

Plataforma multimedia MEX-CULTURE

MIREYA SARAÍ GARCÍA VÁZQUEZ

Instituto Politécnico Nacional, México

YALJA MONTIEL PÉREZ

Instituto Politécnico Nacional, México

JENNY BENOIS-PINEAU

Univeristy Bordeaux, Francia

MICHEL CRUCIANU

Conservatoire National des Arts et Métiers (CNAM), Francia

PLATAFORMA MULTIMEDIA MEX-CULTURE

Con la introducción de los primeros sistemas de búsqueda de información basados en contenido visual (Content Based Visual Information Retrieval, CBVIR), tales como el sistema Virage (M. Flickner, H. Sawhney, W. Niblack, J. Ashley, 1995) o el sistema QBIC (A. Gupta, Bach, C. Fuller, Chiao-Fe Shu, 1996) se desarrolló una investigación fértil en este campo y se implementaron herramientas disponibles de manera pública, como el buscador de imágenes de Google (Google Image Search). Sin embargo, la práctica de más de dos décadas muestra que la información visual es sólo una parte de un dato complejo y heterogéneo que actualmente el usuario está dispuesto a buscar. Además, el uso inteligente de información diferente en un conjunto

heterogéneo que representa a un documento multimedia de hoy en día ayuda a aumentar la eficiencia y la solidez de búsqueda. El segundo reto es la dimensionalidad y el volumen de los datos conocidos en la actualidad como “desafío de grandes datos (*big data challenge*)”. Estas dos dimensiones del problema tienen que ser abordadas no sólo por los algoritmos “escalables” y eficientes, sino también por una arquitectura extensible de un sistema de búsqueda de información, que aprovecharía la heterogeneidad de la información y proporcionaría la calidad de servicio requerida por el usuario. Además, debe garantizar la accesibilidad de los servicios en cualquier momento y en cualquier lugar. Este tipo de sistema debe ser construido sobre tecnologías web. En este trabajo se propone un sistema de búsqueda de la información multimedia, adaptado para la exploración de grandes archivos audiovisuales digitales, en el campo de la preservación y el acceso al patrimonio cultural. La plataforma multimedia Mex-Culture está inscrita en estos términos y es el resultado de un esfuerzo conjunto de investigadores franceses y mexicanos del proyecto Mex-Culture, dirigido a garantizar el acceso al primer repositorio digital de información heterogénea, tal como el contenido de audio, los documentos audiovisuales o las imágenes. Éstos, reunidos y digitalizados por diversas fuentes: la Fonoteca Nacional, la televisión mexicana y otras colecciones de imágenes. En las secciones siguientes, se presenta la arquitectura del sistema, así como sus servicios y herramientas de análisis de contenido que le permiten garantizar estos servicios. Los aspectos de la arquitectura se presentan en la sección arquitectura del sistema Mex-Culture y los servicios, en este marco de trabajo, se describen más adelante, en la sección servicios media. Se presta atención específica a la recolección de da-

tos y a su naturaleza en la sección colección de datos. Finalmente, se presentan conclusiones y perspectivas.

ARQUITECTURA DEL SISTEMA MEX-CULTURE

La infraestructura multimedia de la plataforma Mex-Culture está basada en una Arquitectura Orientada a Servicios (SOA, por sus siglas en inglés). El SOA difiere del modelo general cliente/servidor tanto en su énfasis en el acoplamiento débil entre los componentes de un software como en el uso de interfaces permanentes por separado (Thomas, 2005; R. W. Schutle, 1996). El acoplamiento débil en los sistemas o soluciones en software se refiere a que los componentes (servicios) mantienen una relación que minimiza las dependencias y sólo se requiere mantener un conocimiento de uno al otro.

