

TALES: Integración de tesis en una biblioteca digital avanzada

Lourdes Fernández Ramírez

J. Alfredo Sánchez

Alberto García

Biblioteca. Universidad de las Américas-Puebla

0.1 Resumen

Se presenta la creación de una colección de tesis en el marco del proyecto University Digital Libraries for All (U-DL-A) es una iniciativa que explora aspectos del desarrollo de servicios y colecciones para bibliotecas digitales. Evitará las restricciones de espacio físico y las tesis estarán disponibles para todos los interesados en cualquier momento por Internet. Para hacer posible estos servicios se desarrollaron varias herramientas que incluyen desde la recepción electrónica de la tesis hasta la búsqueda y visualización de la misma. Se enfatizó la navegación y búsqueda por elementos de la tesis (figuras, tablas, y otros). De esta manera el usuario podrá visualizar la sección específica y de ahí ir a cualquier otra parte del documento si así lo desea. Presentar la tesis por partes reduce el tiempo de descarga de la página y ayuda al usuario a darse una idea rápida de su contenido.

Palabras clave: Bibliotecas digitales. Tesis digitales. Archivos digitales.

0.2 Summary

The creation of an electronic theses repository in the frame of the project University Digital Libraries for All (U-DL-A) is an initiative set out to explore the issues in the development of high quality digital services and collections is presented. With it, the library and departments will not have to worry about space restrictions or time schedules; moreover, theses will be available to all interested parties at any time on the Web. In order to make these services possible, various tools were developed. These tools include automated reception, thesis visualization and searching. We emphasize searching and browsing at the level of thesis units (such as figures, tables and others). In this way, the user can access any specific unit and from there go to any other thesis unit or to the whole document. Presenting the theses by units reduces loading time and rapidly gives the user a good idea about their content.

Keywords: Digital libraries. Digital theses. Digital repositories.

1. Introducción

Las bibliotecas digitales ofrecen múltiples colecciones de datos organizados y diversos servicios para que sus usuarios puedan aprovechar un universo de información cambiante. La información puede preservarse, almacenarse de manera compacta y permitir búsquedas rápidas y variadas. Además, es posible acceder a los servicios de una biblioteca digital desde cualquier parte del mundo ya que ofrecen navegación y recuperación de información a través de amplias redes de comunicación. Las bibliotecas digitales cambian radicalmente el concepto de la biblioteca tradicional ampliando los recursos y servicios existentes, permitiendo el desarrollo de nuevas maneras de expresarse y resolver problemas. Por estas razones, las bibliotecas digitales se han convertido en una de las áreas de investigación multidisciplinaria más vigorosas.

1.1. Bibliotecas digitales universitarias para todos

Bibliotecas Digitales Universitarias para Todos (U-DL-A, por sus siglas en inglés) es una iniciativa creada para explorar elementos en el desarrollo de ambientes en bibliotecas digitales que faciliten la colaboración entre usuarios distribuidos respondiendo además a sus necesidades individuales (Sánchez 2001). El proyecto considera el trabajo previamente desarrollado en la definición de componentes para una biblioteca digital distribuida así como interfaces de usuario para el manejo de grandes espacios de información.

U-DL-A se ha enfocado a la investigación y desarrollo de dos áreas principalmente: la creación de colecciones digitales y el desarrollo de servicios y ambientes para usuarios finales. Una de las colecciones más importantes que actualmente se construyen en U-DL-A es la que incluye todas las tesis producidas por los estudiantes graduados de nuestra universidad. Este esfuerzo se ha comenzado dentro de la iniciativa internacional de una Biblioteca Digital para Tesis y Disertaciones (Networked Digital Library of Theses and Dissertations, NDLTD) (Fox et al., 1996).

