

Utilización de metadatos para la descripción documental y recuperación de material audiovisual en los archivos televisivos

Dr. JORGE CALDERA SERRANO

F. de Biblioteconomía y Documentación. U.de Extremadura

RESUMEN:

Se muestra el principal estándar de metadatos para la conservación y acceso a la información audiovisual MPEG-7 con identificación de contenido por medio de descriptores. Previamente se analiza los diferentes métodos de división conceptual del material audiovisual en su vertiente sonora y visual en movimiento necesario para la implementación de dichos metadatos, estudiando la pertinencia de dicho estándar con las necesidades reales de gestión documental en los archivos audiovisuales televisivos.

0. ARCHIVOS ELECTRÓNICOS AUDIOVISUALES.

La transición del *homo typographicus* al *homo videns* (propiada por los medios de comunicación audiovisuales) ha sido claramente alterada por la aparición del *homo digitales*, el cual en lugar de desbancar al anterior lo ha potenciado, aportando mucha mayor capacidad de acceso y multiplicidad de contenidos.

La televisión, como empresa y medio de comunicación, ha encontrado en la difusión digital un nuevo y revolucionario método para la transmisión de sus productos programáticos.

Los archivos de televisión como parte de un engranaje empresarial y activo institucional e informacional no es un extraño a esta nueva realidad, debiendo adaptar sus técnicas y hábitos de gestión analógicos hacia una cultura empresarial digital.

Tratar de metadatos en los archivos de televisión aún puede ser un elemento con demasiadas pretensiones en aquellas cadenas que aún trabajan con material analógico. Es obvio que las cadenas se plantean dicha problemática derivada de la necesidad de adaptación a las nuevas tecnologías de difusión – televisión a la carta – como de conservación en los departamentos de documentación, sin

embargo este proceso de migración de lo analógico a lo digital será un proceso lento especialmente para aquellas televisiones con documentación en diferentes soportes y formatos, y con gran cantidad de horas almacenadas en sus depósitos. Televisión Española supera el millón de horas en sus archivos lo que acarreará una labor de años para la migración de esta información a formato electrónico.

Hasta ahora las cadenas contaban con bases de datos referenciales que enviaban al usuario a la videoteca donde poder recuperar el documento. El registro documental cambia por completo en el ámbito digital ya que el video, en si mismo, es un elemento del propio registro, agilizando tanto el acceso como el tratamiento periodístico. A este nuevo proceso se lo ha denominado “producción en línea”.

1. METADATOS Y AUDIOVISUAL.

No se pretende entrar en consideración sobre la presencia cada vez mayor de material audiovisual en la red Internet aunque tampoco se va a obviar que es una plataforma interesante para el intercambio de material audiovisual; el trabajo se va a centrar en la implementación de dichos recursos por medio de estándares comunes para el acceso tanto a usuarios remotos como en el marco de intranet corporativas. La realidad entre archivos audiovisuales es la escasa relación entre ellos derivado de las cerradas estructuras empresariales y del control del flujo informacional de ahí la escasa repercusión de estándares comunes como ECHO sobre los que la Unión Europea está creando grandes expectativas para el intercambio entre productoras y cadenas televisivas.

De todas formas las estandarizaciones en metadatos ofrece la dualidad en la consecución de dichos objetivos: difusión para usuario/telespectadores como intercambio entre empresas audiovisuales.

La gran cantidad de información audiovisual existente

en las cadenas televisivas hace necesario buscar soluciones para los problemas de catalogación, estructuración y acceso a la información. Los motores de búsqueda suelen llevar a cabo sus interrogaciones sobre elementos textuales que acompañan al documento y no sobre el material audiovisual propiamente dicho. El método para afrontar esta problemática es la estandarización y mejora de los metadatos (información sobre información) comunes. Estas anotaciones que se realizan sobre el multimedia son variadas: título, palabra clave, tema, etc. Así como del contenido de la escena: fondo, objetos, etc. Al final del trabajo se señalan aquellas etiquetas / descriptores con los que cuenta el MPEG-7.

