

**UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”**

**CONSTRUCCION DE UNA HERRAMIENTA DE INTEGRACIÓN Y
BÚSQUEDA EN SISTEMAS DE GESTIÓN DE CONTENIDOS EN
AMBIENTES DISTRIBUIDOS PARA BIBLIOTECAS DIGITALES**

ING. REYNALDO MANTILLA ACEVEDO

Barquisimeto, 2010

UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”
DECANATO DE CIENCIAS Y TECNOLOGÍA
MAESTRÍA EN CIENCIA DE LA COMPUTACIÓN
MENCIÓN: INGENIERÍA DE SOFTWARE

**CONSTRUCCION DE UNA HERRAMIENTA DE INTEGRACIÓN Y
BÚSQUEDA EN SISTEMAS DE GESTIÓN DE CONTENIDOS EN
AMBIENTES DISTRIBUIDOS EN BIBLIOTECAS DIGITALES**

Trabajo presentado para optar al grado de

Magíster Scientiarium

Por: ING. REYNALDO MANTILLA ACEVEDO

Barquisimeto, 2010

**CONSTRUCCION DE UNA HERRAMIENTA DE INTEGRACIÓN Y
BÚSQUEDA EN SISTEMAS DE GESTIÓN DE CONTENIDOS EN
AMBIENTES DISTRIBUIDOS EN BIBLIOTECAS DIGITALES**

Por: ING. REYNALDO MANTILLA ACEVEDO

Trabajo de grado aprobado

<hr/> (Jurado 1) Tutor	<hr/> (Jurado 2)
<hr/> (Jurado 3)	

Barquisimeto, 2010

INDICE GENERAL

	Página
INDICE DE TABLAS	viii
INDICE DE FIGURAS.....	ix
RESUMEN.....	xi
INTRODUCCION	1
CAPITULO I: EL PROBLEMA.....	4
Planteamiento del problema.....	4
Objetivo General.....	11
Objetivos Específicos.....	11
Justificación e Importancia	12
Alcances y Limitaciones	14
CAPITULO II: MARCO TEORICO	15
Antecedentes	15
Bases Teóricas	17
Datos, Información, Conocimiento y Saber.....	18
Base de Datos.....	20
Libro Electrónico	20
Biblioteca: Electrónica, Digital, Híbrida Y Virtual	21
Características de las Bibliotecas Digitales	22
Alianzas estratégicas.....	23
Alianzas inter-bibliotecarias.....	24
Gestión de conocimientos.....	26
Sistema de gestión de conocimientos	26
Sistema de Manejo de Contenido (SMC)	27

Metadatos.....	29
XML (Extensible Markup Language).....	30
Modelo Cliente-servidor.....	34
Computación Grid.....	36
Peer-to-Peer (P2P).....	36
Agentes Inteligentes.....	39
Sistemas Multi-Agente.....	40
Patron.....	41
Clasificación de los Patrones.....	42
Patrones de Arquitectura ó Arquitecturales.....	42
Patrones de diseño.....	43
Patrones elementales (idiomas).....	43
Características de un buen patrón.....	43
Unified Modeling Language (UML).....	44
Estructuras de arquitectura en UML.....	47
Arquitectura del software.....	48
Arquitectura basada en patrones.....	50
La metodología empleada es la XP, descrita posteriormente.....	50
Metodología XP.....	50
Las Historias de Usuario.....	57
Marcos de trabajo (Frameworks).....	57
Spring framework.....	58
Lucene.....	59
JXTA.....	60
Bases Legales.....	64
Operacionalización de las variables.....	68
Variables Conceptuales.....	68
Variables Operacionales.....	69
CAPÍTULO III: MARCO METODOLÓGICO.....	72
Naturaleza del estudio.....	72

Fase I: Estudio Diagnostico	73
Validación del instrumento	75
Procedimiento	75
Aplicación de los instrumentos	76
Fase II: Factibilidad	76
Social.....	76
Operativa.....	77
Institucional.....	77
Tecnológica.....	77
Económica.....	78
CAPITULO IV: PROPUESTA DEL ESTUDIO	79
Introducción	79
Aplicación de los instrumentos	81
Diagnostico	82
Análisis de datos recabados	87
Primera fase: Análisis de requerimientos	89
Requisitos Funcionales	92
Requisitos no Funcionales	93
Diagramas de caso de Uso	93
Segunda fase: Planificación de entregas	94
Descripción de la herramienta	95
Tercera Fase: Iteraciones	96
Iteración 0	96
Arquitectura de Software	96
Infraestructura y Herramientas utilizadas.....	103
Iteración 1	104
Iteración 2	107
Iteración 3	112
Iteración 4	114
CAPITULO V: CONCLUSIONES Y RECOMENDACIONES	117

Conclusiones	117
Recomendaciones	119
REFERENCIAS BIBLIOGRAFICAS.....	120
ANEXOS	125
ANEXO 1:Cuestionario N° 1 Entrevista semiestructurada	126
ANEXO 2: Guía de observación.....	128
ANEXO 3: Carta de Solicitud de Validación a los Expertos.....	129
ANEXO 4: Formato de Validación de los Instrumentos	130
ANEXO 5: Casos de Uso.....	132

INDICE DE TABLAS

	Página
Tabla 1. Operacionalización de variables estudiadas.....	70
Tabla 2. Plan de Publicaciones (Release plan).....	95
Tabla 3. Herramientas usadas.....	104

INDICE DE FIGURAS

	Página
Figura 1. Tipo de ventaja competitiva buscada.....	20
Figura 2. Conexión cliente/servidor.....	35
Figura 3. Las tres arquitecturas P2P.....	39
Figura 4. Patrones según el nivel de detalle.....	42
Figura 5. UML Resumen de la notación.....	47
Figura 6. Ciclos en XP.....	52
Figura 7. Fases en un proyecto en XP.....	53
Figura 8. Diagrama de Caso de Uso de HYPATIA.....	94
Figura 9. Diagrama de contexto de funcionamiento de HYPATIA.....	99
Figura 10. Arquitectura en capas de HYPATIA.....	99
Figura 11. Modelo de Componentes.....	101
Figura 12. Diagrama de despliegue de Hypatia.....	102
Figura 13. Diagrama de casos de Uso para iteración 1.....	105
Figura 14. Diagrama de clases, Definición de catálogo y repositorio.....	106
Figura 15. Diagrama de secuencia de factory de documento.....	106
Figura 16. Diagrama de casos de Uso para iteración 2.....	107
Figura 17. Diagrama de clases, Indexado y búsqueda de contenido.....	108
Figura 18. Diagrama de secuencia de búsqueda local.....	109
Figura 19. Diagrama de secuencia de Indexado.....	110
Figura 20. Diagrama de clases que modela la capa de núcleo de contenido...	111
Figura 21. Diagrama de paquetes de la capa de núcleo de contenido.....	112
Figura 22. Diagrama de casos de Uso para iteración 3.....	112
Figura 23. Diagrama de clases de la capa de núcleo de nodo.....	113
Figura 24. Diagrama de casos de Uso para iteración 4.....	114
Figura 25. Interfaz de búsqueda en SMC Obelisco.....	115
Figura 26. Diagrama de clases de la capa de servicios.....	116

UNIVERSIDAD CENTROCCIDENTAL "LISANDRO ALVARADO"
DECANATO DE CIENCIAS Y TECNOLOGÍA
MAESTRÍA EN CIENCIA DE LA COMPUTACIÓN
MENCIÓN: INGENIERÍA DE SOFTWARE

**CONSTRUCCION DE UNA HERRAMIENTA DE INTEGRACIÓN Y
BÚSQUEDA EN SISTEMAS DE GESTIÓN DE CONTENIDOS EN
AMBIENTES DISTRIBUIDOS EN BIBLIOTECAS DIGITALES**

Autor: Ing. Reynaldo Mantilla
Tutor: M.Sc. Ing. Ramón Valera

RESUMEN

Las bibliotecas y los centros de documentación e información están asociados a la producción y diseminación del conocimiento, destacándose la información como uno de los elementos relevantes en dicho proceso. Para todas las personas, tener acceso a la información es parte indisociable; sin embargo el desarrollo tecnológico y la aparición de Internet han cambiado la perspectiva de desplazamiento para obtención de información. Las bibliotecas han colocado sus contenidos y servicios de manera digital, no obstante, en la mayoría de los casos dichos materiales son insuficientes para todas las necesidades presentes en la sociedad, por lo que la integración de las mismas permitirá acceder a una cantidad de material más significativa.

Esta investigación realizó la construcción de una herramienta bajo la metodología XP, la cual facilita el desarrollo de redes de bibliotecas digitales, permitiendo a los gestores de contenidos o SMC ampliar sus ámbitos de trabajo a ambientes distribuidos, y proveer a las mismas utilidades que faciliten la integración, búsqueda y extracción de contenidos, así como la definición de un catálogo común, e incorporar características de redes P2P que facilitaran su integración.

Palabras claves: Biblioteca Digital, gestores de contenidos, ambientes distribuidos, integración, búsqueda distribuida, metodología XP.

INTRODUCCION

En la actualidad, la información juega un papel de considerable importancia en la sociedad debido en gran medida, a experiencias previas que sirven como guía para capturar ideas o realizar investigaciones. El conocimiento se deriva de la información como la misma de los datos, el resultado de la secuencia inversa origina el saber. El conocimiento ha ocupado siempre un lugar muy importante del crecimiento y desarrollo económico y social, por esto hay que darle a la información el valor que ella merece, por lo que el hombre se ha preocupado por registrar y almacenar conocimientos para su transmisión por medio de diferentes tipos de herramientas.

La biblioteca en torno a la distribución y almacenamiento de la información ha jugado un papel importante como lo describe Orera (2008), donde plantea que las bibliotecas son el corolario de la relación entre el conocimiento parcial y desordenado, y la representación del mundo como manifestación del orden, la formalidad y la perfección. Al evolucionar en el tiempo la foto del universo y las conexiones entre las ciencias humanas, los modelos de la biblioteca se han ido transformando y mostrando, cada vez, la organización del saber y la forma de ordenación de la mente humana; la biblioteca lleva a cabo la difusión de la información. Hoy en día, se hace hincapié en que ésta es la principal tarea de la biblioteca, labor que aunque siempre ha estado presente en la biblioteca, alcanza hoy sus cotas más altas.

No obstante, además de difundir; la biblioteca tiene también la labor de conservar información y/o conocimiento, tarea ésta presente siempre en la misma. Junto a ello deben mantenerse en equilibrio entre cantidad y calidad de servicios así como la confiabilidad de los documentos digitales, además estos se encuentran en función del tipo de biblioteca

La modernización de las bibliotecas está directamente relacionada con la automatización de rutinas y servicios con el objetivo de implementar una infraestructura de comunicación para agilizar y ampliar el acceso a la información, cambiando así el modo en que el usuario utiliza la biblioteca.

En otro orden de ideas, Tómese en cuenta que a pesar de la tecnología actual, no se satisfacen las necesidades con respecto a la cantidad de servicios y material prestado por algunas bibliotecas. Caso que se observa en la Universidad Centro-occidental Lisandro Alvarado (UCLA) la cual se divide en decanatos, por lo que se considera necesario el acceso a contenidos y servicios de manera distribuida.

Por lo antes expuesto, se presenta un estudio denominado “Construcción de una herramienta de integración y búsqueda en sistemas de gestión de contenidos en ambientes distribuidos en bibliotecas digitales”, el cual será aplicado a las bibliotecas de la Universidad Centro-occidental Lisandro Alvarado, en la ciudad de Barquisimeto, Estado Lara, usando para su desarrollo una metodología Ágil XP(eXtreme Programming).

Este trabajo de investigación esboza la problemática existente de las bibliotecas digitales de los decanatos de UCLA en la actualidad, y de la cual se sustenta a través de mecanismos de investigación anteriormente planteados por otros autores que desarrollaron investigaciones y publicaron sobre los temas tratados en este proyecto, lo cual permite darle un soporte válido a la hipótesis planteada, y como bases para definir la arquitectura del mismo.

Adicionalmente se plantea la metodología aplicada, la cual consta de naturaleza y diseño de la investigación, así como, las técnicas e instrumentos de recolección de datos y técnicas de análisis de datos. Estos elementos buscan darle carácter científico y respuesta a los objetivos planteados en esta investigación, considerando una muestra poblacional de las bibliotecas de los diferentes decanatos de la UCLA.

Finalmente, dado el estudio diagnóstico y de factibilidad del capítulo anterior, se presentan todas las fases de la metodología XP, que fueron necesarias para producir la herramienta de integración y búsqueda en sistemas de gestión de contenidos en ambientes distribuidos para bibliotecas digitales HYPATIA, presentado a su vez, los principales artefactos UML que describen algunas vistas del sistema en estudio, para llegar posteriormente a una serie de conclusiones y algunas recomendaciones de la investigación.

CAPITULO I

EL PROBLEMA

Planteamiento del problema

El conocimiento ha ocupado siempre el lugar central del crecimiento económico y de la elevación progresiva del bienestar social. El desarrollo se ha alimentado históricamente de la capacidad de inventar e innovar nuevos conocimientos y nuevas ideas que se materializan luego en productos, procedimientos y organizaciones.

Hoy en día, la sociedad le ha dado mayor importancia a la información, en gran medida, debido a la observación de experiencias que sirven de guía para capturar ideas que ayuda a su propio desarrollo. Al respecto Valentim (1995), indica que la importancia dada a la información es directamente proporcional al desarrollo de la sociedad, planteando que cuando más desarrollado sea un país mayor será el nivel de producción de información. Siendo consecuentemente mayor el valor que la sociedad otorgara a la información.

El aprendizaje es básico para la subsistencia y evolución de la especie como lo plantea Torrou (1969), el cual modifica y transforma las estructuras del pensamiento, permitiendo la realización de nuevos aprendizajes de mayor complejidad. Con la primera gran división social del trabajo, se generó un mayor caudal de conocimientos sobre la naturaleza y la propia actividad humana, lo que condujo a nuevos términos y herramientas para la transmisión de dichos conocimientos, permitiendo a su vez, ampliar y perfeccionar la información sobre su mundo circundante.

La necesidad de la información nace con los primeros hombres que habitaron la tierra y esta le permitió al humano según Engels (1964), adquirir su condición de ser social. Castaño (2002), describe cómo desde el principio de los tiempos parte de la humanidad se ha preocupado por registrar y almacenar conocimientos para la transmisión de los mismos por medios de herramientas desde las pinturas rupestres, hasta los dispositivos de almacenamiento actuales.

La economía del conocimiento se basa en el conocimiento como lo plantea Tang (2000), quien indica que en la era de la economía del conocimiento la gestión se refiere a identificar, adquirir, desarrollar, resolver, usar, almacenar y participar del conocimiento de una manera efectiva, a crear un acercamiento para transformar y compartir el conocimiento tácito y explícito y a aumentar la necesidad y capacidad de novedad utilizando el saber del equipo. Desde que el conocimiento se ha convertido en la fuerza conductora para el desarrollo social, la atención de la sociedad a la información y el conocimiento está creciendo del mismo modo que, poco a poco, se está incrementando la demanda en la gente.

Vieira y otros (2006), citando a Ranganathan señalan que la biblioteca es depositaria del conocimiento humano, el cual crece día a día, y consecuentemente la biblioteca como organismo también. En estas, se generaron cambios significativos en el siglo XX, pasando de ser guardianas del conocimiento a gestores del conocimiento, un intermediario entre el lector o usuario y la información. Las bibliotecas y los centros de documentación e información están asociados a la producción y disseminación del conocimiento, destacándose la información como uno de los elementos relevantes en este proceso. Para todas las personas, tener acceso a la información es parte indisociable

En este contexto, la difusión de este conocimiento es de igual importancia si la comparamos con la innovación del conocimiento. Shanhong (2000), plantea que los agentes creadores del conocimiento no tienen mucho tiempo, ni energía para

preocuparse de los usuarios del conocimiento. Igualmente expone, que aunque hay una multitud de usuarios del conocimiento es muy difícil adquirir el conocimiento que ya existe en las mentes de los creadores del conocimiento restringido por objetivos y condiciones subjetivas; por lo tanto, las bibliotecas deben jugar el papel de mover el conocimiento y usar los diferentes medios y canales para difundir los nuevos conocimientos. En el siglo XXI, la red Internet, con su gran masa de información y extenso contenido, proporciona a la gente el principal acercamiento para buscar el conocimiento y adquirir información.

Con respecto a la integridad de esta, Shanhong (ob. cit.), indica que en la actualidad emerge la información absurda, salaz, falsa y descortés que resulta de la búsqueda de los beneficios comerciales y objetivos políticos en la red; razón por la cual las bibliotecas deben conceder una especial importancia al suministro de Servicios para que la gente adquiera conocimiento de calidad y pueda ejercer máximas funciones y eficiencia de la información del conocimiento. Por tanto, los servicios del conocimiento basados en redes de información rápidos, deben ser llevados a cabo por la creación de bibliotecas virtuales, además de llevar a cabo servicios de digitalización del conocimientos así como digitalizar los recursos de las bibliotecas. Las bibliotecas electrónicas o digitalizadas son los medios técnicos y mercados de desarrollo de las bibliotecas en la era de la economía del conocimiento. Los servicios de conocimiento de las bibliotecas en el futuro comenzarán con la creación de bases de datos que comprendan periódicos electrónicos y libros en diferentes lenguas que tengan características disciplinares y puedan operar en redes de información de alta velocidad. Es de gran importancia la digitalización de material impreso para que el número de información en las bibliotecas aumente como lo plantea Shanhong (2000), se debe hacer un gran esfuerzo para transformar los grandes recursos existentes de información no electrónica en información electrónica e integrarlas en bibliotecas digitales,

La modernización de las bibliotecas está directamente relacionada con la automatización de rutinas y servicios con el objetivo de implementar una infraestructura de comunicación para agilizar y ampliar el acceso a la información por el usuario como es planteado por Vieira y otros (2006), la implementación de tecnologías de la información en bibliotecas es una poderosa arma para cambiar el modo de hacer el trabajo.

Por lo anteriormente mencionado Romero y otros (2003), plantean que el desarrollo de la tecnología y la actividad laboral la cual demanda tener acceso inmediato a la información para la toma de decisiones lo más acertada posible, exige mecanismos que permitan almacenar la información y luego recuperarla y actualizarla con base a peticiones de manera rápida y confiable, para lo cual se han generado a través de la ingeniería del software herramientas informáticas que facilitan el manejo de la información o datos, ayudando a aligerar la sobrecarga de información, definiéndose esta como una colección de información almacenada de manera organizada en un ordenador, que permite la obtención rápida de grandes cantidades de información.

Christensem y otros (1999), expresan que la información natural no suele tener una estructura predefinida y por lo general es presentada como no estructurada, está diseñada para el uso humano y no es eficiente en una máquina, la cual, procesa más fácilmente la información estructurada ya que ésta es ingresada en un modo uniforme para facilitar su procesamiento futuro.

Pero en la actualidad se puede encontrar en una organización, información de tipo estructurada o no estructurada, y el análisis de estas es vital para las mismas, como lo plantean Sebastián y otros (2004), muestran que para la gestión de vastas colecciones de información plasmadas tanto en información estructurada como no estructurada, las compañías han venido empleando en las últimas décadas múltiples herramientas informáticas, desde sencillos sistemas de gestión de bases de datos

relacionales y/o documentales, a programas más potentes y complejos con funcionalidades añadidas, siendo estos últimos años marcados por el tratamiento de la información altamente estructurada y por lo tanto relativamente sencilla de manejar. Pero Sebastián y otros (ob. cit.), plantean que hoy en día, el modelo clásico de información corporativa se rompe, al constatar que un ochenta por ciento de los contenidos de cualquier organización se encuentra en documentos poco estructurados.

Por lo anteriormente planteado se expone que el tratamiento de la información no estructurada es de importancia, por lo que Llauger y otros (2001), exponen que al igual que el software de gestión de base de datos ordena y mantiene la integridad de datos estructurados, la gestión documental proporciona las mismas capacidades para los datos no estructurados y semi-estructurados de los documentos; esto se logra a través de metadatos sobre atributos propios del documento, que permiten optimizar los procesos de búsquedas y extracción, además de ordenar y gestionar documentos compuestos.

Además del tipo de información, Vieira y otros (2006), plantean que la posibilidad de compartir recursos y contenidos por medio de la interconexión de dispositivos, ha posibilitado el intercambio de datos y el uso conjunto de periféricos entre esos equipamientos gracias al desarrollo de las telecomunicaciones, provocando cambios en la producción, la transmisión y en el uso de la información. Internet facilitó el acceso a innumerables servicios, investigaciones, bases de datos, etc. Creando una verdadera revolución ideológica, científica y comercial a gran escala.

Rodríguez y otros (2008), reflejan que con el desarrollo de las nuevas tecnologías de información se afectan directamente las formas de trabajo del investigador, ya que la información está más a su alcance. Antes se recorría biblioteca tras biblioteca para revisar los ficheros de sus catálogos; hoy día, se realiza desde la computadora en un ambiente amigable denominado world wide web (WWW) en conjunto. Es contundente, como el uso de recursos de cómputo ayuda al investigador

en las tareas de manejar la información y en consecuencia a un mejor aprovechamiento de los materiales bibliográficos

Hoy en día existen gran cantidad de herramientas que facilitan el acceso a la información como lo plantean Llauger y otros (2001), indicando que un sistema de gestión de conocimientos permite poner información al alcance de todas las personas o grupos de una organización gracias a un conjunto de elementos y recursos tecnológicos, el uso de los mismos se ha extendido ya que están orientados a facilitar el acceso a la información, permitiendo generar y almacenar nuevos conocimientos; así como, el acceso de sus miembros a un ámbito documental digital de formación y de comunicación.

Wals (2003), describe que la biblioteca universitaria que surge ahora ya no es un edificio lleno de libros y revistas en papel con espacio para su lectura y la posibilidad de llevarse un libro en calidad de préstamo, es y tiene que ser, un sistema gestor de información, basada en una organización humana estructurada, que realiza los procesos y proporciona los servicios utilizando los recursos y las herramientas más adecuados para las exigencias y las posibilidades del momento. Basado en esa descripción y ubicándonos en la biblioteca universitaria de la UCLA, ésta dista mucho de esta afirmación, a pesar de contar con una diversidad de herramientas para facilitar el acceso a la información, y contar con una plataforma tecnológica Valera (2005), enmarca que la misma no tiene políticas acertadas para el uso adecuado sus recursos electrónicos, por lo que este investigador, generó una herramienta para la gestión de contenidos que puede permitir el intercambio de documentos de manera fácil y confiable entre la comunidad universitaria, dicha herramienta podría ser utilizada por la biblioteca de la universidad.

Ahora bien, queda el problema de la cantidad de material del cual disponen las bibliotecas universitarias y los servicios que las mismas pueden prestar, queda a su vez el detalle de aumentar el número y la calidad de recursos que esta pondrá a

disposición de los alumnos. Estando la UCLA dividida en decanatos hace necesario que el material de cada uno de estos, sea compartido a otros. Es posible que aplicando el concepto de biblioteca distribuida, la información y los servicios, se integren en un medio único que acoge tanto la gestión como el acceso ya sea que la misma se efectúe desde lugares muy diversos.

Para solventar el problema de la relevancia que las bibliotecas universitarias aporten información de calidad para el desarrollo de nuestras sociedades el Temario Auxiliar de Bibliotecas de la Universidad de Zaragoza escrito por Técnica y Empleo Aragón (2007), plantea las alianzas inter bibliotecarias, con el propósito de identificar y resolver necesidades comunes surgidas a partir del desarrollo de la tecnología de la información, debido que en algunos casos los recursos de algunas bibliotecas son insuficientes y las bibliotecas deben incrementar el uso de sus recursos. Cooperar es la clave por lo que la globalización de la información, la limitación de recursos y la obligación de responder a la demanda de la comunidad científica lleva a adoptar una actitud cooperativa como única forma de supervivencia.

Por lo anteriormente expuesto, es posible que la integración de las bibliotecas de los diferentes decanatos de la UCLA o la alianza con otras bibliotecas universitarias permita desarrollar con mayor facilidad la calidad de servicios y la integración de las mismas. Ante esto, se plantea la existencia de una herramienta que admita la búsqueda y la compartición de contenidos entre diversas bibliotecas que podrían ser utilizadas por gestores de contenidos, lo cual facilitaría que bibliotecas con objetivos afines y con necesidades de fortalecerse en cuanto a los servicios que esta presentan, sean posiblemente, más realistas y funcionales.

Para dar una posible solución a la problemática planteada, se aplicará el concepto de gestión documental y el uso de tecnologías de búsqueda e integración en redes, para generar una herramienta que pueda ser usada por diversos manejadores de contenidos, que pueda permitir que las bibliotecas universitarias tengan un mayor

impacto en la sociedad actual y faciliten las tareas de investigación de sus estudiantes y profesores.

Objetivo General

Construir una herramienta de integración y búsqueda en sistemas de gestión de contenidos en ambientes distribuidos para bibliotecas digitales.

Objetivos Específicos

1. Diagnosticar la necesidad de integrar las bibliotecas de los decanatos de la universidad centro-occidental Lisandro Alvarado (UCLA).
2. Establecer los requerimientos o características relacionadas a la creación de una herramienta para la integración y búsqueda de contenidos en ambientes distribuidos.
3. Determinar la arquitectura del sistema, que permita la búsqueda, indexación e integración de contenidos en ambientes distribuidos.
4. Construir la herramienta de integración y búsqueda en gestores de contenidos para el sector bibliotecario

Justificación e Importancia

En la actualidad el conocimiento es vital para el desarrollo de la sociedad, y las universidades son factor importante para la distribución de la misma. Para la biblioteca universitaria actual es fundamental distribuir y dar acceso a la información de manera rápida y fácil y no obligar a investigadores o alumnos a tener que desplazarse a diferentes sitios para acceder a un número limitado de información.

Como indico Paez (2009) el sistema educativo debe adecuarse a los cambios que está exigiendo la sociedad actual y el rol que desempeña el docente debe trazarse hacia la sociedad de la información, donde los centros educativos han perdido, en parte, el liderazgo sobre la distribución de la información, supliendo la necesidad creciente de información. Para esto es vital la integración en un mundo globalizado.

Para el acceso a la información en el área universitaria es vital tanto para el estudiante que investiga como para el profesor que corrobora la investigación, la confiabilidad de las fuentes en la Internet, a pesar de proveer de una gran cantidad de información, la mayoría de ella es cuestionable, por lo que el papel de las bibliotecas digitales es muy importante. A pesar de esto, en la actualidad y específicamente en el Estado Lara; Venezuela, el investigador tiene que escudriñar en Internet encontrando fuentes del todo no fiables y algunas veces con dificultad para obtener información acerca de un tema en específico a causa de que las bibliotecas tienen gran cantidad de material no digitalizado y sin una plataforma para distribuirlos. Algunas universidades ofrecen bibliotecas digitales pero en la mayoría de los casos no presentan un número considerable de servicios y sus contenidos son muy limitados, ejemplo de ello se visualiza en la biblioteca del DAC, la cual tiene la necesidad de suscribirse a otras bibliotecas en el extranjero, principalmente a España para darle así la posibilidad de acceso a sus usuarios a libros, papers o documentos que no dispone

la misma, servicio o suscripción ésta que no es del conocimiento de la población universitaria de la misma UCLA.

Ahora bien, considerando la política de independencia tecnológica actual y la necesidad de fortalecer el papel de las bibliotecas digitales en el acceso a contenidos de investigación puede ser necesaria la integración de las bibliotecas de los diferentes decanatos de la misma universidad o alianzas inter-bibliotecarias del sector universitario, integrando sus servicios y contenidos, la realización de una herramienta que pueda ser usada por un gestor de contenidos que facilite la integración entre diversas bibliotecas digitales permitirá eliminar las barreras tecnológicas que pudieran existir y dar mejores servicios tanto a los estudiantes como a los profesores, y de esta manera mejorar la difusión de conocimiento en el país.