1.2. Biblioteca digital de tesis y disertaciones

En la actualidad existen varias universidades en el mundo que desarrollan colecciones digitales de tesis y disertaciones. NDLTD es una iniciativa para la creación de una biblioteca digital de tesis y disertaciones en la que participan más de 90 universidades (NDLTD, 2001). Su propósito es incrementar el intercambio de conocimiento entre universitarios y así mejorar la calidad de la educación; resaltando el valor de las bibliotecas digitales. Otras organizaciones como la UNESCO apoyan también la esta iniciativa.

1.3. Tesis en la UDLA-P

En la Universidad de las Américas-Puebla (UDLA-P) son generados alrededor de 800 documentos de tesis anualmente, de los cuales aproximadamente el 10% corresponden a tesis de maestría (el sistema educativo de México requiere el desarrollo de una tesis para la mayoría de los programas de licenciatura). Actualmente, debido a restricciones de espacio, la Biblioteca de la UDLA-P sólo cataloga y almacena las tesis de maestría. Las tesis de licenciatura están disponibles sólo en las oficinas de los departamentos académicos. Bajo estas circunstancias, encontrar una tesis en un tema dado puede ser una tarea lenta e inclusive infructuosa, ya que frecuentemente los departamentos no ofrecen catálogos actualizados. Siendo uno de los principales objetivos de la universidad la difusión del conocimiento, la Biblioteca de la UDLA-P ha comenzado a requerir tesis en formato digital (aprobado recientemente en la UDLA-P para algunos departamentos). Nuestro proyecto contempla, entre otras, las siguientes ventajas:

- *Disponibilidad*: Las tesis estarán disponibles permanentemente para toda la comunidad universitaria; no estarán restringidas a un número de copias, a un lugar físico o a un horario de consulta.
- *Almacenamiento electrónico*: No será necesario que los departamentos de la universidad almacenen físicamente el “documento”. Algunos departamentos inclusive no cuentan con el espacio suficiente para almacenarlas o para consultarlas.
- *Reducción de gastos*: El estudiante no gastará en impresiones, copias y empastados.
- *Servicios de búsqueda y navegación sobre la colección*: Esta funcionalidad pone al alcance de los usuarios las ventajas más importantes del medio digital.

TALES es un proyecto que considera la construcción de una colección de tesis digitales y diversos servicios de recuperación de información. Así, las tesis estarán disponibles para todos los interesados a cualquier hora, puesto que pueden ser consultadas a través de Web. La biblioteca no tendrá que preocuparse por restricciones de espacio físico y lo que es mejor, los estudiantes e investigadores no tendrán que buscar colección tras colección para encontrar una tesis, ya que la búsqueda en múltiples colecciones será transparente para el usuario (Chevalier et al., 2001).

2. Diseño conceptual de TALEs

Este proyecto contempla el tratamiento de la tesis desde la recepción hasta su publicación (incluyendo preparación para búsquedas) además de el desarrollo de servicios de navegación y búsqueda de tesis.

2.1. Modelo de datos

Para plantear el modelo de datos se tomaron en cuenta dos aspectos: la integración con los demás componentes de U-DL-A y la estructura propia de la tesis. Para este último se aprovechó la definición de tipos de documentos para tesis electrónicas definido en el *Lenguaje para Mercado de Tesis y Disertaciones Electrónicas* (ETD-ML por sus siglas en inglés) (Kipp et al., 1997), formato aceptado por NDLTD. Se consideraron únicamente los elementos relevantes para este proyecto y se tomó en cuenta que éstos estuvieran cubiertos en el reglamento de tesis de nuestra universidad. Para su adaptación con U-DL-A se consideró principalmente el modelado de las personas y los roles que desempeñan dentro de la Universidad; así como los datos escolares: carreras, departamentos, escuelas y otros.