El SOA es un enfoque de arquitectura que permite realizar:

- Implementación distribuida: gestión de datos con unidades estructuradas basadas en estándares.
- Reutilización: servicios de reutilización.
- Componibilidad: montar nuevos procesos de servicios existentes que están expuestos a través de una granularidad deseada bien definida, publicada y de estándares de interfaces compatibles.
- Interoperabilidad: capacidad de compartir y reutilizar los servicios compartidos a través de una red, independientemente de los protocolos subyacentes o la tecnología de aplicación.

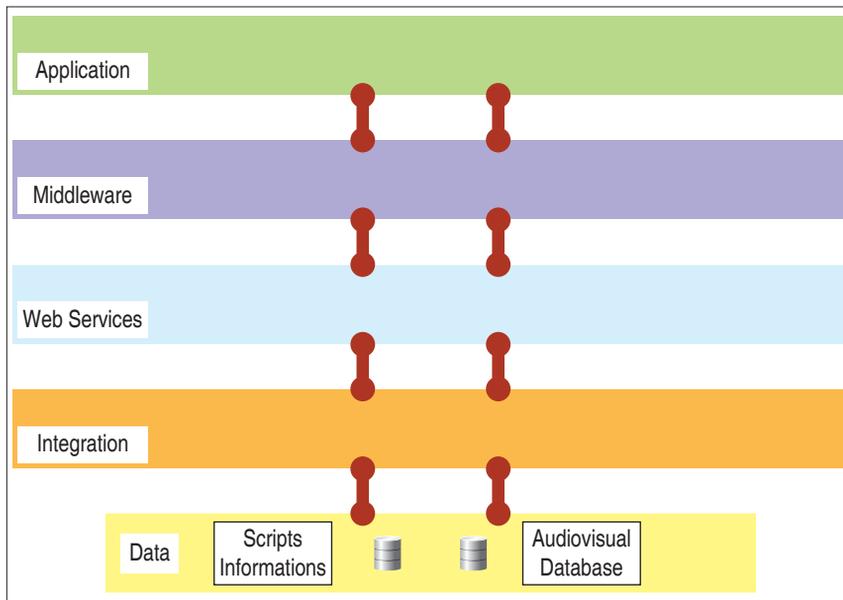
La arquitectura de Mex-Culture está basada en un servicio web específicamente diseñado para componentes de

procesamiento multimedia. Por lo tanto, permite la integración de componentes diseñados de forma independiente y facilita la integración de la mejor variedad de módulos.

Como los servicios web eliminan los problemas técnicos de heterogeneidad, hemos desarrollado un modelo de intercambio específico que soporta la interoperabilidad de los componentes multimedia heterogéneos.

El sistema Mex-Culture, basado en una arquitectura orientada a servicios, incluye cinco niveles que permiten una implementación y operación eficiente (*Figura 1*).

Figura 1.
Sistema Mex-Culture basado en SOA



Fuente: elaboración propia.

El sistema Mex-Culture está compuesto de:

- *Datos*: esta parte está compuesta de información *cross-media* y la información basada en metadatos.
- *Integración*: componente base; esta parte facilita la integración e interacción de los componentes para hacerles el procesamiento de base a los datos no-estructurados de todo tipo (texto, imagen, video, voz, audio).
- *Servicios web*: éstos son servicios de procesamiento e interfaz hombre-máquina. Cada servicio web realiza un procesamiento particular que puede ser insertado en una cadena y asociar éste con otros servicios web para satisfacer una necesidad específica.
- *Middleware*: un *Bus* de comunicaciones que provee la comunicación entre diferentes aplicaciones, las cuales no necesariamente están hechas para trabajar juntas.
- *Aplicación*: les permite a los usuarios acceder a los servicios de una forma oculta a través de un simple navegador web.

La aplicación web del sistema Mex-Culture está desarrollada en un servidor de aplicación JEE (Java Enterprise Edition). Este servidor web es un software de código abierto desarrollado por la Fundación de Software Apache (ASF, por sus siglas en inglés) (Apache-tomcat, 2014).

