup>19</sup> LIEB, Tom. *The X (HTML) Files* (en línea). *Journal of Electronic Publishing*, 2000. Vol 5, núm. 4. <<http://www.press.umich.edu/jep/05-04/lieb0504.html>> (Consulta: 28 de julio de 2000).
- <sup>20</sup> *HighWire Press* (en línea). <<http://highwire.stanford.edu>> (Consulta: 28 de julio de 2004).
- <sup>21</sup> *ScienceDirect* (en línea). <<http://www.sciencedirect.com>> (Consulta: 28 de julio de 2004).
- <sup>22</sup> *SpringerLink* (en línea). <<http://www.springerlink.com>> (Consulta: 28 de julio de 2004).
- <sup>23</sup> *What is Adobe PDF?* [en línea]. Adobe Systems, 2004. <<http://www.adobe.com/products/acrobat/adobepdf.html>> (Consulta: 28 de julio de 2004).
- <sup>24</sup> *PDF Reference : Adobe Portable Document Format 1.3*. Boston : Addison-Wesley 2000.
- <sup>25</sup> Macromedia Director (en línea). <<http://www.macromedia.com/software/director/productinfo/features/>> (Consulta: 28 de julio de 2004).
- <sup>26</sup> Macromedia Flash (en línea) <<http://www.macromedia.com/software/flash/productinfo/features/>> (Consulta: 28 de julio de 2004).
- <sup>27</sup> Showcase. Macromedia (en línea). <<http://www.macromedia.com/cfusion/showcase/index.cfm>> (Consulta: 28 de julio de 2004).
- <sup>28</sup> *Rich Text Format Specification, version 1.6* (en línea). Microsoft Corporation, 1999. <[http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnrtfspec/html/RTFSpec\\_1.asp](http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnrtfspec/html/RTFSpec_1.asp)> (Consulta: 28 de julio de 2004).
- <sup>29</sup> Microsoft Word (en línea). <[http://msdn.microsoft.com/library/en-us/dnanchor/html/odc\\_ancword.asp](http://msdn.microsoft.com/library/en-us/dnanchor/html/odc_ancword.asp)> (Consulta: 28 de julio de 2004).
- <sup>30</sup> Microsoft Reader (en línea). <<http://www.microsoft.com/reader>> (Consulta: 28 de julio de 2004).
- <sup>31</sup> Showcase. Microsoft Reader (en línea). <<http://www.microsoft.com/reader/us/shop/default.asp>> (Consulta: 28 de Julio de 2004).
- <sup>32</sup> HICKS, Tony. *Should we be using ISO 12083?* (en línea). *Journal of Electronic Publishing*, 1998. Vol. 3, núm 4. <<http://www.press.umich.edu/jep/03-04/hicks.html>> (Consulta: 28 de julio de 2004).
- <sup>33</sup> *SGML/XML: Academic Applications* (en línea). CoverPages, 2004. <<http://xml.coverpages.org/acadapps.html>> (Consulta: 28 de julio de 2004).
- <sup>34</sup> *Extensible Markup Language (XML) 1.0 (Third Edition)* (en línea), *World Wide Web Consortium*, 2004. <<http://www.w3.org/TR/REC-xml/>> (Consulta: 28 de julio de 2004).
- <sup>35</sup> *XSL Transformations (XSLT) Versión 1.0* (en línea). *World Wide Web Consortium*, 2004. <<http://www.w3.org/TR/xslt>> (Consulta: 28 de julio de 2004).
- <sup>36</sup> *Text Encoding Initiative* (en línea). <<http://www.tei-c.org/>> (Consulta: 28 de julio de 2004).
- <sup>37</sup> SPERBERG-MCQUEEN, C. M. and BURNARD, L. (eds.) (2002). *TEI P4: Guidelines for Electronic Text Encoding and Interchange* (en línea). Text Encoding Initiative Consortium. XML Version: Oxford, Providence, Charlottesville, Bergen. <<http://www.teic.org/P4X/>> (Consulta: 28 de julio de 2004).
- <sup>38</sup> *Projects using the TEI* (en línea). Text Encoding Initiative, 2004. <<http://www.teic.org/Applications/index.html>> (Consulta: 28 de julio de 2004).
- <sup>39</sup> *Teach yourself TEI* (en línea). Text Encoding Initiative, 2003.

