



Luis Herranz Arribas es Ingeniero de Telecomunicación por la Universidad Politécnica de Madrid, especialidad Comunicaciones.

Durante la realización de su proyecto de fin de carrera en el Grupo de Tratamiento de Imágenes (GTI) de la ETSIT-UPM fue becario de la Cátedra Telefónica de la UPM para Internet de Nueva Generación.

Actualmente se encuentra cursando estudios de doctorado en Ingeniería Informática y Telecomunicación en la Universidad Autónoma de Madrid, en la sección del GTI de la Escuela Politécnica Superior de la UAM, donde trabaja como becario de investigación y participa activamente en el proyecto AceMedia del VI programa Marco de la Unión Europea y el proyecto nacional DYMAS, relacionados con temas de televisión digital, análisis y transcodificación de vídeo, y adaptación de contenidos multimedia.

Datos del Proyecto Fin de Carrera

Datos del autor:

Nombre: Luis Herranz Arribas

Nº Colegiado: 12.382

Nombre del Tutor del Proyecto: José María Martínez Sánchez

Proyecto de Fin de Carrera realizado en el Grupo de Tratamiento de Imágenes del Departamento de Señales, Sistemas y Radiocomunicaciones de la E.T.S.I. Telecomunicación de la Universidad Politécnica de Madrid.

Fecha de lectura: 22 de Julio de 2003

Calificación: Matrícula de Honor

Acceso Multimedia Universal a través de terminales y redes fijas mediante MPEG-7 y MPEG-21

Introducción

En los últimos años, con la evolución tecnológica en el mundo de las comunicaciones, han surgido nuevas formas de producir, filtrar, buscar y administrar información multimedia digitalizada. El acceso de banda ancha es ya una realidad para los usuarios, ofreciendo crecientes velocidades de acceso y calidades de audio y vídeo. Todos estos elementos han cambiado el modelo de la web y han permitido la aparición de nuevos servicios en los que la información multimedia es fundamental. La cantidad de contenidos multimedia disponibles actualmente a través de una gran cantidad de fuentes es tal que el acceso eficiente y preciso se vuelve algo muy valioso. A pesar de que los usuarios cada vez tienen mayor acceso a estos contenidos, identificar y administrarlos eficientemente se está volviendo cada vez más complicado, tanto a nivel profesional como personal.

Ante este panorama, el grupo MPEG (Moving Picture Experts Group) desarrolló el estándar MPEG-7 para hacer frente a las necesidades planteadas. Este estándar establece una serie de herramientas genéricas de descripción de metadatos que puedan ser utilizadas en un amplio rango de aplicaciones. Posteriormente, también relacionado con la gestión del entorno multimedia, se inició el estándar MPEG-21 (actualmente algunas partes siguen en desarrollo), que pretende describir un marco de referencia global y estándar en el que se encuentren todos los agentes que intervienen en la manipulación de contenido multimedia. En este marco se incluyen desde los productores de contenido hasta los usuarios finales, incluyendo los diversos agentes intermedios, así como el propio contenido multimedia.

En este proyecto se ha desarrollado un prototipo de un sistema web en el que se emula el acceso transparente, mediante diversos terminales y redes fijas, a una base de datos multimedia mediante el uso de perfiles de sesión (incluyendo preferencias de usuario, recursos del terminal utilizado y capacidades de la red de acceso). Para ello se han utilizado las herramientas de los mencionados estándares MPEG-7 y MPEG-21 para la selección del contenido, filtrado, la navegación y previsualización de resultados.

Acceso Multimedia Universal

Como indica el título del proyecto, el objetivo principal del proyecto es investigar y evaluar cómo las herramientas de MPEG-7 y MPEG-21 habilitan, en un marco estándar, el desarrollo de sistemas y aplicaciones de Acceso Multimedia Universal (*Universal Multimedia Access, UMA*), y posteriormente desarrollar un sistema que implemente un grado razonable de Acceso Multimedia Universal, centrándose en el caso particular de terminales y redes fijas.

El Acceso Multimedia Universal es un concepto introducido recientemente para referirse a la capacidad de acceder a contenido multimedia enriquecido a través de cualquier terminal y de cualquier red. Estos sistemas deben crear diferentes presentaciones de la información (teniendo en cuenta los dispositivos, redes y preferencias de usuario), de forma transparente al usuario. El ejemplo paradigmático de aplicación UMA es la de un servicio de acceso a noticias, en el que el usuario recibe las mismas noticias (información) accediendo a través de televisión, internet, periódico, teléfono móvil, etc.

Otro detalle clave es que la información ha sido creada una sola vez (las noticias del ejemplo anterior se han creado una sola vez, y en principio almacenadas en un solo archivo), y es responsabilidad del sistema adaptar el contenido a las diferentes características de las redes (por ejemplo, diferentes tasas binarias), capacidades de los terminales (por ejemplo, reducir la resolución para una pantalla de teléfono móvil, y), y preferencias y características del usuario (por ejemplo, enviar texto subtulado en una película en lugar de audio para usuarios con deficiencias o incapacidad auditiva). Para la adaptación del contenido se necesitan técnicas de transcodificación o de cambio de modalidad (*transmoding*, como por ejemplo el paso de voz a texto, o viceversa).