Las disciplinas que pudieran dedicarse a la extracción del contenido de este material audiovisual son variadas, desde la semiótica a la documentación, extrayendo contenidos diferentes para este material dependiendo del prisma con el que es observado y analizado. Desde el punto de vista documental, las bases de datos referenciales cuentan con gran cantidad de información derivada de las solicitudes de los usuarios por lo que la información se encuentra a diferentes niveles.

Existen metadatos sobre elementos de propiedad intelectual como aquellos que se refieren a la distribución y potencialidad de acceso, además de contar con información sobre la producción del contenido (autor, productor, origen, derechos) para intentar contrarrestar la piratería audiovisual¹.

Por lo tanto, el acceso a la información se extrae automáticamente de los descriptores sobre el material (referencial) y sobre el propio material.

2. DESCOMPOSICIÓN DEL VIDEO EN FRAGMENTOS A DESIGNAR METADATOS.

Tres son las principales fórmulas de extracción de la información en videos.

2.1. *Descomposición temporal.* Se realiza sobre el bruto, material sin montar, en forma secuencial en tiempo para audio y video. Este material no está editado y no existe una manipulación por parte del productor como ocurre con el material emitido.

Existen programas que descomponen el material en las secuencias por los cortes, fundidos, cambios de planos, superposición y desplazamientos, etc (*Software Pinacle*). El número de planos varía de los 7 a 14 de la noticia de un informativo a los más de 1000 para películas cinema-

tográficas. Estos mecanismos de detección automática cuenta con fallos que deben ser revisados por el profesional. De todas maneras, ofrece una serie de algoritmos que agrupan planos, creando grupos con propiedades comunes (escenas completas) siendo necesario una revisión final para su mejora, es decir, se crean áreas temáticas visuales. Cuenta por tanto con una descomposición jerárquica que facilita la búsqueda y la navegación, por lo tanto no existe únicamente una descripción lineal sino una descomposición en un árbol semántico. Además confecciona resúmenes de contenidos visuales que ayudan a una toma de información más rápida, elemento que facilita la operación de selección en los archivos audiovisuales cuando debe determinarse la valía de material de horas de duración.

La realidad del movimiento es doble, por lo que analiza tanto movimientos de cámaras y planos, y el movimiento por el que se desarrolla la acción. Ambos movimientos ofrece datos para el sistema ya que la distancia focal, encuadres, profundidad de campo son fuente para extracción de metadatos y, por consiguiente, de descriptores que sirvan como elemento o punto de acceso a la información audiovisual.

2. *Análisis gráfico de imágenes.* Señalar que el ojo humano descompone la información recibida por medio de elementos simples como los bordes, colores, densidad y textura, superficie, entre otros, creando así patrones comunes para agruparlas. Este es el principio de este método de descomposición de imágenes que se analiza ya que agrupa elementos con semejanzas comunes, teniendo en cuenta el movimiento y el cambio de secuencias.

3. *Descomposición espacial.* Este método lleva a cabo la separación de la imagen por medio de los elementos que lo constituyen. Se implementa partiendo de la base del modelo número dos junto con el análisis del movimiento en el video. Su utilidad aún se está desarrollando ya que su funcionamiento no es aún correcto por la separación en secuencias que puede resultar relativamente fácil pero no la segmentación por movimiento.

Existen dos elementos que pueden encontrarse presente en el video que son muy utilizados para la recuperación, con es la "rotulación" (subtítulos, identificar de personas, titulares, etc.) y los "rostros" de personas. La rotulación ayuda a la identificación y clarificación de las imágenes mientras que la identificación de rostros es uno de los mayores requerimientos al software por parte de los archivos audiovisuales derivado de un periodismo

¹ Directiva 2001/2/CE del Parlamento Europeo y del Consejo de 22 de mayo de 2001 relativa a la armonización de determinados aspectos de los derechos de autor y derechos afines a los derechos de autor en la sociedad de la información. Diario Oficial de las Comunidades Europeas, L 167.

declarativo. El software EXCALIBUR puede identificar personas aunque con grandes y graves problemas; dicho software se basa en el segundo modelo analizado y aún la investigación en visión artificial no se acerca al sistema de visión humano en el reconocimiento de objetos y escenas. Clarificar que la lectura y análisis de rótulos es relativamente sencillo por medio de reconocedor de caracteres mientras que para el reconocimiento de rostros es necesario contar con un banco de imágenes donde el software localiza las características del personaje. Funciona con un número reducido de personajes.

