

Agentes y Movilidad en Colecciones Multimediales Heterogéneas

Proyecto apoyado por la Corporación Universitaria para el Desarrollo de Internet (CUDI)

Investigador responsable: Dr. Alfredo Sánchez H.
Universidad de las Américas–Puebla (UDLA)

Investigador co-responsable: Dr. David Garza
Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM)

Resumen

Este proyecto ha tenido como objetivo principal la creación de una infraestructura que aprovecha las características de Internet 2 para proveer servicios de recuperación de información y permitir el acceso uniforme a colecciones heterogéneas a través de dispositivos convencionales y móviles.

El contenido de este informe final se ha organizado en los siguientes apartados que describen los resultados del proyecto: (1) el acceso a colecciones digitales a través de Internet 2 y dispositivos móviles, (2) servicios de recolección y manejo de metadatos implementados, (3) herramientas de apoyo desarrolladas, y (4) documentos generados (artículos, tesis, presentaciones). Se incluyen también referencias específicas a los servicios disponibles como resultado del proyecto, así como un directorio de los participantes.

El proyecto se ha desarrollado en forma conjunta por la Universidad de las Américas, Puebla (UDLA) y el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) campus Monterrey, con el apoyo de la Corporación Universitaria para el desarrollo de Internet (CUDI).

Introducción

Un reto importante del proyecto **Agentes y Movilidad en Colecciones Multimediales Heterogéneas** consistió en integrar dos arquitecturas de bibliotecas digitales con varios años de desarrollo a través de Internet 2,

desarrollando nuevos servicios de consulta e integrando el uso de dispositivos móviles. La infraestructura desarrollada se encuentra operando de manera estable y permite acceder a colecciones digitales utilizando el protocolo para la recolección de metadatos OAI-PMH propuesta por la Iniciativa de Archivos Abiertos (OAI, por sus siglas en inglés). Una vez localizados los documentos de interés por parte del usuario, ya sea utilizando un navegador o dispositivo móvil, es posible acceder a su contenido. En el caso de los dispositivos móviles que presentan restricciones en espacio, memoria y tiempo de conexión, se proponen alternativas como la de enviar el contenido en diferentes formatos a una cuenta de correo electrónico. Es posible además realizar búsquedas en colecciones no disponibles al momento de la consulta, este procedimiento se realiza sin intervención del usuario y provee los resultados al momento de que el usuario regresa al sistema.

El proyecto contempla la incorporación de nuevas colecciones y servicios adicionales a los de las instituciones participantes. Esto ha motivado el desarrollo de una herramienta que permite generalizar y automatizar la construcción de servidores de metadatos bajo el estándar OAI-PMH. Una vez implementado un nuevo servidor de metadatos, se puede generar una representación para navegadores convencionales o dispositivos móviles usando el explorador de metadatos desarrollado en el proyecto, o incorporarse a una colección en un servicio personal. Como parte de la arquitectura desarrollada para el proyecto, se incorporan servicios de recuperación basados en agentes ontológicos que consideran las características semánticas de los datos además de la diversidad de los formatos.

El proyecto **Agentes y Movilidad en Colecciones Multimediales Heterogéneas** enfatiza la implementación de estándares. El uso de protocolo OAI-PMH (basado en XML) a nivel de servicios permite la consulta de metadatos de las colecciones y al mismo tiempo garantiza la interoperabilidad. Para comunicación de aplicaciones se utiliza XML-RPC y HTTP. A nivel de interfaz se promueve uso de XHTML para navegadores y WML para dispositivos móviles. En el caso de aplicaciones funcionando en dispositivos móviles se utiliza J2ME.

Los resultados del proyecto (disponibles al público en <http://ict.udlap.mx/cudi/udlatec>) incluyen el acceso a colecciones digitales, servicios de recolección de metadatos, herramientas para generación de interfaces móviles y servidores de OAI, así como diversas publicaciones.

La Figura 1 ilustra los componentes principales de la infraestructura desarrollada para el presente proyecto. Las secciones subsiguientes describen de manera sintética estos componentes. Mayores detalles se encuentran disponibles en el sitio de web del proyecto (<http://ict.udlap.mx/cudi/udlatec>).

