

El análisis documental de documentos digitales y/o multimedia

Blanca Rodríguez Bravo*

RESUMEN

Se analizan las peculiaridades del documento digital y/o multimedia. Atendiendo a sus características particulares que condicionan su análisis y recuperación se establecen las etapas y procedimientos para representar el documento digital/multimedia con vistas a su localización y reutilización. Se aborda la autodescripción y el análisis de contenido.

ABSTRACT

Analysis of the digital and multimedia material's peculiarities. According with its characteristics that determine its analysis and retrieval, the stages and procedures to represent the multimedia/digital resource, with a view to its retrieval, are settle down. The article focuses in the autodescription and in the content analysis.

* Profesora Titular de Biblioteconomía y Documentación en la Universidad de León, España.

EL DOCUMENTO DIGITAL Y/O MULTIMEDIA

Hasta hace poco más de una década, la principal oposición se establecía entre los documentos en papel, que dominaban el mercado de la edición, a los que se denominaba documentos impresos, bibliográficos o librarios normalmente, y el resto de los documentos, que eran producto de los avances técnicos producidos en el siglo XX, y que se agrupaban, a efectos prácticos, bajo denominaciones en absoluto acertadas: documentos especiales, no impresos o no librarios, traduciendo de las expresiones inglesas *non-print materials* o *non-book materials* (Rodríguez Bravo, 2002:111).

Sin embargo, las diversas clasificaciones tipológicas que se han establecido utilizando como criterio el soporte deben ser revisadas, pues la introducción del documento digital ha hecho que el panorama experimente un vuelco considerable. Según Codina (2000: 21-22), la oposición más significativa no es ya la que existe entre medios electrónicos (los audiovisuales) y no electrónicos (los bibliográficos) sino la que se origina entre medios analógicos y medios digitales, entre átomos y bits en terminología de Negroponte (1995). No obstante, conviene señalar que los adjetivos «electrónico», «digital» e «informático» se intercambian. El concepto de documento electrónico es, no obstante, más amplio y englobaría al documento digital o informático y al audiovisual.

Lo que caracteriza a los medios analógicos es que representan la información mediante una relación de parecido o de analogía, se mantiene cierta semejanza entre la información y su codificación, aunque sea remota. En cambio, en los medios digitales, que utilizan series de bits para representar la información, cualquier semejanza entre la información y su representación queda destruida. Lo mismo un texto que una fotografía o un vídeo,

son series de ceros y unos para el ordenador, es decir que se representan de la misma manera.

En los medios analógicos; sin embargo, cada código y cada soporte de la información exigen una forma de codificación propia. De hecho, ningún soporte análogo resulta adecuado para todas las morfologías de la información al mismo tiempo. Por el contrario, los medios digitales pueden contener cualquier morfología de la información y cualquier combinación entre ellas (Codina, 2000: 22-23).

El documento digital tiene otras peculiaridades, puestas de relieve por Schamber (1996: 669): es fácilmente manipulable, es enlazable interna y externamente, es rápidamente transformable, es fácilmente accesible, instantáneamente transportable e infinitamente replicable.

De estas características se deduce que la primera diferencia de los documentos digitales con respecto a los analógicos estriba en que en los primeros se produce una disociación entre el soporte y el contenido. Las peculiaridades establecidas por Schamber se refieren a los contenidos, los soportes han perdido relevancia, de hecho, los mensajes que contienen se copian con facilidad en otro soporte, lo que los convierte en transformables o manipulables y en transportables.

La separación entre contenido y soporte hace que estos documentos sean compuestos (distintos códigos), distribuidos (en varios ficheros) y dinámicos (fácilmente modificables). De estas características se derivan las nuevas posibilidades que ofrecen, que son muchas. Sus puntos fuertes son su gran capacidad de almacenamiento, su virtualidad y accesibilidad a distancia y su actualización. Sus debilidades son la dificultad de conservación de sus mensajes (su duración) y la mayor dificultad que presentan para garantizar su

integridad y autenticidad. La facilidad que tienen los mensajes para cambiar de soporte y su sencilla manipulación están en el origen tanto de sus ventajas como de sus inconvenientes.