Según el direccionamiento estratégico Institucional de la universidad centroccidental “Lisandro Alvarado” 2006 -2011 aprobado en sesión numero 1636, ordinaria, del consejo Universitario de fecha 13-07-2005, platean varios objetivos al alcanzar se mencionaran algunos que justifican la investigación y que se consideran relevantes para el funcionamiento bibliotecario, como lo son:

- Implementar continuamente el equipamiento de apoyo para el desarrollo de las funciones académicas cónsonas con el número de estudiantes y logística del Decanato (bibliotecas, talleres, laboratorios y equipos).
- Desarrollar canales sistemáticos de comunicación con los núcleos Universitarios a fin de garantizar la satisfacción oportuna de sus necesidades.
- Crear un plan para optimizar el uso de los recursos destinados a la Investigación en el Decanato, de manera tal que se sumen o articulen dichos recursos para facilitar las diferentes actividades de investigación.
- Promocionar y difundir los resultados de investigaciones y desarrollo tecnológico alianzas

Alcances y Limitaciones

La solución planteada se puede aplicar en cualquier contexto de búsquedas distribuidas de documentos digitales, a universidades con múltiples decanatos y distribuciones físicas, pero el ámbito de la investigación se realizara a las bibliotecas de la universidad centroccidental Lisandro Alvarado, ya que dicha institución tiene una infraestructura tecnológica adecuada para el desarrollo del mismo, y que las bibliotecas son fundamentales para la transferencia de conocimientos.

En este trabajo se planteara la alianza estratégica inter-bibliotecarias, pero el trabajo no intentara establecer la factibilidad de esto, solo se referirá a investigaciones previas como opciones de uso, basadas en esta herramienta.

La recolección de datos será para el desarrollo de la herramienta de acuerdo a la metodología de desarrollo seleccionada, no se harán análisis estadísticos de los mismos, ya que este no es el objeto de la misma.

El indexado, y búsqueda se realizara solo sobre publicaciones de tipo electrónico.

CAPITULO II

MARCO TEORICO

Antecedentes

Los antecedentes permitieron definir e interpretar el problema planteado, mediante trabajos vinculados a la investigación y estableciendo un enfoque teórico y metodológico. Tales recursos sirvieron de sustento a todo lo relacionado con gestión de contenidos, bibliotecas virtuales e integración de información, considerados éstos necesarios en el desarrollo de una herramienta para la integración y búsqueda en sistemas de gestión de contenidos en ambientes distribuidos para bibliotecas digitales. Los trabajos considerados fueron los siguientes:

En la investigación “Modelado de Acervos en Bibliotecas Digitales”, Gnosis (2004), se exponen los distintos tipos de datos que son comunes en una biblioteca digital, así como las técnicas y herramientas para su manipulación del material de las mismas, relevantes al exponer los diversos tipos de formatos digitales utilizados por las mismas.

Valera (2005) realizó una investigación sobre las tecnologías de información y la ingeniería de software orientadas a la creación sistemas de gestión de conocimiento, para permitir formar un extenso y homogéneo repositorio de documentos estructurados o semi-estructurados, así como la gestión de contenidos y documentos basados en la Web, facilitando la creación, intercambio, publicación y búsqueda de documentos, además proveer y establecer los niveles de seguridad necesarios para el acceso de los mismos.

La relevancia de este estudio se fundamenta en que el mismo proporciona un posible ambiente para el uso de herramientas como esta, ya que esta investigación permitirá expandir el potencial en el uso de recursos como éstos para ser utilizadas en las bibliotecas digitales.

Brogan (2006) planteo en su investigación sobre librerías distribuidas, la interoperabilidad a través de cuatro dominios que se cruzan: tecnología educativa, servicios de biblioteca, instrumentos de escritorio, y software social. La relevancia de este estudio se fundamenta en que el mismo proporciona metodologías y obstáculos para la creación de colecciones digitales dentro de ambientes de investigación, mostrando aspectos propios del negocio de las bibliotecas, como: digitalización de contenidos y el uso de metadatos para transferencia de fichas en ambientes distribuidos.

Páez (2006) nos muestra en su investigación el diseño de una arquitectura de software, que permite publicar objetos de contenidos, siguiendo las normas de empaquetamiento de información proporcionado por el modelo estándar SCORM y basado en lenguaje de patrones. Esta investigación permitió evaluar esta arquitectura de contenidos, para extrapolar los resultados de la misma y basarla en ambientes distribuidos.

Así como el análisis de cada uno de los patrones de diseño, usados para el desarrollo de su arquitectura y estudiando algunos de los cuales serán usados en el diseño de la arquitectura de la herramienta planteada, esta investigación explica patrones como el Strategy, el Component Configurator, el wrapper Facade, el monitor Object, Acceptor-Connector, Reactor, Composite.

Steffen Staab, Heiner Stuckenschmidt (2006) en su libro “Web Semántica entre iguales (P2P) para Dirección Descentralizada e intercambio de Conocimiento e

Información”, describe dos tecnologías que dirigen una necesidad común en niveles diferentes:

Las tecnologías P2P que apuntan al abandono del control centralizado a favor de principios de organización descentralizados. La misma, permite nuevos grados de libertad para cambiar arquitecturas de información y transferir información entre nodos de una red.

Tal libro se considera de gran importancia como base para esta investigación ya que plantea el uso de la tecnología P2P, que aplicadas al problema planteado, podrían ser muy útiles dado que muestra diversas arquitecturas, problemas de las mismas, métodos prácticos y metodologías, aplicadas a los tópicos de manejo de contenidos y distribución de los mismos.

Martinez (2007) plantea en su investigación “Consideraciones iniciales para implementar funciones de búsqueda federada por medio de una red tipo igual-igual (P2P) de repositorios basados en Web Semántica”, el uso de tecnologías p2p para permitir la integración de bibliotecas, para el proyecto de una red abierta de bibliotecas digitales. Tal trabajo sirve también para evaluar y comparar algunos frameworks para este tipo de redes.

Todos los antecedentes anteriores presentan una relación directa con el estudio, proporcionando conceptos, arquitecturas, experiencias que sirven de marco para el desarrollo de la misma. De igual manera, se consideraron textos bibliográficos que demuestran el estudio ampliado de los elementos a desarrollar en este proyecto y que a continuación se expresan:

Bases Teóricas

En este punto se desarrollan las bases teóricas con las cuales se sustenta el análisis comparativo de la investigación. Donde fueron revisadas y analizadas ciertas

bibliografías relacionadas con el tema a investigar, pretendiendo que sirvan de apoyo para el claro entendimiento e interpretación de los parámetros estudiados. A tal efecto, se inició el referencial teórico con una serie de conceptualizaciones lo que son los datos, la información, y el conocimiento, así como permiten la formación del saber, a través de la definición de base de datos, documento digital, biblioteca digital y los diversos términos con la que es llamada y sus divergencias, incluyendo sus características.

Al mismo tiempo, se pueden encontrar los conceptos de gestión de conocimientos, y sistemas de gestión de conocimientos, considerando sus características principales. A la par, se desarrolla la descripción de web semántica, como está conformada, y su arquitectura, para lo cual será necesario describir que son metadatos, y explicar concepto XML, sus especificaciones y características, así como RDF con una descripción de sus funcionamientos, y modelados.

Para ultimar, se describen los diferentes modelos y o protocolos que permitan la comunicación entre nodos tales como, modelo cliente servidor, sus ventajas, y conceptos; computación grid, P2P, describiéndolo y mostrando los diversos usos que presenta en la actualidad, y sus diversas arquitecturas; asimismo tenemos definición de web service y patrones, su concepto y diferentes clasificaciones: arquitectural, de diseño y elemental.

Para continuar el desarrollo del presente capítulo se muestra a continuación una serie de conceptualizaciones que de manera general son necesarias a la luz del progreso continuado de este proyecto; partiendo de lo general a lo específico tenemos:

Datos, Información, Conocimiento y Saber

Valhondo (2003) define a los datos como un conjunto discreto de hechos objetivos acerca de eventos. En el contexto empresarial, los datos pueden ser descrito

como registros estructurados o transacciones. Estos carecen de sentido porque describen solo parcialmente lo que sucede y no proporcionan juicio ni interpretación, para la toma de decisiones, es decir, un dato en bruto no nos dice que tenemos que hacer.

Fernandez (2005) describe a la información como un conjunto de datos que sirven para tomar una decisión. La información que fluye en una empresa por naturaleza es cíclica, por lo tanto los datos que no sirven para la toma de decisiones no son información.

Valera (2005) citando a Davenport y otros define a la información como un mensaje, normalmente bajo la forma de un documento o algún tipo de comunicación audible o visible.

El conocimiento deriva de la información como esta deriva de los datos, y Valhondo (2003) la define como una mezcla fluida de experiencia, valores, información contextual y apreciaciones expertas que proporcionan un marco para su evaluación e incorporación de nuevas experiencias e información. Siendo entonces una mezcla de elementos.

Valhondo (2003) citando a otros indica que la secuencia de datos, información y conocimiento se extienden hasta un nivel superior, el saber, definiéndolo como la capacidad de comprender los principios, como contraposición al conocimiento, que comprende patrones, y la información, que comprende relaciones y cuya acumulación puede dar lugar al capital intelectual, ver figura 1.

Marín y otros (1995) indican que información significa hablar de datos estructurados y de las condiciones y límites para su transmisión, porque los datos, de no estar estructurados con otros, no transmiten ningún mensaje



Figura 1. Tipo de ventaja competitiva buscada
Fuente: Gestión del conocimiento del mito a la realidad, 2003.
Autor: Domingo Valhondo

Base de Datos

Betancourt (2004) citando a Date, indica que un sistema de base de datos no es más que un sistema computacional de mantenimiento de registros, que se diseña para manejar grandes cantidades de información. El tratamiento de los datos conlleva tanto la definición de las estructuras para el almacenamiento de la información como el suministro de mecanismos para la manipulación de la información.

Libro Electrónico

Las primeras concepciones sobre el libro electrónico fueron producto de todo un contexto tecnológico y cultural que se hizo más evidente en la década de 1990, cuando las diversas innovaciones tecnológicas como la computadora personal, la red internet y el uso de la web, comenzaron a popularizarse. Definiéndose como un libro cuyas páginas son compuestas por medio de un formato electrónico dinámico. Las páginas que lo conforman son organizadas de forma semejante a las de un libro convencional como lo definen Torres y otros (2005).

Biblioteca: Electrónica, Digital, Híbrida Y Virtual

Las definiciones de biblioteca virtual, digital, híbrida, y virtual. Son ambiguas como los conceptos relacionados con las tecnologías de la información por esto se expondrán cada uno de estos conceptos

Biblioteca Electrónica Según Faba y otros (2004) citando a López la describen como aquella que cuenta con sistemas de automatización que le permiten una ágil y correcta administración de los materiales que resguarda, principalmente en papel. Así mismo, cuenta con sistemas de telecomunicaciones que le permiten acceder a su información, en formato electrónico, de manera remota o local. Proporciona principalmente catálogos y listas de las colecciones que se encuentran físicamente dentro de un edificio.

Biblioteca Digital Torres y otros (2005) describen la biblioteca digital como una colección digitalizada que aunada al uso de herramientas tecnológicas para la administración la información. Conjuga el ciclo de la creación, diseminación, uso y preservación de los datos, la información y el conocimiento

Biblioteca Híbrida HyLife Programme, (2001) la describe como una biblioteca donde se conjugan formatos electrónicos e impresos reunidos en un servicio de información integrado al que se accede mediante una combinación de pasarelas electrónicas locales y remotas.

Las bibliotecas virtuales Araceli y otros (2005) la describen como sistemas basados en la tecnología web que ofrecerá un acceso ilimitado a inmensas colecciones de información, compuestas por publicaciones generadas digitalmente.

Faba y otros indican (2004) que el concepto “biblioteca digital” es considerado sinónimo de electrónica, virtual, e incluso híbrida.

Ince (2001) equipara biblioteca digital y virtual considerándola meramente como un depósito de textos electrónicos en línea.

Araceli y otros (2005) indican que el termino biblioteca digital puede ser utilizado como sinónimo de biblioteca virtual, y tal uso es correcto, indicando que biblioteca digital es el término más utilizado tanto en artículos científicos como en conferencias y eventos académicos a nivel mundial.

Características de las Bibliotecas Digitales

La Association of Research Libraries (ARL) (<http://arl.cni.org>) muestra las siguientes características como elementos determinantes en las bibliotecas digitales:

- Son entidades complejas.
- Requieren el uso de una amplia tecnología para reunir los recursos de información distribuidos a través del mundo.
- Las conexiones que establecen con los servicios de información son transparentes para el usuario final.
- Su objetivo es el acceso universal a las bibliotecas y servicios de información.
- Los acervos digitales no se restringen a sustitutos de documentos, también contienen elementos que no pueden ser representados o distribuidos en formato impreso.
- La disponibilidad y fiabilidad de las tecnologías de la información y la comunicación (TIC).
- La tenencia de recursos económicos y herramientas que permitan el acceso a determinados servicios.
- El conocimiento de la lengua en la que están escritos la mayor parte de los recursos informativos (lengua inglesa).

- Las barreras contractuales impuestas por los vendedores, también excluyen a gran parte de los usuarios potenciales.

Faba y otros (2004) plantean que estas limitaciones pueden inhibir a una gran mayoría de usuarios, lo que contribuiría a incrementar la marginalidad de los ya marginados, esto es, de los que no disponen de los recursos económicos necesarios. Esta contradicción entre el acceso para unos pocos y la marginación para muchos plantea un gran reto para las bibliotecas digitales y, por consiguiente, para su gestión.

Finalizada la conceptualización de términos muy puntuales, es imperioso ahondar en otros temas de mayor envergadura e importantes para este trabajo, entre los cuales encontramos las alianzas estratégicas, ya que se persigue ampliar la conexión de las diversas bibliotecas de los decanatos de la UCLA para extender su campo de acción y las alianzas estratégicas pueden ser la alternativa viable en este proceso.

Alianzas estratégicas

Las alianzas estratégicas o estrategias de cooperación, se han incrementado en los últimos años porque han permitido que las empresas puedan ampliar su capacidad para crear nuevos productos, reducir costos, incorporar nuevas tecnologías, penetrar en nuevos mercados, desplazar a la competencia, generar inversiones para tareas esenciales, etc.

Bateman (2004), denomina una alianza estratégica como la relación formal creada entre organizaciones independientes con el objetivo de perseguir metas mutuas en conjunto.

Para Thompson y otros, (2004), son acuerdos de cooperación entre las compañías que van más allá de los tratos normales entre una compañía y otra. Pero

que no llegan a ser una fusión o una sociedad en participación en sentido estricto con lazos de propiedad formales.

Las Alianzas Estratégicas son instrumentos que pueden ponerse en práctica con todas las personas naturales o jurídicas donde exista alguna relación, de allí que lo importante es establecer prioridades y evaluar con quienes resulta más conveniente y hasta imprescindible su establecimiento.

Sólo en lo que se refiere al beneficio directo de la empresa (ya que también pueden plantearse alianzas estratégicas para el desarrollo de la comunidad, el desarrollo de la cultura, la protección del medio ambiente, la defensa de los derechos de la mujer, la prevención de la delincuencia, etc.), los candidatos más obvios son: Proveedores, trabajadores de la propia empresa, clientes reales o clientes potenciales competidores, representantes del gobierno central (ministerios, etc.) o de gobiernos seccionales (juntas parroquiales, municipios, consejos provinciales) representantes de la sociedad civil (fundaciones, organizaciones, entidades culturales, deportivas, religiosas, financieras, etc.) agencias internacionales de desarrollo. En general, toda persona, natural o jurídica, toda institución, que quiera lograr algo similar o coherente a lo que quiere lograr la empresa; en este caso, no es ilógico pensar en una alianza inter-bibliotecaria para el desarrollo del estudiantado y profesorado de la UCLA y no se estaría hablando de un término desconocido, ya que el mismo ha sido considerado por otros autores como los que se nombran a continuación.

Alianzas inter-bibliotecarias

Técnica y Empleo Aragón, S.L.(2007) en su temario Oposiciones Auxiliar de Bibliotecas de la Universidad de Zaragoza plantean que existen una serie de principios validos para todas las bibliotecas, que casi obligan a buscar ayuda cooperativa ya que se considera que toda biblioteca es insuficiente.

La Globalización de la información, la limitación de recursos y la obligación de las bibliotecas de responder a la demanda de la comunidad científica lleva a adoptar una aptitud cooperativa como única forma de supervivencia.

Los factores que favorecen esta cooperación son:

- Las enormes posibilidades tecnológicas para compartir información.
- Grandes cambios en los servicios y los productos que ofrece el mercado de la información.
- El elevado costo de nuevos productos y servicios
- La necesidad del conocimiento muy especializado para poner en marcha algunos de estos servicios.
- Las creciente demanda de usuarios de las bibliotecas.

Formas de cooperación como préstamo inter-bibliotecario, catalogación automatizada, digitalización de material, Desarrollo de colecciones. Podrían traer beneficios a las bibliotecas tales como:

- Ahorro de costos
- Compartir complejidad técnica de los servicios bibliotecarios.
- Incrementar la visibilidad de las biblioteca, hacerlas mas importantes y asegurar así, apoyo político y financiero a sus proyectos.
- Compartir riesgos y errores en una situación en que las inversiones son grandes y los entornos tecnológicos muy cambiantes.
- Aprender de los demás

Sin embargo las alianzas inter-bibliotecarias por si solas no complementan el soporte necesario para garantizar la facilidad en la forma de adquirir todo el conocimiento inmerso en las mismas; por tal motivo, otro tema imprescindible que debe abordarse tiene que ver con la gestión del conocimiento y más aun la utilización

de sistemas de gestión del conocimiento, a través de herramientas tecnológicas para transferir ese conocimiento. Hoy por hoy los sistemas de tecnología se han convertido en apoyo y propagadores del conocimiento, así lo indican los autores señalados con posterioridad.

Gestión de conocimientos

Valhondo (2003), los describe como procesos pre-acordados que permiten mejorar la utilización del conocimiento y de la información que manejan las personas y los grupos. Proporcionando una base tecnológica para el desarrollo de dichos conocimientos.

Las iniciativas de gestión del conocimiento florecen a mediados de los 90 en parte a internet y a la aparición de organizaciones enfocadas hacia la gestión de los recursos de conocimiento explícitos y tácitos para logra ventajas competitivas.

Valera (2005), citando a Carrillo (1999) indica que la gestión de conocimiento (GC) es la aplicación sistemática del entendimiento científico acerca del conocimiento, como estrategia deliberada de individuos, organizaciones o sociedades, para optimizar la producción de valor.

Sistema de gestión de conocimientos

Valera (2005) citando a Ruggles (1997) define a los Sistemas de Gestión de Conocimiento como herramientas tecnológicas que permiten aumentar, generar, codificar y transferir el conocimiento. Los Sistemas de Gestión del Conocimiento deben realizar funciones específicas para apoyar la generación del conocimiento por parte de las personas, así como su almacenamiento, recuperación y presentación.

Dentro de los sistemas de gestión del conocimiento encontramos diversos sistemas de software que dan respuestas a las necesidades planteadas por los usuarios

de los mismos. Este trabajo de investigación persigue idear una herramienta de software que pueda hacer internamente una búsqueda en las bibliotecas de los decanatos de la UCLA, partiendo de conceptos puntuales inmersos en los textos presentes en las bibliotecas digitales con sus respectivas alianzas; para llevar a cabo el mismo se utilizó el sistema de manejo de contenido (SMC) como una alternativa viable por lo que se indica en el siguiente párrafo.

Sistema de Manejo de Contenido (SMC)

Joomla (2006) lo describe como un sistema de software que permite crear sitios web de alta interactividad, profesionalidad y eficiencia, que organiza y facilita la creación de documentos y otros contenidos de un modo cooperativo ya sean páginas de texto, imágenes o archivos multimedia, sin necesidad de editar los archivos y tener que realizar pesadas tareas de mantenimiento.

Valera (2005) indica que los sistemas de manejo de contenidos o SMC deben ser capaces de:

- Definir niveles de acceso, para la publicación y consulta de documentos.
- Almacenar y ordenar todos los tipos de documentos, incluyendo gráficos, audio y video en un repositorio central.
- Crear un flujo de trabajo entre las personas que crean documentos y los responsables de su publicación.
- Cambiar la apariencia general del formato del sitio, a través de la definición y rehusó de plantillas.
- Agregar nuevas categorías de contenido y metadata a los documentos antes y después de que son almacenados en la base de datos de documentos.

Existen sistemas que manejan de manera específica la gestión documental y que debe considerarse conjuntamente con los SMC, para darle un sustento detallado de lo que se persigue en el trabajo aquí planteado.

Sistemas de Gestión documental

La Federación Española de Sociedades de Archivística Biblioteconomía y Documentación (1994), los describe como las herramienta para una oficina sin papeles, los cuales están compuestos por una gran variedad de tecnologías diferentes que integradas consiguen dar respuesta a cada uno de los casos en concreto para la manipulación de documentos. Una de las características más valorada en estos tipos de sistemas es su flexibilidad y posibilidad de integración con otras herramientas

Valhondo (2003) define que las funcionalidades básicas que suelen incluir los sistemas de gestión documental son:

- Sistemas y procedimientos para la incorporación masiva de documentos en diversos formatos electrónicos y procedentes de sistemas informáticos heterogéneos.
- Sistema de escaneo para la incorporación masiva o discreta de documentos que están en formatos no electrónicos.
- Gestión de documentos y sus metadatos.
- Referencia a documentos externos al propio sistema de gestión documental (fotos, videos, radiografías, etc.).
- Creación de relación entre los documentos y entre estos y el resto de componentes del sistema documental (carpetas, almacenes, metadatos, etc.)
- Mantenimiento de histórico de los documentos
- Localización de documentos mediante técnicas de búsqueda
- Edición de archivos utilizando sus herramientas nativas de creación.

- Disposición de un software para la visualización de archivos en los formatos más comunes.
- Control de acceso a documentos según perfiles de usuario.
- Anotación en documentos sin modificar el contenido de los mismos.
- Importación de información contenida en otras aplicaciones.

Por otro lado, no se puede pasar por alto el metadato como fuente fundamental para describir los datos contenidos en la Web Metadatos, y como propuesta en este estudio se encuentra el XML, que proporciona información sobre el significado de los datos, lo que facilita el procesamiento automático de la información contenida y la consiguiente gestión del conocimiento que se desea implementar como alternativa de uso factible para este trabajo de investigación.’

Metadatos

Valholdo (2003) los define como una definición o descripción de los datos. Si se tiene un documento de texto, puede describirse el mismo complementando los campos de: asunto, autor, título, resumen en palabras clave, etc., para definir el propio documento. Los valores de estos campos son metadatos del documento.

Los metadatos aportan una descripción de las fuentes materiales va partir de las cuales derivan los datos, los medios y los métodos empleados en su elaboración

La WC3 plantea que en un principio la World Wide Web se construyó para el uso humano, y a pesar de que todo en ella era legible por máquina estos datos todavía no son legibles por ésta última. Es muy difícil automatizar cualquier cosa en la Web, debido al volumen de información que contiene, no es posible gestionarla manualmente. La solución que se propone aquí es el uso de *metadatos* para describir los datos contenidos en la Web. Los metadatos son "datos sobre los datos" o concretamente en el contexto de esta especificación "datos que describen recursos Web". La distinción entre "datos" y "metadatos" no es incuestionable; es una

diferencia creada en primera instancia por una aplicación particular, y muchas veces el mismo recurso se interpretará de ambas formas (como dato y como metadato) simultáneamente.

La gestión de una biblioteca de objetos digitales requiere la gestión de metadatos sobre esos objetos. Los metadatos necesarios para gestionar y usar con éxito objetos digitales son más complejos que los que se emplean para gestionar colecciones de documentos impresos y materiales con soporte físico. Una biblioteca puede registrar metadatos descriptivos sobre un libro de su colección, pero el libro nunca se disolverá en una serie de páginas independientes, desconectadas, si la biblioteca no registra los metadatos estructurales relativos a la organización del libro; tampoco los usuarios se verán incapacitados para valorar la obra si la biblioteca no registra que el libro se produjo usando una prensa offset de un tipo determinado. Sin embargo, esto mismo no podría afirmarse para la versión digital de ese mismo libro. Sin metadatos estructurales, las imágenes y los archivos de texto que conforman el objeto digital tienen poca utilidad, y sin los metadatos técnicos relativos al proceso de digitalización los usuarios no pueden evaluar en qué medida la obra digital es un fiel reflejo del original impreso. Para la gestión interna, la biblioteca debe conocer los metadatos técnicos para poder refrescar y migrar regularmente los contenidos y asegurar la preservación de estos valiosos recursos.

XML (Extensible Markup Language)

Born (2002) lo define como un conjunto de reglas para definir etiquetas semánticas que nos organizan un documento en diferentes partes. XML es un metalenguaje que define la sintaxis utilizada para definir otros lenguajes de etiquetas estructurados.

Navarra (2006) indica que es necesario tener en cuenta los usos de XML para la gestión del conocimiento organizacional. Ya que este lenguaje presenta simultáneamente el contenido de forma que otro software pueda entender y usar este conocimiento. Gracias a que este proporciona información sobre el significado de los datos, lo que facilita el procesamiento automático de la información contenida y la consiguiente gestión del conocimiento.

Brochard (2001) indica que un documento XML está formado por elementos, marcadores, varios bloques, comentarios y atributos. Todos los documentos XML contienen un prologo o encabezado, un elemento raíz y un árbol. Este árbol está constituido por elementos anidados y por elementos adyacentes, y este tipo de documentos distingue las mayúsculas de las minúsculas.

Navarra (2006) indica que el *metalenguaje* XML consta de cuatro especificaciones

DTD (*Document Type Definition*): Definición del tipo de documento. Es, en general, un archivo/s que encierra una definición formal de un tipo de documento y, a la vez, especifica la estructura lógica de cada documento. Define tanto los elementos de una página como sus atributos. El DTD del XML es opcional. En tareas sencillas no es necesario construir una DTD, entonces se trataría de un documento "bien formado" (*well-formed*) y si lleva DTD será un documento "validado" (*valid*).

XSL (*eXtensible Stylesheet Language*): Define o implementa el lenguaje de estilo de los documentos escritos para XML. Desde el verano de 1997 varias empresas informáticas como Arbortext, Microsoft e Inso vienen trabajando en una propuesta de XSL (antes llamado "xml-style") que presentaron a W3C. Permite modificar el aspecto de un documento. Se puede lograr múltiples columnas, texto girado, orden de visualización de los datos de una tabla, múltiples tipos de letra con amplia variedad en los tamaños. Este estándar está basado en el lenguaje de

semántica y especificación de estilo de documento (DSSSL, *Document Style Semantics and Specification Language*, ISO/IEC 10179) y, por otro lado, se considera más potente que las hojas de estilo en cascada (CSS, *Cascading Style Sheets*), usado en un principio con el lenguaje DHTML. "Se espera que el CSS sea usado para visualizar simples estructuras de documentos XML (actualmente se ha conseguido mayor integración en XML con el protocolo CSS2 (*Cascading Style Sheets, level 2*) ofreciendo nuevas formas de composición y una más rápida visualización) y, por otra parte, XSL pueda ser utilizado donde se requiera más potencia de diseño como documentos XML que encierran datos estructurados (tablas, organigramas, etc.)⁽²⁾".

XLL (*eXtensible Linking Language*): Define el modo de enlace entre diferentes enlaces. Se considera que es un subconjunto de HyTime (*Hipermedia/Timed-based structuring Language* o Lenguaje de estructuración hipermedia/basado en el tiempo, ISO 10744) y sigue algunas especificaciones del TEI (*Text Encoding Initiative* o Iniciativa de codificación de texto). Desde marzo de 1998 el W3C trabajó en los enlaces y direccionamientos del XML. Provisionalmente se le renombró como *Xlink* y a partir de junio se le denomina XLL. Este lenguaje de enlaces extensible tiene dos importantes componentes: *Xlink* y el *Xpointer*. Va más allá de los enlaces simples que sólo soporta el HTML. Se podrá implementar con enlaces extendidos. Jon Bosak establece los siguientes mecanismos hipertextuales que soportará esta especificación:

- Denominación independiente de la ubicación.
- Enlaces que pueden ser también bidireccionales.
- Enlaces que pueden especificarse y gestionarse desde fuera del documento a los que se apliquen
- Hiperenlaces múltiples (anillos, múltiples ventanas, etc.).
- Enlaces agrupados (múltiples orígenes).
- Se pueden aplicar atributos a los enlaces (tipos de enlaces).

XUA (*XML User Agent*): Estandarización de navegadores XML. Todavía está en proceso de creación de borradores de trabajo. Se aplicará a los navegadores para que compartan todas las especificaciones XML.

