

Universidad de las Ciencias Informáticas
Facultad 3



Título: Desarrollo de un módulo de Administración web para el motor de Búsqueda Orión

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

Autor: Eyeris Rodríguez Rueda

Tutor: Ing. Yusniel Hidalgo Delgado

Tutor: Ing. Abdiel Matos Nieto

Ciudad de La Habana, 17 de Junio de 2011

“Año 53 de la revolución”

“Mejor es la sabiduría que las piedras preciosas y todo cuanto se puede desear, no es de compararse con ella”

Santa Biblia, Reina Valera 60, Proverbios 8:11.

DECLARACIÓN DE AUTORÍA

Yo Eyeris Rodríguez Rueda declaro que soy el único autor de este trabajo y autorizo a la Facultad 3 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
Eyeris Rodríguez Rueda

Tutor
Ing. Abdiel Matos Nieto.

Tutor
Ing. Yusniel Hidalgo Delgado

Firma del Autor

Firma del Tutor

Firma del Tutor

Datos de Contacto

Ing. Abdiel Matos Nieto:

Graduado de Ingeniero en Ciencias Informáticas por la UCI. Web Máster del portal de la facultad 3. Posee dos años de experiencia. Sin categoría docente. Desarrollador del Centro de Gobierno Electrónico CEGEL.

Ing. Yusniel Hidalgo Delgado:


Graduado de Ingeniero en Ciencias Informáticas por la Universidad de las Ciencias Informáticas en el 2010. Actualmente se desempeña como Arquitecto de Datos del subsistema de control logístico del proyecto ERP en el Centro de Gestión de Entidades CEIGE. Es jefe del Grupo de Investigación de Inteligencia Artificial de la Facultad 3. No posee categoría docente.

AGRADECIMIENTOS

En primer lugar a Dios por permitirme cumplir este sueño.

A toda mi familia y especialmente a mi madre y amiga Anabel, también a mi hermana, a mi padrastro Pedro y a mi padre Raudel por su confianza y apoyo en los momentos en los que creía que no iba a llegar.

Agradezco a Todos mis hermanos en Cristo que muchas veces han sido instrumentos para hacerme crecer.

Un agradecimiento muy especial a mi novia Daniuska Torres Matos por ser mi ayuda y por hacerme un mejor hombre (JE AMO). 

A mis amigos y más que amigos hermanos Yasel, Boris, Yordanis y Reinier que desde 1er año cuento con su apoyo y comprensión, al igual que a todos mis compañeros de aula.

A la Revolución por permitirme cursar estudios en la UCS.

A mis tutores Yusniel y Abdiel por su ayuda y comprensión.

A todos mis compañeros de proyecto, Juan Javier y otros.

A todos los que de una forma u otra han contribuido a mi preparación como un profesional.

Muchas Gracias

DEDICATORIA

Quiero dedicar este triunfo a mi madre y amiga Anabel por ver éxitos donde muchos veían fracaso, gracias por existir, te quiero titi.

A mi hermana Amanda, que pueda servir de inspiración para ella, eres especial para mi, te quiero mucho.

A mi novia por ser la fuente de mi inspiración cuando me faltaban las fuerzas para continuar, eres mi otro yo.

A la UCS por ser el medio para mi preparación profesional.

RESUMEN

El motor de búsqueda Orión se encuentra desplegado en los servidores de la infraestructura productiva al servicio de toda la comunidad universitaria, facilitando el proceso de búsqueda y recuperación de la información existente en la red interna de la Universidad de las Ciencias Informáticas.

La forma en que se administre cualquier Sistema para la recuperación de información influirá en la calidad de los resultados obtenidos, además, este proceso no debe hacerse sin un estricto control de acceso que permita mayor control en los cambios realizados en la configuración. Actualmente el proceso de administración de Orión se realiza a través de una conexión segura usando el protocolo (SSH) usando la consola del sistema operativo GNU/Linux directamente en el servidor donde está desplegada la solución informática, lo que trae consigo un considerable gasto de tiempo y esfuerzo en dicho proceso, no se tiene un control de los cambios en esta configuración por lo que las probabilidades de ocurrencia de un error crítico en el sistema aumentan considerablemente. Para configurar este motor de búsquedas se deben modificar algunos ficheros distribuidos por todo el sistema operativo con muchas propiedades cada uno de ellos.

Con la realización de este trabajo de diploma se pretende realizar una interfaz web que facilite el proceso de administración del motor de búsqueda Orión.

PALABRAS CLAVE

Administración, interfaz, Motor de búsqueda, Orión.

ÍNDICE

AGRADECIMIENTOS	IV
DEDICATORIA	IV
RESUMEN	V
INTRODUCCIÓN	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA	6
1.1 Introducción	6
1.2 Generalidades	6
1.2.1 Información	6
1.2.2 Recuperación de Información	7
1.2.3 Sistemas de Recuperación de Información	7
1.2.4 Administración	8
1.2.5 Interfaces Gráficas	8
1.2.6 Interfaces Gráficas para la Administración Web	9
1.2.7 Herramientas existentes para la Recuperación de Información	11
1.3 Tecnologías y Herramientas	14
1.3.1 Tecnologías del lado del Cliente	15
1.3.2 Tecnologías del lado del Servidor	18
1.3.3 Entornos de Desarrollo Integrado	20
1.3.4 Herramientas CASE	21
1.3.5 Sistemas Gestores de Base de Datos	23
1.4 Lenguaje de Modelado	26
1.5 Marcos de Trabajo	27
1.6 Metodologías de Desarrollo	30
1.7 Conclusiones Parciales	33
CAPÍTULO 2: PROPUESTA DE SOLUCIÓN	34
2.1 Introducción	34

2.2	Requerimientos del Sistema	34
2.2.1	Requerimientos Funcionales	34
2.2.2	Requerimientos No Funcionales	36
2.3	Diseño del sistema	38
2.3.1	Patrones de Diseño	38
2.3.2	Prototipos de Interfaz de usuario	40
2.4	Diseño de la Base de Datos	41
2.5	Implementación del Sistema	44
2.6	Conclusiones Parciales	44
CAPÍTULO 3: EVALUACIÓN DEL SISTEMA		45
3.1	Introducción	45
3.2	Tipos de pruebas de software	45
3.2.1	Pruebas de caja Negra	46
3.2.2	Pruebas de caja Blanca	46
3.3	Casos de prueba	47
3.4	Conclusiones Parciales	52
CONCLUSIONES GENERALES		54
RECOMENDACIONES		55
REFERENCIAS BIBLIOGRÁFICAS		56
BIBLIOGRAFÍA		58
GLOSARIO DE TÉRMINOS		60
ANEXOS		62

INTRODUCCIÓN

En la segunda mitad del siglo XX como parte de un proyecto de la Agencia de Proyectos de Investigación Avanzada (ARPA por sus siglas en inglés) se crea la red ARPANET siendo esta el medio de comunicación que más trascendería a nuestros días y que sería la base de lo que hoy conocemos como Internet¹.

Creada con fines militares, ARPANET inicialmente conectó aproximadamente una comunidad de mil estudiantes. Esta nueva forma de comunicarse rápidamente se convirtió en una red que cambiaría la vida de las personas, empresas y países completos. Aproximadamente 20 años después ya esta red contaba con millones de usuarios, se abría paso a una nueva era en la informática y las telecomunicaciones.

No es precisamente el desarrollo acelerado de esta gran red la causa del asombro principal, sino su poder de penetración en la sociedad actual, su capacidad de revolucionar desde el puesto de trabajo hasta las relaciones sociales y su capacidad de mediar en casi todos los actos de la sociedad moderna. La interacción de los usuarios con esta gran red no solo ha sido como receptores de conocimiento sino también como generadores del mismo lo que ha provocado que en los últimos 20 años la información generada en esta sea imposible de consultar manualmente.

El surgimiento de internet también ha favorecido un crecimiento en el conocimiento científico y cada día son muchas las publicaciones de artículos y resultados relevantes en diferentes áreas de la ciencia. En algunas ocasiones los usuarios no saben donde buscar la información que necesitan debido a las gran cantidad de documentos existentes en la web. Para resolver esta situación se necesitan sistemas que ayuden a los usuarios en esta tarea.

Existen sistemas para la recuperación de la información en lo adelante SRI que se encargan de recorrer la red en busca de información existente en documentos publicados en sitios web y bibliotecas digitales. Una vez que estos documentos son visitados se procede a su almacenamiento en sistemas gestores de bases de datos que luego permitirán realizar búsquedas sobre ellos

¹ Internet: a veces conocida también como web o también www, siglas de World Wide Web (Red de Banda Ancha en español).

permitiéndole a los usuarios de todo el mundo un rápido acceso a gran parte de la información presente en toda la red.

Como parte de uno de los programas de la Revolución Cubana en el año 2002 surge la Universidad de las Ciencias Informáticas, en lo adelante UCI, que cuenta actualmente con aproximadamente 10 000 estudiantes y más de 5000 profesionales. Debido al desarrollo tecnológico presente en esta Universidad se realizan muchos trabajos investigativos y también eventos científicos que son publicados dentro de la misma, por lo que se cuenta con una red con más de 300 sitios web y más de un millón de documentos publicados entre los que se destacan documentos portables, texto plano, imágenes, páginas web estáticas y dinámicas como se muestra en la **Figura 1**.

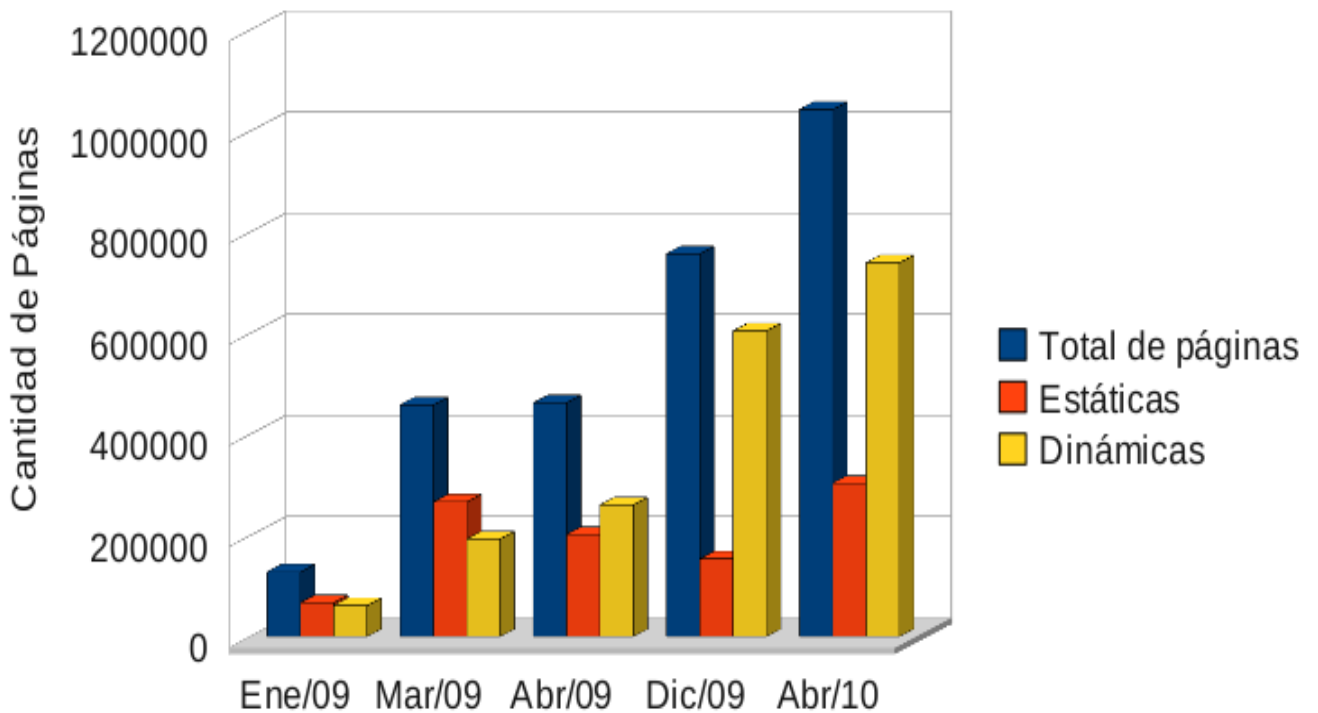


Figura 1. Resultados de los Estudios Webmétricos realizados en la UCI.

La forma en que se administre cualquier SRI influirá en la calidad de las búsquedas, además de que este proceso no debe hacerse sin un estricto control de acceso. Las Interfaces gráficas de usuarios, en lo adelante GUIs, permiten la interacción entre los sistemas informáticos y los usuarios finales de los mismos y muchas veces son usadas en el proceso de administración de estos sistemas. Aunque

siguen ligadas estrechamente a las aplicaciones de escritorio, han ido ganando un espacio en la Web², con las nuevas técnicas que esta ofrece.

El motor de búsqueda Orión se encuentra desplegado en los servidores de la infraestructura productiva al servicio de toda la comunidad universitaria, facilitando el proceso de búsqueda y recuperación de la información existente en la red interna de la Universidad de las Ciencias Informáticas. Actualmente el proceso de administración y configuración de Orión se realiza a través de la consola del sistema operativo GNU/Linux directamente en el servidor donde está desplegada la solución informática, lo que trae consigo un considerable gasto de tiempo y esfuerzo en dicho proceso, no se tiene un control de los cambios en esta configuración por lo que las probabilidades de ocurrencia de un error crítico en el sistema aumentan considerablemente. Para configurar este motor de búsquedas se deben manipular algunos ficheros distribuidos por todo el sistema operativo con muchas propiedades en cada uno de ellos. Cada una de estas opciones tiene repercusión en la calidad de las búsquedas realizadas por los usuarios y no se cuenta con una interfaz para su administración de manera que se pueda hacer este proceso desde cualquier lugar dentro de la Universidad.

Teniendo en cuenta la situación problemática anterior surge el siguiente **problema a resolver**: ¿Cómo lograr una adecuada gestión de los procesos de administración y configuración del motor de búsqueda Orión?

El **objeto de estudio** lo constituyen los Sistemas de Recuperación de Información y el **campo de acción** el motor de búsqueda Orión.

Para dar solución al problema planteado se define el **Objetivo General**: Desarrollar un sistema que permita la administración web del motor de búsquedas Orión.

Para dar cumplimiento al objetivo general se definieron las siguientes **Tareas de Investigación**.

- Investigar sobre el proceso de administración en otros motores de búsqueda.
- Estudiar el proceso de administración del motor de búsqueda Orión.

² Web, Sinónimo de Internet o WWW.

- Implementar los prototipos de Interfaz de usuario necesarios para la administración web del motor de búsqueda.
- Obtener del código fuente correspondiente al módulo de administración del motor de búsqueda Orión.
- Obtener del modelo de datos que permita almacenar todas las configuraciones gestionadas por la solución informática.
- Obtener del script de la base de datos.
- Obtener de los artefactos correspondientes al desarrollo de la solución informática.
- Realizarle pruebas al módulo de administración implementado.

Para dar solución a las tareas de investigación se utilizaron los siguientes métodos de investigación:

El método histórico – lógico: Se usa cuando se estudia el motor de búsqueda Orión desde su surgimiento, así como cada una de las etapas por las que ha transcurrido el mismo para lograr una mayor comprensión de su funcionamiento.

Modelado: Se realiza una reproducción simplificada de la realidad, mostrando las relaciones internas y características del sistema a través de diagramas. La modelación es justamente el proceso mediante el cual se crean modelos con vistas a investigar la realidad.

Implementación: Se encarga de demostrar que la solución tiene ciertas propiedades o que se comporta de una manera en específico y para ello la implementación es comparada con la de otras aplicaciones existentes. Este método es, unido al de modelado uno de los más utilizados pues será empleado durante el proceso de implementación para obtener el sistema que permita la administración de Orión.

Estructura del documento

Capítulo 1. Fundamentación teórica: En este capítulo se realiza un análisis sobre el estado del arte de los principales SRI en el mundo así como las tendencias, las técnicas y tecnologías actuales en el desarrollo de aplicaciones Web.

Capítulo 2. Propuesta de solución: En este capítulo se exponen las características del sistema, incluyendo los requerimientos funcionales y no funcionales, algunos patrones de diseño utilizados así como algunos artefactos que plantea la metodología de desarrollo utilizada.

Capítulo 3. Evaluación del Sistema. En este capítulo se obtienen los casos de prueba que se le aplican al sistema para validar la solución dada.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

1.1 Introducción

En este capítulo se tratarán algunos conceptos relacionados con los Sistemas para la Recuperación de Información, además se abordará acerca de la arquitectura de los motores de búsquedas y la utilización de las GUIs dentro del proceso de administración de los mismos. Se expondrán las tendencias actuales de desarrollo de aplicaciones web para la administración así como las tendencias, las técnicas y tecnologías actuales en el desarrollo de aplicaciones Web.

1.2 Generalidades

1.2.1 Información

Cada vez que se necesita o desea conocer algo acerca de acontecimientos, hechos, sucesos, personas, empresas, instituciones ciudades, países, (ya sea porque se necesita reducir la incertidumbre para tomar decisiones o simplemente porque se quiere incrementar el conocimiento acerca de algo), se busca aquello que se denomina «información». Y esto, es algo que se hace, en mayor o menor medida, todos los días, sin embargo muchas veces no se sabe en realidad que significa esta palabra.

La distinción entre información y documentos es importante. Siempre se necesita información, pero a veces esta se busca más directamente sin importar la organización que tenga o el lugar donde aparezca, por ejemplo si se quisiera saber sobre el lenguaje Java, no importaría si fuera en archivos de texto plano o en sitios web. Otras veces se buscan documentos concretos que se supone la contienen o que interesa en particular conocer.

Según Idalberto Chiavenato, información *"es un conjunto de datos con un significado, o sea, que reduce la incertidumbre, o que aumenta el conocimiento de algo. En verdad, la información es un mensaje con significado en un determinado contexto, disponible para uso inmediato y que proporciona orientación a las acciones por el hecho de reducir el margen de incertidumbre con respecto a nuestras decisiones"* (1).

