

CUDI 2015

REUNIÓN DE PRIMAVERA

21 AL 24 DE ABRIL

Puerto Vallarta, Jal.

MEX - CULTURE

Multimedia libraries indexing for the preservation and dissemination of the Mexican Culture

Proyecto Bilateral Francia-México

Dra. Mireya García-Vázquez

Instituto Politécnico Nacional -CITEDI

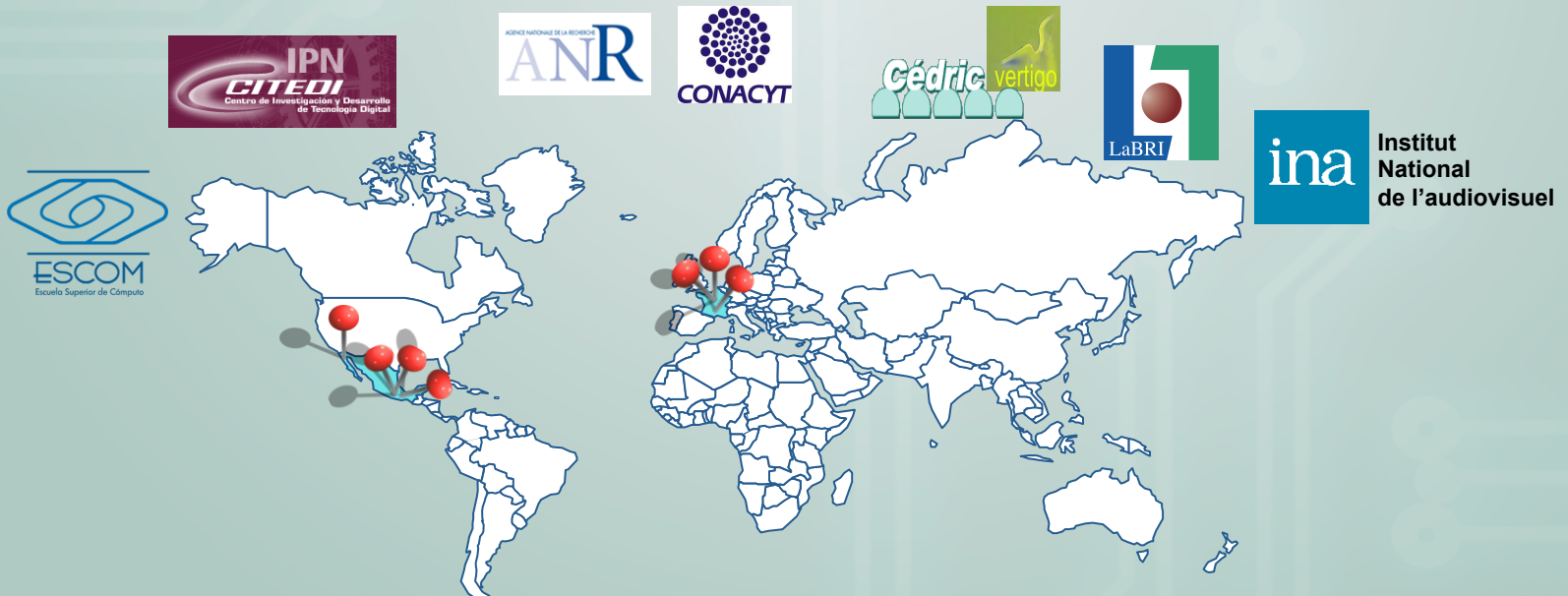
msarai@ipn.mx , mgarciav@citedi.mx



22 de abril del 2015

Consortium

- Start-End Date: 21 November 2012 – 20 November 2015
- Scientific coordinator: CITEDI-IPN, CEDRIC-CNAM
- Consortium: 5 partners from 2 countries
- (MX) CITEDI-IPN, ESCOM-IPN
- (FR) CEDRIC, CNAM, LABRI UMR, CNRS, INA



- ❖ The MEX-CULTURE project aims to:
 - ❖ Devise, with bilateral research efforts, new automated methods for large-scale processing of encoded visual content (video, audio, images and speech) focused on:
 - ❖ Indexing
 - ❖ Retrieval
- } of information

❖ These methods concern:

- ❖ video indexing directly performed on encoded visual content,
- ❖ audio indexing for speech recognition and speaker identification,
- ❖ cross-media indexing (image, video, audio and speech),
- ❖ scalable search and retrieval.

