

Universidad de las Ciencias Informáticas
Facultad 1



**Solución para las Búsquedas Avanzadas
en el Repositorio Institucional.**

Trabajo de diploma para optar por el título de Ingeniero
en Ciencias Informáticas.

Autor:

Yordanis Medero Castellanos.

Tutores:

MSc. Adrian Cid Almaguer.

Ing. Julio Jesús García Coste.

Ciudad de la Habana, Junio 2013

Declaración de Autoría

Declaro que soy el único autor del trabajo “**Solución para las Búsquedas Avanzadas en el Repositorio Institucional.**” y autorizo a la Facultad 1 de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Autor

Yordanis Medero Castellanos

Tutores

MSc. Adrian Cid Almaguer

Ing. Julio Jesús García Coste

Dedicatoria

Este Trabajo de Diploma está dedicado al único Rey y Salvador de mi vida, al único que ha permitido que esta Tesis llegue hasta este lugar, al que me ha guiado y capacitado en medio de las dificultades, al que siempre me ha amado aunque aún no lo sabía, a mi AMIGO, mi PASTOR, mi FORTALEZA nada más y nada menos que ESPÍRITU SANTO DE DIOS.

Agradecimientos

Quiero darle gracias a mi PASTOR por su apoyo incondicional, por estar presente en todo momento.

Gracias a mis padres por su apoyo todos estos años de mi vida y por la educación que me fue dada.

A mis hermanos en Cristo por ser todos tan especiales, llenando mi corazón de alegría y amor en todo momento.

Gracias a los tutores por todo este tiempo que me han estado ayudando y por la preocupación que han tenido en cuando a esta tesis.

También a mi tribunal de tesis que sin duda hay ayudado a mejorar el presente trabajo.

Y de manera muy especial a mi amada esposa, quien me ha ayudado y animado en el transcurso de todo este tiempo, y siempre ha estado ahí cuando la he necesitado.

Resumen

En la Universidad de las Ciencias Informáticas se encuentra implementado un Repositorio Institucional que además de contener información científica permite recuperarla para su posterior utilización por la comunidad universitaria, para ello cuenta con un módulo de búsqueda avanzada con variados criterios de búsqueda para la obtención de documentos. El presente trabajo se enmarca precisamente en ese módulo de búsqueda con el objetivo de mejorar dicho sistema de búsqueda. Para ello se realiza un estudio de varios sistemas que implementan funcionalidades de recuperación de la información, el cual permite analizar los principales criterios de búsqueda que ellos gestionan, para el desarrollo del sistema propuesto en el presente trabajo. Además de esto, se realiza la implementación de funcionalidades que permiten la gestión de estos criterios, brindando la posibilidad de seleccionar los que estarán disponibles en la búsqueda avanzada del Repositorio Institucional de la universidad.

Palabras clave: *búsqueda avanzada, criterios de búsqueda, recuperación de información, repositorio institucional.*

Contenido

Introducción	1
Fundamentación teórica 1	5
1.1 Introducción	5
1.2 Conceptos Fundamentales.....	5
1.2.1 Repositorio Institucional.....	5
1.2.2 Búsqueda y Recuperación de Información.....	7
1.3 Sistemas de Gestión de Repositorios Institucionales	13
1.4 Repositorio de la Universidad de las Ciencias Informáticas	14
1.5 Análisis de herramientas que gestionan búsquedas avanzadas	17
1.5.1 Repositorio E-Archivo de la Universidad Carlos III de Madrid.....	18
1.5.2 Repositorio ScholarsArchive @ OSU	18
1.5.2 Repositorio DSpace @ Cambridge.....	19
1.5.2 Sistemas web que gestionan búsquedas avanzadas y sus criterios.....	19
1.5.3 Selección de criterios	21
1.6 Metodologías y Tecnologías utilizadas.....	22
1.6.1 Proceso de desarrollo de software	22
1.6.2 Lenguaje de Programación	24
1.6.3 Gestor de Base de Datos	25
1.6.4 Herramienta CASE para el modelado	25
1.6.5 Servidor de Aplicación	26
Características del Sistema 2	28
2.1 Introducción	28
2.2 Propuesta de solución.....	28
2.3 Modelo de Dominio.....	30
2.3.1 Clases del modelo de dominio.....	30
2.4 Especificación de Requisitos.....	31
2.4.1 Requisitos Funcionales.....	31
2.4.2 Requisitos no funcionales.....	32
2.5 Modelo de casos de uso del sistema.....	33
2.5.1 Actores del sistema	33

2.5.2 Definición de casos de uso del sistema	33
2.5.4 Descripción textual de los casos de uso del sistema.....	34
2.6 Modelo de análisis	36
2.6.1 Diagramas de clases del análisis	37
2.7 Modelo de Diseño	38
2.7.1 Diagrama de colaboración	38
2.7.2 Diagramas de clases del diseño	39
2.8 Modelo de datos.....	41
2.10 Arquitectura de la aplicación. Arquitectura 3 Capas	42
2.11 Patrones GRASP	43
Implementación y Prueba 3	45
3.1 Introducción.....	45
3.2 Modelo de Implementación	45
3.2.1 Diagrama de despliegue.....	45
3.2.2 Diagrama de componentes	46
3.3 Métodos de Validación	47
3.3.1 Pruebas de Funcionalidad.....	47
3.4 Pruebas de Caja Negra.....	48
3.4.1 Casos de Prueba	48
3.5 Clasificación de las no conformidades	55
3.5.1 Resultados de las pruebas de caja negra.....	56
Conclusiones Generales.....	58
Recomendaciones	59
Referencias Bibliográficas	60

Introducción

En décadas anteriores no se contaba con medios informáticos para la gestión de la información, de modo que todo este proceso de gestión era en extremo limitado y como consecuencia la expansión de la información fue restringida. El surgimiento de las computadoras constituyó un gran avance para la gestión de la información reduciendo al mínimo de tiempo operaciones complejas que su ejecución demoraban horas y hasta días, y esto a su vez facilitaba en gran medida el crecimiento de la información. En la actualidad la informática ha llegado a dominar el mundo, al extremo de crear casi una total dependencia de esta en procesos sociales, económicos, políticos, educativos por solo mencionar algunos. Es notable el crecimiento de la información digital que esto ha traído consigo como imágenes, vídeos, audio, todos ellos en formatos diferentes. Las instituciones no han sido una excepción de esto, sino que la mayor cantidad de los procesos en ellas están totalmente informatizados dando lugar a que la información de las mismas sea digitalizada y difícil de encontrar. Debido a esto surgen los Repositorios: Se denomina repositorio a los archivos o bases de datos que almacenan recursos digitales (texto, imagen y sonido). En general los recursos son depositados por el propio autor, proceso denominado autoarchivo. Pueden ser ponencias de eventos, conferencias, informes de investigación, presentaciones a seminarios, tesis, textos de enseñanza y otros trabajos académicos. (De Volder, 2012)

De ahí que los Repositorios institucionales: Reúnen la producción científica y/o académica de los miembros de una o varias instituciones resultado de su actividad docente e investigadora, almacenando, preservando, divulgando y dando acceso abierto a los recursos depositados en ellos. (De Volder, 2012)

Muchos países han optado por los Repositorios Institucionales y Cuba no se ha quedado atrás en cuanto a ello sino que también ha implementado esta tecnología. Específicamente la Universidad de las Ciencias Informáticas cuenta con un Repositorio donde se almacenan sus documentos intelectuales, estos se encuentran en una clasificación dada por comunidades y colecciones según el tema en que se desempeñe el documento. Los documentos forman parte de su producción científica como por ejemplo el Evento Juvenil Martiano, Fórum Ciencia y Técnica, Mi web por Cuba que son eventos desarrollados en la universidad y de una importancia relevante. Las principales ventajas del Repositorio están constituidas por la centralización y disponibilidad de la información. (Alvarez

Fernández, 2010)

Sin embargo a pesar de las ventajas que ha traído consigo para la universidad, este cuenta en su módulo de búsqueda avanzada con los mismos criterios de búsqueda que implementa por defecto la herramienta con que se implantó el Repositorio Institucional dos años atrás, esto trae como consecuencia que en dicho módulo los criterios vallen quedando atrás a medida que avanza la tecnología y por tanto en desuso por parte de la comunidad universitaria. Esto trae como consecuencia que las búsquedas sean insatisfactorias ya los criterios que describen los documentos son de relevante importancia en el proceso de recuperación de la información del módulo, contando a su vez los gastos de las cuotas para el acceso a internet en busca de información que probablemente se encuentre disponible en el Repositorio Institucional de la universidad y que almacena buena cantidad de información.

Además el módulo de búsqueda avanzada no es configurable ya que no cuenta con funcionalidades que permitan al administrador gestionar los criterios de búsqueda, de modo que para poder agregar o eliminar un criterio se necesita acudir a los especialistas encargados del mantenimiento del sistema para realizar dichos cambios, quienes deben instalar todas las herramientas necesarias para la modificación del módulo, también se debe trabajar directamente en el código fuente de la aplicación pocas veces modificado por los especialistas. Una vez modificado el código fuente es necesaria la recompilación del sistema para guardar los nuevos cambios realizados. Al realizar todas estas gestiones existe un consumo considerable de tiempo para poderlas llevar a cabo dificultando así la realización de cambios en cuanto a los criterios de búsquedas.

De lo antes expuesto surge el siguiente **problema de investigación**: ¿Cómo seleccionar y gestionar los criterios de búsquedas del Repositorio Institucional de la Universidad de las Ciencias Informáticas?

Se define como **objeto de estudio** las técnicas y procedimientos para la recuperación de la información.

Se ha definido el siguiente **objetivo general**: Modificar el módulo de búsqueda avanzada para facilitar las búsquedas y gestionar los criterios de búsqueda que se encuentran en el Repositorio Institucional de la Universidad de las Ciencias Informáticas a través de nuevos criterios descriptivos que permitan la recuperación de información.

Definiendo como **campo de acción** el mecanismo para la búsqueda y recuperación de la información en el Repositorio Institucional de la Universidad de las Ciencias Informáticas.

Para dar cumplimiento a los objetivos planteados, se deben realizar las siguientes **tareas de investigación**:

- Fundamentación Teórica de los principales conceptos y definiciones asociadas al área de la Bibliotecología relacionada con los Repositorios Institucionales para la captura de datos.
- Fundamentación Teórica de las tecnologías utilizadas para el desarrollo del módulo (definidas por el equipo de especialistas de la línea) para caracterizar estas tecnologías.
- Realización del Estado del Arte de los principales Repositorios Institucionales que automaticen procesos de búsquedas avanzadas para la selección de las mejores prácticas de búsqueda y su implantación en el Repositorio Institucional de la Universidad de las Ciencias Informáticas.
- Levantamiento de requisitos necesarios en la realización de búsquedas avanzadas del Repositorio Institucional de la Universidad de las Ciencias Informáticas para capturar las definiciones detalladas de las funcionalidades necesarias.
- Análisis y diseño de las funcionalidades identificadas relativas a la realización de búsquedas avanzadas en el Repositorio Institucional de la Universidad de las Ciencias Informáticas para la definición de la arquitectura, componentes e interfaces que permitan su implementación.
- Implementación de las funcionalidades relativas a la realización de búsquedas avanzadas en el Repositorio Institucional de la Universidad de las Ciencias Informáticas para la ejecución del diseño de las funcionalidades descritas que brindan la solución del problema.
- Realización de pruebas de Caja Negra a las funcionalidades implementadas en el módulo búsquedas avanzadas del Repositorio Institucional de la Universidad de las Ciencias Informáticas para garantizar la calidad del producto.

Obteniendo el **resultado esperado**:

- ◆ Obtención del estado del arte de la automatización de búsquedas avanzadas en Repositorios Institucionales.
- ◆ Obtención del análisis y diseño de funcionalidades que permitan realizar búsquedas avanzadas en el Repositorio Institucional de la Universidad de las Ciencias Informáticas.

- ♦ Obtención de funcionalidades que permitan realizar búsquedas de los recursos que se encuentran en el Repositorio Institucional de la Universidad de las Ciencias Informáticas a través de criterios descriptivos que permitan su rápida recuperación.

Los métodos investigativos empleados en la presente investigación son el método teórico y el método empírico.

Dentro del método teórico se utilizó el **análisis histórico-lógico** para realizar un estudio de la búsqueda y recuperación de la información en diferentes Repositorios y sistemas que hagan uso de ella y el **método analítico-sintético** al identificar y analizar los principales conceptos y definiciones relacionados con los Repositorios Institucionales y la recuperación de la información, así como el análisis de toda la documentación relacionada con el tema. También es empleado el método teórico **Modelación** al confeccionar diagramas que permitan modelar aspectos que faciliten la implementación.

Dentro del método empírico se utilizó la **Observación**, este permitió observar y obtener información sobre cómo se gestionan los módulos de búsquedas avanzadas en otras herramientas existentes que ayuden a resolver la problemática planteada.

El presente trabajo de diploma está estructurado de la siguiente manera:

Capítulo 1: Fundamentación teórica: se documentan los principales conceptos relacionados con los Repositorios, búsquedas avanzadas, así como la descripción de las tecnologías a emplear para la solución del problema.

Capítulo 2: Análisis y Diseño de la solución: se detallan los elementos para comenzar el proceso de desarrollo de la solución propuesta y su diseño para la implementación del modulo de búsqueda avanzada.

Capítulo 3: Implementación y Prueba: aborda todo el proceso de prueba de la solución propuesta.

Fundamentación teórica 1

1.1 Introducción

En el presente capítulo se lleva a cabo la fundamentación teórica del trabajo, teniendo como principal objetivo investigar, estudiar y analizar los principales elementos relacionados con las herramientas, metodologías y tecnologías a utilizar para la creación de una posible solución al problema. Se tratan las principales definiciones de búsqueda y repositorios. Además se realiza un estudio de las principales aplicaciones que implementen módulos de búsquedas avanzadas para la implementación de un módulo para el Repositorio Institucional de la Universidad de las Ciencias Informáticas.

1.2 Conceptos Fundamentales

1.2.1 Repositorio Institucional

Muchas son las definiciones que brindan varios autores acerca de lo que es un Repositorio o Repositorio digital algunos ejemplos de ellos es Texier quien dice:

Los repositorios, también conocidos como repositorios digitales, están constituidos por un conjunto de archivos digitales en representación de productos científicos y académicos que pueden ser accedidos por los usuarios (texier, et al., 2012)

Álvarez por su parte dice:

Un repositorio digital es un sistema completo para la preservación en el tiempo del contenido digital y/o documentación generada por una organización. (Álvarez González, 2011)

Analizando estas definiciones se concluye que de forma general concuerdan en que es un depósito, archivo o base de datos en los cuales se almacena información digital que puede ser académica y/o científica.