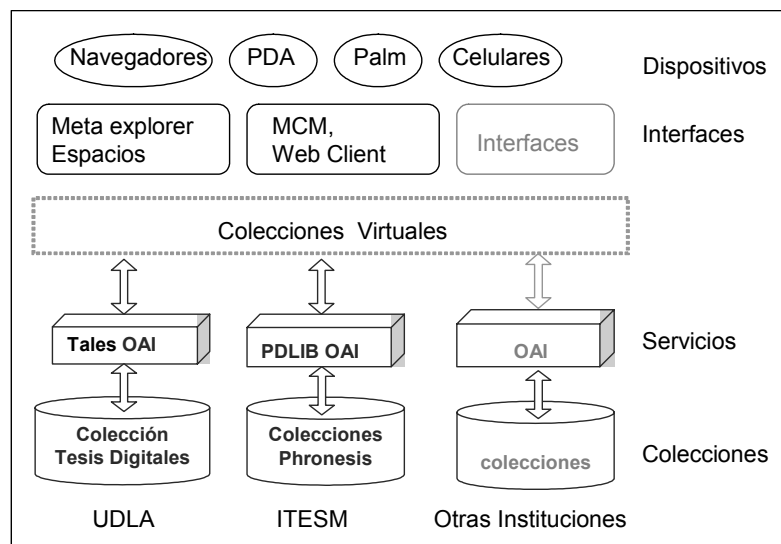


Figura 1. Infraestructura del Proyecto

1. Acceso a colecciones

Como parte del proyecto se implementaron interfaces de acceso, navegación y consulta. A continuación se presenta cada una de las interfaces. Las colecciones se describen en un anexo al final del reporte.

UDLA

Espacios PDA. Componente desarrollado en Java usando Java Server Pages, que permite consultar la colección de Tesis Digitales UDLA generando interfaces de acuerdo al dispositivo. Para acceso vía web y PDA se genera XHTML; para el acceso vía teléfono celular se genera WML. La única restricción de uso es contar con un navegador adecuado.

META-Explorador. A través de este componente es posible navegar la colección de un servidor OAI-PMH con base en sus metadatos. Permite el acceso transparente a colecciones como las de tesis digitales de la UDLA

y de ITESM. Desarrollada en Java, usando JDOM, para acceso vía web y PDA utiliza XHTML, para acceso vía teléfono celular utiliza WML.

ONTO-Explorador. Aplicación en desarrollo que permite la navegación “ontológica” a colecciones con servidores OAI-PMH. Desarrollada en Java, usando JDOM. Para acceso vía web y PDA utiliza XHTML, para acceso vía Celular utiliza WML.

ITESM

WebClient. Aplicación desarrollada en Java usando Servlets, que permite consultar las siguientes colecciones: Tesis del programa de graduados de la DECIC - ITESM, Tesis digitales de la UDLA, y las definidas por el usuario. Esta interfaz hace uso de los metadatos adquiridos a través de un servidor OAI-PMH. No tiene restricciones de uso y puede ser consultado en cualquier navegador.

PDLIB Trader. Aplicación desarrollada en J2ME – CLDC (MIDP) para Asistentes Personales Digitales (o PDAs, como Tugsten C, Ipaq 4150) . Con PDLIB Trader es posible consultar las siguientes colecciones: Tesis del programa de graduados de la DECIC - ITESM, Tesis digitales UDLAP, y las definidas por el usuario. Requiere de una conexión a Internet en el dispositivo móvil. Su instalación y uso es restringido.

2. Servicios

En este apartado se presentan servicios que además de ser utilizados para facilitar el acceso a colecciones de las instituciones participantes, también están disponibles públicamente. Estos servicios implementan el protocolo OAI-PMH lo que garantiza la interoperabilidad de las colecciones.

