

Universidad de las Ciencias Informáticas



Directorio del Buscador CubaSearch
Trabajo para optar por el título de Ingeniería en Informática

Autor:

Alden Hernández Gómez

Tutor:

Lic. José Albert Cruz Almaguer

Ciudad de la Habana.

Marzo de 2006

Declaración de Autoría

Yo: **Alden Hernández Gómez** me declaro como único autor de este trabajo y autorizo a la Universidad de las Ciencias Informáticas (UCI) y a la Entidad CubaSI para que hagan uso de este trabajo de la manera que estimen conveniente.

Y para que así conste firmo la presente a los _____ días del mes de _____ de 2006

Firma del Autor

Firma del Tutor

OPINIÓN DEL USUARIO DEL TRABAJO DE DIPLOMA

El Trabajo de Diploma, titulado: “**Directorio del Buscador CubaSearch**”, fue realizado en la Universidad de las Ciencias Informáticas. Este centro considera que, en correspondencia con los objetivos trazados, el trabajo realizado le satisface:

- Totalmente
- Parcialmente en un ____ %

Los resultados de este Trabajo de Diploma le reportan a CubaSI los beneficios siguientes:

Como resultado de la implantación de este trabajo se reportará un efecto económico que asciende a

Y para que así conste, se firma la presente a los ____ días del mes de _____ del año _____

Representante de la entidad

Cargo

Firma

Cuño

OPINIÓN DEL TUTOR DEL TRABAJO DE DIPLOMA

Título: Directorio del Buscador *CubaSearch*

Autor: Alden Hernández Gómez

El tutor del presente Trabajo de Diploma considera que durante su ejecución el estudiante mostró las cualidades que a continuación se detallan:

Por todo lo anteriormente expresado considero que el estudiante está apto para ejercer como Ingeniero Informático; y propongo que se le otorgue al Trabajo de Diploma la calificación de _____

Firma

Fecha

Agradecimientos

El agradecer a quienes han intervenido en la realización de este trabajo, el cual simboliza la culminación de mis estudios y muchos años de sacrificio no resulta para nada sencillo. Sin embargo si existen personas que merecen tratamiento especial y que no puedo dejar de mencionar, una de ellas es mi madre por ser mi guía y motor impulsor e inspirador en mí, además de que mucho de lo logrado hasta ahora se lo debo a ella por siempre estar presente, guiándome y aconsejándome. Agradecimiento especial merece mi novia, Daniela Hernández Laguna, de la cual he recibido su apoyo incondicional todo el tiempo. Agradezco también a todos mis compañeros de estudio y trabajo con los que diariamente convivo, Ramsés y Luís Alberto en especial, con los que he iniciado el largo camino del profesionalismo y perfeccionamiento humano.

Dedicatoria

El presente trabajo va dedicado a la memoria de mi abuela Mirta Ramos Pagán, mi segunda madre, quien estoy seguro estaría muy feliz y orgullosa al verme en este momento tan importante de mi vida, *“Ojalá y pudieras ver esto abue”*.

De forma general este trabajo lo dedico a aquellas personas que han contribuido de alguna forma en el desarrollo de este trabajo, a quien mejor que a ellos podría dedicarle este resultado.

Resumen

Hoy día la información en la Web se encuentra muy desorganizada y en muchos casos es redundante. Actualmente los sistemas de recuperación de información existentes en Internet suelen ser el punto de partida para todas las búsquedas de información. La entidad CubaSI, en aras de ganar usuarios ha decidido adicionar este servicio en su portal y en conjunto con la Universidad de las Ciencias Informáticas se ha propuesto desarrollar e implementar una herramienta que de soporte a esta necesidad.

Se crea entonces un proyecto encargado de llevar a cabo la materialización de esta herramienta. En este trabajo se explica el desarrollo del “Directorio del Buscador CubaSearch” elemento primordial en la estructura de cualquier sistema de búsqueda en Internet de hoy día y tiene como **objetivo concreto** diseñar una herramienta donde se agrupen los sitios cubanos con dominio .cu por categorías y sea capaz de mostrar esa información al usuario de forma tal que le permita realizar búsquedas sobre los sitios nacionales sin necesidad de recurrir a otro sistema de búsqueda existente en Internet.

Con este trabajo se crea una herramienta tipo Directorio, capaz de mostrar la información de la Web cubana, agrupada por categorías previamente establecidas, contiene asociado un motor de búsqueda que permite realizar búsquedas corrientes, este directorio servirá como complemento al primer buscador mixto de cuba, Buscador CubaSearch, servicio que será ofertado por el portal CubaSI.

En este documento se plasman los estudios realizados para llevar a cabo la construcción del sistema, se explican los principales conceptos necesarios a estudiar para entender el funcionamiento del mismo y se dejan algunas recomendaciones para el mejoramiento futuro de la herramienta.

Índice

INTRODUCCIÓN.....	1
FUNDAMENTACIÓN DEL TEMA.....	6
INTRODUCCIÓN.....	6
RECUPERACIÓN DE INFORMACIÓN	7
EL PROCESO DE RECUPERACIÓN DE INFORMACIÓN	8
SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN	8
<i>Buscadores</i>	10
Componentes Básicos.....	11
Arquitectura de un Motor de Búsqueda	11
<i>Metabuscadores</i>	12
<i>Directorios</i>	13
Formas de recuperar la Información:.....	13
Componentes Básicos.....	14
<i>Buscadores vs. Directorios</i>	15
Diferencias.....	15
OBJETO DE ESTUDIO	16
DESCRIPCIÓN DEL PROCESO ACTUAL.....	17
SITUACIÓN PROBLÉMICA.....	17
SISTEMAS AUTOMATIZADOS EXISTENTES VINCULADOS AL CAMPO DE ACCIÓN.....	18
SOLUCIONES EXISTENTES.....	18
<i>Ámbito Nacional</i>	18
<i>Ámbito Internacional</i>	19
PROPUESTA DE SOLUCIÓN.....	21
FUNDAMENTACIÓN DE LOS OBJETIVOS QUE SE PROPONE EL TRABAJO	22
<i>Objetivo General</i>	22
<i>Objetivos Específicos</i>	22
CONCLUSIONES.....	23
TENDENCIAS Y TECNOLOGÍAS ACTUALES A CONSIDERAR	24
INTRODUCCIÓN.....	24
PRINCIPALES ASPECTOS EN LOS SRI	25
CRITERIOS DE EVALUACIÓN DE INFORMACIÓN.....	25
MODELOS CONCEPTUALES DE RECUPERACIÓN DE INFORMACIÓN	26
<i>Modelos booleanos</i>	27
<i>Modelo vectorial</i>	28
MÉTODOS DE ORGANIZACIÓN DE LA INFORMACIÓN	29
<i>Ficheros Inversos o Invertidos</i>	29
MODELOS DE BÚSQUEDA	31
<i>Operadores Booleanos</i>	32
<i>Operadores Posicionales</i>	32
Operadores de Existencia	33
Operadores de Truncamiento.....	33
Operadores Límite/Comparación.....	33
<i>Criterios de Asignación de Relevancia</i>	34
APLICACIONES WEB	35
<i>Arquitectura de Aplicaciones Web</i>	35
<i>Lenguajes para la Web</i>	36
FUNDAMENTACIÓN DEL LENGUAJE UTILIZADO	40
METODOLOGÍA UTILIZADA.....	41
<i>Unified Modeling Language (UML)</i>	42

HERRAMIENTAS UTILIZADAS.....	43
<i>GoldParser</i>	43
<i>Macromedia Dreamweaver MX 2004</i>	43
<i>Herramientas CASE</i>	43
<i>Visual Studio .NET</i>	43
GESTOR DE BD UTILIZADO.....	44
CONCLUSIONES.....	45
DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA.....	46
INTRODUCCIÓN.....	46
MODELO DE DOMINIO.....	47
REQUISITOS FUNCIONALES.....	49
REQUISITOS NO FUNCIONALES.....	51
DESCRIPCIÓN DEL SISTEMA.....	53
<i>Operadores booleanos</i>	54
<i>Gramática del Lenguaje de Consulta</i>	54
<i>Actores del Sistema</i>	55
<i>Modelo de Casos de Uso del Sistema</i>	56
<i>Expansión de los Casos de Uso</i>	57
CONCLUSIONES.....	70
CONSTRUCCIÓN DE LA SOLUCIÓN PROPUESTA.....	71
INTRODUCCIÓN.....	71
DIAGRAMA DE CLASES.....	72
DISEÑO DE LA BASE DE DATOS.....	77
<i>Modelo de Datos</i>	78
PRINCIPIOS DE DISEÑO.....	79
<i>Estándares de la Interfaz de la Aplicación</i>	79
<i>Formato de Salida de los Reportes</i>	80
<i>Concepción de la Ayuda</i>	80
<i>Tratamiento de Excepciones</i>	81
ESTÁNDARES DE CODIFICACIÓN.....	81
Declaraciones.....	82
MODELO DE DESPLIEGUE.....	82
MODELO DE IMPLEMENTACIÓN.....	84
<i>Explicación de los componentes</i>	87
<i>Otros componentes</i>	87
CONCLUSIONES.....	90
CONCLUSIONES.....	91
RECOMENDACIONES.....	92
REFERENCIAS BIBLIOGRÁFICAS.....	93
BIBLIOGRAFÍA.....	94
ANEXOS.....	95

INTRODUCCIÓN

Se plantea que Internet surgió en la década de los 60 del pasado siglo, pero no es hasta 1991 con la aparición de la World Wide Web, con posibilidades de multimedia y el surgimiento de los browsers (Netscape, Explorer), que toma gran auge y demanda en el mundo entero. En la actualidad son millones los usuarios y computadoras que se conectan y hacen uso de la gran RED DE REDES. En Internet se puede encontrar todo tipo de información, y el número de sitios y páginas Web que la componen está en el orden de los mil millones y crece exponencialmente.

A medida que crecía Internet, y se desarrollaba la WWW. realizar una búsqueda se hacia cada vez más complejo, esto trajo consigo que se desarrollaran nuevas herramientas que ayudaran en este proceso, fue entonces cuando surgieron los Directorios¹ y Buscadores² como alternativa para solucionar este problema ya que ambas herramientas localizan y facilitan el acceso a la información compartida en la red, estos sistemas de búsqueda en la Web hoy día son ampliamente utilizados en Internet, puesto que a través de ellos es posible encontrar la información deseada de manera rápida y eficiente. Actualmente suelen ser el punto de partida para la navegación en Internet, por este motivo la proliferación de la herramienta ha sido tan grande que se estima existen más de 2000 mientras que en 1995 apenas existía una docena.

Los Buscadores o Motores de Búsqueda permiten realizar la búsqueda corriente por palabras o términos, estos muestran la información almacenada en su base de datos relacionada con los términos de la búsqueda. Los Directorios, por su parte, muestran la Web organizada por temas, las búsquedas se realizan navegando en profundidad sobre el directorio, haciéndose cada vez más específicas las categorías que van apareciendo.

¹ Herramienta de búsqueda de Internet, agrupada por categorías donde las búsquedas se hacen descendiendo mediante estas hasta llegar a la información deseada.

² Herramientas de búsqueda de Internet que permite buscar información a través de palabras claves.

En la actualidad los buscadores tienden a ser sistemas mixtos, que no es otra cosa que una mezcla entre motores de búsqueda y directorios, además de tener características de buscadores, presentan la Web agrupadas en catálogos temáticos.

Los directorios temáticos (también llamados índices o directorios) se exploran mediante navegación (*browsing*). Proporcionan además un acceso a la información de forma estructurada (implementada generalmente en forma de árbol), atendiendo a algún criterio de clasificación en categorías. El usuario puede descender por los niveles de especificidad hasta encontrar la información adecuada a sus intereses sin necesidad de formular con términos su consulta. Tanto la selección, como la clasificación e indexación de recursos es efectuada por personas. [1]

Existen directorios que presentan motores de búsqueda en su estructura, lo cual es muy ventajoso y útil ya que el hecho de realizar una búsqueda corriente dentro de una categoría de interés permite rápidamente acotar la búsqueda a las páginas que son relevantes para el usuario.

Actualmente CubaSI, que es uno de los principales portales Web de Cuba no dispone de un buscador mixto, ni tan siquiera de un directorio de búsqueda, esto trae consigo la siguiente **situación problemática**: Es imposible desde CubaSI realizar una búsqueda sobre sitios nacionales con dominio .cu

Por esto se hizo necesario crear un grupo de trabajo integrado por estudiantes y profesores de esta Universidad en conjunto con la Entidad CubaSI, para desarrollar una aplicación denominada **CubaSearch. El Buscador de Cuba** con el objetivo de satisfacer la **necesidad** de encontrar una manera dentro del portal CubaSI de buscar información sobre los sitios de Cuba con dominio .cu, creando un buscador mixto que satisfaga esta necesidad de búsqueda de información.

El **objeto de estudio** de este trabajo es la Entidad CubaSI, prestigioso portal completamente dinámico que dedica sus servicios a la comercialización de sus productos. De aquí se deriva que el **campo de acción** es el desarrollo de una herramienta de búsqueda, en específico un directorio integrado a un buscador mixto para dicha entidad.

Como **Hipótesis** se plantea que si se desarrolla la aplicación basada en un potente Gestor de Base de Datos como Microsoft SQL Server y se realiza un diseño de la aplicación bien elaborado, es posible lograr que la búsqueda a través del directorio se desarrolle de una manera rápida y eficiente.

El **objetivo general** del trabajo es: desarrollar una Herramienta Web de búsqueda de información para el portal CubaSI sobre los sitios cubanos con dominio .cu, que muestre la información de la Web agrupada por categorías y permita realizar búsquedas más específicas, esta herramienta estará contenida en lo que será el buscador mixto de CubaSI.

De este objetivo general se derivan los siguientes **objetivos específicos**.

1. Crear una herramienta de búsqueda Web en forma de directorio que agrupe a los sitios cubanos en categorías y subcategorías previamente definidas, siguiendo una estructura de árbol.
2. Hacer un estudio de las herramientas de recuperación de información existentes en especial de motores de búsqueda y directorios.
3. Diseñar un motor de búsqueda en el directorio como servicio adicional del mismo.
4. Garantizar a los usuarios un servicio de búsqueda rápido y fácil de usar.
5. Ofrecer servicio de suscripción online de sitios Web Cubanos al directorio.

Para dar cumplimiento a los objetivos propuestos se desarrollan un número de **tareas** las cuales se enumeran a continuación:

- ✓ Hacer un análisis de los métodos de búsqueda que actualmente presenta CubaSI.
- ✓ Identificar las necesidades que alberga la Entidad.
- ✓ Analizar el desarrollo que presentan las herramientas de búsqueda de información en el ámbito nacional e internacional.
- ✓ Examinar los requerimientos que debe cumplir el sistema.
- ✓ Describir los procesos que se desarrollan en el sistema.
- ✓ Modelar conceptualmente las clases que se desarrollan en el sistema.
- ✓ Desarrollar los diagramas que ayuden al entendimiento del sistema y contribuyan en la realización de un diseño óptimo.
- ✓ Diseño de la Base de Datos eficiente que soporte las funcionalidades que exige el sistema.

- ✓ Elaborar una interfaz de navegación, agradable, sencilla y fácil de utilizar.

Se pretende obtener un producto que cumpla con los estándares de diseño y satisfaga las exigencias actuales de búsqueda de información. Se utiliza la metodología RUP para el desarrollo del sistema y el UML como lenguaje para describir su estructura.

El directorio que estará integrado al buscador mixto ha sido implementado en una interfaz Web, lo menos sobrecargada posible, pocas imágenes, nada de banners para así lograr mayor rapidez al momento de cargar la página en el instante de realizar una búsqueda. Además contiene un servicio de suscripción online de sitios, que permite a cualquier persona solicitar que se incluya determinado sitio al directorio.

Con este trabajo se crea un directorio automatizado, formará parte del primer buscador mixto cubano, además posee una interfaz Web agradable y sencilla. La principal ventaja es que se crea uno de los componentes fundamentales de un buscador actual, un directorio cubano situado en uno de los portales Web más importantes y visitados de Cuba, que permitirá realizar búsquedas temáticas y por palabras en la red sobre los sitios cubanos con dominio .cu.

El directorio no recorre la Web, solo registra algunos datos de los sitios como, título, url, descripción o resumen, de las categorías que lo componen registra: nombre, descripción, subcategorías y sitios que están dentro de esta. Los resultados de la búsqueda, estarán determinados por la información que se haya suministrado al directorio cuando se realiza la indexación. Esta indexación se realiza de forma manual.

Para realizar una búsqueda en la Web usando el directorio es muy sencillo, este presenta la información organizada en una colección de categorías, por lo tanto el usuario lo único que necesita es ir seleccionando la categoría que el crea contenga el tema que busca. A medida que vaya seleccionando la búsqueda esta se hace más profunda, otra manera de realizar una búsqueda en el directorio es de forma corriente, o sea entrando una frase que registrará el patrón de búsqueda, con el valor agregado de que buscará los sitios que se relacionen con estas frases dentro de la categoría en la que esté ubicado el usuario, esta funcionalidad acota notablemente la búsqueda en el directorio.

Más que buscar información sobre las palabras de la página, los resultados serán presentados haciendo referencia al contenido y temática de la Web registrada. Es decir, si se busca "ordenadores", aparecerán Sitios que de una forma u otra estén relacionados con la temática de ordenadores.

Este trabajo consta de 4 capítulos, y cada uno aborda lo siguiente:

Capítulo 1: Se describen y explican conceptos en el mundo de la recuperación de la información, se explica además como se lleva a cabo el proceso de búsqueda dentro del portal y los principales problemas que presenta. Se menciona y explica el funcionamiento de las principales herramientas de búsqueda existentes hoy día en el mundo.

Capítulo 2: Aborda todo lo referente a las tecnologías a usar, aquí se explica de modo más amplio cómo funcionan los buscadores, se fundamenta la metodología y lenguaje utilizado.

Capítulo 3: Aquí se describe el negocio a través de un modelo de dominio, se realiza el análisis del sistema y se definen los procesos del mismo usando las herramientas de modelación y el lenguaje UML.

Capítulo 4: Capítulo en que se trata la construcción del sistema propuesto, se modelan las clases necesarias en la implementación de la herramienta.



FUNDAMENTACIÓN DEL TEMA

Introducción

En el presente capítulo se brindan los aspectos y conceptos relacionados con la recuperación de información, se explica el funcionamiento y características de las principales herramientas de búsqueda de información existentes. Se abordan de manera detallada los principales conceptos relacionados con el dominio del problema para poder entender mejor el sistema y la propuesta de solución brindada.

Recuperación de Información

La recuperación de información es una actividad práctica que el ser humano realiza, consciente e inconscientemente, casi continuamente y en el marco de cualquier actividad, la necesidad de resolver una duda o documentar una afirmación, son expresiones clásicas de los procesos de recuperación de información. Con el desarrollo de los sistemas digitales de procesamiento de datos y tratamiento de la información, las técnicas de recuperación de información han ido desarrollando un conjunto de teorías y aplicaciones prácticas que subyacen en la actualidad a cualquier búsqueda y recuperación de información que tiene lugar en Internet.

La recuperación de información o Information Retrieval (IR) es el conjunto de tareas mediante el cual el usuario accede a los recursos de información que son pertinentes para dar solución a la necesidad de obtener nueva información sobre un asunto o material de su interés.

Conceptualmente es una operación en la que se interpreta una necesidad de información de un usuario y se seleccionan los documentos más relevantes capaces de solucionarla, es decir, consiste en buscar documentos que respondan de la mejor manera a la pregunta formulada.

En el contexto de la WWW puede definirse como la identificación de una o más referencias de páginas Web que resulten relevantes para satisfacer una necesidad de información.

Un sistema de Recuperación de Información es aquel que, con distintas técnicas, proporciona acceso rápido automático a grandes colecciones de objetos que contienen información de cualquier tipo.

Un modelo de recuperación de información, que se puede considerar “genérico”, tiene las siguientes fases: [1]

1. Definición de las necesidades informativas del usuario.

2. Selección y ordenación de las fuentes y recursos de información a utilizar. La selección de los índices, directorios y motores de búsqueda a utilizar es de suma importancia. Cada vez aparecen más directorios especializados en los más diversos temas.
3. La consulta de los recursos de información que comprende la traducción de la necesidad del usuario a una expresión en el lenguaje de interrogación propio de cada sistema y la ejecución de dicha expresión.
4. Evaluación de los resultados. Se seleccionan y obtienen los documentos que responden a las necesidades del usuario. Si los resultados obtenidos no son pertinentes implica un replanteamiento de la consulta.
5. Presentación al usuario.

El proceso de Recuperación de Información

Consta de dos partes fundamentales:

Interfaz de Usuario: Gestiona toda la interacción con el usuario

- Entrada de las consultas.
- Visualización de los resultados.

Operaciones de Consultas: Transforma la consulta para mejorar la búsqueda.

- Transforma la consulta que se realiza en lenguaje natural para que el sistema entienda lo que se quiere buscar.