Existen dos clasificaciones fundamentales de los Repositorios como son:

- ♦ **Temáticos:** Fueron los primeros repositorios en aparecer. Se crean en torno a una disciplina; imposible que sean completos. No hay forma de obligar a los autores a remitir sus trabajos.
- ♦ **Institucionales:** Recogen la producción de una institución y es la forma más extendida; actualmente, se centran en una organización (universidad, departamento, instituto, sociedades científicas). Es posible definir políticas para que los miembros añadan contenidos. En esta clasificación, se incluyen los repositorios de tesis doctorales. (Flores Cuesta, et al., 2007)

Principales características

Algunas de las características de los repositorios y del los Repositorios Institucionales es que contienen objetos digitales en diferentes formatos los cuales pueden variar en pdf, doc, jpg por mencionar algunos. Además contienen metadatos que permiten describir cada uno de los diferentes datos almacenados en los mismos y que son de mucha utilidad para su posterior recuperación, esto consiste en describir los registros almacenados en el Repositorio a través de varios aspectos (como pueden ser autor, título, tema, lengua) determinados por estándares de descripción de metadato como el estándar Dublin Core que es utilizado por el software Dspace. Asegura la identificación persistente del objeto mediante un identificador único persistente, este proceso se realiza en el momento de almacenar un registro determinado, para su posterior ubicación en el sistema, se le asigna dicho identificador que es único para cada elemento y que funciona como el nombre del registro dentro del Repositorio. Permiten adicionar, modificar y eliminar objetos así como el archivo de los mismos y su preservación. También proporcionan un acceso fácil a la información almacenada en los mismos lo cual no significa que no sea controlado, además ofrecen los sistemas adecuados de seguridad para los objetos y sus metadatos.

Importancia de los Repositorios Institucionales

Los Repositorios Institucionales tienen gran importancia en nuestros días y es una tecnología empleada ya por muchos países, tal es el caso de España, Reino Unido, EEUU, Alemania, Brasil, Japón por solo citar algunos, que son los que más repositorios tienen inscrito en el Registro de Repositorios de Acceso Abierto (ROAR) por sus siglas en inglés. (ROAR, 2013).

Según Aroca el establecimiento de un Repositorio Institucional permite a la institución publicar su investigación y sus programas educativos ofreciendo acceso a las obras de su personal. La calidad de esos resultados académicos constituye una propaganda eficaz para la institución o para la investigación financiada con fondos públicos. Las organizaciones académicas de todo el mundo obtienen ventajas de un acceso más fácil y rápido a las investigaciones desarrolladas en otras instituciones. La inclusión de un trabajo científico en un depósito institucional aumenta la proyección del perfil de un autor a escala mundial, y aumenta el impacto de la investigación que desarrolla. La inclusión en un Repositorio Institucional también puede aligerar la carga administrativa que conllevan los informes sobre publicaciones para los ejercicios de revisión y evaluación de la investigación. Para la institución que lo genera, este tipo de repositorio puede incrementar la visibilidad y el prestigio de la institución. Para los investigadores, aumenta la visibilidad de sus trabajos, ampliando la difusión y el uso de los mismos. (Aroca Cámara, y otros, 2009)

1.2.2 Búsqueda y Recuperación de Información

La Recuperación de Información (RI) no es un área nueva, sino que se viene desarrollando desde finales de la década de 1950. Sin embargo, en la actualidad adquiere un rol más importante debido al valor que tiene la información. Se puede plantear que disponer o no de la información justa en tiempo y forma puede resultar en el éxito o fracaso de una operación. (H. Tolosa, y otros, 2008)

Varios autores emiten su criterio referente a la RI como Baeza-Yates quien dijo que: la Recuperación de Información trata con la representación, el almacenamiento, la organización y el acceso a ítems de información (Baeza-Yates, et al., 1999)

Anterior a este Salton brindó una definición más amplia la cual plantea que la RI es un campo relacionado con la estructura, análisis, organización, almacenamiento, búsqueda y recuperación de información. (Salton, et al., 1983)

Teniendo en cuenta los conceptos anteriormente citados se concluye que la RI consiste en encontrar y establecer un orden por relevancia de los documentos que satisfacen la necesidad de información de un usuario expresada en forma de consulta.

Problema de la RI

Según Baeza-Yates el problema de la RI se divide en dos partes fundamentales: la computacional y la humana (Baeza-Yates, y otros, 1999). La humana no es más que las necesidades del usuario y la computacional es todo lo que el sistema debe de hacer para dar respuesta a la consulta del usuario (dígase estructuras de los datos y algoritmos eficientes) con una calidad aceptable.



Figura 1: Problema de la RI.

En la figura anterior se evidencia de forma clara el problema de la RI: el corpus corresponde al conjunto de documentos que existen en el Sistema de Recuperación de la Información (SRI) y que luego serán accedidos por los usuarios según sus necesidades de información expresadas en consultas. El sistema frente a la petición del usuario determina un conjunto de documentos relevantes que son los que dan respuesta a la consulta del usuario y se los presenta generalmente en una lista según su relevancia, también existe un conjunto de documentos no relevantes que no responden a las necesidades del usuario y por tanto no son presentados como respuesta.

Tolosa plantea que la respuesta “ideal” de un SRI está formada solamente por documentos relevantes a la consulta, pero – en la práctica – esta no es aún alcanzable. Esto se debe a que – entre otros motivos – existe el problema de compatibilizar la expresión de la necesidad de información y el lenguaje y de los documentos. Además, hay una carga de subjetividad subyacente y depende de los usuarios. Entonces, el SRI recupera la mayor cantidad posible de documentos relevantes, minimizando la cantidad de documentos no relevantes (ruido) en la respuesta. En términos de eficiencia, se plantea la idea de

precisión de la respuesta, es decir, cuando más documentos relevantes contengan el conjunto solución (para una consulta dada), más preciso será. (Tolosa, et al., 2007)

Arquitectura de un SRI

Un SRI debe de tener en cuenta la arquitectura expresada en la figura 2, que responden a las siguientes tareas:

- ◆ Representación lógica de los documentos.
- ◆ Representación de la necesidad de información del usuario en forma de consulta.
- ◆ Evaluación de los documentos respecto de una consulta para establecer la relevancia de cada uno.
- ◆ Ranking de los documentos considerados relevantes para formar el “conjunto solución” o respuesta.
- ◆ Presentación de la respuesta al usuario.
- ◆ Retroalimentación o refinamiento de las consultas (para aumentar la calidad de la respuesta) (Tolosa, et al., 2007)

En la figura 2 se detalla la arquitectura básica de los SRI.

La representación lógica de los documentos está dada por términos, frases u otros que permitan caracterizar los documentos almacenados en el sistema (dichos documentos almacenados se denominan corpus) y sin la cual no se pueden realizar operaciones sobre el corpus. Luego es necesario realizar la indexación, la cual consiste en crear un índice que permita la localización dentro del sistema de los documentos.

El algoritmo de búsqueda tiene como entrada la consulta realizada por el usuario, partiendo de esto verifica en el índice los que pueden satisfacer la necesidad del usuario obteniendo una lista de los documentos relevantes los cuales serán analizados por el algoritmo de ranking, quien determina el nivel de relevancia de estos y los presenta al usuario en una lista de orden decreciente. Finalmente la interfaz se emplea para presentar los resultados al usuario además de capturar la consulta del usuario en un lenguaje predeterminado implementado en el SRI.

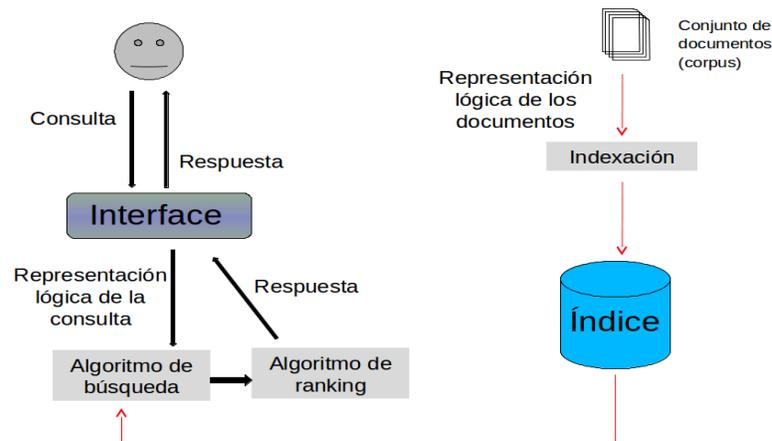


Figura 2: Arquitectura básica de los SRI

Principales modelos para la recuperación de información

Existen varias formas de resolver el problema de la RI las cuales son de vital importancia a la hora de desarrollar un SRI, estas formas son denominadas modelos de RI los cuales constituyen modelos matemáticos que permiten determinar los documentos que representan la respuesta a la consulta del usuario así como su relevancia y ranking. De manera básica existe una clasificación de documentos según varios autores que consiste en documentos estructurados y documentos no estructurados. Según Tolosa los primeros son aquellos en los que se pueden reconocer elementos estructurales con una semántica bien definida, mientras que los segundos corresponden a texto libre, sin formato. La diferencia fundamental de un SRI que procese documentos estructurados se encuentra en que puede extraer información adicional al contenido textual, la cual utiliza en la etapa de recuperación para facilitar la tarea y aumentar las prestaciones. (H. Tolosa, y otros, 2008)

Los modelos para los documentos no estructurados son: booleano, vectorial, probabilístico y para los no estructurados los modelos basados en álgebra de regiones. A continuación se explican los modelos para documentos no estructurados ya que estos son los de interés en el presente trabajo.

Modelo Booleano

En el modelo booleano la representación de la colección de documentos se realiza sobre una matriz binaria documento–término, donde los términos han sido extraídos manualmente o automáticamente de los documentos y representan el contenido de los mismos.

Las consultas se arman con términos vinculados por operadores lógicos (AND, OR, NOT) y los resultados son referencias a documentos donde cuya representación satisface las restricciones lógicas de la expresión de búsqueda. En el modelo original no hay ranking de relevancia sobre el conjunto de respuestas a una consulta, todos los documentos poseen la misma relevancia. (H. Tolosa, y otros, 2008)

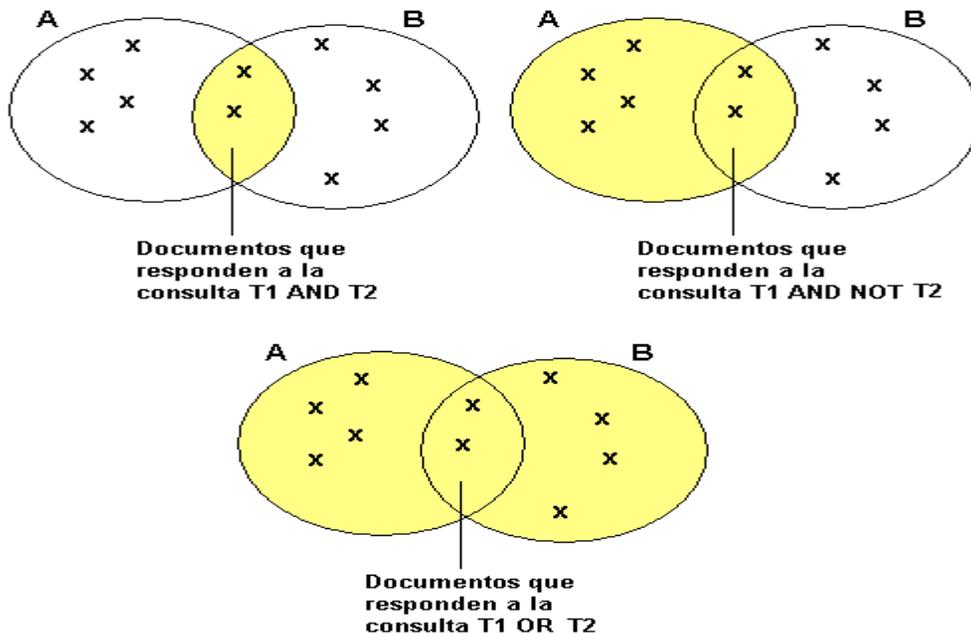


Figura 3: Operaciones realizadas con el método booleano

En la anterior figura el conjunto A representan los documentos que responden al término T1 y el conjunto B los que responden al termino T2, donde T1 y T2 son términos (palabras) que conforman la consulta del usuario.

Modelo Probabilístico

Fue propuesto con el objetivo de representar el proceso de recuperación de información desde el punto de vista de las probabilidades. A partir de una expresión de consulta se puede dividir una colección de N documentos en cuatro subconjuntos distintos: REL conjunto de documentos relevantes, REC conjunto de documentos recuperados, RR conjunto de documentos relevantes recuperados y NN el conjunto de documentos no relevantes no recuperados.

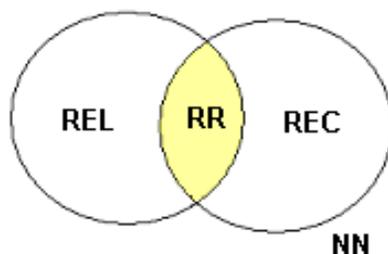


Figura 4: Representación del modelo probabilístico.

El resultado ideal de una consulta se da cuando el conjunto REL es igual REC. Como resulta difícil lograrlo en primera intención, el usuario genera una descripción probabilística del conjunto REL y a través de sucesivas interacciones con el SRI se trata de mejorar la performance de recuperación. Dado que una recuperación no es inmediata dado que involucra varias interacciones con el usuario y que estudios han demostrado que su performance es inferior al modelo vectorial, su uso es bastante limitado. (H. Tolosa, y otros, 2008)

Modelo Vectorial

Este modelo fue planteado y desarrollado por Gerard Salton (Salton, y otros, 1983) y originalmente se implementó en un SRI llamado SMART. Aunque el modelo posee más de treinta años se sigue utilizando debido a sus buenos resultados en la recuperación de documentos.

Conceptualmente, este modelo utiliza una matriz documento–término que contiene el vocabulario de la colección de referencia y los documentos existentes. En la intersección de un término t y un documento d se almacena un valor numérico de importancia del término t en el documento d ; tal valor representa su poder de discriminación. Así, cada documento puede ser visto como un vector que pertenece a un espacio n -dimensional, donde n es la cantidad de términos que componen el vocabulario de la colección. En teoría, los documentos que contengan términos similares estarán a muy poca distancia entre sí sobre tal espacio. De igual forma se trata a la consulta, es un documento más y se la mapea sobre el espacio de documentos. Luego, a partir de una consulta dada es posible devolver una lista de documentos ordenados por distancia (los más relevantes primeros). Para calcular la semejanza entre el vector consulta y los vectores que representan los documentos se utilizan diferentes fórmulas de distancia, siendo la más común la del coseno. (Tolosa, y otros, 2007)

1.3 Sistemas de Gestión de Repositorios Institucionales

Existen diferentes herramientas que permiten crear repositorios, en este trabajo se hará mención a las dos más usadas según las estadísticas de OpenDOAR (Open Access Repositories o Repositorios de Acceso Abierto): DSpace con un 40,9%, en segundo lugar EPrint con el 14,3%.