UDLA

El servicio de recolección de metadatos OAI-PMH 2.0 de la UDLA es una aplicación desarrollada en Java usando Servlets (Tomcat 5.0). Este servicio proporciona los metadatos de la colección de Tesis Digitales UDLAP. El servicio de metadatos implementa los estándares Dublin Core y OAI-PMH. Actualmente se encuentra instalado como un servidor

en una computadora UNIX (Sunfire 480R) con sistema operativo Solaris 9, 8GB RAM, 2 discos de 73GB.

ITESM

El servicio de recolección de metadatos OAI-PMH 2.0 en el ITESM es una aplicación desarrollada en Java y Perl, que proporciona los metadatos de la colección de Tesis del programa de graduados de la DECIC - ITESM. El servicio de metadatos implementa los estándares: XML-RPC, Dublin Core, OAI-PMH. Opera como un servidor en un equipo Windows 2003 Server, 2 procesadores Xeon 3.02 Ghz, 4 Gb en RAM y 750 Gb en disco duro.

3. Herramientas

Otro conjunto importante de resultados lo constituyen herramientas que se han puesto en el dominio público para facilitar la construcción de interfaces y servicios para otras organizaciones.

UDLA

EditMos es un ambiente gráfico desarrollado en Java 2 cuyo fin es permitir la conversión de interfaces y documentos en HTML a estándares como WML y XHTML utilizando hojas de estilo (XSL). Con EditMos se busca facilitar al proveedor de servicios la presentación de sus documentos e interfaces actuales para los usuarios con dispositivos móviles.

Las hojas de estilo utilizadas en EditMos forman parte del proyecto. Se cuenta en línea un tutorial que explica la conversión de XML a WML y XHTML usando XSLT.

Voai es una aplicación Web desarrollada en Java utilizando Servlets, que permite generalizar y automatizar la construcción de servidores OAI. Si se tiene una base de datos relacional y se desea un servidor OAI para compartir los metadatos de esa base de datos, se puede utilizar el generador de servidores *Voai* para crear automáticamente un servidor OAI. Para esto es necesario solamente especificar la estructura de la base de datos y algunas consultas en SQL que permitirán recuperar los metadatos a compartir. Una vez definidas estas opciones, se generará código Java listo para ser colocado en un contenedor de servlets. Este

código es el servidor OAI que está listo para correr en un servidor Web.
La Figura 2 ilustra de manera abreviada este proceso.

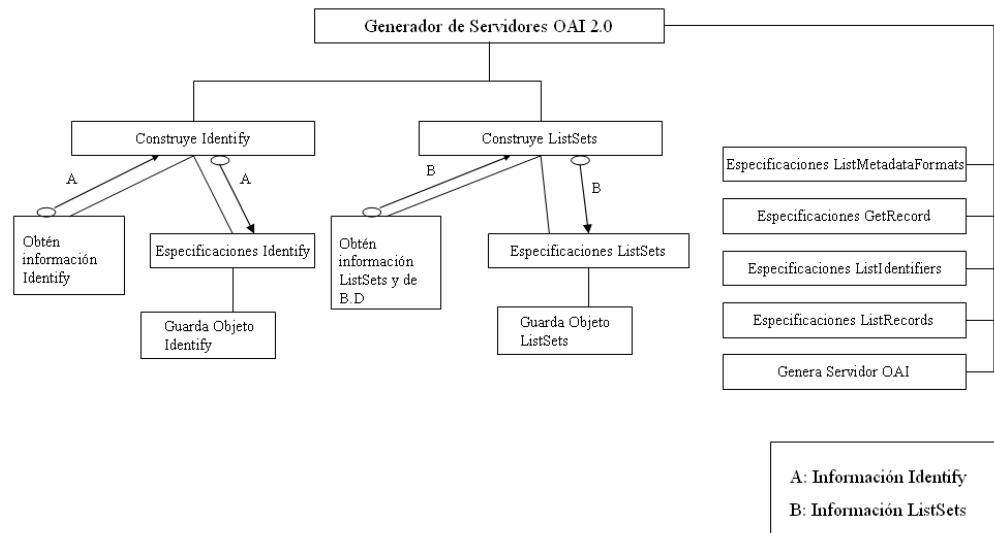


Figura 2. Configuraciones de VOAI

Valida-OAI. Aplicación en desarrollo que permite validar las respuestas de un servidor OAI-PMH en base a sus esquemas (*Schemas*). Desarrollada en Java usando JDOM, esta aplicación realiza pruebas de consultas con todos los verbos y opciones del protocolo y valida las respuestas del servidor con respecto a su esquema. Este enfoque permite encontrar diferencias de las implementaciones con la definición del protocolo.