Sistemas de Recuperación de Información

Un sistema de recuperación de información comprende una serie de pasos básicos:

1. Generar una representación del significado o contenido de cada documento incluido en la colección a partir de su descripción. Este proceso es la **indexación**.
2. Generar una representación del significado de la consulta de usuario, llamado **formulación**.
3. Comparar estas dos representaciones para seleccionar aquellos documentos que tienen mayor probabilidad de satisfacer la consulta (**comparación**). [1]

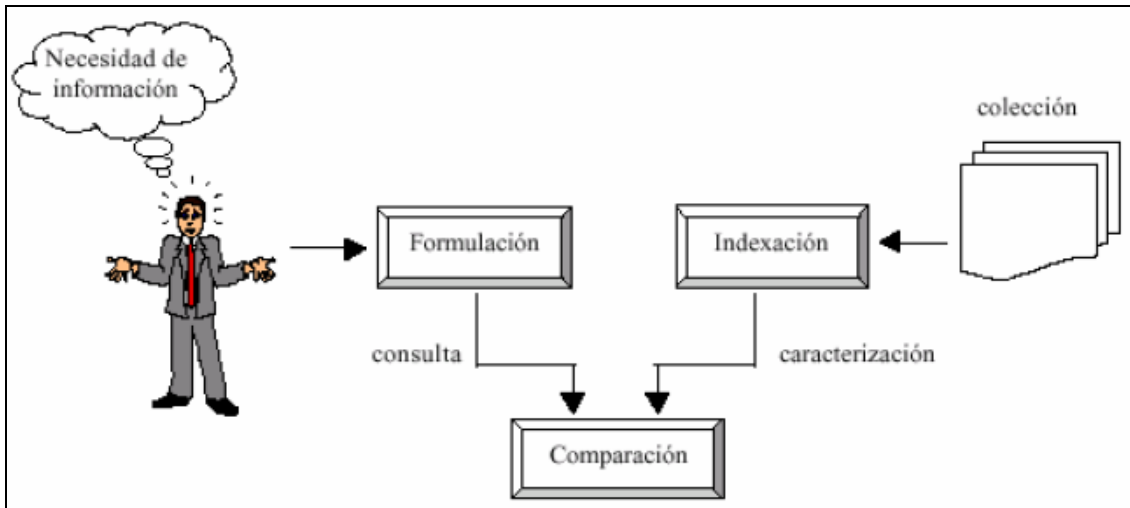


Figura 1 Paradigma de la recuperación de información

Actualmente existen dos formas básicas de recuperación de información en la Web:

- Búsqueda
- Navegación

Entre las principales herramientas utilizadas en Internet para buscar información, se encuentran las siguientes:

- Buscadores o Motores de Búsqueda (Búsqueda)
- Metabuscadore (Búsqueda)
- Directorios (Navegación)
- Buscadores Mixtos (Ambas)

Es válido señalar que actualmente estas herramientas suelen ser el punto de partida en la navegación de Internet, es por esto que su proliferación ha sido tan grande, que actualmente se estima existen mas de 2000 mientras que en 1995 solo se conocía de una docena. [1]

El principal objetivo de estas herramientas es ayudar al usuario a encontrar documentos relevantes a un tema buscado de forma rápida y eficiente. Pero para lograr este objetivo. Se tienen que enfrentar a grandes problemas como son:

- La información que se encuentra en Internet no está ordenada.

- La información en Internet es volátil. Muchos documentos que cambian o desaparece totalmente.
- La información en Internet es en muchos casos redundante.
- Gran volumen de datos que requiere manejar, miles de millones de documentos disponibles, además de que la información en la Web crece exponencialmente.
- El problema de como representar toda esa información en una Base de Datos.
- El problema de como buscar en ella para obtener los documentos recuperados.
- El problema de la precisión y relevancia de los documentos recuperados.

Además de: dada una necesidad de información (consulta + usuario +...) y un conjunto de documentos, se debe de ordenar los documentos de mas a menos relevancia para esta necesidad y presentarlos como resultado.

Buscadores

Consideran la Web como una gran Base de Datos de texto y tienen como objetivo “Indexar toda la Web”.

Actualmente son los más utilizados, emplean un mecanismo que recorre automáticamente la Web en busca de nuevos documentos y actualizaciones en los ya indexados. Este mecanismo es realizado por un software, “Spider” que de cada documento recupera su información y la almacena en una BD. Una de las principales ventajas que presenta es que de cada url visitada es posible guardar la mayoría de sus elementos, título, texto íntegro, texto HTML, vínculos, permitiendo realizar búsquedas sobre cada uno de estos campos, otra ventaja importante es que mediante la Spider es posible actualizar rápidamente los cambios producidos en Internet.

Para realizar una búsqueda utiliza lo que se conoce como “**Motor de Búsqueda**” esta herramienta es muy útil y tiene un gran peso en el éxito de cualquier buscador, el motor de búsqueda trabaja tomando los términos introducidos por la persona que está realizando la búsqueda y devuelve los documentos que se encuentran relacionados con dichos términos, la forma más habitual de hacer esta consulta es con lenguaje natural.

Componentes Básicos [3]

- **Spider:** Recorre la Web buscando las páginas para indexar.
- **Indexador:** Gestiona el índice con la información de las páginas localizadas.
- **Motor de Búsqueda:** Realiza la búsqueda sobre el índice de documentos
- **Interfaz:** Interactúa con el usuario.

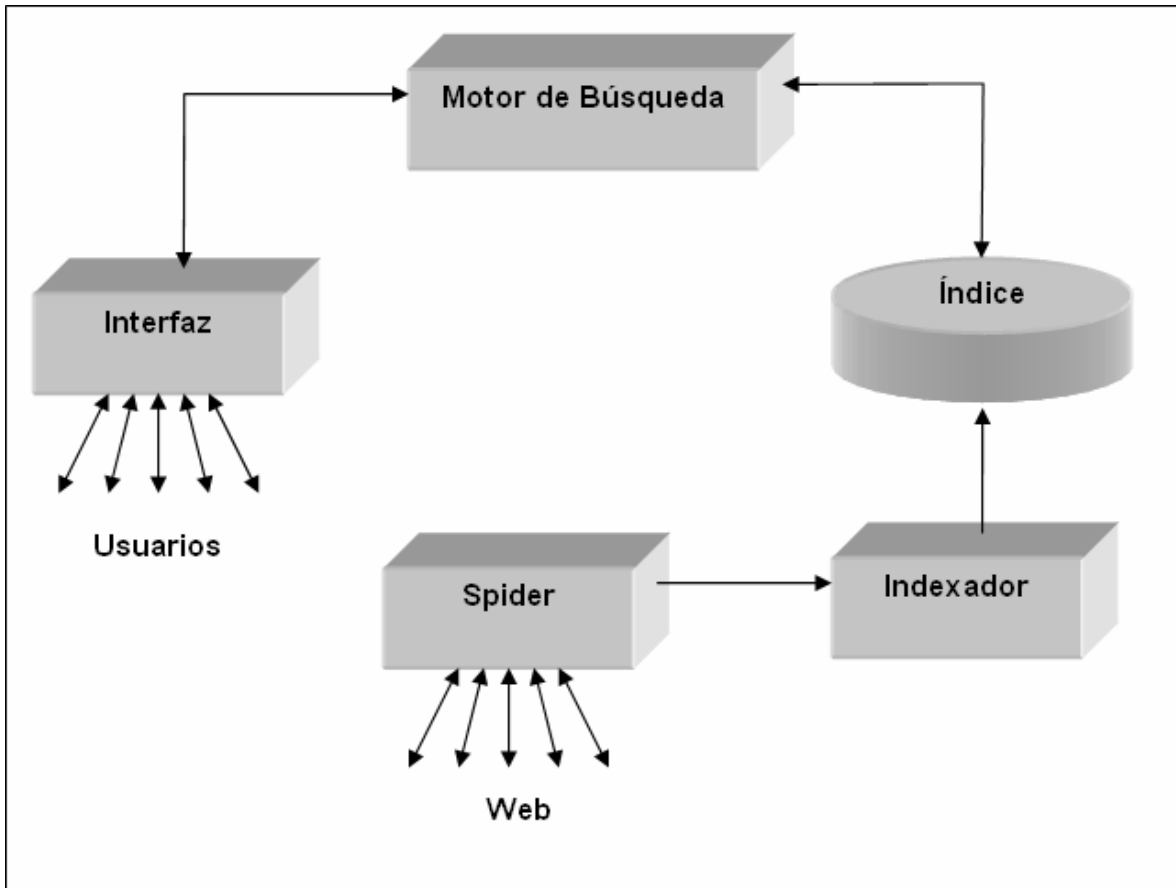


Figura 2 Componentes de un Buscador

Arquitectura de un Motor de Búsqueda.

Un motor de Búsqueda recibe una consulta de un usuario, que consiste en el ingreso de una o varias palabras claves, realiza la búsqueda en la Base de Datos y extrae una lista ordenada de documentos que responde a la consulta de forma total o parcial. El orden de las respuestas depende de una puntuación que se asocia a cada documento cuando se realiza la búsqueda, esta puntuación depende de cómo sea definida por los creadores de

la herramienta, puede ser cuanto antes aparecen las palabras o por proximidad entre estas.

Operacionalmente, para recuperar determinada información, el usuario realiza una consulta (query), la cual llega al motor de búsqueda y donde es tomada por un componente (Analizador de Consultas) que se encarga de transformar la consulta del usuario en un conjunto de incógnitas posibles en base a las palabras claves ingresadas por el usuario. Esto se realiza de acuerdo con el conjunto de términos indexados que posee el buscador. Una vez que se tienen las posibles incógnitas, se construye la información que se mostrará al usuario, para lo cual se envían las palabras claves posibles al generador de frases. El generador de frases selecciona la serie de títulos posibles para cada palabra buscada.

Metabuscadore

Metabuscadore. En realidad, no son buscadores. Lo que hacen, es realizar búsquedas en auténticos buscadores, analizan los resultados de la página, y presentan sus propios resultados. No disponen de estructuras de indexación propias, reenvían las búsquedas a otros sistemas de búsqueda.

No suelen ser bienvenidos por los buscadores. Para utilizar los servicios gratuitos de un buscador de esta forma, es necesario pedir permiso. El motivo es el siguiente: El Buscador, pone el dinero para operar el servicio, los contenidos que utilizará el metabuscador, y no percibe nada a cambio. Al eliminar la publicidad, no se obtienen ingresos. Solo gasto y pérdida de visitantes que utilicen este servicio de búsqueda.

Entre las principales ventajas de este sistema de búsqueda se encuentra que combina resultados de múltiples buscadores, elimina los resultados duplicados y confiere mayor peso a aquellos resultados que aparecen en varios buscadores y como desventaja que es mayor el tiempo de espera, además de que explota los servicios ofrecidos por otros sistemas de búsqueda.

Directorios

Los directorios son unos de los primeros sistemas de búsqueda existentes, se dice que el primero que hubo fue el directorio de Yahoo, el cual opera desde 1993, hoy día se estima que tiene alrededor de 3.000.000 de páginas indexadas.

Este tipo de buscador presenta la información de la Web en forma de una colección de categorías, y las búsquedas se realizan en profundidad, una vez que se selecciona una categoría, se muestran las subcategorías contenidas en ella, además de los sitios que se relacionan directamente con dicha categoría

La indexación se realiza de forma manual por lo editores del sistema. Presenta una estructura en forma de árbol, en el que se agrupa en las ramas principales las categorías o temas más generales y amplios, por ejemplo (artes, deportes, computación), obsérvese que estas categorías son capaces de encerrar dentro de ellas a muchas otras. El hecho de que los sitios se incluyan de forma manual implica que los resultados de la búsqueda sean mejores y que el ruido sea casi nulo, esto quiere decir que los resultados se corresponden en un gran % al tema de búsqueda, una desventaja es que el costo de operación es relativamente alto al ser el factor humano fundamental en el funcionamiento del mismo. A diferencia de los buscadores, se centran en la calidad y no en la cantidad. Como principal ventaja está la gran calidad de sus contenidos y desventaja con respecto al crecimiento que tiene la Web y la velocidad de clasificación.

Un ejemplo importante de Directorio es el Open Directory Project (ODP), basado en miles de editores voluntarios con más de 6.000.000 millones de páginas indexadas y más de 60.000 editores.

Formas de recuperar la Información:

- Búsqueda.
- Navegación.
- Ambas:
 - El usuario navega hasta un punto de la jerarquía.
 - La búsqueda se realiza sobre ese subárbol de jerarquía.

Componentes Básicos [3]

- **Clasificador** Personas que recorren la Web clasificando páginas relevantes. Requiere gran cantidad de recursos humanos.
- **Indexador**
- **Motor de Búsqueda**
- **Interfaz**

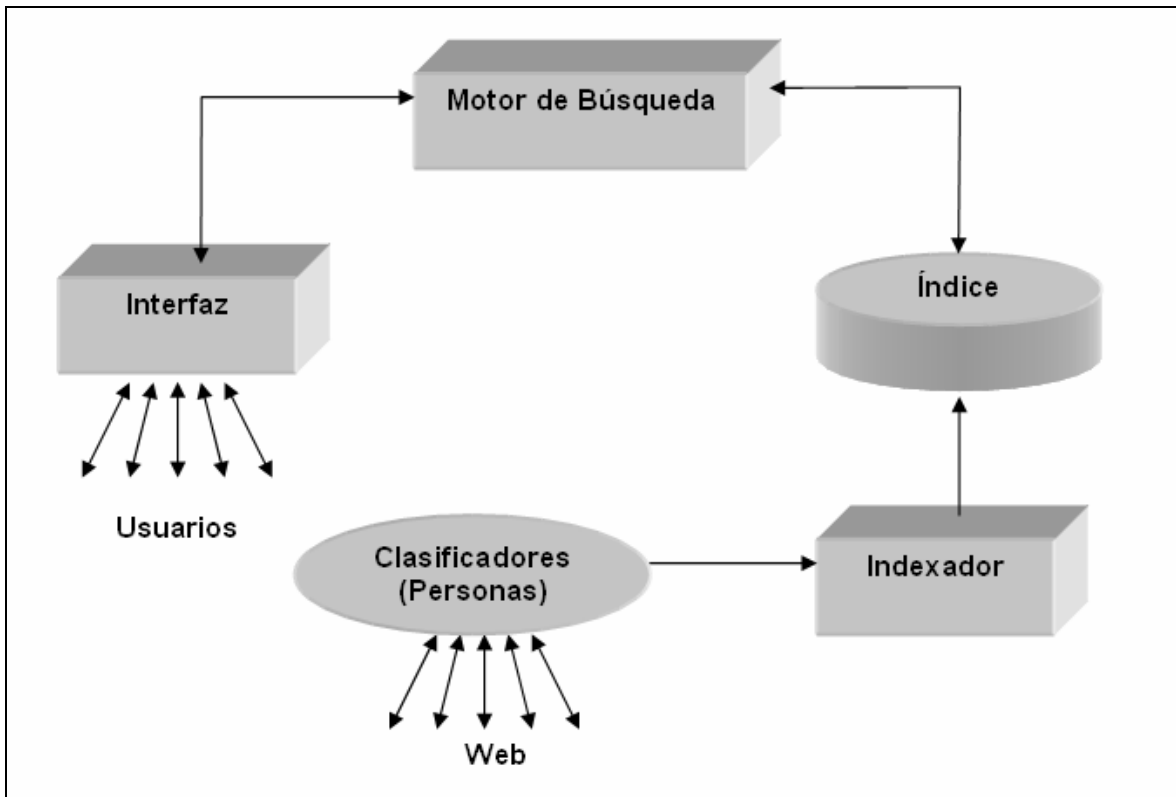


Figura 3 Componentes de un Directorio

Buscadores vs. Directorios

Se emplea el término buscador para referirse a motor de búsqueda.

Diferencias

Si bien no presentan muchas diferencias en la interfaz de usuario puesto que muchos buscadores presentan en su estructura directorios y viceversa, en el momento de indexar la información de la Web las diferencias se hacen notables.

Existen actualmente, dos maneras principales de indexar la información de Internet:

1. Indexación manual:

Es el mecanismo empleado en los sistemas de búsqueda con estructura de Directorio. En este caso son los propios Webmaster o los responsables del sistema los encargados de llevar a cabo la indexación de forma manual de documentos bajo alguna de las categorías previamente establecidas en el directorio. Entre sus problemas se encuentra la lentitud en el análisis y actualización de la información, otro problema es el de “confusión clasificatoria”, esto es que la mayoría de los directorios no tienden a usar ningún sistema de clasificación establecido, sino que los propios responsables son los encargados de establecer las materias, las cuales a veces no son muy intuitivas, confundiendo al usuario.

2. Indexación Automática:

Es el tipo de indexación en los sistemas tipo Buscadores o Motores de Búsqueda. Se auxilian de un programa denominado “spider” que se encarga de ir moviéndose por todas las páginas Web, simulando la confección de una gran telaraña e ir recolectando la información de los documentos por donde va pasando y almacenándola. Entre las principales ventajas se encuentra la rápida actualización de los cambios producidos en Internet, además, al indexar prácticamente todos los términos del documento, la búsqueda se puede llevar a cabo por todos estos. Como principal desventaja, señalar que la precisión en la identificación del documento no es tan exacta como la manual, además muchas páginas pueden atraer la atención hacia sí repitiendo palabras muy solicitadas por los usuarios, otra desventaja es que la indexación de este tipo solo reconoce textos puros, y en ocasiones, ficheros de otro formato, imágenes por ejemplo, brindan información relevante también sobre un tema en específico.

De esta diferencia, indexación manual o automática se derivan muchas otras que a continuación se abordan:

Cobertura de red: Los directorios cubren mucho menos documentos que los buscadores ya que el proceso automático es mucho más rápido.

Profundidad de Clasificación: Los Directorios clasifican los documentos bajo encabezamientos de materias amplios, mientras que los motores permiten clasificarlo por cada una de las palabras (no vacías³) que contiene el documento.

Actualidad de los Enlaces: Los motores tienden a tener los enlaces más actualizados que los directorios.

Resultados de Búsqueda: Generalmente se obtienen más resultados utilizando motores de búsqueda que Directorios, aunque es válido señalar que muchos de estos resultados pueden ser ruidos mientras que en los directorios el nivel de ruido es casi nulo.

Actualmente la tendencia que hay es incorporarle a los directorios motores de búsqueda, de forma que existan dos maneras de utilizar los directorios, la primera es como ya se ha explicado, navegando sobre las categorías y la otra es realizando una búsqueda corriente, se incluye una frase o palabra y se genera una consulta a la Base de Datos del directorio, lo novedoso es que los resultados siempre serán del directorio, por lo que, independientemente que sean menos, (es muy poco probable que exista la misma cantidad de sitios indexados en la BD del directorio que en la de un Motor de Búsqueda) estos resultados serán más específicos.

Objeto de Estudio

El objeto de estudio de este trabajo es la recuperación o búsqueda de información que tiene lugar en el Portal **CubaSI** sobre los sitios nacionales con dominio .cu

CubaSI se presenta como una Unidad de Negocios de Datos. Entidad subordinada a la Empresa de Telecomunicaciones de Cuba S.A. (ETECSA) líder en la provisión de

³Palabras no vacías o stopword, término que se refiere a las palabras muy comunes en el lenguaje y que aparecen un gran número de veces en cualquier texto, ejemplo artículos, preposiciones, conjunciones, etc.

servicios de Internet en Cuba, dedica sus acciones a la comercialización de productos y servicios y brinda las mejores prestaciones a sus clientes. En su portal es posible encontrar información sobre el ámbito cultural, actualidad en general, música, multimedia, presenta vínculos destinados al turismo que responden a la categoría de ventas en línea, ellos son: Cubasishopping, serCUBA, reserva Hoteles, LatinCuba, Llamadas a Cuba, y entre los principales servicios que brinda se encuentran: prensa en Cuba, Clic.Internet, Mapa Interactivo de Cuba, Fotorreportajes, Radio Cubana en vivo, TV en vivo, Horóscopo, Directorio de Discos.

El Portal CubaSI se destaca por la actualidad de sus informaciones en formato noticioso y de multimedia. Ofrece variados servicios y tiendas de artículos en línea, así como espacios para anuncios publicitarios de negocios.

Descripción del Proceso Actual

En la página principal o portada de CubaSI se publican noticias de carácter nacional, en su mayoría culturales (Literatura, Música, Artes). El espacio que corresponde a la noticia principal es para el acontecimiento más importante del momento, no interfiere el tipo de noticia que sea, brinda información de personalidades de cuba en sitios bibliográficos que describen la vida de grandes figuras de la historia de cuba como Martí y el Ché.

Presenta un buscador interno, el cual permite efectuar búsquedas corrientes y ofrece resultados que estén relacionados directamente con el sitio, contenidos en la BD de CubaSI, la búsqueda no vas mas allá de su propia frontera, en su mayoría los resultados que se muestran son noticias que ya han sido publicadas.

Situación Problemática

La principal dificultad en torno a la búsqueda y recuperación de información en CubaSI es que no presenta una herramienta capaz de ofrecer información contenida en sitios nacionales, salvo de aquellos que estén vinculados al portal que en su mayoría no ofrecen mucha información porque están destinados a otro fin, al ser de tipo comercial están concebidos para ofrecer o promocionar servicio y brindar propaganda a sus productos.

El buscador interno que presenta, como ya se explicó en el sub-epígrafe anterior no realiza búsquedas fuera de la frontera del portal, y solamente brinda resultados de noticias que estén vinculadas a los términos de la búsqueda pero que hayan sido publicadas anteriormente, esto tiene 2 inconvenientes principales, el primero es con el tipo de resultado, si bien las noticias son muy importantes porque informan y contribuyen a mantener al usuario al corriente de lo que sucede en la actualidad, es un tipo de información que trata un tema específico resumiendo y abordando lo principal del mismo, en su mayoría presentan un formato corto por lo que generalmente no es usado como fuente bibliográfica en si, no está concebida para eso, el segundo inconveniente tiene que ver con la fecha, en el resultado de la búsqueda se muestran noticias que tuvieron lugar en momentos anteriores y por consiguiente no tienen mucha relevancia.

Sistemas Automatizados existentes vinculados al campo de acción

CubaSI presenta un buscador interno que permite efectuar búsquedas corrientes y ofrece resultados de noticias que ya hayan sido publicadas en el portal.

Soluciones Existentes

Existen en el mundo numerosas herramientas de búsqueda muy capaces y potentes, a continuación se enumeran algunos de los sistemas de recuperación de información más importantes y usados en Cuba e Internet y se explican las principales características de alguno de ellos.

Ámbito Nacional

El portal www.cuba.cu presenta el servicio Sitios Cubanos que brinda la posibilidad de encontrar información cubana de primera mano, organizada por categorías para facilitar la búsqueda sobre los más variados temas del acontecer del país. <http://sitioscubanos.cuba.cu/>, no presenta servicio de suscripción online, para realizar una solicitud de inclusión de sitio debe enviarse un correo a la siguiente dirección sitioscubanos@citmatel.inf.cu. No tiene incorporado ningún motor de búsqueda. Esta hecho en PHP y solo presenta una profundidad en las categorías de nivel 2.

Otro de los portales donde es posible encontrar un directorio es el de www.cubaweb.cu, este portal se presenta como una guía cubana con noticias, referencias y comercio electrónico, el servicio de búsqueda que presenta en forma de directorio contiene asociado un motor de búsqueda que realiza la misma buscando coincidencias exactas entre los términos entrados por el usuario y el título, url o resumen de los sitios que tienen indexados, no permite el uso de operadores booleanos, otra de las dificultades que presenta es que agrupa las categorías en un árbol de nivel 2 de profundidad, al igual que el directorio del portal cuba.cu.