Una vez definido el metadato que se implementará (XML) en este trabajo como alternativa para capturar información de otro sistema o software, también se tiene que definir el tipo de modelo cliente-servidor necesario para conectar las diferentes bibliotecas de los decanatos de la UCLA, como sistema multiusuario distribuido a través de una red de computadores, para trasladar el dato de una biblioteca a otra.

DIDL (*Digital Item Declaration Language*) proporciona un estándar XML para representar documentos digitales complejos, así como, a elementos multimedia.

MOA2 (*Making of America II*): es formato de codificación de metadatos descriptivos, administrativos y estructurales para obras textuales y basadas en imágenes.

METS: The Library of congress (2008) indica que este ofrece un medio flexible para codificar metadatos descriptivos, administrativos y estructurales para un objeto digital, y expresar las complejas relaciones entre estos tipos de metadatos. Ofrece un estándar útil para el intercambio de objetos digitales entre repositorios. Además, METS permite asociar objetos digitales con comportamientos o servicios.

METS o DIDL pueden ofrecer el marco general para metadatos de un objeto digital, aunque la aceptación más amplia del primero en la comunidad de bibliotecas hace que sea la opción preferida. Los metadatos descriptivos pueden gestionarse con el estándar Dublin Core o el más sofisticado MODS. Los metadatos técnicos, que dependen de los tipos de archivo que componen el objeto digital, se gestionan con estándares como MIX (imágenes fijas), AUDIOMD (archivos de audio), VIDEOMD o PBCORE (video) y Encabezamientos TEI (textos). La gestión de derechos puede hacerse con el esquema METS Rights o por esquemas más complejos como XrML o

ODRL. Los metadatos de preservación se gestiona mejor a través de los cuatro esquemas que forman parte del estándar PREMIS .

La integración de estos estándares usando el mecanismo namespace XML es técnicamente sencillo, aunque pueden surgir algunos problemas con namespaces que se definen con diferentes URIs, o como resultado de duplicaciones y consecuente redundancia entre esquemas: estos se pueden resolver mejor con guías de buenas prácticas, varias de las cuales están en construcción actualmente

Modelo Cliente-servidor

Tanenbaum (2003), indica que dentro de esta categoría entran por definición todos aquellos sistemas centralizados donde no existe agrupación lógica o la misma es realizada exclusivamente por el componente central, ni colocación física

Según Ralph y otros (2000), esta arquitectura consiste básicamente en que un programa, el cliente informático realiza peticiones a otro programa, el servidor, que les da respuesta. Aunque esta idea se puede aplicar a programas que se ejecutan sobre una sola computadora es más ventajosa en un sistema multiusuario distribuido a través de una red de computadores

En esta arquitectura la capacidad de proceso está repartida entre los clientes y los servidores, aunque son más importantes las ventajas de tipo organizativo debidas a la centralización de la gestión de la información y la separación de responsabilidades, lo que facilita y clarifica el diseño del sistema.

La separación entre cliente y servidor es una separación de tipo lógico, donde el servidor no se ejecuta necesariamente sobre una sola máquina ni es necesariamente un sólo programa.

Una disposición muy común son los sistemas multicapa en los que el servidor se descompone en diferentes programas que pueden ser ejecutados por diferentes computadoras aumentando así el grado de distribución del sistema.

La arquitectura cliente-servidor sustituye a la arquitectura monolítica en la que no hay distribución, tanto a nivel físico como a nivel lógico, ver figura 2.

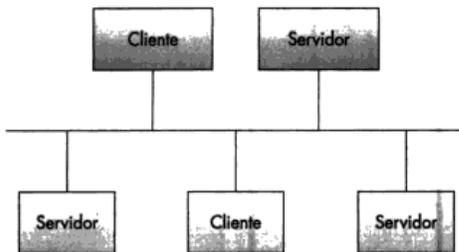


Figura 2. Conexión cliente/servidor

Fuente: Principios de sistemas de información: enfoque administrativo, 2000.

Autores: Ralph M. Stair, George W. Reynolds, Julio Coro Pando, Jorge L Blanco.

Ventajas de la arquitectura cliente-servidor

Centralización del control: los accesos, recursos y la integridad de los datos son controlados por el servidor de forma que un programa cliente defectuoso o no autorizado no pueda dañar el sistema.

Escalabilidad: se puede aumentar la capacidad de clientes y servidores por separado.

Se reduce el tráfico de red considerablemente. Idealmente, el cliente se comunica con el servidor utilizando un protocolo de alto nivel de abstracción como por ejemplo SQL.

Computación Grid

Joseph y otros (2004) describen como la necesidad de aprovechar los recursos disponibles en los sistemas informáticos conectados a Internet y simplificar su utilización ha dado lugar a una nueva forma de tecnología de la información conocida como *Grid Computing*. De este modo, los sistemas distribuidos se pueden emplear como un único sistema virtual en aplicaciones intensivas en datos o con gran demanda computacional.

Di Stefano (2005) señal que ha surgido como un framework para apoyar compilaciones complejas sobre datos grandes, aportando conjunto de recursos hardware y software distribuidos por Internet que proporcionan servicios accesibles por medio de un conjunto de protocolos e interfaces abiertos (gestión de recursos, gestión remota de procesos, librerías de comunicación, seguridad, soporte a monitorización, etc.). Las organizaciones virtuales que se interconectan por medio de un Grid tienen que mantener sus propias políticas de seguridad y gestión de recursos. Esto significa que la tecnología usada para construir un Grid es complementaria a otras tecnologías aprovechando los recursos distribuidos en la intranet de una organización.

Siguiendo el esquema de búsqueda de las herramientas y aplicaciones existentes en la actualidad para dar respuestas a tecnologías ligeras más eficaces, que satisfagan su empleo en este proyecto de investigación, se utilizará el Peer-to-Peer (P2P), como esquema de comunicación para conectar aplicaciones distribuidas en tiempo real. Para lo cual se hace necesario definirlo, caracterizarlo y especificar las razones que lo llevan a considerarse como una arquitectura robusta y adecuada.

Peer-to-Peer (P2P)

Mendoza (2005), indica que el término Peer-to-Peer (P2P) es utilizado para referirse a un esquema de comunicación basado en diferentes tecnologías que permite

que una computadora en red pueda interactuar con otra sin necesidad de la intervención de un servidor central. Este esquema de comunicación es muy conveniente cuando se requiere que sistemas heterogéneos se interconecten para realizar diferentes transacciones de manera de aprovechar y compartir eficazmente recursos distribuidos. Dado que en el mercado existe una amplia gama de tecnologías Middleware para conectar aplicaciones distribuidas en tiempo real como las que se pueden conseguir a través de aplicaciones P2P empotradas, el objetivo de esta investigación es orientar a los Ingenieros de Software en la selección más adecuada de estas tecnologías al momento de requerir diseñar la interconexión de Sistemas P2

Según Theotokis (2004), las arquitecturas P2P han sido utilizadas para una variedad de aplicaciones diferentes, entre las cuales se tienen:

Comunicación y Colaboración. Esta categoría incluye sistemas que proveen la infraestructura para facilitar la comunicación y la colaboración directa, normalmente en tiempo real, entre las computadoras a la par. En esta categoría se incluyen los chat y las aplicaciones de la mensajería instantáneas.

Computación distribuida. Esta categoría incluye sistemas cuyo objetivo es tomar ventaja de la disponibilidad del poder del proceso de computación (ciclos del CPU) del par. Esto se logra desagregando las tareas intensivas de computación en unidades de trabajo pequeñas y distribuyéndolas entre computadoras pares diferentes que ejecutan su unidad de trabajo correspondiente y devuelven los resultados. La coordinación central es requerida, principalmente para la separación y distribución de las tareas y para recolectar los resultados.

Soporte de servicios de Internet. Varias aplicaciones diferentes basadas en las infraestructuras P2P han surgido para apoyar una variedad de servicios de Internet. Los ejemplos de tales aplicaciones incluyen los sistemas de multidifusión P2P y

aplicaciones de seguridad, proporcionando protección contra el rechazo del servicio o ataques de virus.

Sistemas de bases de datos. Se ha realizado un trabajo considerable en diseñar sistemas de bases de datos distribuidos basados en infraestructuras P2P. Ejemplo de esto es el Modelo Correlativo Local (LRM), el motor de búsqueda PIER, el sistema Piazza que proporciona datos (por ejemplo de una base de datos relacional), esquemas (u ontologías) o ambos.

Distribución de contenido. La mayoría de los sistemas actuales P2P cae dentro de esta categoría, ya que incluyen sistemas e infraestructuras diseñadas para compartir de medios digitales y otros datos entre los usuarios. El rango de sistemas de distribución de contenidos va desde aplicaciones para la simple distribución directa de archivos a sistemas más sofisticados que crean un medio del almacenamiento distribuido para publicar, buscar, actualizar y recuperar datos de manera segura, eficaz, organizada e indexada.

Backx (2002), asevera que todas las arquitecturas P2P tienen una cosa en común: la transferencia datos siempre es P2P: una conexión directa de datos entre el par que ofrece el archivo y el que lo requiere. Sin embargo, el control se lleva a cabo de varias maneras, como se puede apreciar en la figura 3.

Arquitectura mediada. Usa un esquema cliente-servidor para sus operaciones de control. Todos los pares se conectan a un servidor central que maneja el archivo y las bases de datos del usuario. Se envían búsquedas para un archivo al servidor; si se encuentra, el archivo puede descargarse directamente del par. En la mayoría de los casos el servidor tendrá una base de datos de archivos compartido por los pares.

Arquitectura pura P2P. Las aplicaciones puras P2P no usarán un servidor central en absoluto (excepto posiblemente para acceder la red). Pueden enviarse búsquedas para los archivos a través de la red o pueden usarse mecanismos más

inteligentes. Las redes puras P2P no han sido populares porque ellas generan mucho tráfico que debe mantenerse arriba y funcionando.

Arquitecturas híbridas. Las arquitecturas híbridas son el último desarrollo en la comunidad P2P. Su meta es ofrecer lo mejor de las propiedades de las arquitecturas mediadas y las puras, a través de la introducción de los llamados ultrapeers. Los ultrapeer realizarán la tarea de un servidor en la arquitectura mediada, pero para sólo un subconjunto de los pares. Los ultrapeers se conectan a través de una red pura P2P. Así, las arquitecturas híbridas introducen dos capas para el control: uno de conexión de pares a los ultrapeers de forma cliente-servidor y uno de conexión entre ultrapeers vía una red pura P2P.

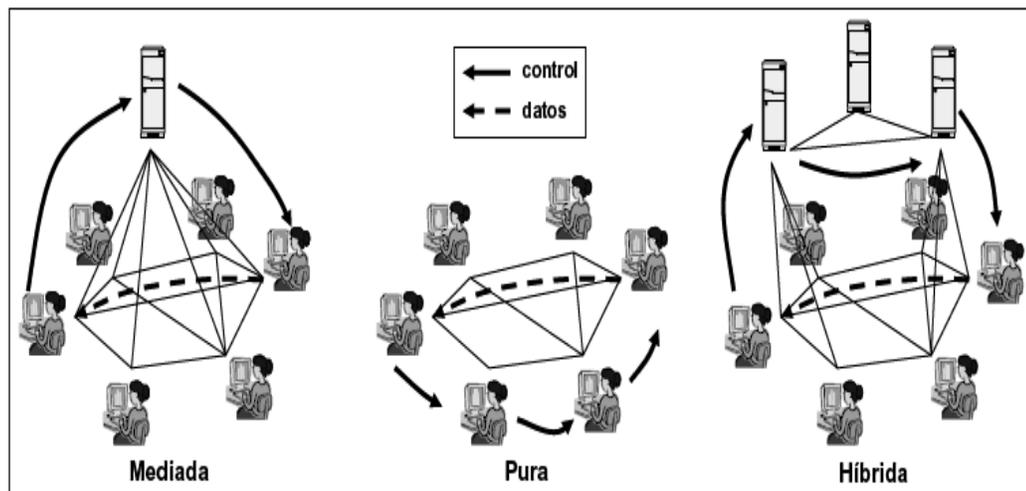


Figura 3, Las tres arquitecturas P2P

Fuente: A comparison of peer-to-peer architectures

Autores: Backx, P., Wauters, T., Dhoedt, B., Demeester, P.

Agentes Inteligentes

Sánchez (2007) haciendo referencia Wooldridge (2000), quien plantea que Un agente es un sistema computarizado que está situado en algún entorno, y que es capaz de actuar de forma autónoma en este entorno para satisfacer sus objetivos de diseño, mostrando como características más resaltantes: la autonomía, su localización en

algún medio con el que tiene que interactuar, la flexibilidad y la actuación en representación de otros, además de enfatizar nuevamente la idea de que todo agente realiza las acciones pertinentes con el único propósito de alcanzar unos objetivos predefinidos.

Wooldridge (2000) destaca que todo agente, para llegar a ser denominado “inteligente”, debe satisfacer el siguiente conjunto mínimo de propiedades:

- **Reactividad:** los agentes inteligentes son capaces de percibir su entorno y responder en un periodo de tiempo adecuado a cambios que ocurren en este para satisfacer sus objetivos de diseño.
- **Proactividad:** los agentes inteligentes son capaces de mostrar un comportamiento dirigido por objetivos tomando la iniciativa para satisfacer sus objetivos de diseño.
- **Habilidad social:** los agentes inteligentes son capaces de interactuar con otros agentes (y posiblemente humanos) para satisfacer sus objetivos de diseño.

Sistemas Multi-Agente

Según Sanchez (2007), los agentes pueden ser útiles como entidades independientes en entornos aislados a las que se les delegan ciertas tareas repetitivas y que se pueden automatizar en representación de unos determinados usuarios. Sin embargo, en la mayoría de los casos, los agentes se encuentran en entornos que contienen otros agentes constituyendo de este modo un Sistema Multi-Agente (SMA), esto es, un sistema constituido por un grupo de agentes que pueden interactuar.

Cuando un grupo de agentes individuales forma un SMA, surge la necesidad de disponer de un mecanismo para coordinar dicho grupo de agentes y de un lenguaje para permitir la comunicación entre ellos. Entre los mecanismos de coordinación, se

pueden distinguir los casos en los que los agentes tienen objetivos comunes y, por tanto, *cooperan*, de los casos en los que los agentes son “auto-interesados” y tienen objetivos conflictivos con los demás, para lo cual precisaremos de mecanismos de *negociación*.

En el caso de agentes cooperativos, los *mecanismos de cooperación* más comunes son las estructuras organizacionales, la planificación multi-agente (centralizada y distribuida), redes de contratos y cooperación funcionalmente exacta. Por otro lado, para el caso de agentes competitivos, se hace necesario un *mecanismo de negociación*. Entre los tipos de mecanismos de negociación más utilizados en la literatura destaca la formación de coaliciones, los mecanismos de mercados, la teoría del regateo, la votación, las subastas y la asignación de tareas entre dos agentes.

Los patrones en los sistemas tecnológicos son de gran ayuda en la reutilización de diseños y arquitecturas con soluciones más ágiles, ya que los mismos permiten ayudar a los desarrolladores a comprender un sistema más rápidamente cuando está documentado con los patrones que se han usado. Es este trabajo se consideraron debido a su gran utilidad dentro del campo de la programación. Algunos de los autores defensores de los mismos se especifican a continuación.

Patron

Valera (2005) lo describe como un modelo a seguir que describe una solución exitosa de un problema particular en un contexto dado. Solucionando problemas recurrentes, por lo que, una de las ventajas de los patrones es fomentar y facilitar el rehusó de soluciones.

Paez (2006) lo expone que es una regla que consta de tres partes básicas, un problema, una solución y el contexto donde se desarrolló; y si el problema es frecuente, proporciona la certeza de que la solución propuesta es ampliamente buena y es garantía para resolver ese problema. De esta manera, Los patrones facilitan la

reutilización de diseños y arquitecturas de software que han tenido éxito, además capturan la experiencia y la hacen accesible a los no expertos, se pueden agrupar y el conjunto de sus nombres forma un vocabulario que ayuda a que los desarrolladores se comuniquen mejor. Ayudan a los desarrolladores a comprender un sistema más rápidamente cuando está documentado con los patrones que usa.

Clasificación de los Patrones

Buschmann (1996), clasifica los patrones en tres grandes grupos, estos en cada grupo varían respecto a su nivel de detalle y abstracción, tal como se muestra a continuación (ver figura 4):



Figura 4, patrones según el nivel de detalle

Fuente: Pattern Oriented Software Architecture, Volume 1, 1996

Autor: Buschmann

Patrones de Arquitectura ó Arquitecturales

Según Valera (2005) el cual cita a Welicki los patrones de arquitectura expresan el esquema fundamental de organización para sistemas de software. Proveen un conjunto de subsistemas predefinidos; especifican sus responsabilidades e incluyen reglas y guías para organizar las relaciones entre ellos. Ayudando a especificar la estructura fundamental de una aplicación.

Estos expresan un paradigma fundamental para estructurar un sistema software y proporcionan un conjunto de subsistemas predefinidos, que especifica sus responsabilidades, e. incluye reglas y guías para organizar las relaciones entre ellos.

Patrones de diseño

Gómez y otros (2003) los describen como un esquema para refinar sus subsistemas, componentes o relaciones entre ellos. Describiendo la estructura de una solución a un problema que aparece repetitivamente, así como, de componentes que se comunican entre estos; resolviendo un problema de diseño general en un contexto particular

Patrones elementales (idiomas)

Páez (2006) describe un patrón de idiomas como un patrón de bajo nivel, específico a un lenguaje de programación. Un patrón de idiomas describe cómo llevar a cabo aspectos particulares de componentes o las relaciones entre ellos, usando las características de lenguaje de programación utilizado.

Características de un buen patrón

Documentar buenos patrones puede ser una tarea muy difícil. Citando a Coplien citado por Hillside Group 2003, un buen patrón:

- **Resuelve un problema:** Los patrones capturan soluciones, no principios o estrategias abstractas.
- **Es un concepto probado:** Capturan soluciones, no teorías o especulaciones. En el caso del “Design Patterns”, el criterio para decidir si algo era un patrón o no, era que éste debía tener al menos 3 implementaciones reales.
- **La solución no es obvia:** Muchas técnicas de resolución de problemas (como los paradigmas o métodos de diseño de software) intentan derivar

soluciones desde principios básicos. Los mejores patrones generan una solución a un problema indirectamente (un enfoque necesario para los problemas de diseño más difíciles).

- **Describe una relación:** Los patrones no describen módulos sino estructuras y mecanismos.
- **Tiene un componente humano significativo:** El software sirve a las personas. Los mejores patrones aplican a la estética y a las utilidades (de hecho, no es casual que varios de los primeros lenguajes de patrones tengan que ver con temas estéticos y utilidades como **Hot Draw** ó **ET++**).

No se puede dejar pasar el lenguaje en que se desarrolla este trabajo, para lo cual se consideró el UML, por ser grafico y permite modelar aplicaciones en diversos dominios de aplicación como se define en el siguiente párrafo.

Unified Modeling Language (UML)

Entre los lenguajes de modelado que define OMG (Object Management Group) el más conocido y usado es sin duda UML (Unified Modeling Language). UML es un lenguaje gráfico para especificar, construir y documentar los artefactos que modelan un sistema.

Cueva (1999), indica que fue diseñado para ser un lenguaje de modelado de propósito general, por lo que puede utilizarse para especificar la mayoría de los sistemas basados en objetos o en componentes, y para modelar aplicaciones de muy diversos dominios de aplicación (telecomunicaciones, comercio, sanidad, etc.) y plataformas de objetos distribuidos (como por ejemplo J2EE, .NET o CORBA).

El hecho de que UML sea un lenguaje de propósito general proporciona una flexibilidad y expresividad a la hora de modelar sistemas. Sin embargo, hay numerosas ocasiones en las que es mejor contar con algún lenguaje más específico

para modelar y representar los conceptos de ciertos dominios particulares. Esto sucede, por ejemplo, cuando la sintaxis o la semántica de UML no permiten expresar los conceptos específicos del dominio, o cuando se desea restringir y especializar los constructores propios de UML, que suelen ser demasiado genéricos y numerosos.

OMG define dos posibilidades a la hora de definir lenguajes específicos de dominio, y que se corresponden con las dos situaciones mencionadas antes: o bien se define un nuevo lenguaje (alternativa de UML), o bien se extiende el propio UML, especializando algunos de sus conceptos y restringiendo otros, pero respetando la semántica original de los elementos de UML (clases, asociaciones, atributos, operaciones, transiciones, entre otros).

Cada una de estas dos alternativas presenta ventajas e inconvenientes. Así, definir un nuevo lenguaje ad-hoc permite mayor grado de expresividad y correspondencia con los conceptos del dominio de aplicación particular, obteniéndose de esta forma una solución a medida.

Hurtado (2008), indica que Los elementos básicos de UML, aquellos que representan principalmente las partes estáticas del sistema, son:

- Nodos
- Paquetes

Las relaciones que se utilizan para establecer conexiones entre los elementos son:

- Dependencia
- Asociación
- Generalización
- Realización

- Diagramas UML

Los elementos y relaciones se agrupan en diagramas que representan diferentes aspectos del sistema, los cuales se pueden apreciar en la figura 5. Los diagramas de UML son

- *Diagrama de clases*: Presenta las clases, junto con sus atributos, operaciones, interfaces y relaciones. También presenta el agrupamiento de clases en paquetes y las relaciones entre ellos.
- *Diagrama de objetos*: Muestra instancias de clases (objetos) con valores en sus atributos y relaciones.
- *Diagrama de casos de uso*: Los escenarios de uso del sistema, incluyendo los roles de los usuarios.
- *Diagramas de interacción*: Comprende los diagramas de secuencia y de colaboración. Presenta objetos y relaciones entre ellos desde el punto de vista dinámico.
- *Diagrama de estado*: Representa los posibles estados, eventos y transiciones entre las clases u objetos.
- *Diagrama de componentes*: Organización y dependencia entre componentes en los dispositivos de hardware.

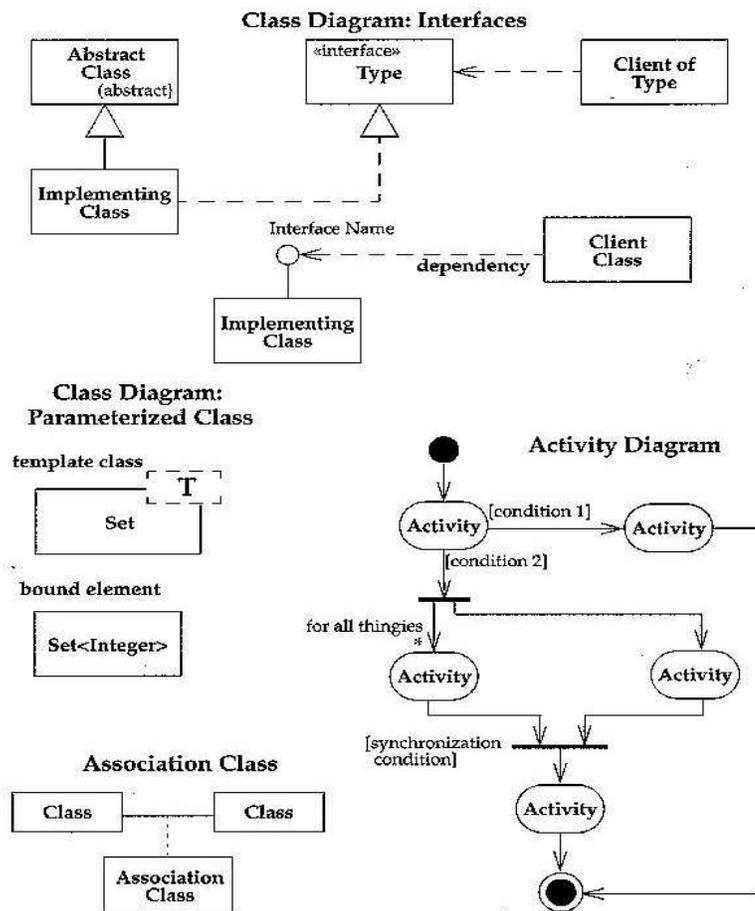


Figura 5, UML, Resumen de la notación
Fuente: Lenguaje para modelar objetos, 1999
Autor: Juan Manuel Cueva Lovelle

Estructuras de arquitectura en UML

Hurtado (2008), indica que la primera fase de un proyecto es la identificación de los componentes que participan en la descripción de la arquitectura de un sistema, y luego sus relaciones en las diferentes estructuras.

Para cada componente y conector se determinan los elementos de UML que los representan, con su sintaxis y semántica. Algunos componentes o estructuras no

tendrán una representación directa y en este caso se utilizarán los mecanismos de extensión que provee UML, como estereotipos o restricciones.

Como parte integral de cada estructura se deben incluir restricciones adicionales que determinan las relaciones y los tipos de componentes y conectores que pueden aparecer en dicha estructura.

Por último en el proyecto se presentan las relaciones existentes entre las diferentes estructuras y la manera de verificar dichas relaciones, lo que ayudará a la persona que modela la arquitectura del sistema a validar la consistencia de esta última.

Arquitectura del software

Clements (1996.), describe la arquitectura de software como una vista del sistema que incluye los componentes principales del mismo, la conducta de esos componentes según se la percibe desde el resto del sistema y las formas en que los componentes interactúan y se coordinan para alcanzar la misión del sistema. La vista arquitectónica es una vista abstracta, aportando el más alto nivel de comprensión y la supresión o diferimiento del detalle inherente a la mayor parte de las abstracciones.

Desde puntos de vista implícitos o explícitos, clasifican los modelos de esta forma:

- 1) Modelos estructurales:** Sostienen que la AS está compuesta por componentes, conexiones entre ellos y (usualmente) otros aspectos tales como configuración, estilo, restricciones, semántica, análisis, propiedades, racionalizaciones, requerimientos, necesidades de los participantes. El trabajo en esta área está caracterizado por el desarrollo de lenguajes de descripción arquitectónica (ADLs).

- 2) **Modelos de framework:** Son similares a la vista estructural, pero su énfasis primario radica en la (usualmente una sola) estructura coherente del sistema completo, en vez de concentrarse en su composición. Los modelos de framework a menudo se refieren a dominios o clases de problemas específicos. El trabajo que ejemplifica esta variante incluye arquitecturas de software específicas de dominios, como CORBA, o modelos basados en CORBA, o repositorios de componentes específicos, como PRISM.
- 3) **Modelos dinámicos:** Enfatizan la cualidad conductual de los sistemas. “Dinámico” puede referirse a los cambios en la configuración del sistema, o a la dinámica involucrada en el progreso de la computación, tales como valores cambiantes de datos.
- 4) **Modelos de proceso:** Se concentran en la construcción de la arquitectura, y en los pasos o procesos involucrados en esa construcción. En esta perspectiva, la arquitectura es el resultado de seguir un argumento (script) de proceso. Esta vista se ejemplifica con el actual trabajo sobre programación de procesos para derivar arquitecturas.
- 5) **Modelos funcionales:** Una minoría considera la arquitectura como un conjunto de componentes funcionales, organizados en capas que proporcionan servicios hacia arriba. Es tal vez útil pensar en esta visión como un framework particular.

Ninguna de estas vistas excluye a las otras, ni representa un conflicto fundamental sobre lo que es o debe ser la AS. Por el contrario, representan un espectro en la comunidad de investigación sobre distintos énfasis que pueden aplicarse a la arquitectura: sobre sus partes constituyentes, su totalidad, la forma en que se comporta una vez construida, o el proceso de su construcción.

Arquitectura basada en patrones.

Shaw (1996), indica que si bien reconoce la importancia de un modelo emanado históricamente del diseño orientado a objetos, esta corriente surgida hacia 1996 no se encuentra tan rígidamente vinculada a UML en el modelado, ni a CMM en la metodología. El texto sobre patrones que esta variante reconoce como referencia lo indicado por Buschmann y otros (1996). En esta manifestación de la Arquitectura de Software donde prevalece cierta tolerancia hacia modelos de proceso tácticos, no tan macroscópicos y las premisas de la programación extrema. El diseño consiste en identificar y articular patrones preexistentes, que se definen en forma parecida a los estilos de arquitectura.

La metodología empleada es la XP, descrita posteriormente.

