UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

VICERRECTORÍA DE FORMACIÓN

DIRECCIÓN DE FORMACIÓN POSTGRADUADA

Concepción del Sistema de Filtrado para Internet (FILPACON v1.0)

Tesis para optar por el

título de Máster en Informática Aplicada

Autor: Ing. Alain Guerrero Enamorado

Tutor: MSc Héctor Rodríguez Figueredo

Ciudad de la Habana, mayo de 2009

DECLARACIÓN DE AUTORÍA

Declaro que soy el único autor de este trabajo y autorizo a la Universidad de las Ciencias Informáticas para que haga el uso que estime pertinente con él.

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

_____	_____
Ing. Alain Guerrero Enamorado	MSc. Héctor Rodríguez Figueredo

Dedicatoria

A mis padres y abuelos, por ser las mejores personas del mundo.
A mi novia, por brindarme todo su amor y su comprensión.
A mi familia y amigos, por su inmenso apoyo.

Agradecimientos

A mis padres, Magnolia y Lázaro: por dedicar gran parte de sus vidas para formarme, por todo su amor y cariño, por estar siempre conmigo, apoyándome, aconsejándome, guiándome, a ellos les estoy eternamente agradecido.

A mis abuelos, Ana y Uliser: por enseñarme tanto, por ser fuente de inspiración para alcanzar mis metas, por estar siempre preocupándose por mí, por su cariño y comprensión.

A mi novia, Daimerys, por todo su apoyo, paciencia y comprensión para conmigo.

A mis compañeros del proyecto, sin los cuales la implementación de este trabajo no hubiera sido posible.

Agradezco además de manera especial a: Luis, Dovier, José Ramón, Siovel, Carpio, Saimel, Yordan, Yanet Salazar y Maidely por todo el esfuerzo y dedicación que compartimos en esta tarea sin que esto signifique menos para el resto del equipo.

A mi tutor, Hector por todo el apoyo incondicional durante estos primeros años de mi experiencia profesional, he aprendido mucho de usted.

A todos aquellos que de una forma u otra han aportado algún grano de arena para la realización de este trabajo, recuerden que para mi los detalles no son solo detalles.

Muchas Gracias.

Resumen

Tras una revisión de algunos de los filtros más conocidos a nivel internacional, en este trabajo se presenta la concepción de un Sistema de Filtrado para Internet, su arquitectura y funcionalidades principales que le permiten adaptarse a diferentes entornos. El sistema propuesto constituye una primera versión sobre la que se sientan las bases para su futuro desarrollo. Con esta aplicación se espera contribuir a una Internet más segura para aquellos sectores de la sociedad más vulnerables ante los contenidos inadecuados. Actualmente se aplica en la Facultad 10 de la UCI, en el centro UCID y en los INFOCENTROS de la República Bolivariana de Venezuela, lugares en donde ha tenido un excelente rendimiento.

Abstract

After reviewing some of the most popular filters at the international level, this paper presents the design of a Filtering System for Internet, as well as its architecture and main functions that allow it to adequate to changeable environments. The proposed system is a first version on which the foundations for future development. With this application is expected to contribute to a safer Internet for those sectors of society most vulnerable to inappropriate content. Currently applied in Faculty 10 of the UCI, in the UCID center and INFOCENTROS of the Bolivarian Republic of Venezuela, places where he has had an excellent performance.

Índice

Índice de Tablas

Introducción

> “El problema de los sistemas de filtrado, es su imperfección: no hay software ni hardware que sustituya la responsabilidad paterna” (Casacuberta, 1998)

Mientras la discusión sobre el origen de Internet se mantiene abierta (Hauben, 2004) y el crecimiento de esta red de redes es de manera exponencial (Netcraft Ltd, 2009), mucha es la variedad de contenidos útiles que alberga, sin embargo, otros pueden resultar ofensivos e incluso ilegales de acuerdo a las legislaciones de muchos países. En su concepción actual, Internet por si misma, no es capaz de diferenciar los materiales de acuerdo a sus contenidos, su funcionamiento distribuido la convierte en una red en la cual muchas páginas son colocadas sin ningún tipo de revisión previa ni reserva. El problema fundamental de este punto radica en que de la misma forma que contenidos de incuestionable valor: educativo, científico, informativo, artístico, recreativo, y muchos más son colocados en esta red, también se introducen algunos de dudoso valor y en otros casos degradantes de normas morales. Cada día el número de sitios web que albergan contenidos pornográficos, pedofílicos, terroristas, xenofóbicos, racistas, extremistas, violentos y otros va en aumento, en el 2006 se estimaron 4.2 millones de sitios pornográficos representando el 12% de la Web, los cuales albergaban 420 millones de páginas con este tipo de contenido (Ropelato, 2006). La complejidad del problema empeora por la no jurisdicción de las leyes. Así por ejemplo, de nada sirve que en Alemania, Francia, Austria e Italia tengan leyes que prohíban la existencia de sitios web con contenidos xenofóbicos o nazis, si en los Estados Unidos esto es perfectamente legal, o que en nuestro país se abogue por la no comercialización del sexo si este negocio arroja más ingresos que la unión de las industrias líderes de la tecnología: Microsoft, Google, Amazon, eBay, Yahoo!, Apple, Netflix y EarthLinkmiles (Ropelato, 2006) y por tanto, tienen los recursos y la justificación de la libertad de expresión para vetar cualquier

ley que intente limitar este denigrante mercado. Otro de los ejemplos son los relativos a la pornografía infantil, en algunos países se prohíbe la distribución y la receptación, en otros solo la distribución, y en algunos ninguna de ambas cosas.

Esta preocupante situación internacional ha obligado a la sociedad a buscar posibles vías de solución a dicha problemática. Hasta el momento son las soluciones técnicas las únicas que han logrado cierto nivel de éxito a la hora de regular un tanto cuáles contenidos son accesibles por los usuarios. En varios países se han tomado iniciativas de ley que han terminado en un rotundo fracaso. Ejemplo de esto fue la controvertida Acta de Decencia de las Comunicaciones de EE.UU, promovida en 1995 por los senadores James Exon y Slade Gorton en el marco de la nueva ley de telecomunicaciones de ese país (Federal Communications Commission, 1996). En Alemania existe una ley sobre los servicios electrónicos que establece que los Proveedores de Servicios de Internet (ISP) pueden ser perseguidos por dar la oportunidad a sus usuarios de visitar recursos con contenidos ilegales (Gaceta Federal, 2004), esta ley no es del todo efectiva ya que muchas veces los proveedores de acceso a Internet no están en condiciones de saber cuáles contenidos están visitando sus usuarios. Otro ejemplo es la experiencia del Reino Unido, en este país, los ISP han adoptado mecanismos de regulación a través de códigos de conducta y la creación de un organismo dedicado a recibir denuncias sobre contenidos ilícitos llamado, Safety Net Foundation (Isabel Guerriero, y otros, 2002). En Australia se aprobó un plan de 162 millones de dólares para desarrollar sistemas de filtrado de contenidos destinado a las familias, escuelas e instituciones públicas (Redacción de Noticiasdot.com, 2007), este sistema se ha concebido para funcionar a nivel de ISP y es impulsado por el ministerio de comunicaciones de ese país (Editorial Board, 2008).

Muchos coinciden en que son los sistemas de filtrado una solución técnicamente viable, aunque no la mejor (Casacuberta, 1998), con ellos sería respetada la diferencia de criterios, valores o costumbres morales entre comunidades, países y culturas. Son los filtros el equivalente a las barreras físicas que existen en la sociedad para segregar contenidos no

adecuados para determinados grupos, así la acción de uno de estos filtros sería equivalente a la que hace el portero de un cine al no dejar entrar a un adolescente con menos de 16 años a ver una película con contenido sexual no apto para menores de edad, o más simple aun, al que hace nuestro jefe al exigirnos que en horario laboral no veamos un partido de fútbol. Algunos de estos sistemas aun se encuentran en un estadio primitivo de su desarrollo, muchos funcionan a nivel de palabras claves, otros a través de listas de inclusión y exclusión, los menos van incorporando tecnología inteligente para realizar el filtrado, por otro lado se encuentra que gran parte de estos sistemas son privativos en cuanto a facilidades de modificaciones en su código fuente y en cuanto a su precio imposibilitando su adaptación y utilización a gran escala en Cuba.

Es válido señalar que se tienen referencias de trabajos relativos al montaje y configuración de Sistemas de Filtrado para Internet en el Instituto Superior Politécnico José Antonio Echeverría (CUJAE), aunque sin llegar al desarrollo de un producto nacional que permita su generalización independientemente de las características de las empresas, centros educacionales u organizaciones en general que brinden este importante servicio de Internet.

De esta manera se identifica como **problema**, la inexistencia de un Sistema de Filtrado que permita regular el acceso a Internet en Cuba de manera flexible y adaptable a las condiciones del lugar donde se despliegue.

Estableciéndose la siguiente **hipótesis**:

La concepción de un Sistema de Filtrado para su posterior implementación permitirá regular de manera flexible y adaptable a las condiciones del lugar donde se despliegue el acceso a Internet y brindará información útil sobre el uso de este servicio lo cual permitirá el proceso de toma de decisiones.

Partiendo del problema se define como **objeto de estudio** la regulación de los contenidos de Internet y el **campo de acción** está delimitado a los Sistemas de Filtrado para regular el acceso a Internet.

De manera general el **objetivo** es:

Concebir una aplicación informática, Sistema de Filtrado para Internet, basada en Software Libre que permita de manera flexible y adaptable a las condiciones del lugar de despliegue regular el acceso a Internet.

Para complementar este objetivo general se puntualizan los siguientes objetivos específicos:

- Concebir una arquitectura del Sistema de Filtrado para Internet que garantice una implementación lo suficientemente flexible y modular que le permita adaptarse a las particularidades de cada lugar donde sea instalado y que garantice la incorporación paulatina de las funcionalidades que a nivel internacional tienen estos sistemas.
- Seleccionar tecnologías que pertenezcan a la categoría de Software Libre tanto para el desarrollo del sistema como para su funcionamiento.
- Concebir un instalador para el sistema, de manera que su instalación y puesta a punto no sea costosa y además pueda distribuirse a manera de producto en un dispositivo de almacenamiento extraíble.
- Concebir una interfaz de administración que permita definir políticas de filtrado flexibles y adaptables a las características del lugar donde se instale.
- Concebir la posibilidad de generar reportes estadísticos que ayuden en el proceso de toma de decisiones.

Para lograr los objetivos propuestos se formulan las siguientes preguntas científicas:

- ¿Cuáles componentes conformarán la arquitectura para que sea suficientemente flexible y modular?
- ¿Cómo garantizar desde el punto de vista de la arquitectura, la incorporación paulatina de funcionalidades que a nivel internacional tienen los Sistemas de Filtrado para Internet más comunes?

- ¿Cuáles tecnologías libres son necesarias para la concepción y desarrollo del sistema?
- ¿Cuál instalador garantiza una correcta detección de una elevada variedad de hardware y puede ser adaptado al sistema?
- ¿Qué elementos y funcionalidades deben tenerse presentes para que la interfaz de administración permita definir políticas de filtrado adaptables a las características del lugar donde se instale?
- ¿Cuáles reportes estadísticos ayudan a la toma de decisiones que promueven el buen uso de Internet?

Para llevar a cabo este proyecto han de combinarse métodos de investigación, teóricos y empíricos.

Métodos Teóricos:

Analítico-Sintético: permite hacer un análisis de los Sistemas de Filtrado y las funcionalidades que estos brindan, a manera que posibilite identificar aquellas que puedan ser aplicadas al desarrollo de este nuevo proyecto.

Inducción-Deducción: para analizar las características del comportamiento de los Sistemas de Filtrado y así poder deducir conclusiones sobre casos particulares que pueden ser verificados en la práctica.

Histórico-Lógico: para estudiar la evolución y desarrollo de los Sistemas de Filtrado y comprender lógicamente cuales son las tendencias actuales.

Métodos Empíricos:

Observación: posibilita el análisis de los resultados para la obtención del conocimiento acerca del comportamiento de los Sistemas de Filtrado mediante la percepción directa de los objetos y fenómenos.

Medición: para obtener información numérica acerca del comportamiento del Sistema de Filtrado y comparar sus valores.

Experimentación: por su importancia decisiva en las pruebas para demostrar el funcionamiento del sistema.

Capítulo 1. Los Sistemas de Filtrado. Conceptos básicos y estado actual.

1.1. Introducción

Como se ha planteado, Internet por su naturaleza distribuida y descentralizada se convierte en una red en la que no son aplicables legislaciones, estas en su gran mayoría dependen de las fronteras de los países, fronteras que no existen para esta red. Esto la convierte en una red vulnerable para prácticas de todo tipo. Las iniciativas de los diferentes estados han resultado en la mayoría de los casos en un rotundo fracaso, o en proyectos casi inaplicables. Defender la no regulación de Internet presupone una contradicción, pues aceptarlo significaría que el único derecho válido es el de aquellos que consideran que debe existir en Internet una libertad sin límites. En muchos casos son los mismos argumentos que se esgrimen para defender la pornografía y la violencia en los medios de comunicación en general. Los intereses son variados, ejemplo de ello es que los individualistas radicales y los empresarios, constituyen obviamente dos grupos muy diferentes. Sin embargo, existe convergencia de intereses entre quienes buscan que Internet se convierta en un lugar apto para cualquier tipo de expresión, sin importar si es vil y destructiva, y quienes quieren que sea un vehículo de actividad sin trabas según un modelo neoliberal que, considera las ganancias y las leyes del mercado como parámetros absolutos, en detrimento de la dignidad, del respeto a las personas y a las diferentes culturas.

Mientras los países no se ponen de acuerdo, son los Sistemas de Filtrado llamados a brindar una solución viable, que si bien no es la solución definitiva, si contribuye de manera importante a mitigar los riesgos de la problemática planteada.

1.2. ¿Qué son los sistemas de filtrado?

Podemos definir a un Sistema de Filtrado en el contexto de Internet como aquella aplicación informática concebida para lograr un bloqueo efectivo sobre aquellos contenidos que así lo requieran. De esta manera los Sistemas de Filtrado determinan que contenido estará disponible para un usuario partiendo de reglas que definieron determinadas autoridades, ejemplo de ello es una escuela donde los usuarios son niños los cuales constituyen un sector vulnerable a la mayoría de los contenidos nocivos que abundan en la red, otro caso puede verse en una empresa donde los directivos de la misma necesitan que Internet solo sea usado con objetivos laborales.

1.3. Tipos de Sistemas de Filtrado

Los Sistemas de Filtrado pueden clasificarse atendiendo al mecanismo que utilizan para bloquear el acceso en:

- **filtros por palabras claves** (Villate, 1997), se bloquean los recursos que contengan algunas palabras preestablecidas. El mayor inconveniente radica en su incapacidad para discernir el contexto en el cual estas palabras son utilizadas. Así por ejemplo si se define bloquear “sexo”, serían bloqueados todos los sitios que contengan esta palabra, pudiendo ser incluidos los sitios que contengan contenidos relativos a la educación sexual.
- **filtros basados en listas** (Villate, 1997), se bloquean los recursos que pertenezcan a la lista de exclusión (lista negra), pueden además tener listas de inclusión (lista blanca) que nunca serán denegadas. Su mayor inconveniente radica en el mecanismo de actualización de las listas, ya que inmediatamente que sean construidas quedan de hecho desactualizadas producto del dinamismo de Internet. Por otro lado, construir estas listas puede representar un esfuerzo enorme si se lleva a cabo por personas.
- **filtros por el contenido contextual** (Villate, 1997), se bloquean los contenidos por el significado contextual. En general, utilizan técnicas de clasificación automática de

contenidos (ejemplo es la Inteligencia Artificial). En realidad este tipo de filtros es un paso evolutivo superior de los basados en listas, la diferencia fundamental radica en el método que se utiliza para actualizar las listas.

Atendiendo al nivel jerárquico donde son instalados, los filtros pueden clasificarse de la siguiente forma:

- **filtros para *backbone* de Internet**, esquemas de filtrado que se colocan a nivel de los países (OpenNet, 2007).
- **filtros para proveedores de acceso** (Villate, 1997), se pueden instalar en un proveedor de acceso a Internet. De esta forma las políticas se aplicarían a todos los usuarios del proveedor, los cuales pueden ser instituciones o usuarios finales.
- **filtros para instituciones**, ideal para empresas y centros educacionales donde es deseable que los usuarios utilicen Internet solo con propósitos acordes a la labor que realizan, este nivel de filtrado puede verse en algunos casos solapado al anterior, pues en ocasiones los clientes de los proveedores de acceso son usuarios finales y no instituciones (OpenNet, 2007).
- **filtros para clientes finales** (Villate, 1997), aplicables en ambientes domésticos o estaciones individuales, es uno de los tipos de filtrado más usados en la actualidad, aunque tienen el inconveniente de que al ser los que más cerca están del usuario final pueden ser desactivados si se dispone del conocimiento y pericia.

1.4. Características de los Sistemas de Filtrado

Durante mucho tiempo los fabricantes de estas aplicaciones han asumido el derecho a decidir, qué es bloqueado y que no. En general las listas de bloqueo de los Sistemas de Filtrado Comerciales no son públicas y constituyen parte de la propiedad del software considerándose secreto comercial.

En un inicio, si a los usuarios no les satisfacían las listas que el fabricante preestablecía, no le quedaba más remedio que optar por otro proveedor de un software similar, muchos fueron

los ejemplos de arbitrariedades por esos tiempos, ya que llegaron a ser bloqueados sitios web por el mero hecho de criticar de alguna forma a los Sistemas de Filtrado o por motivos políticos como el caso del sitio http://pimientanegra.cjb.net/, (Cibercensura, 2003). En nuestros días la mayoría de los sistemas de este tipo brindan las facilidades de modificar las listas de inclusión y exclusión, por lo cual han evolucionado a sistemas mucho más flexibles, en otros casos algunos de los sistemas publican de alguna forma los mecanismos que utilizan para crear sus listas. No obstante cada fabricante tiene su propia iniciativa sobre que dar a los usuarios y que no, el caso es que la gran mayoría de los fabricantes propietarios no permiten el acceso a su base de datos de sitios clasificados, pues lo consideran como un secreto comercial, otros describen de manera detallada cómo fue creada la misma pero tampoco nos dan acceso a la lista completa ni los algoritmos que utilizaron; los menos, y pertenecientes en su mayoría a la categoría de software libre, publican las bases de datos de contenidos[1] aunque no se puede dejar de decir que son bastante pequeñas comparadas con el tamaño de Internet y tienden a tener muchos errores; existen por otro lado fabricantes que tienen modelos de negocio más complicados aun pues dan libre algunas partes y cobran las actualizaciones de bases de datos. A pesar de la variedad de formas en la que los fabricantes establecen los derechos sobre la base de datos de sitios clasificados, en lo que si coinciden la mayoría es en permitir a los usuarios adicionar sitios que ellos consideren o eliminar otros que ya estaban. Menos frecuentes son los filtros que permiten tener niveles dentro de cada categoría de filtrado, teniendo así un filtrado con listas grises. De esta forma, la decisión sobre qué bloquear puede realizarse a partir de contenidos que pertenezcan a varios niveles de clasificaciones.

Otra de las características importantes de los Sistemas de Filtrado es su especialización hacia un protocolo específico, es decir los filtros que se dedican al correo electrónico se

[1] http://urlblacklist.com/

especializan en el protocolo SMTP (Simple Mail Transfer Protocol), los que se dedican a los contenidos web se especializan en el Protocolo de Transferencia de Hipertexto (HTTP).

La evolución de los Sistemas de Filtrado no termina aun, la tendencia actual es la utilización de algoritmos inteligentes para clasificar los contenidos, en este caso no todos los fabricante pero si los líderes mundiales de esta tecnología lo han ido incorporando; el hecho es que la dinámica de Internet lo impone. Son estos filtros, capaces de "entender" el contenido contextual, el futuro del filtrado. El reconocimiento de patrones, la recuperación de información, la minería de datos, el reconocimiento de idioma y los analizadores semánticos son sólo algunos de los campos vinculados al desarrollo de este tipo de filtros; campos que en su mayoría tienen aun elementos importantes en la frontera del conocimiento. Por último decir que a pesar de que algunos sistemas actuales se auto atribuyen que clasifican sus contenidos utilizando análisis contextual, la realidad es que varios de ellos siguen utilizando búsquedas de palabras, falta mucho por hacer y madurar aun en este sentido.

1.5. Normas relacionadas con el filtrado

1.5.1. Plataforma para la Selección del Contenido en Internet (PICS)

Fue desarrollada por el *World Wide Web Consortium* (W3C) y originalmente creada para permitirle a los padres, profesores y bibliotecólogos evaluar materiales de contenido cuestionable que no debían estar al alcance de los menores en Internet (Resnick, 1997), en ese artículo Resnick explica que antes de PICS no existía un formato estándar para las etiquetas, así que las compañías que deseaban proporcionar Filtrado de Contenidos tenían que proporcionar tanto el software como las etiquetas. La idea inicial buscaba lograr que las personas pudieran compartir descripciones de sus trabajos digitales a través de la red, de una forma que pudiera ser interpretada por las computadoras, así se podría lograr que padres y maestros tuvieran los medios para filtrar contenidos que consideraran inadecuados. La filosofía de PICS es diferente a la de las leyes, estas últimas intentan censurar lo que se distribuye, mientras que la plataforma PICS, al igual que los Sistemas de Filtrado, crea el

mecanismo para permitir o no el acceso a los contenidos que llegan al usuario final. La diferencia fundamental de los sistemas PICS con los Sistemas de Filtrado se encuentra en que con estos últimos quedamos sujetos a los criterios que utilizó el fabricante del software para crear sus categorías y listas de sitios. En cambio, con PICS las categorías se encuentran estandarizadas y las decisiones sobre qué se bloquea las toman los usuarios o los proveedores de acceso. Es además un estándar abierto, y la información sobre las clasificaciones de los sitios puede ser brindada por el fabricante del software PICS, o por terceros. La distinción más importante que introdujo PICS fue la separación entre el etiquetado y el filtrado (Resnick, 1999), quedando así dos elementos importantes en esta tecnología:

- el sistema estandarizado de categorías y,
- el software que lo interpreta para filtrar los contenidos.