Uso de Repositorio de Acceso Abierto.

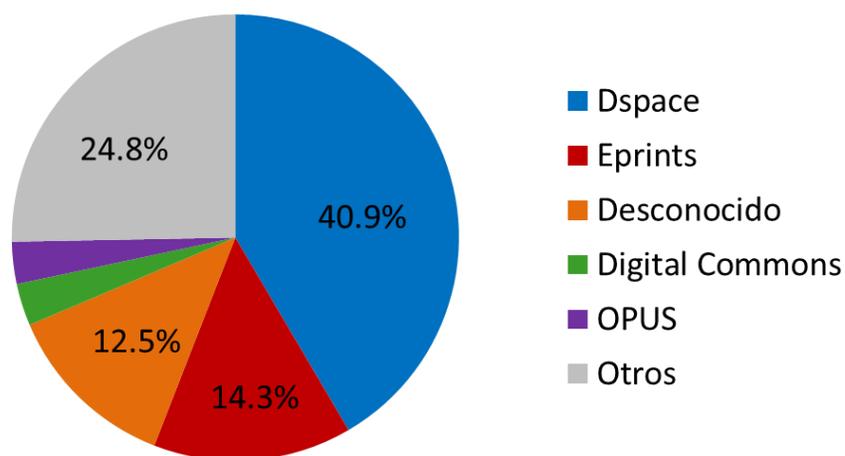


Figura 5: Uso de software de acceso abierto para repositorios (OpenDOAR, 2013)

DSpace

DSpace es un sistema de biblioteca digital diseñado para capturar, almacenar, ordenar, conservar y redistribuir la producción intelectual y de investigación de una Universidad en formato digital. Desarrollado conjuntamente por HP Labs y MIT Libraries. Es un repositorio listo para utilizar. Puede ser instalado como un repositorio institucional. La mayoría de sus usuarios hacen pequeñas modificaciones o simplemente no las hacen a la aplicación. Su estructura interna, que se basa en el concepto de “comunidades” y sus “colecciones” de ítems, consiste de una jerarquía predeterminada que permite organizar el contenido fácilmente. Provee de esquema interno de metadata (basado en DublinCore) para describir el contenido y permite añadir tu propia estructura de metadata.(Álvarez González, 2011)

Se instala sobre sistema operativo Unix. Esta herramienta también incluye un sistema de bases de datos relacionales.

EPrints

EPrints es un software gratuito de código abierto desarrollado en la Universidad de Southampton, Inglaterra. Está diseñado para crear un Repositorio Institucional de edición electrónica para la investigación académica, pero puede usarse para otros propósitos.(Álvarez González, 2011). Es distribuido bajo la Licencia GNU General Public Licence (GPL).

Entre sus características esenciales se encuentran: (Barton, y otros, 2005)

- ◆ Cualquier tipo de contenido aceptado.
- ◆ Puede usar cualquier esquema de metadatos.
- ◆ Características del flujo de trabajo (workflow): el contenido pasa por un proceso de moderación para su aprobación, rechazo, o a otro autor para revisión.
- ◆ Base de datos MySQL.
- ◆ Extensible a través de API usando el lenguaje de programación Perl.
- ◆ Búsqueda de texto completo.
- ◆ Producción RSS.
- ◆ Interfaz basada en la web.
- ◆ Cuenta con apoyo técnico de Eprints-tech y con una Wiki.

1.4 Repositorio de la Universidad de las Ciencias Informáticas

La universidad genera una gran cantidad de información digital en los diferentes eventos científicos que se realizan, y en los cuales participan los miembros de la comunidad universitaria. Antiguamente gran parte de esa información e investigaciones de gran validez, en la mayoría de los casos, no era aprovechada al máximo debido a que no estaba disponible. Con mayor profundidad de esto se habla tres años atrás:

En la Universidad de las Ciencias Informáticas (UCI), el almacenamiento y gestión de la documentación digital está en manos de la biblioteca de la institución. Debido a las publicaciones realizadas por los estudiantes y profesores y en gran medida producto a los Trabajos de Diplomas y Tesis de Maestrías y Doctorados se ha llegado a contar con un gran cúmulo de documentos. Algunas investigaciones son publicadas en la Serie Científica UCI, una publicación electrónica en la que los trabajos enviados son revisados por el Consejo Editorial de la Serie, quienes avalan el trabajo y si cumplen los requisitos, los

publican. Las investigaciones desarrolladas y presentadas en eventos que se realizan en la Universidad como son la Jornada Científica Estudiantil, el Seminario Juvenil Martiano o UCIENCIA no están disponibles(...). Algunos de estos trabajos sólo están disponibles en servidores de proyecto, o en el peor de los casos, la comunidad universitaria no tiene acceso a ellos.(Alvarez Fernández, 2010)

Hoy esta situación ha sido revertida gracias a la creación del Repositorio Institucional de la Universidad de las Ciencias Informáticas que por primera vez comienza a implementarse en el 2011 con grandes ventajas.

Ventajas

Las ventajas han sido expuestas desde que la universidad decidió la creación del Repositorio:

La implantación de un Repositorio Institucional en la UCI facilitaría un acceso centralizado a la información, tanto a los usuarios internos como externos. Uno de los mayores beneficios de una localización centralizada para todos los diferentes contenidos es que la producción intelectual de la Universidad estaría disponible en un lugar. Las ventajas son dobles: para los miembros de la comunidad universitaria el trabajo de una comunidad (departamento, grupo, individuo....) es representado de forma coherente como un todo contextualizado. Para los usuarios externos (visitantes), la producción intelectual de la Universidad está disponible en un lugar centralizado, lo cual les permite encontrar, usar y conocer esta información de una forma mucho más efectiva, además de permitir evaluar los resultados de investigación de la Universidad de forma rápida y eficiente.(Alvarez Fernández, 2010)

Hoy se utiliza la información expuesta en el Repositorio en diferentes sentidos, por ejemplo para los Trabajos de Diplomas realizados por los estudiantes para optar por el título de Ingeniero en Ciencias Informáticas, es de gran utilidad ya que pueden consultar información existente en el Repositorio. También se podría citar su empleo en los diferentes proyectos productivos de la Universidad en trabajos investigativos referentes a sus líneas de investigación. Además se pueden consultar las bibliografías disponibles para realizar cualquier otra actividad como investigaciones de carácter docente, vinculadas al área productiva de la universidad, eventos científicos.

Software utilizado

Cuba cuenta con la Universidad de las Ciencias Informáticas la cual también implementa su Repositorio de tipo Institucional y después de realizado un estudio detallado de varias herramientas para Repositorios, en el año 2010 (parte de un Trabajo de Diploma (Alvarez Fernández, 2010) realizado por estudiantes de la Universidad) se selecciona Dspace 1.5 acorde a las necesidades de dicha institución, atendiendo a los siguientes parámetros: licencias, necesidades del cliente, tecnologías más utilizadas en la Universidad, mantenimiento y soporte internacional.

Contenido

El contenido de este Repositorio está organizado por comunidades y colecciones tal como está estructurado el software sobre el cual se montó y que brinda una mayor organización de los contenidos existentes. Actualmente está compuesto por 5 comunidades (y una comunidad dentro de otra, o sea, una subcomunidad) cada una con sus colecciones correspondientes, donde se almacena toda la información científica de la Institución. A continuación se brinda la descripción jerárquica de sus contenidos por comunidades y colecciones.

Comunidades	Subcomunidades	Colecciones
Eventos	UCIENCIA(con una sola colección UCIENCIA 2012)	Evento Juvenil Martiano, Fórum de Ciencia y Técnica, Fórum de Historia, Jornada Científica Estudiantil, Mi Web por Cuba, Peña Tecnológica, Pedagogía 2013
Otras Publicaciones	----	Publicaciones
Revistas	----	RCCI, Serie Científica, UXI
Tesis	----	Tesis de Doctorado, Tesis de Maestría, Trabajos de Diploma

Tabla 1: Descripción del contenido del Repositorio por comunidades y colecciones. (UCI, 2011)

Los contenidos pueden ser representados con archivos en diferentes formatos ya que Dspace permite importar variedades de formatos e incluso formatos desconocidos para él.

Búsqueda Avanzada

Dspace permite a los usuarios descubrir el contenido en varias maneras, incluyendo:

- Vía de referencia externa usando handles(o manejadores).
- Búsqueda de una o varias palabras clave en los metadatos o extraídas de texto completo.
- Búsqueda por título, autor, fecha o índice de tema, con imagen en miniatura opcional.

La búsqueda es un componente esencial en DSpace. Las expectativas de los usuarios por un motor de búsqueda son bastante altas, por lo que una de las metas de DSpace es suplirlo con muchas características y facilidades en la búsqueda como sea posible. El módulo de búsqueda e indexación de DSpace es un API muy sencillo que permite la indexación de contenidos, la regeneración del índice y realizar búsquedas en el contenido, comunidades y colecciones. Detrás del API se encuentra el motor de búsqueda Lucene, el cual proporciona la búsqueda por campos, por palabras y la capacidad de gradualmente añadir nuevos contenidos indexados sin la necesidad de regenerar todo el índice. Los índices específicos de búsqueda de Lucene son configurables, permitiendo a la institución personalizar cuáles metadatos de DSpace son indexados.

Otro importante mecanismo para la búsqueda en DSpace es la navegación. Este es el proceso mediante el cual el usuario ve un índice en particular, como el índice de títulos, y se desplaza a su alrededor en busca de ítems. El subsistema de navegar proporciona una API sencilla para lograr este objetivo al permitir una llamada para especificar un índice, y una sub-sección de ese índice. El subsistema de navegación, revela la parte del índice de interés. Índices que pueden ser consultados son el título del artículo, fecha de emisión del ítem, autor y tema. Además, el navegar puede limitarse a los elementos dentro de una colección o comunidad en particular.(Alvarez Fernández, 2010)

1.5 Análisis de herramientas que gestionan búsquedas avanzadas

Los repositorios en todo el mundo que usan la herramienta Dspace por lo general tienen el mismo módulo de búsqueda avanzada exceptuando que algunos incluyen nuevos criterios de búsqueda, a continuación se presentan algunos, de los cuales se hace referencia solo a los criterios diferentes a los que trae por defecto Dspace.

1.5.1 Repositorio E-Archivo de la Universidad Carlos III de Madrid

E-Archivo, el Repositorio Institucional de la Universidad Carlos III, tiene como objetivos, reunir, archivar y preservar la producción intelectual resultante de la actividad académica e investigadora de nuestra comunidad universitaria, en formato digital, y ofrecer acceso abierto a dicha producción. La colección abarca tesis doctorales, revistas editadas por la UC3M, documentos de trabajo, artículos, actas de congresos, informes, etc. (Universidad Carlos III de Madrid, 2013)

En este Repositorio la búsqueda avanzada le permite seleccionar distintos campos de búsqueda tal como se realiza en la herramienta Dspace, así como realizar combinaciones entre los distintos términos haciendo uso de operadores booleanos. (Universidad Carlos III de Madrid, 2013)

AND o **+**: Para encontrar documentos que contengan todos los términos de búsqueda combinados con este operador.

OR: Para encontrar documentos que contengan cualquiera de los términos de búsqueda combinados con este operador.

NOT o **-**: Para excluir los documentos que tengan alguna coincidencia con el término de búsqueda introducido.

Este repositorio implementa el criterio identifier (identificador en español) que tiene la herramienta Dspace y que está compuesto de varios tipos de identificadores, de estos tipos solo se implementan en este Repositorio español ISBN e ISSN. Otro criterio que implementan es materia, también id que permite realizar búsqueda por el id que el sistema le asigna a los registros en el momento de realizar un envío determinado.

1.5.2 Repositorio ScholarsArchive @ OSU

ScholarsArchive @ OSU es el servicio digital de la Universidad Estatal de Oregon para la recopilación, clasificación, puesta a disposición y el almacenamiento de la labor académica de la comunidad de la Universidad Estatal de Oregon, Estados Unidos. También incluye materiales de fuera de la institución. (Universidad de Oregon, 2013)

Este implementa criterios como fecha de envío del registro al Repositorio, el tipo de registro y otro que es muy particular de la universidad miembro del comité

1.5.2 Repositorio DSpace @ Cambridge

DSpace @ Cambridge es el repositorio institucional de la Universidad de Cambridge de Reino Unido. El repositorio se estableció para facilitar el depósito de los contenidos digitales de carácter académico, permitiendo a los académicos y los departamentos de la Universidad compartir y preservar este contenido en un entorno administrado. (Universidad de Cambridge, 2013)

Este Repositorio elimina varios criterios de búsqueda como sponsor (patrocinador en español), language (Idioma en español), entre otros, y tiene los criterios tipo de registro y como identificador solo tiene ISBN e ISSN.

1.5.2 Sistemas web que gestionan búsquedas avanzadas y sus criterios

Se hace necesario realizar un estudio a otras herramientas que gestionan búsquedas avanzadas para capturar criterios de búsquedas ya que los Repositorios existentes con Dspace no cuentan con una amplia gama de criterios como se ha visto en el estudio de los mismos.

Muchos son los sistemas que implementan módulos de búsquedas avanzadas, mediante el cual se puede obtener la información que se encuentra almacenada en los mismos. Cada uno de ellos cuenta con una gama de criterios de búsquedas. A continuación se presentan algunos de ellos con sus criterios de búsquedas.

Amazon

Amazon.com, es una compañía estadounidense de comercio electrónico con sede en Seattle, Washington. Fue una de las primeras grandes compañías en vender libros a través de Internet. Fundada por Jeffrey Bezos en 1995. *Amazon.com* comenzó como una librería online, pronto se diversificó en diferentes líneas de productos, añadiendo DVDs, Cds de música, software, videojuegos, electrónica, ropa, muebles, comida y más (Computación Aplicada al Desarrollo S.A. de C.V. , 2012). Este en su módulo de búsqueda avanzada posee criterios tales como:

Palabras Clave, Autor, Título, ISBN, Editorial, Categoría, Idioma, Fecha Publicación y Ordenar por.

Librería Universal

Libreriauniversal.com: es una librería virtual para comprar libros en español, la cual posee librería temática, librería técnica, librería especializada, librería medica, librería cristiana. (Librería Univesal, 1999)

Información del sitio: Idioma-País: Español- España

País: España

Provincia: Zaragoza

Localidad: Barcelona, Madrid, Zaragoza

Dirección: C/Nápoles 260

Teléfono: 902 555 345

Fax: 902 555 543

Los criterios que implementa su búsqueda avanzada son:

ISBN, Título, Autor, Editorial, Año edición, No. pág., Materia, Población e Idioma.