Son muchos y variados los soportes documentales y, además de la distinción entre documentos librarios y no librarios, «normales» y «especiales», otras agrupaciones que se suelen hacer no hacen ya referencia al soporte sino al medio y código que se utiliza para transmitir el mensaje. Entendemos que la sistematización en función del código empleado tiene mayor trascendencia en Documentación, pues los distintos códigos y las diferentes disposiciones de los mensajes en los soportes condicionan el tratamiento de los contenidos y su recuperación. Los códigos son los siguientes (Rodríguez Bravo, 2002a: 81):

- ◆ En el caso de que sea la escritura, nos encontramos con los documentos textuales o escritos, hasta ahora en papel fundamentalmente, con lo que equivalían a los bibliográficos o impresos, sin embargo, en la actualidad, lo textual tiene una presencia considerable en los documentos digitales.
- ◆ Si es el sonido, nos encontramos con los documentos sonoros que engloban discos, cassettes y CD-Audio, principalmente.
- ◆ Si se trata de la imagen, fija o en movimiento, se agrupan las fotografías, películas, diapositivas, etc.
- ◆ Existen por último los documentos mixtos, hasta hace poco audiovisuales exclusivamente, pero el documento digital permite la integración del documento sonoro, visual y textual, y aunque es éste último el que predomina, cada vez la tendencia a la multimedia es mayor.

Una de las singularidades de los documentos digitales es su carácter multimedia. Conviene que realicemos alguna precisión sobre este concepto. Señalan Pinto, García y Agustín (2002: 23) que el concepto multimedia es confuso y depende del enfoque adoptado, pese a que etimológicamente su significado es claro (muchos medios). Desde una perspectiva informativo-comunicacional se define el concepto multimedia como el agrupamiento sobre un soporte o un único modo de acceso de imágenes fijas o animadas, sonidos, textos y datos almacenados en forma digital. Como señalan los autores citados, es común también el término hipermedia para referirse al mismo concepto, aludiendo a la conexión de los multimedia por mediación de los procedimientos hipertextuales.

Últimamente, asistimos a un fenómeno trascendental, la migración de los sistemas textuales y audiovisuales a los nuevos entornos digitales. Precisamente la nueva categoría de documentos «multimedia» surge de la combinación de documentos textuales y audiovisuales posibilitada por los entornos informáticos. Los documentos digitales combinan los dos canales emisores (visual y acústico) y la trilogía de códigos disponibles (textual, icónico y musical). Los CD-ROM interactivos representan la versión digital del libro tradicional, pero el desarrollo de Internet y la *Web* ha dado origen a una alianza entre las aplicaciones multimedia y las redes en línea cuyo alcance supera las anteriores formas documentales disponibles.

Dado que el documento multimedia se ha materializado en el documento digital, es decir, ha sido viable gracias a la tecnología digital, vamos a abordar las peculiaridades del tratamiento de los documentos digitales, en general, y de los multimedia, en particular. Evidentemente, ambos conceptos no son idénticos. Por un lado, en los

documentos digitales con soporte intangible que circulan por Internet, el peso de lo textual es todavía muy grande, hay documentos sólo textuales como los hay sonoros o audiovisuales. Por otro lado, el término multimedia no se aplica en exclusividad a la integración de códigos bajo un mismo soporte o con un único medio de acceso, sino también a la unión de libros, vídeos y cassettes, cd-roms, etc. con un mismo contenido y finalidad.

No obstante, conviene señalar que la aparición del ordenador multimedia ha popularizado el concepto de multimedia y ha provocado un salto cualitativo y cuantitativo en la integración de los medios. La revolución de la WWW, entorno de comunicación multimedia e hipertextual, hipermedia, ha convertido esta realidad en omnipresente.

EL ANÁLISIS DOCUMENTAL DEL DOCUMENTO DIGITAL Y/O MULTIMEDIA

La problemática principal que presenta el documento digital no reside en el análisis de contenido, que, como veremos, seguirá las pautas establecidas para el tratamiento de los distintos códigos de transmisión de contenidos: textual, sonido, imagen fija o en movimiento, sino en su localización, organización y autodescripción para permitir su óptima recuperación.