El sistema estandarizado de categorías no es más que un conjunto de clasificaciones en las que para cada una se asocia un determinado nivel según la fuerza con la que esté presente un determinado contenido. Los sistemas de clasificación más populares son RSACi y SafeSurf. En la Tabla 1 se muestra el sistema RSACi utilizado por MS Internet Explorer para filtrar contenidos usando este estándar de etiquetado de los contenidos para las páginas de la Web.

	DESNUDEZ	LENGUAJE	SEXO	VIOLENCIA
NIVEL 0	Ninguna	Jerga Inofensiva	Solo romance	Sin violencia
NIVEL 1	Atuendos Reveladores	Términos Suaves	Besos apasionados	Peleas con heridas
NIVEL 2	Desnudez Parcial	Términos sin Referencias Sexuales	Roce sexual con ropa puesta	Muerte con daño
NIVEL 3	Desnudez Frontal	Lenguaje vulgar y uso de epítetos	Roce sexual no explícito	Muerte con sangre y horror

	DESNUDEZ	LENGUAJE	SEXO	VIOLENCIA
NIVEL 4	Desnudez Provocativa	Lenguaje chabacano, referencias sexuales explicitas	Actividad sexual explícita	Violencia perversa y gratuita

Tabla 1 Sistema RSACi que implementa el navegador MS Internet Explorer.

De manera general, PICS pudiera ser la solución a muchos de los problemas de los contenidos en Internet. Sin embargo, los usuarios están actualmente limitados a elegir entre un pequeño número de sistemas de clasificación, cada uno de los cuales tiene sus propias inclinaciones y puntos de vista. Los usuarios que están en desacuerdo con los sistemas de clasificación populares no pueden utilizar PICS de forma que responda a sus necesidades y puntos de vista. A estos usuarios sólo les queda desarrollar su propio estándar y clasificar la mayor cantidad de sitios con este.

Otro problema aún mayor es que muchos de los proveedores de contenidos se niegan a clasificar sus sitios, alegando que no están de acuerdo con ninguno de los sistemas PICS existentes. Por otro lado, este tipo de situaciones es la que provoca que la cantidad real de páginas clasificadas con esta tecnología sea muy pequeña. Los sistemas más utilizados se describen a continuación:

- La historia de RSACi (*Recreational Software Advisory Council Internet*) data de 1994 aunque desde 1999 RSACi pasó a ser ICRA (*Internet Content Rating Association*) el cual es parte del *Family Online Safety Institute* (FOSI, 2007), siendo el sistema de clasificación más utilizado hoy en día al estar presente en todos los ordenadores con el Sistema Operativo Windows ya que lo trae implementado el navegador por defecto Internet Explorer. Incluye en su versión original las categorías violencia, desnudos, sexo y lenguaje, con cinco clasificaciones dentro de cada categoría. Las etiquetas no son colocadas por ICRA, esta responsabilidad la asumen los creadores de las páginas Web, además estas etiquetas están representadas por un código numérico. En la

actualidad se han incluido nuevas categorías y se realiza la descripción del vocabulario utilizando RDF, aunque desde el 2006 se mantiene una versión con PICS junto a las etiquetas RDF (FOSI, 2007).

- SafeSurf: Las categorías de este sistema incluyen "Rango de Edades", "Blasfemias", "Temas Heterosexuales", "Temas Homosexuales", "Desnudos", "Violencia", "Sexo, Violencia y Blasfemias", "Intolerancia", "Glorificación del uso de drogas", "Otros temas para adultos" y "Juego", con nueve subdivisiones dentro de cada categoría. SafeSurf al igual que RSACi dependen de auto clasificación de los sitios Internet por parte de los propios creadores de páginas web, aunque se brinda una herramienta Web que permite fácilmente generar las etiquetas de clasificación (SafeSurf, 2007).

Como la mayoría de las páginas no están aún clasificadas, los programas que dan soporte a PICS incluyen una opción de configuración para que los usuarios puedan bloquear todas las páginas que no tengan la clasificación. Más información sobre este estándar puede encontrarse en la página oficial (W3C, 2005).

1.5.2. Marco de descripción de recursos (RDF)

El RDF (*Resource Description Framework*) es un lenguaje de representación de la información acerca de los recursos en la Web (Manola, et al., 2004), fue desarrollado también por el W3C y es continuidad del trabajo en PICS pues aborda un tratamiento mucho más general de los metadatos al incorporar información estructural y relacional entre los recursos Web (Swick, 1997). Está diseñado para la representación de metadatos sobre recursos Web como el título, autor y fecha de modificación de una página Web, o los derechos de autor y licencias de información acerca de un documento Web. Sin embargo, RDF puede utilizarse también para representar información acerca de recursos, incluso cuando no se puedan recuperar directamente en la Web; tal es el caso de la inclusión de información relativa a recursos disponibles en tiendas en línea, por ejemplo, información sobre especificaciones, precios y disponibilidad de los objetos.

RDF es un lenguaje declarativo y proporciona una forma estándar para el uso de metadatos en XML para representar en forma de declaraciones sobre las propiedades y las relaciones de los objetos en la Web. Estos objetos, conocidos como los recursos, pueden ser casi cualquier cosa, siempre que tengan una dirección Web. Esto significa que usted puede asociar metadatos con una página Web, un gráfico, un archivo de audio, una imagen en movimiento, y así sucesivamente. Su basamento esta dado por la conversión de las descripciones de los recursos en expresiones ternarias en la forma sujeto-predicado-objeto. El sujeto es el recurso en si mismo, el predicado es la propiedad o relación que se desea establecer acerca del recurso, y el objeto es o bien un valor de la propiedad o bien otro recurso con el que se quiere establecer una relación.

RDF proporciona un marco (*framework*) en el que comunidades independientes pueden desarrollar vocabularios que se adapten a sus propias necesidades y puedan además compartirlos con otras comunidades, ejemplo fue el desarrollo del *Dublin Core* (DCMI, 2009). Con el fin de compartir los vocabularios, el significado de los términos debe ser expuesto en detalles. Las descripciones de estos conjuntos de vocabularios son llamados Esquemas RDF. Un esquema define el significado, características, y las relaciones de un conjunto de propiedades, y esto puede incluir restricciones sobre los posibles valores y la herencia de propiedades de otros esquemas. El lenguaje RDF le permite a cada documento que contenga metadatos aclarar el vocabulario que utiliza asignando a cada vocabulario una dirección Web (Swick, 1997).

Producto de su generalidad, el estándar RDF, tiene un amplio conjunto de aplicaciones prácticas:

- **Tesauros y sistemas de clasificación de bibliotecas.** Son casos bien conocidos de sistemas jerárquicos de representación de las taxonomías en términos de las relaciones entre conceptos. Las especificaciones del Esquema RDF tiene las características exactas para permitir crear modelos RDF que representen la estructura

lógica de los tesauros y otros sistemas de clasificaciones de la bibliotecología (Swick, 1997).

- **Mapas de navegación.** Un mapa de navegación puede verse internamente como la descripción de un sitio Web. Las especificaciones del Esquema RDF proveen el mecanismo para definir el vocabulario en este tipo de aplicaciones. Con RDF se puede describir como un elemento está relacionado con otro, como una página es descendiente de otra y así sucesivamente (Swick, 1997).
- **Descripción de los contenidos de las páginas Web.** Esta es una de las funciones básicas de la *Dublin Core Metadata Initiative* (DCMI, 2009), El *Dublin Core* es un conjunto de 15 propiedades asociadas con información bibliográfica. Estas pueden utilizarse para describir elementos en la Web de manera que los buscadores puedan trabajar más eficientemente. Los talleres asociados a esta iniciativa han sido la mayor influencia en el desarrollo del RDF (Swick, 1997).
- **Para describir la estructura formal de las descripciones de costumbres privadas (*privacy practice descriptions*).** ¿Cómo un sitio maneja la información personal? ¿Divulga alguna de esta información a otros? ¿Qué obtiene a cambio el usuario? El Proyecto Plataforma para Preferencias de Privacidad del W3C, más conocido por P3P (*Platform for Privacy Preferences Project*), trabaja en una plataforma que le informe al usuario acerca de las prácticas de los sitios Web. De esta manera usuarios o aplicaciones que operen en su nombre podrán negociar con los sitios Web entre diferentes políticas de privacidad, llegando a un acuerdo sobre cuanta información personal podrá ser brindada posteriormente a terceros. RDF se puede utilizar para describir la estructura formal de las descripciones de costumbres privadas (Wenning, 2008).
- **Descripciones de las capacidades de los dispositivos móviles.** RDF provee una vía para describir las capacidades y preferencias asociadas a los usuarios, y el hardware y el software que ellos usan para acceder a la Web. Esto permitirá que el

contenido de la Web se adapte a las necesidades específicas de los usuarios (Kitagawa, et al., 2001).

- **Expresando metadatos de metadatos.** Supongamos que se ha construido un sistema de clasificación de Restaurantes sobre RDF. Puede entonces utilizarse RDF para describir metadatos sobre una determinada clasificación, ejemplo, la fecha en que se dio la clasificación, por cual organización y así sucesivamente (Swick, 1997).
- **Firmas Digitales.** En la misma medida que la Web madura y un mayor número de actividades se realizan sobre ella, las Firmas Digitales se hacen cada vez más necesarias (Reagle, 2002). RDF puede utilizarse para expresar información concerniente a que se está firmando, cual es la importancia de esta firma, el rango de fechas durante la cual la firma es válida y así sucesivamente.
- **Sistemas de Clasificación.** Estos ofrecen una forma de etiquetar los recursos para que las personas o equipos de cómputo puedan filtrar la información. RDF le permite a los programadores diseñar sistemas de clasificación para cualquier número de dominios (Swick, 1997). Esta es la aplicación que más se relaciona con este trabajo, por lo cual se volverá a retomar este tema en el capítulo 2.

1.6. Las soluciones actuales

Tras un breve bosquejo por los tipos, características y estándares relativos a los Sistemas de Filtrado, analizaremos a continuación aquellos sistemas que por sus características han logrado un espacio importante en las direcciones funcional y comercial, siendo las bases para el posterior diseño de la solución.

1.6.1. Net Nanny Parental Controls.

Sistema de filtrado número uno en los análisis que se realizan en (TopTenREVIEWS, 2009), por tanto le han adjudicado el título: "*TopTenREVIEWS Gold Award"*, el cual además lo ha obtenido en cinco ocasiones. Ganador además en cuatro ocasiones del *PC Magazine's Editors' Choice Award* (Rubenking, 2008). NetNanny es un producto de ContentWatch, Inc.

(ContentWatch, 2009) compañía estadounidense que tiene toda una suite de productos relativos a garantizar protección contra contenidos indeseables, en los principales niveles de aplicación, fundada en marzo del 2000. Algunos de los productos que ofrece son:

- NetNanny: Elimina contenidos indeseados mientras monitorea la actividad en Internet, ver Anexo 1.
- ContentCleanup: Analiza, identifica y elimina los archivos de contenido cuestionable.
- ContentProtect ™ Professional Internet Filtering For Business: Elimina contenidos indeseados mientras monitorea la actividad en Internet para empresas, organizaciones gubernamentales, instituciones educacionales y bibliotecas. Aunque funciona a nivel de cada cliente. Por el momento no brinda soporte para usuarios GNU/Linux ni MAC (ContentWatch, 2008).
- ContentProtect™ Professional Suite For Business: Sistema para administrar el acceso a Internet, la privacidad de los datos del personal y las aplicaciones que corren las estaciones del personal en empresas, organizaciones gubernamentales, instituciones educacionales y bibliotecas. Aunque funciona a nivel de cada cliente. Por el momento no brinda soporte para usuarios GNU/Linux ni MAC (ContentWatch, 2008).

El producto insignia de ContentWatch es NetNanny en este se basan el resto de las soluciones, con algunas adiciones de otras aplicaciones, sus principales funcionalidades se muestran a continuación (TopTenREVIEWS, 2009):

- Permite filtrar los sitios Web basándose en algún tipo de análisis de los objetos del sitio. *(toptenreviews: Object Analysis)*
- Permite filtrar los sitios Web basándose en su dirección. *(toptenreviews: URL Based)*
- Permite filtrar los sitios Web basándose en determinadas palabras que el sitio puede contener. *(toptenreviews: Keyword Based)*

- Permite filtrar los sitios Web basándose en la posibilidad de que el contenido del sitio cambie aun cuando mantenga la misma dirección. *(toptenreviews: Dynamic Categorization)*
- Permite filtrar aplicaciones P2P tales como Gnutella, Kazaa, Morpheus, Limewire, Bearshare y otras. *(toptenreviews: Peer-to-Peer (P2P) Blocking)*
- Permite editar las listas de direcciones y palabras a filtrar. *(toptenreviews: Editable Filter Lists)*
- Permite la creación de diferentes perfiles de navegación en dependencia de quien está usando Internet. *(toptenreviews: Individual User Profiles)*
- Permite establecer políticas de navegación horarias. *(toptenreviews: Daily Time Limits)*
- Permite la generación de reportes gráficos de históricos. *(toptenreviews: Graphical Reporting)*
- Permite la generación de reportes resumen de históricos. *(toptenreviews: Summary History Reporting)*
- Permite la generación de reportes detallados de históricos. *(toptenreviews: Detailed History Reporting)*
- Permite la generación de notificaciones y alertas por correo electrónico. *(toptenreviews: Notification Alerts by E-mail)*
- Cantidad de categorías diferentes que identifica el Sistema de Filtrado: 31. *(toptenreviews: Filter Categories)*
- El producto provee documentación y ayudas. *(toptenreviews: Product Documentation)*
- El producto está limitado para algún navegador. *(toptenreviews: Supported Browsers)*

El precio de la última versión del software NetNanny 6.0, para usuarios finales es de $39.99 por una licencia anual, a continuación cada licencia hasta un límite de cinco tiene un costo de $ 19.99, en la versión empresarial el costo es de $ 39,99 más $ 35,99 por cada licencia adiciona (por cada estación cliente). En el precio están incluidos los siguientes elementos:

- La adquisición del software y la inicialización del servicio.

- Servicio de atención al cliente.
- Actualizaciones automáticas de las listas y definiciones de la aplicación.

La licencia de uso es completamente privativa (ContentWatch, 2008) pues nos impide:

- Copiar, distribuir, rentar, arrendar o sub-licenciar todo o parte del producto.
- Modificar o preparar trabajos derivados del producto.
- Compartir el producto por cualquier medio electrónico.
- Realizarle ingeniería inversa, descompilar o desensamblar el producto.

Cada licencia anual de la versión ContentProtect ™ Professional Suite For Business, tiene un costo de **$59,99** por la primera y **$53,99** por cada licencia adicional. Para una red de 1000 usuarios por un año sería de unos: **$53.996,00 dólares norteamericanos** (ContentWatch, 2008)**.**

Cada licencia anual de la versión ContentProtect ™ Professional Internet Filtering For Business, tiene un costo de **$39,99** por la primera y **$35,99** por cada licencia adicional. Para una red de 1000 usuarios por un año sería de unos: **$35.994,00 dólares norteamericanos** (ContentWatch, 2008)**.**

1.6.2. SafeEyes

Sistema de filtrado número dos en los análisis que se realizan en (TopTenREVIEWS, 2009), por tanto le han adjudicado el título: "*TopTenREVIEWS Silver Award".* Ganador en dos ocasiones del *PC Magazine's Editors' Choice Award* (Rubenking, 2006). SafeEyes es un producto de InternetSafety.com™, Inc. (InternetSafety.com, 2009) Compañía estadounidense, con oficina regional en Australia - Asia/Pacífico fundada en 1999, según datos de la propia compañía tiene negocios en más de 140 países (InternetSafety.com, 2009). Brinda una suite de productos relativos a garantizar protección contra contenidos indeseables, en los principales niveles de aplicación. Los principales productos que ofrece son:

- *SafeEyes. Internet Filtering Software* (*For Home*): Esta aplicación brinda soluciones de filtrado para usuarios finales, centros educacionales o empresas, ver Anexo 2. Es una aplicación cliente que se instala en los usuarios finales y se conecta a una infraestructura de servidores en Internet donde son categorizados los contenidos (InternetSafety.com, 2009).
- *EtherShield. Internet Filtering Hardware* (*For Business*): Brinda soluciones de filtrado de Internet por hardware.
- *SafeEyes Mobile* (*Mobile*): Brinda soluciones de filtrado de Internet para dispositivos móviles *iPhone* o *iPod*.

El producto insignia de InternetSafety.com es SafeEyes, sus principales funcionalidades se muestran a continuación (TopTenREVIEWS, 2009):

- Permite filtrar los sitios Web basándose en su dirección. (toptenreviews: URL Based)
- Permite filtrar los sitios Web basándose en determinadas palabras que el sitio puede contener. (toptenreviews: Keyword Based)
- Permite filtrar aplicaciones P2P tales como Gnutella, Kazaa, Morpheus, Limewire, Bearshare y otras. (toptenreviews: Peer-to-Peer (P2P) Blocking)
- Permite editar las listas de direcciones y palabras a filtrar. (toptenreviews: Editable Filter Lists)
- Permite la creación de diferentes perfiles de navegación en dependencia de quien está usando Internet. (toptenreviews: Individual User Profiles)
- Permite la generación de reportes resumen de históricos. (toptenreviews: Summary History Reporting)
- Permite la generación de reportes detallados de históricos. (toptenreviews: Detailed History Reporting)
- Permite la generación de notificaciones y alertas por correo electrónico. (toptenreviews: Notification Alerts by E-mail)

- Cantidad de categorías diferentes que identifica el Sistema de Filtrado: 35. (toptenreviews: Filter Categories)
- El producto proveé documentación, ayudas y soporte técnico. (toptenreviews: Product Documentation).
- El producto está limitado para el navegador Opera. (toptenreviews: Supported Browsers)

El precio de la última versión del software SafeEyes, para usuarios finales es de $49.95 por una licencia anual, en la versión en CD el costo es de $54,95. En el precio están incluidos los siguientes elementos:

- La adquisición del software y la inicialización del servicio.
- Servicio de atención al cliente 24x7.
- Actualizaciones a la aplicación.
- 30 días de garantía según los términos de la licencia (InternetSafety.com, 2009).

La licencia de uso es completamente privativa (InternetSafety.com, 2009) pues nos impide:

- Copiar, distribuir, rentar o arrendar el producto.
- Modificar, adaptar o preparar trabajos derivados del producto.
- Compartir el producto por cualquier medio electrónico.
- Realizarle ingeniería inversa, descompilar o desensamblar el producto.
- Usar componentes del software en otros contextos.

No se encontraron referencias de precios preferenciales para SafeEyes en la medida que aumenta la cantidad de usuarios. Solo para usuario final, tiene un costo de **$49,95** por cada licencia anual. Suponiendo un 20% de descuento para las licencias adicionales, podemos calcular que para una red de 1000 usuarios por un año sería de unos: **$40.009,95 dólares norteamericanos**, el precio real nunca estaría por debajo de este pues se asumieron tarifas preferenciales.

1.6.3. CYBERsitter

Sistema de filtrado número tres en los análisis que se realizan en (TopTenREVIEWS, 2009), por tanto le han adjudicado el título: “TopTenREVIEWS Bronze Award”. CYBERsitter es un producto de Solid Oak Software, Inc. (Solid Oak Software, 2008) Compañía estadounidense, fundada en 1990, según datos de la propia compañía fue el primer filtro comercial para Internet (Solid Oak Software, 2008). Ganador en cinco ocasiones del *PC Magazine's Editors' Choice Award*. Su funcionamiento se concentra en los niveles de usuario final y empresas.

En general el producto CYBERsitter, ver Anexo 3, ofrece el siguiente listado de funcionalidades (TopTenREVIEWS, 2009):

- Permite filtrar los sitios Web basándose en su dirección. (toptenreviews: URL Based)
- Permite filtrar los sitios Web basándose en determinadas palabras que el sitio puede contener. (toptenreviews: Keyword Based)
- Permite filtrar los sitios Web basándose en la posibilidad de que el contenido del sitio cambie aun cuando mantenga la misma dirección. (toptenreviews: Dynamic Categorization)
- Permite filtrar aplicaciones P2P tales como Gnutella, Kazaa, Morpheus, Limewire, Bearshare y otras. (toptenreviews: Peer-to-Peer (P2P) Blocking)
- Permite filtrar determinados puertos. (toptenreviews: Customizable Port Blocking)
- Permite editar las listas de direcciones y palabras a filtrar. (toptenreviews: Editable Filter Lists)
- Permite la creación de diferentes perfiles de navegación en dependencia de quien está usando Internet. (toptenreviews: Individual User Profiles)
- Permite la administración remota del sistema. (toptenreviews: Remote Management)
- Permite establecer políticas de navegación horarias. (toptenreviews: Daily Time Limits)
- Permite la generación de reportes resumen de históricos. (toptenreviews: Summary History Reporting)

- Permite la generación de reportes detallados de históricos. (toptenreviews: Detailed History Reporting)
- Permite la generación de notificaciones y alertas por correo electrónico. (toptenreviews: Notification Alerts by E-mail)
- Cantidad de categorías diferentes que identifica el Sistema de Filtrado: 35. (toptenreviews: Filter Categories)
- El producto provee documentación y ayudas. (toptenreviews: Product Documentation)
- El producto está limitado para algún navegador. (toptenreviews: Supported Browsers)

La empresa que comercializa CYBERsitter (en su actual versión 10), tiene tres formas de comercializar el producto, según el tipo de cliente: el precio para usuarios finales es de $34.95 por cada licencia, para entidades educacionales el precio es de $9,95 por cada estación, para licencias empresariales el precio es de $199,50 por cada 10 estaciones. En todos los casos la licencia no tiene límite de tiempo (Solid Oak Software, 2008). La empresa garantiza los siguientes elementos de soporte:

- Servicio de atención al cliente por correo electrónico y vía telefónica.
- Actualizaciones a la aplicación.
- 30 días de garantía con la devolución de los fondos.