Uniliber

Uniliber: es una Asociación Nacional española de profesionales del Libro antiguo, viejo, usado y agotado, así como de profesionales del Coleccionismo, constituida el 26 de octubre de 2005, de la mano de un grupo de personas procedentes de la práctica totalidad de las Comunidades Autónomas que conforman la geografía nacional. Actualmente tiene como presidente Enrique Rontomé. Ha sido creada con el objetivo prioritario de potenciar y favorecer el comercio a través de Internet de libros y objetos de coleccionismo antiguos, a través de la puesta en marcha de su Portal en Internet. (Libros Antiguos, 2013).

Los criterios que se exponen en su modulo de búsqueda son los siguientes:

Autor, Título, Palabras Clave, ISBN, Librería, Lugar de Edición, Editorial, Provincia, Año, Precio, Ordenado por, Resultados por Página, 1ra Edición.

1.5.3 Selección de criterios

Después de realizado un estudio de los criterios anteriormente planteados en los diferentes sistemas analizados se seleccionaron algunos criterios atendiendo a su nivel de utilización en la actualidad por los diferentes sistemas que gestionen búsquedas avanzadas y a los que puede soportar el estándar de descripción de metadatos Dublin Core que es el que soporta la herramienta Dspace. Estos criterios se implementan en el módulo de búsqueda avanzada del Repositorio Institucional de la universidad. De los que ya se encuentran en el Repositorio se mantienen:

Fecha de Publicación: estará disponible con un cuadro de texto y un selector de fecha (este criterio permite realizar búsquedas a través de la fecha en que se publicó un archivo determinado en el Repositorio Institucional)

Palabra clave: estará disponible con un cuadro de texto (este criterio permite realizar búsquedas a través de una palabra especificada por el usuario).

Autor: estará disponible con una lista desplegable (este criterio permite realizar búsquedas a través de una lista de autores presentados por el sistema).

Título: estará disponible con un cuadro de texto (este criterio permite realizar búsquedas a través de un título especificado por el usuario).

Tema: estará disponible con una lista desplegable (este criterio permite realizar búsquedas a través de una lista de temas presentados por el sistema).

Identificador: estará disponible con un cuadro de texto (este criterio permite realizar búsquedas a través de un identificador especificado por el usuario que puede ser el ISBN del documento u otro).

Sponsor: estará disponible con un cuadro de texto (este criterio permite realizar búsquedas a través de un patrocinador especificado por el usuario).

Resumen: estará disponible con un cuadro de texto (este criterio permite realizar búsquedas a través de palabras especificado por el usuario y que el sistema busca en el resumen del archivo).

Lengua: estará disponible con un cuadro de texto (este criterio permite realizar búsquedas a través de un idioma especificado por el usuario).

Los criterios seleccionados de los sistemas analizados son:

Tipo: estará disponible con un cuadro de texto (este criterio permite realizar búsquedas a través de un tipo especificado por el usuario, ejemplo un libro una revista entre otros).

Editorial: estará disponible con una lista desplegable (este criterio permite realizar búsquedas a través de una lista de editoriales presentados por el sistema).

1.6 Metodologías y Tecnologías utilizadas

El Repositorio Institucional es un proyecto previamente concebido en la Universidad de las Ciencias Informáticas y que además como se ha presentado en el presente trabajo está en manos de la herramienta Dspace, debido a que en el presente se trata del desarrollo de un módulo de ese sistema se hace necesario la utilización de la metodología y las tecnologías descritas previamente definidas por los especialistas sobre las cuales está desarrollado el sistema.

1.6.1 Proceso de desarrollo de software

El proceso de desarrollo está guiado por la metodología RUP con el proceso de mejora de CMMI en su nivel 2.

RUP

Proceso de desarrollo de software que define quién está haciendo qué, cuándo y cómo alcanzar un determinado objetivo, además es la definición del conjunto de actividades que guían los esfuerzos de las personas implicadas en el proyecto, a modo de plantilla que explica los pasos necesarios para terminar el proyecto (Ivar, y otros, 2000)

CMMI

CMMI es un modelo de madurez de mejora de los procesos para el desarrollo de productos y de servicios. Consiste en las mejores prácticas que tratan las actividades de desarrollo y de mantenimiento que cubren el ciclo de vida del producto, desde la concepción a la entrega y el mantenimiento.(Chrissis, y otros, 2009)

La Universidad se encuentra inmersa en un proceso de mejora de los procesos que se desarrollan en la construcción del software basado en el modelo de CMMI específicamente el segundo nivel de madurez. Y

por tanto todos los proyectos productivos de la entidad deben de adoptar este proceso de mejora el cual propone el siguiente ciclo de vida: (Galardy Tarragó, 2012)

Estudio preliminar: se realiza un estudio profundo de la organización cliente que posibilita obtener la información requerida para determinar el alcance del proyecto, así como la estimación del costo, tiempo y el esfuerzo

Modelado de negocio: se comprende el negocio de la entidad con el objetivo de que el software a desarrollar cumpla con lo que realmente quiere el cliente. Se pueden utilizar técnicas para la descripción del modelado del negocio como la Notación de Modelado de Proceso de Negocio.

Requisitos: el objetivo fundamental es desarrollar el modelo del sistema, identificando los requisitos funcionales y no funcionales con las descripciones correspondientes en cada caso.

Análisis y diseño: se realiza el análisis y el modelado del sistema a partir de los requisitos definidos previamente.

Implementación: a partir de los artefactos obtenidos durante el análisis y diseño se procede a realizar la implementación del software en términos de componentes de implementación.

Pruebas internas: se realizan las pruebas internas con el equipo del proyecto en cada una de las iteraciones o versiones finales próximas a ser liberadas, según lo defina el proyecto. Se desarrollan artefactos de pruebas como: diseños de casos de prueba y listas de chequeo.

Pruebas de liberación: pruebas realizadas por parte de la oficina o institución encargada de la calidad y de la certificación del proyecto a todos los entregables de los proyectos antes de ser entregados al cliente para su aceptación.

Despliegue: se realiza la entrega de la aplicación al cliente, así como la configuración y prueba en el ámbito del cliente. Las pruebas realizadas durante esta fase incluye pruebas de aceptación y pruebas pilotos. Se debe realizar además capacitaciones a los trabajadores del sistema.

Soporte: por un tiempo limitado el proyecto ofrecerá un servicio para resolver conflictos y problemas de usabilidad y rendimiento del software entregado al cliente, suministrándole actualizaciones y parches a errores.

1.6.2 Lenguaje de Programación

Un lenguaje de programación es aquel elemento dentro de la informática que nos permite crear programas mediante un conjunto de instrucciones, operadores y reglas de sintaxis; que pone a disposición del programador para que este pueda comunicarse con los dispositivos hardware y software existentes.

JSP 2.1

La tecnología JSP (*Java Server Pages*) es una especificación abierta desarrollada por Sun Microsystems como un alternativa a Active Server Pages (ASP) de Microsoft, y son un componente dominante de la especificación de Java 2 Enterprise Edition (J2EE). Esta tecnología permite desarrollar páginas web con contenido dinámico. Un fichero JSP puede contener etiquetas HTML normales, y elementos especiales para generar el contenido dinámico. Una página JSP es básicamente una página Web con HTML y código Java. (Villafranca, 2008)

CSS 2.1

CSS son las siglas de Cascading Style Sheets, en español, Hojas de estilo en Cascada. Es una tecnología que permite crear páginas Web de una manera más exacta representando un avance importante para los diseñadores, quienes tienen un mayor rango de posibilidades para mejorar la apariencia de las páginas. Es la mejor forma de separar los contenidos y su presentación, imprescindible para que puedan ser creadas páginas Web complejas. El hecho de separar el contenido del diseño presenta entre otras ventajas la de brindar una mejor accesibilidad al documento así como que el mismo pueda ser visualizado en gran cantidad de dispositivos diferentes. Por otra parte uno de los antecedentes más notables que ha conllevado al desarrollo de esta tecnología consiste en que las páginas Web tienen embebidas en su código HTML el contenido del documento con las etiquetas necesarias para darle forma, trayendo como inconveniente que la lectura del código se haga más pesada y difícil a la hora de buscar errores y depurar la página. El uso de las CSS reduce la complejidad del mantenimiento de la página, pues una vez creados los contenidos se utilizan para definir el aspecto de cada elemento. (Torres Valdés, y otros, 2010)

JavaScript 1.8

JavaScript es un lenguaje de programación que ha permitido el gran desarrollo de la Web, ha sido el avance más significativo en el logro de páginas Web dinámicas y exactas en cuanto a posición y presentación de su contenido, es un lenguaje robusto y a la vez ligero, el cual a pesar de ser considerado por muchos como un lenguaje no orientado a objetos permite implementar varias de las características de este paradigma de programación, basado en prototipos, las nuevas clases se generan clonando las clases bases y extendiendo su funcionalidad, cualquier navegador puede interpretar el código JavaScript dentro de las páginas Web, permite la máxima interactividad entre el usuario y la página, además la verificación de los datos introducidos por el usuario antes de enviar el formulario al servidor, el manejo de applets y plugins dentro de múltiples marcos de HTML. (Castro Mecías., y otros, 2008)

1.6.3 Gestor de Base de Datos

Un Sistema de gestión de bases de datos es una aplicación que permite a los usuarios definir, crear y mantener la base de datos, y proporciona acceso controlado a la misma. (Date, 2001)

PostgreSQL 9.1

PostgreSQL es un sistema gestor de base de datos objeto-relacional libre, liberado bajo la licencia BSD. Como muchos otros proyectos Open Source, el desarrollo de PostgreSQL no es manejado por una sola compañía sino que es dirigido por una comunidad de desarrolladores y organizaciones comerciales las cuales trabajan en su desarrollo, dicha comunidad es denominada el PostgreSQL Grupo Global de Desarrollo (PGDG), sus siglas en inglés se definen como: PostgreSQL Global Development Group. PostgreSQL ha tenido una larga evolución, comenzando con el proyecto Ingres en la Universidad de Berkeley. (The PostgreSQL Global Development Group, 2005)

1.6.4 Herramienta CASE para el modelado

CASE es una sigla, que corresponde a las iniciales de: Computer Aided Software Engineering; y en su traducción al español significa Ingeniería de Software Asistida por Computación. Son diversas aplicaciones informáticas destinadas a aumentar la productividad en el desarrollo de software reduciendo el coste de las mismas en términos de tiempo y dinero. Estas herramientas permitirán organizar y manejar la información de un proyecto informático. Permiéndole a los participantes del mismo, que los sistemas

(especialmente los complejos), se tornen más flexibles, más comprensibles y además mejorar la comunicación entre los participantes. (Sosa Vázquez, y otros, 2010)

Visual Paradigm 5.0

Es una suite de aplicaciones, herramienta profesional que soporta el ciclo completo de desarrollo de software. Soporta el modelado visual con UML, que ayuda a una rápida construcción de aplicaciones de calidad, mejores y a un menor costo. Permite generar documentación en html/pdf con los diagramas realizados. Permite la generación de bases de datos, conversión de diagramas entidad-relación a tablas de base de datos, mapeos de objetos y relaciones, ingeniería inversa desde gestores de bases de datos.

Proporciona una plataforma de modelado colaborativo para el trabajo en equipo. Con las características de colaboración en equipo, los miembros del equipo pueden ver y editar el mismo proyecto, o el mismo esquema, incluso de forma simultánea. Todos los cambios se almacenan en el servidor de Visual Paradigm en función de revisión. (Galaray Tarragó, 2012)

Se caracteriza por:

- ◆ Ser un software libre.
- ◆ Presentar una licencia: gratuita y comercial.
- ◆ Estar disponible para múltiples plataformas (Windows, Linux).
- ◆ Permitir el uso de un lenguaje estándar común a todo el equipo de desarrollo que facilita la comunicación.
- ◆ Permitir la realización de ingeniería directa e inversa.

1.6.5 Servidor de Aplicación

Un servidor de aplicaciones es un software que proporciona aplicaciones a los equipos o dispositivos cliente, por lo general a través de Internet y utilizando el protocolo http. Los servidores de aplicación se distinguen de los servidores web por el uso extensivo del contenido dinámico y por su frecuente integración con bases de datos, además es un producto basado en un componente que se encuentra en el plano medio de la arquitectura central de un servidor y trabaja como un intermediario para la seguridad y el mantenimiento, y además de proveer acceso a los datos. Maneja la mayoría de las transacciones relacionadas con la lógica y el acceso a los datos de la aplicación. (La Rosa Agramonte, y otros, 2009)

Apache Tomcat 7

Tomcat es un servidor web con soporte de servlets y JSPs. Incluye el compilador Jasper, que compila JSPs convirtiéndolas en servlets. El motor de servlets de Tomcat a menudo se presenta en combinación con el servidor web Apache. Tomcat puede funcionar como servidor web por sí mismo. En sus inicios existió la percepción de que el uso de Tomcat de forma autónoma era sólo recomendable para entornos de desarrollo y entornos con requisitos mínimos de velocidad y gestión de transacciones. En la actualidad ya no existe esa percepción y Tomcat es usado como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad. Dado que Tomcat fue escrito en Java, funciona en cualquier sistema operativo que disponga de la máquina virtual Java. Tomcat es una implementación libre y open source de las tecnologías Java Servlet y JavaServer Pages desarrolladas en el marco del proyecto de Jakarta en la Apache Software Foundation. (Mesa Bles, y otros, 2009)

Características del Sistema 2

2.1 Introducción

El presente capítulo constituye la propuesta de solución para mejorar la búsqueda avanzada en el Repositorio Institucional de la Universidad de las Ciencias Informáticas, en el mismo se desarrollan los primeros flujos de trabajo de la metodología antes propuesta (RUP). Se realiza la confección del modelo de dominio, además se definen los requerimientos funcionales y no funcionales del sistema. Se diseñan los casos de uso del sistema con sus descripciones textuales, también se muestran los diagramas y descripciones correspondientes.

2.2 Propuesta de solución

El sistema propuesto formará parte del Repositorio Institucional de la Universidad de las Ciencias Informáticas mejorando en el mismo el módulo de búsqueda avanzada para la recuperación de la información almacenada en el mismo. El mismo cuenta con una selección de criterios según el estudio de sistemas que implementan módulos de búsquedas avanzadas. El sistema es de fácil acceso a los usuarios sin necesidad de emplear espacio en disco duro para su instalación y uso ya que es un sistema web, proporciona dos sencillas interfaces que le permita al usuario interactuar de forma cómoda y facilitar las acciones a ejecutar, una de las interfaces del modulo es como tal la de búsqueda avanzada donde estarán disponibles todos los criterios de búsquedas que permitirán la recuperación de la información según las entradas especificadas por el usuario, la otra interfaz es la de administración referente a la selección de los criterios que estarán disponibles en el sistema, desde el momento en que el administrador decida que algún criterio determinado va a formar parte de la búsqueda avanzada del Repositorio Institucional, esto se hace con el objetivo de que el sistema sea lo más configurable posible y por tanto que no sea estático.