La importancia del acceso universal se hace patente sobre todo si consideramos la tendencia de crecimiento cada vez mayor del número de dispositivos de capacidades que pueden hacer uso de ello y de sus posibilidades, como pueden ser PDAs, ordenadores de bolsillo, teléfonos inteligentes de nueva generación, etc. Mediante el uso de descriptores y esquemas de descripción estandarizados se tendrá un acceso realmente universal y avanzado al contenido multimedia, como ya el grupo MPEG lo ha tenido en cuenta y mediante el estándar MPEG-7 de descripción de contenidos y el estándar en desarrollo MPEG-21 que define un marco común de interoperabilidad para los recursos de red y los dispositivos.

MPEG-7

MPEG-7 es un estándar internacional de la ISO/IEC, desarrollado por el grupo MPEG, y conocido como Interfaz de Descripción de Contenido Multimedia. Proporciona un amplio conjunto de herramientas estándar para describir contenido multimedia, tanto para usuarios humanos como para sistemas automáticos que procesen información audiovisual. Estas herramientas de descripción (*Description Tools*, que son los elementos de metadatos, su estructura y relaciones, definidas en el estándar, y que se dividen en Descriptores (D) y Esquemas de Descripción (DS)) sirven para crear descripciones que serán la base para aplicaciones que permitan este necesario acceso eficiente a contenido multimedia.

MPEG-7 utiliza XML *Schema* (XML, *Extensible Markup Language*) como lenguaje para la representación textual del contenido, lo cual le permite ser flexible y aumentar las herramientas de descripción existentes. Para poder dar al estándar mayores funcionalidades, no se utiliza XML *Schema* puro, sino que se han añadido algunas extensiones de las que éste carece, dando lugar al Lenguaje de Definición de Descripciones (DDL), con el cual se definen los Descriptores y los Esquemas de Descripción.

El alcance de MPEG-7 se limita al formato de las descripciones, y no estandariza ni la producción de esas descripciones ni el uso que se va a dar.

De entre las partes de que consta MPEG-7 las más importantes son:

- *MPEG-7 Systems*: arquitectura del estándar y herramientas necesarias para preparar las descripciones de MPEG-7 para su transporte y almacenamiento eficiente.
- *MPEG-7 Description Definition Language*: lenguaje para definir nuevos DSs y, quizás eventualmente, también para nuevos Ds.
- *MPEG-7 Audio*: proporciona estructuras para describir material sonoro. En ellas se tiene un conjunto para de descriptores de bajo nivel (para características espectrales, temporales, etc), y otro de herramientas de alto nivel.
- *MPEG-7 Visual*: estructuras básicas y descriptores que cubren las siguientes características visuales básicas: color, textura, forma, movimiento, localización y reconocimiento facial.

- *MPEG-7 Multimedia Description Schemes*: elementos (DSs y Ds) que describen información genérica, es decir, que no es ni puramente visual ni de audio.

Esquemas de descripción multimedia

Los Esquemas de Descripción Multimedia (Multimedia Description Schemes) son estructuras de metadatos para describir y anotar contenido audiovisual. Estos DSs proporcionan un modelo estándar para describir en XML aspectos importantes relacionados con la descripción de contenido audiovisual y gestión de contenido, para facilitar las tareas de búsqueda, indexado, filtrado y acceso a dicho contenido. Estos esquemas de descripción se definen utilizando el lenguaje descrito en MPEG-7 DDL. Las descripciones MPEG-7 resultantes pueden ser expresadas bien en formato textual, legible por personas, y útil para búsqueda, edición y filtrado, o bien comprimidas en formato binario, más adecuado para almacenamiento y transmisión.

MPEG-21

MPEG-21 es el estándar más reciente del grupo MPEG. Su es definir un marco de trabajo que permita un uso transparente y mejor de los recursos multimedia a través de un amplio rango de redes y dispositivos.

La cadena del contenido multimedia va desde los creadores de contenido hasta la consumición del mismo. Para ello, el contenido tiene que ser identificado, descrito, administrado y protegido. El transporte y entrega del contenido tendrá lugar sobre un conjunto muy heterogéneo de terminales y redes, dentro del cual ocurrirán eventos que tendrá que ser registrados e informados. Aquí se incluye envío fiable, administración de datos personales y preferencias, de forma segura para garantizar la privacidad de éstos.

Se requiere un nuevo marco multimedia para soportar este nuevo tipo de uso del material multimedia, definido de forma que los sistemas que manejan contenidos multimedia sean interoperables, y para que las transacciones se simplifiquen, y si es posible estén automatizadas.