LA AUTODESCRIPCIÓN DE LOS DOCUMENTOS DIGITALES Y/O MULTIMEDIA

En el documento digital han desaparecido las limitaciones espaciales y temporales intrínsecas al resto de los documentos. Ahora la preocupación se orienta a que puedan ser localizados en la ilimitada selva digital donde los contenidos se atomizan en un mosaico de elementos cuyo sentido es reconstruido libremente por el usuario gracias al hipertexto.

La colección de recursos multimedia *Web* de Internet no se han diseñado para permitir la recuperación organizada de la información. Sus contenidos están almacenados en una gran base de datos descentralizada, heterogénea, abierta, en evolución y sin grandes filtros de validación (Pinto, 1999: 277). De ahí que los mayores esfuerzos de los implicados en la nueva industria de la información se concentren en el tratamiento de este inmenso caudal informativo para dotarlo de una estructura que facilite su automatización.

La mayor carencia actual de la red es un sistema universal de etiquetado, representación y estructuración de la información que permita la búsqueda y el procesamiento automático más adecuado de cualquier documento *Web*.

La opción más utilizada para buscar información en la red es el uso de los motores de búsqueda, cuya eficacia en cuanto a pertinencia de los resultados es baja en comparación con los sistemas de recuperación de la información tradicionales, debido entre otros factores, a la falta de un mínimo tratamiento documental de la mayoría de los recursos de información almacenados en servidores *Web*.

Para solventar esta situación han aparecido diversas iniciativas dirigidas a la localización, identificación y descripción de los documentos digitales en Internet basadas, éstas últimas, en el uso de esquemas normalizados de metadatos. Los metadatos ayudan a la identificación, descripción y localización de los recursos en red, por medio de la estructuración de los datos de una forma similar a la que tradicionalmente se venía empleando con los restantes documentos, véase las normas ISBD, la norma ISAD, o el formato MARC.

Se han definido los metadatos de forma redundante como datos sobre datos o información sobre información. Muy ilustrativa es la definición que sugiere Garduño (2000: 125): «...los metadatos, considerados como el conjunto de elementos que pueden generar una semántica internacionalmente aceptada con el propósito de representar la información digital, evitar su dispersión a través de una sistematización apropiada y asegurar su recuperación. De manera general se puede señalar que los metadatos representan, por una parte, datos acerca de recursos informativos disponibles en redes, y contienen, por la otra, los elementos útiles para facilitar la identificación y localización de recursos digitales. En consecuencia los metadatos deben considerar el contenido, la condición, la cualidad y la calidad, entre otras características de la información digital».

Antes de introducirnos en el mundo de los metadatos vamos a mencionar las principales actuaciones conducentes a la identificación y localización del documento digital de Internet de forma inequívoca y para poder localizarlo con facilidad, que han sido las siguientes (Méndez; Merlo, 2000: 223-224):

Uniform Resource Identifiers (URI) (Internet Engineering Task Force): sistema para identificar documentos que combina URLs y URNs, consiste en una secuencia de caracteres con sintaxis controlada mediante la cual se reconoce un documento de forma única.

Uniform Resource Name (URN) (IETF): tentativa normativa que pretende suplir las limitaciones del sistema URL. Su finalidad es la consecución de un método único y duradero de identificación de recursos. Su sintaxis se estructura en tres bloques separados por dos puntos: el identificador URN, el NID o nombre de la categoría en la que se incluye el documento que se pretende identificar y el NSS o

cadena específica. Trabaja en paralelo con el *Uniform Resource Characteristics (URC)*, esquema para la descripción de metadatos.

Uniform Resource Locator (URL): es el actual sistema de localización de documentos en la Red. Indica el protocolo de acceso, el servidor donde está alojado el documento, sirviéndose del Sistema de Nombre de Dominios, el directorio o directorios, y por último el nombre y extensión del archivo a localizar. Se prevé su pronta sustitución por el sistema URN.