En el Diccionario de la Real Academia Española, se encuentran, entre varios significados, los siguientes: “(Del lat. *informatio*, -ōnis). 1. f. Acción y efecto de informar. 2. f. Oficina donde se informa sobre algo. 3. f. Averiguación jurídica y legal de un hecho o delito. 4. f. Pruebas que se hacen de la calidad y circunstancias necesarias en una persona para un empleo u honor. U. m. en pl. 5. f. Comunicación o adquisición de conocimientos que permiten ampliar o precisar los que se poseen sobre una materia determinada. 6. f. Conocimientos así comunicados o adquiridos” (2).

Se Puede llegar a la conclusión que los datos por si solos no aportan información, sino que para obtenerla es necesario un procesamiento de estos como se muestra en la **Figura 2**.



Figura 2. La información como resultado del procesamiento de los datos.

1.2.2 Recuperación de Información

La Recuperación de Información o *Information Retrieval*, en lo adelante RI es una disciplina que se encarga de estudiar las técnicas para buscar información dentro de documentos que generalmente no se encuentran organizados o que debido a la cantidad de estos resulta difícil buscar manualmente dentro de los mismos. Algunos otros autores han dado su definición de lo que consideran Recuperación de Información.

Para Raquel Gómez la RI “se entiende el proceso por el cual, una vez analizado un documento e identificada la necesidad de información, se produce una comparación entre ambos para obtener resultados satisfactorios para el usuario “ (3).

Según esta definición, el documento ha de estar procesado para su correcta recuperación, o sea debe tenerse indizado en alguna Base de Datos para que el SRI pueda procesarlo y luego ordenarlo por relevancia para mostrarlo al usuario.

1.2.3 Sistemas de Recuperación de Información

Los SRI se encargan de buscar información dentro de un grupo de documentos que han sido indizados previamente. Estos sistemas no se limitan a buscar dentro de documentos solamente, sino que pueden ampliar su dominio de búsqueda a Bases de Datos y también a Repositorios Institucionales. Es importante señalar que estos sistemas pueden obtener información a partir de búsquedas dentro de archivos multimedia como pueden ser videos y música. El objetivo fundamental de este tipo de sistemas es brindar a los usuarios un conjunto de documentos ordenados por relevancia.

La primera parte del proceso de RI es realizada por un componente de los SRI llamado *Spider* o *Crawler*. Este se encarga de escanear la red en busca de sitios o documentos publicados con el fin de obtener su ruta para luego realizar el proceso de indización del contenido.

1.2.4 Administración

La administración es el proceso donde se dirige y controla el uso de los recursos para lograr algunos objetivos. En este caso un administrador de un Sistema Informático es la persona que tiene la responsabilidad de asegurar el correcto funcionamiento y mantenimiento del mismo, o algún aspecto de este. Una de las tareas más importantes que se deben tener en cuenta en cualquier proceso de administración es llevar a cabo la seguridad mediante la gestión de perfiles y controles de acceso a los sistemas informáticos.

1.2.5 Interfaces Gráficas

El término interfaz, proviene del Inglés *interface* y pudiera definirse, de forma general, como el medio que posibilita la interacción entre dos elementos para que puedan trabajar juntos. (4) Sin embargo, en la esfera de la informática, hay que destacar que existen diversos tipos de interfaces que trabajan a distintos niveles. Existen las interfaces que hacen posible el trabajo del software con el hardware, es decir, que posibilita a los sistemas operativos el funcionamiento con los dispositivos de hardware, tanto interno como externos. También existe la interfaz de línea de comandos para interpretar las sentencias introducidas por el usuario, la interfaz controlada por menús, que le brinda al usuario mayores funcionalidades para la interpretación de combinaciones de teclas en el trabajo con la aplicación y la interfaz gráfica de usuario, en lo adelante GUI, que además de controlar las opciones

de los menús, también posibilita controlar la posición, el tamaño y el contenido de las áreas de trabajo, en pantalla. (4)

Algunos autores han dado otros conceptos de lo que ellos consideran GUI a continuación la definición de Lewis y Rieman *“Las interfaces gráficas de usuario son aquellas que incluyen cosas como menús, ventanas, teclado, ratón y algunos otros sonidos que la computadora hace, en general, todos aquellos canales por los cuales se permite la comunicación entre el hombre y la computadora.”* (5)

1.2.6 Interfaces Gráficas para la Administración Web

La aparición de la Web hizo posible que cualquier persona pueda ofrecer información particularizada a los demás y encontrar documentos interactivos sobre cualquier tema, relacionados unos con otros mediante enlaces que permiten navegar de página en página alrededor del mundo.

Las páginas Web impulsaron la aparición de las interfaces Web, interfaces gráficas de usuario con elementos comunes de presentación y navegación que pronto se convirtieron en estándares. Este tipo de interfaces sirven de intermediarias entre usuarios genéricos, no acostumbrados generalmente al uso de aplicaciones informáticas, y unos sistemas de información y procesos transaccionales que son transparentes para el usuario, posibilitando la localización de información deseada, el entendimiento claro de las funcionalidades ofrecidas, la realización práctica de tareas específicas por parte de los usuarios y la navegación intuitiva por las diferentes páginas que forman la aplicación Web (5).

Una interfaz de administración web es considerada también una aplicación web, que no es más que aquella aplicación que se encuentran en un servidor, a la cual los usuarios pueden acceder a través de Internet o de una Intranet mediante un navegador. Estas aplicaciones se han hecho muy populares debido a lo práctico que se ha vuelto el uso de un navegador web y a las facilidades de actualizarlas y mantenerlas sin la necesidad de distribuir las e instalarlas en miles de usuarios potenciales.

Es importante mencionar que una página web puede contener elementos que permiten una comunicación activa entre el usuario y la información. Esto permite que el usuario acceda a los datos

de modo interactivo, gracias a que la página responderá a cada una de sus acciones, como por ejemplo rellenar y enviar formularios, participar en juegos y acceder a gestores de base de datos de todo tipo, lo cual puede ser más seguro y configurable gracias a la estructura que normalmente debe tener una aplicación web.

Una Aplicación Web es una interfaz diseñada para cubrir las necesidades de un negocio y gestionar su información (la información puede ser de dominio público o restringido a ciertas personas a través de un nombre de usuario y contraseña) con el objetivo de que cualquier persona pueda consultarla e interactuar con ella a través de la red. Se pueden adaptar a muchas situaciones, su objetivo es mejorar la forma de trabajo y la productividad de una empresa o grupo de personas de una manera sencilla.

Beneficios

- Permiten reunir las diferentes áreas de una entidad.
- Tener mayor control de datos y mejor seguridad en las diferentes secciones del sitio web.
- Permite tener información siempre actualizada.
- Otorga la flexibilidad de determinar niveles de acceso según la confidencialidad de los datos así como la posibilidad de realizar transacciones en línea.

Ventajas de una aplicación web

- Ahorra tiempo: Se pueden realizar diferentes tareas, sin necesidad de descargar ni instalar ningún programa.
- Se reducen los problemas de compatibilidad, solo basta tener una computadora con mínimos recursos para poder utilizarlas.
- No ocupan mucho espacio en el disco duro.
- Actualizaciones inmediatas: Al conectarse se está utilizando siempre la última versión que se haya lanzado.
- Consumo de recursos bajo: Dado que la aplicación se encuentra instalada en otro ordenador, muchas de las tareas que realiza el software no consumen recursos de la computadora que se está usando.
- Multiplataforma: No importa el sistema operativo para que esta funcione correctamente.

- Portables: Son independiente de la computadora donde se utilicen, sólo es necesario disponer de acceso a Internet o a una red local.
- La disponibilidad suele ser alta: Porque el servicio se ofrece desde múltiples localizaciones para asegurar la continuidad del mismo.
- Los virus no dañan los datos: Estos están guardados en el servidor de la aplicación.

1.2.7 Herramientas existentes para la Recuperación de Información

A continuación se realizará un análisis de algunos SRI existentes, así como algunas de sus características más importantes en una tabla comparativa, mostrando sus ventajas y desventajas, **Ver Tabla 1.**

Indri

El motor de búsqueda Indri forma parte del proyecto Lemur³ comenzado en el año 2000 por el Centro de Inteligencia para la Recuperación de la Información de la Universidad de Massachusetts. Indri fue desarrollado conjuntamente entre la Universidad de Massachusetts y la Universidad de Carnegie. Este proyecto fue terminado en el año 2004 y se usa generalmente para entornos académicos aunque puede ser usado para pequeñas y medianas empresas (6).

Apache Lucene

Lucene es una novedosa herramienta que permite tanto la indización como la búsqueda de documentos. Creada bajo una metodología orientada a objetos e implementada completamente en Java, no se trata de una aplicación que pueda ser descargada, instalada y ejecutada sino de una API⁴ flexible, muy potente y fácil de utilizar, a través de la cual se pueden añadir, con pocos esfuerzos de programación, capacidades de indización y búsqueda a cualquier sistema que se esté desarrollando.

³ <http://www.lemurproject.org/>

⁴ API: Acrónimo de *Interfaz de Programación de Aplicaciones* (*Application Programming Interface* en inglés)

Originalmente escrita por Doug Cutting, en Septiembre de 2001 pasó a formar parte de la familia de Apache. A partir del año 2005 este SRI tomó gran auge en el mundo y comenzó a ser usado en algunos sitios y aplicaciones (6).

Este motor presenta una interfaz de configuración como se muestra en la **Figura 3**.

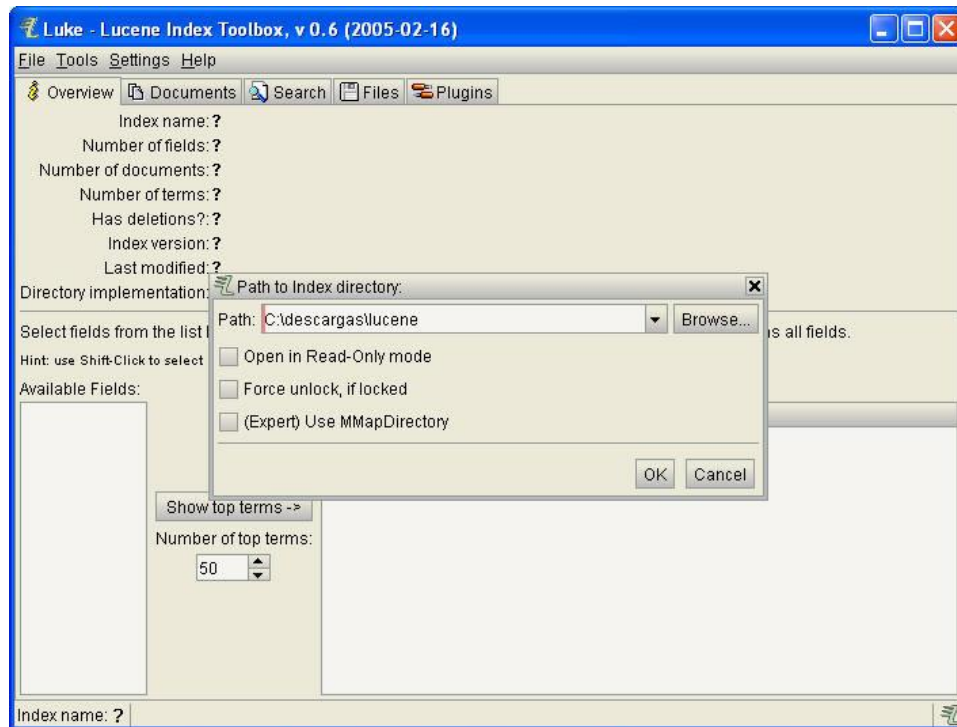


Figura 3. Interfaz de configuración de Apache Lucene.

Minion

Este Sistema está diseñado para ser altamente configurable y se propone que sea usado en investigaciones así como en entornos empresariales. Presenta indización y recuperación concurrente. En mayo de 2008 fue publicado su código fuente y la versión desarrollada hasta el momento es parte del proyecto de Tecnologías para la Búsqueda Avanzada en los laboratorios de la compañía Sun Microsystem/Oracle. Actualmente es utilizado en el sistema Aura, desarrollado en los laboratorios de Sun así como en aplicaciones web de la misma compañía (6).

Terrier

Se comenzó a desarrollar en el año 2001 en la Universidad de Glasgow, Escocia, y ha estado disponible al público desde noviembre de 2004. Su nombre es un acrónimo de **TER**abyte **reTR**ieV**ER**. Permite realizar experimentación con fuentes experimentales de gran escala. Desde la publicación de su API, alrededor de 40 instituciones académicas y empresariales manifiestan interés en su utilización y mejora continua (6).

MnoGosearch

MnoGoSearch es un motor de búsqueda completo de código abierto y basado en SQL. MnoGoSearch consiste en dos partes. La primera parte es un mecanismo de indización (indexer) el cual se mueve a través de vínculos de hipertexto HTML y almacena información acerca de los documentos en la base de datos. La segunda parte es una interfaz web CGI⁵ -search.cgi- la cual muestra en el navegador un formulario HTML y los resultados de búsquedas. search.cgi utiliza información recopilada por el indizador (7).

Entre sus principales características destacan:

- Soporte para diversos protocolos: HTTP, HTTPS, FTP, NNTP.
- Analizadores incorporados para diversos formatos de archivo: text/html, text/xml, text/plain y audio/mpeg.
- Soporte para autenticación de proxy.
- Indización multihilo.
- Interfaces web CGI, Perl y PHP.
- Lenguaje de consulta booleano.
- Soporte para la mayoría de los conjuntos de caracteres modernos.
- Soporte para múltiples bases de datos: MySQL, PostgreSQL, SQLite, Mimer, Virtuoso, Interbase, Oracle, MS SQL, DB2, Sysbase.
- Posee una API externa para PHP.
- Configuración a través de ficheros.
- Manejo de clústeres de base de datos.

⁵ CGI: Acrónimo de Interfaz de entrada común (*Common Gateway Interface en inglés*).

Características	Apache Lucene	Minion	Terrier	MnoGoSearch
Estructura del Índice	Índice invertido	Índice invertido	Índice invertido	Índice invertido
Índice Incremental	Si	Si	No	Si
Documentos que indiza	HTML, TXT, PDF, XML, PPT, DOC, RTF	Texto Plano	HTML, TXT, PDF, XML, PPT, DOC, RTF y Fuentes Documentales TREC	HTML, TXT, PDF, XML, PPT, DOC, RTF, JPG, GIF, PNG, ODT
Modelo RI	Combinación del Espacio Vectorial y el Booleano	Combinación del modelo booleano, consultas relacionales y de proximidad	Divergencia de aleatoriedad, basado en el modelo probabilístico	Combinación del Espacio Vectorial y el Booleano
Operadores de Búsqueda	Booleanos, frases, términos wildcard, campos.	Booleanos, Proximidad, Pasajes y campos	Booleanos, Operadores +/-, frases, Proximidad y campos	Booleanos, Operadores +/-, frases, Proximidad y campos
Lenguajes de Programación	Java, C++, C#, Pyton, Perl, Ruby, Php	Java	Java	Java, C++, C#, Pyton, Perl, Ruby, Php
Configuración	Interfaz	Interfaz(En desarrollo)	Interfaz	Ficheros de Configuración
Lematización	Si	Si	Si	Si
Plataforma	Windows, Mac Os, Gnu/Linux, Unix	Windows, Mac Os, Gnu/Linux, Unix	Windows, Mac Os, Gnu/Linux, Unix	Windows, Unix Gnu/Linux.
Licencia	Apache Software Foundation(ASF)	GNU Public licence GPL	Mozilla Software Foundation(MSF)	GNU Public licence GPL

Tabla 1. Comparativa entre estos Motores de Búsquedas (6)

1.3 Tecnologías y Herramientas

Una de las características más importantes de las aplicaciones web es que generalmente el cliente no tiene que instalar nada para acceder al sistema, solo un navegador. Generalmente el uso de este tipo de sistema se divide en dos partes, la parte del cliente y la parte del servidor. Cada una de estas partes tiene sus características y tecnologías, el servidor se encarga de gestionar toda la parte del negocio y el acceso a datos y para garantizar esto se requieren de algunas tecnologías llamadas tecnologías del lado del servidor. En cambio las tecnologías del lado de cliente se encargan de la

parte de la vista, o sea, la apariencia de las páginas y las validaciones de los datos de entrada, entre otras funciones. A continuación se abordarán algunas de las tecnologías más frecuentes.

1.3.1 Tecnologías del lado del Cliente

La programación del lado del Cliente se basa fundamentalmente en la rapidez que brinda a los usuarios en comparación con la programación del lado del Servidor, se validan posibles errores antes que los datos sean enviados al servidor y de esta forma se gana en consistencia y velocidad, aunque es siempre recomendable hacer validaciones en el servidor. La programación del lado del Cliente depende del navegador o explorador Web, si existe una mínima alteración la aplicación puede no recuperarse al cambio y colapsar.

JavaScript

JavaScript en lo adelante JS, es un lenguaje de programación script que se utiliza fundamentalmente para crear páginas Web dinámicas. Es un lenguaje interpretado, por lo que no es necesario compilar los programas para ejecutarlos. Los programas escritos en JS se pueden probar directamente en cualquier navegador sin necesidad de utilizar herramientas de compilación o cualquier proceso intermediario, solo basta con incluir las etiquetas correspondientes y realizar algunas funciones dentro de nuestra página HTML o PHP y estas funciones estarán accesibles para los usuarios.

¿Qué es una página Web dinámica?

Es una página Web que implementa funcionalidades, acciones que se ejecutan al pulsar botones u otros elementos HTML, operaciones sencillas o complejas que luego son mostradas al usuario, así como mensajes de aviso, efectos visuales o animaciones. JS es un lenguaje interpretado, inspirado en Java, que se incluye en los documentos HTML para añadir cierta interactividad a sus contenidos, evitando tener que realizar programación en el servidor. Es un lenguaje de sintaxis similar a Java, en cuanto a tipos de datos y estructuras de control; sin embargo, al no ser compilado, realiza un control de tipos menos estricto. Por ejemplo, no es necesario declarar las variables, y su tipo puede cambiar durante la ejecución del programa. Todas las referencias entre objetos se resuelven en tiempo de ejecución; Java lo realiza en tiempo de compilación. Utiliza un gestor automático de memoria dinámica, que reserva espacio para crear objetos y los elimina cuando éstos ya no se utilizan llamado Recolector de Basura (*Garbage Collector en inglés*). Está basado en un conjunto predefinido de objetos, que pueden ser extendidos. Sin embargo, no es posible crear nuevas clases,

o establecer relaciones de herencia. Permite utilizar funciones, al estilo de los lenguajes de programación orientados a procedimientos.