La licencia de uso es privativa pues nos impide:

- Copiar, distribuir, rentar o arrendar el producto.
- Modificar, adaptar o preparar trabajos derivados del producto.
- Compartir el producto por cualquier medio electrónico.
- Realizarle ingeniería inversa, descompilar o desensamblar el producto.
- Usar componentes del software en otros contextos.

Teniendo en cuenta los precios preferenciales para una red de 1000 usuarios empresariales sin límite de tiempo el precio total sería de: **$19.950,00 dólares norteamericanos**.

1.6.4. Optenet

Optenet (OPTENET, 2008) es una multinacional de las soluciones de filtrado para Internet, situada originalmente en San Sebastián, España, tiene oficinas en los Estados Unidos, Argentina, Brasil, Colombia, México, Reino Unido, Francia, Italia y Australia (OPTENET, 2008). Optenet nace en 1997 en pleno apogeo de Internet de la mano de su fundador, Francisco Martín Abreu, con una larga trayectoria en proyectos de las TIC.

Se trata de una solución multilingüe capaz de analizar cualquier información en tiempo real, a partir de técnicas de inteligencia artificial. Provee una completa *suite* de productos y soluciones en todos los niveles, es decir, a nivel *backbone* de Internet, a nivel de proveedores de acceso e instituciones y para usuarios finales. Algunos de los ejemplos son:

- Optenet WebFilter™[2]
- Optenet WebSecure™[3]
- Optenet MailSecure™[4]
- Optenet NetSecure™[5], es la solución integradora ver Anexo 4.
- Optenet ChildSecure[6]
- Optenet ContentSecure[7]
- Optenet End Point Protection[8]

---

[2] http://www.optenet.com/en-us/webfilter.asp

[3] http://www.optenet.com/en-us/websecure.asp

[4] http://www.optenet.com/en-us/mailsecure.asp

[5] http://www.optenet.com/en-us/netsecure.asp

[6] http://www.optenet.com/en-us/ispchildsecure.asp

[7] http://www.optenet.com/en-us/ispcontentsecure.asp

[8] http://www.optenet.com/en-us/ispend.asp

En general la *suite* de productos de Optenet ofrece el siguiente listado de funcionalidades (OPTENET, 2008):

- Permite la creación de diferentes perfiles de navegación.
- Permite la configuración de políticas de navegación flexibles; se puede escoger entre las funciones: permitir, bloquear, continuar, restringir y bloquear por tipo de fichero para administrar el acceso a la Web y filtrar sitios dependiendo además de políticas horarias.
- Puede adaptarse a las regulaciones de diferentes países.
- Permite el control de la descarga a través de protocolos P2P, de aplicaciones como Kazaa, eMule, Gnutella, etc.
- Permite la integración a infraestructuras existentes reduciendo el tiempo de instalación de la solución.
- La interfaz de administración está diseñada para un ambiente profesional, intuitivo y atractivo al mismo tiempo.
- No se requiere conocimiento técnico para configurar las soluciones. No obstante existen algunas funcionalidades para usuarios avanzados tales como listas personales y la configuración de distintos perfiles de usuarios.
- Incorpora detección de virus, spyware y otros tipos de malware por medio de sociedad e incorporación del sistema de detección de Kaspersky Lab antivirus.
- Tiene una taza mínima de errores en el bloqueo de páginas, gracias a que existe un servicio en línea de reclasificación que funciona 24x7.
- Incorpora un completo módulo de reportes, que identifica amenazas externas, usos indebidos del ancho de banda o violaciones de la política de Internet de la organización, reflejando estos eventos en gráficos y lanzando alertas para los administradores.
- Tiene un filtrado Web con alta efectividad, gracias a que incorpora analizadores multilingües (soporta 180 lenguajes) con técnicas avanzadas de inteligencia artificial

que dinámicamente recolectan y clasifican contenidos de Internet, este sistema está bautizado como MIDAS, ver Anexo 5, del ingles Multicontent Inspection and Dynamic Analysis System (OPTENET, 2008).

- Combina técnicas de análisis propietarias que tributan a una solución de protección multicapa que detecta correo basura, virus y códigos maliciosos en los correos electrónicos.

En estos momentos según datos de la propia compañía (OPTENET, 2009), tiene alrededor de 75 millones de usuarios finales en todo el mundo distribuidos en alrededor de 60 clientes entre proveedores de acceso a Internet y empresas (OPTENET, 2008). El modelo de negocio de esta suite de productos es variado, puesto que para diferentes ámbitos de instalación los precios se fijan de manera diferente, lo que si es común es que las licencias para uso del software se dan por períodos múltiplos de un año, periodo que al terminar el sistema deja de funcionar a menos que se pague por otra licencia, por ejemplo la suite completa para usuario final tiene un costo de $ 88,50 por cada estación de trabajo donde se instale (Gonzalo, 2008), y del producto Optenet WebFilter de $ 39,00 en estos precios se incluye:

- La adquisición del software y la inicialización del servicio.
- Actualizaciones al software.
- Actualizaciones diarias de las listas y los parámetros del software.
- Servicio de desbloqueo en línea.
- Servicio de atención al cliente.

En la versión de Optenet WebFilter empresarial se ofrecen precios preferenciales de $9,00 por cada licencia anual en una red de 1000 usuarios en total sería de unos: **$9.000,00 dólares norteamericanos**.

1.6.5. Websense

Websense (Websense, 2009) es una multinacional de las soluciones de filtrado para Internet, situada originalmente en los Estados Unidos, tiene oficinas comerciales en: Australia, China,

Francia, Reino Unido, Alemania, Hong Kong, India, Irlanda, Israel, Italia, Japón, Holanda, Singapur y España (Websense, 2009). Websense, Inc. nace en el 1994 aunque con otro nombre, adopta el nombre actual en 1999 y se hace pública en el 2000. Se trata de una suite de productos enfocados en tres grandes grupos: Seguridad para Datos, Web y Correo Electrónico, las soluciones que brinda funcionan en todos los niveles, es decir, a nivel *backbone* de Internet, a nivel de proveedores de acceso e instituciones. Algunos de los productos y soluciones que oferta relativos a la Seguridad Web son (Websense, 2009):

- Websense Web Security Gateway: Analiza el tráfico Web en tiempo real, categorizando instantáneamente contenidos dinámicos, identifica riesgos de seguridad y bloquea software malware. Incorpora además elementos de Web 2.0.
- Websense Web Security: Alerta proactiva y rápida cuando los sitios Web de la empresa están siendo atacados e informa sobre las vulnerabilidades de los servidores Web.
- Websense Web Filter: Es la solución de Filtrado Web de Websense, ver Anexo 6 y Anexo 7, esta permite aumentar la productividad de los empleados, mitiga la responsabilidad legal y optimiza el uso de los recursos de la empresa. Permite escalar desde 50 hasta 250 mil usuarios.
- Websense Express: Es también una solución de Filtrado Web pero pensada para pequeñas empresas, nunca más de 250 usuarios.

Los productos para el Filtrado Web cumplen con las siguientes características:

- Permite filtrar los sitios Web basándose en algún tipo de análisis de los objetos del sitio. (toptenreviews: Object Analysis)
- Permite filtrar los sitios Web basándose en su dirección. (toptenreviews: URL Based)
- Permite filtrar los sitios Web basándose en determinadas palabras que el sitio puede contener. (toptenreviews: Keyword Based)

- Permite filtrar los sitios Web basándose en la posibilidad de que el contenido del sitio cambie aun cuando mantenga la misma dirección. (toptenreviews: Dynamic Categorization)
- Permite filtrar aplicaciones P2P tales como Gnutella, Kazaa, Morpheus, Limewire, Bearshare y otras. (toptenreviews: Peer-to-Peer (P2P) Blocking)
- Permite filtrar determinados puertos. (toptenreviews: Customizable Port Blocking)
- Permite editar las listas de direcciones y palabras a filtrar. (toptenreviews: Editable Filter Lists)
- Permite la creación de diferentes perfiles de navegación en dependencia de quien está usando Internet. (toptenreviews: Individual User Profiles)
- Permite la administración remota del sistema. (toptenreviews: Remote Management)
- Permite establecer políticas de navegación horarias. (toptenreviews: Daily Time Limits)
- Permite la generación de reportes gráficos de históricos. (toptenreviews: Graphical Reporting)
- Permite la generación de reportes resumen de históricos. (toptenreviews: Summary History Reporting)
- Permite la generación de reportes detallados de históricos. (toptenreviews: Detailed History Reporting)
- Cantidad de categorías diferentes que identifica el Sistema de Filtrado: 90. (toptenreviews: Filter Categories)
- El producto provee documentación y ayudas. (toptenreviews: Product Documentation).
- El producto no está limitado para ningún navegador. (toptenreviews: Supported Browsers)

En estos momentos según datos de la propia compañía (Websense, 2009) se le brinda protección con sus productos a 42 millones de usuarios finales en todo el mundo distribuidos en alrededor de 50 mil organizaciones. Es una compañía que utiliza el modelo Investigación-Desarrollo. En general los productos de Websense utilizan la tecnología ThreatSeeker que

no es más que la aglutinación de un conjunto de técnicas avanzadas: defensa proactiva ante amenazas, utilización de Inteligencia Artificial, minería de datos y otras que le permiten a la empresa revisar más de 40 millones de sitios cada hora en busca de códigos maliciosos, asignar diferentes clasificaciones a más de 2 millones de dominios cada hora, clasificar en basura o no 10 millones de correos cada hora, analizar con minería de datos 100 millones de sitios Web cada día, mantener en funcionamiento 50 millones de sistemas de recolección de datos en Internet que separan un billón de piezas de contenidos cada día, mantener un arreglo de computadoras realizando minería de datos sobre los sitios Web logrando que cada sitio de la base de datos global sea revisado como promedio cada 12 horas.

El modelo de negocio de esta suite de productos por la venta de licencias por el período de uno, dos o tres años, el precio de una licencia de Websense Web Filter para 25 usuarios por un año es de $1.000,00 (Websense, 2009), esto incluye:

- La adquisición del software y la inicialización del servicio.
- Actualizaciones al software.
- Actualizaciones diarias de las listas y los parámetros del software.
- Servicio de atención al cliente.

El modelo de licencia es privativo. En la versión de Websense Web Filter el precio de la licencia por un año para 500 usuarios es de $10.000,00 (Websense, 2009) por lo cual para una red de 1000 usuarios por un año el precio oscilaría entre los $15.000,00 y los **$20.000,00 dólares norteamericanos**.

1.6.6. Proventía Web Filter

Proventía Web Filter (IBM Corp, 2009) es un producto de la Corporación multinacional IBM (conocida también como “El Gigante Azul”), tiene su cede en Armonk en los Estados Unidos, tiene empleados en alrededor de 161 países, fue constituida oficialmente en el 1911. En realidad IBM tiene una división que se encarga de las soluciones de seguridad: Internet Security System, esta tiene toda una suite de productos uno de los cuales es Proventia Web

Filter (Internet Security Systems, 2007). Las características fundamentales de este producto son:

- Permite filtrar los sitios Web basándose en algún tipo de análisis de los objetos del sitio. (toptenreviews: Object Analysis)
- Permite filtrar los sitios Web basándose en su dirección. (toptenreviews: URL Based)
- Permite filtrar los sitios Web basándose en determinadas palabras que el sitio puede contener. (toptenreviews: Keyword Based)
- Permite filtrar los sitios Web basándose en la posibilidad de que el contenido del sitio cambie aun cuando mantenga la misma dirección. (toptenreviews: Dynamic Categorization)
- Permite filtrar aplicaciones P2P tales como Gnutella, Kazaa, Morpheus, Limewire, Bearshare y otras. (toptenreviews: Peer-to-Peer (P2P) Blocking)
- Permite filtrar determinados puertos. (toptenreviews: Customizable Port Blocking)
- Permite editar las listas de direcciones y palabras a filtrar. (toptenreviews: Editable Filter Lists)
- Permite la creación de diferentes perfiles de navegación en dependencia de quien está usando Internet. (toptenreviews: Individual User Profiles)
- Permite la administración remota del sistema. (toptenreviews: Remote Management)
- Permite establecer políticas de navegación horarias. (toptenreviews: Daily Time Limits)
- Permite la generación de reportes gráficos de históricos. (toptenreviews: Graphical Reporting)
- Permite la generación de reportes resumen de históricos. (toptenreviews: Summary History Reporting)
- Permite la generación de reportes detallados de históricos. (toptenreviews: Detailed History Reporting)
- Cantidad de categorías diferentes que identifica el Sistema de Filtrado: 68. (toptenreviews: Filter Categories)

- El producto provee documentación y ayudas. (toptenreviews: Product Documentation).
- El producto no está limitado para ningún navegador. (toptenreviews: Supported Browsers)

En estos momentos según datos de la propia compañía (IBM Corp, 2009) la empresa ISS tiene más de 11 mil clientes corporativos. Por otro lado la base de datos de Proventía tiene 87 millones de URL analizadas y más de 7 billones de páginas e imágenes, esta base de datos es mantenida por una granja de mil servidores. Cada día alrededor de 1.200 robots web navegan por Internet en busca de nuevos contenidos o actualizar los que existen, de esta manera se logran actualizar o adicionar 150 mil contenidos cada día. Las tecnologías de análisis incluyen texto e imágenes.

El modelo de negocio de este productos es por la venta de licencias por el período de uno, dos o tres años, el precio de una licencia de Proventía Web Filter para 1 usuario por un año es de $6.55 (AlphaStore, 2009), la empresa provee:

- Actualizaciones al software.
- Actualizaciones diarias de las listas y los parámetros del software.
- Servicio de atención al cliente.

El modelo de licencia es privativo, su precio por un año para 1000 usuarios (AlphaStore, 2009) es de **$25.935,91 dólares norteamericanos.**

1.6.7. POESIA (*Public Open-source Environment for a Safer Internet Access*)

El proyecto POESIA[9] (2002-2004) fue financiado por la Comisión Europea y tiene lugar en el marco de la Sociedad de la Información y un Plan de Acción Tecnológica para una Internet más segura. POESIA[10] busca desarrollar, probar, evaluar y promover un filtro completamente

[9] http://www.poesia-filter.org/

[10] http://sourceforge.net/projects/poesia/

Open Source y que se encuentre en el estado del arte. Desde su concepción se diseña para que sea capaz de filtrar contenidos inadecuados en diferentes protocolos combinando tecnologías innovadoras presentes en los filtros más efectivos de la actualidad. Se realiza el filtrado del lenguaje natural en varios idiomas, filtrado de imágenes, filtrado por URL, PICs y JavaScript. El proyecto pretende ser la solución estándar de filtrado para instituciones educacionales, bibliotecas, cibercafés y el hogar.

En la implementación de este producto las características que se incluyeron fueron (Donert, et al., 2002):

- Permite filtrar los sitios Web basándose en algún tipo de análisis de los objetos del sitio. (toptenreviews: Object Analysis)
- Permite filtrar los sitios Web basándose en su dirección aunque el rendimiento para este mecanismo no es bueno. (toptenreviews: URL Based)
- Permite filtrar los sitios Web basándose en la posibilidad de que el contenido del sitio cambie aun cuando mantenga la misma dirección. (toptenreviews: Dynamic Categorization)
- Permite filtrar aplicaciones P2P tales como Gnutella, Kazaa, Morpheus, Limewire, Bearshare y otras. (toptenreviews: Peer-to-Peer (P2P) Blocking)
- Permite filtrar determinados puertos. (toptenreviews: Customizable Port Blocking)
- Permite editar las listas de direcciones a filtrar. (toptenreviews: Editable Filter Lists)
- Permite establecer políticas de navegación horarias. (toptenreviews: Daily Time Limits)
- Cantidad de categorías diferentes que identifica el Sistema de Filtrado: 7. (toptenreviews: Filter Categories)
- El producto provee documentación y ayudas. (toptenreviews: Product Documentation).
- El producto no está limitado para ningún navegador. (toptenreviews: Supported Browsers)

Este filtro se permite tener una característica que no tienen los sistemas de filtrado privativos y es que su Base de Datos no es secreta, por lo cual puede conocerse que URL pertenece a cada categoría. Entre las funcionalidades que incluye esta el soporte para PICS. El punto débil de este sistema es la no existencia de un grupo que se mantenga desarrollando mejoras y manteniendo el sistema, por otro lado una base de datos centralizada pudiera aprovechar las adiciones de todos los usuarios, en el área de reportes tampoco tiene mucho que decir. En este estudio no se ha identificado ningún modelo de negocio para este filtro, solo una iniciativa para extender su utilización.

1.6.8. DansGuardian

DansGuardian en su versión 2 es un Sistema de Filtrado licenciado bajo GPLv2 siempre que no se utilice con fines comerciales. Su primera versión alfa sale a la luz en enero del 2001 liberándose como estable en junio de ese mismo año (Barron, 2009). DansGuardian lo usan más de 30 clientes los cuales son centros educacionales fundamentalmente (Barron, 2007). Las características fundamentales de este producto son:

- Permite filtrar los sitios Web basándose en su dirección. (toptenreviews: URL Based)
- Permite filtrar los sitios Web basándose en determinadas palabras que el sitio puede contener. (toptenreviews: Keyword Based)
- Permite filtrar los sitios Web basándose en la posibilidad de que el contenido del sitio cambie aun cuando mantenga la misma dirección. (toptenreviews: Dynamic Categorization)
- Permite filtrar aplicaciones P2P tales como Gnutella, Kazaa, Morpheus, Limewire, Bearshare y otras. (toptenreviews: Peer-to-Peer (P2P) Blocking)
- Permite filtrar determinados puertos. (toptenreviews: Customizable Port Blocking)
- Permite editar las listas de direcciones y palabras a filtrar. (toptenreviews: Editable Filter Lists)

- Permite la creación de diferentes perfiles de navegación en dependencia de quien está usando Internet. (toptenreviews: Individual User Profiles)
- Permite establecer políticas de navegación horarias. (toptenreviews: Daily Time Limits)
- Cantidad de categorías diferentes que identifica el Sistema de Filtrado: 70 usando la base de datos de listas negras urlblacklist.com[11] aunque también está disponible Shalla's Blacklists[12]. (toptenreviews: Filter Categories)
- El producto provee documentación y ayudas. (toptenreviews: Product Documentation).
- El producto no está limitado para ningún navegador. (toptenreviews: Supported Browsers)

Este filtro se permite tener una característica que no tienen los sistemas de filtrado privativos y es que su Base de Datos no es secreta pues usa urlblacklist.com o Shalla's Blacklists, por lo cual puede conocerse que URL pertenece a cada categoría. Entre las funcionalidades que incluye esta el soporte para PICS. Los puntos débiles de esta aplicación son:

- Carencia de una interfaz para administrar todas las funcionalidades del sistema.
- No tiene sistema de reportes propios.
- No utiliza mecanismos de Inteligencia Artificial para el análisis de los contenidos.

El modelo de negocio de este producto es por la venta de licencias a usuarios con intereses comerciales y por la promoción de servicios de descargas de Bases de Datos de listas negras y soporte técnico. Además este producto le pone precio a la descarga del software con fines comerciales, así para una red comercial de 1000 usuarios habría que pagar **$210,00 dólares norteamericanos**.

[11] http://urlblacklist.com/

[12] http://www.shallalist.de/

1.6.9. Squid-Guard

Los primeros conceptos de SquidGuard fueron desarrollados por Pål Baltzersenen y Lars Erik Håland alrededor de 1998, en Tele Danmark InterNordia. Resultó ser una aplicación muy rápida en la época y además libre por lo que suscitó atención por entonces. Con el paso del tiempo y tras la versión 1.2.0 en 2001 su desarrollo se descontinuó casi por completo, hasta el punto que en el verano del 2006 Tele Danmark InterNordia dejó de soportar completamente la aplicación eliminándose incluso las entradas DNS para squidguard.org. Algunos meses después el código fuente se le traspasó a Shalla Secure Services[13] quienes desde entonces mantienen SquidGuard. Está liberado bajo la licencia GPLv2, su funcionamiento es acoplado al squid-cache[14] proxy. Las características fundamentales de este producto son:

- Permite filtrar los sitios Web basándose en su dirección. (toptenreviews: URL Based)
- Permite filtrar los sitios Web basándose en determinadas palabras que el sitio puede contener. (toptenreviews: Keyword Based)
- Permite filtrar aplicaciones P2P tales como Gnutella, Kazaa, Morpheus, Limewire, Bearshare y otras. (toptenreviews: Peer-to-Peer (P2P) Blocking)
- Permite filtrar determinados puertos. (toptenreviews: Customizable Port Blocking)
- Permite editar las listas de direcciones y palabras a filtrar. (toptenreviews: Editable Filter Lists)
- Permite la creación de diferentes perfiles de navegación en dependencia de quien está usando Internet. (toptenreviews: Individual User Profiles)
- Permite establecer políticas de navegación horarias. (toptenreviews: Daily Time Limits)

---

[13] http://www.shalla.de/

[14] http://www.squid-cache.org/

- Cantidad de categorías diferentes que identifica el Sistema de Filtrado: 70 usando la base de datos de listas negras urlblacklist.com aunque también está disponible Shalla's Blacklists. (toptenreviews: Filter Categories)
- El producto provee documentación y ayudas. (toptenreviews: Product Documentation).
- El producto no está limitado para ningún navegador. (toptenreviews: Supported Browsers)

Este filtro se permite tener una característica que no tienen los sistemas de filtrado privativos y es que su Base de Datos no es secreta pues usa urlblacklist.com o Shalla's Blacklists, por lo cual puede conocerse que URL pertenece a cada categoría. Los puntos débiles de esta aplicación son:

- Carencia de una interfaz para administrar todas las funcionalidades del sistema.
- No tiene sistema de reportes propios.
- No realiza ningún tipo de análisis de los contenidos, solo trabaja a nivel de URL.