Metodología XP

Amaro y otros (2007) describen a XP es una metodología ágil centrada en potenciar las relaciones interpersonales como clave para el éxito en el desarrollo de software, promoviendo el trabajo en equipo, preocupándose por el aprendizaje de los desarrolladores, y propiciando un buen clima de trabajo. XP se basa en realimentación 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. XP se define como especialmente adecuada para proyectos con requisitos imprecisos y muy cambiantes, y donde existe un alto riesgo técnico. A continuación presentaremos las características esenciales de XP organizadas en los tres apartados siguientes: historias de usuario, roles, proceso y prácticas.

Roles en XP

Los roles de acuerdo con la propuesta original de Beck son:

- *Programador*. El programador escribe las pruebas unitarias y produce el *código del sistema*.
- *Cliente*. Escribe las historias de usuario y las pruebas funcionales para validar su implementación. Además, asigna la prioridad a las historias de usuario y decide cuáles se implementan en cada iteración centrándose en aportar mayor valor al negocio.
- *Encargado de pruebas (Tester)*. Ayuda al cliente a escribir las pruebas funcionales. Ejecuta las pruebas regularmente, difunde los resultados en el equipo y es responsable de las herramientas de soporte para pruebas.
- *Encargado de seguimiento (Tracker)*. Proporciona realimentación al equipo. Verifica el grado de acierto entre las estimaciones realizadas y el tiempo real dedicado, para mejorar futuras estimaciones. Realiza el seguimiento del progreso de cada iteración.
- *Entrenador (Coach)*. Es responsable d proceso global. Debe proveer guías al equipo de forma que se apliquen las prácticas XP y se siga el proceso correctamente.
- *Consultor*. Es un miembro externo del equipo con un conocimiento específico en algún tema necesario para el proyecto, en el que puedan surgir problemas.
- *Gestor (Big boss)*. Es el vínculo entre clientes y programadores, ayuda a que.

Procesos en XP

El ciclo de desarrollo consiste (a grandes rasgos) en los siguientes pasos:

1. El cliente define el valor de negocio a implementar.
2. El programador estima el esfuerzo necesario para su implementación.
3. El cliente selecciona qué construir, de acuerdo con sus prioridades y las restricciones de tiempo.

4. El programador construye ese valor de negocio.
5. Vuelve al paso 1.

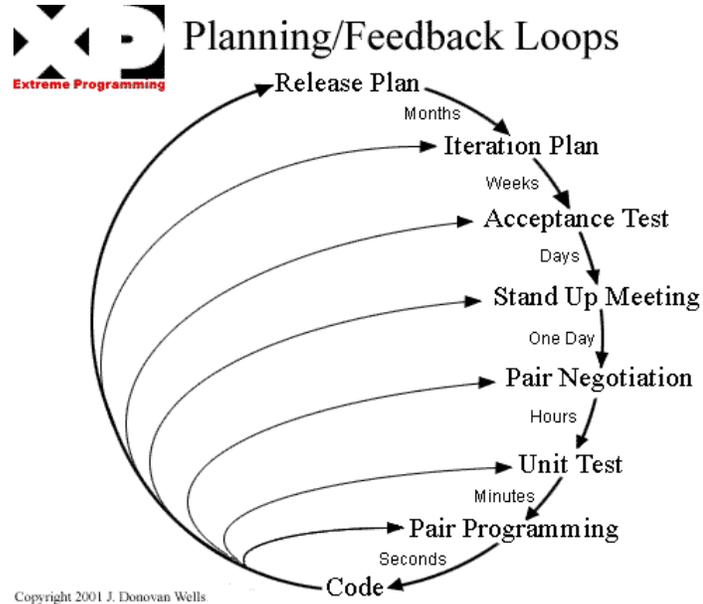


Figura 6. Ciclos en XP
Fuentes: The success of the XP. 2001
Autor: Donovan Wells

En todas las iteraciones de este ciclo tanto el cliente como el programador aprenden. No se debe presionar al programador a realizar más trabajo que el estimado, ya que se perderá calidad en el software o no se cumplirán los plazos. De la misma forma el cliente tiene la obligación de manejar el ámbito de entrega del producto, para asegurarse que el sistema tenga el mayor valor de negocio posible con cada iteración.

El ciclo de vida ideal de XP consiste de seis fases: Exploración, Planificación de la Entrega (Release), Iteraciones, Producción, Mantenimiento y Muerte del Proyecto como se puede apreciar en la figura 6.



Extreme Programming Project

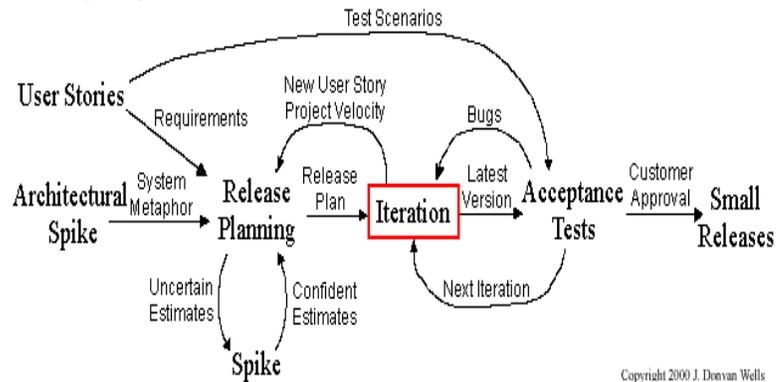


Figura 7. Fases de un proyecto en XP

Fuentes: The success of the XP. 2001

Autor: Donovan Wells

A continuación describiremos las diferentes fases de desarrollo en la metodología XP como se puede apreciar en la figura 7.

1. Exploración

En esta fase, los clientes plantean a grandes rasgos las historias de usuario que son de interés para la primera entrega del producto. Al mismo tiempo el equipo de desarrollo se familiariza con las herramientas, tecnologías y prácticas que se utilizarán en el proyecto. Se prueba la tecnología y se exploran las posibilidades de la arquitectura del sistema construyendo un prototipo. La fase de exploración toma de pocas semanas a pocos meses, dependiendo del tamaño y familiaridad que tengan los programadores con la tecnología.

2. Planificación de la Entrega (Release)

En esta fase el cliente establece la prioridad de cada historia de usuario, y correspondientemente, los programadores realizan una estimación del esfuerzo

necesario de cada una de ellas. Se toman acuerdos sobre el contenido de la primera entrega y se determina un cronograma en conjunto con el cliente. Una entrega debería obtenerse en no más de tres meses. Esta fase dura unos pocos días. Las estimaciones de esfuerzo asociado a la implementación de las historias la establecen los programadores utilizando como medida el punto. Un punto, equivale a una semana ideal de programación. Las historias generalmente valen de 1 a 3 puntos. Por otra parte, el equipo de desarrollo mantiene un registro de la “velocidad” de desarrollo, establecida en puntos por iteración, basándose principalmente en la suma de puntos correspondientes a las historias de usuario que fueron terminadas en la última iteración. La planificación se puede realizar basándose en el tiempo o el alcance. La velocidad del proyecto es utilizada para establecer cuántas historias se pueden implementar antes de una fecha determinada o cuánto tiempo tomará implementar un conjunto de historias. Al planificar por tiempo, se multiplica el número de iteraciones por la velocidad del proyecto, determinándose cuántos puntos se pueden completar. Al planificar según alcance del sistema, se divide la suma de puntos de las historias de usuario seleccionadas entre la velocidad del proyecto, obteniendo el número de iteraciones necesarias para su implementación.

3. Iteraciones

Esta fase incluye varias iteraciones sobre el sistema antes de ser entregado. El Plan de Entrega está compuesto por iteraciones de no más de tres semanas. En la primera iteración se puede intentar establecer una arquitectura del sistema que pueda ser utilizada durante el resto del proyecto. Esto se logra escogiendo las historias que fueren la creación de esta arquitectura, sin embargo, esto no siempre es posible ya que es el cliente quien decide qué historias se implementarán en cada iteración (para maximizar el valor de negocio). Al final de la última iteración el sistema estará listo para entrar en producción. Los elementos que deben tomarse en cuenta durante la elaboración del Plan de la Iteración son: historias de usuario no abordadas, velocidad del proyecto, pruebas de aceptación no superadas en la iteración anterior y tareas no

terminadas en la iteración anterior. Todo el trabajo de la iteración es expresado en tareas de programación, cada una de ellas es asignada a un programador como responsable, pero llevadas a cabo por parejas de programadores.

4. Producción

La fase de producción requiere de pruebas adicionales y revisiones de rendimiento antes de que el sistema sea trasladado al entorno del cliente. Al mismo tiempo, se deben tomar decisiones sobre la inclusión de nuevas características a la versión actual, debido a cambios durante esta fase. Es posible que se rebaje el tiempo que toma cada iteración, de tres a una semana. Las ideas que han sido propuestas y las sugerencias son documentadas para su posterior implementación (por ejemplo, durante la fase de mantenimiento).

5. Mantenimiento

Mientras la primera versión se encuentra en producción, el proyecto XP debe mantener el sistema en funcionamiento al mismo tiempo que desarrolla nuevas iteraciones. Para realizar esto se requiere de tareas de soporte para el cliente. De esta forma, la velocidad de desarrollo puede bajar después de la puesta del sistema en producción. La fase de mantenimiento puede requerir nuevo personal dentro del equipo y cambios en su estructura.

6. Muerte del Proyecto

Es cuando el cliente no tiene más historias para ser incluidas en el sistema. Esto requiere que se satisfagan las necesidades del cliente en otros aspectos como rendimiento y confiabilidad del sistema. Se genera la documentación final del sistema y no se realizan más cambios en la arquitectura. La muerte del proyecto también ocurre cuando el sistema no genera los beneficios esperados por el cliente o cuando no hay presupuesto para mantenerlo.

Prácticas en XP

La principal suposición que se realiza en XP es la posibilidad de disminuir la mítica curva exponencial del costo del cambio a lo largo del proyecto, lo suficiente para que el diseño evolutivo funcione. Esto se consigue gracias a las tecnologías disponibles para ayudar en el desarrollo de software y a la aplicación disciplinada de las siguientes prácticas.

Schenone (2004) describe Los principios de XP de Beck:

- *El juego de Planeamiento*: Rápidamente determinar el alcance del próximo release mediante la combinación de prioridades del negocio y estimaciones técnicas. A medida que la realidad va cambiando el plan, actualizar el mismo.
- *Pequeños Releases*: Poner un sistema simple en producción rápidamente, luego liberar nuevas versiones en ciclos muy cortos.
- *Metáfora*: Guiar todo el desarrollo con una historia simple y compartida de cómo funciona todo el sistema.
- *Diseño Simple*: El sistema deberá ser diseñado tan simple como sea posible en cada momento. Complejidad extra es removida apenas es descubierta.
- *Testing*: Los programadores continuamente escriben pruebas unitarias, las cuales deben correr sin problemas para que el desarrollo continúe. Los clientes escriben pruebas demostrando que las funcionalidades están terminadas.
- *Refactoring*: Los programadores reestructuran el sistema sin cambiar su comportamiento para remover duplicación, mejorar la comunicación, simplificar, o añadir flexibilidad.
- *Programación de a Pares*: Todo el código de producción es escrito por dos programadores en una máquina.

- *Propiedad Colectiva del Código:* Cualquiera puede cambiar código en cualquier parte del sistema en cualquier momento.
- *Integración Continua:* Integrar y hacer builds del sistema varias veces por día, cada vez que una tarea se completa.
- *Semana de 40-horas:* Trabajar no más de 40 horas semanales como regla. Nunca trabajar horas extras durante dos semanas consecutivas.
- *Cliente en el lugar de Desarrollo:* Incluir un cliente real en el equipo, disponible de forma full-time para responder preguntas.

Las Historias de Usuario

Canós y otros (2008), la describe como la técnica utilizada para especificar los requisitos del software. Se trata de tarjetas de papel en las cuales el cliente describe brevemente las características que el sistema debe poseer, sean requisitos funcionales o no funcionales. El tratamiento de las historias de usuario es muy dinámico y flexible. Cada historia de usuario es lo suficientemente comprensible y delimitada para que los programadores puedan implementarla en unas semanas.

Marcos de trabajo (Frameworks)

Cabrera (2005) lo describe como la reutilización de las arquitecturas de software se definen dentro del marco de trabajo (framework), el cual es un patrón arquitectónico que proporciona una plantilla extensible para aplicaciones dentro de un dominio. De hecho, se puede pensar en un marco de trabajo como una microarquitectura que incluye un conjunto de mecanismos que colaboran para resolver un problema en común en un dominio común. Cuando se especifica un marco de trabajo, se especifica el esqueleto de una arquitectura, junto a los elementos variables, que se muestran a los usuarios que quieren adaptar el marco de trabajo a su propio.

Spring framework

Spring es un framework de aplicaciones Java/J2EE, que proporciona las siguientes características:

- Una potente gestión de configuración basada en JavaBeans, aplicando los principios de Inversión de Control (*IoC*). Esto hace que la configuración de aplicaciones sea rápida y sencilla. Ya no es necesario tener *singletons* ni ficheros de configuración, una aproximación consistente y elegante. Esta factoría de *beans* puede ser usada en cualquier entorno, desde *applets* hasta contenedores J2EE. Estas definiciones de *beans* se realizan en lo que se llama el contexto de aplicación.
- Una capa genérica de abstracción para la gestión de transacciones, permitiendo gestores de transacción enchufables (*pluggables*), y haciendo sencilla la demarcación de transacciones sin tratarlas a bajo nivel. Se incluyen estrategias genéricas para JTA y un único JDBC DataSource. En contraste con el JTA simple o EJB CMT, el soporte de transacciones de Spring no está atado a entornos J2EE.
- Una capa de abstracción JDBC que ofrece una significativa jerarquía de excepciones (evitando la necesidad de obtener de SQLException los códigos que cada gestor de base de datos asigna a los errores), simplifica el manejo de errores, y reduce considerablemente la cantidad de código necesario.
- Integración con Hibernate, JDO e iBatis SQL Maps en términos de soporte a implementaciones DAO y estrategias con transacciones. Especial soporte a Hibernate añadiendo convenientes características de *IoC*, y solucionando muchos de los comunes problemas de integración de Hibernate. Todo ello cumpliendo con las transacciones genéricas de Spring y la jerarquía de excepciones DAO.
- Funcionalidad AOP, totalmente integrada en la gestión de configuración de Spring. Se puede aplicar AOP a cualquier objeto gestionado por Spring,

añadiendo aspectos como gestión de transacciones declarativa. Con Spring se puede tener gestión de transacciones declarativa sin EJB, incluso sin JTA, si se utiliza una única base de datos en un contenedor web sin soporte JTA.

- Un framework MVC (*Model-View-Controller*), construido sobre el núcleo de Spring. Este framework es altamente configurable vía interfaces y permite el uso de múltiples tecnologías para la capa vista como pueden ser JSP, Velocity, Tiles, iText o POI. De cualquier manera una capa modelo realizada con Spring puede ser fácilmente utilizada con una capa web basada en cualquier otro framework MVC, como Struts, WebWork o Tapestry.

Toda esta funcionalidad puede usarse en cualquier servidor J2EE, y la mayoría de ella ni siquiera requiere su uso. El objetivo central de Spring es permitir que objetos de negocio y de acceso a datos sean reusables, no atados a servicios J2EE específicos. Estos objetos pueden ser reutilizados tanto en entornos J2EE (web o EJB), aplicaciones standalone, entornos de pruebas,... sin ningún problema.

La arquitectura en capas de Spring ofrece cantidad de flexibilidad. Toda la funcionalidad está construida sobre los niveles inferiores. Por ejemplo se puede utilizar la gestión de configuración basada en JavaBeans sin utilizar el framework MVC o el soporte AOP.

Lucene

Gospodnetic y otro (2005), describe a Lucene como un API open source, para la indexación y búsqueda de documentos o datos no estructurados, presentando un alto performance, escalabilidad y avanzadas tecnologías de búsquedas.

Para aplicar los modelos de recuperación y búsqueda Lucene requiere de indexar los documentos estructurando los mismos y permitiendo hacer todo tipos de búsquedas rápidas y aplicación de queries de acuerdo a dicha indexación.

un índice Lucene contiene una secuencia de documentos, un documento una secuencia de campos, un campo es una de términos nombrados, y un término es una secuencia de texto String identificado por el nombre del campo.

JXTA

El proyecto JXTA es un conjunto de protocolos que fueron diseñados para solventar mucho de los problemas de la computación distribuida, especialmente de las redes peer-to-peer, o más conocidas como aplicaciones P2P. La primera implementación 1.0 fue liberada en 2001, y a día de hoy, la última implementación de JXTA corresponde a la versión 2.5. Básicamente, lo que ofrece JXTA, es un marco de trabajo que facilita el desarrollo e implementación de aplicaciones P2P.

Algunas de las ventajas de usar JXTA para el desarrollo de aplicaciones P2P:

1. Permite encontrar a otros nodos en la red de forma dinámica a través incluso de cortafuegos y NATS.
2. Facilidad para compartir recursos entre los nodos
3. Crear grupos de nodos que tengan algún objetivo en común
4. Poder monitorizar las actividades de los nodos de forma remota
5. Establecer comunicaciones seguras entre los nodos
6. Independencia del sistema operativo, topología de red, lenguaje de programación o hardware
7. Ventajas propias de utilizar el modelo de red P2P

Componentes básicos de JXTA

Peers

El componente más común y extendido de cualquier sistema P2P es el peer, que es un nodo en una red P2P que conforma la unidad de procesamiento fundamental de

cualquier solución P2P. Un peer es simplemente una aplicación, que se ejecuta en un dispositivo, generalmente, en un computador, y que tiene la habilidad de poder comunicarse con otros semejantes. Y aunque puede ser una aplicación en un solo ordenador, también se extiende a aplicaciones distribuidas que hacen uso de varios ordenadores o diversos dispositivos.

Identificadores

Se usan UUID de 128-bits para referirse a las entidades (un peer, un grupo, un servicio, etc.) de forma única. Cada identificador representa una dirección independiente de la dirección física real de la entidad.

Peer groups

Es un conjunto de peers que tienen en común una serie de intereses o objetivos comunes. Dentro del conjunto de peers se pueden ofrecer servicios a sus miembros que no son accesibles por otros peer que estén incluso en la misma red.

Los intereses en común que pueden compartir suelen ser de colaboración en aplicaciones específicas, en requerimientos de seguridad o de monitorización.

Un peer group es una entidad virtual que habla el conjunto de protocolos de grupo. Típicamente, un grupo de nodos es una colección de nodos cooperativos que tienen en común un conjunto de servicios.

Advertisements

Un peer o un grupo de peers tienen un servicio que desean ofrecer en la red, utilizan un advertisement o anuncio para ofrecerlo. Un advertisement no es más que un documento XML que describe un recurso, ya sea un peer, un peer group o un pipe, etc. Existen varios tipos de anuncios (como Peer Services y Peer Group Services)

aunque se pueden crear nuevos tipos. Los anuncios es la forma que tienen las entidades de publicar los recursos que disponen en la red. Los anuncios llevan aparejados un tiempo de vida de forma que, se puedan eliminar recursos obsoletos sin requerir un control centralizado sobre ellos.

Capa de transporte

Para intercambiar la información, los peers suelen utilizar algunos mecanismos que permiten realizar la transmisión de la información a través de la red. Puede ser un protocolo de bajo nivel, como UDP o TCP, o un protocolo de alto nivel, como HTTP o SMTP.

Se suele dividir en tres componentes:

- **Endpoints**(Extremos): El origen inicial o destino final de cualquier fragmento de datos transmitido a través de la red. Un endpoint se corresponde con la interfaces de red usadas para enviar y recibir datos.
- **Pipes**(Cauces): Canales virtuales de comunicación asíncronos y unidireccionales que conectan uno o más extremos. En ellos se puede transferir cualquier tipo de objeto: código binario, string, etc. Permiten conectar peers que no están conectados directamente de forma física. Además pueden conectar peers que tampoco compartan el mismo servicio de conexión(http con bluetooth).
- **Messages**(Mensajes): Contenedores de datos(datagramas) transmitidos a través de un cauce de un extremo a otro.

Servicios

Funcionalidades que los peers pueden ofrecer otros peers. Las funcionalidades suelen ser del tipo de transferencia de archivos, procesamiento, etc Los servicios es el motivo principal de conformar una red P2P entre varios dispositivos.

Normalmente los servicios se dividen en dos categorías:

- **Peer services:** Son funciones que un peer determinado ofrece a otros peers de la red. Es obvio, que las funciones que pueda ofrecer ese peer específico solo estarán disponibles mientras éste permanezca conectado a la red. Cuando el peer se desconecte de la red, el servicio dejará de estar disponible.
- **Peer group services:** Funcionalidades ofrecidas por un grupo de peers a los miembros que pertenecen al grupo. Las funcionalidades ofrecidas por varios miembros permite acceso redundante a servicios, donde estará disponible mientras exista algún miembro conectado que la ofrezca.

Protocolos

Un protocolo es una forma de estructurar un intercambio de información entre dos o más participantes usando unas reglas previamente acordadas por todos ellos. En JXTA, un protocolo está definido por mensajes en XML. Los peers usan estos protocolos para descubrir a otros peers, anunciar o descubrir recursos de la red, y para comunicarse y enviar mensajes.

En JXTA existen un total de seis protocolos:

- **Peer Discovery Protocol (PDP):** Usado por los peers para anunciar mediante advertisement sus recursos, y descubrir los publicados por otros peers.
- **Endpoint Routing Protocol (ERP):** Usado por los peers para encontrar el camino hacia otro peer.
- **Peer Resolver Protocol (PRP):** Los peers pueden enviar una consulta genérica a uno o más peers y recibir una respuesta a dicha consulta.
- **Peer Information Protocol (PIP):** Usado por los peers para obtener cierta información de estado(estado, tráfico reciente, etc) de otros peers.

- **Rendezvous Protocol (RVP):** Usado por los edge peers para encontrar recursos, propagar mensajes y anunciar recursos propios. Usado también por los rendezvous peers para compartir los servicios ofertados de sus peers con otros rendezvous peers vecinos.
- **Pipe Binding Protocol (PBP):** Usado por los peers para establecer una comunicación virtual mediante pipes entre uno o más peers. Además, es el mecanismo que permite a un peer descubrir la ruta a un peer destino.

Todos estos protocolos asíncronos están basados en el modelo llamada / respuesta. Las implementaciones de JXTA no tienen porque soportar, como ya comente antes, los seis protocolos completos.

Para complementar el marco teórico, en este trabajo se consideró necesario especificar las bases legales que los sustentan, ya que se está trabajando con un problema que puede presentarse como es el derecho de autor que limitan la publicación libremente de textos bibliográficos.

Bases Legales

Dadas las características de accesibilidad a través de Internet de libros o material digital, es evidente que existe un gran riesgo para los autores y editores de libros electrónicos en frente de reproducciones fraudulentas o copias no autorizadas de dichos libros. En efecto, si cualquier persona pudiera descargar un libro completo a su computador, Laptop o Palm sin ninguna restricción para su uso posterior, ello facilitaría las reproducciones ilegales. Así por ejemplo podrían reproducirse cuantas copias se quisiera a fin de venderlas, modificarlas, etc., sin necesidad ni siquiera de tener que re-digitar el contenido del libro, producir planchas de impresión u otros tantos inconvenientes o trabas que tienen que superar los infractores de derechos de autor de libros físicos.

En general cuando una persona compra un libro en su forma tradicional, tiene el derecho a disponer de esa copia física del libro, pudiendo venderla, prestarla, usarla para fines de estudio, obtener una copia solo para su uso personal etc. Esos derechos no lesionan los derechos del autor de la obra y en la legislación Americana y del sistema legal Anglosajon se denomina “la doctrina del fair use” o la doctrina del uso legal de una obra protegida. Bajo la legislación Venezolana la ley de derecho de autor del 14 de agosto de 1993, Reglamento de la Ley Sobre el Derecho de Autor y de la Decisión 351 de la Comisión del Acuerdo de Cartagena que contiene el Régimen Común sobre Derecho de Autor y Derechos Conexos en su artículo 44 indica como reproducciones lícitas :

- La reproducción de una copia de la obra impresa, sonora o audiovisual, salvo en el programa de computación que se regirá conforme al numeral 5 de este artículo, siempre que sea realizada para la utilización personal y exclusiva del usuario, efectuada por el interesado con sus propios medios.
- Las reproducciones fotomecánicas para el exclusivo uso personal, como la fotocopia y el microfilme, siempre que se limiten a pequeñas partes de una obra protegida o a obras agotadas, y sin perjuicio de la remuneración equitativa que deban abonar las empresas, instituciones y demás organizaciones que presten ese servicio al público, a los titulares del respectivo derecho de reproducción. Se equipara a la reproducción ilícita, toda utilización de las piezas reproducidas para un uso distinto del personal que se haga en concurrencia con el derecho exclusivo del autor de explotar su obra.
- La reproducción por medios reprográficos, para la enseñanza o la realización de exámenes en instituciones educativas, siempre que no haya fines de lucro y en la medida justificada por el objetivo perseguido, de artículos, breves extractos de obras u obras breves lícitamente publicadas, a condición de que tal utilización se haga conforme a los usos honrados.

- La reproducción individual de una obra por bibliotecas o archivos que no tengan fines de lucro, cuando el ejemplar se encuentre en su colección permanente, para preservar dicho ejemplar y sustituirlo en caso de necesidad o para sustituir en la colección permanente de otra biblioteca o archivo, un ejemplar que se haya extraviado, destruido o inutilizado, siempre que no resulte posible adquirir tal ejemplar en plazo y condiciones razonables.
- La reproducción de una sola copia del programa de computación, exclusivamente con fines de resguardo o seguridad.
- La introducción del programa de computación en la memoria interna del equipo, a los solos efectos de su utilización por el usuario lícito, y sin perjuicio de su participación al titular del derecho cuando así se haya pactado en el contrato de enajenación del soporte material o en la licencia de uso.
- La reproducción de una obra para actuaciones judiciales o administrativas, en la medida justificada por el fin que se persiga.
- La copia de obras de arte efectuada a los solos fines de un estudio.
- La reproducción de una obra de arte expuesta permanentemente en las calles, plazas u otros lugares públicos, por medio de un arte diverso del empleado para la elaboración del original. Respecto de los edificios, dicha facultad se limita a la fachada exterior.

Por su parte la decisión 351 del Acuerdo de Cartagena, norma comunitaria, también incluye una serie de previsiones, respecto de usos legítimos que no constituyen infracción de derechos de autor y cuyo fin específico es que no se sacrifique los intereses educativos y de la comunidad por la protección que la ley otorga a los titulares de derechos de autor.

2 Artículo 22 Decisión 351 del Acuerdo de Cartagena.- “Sin perjuicio de lo dispuesto en el Capítulo V y en el artículo anterior, será lícito realizar, sin autorización del autor y sin el pago de remuneración alguna, los siguientes actos:

- Citar en una obra, otras obras publicadas, siempre que se indique la fuente y el nombre del autor, a condición de que tales citas se hagan conforme a los usos honrados y en la medida justificada por el fin que se persiga;
- Reproducir por medios reprográficos para la enseñanza o para la realización de exámenes en instituciones educativas, en la medida justificada por el fin que se persiga, artículos lícitamente publicados en periódicos o colecciones periódicas, o breves extractos de las obras lícitamente publicadas, a condición que tal utilización se haga conforme a los usos honrados y que la misma no sea objeto de venta u otra transacción a título oneroso, ni tenga directa o indirectamente fines de lucro;
- Reproducir en forma individual, una obra por una biblioteca o archivo cuyas actividades no tengan directa o indirectamente fines de lucro, cuando el ejemplar respectivo se encuentre en la colección permanente de la biblioteca o archivo, y dicha reproducción se realice con los siguientes fines:
 - Preservar el ejemplar y sustituirlo en caso de extravío, destrucción o inutilización; o,
 - Sustituir, en la colección permanente de otra biblioteca o archivo, un ejemplar que se haya extraviado, destruido o inutilizado.
- Reproducir una obra para actuaciones judiciales o administrativas, en la medida justificada por el fin que se persiga;
- Reproducir y distribuir por la prensa o emitir por radiodifusión o transmisión pública por cable, artículos de actualidad, de discusión económica, política o religiosa publicados en periódicos o colecciones periódicas, u obras radiodifundidas que tengan el mismo carácter, en los casos en que la reproducción, la radiodifusión o la transmisión pública no se hayan reservado expresamente;
- Reproducir y poner al alcance del público, con ocasión de las informaciones relativas a acontecimientos de actualidad por medio de la fotografía, la cinematografía o por la radiodifusión o transmisión pública por cable, obras

vistas u oídas en el curso de tales acontecimientos, en la medida justificada por el fin de la información;

- Reproducir por la prensa, la radiodifusión o la transmisión pública, discursos políticos, así como disertaciones, alocuciones, sermones, discursos. Pues bien, la principal preocupación para la comercialización exitosa de libros electrónicos era poder restringir que personas inescrupulosas tuvieran acceso a ellos y procedieran a obtener copias ilegales en detrimento de los derechos de los distribuidores vendedores y del mismo autor. Para ello compañías como Adobe Systems Inc desarrollaron un programa conocido como “eBook Reader” que le permite al poseedor legal de un libro electrónico, sea por que lo hubiere comprado o por que hubiere tenido acceso al mismo en forma gratuita pero autorizada, leer el contenido del libro exclusivamente en el PC o en la Palm en donde hubiere sido descargado el e-book.