Persistent Uniform Resource Locator (PURL) (OCLC): método cuyos objetivos son los mismos del sistema URN. Se trata de un intermediario entre la dirección antigua de un documento *Web* y su nueva dirección. Está operativo en el servicio OCLC PURL. Esta iniciativa pretende paliar las deficiencias de la URL.

Digital Object Identifier (DOI): iniciativa de una serie de editores que persiguen también la identificación única y persistente, con el objetivo de facilitar el intercambio de información con fines fundamentalmente comerciales. Se estructura en un prefijo, en el que se identifica el directorio donde se localiza el documento, y un sufijo, en el que se expresa el código del objeto digital en cuestión.

Si abordamos ahora la descripción de los documentos *Web*, tanto en aspectos formales como semánticos, nos introducimos en el mundo de los metadatos. Existen al respecto varios modelos que constituyen proyectos normativos, siendo los principales: el modelo *Dublin Core (DC)*, el *Text Encoding Initiative (TEI)* y el *Resource Description Framework (RDF)*.

El establecimiento de metainformación no es algo totalmente nuevo, en los primeros lenguajes de

marcado ya existe la posibilidad de introducir materias o de identificar al autor de un documento, como es el caso del lenguaje HTML. La novedad reside en la tendencia a aumentar el número de campos que informen sobre el documento, normalizando el tipo de informaciones y el etiquetado de las mismas.

Factores fundamentales en el éxito de Internet han sido los lenguajes HTML, SGML y XML, que permiten la incorporación de estos metadatos en forma de etiquetas o *tags* que sirven para marcar las características fundamentales de los documentos digitales. La norma HTML no proporciona una lista cerrada de metaetiquetas, sino un método para declarar metaetiquetas con el propósito de representar la información sobre el documento.

Si bien no existe una lista cerrada de metaetiquetas, se está imponiendo el acuerdo tácito entre la comunidad de creadores de páginas *Web* en torno al conjunto mínimo siguiente de tres etiquetas: *<author>*, *<keywords>* y *<description>*. Como señala Codina (2000a: 56), cada vez hay más motores de búsqueda que entienden esas tres etiquetas y que les conceden una triple utilidad: en primer lugar, suelen otorgar una mayor relevancia a los documentos cuyas palabras están en esa sección; en segundo lugar, permiten una búsqueda más similar a la búsqueda por campos y, en tercer lugar, cuando un motor de búsqueda o un directorio encuentra la etiqueta utiliza el texto que contiene esa etiqueta como resumen del documento, en lugar de intentar generarlo de manera automática.

Gracias a los metadatos los objetos en una colección pueden distribuirse, pero la colección se presenta cohesionada lógicamente, porque las estructuras de metadatos están asociadas a ella y soportan una navegación coherente (Méndez, 2002: 303).

Abordaremos ahora brevemente las principales características de *Dublin Core*¹, intento multidisciplinar e internacional de especificación de un conjunto estándar de metadatos que sirvan para identificar con mayor precisión el contenido de los documentos y recursos *Web*. Los metadatos que especifica *Dublin Core* tienen la forma de metaetiquetas que se escriben como código HTML en la sección *<HEAD>* de los documentos *Web*; pero, a diferencia de las tres metaetiquetas señaladas antes, la norma *Dublin Core* propone un conjunto de 15 metadatos sobre los que existe suficiente acuerdo internacional para confiar en que pueden ser un instrumento de gran capacidad para describir documentos y recursos digitales en Internet. Su finalidad es facilitar el descubrimiento de recursos a través del uso de motores de búsqueda y bases de datos. Su uso por tanto, no sustituye a las operaciones de catalogación de documentos tal como suelen hacerse en el seno de las bibliotecas del sector público, que utilizarían la norma ISBD (ER) (Codina, 2000a: 56-57).

