

Documentación del sistema de apoyo a la catalogación de archivos audiovisuales

Rodrigo Eduardo Colín Rivera

29 de mayo de 2015

Resumen

El Laboratorio Audiovisual de Investigación Social (LAIS) del Instituto Mora consta de una colección archivística para materiales audiovisuales cuyos registros actuales se encuentran en formato de hojas de cálculo. Un mejor tratamiento de esta información requiere el uso de bases de datos y una interfaz (preferentemente web) para realizar la consulta e inserción de los registros e información que describen dichos materiales con el objetivo de facilitar las tareas de catalogación dentro y fuera del laboratorio.

Se mencionará la forma de catalogación actual, la propuesta del sistema, los problemas y soluciones que brinda el sistema, la elección de las tecnologías involucradas en la implementación, el alcance o usos del sistema, los prototipos y las propuestas a considerar para el futuro desarrollo.

Antecedentes y problemática actual

El Laboratorio Audiovisual de Investigación Social (LAIS) del Instituto Mora consta de una colección archivística para materiales audiovisuales cuyos registros actuales se encuentran en formato de hojas de cálculo a través del programa Microsoft Excel ingresando diversos rubros que integran la ficha de documentación de cada material audiovisual con base en la adaptación de la norma ISAD(G) (*General International Standard Archival Description*), de la cual se mencionará a detalle más adelante.

La problemática principal con esta manera de catalogación es que resulta complicado buscar o filtrar los registros. Aunque este problema es la base de la propuesta del sistema, existen otros problemas relacionados tales como: La integridad de la información entre los libros de las hojas de cálculo, la cantidad de usuarios que pueden consultar estos registros y la manipulación de los datos.

Propuesta para el sistema de apoyo a la catalogación del LAIS

La propuesta principal está apoyada en la idea de crear una base de datos para el manejo actual y futuro de las fichas de documentación de los materiales

audiovisuales del LAIS.

Diferencias entre hojas de cálculo y bases de datos

Aunque ambas opciones tienen como objetivo almacenar conjuntos de datos, hay diferencias importantes a considerar y que se mencionan a continuación:

	Hoja de cálculo	Base de datos
Objetivo	Realizar cálculos numéricos a través del uso de diversas fórmulas aritméticas.	Almacenar datos en un servidor.
Almacenamiento	En archivos (libros) que contienen una o varias hojas de cálculo con filas y columnas para describir los datos.	En tablas que contienen registros (renglones) donde cada valor le corresponde un tipo de dato (columna)
Manejo	Cualquier persona con un conocimiento básico puede manipular, ordenar o filtrar datos.	Programador o persona encargada de la base de datos realiza consultas a través del lenguaje de programación SQL (<i>Structured Query Language</i>). Requiere una interfaz de usuario para una fácil manipulación de la base.
Complejidad (datos)	Los datos son sencillos de manejar pero a mayor cantidad, mayor complejidad.	El modelo relacional permite mantener unidades organizadas para los datos (tablas)
Repetición de datos	Se pueden presentar repetición o redundancia de datos.	Si el diseño de la base de datos es correcta, no se permite redundancia de información.
Modificaciones	Una sola persona puede modificar a la vez. Varias personas con copia del original pueden hacer modificaciones pero al combinar la información queda sujeta a errores humanos.	Los manejadores (programas) de bases de datos permite multi-usuarios y mecanismos de control en las modificaciones.
Presentación	Adecuado para incrustar en documentos escritos o mostrar gráficas relativas a los datos contenidos.	Solamente a través de una interfaz de usuario se pueden presentar los datos. El formato varía según la implementación.

Esquema general del sistema

Debido a que no la manipulación de la base de datos requiere de instrucciones en el lenguaje de consulta SQL, para que los usuarios puedan realizar las operaciones comunes como consultar, buscar, insertar y borrar registros, se debe crear una interfaz que facilite estas tareas.

La figura 1 muestra el diseño general separando la base de datos de la aplicación web. La aplicación web se comunica con la base de datos a partir de peticiones que devuelven los datos necesarios (en caso de consulta, búsqueda o inserción de datos) o la modificación correspondiente (en caso de modificar o eliminar datos).

También muestra que la base de datos está basada en las hojas de cálculo existentes, de esta manera no hay necesidad de reescribir toda la información, más bien es importada a la base de datos.