Ámbito Internacional

Buscadores o Motores de Búsqueda

Nombre	Url
Google	www.google.com
AltaVista	www.es.altavista.com
MSN	www.search.msn.com
Webcrawler	www.webcrawler.com

Directorios

Nombre	Url
Directorio de Google	www.directory.google.com
The Open Directory	http://dmoz.org (editado por miles de voluntarios)
Yahoo	www.yahoo.com

MetaBuscadores

Nombre	Url
Vivísimo	www.vivisimo.com
Metacrawler	www.metacrawler.com

Buscadores Especializados

Nombre	Url
En Noticias -Google News	www.news.google.com
Científico -Scirus	www.scirus.com
Compras -Yahoo compras	www.shopping.yahoo.com

Google - www.google.com

¿Como activa su éxito google.com? Es un resultado alfanumérico que se inicia al consultar una duda, ahí, en una espiral de resultados positivos, se activan 54.000 servidores, luego 100.000 procesadores a los que activan 8.000 millones de páginas Web e inducen a una potencia a la "n" de cifra digital, que en una abrir y cerrar de ojos se obtiene una respuesta, en la mayoría de los casos afirmativa. Es una ecuación de 500 millones de variables posibles que Google resuelve en un clic. Estas consultas ascienden a más de 200 millones por día.

Yahoo – www.search.yahoo.com

Mención especial merece la transformación de yahoo!, cuya novedad más destacada es que el usuario puede personalizar sus búsquedas a través del servicio "my yahoo! Search". De este modo, el internauta puede ir guardando así los enlaces que desee e incluso introducir comentarios personales. Posteriormente, en la opción "mi Web", el internauta encontrará sólo aquellas páginas que haya consultado con anterioridad, similar a un servicio de favoritos. Además, el portal permite almacenar direcciones de sindicación de contenidos con el fin de rescatarlas y utilizarlas más tarde.

MSN - www.search.msn.com

MSN ha presentado su nuevo motor, una versión más simplificada que parece estar inspirada en Google, el nuevo MSN Search no solo proporciona a los usuarios una larga lista de enlaces Web, sino también en proporcionarles acceso a las respuestas e informaciones que se busca, el nuevo MSN Search incluye un nuevo motor, índice y rastreador de búsquedas, todo ello creado desde cero con tecnología Microsoft.

Altavista – www.altavista.com

Buscador de Web, imágenes, MP3/audio, video, noticias. También ofrece otras herramientas como un traductor, filtro familiar, servicio de páginas amarillas, buscador de personas y comparador de precios de diferentes productos a través de DealTime (<http://altavista.dealtime.com>). La base de datos actual está ordenada según la fórmula de relevancia de AltaVista.

Propuesta de Solución

Luego de haber realizado un análisis sobre las herramientas de búsqueda de información existentes y haber estudiado las necesidades y características del objeto de estudio se infiere la necesidad de implementar y adicionar una herramienta que satisfaga la necesidad de búsqueda de información que tiene lugar en el portal CubaSI, y se ajuste a la política y requerimientos de la Entidad.

Independientemente que existan en el mundo y en Cuba, sistemas de búsqueda capaces de realizar consultas y mostrar la información de la Web agrupada por categorías, se hace necesario crear una herramienta de búsqueda que se adecue a las características y políticas de trabajo establecidas en CubaSI, que contenga un directorio el cual mostrará la información de la Web cubana agrupada en categorías de la forma que decidan los editores y administradores de la Herramienta y sea capaz de ofrecer resultados de sitios cubanos con dominio .cu.

La aplicación se corresponderá con las políticas de Seguridad, confiabilidad e integridad de los datos, definidos en CubaSI.

Fundamentación de los objetivos que se propone el trabajo

A continuación se abordan los objetivos que se proponen para dar respuesta a la situación problemática planteada.

Objetivo General

El objetivo es desarrollar una herramienta Web en forma de Directorio que se integre a un buscador mixto para el portal CubaSI, donde estarán organizados jerárquicamente por categorías y subcategorías todos los sitios cubanos con dominio .cu

Objetivos Específicos

1. Crear una herramienta de búsqueda Web en forma de directorio que agrupe a los sitios cubanos con dominio .cu en categorías y subcategorías, siguiendo una estructura de árbol.
2. Hacer un estudio de las herramientas de recuperación de información existentes, en especial motores de búsqueda y directorios.
3. Hacer un estudio de las ventajas que conlleva implementar un motor de búsqueda en el directorio.
4. Facilitar a los usuarios nacionales e internacionales el acceso al índice de todos los sitios cubanos mediante el diseño de una herramienta que permita realizar búsquedas temáticas por categorías y subcategorías, previamente definidas y que sea estable y fácil de usar.
5. Ofrecer servicio de suscripción online de sitios Web Cubanos al directorio.

Conclusiones

En este capítulo se han explicado las cuestiones importantes que rodean al objeto de estudio, se determinaron las condiciones específicas en el entorno al problema y en base a esto se obtuvo un objetivo general y otros específicos para este trabajo. Independientemente que este capítulo sea totalmente teórico es sumamente importante que sea elaborado con sumo detalle puesto que las ideas y conceptos que se estudian aquí, contribuyen notablemente a fomentar los cimientos del desarrollo de este trabajo.

Capítulo 2

TENDENCIAS Y TECNOLOGÍAS ACTUALES A CONSIDERAR

Introducción

En el presente capítulo se realiza un estudio de los mecanismos y observaciones a tener en cuenta para implementar un sistema de recuperación de información (SRI), los aspectos más importantes a considerar en la elaboración del mismo. Se explica el porqué de la metodología y lenguaje utilizado para implementar el sistema y se detallan algunas características que justifican tal elección.

Principales aspectos en los SRI

Cuando se usa un Sistema de Recuperación de Información se chequean aspectos importantes que hacen de este un elemento útil y eficiente a simple vista para las personas que lo utilizan, por ejemplo, la rapidez con que muestra los resultados y sobre todas las cosas el nivel o medida en que este satisface la consulta realizada, es por esto que en el momento de desarrollar un sistema de búsqueda, debe diseñarse de forma tal que se optimicen lo mejor posible estos aspectos.

Criterios de Evaluación de Información.

Un Sistema de Recuperación de Información puede ser evaluado por diversos criterios, incluyendo entre los mismos a: *eficaz ejecución*, *efectivo almacenamiento*, *efectividad en la recuperación* y otra serie de características que puede ofrecer al usuario y hacen del sistema una herramienta cómoda y fácil de usar. La relativa importancia de estos factores debe ser decidida por el diseñador del sistema y la selección de la apropiada estructura de datos y algoritmos para su implementación dependerá de esa decisión.

La *eficacia en la ejecución* es medida por el tiempo que se toma un sistema o una parte de un sistema para realizar una operación, es decir, el tiempo entre el pedido y la respuesta. Este parámetro ha sido siempre una de las preocupaciones principales en un Sistema de Recuperación de Información, un largo tiempo en obtener la respuesta interfiere con la utilidad del sistema, y llega a alejar a los usuarios del mismo.

La *eficiencia del almacenamiento* es medida por el número de bytes que se precisan para almacenar los datos. Una medida común para medir la eficacia de almacenamiento, es la razón del tamaño del índice de los ficheros además del tamaño de los archivos del documento sobre el tamaño de los archivos del documento.

$$\text{Exceso de espacio} = \frac{\text{tamaño}_{\text{índice}} + \text{tamaño}_{\text{documentos}}}{\text{tamaño}_{\text{documentos}}}$$

De forma tradicional se ha conferido mucha importancia a la *efectividad de la recuperación*, normalmente basada en la relevancia de los documentos, lo cual ha

representado un problema ya que medir la relevancia es un proceso subjetivo y sin confianza. Esto es, diferentes juicios personales asignarían diferentes valores de relevancia a un documento recuperado en respuesta a la búsqueda hecha. La seriedad del problema es la materia de debate, bastantes investigadores señalan que la subjetividad del juicio sobre la relevancia no es suficiente para invalidar el sistema. Muchas medidas de la efectividad de la recuperación han sido propuestas. Las más empleadas, de forma general, son las conocidas como *relevancia* y *precisión*.

La *relevancia* (*recall*) es la proporción entre el número de los documentos relevantes recuperados en una búsqueda dada y el número de documentos relevantes para esa búsqueda en la base de datos. Excepto para pruebas realizadas sobre pequeñas colecciones, este denominador es generalmente desconocido y debe ser estimado por muestreo o por otros métodos.

$$\text{Recall} = \frac{\text{Cantidad de documentos relevantes recuperados}}{\text{Cantidad de documentos relevantes}}$$

La precisión es la proporción entre el número de documentos relevantes recuperados y el número total de documentos recuperados.

$$\text{Precisión} = \frac{\text{Cantidad de documentos relevantes recuperados}}{\text{Cantidad de documentos recuperados}}$$

El rango de valores de ambas medidas está entre 0 y 1.

Otras medidas para evaluar la calidad de la recuperación son:

- Cubrimiento, que es la cantidad de información relevante que incluye el sistema.
- La forma de presentación de los resultados.
- El esfuerzo realizado por el usuario para lograr las respuestas a su pedido.

Modelos Conceptuales de Recuperación de Información

Un modelo de Recuperación de Información queda definido cuando se fijan en un sistema en particular lo siguiente: cómo se obtienen las representaciones de los documentos y de la consulta, la estrategia para evaluar la relevancia de un documento respecto a una

consulta, los métodos para establecer la importancia (orden) de los documentos recuperados y los mecanismos que permiten una retroalimentación por parte del usuario para mejorar la consulta.

Por lo tanto es posible afirmar que un modelo de recuperación de información es una tupla $\langle D, Q, F, R(q_i, d_j) \rangle$ donde:

D: Representación lógica de los documentos.

Q: Representación lógica de los requerimientos de información (queries).

F: Marco para modelar documentos, consultas y sus relaciones.

$R(q_i, d_j)$: Función de ranking que asocia la relevancia de d_j con respecto a q_i

Algunos ejemplos de modelos de recuperación de información son:

- Modelos booleanos.
- Modelo vectorial.

Modelos booleanos.

Este modelo representa los documentos por un conjunto finito de términos de indexación, cada uno de los cuales se trata de una variable booleana que vale verdadero si la característica asociada está presente en el documento y falso si no lo está. En general las características son las palabras de la colección por lo tanto esta variable tomara su valor en dependencia si la palabra aparece o no en el documento, esta variable no tiene asociada ningún peso, esto quiere decir que no determina si la palabra aparece muchas veces o solo una vez.

Las consultas de usuario se expresan en términos de expresiones booleanas arbitrarias, formadas por palabras enlazadas por los operadores lógicos estándar AND, OR, y NOT. En los Sistemas de Recuperación de Información mas sofisticados el usuario emplea lenguaje natural, que el sistema convierte a expresiones booleanas.

La relación entre un documento y una consulta es una función booleana que devuelve 1 si el documento contiene las palabras expresadas en la consulta según los operadores relacionales y 0 si no la satisface.

En general, los modelos booleanos son preferibles a otros modelos en términos de los requerimientos computacionales, que son bajos, tanto en el espacio requerido como en la rapidez de consulta. Sin embargo fallan en la eficiencia de la recuperación.

Modelo vectorial

Este modelo está basado en que cada documento de la colección está representado por un vector n-dimensional, donde n es la cardinalidad del conjunto de términos de indexación elegidos para la colección de documentos, en el que cada componente representa el peso del término previamente asociado a esa dimensión. Este peso representa un estimado de la utilidad del término como descriptor del documento, es decir, de la utilidad para distinguir ese documento del resto de los presentes en la colección (un término recibe un peso igual a 0 en los documentos en los que no aparece). Los pesos pueden reflejar diferentes medidas, por ejemplo, la frecuencia de aparición de un término en un conjunto dado (técnica conocida como TF, de uso generalizado). Es bueno señalar que una representación como ésta, basada sólo en la existencia de los términos, sacrifica información sobre el orden en que los términos ocurren, información sintáctica, etc.

Las consultas formuladas por los usuarios, al estar compuestas también por términos, son representadas de la misma forma que los documentos; posibilitando el cálculo de la semejanza entre una consulta y un documento a partir de la obtención del producto escalar de ambos vectores.

Este modelo es el que más resultados está dando en cuanto a eficacia hoy día. Su inconveniente es que requiere un esfuerzo de procesamiento muy superior al resto de los modelos.

Se conoce que la frecuencia de un término mejora la relevancia en la recuperación de información, pero no siempre mejora la precisión. Como los términos frecuentes tienden a aparecer en muchos textos, estos tienen poco poder discriminante. Para resolver esto se eliminan los términos de mucha frecuencia. Encontrar los umbrales óptimos es el principal problema de esta solución. [1]

Métodos de organización de la Información

Después de elegir el modelo conceptual a utilizar en el sistema de Recuperación de Información es necesario optar por un modelo de organizar la información de los documentos, es decir, seleccionar que tipo de estructura de fichero se va a usar para la Base de Datos subyacente. Y esta decisión no es independiente del modelo conceptual considerado, ya que ciertos modelos conceptuales implican el uso de una determinada organización de la información y viceversa.

La información puede o no estar indexada. Y dentro de la primera, es decir de la información indexada, se encuentra principalmente:

- **Ficheros invertidos**

Los ficheros no indexados se conocen como ficheros planos.

La idoneidad de indexar la información viene dada por varios factores, siendo los más importantes la frecuencia de actualización y el volumen de la misma. Así, una colección de documentos que varía frecuentemente no parece adecuada para ser indexada, ya que no resulta eficiente, puesto que el tiempo que se pierde en el esfuerzo invertido en la indexación de la colección de documentos no se corresponde con el tiempo ganado en la hora de resolver la consulta. Por otra parte, en el caso donde el tamaño de la colección de documentos sea demasiado grande, el único modo de responder a la consulta en un tiempo razonablemente será disponiendo de un índice. En el caso de una colección muy grande con una tasa de actualización muy alta se tendrán que utilizar técnicas de actualización de índices o alguna solución de compromiso.

Ficheros Inversos o Invertidos

La utilización de la técnica del fichero invertido es un elemento clásico en los sistemas de recuperación de información textual. Dada la gran cantidad de información contenida en los documentos textuales, una página Web por ejemplo, los procedimientos clásicos de búsqueda secuencial o de ficheros indexados no son capaces de responder de manera adecuada a los requerimientos de velocidad y exactitud en la respuesta necesarios para satisfacer al usuario. Por esta razón, estos sistemas usan una especialización de los ficheros indexados, desarrollando un nuevo tipo de fichero, al que se conoce como fichero invertido.

Un fichero invertido es un tipo de fichero indexado donde cada ítem tiene la forma:

Identificador índice	Identificadores de Documento	Identificador de Campo
----------------------	------------------------------	------------------------

Un Identificador índice es único para cada palabra clave, frase o término con significado semántico que describe al menos un documento de la colección de documentos. El fichero invertido contiene a estos términos índices ordenados de forma alfabética. El identificador de documento es único para cada documento. Cada término índice tiene asociado una lista de los identificadores de documentos que contienen ese término. El identificador de campo se utiliza para almacenar información útil dependiendo de los requerimientos del sistema de recuperación de información. Por ejemplo, el identificador de campo puede indicar dentro de qué campo del documento aparece el término índice. Algunos sistemas incluyen también información acerca de la localización del término en el documento, por ejemplo, el párrafo o frase en que el término aparece dentro del documento.

Si se utiliza el modelo conceptual de espacio vectorial, para cada identificador de documento asociado a un término índice puede incluirse su peso en el documento. Un ejemplo de fichero invertido es el que se muestra a continuación:

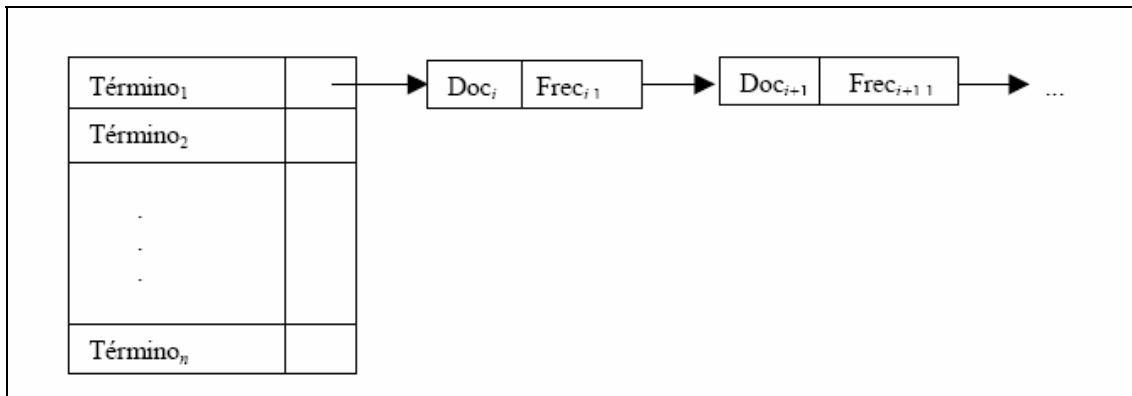


Figura 4 Esquema general de un fichero invertido

La construcción de un fichero invertido comprende a grandes rasgos los siguientes pasos:

- Conseguir una lista de las palabras que aparecen en el texto, junto con su localización en el mismo.
- Invertir la lista anterior, es decir, conseguir una lista de términos ordenados lexicográficamente y por orden de aparición con las localizaciones asociadas a los términos.

- Opcionalmente procesar posteriormente el fichero invertido, añadiendo pesos a los términos, reorganizándolos o comprimiéndolos.

El fichero invertido puede usarse asociado al modelo conceptual booleano o al modelo de espacio vectorial.

En este tipo de estructura se suele aplicar:

- Operaciones de normalización: eliminar signos de puntuación, múltiples espacios, todo a minúsculas.
- Eliminación de palabras comunes (stopwords), esto contribuye considerablemente a disminuir el tamaño del índice.

¿Qué información se indexa?

- Palabras
- +su posición en el texto
- +información de relevancia
- +información de enlaces entre documentos (PageRank)

Y además de una breve descripción de cada documento indexado, título, url, resumen, tamaño, fecha de actualización, etc.

Modelos de Búsqueda

Las necesidades de información del usuario son expresadas a través de una ecuación de búsqueda, que en el caso más simple consta de palabras sueltas, denominadas *palabras clave* y, en otros, puede ser una complicada combinación de operadores, palabras clave y paréntesis. Algunos buscadores permiten expresar las consultas en lenguaje natural siendo el motor de búsqueda el responsable de traducir dicha expresión a un formato estructurado sobre el cuál aplicar los algoritmos correspondientes.

Independientemente de la forma en que el usuario exprese la consulta, esta será analizada por el motor de búsqueda y traducida a una representación interna que permita compararla con los términos recogidos en la base de datos y seleccionar las direcciones URL que sean más relevantes

En general los buscadores ofrecen la posibilidad de realizar dos tipos de búsquedas: una simple, en la cual se especifican tan solo las palabras claves a buscar, y otra avanzada en

la que el usuario puede, usando operadores, construir expresiones bastante complejas, para la confección de estas últimas normalmente se provee al sistema de una interfaz amigable que no le exija a los usuarios el tener que memorizar los operadores, ni la sintaxis.

Las búsquedas avanzadas suelen ofrecer la posibilidad de utilizar distintos tipos de operadores:

- Operadores booleanos.
- Operadores posicionales.
- Operadores de existencia.
- Operadores de truncamiento.
- Operadores de límite/comparación.

Operadores Booleanos

Estos operadores son muy utilizados en los sistemas de recuperación de información. Los tres operadores básicos son: suma/unión (OR), producto/intersección (AND) y el operador resta o división (NOT). La combinación de dichos operadores da como resultado operaciones más complejas.

Para ayudar a los usuario a usar estos operadores es conveniente facilitar algunos servicios que ayuden a generar las consultas usando frases como “*Buscar resultados con todas las palabras*” en vez del operador AND o “*Buscar resultados con alguna de las palabras*” en vez del operador OR o “*Buscar resultados que no contengan*” para hacer uso del operador NOT.

Operadores Posicionales

Estos operadores tienen como objetivo superar algunas limitaciones que tienen los operadores booleanos. Toman como punto de partida la valoración del término dentro del contexto en el que se encuentre, se dividen en dos grupos:

Posicionales absolutos: Permiten buscar un término en un lugar específico del documento, ejemplo el título o url.

Posicionales relativos o de proximidad: Son operadores para establecer la posición de un término respecto a otro. Se pueden buscar palabras que estén juntas, separadas por varias palabras o caracteres, que se encuentren en la misma frase o párrafo e incluso si

se debe o no respetar el orden en que se han introducido los términos. Un operador posicional muy común es NEAR.

Operadores de Existencia

Estos operadores tienen como fin forzar la presencia de determinados términos en los documentos recuperados, para precisar que aparezca un determinado término se usa el operador '+' y para obligar la exclusión de un determinado término, suele usarse el operador '-'

Operadores de Truncamiento

En ocasiones se hace necesario a la hora de formular una búsqueda usar no solo un término simple sino sus derivados, fijados por prefijación o sufijación, mínimas variantes léxicas, etc. Los operadores de truncamiento facilitan este tipo de búsqueda. Se trata de operadores, generalmente símbolos como *, \$, cuya presencia puede sustituir uno o un conjunto de caracteres, situados a la derecha, izquierda o centro del término especificado.

Operadores Límite/Comparación

Estos operadores fijan las cotas para la búsqueda, estas pueden ser tanto numéricas como alfabéticas, dichos operadores suelen ser del tipo "mayor que", "menor o igual que". Estos operadores están destinados a documentos contenedores de información numérica.

Criterios de Asignación de Relevancia

Los buscadores utilizan un algoritmo para darle un orden a los documentos que devolverán en la consulta, este aspecto contribuye enormemente en su éxito, el orden en que un resultado de búsqueda satisface a un usuario constituye la piedra angular de un buscador. Generalmente cada buscador se reserva la forma en que está implementado dicho algoritmo. A continuación se listan una serie de aspectos posibles a tener en cuenta para medir el grado en que un determinado documento puede tener más relevancia en una búsqueda que otro.

- Frecuencia de la palabra o frase de la consulta en el documento. Generalmente se da prioridad a las páginas que contienen un mayor número de veces la palabra clave que aparece en la consulta.
- Longitud del documento. Si es corto y contiene repetidamente los términos de la búsqueda, tiene prioridad sobre otro más extenso que también contenga reiteraciones de la palabra clave.
- Grado de presencia del conjunto de palabras claves en el documento recuperado.
- Presencia de las palabras en importantes partes del documento como, título, url, o al inicio del texto.
- Presencia de las palabras en la etiqueta META del documento, utilizada para poner información que describa el documento.
- Grado de popularidad del documento, es decir si el documento es muy citado por otros documentos, este criterio constituye la base de éxito del famoso buscador Google con su tecnología de PageRank.