AJAX

AJAX⁶ no es una tecnología en sí mismo. En realidad, se trata de la unión de varias tecnologías que se desarrollan de forma autónoma y que se unen de formas nuevas y sorprendentes (8).

Las tecnologías que forman AJAX son:

- XHTML y CSS, para crear una presentación basada en estándares.
- DOM, para la interacción y manipulación dinámica de la presentación.
- XML, XSLT y JSON, para el intercambio y manipulación de información.
- XMLHttpRequest, para el intercambio asíncrono de información.
- JS, para unir todas las demás tecnologías.

Para el desarrollo de aplicaciones AJAX se requiere de un conocimiento avanzado de todas y cada una de las tecnologías anteriores. Pero ¿cómo funciona una aplicación AJAX?

En una aplicación Web tradicional las acciones del usuario en la página desencadenan llamadas al servidor. Una vez procesada la petición del usuario, el servidor devuelve una página HTML al navegador del usuario. Esta técnica tradicional para desarrollar aplicaciones Web funciona correctamente, pero no muestra una buena sensación al usuario. Cuando se realizan peticiones continuas al servidor el usuario tiene que esperar que se recargue la página con los cambios solicitados, **Ver Figura 4.**

AJAX se ha desarrollado para mejorar la interacción entre el usuario y la aplicación, evitando las constantes recargas de la página, debido a que el intercambio de información con el servidor se produce en un segundo plano. Este segundo plano es la creación de un elemento intermedio, una nueva capa que mejora la respuesta de la aplicación, y que no permite que el usuario encuentre una ventana del navegador en blanco, esperando la respuesta del servidor.

⁶ AJAX es un acrónimo de Asynchronous JavaScript + XML, que se puede traducir como “JavaScript Asíncrono + XML”.

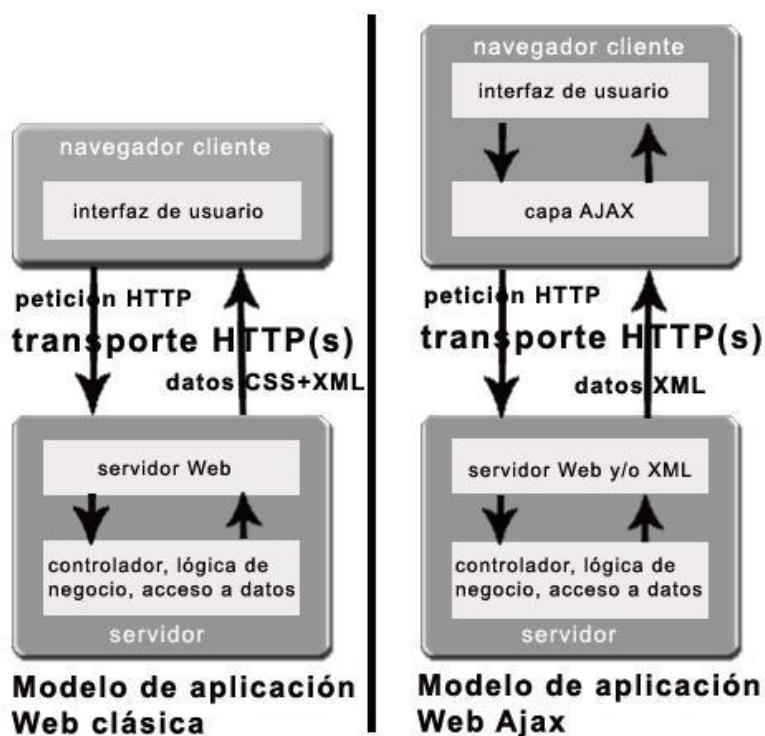


Figura 4. A la izquierda muestra el modelo tradicional de las aplicaciones Web. A la derecha muestra el modelo de AJAX.

CSS

CSS⁷ es un lenguaje de hojas de estilos creado para controlar la presentación de los documentos electrónicos definidos con HTML⁸ y XHTML. CSS es la mejor forma de separar los contenidos y su presentación y es imprescindible para la creación de páginas web complejas. El W3C⁹ es el encargado de formular la especificación de las hojas de estilo que servirán de estándar para los navegadores.

La separación de los contenidos y su presentación tiene numerosas ventajas, ya que obliga a crear documentos HTML/XHTML bien definidos y con significado completo (también llamados “documentos semánticos”). Además, mejora la accesibilidad del documento, reduce la complejidad

⁷ CSS: Siglas de Cascade Style Sheet (Hojas de Estilos en Cascada en español).

⁸ HTML: Acrónimo de Hyper Text Markup Language

⁹ Acrónimo de World Wide Web Consortium.

de su mantenimiento y permite visualizar el mismo documento en infinidad de dispositivos diferentes. (9)

Este lenguaje permite a los desarrolladores crear hojas de estilo que pueden ser usadas en las páginas HTML. Tradicionalmente se usaba dentro de las páginas HTML un estilo particular para cada etiqueta, esto traía la desventaja de que cuando la página HTML incluía muchas etiquetas había que incluir cada estilo por separado, y si se necesitaba modificar la forma en que se veían la cabeceras por ejemplo, había que modificar todas las entradas de esta etiqueta en nuestra página, por lo que el proceso se hacía muy engorroso. Con la llegada del CSS se logró separar y agrupar los estilos en documentos por separado, ahora se le puede decir a las cabeceras que usarán un estilo de nuestra hoja de CSS y si se desea modificar la forma en que se verán en nuestra página solo se tiene que modificar la hoja mencionada anteriormente y todas las cabeceras cambiarán su estilo simultáneamente.

1.3.2 Tecnologías del lado del Servidor

PHP

Constituye un lenguaje script de alto nivel interpretado del lado del servidor. Es un lenguaje de programación (originario del nombre PHP Hypertext Preprocessor) que sirve principalmente para proporcionar características dinámicas a una página Web. Al ser ejecutado del lado del servidor, PHP permite acceder a los recursos internos del mismo o a otros externos, como por ejemplo a una base de datos, siendo el resultado normalmente una página HTML con los datos y acciones enviadas al cliente.

Ventajas

- Es un lenguaje multiplataforma.
- Completamente orientado al desarrollo de aplicaciones web dinámicas con acceso a información almacenada en una Base de Datos.
- El código fuente escrito en PHP es invisible al navegador y al cliente, es el servidor el que se encarga de ejecutar el código y enviar su resultado HTML al navegador. Esto hace que la programación en PHP sea segura y confiable.

- Capacidad de conexión con la mayoría de los motores de base de datos que se utilizan en la actualidad, destaca su conectividad con MySQL y PostgreSQL.
- Posee una amplia documentación en su página oficial, entre la cual se destaca que todas las funciones del sistema están explicadas y ejemplificadas en un único archivo de ayuda.
- Es libre, por lo que se presenta como una alternativa de fácil acceso para todos.
- Permite aplicar técnicas de programación orientada a objetos.
- Biblioteca nativa de funciones sumamente amplia e incluida.
- No requiere definición de tipos de variables aunque sus variables se pueden evaluar también por el tipo que estén manejando en tiempo de ejecución.
- Tiene manejo de excepciones (desde PHP5).

En los últimos años, PHP ha crecido hasta ser el más generalizado en Internet, funcionando en más de un tercio de los servidores web a través del mundo. PHP ha tenido un crecimiento no solo cuantitativo sino también cualitativo. Cada vez más empresas, se basan en PHP para administrar sus aplicaciones críticas (10).

PHP5 ha mantenido el liderazgo en la web. PHP5 no solo revoluciona el modelo orientado a objetos sino que también contiene muchas nuevas características que lo convierten en un lenguaje mucho más maduro (10).

La llegada de PHP5 vino emparejada de una reestructuración del Core de PHP, lo que los creadores de PHP llaman Zend Engine3. Así, como el lejano PHP3 incluye su Zend Engine 0.5, y PHP4 el Zend Engine 1.0, tenemos Zend Engine 2.0 en PHP5. El cambio de versión no fue trivial, incluye la reescritura casi total del modelo de objetos, entre sus cambios más sustanciales (11).

La programación orientada a objetos y todos sus estándares está siendo completada, ya están añadidos los modificadores de acceso a los métodos, public/private/protected, nombre unificado para el constructor de la clase, soporte de un destructor de la clase con un nombre genérico, soporte de Interface, que le da la posibilidad a las clases de tener varias relaciones, nuevo operador para conocer el tipo de una variable, el paso de parámetros por referencia de forma implícita, declaración de métodos abstractos, entre otras (10).

Se usará la versión 5.3.3 para el desarrollo de este módulo de administración.

1.3.3 Entornos de Desarrollo Integrado

Un IDE¹⁰ es un programa compuesto por un conjunto de herramientas que facilitan el desarrollo de aplicaciones WEB, puede dedicarse en exclusiva a un solo lenguaje de programación o bien, poder utilizarse para varios. El entorno de desarrollo es imprescindible en la producción de un software. Es donde se definen el conjunto de herramientas, tecnologías y versiones a usar que intervienen en un proceso de desarrollo del software. A continuación se presentan un conjunto de herramientas utilizadas en el desarrollo de este trabajo, se exponen sus características y ventajas para poder tener un mayor conocimiento de su utilidad.

Eclipse

Eclipse es un IDE para todo tipo de aplicaciones, entiéndase web o de escritorio, inicialmente desarrollado por IBM, y actualmente gestionado por la Fundación Eclipse. La característica clave de este IDE es la extensibilidad. Presenta una estructura formada por un núcleo y muchos plugins que van conformando la funcionalidad final. La forma en que los plugins interactúan es mediante interfaces o puntos de extensión, así los nuevos aportes se integran sin dificultad ni conflictos.

Entre las principales características de este IDE se encuentran:

- Multiplataforma.
- Soporte para distintas arquitecturas.
- Resaltado de sintaxis.
- Auto completado.
- Tabulador de un bloque de código seleccionado.
- Asistentes (wizards): para la creación, exportación e importación de proyectos y para generar plantillas de códigos (templates).

NetBeans

NetBeans IDE es una herramienta para programadores de código abierto escrito en Java usando la plataforma NetBeans. Soporta el desarrollo de todos los tipos de aplicación Java, tales como: JEE

¹⁰ IDE: Acrónimo de Integrated Development Environment (Entorno de Desarrollo Integrado en español).

(*Java Platform, Enterprise Edition*), Web, EJB (*Enterprise JavaBeans*) y aplicaciones móviles. NetBeans provee una estructura para los proyectos que podemos crear, propone un esqueleto para organizar el código fuente, el editor conjuntamente integra los lenguajes como HTML, JS y CSS. Este IDE posee un sistema para examinar todo los directorios de cada proyecto, haciendo reconocimiento y carga de clases, métodos y objetos, para acelerar la programación.

La plataforma NetBeans es una base modular y extensible usada como una estructura de integración para crear aplicaciones de escritorio grandes. Ofrece servicios comunes permitiéndole al desarrollador enfocarse en la lógica específica de su aplicación. Entre las principales características de la plataforma se encuentran:

- Extensible mediante plugins.
- Administración de las interfaces de usuario.
- Administración de las configuraciones del usuario.
- Integración con repositorios y Base de Datos.
- Integración con los proyectos creados en Eclipse.
- Administración de almacenamiento.
- Administración de ventanas.
- Marco de trabajo basado en asistentes.
- Soporte al marco de trabajo Symfony usado en este trabajo.

Teniendo en cuenta todas las potencialidades que brinda este IDE se decidió la utilización de la versión 6.9 en el desarrollo de este módulo de administración.

1.3.4 Herramientas CASE

Las herramientas CASE¹¹ 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 de dinero. Estas herramientas nos pueden ayudar en todos los aspectos del ciclo de vida de desarrollo del software en tareas como el proceso de realizar un diseño del proyecto, cálculo de costes, implementación de parte del código automáticamente con el diseño dado, compilación automática, documentación o detección de errores entre otras (12).

¹¹ CASE: Acrónimo de Computer Aided Software Engineering (Ingeniería de Software Asistido por Computadoras).

Visual Paradigm

Visual Paradigm es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. El software de modelado UML ayuda a una más rápida construcción de aplicaciones de calidad, mejores y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. La herramienta UML CASE también proporciona abundantes tutoriales de UML, demostraciones interactivas de UML y proyectos UML. Brinda la posibilidad de generar código a partir de los diagramas, para plataformas como .Net, Java y PHP, así como obtener diagramas a partir de código. Esta es precisamente una gran ventaja puesto que el sistema será desarrollado en PHP.

Tiene disponibilidad en múltiples plataformas y en múltiples versiones. Esta característica es muy importante pues por ejemplo el Rational Rose, es una herramienta muy recomendada y además profesional, tiene una desventaja en su contra pues obliga al usuario a desarrollar en máquinas con el sistema operativo Windows, mientras que el Visual Paradigm está disponible para varios sistemas operativos como Windows, Linux, Unix. Teniendo en cuenta estas características será usado en su versión 6.4, para modelar los artefactos generados.

Rational Rose

Es una herramienta CASE que cubre todo el ciclo de ayuda de un proyecto: concepción y formalización del modelo, construcción de los componentes y certificación de las distintas fases. Permite trazabilidad real entre modelo (análisis y diseño) y el código ejecutable.

Sus principales características son:

- Mantiene la consistencia de los modelos del sistema software.
- Diseño centrado en casos de uso y enfocado al negocio que genera un software de mayor calidad.
- Chequeo de la sintaxis UML.
- Genera documentación automáticamente.
- Generación de código a partir de los modelos.

- Disponibilidad en múltiples plataformas.
- Capacidades de Ingeniería Inversa.

Rational Rose es una herramienta que se puede encuadrar dentro del grupo de herramientas más técnicas debido a que se encarga de llevar a cabo la automatización de los sistemas para la posterior generación de código. Entre las diferentes ventajas que ofrece el Rational Rose, la más importante es que utiliza la notación estándar en la arquitectura de software UML, permitiendo a los desarrolladores y arquitectos de software utilizar un lenguaje común para poder visualizar el sistema completo. También pueden los diseñadores modelar sus componentes e interfaces en forma individual y luego unirlos con otros componentes del proyecto. Una de las desventajas de esta herramienta es que es privativa lo que en muchas ocasiones limita el uso de esta en la realización de los artefactos de algunos proyectos.

Embarcadero Erwin Studio

Es una herramienta para la arquitectura y modelado de Base de Datos, en lo adelante BD. Provee de mejores reportes y mejor soporte para la implementación de prácticas y estándares en modelos complejos que contienen metadatos críticos. Mejora la calidad del modelo de datos y reduce los riesgos asociados a la seguridad de la información. Es una herramienta multinivel destinada para la construcción del modelo físico y lógico. Está equipado para crear y manejar diseños de BD grandes y complejos. Cuenta con grandes capacidades de diseño lógico, construcción automática de la BD, permite realizar ingeniería inversa partiendo del modelo físico al modelo lógico. Brinda mayor soporte en la integración de almacenamiento de datos diseñados para visualizar, documentar y compartir el conocimiento. Soporta como sistemas de gestión de BD a Oracle, Sybase System, Microsoft SQL Server, IBM D B/2 Universal, InterBase 4, Microsoft Access y Microsoft Visual Fox Pro (13).

1.3.5 Sistemas Gestores de Base de Datos

Se define un Sistema Gestor de Bases de Datos en lo adelante SGBD, también llamado DBMS¹² como una colección de datos relacionados entre sí, estructurados y organizados, y un conjunto de programas que acceden y gestionan esos datos (14).

En efecto un SGBD es una aplicación constituida por datos donde los usuarios definen, crean y mantienen la información coleccionada dentro de una BD y se gestiona por programas específicos.

Un Sistema de Gestión de Base de Datos se divide en módulos que tratan cada una de las responsabilidades del sistema general. Los componentes funcionales de un SGBD incluyen:

- **Procesador de Consultas:** Traduce sentencias en un lenguaje de consultas a instrucciones de bajo nivel que entiende el gestor de la BD.
- **Gestor de la Base de Datos:** Proporciona la interfaz entre los datos de bajo nivel almacenados en la BD, los programas de aplicación y las consultas que se hacen en el sistema.
- **Gestor de Archivos.** Gestiona la asignación de espacio en la memoria del disco y de las estructuras de datos usadas para representar la información almacenada en disco.
- **Pre compilador del Lenguaje de Manipulación de Datos DML:** Convierte las sentencias en DML¹³ incorporadas en un programa de aplicación en llamadas normales a procedimientos en el lenguaje principal.
- **Compilador del Lenguaje de Definición de Datos DDL:** Convierte sentencias en DDL¹⁴ en un conjunto de tablas metadatos o "datos sobre datos".
- **Gestor del Diccionario de Datos:** Almacena metadatos sobre la estructura de la BD (15).

Los SGBD se desglosan en:

- SGBD Libres.
- SGBD Proprietarios.

¹² DBMS: Acrónimo de Data Base Management System (Sistema de administración de Base de Datos en español.)

¹³ DML: Acrónimo de Data Manipulation Language (Lenguaje de Manipulación de Datos en español).

¹⁴ DDL: Acrónimo de Data Definition Language (Lenguaje de Definición de Datos en español).

Dentro de estos dos grupos se analizan brevemente:

Oracle

Es uno de los manejadores de BD más potente que existe actualmente. Brinda las funcionalidades básicas de un gestor sin olvidar las opciones de sub-consultas, no es lento al ser ejecutadas las mismas. Es muy vendido a nivel mundial, y su alto precio es debido a que Oracle es un manejador de BD muy poderoso y solo está al alcance de medianas o grandes empresas como las multinacionales. Consume gran cantidad de recursos en el servidor que trae consigo que se necesite de un buen sistema de hardware, RAM, microprocesador, disco duro. Es capaz de almacenar gran cantidad de información, existen servidores que gestionan hasta muchos terabytes de capacidad. En el desarrollo de aplicaciones web no está tan extendido debido a que es un sistema muy caro a diferencias de otros sistemas de gestión de BD como MySQL, SQL Server. Oracle se basa en tecnología cliente-servidor, cuenta con herramientas necesarias para su utilización en el servidor, así como con herramientas de programación básicas de Oracle (13)

MySQL

Es un sistema de gestión de BD relacional, cuyo principal objetivo de creación fue la velocidad. Cuenta con una licencia para uso en software privativos, de esta forma las empresas que deseen adquirirla para el desarrollo de software privados es necesario comprarla con este fin. El MySQL es propietario, pertenece a la empresa ORACLE Su diseño multi-hilo y multiusuario le permite una gran carga de trabajo, procesando gran cantidad de información.