Finalmente se muestra las responsabilidades del programador para el diseño y creación de la base de datos y la interfaz web, mientras que los usuarios solamente requieren conocer el resultado de la interfaz web, es decir, una página en internet amigable y funcional para manipular la base de datos.

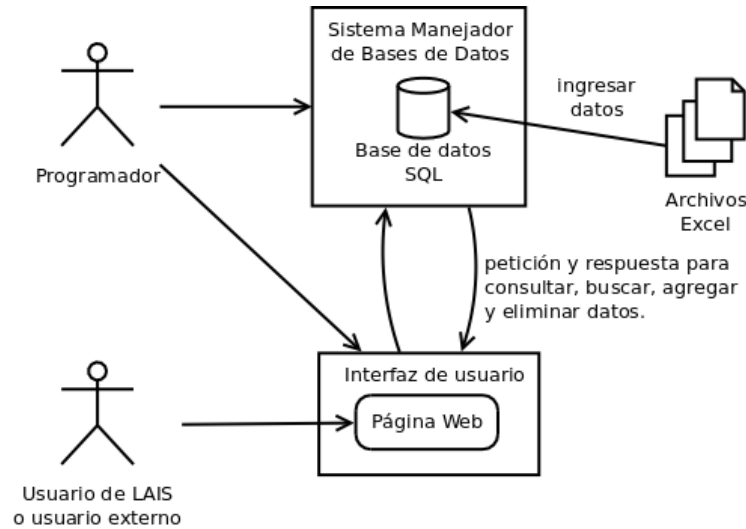


Figura 1: Esquema general del sistema de apoyo a la catalogación

Casos de uso

La figura 2 muestra el diagrama de casos de uso que representa a los usuarios y las acciones principales del sistema.

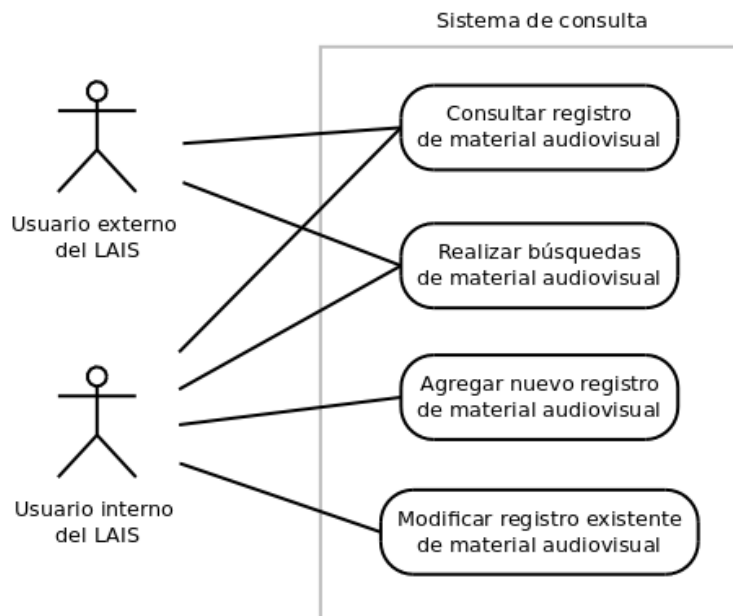


Figura 2: Diagrama de casos de uso

Selección de tecnologías

Existen diversas maneras de implementar el sistema y la selección de las tecnologías necesarias para su implementación es importante para que cumplan con los requisitos y se adapte adecuadamente a las necesidades del LAIS.

MySQL

MySQL es el Sistema Manejador de Bases de Datos Relacionales de código abierto (open source) más ampliamente utilizado y que actualmente es respaldada para su desarrollo por la corporación Oracle. MySQL es una elección común para uso de bases de datos para aplicaciones web, pues es un componente del conjunto de software denominado LAMP (acrónimo de "Linux, Apache, MySQL, Perl/PHP/Python"). Debido a que en el LAIS se cuenta con un servidor Apache en Linux, es conveniente que la elección para el manejo de la base de datos sea con MySQL.

MySQL Workbench

Este entorno de desarrollo integrado (IDE por sus siglas en inglés) permite el diseño, desarrollo y administración de manera visual de la base de datos. Especialmente útil para programadores ya que está disponible para todos los

sistemas operativos (Linux, Mac y Windows) y solo requiere hacer la conexión con el servidor para poder administrarla de manera remota.