ITESM

PDLib DataServer. El servidor de datos del PDLib proporciona servicios de almacenamiento y recuperación del contenido de las bibliotecas digitales personales. Se ha creado una implementación del servidor de datos en la plataforma J2SE. Para el almacenamiento físico de los documentos y las colecciones de las bibliotecas personales, se emplea el manejador de base de datos MySQL. La implementación actual utiliza como motor de búsqueda de texto completo a los índices del manejador de base de datos MySQL pero se espera reemplazar a este motor de búsqueda con el motor de búsqueda Lucene del proyecto Jakarta del grupo Apache, el servidor de datos cuenta con facilidades para

comunicarse con otros servidores utilizando protocolos como OAI (Open Archives Initiative) y estándares de metadatos como Dublin Core.

Mobile Connection Middleware. Se cuenta con una implementación del MCM realizada sobre la plataforma J2SE. Actualmente, los servicios del MCM son accedidos por medio del protocolo XML-RPC implementado mediante las librerías XML-RPC del proyecto <Web Services /> del grupo Apache. La figura 3 muestra los roles de las distintas capas de la arquitectura

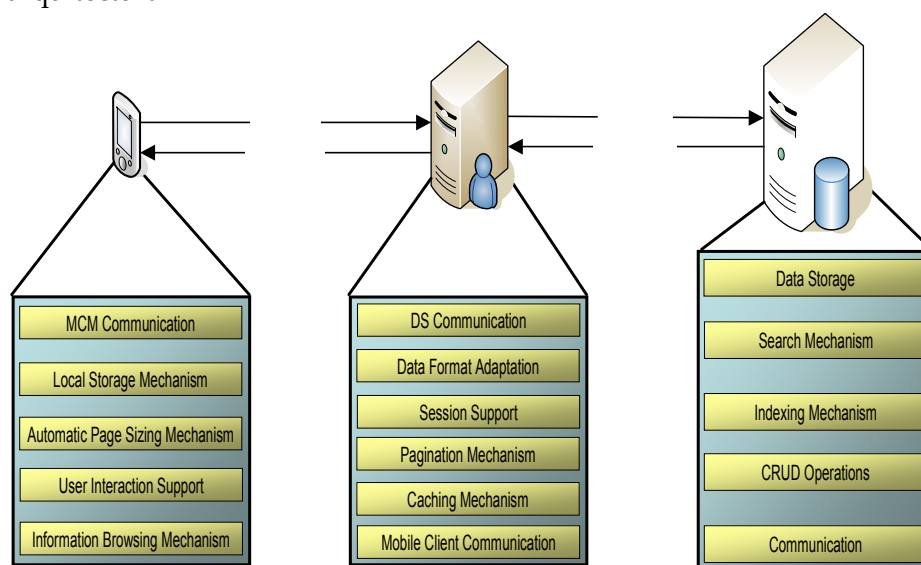


Figura 3. Detalle de los componentes de la arquitectura de PDLib

Los servicios desarrollados en conjunto (Intercambiador, MCM y Servidor de datos) son:

- Descarga total y parcial de documentos digitales en los dispositivos clientes.
- Navegación en bibliotecas que brindan el soporte a protocolos OAI (Open Archives Initiative).
- Adaptación automática a la red.

