

Universidad de las Ciencias Informáticas

Facultad 10



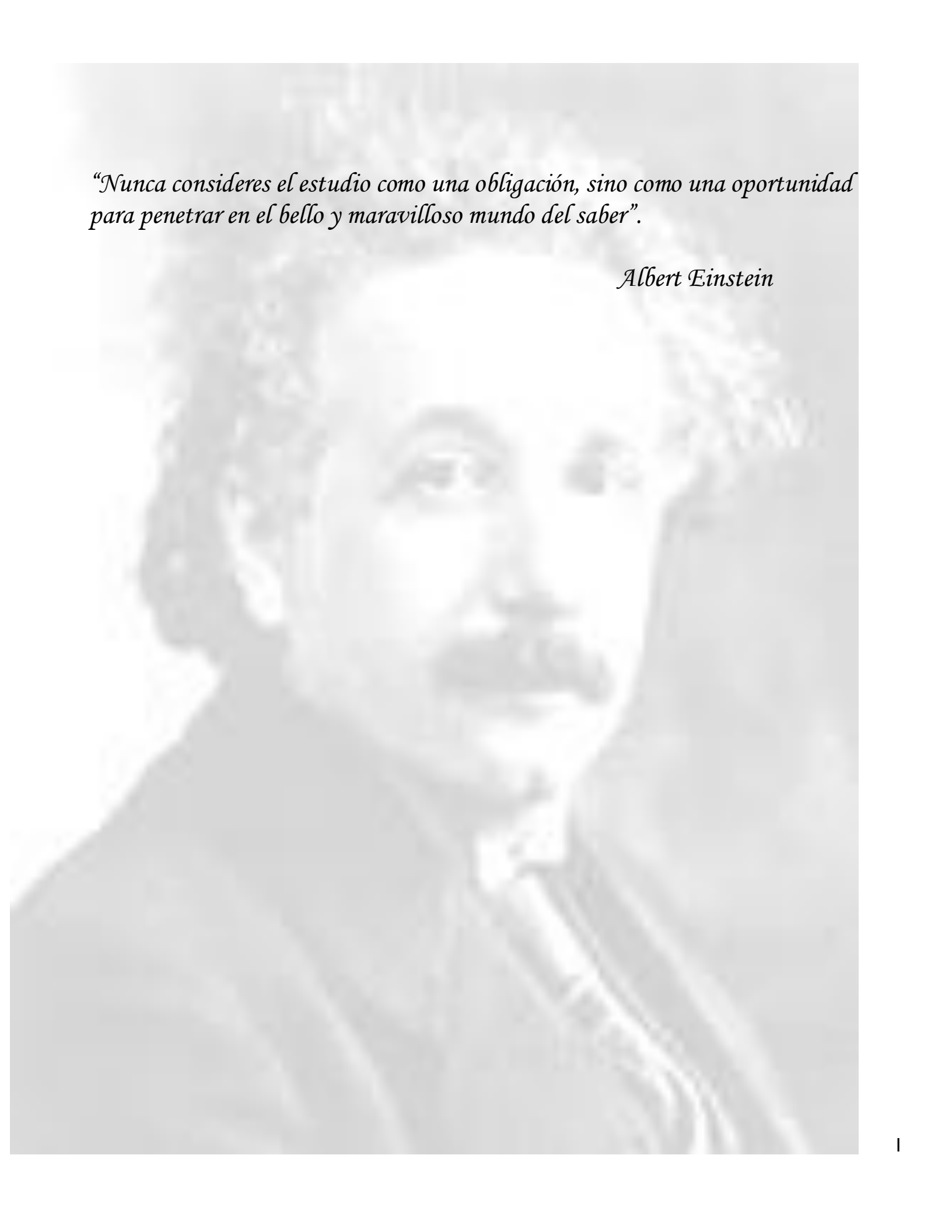
**Título: "Propuesta para Actualizar la Base de Datos de URLs
Categorizadas de Filpacon v2.0"**

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

Autora: Dania Fernández Aguilar

Tutora: MCs. Damaris Cruz Amarán.

Ciudad de la Habana, 17 de junio de 2009



“Nunca consideres el estudio como una obligación, sino como una oportunidad para penetrar en el bello y maravilloso mundo del saber”.

Albert Einstein

DECLARACIÓN DE AUTORÍA

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

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

Dania Fernández Aguilar

MCs. Damaris Cruz Amarán

AGRADECIMIENTOS

Hay momentos en la vida que son especiales por si solos, compartirlos con las personas que quieres, los convierte en inolvidables. Cuando la gratitud es tan absoluta las palabras sobran y precisamente en este momento no tengo palabras suficientes para agradecer a tantas personas que han tenido que ver con este momento tan especial en mi vida.

En primer lugar quiero agradecer a mi familia, las personas por las que me he convertido en todo lo que soy, por ser la razón de mi existir.

A mimi, por estar siempre a mi lado cuando todo se me pone oscuro, por su entrega a mi educación y formación. Porque nada de esto hubiese sido posible sin su amor, su comprensión, su apoyo y sus deseos de verme feliz. Por ser la mejor mamita del mundo.

A mis abuelos, mis tíos y primos por haber depositado en mí toda su confianza, por su interés y preocupación.

A mis fieles amigos Ledita, Yalena y Molina, por conservar nuestra amistad a pesar de las distancias, por apoyarme cuando mi familia no podía hacerlo, por ayudarme cada vez que lo necesité, por siempre darme ánimos para alcanzar esta meta.

A mis compañeras de cuarto en estos cinco años y mis amigas, Mirneyis, Ariannis y Wendy, por aguantar mi carácter, por ayudarme en todo momento incondicionalmente, por acompañarme en mis alegrías y tristezas durante esta importante etapa de mi vida.

A Damaris, mi tutora, por su tiempo, por darme la oportunidad de crear y de equivocarme por tener tanta paciencia y tratar de hacer de mi una excelente profesional. Muchas gracias.

Agradezco a todas y cada una de las personas que contribuyeron en mi formación como profesional y como ser humano, a todos los que de una forma u otra aportaron un granito de arena para que este sueño

Agradecimientos

se hiciera realidad, que de una manera u otra me ayudaron a lo largo de esta carrera, unos ya no están, otros sí, muchas gracias por estar siempre presente. A todos ustedes muchas gracias...

DEDICATORIA

Dedico este trabajo a mi familia que me ha apoyado en los momentos más difíciles por los que he pasado, por ser pacientes y consecuentes conmigo, por seguirme en mis ideas.

A mi mamá por estar siempre pendiente de mí, por su ejemplo, por querer que fuese siempre alguien en este mundo, por sus consejos, por entenderme, por soportar todos mis egoísmos y malcriadeces, por ser la inspiración de mi vida, una gran amiga, la mejor mamá del mundo, por haberme alentado en aquellos momentos en los que no tenía ni fuerzas para levantarme, por haberse mantenido a mi lado todo este tiempo luchando por mí y dando todo de sí y sobre todo por su plena confianza en mí: mi graduación, es mitad tuya.

A mis queridos abuelos maternos mis segundos padres otras de las personas que son especiales en mi vida y en mi corazón por su incondicionalidad y su eterno amor, por todos los regaños que me han dado, por aconsejarme en todos los momentos que lo necesité, por haberme guiado por un buen camino.

A mis tíos del alma por haberme ayudado en todos estos años, a mis primos que son los hermanos que no tuve los quiero mucho.

A todas mis amistades de la infancia y Universidad que nunca los olvidaré.

Dania Fernández Aguilar

RESUMEN

Filpacon es un sistema de filtrado de contenido que permite regular (permitir o denegar) el acceso de usuarios a determinados contenidos de la Web, lo cual facilita la utilización de la red de forma responsable. Para el buen funcionamiento del software, se necesita un sistema de actualización de su base de datos de URLs categorizadas de forma tal, que la renovación de su información se realice de manera sistemática, y aporte una mayor efectividad en su servicio.

Para lograr dicho objetivo se realiza la presente propuesta de solución, basada en la realización de forma semi-automática de la actualización de la base de datos de URLs categorizadas, la toma de los ficheros de las listas de Isak y Toulouse y la utilización de un Gestor de Archivos de URLs Categorizadas (GARUC) para su almacenamiento. El trabajo aporta una base teórica a partir de la investigación realizada y la documentación necesaria desarrollada para el uso del sistema de actualización que se implementará.

Palabras claves

Filpacon, URLs, GARUC, Sistema de filtrado de contenidos

TABLA DE CONTENIDOS

INTRODUCCIÓN.....	1
CAPÍTULO 1: Fundamentación teórica.....	6
1.1 Introducción	6
1.2 Fundamentos teóricos	6
1.2.1 Internet	6
1.2.2 Filtrado de contenidos	6
1.2.3 ¿Qué son las URLs?	7
1.3 Trabajos similares	8
1.3.1 Ámbito internacional	8
1.3.2 Ámbito nacional	9
1.4 Sistema de actualización.....	10
1.5 Archivos de URLs Categorizados	12
1.6 Generador de Archivos URLs Categorizados.....	13
1.6.1 ¿Qué son las listas?	13
1.6.2 Licencias para las URLs	14
1.6.3 Listas para actualizar las URLs	17
1.7 Tecnologías a utilizar.....	20
1.7.1 Lenguajes de programación	20
1.7.2 Entorno de desarrollo integrado	22
1.7.3 Control de versiones	24
1.7.4 Sistema Operativo	27
1.8 Conclusiones	29
CAPÍTULO 2: Descripción de la propuesta.....	30
2.1 Introducción	30
2.2 Descripción del sistema propuesto	30
2.3 Funcionalidades de los Módulos de Categorías y URL de la Interfaz Web de Filpacon	33
2.4 Herramienta y lenguaje para el desarrollo del software	33
2.4.1 Entorno de desarrollo integrado	33
2.4.2 Control de versiones	34

2.4.3	Lenguaje de programación.....	35
2.4.4	Sistema Operativo	36
2.5	Evaluación de la propuesta	36
2.5.1	Definición de indicadores	37
2.5.2	Elección de los especialistas.....	38
2.5.2.1	Determinar la cantidad de especialistas	38
2.5.2.2	Confeccionar el listado de especialistas	38
2.5.2.3	Confirmar la participación de los candidatos	39
2.6	Elaboración de los cuestionarios	39
2.7	Resultados de la evaluación	40
2.8	Conclusiones	43
	CONCLUSIONES GENERALES	44
	RECOMENDACIONES	45
	REFERENCIAS BIBLIOGRÁFICAS	44
	BIBLIOGRAFÍA.....	46
	ANEXOS.....	48
	GLOSARIO DE TÉRMINO.....	54

INTRODUCCIÓN

El avance tecnológico en materia de desarrollo de software a nivel mundial ha ido en incremento. Países desarrollados y otros no tan desarrollados se encuentran inmersos en la búsqueda de alternativas tecnológicas que hagan frente a la tan globalizada Era de la Comunicación. Hoy en día contenidos que se encuentran alojados en internet se han convertido en una amenaza para empresas y hogares, su consumo conlleva riesgos principalmente para grupos más vulnerables como el de los niños, que alejados de una atención adecuada por parte de padres y maestros les deja abierta una brecha para su acceso sin supervisión.

En la actualidad existen diversos programas, como los sistemas de filtrado de contenido, que aportan una solución a esta problemática y que son utilizados para prevenir sus efectos nocivos. Es una solución técnica que se ofrece a través de la cual, se regula el acceso de los usuarios en empresas, centros de enseñanza y hogares (Sánchez, 2008).

Los usuarios de Internet, generalmente, presentan vicios derivados de la dependencia que genera su uso constante, descuidándose aspectos personales o laborales. En la red es fácil encontrar información adecuada, y al mismo tiempo otras con características diferentes y desagradables como pornografía, violencia explícita, terrorismo entre otras que pueden afectar el normal desarrollo de los menores.

El acceso a la red de redes se realiza desde cualquier parte del mundo, lo cual incluye a Cuba, y una situación similar ante la problemática que se aborda es recurrente para los que hacen uso de este medio, en especial para la segunda red de computadoras más grande de Latinoamérica donde se encuentran conectadas alrededor de ocho mil computadoras: la Universidad de las Ciencias Informáticas y un riesgo para el programa de desarrollo de informatización de la sociedad cubana. Debido a lo anterior y a partir de la solicitud de la Oficina de Seguridad para las Redes Informáticas (OSRI) en el año 2004, la UCI comienza el desarrollo de un software de filtrado que lleva por nombre Filpacon.