La OMPI (organización Mundial de la Propiedad Intelectual) (WIPO en inglés) www.ompi.org, desde 1996 se ocupó del tema de la protección legal de los derechos de autor y el desafío que pronunciados durante actuaciones judiciales u otras obras de carácter similar pronunciadas en público, con fines de información sobre los hechos de actualidad, en la medida en que lo justifiquen los fines perseguidos, y conservando los autores sus derechos a la publicación de colecciones de tales obras.

Operacionalización de las variables

Con la finalidad de establecer los aspectos a medir en la presente investigación, se identifican las variables según el siguiente esquema:

Variables Conceptuales

- Documento: Es una agrupación de material digital que maneja y ofrece la biblioteca para ser consumida como servicio.

- Repositorio de Documentos: Es el proceso de crear, publicar y controlar Material digitalizado, disponible para prestar como servicio dentro de la biblioteca.
- Búsqueda Distribuida: Es el proceso de ubicar por características un material digital específico para acceder a este, en las diferentes bibliotecas de los decanatos o incluso bibliotecas aliadas.

Variables Operacionales

Para efectos de la operacionalización de las variables conceptuales se procedió de la siguiente manera:

La variable “Documento” se operacionalizó considerando dos dimensiones, la dimensión catalogación de material y Derechos de autor de los mismo.

La dimensión catalogación está constituida por los atributos que permitirán definir un material digital. La dimensión derechos de autor está constituida por las limitaciones legales y acuerdos con autores o editoriales para poder disponer de estos.

La variable “Repositorio de Documentos” se operacionalizó considerando tres dimensiones, la dimensión Digitalización de documento, Publicación y acceso del servicio y control del mismo. La dimensión Digitalización del material está constituida por la apreciación acerca de cómo cuando y de qué manera debe digitalizarse el documento, así como en que formato se digitaliza y la infraestructura para almacenamiento de los mismos. La dimensión publicación y acceso del servicio, recoge la manera que el bibliotecario transformara el documento en un servicios a disposición de los usuarios así como la manera que usuario podrá acceder al mismo. Por último la dimensión control la cual recogerá las restricciones de acceso a los servicios publicados.

La variable “Búsqueda Distribuida” se operacionalizó considerando tres dimensiones, la dimensión criterios de búsqueda, servicio de búsqueda y alianzas inter-bibliotecarias. La dimensión criterios de búsqueda, está constituida la búsqueda del material digital a realizar en base a catalogación, contenido de los mismos o ambos. La dimensión, servicio de búsqueda, recoge la manera que se podría realizar las búsquedas de manera local y distribuida así como la estructura tecnológica para el desarrollo de las mismas y finalmente la dimensión alianzas inter-bibliotecarias la cual recogerá las posibilidad de la realización de alianzas entre bibliotecas digitales para maximizar la cantidad de contenidos o servicios, determinando restricciones tecnológicas y alianzas existentes o futuras posibles.

A continuación se presenta la tabla N° 1 que resume la operacionalización de las variables en estudio.

Tabla 1: Operacionalización de las variables estudiadas.

Variable	Dimensión	Indicadores	Item en cuestionario
Documento	Catalogación de material	Atributos a Catalogar	3
		Manera a catalogar, según tipo documento	4
		Estandarización de catalogación entre decanatos	2
	Derechos de autor	Acuerdos con editoriales para publicación de libros digitales.	5
		Variación de Acceso o digitalización.	6
Repositorio de	Digitalización de documento	Técnicas implementadas Actual	8
		Formatos de digitalización	9

Documentos		Capacidad de almacenamiento del material	10	
	Publicación y acceso del servicio	Publicación de Material Digital	11	
		Modo de Acceso a servicios	12	
	Control de servicios	Definición de perfiles por documento	13	
		Definición de perfiles para acceso a servicios	14	
		Formas de acceso	15	
	Búsqueda Distribuida	Criterios de búsqueda	Criterios por catalogación.	17
			Búsqueda por contenido	18
			Existencia Atributos de solo control en Catálogo	19
Servicio de búsqueda		Búsqueda local	16	
		Búsqueda remota	20	
		Estructura tecnológica	21	
Alianzas inter-bibliotecarias		Inconvenientes tecnológicos	23	
		Posibilidad o existencia de alianzas	22	

Fuente: El autor de la investigación.

CAPÍTULO III

MARCO METODOLÓGICO

Naturaleza del estudio

Es imprescindible en toda investigación científica que los hechos estudiados, sus relaciones y los resultados obtenidos, sean evidencias significativas con respecto al problema investigado y los nuevos conocimientos reúnan la condición de validez; para la cual es necesario demostrar los procedimientos de orden metodológico, a través de los cuales se propone dar respuesta a las interrogantes de la investigación.

Esta investigación es de tipo experimental, debido a que las variables existentes tales como las bibliotecas digitales, estándares comunes entre las bibliotecas digitales, modelos, no son manipuladas, ni se ejerció control sobre las mismas. En este diseño de investigación su propósito fue describir variables y analizar su incidencia e interrelación en un momento dado (Hernández y otros, 2003. P. 270), por lo que se diseñó una herramienta que integra las bibliotecas digitales de las universidades de los decanatos de la UCLA y otra con las que se pudiera establecer una alianza.

El estudio se apoyó en una investigación bajo la modalidad proyectos tal como lo especifica UCLA (2002), donde se descubrieran valores reales que representaron el diseño de un modelo factible para la solución de un problema de tipo práctico, en este caso el diseño y creación de una herramienta para búsqueda de contenidos distribuidos de bibliotecas digitales.

De la misma manera esta investigación es de campo, ya que estudia los fenómenos en su ambiente natural, los datos se recogerán en forma directa permitiendo los efectos de la interrelación de variables en su propio ambiente donde acontece. Igualmente, Sabino (2000), define diseño de campo, como los métodos empleados cuando los datos de interés se recogen en forma directa de la realidad mediante el trabajo concreto del investigador y su equipo, estos datos permiten la obtención de la información necesaria, para realizar una herramienta para la integración y búsqueda en sistemas de gestión de contenidos en ambientes distribuidos para bibliotecas digitales. Siendo trabajo de desarrollo se usa una metodología de desarrollo, en este caso en particular se eligió XP.

Fase I: Estudio Diagnostico

Atendiendo a la modalidad de investigación de este trabajo, es aquí donde se establece el problema que se quiere solventar, analizando a una población, por medio de una muestra, a la cual se le aplicaran una serie de instrumentos para identificar, explicar y comprender la existencia de un problema.

Hernández y Otros (2001), refieren que está dirigida a “...proporcionar una información adecuada y confiable que sirva de base para una acción (realización de un plan, programa o proyecto) y para fundamentar las estrategias que se han de expresar en la práctica concreta” (p. 55).

Lee (1992), indicó varios investigadores dedicados a hacer estudios tanto cuantitativos como cualitativos, sugieren combinar más de un método en una misma investigación; por esto, las técnicas en las que se incurrió para la recolección de la información necesaria para el desarrollo de esta investigación fueron las siguientes:

Entrevista semi-estructurada focalizada (Anexo 1)

Según Sabino (ob.cit.), es una forma específica de interacción social, donde el investigador formula preguntas a las personas capaces de aportarles datos de interés estableciendo un diálogo peculiar donde una de las partes busca recoger informaciones y otra es la fuente de la misma. Por tal razón, el propósito de esta técnica es conferenciar de manera informal, sobre algún tema establecido previamente y, a la vez reunir datos.

Por lo tanto, la entrevista que se realiza en la investigación, será de tipo semiestructurada, por cuanto se caracteriza por la libertad para formular las preguntas siguiendo unos lineamientos y profundizar de acuerdo a las respuestas.

El investigador consideró pertinente aplicar esta entrevista para poder levantar las historias de usuario de la metodología XP. En virtud a ello, se entrevistó al nivel estratégico de las bibliotecas de los decanatos de la UCLA, para de esta manera conocer los requerimientos para el manejo y control integrado del material digital por parte de las bibliotecas de los mismos.

Observación directa (Anexo 2)

Padua (1989) señala que: “la observación directa puede ser no participante, la cual es aquella en la que el investigador hace uso de la observación directa sin ocupar un determinado status o función dentro de la organización en donde se hace la investigación” (p.55).

Se empleó como técnica de recolección de datos la observación directa, la cual permitió al autor del trabajo observar y recoger datos mediante su propia observación, lo que le permitió obtener una visión más clara del problema y determinar la situación real de las bibliotecas digitales y la digitalización de material de las mismas. Como instrumento se empleó una guía de observación.

Revisión bibliográfica

Esta permitió al autor conocer definiciones, características y funcionamiento de las bibliotecas digitales, tendencias en cuanto a la distribución de contenidos y su manejo en ambientes distribuidos. Forma de trabajo de bibliotecas digitales en otros países, técnicas de búsquedas actuales.

Validación del instrumento

Se realizó mediante el método de Juicio de un Experto, para lo cual se seleccionaron dos especialistas en análisis de requerimientos, al cuales se les entregó una copia de la matriz de variables, objetivos de la investigación para que determinara la claridad, la congruencia de las preguntas y si medían lo que se pretendía medir en la investigación.

Procedimiento

En el siguiente aparte se listan los pasos de la presente investigación:

- Elaborar los instrumentos.
- Aplicar los instrumentos.
- Analizar los datos recabados por los instrumentos.
- Presentar las conclusiones del diagnóstico.
- Presentar las recomendaciones del diagnóstico.

La población en esta investigación estará conformada por miembros de las diferentes bibliotecas de los diferentes decanatos de la universidad Lisandro Alvarado en la ciudad de Barquisimeto.

La muestra se realizara a diferentes perfiles como asistentes de bibliotecas, especialistas de información, miembros de personal administrativo y miembros del

departamento de procesos técnicos, de acuerdo disponibilidad de estos actores, y su capacidad y compromiso para aportar a esta investigación.

Aplicación de los instrumentos.

En la aplicación de los instrumentos se realizaron las siguientes actividades:

- Se contacto al personal de las bibliotecas.
- Se efectuaron las entrevistas semi-estructuradas.
- Se observo el funcionamiento de la biblioteca, por medio de visitas a sus diferentes páginas web, y salas de bibliotecas en sus respectivos decanatos.
- Analizar los datos recabados por los instrumentos

En base a los datos recabados la información se analizara, definiendo el modelo genera y posteriormente elaborar una lista en base a requerimientos.

Fase II: Factibilidad

Social

La importancia de la gestión del conocimiento en esta era, y la relevancia de la distribución de conocimiento de las universidades por medio de sus bibliotecas. Para presentar una posible solución a este problema de relevancia, permitiendo que las bibliotecas universitarias aporten información de calidad para el desarrollo de nuestras sociedades, y por medio de la integración o alianzas puedan tal vez fortalecerlas y poder así, prestar mayor cantidad y calidad de información y servicios.

Operativa

Los usuarios de la solución planteada, en primera instancia, van a ser todas las personas que laboran en las bibliotecas digitales universitarias de la Universidad Centro-Occidental Lisandro Alvarado de la ciudad de Barquisimeto, alumnado de las mismas y profesores, sin embargo el impacto que genere la investigación puede trascender a esta Universidad, sector y en el espacio físico de Barquisimeto, ya que entre otras cosas va a establecer un marco para permitir las alianzas entre las diversas bibliotecas sin restricciones de ubicación, además de estar basada en estándares libres posibilitando su implantación dentro de cualquier ambiente tecnológico y ser anexada por cualquier herramienta de manejo de contenidos y ambientes distribuidos.

Institucional

En relación al aspecto institucional el desarrollo de esta herramienta, podrá fortalecer el acceso a la información para alumnos y profesores, garantizando a la institución mejorar la efectividad, en cuanto al acceso a la información y posiblemente mejorando la calidad de futuros trabajos en el ámbito de investigación.

Tecnológica

El software requerido para el desarrollo de la investigación consiste principalmente en frameworks que soporten los principios enunciados en ambientes distribuidos, así como también herramientas que permitan documentar el modelado.

Dada la iniciativa del gobierno nacional en promover el uso del software libre por medio del decreto 3390 en los organismos gubernamentales, se usarán herramientas de código abierto para la implementación del mismo ya que esta está enfocada a las bibliotecas universitarias del sector público. En este sentido es importante acotar, que dado los alcances y limitaciones de esta investigación, existe

en el mundo del software libre una amplia gama de componentes que permitirán el desarrollo del sistema.

La herramienta no será realizada como anexo a un manejador de contenidos específico, permitiendo a la institución usar el manejador de contenidos de su preferencia, pero para efectos prácticos el mismo será probado en el gestor de contenidos “Obelisco”

Económica

En lo que respecta a la adquisición de licencias de software para el desarrollo del proyecto, el costo podría reducirse a cero, ya que este producto estará basado en tecnologías libres y su diseño estará enfocado a que pueda ser ejecutado por diversos gestores de contenidos, el costo dependerá del gestor de contenidos que quiera utilizar la institución y su adaptación para uso de la herramienta.

CAPITULO IV

PROPUESTA DEL ESTUDIO

Introducción

Partiendo de la existencia de problemas en las búsqueda distribuidas e integración de información en las biblioteca digitales de los diferentes decanatos de la Universidad Centro Occidental Lisandro Alvarado, se procedió a la elaboración de instrumentos para la captura de datos que permitan un diagnostico de este problema, utilizando técnicas de recolección tales como: entrevista semiestructura al personal de las bibliotecas, observación directa al funcionamiento de las mismas de modo físico, y a accesos por internet, así como verificación de material bibliográfico que serviría de aporte.

Posterior a esto se procedió a la aplicación de los estos, y al análisis de los datos recabados por dichos instrumentos, dicho análisis servirá de entrada a una metodología de desarrollo ágil, con el fin de desarrollar una herramienta que permita la solución tecnológica del problema detectado.

En la actualidad se cuenta con una amplia gama de modelos de desarrollos de software, los cuales nos permiten organizar las actividades necesarias para la generación del producto, estas imponen un proceso disciplinado sobre el desarrollo de software con el fin de hacerlo más predecible y eficiente. Se aplica desarrollando un proceso detallado con un fuerte énfasis en planificar inspirado por otras disciplinas de la ingeniería. Por lo tanto se selecciono la metodología XP como una metodología

de desarrollo ágil, para la elaboración de este proyecto, debido a que esta se basa en la simplicidad, la comunicación y el reciclado continuo de código.

Partiendo de que la metodología XP está dividida en fases iterativas; en este trabajo de investigación aplicaran dichas fases de la siguiente manera:

La primera fase Planificación del proyecto; se definió las historias de usuario con el cliente. Estas serán generadas por medio del análisis de los datos recabados por los instrumentos.

Después de tener definidas las historias de usuario, se creó el plan de publicaciones (Release plan), donde se indican las historias de usuario que se crearán para cada versión del programa y las fechas en las que se publicarán estas versiones. Este permite tener claros estos cuatro factores críticos como los son: los objetivos que se deben cumplir (que son principalmente las historias que se deben desarrollar en cada versión), el tiempo que tardarán en desarrollarse y publicarse las versiones del programa, el número de personas que trabajarán en el desarrollo y cómo se evaluará la calidad del trabajo realizado

Posteriormente se procedió a determinar cinco iteraciones para este proyecto, con una duración aproximada de 3 semanas. También se seleccionan las historias de usuario que no pasaron el test de aceptación.

La velocidad del proyecto es una medida que representa la rapidez con la que se desarrolla el proyecto, en el caso de este proyecto será de uno a dos casos de uso en cada iteración dependiendo el caso y la iteración.

La segunda fase es la de Diseño; los diseños serán simples y sencillos, para poder controlarlos rápidamente, además se usaran diagramas UML para entender mejor el diseño de los mismos.

La tercera fase de codificación; se codificaran las historia de usuario definidas en la fase anterior para cada iteración, para lo cual se seguirán a estándares de codificación ya creados para facilitar su comprensión y escalabilidad.

La cuarta y última fase es la de Pruebas es uno de los pilares de la metodología X.P es el uso de test para comprobar el funcionamiento de los códigos que vayamos implementando. Sometiendo a test las distintas clases del sistema omitiendo los métodos más triviales. Creando los test que pasarán los códigos antes de implementarlos.

El producto final de la presente investigación, será una herramienta de integración y búsqueda en sistemas de gestión de contenidos en ambientes distribuidos para bibliotecas digitales, que podrá ser utilizada por las bibliotecas de la universidad Centro-occidental Lisandro Alvarado (UCLA) permitiendo integrar los diferentes decanatos por medio del el manejador de contenidos “Obelisco”; para el control, desarrollo adecuado y estructurado de las actividades del mismo, se utilizara el modelo de desarrollo XP(programación extrema).

A continuación se desarrolla detalladamente cada fase e iteración, las cuales cubren una lista de historia de usuarios y requerimientos que son establecidos en el plan de entregas.

Aplicación de los instrumentos

Para esta fase se procedió a realizar la entrevista semiestructurada (Anexo1) al personal bibliotecario de los diferentes decanatos, donde encontramos responsables por área y a los cuales se les aplicó este instrumento. Las preguntas aplicadas fueron abiertas y su audio fue almacenado, se dio mayor interés a las áreas que manejaban material digital o se responsabilizaban por la digitalización del mismo.

Igualmente se realizó visitas de reconocimiento físico, para apreciar su funcionamiento desde otra óptica, antes de la realización de cada entrevista. Así como acceder a las diferentes páginas de las bibliotecas dentro y fuera del campus universitario para apreciar su funcionamiento, sus estándares, integración, y material digital disponible; así como la revisión del material documental que permitiera tener una visión más clara de la realidad actual de las bibliotecas digitales.

Diagnostico

Luego de generados los instrumentos se procedió al análisis de los mismos, para de esta manera determinar los requerimientos tecnológicos y problemas de las bibliotecas, enfocándolos en la manipulación y búsqueda de documentos digitales, para lo cual primero se planteo la realidad o situación actual de las mismas.

A continuación se describirá en base a la información recaudada el estado actual y funcional de la biblioteca enfocado en los siguientes aspectos: acceso al material digital o físico, gestión del documento Digital, Manejo de contenido, control de material digital (repositorio de Documentos) y Búsqueda distribuida.

Acceso al material físico o digital en las Bibliotecas de la UCLA:

En la actualidad las bibliotecas físicas de los diversos decanatos de la UCLA están estructurados por salas, donde el usuario tiene acceso al material físico bajo dos (2) posibles modalidades dependiendo de las características de dicha sala y decanato, la modalidad de estantería cerrada (el usuario necesita un Intermediario para acceder a los recursos) y abierta (se accede directamente al recurso).

La mayoría de las salas cuentan con computadores que permiten al usuario realizar búsquedas al material existente y la disponibilidad del mismo por medio del sistema Alejandría; éste es el sistema común de las bibliotecas de los decanatos para catalogar y controlar el material físico de las mismas, permitiendo dar soporte interno

para fichas y control de material físico, así como definir usuarios, establecer reportes y estadísticas de acceso y disponibilidad del material.

Además de las infraestructuras físicas cada decanato cuenta con una página propia de acceso a sus bibliotecas digitales, donde el usuario podrá realizar búsquedas del material físico catalogado en el Alexandria así como a material digitalizado alojado en los repositorios de los decanatos de la red UCLA. Existe además una página Web general de biblioteca de la UCLA, la cual tiene accesos a las de los decanatos.

El material digitalizado texto completo, presenta un acceso directo donde se puede acceder al documento en su totalidad en formato protegido, en la mayoría de los casos estos son tesis; este acceso por lo general cuenta con un servicio de búsqueda, sin embargo no todos los resultados arrojan el material a texto completo, también muestra documentos solo catalogados, que tienen que consultarse de manera física en la institución y limitarse al resumen del mismo.

Además de la data digital propia de la biblioteca de los decanatos, las bibliotecas UCLA cuentan con suscripción a diversos repositorios de documentos digitales como lo son ProQuest, EBSCOhost, E-Journals, E-ebrary, y E-libro, etc. Estos dentro del campus Universitario tendrán acceso sin necesidad de identificación, pero fuera del área del campo universitario deberá el usuario acceder por medio de un login y clave, que en la mayoría de los casos se desconoce, pero la misma puede ser solicitado por el personal o estudiantes de cada decanato a través de un correo electrónico o de forma personal. A continuación se dará una breve descripción de los diferentes repositorios a los que están suscritas las bibliotecas de la UCLA.

- ProQuest: Base de Datos multidisciplinaria de revistas científicas internacionales con gran cantidad de Artículos, papers y tesis a texto

completo, en todas las áreas del conocimiento, en la actualidad este servicio se encuentra personalizado.

- EBSCOhost: Sistema en línea que ofrece una variedad de bases de datos propietarias con una gran cantidad de Artículos a Texto Completo de los principales proveedores de información.
- E-Journals: Base de datos contenida en EbscoHost la cual permite consultar los texto completo de revistas que se adquieren en formato impreso (disponibles en físico en la hemeroteca de su decanato) y para lo cual se tienen algunos accesos OnLine vía ATHENS.
- E-ebrary: Biblioteca de Libros Electrónicos en Inglés a texto completo, provenientes de más de 100 editoriales líderes en el mercado, incluyendo The McGraw-Hill Companies, Random House, Penguin Classics, John Wiley & Sons, Cambridge University Press y más.
- Libro (Intranet) - E-Libro (Acceso Remoto): Biblioteca de Libros Electrónicos en español a texto completo, provenientes de más de 100 editoriales líderes en el mercado, incluyendo The McGraw-Hill Companies, Random House, Penguin Classics, John Wiley & Sons, Cambridge University Press y otras.
- Biblioteca Virtual en Salud: (usada por el decanato de Medicina) presentando un Catálogo de revistas científicas, documentos y eventos relacionados a la salud y medicina.
- Biblioteca.universia.net: tiene un alianza con el decanato de contabilidad y administración, permitiendo el acceso total o resúmenes de investigaciones, papers y trabajos de grado de diversas universidades incluyendo el decanato de administración y contabilidad de la UCLA.
- El decanato de Agronomía tiene diversos enlaces a diversas paginas con información agrícola como la son la biblioteca de agronomía de la UCV, biblioteca de la Luz, etc.

Gestión del documento digital:

Tanto el material físico como el digitalizado, es catalogado por medio del sistema Alejandría, con una ficha cerrada que presenta los siguientes atributos: autores, tutor, título, materia, cota, fecha, páginas, grado académico, disciplina, descriptor temático, tipo de documento, resumen y enlace.

Como todos los decanatos trabajan con la misma herramienta y en esta catalogación es cerrada, los decanatos deben mantener un estándar en la catalogación del material, mas sin embargo la codificación de los mismos es a criterio de cada decanato; el número de referencia debe ser única y es equivalente al ISBN en los libros pero en la digitalización de tesis no hay un estándar común definido o adoptado.

En el caso del material digital solo se presentan trabajos de grado de los mismos decanatos, no todos los decanatos están digitalizando material; los libros y las revistas no son digitalizados por los posibles problemas legales por derechos de autor.

En el momento no se presenta ningún acuerdo de acceso parcial o total condicionado a material digital con editoriales o autores, ni se proyectan los mismos para un futuro cercano.

Manejo de contenido y control de material digital (repositorio de Documentos)

En este momento el material completo digitalizado casi en su totalidad son Tesis de grado, la digitalización de revistas y libros se realiza al descriptor temático o resumen para el catálogo del mismo, y los datos incluidos en el Alejandría.

En cuanto a los trabajos de grado de postgrado a partir del año 2005 son entregados en formato digital, y los de pregrado o postgrado anteriores a las fecha que

por petición sea necesaria su digitalización se realiza en forma total por medio del escaneado, el formato de digitalización usado para estas es el PDF protegido y con marcas de agua.

Luego de digitalizado el documento, es almacenado en el repositorio de documentos de la red, y se procede a la modificación de su catálogo anexando que es tipo digital y su enlace.

En la actualidad el acceso al material digitalizado va dirigido a todo público a través del portal web del decanato sin restricciones de acceso. En este se pueden realizar las búsquedas y acceder a su catálogo o descripción por medio de un enlace y de tener texto completo muestra un enlace que nos dará acceso al PDF con la información, este modo de acceso para el material digitalizado hasta el momento es factible, pero si en un futuro se realizaran convenios con autores dando restricciones al acceso de los mismos, este debería ser cambiado por un manejador de contenidos como tal.

El control y estadísticas del acceso o manipulación de los documentos digitales no se realizan, por lo que el usuario trabaja con el archivo digital y no con manejador de contenidos, aunque si se tiene controles sobre la consulta de catálogo o fichas.

Búsqueda distribuida

Los criterios de búsqueda varían de acuerdo al portal de cada decanato, pero por lo general presentan los siguientes: Búsqueda simple la más común, busca coincidencias entre el texto administrado y cada uno de los atributos del catálogo; búsqueda por título, autor o materia solo busca coincidencias por el valor seleccionado, la general se indica el atributo por el cual buscar y el valor a buscar para ese atributo en específico.

La posibilidad de búsqueda del contenido del documento digital como tal no se realiza, como se menciona anteriormente se realiza es por medio de la información en el catálogo.

Aunque la UCLA presenta una infraestructura tecnológica buena y una intranet común, cada biblioteca es una isla y cada portal busca su propio material, ya que no cuenta con búsquedas integradas. En la página General de Biblioteca de la UCLA o rebin, este tipo de búsqueda no existe y la misma solo redirecciona a los diversos decanatos.

Actualmente la UCLA no tiene alianzas tecnológicas de colaboración e integración con ninguna otra biblioteca, esta necesidad es remplazada con suscripciones a bases de datos de documentos electrónicos ya mencionados anteriormente. Ahora bien sería posible la realización de dichas alianzas de acuerdo a los términos planteados para la ejecución de las mismas.

Análisis de datos recabados

Se aplicó el análisis de las bibliotecas de los decanatos de la UCLA, en base a las variables conceptuales definidas anteriormente:

Documento:

Se pudo apreciar que el enfoque general de la biblioteca es el acceso a material físico, solo en unos pocos decanatos están digitalizando o publicando material digital completo, el catálogo de los documentos es estándar ya que este en la herramienta actual es cerrado, mas no hay un estándar aplicado entre los diferentes decanatos para la codificación de los documentos digitalizados, cada decanato tiene su propio estándar, no existen acuerdos con editorial solo suscripciones con repositorios de documentos, el servicio de acceso a material digital propio de las bibliotecas es reducido y no existen por el momento restricciones impuestas por autores en la

publicación del material, el acceso del mismo es un enlace, que no refleja control, ni estadísticas del acceso de estos,

Repositorio de documentos:

Por el momento solo se incorporan tesis como material digital, propio de los repositorios de las bibliotecas de los decanatos, las cuales en la mayoría de los casos vienen digitalizadas, de no ser así se escanean, las tesis son publicadas en formatos PDF protegidos y con marca de agua incorporada, y el acceso a este material no presenta restricciones más que las definidas en el documento.

Los documentos a digitalizar no son sometidos a restricciones de tamaño o peso y disponen de varios posibles servidores para almacenar dichos repositorios, y son almacenados en rutas específicas de disco y anexada esta información en el catálogo del sistema Alejandría.