4. Publicaciones

Artículos

- Medina Nieto, M. A. Sánchez, J. A., Chávez, A., Benítez, A. 2004. Designing ontological agents: An alternative to improve information retrieval in federated digital libraries. In *Advances in Web Intelligence: Proceedings of the Second International Atlantic Web Intelligence Conference (AWIC 2004, Cancun, Mexico, May 16-19)*. Lecture Notes in Computer Science 3034, Springer-Verlag, 155-163. ISBN 3-540-22009-7
- Sánchez J.A., An Exploration of Network Technologies for Mobile Data Access in Digital Libraries, *WSEAS Transactions on Communications* (Issue 1, Vol. 3, January 2004, -ISSN 1109-2742, pp. 104-109 (publicado simultáneamente en *Proceedings del 6th WSEAS International Conference on Telecommunications and Informatics (TELE-INFO'04, ISBN 960-8052-98-X)* pp. 485-213: 1-6.
- Castellanos, N., Sánchez, J. A. 2004. Mobile access to digital libraries via generic interfaces. *Avances en Ciencias de la Computación (Taller de Interacción Humano-Computadora en el ENC 2004, Colima, Mexico)*. 571-576. ISBN 970-692-170-2.
- Álvarez F., Lavariega J.C. 2004 A 3-tiered client-server distributed database system component-based *Proceedings of the Second winter international symposium on Information and communication technologies. ACM International Conference Proceeding Series*, pages 501–506, Cancun, Mexico, January 2004. Workshop Session: Distributed Objects Research, Experiences and Applications, Trinity College Dublin. ISBN 0954414535.2004
- Álvarez F., Lavariega J.C. 2004 Una Arquitectura Basada en Componentes de Software de un Sistema de Bases de Datos Distribuidas con una Estructura Cliente-Servidor de 3 Capas *Memorias del XXXIV Congreso de Investigación y Extensión del Sistema Tecnológico de Monterrey*. Febrero 2004

Reportes

- Sánchez J.A., Garza D. 2004. Agentes y movilidad en colecciones multimediales heterogéneas- Reporte de Avance Técnico, Corporación Universitaria para el desarrollo de Internet A.C. (CUDI)
- Álvarez F. 2004 A Universally Available Peer-to-Peer Personal Digital Library System. Reporte interno. Centro de Investigación en Informática (CII). Instituto Tecnológico y de Estudios Superiores de Monterrey. Diciembre
- Álvarez F. 2004. PDLib: Universally Available Personal Digital Libraries Reporte interno. Centro de Investigación en Informática (CII). Instituto Tecnológico y de Estudios Superiores de Monterrey. Diciembre
- Guevara M. 2004. PDLib Web Front-End. Reporte interno. Centro de Investigación en Informática (CII). Instituto Tecnológico y de Estudios Superiores de Monterrey. Diciembre
- Guevara M. 2004 Herramientas Convertidoras de Formato para su uso con Lucene. Centro de Investigación en Informática (CII). Instituto Tecnológico y de Estudios Superiores de Monterrey. Diciembre
- García R., Garza D. 2004 Técnicas para Acceso a Bibliotecas Digitales usando Dispositivos Móviles [Adaptación Automática a la Conexión] Reporte interno. Centro de Investigación en Informática (CII). Instituto Tecnológico y de Estudios Superiores de Monterrey. Diciembre
- García R., Garza D. 2004 J2ME Developers Roadmap. Reporte interno. Centro de Investigación en Informática (CII). Instituto Tecnológico y de Estudios Superiores de Monterrey. Diciembre

Presentaciones

- Castellanos, N., Espacios personales genéricos en el contexto del Proyecto: "Agentes y movilidad en colecciones multimediales heterogéneas". Congreso Amigos 2004: Consorcios en la Sociedad del Conocimiento, Universidad de las Américas, Puebla, Feb. 2004.
- Fernández, L. Proyecto: Agentes y movilidad en colecciones multimediales heterogéneas, Reunión de Primavera de CUDI, Manzanillo, Colima, Abril 2004.

Sánchez, J. A., Lavariega, J. C. Bibliotecas Digitales e Internet 2. Reunión de Otoño de CUDI, Acapulco, Guerrero, Octubre 2004.