En este producto no se identificó ningún tipo de visión comercial, es completamente libre y gratis.

1.7. Conclusiones

En el epígrafe anterior se expusieron nueve sistemas de filtrado de los más reconocidos a nivel mundial, en este estudio no se encontraron referencias de software similares en nuestro país. De los nueve seis son productos Estadounidenses, uno Español, otro financiado por la Unión Europea y el tercero restante Alemán. Los seis primeros que se presentan son productos comerciales de licencias privativas y con distintos modelos de negocio, los últimos tres pertenecen a la categoría de Software Libre. En el caso de los privativos vale destacar el caso de NetNanny para usuarios finales pues es uno de los más completos, para el sector empresarial y a nivel de proveedores de acceso a Internet se destacan los casos de Websense, Proventía y Optenet, productos muy completos, en estos la compañía fabricante dispone de granjas de servidores dedicados a las tareas de recuperación y clasificación de

páginas Web con tecnologías que incorporan técnicas de Inteligencia Artificial que les permiten mantener una base de datos de contenidos categorizados y brindar el servicio de actualizaciones de las bases de datos locales de sus clientes, con estos modelos de negocio **el cliente siempre depende de las actualizaciones de la compañía fabricante del software y de pagos cíclicos para renovar la licencia de uso**, de lo contrario no solo dejaría de actualizarse la aplicación sino que también dejaría de funcionar por completo. Para el caso de los productos que pertenecen a la categoría de Software Libre vale la pena destacar el esfuerzo de la Unión Europea en la financiación del proyecto POESIA, pues se identifica como uno de los más se acerca a los líderes de software comercial de su tipo, el mayor problema que tiene es que en la actualidad no tiene un grupo de personas encargadas de darle continuidad en el tiempo al proyecto, pues por principio fue un proyecto financiado con un periodo de duración de dos años (2002-2004) que logró ponerse en el estado del arte y alcanzar un nivel funcional bastante elevado. Los productos libres de filtrado: SquidGuard y DansGuardian están lejos de los productos comerciales tanto en nivel funcional como en estado del arte pues solo incorporan mecanismos de filtrado por listas negras y palabras claves, no tienen interfaces de administración y su instalación y configuración es a nivel de líneas de comandos Unix.

Capítulo 2. Características de la solución

2.1. Introducción

En este trabajo proponemos una solución de Filtrado para Internet coherente con el estado del arte Internacional en esta tecnología, de esta forma se presenta la concepción del producto FILPACON en su versión 1.0 registrado en el Centro Nacional de Derechos de Autor (CENDA) con número de registro 3421-2008, ver Anexo 8, FILPACON es una marca registrada por la Universidad de las Ciencias Informáticas con número 2008-0303 (Boletín Oficial 247 Noviembre, 2008).

FILPACON en un inicio se pensó como una solución de Filtrado de Paquetes por Contenidos, de ahí el acrónimo, más tarde tras el estudio del estado del arte de la problemática de los contenidos en Internet y de las soluciones que a nivel internacional se daban, se decidió cambiar la visión y el alcance de la solución para un Filtro de Contenido Web; sin embargo el acrónimo aunque técnicamente incorrecto por costumbre ha perdurado.

Durante este capítulo se presentarán primero los requerimientos no funcionales que se tuvieron en cuenta desde un principio en el diseño de la solución y en un segundo momento se presenta la arquitectura del sistema junto a los requerimientos funcionales de cada uno de los subsistemas que lo componen.

2.2. Requerimientos no funcionales

En general existen algunos requerimientos que no tienen que ver directamente con el filtrado en si mismo, pero que se deben tener en cuenta durante el diseño del sistema para lograr la satisfacción del usuario:

- **Tiempo de respuesta**: es probablemente una de las características más importantes para los usuarios, no serviría de mucho un filtrado sin errores si las páginas demoraran más de 30 segundos en visualizarse. Lograr que la latencia que introducen

los algoritmos de filtrado sea despreciable es uno de los elementos a tener en cuenta en todo momento.

- **Fiabilidad**: el resultado del filtrado debe tener una baja tasa de errores, en este tipo de aplicaciones los errores por encima del 10% repercutirán directamente en el prestigio del sistema entre la comunidad de usuarios.
- **Escalabilidad**: otra de las características imprescindibles para que los usuarios se sientan cómodos ante el crecimiento o cambios en su organización. Una solución escalable siempre es más económica de mantener, la mayoría de los usuarios ya se enteraron de esta peculiaridad.
- **Compatibilidad**: los usuarios de seguro ya tienen una infraestructura que funciona, por lo tanto no van a querer tener que cambiarla toda solo para incorporar un nuevo elemento, para ellos la mejor solución es la que se integre de la manera más transparente posible a las soluciones que tienen en producción.
- **Fácil de usar**: una aplicación demasiado difícil de poner a punto o configurar de seguro será rechazada por la mayoría de los clientes potenciales, por otro lado los clientes no tienen por que ser expertos en temas informáticos.

2.3. Arquitectura general del sistema

El producto FILPACON ha pasado por varias versiones: v0.5, v0.6, v0.7 y v0.8 hasta finalmente llegar hasta la versión v1.0, no obstante la ideas basales de la arquitectura general fueron concebidas desde la versión 0.5, ver Anexo 9, esta se le presentó a los especialistas de la Oficina de Seguridad para las Redes Informáticas (OSRI)[15] alrededor del mes de abril del 2005. La OSRI desde el principio estuvo muy ligada al desarrollo del producto, fue quien solicitó inicialmente a la Universidad el desarrollo de un software que

[15] Cuyo objetivo central será el de proponer medidas y tomar las acciones necesarias para lograr la invulnerabilidad de las redes informáticas cubanas (Lage Dávila, 2000).

permitiese regular el acceso a Internet, como parte de un programa que pretende llevar Internet de manera segura a todos los puntos del país.

Durante este epígrafe no se profundizará en las características particulares de cada versión sino que se presentarán aquellas que han tenido una repercusión importante en la v1.0. El sistema integra para su funcionamiento al proxy squid-cache.org[16], como servidor de bases de datos al gestor PostgreSQL[17], como servidor http al apache2[18], como sistema operativo base a Debian[19] GNU/Linux del cual además se utilizó el instalador, para el desarrollo web el lenguaje PHP y el framework Symfony[20], para el desarrollo de la lógica de filtrado al lenguaje Perl[21] y librerías todos estos aspectos son detallan en profundidad en la documentación del proyecto. Todo el software que se utilizó pertenece a la categoría de Software Libre, el *framework* y librerías tienen modelos de licencia sin *copyright* lo cual permite liberar FILPACON bajo cualquier modelo de licencia incluso modelos no-libres.

El principio de funcionamiento del sistema se basa en que las peticiones de recursos de Internet de los usuarios son capturadas por el squid-cache proxy y le son entregadas al Subsistema de Filtrado, ver epígrafe 2.3.1, en el Anexo 9 este componente tiene el nombre de redirector; a partir de este momento dependiendo de la política de navegación que tiene definida el usuario y el tipo de contenido al que pertenece el objeto se toma la decisión sobre denegar o permitir el acceso al contenido solicitado.

---

[16] http://www.squid-cache.org/

[17] http://www.postgresql.org/

[18] http://httpd.apache.org/

[19] http://www.debian.org/

[20] http://www.symfony-project.org/

[21] http://www.perl.org/

Desde sus inicios el sistema se ha concebido para lograr escalabilidad puesto que cada subsistema[22] puede funcionar en servidores independientes y se comunica con el resto utilizando un paradigma arquitectónico cliente-servidor, de esta manera se pueden realizar despliegues simples, ver Anexo 10, o despliegues para entornos de alta demanda, ver Anexo 11.

2.3.1. Subsistema de Filtrado

Este subsistema de manera conceptual incluye los módulos el de autenticación y el de redirección. El módulo de autenticación garantiza la validación de los usuarios, esta puede realizarse contra la base de datos local del sistema o a través de la integración en un Directorio Activo de Microsoft Windows[23]. En el módulo de redirección es donde se encuentra la lógica para el filtrado, partiendo de la política de navegación que tenga definida un usuario, si es mayor de edad o no y el tipo de contenido al que accede según el Subsistema de Almacenamiento. La **política de navegación** es una propiedad de los usuarios y los valores posibles son:

0. **Ninguna**: significa que al usuario se le deniega siempre el servicio de Internet.
1. **Restrictiva**: se permite el acceso bajo determinadas condiciones que dependen de la edad y el tipo de contenido al que accedió.
2. **Permisiva**: se permite el acceso bajo determinadas condiciones que dependen de la edad y el tipo de contenido al que accedió.
3. **Básica**: se permite el acceso solo a una lista de dominios definidos en el Subsistema de Almacenamiento.
4. **Plena**: se permite siempre el acceso sin tener en cuenta ninguna otra restricción.

[22] Excepto el Subsistema de Instalación que solo funciona en el momento que el sistema es instalado.

[23] http://www.microsoft.com/

La diferencia entre la política de navegación permisiva y restrictiva tiene que ver con que decisión tomar en el caso de que un contenido no se encuentre registrado en el Subsistema de Almacenamiento, así cuando la política de navegación es permisiva todos los objetos de Internet que no se encuentren en la base de datos son permitidos.

Los tipos de contenido pueden ser:

0. **Deshabilitado**: se le asigna a las categorías que se quieran mantener en la base de datos pero que no influyan en las decisiones de filtrado.
1. **Adecuado**: son aquellos contenidos inocuos para todo tipo de usuarios, mayores y menores de edad.
2. **Ilícito**: son aquellos contenidos que deben serles denegados a todos los tipos de usuarios, mayores y menores de edad.
3. **Nocivo**: son los contenidos que solo se le deben denegar a los usuarios menores de edad.

En la Tabla 2 se detalla que decisión se toma en cada caso.

Edad (M/m)	Política de Navegación	Decisión a tomar	Subsistema de Almacenamiento (BD)
M/m	Ninguna ('0')	No tiene acceso a Internet	no influye
M/m	Básica ('3')	Solo se tiene acceso a las direcciones definidas como básicas	la tabla dominios básicos define las direcciones que pertenecen a esta política de navegación
m	Restrictiva ('1')	Permite solamente las categorías adecuadas	si la dirección no existe en la BD se deniega por defecto
M	Restrictiva ('1')	Permite solamente las categorías adecuadas y nocivas	si la dirección no existe en la BD se deniega por defecto
m	Permisiva ('2')	Deniega solamente las categorías ilícitas y nocivas	si la dirección no existe en la BD se permite por defecto
M	Permisiva ('2')	Deniega solamente las categorías ilícitas	si la dirección no existe en la BD se permite por defecto
M/m	Plena ('4')	Acceso Total	no influye

Tabla 2 Tabla descriptiva de la decisión para el filtrado

Por otro lado en el Anexo 12 se muestra el diagrama en bloques del algoritmo de filtrado que implementa todos los casos de la Tabla 2, además de incorporar los elementos necesarios para recoger la información que se utiliza en el módulo de reportes del Subsistema de Administración Web, la gestión de los accesos indebidos (incidencias) y en el caso de que se le deniegue un contenido al usuario construye la nueva URL pasando los parámetros a la página de denegación, esta última le muestra al usuario los motivos por los cuales se le denegó el acceso como se muestra en un ejemplo en el Anexo 13.

Por último decir que este subsistema es el cuello de botella del Sistema de Filtrado, por lo cual se le realizaron optimizaciones al código y pasó por un riguroso conjunto de pruebas de rendimiento, estas actividades fueron acometidas por los desarrolladores expertos del proyecto. La meta fue de lograr que el Sistema de Filtrado no incluyera tiempos de latencia mayores de 50 ms por cada petición de usuario, el promedio que se logró fue de alrededor de 30 ms.

2.3.2. Subsistema de Administración Web

Es el subsistema más grande en cuanto a número de funcionalidades, está compuesto por cinco módulos, en realidad es una interfaz gráfica de usuario que se comienza a incorporar paulatinamente desde la versión 0.7 de FILPACON, en (Hermosilla Moreno, y otros, 2008) se realiza una tesis de ciclo completo de este Subsistema de Administración Web, detallándose en cada caso las funcionalidades que tras la revisión de los productos homólogos a nivel internacional se presentan en este trabajo:

- Módulo de Gestión de Usuarios (Anexo 14)
    - Iniciar sesión de usuario
    - Finalizar sesión de usuario
    - Insertar usuario
    - Eliminar usuario
    - Actualizar los datos del usuario

    - Cambiar contraseña
    - Denegar la navegación
    - Habilitar la navegación
- Módulo de Gestión de Categorías (Anexo 15)
    - Insertar nueva categoría
    - Eliminar categoría existente
    - Seleccionar tipo de contenido para cada categoría
- Módulo de Gestión de URLs (Anexo 16)
    - Insertar o actualizar una de las categorías de una URL
    - Comprobar las categorías a las que pertenece una URL
    - Eliminar una o todas las categorías de una URL
- Módulo de Gestión de Dominios
    - Adicionar un dominio de Navegación Básica
    - Eliminar un dominio de Navegación Básica
- Módulo de Reportes (Anexo 17)
    - Incidencias por usuario
    - Ranking de URLs denegadas
    - Ranking de usuarios por incidencias
    - Ranking de días de la semana por incidencias
    - Cantidad de incidencias por tipos de usuarios
    - Cantidad de sitios por categorías
    - Posibles denegaciones incorrectas
    - Ranking de usuarios por cuota usada

Con esta interfaz se persigue el objetivo de facilitar al máximo todas las tareas de gestión que provee el sistema, por otro lado al utilizarse para su implementación las tecnologías Web se garantiza que la administración del sistema se logre de manera remota, sin necesidad instalar ninguna aplicación pues basta con contar con un navegador Web.

2.3.3. Subsistema de Instalación

Es el primer subsistema al cual se enfrentan los administradores del sistema, con el se garantiza que este proceso inicial sea relativamente fácil de llevar a cabo y que los conocimientos que se necesiten para utilizarlo sean los que dispone cualquier persona que halla instalado un sistema operativo común como es el caso de Debian GNU/Linux (Debian, 2009). En la implementación actual de este subsistema se reutilizó el instalador de esta distribución, producto de que la misma soporta 11 arquitecturas de hardware diferentes[24], quedando solamente para el proyecto el desarrollo de las funcionalidades relativas a la instalación del Sistema de Filtrado y dejando toda la detección de hardware y preparación del servidor al instalador de esta distribución. De esta manera se incluye FILPACON como un paquete más entre los que se encuentran en el disco de instalación de debían GNU/Linux 4.0 esto se puede ver en el Anexo 18.

2.3.4. Subsistema de Almacenamiento

Este subsistema en su concepción consta de tres elementos cada uno de los cuales tiene una función específica dentro del Sistema de Filtrado:

1. **Base de datos de FILPACON**: es la base de datos que se incorpora en el disco compacto del producto FILPACON, se puede ver el diagrama físico en el Anexo 19, está formada por siete tablas en (Hermosilla Moreno, y otros, 2008) se detalla cada atributo de cada tabla y las relaciones entre ellas:
    - **t_usuario**: donde se guardan los datos de los usuarios del sistema, incluyendo todo lo relativo a sus políticas de navegación.
    - **t_dominio_basico**: donde se guardan los dominios para el tipo de navegación básica.

[24] Son soportadas las arquitecturas: alpha, amd64, arm, armel, hppa, i386, ia64, mips, mipsel, powerpc y sparc, ver http://www.debian.org/ports/

- **t_rango_restringido**: en esta tabla se guarda información sobre los usuarios a los cuales se le deniega totalmente el acceso a Internet durante un período de tiempo.
- **t_bitacora**: donde se registran todas las actividades de los usuarios, siempre que la propiedad de monitorizar del usuario esté habilitada para el usuario en cuestión.
- **t_categoria**: es el lugar donde se encuentran registradas todas las categorías que soporta el sistema, así como el tipo de contenido al que pertenece cada una, ver 2.3.1.
- **t_principal**: es la tabla donde se encuentran almacenados los *hash* de todas las URL con que cuenta la base de datos.
- **t_principal_has_t_categoria**: es una tabla que aparece en el modelo físico de la base de datos, puesto que en el modelo lógico existía una relación mucho a mucho entre las tablas *t_principal* y *t_categoria* que genera esta última tabla. En ella se encuentra la información relativa a que categoría pertenece cada hash(URL) de la tabla principal.

2. **Base de datos maestra:** es la base de datos concebida para ser administrada solamente en el seno del proyecto, en ella está guardada la información relativa a cual *hash*(URL) se corresponde con la URL, información que de ser pública se perdería gran parte del potencial comercial del producto. Por otro lado en ella se registran las modificaciones que tienen lugar sobre las categorías a las que pertenece una URL a manera que se puedan generar paquetes de actualizaciones, su diagrama físico se puede ver en el Anexo 20, varios son los entes que pueden realizar modificaciones sobre esta base de datos: revisores humanos, algoritmos inteligentes de clasificación de información y las aplicaciones que procesan las listas de contenidos públicas en Internet, este último caso se aborda un poco más a continuación.
3. **Módulo de generación de actualizaciones:** son un conjunto de aplicaciones que se encargan de construir la base de datos, tienen vital importancia hasta tanto el proyecto

no cuente con mecanismos propios para actualizarlas, se basan en procesar listas de sitios Web que existen públicas en Internet, se utilizó en este caso URLBlacklist.com[25] y las que provee el Open Directory Project[26] (ODP), este último utiliza RDF para describir las categorías a las que pertenecen las URLs. Este módulo también abarca las aplicaciones que una vez generada la base de datos maestra se creen los ficheros de actualizaciones diarias, los cuales permiten que toda base de datos de los clientes finales del producto sea sincronizada con los cambios que hallan ocurrido en la base de datos maestra, similar a como los antivirus actuales crean paquetes de actualización para las nuevas definiciones de virus.

2.3.5. Subsistema de Recuperación y Clasificación de Información (SRCI)

En el Anexo 9 se le nombra como "*Recuperador y Analizador Inteligente*", su implementación aun está a nivel de investigaciones o de prototipo, en el Polo de Soluciones Informáticas para Internet[27] se trabaja arduamente en la implementación de varios componentes del SRCI. De esta manera se encuentran en faces de: valoración, experimentación y pruebas, variantes de algoritmos para acometer las tareas de recuperar información de Internet y posteriormente clasificarla. En la concepción del Sistema de Filtrado hasta este punto, se han planteado básicamente tres posibilidades de cómo implementar este subsistema:

1. El SRCI se incorpora a la solución de filtrado y cada petición de usuario es analizada por este.

---

[25] http://urlblacklist.com/?sec=download

[26] http://www.dmoz.org/

[27] Polo productivo perteneciente a la Facultad 10 de la Universidad de las Ciencias Informáticas, que tiene como misión: "Producir conocimientos, productos y servicios informáticos para Internet que tributen a la informatización de la sociedad y al desarrollo económico del país; mediante la integración de los procesos de formación, investigación, producción y comercialización de la Universidad de las Ciencias Informáticas"

2. El SRCI se mantiene del lado del equipo de desarrollo y se generan actualizaciones que se le hacen llegar a los clientes en dependencia del modelo de negocio que finalmente se defina para FILPACON.
3. Combinaciones de los casos anteriores.

Cualquiera sea la forma que finalmente se implemente en FILPACON, esta debe estar sincronizada con el modelo de negocio que se pretenda implantar, este establecerá los principios que guíen el futuro desarrollo de esta herramienta. Por otro lado, en el alcance de este trabajo no se presentan todas las variantes posibles de formas de cómo acometer la tarea de mantener actualizadas las bases de datos de FILPACON, ejemplo de ello es el desarrollo de una herramienta para categorizar páginas Web que se puede revisar en detalle en (Pérez Clemente, y otros, 2008) y la propuesta de un modelo neuronal que permita la categorización de texto en (Ceballo Gastell, y otros, 2008), otros son componentes esenciales como el identificador de idiomas que se implementó en (Hernández Moya, y otros, 2008). No obstante la mayoría de los trabajos no han llegado a la etapa final de su desarrollo o no han sido publicados aun, se espera que en un futuro próximo se cuente con un conjunto de componentes que sirvan de base para generar las tan necesarias actualizaciones.

Capítulo 3. Análisis de los resultados

Tras la concepción del sistema y su implementación se pueden presentar algunos de los resultados obtenidos con este, así en este capítulo se muestra en un primer momento una comparación a nivel funcional entre Sistemas de Filtrado para Internet, posteriormente se especifican los lugares fundamentales donde el sistema ha sido desplegado hasta el momento y por último se abunda en los aportes de este trabajo para el país.

3.1. Comparación funcional de algunos Sistemas de Filtrado

En la Tabla 3 se realiza una comparación a nivel funcional entre algunos de los principales productos para el filtrado de Internet, en la cual se incluyó a FILPACON en su v1.0, en ella se incluyeron productos comerciales con licencias privativas y otros de licencias GPL. Otros productos también importantes no fueron incluidos por motivos de espacio, aunque se analizaron en los epígrafes del capítulo uno de este trabajo.