Búsqueda distribuida:

En cuanto a la integración del material o búsqueda distribuida no existe, la Biblioteca presenta una página general de biblioteca que agrupa los decanatos pero esta no presenta un servicio de búsqueda global de la información, y la página de la biblioteca de cada decanato te da acceso al material digital exclusivo de la misma.

Como lo indica Faba y otros (2004) las bibliotecas digitales ofrecen muchos beneficios. En lo ideal, un buscador de información puede tener acceso a los materiales siempre y cuando los materiales existan. No habrá escasez de copias, la recuperación puede ser instantánea y los materiales no decaerán, ni se decolorarán. Un usuario no necesitaría desplazarse a la biblioteca más cercana.

En el objeto de estudio que son las bibliotecas de la UCLA, el usuario debe desplazarse a la misma para acceder al material físico, rompiendo con la definición dada anteriormente. Además, de que las búsquedas al material digital no están integrada y el material es muy escaso, ya que solo se publican tesis de grado de pregrado o postgrado y no en todos los decanatos, argumentado esto, a posibles problemas de derechos de autor copyright o que no se han digitalizado documento con otro tipo de derecho de autor, de igual manera no se tiene previsto hacer convenios para publicar libros digitales con restricciones que imponga el autor.

La UCLA además de contar con una buena infraestructura de red y tecnológica, genera una gran cantidad de material digital que podrá ser anexado a las bibliotecas y una posible característica idónea para establecer alianzas con otras bibliotecas, unirse a otras redes de bibliotecas permitiendo que la cantidad de material y servicios a prestar sea mayor.

El servicio de búsqueda y catalogación es común entre el físico y el digital, haciendo que el usuario se confunda o espere un documento y recibe solamente la referencia o catálogo del mismo.

Primera fase: Análisis de requerimientos

Como se menciona inicialmente en esta investigación se enfoca al control, administración y búsqueda de documentos digitales y no a la gestión del material físico, los cuales el sistema Alejandría ya satisface en las bibliotecas del mismo.

A partir del material recogido y al análisis del mismo, se procedió a la generación de los requerimientos funcionales para que la biblioteca preste servicios digitales, para el cual se detectaron dos procesos críticos en el ambiente tecnológico como los son: la gestión de reposito de documentos y la búsqueda local y distribuida de los documentos.

Repositorio de documentos

- Digitalización o creación de Documentos.
- Permitir definir libremente atributos de catálogo.
- Catalogación del Material.
- Generar repositorio de documentos
- Control de actividades realizadas al documento.
- Seguimiento del Documento
- Generar estadísticas de acceso a Documentos.
- Control de acceso.

Búsqueda Distribuida

- Búsqueda por Atributos de catálogo definido.
- Búsqueda de contenido a documentos en diversos formatos PDF, DOC, XLS, etc.
- Posibilidad de generar búsquedas a material en los diferentes decanatos, haciéndolo transparente al usuario.
- Poder realizar la búsqueda en material accesible por bibliotecas aliadas.
- Gestionar datos tanto locales como remotos.

Los SMC son software que permiten y facilitan la organización, creación y gestión de documentos y otros contenidos de un modo cooperativo, como documentos digitales, páginas de texto, imágenes o archivos multimedia.

Existen diversos procesos de gestión de documentos que contemplan reglas funcionales específicas como los requerimientos de este tipo mencionados anteriormente, necesarias para la solución del problema, esto amerita la implementación de módulos de control y seguimiento. En la actualidad existen

múltiples SMC para satisfacer estas necesidades como los son Joomla, Drupal, Obelisco, entre otros.

Para resolver el control y seguimiento de material digital se va a utilizar un SMC para lo cual se seleccionó el “Obelisco”, esta decisión fue tomada siguiendo los siguientes aspectos o características:

- Su flexibilidad al catalogar o caracterizar un documento digital.
- Su accesibilidad Web.
- Se presenta dentro del catálogo de software de la UCLA, y es de código abierto.
- Se tiene disponibilidad al autor del proyecto.
- Administra de manera centralizada procesos organizacionales que producen documentos.
- Define estructuras de seguridad y Perfiles de Usuario necesarias para la configuración de acceso a las distintas opciones y documentos.
- Registro de cada una de las modificaciones realizadas sobre los documentos generados en el sistema.
- Permite el registro de los flujos de trabajo que establecen los roles y actividades necesarios para la generación y concreción de un documento en particular.
- Permite dar seguimiento y monitoreo a las actividades desarrolladas en un Documento.
- Genera notificaciones vía correo electrónico y SMS, en situaciones de cambios de estado y asignación de responsabilidades en un Documento.
- Permite construir una estructura de significados para la clasificación y catalogación de un documento.
- Cambia la apariencia general del formato del sitio, a través de la definición y rehúso de plantillas.

Dado que la construcción del repositorio, manejo, control y seguimiento de los documentos digitales, será responsabilidad del SMC Obelisco queda por resolver la búsqueda distribuida, por lo que se realizó el diseño y desarrollo de una herramienta de integración y búsqueda en sistemas de contenidos en ambientes distribuidos para bibliotecas digitales permitiendo así a los SMC prestar estas funcionalidades.

Requisitos Funcionales

Los requisitos funcionales son las funcionalidades que requieren los SMC para cumplir con los requerimientos adquiridos en la fase de análisis de requerimientos.

De acuerdo a las características de negocio de la herramienta, los requerimientos son los siguientes.

Definición de las características de la biblioteca en la red de bibliotecas.

- **Búsqueda de Nodos:** descubrimiento de nodos (elementos de la red con un SMC) activos que poseen repositorio de documento donde realizar las búsquedas distribuidas.
- **Indexación y búsqueda de catálogos locales:** capacidad de indexar los datos anexados por el SMC y estructurarlos así como realizar búsquedas locales de los mismos.
- **Comunicación entre Nodos:** cada nodo debe tener la capacidad de transferir mensajes, cambios de estado e información entre los diferentes nodos
- **Búsqueda Distribuida:** cada nodo debe tener la capacidad de realizar búsquedas locales y distribuidas en los nodos descubiertos.
- **Acceso a Recurso remotos:** capacidad de acceder a recursos solicitados de otro nodo.

Requisitos no Funcionales

Los requisitos no funcionales son primordiales en el desarrollo de esta herramienta. Procedemos a nombrar los factores de calidad del software para el sistema que se desarrollaran.

- **Concurrente** Se desea ofrecer un sistema distribuido, sin restricciones tecnológicas notables. El motivo de este requisito es permitir el fácil acceso a un gran número de usuarios simultáneamente. Y a una mayor cantidad de material integrando a las diversas bibliotecas.
- **Libre de errores** Este requisito es deseable en todos los sistemas, aunque en algunos es especialmente crítico su cumplimiento.
- **Fiable** Directamente relacionado con el requisito anterior, el sistema deberá ser tolerante a fallos de subsistemas externos, incluido caídas de nodos
- **Mantenible** Será deseable que la implementación del sistema ofrezca suficiente modularidad para poder reemplazar partes del mismo con facilidad. Este factor de calidad será importante a dos niveles: fácil acceso a los mecanismos de configuración de la Herramienta, y bajo acoplamiento entre las diferentes estructuras arquitectónicas.
- **Escalable** la herramienta debe permitir acceder a documentos de diferentes tipos creciendo al crecer los tipos de documentos.

En consideración a los requisitos funcionales tenemos:

Diagramas de caso de Uso

En función a la lista de Funcionalidades definidas, se creó un diagrama de casos de uso, en el cual se presentan de manera grafica las actividades generales a ser desarrolladas en la Herramienta Hypatia (ver figura 8), como lo son definición de documentos y nodo, indexar documento, buscar nodos activos, buscar documentos en otros nodos, y acceder a recursos remotos.

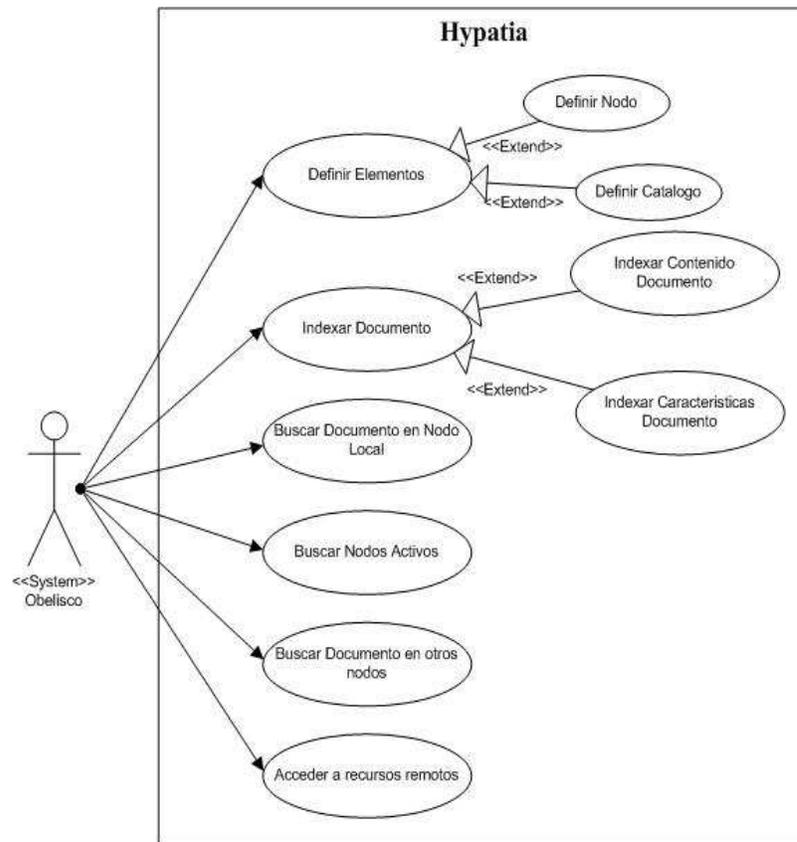


Figura 8. Diagrama de Casos de Uso de Hypatia.

Autor: El autor de la investigación

Diccionario de Actores

En el sistema se interactúan el personal de Biblioteca y los usuarios del mismo, pero estos interactúan con el gestor de contenidos, es este el que interactuara con la herramienta indicando solicitudes y recibiendo las entradas.

Segunda fase: Planificación de entregas

La metodología de desarrollo XP sugiere establecer prioridades sobre las tareas que ejecutará el sistema, con el fin de organizarlas para construirlas en cada iteración del proceso de desarrollo.

Una vez definidas las prioridades sobre las tareas y habiéndolas organizado en iteraciones, La Tabla 2 muestra la planificación de las iteraciones, mostrando el esfuerzo necesario en semanas para completar cada tarea del sistema.

Tabla 2: Plan de publicaciones (reléase plan)

Iteración	Tareas a Desarrollar	Semanas
0	<ul style="list-style-type: none"> • Describir Arquitectura de Software 	3
1	<ul style="list-style-type: none"> • Definición de Catálogo de búsqueda 	2
2	<ul style="list-style-type: none"> • Indexado de catálogos de Documentos • Búsqueda local de documentos. 	3
3	<ul style="list-style-type: none"> • Definición de Nodo • Búsqueda de Nodos • Comunicación entre Nodos. 	3
4	<ul style="list-style-type: none"> • Búsqueda Distribuida 	5

Fuente: El autor de la investigación.

Descripción de la herramienta

La herramienta para la integración y búsqueda en sistemas de gestión de contenidos para bibliotecas digitales, se denominó “HYPATIA”, y es un software que permitirá a los gestores de contenidos integrar documentos y realizar búsquedas distribuidas de contenido.

Esta debe estar desarrollada con tecnología que le permita su portabilidad a distintas plataformas de sistemas operativos, con el objeto de que no se limite al usuario a usar una sola plataforma y debe solventar los siguientes requerimientos funcionales y no funcionales.

Esta fue desarrollada como una Herramienta P2P para a los diferentes SMC con repositorio de datos, para compartir su material digital y realizar búsquedas de los mismos de manera distribuida.

Tercera Fase: Iteraciones

Iteración 0

Esta iteración se centra en definir la arquitectura que soportara el desarrollo de software. Así como establecer las herramientas tecnológicas más idóneas para su construcción y seleccionar los patrones de diseños más apropiados para el desarrollo de la misma.

Arquitectura de Software

Se dispone a construir la arquitectura de una herramienta que permita a un manejador de contenido ya existente, en este caso Obelisco, expandir sus funcionalidades, permitiéndole realizar búsquedas distribuidas por contenidos, como por atributos o catálogo. Esta requiere que pueda adaptarse a una amplia gama de formatos en la que pueden ser digitalizados los documentos, y una arquitectura para la búsqueda distribuida de contenido.

Esta arquitectura se basa en las características de una aplicación P2P, donde cada nodo es independiente, autocontenido, y maneja su propio repositorio con búsquedas locales, y podrá realizar búsqueda de nodos e información disponible en los

otros nodos, cada nodo requerirá de un SMC para el control y gestión de su repositorio de datos.

A continuación se describirán los componentes de la arquitectura y la conducta de esos componentes según se la percibe desde el resto del sistema y las formas en que los componentes interactúan y se coordinan para alcanzar la misión del sistema.

Nodo: el cual representara la biblioteca digital, con repositorio de datos, y que trabaja con SMC. Dichos nodos deberán tener las siguientes características:

- Este es autónomo e independiente.
- Tiene su propio motor de búsqueda.
- Puede publicar sus contenidos a grupos específicos.
- Tiene la capacidad descubrir otros nodos de sus grupos.
- Puede buscar material digital en otros nodos descubiertos.

Catálogo: el cual representa la ficha del documento digital con una serie de atributos; este puede estar presente dentro del SMC, pero para la búsqueda, Hypatia requerirá un catálogo propio, este será estructurado por medio de la indexación del mismo, permitiendo la búsqueda del documento por medio de los atributos definidos o el contenido del mismo de acuerdo a la configuración de este. Dicho catálogo presenta las siguientes características.

- Debe poder ser configurado.
- Su estructura de atributos debe ser fija y configurable.
- Los atributos definidos en catálogo de los nodos del grupo deben ser iguales.
- Permitirá búsquedas por medio del contenido del documento digital, esto de acuerdo a la configuración del mismo.

Grupo: este representa las redes a las que pertenece el nodo, discriminando los nodos que las mismas puedan descubrir, aportando un atributo de seguridad, y sus características son las siguientes.

- Cada grupo puede tener subgrupos.
- El grupo va a determinar los servicios a prestar entre los nodos.
- La definición de catálogo de documento de los nodos debe ser igual en características.

Además de los elementos mencionados anteriormente, esta requiere para su funcionamiento, de un manejador de contenidos, que pueda usar los servicios prestados por Hypatia, por lo que la misma no presentara una interfaz grafica (UI), por no ser usada por un actor como tal, sino por una aplicación. La lógica de negocio está enfocada al comportamiento del documento y su catálogo así como las características de los diferentes nodos o bibliotecas que van a utilizar dicha herramienta.

El funcionamiento de la herramienta es el siguiente, el acceso a los documentos digitales va ser por medio de un SMC en cada nodo con repositorio, en el cual se podrán efectuar búsquedas, estas van a ser efectuada de manera local y en los otros nodos descubiertos en el grupo o grupos comunes, por medio de mensajes, que solicitan una búsqueda local en cada nodo y las coincidencias son retornadas al nodo que las solicita para que de esta manera la herramienta genere un evento que pueda ser escuchado por el SMC e incorpore a los resultados de la búsqueda, los de los otros nodos. Esto se puede ver gráficamente en la figura 9, estas búsquedas podrán ser efectuadas por cualquiera de los elementos definidos en el catálogo del documento.

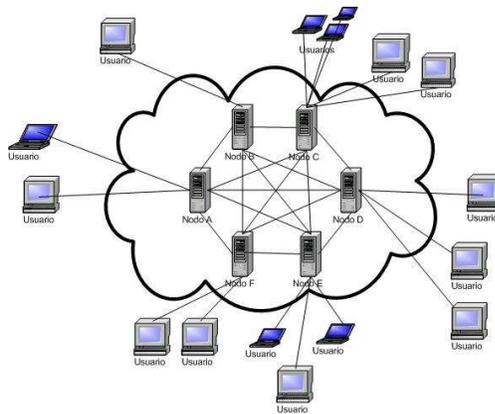


Figura 9. Diagrama de Contexto de funcionamiento de Hypatia
Autor: El autor de la investigación

La arquitectura basada en capas, se enfoca en la distribución de roles y responsabilidades de los componentes de forma jerárquica proveyendo una forma muy efectiva de separación de responsabilidades. El rol indica el modo y tipo de interacción con otras capas, y la responsabilidad indica la funcionalidad que está siendo desarrollada.

La arquitectura de Hypatia, va a ser una arquitectura de capas, se desarrollo en tres capas: El núcleo del documento, el Núcleo del Nodo, y Una capa de Servicios como se puede apreciar en la figura 10.

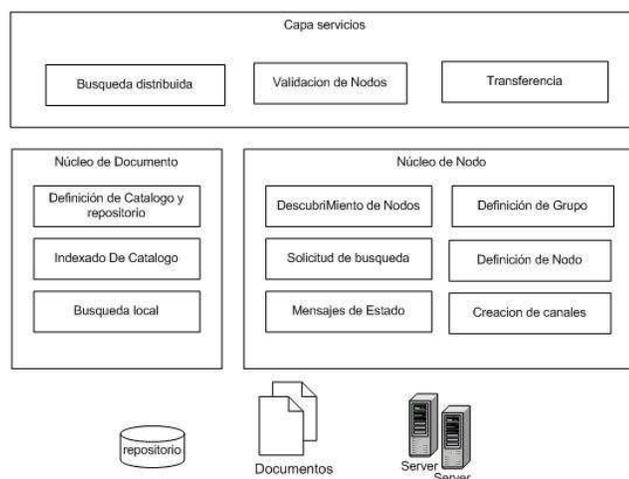


Figura10. Arquitectura en capas de Hypatia.
Autor: El autor de la investigación

Capa del Núcleo de Documento

Esta capa es la responsable de la estructuración de los documentos digitales, así como de la búsqueda local de los mismos y contiene los elementos esenciales para la manipulación de documentos y su catálogo tales como:

- Definir catálogo de documentos, atributos y características.
- Realizar indexado de documentos por catálogo y contenido del mismo de ser requerido.
- Realizar búsquedas locales de contenido, condicionado a las características del catálogo definido.

En esta capa está presente además, el código referente a la indexación de el contenido de cualquier documento, el cual podrá ser enriquecido manteniendo la interfaz definida para esto.

Capa del Núcleo del nodo

Esta capa es la responsable de la manipulación de nodos y la comunicación de los mismos por medio de arquitectura P2P, Esta contiene los servicios esenciales para prestar servicios tales como:

- Definición de grupos, características del mismo.
- Definición de nodo, características y grupos a los que pertenece.
- Descubrimiento de nodos, el cual le permitirá detectar nodos con servicios disponibles, para ser usados dentro de los grupos.
- Verificación de grupo de trabajo, para permisologías y restricciones de acceso.
- Solicitud de búsqueda a nodos descubiertos.
- Mensajes de estado, para transferir la petición de búsqueda o servicios o indicar cambios en estado de una solicitud o servicios.

- Creación de canales de comunicación para transferencia de información o servicios.
- Envío de Advertisement con resultados de la búsqueda solicitada y actualización del estado de la consulta

Capa de Servicios

Esta capa es la responsable de la búsqueda distribuida de documentos y contiene los elementos esenciales para los servicios que mezclen los dos núcleos, tales como:

- Servicio de búsqueda distribuida, permite realizar búsqueda en nodo en el que se realiza así como en los nodos descubiertos.
- Validación de nodos, verifica que la definición de catálogo se la misma a los nodos descubiertos.
- Transferencia de documento entre nodos.

Como se mencionó anteriormente la arquitectura está dividido en tres capas, cada capa va a estar definida en un componente, el núcleo de documento en el componente hypatiaDocumento, el núcleo de nodo en el componente hypatiaRed, y la capa de servicio en el hypatiaServicio como se puede apreciar en la figura 11, así como, su relaciones entre ellos.

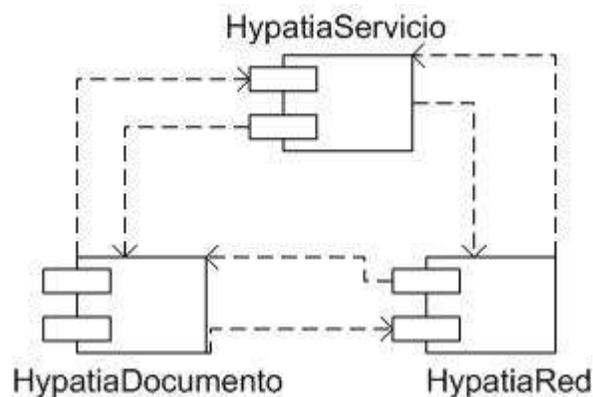


Figura 11. Modelo de Componentes

Fuente: El autor de la Investigación.

El componente ‘hypatiaDocumento’ se centra en el desarrollo de una serie de elementos, que permitirán la configuración del catalogo, indexación y búsqueda de documentos de manera local.

El componente ‘hypatiaRed dentro de la arquitectura es quien se encarga del manejo del P2p, la comunicación entre los nodos y el control de los mensaje entre los mismos.

Por último, el componente “hypatiaServicio” se encarga de la búsqueda distribuida, aprovechando los dos componentes mencionados anteriormente.

El SMC obelisco trabajara con el componente HypatiaServicio, para las búsquedas distribuidas y para el indexado lo hará por medio de hypatiaDocumento a través de un web service clienteObelisco que lo realizara un indexado definido por tiempo, con el nuevo material incorporado en ese intervalo.

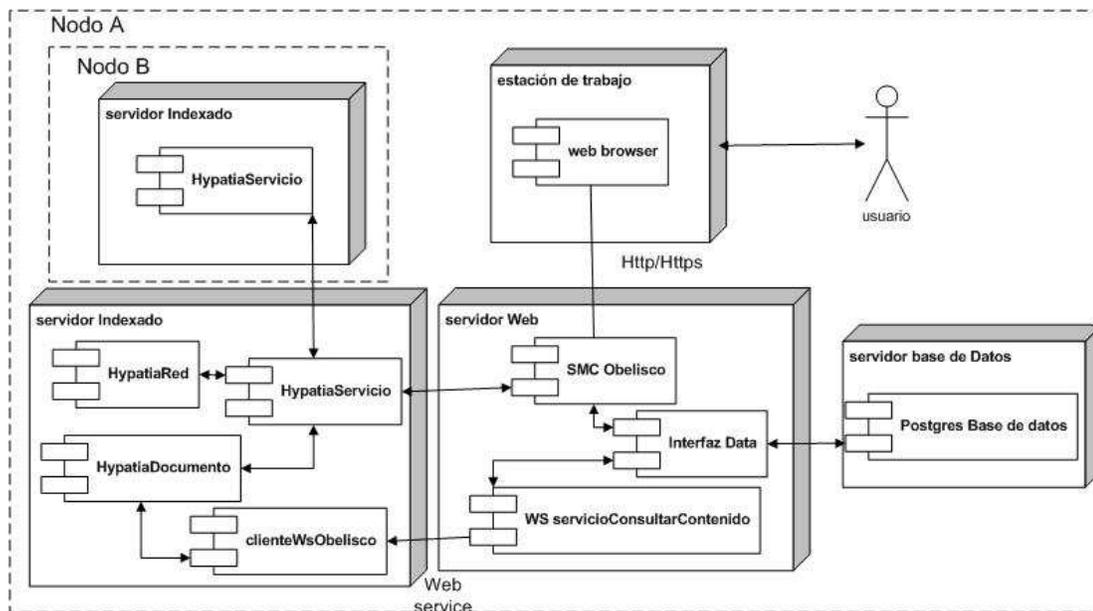


Figura 12. Diagrama de Despliegue de Hypatia.

Fuente: El autor de la investigación.

Esta arquitectura está basada en el uso de patrones para solventar características del sistema, así como atributos no funcionales, y estos patrones son los siguientes:

El patrón Strategy; debido a la extensa cantidad de tipos de documentos y que pueden aparecer nuevos formatos, este permitirá así, una estrategia para cada tipo de documento, garantizando la escalabilidad del mismo.

El patrón de inyección de dependencia, en el que se suministran objetos a una clase en lugar de ser la propia clase quien cree el objeto, será usado para suministrar la estrategia adecuada para un tipo de documento, aumentando la mantenibilidad del mismo

El patrón Factory para la construcción de Entidades como Nodo, Grupo y Catálogo de documento, para mejorar la mantenibilidad del mismo.

El patrón observador para avisar a la aplicación o SMC que solicito la búsqueda que un nodo ha enviado su resultado para ser anexado a la lista de documentos encontrados.

Infraestructura y Herramientas utilizadas

Un framework es una estructura conceptual y tecnológica de soporte definida, normalmente con artefactos o módulos de software concretos, en base a la cual otro proyecto de software puede ser organizado y desarrollado. Típicamente, puede incluir soporte de programas, bibliotecas y un lenguaje interpretado entre otros programas para ayudar a desarrollar y unir los diferentes componentes de un proyecto. Para el desarrollo de hypatia se usaron una serie de frameworks, los cuales son descritas a continuación, justificando su elección y para qué son usadas.

Tabla 3. Herramientas usadas

	Justificación	Uso
Spring	Es una herramienta de integración, de código abierto, muy modular, enfocada en el manejo de objetos de negocio, dentro de una arquitectura en capas.	Para la inyección de dependencia en el sistema.
Lucene	Es un proyecto de gran madurez y eficiencia en sus búsquedas de documentos digitales no estructurados, o semi estructurados.	Para la máquina de búsqueda de documentos catalogados, y estructuración de documentos por medio de indexado.
JXTA	Por ser la Herramienta más madura, para manejo de protocolos P2P y presenta características como computación distribuida, intercambio de archivos, seguridad, disponibilidad, manejo de clúster, etc.	Para la definición de nodos y grupos, así como comunicación, administración de información entre los nodos por protocolo P2P.
XDOM	Es una biblioteca de código abierto, que incluye sobrecarga de métodos, colecciones, entre otros, y esta optimizada para java.	Para manipulación de datos XML.

Fuente: El autor de la Investigación.

Iteración 1

Objetivo

Configuración de las entidades del documento, así como del repositorio.

A continuación se presenta un diagrama de caso de usos en la figura 13, con las actividades que se presentan para la ilustración de definición de Catálogo y manejo del repositorio:

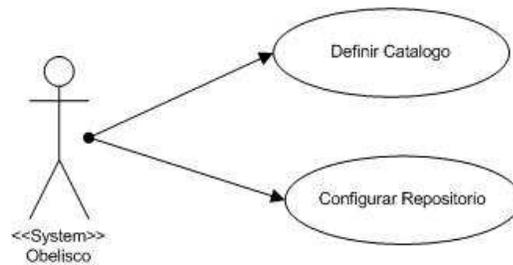


Figura 13. Diagrama de Casos de Uso para Iteración N° 1.

Fuente: El autor de la Investigación.

Para el desarrollo de esta Iteración se procedió a la definición de las Clases entidad del documento, como lo es la clase RepositorioIndexado del repositorio, y CatalogoDocumento que representa el catálogo del documento e implementa la Interfaz ICatalogo.

Cada objeto de tipo ICatalogo tiene una estructura o atributos definidos en el repositorio, dando la potencialidad de que dichos atributos puedan ser configurables, pero ya definidos el objeto solo podrá cargar estos atributos, y cada objeto será creado y manipulado por una clase encargada del control y factory de la misma, como los son CatalogoManager y RepositorioManager, esto se puede observar en el diagrama de clases de la figura 14.

La lista de Atributos a catalogar es almacenada en un documento XML, de no existir dicha configuración al crear un catálogo para un documento lanzara una excepción, por lo que es necesario la creación de este archivo y este será el único medio donde se definirá cada uno de los atributos a catalogar de cualquier documento en el sistema, así como si este va a manejar el contenido del mismo para las búsquedas, para la manipulación y creación de los archivos de configuración fue usado XDOM.