Filpacon es un sistema de filtrado de contenido que permite regular (permitir o denegar) el acceso de usuarios a determinados contenidos de la WEB. Entre sus principales características se destacan su precisión y rapidez en el proceso de filtrado, la flexibilidad y adaptabilidad al lugar donde se aplique y una

de las más importantes es software libre. Además, de su uso en el ámbito empresarial, se aplica en escuelas, bibliotecas y lugares públicos de acceso a Internet.

El uso del software brinda ventajas que se manifiestan de la siguiente manera:

En las empresas:

- ◆ Mejora la productividad de sus empleados, mediante la denegación de los contenidos que les distraen en sus jornadas laborales.
- ◆ Reduce su responsabilidad legal por el acceso de sus empleados a contenidos ilegales o ilícitos de Internet.
- ◆ Propicia un medio viable para definir políticas o normas para el uso correcto de Internet.
- ◆ Mejora el rendimiento de la red, mediante la reducción del consumo de ancho de banda por contenidos pesados que, en la mayoría de los casos, carecen de interés para las empresas.

Para la sociedad:

- ◆ Proporciona una herramienta necesaria para regular el uso de Internet, dependiendo del contenido y las características del usuario.
- ◆ Propicia el uso creativo y beneficioso de Internet.
- ◆ Establece procedimientos adecuados para aprovechar plenamente las innumerables posibilidades que ofrece la red, favoreciendo la correcta utilización de la tecnología.

El producto permite manejar la red de una forma responsable y garantiza que los usuarios no accedan a contenidos que puedan tener efectos negativos en su aprendizaje, conducta y salud haciendo del ciberespacio un medio seguro para el crecimiento intelectual. Protege a grupos vulnerables, como el de los niños, de acceder a materiales inadecuados que pueden ser devastadores en el futuro desarrollo de su comportamiento como ciudadanos.

No obstante a los logros obtenidos en el desarrollo de la solución, en la UCI no se ha desarrollado un sistema capaz de actualizar su base de datos de manera sistemática, y que aporte una mayor efectividad en su servicio lo cual conlleva a que existan brechas en el diseño de la seguridad y en la regulación del

acceso de los usuarios a determinados contenidos de la WEB, desactualización en la información sobre el uso que de Internet hacen los usuarios y todo esto trae como consecuencia que exista una herramienta obsoleta para la correcta toma de decisiones por parte de los clientes.

Teniendo en cuenta la situación antes señalada, se identificó el siguiente **problema científico**: ¿Cómo lograr un sistema de actualización del producto informático Filpacon en la regulación del acceso de usuarios a determinados contenidos de la Web?

En consecuencia con lo anterior se determinó como **objeto de estudio** los mecanismos de actualización de la base de datos de URLs categorizadas de los sistemas de filtrado y como **campo de acción** la actualización de la base de datos de Filpacon v2.0.

Como **Objetivo General**: Diseñar una propuesta para la actualización de la Base de Datos de URLs Categorizadas de Filpacon v2.0 que permita mejorar la regulación del acceso de usuarios a determinados contenidos de la Web.

La **idea a defender** que le dio curso a la investigación es la siguiente: El diseño de una propuesta para la actualización de la Base de Datos de URLs Categorizadas de Filpacon v2.0 podrá garantizar mayor efectividad en el servicio de la regulación del acceso de usuarios a determinados contenidos de la Web.

Las **Tareas de investigación** que permitieron el desarrollo de la investigación estuvieron dirigidas a:

1. Determinación de los precedentes investigativos en la actualización de la base de datos de URLs categorizadas de los sistemas de filtrado de contenido.
2. Identificación de las fuentes (URL) a partir de las cuales se van a realizar las actualizaciones
3. Definición de la estructura que van a tener los ficheros de actualización y del proceso de actualización.
4. Diseño de una propuesta para la actualización de la Base de Datos de URLs Categorizadas de Filpacon v2.0

Aplicación de métodos científicos de investigación.

Los **métodos** utilizados en la investigación del nivel **teórico** son:

Método analítico-sintético: permitió el procesamiento de la información y la precisión de características de un sistema de actualización y con el estudio detallado de los conceptos y documentos relacionados con el tema se pudieron resumir los elementos más importantes y de mayor utilidad para el desarrollo del trabajo.

Método histórico-lógico: permitió determinar las tendencias actuales de desarrollo de las base de datos y analizar la trayectoria completa del trabajo.

Del nivel **empírico** se utilizaron los siguientes:

Entrevista: es una entrevista planificada entre el investigador y el entrevistado para obtener información cualitativa sobre el tema que se investiga.

Criterio de especialistas con los que se valida teóricamente los resultados de la investigación.

La significación práctica lo constituye la propuesta de solución, basada en la realización de forma semi-automática de la actualización de la base de datos de URLs categorizadas, la toma de los ficheros de las listas de Isak y Toulouse y la utilización de un Gestor de Archivos de URLs Categorizadas (GARUC) para su almacenamiento. El trabajo aporta una base teórica a partir de la investigación realizada y la documentación necesaria desarrollada para el uso del sistema de actualización que se implementará.

El documento final está estructurado de la siguiente manera:

Capítulo 1: Fundamentación teórica. Estado del Arte

En este capítulo se abordan conceptos fundamentales en el análisis del tema. Se hace un estudio del estado del arte relacionado con la existencia de soluciones a nivel internacional y nacional referente al sistema propuesto y de las herramientas y tecnología actuales con el fin de sustentar la propuesta más adecuada para la solución del problema.

Capítulo 2: Descripción de la propuesta

En este capítulo se desarrolla la propuesta de actualización de la base de datos de URLs categorizadas de Filpacon. Se especifica la mejor manera para la estructura de la actualización, cómo se debe realizar y el contenido que debe generarse. Se ofrecen los resultados de validación a que se somete la propuesta de solución.

CAPÍTULO 1: Fundamentación teórica

1.1 Introducción

En este capítulo se describen y explican las características y herramientas que puedan ser utilizadas para desarrollar una correcta actualización de la base de datos de Filpacon con el fin de proponer la más adecuada para darle solución al problema planteado. Se realiza un análisis de las actualizaciones de base de datos a nivel nacional e internacional.

1.2 Fundamentos teóricos

1.2.1 Internet

Algunos definen Internet como la Red de Redes y otros como la Autopista de la Información es una red mundial de computadoras interconectadas, la cual está hecha para unir muchas redes locales de ordenadores y gracias a ella, millones de personas tienen acceso a una cantidad extensa y diversa de información en línea. (Clarín, 1997)

Internet crece a un ritmo sorpresivamente rápido, constantemente se mejoran los canales de comunicación con el fin de aumentar la rapidez de envío y recepción de información. Cada día que pasa se publican en la red miles de documentos nuevos y constantemente aparecen nuevas posibilidades de su uso de Internet y se trazan nuevos términos para poder entenderse en ese mundo que no para de prosperar.

1.2.2 Filtrado de contenidos

Poco a poco Internet se está convirtiendo en un elemento imprescindible para el entretenimiento en la casa y para la educación de los niños y como es libre los menores pueden navegar en sitios dedicados al juego, pornografía, drogas, violencia o a otros sitios de riesgos. Además el 70% de las visitas a páginas web relacionadas con la pornografía se efectúan durante la jornada laboral. Para evitar este tipo de situaciones existen sistemas de filtrados de contenidos que son programas o servicios que permiten controlar el acceso a determinados contenidos en Internet.

Los programas de filtrado de contenido no se ocupan únicamente de “censurar” páginas en el ámbito privado o empresarial. Hay otros que además se ocupan de limitar la entrada de contenidos publicitarios y ventanas emergentes que aumenta el tiempo de descarga. El más conocido es Internet Watcher 2000, aunque también incluye sistemas de limitación de contenidos no publicitarios.

Existen en el mercado software de filtrado de acceso a Internet que son específicos para bloquear el acceso a diferentes sitios de la Web, por ejemplo están: Semáforo.net, OPTNET, Netfilter entre otros. Asimismo navegadores web como Internet Explorer y Netscape permiten el filtrado de contenidos de forma gratuita. (Pérez, 2005) Entre las ventajas que aporta un sistema de filtrado de contenidos pueden destacarse:

- Optimizar la navegación y descarga de recursos en Internet.
- Evitar el acceso de menores a Web inapropiadas.
- Obtener informes de la navegación realizada desde el centro.

Los sistemas de filtrado de contenido permiten que los administradores de red restrinjan el uso que los empleados hacen de Internet mediante categorías, grupos, usuarios, horas del día, días de la semana o cuotas de ancho de banda. Filtra el contenido Web con ayuda de una base de datos en continua expansión que incluye millones de URLs categorizadas para identificar y bloquear eficazmente los sitios Web no adecuados.

1.2.3 ¿Qué son las URLs?

Las URLs son una secuencia de caracteres que se usa para nombrar recursos de acuerdo a un formato estándar, como documentos e imágenes en Internet por su localización y se le asignan una única dirección a cada uno de los recursos de información disponibles en Internet. La URL de un recurso de información es su dirección en Internet, lo que permite que se encuentre y se muestre de forma adecuada.

Las URLs se identifican por su esquema, las cuales indican el protocolo de red que se necesita para recuperar la información del recurso identificado, a través de la red. Uno de los esquemas URL como mailto, http, ftp y file se iniciaron en 1994 Reques for Comments RFC 1630, pero el estándar que se utiliza en la actualidad es el STD 66 / RFC 3986. (Wikipedia, 2009)

Las URIs es un concepto que ha sido incorporado dentro de otro más general que es la URI (Identificador Uniforme de Recursos). Las URIs fueron definidas en las normas IETF RFC 2397 y pueden ser más simples que otros métodos de inclusión alternativos como MIME. Actualmente las URIs están soportadas por: Gecko en el cual se encuentra Firefox, Opera y XHTML y sus derivados, Safari, Konqueror y Google Chrome. (Wikipedia, 2009)

Una URL tiene tres partes:

- Siglas del Protocolo: indican qué protocolo vamos a usar para la transmisión de datos. Lo normal es usar el protocolo de hipertexto, o sea, páginas Web, que es el HTTP (HyperText Transfer Protocol: Protocolo de Transferencia de Hipertexto). Normalmente, se pone después del protocolo los siguientes caracteres: "://", por lo que las siglas del protocolo de páginas Web serán: "http://".
- Nombre de Dominio del ordenador servidor: indica qué ordenador servidor de Internet nos ofrece la información solicitada. Normalmente, los ordenadores servidores de páginas Web tienen por nombre de dominio algo parecido a lo siguiente: www.nombre.es. Las tres letras "w" del principio indican que el ordenador está destinado a servir páginas Web en Internet.
- Ruta en el servidor: indica la posición del fichero que específicamente se ha solicitado dentro del ordenador servidor. Los ficheros están en el servidor ordenados jerárquicamente en una estructura de directorios o carpetas, como si fuera un disco duro. Una carpeta se separa de su sucesora mediante una barra de dividir.

1.3 Trabajos similares

En nuestros días existen en el mercado disímiles técnicas de selección y filtrado de contenidos capacitados para evitar que materiales dañinos que están en internet lleguen a manos de los usuarios. El producto Filpacon es una solución a esta problemática y ante la necesidad de una actualización para las URLs categorizadas de la base de datos, hace falta el estudio de trabajos similares.

1.3.1 Ámbito internacional

En los sistemas de filtrado de contenidos el proceso de actualización de la base de datos de URLs categorizadas para el proceso de administración está ampliamente desarrollado. Desde este proceso se

puede producir archivos fundados en la información que se almacena por parte del sistema. Uno de los ejemplos de herramientas donde actualizan su base de datos es:

Panda Security for Domino Servers:

Filtro que detecta en memoria y en tiempo real los virus, gusanos, programas espía, ataques phishing, entre otros. Panda Security se actualiza a través de la consola de Administración AdminSecure. Con AdminSecure el administrador se ahorra la tarea de crear manualmente certificados cruzados en cada servidor antes de proceder a la instalación de la protección. Panda for Domino se actualiza por si solo de forma automática y transparente. Las actualizaciones son incrementales, por lo que solo se descargan los identificadores del malware aparecido desde la última actualización. (Security, 2009)

Optenet:

Filtro de selección de contenidos para ámbitos domésticos, educativos y laborales, de fácil instalación. Funciona a nivel de usuario y a nivel de servidor. Optenet dispone del motor de análisis de páginas web más potentes del mercado, listas de protección predefinidas y un sistema de personalización, que permite añadir o sacar direcciones de manera rápida. También permite filtrar la descarga de artículos, determinar horarios de navegación según perfiles de usuario y ver el histórico de navegación de cada usuario. Está disponible en español, inglés, francés e italiano. (González, 2009)

CyberSitter:

Filtro fácil de usar que bloquea 10 categorías de contenido. Las actualizaciones son automáticas una vez por semana, pero también pueden hacerse manualmente cuando se quiera. Además del filtro, se incorporan las opciones de grabar las webs visitadas por el menor y de bloquear información que entra o sale (preguntas, datos personales, palabras nocivas). (González, 2009)

1.3.2 Ámbito nacional

Filpacon constituye la única solución existente en el país, implicando novedad en cada uno de sus componentes. Dado que es imposible aprovechar alguna solución existente a nivel internacional, es

preciso crear un proceso para actualizar su base de datos. Es necesario un modo sencillo de analizar los datos almacenados por Filpacon y que a su vez brinde información útil.