En el caso del directorio un criterio posible a tener en cuenta para asignar la relevancia es que los sitios situados en categorías más generales, se consideran más relevantes que los que estén situados en categorías más específicas.

Aplicaciones Web

Una aplicación Web puede ser considerada como un sistema donde la entrada del usuario (entrada de datos y navegación), afecta el estado del negocio. Esta definición intenta establecer que una aplicación Web es un sistema de software con estado de negocio, y por tanto puede elaborarse utilizando la metodología de trabajo que propone RUP, y utilizando el UML para los modelos.

Para UML una Aplicación Web es un software diseñado para automatizar procesos cuyo resultado se entrega a través de Internet o una Intranet. [4]

Arquitectura de Aplicaciones Web

La arquitectura básica de una aplicación Web incluye los navegadores, la red y el servidor Web.

El servidor Web distribuye páginas de información formateada a los clientes que las solicitan. Los requerimientos son hechos a través de una conexión de red, y para ello se usa el protocolo HTTP.

En algunos ambientes de desarrollo de aplicaciones Web, las páginas contienen código HTML y scripts dinámicos, que son ejecutados por el servidor antes de entregar la página, ejemplos: ASP (Active Server Pages), JSP (Java Server Pages), PHP (Hypertext Preprocessor), etc.

Algunas páginas incluyen scripts del lado cliente que son interpretadas por el navegador (código JavaScript o VBScript, mezclados con código HTML). Estos scripts definen el comportamiento dinámico adicional para la página y pueden interactuar con el navegador, el contenido de la página y los controles adicionales contenidos en ella (Applets de Java, controles Active X y plug-ins).

El usuario ve e interactúa con el contenido de la página; la forma más común de interacción es a través de formularios (Form), donde el usuario introduce la información en campos de entrada (textbox, textarea, checkbox, etc.) y los envía al servidor para

procesarlos. El usuario también puede interactuar con el sistema navegando a diferentes páginas a través de los hipervínculos.

Lenguajes para la Web

Los lenguajes de programación para la Web se clasifican en dos partes fundamentales que reconocen la propia arquitectura Cliente/Servidor de esta plataforma de desarrollo: los lenguajes del lado del Servidor y los lenguajes del lado del Cliente.

Un lenguaje del lado del servidor es aquel que se ejecuta en el servidor Web, justo antes de que se envíe la página a través de Internet al cliente. Las páginas que se ejecutan en el servidor pueden realizar accesos a bases de datos, conexiones en red, y otras tareas para crear la página final que verá el cliente. Como ventaja señalar que el cliente no puede ver los scripts, ya que se ejecutan y transforman en HTML antes de enviarlos. Además son independientes del navegador del usuario, ya que el código que reciben es HTML fácilmente interpretable.

Entre los principales lenguajes del lado del servidor se encuentran por el auge que estos han tenido el: Common Gateway Interface (CGI) comúnmente escritos en Perl, Active Server Pages (ASP y ASP.NET), Hipertext Preprocesor (PHP), y Java Server Pages (JSP). Estos se caracterizan por desarrollar la lógica de negocio dentro del Servidor, además de ser los encargados del acceso a Bases de Datos, tratamiento de la Información etc. Del lado del cliente se encuentran principalmente el JavaScript y el Visual Basic Script, que son los encargados de aportar dinamismo a la aplicación en los navegadores.

CGI

Es el sistema más antiguo que existe para la programación de las páginas dinámicas de servidor. Actualmente se encuentra un poco desfasado por diversas razones entre las que destaca la dificultad con la que se desarrollan los programas y la pesada carga que supone para el servidor que los ejecuta.

Los CGI se escriben habitualmente en el lenguaje Perl, sin embargo, otros lenguajes como C, C++ o Visual Basic pueden ser también empleados para construirlos.

Algunas desventajas de la programación en CGI son las siguientes:

- Los resultados se escriben directamente con el CGI, así que el código del programa se mezcla con el del HTML haciendo difícil su comprensión y mantenimiento.
- Cada programa CGI que se pone en marcha lo hace en un espacio de memoria propio. Así, si tres usuarios ponen en marcha un CGI a la vez se multiplicará por tres la cantidad de recursos que ocupe ese CGI. Esto significa una grave ineficiencia. [5]

Perl

Es un lenguaje de programación muy utilizado para construir aplicaciones CGI para la Web. Perl es un acrónimo de Practical Extracting and Reporting Language, indicando que se trata de un lenguaje de programación muy práctico para extraer información de archivos de texto y generar informes a partir del contenido de los ficheros.

Es un lenguaje libre de uso, eso quiere decir que es gratuito. Antes estaba muy asociado a la plataforma Unix, pero en la actualidad está disponible en otros sistemas operativos como Windows.

PHP

Es el acrónimo de Hipertext Preprocesor. Es un lenguaje de programación del lado del servidor gratuito e independiente de plataforma, rápido, con una gran librería de funciones y mucha documentación.

Fue creado originalmente en 1994 por Rasmus Lerdorf, pero como PHP está desarrollado en política de código abierto, a lo largo de su historia ha tenido muchas contribuciones de otros desarrolladores. Actualmente PHP se encuentra en su versión 4, que utiliza el motor Zend, desarrollado con mayor meditación para cubrir las necesidades de las aplicaciones Web actuales.

Este lenguaje de programación está preparado para realizar muchos tipos de aplicaciones Web gracias a la extensa librería de funciones con la que está dotado. La librería de

funciones cubre desde cálculos matemáticos complejos hasta tratamiento de conexiones de red, por poner dos ejemplos.

Algunas de las más importantes capacidades de PHP son: compatibilidad con las bases de datos más comunes, como MySQL, Oracle, Informix, por ejemplo. Incluye funciones para el envío de correo electrónico, upload de archivos, crear dinámicamente en el servidor imágenes en formato GIF, incluso animadas y una lista interminable de utilidades adicionales.

JSP

Es una tecnología orientada a crear páginas Web con programación en Java. Se pueden crear aplicaciones Web que se ejecuten en variados servidores Web, de múltiples plataformas. Las páginas JSP están compuestas de código HTML/XML mezclado con etiquetas especiales para programar scripts de servidor en sintaxis Java.

ASP

Active Server Pages es la tecnología para la creación de páginas dinámicas del lado del servidor desarrollada por Microsoft.

El tipo de servidores que emplean este lenguaje son aquellos que funcionan con sistema operativo de la familia de Windows NT. Aunque es posible utilizarlo en una PC con Windows 98 si se instala un servidor denominado Personal Web Server. Incluso en sistemas Linux es posible utilizar esta tecnología si se instala un componente denominado Chilisoft, aunque lo mejor para trabajar esta tecnología es sobre el servidor para el cual está pensado, **Internet Information Server**. [7]

ASP.NET

ASP.NET es la tecnología que permite hacer formularios Web y Servicios Web. Una aplicación Web ASP.NET puede estar formada por varios formularios Web.

Es el sucesor de las páginas ASP, son páginas dinámicas que pueden acceder a recursos del servidor, permite la ejecución de páginas Web del lado del servidor, permite hacer

servicios Web, es independiente del explorador de Internet. Es independiente del lenguaje de programación.

ASP.NET combina una productividad sin precedentes de los desarrolladores con rendimiento, fiabilidad y facilidad de despliegue.

Aplicaciones Web Asp.NET

Las aplicaciones Web ASP.NET tienen varios componentes:

Formularios Web o páginas .ASPX: Proveen de la interfase visual. No tienen código ejecutable.

Páginas de código en archivo aparte: Están asociadas con cada formulario y son las que proveen del código ejecutable. A diferencia de las páginas ASP con la tecnología anterior, no se mezcla código y etiquetas en la misma página.

Archivos de configuración: Son archivos que permiten configurar la aplicación, por ejemplo el archivo web.config y el servidor, por ejemplo el archivo machine.config.

Global.asax: Es un archivo que contiene código. Este código responde a eventos que se disparan en la aplicación Web.

Enlaces a Servicios Web XML: Permite a la aplicación Web transferir datos XML desde y hacia Servicios Web.

Conectividad a Bases de Datos: Permite a la aplicación Web transferir datos desde y hacia distintas Bases de datos.

Caching: Permite a la aplicación Web devolver formularios Web más rápidamente después del primer acceso.

Fundamentación del Lenguaje Utilizado

CubaSI es un portal totalmente renovado y dinámico, desarrollado en .Net Framework

.NET Framework es una plataforma informática que simplifica el desarrollo de aplicaciones en un entorno altamente distribuido como es Internet. Ofrece un entorno de desarrollo integrado (IDE) único que ayuda a los programadores a crear sus soluciones de forma más rápida y con acceso a la productividad clave, independientemente del lenguaje utilizado, tiene dos elementos principales:

- **El Common Language Runtime:** Administra servicios en tiempo de ejecución como por ejemplo, la seguridad, la administración de la memoria y la integración de los lenguajes.
- **Las bibliotecas de clases:** Proveen de código reusable para las tareas más comunes como por ejemplo: acceso a datos, desarrollo de Servicios Web y formularios Web o de Windows.

Se decidió implementar la aplicación usando esta plataforma para que fuera más sencillo poder integrarla al portal, por tanto el trabajo se desarrolla utilizando Visual Studio.NET que admite el entorno de Microsoft .NET Framework, El IDE de Visual Studio .NET es un centro de control totalmente personalizable que proporciona el máximo rendimiento para los programadores, incluye diversas tecnologías suplementarias para simplificar el diseño, desarrollo e implementación en equipo de las soluciones, así como herramientas de desarrollo basadas en componentes.

Como lenguaje de programación se toma **C#**, ya que los administradores de CubaSI están familiarizados con el, además el portal en su totalidad está programado en C sharp.

A continuación se explican algunas características importantes del lenguaje que contribuyen a justificar la elección.

Es un lenguaje de programación sencillo y moderno, orientado a objetos y con seguridad de tipos. Combina la gran productividad de los lenguajes de desarrollo rápido (RAD, *Rapid Application Development*) con la eficacia de C++. Es un lenguaje de programación simple pero eficaz. Es una evolución de los lenguajes C y C++. Utiliza mucha de las características de C++ en las áreas de instrucciones, expresiones y operadores. Presenta

considerables mejoras e innovaciones en áreas como seguridad de tipos, control de versiones, eventos y recolección de elementos no utilizados.

Metodología utilizada

En su mayoría el desarrollo de software es riesgoso y difícil de controlar, pero si no se lleva una metodología de por medio, los clientes del software quedan insatisfechos y los desarrolladores aún mas.

Rational Unified Modeling (RUP) es la metodología utilizada para modelar el contexto del sistema. Es un proceso que define claramente *quien, cuando y que debe hacerse*, y como su enfoque está basado en modelos utiliza un lenguaje bien definido para tal fin, el **UML**.

Las características más importantes de este proceso son:

- Guiado por Casos de Uso
- Centrado en la Arquitectura
- Iterativo en Incremental

A través de un proyecto guiado por RUP, los requerimientos funcionales son expresados en la forma de Casos de Uso, que guían la realización de una arquitectura ejecutable de la aplicación. Además el proceso focaliza el esfuerzo del equipo en construir los elementos críticos estructuralmente y del comportamiento (llamados Elementos Arquitecturales) antes de construir elementos menos importantes. Finalmente RUP particiona el ciclo de vida en iteraciones que producen versiones incrementales del sistema.

La metodología RUP divide en 4 fases el desarrollo de software. Cada fase tiene definido un conjunto de objetivos.

Fase	Objetivos
Inicio	Definir el alcance del proyecto. Entender que se va a construir.
Elaboración	Construir una versión de la arquitectura de la aplicación. Entender como se va a construir.
Construcción	Completar el esqueleto de la aplicación con funcionalidad.

	Construir una versión Beta.
Transición	Disponibilizar la aplicación para los usuarios finales.
	Construir una versión final.

Cada una de estas etapas es desarrollada mediante el ciclo de iteraciones, la cual consiste en reproducir el ciclo de vida en cascada a menor escala. Los Objetivos de una iteración se establecen en función de la evaluación de las iteraciones precedentes.

En cada ciclo de vida que se desarrolla en cada iteración es llevado por las siguientes disciplinas:

1. Modelamiento del negocio: entendimiento de las necesidades del negocio.
2. Requerimientos: Traslado de las necesidades del negocio a un sistema automatizado.
3. Análisis y Diseño: Traslado de los requerimientos dentro de la arquitectura de software.
4. Implementación: Crear el software que se ajuste a la arquitectura y tenga el comportamiento deseado.
5. Pruebas: Asegurar que el comportamiento obtenido es el correcto y que todo lo solicitado está contemplado.
6. Configuración y administración de cambios: Guardar todas las versiones del proyecto.
7. Administración del proyecto: Administrando horarios y recursos
8. Ambiente: Administrando el ambiente de desarrollo.

Unified Modeling Language (UML)

El UML (Lenguaje Unificado para la construcción de Modelos) se define como un lenguaje que permite especificar, visualizar y construir los artefactos de los sistemas de software. Es un sistema rotacional (que, entre otras cosas, incluye el significado de sus notaciones) destinado a los sistemas de modelado que utilizan conceptos orientados a objetos.

El UML es un estándar para construir modelos orientados a objetos surgió en respuesta a una petición de OMG (*Object Management Group*, una asociación para fijar los estándares de la industria) para definir un lenguaje de construcción de modelos.

“El UML es un lenguaje para la construcción de modelos; no guía al desarrollador en la forma de realizar el análisis y diseño orientados a objetos ni le indica cual proceso de desarrollo adoptar”. [2]

Herramientas Utilizadas

GoldParser. Esta herramienta es usada para generar dada la gramática del lenguaje para la consulta la tabla LARL correspondiente, en conjunción con las librerías open source goldnet, consistentes en el motor de ejecución encargado de interpretar las tablas producidas por el GoldParser.

Macromedia Dreamweaver MX 2004. Se utilizó en el diseño Web, dado las potencialidades de la herramienta y la comodidad de uso. Dreamweaver MX 2004 ofrece una combinación muy útil de herramientas de diseño visual funciones de desarrollo de soporte para la edición de código.

Herramientas CASE (Computer Aided Software Engineering). Son herramientas que ayudan al analista, o a todo el equipo de trabajo para administrar el proceso durante una o más fases del desarrollo de software y facilitar el mantenimiento o la creación de nuevas versiones en otros momentos.

- Embarcadero Erwin Studio: Herramienta muy útil para diseñar y crear Bases de Datos.
- Rational Rose Enterprise Edition: Herramienta de Modelación por excelencia muy usada por las facilidades que usa en la utilización del UML.

Visual Studio .NET Siendo la herramienta esencial para desarrollar cualquier aplicación .NET. Utiliza un único Entorno de Desarrollo Integrado (IDE) para múltiples lenguajes, que les permite compartir herramientas y facilita la creación de soluciones en varios lenguajes permitiendo que estos aprovechen las funciones de .NET Framework, que ofrece acceso a tecnologías claves para el desarrollo de aplicaciones Web y Servicios Web XML. Visual Studio .NET ofrece una funcionalidad superior para optimizar procesos empresariales, abriendo así las puertas a nuevas oportunidades para las empresas,

también mejora notablemente la productividad del programador, permitiendo generar soluciones para una amplísima gama de clientes, incluidas aplicaciones para la Web, Windows y dispositivos de cliente ligero.

Gestor de BD utilizado

Una Base de Datos (BD) es una colección de datos relacionados entre si almacenadas para su uso posterior, perfectamente se puede considerar una biblioteca como una Base de Datos. En la actualidad la mayoría de las BD tienen formato electrónico, en lo adelante cuando se haga alusión a una BD se refiere a un formato digital de la misma.

Las Bases de Datos y sus tecnologías tienen un impacto decisivo en el creciente uso de las computadoras. Desempeñan un papel crucial en casi todas las áreas de la informática y sistemas automatizados.

Las principales ventajas de las BD son debido a los Sistemas Gestores de Bases de Datos (SGBD) que permiten y posteriormente acceder a los datos de forma rápida y estructurada. Un SGBD es el conjunto de programas que se encargan de manejar la creación y todos los accesos a la BD.

Entre los principales SGBD existentes se encuentran: *Oracle, SQLServer, MySQL.*

Se ha decidido usar **Microsoft SQL Server**, puesto que es el (SGBD) utilizado en el portal, a continuación se explican algunas de sus características importantes que contribuyen a justificar su elección.

Es una aplicación poderosa, robusta, que permite gran seguridad de los datos, ostenta marcas de referencia en cuanto a escalabilidad y confiabilidad, que son críticas para el éxito de bases de datos de gran tamaño. El SQL Server permite lograr una gran velocidad en el procesamiento de transacciones, y agilidad en todas sus operaciones.

Conclusiones.

En este capítulo se profundizó sobre los principales aspectos necesarios a reconocer en los sistemas de recuperación de información, se justificó la tecnología y el lenguaje utilizado, con aspectos técnicos de los mismos.

Capítulo

3

DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

Introducción

En el presente capítulo se realiza la descripción de la solución propuesta, se utiliza un modelo de dominio puesto que el sistema no da soporte a negocio alguno, ya que CubaSI, no cuenta con una herramienta tipo directorio en su sistema de búsqueda. Este modelo de dominio contiene la información necesaria: las principales entidades del dominio y sus relaciones.

En otro acápite se describen los requisitos funcionales y no funcionales, pasos importantes para identificar y documentar lo que se necesita realmente, y por último se explican los principales casos de uso, técnica que permite mejorar la comprensión de los requerimientos. Los casos de uso son casos de utilización del sistema, ejemplifican e incluyen tácticamente los requerimientos en sus narraciones. Se utiliza un diagrama de Casos de Uso para entenderlos gráficamente, y explicar la relación que existe entre estos y los actores (agentes externos). Su principal objetivo es conocer los actores de un sistema y las formas básicas en que lo utilizan.

Modelo de Dominio

Un modelo del dominio describe los conceptos importantes del contexto como objetos del dominio, y enlaza estos objetos unos con otros. La identificación y la asignación de un nombre para estos objetos ayudan a desarrollar un glosario de términos que permitirán comunicarse mejor a todos los que están trabajando en el sistema. Mas adelante, los objetos del dominio ayudan a identificar algunas de las clases que se analizan y diseñan en el sistema. [2]

Se decidió utilizar solamente un modelo de dominio, puesto que la aplicación no da soporte a ningún negocio en CubaSI.

El sistema utilizará la Web para mostrar al usuario (navegante) la estructura del directorio desglosado por categorías, el usuario podrá realizar búsquedas y el sistema devolverá los resultados acorde a la búsqueda realizada, permite además recoger las solicitudes de inclusión de un sitio que se hayan confeccionado, evaluarlas e incluirlas siempre y cuando cumpla con determinadas condiciones.

Una **solicitud** es un pedido que el cliente hace de inclusión de un sitio Web en el directorio.

Se considera **índice de documento** a un elemento que contenga los principales datos de un sitio Web, título, url, resumen, palabras claves.

Se denomina **servidor de documento** a aquel elemento que contenga almacenado una lista de índices de elementos

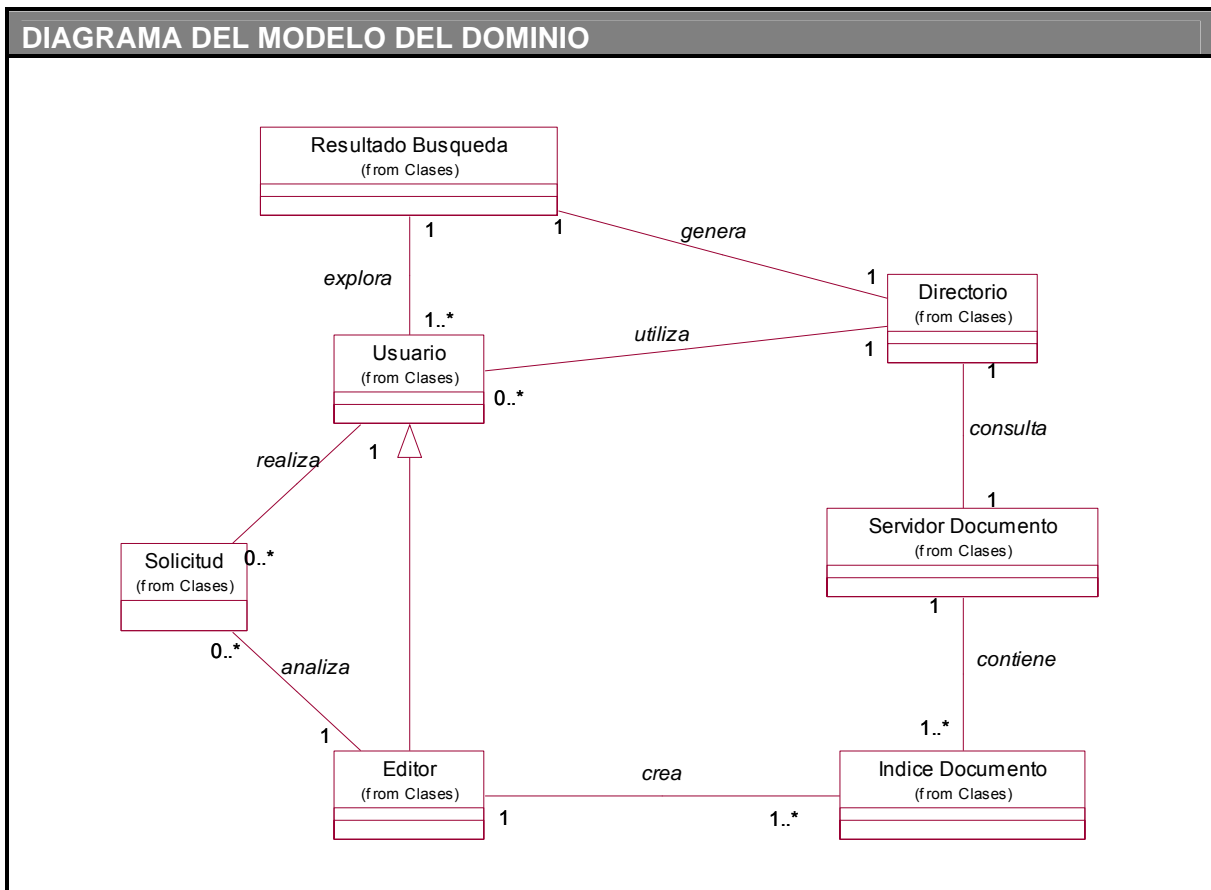
Se reconoce como **usuario** a aquella persona que se encuentre utilizando el buscador.