El componente sonoro es igualmente importante en el análisis y tratamiento en los archivos audiovisuales. Dicho componente sonoro puede contar con gran cantidad de información, a veces más que en la propia imagen. Para el reconocimiento de voz existen software para reconocer la locución, por lo que se lleva a cabo una transcripción del lenguaje oral a texto, pudiendo llevar a cabo técnicas de indexación automatizada.

Los elementos que pueden tratarse, siendo susceptible de describir con metadatos sería tanto el contenido sonoro como el tipo de sonido (música, efectos, palabra), idioma, identificación de personaje por la voz, entre otros factores.

3. ESTÁNDAR NORMALIZADO DE METADATOS PARA TELEVISIÓN: MPEG-7².

Con la llegada de la digitalización a la producción y difusión de programas televisivos han surgido iniciativas para incorporar metainformación en la producción y distribución de contenidos, tanto de cara a los usuarios finales como para el intercambio de material entre instituciones y empresas. El Instituto Europeo de Normalización publicó un estándar para la incorporación de información sobre contenidos en sistema de difusión de televisión digital³ para el envío junto con la señal televisiva de otra para informar sobre la programación de la cadena.

La Unión Europea de Radiodifusión (UER o EBU) y la estadounidense SMPTE (*Society of Motion Picture and Television Engineers*) conformaron un grupo de investigación denominado *Task Force* cuyo fin era la implementación de estándares para intercambio de material audiovisual entre empresas, especialmente productoras y televisión. Aunque se sigue trabajando en dicha línea, cuenta con un informe publicado en 1997 con el nombre *Task*

Force for Harmonized Standards for the Exchange of Program Material as Bit Streams, en el que se puede evaluar el estándar alcanzado.

MPEG-7 (Moving Pictures Experts Group).

Sin lugar a dudas este estándar es el más destacado, importante y ambicioso en el ámbito de la metainformación para material multimedia especialmente para el audiovisual. Implementado por comité de MPEG de ISO para incorporar descripción de contenidos a sus listas de estándares es la evolución y desarrollo de formatos anteriores comenzada con MPEG-1 en 1992 (almacén y descarga de archivos audiovisuales), MPEG-2 en 1994 (televisión digital), sirviendo de base para estándares muy conocidos en la actualidad como el Vídeo CD, el Digital Audio Broadcasting y para la codificación audio MP3. Con posterioridad, en 1999 se presenta la segunda versión de MPEG-4 (codificación audio y vídeo en forma de objetos). Éstos contaban con el objetivo de la representación del contenido por medio de una capa semántica por encima del material audiovisual para acceder a la información. Especialmente MPEG-4 incluye la noción de objeto a una representación audiovisual facilitando el proceso para material audiovisual desde el punto de vista del contenido. Se estableció como base para nuevos servicios por las aplicaciones a móviles de carácter multimedia y muy especialmente para la potenciación de la televisión digital.

MPEG-7, publicado como Estándar Internacional en el año 2001, incluye en su estructuración el concepto novedoso –no desde el punto de vista documental– de *descriptor* como representación de una propiedad de la información audiovisual que muestra el contenido del mismo en alguna de sus facetas. Por lo tanto no se podrá hablar del concepto de metadatos y MPEG hasta la inclusión de este nuevo estándar. Aunque en este trabajo se adapta al contenido de material audiovisual puede también ser utilizado para material gráfico fotoperiodístico, material en tres dimensiones, audio, como para la unión de dichos elementos en montajes multimedia.