1.4 Sistema de actualización

La Internet facilita acceso instantáneo a la información, y ha cambiado para siempre las formas en que las personas se comunican, aprenden y dirigen los negocios. Sin embargo, muchas de las más de 800 millones de páginas Web, no son apropiadas para ser visualizadas por todos los usuarios. Por ejemplo, las escuelas necesitan asegurarse de que los estudiantes no acceden a contenidos violentos o pornográficos, y las corporaciones desearían restringir a los empleados el acceso a sitios de búsqueda de trabajos.

Si bien es muy esperado el control de acceso a contenidos específicos, la tarea de implementar dicho control es ardua. Los métodos manuales son poco efectivos, debido a la naturaleza evolutiva de la Web y a su tamaño.

Se requiere una solución que determine que sitios Web hay que bloquear y que continuamente actualice la lista de dichas restricciones. Esta solución es un filtrado de contenidos basado en una suscripción. El siguiente cuadro muestra las ventajas de los filtrados de contenidos en cuanto a los proveedores, corporaciones y escuelas. (2003)

Proveedores de Servicios	Corporaciones	Escuelas
<ul style="list-style-type: none"> ● Ofrecer Clases de Servicio Restringidas 	<ul style="list-style-type: none"> ● Reducir las Responsabilidades Legales ● Maximizar la productividad de los empleados 	<ul style="list-style-type: none"> ● Cumplir las normas Gubernativas ● Blindar a los estudiantes de contenidos inapropiados
<ul style="list-style-type: none"> ● Descargar las redes congestionadas de tráfico NO esencial, ahorrando ancho de banda 		

Cuadro 1: ventajas del filtrado de contenidos

Las herramientas de filtrado de contenidos obligan a que los accesos a la Web de las organizaciones cumplan determinadas políticas, verificando cada usuario en una base de datos de URLs. Dado que esta base de datos contiene cientos de miles (y creciendo) de URLs, esta organizada por categorías, para facilitar las políticas de administración. Por ejemplo, una organización puede decidir el bloqueo de todos los sitios relativos a Apuestas, Sexo, o docenas de áreas relativas a otros sujetos. Dado que diariamente se construyen nuevos sitios Web, los servicios de filtrado de contenidos añaden, continuamente, nuevos sitios a la base de datos maestra.

Los líderes de los servicios de filtrado de contenidos basados en suscripción, son Secure Computing Corporation SmartFilter y WebSENSE Inc. (2003). El siguiente cuadro 2 muestra las características fundamentales de CacheFlow:

Característica	SmartFilter para CacheFlow	WebSENSE para CacheFlow
Integrado en el dispositivo	✓	✓
Base de datos de URLs potencialmente no deseables, organizadas por categorías como Apuestas, Armas, Drogas, Juegos, Personales, Sexo, Trabajos, etc.	✓	✓
Actualizaciones de la base de datos de filtros	A través de Internet, automáticamente, semanalmente	A través de Internet, automáticamente, diariamente
Filtrado de URLs definidos por el usuario	✓	✓
Políticas de administración en línea de comandos	✓	✓
Aplicación dinámica de políticas	✓	✓
Mensajes definibles tipo "contenido bloqueado" o redirección a una determinada página Web	✓	✓

Funcional tanto en modo Proxy como en modo Transparente (redirección HTTP mediante un conmutador Nivel 4 o router basado en WCCP)

Cuadro 2: Filtrado de Contenidos para CacheFlow: Relación de Características

En la actualidad el enfoque utilizado por Filpacon, es un sistema de filtrado de contenido mediante BDUC, el cual se materializa del siguiente modo: (Pérez, 2008)

- el usuario solicita una página web a través del filtro.
- el filtro busca en su BDUC la URL solicitada.
- si la URL no existe en la BDUC, dependiendo de las opciones de configuración del filtro, puede ser permitida o denegada.
- si la URL existe en la BDUC, dependiendo de la naturaleza de su categoría asociada, puede ser permitida o denegada.

1.5 Archivos de URLs Categorizados

Archivos

Un archivo informático es un conjunto de datos que se almacenan bajo un formato determinado el cual es una manera de codificar la información almacenada en el mismo. Dependiendo del tipo de contenido con el que se trabaje, los datos pueden ser almacenados en distintos formatos, por ejemplo:

ASCII: Debe su nombre al término inglés American Standard Code for Information Interchange y es un formato que almacena los datos a través de texto simple, es por ello que se utiliza para intercambio de datos o bien para generar archivos modificables por el usuario. Su contenido se puede visualizar y editar con cualquier procesador de palabras, como es el caso del Bloc de notas de Windows o el Simple Text de

Macintosh. Es un estándar que asigna un valor numérico a cada carácter, con lo que se pueden representar los documentos llamados de Texto Plano, es decir, los que son legibles.

Binario: A diferencia del ASCII, este formato almacena cualquier información como colores, órdenes, texto, imágenes, sonidos, etcétera, en un lenguaje entendible únicamente por la computadora. Con ello, se obtienen archivos de menor peso a los generados en formato ASCII. Al mostrar el contenido de un archivo binario con algún procesador de palabras, solamente se visualizará un conjunto de caracteres y símbolos extraños.

1.6 Generador de Archivos URLs Categorizados

1.6.1 ¿Qué son las listas?

Las blacklist son listas donde se registran las direcciones de IP que generan spam de forma voluntaria o involuntaria y son de libre distribución. (Wikipedia, 2009) Algunos de los términos relacionados con las blacklist son:

RBL es un acrónimo de "Real-time Blackhole List". Es el nombre del primer sistema que usó esta tecnología. Es una marca registrada de MAPS. Hay otras listas que son de libre distribución y que cumplen con los mismos requerimientos.

DNSBL es un acrónimo de "DNS blacklist". El uso de la palabra "lista negra" genera controversia. El término "rejectlist" ha sido usado por muchos desde sus inicios.

DRBL es un acrónimo de "Distributed Realtime Block List".

DNSWL es un acrónimo de "DNS whitelist". Es una lista de direcciones IP que se quiere tratar de manera correcta.

RHSBL es un acrónimo de "Right Hand Side Blacklist". Es similar a DNSBL con la diferencia que utiliza nombre de dominio en vez de direcciones IP. El término viene del lado derecho de las direcciones de mail "@dominio".

URIBL es un acrónimo de "Uniform Resource Identifier Blacklist". Al contrario de RHSBL, enumera los nombres del dominio usados en URIs y no en direcciones del E-mail.

1.6.2 Licencias para las URLs

Dada la necesidad e importancia de utilizar una licencia para liberar las listas de URLs de internet se describen a continuación las características de las licencias libres que son las que se tuvieron en cuenta para el desarrollo del trabajo.

Una licencia de software es un contrato que se lleva entre el propietario y el usuario del programa. Las licencias de software según los derechos de cada autor se pueden clasificar en:

a) Licencias de software libres sin protección heredada: en estas se puede crear un programa derivado sin que tenga obligación de protección. Algunas de las licencias que pertenecen a esta clasificación son:

- ◆ Academic Free License v.1.2.
- ◆ Apache Software License v.1.1.
- ◆ Artistic.
- ◆ Attribution Assurance license.
- ◆ BSD License.
- ◆ MIT License.
- ◆ University of Illinois/NCSA Open Source License;
- ◆ W3C Software Notice and License.

b) Licencias de software libre con protección heredada: Tienen algunas restricciones que se aplican a los programas derivados. Las licencias que se encuentran en esta categoría son:

- ◆ Artistic License.
- ◆ Common Public License v.1.0.

- ◆ GNU General Public License v.2.0.
- ◆ GNU General Public License v.3.0.
- ◆ GNU Lesser General Public License v.2.1.
- ◆ Mozilla Public License

c) Licencia de software no libre: Estas licencias también son conocidas como software privativo. Este tipo de licencias no permiten que el software sea modificado, desensamblado, copiado o distribuido de forma ilegal.

Software de dominio público: Son los que no tienen licencias y permiten el uso, copia, modificación o redistribución.

Existen diferentes tipos de licencias, en la tabla 1.1 se realiza una muestra comparativa de las licencias más importantes para el software no propietario, las cuales indican si son compatibles con la licencia GNU (GPL), y si están aprobadas por la Open Source Initiative. Además se puede deducir que el software libre y el de código abierto no son estrictamente similares y que la cualidad del copyleft no es una condición necesaria en estas licencias. (García, 2007).

Nombre	Descripción	Compat. GNU	Certif. OSI
Apache Software	Libre y abierta, con patentes	No	Si
Apple Public Software (APSL)	Libre, permite enlazar con ficheros propietarios	No	Si
Common Public (CPL)	Libre, con patentes	No	Si
Eclipse Public (EPL)	Libre, con patentes (menos agresiva que	No	Si

	CPL)		
Eiffel Forum (EFL)	Libre y abierta (la versión 1 no es compatible con GPL)	Si	-
GNU Public (GPL)	Libre, abierta, con copyleft	Si	Si
GNU Reducida (LGPL)	GPL sin copyleft, permite enlazar con módulos no libres	Si	Si
Lucent Public (Plan9)	Libre, incompatible GPL	No	Si
MIT/X Window	Libre, permisiva, copyleft limitado	Si	Si
Mozilla Public (MPL)	Libre, copyleft limitado, no enlazable con GPL	No	Si
Netscape Public (NPL)	Como MPL pero puede usar código propietario	No	No
OpenLDAP	Libre, permisiva, sin copyleft	V2.7	No
Open Software (OSL)	Libre, abierta, con copyleft reducido (según FSF)	No	Si
Perl	Licencia dual AL/GPL	Si	
PHP	Libre, sin copyleft (similar a BSD Original)	No	Si
Python	Libre (compatible GPL).	Si	Si

Tabla 1.1 Tabla comparativa de las principales licencias de software

1.6.3 Listas para actualizar las URLs

DMOZ:

Es un proyecto colaborativo que permite a editores listar y categorizar enlaces a páginas web, es también conocido por ODP (Open Directory Project), y en la actualidad es uno de los directorios más importantes de Internet. El derecho de copia de esta compilación le pertenece a Netscape Communications Corporation ('Netscape') y permite usarlo bajo los términos de la licencia y condiciones de uso del Open Directory. La licencia básica de Netscape concede una licencia gratuita y no exclusiva en la que se puede distribuir y publicar el Open Directory y sus trabajos, sometidos a todos los términos y condiciones de esta Licencia. (Project, 1999)

Cuando Nestcape asumió la administración del Open Directory Project este tenía alrededor de 100.000 URLs indexadas con contribuciones de aproximadamente 4.500 editores y ya para el 2007 habían contribuido al directorio 75.151 editores y tenía más de cuatro millones de URLs indexadas. El OPD distribuye sus datos en el vocablo RDF que es una colección de archivos acumulados por el Open Directory Project, tienen un formato XML y sus sitios están codificados en UTF-8, antes de esta las que utilizaban eran ISO 8859-1.

Para descargar los ficheros de DMOZ puede hacerse desde <http://rdf.dmoz.org/rdf> y una vez descargados los RDF se deberá tener un script propio para analizar la fecha, el ODP las etiquetas que utiliza actualmente se pueden encontrar en <http://rdf.dmoz.org/rdf/tags.html>. El proceso de creación de un nuevo horno comienza los lunes y actualmente se completa los viernes por la mañana todas las semanas. Los datos del ODP están disponibles para cualquier persona y son 100% libres de usar y deben de estar de acuerdo con la licencia ODP. (Corporation, 2005)

MESD:

Esta lista pertenece a Squidguard el cual es un sistema de filtrado web por listas negras, esto significa que tiene una larga lista de URLs y dominios para denegar o dar acceso al usuario, bloquea el acceso a URLs que tienen en ventas palabras o expresiones regulares para los usuarios.

