



UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

Facultad 1

Sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.

Trabajo de Diploma para optar por el título de
Ingeniero en Ciencias Informáticas

Autores: Norvelis Legrá Matos.
Yurelkys de los Angeles Carreras Riopedre.

Tutores: Ing. Maikel Manuel Fernández Fernández
Ing. Yanedi Abreu Bartomeo

**Ciudad de la Habana, Cuba.
Junio, 2011**



*"Puesto que a vivir viene el hombre, la educación ha de prepararlo para vivir".
José Martí*

Declaración de autoría

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmamos la presente a los 29 días del mes de junio del año 2011.

Norvelis Legra Matos

Firma del Autor

Yurelkys de los Angeles Carreras Riopedre

Firma del Autor

Ing. Maikel Manuel Fernández Fernández

Firma del Tutor

Ing. Yanedi Abreu Bartomeo

Firma del Tutor

Resumen

El Ministerio del Poder Popular para la Educación Universitaria en Venezuela (MPPEU) se encuentra desarrollando el proceso de municipalización de la Educación Superior. En apoyo a este proceso se propone la creación de la Biblioteca Digital “Alma Mater”, como espacio para acceder a los repositorios de las universidades y a los artículos de revistas científicas de interés para los distintos programas de formación.

El objetivo de este trabajo es desarrollar un sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”, motivado por la necesidad de dicha biblioteca de poseer una herramienta que le permita la recolección de metadatos desde fuentes externas, para complementar la creación de contenidos dentro de la biblioteca. El sistema desarrollado es una herramienta flexible y configurable basadas en el protocolo de recolección de metadatos de la iniciativa abierta de archivos (OAI-PMH) que posibilita un crecimiento acelerado de los fondos de la biblioteca y la familiarización del usuario con una única interfaz de búsqueda.

Para el desarrollo del sistema se elabora el estado del arte de los mecanismos para el intercambio de información en sistemas tipo biblioteca, profundizando en el estudio del protocolo OAI-PMH que condiciona el proceso de indexación de metadatos dentro de la biblioteca. Se definen las características y funcionamiento del sistema y se prueba la factibilidad del mismo.

Palabras Claves

Indexación, biblioteca, digital, documentos, protocolo.

Índice

Resumen.....	II
Índice.....	III
Introducción.....	1
Capítulo1. Estado actual de las tecnologías para la gestión de documentos digitales.	6
1.1. Evolución de las bibliotecas.	6
1.1.1. Conceptualización de biblioteca electrónica, virtual y digital.	7
1.1.2. Características de las bibliotecas digitales.	8
1.2. Representación de metadatos.....	9
1.2.1. Formatos para la representación de metadatos.	9
1.2.1.1. Formato Dublin Core.	10
1.3. Protocolos para la transmisión de Metadatos.....	12
1.3.1. Protocolos de Internet.....	12
1.3.1.1. Protocolo TCP/IP.	12
1.3.1.2. Protocolo HTTP.	13
1.3.2. Protocolo z39.50.	14
1.3.3. Protocolo OAI-PMH.	14
1.4. Sistemas de gestión de contenidos especializados en repositorios digitales.	16
1.4.1. Green Stone.....	16
1.4.2. Fedora.....	19
1.4.3. E-Prints.	20
1.4.4. DSpace.	21
1.5. Lenguajes y herramientas informáticas de apoyo para el desarrollo.	24
1.5.1. Sistemas de Gestión de Contenidos.....	24
1.5.1.1. CMS Drupal.	25
1.5.2. HTML y CSS.....	27
1.5.3. XML.....	27
1.5.4. RDF.....	28

1.5.5. Lenguajes dinámicos.....	28
1.5.5.1. PHP.....	29
1.5.6. Gestores de Base de datos.....	29
1.5.6.1. PostgreSQL.....	29
1.5.6.2. MySQL.....	30
1.5.7. IDE NetBeans 6.8.....	30
1.6. Metodologías de Desarrollo.....	31
1.6.1. XP.....	31
1.6.2. SXP.....	31
1.6.3. Proceso Unificado de Desarrollo de Software (RUP).....	32
1.7. Estructuras de datos.....	33
Capítulo 2. Caracterización del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.....	35
2.1. Propuesta del sistema.....	35
2.2. Modelo de dominio.....	35
2.2.1. Definición de las entidades y conceptos principales.....	35
2.3. Levantamiento de requisitos.....	37
2.3.1. Requisitos funcionales.....	37
2.3.2. Requisitos no funcionales.....	37
2.4. Modelo de Casos de Uso.....	39
2.5. Especificaciones de Casos de Uso.....	40
2.6. Modelo de diseño.....	47
2.6.1. Diagramas de clases del diseño.....	49
2.7. Modelo de datos.....	50
Capítulo 3. Implementación y pruebas del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.....	52
3.1. Diagrama de Despliegue.....	52
3.2. Diagrama de Componentes.....	53
3.3. Pruebas.....	54

3.3.1. Caso de Prueba para el caso de uso: Adicionar fuentes OAI	55
3.3.2. Caso de Prueba para el caso de uso: Importar documento de archivo local. ...	56
3.3.3. Caso de Prueba para el caso de uso: Importar documentos de Internet.	57
3.3.4. Caso de Prueba de integración.	58
Conclusiones.....	61
Recomendaciones	62
Glosario de Términos.....	63
Referencias Bibliográficas	65
Bibliografía	69

Introducción

La búsqueda, disponibilidad y organización de la información es uno de los problemas que el ser humano ha buscado resolver desde sus inicios. La necesidad de contar con una fuente de información provocó el surgimiento de las primeras bibliotecas, que nacieron en los templos y su primera función fue solamente conservadora, pues la información solo podía ser accedida por los escribas y sacerdotes. (Setién, 2005)

Con el tiempo se diversificó la información guardada y se generalizó el acceso a la lectura y al libro, de esa manera aparecieron las bibliotecas desvinculadas de los templos cuya función se enfocaba en suplir las necesidades informativas de los visitantes. La biblioteca se convirtió finalmente en una “institución cuya finalidad consiste en la adquisición, conservación, estudio y exposición de libros y documentos” (Diccionario de la Lengua Española, 2010)

Con el surgimiento de la informática se ha encontrado una vía para almacenar, compartir y gestionar grandes colecciones de datos digitalizados. La posibilidad de digitalizar los documentos conllevó al surgimiento de lugares donde se gestionan documentos electrónicos tangibles, conocidos como bibliotecas electrónicas.(Pérez, 2007)

Más adelante, con el uso creciente del acceso en línea derivado de los progresos en las comunicaciones y la tecnología informática, aparecieron repositorios de documentos de todo tipo, generados por diferentes instituciones. Repositorios que aumentaron considerablemente y cambiaron notablemente su visualización, adoptando interfaces cada vez más sencillas de utilizar. Se caracterizaron por incorporar nuevos elementos: texto, datos, imágenes, figuras 3D, gráficos, vídeos y audio.

Estos avances dieron paso al surgimiento de las bibliotecas digitales, que según definiciones provenientes del ámbito bibliotecario son organizaciones que ofrecen servicios de valor añadido sobre recursos de información digital, al tiempo que insisten en la problemática de acceso y perduración de las colecciones.(Pérez, 2007)

El Dr Jesús Tramullas Saz en su artículo Bibliotecas Digitales: Una revisión de conceptos y técnicas hace un análisis detallado de los diferentes conceptos de biblioteca digital y expresa que “*dada la dificultad de*

proponer una definición debe considerarse que una biblioteca digital es un espacio de información digital, que reúne un conjunto de recursos de información, en el marco de un proceso multifacetado de transferencia de información desde la fuente al usuario, proceso que se caracteriza por su valor añadido”.

(Tramullas, 2002)

En la actualidad con el desarrollo de las tecnologías, grupos de personas con objetivos comunes, han encontrado la vía para socializar conocimiento entre ellos y permitir que los documentos se encuentren disponibles en el formato digital para ser consultados cuando sea necesario.

La disponibilidad de documentos en formato digital y el desarrollo de Internet, han permitido que las bibliotecas digitales que en principio sus documentos eran introducidos manualmente ahora se nutran de fuentes ya existentes, a través de protocolos de comunicación que permiten el intercambio de metadatos y brindan la posibilidad de recolectar y almacenar información de diversos orígenes.

En la mayoría de los países existen instituciones que tienen sus propias bibliotecas digitales. Venezuela es uno de estos países, cuyo Ministerio del Poder Popular para la Educación Universitaria (MPPEU) se encuentra desarrollando el proceso de municipalización de la Educación Superior, que orienta de manera integral y democrática a toda la población que desee ingresar a la educación superior desde su propia región, localidad y parroquia. La municipalización tiene una nueva concepción de la formación universitaria en la que se crean espacios educativos en cada uno de los municipios de la Nación propiciando el trabajo compartido de las comunidades con la universidad. (Guzmán, 2005)

En apoyo a la municipalización el MPPEU ha puesto a disposición de estudiantes, profesores e investigadores un conjunto de aplicaciones informáticas que complementan los procesos de formación e investigación. Entre estas aplicaciones se propone la creación de la Biblioteca Digital “Alma Mater”, como espacio para acceder a los repositorios de universidades y a los artículos de revistas científicas de interés para los distintos programas de formación. Se necesita que los fondos de la biblioteca crezcan de forma acelerada y que los usuarios desde una única interfaz de búsqueda puedan acceder a documentos disponibles en **fuentes externas** específicamente en canales de información de acceso abierto.

A partir de la necesidad social antes expresada, se plantea el siguiente **problema científico**: ¿Cómo garantizar la indexación de metadatos de **fuentes externas** para la creación de contenidos dentro de la Biblioteca Digital “Alma Mater”?

El diseño de la investigación permite declarar como **objeto de estudio**: Las Tecnologías para el intercambio de información entre sistemas tipo biblioteca. Centrando el **campo de acción en**: Los protocolos para la transmisión de metadatos entre sistemas tipo biblioteca.

En el contexto investigativo se determina como **objetivo general**: Desarrollar un sistema para la indexación de metadatos desde **fuentes externas** que complemente la creación de contenidos en la Biblioteca Digital “Alma Mater”.

Para darle cumplimiento al objetivo general, este ha sido desglosado en los siguientes **objetivos específicos**:

- Elaborar el estado del arte de los mecanismos para el intercambio de información entre sistemas tipo bibliotecas.
- Definir las características y funcionamiento del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.
- Implementar el sistema de indexación automática de contenidos para la Biblioteca Digital “Alma Mater”.
- Probar la factibilidad del sistema.

La idea que se defiende con el trabajo es: El Desarrollo de un sistema para la indexación de metadatos desde **fuentes externas**, permitirá contar con una herramienta que complemente la creación de contenidos en la Biblioteca Digital “Alma Mater”, proporcionando un crecimiento acelerado de los fondos de la misma y la familiarización del usuario con una única interfaz de búsqueda.

Para guiar el proceso investigativo se determinan las siguientes **tareas científicas**:

- Fundamentar la necesidad de desarrollar el sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.

- Describir el funcionamiento del protocolo OAI-PMH para el intercambio de información que implementará el sistema.
- Determinar el formato de metadatos que soportará el sistema de indexación.
- Caracterizar los principales sistemas de gestión de contenidos especializados en repositorios digitales.
- Seleccionar herramientas y tecnologías para el desarrollo del trabajo.
- Describir las funcionalidades del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.
- Crear las estructuras de datos necesarias para el funcionamiento del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.
- Implementar las distintas funcionalidades del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.
- Realizar pruebas de funcionamiento al sistema de indexación.

La realización de las tareas de investigación se concreta sobre la base del uso de **Métodos Científicos** tanto teóricos como empíricos. Estos permiten conocer la esencia del problema abordado.

Uno de los métodos teóricos empleado en la presente investigación es el método analítico – sintético, para hacer un análisis detallado de los componentes de los protocolos de intercambio de metadatos y a partir de este análisis diseñar un sistema que permita la indexación de documentos. También se utiliza en el estudio de las herramientas para desarrollar este tipo de sistemas y llegar a una conclusión para seleccionar el entorno tecnológico de desarrollo más conveniente.

El método histórico-lógico permite a partir de elementos fehacientes en el desarrollo de las bibliotecas digitales y los sistemas de indexación de documentos, observar detalles de su uso a escala mundial,

desde su surgimiento hasta el desarrollo actual. Al estudiar particularidades básicas de su esencia se siguió una línea lógica en la indagación sobre el funcionamiento y progreso de los mismos.

Los métodos empíricos empleados son la encuesta para obtener información sistemáticamente

El presente informe está estructurado en 3 capítulos:

Capítulo 1. Estado actual de las tecnologías para la gestión de documentos digitales.

En este capítulo se exponen los principales conceptos necesarios para comprender el posterior desarrollo de este proyecto. Se realiza un estudio de los sistemas de indexación de documentos utilizados en las bibliotecas digitales y de los protocolos de comunicación que emplean. También se caracterizan las tecnologías informáticas usadas en la solución del problema.

Capítulo 2. Caracterización del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.

En este capítulo se definen las características del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”. Se hace una propuesta general de cómo debe funcionar. Se describe el modelo de dominio. Se elabora el modelo de casos de uso y se muestran los diagramas de clases correspondientes al diseño y se elabora y describe el modelo de datos.

Capítulo 3. Implementación y pruebas del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.

En este capítulo se define el modelo de despliegue. Se modela la vista estática del sistema representada como un grafo de componentes de software y se construye el modelo de prueba que permite constatar que las funcionalidades implementadas son correctas.

Capítulo 1. Estado actual de las tecnologías para la gestión de documentos digitales.

El desarrollo tecnológico ha traído consigo una revolución en el trabajo de las bibliotecas. La constante necesidad de mejorar la organización, acceso y distribución de los documentos ha provocado el desarrollo de sistemas especializados en la gestión de los repositorios donde se almacena la información, específicamente sistemas indexadores de metadatos.

El presente capítulo tiene el objetivo de ilustrar los aspectos más importantes del soporte teórico en la construcción del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”. Se exponen los principales conceptos relacionados con el ámbito del problema. Como elemento indispensable se aborda el estado del arte de estos tipos de sistemas, así como la selección de la metodología a utilizar para el desarrollo del sistema, partiendo de las facilidades que puede aportar al trabajo. También se definen los estándares, protocolos y herramientas a utilizar para la implementación del proyecto teniendo en cuenta las necesidades existentes y el entorno donde se enmarca el objeto de estudio.

1.1. Evolución de las bibliotecas.

El atesoramiento de colecciones y su uso han sido rasgos distintivos de las bibliotecas durante toda su historia, desde aquellos tiempos de las civilizaciones fluviales, primera fase de la sociedad esclavista, cuando la escritura, la lectura y las bibliotecas estuvieron a cargo principalmente de escribas y sacerdotes, hasta las sociedades contemporáneas, donde existe libre acceso a las bibliotecas de todos los sectores de la población. (Setién, 2005)

El desarrollo tecnológico ha revolucionado el trabajo propio de una biblioteca, sobre todo, como resultado de las demandas de los usuarios actuales, quienes exigen una respuesta rápida, directa y relevante a sus necesidades de información. Estos factores han determinado la aparición de las bibliotecas electrónicas, digitales y virtuales, sin embargo, dichos términos se emplean indistintamente con diversas acepciones. (Sánchez, 2002)

1.1.1. Conceptualización de biblioteca electrónica, virtual y digital.

La biblioteca electrónica sería aquella que permite acceder a bancos de información en formato electrónico. Este tipo de bibliotecas incluiría también los catálogos automatizados de bibliotecas tradicionales. Según esta definición la biblioteca electrónica intentaría reproducir la producción impresa pero utilizando un medio diferente del soporte papel.(Pérez, 2007)

Se habla de bibliotecas electrónicas, cuando el acceso a la información se ofrece por medios informáticos de modo local.(Sánchez, 2002)

La biblioteca virtual es aquella en la que interviene, de forma constante el elemento humano, brindando servicios a distancia y aprovecha la respuesta virtual de los usuarios para organizar la información y los contenidos que éste necesita, o para darle las herramientas necesarias para obtener información.(Pérez, 2007)

La biblioteca virtual se basa en la prestación de servicios bibliotecarios y documentales a través de Internet y, más concretamente, de la tecnología Web.(Faba, 2004)

Se entiende como biblioteca digital, una colección de documentos disponibles en formato electrónico y que son accesibles desde una red telemática es decir aquella que contiene elementos digitalizados (...) de todo tipo, que incluso, a veces tienen una forma cambiante a causa de su constante actualización, que aplica la tecnología en cualquiera de sus formas pero en la que no actúa el elemento humano de una manera directa y simultánea al servicio que da.(Pérez 2007)

Las bibliotecas digitales son bibliotecas electrónicas en las que un gran número de usuarios distribuidos geográficamente tienen acceso a una gran cantidad de contenidos diversos de objetos electrónicos.(Faba, 2004)

Las bibliotecas electrónicas, virtuales y digitales difieren en su definición sin embargo usan las tecnologías de la informática y las comunicaciones para su funcionamiento. Además las virtuales y digitales emplean las redes como canal de acceso a sus servicios.