Hay dos conceptos esenciales en los que se basa el modelo de MPEG-21: la definición de un unidad fundamental de transacción y distribución (*Elementos Digitales*) y el concepto de usuarios interactuando con estas unidades:

- **Elemento Digital**: es un objeto digital estructurado con una representación e identificación estándar, y metadatos asociados dentro del marco MPEG-21. Esta entidad es la unidad fundamental de distribución y transacción dentro de este marco.
- **Usuario**: en MPEG-21 es una entidad que interactúa dentro del entorno MPEG-21 y puede hacer uso de Elementos Digitales. Tales Usuarios son individuos, consumidores, comunidades, organizaciones, corporaciones, consorcios, etc.

De las partes de MPEG-21, las más relevantes para este proyecto son Digital Item Declaration (DID), y Digital Item Adaptation (DIA). DID describe un conjunto de términos y conceptos abstractos para formar un modelo útil para definir Elementos Digitales. El modelo pretende ser tan flexible y general como sea posible. DIA proporciona las herramientas de metadatos necesarias para llevar a cabo la adaptación del elemento digital.

Otras tecnologías utilizadas

Aparte de los mencionados estándares MPEG-7 y MPEG-21, para la implementación del sistema se ha utilizado un amplio abanico de tecnologías software adicionales. El sistema web cliente-servidor se ha desarrollado sobre la plataforma Java. En el lado

servidor se utilizan servlets y páginas JSP. En el lado del cliente se utilizan varias applets de Java dentro de las páginas web para proporcionar un interfaz gráfico al usuario más cómodo y potente.

Además, debido a que MPEG-7 y MPEG-21 se construyen sobre XML, se utiliza una variedad de tecnologías asociadas a XML, como son las JAXP (APIs de Java para XML), el modelo DOM y las transformaciones XSLT, como pieza fundamental del motor de búsqueda y filtrado.

Perfiles de metadatos

El primer paso es considerar los metadatos sobre los que se va a basar el sistema, para definir las funcionalidades que se quieren implementar, y desarrollando los perfiles de metadatos necesarios. En este caso se utilizan tres tipos de descripciones: contenido, sesión y petición de búsqueda (*query*).

Perfil de contenido

Las descripciones de contenido se realizan mediante MPEG-7. En particular, este perfil se diseña de acuerdo con las herramientas de descripción de la parte 5 de MPEG-7 (Esquemas de Descripción Multimedia, MPEG-7 MDS). Como las herramientas de MDS son muy numerosas y complejas, se parte de la propuesta de perfil *Simple Profile* que hace el grupo MPEG. Esta propuesta simplifica en el sentido de que sólo utiliza descripciones multimedia (no utiliza herramientas de descripción de MPEG-7 Audio ni MPEG-7 Video) y no utiliza algunas de las herramientas más complejas de MDS. Sin embargo, el *Simple Profile* sigue siendo muy extenso, y contiene un conjunto muy grande de herramientas de descripción genéricas, y la mayoría de ellas no tienen mucha utilidad dentro del sistema planteado. Por esta razón, partiendo de el perfil *Simple Profile*, se seleccionan las herramientas de descripción necesarias, imponiendo las restricciones necesarias, para poder reducir la complejidad cuando sea oportuno.

Los dos conjuntos de herramientas de descripción más importantes que se utilizan en este perfil son:

- Herramientas de Información sobre la Creación del contenido (*CreationInformation*). Permiten anotar datos sobre la creación (como título, autor, lugar de creación, etc.) y clasificación del contenido multimedia, y otros contenidos relacionados con ese contenido.
- Herramientas de Información sobre *Media*¹ (*MediaInformation*). Permiten describir información sobre el almacenamiento del contenido, como puede ser la compresión, codificación y formato de los datos de cada uno de los *media*.

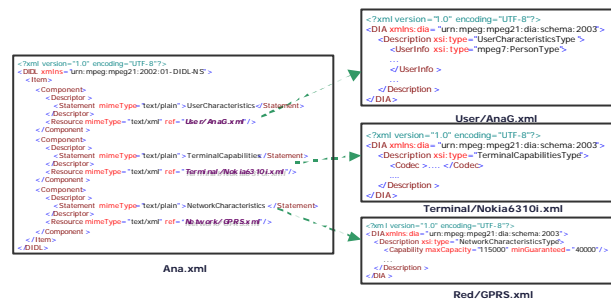
En la memoria del proyecto se detalla la semántica y las restricciones impuestas sobre el *Simple Profile* respecto al perfil de contenido utilizado.

Perfil de sesión

El perfil de sesión se utiliza para personalizar la aplicación para cada uno de los usuarios, y el entorno en el que se realiza la interacción. En este caso, el entorno de uso se refiere a la red y al terminal a través de los cuales se accede al contenido multimedia. Para ello se han utilizado MPEG-21, en concreto las partes *Digital Item Declaration* (DID) y *Digital Item Adaptation* (DIA).

