

CUDI - Corporación Universitaria para el Desarrollo de Internet - Reporte de avances

Proyecto: “*Digitalización y Divulgación Digital de Acervos Antiguos*”

En este proyecto se plantea como objetivo la creación de una biblioteca digital constituida por acervos de fondo antiguo pertenecientes a instituciones miembros de CUDI, con la finalidad de preservar la riqueza cultural, histórica y artística contenida en dichos acervos, así como ponerla al alcance de investigadores y público en general interesado en este tipo de documentos para su aprovechamiento.

Las instituciones miembros de CUDI que se encuentran actualmente participando en este proyecto son:

- Biblioteca Franciscana (UDLA-P)
- Biblioteca José María Lafragua (BUAP)
- Biblioteca UDLA-P (Sala de Colecciones Especiales)

El proyecto está dividido en tres fases, estableciéndose en cada una de ellas los objetivos que a continuación se describen así como los avances hasta el momento:

- Fase I: Construcción de acervo digital piloto
- Fase II: Construcción de acervos adicionales
- Fase III: Integración del Acervo Digital CUDI de Libros Antiguos

Fase I

El acervo piloto corresponde a la Biblioteca Franciscana, siendo el objetivo la digitalización de los 100 impresos mexicanos más importantes de esta institución. Tanto en la Fase I como en la Fase II se distinguen 5 actividades principales:

Selección:

Proceso que ya ha sido realizado por parte de la Biblioteca Franciscana.

Digitalización:

Actividad que continúa en desarrollo, teniendo 62 libros totalmente digitalizados de los 100 seleccionados de la Biblioteca Franciscana (Figura 1).