Tutoriales

Razo A. 2004 Tecnologías XML. Disponible en línea:
<http://ict.udlap.mx/cudi/udlatec>

Castellanos N. 2004 Conversión de XML a WML y XHTML usando XSLT. Disponible en línea: <http://ict.udlap.mx/cudi/udlatec>

Tesis

González M. 2004 Interfaz gráfica para conversión de interfaces para dispositivos móviles. Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales. Escuela de Ingeniería, Universidad de las Américas Puebla. Diciembre.

García R. 2004 Técnicas de Adaptación a la Conexión para Clientes Móviles que accesan Servicios de Biblioteca Digital. Tesis de Licenciatura. Instituto Tecnológico y de Estudios Superiores de Monterrey. Diciembre

Tesis en desarrollo

Villegas A. 2005 Servicios de interoperabilidad OAI. Tesis de Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales. Escuela de Ingeniería, Universidad de las Américas Puebla. Mayo.

Hurtado L. 2004. Modelo de datos y repositorios de información para bibliotecas digitales en ambientes móviles. Tesis de Licenciatura. Instituto Tecnológico y de Estudios Superiores de Monterrey. Diciembre

Álvarez F. 2004 Distribución de Datos para Bases de Datos Distribuidas, Una Arquitectura Basada en Componentes de Software. Tesis de Doctorado. Instituto Tecnológico y de Estudios Superiores de Monterrey. Diciembre

Medina M.A. 2005 Interoperabilidad de Bibliotecas Digitales basada en Agentes Ontológicos. Tesis de Doctorado. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas

Computacionales. Escuela de Ingeniería, Universidad de las Américas Puebla. Mayo.

5. Referencias

- Corporación Universitaria para el desarrollo de Internet A.C. (CUDI)
<http://www.cudi.edu.mx>
- Bibliotecas de la Universidad de las Américas, Puebla
<http://biblio.udlap.mx>
- Universidad de las Américas, Puebla
<http://www.udlap.mx>
- Laboratorio de Tecnologías Interactivas y Cooperativas ICT
<http://ict.udlap.mx>
- Open Archives Initiative
<http://www.openarchives.org>
- Espacios PDA** *<http://ict.udlap.mx/cudi/espacios>* responsable Antonio Felipe Razo Rodríguez *anrazo@mail.udlap.mx*
- META-Explorador** *<http://ict.udlap.mx/cudi/udlatec>* responsable Antonio Felipe Razo Rodríguez *anrazo@mail.udlap.mx*
- ONTO-Explorador** *<http://ict.udlap.mx/cudi/udlatec>* responsable María Auxilio Medina Nieto *sp098988@mail.udlap.mx*
- WebClient** *<http://newton.mty.itesm.mx:8080/pdlib/>* responsable Marcos Eliud Guevara Camacho *A00611038@itesm.mx*
- Servidor de Metadatos UDLA** *http://ict.udlap.mx/Tales/Oai_tesis*
responsable Isaac Abraham Villegas Ortega *shaka_mx@hotmail.com*
- Servidor de Metadatos ITESM**
<http://copernico.mty.itesm.mx/bibliotecas/oai/oa/request.cgi>
responsable Isaac Luis Alberto Hurtado Alvarado
A00788662@itesm.mx
- Tutorial de conversión de xml a wml y xhtml usando XSLT**
<http://ict.udlap.mx/cudi/udlatec>
- Valida-OAI** *<http://ict.udlap.mx/cudi/udlatec>* responsable Antonio Felipe Razo Rodríguez *anrazo@mail.udlap.mx*

7. Participantes

Directores del Proyecto

Dr. J. Alfredo Sánchez Huitrón
Laboratorio de Tecnologías Interactivas y Cooperativas ICT

Correo electrónico: alfredo@mail.udlap.mx
Página personal: <http://ict.udlap.mx/people/alfredo>
Teléfono: +52(222)229-2666
Fax: +52(222)229-2431

Universidad de las Américas Puebla.
Sta. Catarina Mártir s/n, San Andrés Cholula.
Puebla, México C.P. 72820
<http://www.udlap.mx>