❖ Project framework:

- ❖ Strictly academic and research.

❖ Application

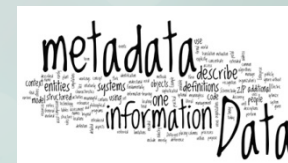
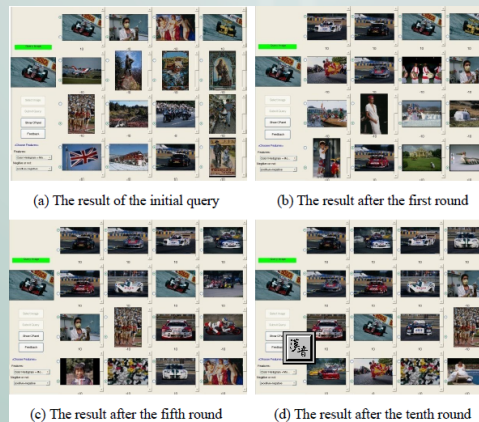
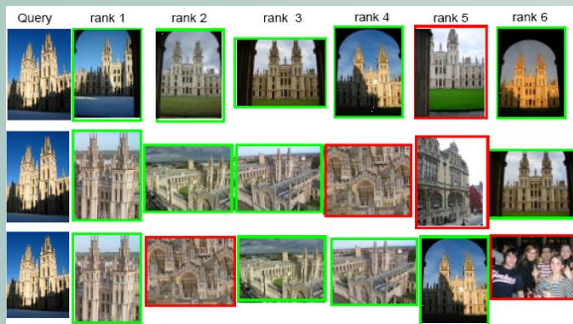
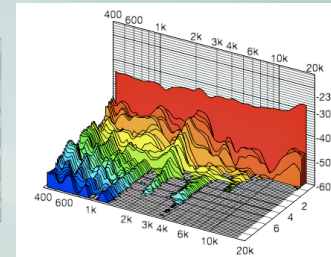
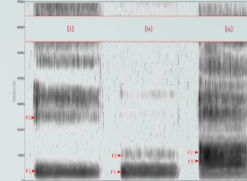
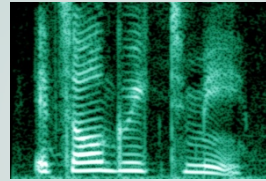
- ❖ Support the **preservation and dissemination of Mexican Culture** by applying the methods devised in the project to the **large databases** more than **100.000 hours** of video/audio (, “Fonoteca Nacional”, “DL-INAH”) .



Develop:

- ❖ The tools for advanced multimedia search
 - ❖ Automatic indexing,
 - ❖ Locating content in an efficient,
 - ❖ Location of contents without prior manual annotation.

Audio / Speech



```
<meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1">
<meta name="description" content="Página Web desarrollada para la asignatura
Sistemas avanzados de recuperación de información, contiene información sobre
metadatos como XML, Dublin Core, RDF, etc">
<meta name="author" content="María José Martínez Pérez"/>
<meta name="Keywords" content="Sistemas avanzados de recuperación de
información, SARI, XML, Metadatos, etiquetas Meta, RDF, recuperación de
información, Dublin Core, Dublincore, DTD, SGML, HTML">
<meta http-equiv="content-type" content="text/html"/>
```

21 AL 24 DE ABRIL

Puerto Vallarta, Jal.

MEX-CULTURE

FONOTECA NACIONAL

PRESERVAMOS LA MEMORIA SONORA PARA EL FUTURO




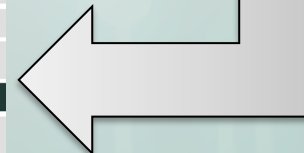
- ❖ 1. Digitalization 96 kHz, 14-bit WAV format.
- ❖ 2. Digitalization low resolution MP3 format.