Es una de las BD más populares de Internet. Es usada actualmente por muchas empresas debido a su fiabilidad y facilidad de uso, por lo que persigue cubrir todas las necesidades del cliente, principalmente la velocidad. Funciona sobre varias plataformas, permitiendo su uso en disímiles lenguajes de programación al acceder a las BD, este gestor cuenta con una gran cantidad de librerías y herramientas que alcanzan una mayor popularidad entre los sistemas de su tipo. Es desarrollado cumpliendo con el estándar SQL, logrando así una gran fiabilidad por los desarrolladores (13).

Postgres

Es un servidor de BD relacional de código abierto orientado a objetos, publicado bajo la licencia BSD y creado en software libre por una comunidad de desarrolladores y organizaciones comerciales. PostgreSQL permite que mientras unos usuarios accedan a una tabla otros pueden trabajar en ella sin necesidad de bloqueos. Cuenta con un entorno administrativo amigable y fácil de usar. Es un sistema objeto-relacional, ya que incluye la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional.

Es el sistema libre más avanzado, soporta la gran mayoría de las transacciones SQL. Trabaja con varios lenguajes de programación como son: C, C++, Java, Python, PHP.

Ventajas

- Soporta múltiples tipos de datos: como los tipos de fecha, monetarios, elementos gráficos, datos de redes y también creación de tipos de datos propios.
- Permite la creación de funciones personalizadas y el manejo y configuración de disparadores.
- Posee una gran escalabilidad.
- Es capaz de ajustarse al número de CPU y la cantidad de memoria que ofrece el sistema de forma óptima, siendo capaz de soportar mayor cantidad de peticiones simultáneas sin provocar errores.
- Comprueba la integridad referencial y almacena procedimientos en su propia BD (13).

Teniendo en cuenta las características que presenta este SGBD además de la integración que tiene con el marco de trabajo Symfony y Ext JS, se decidió usar para la implementación de este módulo de administración en su versión 8.2.4.

1.4 Lenguaje de Modelado

UML

Un lenguaje de modelado es un lenguaje cuyo vocabulario y reglas se centran en la representación conceptual y física de un sistema, por lo tanto, es un lenguaje estándar para los planos del software (16).

El Lenguaje Unificado de Modelado (Unified Modeling Language, UML), es un lenguaje estándar para escribir planos de software. UML puede utilizarse para visualizar, especificar, construir y documentar los artefactos de un sistema que involucra una gran cantidad de software. UML cubre la documentación de la arquitectura y proporciona requisitos, pruebas y las actividades de planificación de proyectos (16).

Entre las principales características de UML se encuentran:

- El sistema de software es diseñado y documentado antes de que sea codificado.
- Los lógicos “agujeros” en la etapa de diseño podrán ser detectados con anterioridad. El software se comportará de la forma esperada y surgirán menos imprevistos.
- El diseño total del sistema dicta el modo en que se desarrollará el software. Las decisiones finales se harán antes de que se encuentre código mal escrito.

Cualquier modificación en el sistema será más fácil de llevar a cabo sobre la documentación UML (15).

1.5 Marcos de Trabajo

Marco de trabajo (Framework): no es más que una estructura de soporte definida mediante la cual un proyecto de software puede ser desarrollado y organizado. Pueden incluir soporte de programas, bibliotecas y lenguajes interpretados entre otros software para ayudar a desarrollar y unir los diferentes componentes de un proyecto. Representa una arquitectura de software que modela las relaciones generales de las entidades del dominio. Provee una estructura y una metodología de trabajo la cual extiende o utiliza las aplicaciones del dominio. Son diseñados con el intento de facilitar el desarrollo de software, permitiendo a los diseñadores y programadores pasar más tiempo identificando requerimientos de software que tratando con los tediosos detalles de bajo nivel de proveer un sistema funcional. Es el esqueleto sobre el cual varios objetos son integrados para una solución dada. No es más que una base de programación que atiende a sus descendientes (manejado de una forma estructural y/o en cascada) posibilitando cualquier respuesta ante las necesidades de sus miembros, o secciones de una aplicación web.

Symfony

Symfony es un completo marco de trabajo diseñado para optimizar, gracias a sus características, el desarrollo de las aplicaciones web. Para empezar, separa la lógica de negocio, la lógica de servidor y la presentación de la aplicación web. Proporciona varias herramientas y clases encaminadas a reducir el tiempo de desarrollo de una aplicación web compleja. Además, automatiza las tareas más comunes, permitiendo al desarrollador dedicarse por completo a los aspectos específicos de cada aplicación. El resultado de todas estas ventajas es que no se debe reinventar la rueda cada vez que se crea una nueva aplicación web (17).

Características de Symfony:

- Fácil de instalar y configurar en la mayoría de las plataformas (y con la garantía de que funciona correctamente en los sistemas Windows y *nix estándares).
- Independiente del sistema gestor de bases de datos.
- Sencillo de usar en la mayoría de casos, pero lo suficientemente flexible como para adaptarse a los casos más complejos
- Basado en la premisa de "convenir en vez de configurar", en la que el desarrollador solo debe configurar aquello que no es convencional.
- Sigue la mayoría de mejores prácticas y patrones de diseño para la web.
- Preparado para aplicaciones empresariales y adaptable a las políticas y arquitecturas propias de cada empresa, además de ser lo suficientemente estable como para desarrollar aplicaciones a largo plazo
- Código fácil de leer que incluye comentarios de phpDocumentor¹⁵ y que permite un mantenimiento muy sencillo (17).

Teniendo en cuenta todas estas características además de la gran cantidad de componentes llamados helpers que facilitan el desarrollo web, se decidió usar este marco de trabajo para la realización del módulo de administración de Orión en su versión 1.4 .

Propel

¹⁵ Sistema para crear documentación de aplicaciones creadas con PHP. Conocido también como phpdoc o phpdocu, PhpDocumentor está escrito en PHP.

Propel es un Mapeado de Objetos Relacionales, en lo adelante ORM¹⁶, que funciona como una capa de abstracción entre el usuario y la BD. Se ha vuelto muy común el uso de este tipo de sistemas, principalmente cuando los proyectos requieren una gran dependencia de la BD como es el caso del motor de búsqueda Orión. Básicamente Propel se encarga de abstraer o separar al desarrollador del lenguaje estructurado de consultas (SQL), desarrollando clases a partir de las tablas y realizando relaciones entre ellas que facilitan el trabajo con las tablas y atributos. Symfony incluye Propel en la capa de abstracción, por lo que se usará en la realización de esta aplicación web.

Doctrine

Doctrine es un potente y completo sistema ORM para PHP 5.2+ con un DBAL (Capa de Abstracción de Base Datos) incorporado.

Entre muchas otras cosas tienes la posibilidad de exportar una base de datos existente a sus clases correspondientes y también a la inversa, es decir, convertir clases (convenientemente creadas siguiendo las pautas del ORM) a tablas de una base de datos. Por otro lado, como la librería es bastante grande, esta tiene un método para ser “compilada” al pasar a producción. Su principal ventaja radica en poder acceder a la base de datos utilizando la programación orientada a objetos (POO) debido a que doctrine utiliza el patrón Active Record para manejar la base de datos, tiene su propio lenguaje de consultas y trabaja de manera rápida y eficiente. Es fácilmente integrado a los principales marcos de trabajo de desarrollo utilizados actualmente.

ExtJS

Ext JS es una potente biblioteca de JS que simplifica el desarrollo asincrónico JavaScript + XML (Ajax) mediante el uso de objetos reutilizables y widgets (18).

Con la amplia variedad de marcos de desarrollo web disponibles hoy día, resulta en algunas ocasiones difícil para los desarrolladores determinar cual usar. ExtJS, es un marco de desarrollo basado en JS, que se destaca como una herramienta que los desarrolladores de aplicaciones Web deben considerar seriamente. Principalmente se basa en el desarrollo de la capa de presentación de las aplicaciones web a través de componentes dinámicos. Es

¹⁶ ORM: Acrónimo de Object Relational Mapping (Mapeo de Objetos Relacionales).

completamente orientado a objetos y extensible. Ext JS comenzó como un grupo de extensiones a la interfaz de usuario Yahoo! (YUI) Library por Jack Slocum. Con el reciente lanzamiento de la versión 3.0, sin embargo, se ha convertido en uno de los más sencillos y más potente bibliotecas de JS en el mercado (18).

Se usará este marco de trabajo en su versión 3.0.

1.6 Metodologías de Desarrollo

Se conoce que la construcción de un sistema de software implica la toma de decisiones sobre la arquitectura del mismo, así como definir los componentes y sus interacciones. Estas decisiones pueden ser cruciales para el éxito o fracaso del sistema resultante, por lo que se requiere seleccionar un proceso de desarrollo de software con el fin de obtener la calidad del sistema de software deseada y que cumpla con los requerimientos establecidos (19).

Las metodologías de desarrollo se pueden clasificar dentro de 2 grandes categorías: Las metodologías tradicionales y las metodologías ágiles, teniendo sus ventajas y sus desventajas cada una de ellas.

XP

XP¹⁷ es una metodología ágil centrada en potenciar las relaciones interpersonales como clave para el éxito en desarrollo de software, promoviendo el trabajo en equipo, preocupándose por el aprendizaje de los desarrolladores, y propiciando un buen clima de trabajo. XP se basa en realimentación continua entre el cliente y el equipo de desarrollo, comunicación fluida entre todos los participantes, simplicidad en las soluciones implementadas y coraje para enfrentar los cambios. XP se define como especialmente adecuada para proyectos con requisitos imprecisos y muy cambiantes, y donde existe un alto riesgo técnico (20).

RUP

RUP¹⁸ Es una de las metodologías de software más usada en el mundo. Está considerada como una metodología tradicional y generalmente se aplica a los proyectos grandes, por lo que el ritmo de

¹⁷ XP: Acrónimo de Extreme Programming.

¹⁸ Acrónimo de Rational Unified Process (Proceso Unificado de Desarrollo en español).

desarrollo es más lento que otras ágiles como SCRUM y XP. Esta metodología es iterativa e incremental y está compuesta por cuatro fases denominadas Inicio, Elaboración, Construcción y Transición. Además está estructurada en 9 flujos de trabajo de los cuales 6 son de proceso y 3 son de soporte como se muestra en la **Figura 4**.

Esta metodología está basada en componentes e interfaces bien definidas y utiliza el Lenguaje Unificado de Modelado (UML) para visualizar, especificar, construir y documentar el sistema de software.

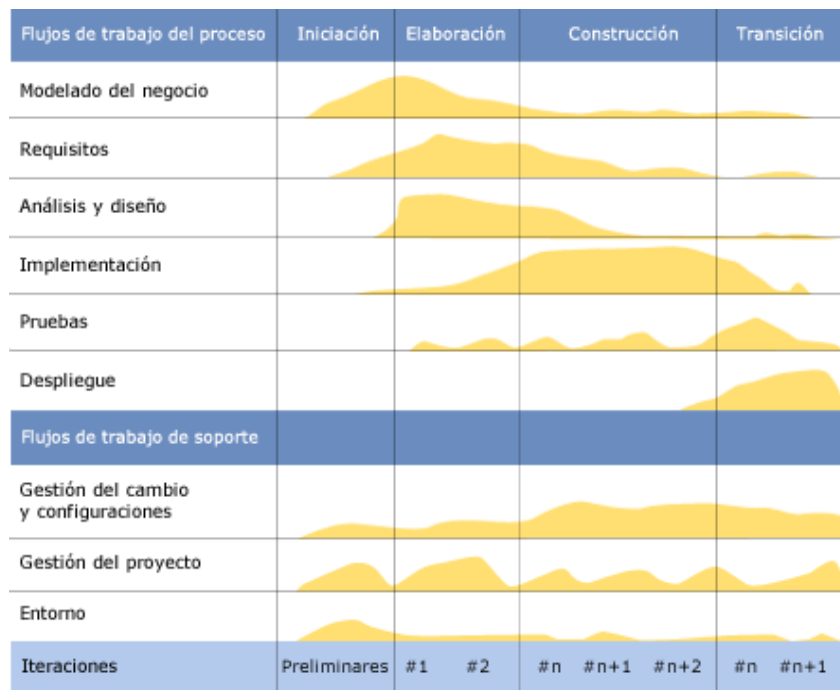


Figura 5. Metodología RUP con sus 9 flujos de trabajo.

Características de RUP:

- Dirigido por los casos de uso: los casos de usos son una forma abstracta de representar lo que los usuarios finales necesitan, por lo cual ellos son los que deben guiar el proceso de desarrollo. Esto se garantiza con la obtención de los diferentes modelos que son el resultado de los flujos de trabajo que propone RUP.
- Centrado en la arquitectura: la arquitectura de un proyecto muestra una visión común del sistema completo, visión en la que deben estar de acuerdo tanto el equipo de desarrollo como los clientes. Es por ello que la arquitectura describe los elementos del modelo que son más

importantes para la construcción del sistema, así como los cimientos para comprenderlo y desarrollarlo de forma económica.

- Iterativo e incremental: RUP divide el proceso en cuatro fases como se mencionaba anteriormente, dentro de las cuales se realizan varias iteraciones en número variable según el proyecto y las cuales se definen según el nivel de madurez que alcanzan los productos que se van obteniendo con cada actividad ejecutada. La terminación de cada fase ocurre en el hito correspondiente a cada una, donde se evalúa que se hayan cumplido los objetivos de la fase en cuestión.
- Enfocado en los riesgos: El Proceso Unificado requiere que el equipo del proyecto se centre en identificar los riesgos críticos en una etapa temprana del ciclo de vida. Los resultados de cada iteración, en especial los de la fase de Elaboración, deben ser seleccionados en un orden que asegure que los riesgos principales son considerados primero.

Scrum

Scrum es una metodología de desarrollo de software que generalmente se aplica a proyectos que no son tan grandes por lo que se le conoce como una metodología ágil. Tiene su nombre debido a una de las jugadas presente en el juego de rugby, donde el equipo entero rodea la pelota y actúa como un sólo hombre para intentar llegar al otro lado del campo, pasando el balón de uno a otro, dando la idea de que todo el equipo se concentra en un solo objetivo. En esta metodología se define un conjunto de prácticas y roles, y que puede tomarse como punto de partida para definir el proceso de desarrollo que se ejecutará durante un proyecto. Los roles principales en Scrum son el ScrumMaster, que mantiene los procesos y trabaja de forma similar al director de proyecto, el ProductOwner, que representa a los stakeholders (clientes externos o internos), y el Team que incluye a los desarrolladores.

Una de las cualidades de Scrum son los sprint, o sea, un período entre 15 y 30 días (la magnitud es definida por el equipo), el equipo crea un incremento de software potencialmente entregable (utilizable). El conjunto de características que forma parte de cada sprint viene del Product Backlog, que es un conjunto de requisitos de alto nivel priorizados que definen el trabajo a realizar. **Figura 6.** Teniendo en cuenta las características de esta metodología de desarrollo es la que se usará para la realización del módulo de administración web propuesto.

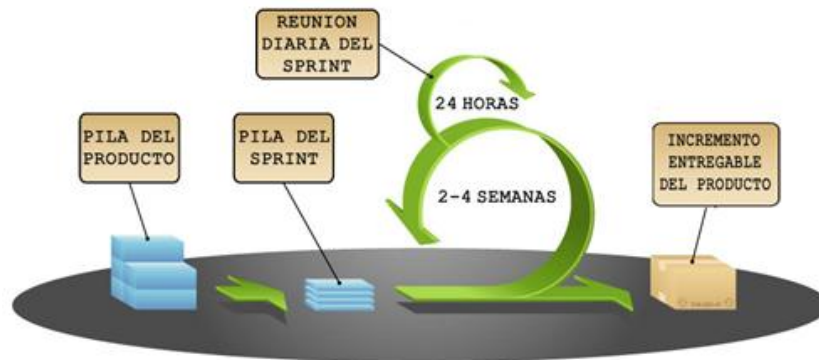


Figura 6. Metodología de desarrollo SCRUM.

1.7 Conclusiones Parciales

Luego de estudiadas las principales tendencias actuales relacionadas con el desarrollo de software, especialmente con las aplicaciones web, se adoptan un conjunto de herramientas y tecnologías que se describen a continuación. Se usará el patrón arquitectónico modelo-vista-controlador. Para la programación del lado del servidor se seleccionó PHP 5.3.3 por poseer características ideales que lo distinguen del resto de los lenguajes de programación. En aras de agilizar el desarrollo de la solución informática, se adopta Symfony 1.4 como marco de trabajo el cual se destaca por poseer una arquitectura flexible, la cual es extensible mediante plugins. Teniendo en cuenta la metodología usada y que en esta, el proceso de desarrollo está más orientado al desarrollo rápido y no a los artefactos que plantean otras metodologías como RUP, se usará solamente la herramienta case visual paradigm en la versión 6.4 para el modelado para el diseño de la base de Datos del sistema. Teniendo en cuenta la integración existente entre el NetBeans IDE 6.9 y el marco de trabajo Symfony, se adopta dicho IDE para el desarrollo de la solución. Se adopta la metodología de desarrollo Scrum por su rapidez en el desarrollo de proyectos informáticos.