1.1.2. Características de las bibliotecas digitales.

La biblioteca digital funciona de modo similar a una biblioteca tradicional tiene secciones, donde están ubicados diferentes clases de documentos, para seleccionar, estructurar, distribuir, controlar el acceso, conservar la integridad y asegurar la persistencia a través del tiempo de colecciones de trabajos digitales.

Su objetivo es el acceso universal a la información, sin limitantes de tiempo ni espacio, y se caracteriza por los siguientes atributos: disponibilidad, recuperación, autenticidad, utilización, asequibilidad y tecnología. (Plaza, 2003)

En la medida en que dan acceso a grandes repositorios de información, tienen que desarrollar herramientas y tecnologías para dar valor añadido al acceso al conocimiento y al significado inherente a los contenidos de las colecciones digitales.

Una de sus características es que en ellas se almacenan materiales en diferentes formatos, distintas versiones y ubicados en diferentes lugares que pueden ser accedidos por un gran número de personas.

La digitalización de los contenidos brinda facilidad de acceso, velocidad en la recuperación de la información, transmisión en línea de forma remota y gran capacidad de almacenamiento.

Esta nueva forma de representación permite que a la información textual se añada información no textual, por ejemplo: fotografías, dibujos, ilustraciones; gran cantidad de datos numéricos; sonidos digitalizados; imágenes en movimiento; representaciones multidimensionales entre otros elementos.

El trabajo interdisciplinar es un elemento fundamental para el crecimiento exponencial de los datos en una biblioteca digital. Una coordinación a gran escala de científicos, ingenieros, informáticos, editores y bibliotecarios es esencial para crear, adaptar nuevas tecnologías y desarrollar nuevas normativas.

Los derechos de autor y la legislación sobre la propiedad intelectual son aspectos de vital importancia tanto para la creación de estas bibliotecas como para su protección.

Para la confección de bibliotecas digitales, se usan varios formatos que pueden ser de tipo abierto o cerrado: Son de tipo abierto los que pueden manipularse con más facilidad y de tipo cerrado los que presentan más dificultades para ser cargados, modificados o manipulados.

La biblioteca digital debe permitir la recuperación de información mediante metadatos, los que permiten una búsqueda efectiva y precisa proporcionando valor añadido a la mera acumulación de información.(Pérez, 2007)

En la actualidad existe tendencia a crecer anexando contenido, debido al desarrollo de mecanismos que permiten la comunicación entre las bibliotecas digitales. Estos mecanismos han brindado la posibilidad de indexar gran cantidad de datos de diferentes tipos, sin necesidad de utilizar un gran ancho de banda ya que se indexa la representación del objeto y no el objeto en sí.

1.2. Representación de metadatos.

Los mecanismos de comunicación que han impulsado la tendencia de las bibliotecas digitales a crecer anexando contenidos, parten de la estandarización de la información. Donde juegan un papel importante los formatos para la representación de metadatos o estándares de metadatos.

Los estándares permiten, en cualquier área, definir modelos o patrones de hechos objetivos o subjetivos. No solo definen la forma en que está descrito un conjunto de datos y los valores individuales, sino también las bases en que una institución puede interactuar e interoperar con las colecciones de instituciones similares (Marshall, 1998).

1.2.1. Formatos para la representación de metadatos.

El término de metadatos se refiere a aquellos datos que describen las características de un recurso y son sumamente importantes cuando se quiere tener un conocimiento avanzado total de las características de los recursos (Dempsey, Heery, 1999).

Los metadatos son utilizados para estructurar información de tal manera que permitan mostrar todas las características propias. (Kuang, 2000).

Los metadatos son datos que se utilizan para ayudar a la identificación, descripción y localización de recursos electrónicos enlazados en red. Existen diversos formatos de metadatos, que van desde estructuras simples hasta otras más complejas. (Sené, 2009)

Se puede decir entonces, con base en las definiciones vistas anteriormente, que los metadatos son útiles para estructurar la información de tal manera que sea presentada como un desglose general de todos los datos que la integran.

Dentro de los estándares de metadatos más conocidos se encuentran: Dublin Core, Warwick Framework, Global Information Locator Service (GILS) y Content Standard for Digital Geospatial Metadata (CSDGM), los cuales se caracterizan por tener distintas áreas de aplicación.

Las fuentes de las que se nutre la Biblioteca Digital “Alma Mater” emplean el protocolo de recolección de metadatos de la iniciativa abierta de archivos (OAI-PMH), que requiere que todos los servidores ofrezcan sus servicios utilizando Dublin Core como formato para la representación de metadatos. En aras de proporcionar una buena comunicación entre estos sistemas se emplea el formato Dublin Core.

1.2.1.1. Formato Dublin Core.

El Dublin Core es la forma abreviada para el Dublin Metadata Core Element Set (Conjunto básico de elementos metadatos de Dublín).

Nació en 1995 en Dublin, Ohio, Estados Unidos. Fue creado por bibliotecarios, científicos en computación, archivistas, y proveedores de servicios en línea en la convención del Online Computer Library Center (OCLC), National Center for Supercomputing Applications (NCSA) y Softquad.

Actualmente varias organizaciones han contribuido en el desarrollo de este estándar como son: Research Libraries Group, Coalition for Networked Information, Internet Engineering Task Force, UK Office for Library and Information Networking, y el National Library of Australia (Kuang, 2000).

Dublin Core permite describir los elementos básicos de identificación de cualquier tipo de documento. Su uso está ampliamente relacionado con las bibliotecas tradicionales, digitales y con el Internet.

La amplia difusión de dicho estándar depende de tres características potenciales:

- Es un formato de descripción simple de un gran número de fuentes.
- Brinda una base semántica de metadatos para HTML (Hypertext Markup Language), analizado en un apartado posterior.
- Define una base para la interoperabilidad semántica entre dominios de metadatos. (Endemi, 2005)

Su objetivo principal es crear un conjunto de elementos de datos que describan los documentos electrónicos de las redes con el fin de facilitar su búsqueda y recuperación. Está compuesto por un conjunto mínimo de elementos que facilitan la recuperación de información en la red. Está diseñado para facilitar la recuperación de recursos en las redes de una forma similar a un catálogo de biblioteca, pero con una estructura mucho más sencilla. Está formado por 15 elementos de datos y su punto más fuerte es que el diseño es tan intuitivo que los propios proveedores de información pueden codificar sus documentos al mismo tiempo que los crean. (Repiso, 1999)

Hay muchos proyectos que están experimentando con el Dublin Core, la mayoría incluye la interoperabilidad entre diferentes bases de datos, objetos digitales y otras formas de metadatos. Entre ellos cabe destacar los que se están desarrollado en:

- La Biblioteca Nacional de Australia. Esta institución junto con otros organismos oficiales está utilizando el Dublin Core para proporcionar recuperación y acceso a 70 bases de datos de gran variedad y construidas en diferentes formatos.

- En varias bibliotecas nórdicas y en el Centro de Tecnologías y Sistemas Distribuidos de Australia se realizan proyectos para convertir el MARC en registros Dublin Core y facilitar, así, la recuperación de información en las redes. (Repiso, 1999)

1.3. Protocolos para la transmisión de Metadatos.

La existencia de formatos de metadatos para describir los documentos es de suma importancia, pero no solo basta con que estén descritos, hace falta que esa información se transmita lo cual se hace a través de los protocolos.

Los protocolos son un conjunto de hábitos y procedimientos utilizados en las relaciones interpersonales. Cuando es usado bajo el contexto de redes de comunicación, el término protocolo tiene un significado similar pero a un nivel más específico. Un protocolo de red es un conjunto de reglas, secuencias, formatos de mensajes y procedimientos bien detallados que posibilitan la transferencia de datos entre dos o más sistemas de computación, el término utilizado para describir como los sistemas de computación se comunican con otros a nivel de bit y de byte. (Duran, 2007)

1.3.1. Protocolos de Internet.

El sistema de indexación de documentos para la Biblioteca Digital “Alma Mater” es un módulo para una aplicación Web, por lo que emplea protocolos de Internet, como TCP/IP y HTTP para garantizar la disponibilidad vía http y el acceso a la biblioteca desde cualquier lugar.

1.3.1.1. Protocolo TCP/IP.

TCP/IP es resultado de la investigación y desarrollo llevados a cabo en la red experimental de conmutación de paquetes ARPANET, financiada por la Agencia de Proyectos de Investigación Avanzada para la Defensa (DARPA), y se denomina globalmente como la familia de protocolos TCP/IP. Esta familia consiste en una extensa colección de protocolos que se han establecido como estándares de Internet. (Atallings, 2000)

Proporciona la base para el Internet global, una gran red de redes operacional que interconecta universidades, corporaciones y dependencias gubernamentales en muchos países alrededor del mundo. Está basado en una tecnología convencional de conmutación de paquetes y es independiente de cualquier marca de hardware en particular. Una de las mayores cualidades del TCP/IP radica en la variedad de tecnologías de red física sobre las que se puede utilizar.

Está diseñado para proporcionar interconexión universal entre maquinas, independientemente de las redes en particular a las que están conectadas. Utiliza direcciones binarias de 32 bits como identificadores universales de máquinas. Llamadas direcciones IP o direcciones de red de redes. (Comer, 1996)

1.3.1.2. Protocolo HTTP.

El Protocolo de Transferencia de Hipertexto es un sencillo protocolo cliente-servidor que articula los intercambios de información entre los clientes Web y los servidores HTTP. Fue propuesto por Tim Berners-Lee, atendiendo a las necesidades de un sistema global de distribución de información como el World Wide Web.

Desde el punto de vista de las comunicaciones, está soportado sobre los servicios de conexión TCP/IP, y funciona de la siguiente forma: un proceso servidor escucha en un puerto de comunicaciones TCP, por defecto el 80 y espera las solicitudes de conexión de los clientes Web. Una vez que se establece la conexión, el protocolo TCP se encarga de mantener la comunicación y garantizar un intercambio de datos libre de errores.

Se basa en sencillas operaciones de solicitud/respuesta. Un cliente establece una conexión con un servidor y envía un mensaje con los datos de la solicitud. El servidor responde con un mensaje similar, que contiene el estado de la operación y su posible resultado. Todas las operaciones pueden adjuntar un objeto o recurso sobre el que actúan; cada objeto Web es conocido por su URL. (Romero, 1999)

1.3.2. Protocolo z39.50.

El Z39.50 es un protocolo para la recuperación de información basado en la estructura cliente-servidor que facilita la interconexión de sistemas informáticos y que podría considerarse como la norma más relevante para el mundo de las bibliotecas y de la documentación desde la aparición del formato MARC.

El objetivo principal de este protocolo consiste en permitir al usuario realizar búsquedas en bases de datos que cuenten con un servidor Z39.50, sin tener que conocer para ello las sintaxis de búsqueda que utilizan dichos sistemas.

Facilita la interconexión entre los usuarios y las bases de datos donde se encuentra la información que necesitan a partir de una interfaz común y brinda fácil manejo, independientemente del lugar en que se encuentren las bases de datos así como la estructura y la forma de acceso de éstas.

Especifica un conjunto de reglas para gestionar las formas y procedimientos de interconexión remota de computadoras, con el propósito de buscar y recuperar información, aunque su aplicación actual es más amplia pues incluye la consulta y el intercambio de datos bibliográficos, la intercomunicación de índices y resúmenes de: información geoespacial, documentos oficiales, objetos digitales o de metadatos que describen los documentos de las bibliotecas electrónicas y digitales. (Guajardo, 2010)

1.3.3. Protocolo OAI-PMH.

El protocolo de recolección de metadatos de la incitativa abierta de archivos (OAI-PMH) es una herramienta de interoperabilidad que posibilita el intercambio de metadatos sobre cualquier material almacenado en soporte electrónico. (Guajardo, 2010)

La arquitectura del protocolo rechaza la búsqueda distribuida, a diferencia del Z39.50. Simplemente posee servidores proporcionado metadatos, sujetos sólo a criterios de alcance bastante simples, como proporcionar todos los registros añadidos o cambiados desde una fecha específica.

Básicamente utiliza transacciones HTTP para emitir preguntas y obtener respuestas entre un servidor o archivo y un cliente o servicio recolector de metadatos. El segundo puede pedir al primero que le envíe metadatos según determinados criterios como la fecha de creación de los datos. En respuesta el primero devuelve un conjunto de registros en formato XML, incluyendo identificadores de los objetos descritos en cada registro. (Barrueco, 2003)

Las peticiones se emiten utilizando los métodos GET o POST del protocolo HTTP y constan de una lista de opciones con la forma de pares del tipo: clave=valor. Existen seis peticiones que un cliente puede realizar a un servidor:

- **GetRecord:** Utilizado para recuperar un registro concreto. Necesita dos argumentos, identificador del registro pedido y especificación del formato bibliográfico en que se debe devolver.
- **Identify:** Utilizado para recuperar información sobre el servidor: nombre, versión del protocolo que utiliza, dirección del administrador, etc.
- **ListIdentifiers:** Recupera los encabezamientos de los registros, en lugar de los registros completos. Permite argumentos como el rango de fechas entre los que queremos recuperar los datos.
- **ListRecords:** Igual que el anterior pero recupera los registros completos.
- **ListSets:** Recupera un conjunto de registros. Estos conjuntos son creados opcionalmente por el servidor para facilitar una recuperación selectiva de los registros. Sería una clasificación de los contenidos según diferentes entradas. Un cliente puede pedir que se recuperen solo los registros pertenecientes a una determinada clase. Los conjuntos pueden ser simples listas o estructuras jerárquicas.
- **ListMetadataFormats:** Devuelve la lista de formatos bibliográficos que utiliza el servidor.

El protocolo soporta múltiples formatos para expresar los metadatos, no obstante requiere que todos los servidores ofrezcan los registros utilizando Dublin Core codificado en XML. Además de éste formato cada servidor es libre de ofrecer los registros en otros formatos adicionales como MARC. Un cliente puede pedir que los registros se le sirvan en cualquiera de los formatos soportados por el servidor. (Barrueco, 2003)

La arquitectura de este protocolo se basa en proveedores de datos y proveedores de servicios; los primeros son los archivos que proporcionan la información y los segundos son los recolectores o servicios que toman los datos, y los presentan a los usuarios finales.

La aplicación de OAI-PMH permite realizar el intercambio de información para que desde puntos centralizados, proveedores de servicios puedan realizar búsquedas conjuntas sobre los metadatos de todos los repositorios asociados, específicamente los Open Access (Acceso Abierto) que promueve eliminar las barreras económicas, legales y tecnológicas, y trata de obtener a cambio, como beneficios, una mayor accesibilidad para los documentos y una mayor visibilidad para los autores. Los documentos que están disponibles en Acceso Abierto son más consultados y tienen más posibilidades de ser citados.