El *Dublin Core* es la forma abreviada para el *Dublin Metadata Core Element Set*, proyecto cooperativo de ámbito internacional, promovido por OCLC y NCSA (*National Center for Supercomputer Applications*). Desde 1995 en que nació el proyecto, se han celebrado varios seminarios para llegar a acuerdos sobre el *Dublin Core*. Uno de los objetivos principales del *Dublin Core* es proporcionar un método estándar para que sean los propios organismos creadores o editores de documentos HTML quienes describan los recursos de manera que sea fácil realizar después búsquedas mucho más precisas de lo que permiten los motores de búsqueda actuales.

Su uso está previsto, principalmente, en el seno de instituciones académicas, culturales y de

1 *Dublin Core Metadata Initiative*. <http://dublincore.org>

investigación, es decir, en general en contextos donde el esfuerzo de descripción venga justificado por la calidad intrínseca y la condición no volátil de los recursos, por lo cual es fácil prever que también será utilizado por todas aquellas instituciones y empresas que deseen beneficiarse de un estándar de descripción de datos de estas características.

Aunque es el *Dublin Core* el formato de metainformación más divulgado no es el único, como hemos señalado. Es preciso hacer mención también del proyecto *Text Encoding Initiative* (TEI), del *Resource Description Framework* (RDF) y del *MARC DTD*.

El TEI² es una DTD de SGML para asegurar un formato estándar con amplias capacidades de marcado para la indización y el intercambio de información textual, pero que sirve para estructurar información de cualquier tipo. Es significativa su cabecera (TEIH) que permite definir una descripción bibliográfica detallada para cada texto codificado. Desde 1998 se está trabajando en una versión de la TEIH en XML, que previsiblemente ampliará las posibilidades de aplicación de este formato de descripción de contenido *Web*.

El RDF³ es el estándar en desarrollo más importante para la descripción de contenidos *Web*, no es sólo un formato de metadatos. Promovido por el Consorcio (W3C) es una aplicación de metadatos que utiliza XML con el fin de proporcionar un marco normativo donde puedan interoperar los distintos modelos de metadatos para la descripción de contenido, como el *Dublin Core*, formato que más ha influido en el desarrollo del RDF. Su ventaja reside en su capacidad para soportar diferentes

sistemas de descripción de recursos. La ventaja de XML, como sabemos, es su flexibilidad lo que puede contribuir a la creación de un *Web* organizado finalmente. Los tres aspectos de la semántica funcional del RDF son: un modelo de datos, una sintaxis y un esquema (Méndez, 1999: 487-498).

El MARC DTDs⁴ (*Machine Readable Cataloging Document Type Definitions*) es un proyecto desarrollado por la *Network Development and MARC Standards Office* de la *Library of Congress*. El propósito inicial era crear DTDs de SGML para soportar la conversión de los datos catalográficos desde la estructura del formato MARC a SGML y viceversa sin perder información. El proyecto se inició en 1995 y en estos momentos está disponible en XML. Es posible descargar las DTDs desde la página correspondiente de la *Library of Congress* (Ortiz-Repiso, 1999: 212).

Para terminar este apartado creemos necesario referirnos al ya mencionado EAD (Descripción Archivística Codificada)⁵. Su origen fue el proyecto de instrumentos de descripción de Berkeley, situado en la *University of California*, en Berkeley. El proyecto de dos años de duración comenzó en 1993. Su finalidad era investigar la viabilidad y conveniencia de establecer un estándar codificado para instrumentos de descripción electrónicos sin marca registrada, que no sólo proveyera un mecanismo de depósito estable y a largo plazo, sino que también mejorara la navegación y la búsqueda en la *World Wide Web* (Kiesling, 2001: 74).

La versión 1.0 de EAD prácticamente fue completada a fines de enero de 1998, pero el grupo de trabajo decidió posponer su publicación a que

1 Dublin Core Metadata Initiative. [en línea] <http://dublincore.org>

2 <http://www.tei-c.org>

3 <http://www.w3.org/TR/rdf-schema>

4 <http://lcWeb.loc.gov/marc/marcdtd/marcdtdback.html>

5 <http://lcWeb.loc.gov/ead/>

se completara la norma XML ya que se consideró conveniente, dado el previsible desarrollo de la *Web* en este estándar, desarrollar una DTD compatible con XML. En agosto de 1998 apareció la versión 1.0 de EAD en ambas sintaxis, SGML y XML, acompañado del Repertorio de Etiquetas. Las Directrices de Aplicación de EAD se publicaron un poco antes de la reunión anual de la *Society of American Archivists* de agosto de 1999, y en esta reunión el consejo de la asociación aprobó el EAD como estándar descriptivo.