Productos / Características	Proventía	Websense	NetNanny	POESIA	Squid-Guard	Dans-Guardian	FilPaCon
Permite filtrar los sitios Web basándose en algún tipo de análisis de los objetos del sitio.	SI	SI	Sin IA	SI	NO	Sin IA, análisis de frases	Implem.
Permite filtrar los sitios Web basándose en su dirección.	SI	SI	SI	Ineficiente	SI	SI	SI
Permite filtrar los sitios Web basándose en determinadas palabras que el sitio puede contener.	SI	SI	SI	NO	SI	SI	SI
Permite filtrar determinados puertos.	SI	SI	NO	SI	SI	SI	SI
Permite editar las listas de direcciones y palabras a filtrar.	SI	SI	SI	SI	SI	SI	SI

Productos / Características	Proventía	Websense	NetNanny	POESIA	Squid-Guard	Dans-Guardian	FilPaCon
Permite la creación de diferentes perfiles de navegación en dependencia de quien está usando Internet.	SI	SI	SI	NO	NO	NO	SI
Posee una interfaz gráfica para administrar el sistema.	SI	SI	SI	NO	NO	NO	SI
Permite la administración remota del sistema.	SI	SI	NO	NO	NO	NO	SI
Permite establecer políticas de navegación horarias.	SI	SI	SI	SI	SI	SI	SI
Permite la generación de reportes gráficos de históricos.	SI	SI	SI	NO	NO	NO	SI
Cantidad de categorías diferentes que identifica el Sistema de Filtrado.	68	90	31	7	Depende de la BD	Depende de la BD	28
El producto provee documentación y ayudas (técnica o comercial).	comercial	comercial	comercial	técnica	técnica, limitada	técnica, limitada	técnica
El producto está limitado para algún navegador.	No limitado	No limitado	Soporta: IE, Firefox, Opera, NetScape.	No limitado	No limitado	No limitado	No limitado
Sistemas operativos que soporta en los clientes de Internet.	No limitado	No limitado	Windows Vista, XP, 2000, MAC	No limitado	No limitado	No limitado	No limitado
El producto permite a los usuarios que puedan realizar reportes sobre posibles regulaciones incorrectas.	No en la aplicación	No en la aplicación	NO	NO	NO	NO	SI
El producto facilita su instalación y configuración inicial.	SI	SI	SI	Necesita de expertos	NO	NO	SI
Nivel jerárquico donde se instala.	Backbone, ISP, Org.	ISP, Org.	Cliente final	ISP, Org.	ISP, Org.	ISP, Org.	ISP, Org.

Productos / Características	Proventía	Websense	NetNanny	POESIA	Squid-Guard	Dans-Guardian	FilPaCon
Permite escalabilidad en la solución.	SI	SI	NO	NO	NO	NO	SI
Costo aproximado para 1000 usuarios en un año.	$ 26 mil	$ 20 mil	$ 54 mil	Sin costo	Depende de la BD	Uso comercial: $210 + BD usa una GPL-Proxy	Sin costo para Cuba, reporta ingresos
Empresa que lo respalda.	ISS, IBM, Corp.	Websense Inc.	ContentWatch, Inc.	UE	Shalla Secure Services	comunidad	UCI
Permite gestión de cuotas de tráfico por usuarios.	NO	NO	NO	NO	NO	NO	SI

Tabla 3: Tabla comparativa de funcionalidades entre Sistemas de Filtrado para Internet reconocidos a nivel Internacional y FilPaCon v 1.0.

3.2. Despliegue y resultados de la solución

La versión 1.0 de FILPACON, fue liberada por el Laboratorio de Calidad de la Universidad de las Ciencias Informáticas ha obtenido importantes resultados y ha sido desplegada en varios entornos reales, a continuación los más importantes en cuanto a número de usuarios:

- Prueba de campo en la Universidad de las Ciencias Informáticas, se pusieron a navegar todos los estudiantes de las Facultades 1 y 10, para un total aproximado de 2000 usuarios (ver Anexo 24), se tomó una muestra de 31 días, desde el 15 de enero del 2009 hasta el 15 de febrero del 2009, de esta manera se determinó que el servidor soportó una carga promedio de 117 peticiones por minuto. Sin embargo, con la configuración del servidor que se dispuso se llegó al límite de consumo de los recursos computacionales con estas dos facultades (2000 usuarios).
- Servidor de experimentación del proyecto en la Facultad 10 de la Universidad de las Ciencias Informáticas, manteniendo un funcionamiento estable durante varios años, se tienen registros desde el 23 de marzo del 2007 hasta nuestros días. En este servidor

se han puesto en un par de ocasiones todos los estudiantes de la facultad 10 (ver Anexo 22), para estimar la carga soportada en un día se tomó una muestra desde el 25 de septiembre del 2008 hasta el 25 de octubre del 2008, en el cual toda la facultad se encontraba navegando por este servidor de manera alternativa pues tenían servicio también a través del servidor oficial de la universidad, en total alrededor de 1000 estudiantes y más de 150 profesores, la carga promedio soportada fue de 41,3 peticiones por minuto, con un máximo el 1ro de octubre del 2008 de 114 peticiones por minuto.

- En el MegaInfocentro que se encuentra en el Ministerio de Ciencia y Tecnología en la República Bolivariana de Venezuela (ver Anexo 22), con alrededor de 100 estaciones clientes; en este caso se les brindó capacitación a los especialistas venezolanos para que ellos pudieran instalar FILPACON en el resto de los Infocentros del país.
- Se encuentra instalado desde enero del 2008 en el Centro UCID y le brinda servicios a más de 500 usuarios (ver Anexo 21).
- Premio del Rector 2009 en la categoría de Mejor Producto (ver Anexo 23).

3.3. Elementos de novedad

La novedad de este sistema puede verse fundamentalmente desde tres aristas basales: el primero radica en que las soluciones más completas a nivel internacional que permiten regular el acceso a Internet, utilizan modelos de licencias privativas y por otro lado tienen precios muy elevados por lo cual la posibilidad de hacerles adaptaciones a nuestras particularidades se complican sobremanera y su posibilidad de generalización a nivel nacional se convierte en económicamente inviable en las condiciones actuales de nuestro país; el segundo elemento radica en que utilizar aplicaciones libres nos restringe a muy pocas aplicaciones de este tipo y todas bajo licencias GPL, esta licencia impone un efecto contaminante al obligarnos a licenciar también con GPL todo trabajo derivado de esta, imposibilitando desarrollar al máximo el potencial comercial de un producto de este tipo, en cambio reutilizando componentes y librerías que usan modelos de licencia completamente

libres como la MIT permitió el desarrollo de un producto que es propiedad de la UCI y que por tanto puede ser licenciado bajo el modelo que la institución estime conveniente, además los productos existentes hasta la fecha con licencias libres no satisfacen las expectativas de requerimientos funcionales que nuestro país necesita para regular el acceso a Internet. La última arista es el aporte del sistema al proceso de toma de decisiones, así por ejemplo:

- al obtener el reporte de URL denegadas en un período de tiempo podemos averiguar cuales sitios de los no permitidos son visitados con mayor asiduidad,
- al obtener el listado ordenado de usuarios por cantidad de incidencias en un período podemos averiguar cuales usuarios no están realizando un buen uso de la Internet de acorde a las políticas de la organización,
- con el listado de incidencias por días podemos averiguar cuales días la Internet se utiliza con otros fines no acordes a las políticas de la organización,
- con el listado ordenado de los reportes de usuarios de posibles denegaciones incorrectas el administrador del sistema puede ajustar aun más la efectividad del sistema.

Conclusiones

En este trabajo se presentó la concepción de la arquitectura y funcionalidades fundamentales del producto FILPACON en su versión 1.0. Proponiendo una solución viable a la problemática de cómo regular el acceso a Internet. En todo momento la aplicación que se desarrolló utilizó herramientas y aplicaciones pertenecientes a la categoría de Software Libre, sin llegar en ningún caso a modificar el código de ninguna de ellas evitando de esta manera el efecto contaminante de algunas de estas licencias, esto posibilita que pueda ser liberado bajo el modelo de licencia que la Universidad estime conveniente. El sistema hasta su versión actual tiene un elevado nivel de flexibilidad demostrado en la posibilidad de configurar el mismo producto en disímiles entornos donde ha sido instalado con buenos resultados como se presentó en el último capítulo de este trabajo, dispone de una interfaz gráfica para la administración de las funcionalidades que brinda facilitando de esta forma una mejor administración de las políticas de uso de Internet, por otro lado el sistema desde su concepción posibilita que se pueda instalar en varios sistemas operativos como pudieran ser los de la familia GNU/Linux, FreeBSD, incluso hasta en Windows pues todos sus componentes tienen versiones para estos sistemas operativos, en la implementación actual el instalador soporta 11 arquitecturas de Hardware desde una interfaz gráfica muy cómoda e intuitiva, así el sistema puede ser empaquetado y distribuido en un medio de almacenamiento extraíble. El subsistema de reportes estadísticos que provee el sistema genera información valiosa para el proceso de toma de decisiones y permite la incorporación paulatina de nuevos reportes. Por otro lado debido a su concepción en forma de producto genérico pudo ser comercializado ingresando una importante suma a la economía nacional, además en el estado actual puede seguir comercializándose sin mucho esfuerzo adicional en el desarrollo de nuevas funcionalidades, además está pendiente a ser instalado en el Ministerio de la Informática y las Comunicaciones de Cuba como próximo paso para su generalización. De esta manera se dan por cumplidos los objetivos planteados inicialmente.

Recomendaciones

El producto FILPACON tiene una versión estable y probada, la 1.0, pero esto no significa que sea un producto terminado ni siquiera a nivel conceptual, a continuación se brindan un conjunto de puntos que nos indican que falta mucho por hacer todavía:

- Implementar un instalador genérico que sea completamente independiente del Sistema Operativo.
- Concebir e implementar la posibilidad de analizar videos, sonido, pdf, documentos de suites ofimáticas y otros archivos de uso común.
- Implementar la posibilidad de gestión de grupos de usuarios.
- Incorporar el Subsistema de Recuperación y Clasificación Inteligente (SRCI) al sistema.
- Estudiar e implementar elementos del SRCI a nivel de Hardware.
- Implementar la internacionalización de la interfaz de administración.
- Implementar las configuraciones de integración a dominios y de las políticas de navegación horarias desde la interfaz de administración.
- Implementar las funcionalidades que permitan actualizar la base de datos de manera remota.
- Definir un modelo de negocio competitivo para el producto.
- Definir un modelo de licencia internacional para el producto.
- Continuar el proceso de generalización del producto en instituciones educacionales del país con acceso a Internet.
- Realizar campañas para posicionarse en determinados sectores del mercado internacional.

Bibliografía

AlphaStore. 2009. AlphaStore. [Online] 2009. [Cited: abril 1, 2009.] http://www.alphastore.com.au/catalogue/?tier4=282244.

Barron, Daniel. 2009. DansGuardian - True Web Content Filtering for All. *History (and future plans).* [Online] January 21, 2009. [Cited: abril 1, 2009.] http://dansguardian.org/?page=history.

—. 2007. DansGuardian - True Web Content Filtering for All. *Who Uses DansGuardian?* [Online] November 22, 2007. [Cited: abril 1, 2009.] http://dansguardian.org/?page=whousesit.

Boletín Oficial 247 Noviembre. **OCPI. 2008.** 247, s.l. : Oficina Cubana de la Propiedad Industrial, 2008, Vol. CVII. ISSN 1028-1452.

Casacuberta, David. 1998. Presidente de Fronteras Electrónicas España, comparecencia ante el Senado, Comisión especial sobre redes informáticas. [En línea] 16 de junio de 1998. [Citado el: 25 de marzo de 2009.] http://www.senado.es/legis6/expedientes/index_715000230.html.

Ceballo Gastell, Daimerys y de Diego Ceruto, Yanet del Carmen. 2008. *Categorización de texto usando Redes Neuronales Artificiales.* Ciudad de la Habana : Universidad de las Ciencias Informáticas, 2008.

Cibercensura. 2003. Pimienta negra, censurada en Internet. *Rebelión.* [Online] marzo 12, 2003. [Cited: marzo 28, 2009.] http://www.rebelion.org/hemeroteca/cibercensura/pimientanegra120303.htm.

Comisión Europea. 2002. Poesia Project. [Online] Safer Internet Action Plan, 2002. [Cited: abril 1, 2009.] http://www.poesia-filter.org/.

ContentWatch. 2008. ContentProtect™ Professional Internet Filtering For Business. [Online] ContentWatch Inc., 2008. [Cited: marzo 31, 2009.] http://www.contentwatch.com/products/contentprotect_pro/detail/technical#it.

—. 2008. ContentProtect™ Professional Suite For Business. [Online] ContentWatch, Inc., 2008. [Cited: marzo 31, 2009.] http://www.contentwatch.com/products/contentprotect_pro_suite/detail/technical.

—. 2008. Home use license agreement and limited warranty. *Net Nanny License Agreement.* [Online] ContentWatch Inc., 2008. [Cited: marzo 31, 2009.] http://www.netnanny.com/eula/netnanny.

—. 2009. Internet Filter, Parental Control & Filter Software | Net Nanny. [Online] ContentWatch, Inc., 2009. [Cited: marzo 30, 2009.] http://www.netnanny.com/.

DCMI. 2009. About the Dublin Core Metadata Initiative. [Online] Dublin Core Metadata Initiative, January 5, 2009. [Cited: marzo 30, 2009.] http://dublincore.org/about/.

Debian. 2009. Debian -- The Universal Operating System. [Online] Debian is a registered trademark of Software in the Public Interest, Inc. , April 6, 2009. [Cited: abril 8, 2009.] http://www.debian.org/.

Deibert, Ronald, et al. 2008. *Access Denied: The Practice and Policy of Global Internet Filtering.* s.l. : Cambridge: MIT Press, 2008. ISBN-13:978-0-262-54196-1.

Donert, Karl and Carro, Sara Martinez. 2002. End-User Requirements: Final Report Deliverable 2.1. *POESIA-WP2-2.1.* [Online] December 23, 2002. [Cited: abril 1, 2009.] http://www.poesia-filter.org/pdf/Deliverable_2_1.pdf.

Editorial Board. 2008. *Internet filtering technology.* [En línea] Official Website of the Australian Labor Party, 29 de julio de 2008. [Citado el: 26 de marzo de 2009.] http://www.abc.net.au/news/stories/2007/12/31/2129471.htm.

Federal Communications Commission. 1996. Telecommunications Act of 1996. [Online] 1996. [Cited: marzo 26, 2009.] http://www.fcc.gov/Reports/tcom1996.txt.

FOSI. 2007. ICRA Tools. *Family Online Safety Institute.* [Online] FOSI.org, 2007. [Cited: marzo 30, 2009.] http://www.fosi.org/icra/.

Gaceta Federal. 2004. Ley de Telecomunicaciones de Alemania. [Online] junio 25, 2004. [Cited: marzo 26, 2009.] http://217.160.60.235/BGBL/bgbl1f/bgbl104s1190.pdf.

Gonzalo, Ibán Pravos. 2008. Optenet: Optimal Internet - Buy Optenet Security Suite PC: General conditions. [Online] 2008. [Cited: marzo 30, 2009.] https://seguro.optenet.com/shop/impresion.asp.

Hauben, Ronda. 2004. The Internet: On its International Origins and Collaborative Vision (A Work In Progress). *The Amateur Computerist Newsletter.* [En línea] 2004. [Citado el: 25 de marzo de 2009.] http://www.ais.org/~jrh/acn/ACn12-2.a03.txt.

Hermosilla Moreno, José Ramón y Sánchez Arce, Luis Enrique. 2008. *Interfaz de Administración Web para el Sistema de Filtrado Filpacon.* Ciudad de la Habana : Universidad de las Ciencias Informáticas, 2008.

Hernández Moya, Yurisleidy y Cardoso Carmona, Dionny. 2008. *Identificador Automatizado de Idiomas para Textos.* Ciudad de la Habana : Universidad de las Ciencias Informáticas, 2008.

IBM Corp. 2009. IBM - Proventia Web Filter. [Online] IBM Corporation - Internet Security Systems, 2009. [Cited: abril 1, 2009.] http://www-935.ibm.com/services/us/index.wss/offering/iss/a1027244.

Internet Security Systems. 2007. Proventía Web Filter. Getting Started Guide. [Online] April 2007. [Cited: abril 1, 2009.] http://documents.iss.net/literature/proventia/pvwebfilter_gs_143.pdf.

InternetSafety.com. 2009. About Internet Safety - Internet Safety Company Information. [Online] InternetSafety, Inc., 2009. [Cited: marzo 31, 2009.] http://www.internetsafety.com/about-internet-safety.php.

—. 2009. Internet Safety Software at InternetSafety.com. [Online] InternetSafety.com, Inc, 2009. [Cited: marzo 31, 2009.] http://www.internetsafety.com/.

—. 2009. Safe Eyes Software License Agreement. [Online] 2009. [Cited: marzo 31, 2009.] https://secure.parentalctrl.com/signup/softwarelicenseagreement.pdf.

Isabel Guerriero, Gabriela and Mongiardino, Marina. 2002. Aspectos de la responsabilidad jurídica de los proveedores de servicios y contenidos de Internet. *VII Congreso Internacional de Derecho de Daños y Responsabilidades en el Siglo XXI.* [Online] octubre 2002. [Cited: marzo 26, 2009.] http://www.aaba.org.ar/bi20op14.htm.

Kitagawa, Kazuhiro, Hjelm, Johan and Hagino, Tatsuya. 2001. Mobile Access Activity Statement. [Online] World Wide Web Consortium, July 3, 2001. [Cited: marzo 30, 2009.] http://www.w3.org/Mobile/Activity.

Lage Dávila, Carlos. 2000. Acuerdo No. 3736. *Ministerio de la Informática y las Comunicaciones.* [En línea] 18 de julio de 2000. [Citado el: 4 de abril de 2009.] http://www.cubagob.cu/ingles/des_eco/mic/mic_regulaciones/decretos/general/acuerdo_3736%20.htm.

Manola, Frank and Miller, Eric. 2004. RDF Primer. *W3C Recommendation.* [Online] World Wide Web Consortium, February 10, 2004. [Cited: marzo 29, 2009.] http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

Netcraft Ltd, 2009. Netcraft. [En línea] [Citado el: 25 de marzo de 2009.] http://www.netcraft.com/.

OpenNet. 2007. About Filtering. [Online] OpenNet Initiative, 2007. [Cited: marzo 30, 2009.] http://opennet.net/about-filtering.

OPTENET. 2008. Offices. *Optenet: Get Optimal Internet.* [Online] OPTENET, 2008. [Cited: marzo 30, 2009.] http://www.optenet.com/en-us/offices.asp.

—. 2009. Optenet Solutions for the Enterprise. *Intelligent Content Security for the Enterprise.* [Online] 2009. [Cited: marzo 30, 2009.] http://www.optenet.com/en-us/pdf/PSEN09_Enterprise.pdf.

—. 2008. Optenet: Get Optimal Internet - Products. [Online] OPTENET, 2008. [Cited: marzo 30, 2009.] http://www.optenet.com/en-us/products.asp.

—. 2008. Optenet: Get Optimal Internet - Enterprise Web and Mail Protection | Security as a Service for Service Providers. *Enterprise Web and Mail Protection* | *Security as a Service for Service Providers.* [Online] OPTENET, 2008. [Cited: marzo 30, 2009.] http://www.optenet.com/en-us/index.asp.

—. 2008. Optenet: Get Optimal Internet - Partners - Sales Partners. *Partners - Sales Partners.* [Online] 2008. [Cited: marzo 30, 2009.] http://www.optenet.com/en-us/comercialpartners.asp.

Pérez Clemente, Yandry y Ripoll Méndez, Dovier Antonio. 2008. *Asignación automatizada de categorías temáticas al contenido textual de documentos HTML.* Ciudad de la Habana : Universidad de las Ciencias Informáticas, 2008.

Reagle, Joseph M. 2002. XML Digital Signatures Activity Statement. [Online] World Wide Web Consortium, December 31, 2002. [Cited: marzo 30, 2009.] http://www.w3.org/Signature/Activity.html.

Redacción de Noticiasdot.com. 2007. Noticiasdot.com. *Noticiasdot.com es un producto de Noticias Digitales S.L.* [En línea] 14 de agosto de 2007. [Citado el: 26 de marzo de 2009.] http://www.noticiasdot.com/wp2/2007/08/14/cruzada-moral-contra-la-pornografa-online/.

Resnick, Paul. 1997. *Filtering Information On The Internet.* s.l. : Scientific American Magazine, 1997.

—. 1997. *Filtering Information On The Internet.* s.l. : Scientific American Magazine, 1997.

—. 1999. PICS, Censorship, & Intellectual Freedom FAQ. *World Wide Web Consourtion.* [Online] agosto 4, 1999. [Cited: marzo 28, 2009.] http://www.w3.org/PICS/PICS-FAQ-980126.html.

Ropelato, Jerry. 2006. Internet Pornography Statistics. [Online] 2006. [Cited: marzo 26, 2009.] http://internet-filter-review.toptenreviews.com/internet-pornography-statistics.html.

Rubenking, Neil J. 2008. Net Nanny 6.0. *PC Magazine.* [Online] Ziff Davis Publishing Holdings Inc., November 24, 2008. [Cited: marzo 31, 2009.] http://www.pcmag.com/article2/0,2817,2335480,00.asp.

—. 2006. Safe Eyes 2006. *PC Magazine.* [Online] Ziff Davis Publishing Holdings Inc., April 4, 2006. [Cited: marzo 31, 2009.] http://www.pcmag.com/article2/0,2817,1945995,00.asp.

SafeSurf. 2007. SafeSurf Rating System. [Online] 2007. [Cited: marzo 30, 2009.] http://www.safesurf.com/classify/.

Solid Oak Software. 2008. CYBERsitter Official Web Site #1 Internet Filter. [Online] Solid Oak Software, Inc., 2008. [Cited: marzo 31, 2009.] http://www.cybersitter.com/.

Swick, Ralph. 1997. Metadata Activity Statement. [Online] World Wide Web Consortium, December 1997. [Cited: marzo 29, 2009.] http://www.w3.org/Metadata/Activity.

TopTenREVIEWS. 2009. 2009 Internet Filter Software Product Comparisons. [Online] TopTenREVIEWS, Inc, 2009. [Cited: marzo 30, 2009.] http://www.internet-filter-review.toptenreviews.com/.

Villate, Javier. 2001. Libertad de expresión en Internet. *ARCHIVO del Observatorio para la CiberSociedad.* [En línea] 2001. [Citado el: 26 de marzo de 2009.] http://www.cibersociedad.net/archivo/articulo.php?art=37.