Las fuentes principales de las que se nutre la Biblioteca Digital “Alma Mater” son los archivos de acceso abierto quienes basan su comunicación en el protocolo OAI-PMH por lo que el sistema de indexación de documentos para la Biblioteca Digital “Alma Mater” utiliza este protocolo e implementa específicamente un proveedor de servicios. No hay necesidad de implementar el proveedor de datos, pues la comunidad de Drupal ha desarrollado el módulo OAI2forCCK que implementa perfectamente el proveedor de datos.

Este módulo proporciona una implementación del protocolo OAI para indexación de metadatos con soporte para tipos de contenido CCK y sus campos. Su implementación permite exponer los metadatos como repositorio OAI-PMH, recibir peticiones OAI-PMH de proveedores de servicios en forma de petición HTTP y responder a dichas peticiones en formato XML. (Drupal.org, 2010)

1.4. Sistemas de gestión de contenidos especializados en repositorios digitales.

Toda biblioteca digital debe tener un repositorio donde almacenar la información que posee. Existen varias herramientas que permiten la gestión de repositorios y cada una de ellas cuenta con sus propios métodos para la indexación de contenidos.

1.4.1. Green Stone.

Greenstone es una suite completa de software para la organización de la información y la distribución de colecciones digitales. Este sistema provee una nueva forma de organizar la información a través de

Internet o distribuirlo en CD-ROM. Greenstone ha sido producido por el proyecto de la Biblioteca Digital de Nueva Zelanda, en la universidad de Waikato.(Albarado, 2010)

Trabaja con documentos con contenido textual y con metadatos, extraídos automáticamente o introducidos manualmente, organizándolos en colecciones digitales. Permite desarrollar estructuras de organización de la colección. Es multilingüe y está implementado en Perl, usa el motor de representación y recuperación de información MG, o su versión avanzada MGPP, que usan el modelo vectorial. Codifica toda la información en XML. Necesita de un servidor Web de soporte e incorpora una interfaz gráfica en Java.(Tramullas 2008)

El sistema Greenstone es una aplicación pensada para la construcción y administración de bibliotecas digitales accesible desde la web, proporcionando de esta forma el acceso tanto a documentos textuales como a imágenes, vídeos y sonidos. Su organización permite distribuir los documentos en colecciones que se pueden crear atendiendo a sus características. Por otra parte, favorece la importación y asignación de metadatos, de manera manual o automática. Permite configurar el tipo de indexación a utilizar, modalidades de búsqueda y obtener listados de resultados en función de la relevancia con los términos consultados.(Cyraneck, 2010)

Las implementaciones mediante este sistema pueden publicarse en internet y en entornos más específicos, las intranets, tanto para recursos digitales o la captura del acervo intelectual producido internamente en una organización. Es totalmente compatible con el protocolo OAI-PMH.

Funciona en todas las versiones de Windows, Unix/Linux y también en Mac OS-X. Es muy sencillo de instalar y trabaja con otros programas libres de software, incluyendo el servidor Web Apache y Perl. Para trabajar con esta aplicación existen dos interfaces interactivas separadas: la página del lector y la interfaz de bibliotecario de Greenstone.

La descripción de los documentos se puede realizar de manera individual, documento a documento, o describir las distintas agrupaciones en conjunto. De esta última manera se asigna a todos los recursos incluidos en cada agrupación la misma descripción de manera automática, desarrollando así un práctico sistema multinivel.(Sola, 2008)

Greenstone permite organizar los documentos de la biblioteca a partir de las colecciones que posee por defecto o a partir de la creación de nuevas colecciones.

Una biblioteca digital típica creada con esta herramienta contendrá muchas colecciones, organizadas individualmente, pero muy parecidas entre sí. Las colecciones pueden aumentarse y reorganizarse automáticamente.

El sistema requiere un usuario y contraseña para registrarse, además una dirección de correo electrónico para utilizarla en caso de que su contraseña sea olvidada.

Greenstone es capaz de procesar un considerable número de documentos de diversa índole: textos (.txt, .doc, .pdf, .rtf), páginas Web (html, htm, xhtml, mhtml), imágenes (.jpeg, .jpg, .gif, .png), bases de datos (ISIS), vídeos, audio, mensajes de correo y otros. Tal procesamiento depende, en gran medida, de la utilización de conectores (plugins). A cada tipo de documento corresponde un tipo de conector.

La arquitectura del programa permite la utilización de plugins y clasificadores aun para datos no textuales. Una colección puede tener documentos de origen de diferentes formas. Esta aplicación es capaz de procesar archivos zip, descomprimiéndolos al momento de importarlos.

Greenstone soporta la utilización de esquemas de metadatos estándar y no estándar. El usuario es quien define el esquema a utilizar a través de la interfaz de Bibliotecario. Estos conjuntos de metadatos ya vienen predefinidos y pueden seleccionarse para ser utilizados en las colecciones: Dublin Core (esquema básico y cualificado), RFC 1807, NZGLS (New Zealand Government Locator Service) y AGLS (Australian Government Locator Service) (Cyraneck, 2010)

Dentro de las ventajas aportadas por Greenstone se tiene:

- Integración de todos los procesos necesarios para gestionar miles o millones de recursos digitales.
- Potente búsqueda a texto completo y múltiples índices de búsqueda basados en metadatos.
- Disponibilidad de vasta documentación, soporte y mejoramiento.
- Seguridad y control de acceso.
- Integración con otros sistemas.
- Acceso ilimitado a las colecciones vía Internet o intranet o por un CD-ROM

- Posibilidad de automatizar un servicio del tipo biblioteca digital con un mínimo de inversión y sin costos ocultos.
- Posibilidad de compartir experiencias sin restricciones dentro de la comunidad internacional dado su carácter de software fuente abierta.
- Comprende una función de auto instalación que permite instalar fácilmente el programa en Windows y Unix.
- Su aplicación abarca un amplio abanico de posibilidades que cubre las necesidades tanto de pequeñas bibliotecas como las de importantes entidades que requieran programas capaces de satisfacer exigencias más elevadas. (Arias, 2008)
- Posibilidad de importar directamente bases de datos cds/isis (greenstone, e-print, d-space).

1.4.2. Fedora.

Es el acrónimo de arquitectura digital de repositorio de objetos digitales flexible y extensible. Este repositorio de documentos digitales requiere de la instalación previa del programa Java en el servidor en el que funcionará; además, incluye en su paquete de instalación una base de datos que en caso de que los especialistas decidan no utilizarla, puede sustituirse por otras como Oracle y MySQL. Este software funciona con los sistemas operativos Windows y Unix y sobre servidores Web Apache y Tomcat.

Al ser uno sistema open source gratuito, ofrece a los programadores su código fuente. Se distribuye bajo la Licencia de la Comunidad Educativa, que permite que el programa se propague a todos los que lo requieran y que se hagan modificaciones, siempre que se coloquen en forma visible los términos de esta licencia para que otras personas puedan conocer sobre ella.

El procesamiento de los documentos se realiza según los metadatos asignados por los especialistas en formato Dublin Core. La interfaz de presentación de Fedora es distinta para cada una de sus sesiones, es decir, una para el procesamiento de los documentos y otra para los usuarios finales, que además tendrán la posibilidad de recuperar los contenidos mediante búsquedas en varios índices, previamente declarados por los procesadores o mediante la navegación por las listas de las colecciones. Fedora permite crear colecciones digitales en varios formatos de documentos, como son: texto, imagen, sonido. El lenguaje de la interfaz de trabajo y de presentación a los usuarios, es inglés, aunque es posible configurar, por medio

de la agregación de aplicaciones adicionales que ofrece el sistema, el programa en varios idiomas.(Sarduy, 2006)

Este sistema provee una tecnología escalable para crear, manejar, publicar, compartir y preservar contenidos digitales. Ha sido desarrollado conjuntamente por la Universidad de Cornell y la Universidad de Virginia, basado en sus inicios en el proyecto DARPA y la Fundación Nacional de Ciencia.(Albarado, 2010)

1.4.3. E-Prints.

EPrints es uno de los software para la gestión de repositorios más usado, se distribuye bajo la licencia GNU, lo cual significa que el código fuente es accesible y modificable por cualquier programador, con la condición que las modificaciones se hagan también accesibles públicamente.

Un repositorio creado con esta herramienta, ofrece un conjunto de servicios a todos los miembros de una institución, proporcionando acceso abierto a los datos. Asegura la preservación, enseñanza y colaboración de las publicaciones e investigaciones. Permite subir actualizaciones de un documento en el mismo formato de metadatos y se crea un árbol de versiones del documento posibilitando el acceso a cualquiera de ellos.

Los metadatos ingresados para un documento se pueden reutilizar para la carga rápida de otro documento con iguales características y es totalmente compatible con el protocolo OAI-PMH. El software ofrece la operación de copia o la clonación. Para evitar duplicación de objetos el sistema alerta al usuario en caso de que un objeto existente se vea similar al que se está ingresando.(Belladonna, 2009)

El código está escrito en Perl, utiliza el servidor web Apache y MySQL como base de datos, funciona sobre los sistemas operativos Linux, GNU o Solaris y Windows gracias a un módulo adicional desarrollado en su última versión.(evalEPrints, 2009)

EPrints provee posibilidades de organizar los documentos del repositorio en colecciones. Este sistema puede almacenar cualquier tipo de ficheros que el administrador decida aceptar, incluyendo HTML, Adobe PDF, y PostScript, cada documento puede ser almacenado en distintos formatos. No impone límite alguno

sobre el tamaño de los archivos que se pueden subir, quedando esto determinado entonces únicamente por el sistema de archivos subyacente o el servidor web.(Barrueco, 2010.)

EPrints emplea por defecto Dublin Core para representar los metadatos, sin embargo, permite aplicar cualquier esquema de metadatos y el administrador decide qué campos se asignan a cada tipología documental.

Dentro de las ventajas se tiene:

- Posibilita ingresar nuevos módulos que se desarrollen.
- La interfaz es modificable a las características de la Unidad de Información.
- El proceso de la carga de la descripción de un libro en el sistema es muy fácil e intuitivo de usar.
- Permite la pre catalogación, de manera que sea solo un editor el que autorice que un elemento suba al sistema.
- Posibilita indicar los tipos de documentos que podrá cargar el sistema.
- Permite modificar la estructura de la clasificación que viene predeterminada a conveniencia de la Unidad de Información.
- Realiza búsquedas a partir de otras ya realizadas.
- Permite limpiar el formulario para hacer nuevas búsquedas.

Entre sus desventajas se encuentra que:

- No soporta formatos para objetos complejos como METS, MPEG21 o IMS CP (para el empaquetamiento de objetos de aprendizaje).
- Funciona mejor para documentación textual y no multimedia.(Ailefi, 2011)

1.4.4. DSpace.

DSpace es un conjunto de programas de software diseñado para crear y distribuir colecciones digitales, proporcionando así una nueva forma de organizar y publicar la información a través de Internet. Es un sistema de biblioteca digital diseñado para capturar, almacenar y publicar información en formatos digitales.(Albarado, 2010)

Provee una forma de gestionar materiales y publicaciones resultantes, tanto de la actividad de investigación como de educación, almacenadas en un repositorio que ofrece a los usuarios accesibilidad en todo momento. Creado en colaboración entre la empresa Hewlett-Packard y las bibliotecas del Instituto de Tecnología de Massachusetts, el programa en sus inicios satisfizo las necesidades de estos centros de información pero luego se colocó a disposición de la comunidad internacional, como una herramienta open source, gratuita y de licencia GPL.

Necesita de la aplicación Java para su funcionamiento. Esta herramienta también incluye un sistema de bases de datos relacionales; este repositorio de colecciones digitales puede montarse sobre servidores Web Apache o Tomcat.

La interfaz en ambiente Web varía en dependencia de la persona que la utilice; así, los responsables de la colección tienen una, los administradores del sistema otra y los usuarios finales otra; aunque todas coinciden en que son en idioma inglés. Los usuarios pueden hacer sus búsquedas a partir de los metadatos declarados por los especialistas o simplemente mediante las listas de documentos por campos específicos, como son: autor, título y fecha. (Sarduyz, 2006)

DSpace impone una estructura de organización de la información que pretende ser un reflejo de la estructura de la institución que implementa el sistema.

Cada instancia está compuesta por comunidades, las cuales representan el nivel más alto de la jerarquía de contenido. Una comunidad es una agrupación lógica de una o más colecciones y puede a su vez ser dividida en sub-comunidades

El sistema requiere una dirección de correo electrónico y no permite a múltiples usuarios usar la misma dirección de correo electrónico. Brinda la función de recordatorio de contraseña.

DSpace no impone restricciones sobre formatos de archivos a subir y permite al administrador mantener una lista de formatos soportados por el repositorio. Tampoco impone límite alguno sobre el tamaño de los archivos que se pueden subir, aunque da al administrador la posibilidad de hacerlo mediante edición del archivo de configuración principal. (Belladonna, 2009)

Aunque soporta todo tipo de formatos de documentos, recomienda a los administradores de cada sistema decidir qué formatos aceptar, diferenciando tres categorías de formatos: soportados, conocidos y no soportados. Los documentos se mantienen accesibles con formatos actuales, mientras se desarrollan y actualizan formatos nuevos.

DSpace provee por defecto un esquema de metadatos Dublin Core. Permite configurar múltiples esquemas y seleccionar campos de metadatos de una combinación de esquemas configurados para describir un ítem. (Mendoza, 2009)

La información sobre los esquemas es mantenida en lo que la aplicación llama registro de metadatos. El registro de metadatos puede ser administrado directamente desde la interfaz web permitiendo editar esquemas agregando, quitando o modificando elementos. Cualquier colección de DSpace puede ser importada en Greenstone.

Algunas de las ventajas que ofrece DSpace son:

- El Código es limpio y utiliza los estándares.
- Emplea estándares como Dublin Core, METS, OAI, CNI Handles para la persistencia de las URL.
- El Código se encuentra bien comentariado. Se puede ver lo que el programador ha hecho y se puede modificar o corregir fácilmente.
- Posee una arquitectura en capas, donde la capa de negocios está separada de la de aplicación.
- La utilización del motor de bases de datos Postgresql.
- La base de datos permite las búsquedas normales de SQL y facilita la corrección de problemas.
- Posee identificadores persistentes.
- Es orientado a la preservación.
- Soporta gran cantidad de formatos.
- Permite realizar suscripciones a colecciones.
- Cuenta con suficiente documentación técnica actualizada. (Medina, 2010)

Luego de analizar los sistemas para la gestión de repositorios se concluye que su utilización no soluciona el problema planteado pues estos permiten montar y compartir un repositorio, basado en una estandarización. Y lo que se necesita es leer de diferentes repositorios y a partir de estas lecturas proveer

a la biblioteca de información. Por lo que se implementa un sistema para consultar repositorios, basado en el protocolo OAI-PMH y no un sistema de gestión de repositorios.

1.5. Lenguajes y herramientas informáticas de apoyo para el desarrollo.

Para desarrollar el sistema de indexación de documentos para la Biblioteca Digital “Alma Mater “se hace necesario el uso de un conjunto de tecnologías.

1.5.1. Sistemas de Gestión de Contenidos.

La Gestión de Contenido es un complejo proceso de colección, gestión y publicación de contenido informativo en entorno digital.

Un CMS es una aplicación responsable de la colección, gestión y publicación de bloques de información, llamados componentes de contenido.(Tramullas, 2007)

Son aplicaciones prefabricadas altamente configurables que brindan la posibilidad de manipular contenidos de propósito general, aunque se pueden personalizar todo lo que se quiera. El principal objetivo de los CMS es proveer al desarrollador de una herramienta para la construcción de aplicaciones Web que manipulen contenidos de forma dinámica minimizando la necesidad de conocimientos técnicos en cuanto a programación se refiere.(Guía CMS-Spain, 2004)

Consiste en una interfaz que controla una o varias bases de datos donde se aloja el contenido del sitio. El sistema permite manejar de manera independiente el contenido por una parte y el diseño por otra. Así, es posible manejar el contenido y darle en cualquier momento un diseño distinto al sitio sin tener que darle formato al contenido de nuevo, además de permitir la fácil y controlada publicación en el sitio a varios editores.(Guía CMS-Spain, 2004)

1.5.1.1. CMS Drupal.

Drupal es un sistema de gestión de contenidos que sirve para administrar recursos Web. Es una herramienta de software que se instala en un servidor de información y se usa a través de un cliente o navegador Web.(Tramullas, 2007)

Es un sistema configurable, escrito en PHP, desarrollado y mantenido por una activa comunidad de usuarios. Destaca por la calidad de su código y de las páginas generadas, el respeto de los estándares de la Web, y un énfasis especial en la usabilidad y consistencia de todo el sistema.