¹ El término *Media* hace referencia a cada uno de los tipos de contenido audiovisual (audio, vídeo, imagen, etc.). En ese sentido, multimedia se refiere a contenido (generalmente sincronizado) con varios *media*

Para definir el perfil de sesión se ha optado por dividirlo en tres partes independientes (tres subperfiles), correspondientes a las características del usuario, las características de la red y las características del terminal. Cada una de estas descripciones se almacena independientemente de las otras en el servidor, y están relacionadas mediante otra descripción que simplemente señala cada una de las descripciones específicas que forman el perfil de sesión. Esta división permite gestionar independientemente cada una de estas descripciones específicas, al mismo tiempo que simplifica la tarea de definir su perfil al usuario, ya que puede reutilizar descripciones realizadas previamente (por ejemplo, puede seleccionar un modelo de terminal o red ya descrito con anterioridad).



Se tienen por tanto tres subperfiles:

- Perfil de usuario. Descrito en MPEG-21 DIA, que a su vez utiliza la parte de MPEG-7 dedicada a la anotación de preferencias de usuario.
- Perfil de terminal. Descrito en MPEG-21 DIA.
- Perfil de red. Descrito también en MPEG-21 DIA.

Perfil de búsqueda

Dado que aún no están desarrollados completamente (y en el momento del desarrollo del proyecto menos) los estándares de *queries* en documentos XML (como XQuery), y después de considerar otros formatos, se optó por tomar un enfoque diferente, en el cual se considera una petición de búsqueda como una descripción de un contenido al que el usuario quiere acceder. En este sentido se describe de forma similar en MPEG-7, anotando con las herramientas de descripción de MDS los campos en los que se quiere buscar de forma estructurada, similar al perfil de contenido, en un documento XML que será el que se envíe al servidor como petición de búsqueda.

Este tipo de enfoque permite realizar la búsqueda a partir de transformaciones XSLT, como se verá más adelante.

Sin embargo, el perfil no es compatible con MPEG-7 MDS, ya que éste no fue pensado para este tipo de descripciones de búsqueda, por lo que ha sido necesario utilizar un esquema modificado, en el que se relajan algunas restricciones y se utilizan algunos atributos adicionales para que la búsqueda sea más flexible.

Arquitectura del sistema

El sistema sigue el esquema clásico cliente-servidor, comunicados a través de Internet. En la Figura 1 se muestra la arquitectura general del sistema, en la que se prevé que el contenido sea accesible a través de diferentes terminales y redes, tanto móviles como fijas, mediante redes conectadas a Internet. En este proyecto se implementa la parte de redes fijas para terminales fijas (PCs), aunque el diseño del servidor es lo suficientemente independiente como para poder ser utilizado por otros tipos de terminales (por ejemplo, móviles).

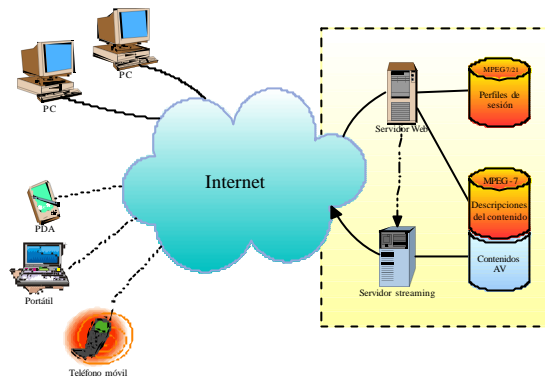


Figura 1: Esquema general del sistema

El sistema está a su vez compuesto por dos subsistemas diferenciados, aunque relacionados entre sí. En primer lugar está el subsistema web, que es aquel que se encarga de la búsqueda, filtrado a través de las descripciones almacenadas, y de la edición de estas y de los perfiles de sesión. Y por otro lado se encuentra el subsistema de *streaming*, que se encarga de adaptar el contenido multimedia, almacenado en la base de datos de contenidos, a partir de la información suministrada por el otro subsistema. El contenido que se sirve al cliente no es el que se encuentra almacenado en la base de datos de contenidos, sino una versión de éste convenientemente transcodificada en tiempo real a un formato que pueda utilizar el terminal y aprovechando al máximo las posibilidades que permita el terminal y la web.

Subsistema web

Esta parte del sistema está compuesta por una serie de aplicaciones web, compuestas básicamente por páginas HTML sobre las que puede ir incrustado objeto ActiveX o un applet de Java, según la página en cuestión, para dotar de mayor capacidad de procesado al cliente, especialmente para la edición y visualización intuitiva de las descripciones XML.

En la parte de servidor se utilizan servlets, y que se encargan de recoger los datos enviados por el cliente procesarlos e interactuar con las bases de datos de perfiles de sesión y/o descripciones de contenido.

En la Figura 2 se ilustra la relación entre las aplicaciones cliente y los servlets de servidor en el sistema web, y los formatos de datos utilizados en la comunicación entre ambas partes.