- <http://www.teic.org/Tutorials/index.html> (Consulta: 28 de julio de 2004).
- 40 Open eBook Forum (en línea). <<http://www.openebook.org>> (Consulta: 28 de julio de 2004).
- 41 *Open eBook Publication Structure* (en línea). Open e-book Forum, Publication Structure Working Group, 2002. <<http://www.openebook.org/oebps/oebps1.2/index.htm>> (Consulta: 28 de julio de 2004).
- 42 *Cascading Style Sheets home page* (en línea). W3C, 2004. <<http://www.w3.org/Style/CSS/>> (Consulta: 28 de julio de 2004).
- 43 [http://www.openebook.org/oebps/oebps1.2/samples/Wuthering%20Heights%20-\(OEBPS%201.2\).zip](http://www.openebook.org/oebps/oebps1.2/samples/Wuthering%20Heights%20-(OEBPS%201.2).zip) (Consulta: 28 de julio de 2004).
- 44 CervanTeX (en línea). <<http://www.cervantex.org>> (Consulta: 28 de julio de 2004).
- 45 CASCALES SALINAS, Bernardo. *LaTeX: una imprenta en sus manos*. ADI, 2000.
- 46 MITTELBAACH, Frank et al. *The LaTeX companion, 2/E*. Addison Wesley Professional, 2004.
- 47 JENNINGS, Tom. *ASCII: American Standard Code for Information Interchange* (en línea) 2001. <<http://www.wps.com/projects/codes/>> (Consulta: 28 de julio de 2004).
- 48 Unicode Home Page (en línea). <<http://www.unicode.org>> (Consulta: 28 de julio de 2004).
- 49 *The Unicode Standard, version 4.0* (en línea). Unicode Consortium, 2003. <<http://www.unicode.org/versions/Unicode4.0.1/>> (Consulta: 28 de julio de 2004).
- 50 <http://www.wordiq.com/definition/UTF-2000> (Consulta: 28 de julio de 2004).
- 51 [http://www.wordiq.com/definition/Giga\\_Character\\_Set](http://www.wordiq.com/definition/Giga_Character_Set) (Consulta: 28 de julio de 2004).
- 52 NIELSEN, Jacob. *Effective use of style sheets* (en línea). 1997. <<http://www.useit.com/alertbox/9707a.html>> (Consulta: 28 de julio de 2004).
- 53 *Cascading Style Sheets, level 2 revision 1* (en línea). World Wide Web Consortium, 2004. <<http://www.w3.org/TR/CSS21/>> (Consulta: 28 de julio de 2004).
- 54 *Style master online tutorial* (en línea). Western Civilisation Pty. Ltd., 2004. <[http://www.westciv.com/style\\_master/academy/hands\\_on\\_tutorial/index.html](http://www.westciv.com/style_master/academy/hands_on_tutorial/index.html)> (Consulta: 28 de julio de 2004).
- 55 *XSL Transformations (XSLT) Version 1.0* (en línea). World Wide Web Consortium, 1999. <<http://www.w3.org/TR/xslt>> (Consulta: 28 de julio de 2004).
- 56 KAY, Michael. *XSLT Programmer's reference*. Wrox Press, 2000.
- 57 MANGANO, Sal. *XSLT Cookbook*. O'Reilly & Associates, 2002.
- 58 *THE EXTENSIBLE STYLESHEET LANGUAGE FAMILY (XSL)* (en línea). WORLD WIDE WEB CONSORTIUM, 1999. <<http://www.w3.org/Style/XSL/>> (Consulta: 28 de julio de 2004).
- 59 COUTIN DOMÍNGUEZ, Adrián. *Arquitectura de información para sitios web*. Anaya Multimedia, Madrid, 2002.
- 60 *Cataloguing in Publication Program* (en línea). The Library of Congress. <<http://cip.loc.gov/>> (Consulta: 28 de julio de 2004).
- 61 *MARC Standards* (en línea). Library of Congress. <<http://www.loc.gov/marc/>> (Consulta: 28 de julio de 2004).
- 62 Dublin Core Metadata Initiative (en línea). <<http://www.dublincore.org>> (Consulta: 28 de julio de 2004).
- 63 MASA, Javier. *Metainformación Dublin Core : Elementos del conjunto de metadatos de Dublin Core: descripción de referencia* (en línea). RedIris, 2003. <<http://www.rediris.es/-metadata/>> (Consulta: 28 de julio de 2004).
- 64 *Conjunt dels elements de metadades del Dublin Core, Versió 1.1: Descripció de referència* (en línea). Biblioteca de Catalunya, 2001. <<http://www.gencat.net/bc/dces.htm>>