Se constata la aproximación de intereses, objetivos y métodos de trabajo que se está produciendo entre las distintas instituciones documentales, principalmente archivos y bibliotecas, y en ello está jugando un papel crucial el documento digital. Se comprueba también que en el estadio en que nos encontramos se debiera hablar de archivos y bibliotecas híbridos, más que de archivos y bibliotecas digitales, por lo que es preciso integrar los documentos digitales con los analógicos y someterlos a los mismos procedimientos. La experiencia normalizadora de los profesionales de la información sigue siendo fundamental para el tratamiento de los documentos digitales.

EL ANÁLISIS DE CONTENIDO DE LOS DOCUMENTOS DIGITALES Y/O MULTIMEDIA

Para el análisis del contenido, los componentes formales de los documentos, el medio de transmisión del mensaje y el código son condicionantes significativos. El código icónico de una fotografía o la mezcla de códigos en los documentos audiovisuales o digitales obligan a adaptar el Análisis de Contenido concebido para representar y recuperar el documento escrito.

A efectos exclusivamente de análisis de contenido, resulta útil clasificar los documentos digitales en tres tipos (Pinto; García y Agustín, 2002: 314-315):

- ◆ Documentos digitalizados
- ◆ Programas de ordenador
- ◆ Documentos multimedia

En el primer grupo cabe incluir aquellos documentos que suponen una simple traslación a formatos, soportes y medios de reproducción digitales de mensajes que son documentos textuales, sonoros o visuales y que, por tanto, requieren el análisis de contenido propio de su código, recogiendo, como novedad, el medio, formato y soporte digital.

Los programas de ordenador son herramientas para procesar información, y necesitan ser identificados y localizados, pero carecen de mensajes informativos y no requieren un análisis de contenido.

Por lo que se refiere a los documentos multimedia, son documentos en que se combinan, bajo las herramientas interactivas usuales, diferentes elementos comunicativos: texto, archivos sonoros, vídeo digital, etc. En los documentos multimedia interactivos el medio modifica el mensaje lo que origina una clase de documentos distintiva a efectos de análisis documental.

El análisis de contenido de los documentos multimedia es complejo debido a la ya citada yuxtaposición de códigos pero también a la diversidad de documentos de que se ocupa. Entre ellos se encuentran documentos con contenidos pedagógicos, científicos, técnicos o de divulgación, así como los generados por medios de comunicación de masas de temática más relacionada con la actualidad y la creación.

Frente al discurso textual y al documento visual fijo, los medios que descansan en la imagen móvil y en el sonido plantean problemas en el proceso de análisis, debido al carácter móvil y transitorio de los mensajes que emiten. El documentalista se verá obligado al «revisiónado» o «reaudición» y las secuencias se definirán y localizarán mediante cronómetro.

Conviene considerar, igualmente, las peculiaridades de la imagen, su carácter sintético y polisémico, que implican su rápida selección en la recuperación y flexibilidad en su utilización. Estas características derivan de su código icónico menos formalizado que el verbal. La descripción de la imagen es intrínsecamente subjetiva en virtud de la separación entre lo que denota - el significado común del signo - y connota - el significado cultural del signo - y requiere, con frecuencia, del texto o el sonido para fijar su significado, para su contextualización. En los documentos multimedia la función contextualizadora podrá ser realizada tanto por el texto como por el sonido que acompaña a las imágenes. El código verbal resulta más preciso semánticamente que el visual.

Dado que la información multimedia está integrada por elementos que pertenecen al ámbito textual, de las imágenes y del sonido, su estudio debe abordarse desde una doble dimensión:

- ◆ Considerando cada uno de los niveles separadamente.
- ◆ Considerando todos los niveles conjuntamente, observando las transformaciones que experimentan como resultado de la combinación de códigos.