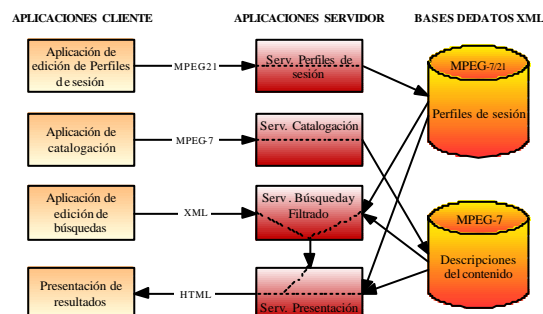


Figura 2: Esquema del sistema web

Cliente web. Applets de edición

La mayoría de las páginas web del sistema corresponden a herramientas para la edición de diferentes tipos de descripciones en XML, correspondientes a las descripciones de perfil de sesión (MPEG-7/21), contenido (MPEG-7) o búsqueda (MPEG-7 modificado). Por esta razón, se ha creado una interfaz de usuario genérica para este tipo de descripciones XML que luego se particulariza para cada herramienta en concreto.

Esta interfaz genérica se compone de los controles HTML (formularios), en la parte izquierda de la página, y un applet en la parte derecha. Los controles indican qué tipo de anotaciones se pueden añadir, y dependen de cada herramienta. En el applet se puede visualizar la descripción gráficamente, a medida que se van añadiendo campos a la estructura de datos de la descripción (que se corresponden con diferentes nodos XML). La visualización de la descripción se hace mediante un árbol gráfico en la interfaz, de forma que presenta la estructura de una forma más intuitiva y sencilla de manejar. Internamente esta estructura dará lugar al fichero MPEG-7/21 que será enviado al servidor.

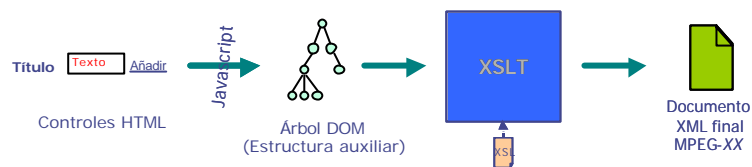


Figura 3: Funcionamiento del applet de edición

En la Figura 3 se resume el modelo de procesamiento de estas herramientas. Mediante el control web se añade un determinado campo en la estructura XML de la descripción almacenada en el applet (representada según el modelo DOM del W3C). La comunicación entre ambos se realiza mediante Javascript. Dada la enorme complejidad de los esquemas de los estándares utilizados (en especial de MPEG-7), esta estructura DOM se corresponde con una estructura XML auxiliar, mucho más simple y sencilla de gestionar, que además permite simplificar la interfaz de usuario. Para obtener el fichero XML final se utiliza una transformación XSLT, que transforma la estructura XML auxiliar en el fichero XML final, conforme al perfil de metadatos correspondiente.

Servidor web

Por una parte, el servidor web se encarga de realizar las tareas relacionadas con el almacenamiento de las descripciones de contenido y perfiles de sesión.

Otra tarea de la que se encarga el servidor web es la relacionada con la búsqueda y filtrado (ver Figura 4). Cuando el servidor recoge una petición de búsqueda (en XML según el perfil explicado anteriormente), la pasa a través de un módulo de búsqueda y filtrado. En el filtrado la búsqueda se hace sobre las preferencias que el usuario tiene almacenadas sobre su perfil de sesión. Tal como está implementada la búsqueda, se hace mediante transformaciones XSLT, utilizando dos transformaciones encadenadas. Como resultado de estas dos transformaciones, el resultado es un documento XML con la lista de contenidos que satisfacen los criterios de la búsqueda y filtrado, con un valor de relevancia asociado calculado a partir de los criterios de búsqueda establecidos para cada regla de transformación (por ejemplo, se valora más una coincidencia en el título que una coincidencia en el año, y por supuesto, se valora más la coincidencia en ambos). Este documento de resultados, pasa a través de otra transformación XSLT de presentación, que se encarga de transformar el documento en un documento final según un formato (en este caso HTML, aunque también podría ser WML, PDF, etc.) y una composición y diseño concretos. La plantilla de presentación (XSL de la transformación

de presentación) se determina a partir del perfil de sesión. El documento ya formateado se envía de vuelta al cliente para que pueda visualizar los resultados de la búsqueda.

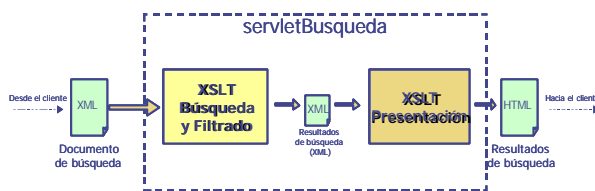


Figura 4: Funcionamiento de la búsqueda mediante XSLT

El procesado de la búsqueda y filtrado en XSLT se basa en las reglas de transformación definidas en el modelo de XSLT. Se basan en la correspondencia de patrones (*pattern-matching*), y una regla concreta básicamente define la dirección XPath dentro de la descripción MPEG-7 donde se encuentra un campo concreto (por ejemplo, el título). Se encarga de añadir un contenido como resultado de la búsqueda si lo que está anotado en su descripción MPEG-7 cumple el criterio de búsqueda para esa dirección XPath.. Se utiliza Xpath para direccionar los nodos de forma similar a como se hace en el reciente lenguaje XQuery.