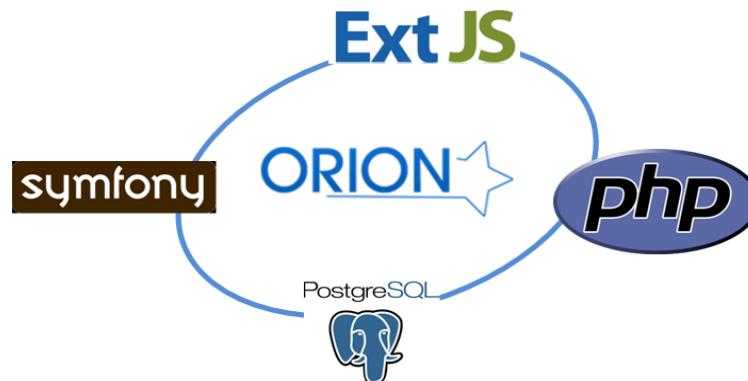


Figura 7. Base tecnológica para el módulo de administración propuesto.

CAPÍTULO 2: PROPUESTA DE SOLUCIÓN

2.1 Introducción

En este capítulo se abordarán los principales requerimientos funcionales y no funcionales tenidos en cuenta para la realización de la solución propuesta, se incluirán algunos patrones de diseño utilizados para lograr buenas prácticas de diseño y programación, se explicará la estructura de la base de datos desarrollada, así como la explicación de las principales opciones de configuración del motor de búsqueda Orión.

2.2 Requerimientos del Sistema

Los requerimientos del sistema son características, necesidades o funcionalidades que el sistema de administración deberá responder una vez terminada la solución propuesta. Scrum asume como punto de partida un documento denominado Pila del Sprint (Sprint Backlog) y Pila del Producto (Product Backlog) en los cuales se identifican los requerimientos funcionales generales del producto y los requerimientos que serán tenidos en cuenta para el sprint respectivamente, además de un conjunto de tareas que también son tenidas en cuenta para el primer sprint. Estos artefactos se muestran en el **Anexo 1** y **Anexo 2** respectivamente.

2.2.1 Requerimientos Funcionales

- **Autenticar Usuario**

La autenticación de usuario permitirá en primer lugar que el usuario se identifique correctamente en el sistema con su identificador de usuario y su clave de acceso y una vez realizado este proceso se procederá al acceso al módulo de administración con los privilegios asignados para ejercer tal función.

- **Gestionar Sección**

- **Insertar Sección**
- **Modificar Sección**
- **Eliminar Sección**

Las secciones son las partes en las que se divide un documento, algunas de estas son head, body, title, entre otras. A cada una de estas secciones se le asignará un peso, o sea, un valor numérico

que será tenido en cuenta a la hora de calcularle la relevancia a algún resultado encontrado. **Ver Anexo 3**

- **Restablecer Secciones**

Este requerimiento permitirá volver a la configuración original que tenían estas secciones.

- **Gestionar filtro**

Permitirá configurar los filtros realizados a las secciones de los documentos.

- **Gestionar Conversores**
 - **Insertar Conversor**
 - **Modificar Conversor**
 - **Eliminar Conversor**

Este requerimiento está relacionado con los programas que convertirán los documentos encontrados en la red de sitios web a un formato entendible para el indizador, por ejemplo, al encontrar algún documento pdf, este documento es convertido a texto plano con un Conversor definido anteriormente, este requerimiento será el que permitirá esta configuración. **Ver Anexo 4**

- **Restablecer Conversores**

Este requerimiento es para poder volver a la configuración original la configuración sobre los conversores de documentos utilizados.

- **Gestionar Tipos Mime**
 - **Insertar nuevos Tipos Mime**
 - **Modificar Tipos Mime**
 - **Eliminar Tipos Mime**

Es posible agrupar las extensiones de los documentos por tipos, estos tipos son llamados tipos Mime, o sea es la asociación de un tipo de documento con su extensión, este requerimiento permitirá la gestión de los tipos Mime que Orión tendrá en cuenta en las búsquedas realizadas. **Ver**

Anexo 5

- **Gestionar Documentos**
 - **Insertar Documentos**

- **Modificar Documentos**
- **Eliminar Documentos**

Cada uno de los tipos de documentos admitidos en el proceso de indización serán configurados a través de este requerimiento funcional. Los documentos pueden estar permitidos (allowed), denegados (deny) o solo chequeados (CheckOnly). **Ver Anexo 6.**

- **Restablecer Tipos Mime**

Este requerimiento está relacionado con el restablecimiento de la configuración original que poseía el motor de búsqueda antes de haberle hecho algún cambio

- **Gestionar URLs iniciales**

- **Insertar URL**
- **Modificar URL**
- **Eliminar URL**

Las URLs iniciales son el semillero o punto de partida a la hora de indizar la web, siempre que el *spider* encuentre nuevos enlaces, estos serán incluidos para su visita posterior, por lo que este requerimiento incluye la inserción, modificación y eliminación de estas URL iniciales.

- **Gestionar Configuración General**

Este requerimiento es para permitir la configuración de algunas propiedades que son básicas para el correcto funcionamiento del motor de búsqueda Orión

- **Restablecer Configuración General**

Este requerimiento está relacionado con la aplicación de la configuración básica con la que se instala el motor de búsqueda inicialmente.

- **Importar Ficheros**

Este requerimiento es el que incluye la carga de los ficheros de sinónimos y ficheros de palabras que no son tenidas en cuenta (stopword).

2.2.2 Requerimientos No Funcionales

Rapidez

El módulo de administración debe ser rápido aunque parte de lo relacionado con la rapidez de respuesta dependerá de las características del hardware donde se instale el sistema, pero también dependerá de optimizar el código lo más posible para su desarrollo. Las respuestas no deben demorar más de tres segundos.

Usabilidad

Debe poseer una funcionabilidad adecuada, o sea, satisfacer los requisitos funcionales declarados. Debe tener buena usabilidad, de modo que el esfuerzo para usarlo sea mínimo. También Debe contar con la portabilidad necesaria para poder ser transferido de un ambiente a otro o reemplazado por nuevas versiones. No deben existir enlaces con una profundidad mayor que tres.

Interfaz

Interfaz simple, intuitiva y agradable al usuario. No hacer uso de más de tres tonalidades de colores. Textos visibles y con buena tipografía preferentemente sin San Serif.

Disponibilidad

El sistema debe permanecer disponible las 24 horas del día. En caso de la ocurrencia de una falla técnica o eventualidad no prevista, se cuenta con un máximo de dos horas para la corrección del problema y la puesta en línea del sistema nuevamente.

Seguridad

El sistema debe verificar las credenciales de los usuarios que intenten acceder para garantizar el menor privilegio contra intentos mal intencionados. Además debe contar con mecanismos para el manejo de sesiones.

Hardware

- 2 Procesadores Intel Dual Core o Core 2 Duo a 3.0 Ghz.
- 2 GB de memoria RAM mínimo.
- 5 GB de espacio disponible en disco.
- Conexión de red Ethernet.

Software

- Sistema Operativo: Debian Lenny o Ubuntu Server
- Servidor Web Apache 2.2
- Servidor de Base de Datos: PostgreSQL
- Lenguaje de Programación PHP 5.

- Navegador Web Mozilla (versión 3.5.19 o superior) o Navegador Internet Explorer 7 (o superior).

2.3 Diseño del sistema

En el desarrollo del sistema se utilizaron un grupo de plugins de symfony previamente implementados por terceros y los cuales daban solución a distintas funcionalidades importantes para el sistema. A continuación se describen brevemente los plugins de symfony utilizados:

SfJqueryReloadedPlugin: Plugin que permite la integración de la librería de JavaScript Jquery en el marco de trabajo Symfony.

jQuery es una biblioteca o marco de trabajo para JavaScript, creada inicialmente por John Resig, que permite simplificar la manera de interactuar con los documentos HTML, manipular el árbol DOM, manejar eventos, desarrollar animaciones y agregar interacción con la tecnología AJAX a páginas web.

SfWebBrowserPlugin: integra un potente navegador web escrito en PHP. Muy útil para realizar peticiones web a servidores distantes sin la intervención del usuario. Es capaz de detectar los distintos códigos de estados y cabeceras HTTP devueltas por el servidor web.

SfMnogosearchPlugin: es un plugin que encapsula la lógica del motor de búsqueda mnoGoSearch, haciendo más fácil e intuitivo el desarrollo de aplicaciones basadas en el mismo. Para la correcta utilización del plugin, es necesario que estén previamente instaladas y configuradas las librerías de mnoGoSearch en el sistema.

2.3.1 Patrones de Diseño

Los patrones de diseño son soluciones simples y elegantes a problemas específicos y comunes del diseño orientado a objetos. Son soluciones basadas en la experiencia y que se ha demostrado que funcionan.

Un patrón de diseño es:

- ✓ Una solución estándar para un problema común de programación.
- ✓ Una técnica para flexibilizar el código haciéndolo satisfacer ciertos criterios.
- ✓ Un proyecto o estructura de implementación que logra una finalidad determinada.
- ✓ Un lenguaje de programación de alto nivel.
- ✓ Una manera más práctica de describir ciertos aspectos de la organización de un programa.
- ✓ Conexiones entre componentes de programas.

✓ La forma de un diagrama de objeto o de un modelo de objeto.

Singleton: Garantiza que una sola clase sólo tenga una instancia y proporciona un punto de acceso global a ella. El uso de este patrón consiste en definir una operación Instancia que permite que los clientes acceder a su única instancia. Instancia es una operación de clase que puede ser responsable de crear su única instancia. (21)

Este patrón es muy utilizado ya que entre sus ventajas está la del acceso controlado a la única instancia, espacio de nombres reducido, permite el refinamiento de operaciones y la representación, permite un número variable de instancias y es más flexible que las operaciones de clases.

La aplicación de este patrón se pone de manifiesto en el manejo de las sesiones de usuario y el uso del singleton ***sfcontext***.

Bajo Acoplamiento: El acoplamiento es una medida de fuerza con que un elemento está a, tiene conocimiento de, confía en, otros elementos. Este patrón es un principio que asigna la responsabilidad de controlar el flujo de eventos del sistema, a clases específicas. Esto facilita la centralización de actividades (validaciones, seguridad, etc.). El controlador no realiza estas actividades, las delega en otras clases con las que mantiene un modelo de alta cohesión.

Un error muy común es asignarle demasiada responsabilidad y alto nivel de acoplamiento con el resto de los componentes del sistema. (21)

Alta Cohesión: La cohesión es una medida de la fuerza con la que se relacionan las clases y el grado de focalización de las responsabilidades de un elemento, Cada elemento de nuestro diseño debe realizar una labor única dentro del sistema, no desempeñada por el resto de los elementos y auto-identificable, una clase con baja cohesión hace muchas cosas no relacionadas o hace demasiado trabajo. (21)

Controlador: Es un evento generado por actores externos. Se asocian con operaciones del sistema, operaciones del sistema como respuestas a los eventos del sistema, tal como se relacionan los mensajes y los métodos. Normalmente un controlador delega en otros objetos el trabajo que se necesita hacer; coordina o controla la actividad. No realiza mucho trabajo por sí mismo. (21)

Decorator: Responde a la necesidad de añadir dinámicamente funcionalidad a un Objeto. Esto nos permite no tener que crear sucesivas clases que hereden de la primera incorporando la nueva funcionalidad, sino otras que la implementan y se asocian a la primera. (22)

Su aplicación se basa principalmente en los siguientes casos:

- ✓ Añadir objetos individuales de forma dinámica y transparente
- ✓ Responsabilidades de un objeto pueden ser retiradas
- ✓ Cuando la extensión mediante la herencia no es viable.
- ✓ Hay una necesidad de extender la funcionalidad de una clase, pero no hay razones para extenderlo a través de la herencia.
- ✓ Hay la necesidad de extender dinámicamente la funcionalidad de un objeto y quizás quitar la funcionalidad extendida.

2.3.2 Prototipos de Interfaz de usuario

A continuación se muestran algunos prototipos correspondientes a algunos de los requerimientos funcionales más importantes.



Figura 9. Prototipo de interfaz de usuario del requerimiento Gestionar Sección.



Figura 10. Prototipo de interfaz de usuario del requerimiento Importar Ficheros.

2.4 Diseño de la Base de Datos

El diseño de la base de datos es una de las tareas más importantes que se realiza en el desarrollo de un sistema que necesite trabajar con información persistente. En general, el objetivo del diseño de una base de datos relacional es generar un conjunto de esquemas de relaciones que permitan almacenar la información con un mínimo de redundancia, pero que a la vez faciliten la recuperación de la información. En dependencia del modo en el que se esté usando el motor de búsqueda (single, multi, blob), serán las tablas con las que contará el sistema base para almacenar las búsquedas realizadas por los usuarios. Además de las tablas del sistema base, el módulo de administración agrega un total de 15 tablas para almacenar los valores que presenta cada una de las propiedades que serán configuradas por los administradores. La Base de datos se encuentra normalizada en tercera forma normal como muestra el diagrama entidad-relación de la *Figura 8*.

Descripción de algunas las tablas:

A continuación se muestra la descripción de algunas tablas de la base de datos del módulo de administración, para una revisión de todas las tablas remitirse al **Anexo 7**

Nombre: admin_servidor_local		
Descripción: Tabla que contendrá los atributos del servidor local donde estará el sistema.		
Atributo	Tipo	Descripción
Id_servidor_local	integer	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica al servidor local.
ip	varchar	Cadena que tendrá la dirección IP ¹⁹ del servidor local.

Tabla 1 Descripción de la tabla admin_servidor_local.

¹⁹ IP: Acrónimo de Protocolo de Internet (Internet Protocol en inglés). Número que identifica cada computadora en la red.

Nombre: admin_bd			
Descripción: Tabla que contendrá los atributos relacionados con la base de datos donde se guardarán las configuraciones de Orión de forma persistente.			
Atributo	Tipo	Null	Descripción
Id_bd	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada base de datos.
ruta	varchar	si	Dirección de la base de datos.
usuario	varchar	si	Nombre del usuario con acceso a la bd
clave	varchar	si	Contraseña del usuario
tipo	varchar	si	El tipo de base de datos (postgres, mysql, oracle).
modo	varchar	si	Modo de funcionamiento de la Base de datos (single, multi, blob).
activa	boolean	si	Muestra el estado de la Base de datos.
Id_servidor_local	varchar	si	Identificador del servidor local del indexer
Id_codificación	varchar	si	Identificador de la codificación usada en la Base de Datos

Tabla 3. Descripción de la tabla admin_bd.

Nombre: admin_convertor			
Descripción: Tabla que contendrá los atributos relacionados con los conversores auxiliares que usará el motor de búsqueda.			
Atributo	Tipo	Null	Descripción
Id_convertor	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada convertor.
activo	boolean	no	Estado que tendrán los conversores(activo o no)
tipo_origen	varchar	no	Tipo de documento origen
tipo_destino	varchar	no	Tipo de documento destino
comando	varchar	no	Comando para convertir el documento
descripcion	varchar	si	Descripción del convertor
Id_servidor_local	integer	no	Identificador del servidor local del indexer

Tabla 4. Descripción de la tabla admin_convertor.

Código de la base de datos

El script fue generado para el SGBD PostgreSQL

2.5 Implementación del Sistema

Durante el desarrollo del módulo de administración implementado se pudo aplicar el patrón arquitectónico modelo-vista-controlador propuesto por el marco de trabajo Symfony, realizando las funcionalidades que interactúan con el fichero de configuración en la clase ***action.class.php*** y a su vez esta hizo uso de las clases generadas automáticamente por el ORM para la abstracción de la Base de Datos.

Se hizo uso de las expresiones regulares para identificar los patrones dentro del fichero y poder realizarle los filtros a las líneas que pertenecen a cada requerimiento funcional.

Durante el desarrollo del sistema, se reutilizaron los componentes de software que se describen a continuación:

- MnoGoSearch: constituye las librerías de mnoGoSearch para el lenguaje de programación PHP 5.

Dicha librería contiene las principales funciones utilizadas en la interacción con las opciones que brinda el software mnoGosearch.

- Symfony: constituyen las librerías utilizadas del framework symfony.
- Plugins: paquete de plugins implementados por terceros que sirvieron de base para la implementación de algunas funcionalidades necesarias para el sistema.
- Orión: constituye las clases y estructuras de datos del sistema implementado.

2.6 Conclusiones Parciales

Con el desarrollo del sistema, se obtuvieron algunas clases para lograr la configuración del motor de búsqueda, además se llegó a la conclusión de que existían muchas propiedades que no se tenían en cuenta a la hora de configurar Orión. Se logró el modelado y programación de las clases y estructuras de datos fundamentales sobre las que se continuará el desarrollo del sistema. Se logró además, la obtención de un diagrama entidad-relación que satisface las principales necesidades de almacenamiento que requiere el sistema.

CAPÍTULO 3: EVALUACIÓN DEL SISTEMA

3.1 Introducción

El desarrollo del software implica una serie de actividades de producción en las que las posibilidades de que aparezca la falibilidad humana son comunes. Los errores pueden empezar a darse desde el primer momento del proceso en el que los objetivos pueden estar especificados de forma errónea. Debido a la imposibilidad humana de trabajar y comunicarse de forma perfecta, el desarrollo del software ha de ir acompañado de una actividad que garantice la calidad.

Las pruebas son una actividad en la cual un sistema o componente es ejecutado bajo unas condiciones o requerimientos especificados, los resultados son observados y registrados, y una evaluación es hecha de algún aspecto del sistema o componente.

La prueba de software es un elemento crítico para la garantía de la calidad del software y representa una revisión final de las especificaciones del diseño y de la codificación.

El método de prueba que se va a aplicar al sistema es el de caja negra el cual se refiere a las pruebas que se llevan a cabo sobre la interfaz del software. Los casos de prueba pretenden demostrar que las funciones del software son operativas, que la entrada se acepta de forma adecuada y que se produce un resultado correcto, así como que la integridad de la información externa se mantiene.

3.2 Tipos de pruebas de software

La calidad de un sistema está determinada, entre otras cosas, por la coincidencia entre lo que se programó y los requerimientos establecidos en la primera parte de su desarrollo. Para comprobar el grado de cumplimiento de estos requisitos se usan las pruebas del sistema. La prueba de unidad es la prueba enfocada a los elementos probables más pequeños del software, consiste en una prueba estructural (o caja blanca), lo cual requiere conocer el diseño interno de la unidad puesto que verifica la lógica interna y el flujo de datos; y una prueba de especificación (de caja negra), basada sólo en la especificación del comportamiento externamente visible de la unidad, generalmente relacionada con las interfaces del sistema.