—. 1997. Preguntas frecuentes sobre los filtros de contenido. [En línea] 1997. [Citado el: 5 de octubre de 2007.] http://peru.cpsr.org/filtro/faq.

W3C. 2005. Platform for Internet Content Selection (PICS). [Online] World Wide Web Consortium, julio 8, 2005. [Cited: marzo 28, 2009.] http://www.w3.org/PICS/.

Websense. 2009. Web Filtering, Web Security, Filtering Software - About - Websense, Inc. *Websense office locations and phone numbers.* [Online] Websense, Inc., 2009. [Cited: abril 1, 2009.] http://www.websense.com/global/es/AboutWebsense/ContactUs/index.php.

—. 2009. Websense - Products. [Online] Websense, Inc., 2009. [Cited: abril 1, 2009.] http://www.websense.com/content/Products.aspx.

—. 2009. Websense Web Filter (Websense Enterprise) | Web Filtering and Web Security | GuardSense.com. *List Price.* [Online] Websense, Inc., 2009. [Cited: abril 1, 2009.] http://www.guardsense.com/Websense-Web-Filter.asp.

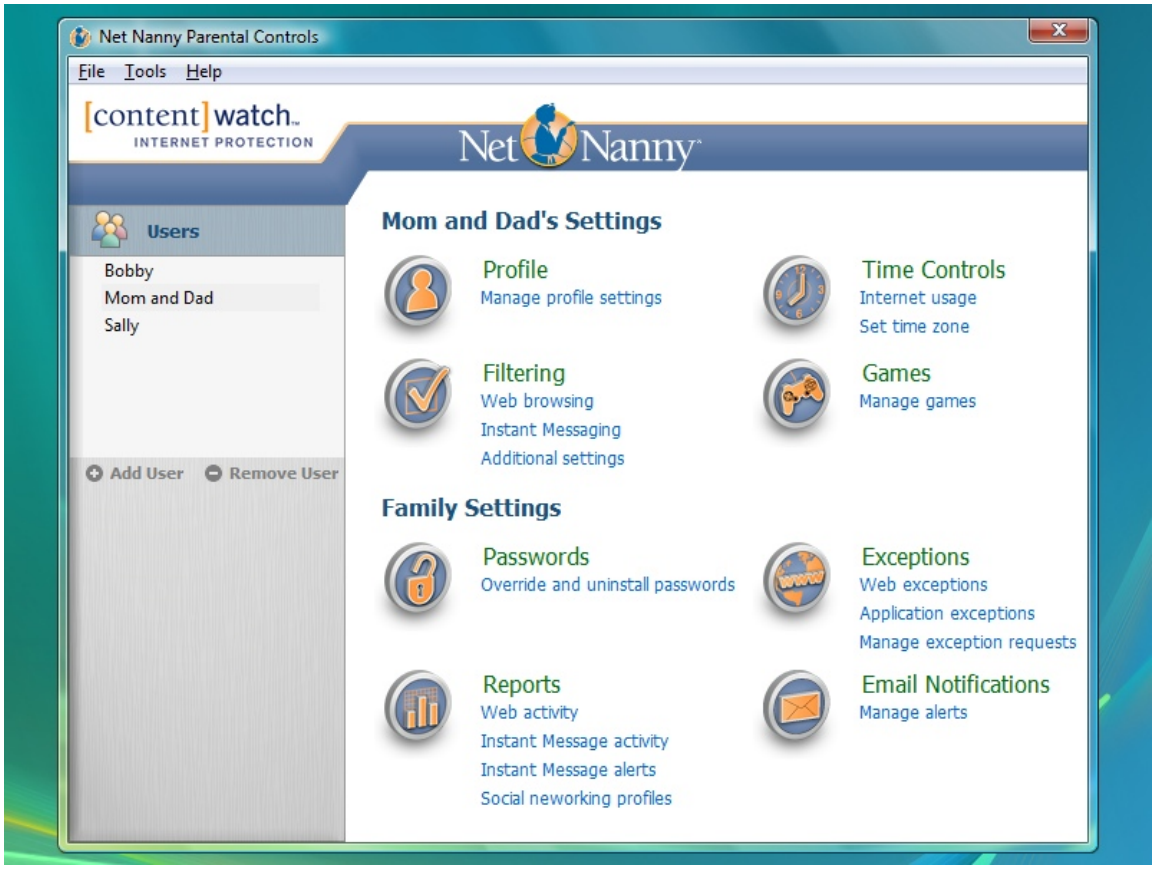
—. 2009. Websense.com. [Online] Websense, Inc., 2009. [Cited: abril 1, 2009.] http://www.websense.com/.

—. 2009. Websense.com - About Us. [Online] Websense, Inc., 2009. [Cited: abril 1, 2009.] http://www.websense.com/site/aboutus/index.html.

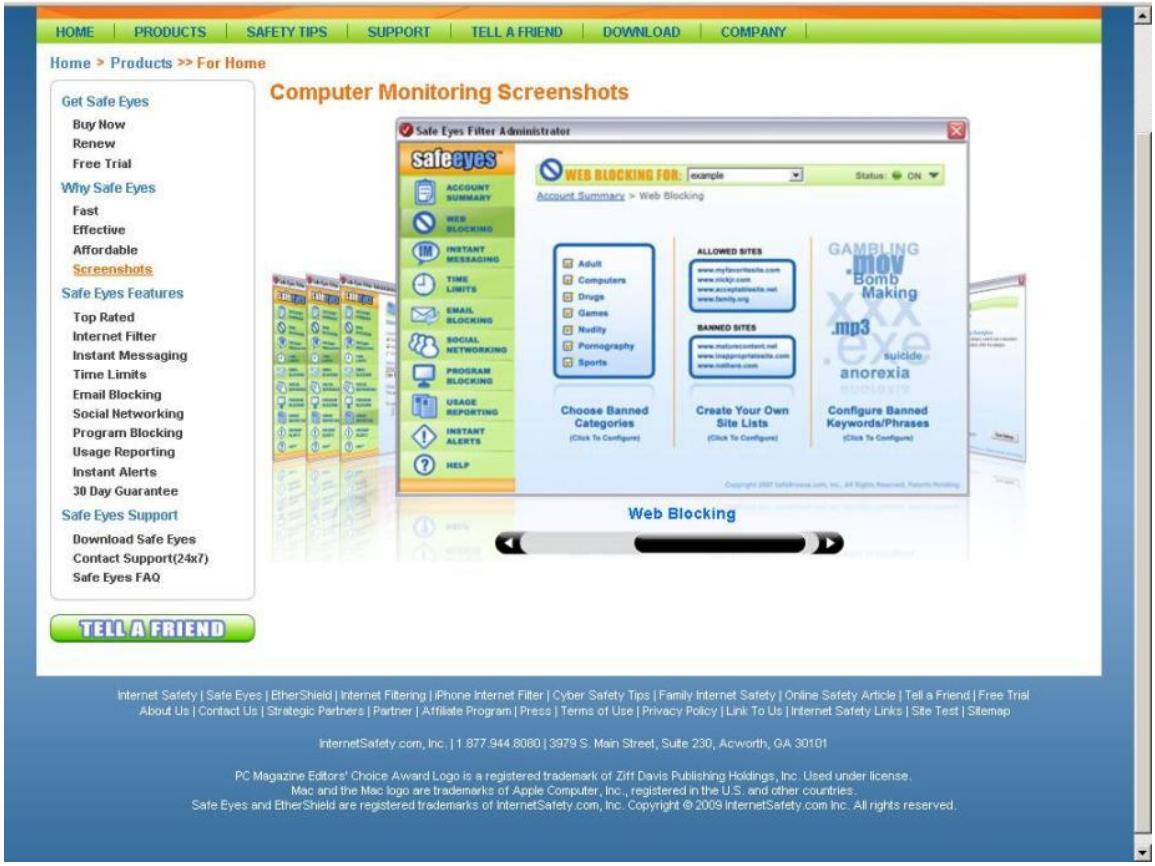
Wenning, Rigo. 2008. Privacy Activity Statements. [Online] World Wide Web Consortium, October 2008. [Cited: marzo 30, 2009.] http://www.w3.org/Privacy/Activity.html.

Anexos

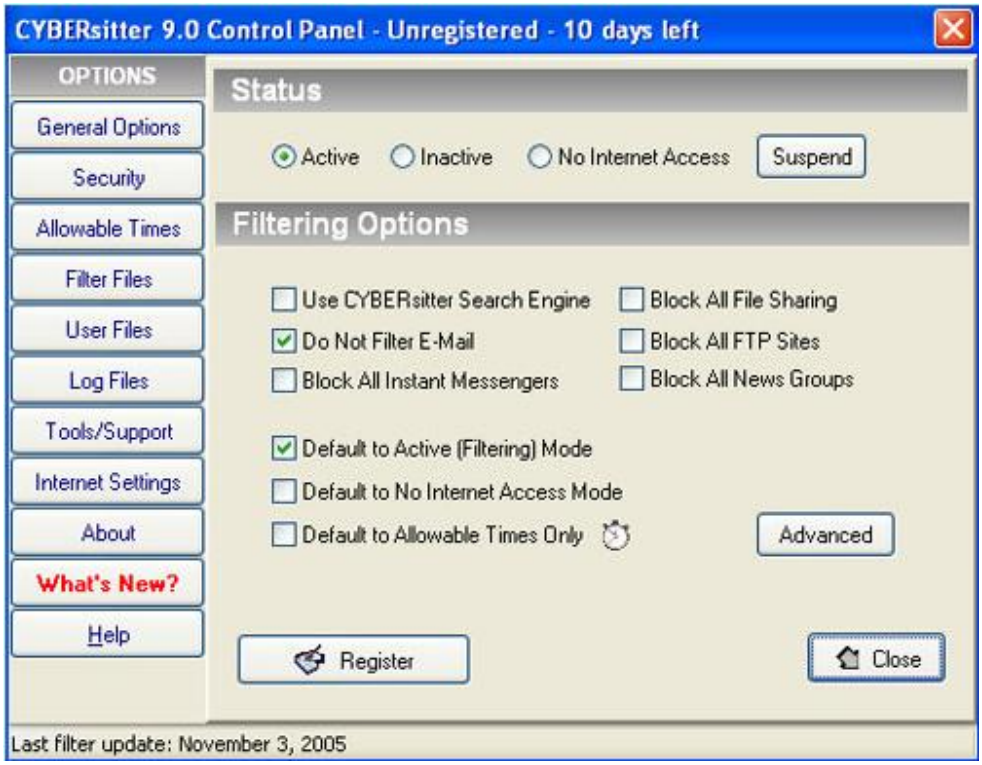
Anexo 1. NetNanny en su versión 6.0



Anexo 2. SafeEyes de InternetSafety

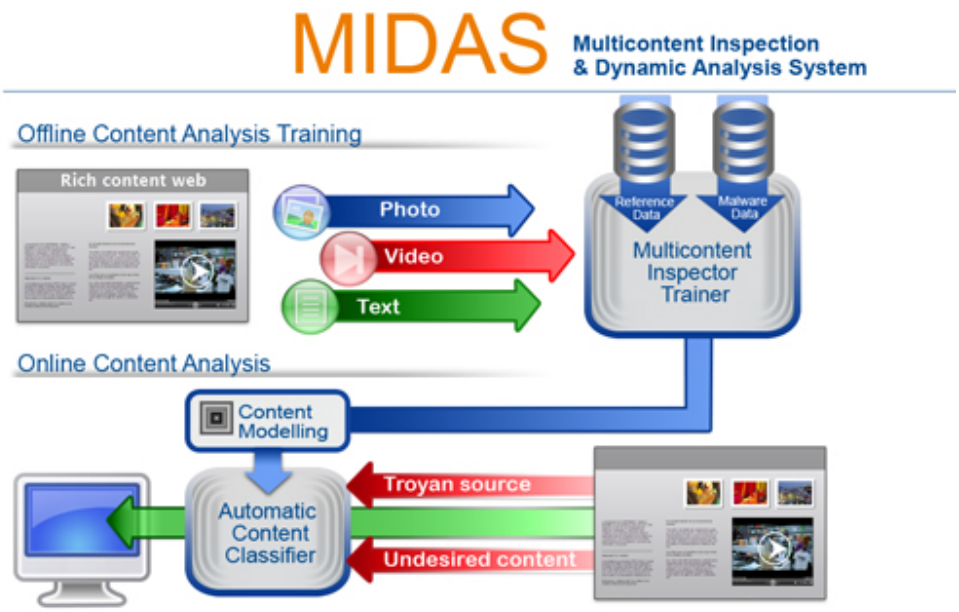


Anexo 3. CYBERsitter de Solid Oak Software

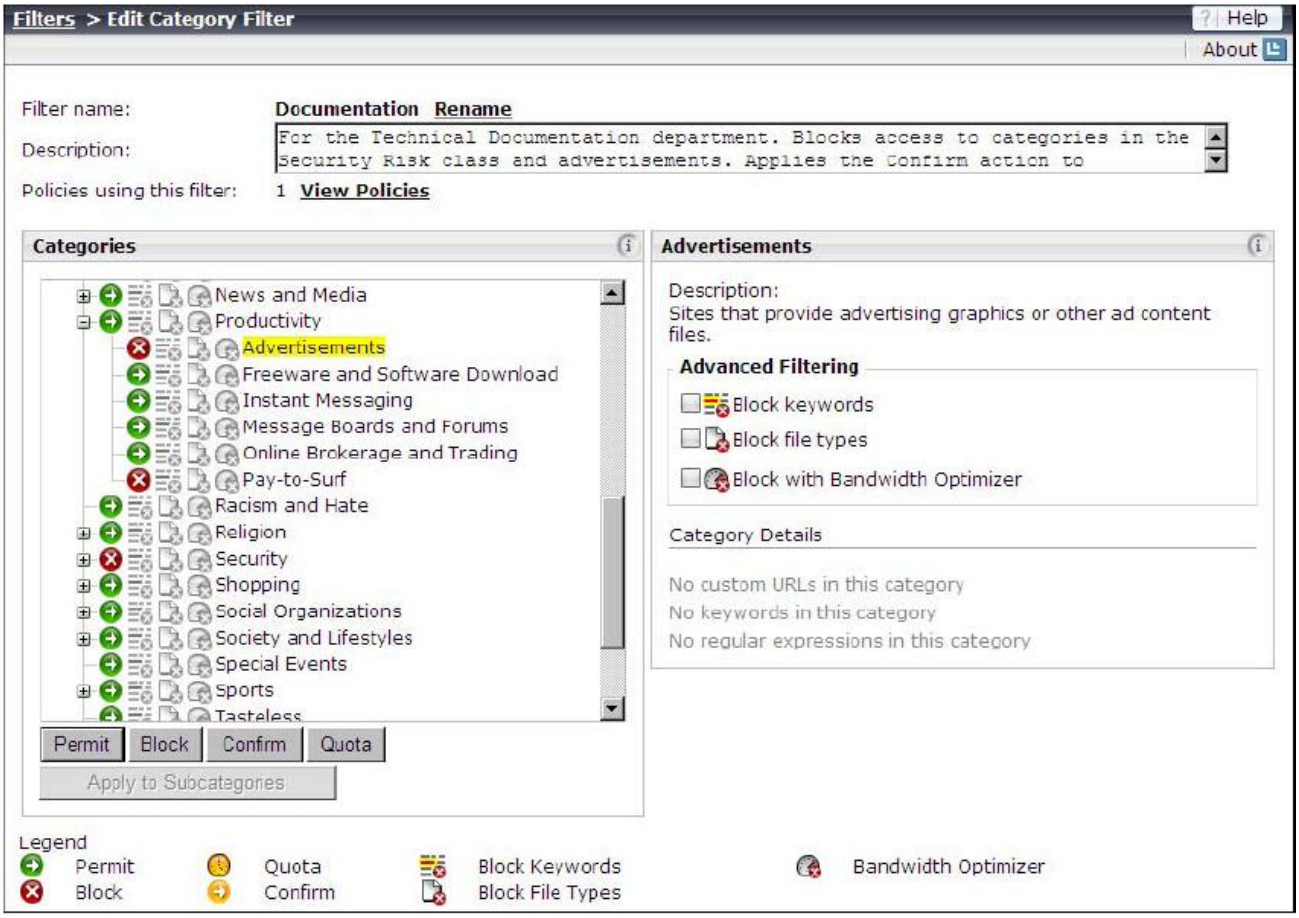


Anexo 4. Solución Enterprise Optenet NetSecure

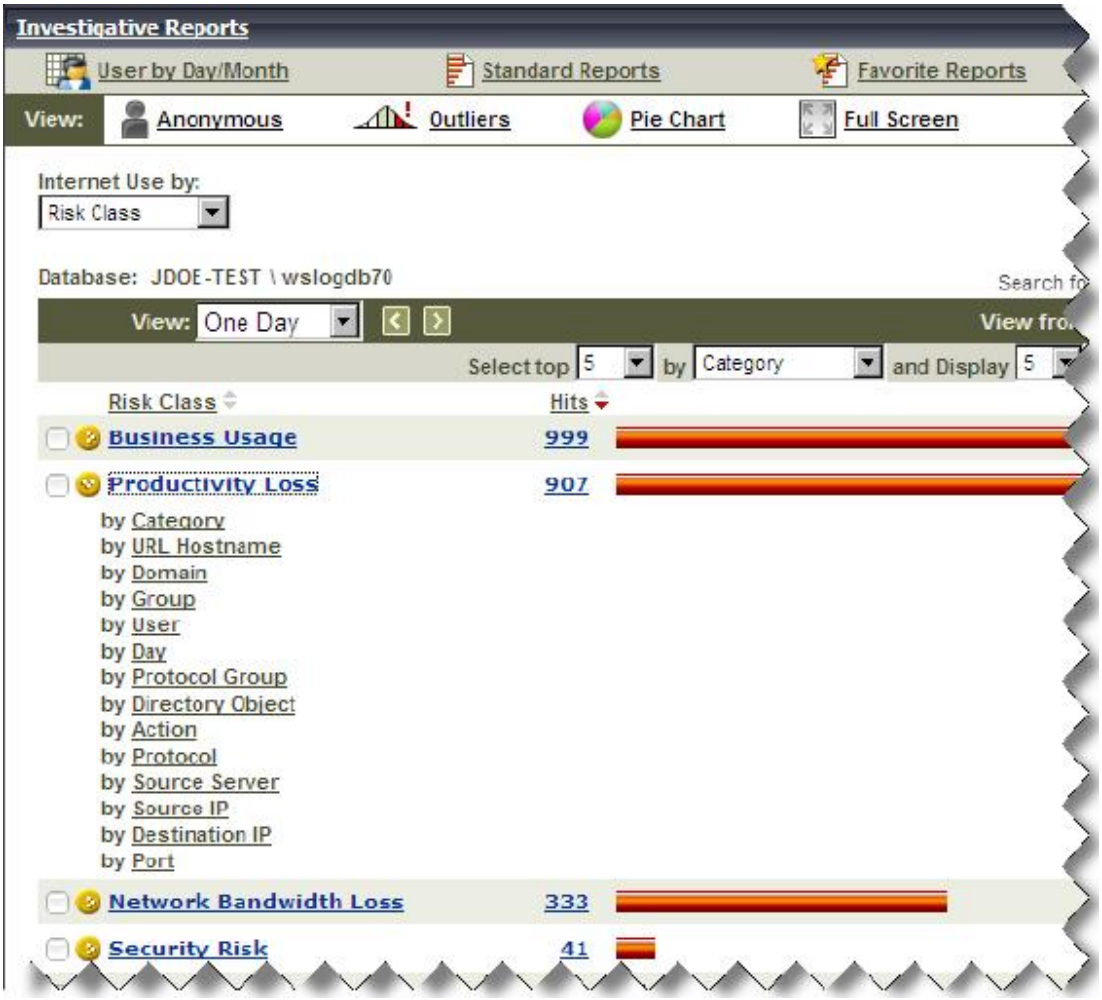
Anexo 5. Multicontent Inspection and Dynamic Analysis System



Anexo 6. Websense Web Filter - Categorías



Anexo 7. Websense Web Filter - Reportes



Anexo 8. Certifico de registro, del producto FILPACON v1.0 y la marca FILPACON, emitido por la Dirección de Servicios Legales.