PHP

Debido a que se requiere de una interfaz web para hacer uso de la base de datos, dicha interfaz requiere un lenguaje de programación que permita hacer desarrollo web del lado del servidor. PHP es un lenguaje ampliamente usado y probado.

En particular, el uso del complemento PDO (PHP Database Object) permite realizar operaciones comunes de la base de datos (como conectarse, agregar registros y realizar consultas) desde una página web utilizando la filosofía de la orientación a objetos, lo cual también facilita la programación de la interfaz o página web.

HTML5 y Bootstrap

Así como PHP es el intermediario con la base de datos, el intermediario que hace uso de esta tecnología y la pone a disposición del usuario es HTML. HTML es un lenguaje de marcado para crear páginas web que facilitan el uso de la base de datos desde un ambiente más amigable. El estándar actual recomendado se conoce como HTML5 y es el estándar con el que se desarrollará.

Bootstrap es un conjunto de herramientas para la creación de páginas web y aplicaciones web, principalmente usado para darle estilo a componentes como la tipografía, formularios, botones y otros componentes de la interfaz. Además permite diseño responsivo para que la interfaz web sea adecuada tanto en monitores grandes de computadoras de escritorio como para pantallas pequeñas de dispositivos móviles (smartphones y tablets).

Diseño de la base de datos

Las tablas y los campos o columnas que lo conforman están basadas en la descripción archivística existente en hojas de cálculo que hasta el momento han sido el modelo a seguir para la catalogación archivística.

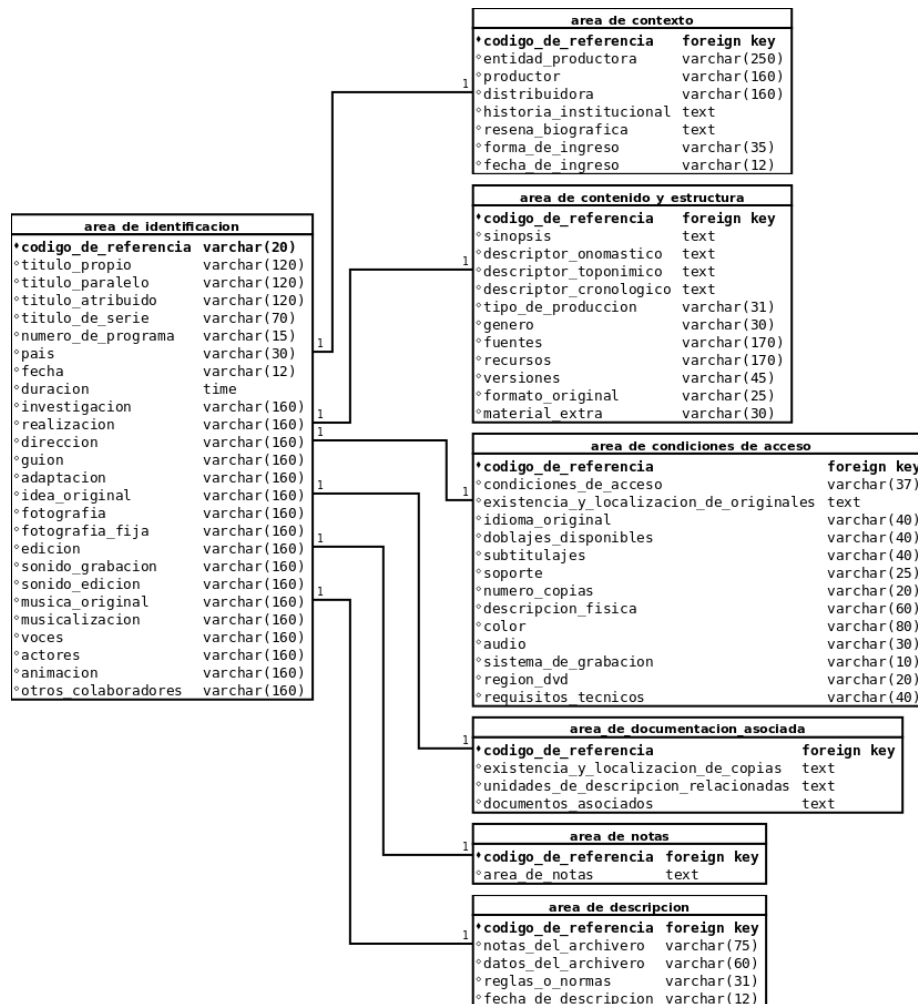


Figura 3: Diagrama Entidad-Relación que muestra el diseño de la base de datos

Explicación del diseño de la base de datos

Cada tabla muestra los campos que lo componen y el tipo de dato asociado, los tipos de datos que se emplean son los siguientes:

varchar

Representa una cadena de caracteres (texto) de longitud variable. Se agrega entre paréntesis la longitud máxima que puede tener el texto. Este es el tipo de dato más común en la base de datos.