Estas definen un conjunto amplio de acciones de comprobación que abarcan todas las características que determinan la calidad de un software. Se comprueban las funcionalidades diseñando casos de prueba que definen cómo proceder. Estos casos de prueba incluyen los juegos

de datos a usar que son los válidos o esperados y los no válidos o no esperados por el programa. Las pruebas se deben aplicar durante todo el ciclo de vida del software e invariablemente se le debe dedicar una gran parte del esfuerzo total del desarrollo. Se deben planificar correctamente desde el inicio y establecer qué hacer, cómo hacer, quién va a hacer y en qué condiciones hacer las comprobaciones.

3.2.1 Pruebas de caja Negra

Este tipo de pruebas se llevan a cabo sobre la interfaz del software, por lo que los casos de prueba pretenden demostrar que las funciones del software son operativas, que la entrada se acepta de forma adecuada y que se produce una salida correcta, así como que la integridad de la información externa se mantiene. Esta prueba examina algunos aspectos del modelo fundamentalmente del sistema sin tener mucho en cuenta la estructura interna del software. Generalmente este tipo de pruebas intentan encontrar funciones incorrectas o ausentes, errores en la interfaz o errores en los valores de entrada por lo que las validaciones de estos es algo que se debe tener en cuenta por los desarrolladores. (23)

La prueba de Caja Negra no es una alternativa a las técnicas de prueba de la Caja Blanca, sino un enfoque complementario que intenta descubrir diferentes tipos de errores a los encontrados en los métodos de la Caja Blanca. Teniendo en cuenta estas características se decidió aplicarlas para evaluar la solución implementada.

3.2.2 Pruebas de caja Blanca

Se basa en el minucioso examen de los detalles procedimentales. Se comprueban los caminos lógicos del software proponiendo casos de prueba que examinen que están correctas todas las condiciones y/o bucles para determinar si el estado real coincide con el esperado o afirmado. Esto genera gran cantidad de caminos posibles por lo que hay que dedicar esfuerzos a la determinación de las condiciones de prueba que se van a verificar.

La prueba de caja blanca se basa en el diseño de casos de prueba que usa la estructura de control del diseño procedimental para derivarlos. Mediante la prueba de la caja blanca el ingeniero del software puede obtener casos de prueba que:

- Garanticen que se ejerciten por lo menos una vez todos los caminos independientes de cada módulo, programa o método.
- Ejerciten todas las decisiones lógicas en las vertientes verdadera y falsa.
- Ejecuten todos los bucles en sus límites operacionales.
- Ejerciten las estructuras internas de datos para asegurar su validez.

Es por ello que se considera a la prueba de Caja Blanca como uno de los tipos de pruebas más importantes que se le aplican a los software, logrando como resultado que disminuya en un gran porcentaje el número de errores existentes en los sistemas y por ende una mayor calidad y confiabilidad. (24)

3.3 Casos de prueba

Un caso de prueba no es más que una evaluación que se le realiza a las interfaces que responden a un requerimiento funcional del sistema. Para preparar los casos de pruebas hacen falta un número de datos que ayuden a la ejecución de estos casos y que permitan que el sistema se ejecute en todas sus variantes, pueden ser datos válidos o inválidos en dependencia de si lo que se desea es hallar un error o probar una funcionalidad. Los datos se escogen atendiendo a las especificaciones del problema, sin importar los detalles internos del programa, a fin de verificar que el programa se ejecute correctamente.

RF Autenticar Usuario

Caso de Prueba del requerimiento Autenticar Usuario
Descripción: Este caso de prueba es para verificar la entrada de los campos usuario y clave
Condiciones: <ul style="list-style-type: none"> • La estación de trabajo debe tener algún navegador con la aplicación en ejecución.
Entrada: <ul style="list-style-type: none"> • Se introduce el usuario en un campo de texto • Se introduce la clave en un campo de contraseñas
Resultados Esperados: <ul style="list-style-type: none"> • Se presiona el botón Aceptar y se muestra la página de administración. • En caso de que el usuario no exista en la Base de Datos se muestra un mensaje de error indicando que ese usuario no está registrado. • Si deja el campo de usuario o clave vacío se muestra un error indicando que debe llenarlo • Si la contraseña introducida no coincide con la contraseña para el usuario, se muestra un mensaje de error, indicando que la contraseña es incorrecta.
Juego de Datos: <ul style="list-style-type: none"> • Al introducir un usuario correcto y su clave incorrecta el sistema muestra el siguiente mensaje

“Asegúrese de que su clave sea correcta”

- Al dejar el campo de usuario vacío o el de contraseña el sistema muestra el siguiente mensaje “No se permiten campos vacíos”
- Al introducir un usuario que no esté registrado en el sistema se muestra el siguiente mensaje “Este usuario no está registrado”

RF Insertar Sección

Caso de Prueba del requerimiento Insertar Sección

Descripción: Este caso de prueba es para verificar la entrada de los campos referentes a las secciones.

Condiciones:

- Se debe estar autenticado en el sistema

Entrada:

- Se introduce el nombre de la sección en un campo de texto
- Se introduce el identificador de la sección en un campo de números enteros
- Se introduce el peso de la sección en un campo de números enteros
- Se introduce la descripción de la sección en un campo de texto

Resultados Esperados:

- Se presiona el botón Adicionar dentro del tab sección y se muestra una ventana para adicionar la sección.
- El campo id y peso no permiten entrada de otra cosa que no sea números por estar validados con ExtJS.
- En caso de que deje el campo nombre, id o peso vacío se mostrará el mensaje “Solo se permite vacío el campo descripción”.
- Puede dejar el campo descripción vacío teniendo en cuenta que es opcional en la Base de datos.
- Se presiona el botón Aceptar y se inserta la nueva sección.

Juego de Datos:

- El sistema no permite la entrada de caracteres incorrectos en los campos id y peso.
- Al dejar solo el campo de id vacío y llenar los otros se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”.
- Al dejar solo el campo de nombre vacío y llenar los otros se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”
- Al dejar solo el campo de peso vacío y llenar los otros se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”

RF Insertar Documentos

Caso de Prueba del requerimiento Insertar Documentos
Descripción: Este caso de prueba es para verificar la entrada de los campos referentes a los documentos.
Condiciones: <ul style="list-style-type: none"> • Se debe estar autenticado en el sistema
Entrada: <ul style="list-style-type: none"> • Se escoge el método de indización a través de un Combobox. • Se escoge el Tipo de expresión a través de un Combobox. • Se escoge la exactitud o no a través de un Combobox(opcional). • Se escoge la sensibilidad o no a través de un Combobox(opcional). • Se introducen los tipos de documentos en un campo de texto • Se introduce la descripción de la sección en un campo de texto(opcional).
Resultados Esperados: <ul style="list-style-type: none"> • Se presiona el botón Adicionar dentro del tab documentos y se muestra una ventana para adicionar los tipos de documentos. • Los campos método, Tipo Expresión, exactitud y sensibilidad ya están validados en extjs para no permitir valores incorrectos, teniendo en cuenta que a través de un combo el usuario solo podrá seleccionar las opciones permitidas para estos tipos. • En caso de que deje el campo tipos vacío se mostrará el mensaje “Solo se permite vacío el campo descripción”. • Puede dejar el campo descripción vacío teniendo en cuenta que es opcional en la Base de datos.
Juego de Datos: <ul style="list-style-type: none"> • Los campos método, Tipo Expresión, exactitud y sensibilidad ya están validados en extjs para no permitir valores incorrectos, teniendo en cuenta que a través de un combo el usuario solo podrá seleccionar las opciones permitidas para estos tipos. • Al dejar el campo tipos vacío y llenar los otros se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”.

No Conformidad: Esta interfaz tenía un problema con el campo descripción y no permitía dejarlo vacío por lo que hubo que volver a realizar el caso de prueba después de corregir el error.

RF Insertar Conversores

Caso de Prueba del requerimiento Insertar Conversores
Descripción: Este caso de prueba es para verificar la entrada de los campos referentes a los conversores externos.
Condiciones: <ul style="list-style-type: none"> • Se debe estar autenticado en el sistema

<p>Entrada:</p> <ul style="list-style-type: none"> • Se introduce el tipo origen en un campo de texto • Se introduce el tipo destino en un campo de texto • Se introduce el comando en un campo de texto • Se introduce la descripción de la sección en un campo de texto(opcional).
<p>Resultados Esperados:</p> <ul style="list-style-type: none"> • Se presiona el botón Adicionar dentro del tab conversores y se muestra una ventana para adicionar los conversores externos. • En caso de que deje el campo origen, destino o peso vacío se mostrará el mensaje “Solo se permite vacío el campo descripción”. • Puede dejar el campo descripción vacío teniendo en cuenta que es opcional en la Base de datos. • Se presiona el botón Aceptar y se inserta el nuevo conversor.
<p>Juego de Datos:</p> <ul style="list-style-type: none"> • Al dejar solo el campo de origen vacío y llenar los otros se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”. • Al dejar solo el campo de destino vacío y llenar los otros se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”. • Al dejar solo el campo comando vacío y llenar los otros se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”. • Se permite cualquier dato de entrada teniendo en cuenta la complejidad del texto a entrar y que puede ser números, letras, incluso caracteres especiales.

RF Insertar Tipos Mime

<p>Caso de Prueba del requerimiento Insertar Tipos Mime</p>
<p>Descripción: Este caso de prueba es para verificar la entrada de los campos referentes a los Tipos Mime</p>
<p>Condiciones:</p> <ul style="list-style-type: none"> • Se debe estar autenticado en el sistema
<p>Entrada:</p> <ul style="list-style-type: none"> • Se introduce el tipo mime en un campo de texto • Se escoge el Tipo de expresión a través de un Combobox. • Se escoge la sensibilidad o no a través de un Combobox(opcional). • Se introduce las extensiones asociadas a un tipo mime en un campo de texto • Se introduce la descripción del tipo mime en un campo de texto(opcional).
<p>Resultados Esperados:</p> <ul style="list-style-type: none"> • Se presiona el botón Adicionar dentro del tab Tipos Mime y se muestra una ventana para adicionar los conversores externos. • Se escoge el Tipo de expresión a través de un Combobox. • Se escoge la sensibilidad o no a través de un Combobox(opcional). • En caso de que deje el campo Tipo mime o los tipos vacío se mostrará el mensaje “Solo se

<p>permite vacío el campo descripción”.</p> <ul style="list-style-type: none"> • Puede dejar el campo descripción vacío teniendo en cuenta que es opcional en la Base de datos. • Se presiona el botón Aceptar y se inserta el nuevo Tipo mime.
<p>Juego de Datos:</p> <ul style="list-style-type: none"> • Los campos Tipo Expresión y sensibilidad ya están validados en extjs para no permitir valores incorrectos, teniendo en cuenta que a través de un combo el usuario solo podrá seleccionar las opciones permitidas para estos tipos. • Al dejar solo el campo de Tipo mime vacío y llenar los otros se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”. • Al dejar solo el campo tipos vacío y llenar los otros se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”. • El campo tipo mime puede tener letras, números, incluso caracteres especiales

RF Insertar URLs Iniciales

<p>Caso de Prueba del requerimiento Insertar URLs iniciales</p>
<p>Descripción: Este caso de prueba es para verificar la entrada de los campos referentes a las Urls iniciales</p>
<p>Condiciones:</p> <ul style="list-style-type: none"> • Se debe estar autenticado en el sistema
<p>Entrada:</p> <ul style="list-style-type: none"> • Se escoge el método a través de un Combobox (opcional). • Se escoge la subsección a través de un Combobox (opcional). • Se introduce la dirección url en un campo de texto. • Se introduce el alias en un campo de texto (opcional). • Se introduce la descripción de la url en un campo de texto(opcional).
<p>Resultados Esperados:</p> <ul style="list-style-type: none"> • Se presiona el botón Adicionar dentro del tab Url iniciales y se muestra una ventana para adicionar las url que servirán como semillero para el motor de búsqueda. • Se escoge el método a través de un Combobox(opcional). • Se escoge la subsección a través de un Combobox(opcional). • En caso de que deje el campo url vacío se mostrará el mensaje “Solo se permite vacío el campo descripción”. • Puede dejar el campo descripción vacío teniendo en cuenta que es opcional en la Base de datos. • Se presiona el botón Aceptar y se inserta la nueva url.
<p>Juego de Datos:</p> <ul style="list-style-type: none"> • Los campos método y subsección ya están validados en extjs para no permitir valores incorrectos, teniendo en cuenta que a través de un combo el usuario solo podrá seleccionar las opciones permitidas para estas url.

- Al dejar el campo url vacío se muestra el siguiente mensaje “Solo se permite vacío el campo descripción”.
- El campo url puede tener letras, números, incluso caracteres especiales

RF Gestionar configuración general

Caso de Prueba del requerimiento Gestionar configuración general
Descripción: Este caso de prueba es para verificar la entrada de los campos referentes a los parámetros generales del motor de búsqueda Orión.
Condiciones: <ul style="list-style-type: none"> • Se debe estar autenticado en el sistema
Entrada: <ul style="list-style-type: none"> • Se introducen algunas propiedades a través de campos de texto. • Ningún campo de texto es opcional en este caso de prueba.
Resultados Esperados: <ul style="list-style-type: none"> • No se permite ningún campo de texto vacío. • Los campos que son para números están validados con EXTJS para no permitir valores incorrectos. • Los campos que son para letras solamente están validados con EXTJS para no permitir valores incorrectos. • Se guarda la configuración general • Se presiona el botón Aceptar y se inserta la nueva url.
Juego de Datos: <ul style="list-style-type: none"> • Al dejar algún campo vacío se muestra el siguiente mensaje “No se permiten campos vacíos”.

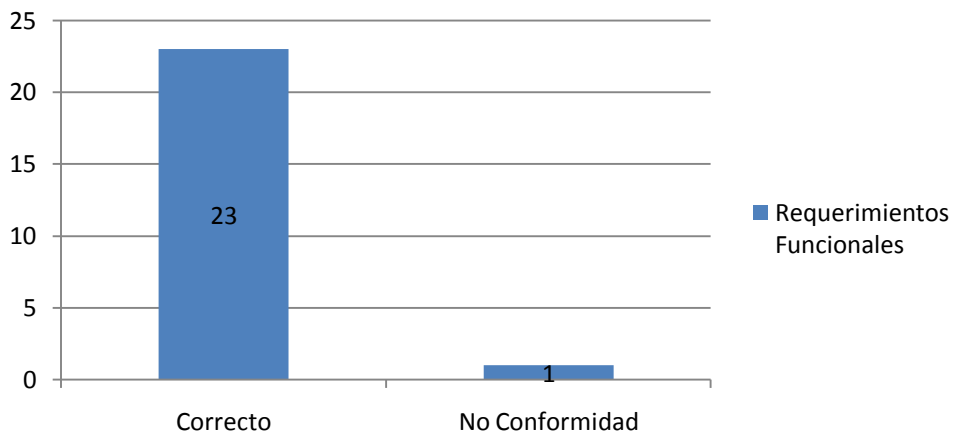
Nota: Todos los requerimientos funcionales que son modificar utilizan las mismas interfaces que ya han sido probadas por lo que se hace innecesaria su prueba nuevamente.

Existe otro conjunto de requerimientos funcionales que solamente llaman funciones al pulsar botones y no requieren de ningún parámetro de entrada, todos estos enlaces fueron revisados para evitar enlaces rotos, en todos los casos funcionaban correctamente.

3.4 Conclusiones Parciales

Luego de realizadas las pruebas de caja negra a las interfaces se concluye que el sistema cuenta con muy buena calidad, de un total de 23 requerimientos funcionales solo uno tuvo una no conformidad que también fue subsanada lo que representa un 95.65 % de calidad, siendo fácil para los administradores realizar la administración a través de las interfaces implementadas.

Requerimientos Funcionales



CONCLUSIONES GENERALES

El presente trabajo se enfocó hacia el desarrollo de un módulo de administración para el motor de búsqueda Orión. Como resultado se obtuvo un módulo que facilita el proceso antes mencionado a los administradores, permitiendo gestionar las secciones, los documentos, las URLs iniciales, entre otras opciones generales de configuración, aportando las siguientes ventajas:

- Ahorro en tiempo y esfuerzo a los administradores.
- Mayor usabilidad que la variante antes usada para administrar Orión.
- Se obtuvo un estudio acerca de las interfaces gráficas de usuario en los principales sistemas para la recuperación de información actuales, tales como Indri, Lemur, Apache Lucene, entre otros.
- Se realizó un estudio sobre las herramientas definidas para el desarrollo del sistema.
- Después del desarrollo del sistema se pudo evidenciar que existían muchas propiedades que los administradores no tenían en cuenta por el gran tamaño que tiene el fichero de configuración y que ahora si son configuradas a través del módulo implementado.
- Se consideran cumplidos todos los objetivos trazados al comenzar la investigación.

RECOMENDACIONES

Se recomienda implementar un módulo de estadísticas para trabajar con los datos almacenados en la Base de Datos.

Se recomienda la publicación del código fuente a toda la comunidad de internet que utiliza el spider MnoGosearch para facilitar la administración del mismo.

Se recomienda luego de traducir la documentación incluirla en este módulo para informar más a los administradores acerca de las propiedades tenidas en cuenta por el indizador.