A continuación se presentan los prototipos de interfaz de usuario del sistema:



Figura 7: Prototipo de Interfaz de Usuario "Realizar búsqueda".

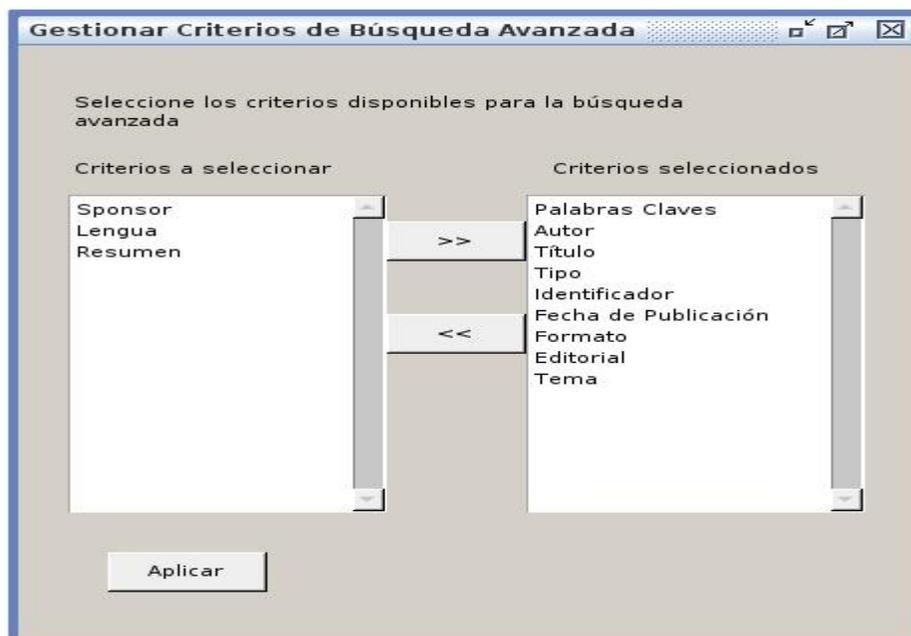


Figura 8: Prototipo de Interfaz de Usuario "Gestionar criterios".

2.3 Modelo de Dominio

Debido al bajo nivel de estructuración que presenta el negocio sobre el que se está investigando y que está altamente centrado en tecnologías informáticas, se propone un modelo conceptual o modelo de dominio, ya que de manera visual permite mostrar los principales conceptos que se manejan en el sistema en desarrollo.

A continuación se muestra la imagen que representa el modelo de dominio del sistema.

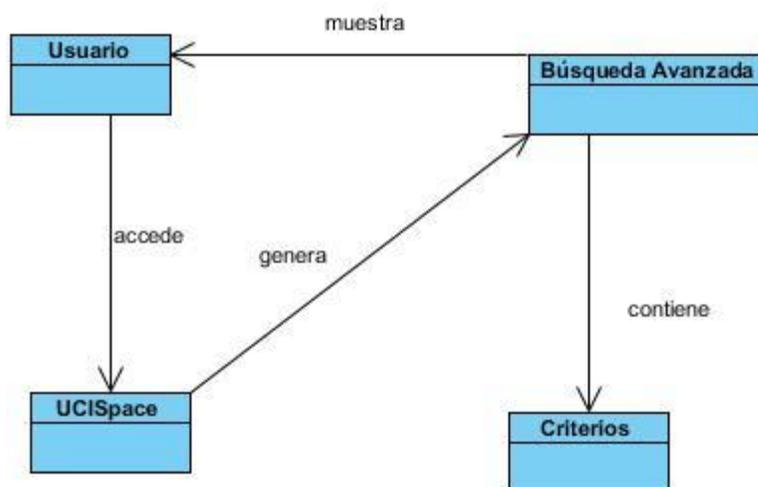


Figura 9: Modelo de Dominio.

El usuario accede al sistema que genera el modulo de búsqueda avanzada, donde están los criterios mediante los cuales se realiza la consulta, realiza su consulta y el sistema le presenta los resultados de su búsqueda.

2.3.1 Clases del modelo de dominio

Usuario: representa al usuario que va a realizar la búsqueda en el módulo.

Búsqueda Avanzada: representa el módulo de búsqueda avanzada encargado de recuperar la información según las especificaciones del usuario.

UCISpace: representa el Repositorio Institucional que almacena toda la información a recuperar.

Criterios: representa los criterios de búsquedas que serán seleccionados por el administrador del sistema.

2.4 Especificación de Requisitos

2.4.1 Requisitos Funcionales

El sistema presenta los siguientes requisitos funcionales, de los cuales se agregan realizar búsqueda por editorial, y tipo:

RF1: Realizar búsqueda.

RF1.1 Realizar búsqueda por Autor.

RF1.2 Realizar búsqueda por Tema.

RF1.3 Realizar búsqueda por Título.

RF1.4 Realizar búsqueda por Fecha de Publicación.

RF1.5 Realizar búsqueda por Sponsor.

RF1.6 Realizar búsqueda por Identificador.

RF1.7 Realizar búsqueda por Editorial.

RF1.8 Realizar búsqueda por Lengua.

RF1.9 Realizar búsqueda por Tipo.

RF1.10 Realizar búsqueda por Resumen.

RF1.11 Realizar búsqueda por Palabras Claves.

RF2: Seleccionar criterios de búsqueda.

RF2.1: Listar criterios de búsqueda existentes.

RF2.2: Listar criterios de búsqueda disponibles.

RF2.3: Insertar criterios de búsqueda.

RF2.4: Eliminar criterios de búsqueda.

RF2.5: Eliminar índice del fichero de configuración.

RF2.6: Insertar índice en el fichero de configuración.

2.4.2 Requisitos no funcionales

Los requisitos no funcionales son propiedades o cualidades que debe cumplir un sistema.

- ◆ Seguridad

RNF1 El acceso al sistema será controlado con nombre de usuario y contraseña para el módulo de administración.

RNF2 El sistema permitirá acceso a los usuarios solo a las funcionalidades que se correspondan con los permisos asignados.

- ◆ Restricciones del Diseño

RNF3 Se utilizará para la construcción del sistema el lenguaje de programación JSP y herramientas que se distribuyan bajo licencias libres.

RNF4 Las librerías usadas no deben ser propietarias.

RNF5 El sistema podrá ser accedido a través del navegador web Firefox, Opera que deben de tener el soporte Java Script habilitado desde los sistemas operativos Windows y Gnu/Linux.

- ◆ Hardware

RNF6 La PC del servidor debe de tener al menos 512 MB de RAM, procesador con 1.30 GHz, 40 GB de disco duro.

- ◆ Software

RNF7 Se utiliza PostgreSQL versión 9.1 como Gestor de Base de Datos.

RNF8 Se utiliza Apache Tomcat 7.0 como servidor web.

2.5 Modelo de casos de uso del sistema

2.5.1 Actores del sistema

Los actores representan personas, sistemas o hardware externo que interactúan con el sistema, o sea, los actores son agentes externos al subsistema los cuales tienen relación con el mismo.

A continuación se definen los actores del sistema:

Actor	Descripción
Usuario	Es la persona que puede realizar búsquedas interactuando con el módulo de búsqueda avanzada.
Administrador	Es la persona que realiza la selección de los criterios que se presentarán luego en el módulo de búsqueda avanzada.

2.5.2 Definición de casos de uso del sistema

CU1:	Realizar Búsqueda
Actor:	Usuario
Descripción:	Permite realizar búsquedas a través de criterios descriptivos para la recuperación de la información almacenada en el Repositorio Institucional
Referencia:	RF1

CU2:	Seleccionar criterios de búsqueda
Actor:	Administrador
Descripción:	Permite seleccionar los criterios de búsqueda que estarán disponibles en la búsqueda avanzada.
Referencia:	RF2

2.5.3 Diagramas de caso de uso del sistema

El diagrama de casos de uso del sistema brinda las funcionalidades que el sistema debe ofrecer para aportar un resultado de valor, representando gráficamente a los procesos y su interacción con los actores del sistema.

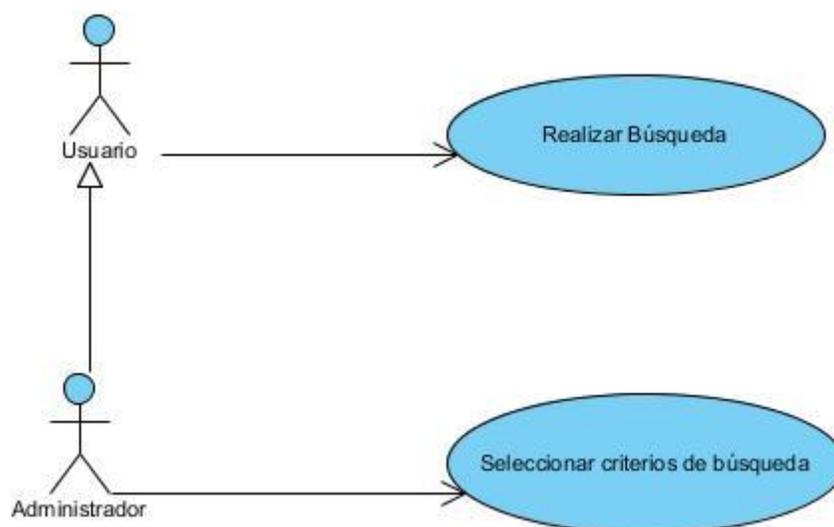


Figura 10: Diagrama de casos de uso del sistema.

2.5.4 Descripción textual de los casos de uso del sistema

Caso de Uso:	Realizar búsqueda.
Actores:	Usuario
Resumen:	El caso de uso inicia cuando el Usuario decide realizar búsquedas a través de criterios descriptivos para la recuperación de la información almacenada en el Repositorio Institucional. El mismo concluye cuando se muestran los resultados.

Precondiciones:	El Usuario debe estar dentro del sistema.	
Referencias	RF1	
Prioridad	Crítico	
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
<p>En el menú izquierdo de la página principal aparece la opción:</p> <ul style="list-style-type: none"> - Buscar en DSpace <p>En dicho menú aparecen la opción:</p> <ul style="list-style-type: none"> -Búsqueda avanzada <p>1. El Usuario selecciona la opción “Búsqueda avanzada”.</p>	2.	<p>El sistema muestra una interfaz que permite realizar la Búsqueda avanzada atendiendo a varios criterios de búsqueda: Autor, Palabras Clave, Identificador, Título, Editorial, Tema, Resumen, Sponsor, Tipo, Fecha de Publicación y Lengua.</p> <p>Aparece el tipo de documento que puede ser una Tesis, Libro, Revista, Animación o Video además de los botones “Buscar” y “Limpiar”.</p>
<p>3. El Usuario llena los campos referentes a los criterios de búsqueda y selecciona la opción “Buscar”.</p>	4.	<p>El sistema busca la información referente a la búsqueda especificada por el usuario y visualiza el resultado de la búsqueda obtenida.</p>
	5.	<p>Finaliza el caso de uso.</p>
Flujo Alterno		
Acción del Actor	Respuesta del Sistema	
<p>3. El usuario llene los campos referentes a los criterios de búsqueda y seleccione la opción “Limpiar”.</p>	4.	<p>El sistema borra todos los campos que hayan sido llenados.</p>

Tabla 2: Descripción del Caso de uso Realizar búsqueda.

Caso de Uso:	Seleccionar criterios de búsqueda.
Actores:	Administrador

Resumen:	El caso de uso inicia cuando el Administrador decide seleccionar los criterios de búsquedas descriptivos que estarán disponibles en el módulo de búsqueda avanzada. El mismo concluye terminada esta acción.	
Precondiciones:	El Administrador debe estar autenticado en el sistema.	
Referencias	RF2	
Prioridad	Crítico	
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
En el menú izquierdo de la página principal aparece la opción: - Administrar 1. El Administrador selecciona la opción "Administrar".	2. El sistema muestra la interfaz que permite la administración del sistema.	
En el menú izquierdo de esta interfaz de administración aparece la opción: -Seleccionar Criterios de búsquedas 3. El Administrador selecciona la opción "Seleccionar Criterios de búsquedas".	4. El sistema muestra una interfaz que le permite al Administrador seleccionar los criterios de búsqueda que estarán disponibles en el módulo de búsqueda avanzada.	
5. El Administrador selecciona los criterios que estarán disponibles en el módulo y selecciona la opción "Aplicar".	6. El sistema guarda los cambios realizados por el Administrador.	
	7. Finaliza el caso de uso.	

Tabla 2: Descripción del Caso de uso.

2.6 Modelo de análisis

El análisis es uno de los flujos de trabajos desarrollado en la fase de elaboración. Consiste en obtener una visión del sistema que se preocupa de ver qué hace, de modo que solo se interesa por los requisitos funcionales. Representa las cosas del mundo real, no de la implementación automatizada de ellas.

2.6.1 Diagramas de clases del análisis

A continuación se presentan los diagramas de clases del análisis correspondiente a los casos de uso descritos anteriormente donde:

CI: tiene como significado clase interfaz, mediante la cual el usuario interactúa con el sistema.

CC: representa la clase controladora, la cual construye las instancias de las clases necesarias.

CE: es la clase entidad donde se encuentran los datos almacenados.

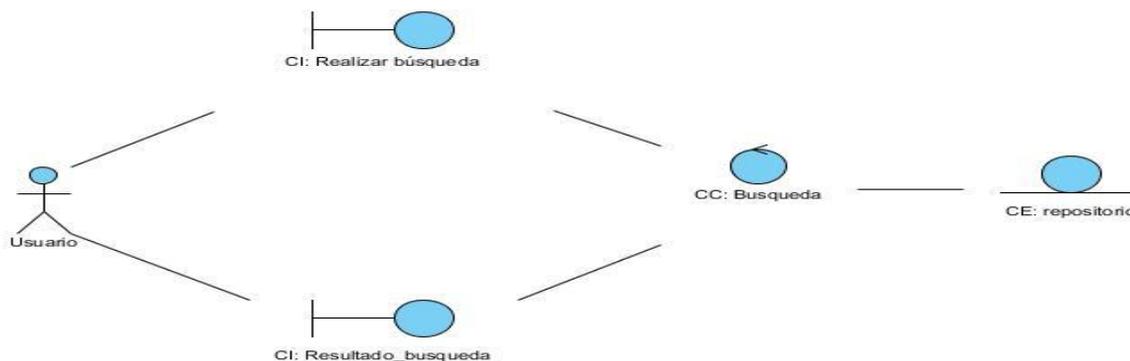


Figura 11: Diagrama de clases del análisis para el CU "Realizar Búsqueda".

Realizar_búsqueda: es la clase interfaz que permite recopilar la información mediante la cual se construye la consulta especificada por el usuario.

Resultado_búsqueda: clase interfaz donde se muestran los resultados propuestos por el sistema.