El diseño del sistema Mex-Culture consta de un conjunto extenso de servicios, pero no todos ellos tienen la misma madurez desde el punto de vista de su integración. Tales servicios están dirigidos a diferentes tareas del usuario, y responden al acceso fácil y escalable del contenido en su forma original o desde su forma previa. En la siguiente sección presentaremos algunos de ellos.

SERVICIOS MEDIA

Servicio resumen

El resumen del video es una representación compacta del contenido de éste, el cual provee el acceso a la información más relevante basada en similaridad. Dada la importancia del problema de búsqueda y recuperación de la información en datos de gran escala, se han propuesto varios enfoques para hacer el resumen del video (X. Jinm, J. Han, L. Cao *et al.*, 2010; J. Almeida, N. J. Leite, R. da S. Torres, 2013); y la campaña TRECVID ejecutó una tarea específica llamada *sumarization rushes* (E. Rossi, S. Benini, R. Leonardi, B. Mansencal, 2009).

El servicio de resumen provee un resumen del video escalable en términos del contenido de la media. Este enfoque está inspirado de las operaciones del cubo de datos en Procesamiento Analítico en Línea (OLAP, por sus siglas en inglés) (J. Gray, 1997).

El concepto de cubo de datos se eligió para facilitar la navegación al usuario a través del espacio multidimensional de descriptores de contenido de cada documento audiovisual.

Las características del video y el audio son calculadas a diferentes tasas de muestreo, y su sincronización está asegurada por una interpolación (K. Perez-Daniel, 2014).

Para el nivel de integración se han desarrollado tres componentes: i) extracción de las características visuales del *keyframe* (imagen clave o de referencia), ii) extracción de las características del audio y iii) producción del cubo de datos. A continuación se presenta la arquitectura de este servicio. En general, la descripción de los otros servicios sigue la misma arquitectura.

Nivel de integración: Componente - Características de los Key Frames

Características visuales: para describir la información visual nos limitamos a los descriptores globales de la imagen. Para el color, usamos las características bien conocidas de MPEG-7, como el Descriptor de la Estructura de Color (CSD, por sus siglas en inglés) (D. S. Messing, P. Van Beek, J. H. Errico, 2001) y el Descriptor Escalable de Color (SCD, por sus siglas en inglés) (W. Yong-ge, 2012). Las características estructurales utilizadas son los Histogramas Piramidales de Gradientes Orientados (PHOG, por sus siglas en inglés), como se describe en K. Perez-Daniel, M. Nakano Miyatake y J. Benois-Pinneau (2014).

Nivel de integración: Componente - Características del audio

Características del audio: nosotros consideramos las descripciones del audio a través de los descriptores Mel-Frequency Cepstral Coefficients (MFCC) y los vectores Chroma. Un MFCCS es un grupo de coeficientes que codifica la forma del envolvente espectral de la señal (V. Peltonen, 2002). Un vector Chroma es un grupo de coeficientes que cuantifica la energía asociada a los doce semi-tonos del espectro en la teoría de la música occidental (V. Peltonen, 2002; M. V. Peltonen, J. Tuomi, A. Klapuri, J. Huopaniemi, T. Fillon, J. Prado, G. Richard, 2010).

Nivel de integración: Componente - Producción del cubo de datos

El cubo de datos ofrece flexibilidad para la navegación en los datos, al desplegar un resumen de diferente nivel de granularidad.

Nivel de aplicación: Portlet - Resúmenes del video

La presentación y la navegación a través de resúmenes escalables se realizan utilizando una interfaz basada en web. Esta interfaz le permite al usuario navegar entre varios niveles de detalle y dimensiones del resumen escalable.

La interfaz web es desarrollada como una *portlet* para el servidor del portal de código fuente abierto escrito en Java.

El lado del cliente utiliza un reproductor de video, un deslizador y una representación estructural en árbol para permitirle al usuario una navegación fácil en el cubo de datos que representan el resumen del video.