Para actualizar las listas de MESD la RED HAT 7.2 RPM está configurando para que se sincronicen automáticamente las blacklists cada noche. MESD combina la blacklists de ftp://ftp.univ-tlse1.fr/pub/reseau/cache/squidguard_contrib/blacklists.tar.gz y <http://www.bn-paf.de/filter/de-blacklists.tar.gz>. La combinación de estas listas están disponible en: <http://squidguard.mesd.k12.or.us/blacklists.tgz>. El formato que utiliza MESD esta formado por una carpeta organizada por categorías en la cual se almacenan dos ficheros más de texto plano, donde están las URLs y los dominios y la extensión del archivo blacklists es tgz.gz.

Toulouse:

Puede ser utilizada con varios software comerciales y libres. Fecha de actualización: se renueva/actualiza, más o menos, dos veces por semana. Fecha y horario para descargas: domingos. (d'information, 2009)
Cómo se constituye:

Contribuidores, los cuales envían modificaciones.

Un Agente o Robot Web que busca URLs de adultos.

URL(s) para descargas:

- ◆ <http://cri.univ-tlse1.fr/blacklists/download/blacklists.tar.gz> y
- ◆ <ftp://ftp.univ-tlse1.fr/blacklist/blacklists.tar.gz>.

El fichero contiene una carpeta por cada categoría, dentro de la cual se almacenan 3 ficheros de texto plano:

urls (contiene un listado URLs). Por ejemplo: 20six.fr/VIX/weblogCategory/10alsxd6z3vsz

domains (contiene un listado de dominios). Por ejemplo: ameeotecarms.com

usage (especifica el tipo de contenido de la categoría). Por ejemplo: black.

Utiliza como licencia Creative Commons (<http://creativecommons.org/licenses/by-sa/2.0/es/>)

Reconocimiento. Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciante (pero no de una manera que sugiera que tiene su apoyo o apoyan el uso que hace de su obra).

Compartir bajo la misma licencia. Si altera o transforma esta obra, o genera una obra derivada, sólo puede distribuir la obra generada bajo una licencia idéntica a esta.

Al reutilizar o distribuir la obra, tiene que dejar bien claro los términos de la licencia de esta obra.

Alguna de estas condiciones puede no aplicarse si se obtiene el permiso del titular de los derechos de autor.

Nada en esta licencia menoscaba o restringe los derechos morales del autor.

iSAK:

En la siguiente página se puede encontrar las listas de iSAK <http://isak.gplindustries.com/wiki/Isakurldb>. Utiliza la licencia GPL en condiciones de uso.

La fecha de actualización: es actualizada cada día a las 3 am EDT y la fecha y horario para descargas es en la noche de cada día. (Kit, 2006) Las URL(s) para descargas son: <http://download.isak.gplindustries.com/isakurldbtext.tar.gz>. Se constituye de la siguiente manera:

List of domains and URLs from Open Directory Project (ODP – dmoz.org).

SquidGuard Blacklists.

http://cri.univ-tlse1.fr/documentations/cache/squidguard_en.html

<http://www.bn-paf.de/filter/>

<http://www.ingrid.org/~harada/filtering/>

iSAK administrators through the web interface.

1.7 Tecnologías a utilizar

Decidir desde el inicio del desarrollo de un software las tecnologías adecuadas a utilizar, contribuye a evitar el aumento del tiempo requerido y el descenso de la calidad del producto. Analizar las características particulares de la aplicación así como las ventajas que reportaría el uso de cada herramienta son elementos importantes a tener en cuenta en la selección.

1.7.1 Lenguajes de programación

Filpacon es un software que está basado en el sistema operativo Debian GNU/Linux, para el desarrollo de las listas existen diferentes lenguajes de programación, como Perl, JavaScript, Ruby y Python. Para elegir uno de ellos se deben tener en cuenta las características de la solución que se proponga y los conocimientos de los desarrolladores sobre estas. A continuación se describen algunas de las características de los lenguajes antes mencionados.

Perl:

Es un lenguaje de programación que originalmente fue desarrollado para procesar textos y que en la actualidad es utilizado para un amplio rango de tareas como administración de sistemas, desarrollo web, programación en red, desarrollo de GUI entre otras. Se considera como la herramienta más importante para la expansión de los sitios Web y más de la mitad de estos se construyen basados en el lenguaje Perl. Entre sus características está la facilidad de su uso y el lenguaje práctico que posee (fácil de usar, eficiente, y completo) en lugar de pequeño, elegante y mínimo. (Robert, 2009)

Es soporte de diferentes paradigmas de programación tanto estructural como orientada a objetos, es fácil de utilizar y tiene un robusto sistema de procesamiento de texto y una gran cantidad de módulos que constituyen un conjunto de funciones u objetos que son accedidos por los programas / scripts para realizar determinadas funcionalidades. Es un lenguaje extensible, pues permite hacer llamadas a múltiples programas desarrollados en otros lenguajes de programación. Su estructura está basada en bloques al estilo de C, lo que lo convierte en un lenguaje imperativo y toma características del lenguaje C/C++, el lenguaje interpretado shell, AWK, Lisp.

Entre sus ventajas se encuentran: rápido de crear, está disponible para distintos sistemas operativos, lo cual lo convierte en un lenguaje accesible a cualquier usuario y abre la posibilidad de ser muy usado y confiable. Perl tiene una gran cantidad de aplicaciones que se le pueden dar al lenguaje en campos como la administración de sistemas, aplicaciones web, entre otros.

JavaScript:

Es un lenguaje interpretado lo que facilita que no necesite compilación para obtener respuesta, utiliza una sintaxis parecida al del lenguaje Java y C. Entre sus características están que es un lenguaje interpretado por el navegador y no necesita tener instalado un Framework, y admite la programación orientada a objetos. Los navegadores como Internet Explorer, Netscape, Opera, Mozilla Firefox son soportados por JavaScript. (Valdés, 2007)

De sus ventajas se pueden destacar que es un lenguaje seguro y fiable, tiene capacidades limitadas por lo que se hace necesario utilizarlo conjuntamente con otros lenguajes más seguros como es el caso de Java. Su código puede ser visto y leído por cualquier usuario aunque esté protegido con las leyes del copyright.

Ruby:

Es un lenguaje reflexivo, interpretado y orientado a objetos, con características de SmallTalk, Lisp, Eiffel, Ada y Perl y su sintaxis está basada en Perl y Python. Entre sus características se pueden destacar la forma simple en que integra mecanismos como orientación a objetos, clausuras y macros y posee un potente vocablo sobre cadenas de caracteres y expresiones regulares, facilidad de desarrollo incremental y posibilidades de llamadas directas al sistema operativo. En la actualidad está entre los diez lenguajes de programación más importantes a nivel mundial. (Valdés, 2007)

Python:

Es un lenguaje dinámico orientado a objetos, que se puede desarrollar en distintas plataformas como son Unix, Windows, OS/2 y Mac OS X y ha sido portada a Java y .NET. Está distribuida bajo la licencia de código abierto lo que hace que sea fácil de usar (community, 2009). Entre sus características están que es

un lenguaje interpretado lo que significa que para ejecutarlo no es necesario compilar el código fuente que tiene como ventaja la rapidez de desarrollo, es un lenguaje de alto nivel y posee una extensa librería.

1.7.2 Entorno de desarrollo integrado

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

Eclipse:

Es un entorno de desarrollo integrado de software open-source (código abierto) independiente de una plataforma para desarrollar, está dividido en tres partes: el proyecto Eclipse Project, Eclipse Tools, y Eclipse Technology Project.

El Eclipse Project está subdividido a su vez en tres sub-proyectos que son la propia plataforma, JDT (Java Development Tool) y PDE (Plugin Development Environment). Mediante este IDE se pueden crear diversas aplicaciones como pueden ser, sitios web, programas Java, PHP, C++ y Enterprise Java Beans. Su principal aplicación es JDT que es la herramienta para crear aplicaciones en Java, además como su plataforma está construida en base de módulos o plug-ins esto permite que otras aplicaciones pueden ser integradas a eclipse en forma de plug-ins y ser reconocidos automáticamente por el IDE al iniciarse. Su interfaz de usuario está compuesta de un conjunto de vistas, editores y perspectivas. Los editores permiten crear, modificar y salvar objetos, las vistas proveen información acerca de los objetos con los que se están trabajando y las perspectivas proveen distintas formas de organización del proyecto. (Wikipedia, 2009)

Eclipse es administrado y dirigido por un consorcio de compañías de desarrollo de software con un interés comercial en promover Eclipse como plataforma compartida para herramientas de desarrollo de software.

Ventajas:

- ◆ Soporta la construcción de una variedad de herramientas para el desarrollo de aplicaciones.
- ◆ Es una herramienta de código abierto.
- ◆ Soporta el desarrollo de aplicaciones basadas en GUI y non-GUI.
- ◆ Corre en una gran cantidad de sistemas operativos incluyendo Windows y Linux.
- ◆ Soporta herramientas que manipulan diferentes tipos de archivos como por ejemplo Java, C, PHP, C++, EJB, HTML, GIF.
- ◆ Provee a los desarrolladores, herramientas que facilitan la creación de plugins.

Desventajas:

- ◆ Si bien Eclipse es multiplataforma, los plugins no tienen por qué serlo.
- ◆ Al ser una herramienta de código abierto, se desarrollan plugins que no tienen todas las funcionalidades que tienen en otras herramientas comerciales.
- ◆ Existen plugins que sólo corren en una plataforma, que aún no han sido desarrollados para más de una.

Plugins:

Un plugin es una aplicación que interactúa con otra para agregarle una funcionalidad específica y es ejecutada por la aplicación principal. En el caso particular de Eclipse son un conjunto de clases que permiten hacerlo más extensible.

ZendStudio:

ZendStudio divide sus funcionalidades en dos partes: la del cliente y la del servidor las cuales se instalan por separado. La parte del cliente contiene la interfaz de edición y la ayuda y además permite hacer depuraciones simples de scripts. Para disfrutar de toda la potencia de la herramienta de depuración se hace necesaria la parte del servidor que instala Apache y el módulo PHP o los configura para trabajar juntos en depuración en caso de que ya estén instalados.

El programa, además de servir de editor de texto para páginas PHP, proporciona una serie de ayudas que pasan desde la creación y gestión de proyectos hasta la depuración de código. El editor, la parte del programa que permite escribir los scripts, es muy útil para la programación PHP. A la hora de escribir brinda gran ayuda puesto a que permite:

Editar varios archivos y moverse fácilmente entre ellos.

Marcar a qué elementos corresponde los inicios y cierres de las Etiquetas, paréntesis o llaves.

Moverse al principio o al final de una función.

Identificar automáticamente el código.

En la interfaz encontramos un explorador de archivos, una ventana de depuración, otra para mostrar el código de las páginas y los menús. Además coloca puntos de parada en los scripts y realiza las acciones típicas de depuración. Dispone de una herramienta muy interesante de debug o depuración que permite ejecutar páginas y conocer en todo momento el contenido de las variables de la aplicación y las variables del entorno como las cookies, las recibidas por formulario o en la sesión.

NetBeans:

NetBeans es una herramienta para programadores pensada para escribir, compilar, depurar y ejecutar programas. Está escrito en Java, pero puede servir para cualquier otro lenguaje de programación. Existe además un número importante de plugins para extender el IDE NetBeans. Es un producto libre y gratuito sin restricciones de uso.

1.7.3 Control de versiones

El control de versiones es la gestión de versiones de todos los elementos de configuración que forman la línea base de un producto o una configuración del mismo. Los sistemas para el control de versiones facilitan la administración de las distintas versiones de cada producto desarrollado junto a las posibles especializaciones realizadas para algún cliente en específico. Sin embargo, los mismos conceptos son

aplicables a otros ámbitos y no solo para código fuente sino para documentos, imágenes, archivos xml, html, jsp.

Un sistema de control de versiones debe proporcionar un mecanismo de almacenaje de cada uno de los recursos que deba gestionarse (archivos de texto, documentación), posibilidad de mover, modificar, borrar cada uno de los elementos, un histórico de las acciones realizadas con cada elemento pudiendo volver a un estado anterior dentro de ese historial. Aunque un sistema de control de versiones puede realizarse de forma manual es muy aconsejable disponer de herramientas que faciliten esta gestión por ejemplo:

Subversión:

Es un software de control de versiones diseñado específicamente para remplazar al popular CVS, el cual posee varias deficiencias. El Subversión es un sistema de software libre y de código fuente abierta que maneja ficheros y directorios a través del tiempo, es empleado en la administración de archivos utilizados en el desarrollo del software; es uno de los controladores más utilizados en proyectos de software libre.