Estos descriptores organizados en esquemas de descripción utilizando como medio el *Lenguaje de Definición de Descriptores* (DDL), definido igualmente por MPEG, que utiliza como base el lenguaje *XML Schema* de la WWW. Este lenguaje (DDL) aporta una gran flexibilidad para la especificación de descriptores no previamente definidos.

² VILLEGAS NÚÑEZ, Paulo; GARCÍA FERNÁNDEZ, Daniel. Sistemas de descripción de contenidos multimedia. Telefónica Investigación y Desarrollo, número 24 – Enero 2002, 12 p.

³ European Telecommunications Standards Institute: *Digital broadcasting systems for television, sound and data services: Specification for Service Information (SI) in Digital Video Broadcasting (DVB) systems*. ETS 300 468, mayo 1996.

Este estándar cuenta con la clara orientación de creación de modelos de descodificación general para detallar la sintaxis y la semántica de dichos elementos descriptivos.

Por lo tanto, el MPEG-7 permite la descripción por palabras claves y por significación semántica de los contenidos más interesantes para los archivos de televisión (quién, qué, cuándo, dónde) además de aportar elementos estructurales (colores, textura, movimientos, sonidos, formas), ayudando a las herramientas de indización para la creación de grandes bases de datos de material audiovisual.

Recordar que MPEG-7 trabaja en diferentes secciones que describen el material tales como: Sistemas, el cual analiza aspectos generales sobre el acceso a la información, tipología en formatos, sincronismo entre ficheros. DDL, definiendo un lenguaje de descriptores. Sección visual, que identifica y describe los elementos visuales por sus propiedades espaciales ya mencionadas. Audio, delimitando y especificación los diferentes elementos de descripción del sonido como timbre, distribución, tempo, ritmo. Esquemas de descripción, que trabaja no solo con video sino con la organización de elementos individuales que describe las estructuras físicas y de contenido semántico.

Destacar también la labor de investigación de MPEG sobre el nuevo estándar *MPEG-21*, implementado para ofrecer soluciones a los diferentes problemas planteados por los estándares anteriores y muy especialmente orientado a presentar una perspectiva general y global desde el punto de vista de la infraestructura necesaria para la distribución y uso de contenidos digitales, abarcando aspectos tales como la producción, descripción, gestión, protección como acceso.

MPEG-7 ofrece un número importante de descriptores, sin embargo se presentan aquellos que se estiman más importantes⁴:

Tipo. VISUAL

Características: ESTRUCTURA BÁSICA

Descriptores: DISEÑO DE CUADRÍCULA, HISTOGRAMA

Características: COLOR

Descriptores: ESPACIO DE COLOR, COLOR DOMINANTE, HISTOGRAMA DE COLOR, CUANTIFICACIÓN DEL COLOR.

Características: TEXTURA

Descriptores: DISTRIBUCIÓN DE INTENSIDAD ESPACIAL, TEXTURA HOMOGÉNEA

Característica: FORMA

Descriptores: CAJA QUE ENGLOBA UN OBJETO, FORMAS BASADAS EN REGIONES, FORMAS BASADAS EN CONTORNO, DESCRIPTORES 3D

Características: MOVIMIENTO

Descriptores: MOVIENTOS DE CÁMARA, TRAYECTORIA DEL OBJETO, VELOCIDAD DEL OBJETO, ACELERACIÓN DEL OBJETO

Tipo: AUDIO.

Características: ANOTACIÓN DEL HABLA

Descriptores: PALABRA Y FONEMAS

Características: TIMBRE

Descriptores: ARMÓNICOS, COHERENCIA DEL ATAQUE ARMÓNICO.

Características: MELODÍA

Descriptores: CONTORNO, RITMO.

⁴ PEIG OLIVÉ, Enric. Interoperabilidad de metadatos en sistemas distribuidos. Barcelona: Universitat Pompeu Fabra, 2003, p. 22-23.