Servicio por consulta de acción

Este servicio está diseñado específicamente para investigaciones audiovisuales de contenido cultural mexicano. Uno de los elementos típicos en la programación televisiva de la cultura mexicana son los reportajes de los espectáculos tradicionales sobre corridas y bailes populares mexicanos. Por lo tanto, la consulta por acción, como “embestida del toro a la capa” o “equitación de caballos”, es requerida para el escenario de búsqueda del usuario.

Nuestro sistema le permite al usuario localizar con precisión los casos de acciones en nuestra base de datos de alta escala.

COLECCIÓN DE DATOS

En el proyecto Mex-Culture son exploradas dos fuentes de contenido audiovisual. Estas dos colecciones de datos han sido formadas por ser representativas de la preservación y difusión del contenido. El primer corpus fue seleccionado por el Instituto Nacional del Audiovisual en Francia (INA), que es también una institución de investigación. El segundo corpus fue compuesto por diferentes fuentes de organismos mexicanos con canales digitales del tipo transmisión de dominio público, Fonoteca Nacional, Canal 11, televisión mexicana, etcétera.

Corpus INA

Ésta es una base de datos que incluye 1 249 videos sobre 14 categorías de contenido cultural, tiene una duración de 2 000 horas de video y fue diseñada por el INA. La selección del contenido audiovisual siguió el criterio definido por el objetivo de la plataforma Mex-Culture, así como el almacenamiento y la facilidad de acceso a los contenidos culturales de México.

México – se realizaron búsquedas en el contenido audiovisual de temas relacionados en los archivos y categorías audiovisuales del INA.

Entre estas categorías se pueden encontrar: “danza mexicana”, “danza folclórica”, “huapango”, “mariachis”, “norteño”, “paisajes mexicanos”, “sitios arqueológicos mexicanos”, “corrida”, “Thalassa...” Finalmente, se seleccionaron 828 archivos de contenido codificado en MP4.

Colección nativa mexicana

Las 32 entidades territoriales de México están agrupadas en regiones de acuerdo con el INEGI (Instituto Nacional de Estadística y Geografía, organismo del gobierno mexicano) (Regiones de México, 2015). Esta división regional de México fue establecida por la combinación de factores naturales, históricos y culturales (población, economía, etcétera). Por tanto, ésta es la base del sistema Mex-Culture para caracterizar a la cultura mexicana (*Figura 2*).

Figura 2.
Las ocho regiones de México



Fuente: Regiones de México, 2015.

Una de las características del sistema Mex-Culture es permitir la creación de base de datos de documentos audiovisuales con contenido variado de la cultura mexicana.

Además, se ha trabajado en la automatización de diversas tareas de adquisición y procesamiento, así como en el diseño del uso de estos documentos digitales. La caracterización de las bases de datos consiste en tres aspectos principales de contenidos basados en imagen y video, y tres aspectos con base en el contenido de la voz y el audio.

Aspectos de caracterización de la imagen y el video:

Paisajes naturales

- Descripción de las principales características de cada región. Los factores físicos son tomados en cuenta, y todos los elementos de la naturaleza, como la vegetación, el agua, el cielo, el clima, las características geográficas, etcétera.

Población

- Descripción de las principales características de cada región. Grupos étnicos, bailes mexicanos y corridas de toros.

Arquitectura

- Descripción de las principales características de cada región: prehispánica, colonial y moderna.

Aspectos de caracterización de la voz y el audio:

Lenguas indígenas

- Descripción de las principales características de cada región.

Clases de audio Fonoteca Nacional

- Caracterización de las cinco clases sonoras establecidas por la Fonoteca Nacional.

Figuras importantes en la vida de México

- Reconocimiento de figuras importantes de la vida nacional.

La selección del contenido audiovisual de acuerdo con estos criterios ha sido realizada utilizando la información pública disponible en los canales digitales del tipo web (*dailymotion, youtube, metacafe* y otros) del IPN Canal Once, México Hoy, O2 México, etcétera.