Características:

Fuerte integración con Apples, lo cual permite definir controles de acceso avanzados y navegación vía web para consultar el depósito de archivos.

En las copias ligeras sobre ramificaciones este controlador de versiones independientemente del número de ramificaciones creadas mantiene un árbol diferencial de cambios, minimizando así el espacio consumido en el depósito.

Contempla y corrige con éxito la transparencia al eliminar y cambiar nombres de archivos.

En las copias diferenciales de archivos binarios este controlador de versiones es capaz de mantener un control diferencial sobre cualquier archivo binario del depósito así reduciendo el consumo de espacio.

Gforge:

Es un software libre creado originalmente para SourceForge, permite organizar y administrar gran cantidades de proyectos y está licenciado bajo la licencia GPL (GNU General Public License). GForge ofrece alojamiento de control de versiones (CVS y Subversion), de seguimiento de fallos, mensajería y plug-ins para Eclipse IDE. Es una plataforma para el desarrollo de aplicaciones orientada a grupos de trabajo, es el mayor sitio web de desarrollo de software el cual cuenta con diversos servicios como:

Proporciona un entorno configurable con control de versiones.

Herramientas para comunicación entre desarrollos y servidor web por proyecto.

Permite a los miembros del equipo desarrollar una base de conocimiento compartida por proyecto.

Entre sus características están las diversas técnicas que tiene como: es libre, escalable y robusta, sus funcionalidades son la comunicación, centralización, agrupa fuentes de información y homogeneiza su modo de acceso y control. Una de las ventajas es que permite centralizar y homogeneizar la gestión de proyectos, es una página única y la centralización de los recursos técnicos en un servidor en vez de tener que soportar múltiples máquinas por proyecto. Actualmente se gestionan aproximadamente alrededor de 80000 a través del Gforge y miles de colaboradores participan en él todos los días. Por los menos 93 sitios web utilizan este sistema.

A continuación una figura que representa la interfaz gráfica del Gforge que utiliza el proyecto el cual se utiliza para registrar ítems de los proyectos que se realicen y darles seguimientos, tales como errores, correcciones, solicitudes de nuevas funcionalidades, campo de texto, entre otras cosas.



Cuadro 3: La interfaz del Gforge

1.7.4 Sistema Operativo

Windows

Windows XP es una opción de sistema operativo para empresas de todo tamaño, así como para usuarios de cómputo que desean sacar el mayor provecho de su experiencia de cómputo.

Ventajas

Incorpora tecnologías que mejoran su confiabilidad y seguridad, mejorando así la productividad general del usuario y garantizando la seguridad de los datos corporativos.

Es seguro, gracias a que tiene como base el núcleo de Windows 2000, Windows XP ofrece funciones diseñadas para proteger las computadoras.

Incluye un diseño innovador y fácil de usar.

Está diseñado para mejorar la experiencia del usuario, así como para ser más fácil de administrar y ser administrado. Es fácil de usar.

Linux

Linux es la denominación de un sistema operativo tipo Unix y el nombre de un núcleo. Es uno de los ejemplos más prominentes del software libre y del desarrollo del código abierto, cuyo código fuente está disponible públicamente, para que cualquier persona pueda libremente usarlo, estudiarlo, redistribuirlo, comercializarlo y, con los conocimientos informáticos adecuados, modificarlo. Las ventajas que presenta Linux se debe, entre otras razones, a su estabilidad, al acceso a las fuentes, que permite personalizar el funcionamiento y auditar la seguridad y privacidad de los datos tratados, a la independencia de proveedor, seguridad, rapidez con que incorpora los nuevos adelantos tecnológicos, escalabilidad, activa comunidad de desarrollo que hay a su alrededor, interoperabilidad y abundancia de documentación relativa a los procedimientos. Algunas distribuciones muy utilizadas son Debian, Ubuntu, Fedora y Red Hat, todas de libre distribución que pueden conseguirse fácilmente a través de las páginas Web de sus fabricantes o a través de archivos torrent.

Debian GNU/Linux

Debian GNU/Linux es un sistema operativo libre que soporta un total de doce arquitecturas de procesador e incluye los entornos de escritorio KDE, GNOME. Los sistemas Debian usan el núcleo Linux de ahí el nombre GNU/Linux. La distribución Debian es de muy buena calidad, tiene gran estabilidad y confianza. Aunque el proceso de instalación es un poco más complicado.

Ventajas:

- La disponibilidad en varias plataformas hardware.
- Una amplia colección de software disponible.
- Contiene un grupo de herramientas para facilitar el proceso de instalación y actualización del software.
- Su compromiso con los principios y valores involucrados en el movimiento del Software Libre.

1.8 Conclusiones

El análisis realizado permitió a la autora determinar los conceptos fundamentales como son los filtrados de contenidos, internet y URLs y las características e importancia los cuales sirvieron como puntos de reflexión para la toma de decisiones

El estudio de las diferentes metodologías y herramientas utilizadas internacionalmente permitió conocer los sistemas de actualización de base de datos para lograr una propuesta que permita que la renovación de la información se realice de manera sistemática, lo cual aporta una mayor efectividad en el servicio que ofrece Filpacon.

CAPÍTULO 2: Descripción de la propuesta

2.1 Introducción

En este capítulo se establece la descripción de la solución al problema propuesto, se selecciona y argumenta la herramienta, lenguaje y tecnología que sustenta la elaboración del presente trabajo y se analizan los resultados del diseño para asegurar la calidad del trabajo.

2.2 Descripción del sistema propuesto

Filpacon dispone de una base de datos en la que se encuentra una gran cantidad de URLs preclasificadas en distintas categorías que pueden pertenecer a su vez a tres tipos de contenidos diferentes: adecuados, nocivos e ilícitos, según las características propias del lugar donde se encuentre instalado. Ante la solicitud de un recurso de Internet por parte de un usuario, el servidor proxy squid-cache, que es parte inseparable del sistema, analiza los parámetros (identificador del usuario, dirección IP de origen y URL) a través de un programa hijo llamado redirector que interactúa con la BD. Según características del usuario como edad y política de navegación, la existencia o no de la URL en la BD y las categorías y tipo de contenido al que pertenezca la URL, se toma la decisión de permitir o denegar el recurso solicitado. (Sánchez, 2008)

Para lograr que la base de datos esté previamente actualizada se crea un sistema capaz de actualizar la base de datos de URLs categorizadas y así los administradores podrán eliminar, adicionar y cambiar las URLs y categorías de acuerdo a lo que sea mejor para el usuario e igualmente filtrar las URLs que proporcionan información inadecuada para el usuario. Con la colaboración de esta solución se estarían agregando funcionalidades que la base de datos de Filpacon anteriormente carecía que son muy necesarias para llevar un mejor control del sistema de filtrados de contenidos y será desarrollado con herramientas de software libre.

La propuesta de solución cuenta con dos módulos: actualización manual por parte de los administradores desde la Interfaz de Administración Web (IAW) y actualización semi-automática desde la IAW.

Actualización manual por parte de los administradores desde la IAW:

Filpacon cuenta con una base de datos de más de 3 millones de URLs categorizadas, estas se obtienen de sitios que las publican libres como son: DMOZ, Toulouse, Isak y MESD. Ese proceso se hacía manual, y para usuarios que no tienen conocimientos avanzados en tecnologías informáticas y para un administrador era muy poco factible, por lo que surge la necesidad de hacer un mecanismo que lo hiciera de forma automática o semi-automática y que los clientes de Filpacon las pudieran descargar de un sitio web y con el lenguaje de programación Perl se buscan y se añaden a la base de datos.

La descarga y automatización de los ficheros de actualización de la base de datos se realizará a través de cron el cual es un administrador regular de procesos en segundo plano que ejecuta procesos o scripts a intervalos regulares. Los procesos que deben ejecutarse y la hora en la que deben hacerlo se especifican en el archivo crontab. Para abrir un archivo en el crontab se hace con el comando crontab – el cual abrirá el editor definido en la variable EDITOR, para editar la tarea se especifica fecha y hora en la que se quiere se ejecute esta y se pone la ruta del script a ejecutar por ejemplo "0 1 24 may mon /scripts/comando.sh", Ctrl+o y luego Enter para guardar la tarea y Ctrl+x para salir del editor, si se quiere chequear que se ha programado la tarea poner en consola crontab -l. (Daishonin, 2006).

Actualización semi-automática desde la IAW:

Las listas que se manejarán para la actualización de la base de datos de URLs categorizadas de Filpacon serán Isak y Toulouse las cuales son las más confiables y de licencia libre, las mismas se bajarán los domingos cad

Para descargar las listas de los sitios se utilizará el siguiente módulo de Perl `www::curl::Easy`; y el módulo `archive::tar` para descompactar las listas que están en el formato `tar.tgz`, el código que se genera en Perl para el desarrollo de este procedimiento se muestra en el Anexo 4.

Para llevar las listas a la base de datos se emplea un archivo XML para guardar las categorías, el camino y el tipo de cada URLs. Se recorre el XML y se buscan las categorías y el camino el cual es el que da el fichero de las URLs de la categoría, posteriormente se busca en el fichero URL por URL que se encuentra en la base de datos y se comprueba la categoría que tiene la URL que está en la base de datos

con la nueva que se le está pasando y si el tipo de las categorías no son iguales no se adicionan y si son iguales se adiciona en la tabla de las categorías. El archivo XML se muestra en el Anexo 5.

Como formato interno se decidió utilizar el archivo de texto plano los cuales están compuesto por caracteres que van a ser codificados por el sistema de codificación UTF-8 que actualmente está estandarizado por RFC 3629. Entre las ventajas que tiene UTF-8 están que puede codificar cualquier carácter, se representan como 1 byte, aunque algunos símbolos pueden tomar más, de 2 de 4. Así, UTF-8 generalmente ahorrará espacio comparado con UTF-16 o UTF-32.



Gráfico 2.1: Actualización de la Base de Datos.

2.3 Funcionalidades de los Módulos de Categorías y URL de la Interfaz Web de Filpacon

En un principio se determinó que los módulos de Categoría y URL contarán con las siguientes funcionalidades:

Categoría

- ◆ Adicionar una nueva categoría.
- ◆ Modificar el tipo de contenido que una categoría existente tiene asociado.
- ◆ Eliminar una categoría existente.

URL

- ◆ Categorizar una URL
- ◆ Modificar la(s) categoría(s) asociada(s) a una URL
- ◆ Eliminar la(s) categoría(s) asociada(s) a una URL
- ◆ Conocer la(s) categoría(s) asociada(s) a una URL

En su nueva versión se proponen las funcionalidades con la diferencia de que en el módulo de categorías se le agregue un tipo de categoría y en el módulo de URL una descripción de lo que sería la URL. Para consultar anexo 2 y 3 respectivamente.

2.4 Herramienta y lenguaje para el desarrollo del software

A continuación se explican detalladamente las herramientas y lenguajes que serán utilizados en la construcción del sistema propuesto, basándose en las tendencias del país y de la Universidad de las Ciencias Informáticas dentro del mercado del desarrollo en software libre y se fundamenta su selección.

2.4.1 Entorno de desarrollo integrado

Se sugiere utilizar entornos de desarrollo para obtener aplicaciones robustas y flexibles ya que estos entornos te ayudan a organizar, probar y escribir el código, aumentando la calidad y productividad de la

aplicación que se va a producir. Entre los diferentes tipos de entornos de desarrollo que existen se va a utilizar el Eclipse 3.2. Es un potente editor con un buen completamiento de código y es basado en plugins .

Un plugin es una aplicación que interactúa con otra para agregarle una funcionalidad específica y es ejecutada por la aplicación principal. En el caso particular de Eclipse es un conjunto de clases que permiten hacerlo más extensible. Los módulos o plugins basado en la plataforma Eclipse que se utilizarán son los siguientes:

EPIC: Está licenciado bajo la licencia Common Public License 1.0. El EPIC tiene código fuente abierta que soporta como característica el resaltado de sintaxis, apoyo de perldoc y un depurador de Perl.