Su diseño es especialmente idóneo para construir y gestionar comunidades en Internet. No obstante, su flexibilidad y adaptabilidad, así como la gran cantidad de módulos adicionales disponibles, hace que sea adecuado para realizar diferentes tipos de sitio Web.(Reyero, 2006)

Este sistema de gestión de contenidos es capaz de soportar un amplio abanico de páginas Web, desde simples weblogs a páginas profesionales para empresas, usa una base de datos para guardar los contenidos. El código fuente completo del sistema está bajo licencia GPL y puede por tanto usarse libremente, modificarse y distribuirse.(Linux-Magazine, 2005)

Es un sistema multiusuario, multiplataforma, multilenguaje, extensible y modular. Drupal brinda un buen diseño y un código HTML extremadamente limpio y bien estructurado. Es una buena base en la que puedes integrar todo tipo de módulos. Proporciona gestión de usuarios, clasificación y edición del contenido.(Cruañez, Rey, Viltres)

Drupal provee estructuras para el trabajo en un entorno bibliotecario, como son:

Content Construction Kid (CCK): Permite crear las distintas estructuras de datos para nuevos tipos de contenido.

VIEW: Sirve para la recuperación de información, basado en un asistente para crear las expresiones de búsqueda o selección (tanto en CCK como en VIEW se habla de la familia de módulos, no solo de un módulo).

Faceted Search: Permite hacer búsquedas por facetas sin necesidad de instalar Java.

Bibliography: Manipula bibliografías con la capacidad de importar formatos XML, MARC, Bibtex y otros, maneja diversidad de estilos bibliográficos y se integra con los manejadores de referencias bibliográficas más comunes (endnote, zotero, procite).

Biblio Facets: Integra los dos anteriores, para búsquedas facetadas en las bibliografías.

Book Post: Obtiene información de libros mediante su ISBN, tomada de Open Library Project.

MARC: Importa registros MARC, a partir de ficheros mrc ó mrk.

Millenium Integration: Desarrolla un catálogo automatizado de acceso público en línea (OPAC) contra un sistema propietario Millenium, enlaza con información de Google Books e importa índices de la Biblioteca del Congreso.

OAI2 for CCK: Crea un proveedor de datos OAI para tipos de contenidos creados con CCK. (Fernández, Abreu, 2010)

El sistema de indexación de documentos para la Biblioteca Digital “Alma Mater” se desarrolla como un módulo para un sistema hecho con Drupal, por lo que responde a ésta herramienta. La selección de Drupal para desarrollar el sistema de indexación de documentos condiciona el uso de determinados lenguajes de programación, que emplea el propio CMS.

1.5.2. HTML y CSS.

HTML es un lenguaje de composición de documentos y especificación de ligas de hipertexto que define la sintaxis y coloca instrucciones especiales que no muestra el navegador, aunque sí le indica cómo desplegar el contenido del documento, incluyendo texto, imágenes y otros medios soportados. Indica cómo hacer un documento interactivo a través de ligas especiales de hipertexto, las cuales conectan diferentes documentos y otros recursos de Internet. (Musciano, Kennedy, 1999)

Las hoja de estilo en cascada CSS, proporcionan una manera muy eficiente de aplicar estilos y formato a las páginas Web y a prácticamente cualquiera de sus elementos. Además, permiten definir nuevas clases de los estilos y subclases de las etiquetas HTML o de otras clases. Estas opciones las convierten en una herramienta muy poderosa y muy flexible.

CSS se basa en un código estándar desarrollado por la World Wide Web Consortium (W3C) mediante la cual es posible definir anticipadamente el estilo que tendrá cada etiqueta HTML en la página. (Ruiz, 2002)

1.5.3. XML.

XML es un conjunto de normas que permiten tratar información muy diversa desde muchos puntos de vista y sistemas diferentes, siendo el propio diseñador el encargado de decidir el proceso más adecuado a cada caso y, en consecuencia, XML es un sistema complejo de descripción de información libre y rigurosa. (Montero, 2001)

También es considerado como un lenguaje de etiqueta poderoso y organizado, con muchas aplicaciones. El hecho de poder convertir a una página Web en un documento XML, da a los programadores una flexibilidad muy grande. (Hernán, 2002)

Además de describir los documentos y garantizar la interoperabilidad entre los sistemas, es el medio que establece el protocolo OAI-PMH para la comunicación.

1.5.4. RDF

RDF es una especificación de la W3C para la definición mediante metadatos, generalmente en XML, de los recursos que se pueden encontrar en un sitio. Dublín Core es uno de los ejemplos de RDF usando XML.

Es muy útil en situaciones en las que la información necesita ser procesada por aplicaciones que intercambian información legible por máquina, más que por humanos. Provee una sintaxis basada en XML, llamada RDF/XML, para guardar e intercambiar la información.

Este lenguaje toma un tanto de terminología abstracta y otro tanto de sintaxis XML para definir los documentos, de manera que se puedan escribir programas para procesarlos.

Está cuidadosamente diseñado para tener las siguientes características: Independencia, intercambio, escalabilidad.(Ayllón, 2007)

Este lenguaje especifica una semántica para los datos basados en XML, de manera interoperable y estandarizada.

En RDF la construcción básica es la tripleta: sujeto, propiedad, objeto que se representa como nodos conectados por línea con etiquetas. Los nodos representan recursos y las líneas con etiquetas propiedades de estos recursos .Los tres elementos de una tripleta se representan mediante URIs.(Abián, 2005)

1.5.5. Lenguajes dinámicos.

Un lenguaje dinámico es aquel que en tiempo de ejecución realiza tareas que otros lenguajes hacen en tiempo de compilación, facilitando enormemente la labor del programador y haciendo el trabajo más productivo. Permiten cargar tipos de datos desde estructuras externas, usar datos sin necesidad de definir contratos, disminuir la cantidad de código requerido, evaluar en tiempo real fórmulas y código, además de modificar el comportamiento de un proceso o flujo de trabajo en tiempo real sin necesidad de recompilar.

1.5.5.1. PHP

PHP es un lenguaje dinámico de secuencia de comandos de servidor, diseñado específicamente para la Web. Dentro de una página Web puede incrustar código PHP que se ejecutará cada vez que se visite una página. Su código es interpretado en el servidor Web y genera código HTML y otro contenido que el visitante verá.

Fue concebido en 1994 y es fruto del trabajo de Rasmus Lerdorf. Ha sido adoptado por otras personas de talento y ha experimentado tres transformaciones importantes hasta convertirse en el producto actual. En octubre de 2002, era utilizado por más de nueve millones de dominios de todo el mundo y su número crece rápidamente. Es un producto de código abierto. Las siglas PHP equivalían inicialmente a Personal Home Page, se modificaron de acuerdo con la convención de designación de GNU y ahora equivale a Hipertext Preprocessor. (Wellin, Thomson, 2007)

El CMS Drupal está implementado en PHP y el sistema de indexación de documentos para la Biblioteca Digital “Alma Mater” como un módulo para Drupal, también se implementa en PHP.

1.5.6. Gestores de Base de datos

Un sistema de gestión de bases de datos (SGBD) consiste en una colección de datos interrelacionados y un conjunto de programas que permiten a los usuarios acceder y modificar dichos datos. La colección de datos se denomina base de datos. El primer objetivo de un SGBD es proporcionar un entorno que sea tanto práctico como eficiente de usar en la recuperación y el almacenamiento de la información de la base de datos. Otro de los objetivos principales de un SGBD es proporcionar al usuario una visión abstracta de la información, es decir, el sistema oculta detalles de la forma de almacenar y mantener los datos. (González, 1999)

1.5.6.1. PostgreSQL

PostgreSQL es un sistema de gestión de bases de datos que soporta una gran parte del estándar SQL y ofrece muchas características modernas como consultas complejas, llaves foráneas, vistas, triggers y control de concurrencia multiversión.

Puede ser ampliado por el usuario en muchos aspectos, por ejemplo mediante la adición de nuevos tipos de datos, funciones, operadores. Está patentado bajo licencia libre por lo que puede ser usado, modificado y distribuido por cualquier persona de forma gratuita para cualquier propósito, ya sea privado, comercial o académico. (PostgreSQL.org, 2010)

1.5.6.2. MySQL

MySQL es un sistema para la administración de bases de datos relacional. Las bases de datos permiten almacenar, buscar, ordenar y recuperar datos de forma eficiente. Su servidor controla el acceso a datos para garantizar el uso simultáneo de varios usuarios y proporcionar acceso a dichos datos. Es un servidor multiusuario y de subprocesamiento múltiple. Utiliza SQL, el lenguaje estándar para la consulta de bases de datos utilizado en todo el mundo.

Lleva disponible desde 1996 pero su nacimiento se remonta a 1979. Ha obtenido el galardón Choice Award del Linux Journal Readers en varias ocasiones. MySQL se distribuye bajo una licencia de código abierto en la actualidad, pero también existen licencias comerciales. (Wellin, Thomson, 2007)

El sistema de indexación para la Biblioteca Digital “Alma Mater” es un módulo para Drupal por lo que funciona para cualquier gestor de base de datos en uso.

1.5.7. IDE NetBeans 6.8

NetBeans IDE es un reconocido entorno de desarrollo integrado disponible para Windows, Mac, Linux y Solaris, hecho principalmente para el lenguaje de programación Java, sin embargo permite crear aplicaciones Web con PHP 5 y posee un potente debugger integrado. Es un proyecto de código abierto de gran éxito con una gran base de usuarios y una comunidad en constante crecimiento.

La programación mediante NetBeans se realiza a través de componentes de software modulares, llamados módulos, que están a disposición del usuario, en su página oficial, para conseguir mejoras en las aplicaciones. Además permite importar proyectos completos hechos en Drupal y brinda auto completamiento de código. (netbeans.org, 2010)

1.6. Metodologías de Desarrollo.

Una Metodología de desarrollo es un proceso de software detallado y completo. Se basa en una combinación de los modelos de procesos genéricos. Define artefactos, roles y actividades involucradas, junto con prácticas y técnicas recomendadas.

Una Metodología puede ser estructurada u orientada a objeto y además según su filosofía de desarrollo puede ser tradicional o ágil.(Colectivo de EVA, 2010)

1.6.1. XP

XP es una metodología ágil centrada en potenciar las relaciones interpersonales como clave para el éxito en desarrollo de software, promoviendo el trabajo en equipo, preocupándose por el aprendizaje de los desarrolladores y proporcionando un buen clima de trabajo. Se basa en la retroalimentación continua entre el cliente y el equipo de desarrollo, comunicación fluida entre todos los participantes, simplicidad en las soluciones implementadas y coraje para enfrentar los cambios. Está formado por valores, principios y prácticas. Las prácticas son actividades concretas que un equipo puede realizar día a día, mientras que los valores representan el conocimiento fundamental que sustenta dichas prácticas. Tanto los valores como las prácticas son necesarios, pero hay un gran espacio entre ambos.(Peñalver, 2008)

1.6.2. SXP

Es una metodología compuesta por las metodologías SCRUM y XP que ofrece una estrategia tecnológica, a partir de la introducción de procedimientos ágiles que permitan actualizar los procesos de software para el mejoramiento de la actividad productiva fomentando el desarrollo de la creatividad, aumentando el nivel de preocupación y responsabilidad de los miembros del equipo, ayudando al líder del proyecto a tener un mejor control del mismo. SCRUM es una forma de gestionar un equipo de manera que trabaje de forma eficiente y de tener siempre medidos los progresos, de forma que sepamos por dónde andamos. XP más bien es una metodología encaminada para el desarrollo; consiste en una programación rápida o extrema, cuya particularidad es tener como parte del equipo, al usuario final, pues es uno de los requisitos para llegar al éxito del proyecto.(Peñalver, 2008)

Está especialmente indicada para proyectos de pequeños equipos de trabajo, rápido cambio de requisitos o requisitos imprecisos, muy cambiantes, donde existe un alto riesgo técnico y se orienta a una entrega rápida de resultados y una alta flexibilidad. Esta metodología ayuda a que trabajen todos juntos, en la misma dirección, con un objetivo claro, permitiendo además seguir de forma clara el avance de las tareas a realizar, de forma que los jefes pueden ver día a día cómo progresa el trabajo .

Consta de 4 fases principales:

- **Planificación-Definición:** Donde se establece la visión, se fijan las expectativas y se realiza el aseguramiento del financiamiento del proyecto.
- **Desarrollo:** Es la fase donde se realiza la implementación del sistema hasta que esté listo para ser entregado.
- **Entrega:** Fase donde se pone en marcha el producto desarrollado y se hace la entrega al cliente.
- **Mantenimiento:** En esta se realiza el soporte para el cliente. (Leyva, 2009)

1.6.3. Proceso Unificado de Desarrollo de Software (RUP).

El Proceso Unificado RUP, constituye la metodología estándar más utilizada para el análisis, diseño, implementación y documentación de sistemas orientados a objetos. Proporciona una aproximación disciplinada a la asignación de tareas y responsabilidades. Actúa como modelo y puede ser adaptado y extendido. (Jacobson, 2000)

Características del ciclo de vida de RUP

- Dirigido por casos de uso
- Centrado en la arquitectura.
- Iterativo e Incremental

Fases:

- **Concepción o inicio:** Comprender los requisitos y determinar visión y alcance del proyecto.
- **Elaboración:** Asignar recursos, especificar las características y definir la arquitectura.

- **Construcción:** Implementación, construir el producto operacional.
- **Transición:** Hacerlo operativo para los usuarios, nivel correcto de calidad para entregar.

En RUP se han agrupado las actividades en grupos lógicos definiéndose 9 flujos de trabajo principales. Los 6 primeros son conocidos como flujos de ingeniería y los tres últimos como de apoyo.

Luego de analizar las características de XP, SXP y RUP se decidió utilizar RUP, por ser requerimiento del cliente y adaptarse a una gran diversidad de sistemas de software independientemente del tamaño del proyecto, el tipo de organización y los diferentes niveles de aptitud.

Como lenguaje de modelado se usa UML (Lenguaje Unificado de Modelado) y la herramienta empleada es el Visual Paradigm 6.4.

1.7. Estructuras de datos.

El sistema de indexación de documentos surge como un módulo para la Biblioteca Digital “Alma Mater”. Su función principal es indexar metadatos desde fuentes externas para complementar la creación de contenidos en dicha biblioteca, motivo por el cual necesita estructuras de datos para almacenar los datos indexados.

La biblioteca posee el tipo de contenido documento, y sus respectivas tablas donde es almacenada la información de los documentos. Sin embargo para su correcto funcionamiento el sistema de indexación debe manipular y almacenar información referente a los proveedores de datos para la posterior indexación y la biblioteca no cuenta con estructura alguna donde se almacenen estos datos. Por tanto, como parte del funcionamiento del sistema de indexación es necesario la creación de una estructura de datos para almacenar esta información.

Para la creación de estructuras de datos en Drupal desde un módulo se emplea la función o el hook `schema`, que define la estructura representativa para uno o más tablas con sus llaves relacionadas e índices. La función se plantea nombre del módulo `_schema()`.