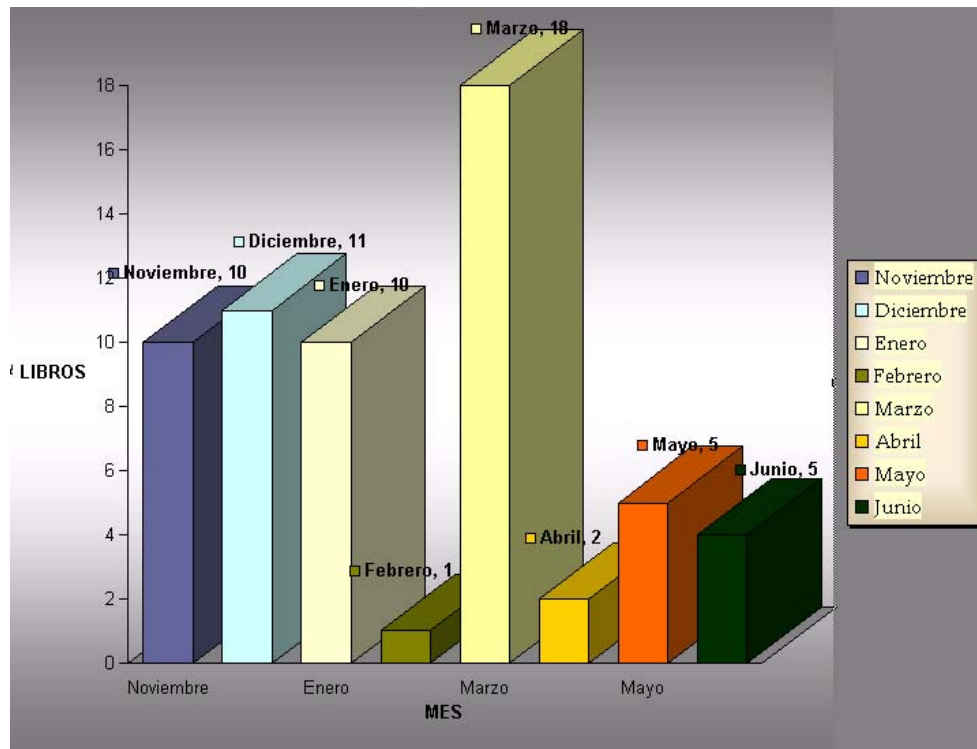


Figura 1. Gráfica de avances por mes del proceso de digitalización

- Reconocimiento: en este punto se propuso la integración de un componente para el reconocimiento de caracteres antiguos denominado ANTIQUE, Que es parte de investigaciones desarrolladas por el grupo de reconocimiento de patrones y redes neuronales artificiales de la Universidad de las Américas-Puebla que trabaja conjuntamente con los investigadores del presente proyecto y que incluye:
 - El trabajo de Segmentación de Caracteres
 - Reconocimiento de caracteres previamente segmentados
 - Identificación de palabras
 - Corrección de Errores
 - Entrenamiento en base a las editoriales

Actualización:

Previo a la actualización en la base de datos, el material digitalizado pasa por un proceso de edición y renombramiento de cada página para su correcta visualización. El estado actual en que se encuentran los libros ya digitalizados se muestra en la Figura 2. Finalmente el almacenamiento del material digitalizado, revisado y editado es almacenado en la base de datos a través del sistema CIText.

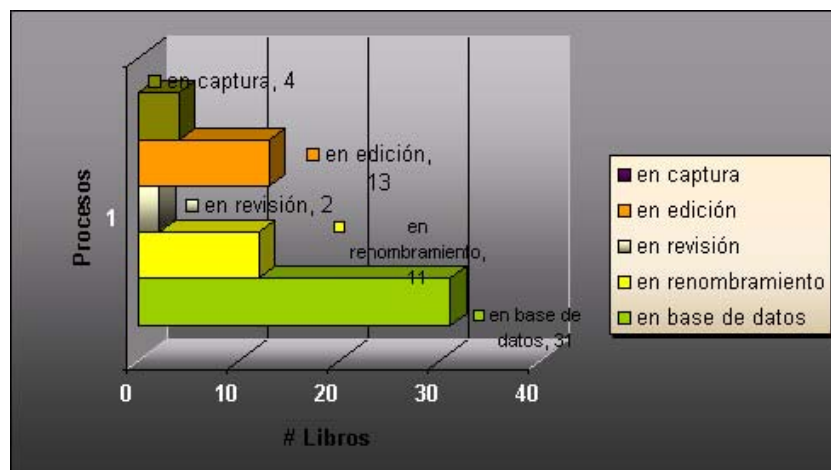


Figura 2. Gráfica de estado del material ya digitalizado

Fase II

En la Fase II se estableció como objetivo la construcción de acervos adicionales, la continuación en el proceso de digitalización de los libros de la Biblioteca Franciscana y la integración del servidor de modelos de búsqueda (Hermes) y del protocolo OAI. En este sentido, se ha iniciado ya la construcción del acervo digital de la Biblioteca José María Lafragua (BUAP), constituido inicialmente por 20 obras. Hasta el momento se cuenta con 1 libro digitalizado disponible vía web, sin embargo, para mejorar el rendimiento del mismo ésta institución se encuentra en procesos de migración de sus aplicaciones a un servidor UNIX, dado que inicialmente se encontraban trabajando en plataforma PC con LINUX Susie.

En cuanto a la digitalización de los 38 libros restantes de la Biblioteca Franciscana, se continúa con el proceso, esperando concluirlo en el mes de Octubre de este año.

Finalmente, para la integración del servidor Hermes con el sistema CIText, es necesario realizar la instalación y configuración del servidor así como las pruebas de integración finales, aspecto sobre el cual se estará trabajando en los próximos meses.

Fase III

Integración de protocolos y aplicaciones:

Para la interoperabilidad de las bibliotecas participantes se estableció la utilización de protocolos OAI (<http://www.openarchives.org/index.html>), aspecto que está siendo definido actualmente. A su vez, el proceso de integración del servidor de modelos de recuperación de información (Hermes) está en desarrollo.

Finalmente cabe agregar que para mejorar el acceso y la consulta a los acervos de las bibliotecas participantes, se están realizando actualizaciones de las páginas de acceso, así como del software de búsqueda y visualización de la información utilizando como plataforma de desarrollo php.