Visual Editor: Es un editor de código que brinda una ayuda para el diseño de aplicaciones que contienen interfaz gráfica de usuario (GUI). Permite editar el código fuente y el diseño simultáneamente además puede ser utilizado no sólo como una herramienta para generar código, sino como un editor para mostrar el efecto de las modificaciones de código fuente durante el desarrollo.

Permite un ciclo de desarrollo más transparente y predecible. Tiene muy buen soporte de refactorización mediante técnicas. Con esta versión se pueden utilizar todas las herramientas que un buen entorno de desarrollo integrado ofrece. Posee una agradable interfaz, es confiable y muy utilizado por los programadores de hoy en día.

2.4.2 Control de versiones

Se usa el control de versiones GForge siendo uno de los controladores usados en proyectos de software libre. La herramienta tiene integrada características como la creación de tareas y asignación de las mismas a usuarios, creación de Bugs y asignación a usuarios, permite la vinculación de commits con tareas/Bugs para poder trazar las revisiones en ambos sentidos, tiene plugins para visio, eclipse, MS Project y Maven y posee integrado un servidor Svn para poder crear un repositorio para el proyecto

Un usuario del GForge puede hospedar proyectos de desarrollo de software, si se está desarrollando algún producto de software, el GForge proporciona una gran cantidad de herramientas para administrar el

desarrollo y puede ser usada para compartir el código, si es que el desarrollo ya está terminado. Brinda distintos servicios tales como:

- ◆ Sitio web para el proyecto
- ◆ Lista de correo del proyecto.
- ◆ Foro de discusión del proyecto
- ◆ Un administrador de tareas por hacer, en el proyecto
- ◆ Espacio en el servidor web para subir los archivos fuentes del proyecto

2.4.3 Lenguaje de programación

El lenguaje de programación a utilizar es Perl escogido por ser un lenguaje interpretado y optimizado para la lectura y extracción de información de archivos de texto, generando reportes basados en la información proporcionada por ellos. Perl es un software libre y su código está licenciado bajo la Licencia Artística y la GNU General Public License, está disponible para distintos sistemas operativos, lo cual lo hace un lenguaje accesible a cualquier usuario y así convertirlo en un lenguaje confiable.

Perl se usa a menudo como un "lenguaje pegamento", ligando sistemas e interfaces que no fueron diseñados específicamente para interoperar, su combinación la hace una popular herramienta para los administradores de sistemas. Grandes proyectos escritos en Perl son Slash, IMDb y UseModWiki, un motor de Wiki y muchos sitios web con alto tráfico, como Amazon.com y Ticketmaster.com usan Perl ampliamente. Es un lenguaje multiplataforma y uno de los más populares y usados para la creación de aplicaciones Web, junto a Python y PHP.

Una de las grandes ventajas de utilizarlo, es que dispone de CPAN que es un directorio central donde se pueden encontrar módulos y programas creados por terceras personas en Perl que podemos usar en nuestros propios desarrollos. (Lizama, 2009) Actualmente existen más de 13.561 módulos disponibles, lo que se publica está disponible de forma libre, estandarizada y documentada.

2.4.4 Sistema Operativo

El sistema operativo que se va a utilizar es el Debian GNU/Linux 5.0 es la última versión que salió como versión estable y su nombre en clave es Lenny. Esta trae muchos más programas que su anterior etch, incluye más de 7700 paquetes nuevos y se actualizó el conjunto de aplicaciones de Mozilla: iceweasel (versión 3.0.6) es el navegador web sin marca de Firefox e icedove (versión 2.0.0.19) es el cliente de correo.

2.5 Evaluación de la propuesta

Para la validación y aprobación de la propuesta de actualización de la base de datos de URLs categorizadas de Filpacon, definida con anterioridad en este capítulo, se utilizó Criterio de un Panel de Especialistas, y para guiarse se utilizaron las técnicas que propone el Método Delphi. Una Delphi es una técnica que se puede utilizar para estructurar un proceso de comunicación en grupo, para ocuparse de un problema y los participantes en el mismo no tienen la necesidad de ir en persona a la reunión. Un panel de especialistas formula un grupo de hipótesis sobre el estado futuro del asunto en cuestión (Manage, 2009). Las estimaciones de los especialistas se realizan en sucesivas rondas, anónimas, al objeto de tratar de conseguir consenso, pero con la máxima autonomía por parte de los participantes.

En esta técnica se realiza una selección del grupo de especialistas que participará en el proceso de evaluación, teniendo en cuenta que ningún conoce la identidad y las respuestas individuales de los otros que componen el grupo. Se debe hacer una buena elección, para que el resultado sea el mejor.

PROCESO DELPHI

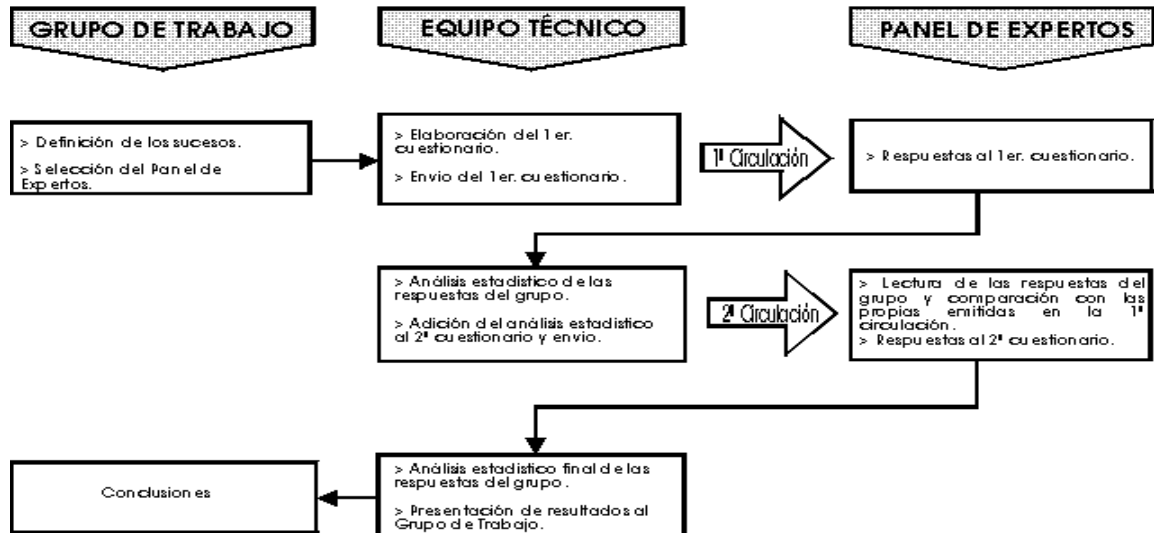


Figura 2.1 Método Delphi

El panel de especialistas se conformó con personas que poseen una vasta experiencia como administradores de red y en temas relacionados con este trabajo. En el presente capítulo se hará la descripción de los pasos utilizados en la selección del panel y los resultados obtenidos.

2.5.1 Definición de indicadores

Para realizar la validación o evaluación de la propuesta primeramente se definen los indicadores o atributos que serán evaluados por los especialistas. A partir de estos se confecciona una encuesta mediante la cual se expresarán los juicios o valoraciones en relación al sistema propuesto. Los atributos identificados son:

A1. Importancia o necesidad de la propuesta.

A2. Consistencia de la estructura de la propuesta.

A3. Nivel de completitud de la propuesta.

A4. Grado de utilidad de la propuesta.

A5. Nivel de vinculación de la propuesta con el desarrollo de software.

Una vez definidos los criterios a evaluar, se procederá a la selección del grupo de especialistas que desarrollarán la evaluación.

2.5.2 Elección de los expertos

Se entiende por especialista a una persona que cultiva o se dedica a un ramo de determinada arte o ciencia de la que tiene especiales conocimientos o habilidades, posee una gran experiencia en el tema y es capaz de ofrecer valoraciones conclusivas y hacer recomendaciones al respecto.

Para el procedimiento de la selección de los especialistas se llevaron a cabo los siguientes pasos:

2.5.2.1 Determinar la cantidad de especialistas

Para la determinación de la cantidad de especialistas, no existe una norma generalizada que determine un número óptimo. En este trabajo se decidió contar con un número de cinco para la confección del panel, teniendo en cuenta el nivel de complejidad y profundidad del contenido.

2.5.2.2 Confeccionar el listado de especialistas

La confección del listado se realizó atendiendo a la posibilidad reales de participación de los candidatos, todos son profesionales de la UCI que tienen experiencia en el tema tratado, poseen además, amplios conocimientos en temas relacionados con el proceso de actualización.

Las cualidades que caracterizan estos especialistas seleccionados, que se tuvieron en cuenta por parte de la autora de esta investigación para la confección del listado son:

Sinceridad: Al responder las preguntas que se hacen en la encuesta.

Seriedad: En la consistencia de las respuestas emitidas.

Creatividad: En las opiniones pedidas en la encuesta.

Responsabilidad: A la hora de medir el tiempo de entrega de la encuesta.

Capacidad de análisis: De acuerdo a los temas relacionados en la encuesta.

Estas cualidades han permitido que las opiniones obtenidas brinden confiabilidad y valides para el objetivo propuesto con esta encuesta.

2.5.2.3 Confirmar la participación de los candidatos

Una vez conformado el listado se invitó a cada uno a participar en la evaluación. Una vez recibida la respuesta de aceptación, se conformó el listado final de los especialistas que participarían en la investigación. De esta forma culmina el proceso de selección, logrando la participación de los escogidos.

A continuación se mencionan los especialistas participantes en este proceso de validación.

- ✓ Est. Saimel Sáez Estrada
- ✓ Ing. Dovier Antonio Ripoll Méndez
- ✓ Ing. Alain Guerrero Enamorado
- ✓ Ing. Luis Enrique Sánchez Arce
- ✓ Est. Raúl Amambay

2.6 Elaboración de los cuestionarios

La encuesta se llevó a cabo de una forma anónima y para su elaboración se tuvieron en cuenta los objetivos que debería cumplir el sistema de actualización de la base de datos de URLs categorizadas propuesto para su implantación en el proyecto Filpacon.

La encuesta establece una serie de preguntas de enfoque investigativo, que permitieron ver la posibilidad real de que pueda ser aplicada la propuesta, según las características actuales del proyecto Filpacon (Ver anexo1).

Este conjunto de preguntas aportaron criterios que permiten evaluar los indicadores anteriormente seleccionados. La relación que existe entre estos se refleja en el Gráfico 2.1 mostrado a continuación.

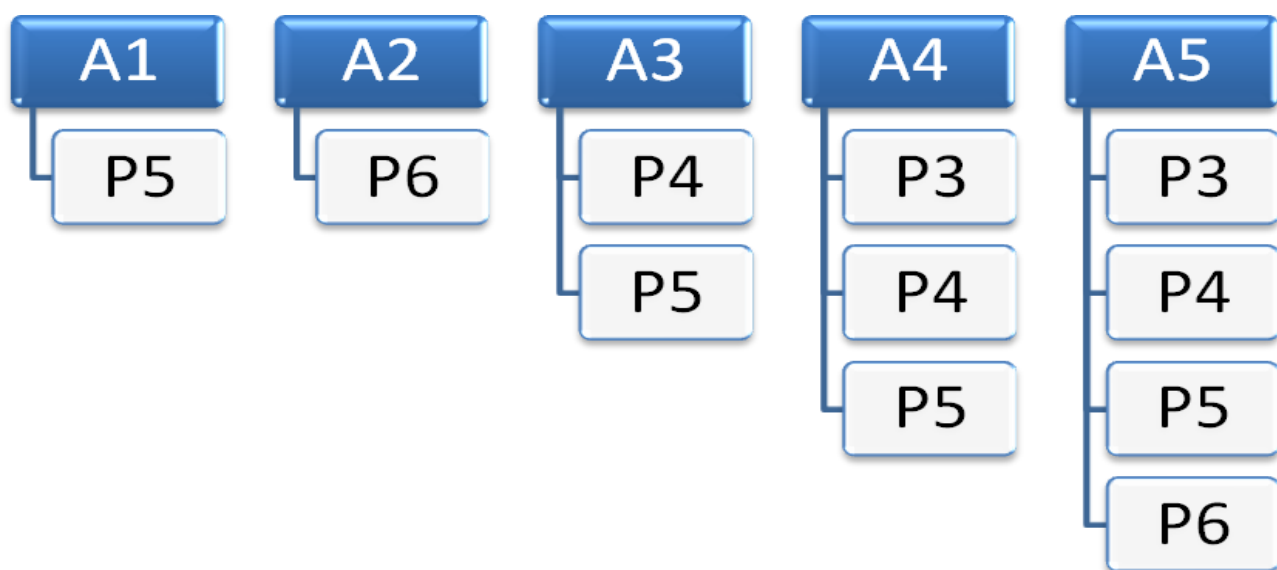


Gráfico 2.2 Interrelación entre los indicadores definidos y las preguntas del cuestionario.