En el desarrollo de este capítulo se realiza un estudio detallado de las bibliotecas y su evolución a partir de la aparición de los mecanismos para la comunicación entre sistemas y de la estandarización de la información, destacando el importante papel de los formatos para la descripción de documentos como es el caso de MARC y Dublin Core. Además se analizan los protocolos para la transmisión de metadatos y los sistemas para la gestión de repositorios. Así como los sistemas gestores de base de datos y las metodologías de desarrollo de software de mayor impacto a nivel internacional.

Este análisis se toma como punto de partida en la elección de las herramientas a utilizar en el desarrollo del sistema informático que dará solución a la problemática planteada. Para implementar el módulo se utiliza el CMS Drupal versión 6.17, el lenguaje de programación PHP 5.1 y el IDE de programación NetBeans 6.8. Se emplea RUP como metodología de desarrollo de software y UML para el modelado visual, con la herramienta Visual Paradigm. Como protocolo para la transmisión de metadatos se usa OAI-PMH y Dublin Core como formato de metadatos. No se emplea un SGBD específico pues el módulo funciona con cualquiera que esté usando Drupal.

Capítulo 2. Caracterización del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.

En el presente capítulo se exponen las características que debe poseer el sistema. Se describen los procesos que serán automatizados. Se determinan los requisitos funcionales y no funcionales a tener en cuenta para el desarrollo del sistema, además se elabora el modelo de casos de uso y el modelo de datos.

2.1. Propuesta del sistema.

El sistema propuesto para dar solución al problema planteado cuenta con varias interfaces que permiten agregar fuentes que respondan a la iniciativa OAI, importar documentos de archivo local e importar documentos de Internet propiciando un crecimiento acelerado de los fondos de la biblioteca.

Para acceder al sistema es necesario poseer permisos de administración y su correcto funcionamiento solo es posible con la existencia de determinadas estructuras de datos proporcionas por la biblioteca.

2.2. Modelo de dominio

Después de realizar un estudio de los procesos que se efectúan, se concluye que el negocio estudiado presenta bajo nivel de estructuración, los flujos de información se encuentran difusos y es difícil establecer las reglas de funcionamiento, por lo que se propone realizar un modelo de dominio .

El Modelo de Dominio es una representación visual estática del entorno real objeto del proyecto y se centra en una parte del negocio, la relacionada con el ámbito del proyecto. Es una representación de los conceptos significativos en el dominio del problema e incluye clases de objetos, asociaciones entre clases de objetos y atributos de las clases de objetos. (Sánchez, 2006)

2.2.1. Definición de las entidades y conceptos principales

Biblioteca_Digital: Sistema para acceder a repositorios de universidades y a los artículos de revistas científicas de interés para los distintos programas de formación del Ministerio del Poder Popular para la Educación Universitaria en Venezuela.

Proveedor_de_Servicios: Mecanismo que permite la indexación de datos desde fuentes externas.

Proveedor_de_Datos: Sistemas externos que son encuestados por el proveedor de servicios pues proveen datos que son indexados a la biblioteca.

Sistema_de_Búsqueda: Sistema que permite hacer búsquedas sobre los documentos indexados en las bases de datos.

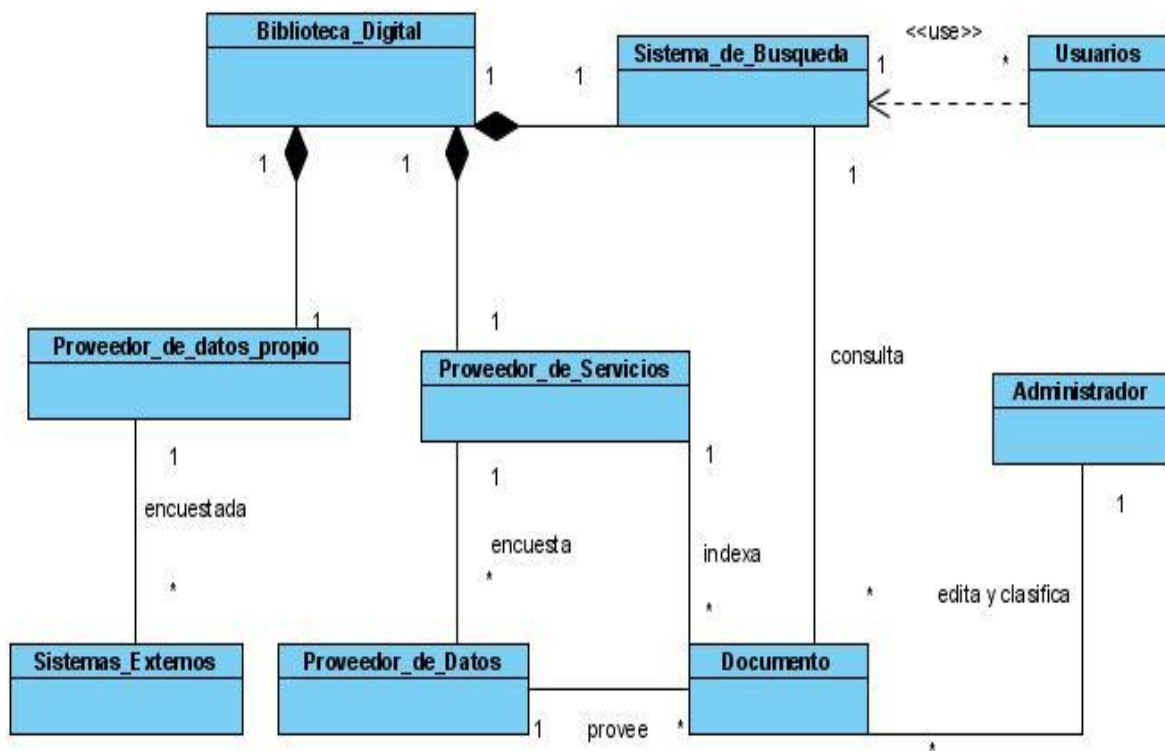
Usuarios: Personas que usan el sistema de búsquedas.

Documento: Es una información estructurada que es almacenada y gestionada por la biblioteca.

Proveedor_de_datos_propio: Mecanismo que le permite a la Biblioteca Digital brindar los metadatos de sus documentos a otros sistemas externos bajo los estándares del protocolo OAI-PMH.

Sistemas_Externos: Sistemas que encuestan al proveedor de datos propio de la biblioteca para recolectar datos.

Administrador: Rol encargado de clasificar y editar los documentos.



Img.1 Modelo de Dominio

2.3. Levantamiento de requisitos

2.3.1. Requisitos funcionales

El sistema debe:

RF1	Crear estructura de datos.
RF2	Eliminar estructura de datos.
RF3	Adicionar fuentes OAI.
RF4	Importar documentos desde ficheros locales.
RF5	Importar documentos desde Internet.
RF6	Leer XML cargado.
RF7	Salvar documentos en la base de datos.
RF8	Chequear permisos de administración.
RF9	Permitir integrarse con Drupal 6.x.

2.3.2. Requisitos no funcionales

Fiabilidad

- El tiempo medio de corrección de errores no debe exceder las 72 horas.
- Se espera un tiempo medio entre fallos de un mes.
- En caso de que el sistema presente alguna falla, los errores se deben mostrar sin detalles de información que pueda comprometer la seguridad e integridad del mismo.

Eficiencia

- El sistema debe demorar como promedio en una transición, de uno (1) a tres (3) segundos aproximadamente.
- Recursos tecnológicos requeridos.

- **Software:**

Servidor apache (versión 2.0).

Intérprete de aplicaciones Web (Navegador).

Sistema Gestor de Bases de Datos PostgreSQL 8.1.4 o MySQL 5.0 o superior.

- **Hardware mínimo para el cliente:**

512 MB de memoria RAM.

20 GB de disco duro.

Procesador Pentium IV.

- **Hardware mínimo para el servidor Web:**

6 GB de memoria RAM.

60 GB de disco duro.

Intel Core Dos Duo o equivalente.

Soporte

- Librerías de xml2array.
- El sistema debe dar la posibilidad de ser mejorado, así como de incorporarle nuevos servicios en caso de ser necesarios.

Interfaz

- Deberá visualizarse de manera correcta en los navegadores más usados en la actualidad, como: Internet Explorer en sus versiones 6.0 y 7.0, Mozilla Firefox, Opera, Netscape, entre otros.

- Uso del protocolo http.

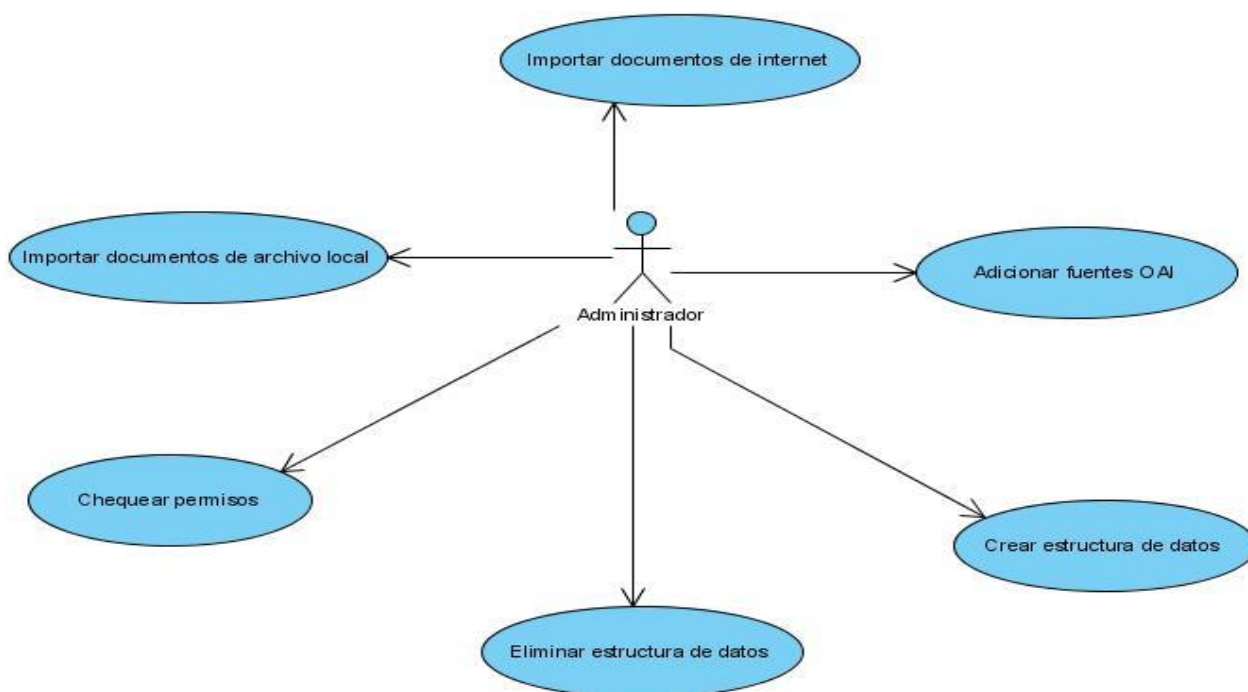
Interfaces de Software

- Bases de datos especializadas adquiridas por compra.
- Repositorios libres de universidades venezolanas.
- Fuentes de acceso abierto.

Interfaces de Comunicación

- La comunicación entre el cliente y el servidor Web será realizada a través del protocolo HTTP.
- La comunicación entre la biblioteca digital y las fuentes de acceso libre se realizarán a través de HTTP request.

2.4. Modelo de Casos de Uso



Img.2 Diagrama de Casos de Uso

2.5. Especificaciones de Casos de Uso

Caso de Uso:	Adicionar fuentes OAI	
Actores:	Administrador	
Resumen:	El caso de uso se inicia cuando el administrador accede a la biblioteca y decide agregar fuentes OAI.	
Precondiciones:	<p>El administrador debe estar autenticado.</p> <p>Debe de existir el tipo de contenido documento.</p> <p>Debe haberse creado la estructura de datos necesaria en la base de datos</p>	
Referencias	RF3.	
Prioridad	Crítico	
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
<ol style="list-style-type: none"> 1. El administrador selecciona la opción Crear contenido del menú Administración de contenido. 2. El administrador selecciona la opción OAI. 3. El administrador selecciona la opción Agregar fuentes OAI. 	<ol style="list-style-type: none"> 4. El sistema muestra el formulario para agregar fuentes OAI. 	

5. El administrador completa los datos y presiona el botón Adicionar.

6. El sistema comprueba que es un proveedor de datos válido y que no ha sido registrado anteriormente, guarda los datos y muestra el mensaje “Se ha registrado la URL fuente correctamente” y se termina el caso de uso.

Prototipo de Interfaz



Flujo Alternativo al paso 6

6.1. El sistema comprueba que no es un proveedor de datos válido y emite un mensaje diciendo: “No cumple con el estándar de metadatos Dublin Core” o verifica que ya ha sido registrado y emite un mensaje diciendo: “Ya existe el proveedor de datos”.

Pos condiciones

Se adicionan fuentes OAI satisfactoriamente.

Caso de Uso:	Crear estructura de datos.	
Actores:	Administrador	
Resumen:	El caso de uso se inicia cuando el administrador accede a la biblioteca y activa el módulo dataprovider_reader.	
Precondiciones:	El administrador debe estar autenticado. Debe de existir el tipo de contenido documento.	
Referencias	RF1.	
Prioridad	Crítico	
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
1. El administrador activa el módulo dataprovider_reader en la sección modules.	2. El sistema crea la tabla source_oai y se termina el caso uso.	
Prototipo de Interfaz		
Pos condiciones	Se crea la estructura de datos satisfactoriamente.	

Caso de Uso:	Eliminar estructura de datos.	
Actores:	Administrador	
Resumen:	El caso de uso se inicia cuando el administrador desinstala el módulo dataprovider_reader.	
Precondiciones:	El administrador debe estar autenticado.	
Referencias	RF2.	
Prioridad	Secundario	

Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
3. El administrador desactiva el módulo dataprovider_reader en la sección modules.	5. El sistema elimina la tabla source_oai, que creo al instalar el módulo y se termina el caso de uso.
4. El administrador desinstala el módulo dataprovider_reader.	
Prototipo de Interfaz	
Pos condiciones	Se elimina la estructura de datos satisfactoriamente.

Caso de Uso:	Importar documentos de Internet.
Actores:	Administrador
Resumen:	El caso de uso se inicia cuando el administrador accede a la biblioteca y desea importar los documentos desde Internet.
Precondiciones:	El administrador debe estar autenticado. Debe de existir el tipo de contenido documento.
Referencias	RF 5.
Prioridad	Crítico
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema

<ol style="list-style-type: none"> 1. El administrador selecciona la opción Crear contenido del menú Administración de contenido. 2. El administrador selecciona la opción OAI. 3. El administrador selecciona la opción colectar fuentes OAI. 	<ol style="list-style-type: none"> 4. El sistema muestra las fuentes disponibles.
<ol style="list-style-type: none"> 5. El administrador selecciona las fuentes de las que desee colectar y presiona el botón Colectar. 	<ol style="list-style-type: none"> 6. El sistema importa los documentos de los proveedores de datos seleccionados directamente de Internet, muestra el mensaje “Se han indexado correctamente todos los documentos” y termina el caso de uso.

Prototipo de Interfaz



Flujo Alternativo al paso 5

<p>5.1. El administrador no selecciona ninguna fuente y presiona el botón Colectar.</p>	<p>6.1. El sistema muestra en mensaje “Debe seleccionar al menos una fuente”.</p>
<p>Pos condiciones</p>	<p>Se importan documentos de Internet satisfactoriamente.</p>

Caso de Uso:	<p>Importar documentos de archivo local.</p>
Actores:	<p>Administrador</p>
Resumen:	<p>El caso de uso se inicia cuando el administrador accede a la biblioteca para importar documento desde un archivo local.</p>
Precondiciones:	<p>El administrador debe estar autenticado. Debe de existir el tipo de contenido documento.</p>
Referencias	<p>RF4.</p>
Prioridad	<p>Crítico.</p>
<p>Flujo Normal de Eventos</p>	
Acción del Actor	Respuesta del Sistema
<ol style="list-style-type: none"> 1. El administrador selecciona la opción Crear contenido del menú Administración de contenido. 2. El administrador selecciona la opción OAI. 3. El administrador selecciona la opción desde fichero local. 	<ol style="list-style-type: none"> 4. El sistema muestra un formulario en el que se debe llenar los datos de la fuente, el tipo de recurso y subir el XML.