REFERENCIAS BIBLIOGRÁFICAS

1. **Chiavenato, Idalberto.** *Introducción a la Teoría General de la Administración.* s.l. : McGraw-Hill, 2006. ISBN: 9701055004.
2. **Real Academia Española.** Diccionario de la Real Academia Española. [En línea] [Citado el: 8 de Marzo de 2011.] http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=informaci%C3%B3n.
3. **Gómez Díaz, Raquel.** *Estudio de la incidencia del conocimiento lingüístico en los Sistemas de Recuperación de Información.* Salamanca : Universidad de Salamanca, 2002. ISBN:84-7800-831-4.
4. **Stephanidis, Constantine.** *User Interfaces for All.* Estados Unidos : Lawrence Erlbaum, 2001. ISBN:0-8058-2967-9.
5. **Lewis, Clayton y Rieman, John.** Task-Centered User Interface Design. [En línea] -, 1995. [Citado el: 10 de Marzo de 2011.] <http://hcibib.org/tcuid/>.
6. **Fernández Luna, Juan Manuel y Torres López, Carmen.** *Empleo de motores de búsqueda de código abierto para la recuperación de información vertical.* Granada : Universidad de Granada, 2010. ISBN: 470-1758-1.
7. **Barkov, Alexander.** mnoGoSearch 3.3.9 reference manual. [En línea] 2009. [Citado el: 10 de Marzo de 2011.] <http://www.mnogosearch.org/doc33/msearch-intro.html#features>.
8. **Garret, Jesse James.** Ajax: A New Approach to Web Applications. [Online] Adaptive Path, 2005. [Cited: marzo 5, 2011.] http://www.robertspahr.com/courses/web1/ajax_web_applications.pdf.
9. **Eguíluz Pérez, Javier.** Introducción a CSS. [En línea] Junio de 2007. [Citado el: 14 de Marzo de 2011.] <http://www.librosweb.es/css>.
10. **Derick Rethans Andi Gutmans, Stig Sather Bakken.** PHP 5 Power Programming, chapter PHP 5 Basic Language, pages 10–13. s.l. : PRENTICE HALL, 2005.
11. **López, Daniel.** PHP4 y PHP5: ¿Cuál elegir? ¿Migrar o no Migrar? El advenimiento de PHP6 [Internet]. [En línea] [Citado el: 28 de Enero de 2011.] <http://www.maestrosdelweb.com/editorial>.
12. **Instituto Nacional de Estadística e Informática.** *Herramientas Case.* 1999.
13. **Mató Rodríguez, Raúl.** *Diseño de Bases de Datos para la Empresa de Gas Manufacturado.* Habana. s.l. : Casa Editorial La Habana, 2009.
14. **Editorial McGraw-Hill .** SGBD. Sistemas gestores de bases de datos. Introducción. [En línea] Editorial McGraw-Hill , 21 de Diciembre de 2008. [Citado el: 13 de Marzo de 2011.] <http://www.mailxmail.com/curso-sistemas-bases-datos/sghbd-sistemas-gestores-bases-datos-introduccion>. ISSN:1699-4914.
15. **Martínez Capetillo, Cristian.** Idalicius. [En línea] 2 de Febrero de 2010. [Citado el: 13 de Marzo de 2011.] <http://www.idalicius.net/2010/02/sistema-gestion-de-bases-de-datos-sghbd.html>.
16. **Márquez, Juan Manuel.** *SISTEMAS DE INFORMACION.* 2003.
17. **Potencier, Fabien y Zaninotto, François.** LibrosWeb.es - Libros gratuitos sobre HTML, CSS, JavaScript, AJAX y otras tecnologías web. [En línea] Apress, 12 de Diciembre de 2008. [Citado el: 2 de Marzo de 2011.] http://www.librosweb.es/symfony_1_2. ISBN-13: 978-1590597866.
18. **W. Fronckowiak, John.** *Build Ajax applications with Ext JS.* s.l. : IDC Consulting, 2008.
19. Rational Unified Process. [En línea] 2006. [Citado el: 12 de Marzo de 2011.] http://es.wikipedia.org/wiki/Rational_Unified_Process.

20. **Letelier, Patricio, Penadés, Carmen y Canós, José H.** *Métodologías Ágiles en el Desarrollo de Software*. Valencia : DSIC -Universidad Politécnica de Valencia, 2000.
21. **Larman, Craig.** *UML y patrones*. 1999. ISBN 970-1 7-0261-1.
22. **Fowler, Martin.** *Patterns of Enterprise Application Architecture*. ISBN: 0-32112-742-0.
23. **Serrano Rosales, Carlos Luis.** *Sobre la disciplina de Prueba*. Ciudad de la habana : s.n., 2010.
24. **S. Pressman, Roger.** *Ingeniería del Software:Un enfoque práctico*. s.l. : McGraw-Hill, 2000. ISBN:970-10-5473-3.

BIBLIOGRAFÍA

1. **Chiavenato, Idalberto.** *Introducción a la Teoría General de la Administración.* s.l. : McGraw-Hill, 2006. ISBN: 9701055004.
2. **Real Academia Española.** Diccionario de la Real Academia Española. [En línea] [Citado el: 8 de Marzo de 2011.] http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=informaci%C3%B3n.
3. **Gómez Díaz, Raquel.** *Estudio de la incidencia del conocimiento lingüístico en los Sistemas de Recuperación de Información.* Salamanca : Universidad de Salamanca, 2002. ISBN:84-7800-831-4.
4. **Stephanidis, Constantine.** *User Interfaces for All.* Estados Unidos : Lawrence Erlbaum, 2001. ISBN:0-8058-2967-9.
5. **Lewis, Clayton y Rieman, John.** Task-Centered User Interface Design. [En línea] -, 1995. [Citado el: 10 de Marzo de 2011.] <http://hcibib.org/tcuid/>.
6. **Fernández Luna, Juan Manuel y Torres López, Carmen.** *Empleo de motores de búsqueda de código abierto para la recuperación de información vertical.* Granada : Universidad de Granada, 2010. ISBN: 470-1758-1.
7. **Barkov, Alexander.** mnoGoSearch 3.3.9 reference manual. [En línea] 2009. [Citado el: 10 de Marzo de 2011.] <http://www.mnogosearch.org/doc33/msearch-intro.html#features>.
8. **Garret, Jesse James.** Ajax: A New Approach to Web Applications. [Online] Adaptive Path, 2005. [Cited: marzo 5, 2011.] http://www.robertspahr.com/courses/web1/ajax_web_applications.pdf.
9. **Eguíluz Pérez, Javier.** Introducción a CSS. [En línea] Junio de 2007. [Citado el: 14 de Marzo de 2011.] <http://www.librosweb.es/css>.
10. **Derick Rethans Andi Gutmans, Stig Sather Bakken.** PHP 5 Power Programming, chapter PHP 5 Basic Language, pages 10–13. s.l. : PRENTICE HALL, 2005.
11. **López, Daniel.** PHP4 y PHP5: ¿Cuál elegir? ¿Migrar o no Migrar? El advenimiento de PHP6 [Internet]. [En línea] [Citado el: 28 de Enero de 2011.] <http://www.maestrosdelweb.com/editorial>.
12. **Instituto Nacional de Estadística e Informática.** *Herramientas Case.* 1999.
13. **Mató Rodríguez, Raúl.** *Diseño de Bases de Datos para la Empresa de Gas Manufacturado.* Habana. s.l. : Casa Editorial La Habana, 2009.
14. **Editorial McGraw-Hill .** SGBD. Sistemas gestores de bases de datos. Introducción. [En línea] Editorial McGraw-Hill , 21 de Diciembre de 2008. [Citado el: 13 de Marzo de 2011.] <http://www.mailxmail.com/curso-sistemas-bases-datos/sghbd-sistemas-gestores-bases-datos-introduccion>. ISSN:1699-4914.
15. **Martínez Capetillo, Cristian.** Idalicius. [En línea] 2 de Febrero de 2010. [Citado el: 13 de Marzo de 2011.] <http://www.idalicius.net/2010/02/sistema-gestion-de-bases-de-datos-sghbd.html>.
16. **Márquez, Juan Manuel.** *SISTEMAS DE INFORMACION.* 2003.
17. **Potencier, Fabien y Zaninotto, François.** LibrosWeb.es - Libros gratuitos sobre HTML, CSS, JavaScript, AJAX y otras tecnologías web. [En línea] Apress, 12 de Diciembre de 2008. [Citado el: 2 de Marzo de 2011.] http://www.librosweb.es/symfony_1_2. ISBN-13: 978-1590597866.
18. **W. Fronckowiak, John.** *Build Ajax applications with Ext JS.* s.l. : IDC Consulting, 2008.
19. Rational Unified Process. [En línea] 2006. [Citado el: 12 de Marzo de 2011.] http://es.wikipedia.org/wiki/Rational_Unified_Process.

20. **Letelier, Patricio, Penadés, Carmen y Canós, José H.** *Métodologías Ágiles en el Desarrollo de Software*. Valencia : DSIC -Universidad Politécnica de Valencia, 2000.
21. **Larman, Craig.** *UML y patrones*. 1999. ISBN 970-1 7-0261-1.
22. **Fowler, Martin.** *Patterns of Enterprise Application Architecture*. ISBN: 0-32112-742-0.
23. **Serrano Rosales, Carlos Luis.** *Sobre la disciplina de Prueba*. Ciudad de la habana : s.n., 2010.
24. **S. Pressman, Roger.** *Ingeniería del Software: Un enfoque práctico*. s.l. : McGraw-Hill, 2000. ISBN:970-10-5473-3.
25. **<http://www.verypdf.com>.** *Clasificación de documentos con BlindLight*.
26. **Ciudad Ricardo, MSc. Febe Angel.** *Buenas prácticas para una exposición exitosa*. Ciudad de La Habana : UCI, 2008.
27. **Eito Brun, Ricardo.** *Recuperación de Información y sistemas para la gestion del conocimiento*. Madrid : Universidad Carlos III, 2009.
28. **Hernández León, Rolando Alfredo.** *EL PARADIGMA CUANTITATIVO DE LA INVESTIGACIÓN CIENTÍFICA*. Ciudad de la Habana : s.n., 2002. ISBN: 959-16-0343-6.
29. **Hidalgo Delgado, Ing. Yusniel.** *Orión, un motor de búsqueda para la web de la uci*. Ciudad de La Habana : UCI, 2010.
30. **Jacobson, Ivar, Booch, Grady y Rumbaugh, James.** *El Proceso unificado de software*. Madrid : Addison Wesley, 2000. ISBN:84-7829-036-2.
31. **James Garrett, Jesse.** *Ajax: A New Approach to web Applications*. 2005.
32. **Jorge Fernández, Francisca Miriam.** *Comunicación Científica*. Ciudad de La Habana : Editorial Universitaria, 2007, 2007. ISBN 978-959-16-0608-2.
33. **Labrada Sedeño, Roberto Oscar.** *Clasificación y Agrupamiento de textos de noticias*. Ciudad de La Habana : s.n., 2010.
34. **Mampel Laboira, Luis y Cortés García, Angel Luis.** *Los motores de búsqueda de información en Internet como herramienta de aprendizaje: análisis de contenidos sobre el concepto “dinosaurio”*. s.l. : Revista Electrónica de Enseñanza de las Ciencias, 2009.
35. **Matos Nieto, Abdiel.** *Coliseo Virtual, Sistema de apoyo al proceso de enseñanza-aprendizaje desde juegos didácticos*. Ciudad de La Habana : UCI, 2009.
36. **Peyton Jones, Simon.** *How to write a great research paper*. Ciudad de la Habana : s.n., 2010.
37. **Pupo Mulgado, Marledis y Reyna Castro, Yisel.** *Propuesta de pronósticos y tendencias sobre el comportamiento de la Web en la UCI*. Ciudad de La Habana : UCI, 2010.
38. **Segoviano Martínez, Zamira.** *ZOTERO: UNA ALTERNATIVA LIBRE PARA LA GESTIÓN*. Ciudad de la Habana : s.n., 2010.
39. **Valero de La Rosa, Damicel y Abella Pérez, Raul.** *Propuesta de estándares para el Posicionamiento de Sitios Web*. Ciudad de La Habana : UCI, 2007.

GLOSARIO DE TÉRMINOS

SRI: Sistema para la recuperación de Información.

URL: (Uniform Resource Locator): Dirección de un recurso en la Web.

Web: La Web es la parte de Internet a la que accedemos a través del protocolo HTTP, es donde se publican y visitan los sitios y los documentos HTML.

Symfony: Marco de trabajo para la realización de aplicaciones web escrito en PHP5.

UCI: Universidad de las Ciencias Informáticas.

RUP: (Proceso Unificado de desarrollo de Software) Es un proceso de desarrollo de software y junto con el Lenguaje Unificado de Modelado UML, constituye la metodología estándar más utilizada para el análisis, implementación y documentación de sistemas orientados a objetos.

UML: Es el Lenguaje de Modelado Unificado para detallar, construir, visualizar y documentar las partes o artefactos de un software.

HTML: (Lenguaje de Marcado de Hipertexto) Es el lenguaje de marcado predominante para la elaboración de páginas web. Es usado para describir la estructura y el contenido en forma de texto, así como para complementar el texto con objetos tales como imágenes.

PHP: Lenguaje de programación interpretado, diseñado originalmente para la creación de páginas web dinámicas. Es usado principalmente en interpretación del lado del servidor (server-side scripting) pero actualmente puede ser utilizado desde una interfaz de línea de comandos o en la creación de otros tipos de programas incluyendo aplicaciones con interfaz gráfica usando las bibliotecas Qt o GTK+.

ORM: Componente de software que permite trabajar con las tablas de una Base de Datos como si ellas fueran orientadas a objetos (en este caso virtual). Debido a que lo standard es trabajar con BD relacionales, se deben realizar operaciones que permitan transformar un registro en objeto y viceversa. A esta funcionalidad se la llama Mapeo objeto-relacional (ORM).

UNIX: Es un sistema operativo portable, multitarea y multiusuario; desarrollado, en principio, en 1969 por un grupo de empleados de los laboratorios Bell de AT&T.

Windows: Es el nombre de la familia de sistemas operativos desarrollados por Microsoft desde 1981.

CGI: Es una importante tecnología de la World Wide Web que permite a un cliente (navegador web) solicitar datos de un programa ejecutado en un servidor web. CGI especifica un estándar para transferir datos entre el cliente y el programa. Es un mecanismo de comunicación entre el servidor web y una aplicación externa.

ANEXOS
Anexo 1: Pila del Sprint


Pila del Sprint			
Id. Tarea	Nombre y/o Descripción de la tarea.	Días	Estado
Sprint 1	Conjunto de tareas con mayor prioridad.	26	Resuelto
1	Investigar sobre el proceso de administración en otros motores de búsqueda.	3	Resuelta
2	Analizar el proceso de administración del motor de búsqueda Orión	5	Resuelta
3	Obtener del modelo de datos que permita almacenar todas las configuraciones gestionadas por la solución informática.	7	Resuelta
4	Obtener del script de la base de datos.	1	Resuelta
5	Implementar los prototipos de Interfaz de usuario necesarios para la administración web del motor de búsqueda.	10	Resuelta

Anexo 2: Pila del Producto

Pila del Producto			
Id. Tarea	Nombre y/o Descripción de la tarea.	Días	Estado
Sprint 1	Conjunto de tareas con mayor prioridad.	26	Resuelto
1	Investigar sobre el proceso de administración en otros motores de búsqueda.	3	Resuelta
2	Analizar el proceso de administración del motor de búsqueda Orión	5	Resuelta
3	Obtener del modelo de datos que permita almacenar todas las configuraciones gestionadas por la solución informática.	7	Resuelta
4	Obtener del script de la base de datos.	1	Resuelta
5	Implementar los prototipos de Interfaz de usuario necesarios para la administración web del motor de búsqueda.	10	Resuelta
Sprint 2	Conjunto de tareas con menor prioridad respecto al sprint 1.	35	Resuelta
6	Autenticar usuario	1	Resuelta
7	Gestionar Sección	5	Resuelta
8	- Insertar Sección	2	Resuelta
9	- Modificar Sección	2	Resuelta
10	- Eliminar Sección	1	Resuelta
11	Restablecer Secciones	3	Resuelta
12	Gestionar filtro	2	Resuelta
13	Gestionar Conversores	6	Resuelta
14	- Insertar Conversor	2	Resuelta
15	- Modificar Conversor	2	Resuelta
16	- Eliminar Conversor	2	Resuelta
17	Restablecer Conversores	1	Resuelta
18	Gestionar tipos Mime	6	Resuelta
19	- Insertar nuevos tipos Mime	2	Resuelta

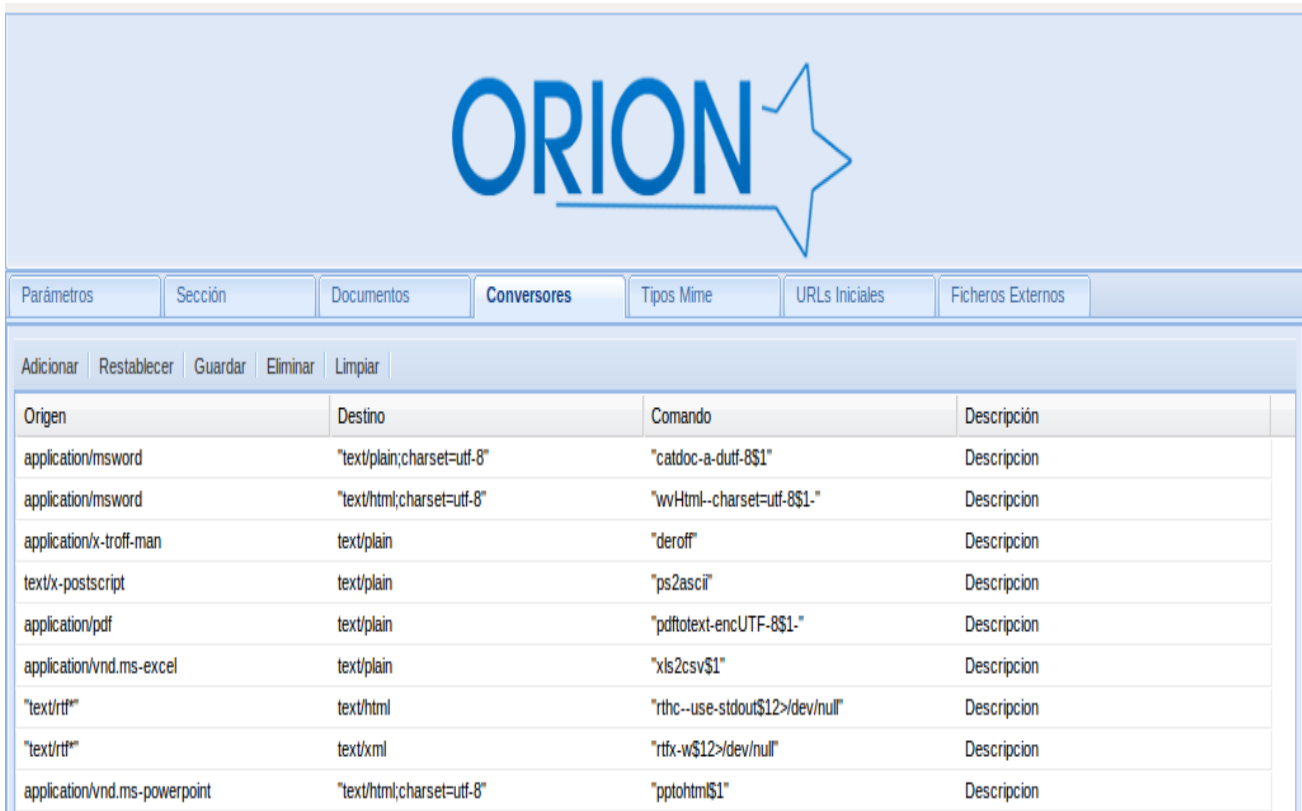
20	- Modificar tipos Mime	2	Resuelta
21	- Eliminar tipos Mime	2	Resuelta
22	Restablecer Tipos Mime	2	Resuelta
23	Gestionar URLs iniciales	6	Resuelta
24	- Insertar URL	2	Resuelta
25	- Modificar URL	2	Resuelta
26	- Eliminar URL	2	Resuelta
27	Gestionar Configuración General	5	Resuelta
28	Restablecer Configuración General	5	Resuelta
29	Importar Ficheros	3	Resuelta

Anexo 3: Vista del Requerimiento Funcional Gestionar Sección.