Por último, el servidor web se encarga de presentar la información de un contenido de un contenido concreto cuando éste es seleccionado entre los resultados de una búsqueda. Se utiliza una transformación XSLT de presentación análoga a la descrita para la búsqueda y filtrado, en la que se transforma la descripción MPEG-7 del contenido a un formato tal como HTML. En esta página además se encuentran incrustados el applet y el objeto ActiveX del reproductor multimedia necesarios para reproducir el contenido multimedia, y que suponen la interfaz del subsistema de *streaming*.

Arquitectura basada en XSLT

Como se ha visto, las transformaciones XSLT tienen un papel fundamental en el sistema web, tanto en el servidor como en el cliente. De hecho, todo el sistema web, extremo a extremo, se puede entender como una cadena de transformaciones XSLT. Por esta razón, se han diseñado las transformaciones en el lenguaje declarativo de XSLT, y se deja en manos del procesador XSLT la implementación concreta de las mismas. Los procesadores XSLT están diseñados para procesar en paralelo diferentes reglas al mismo tiempo y optimizados para realizar las transformaciones de forma muy eficiente. El diseño del sistema basado en transformaciones XSLT se debe en parte a la poca madurez de las soluciones de bases de datos XML en el momento de realizar el diseño del sistema (y actualmente siguen siendo muy ineficientes y limitadas). No existe un modelo de bases de datos suficientemente satisfactorio que represente la estructura y tipos de datos de documentos XML/XML Schema, y menos para MPEG-7². Además tampoco está suficientemente claro que vaya a ser más eficiente que hacerlo mediante el modelo utilizado finalmente. Además el modelo de transformaciones utilizado se basa en localizar los nodos mediante XPath de forma similar a como lo hace XQuery, lo que permitiría en un futuro utilizar XQuery como lenguaje de query en el sistema.

Por esta razón se optó por el diseño del sistema en términos de transformaciones XSLT. El uso de transformaciones XSLT para realizar búsquedas es también un enfoque

² Para un análisis exhaustivo de las actuales soluciones de bases de datos XML, y sus limitaciones para gestionar descripciones MPEG-7 (y XML en general) véase “*Analysis of XML Database Solutions for the Management of MPEG-7 Media Descriptions*”, Utz WesterMan, Wolfgang Klas, Universidad de Viena

novedoso. Casi toda la literatura referente a XSLT se encuentra relacionada con las transformaciones orientadas simplemente a formatear el documento, generalmente a HTML (transformaciones que también se utilizan en este sistema), de forma más o menos simple. Pero no se encontró ninguna referencia a propuestas similares, en las que se utilice XSLT para realizar transformaciones más o menos complejas del tipo a las utilizadas en el módulo de búsqueda y filtrado.

No pretende ser el enfoque más eficiente, sino simplemente el que se consideró más oportuno en su momento. En cualquier caso, los resultados obtenidos son bastante satisfactorios, con un retardo de procesamiento muy pequeño, al menos en el sistema implementado.

Subsistema de streaming

El subsistema de *streaming* se encarga de adaptar/transcodificar el contenido multimedia de la base de datos de contenido según los recursos de red y terminal que utilice el usuario (anotados en su perfil de sesión), y las características del contenido seleccionado. Se utiliza transcodificación en tiempo real, utilizando la tecnología de la plataforma Windows Media Series 9, por las posibilidades que ofrece (aunque limita por el momento la aplicación a sistemas Windows).

Cliente

El cliente es básicamente el reproductor Windows Media Player 9 incrustado en la interfaz de la página web de visualización del contenido. Para comunicar y sincronizar la petición de visualización del contenido transcodificado entre el cliente y el servidor web, se utiliza un applet auxiliar, que además cuenta con un interfaz que permite modificar los parámetros de transcodificación.

Servidor

En el servidor se utiliza Windows Media Encoder 9, programado a través de su SDK para que de forma automática cree un proceso en el que transcodifique un contenido multimedia de acuerdo con los parámetros especificados en el perfil de sesión o bien los especificados por el cliente.

Para que el servidor pueda interactuar con el cliente mediante la página web se utiliza un servlet auxiliar para la sincronización con éste, y que será el que lance el proceso (externo al servidor web) de transcodificación.

Resultados

La interfaz de usuario

Para resumir el funcionamiento del sistema desde el punto de vista de usuario, se van a presentar a continuación algunas capturas de las diferentes herramientas web de la interfaz, mostrar el resultado final.

La interfaz está organizada mediante un portal web, desde el cual se pueden acceder a las diferentes herramientas una vez se haya introducido el nombre del usuario. Este nombre está asociado a una descripción MPEG-21 del perfil de sesión, y será la que se utilice para personalizar el acceso y la búsqueda y filtrado.