5. El administrador llena los datos correctamente y presiona el botón Importar.

6. El sistema importa los documentos, muestra el mensaje “Se ha realizado la operación correctamente” y se termina el caso de uso.

Prototipo de Interfaz



Flujo Alternativo al paso 5

5.1 El administrador llena los datos incorrectamente y presiona el botón Importar.

6.1 El sistema muestra el mensaje. "Ha ocurrido un error en el campo XML"

Pos condiciones

Se importan documentos de archivo local satisfactoriamente.

Caso de Uso:	Chequear permisos.			
Actores:	Administrador			
Resumen:	El caso de uso se inicia cuando el administrador accede a la biblioteca.			
Precondiciones:	El administrador debe estar autenticado.			
Referencias	RF8			
Prioridad	Crítico			
Flujo Normal de Eventos				
Acción del Actor		Respuesta del Sistema		
1. El administrador otorga el permiso para Gestionar Fuente OAI.		2. El sistema verifica que un determinado rol tenga el permiso a Gestionar Fuente OAI para permitirle acceder a las opciones del menú. Termina el caso de uso		
Prototipo de Interfaz				
Permiso	usuario anónimo	usuario autenticado	Administrador	Editor
módulo dataprovider_reader				
Gestionar Fuente OAI	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Pos condiciones	Se chequean los permisos satisfactoriamente.			

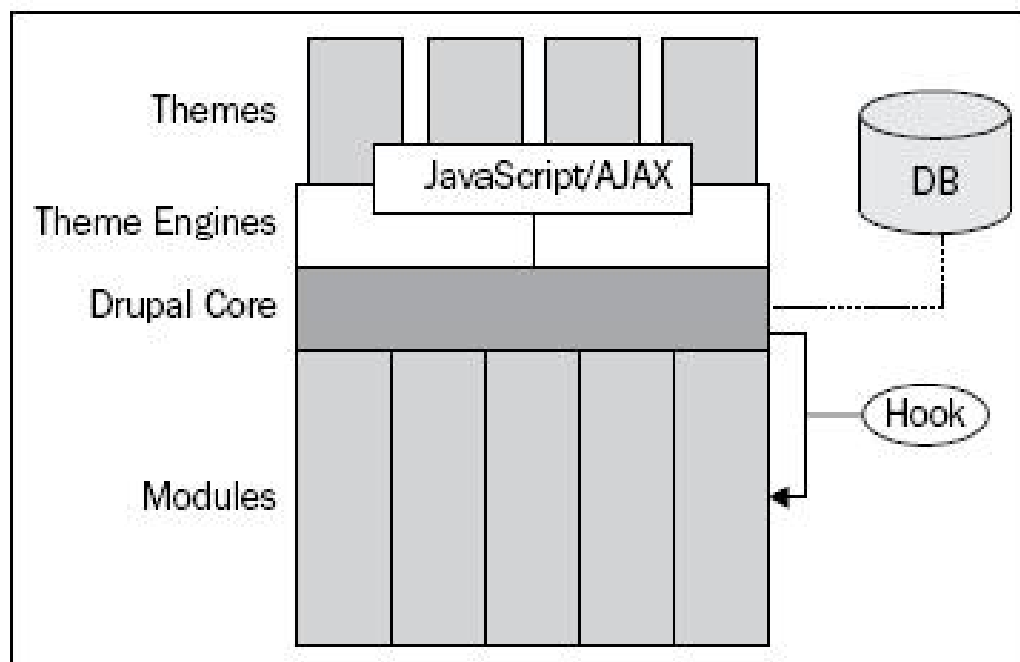
2.6. Modelo de diseño

El propósito del diseño es especificar una solución que trabaje y pueda ser fácilmente convertida en un código fuente y construir una arquitectura simple y fácilmente extensible. (Mendoza, 2008)

Para la implementación del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater” se utiliza el CMS Drupal y por tanto el patrón de arquitectura usado es el Modelo Vista Controlador (MVC).

El Modelo Vista Controlador separa el código del programa en tres capas, dependiendo de su naturaleza. La lógica relacionada con los datos se incluye en el **modelo**, el código de la presentación en la **vista** y la lógica de la aplicación en el **controlador**. (Colectivo de EVA 2010)

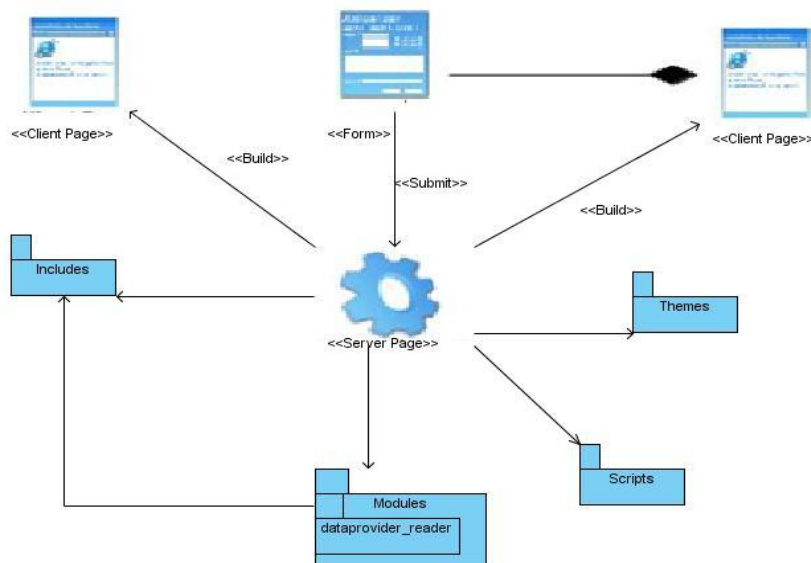
En la figura se observa cómo está estructurado en capas el CMS Drupal, donde el modelo son las clases de conexión a la base de datos (DB), la vista son las clases correspondientes a los temas, motores de temas (Themes, Themes Engines) y JavaScript/Ajax. El controlador son las clases del núcleo de Drupal y de los módulos (Drupal Core y Modules). Además están los ganchos (Hook) que son funciones que facilitan la comunicación entre el núcleo de Drupal y los módulos.



Im g.3 Representación de la arquitectura de Drupal en tres capas.(Butcher, 2008)

2.6.1. Diagramas de clases del diseño

Para la comprensión de los diagramas de diseño Web se modela primeramente el diagrama general del Drupal y luego los diagramas correspondientes al módulo `dataprovider_reader`. El diagrama general muestra el funcionamiento del CMS Drupal con sus clases y módulos.



Img.4 Diagramas de clases del diseño de Drupal

Drupal depende de un grupo de clases o módulos que permiten a los usuarios finales interactuar con la aplicación, estos están estructurados de acuerdo a sus funcionalidades y se modelan en el diagrama como paquetes:

Modules: Incluye todos los módulos, que permiten las distintas funcionalidades del CMS.

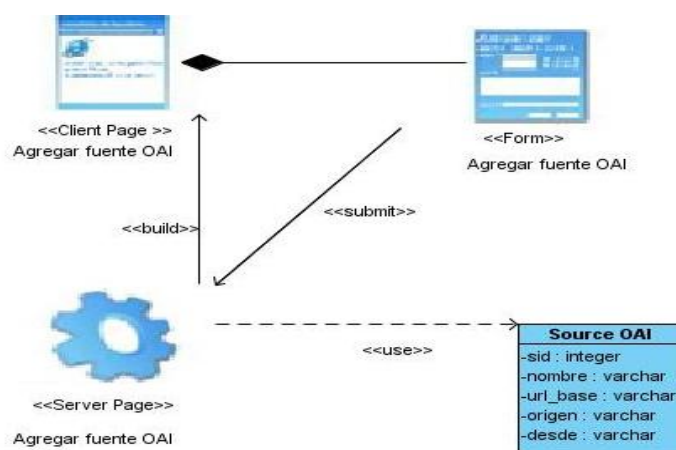
Includes: Contiene un conjunto de ficheros indispensables para el funcionamiento, como son las conexiones a base datos del sistema, APIs generales del CMS utilizadas por los módulos.

Themes: Incluye el motor que genera la presentación o interfaz del CMS, incluye las distintas plantillas de temas.

Scripts: Contiene los ficheros correspondientes al manejo visual del sistema, CSS y JavaScript.

El paquete desarrollado es `datapreader_reader` que se incluye dentro del paquete **modules** para habilitar su funcionamiento.

Las páginas que genera Drupal pueden tener o no formularios, en dependencia de lo que se desee. Es importante señalar que este CMS contiene una única página servidora que a través de funciones o ganchos (hook) se conecta con los módulos, a su vez los módulos se conectan a la base datos a través de los includes, para finalmente mostrar la información requerida.



Img.5 Diagrama de clases del diseño para el CU Adicionar Fuentes OAI

2.7. Modelo de datos

Un modelo de datos describe la representación lógica y física de los datos persistentes de la base de datos. Básicamente consiste en una descripción de un contenedor de datos donde se guarda la información, así como los métodos a almacenar y recuperar información de dichos contenedores. Consiste en objetos, atributos y relaciones.

Drupal posee varias tablas que emplea para su funcionamiento, en el presente modelo de datos se representan las tablas con las que interactúa el módulo `dataproviver_reader` y las que son creadas como parte del funcionamiento del mismo.



Img.6 Modelo de datos

Con la realización de este capítulo se trazan las bases para el desarrollo del sistema pues son descritas las características del módulo a través de los artefactos generados. Se elabora el modelo de dominio que representa los conceptos significativos en el dominio del problema, se plasman los requisitos funcionales y no funcionales a tener en cuenta. Se construye el diagrama de casos de Uso obteniendo una idea general de las funcionalidades que debe cumplir la aplicación. Además se describe y representa el modelo de datos al igual que el modelo de diseño, lo que permite obtener una visión avanzada del sistema.

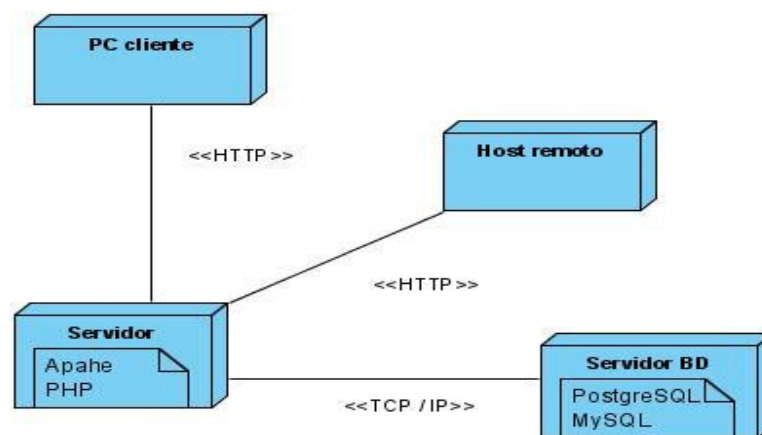
Capítulo 3. Implementación y pruebas del sistema de indexación de documentos para la Biblioteca Digital “Alma Mater”.

En este capítulo se describen los elementos necesarios para la implementación, partiendo del resultado obtenido del diseño. Se muestra la distribución del sistema en nodos mediante el diagrama de despliegue y la organización de los componentes y las relaciones lógicas entre ellos a través del diagrama de componentes, quedando así conformado el modelo de implementación.

Se abordan los métodos de pruebas usados en el desarrollo de la aplicación, describiendo las pruebas de aceptación con el objetivo de asegurar que el sistema cumpla con los requisitos definidos y tenga la calidad requerida.

3.1. Diagrama de Despliegue

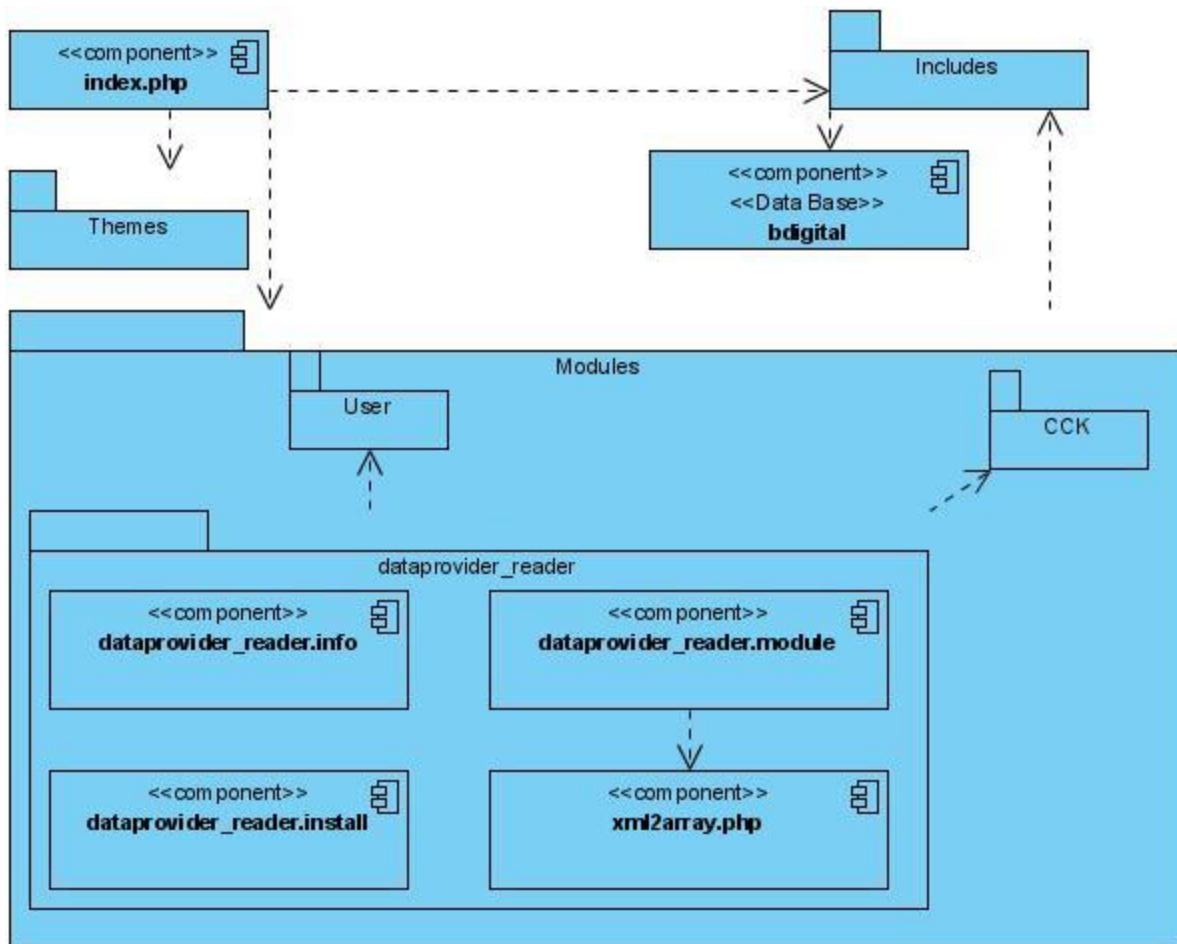
El diagrama de despliegue describe como una aplicación se despliega a través de una infraestructura y muestra la configuración ó relaciones físicas de los nodos que participan en la ejecución y de los componentes hardware y software que residen en ellos, contiene: nodos que son servidores o procesadores y dispositivos, además de relaciones de dependencia y asociación.(Vilas, 2001)



Img.7 Diagrama de despliegue

3.2. Diagrama de Componentes

Los Diagramas de Componentes modelan la vista estática de un sistema. Se representan como un grafo de componentes software unidos por medio de relaciones de dependencia. En un Diagrama de Componentes se tienen en consideración los requisitos relacionados con la facilidad de desarrollo, la gestión del software, la reutilización y las restricciones impuestas por los lenguajes de programación y las herramientas utilizadas en el desarrollo. (Colectivo de EVA, 2010)



Img.8 Diagrama de componentes

Index.php	Es el núcleo de Drupal, es el que inicia el sistema.
Themes	Se encuentra toda la capa de presentación y diseño.
Modules	Se encuentran todos los módulos de Drupal.
Includes	Se encuentran funciones y scripts necesarios para el funcionamiento de los módulos y el sistema Drupal, incluyendo las APIs de conexión a base de datos.