UCI Universidad de las Ciencias Informáticas | Dirección de Servicios Legales DSL-C-013/2008

En uso de las atribuciones que me están conferidas emito el siguiente:

CERTIFICO

PRIMERO: Que el Producto de Software **Filtrado de Paquetes por Contenidos (FILPACON) v.1.0.** fue objeto de **Registro** en el Centro Nacional de Derecho de Autor de Cuba (CENDA) con **Número de Registro: 3421-2008.**

SEGUNDO: Que la marca **FILPACON** fue presentada para su Registro en la Oficina Cubana de la Propiedad Industrial (OCPI) con **Número de Solicitud: 2008-0303.**

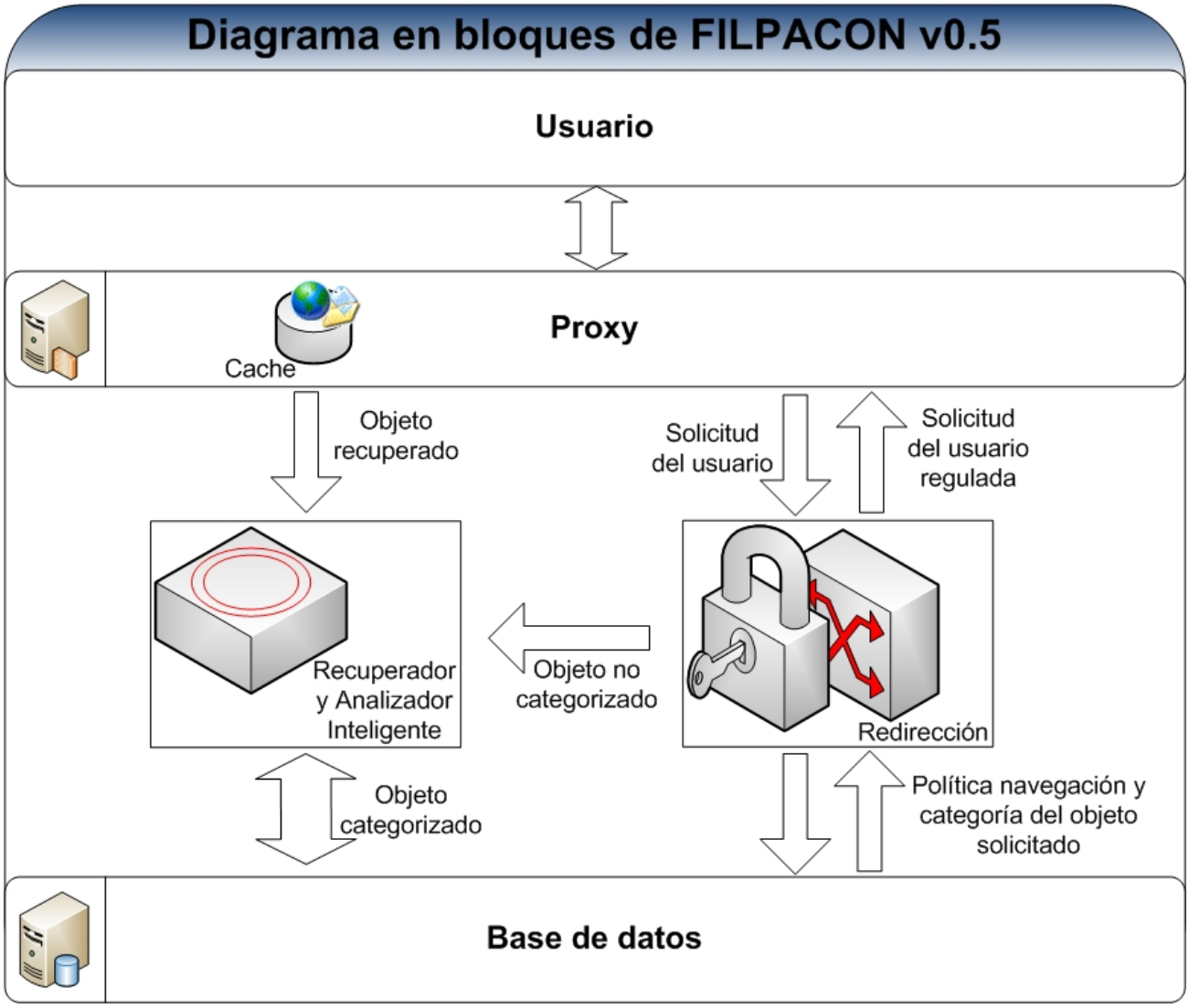
Y para su protocolización, se emite la presente documento, en la Dirección de Servicios Legales de la Infraestructura Productiva de la Universidad de las Ciencias Informáticas, en la Ciudad de La Habana, a los 10 días del mes de diciembre de 2008.

UCI SERVICIOS LEGALES INFRAESTRUCTURA PRODUCTIVA

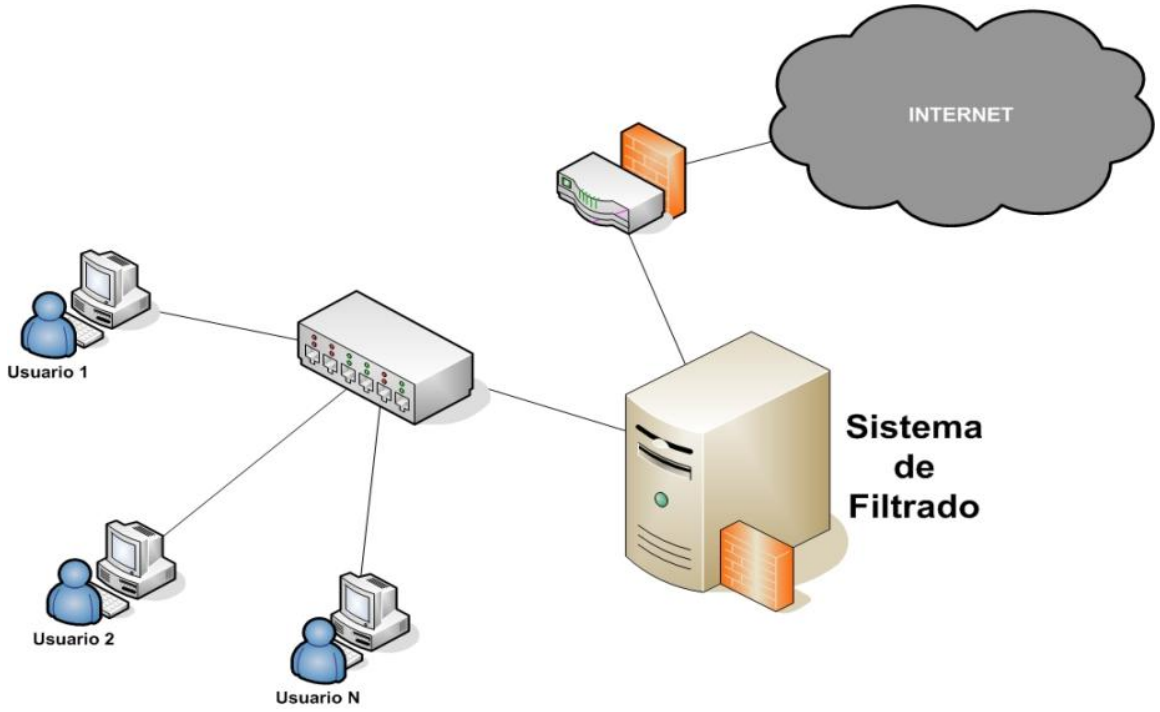
Lic. Ignacio Pérez Valdés.
Jefe de Grupo de Propiedad Intelectual

Universidad de las Ciencias Informáticas.
Carretera San Antonio de los Baños Km. 2½, Torrens, Municipio Boyeros, Ciudad de La Habana, Cuba.
Teléfono: +53 (7) 837 2402 E-mail: ip.dsl@uci.cu

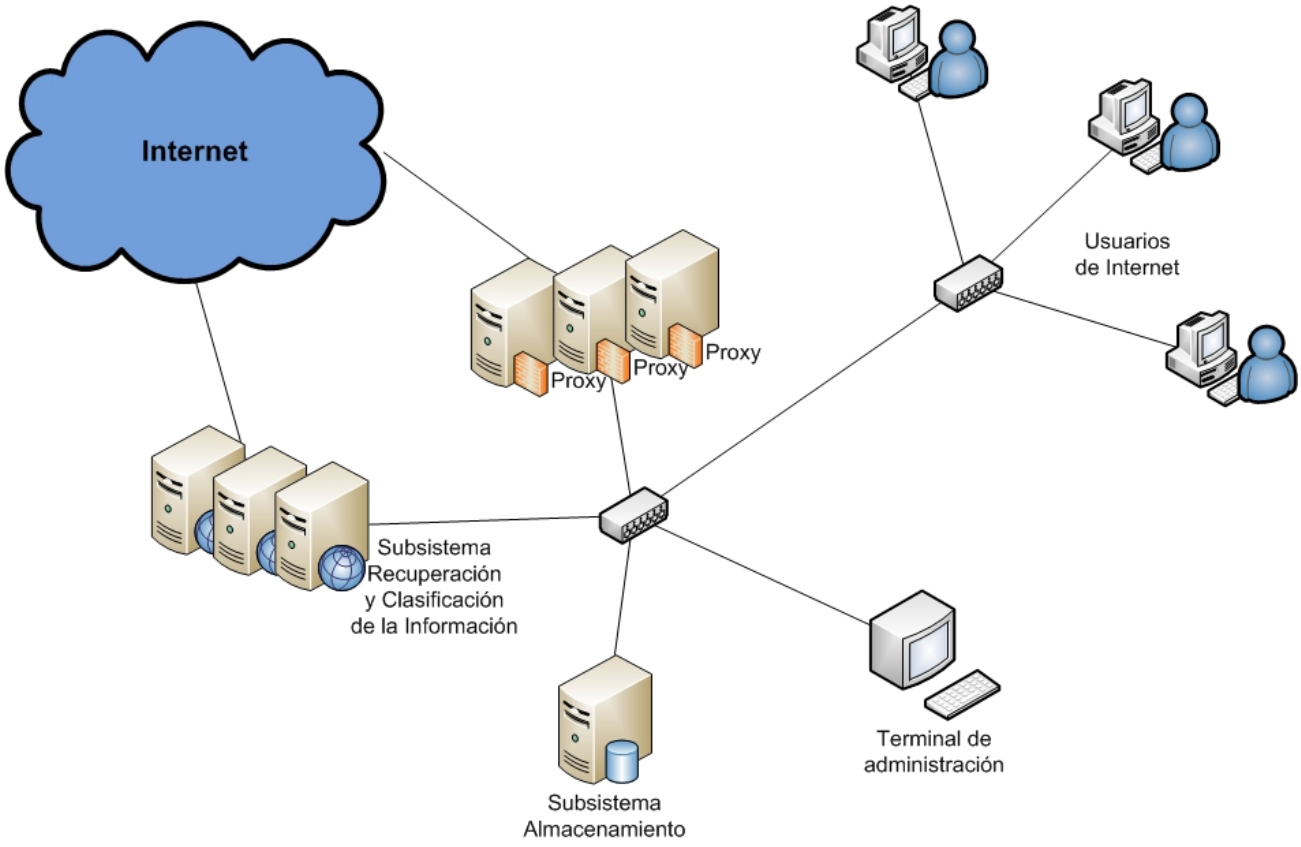
Anexo 9. Concepción General de la Arquitectura (FILPACON v0.5)



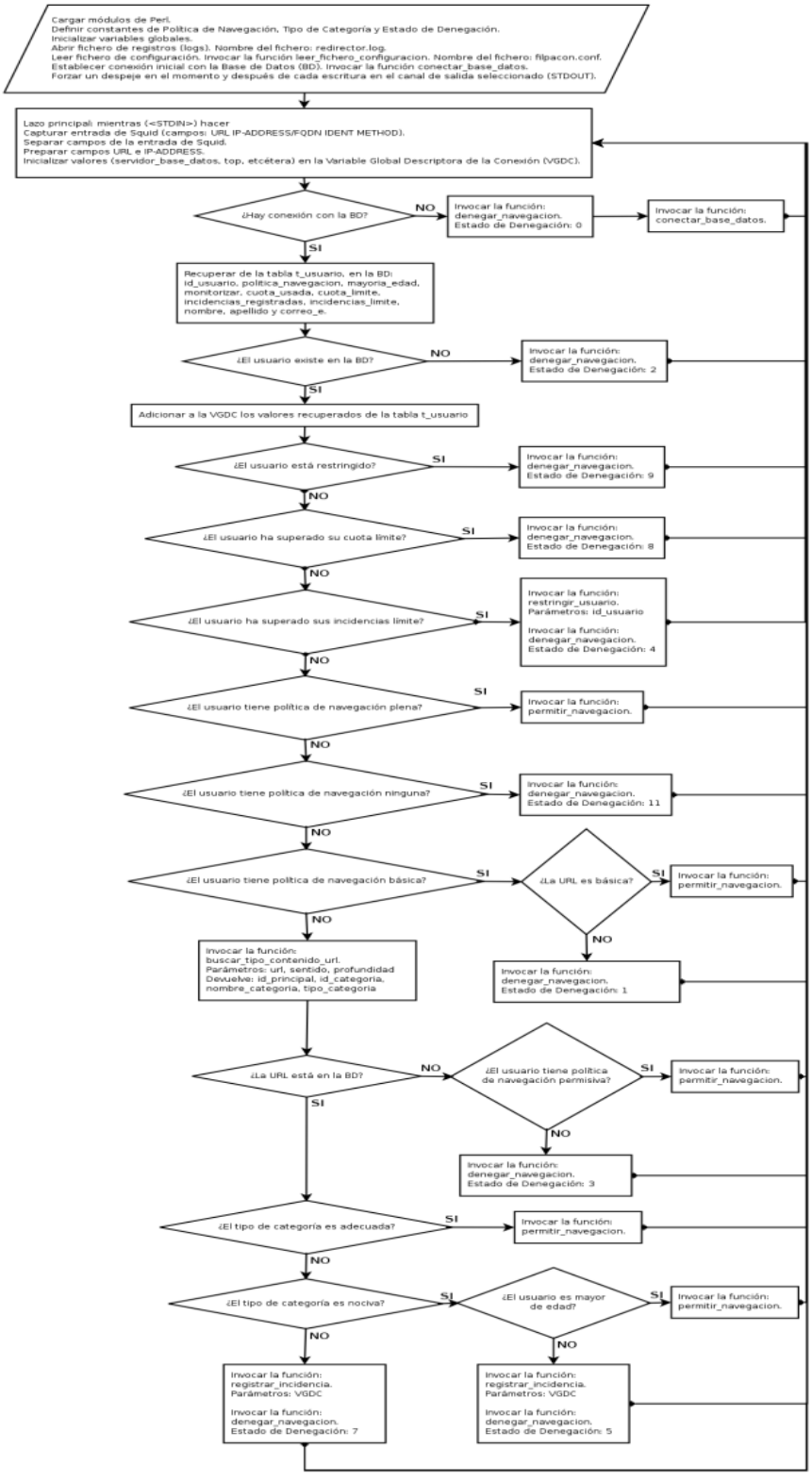
Anexo 10. Concepción para un despliegue de baja demanda (FILPACON v0.5)



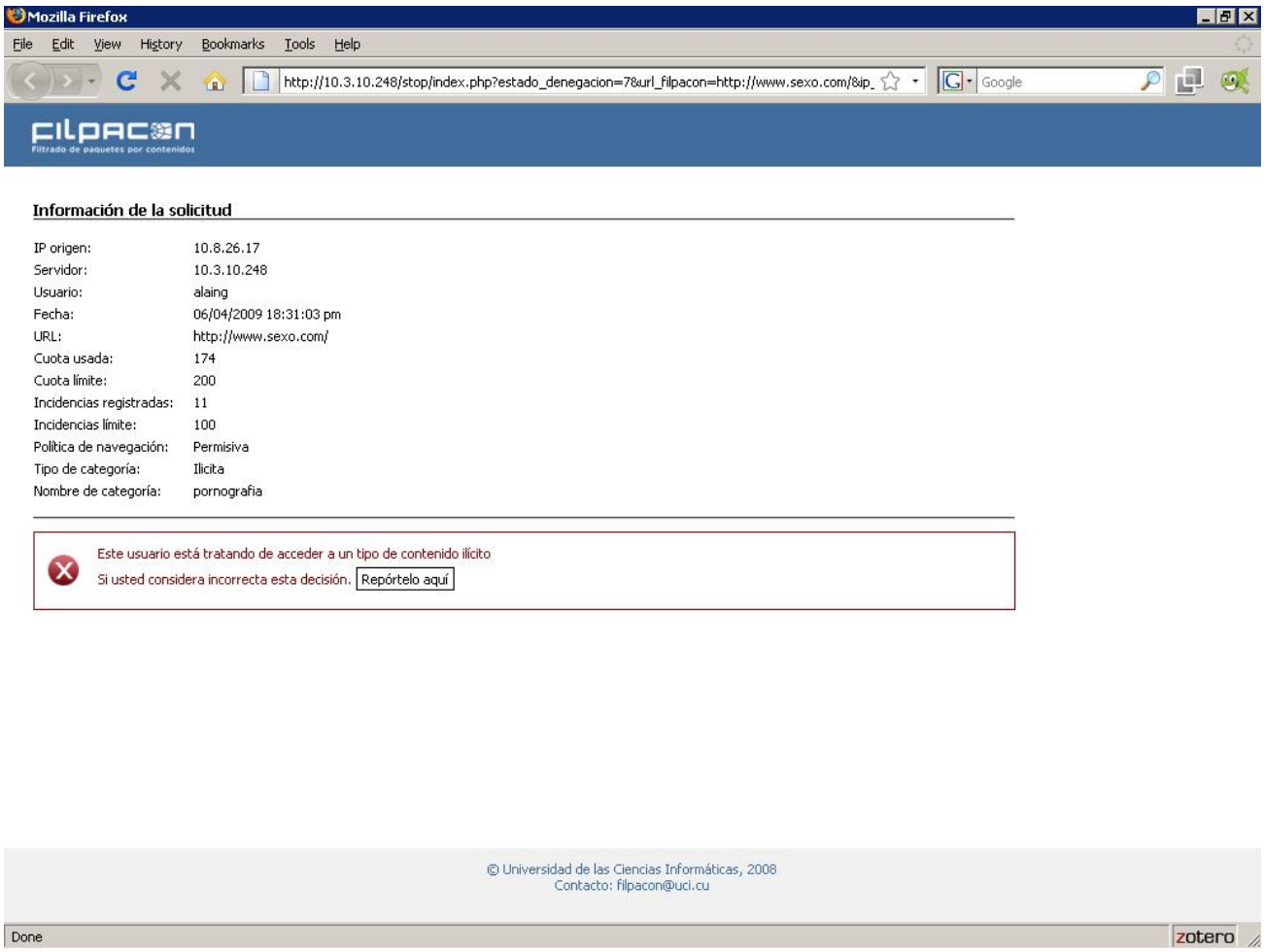
Anexo 11. Concepción para un despliegue de alta demanda con posibilidad de escalabilidad (FILPACON v0.5)



Anexo 12. Diagrama en bloques del algoritmo de filtrado

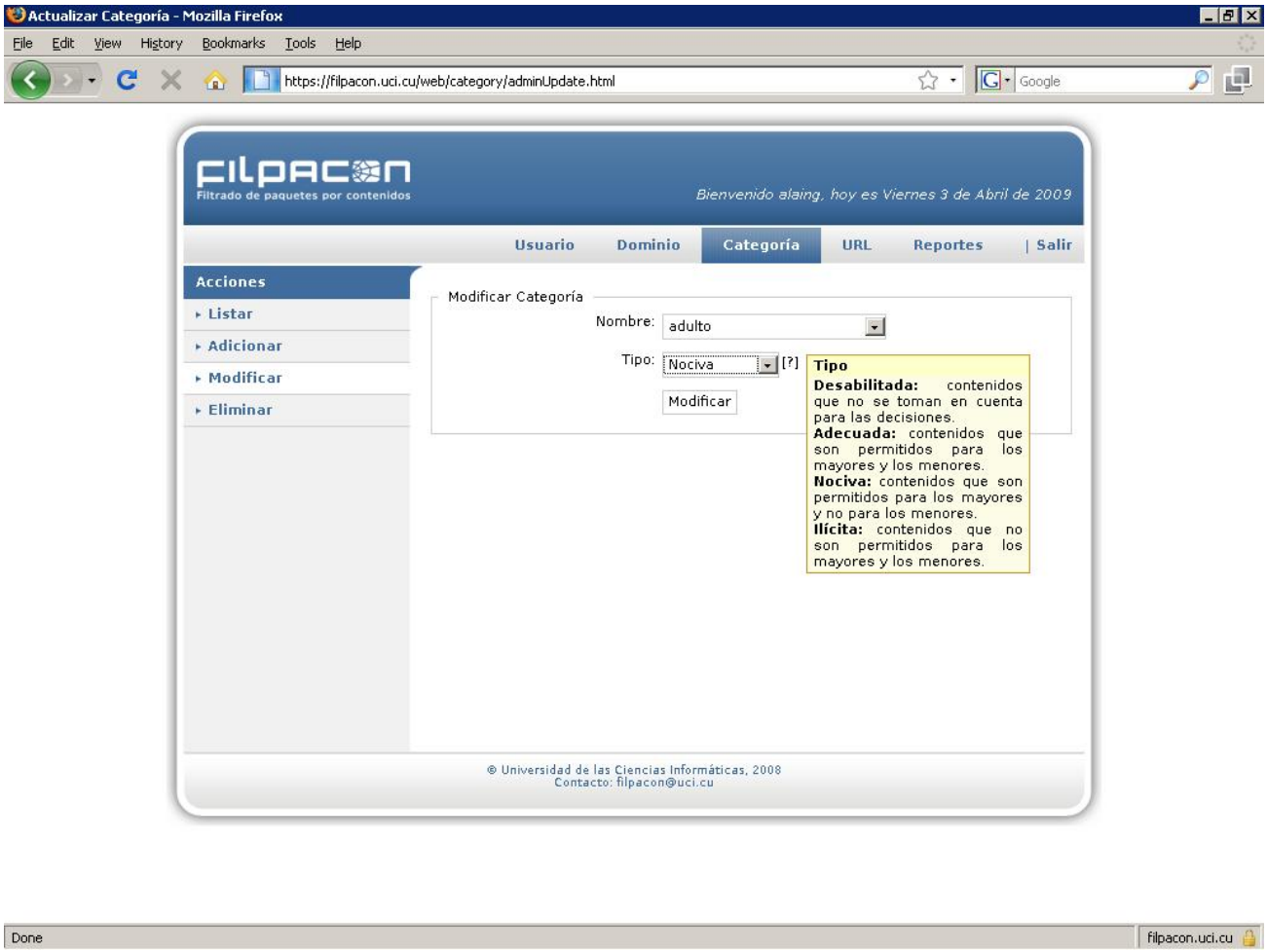


Anexo 13. Página de denegación

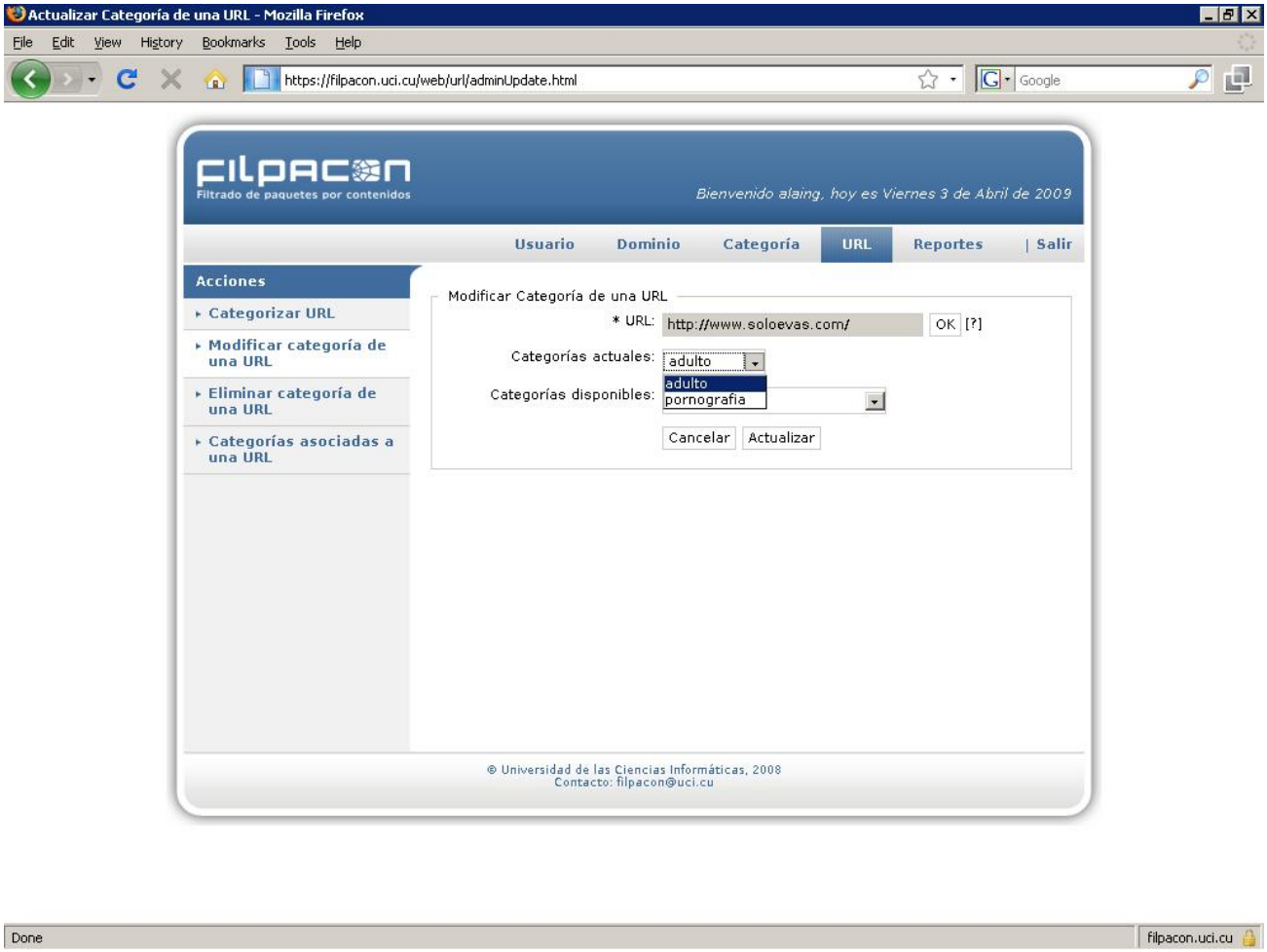


Anexo 14. Módulo de Gestión de Usuarios. Subsistema de Administración FILPACON 1.0