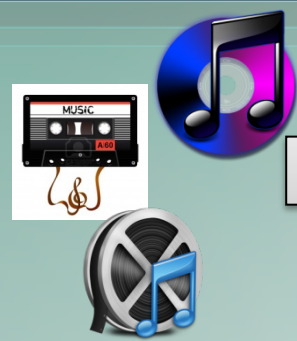
❖ Cataloging Manual



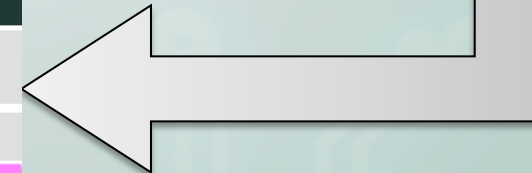
Ficha catalográfica del archivo sonoro	
Título del Record	[Ahí viene Francisco Villa. Las abandonadas]
Edición	Edición Columbia
Lugar	México Estado de México
Idioma	Español
Temas	General Música popular y tradicional mexicana
Notas	En el soporte: Discoteca XESM 71.
Id	FNM0012632
Ficha catalográfica track	
Título del Track	Ahí viene Francisco Villa 
Mención de Responsabilidad	Composición Reyna Cordero
Género	Corrido
Notas	Registrada como: Mex-3577-1. En el soporte esta pista corresponde a todo el lado.
Id	FN08030041921_01



❖ Automatic Indexing and Cataloging



Ficha catalográfica del archivo sonoro	
Título del Record	[Ahí viene Francisco Villa. Las abandonadas]
Edición	Edición Columbia
Lugar	México Estado de México
Idioma	Español
Temas	General Música popular y tradicional mexicana
Notas	En el soporte: Discoteca XESM /1.
Id	FNM0012632
Ficha catalográfica track	
Título del Track	Ahí viene Francisco Villa
Mención de Responsabilidad	Composición Reyna Caldero
Género	Corrido
Notas	Registrado como: Mex-3577-1. En el soporte esta pista corresponde a todo el lado.
Id	FNM0012632_01
Cantante(s)	1: Hombre Tiempo: 00:02 - 00:10, 00:14 - 00:25, 00:30 - 00:45 2: Hombre Tiempo: 00:08 - 00:10, 00:26 - 00:28, 00:43 - 00:50 1: Mujer Tiempo: 00:08 - 00:10, 00:38 - 00:47
Instrumentos	1: Guitarra Tiempo: 00:02 - 00:25, 00:40 - 00:55 2: Piano Tiempo: 00:08 - 00:26, 00:40 - 00:55 3: Trompeta Tiempo: 00:10 - 00:20, 00:30 - 00:47



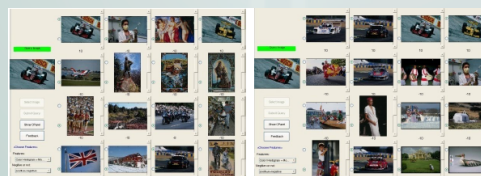
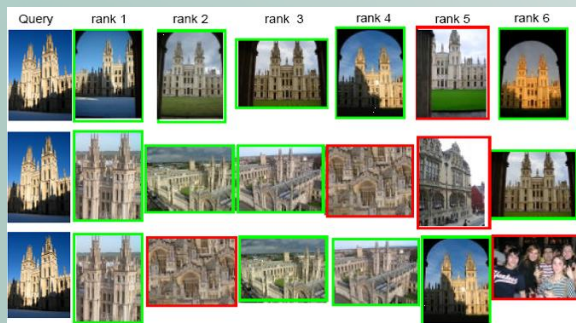
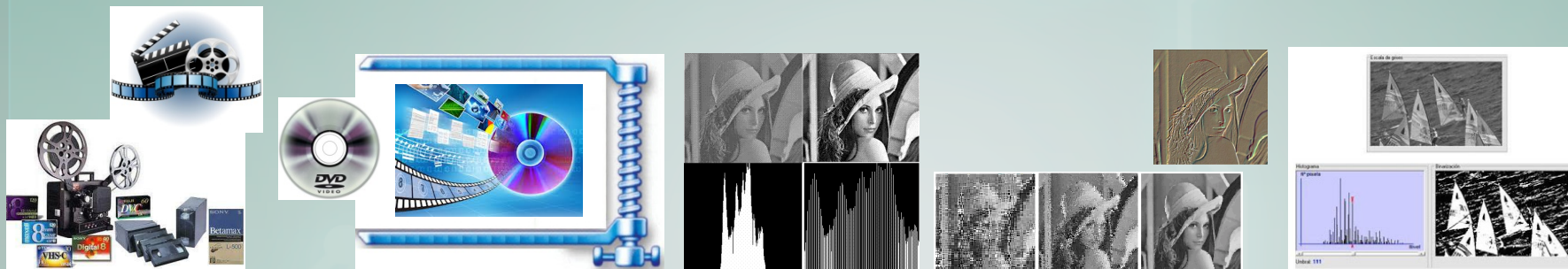