2.2. Creación de la colección

La creación de la colección se inicia aprovechando el nacimiento digital de la tesis. La mayoría de los estudiantes de esta universidad utilizan un procesador de palabras para la generación de su documento de tesis. La entrega de tesis debe ser de una manera sencilla, para motivar a los estudiantes a participar en la creación de este nuevo servicio. Los formatos más utilizados por NDLTD son PDF, HTML y PostScript. El formato óptimo para proveer nuestros servicios es HTML, porque sólo requiere de un navegador (cualquiera) para su visualización y nos permite seccionar el documento en la manera que lo requiramos. En etapa de pruebas, puede considerarse recibir la tesis en formato DOC (Microsoft Word) ya que para los autores, es la forma más sencilla de hacerlo, no implica trabajo extra o conocimientos nuevos. Esto es manejable cuando se reciben pocas tesis (aproximadamente 50 al semestre) pero no para más por el trabajo que implica su conversión a HTML y revisión.

El proceso comienza con la recepción de tesis. Ya que se planea recibir una gran cantidad de tesis cada semestre, la mejor opción es la recepción electrónica de tesis. A través de una herramienta disponible vía Web, los estudiantes introducirán los archivos que componen su tesis. El administrador (revisor de tesis, bibliotecario) revisará el contenido de los archivos y comentará si los archivos recibidos están completos o falta algún requisito. El usuario será notificado de cualquier comentario realizado por el administrador. Una vez terminado este proceso la tesis pasará al catalogador y a un tratamiento para ser integrada a la colección. El proceso de añadir una tesis a la colección termina cuando la tesis es insertada en la base de datos. Este paso requiere de una revisión léxico-sintáctica en el formato del documento para identificar los elementos definidos en el modelo de datos mencionados en la sección 2.1. El proceso se resume en la figura 1.

Actualmente, en U-DL-A se exploran otros métodos para la creación de la colección, tal es el caso de Zeus, una herramienta desarrollada para auxiliar a estu-

diantes y evaluadores en el proceso de edición de la tesis (Fernández et al., 2000). El director de la tesis (y otros miembros del comité evaluador) realizan correcciones al documento de manera electrónica. El estudiante revisa sus correcciones y actualiza en Zeus la versión corregida del documento. Una vez que la tesis se considere completa, se integrará de manera automática a la colección.

2.3. Servicios de consulta

Los usuarios accederán a la colección de tesis por medio de dos interfaces, una interfaz de búsqueda y una de navegación. La interfaz de búsqueda, especial para tesis, debe mostrar las secciones de la tesis en las que se puede buscar por autor, resumen, lista de figuras, lista de tablas, y referencias entre otros. El usuario podrá seleccionar algunas, o todas, según sus necesidades. El resultado presentará una lista de las tesis que cumplan con los requisitos planteados (palabras clave) resaltando la sección en que se encontraron (resumen, referencias, etcétera). De esta manera el usuario podrá visualizar la sección específica y de ahí ir a cualquier otra parte del documento si así lo deseara. Otros tipos de búsquedas como tema, clasificación, título y autor, pueden ser realizadas desde el sistema de búsqueda de biblioteca, que considera la colección física y la digital.

2.4. Servicios de navegación

Se requiere que la interfaz de navegación presente las tesis clasificadas por elementos como carrera, autor, título de la tesis y año de presentación. Puesto que la cantidad de tesis se incrementa constantemente, es necesario incluir algún tipo de filtro (carrera, año, letra de inicio, etc.). Cada página de la tesis debe contener los datos completos de la tesis y una referencia a las demás secciones del documento (o documento completo) para que sea identificada. Esto es importante ya que se accederá a ellas vía Web. Presentar la tesis por partes reducirá el tiempo de descarga de la página y dará al usuario una idea rápida del contenido de la tesis.