Un **editor** es aquel usuario que tiene determinados permisos para analizar una solicitud, construir o crear un índice de documento y almacenarlo en el servidor de documento.

Se entiende por **directorio** a la aplicación de búsqueda, con la que el usuario interactúa directamente.

Y por ultimo se entiende por **resultado de búsqueda** a una lista de índices de documentos generada por el directorio en correspondencia a la búsqueda que realiza el usuario.

El modelo de dominio se describe mediante un diagrama de clases conceptuales en UML y su objetivo es comprender y describir algunas de las clases más importantes dentro del contexto del sistema.



Requisitos Funcionales

1. Realizar búsquedas.

1.1. Permitir realizar 2 tipos de búsqueda: Corriente o por Navegación en el Directorio

1.1.1. Realizar Búsqueda Corriente (por palabras)

- El usuario introduce las palabras por las que quiere buscar y se mostraran los índices de los sitios contenidos en el directorio que se relacionen con los términos de la búsqueda.
- Los resultados que se mostraran estarán contenidos en la categoría en la que el usuario esté en ese momento ubicado.

1.1.2. Realizar Búsquedas por Navegación en el Directorio

- El usuario realiza la búsqueda seleccionando algunas de las categoría que el sistema muestre.
- El sistema mostrará el rastro de exploración de las categorías de manera que permita dirigirse directamente a una categoría que ya haya sido seleccionada por el usuario en ese instante de búsqueda.

1.2. Mostrar Título, Url, Resumen de los sitios que aparezcan como resultado, sea de una búsqueda corriente o por navegación.

2. Realizar solicitud de inclusión de un sitio.

2.1. El sistema mostrará la opción de sugerir sitio en la interfaz de búsqueda del directorio.

2.2. EL sistema debe permitir que el usuario introduzca los datos del sitio que desea adicionar al directorio, Url, Resumen. Debe validar que la url suministrada sea válida.

2.3. Validar que el usuario introduzca una dirección de correo valida, único campo obligatorio que se solicita al usuario al realizar una solicitud, brindarle la posibilidad de agregar el nombre.

3. Administrar el sistema.

3.1. Administrar Editores del Sistema

- 3.1.1. Adicionar editores, especificando los datos y el rol que desempeña.
- 3.1.2. Actualizar datos de un editor, incluido cambiar contraseña.
- 3.1.3. Eliminar editores del sistema.

3.2. Administrar las Categorías del Sistema

- 3.2.1. Adicionar una Categoría al directorio, incluyendo una raíz.
- 3.2.2. Permitir eliminar una categoría del directorio.
- 3.2.3. Eliminar un sitio de una categoría determinada.

4. Autenticar Usuario. Este debe introducir su nombre y contraseña.

- 4.1. Una vez logeado en el sistema podrá realizar solo las opciones que se corresponden con el rol que juega en el mismo.
- 4.2. El usuario logeado podrá acceder a sus datos personales y realizar alguna actualización o modificación sobre los mismos.

5. Editar Categorías. Permitir que se editen y se de mantenimiento a las categorías del directorio.

- 5.1. Cada categoría debe tener asociada una persona responsable de darle mantenimiento a la misma.

6. Gestionar Solicitudes, Cada solicitud es revisada por el editor encargado de la categoría a la que se desea añadir el sitio.

- 6.1. En dependencia de su contenido, la solicitud puede ser aceptada o denegada.
- 6.2. Si la solicitud es aceptada se adicionará el sitio a la categoría correspondiente.
- 6.3. Enviar un correo electrónico al usuario que realizó la solicitud notificándole la decisión tomada sobre la misma.
- 6.4. Actualizar lista de solicitudes pendientes.

Requisitos no Funcionales

Los Requerimientos no funcionales son propiedades o cualidades que el producto debe cumplir. Debe pensarse en estas propiedades como las características que hacen al producto atractivo, usable, rápido o confiable.

Los requerimientos no funcionales forman una parte significativa de la especificación. Son importantes para que clientes y usuarios puedan valorar las características no funcionales del producto, pues si se conoce que el mismo cumple con la toda la funcionalidad requerida, las propiedades no funcionales, como cuán usable, seguro, conveniente y agradable, pueden marcar la diferencia entre un producto bien aceptado y uno con poca aceptación.

Los requerimientos no funcionales incluyen:

- Conjunto de Facilidades.
- Capacidades.

Apariencia o Interfaz externa:

- Diseño sencillo, permitiendo que no sea difícil maniobrar con el sistema.
- Paginación en los listados de reportes, y búsquedas.

Usabilidad

- El sistema podrá ser usado por cualquier persona que posea conocimientos básicos en un ambiente Web.
- Deberá visualizarse bien en los principales navegadores en cuba.

Rendimiento

- El sistema operará con grandes volúmenes de información además con gran cantidad de usuarios conectados al mismo tiempo, realizando pedidos.

Soporte

- Se requiere de un potente servidor de Base de Datos capaz de almacenar grandes volúmenes de información, con velocidad de procesamiento eficiente y tiempo de respuesta rápido.
- Plataforma .NET, versión 2.0

- Se requiere un navegador por parte del cliente capaz de interpretar JavaScript.

Seguridad

- Identificar al usuario antes de efectuar cualquier operación en el directorio que afecte su organización, no todas las operaciones requieren de este proceso, para realizar una búsqueda no es necesario haber identificado al usuario antes.
- Garantizar que la información sea vista y actualizada solamente por quien tiene derecho a verla.
- Verificación sobre acciones que no tienen marcha atrás, ejemplo, eliminación.

Funcionalidad

- Capacidad de Búsqueda, muestreo de los resultados en el menor tiempo posible.
- Mínima cantidad de páginas posibles, agrupar en las páginas la mayor cantidad de funciones fuertemente relacionadas.

Software

- Navegador Internet Explorer 5.0 o superior.
- Microsoft SQL Server 2000.
- Internet Information Server 5.0 o posterior.
- Plataforma .NET 2.0.

Descripción del Sistema

Para dar cumplimiento a los objetivos del sistema y a los requerimientos antes mencionados, se divide el sistema en dos módulos principales, Directorio y Buscador. En el primero se gestiona todo lo referente al funcionamiento del sistema relacionado con la funcionalidad de directorio. O sea, se le da mantenimiento y seguimiento al directorio, se realizan las modificaciones a las categorías, se analizan las solicitudes pendientes de inclusión de sitios, además se gestiona todo lo referente a la administración, tanto de los editores, que son las personas encargadas de darle funcionamiento y mantenimiento al directorio y de las categorías que son las que componen y organizan el directorio. El objetivo de desarrollar una estructura de categorías es crear un sistema que permita al usuario encontrar fácilmente lo que busca.

Los editores seleccionan, evalúan, describen y organizan los sitios Web, son los responsables de revisar las sugerencias enviadas a sus categorías y listan los sitios según las líneas editoriales preescritas. El sistema permitirá avisarles a los solicitantes cuando la solicitud sea aprobada o denegada. Una persona que esté listada como editor de una categoría, puede editar también en cualquiera de las categorías que estén por debajo como “editor superior”, e incluso renombrar algunas de las categorías en las que el juegue este rol, no necesita nombrarse a si mismo como editor de las subcategorías que cree. Pueden existir categorías que no contengan editores, en tal caso su editor superior asumirá la responsabilidad de mantenerla. Solo los editores listados como “editores avanzados” tienen permisos para mover, eliminar o renombrar alguna de las categorías del directorio.

En el módulo de Buscador se gestiona todo lo referente al motor de búsqueda, se encarga de devolver los resultados de los sitios relacionados con las búsquedas formuladas por los usuarios. Cuenta con un analizador de consultas encargado de realizar un análisis lexicográfico y sintáctico de la consulta elaborada por el usuario, y de la generación de la forma interna correspondiente.

El lenguaje de la consulta está formado por operadores booleanos, a continuación se explica como funcionan estos operadores en la consulta.

Operadores booleanos

Los operadores incluidos en el lenguaje son: *OR* (para indicar necesidad de documentos en los que aparezcan algunos de los términos) *AND* (interés en los documentos en los que estén todos los términos) y *NOT* (para indicar la exclusión de los términos). En el caso del operador *AND* es bueno destacar el carácter no restrictivo que presenta: dada una expresión de la forma *pal1 AND pal2 AND pal3*, no solo son relevantes los documentos que contienen estas tres palabras sino aquellas que contengan un subconjunto de las mismas.

Gramática del Lenguaje de Consulta

El lenguaje de consulta utilizado es libre de contexto o de tipo II y la gramática que lo genera queda definida de la siguiente manera:

$G = \langle \{S, T, K, F\}, \Sigma, P, I \rangle$ Donde:

El alfabeto es: $\Sigma = \{\varepsilon, or, and, not, \text{término}, (,.)\}$

En el que ε constituye la cadena vacía y los tokens quedan especificados como sigue:

Or $\leftarrow [Oo][Rr]$

And $\leftarrow [Aa][Nn][Dd]$

Not $\leftarrow [Nn][Oo][Tt]$

término $\leftarrow [A\dots Za\dots z0\dots 9ÁÉÍÓÚáéíóú, \%]$

Y las producciones son las siguientes:

$$P = \{$$

$$S \rightarrow T \mid T \text{ or } S \mid T \text{ and } S$$

$$T \rightarrow F \mid \text{not } F \mid$$

$$F \rightarrow \text{término } F \mid \text{término} \mid (S)$$

$$\}$$

Una vez creada la forma interna de la consulta se extraen de la BD todos los posibles documentos que han de satisfacerla, estos son los que contengan al menos uno de los términos de la consulta, y se obtienen ordenados por la relevancia que tiene cada documento.

Actores del Sistema

ACTORES	JUSTIFICACIÓN
Administrador	Representa una persona que puede realizar operaciones de configuración sobre el directorio, como configurar categorías, sitios, usuarios, etc.
Usuario	Representa a la persona que ha accedido al directorio desde la Web para realizar una búsqueda.
Editor	Es aquella persona responsable directa de una categoría, es la encargada de Analizar las solicitudes hechas hacia dicha categoría Puede también editar en cualquiera de las subcategorías o sub-subcategorías bajo ella como editor superior.
Editor Avanzado	La persona con este rol, es responsable de una categoría raíz del directorio, puede realizar múltiples operaciones en ella como, mover categorías, adicionar editores, categorías, sitios, etc.

Modelo de Casos de Uso del Sistema

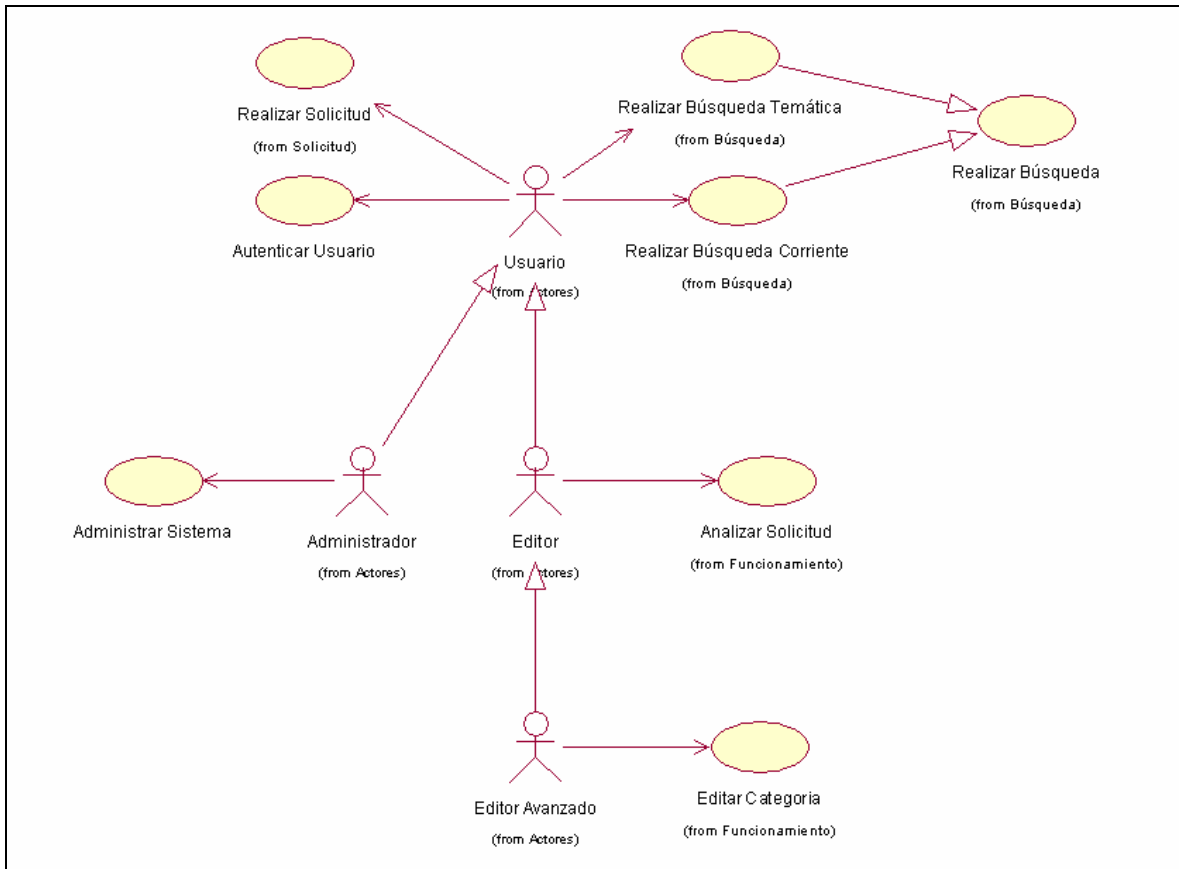


Figura 5 Modelo de CU del Sistema

Expansión de los Casos de Uso

Caso de Uso	Realizar Búsqueda	
Actores	Usuario	
Propósito	Realizar una búsqueda en la Web y obtener resultados de sitios que contengan información referente al tema de búsqueda.	
Resumen	El caso de uso se inicia cuando el Caso de Uso “Realizar Búsqueda Corriente” o “Realizar Búsqueda Temática”, invoca al comportamiento del caso de uso en cuestión, que generaliza acciones comunes de ambos casos.	
Referencias	R1	
Precondiciones		
Curso Normal de Eventos		
Sección: “Iniciar Búsqueda”		
Acción del Actor	Respuesta del Sistema	
1. El usuario accede a la interfaz principal del directorio con el objetivo de efectuar una búsqueda.	1.1 Muestra la estructura del directorio desde las categorías raíces así como la posibilidad de que el usuario introduzca las palabras por la que desee efectuar la búsqueda. <i>(Ver Pantalla 1)</i>	
Sección: Realizar Búsqueda		
Se invoca a la sección “Realizar búsqueda” del caso de uso búsqueda temática o del caso de uso búsqueda corriente, en dependencia del tipo de búsqueda.		
Sección: Obtener Resultados		
3. Selecciona en (A) alguno de los sitios encontrados como resultado de la búsqueda. <i>(Ver Pantalla 2)</i>	3.1. Muestra el sitio seleccionado en una nueva ventana del navegador.	
Poscondiciones		
Requisitos Especiales		

Prototipo

Directorio CubaSearch®

Buscar en la categoría actual Buscar en el directorio

Categorías

ARTES	EDUCACION	PAISES	CIENCIA Y TECNOLOGÍA
REFERENCIA	COMPRAS	JUEGOS	SALUD
MEDIOS DE COMUNICACION	SOCIEDAD	DEPORTES	NEGOCIOS

Pantalla 1

Sitios Web relacionados A

[Universidades de Cuba](#)
Sitio que aborda el tema de la educacion universitaria en cuba
<http://universidades.edu.cu>

[Educacion General Integral](#)
Todo lo referente a la importancia de una buena educacion integral en el joven que se forja al calor de la batalla de ideas que libra nuestro pais actualmente.
<http://educaciongeneral.edu.cu>

Pantalla 2

Caso de Uso	Búsqueda temática (Especialización)	
Actores	Usuario	
Resumen	El caso de uso inicia cuando el usuario accede a la página principal del directorio y se dispone a realizar una búsqueda seleccionando las categorías que aparece y finaliza cuando selecciona uno de los sitios encontrados en la búsqueda y lleva al usuario a la página del sitio.	
Referencias	R1	
Precondiciones	.	
Curso Normal de Eventos		
Sección: “Iniciar Búsqueda”		
Acción del Actor	Respuesta del Sistema	
Se invoca a la sección: “Iniciar Búsqueda” del Caso de Uso Generalizado: “Realizar Búsqueda”		
Sección: “Realizar Búsqueda”		
2. El usuario selecciona una categoría en el directorio. (A) (Ver Pantalla 1)	1.1. Muestra el rastro de exploración de las categorías. (B) 1.2. La categoría seleccionada no es hoja, muestra los sitios relacionados con la categoría.	
Cursos Alternos		
La categoría seleccionada no es hoja, muestra la lista de subcategorías, además de los sitios relacionados con la misma. <i>Si selecciona una categoría se invoca a la actividad 2</i>		
Sección: “Obtener Resultados”		
Se invoca a la sección: “Obtener Resultados” del Caso de Uso Generalizado: “Realizar Búsqueda”.		
Poscondiciones		
Requisitos Especiales		
Prototipo		

The screenshot shows the CubaSearch directory interface. At the top, there is a search bar with a 'Buscar' button and a link to 'Sugerir un sitio'. Below the search bar, there are two radio buttons: 'Buscar en la categoría actual' (selected) and 'Buscar en el directorio'. A callout 'B' points to the search bar area. The main content area is titled 'EDUCACION' and includes a '<< atrás' link. A red bar highlights the 'Categorías' section, with a callout 'A' pointing to it. Below this, there is a grid of category links: UNIVERIDADES, A DISTANCIA, EL SALVADOR, NICARAGUA, BIBLIOTECAS, RECURSOS, SOFTWARE, LIBROS, IDIOMAS, CIENCIA Y TECNOLOGÍA, and COMPUTADORAS. A second red bar highlights the 'Sitios Web relacionados' section, which contains two entries: 'Universidades de Cuba' with a description and a URL, and 'Educación General Integral' with a description. A 'Local intranet' logo is visible in the bottom right corner of the page content.

Pantalla 1

Caso de Uso	Búsqueda Corriente (Especializado)	
Actores	Usuario	
Propósito	Realizar una búsqueda en el directorio, a través de una frase entrada por el usuario.	
Resumen	El caso de uso inicia cuando un usuario se dispone a realizar una búsqueda en el directorio, introduce los términos de la búsqueda y obtiene como resultado los sitios relacionados con dichos términos contenidos en el directorio.	
Referencias	R1	
Precondiciones	El usuario se encuentra navegando en el directorio.	
Curso Normal de Eventos		
Sección: Iniciar Búsqueda		
Acción del Actor	Respuesta del Sistema	
Se invoca la sección: "Iniciar Búsqueda" del Caso de Uso Generalizado: "Realizar Búsqueda".		
Sección: "Realizar Búsqueda"		
2. El usuario introduce en (A) los términos de la búsqueda. (Ver Pantalla 1)	1.1. Realiza la búsqueda dentro de la categoría en la que el usuario esté ubicado en ese momento. 2.2. Encuentra los sitios relacionados con los términos de la búsqueda (B)	
Cursos Alternos		
2.2. No encuentra sitios relacionados con los términos de la búsqueda, muestra un mensaje al usuario notificándole que su búsqueda no ha producido resultados. (Ver Pantalla 2)		
Sección: "Obtener Resultados"		
Se invoca a la Sección: "Obtener Resultados" del Caso de Uso Generalizado: "Realizar Búsqueda".		
Poscondiciones		
Requisitos especiales		
Prototipo		

The screenshot shows the CubaSearch directory interface. At the top left, a callout bubble labeled 'A' points to the search input field. The search term 'universidad' is entered in the field. To the right of the field is a 'Buscar' button and a link 'Sugerir un sitio'. Below the search field are two radio buttons: 'Buscar en la categoría actual' (selected) and 'Buscar en el directorio'. The main content area is titled 'EDUCACION' and includes a '<< atrás' link. A red header bar contains the word 'Categorías'. Below this, there are several blue underlined links: 'UNIVERSIDADES', 'A DISTANCIA', 'EL SALVADOR', 'NICARAGUA', 'BIBLIOTECAS', 'RECURSOS', 'SOFTWARE', 'LIBROS', 'IDIOMAS', 'CIENCIA Y TECNOLOGÍA', and 'COMPUTADORAS'. A second callout bubble labeled 'B' points to the 'SitiOS Web relacionados' section. This section lists two related websites: 'Universidades de Cuba' with a description and URL 'http://universidades.edu.cu', and 'Educación General Integral' with a description and URL 'http://educaciongeneral.edu.cu'.

Pantalla 1

The screenshot shows the CubaSearch directory interface with a search for 'estupefacto'. The search input field contains 'estupefacto'. To the right of the field is a 'Buscar' button and a link 'Sugerir un sitio'. Below the search field are two radio buttons: 'Buscar en la categoría actual' (selected) and 'Buscar en el directorio'. The main content area displays the message 'SU BÚSQUEDA NO HA PRODUCIDO NINGUN RESULTADO' in bold, followed by a '<< atrás' link.

Pantalla 2

Caso de Uso	Realizar Solicitud
Actores	Usuario.
Propósito	Solicitar se añada un sitio para que aparezca en los resultados de búsqueda que se realicen en el directorio.
Resumen	El caso de uso inicia cuando el usuario selecciona la opción de adicionar sitio en el directorio, formula los datos necesarios para llenar la solicitud, al finalizar se añade la solicitud a la lista de solicitudes pendientes
Referencias	R2
Precondiciones	El usuario se encuentra navegando en el directorio, y no se encuentra ubicado en la página principal de directorio, solo podrá agregar solicitudes en cualquiera de las subcategorías que se deriven de las categorías raíces.

Curso Normal de Eventos

Acción del Actor	Respuesta del Sistema
1. Selecciona la opción sugerir sitio (A). (Ver pantalla 1)	1.1. El sistema lleva al usuario a la página que contiene el formulario necesario para llenar la solicitud.
2. Llena todos los datos requeridos y presiona el botón adicionar (A). (Ver Pantalla 2)	2.1. Comprueba que los datos estén correctos. 2.2. Crea un registro de solicitud pendiente y la adiciona a la lista de solicitudes. 2.3. Redirecciona al usuario a la página principal del directorio.