Busqueda: clase controladora que permite gestionar todos los datos tanto los introducidos por el usuario como los generados por el sistema.

repositorio: clase donde está almacenada toda la información de los registros del sistema y sobre los cuales se realiza la búsqueda.

Usuario: representa a cualquier usuario que accede al sistema y realiza la búsqueda según su necesidad.

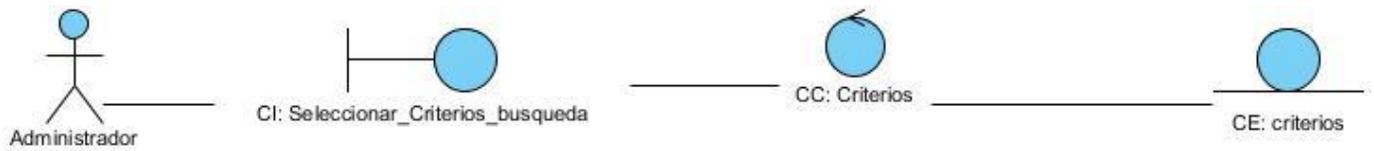


Figura 12: Diagrama de clases del análisis para el CU “Seleccionar criterios de búsqueda”.

Administrador: representa al administrador del sistema el cual lleva a cabo la selección de los criterios.

Seleccionar_Criterios_búsqueda: clase interfaz que permite al administrador visualizar los criterios disponibles y los criterios a seleccionar.

Criterios: clase controladora que maneja los criterios según las especificaciones del usuario.

criterios: contiene los datos que son manejados por el sistema en el proceso de selección de criterios.

2.7 Modelo de Diseño

El Modelo de Diseño se utiliza para documentar el diseño de un sistema. Es un modelo de objeto que describe la realización de los casos de uso y sirve como una abstracción del Modelo de Implementación y del código fuente. Se utiliza como entrada esencial para las actividades en los flujos de trabajo Implementación y Prueba. Es un artefacto integral que abarca todas las clases del diseño y sus relaciones, e incluye los diagramas de clases y de interacción del diseño. (Pressman, 2005)

2.7.1 Diagrama de colaboración

En el análisis, las interacciones entre objetos se pueden representar a través de diagramas de colaboración o de secuencia. Para la solución del módulo en desarrollo se emplean los diagramas de colaboración, pues el objetivo fundamental es identificar los requisitos y las responsabilidades sobre los objetos, y no identificar secuencias de interacción detalladas y ordenadas cronológicamente, para lo cual serían más factibles los diagramas de secuencia.

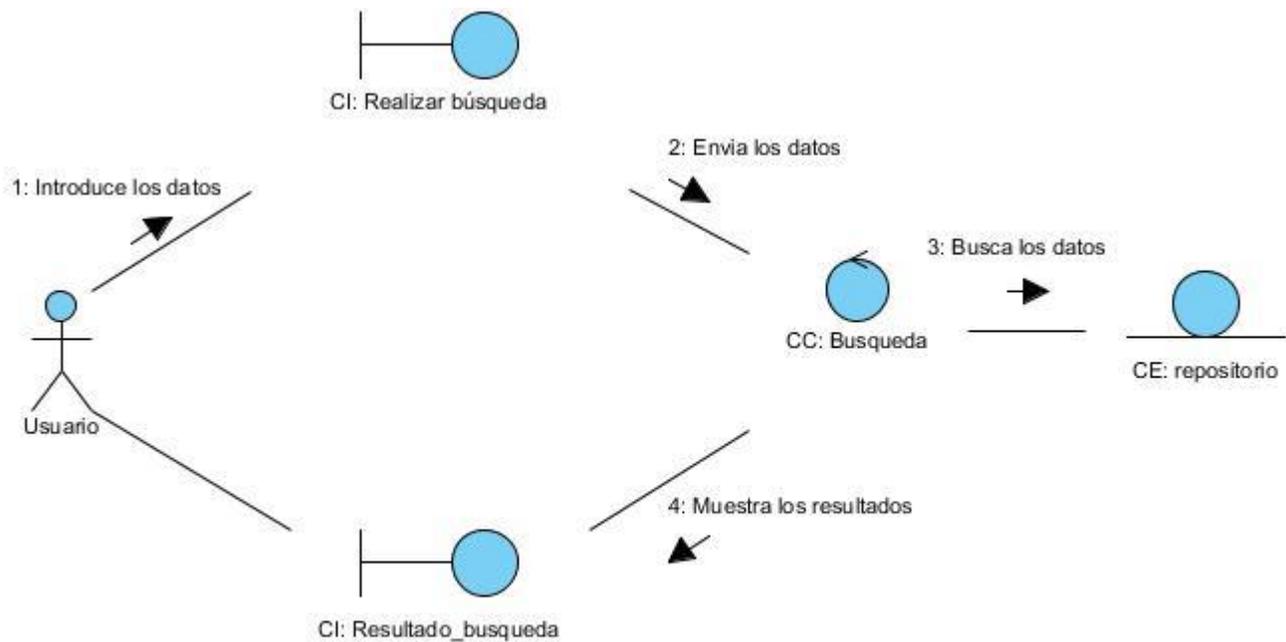


Figura 13: Diagrama de colaboración para el CU "Realizar Búsqueda".

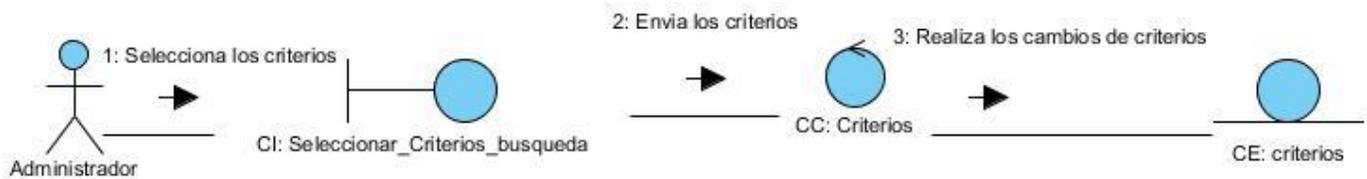


Figura 14: Diagrama de colaboración para el CU "Seleccionar criterios de búsqueda".

2.7.2 Diagramas de clases del diseño

En la fase de elaboración de la metodología RUP se comienza con un análisis de los elementos significativos de la arquitectura como parte de la primera iteración de elaboración y en las siguientes iteraciones se refina la arquitectura hasta diseñar todos sus elementos. El diseño es el centro de atención al final de la fase de elaboración y el comienzo de las iteraciones de construcción. Esto contribuye a una arquitectura estable, sólida y crea un plano al modelo de implementación. En el diseño se confeccionan

los diagramas de clases del diseño. Los elementos básicos que se pueden encontrar en este diagrama son clases y relaciones entre estas.

A continuación se presentan los diagramas de clases del diseño del sistema donde:

CP: representa una clase interfaz del sistema.

SP: es una clase servidora, que se encarga de manejar los datos.

Form: es un formulario que permite recopilar información.

CC: representa una clase controladora.

CE: representa un clase entidad

La página principal mediante un enlace solicita a la clase servidora SimpleSearchServlet mostrar la interfaz de búsqueda avanzada (CP_Busqueda_avanzada), quien que es enlazada por dicha clase servidora y que además contiene un formulario mediante el cual el usuario especifica la consulta que desea realizar. Los datos introducidos por el usuario son procesados por la servidora quien consulta las clase Community y QueryArg para obtener los datos necesarios que dan respuesta al usuario.

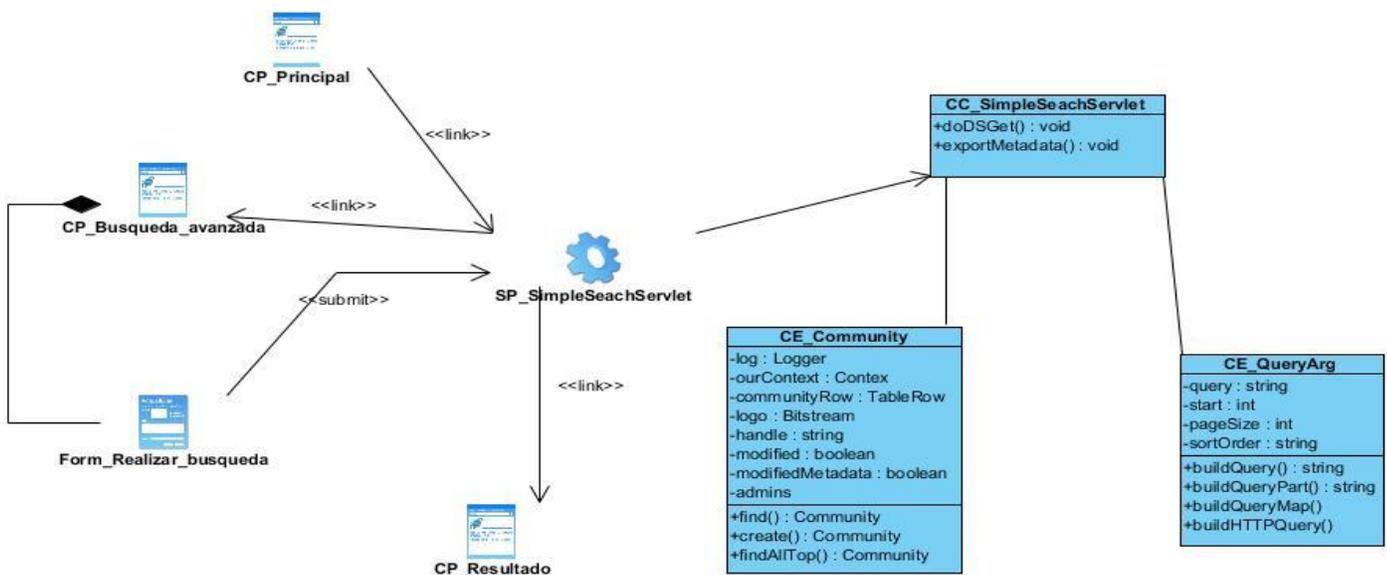


Figura 15: Diagrama de clases del diseño para el CU "Realizar Búsqueda".

En la siguiente figura se representa el diagrama de diseño correspondiente al CU Seleccionar Criterios de búsqueda, el mismo contiene la página principal que solicita a la servidora Criterios, a través de un enlace, mostrar la página Seleccionar_Criterios, que permite al administrador después de autenticado en el sistema seleccionar los criterios que estarán disponibles en la búsqueda avanzada y que son obtenidos a través de las clases Criteria_Sel y Criteria_a_Sel. Finalmente el sistema almacena en la Base de Datos Dspace los criterios seleccionados.

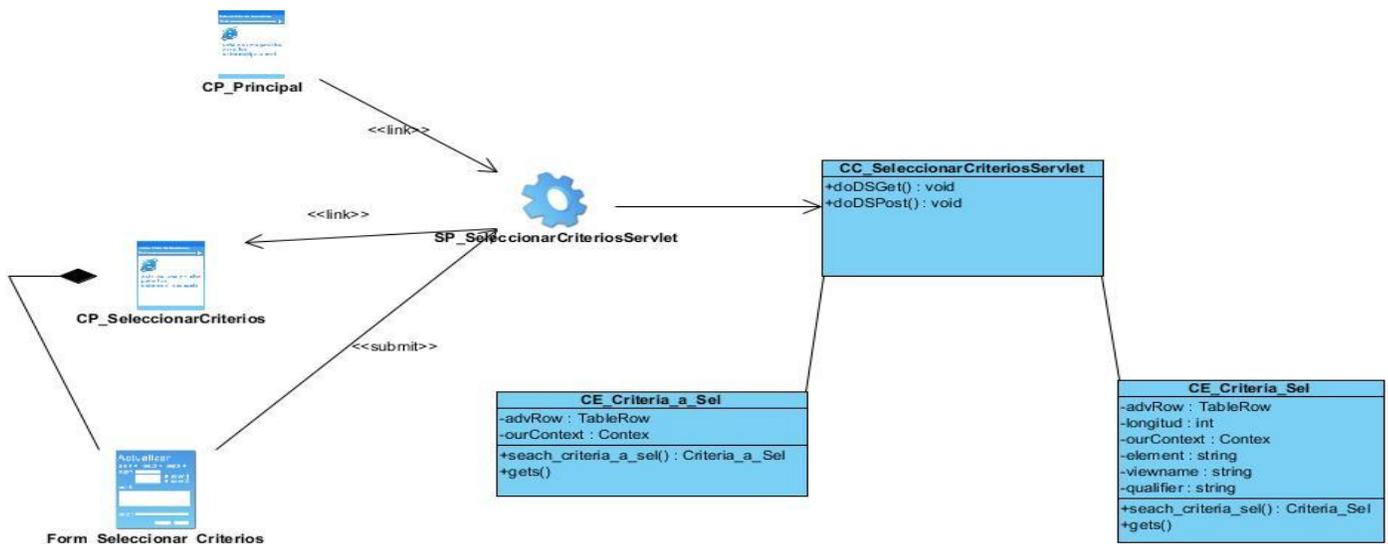


Figura 16: Diagrama de clases del diseño para el CU "Seleccionar Criterios de búsqueda".

2.8 Modelo de datos

El modelo de datos en el sistema consiste en una tabla donde se almacenan los criterios que se encuentren disponibles, seleccionados previamente por el administrador.

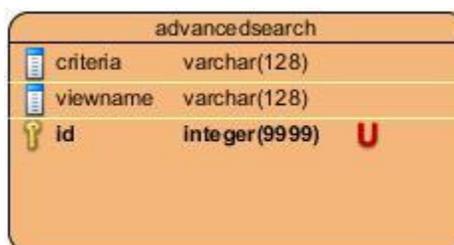


Figura 17: Diagrama del Modelo de Datos.

id: es un campo de valor entero y único para cada criterio disponible.

criteria: es un campo donde se almacenan los criterios de búsqueda disponible.

viewname: representa la disponibilidad del criterio de búsqueda (o sea si está o no disponible ese criterio en el módulo).

2.10 Arquitectura de la aplicación. Arquitectura 3 Capas

La arquitectura en capa define el patrón en capas como una organización jerárquica. Lo que posibilita un diseño basado en niveles de abstracción creciente, posibilitando a los implementadores, particionar un problema en una secuencia de pasos incrementales. Este estilo de desarrollo en varios niveles, facilita que en caso que ocurra algún cambio, sólo se tendrían que realizar las correcciones necesarias en el nivel requerido sin tener que revisar código de otros niveles. La programación por capas es un estilo de programación en la que el objetivo primordial es dividir, fraccionar y llegar a separar la Presentación (donde muestras y obtienes datos), de la Lógica de Negocio (donde realizas operaciones) y con todo esto se obtendría independencia, por lo que si hay cambios de arquitectura se puede acceder a los datos o cambiar el negocio o modificar la presentación y solo sería en esa fracción de la capa. En dichas arquitecturas a cada nivel se le confía una misión simple, lo que permite el diseño de arquitecturas escalables (que pueden ampliarse con facilidad en caso de que las necesidades aumenten). (Santos Pérez, y otros, 2008)

Contiene tres capas:

1 - **Capa de presentación:** En esta capa se diseña todo lo que constituye la interfaz gráfica y la interacción del usuario con el software. Se comunica únicamente con la capa de negocio. Representa el conjunto de componentes que genera la información que se representará en la interfaz de usuario del cliente.