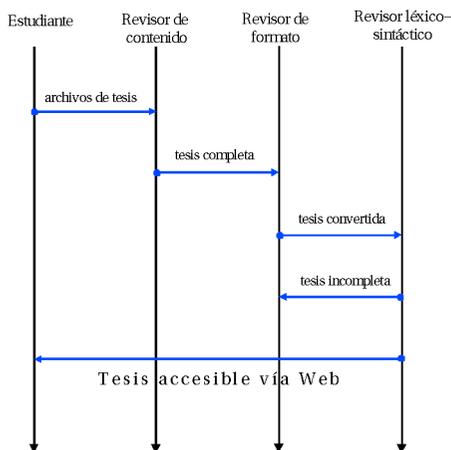


Figura 1. Proceso de creación de la colección

3. Implementación

Para hacer posible estos servicios se desarrollaron varias herramientas que abarcan desde la recepción electrónica de la tesis hasta la búsqueda y visualización de la misma. Estas herramientas fueron implementadas en el lenguaje de programación Java utilizando librerías definidas para U-DL-A. Para que el usuario del sistema sólo requiera un navegador para acceder a los servicios se utilizaron Java Servlets que presentan la información en formato HTML. Como manejador de base de datos se utilizó Informix Universal Server. Las páginas son generadas sobre demanda. En la figura 2, se muestra parcialmente la lista de todas las tesis ordenadas por autor. Se muestran cuatro atributos principales de la tesis, autor(es), título, carrera y año de publicación. La lista puede ser ordenada por cada uno de estos atributos. En la parte superior se muestra una barra con el abecedario para filtrar las tesis por la letra con la que inicie el atributo seleccionado. En este caso, al seleccionar la "A" se mostrarían las tesis ordenadas por autor cuyo primer apellido comience con dicha letra. En cualquier caso, el nombre del autor está referenciado a la portada de la tesis. En la parte inferior de dicha página se muestra el número total de tesis que contiene la colección y la fecha en que se actualizó este catálogo.

La interfaz de búsqueda (figura 3) muestra un campo de texto para introducir la palabra (o frase) a buscar y los elementos en donde se puede buscar. El usuario seleccionará 1 ó más elementos (o se buscará en todos). En este ejemplo se

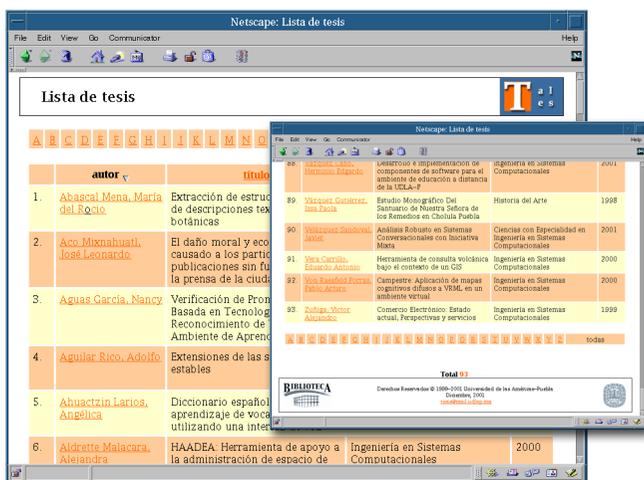


Figura 2. Lista de todas las tesis ordenadas por autor.

buscará la palabra 3D en figuras y referencias. La lista resultante incluye el atributo encontrado en además de los mostrados en la interfaz de navegación (figura 2). Este atributo está referenciado a la sección que describe, misma donde fue encontrada la palabra clave. En este caso, el primer elemento de la lista (figura 3) contiene la palabra como título de alguna figura. La figura 4 muestra la sección del índice de figuras resaltando la palabra buscada. La selección de cualquier elemento dará como resultado la figura misma con un hipervínculo a la sección del documento donde se le hace referencia. En la parte inferior de las páginas se puede observar la barra de navegación que permitirá al usuario acceder a otras secciones de la tesis (resaltando la sección en la que se encuentra actualmente) e inclusive contactar al autor vía correo electrónico.