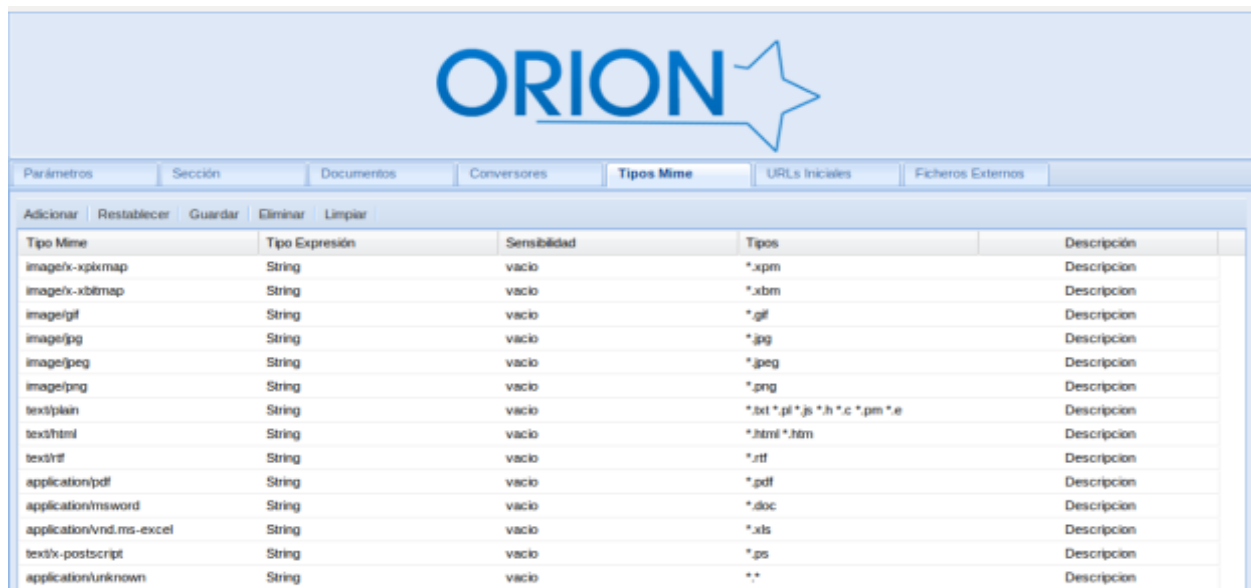
Parámetros			
Sección	Documentos	Conversores	Tipos Mime
URLs Iniciales	Ficheros Externos		
Adicionar Restablecer Guardar Eliminar Limpiar			
Nombre	ID_Sección	Peso	Descripción
body	1	300	Descripción
title	2	128	Descripción
meta.keywords	3	128	Descripción
meta.description	4	250	Descripción
url.file	6	0	Descripción
url.path	7	0	Descripción
url.host	8	0	Descripción
url.proto	9	0	Descripción
crosswords	10	0	Descripción
Charset	11	32	Descripción
Content-Type	12	64	Descripción
Content-Language	13	16	Descripción
CachedCopy	25	64000	Descripción

Anexo 4: Vista del Requerimiento Funcional Gestionar Conversores.




Origen	Destino	Comando	Descripción
application/msword	"text/plain;charset=utf-8"	"catdoc-a-dutf-8\$1"	Descripción
application/msword	"text/html;charset=utf-8"	"wvHtml-charset=utf-8\$1."	Descripción
application/x-troff-man	text/plain	"deroff"	Descripción
text/x-postscript	text/plain	"ps2ascii"	Descripción
application/pdf	text/plain	"pdftotext-encUTF-8\$1."	Descripción
application/vnd.ms-excel	text/plain	"xls2csv\$1"	Descripción
"text/rtf"	text/html	"rthc--use-stdout\$12>/dev/null"	Descripción
"text/rtf"	text/xml	"rtx-w\$12>/dev/null"	Descripción
application/vnd.ms-powerpoint	"text/html;charset=utf-8"	"ppthtml\$1"	Descripción

Anexo 5: Vista del Requerimiento Funcional Gestionar Tipos Mime.



Tipo Mime	Tipo Expresión	Sensibilidad	Tipos	Descripción
image/x-xpixmap	String	vacio	*.xpm	Descripción
image/x-xbitmap	String	vacio	*.xbm	Descripción
image/gif	String	vacio	*.gif	Descripción
image/jpeg	String	vacio	*.jpg	Descripción
image/peg	String	vacio	*.jpeg	Descripción
image/png	String	vacio	*.png	Descripción
text/plain	String	vacio	*.txt *.pl *.js *.h *.c *.pm *.e	Descripción
text/html	String	vacio	*.html *.htm	Descripción
text/rtf	String	vacio	*.rtf	Descripción
application/pdf	String	vacio	*.pdf	Descripción
application/msword	String	vacio	*.doc	Descripción
application/vnd.ms-excel	String	vacio	*.xls	Descripción
text/x-postscript	String	vacio	*.ps	Descripción
application/unknown	String	vacio	**	Descripción

Anexo 6: Vista del Requerimiento Funcional Gestionar Documentos



Parámetros						
Sección	Documentos		Conversores	Tipos Mime	URLs Iniciales	Ficheros Externos
Adicionar Restablecer Guardar Eliminar Limpiar						
Acción	Tipo Expresión	Exactitud	Sensibilidad	Tipos	Descripción	
Allow	String	0	Case	*.HTM	Descripción	
Disallow	String	0	0	*.b *.sh *.md5 *.rpm	Descripción	
Disallow	String	Match	NoCase	*.vdo *.mpeg *.mpe *.mpg *.avi *	Descripción	
Disallow	String	0	0	*.mid *.mp3 *.rm *.ram *.wav *.ail	Descripción	
Disallow	String	0	0	*.tex *.texi *.xls *.doc *.texinfo	Descripción	
Disallow	String	0	0	*.cpt *.bms *.oda *.tcl	Descripción	
Disallow	String	0	0	*.o *.a *.la *.so	Descripción	
Disallow	String	0	0	*.pat *.pm *.m4 *.am *.css	Descripción	
Disallow	String	0	0	*.map *.aif *.sit *.sea	Descripción	
Disallow	String	0	0	*.m3u *.qt	Descripción	
Disallow	String	0	0	*D=A *D=D *M=A *M=D *N=A *N	Descripción	
Disallow	Regex	0	0	\r[0-9][0-9]\$ \a[0-9][0-9]\$ \s\.	Descripción	
CheckOnly	String	0	0	*.arj *.tar *.zip *.tgz *.gz	Descripción	
CheckOnly	String	0	0	*.lha *.lzh *.rar *.zoo *.tar*.Z	Descripción	
CheckOnly	String	0	0	*.gif *.jpg *.jpeg *.bmp *.tiff	Descripción	
CheckOnly	String	0	0	*.vdo *.mpeg *.mpe *.mpg *.avi *	Descripción	
CheckOnly	String	0	0	*.exe *.cab *.dll *.bin *.class	Descripción	
CheckOnly	String	0	0	*.tex *.texi *.xls *.doc *.texinfo	Descripción	
CheckOnly	String	0	0	*.rtf *.pdf *.cdf *.ps	Descripción	
HrefOnly	String	0	0	*/mail*.html */thread*.html	Descripción	
CheckMP3	String	0	0	*.bin *.mp3 *.mio	Descripción	
CheckMP3Only	String	0	0	*.bin *.mp3	Descripción	

Anexo 7: Descripción de las tablas
Descripción ampliada de las tablas.

Nombre: admin_servidor_local		
Descripción: Tabla que contendrá los atributos del servidor local donde estará el sistema.		
Atributo	Tipo	Descripción
Id_servidor_local	integer	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica al servidor local.
ip	varchar	Cadena que tendrá la dirección ip del servidor local.

Tabla 1. Descripción de la tabla admin_servidor_local.

Nombre: admin_acceso					
Descripción: Tabla que contendrá los atributos relacionados con la forma en Orión tratará de acceder a los sitios web.					
Atributo	Tipo	Null	Descripción		
Id_acceso	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada acceso.		
nombre	varchar	si	Cadena que tendrá el nombre del acceso (nombre del servidor).		
url	varchar	si	Dirección url del acceso		
alias	varchar	si	Alias que tendrá el acceso		
tipo	varchar	si	El tipo de acceso (exp reg o string).		
usuario	varchar	si	Nombre de usuario general		
contrasenna	varchar	si	Contraseña general		
direccion_proxy	varchar	si	Dirección url del proxy		
proxy_usuario	varchar	si	Nombre de usuario del proxy específico		
proxy_contrasenna	varchar	si	Contraseña para acceder al proxy específico		
proxy_puerto			integer	si	puerto para conectarse al proxy específico
Id_servidor_local			integer	no	Llave heredada de la tabla admin_servidor_local

Tabla 2. Descripción de la tabla admin_acceso.

Nombre: admin_bd					
Descripción: Tabla que contendrá los atributos relacionados con la base de datos donde se guardarán las configuraciones de Orión de forma persistente.					
Atributo	Tipo	Null	Descripción		
Id_bd	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada base de datos.		
ruta	varchar	si	Dirección de la base de datos.		
usuario			varchar	si	Nombre del usuario con acceso a la bd
clave			varchar	si	Contraseña del usuario
tipo			varchar	si	El tipo de base de datos(postgres, mysql, oracle).
modo			varchar	si	Modo de funcionamiento de la Base de datos(single, multi, blob).
activa			boolean	si	Muestra el estado de la Base de datos.
Id_servidor_local			varchar	si	Identificador del servidor local del indexer
Id_codificación			varchar	si	Identificador de la codificación usada en la Base de Datos

Tabla 3. Descripción de la tabla admin_bd.

Nombre: admin_convertor				
Descripción: Tabla que contendrá los atributos relacionados con los conversores auxiliares que usará el motor de búsqueda.				
Atributo	Tipo	Null	Descripción	

Id_convertor	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada convertor.
activo	boolean	no	Estado que tendrán los conversores(activo o no)
tipo_origen	varchar	no	Tipo de documento origen
tipo_destino	varchar	no	Tipo de documento destino
comando	varchar	no	Comando para convertir el documento
descripcion	varchar	si	Descripción del convertor
Id_servidor_local	integer	no	Identificador del servidor local del indexer

Tabla 4. Descripción de la tabla admin_acceso.

Nombre: admin_documentos			
Descripción: Tabla que contendrá los tipos de documentos permitidos en el proceso de indexación.			
Atributo	Tipo	Null	Descripción
Id_documentos	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada tipo de documento.
accion	varchar	no	acción que tendrán los conversores(permitidos o denegados)
sensible_minusculas	boolean	no	Sensibilidad a mayúsculas
descripcion	varchar	si	Descripción sobre los documentos
tipos	varchar	no	Extensiones de los documentos permitidos o denegados
Id_servidor_local	integer	no	Identificador del servidor local del indexer

Tabla 5. Descripción de la tabla admin_documentos.

Nombre: admin_fichero			
Descripción: Tabla que contendrá los tipos de documentos permitidos en el proceso de indexación.			
Atributo	Tipo	Null	Descripción
Id_fichero	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada fichero externo.
activo	boolean	no	Estado de los fichero (activo, inactivo)
nombre	varchar	no	Nombre del fichero
ruta	varchar	no	Dirección del fichero en el disco duro.
descripcion	varchar	no	Descripción sobre los ficheros
extension	varchar	no	Extension del fichero
Id_servidor_local	integer	no	Identificador del servidor local del indexer

Tabla 6. Descripción de la tabla admin_fichero.

Nombre: admin_filtro			
Descripción: Tabla que contendrá los filtros asociados a las secciones.			
Atributo	Tipo	Null	Descripción
Id_filtro	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable

			que identifica cada filtro de las secciones.
accion	varchar	no	Acción aplicada a este filtro(bloquear, permitir)
exactitud	boolean	no	Exactitud del valor a filtrar(incluido, exacto)
valor	varchar	no	Criterio a filtrar
id_seccion	integer	no	Identificador de la sección a la que pertenece este filtro

Tabla 7. Descripción de la tabla admin_filtro.

Nombre: admin_mime			
Descripción: Tabla que contendrá los tipos mime.			
Atributo	Tipo	Null	Descripción
Id_mime	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada mime.
tipo	varchar	no	Acción aplicada a este filtro(bloquear, permitir)
tipo_expresión	varchar	no	Exactitud del valor a filtrar(incluido, exacto)
sensible_minusculas	boolean	no	Criterio a filtrar
exactitud	boolean	no	Exactitud del tipo mime.
extensiones	varchar	no	Tipos de ficheros asociados a cada tipo mime
descripcion	varchar	si	Descripción de cada tipo mime.
id_servidor_local	integer	no	Identificador del servidor local del indexer

Tabla 8. Descripción de la tabla admin_mime.

Nombre: admin_parametros			
Descripción: Tabla que contendrá los tipos mime.			
Atributo	Tipo	Null	Descripción
Id_parametros	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada parámetro o propiedad.
parametro	varchar	no	parámetro o propiedad a configurar
valor	varchar	no	Valor que contendrá la propiedad
descripcion	boolean	si	Descripción de la propiedad configurada
id_servidor_local	integer	no	Identificador del servidor local del indexer
id_lenguaje	integer	no	Identificador del lenguaje utilizado

Tabla 9. Descripción de la tabla admin_parametros.

Nombre: admin_seccion			
Descripción: Tabla que para almacenar las secciones de los documentos y sus pesos asociados.			
Atributo	Tipo	Null	Descripción
Id_seccion	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada parámetro o propiedad.
nombre	varchar	no	Nombre de la sección

peso	integer	no	Peso asignado a cada sección
tamaño_maximo	integer	si	Tamaño máximo de la sección
id_servidor_local	integer	no	Identificador del servidor local del indexer

Tabla 10. Descripción de la tabla admin_seccion.

Nombre: admin_servidor_local			
Descripción: Tabla que las secciones de los documentos y sus pesos asociados.			
Atributo	Tipo	Null	Descripción
id_servidor_local	integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica el servidor local del indexer
ip	varchar	no	Dirección ip del servidor

Tabla 11. Descripción de la tabla admin_servidor_local.

Nombre: admin_tag			
Descripción: Tabla que para almacenar los tag para agrupar sitios.			
Atributo	Tipo	Null	Descripción
Id_tag	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada tag para agrupar los sitios web.
nombre	varchar	no	Nombre de cada tag
periodo	integer	no	Tiempo de indización de cada tag
id_servidor_local	integer	no	Identificador del servidor local del indexer

Tabla 12. Descripción de la tabla admin_tag.

Nombre: admin_tc_codificacion			
Descripción: Tabla para almacenar las distintas codificaciones existentes.			
Atributo	Tipo	Null	Descripción
id_lenguaje	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada codificación.
valor	varchar	si	Valor de la codificación
nombre	varchar	si	Nombre de la codificación
descripcion	varchar	si	Identificador del servidor local del indexer

Tabla 13. Descripción de la tabla admin_tc_codificacion.

Nombre: admin_tc_lenguaje			
Descripción: Tabla que para almacenar los distintos lenguajes existentes.			
Atributo	Tipo	Null	Descripción
id_lenguaje	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada lenguaje.

nombre	varchar	no	Nombre del lenguaje
valor	varchar	no	valor del lenguaje
descripcion	varchar	si	Descripción de cada lenguaje

Tabla 14. Descripción de la tabla admin_tc_lenguaje.

Nombre: admin_url_iniciales			
Descripción: Tabla que para almacenar las url iniciales del motor de búsqueda.			
Atributo	Tipo	Null	Descripción
id_url_iniciales	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada url inicial.
metodo	varchar	si	Atributo que indicará el método de indización, check only, allow, entre otros.
subseccion	varchar	si	Este atributo es para indicar si es una pagina solamente o todo el sitio a partir de esta página.
url	varchar	no	Dirección url inicial
periodo	integer	si	Tiempo de reindización de esta url
Id_servidor_local	integer	no	Identificador del servidor local del indexer
Id_acceso	integer	si	Identificador del acceso a esta url
Id_tag	integer	si	Identificador del tag al que pertenece esta url

Tabla 15. Descripción de la tabla admin_url_iniciales.

Nombre: admin_restablecer			
Descripción: Tabla que para almacenar las configuraciones iniciales de las propiedades.			
Atributo	Tipo	Null	Descripción
id_restablecer	Integer	no	Es la llave primaria de la tabla. Valor numérico autoincrementable que identifica cada acción de restablecimiento.
propiedad	varchar	si	propiedad que será restablecida
valor	varchar	si	Valor inicial de la propiedad que será restablecida
descripcion	varchar	si	Descripción de la del restablecimiento
fecha	date	si	Fecha del valor de esta propiedad
id_servidor_local	integer	no	Identificador del servidor local del indexer

Tabla 16. Descripción de la tabla admin_restablecer.