2 - **Capa de lógica de negocio:** En esta capa residen los objetos de negocio que interactúan con los objetos de dominio. Se denomina capa de negocio (e incluso de lógica del negocio) pues es aquí donde se establecen todas las reglas que deben cumplirse para un correcto funcionamiento lógico de la aplicación. Los objetos de negocio, separan los datos y la lógica de negocio haciendo uso del modelo de

objetos definido, dichos objetos de negocio son los encargados de procesar la información proveniente tanto de la Presentación, como de la Capa de Acceso a Datos.

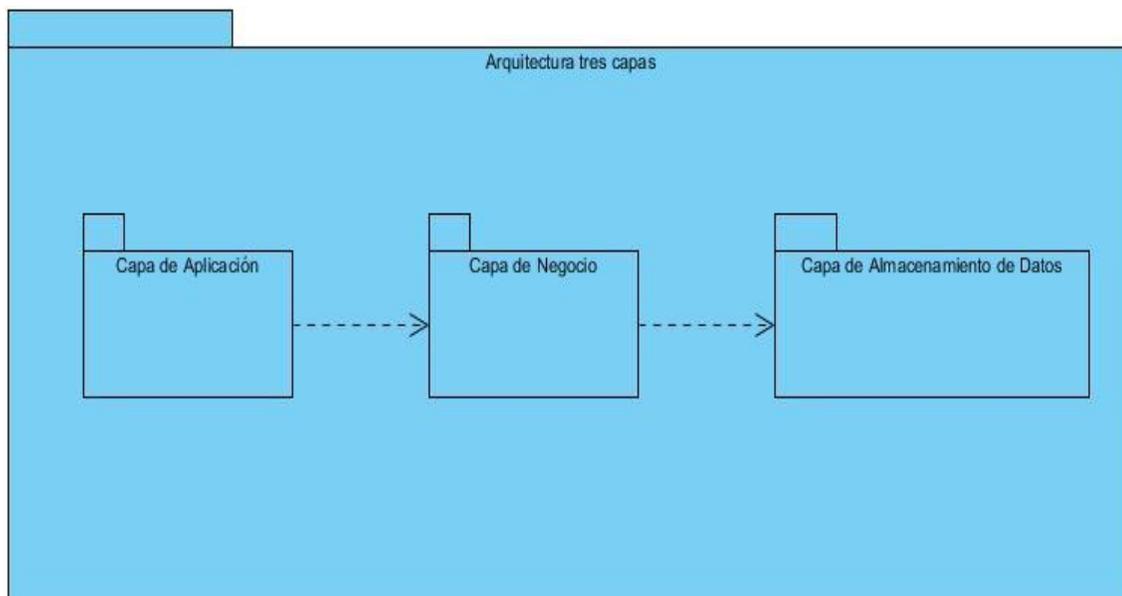


Figura 18: Arquitectura tres capas.

3 - **Capa de Acceso a Datos:** La capa de Acceso a Datos es el puente entre la capa de Lógica de Negocio y el Sistema de Base de Datos. Encapsula la lógica de acceso a datos. Aquí se encuentran componentes que hacen transparente el acceso a la base de datos. Este es el lugar idóneo para implementar los objetos de acceso a datos, permitiendo ingresar, obtener, actualizar y eliminar información del Sistema de Bases de Datos. Los Objetos de Acceso a Datos (DAOs) encapsulan la persistencia de los objetos de dominios, proveen la persistencia de los objetos transitorios y las actualizaciones de los objetos existentes en la base de datos. Las implementaciones de los DAOs estarán disponibles para los objetos de negocio.

2.11 Patrones GRASP

Los patrones GRASP (*General Responsibility Assignment Software Patterns*, en español Patrones Generales de Software para la Asignación de Responsabilidades) describen los principios fundamentales de la asignación de responsabilidades a objetos (Larman, 1999)

- ◆ El patrón Experto: soluciona el problema de asignar una responsabilidad de forma general, tomamos decisiones sobre la asignación de responsabilidades a las clases. Si se hacen en forma adecuada, los sistemas tienden a ser más fáciles de entender, mantener y ampliar y se presenta la oportunidad de reutilizar los componentes en futuras aplicaciones.
- ◆ El patrón Controlador: asigna la responsabilidad de recibir o manejar los mensajes de evento del sistema.
- ◆ El patrón Bajo Acoplamiento: da soporte a una dependencia escasa y a un aumento de la reutilización, manteniendo las clases que no dependan de muchas otras clases.
- ◆ El patrón Alta Cohesión: aconseja mantener la clase lo más relacionada posible para controlar la complejidad de la misma.

Implementación y Prueba 3

3.1 Introducción

En el presente capítulo se estará abordando sobre los flujos de trabajo de Implementación y Prueba, donde en la Implementación se estará describiendo la estructura del sistema y la manera en que se agrupan sus elementos fundamentales mediante el empleo de los diagramas de paquetes para su representación más general. Y para la descripción de las funcionalidades más complejas de manera más exquisita se utilizan los diagramas de componentes.

Mientras que las pruebas son de vital importancia en el desarrollo de software debido a que propician la obtención de buenos resultados. Se realizan con el objetivo de revisar que el software tenga el nivel de calidad requerido. Permiten detectar errores durante su funcionamiento, probando así la entrada y salida correcta de datos, tomando las especificaciones como referencia de comportamiento de la aplicación. Este proceso debe comenzar en la fase de requerimientos y terminar con la finalización de la aplicación. En el proceso de pruebas se definen varios métodos, técnicas y tipos de pruebas, las cuales se abordarán más adelante.

3.2 Modelo de Implementación

El modelo de implementación describe cómo los elementos del modelo de diseño se implementan en términos de componentes, cómo se organizan los componentes de acuerdo con los mecanismos de estructuración y modularización disponibles en el entorno de implementación y en el lenguaje o lenguajes de programación utilizados y cómo dependen los componentes unos de otros (Acuña, 2009).

3.2.1 Diagrama de despliegue

El Diagrama de Despliegue muestra la configuración de los nodos de procesamiento en tiempo de ejecución, los enlaces de comunicación entre ellos y las instancias de los componentes y objetos que

residen en ellos. El propósito del diagrama de despliegue es capturar la configuración de los elementos de procesamiento y las conexiones entre estos elementos en el sistema. El diagrama consiste en uno o más nodos, dispositivos, conectores entre nodos y conectores entre nodos y dispositivos. A continuación se presenta el diagrama de despliegue propuesto para el sistema en cuestión:



Figura 19: Diagrama de despliegue.

PC Cliente: representa la PC desde donde el usuario se conecta al sistema a través del protocolo http.

Servidor de Aplicaciones Web Tomcat: representa al servidor Tomcat el cual permite al usuario interactuar con el sistema brindando todos los servicios necesarios, el mismo se conecta a través del protocolo TCP/IP al servidor de base de datos PostgreSQL.

Servidor de Base de Datos PostgreSQL: es el servidor de Base de Datos PostgreSQL en el cual están almacenados todos los datos necesarios para el momento en que el sistema realice la solicitud de los mismos.

3.2.2 Diagrama de componentes

Los diagramas de componentes modelan la vista estática del sistema y son usados para estructurar el modelo de implementación en términos de subsistema de implementación y mostrar las relaciones entre los elementos de implementación. El uso más importante de estos diagramas es mostrar la estructura de alto nivel del modelo de implementación, especificando los subsistemas de implementación y sus dependencias al importar un código y organizar los subsistemas de implementación en capas.

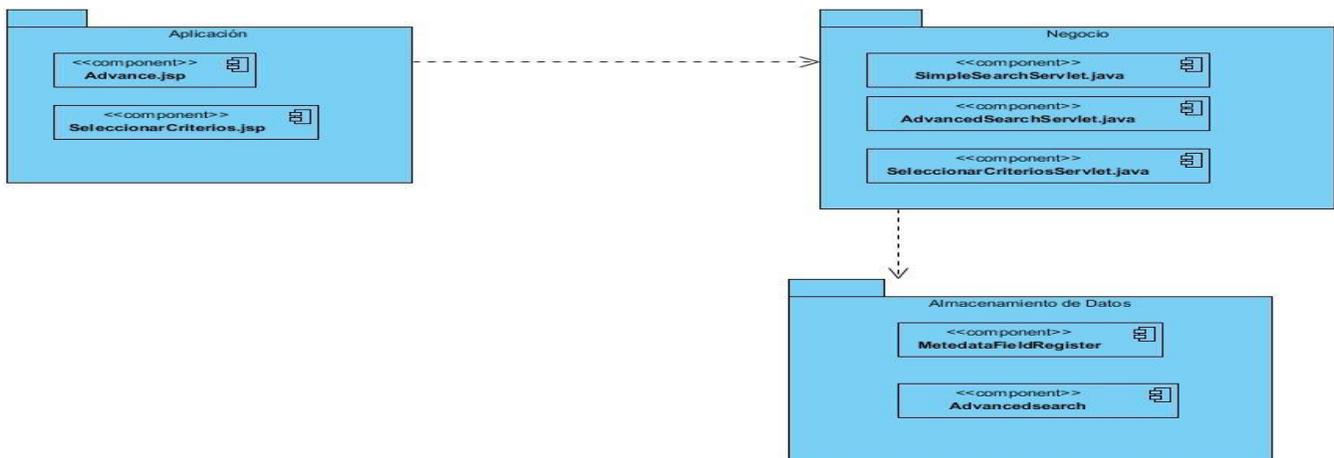


Figura 20: Diagrama de Componentes.

3.3 Métodos de Validación

En la actualidad existen métodos con diferentes enfoques que posibilitan la validación de software. Para la presente investigación se utilizan las pruebas de seguridad y de funcionalidad.

3.3.1 Pruebas de Funcionalidad

La realización de pruebas es un elemento crítico para garantizar la calidad del software. Es una actividad en la cual un sistema o componente es ejecutado bajo condiciones o requerimientos específicos. Las pruebas se aplican en las diferentes fases del desarrollo del software, con el objetivo limitar gradualmente los posibles errores, a medida que el desarrollo avanza. En este acápite se verán los dos grupos en los cuales se dividen como son:

Las pruebas de Caja Blanca: Se nombran de esta forma porque prueban la parte interna del software, específicamente el código fuente. Se comprueban los caminos lógicos del sistema generando casos de prueba que ejerciten las estructuras condicionales y los bucles. Existen varios métodos que analizan diferentes partes del software y se complementan entre sí para garantizar la calidad del sistema.

Las Pruebas de Caja Negra: Se basan en los requerimientos funcionales del sistema y se llevan a cabo desde el exterior de la aplicación. Estas pruebas permiten medir en qué grado se cumplen los

requerimientos solicitados por el cliente y se aplican sobre la interfaz del software observando las respuestas del sistema antes determinadas acciones. Son realizadas mediante casos de prueba con el objetivo principal de demostrar que las funcionalidades son correctas.

En la validación de la solución propuesta se utilizaron las pruebas de caja negra, donde se aplicó el método de partición equivalente, que divide el campo de entrada de un programa en clases de datos de los que se pueden derivar casos de pruebas. Este método está orientado a los requisitos funcionales del software y permite la creación de juegos de datos para probar a profundidad todas las funcionalidades del software e identificar los diseños de casos de prueba (Pressman, 2010).

3.4 Pruebas de Caja Negra

El método de prueba de caja negra se aplica a la interfaz de la aplicación. Pretende demostrar que las funcionalidades del módulo son operativas, las entradas se aceptan correctamente y que se producen los resultados esperados. Dentro del método de caja negra se utiliza la técnica “Partición Equivalente” siendo considerada como una de las más efectivas en la evaluación de los valores válidos, inválidos y los que no es necesario proporcionar un valor del dato en las entradas existentes en la aplicación. La Partición Equivalente divide el dominio de entrada de un programa en clases de datos a partir de las cuales se derivan casos de prueba (Pressman, 2010).

3.4.1 Casos de Prueba

Diseño de caso de prueba basado en el Caso de Uso “Seleccionar criterios de búsqueda”.

Descripción General: El caso de uso inicia cuando el Administrador decide seleccionar los criterios de búsquedas descriptivos que estarán disponibles en el módulo de búsqueda avanzada. El mismo concluye terminada esta acción.

Condiciones de ejecución: El Administrador debe estar autenticado en el sistema.

SC< Seleccionar criterios de búsqueda >

Escenario	Respuesta del sistema	Criterios a Seleccionar	Criterios Seleccionados	Flujo central
El Administrador selecciona la opción "Administrar".	El sistema muestra la interfaz que permite la administración del sistema.			El administrador accede en la parte lateral izquierda del sistema la opción "Administrar"
El Administrador selecciona la opción "Seleccionar Criterios de búsquedas".	El sistema muestra una interfaz que le permite al Administrador seleccionar los criterios de búsqueda que estarán disponibles en el módulo de búsqueda avanzada.			El administrador accede en la parte lateral izquierda del sistema la opción "Seleccionar Criterios de búsquedas".
El Administrador selecciona los criterios correctamente	El sistema guarda los cambios realizados por el Administrador.	Editorial	Tema	El administrador selecciona los criterios que estarán disponibles en el módulo de búsqueda y presiona el botón "Aplicar"
El Administrador selecciona los criterios incorrectamente	El sistema muestra el siguiente mensaje "Debe de haber al menos un criterio de búsqueda seleccionado".	N/A		El administrador selecciona los criterios que estarán disponibles en el módulo de búsqueda dejando los criterios seleccionados vacíos y presiona el botón "Aplicar"

Descripción de variable:

No	Nombre del campo	Clasificación	Valor Nulo	Descripción
1	Título	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
2	Palabras Claves	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
3	Identificador	Campo de texto	Si	En este campo de texto solo pueden aparecer valores alfanuméricos.
4	Editorial	Campo de texto	Si	En este campo solo pueden aparecer valores alfanuméricos.
5	Tema	Campo de texto	Si	En este campo solo pueden aparecer valores alfanuméricos.
6	Resumen	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
7	Sponsor	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
8	Fecha de Publicación	Campo de texto	Si	En este campo de texto solo pueden aparecer números.
9	Lengua	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.

Diseño de caso de prueba basado en el Caso de Uso “Realizar Búsqueda”.

Descripción General: El caso de uso inicia cuando el Usuario decide realizar búsquedas a través de criterios descriptivos para la recuperación de la información almacenada en el Repositorio Institucional. El mismo concluye terminada esta acción.

Condiciones de ejecución: El Usuario debe estar dentro del sistema.