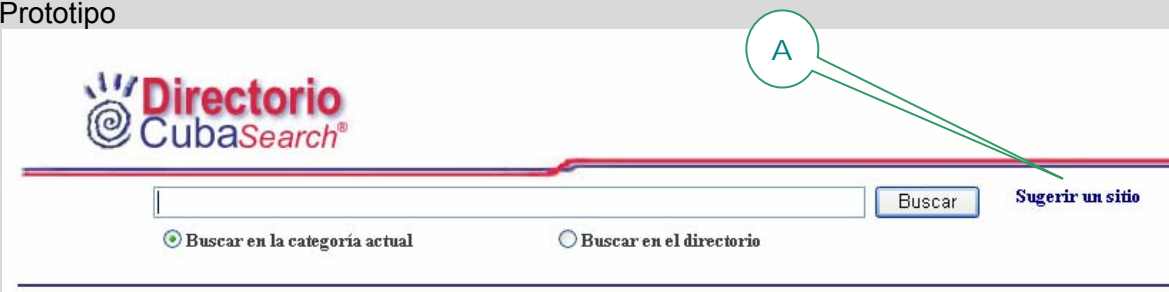
Curso Alterno

2.1. Los datos no son correctos, muestra un mensaje para que el usuario corrija la falla

Poscondiciones

Requisitos especiales

Prototipo



Pantalla 1

Directorio CubaSearch®

EDUCACION

La solicitud que usted realice, será evaluada por una editor que analizará el contenido del sitio en cuestion y despues de evaluada su solicitud se le notificará via email siempre y cuando usted introduzca esta informacion

Si desea ver informacion de la categoría marque aqui

Url del sitio que desea adicionar

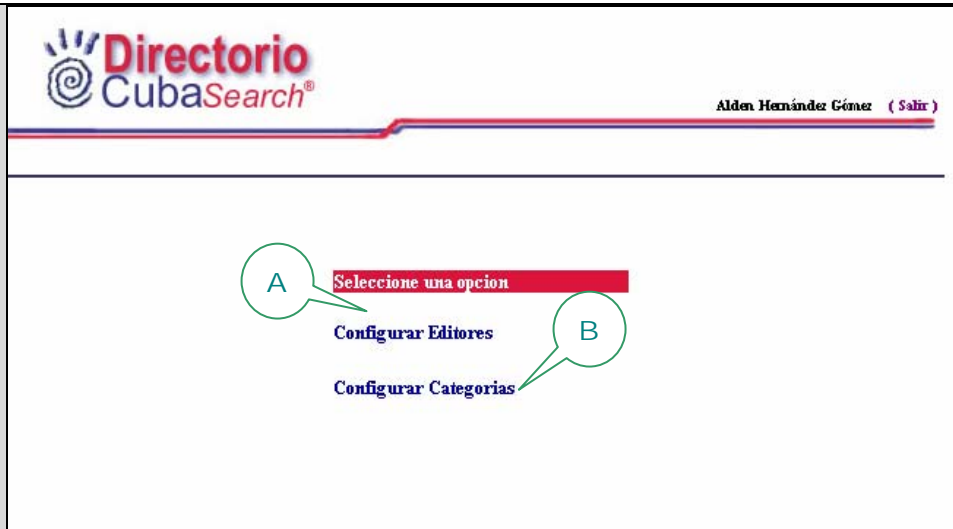
Resumen que quiere se muestre de su sitio, si lo desea puede estar en blanco

Favor de introducir su nombre

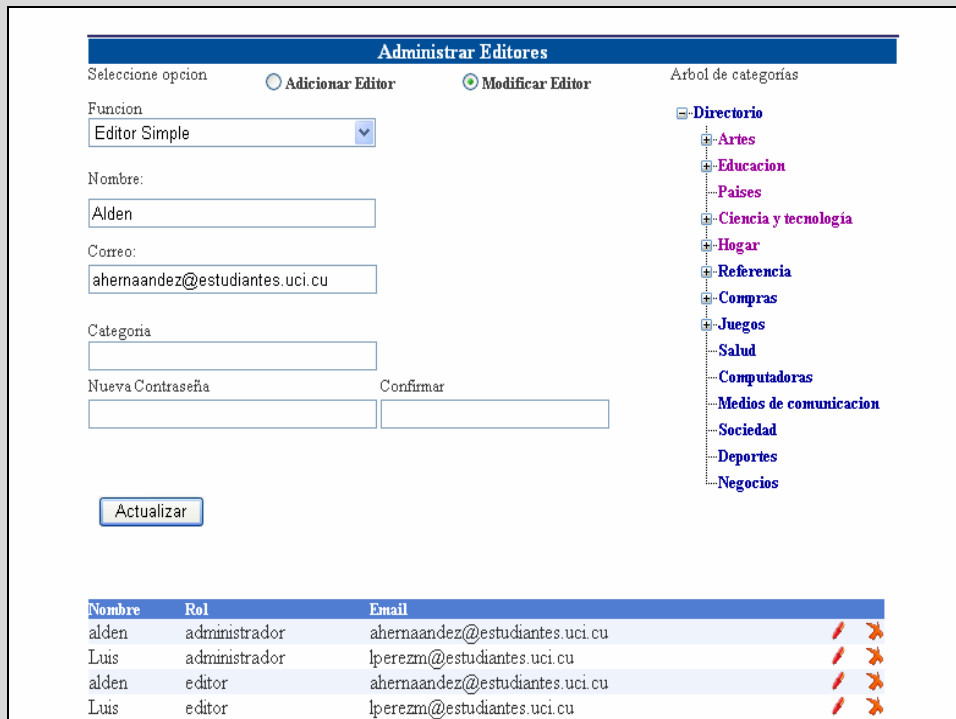
Introduzca su correo para que se le notifique de la decision tomada

Pantalla 2

Caso de Uso	Administrar Sistema	
Actores	Administrador	
Propósito	El administrador puede configurar el sistema y realizar una serie de operaciones como son, listar editores, categorías, permitiendo hacer variaciones en las mismas.	
Resumen	El caso de uso inicia cuando el administrador accede a la página administrativa. Aquí puede realizar varias operaciones, Listar Editores, Listar Categorías, etc.	
Referencias	R3	
Precondiciones	Debe estar identificado como administrador del sistema	
Curso Normal de Eventos		
	Acción del Actor	Respuesta del Sistema
	1. El administrador accede a la página administrativa.	1.1. El sistema muestra en pantalla las opciones que puede realizar el administrador <i>(Ver pantalla 1)</i>
	2. Selecciona la opción: a) Configurar Editores. Ver sección <i>Configurar Editores (A)</i> b) Configurar Categorías del Sistema. Ver sección <i>Configurar Categorías (B)</i>	
Sección “Configurar Editores”		
	1. Puede realizar cualquier operación sobre los editores del sistema, con las opciones de editar y eliminar, además los campos vacíos de nombre, correo, rol que desempeñará en el sistema, además de la categoría que será especialista, esto es para el caso que desee agregar un nuevo editor al directorio. <i>(Ver Pantalla 2)</i>	
Sección “Configurar Categorías”		
	1. Puede realizar cualquier operación sobre las categorías existentes, incluido adicionar una categoría raíz al directorio. O eliminar algún sitio del sistema. <i>(Ver pantalla 3)</i>	
Poscondiciones		
Requisitos Especiales		
Prototipo		



Pantalla 1



Pantalla 2

Configurar Categorías

Seleccione opción

Adicionar Categoría
 Modificar Categoría
 Mover Categoría

Nombre de la categoría

Introduzca el resumen

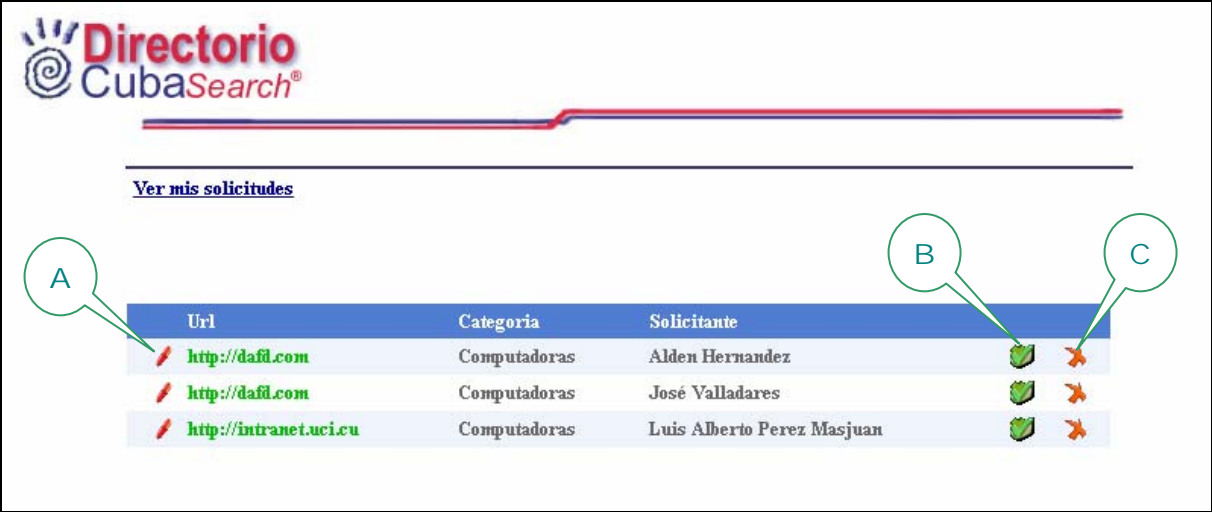
Categoría a la que pertenece, seleccionar categoría del árbol

Directorio

- Artes
- Educación
- Países
- Ciencia y tecnología
- Hogar
- Referencia
- Compras
- Juegos
- Salud
- Computadoras
- Medios de comunicación
- Sociedad
- Deportes
- Negocios

Sitios relacionados			Mod	Del
Titulo:	departamento de matematicas.	Url: http://matematica.uci.cu/docum/clases_ioperaciones.htm	✍	✖
Titulo:	portal de software libre uci-bienvenido!	Url: http://softwarelibre.uci.cu/userinfo.php?uid=1102	✍	✖
Titulo:	portal de software libre	Url: http://softwarelibre.uci.cu/modules/news/article.php	✍	✖

Pantalla 3

Caso de Uso	Analizar Solicitud																					
Actores	Editor																					
Propósito	Analizar una solicitud de inclusión de un sitio al directorio y darle tratamiento a la misma, puede ser aceptada o denegada.																					
Resumen	El caso de uso inicia un editor accede a la página de Edición de Categoría, y selecciona una solicitud de la lista de solicitudes pendientes en su categoría, finaliza cuando la solicitud queda aceptada o denegada.																					
Precondiciones	Debe estar identificado como editor del sistema																					
Referencias	R5																					
Curso Normal de Eventos																						
Acción del Actor	Respuesta del Sistema																					
1. Selecciona la opción de analizar. (A)	1.1. El sistema lleva al editor directamente a la página cuyo url es el suministrado.																					
2. Analiza el estado y contenido del sitio en cuestión y selecciona la opción aprobar.(B)	2.2. Se produce la indexación del mismo, se elimina la solicitud de la lista de solicitudes pendientes.																					
Curso Alterno																						
3. No está de acuerdo con el contenido del sitio, selecciona la opción denegar, se elimina la solicitud de la lista y no se procede a la indexación del documento. (C)																						
Poscondiciones																						
Requisitos Especiales																						
Prototipo																						
 <p>The screenshot shows the 'Directorio CubaSearch' interface. At the top, there is a logo and a navigation bar. Below the navigation bar, there is a link 'Ver mis solicitudes'. The main content is a table with three columns: 'Url', 'Categoría', and 'Solicitante'. The table contains three rows of data. Callout 'A' points to the first row's 'Url' column. Callout 'B' points to the 'Aprobar' (green checkmark icon) column. Callout 'C' points to the 'Denegar' (red X icon) column.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Url</th> <th>Categoría</th> <th>Solicitante</th> <th>Aprobar</th> <th>Denegar</th> </tr> </thead> <tbody> <tr> <td>http://dafid.com</td> <td>Computadoras</td> <td>Alden Hernandez</td> <td></td> <td></td> </tr> <tr> <td>http://dafid.com</td> <td>Computadoras</td> <td>José Valladares</td> <td></td> <td></td> </tr> <tr> <td>http://intranet.uci.cu</td> <td>Computadoras</td> <td>Luis Alberto Perez Masjuan</td> <td></td> <td></td> </tr> </tbody> </table>			Url	Categoría	Solicitante	Aprobar	Denegar	http://dafid.com	Computadoras	Alden Hernandez			http://dafid.com	Computadoras	José Valladares			http://intranet.uci.cu	Computadoras	Luis Alberto Perez Masjuan		
Url	Categoría	Solicitante	Aprobar	Denegar																		
http://dafid.com	Computadoras	Alden Hernandez																				
http://dafid.com	Computadoras	José Valladares																				
http://intranet.uci.cu	Computadoras	Luis Alberto Perez Masjuan																				
Pantalla 1																						

Caso de Uso	Editar Categoría	
Actores	Editor Avanzado	
Propósito	Dar mantenimiento a las categorías del directorio	
Resumen	Cada editor avanzado es responsable de una categoría raíz del directorio, en la que puede realizar diversas operaciones sobre la misma. Listar categorías, Adicionar Editor, etc.	
Precondiciones		
Referencias		
Curso Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
1. Accede a la página de edición	1.1 Muestra las opciones posibles a realizar. Adicionar (A), Modificar (B) y Mover Categoría (C), así como listado de las solicitudes (D), para manipularlas, y listado de los editores (E), puede adicionar editores también.	
2. De los sitios mostrados de cada categoría puede modificar (F) o eliminar algún sitio de la categoría (G)		
Poscondiciones		
Requisitos Especiales		
Prototipo		
Pantalla 1		

Conclusiones

En este capítulo se utilizó un modelo de dominio como ayuda para definir el contexto del sistema. El estudio de los requisitos que debe cumplir el sistema y seguido a esto el análisis de los casos de usos, son etapas que contribuyen notablemente en la elaboración del sistema. Se detallaron los casos de uso y se observó como estos ayudan a enlazar todas las actividades del desarrollo.

El presente capítulo representa la primera fase en el desarrollo de la propuesta de solución, contribuye notablemente a la construcción del sistema.

Capítulo

4

CONSTRUCCIÓN DE LA SOLUCIÓN PROPUESTA

Introducción

En el presente capítulo se modela el sistema y se describe la arquitectura del mismo. Se muestra el modelo de datos y diagrama de clases persistentes para comprender mejor que parte de la información ha de ser almacenada.

En la descripción de la arquitectura se realiza la vista del modelo de despliegue, donde se muestran los artefactos relevantes para la arquitectura y se describe la configuración sobre la cual deberá implementarse el sistema. Se exponen las pautas seguidas para el diseño de la capa de presentación de la aplicación. Se muestra además, cómo las interfaces han sido elaboradas, la estructura de la ayuda y los reportes realizados.

Además, se especifican los estándares a seguir en la codificación, para lograr una mejor organización y un estilo propio del software. Y por último se presenta un modelo de despliegue donde se definen los componentes que conforman la estructura física de la aplicación.

Diagrama de Clases

El sistema se ha modelado usando el estilo Capas sobre una arquitectura cliente servidor, específicamente la variante de las 3 capas, Cada una puede contener paquetes, en dependencia del nivel de complejidad que presenta, el propósito de dividir en subpaquetes, posibilita mayor entendimiento, independencia, legibilidad y organización a la hora de modelar y entender el sistema.

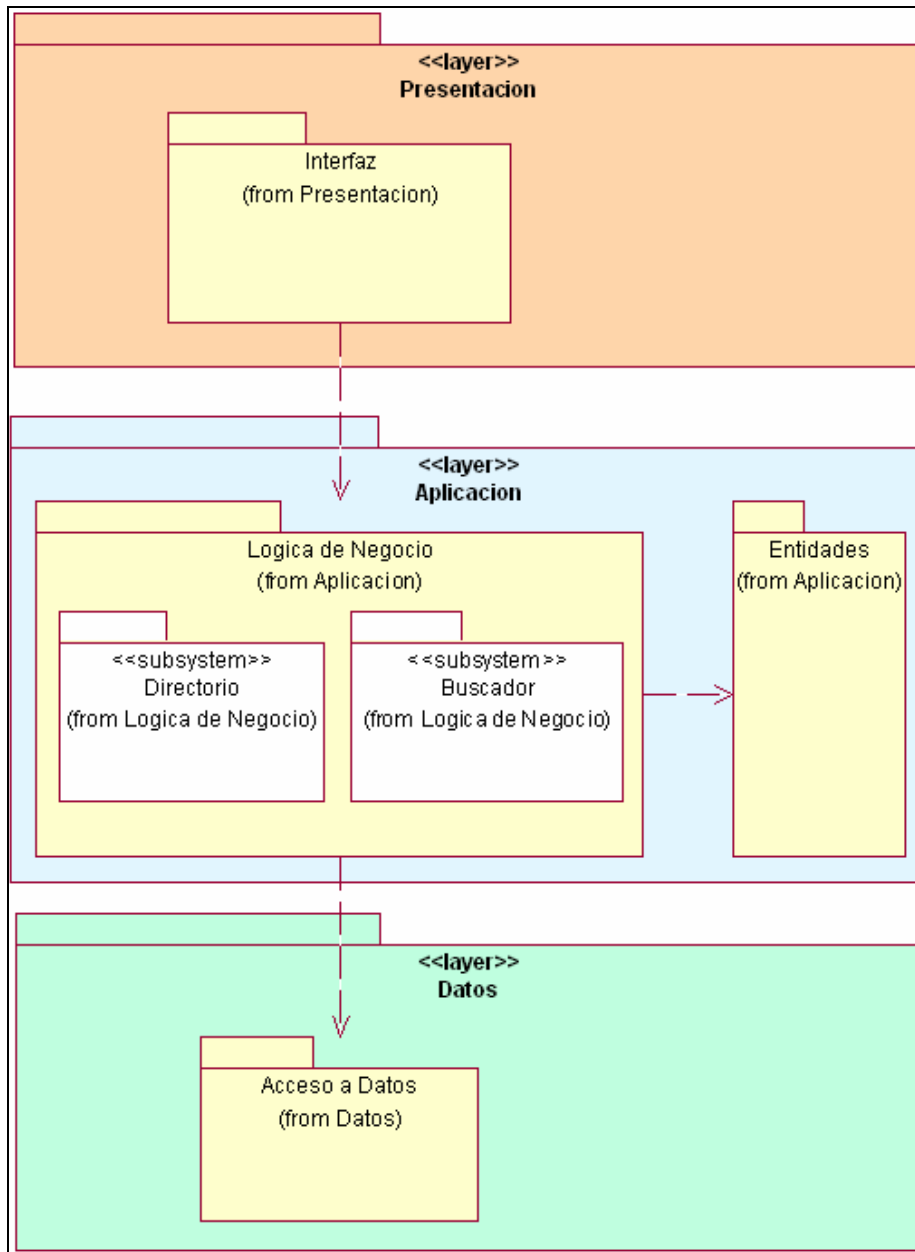


Figura 6 Arquitectura del Sistema

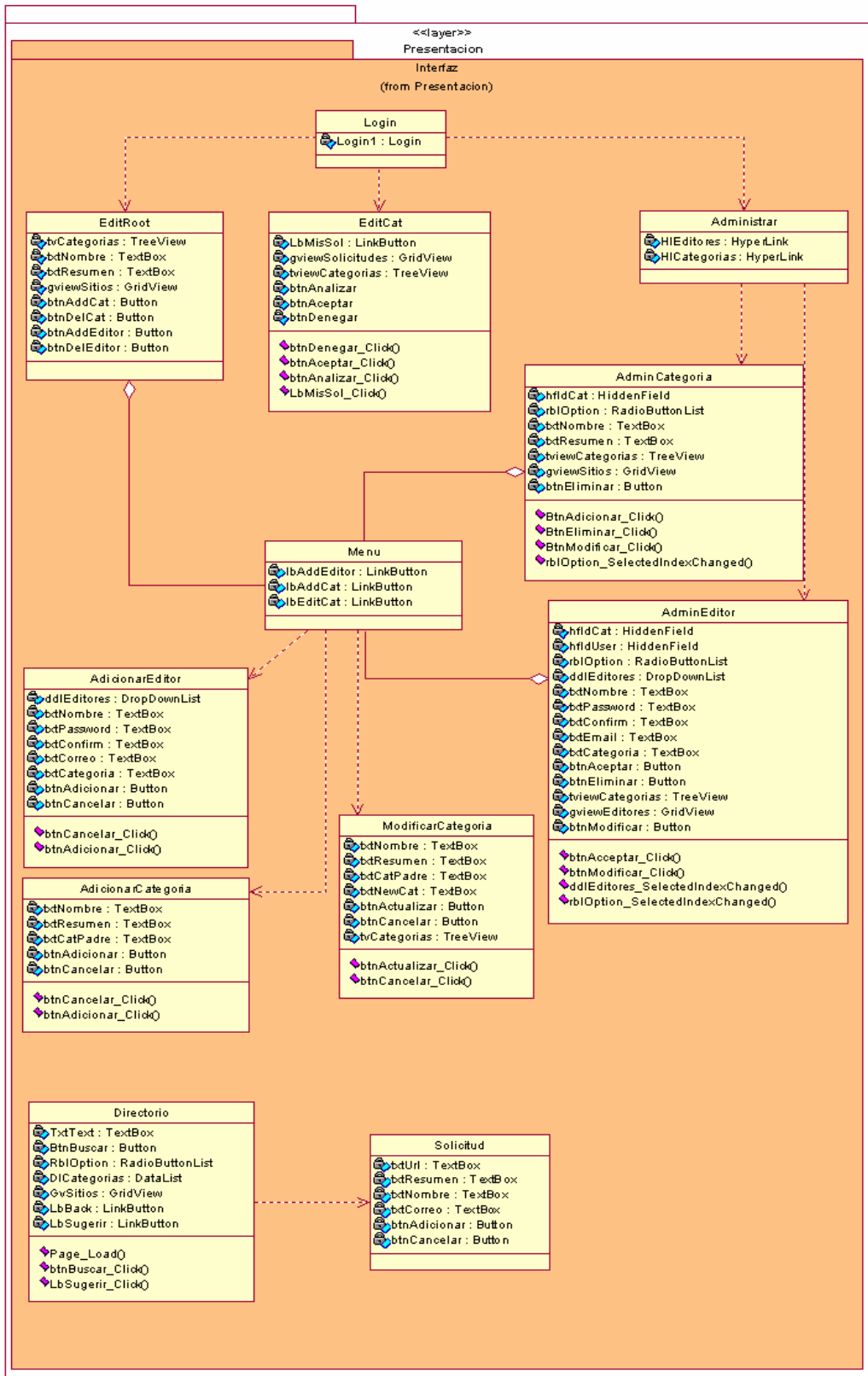
La **Capa de Presentación** es la encargada de interactuar con los usuarios y se corresponde con lo que normalmente se conoce como interfaz de usuario. Contiene un subpaquete Páginas que abarca las clases de interfaz, las que interactúan directamente con los usuarios.