2.7 Resultados de la evaluación

Para el análisis y procesamiento de los resultados, inicialmente se escogió un rango de evaluación [1; 5] para la primera pregunta de la encuesta, donde 5 representa la evaluación máxima. En la misma se comprueba el grado de conocimiento que tiene el experto sobre el software Filpacon. Los resultados de la encuesta aplicada aparecen recogidos en la tabla 2.1 que se muestra a continuación:

Pregunta	Saimel Sáez	Dovier Ripoll	Luis E. Sánchez	Alain Guerrero	Raúl Amambay
P1	4	5	5	3	4
P2	5	5	5	4	5
P3	5	3	3	5	4
P4	4	3	3	5	4
P5	5	5	5	5	5
P6	4	4	3	5	4

Tabla 2.1 Evaluación de los expertos a las preguntas de la encuesta

Una vez recogidos los resultados se promediaron para cada indicador las preguntas correspondientes de manera que se obtuviera un valor lo más representativo posible y cercano a la evaluación máxima. Los resultados de este paso se muestran en la tabla 2.5 expuesta a continuación.

Indicadores	Saimel Sáez	Dovier Ripoll	Luis E. Sánchez	Alain Guerrero	Raúl Amambay
A1	5,0	5,0	5,0	5,0	5,0
A2	4,0	4,0	3,0	5,0	4,0
A3	4,5	3,0	3,0	5,0	4,0
A4	4,7	3,6	3,6	5,0	4,3
A5	4,5	3,9	3,6	5,0	4,2

Tabla 2.2 Promedio de evaluación por indicadores

En el gráfico 2.2 que aparece a continuación se representan los datos de la tabla anterior para una mejor visualización de los resultados.

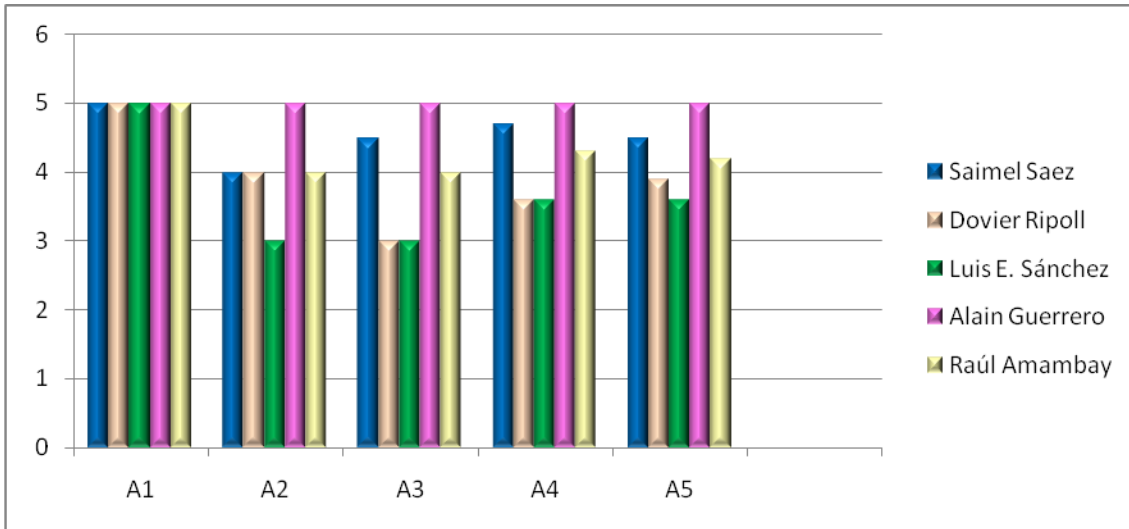


Gráfico 2.2 Evaluación de indicadores por expertos.

Para otorgarle una evaluación final a cada indicador se calculó el promedio a partir de la evaluación dada por los expertos. Los resultados se muestran en el Gráfico 2.3.

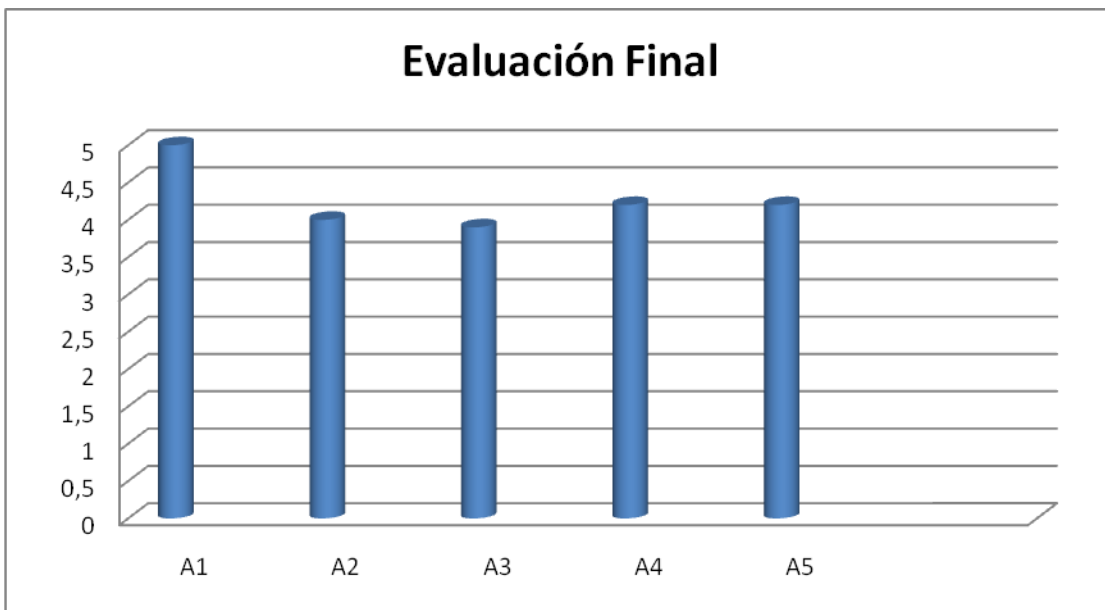


Gráfico 2.3: Evaluación final por indicadores.

De manera general se puede constatar que todos los indicadores tienen una evaluación superior a 4 puntos, lo que evidencia el grado de calidad y eficiencia de la propuesta de actualización de la base de datos definida.

Si bien las actividades definidas pueden cambiar o ampliarse según las características del proyecto, para los objetivos planteados la propuesta incluye las necesarias para garantizar la calidad de su aplicación.

2.8 Conclusiones

La propuesta de solución diseñada se basa en la realización de forma semi-automática de la actualización de la base de datos de URLs categorizadas, la toma de los ficheros de las listas de Isak y Toulouse y la utilización de un Gestor de Archivos de URLs Categorizadas (GARUC) para su almacenamiento.

El diseño propuesto permite el buen funcionamiento del software, a partir del nuevo sistema de actualización de su base de datos de URLs categorizadas de forma tal, que la renovación de su información se realice de manera sistemática, lo cual aporta una mayor efectividad en su servicio

El trabajo aporta una base teórica a partir de la investigación realizada y la documentación necesaria desarrollada para el uso del sistema de actualización que se implementará.

CONCLUSIONES GENERALES

- El análisis realizado permitió determinar los conceptos fundamentales referentes a los filtrados de contenidos, internet y URLs y las características e importancia los cuales sirvieron como puntos de reflexión para la toma de decisiones y de las diferentes metodologías y herramientas utilizadas internacionalmente para evaluar los sistemas de actualización de base de datos.
- La propuesta de solución diseñada se basa en la realización de forma semi-automática de la actualización de la base de datos de URLs categorizadas, la toma de los ficheros de las listas de Isak y Toulouse y la utilización de un Gestor de Archivos de URLs Categorizadas (GARUC) para su almacenamiento.
- El diseño propuesto permite el buen funcionamiento del software, a partir del nuevo sistema de actualización de su base de datos de URLs categorizadas de forma tal, que la renovación de su información se realice de manera sistemática, lo cual aporta una mayor efectividad en su servicio
- **El método de “Criterio de especialistas”** indicó consenso sobre las dimensiones e indicadores propuestos, permitió mejorar los indicadores y los criterios del contenido de la evaluación, obtener criterios y una evaluación superior a 4 puntos, lo que evidencia el grado de calidad y eficiencia de la propuesta de actualización de la base de datos definida y su factibilidad para ser aplicada.

RECOMENDACIONES

Se recomienda profundizar en otros aspectos para mejorar la propuesta tales como:

- Efectuar la implementación del sistema de actualización a partir de lo reglamentado en este trabajo.
- Profundizar en las actividades presentadas en el sistema propuesto.
- Profundizar en estudios del consumo del ancho de banda de Internet para tener claridad de los costos de mantenimiento de la base de datos de Filpacon.
- Darle continuidad al estudio del sistema propuesto a partir de la designación a otro grupo de investigadores.

REFERENCIAS BIBLIOGRÁFICAS

1. **Clarín, Dario. 1999.** Monografía.com. *Monografía.com*. [Online] 1999. <http://www.monografias.com/trabajos5>.
2. **Community, Python. 2009.** Python. *Python*. [Online] 2009. <http://www.python.org/>.
3. **Consultel. 2003.** Consultel. *Consultel*. [Online] 01 11, 2003. <http://www.consultel.es/Html/productos/Cacheflow/filtrado.htm>.
4. **Daishonin, Nichiren. 2006.** “Encuentro Alternativo. Programando tareas en cron”. Uruguay, septiembre 9, 2006.
5. **d'information, DSI Direction du Système. 2009.** Blacklists UT1. *Blacklists UT1*. [Online] mayo 2009. <http://cri.univ-tlse1.fr/blacklists/>.
6. **García, Leonardo Muro. 2007.** monografias.com. *monografias.com*. [Online] 11 21, 2007. <http://www.monografias.com/trabajos55/licencias-de-software/licencias-de-software2.shtml#tabla>.
7. **González, Antonio. 2009.** Filtros para Internet. *Filtros para Internet*. [Online] 05 11, 2009. <http://www.opusdealdia.org/200707151705/filtros-para-internet-optenet-canguro-net-y-otros.html>.
8. **Kit, iSAK Internet Secure Access. 2006.** iSAK Internet Secure Access Kit. *iSAK Internet Secure Access Kit*. [Online] 03 11, 2006. <http://isak.gplindustries.com/wiki/Isakurldb>.
9. **Lizama, Uriel. 2009.** Perl en Español. *Perl en Español*. [Online] 2009. <http://perlenespanol.com/articulos/archivo/000134.html>.
10. **Manage, 12. 2009.** Metodo Delphi. *Metodo Delphi*. [Online] 05 07, 2009. http://www.12manage.com/methods_helmer_delphi_method_es.html.
11. **Pérez, Dovie Ripoll y Yandri. 2008.** *Asignación automatizada de categorías temáticas*. 2008.

12. **Pérez, Leo González. 2005.** Clarin.com. *Clarin.com*. [Online] 2005.
<http://www.clarin.com/suplementos/informatica/2005/06/08/f-00211.htm>.
13. **Sánchez, Jose Ramon Hermosilla y Luis Enrique. 2008.** Interfaz de Administración Web para el Sistema de Filtrado Filpacon. . *Interfaz de Administración Web para el Sistema de Filtrado Filpacon*. . [Online] 2008.
14. **Sánchez, Yolanda. 2000.** iWorl. *iWorl*. [Online] 1 10, 2000.
<http://www.idg.es/iworld/articulo.asp?id=113038>.
15. **Security, Panda. 2009.** Panda Security for Domino Servers. *Panda Security for Domino Servers*. [Online] 2009. http://www.pandasecurity.com/NR/rdonlyres/9B8C155E-91F1-44A7-9B6D-C294A6E98145/0/01dwn_ps_PS4DOM.pdf.
16. **Valdés, Damián Pérez. 2007.** Maestros del web. [Online] 07 03, 2007.
<http://www.maestrosdelweb.com/editorial/%C2%BFque-es-javascript/>.
17. **Wikipedia. 2009.** Localizador Uniforme de Recursos. *Localizador Uniforme de Recursos*. [Online] 02 8, 2009. <http://es.wikipedia.org/wiki/Urls>
18. **Wikipedia 2009.** Wikipedia. *Wikipedia*. [Online] 02 09, 2009.
<http://es.wikipedia.org/wiki/Blacklist>.
19. **Wikipedia. 2009.** Eclipse. [En línea] 07 de 05 de 2009.
[http://es.wikipedia.org/wiki/Eclipse_\(software\)](http://es.wikipedia.org/wiki/Eclipse_(software)).