3.3. Pruebas

Se traza una estrategia de prueba que incluye el desarrollo de las pruebas de forma manual, utilizando el método de prueba de caja negra que se refiere a las pruebas que se llevan a cabo sobre la interfaz del software, donde los casos de prueba pretenden demostrar que las funciones del software son operativas, que la entrada se acepta de forma adecuada y que se produce un resultado correcto, así como que la integridad de la información externa se mantiene.

Se realizan pruebas de integración para asegurar que los componentes en el modelo de implementación operen correctamente cuando son combinados en la ejecución de un caso de uso y pruebas de aceptación para verificar que el software está listo y que puede ser usado por usuarios finales para ejecutar aquellas funciones y tareas para las cuales fue construido. El tipo de prueba que se realiza es de funcionalidad.

3.3.1. Caso de Prueba para el caso de uso: Adicionar fuentes OAI

Escenario	URL Base	Respuesta del Sistema	Resultado de la Prueba	Flujo Central
Agregar fuente OAI satisfactoriamente.	http://www.scielo.br/oai/scielo-oai.php	Se agrega el proveedor correctamente y el sistema muestra el mensaje "Se ha realizado la operación correctamente."	Satisfactorio.	<ol style="list-style-type: none"> 1. Acceder a la opción Agregar fuentes OAI del menú Administración de contenido. 2. Completar Formulario. 3. Enviar los datos
Agregar fuente OAI insatisfactoriamente.		El sistema muestra un mensaje diciendo que "El campo URL Base es obligatorio."	Satisfactorio.	
	http://10.33.14.240:5901/sitios/otro/add/enlace.doc	El sistema muestra un mensaje diciendo que "No cumple con el estándar de metadatos Dublin Core."		

3.3.2. Caso de Prueba para el caso de uso: Importar documento de archivo local.

Escenario	Fuente	Tipo recurso	XML	Respuesta del Sistema	Resultado de la Prueba	Flujo Central
Importar documento de archivo local satisfactoriamente	Universidad Central de Venezuela.	libro	Archivo XML con la estructura correcta.	El sistema muestra el mensaje "Se ha realizado la operación"	Satisfactorio.	1. Seleccionar la opción Crear contenido del menú Administración de contenido y luego la opción OAI.
	Universidad Central de Venezuela.		Archivo XML con la estructura correcta.	El sistema muestra el mensaje "Se ha realizado la operación correctamente".		
Importar documento de archivo local insatisfactoriamente		Tesis	Archivo XML con la estructura correcta.	El sistema muestra el mensaje "El campo Fuente es obligatorio."	Satisfactorio.	2. Selecciona la opción desde fichero local. 3. Llenar los datos y presionar el botón Importar.
	Universidad Central de Venezuela.	Artículo		El sistema muestra el mensaje "Ha ocurrido un error en el campo XML."		
	Universidad Central de Venezuela.		Archivo XML con la estructura incorrecta.	El sistema muestra el mensaje "Ha ocurrido un error en el campo XML."		

3.3.3. Caso de Prueba para el caso de uso: Importar documentos de Internet.

Escenario	Fuente	Respuesta del Sistema	Resultado de la Prueba	Flujo Central
Importar documentos de Internet satisfactoriamente	El administrador selecciona al menos una fuente del listado.	El sistema muestra el mensaje "Se han indexado correctamente todos los documentos"	Satisfactorio	<ol style="list-style-type: none"> 1. Seleccionar la opción Crear contenido del menú Administración de contenido y luego la opción OAI. 2. Selecciona la opción desde fichero local. 3. El sistema muestra las fuentes disponibles e importa los documentos de los proveedores de datos seleccionados directamente de Internet.
Importar documentos de Internet insatisfactoriamente	El administrador no selecciona ninguna fuente para coleccionar datos.	El sistema muestra el mensaje "Debe seleccionar al menos una fuente."	Satisfactorio.	

3.3.4. Caso de Prueba de integración.

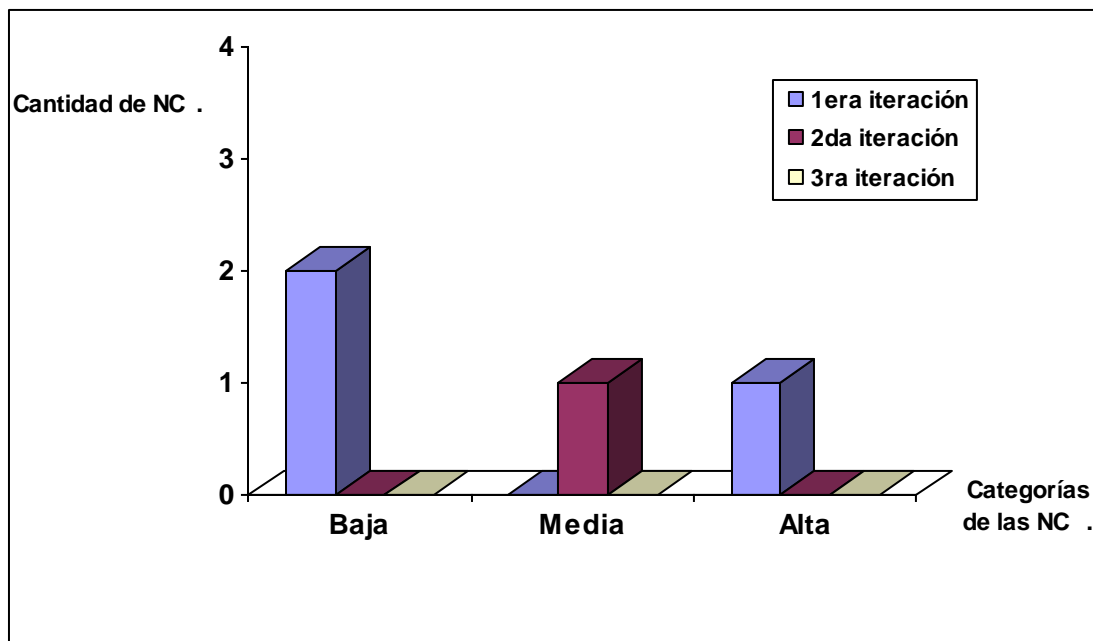
Escenario	XML de entrada	Respuesta del Sistema	Resultado de la Prueba
Importar XML satisfactoriamente	El módulo recibe el XML en respuesta a las peticiones realizadas a los proveedores de datos.	La biblioteca almacena correctamente los datos del XML en la base de datos.	Satisfactorio.
Importar XML insatisfactoriamente	El módulo recibe el XML en respuesta a las peticiones realizadas a los proveedores de datos.	La biblioteca no permite almacenar los datos del XML en la base de datos.	Satisfactorio.

A continuación se muestra el listado de no conformidades que fue generado durante el proceso de pruebas en el que se identifica el elemento afectado así como la descripción de las no conformidades y la ubicación de la misma. También se especifica la etapa en que fueron detectadas. Para identificar el grado de complejidad de estas se clasificaron en tres categorías: alta, media y baja.

No	Elemento	No Conformidad	Ubicación de la No Conformidad	Etapas de detección	Categoría
1	Aplicación	En el mensaje "No cumple con el estándar de metadatos OAI_DC" es poco sugerente.	En la funcionalidad Adicionar Fuentes OAI.	1ra iteración	Baja
2	Aplicación	Falta de ortografía en el mensaje "Se ha realizado la operacion correctamente".	En la funcionalidad Adicionar Fuentes OAI.	1ra iteración	Baja
3	Aplicación	No se valida que el campo fuente es obligatorio.	En la funcionalidad importar documentos de archivo local.	1ra iteración	Alta
4	Aplicación	El sistema no muestra un mensaje si no se selecciona ninguna fuente para indexar documentos.	En la funcionalidad Adicionar Fuentes OAI.	2da iteración	Media

Al concluir cada iteración de las pruebas se analizaron las no conformidades encontradas. Las pruebas se realizaron de forma iterativa e incremental, comprobando en cada iteración la corrección de los errores detectados en la iteración anterior, lo que contribuyó a mejorar la calidad y eficiencia del software.

A continuación se muestra un resumen en forma de gráfico teniendo en cuenta la cantidad de no conformidades encontradas (eje vertical de la figura), las categorización de los errores (eje horizontal de la gráfica) y las iteraciones de prueba realizadas.



Img.9 Resumen general de las pruebas.

En este capítulo se elabora el modelo de implementación que permite obtener una visión general de los componentes a implementar, facilitando que el proceso de implementación se realizara de forma más simple y con una mayor calidad. Queda representado como se despliega la aplicación a través del diagrama de despliegue. Además se construye la estrategia de prueba que incluye el desarrollo de las pruebas de integración y aceptación, permitiendo constatar que el módulo se integra correctamente a la biblioteca y que las funcionalidades implementadas se corresponden con los requisitos establecidos por el cliente.

Conclusiones

El estudio de los sistemas de gestión de contenidos especializados en repositorios digitales, arrojó que su utilización no soluciona el problema planteado, pues estos solo permiten montar y compartir un repositorio y no indexar documentos de **fuentes externas**.

A partir de un estudio de los protocolos más usados para el intercambio de información entre sistemas de tipo biblioteca se selecciona el protocolo OAI-PMH por ser empleado por las fuentes de las que se nutre la Biblioteca Digital “Alma Mater”.

La utilización del protocolo OAI-PMH conlleva al uso del lenguaje XML y del formato de representación de metadatos Dublin Core, debido a que estos, especifican la forma de comunicación del protocolo.

La descripción detallada de los casos de uso permite contar con una guía para la correcta implementación del sistema, asegurando un entendimiento claro por parte de los usuarios.

La implementación del módulo `datapvider_reader` dota a la Biblioteca Digital “Alma Mater” de una herramienta flexible y configurable para importar documentos desde fuentes externas.

La posibilidad de contar con documentos indexados desde varias **fuentes externas** posibilita un crecimiento acelerado de los fondos de la biblioteca y la familiarización del usuario con una única interfaz de búsqueda.

La aplicación de pruebas de integración y aceptación permite demostrar que el sistema de indexación de documentos desarrollado se integra correctamente a la biblioteca para la cual fue creado y que funciona según las funcionalidades definidas.

Recomendaciones

Desplegar el sistema de indexación, para validar su uso y aceptación.

La publicación de este documento por parte de la universidad como consulta y guía para el personal que proporcionará mantenimiento al sistema.

Glosario de Términos

API - Acrónimo de Application Programming Interface (interfaz de programación de aplicaciones).

ARPANET - Acrónimo de la expresión en inglés (Advanced Research Projects Agency Network - Red de la Agencia de Proyectos de Investigación Avanzada).

CCK - Content Construction Kit (kit de construcción de contenido).

DARPA - Acrónimo de la expresión en inglés Defense Advanced Research Projects Agency (Agencia de Investigación de Proyectos Avanzados de Defensa).

Debugger – Depurador en español, es un programa usado para probar y eliminar los errores de otros programas.

Dublin Core - Forma abreviada para el Dublin Metadata Core Element Set (Conjunto básico de elementos de metadatos).

Fedora - Acrónimo de Flexible Extensible Digital Object Repository Architecture (Arquitectura digital de repositorio de objetos digitales flexible y extensible).

Fuentes externas: Universidades e Instituciones de todo el mundo que responden a la Iniciativa de Archivos Abiertos (OAI) y proveen los metadatos de un gran número de materiales digitales y publicaciones electrónicas: tesis en línea, informes, reportes, artículos e imágenes, haciendo uso del protocolo OAI-PMH.

GET/POST - Métodos para enviar los datos de los formularios desde el navegador al servidor Web.

GPL - Acrónimo de General Public License (Licencia Pública General de GNU).

MARC - Un conjunto de estándares para manipular datos bibliográficos legibles por ordenador.

MG/MGPP - Motores de búsqueda.

OPAC- Acrónimo de (Online public access catalog) es un catálogo automatizado de acceso público en línea.

Protocolo - Conjunto de reglas, secuencias, formatos de mensajes y procedimientos bien detallados que posibilitan la transferencia de datos entre dos o más sistemas de computación.

TCP/IP - Familia de protocolos Esta familia que se han establecido como estándares de Internet.

Triggers - Bloque asociado a una tabla, que se ejecuta como consecuencia de una determinada instrucción SQL.

UCLM - Universidad de Castilla-La Mancha.

URI - Acrónimo de Uniform Resource Identifier (identificador uniforme de recurso).

World Wide Web - Sistema de distribución de información basado en hipertexto, accesibles a través de Internet.

Weblogs - Página web de fácil actualización para publicar contenido.

W3C - Comunidad internacional que desarrolla estándares Web.

Z39.50 - Protocolo para la recuperación de información basado en la estructura cliente/servidor.

Referencias Bibliográficas

- Abián, Miguel Ángel. EL FUTURO DE LA WEB XML, RDF/RDFS, ontologías y la Web semántica virtual. 2005. [citado 15 noviembre 2010]. Available from World Wide Web: <http://www.javahispano.org/licencias/>
- Ailefi. Eprints – Sistema de Gestión Documental. 2011. [citado 19 junio 2011]. Available from World Wide Web: <http://ailefi.net.ve/?p=88>.
- Arias, José. Análisis para la creación y desarrollo de la biblioteca digital de Colombia. 2008. Available from World Wide Web: <http://www.tesisenred.net/bitstream/handle/10803/10913/AriasOrdonez.pdf?sequence=1>.
- Atallings, William .Comunicaciones y Redes de Computadoras. 6 Prentice Hall, 2000 Available from World Wide Web: <ftp://10.0.0.22/estudiantes/3ro/Tele1/Comunicaciones%20y%20redes/20de/20computadores.pdf>.
- Ayllón Bonet, Julio César. Metadatos y documentos XML/RDF para recuperación. 2007. [citado 9 diciembre 2010]. Available from World Wide Web: <http://metadatos-xml-rdf.awardspace.com/rdf.html>.
- Belladonna. Evaluación de software para organización de recursos digitales: Dspace. 2009. Available from World Wide Web: http://ricabib.cab.cnea.gov.ar/90/1/evalDSpace_v01.pdf.
- Belladonna. Repositorio Institucional del CAB-IB. Evaluación de software para organización de recursos digitales: EPrints 3.1 2009. [citado 19 junio 2011]. Available from World Wide Web: <http://ricabib.cab.cnea.gov.ar/91/>.
- Butcher, Matt. Learning Drupal 6 Module Development. Editorial Team Leader, 2008.
- Chuck Musciano, Bill Kennedy. HTML La guía completa. 2da [Mexico]: McGRAW-HILL, 1999.
- Colectivo del EVA. EVA. 2010. Available from World Wide Web: <http://eva.uci.cu>.
- Cruañez, Roberkys Martín, Rey, Derick Enriquez, Viltres Sala, Hubert. Trabajo Investigativo sobre Sistemas de Gestión de Contenido. CMS Drupal.
- Cyranek, Günther, y UNESCO Montevideo. Greenstone: Un software libre de código abierto para la construcción de bibliotecas digitales. 2010 Available from World Wide Web: <http://www.unesco.org/uy/ci/fileadmin/comunicacion-informacion/Greenstone-paraweb.pdf>.