A través de las páginas web se accede a las distintas herramientas de catalogación, edición de perfiles y búsqueda y filtrado (ver Figura 5). A través de ellas se anotan, viendo la estructura del documento editado, las diferentes descripciones que serán posteriormente enviadas al servidor. En el caso de la búsqueda esta descripción se procesa y se recibe en el cliente una lista con los resultados de la búsqueda/filtrado (ver Figura 6).

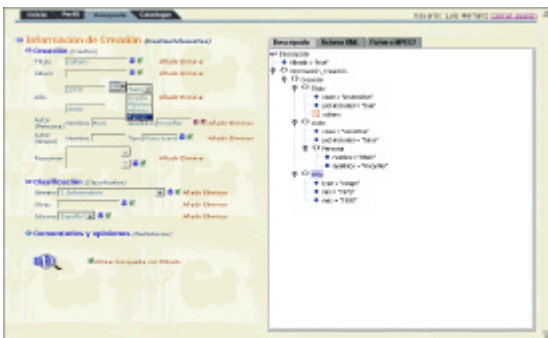


Figura 5: Herramienta de búsqueda



Figura 6: Pantalla de resultados de búsqueda

Cuando se selecciona un contenido, aparece un pantalla con información sobre el contenido seleccionado, y desde ella se puede reproducir el propio contenido multimedia, a través una reproductor incrustado en la propia página. Además, si el contenido cuenta con subtítulos, se visualizan automáticamente (ver Figura 7).



Figura 7: Pantalla de navegación y reproducción del stream



Figura 8: Otro perfil de presentación de los resultados de búsqueda

Por último, como ya se ha mencionado, la interfaz web está personalizada para cada usuario mediante el uso de plantillas XSL de presentación. La Figura 8 muestra la pantalla de los resultados de búsqueda con otra plantilla diferente, en la que se utiliza una composición diferente y el interfaz en otro idioma (inglés).

Conclusiones sobre el sistema

En general el sistema desarrollado cumple con los objetivos propuestos. Aunque el sistema no es realmente de acceso universal, ya que se limita a terminales PC con sistema Windows, sirve como prototipo de escenario UMA a través de las herramientas de MPEG-7 y MPEG-21. Desde el punto de vista de servidor, el sistema web es mucho más independiente, ya que se basa en tecnologías portables y no dependientes del sistema operativo, como son XML y Java, lo que permite una extensión sencilla para el acceso desde terminales diferentes³.

La búsqueda y el filtrado son muy rápidas, aunque hay que tener en cuenta que el número de descripciones utilizadas es reducido (unas 15-20). No se ha probado con un número de descripciones suficientemente elevado, propio de un sistema comercial, pero posiblemente la búsqueda y filtrado fuesen algo lentas, y sería necesario incorporar alguna solución de bases de datos. Sin embargo, como ya se ha mencionado, las soluciones de bases de datos XML no están del todo maduras, y es posible que la mejora en velocidad no fuese sustancial.

El mayor inconveniente del sistema es la gran cantidad de recursos empleados en la adaptación del contenido, debido a que el módulo de transcodificación debe decodificar el vídeo original y volver a codificarlo según los parámetros de adaptación y servirlo en *streaming* en tiempo real. Sin embargo, desde el punto de vista de resultados es bastante satisfactorio, ya que se adapta bien a los parámetros fijados y el retardo no es demasiado elevado, obteniendo además una calidad bastante buena, incluso en casos con tasas binarias bajas.

Hay que destacar que este sistema no sería un sistema comercial, sino que es simplemente un proyecto de demostración de las capacidades que tiene tanto MPEG-7 como MPEG-21 en un escenario UMA. Por esta razón, las herramientas de catalogación y perfiles son más complejas y permiten añadir anotaciones y datos más avanzados que no tienen sentido para un usuario común.

El sistema desarrollado también sirve de demostración de las múltiples posibilidades de las tecnologías Java y XML, y el resto de tecnologías asociadas, para crear aplicaciones web de nueva generación. En concreto, la tecnología XSLT ha sido especialmente importante en el desarrollo del sistema.

Y en relación con XSLT, destacar la personalización del interfaz a través de la hojas de estilo de presentación, que además de la personalización estética, permite variantes más funcionales, como la personalización del lenguaje, estructura, etc.

Por último destacar las posibilidades que permite el uso de subtítulos sincronizados con el vídeo, que han sido añadidos como una característica adicional al sistema. En primer lugar, el uso más inmediato, e importante, es el de proporcionar acceso al contenido hablado a personas con discapacidades auditivas, necesario si se quiere proporcionar un acceso realmente universal. Sin embargo, este tipo de características permiten también crear aplicaciones educativas y recreativas, por ejemplo, aplicaciones para aprender idiomas, en el primer caso, o aplicaciones de tipo karaoke, en el segundo.