Asimismo, será preciso considerar las finalidades del tratamiento de estos documentos, que pueden ser dos:

- ◆ Recuperación unitaria de documentos.
- ◆ Recuperación fraccionada en microunidades informativas.

Esta realidad implica que la unidad documental podrá ser cualquier unidad temática de información susceptible de ser indizada por sí misma, forme parte o no de un conjunto más amplio. El nivel de análisis estará en relación con la explotación posterior que este material pueda tener. Los contenidos de los documentos con un alto potencial de ser reutilizados, exigirán una descripción y representación más detallada para facilitar su recuperación por cualquier concepto.

Dada la inestabilidad de los documentos digitales, los productos del análisis de contenido se integraran entre los metadatos, lo que favorecerá su localización. El objetivo del análisis documental es obtener una representación del documento multimedia que ha de cumplir dos funciones básicas:

- ◆ Proporcionar un conjunto de datos que identifiquen inequívocamente el documento.
- ◆ Proporcionar una representación de su contenido semántico.

En el modelo Dublin Core existen elementos dirigidos tanto a identificar el documento, elementos catalográficos, como a poner de relieve su contenido.

Consideramos que los niveles del análisis de contenido de los documentos multimedia serían los siguientes:

1. Elementos de identificación o catalográficos.

2. Análisis del contenido (visionado):

- ◆ Lectura banda texto.
- ◆ Lectura banda imagen.
- ◆ Lectura banda sonido.

3. Indización.

4. Resumen.

La descripción de los documentos multimedia sólo puede hacerse a partir de un visionado completo del documento. La diacronía del documento textual, sonoro y audiovisual obliga a ello. Para facilitar la localización de las imágenes es necesario indicar el punto del documento (Time Code o código de tiempo) en el que se localiza cada información. Durante el visionado el documentalista debe «leer» las imágenes y crear una descripción textual, traduciendo en palabras el contenido visual y sonoro del documento, de forma lo más objetiva posible, y teniendo presente tanto el aspecto técnico y compositivo como el del contenido. Se trata no sólo de ordenar aquello que el ojo percibe en muy pocos segundos sino de identificar los elementos, para lo que es necesario poseer los referentes adecuados que permitan identificar los objetos, sujetos y acciones de las imágenes cuando no aparecen explicadas por el texto o el sonido.

El nivel de detalle de la descripción dependerá del tipo de documento, de su estabilidad o longevidad prevista, interés y del tipo de búsquedas que posteriormente puedan establecerse. Fijar el nivel de descripción de los mensajes que el documento ofrece es fundamental para una buena gestión. Dadas la facilidad de publicar en Internet y, por tanto, la diversidad en cuanto a calidad de los documentos que circulan por la red, será preciso realizar una selección que priorice la catalogación e indización

de documentos merecedores del esfuerzo y coste que el análisis documental implica.

Concluyendo, los documentos digitales de carácter informativo, considerados prioritarios, habrán de ser catalogados e incorporados al OPAC, realizando una gestión integrada de la colección analógica y digital, independientemente de que hayan sido descritos en el momento de su creación por medio de metadatos.

Con este fin, las bibliotecas tienen que preocuparse de su conservación a largo plazo. No obstante, los mayores esfuerzos se han concentrado en elaborar herramientas que permitan la recuperación de la información que circunstancialmente circula por la red sobre un determinado asunto. Esto ha contribuido al enorme desarrollo de los buscadores que adolecen, sin embargo, de carencias que los invalidan para funcionar como una gran biblioteca digital, fundamentalmente que no son capaces de custodiar todos los contenidos.

Las bibliotecas nacionales han empezado a archivar los contenidos de las páginas *Web*. En algunos casos se han buscado mecanismos similares al depósito legal para garantizar la recopilación de todas las publicaciones electrónicas de un país. Uno de los proyectos pioneros ha sido PANDORA (*Preserving and Accessing Networked Documentary Resources of Australia*)⁶, mediante el cual la Biblioteca Nacional Australiana, en colaboración con otras nueve bibliotecas de dicho país y algunas organizaciones culturales, pretende recoger todas las publicaciones electrónicas, almacenando siempre la versión en línea junto con las versiones ya retiradas. Este ejemplo ha de ser seguido por las principales instituciones bibliotecarias de los distintos países si queremos conservar la cultura digital que desde hace más de una década se ha venido originando.