- Dempsey L, Heery R. ESIRE: Development of a European service for information on research and education, UKOLN Metadata Group. Available from World Wide Web: <<http://www.ukoln.ac.uk/metadata/desire/overview/overview.pdf>>.
- Duran Gil, Carmen L. UCAB, Dirección de postgrados. Sistemas de informatización: Protocolos. 2007. Available from World Wide Web: <<http://www.fortunecity.es/imaginapoder/nada/617/PT111.htm>>.
- Edemi Martínez, Lyz. Implementación de un sistema de información para el manejo de expedientes clínicos utilizando el estándar HL7 en el hospital regional universitario. 2004. [citado 18 junio 2011]. Available from World Wide Web: <http://digeset.ucol.mx/tesis_posgrado/Pdf/Lyz_Edemi_Martinez_Angulo.pdf>.
- Faba Pérez, Cristina. La nueva gestión en las bibliotecas virtuales. 2004. Available from World Wide Web: <<http://www.aab.es/pdfs/baab74/74a2.pdf>>.
- Fernández Fernández , Maikel Manuel. Abreu Bartomeo, Yanedi .DESARROLLO DE UNA BIBLIOTECA DIGITAL CON EL USO DEL SISTEMA DE GESTIÓN DE CONTENIDOS DRUPAL.2010.
- Fernández Vilas, Ana. Diagrama de Despliegue. 2001. Available from World Wide Web: <<http://tvdi.det.uvigo.es/~avilas/UML/node50.html>>.
- Guía CMS-Spain. GUIAS-IT.COM. 2004. [citado 11 enero 2011]. Available from World Wide Web: <<http://www.ecm-spain.com/interior.asp?IdItem=1851>>.
- Guzmán, Teresa. La UBV y la municipalización de la educación superior. febrero 2005. [citado 12 noviembre 2010]. Available from World Wide Web :<<http://www.aporrea.org/educacion/a11969.html>>.
- Kuang, Hwei. Finding the needle: controlled vocabularies, resource discovery, and Dublin Core. Library collections, acquisitions, and technical services. 2000.
- Linux-magazine. Drupal, el gestor de contenido_ para portales. 2005.
- Marshall, C. Making Metadata: a study of metadata creation for a mixed physical-digital collection. International conference on digital libraries, Pittsburgh. 1998.
- Medina, Sandra. Resumen de requerimientos e instalación de Dspace. 2010. [citado 19 junio 2011]. Available from World Wide Web: <<http://es.scribd.com/doc/53889108/52>>.

- Mendoza Navarro, Javier. Diseño del sistema de tarjeta de Crédito com UML. 2008. Available from World Wide Web: <http://sisbib.unmsm.edu.pe/bibvirtualdata/tesis/basic/mendoza_nj/cap5.pdf>.
- Mendoza, Tejerina. Analisis de DSpace como repositorio institucional. 2009. [citado 19 junio 2011]. Available from World Wide Web: <<http://es.scribd.com/doc/48250416/Analisis-de-DSpace-como-repositorio-institucional>>.
- Netbeans.org. Welcome to NetBeans. 2010. [citado 3 febrero 2011]. Available from World Wide Web: <<http://netbeans.org/>>.
- Ortiz. Virginia. Repiso Jiménez. Nuevas perspectivas para la Catalogación: Metadatos versus MARC. 1999. Available from world wide Web: <<http://redc.revistas.csic.es>>.
- Pérez, Dora. La biblioteca digital. 2007 Available from World Wide Web: <<http://www.uoc.edu/Web/esp/articulos/La_biblioteca_digital.htm>>.
- Plaza, Giraldo. Diseño e implementación de una ontología para el mercado semántico de objetos digitales en bibliotecas digitales 2003.
- Reyero, José. Sobre Drupal | Comunidad de usuarios de Drupal. 2006. [citado 11 enero 2011]. Available from World Wide Web: <<http://drupal.org.es/drupal>>.
- Sánchez Díaz, Marlery. Bibliotecas electrónicas, digitales y virtuales: tres entidades por definir. Vol 10, 2002. [citado 4 noviembre 2010]. Available from World Wide Web: <http://bvs.sld.cu/revistas/aci/vol10_6_02/aci05602.htm>.
- Sánchez, Martha. Definición del modelo del negocio y del dominio utilizando Razonamiento Basado en Casos. 2006. Available from World Wide Web: <<http://www.inf.udec.cl/~revista/ediciones/edicion8/Rbc.pdf>>.
- Scott Hommel. Convenciones de código para el lenguaje de programación JAVA™. 1999. Sun Microsystems Inc.
- Sené, María Luisa. Los metadatos y su lugar en la arena internacional. [Prado y San José, Capitolio Nacional Ciudad de La Habana 862-6501 Cuba]: IDICT, 2009 Available from World Wide Web: <<http://www.bibliociencias.cu/gsd/collect/eventos/index/assoc/HASH01b0/88efe511.dir/doc.pdf>>.
- Setién Quesada, Emilio. El acceso a las bibliotecas y la evolución económica social de la humanidad.

2005.

Sola-Martínez. Greenstone: distribución de colecciones digitales. 2008. [citado 18 junio 2011]. Available from World Wide Web: <<http://www.ub.edu/bid/21/sola2.htm>>.

Tramullas Saz, Jesús. Bibliotecas digitales: una revisión de conceptos y técnicas. 2002. Available from World Wide Web: <<http://tramullas.com/papers/bidipe.pdf>>.

Tramullas, Jesús. Gestión de contenidos con Drupal. 2007.

Tramullas, Jesús. Greenstone ver. 2.51. 2008. Available from World Wide Web: <<http://tramullas.com/pdf/greenst04.pdf>>.

Wellin, Luke, Thomson, Laura. Programación: Desarrollo web con PHP y Mysql. 2007 Available from World Wide Web: <<http://bibliodoc.uci.cu/pdf/reg02819.pdf>>.

Bibliografía

- Abián, Miguel Ángel .EL FUTURO DE LA WEB XML, RDF/RDFS, ontologías y la Web semántica. virtual.2005. [citado 15 noviembre 2010]. Available from World Wide Web: <http://www.javahispano.org/licencias/>
- ACIMED .Los repositorios institucionales: análisis de la situación internacional y principios generales para Cuba. 2007. [citado 10 enero 2011]. Available from World Wide Web: http://scielo.sld.cu/scielo.php?pid=S1024-94352007001200006&script=sci_arttext.
- Acuña Soto Mateo. Repositorios de objetos educativos. [citado 9 enero 2011]. Available from World Wide Web: <http://www.slideshare.net/ac.sotom/repositorios-de-objetos-educativos-acua-soto-mateo>.
- ADLRepositoryTIR.pdf. 2002. [citado 9 enero 2011]. Available from World Wide Web: <http://xml.coverpages.org/ADLRepositoryTIR.pdf>.
- Archivos en la era digital: problema (y solución) de los recursos electrónicos.pdf. [citado 10 enero 2011]. Available from World Wide Web: <http://www.elprofesionaldelainformacion.com/contenidos/2005/noviembre/2.pdf>.
- A working definition of digital library [1998]. [citado 30 octubre 2010]. Available from World Wide Web: <http://www.diglib.org/about/dldefinition.htm>.
- Barité M. Diccionario de Organización y Representación del Conocimiento. KODictionary. 2005. [citado 28 Febrero 2011]. Available from World Wide Web: [http://www.eubca.edu.uy/diccionario/letra_r.htm#Le tra%20A](http://www.eubca.edu.uy/diccionario/letra_r.htm#Le%20tra%20A).
- Barton, Mary R. y Waters, Margaret M. Cómo crear un Repositorio Institucional.pdf. [citado 28 Febrero 2011]. Available from World Wide Web: <http://www.recolecta.net/buscador/documentos/mit.pdf>.
- Barrionuevo, Leticia. Repositorio || Repository - Glossarium-BITri. 2009. [citado 9 enero 2011]. Available from world wide web: <http://glossarium.bitrum.unileon.es/Home/repositorio>.
- Barrueco, José Manuel. OAI-PMH: Protocolo para la transmisión de contenidos en Internet.2010.
- Beitra Oliva, Elena Margarita Selección de recursos disponibles en Internet para el desarrollo de colecciones de la biblioteca virtual. [citado 15 noviembre 2010]. Available from World Wide Web:

<http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S156247302008000200007&lng=es&nr=iso>.

BNJM. Biblioteca Nacional José Martí - Portal del Sistema. [citado 10 enero 2011]. Available from World Wide Web: <<http://www.bnjm.cu/salas.htm>>.

Cabral, Brenda. Ramírez, Reyna. Figueroa, Hugo. - La biblioteca digital.pdf. [citado 8 de enero 2011]. Available from World Wide Web: <http://ru.ffyl.unam.mx:8080/jspui/bitstream/10391/949/1/Brenda_Cabral_Hugo_Figueroa_La_biblioteca_digital.pdf>.

CLACSO. Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe de la Red de Centros Miembros de CLACSO - Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe de la red CLACSO. [citado 10 enero 2011]. Available from World Wide Web: <<http://www.biblioteca.clacso.edu.ar/>>.

Clifford A. Lynch. Association of Research Libraries: Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. 2003. [citado 9 enero 2011]. Available from World Wide Web: <<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>.

CUEGreenstone. Comunidad de usuarios Greenstone en Español. [citado 27 enero 2011]. Available from World Wide Web: <<http://www.prodigioconsultores.com/?p=72>>.

Diccionario de la lengua española - Vigésima segunda edición. [citado 6 noviembre 2010]. Available from World Wide Web: <http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=biblioteca>

Dspace-general Info Page. [citado 28 mayo 2011]. Available from World Wide Web: <<http://mailman.mit.edu/mailman/listinfo/dspace-general>>.

Escolar Sobrino, H. 1990. "Historia de las bibliotecas". Madrid: Fundación G. Sánchez Ruipérez

EcuRed. Enciclopedia cubana - EcuRed. [citado 10 enero 2011]. Available from World Wide Web: <http://www.ecured.cu/index.php/EcuRed:Enciclopedia_cubana>.

E-Prints C. Archivo Institucional E-Prints Complutense - E-Prints Complutense. [citado 10 enero 2011]. Available from World Wide Web: <<http://eprints.ucm.es/>>.

evalDSpace_v01.pdf. 2009. [citado 25 Enero 2011]. Available from World Wide Web:

<http://ricabib.cab.cnea.gov.ar/90/1/evalDSpace_v01.pdf>.

evalEPrints_v02.pdf. 2009. [citado 25 Enero 2011]. Available from World Wide Web:
<http://ricabib.cab.cnea.gov.ar/91/1/evalEprints_v02.pdf>.

Fernández Fernández, Maikel Manuel. Abreu Bartomeo, Yanedi. DESARROLLO DE UNA BIBLIOTECA DIGITAL CON EL USO DEL SISTEMA DE GESTIÓN DE CONTENIDOS DRUPAL. 2010.

Fernández Fernández, Maikel Manuel. Abreu Bartomeo, Yanedi. Torricella Morales, Raul G. Definición de los enfoques y las soluciones tecnológicas. 2010.

Furrie, B. Understanding MARC bibliographic: Machine-Readable Cataloging (16a). Cataloging Distribution Service, Library of Congress, en colaboración con TheFollett Software Company. 2003. Available from world wide web: <<http://www.loc.gov/marc/umb/>>.

GUGreenstone. Grupo de Usuarios Greenstone de Latinoamérica. [citado 27 enero 2011]. Available from World Wide Web: <<http://www.greenstone-la.org/>>.

Harold y E. R. Bases de datos: modelos, lenguajes, diseño. 2008.

Laureano Felipe. 2007. Interoperabilidad en los Sistemas de Información Documental (SID): La Información Debe Fluir.pdf. [citado 10 enero 2011]. Available from World Wide Web:
<<http://redalyc.uaemex.mx/pdf/953/95330103.pdf>>.

Le Crosnier, Hervé. Bibliotecas digitales. [citado 8 de enero 2011]. Available from World Wide Web:
<<http://vecam.org/article627.html>>.

López Guzmán, Clara y Estrada Corona, Adrián. Bibliotecas digitales. 2006. [citado 15 noviembre 2010]. Available from World Wide Web:
<http://www.mati.unam.mx/index.php?option=com_content&task=view&id=80&Itemid=51>.

Los Repositorios de Objetos de Aprendizaje. 2010. [citado 9 Enero 2011]. Available from World Wide Web:
<<http://elearning-formaciononline.blogspot.com/2010/03/los-repositorios-de-objetos-de.html>>.

Melero, R. 2005. "Acceso abierto a las publicaciones científicas: definición, recursos, copyright e impacto," El Profesional de la Información.

- Menús. [cited 27 Enero 2011]. Available from World Wide Web: <http://drupal.org/handbook/modules/menu> y <http://drupal.org/node/120632>
- Monge Benito S, Ovelar Beltrán R. Repositorio 2.0: Dinámicas sociales para favorecer el desarrollo de comunidad en torno a un repositorio de contenidos educativos digitales. 2007. [citado 28 Febrero 2011]. Available from World Wide Web: <[http://spdece07.ehu.es/actas/ Monge.pdf](http://spdece07.ehu.es/actas/Monge.pdf)>.
- Muñoz Casals, Velmour. Repositorio de documentos. 2008. [citado 9 enero 2011]. Available from World Wide Web: <<http://www.monografias.com/trabajos-pdf/repositorio-documentos/repositorio-documentos.shtml>>.
- OpenDOAR - Home Page - Directory of Open Access Repositories. [citado 9 enero 2011]. Available from World Wide Web: <<http://www.opendoar.org/>>.
- Ortiz Repiso, Virginia y Moscoso, Purificación. La Biblioteca Digital.pdf. [citado 9 enero 2011]. Available from World Wide Web:<<http://www.bibliociencias.cu/gsd/collect/eventos/import/La%20Biblioteca%20Digital.pdf>>.
- Peñalver Romero, Gladys Marsi. Metodología ágil para proyectos de software libre.2008.
- Pérez, M. y L.F. Silva. 2007. “Cómo funciona el protocolo OAI – PMH en la recuperación de información”.
- Pérez López. Ana. 1995. ECOBIBLIOTECA.pdf. [citado 15 noviembre 2010]. Available from World Wide Web: <<http://eprints.rclis.org/bitstream/10760/7745/1/ECOBIBLIOTECA.pdf>>.
- Remedios Melero. EPI-rmelero.pdf. [citado 9 Enero 2011]. Available from World Wide Web: <<http://eprints.rclis.org/bitstream/10760/6571/1/EPI-rmelero.pdf>>.
- RTD. Acerca de - Repositorio de Tesis Doctorales. [citado 10 enero 2011]. Available from World Wide Web: <<http://tesis.repo.sld.cu/information.html>>.
- Sola Martínez, María José. Greenstone: distribución de colecciones digitales. [citado 27 enero 2011]. Available from World Wide Web: <<http://www.ub.edu/bid/21/sola2.htm>>.
- Shreves, Ric. 2008. Drupal 6 Themes. Packt Publishing.
- Terán, Jorge y González, Víctor. Dspace. [citado 25 enero 2011]. Available from World Wide Web: <<http://www.slideshare.net/jvelizster/dspace-2004449>>.

- Tramullas, J., Drupal para Bibliotecas y Archivos. 2010. Información sobre módulos. [citado 27 enero 2011]. Available from World Wide Web: < <http://drupalmodules.com/>>
- UASB. Repositorio UASB-Digital: Página de inicio. [citado 10 enero 2011]. Available from World Wide Web: <<http://repositorio.uasb.edu.ec/>>.
- UNESCO Greenstone: Un software libre de código abierto para la construcción de bibliotecas digitales. Experiencias en América Latina y el Caribe. .pdf. [citado 27 enero 2011]. Available from World Wide Web: <<http://unesdoc.unesco.org/images/0018/001887/188719s.pdf>>.
- Valdés Guerrero. Herramienta para la migración de bases de datos ISIS a formato MARC. 2009. Available from world wide web: <<http://biblioteca.uci.cu>>
- Vera, F. 2006. "Construcción de Bibliotecas Digitales, modelo de capa piramidal".