BIBLIOGRAFÍA

1. **cuidanía, Educación para la. 2008.** Educación para la ciudadanía. *Educación para la ciudadanía*. [En línea] 24 de 03 de 2008. <http://educacionparalaciudadania.wordpress.com/2008/03/24/filtrado-de-contenidos-en-internet/>.
2. **Foundation, Eclipse. 2009.** Eclipse. *Eclipse*. [En línea] 2009. <http://www.eclipse.org/>.
3. **Harada. 2002.** A tiny guide to Web Content Filtering. *A tiny guide to Web Content Filtering*. [En línea] 04 de 05 de 2002. <http://www.ingrid.org/~harada/filtering/>.
4. **Pérez, Dovier Ripoll y Yandri. 2008.** *Asignación automatizada de categorías temáticas*. 2008.
5. **platform, Eclipse - an open development. 2008.** Eclipse - an open development platform. *Eclipse - an open development platform*. [En línea] 17 de 03 de 2008. <http://www.eclipse.org/>.
6. **Sánchez, José Ramón Hermosilla y Luis Enrique. 2008.** Interfaz de Administración Web para el Sistema de Filtrado Filpacon. . *Interfaz de Administración Web para el Sistema de Filtrado Filpacon.* . [En línea] 2008.
7. **Sánchez, Yolanda. 2000.** iWorl. *iWorl*. [En línea] 10 de 01 de 2000. <http://www.idg.es/iworld/articulo.asp?id=113038>.
8. **Squidguard. 2009.** Squidguard. *Squidguard*. [En línea] 2009. http://cri.univ-tlse1.fr/documentations/cache/squidguard_en.html.
9. **TopTenReviews. 2009.** TopTenReviews. *TopTenReviews*. [En línea] 2009. <http://internet-filter-review.toptenreviews.com/>.

10. **Wikilearning. 2007.** Wikilearning. *Wikilearning*. [Online] 2007.
<http://www.wikilearning.com/buscador.php?txtPalClave=Perl&tipo>.
11. **Wikipedia. 2009.** Wikipedia. *Wikipedia*. [En línea] 02 de 09 de 2009.
<http://es.wikipedia.org/wiki/Blacklist>.
12. **Yavor. 2008.** Diversas licencias y comentarios sobre ellas. [En línea] 23 de 07 de 2008.
<http://www.gnu.org/licenses/license-list.es.html>

ANEXOS

Anexo 1: Encuesta para la validación de la propuesta.

Estimado Experto:

La presente encuesta forma parte de la validación de la Propuesta para actualizar la base de datos de URLs categorizadas de Filpacon v2.0 mediante el Método de Valoración de Especialistas o Expertos. Con este fin solicitamos su valiosa colaboración.

Muchas gracias de antemano.

Nombre y Apellidos: _____

Año de graduación: _____ Puesto de trabajo actual: _____

Calificación profesional: Ingeniero____ Licenciado____ Máster____ Doctor____

Responda las siguientes preguntas

1. Seleccione de escala el valor que corresponda al grado de conocimientos que usted posee sobre Filpacon, siendo 1 el de menor valor así como 10 el máximo.

1	2	3	4	5	6	7	8	9	10

2. Valore el grado de influencia de cada una de las fuentes que se presentan a continuación ha tenido en su conocimiento.

Fuentes de argumentación	Grado de influencia de cada fuente		
	Alta	Medio	Bajo
Conocimiento que posee sobre Filpacon			
Participación en el mismo			
Su experiencia obtenida en la actividad práctica.			

3. ¿Considera usted que la propuesta cuenta con un orden lógico y claro para la implementación del sistema de actualización de la base de datos de Filpacon?

Si_____ No_____ No sé_____

4. ¿Cree que las actividades descritas en la propuesta son suficientes para realizar un buen sistema de actualización? ¿Son las mismas precisas?

Si_____ No_____ No sé_____

Argumente:_____

5. ¿Considera usted importante diseñar un sistema de actualización para la base de datos de Filpacon? Otórguele un valor al grado de importancia que usted considere, siendo 1 el valor mínimo y 5 el valor máximo.

1	
2	
3	
4	
5	

6. ¿Resolvería la misma los problemas organizativos de Filpacon? Otórguele un valor al grado de importancia que usted considere, siendo 1 el valor mínimo y 5 el valor máximo.

1	
2	
3	
4	
5	

7. Haga un comentario o aporte sobre la propuesta que usted esta evaluando

Anexo 2: Funcionalidades del Módulo de Categoría de Filpacón v1.0.

The screenshot displays the Filpacón web application interface. At the top left, the logo 'FILPACÓN' is shown with the tagline 'Filtrado de paquetes por contenidos'. To the right, a welcome message reads 'Bienvenido. Hoy es Martes 27 de Noviembre de 2007'. Below this is a navigation menu with options: 'Usuario', 'Dominio', 'Categoría' (highlighted), 'URL', 'Reportes', and 'Salir'. On the left side, there is a sidebar titled 'Acciones' with three items: 'Adicionar', 'Modificar', and 'Eliminar'. The main content area is titled 'Funcionalidades del módulo Categoría' and contains the text 'Permite realizar las siguientes acciones:' followed by a bulleted list: 'Adicionar una nueva categoría.', 'Modificar el tipo de contenido que una categoría existente tiene asociado.', and 'Eliminar una categoría existente.'. At the bottom of the page, the footer text reads 'GNU General Public License | FILPACÓN 2007' and 'Gobierno Bolivariano de Venezuela'.

Anexo 3: Funcionalidades del Módulo de URL de Filpacon v1.0.

The screenshot displays the Filpacon web application interface. At the top left, the logo 'FILPACON' is shown with the tagline 'Filtrado de paquetes por contenidos'. To the right, a welcome message reads 'Bienvenido. Hoy es Martes 27 de Noviembre de 2007'. Below this is a navigation menu with tabs for 'Usuario', 'Dominio', 'Categoría', 'URL', 'Reportes', and 'Salir'. The 'URL' tab is currently selected. On the left side, there is a sidebar titled 'Acciones' with four menu items: 'Categorizar URL', 'Modificar categoría de una URL', 'Eliminar categoría de una URL', and 'Categorías asociadas a una URL'. The main content area is titled 'Funcionalidades del módulo URL' and contains the text 'Permite realizar las siguientes acciones:' followed by a bulleted list of four actions: 'Categorizar una URL.', 'Modificar la(s) categoría(s) asociada(s) a una URL.', 'Eliminar la(s) categoría(s) asociada(s) a una URL.', and 'Conocer la(s) categoría(s) asociada(s) a una URL..'. At the bottom of the page, the footer text reads 'GNU General Public License | FILPACON 2007' and 'Gobierno Bolivariano de Venezuela'.

Anexo 4: Código en Perl.

```
#!/usr/bin/perl

use strict;
use WWW::Curl::Easy;
use Archive::Tar;
use LWP::UserAgent;
#use Term::ReadKey;

my $curl = WWW::Curl::Easy->new;
my $tar = Archive::Tar->new;
=scip
print "User:";
my $user = <STDIN>;
print "Password:";
readMode('noecho');
my $password = ReadLine(0);
=cut

$curl->setopt(CURLOPT_PROXY,"http://10.3.10.248:3128");
$curl->setopt(CURLOPT_PROXYUSERNAME,'ienez');
$curl->setopt(CURLOPT_PROXYPASSWORD,'tater');
$curl->setopt(CURLOPT_NOPROGRESS, 0);
$curl->setopt(CURLOPT_WRITEDATA, 0);

=print
_lista de ISAK
=cut

open ISAK, ">/home/indira/Desktop/ISAK.tgz";
my $url = "http://download.isak.gplindustries.com/isakurldbtext.tar.gz";
$curl->setopt(CURLOPT_URL, $url);
$curl->setopt(CURLOPT_FILE, \*ISAK);
my $return_code = $curl->perform;
if ($return_code == 0){
    print("\nThe transfer of the list ISAK was completed.\n");
} else {
    print("An error happened: ".$curl->sterror($return_code)."\n");
}
=close (ISAK);
$curl->cleanup();

$tar->read("ISAK.tgz");
$tar->extract();

rename "isakurldbtext", "ISAK";

=print
_lista de TOULOUSE
=cut

open TOU, ">/home/indira/Desktop/TOULOUSE.tgz";
my $url = "http://cri.univ-tlse1.fr/blacklists/download/blacklists.tar.gz";
$curl->setopt(CURLOPT_URL, $url);
$curl->setopt(CURLOPT_FILE, \*TOU);
my $return_code = $curl->perform;
if ($return_code == 0){
    print("\nThe transfer of the list TOULOUSE was completed.\n");
} else {
    print("An error happened: ".$curl->sterror($return_code)."\n");
}
=close (TOU);
$curl->cleanup();

$tar->read("TOULOUSE.tgz");
$tar->extract();

rename "blacklists", "TOULOUSE";

print "Download complete";
```

Anexo 5: Archivo XML.

```
<order>
  <category>
    <name>Audio_video</name>
    <source>
      <path>/usr/local/GARUC/TOULOUSE/audio_video/urls</path>
      <list>TOULOUSE</list>
    </source>
    <source>
      <path>/usr/local/GARUC/TOULOUSE/audio_video/domains</path>
      <list>TOULOUSE</list>
    </source>
    <source>
      <path>/usr/local/GARUC/ISAK/Audioandvideo_urls.txt</path>
      <list>ISAK</list>
    </source>
    <source>
      <path>/usr/local/GARUC/ISAK/Audioandvideo_domains.txt</path>
      <list>ISAK</list>
    </source>
    <type>0</type>
  </category>

  <category>
    <name>Arts</name>
    <source>
      <path>/usr/local/GARUC/ISAK/Arts_urls.txt</path>
      <list>ISAK</list>
    </source>
    <source>
      <path>/usr/local/GARUC/ISAK/Arts_domains.txt</path>
      <list>ISAK</list>
    </source>
    <type>0</type>
  </category>

  <category>
    <path>/usr/local/GARUC/ISAK/Society_domains.txt</path>
    <list>ISAK</list>
  </source>
  <type>0</type>
</category>

  <category>
    <name>Anti_virus</name>
    <source>
      <path>/usr/local/GARUC/ISAK/Anti_virus_urls.txt</path>
      <list>ISAK</list>
    </source>
    <source>
      <path>/usr/local/GARUC/ISAK/Anti_virus_domains.txt</path>
      <list>ISAK</list>
    </source>
    <type>0</type>
  </category>

  <category>
    <name>Search_Engines</name>
    <source>
      <path>/usr/local/GARUC/ISAK/Searchengines_urls.txt</path>
      <list>ISAK</list>
    </source>
    <source>
      <path>/usr/local/GARUC/ISAK/Searchengines_domains.txt</path>
      <list>ISAK</list>
    </source>
    <type>0</type>
  </category>

  <category>
    <name>Child</name>
    <source>
      <path>/usr/local/GARUC/TOULOUSE/child/urls</path>
      <list>TOULOUSE</list>
    </source>
    <source>
      <path>/usr/local/GARUC/TOULOUSE/child/domains</path>
      <list>TOULOUSE</list>
    </source>
  </category>
```


GLOSARIO DE TÉRMINOS

Blacklist: son listas donde se registran las direcciones de IP que generan spam de forma voluntaria o involuntaria.

Filpacon: Filtrado de Paquetes por Contenidos

GUI: Interfaz Gráfica de Usuario

IAW: Interfaz de Administración Web

Software libre: Software que brinda libertad a los usuarios sobre su producto y por tanto, puede ser usado, copiado, estudiado, modificado y redistribuido libremente.

Software privativo: Software no libre, es un software o programa al que por medio de una licencia no libre se le ha privado de ciertas libertades de uso, modificación o distribución.

URL: indican el protocolo de red que se necesita para recuperar la información del recurso identificado, a través de la red.

XML: Extensible Markup Language. Metalenguaje de etiquetado en SGML. Diseñado específicamente para la WWW por W3C. Permite que un usuario diseñe sus propias etiquetas, con sus atributos y reglas de construcción de documentos.