Dr. David A. Garza-Salazar
Director de Investigación y Posgrado DTIE
División de Tecnologías de Información y Electrónica

Correo electrónico: dgarza@itesm.mx
Página personal: <http://copernico.mty.itesm.mx/~dgarza>
Teléfono: +52(81)8358-2000, ext. 5011.
Fax: +52(81)8358-1400, ext. 5011

Tecnológico de Monterrey, Campus Monterrey
Cetec Torre norte 6° piso.
Eugenio Garza Sada 2501, Col. Tecnológico,
Monterrey, N.L. México C.P. 64849
<http://www.itesm.mx>

Participantes UDLAP

Investigadores

Ing. Lourdes Fernández Ramírez

Asistentes de Investigación

M.C. Antonio Felipe Razo Rodríguez
M.C. Nohema Castellanos

Agentes y Movilidad en Colecciones Multimediales Heterogéneas
J. A. Sánchez

Estudiantes de Posgrado

M.C. María Auxilio Medina Nieto

Estudiantes de Licenciatura

Isaac Abraham Villegas Ortega

Marisol González Rojas

*Participantes ITESM***Investigadores**

Dr. Juan Carlos Lavariega

Mtra. Martha Sordia Salinas

Asistentes de Investigación

M.C. Adán Salinas

Eugenio Flores Centeno

Estudiantes de Posgrado

M.C. Francisco Álvarez Cavazos

M.C. Roberto García Sánchez

Estudiantes de Licenciatura

Marcos Eliud Guevara Camacho

Luis Alberto Hurtado Alvarado

8. Anexos*Colecciones Digitales*

Aquí se describen las colecciones ofrecidas por las bibliotecas digitales en cada una de las instituciones. Las bibliotecas digitales en cada institución cuentan con su propia dirección de acceso y servicios de consulta y búsqueda a estas colecciones.

ITESM -Tesis del programa de graduados de la DECIC

En esta colección se incluyen tesis de alumnos a nivel maestría de la DECIC- DIVISIÓN DE ELECTRÓNICA, COMPUTACIÓN, INFORMACIÓN Y

Agentes y Movilidad en Colecciones Multimediales Heterogéneas
J. A. Sánchez

COMUNICACIÓN. Cuenta con un total de 387 tesis completas en formato PDF y texto, distribuidas en 7 colecciones. No existe restricción alguna para su consulta y su página de Internet es la siguiente:

<http://copernico.mty.itesm.mx/phronesis/mty>

PDLIB fue desarrollada en ANSI-C y Perl se basa en el proyecto Phronesis (<http://copernico.mty.itesm.mx/phronesis/mty>). Implementa estándares como HTML, CGI, PDF, Zip, Gzip y Dublin Core (para metadatos).

Para mayor información contactar:

Ing. Luis Basto

Asistente de la Dirección del programa de Graduados de la DECIC.

Correo electrónico: luisbasto@gmail.com

Teléfono: +52(81)8358-2000 Ext. 5010, 5011

Dr. David A. Garza-Salazar

Director de Investigación y Posgrado DTIE División de Tecnologías de Información y Electrónica Tecnológico de Monterrey, Campus Monterrey

Correo electrónico: dgarza@itesm.mx

Teléfono: +52(81) 8358-2000 Ext. 5010, 5011

UDLA - Tesis Digitales

En esta colección se incluyen tesis de alumnos a nivel licenciatura y posgrado de todas las carreras en la UDLA. A la fecha de este reporte cuenta con un total de 1045 tesis en formato PDF.

El servicio de acceso a Tesis Digitales fue desarrollado en Java, utiliza Servlets (Apache-Tomcat) y una base de datos MySQL. Implementa estándares como HTML, PDF y Dublin Core (para metadatos). Para mayor información contactar:

Ing. María de Lourdes Fernández Ramírez

Coordinación de Colecciones Digitales

Biblioteca, Universidad de las Américas-Puebla

Correo electrónico: lulu@mail.udlap.mx

Teléfono: +52(222) 229-2000 ext. 4141