La colección física y la digital son integradas de manera transparente para el usuario. El índice de la tesis muestra los temas que la describen, catalogados de acuerdo a la clasificación de Biblioteca del Congreso. La selección de cualquiera de estos temas generará una búsqueda en el sistema de biblioteca que dará como resultado todos los libros (u otros elementos) que se encuentran bajo ese mismo tema (parte izquierda de la figura 5). Junto a los temas aparece un vínculo que lleva a la lista de temas que está al mismo nivel que el mencionado. En el ejemplo de la figura 5, al seleccionar esta opción, en el tema *Digital Libraries*, el resultado es: *Digital Libraries Congress*, *Digital Libraries US*, *Digital Libraries Congress US* (figura 5 parte derecha). Por otra parte, cuando el usuario realiza una búsqueda desde el catálogo de biblioteca, por cualquiera de sus opciones, sabrá que ha

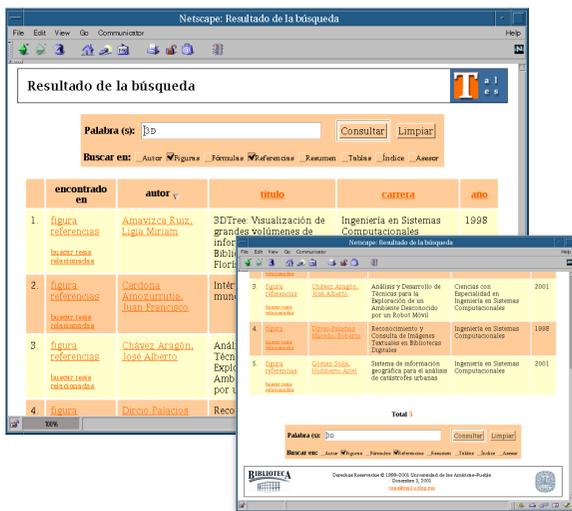


Figura 3. Interfaz resultado de una búsqueda.

4. Trabajo en proceso

Actualmente, el servicio de búsquedas se encuentra en etapa de pruebas mientras que el servicio de navegación está disponible para todo el mundo. Las estadísticas nos muestran que la mayor parte de las visitas provienen de América Latina, teniendo un promedio de 75 visitas por día (<http://biblio.udlap.mx/tesis/>). A la fecha, la colección se compone de 93 tesis (incluyendo licenciatura y maestría). Este proyecto se inició involucrando a un departamento piloto en diciembre de 1999 pero se planea que en los próximos dos años se integre a todos los departamentos de la universidad. A pesar de sólo ser requisito para algunos departamentos, se han recibido voluntariamente tesis pertenecientes a otros departamentos e inclusive a otras universidades, lo que nos demuestra interés de participación por parte de los estudiantes. En este momento, la mayoría de las tesis sólo contienen texto e imágenes y se considera que en un futuro cercano puedan incluir otro tipo de elementos como audio, video, modelos tridimensionales, y más. Se exploran también otras alternativas para la descripción e identificación de los datos como XML. Actualmente, nuestra universidad colabora con instituciones como el ITESM y el Tecnológico de Virginia en la creación de una federación de documentos y de los servicios para accederla. Las tesis digitales de nuestra universidad formarán parte de esta federación. Los usuarios podrán consultar cualquiera de las colecciones existentes como si fuera una sola (Chevalier et al 2001). En el contexto de este proyecto, se ha logrado que la colección de tesis sea proveedora de datos bajo la *Iniciativa de Archivos Abiertos* (OAI).

5. Conclusiones

El área de tesis digitales es un esfuerzo internacional que cada día cobra mayor importancia. Anualmente se organiza un congreso para difundir sus objetivos, atraer más miembros y presentar los avances logrados en esta área (ETD 2001). Siendo uno de los principales objetivos de la Universidad la difusión del conocimiento, se consideró la participación en la iniciativa de NDLTD. Así, la universidad puede brindarle a toda su comunidad un servicio de consultas sobre tesis vía Internet de todas las carreras de licenciatura y maestría que ofrece. Los autores tendrán la oportunidad de publicar su tesis de manera electrónica y dar a conocer su trabajo a todo el mundo (vía Web). Los servicios de navegación y búsqueda planteados permitirán a los usuarios del sistema, consultar la tesis por secciones, reduciendo el tiempo de descarga de la página y le ayudarán a darse una idea rápida del contenido del tesis.