SC<Realizar Búsqueda>

Escenario	Respuesta del sistema	Buscar	Título	Autor	Lengua	Palabras clave	Editorial	Identificador	Resumen	Tema	Sponsor	Fecha de Publicación	Tipo	Flujo central
El Usuario selecciona la opción "Búsqueda avanzada".	El sistema muestra una interfaz que permite realizar la Búsqueda avanzada atendiendo a varios criterios de búsqueda: Autor, Palabras Clave, Identificador, Título, Editorial, Formato, Tema, Resu													El usuario presiona el enlace "Búsqueda avanzada" en la parte superior izquierda en el panel "Buscar en Dspace"

	men, Sponsor, Tipo, Fecha de Publicación y Lengua.													
El Usuario llena los campos correctamente	El sistema busca la información referente a la búsqueda especificada por el usuario y visualiza el resultado de la búsqueda obtenida.	Todo Disponible	Implementación de algoritmos de agrupamiento.	Daniel Campos Diaz	en	Implementación	Generativa	5	La digitalización de la información y la automatización de los servicios, han provocado la generación de grandes volúmenes de información.	computación	UCI12	###	video	El usuario selecciona cada uno de los criterios por los que desea realizar la búsqueda y llena los campos referentes a esos crite

El usuario llena los campos incorrectamente	El sistema no encuentra ningun resultado u muestra el mensaje "La búsqueda no ha producido resultados."	N/A	implementación/*	Dat*	en glis h	implementaciónbas ica	sgd asdf	lhghd	digitalización /*	dfshd vsdrt y	uciuh oriente	###	N/A	rios El usuario selecciona cada uno de los criterios por los que desea realizar la búsqueda y llena los campos referentes a esos criterios con entradas inválidas
---	---	-----	------------------	------	-----------	-----------------------	----------	-------	-------------------	---------------	---------------	-----	-----	--

Descripción de variable:

No	Nombre del campo	Clasificación	Valor Nulo	Descripción
----	------------------	---------------	------------	-------------

1	Búsqueda	Lista desplegable	No	En este campo se selecciona donde se realizara la búsqueda. Brinda opciones para la selección.
2	Autor	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
3	Título	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
4	Lengua	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
5	Palabras Clave	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
6	Identificador	Campo de texto	Si	En este campo de texto solo pueden aparecer valores alfanuméricos.
7	Editorial	Campo de texto	Si	En este campo solo pueden aparecer valores alfanuméricos.
8	Tema	Campo de texto	Si	En este campo solo pueden aparecer valores alfanuméricos.
9	Resumen	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
10	Sponsor	Campo de texto	Si	En este campo de texto solo pueden aparecer letras.
11	Fecha de Publicación	Campo de texto	Si	En este campo de texto solo pueden aparecer números.
12	Tesis	Lista desplegable	Si	En este campo se seleccionan tesis y brinda opciones para la selección.
13	Libro	Lista desplegable	Si	En este campo se seleccionan libros y brinda opciones para la selección.
14	Animación	Lista desplegable	Si	En este campo se seleccionan animaciones y brinda opciones para la selección.
15	Video	Lista desplegable	Si	En este campo se seleccionan videos y brinda opciones para la selección.

3.5 Clasificación de las no conformidades

La descripción de las clasificaciones de las no conformidades ayudará a que el proceso de pruebas se realice en un tiempo reducido y planificado, sin atentar contra el cronograma de desarrollo y los compromisos con el cliente. Se podrá realizar una evaluación más profunda del software a revisar, se tendrán mejores estadísticas de cuáles son las no conformidades más comunes en los diferentes tipos de software y se le podrá dar un mejor tratamiento a la no conformidad en el proceso de liberación del software (Mares Amézquita, 2007)

Según lo planteado en la Norma UNE-EN ISO 9001:2000 una no conformidad. “Es el incumplimiento de algunos de los requisitos”(Mares Amézquita, 2007)

Se trata de una desviación entre lo que hay escrito (lo que hemos dicho que vamos a hacer) y lo que ha ocurrido (lo que hemos hecho). Este fallo queda registrado en un informe y se establecen las acciones preventivas y correctivas necesarias para arreglar lo que no funcione y evitar que vuelva a ocurrir. (Fernández Pereda, 2010)

Las mismas se clasifican de acuerdo al nivel de importancia en:

1. **Las No Conformidades Significativas:** Son aquellas que afectan la calidad del producto o servicio de manera visible, impidiendo o no el cumplimiento de algún requisito.
2. **Las No Conformidades No Significativas:** Son aquellas que resultan menos visibles, que no atentan el cumplimiento de algún requisito.
3. **Las Recomendaciones:** Son aquellas que quedan en función de la apreciación del probador para oportunidades de mejoras del producto o servicio.

Las no conformidades se clasifican además en función de sus características acuerdo al tipo de artefacto como:

1. No Conformidades de Documentación:

- ◆ Formato.
- ◆ Error técnico.
- ◆ Otros errores.

- ◆ Correspondencia con otra documentación.

2. No Conformidades de Aplicación:

- ◆ Validación.
- ◆ Opciones que no funcionan.
- ◆ Errores de interfaz.
- ◆ Funcionalidad.
- ◆ Excepciones.
- ◆ Correspondencia de lo implementado con lo documentado.

3. No Conformidades comunes (para ambos artefactos):

- ◆ Ortografía.
- ◆ Redacción.
- ◆ Errores de idioma.

3.5.1 Resultados de las pruebas de caja negra

Los resultados obtenidos durante las iteraciones fueron los descritos en la Tabla 4. En la misma se pueden apreciar la cantidad de no conformidades, las que fueron solucionadas (cerradas) y las que por algún motivo justificado no pudieron ser resueltas (no proceden), obteniéndose los siguientes resultados:

Iteraciones	No Conformidades	Cerradas	No Proceden
1ra Iteración	5	5	0
2da Iteración	3	3	0
3ra Iteración	2	2	0

Tabla 2: Resumen de las No Conformidades.

Como se puede apreciar a partir de las pruebas realizadas a todos los casos de uso se obtuvieron 5 no conformidades en la primera iteración de las cuales 2 resultaron significativas, 1 no significativas y el resto recomendaciones. Para la segunda iteración se disminuyo a 1 significativas y 2 recomendaciones. Y en la tercera y última iteración 2 recomendaciones.

Conclusiones Generales

Una vez finalizada la investigación se puede afirmar que se cumplieron los objetivos propuestos, arribando a las siguientes conclusiones:

- ◆ Los Repositorios son archivos o Bases Datos donde se almacenan información digital.
- ◆ La Recuperación de la Información consiste en encontrar y establecer un orden por relevancia de los documentos que satisfacen la necesidad de información de un usuario expresada en forma de consulta.
- ◆ El análisis y diseño de las funcionalidades expuestas propicia una vía de solución que guía el proceso de desarrollo del módulo de búsqueda avanzada para el Repositorio Institucional de la Universidad de las Ciencias Informáticas.
- ◆ Con la realización de este trabajo en la UCI se logró desarrollar un módulo para el Repositorio Institucional que permite la búsqueda avanzada.
- ◆ Se obtuvieron funcionalidades para la realización de búsquedas a través de nuevos criterios de búsqueda y que permiten gestionar dichos criterios para la búsqueda de los recursos que se encuentran en el Repositorio Institucional de la Universidad de las Ciencias Informáticas a través de criterios descriptivos que permitan su rápida recuperación.

Recomendaciones

Una aplicación informática requiere de una constante actualización por lo que al concluir la investigación realizada se recomienda:

- Incorporar nuevas funcionalidades al Modulo de Búsqueda Avanzada del Repositorio Institucional de la universidad.
- Incorporar al Repositorio otro estándar de descripción de metadato diferente a Dublin Core que ya existe.

Referencias Bibliográficas

- Álvarez González, Luis Alberto. 2011.** [En línea] 07 de Diciembre de 2011. [Citado el: 12 de 2 de 2013.]
<http://www.adpt.cl:8080/jspui/bitstream/123456789/52/3/UBB0711,%20Informe%20Etapa%203%2007.12.2011.docx>.
- Flores Cuesta, Gilda y Sánchez Tar, Nancy. 2007.** ACIMED. [En línea] Diciembre de 2007. [Citado el: 12 de 2 de 2013.] http://scielo.sld.cu/scielo.php?pid=S1024-94352007001200006&script=sci_arttext.
- Acuña, Karenny Brito. 2009.** Por qué utilizar RUP para desarrollar aplicaciones web. [En línea] 2009. [Citado el: 06 de 05 de 2013.]
<http://www.eumed.net/libros/2009c/584/Por%20que%20utilizar%20RUP%20para%20desarrollar%20aplicaciones%20web.htm>.
- Alvarez Fernández, Luis Carlos. 2010.** [En línea] 2010. [Citado el: 11 de 2 de 2013.]
http://repositorio_institucional.uci.cu/jspui/handle/ident/TD_03552_10.
- Álvarez González, Luis Alberto. 2011.** *Recursos Pedagógicos y Tecnológicos*. 2011.
- Aroca Cámara, Mercedes, Checa Claudel, José y Pérez, Guzmán. 2009.** Helvia: la apuesta de la Universidad de Córdoba por el conocimiento abierto. [En línea] 2009.
<http://helvia.uco.es/xmlui/handle/10396/2344>.
- Baeza-Yates, Ricardo y Ribeiro-Neto. 1999.** *Modern Information Retrieval*. 1999.
- Barton, Mary y Waters, Margaret. 2005.** *Como crear un Repositorio Institucional*. 2005.
- Castro Mecías., Lilian Teresa y Bermúdez Peña, Rolando. 2008.** *Desarrollo de una Librería de Componentes JavaScript basado en Dojo Toolkit*. La Habana : s.n., 2008.
- Chrissis, Mary Beth y Konrad y y otros. 2009.** *CMMI. Guía para la integración de procesos y la mejora de productos*. 2009. ISBN: 9788478290963..
- Computación Aplicada al Desarrollo S.A. de C.V. . 2012.** [En línea] 2012. www.cad.com.mx.
- Date, Christopher. 2001.** *Introducción a los Sistemas de Base de Datos*. 2001.
- De Volder, Carolina. 2012.** [En línea] 2 de junio de 2012. [Citado el: 11 de 02 de 2013.]
eprints.rclis.org/bitstream/10760/17227/1/oaparaboletinABGRA.pdf.
- Fernández Pereda, Héctor. 2010.** [En línea] 2010. [Citado el: 07 de 05 de 2013.]
http://www.buscarportal.com/articulos/iso_9001_2000_gestion_calidad.html...
- Galardy Tarragó, Dianela. 2012.** *Propuesta de solución para el proceso de evaluación y control docente en el Sistema de Gestión Académica de Pregrado*. La Habana : s.n., 2012.

H. Tolosa, Gabriel y R.A. Bordignon, Fernando. 2008. *Introducción a la Recuperación de la Información. Conceptos, modelos y algoritmos básicos.* 2008.

Ivar, Jacobson, Booch, Grady y Rumbaugh, James. 2000. *El proceso unificado de desarrollo de software.* 2000.

La Rosa Agramonte, Alberto Jesús y del Castillo Rodríguez, Hiran. 2009. *Herramienta para la instalación y configuración de clústeres del contenedor servlets Apache Tomcat.* 2009.

Larman, C. 1999. *UML y Patrones.* 1999.

Librería Univesal. 1999. Librería Universal. [En línea] 1999.

Libros Antiguos. 2013. Libros Antiguos. [En línea] 2013. [Citado el: 20 de 02 de 2013.] <http://uniliber.com/quienes.php>.

Mares Amézquita, Saúl. 2007. Gestión de No Conformidades. [En línea] 2007. [Citado el: 07 de 05 de 2013.] <http://www.mex.ops-oms.org/contenido/tuberculosis/cdtaller/presentaciones/M%C3%B3dulo%20%20..>

Mesa Bles, Yordel y Perez Boniche, Ainel. 2009. *alasClínicas: Adaptación del sistema OpenClinica al control de procesos de los Ensayos Clínicos cubanos.* 2009.

OpenDOAR. 2013. OpenDOAR. [En línea] 2013. <http://www.opendoar.org/onechart.php?clID=&ctID=&rtID=&clID=&IID=&potID=&rSoftWareName=&search=&groupby=r.rSoftWareName&orderby=Tally%20DESC&charttype=pie&width=600&height=300&caption=Usage%20of%20Open%20Access%20Repository%20Software%20-%20Worldwide>.

Pressman, Roger S. 2010. *Ingeniería de Software, Un enfoque práctico.* Séptima. New York : McGraw-Hill, 2010. ISBN 978-0-07-337 597 -7.

Pressman, Rogers. 2005. *Ingeniería de Software. Un Enfoque Práctico.* 2005.

ROAR. 2013. Registro de Repositorios de Acceso Abierto. [En línea] 2013. [Citado el: 11 de 2 de 2013.] <http://roar.eprints.org>.

Salton, Gerard y Mc Gill, M.J. 1983. *Introduction to Modern Information Retrieval.* New York : s.n., 1983.

Santos Pérez, Danais y Sánchez Baños, Daris. 2008. *propuesta de la arquitectura de software para El Programa Nacional de Informa conocimiento geológico.* 2008.

Sosa Vázquez, Yailis y Ruiz Cardoso, Alberto. 2010. *Análisis y diseño de los procesos de Agregación, Notificaciones y Ajustes al Plan del subsistema de Planificación perteneciente al Sistema Integral de Gestión de Entidades Cedrux.* La Habana : s.n., 2010.

texier, Jose, y otros. 2012. *El Uso de Repositorios y su Importancia para la Educación en Ingeniería.* 2012.

The PostgreSQL Global Development Group. 2005. The PostgreSQL Global Development Group. [En línea] 2005. <http://www.postgresql.org/docs/8.0/interactive/index.html>.

Tolosa, Gabriel y R.A. Bordignon, Fernando. 2007. *Introducción a la Recuperación de Información. Conceptos, modelos y algoritmos básicos.* Buenos Aires : s.n., 2007.

Torres Valdés, Abimael y Espinosa Menéndez, Raicel. 2010. *Detección de términos relevantes en páginas Web analizando las Hojas de Estilo en Cascada para la clasificación automática de documentos en el Motor de Clasificación Inteligente por Contenido.* La Habana : s.n., 2010.

UCI. 2011. Dirección de Información. [En línea] 2011. http://repositorio_institucional.uci.cu/jspui/.

Universidad Carlos III de Madrid. 2013. [En línea] 2013. <http://e-archivo.uc3m.es:8080/>.

Universidad de Cambridge. 2013. DSpace@Cambridge. [En línea] 2013. <http://www.dspace.cam.ac.uk/>.

Universidad de Oregon. 2013. ScholarsArchive@OSU. [En línea] 2013. <http://ir.library.oregonstate.edu/xmlui/>.

Villafranca, Daniel. 2008. *Introducción a las aplicaciones Web con Java.* 2008.