6 <http://pandora.nla.gov.au/index.html>

BIBLIOGRAFÍA

- Codina, L. (2000). *El libro digital y la WWW*. Madrid: Tauro.
- Codina, L. (2000 a). Evaluación, descripción y representación de recursos digitales. En Rovira, C. y Codina, L. (dir.). *Documentación digital 2000*. [En línea]. Barcelona: UPF. <<http://docdigital.upf.es>>
- Garduño Vera, R. (2000). Paradigmas normativos para la organización documental en los albores del siglo XXI. *Investigación Bibliotecológica*, 14, n. 28, p. 115-149.
- Méndez Rodríguez, E. (1999). RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. En *7es Jornades Catalanes de Documentació*. Barcelona: Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya, p. 487-498.
- Méndez Rodríguez, E. (2002). Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea.
- Méndez Rodríguez, E.; Merlo Vega, J. A. (2000). Localización, identificación y descripción de recursos Web: tentativas hacia la normalización. En *7es Jornades Espanyolas de Documentación*. Bilbao: Universidad del País Vasco. P. 221-231.
- Moreiro González, J. A. (2004). *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. Gijón: Trea.
- Negroponte, N. (1995). *El mundo digital*. Barcelona: Ediciones B.
- Ortiz-Repiso Jimenez, V. (1999). Nuevas perspectivas para la catalogación: metadatos versus MARC. *Revista Española de Documentación Científica*, 22, n. 2, p. 198-219.
- Kiesling, K.L.R. (2001). Descripción archivística codificada (EAD): desarrollo y potencial internacional. *Lligall*, 17, p. 73-88.
- Pinto Molina, M. (1999). Tratamiento de los contenidos en la Sociedad de la Información. En Caridad Sebastián, M. (coord.). *La sociedad de la información: política, tecnología e industria de los contenidos*. Madrid: Centro de Estudios Ramón Areces. P. 267-288.
- Pinto Molina, M.; García Marco, F.J.; Agustín Lacruz, M^a C. (2002). *Indización y resumen de documentos digitales y multimedia: técnicas y procedimientos*. Gijón: Trea.
- Rodríguez Bravo, B. (1996) La indización de documentos no bibliográficos. *Accademie e Biblioteche d'Italia*. Anno LXIV, n. 3, 1996, p. 47-54.
- Rodríguez Bravo, B. (2000). El tratamiento documental de la imagen fotográfica: su enseñanza en la Universidad de León. En *6es Jornades Antoni Varés*. Imatge i recerca. Girona: Ajuntament de Girona. P. 205-211.
- Rodríguez Bravo, B. (2002). *El documento, entre la tradición y la renovación*. Gijón: Trea.
- Rodríguez Bravo, B. (2002a). Revisión de las clasificaciones documentales basadas en el soporte. *Revista española de documentación científica*, vol. 25, n.1, enero-marzo, p.74-86.

Rodríguez Bravo, B. (2004). El documento audiovisual en las emisoras de televisión: selección, conservación y tratamiento. *Biblios: revista electrónica de ciencias de la información*, Año 5, n.20.

Rodríguez Bravo, B., Diez Diez, A. (2003). Los contenidos audiovisuales y multimedia en el curriculum de Biblioteconomía y Documentación de la Universidad de León. Ponencia presentada en la 69th IFLA World Conference 2003. Berlin, agosto de 2003. [Accesible en: http://www.ifla.org/IV/ifla69/papers/001s-Bravo_Diez.pdf]

Schamber, L. (1996). What is a document? Rethinking the concept in uneasy times. *Journal of the American Society for Information Science*, vol.47, n.9, p. 669-671.

Vianello Osti, M. (2004). *El hipertexto entre la utopía y la aplicación: identidad, problemática y tendencias de la Web*. Gijón: Trea.