La colección de tesis digitales se integra en la Biblioteca y en la Biblioteca Digital recibiendo los servicios que ambas otorgan. Por ejemplo, de la Biblioteca recibe la catalogación de temas e “integración al librero” y de la Biblioteca Digital puede recibir los servicios de recomendación (Ramírez, 2001), referencia

virtual (Sánchez et al., 2001) y navegación tridimensional en colecciones (Proal et al., 2000).

6. Agradecimientos

Este trabajo es parcialmente financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT). Proyecto número G 33009-A.

7. Referencias

- Chevalier Dueñas, G.; Sánchez, J. A.; Fernández Ramírez, L.; Nava Muñoz, S. (2001). *Viajerus: A framework based on mobile agents for distributed information retrieval*. // *Memorias del Encuentro Internacional de Computación, ENC 01, Aguascalientes, México, Septiembre 2001*. 673-682.
- ETD 2001: Cuarto Simposium Internacional de Tesis y Disertaciones Electrónicas, Marzo 22-24. Pasadena, CA.: Instituto Tecnológico de California, 2001. URL: <<http://library.caltech.edu/etd/>>. Consultado: 16-07-2001.
- Fernández, L.; Sánchez, J. A.; Flores, L. A. (2000). An environment for the collaborative revision of digital theses. // *Proceedings of the 6th Internacional Workshop of Groupware (CRIWG 2000, Madeira Island, Portugal, Octubre 2000)*. Los Alamitos, Ca.: IEEE Computer Society Press, 2000. 150-153.
- Fox, E.; Eaton, J.; McMillan, G.; Kipp, N. A.; Mather, P.; McGonigle, T.; Schweiker, W.; DeVane, B. (1997). *Networked digital library of theses and dissertations: An international effort unlocking university resources*. // *D-Lib Magazine*. (Sept. 1997). URL: <<http://www.dlib.org>>.
- Kipp, N. A.; Fox, E.; Eaton, J. L.; McMillan, G.; Arce, J. E. *Document Type Definition for Electronic Theses and Dissertations*. NDLTD, Scholarly Communications Project, Universities Libraries, Virginia Polytechnic Institute and State University, Blacksburg, VA. URL: <<http://etd.vt.edu/etd-ml/dtdets.htm>>. Consultado: junio 1999.
- Fox, E.; Eaton, J. L.; McMillan, G.; Kipp, N. A.; Mather, P.; Phanouriou, C. (2001). *Networked Digital Library of Theses and Dissertations*. NDLTD Project. 2001. URL: <<http://www.ndltd.org/>>. Consultado: 3-12-2001.
- Proal Aguilar, C.; Sánchez, J. A.; Fernández, L. (2000). *UVA: 3D representations for visualizing digital collections*. // *Proceedings of the Third International Conference on Visual Computing (Visual 2000, Mexico City, Sept. 18-22)*. 185-192.
- Ramírez Vargas, M. (2001). *Servicios de recomendación en la biblioteca digital de la UDLA-P [Tesis Licenciatura]*. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas-Puebla. Mayo 2001.
- Sánchez, J. A. (2001). *HCI and CSCW in the context of digital libraries*. // *Conference on Human Factors in Computing Systems (Extended Abstracts, CHI 2001, Seattle, Wash., April)*. 63-64.
- Sánchez, J. A.; García, A. J.; Proal, C.; Fernández, L. (2001). *Enabling the collaborative construction and reuse of knowledge through a virtual reference environment*. // *Proceedings of the Seventh International Workshop on Groupware (Darmstadt, Germany, Sept. 6-8)*. Los Alamitos, Ca.: IEEE Computer Society Press, 2001. 90-97.