❖ The interest of the Fonoteca Nacional and DL-INAH in the MEX-CULTURE project in a timely manner lies in the field of digital and automatic indexing of audio documents

- ❖ Mexican indigenous languages indexing
- ❖ Indexing the sound of music library catalog
- ❖ Indexing voice and speech recognition Mexican personages

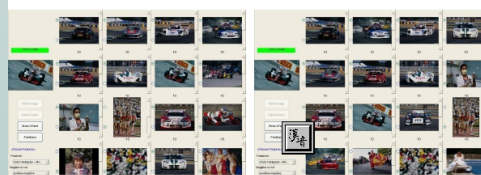


Image / Video



(a) The result of the initial query

(b) The result after the first round



(c) The result after the fifth round

(d) The result after the tenth round



```
<meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1">
<meta name="description" content="Página Web desarrollada para la asignatura
Sistemas avanzados de recuperación de información, contiene información sobre
metadatos como XML, Dublin Core, RDF, etc">
<meta name="author" content="Maria José Martínez Pérez"/>
<meta name="Keywords" content="Sistemas avanzados de recuperación de
información, SARI, XML, Metadatos, etiquetas Meta, RDF, recuperación de
información, Dublin Core, Dublincore, DTD, SGML, HTML"
<meta http-equiv="content-type" content="text/html"/>
```

21 AL 24 DE ABRIL

Puerto Vallarta, Jal.

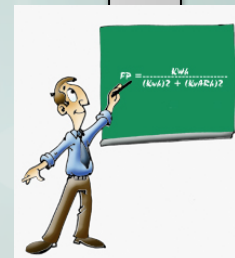
MEX-CULTURE

❖ Indexing and Cataloging Manual

ONCE
TVMÉXICO
INSTITUTO POLITÉCNICO NACIONAL



❖ Digitalization of visual content on Cassette ¾ to MPEG-2 format.



Ficha catalográfica del archivo video	
Título	[Bélgica vs Armenia]
Edición	Edición Columbia
Lugar	Bélgica Bélgica
Idioma	Español
Temas	General Futbol
Notas	En el soporte: Mediateca XESM 71
Id	ONCTVM0012632

Ficha catalográfica capítulo	
Título del capítulo	Bélgica vs Armenia
Mención de responsabilidad	Composición Reyna Cordero
Género	Deportivo
Notas	Registrada como: Mex-3577-1
Id	ONCTVM08030041921_01



❖ Automatic Indexing and Cataloging

MEX-CULTURE



Fecha catalográfica del archivo video	
Título	[Bélgica vs Armenia]
Edición	Edición Columbia
Lugar	Bélgica Bélgica
Idioma	Español
Temas	General Fútbol
Notas	En el soporte: Mediateca XESM 71
Id	ONCT/M0012_32
Fecha catalográfica capítulo	
Título del capítulo	Bélgica vs Armenia
Mención de responsabilidad	Composición Reyna Cordero
Género	Deportivo
Notas	Registrada como: Mex-3577-1
Id	ONCT/M0013004921_01
Inf A	
Inf B	
Inf C	
Inf D	
Inf E	
Inf F	





- ❖ **Beneficios** en diferentes aspectos tanto sociales, educativos y científicos, algunos de estos son:
- ❖ Ser la entidad de prueba de las herramientas de indexado del proyecto,
- ❖ Ser el modelo en la dirección del desarrollo de sistemas de indexación a gran escala.

MEX-CULTURE



El desarrollo de los métodos automáticos de vanguardia para el procesamiento e indexación a gran escala del contenido multimedia en las bases de datos,

- ❖ les ha permitido a patrocinadores de contenido tener el conocimiento sobre la inclusión de tecnologías de vanguardia apropiadas para el buen manejo de sus bases de datos,
- ❖ explotar estos conocimientos para la migración hacia su biblioteca multimedia digital de gran escala,

21 AL 24 DE ABRIL

Puerto Vallarta, Jal.

MEX-CULTURE

Thank You!

MEX - CULTURE

Multimedia libraries indexing for the preservation and dissemination of the Mexican Culture