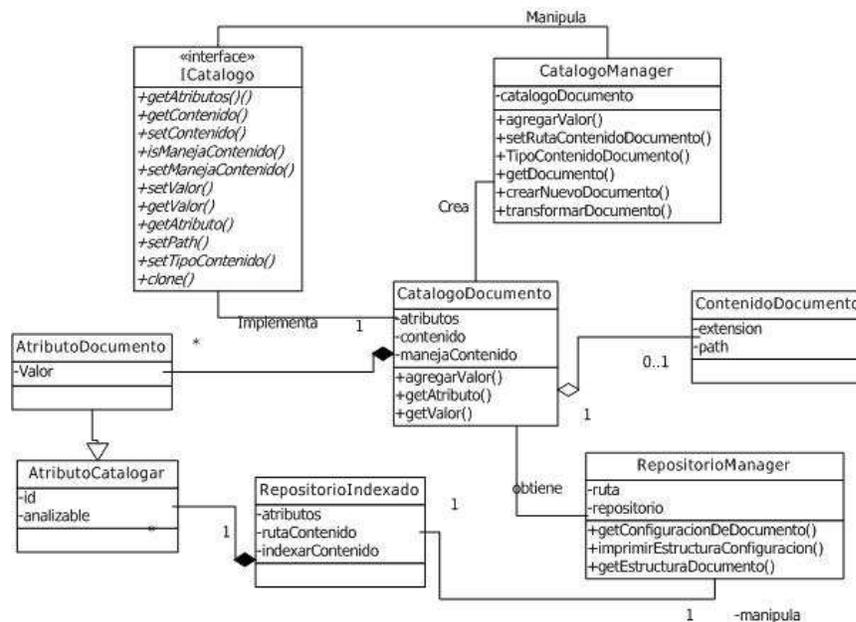


Figura 14. Diagrama de Clases definición de Catálogo y Repositorio
Fuente: El autor de la investigación.

Para la creación de cualquier Catálogo de documento, se realiza el siguiente proceso vea figura 15. Se realiza un llamado del método nuevoCatalogo() de la clase CatalogoManager, el cual obtiene una instancia del documento del repositorio(singleton), carga la plantilla del documento de acuerdo a los atributos definidos en el XML de configuración, luego la se clona la plantilla para generar un documento con la estructura de atributos, sin acceso a modificación de la estructura de los mismos.

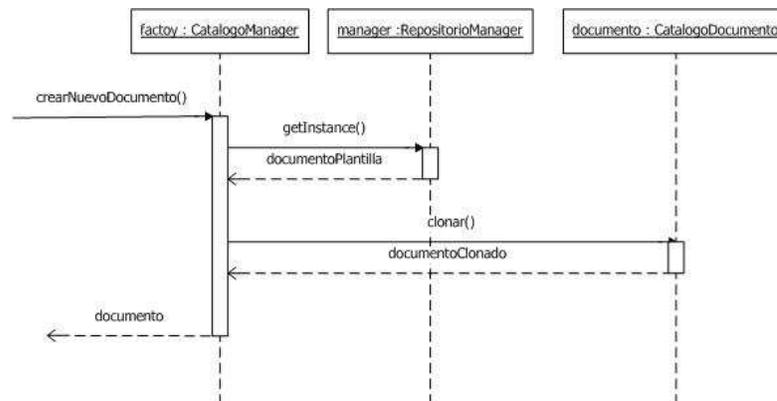


Figura 15. Diagrama de Secuencia de factory de documento
Fuente: El autor de la investigación.

Iteración 2

Objetivo

Permitir el indexado de contenido de cualquier tipo de documento, su búsqueda local como se puede observar en la figura 16.

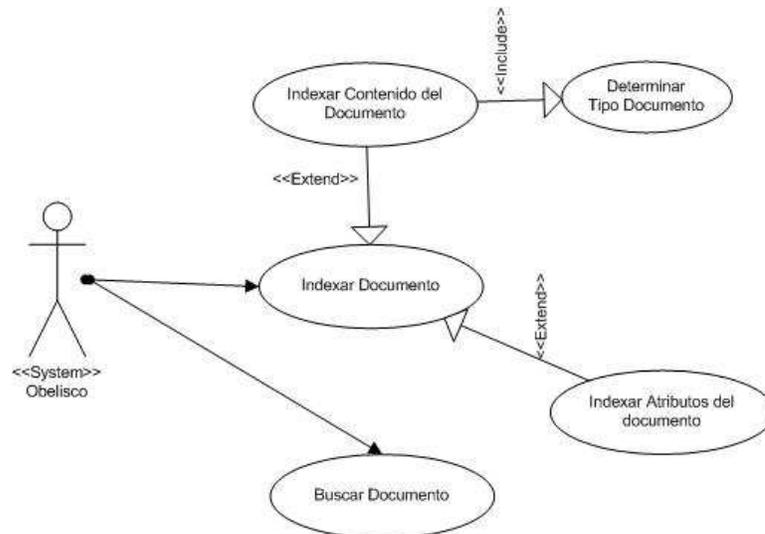


Figura 16. Diagrama de Casos de Uso para Iteración N° 2.

Fuente: El autor de la Investigación.

Como se puede observar en la figura 17, se puede apreciar el catálogo de clases de las funcionalidades de búsqueda local, e indexado.

En esta iteración se generan dos servicios por medio de Lucene a las entidades definidas en la primera iteración, como lo son el indexado de catálogos o documentos por medio de la clase Indexador y la búsqueda local con la clase Buscador.

El indexado permite la estructuración del documento, tanto de los atributos de catálogo, así como, de la información del documento en cuestión, aplicando una estrategia particular para la extracción de la información del mismo, por medio del patrón strategy, apoyado por el patrón inyección de dependencia el cual dependiendo

del tipo de documento carga la interfaz usada con la estrategia a implementar, a continuación se muestra el diagrama de clases de esta iteración en la figura 16.

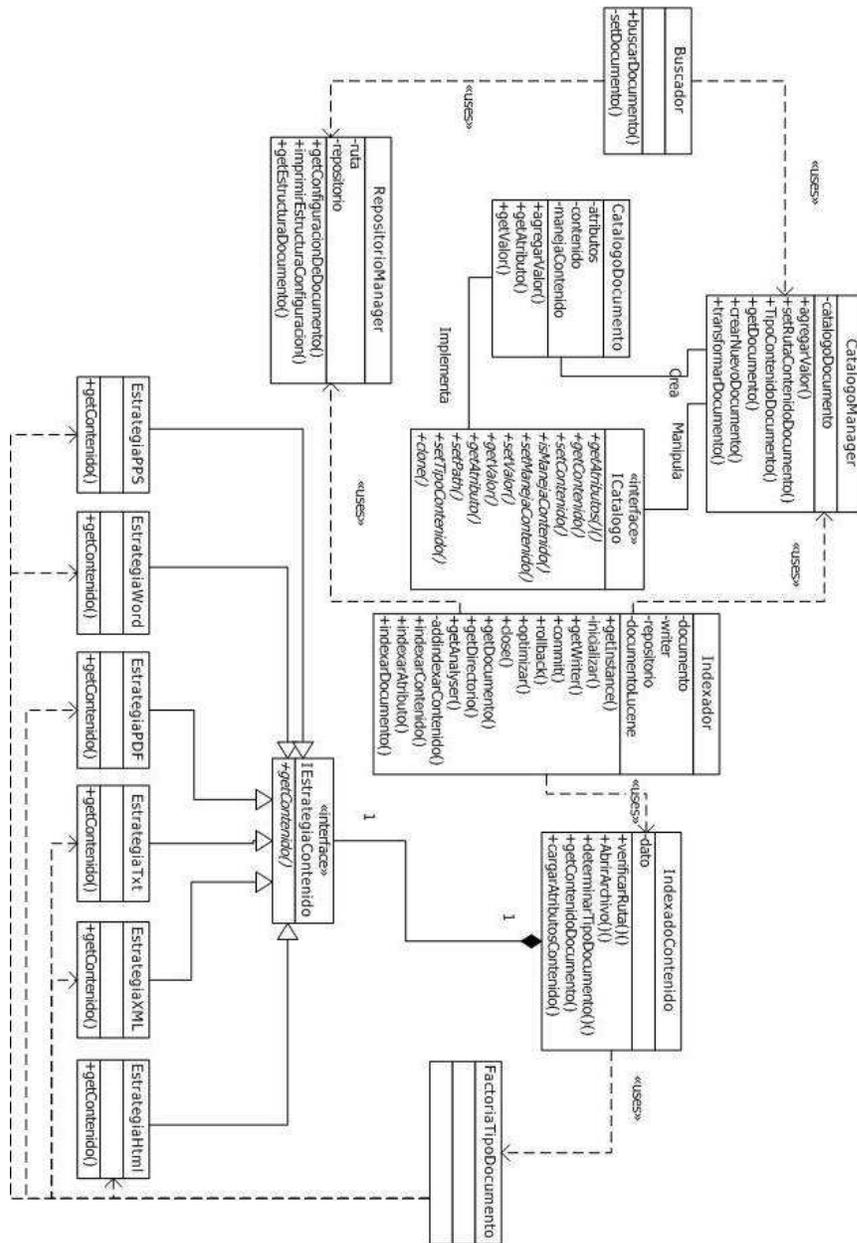


Figura 17. Diagrama de Clases Indexado y búsqueda de contenido
Fuente: El autor de la investigación.

El servicio de Búsqueda local se realiza por medio de la clase Buscador que implementa el método buscarDocumento(atributo, Valor) la cual buscara en valor en el atributo indicado en los documentos del repositorio y las coincidencias serán enviados a un lista de tipo ICatalogo por medio de la clase DocumentoManager, este proceso se puede apreciar en la figura 18

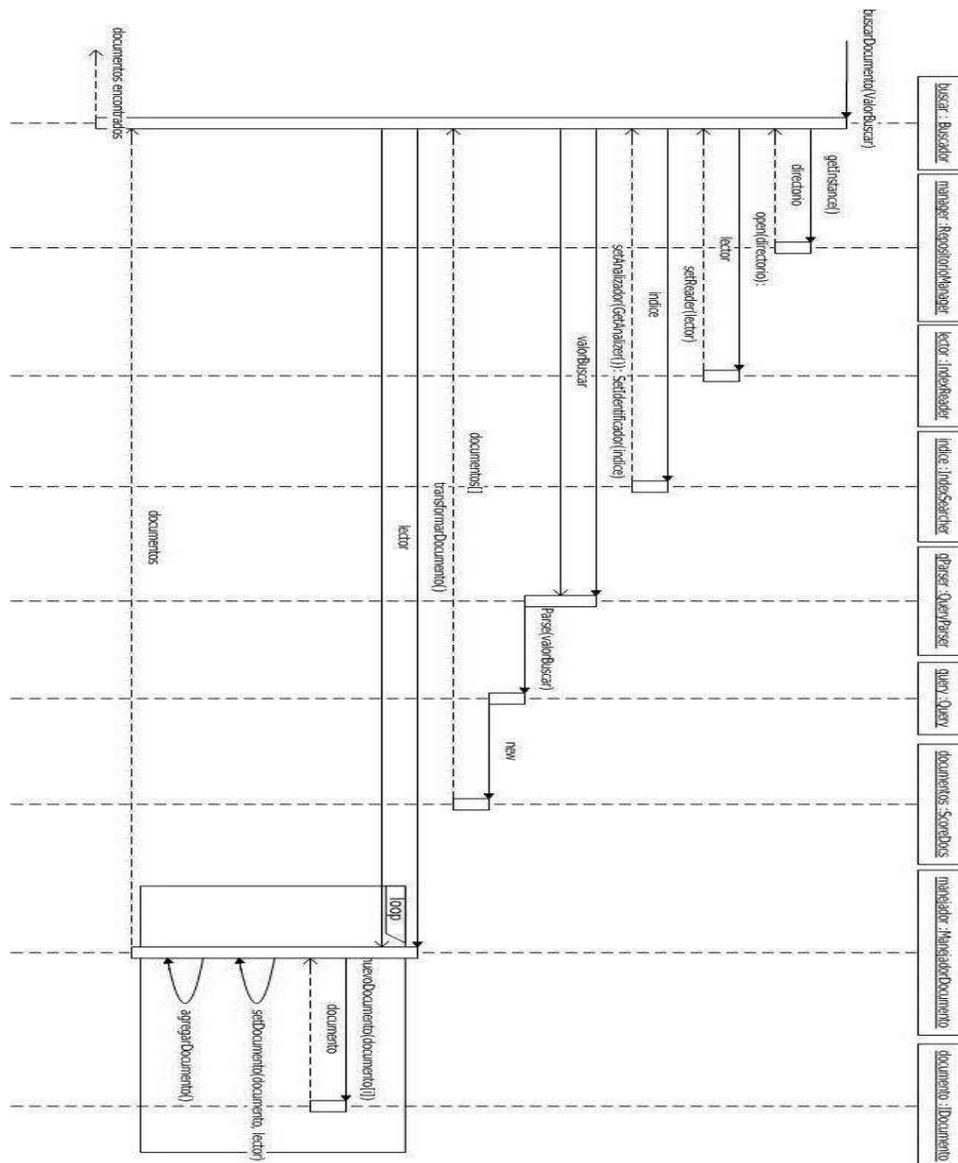


Figura 18. Diagrama de Secuencia de búsqueda local
Fuente: El autor de la investigación.

El servicio de indexado se realiza por medio de la clase Indexador que implementa el método indexarDocumento() como se puede apreciar en la figura 19.

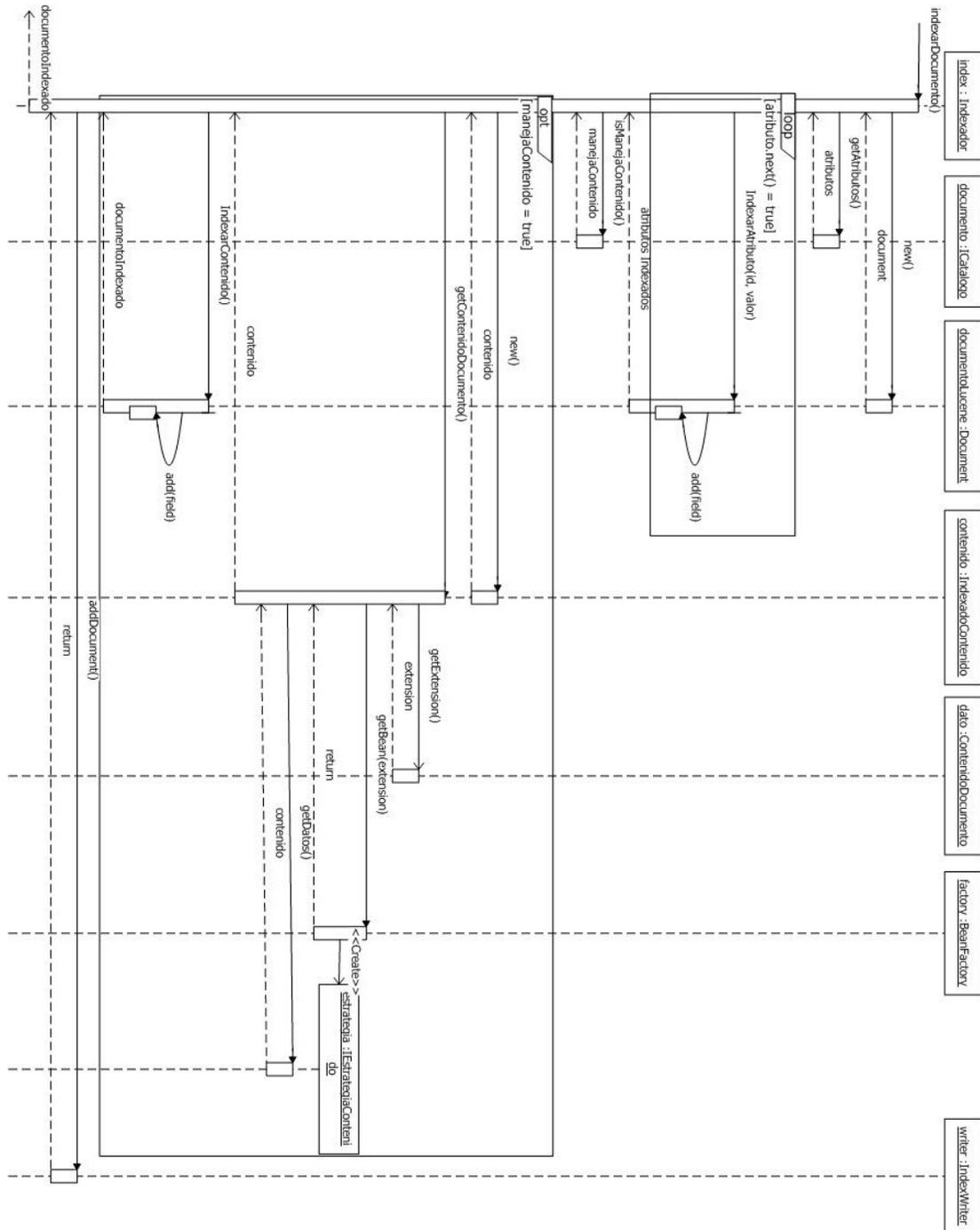


Figura 19. Diagrama de Secuencia de Indexado

Fuente: El autor de la investigación.

Esta iteración culmina la capa de núcleo de documento, que controla la catalogación, indexado y búsqueda local de documentos. El diagrama de clases se visualiza en la figura 20 y el diagrama de paquetes de la fase en la figura 21.

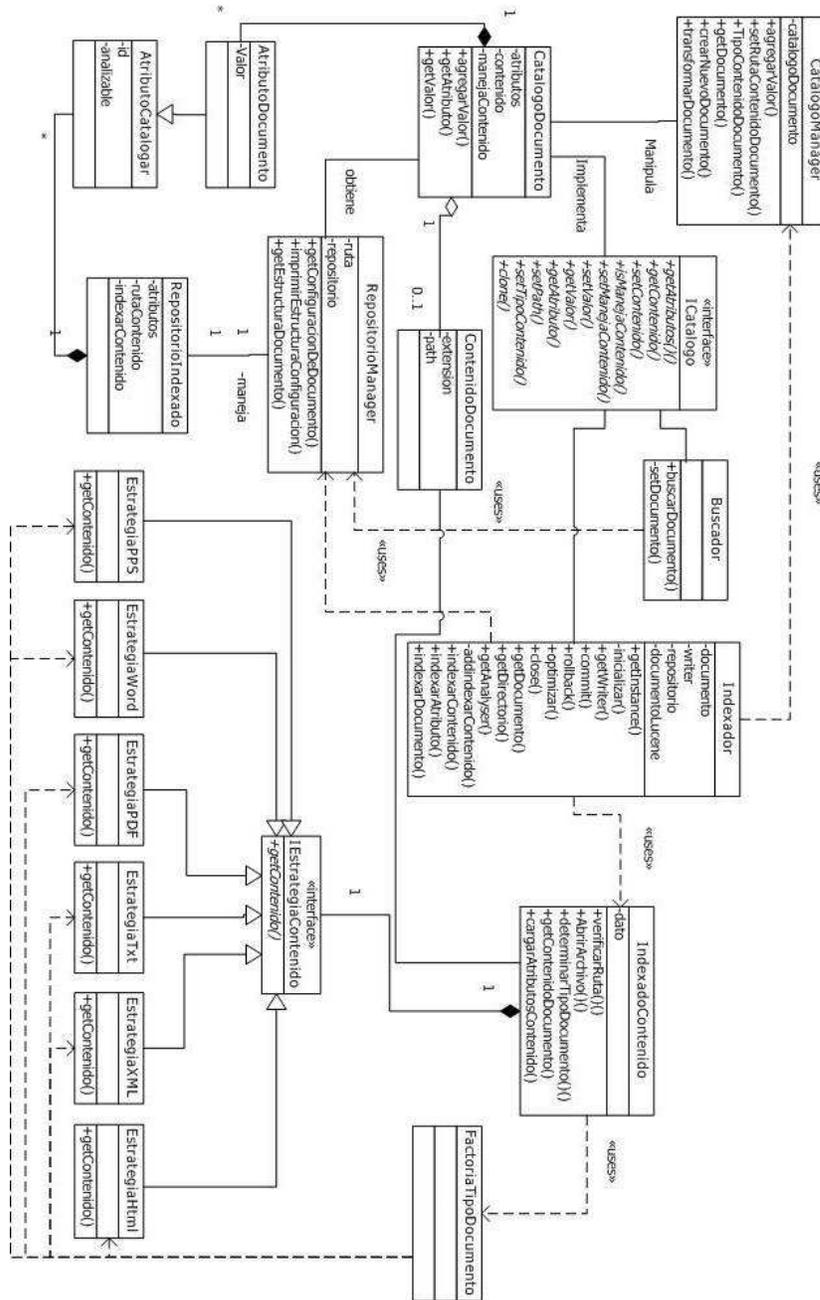


Figura 20. Diagrama de Clases que modela la capa Núcleo de contenido
Fuente: El autor de la investigación.

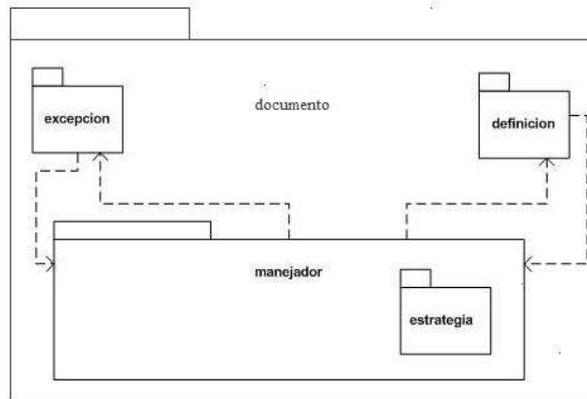


Figura 21. Diagrama de paquetes de la capa Núcleo de contenido
Fuente: El autor de la investigación.

Iteración 3

Objetivo

Definición de nodo y grupo así como, el descubrimiento de Nodos o bibliotecas asociadas y establecer la comunicación de las mismas, desarrollando la capa de nodo.

A continuación se presenta un diagrama de caso de usos en la figura 22, con las actividades que se presentan para la definición de Nodo, así como buscar nodos activos.

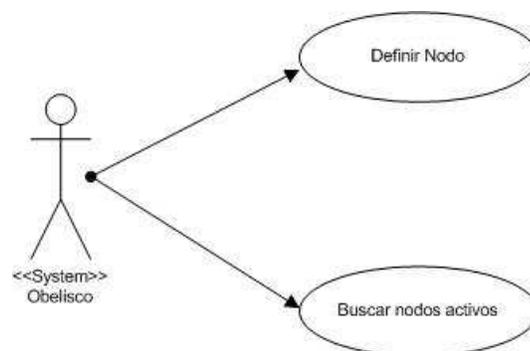


Figura 22. Diagrama de Casos de Uso para Iteración N° 3.
Fuente: El autor de la Investigación.

Esta iteración implementa los servicios P2P de la herramienta, por medio del JXTA, Configurando y definiendo las Entidades Nodo y grupo, así como, servicios propios de JXTA como el monitoring service para monitorear a otros miembros del mismo grupo de iguales, así como el control de los pipes definiendo una tubería de entrada Controlada ClienteManager y una de salida por la clase ServidorManager las cuales capturan eventos descubiertos , permitiendo la transferencia de mensajes, peticiones, como se puede apreciar en la figura 23

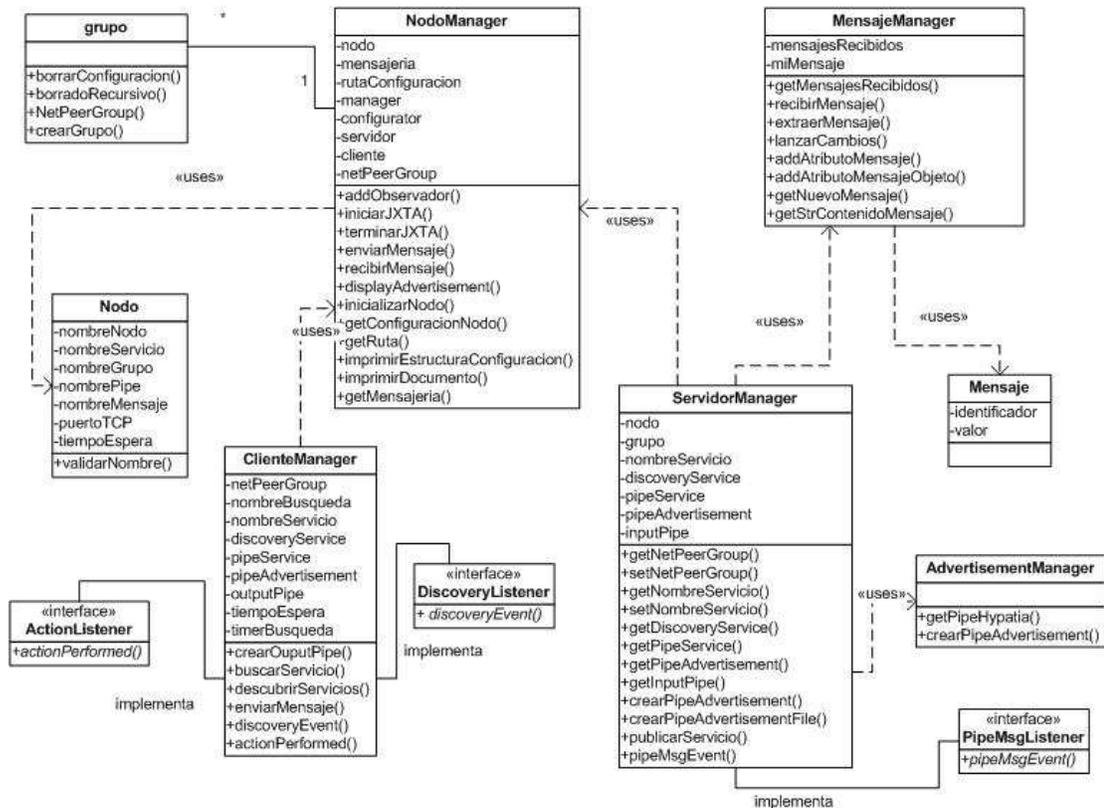


Figura 23. Diagrama de Clases de la capa núcleo de nodo
Fuente: El autor de la investigación.

Iteración 4

Objetivo

Permitir la búsqueda distribuida y acceso a la información, desarrollando la capa de servicios.

A continuación se muestra un diagrama de caso de usos en la figura 24, con las actividades que presenta en la capa de servicios.

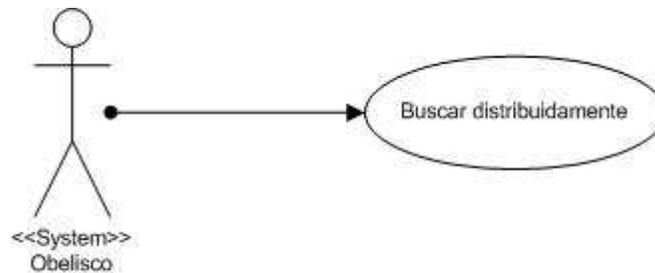


Figura 24. Diagrama de Casos de Uso para Iteración N° 4.

Fuente: El autor de la Investigación.

En esta fase se integrara el manejo de documentos, con los servicios de red para generar la búsqueda distribuida, para esto se usaron algunas clases de las Iteraciones anteriores y se le anexara la clase `BuscadorIntegrador` la cual solitaria una búsqueda local a `Buscador` y publica una petición en los nodos por medio del `ServidorManager` la cual será enviada en forma de mensaje.

En la figura 25 se puede apreciar el llamado a la función de búsqueda, efectuada por el SMC, el cual invocara la clase `BuscadorIntegrador`, el cual generar la búsqueda local y enviara mensajes de búsqueda a cada uno de los nodos descubiertos, los cuales le enviaran los documentos con las coincidencias de los mismos, siendo estos retornados por medio de un evento de nueva búsqueda detectada que debe ser escuchado en el SMC.

JXTA implementa el patrón publisher-suscriber el cual será usado en esta aplicación, para el control de mensajes y respuestas a las búsquedas.

Si un nodo captura un evento de solicitud de búsqueda en el ClienteManager ente instanciará la clase BusquedaRemota que generara la búsqueda en ese nodo y el resultado lo publicara.



Figura 25. Interfaz de búsqueda en SMC Obelisco

Fuente: El autor de la investigación.

Si el nodo captura un evento de obtenerResultado este será procesado por el ResultadoManager el cual generara todos los resultados de los nodos y maneja el estado de la solicitud, actualizado el resultado solicitado por BuscadorIntegrador, La clase ResultadoManager es la responsable de controlar la información obtenida de las búsquedas de los nodos, disparando eventos que serán capturados por el SMC a medida que reciba información de los nodos, dependiendo del estado de la solicitud, de existir un cambio en el mismo suspende la emisión de eventos, para esto se

implemento el patrón observador, a continuación se muestra el diagrama de clase de esta fase en la figura 26.