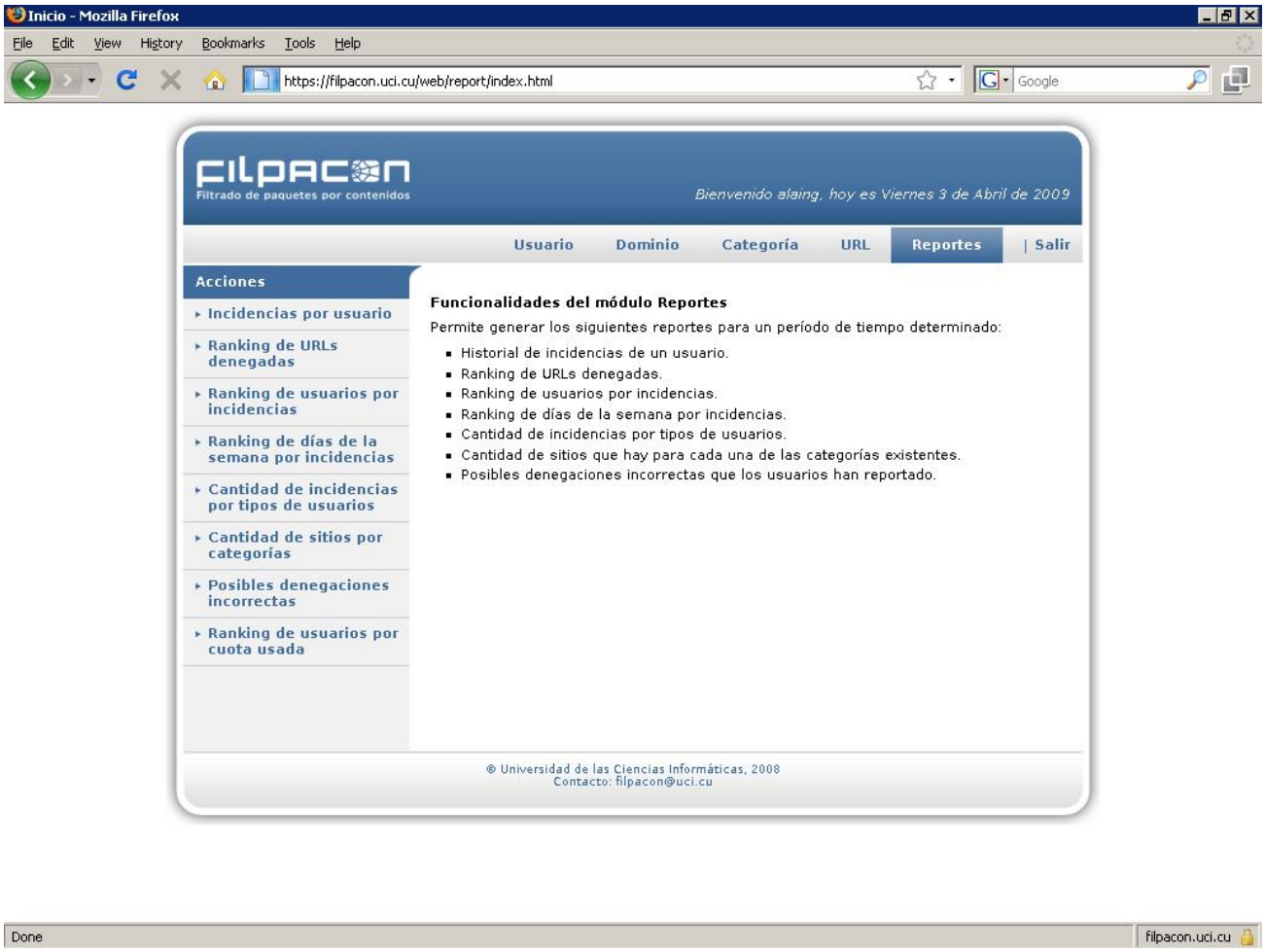
Anexo 15. Módulo de Gestión de Categorías. Subsistema de Administración FILPACON 1.0



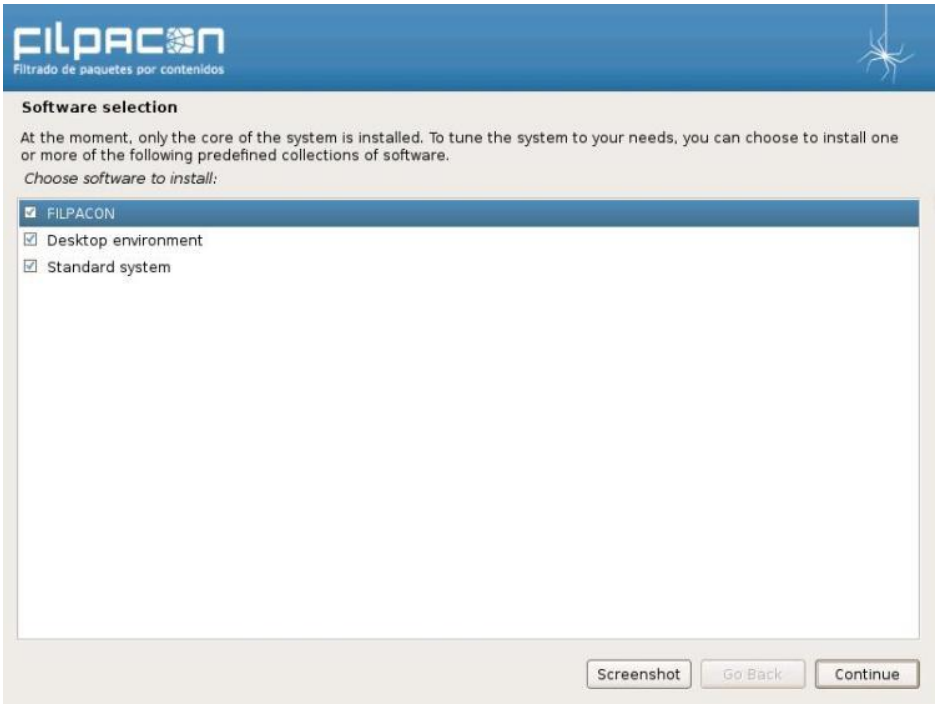
Anexo 16. Módulo de Gestión de URLs. Subsistema de Administración FILPACON 1.0



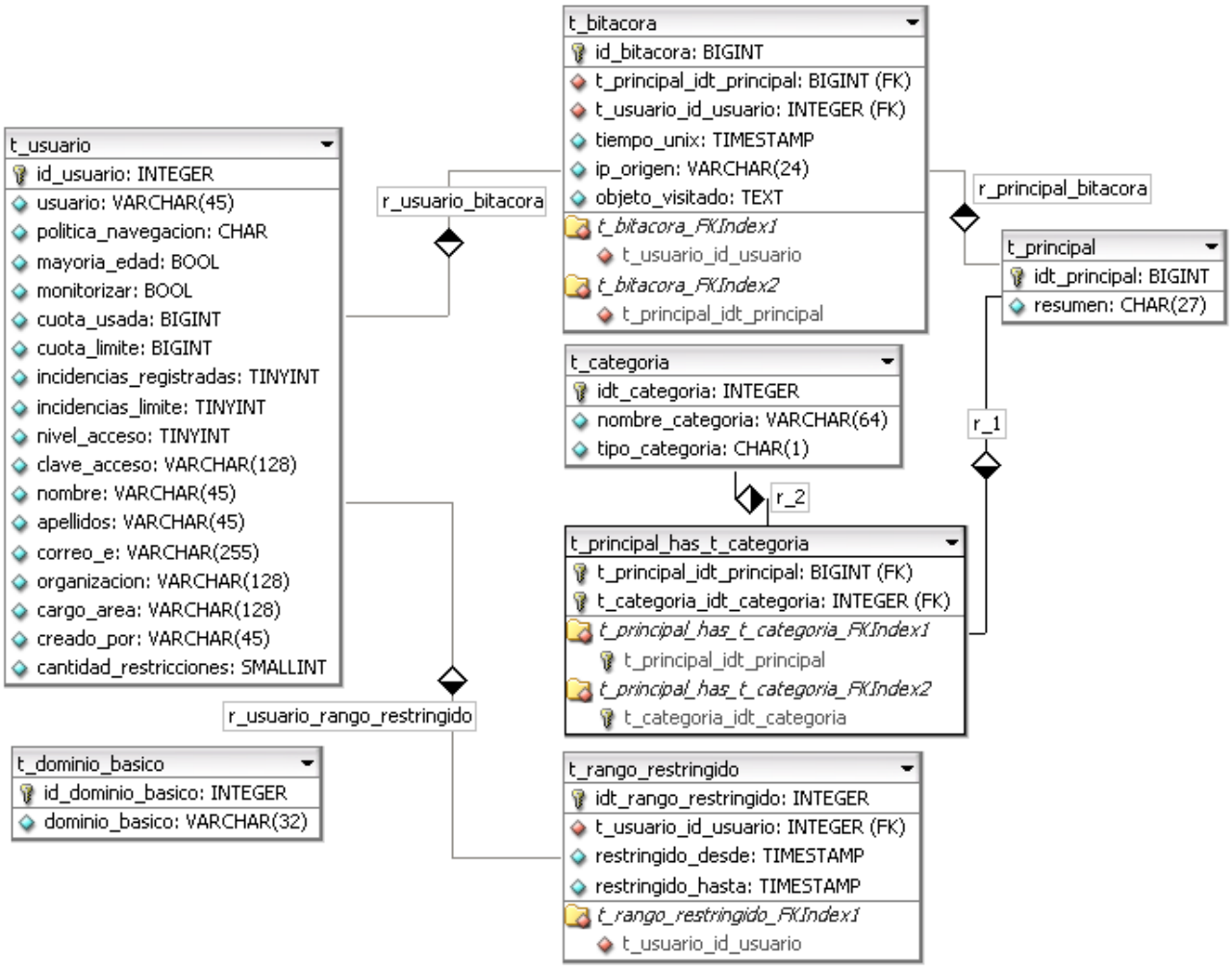
Anexo 17. Módulo de Reportes. Subsistema de Administración FILPACON 1.0



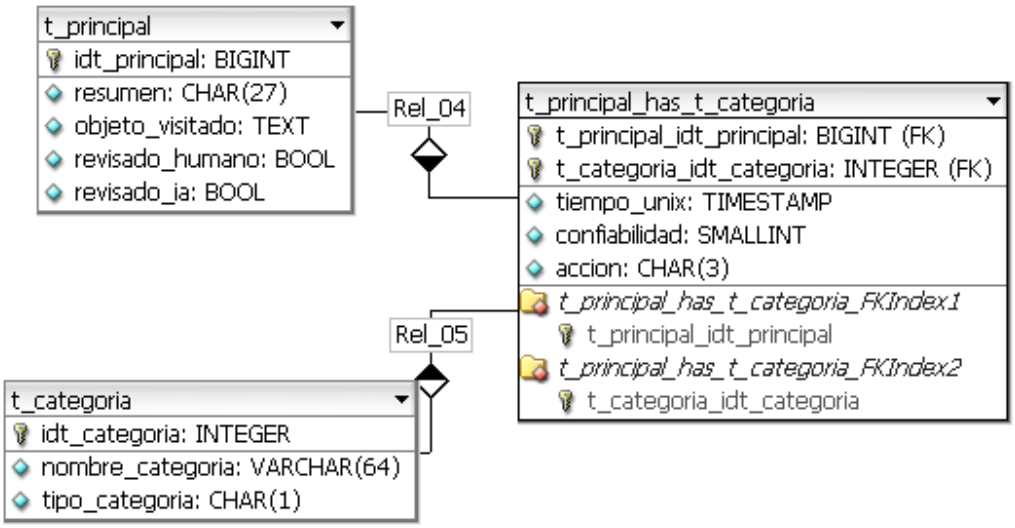
Anexo 18. Subsistema de Instalación. FILPACON 1.0



Anexo 19. Subsistema de Almacenamiento – Base de Datos FILPACON 1.0



Anexo 20. Subsistema de Almacenamiento – Base de Datos Maestra



Anexo 21. Aval de FILPACON 1.0 en el centro UCID.

Universidad de las Ciencias Informáticas, 8 de Abril de 2009
"Año del 50 aniversario del Triunfo de la Revolución"

A: Quien pueda interesar.

AVAL

El sistema informático Filpacon, desarrollado en el Polo Productivo de Soluciones Informáticas para Internet (SINI), perteneciente a la Facultad 10, fue instalado en la UCID y puesto en explotación en enero del 2008 con el objetivo de gestionar el acceso y la navegación en Internet de los más de 500 usuarios del centro.

Los resultados han sido satisfactorios; contribuido a:

1. Aplicar la Política establecida para el uso de Internet en la institución.
2. Realizar estudios y análisis sobre el uso de Internet por los usuarios.
3. Mejorar la productividad de los usuarios, mediante la denegación de recursos de Internet no afines a la institución.
4. Reducir la responsabilidad legal de la institución, mediante la denegación de recursos ilícitos de Internet (por ejemplo: pornografía).
5. Prevenir la invasión de virus, mediante la denegación de recursos contaminados de Internet.
6. Prevenir el robo o fuga de información, mediante la denegación de recursos de Internet que contienen spywares y otros que propician el intercambio de información (foros no afines a la institución, correos electrónicos internacionales, etc.).
7. Mejorar el aprovechamiento del Ancho de Banda de Internet y otros recursos de Red, mediante la denegación de recursos pesados (imágenes, videos, música, etc.) no afines a la institución.

Considero que Filpacon es un producto excelente, útil y con buenas perspectivas.

Atentamente.

Jefe UCID
Tte. Cor.
Raul M. Triana Díaz

Anexo 22. Aval de FILPACON 1.0 por el Decano de la Facultad 10

Universidad de las Ciencias Informáticas (UCI), 31 de Marzo de 2009
"Año del 50 aniversario del Triunfo de la Revolución"

AVAL

El producto Filpacon, desarrollado en el Polo Productivo de Soluciones Informáticas para Internet (SINI), perteneciente a la Facultad 10, ha sido durante mucho tiempo el producto insgnia de dicho Polo Productivo. Se desarrolló a solicitud de la Oficina para la Seguridad de las Redes Informáticas (OSRI) del Ministerio de la Informática y las Comunicaciones (MIC) y se ha instalado en par de ocasiones (4 y 3 meses) para los estudiantes de la Facultad 10, con el objetivo de gestionar el acceso y la navegación en Internet de aproximadamente 1000 usuarios.

Los resultados de ambas pruebas fueron muy satisfactorios, lográndose rendimientos superiores a los tradicionales con los proxies que usualmente utiliza la UCI. Durante los períodos de prueba permitió: (1) un mejor aprovechamiento del ancho de banda de Internet y recursos de Red, mediante la denegación de recursos pesados y ajenos a los fines de la UCI; (2) reducir la responsabilidad legal, mediante la denegación de contenidos ilícitos de Internet; (3) reducir la pérdida de productividad, mediante la denegación de contenidos ajenos a los fines de la UCI y (4) obtener una muy valiosa información estadística de la navegación, muy útil para la toma de decisiones.

Una versión comercial de Filpacon fue instalada en Venezuela, comercializada a la Fundación INFOCENTROS a través del Centro Nacional de Tecnologías de Información (CNTI) del Ministerio de Telecomunicaciones e Informática como parte del Convenio Cuba-Venezuela en el marco de las Comisiones Intergubernamentales mixtas de los dos países. De igual manera los resultados en los INFOCENTROS de Venezuela han sido muy estimulantes y útiles para aquella organización pues además el producto, debido a la posibilidad de personalización que tiene, se diseñó a la medida de cada INFOCENTRO considerando las características regionales que tiene aquel país.

Considero que Filpacon es un producto excelente, con un nivel de utilidad muy elevado.

Respetuosamente,

Msc. Héctor Rodríguez Figueredo
Decano
Facultad 10, UCI

Carretera a San Antonio de los Baños. Km 5 ½ Reparto Torrens. Boyeros. Ciudad Habana.
Teléfono (53 7) 8372519 e-mail: decano.f10@uci.cu

Anexo 23. Premio del RECTOR 2009 al Mejor Producto

2009

PREMIO DEL RECTOR

El Consejo Universitario
de la Universidad de las Ciencias Informáticas
otorga el presente premio a:

Filpacon

en la categoría de

Mejor Producto

"El deber debe cumplirse sencilla y naturalmente".

José Martí

UCI
Universidad de las Ciencias Informáticas

Se entrega en La Habana, a los 19 días del mes de febrero de 2009

Anexo 24. Aval de FILPACON 1.0 por la Dirección de Redes y Seguridad Informática de la UCI

Universidad de las Ciencias Informáticas (UCI), 6 de Abril de 2009

"Año del 50 aniversario del Triunfo de la Revolución"

AVAL

El producto Filpacon, desarrollado en el Polo Productivo de Soluciones Informáticas para Internet (SINI), perteneciente a la Facultad 10, fue instalado en el Nodo Central de la UCI y estuvo en uso durante aproximadamente 3 meses para gestionar el acceso y la navegación en Internet de aproximadamente 2000 estudiantes de las facultades 1 y 10.

Los resultados fueron satisfactorios; entre otros beneficios, durante el período de pruebas contribuyó a:

1. Aplicar la Política Aceptable de Uso de Internet de la institución.
2. Obtener valiosos reportes estadísticos para conocer el uso que de Internet hacían los usuarios.
3. Mejorar la productividad de los usuarios, mediante la denegación de recursos de Internet no afines a la institución.
4. Reducir la responsabilidad legal de la institución, mediante la denegación de recursos ilícitos de Internet (por ejemplo: pornografía).
5. Prevenir la invasión de virus, mediante la denegación de recursos contaminados de Internet.
6. Prevenir el robo o fuga de información, mediante la denegación de recursos de Internet que contienen spywares y otros que propician el intercambio de información (foros no afines a la institución, correos electrónicos internacionales, etc.).
7. Mejorar el aprovechamiento del Ancho de Banda de Internet y otros recursos de Red, mediante la denegación de recursos pesados (imágenes, videos, música, etc.) no afines a la institución.
8. Identificar posibles mejoras al software para su versión 2.0.

Considero que Filpacon es un producto excelente, útil y con buenas perspectivas; debe convertirse, dentro de algún tiempo, en el Sistema de Filtrado de Contenido de Internet que se utilice para los usuarios de la UCI.

Atentamente,

UCI VICERRECTORIA DE TECNOLOGÍA

Raydel Montesino Perurena
Director de Redes y Seguridad Informática
Universidad de las Ciencias Informáticas

Carretera a San Antonio de los Baños. Km 5 ½ Reparto Torrens. Boyeros. Ciudad Habana.
Teléfono (53 7) 8358059 e-mail: raydelmp@uci.cu