Obviamente, en una colección multimedia de gran escala, el problema más difícil es la anotación (verdad absoluta, *ground truth*) del contenido para evaluar comparativamente los diferentes servicios de la media. Para desarrollar la plataforma se siguió una tarea de anotación para crear las bases de datos para evaluar las partes de la plataforma Mex-Culture.

CONCLUSIONES

Este trabajo ha presentado la plataforma multimedia Mex-Culture para almacenar y acceder a los contenidos digitales de la cultura mexicana. Se describieron las capas de la arquitectura para la plataforma multimedia Mex-Culture y se explicaron con detalle los servicios, utilizando como ejemplo el servicio de resumen. Los servicios y los componentes de cada capa se han diseñado con una escalabilidad, ya que el prototipo de la plataforma multimedia Mex-Culture es el primer paso hacia la introducción de la indexación multimedia y la recuperación de información de la cultura en el país bajo este esquema. La fase actual del proyecto ha integrado sólo parcialmente los servicios visualizados, por lo que la finalización de la integración será uno de los trabajos futuros.

RECONOCIMIENTO

Esta investigación fue apoyada por el programa bilateral ANR-CONACYT ANR BLANC II Mexculture núm. 11IS02 00102. Registro en el Instituto Politécnico Nacional: SIP-2012-RE/64.

BIBLIOGRAFÍA

- Almeida, J., Leite, N. J., Torres, R. da S. (2013). Online video summarization in compressed domain. *Journal of Visual Communication and Image Representation*, 24, 729-738.
- Apache-tomcat (2014). Stable release 7.0.40 [en línea], <http://tomcat.apache.org>

- Benoit, M., Essid, S., Fillon, T., Prado, J., Richard, G. (2010). YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software. En *Proceedings of the 11th International Society for Music Information Retrieval (ISMIR)* (pp. 441-446).
- Erl, T. (2005). *Service-Oriented Architecture: concepts, technology and design*. Boston: Pearson Education.
- Flickner, M., Sawhney, H., Niblack, W., Ashley, J. (1995). Query by image and video content: the QBIC system. *IEEE Computer*, 28 (9), 23-32.
- Gray J. *et al.* (1997). Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub totals. *Data Min. Knowl. Discov*, 1(1), 29-53.
- Gupta, A., Bach, J., Fuller, C., Chiao-Fe Shu (1996). The Virage image search engine: an open framework for image management. En *Storage and Retrieval for Image and Video Databases IV*, Proc. SPIE 2670 (pp. 76-87).
- Jin, X., Han, J., Cao, L. *et al.* (2010). Visual cube and on-line analytical processing of images. En *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM)* (pp. 849-858).
- Messing, D. S., Van Beek, P., Errico J. H. (2001). The mpeg-7 color structure descriptor: Image description using color and local spatial information. En *Proceedings of the International Conference on Image processing* (pp. 670-673).
- Peltonen, V., Tuomi, J., Klapuri, A., Huopaniemi, J. (2002). Computational auditory scene recognition. En *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 1941-1944).

- Perez-Daniel, K., Nakano Miyatake, M., Benois-Pineau, J. (2014). Scalable video summarization of cultural video documents in cross-media space based on data cube approach. *CBMI*: 1-6.
- Regiones de México (2015) [en línea], http://es.wikipedia.org/wiki/Regiones_de_M%C3%A9xico.
- Rossi, E., Benini, S., Leonardi, R., Mansencal, B. (2009). Clustering of scene repeats for essential rushes preview. En *Workshop on Image Analysis for Multimedia Interactive Services* (pp. 234–237).
- Schutle, R. W., Natis, Y. V. (1996). Service Oriented Architectures. *Part 1. SSA Research Note SPA*. Gartner, Abril 12, 401-468.
- Yong-ge, W., Peng Sheng-ze, (2012). Research on image retrieval based on scalable color descriptor of mpeg7. En *Advances in Control and Communications* (pp. 91-98).