text

Representa una cadena de caracteres (texto) de longitud muy extensa. Es útil para descripciones como la *historia institucional* o la *reseña biográfica*. Aunque ofrece una longitud teóricamente ilimitada de texto, las consultas o búsquedas en este tipo de datos son más lentas que en **varchar**.

time

Representa tiempo y tiene dos utilidades: representar una hora del día (por ejemplo, las 2:15 pm) y representar un lapso de tiempo (por ejemplo, 2 horas y 15 minutos). Nos interesa el último uso para la representación de *duración*.

date

Representa fechas y son especialmente útiles para campos como *fecha de ingreso* o *fecha de descripción*.

Cada tabla contiene el campo **código de referencia** porque todas las tablas hacen referencia a un archivo audiovisual que debe aparecer una sola vez en cada tabla, este tipo de relación es llamada “1 a 1” y es por eso que en la figura 3 se muestran números ‘1’ al inicio y fin de cada línea. También el **código de referencia** es el único elemento obligatorio para todas las tablas, el resto de los campos o columnas de las tablas son opcionales (es decir, pueden quedar en blanco).

La descripción que existe en los archivos Excel repiten el campo de *título propio* para cada hoja. Este detalle no se permite en la base de datos porque crearía redundancia de información (repetición de datos).

Prototipos de la interfaz web

Los prototipos o ejemplos de prueba son un diseño inicial para mostrar las características y funcionalidad principalmente de la interfaz web.

La figura 4 muestra un ejemplo de conexión y consulta exitoso entre la página web y la base de datos. La petición resulta en obtener ciertos datos y mostrarlos en una tabla a modo de ejemplo. El detalle importante dentro del código fuente de la página web de este prototipo es que rescata valores de tres diferentes tablas con PHP y las integra en una tabla preformateada con Bootstrap.

Consulta LAIS

localhost/mysql/select3.php

Pruebas de conexión Mediante HTML5, PHP y Bootstrap

Hola mundo desde el [LAIS](#).

Esta es una página de prueba que consulta la base de datos creada en MySQL con la información de los registros audiovisuales del [LAIS](#).

Tabla de muestra

Esta tabla realiza una petición mediante PHP al servidor (que contiene el manejador de bases de datos MySQL) para realizar una consulta. La consulta requiere información de varias tablas dentro de la base de datos, las tablas involucradas son "area_de_identificacion", "area_de_contexto" y "area_de_contenido_y_estructura".

Código de referencia	Título propio	Productor	Tipo de producción
MXIM-AV-1-4-1	Chelovek kinoapparatom	Gobierno ucraniano de la Unión Soviética	Película documental
MXIM-AV-1-4-2	Drifters		Película documental
MXIM-AV-1-4-3	Kino-Glaz	Gobierno ucraniano de la Unión Soviética	Película documental
MXIM-AV-1-4-4	Nanook of the north	Divisa Home Video	Película documental
MXIM-AV-1-4-5	Nogent, eldorado du diamanche		Película documental
MXIM-AV-1-4-6	Regen		Película documental

Esta página es únicamente de muestra y contiene un posible uso para la consulta de información.

Figura 4: Prueba de conexión y consulta con la base de datos

La figura 5 muestra un formulario básico de consulta que permite hacer búsquedas por título y filtrar la información importante que se desea ver. La figura 6 muestra mas elementos de la consulta realizada en formato de lista, y muestra que es posible manipular los datos para darles el formato que queramos.

Consulta base de datos: x Perfil predete...

localhost/mysql/select4.php?titulo=&codigoReferencia=on&tituloPropio=or

Pruebas de conexión Mediante HTML5, PHP y Bootstrap

Hola mundo desde el LAIS.