³ De hecho el sistema desarrollado en este PFC se ha extendido a terminales y redes móviles, utilizando el mismo servidor web sin prácticamente ninguna modificación. Para más información ver "Acceso Multimedia Universal mediante MPEG-7 y MPEG-21 a través de redes y terminales móviles", Ana García Puerta, Proyecto de Fin de Carrera, E.T.S.I.T. UPM, Marzo 2004

Publicaciones relacionadas

El presente proyecto de fin de carrera ha dado lugar a dos publicaciones en congresos de relevancia internacional:

- “Towards universal and personalized access to audiovisual content in the DYMAS system”, L. Herranz and J. M. Martínez, Proceedings of the International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS’2004, Lisboa, Portugal, abril 2004. (en prensa)
- “Universal and Personalized Access to Content via J2ME terminals in the DYMAS system”, A. García, J. M. Martínez and L. Herranz, International Conference on Image and Video Retrieval CIVR2004, Dublín, Irlanda, julio 2004 (en prensa)

Utilización del sistema en otros proyectos

DYMAS

Este proyecto ha sido incorporado como parte de las infraestructuras del proyecto TIC2002-03692 DYMAS “Obtención e inserción automática de descripciones MPEG-7 en flujo DVB” (2002 – 2004), del Ministerio de Ciencia y Tecnología. En este proyecto participan la Universidad Politécnica de Madrid y la Universidad Autónoma de Madrid. El sistema DYMAS permite la generación de aplicaciones a partir de flujos de televisión digital (*MPEG-2 Transport Streams* sobre DVB). Consiste en dos entornos (ver Figura 9), el primero (Entorno de Tiempo Real, RTE) para la extracción de características y generación de metadatos a partir de los flujos MPEG-2 de entrada, y la provisión de un Transport Stream de salida enriquecido mediante metadatos, aplicaciones MHP sobre DVB, etc., de forma que se puede reproducir en un receptor de televisión digital. El segundo entorno (Entorno de Tiempo Diferido, DTE) se encarga de aplicaciones interactivas en las que se permite un cierto retardo. De esta forma se pueden generar aplicaciones para diversos entornos diferentes, como son portales web, vídeo bajo demanda, acceso multimedia desde terminales móviles, mensajería multimedia, etc.

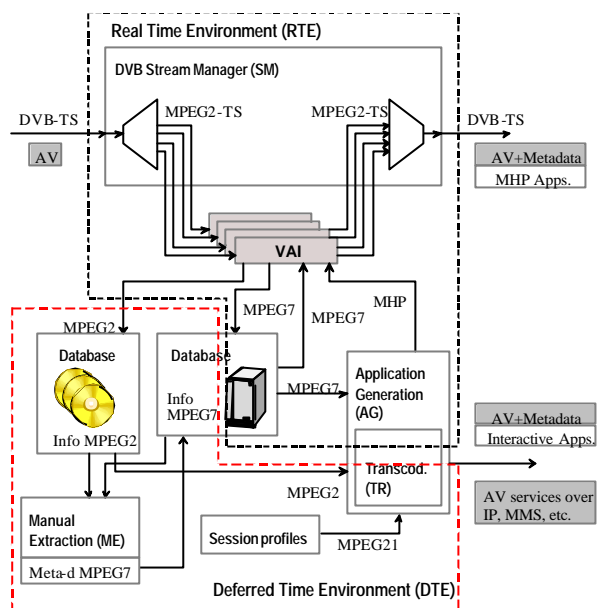


Figura 9: Arquitectura del sistema DYMAS

El sistema desarrollado en este Proyecto de Fin de Carrera se ha incorporado en el Entorno de Tiempo Diferido para proporcionar acceso multimedia mediante perfiles de sesión MPEG-21 y a través de las herramientas web desarrolladas, y utilizando los servicios de transcodificación y adaptación implementados en él.

aceMedia

Parte de este proyecto se está incluyendo en el proyecto aceMedia (“Integrating knowledge, semantics and content for user-centred intelligent media services”), proyecto integrado IST del VI Programa Marco de la Unión Europea (Febrero 2004 - 2008), como parte del desarrollo que hace la Escuela Politécnica Superior de la Universidad Autónoma de Madrid sobre tareas de transcodificación de contenido multimedia utilizando metadatos MPEG-7 y MPEG-21. En el proyecto participan las siguientes entidades: Motorola Ltd, Philips Electronics Nederland, Thomson, Queen Mary University of London, Fraunhofer FIT, Universidad Autónoma de Madrid, Fratelli Alinari, Telefónica Investigación y Desarrollo, Dublin City University, Informatics and Telematics Institute, INRIA, France Télécom, Belgavox, University of Karlsruhe, Motorola SAS.

Otros méritos

Becas

El proyecto fue seleccionado en la convocatoria 2002 de las becas para la realización del Proyecto de Fin de Carrera de la cátedra Telefónica de la UPM para Internet de Nueva Generación. Duración de la beca: desde septiembre de 2002 hasta mayo de 2003.