(Consulta: 28 de julio de 2004).

<sup>65</sup> *Dublin Core Qualifiers* (en línea). Dublin Core Initiative, 2000. <<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>> (Consulta: 28 de julio de 2004).

<sup>66</sup> ALVESTRAND, H. *Tags for the Identification of Languages RFC 1766* (en línea). IETF, 1995. <<http://www.ietf.org/rfc/rfc1766.txt>> (Consulta: 28 de julio de 2004).

<sup>67</sup> KUNZE, J. *Encoding Dublin Core Metadata in HTML RFC 2731* (en línea). IETF, 1999. <<http://www.ietf.org/rfc/rfc2731.txt>> (Consulta: 28 de julio de 2004).

<sup>68</sup> POWELL, Andy. *Guidelines for implementing Dublin Core in XML* (en línea). Dublin Core Metadata Initiative, 2003. <<http://dublincore.org/documents/dc-xml-guidelines/>> (Consulta: 28 de julio de 2004).

<sup>69</sup> *Caslon Analytics online accessibility guide* (en línea). <<http://www.caslon.com.au/accessibilityguide1.htm>> (Consulta: 28 de julio de 2004).

<sup>70</sup> *Web Content Accessibility Guidelines 1.0 [en línea]*. World Wide Web Consortium, 1999. <<http://www.w3.org/TR/1999/WAI-WEBCONTENT-19990505/>> (Consulta: 28 de julio de 2004).

<sup>71</sup> *Design guidelines for the web* (en línea). UsabilityNet, 2003. <<http://www.usability-net.org/tools/webdesign.htm>> (Consulta: 28 de julio de 2004).

<sup>72</sup> SCAPIN, Dominic et al. *A Framework for Organizing Web Usability Guidelines* (en línea). <<http://www.tri.sbc.com/hfweb/scapin/Scapin.html>> (Consulta: 28 de julio de 2004).

<sup>73</sup> NIELSEN, J. ; MACK, R.M., *Usability Inspection Methods*. John Wiley & Sons, New York, 1994.

<sup>74</sup> *Web Design Guidelines: Design in Action* (en línea). <[http://www-3.ibm.com/ibm/easy/eou\\_ext.nsf/Publish/572](http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/572)> (Consulta: 28 de julio de 2004).

<sup>75</sup> *Web Style Guide* (en línea). <<http://www.webstyleguide.com/index.html>> (Consulta: 28 de julio de 2004).

<sup>76</sup> NIELSEN, J. *Useit.com* (en línea). <<http://useit.com>> (Consulta: 28 de julio de 2004).

<sup>77</sup> *Research-Based web design and usability guidelines* (en línea). National Cancer Institute, 2001. <<http://usability.gov/guidelines/>> (Consulta: 28 de julio de 2004).