Título:

Agregar texto si se desea buscar por título

☒ Código de referencia
☒ Título
☒ País
☒ Fecha

Consultar

```
SELECT codigo_de_referencia, titulo_propio, pais, fecha FROM area_de_identificacion
```

Resultados

Título de ejemplo Código de referencia: MXIM-AV-1-10-28 País: México Fecha: 2015
Otro ejemplo Código de referencia: MXIM-AV-1-10-29 País: México

Figura 5: Prueba de búsqueda selectiva que muestra un formulario básico de consulta



Otro ejemplo	Código de referencia: MXIM-AV-1-10-29	País: México	Fecha: 2015
Chelovek kinoapparatom	Código de referencia: MXIM-AV-1-4-1	País: Unión Soviética	Fecha: 1929
Drifters	Código de referencia: MXIM-AV-1-4-2	País: Reino Unido	Fecha: 1929
Kino-Glaz	Código de referencia: MXIM-AV-1-4-3	País: Unión Soviética	Fecha: 1924
Nanook of the north	Código de referencia: MXIM-AV-1-4-4	País: Estados Unidos	Fecha: 1922
Nogent, eldorado du diamanche	Código de referencia: MXIM-AV-1-4-5	País: Francia	Fecha: 1929

Figura 6: Algunos ejemplares en un formato de prueba para mostrar resultados de búsqueda selectiva

Análisis sobre la base de datos

Futuro desarrollo

Se requiere de la supervisión del equipo del LAIS para aprobar el diseño de la base de datos y crear nuevos prototipos para la interfaz web que sean adecuados para las necesidades del LAIS.

En la programación se requiere crear una clase para definir un recurso audiovisual en lenguaje de programación PHP e investigar la creación dinámica de páginas que representen a cada recurso audiovisual de manera individual con toda la información posible en base a las fototecas digitales del proyecto Huellas de Luz (<http://lais.mora.edu.mx/huellasdeluz/>).

Se debe organizar el proyecto para subirlo a un repositorio Git para tener un manejo adecuado de versiones y el desarrollo esté organizado y sea eficiente de manejar.

Problemas

Listado de algunos problemas que se han encontrado durante el desarrollo y pruebas:

Longitud de texto

Se debe revisar la longitud de las cadenas de texto para los campos de la base de datos (tipos de dato **varchar**). Muchos de los campos son para describir nombres de personas y tienen longitud máxima de 70 caracteres¹.

Fechas

El campo *fecha* tiene el problema de que permite almacenar un periodo de tiempo en formato [YYYY-YYYY], sin embargo, para la base de datos no hay tipo específico para determinar rangos o periodos de tiempo.

Existen dos posibles soluciones: Dejar el tipo de dato para *fecha* como *varchar(11)* (es decir, como texto) pero las búsquedas por lapsos de tiempo serán complicadas y posiblemente inexactas. Otra alternativa es separar el campo *fecha* en *fecha_inicio* y *fecha_fin*, pero requiere el preprocesamiento de la información (requiere programación) y un cambio en el diseño de la base de datos; la ventaja radica en que las búsquedas por lapsos de tiempo serán viables.

Duración

El campo *duración* de la base de datos requiere que el formato sea el siguiente HH:MM:SS o HHMMSS. Sin embargo hay restricciones en cuanto a los números válidos para los minutos y segundos porque no pueden rebasar la cifra 60. Sin embargo, hay registros en las hojas de cálculo con valores como 66:06 que significa duración de 66 minutos y 6 segundos, pero la base de datos espera que se encuentren en el formato 01:06:05. El problema se puede resolver con preprocesamiento de la información (requiere programación).

Datos nulos o cadenas vacías

En la base de datos, el valor por default es la constante NULL y representa un elemento vacío. Sin embargo, en las hojas de cálculo el valor por default es la cadena vacía. Hay que determinar cuál será el elemento por default en el sistema.

Notas

La base de datos y los prototipos aun no están en línea, están en desarrollo y por el momento solo se pueden ejecutar en las máquinas del LAIS con la restricción de que el equipo de cómputo de Rodrigo (programador del LAIS) se encuentre prendido para hacer la función de servidor. Próximamente, cuando

¹UK Government Data Standards Catalogue http://www.datadictionary.nhs.uk/data_dictionary/data_field_notes/p/pers/person_full_name_de.asp?shownav=1

el sistema se encuentre en una etapa más avanzada de desarrollo, se subirá al servidor del LAIS.