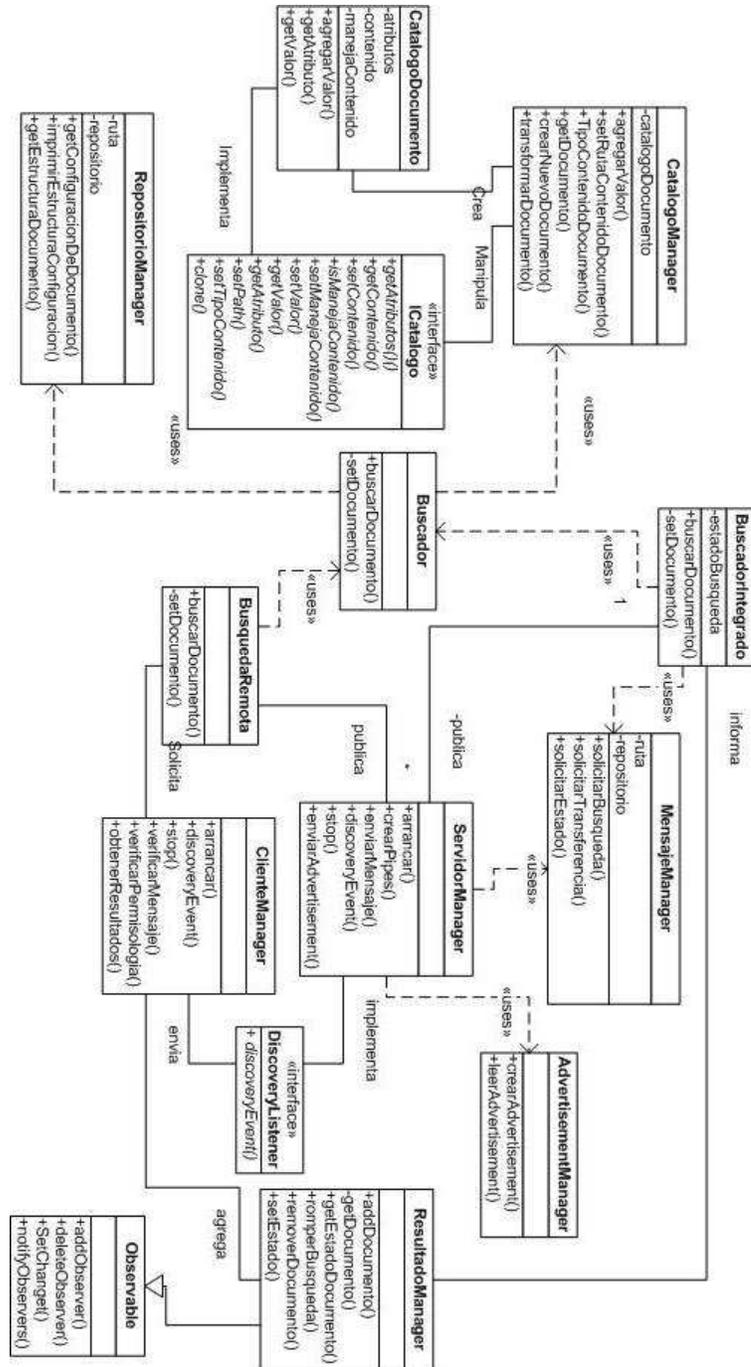


Figura 26. Diagrama de Clases de la capa de servicios
Fuente: El autor de la investigación.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

Al desarrollar el presente trabajo de investigación, el autor puede establecer las siguientes conclusiones:

Los métodos de análisis de requerimientos son críticos para el desarrollo del sistema, una visión más clara de lo que se va a desarrollar, las entrevistas así como la observación y la revisión bibliográfica ayudaron a detectar las necesidades y requerimientos relacionados al proceso de búsqueda distribuida y control de material digital de una biblioteca digital y su aplicación en las bibliotecas de la UCLA.

Este análisis de requerimientos apoyado por el empleo de una metodología ágil para el desarrollo de una aplicación facilitó el desarrollo de misma, la aplicación de prácticas de XP permitió concentrarse en la obtención del producto de acuerdo a las expectativas de las Bibliotecas.

Después de haber analizado la información referente a SMC, protocolos P2P y agentes y metodologías de búsqueda, puede apreciarse que la integración de estos conceptos al contexto de bibliotecas digitales sugiere el desarrollo de sistemas que incrementan de una manera significativa la utilidad y la eficiencia de éstas.

La colaboración entre bibliotecas es un elemento muy importante en el proceso de búsqueda y recopilación de información. El contexto digital ofrece oportunidades para nuevas maneras de colaboración y comunicación permitiendo la creación de grupos virtuales de bibliotecas que cooperan en el uso de material digital.

Para este tipo de alianzas una red P2P es más robusta en el sentido de que si falla un nodo, los otros nodos no se ven afectados, solo se ausentaron los servicios del mismo, pero no afectará al sistema como tal. Para esto JXTA demostró ser un protocolo P2P muy robusto, ofreciendo una gama de servicios muy amplia para el control de nodos de una red y los servicios de la misma.

El uso de distintos patrones de diseño permitió garantizar algunos atributos no funcionales requeridos para el funcionamiento de HYPATIA.

La plataforma de implementación de HYPATIA es económica, dado que su desarrollo está basado en herramientas de código abierto, así como su implementación es libre de restricciones de sistemas y hardware donde se instale.

HYPATIA plantea un flujo de trabajo sencillo y controlable para las búsquedas distribuidas, esta herramienta puede ser usada por diversos manejadores de contenidos, permitiendo expandir su potencialidad.

La herramienta HYPATIA manejada por un SMC como el OBELISCO constituye un aporte importante, para el desarrollo de las bibliotecas digitales de la UCLA, incidiendo esto en el proceso enseñanza-aprendizaje e investigación dentro de la UCLA o bibliotecas que lo usen ya que permite expandir la cantidad de material disponible con ayuda de posibles alianzas.

Recomendaciones

Se recomienda aumentar la cantidad de material a digitalizar, así como anexar material digital de dominio público, como los son las obras de la antigüedad, las cuales incluye un amplio número de ediciones publicadas antes de 1923, así como el acuerdo con aquellas empresas o instituciones que fomentan y generan material de libre acceso, por no tener problemas de derecho de autor.

Se aconseja definir un catálogo común con mayor adaptación para material digital y diferenciarlo del material físico, estandarizar la codificación interna asociada al material generado en cada decanato, así como implementar el uso de un manejador de contenido como el obelisco para el manejo y control del material digital de las mismas, de este modo se pueden evitar problemas a la hora de integrar las bibliotecas de los decanatos.

La alianza entre bibliotecas permitirá por medio de la implementación de HYPATIA y un gestor de contenidos, la expansión de la cantidad de material que estas prestan y el cambio de paradigma de tener que ir a una ubicación física para acceder a la información confiable como es la dada por las bibliotecas.

Por último, es recomendable expandir las potencialidades de la herramienta para que esta pueda ser más versátil y adaptarse a otros medios diferentes a las bibliotecas, así como incorporar Web semántica en las mismas para expandir su potencial de búsqueda.

REFERENCIAS BIBLIOGRAFICAS

Amaro, S y otro 2007. Metodologías Ágiles.
Universidad Nacional de trujillo. Peru.

Balestrini, M. y otros 1997. Cómo se elabora el proyecto de investigación.
Fotolito Quintana. Caracas

Backx, P. y otros 2002. A comparison of peer-to-peer architectures.
Proceedings of Eurescom Summit 2002. Alemania, Heidelberg.

Barceló, M y otros 2001, Hacia una economía del conocimiento,
ESIC Editorial,

Bateman, T. y Otros 2004. Administración Un nuevo panorama competitivo.
McGraw-Hill Interamericana. México.

Batini, C y otros 2004, Diseño conceptual de bases de datos.
Ediciones Díaz de Santos.

Berners-Lee T. 2000, Semantic Web -XML2000. Architecture.
URL: <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide11-0.html>

Born, G. 2001, Compendium HTML: Con XHTML, DHTML, CSS, XML, Xsl y
Wml, Marcombo

Brochard, J 2001, XML Conceptos E Implementación.
ENI

Brogan, M. 2006, Contexts and Contributions: Building the Distributed Library.
Digital Library Federation, Washington, D.C.

Buschmann, F 1996, Pattern Oriented Software Architecture, Volume 1: A System
of Patterns.
Wiley & Sons

Cabrera, J 2005, Diseño e implementación de un sistema en Web de biblioteca digital
de documentos de literatura científica.
UNIVERSIDAD TECNOLÓGICA DE LA MIXTECA

Canós, L y otro 2002. Metodologías Ágiles en el Desarrollo de Software.
Universidad Politécnica de Valencia, Valencia. España.

- Casas, R y otra 2001, La formación de redes de conocimiento.
Anthropos., Mexico.
- Christensen, C. y otro 1999, El dilema de los innovadores
Ediciones Granica S.A.
- Clements, P 1996 . “A Survey of Architecture Description Languages”.
Proceedings of the International Workshop on Software Specification and Design,
Alemania.
- David, P. y otros 2002, La Sociedad del conocimiento
Revista internacional de ciencias sociales, Número 171.
- Di Stefano, M. 2005, Distributed data management for grid computing.
John Wiley and Sons.
- Fava, C y otros 2004, La nueva gestión en las bibliotecas virtuales.
Boletín de la Asociación Andaluza de Bibliotecarios,
- Federación Española de Sociedades de Archivística Biblioteconomía y
Documentación 1994, Documat 94
Universidad de Oviedo
- Feria, L. 2002 Bibliotecas digitales
UCOL
- Fernández, M 2005, Introducción a la gestión (management).
Universidad Politécnica de Valencia. Valencia.
- Giner, F y otro 2004, Los sistemas de información en la sociedad del conocimiento.
ESIC Editorial
- Gnosis (2004), Modelado de Acervos en Bibliotecas Digitales,
Universidad de Guadalajara
- Gospodnetic, O y otro 2005, Lucene in Action
Editorial Manning Publications Co.
- Gómez, C y otros 2003, Diseño de sistemas software en UML
Ediciones UPC
- Guaracy, J. y otra 2006, Impacto tecnológico y arquitectura en bibliotecas
Publicado por Alfagrama Ediciones
- Herman, I. 2009, Semantic Web Activity, URL: <http://www.w3.org/2001/sw/>

W3C, Junio, 2009

Hernández, R y Otros. 2001. Metodología de la Investigación.
McGraw Hill. Interamericana S.A, México.

Hillside Group 2003, Home of the Patterns Library, URL: <http://hillside.net/> .

Howard, J 2003, When is Web Content Management Right for an Intranet?.
URL: http://www.intranetjournal.com/articles/200312/pij_12_17_03a.html
Intranet Journal Diciembre, 2003

Hurtado S 2008, Representación de la arquitectura de software usando UML.
Universidad Icesi

Hurtado I. 1997. Metodología de la Investigación Holística.
SYPAL, Caracas.

Ince, D. 2001, A dictionary of the Internet.
University Press. Oxford.

Joseph, J. y otro 2004, Grid computing.
Prentice Hall PTR

Joomla spanish, Administrador 2009, Sistema de manejo de contenido (smc),
URL:http://www.fguadalupe.com/joomla/index.php?option=com_content&view=article&id=11&Itemid=10
CMS - Manejadores de Contenido, Mayo, 2009

JXTA 2008, JXTA Community. URL: <https://jxta.dev.java.net/>
Julio, 2008

Lara, P y otro 2006, La organización del conocimiento en Internet.
Editorial UOC, Barcelona.

Laudon, K y otro 2004, Sistemas de Información Gerencial.
Pearson Prentice Hall, Edición 8

Lee, A 1992. The Thirteen International Conference on Information Systems.
Workshop on two techniques for qualitative data analysis: Actino Research and
Ethnography

López, J. 2005, Planificar la formación con calidad.
WK Educación

Magán, J 2003, Temas de Biblioteconomía Universitaria y General,

publicado por Editorial Complutense

Martínez, S. 2007, Biblioteca digital.
Alfagrama Ediciones

Méndez, A. y otro 2008 La investigación en la era de la INFORMACION.
Publicado por UNAM.

Méndez, E (1999), RDF: Un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio.

URL:<http://www.bib.uc3m.es/~mendez/publicaciones/7jc99/rdf.htm>
Jornadas Catalanes de Documentación, Noviembre, 1999

Mendoza, L. 2005, Estudio de Tecnologías Middleware para Sistemas Peer-to-Peer. Memorias de las VII Jornadas Internacionales de las Ciencias Computacionales", LISI, Universidad Simón Bolívar, Vol. 1.

Nuño, M y otra 2004 , La nueva gestión en las bibliotecas virtuales.
Boletín de la Asociación Andaluza de Bibliotecarios, ISSN 0213-6333, Año nº 19, Nº 74, Dialnet.

Orera, L. 2008, Preservación digital y bibliotecas: *un nuevo escenario*,
Revista general de información y documentación, ISSN 1132-1873, Vol. 18, Nº 1.

Páez, Y 2006, Esquema arquitectural de software para educación a distancia.
Trabajo de Grado, Universidad centro-Occidental Lisandro Alvarado, Barquisimeto

RDF Core Working Group 2004, Resource Description Framework (RDF).
URL: <http://www.w3.org/RDF/> WC3, 2004

Romero, M y otros 2003, Nuevas tecnologías y Terapia ocupacional.
Elsevier España

Sabino, C. 2004, El proceso de Investigación.
El Cid Edit. Caracas.

Saffady, W y otro 1986, Informática documental para bibliotecas.
Ediciones Díaz de Santos.

Sebastián M. y otros 2004, La información en la posmodernidad.
Editorial Ramón Areces.

Schenone M. 2004. Diseño de una Metodología Ágil de Desarrollo de Software.
Facultad de Ingeniería de Universidad de Buenos Aires, Argentina.

- Stair, R 2000, Principios de sistemas de información: Enfoque administrativo. Cengage Learning Editores.
- Steffen Staab, Heiner Stuckenschmidt 2006, SemanticWeb and Peer-to-Peer Decentralized Management and Exchange of Knowledge and Information. University of Mannheim Institute for Practical Computer Science
- Tamayo, A. 1998. Metodología de la Investigación. Mc Graw-Hill Interamericana, México.
- Tanenbaum, A 2003, Redes de computadoras. Pearson Educación.
- Técnica y Empleo Aragón, S.L. 2007, Temario Oposiciones Auxiliar de Bibliotecas de la Universidad de Zaragoza, EDITORIAL CEP, Zaragoza.
- The library of congress 200), Metadata Encoding & Transmission Standard. URL: http://www.loc.gov/standards/mets/METSOverview_spa.html Noviembre, 2008
- Thompson, A. y Otros. 2004. Administración Estratégica. McGraw-Hill Interamericana. México.
- Torres, G. y otro 2005 La biblioteca digital, UNAM.
- UCLA 2002. Manual para la Presentación del Trabajo Conducente al Grado Académico de: Especialización, Maestría, Doctorado. Universidad centro-occidental Lisandro Alvarado. Barquisimeto
- Valera, R 2005, Desarrollo de una herramienta para la gestión de documentos y contenidos en el proceso enseñanza aprendizaje dentro del decanato de ciencias y tecnología de la UCLA. Trabajo de grado, Universidad centro-occidental Lisandro Alvarado. Barquisimeto
- Valhondo, D. 2003, Gestión del conocimiento, del mito a la realidad. Díaz de Santos
- Vieira, M y otro 2006, Impacto tecnológico y arquitectura en biblioteca. Alfagrama Ediciones.

ANEXOS

ANEXO 1

Cuestionario N° 1 Entrevista semiestructurada
UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”
DECANATO DE CIENCIAS Y TECNOLOGIA
POSTGRADO EN CIENCIAS DE LA COMPUTACION
MENCION INGENIERÍA DEL SOFTWARE

DESARROLLO DE UNA DE UNA HERRAMIENTA DE INTEGRACIÓN Y
BÚSQUEDA EN SISTEMAS DE GESTIÓN DE CONTENIDOS EN AMBIENTES
DISTRIBUIDOS PARA BIBLIOTECAS DIGITALES
DE LA UCLA

ENTREVISTA ESTRUCTURADA

Nombre:	
Cargo:	

1. ¿Como está Estructurada la biblioteca?
2. ¿Existe algún estándar para catalogar documentos?
3. ¿Que Atributos presentan los catálogos?
4. ¿Qué diferencia presenta la catalogación de acuerdo al tipo de documento?
5. ¿La biblioteca tiene algún acuerdo para publicar libros o documentos digitales?
6. ¿Existen documentos digitales con restricción de acceso, de existir que casos y cómo?
7. ¿Digitalizan material?
8. ¿Que técnicas utilizan para la digitalización del material?
9. ¿Que tipos de formatos aplican al material digitalizado y porque?
10. ¿Existen restricciones de tamaño o peso para la digitalización de material?
11. ¿Como manera se publica, el material digitalizado?
12. ¿De qué modo se accede al material digitalizados?

13. ¿Existen perfiles de acceso para el material?
14. ¿De existir perfiles, cuáles son sus tipos?
15. ¿Cuál es la manera de acceder al material digital?
16. ¿Se cuenta con un servicio de búsqueda de documentos?
17. ¿Para la búsqueda se usan los criterios de catalogación?
18. ¿Se realizan búsquedas por el contenido del material?
19. ¿De usarse los atributos del catálogo para búsqueda, existen atributos exclusivos para el control del documento?
20. ¿Se cuenta con un servicio de búsqueda integrada con otros decanatos o bibliotecas?
21. ¿Con que Estructura tecnológica cuentan ¿
22. ¿Existe la posibilidad de establecer alianzas con otras bibliotecas o redes de bibliotecas al corto o mediano plazo?
23. ¿Existe alguna limitante para la integración con otras bibliotecas fuera de la UCLA?

ANEXO 2
UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”
DECANATO DE CIENCIAS Y TECNOLOGIA
POSTGRADO EN CIENCIAS DE LA COMPUTACION
MENCION INGENIERÍA DEL SOFTWARE

DESARROLLO DE UNA DE UNA HERRAMIENTA DE INTEGRACIÓN Y
BÚSQUEDA EN SISTEMAS DE GESTIÓN DE CONTENIDOS EN AMBIENTES
DISTRIBUIDOS PARA BIBLIOTECAS DIGITALES
DE LA UCLA

GUÍA DE OBSERVACIÓN

DECANATO.

TIPO DE OBSERVACION:

PROCESO O ACTIVIDAD OBSERVADA:

FECHA.

ASPECTO OBSERVADO

- Modo de acceso al material
- Datos mostrados del catálogo
- Modo de publicación del material digital.
- Uso de catálogo para búsqueda
- Métodos para la búsqueda del material
- Tipos de búsqueda, por alcance y por atributos
- Material digital disponible
- Alternativas no propias para acceso al material

ANEXO 3

Carta de Solicitud de Validación a los Expertos

UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”

DECANATO DE CIENCIAS Y TECNOLOGÍA

Barquisimeto, Noviembre de 2009

Ciudadano (a): _____

Reciba un cordial saludo en mi nombre. Sirva la presente para solicitar ante Ud. su colaboración en calidad de experto para determinar la validez de contenido de los cuestionarios con los cuales se pretende recabar información necesaria para el Trabajo de Grado titulado “Desarrollo de una herramienta para la integración y búsqueda en sistemas de gestión de contenidos para bibliotecas digitales”, el cual será aplicado a personal de las bibliotecas de los diversos decanatos de la UCLA.

Para tal propósito se anexa a la presente: 1) Matriz de peracionalización de las Variables; 2) Formatos de Validación con sus respectivas instrucciones; 3) Cuestionarios que se aplicarán.

Su opinión al respecto representa un gran aval para esta investigación, dada la excelente labor docente, de investigación y extensión que Ud. ha realizado en pro de la formación profesional en la Universidad.

Sin otro particular al cual hacer referencia y agradeciendo de antemano su colaboración, quedo de Ud.

Atentamente,

Ing. Reynaldo Mantilla Acevedo

C.I.: 22.328.626

ANEXO 4

Formato de Validación de los Instrumentos

UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”
DECANATO DE CIENCIAS Y TECNOLOGIA
POSTGRADO EN CIENCIAS DE LA COMPUTACION
MENCION INGENIERÍA DEL SOFTWARE

DESARROLLO DE UNA HERRAMIENTA DE UNA HERRAMIENTA DE
INTEGRACIÓN Y BÚSQUEDA EN SISTEMAS DE GESTIÓN DE
CONTENIDOS EN AMBIENTES DISTRIBUIDOS PARA BIBLIOTECAS
DIGITALES

Formato de Validación de los Instrumentos

Ciudadano (a): _____

Para efectos de la evaluación correspondiente a los ítems planteados se determinará la validez de cada instrumento en los siguientes términos:

Se tomarán en cuenta los siguientes aspectos: a) Pertinencia: Es la correspondencia del ítem con el aspecto; b) Claridad: se refiere a la redacción precisa y sencilla del ítems; y c) Congruencia: entendida como la lógica interna del ítem.

Se le agradece seleccionar una de las 2 posibles opciones (Si/No) para cada ítems con el objetivo de señalar el grado de pertinencia, claridad y congruencia de los ítems.

Cuestionario

Item	Pertinencia		Claridad		Congruencia		Observación
	Si	No	Si	No	Si	No	
1							
2							
3							
4							

5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							

ANEXO 5

Casos de Uso

Descripciones textuales de los casos de uso de la herramienta para la integración y búsqueda en sistemas de gestión de contenidos para bibliotecas digitales

Caso de Uso ‘Determinar Estructura de Indexado en Documento’.

Nombre	Determinar Estructura de Indexado en Documento
Descripción:	Permite determinar las características o atributos, definidas en el repositorio para cada documento.
Precondiciones:	Las características deben estar almacenadas en un XML de configuración.
Flujo Normal	
Obelisco	Sistema <ol style="list-style-type: none">1. comprueba la existencia del archivo de configuración.2. Se extrae la información de dicho archivo.3. Se crean los atributos del documento a ser leídos desde la interfaz.4. Se cargan los valores de comportamiento del repositorio como: uso de contenido, cantidad de resultados, directorio de indexación.
Flujo Alternativo	
	<ol style="list-style-type: none">2. de no existir una excepción de configuración con el mensaje “Problemas con archivo de Configuración”5. El sistema comprueba la validez de

	los datos, si los datos no son correctos dispara una excepción de configuración con el Mensaje “Problemas con datos de Configuración”
--	---

Fuente: El autor de la investigación.

Caso de Uso ‘Configurar repositorio’.

Nombre	Configurar repositorio
Descripción: Permite almacenar en un archivo la configuración y determinar las características o atributos, definidas en este.	
Actores: Sistema Obelisco	
Precondiciones: Definir la ruta en la que se generara y revisara el archivo de configuración en el IndexBeans	
Flujo Normal	
Obelisco	Sistema
3. setea los valores de configuración en la plantilla.	<ol style="list-style-type: none"> 1. Buscar la ruta del archivo de configuración. 2. verifica que la plantilla del documento, este cargada en memoria. 3. 4. Se almacena la información del documento y del repositorio en el archivo de configuración. 5. retorna la configuración del

	documento por medio del caso de uso anterior.
Flujo Alternativo	
	<ol style="list-style-type: none"> 3. si la plantilla no existe, dispara una excepción de configuración “Plantilla de documento no definida para ser almacenada”. 4. 5. Si el documento de configuración no puede ser almacenado se dispara una excepción de configuración, con el mensaje “Problema al generar Archivo de configuración”.

Fuente: El autor de la investigación.

Caso de Uso ‘Indexar atributos del documento’.

Nombre	Indexar atributos del documento
Descripción: Permite incluir en el documento indexado cada una de los atributos del documento.	
Precondiciones: <ul style="list-style-type: none"> • Que la configuración del documento este definida. • Que los valores de cada atributo estén seteados 	
Flujo Normal	
Obelisco	Sistema <ol style="list-style-type: none"> 1. verifica la configuración del documento. 2. Recorre el map de atributos. 3. Verifica los valores de los atributos. 4. Agrega el campo al documento de indexado.

Flujo Alternativo	
•	2. De no existir la configuración del documento, genera una excepción de Indexado con el mensaje “Configuración de documento no definida”. 3. 4. De detectar algún problema en el mismo este será seteado una cadena de longitud 0.

Fuente: El autor de la investigación.

Caso de Uso ‘Determinar tipo de documento’.

Nombre	Determinar tipo de documento
Descripción: Determina el tipo de documento digital a indexar para aplicar la estrategia de extracción del documento a indexar.	
Precondiciones: Que la ruta del documento o el tipo estén seteados	
Flujo Normal	
Obelisco	Sistema 1. Se verifica el tipo de documento 2. Se verifica que el tipo este definido como estrategia valida 3. Se retorna valides de indexar.
Flujo Alternativo	
•	1. 2. 3. Si el tipo de documento, no cuenta con una estrategia definida valida,

	genera una excepción de Contenido de documento con el mensaje “No existe algoritmo definido, para un documento de este tipo”.
--	---

Fuente: El autor de la investigación.

Caso de Uso ‘Indexar contenido del documento’.

Nombre	Indexar contenido del documento
Descripción: Extraer contenido del documento e indexarlo, ruta de ubicación y tipo.	
Precondiciones: <ul style="list-style-type: none"> • El documento debe existir en la ruta indicada. • El documento digital debe contener texto, los documentos digitales de imágenes no podrán ser indexados por contenido, así como los protegidos. En la configuración del documento se indicara que indexaba contenido.	
Flujo Normal	
Obelisco	Sistema <ol style="list-style-type: none"> 1. Abrir el documento en la ruta especificada. 2. Extraer contenidos del documento con estrategia para su tipo. 3. Agregar contenido a documento de indexado. 4. Agregar ruta y tipo al documento de indexado.
Flujo Alternativo	
	<ol style="list-style-type: none"> 1. 2. Problema al abrir documento, dispara una excepción de acuerdo al problema presentado, las

	<p>excepciones serian:</p> <ul style="list-style-type: none"> i. Ruta no valida con el Mensaje “Ruta no valida, Archivo inexistente”. ii. Permisologia del documento, "Documento no presenta los permisos para realizar esta operación". <p>3. Problema al extraer contenido del documento, generara una Excepción de Tipo no definido “No existe algoritmo definido, para un documento de este tipo”</p>
--	---

Fuente: El autor de la investigación.

Caso de Uso ‘Indexar Documento’.

Nombre	Indexar Documento	
Descripción:		
Indexa Atributos y contenido de un documento indicado de acuerdo a la configuración definida.		
Actores:		
Sistema Obelisco		
Precondiciones:		
Debe existir documento plantilla en la configuración.		
Flujo Normal		
Obelisco		Sistema
1.El sistema Obelisco transfiere un documento del tipo configurado con los valores del mismo.		1. 2. verifica los datos recibidos.

	3. Indexa los datos del documento <ul style="list-style-type: none"> • Atributos • Contenido
Flujo Alternativo	
•	1. 2. 3. Errores en los datos, genera una excepción de tipo Documento con el mensaje de “Los datos emitidos, presentan errores por favor verifique”

Fuente: El autor de la investigación.

Caso de Uso ‘Buscar Documento’.

Nombre	Buscar Documento	
Descripción:	Permite generar búsqueda de contenido en base a los atributos definidos en el repositorio de indexación local.	
Precondiciones:	<ul style="list-style-type: none"> • Existencia de documentos indexados. • Indicar datos a buscar 	
Actores	Sistema Obelisco	
Flujo Normal		
	Obelisco	Sistema
	1. El sistema obelisco envía atributo o valor a buscar y dato a coincidir.	1. 2. ubica el repositorio gracias a la información en configuración 3. realiza la búsqueda en repositorio. 4. Trasforman las coincidencias de

	<p>tipo indexado, en una lista de documentos.</p> <p>5. Retorna la lista de documentos con coincidencia.</p>
Flujo Alternativo	
•	<ol style="list-style-type: none"> 1. 2. 3. El sistema detecta Problemas en el repositorio, genera una excepción de tipo Indexar “El repositorio presenta problemas”

Fuente: El autor de la investigación.