La **Capa Aplicación** Aquí se localiza la Lógica del Negocio, esta capa recibe la petición del usuario a través de la capa de presentación y se encarga de darle curso a dichas peticiones, esta capa contiene todas las clases que se encargan de darle el principal funcionamiento al sistema, se realizan las validaciones, cálculos, etc. Contiene dos paquetes, uno Lógica del Negocio que contiene a su vez dos subsistemas en el que cada uno representa los principales componentes del directorio, un subsistema es el del *Directorio* que contiene todas las clases controladoras de las entidades del Directorio y el otro subsistema es *Buscador* que contiene las clases controladoras que definen el comportamiento del motor de búsqueda asociado al directorio. El otro paquete es el de las Entidades del Negocio, en este paquete se agrupan las clases que no tienen comportamiento, solamente son contenedoras de información.

La **Capa de Datos** Esta es la capa mediante la cual se accede a los repositorios de información. El subpaquete Acceso a Datos contiene las clases que posibilitan el acceso a la Base de Datos y manipular la persistencia de las Entidades.

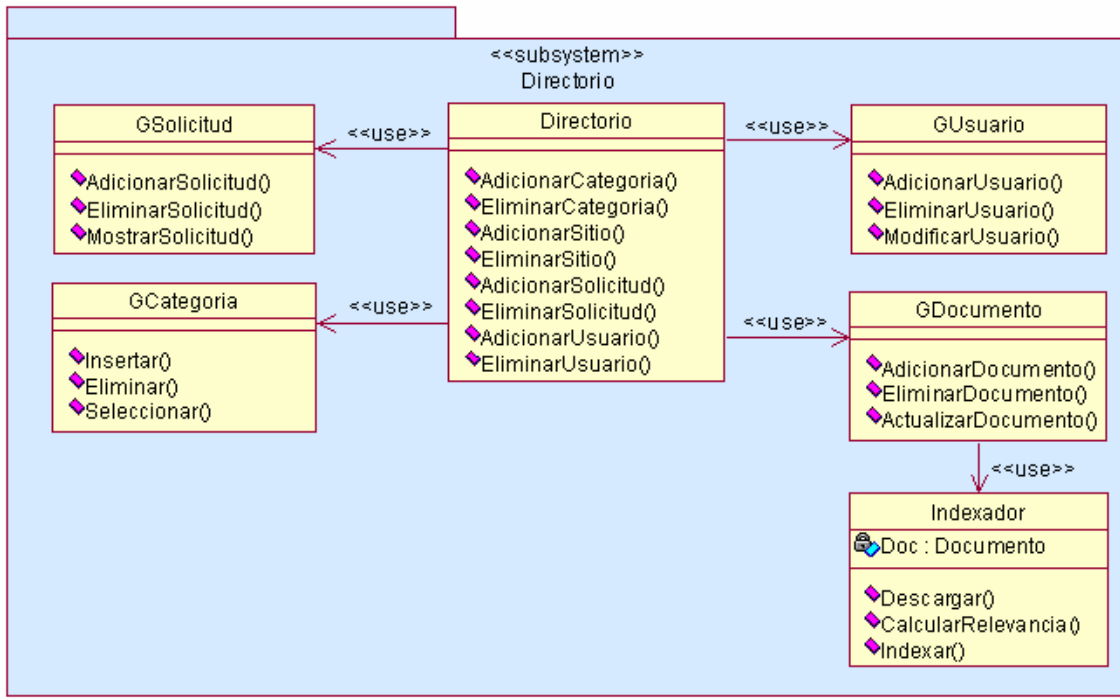
A continuación se muestran los diagramas de clases correspondientes a cada paquete.

Capa: Presentación
Subpaquete: Interfaz



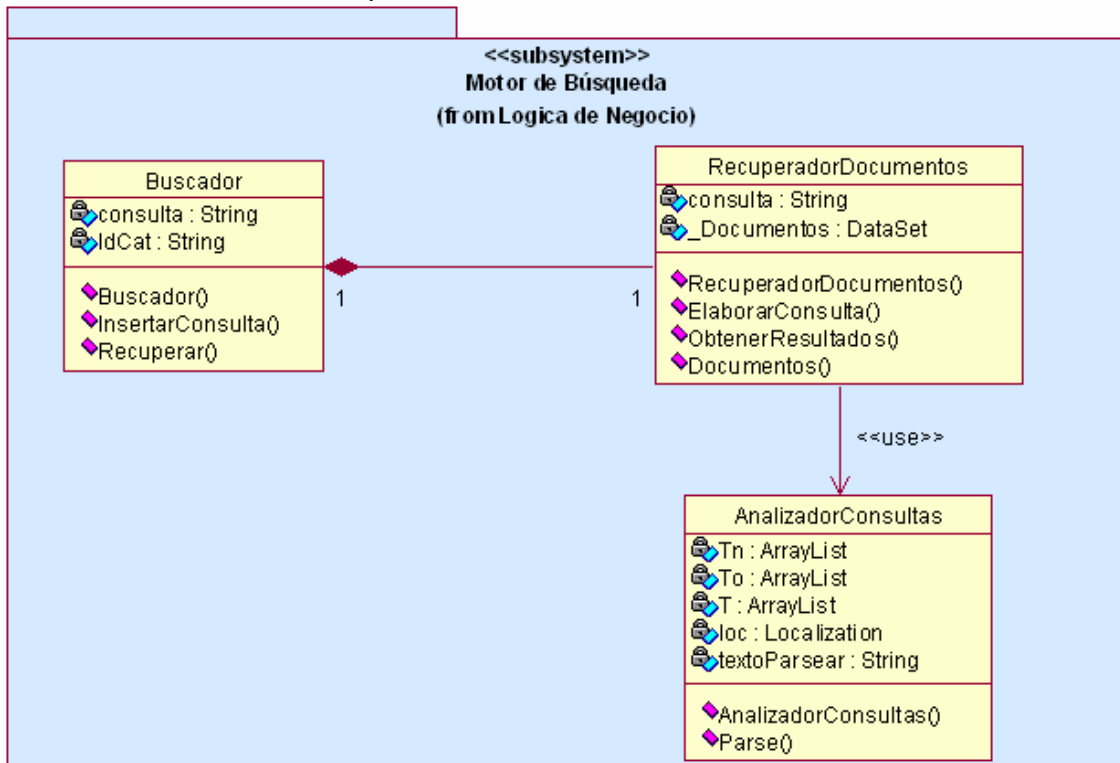
Capa: Aplicación

Subpaquete: Lógica del Negocio
Subsistema: Directorio

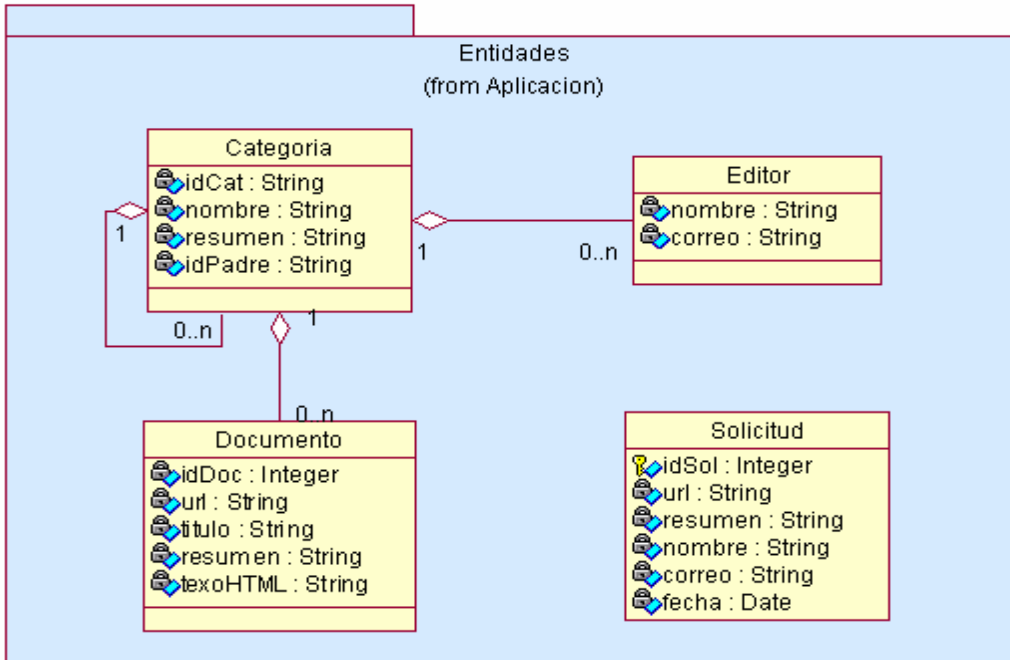


Capa: Aplicación

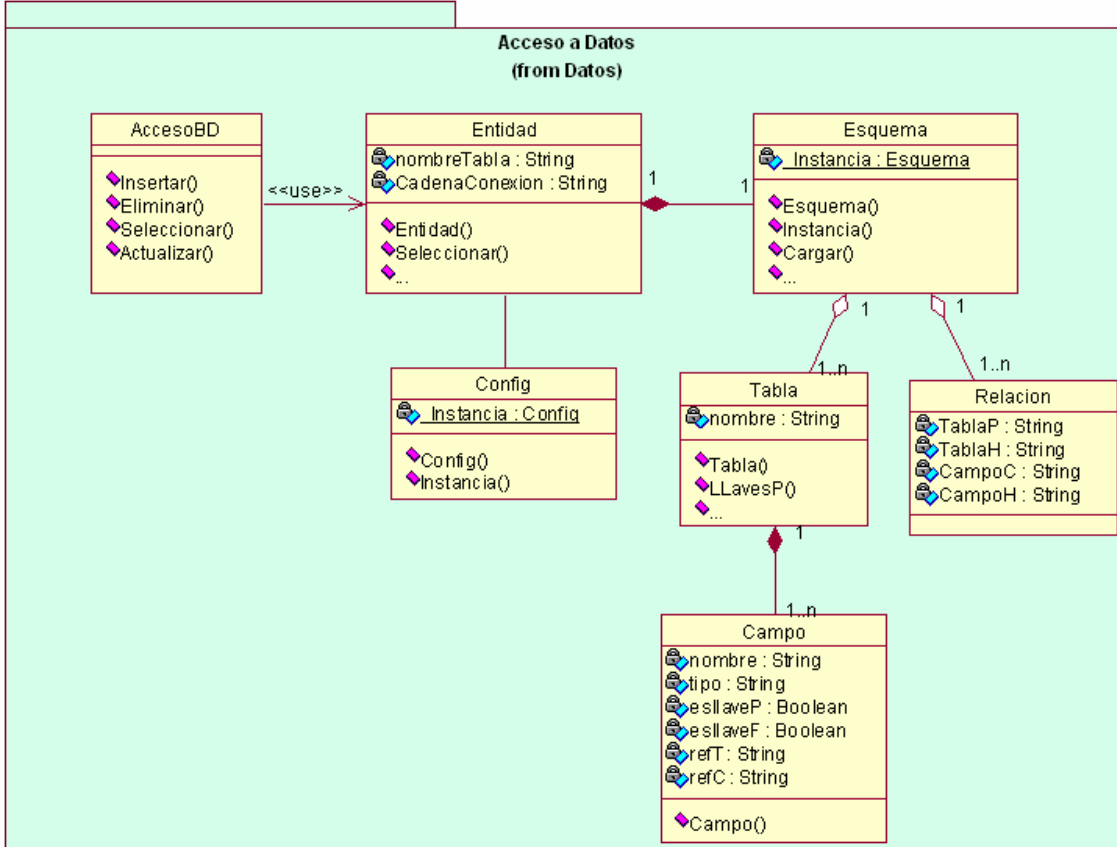
Subpaquete: Lógica del Negocio
Subsistema: Motor de Búsqueda



Capa: Aplicación
Subpaquete: Entidades



Capa: Datos
Subpaquete: Acceso a Datos



Diseño de la Base de Datos

No se muestra el diagrama de clases persistentes porque se corresponde exactamente con el de entidades. Si el del modelo de datos, donde se representa la estructura de la Base de Datos.

Modelo de Datos

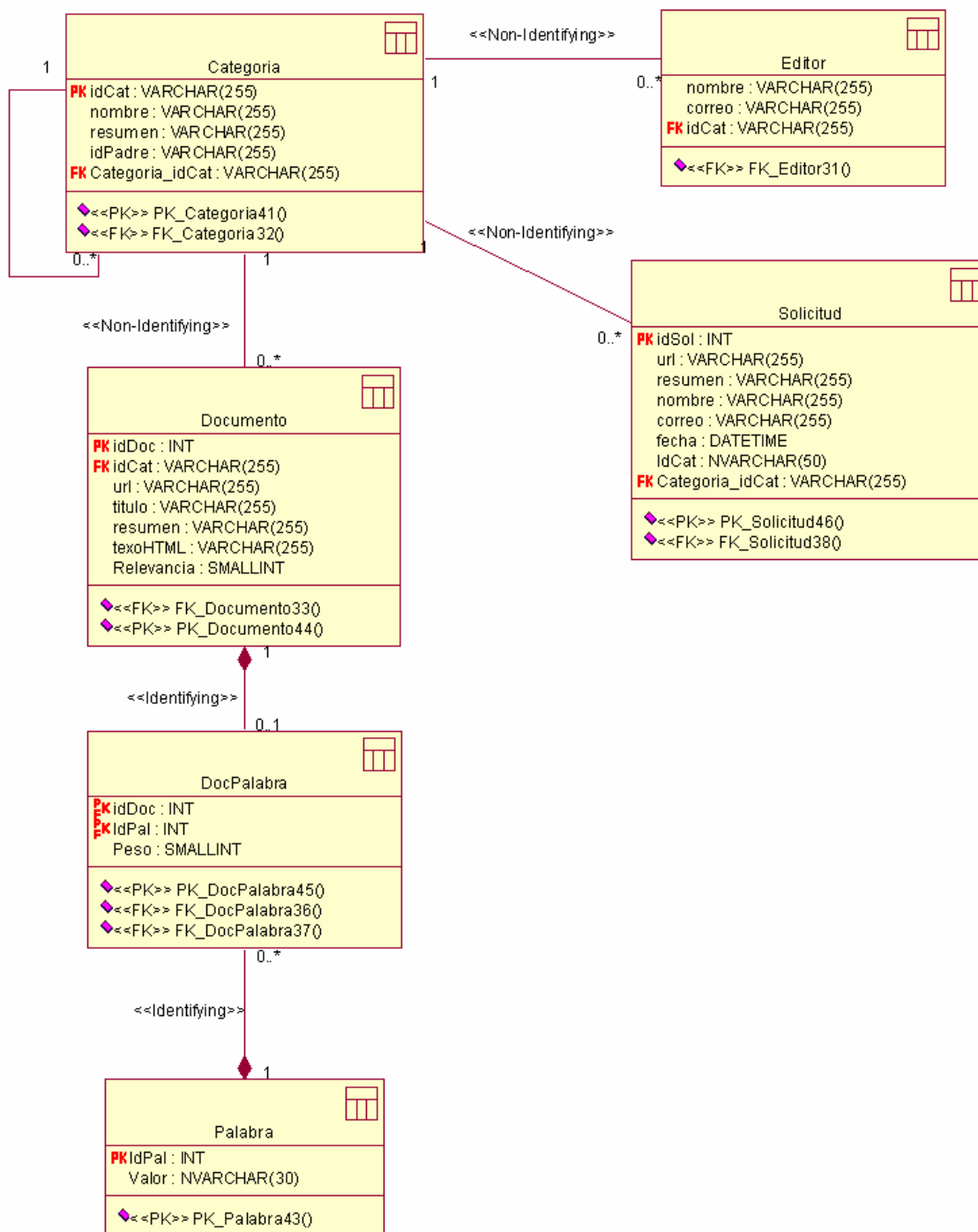


Figura 7 Modelo de Datos

Principios de Diseño

Presenta elementos que lo identifican como un sistema para el portal CubaSI, pero ha sido pensado y elaborado para que no contenga mucha sobrecarga la página en la que se muestren los resultados, esto contribuye a que sea más rápido en el momento de cargar la página al hacer un pedido al servidor.

Estándares de la Interfaz de la Aplicación

Cada página contiene en la parte superior una imagen representativa del Logo del Directorio.



Figura 8 Logo del Directorio

Cada página consta con alrededor de 3 colores. Como color primario se establece el blanco, color que actúa como equilibrio maravilloso entre los colores, hace las páginas agradables a la visión, "da sensación de elegancia", este es el tono básico de la página, el color que la define, se establece como atributo "background". Como color secundario el azul, es el que se utiliza para los links que se muestran en las páginas y como color destacado se emplea el rojo, utilizado para resaltar aspectos concretos de las páginas así como para mostrar mensajes, sea de error o de campos obligatorios.

En cada página se pueden realizar múltiples operaciones, posibilitando que el usuario no tenga que moverse sobre tantas páginas a menos que sea estrictamente necesario. La actualización y eliminación de los componentes se realiza en la misma página que muestra los listados.































Cada vez que sea necesario la entrada de datos por parte de un usuario, se le piden solamente los necesarios, todos aquellos que se puedan inferir o calcular son determinados por el sistema, por ejemplo si se va a sugerir un sitio, solo se le pide el url y

el resumen ya que los otros datos como título o keyword pueden determinarse si se tiene el url.

Formato de Salida de los Reportes

Los reportes se muestran en una de tabla y las categorías se muestran en forma de árbol, donde cada categoría tiene tantas categorías hijas como sean definidas.

En el caso de los reportes en tablas se brinda la posibilidad de ordenar por campos, esto facilita la búsqueda. Las filas de las tablas alternan en colores que no tienen mucho contraste, cada tabla presenta una fila título, con los nombres de los campos que resalta de las otras, se utiliza el color azul fuerte de fondo y las letras en blanco. Además las tablas presentan paginado, con un número de documentos por páginas, permite al final mediante el uso de flechas de dirección moverse hacia la siguiente o anterior página.

Url	Categoría	Solicitante		
 http://dafil.com	Computadoras	Alden Hernandez		
 http://dafil.com	Computadoras	José Valladares		
 http://intranet.uci.cu	Computadoras	Luis Alberto Perez Masjuan		
 http://jclub.cu	Computadoras	Daniela Hernandez		
 http://jclub.cu	Computadoras	Daniela Hernandez		
 http://feu.cu	Computadoras	Danayis Hdez		
 http://casa.cu	Computadoras	Danayis Hdez		
 http://hola.cu	Computadoras	Danayis Hdez		
 http://UJC.cu	Computadoras	Danayis Hdez		
 http://batalla.cu	Computadoras	Danayis Hdez		

1 2

Figura 9 Formato de salida de los reportes

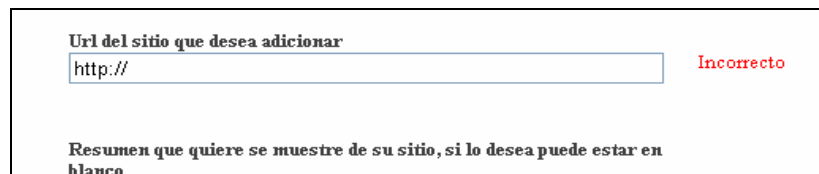
Concepción de la Ayuda

El sistema contará con un manual de usuario que estará disponible desde cualquier página del mismo, en dicho manual, se explicará el funcionamiento del sistema, y se mostrará una lista con preguntas frecuentes y sus respuestas.

Tratamiento de Excepciones

El tratamiento de excepciones es sumamente importante en cualquier aplicación. Las excepciones que puedan ocurrir en tiempo de ejecución se tratan haciendo uso del mecanismo de tratamiento de excepciones que brinda .NET, sin embargo se realizan una serie de validaciones del lado del cliente que contribuyen a evitar errores en tiempo de ejecución, por ejemplo se habilitan y deshabilitan los botones en dependencia del momento y si pueden o no ser utilizados, se realizan también validaciones que garantizan que los datos suministrados por los usuarios no tengan inconsistencias, se verifican los campos obligatorios y el tipo de datos a suministrar.

Todos los mensajes de error se muestran de color rojo, que resalta mucho a la vista.



Url del sitio que desea adicionar

http:// Incorrecto

Resumen que quiere se muestre de su sitio, si lo desea puede estar en blanco

Figura 10 Verificación del correcto formato de los campos

Estándares de Codificación

Los estándares de codificación son reglas específicas a un lenguaje que reducen perceptiblemente el riesgo a introducir errores, no destapan problemas existentes, mas bien ayudan a evitar que estos ocurran. Durante el desarrollo estos estándares ayudan a crear un código de alta calidad y a realzan considerablemente la capacidad de mantenimiento y reuso a largo plazo del producto final. Entre otras ventajas de programar usando estándares de codificación están:

- ✓ Reducir la posibilidad de producir errores.
- ✓ Hacer el código más uniforme y fácil de entender.
- ✓ Hacer el software más fácil de entender.

En esta aplicación se ha utilizado el estándar de codificación “Camel Case” que principalmente tiene que ver con la capitalización de los caracteres. Así mismo, se ha seguido el estilo de codificación propuesto por Microsoft para programar con C#.

```
public string AdicionarCategoria(string nombre, string resumen, string IdPadre)
{
    string IdCat;
    GCAtegoria Gestion = new GCAtegoria();
    DataSet dsCat = Gestion.Seleccionar("IdPadre= '" + IdPadre + "' AND IdCat <> '0'");
    int cant = dsCat.Tables[0].Rows.Count;

    if (IdPadre != "0")
        IdCat = IdPadre + (Convert.ToChar(cant + 97)).ToString();

    else
        IdCat = (Convert.ToChar(cant + 97)).ToString();

    Gestion.Insertar(IdCat, nombre, resumen, IdPadre);
    return IdCat;
}
```

Figura 11 Ejemplo de codificación utilizada

Declaraciones

En el momento de nombrar las clases, variables y funciones se ha tenido especial cuidado, en el caso de las variables se le antepone un prefijo al nombre indicando el tipo de variable que es, la siguiente lista muestra ejemplo de algunas declaraciones de tipos de variables.

- Campos de edición: **txt**Nombre.
- Campos de selección: **lbox**Nombre o **chk**Nombre
- Botones de acción: **btn**Nombre.
- Variables de control de ciclos: **i, j, k**
- Datalist: **dtn**Nombre
- Datasets: **ds**Nombre
- Conexión: **conn**Nombre

Modelo de Despliegue

El modelo de despliegue es un modelo de objetos que describe la distribución física del sistema en términos de cómo se distribuye la funcionalidad entre los nodos de cómputo, se utiliza como entrada fundamental en las entradas del modelo de diseño e implementación debido a que la distribución del sistema tiene una influencia principal en el diseño.

El sistema se ha construido siguiendo la arquitectura de 3 Capas, logrando así una independencia entre los componentes de Presentación, Aplicación y Datos. Dentro de la capa de Aplicación se han agrupado en dll diferentes los componentes relacionados directamente con la estructura del directorio y las clases que intervienen en el funcionamiento del motor de búsqueda asociado al directorio, de esta forma se consigue una independencia entre ambas, y se logra si cambia el funcionamiento del buscador no se afecte el funcionamiento del directorio y viceversa.

Existe un servidor SQL 2000 que contiene la BD, el servidor de aplicaciones reside en un Server separado que contiene el Internet Information Server (IIS), este Server tiene instalada la versión 2.0 del Framework de .NET. Los clientes son los distintos usuarios que hacen uso del sistema accediendo a través de la red mediante un navegador.

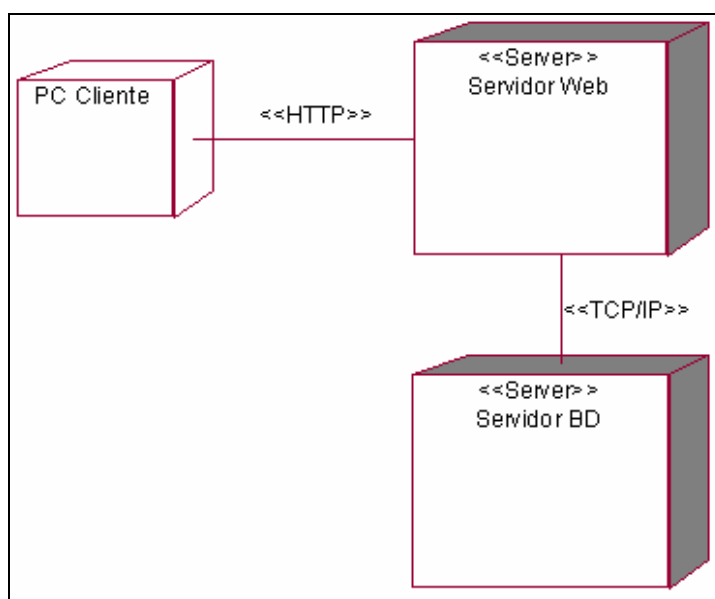
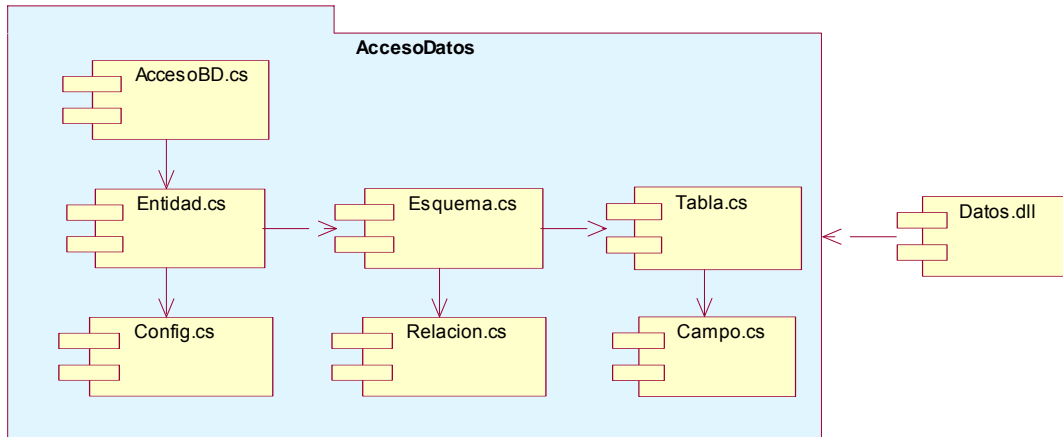


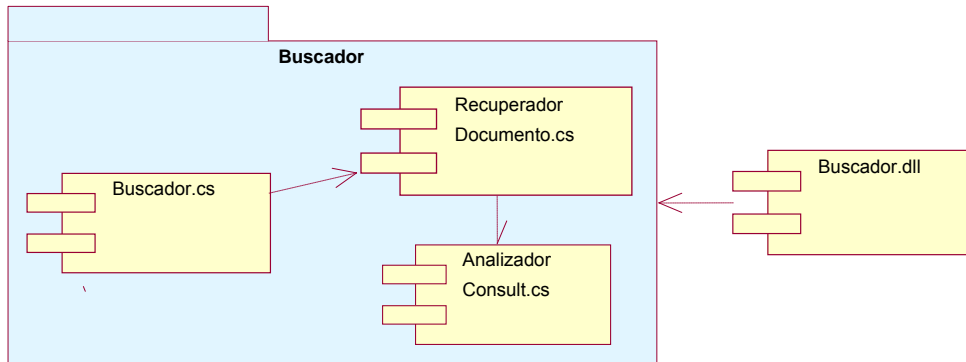
Figura 12 Diagrama de Despliegue

Modelo de Implementación

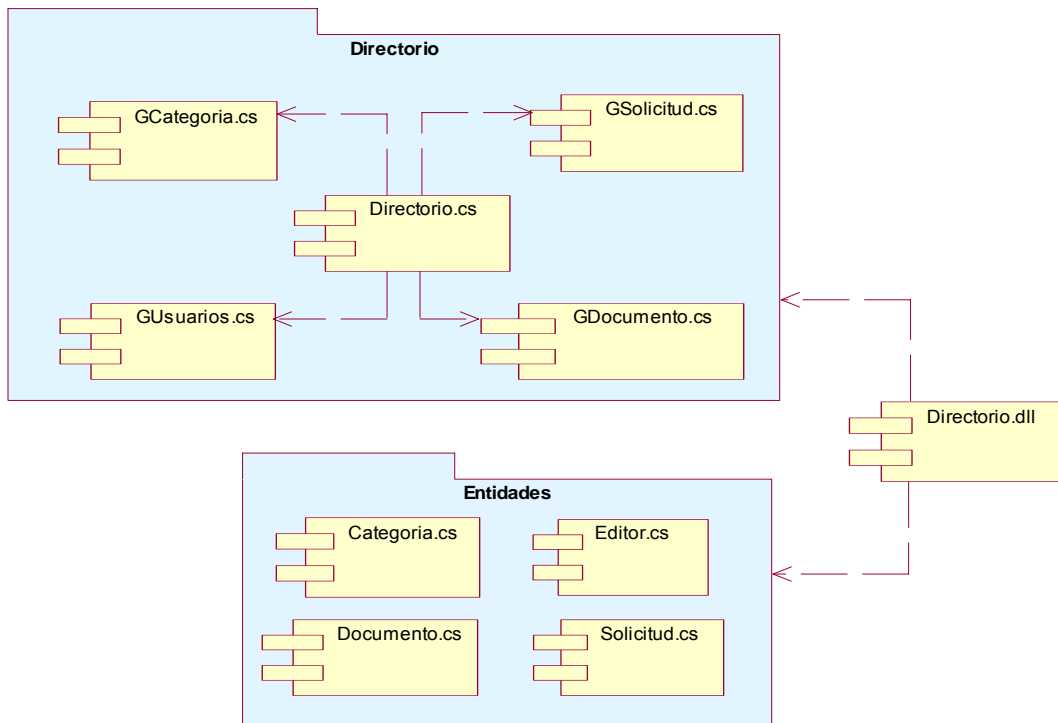
Paquete Acceso a Datos



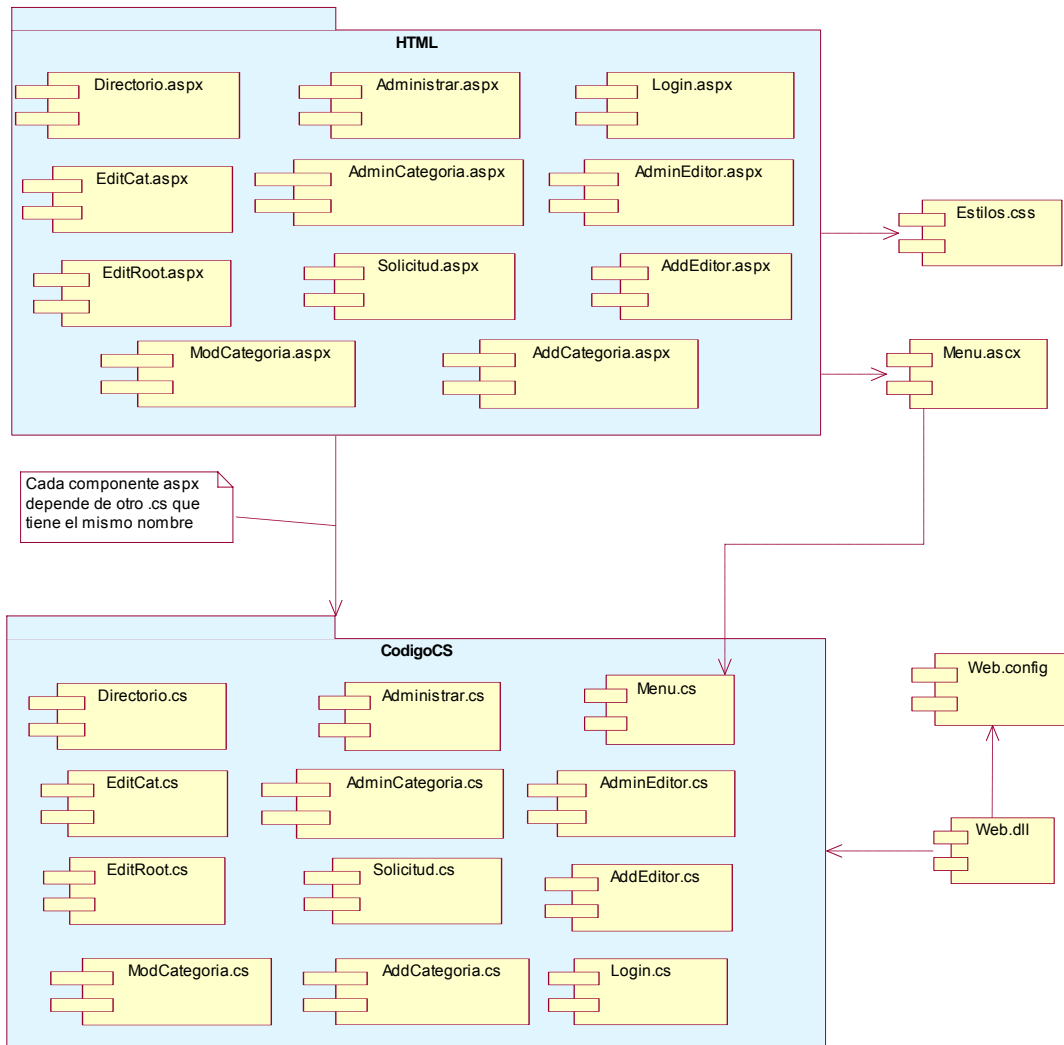
Paquete Buscador



Paquetes Entidades y Directorio



Paquete Presentación



Explicación de los componentes.

Se explicará cada paquete en el mismo orden en que se presentó.

Paquete **Acceso a Datos**. En este paquete se encuentran los componentes que contienen las implementaciones de las clases que se ocupan de manipular la persistencia de las clases persistentes del sistema.

Paquete **Buscador**: Aquí se encuentran los componentes que contienen las implementaciones de las clases encargadas de manejar lo referente al motor de búsqueda asociado al directorio.

Paquete **Directorio** Contiene los componentes que contienen las implementaciones de las clases que dan funcionamiento al directorio, las principales clases controladoras encargadas de gestionar lo referente a las categorías, sitios, solicitudes, etc.

El paquete **Entidades** contiene los componentes donde se encuentran las implementaciones de las clases entidades del sistema.

En el paquete **Código CS** se encuentran los códigos fuentes de las páginas aspx,

Otros componentes

Datos.dll Librería que agrupa las clases encargadas de manejar la persistencia de las clases.

Buscador.dll En esta librería se agrupan las clases encargadas del comportamiento del motor de búsqueda. Necesaria para el funcionamiento del motor de búsqueda del sistema.

Directorio.dll Librería necesaria para gestionar las funciones básicas del directorio, contiene la implementación de las clases controladoras del directorio.

Web.dll Librería necesaria para el funcionamiento de las páginas de la aplicación, contiene la implementación de las clases del paquete de presentación.

Estilos.css: Hoja de estilos para las páginas, contiene los estilos definidos para el sitio.

Menu.ascx: Este elemento es un UserControl, que tiene como objetivo el poder ser reutilizado en algunas páginas con el propósito de brindar un control reutilizable que facilite la navegación del sitio para los administradores y editores avanzados, preemitiéndole las opciones de adicionar un editor, y adicionar y modificar categorías.

Web.config: Archivo de configuración del sitio con formato de XML que contiene datos globales necesarios para diversas páginas.

Esquema.xml: Elemento que contiene el esquema de la Base de Datos. Este archivo es utilizado en el paquete de Acceso a Datos.

Config.xml: Archivo que contiene algunos datos necesarios para el correcto funcionamiento del paquete de acceso a datos, como por ejemplo la conexión a la BD

Páginas Web

Componente	Propósito	Contenido
Directorio.aspx	Página principal donde se lleva a cabo la navegación del directorio.	Ver CU realizar Búsqueda
Login.aspx	Página donde se lleva a cabo la autenticación del sitio.	Ver CU Autenticar Usuario (Anexo 1)
Administrar.aspx	Página a la que accede el administrador y a la que se le presentan las opciones disponibles	Ver CU Configurar Sistema
EditCat.aspx	Página a la que accede un editor cuando se autentifica en el sistema	Ver CU Editar Categoría
AdminCategoría.aspx	Página a la que accede el administrador cuando selecciona la opción de configurar categorías.	Ver CU Configurar Sistema
AdminEditor.aspx	Página a la que accede el administrador cuando selecciona la opción de configurar editores.	Ver CU Configurar Sistema
Solicitud.aspx	En esta página un usuario cualquiera del directorio realiza una solicitud de inclusión de un sitio en el directorio.	Ver CU Realizar Solicitud
AddCategoría.aspx	En esta página se lleva a cabo la adición de una categoría, esta operación puede ser realizada por un administrador o editor Avanzado del	Ver CU Editar Categoría

	sistema.	
AddEditor.aspx	En esta página se lleva a cabo la adición de un editor al directorio, la cual puede ser realizada por un administrador o un editor Avanzado del sistema	<i>Ver CU Editar Categoría o CU Configurar Sistema</i>
EditEditor.aspx	En esta página se procede a realizar alguna modificación, incluido eliminación de un editor del sistema	<i>Ver CU Configurar Sistema</i>
ModCategoria.aspx	En esta página se lleva a cabo cualquier operación sobre las categorías del directorio	<i>Ver CU Configurar Sistema.</i>

Conclusiones

En este capítulo se ha llevado a cabo la descripción de las clases y demás elementos necesarios para la implementación. Se obtuvo el diagrama de clases del sistema, separado por paquetes. Se definieron, cuáles son las clases persistentes, luego, a partir de esto, se construyó el modelo de datos. Se expusieron las pautas seguidas para el diseño de la interfaz, y se explicó cómo está estructurada la aplicación físicamente, mediante los modelos de despliegue y de componentes.

Conclusiones

Con este trabajo se crea un directorio automatizado que muestra la información de los sitios cubanos con dominio .cu agrupados por categorías, servicio que será suministrado por la entidad CubaSI, además permite realizar búsquedas corrientes, ya que presenta asociado un motor de búsqueda, capaz de ofrecer los resultados relacionados con las palabras suministradas por el usuario. Forma parte de lo que será el primer buscador mixto cubano, además posee una interfaz Web agradable y sencilla. La principal ventaja es que se crea uno de los componentes fundamentales de un buscador actual, un directorio cubano situado en uno de los portales Web más importantes y visitados de Cuba

El sistema se desarrolló siguiendo la metodología RUP, y se utilizaron representaciones UML para la modelación de todas las fases del proyecto.

El sistema resultante está provisto de un ambiente cómodo, fácil de entender, que cumple los estándares de diseño y utiliza técnicas modernas de programación orientada a objetos.

Por todo lo anterior se concluye que los objetivos propuestos para el presente proyecto han sido cumplidos satisfactoriamente. Se incluyen una serie de recomendaciones que deben tenerse en cuenta para el trabajo futuro.

Recomendaciones

Los objetivos de este trabajo han sido logrados, pero a lo largo de su desarrollo, han surgido ideas que podría implementarse en un futuro que mejorarían el rendimiento y harían que la aplicación fuese más útil y efectiva, para lo que se recomienda.

1. Utilizar el mecanismo de Caché de .NET para acelerar el proceso de búsqueda de forma que a medida que el usuario vaya realizando la búsqueda temáticas se guarden en caché las categorías mas probables a ser utilizadas.
2. Hacer un estudio de los principales navegadores existentes en Cuba y permitir que la aplicación pueda visualizarse en cada uno de ellos.
3. Estudiar las ventajas que traería desarrollar un servidor de Base de Datos con arquitectura distribuida donde existan varios servidores colaborando a las respuestas de las peticiones.
4. Idear algún mecanismo que sugiera la clasificación de sitios y de esta forma automatizar una parte de la construcción del directorio.
5. Estudiar las ventajas que conllevaría desarrollar la herramienta utilizando otras tecnologías en aras de mejorar el rendimiento del mismo.
6. Continuar el desarrollo de este sistema, adicionándole nuevas funcionalidades, adecuándolo más a las demandas de la creciente y dinámica de las aplicaciones Web que tiene lugar hoy día.

Referencias Bibliográficas


- [1]. Diseño e implementación de una Arquitectura multiplataforma para el estudio de Motores de búsqueda en Internet. <http://acoruna.tuportal.com>.
- [2]. Craig Larman. *UML y Patrones. Introducción al Análisis y Diseño Orientado a Objetos*. Prentice Hall Hispanoamérica, SA..., 1999.
- [3]. Temas avanzados de Enxeñería Telemática.
<http://www.tic.udc.es/~fidel/doctorado.htm>
- [4]. Conferencia # 9. *Extensiones para la Web*. Ingeniería de Software. (Curso 2004-2005)
http://docencia.uci.cu/is/cgi-bin/list_clases.pl?option=72
- [5]. Que es CGI <http://www.desarrolloweb.com/articulos/758.php?manual=27> (20/3/2006)
- [6]. Que es Perl <http://www.desarrolloweb.com/articulos/541.php?manual=27> (20/3/2006)
- [7]. Introducción a la Programación en ASP
<http://www.desarrolloweb.com/articulos/244.php?manual=8> (20/3/2006)

Bibliografía

1. Booch, G., Rumbaugh, J., Jacobson, I.: *El Proceso Unificado de Desarrollo de Software*. Addison-Wesley 2000.
2. Craig Larman. *UML y Patrones. Introducción al Análisis y Diseño Orientado a Objetos*. Prentice Hall Hispanoamérica, SA..., 1999.
3. Cruz A, José A. *Buscador Web* Trabajo de Diploma para optar por el título de Licenciado en Ciencia de la Computación. Universidad de las Ciencias Informáticas. Ciudad de la Habana, julio de 2004.
4. Diseño e implementación de una Arquitectura multiplataforma para el estudio de Motores de búsqueda en Internet. <http://acoruna.tuportal.com>. (10/1/2006)
5. Microsoft Visual Studio .NET Professional: información general acerca de las características.
<http://www.microsoft.com/spanish/msdn/vstudio/productinfo/features/profeatures.asp>
(12/1/2006)
6. Search Engine Watch: Tips About Internet Search Engines & Search Engine Submission <http://www.searchenginewatch.com/> (15/1/2006)
7. Soto, N., Saborit, Y Hubble: Propuesta para un sistema de catalogación y recuperación de recursos de información. Propuesta para optar por el título de Ingeniería en informática. Instituto Superior Politécnico “José Antonio Echeverría” Ciudad de la Habana, Junio de 2004.
8. Temas avanzados de Enxeñería Telemática.
<http://www.tic.udc.es/~fidel/doctorado.htm> (15/3/2006)
9. Manuales de desarrollo Web, diseño, programación.
<http://www.desarrolloweb.com/manuales/> (20/3/2006)

ANEXOS

Anexo 1 Caso de Uso expandido Autenticar Usuario

Caso de Uso	Autenticar Usuario	
Actores	Usuario	
Propósito	Registrarse en el sistema, con el propósito de realizar alguna operación de modificación o edición sobre el mismo.	
Resumen	El caso de uso se inicia un usuario accede a la página de Login con motivo de editar o administrar el sistema, introduce su nombre y contraseña, el sistema valida los datos y en dependencia del rol que tenga, dirige al usuario a la página que le corresponda.	
Referencias	R1	
Precondiciones		
Curso Normal de Eventos		
	Acción del Actor	Respuesta del Sistema
	1. El usuario accede a la página de login.	1.1. Muestra la ventana de entrada al sistema con los campos para que el usuario introduzca su nombre y password. <i>(Ver Pantalla 1)</i>
	2. Introduce los datos necesarios	2.1. Chequea los datos entrados y valida que el usuario se encuentre registrado en el sistema, en caso afirmativo, redirecciona al usuario a la página que le corresponde de acuerdo a su rol.
Poscondiciones		
Requisitos Especiales		
Prototipo		
	Pantalla 1	

